

ORIGINAL ARTICLE

Adaptation and spectral tuning in divergent marine proteorhodopsins from the eastern Mediterranean and the Sargasso Seas

Gazalah Sabehi^{1,4}, Benjamin C Kirkup^{2,4}, Mira Rozenberg¹, Noga Stambler³, Martin F Polz² and Oded Béjà¹

¹Department of Biology, Technion-Israel Institute of Technology, Haifa, Israel; ²Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA and ³Faculty of Life Sciences, Bar-Ilan University, Ramat-Gan, Israel

Proteorhodopsins (PRs) phototrophy was recently discovered in oceanic surface waters. PRs have been observed in different marine environments and in diverse taxa, including the ubiquitous marine alphaproteobacterial SAR11 group and the uncultured gammaproteobacterial SAR86 group. Previously, two SAR86 PR subgroups, discovered in the Pacific Ocean, were shown to absorb light with different maxima, λ_{\max} 527 nm (green) and λ_{\max} 490 nm (blue) and their distribution was explained by prevailing light conditions – green pigments at the surface and blue in deeper waters. Here, we show that PRs display high diversity in geographically distinct patterns despite similar physical water column properties such as mixing and light penetration. We compared summer and winter samples representing stratified and mixed conditions from both the Mediterranean and Sargasso Sea. As expected, in the Mediterranean Sea, green pigments were mainly confined to the surface and the percentage of blue pigments increased toward deeper samples; in the Sargasso Sea, unexpectedly, all PRs were of the blue type. As an additional result, both locations show seasonal dependence in the distribution of different PR families. Finally, spectral tuning was not restricted to a single PR family as previously reported but occurs across the sampled PR families from various microbial taxa. The distribution of tunable PRs across the PR tree suggests that ready adaptability has been distributed widely among microorganisms, and may be a reason that PRs are abundant and taxonomically widely dispersed.

The ISME Journal (2007) 1, 48–55; doi:10.1038/ismej.2007.10

Subject Category: microbial population and community ecology

Keywords: flavobacteria; proteorhodopsin; retinal; SAR11; SAR86; spectral-tuning

Introduction

Proteorhodopsins (PRs) (Béjà *et al.*, 2000, 2001) are bacterial retinal-binding membrane pigments that belong to the microbial rhodopsin superfamily (Spudich *et al.*, 2000) and are predicted to have an important role in supplying light energy for microbial metabolism in marine ecosystems (Béjà *et al.*, 2000, 2001; Sabehi *et al.*, 2005; Martinez *et al.*, 2007;

Walter *et al.*, 2007). PRs have been observed in different ocean regions (Béjà *et al.*, 2000, 2001; de la Torre *et al.*, 2003; Man *et al.*, 2003; Sabehi *et al.*, 2003, 2004, 2005; Venter *et al.*, 2004; Frigaard *et al.*, 2006) and are found in diverse taxonomic backgrounds, including the ubiquitous marine gamma-proteobacterial SAR86 (Béjà *et al.*, 2000; Sabehi *et al.*, 2004, 2005) and alphaproteobacterial SAR11 (Giovannoni *et al.*, 2005a,b; Sabehi *et al.*, 2005) groups, as well as in marine *Bacteroidetes* (Venter *et al.*, 2004; Gómez-Consarnau *et al.*, 2007), planktonic *Archaea* (Frigaard *et al.*, 2006) and other microbial taxa (de la Torre *et al.*, 2003; Sabehi *et al.*, 2003, 2005; Venter *et al.*, 2004; McCarren and DeLong, 2007). While previous work did not detect light enhanced growth in PR-containing SAR11 (Giovannoni *et al.*, 2005a), or in the gammaproteobacterial SAR92 (Stingl *et al.*, 2007) isolates grown

Correspondence: Professor MF Polz, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

E-mail: mpolz@mit.edu or

Professor O Béjà, Department of Biology, Technion-Israel Institute of Technology, Haifa 32000, Israel.

E-mail: beja@tx.technion.ac.il

⁴These authors contributed equally to this work.

Received 15 January 2007; revised and accepted 19 February 2007

in seawater, significant enhancement of both growth rate and yield was recently reported in PR-expressing marine *Bacteroidetes* (Gómez-Consarnau *et al.*, 2007).

PR spectral tuning was first observed among genes amplified directly from the environment by using highly specific polymerase chain reaction (PCR) primers (Béjà *et al.*, 2001) to the gammaproteobacterial SAR86 group. These genes produced pigments of two distinct absorption spectra where 'green-absorbing' PR (GPR) and 'blue-absorbing' PR (BPR) types shared >78% of their amino-acid residues (200 out of 247 are identical). Moreover, a single amino-acid change at position 105 (leucine for green, glutamine for blue) functions as a spectral tuning switch to account for most of the spectral difference (Man *et al.*, 2003). Here, we sought to determine how common spectral tuning is among PRs from diverse taxonomic backgrounds, and to what extent their distribution correlates to distinct oceanic regimes by comparing PR sequence diversity from different depths and different seasons from two oligotrophic seas, the eastern Mediterranean and Sargasso Seas.

