A peer-reviewed version of this preprint was published in PeerJ on 23 October 2018.

Sieradzki ET, Fuhrman JA, Rivero-Calle S, Gómez-Consarnau L. 2018. Proteorhodopsins dominate the expression of phototrophic mechanisms in seasonal and dynamic marine picoplankton communities. PeerJ 6:e5798 https://doi.org/10.7717/peerj.5798

Proteorhodopsins dominate the expression of phototrophic mechanisms in seasonal and dynamic marine picoplankton communities

Ella T Sieradzki [Corresp., 1], **Jed A Fuhrman** [1], **Sara Rivero-Calle** [1], **Laura Gómez-Consarnau** [1]

[1] Department of Biological Sciences, University of Southern California, Los Angeles, CA, United States

Corresponding Author: Ella T Sieradzki
Email address: ellasiera@berkeley.edu

The most abundant and ubiquitous microbes in the surface ocean use light as an energy source, capturing it via complex chlorophyll-based photosystems or simple retinal-based rhodopsins. Studies in various ocean regimes compared the abundance of these mechanisms, but few investigated their expression. Here we present the first full seasonal study of abundance and expression of light-harvesting mechanisms (proteorhodopsin, PR; aerobic anoxygenic photosynthesis, AAnP; and oxygenic photosynthesis, PSI) from deep-sequenced metagenomes and metatranscriptomes of marine picoplankton (< 1 µm) at three coastal stations of the San Pedro Channel in the Pacific Ocean. We show that, regardless of season or sampling location, the most common phototrophic mechanism in metagenomes of this dynamic region was PR (present in 65-104% of the genomes as estimated by single-copy recA), followed by PSI (5-104%) and AAnP (5-32%). Furthermore, the normalized expression (RNA to DNA ratio) of PR genes was higher than that of oxygenic photosynthesis (average±standard deviation 26.2±8.4 vs. 11±9.7), and the expression of the AAnP marker gene was significantly lower than both mechanisms (0.013±0.02). We demonstrate that rhodopsin expression was dominated by the SAR11-cluster year-round, followed by other Alphaproteobacteria, unknown-environmental clusters and Gammaproteobacteria. This highly dynamic system further allowed us to identify a trend for PR spectral tuning, in which blue-absorbing PR genes dominate in areas with low chlorophyll-*a*concentrations (< 0.25 µg/L). This suggests that PR phototrophy is not an accessory function but instead a central mechanism that can regulate photoheterotrophic population dynamics.

Proteorhodopsins dominate the expression of phototrophic mechanisms in seasonal and dynamic marine picoplankton communities

Ella T. Sieradzki, Jed A. Fuhrman, Sara Rivero-Calle, Laura Gómez-Consarnau

Abstract

The most abundant and ubiquitous microbes in the surface ocean use light as an energy source, capturing it via complex chlorophyll-based photosystems or simple retinal-based rhodopsins. Studies in various ocean regimes compared the abundance of these mechanisms, but few investigated their expression. Here we present the first full seasonal study of abundance and expression of light-harvesting mechanisms (proteorhodopsin, PR; aerobic anoxygenic photosynthesis, AAnP; and oxygenic photosynthesis, PSI) from deep-sequenced metagenomes and metatranscriptomes of marine picoplankton (< 1 µm) at three coastal stations of the San Pedro Channel in the Pacific Ocean. We show that, regardless of season or sampling location, the most common phototrophic mechanism in metagenomes of this dynamic region was PR (present in 65-104% of the genomes as estimated by single-copy recA), followed by PSI (5-104%) and AAnP (5-32%). Furthermore, the normalized expression (RNA to DNA ratio) of PR genes was higher than that of oxygenic photosynthesis (average±standard deviation 26.2±8.4 vs. 11±9.7), and the expression of the AAnP marker gene was significantly lower than both mechanisms (0.013±0.02). We demonstrate that rhodopsin expression was dominated by the SAR11-cluster year-round, followed by other Alphaproteobacteria, unknown-environmental clusters and Gammaproteobacteria. This highly dynamic system further allowed us to identify a trend for PR spectral tuning, in which blue-absorbing PR genes dominate in areas with low chlorophyll-*a*concentrations (< 0.25 µg/L). This suggests that PR phototrophy is not an accessory function but instead a central mechanism that can regulate photoheterotrophic population dynamics.

Introduction

Sunlight is the most readily available source of energy in the photic zone of the ocean. Light utilization in marine microorganisms is divided between complex, high-yield photosystems (oxygenic and anoxygenic photosynthesis) and simple, low-yield rhodopsins (Finkel et al., 2013). Light-harvesting mechanisms span the entire visible light spectrum, with bacteriochlorophyll-a and chlorophyll-a (Chl-a) utilizing its extremes, and various types of rhodopsins absorbing intermediate frequencies (Fuhrman et al., 2008). Since the discovery of proteorhodopsin proteins (PR; Beja et al., 2000; 2001), several surveys in various oceanic regimes demonstrated their global abundance (Rusch et al., 2007; Boeuf et al., 2016; Brindefalk et al., 2016; Dubinsky et al., 2017). Furthermore, genomic studies showed that the gene coding for PR is present in some of the most abundant bacteria in the ocean, e.g. SAR11 and SAR86 (Beja et al., 2000; Sabehi et al., 2004; Giovannoni, 2005). PR-coding genes have also been found in some microbial eukaryotes such as fungi and photosynthetic protists, as well as in archaea and even in viruses as an auxiliary metabolic gene (Philosof and Beja, 2013; reviewed by Pinhassi et al., 2016). Microbial rhodopsins are the simplest light-harvesting mechanisms known to date, containing only one membrane protein and a retinal chromophore (Béjà et el. 2000). Light-

driven proton pump PRs can increase the membrane potential of the cell, ultimately supporting a variety of processes such as ATP synthesis (Béjà et al, 2000; Walter et al. 2007; Steindler et al. 2011), substrate uptake (Steindler et al. 2011; Gómez-Pereira et al. 2013; Gómez-Consarnau et al. 2016), survival during starvation (Gómez-Consarnau et al. 2010; Steindler et al. 2011) and/or salinity stress response (Feng et al. 2013). Taken together, their structural simplicity and the range of functions they can support seem to have promoted the expansion of microbial rhodopsins in the sunlit ocean. However, estimates of the relative abundance of PR genes using metagenomics (Finkel et al. 2013; Brindefalk et al. 2016; Dubinsky et al. 2017) or metatranscriptomics (Shi et al. 2011; Kopf et al. 2015) have only been examined recently. In contrast to qPCR methods, next generation sequencing techniques can provide more reliable estimates without introducing qPCR and cloning biases that would miss certain PR gene types (Nguyen et al., 2015; Boeuf et a. 2016).

Rhodopsin genes are highly expressed in the photic zone (Frias-Lopez et al., 2008; Poretsky et al., 2009; Satinsky et al., 2014). While generally transcription is not always an indicator of protein activity, one study shows good correlation between transcription of proteorhodopsin and PR synthesis in a diatom (Marchetti et al., 2015) and another shows diel oscillations in proteorhodopsin transcription that peak before dawn, implying a preparation for light harvesting during the day (Ottesen et al., 2014). Combined, these results may indicate that transcription levels are a good proxy for PR synthesis. With that, there is limited information on the expression of rhodopsins compared to other light-harvesting mechanisms. In contrast with the established global distribution and abundance of rhodopsin taxonomic clades, very few studies have compared their expression in environmental samples (Shi et al., 2011; Kopf et al., 2015; Boeuf et al., 2016; Brindefalk et al., 2016; Vader et al., 2018). Additionally, the vast majority of studies were based on single time-points, with the exception of Sabehi et al. (2007), which compared winter and summer expression at two sites (Mediterranean and Sargasso Sea) and Nguyen et al. (2015), which compared early- and late-winter expression in the arctic. Thus, information on temporal expression patterns of different rhodopsin clades remains scarce.

Proteorhodopsin has two main variants that differ in their light absorption spectrum (Béjà et al., 2001; Man et al., 2003). This spectral tuning is determined by a single residue at the frequency-tuning site (FTS) (Man et al., 2003). It has been proposed that spectral tuning is related to the spectral quality and quantity of light in the water, i.e. water color. Consistent with this pattern, green-tuned proteorhodopsin are generally common in coastal waters, whereas the blue-tuned counterparts are typical of open ocean or deeper water (Fuhrman et al., 2008 and references therein; Pinhassi et al., 2016). For instance, while more than 70% of the PR sequences retrieved from the ultraoligotrophic Eastern Mediterranean were classified as blue-absorbing (Dubinsky et al. 2017), less than 10% belonged to this group in the eutrophic Baltic sea (Brindefalk et al. 2016). No study to date has evaluated: 1) whether this distribution pattern applies to more dynamic environments with contrasting trophic conditions associated to seasonal and spatial gradients, or 2) whether the specific underwater light field could be an important ecological driver for photoheterotrophic populations in nutrient dynamic regions.
Here we present the first seasonal study of PR in metagenomes and metatranscriptomes of surface water microbial communities at three contrasting locations of the San Pedro Channel (fig. 1). The transect spanned 23 miles between the highly polluted Port of Los Angeles and the mildly impacted Santa Catalina Island, with the largely oligotrophic San Pedro Ocean Time-

series (SPOT) halfway between them. We further compared PR abundance and distribution to the other two main phototrophic metabolisms in surface waters: oxygenic photosynthesis and aerobic anoxygenic phototrophy. Our data show that PR is the dominant phototrophic metabolism in microbial metagenomes and metatranscriptomes of this dynamic environment year-round.

