

EVE 161: Microbial Phylogenomics

Era IV: Genomes from the Environment

- Term first use in 1998 paper by Handelsman et al. Chem Biol. 1998 Oct;5(10):R245-9. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products
- Collective genomes of microbes in the soil termed the “soil metagenome”
- Good review is Metagenomics: Application of Genomics to Uncultured Microorganisms by Handelsman Microbiol Mol Biol Rev. 2004 December; 68(4): 669–685.



Bacterial Rhodopsin: Evidence for a New Type of Phototrophy in the Sea

Oded Bèjà, *et al.*

Science **289**, 1902 (2000);

DOI: 10.1126/science.289.5486.1902

<http://www.sciencemag.org/content/289/5486/1902.full>

.....

**Proteorhodopsin phototrophy
in the ocean**

**Oded Bèjà^{*†}, Elena N. Spudich^{†‡}, John L. Spudich[‡], Marion Leclerc^{*}
& Edward F. DeLong^{*}**

<http://www.nature.com/nature/journal/v411/n6839/abs/411786a0.html>



Bacterial Rhodopsin: Evidence for a New Type of Phototrophy in the Sea

Oded Bèjà, *et al.*

Science **289**, 1902 (2000);

DOI: 10.1126/science.289.5486.1902

<http://www.sciencemag.org/content/289/5486/1902.full>

Extremely halophilic archaea contain retinal-binding integral membrane proteins called bacteriorhodopsins that function as light-driven proton pumps. So far, bacteriorhodopsins capable of generating a chemiosmotic membrane potential in response to light have been demonstrated only in halophilic archaea. We describe here a type of rhodopsin derived from bacteria that was discovered through genomic analyses of naturally occurring marine bacterioplankton. The bacterial rhodopsin was encoded in the genome of an uncultivated - proteobacterium and shared highest amino acid sequence similarity with archaeal rhodopsins. The protein was functionally expressed in *Escherichia coli* and bound retinal to form an active, light-driven proton pump. The new rhodopsin exhibited a photochemical reaction cycle with intermediates and kinetics characteristic of archaeal proton-pumping rhodopsins. Our results demonstrate that archaeal-like rhodopsins are broadly distributed among different taxa, including members of the domain *Bacteria*. Our data also indicate that a previously unsuspected mode of bacterially mediated light-driven energy generation may commonly occur in oceanic surface waters worldwide.

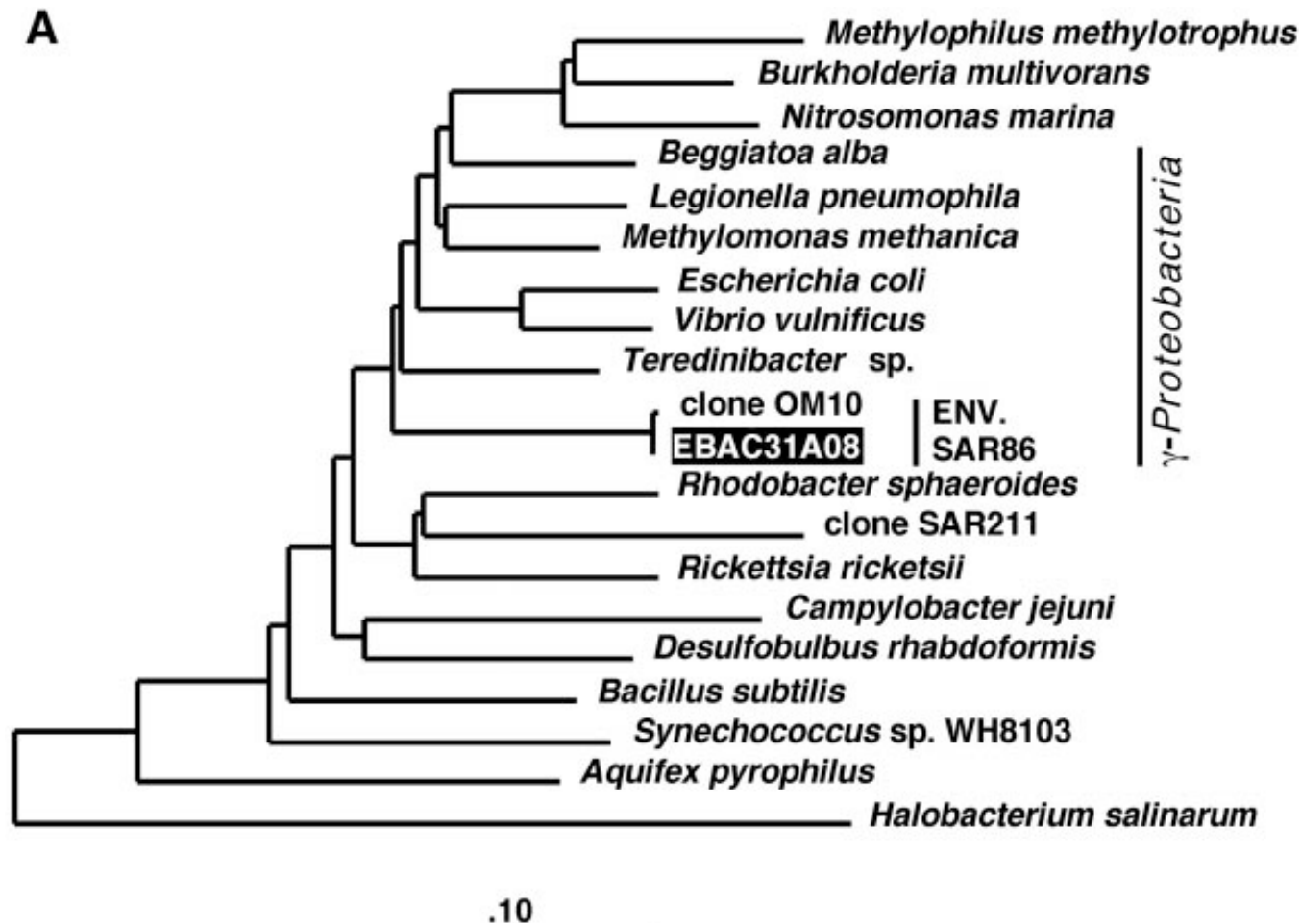


Fig. 1. (A) Phylogenetic tree of bacterial 16S rRNA gene sequences, including that encoded on the 130-kb bacterioplankton BAC clone (EBAC31A08) (16).

The 16S ribosomal RNA neighbor-joining tree was constructed on the basis of 1256 homologous positions by the “neighbor” program of the PHYLIP package [J. Felsenstein, *Methods Enzymol.* **266**, 418 (1996)]. DNA distances were calculated with the Kimura’s 2-parameter method.

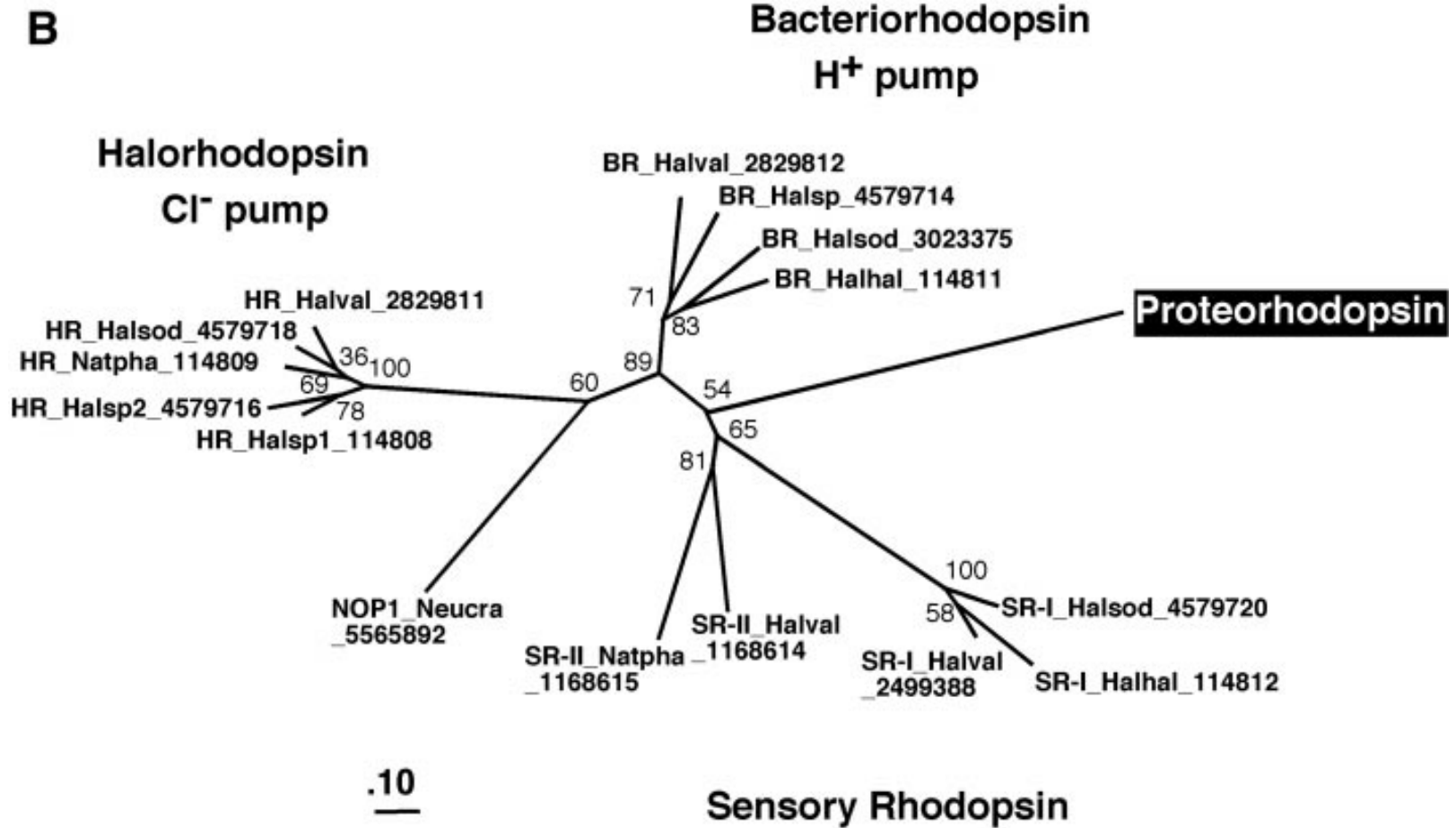


Fig. 1. (B) Phylogenetic analysis of proteorhodopsin with archaeal (BR, HR, and SR prefixes) and *Neurospora crassa* (NOP1 prefix) rhodopsins (16). Nomenclature: Name_Species.abbreviation_Genbank.gi (HR, halorhodopsin; SR, sensory rhodopsin; BR, bacteriorhodopsin). Halsod, *Halorubrum sodomense*; Halhal, *Halobacterium salinarum* (halobium); Halval, *Haloarcula vallismortis*; Natpha, *Natronomonas pharaonis*; Halsp, *Halobacterium* sp; Neucra, *Neurospora crassa*.

The proteorhodopsin tree was constructed on the basis of the archaeal rhodopsin alignment used by Mukohata *et al.* [Y. Mukohata, K. Ihara, T. Tamura, Y. Sugiyama, *J. Bio-chem. (Tokyo)* **125**, 649 (1999)] and the neighbor-joining method as implemented in the neighbor program of the PHYLIP package. The numbers at forks indicate the percent of bootstrap replications (out of 1000) in which the given branching was observed. The least-square method (the Kitsch program of PHYLIP) and the maximum likelihood method (the Puzzle program) produced trees with essentially identical topologies, whereas the protein parsimony method (the Protpars program of PHYLIP) placed proteorhodopsin within the sensory rhodopsin cluster. The Yro/Hsp30 subfamilies sequences (6) were omitted from the phylogenetic analysis to avoid the effect of long branch attraction.

Cloning of proteorhodopsin. Sequence analysis of a 130-kb genomic fragment that encoded the ribosomal RNA (rRNA) operon from an uncultivated member of the marine *g-Proteobacteria* (that is, the “SAR86” group) (8, 9) (Fig. 1A) also revealed an open reading frame (ORF) encoding a putative rhodopsin (referred to here as proteorhodopsin) (10).

Cloning of proteorhodopsin. Sequence analysis of a 130-kb genomic fragment that encoded the ribosomal RNA (rRNA) operon from an uncultivated member of the marine *g-Proteobacteria* (that is, the “SAR86” group) (8, 9) (Fig. 1A) also revealed an open reading frame (ORF) encoding a putative rhodopsin (referred to here as proteorhodopsin) (10).

8. T. D. Mullins, T. B. Britcshgi, R. L. Krest, S. J. Giovannoni, *Limnol. Oceanogr.* **40**, 148 (1995).
9. O. Béjà *et al.*, *Environ. Microbiol.*, in press.



Association for the Sciences of
Limnology and Oceanography

[About ASLO](#)

[Members](#)

[Awards](#)

[Publications](#)

[Meetings](#)

[Employment](#)

[Activities](#)

[Search](#)

[Join!](#)

Genetic comparisons reveal the same unknown bacterial lineages in Atlantic and Pacific bacterioplankton communities

MULLINS, THOMAS D., THERESA B. BRITSCHGI, ROBIN L. KREST, STEPHEN J. GIOVANNONI

Limnol. Oceanogr., 40(1), 1995, 148-158 | DOI: 10.4319/lo.1995.40.1.0148

ABSTRACT: This article is in Free Access Publication and may be downloaded using the "Download Full Text PDF" link at right.

Article Links

[Download Full-text PDF](#)

[Return to Table of Contents](#)

Please Note

Articles in L&O appear in PDF format. Open access articles may be freely downloaded by anyone. Other articles are available for download to subscribers only, or may be purchased for \$10 per article. All L&O articles are moved into Open Access after three years.

Genetic comparisons reveal the same unknown bacterial lineages in Atlantic and Pacific bacterioplankton communities

Thomas D. Mullins, Theresa B. Britschgi, Robin L. Krest, and Stephen J. Giovannoni¹

Department of Microbiology, Oregon State University, Corvallis 97331

Abstract

The phylogenetic diversity of oligotrophic bacterioplankton communities was compared with 16S ribosomal RNA genes cloned from natural populations. The data reported here extend a previous analysis of a bacterioplankton 16S rRNA clone library with 15 additional nucleic acid clone sequences, to provide information on 60 16S rDNA clones from hydrostation S in the Sargasso Sea. The data were compared to partial sequences of 37 *Bacterial* 16S rDNA clones reported from a surface picoplankton population collected at the Aloha station in the North Pacific gyre, and partial sequences of 29 *Bacterial* 16S rRNA clones obtained from sites near Bermuda and the western California Current. The results support reports of diverse groups of previously unknown α -proteobacteria, γ -proteobacteria, and cyanobacteria in oceanic surface samples. Three novel lineages (SAR121, 125, 145) of proteobacteria were found. Several genes cloned from the Sargasso Sea were nearly identical to genes cloned from the Pacific samples, suggesting that these previously unrecognized bacteria groups (SAR11, SAR122, SAR86) are distributed widely in the surface waters of subtropical oceans. Two gene clones closely matched nucleotide sequences from the cultivated bacterial species *Photobacterium phosphoreum* and *Alteromonas haloplanktis*.

- rRNA PCR studies of marine microbes have been extensive
- Comparative analysis had revealed many lineages, some very novel, some less so, that were dominant in many, if not all, open ocean samples
- Lineages given names based on specific clones: e.g., SAR11, SAR86, etc

Genetic diversity in Sargasso Sea bacterioplankton

**Stephen J. Giovannoni, Theresa B. Britschgi,
Craig L. Moyer & Katharine G. Field**

<http://www.nature.com/nature/journal/v345/n6270/abs/345060a0.html>

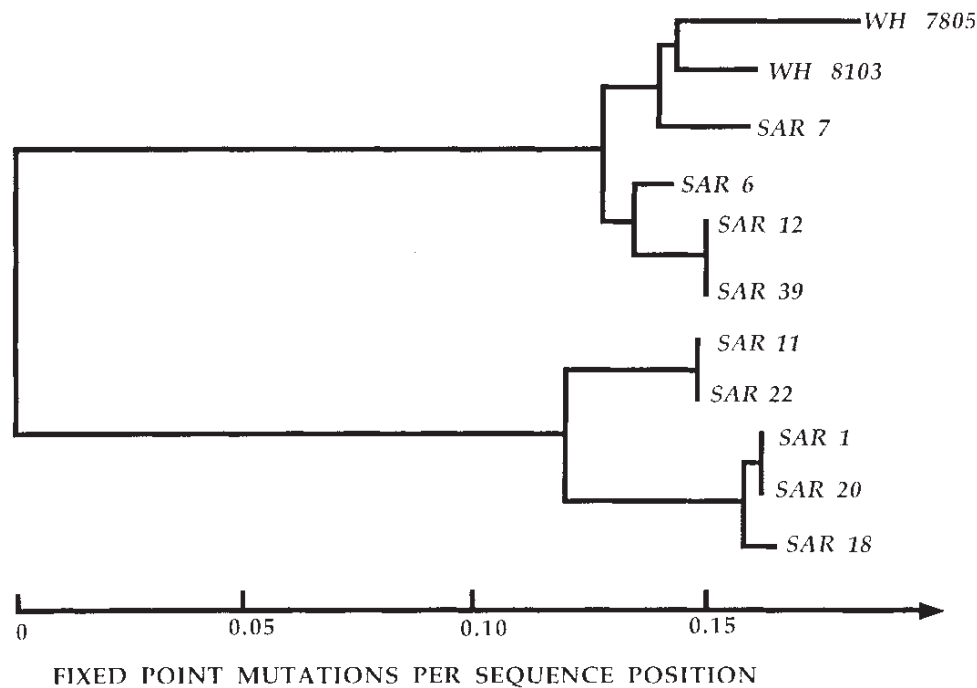


FIG. 1 Unrooted phylogenetic tree depicting relationships among Sargasso Sea bacterioplankton 16S rRNAs. The analysis includes the 16S rRNA sequences of two cultivated strains of marine cyanobacteria, WH8103 and WH7805. The sequences of clones SAR67 and SAR78 were identical to those of SAR1 and SAR20; SAR77 differed at one position (not shown). A total of 230 positions located at the 5' region of the gene, including both conserved and hypervariable domains, were included within the analysis (*E. coli* positions 101–344). Phylogenetic trees were constructed by a distance matrix method^{23,24,25}.

METHODS. The picoplankton samples were collected in April by tangential flow filtration on Durapore 0.1 μ m fluorocarbon membranes from a depth of 1–2 m at hydrostation S (32°4' N 64°23' W). The small subunit ribosomal RNA genes were amplified from bulk picoplankton DNA by a modification of the polymerase chain reaction²⁶. The reaction conditions were: 1 μ g template DNA; 2' at 94 °C, 2' at 37 °C, 7' at 72 °C; 30 cycles. The amplified genes were cloned as *Bam*HI/*Pst*I fragments into M13 phage mp18 and sequenced twice, once with inosine substituting for guanosine, using the dideoxy chain termination method²². The amplification primers (OX1 and OX2) were designed with a bias towards the cyanobacterial phylum of the eubacteria. Primers OX1 and OX2 are, respectively, 90% and 82% similar to eubacterial consensus sequences. The sequences of the amplification primers are: OX1, GTGCTGCAG**AGAGT**TYGATCCTGGCTAGG; OX2, CACGGATCCA**AAGGAGGT-GATCCANCCNCACC**, where the domain complementary to the coding region is indicated in bold.

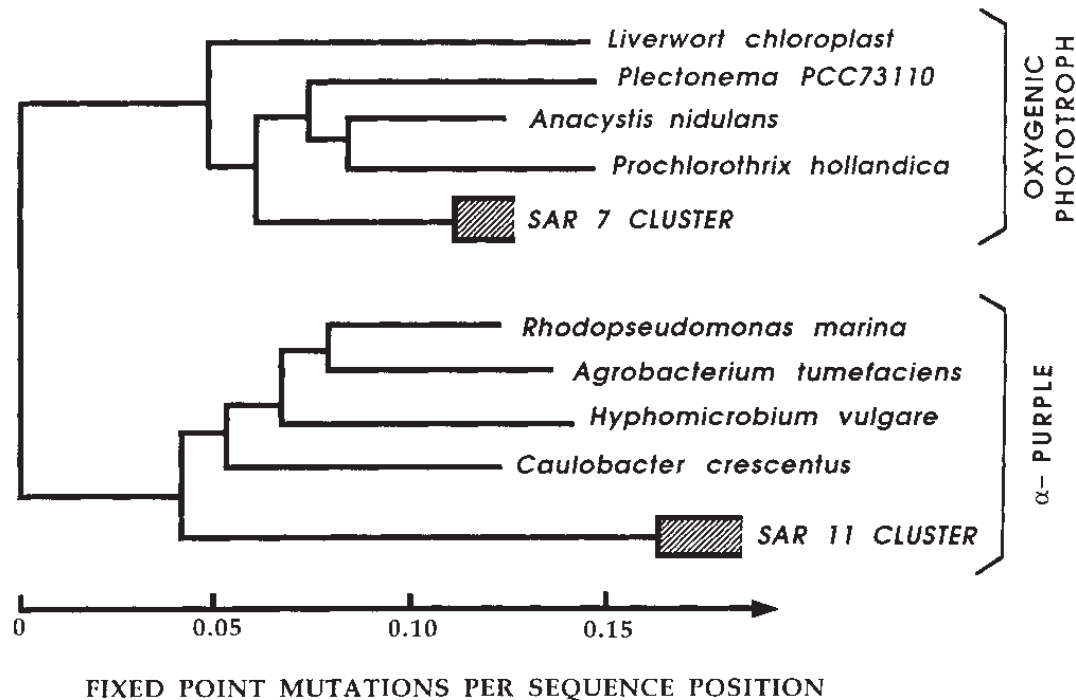


FIG. 3 Phylogenetic relationships of SAR7 and SAR11 16S rDNA sequence clusters to a collection of 16S rRNA sequences representing the oxygenic phototroph^{8,28} and α -purple eubacterial phyla²⁹. Four clones were sequenced completely (SAR6, SAR7, SAR1 and SAR11) and used for the inference of distant relationships. The tree was rooted using the sequences of *Bacillus subtilis* and *Heliobacterium chlorum*^{8,30}. The analysis was restricted to 900 sequence positions. Regions of uncertain homology between phyla, including hypervariable domains, were excluded from this analysis. Hence, the variability within the clusters (indicated by the hatched boxes) is about 0.01 similarity units less in this figure than in Fig. 1. The 3'-terminal domain of the 16S rRNAs was excluded from the analysis because of an internal *Bam*HI restriction site in clones SAR1 and SAR11 at position 1,190.

