environmental microbiology

Environmental Microbiology (2017) **19**(3), 1077–1090 doi:10.1111/1462-2920.13624

Metagenomic analysis reveals unusually high incidence of proteorhodopsin genes in the ultraoligotrophic Eastern Mediterranean Sea

Vadim Dubinsky,[1#] Markus Haber,[1#] Ilia Burgsdorf,[1] Kumar Saurav,[1] Yoav Lehahn,[2] Assaf Malik,[3] Daniel Sher,[1] Dikla Aharonovich[1] and Laura Steindler[1*]

[1]*Department of Marine Biology, Leon H. Charney School of Marine Sciences, University of Haifa, Haifa, Israel.*

[2]*Department of Earth and Planetary Sciences, Weizmann Institute of Science, Rehovot, Israel.*

[3]*Bioinformatics Service Unit, University of Haifa, Haifa, Israel.*

Summary

Sunlight can be directly harvested by photoheterotrophic bacteria to create a pH gradient across the membrane, which can then be utilized to produce ATP. Despite the potential importance of this trophic strategy, when and where such organisms are found in the seas and oceans is poorly described. Here, we describe the abundance and taxonomy of bacteria with different trophic strategies (heterotrophs, phototrophs and photoheterotrophs) in contrasting water masses of the ultra-oligotrophic eastern Mediterranean Sea. These water bodies, an anticyclonic eddy and a high-chlorophyll patch resulting from transport of nutrient-rich coastal waters into offshore oligotrophic waters, each supported different microbial populations in surface waters. Based on infrared microscopy and metagenomics, aerobic anoxygenic photoheterotrophic (AAP) bacteria represented up to 10.4% of the microbial community. In contrast, the proteorhodopsin (PR) gene was found in 78.6%–118.8% of the bacterial genome equivalents, the highest abundance reported to date. These results suggest that PR-mediated photoheterotrophy may be especially important in oligotrophic, potentially phosphate-limited conditions.

*For correspondence. E-mail lsteindler@univ.haifa.ac.il; Tel. +972-4-8288987; Fax +972-4-8288267. [#]Vadim Dubinsky and Markus Haber have contributed equally to this work.

Introduction

The South-Eastern Mediterranean Sea is recognized as a region of exceptionally low primary productivity (e.g., Tanaka *et al.*, 2007), which results from the ultraoligotrophic conditions that characterize this area. Inorganic phosphate is mostly below 50 nM (Rusch *et al.*, 2007) and, correspondingly, microbial growth in this area is likely limited by the availability of soluble phosphate (Thingstad *et al.*, 2005; Ghai *et al.*, 2010; Tsiola *et al.*, 2015). Under such conditions, the employment of high-affinity ATP-binding cassette transporters (ABC transporters) may convey a competitive advantage, providing that energy (ATP) is not an additional limiting factor. Thus energy for transport (including transport of phosphates and phosphonates) may be an important factor in areas of very low primary productivity, such as the Eastern Mediterranean Sea. Heterotrophic bacteria that are able to complement their energetic budget with light-mediated ATP production (photoheterotrophs) may have a competitive advantage in such oligotrophic environments. Accordingly, photoheterotrophic bacteria are expected to be highly abundant in the Eastern Mediterranean, one of the most oligotrophic sites worldwide.

The main microbial photoheterotrophic groups include proteorhodopsin (PR) containing bacteria and aerobic anoxygenic phototrophic (AAP) bacteria. A previous study provided evidence of the incredible abundance of PR-containing bacteria, showing that, on a global scale, PR-based bacteria exceed photosynthetic bacteria by threefold (Finkel *et al.*, 2013). PR-containing and AAP bacteria represent assemblies of species from very diverse taxonomical origins. PR-containing bacteria include Actinobacteria, Bacteroidetes, Alpha-, Beta- and Gammaproteobacteria (Beja *et al.*, 2000; Giovannoni *et al.*, 2005; Gomez-Consarnau *et al.*, 2007; Rusch *et al.*, 2007; Yoshizawa *et al.*, 2012; Courties *et al.*, 2013; Mizuno *et al.*, 2015). AAP bacteria include Alpha-, Beta- and Gammaproteobacteria (reviewed in Tsiola *et al.*, 2015) and require organic substrate for growth, as they lack carbon fixation pathways. They are thus heterotrophs that can utilize light as an additional source of energy and save organic substrate that would otherwise need to be respired. Consequently, light exposure can also assist AAP bacteria in survival during

starvation (e.g., (Courties *et al.*, 2013)). Similar to AAP bacteria, PR-containing bacteria also combine the use of organic substrate with light energy (Beja *et al.*, 2000; 2001; DeLong and Beja, 2010). As predicted by Fuhrman *et al.* (2008), several studies have provided evidence for a functional plurality of PRs in different bacterial types. Only in a few copiotrophic bacterial species could PR function be associated with light-enhanced growth (e.g., Gomez-Consarnau *et al.*, 2007; Palovaara *et al.*, 2014), while in other bacteria PR was shown to provide enhanced survival (Gomez-Consarnau *et al.*, 2010), or reduced respiration rate and enhanced substrate uptake levels in carbon starved conditions (Steindler *et al.*, 2011). Yet, no clear effect on growth or survival by light was shown in most studies (e.g., Stingl *et al.*, 2007; Riedel *et al.*, 2010; Efrati *et al.*, 2013). A recent comparison of the genomes and physiology of taxonomically close PR-containing Flavobacteria suggested that PR may be advantageous to strains that are auxotrophic for vitamins and thus require proton motif force dependent transport (Gomez-Consarnau *et al.*, 2015). Based on estimates of energy benefits versus costs of maintenance of PR phototrophy, PR was suggested to provide an energetic advantage only at high light intensities and when large numbers of PR molecules are present per cell (Kirchman and Hanson, 2013). However, the sustained expression of PRs also at the low-light conditions typical of polar regions in winter, suggests additional roles for PR (Nguyen *et al.*, 2015). Taken together, PR-photoheterotrophy appears to provide diverse benefits according to the genomic environments (different bacterial hosts) and the geographic environments (niches characterized by different conditions).

The environmental variables that control microbial abundances, taxonomy and dynamics can be studied by sampling throughout environments that exhibit diverse trophic regimes. In this study, we used multiple methods to estimate the abundance of light-utilizing microorganisms (phytoplankton, PR-containing bacteria and AAP bacteria) at six sites encompassing several different water masses in the Eastern Mediterranean. Our results show that the abundance of photoheterotrophic bacteria, in particular of PR-containing bacteria, in this region are among the highest measured anywhere in the world, raising the question on whether the drivers of this extreme abundance may be related to phosphate and/or carbon limitation.

Results

Characteristics of the study area

During November 2013, we sampled six offshore stations in the southeastern Mediterranean Sea (Fig. 1A). *In situ* CTD vertical profiles of temperature and salinity indicated a mixed layer depth of approximately 50 m (Supporting Information Fig. S1A, B). Temperatures in the mixed layer ranged from 23°C to 23.5°C. Fluorescence vertical profile displayed a deep chlorophyll maximum (DCM) between 90 and 100 m (Supporting Information Fig. S1C). Concentrations of dissolved inorganic nutrients in the surface water (10 m depth) were within previously published ranges for the Eastern Mediterranean Sea (Supporting Information Table S1; Krom *et al.*, 2005; Tanaka *et al.*, 2007). Phosphate concentrations were near or below the detection limit (40 nmol l^{-1}). Silica concentrations were elevated in stations G1 and G2 compared with the other stations (Supporting Information Table S1).

Stations G2 and G6 were located in close proximity to or on transport barriers between distinct water masses, in the form of Lagrangian coherent structures. Lagrangian methods are very effective quantifying the structuring effect of horizontal stirring on mesoscale ($\sim$10–100 km^2) patches with potentially different physical and biological properties (D'Ovidio *et al.*, 2010; Lehahn *et al.*, 2014). Stations G3-G5 were enclosed within an anticyclonic eddy with transport barriers that isolated them from the surrounding waters (Fig. 1C). Stations G1 and G2 were characterized by relatively high chlorophyll concentrations (Fig. 1). Analysis of satellite-derived surface chlorophyll and model-derived surface currents indicated that the relatively high chlorophyll concentrations in these stations were associated with westward transport of a patch of presumably nutrient-rich coastal waters. The important contribution of the coastal (chlorophyll-rich) waters intrusion to structuring the regional chlorophyll field was emphasized by comparing concentrations at stations G1, G2 and G3. These stations were located at the core of the patch (G2), periphery of the patch (G1) and outside of the patch (G3), and were characterized by relatively high, medium and low chlorophyll concentrations, respectively (Fig. 1B). Lower chlorophyll *a* (chl *a*) concentrations estimated from the satellite images at stations G3, G4 and G5 were also confirmed via measurements of chl *a* extracted from samples collected at 10 m depth (Supporting Information Table S1). Although all stations can be considered as offshore stations (water column depth $>$ 1000 m), for the purpose of this study, and based on the differences described above, we defined stations G1 and G2 as 'coastal-impacted' and station G3-G6 as 'offshore' stations, and tested for differences between these two groups in all the following comparisons.