Materials and methods

Sample collection

Surface seawater was collected from the Port of Los Angeles (33°42.75'N 118°15.55'W), the San Pedro Ocean Time-series (33°33.00'N 118°24.01'W) and Two Harbors, Santa Catalina Island (33°27.18'N 118°28.51'W) in four seasons: July 2012, October 2012, January 2013 and April 2013. The water was prefiltered through a 1 µm glass fiber filter and then through a 0.22 µm Sterivex polyethersulfone (PES, Millipore, SVGPL10RC) filter, preserved in RNAlater, flash frozen and kept in -80°C until extraction.

Chemical and biological parameters

Whole seawater for nutrients analysis were collected and kept in -20°C until they were sent to analysis at the Marine Sciences Institute Analytical Lab at University of California, Santa Barbara. Bacteria and viruses per ml seawater were counted using SYBR green epifluorescence microscopy (Noble and Fuhrman, 1998; Patel et al., 2007). Chlorophyll-a concentration measurements were courtesy of the Caron lab at the University of Southern California.

16S/18S-rRNA amplification and sequencing

Hypervariable regions V4-V5 were amplified from DNA and cDNA of all samples following the protocol described in Parada et al. (2015). The products were bead-cleaned with Ampure beads at a 1x ratio, diluted to 1 ng/µl and pooled. The pool was bead cleaned again at a 0.8x beads to pool ratio, insert size was verified on an Agilent 2100 Bioanayzer and sequenced on Illumina MiSeq 300 bp paired-end at the UC Davis genome core.

The resulting reads were quality-trimmed using Trimmomatic 0.3 (Bolger et al., 2014) with parameters set to Leading:20 Trailing:20 Slidingwindow:15:25. The reads were then merged with Usearch 7 (Edgar, 2010) and analyzed using Mothur (Kozich et al., 2013) following the Miseq SOP (https://www.mothur.org/wiki/MiSeq_SOP).

Reads that failed to merge, and were therefore more likely to represent 18S-rRNA, were first concatenated with an additional N base between the forward and reverse-complemented reverse read as described in Needham et al., 2016, and then clustered with usearch 6.1 (Edgar, 2010) at 97% identity via the Qiime framework. Taxonomy was assigned using SILVA release 132 (Yilmaz et al., 2014). These reads were used only to estimate the relative abundance of diatoms in all samples.

Metagenomic and metatranscriptomic Library preparation

RNAlater was removed from the Sterivex filters as much as possible in order to improve DNA yield. Cells on the filters were lysed by bead-beating for 2 cycles of 10 minutes each with glass beads in 1.5 ml STE buffer injected into the Sterivex. DNA and RNA were then extracted from the flow-through using the AllPrep kit (Qiagen) that yields RNA and DNA from the same sample simultaneously. After extraction and quality assessment with Qubit HS (Thermo-Fisher

Scientific) and Bioanalyzer 2100 (Agilent) nucleic acids were stored at -80°C until further processing. RNA samples were spiked with an internal standard (ERCC RNA Spike-In Mix, Thermo-Fisher 4456740). Libraries were prepared using Ovation Ultralow library system V2 (Nugen, 0344) and sequenced on Illumina HiSeq 2x125 bp or 2x150 bp for metagenomes and 2x250 bp for metatranscriptomes.

Sequence quality trimming

Quality trimming was performed using Trimmomatic 0.33 (Bolger et al., 2014) with parameters set to Leading:20 Trailing:20 Slidingwindow:15:25. Internal standard reads were removed from the metatranscriptomes informatically. Metatranscriptomic reads were merged with PEAR (Zhang et al., 2013). Metagenomic reads could not be merged due to insert length and only the forward read was used. Sequencing depth after quality control is detailed in sup. table S2.

Marker genes selection

Each of the light-harvesting mechanisms has established marker genes (Finkel et al., 2013), which are both single-copy in most cases and can be found in all organisms that use said mechanisms. These genes can also be correlated to phylogeny, albeit not at very a high resolution and with the caveat that they (mainly proteorhodopsin) can be laterally transferred. Those marker genes can be used to track global distribution as well as expression. The *prd* gene codes for the membranal protein of proteorhodopsin, which anchors the retinal antenna. *psaA* codes for apoprotein a1 which binds P700, the main electron donor of photosystem-I and *pufM* codes for chain M in the reaction center of the anoxygenic bacteriochlorophyll (Finkel et al., 2013).

Assembly

Contigs were assembled within each metagenome/metatranscriptome separately with Megahit v1.0.4 (Li et al., 2016) and clustered with cd-hit (Fu et al., 2012) at 99% identity. Contigs longer than 2000 kbp were co-assembled with Minimus2 (Sommer et al., 2007) and shorter contigs were co-assembled with Newbler (Margulies et al., 2005). Both co-assemblies required minimum overlap 40 bp minimum identity 99% and clustered again with cd-hit at 99% identity.

Marker gene extraction from assemblies

Open reading frames (ORFs) were identified using prodigal version 2.6.2 (Hyatt et al., 2010). The resulting translated ORFs were then searched for Prd, PsaA, PufM, PufL and RecA proteins via Anvi'o (Eren et al. 2015), and ORF sequences long enough to not affect the curated alignment were added to the protein dataset used for phylogenetic placement (see below). Assembled PsaA ORFs were all placed in the eukaryotic clade (sup. fig. S4), assembled Prd ORFs represented multiple clades and contributed significantly to recruitment (sup. table S3), and no assembled PufM ORFs matched our criteria.

Phylogenetic trees

Curated protein subsets limited to aquatic bacteria, archaea, viruses and picoeukaryotes of PsaA, PufM and PufL were downloaded from Pfam (Finn et al., 2016) and RefSeq. These sets were supplemented by those protein sequences found in the assembled contigs. Two sets of sequences were aligned using mafft (Katoh et al., 2013) (globalpair, gap open penalty 1.5, gap extension penalty 0.5 and scoring matrix BLOSUM30) and alignment trimming (Gblocks b3=50, b4=5,

b5=h, Castresana, 2007): one set of *psaA* only and the other of *psbA*, *pufM* and *pufL* which are homologous. Each alignment was then used to build a Hidden Markov Model (HMM) using HMMER 3.0 (Johnson et al., 2010) and a maximum likelihood tree with RAxML v8.2.5 (Stamatakis, 2014) using WAG substitution matrix and Gamma model (Ignacio-Espinoza et al., 2012). The trees are provided under supplementary figures S4 and S6.
A curated alignment and a phylogenetic tree of proteorhodopsin gene *prd* was graciously provided by the MicRhoDE project (Boeuf et al., 2015). The alignment was used to build an HMM of the Prd amino acid sequence via hmmbuild. The protein sequences from the assembled contigs were first placed into the MicRhoDE tree (see sup. table S3 for placements) and the resulting tree was used for placement of short reads.

Short reads placement
Most studies use either best blast hit or reciprocal blast to recruit reads to rhodopsins, whereas we used a combination of blastx, a HMM (Hidden Markov Model) and placement of short translated reads into a phylogenetic protein tree. This method almost always yielded many more reads than reciprocal blast, which is intentionally a very conservative estimate (sup. fig. S5).
All reads from the metagenomes and metatranscriptomes were searched against the curated protein datasets using blastx (Camacho et al., 2009) and requiring an e-value of 10^{-5}. Reads that hit those genes were translated and searched again using the HMMs with hmmsearch, and aligned to the dataset using hmmalign. The remaining reads were placed into the phylogenetic trees with pplacer 1.1 (Matsen et al., 2010).
The same process with the exception of placement into a phylogenetic tree was performed for RecA and used for normalization of the functional genes.
Gene abundances were determined by the formula
(funcAbun/funcLen)/(RecAAbun/RecALen)
where func is any functional gene (*psaA*, *prd* or *pufM*), funcAbun is the relative abundance of reads placed into leaves in the phylogenetic tree of this functional gene per sample, funcLen is the length of the HMM built for the functional gene, RecAAbun is the relative abundance of reads mapped to RecA per sample by HMM and RecALen is the length of the RecA HMM.

Reciprocal blast
For comparability to previous papers, reads mapping to all genes were also extracted from the metagenomes and metatranscriptomes using reciprocal blast. First, we built a blast database from every metagenome and metatranscriptome. Then we used the curated sequences as a query to search these databases using tblastn (Camacho et al., 2009). The reads that resulted from this search were then searched against the NCBI non-redundant database (nr) using blastx (Camacho et al., 2009), and only reads that hit the desired genes were retained. General trends between genes were similar using this method compared to HMMs but the number of recruited reads was almost always significantly lower for the functional genes (sup. table S2, sup. fig. S5).

Blue/green tuning of proteorhodopsin
Reads that mapped to the frequency tuning site (FTS) in the protein alignment (using HMMalign) were analyzed to determine tuning relative abundance of blue (glutamine) or green (leucine or methionine). While other residues were observed, they were extremely rare and therefore not included in the analysis.

Remote sensing data acquisition and analysis

MODIS level 3 mapped daily 4km resolution satellite products of remote sensing reflectance (RRS) were downloaded from the NASA ocean color distribution website (https://oceancolor.gsfc.nasa.gov) in October 2017. Satellite products for this study were extracted for each location and date spanning six days before and one day after sampling, therefore exploring the short temporal variability as well. To account for cloud cover in some dates and locations on the day of sampling, we used the corresponding data from the next day. Additionally, we averaged our satellite estimates within a 0.15 degree radius from each sampling location.

Statistics

Shannon index of evenness was calculated using the R package RAM and the one-sided paired t-test between gene abundance evenness and expression evenness was run using R basic package with mu=0.

Spearman correlations were calculated using the corr.test function within the R package psych (https://cran.r-project.org/web/packages/psych/psych.pdf). This function can calculate Spearman correlations with p-value correction for multiple tests (we used option "fdr" for the correction).

Data availability

All raw data (16S/18S, metagenomes and metatranscriptomes) can be found on EMBL-ENA under project number PRJEB12234. Metatranscriptomics sequences accession numbers are ERS1864892-ERS1864903, and negative control library sequences accession number is ERR2089009. Metagenomic sequences accession numbers are ERS1869885-ERS1869896 and negative control accession number is ERS1872073.