Materials and methods

Sample collection, DNA preparation and storage

Environmental samples were collected from the BATS station (31° 40'N, 64° 10'W) in the Sargasso Sea (mixed BATS114a station (March 1998) and stratified BATS118 station (July 1998)) and from the H01 station (32° 54'N, 34° 55'E) in the Mediterranean Sea (mixed H01 station in January 2006 and stratified H01 station in May 2003) from depths of 0, 40, 80 m and 0, 20, 55 m, respectively (see Supplementary Figure S1 for temperature profiles collected in the different stations). BATS DNA was prepared from 50 to 70 l according to Gordon and Giovannoni (1996), while Mediterranean Sea DNA was prepared from 20 l according to Massana *et al.* (1997). All samples were collected on 0.2 µm-pore-size filters. Samples were stored in liquid nitrogen or in a -80°C freezer before and after DNA extraction.

PR PCR amplification

PR fragments (~330 bp) were amplified by PCR from DNA extracts using a subset of the degenerate primers previously used by us to amplify environmental PRs (Sabehi *et al.*, 2005). The degenerated primers RYIDWLFwd, (5'-MGNTAYATHGAYTGGYT-3') and GWAIYPrev, (5'-GGRTADATNGCCCANCC-3'), targeting the conserved RYIDWL and GWAIYP regions in the PR proteins, respectively, using a high-fidelity proof reading polymerase mixes BIO-X-ACT (Bioline, Berlin, Germany). PCRs were performed in a total volume of 25 µl containing 10 ng of template DNA, 200 µM dNTPs, 2.6 mM MgCl₂, 0.8 µM primers and 1.2 U of DNA polymerase. The amplification

conditions comprised steps at 95°C for 4 min, and 30 cycles at 94°C for 30 s, 52°C for 30 s, and 68°C for 30 s followed by one step of 2 min at 68°C. PCR products were cloned using the QIAGEN-PCR cloning kit (Qiagen, Hilden, Germany).

PR phylogeny

The PR tree was constructed according to Sabehi *et al.* (2005). PR proteins were identified in GenBank including predicted proteins from the Sargasso Sea assemblies (Venter *et al.*, 2004) using BLAST (Altschul *et al.*, 1990) searches with representatives of previously identified PR-like protein families as query sequences. PR proteins were aligned using CLUSTALx (Thompson *et al.*, 1997), and a neighbor-joining phylogenetic tree was inferred using the neighbor programs of PAUP* (Swofford, 2002).

Estimation of sampling efficiency

The efficiency of sequence recovery from the different samples was estimated based on DNA and protein sequences, respectively. In both cases, non-parametric richness estimators were calculated to estimate the total number of operational taxonomic units (OTUs) defined as clusters of different DNA or protein sequence similarity thresholds. These estimates were then compared to the observed number of OTUs to determine the percent coverage. The OTUs were derived from the complete data set after which the individual samples were broken out so that OTUs are defined consistently and can be traced across all samples. We report the richness obtained by using the abundance-based coverage estimator (ACE) richness estimator since it consistently gave the highest diversity estimates and can thus be treated as an upper bound of diversity estimates. For DNA-based estimates, OTUs were assembled at 1%, 5% and 20% DNA sequence divergence using CLUSTERER (Klepac-Ceraj *et al.*, 2006). For different samples, total coalescence of the OTUs occurred at between 40% and 50% DNA sequence divergence. The EstimateS program (Colwell, 2005) was used to determine non-parametric richness estimators of total diversity in each sample. The second method used protein alignments to create distance measures, employing a JTT dynamic matrix model. Such correction does not assign constant weights to amino-acid changes; however, a distance of 0.01–0.02 roughly corresponds to a single amino-acid change, and a distance of 2.81 was sufficient to coalesce all the OTUs. The number of OTUs was estimated for JTT distances of 0.02, 0.4 and 0.6 – roughly equivalent to 1%, 20% and 30% amino-acid difference. The distance matrix was created using PROTDIST within the Phylip package (Felsenstein, 1989) and then imported into DOTUR to calculate the non-parametric richness estimates (Schloss and Handelsman, 2005).

Irradiance measurements in the Sargasso and Mediterranean Seas

Irradiance at 5 and 40 m depth in the Sargasso B114a BATS station were measured at 12 wavelengths (410, 441, 465, 488, 510, 520, 555, 565, 589, 625, 665 and 683 nm). The data was extracted from the Bermuda Bio-Optics Project (<http://www.icess.ucsb.edu/bbop/bbop.html>) and was collected from the monthly BBOP casts at station B114a (12 March 1998). Irradiance, at 5 and 55 m depth from the Mediterranean Sea station To (32° 09'N 34° 14'E) was measured at six wavelengths (412, 443, 490, 555, 665 and 694 nm). The Mediterranean spectrums (3 October 2006) were measured further south to the station where the DNA samples were retrieved due to instrument availability.