16S rRNA community structure and phytoplankton composition

To study the bacterial community structures, we performed metagenomic sequencing on DNA extracted from the surface waters in stations G1-G6 (see Experimental Procedures). Overall 80,335 partial bacterial 16S rRNA gene sequences were extracted from the unassembled

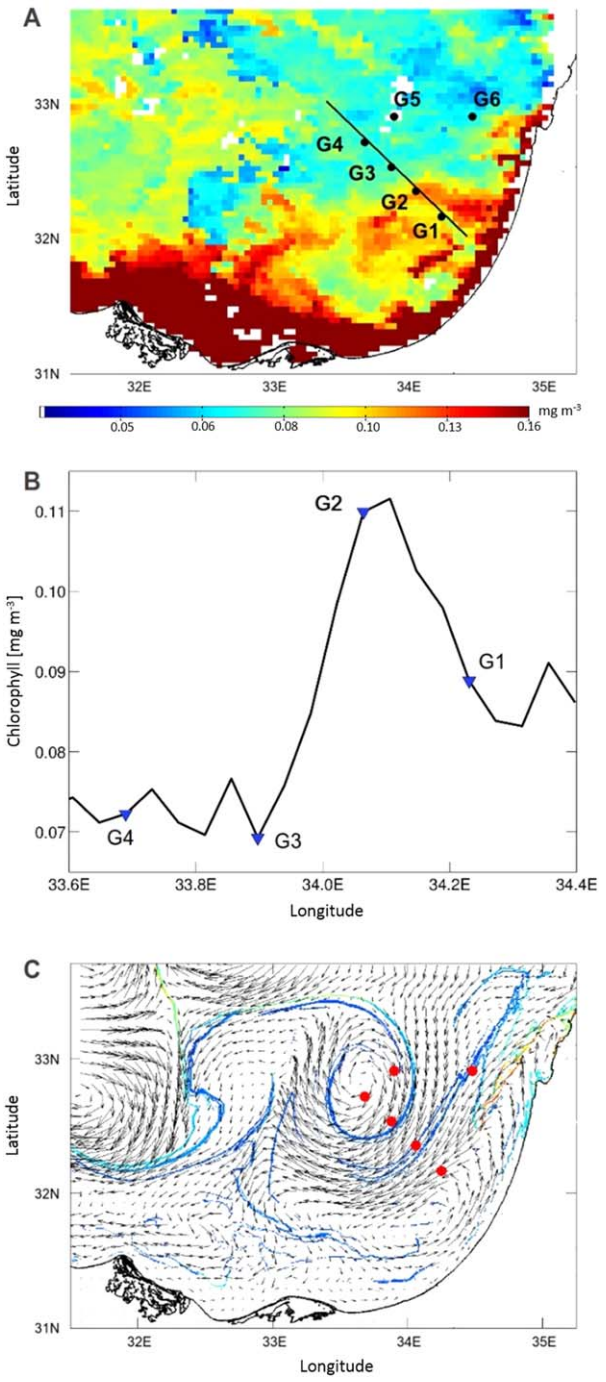

Fig. 1. Satellite-derived surface chlorophyll *a* data and model-derived surface currents of the study-area from November 14th, 2013. (A) Map of surface chlorophyll concentrations. Black dots (G1 to G6) mark the sampling locations. Straight black line marks the section presented in panel B; (B) Surface chlorophyll (satellite) along the transect marked in panel A, showing the locations of two chlorophyll fronts. Blue triangles mark the locations of sampling stations along the transect (G1 to G4); (C) SELIPS surface velocity field of the study area. Arrows represent ocean surface currents, blue lines are attracting Lagrangian coherent structures (LCS), and red dots mark the sampling locations. [Colour figure can be viewed at wileyonlinelibrary.com]

metagenomes using Metaxa2, of which 34,255 sequences could be classified after quality control and re-annotation by the SILVAngs analysis pipeline (Quast *et al.*, 2013). Alphaproteobacteria was the dominant class (average of 50% of the sequences in all the stations), followed by Gammaproteobacteria (23%), Flavobacteriia (9%) and Cyanobacteria (6%). At higher taxonomy resolution, the SAR11 clade was the most abundant group (30% of the 16S rRNA sequences on average). The SAR11 clade appeared to be more abundant at the 'offshore' stations (average 36%) than at the 'coastal-impacted' ones (average 18%, Fig. 2), however, this was mostly true for station G1. The SAR86 clade was the second most abundant group (13%–18% of the community), but did not show any considerable differences between the 'coastal-impacted' and 'offshore' stations. Subsection I cyanobacteria (*Prochlorococcus* and *Synechococcus*) were more abundant at 'coastal-impacted' (12%) than at 'offshore' stations (4%, Fig. 2). A phylogenetic tree based on near full-length 16S rRNA gene sequences reconstructed by EMIRGE (assembled 16S rRNA of high-coverage taxa from the metagenome) is shown in Supporting Information Fig. S2 and includes the calculated abundance data for these taxa. Hierarchical clustering based on Bray–Curtis dissimilarity of the 16S rRNA sequence data confirmed the assumed difference in microbial community structure between 'coastal-impacted' and other stations ($p < 0.05$) (Fig. 2). SIMPER analysis of the 16S rRNA data showed SAR11 and Cyanobacteria to be the top two contributors to the observed dissimilarity (together 53%), with SAR11 being twice more abundant in the 'offshore' stations, while Cyanobacteria were three times more abundant in the 'coastal-impacted' stations (Table 1). Flavobacteriales and Rhodobacterales, both generally regarded as coastal associated orders, were more abundant in the 'coastal-impacted' stations, and Oceanospirillales of the SAR86 clade were slightly more abundant in 'offshore' stations (Table 1). It is noteworthy that the same clustering of stations was found when analysing another housekeeping gene, the *recA* gene, in the SIMPROF analysis (Supporting Information Fig. S3B).

The results of the 16S rRNA gene analysis were supported by microscopic analyses. Heterotrophs were an order of magnitude more abundant than photoautotrophs and contributed between 84.6% and 94.1% to the total cells counts (Supporting Information Table S2). Their total numbers were higher at 'coastal-impacted' than at 'offshore' stations (Fig. 3). Picophytoplankton abundance (including *Prochlorococcus* and picoeukaryotic phytoplankton, as they were recognized based on the same criteria, i.e., chlorophyll fluorescence signal, lack of phycoerythrin signal, and similar sizes) varied across the stations. It was highest at G1 and G2 (2.44×10^4 and 5.9×10^4

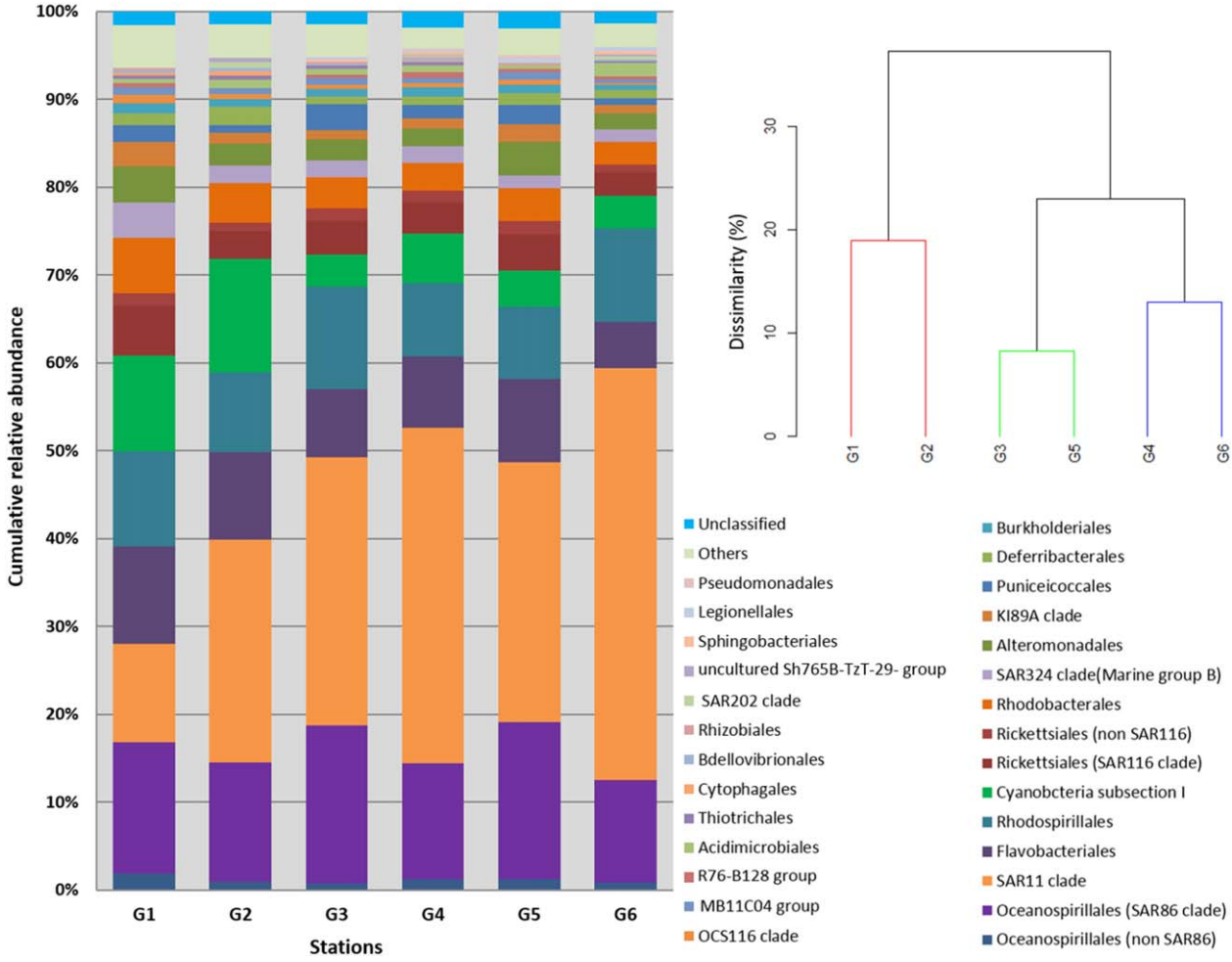

Fig. 2. Relative abundance of bacteria (16S rRNA extracted using Metaxa2) at the order level in the sampled stations. Taxonomic annotations were assigned based on the SILVA SSU database v115 using 97% cutoff for OTUs. Inset shows a dendrogram of the community structure from the six metagenomes using hierarchical clustering dendrogram and Bray–Curtis dissimilarity. Differently coloured branches indicate significantly different ($p < 0.05$) clusters as tested by SIMPROF analysis. [Colour figure can be viewed at wileyonlinelibrary.com]

cells ml^{-1} respectively), and ranged between 9.52 $\times$ 10^3 and 1.82 $\times$ 10^4 cells ml^{-1} at 'offshore' stations (Fig. 3). *Synechococcus* numbers followed the same trend. Picophytoplankton together with *Synechococcus* (the major members of the photoautotrophic community) comprised 10.4%–15.4% fraction of total cells at the 'coastal-impacted' stations, whereas at the 'offshore' stations their relative abundance decreased to 5.9%–7.9% (Supporting Information Table S2). Similar overall trends were observed based on flow-cytometry (Supporting Information Table S3).