Assembled amino acid sequences of Prd, PsaA and RecA can be found at https://figshare.com/collections/Dimensions_of_Biodiversity_-_San_Pedro_Channel/4099757 .

Results

Abundance and expression of light-harvesting mechanisms

We estimated the fraction of microbial cells that contain each of the light-harvesting mechanisms by normalizing the relative abundance of genes coding for rhodopsins (*prd*), photosystem-I (*psaA*) and aerobic anoxygenic photosynthesis (*pufM*) to the relative abundance of the single-copy housekeeping gene *recA* (Finkel et al., 2013; Brindefalk et al., 2016; Dubinsky et al., 2017) which is also used as a baseline for RNAseq experiment normalization (Rocha et al., 2015). The relative abundance of each of these genes in metatranscriptomes was then divided by their relative abundance in metagenomes to generate a normalized RNA to DNA ratio for each gene within each sample. Regardless of season or sampling location, the most common phototrophic mechanism in metagenomes was PR (65-104% compared to *recA*), followed by PSI (5-104%) and AAnP (5-32%) (fig. 2A). In fact, in 10 out of 12 samples the *prd* gene exceeded 80% of *recA* abundance (sup. fig. S2). The percentage for PSI was variable while the percentages of PR remained within a narrower range through the different seasons and stations (fig. 2A). The expression of both *prd* and *psaA* was consistently 1-2 orders of magnitude higher than their respective gene abundance (fig. 2B). The RNA to DNA ratio also revealed that while *pufM* gene abundance was sometimes comparable to *psaA* (fig. 2A), its expression was 2-3 orders of

magnitude lower (fig. 2B). *pufM* expression was also 2-3 orders of magnitude lower than *pufM* gene abundance, suggesting that most of the AAnP bacteria in our samples were not actively performing this type of phototrophy (fig. 2B). Interestingly, we did not observe any geographical or seasonal trends for the presence of any light-harvesting strategies (fig. 2).
The combined normalized relative abundance of PR and PSI was higher than 100% in all samples analyzed, suggesting multiple gene copy numbers or coexistence of these mechanisms within the same prokaryotic cells (Finkel et al. 2013; Dubinsky et al 2016), as previously shown in marine eukaryotic algae (Marchetti et al., 2012; 2015). While we did not observe any significant correlation between abundance of PR and PSI genes or transcripts, strong negative correlations between PR expression and total chlorophyll-*a* (Chl-*a*) concentrations were clearly identified at POLA and SPOT (fig. 3).

Proteorhodopsin distribution by cluster

While our metagenomes revealed high taxonomic evenness of rhodopsin clusters (fig. 4B, sup. fig. S1, average Shannon index of evenness 0.7±0.06), expression was dominated by the SAR11 cluster (52±14%) followed by other Alphaproteobacteria (12±7%), Gammaproteobacteria (15±6%) and unknown environmental clusters (8±2%) (fig. 4 A). Evenness in expression was always lower than evenness in gene abundance within a sample (one-sided paired t-test on Shannon index of evenness, p=0.0003) except at POLA in April 2013 (sup. fig. S1). This particular sample was collected during a localized algal bloom with the highest Chl-a concentration measured in this study (12.7 µg/L). The AAnP RNA to DNA ratio in this sample was the highest we detected, and this was the only time in which expression of PSI surpassed that of PR (fig. 2B). PR expression in this sample was dominated by Gammaproteobacteria (30%) and SAR11 *prd* expression dropped to 16%.

The SAR11 cluster. We further examined the *prd* gene abundance and expression patterns of the SAR11 cluster, as this was the most abundant PR-containing group overall. SAR11 *prd* expression correlated positively with expression of SAR11 OTUs determined by 16S-rRNA (fig. 5B). This correlation was also observed at the gene level after removing one outlier (POLA April 2013) (fig. 5A). We further examined the presence and expression of specific SAR11 strains within the SAR11 cluster by calculating recruitment per leaf on the MicRhoDE phylogenetic tree of rhodopsins (Boeuf et al., 2015). We found that the high-resolution expression patterns were also much less even than gene distribution, where the top 10 most highly expressed leaves generally accounted for >70% of the SAR11 *prd* transcripts compared to less than 50% of gene abundance (fig. 6). The mean Shannon index of evenness for gene abundance was 0.80±0.09, and significantly lower for expression: 0.57±0.08 (Wilcoxon rank sum test, paired one-sided, p=0.0005). Only 4 of the 10 most expressed *prd* transcripts were also in the top 10 highest gene abundances (fig. 6C,D). Thus, it would appear that there are SAR11 strains in our samples that have low to no PR activity or do not contain a proteorhodopsin gene. In order to address the possibility that there are SAR11 strains which do not carry the *prd* gene and could decouple SAR11 16S-rRNA relative gene abundance from *prd* gene abundance, we attempted to find Spearman correlations between the relative abundance of specific SAR11 OTUs and the sum of all SAR11 *prd* gene abundance, but none were significant.

Other clusters. Although we found 13 viral rhodopsin open reading frames (ORFs) in our assemblies, the viral PR cluster does not appear to be expressed (no more than 2.2% of the total rhodopsin transcripts per sample). Archaeal clusters were extremely rare (<1% of the metagenomic and <2% of metatranscriptomic reads recruiting to *prd*).
Expression and gene abundance of eukaryotic rhodopsins were not apparent in any of the samples. Most of the eukaryotes that are known to carry rhodopsins, such as diatoms and dinoflagellates, are large and not expected to be present in our <1 µm size-fraction (Marchetti et al., 2015; Vader et al. 2018). The picoeukaryotes *Micromonas* spp. and *Bathycoccus* spp. that have been shown to contain rhodopsin genes and are sometimes found in the San Pedro Channel were not present in the small size-fraction of our samples.

PR Spectral tuning

We analyzed the spatiotemporal distribution of the two main PR variants (blue and green) in the metagenomes and compared our results to previously reported data (Brindefalk et al. 2016; Dubinsky et al. 2017). Consistent with being a coastal environment, the majority of our samples were dominated by green-absorbing PR genes (fig. 7A). However, samples collected at CAT and SPOT in October were dominated by the blue absorbing type, with 63% and 62% respectively. These two particular samples were collected on dates when Chl-*a* levels were the lowest measured in this study, below 0.25 μgL^{-1}. Furthermore, the compilation of our data with values measured in the Eastern Mediterranean (Dubinsky et al., 2017) revealed a strong correlation between percent of blue-absorbing PR genes and the Chl-a concentrations only below 0.25 μgL^{-1} (Figure 7B). However, the concentration of Chl-a in the water is only one of the components that determine water color. To fully evaluate the role of the underwater light field in the spectral tuning of PR, we compared the proportion of green and blue PR gene variants to the corresponding satellite products of remote sensing reflectance (Rrs). The satellite remote sensing reflectance data shows predominantly blue reflectance spectra for all locations in July and October and green spectra in January and April (fig 8).

Discussion

Proteorhodopsin dominates abundance and expression of picoplankton light-harvesting mechanisms year-round

High abundance of rhodopsin genes was previously described in the Global Ocean Sampling data (Rusch et al., 2007) and in various other marine datasets (reviewed by Pinhassi et al., 2016; Brindefalk et al., 2016; Boeuf et al., 2016; Dubinsky et al., 2017, Maresca et al. 2018), and our observations further support this observation. However, most studies so far have focused on single time-points, and information on seasonal distribution of rhodopsin genes is lacking. Our experimental design allowed us to compare different contrasting locations and seasons, with the potential to identify patterns of phototrophy and resource availability. Previous studies in several marine environments (i.e. the North Atlantic and Arctic oceans as well as the Eastern Mediterranean Sea) found a correlation between *prd* gene abundance in genomes and Chl-a levels in seawater (Campbell et al. 2008; Boeuf et al., 2016; Dubinsky et al., 2017). This trend was further observed at the transcript level, showing higher *prd* gene transcripts in the Arctic (Boeuf et al., 2016). Unexpectedly, we found that the presence of the *prd* gene in genomes was

high year-round in this dynamic ecosystem, even at the eutrophic station of POLA (fig. 2A, sup. table S1), suggesting that the trophic state of the water is not always a good predictor of PR abundance.

Rhodopsin expression in the San Pedro Channel was consistently higher than oxygenic photosynthesis in the picoplankton over different seasons, with the exception of one sample taken during a localized algal bloom (fig. 2B). While a considerable amount of photosynthesis is performed by large photosynthetic eukaryotes, picoplankton can, in fact, represent the majority of the photosynthetic community at SPOT and CAT (Connell et al., 2017; Needham et al., 2017), and rhodopsin genes were shown to be more abundant in this size fraction as well (Finkel et al., 2013). Our results support the previously reported low gene abundances of aerobic anoxygenic photosynthesis (AAnP) (Boeuf et al., 2013, 2016; Dubinsky and Haber et al., 2017), but we further show that its expression is negligible compared to the other light-harvesting mechanisms (fig 2). Overall, we found no spatial trends in abundance or expression of rhodopsin genes, but oxygenic photosynthesis and AAnP were significantly lower at POLA (sup. fig. S3). This highlights the importance of examining not just abundance of genes but also their expression when comparing these ubiquitous phototrophic strategies.

