**­Transcriptional response of *Prochlorococcus* to co-culture with a marine *Alteromonas*: differences between strains and the involvement of putative infochemicals**

Dikla Aharonovich and Daniel Sher

Department of Marine Biology, Leon H. Charney School of Marine Sciences, University of Haifa, Haifa, Israel.

[dsher@univ.haifa.ac.il](mailto:dsher@univ.haifa.ac.il)

**Supplementary methods**

**Detailed description of co-culture vessels, sample collection and preservation.**

The experiment testing the effect of different *Alteromonas* HOT1A3 cell densities on *Prochlorococcus* growth (Fig. 1) was performed using duplicate 25ml cultures in borosilicate test tubes (2.5cm diameter, 15cm length), and the fluorescence was measured using a Turner 10AU fluorometer. The co-cultures used for transcriptome analyses were grown in a volume of 300ml in 1 liter Nalgene bottles. Samples for flow cytometry and RNA-seq were collected at 4, 20 and 44 hours after the start of the experiment. At each time-point 1ml samples of the cultures were preserved with 0.125% glutaraldehyde (Sigma) for 10 min in the dark and 25ml were filtered on 0.22µm filters (Supor, Pall) and preserved in 1ml RNAlater (Ambion). RNA Filters and flow cytometry samples were kept frozen at −80°C until analysis. For the experiments described in supplementary figure S2D, a four day old co-culture of MIT9313 and HOT1A3 in which the culture fluorescence had declined, as well as a stationary stage culture of HOT1A3 alone, were centrifuged as described above and the supernatant filter-sterilized (0.22m). These spent media were used either as-is (no nutrients added) or as a base for making Pro99. The experiment was performed in duplicate 96 well plates as described in ([Sher et al., 2011](#_ENREF_22)).

**Transcriptome analysis**

The co-cultures between *Prochlorococcus* MIT9313 and MED4 and *Alteromonas* HOT1A3 contained the same initial number of *Prochlorococcus* cells (~3x106 cells/ml) but the number of *Alteromonas* cells varied over an order of magnitude (5x106 cells/ml for the 9313+1A3 and MED4+1A3 co-cultures, and 0.5x106 for the 9313 + diluted 1A3 co-culture), and levels of both cell types changed over the course of the experiment. Additionally, the *Alteromonas* genome is ~2-3 times larger than the *Prochlorococcus* genomes (~4.7Mbp for *Alteromonas* HOT1A3 compared to 1.65Mbp for MED4 and 2.4Mbp for MIT9313). Finally, the physiological condition of both organisms likely changes during the co-culture, potentially affecting the levels of transcription of both rRNA and functional genes. Two challenges face the analysis of the transcriptome of such a co-culture: a) separating the reads originating from each of the interacting organisms, and b) verifying that the changes in relative levels of transcription are robust to differences in the number of reads sequenced from each organism under each condition.

To make sure that the reads from each organism are treated separately, with no “cross contamination”, we performed a pre-screening step in Galaxy ([Goecks et al., 2010](#_ENREF_7)). We trimmed the reads (removing the first 11bp), clipped them to remove Illumina adapters, and removed reads mapping to the rRNA genes of *Prochlorococcus* and *Alteromonas*. We also removed all reads which after clipping were shorter than 20bp. These quality control steps typically removed 25-40% of the reads. We then mapped the remaining reads from each co-culture separately against the genomes of the relevant *Prochlorococcus* strain (MED4 or MIT9313) and against the sequenced and assembled genome of *Alteromonas* HOT1A3, which will be described elsewhere, using Bowtie Galaxy Tool Version 1.1.2 ([Langmead et al., 2009](#_ENREF_13)). We used highly stringent cutoffs, namely no mismatches at all allowed and a seed length of 28. Reads that did not map to either of these organisms, or that mapped to both of them under these conditions were discarded. Notably, essentially all of the reads that mapped to both organisms were from ribosomal or tRNA genes, and we therefore do not discuss changes in the expression of these genes further.

