FEMS Microbiology Ecology, 94, 2018, fiy182

doi: 10.1093/femsec/fiy182
Advance Access Publication Date: 5 September 2018
Research Article

RESEARCH ARTICLE

The genome of a novel isolate of *Prochlorococcus* from the Red Sea contains transcribed genes for compatible solute biosynthesis

Ahmed A. Shibl[1,3], David K. Ngugi[1,2], Agathe Talarmin[1,4], Luke R. Thompson[1,5,6], Jochen Blom[7] and Ulrich Stingl[1,8,*]

[1]Marine Microbial Ecology Lab, Red Sea Research Center, King Abdullah University of Science and Technology (KAUST), Thuwal 23955–6900, Kingdom of Saudi Arabia, [2]Department of Microorganisms, Leibniz Insitute DSMZ, Inhoffenstrasse 7B, 38124 Braunschweig, Germany., [3]Biology Department, New York University Abu Dhabi, P.O. Box 129188, Abu Dhabi, UAE, [4]Petroleum Geosciences Department, Petroleum Institute, P.O. Box 2533 Abu Dhabi, UAE, [5]Department of Biological Sciences and Northern Gulf Institute, University of Southern Mississippi, Hattiesburg, Mississippi, USA, [6]Ocean Chemistry and Ecosystems Division, Atlantic Oceanographic and Meteorological Laboratory, National Oceanic and Atmospheric Administration, stationed at Southwest Fisheries Science Center, La Jolla, California, USA, [7]Bioinformatics and Systems Biology, Justus Liebig University, D-35392 Giessen, Germany and [8]University of Florida, UF/IFAS, Department for Microbiology & Cell Science, Fort Lauderdale Research and Education Center, Davie, FL 33314, USA

*Corresponding author: Ulrich Stingl, 3205 College Ave., University of Florida, UF/IFAS, Department for Microbiology & Cell Science, Fort Lauderdale Research and Education Center, Davie, FL 33314, USA; Tel: +1-954577-6326; E-mail: ustingl@ufl.edu

One sentence summary: A novel isolate of *Prochlorococcus* from the saline Red Sea contains transcribed genes for the biosynthesis of an osmolyte that had not been reported before in this genus.

Editor: Gary King
†Jochen Blom, http://orcid.org/0000-0001-6455-3622

ABSTRACT

Marine microbes possess genomic and physiological adaptations to cope with varying environmental conditions. So far, the effects of high salinity on the most abundant marine photoautotrophic organism, *Prochlorococcus*, in marine oligotrophic environments, are mostly unknown. Here, we report the isolation of a new *Prochlorococcus* strain (RSP50) belonging to high-light (HL) clade II from the Red Sea, one of the warmest and most saline bodies of water in the global oceans. A comparative genomic analysis identified a set of 59 genes that were exclusive to RSP50 relative to currently available *Prochlorococcus* genomes, the majority of which (70%) encode for hypothetical proteins of unknown function. However, three of the unique genes encode for a complete pathway for the biosynthesis of the compatible solute glucosylglycerol, and are homologous to enzymes found in the sister lineage *Synechococcus*. Metatranscriptomic analyses of this metabolic pathway in the water column of the Red Sea revealed that the corresponding genes were constitutively transcribed, independent of depth and light, suggesting that osmoregulation using glucosylglycerol is a general feature of HL II *Prochlorococcus* in the Red Sea.

Received: 17 April 2018; **Accepted:** 4 September 2018

Keywords: *Prochlorococcus*; Red Sea; genomic adaptation; compatible solute; glucosyglycerol; metatranscriptomics

INTRODUCTION

The long and narrow main basin of the Red Sea is highly saline (36–41 psu; Edwards 1987; Alraddadi 2013) and warm (from 24 to 32°C; Raitsos *et al.* 2013) all year long. Minimal precipitation in the region and continuous exposure to solar irradiation renders the Red Sea physicochemically distinct from other marine environments (Rasul and Stewart 2015). The input of cooler water masses from the Indian Ocean generates opposing gradients of temperature and salinity from the South to the North (Ngugi and Stingl 2011). The southern region of the Red Sea contains the highest levels of productivity (as indicated by high concentrations of chlorophyll *a*), likely linked to the influx of nutrient-rich waters through the strait of Bab Al Mandab (Raitsos *et al.* 2013; Churchill *et al.* 2014). These latitudinal environmental gradients across the basin appear to affect different marine life. For instance, a break in genetic diversity of reef fish communities and a disruption in larval connectivity patterns were triggered by a shift in physicochemical conditions along the northern and central Red Sea (Nanninga *et al.* 2014). Similarly, the distribution of pelagic prokaryotic communities shows substantial differences along the southward temperature-salinity gradient, as reflected by spatial shifts in the abundances of the dominant phylotypes (Ngugi and Stingl 2011; Ngugi *et al.* 2012).

Prochlorococcus (Chisholm *et al.* 1988) is one of the most abundant pelagic bacterioplankton in the Red Sea (Ngugi and Stingl 2011). The *Prochlorococcus* ecotypes are divided among high-light (HL) and low-light (LL) adapted clades, based on their cellular ratio of divinyl chlorophyll *a* to divinyl chlorophyll *b* pigments (Moore and Chisholm 1999), temperature adaptation (Zinser *et al.* 2007), spatial distribution patterns (Johnson *et al.* 2006), pangenomics (Delmont and Eren 2018) and phylogeny (Rocap *et al.* 2002). The HL-adapted ecotype (HL II) dominates the pelagic layer (0–50 m) and the microdiversity of *Prochlorococcus* at 100 m is higher relative to other depths across the main basin of the Red Sea, according to studies using the internal transcribed spacer (ITS) region (Shibl *et al.* 2014) and, more extensively, using *rpoC1* gene analyses (Shibl *et al.* 2016). The biogeographical distribution of *Prochlorococcus* ecotypes throughout the water column seemed homogenous (Shibl *et al.* 2014, 2016). According to these studies based on *rpoc1* and ITS, the *Prochlorococcus* community structure also appears to be fairly stable across the strong environmental gradients.

At global scales, the phylogenetic and genomic diversity of *Prochlorococcus* ecotypes strongly correlate with the prevailing environmental parameters. Indeed, several gene families identified in *Prochlorococcus* populations significantly correspond to specific physicochemical variables such as nitrate concentrations and temperature (Kent *et al.* 2016). In comparison with other oligotrophic marine environments, Thompson *et al.* (2013b) reported the enrichment of gene families for phosphorus acquisition, DNA repair, and light stress in *Prochlorococcus* metagenomic sequences from the surface waters of the Red Sea, reflecting adaptive strategies specific to their local environment. Notably, Thompson *et al.* (2013b) also found enrichment of gene families for compatible solute (osmolyte) utilization in *Pelagibacter* (SAR11) metagenomic sequences, suggesting that the production of compatible solutes may be relatively higher in the Red Sea than in less saline environments.

Here, we report on the cultivation and genomic inventory of the first HL-adapted *Prochlorococcus* strain (RSP50) isolated from surface waters of the main basin of the Red Sea. Given the unusual environmental parameters of the main basin of the Red Sea, we hypothesized the presence of specific adaptive strategies potentially implemented by unique genes found in the RSP50 genome. Among the annotated RSP50-specific gene pool, we focused our analyses on three genes related to the synthesis of compatible solutes that are expected to contribute to *Prochlorococcus*'s adaptation to high salinity, and analyzed their abundance in metatranscriptomic and metagenomics datasets from the Red Sea.

MATERIALS AND METHODS

Sample collection for isolation of *Prochlorococcus*

Seawater was collected in December 2014 from the Red Sea at the Atlantis II site (22°10'48" N, 38°29'24" E) during an expedition on board the R/V *Thuwal*. A rosette sampler was used for water collection along with a CTD unit with sensors that additionally record salinity and fluorescence readings (Sea-Bird Electronics, USA). Four dark acid-washed polyethylene bottles (150 mL each) were rinsed and filled with seawater collected at 10 m water depth, and stored at room temperature on board for less than 4 h.