HPLC pigment analysis was used to characterize the phytoplankton community, providing information on presence of picophytoplanktonic cyanobacteria and prymnesiophytes (Supporting Information Fig. S4 and Supporting Information Table S4).

Abundances of microbial groups with different trophic strategies

To compare the abundance of photoautotrophic, photoheterotrophic and heterotrophic cells in the samples, we extracted from the metagenomes the sequences of three indicator genes: *psaA* (encoding a core photosystem I protein and thus representing phytoplankton), proteorhodopsin (*prd*) and *pufM* (encoding the M subunit of the photosynthetic reaction centre in anoxygenic phototrophs) (Table 2; Supporting Information Table S6). In addition, we also extracted separately two gene variances of *prd*, xanthorhodopsin and marine actinobacterial clade rhodopsins. Four single-copy, core genes (*recA*, *gyrB*, *rpoB*, *tuf*) found in all bacteria were extracted the same way and used for normalization (see Experimental Procedures).

Table 1. Similarity percentages (SIMPER) analysis of microbial community genes showing which of the *prd* and 16S rRNA genes contributed most to the dissimilarity between the groups ('coastal-affected' and 'offshore').

Gene compared	Average relative abundance (%)		Contribution to dissimilarity (%)		Average dissimilarity between groups (%)
	Coastal group	Offshore group	Single	Cumulative	
16S rRNA					
SAR11 clade	0.18	0.36	37.19	37.19	24.2
Cyanobacteria SubsectionI	0.12	0.04	15.85	53.04	
Flavobacteriales	0.11	0.08	5.96	59	
Oceanospirillales (SAR86 clade)	0.14	0.15	5.65	64.65	
Rhodobacterales	0.05	0.03	4.42	69.07	
prd					
AAY68047.1 uncultured bacterium[a]	0.05	0.08	5.35	5.35	33.7
ABL97421.1 uncultured marine bacterium EB80_69G07[a]	0.02	0.03	2.28	7.63	
ABL97827.1 uncultured marine bacterium HF10_49E08[a]	0.02	0.02	1.55	9.18	
ALS56212.1 uncultured bacterium EIL27G07[b]	0.02	0.02	1.16	10.34	
KZX58463.1 *Halioglobus* sp. HI00S01	0.01	0	1.06	11.4	

Numbers are the relative abundances that each category contributes to the dissimilarity between the groups. Only the 5 highest contributors for each gene are shown.
a. Best cultured blastp hits: *Candidatus* Pelagibacter strains with > 90% similarity.
b. Best cultured blastp hit: *Candidatus* Pelagibacter strain IMCC9063 with 59% similarity.

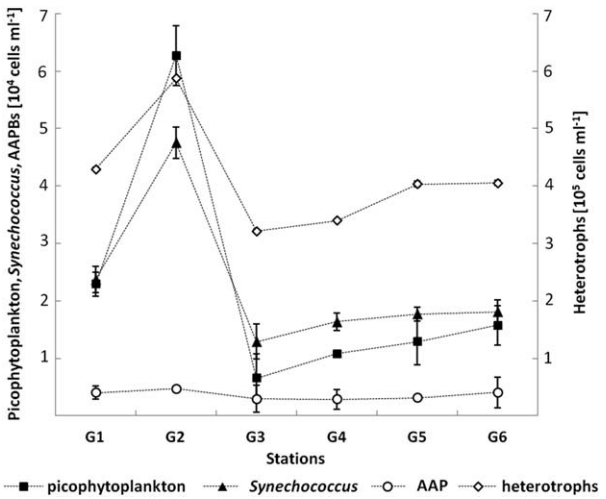

Fig. 3. Abundance of picophytoplankton, *Synechococcus*, AAP bacteria and heterotrophic bacteria as determined by epifluorescence microscopy. Error bars denote standard deviation of duplicate samples. The data for each replicate consisted of an average of 15 frames.

Based on normalized abundance of the *psaA* gene, photoautotrophs were exceedingly more abundant in the 'coastal-impacted' stations (9.9% and 14.2% in G1 and G2 respectively) than in the 'offshore' stations (2.3%–4.6%, Table 2), which was in agreement with the 16S rRNA gene data (Fig. 2), microscopy results (Fig. 3, Supporting Information Table S2) and flow cytometry data (Supporting Information Table S3). SIMPROF analysis (using *psaA*) clustered stations G2 and G6 together and significantly apart from the other stations (Supporting Information Fig. S3D). Given that the hierarchical structuring of this gene did not support a grouping of 'coastal-impacted' vs. 'offshore' stations, no SIMPER analysis was performed.

Table 2. Averaged normalized relative abundance (in percent) ± standard error of the mean of proteorhodopsin gene *prd*, xanthorhodopsin (XR) gene, marine actinobacterial clade (MAC) rhodopsin gene, photosystem I gene *psaA* and anoxygenic photosynthetic reaction centre gene *pufM* in the bacterial community based on reciprocal BLAST analysis.

Station	*prd*	XR	MAC	*pufM*	*psaA*
G1	78.6 ± 9.3	1.8 ± 0.2	1.6 ± 0.2	10.4 ± 1.2	9.9 ± 1.2
G2	80.7 ± 10.2	0.8 ± 0.1	4.8 ± 0.6	5.6 ± 0.7	14.2 ± 1.8
G3	106.2 ± 12.6	1.6 ± 0.2	6.3 ± 0.8	7.7 ± 0.9	2.4 ± 0.3
G4	96.1 ± 11.5	0.8 ± 0.1	5.2 ± 0.6	6.1 ± 0.7	4.6 ± 0.6
G5	105.7 ± 12.5	1.3 ± 0.2	3.9 ± 0.5	8.3 ± 1.0	2.3 ± 0.3
G6	118.8 ± 16.5	0.5 ± 0.1	5.1 ± 0.7	3.5 ± 0.5	2.5 ± 0.4

Abundances are normalized using four single copy house-keeping genes (*recA*, *gyrB*, *rpoB*, *tuf*).

The two kinds of photoheterotrophic bacteria differed greatly in their abundances. Bacteria possessing a *prd* gene were remarkably abundant at all the stations, with a minimum of 78.6% (G1) to a maximum of 118.8% (G6) out of the total bacterial community. Higher than 100% abundances indicate that some bacteria have more than one *prd* gene copy per genome. In addition, two other bacterial proton-pumping rhodopsins were present in low abundance: xanthorhodopsin was present in 0.5%–1.8% of all bacterial cells and marine actinobacterial clade rhodopsin in 1.6%–6.3%, neither showed a clear difference in abundance between coastal-impacted and offshore stations. Other detected rhodopsin types included archaeal and viral

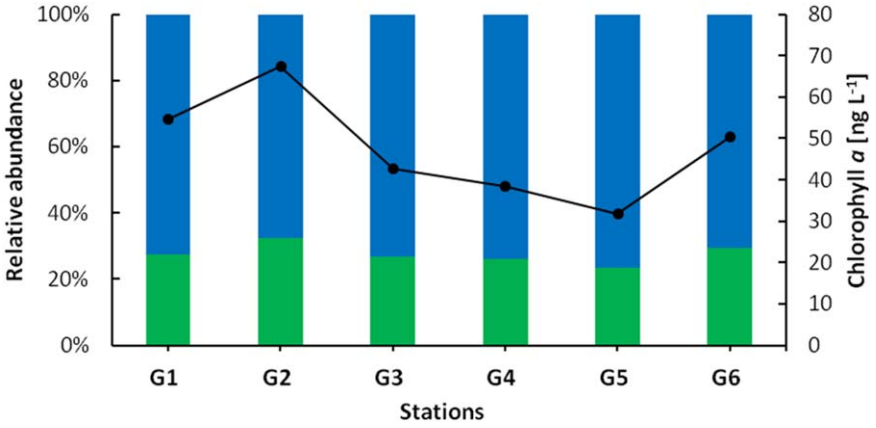

Fig. 4. Relative abundance of blue and green spectrally tuned PR proteins assembled from the metagenome compared with extracted chlorophyll *a* (black line) in the six sampled stations. [Colour figure can be viewed at wileyonlinelibrary.com]

proteorhodopsins (see Supporting Information Fig. S7). The abundance of bacteria bearing a *prd* gene was between 15.4% and 40.2% higher in the 'offshore' stations compared with 'coastal-impacted' ones (Table 2). Clustering analysis with SIMPROF based on *prd* gene showed that G1 and G2 grouped together and apart from other stations (Supporting Information Fig. S3A). SIMPER analysis indicated that the average dissimilarity between these two groups was 33.7%. The top two contributors to the dissimilarity between groups (accounting for 7.63% of the dissimilarity) were more abundant in the 'offshore' group (Table 1). Their best-identified Blast hit with an identity of more than 90% belonged to *prd* genes of the SAR11 clade, which was more abundant in the 'offshore' stations. On the phylum level, the best taxonomic Blast hit for most reads belonged to Proteobacteria (77.1%–89.7% of the *prd* reads with taxonomic annotation), followed by Bacteroidetes (7.8%–17.1%) (Supporting Information Fig. S5). Together these two phyla represented 94.0%–97.5% of taxonomically assigned proteorhodopsin reads found. Detailed results at the class and order level are given in the supplemental material.