Following the pre-screening step, all of the uniquely-mapping reads were re-mapped to the relevant *Prochlorococcus* genome and analyzed using Rockhopper ([McClure et al., 2013](#_ENREF_16)). The number of uniquely-mapped reads used for each organism is shown in Supplemental Table S1. Due to the reasons discussed above, the number of reads mapped to *Prochlorococcus* (and analyzed in this study) differed widely between conditions. Several studies have shown that the ability to identify differentially expressed genes depends on the number of sequencing reads (read depth), but that increasing sequencing depth may also lead to genes being falsely identified as differentially expressed, e.g. due to contamination from genomic DNA ([Sims et al., 2014](#_ENREF_23), [Haas et al., 2012](#_ENREF_8), [Atallah et al., 2013](#_ENREF_3)). However, to what extent comparing libraries with vastly differing coverage affects robust identification of differential expression is still unclear. Specifically, for metagenome analyses, a common procedure is to normalize the number of available sequences (sub-sample) before comparing samples to remove biases associated with differential coverage (e.g. ([Oh et al., 2014](#_ENREF_18)) but see ([McMurdie and Holmes, 2014](#_ENREF_17))). We decided to use all of the reads from each library, rather than sub-sample to the lowest library, as we expected to obtain better statistical support as well as better spatial representation of transcript start and end points, although these aspects are not discussed further in the current study. To test the validity of this approach, we re-analyzed the comparison of 9313+1A3 to the control axenic 9313 culture, in which there are large differences in the number of reads mapping to MIT9313 (~7.6-9 million reads for the axenic MIT9313 and 146,901-222,393 reads for MIT9313 in the 9313+1A3 co-culture). We compared the analysis using all of the reads from all biological replicates with two independent analyses where all libraries were randomly sub-sampled to 100,000 reads. As shown in Supplemental Figure 1A and B, the major patterns of differential gene expression were evident in both sub-sampled analyses, including the higher number of over-expressed (relative to under-expressed) genes and the spatial organization along the genome. About 50% of the genes differentially expressed when all reads were analyzed were also differentially expressed when reads were sub-sampled (Sup Fig. 1C), a within the range observed when different analysis platforms are used to analyze the same biological data ([Seyednasrollah et al., 2013](#_ENREF_21), [Soneson and Delorenzi, 2013](#_ENREF_24)). Importantly, all of the CCRG-1 genes and 5/8 CCRG-2 genes that were identified as upregulated in the analysis of all transcripts were also identified when fewer reads were used, attesting to the robustness of our results even for short genes. Finally, the total number of reads used even in the samples with the lowest number of reads was significantly higher than shown by Haas et al ([Haas et al., 2012](#_ENREF_8)) to be sufficient to identify many differentially-expressed genes in transcriptomes of *Vibrio cholerae*, a bacterium with a significantly larger genome than both strains of *Prochlorococcus* . Taken together, these results support our use of the full number of reads for the analyses, as well as the robustness of the differential expression patterns observed. However, we note that the coverage of some of the co-cultures, while sufficient for assessing changes in the level of gene expression, are not sufficient to provide accurate estimates of transcription start and end sites. For this reason we limit, in most cases, our analysis to fold change. All raw reads can be downloaded from the NCBI Sequence Read Archive under accession numbers XXXXX

**Searching for homologs of CCRG-1 and CCRG-2**

To search for homologs of CCRG-1 and CCRG-2 in bacterial genomes, we used the pre-computed trees in microbesonline.org ([Dehal et al., 2009](#_ENREF_6)), with additional homologs identified using tBLASTn agains the ncbi non-redundant dataset (July 2015 release) ([Altschul et al., 1997](#_ENREF_2)). Aiming to identify structural homologs of CCRG-1 and CCRG-2 that could shed light on their potential molecular function`, we used I-TASSER ([Yang et al., 2015](#_ENREF_32)) and Phyre2 ([Kelley et al., 2015](#_ENREF_11)). However, none of the predictions obtained was of high quality and reliability: Phyre2 and I-TASSER did not predict consistent or similar structural homologs for the same CCRG proteins, and the confidence of the predictions that were suggested was typically low (e.g., in I-TASSER, no C-score was above 0.25). This may be a result of the short length of the proteins and, according to the Phyre2 server, the prediction that large parts of the proteins are unstructured. Therefore, we conclude that these proteins are not similar to any known proteins, and further study is needed to elucidate their functions.