Accession numbers

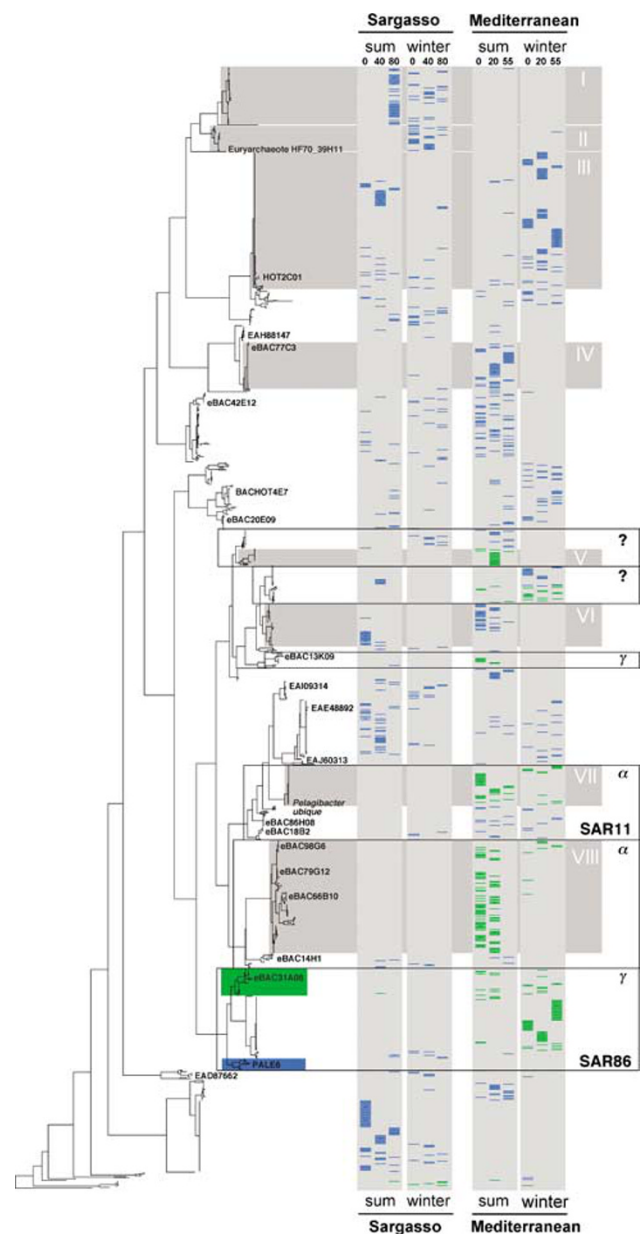
The sequences reported in this study are deposited with GenBank under accession numbers DQ203297–DQ203852, DQ339158–DQ339245, DQ339247–DQ339288, DQ339290–DQ339341, DQ339343–DQ339417, DQ422150–DQ422447.

Results and discussion

Diverse PR sequences were observed in both Mediterranean and Sargasso Sea samples using newly designed degenerate PR primers (Sabehi *et al.*, 2005) (Figure 1 and Supplementary Figure S2). The PR gene fragments fell into several families not restricted to the PR groups previously reported using non-degenerate PCR primers (Béjà *et al.*, 2001; Sabehi *et al.*, 2003) and the overall diversity of the PR sequences was comparable to that previously detected in the Sargasso Sea metagenome (Venter *et al.*, 2004). No apparent bias was also observed regarding any potential preference of our primers

Figure 1 Relationship between Mediterranean Sea and Sargasso Sea GPR and BPR proteins retrieved from different depths and seasons. A neighbor-joining phylogenetic tree of Mediterranean PR protein sequences amplified from 0, 20 and 55 m and Sargasso PRs from 0, 40 and 80 m from mixed and stratified samples. Names of the PR sequences were removed for clarity. Environmental sequences are represented by blue or green bands corresponding to their predicted absorption maxima to illustrate their distribution in different depths and seasons. The two subfamilies previously reported to be spectrally tuned (Béjà *et al.*, 2001) are marked in blue and green boxes. SAR86 and SAR11 PR families, as well as few different PR representative's names from the Pacific Ocean (Béjà *et al.*, 2001), Sargasso Sea shotgun-sequencing project (Venter *et al.*, 2004), Mediterranean Sea bacterial artificial chromosomes (BACs) (Sabehi *et al.*, 2003, 2004, 2005), *Pelagibacter ubique* (Giovannoni *et al.*, 2005a) and marine *Archaea* (Frigaard *et al.*, 2006) are indicated to serve as orientation marks. Different groups discussed in the text are marked with gray boxes while groups in which spectral tuning is suggested are marked with an empty box. *Pyrocystis lunula* (Okamoto and Hastings, 2003) was used to root the tree and is not shown. A modified B&W version of this figure, which could be visualized by people suffering from color blindness, is presented in Supplementary Figure S5.

toward low G + C amplicons (Suzuki and Giovannoni, 1996) (Supplementary Figure S2), further supporting the adequacy of these primers to sample across different subpopulation of the PR family. However, while new groups were retrieved in samples from both Mediterranean and Sargasso Sea environments, some others previously reported from the Sargasso Sea metagenome project were not detected (Venter *et al.*, 2004). This difference may at least be partly explained due to the different collection time of the samples (March and July 1998 vs February 2003 (Venter *et al.*, 2004)), as marine microbial populations are known to change in a seasonal manner (see Giovannoni and Rappé (2000) for a review).