PR expression is negatively correlated with chlorophyll-a concentrations

Despite the co-occurrence of rhodopsin *prd* genes and oxygenic photosynthesis *psaA* genes (fig. 2), *prd* expression appeared to be negatively correlated to Chl-a concentrations at SPOT and POLA (fig. 3), as previously reported in the Arctic Ocean (Boeuf et al., 2016) and the Eastern Mediterranean (Dubinsky et al., 2017). We speculate that the slope of this correlation was steeper at POLA compared to SPOT due to the higher abundance of large photosynthetic eukaryotes at POLA (Connell et al., 2017), which can lead to more available organic carbon due to leaky cells and sloppy feeding. Availability of organic carbon enables cells to acquire energy heterotrophically rather than harvesting light. Consistent with this hypothesis, in terms of physiology, PR phototrophy has been shown to be important particularly under DOM-limiting conditions typical of oligotrophic/low chlorophyll regimes (Steindler et al. 2011; Gómez-Consarnau et al. 2007, 2010, 2016). The negative correlation between *prd* expression and Chl-*a* concentrations was particularly clear at POLA in April 2013, when a localized diatom bloom was observed which did not extend to the other sites and led to the composition of the microbial eukaryotic community entirely diverge from all other samples (Hu et al., 2016; Connell et al., 2017). This was the only sample in which the expression of *psaA* genes for oxygenic photosynthesis exceeded that of *prd* genes. The high expression of *psaA* in this sample might be explained in part by the presence of chloroplasts released from diatoms that broke during the filtration and ended up being collected on the 0.2 µm filter. As the cumulative abundance of photosynthetic eukaryotes was significantly higher in this particular sample (Connell et al., 2017), and the relative abundance of diatom 18S DNA and RNA in it was higher than or equal to that of all other samples combined, this artifact is much more likely to have occurred in this sample. While the relative abundance of *prd* genes by clusters in this sample was similar to others, the expression pattern was unique (fig. 4AB, sup. fig. S1). This observation could be explained by the fact that our *prd* clusters are grouped at a rather coarse, mostly phylum level, while the changes in expression may have been influenced a finer-level shift of the microbial succession during the bloom (Needham et al., 2016; 2017).

SAR11 is the most highly expressed rhodopsin cluster

Owing to the differences between the present rhodopsin-bearing community and its active subset, *prd* gene abundance distribution by cluster was more even at the gene level (sup. fig. S1), as earlier observed in the Red Sea (Philosof and Beja, 2013). The vast majority of rhodopsins in our sites were proteorhodopsins from the SAR11 cluster. The SAR11-cluster of proton-pump type proteorhodopsins dominated *prd* transcripts (fig. 4B). Furthermore, the expression of this cluster correlated positively with the expression of SAR11 16S-rRNA operational taxonomic units (OTUs, 99%). This correlation was maintained at the gene level with the exception of POLA in April, potentially due to the interference introduced by the diatom bloom. These matching trends of presence and expression of the SAR11 proteorhodopsins aligns with the streamlined nature of SAR11 genomes and their reported constitutive expression of this protein (Giovannoni et al., 2005).
Other rhodopsins (e.g. actinorhodopsin, bacteriorhodopsin, xanthorhodopsin, halorhodopsin and xenorhodopsin) were rare (fig. 4AB). The abundance and expression of viral PRs was also very low, consistent with the fact that viral PRs were so far detected only in giant Phycodnaviruses of freshwater eukaryotes (Yutin and Koonin, 2012) and in low-salinity water (Brindefalk et al., 2016). While there is an inherent problem with normalizing both abundance and expression of viral proteorhodopsin, as there are no universal viral marker genes to normalize to, this limitation was unlikely to affect our results due to the low relative abundance of viral PR. However, this should be taken into account in studies where the viral cluster is more highly represented.

Proteorhodopsin spectral tuning has a key role in population dynamics

Some studies show that blue light absorbing PR variants dominate in open ocean, oligotrophic conditions, whereas the green variants are more abundant in shallow or coastal water (Man et al., 2003; Rusch et al., 2007). However, these observations are not consistent in the literature, as some studies show the different PR types to be decoupled (Sabehi et al., 2007). As conditions at SPOT and CAT are dynamic and fluctuate seasonally between oligotrophic and mesotrophic (Connell et al., 2017), we expected the spectral tuning of PRs to vary throughout the year. We observed a majority of blue absorbing PRs only in October of 2012, which coincided with the lowest Chl-*a* concentrations recorded in this study. Comparing our results with published data from the ultraoligotrophic Eastern Mediterranean Sea (Chl-a < 0.01 µg/L, Dubinsky et al., 2017), we observed a Chl-a concentration pivot point of about 0.25 µg/L. When Chl-a concentration was below this threshold, as in the Mediterranean and at SPOT and CAT in October, there was a highly significant negative correlation of chlorophyll concentration with *prd* expression, as well as high (60-75%) percent blue absorption (fig. 4C). However, above this threshold the percentage of cells with the blue variant dropped to an average of about 30% with no clear correlation to chlorophyll (fig. 3B), further supported by even higher relative abundance of the green variant (>90%) in the Baltic Sea (Brindefalk et al., 2016). However, Chl-a concentration is only a proxy for the quality and quantity of light in the water column and the dissolved organic material (DOM) available. Under bloom conditions, Chl-a concentrations increase, turning the water greener and reducing its transparency. The consequent increase in available DOM that follows blooms will further attenuate light in the water column, particularly in the blue and UV region of the spectrum. The resulting combination of reflectance and absorption by algal pigments, dissolved organic matter, water and inorganic particles is what determines the

available light in the water column. When DOM is low, light harvesting and spectrally tuning PR may play a crucial role in survival or fitness of photoheterotrophic bacteria populations. This is evident in the most oligotrophic locations such as the Eastern Mediterranean (Dubinsky et al. 2016). Consistent with this, in year-round eutrophic locations such as the Baltic Sea, the dominant variant is green whereas in dynamic locations such as SPOT and POLA, a mix of the two variants can be expected. Since we can readily identify these patterns at the gene level, our data indicates that the light regime is a key factor selecting PR-containing populations, as suggested in the past (Sabehi et al., 2007). Future studies will be needed to better define and increase the resolution of these thresholds to better understand the role of light wavelength availability in population dynamics, survival and competition.

Conclusion

Our spatial time-series analysis of proteorhodopsin and oxygenic and anoxygenic photosynthesis in marine picoplankton revealed that (1) expression of rhodopsin-based photoheterotrophy exceeded that of oxygenic photoautotrophy, (2) between rhodopsin clusters expression did not necessarily follow the patterns of gene abundance, and (3) aerobic anoxygenic photosynthesis is not an important process in this system despite detectable gene abundances. It is highly important to continue collecting more deeply-sequenced metatranscriptomic data in order to begin to elucidate the local adaptations of photoheterotrophs in the ocean that lead to their global success. Finally, our results reinforce the conclusion that the differences in the light spectrum are an important selective force, defining the abundance of different PR photoheterotrophic types.

Acknowledgements

The authors would like to thank Dominique Boeuf for kindly providing supporting files for phylogenetic trees analysis. We would like to thank Erin Beers-Fichot for her help in the sampling cruises and metatranscriptome generation, as well as Catherine Roney-Garcia, Alma Parada, Jacob Cram, David Needham and the Sundiver crew for their assistance in cruises and sample processing. The authors acknowledge the USC Wrigley Institute of Environmental Studies. This work was funded by NSF grants OCE1335269 and 1136818 as well as by the Gordon and Betty Moore Foundation Marine Microbiology Initiative grant GBMF3779.

References

1. Béja, O., Aravind, L., Koonin, E. V., Suzuki, M. T., Hadd, A., Nguyen, L. P., ... & Spudich, E. N. (2000). Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science*, *289*(5486), 1902-1906.
2. Béjà, O., Spudich, E. N., Spudich, J. L., Leclerc, M., & DeLong, E. F. (2001). Proteorhodopsin phototrophy in the ocean. *Nature*, *411*(6839), 786-789.
3. Boeuf, D., Cottrell, M. T., Kirchman, D. L., Lebaron, P., & Jeanthon, C. (2013). Summer community structure of aerobic anoxygenic phototrophic bacteria in the western Arctic Ocean. *FEMS microbiology ecology*, *85*(3), 417-432.
4. Boeuf, D., Audic, S., Brillet-Guéguen, L., Caron, C., & Jeanthon, C. (2015). MicRhoDE: a curated database for the analysis of microbial rhodopsin diversity and evolution. *Database*, *2015*, bav080.

5. Boeuf, D., Lami, R., Cunnington, E., & Jeanthon, C. (2016). Summer Abundance and Distribution of Proteorhodopsin Genes in the Western Arctic Ocean. *Frontiers in microbiology*, *7*.
6. Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114-2120.
7. Brindefalk, B., Ekman, M., Ininbergs, K., Dupont, C. L., Yooseph, S., Pinhassi, J., & Bergman, B. (2016). Distribution and expression of microbial rhodopsins in the Baltic Sea and adjacent waters. *Environmental microbiology*, *18*(12), 4442-4455.
8. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC bioinformatics*, *10*(1), 421.
9. Campbell, B. J., Waidner, L. A., Cottrell, M. T., & Kirchman, D. L. (2008). Abundant proteorhodopsin genes in the North Atlantic Ocean. *Environmental microbiology*, *10*(1), 99-109.
10. Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular biology and evolution*, *17*(4), 540-552.
11. Connell, P. E., Campbell, V., Gellene, A. G., Hu, S. K., & Caron, D. A. (2017). Planktonic food web structure at a coastal time-series site: II. Spatiotemporal variability of microbial trophic activities. *Deep Sea Research Part I: Oceanographic Research Papers*, *121*, 210-223.
12. Dubinsky, V., Haber, M., Burgsdorf, I., Saurav, K., Lehahn, Y., Malik, A., ... & Steindler, L. (2017). Metagenomic analysis reveals unusually high incidence of proteorhodopsin genes in the ultraoligotrophic Eastern Mediterranean Sea. *Environmental microbiology*, *19*(3), 1077-1090.
13. Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, *26*(19), 2460-2461.
14. Eren, A. M., Esen, Ö. C., Quince, C., Vineis, J. H., Morrison, H. G., Sogin, M. L., & Delmont, T. O. (2015). Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ*, *3*, e1319.
15. Feng, S., Powell, S. M., Wilson, R., & Bowman, J. P. (2013). Light-stimulated growth of proteorhodopsin-bearing sea-ice psychrophile Psychroflexus torquis is salinity dependent. *The ISME Journal*, *7*(11), 2206.
16. Finkel, O. M., Béjà, O., & Belkin, S. (2013). Global abundance of microbial rhodopsins. *The ISME journal*, *7*(2), 448-451.
17. Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., ... & Salazar, G. A. (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic acids research*, *44*(D1), D279-D285.
18. Frias-Lopez, J., Shi, Y., Tyson, G. W., Coleman, M. L., Schuster, S. C., Chisholm, S. W., & DeLong, E. F. (2008). Microbial community gene expression in ocean surface waters. *Proceedings of the National Academy of Sciences*, *105*(10), 3805-3810.
19. Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, *28*(23), 3150-3152.
20. Fuhrman, J. A., Schwalbach, M. S., & Stingl, U. (2008). Proteorhodopsins: an array of physiological roles?. *Nature Reviews Microbiology*, *6*(6), 488-494.