The absorption maxima of PRs can be tuned to 525 nm (green) and 490 nm (blue), based on differences in a single amino acid at position 105 (Beja *et al*., 2001; Man *et al*., 2003). The distribution of differently tuned PRs was shown to be related to the type of available light (Man *et al*., 2003). We analysed the abundance of the differently tuned *prd* sequences using assembled *prd* gene sequences translated to amino acid that covered the amino acid position 105. Raw metagenomic reads of the different stations were then mapped against these assembled sequences to calculate relative abundances of *prd* genes with different spectral tunings. Overall, blue light adapted *prd* sequences were two to three times more abundant than green light adapted ones (blue:green-adapted *prd* sequence ratio ranged from 2.1 to 3.2). The highest abundance of green-tuned *prd* sequences was found at station G2, which is situated at a high surface chlorophyll patch (Fig. 4).

Compared with PR-containing bacteria, AAP (*pufM*-bearing) bacteria were present in lower abundances at all stations, representing 3.5%–10.4% of all bacteria. AAP bacteria abundances based on *pufM* sequences were higher than those based on epifluorescence microscopy data (3.2×10^3 to 6×10^3 AAP bacteria cells ml^{-1} (Fig. 3), representing 0.7%–1.3% of the total microbial cells (Supporting Information Table S2). No apparent differences between 'coastal-impacted' and 'offshore' stations could be detected. G1 clustered together with G4 and next to a not significantly different cluster made up by G3 and G5 in the hierarchical clustering analysis of the *pufM* gene with SIMPROF, while station G2 and G6 clustered significantly apart from each other and the cluster containing all other stations (Supporting Information Fig. S3C). The best taxonomic hits for *pufM* reads were almost exclusively Proteobacteria (97.1%–98% of all taxonomical assigned *pufM* reads). Alphaproteobacteria were most common (69.3%–81.1%) followed by Gammaproteobacteria (9.0%–14%) and Betaproteobacteria (7.6%–15.1%). At the order level, the Rhodobacterales dominated (45.4%–64.4%), followed by Rhizobiales (8.0%–13.5%), Burkholderiales (5.7%–10.0%), Cellvibrionales (5.0%–8.5%), Chromatiales (3.6%–6.1%), Sphingomonadales (3.5%–5.4%) and Rhodospirillales (2.4%–4.4%). No obvious pattern regarding 'coastal-impacted' versus 'offshore' stations was evident at either class or order level (Supporting Information Fig. S6).

Discussion

Our results provide an overview of the microbial groups using light energy via different trophic strategies (PR-containing bacteria, AAP bacteria and photosynthetic microorganisms) at six offshore stations located in the Eastern Mediterranean Sea. Among these stations, two were affected by intrusions of coastal waters at the time of sampling, and thus were characterized by different biotic and likely abiotic conditions. Intrusions of coastal waters into offshore waters have been reported before and appear

to be a recurrent phenomenon at the Eastern Mediterranean coast of Israel (Barale *et al*., 2008; Efrati *et al*., 2013).

The most unexpected result was the unusually high incidence of *prd* genes in our datasets. When normalized by four single copy core genes, we estimated that 78.6%–80.7% of the bacterial community harboured the *prd* gene at the sampling locations that were affected by coastal water intrusions, and 96.1%–118.8% at the sampling sites characterized by low-chlorophyll/low-nutrient conditions. Even if these values were an over-estimation related to the potential presence of more than one *prd* gene in certain bacteria, similar biases affected also previous studies, and these are the highest estimates for PR-containing bacteria reported thus far for microbial communities. Surprisingly, much lower estimates of PR-containing bacteria (13%) were previously reported in the Eastern Mediterranean Sea (Sabehi *et al*., 2005). However, this previous estimate was based on the sequencing of BAC libraries (which involves cloning the large DNA fragments and maintaining the BACs in *E. coli*), rather than direct metagenomic sequencing. A comparison of direct sequencing versus sequencing after cloning in fosmids revealed a strong bias against dominant members of marine microbial communities, including SAR11 (Ghai *et al*., 2010), suggesting that PR-containing bacteria abundance in the Eastern Mediterranean Sea was previously underestimated due to cloning bias. Diversity and geographic distribution of *prd* genes have been estimated also in other marine environments (e.g., De la Torre *et al*., 2003; Sabehi *et al*., 2003; 2007; Philosof and Beja, 2013; Nguyen *et al*., 2015), but only a few studies specifically calculated PR-bacteria abundances in environmental samples. According to qPCR estimates, the Northern Atlantic Ocean is characterized by PR abundances up to 50% of the microbial community (in the Western Sargasso Sea). Interestingly, PR abundances in this study are negatively correlated with chlorophyll, and 3 out of 4 PR-types (the exception being a Flavobacteria-like PR) have a negative correlation with nutrients including phosphate (Campbell *et al*., 2008). These estimates correspond well with shotgun sequence datasets for the Sargasso Sea, where at least 65% of the bacterial community are calculated to contain a *prd* gene (Venter *et al*., 2004; Rusch *et al*., 2007). The observed high abundances of *prd* genes and the inverse relationship of these abundances with chlorophyll and nutrients suggest that conditions of oligotrophy could select for PR-containing bacteria, which is in line with our results.

Alphaproteobacteria, and in particular the SAR11 clade, were the most common PR-containing bacteria in our dataset. In terms of spectral tuning, based on sequence analysis, most PRs in our study were blue-tuned. A mix of green and blue-tuned PRs was reported before for the Eastern Mediterranean Sea, with higher incidence of blue-tuning for surface samples in winter versus summer (Sabehi *et al*., 2007). The highest percentage of green PRs coincided with the highest chlorophyll concentration (based on both satellite data and extracted chlorophyll from 10 m depth sample) (Fig. 4), suggesting that the higher abundance of photosynthetic microbes at G2 utilizing blue light for photosynthesis may have provided an advantage to PR-bearing bacteria that can use green wavelengths, compared with the other sites. An analysis of correlation between PR-spectral tuning and chlorophyll concentration was not conclusive (data not shown), because the extracted chlorophyll concentration at station G6 showed variable values in the replicates (Supporting Information Table S1).

In comparison to the high abundances of PR-containing bacteria, picophytoplankton was less common (2.3%–14.2% of the bacterial community based on normalized *psaA* gene abundance) at all sampled stations. They were enriched in the stations affected by coastal waters intrusion (9.9%–14.2%), where they were two to six times more abundant than in the 'offshore' stations (2.3%–4.6%). Similarly, microscopy, flow cytometry, satellite chlorophyll-fluorescence imaging and extracted chlorophyll data confirmed higher picophytoplankton presence at 'coastal-impacted' stations, in particular at station G2. It was perhaps surprising that *Prochlorococcus* was the dominant phytoplankton at station G2 (see Supporting Information Tables S2–S4), the 'coastal-impacted' station, since these picocyanobacteria are usually considered to be associated with open ocean rather than coastal sites (although exceptions have been described (Shimada *et al*., 1995). Regardless of the identity of the primary producers, higher primary productivity related to the intrusion of nutrient-rich coastal waters may be one of the factors affecting the different microbial community structures at the 'coastal-impacted' stations, represented by higher proportions of Flavobacterales and Rhodobacterales (copiotrophs) compared with the 'offshore' stations that showed higher proportions of SAR11 (oligotrophs). It was unexpected that SAR11, given its typical oligotrophic lifestyle (Lauro *et al*., 2009), was also relatively abundant at the high chlorophyll station G2 (Fig. 2). A previous study indicated that SAR11 could surprisingly be equally sensitive to changes in chlorophyll *a* as other groups of heterotrophic bacterioplankton (Sjostedt *et al*., 2014). Evidence for correlation between abundance of SAR11 and phytoplankton is also reported at the HOT site (Eiler *et al*., 2009) and in the Southern Atlantic gyre (Schattenhofer *et al*., 2009). Future studies on co-cultures between diverse SAR11 ecotype-isolates and phytoplankton may help resolve the inconsistencies between the oligotrophic nature of SAR11 and its sometimes observed correlation with phytoplankton abundance.