Traditionally, rarefaction analysis has been used to estimate the proportion of the natural population that has been sampled at various sites. This method becomes more accurate in estimating the total diversity the more samples are taken. However, it is required that the samples be uniform from within the same population. In the current case, there are two concerns. First, the primers do not represent the full degeneracy available within the presumed conserved amino-acid sequences (GWAIP and RYIDWL). Thus, the real entity being sampled is potentially a subset of the PRs present in the environment. The assumption that these primers adequately represent the PR diversity at those sites is thus noted. Second, no repeated samples were taken from the same depth, location and time. This presents two options for rarefaction analysis: estimate the unsampled population from a single sample at each location or assume that the samples are from a relatively unstructured population and use them as 12 replicates from a single source. In this study, each of these methods was used.

To estimate how well the PR diversity in the clone libraries has been sampled, we clustered the sequences into OTUs and applied the ACE non-parametric richness estimator. Treating all samples as a uniform population shows an estimated coverage of 43%, 59% and 79% for OTUs defined as clusters of 1%, 5% and 20% sequence divergence, respectively (Table 1). Similar results were obtained for protein sequence-based OTUs where the coverage estimates rise up to 93% for distances roughly equivalent to amino-acid sequence divergence of 30% (Supplementary Table S1). We further calculated OTU diversity estimates for each individual sample to determine whether assessing the

sequences as a single population would introduce a systematic bias in the estimates (Table 2). This showed varying degree of coverage but good agreement with the overall picture of sample coverage of ~50% of OTUs with internal divergence of 1% DNA sequence. Thus, the analyses suggest that roughly half of the sequence diversity was sampled when OTUs are stringently defined but also indicates that at amino-acid differences of ~30% nearly complete coverage was achieved.

f-LibShuff (Schloss *et al.*, 2004) was employed to estimate whether the populations from the different samples were overlapping or non-overlapping. This analysis (Supplementary Figure S3) indicates that while most samples are drawn from distinct populations of PRs, samples at different depths during the times of ocean mixing (January/March, respectively) were drawn from statistically indistinguishable populations. Samples from the same locations and times at the stratified time points appear to have significantly more structure, particularly when comparing the extremes of depth. These results

Table 1 Coverage estimates in combined samples based on comparison of number of OTUs calculated by the ACE non-parametric richness estimator and OTUs detected in samples for varying percent DNA sequence differences

Divergence within OT (%)	Observed number of OTUs	Estimated number of OTUs	Coverage (%)
1	489	1127	43
5	163	276	59
20	74	95	79

Abbreviations: ACE, abundance-based coverage estimator; OTUs, operational taxonomic units.

Table 2 Comparison of diversity estimates and observed numbers of OTUs in individual samples by the ACE non-parametric richness estimator

	1% within OTU divergence		5% within OTU divergence		20% within OTU divergence	
	Number detected	Number estimated	Number detected	Number estimated	Number detected	Number estimated
<i>Sargasso Sea</i>						
Summer						
0 m	14	35	6	6	3	3
40 m	20	53	9	10	4	5
80 m	31	61	13	21	10	12
Winter						
0 m	29	52	13	17	6	7
40 m	18	46	11	14	6	7
80 m	22	29	9	10	5	5
<i>Mediterranean</i>						
Summer						
0 m	29	36	14	15	5	5
20 m	52	66	15	19	6	7
55 m	31	55	11	18	7	10
Winter						
0 m	27	50	13	17	7	10
20 m	22	31	7	8	3	4
50 m	20	24	9	10	5	6

Abbreviations: ACE, abundance-based coverage estimator; OTUs, operational taxonomic units.

indicate that a more detailed analysis of molecular variance will have a large amount of variance with which to work; further the limited degree of similarity among the samples presents a challenge to that analysis. It is also probable that a large fraction of PR sequences in these environments remain unknown, stimulating further research to shed light on this interesting group of proteins. Analysis of distribution of these OTUs and their richness in different locations awaits more complete sampling.

While some PR families were detected in both the Mediterranean Sea and the Sargasso Sea (see for example groups III and VI in Figure 1), others were unique to one location (groups I, II, VII and VIII in Figure 1). However, geographic isolation does not appear to play a strong role in PR distribution since representatives of most families appear in both locations. This may be the result of high potential for lateral transfer of PR genes (Frigaard *et al.*, 2006; Sharma *et al.*, 2006; Martinez *et al.*, 2007; McCarren and DeLong, 2007); alternatively, it may be possible that PR genes reside in many widely distributed ecological generalists in addition to locally adapted specialists.

Similarly, some PR families were distributed differently with respect to seasons within a given location. However, a much stronger seasonal pattern is evident in the Mediterranean than in the Sargasso Sea consistent with stronger seasonal variation in the former than in the latter. Some families were abundant in winter samples (e.g. group II in Figure 1) or summer samples only (e.g. group IV in Figure 1), while others were detected in both seasons. The composition of the microbial community harboring PR genes may be strongly influenced by water column stratification. During the summer stratification, gradients in light spectrum and intensity, temperature, salinity and nutrient concentration may develop, while conditions during winter are much more uniform due to deep mixing of the water column. This uniformity is reflected in the distribution of the different PR families as the families detected in winter samples are found at all three depths (Figure 1).