21. Giovannoni, S. J., Tripp, H. J., Givan, S., Podar, M., Vergin, K. L., Baptista, D., ... & Rappé, M. S. (2005). Genome streamlining in a cosmopolitan oceanic bacterium. *science*, *309*(5738), 1242-1245.
22. Gómez-Consarnau, L., González, J. M., Coll-Lladó, M., Gourdon, P., Pascher, T., Neutze, R., ... & Pinhassi, J. (2007). Light stimulates growth of proteorhodopsin-containing marine Flavobacteria. *Nature*, *445*(7124), 210-213.
23. Gómez-Consarnau, L., Akram, N., Lindell, K., Pedersen, A., Neutze, R., Milton, D. L., ... & Pinhassi, J. (2010). Proteorhodopsin phototrophy promotes survival of marine bacteria during starvation. *PLoS biology*, *8*(4), e1000358.
24. Gómez-Consarnau, L., González, J. M., Riedel, T., Jaenicke, S., Wagner-Döbler, I., Sañudo-Wilhelmy, S. A., & Fuhrman, J. A. (2016). Proteorhodopsin light-enhanced growth linked to vitamin-B1 acquisition in marine Flavobacteria. *The ISME journal*, *10*(5), 1102-1112.
25. Gómez-Pereira, P. R., Hartmann, M., Grob, C., Tarran, G. A., Martin, A. P., Fuchs, B. M., ... & Zubkov, M. V. (2013). Comparable light stimulation of organic nutrient uptake by SAR11 and Prochlorococcus in the North Atlantic subtropical gyre. *The ISME journal*, *7*(3), 603-614.
26. Hu, S. K., Campbell, V., Connell, P., Gellene, A. G., Liu, Z., Terrado, R., & Caron, D. A. (2016). Protistan diversity and activity inferred from RNA and DNA at a coastal ocean site in the eastern North Pacific. *FEMS microbiology ecology*, *92*(4), fiw050.
27. Hyatt, D., Chen, G. L., LoCascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC bioinformatics*, *11*(1), 119.
28. Ignacio-Espinoza, J. C., & Sullivan, M. B. (2012). Phylogenomics of T4 cyanophages: lateral gene transfer in the 'core'and origins of host genes. *Environmental microbiology*, *14*(8), 2113-2126.
29. Johnson, L. S., Eddy, S. R., & Portugaly, E. (2010). Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC bioinformatics*, *11*(1), 431.
30. Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*, *30*(4), 772-780.
31. Kiefer, D. A., Chamberlin, W. S., & Booth, C. R. (1989). Natural fluorescence of chlorophyll a: Relationship to photosynthesis and chlorophyll concentration in the western South Pacific gyre. *Limnology and Oceanography*, *34*(5), 868-881.
32. Kopf, A., Kostadinov, I., Wichels, A., Quast, C., & Glöckner, F. O. (2015). Metatranscriptome of marine bacterioplankton during winter time in the North Sea assessed by total RNA sequencing. *Marine genomics*, *19*, 45-46.
33. Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K., & Schloss, P. D. (2013). Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Applied and environmental microbiology*, *79*(17), 5112-5120.
34. Li, D., Liu, C-M., Luo, R., Sadakane, K., and Lam, T-W., (2015) MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*, 31(10):1674-1676

35. Man, D., Wang, W., Sabehi, G., Aravind, L., Post, A. F., Massana, R., ... & Béjà, O. (2003). Diversification and spectral tuning in marine proteorhodopsins. *The EMBO journal*, *22*(8), 1725-1731.
36. Marchetti, A., Schruth, D. M., Durkin, C. A., Parker, M. S., Kodner, R. B., Berthiaume, C. T., ... & Armbrust, E. V. (2012). Comparative metatranscriptomics identifies molecular bases for the physiological responses of phytoplankton to varying iron availability. *Proceedings of the National Academy of Sciences*, *109*(6), E317-E325.
37. Marchetti, A., Catlett, D., Hopkinson, B. M., Ellis, K., & Cassar, N. (2015). Marine diatom proteorhodopsins and their potential role in coping with low iron availability. *The ISME journal*, *9*(12), 2745-2748.
38. Maresca, J. A., Miller, K. J., Keffer, J. L., Sabanayagam, C. R., & Campbell, B. J. (2018). Distribution and diversity of rhodopsin-producing microbes in the Chesapeake Bay. *Applied and environmental microbiology*, AEM-00137.
39. Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., ... & Dewell, S. B. (2005). Genome sequencing in open microfabricated high density picoliter reactors. *Nature*, *437*(7057), 376.
40. Matsen, F. A., Kodner, R. B., & Armbrust, E. V. (2010). pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC bioinformatics*, *11*(1), 538.
41. Needham, D. M., & Fuhrman, J. A. (2016). Pronounced daily succession of phytoplankton, archaea and bacteria following a spring bloom. *Nature microbiology*, *1*(4), 16005.
42. Needham, D. M., Sachdeva, R., & Fuhrman, J. A. (2017). Ecological dynamics and co-occurrence among marine phytoplankton, bacteria and myoviruses shows microdiversity matters. *The ISME Journal*.
43. Nguyen, D., Maranger, R., Balagué, V., Coll-Llado, M., Lovejoy, C., & Pedros-Alio, C. (2015). Winter diversity and expression of proteorhodopsin genes in a polar ocean. *The ISME journal*, *9*(8), 1835-1845.
44. Noble, R. T., & Fuhrman, J. A. (1998). Use of SYBR Green I for rapid epifluorescence counts of marine viruses and bacteria. *Aquatic Microbial Ecology*, *14*(2), 113-118.
45. Ottesen, E. A., Young, C. R., Gifford, S. M., Eppley, J. M., Marin, R., Schuster, S. C., ... & DeLong, E. F. (2014). Multispecies diel transcriptional oscillations in open ocean heterotrophic bacterial assemblages. *Science*, *345*(6193), 207-212.
46. Parada, A. E., Needham, D. M., & Fuhrman, J. A. (2016). Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environmental microbiology*, *18*(5), 1403-1414.
47. Patel, A., Noble, R. T., Steele, J. A., Schwalbach, M. S., Hewson, I., & Fuhrman, J. A. (2007). Virus and prokaryote enumeration from planktonic aquatic environments by epifluorescence microscopy with SYBR Green I. *Nature protocols*, *2*(2), 269-276.
48. Philosof, A., & Béjà, O. (2013). Bacterial, archaeal and viral-like rhodopsins from the Red Sea. *Environmental microbiology reports*, *5*(3), 475-482.
49. Pinhassi, J., DeLong, E. F., Béjà, O., González, J. M., & Pedrós-Alió, C. (2016). Marine bacterial and archaeal ion-pumping rhodopsins: genetic diversity, physiology, and ecology. *Microbiology and Molecular Biology Reviews*, *80*(4), 929-954.

50. Poretsky, R. S., Hewson, I., Sun, S., Allen, A. E., Zehr, J. P., & Moran, M. A. (2009). Comparative day/night metatranscriptomic analysis of microbial communities in the North Pacific subtropical gyre. *Environmental microbiology*, *11*(6), 1358-1375.
51. R package RAM, https://cran.r-project.org/web/packages/RAM/index.html
52. R package psych, https://cran.r-project.org/web/packages/psych/psych.pdf
53. Rocha, D. J., Santos, C. S., & Pacheco, L. G. (2015). Bacterial reference genes for gene expression studies by RT-qPCR: survey and analysis. *Antonie Van Leeuwenhoek*, *108*(3), 685-693.
54. Rusch, D. B., Halpern, A. L., Sutton, G., Heidelberg, K. B., Williamson, S., Yooseph, S., ... & Beeson, K. (2007). The Sorcerer II global ocean sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS biology*, *5*(3), e77.
55. Sabehi, G., Béjà, O., Suzuki, M. T., Preston, C. M., & DeLong, E. F. (2004). Different SAR86 subgroups harbour divergent proteorhodopsins. *Environmental Microbiology*, *6*(9), 903-910.
56. Sabehi, G., Kirkup, B. C., Rozenberg, M., Stambler, N., Polz, M. F., & Béja, O. (2007). Adaptation and spectral tuning in divergent marine proteorhodopsins from the eastern Mediterranean and the Sargasso Seas. *The ISME journal*, *1*(1), 48-55.
57. Satinsky, B. M., Crump, B. C., Smith, C. B., Sharma, S., Zielinski, B. L., Doherty, M., ... & Coles, V. J. (2014). Microspatial gene expression patterns in the Amazon River Plume. *Proceedings of the National Academy of Sciences*, *111*(30), 11085-11090.
58. Shi, Y., Tyson, G. W., Eppley, J. M., & DeLong, E. F. (2011). Integrated metatranscriptomic and metagenomic analyses of stratified microbial assemblages in the open ocean. *The ISME journal*, *5*(6), 999-1013.
59. Sommer, D. D., Delcher, A. L., Salzberg, S. L., & Pop, M. (2007). Minimus: a fast, lightweight genome assembler. *BMC bioinformatics*, *8*(1), 64.
60. Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312-1313.
61. Steindler, L., Schwalbach, M. S., Smith, D. P., Chan, F., & Giovannoni, S. J. (2011). Energy starved Candidatus Pelagibacter ubique substitutes light-mediated ATP production for endogenous carbon respiration. *PLoS One*, *6*(5), e19725.
62. Vader, A., Laughinghouse IV, H. D., Griffiths, C., Jakobsen, K. S., & Gabrielsen, T. M. (2018). Proton-pumping rhodopsins are abundantly expressed by microbial eukaryotes in a high-Arctic fjord. *Environmental microbiology*, *20*(2), 890-902.
63. Walter, J. M., Greenfield, D., Bustamante, C., & Liphardt, J. (2007). Light-powering Escherichia coli with proteorhodopsin. *Proceedings of the National Academy of Sciences*, *104*(7), 2408-2412.
64. Yilmaz, P., Parfrey, L. W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., ... & Glöckner, F. O. (2013). The SILVA and "all-species living tree project (LTP)" taxonomic frameworks. *Nucleic acids research*, *42*(D1), D643-D648.
65. Yutin, N., & Koonin, E. V. (2012). Proteorhodopsin genes in giant viruses. *Biology direct*, *7*(1), 34.
66. Zhang, J., Kobert, K., Flouri, T., & Stamatakis, A. (2013). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics*, *30*(5), 614-620.