**Identification of CCRG-1 and CCRG-2 peptides using tandem mass spectrometry (MS/MS)**

An axenic culture of MIT9313 (1 liter) was grown to mid exponential stage, the cells were collected by centrifugation (5000 g, 10 minutes) and the spent media was filtered through a 0.22m filter (Nalgene) under gentle vacuum to remove residual cells. The spent media was acidified with Trifluoroacetic Acid (TFA, 0.1% final concentration) and dissolved peptides and proteins were extracted by passing it twice through either a Strata X or Strata XL solid phase extraction (SPE) column (resin weight, 100mg, 0.5 liters of spent media for each column). The SPE columns were, eluted with 1 ml of 40% acetonitrile, followed by 1 ml of 80% acetonitrile and the eluents combined and dried by speed-vac. Protein from the pelleted cells were extracted using an urea buffer (9M urea, 400mM ammonium bicarbonate and 10mM DTT, pH 8.0). The resulting protein samples were digested by trypsin, analyzed by LC-MS/MS on a Q-Exactive tandem mass spectrometer (Thermo) and identified using the Discoverer software against a custom database containing the CCRG-1 and CCRG-2 sequences. A 1% false discovery rate was used.

**Identification of CCRG-2 transcripts in environmental samples from the Red Sea**

We searched for sequences encoding CCRG-2 proteins, conservatively defined as reciprocal best BLAST hits (RBBHs([Altenhoff and Dessimoz, 2009](#_ENREF_1))), in a published meta-transcriptomic dataset from samples collected during September 2010 at station A in the Gulf of Aqaba, Red Sea (29°28′N34°55′E) ([Steglich et al., 2015](#_ENREF_27)). At the time of sample collection, the water column was stratified, and *Prochlorococcus* were the dominant phytoplankton at all depths, with ca 10-100 fold fewer *Synechococcus* cells. The reads from two depths, 60m and 100m, were downloaded from the Sequence Read Archive at the ncbi (accession numbers SRR1582030 and SRR1582031, the dataset from 130m was not analyzed as it was processed differently, ([Steglich et al., 2015](#_ENREF_27))). BLASTn searches using all CCRG-2 gene sequences as queries were performed locally with the following parameters: Expect threshold = 10, word size =11, max matches in a query range 0, match/mismatch 2, -3, gap existence 5 extension 2, filter low complexity and mask for lookup table. Reciprocal BLAST was performed using BLASTx against the non-redundant protein database at the ncbi (December 2015 release) using default parameters. Sequences were assembled into contigs using the CAP3 server online (<http://doua.prabi.fr/software/cap3>, ([Huang and Madan, 1999](#_ENREF_9))).

**Multiple sequence alignments**

Multiple sequence alignments were performed using MAFFT with GUIDANCE ([Katoh and Standley, 2013](#_ENREF_10), [Sela et al., 2015](#_ENREF_20)), modified manually in MEGA 6 ([Tamura et al., 2013](#_ENREF_28)) and colored using BoxShade (<http://www.ch.embnet.org/software/BOX_form.html>). Phylogenetic trees were produced using MrBayes 3.2 ([Ronquist and Huelsenbeck, 2003](#_ENREF_19)) using the WAG+I+G model for CCRG-1 proteins, WAG+G model for the CCRG-2 proteins and the 4by4 model for the CCRG-2 nucleotide tree in supplementary figure S4A (protein models were selected by ProtTest 3.0 ([Darriba et al., 2011](#_ENREF_5))). Trees were visualized using FigTree (<http://tree.bio.ed.ac.uk/software/figtree/>) or iTOL ([Letunic and Bork, 2011](#_ENREF_14)).

**Supplementary text**

**Detailed discussion of phycoerythrin-related genes**

In the 9313+diluted 1A3 co-culture, three genes involved in the biosynthesis of pigments associated with phycoerythrin (PE) were differentially regulated, with the expression levels of the dihydrobiliverdin:ferredoxin oxidoreductase PebA decreasing while the expression levels of two genes downstream in the pigment biosynthesis pathway, CpeT and CpeY, increasing ([Wiethaus et al., 2010](#_ENREF_31)). These changes suggest either a shift in the composition of the PE pigments (potentially from phycoerythrolobin or phycourobilon (PEB and PUB) to phycocyanobilin (PCB)), or that the cells are actively reducing the pool of free phycoerythrobilin. PE is typically found in cyanobacterial light harvesting phycobilisomes, yet these structures absent in *Prochlorococcus* and the contribution of PE to light harvesting is minimal ([Steglich et al., 2003](#_ENREF_26)). Thus, the functional role of this change is currently unclear.