Culturing conditions

On the same day in the lab, 500 mL of the collected seawater was filtered through 0.6-μm polycarbonate membrane filters (MilliporeSigma, Germany). The filtrate was used for initial inoculation (20 mL) into 5 mL PRO2 medium (Moore and Chisholm 1999; Moore *et al.* 2007) and incubated at 27°C under continuous cool white fluorescent light at 40 μE m^{-2} s^{-1} in a Multitron Standard incubator (INFORS HT, Switzerland). PRO2 was used for the initial isolation phase (one week) because it contains extremely low macronutrient concentrations. Light intensity was measured using the Quantum Scalar Laboratory sensor (QSL-2100; Biospherical Instruments Inc., CA, USA). Growth was monitored daily by measuring chlorophyll *a* fluorescence using a Turner-10 AU fluorometer (Turner Designs, CA, USA). After one week of growth, the enrichments were transferred to 96-well Teflon plates (2-mL wells) using a dilution-to-extinction approach (Stingl, Tripp and Giovannoni 2007) in ProMM medium and kept under the same light and temperature conditions. ProMM contains organic compounds and vitamins that allow *Prochlorococcus* to grow from low cell numbers (Sher *et al.* 2011). All wells with visible color were then transferred into glass tubes containing 100% Red Sea seawater-based Pro99 media (modified after Moore *et al.* 2007). After two months of weekly transfers into fresh medium, we maintained the isolates in 75% Red Sea seawater-based Pro99. Cultures were regularly checked for heterotrophic contaminants using the marine purity broth as described in Saito *et al.* (2002). Larger volumes of the purified RSP50 culture were prepared in 200-mL polycarbonate flasks to collect sufficient biomass for growth measurements, transmission electron microscopy (TEM) preparations, and DNA extraction. All culture-ware and filtration units were acid-washed prior to use.

Growth measurements with flow cytometry

Three replicates of *Prochlorococcus* RSP50 were subsampled daily for 14 consecutive days and fixed using a 25% glutaraldehyde solution upon sampling into cryotubes (1% final concentration).

All samples were flash frozen in liquid nitrogen and stored at –80°C until processing with a LSR Fortessa II analyzer (Becton Dickinson, CA, USA). Samples were analyzed in 96-well plates, each containing 1 μL of a 1-μm yellow-green bead solution (Fluospheres, Life Technologies, CA, USA) as a standard for fluorescence and scatter. *Prochlorococcus* cells were identified according to the fluorescence of their natural pigments after excitation by a solid-state blue (488 nm) and a solid-state red (640 nm) laser, using two bandpass filters (670/30 nm and 695/40 nm). Cytograms were analyzed using the FCS Express 5 software (De Novo software, CA, USA). Growth rates were estimated based on the method described by Wood, Everroad and Wingard (2005).

Transmission electron microscopy

To obtain high-resolution images of the cells, 40-mL aliquots of *Prochlorococcus* RSP50 in exponential growth phase were pelleted by centrifugation. Pellets were fixed with glutaraldehyde (2.5% final concentration) in cacodylate buffer (0.1 mol L^{-1}, pH 7.4) and treated with reduced osmium (1:1 mixture with 2% aqueous potassium ferrocyanide) for 1 h, as described in Karnovsky (1971). Cells were pre-embedded in agar, dehydrated using an ethanol series (70, 80, 90, 95 and 100%), and subsequently embedded in epoxy resin. Sections (80–120 nm thick) were placed onto copper grids and contrasted with lead citrate. Images were visualized and recorded using a Titan Transmission Electron Microscope (80–300 kV, FEI, USA) and a charge-coupled device (CCD) camera (Gatan Inc., USA). Cell size measurements were obtained using the DigitalMicrograph software (Gatan Microscopy Suite) and the ImageJ software (www.imagej.nih.gov).

DNA extraction and library preparation

The total genomic DNA from 50 mL *Prochlorococcus* RSP50 cell culture in 75% Pro99 media was extracted using the DNeasy Kit (Qiagen) following the manufacturer's protocol under the recommended pre-treatment for Gram-negative bacteria. A fraction of the extracted DNA was used to amplify the 16S rRNA gene and the intergenic transcribed spacer (ITS) gene sequences with 16S rRNA universal primers 27F (5′-AGAGTTTGATCMTGGCTCAG-3′) and 1492R (5′-TACGGYTACCTTGTTACGACTT-3′) and ITS primers Picocya-16S-Forward (5′-TGGATCACCTCCTAACAGGG-3′) and Picocya-23-Reverse (5′-CCTTCATCGCCTCTGTGTGCC-3′), respectively (Cai *et al.* 2010; Huang *et al.* 2012; Jiao *et al.* 2014). The purified PCR products were Sanger-sequenced for phylogenetic assessment and verifying the purity of the strains.

Libraries for whole-genome sequencing were prepared using the TruSeq paired-end DNA sample preparation kit (Illumina, USA) and indexed using TruSeq adapters. Quantification and quality control of the libraries was conducted using a Qubit assay (dsDNA kit, Qubit, USA), and analysis using a DNA 2100 Bioanalyzer (Agilent Technologies Inc., USA).

Whole genome sequencing, assembly and annotation

The libraries were sequenced on an Illumina MiSeq platform at the BCL facility at KAUST. Raw reads were quality-checked using FastQC v0.11.4 (Andrews 2010) and quality-trimmed using Trimmomatic v0.36 (Bolger, Lohse and Usadel 2014) before being assembled into contigs with SPAdes v3.1.1, employing the read error-correction and mismatch repair modes (Bankevich *et al.* 2012). This resulted in a single 1.66-Mbp high-quality contig, which was subsequently annotated using the INDIGO pipeline (Alam *et al.* 2013). Individual gene annotations were further manually curated using the Argot2 server (Falda *et al.* 2012) following the protocol by Biller *et al.* (2014). Genome completeness and fidelity were estimated by determining the presence (or absence) of a suite of 104 single-copy marker genes universally present in complete bacterial genomes as implemented in CheckM v0.9.7 (Parks *et al.* 2015). The *Prochlorococcus* RSP50 genome was deposited in NCBI GenBank under accession number CP018344.

Comparative genomics was conducted using the phylogenomic analysis server EDGAR v2 (Blom *et al.* 2016). For this purpose, 43 genomes of publicly available *Prochlorococcus* strains (Biller *et al.* 2014) were downloaded from ProPortal (Kelly *et al.* 2011) and included together with RSP50 for downstream analyses. Average amino acid identity values, genome-to-genome distances and Venn diagrams were also calculated in EDGAR.

Phylogeny

For phylogenetic inference, the PCR-amplified 16S rRNA gene sequences of RSP50 and those retrieved from the genomes of representative *Prochlorococcus* strains (Table S1) were aligned using ClustalW in GeneiousPro v8.0 (Kearse *et al.* 2012). Neighbor-joining trees were generated using p-distance and 1000 bootstraps, while maximum likelihood trees were generated using the General Time Reversible model of nucleotide substitution with the gamma model of rate heterogeneity (GTR-GAMMA), as estimated in RAxML (Stamatakis 2014), with 1000 bootstraps. All phylogenetic trees were generated in MEGA v6.0 (Tamura *et al.* 2013). To delineate ecotypes within *Prochlorococcus* at a higher resolution, the same approach was applied for full length ITS sequences.

To construct the phylogenomic tree, 239 conserved single-copy genes among all *Prochlorococcus* genomes were identified with Hal v2.0 (Robbertse *et al.* 2011). Neighbor-joining and maximum likelihood trees were constructed using p-distance and the Jones-Taylor-Thornton (JTT) model of amino acid substitution as estimated in ProtTest v3.2 (Darriba *et al.* 2011) with 1000 bootstraps.

To estimate divergence time of the unique genes in question, phylogenetic time trees were generated using Reltime in MEGA v7.0 (Tamura *et al.* 2012; Kumar, Stecher and Tamura 2016). The protein sequences from RSP50 and marine *Synechococcus* strains were aligned with MUSCLE, and maximum likelihood trees using the Le-Gascuel (LG+I+G) model were constructed with 1000 bootstraps.