Figure 1

Map of the sampling sites

Port of Los Angeles (POLA, 33°42.75′N, 118°15.55′W), San Pedro Ocean Time-series (SPOT, 33°33′N, 118°24′W) and Catalina Island (CAT, 33°27.17 N, 118°28.51′W). Adapted with author permission from Connell et al., 2017.

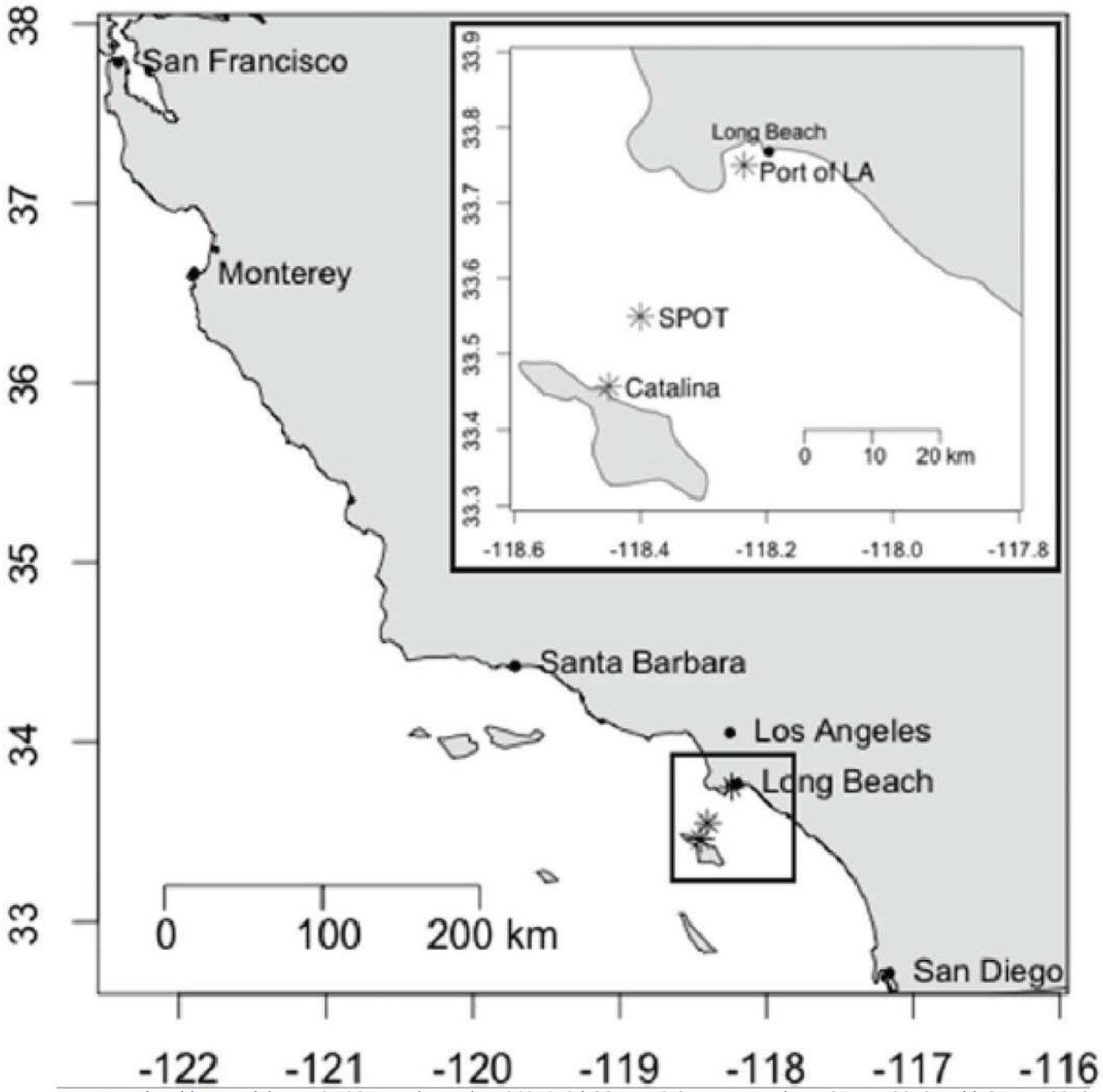

Figure 2

Normalized abundance and expression of phototrophic mechanisms

(**A**) Relative gene abundance of photosystem-I (PSI, *psaA*, green), rhodopsin (PR, *prd*, dotted blue) and aerobic anoxygenic photosynthesis (AAnP, *pufM*, grey) normalized to *recA* in metagenomes and (**B**) the ratio between relative abundance in metatranscriptomes to relative abundance in metagenomes (RNA to DNA ratio) per gene. Note that the Y axis in B is logarithmic.

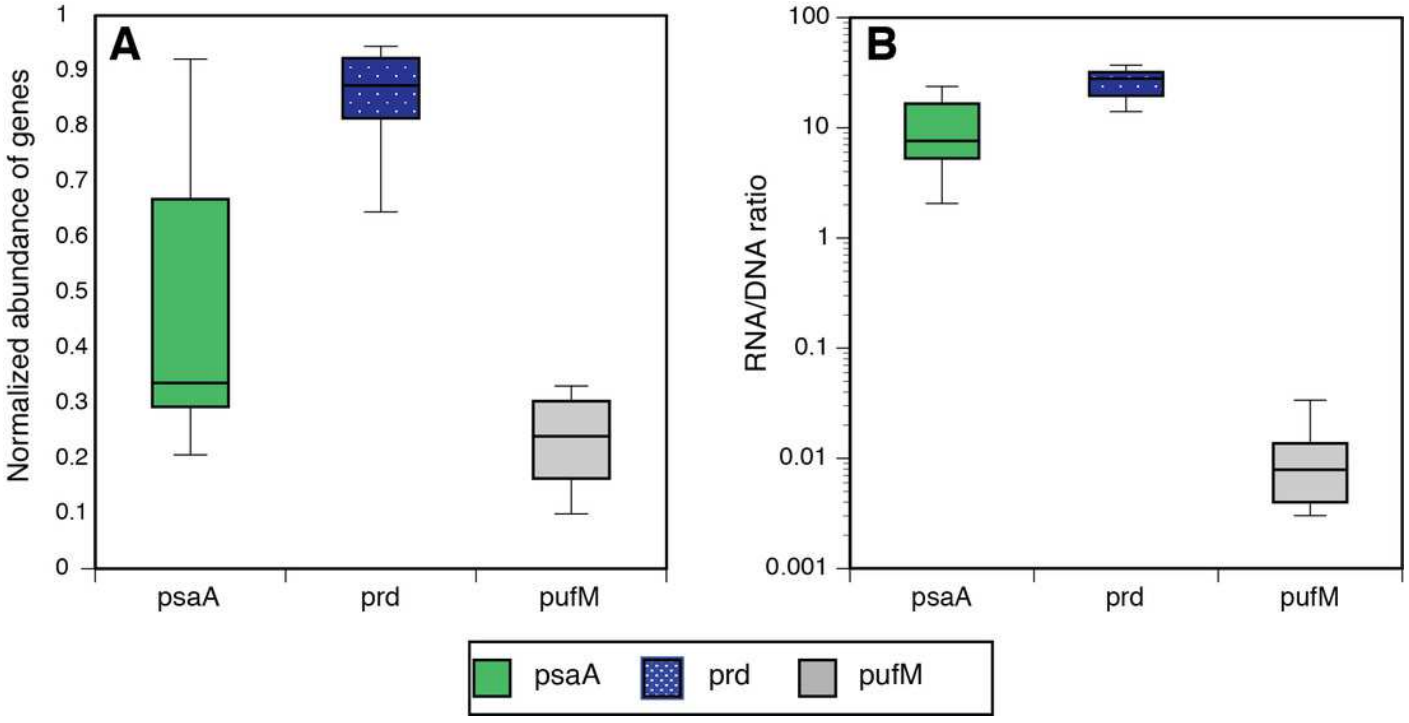

Figure 3

Negative correlation between proteorhodopsin expression and Chlorophyll-a at POLA and SPOT

Prd expression was normalized to RecA expression. Trendline equations and R^2 values are indicated on the plot. No correlation was found in CAT.