Light intensity in the water column declines with increasing depth, and is absorbed or scattered by seawater, yellow substance dissolved matter and particles, including phytoplankton (Thurman and Trujillo, 1999). Red light, absorbed mainly by water molecules, disappears in the first few meters; the green light penetrates further but not to deep water, while blue light is maintained throughout the euphotic zone (Figure 2). Therefore, PR sequences were further sorted based on their 'blue' and 'green' signatures (i.e. leucine or glutamine at position 105). Eight environmental sequences (0.7%) were found to contain a methionine in this position, which preliminary data indicate to absorb in the green range (Gómez-Consarnau *et al.* (2007) and Sabehi and Béjà, data not shown). Methionine contains a

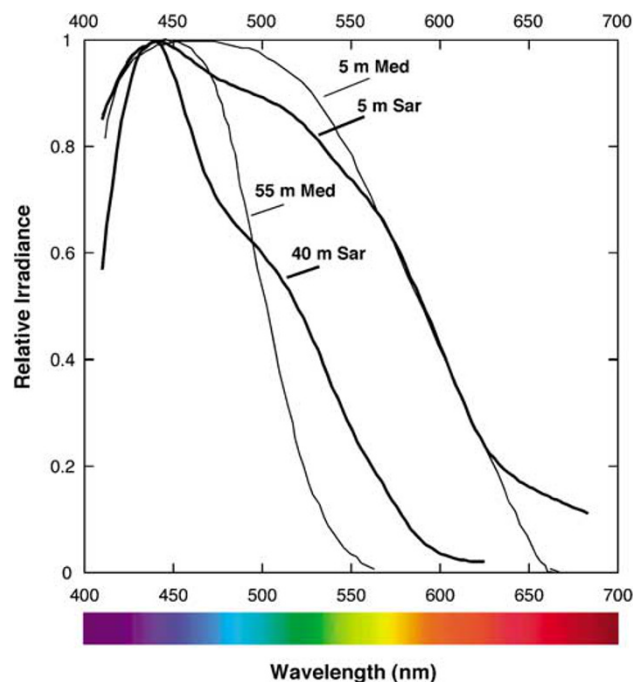


Figure 2 Relative irradiance in the Sargasso and Mediterranean Seas. Relative irradiance at 5 and 40 m depth in the Sargasso B114a BATS station (bold lines). Irradiance is plotted relative to irradiance at 441 nm; and relative irradiance, at 5 and 55 m depth at the Mediterranean Sea T0 station (thin lines). Irradiance is plotted relative to irradiance at 443 nm.

non-polar side chain similar to leucine, which might explain the similarity in green light absorption. However, sequences with methionine at position 105 were not included in the spectral tuning statistics because no further spectral analyses were performed on cells containing those exact PR sequences and the inference is thus relatively uncertain.

PR sequences from the Mediterranean Sea sample collected in May 2003 showed the expected stratification with respect to depth; GPRs were found mainly in surface waters (0 and 20 m, 61% and 44% respectively), while BPRs were found at all depths (Figure 3). The detection of both BPRs and GPRs in Mediterranean surface waters suggests that both types can contribute to growth advantage from utilization of light in the blue and green range and that both pigment types may thus stably coexist. A similar explanation was recently suggested by Stomp *et al.* (2004) when reporting the coexistence of closely related picocyanobacteria of the *Synechococcus* group. During the winter mixing, the bacterial population would be primarily exposed to blue light conditions as they cycle between deeper and surface water. Consistent with this expectation, BPR dominate and are stably distributed with depth in samples collected from the Mediterranean Sea during January 2006 when the water column mixes deeply.

Phylogenetic analysis of the PR sequences retrieved from the Mediterranean sample ($n=396$)

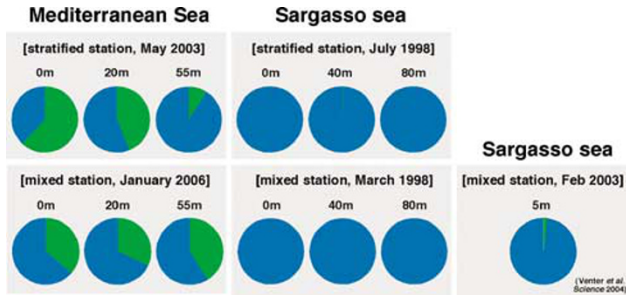


Figure 3 GPRs and BPRs depth distribution in the Mediterranean and Sargasso Seas. Pie charts represent percentages of GPRs (marked in green) and BPRs (marked in blue) at different depths in the stations sampled. Data extracted from the Sargasso Sea shotgun-sequencing project (Venter *et al.*, 2004) are presented on the right. The percentages presented in the different pies are (GPRs%, BPRs%, n = number of clones), (61, 39, n = 147), (44, 56, n = 169), (9, 91, n = 75) and (36, 64, n = 90), (31, 69, n = 103), (41, 59, n = 103) in the 0, 20 and 55 m samples of the stratified and mixed Mediterranean station, respectively. In the Sargasso Sea station, only BPRs were detected except for the stratified 40 m sample, which contained 0.01% GPRs (one out of total 414 sequences). In the data from the Sargasso Sea metagenome, 2% GPRs were detected (n = 291). A modified B&W version of this figure, which could be visualized by people suffering from color blindness, is presented in Supplementary Figure S6.