Sample collection for total community RNA analysis

Seawater samples were collected in March 2013 at the Kebrit Deep site in the Red Sea (24°43'12" N, 36°16'48" E) during the KAUST Red Sea expedition (KRSE2013) on board the *R/V Aegaeo*. Four depths in the water column were sampled: surface (10 m), below the mixed layer (BML; 40 m), deep chlorophyll maximum depth (DCM; 75 m), and oxygen minimum zone (OMZ; 420 m). Duplicate samples were collected using a rosette sampler with 24 Niskin bottles (10 L each) every 4 h over a 48-h period. An attached CTD unit and additional sensors were used to record salinity, temperature, fluorescence, and dissolved oxygen readings at the sampling sites. For each time point and depth, 1 L of seawater was filtered using a peristaltic pump with two in-line filters in series consisting of a 1.6-μm pre-filter (Whatman GF/A,

Whatman Inc., USA) followed by a 0.22-μm Sterivex filter (Millipore Corp., USA). A 2-mL RNAlater solution (Qiagen Inc., USA) was immediately added to completely fill the filter columns. The Sterivex filters were then sealed with Blu-Tack, flash frozen in liquid nitrogen and stored at –20°C on board.

RNA extraction, purification and library preparation

RNA was extracted according to a protocol adapted from Massana *et al.* (1997), Béjà *et al.* (2002), and Stewart, Ulloa and DeLong (2012). Briefly, the RNAlater solution in the Sterivex filter was expelled and a 2-mL lysozyme solution (1 mg mL^{-1} in lysis buffer: 40 mmol L^{-1} EDTA, 50 mM Tris pH 8.3, 0.73 M sucrose) was added. The Sterivex filter was left to incubate at 37°C for 45 min with rotation. Then, 50 μL of Proteinase K solution (20 mg/mL; 5PRIME) and 100 μL of a 20% SDS solution were added. The filter was then left to incubate at 55°C for 2 h with rotation. After incubation, 1 mL of lysis buffer was added to the filter at 55°C for 15 min for washing. The lysate was expelled to a separate tube and 1.5 mL of absolute ethanol was added. From this solution, RNA was extracted using the RNeasy Protect Bacteria Mini Kit (Qiagen Inc.) according to the manufacturer's protocol and eluted in RNase-free water. The resulting RNA was concentrated from 250 μL to 60 μL using a speed vacuum (Thermo Scientific, Bremen, Germany). To remove DNA, the samples were treated with DNaseI (Turbo DNA-free kit, Ambion) by adding 6 μL of 10 × buffer and left to incubate at 37°C for 30 min. To purify the samples, the RNeasy MinElute Cleanup Kit (Qiagen Inc.) was used and RNA was eluted with 14 μL of RNase-free water. To preferentially amplify mRNA (independent of poly-A tail) and select against rRNA from total RNA (range of 2–4 ng), the ExpressArt C&E Bacterial NANO RNA Amplification Kit (AMSBIO, UK) was used. Sequencing libraries were prepared using the paired-end TruSeq RNA sample preparation kits (Illumina, USA) following the manufacturer's protocol.

RNA library sequencing and genome mapping

Libraries were pair-end sequenced on the Illumina HiSeq 2000 platform (2 × 100 bp). Low quality reads and sequencing adapters were removed using Trimmomatic v0.32 (Bolger, Lohse and Usadel 2014). Sequence reads shorter than 50 bp were discarded. Bowtie2 v2.2.4 (Langmead and Salzberg 2012) was used to identify and remove PhiX contamination sequences. The remaining sequences were error-corrected using the BayesHammer algorithm (Nikolenko, Korobeynikov and Alekseyev 2013) implemented in the SPAdes v3.5.0 (Bankevich *et al.* 2012), followed by removal of putative Ribosomal RNA (rRNA) gene transcripts with SortMeRNA v2.0 (Kopylova, Noé and Touzet 2012).

The high-quality mRNA reads were subsequently mapped against the RSP50 genome using Bowtie2 with default settings. The resulting read counts were normalized based on the FPKM metric (fragments per kilobase of gene per million mapped reads; Pachter, 2011), which was used for differential gene expression analysis. The raw RNASeq sequences (164 Gbp) were deposited in NCBI GenBank under Bioproject number PRJNA289956.

RESULTS

Growth and morphology of *Prochlorococcus* RSP50

Open-water samples of the Red Sea collected for culturing and media preparation had an *in-situ* temperature of 28.4°C and a salinity of 39.5. Isolation procedures and successive transfers of the enrichments resulted in a bright green-yellow culture of *Prochlorococcus* (RSP50, Fig. S1). Fluorescence microscopy and flow cytometry using SYBR Green I-staining confirmed the absence of other microbes. No growth of heterotrophic cultures was observed in Marine Purity Test Broth (MPTB) in the dark (Saito *et al.* 2002), and thus the culture was deemed putatively pure. RSP50 in 75% Pro99 media (Moore *et al.* 2007) had a growth rate of 0.57 ± 0.01 (doubling time of 1.22 days) and a maximum abundance of 5 × 10^5 cells mL^{-1} (Fig. S2). The RSP50 cells were ovoid and 0.6 μm long, with a width-to-length ratio of 0.75 (Fig. 1).

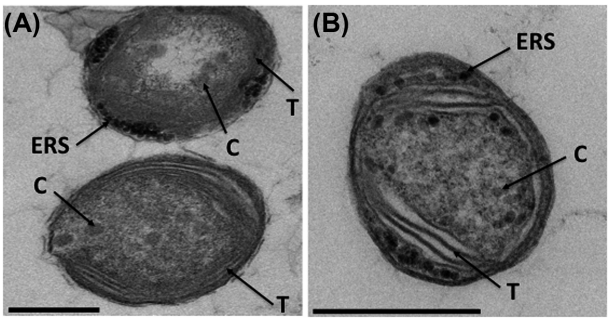

Figure 1. TEM images of the strain RSP50. Thylakoids (T), carboxysomes (C) and electron-rich structures (ERS) are labelled. The scale bar indicates a length of 500 nm.

16S rRNA, ITS, and core-gene phylogeny of *Prochlorococcus* RSP50

Phylogenetic analysis of 16S rRNA and ITS gene sequences places strain RSP50 within the HL II clade of *Prochlorococcus*, in close proximity to strains AS9601, MIT9301 and SB (Fig. S3). The strain RSP50 also clustered with the HL II lineage containing AS9601, MIT9301, SB, MIT0604 and MIT9314, based on phylogenetic analyses inferred using a concatenated alignment of 239 conserved, single-copy genes (SCGs) shared between RSP50 and representative *Prochlorococcus* and *Synechococcus* strains (Fig. 2).

Genome characteristics

The genome of strain RSP50 encompasses a single contig of 1656 033 base pairs in length, with 1939 predicted protein-coding genes and a DNA G+C content of 31%. Table S1 summarizes these features in relation to other sequenced *Prochlorococcus* representatives. This genome is estimated to be ~97% complete based on the profile of single-copy marker genes present in the *Prochlorococcus* lineage (CheckM v0.9.7; Parks *et al.* 2015). The highest percentage of genes clustered in the following categories: translation, ribosomal structure and biogenesis, followed by amino acid transport and metabolism (Fig. S4). Eighteen per cent of genes assigned to COG categories had an unknown function or only a general function prediction. Pairwise comparison of the orthologous coding genes of RSP50 with other HL II strains revealed that it has the highest average amino acid identity (AAI) with strain AS9601 from the Arabian Gulf (94.9%), SB (94.8%), MIT9314 (94.7%) and MIT9301 (94.5%). The identity with LL strains is low (62–64%; Table S2), in comparison with other HL strains.

Comparative genomics indicated that 93.3% of the protein-coding genes of strain RSP50 (1810 genes) are conserved relative to the pangenome of the currently available HL- and LL-adapted *Prochlorococcus* genomes (Fig. 3). Strain RSP50 harbors all