Phylogenetic Analysis of a Natural Marine Bacterioplankton Population by rRNA Gene Cloning and Sequencing

THERESA B. BRITSCHGI AND STEPHEN J. GIOVANNONI*

Department of Microbiology, Oregon State University, Corvallis, Oregon 97331

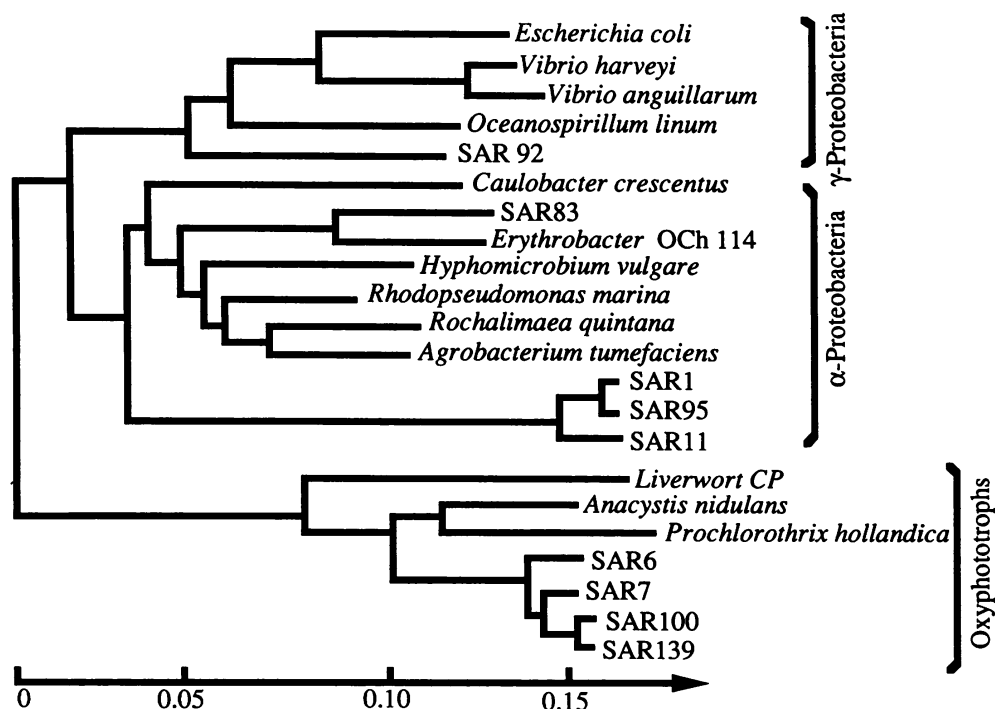


FIG. 4. Phylogenetic tree showing relationships of the rDNA clones from the Sargasso Sea to representative, cultivated species (2, 6, 18, 30, 31, 32, 35, 40). Positions of uncertain homology in regions containing insertions and deletions were omitted from the analysis. Evolutionary distances were calculated by the method of Jukes and Cantor (15), which corrects for the effects of superimposed mutations. The phylogenetic tree was determined by a distance matrix method (20). The tree was rooted with the sequence of *Bacillus subtilis* (38). Sequence data not referenced were provided by C. R. Woese and R. Rossen.

JOURNAL OF BACTERIOLOGY, July 1991, p. 4371–4378
0021-9193/91/144371-08\$02.00/0
Copyright © 1991, American Society for Microbiology

Vol. 173, No. 14

Analysis of a Marine Picoplankton Community by 16S rRNA Gene Cloning and Sequencing

THOMAS M. SCHMIDT,[†] EDWARD F. DELONG,[‡] AND NORMAN R. PACE*

*Department of Biology and Institute for Molecular and Cellular Biology, Indiana University,
Bloomington, Indiana 47405*

Received 7 January 1991/Accepted 13 May 1991

Molecular diversity and ecology of microbial plankton

Stephen J. Giovannoni¹ & Ulrich Stingl¹

The history of microbial evolution in the oceans is probably as old as the history of life itself. In contrast to terrestrial ecosystems, microorganisms are the main form of biomass in the oceans, and form some of the largest populations on the planet. Theory predicts that selection should act more efficiently in large populations. But whether microbial plankton populations harbour organisms that are models of adaptive sophistication remains to be seen. Genome sequence data are piling up, but most of the key microbial plankton clades have no cultivated representatives, and information about their ecological activities is sparse.

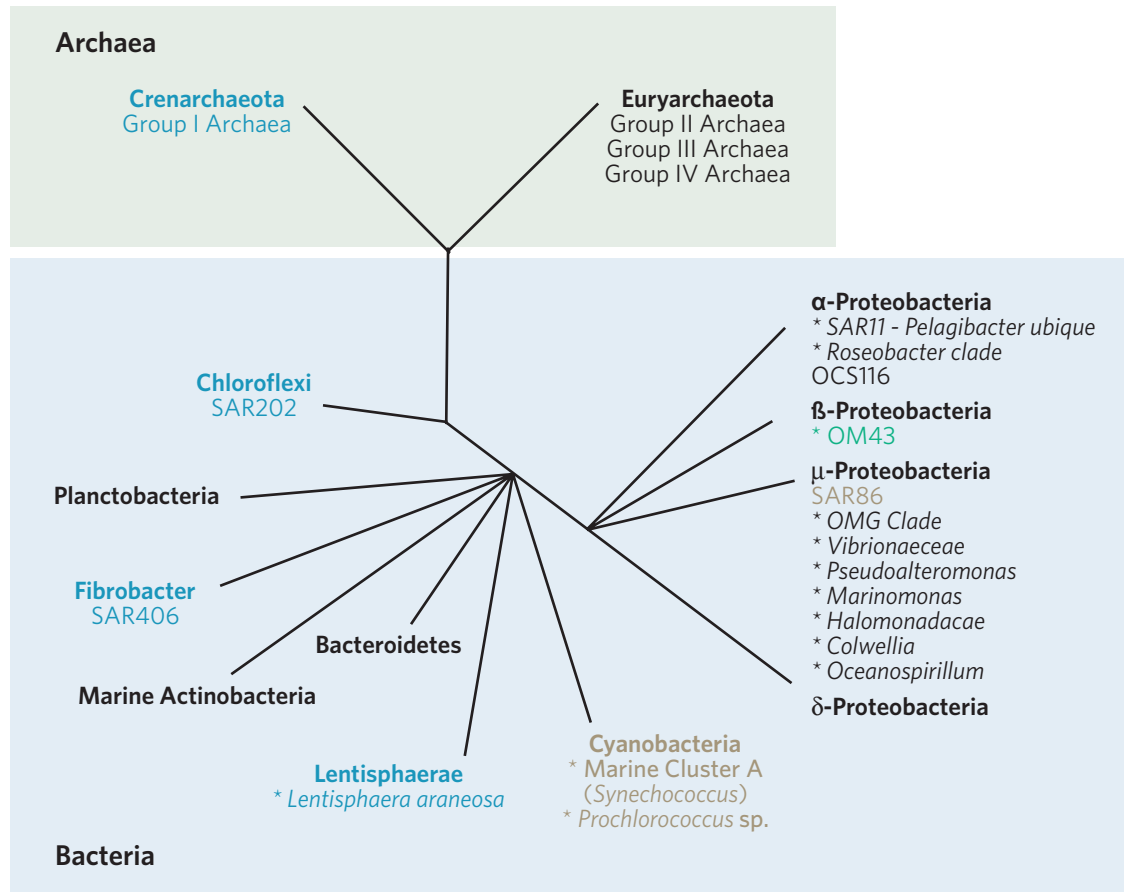
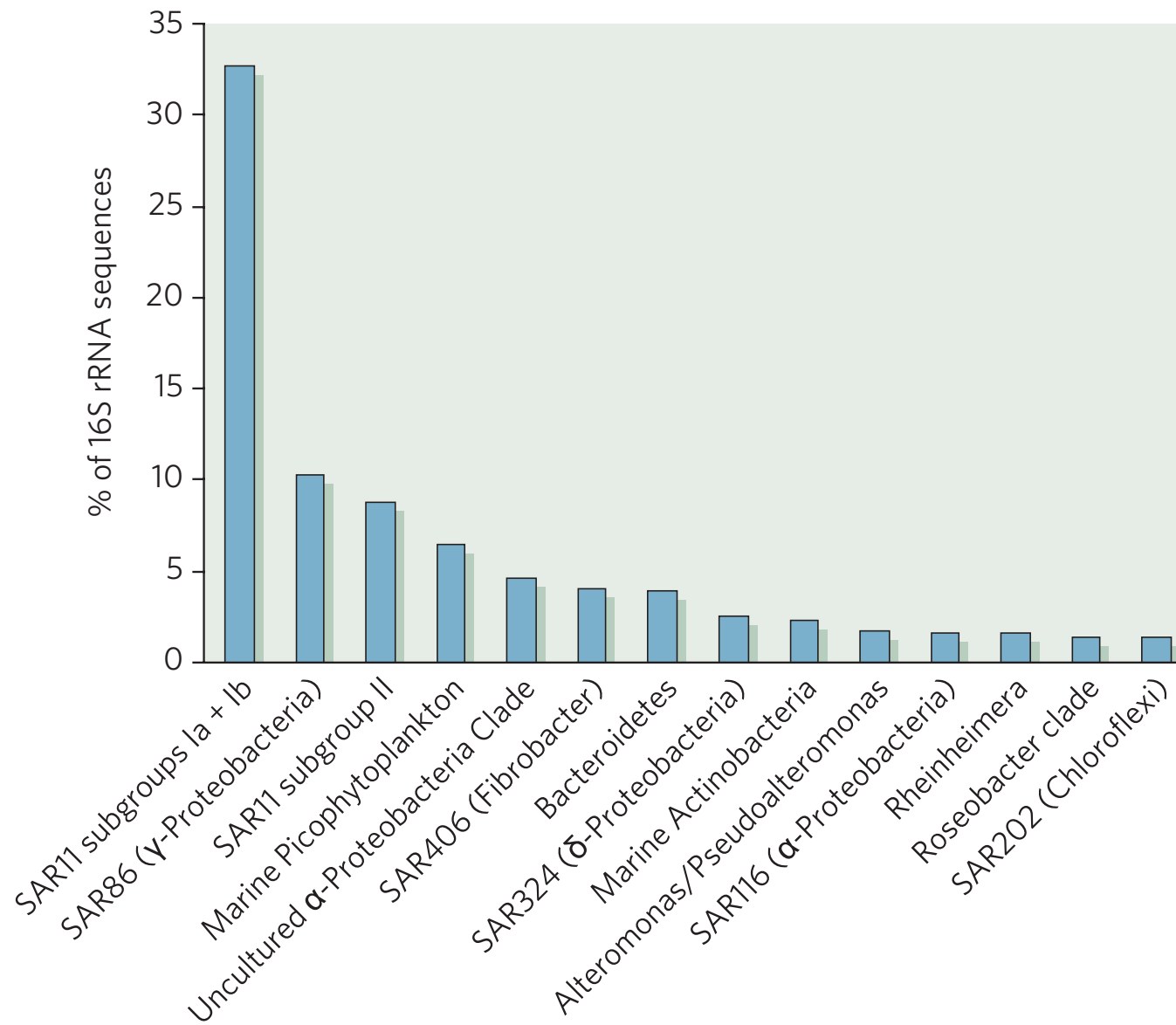


Figure 1 | Schematic illustration of the phylogeny of the major plankton clades. Black letters indicate microbial groups that seem to be ubiquitous in seawater. Gold indicates groups found in the photic zone. Blue indicates groups confined to the mesopelagic and surface waters during polar winters. Green indicates microbial groups associated with coastal ocean ecosystems.



- Studying Sar86 and other marine plankton
- Note - published one of first genomic studies of uncultured microbes - in 1996

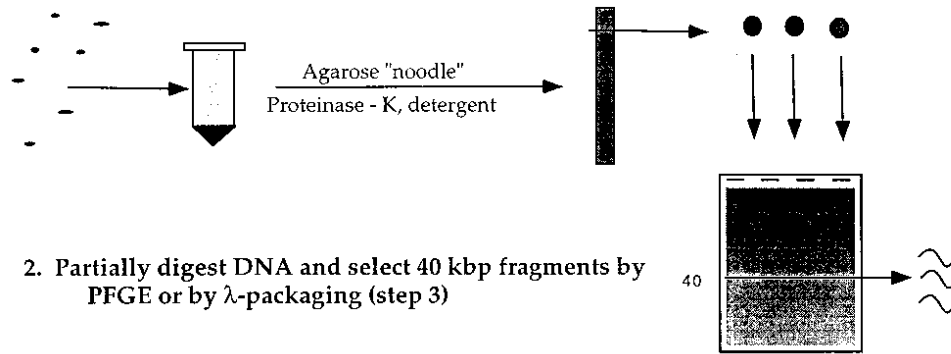
JOURNAL OF BACTERIOLOGY, Feb. 1996, p. 591–599
0021-9193/96/\$04.00+0
Copyright © 1996, American Society for Microbiology

Vol. 178, No. 3

Characterization of Uncultivated Prokaryotes: Isolation and Analysis of a 40-Kilobase-Pair Genome Fragment from a Planktonic Marine Archaeon

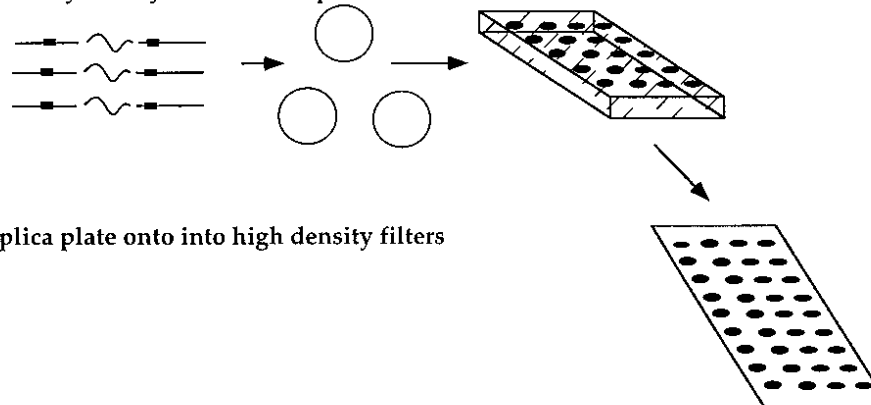
JEFFEREY L. STEIN,^{1*} TERENCE L. MARSH,² KE YING WU,³ HIROAKI SHIZUYA,⁴
AND EDWARD F. DELONG^{3*}

1. Concentrate bacteria, digest protein and preserve high MW DNA



2. Partially digest DNA and select 40 kbp fragments by PFGE or by λ -packaging (step 3)

3. Ligate to fosmid arms, package and transfect to *E. coli*.
Array library in microtiter plates.



4. Replica plate onto into high density filters

5. Probe and "walk" to identify contiguous fragments

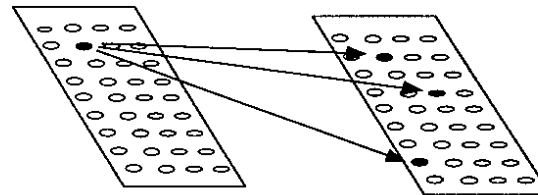


FIG. 1. Flowchart depicting the construction and screening of an environmental library from a mixed picoplankton sample. MW, molecular weight; PFGE, pulsed-field gel electrophoresis.

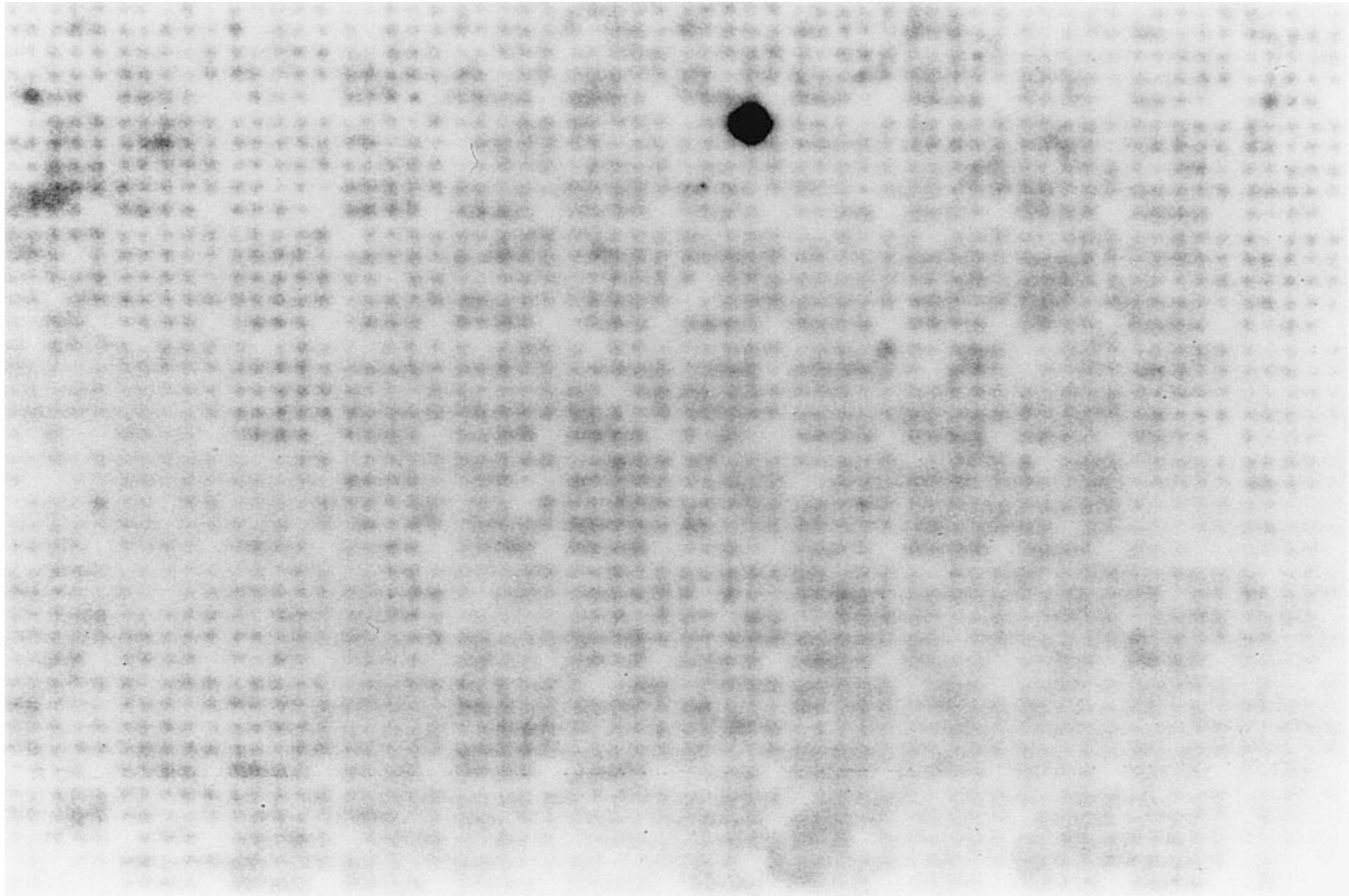


FIG. 4. High-density filter replica of 2,304 fosmid clones containing approximately 92 million bp of DNA cloned from the mixed picoplankton community. The filter was probed with the labeled insert from clone 4B7 (dark spot). The lack of other hybridizing clones suggests that contigs of 4B7 are absent from this portion of the library. Similar experiments with the remainder of the library yielded similar results.

9. Chojas et al., *Environ. Microbiol.*, in press.
10. A subclone shotgun library was constructed from BAC clone 31A08, and subclones were sequenced in both directions on the MegaBACE 1000 capillary array electrophoresis DNA sequencing instrument (Molecular Dynamics, Sunnyvale, CA). The contiguous sequence was assembled with SEQUENCHER 3.1.1 software (Gene Codes, Ann Arbor, MI). The sequence of the proteorhodopsin-containing contig has been deposited in GenBank under accession number AF279106.

10. A subclone shotgun library was constructed from BAC clone 31A08, and subclones were sequenced in both directions on the MegaBACE 1000 capillary array electrophoresis DNA sequencing instrument (Molecular Dynamics, Sunnyvale, CA). The contiguous sequence was assembled with SEQUENCHER

3.1.1 software (Gene Codes, Ann Arbor, MI). The sequence of the proteorhodopsin-containing contig has been deposited in GenBank under accession number AF279106.

GenBank

Nucleotide ▾

AF279106



Search

GenBank ▾ Submit ▾ Genomes ▾ WGS ▾ HTGs ▾ EST/GSS ▾ Metagenomes ▾ TPA ▾ TSA ▾ INSDC ▾

GenBank Overview

What is GenBank?

GenBank[®] is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences ([Nucleic Acids Research](#), 2013 Jan;41(D1):D36-42). GenBank is part of the [International Nucleotide Sequence Database Collaboration](#), which comprises the DNA DataBank of Japan (DDBJ), the European Molecular Biology Laboratory (EMBL), and GenBank at NCBI. These three organizations exchange data on a daily basis.

The complete [release notes](#) for the current version of GenBank are available on the NCBI ftp site. A new release is made every two months. GenBank growth [statistics](#) for both the traditional GenBank divisions and the WGS division are available from each release.

An [annotated sample GenBank record](#) for a *Saccharomyces cerevisiae* gene demonstrates many of the features of the GenBank flat file format.

GenBank Resources

[GenBank Home](#)

[Submission Types](#)

[Submission Tools](#)

[Search GenBank](#)

[Update GenBank Records](#)

GenBank

[GenBank](#) ▾
 [Submit](#) ▾
 [Genomes](#) ▾
 [WGS](#) ▾
 [HTGs](#) ▾
 [EST/GSS](#) ▾
 [Metagenomes](#) ▾
 [TPA](#) ▾
 [TSA](#) ▾
 [INSDC](#) ▾

GenBank Overview

What is GenBank?

GenBank[®] is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences (*Nucleic Acids Research*, 2013 Jan;41(D1):D36-42). GenBank is part of the [International Nucleotide Sequence Database Collaboration](#), which comprises the DNA DataBank of Japan (DDBJ), the European Molecular Biology Laboratory (EMBL), and GenBank at NCBI. These three organizations exchange data on a daily basis.

The complete [release notes](#) for the current version of GenBank are available on the NCBI ftp site. A new release is made every two months. GenBank growth [statistics](#) for both the traditional GenBank divisions and the WGS division are available from each release.

An [annotated sample GenBank record](#) for a *Saccharomyces cerevisiae* gene demonstrates many of the features of the GenBank flat file format.