suggests that spectral tuning has arisen independently multiple times (Figure 1). The previously reported spectral tuning in SAR86 group family from the Pacific Ocean (Béjà *et al.*, 2001) was shown to be driven by Darwinian selection (Bielawski *et al.*, 2004) and we speculate that similar analyses will show that the same selection applies to other PR families with spectral tuning reported here. Of special interest is the spectral tuning observed within the cosmopolitan SAR11 family (Figure 4). In this abundant family, BPRs and GPRs share as much as 86% of their amino acids (e.g. *Pelagibacter* GPR vs Sargasso Sea EAI06928 BPR), a value somewhat higher than observed in the SAR86 group (with a maximum of 78% identity; Béjà *et al.*, 2001). Interestingly, several BPRs and GPRs containing SAR11 members also share common gene organization in the proximity of the PR gene (a ferredoxin and a thioredoxin reductase, compare, for example, *Pelagibacter ubiquus* and eBAC86H08 gene organizations on the right panel of Figure 4).

In contrast to the observation from the Mediterranean waters, no ‘green’ signatures were observed in the Sargasso Sea samples despite similarity in patterns of downwelling irradiance or relative irradiance with depth (Figure 2 and Supplementary Figure S4). All sequences (n = 414, collected at the same station in different seasons) contained the ‘blue’ signature only, irrespective of depth and stratification condition they were collected from (mixed BATS 114a station (March 1998) and stratified BATS 118 station (July 1998); see Supplementary Figure S1 for temperature profiles collected in the different stations). We performed a similar analysis on the data from the Sargasso shotgun-sequencing project (Venter *et al.*, 2004; collected in

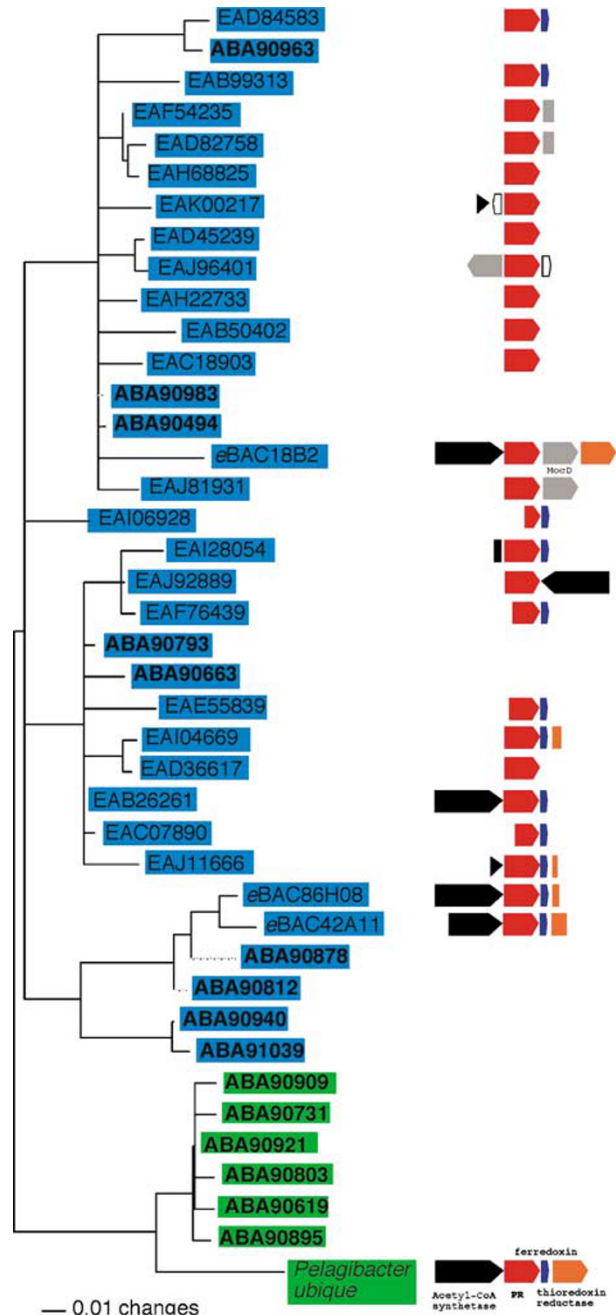


Figure 4 SAR11-like GPRs and BPRs. Neighbor-joining phylogenetic tree of PR proteins from the SAR11 family. Next to each BAC (Sabehi *et al.*, 2005), Sargasso Sea shotgun-sequencing singleton (Venter *et al.*, 2004) or *Pelagibacter ubiquus* (Giovannoni *et al.*, 2005a) PR sequence is a schematic representation of the surrounding PR neighboring proteins. PR sequences retrieved via PCR in this study are marked by boldface letters. BPRs are marked in white letterheads on blue background while GPRs are marked with black letterheads on green background.