The lower productivity of the 'offshore' stations may relate to the highest abundances of PR-containing bacteria

detected at the same sites, as the energy obtained by PR function could favour nutrient transport at sites where nutrients limit bacterial growth. In contrast to the PR-photoheterotrophs, AAP-photoheterotrophs represented a much lower percent of the microbial communities here investigated. The abundances of AAP bacteria at the six sampled sites, based on infrared epifluorescence microscopy, were in the same range as previously reported for photic zones in the Mediterranean Sea, representing 0.7%–1.3% of the microbial community (Lamy *et al*., 2011). AAP bacteria abundances as calculated based on of *pufM* sequences quantification (and normalization by four single copy core genes) were higher (3.5% up to 10.4%). This may be in part due to the difference in sample treatment. In contrast to the DNA samples, microscopy samples were not pre-filtered. Thus the relative abundance of AAP bacteria is reduced as more organisms are counted. This discrepancy may also be related to the difference between gene presence and its expression. Whereas epifluoresence microscopy relies on the expression of BChl *a*, the metagenomic data are independent of BChl *a* production and use the normalized counts of *pufM* genes as proxy for AAP bacteria abundance. The sampling was performed in November and the expression of BChl *a* in part of the microbial taxa may have been lower than detection level. In the Western Mediterranean Sea, lowest BChl *a* concentrations are reported for winter months and it is suggested that this trophic strategy may be favoured in summer months (Ferrera *et al*., 2014). Likewise, higher AAP bacteria numbers are observed in late spring and summer compared with winter also in studies performed at different geographic locations (Masin *et al*., 2006; Zhang and Jiao, 2007). In the Mediterranean, AAP bacteria abundances reportedly decrease from the oligotrophic Western sites to the ultraoligotrophic Eastern basin, suggesting no advantage of AAP bacteria in nutrient limiting conditions (Lamy *et al*., 2011) and contrasting a previous hypothesis that light-generated energy could enhance phosphate acquisition by AAP bacteria in phosphate-limited conditions (in lakes; Masin *et al*., 2008). Similarly, Hojerova and colleagues report higher AAP bacteria abundances in a post-diatom bloom situation and suggest that AAP bacteria may thrive better in higher trophic environments providing sufficient substrates to maintain high growth rates (Hojerova *et al*., 2011). Based on the latter hypothesis we would have expected higher abundances of AAP bacteria in the 'coastal-impacted' stations, characterized by higher abundance of photosynthetic bacteria and thus productivity, than in the 'offshore' stations. However, neither influence on AAP bacteria abundance (based on microscopy and metagenomic data) nor on their diversity (based on clustering of *pufM* genes) was observed as an effect of the intrusion of coastal waters in offshore waters. AAP bacteria abundances and diversity at over forty sampling stations across the Global Ocean Sampling (GOS) transect vary significantly between different oceanic regions, yet similar proportions of AAP bacteria are reported in both oligotrophic and eutrophic environments (Yutin *et al*. 2007). In our study, the most common AAP bacteria in the Eastern Mediterranean belonged to the Alphaproteobacteria (69.3%–81.1% of the taxonomical assigned *pufM* hits). This result differed from previous results for the Western Mediterranean, where AAP bacteria are dominated by Gammaproteobacteria (especially in mesotrophic and eutrophic environments (Lehours *et al*., 2010), and in the summer (Ferrera *et al*., 2014)). Different AAP bacteria are likely not functionally and ecologically equal, thus more investigation is needed to determine the abiotic factors that influence Alphaproteobacteria AAP bacteria dominance in the Eastern Mediterranean. An analysis of seasonal dynamics of AAP bacteria in this ultraoligotrophic region may provide further insight on the conditions enhancing different types of AAP bacteria.

In summary, abundances of AAP bacteria in the Eastern Mediterranean were in agreement with previous studies, whilst the here discovered unusually high incidence of proteorhodopsin genes emphasizes the need to revisit their biological importance in oligotrophic systems. Phosphate is considered a limiting nutrient in the Eastern Mediterranean Sea, and it will be of interest to determine whether and how PR-derived energy is partitioned between phosphates/phosphonates ATP-dependent transport and other metabolic requirement in PR-bacteria from this region. Undoubtedly, an understanding of how PR affects metabolism in abundant members of the marine microbial community will be pivotal to define their biogeochemical function. *In-situ* functional studies and the isolation of oligotrophic photoheterotrophs from the Eastern Mediterranean will open new avenues to follow up on these questions.

Experimental procedures

Sampling and measurements of environmental parameters

Samples were collected from six stations (G1 to G6) during a cruise onboard the R/V Mediterranean Explorer on November 12–13th, 2013. Stations G1 to G4 form a line transect with distance between stations ranging from 26 to 29 km, G5 was separated by 29 km from G4, and G6 was 54 km from G5 (Fig. 1; Supporting Information Table S1). Vertical profiles of temperature, salinity and dissolved oxygen were obtained using a SeaBird conductivity, temperature and depth profiler (CTD, SBE19 plus) and a fluorescence profile of chlorophyll *a* using a Seapoint fluorometer mounted onto the CTD, calibrated with bottle chlorophyll measurements. Seawater samples for subsequent DNA, pigment, nutrient, flow cytometry and microscopy analyses were collected from 10 m depth using Niskin bottles mounted on the CTD.

Surface chlorophyll *a* concentrations were derived from the ocean colour Moderate Resolution Imaging Spectroradiometer

(MODIS) aboard the Aqua satellite. We used the standard Level 3 Chlorophyll products with a spatial resolution of 4 km, downloaded from the ocean colour data distribution site (http://oceandata.sci.gsfc.nasa.gov/). Since during the period of the cruise clouds masked much of the study area, we used a relatively cloud-free image from November 14th, 2013 for the analysis.

The surface velocity field (sea currents) was calculated from the SELIPS (South Eastern Levantine Israeli Prediction System) 3D numerical model for ocean dynamics (Goldman *et al.*, 2015). Transport barriers between water masses were determined by identifying attracting Lagrangian coherent structures (LCS, Haller and Yuan, 2000) from calculation of finite size Lyapunov exponents (Boffetta *et al.*, 2001).

Flow cytometry analysis

Seawater samples (1.5 ml) were fixed (0.25% glutaraldehyde) for 10 min in the dark at room temperature, stored in liquid nitrogen onboard and transferred to −80°C upon arrival at the laboratory until analysis on a flow cytometer (BD FACSCanto II, details in Supporting Information).

Epifluorescence microscopy analysis

Seawater samples (40 ml) were fixed (1.85% formaldehyde) and kept at 4°C until filtration on 0.2 μm black polycarbonate filters (Millipore). Cells were stained for 10 min in the dark with a 1 μg ml^{-1} 4′,6-diamidino-2-phenylindole (DAPI) solution (Sigma) in the filtration manifold (Pall), and rinsed by filtering 50 ml Milli-Q water. The rinsed filters were placed on a glass slide, a drop of 3:1 mixture oil solution (Citifluor AF1 and Vectashield) was applied onto the filter and a cover slip was put and gently flattened. Cells were counted under a Carl Zeiss epifluorescence microscope equipped with AxioCam MRM camera. For each sample, 15 frames (each consisting of 4 images) were captured using the following optical filters and settings: (i) phycoerythrin (PE, excitation: 525–555 nm; emission: 560–650 nm; 100 ms exposure) to detect *Synechococcus* cells; (ii) Chl *a* (excitation: 446–486 nm; emission: 668–726 nm; 500 ms exposure) to detect chlorophyll-containing organisms; (iii) DAPI (excitation: 320–400 nm; emission: 410–510 nm; 200 ms exposure) for total bacteria count; (iv) IR (excitation: 290–490 nm; emission: 750–850 nm; 5 s exposure) showing both AAP bacteria and phytoplankton. The acquired images were analysed semi-automatically with Fiji-ImageJ software (http://fiji.sc/Fiji). The four images of each field were overlaid, and cells with DAPI and IR signals, but no Chl *a* nor PE signal were counted as AAP bacteria. Cells with PE, Chl *a* and DAPI signals were considered as *Synechococcus* and cells with only Chl *a* and DAPI as picophytoplankton. Finally, heterotrophic bacteria were identified as cells having only DAPI fluorescence. These counts were used to calculate the relative abundances of the four groups (AAP bacteria, *Synechococcus*, picophytoplankton and heterotrophic bacteria) and absolute numbers were obtained by multiplying by the total number of cells for each sample, determined according to flow cytometry.

DNA extraction, library preparation and illumina metagenomic sequencing

Seawater samples (10–15 L) for DNA analysis were prefiltered through GF/D filters (Whatman) (nominal pore size 2.7 μm) to remove eukaryotic cells and cells were collected on 0.22 μm Sterivex filters (Millipore). Excess water was removed and 1 ml lysis buffer (0.75 M sucrose, 50 mM Tris-HCl, 40 mM EDTA, pH = 8.3) was added. Filters were stored at −80°C until extraction. Nucleic acids were extracted from the samples as previously described (Massana *et al.*, 1997). The metagenomic libraries were prepared and sequenced at Mr. DNA Molecular Research LP, USA. For each of the six sampled stations (G1–G6) 1 μg of purified genomic DNA was used for Illumina Nextera library preparation following the manufacturer's instructions and libraries were sequenced on one lane of a flow cell using the HiSeq2500 high-throughput system with 2 × 150 bp paired-end technology. The base-calling pipeline (version Illumina Pipeline-0.3) was used to process the raw fluorescence images and call sequences. The metagenomic sequencing generated on average 50 million paired-end reads per sample, 308 million reads and 46 Gbp of raw sequence data in total for the six samples.

Analysis of 16S rRNA community structure

16S rRNA reads were extracted and classified with Metaxa2 (Bengtsson *et al.*, 2011), run using default parameters with enabled paired end reads option. All extracted sequences were classified using Blast matches against SILVA SSU database release 111. The extracted reads were further analysed and re-annotated in the SILVAngs analysis pipeline v1.1 (Quast *et al.*, 2013). Reads shorter than 50 aligned nucleotides, with more than 2% of ambiguities, or 2% of homopolymers, and reads with a low alignment quality were excluded from further processing. Identical reads were identified, unique reads were clustered (OTUs by 97% identity), and the reference read of each OTU was classified. The classification was performed by BLASTn against the SILVA SSU database release 115. Further, full length 16S rRNA genes were reconstructed from the short unassembled metagenomic pair end reads using EMIRGE version 3 (Miller *et al.*, 2011). Details on the phylogenetic analysis methods used for Supporting Information Fig. S2 are given in Supporting Information material.

Functional analysis

Raw unassembled metagenomic sequences were annotated using the Metagenomics Rapid Annotation (MG-RAST) pipeline version 3.5 (Meyer *et al.*, 2008) to determine the abundance of archaeal, eukaryotic, viral and bacterial reads (see Supporting Information Table S5).