GenBank Resources

[GenBank Home](#)

[Submission Types](#)

[Submission Tools](#)

[Search GenBank](#)

[Update GenBank Records](#)

GenBank ▾

Uncultured marine gamma proteobacterium EBAC31A08 BAC sequence

GenBank: AF279106.2


[FASTA](#) [Graphics](#)

[Go to:](#) ☐

LOCUS AF279106 128758 bp DNA linear ENV 15-NOV-2007
DEFINITION Uncultured marine gamma proteobacterium EBAC31A08 BAC sequence.
ACCESSION AF279106
VERSION AF279106.2 GI:34112904
DBLINK BioProject: [PRJNA27781](#)
KEYWORDS ENV.
SOURCE uncultured marine gamma proteobacterium EBAC31A08
ORGANISM [uncultured marine gamma proteobacterium EBAC31A08](#)
Bacteria; Proteobacteria; Gammaproteobacteria; environmental
samples.
REFERENCE 1 (bases 1 to 128758)
AUTHORS Beja,O., Aravind,L., Koonin,E.V., Suzuki,M.T., Hadd,A.,
Nguyen,L.P., Jovanovich,S.B., Gates,C.M., Feldman,R.A.,
Spudich,J.L., Spudich,E.N. and DeLong,E.F.
TITLE Bacterial rhodopsin: evidence for a new type of phototrophy in the
sea

Uncultured proteobacterium EBAC31A08 clone BAC EBAC31A08, complete sequence

GenBank: AF279106.1

 This sequence has been updated. [See current version.](#)

[FASTA](#) [Graphics](#)

Go to: 

LOCUS	AF279106	105184 bp	DNA	linear	BCT 23-OCT-2000
DEFINITION	Uncultured proteobacterium EBAC31A08 clone BAC EBAC31A08, complete sequence.				
ACCESSION	AF279106				
VERSION	AF279106.1 GI:9971876				
KEYWORDS	.				
SOURCE	uncultured proteobacterium EBAC31A08				
ORGANISM	uncultured proteobacterium EBAC31A08 Bacteria; Proteobacteria; environmental samples.				
REFERENCE	1 (bases 1 to 105184)				
AUTHORS	Beja,O., Aravind,L., Koonin,E.V., Suzuki,M.T., Hadd,A., Nguyen,L.P., Jovanovich,S.B., Gates,C.M., Feldman,R.A., Spudich,J.L., Spudich,E.N. and DeLong,E.F.				
TITLE	Bacterial rhodopsin: evidence for a new type of phototrophy in the sea				

1 ttgttatato agtaattggct attgctocaa taacttaata ctaatatata attagtttat
61 gaataaattt tatatatitt ggttattgtt ttttacacta aatgcatttt cttgctocaga
121 tctcttaagt acagacatga gagttcttga ttccgcctgag tcaagaaaacc ttgpgaggtt
181 tgaaggaana gctttactag ttgtgaagt tgcuaataga tvggtttaca ctttatcaata
241 tgetggtcct caaagattat algaagatta taaagatgaa gattttctag taattgggat
301 cccatctaga gattttcttc aagaatacto tgatgaaagc gatgttgcag aattttgttc
361 tacagaatac ggtgttgaat ttccatgtt ctcaactgct aaagtcaaac gaataaaagc
421 acocccattt tataaaaaac ttattgcaga atcaagttttt actccctcat ggaactttaa
481 taaactacta atctcaaaag agggcaaggt tgtatccaca tatggatcaa agytaaaagc
541 tgaatcaaaa gagtttatat cagctataga agcgttgctg taaatattt acttaagaa
601 taatacagtt ttagctgtgt ttgctgcaaa tattccatta tctacaactc caggaatttt
661 attaatcaaa gcttccattt cagtggggtt tgaatatcc atattagaga tatctaaat
721 gtgattacct ttgttgttta taaatccagt tctatatgtt ggtattccac cgtacagagt
781 tatttttctt gcaacaaggc tctacttcto aggtatccac totataggca ttggaaaagc
841 tcccaaaaga ttaaccatct ttgactgato aactatacat ataaactcgt tagaggcaga
901 agcaactatc ttcttctctg tatgtggcgc accacacact ttaataagac aattttcag
961 agacacctca tctgcacctat ctatgtaata agctatatca actacatcat taaggtctaa
1021 gacctctatc ccaatttcat ttaataattt tgatgaagca tctgaactag aaacagctcc
1081 agcaaatttg tgccatgctt cctttagtto ttctataaaa aaattaactg ttgagccggt
1141 tccaatacctt aaatcatctt caggatgaag attatttttt atatatctta tagctttgtt
1201 agcaacattt atctttgagc cactcataga gctataaac aagaaaatat agytagttaa
1261 ttatttttag actaaaaatt aaaaaaacag gtctttttta gaattccacg aagtacctaa
1321 agcttatctc taatgcactt gtttatgagc tagcattaaa atcacctata acatttgctc
1381 taaatatttc ttcaaaagct gggataaag ttctctaaa aagagagat ctgcaaccta
1441 tattttcttt taaaaacaga ggagcgtata acaagattgt aaatttatcc gatgccgaaa
1501 aagaagagggg ggttattgct gcatcagcag gaactcagc tcaaggggtta gccagtgcac
1561 tgaagaanaa aaaaataat ttggttgatg ttatgcccat acaactccca gaaataaaaa
1621 taaaagattt aaaaagattt gpagcaaaa tactcccaaa tggggacacg gtatagtcaag
1681 cattaanaaga ggcactgttt attgcaaga aaaaaaaatt gtctttgtt cactctttg
1741 acgacctctt aacaattgct ggccaaggga ctatagacca agaattctt gaagataaaa
1801 ataattttga tgttgtcttt gttccggtgg gaggaaggag tatttagctt ggttatctg
1861 cctggatagc acagaataat aagaaaataa aaattgttgg ttgtgaggtt gaggattccg
1921 cttgtctctg tgaagccgta aaagctataa aaaggtctat tttaaaaga ttgggctctt
1981 ttgctgtggt gtagagcta tcaaggggtt gaaaataaa ttgtgtgtt attaaaggt
2041 gcgtagatga agtcatataa gttagcgtt atgaagctgt caccgctgta aaagatatct
2101 ttgaagatag aagggttcta tcagaacctg ctggggcatt agcacttcca ggttaaaag
2161 cctacgcaag gaaagttaa aataaaaaac ttattgctat aagtctctgc gctaatgtaa
2221 atttccaaag acttaatttt attgttgcag gatcagagat tggtaaaat agagaaaaaa
2281 tattaagtat caaatccca gagatacctg gaagttctct taagctttca agatgtttg
2341 gcagctctca agttacagag tttaactaca ggaactcag cttaagcgtt gcaattgtt
2401 tagttgtgtt tagaactaaa actgaaaaat catttgaat cttaagctcc aaattaaaaa
2461 aagcagcctt cactcttagc gactttactc gaaatgaat atccaatgat catctgaggc
2521 atatggttgg tggcagaagt agtgaactag gctctataa caatgaagaa atatttaggg
2581 gagagtttcc tgagaagacc ggccgctgtt taaattttct agagaaattt ggaataaat
2641 ggaatatttc ctattttcat taagaagacc taggttcagc ttittgaaag atattaatg
2701 gcatcgaga taaggttaa gacaagctaa taaatcatt aataagtica gmatctatt
2761 ttacagaaga aacctctaac aagcatatac aagatttttt aaaaagaag gtaataactt
2821 taactaaat ttaattgaaa aaagctatc gtagggttt tccacgctt ctttgaacaa
2881 ctcgatttga gatctatcat cctctctgtc gtaaatcttc ccaactttag aatagacca
2941 aaatagatat gacaaaaggag ctagctcata ttatatata atttgaagc aagctacgac
3001 agtatctaat tcaataacta tatattttcc taataaagc tatccattag catctgaaaa
3061 aatactataa ggaatttttt ctttcaagc acaaatga cctttaagtc tgatctctgt
3121 ttattatttt ttaaacocat ttagatcaaa agataaggta tctgtctgtt aatcatatga
3181 ggcaagatta ttattatctt gccatctacg ccaattattt tctttotta ttctatatg
3241 cgcattgatt cttaagttat catttggaaa tattgaacct gctatctgtt aaaaattttt
3301 accataccca ttogaatccc accgattatc ttctctccc ttaaagaagc taactctcca
3361 gtcatatgct cagaatgaat agttctttgc tccaagctct gctgtaaac ctattctctt
3421 ttgtattgtt ataaaggggt aggtcttaatt ttittctgt atagtgtat ttctccag
3481 agatetaaag ttaaatctta attgaaattt agagttgtcc ttaaaactaa aagaattttt
3541 ttgatcgatg cctattggat ttgaattacc agctgtgtca gcatcataat ttagatcaat
3601 tcaataatct atttgtttta atatcgagct attataaaat tcaatttttt ttcgatttcc
3661 tcaataacca gcatgaatcc agtctcttct ttgcagataa ccaagatcat ttaactcgaa
3721 gtctcttcca aataaagaa ggcctccact tatgttagat agttattttg gaagatagt
3781 aaactgagtc ctataccaaa gccatttttt accctctttt cctgaggtca acagatgtta
3841 atatgtaatt attttttttg aacgaataat gatgtaatca ataacattga cttgtgatga
3901 ttccgctgtc atttctatct caaactctgt taccatgaag ccaagcgttt tatttccaa
3961 ctttgtttga gatcgaaag cataatagtc tcttccact gaagaagctt catagcctc
4021 actgtctaca aatactccaa attcattatt attacttttt ttgtaagtc ttaatgcaaa
4081 atcaatataa gaatagtttt tttaagtcgc ctgcgaacct tctcttaact ttctcttga
4141 gcaattatag ctgggggag cttcaactct cccgtatttt ataacaggt attctatata
4201 attactataa tcaaatagtt attggttttc attgaaaaat gctctttttt ctgagtaaaa
4261 agttttttga gcagaaaagt taataaccac atcatcactc tcaagttgtc cgaactctg
4321 ataatagctt aaatttattt gacacacttt tcaagtctta taaagatttt cagccccaat
4381 actcgacctt tcttggtttg taactgaatt ttattttgaa gatatatg gaataaaagt
4441 aagctttgat ttgttatagt ttgtatttcc taagctatct aactctgaa agtagctatt
4501 tctactagat attgttccgg cactgttaac ccatgactca ttctcgac tataagtaa
4561 tggcgtgtga ttaatttttc ttatatccac atcaagctgt tcaattaaag ttacatccca
4621 aggaataaaa aactocagaa cccaataccc atcaaatttt ttgtttttt caatccaatc
4681 tcaatccag ttctgttttaa agtctcctgc ttgcttttt atgcatoga aaagcaggtt
4741 ccaagatttt atagcaagaa tgaagctttt gttaccatca ccaatcaagt ctatattat
4801 agagttttta tggtaagtg atttattttg atctctaaag gttctctctg agaatatga
4861 atcaattact tgaataattt taaatccaa atatttacca tcttatttg agaaataa
4921 agcgtttgta agaatctcat ttctttaag agtaaaaga gatgttcat aaaaactgt
4981 aatttcaaat gcaattatcc actcaagctc atcaagagag ccaatcaata caattgagt
5041 tgaccaaatc agcacagag taagtaaaag tgataatgag gctaaagtta ttctcataga
5101 tagatttaat tcaagattat ttaactatg aagagttaga aatcatcoga tcaattcag
5161 aaaaaacata atcaattata aggtctatgt ttcttccaaa atgcattata caaattctgt
5221 attgataac aagtcacac aataaagtc ttgatccctc ctgtttata gagaataa
5281 ttctatcgaa atatccagag ccaatagccaa gctgtatcc atttaactca actcctgca
5341 taggaataaa cattaaatca atttcaatta ttgtacata atcttcaatt ttaactctt

cat /Users/jeisen/Desktop/sequence.gb.txt | grep product

```
/product="predicted threonine dehydratase"
/product="predicted secreted protein"
/product="predicted 5-formyltetrahydrofolate cyclo-ligase"
/product="predicted Xaa-Pro aminopeptidase"
/product="predicted quinone biosynthesis monooxygenase"
/product="predicted 50S ribosomal protein L28"
/product="predicted deoxyuridine 5'-triphosphate
/product="predicted primosomal protein N' (replication
/product="predicted N-acetylmuramoyl-L-alanine amidase"
/product="predicted kinase of the phosphomethylpyrimidine
/product="predicted kinase of P-loop ATPase superfamily"
/product="predicted eukaryote-type paraoxonase"
/product="predicted alpha/beta hydrolase"
/product="predicted MutT superfamily hydrolase"
/product="predicted metallobeta lactamase fold protein"
/product="predicted acetyl-coenzyme A synthetase"
/product="predicted deoxypurine kinase"
/product="predicted poly(A) polymerase"
/product="predicted DskA family protein with conserved
/product="predicted ribokinase family sugar kinase"
/product="predicted glutamate-1-semialdehyde
/product="predicted enzyme with conserved cysteine"
/product="predicted metallopeptidase of the G-G peptidase
/product="predicted proline dehydrogenase"
/product="predicted tyrosyl-tRNA synthetase"
/product="predicted prolyl endopeptidase"
/product="tRNA-Ile"
/product="predicted unknown protein"
/product="predicted cation efflux pump"
/product="predicted TonB-dependent receptor protein"
/product="predicted conserved membrane protein"
/product="predicted LicA, choline kinase involved in LPS
/product="predicted NAD-dependent formate dehydrogenase"
/product="predicted 2-nitropropane dioxygenase"
/product="predicted oxidoreductase"
/product="predicted CsgA, Rossman fold oxidoreductase"
/product="predicted acid--CoA ligase fadD13"
/product="proteorhodopsin"
/product="predicted phosphopantetheine
/product="predicted formamidopyrimidine-DNA glycosylase"
/product="predicted outer membrane protein TolC"
/product="predicted topoisomerase IV subunit B"
/product="predicted topoisomerase IV subunit A"
/product="predicted succinylornithine aminotransferase"
/product="predicted 6-phosphofructokinase"
/product="predicted ORF"
/product="predicted penicillin-binding protein 2"
/product="predicted membrane-bound lytic transglycosylase"
/product="predicted penicillin-binding protein 6
/product="predicted D-alanine aminotransferase"
/product="predicted DNA polymerase III, delta subunit"
/product="predicted leucyl-tRNA synthetase (leuS)"
/product="predicted acetyl-CoA carboxylase, biotin
/product="predicted esterase of alpha/beta hydrolase fold"
/product="predicted ABC transporter ATPase subunit"
/product="predicted dihydroxyacid dehydratase"
/product="predicted porphobilinogen synthase"
/product="predicted transcription termination factor Rho"
/product="predicted porphobilinogen deaminase"
/product="predicted transporter (ACRA/ACRE type)"
/product="predicted cation efflux system (AcrB/AcrD/AcrF
/product="predicted secreted lipoprotein"
/product="predicted ribosomal large subunit pseudouridine
/product="predicted acetolactate synthase III large chain"
/product="predicted ketol-acid reductoisomerase"
/product="16S ribosomal RNA"
/product="23S ribosomal RNA"
/product="predicted YacE family of P-loop kinases"
/product="predicted preprotein translocase secA subunit"
```

```
/product="proteorhodopsin"  
/protein_id="AAG10475.1"  
/db_xref="GI:9971913"  
/translation="MKLLLILGSVIALPTFAAGGGDLASDYTGVSFWLVTAALLAST  
VFFFVERDRVSAKWKTSLTVSGLVTGIAFWHYMYMRGVWIETGDSPTVFRYIDWLLTV  
PLLICEFYLLILAAATNVAGSLFKKLLVGSLVMLVFGYMGEAGIMAAWPAFIIGCLAWV  
YMIYELWAGEGKSACNTASPAVQSAYNTMMYIIIFGWAIYPVGYFTGYLMGDGGSALN  
LNLIYNLADFDVNKILFGLIIWNVAVKESSNA"  
51815..52402
```

The inferred amino acid sequence of the proteorhodopsin showed statistically significant similarity to archaeal rhodopsins (*11*).

The inferred amino acid sequence of the proteorhodopsin showed statistically significant similarity to archaeal rhodopsins (11).

11. The protein most similar to proteorhodopsin was the sensory rhodopsin from *Natronomonas pharaonis* and bacteriorhodopsin from *Halobacterium halobium*, with a random expectation value of 2×10^{-10} and 30% identity in a 224–amino acid alignment and a random expectation value of 9×10^{-9} and 27% identity over 228 amino acids, respectively.

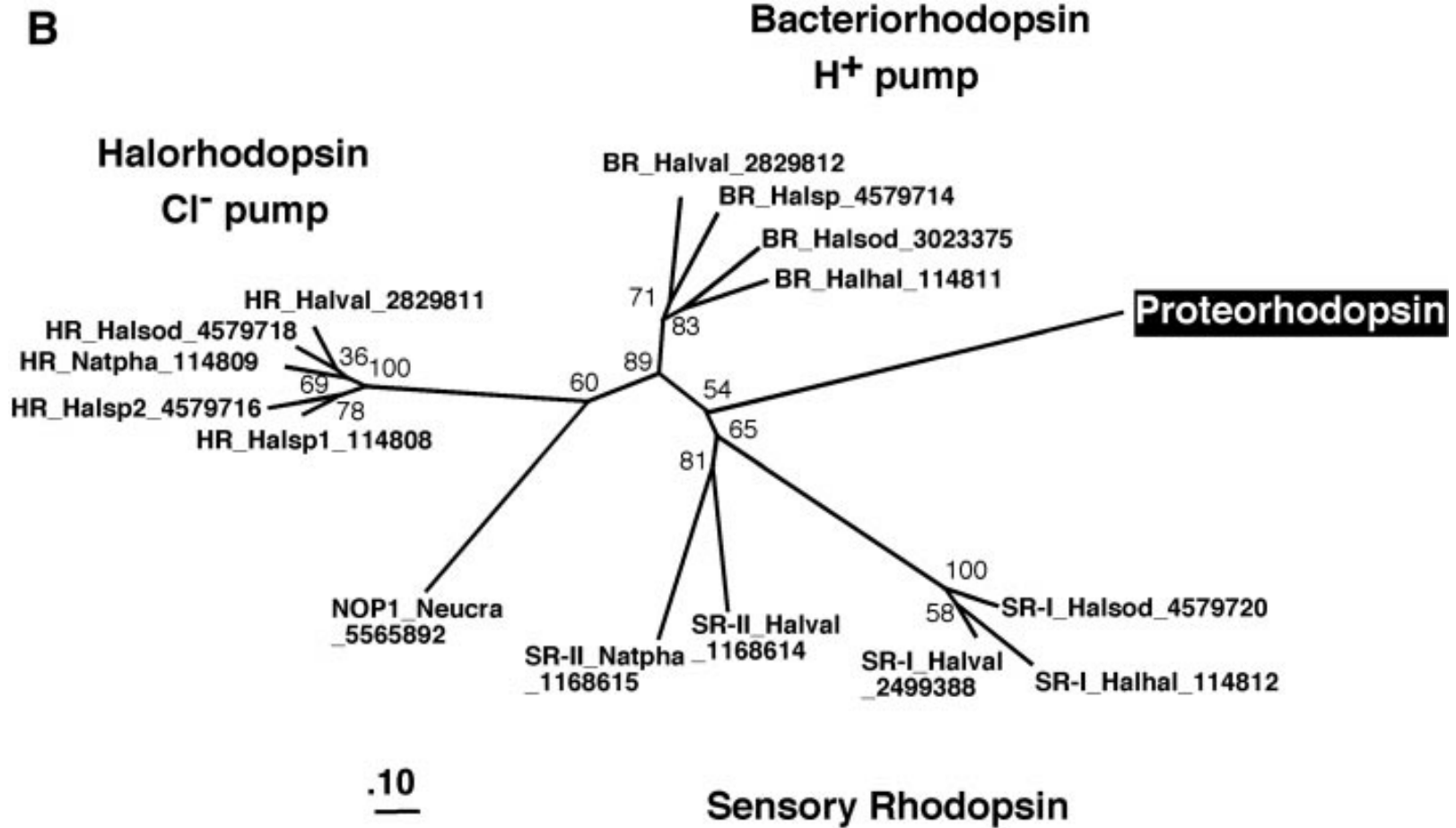


Fig. 1. (B) Phylogenetic analysis of proteorhodopsin with archaeal (BR, HR, and SR prefixes) and *Neurospora crassa* (NOP1 prefix) rhodopsins (16). Nomenclature: Name_Species.abbreviation_Genbank.gi (HR, halorhodopsin; SR, sensory rhodopsin; BR, bacteriorhodopsin). Halsod, *Halorubrum sodomense*; Halhal, *Halobacterium salinarum* (halobium); Halval, *Haloarcula vallismortis*; Natpha, *Natronomonas pharaonis*; Halsp, *Halobacterium* sp; Neucra, *Neurospora crassa*.