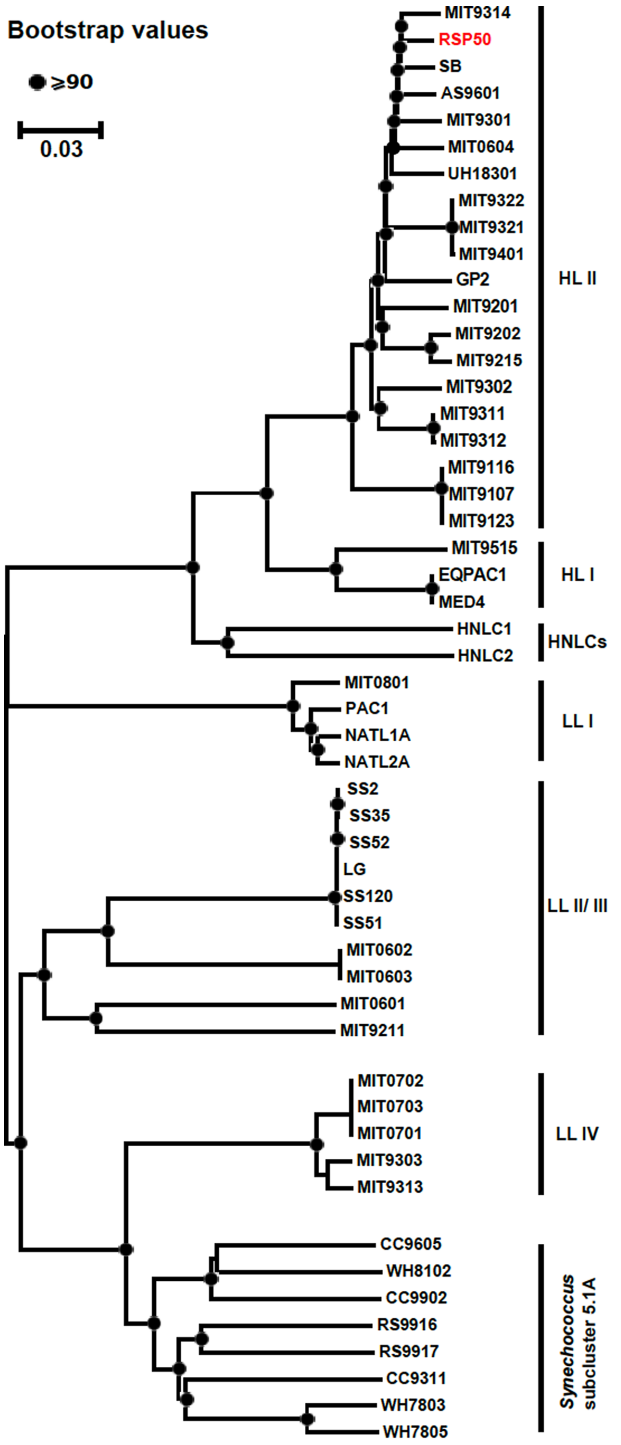

Figure 2. Phylogenomic placement of strain RSP50 based on 239 conserved single-copy genes. Sequences were retrieved from all lineages available to date that represent the ecotypes of *Prochlorococcus*. Strain RSP50 is shown in red. Branches with bootstrap values ≥90% are indicated with a solid black circle, while the scale bar indicates an estimated sequence divergence of 3%. Strains belonging to *Synechococcus* subcluster 5.1A were used as an outgroup.

necessary genes reported within the HL II clade for photosynthesis, carbon fixation and nutrient acquisition. The main components of the photosynthetic apparatus (*psbA1* gene encoding the D1 protein of Photosystem II) and light-harvesting complexes (*pcbD* gene encoding a chlorophyll *a*-*b* binding protein and

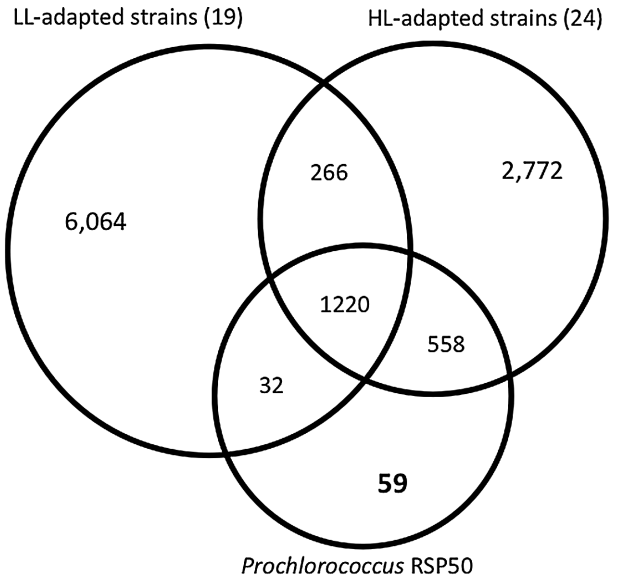

Figure 3. Venn diagram of gene clusters. The number of genes present in one or more genomes of the 24 HL-adapted strains and 19 LL-adapted strains (see Table S1 for list of genomes used) and strain RSP50 are shown inside the circles. The number of unique genes found only in strain RSP50 (59) is also shown.

pebA/*pebB* genes encoding phycoerythrobilin) are also present. All genes encoding carboxysome shell proteins and the ribulose-1,5-bisphosphate carboxylase (RuBisCo) subunits for carbon fixation (Scanlan *et al.* 2009) were found. For nutrient acquisition, the RSP50 genome carries genes necessary for nitrogen regulation and sensing (*ntcA* and *glnB*) and has a complete glutamine synthetase-glutamate synthase pathway for ammonium assimilation and a complete transport system for urea. The high-affinity phosphate transport protein system encoded by *pstSCAB* genes and a putative alkaline phosphatase-like protein (encoded for by *dedA*) are found in RSP50. Additionally, the genome has a phosphonate transport system encoded by the *phnCDE* gene set. A plastoquinol terminal oxidase, a nickel-containing superoxide dismutase and several peroxidases make up the existing reaction oxygen species (ROS)-scavenging system (Scanlan *et al.* 2009). Photoacclimation-related genes such as photolyases and a number of high-light inducible proteins (encoded by *hli* genes) are also present. Similar to other *Prochlorococcus* strains (Scanlan *et al.* 2009), RSP50 possesses the *spsA* gene encoding for sucrose-phosphate synthase and the *stpA* gene encoding for glucosyl-glycerolphosphate phosphatase, which is involved in the biosynthesis of the compatible solute 2-*O*-(α-D-glucopyranosyl) glycerol.

Unique genes of RSP50 and transcription of the glucosylglycerol biosynthetic pathway

The total number of unique genes found in RSP50 was 59 (i.e. 3% of all genes; Fig. 3), and the majority of these unique genes were either hypothetical or uncharacterized genes. Less than half the genes (23) returned annotations, and their functional predictions were further verified using the independent Argot2 pipeline (Falda *et al.* 2012). Based on COG functional categories, the majority of these genes are assigned to cell wall/membrane/envelope biogenesis and carbohydrate transport and metabolism, including starch and sucrose, amino sugars and nucleotide sugars (for the full list of genes see Table S3). Three of those genes specific to RSP50 are likely involved

in the biosynthesis of glucosylglycerol (GG), a compatible solute that is produced by moderately halophilic *Synechococcus* strains and other cyanobacteria (Klähn & Hagemann 2011). In this pathway, GG is putatively synthesized in a two-step reaction by two enzymes: glucosylglycerol-phosphate synthase (*ggpS*) generates glucosylglycerol-phosphate from glycerol-3-phosphate and ADP-α-D-glucose, followed by glucosylglycerol-3-phosphatase (*stpA*), which dephosphorylates glucosylglycerol-phosphate to glucosylglycerol (Pade & Hagemann 2014). Glycerol-3-phosphate is likely produced by a putative glycerol-3-phosphate dehydrogenase (*gspA*) involved in the glycerol biosynthesis pathway. A third gene encodes a trehalose synthase (also known as maltose alpha-D-glucosyltransferase or alpha amylase catalytic protein). These three RSP50-specific genes (*ggpS*, *gspA* and trehalose synthase) are orthologous to those in *Synechococcus* species (Fig. 4A and Figs S5–S6). The RSP50-specific *ggpS* gene shows an average amino acid identity of 70% against the *Synechococcus* orthologs, whereas *gspA* and trehalose synthase displayed average identities ranging from 57 to 77% (Table S3). The geolocalized abundances of the genes, derived from the Ocean Gene Atlas (OGA; Villar *et al.* 2018) using Red Sea samples of the *Tara* Oceans datasets (Pesant *et al.* 2015), ranges from 0.04 to 1.82 copies per cell across the different size fractions (0–0.22 μm, 0.22–1.6 μm, and 0.1–0.22 μm) and sampling depths (surface and deep chlorophyll maximum) (Table S4).

For *gspA*, the divergence time between RSP50 and its *Synechococcus* counterparts was approximately 0.68 MYA (Fig. 4B), while trehalose synthase and *ggpS* seem to have diverged for approximately 0.31 and 0.36 MYA, respectively (Figs S7–S8).