To determine relative abundances of bacteria with different trophic strategy (phototrophs, aerobic anoxygenic photoheterotrophs and photoheterotrophs with proteorhodopsin, xanthorhodopsin and marine actinobacterial clade rhodopsin) the read counts of five functional genes (*psaA, pufM, prd*, xanthorhodopsin, marine actinobacterial clade rhodopsin) and four single copy housekeeping genes (*recA*, *gyrB*, *rpoB*, *tuf*) were extracted using a reciprocal blast approach.

Query databases for the nine genes were created by downloading their sequences from the UniProt Reference database clustered at 90% similarity (UniRef90) using gene name (except for *prd*, xanthorhodopsin and marine actinobacterial clade rhodopsins for which protein name was used), taxonomy: bacteria, and length $\geq$ 200 amino acid as parameters. The databases were quality checked using EggNOG 4.5, hmmer searches against the pfam database and phylogenetic analysis (data not shown). Given the much larger size of the databases for the housekeeping genes, these were further reduced by clustering them at 70% similarity using Blastclust. In a first step, the resulting databases were queried with tBLASTn against the metagenomes of each station, separately for each read direction. All obtained unique metagenomic reads were then queried in a second step with BLASTx against the NCBI NR protein database and only the top hit of each read was retained. In both steps of the reciprocal blast, a default e-value cutoff of 10 was used. Non-bacterial reads were filtered out based on the NCBI taxonomy of their NCBI NR protein database top hit. To ensure that only reads to NR sequences of the target gene were counted, we examined the annotated regions, conserved domains and best 100 BLASTx hits of all unique NR hits with 10 or more reads for any read direction in any station. In addition, all hits annotated as hypothetical, unknown or uncharacterized protein were examined this way and at least one sequence of each unique annotation. For *prd*, xanthorhodopsin and marine actinobacterial clade rhodopsin the decision of the type of rhodopsin was also based on the closest known sequence from cultured isolates.

The abundance of bacteria with the functional genes was calculated using the following equation adopted from (Kwon *et al.*, 2013): $P = (N_i/L_i)/(N_m/L_m)$, in which P is the normalized proportion of cells containing the gene of interest, N_i is the hit count assigned to the gene of interest, N_m is the hit count assigned to housekeeping gene used for normalization, and L_i and L_m are the lengths of the gene of interest and the housekeeping gene respectively. The selected genes of interest were assumed to occur as single copy genes. These calculations were first performed for each read direction and housekeeping gene separately. The results for the four housekeeping genes were first average for each read direction separately in every station and then the two read direction per station were averaged to give the normalized relative abundance for each station.

For taxonomic identification of *psaA*, *pufM*, *prd* and *recA*, all reads regarded as positive hits from the second step of the reciprocal blast were used as a query for another BLASTx against NCBI NR protein database. This time uncultured bacteria, environmental samples, artificial sequences and synthetic constructs were excluded and the taxonomy of the best hit was assigned to the positive read.

Multivariate statistical analysis

Multivariate statistical analysis was used to determine the degree of similarity between the metagenomic samples. Similarity profiles of *prd*, *pufM*, *psaA*, *recA* and 16S rRNA genes (derived from Metaxa2) were constructed using relative abundance matrices. For all the analysed genes relative abundance of best hits was used, except 16S rRNA for which taxonomic level of order was used. Bray–Curtis dissimilarity pairwise distance matrix was calculated for each profile and then clustered by applying hierarchical clustering based on the Ward agglomeration method. Significantly different clusters were defined by $p < 0.05$ after 2000 permutations using similarity profile analysis, SIMPROF (Clarke *et al.*, 2008), which was run with the clustsig package for R (http://cran.r-project.org/web/packages/clustsig/index.html). In case of significant clustering of G1 and G2 apart from G3 to G6 ('coastal-impacted' vs. 'offshore' groups), we determined which taxonomic group (within *prd* and 16S rRNA genes) contributed most to the observed pattern of dissimilarity between the clusters, by similarity percentage analysis, SIMPER (Clarke, 1993), performed in PRIMER v6 (Clarke and Gorley, 2006) based on the Bray-Curtis pairwise distance matrices.

Spectral tuning of proteorhodopsins

To predict the spectral tuning of the proteorhodopsins and assess the abundance of the different forms, we assembled the metagenome (assembly details are given in the Supporting Information). The prediction of the protein-coding genes and assessment of their abundance was done in four steps: (i) scaffolds $>$ 1 kb (215,828 in total) from the single combined assembly of G1–G6 metagenomes were used as an input for Prodigal to predict protein coding genes (Hyatt *et al.*, 2010, 2012) in total and (ii) the obtained 620,425 gene sequences were translated and annotated by BLASTp searches against the non-redudant NCBI database and by Pfam domain calling (Punta *et al.*, 2012) against Pfam-A database. (iii) For each sample, the raw reads of each station were mapped separately to the single assembly with Bowtie2, (iv) read counts per gene per sample were counted with Bedtools (Quinlan and Hall, 2010). Spectral tuning analysis was performed on the 133 PR protein sequences that covered the seventh amino acid position after the conserved 'RYIDW' region responsible for spectral tuning, were more than 140 amino acid long and belonged to bacterial proteorhodopsin based on the phylogenetic analysis (Supporting Information Fig. S7). PRs absorbing green light have leucine or methionine at the spectral tuning position, whereas PRs absorbing blue light have glutamine. To correct for the different length assembled for each PR sequence, the raw read counts were multiplied by the full *prd* gene length and divided by the length of the assembled sequence.

Data deposition

Metagenome sequences can be accessed from the European Nucleotide Archive (https://www.ebi.ac.uk/ena) which is part of the EMBL-EBI with the study accession number PRJEB6559, the separate samples can be accessed with the following accession numbers: ERS500138, ERS500137, ERS500136, ERS500131, ERS500083, ERS481549. The assembled 16S rRNA and assembled *prd* sequences were deposited in NCBI GenBank with following accession numbers: KU937394–KU937453 (for 16S rRNA) and KU951281–KU951433 (for *prd*).

Acknowledgements

This studied was funded by the Israel Science Foundation (ISF# 897/12 to LS). MH received a post-doctoral fellowship by the Helmsley Trust and the Haifa University. The post-doctoral scholarship of KS was provided by the Israeli Council for Higher Education (VATAT). We would like to thank Dr. Itai Sharon and Dr. Noa Sher for helpful discussions on metagenomics analysis methods, and the Mediterranean Explorer team and participants to the cruise, Dalit Roth, Eddie Fadeev, and Elad Rachmilovitz for assistance with sampling. We would also like to thank Ron Goldman for assistance with the circulation and FSLE data.

Conflict of interest: The authors declare no conflict of interest.

References

Barale, V., Jaquet, J.M., and Ndiaye, M. (2008) Algal blooming patterns and anomalies in the Mediterranean Sea as derived from the SeaWiFS data set (1998-2003). *Remote Sens Environ* **112:** 3300–3313.

Beja, O., Aravind, L., Koonin, E.V., Suzuki, M.T., Hadd, A., Nguyen, L.P., *et al*. (2000) Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* **289:** 1902–1906.

Beja, O., Spudich, E.N., Spudich, J.L., Leclerc, M., and DeLong, E.F. (2001) Proteorhodopsin phototrophy in the ocean. *Nature* **411:** 786–789.

Bengtsson, J., Eriksson, K.M., Hartmann, M., Wang, Z., Shenoy, B.D., Grelet, G.A., *et al*. (2011) Metaxa: a software tool for automated detection and discrimination among ribosomal small subunit (12S/16S/18S) sequences of archaea, bacteria, eukaryotes, mitochondria, and chloroplasts in metagenomes and environmental sequencing datasets. *Anton Van Leeuw Int J Gen Mol Microbiol* **100:** 471–475.

Boffetta, G., Lacorata, G., Radaelli, G., and Vulpiani, A. (2001) Detecting barriers to transport: a review of different techniques. *Physica D* **159:** 58–70.

Campbell, B.J., Waidner, L.A., Cottrell, M.T., and Kirchman, D.L. (2008) Abundant proteorhodopsin genes in the North Atlantic Ocean. *Environ Microbiol* **10:** 99–109.

Clarke, K.R. (1993) Non-parametric multivariate analyses of changes in community structure. *Aust J Ecol* **18:** 117–143.

Clarke, K.R., and Gorley, R.N. (2006) *PRIMER v6: User Manual/Tutorial*. Plymouth: PRIMER-E.

Clarke, K.R., Somerfield, P.J., and Gorley, R.N. (2008) Testing of null hypotheses in exploratory community analyses: similarity profiles and biota-environment linkage. *J Exp Mar Biol Ecol* **366:** 56–69.

Courties, A., Riedel, T., Jarek, M., Intertaglia, L., Lebaron, P., and Suzuki, M.T. (2013) Genome sequence of strain MOLA814, a proteorhodopsin-containing representative of the betaproteobacteria common in the ocean. *Genome Announc* **1:** 13.

D'Ovidio, F., De Monte, S., Alvain, S., Dandonneau, Y., and Levy, M. (2010) Fluid dynamical niches of phytoplankton types. *Proc Natl Acad Sci U S A* **107:** 18366–18370.

De la Torre, J.R., Christianson, L.M., Beja, O., Suzuki, M.T., Karl, D.M., Heidelberg, J., and DeLong, E.F. (2003) Proteorhodopsin genes are distributed among divergent marine bacterial taxa. *Proc Natl Acad Sci U S A* **100:** 12830–12835.

DeLong, E.F., and Beja, O. (2010) The light-driven proton pump proteorhodopsin enhances bacterial survival during tough times. *PLoS Biol* **8:** e1000359.