Bacteriorhodopsin and its relatives

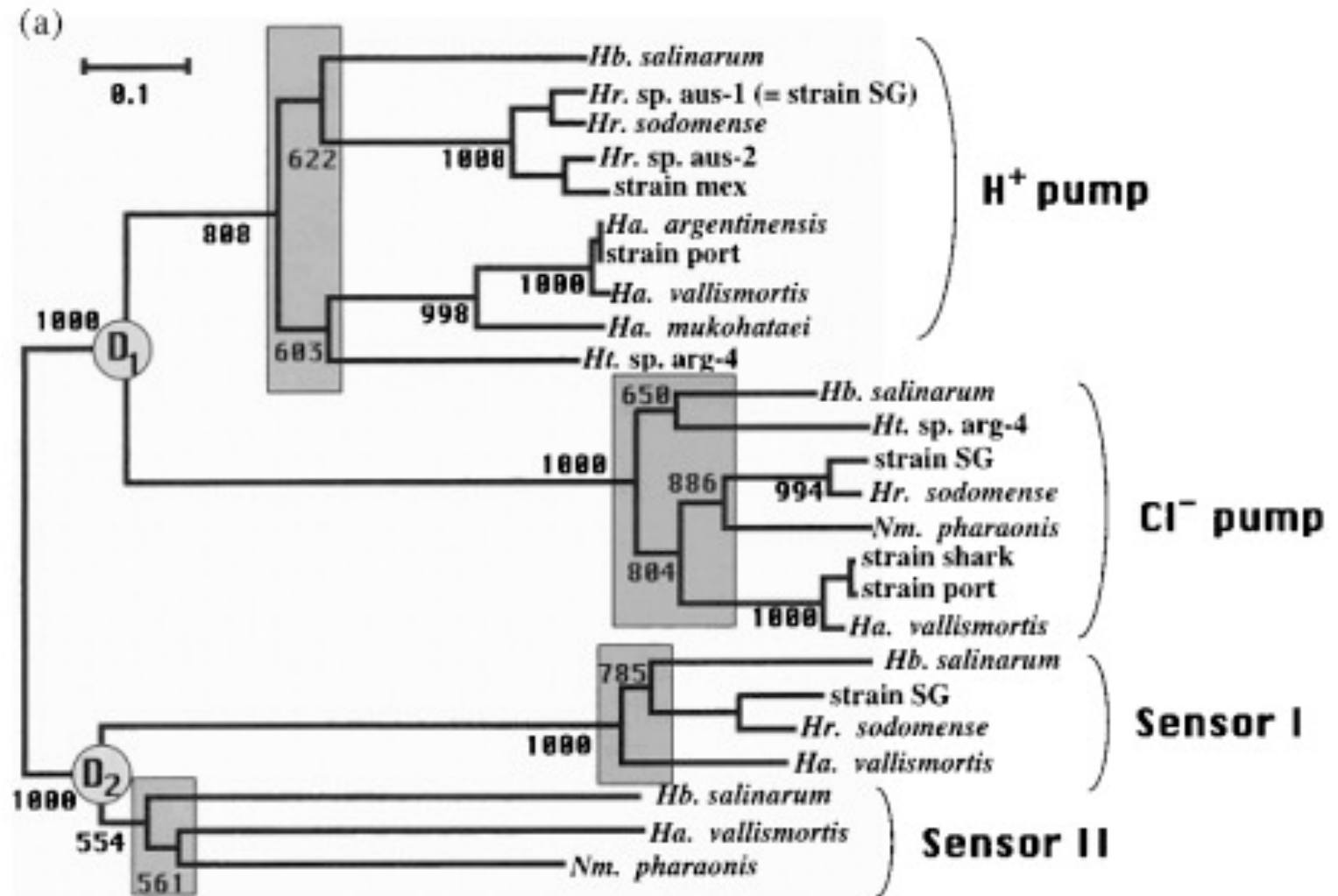


Figure 3. Phylogenetic tree based on the amino acid sequences of 25 archaeal rhodopsins. (a) NJ-tree. The numbers at each node are clustering probabilities generated by bootstrap resampling 1000 times. D1 and D2 represent gene duplication points. The four shaded rectangles indicate the speciation dates when halobacteria speciation occurred at the genus level. (b) ML-tree. Log likelihood value for ML-tree was -6579.02 (best score) and that for topology of the NJ-tree was -6583.43. The stippled bars indicate the 95% confidence limits. Both trees were tentatively rooted at the mid-point of the longest distance, although true root positions are unknown.

What Does It Do?

[illegible]

The amino acid residues that form a retinal binding pocket in archaeal rhodopsins are also highly conserved in proteorhodopsin (Fig. 2). In particular, the critical lysine residue in helix G, which forms the Schiff base linkage with retinal in archaeal rhodopsins, is present in proteorhodopsin. Analysis of a structural model of proteorhodopsin (14), in conjunction with multiple sequence alignments, indicates that the majority of active site residues are well conserved between proteorhodopsin and archaeal bacteriorhodopsins (15).

14. Considering differences in the ionic environment, sequence differences between archaeal and bacterial rhodopsins might be expected to accumulate at residues near the solvent interface. To understand the spatial distribution of these amino acid substitutions, a model of the proteorhodopsin was constructed using the known crystal structure of bacteriorhodopsin. The model was constructed by threading the proteorhodopsin sequence on the 1.55 Å resolution structure of bacteriorhodopsin [H. Luecke, B. Schobert, H. T. Richter, J. P. Cartailler, J. K. Lanyi, *J. Mol. Biol.* **291**, 899 (1999)]. This pairwise alignment was adjusted to minimize the Sippl mean force field energy [M. J. Sippl, *J. Mol. Biol.* **213**, 859 (1990)]. This was followed by model construction with PROMODII and energy minimization with the CHARMM force field. The structure was visualized with the Swiss PDB viewer [N. Guex and M. C. Peitsch, *Electrophoresis* **18**, 2714 (1997)]. In comparison to the most closely related archaeal sequences, the majority of the nonconservative substitutions resulted in changes in residue polarity and were typically localized to regions of solution contact (19). In contrast to archaeal rhodopsins, proteorhodopsin lacks a region between helix B and C that adopts an extended conformation and is in contact with solution on the extracellular surface.

The amino acid residues that form a retinal binding pocket in archaeal rhodopsins are also highly conserved in proteorhodopsin (Fig. 2). In particular, the critical lysine residue in helix G, which forms the Schiff base linkage with retinal in archaeal rhodopsins, is present in proteorhodopsin. Analysis of a structural model of proteorhodopsin (14), in conjunction with multiple sequence alignments, indicates that the majority of active site residues are well conserved between proteorhodopsin and archaeal bacteriorhodopsins (15).

15. Conserved amino acid residues include Lys²³¹ (Lys²¹⁶), which forms the Schiff base with retinal; Arg⁹⁴ (Arg⁸²); Asp⁹⁷ (Asp⁸⁵); Thr¹⁰¹ (Thr⁸⁹); and Asp²²⁷ (Asp²¹²). The numbering begins with the NH₂-terminal residue of proteorhodopsin; the numbers of the corresponding residues in the well-characterized bacteriorhodopsin from *Halobacterium salinarum* are indicated in parentheses. Of these residues, Asp⁹⁷ (Asp⁸⁵) is critical for proton transfer from the photoactivated retinal. Residue Asp²²⁷ (Asp²¹²) probably interacts with another conserved site, Tyr²⁰⁰ (Tyr¹⁸⁵), which contributes to the required

environment for the proton conduit from the Schiff base [J. Heberle, *Biochim. Biophys. Acta* **1458**, 135 (2000)]. Arg⁹⁴ (Arg⁸²) appears to regulate the process of proton release that occurs before the proton uptake from the cytoplasm [R. Govindjee *et al.*, *Biophys. J.* **71**, 1011 (1996)]. Thus, the essential steps of light-driven proton pumping by proteorhodopsin likely follow the same path as in the archaeal bacteriorhodopsins. Glu¹⁰⁸ of proteorhodopsin corresponds to Asp⁹⁶ of bacteriorhodopsin, the proton donor to the Schiff base in the reprotonation phase of the pumping cycle. The carboxylic acid residue in this position is shared by bacteriorhodopsins and proteorhodopsin, but not by sensory rhodopsins and halorhodopsins, which suggests mechanistic similarity between proteorhodopsin and bacteriorhodopsins in the proton uptake stage of their pumping cycles. A major difference between proteorhodopsin and bacteriorhodopsins is apparent in the pathway of proton release to the exterior of the cell. In bacteriorhodopsin, this step is largely mediated by two glutamates, Glu¹⁹⁴ and Glu²⁰⁴ [L. S. Brown *et al.*, *J. Biol. Chem.* **270**, 27122 (1995); S. P. Balashov *et al.*, *Biochemistry* **36**, 8671 (1997)]. One or both of these residues are missing in other members of the prokaryotic rhodopsin family. Examination of the distribution of negatively charged residues on the external face of proteorhodopsin using a structural model (14) shows that Asp²⁷, Glu⁶⁵, Asp⁸⁸, Glu¹⁴², and Asp²¹² are candidates for forming the proton path. Of these residues, Glu¹⁴² and Glu⁶⁵ appear to be optimally positioned to act as the principal proton acceptors in the release process.

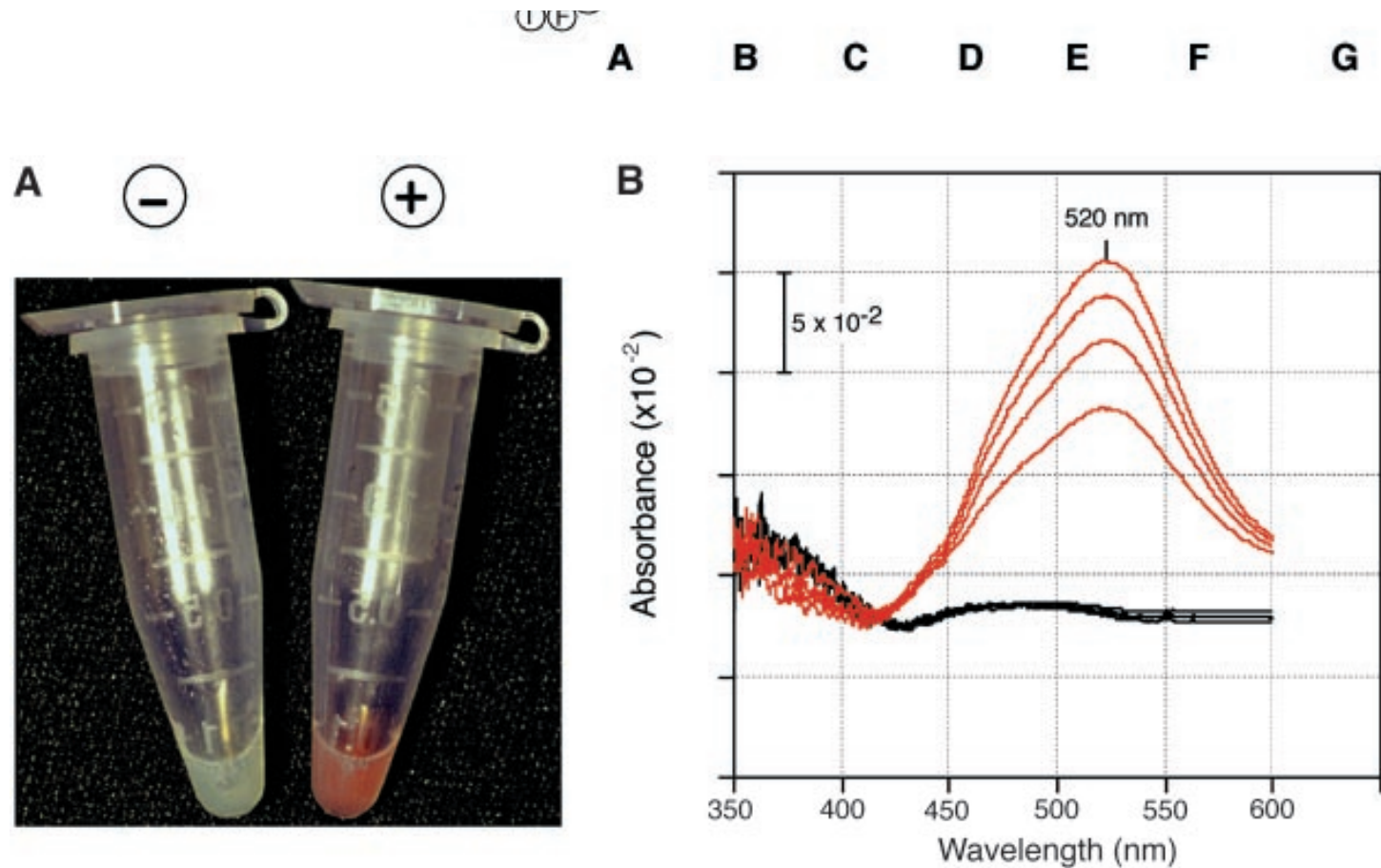


Fig. 3. (A) Proteorhodopsin-expressing *E. coli* cell suspension (+) compared to control cells (-), both with all-*trans* retinal. (B) Absorption spectra of retinal-reconstituted proteorhodopsin in *E. coli* membranes (17). A time series of spectra is shown for reconstituted proteorhodopsin membranes (red) and a negative control (black). Time points for spectra after retinal addition, progressing from low to high absorbance values, are 10, 20, 30, and 40 min.

Fig. 3. (A) Proteorhodopsin-expressing *E. coli* cell suspension (+) compared to control cells (-), both with all-*trans* retinal. (B) Absorption spectra of retinal-reconstituted proteorhodopsin in *E. coli* membranes (17). A time series of spectra is shown for reconstituted proteorhodopsin membranes (red) and a negative control (black). Time points for spectra after retinal addition, progressing from low to high absorbance values, are 10, 20, 30, and 40 min.

Proteorhodopsin was amplified from the 130-kb bacterioplankton BAC clone 31A08 by polymerase chain reaction (PCR), using the primers 5'-ACCATGGGTA-AAT TAT TACTGATAT TAGG-3' and 5'-AGCAT TA-GAAGAT TCT T TAACAGC-3'. The amplified fragment was cloned with the pBAD TOPO TA Cloning Kit (Invitrogen). The protein was cloned with its native signal sequence and included an addition of the V5 epitope and a polyhistidine tail in the COOH-terminus. The same PCR product in the opposite orientation was used as a negative control. The protein was expressed in the *E. coli* outer membrane protease-deficient strain UT5600 [M. E. Elish, J. R. Pierce, C. F. Earhart, *J. Gen. Microbiol.* **134**, 1355 (1988)] and induced with 0.2% arabinose for 3 hours. Membranes were prepared according to Shimono *et al.* [K. Shimono, M. Iwamoto, M. Sumi, N. Kamo, *FEBS Lett.* **420**, 54 (1997)] and resuspended in 50 mM Tris-Cl (pH 8.0) and 5 mM MgCl₂. The absorbance spectrum was measured according to Bieszke *et al.* (7) in the presence of 10 M all-*trans* retinal.

Proteorhodopsin function

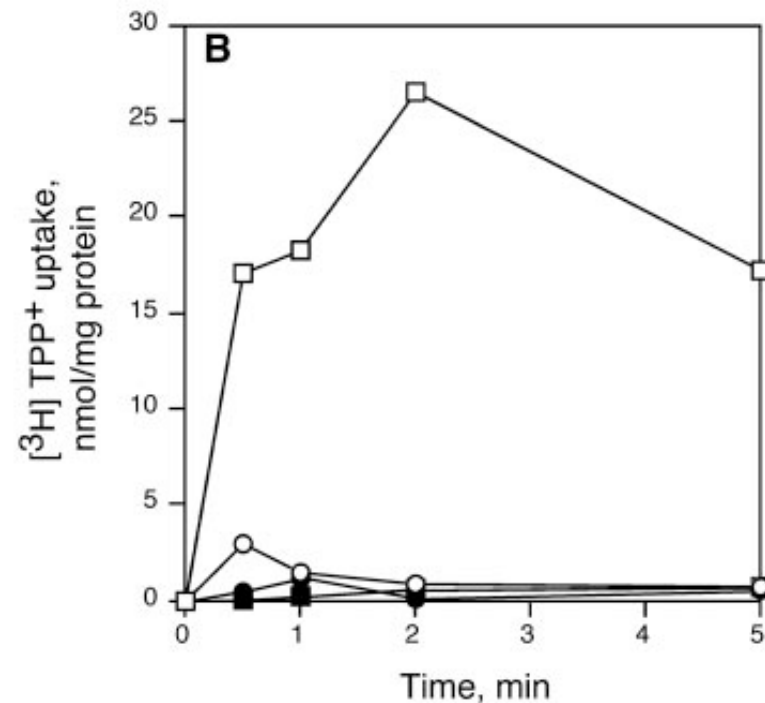
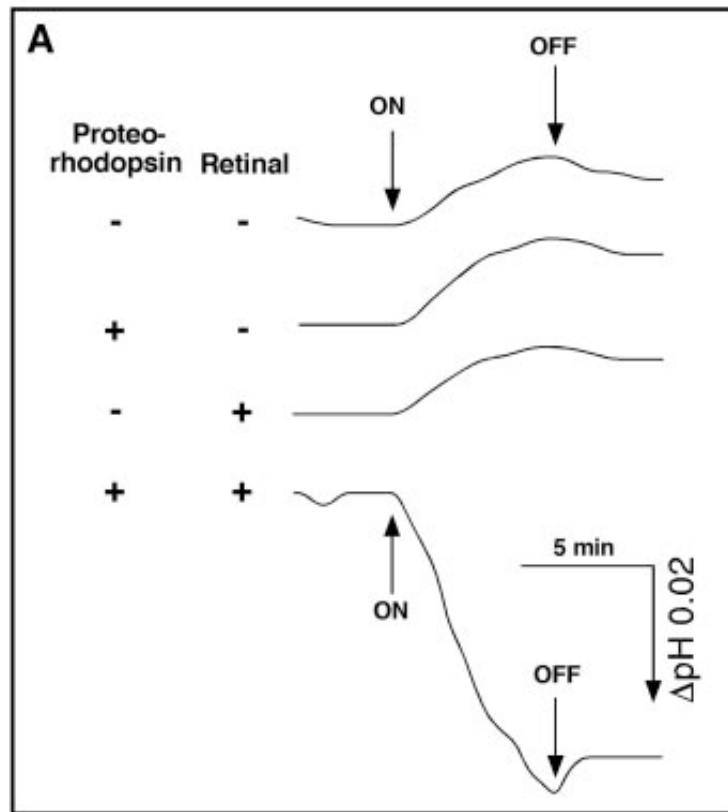


Fig. 4. (A) Light-driven transport of protons by a proteorhodopsin-expressing *E. coli* cell suspension. The beginning and cessation of illumination (with yellow light >485 nm) is indicated by arrows labeled ON and OFF, respectively. The cells were suspended in 10 mM NaCl, 10 mM $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, and 100 μM CaCl_2 . **(B)** Transport of $^3\text{H}^+$ -labeled tetraphenylphosphonium ($[\text{}^3\text{H}^+]\text{TPP}$) in *E. coli* right-side-out vesicles containing expressed proteorhodopsin, reconstituted with (squares) or without (circles) 10 μM retinal in the presence of light (open symbols) or in the dark (solid symbols) (20).

20. Right-side-out membrane vesicles were prepared according to Kaback [H. R. Kaback, *Methods Enzymol.* **22**, 99 (1971)]. Membrane electrical potential was measured with the lipophilic cation TPP by means of rapid filtration [E. Prossnitz, A. Gee, G. F. Ames, *J. Biol. Chem.* **264**, 5006 (1989)] and was calculated according to Robertson *et al.* [D. E. Robertson, G. J. Kaczorowski, M. L. Garcia, H. R. Kaback, *Biochemistry* **19**, 5692 (1980)]. No energy sources other than light were added to the vesicles, and the experiments were conducted at room temperature.

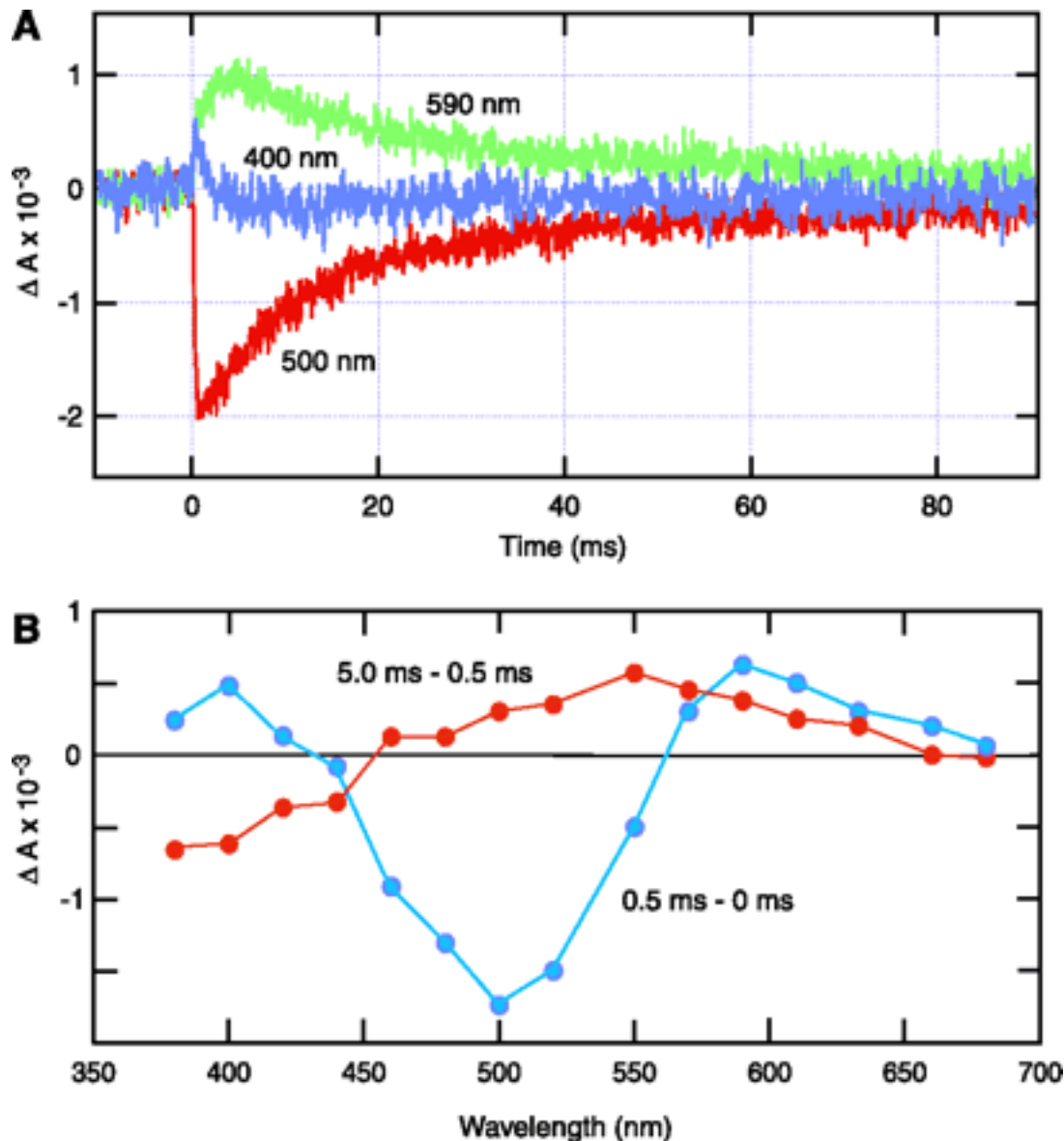


Figure 5

Laser flash-induced absorbance changes in suspensions of *E. coli* membranes containing proteorhodopsin. A 532-nm pulse (6 ns duration, 40 mJ) was delivered at time 0, and absorption changes were monitored at various wavelengths in the visible range in a lab-constructed flash photolysis system as described (34). Sixty-four transients were collected for each wavelength. (A) Transients at the three wavelengths exhibiting maximal amplitudes. (B) Absorption difference spectra calculated from amplitudes at 0.5 ms (blue) and between 0.5 ms and 5.0 ms (red).