To gain more insight into the role of these genes in osmoregulation of *Prochlorococcus* in the Red Sea, we mapped them against several metatranscriptomes from the water column of the Red Sea. The transcriptional profile of *ggpS* and *gspA* genes revealed that both were constitutively being expressed at higher levels in surface waters and BML compared with the DCM. The expression level of the *gspA* gene was always higher than that of *ggpS*, but no clear diel pattern was observed. Expression of the gene encoding a trehalose synthase dropped in the DCM compared with surface and BML and seemed to peak during the night (Fig. 5).

DISCUSSION

Characteristics of *Prochlorococcus* RSP50

We describe the isolation and genome characterization of strain RSP50, a *Prochlorococcus* strain inhabiting the surface waters of the Red Sea. Flow cytometry and cell sorting enabled a reliable discrimination of *Prochlorococcus* from its sister group *Synechococcus* as well as other marine heterotrophs (Fig. S1), which was important to detect contaminants.

The growth rates of RSP50 at 27°C (0.57 $\pm$ 0.01, Fig. S2) are comparable with the growth rates of strains MIT9312 at 27°C (0.61 $\pm$ 0.01) and MIT9215 at 25°C (0.5 $\pm$ 0.01) (Johnson *et al.* 2006). In contrast to HL II ecotypes, HL I and LL ecotypes generally have lower temperature optima and growth rates (Johnson *et al.* 2006; Zinser *et al.* 2007).

The size of RSP50 (~0.6 μm diameter, Fig. 1) is typical of HL-adapted *Prochlorococcus* cells. The smaller size of HL ecotypes, including strain RSP50, relative to LL ecotypes (up to 0.8 μm) has been invoked in the hypothesis that cell size and architecture are factors in niche adaptation (Ting *et al.* 2007).

Phylogeny of RSP50

Phylogenetic inference using 16S rRNA and ITS sequences of RSP50 confirmed that it belongs to HL II (Fig. S3), the most abundant *Prochlorococcus* clade in the Red Sea water column (Ngugi and Stingl 2011; Shibl *et al.* 2014, 2016). The phylogeny based on conserved SCGs placed RSP50 near to strains AS9601 and MIT9314 that were isolated from the Arabian Sea and the Gulf Stream, respectively. Although highly debatable, AAI based on whole genomes of *Prochlorococcus* was reported to be the method with the highest resolution for genomic taxonomy (Thompson *et al.* 2013a). RSP50 shares the highest AAI (>94%) with HL II strains dominating warm, nutrient-poor environments (AS9601 [Arabian Sea], MIT9314 [Gulf Stream], SB [Western Pacific] and MIT9301 [Sargasso Sea]), corroborating the phylogenetic inference from 16S and ITS sequences.

Comparative genomics

Prochlorococcus evolved through several genome-streamlining events, allowing ecotypes in the water column to survive with a minimal set of genes (Partensky and Garczarek 2010). *Prochlorococcus* RSP50 has both a genome size and DNA G+C content within the typical range defined for other HL II strains (Kettler *et al.* 2007) (Table S1). The large number of genes with unknown function is a common feature among various *Prochlorococcus* ecotypes (Coleman *et al.* 2006).

The core functionality of *Prochlorococcus* genes has been covered broadly and compared extensively (Scanlan *et al.* 2009). The RSP50 genome displays the complete glutamine synthetase-glutamate synthase pathway for ammonium assimilation and a complete transport system for urea. Genes that encode specific ABC phosphonate transporters and putative alkaline phosphatase-like proteins likely enable access to phosphorus sources other than inorganic phosphate (Moore *et al.* 2005). For potential ROS protection, RSP50 carries the gene encoding for a nickel-binding superoxide dismutase (Ni-SOD), which was previously linked to iron and manganese conservation (Palenik *et al.* 2003).

Salt tolerance and gene expression patterns

In comparison with all sequenced *Prochlorococcus* genomes used in this study, several genes were identified exclusively in the RSP50 genome (Fig. 3, Table S3). Despite the majority of the unique gene pool being uncharacterized, three genes are likely to be involved in the biosynthesis of compatible solutes to counter the effect of high salinity levels: glucosylglycerol-phosphate synthase (*ggpS*), glycerol-3-phosphate dehydrogenase and trehalose synthase/alpha amylase catalytic protein. In cyanobacteria, the choice of compatible solutes under high salinity is sucrose, trehalose and glucosylglycerol; some even accumulate glycine/homoserine betaine (Kirsch *et al.* 2017).

Glucosylglycerol-phosphate synthase (*ggpS*) catalyzes the synthesis of the intermediate glucosylglycerol-3-phosphate (GG-3P). Glucosyl-glycerolphosphate phosphatase (*stpA*) then dephosphorylates GG-3P into the compatible solute glucosylglycerol (Klähn *et al.* 2010). Both *ggpS* and *stpA* are found in *Synechococcus*, while all previously sequenced *Prochlorococcus* genomes possess *stpA* but lack *ggpS* (Scanlan *et al.* 2009). The role of *stpA* in *Prochlorococcus* currently remains unknown but is presumably able to dephosphorylate alternative sugar phosphates (Hagemann 2013). It is likely that RSP50 carries *ggpS* to regain the

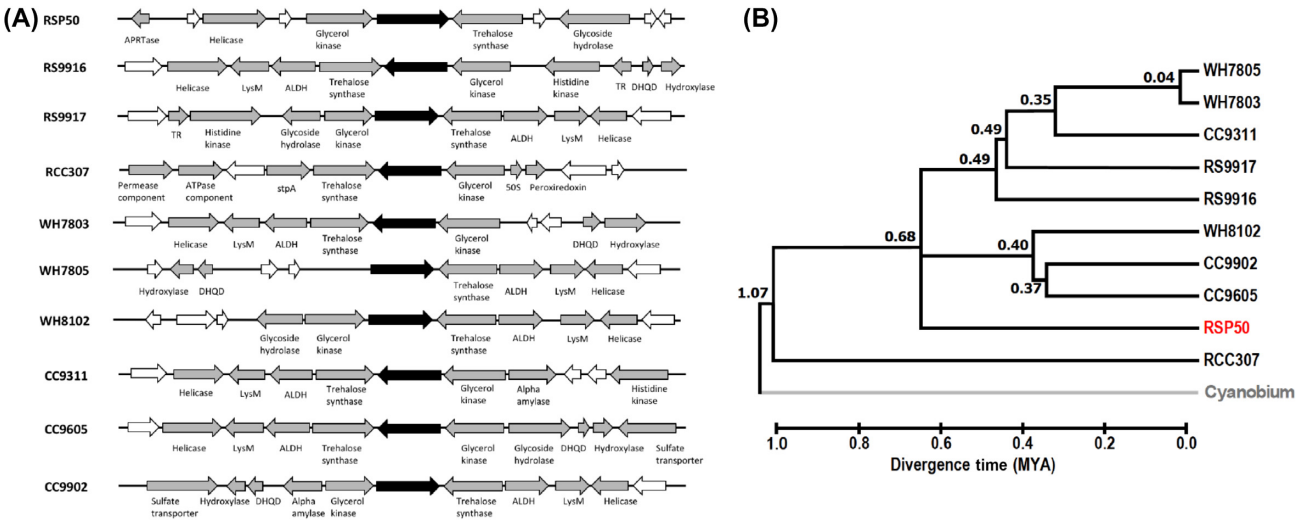

Figure 4. **(A)** Gene neighborhood diagram showing genes in close proximity to glycerol-3-phosphate dehydrogenase (black arrows). Grey arrows denote annotated flanking genes and white arrows denote hypothetical/putative or uncharacterized genes. APRTase: adenine phosphoribosyl-transferase, LysM: peptidoglycan-binding protein, ALDH: aldehyde dehydrogenase, TR: transcriptional regulator, DHQD: 3-dehydroquinate dehydratase, stpA: glucosylglycerol-phosphate phosphatase. **(B)** The phylogenetic time tree for glycerol-3-phosphate dehydrogenase. The tree was generated using the Reltime method in MEGA v7.0 (Kumar, Stecher and Tamura 2016). The gene from RSP50 is colored in red and the outgroup (Cyanobium sp. PCC7001) is colored in grey. Numbers on branches are relative times of divergence between lineages. The scale indicates divergence time of genes (Million Years Ago).

ability to synthesize the N-free compatible solute glucosylglycerol.