**Detailed discussion of changes in expression of key genes involved in nutrient uptake.**

In order to compare the gene expression changes observed in co-culture to those observed in previous studies of nutrient limitation (N-, ([Tolonen et al., 2006](#_ENREF_30)), P-, ([Martiny et al., 2006](#_ENREF_15)) and Fe-limitation, ([Thompson et al., 2011](#_ENREF_29))) we downloaded the expression profiles of selected genes from ProProtal ([Kelly et al., 2011](#_ENREF_12)). As shown in Supplementary figure S2A-C, the transcriptomic responses of the two *Prochlorococcus* strains to co-culture were clearly different from those previously shown to occur under conditions of N, P or Fe starvation. Specifically, most of the key genes previously shown to be up-regulated in response to N and Fe starvation were not differentially expressed in response to co-culture, including those encoding the global nitrogen stress regulator ntcA (which in fact is downregulated in MED4+1A3 and 9313+diluted 1A3), the nitrogen regulatory protein glnB, the transporters for ammonia, urea and formate ([Tolonen et al., 2006](#_ENREF_30)) and the idiA operon and flavodoxin shown to respond to Fe stress ([Thompson et al., 2011](#_ENREF_29))). While the response regulator phoB, a key element in the P starvation response, is upregulated in 9313+diluted 1A3 (see table 1), none of the genes encoding the orthophosphate uptake system (pstABCS operon) are upregulated, nor are other genes suggested to form the core P starvation response such as gap1 and the porin phoE ([Martiny et al., 2006](#_ENREF_15)). Further supporting the difference between stresses induced by co-culture and nutrient starvation, the HLIP-encoding genes differentially expressed under co-culture differ from those whose expression changes following changes in light intensity or quality, or under nitrogen starvation ([Steglich et al., 2006](#_ENREF_25)) ([Berg et al., 2011](#_ENREF_4), [Tolonen et al., 2006](#_ENREF_30)), and only partly overlap the response to iron starvation ([Thompson et al., 2011](#_ENREF_29)). Notably, the results shown in Figures 1, 2 and Supplementary Figure S2D were obtained using cultures grown at different scales (test tubes, 1-liter bottles and 96 well plates, respectively), which have very different surface to volume ratios, suggesting that gas exchange is also not a major factor in determining the outcome of interactions. Taken together, these results rule out nutrient or carbon starvation as major mechanisms underlying the inhibition of MIT9313 strain by high cell densities of *Alteromonas* HOT1A3.

**Supplementary Figure Legends**

**Supplementary figure S1: Comparison of gene expression patterns of 9313+1A3 using all of the uniquely-mapping reads and two independent analyses where each library was sub-sampled to 100,000 reads. A) Synopsis of gene expression results at t=20.** Each dot represents a gene, with colored dots significantly up- or down-regulated. Green dots reveal no significant changes in expression. Note the bias towards up-regulation compared to down-regulation of genes in all three analyses. **B) Genome position and fold change of MIT9313.** Note the clustering of differentially expressed genes in 9313+1A3 in several distinct regions of the genome, which is consistent between all three analyses **C) Venn diagram of differentially expressed genes identified with all of the reads and the sub-sampled analyses,** revealing that about 50% of the genes are identified in all three analyses.

**Supplementary Figure S2: *Prochlorococcus are not nutrient-staved in co-culture.* A-C) Heat-maps of the changes in gene expression during co-culture compared to the expression changes of the same genes during P, N and Fe starvation (**[**Martiny et al., 2006**](#_ENREF_15)**,** [**Thompson et al., 2011**](#_ENREF_29)**)*.*** The genes are ordered by their fold change in the co-culture experiments described in this paper, and are colored based on their fold change, with the highest and lowest log2 fold change shown next to each panel. The x-axis shows the conditions from each published experiment (times after nutrient starvation or alternative nitrogen sources). Only genes deemed to be differentially expressed during co-culture are shown. Note that the changes in gene expression during co-culture were measured using RNA-seq, whereas the changes during nutrient starvation were measured using microarrays. Therefore, qualitative comparisons only can be made. **D) Spent media from pure *Alteromonas* HOT1A3 cultures and from co-cultures in which MIT9313 was inhibited by HOT1A3 does not inhibit MIT9313.** Nutrients were added to some experiments to differentiate between inhibition due to antimicrobial substances (allelopathy) and competition for inorganic nutrients. MIT9313 grew in a manner similar to the controls in all spent media regardless of whether nutrients were added or not, requiring live HOT1A3 cells for inhibition occur. This suggests that competition for nutrients is not responsible for the inhibition of MIT9313 by HOT1A3. It also suggests that if an allelopathic substance produced by HOT1A3 is responsible for the inhibition of MIT9313, this substance is rather unstable. Results shown are means and ranges from duplicate cultures in separate 96 well plates.