February 2003 from 5 m surface water), which confirmed these observations. Out of 291 PR containing the ‘105 position’ in the Sargasso Sea shotgun-sequencing data, only six were predicted to absorb in the green range while the remaining 285 sequences contained a glutamine in the ‘105 posi-

tion' (98% blue PRs in the Sargasso shotgun-sequencing project). In the Sargasso Sea (in the selected collection days), the light energy peak in 5 and 40 m is in the blue range with maximal intensity near 440 nm (blue shifted compared to Mediterranean waters; see Figure 2 and Supplementary Figure S4). At 5 m, there is about 82% at 525 nm and it might be that even 20% reduction in the green light range is enough for the BPRs to outcompete the GPRs under these conditions in the Sargasso Sea. It could also be that over time (and mixing), cells will be exposed to deeper waters and one would assume that evolutionary selection would lead to exclusively blue-tuned PRs as being advantageous under the light available in the Sargasso Sea. However, it is somewhat puzzling that GPRs appear underrepresented at this station when green light is still available. This might suggest that BPRs harboring Sargasso Sea bacteria have additional adaptations, which make them abundant in this station in comparison to GPR harboring bacteria found in the Mediterranean Sea or in the Pacific Ocean (Béjà *et al.*, 2001).

Several BPR families were distributed differently in the water column at both locations even when there was little or no GPR–BPR stratification, suggesting depth-specific adaptations in addition to differentiation in absorption spectrum. Some groups occurred primarily in surface samples (e.g. group VI in Figure 1) while others in deep waters (e.g. group I in Figure 1); however, some groups were found throughout the euphotic zone (e.g. group IV in Figure 1). The GPR families were mainly found in the surface with some minor stratification where some were found at both 0 and 20 m (e.g. group VIII in Figure 1), others at 20 m only (e.g. group V in Figure 1), and still others were found along the water column (e.g. group VII in Figure 1). The different distribution among PR families that absorb the same wavelength in the water column may result from competition between the bacteria harboring these PRs. However, the comparison among the different locations suggests that genomic background may be of higher importance in determining the distribution of the different groups.

Overall, the data show that spectral tuning is a common trait shared by different groups of bacteria carrying PR pigments and is widespread in environments where different light quantities and qualities are present. However, our data from the Sargasso Sea, where only BPRs are observed, suggest that either GPR are competitively disfavored under conditions of the Sargasso Sea or PR distribution is complicated by genomic background correlated to other, yet unknown, environmental parameters.

Note added in proof

The distribution of different surface water GPR and BPR variants across the *Sorcerer II* Global Ocean

Sampling expedition transect (Northwest Atlantic through Eastern Tropical Pacific) was recently released. GPR variants were found to be highly abundant in the North Atlantic samples and in non-marine environments, while BPR variants dominated in the remaining mostly open ocean samples (Rusch DB, Halpern AL, Heidelberg KB, Sutton G, Williamson SJ, Yooseph S *et al.* (2007). The *Sorcerer II* Global Ocean Sampling expedition: I, The Northwest Atlantic through the Eastern Tropical Pacific. *PLoS Biol* 5: e77).

Acknowledgements

We thank A Oz and A Wafrik for collecting the Mediterranean Sea samples and the captain and crew of the RV *Shikmona* for expert assistance at sea. Special thanks to S Giovannoni and K Vergin for providing the Sargasso Sea DNA samples and C Carlson and N Nelson for access to the BBOP data. We acknowledge M Suzuki, C Carlson and D Bryant for discussion and ideas. This work was supported by grant 434/02 from the Israel Science Foundation and an EMBO YIP award to OB, and by grants from the NSF-OCE and DOE-GTL programs to MFP.

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. (1990). Basic local alignment search tool. *J Mol Biol* 215: 403–410.
- Béjà O, Aravind L, Koonin EV, Suzuki MT, Hadd A, Nguyen LP *et al.* (2000). Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* 289: 1902–1906.
- Béjà O, Spudich EN, Spudich JL, Leclerc M, DeLong EF. (2001). Proteorhodopsin phototrophy in the ocean. *Nature* 411: 786–789.
- Bielawski JP, Dunn KA, Sabehi G, Béjà O. (2004). Darwinian adaptation of proteorhodopsin to different light intensities in the environment. *Proc Natl Acad Sci USA* 141: 14824–14829.
- Colwell RK. (2005). EstimateS: statistical estimation of species richness and shared species from samples, version 7.5. <http://purl.oclc.org/estimat>.
- de la Torre JR, Christianson L, Béjà O, Suzuki MT, Karl D, Heidelberg JF *et al.* (2003). Proteorhodopsin genes are widely distributed among divergent bacterial taxa. *Proc Natl Acad Sci USA* 100: 12830–12835.
- Felsenstein J. (1989). PHYLIP-phylogeny interference package (version 3.2). *Cladistics* 5: 164–166.
- Frigaard N-U, Martinez A, Mincer TJ, DeLong EF. (2006). Proteorhodopsin lateral gene transfer between marine planktonic Bacteria and Archaea. *Nature* 439: 847–850.
- Giovannoni SJ, Rappé M. (2000). Evolution, diversity and molecular ecology of marine prokaryotes. In: Kirchnerman DL (ed). *Microbial Ecology of the Oceans*. John Wiley & Sons Inc.: New York, NY, USA, pp 47–84.
- Giovannoni SJ, Bibbs L, Cho J-C, Stapels MD, Desiderio R, Vergin KL *et al.* (2005a). Proteorhodopsin in the ubiquitous marine bacterium SAR11. *Nature* 438: 82–85.