Glycerol-3-phosphate dehydrogenase catalyzes the reversible conversion of dihydroxyacetone phosphate (DHAP) into glycerol-3-phosphate. Glycerol-3-phosphate is the precursor molecule in the biosynthesis of GG-3P and, eventually, the compatible solute glucosylglycerol. We hypothesize that glycerol-3-phosphate dehydrogenase—found only in RSP50 thus far—likely regulates the cellular levels of glycerol-3-phosphate (Rawls, Martin and Maupin-Furlow 2011). Many *Prochlorococcus* strains have the ability to accumulate more than one compatible solute (Scanlan *et al.* 2009) and they all (including RSP50) carry genes for sucrose biosynthesis by the sucrose-phosphate synthase pathway. It has been shown that there is an enrichment of compatible solute-degrading (oxidizing) genes in *Pelagibacter* (SAR11) from the Red Sea relative to the Mediterranean Sea, Sargasso Sea and North Pacific Subtropical Gyre (Thompson *et al.* 2013b). This led to the hypothesis that there are higher compatible solute concentrations in the Red Sea relative to these other seas. As Thompson *et al.* (2013b) relied on existing *Prochlorococcus* reference genomes for metagenomics mapping (none of which came from Red Sea strains), the specific (unique) genetic repertoire of Red Sea populations remains as of yet unidentified. The high salinity levels of the Red Sea may have required glucosylglycerol-phosphate synthase and glycerol-3-phosphate dehydrogenase to synthesize and regulate an additional compatible solute in *Prochlorococcus*.

Trehalose synthase, also known as maltose α-D-glucosyltransferase or alpha-(α)-amylase, catalyzes the conversion of maltose to the compatible solute trehalose (Ruhal, Kataria and Choudhury 2013). Maltose can be taken up from the environment via ABC transporters (Boos and Shuman 1998); however, the substrate specificity for the majority of those transporters in *Prochlorococcus* remains unknown (Hagemann 2010). Despite the general absence of a trehalose uptake system in picocyanobacteria, *Prochlorococcus* MED4 possess a trehalase gene (*treH*) (Scanlan *et al.* 2009), while some *Synechococcus* strains carry the trehalose synthase gene (Hagemann 2010). In a more environmental context, homologs of all three genes identified in the *Tara* Oceans datasets showed higher abundances related to higher salinity levels in the main basin of the Red Sea (Table S4), likely due to the requirement of more osmolyte biosynthesis. The organization and content of the flanking regions of these genes in RSP50 show significant similarity to *Synechococcus* genomes (Figs 4A and S5–S6). Moreover, divergence time estimates (Figs 4B and S7–S8) suggest that these three genes may have been lost during genome reduction events in other *Prochlorococcus* lineages and were retained in RSP50. Interestingly, this is consistent with previous observations that indicated that these genes were among those lost during the early divergence of *Synechococcus* and *Prochlorococcus* (Kettler *et al.* 2007).

Mapping the RSP50 genome to metatranscriptomes collected from different depths over two days revealed the expression patterns of those unique genes. High transcriptional activity of *Prochlorococcus* populations detected at the DCM of Station ALOHA (125 m) in the North Pacific (Shi *et al.* 2011) and at Station A (85 m) in the Gulf of Aqaba (Miller *et al.* 2017) suggested that HL-like cells are likely viable at this depth. The light-dependent gene *psbA*, encoding a component of the photosynthetic apparatus in all strains, follows the expected diel pattern, where it is highly expressed during the daytime but significantly less as it gets dark (Fig. 5A). Glucosylglycerol-phosphate synthase and glycerol-3-phosphate dehydrogenase were both consistently expressed at all depths, with higher levels of expression for the latter (Fig. 5B and C). The absence of significant changes in expression at different light levels of these two genes suggest they are independent from the diel cycle. Expression levels likely depend on cytosolic levels of the precursor glycerol-3-phosphate (Rawls, Martin and Maupin-Furlow 2011). Similarly, the *spsA* gene involved in sucrose synthesis showed only minor changes in its expression pattern along the diel cycle (Fig. S9). Strain AS9601 grown under continuous light showed an increased expression of the genes involved in the biosynthesis of glucosylglycerate (GGA) and sucrose under salt stress (Al-Hosani *et al.* 2015). On the other hand, trehalose synthase peaks

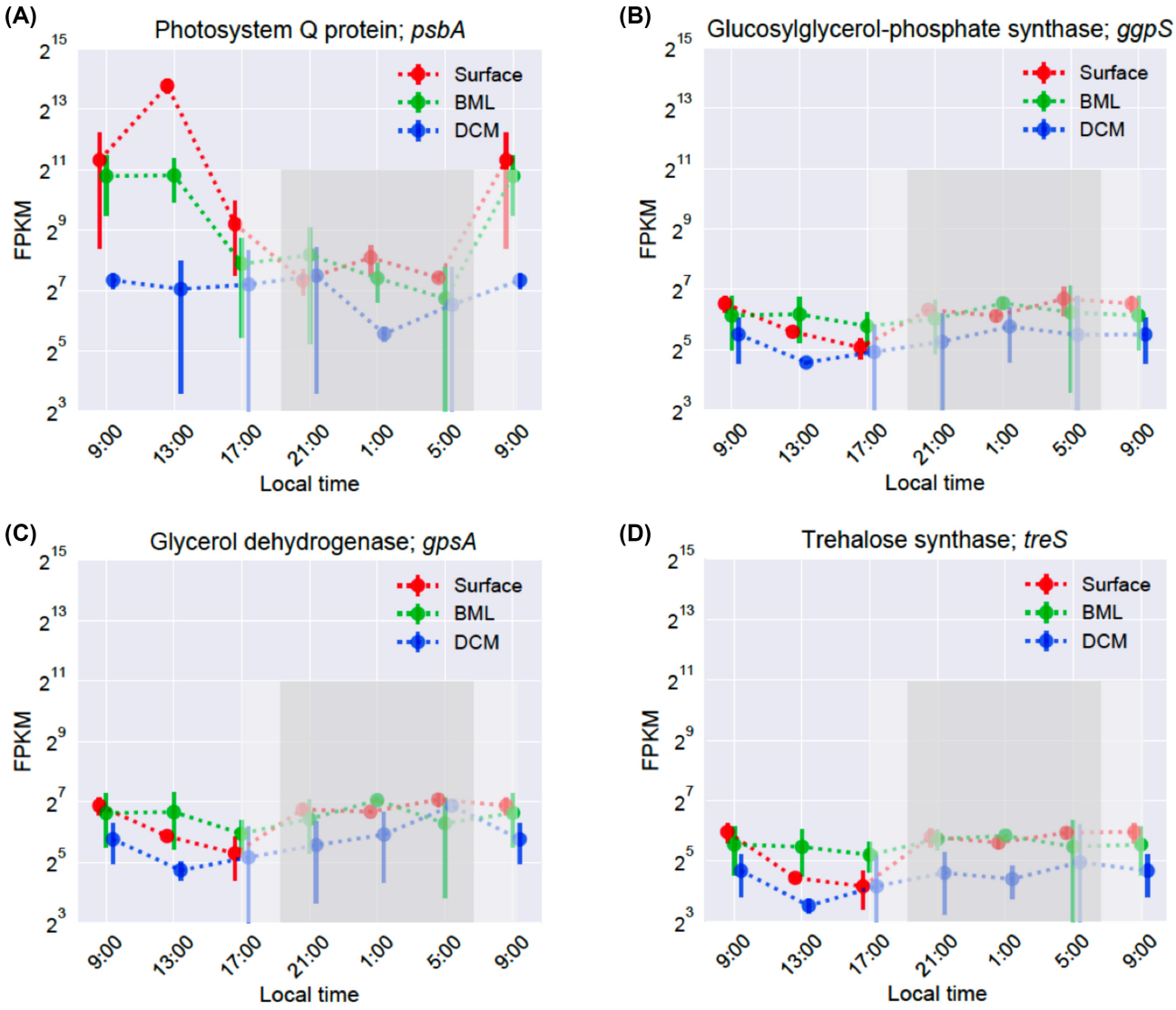

Figure 5. Expression patterns of the *psbA* gene (photosystem Q protein) and unique genes discussed in this study: **(A)** photosystem Q protein (*psbA*), **(B)** glucosylglycerol-phosphate synthase (*ggpS*), **(C)** glycerol dehydrogenase (*gpsA*) and **(D)** trehalose synthase (*treS*). Colored lines represent different depths in the water column. The white areas represent early and late light phases of the diel cycle while grey areas represent the dark phase.

during the night and could possibly follow the diel cycle for optimal functioning (Fig. 5D).