**Supplementary Figure S3: Many short genes (encoding products <100 amino acids long) that do not belong to the four defined gene families discussed in the text (HLI, prochlorosin, CCRG-1 and CCRG-2) are differentially expressed in Prochlorococcus MIT9313 in response to co-culture. A) The distribution of predicted protein length of the differentially-expressed (DE) genes in MIT9313 compared to the full predicted proteomes (same as Figure 4A).** Short proteins form almost half of the differentially expressed genes in 9313+1A3 **B) The same distribution calculated without taking into account those encoding HLI, prochlorosin and CCRG-1 and CCRG-2 proteins**, suggesting that short genes in general are an important part of the response of MIT9313 to co-culture. See also Table 1, showing the ten most highly differentially expressed genes in each condition, where short genes are shown in bold and italicized. None of the genes in table 1 belong to the four gene families.

**Supplementary Figure S4: CCRG-2-like genes are expressed in a Prochlorococcus-rich sample from the Red Sea. A) Phylogentic analysis of nucleotide sequences from 53 transcripts from the Red Sea (black text) and CCRG-2 genes from *Prochlorococcus* (blue text) and *Synechococcus* (green text).** The inner ring shows the water depth from which the samples were collected, and the outer ring shows the identify of the reciprocal best BLAST hit (RBBH) against the non-redundant database. The tree was rooted at mid-point, and circles on the branches denote Bayesian posterior probabilities above 0.5. No clear segregation is observed by depth or the identity of the putative producing organism (RBBH). The contig numbers refer to the sequences used to assemble the contigs shown in panel B, below. **B) Multiple sequence alignment of the predicted protein products of CCRG-2 and assembled contigs from the meta-transcriptome.** Only contigs whose predicted core peptide is longer than 15 amino acids are shown. Note the conservation of the region immediately upstream to the putative GG motif and the lack of conservation in the putative core peptide. The names of the sequences from *Prochlorococcus* genomes are colored in blue and from *Synechococcus* in green.

.

**References for Supplementary Information**

Altenhoff, A. M. & Dessimoz, C. 2009. Phylogenetic and Functional Assessment of Orthologs Inference Projects and Methods. *PLoS Computational Biology,* 5**,** e1000262.

Altschul, S., Madden, T., Schaffer, A., Zhang, J., Zhang, Z., Miller, W., et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res,* 25**,** 3389 - 402.

Atallah, J., Plachetzki, D. C., Jasper, W. C. & Johnson, B. R. 2013. The Utility of Shallow RNA-Seq for Documenting Differential Gene Expression in Genes with High and Low Levels of Expression. *PLoS ONE,* 8**,** e84160.

Berg, G., Shrager, J., Van Dijken, G., Mills, M., Arrigo, K. & Grossman, A. 2011. Responses of psbA, hli and ptox genes to changes in irradiance in marine Synechococcus and Prochlorococcus. *Aquatic Microbial Ecology,* 65**,** 1-14.

Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics,* 27**,** 1164-5.

Dehal, P. S., Joachimiak, M. P., Price, M. N., Bates, J. T., Baumohl, J. K., Chivian, D., et al. 2009. MicrobesOnline: an integrated portal for comparative and functional genomics. *Nucleic Acids Research*.

Goecks, J., Nekrutenko, A. & Taylor, J. 2010. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol,* 11**,** R86.

Haas, B. J., Chin, M., Nusbaum, C., Birren, B. W. & Livny, J. 2012. How deep is deep enough for RNA-Seq profiling of bacterial transcriptomes? *BMC Genomics,* 13.