- Giovannoni SJ, Tripp HJ, Givan S, Podar M, Vergin KL, Baptista D *et al.* (2005b). Genome streamlining in a cosmopolitan oceanic bacterium. *Science* **309**: 1242–1245.
- Gómez-Consarnau L, González JM, Coll-Lladó M, Gourdon P, Pascher T, Neutze R *et al.* (2007). Light stimulates growth of proteorhodopsin-containing marine Flavobacteria. *Nature* **445**: 210–213.
- Gordon DA, Giovannoni SJ. (1996). Detection of stratified microbial populations related to *Chlorobium* and *Fibrobacter* species in the Atlantic and Pacific Oceans. *Appl Environ Microbiol* **62**: 1171–1177.
- Klepac-Ceraj V, Ceraj I, Polz MF. (2006). CLUSTERER: extendable java application for sequence grouping and cluster analyses. *Online J Bioinf* **7**: 15–21.
- Man D, Wang W, Sabehi G, Aravind L, Post AF, Massana R *et al.* (2003). Diversification and spectral tuning in marine proteorhodopsins. *EMBO J* **22**: 1725–1731.
- Martinez A, Bradley AS, Waldbauer J, Summons RE, DeLong EF. (2007). Proteorhodopsin photosystem gene expression enables photophosphorylation in heterologous host. *Proc Natl Acad Sci USA*, 10.1073/pnas.0611470104.
- Massana R, Murray AE, Preston CM, DeLong ED. (1997). Vertical distribution and phylogenetic characterization of marine planktonic *Archaea* in the Santa Barbara channel. *Appl Environ Microbiol* **63**: 50–56.
- McCarren J, DeLong EF. (2007). Proteorhodopsin photosystem gene clusters exhibit co-evolutionary trends and shared ancestry among diverse marine microbial phyla. *Environ Microbiol* **9**: 846–858.
- Okamoto OK, Hastings aJW. (2003). Novel dinoflagellate clock-related genes identified through microarray analysis. *J Phycol* **39**: 519–526.
- Sabehi G, Béjà O, Suzuki MT, Preston CM, DeLong EF. (2004). Different SAR86 subgroups harbour divergent proteorhodopsins. *Environ Microbiol* **6**: 903–910.
- Sabehi G, Massana R, Bielawski JP, Rosenberg M, DeLong EF, Béjà O. (2003). Novel proteorhodopsin variants from the Mediterranean and Red Seas. *Environ Microbiol* **5**: 842–849.
- Sabehi G, Loy A, Jung KH, Partha R, Spudich JL, Isaacson T *et al.* (2005). New insights into metabolic properties of marine bacteria encoding proteorhodopsins. *PLoS Biol* **3**: e273.
- Schloss PD, Handelsman J. (2005). Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl Environ Microbiol* **71**: 1501–1506.
- Schloss PD, Larget BR, Handelsman J. (2004). Integration of microbial ecology and statistics: a test to compare gene libraries. *Appl Environ Microbiol* **70**: 5485–5492.
- Sharma AK, Spudich JL, Doolittle WF. (2006). Microbial rhodopsins: functional versatility and genetic mobility. *Trends Microbiol* **14**: 463–469.
- Spudich JL, Yang CS, Jung KH, Spudich EN. (2000). Retinylidene proteins: structures and functions from archaea to humans. *Annu Rev Cell Dev Biol* **16**: 365–392.
- Stingl U, Desiderio RA, Cho J-C, Vergin KL, Giovannoni SJ. (2007). The SAR92 clade: an abundant coastal clade of culturable marine bacteria possessing proteorhodopsin. *Appl Environ Microbiol*, doi:10.1128/AEM.02559-06.
- Stomp M, Huisman J, De Jongh F, Veraart AJ, Gerla D, Rijkeboer M *et al.* (2004). Adaptive divergence in pigment composition promotes phytoplankton biodiversity. *Nature* **432**: 104–107.
- Suzuki MT, Giovannoni SJ. (1996). Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. *Appl Environ Microbiol* **62**: 625–630.
- Swofford DL. (2002). PAUP*. Phylogenetic analysis using parsimony (*and other methods). *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Sinauer Associates: Sunderland, MA, USA.
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**: 4876–4882.
- Thurman HV, Trujillo AP. (1999). *Essentials of Oceanography*. Prentice Hall: NJ, USA.
- Venter JC, Remington K, Heidelberg J, Halpern AL, Rusch D, Eisen JA *et al.* (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**: 66–74.
- Walter JM, Greenfield D, Bustamante C, Liphardt J. (2007). Light-powering *Escherichia coli* with proteorhodopsin. *Proc Natl Acad Sci USA* **104**: 2408–2412.

Supplementary Information accompanies the paper on The ISME Journal website (<http://www.nature.com/ismej>)