**Note: Auto Gamma Correction was used for the image. This only affects the reviewing manuscript. See original source image if needed for review.*

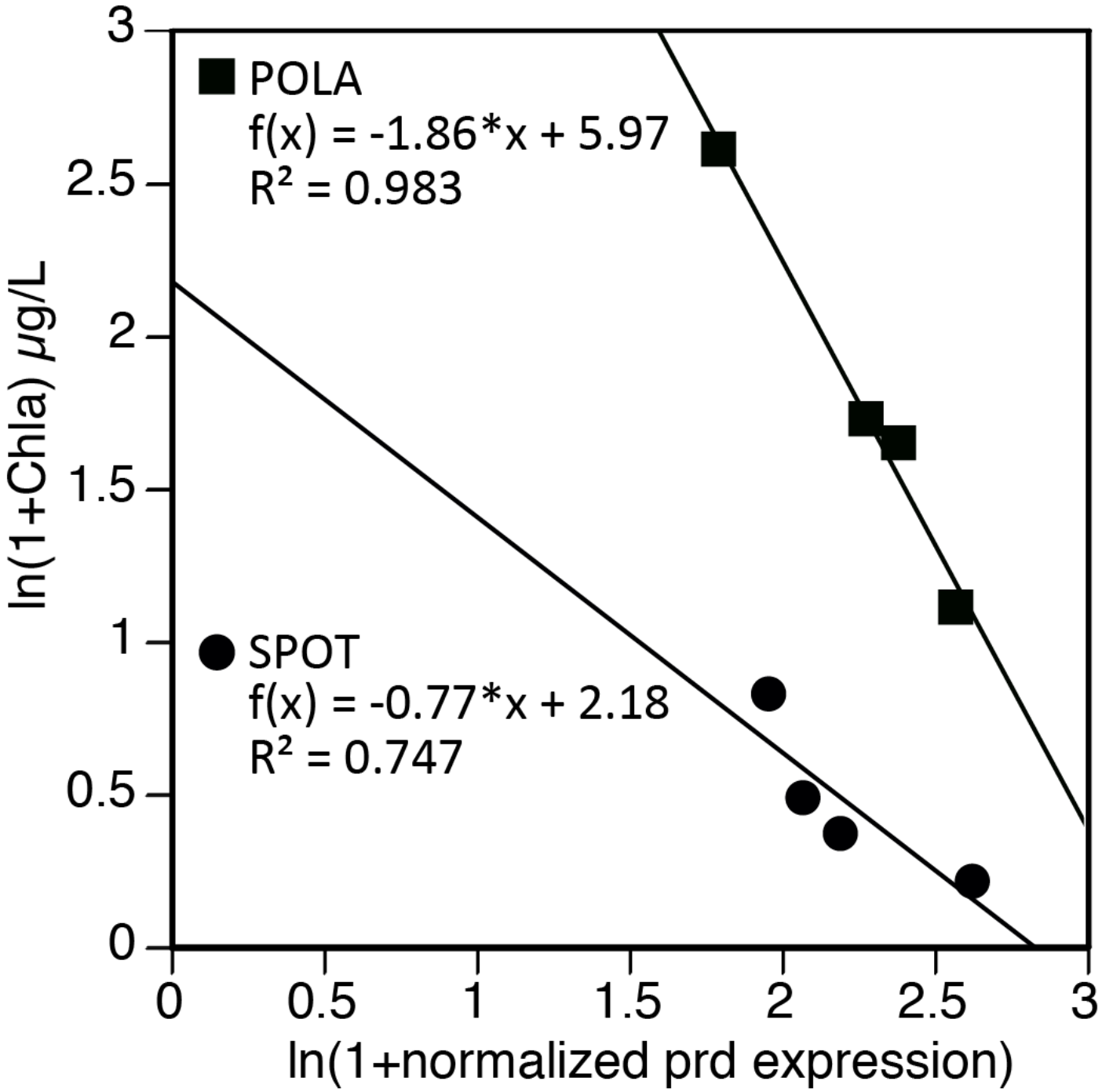

Figure 4

Relative abundance of rhodopsin clades by site and date

Relative abundance was calculated out of all rhodopsin-assigned reads per sample in (**A**) metagenomes (average *prd* reads per sample 6485, standard deviation 3927) and (**B**) metatranscriptomes (average *prd*reads per sample 43005, standard deviation 22367). The evenness of rhodopsin gene abundance was consistently high (see also sup. fig. S1), whereas expression was dominated by SAR11 (dotted blue), other Gammaproteobacteria (dotted red), Alphaproteobacteria (grey) and unknown environmental clade (black).

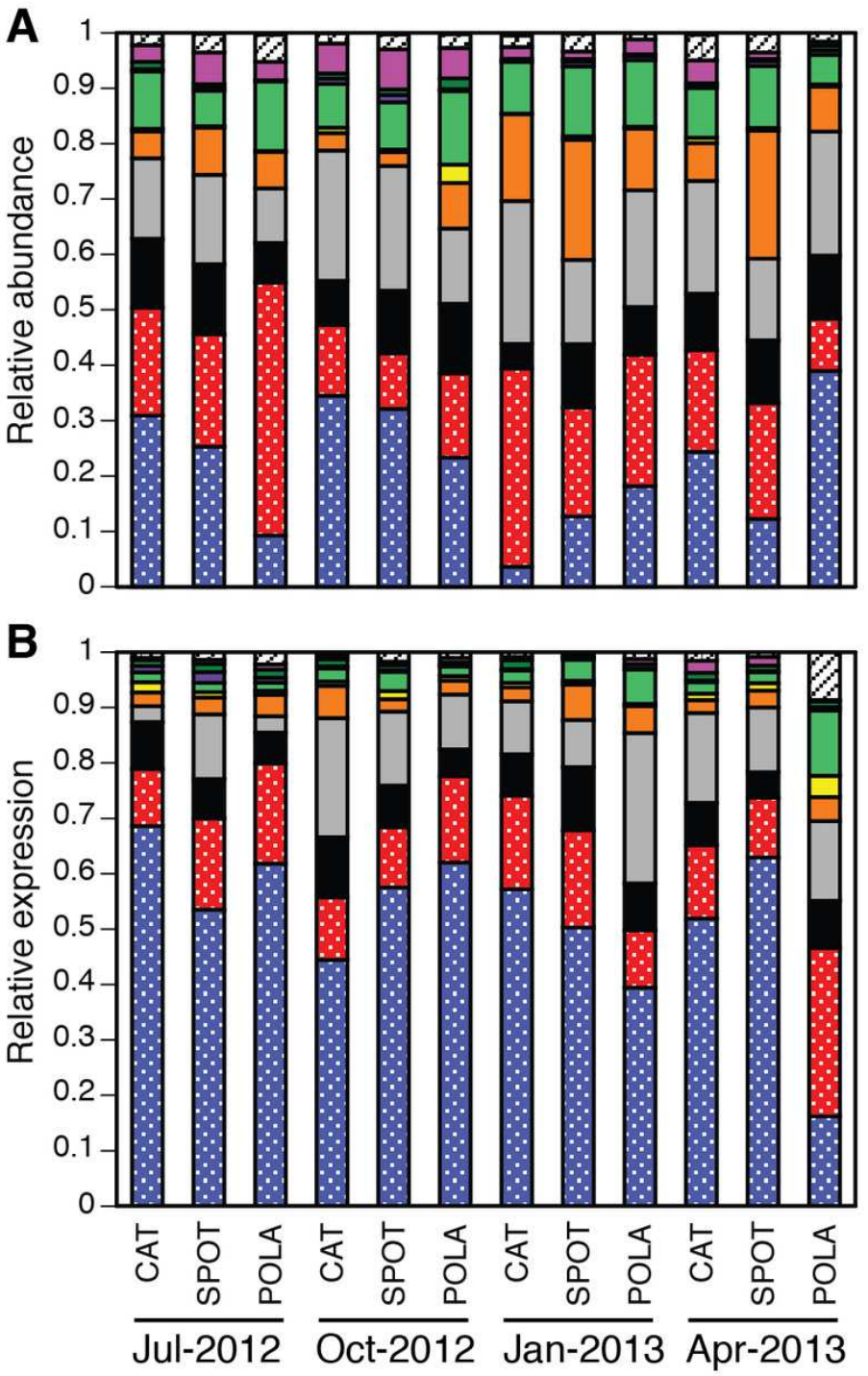

Figure 5

SAR11 correlation between 16S-rRNA and *prd*

(**A**) relative gene abundance and (**B**) relative transcript abundance

**Note: Auto Gamma Correction was used for the image. This only affects the reviewing manuscript. See original source image if needed for review.*