Huang, X. & Madan, A. 1999. CAP3: A DNA Sequence Assembly Program. *Genome Research,* 9**,** 868-877.

Katoh, K. & Standley, D. M. 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution,* 30**,** 772-780.

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protocols,* 10**,** 845-858.

Kelly, L., Huang, K. H., Ding, H. & Chisholm, S. W. 2011. ProPortal: a resource for integrated systems biology of Prochlorococcus and its phage. *Nucleic Acids Research*.

Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol,* 10**,** R25.

Letunic, I. & Bork, P. 2011. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Research,* 39**,** W475-W478.

Martiny, A. C., Coleman, M. L. & Chisholm, S. W. 2006. Phosphate acquisition genes in Prochlorococcus ecotypes: evidence for genome-wide adaptation. *Proc Natl Acad Sci U S A,* 103**,** 12552-7.

Mcclure, R., Balasubramanian, D., Sun, Y., Bobrovskyy, M., Sumby, P., Genco, C. A., et al. 2013. Computational analysis of bacterial RNA-Seq data. *Nucleic Acids Res,* 41**,** e140.

Mcmurdie, P. J. & Holmes, S. 2014. Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. *PLoS Comput Biol,* 10**,** e1003531.

Oh, J., Byrd, A. L., Deming, C., Conlan, S., Program, N. C. S., Kong, H. H., et al. 2014. Biogeography and individuality shape function in the human skin metagenome. *Nature,* 514**,** 59-64.

Ronquist, F. & Huelsenbeck, J. P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics,* 19**,** 1572-1574.

Sela, I., Ashkenazy, H., Katoh, K. & Pupko, T. 2015. GUIDANCE2: accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. *Nucleic Acids Research*.

Seyednasrollah, F., Laiho, A. & Elo, L. L. 2013. Comparison of software packages for detecting differential expression in RNA-seq studies. *Briefings in Bioinformatics*.

Sher, D., Thompson, J. W., Kashtan, N., Croal, L. & Chisholm, S. W. 2011. Response of Prochlorococcus ecotypes to co-culture with diverse marine bacteria. *Isme J,* 5**,** 1125-32.

Sims, D., Sudbery, I., Ilott, N. E., Heger, A. & Ponting, C. P. 2014. Sequencing depth and coverage: key considerations in genomic analyses. *Nat Rev Genet,* 15**,** 121-132.

Soneson, C. & Delorenzi, M. 2013. A comparison of methods for differential expression analysis of RNA-seq data. *BMC Bioinformatics,* 14**,** 91.

Steglich, C., Futschik, M., Rector, T., Steen, R. & Chisholm, S. W. 2006. Genome-wide analysis of light sensing in Prochlorococcus. *J Bacteriol,* 188**,** 7796-806.

Steglich, C., Mullineaux, C. W., Teuchner, K., Hess, W. R. & Lokstein, H. 2003. Photophysical properties of Prochlorococcus marinus SS120 divinyl chlorophylls and phycoerythrin in vitro and in vivo. *FEBS Letters,* 553**,** 79-84.

Steglich, C., Stazic, D., Lott, S. C., Voigt, K., Greengrass, E., Lindell, D., et al. 2015. Dataset for metatranscriptome analysis of Prochlorococcus-rich marine picoplankton communities in the Gulf of Aqaba, Red Sea. *Marine Genomics,* 19**,** 5-7.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution,* 30**,** 2725-2729.

Thompson, A. W., Huang, K., Saito, M. A. & Chisholm, S. W. 2011. Transcriptome response of high- and low-light-adapted Prochlorococcus strains to changing iron availability. *Isme J,* 5**,** 1580-94.

Tolonen, A. C., Aach, J., Lindell, D., Johnson, Z. I., Rector, T., Steen, R., et al. 2006. Global gene expression of Prochlorococcus ecotypes in response to changes in nitrogen availability. *Mol Syst Biol,* 2**,** 53.

Wiethaus, J., Busch, A. W. U., Dammeyer, T. & Frankenberg-Dinkel, N. 2010. Phycobiliproteins in Prochlorococcus marinus: Biosynthesis of pigments and their assembly into proteins. *European Journal of Cell Biology,* 89**,** 1005-1010.

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J. & Zhang, Y. 2015. The I-TASSER Suite: protein structure and function prediction. *Nat Meth,* 12**,** 7-8.