Efrati, S., Lehahn, Y., Rahav, E., Kress, N., Herut, B., Gertman, I., *et al*. (2013) Intrusion of coastal waters into the pelagic eastern Mediterranean: in situ and satellite-based characterization. *Biogeosciences* **10:** 3349–3357.

Eiler, A., Hayakawa, D.H., Church, M.J., Karl, D.M., and Rappe, M.S. (2009) Dynamics of the SAR11 bacterioplankton lineage in relation to environmental conditions in the oligotrophic North Pacific subtropical gyre. *Environ Microbiol* **11:** 2291–2300.

Ferrera, I., Borrego, C.M., Salazar, G., and Gasol, J.M. (2014) Marked seasonality of aerobic anoxygenic phototrophic bacteria in the coastal NW Mediterranean Sea as revealed by cell abundance, pigment concentration and pyrosequencing of *pufM* gene. *Environ Microbiol* **16:** 2953–2965.

Finkel, O.M., Beja, O., and Belkin, S. (2013) Global abundance of microbial rhodopsins. *Isme J* **7:** 448–451.

Fuhrman, J.A., Schwalbach, M.S., and Stingl, U. (2008) Proteorhodopsins: an array of physiological roles? *Nat Rev Microbiol* **6:** 488–494.

Ghai, R., Martin-Cuadrado, A.B., Molto, A.G., Heredia, I.G., Cabrera, R., Martin, J., *et al*. (2010) Metagenome of the Mediterranean deep chlorophyll maximum studied by direct and fosmid library 454 pyrosequencing. *Isme J* **4:** 1154–1166.

Giovannoni, S.J., Bibbs, L., Cho, J.C., Stapels, M.D., Desiderio, R., Vergin, K.L., *et al*. (2005) Proteorhodopsin in the ubiquitous marine bacterium SAR11. *Nature* **438:** 82–85.

Goldman, R., Biton, E., Brokovich, E., Kark, S., and Levin, N. (2015) Oil spill contamination probability in the southeastern Levantine basin. *Mar Pollut Bull* **91:** 347–356.

Gomez-Consarnau, L., Gonzalez, J.M., Coll-Llado, M., Gourdon, P., Pascher, T., Neutze, R., *et al*. (2007) Light stimulates growth of proteorhodopsin-containing marine Flavobacteria. *Nature* **445:** 210–213.

Gomez-Consarnau, L., Akram, N., Lindell, K., Pedersen, A., Neutze, R., Milton, D.L., *et al*. (2010) Proteorhodopsin phototrophy promotes survival of marine bacteria during starvation. *PLoS Biol* **8:** e1000358.

Gomez-Consarnau, L., Gonzalez, J.M., Riedel, T., Jaenicke, S., Wagner-Dobler, I., Sanudo-Wilhelmy, S.A., and Fuhrman, J.A. (2015) Proteorhodopsin light-enhanced growth linked to vitamin-B acquisition in marine Flavobacteria. *Isme J* **10:** 1102.

Haller, G., and Yuan, G. (2000) Lagrangian coherent structures and mixing in two-dimensional turbulence. *Physica D* **147:** 352–370.

Hojerova, E., Masin, M., Brunet, C., Ferrera, I., Gasol, J.M., and Koblizek, M. (2011) Distribution and growth of aerobic anoxygenic phototrophs in the Mediterranean Sea. *Environ Microbiol* **13:** 2717–2725.

Hyatt, D., Chen, G.L., LoCascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform* **11:**

Hyatt, D., LoCascio, P.F., Hauser, L.J., and Uberbacher, E.C. (2012) Gene and translation initiation site prediction in metagenomic sequences. *Bioinformatics* **28:** 2223–2230.

Kirchman, D.L., and Hanson, T.E. (2013) Bioenergetics of photoheterotrophic bacteria in the oceans. *Environ Microbiol Rep* **5:** 188–199.

Krom, M.D., Woodward, E.M.S., Herut, B., Kress, N., Carbo, P., Mantoura, R.F.C., *et al*. (2005) Nutrient cycling in the south east Levantine basin of the Eastern Mediterranean: results from a phosphorus starved system. *Deep-Sea Res Part II-Top Stud Oceanogr* **52:** 2879–2896.

Kwon, S.K., Kim, B.K., Song, J.Y., Kwak, M.J., Lee, C.H., Yoon, J.H., *et al*. (2013) Genomic Makeup of the marine *flavobacterium Nonlabens* (*Donghaeana*) *dokdonensis* and identification of a novel class of rhodopsins. *Genome Biol Evol* **5:** 187–199.

Lamy, D., Jeanthon, C., Cottrell, M.T., Kirchman, D.L., Van Wambeke, F., Ras, J., *et al*. (2011) Ecology of aerobic anoxygenic phototrophic bacteria along an oligotrophic gradient in the Mediterranean Sea. *Biogeosciences* **8:** 973–985.

Lauro, F.M., McDougald, D., Thomas, T., Williams, T.J., Egan, S., Rice, S., *et al*. (2009) The genomic basis of trophic strategy in marine bacteria. *Proc Natl Acad Sci U S A* **106:** 15527–15533.

Lehahn, Y., Koren, I., Schatz, D., Frada, M., Sheyn, U., Boss, E., *et al*. (2014) Decoupling physical from biological processes to assess the impact of viruses on a mesoscale algal bloom. *Curr Biol* **24:** 2041–2046.

Lehours, A.C., Cottrell, M.T., Dahan, O., Kirchman, D.L., and Jeanthon, C. (2010) Summer distribution and diversity of aerobic anoxygenic phototrophic bacteria in the Mediterranean Sea in relation to environmental variables. *Fems Microbiol Ecol* **74:** 397–409.

Man, D., Wang, W., Sabehi, G., Aravind, L., Post, A.F., Massana, R., *et al*. (2003) Diversification and spectral tuning in marine proteorhodopsins. *Embo J* **22:** 1725–1731.

Masin, M., Zdun, A., Ston-Egiert, J., Nausch, M., Labrenz, M., Moulisova, V., and Koblizek, M. (2006) Seasonal changes and diversity of aerobic anoxygenic phototrophs in the Baltic Sea. *Aquat Microb Ecol* **45:** 247–254.

Masin, M., Nedoma, J., Pechar, L., and Koblizek, M. (2008) Distribution of aerobic anoxygenic phototrophs in temperate freshwater systems. *Environ Microbiol* **10:** 1988–1996.

Massana, R., Murray, A.E., Preston, C.M., and DeLong, E.F. (1997) Vertical distribution and phylogenetic characterization of marine planktonic Archaea in the Santa Barbara Channel. *Appl Environ Microbiol* **63:** 50–56.

Meyer, F., Paarmann, D., D'souza, M., Olson, R., Glass, E.M., Kubal, M., *et al*. (2008) The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinform* **9:** 386.

Miller, C.S., Baker, B.J., Thomas, B.C., Singer, S.W., and Banfield, J.F. (2011) EMIRGE: reconstruction of full-length ribosomal genes from microbial community short read sequencing data. *Genome Biol* **12:** R44.

Mizuno, C.M., Rodriguez-Valera, F., and Ghai, R. (2015) Genomes of planktonic Acidimicrobiales: widening horizons for marine Actinobacteria by metagenomics. *MBio* **6:** 14.

Nguyen, D., Maranger, R., Balague, V., Coll-Llado, M., Lovejoy, C., and Pedros-Alio, C. (2015) Winter diversity and expression of proteorhodopsin genes in a polar ocean. *Isme J* **9:** 1835–1845.

Palovaara, J., Akram, N., Baltar, F., Bunse, C., Forsberg, J., Pedros-Alio, C., *et al*. (2014) Stimulation of growth by proteorhodopsin phototrophy involves regulation of central metabolic pathways in marine planktonic bacteria. *Proc Natl Acad Sci U S A* **111:** E3650–E3658.

Philosof, A., and Beja, O. (2013) Bacterial, archaeal and viral-like rhodopsins from the Red Sea. *Environ Microbiol Rep* **5:** 475–482.

Punta, M., Coggill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C., *et al*. (2012) The Pfam protein families database. *Nucleic Acids Res* **40:** D290–D301.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., *et al*. (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41:** D590–D596.

Quinlan, A.R., and Hall, I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26:** 841–842.

Riedel, T., Tomasch, J., Buchholz, I., Jacobs, J., Kollenberg, M., Gerdts, G., *et al*. (2010) Constitutive expression of the proteorhodopsin gene by a flavobacterium strain representative of the proteorhodopsin-producing microbial community in the North Sea. *Appl Environ Microbiol* **76:** 3187–3197.

Rusch, D.B., Halpern, A.L., Sutton, G., Heidelberg, K.B., Williamson, S., Yooseph, S., *et al*. (2007) The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol* **5:** e77.

Sabehi, G., Massana, R., Bielawski, J.P., Rosenberg, M., Delong, E.F., and Beja, O. (2003) Novel proteorhodopsin variants from the Mediterranean and Red Seas. *Environ Microbiol* **5:** 842–849.

Sabehi, G., Loy, A., Jung, K.H., Partha, R., Spudich, J.L., Isaacson, T., *et al*. (2005) New insights into metabolic properties of marine bacteria encoding proteorhodopsins. *PLoS Biol* **3:** e273.

Sabehi, G., Kirkup, B.C., Rozenberg, M., Stambler, N., Polz, M.F., and Beja, O. (2007) Adaptation and spectral tuning in divergent marine proteorhodopsins from the Eastern Mediterranean and the Sargasso Seas. *Isme J* **1:** 48–55.

Schattenhofer, M., Fuchs, B.M., Amann, R., Zubkov, M.V., Tarran, G.A., and Pernthaler, J. (2009) Latitudinal distribution of prokaryotic picoplankton populations in the Atlantic Ocean. *Environ Microbiol* **11:** 2078–2093.