Genomic variability between HL- and LL-adapted *Prochlorococcus* clades, and within distinct HL (I-VI) and LL (I-VII) lineages, arises from the preferential retention of gene sets necessary to occupy contrasting environments and particular niches in the water column (Kettler *et al.* 2007; Scanlan *et al.* 2009; Partensky and Garczarek 2010). For example, several *Prochlorococcus* strains at Station ALOHA in the North Pacific express genes involved in the complete and active nitrogen metabolism pathways *in situ* (Berube *et al.* 2015, 2016). Similarly, transcripts in *Prochlorococcus* MED4 cells related to phosphorus sensing (*phoR*) and uptake (*pstS*), in addition to phosphatase (*phoA*), were upregulated under phosphorus limitation, the latter likely to scavenge phosphorus from alternative, organic sources (Reistetter *et al.* 2013). Under salt stress, carbon fixation and biosynthesis of compatible solutes were upregulated in *Prochlorococcus* AS9601 cells, whereas photosynthesis and cell division were inhibited (Al-Hosani *et al.* 2015).

CONCLUSIONS

The global dominance of *Prochlorococcus* is likely to expand over higher latitudes as a result of warmer waters (Flombaum *et al.* 2013). It was suggested, through scanning of 111 marine metagenomes from the global oceans, that a number of hitherto undiscovered *Prochlorococcus* species exist (Thompson *et al.* 2013a). *Prochlorococcus* RSP50, a Red Sea-specific strain belonging to HL II, potentially belongs to an existing species based on its AAI compared to several previously sequenced strains. Despite having a large number of genes lacking functional annotation, RSP50-specific genes included three genes that encode enzymes potentially involved in the biosynthesis of compatible solutes. Their presence presumably adds to the cells' salt acclimation system and possibly allows them to survive the Red Sea environment. The organization of these genes on the genome is highly similar to those on genomes of *Synechococcus* strains, and were likely progressively lost in *Prochlorococcus* lineages but kept in the RSP50 genome. Along the water column of the Red

Sea, the expression of those unique genes involved in compatible solute biosynthesis was more pronounced in the surface and BML than in the DCM, and their expression pattern over a diel cycle did not show any significant correlation. The observations made here provide a glimpse of how selective pressures of the Red Sea can shape the adaptive traits of local *Prochlorococcus* cells. Further experiments to characterize the enzymes and test their functionality under different salt concentrations are required. This is the first report of a cultivated *Prochlorococcus* strain from the main basin of the Red Sea, which should provide a resource to understand the ecology and adaptations of abundant phytoplankton in the changing marine ecosystem under different global warming scenarios.

SUPPLEMENTARY DATA

Supplementary data are available at *FEMSEC* online.

ACKNOWLEDGEMENT

We thank AbdulAziz Al-Suwailem from the Coastal and Marine Resources Core Lab (CMRC) and the scientists, captain and crew of the *R/V Aegaeo* from the KAUST Red Sea Expedition 2013 for their logistical and technical support. We also thank Rachid Sougrat from the Imaging Core Lab as well as Manal Andijani, Hicham Mansour, Alaguraj Dharmarajnadar and Hatim Talal Almansouri from the Biosciences Core Laboratory (BCL) at KAUST for their help with sample analysis.

FUNDING

This work was supported by the SABIC Career Development Chair from KAUST to US. Additional funding to support the publication of this manuscript was provided to US by the USDA National Institute of Food and Agriculture under Hatch project FLA-FTL-0 05631.

REFERENCES

Alam I, Antunes A, Kamau AA *et al.* INDIGO–INtegrated Data Warehouse of MIcrobial GenOmes with examples from the red sea extremophiles. *PLoS One* 2013;**8**:e82210.

Al-Hosani S, Oudah MM, Henschel A *et al.* Global transcriptome analysis of salt acclimated *Prochlorococcus* AS9601. *Microbiol Res* 2015:**176**:21–8.

Alraddadi TM. Temporal changes in the Red Sea circulation and associated water masses. *University of Southampton, Ocean and Earth Science, Doctoral Thesis* 2013:198.

Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010. http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Bankevich A, Nurk S, Antipov D *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012;**19**:455–77.

Béjà O, Suzuki MT, Heidelberg JF *et al.* Unsuspected diversity among marine aerobic anoxygenic phototrophs. *Nature* 2002;**415**:630–3.

Berube PM, Biller SJ, Kent AG *et al.* Physiology and evolution of nitrate acquisition in *Prochlorococcus*. *ISME J* 2015;**9**:1195.

Berube PM, Coe A, Roggensack SE *et al.* Temporal dynamics of *Prochlorococcus* cells with the potential for nitrate assimilation in the subtropical Atlantic and Pacific Oceans. *Limnol Oceanogr* 2016;**61**:482–95.

Biller SJ, Berube PM, Berta-Thompson JW *et al.* Genomes of diverse isolates of the marine cyanobacterium *Prochlorococcus*. *Scientific Data* 2014;**1**:140034.

Blom J, Kreis J, Spänig S *et al.* EDGAR 2.0: an enhanced software platform for comparative gene content analyses. *Nucleic Acids Res* 2016;**44**:W22–8.

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;**30**:2114–20.

Boos W, Shuman H. Maltose/maltodextrin system of *Escherichia coli*: transport, metabolism, and regulation. *Microbiol Mol Biol Rev* 1998;**62**:204–29.

Cai H, Wang K, Huang S *et al.* Distinct patterns of picocyanobacterial communities in winter and summer in the Chesapeake Bay. *Appl Environ Microbiol* 2010;**76**:2955–60.

Chisholm SW, Olson RJ, Zettler ER *et al.* A novel free-living prochlorophyte abundant in the oceanic euphotic zone. *Nature* 1988;**334**:340–3.

Churchill JH, Bower AS, McCorkle DC *et al.* The transport of nutrient-rich Indian Ocean water through the Red Sea and into coastal reef systems. *J Mar Res* 2014;**72**:165–81.

Coleman ML, Sullivan MB, Martiny AC *et al.* Genomic islands and the ecology and evolution of *Prochlorococcus*. *Science* 2006;**311**:1768–70.

Darriba D, Taboada GL, Doallo R *et al.* ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* 2011;**27**:1164–5.

Delmont TO, Eren AM. Linking pangenomes and metagenomes: the *Prochlorococcus* metapangenome. *PeerJ*. 2018;**6**:e4320.

Edwards FJ. Climate and oceanography. *Red Sea* 1987;**1**:45–68.

Falda M, Toppo S, Pescarolo A *et al.* Argot2: a large scale function prediction tool relying on semantic similarity of weighted Gene Ontology terms. *BMC Bioinformatics* 2012;**13**:S14.

Flombaum P, Gallegos JL, Gordillo RA *et al.* Present and future global distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proc Natl Acad Sci* 2013;**110**:9824–9.

Hagemann M. Molecular biology of cyanobacterial salt acclimation. *FEMS Microbiol Rev* 2010;**35**:87–123.

Hagemann M. Genomics of salt acclimation: synthesis of compatible solutes among cyanobacteria. *Advances in Botanical Research* 2013:27–55.

Huang S, Wilhelm SW, Harvey HR *et al.* Novel lineages of *Prochlorococcus* and *Synechococcus* in the global oceans. *ISME J* 2012;**6**:285–97.

Jiao N, Luo T, Zhang R *et al.* Presence of *Prochlorococcus* in the aphotic waters of the western Pacific Ocean. *Biogeosciences* 2014;**11**:2391.

Johnson ZI, Zinser ER, Coe A *et al.* Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. *Science* 2006;**311**:1737–40.

Karnovsky MJ. Use of ferrocyanide-reduced osmium tetroxide in electron microscopy. *J Cell Biol* 1971;**51**:146A.

Kearse M, Moir R, Wilson A *et al.* Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 2012;**28**:1647–9.

Kelly L, Huang KH, Ding H *et al.* ProPortal: a resource for integrated systems biology of *Prochlorococcus* and its phage. *Nucleic Acids Res* 2011;**40**:632–40.

Kent AG, Dupont CL, Yooseph S *et al.* Global biogeography of *Prochlorococcus* genome diversity in the surface ocean. *ISME J* 2016;**10**:1856–65.