Implications. The γ -proteobacteria that harbor the proteorhodopsin are widely distributed in the marine environment. These bacteria have been frequently detected in culture-independent surveys (24) in coastal and oceanic regions of the Atlantic and Pacific Oceans, as well as in the Mediterranean Sea (8, 25–29). In addition to its widespread distribution, preliminary data also suggest that this γ -proteobacterial group is abundant (30,

31) and specifically localized in marine surface waters. Preliminary data (30) also indicate that the abundance of SAR86-like bacteria positively correlates with proteorhodopsin mRNA expression. The absorbance λ_{max} of proteorhodopsin at 520 nm matches well with the photosynthetically available irradiance in the ocean's upper water column. Furthermore, some phylogenetic relatives of the proteorhodopsin-containing bacteria are chemolithoautotrophs that use CO_2 as a sole carbon source (32). Proteorhodopsin could support a photoheterotrophic lifestyle, or it might in fact support a previously unrecognized type of photoautotrophy in the sea. Either of these alternatives suggests the possibility of a previously unrecognized phototrophic pathway that may influence the flux of carbon and energy in the ocean's photic zone worldwide.

.....

Proteorhodopsin phototrophy in the ocean

**Oded Béjà^{*†}, Elena N. Spudich^{†‡}, John L. Spudich[‡], Marion Leclerc^{*}
& Edward F. DeLong^{*}**

<http://www.nature.com/nature/journal/v411/n6839/abs/411786a0.html>

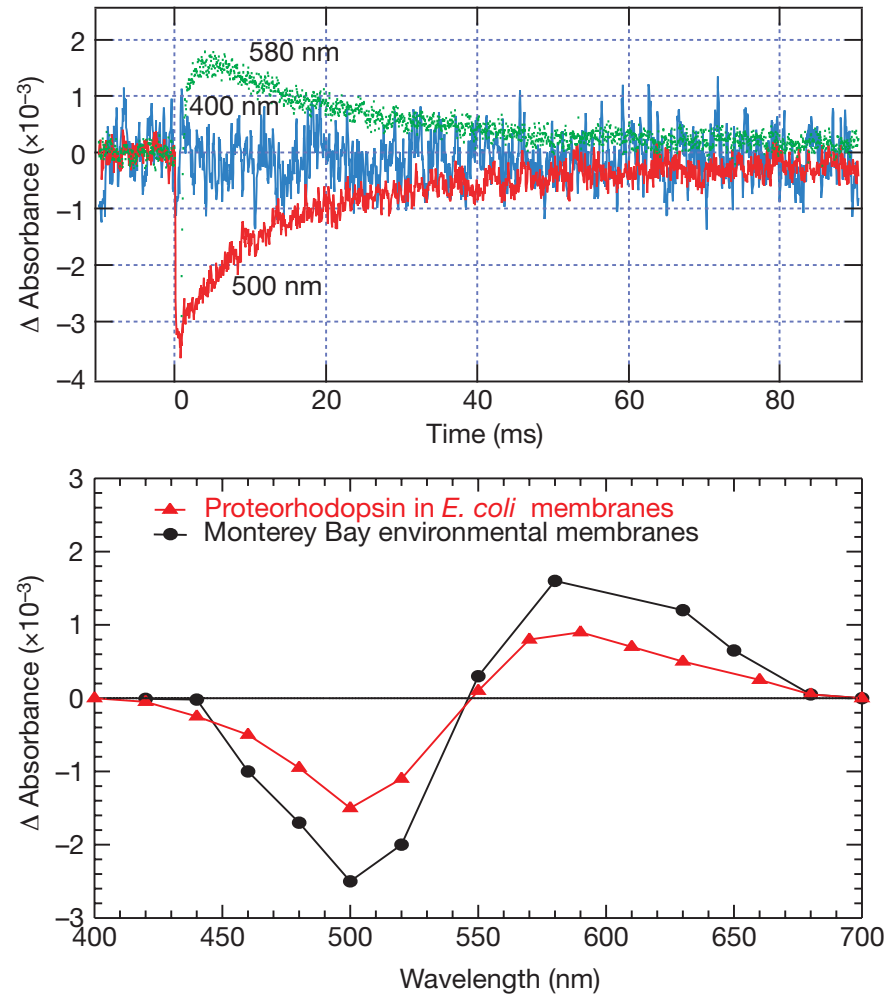


Figure 1 Laser flash-induced absorbance changes in suspensions of membranes prepared from the prokaryotic fraction of Monterey Bay surface waters. Top, membrane absorption was monitored at the indicated wavelengths and the flash was at time 0 at 532 nm. Bottom, absorption difference spectrum at 5 ms after the flash for the environmental sample (black) and for *E. coli*-expressed proteorhodopsin (red).

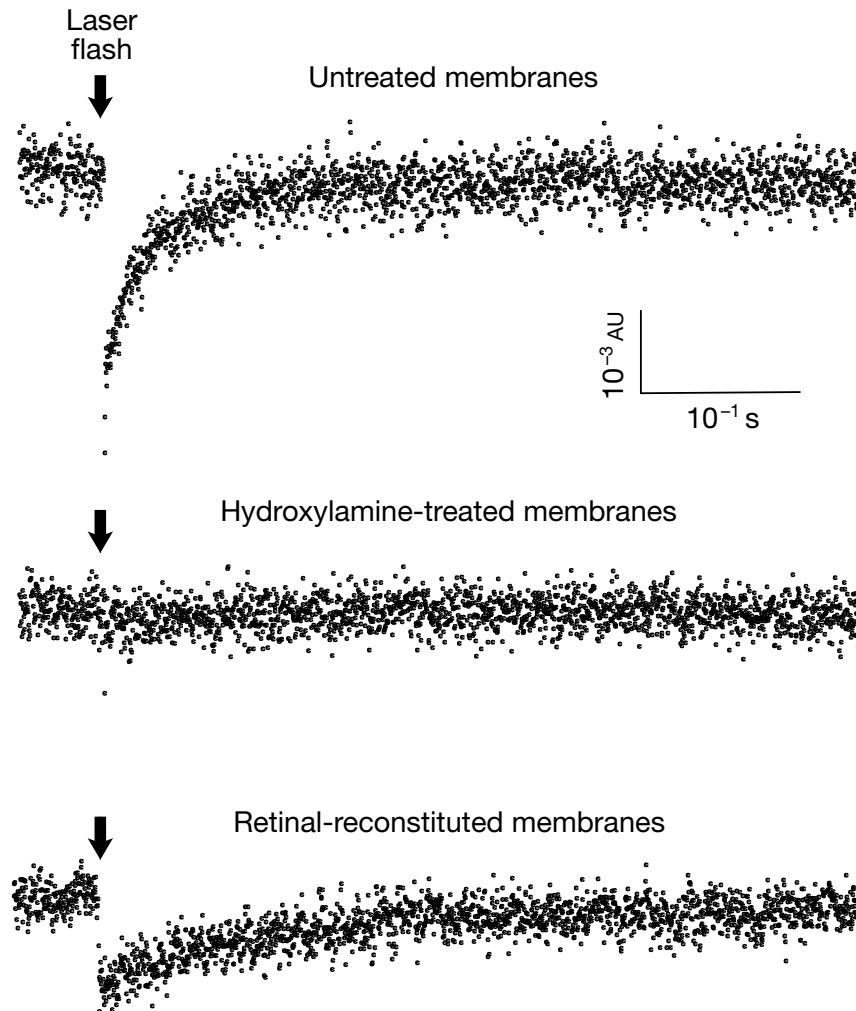


Figure 2 Laser flash-induced transients at 500 nm of a Monterey Bay bacterioplankton membrane preparation. Top, before addition of hydroxylamine; middle, after 0.2 M hydroxylamine treatment at pH 7.0, 18 °C, with 500-nm illumination for 30 min; bottom, after centrifuging twice with resuspension in 100 mM phosphate buffer, pH 7.0, followed by addition of 5 μ M all-*trans* retinal and incubation for 1 h.

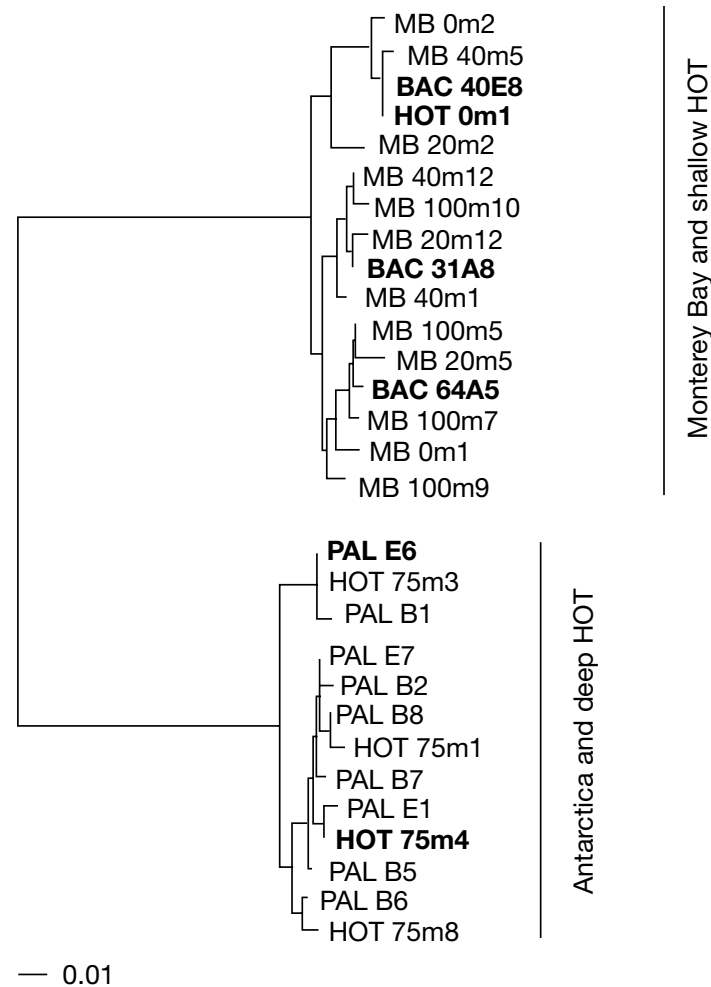


Figure 3 Phylogenetic analysis of the inferred amino-acid sequence of cloned proteorhodopsin genes. Distance analysis of 220 positions was used to calculate the tree by neighbour-joining using the PaupSearch program of the Wisconsin Package version 10.0 (Genetics Computer Group; Madison, Wisconsin). *H. salinarum* bacteriorhodopsin was used as an outgroup, and is not shown. Scale bar represents number of substitutions per site. Bold names indicate the proteorhodopsins that were spectrally characterized in this study.

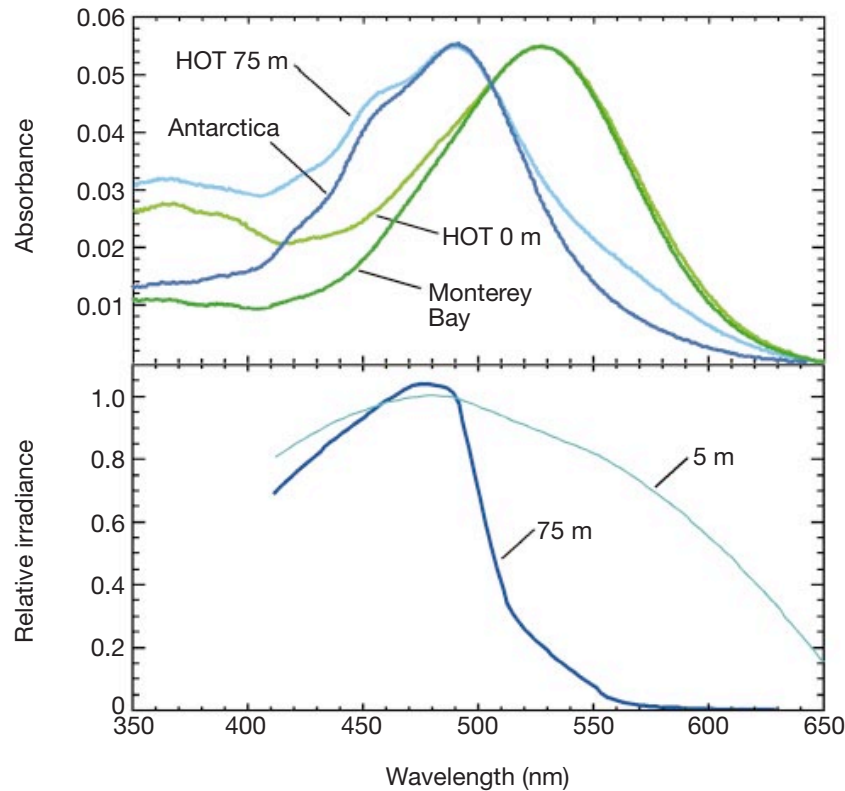


Figure 5 Absorption spectra of retinal-reconstituted proteorhodopsins in *E. coli* membranes. All-*trans* retinal (2.5 μM) was added to membrane suspensions in 100 mM phosphate buffer, pH 7.0, and absorption spectra were recorded. Top, four spectra for palE6 (Antarctica), HOT 75m4, HOT 0m1, and BAC 31A8 (Monterey Bay) at 1 h after retinal addition. Bottom, downwelling irradiance from HOT station measured at six wavelengths (412, 443, 490, 510, 555 and 665 nm) and at two depths, for the same depths and date that the HOT samples were collected (0 and 75 m). Irradiance is plotted relative to irradiance at 490 nm.



Figure 4 Multiple alignment of proteorhodopsin amino-acid sequences. The secondary structure shown (boxes for transmembrane helices) is derived from hydropathy plots. Residues predicted to form the retinal-binding pocket are marked in red.

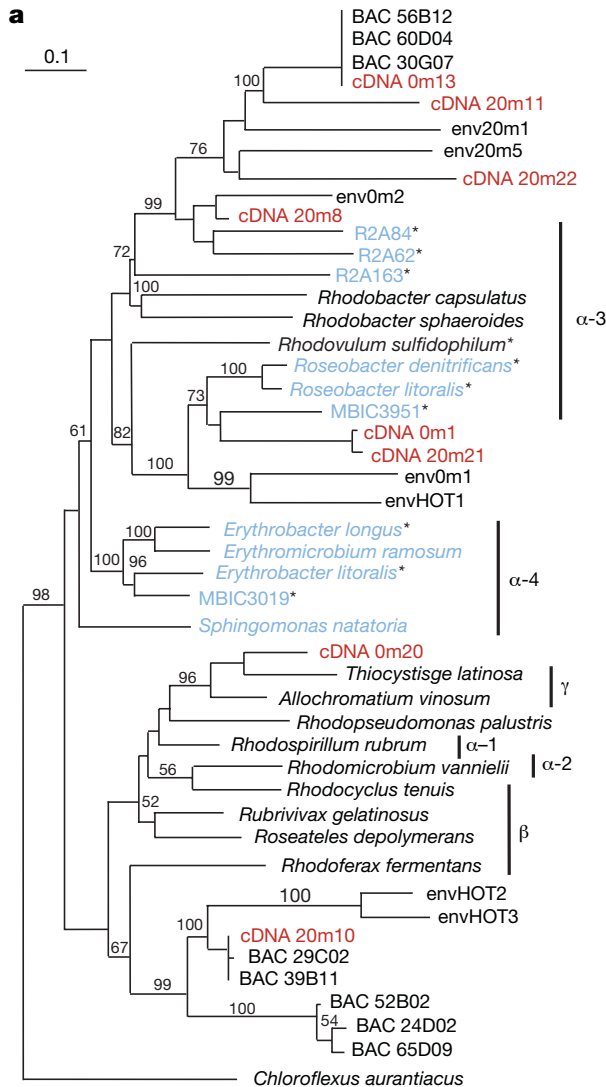
.....

Unsuspected diversity among marine aerobic anoxygenic phototrophs

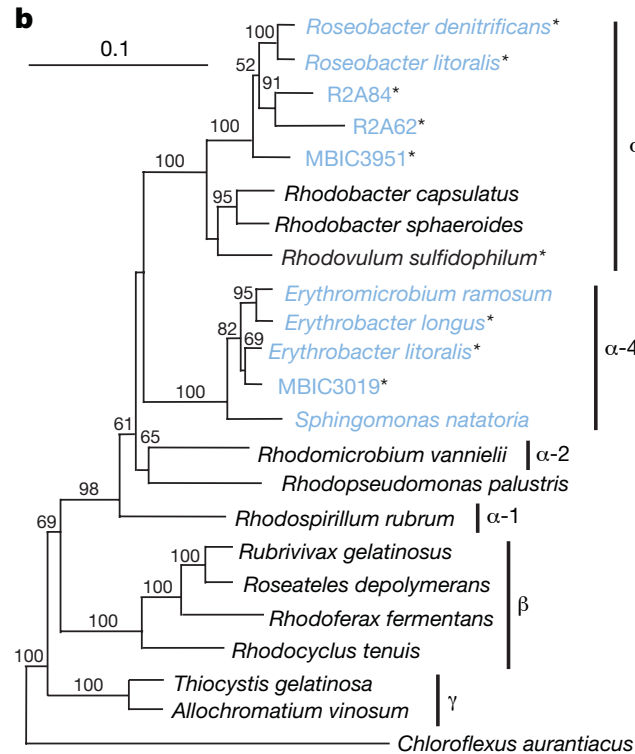
**Oded Béjà^{*†}, Marcelino T. Suzuki^{*}, John F. Heidelberg[‡],
William C. Nelson[‡], Christina M. Preston^{*}, Tohru Hamada^{§†},
Jonathan A. Eisen[‡], Claire M. Fraser[‡] & Edward F. DeLong^{*}**

<http://www.nature.com/nature/journal/v415/n6872/full/415630a.html>

rRNA



PUF



a, b, Evolutionary distances for the *pufM* genes (**a**) were determined from an alignment of 600 nucleotide positions, and for rRNA genes (**b**) from an alignment of 860 nucleotide sequence positions. Evolutionary relationships were determined by neighbour-joining analysis (see Methods). The green non-sulphur bacterium *Chloroflexus aurantiacus* was used as an outgroup. *pufM* genes that were amplified by PCR in this study are indicated by the env prefix, with 'm' indicating Monterey, and HOT indicating Hawaii ocean time series. Cultivated aerobes are marked in light blue, bacteria cultured from sea water are marked with an asterisk, and environmental cDNAs are marked in red. Photosynthetic -, - and - proteobacterial groups are indicated by the vertical bars to the right of the tree. Bootstrap values greater than 50% are indicated above the branches. The scale bar represents number of substitutions per site.

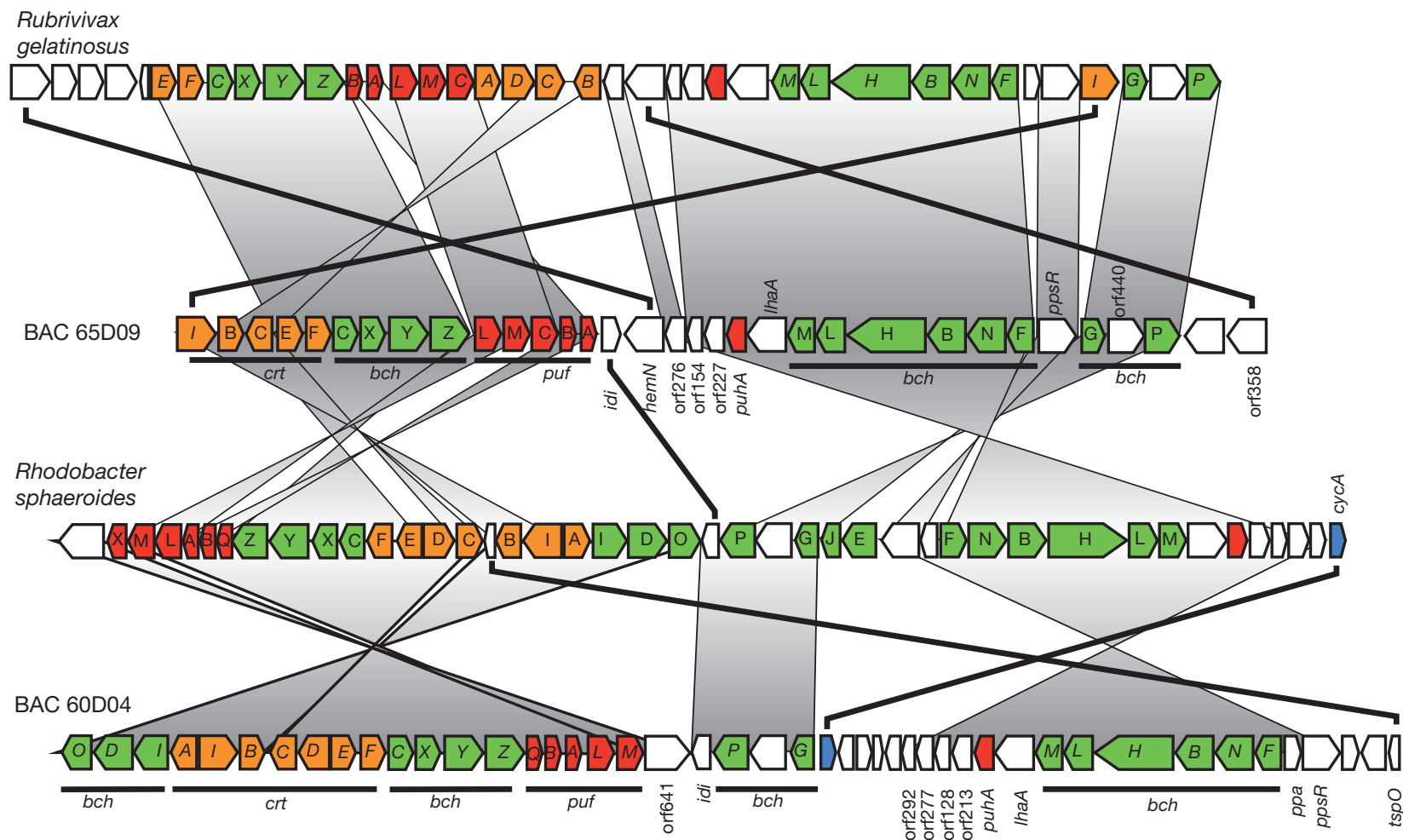


Figure 2 Schematic comparison of photosynthetic operons from *R. gelatinosus* (β -proteobacteria), *R. sphaeroides* (α -proteobacteria) and uncultured environmental BACs. ORF abbreviations use the nomenclature defined in refs 13, 14 and 24. Predicted ORFs are coloured according to biological category: green, bacteriochlorophyll biosynthesis

genes; orange, carotenoid biosynthesis genes; red, light-harvesting and reaction centre genes; and blue, cytochrome c_2 . White boxes indicate non-photosynthetic and hypothetical proteins with no known function. Homologous regions and genes are connected by shaded vertical areas and lines, respectively.