Shimada, A., Nishijima, M., and Maruyama, T. (1995) Seasonal Appearance of Prochlorococcus in Suruga Bay, Japan, in 1992-1993. *J Oceanogr* **51:** 289–300.

Sjostedt, J., Martiny, J.B., Munk, P., and Riemann, L. (2014) Abundance of broad bacterial taxa in the sargasso sea explained by environmental conditions but not water mass. *Appl Environ Microbiol* **80:** 2786–2795.

Steindler, L., Schwalbach, M.S., Smith, D.P., Chan, F., and Giovannoni, S.J. (2011) Energy starved *Candidatus* Pelagibacter ubique substitutes light-mediated ATP production for endogenous carbon respiration. *PLoS One* **6:** e19725.

Stingl, U., Desiderio, R.A., Cho, J.C., Vergin, K.L., and Giovannoni, S.J. (2007) The SAR92 clade: an abundant coastal clade of culturable marine bacteria possessing proteorhodopsin. *Appl Environ Microbiol* **73:** 2290–2296.

Tanaka, T., Zohary, T., Krom, M.D., Law, C.S., Pitta, P., Psarra, S., *et al*. (2007) Microbial community structure and function in the Levantine Basin of the Eastern Mediterranean. *Deep-Sea Res Part I-Oceanogr Res Pap* **54:** 1721–1743.

Thingstad, T.F., Krom, M.D., Mantoura, R.F., Flaten, G.A., Groom, S., Herut, B., *et al*. (2005) Nature of phosphorus limitation in the ultraoligotrophic Eastern Mediterranean. *Science* **309:** 1068–1071.

Tsiola, A., Pitta, P., Fodelianakis, S., Pete, R., Magiopoulos, I., Mara, P., *et al*. (2015) Nutrient limitation in surface waters of the oligotrophic eastern mediterranean sea: an enrichment microcosm experiment. *Microb Ecol* **71:** 575.

Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., *et al*. (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304:** 66–74.

Yoshizawa, S., Kawanabe, A., Ito, H., Kandori, H., and Kogure, K. (2012) Diversity and functional analysis of proteorhodopsin in marine Flavobacteria. *Environ Microbiol* **14:** 1240–1248.

Yutin, N., Suzuki, M.T., Teeling, H., Weber, M., Venter, J.C., Rusch, D., and Beja, O. (2007) Assessing diversity and biogeography of aerobic anoxygenic phototrophic bacteria in surface waters of the Atlantic and Pacific Oceans using the Global Ocean Sampling expedition metagenomes. *Environ Microbiol* **9:** 1464–1475.

Zhang, Y., and Jiao, N. (2007) Dynamics of aerobic anoxygenic phototrophic bacteria in the East China Sea. *Fems Microbiol Ecol* **61:** 459–469.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Fig. **S1.** Vertical section profiles of oceanographic parameters along a longitudinal transect between stations G1 and G4 measured *in situ*. White dotted lines denote the CTD sampling points, black lines are contours. The profiles are: (A) temperature, (B) salinity, (C) fluorescence and (D) dissolved oxygen. Figures were made using Ocean Data View version 4.6.2 (http://odv.awi.de).

Fig. **S2.** Phylogenetic tree of 16S rRNA gene sequences based on 97% OTUs reconstructed from the metagenome reads by EMRIGE. Sequence alignment was performed with SINA v1.2.11 (Pruesse et al., 2012), common gaps were deleted and the alignment was improved manually. Gblocks (v. 0.91) (Castresana, 2000) with the relaxed setting was used to remove ambiguously aligned positions. The final dataset (65 OTU from EMIRGE and 21 reference sequences) with 1189 positions was used to reconstruct the phylogenetic tree using MEGA v6.06 (Tamura et al., 2013) with Neighbor-Joining algorithm. Pairwise distances were based on the Kimura-2-parameter model. Bootstrap values (1000 replicates) greater than 70% (0.7) are displayed. The scale bar indicates nucleotide changes per position. Names on the right represent the bacterial orders and classes as determined by the closest affiliation to sequences of known phylogeny. Vertical colored bars on the right of the tree represent the total relative abundance of all the OTUs included in a specific order (e.g. Flavobacteria) or clade (e.g. SAR11), at the different stations (represented by colors, legend on the top left). The tree was rooted with the clade containing OTU1, an archaeon phylogenetically affiliated with the euryarchaeote *Candidatus* Thalassoarchaea mediterranii.

Fig. **S3.** Clustering of microbial community genes deriving from the six metagenomic samples examined using hierarchical clustering dendogram based on Bray-Curtis dissimilarity: (A) based on *prd* gene, (B) *recA* gene, (C) *pufM* gene and (D) *psaA* gene. Differently colored branches indicate significantly different ($p<0.05$) clusters of samples as tested by SIMPROF analysis.

Fig. **S4.** Ratios of representative diagnostic phytoplankton pigments indicative of: (A) diatoms, (B) cyanobacteria, (C) *Prochlorococcus*. Abbreviations are as follows: Fuco = Fucoxanthin, 19-BF = 19-Butanoyloxyfucoxanthin, 19-HF = 19-Hexanoyloxyfucoxanthin, Zea = Zeaxanthin, Div chl *a* = divinyl chlorophyll *a*, Chl *a* = Chlorophyll *a*, Chl c_1/c_3 = Chlorophyll c_1/c_3.

Fig. **S5.** Relative abundance of PR-bearing bacteria at the order level in the sampled stations. Taxonomic annotations were assigned based on BLASTx of extracted *prd* reads against the NCBI NR protein database (see Experimental Procedures for details). Orders with relative abundance below 0.5% at each station were grouped under "Low abundance orders". Further details on *prd* taxonomy at class and order level and the comparison of *prd* abundance between 'coastal impacted' versus 'offshore' stations are given below in the supplementary results.

Fig. **S6.** Relative abundance of AAP bacteria at the order level in the sampled stations. Taxonomic annotations were assigned based on BLASTx of extracted *pufM* reads against the NCBI NR protein database (see Experimental Procedures for details).

Fig. **S7.** Neighbor joining tree of translated assembled *prd* gene sequences with more than 140 amino acid: A) overview of all groups, B) extended view of the collapsed proteorhodopsin group I and C) of the collapsed proteorhodopsin group II. Sequences from this study are marked by a red dot. Sequences from Uniref90 and the NCBI NR protein databases were added for context. The sequences for this tree were aligned with Muscle (Edgar, 2004) using the EMBL-EBI analysis tool web services (McWilliam et al., 2013). The obtained alignment was manually improved. The neighbor-joining tree was constructed in MEGA v6.06 (Tamura et al., 2013). Distances between sequences were calculated using the Dayhoff model for each pair of sequences after removing all ambiguous positions for each pair. Numbers at nodes indicate bootstrap support >70% based on 100 repetitions. Scale bars indicate number of amino acid substitutions per site. Abbreviations are as follows: PR: proteorhodopsins, ActR: marine actinobacterial clade rhodopsins, XR: xanthorhodopsins, VR: viral proteorhodopsins, BR: bacteriorhodopsins, HR: halorhodopsins, SR: sensory rhodopsins (with I and II identifying groups).

Table S1. Sampling station locations and concentrations of dissolved inorganic nutrients and extracted chlorophyll *a* (10 m depth). Note: Chlorophyll *a* was extracted from two separate filters (a comma separates the value for each replicate), the first value is from a filter that was stored for four months and the second value is from a filter stored for nine months before extraction. The standard error of the instrument was 0.05 μM for $NO_2^- + NO_3^-$ and SiO_4, and 0.04 μM PO_4^{3-}.

Table S2. Abundance (fraction out of total cells) of picophytoplankton (including picoeukaryotes and *Prochlorococcus*),

Synechococcus, AAP and heterotrophic bacteria as determined by epifluorescence microscopy cell counts in the sampled station.

Table S3. Abundance of picoeukaryotic phytoplankton, *Prochlorococcus*, *Synechococcus*, picophytoplankton as determined by flow cytometry.

Table S4. HPLC separated and identified pigments of the sampled stations. Area refers to the intensity of the peaks. ND stands for: not detected. The two most prominent pigments were 19- hexanoyloxyfucoxanthin (19-HF) and zeaxanthin, the former indicative of prymnesiophytes (including coccolithophores) and the latter of picophytoplanktonic cyanobacteria (*Synechococcus* and *Prochlorococcus*). Ratios between diagnostic pigments (Fig. S4, Table S4) were used as indication of the phytoplankton they represent (Sweeney et al., 2003; Mendes et al., 2007; Moore et al., 2007b). The high zeaxanthin:chl *a* and divinyl chl *a*:chl *a* ratios at station G2 suggest a higher abundance of picophytoplanktonic cyanobacteria and *Prochlorococcus*, respectively (Fig. S4). This is supported by 16S rRNA data (Metaxa2 analysis), microscopy (with the caveat that the relevant category contains also eukaryotic picopyhtoplankton), and flow cytometry (Table S3).

Table S5. Total read counts and relative abundance of reads belonging to different domains of life in the metagenomes of the six different stations and two read directions obtained by MG-RAST for predicted proteins and ribosomal RNA genes using all annotation source databases. Only reads that passed the quality control of MG-Rast were considered.

Table S6. Overview of bacterial hits in each station and read direction (R1 and R2) for all analyzed genes. Gene abbreviation are as follows: *prd*: proteorhodopsin, XR: xanthorhodopsin gene, MAC: marine actinobacterial clade rhodopsin gene, *pufM*: photosynthetic reaction center subunit M, *psaA*: Photosystem I P700 chlorophyll *a* apoprotein A1, *recA*: recombinase A, *gyrB*: DNA gyrase subunit B, *rpoB*: RNA polymerase subunit B, *tuf*: elongation factor tu.