Kettler GC, Martiny AC, Huang K *et al.* Patterns and implications of gene gain and loss in the evolution of *Prochlorococcus*. *PLos Genet* 2007;**3**:e231.

Kirsch F, Pade N, Klähn S *et al*. The glucosylglycerol-degrading enzyme GghA is involved in acclimation to fluctuating salinities by the cyanobacterium *Synechocystis* sp. strain PCC 6803. *Microbiology* 2017;**163**, 1319–28.

Klähn S, Steglich C, Hess WR *et al*. Glucosylglycerate: a secondary compatible solute common to marine cyanobacteria from nitrogen-poor environments. *Environ Microbiol* 2010;**12**:83–94.

Klähn S, Hagemann M. Compatible solute biosynthesis in cyanobacteria. *Environ Microbiol* 2011;**13**:551–62.

Kopylova E, Noé L, Touzet H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* 2012;**28**:3211–7.

Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* 2016;**33**:1870–4.

Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie2. *Nat Methods* 2012;**9**:357–9.

Massana R, Murray AE, Preston CM *et al*. Vertical distribution and phylogenetic characterization of marine planktonic *Archaea* in the Santa Barbara Channel. *Appl Environ Microbiol* 1997;**63**:50–6.

Miller DR, Pfreundt U, Elifantz H *et al*. Microbial metatranscriptomes from the thermally stratified Gulf of Aqaba/Eilat during summer. *Mar Geonomics* 2017;**32**:23–6.

Moore LR, Chisholm SW. Photophysiology of the marine cyanobacterium *Prochlorococcus*: ecotypic differences among cultured isolates. *Limnol Oceanogr* 1999;**44**:628–38.

Moore L, Ostrowski M, Scanlan DJ *et al*. Ecotypic variation in phosphorus acquisition mechanisms within marine picocyanobacteria. *Aquat Microb Ecol* 2005, DOI: 10.3354/ame039257.

Moore LR, Coe A, Zinser ER *et al*. Culturing the marine cyanobacterium *Prochlorococcus*. *Limnology and Oceanography: Methods* 2007;**5**:353–62.

Nanninga GB, Saenz-Agudelo P, Manica A *et al*. Environmental gradients predict the genetic population structure of a coral reef fish in the Red Sea. *Mol Ecol* 2014;**23**:591–602.

Nikolenko SI, Korobeynikov AI, Alekseyev MA. BayesHammer: Bayesian clustering for error correction in single-cell sequencing. *BMC Genomics* 2013;**14**:S7.

Ngugi DK, Antunes A, Brune A *et al*. Biogeography of pelagic bacterioplankton across an antagonistic temperature-salinity gradient in the Red Sea. *Mol Ecol* 2012;**21**:388–405.

Ngugi DK, Stingl U. Combined analyses of the ITS loci and the corresponding 16S rRNA genes reveal high micro-and macrodiversity of SAR11 populations in the Red Sea. *PLoS One* 2011;**7**:e50274.

Pachter L. Models for transcript quantification from RNA-Seq. *ArXiv Preprint* 2011; arXiv:1104.3889.

Pade N, Hagemann M. Salt acclimation of cyanobacteria and their application in biotechnology. *Life* 2014;**5**:25–49.

Palenik B, Brahamsha B, Larimer FW *et al*. The genome of a motile marine *Synechococcus*. *Nature* 2003;**424**:1037.

Parks DH, Imelfort M, Skennerton CT *et al*. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 2015;**25**:1043–55.

Partensky F, Garczarek L. *Prochlorococcus*: advantages and limits of minimalism. *Annual Review of Marine Science* 2010;**2**:305–31.

Pesant S, Not F, Picheral M *et al*. Open science resources for the discovery and analysis of *Tara* Oceans data. *Scientific Data* 2015;**2**:150023.

Raitsos DE, Pradhan Y, Brewin RJ *et al*. Remote sensing the phytoplankton seasonal succession of the Red Sea. *PLoS One* 2013;**8**:e64909.

Rasul NM, Stewart IC. The Red Sea: the formation, morphology, oceanography and environment of a young ocean basin. *Springer* 2015.

Rawls KS, Martin JH, Maupin-Furlow JA. Activity and transcriptional regulation of bacterial protein-like glycerol-3-phosphate dehydrogenase of the haloarchaea in *Haloferax volcanii*. *J Bacteriol* 2011;**193**:4469–76.

Reistetter EN, Krumhardt K, Callnan K *et al*. Effects of phosphorus starvation versus limitation on the marine cyanobacterium *Prochlorococcus* MED4 II: gene expression. *Environ Microbiol* 2013;**15**:2129–43.

Robbertse B, Yoder RJ, Boyd A *et al*. Hal: an automated pipeline for phylogenetic analyses of genomic data. *PLoS Currents* 2011. Doi: 10.1371/currents.RRN1213.

Rocap G, Distel DL, Waterbury JB *et al*. Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S-23S ribosomal DNA internal transcribed spacer sequences. *Appl Environ Microbiol* 2002;**68**:1180–91.

Ruhal R, Kataria R, Choudhury B. Trends in bacterial trehalose metabolism and significant nodes of metabolic pathway in the direction of trehalose accumulation. *Microb Biotechnol* 2013;**6**:493–502.

Saito MA, Moffett JW, Chisholm SW *et al*. Cobalt limitation and uptake in *Prochlorococcus*. *Limnol Oceanogr* 2002;**47**:1629–36.

Scanlan DJ, Ostrowski M, Mazard S *et al*. Ecological genomics of marine picocyanobacteria. *Microbiol Mol Biol Rev* 2009;**73**:249–99.

Sher D, Thompson JW, Kashtan N *et al*. Response of *Prochlorococcus* ecotypes to co-culture with diverse marine bacteria. *ISME J* 2011;**5**:1125.

Shi Y, Tyson GW, Eppley JM *et al*. Integrated metatranscriptomic and metagenomic analyses of stratified microbial assemblages in the open ocean. *ISME J* 2011;**5**:999.

Shibl AA, Thompson LR, Ngugi DK *et al*. Distribution and diversity of *Prochlorococcus* ecotypes in the Red Sea. *FEMS Microbiol Lett* 2014;**356**:118–26.

Shibl AA, Haroon MF, Ngugi DK *et al*. Distribution of *Prochlorococcus* ecotypes in the Red Sea basin based on analyses of *rpoC1* sequences. *Frontiers in Marine Science* 2016;**3**:104.

Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014;**30**:1312–3.

Stewart FJ, Ulloa O, DeLong EF. Microbial metatranscriptomics in a permanent marine oxygen minimum zone. *Environ Microbiol* 2012;**14**:23–40.

Stingl U, Tripp HJ, Giovannoni SJ. Improvements of high-throughput culturing yielded novel SAR11 strains and other abundant marine bacteria from the Oregon coast and the Bermuda Atlantic Time Series study site. *ISME J* 2007;**1**:361.

Tamura K, Battistuzzi FU, Billing-Ross P *et al*. Estimating divergence times in large molecular phylogenies. *Proc Natl Acad Sci* 2012;**109**:19333–8.

Tamura K, Stecher G, Peterson D *et al*. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 2013;**30**:2725–9.

Thompson CC, Silva GG, Vieira NM *et al*. Genomic taxonomy of the genus *Prochlorococcus*. *Microb Ecol* 2013a;**66**:752–62.

Thompson LR, Field C, Romanuk T *et al*. Patterns of ecological specialization among microbial populations in the Red Sea and diverse oligotrophic marine environments. *Ecology and Evolution* 2013b;**3**:1780–97.

Ting CS, Hsieh C, Sundararaman S *et al.* Cryo-electron tomography reveals the comparative three-dimensional architecture of *Prochlorococcus*, a globally important marine cyanobacterium. *J Bacteriol* 2007;**189**:4485–93.

Villar E, Vannier T, Vernette C *et al.* The ocean gene atlas: exploring the biogeography of plankton genes online. *Nucleic Acids Res* 2018;**46**:289–95.

Wood AM, Everroad RC, Wingard LM. Measuring growth rates in microalgal cultures. *Algal Culturing Technique* 2005;**18**:269–85.

Zinser ER, Johnson ZI, Coe A *et al.* Influence of light and temperature on *Prochlorococcus* ecotype distributions in the Atlantic Ocean. *Limnol Oceanogr* 2007;**52**:2205–20.