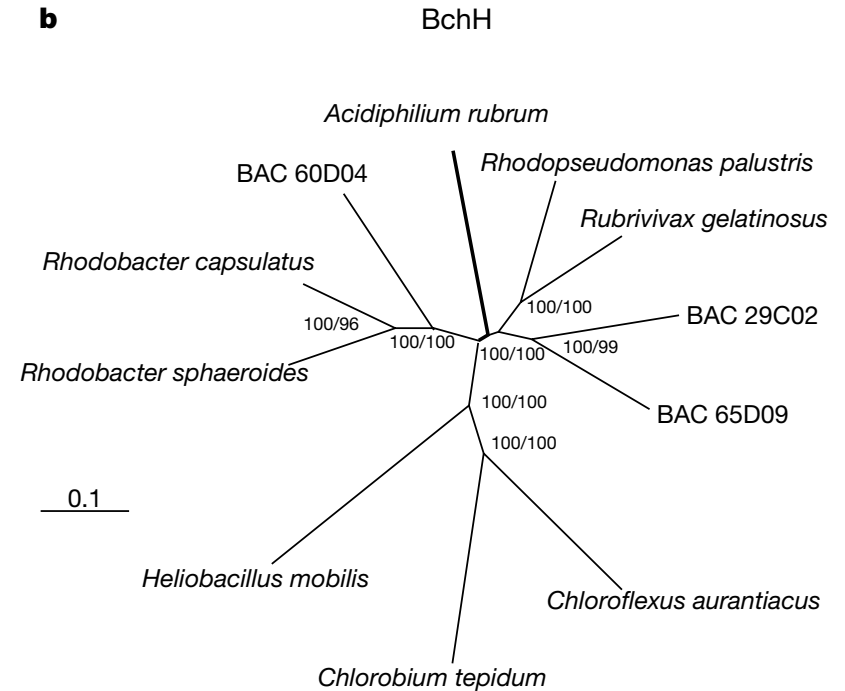
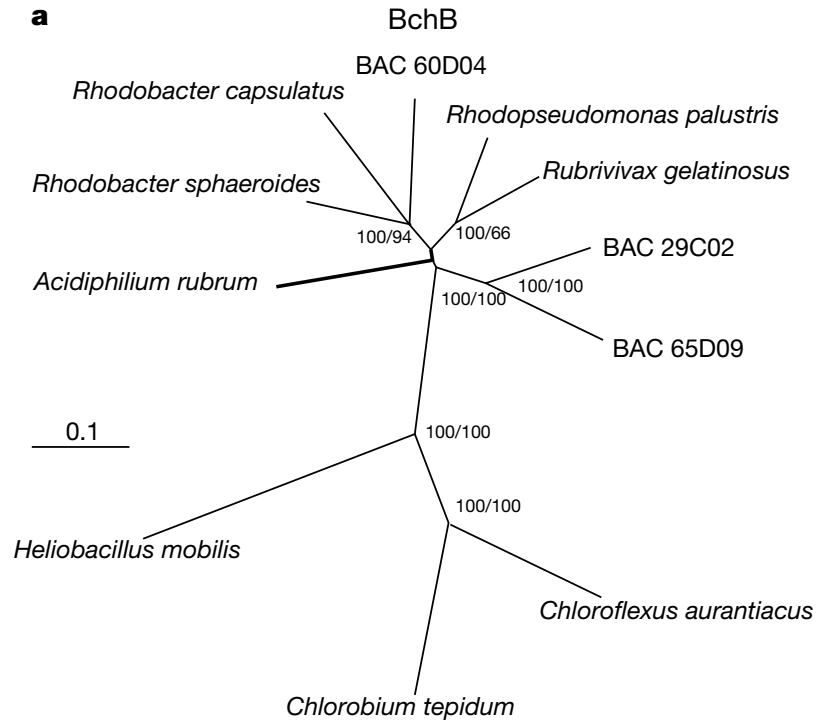


Figure 3 Phylogenetic analyses of BchB and BchH proteins. **a**, Phylogenetic tree for the BchB protein. **b**, Phylogenetic tree for the BchH protein. The BchH sequences from *Chlorobium vibrioforme*²⁵ and BchH2 and BchH3 from *C. tepidum*¹⁸ were omitted from the tree because these genes potentially encode an enzyme for bacteriochlorophyll *c*

biosynthesis and are probably of distinct origin (J. Xiong, personal communication). Bootstrap values (neighbour-joining/parsimony method) greater than 50% are indicated next to the branches. The scale bar represents number of substitutions per site. The position of *Acidiphilium rubrum* (bold branch) was not well resolved by both methods.

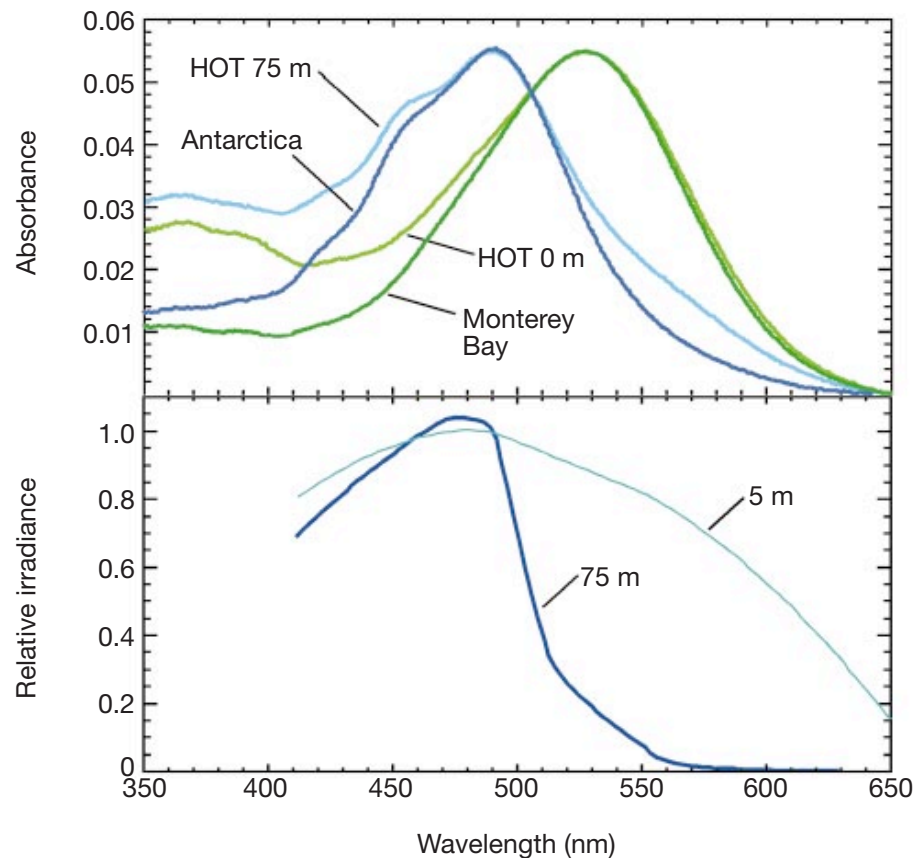


Figure 5 Absorption spectra of retinal-reconstituted proteorhodopsins in *E. coli* membranes. All-*trans* retinal ($2.5 \mu\text{M}$) was added to membrane suspensions in 100 mM phosphate buffer, pH 7.0, and absorption spectra were recorded. Top, four spectra for palE6 (Antarctica), HOT 75m4, HOT 0m1, and BAC 31A8 (Monterey Bay) at 1 h after retinal addition. Bottom, downwelling irradiance from HOT station measured at six wavelengths (412, 443, 490, 510, 555 and 665 nm) and at two depths, for the same depths and date that the HOT samples were collected (0 and 75 m). Irradiance is plotted relative to irradiance at 490 nm.

Community structure and metabolism through reconstruction of microbial genomes from the environment

Gene W. Tyson¹, Jarrod Chapman^{3,4}, Philip Hugenholtz¹, Eric E. Allen¹, Rachna J. Ram¹, Paul M. Richardson⁴, Victor V. Solovyev⁴, Edward M. Rubin⁴, Daniel S. Rokhsar^{3,4} & Jillian F. Banfield^{1,2}

¹Department of Environmental Science, Policy and Management, ²Department of Earth and Planetary Sciences, and ³Department of Physics, University of California, Berkeley, California 94720, USA

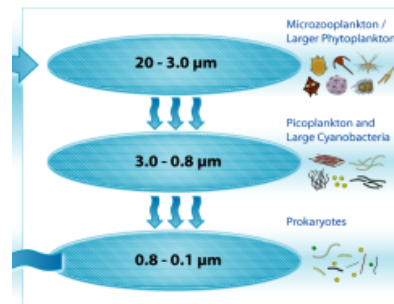
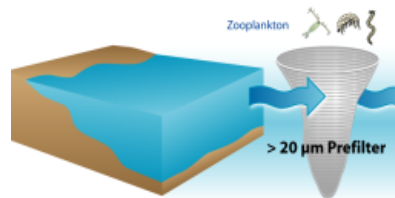
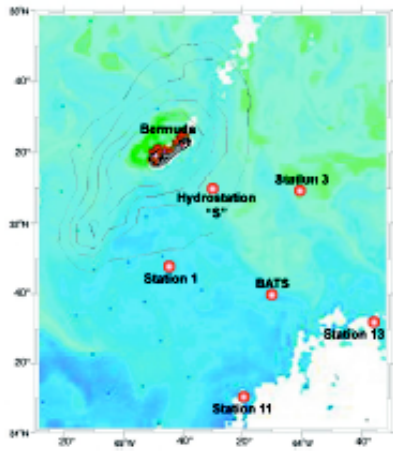
⁴Joint Genome Institute, Walnut Creek, California 94598, USA

RESEARCH ARTICLE

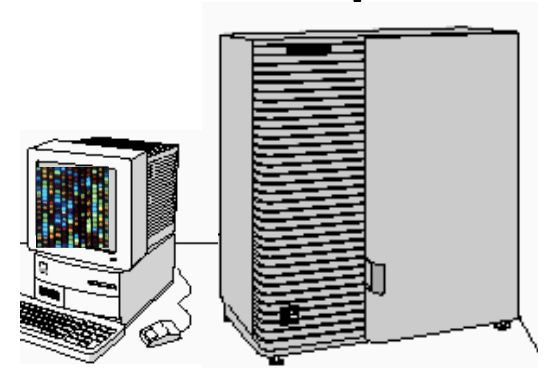
Environmental Genome Shotgun Sequencing of the Sargasso Sea

J. Craig Venter,^{1*} Karin Remington,¹ John F. Heidelberg,³
Aaron L. Halpern,² Doug Rusch,² Jonathan A. Eisen,³
Dongying Wu,³ Ian Paulsen,³ Karen E. Nelson,³ William Nelson,³
Derrick E. Fouts,³ Samuel Levy,² Anthony H. Knap,⁶
Michael W. Lomas,⁶ Ken Nealson,⁵ Owen White,³
Jeremy Peterson,³ Jeff Hoffman,¹ Rachel Parsons,⁶
Holly Baden-Tillson,¹ Cynthia Pfannkoch,¹ Yu-Hui Rogers,⁴
Hamilton O. Smith¹

Shotgun metagenomics



shotgun
sequence



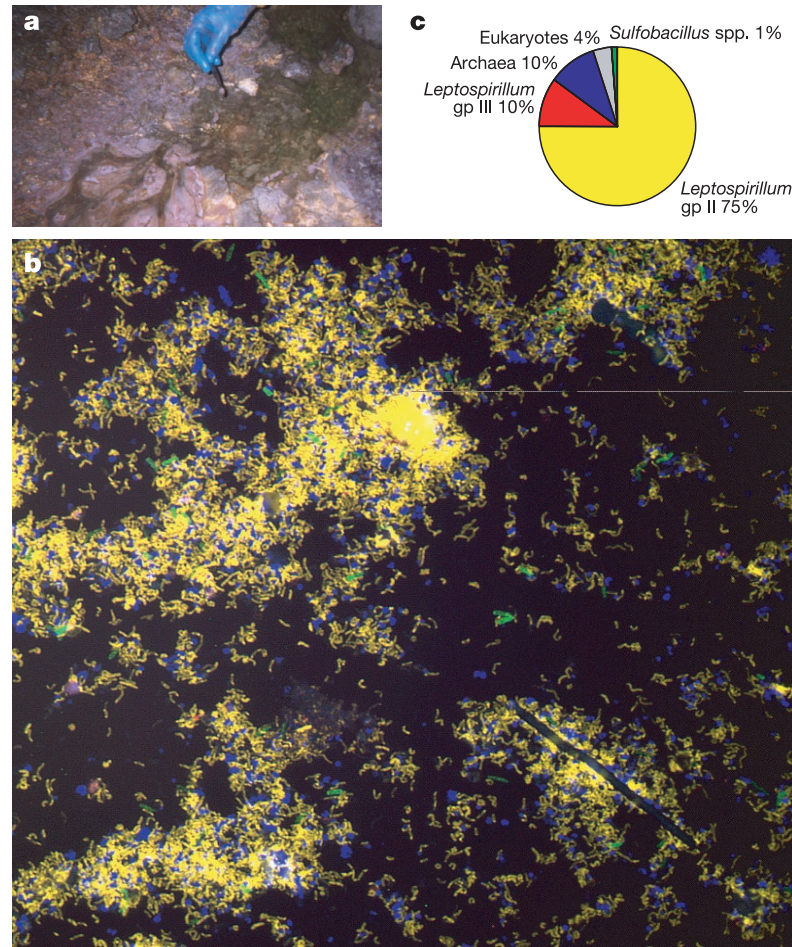


Figure 1 The pink biofilm. **a**, Photograph of the biofilm in the Richmond mine (hand included for scale). **b**, FISH image of **a**. Probes targeting bacteria (EUBmix; fluorescein isothiocyanate (green)) and archaea (ARC915; Cy5 (blue)) were used in combination with a probe targeting the *Leptospirillum* genus (LF655; Cy3 (red)). Overlap of red and green (yellow) indicates *Leptospirillum* cells and shows the dominance of *Leptospirillum*. **c**, Relative microbial abundances determined using quantitative FISH counts.

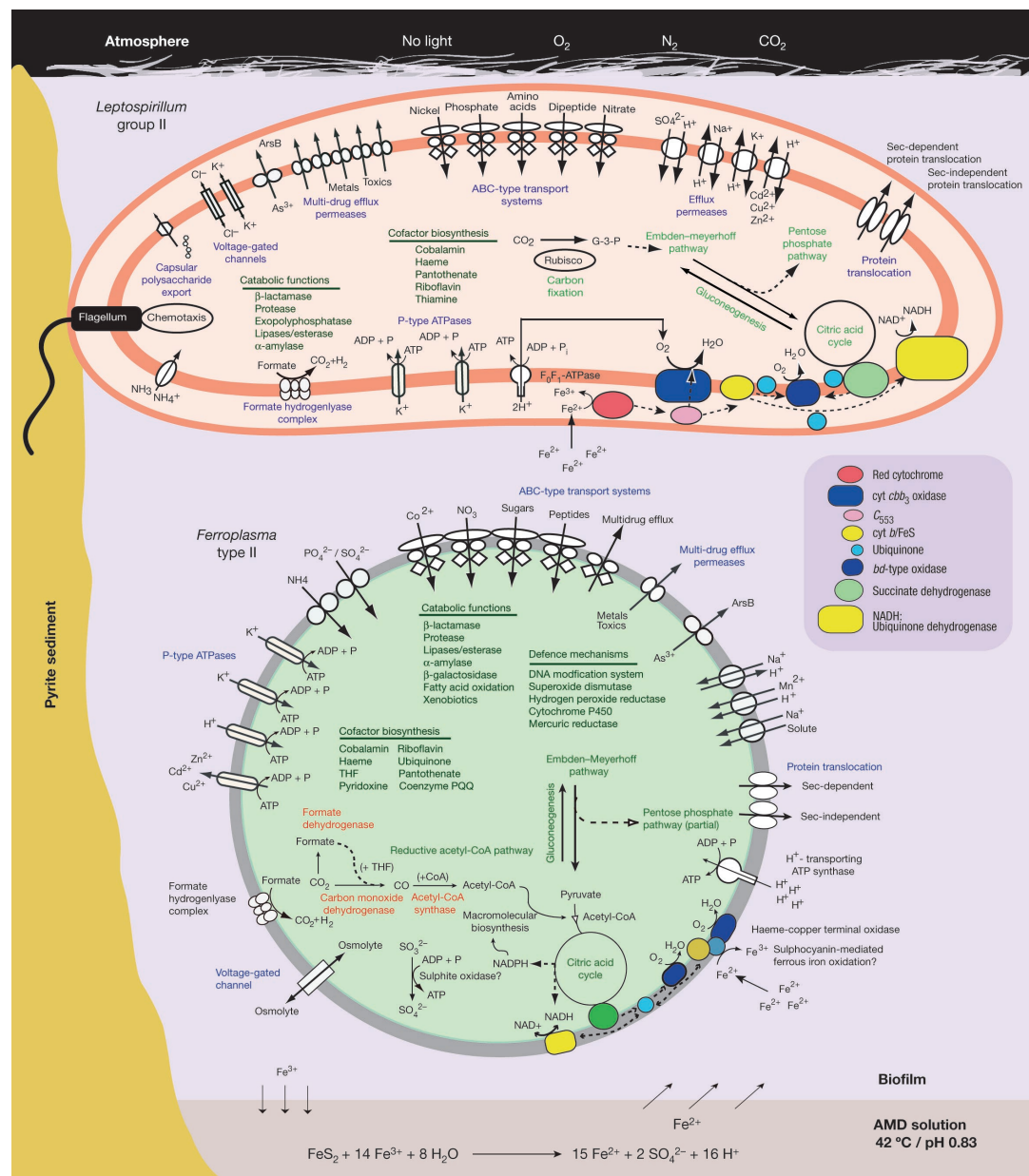


Figure 4 Cell metabolic cartoons constructed from the annotation of 2,180 ORFs identified in the *Leptospirillum* group II genome (63% with putative assigned function) and 1,931 ORFs in the *Ferroplasma* type II genome (58% with assigned function). The cell

drainage stream (viewed in cross-section). Tight coupling between ferrous iron oxidation pyrite dissolution and acid generation is indicated. Rubisco, ribulose 1,5-bisphosphate carboxylase-oxygenase. THF, tetrahydrofolate.

RESEARCH ARTICLE

Environmental Genome Shotgun Sequencing of the Sargasso Sea

J. Craig Venter,^{1*} Karin Remington,¹ John F. Heidelberg,³
Aaron L. Halpern,² Doug Rusch,² Jonathan A. Eisen,³
Dongying Wu,³ Ian Paulsen,³ Karen E. Nelson,³ William Nelson,³
Derrick E. Fouts,³ Samuel Levy,² Anthony H. Knap,⁶
Michael W. Lomas,⁶ Ken Nealson,⁵ Owen White,³
Jeremy Peterson,³ Jeff Hoffman,¹ Rachel Parsons,⁶
Holly Baden-Tillson,¹ Cynthia Pfannkoch,¹ Yu-Hui Rogers,⁴
Hamilton O. Smith¹

<http://www.sciencemag.org/content/304/5667/66>

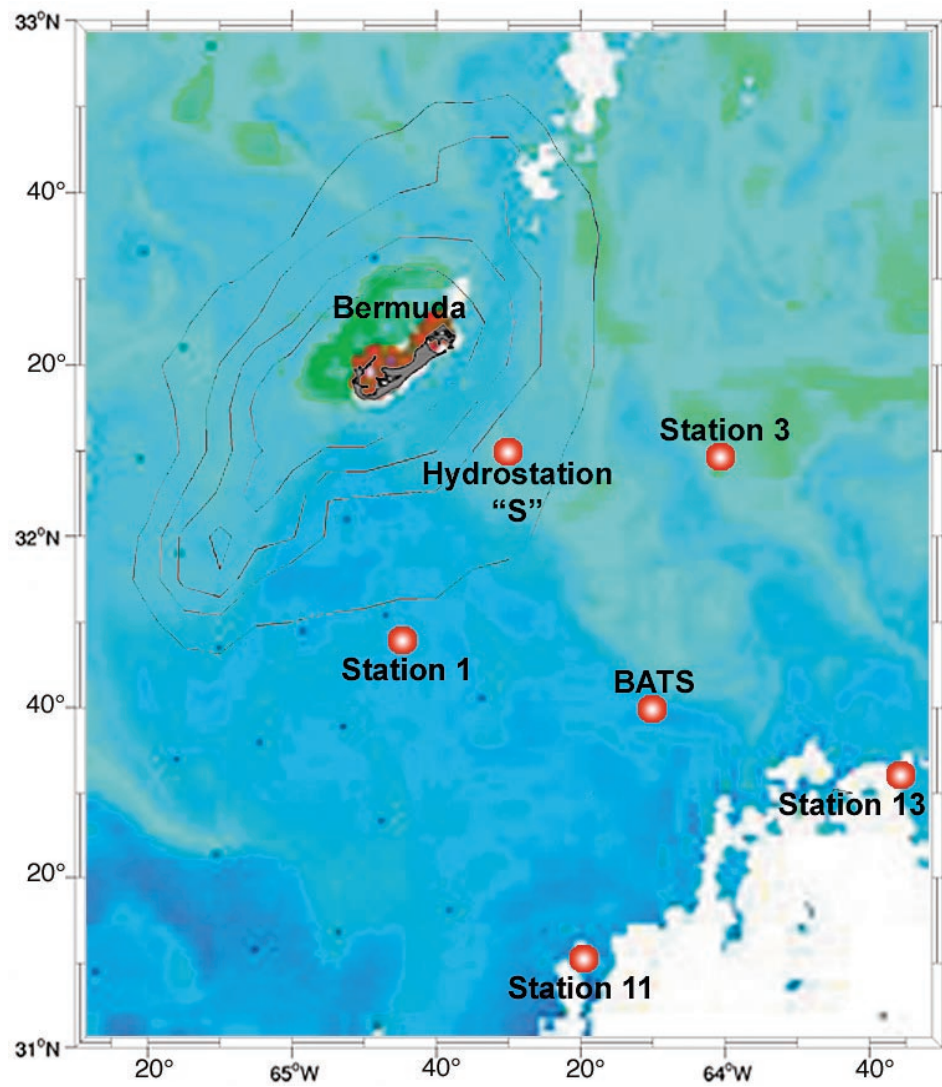


Fig. 1. MODIS-Aqua satellite image of ocean chlorophyll in the Sargasso Sea grid about the BATS site from 22 February 2003. The station locations are overlain with their respective identifications. Note the elevated levels of chlorophyll (green color shades) around station 3, which are not present around stations 11 and 13.

<http://www.sciencemag.org/content/304/5667/66>

Table 1. Gene count breakdown by TIGR role category. Gene set includes those found on assemblies from samples 1 to 4 and fragment reads from samples 5 to 7. A more detailed table, separating Weatherbird II samples from the Sorcerer II samples is presented in the SOM (table S4). Note that there are 28,023 genes which were classified in more than one role category.