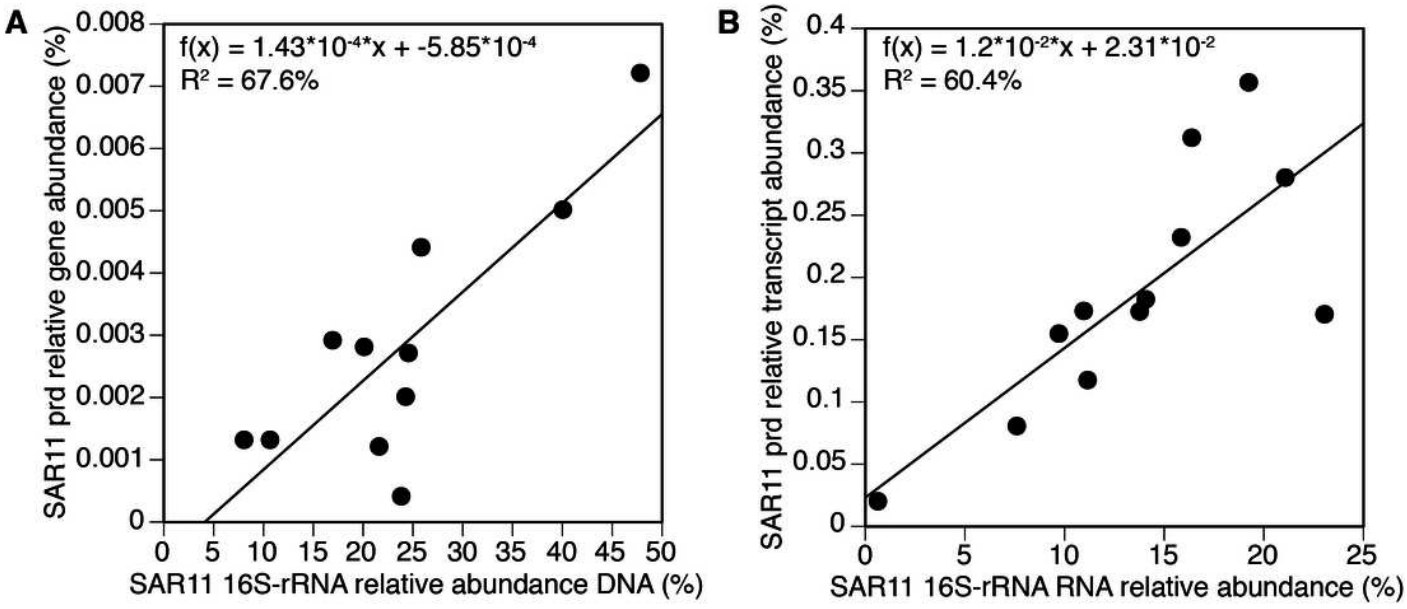

Figure 6

High resolution breakdown of the SAR11 cluster of proteorhodopsin

SAR11 cluster of proteorhodopsins at the single protein level (per leaf in the MicRhoDE phylogenetic tree) had higher evenness in gene abundance than in expression (Wilcoxon rank sum test, paired one-sided, p=0.0005). (**A**) SAR11-cluster *prd* total gene abundance over time and sites, (**B**) SAR11-cluster *prd* total expression over time and sites, (**C**) relative abundance of the top 10 most abundant SAR11-cluster *prd* genes (all other SAR11-cluster *prd* were added up and represented as "other", total 12,852 reads) and (D) relative abundance of the top 10 most highly expressed SAR11-cluster *prd* (total 185,019 reads). Note that C and D have separate legends and that only 4 *prd* leaves are shared between them. Leaves are denoted by accession numbers.

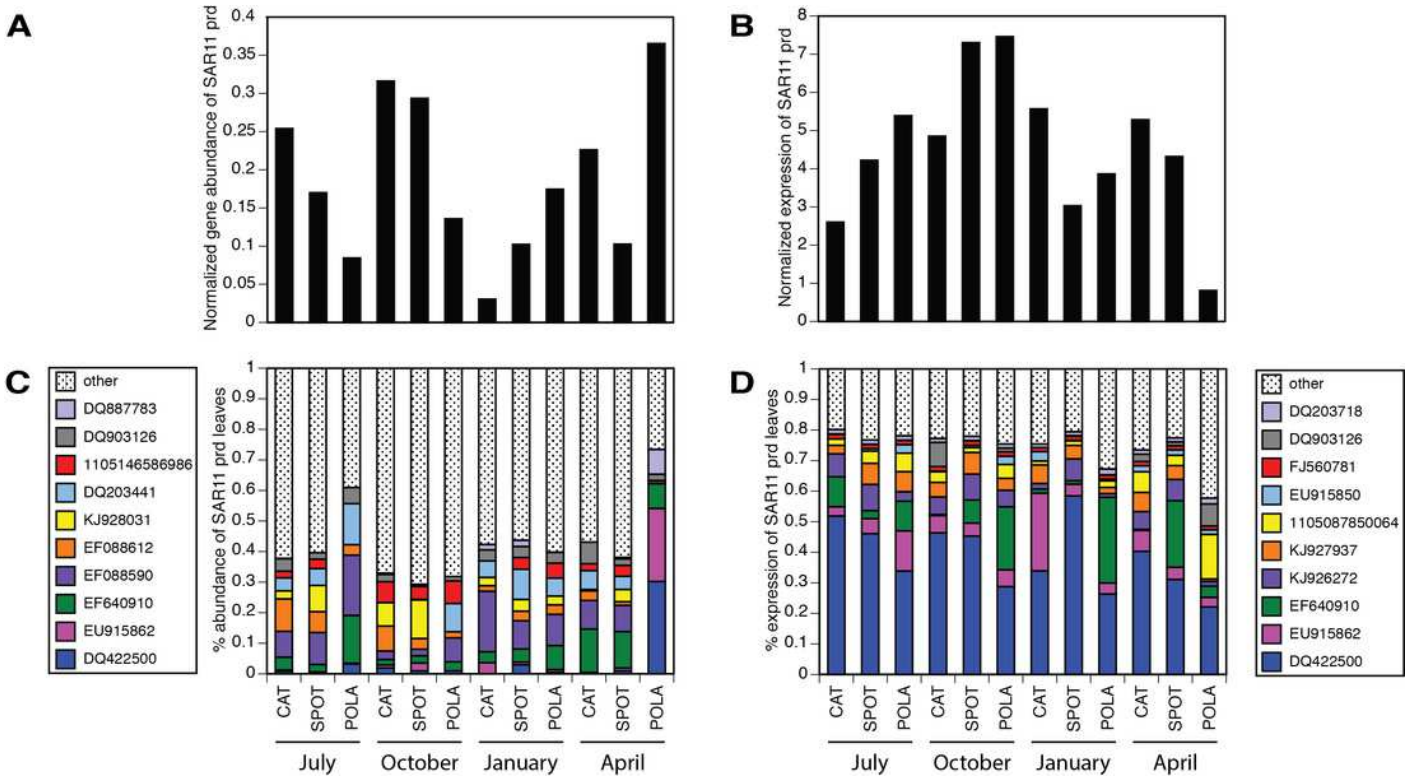

Figure 7

Spectral tuning of rhodopsins in our dynamic system and a comparison to previous studies.

(**A**) Relative abundance of blue/green variants of proteorhodopsin in metagenomes (294-2112 reads per sample): leucine (L) is represented by light green, methionine (M) by dark green and glutamine (Q) by blue. There was not enough data to plot tuning distribution at POLA in April 2013. (**B**) 0.25 µg/L is the Chl-a threshold between environments dominated by blue-tuned proteorhodopsin or green-tuned based on data from this paper and previous publications.

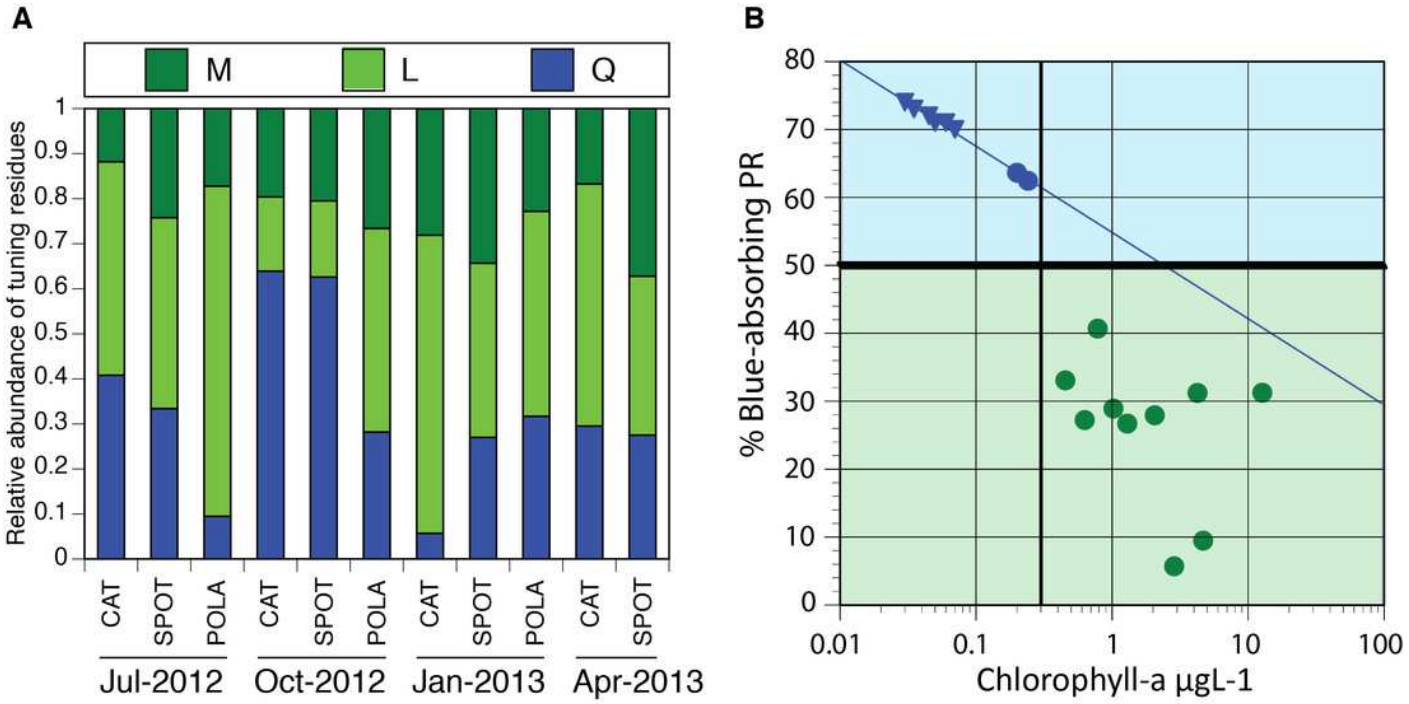

Figure 8

Average Remote sensing reflectance spectra (Rrs) per location and season

(**A**) July 15 and 16, 2012, (**B**) October 19 and 20, 2012, (**C**) January 9 and 10, 2012, (**D**) April 24 and 25, 2013. Measurements were averaged over 2 days including the sampling day and the day before or after, depending on satellite data availability.