TIGR role category	Total genes
Amino acid biosynthesis	37,118
Biosynthesis of cofactors, prosthetic groups, and carriers	25,905
Cell envelope	27,883
Cellular processes	17,260
Central intermediary metabolism	13,639
DNA metabolism	25,346
Energy metabolism	69,718
Fatty acid and phospholipid metabolism	18,558
Mobile and extrachromosomal element functions	1,061
Protein fate	28,768
Protein synthesis	48,012
Purines, pyrimidines, nucleosides, and nucleotides	19,912
Regulatory functions	8,392
Signal transduction	4,817
Transcription	12,756
Transport and binding proteins	49,185
Unknown function	38,067
Miscellaneous	1,864
Conserved hypothetical	794,061
Total number of roles assigned	1,242,230
Total number of genes	1,214,207

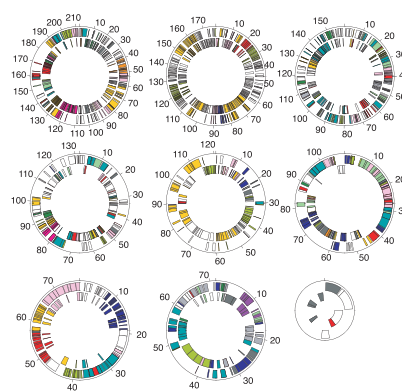
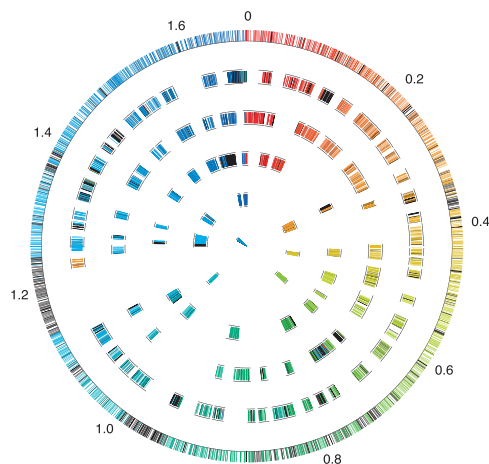
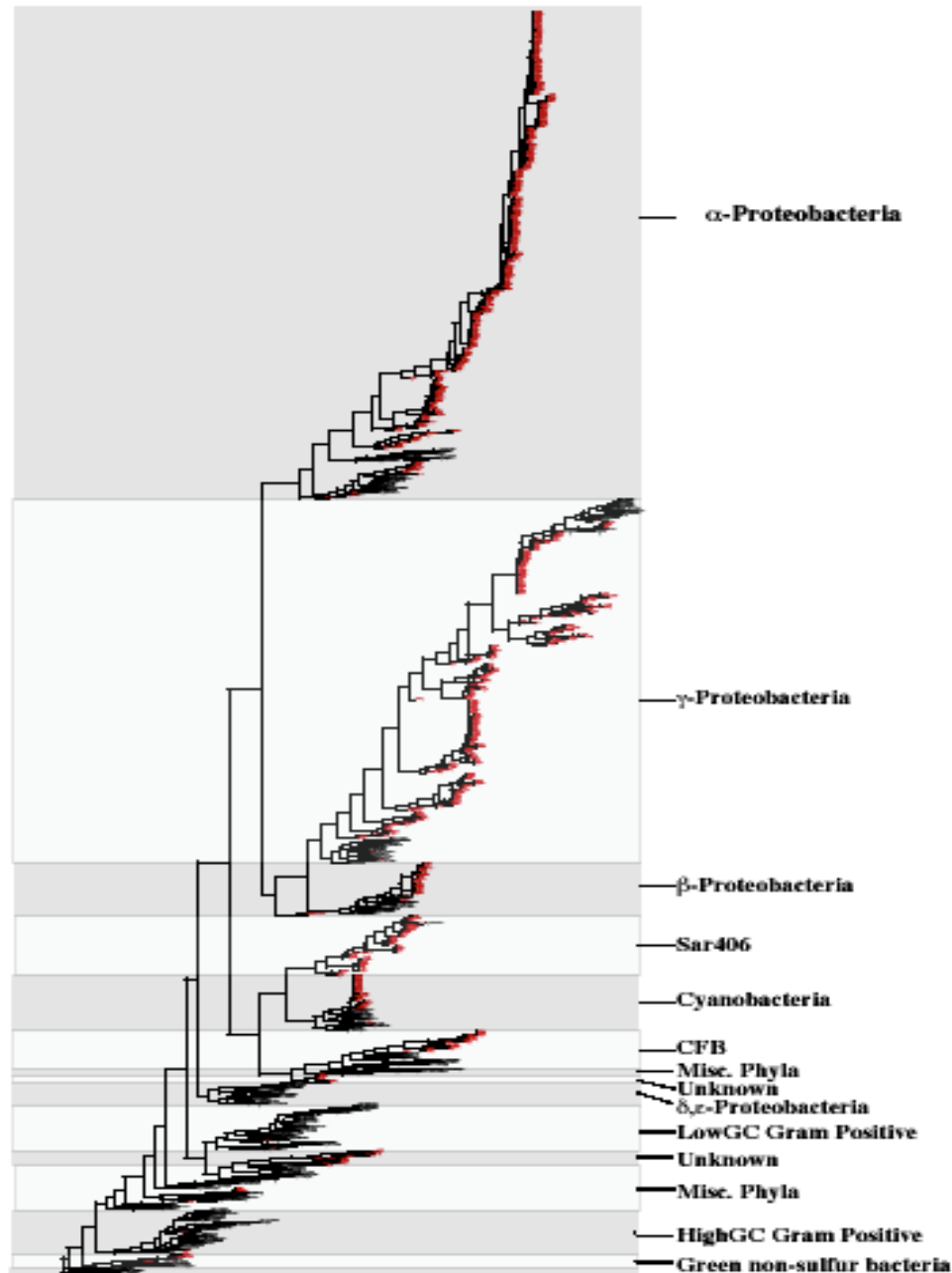


Fig. 4. Circular diagrams of nine complete megaplasmids. Genes encoded in the forward direction are shown in the outer concentric circle; reverse coding genes are shown in the inner concentric circle. The genes have been given role category assignment and colored accordingly: amino acid biosynthesis, violet; biosynthesis of cofactors, prosthetic groups, and carriers, light blue; cell envelope, light green; cellular processes, red; central intermediary metabolism, brown; DNA metabolism, gold; energy metabolism, light gray; fatty acid and phospholipid metabolism, magenta; protein fate and protein synthesis, pink; purines, pyrimidines, nucleosides, and nucleotides, orange; regulatory functions and signal transduction, dark green; transport and binding proteins, blue-green; genes with no known homology to other proteins and genes with homology to genes with no known function, white; genes of unknown function, gray; Tick marks are placed on 10-kb intervals.

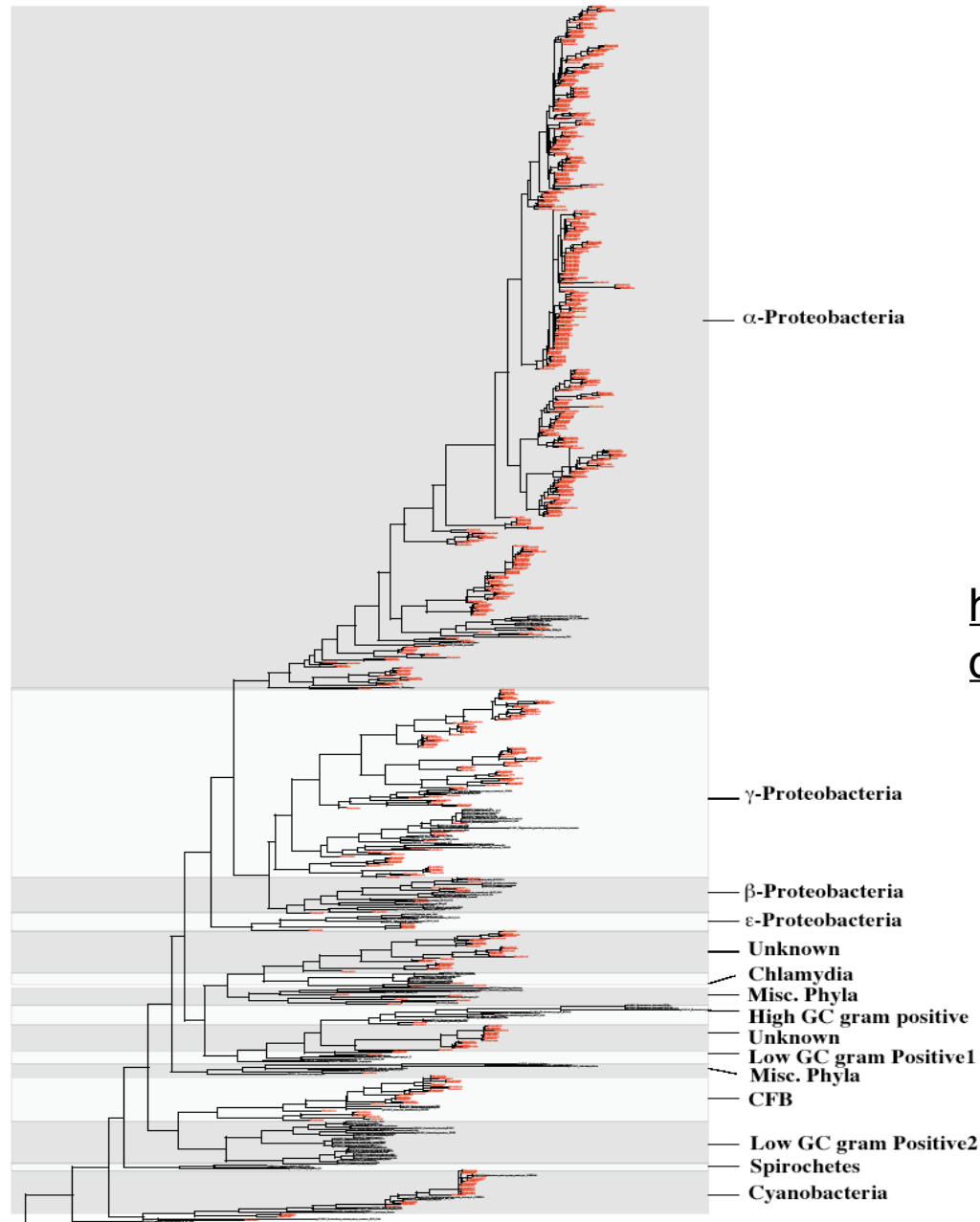
Fig. 2. Gene conservation among closely related *Prochlorococcus*. The outermost concentric circle of the diagram depicts the completed genomic sequence of *Prochlorococcus marinus* MED4 (11). Fragments from environmental sequencing were compared to this completed *Prochlorococcus* genome and are shown in the inner concentric circles and were given boxed outlines. Genes for the outermost circle have been assigned pseudospectrum colors based on the position of those genes along the chromosome, where genes nearer to the start of the genome are colored in red, and genes nearer to the end of the genome are colored in blue. Fragments from environmental sequencing were subjected to an analysis that identifies conserved gene order between those fragments and the completed *Prochlorococcus* MED4 genome. Genes on the environmental genome segments that exhibited conserved gene order are colored with the same color assignments as the *Prochlorococcus* MED4 chromosome. Colored regions on the environmental segments exhibiting color differences from the adjacent outermost concentric circle are the result of conserved gene order with other MED4 regions and probably represent chromosomal rearrangements. Genes that did not exhibit conserved gene order are colored in black.



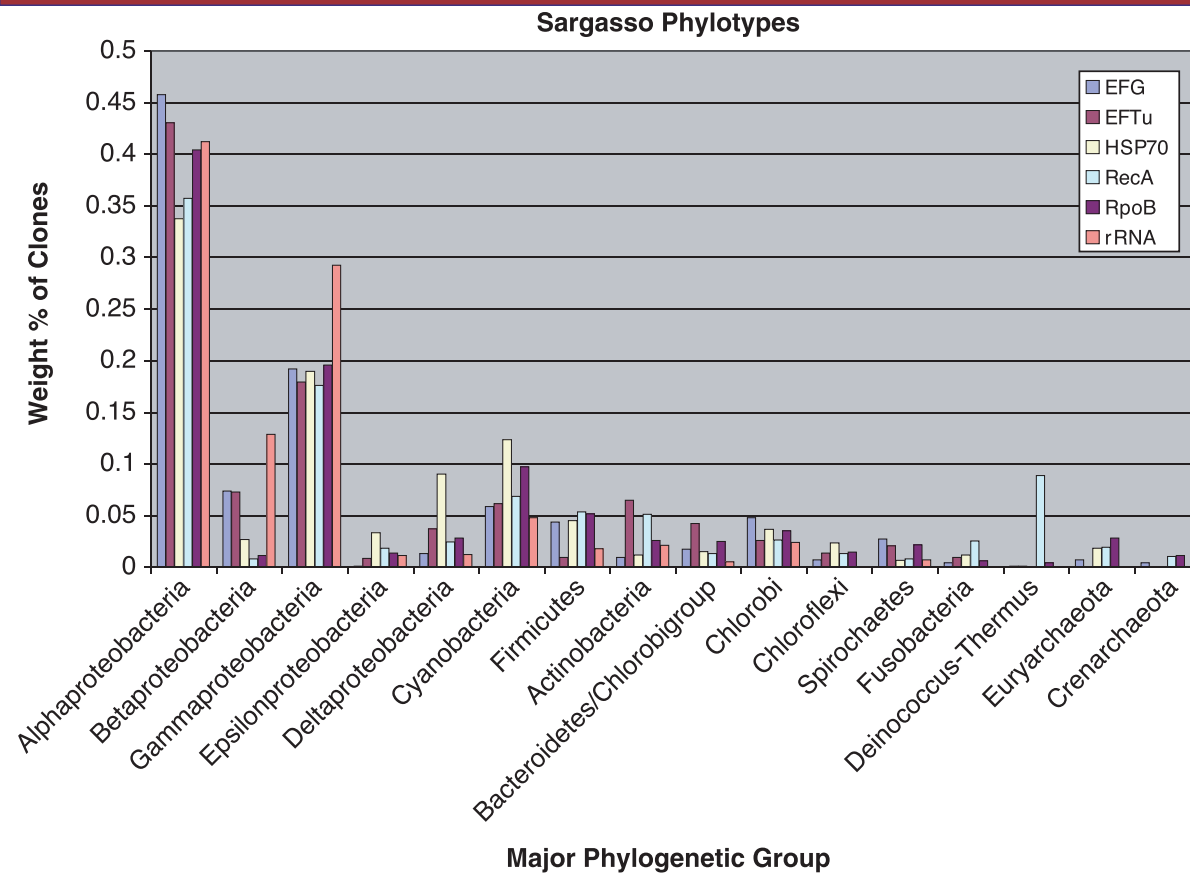


[http://www.sciencemag.org/
content/304/5667/66](http://www.sciencemag.org/content/304/5667/66)

Shotgun Sequencing Allows Alternative Anchors (e.g., RecA)



[http://www.sciencemag.org/
content/304/5667/66](http://www.sciencemag.org/content/304/5667/66)

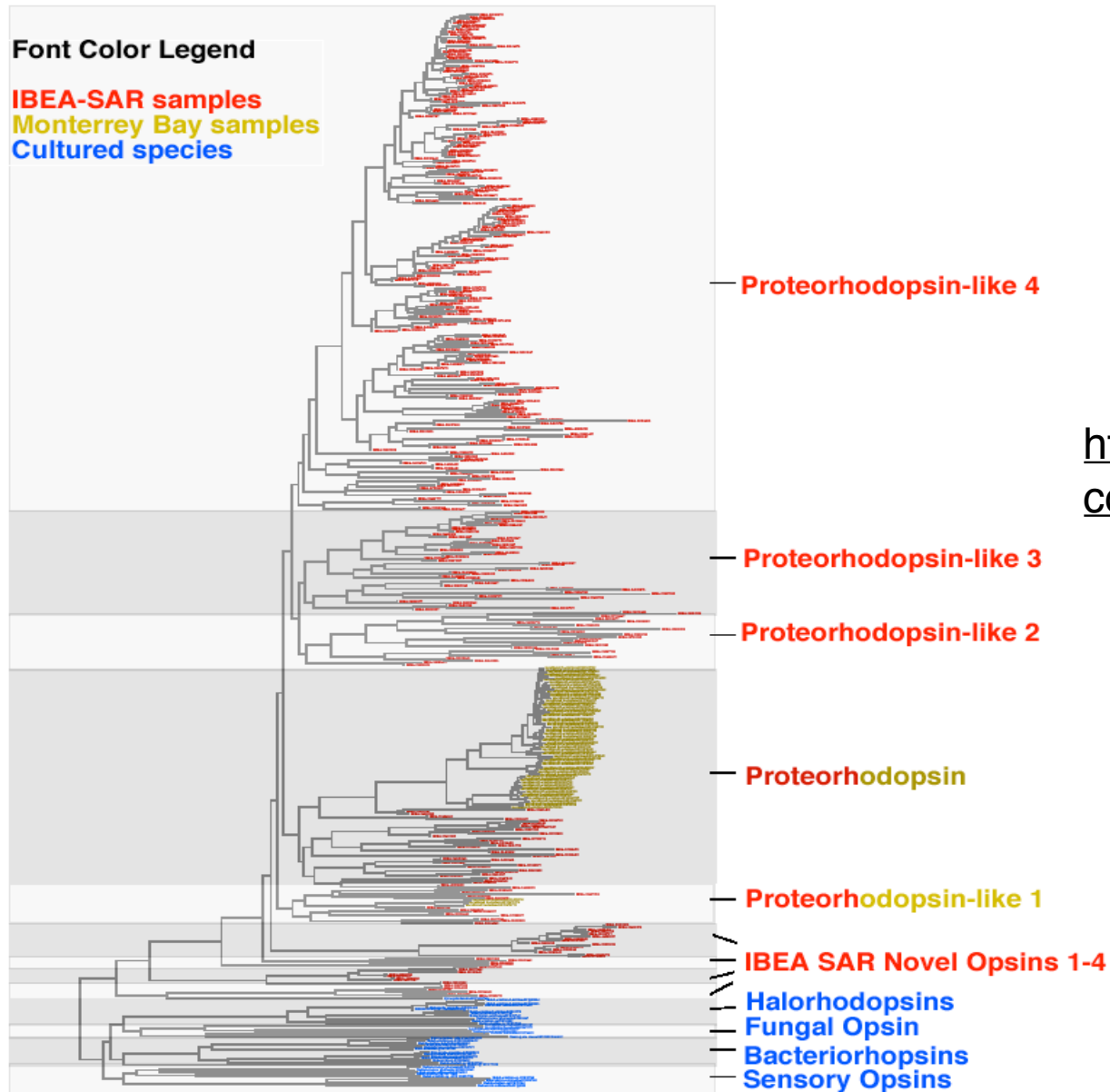


[http://www.sciencemag.org/
content/304/5667/66](http://www.sciencemag.org/content/304/5667/66)

Fig. 6. Phylogenetic diversity of Sargasso Sea sequences using multiple phylogenetic markers. The relative contribution of organisms from different major phylogenetic groups (phylotypes) was measured using multiple phylogenetic markers that have been used previously in phylogenetic studies of prokaryotes: 16S rRNA, RecA, EF-Tu, EF-G, HSP70, and RNA polymerase B (RpoB). The relative proportion of different phylotypes for each sequence (weighted by the depth of coverage of the contigs from which those sequences came) is shown. The phylotype distribution was determined as follows: (i) Sequences in the Sargasso data set corresponding to each of these genes were identified using HMM and BLAST searches. (ii) Phylogenetic analysis was performed for each phylogenetic marker identified in the Sargasso data separately compared with all members of that gene family in all complete genome sequences (only complete genomes were used to control for the differential sampling of these markers in GenBank). (iii) The phylogenetic affinity of each sequence was assigned based on the classification of the nearest neighbor in the phylogenetic tree.

- Can work well for individual genes

Functional Diversity of Proteorhodopsins?



[http://www.sciencemag.org/
content/304/5667/66](http://www.sciencemag.org/content/304/5667/66)

ARTICLE

Received 3 Nov 2010 | Accepted 11 Jan 2011 | Published 8 Feb 2011

DOI: 10.1038/ncomms1188

A bacterial proteorhodopsin proton pump in marine eukaryotes

Claudio H. Slamovits^{1,†}, Noriko Okamoto¹, Lena Burri^{1,†}, Erick R. James¹ & Patrick J. Keeling¹

<http://www.nature.com/ncomms/journal/v2/n2/abs/ncomms1188.html>

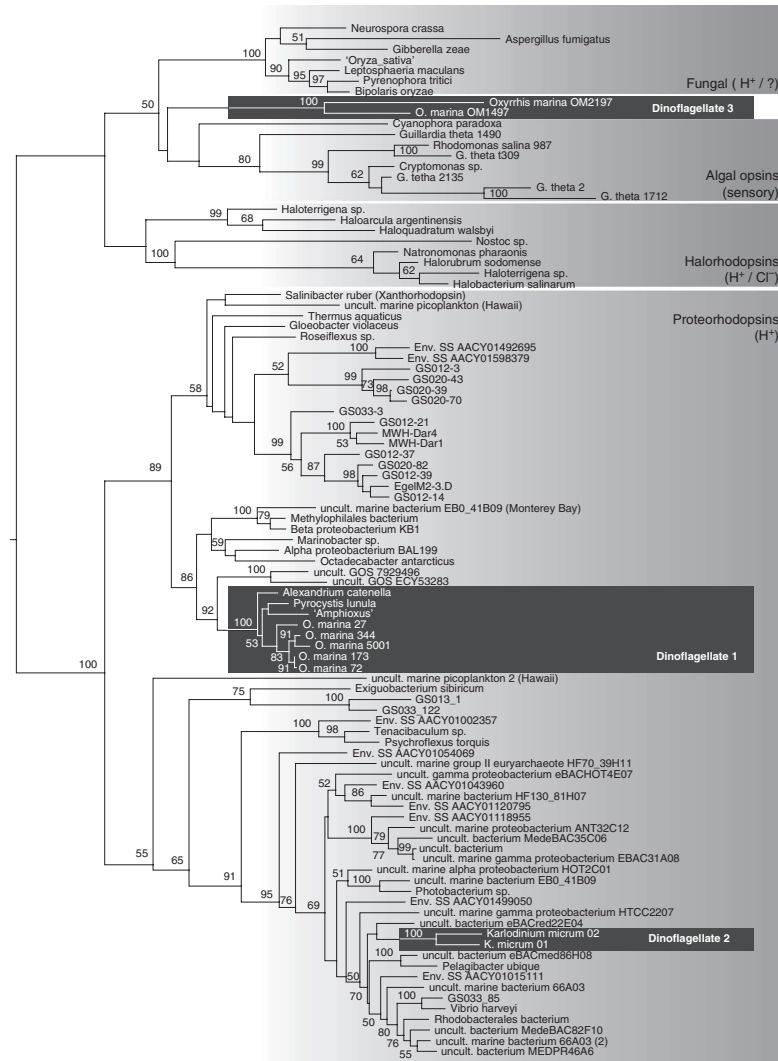


Figure 1 | Phylogenetic distribution of dinoflagellate rhodopsins. Protein sequences of 96 rhodopsins encompassing the known diversity of microbial (type I) rhodopsins from the three domains of life³⁰ were used to generate a maximum likelihood phylogenetic tree (See Supplementary Table S1 for accession numbers). Grey boxes distinguish the recognized groupings of type I rhodopsin, and the prevalent function in each group is shown: proton pumps (H⁺), sensory, chlorine pumps (Cl⁻) and unknown (?). Numbers indicate bootstrap support when ≥50% (over 300 replicates). Black boxes highlight dinoflagellate genes: Dinoflagellate 1 is a large group of proteorhodopsins present in diverse dinoflagellates; dinoflagellate 2 includes two *K. micrum* proteorhodopsin genes of independent origin; dinoflagellate 3 includes two *O. marina* genes related to algal sensory rhodopsins, probably of endosymbiotic origin.

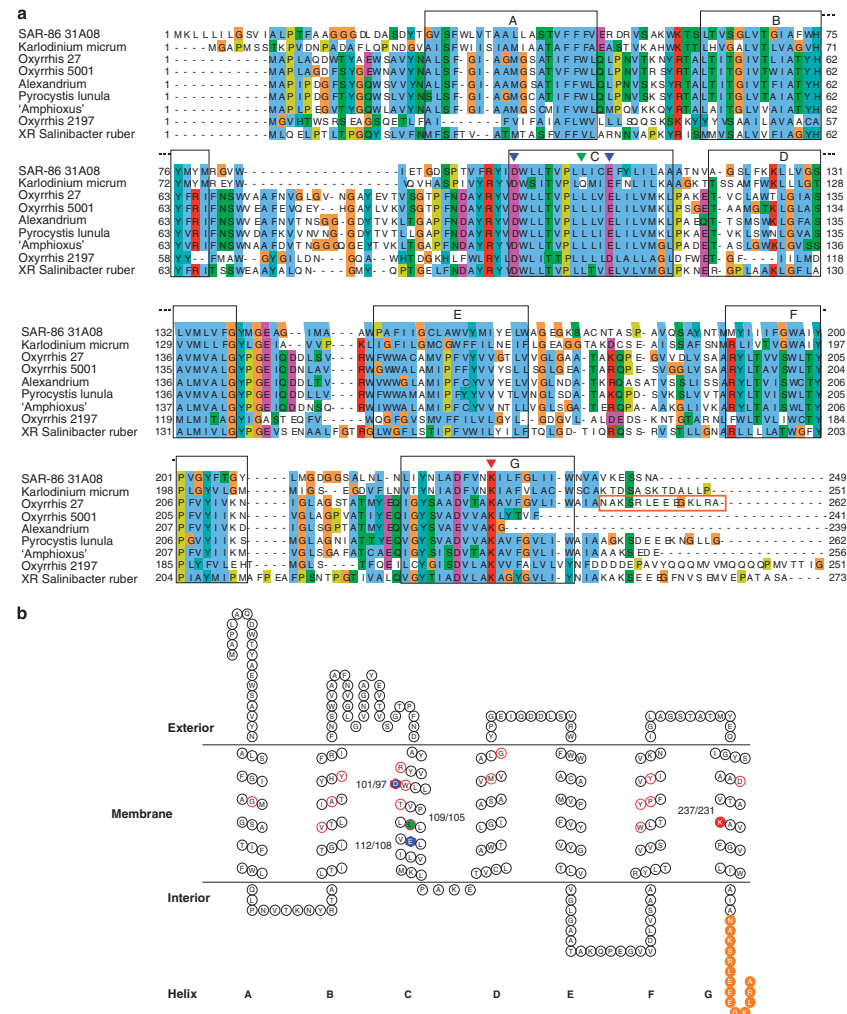


Figure 2 | Structural and comparative analysis of a dinoflagellate proteorhodopsin. (a) Amino-acid alignment of various rhodopsins from bacteria and dinoflagellates: SAR86-31A08 is a functionally characterized proteorhodopsin; *K. micrum* is a 'Dinoflagellate group 2' in Figure 1; *Oryziris* 2197 represents 'Dinoflagellate 3'; *XR S. ruber* is a Xanthorhodopsin; the remaining sequences belong to 'Dinoflagellate group 1'. Numbers on the right indicate residue number. Black rectangles show predicted transmembrane segments. Functional sites (coloured triangles): blue, proton acceptor and donor; green, spectral tuning ($105/105$ for green and Q for blue); red, lysine linked to the cofactor retinal. Epitope for antibody preparation is indicated with an orange rectangle. (b) Secondary structure of the *O. marina* proteorhodopsin predicted using PHD (<http://www.predictprotein.org/>). Single-letter amino-acid codes are used, and numbers correspond to the positions in *O. marina* and 31A08, respectively. Functional residues are highlighted as follows. Blue: proton acceptor ($D^{105/Q7}$) and proton donor ($E^{102/Q58}$); green: spectral tuning; red: retinal pocket; red filled: lysine linked to retinal; orange: epitope for antibody.

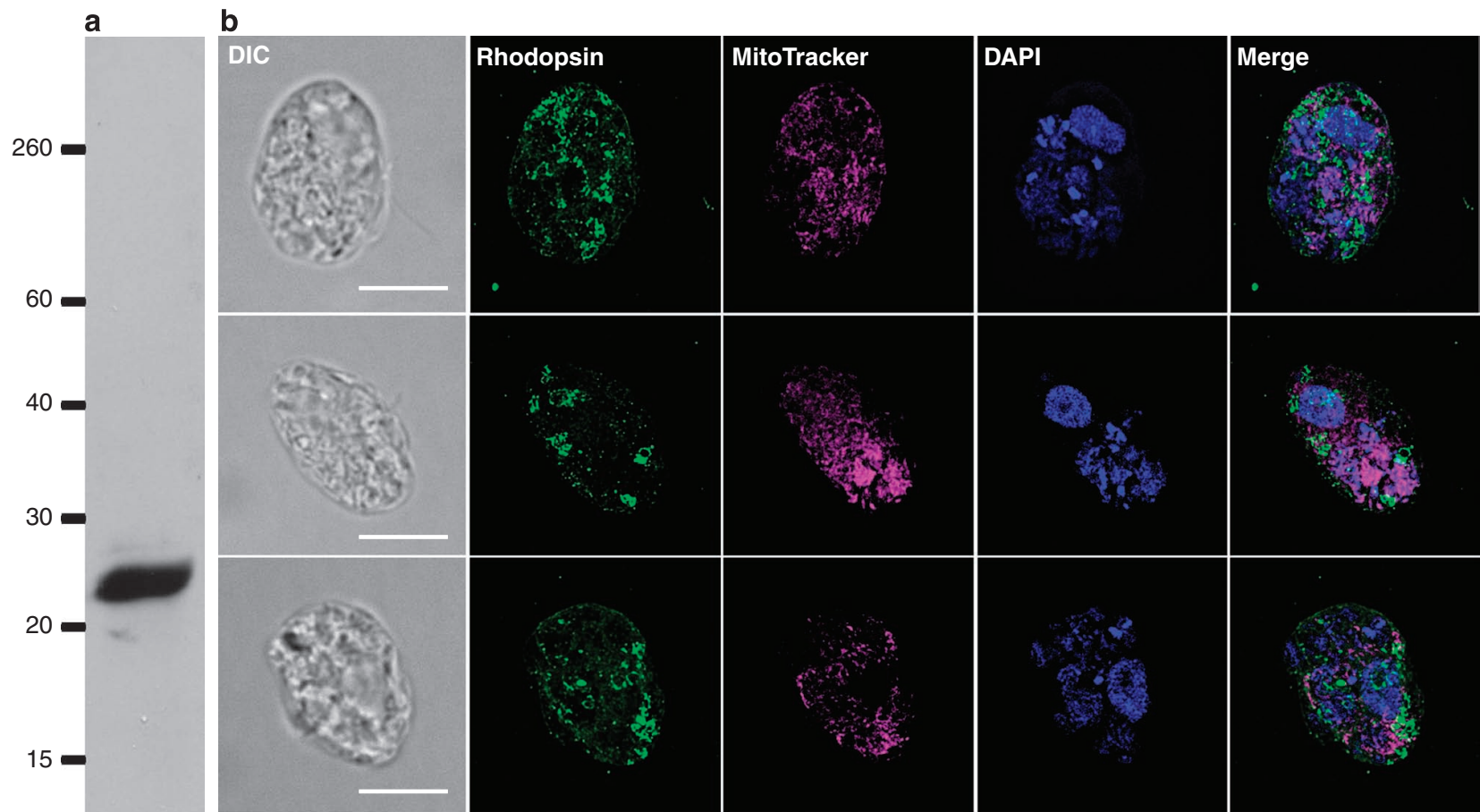


Figure 3 | Cellular localization of proteorhodopsin in *O. marina* cells. (a) Western blot of total *O. marina* protein probed with an antibody raised against the C-terminal peptide of the proteorhodopsin OM27 from *O. marina*. Expected protein size is 28 kDa. (b) Localization of proteorhodopsin in *O. marina* cells using immunofluorescence assay with the same antibody. Antibody signal forms small irregular and ring-like structures independent of mitochondria in *O. marina*. Three independent cells are shown, each showing (left to right) differential interference contrast (DIC) light micrograph, anti-OM27 proteorhodopsin, MitoTracker staining, Hoechst 33258 staining for DNA and a merge of all four. White bar = 10 μm. See Supplementary Movies 1–4 for a 360° video rendering. DAPI, 4,6-diamidino-2-phenylindole.

Primer

The Light-Driven Proton Pump Proteorhodopsin Enhances Bacterial Survival during Tough Times

Edward F. DeLong^{1,2*}, Oded Béjà^{3*}

<http://www.plosbiology.org/article/info%3Adoi%2F10.1371%2Fjournal.pbio.1000359#pbio-1000359-g002>

Table 1. Marine bacterial isolates and genome fragments containing proteorhodopsins.

Organism	Strain	General Group	Reference
Genomes			
Methylophilales	HTCC2181	Betaproteobacteria	GBMF
Rhodobacterales sp.	HTCC2255	Alphaproteobacteria	GBMF
<i>Vibrio angustum</i>	S14	Gammaproteobacteria	GBMF
<i>Photobacterium</i>	SKA34	Gammaproteobacteria	GBMF
<i>Vibrio harveyi</i>	ATCC BAA-1116	Gammaproteobacteria	GenBank # CP000789
Marine gamma	HTCC2143	Gammaproteobacteria	GBMF
Marine gamma	HTCC2207	Gammaproteobacteria	GBMF
<i>Cand. P. ubique</i>	HTCC1002	Alphaproteobacteria	GBMF
<i>Cand. P. ubique</i>	HTCC1062	Alphaproteobacteria	[26]
<i>Rhodospirillales</i>	BAL199	Alphaproteobacteria	GBMF
<i>Marinobacter</i>	ELB17	Gammaproteobacteria	GBMF
<i>Vibrio campbelli</i>	AND4	Gammaproteobacteria	GBMF
<i>Vibrio angustum</i>	S14	Gammaproteobacteria	GBMF
<i>Dokdonia donghaensis</i>	MED134	Flavobacteria	GBMF
<i>Polaribacter dokdonensis</i>	MED152	Flavobacteria	GBMF
<i>Psychroflexus</i>	ATCC700755	Flavobacteria	GBMF
<i>Polaribacter irgensii</i>	23-P	Flavobacteria	GBMF
<i>Flavobacteria bacterium</i>	BAL38	Flavobacteria	GBMF
BACs and fosmids			
HF10_05C07		Proteobacteria	[24]
HF10_45G01		Proteobacteria	[24]
HF130_81H07		Gammaproteobacteria	[24]
EB0_39F01		Alphaproteobacteria	[24]
EB0_39H12		Proteobacteria	[24]
EB80_69G07		Alphaproteobacteria	[24]
EB80_02D08		Gammaproteobacteria	[24]
EB0_35D03		Proteobacteria	[24]
EB0_49D07		Proteobacteria	[24]
EB0_50A10		Gammaproteobacteria	[24]
EB0_55B11f		Alphaproteobacteria	[24]
EB0_41B09		Betaproteobacteria	[24]
HF10_19P19		Proteobacteria	[17]
HF10_25F10		Proteobacteria	[17]
HF10_49E08		Planctomycetes	[24]
HF10_12C08		Alphaproteobacteria	[24]
HF10_29C11		Euryarchaea	[24]
MED13K09		unknown	[10]
MED18B02		unknown	[10]
MED35C06		unknown	[10]
MED42A11		unknown	[10]
MED46A06		unknown	[10]
MED49C08		unknown	[10]
MED66A03		unknown	[10]
MED82F10		unknown	[10]
MED86H08		unknown	[10]
RED17H08		unknown	[10]
RED22E04		unknown	[10]
eBACHOT4E07		Gammaproteobacteria	[25]
EBAC20E09		Gammaproteobacteria	[25]

Table 1. Cont.

Organism	Strain	General Group	Reference
HOT2C01		unknown	[8]
EBAC31A08		Gammaproteobacteria	[4]
ANT32C12		unknown	[8]
HF70_39H11_ArchHighGC		unknown	[12]
HF10_3D09_mediumGC		unknown	[12]
HF70_19612_highGC		unknown	[12]
HF70_59C08		unknown	[12]

Marine microbial isolates and large genome fragments from the environment GBMF, microbial genomes sequenced as part of the Gordon and Betty Moore Foundation microbial genome sequencing project (<http://www.moore.org/microgenome>), found to encode proteorhodopsin genes. The list includes whole genome sequences from a wide array of cultivated marine microorganisms (Genomes), as well as cloned large DNA fragments (BACs and fosmids) recovered directly from the environment. doi:10.1371/journal.pbio.1000359.t001

Box 1. A Decade of Proteorhodopsin Milestones

- 2000** •First proteorhodopsin gene found in uncultured SAR86 using metagenomics; proteorhodopsin light-driven proton pump activity confirmed in heterologous *E. coli* cells [4].
- 2001** •Proteorhodopsin presence confirmed directly in the ocean using laser flash photolysis [5].
- 2003** •Proteorhodopsin genes also found in other bacterial groups [8].
- 2004** •Enormous diversity of proteorhodopsin genes found in the Sargasso Sea using metagenomics [9].
- 2005** •Retinal biosynthesis pathways found in metagenomic data and confirmed using *E. coli* cells [10].
•Proteorhodopsin genes are found in '*Candidatus Pelagibacter ubique*' (SAR11), the most abundant bacterium on earth; environmental SAR11 proteorhodopsin presence confirmed using metaproteomics [11].
- 2006** •Proteorhodopsin genes found in uncultured marine Archaea [12].
- 2007** •First indication of proteorhodopsin light-dependent growth in cultured *Flavobacteria* [13] (see Figure 1 for colony morphologies and pigmentation).
- 2008** •Proteorhodopsin genes found in non-marine environments [14,15].
- 2010** •Proteorhodopsin phototrophy directly confirmed using a genetic system in marine *Vibrio* sp. [16]

Pretty proteorhodopsin

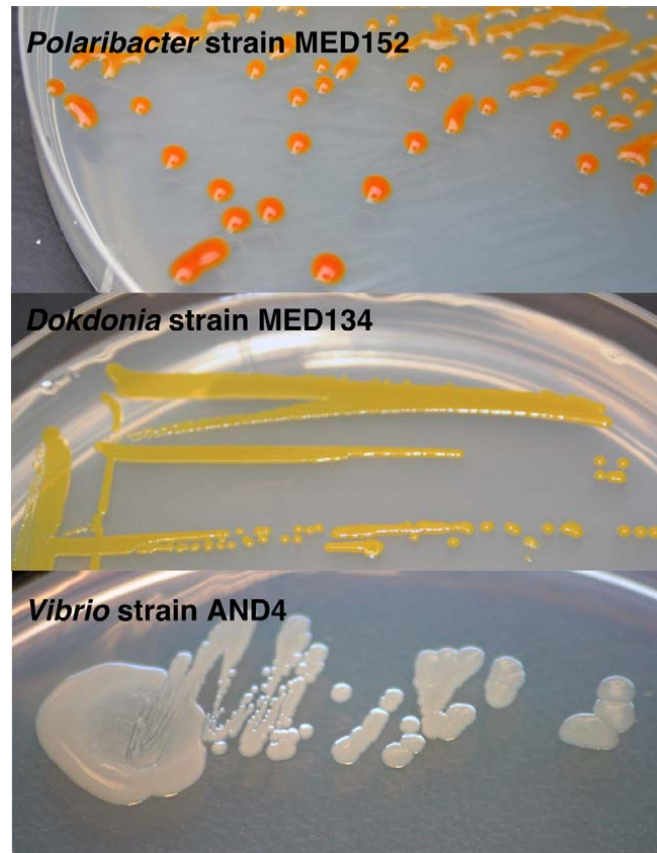


Figure 1. Various colony morphologies and coloration of different proteorhodopsin-containing bacteria used to study proteorhodopsin phototrophy. From top to bottom, the flavobacterium *Polaribacter dokdonensis* strain MED152 used to show proteorhodopsin light stimulated growth [13]; the flavobacterium *Dokdonia donghaensis* strain MED134 used to show proteorhodopsin light stimulated CO₂-fixation [23]; and *Vibrio* strain AND4 used to show proteorhodopsin phototrophy [16]; note the lack of detectable pigments in *Vibrio* strain AND4. However, when these vibrio cells are pelleted, they do show a pale reddish color, which is the result of proteorhodopsin pigments presence in their membranes. Photos are courtesy of Jarone Pinhassi.
doi:10.1371/journal.pbio.1000359.g001

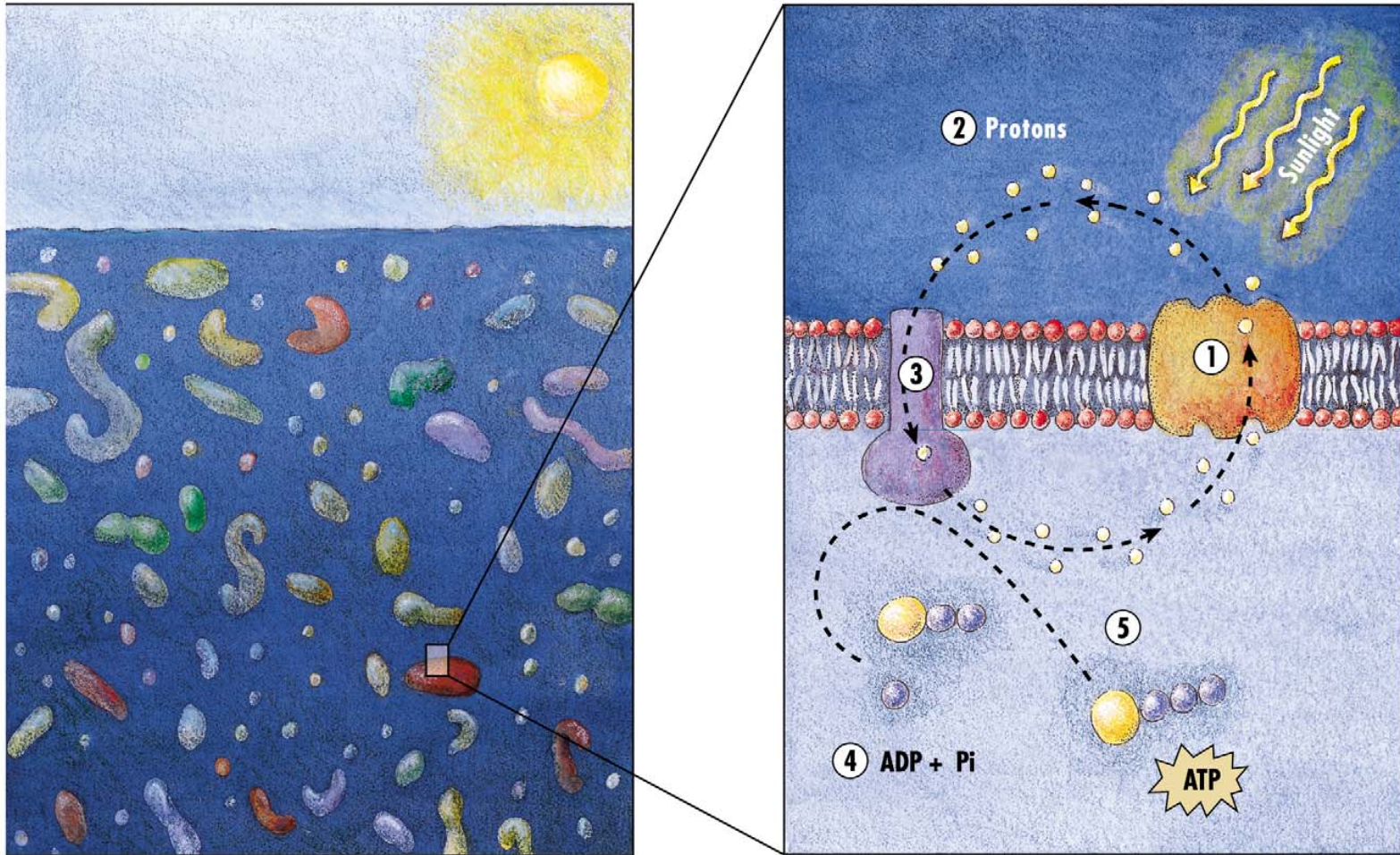


Figure 2. An artist's rendition of the fundamental arrangement of proteorhodopsin in the cell membrane. Left panel: a cartoon (not to scale) of planktonic bacteria in the ocean water column. Right panel: a simple view of one potential proteorhodopsin energy circuit. (1) Proteorhodopsin – uses light energy to translocate protons across the cell membrane. (2) Extracellular protons – the excess extracellular protons create a proton motive force, that can energetically drive flagellar motility, transport processes, or ATP synthesis in the cell. (3) Proton-translocating ATPase – a multi-protein membrane-bound complex that can utilize the proton motive force to synthesize 5. Adenosine triphosphate (ATP, a central high energy biochemical intermediate for the cell) from 4. Adenosine triphosphate (ADP, a lower energy biochemical intermediate). Illustration by Kirsten Carlson, © MBARI 2001.

doi:10.1371/journal.pbio.1000359.g002