A novel oceanic uncultured temperate cyanophage lineage

José Flores-Uribe[1], Alon Philosof[1], Itai Sharon[2,3], and Oded Béjà[1]*

[1]Faculty of Biology, Technion - Israel Institute of Technology, Haifa 32000, Israel; [2]Migal Galilee Research Institute, Kiryat Shmona, 11016, Israel; [3]Tel Hai College, Upper Galilee 12210, Israel.

*Corresponding author: beja@technion.ac.il

Abstract

Marine cyanobacteria are important contributors to primary production in the ocean and their viruses (cyanophages) affect the ocean microbial communities. Despite reports of lysogeny in marine cyanobacteria, isolation of lysogenic marine cyanophages has not been reported yet. Using assemblies from Red Sea and *Tara* Oceans metagenomes, we recovered genomes of a novel uncultured marine cyanophage lineage, which contain genes coding for the phycobilisome degradation protein NblA, an integrase, and a split DNA polymerase. The DNA polymerase forms a monophyletic clade with a DNA polymerase from a putative prophage in *Synechococcus* WH8016. The putative prophage does not resemble any known cyanophage but shares several genes with the newly identified cyanophages. Metagenomic recruitment indicates that the novel cyanophages are widespread, albeit at low abundance. This is the first report of a putative prophage in a cultured marine *Synechococcus* and the genomic characterization of its cyanophage relatives.

Introduction

Viruses are the most abundant biological entities in the planet (Breitbart, 2012; Paez-Espino et al., 2016). Cyanophages, viruses infecting cyanobacteria, regulate cyanobacterial communities and influence the global nutrient cycles. Viruses also represent major genetic reservoirs and rewire their hosts metabolism (Breitbart, 2012).

Viral infections can be lytic or lysogenic. Lytic infections finish with host lysis and virions release, while in lysogenic infections the viral genomes integrate into the host chromosomes as prophages and replicate without virion production (Howard-Varona et al., 2017). Although there are no reports of prophages in cultured cyanobacteria induction of prophages from marine cyanobacterial environmental populations has been studied (McDaniel et al., 2002; Ortmann et al., 2002). Genomic analysis of *Prochloroccoccus* phage P-SSP7 suggests that the virus may reproduce in a lysogenic way (Sullivan et al., 2005), additionally a partial P-SSP7-like prophage sequence was identified during a study of environmental *Prochlorococcus* communities from the Pacific Ocean (Malmstrom et al., 2013).

We previously screened Bacterial Artificial Chromosome (BAC) libraries for photosystem II (PSII) genes. While some BACs originated from cyanobacteria, other resembled known cyanophage genomes (Zeidner et al., 2005). However, one clone from the Mediterranean Sea (BAC21E04) was notable as it contained numerous ORFs with weak or no homology to anything reported in GenBank, including both non-redundant and environmental-non-redundant entries. BAC21E04 did contain a full length viral-like D1 gene, a partial-length viral-like *talC* transaldolase gene and a putative ribonucleotide reductase (RNR) class II gene (Zeidner et al., 2005). This precluded the possibility to assign affiliation to this BAC at the time.

Here we report the identification of a novel lineage of cyanophages related to BAC21E04. These cyanophages possess properties from temperate phages and shares synteny with a putative prophage in *Synechococcus* WH8016.

Results and Discussion

To increase our knowledge regarding uncultured cyanophages, we examined metagenomics assemblies from the Red Sea (Philosof et al., 2017) and *Tara* Oceans expedition microbiomes and viromes (Brum et al., 2015; Sunagawa et al., 2015). Four contigs of about 86 kbp representing putative viral genomes related to BAC21E04 were identified (File S1). These contained overlapping terminal regions suggesting that they represent complete, terminally redundant viral Metagenome Assembled Genomes (MAGs) (Table S1). The genomic content of the MAGs was compared to 82 genomes from cyanophages infecting *Synechococcus* and *Prochlorococcus* (Table S2), identifying 14564 proteins clustered into 2376 orthologs. Analysis of the shared orthologs between viruses allowed to delineate four clusters of cyanophages corresponding to the viral families *Myoviridae*, *Podoviridae*, *Siphoviridae* and one composed exclusively by the MAGs (Fig 1a), such arrangement remains even when comparing genomic nucleotide distances (Fig S1). The clustering results suggested that the MAGs presented here belong to a new cyanophage lineage.

Mapping of the *Tara* Oceans metagenomes into the MAGs indicated the presence (>0.01% relative abundance) of these cyanophages in 36 samples from 16 stations with an average of 0.07% (Fig 1c). While a viral fraction (<0.2 μm) sample from the South Atlantic (TARA_076) had the highest abundance of the new cyanophages (0.39%), suggesting the presence of cell-free virions from the new family in the environment, two thirds of the samples where the MAGs were detected correspond to the bacterial fraction (0.2–3.0 μm) probably indicating ongoing infections during sampling (Philosof et al., 2017).

All four MAGs contained a cyanophage-like ribonucleotide reductase (RNR) class II gene (Fig S2), a *talC* transaldolase gene (Fig S2), an integrase gene, and three contained a D1 gene with viral signatures (Sharon et al., 2007). Also, for the first time in marine cyanophages two genes coding for the phycobilisome degradation protein NblA were

detected in each of the MAGs. Functional *nblA* genes can be found in freshwater cyanophages (Gao et al., 2012), however their presence in marine viruses is intriguing as no sequenced marine cyanobacteria carries the gene (Supp Info). Besides of the terminase large subunit (TerL), portal, major capsid protein (MCP), and putative tail proteins identified based on structural predictions, no other known phage structural proteins could be classified supporting the classification of these cyanophages into a new family (Table S3).

The new cyanophages also encode a split DNA polymerase, which formed a new cluster within the T7-like cyanophage DNA polymerase family (Fig 1b). Unexpectedly, a split DNA polymerase from *Synechococcus* WH8016 also clustered with the new phage family. Examination of the vicinity of the DNA polymerase genes in the genome of *Synechococcus* WH8016 revealed the possible existence of a large prophage-associated genomic island of approximately 100 kbp within the genome of *Synechococcus* WH8016 (Fig 2) with distinct nucleotide usage. The putative prophage does not resemble any known cultured marine cyanophage genome reported to date but do shares several genes with the newly identified cyanophage family (split DNA polymerase genes, integrase, phage/plasmid related protein, and different ORFs that are also found on BAC21E04; see Fig 2). The putative *attB* site for WH8016 could be located between a tRNA-Pro located next to the integrase in the borders of the prophage area, however the region does not contain detectable known viral structural protein coding genes.

Genomic evidence for a partial prophage in *Prochlorococcus* cells was recently observed using single cell genomics (an ~8 kbp fragment resembling part of the podophage P-SSP7 genome (Malmstrom et al., 2013)). Based on our observations we suggest that the newly identified marine phages form a new cyanophage family related to the identified putative prophage (~ 100 kbp) detected in *Synechococcus* WH8016.

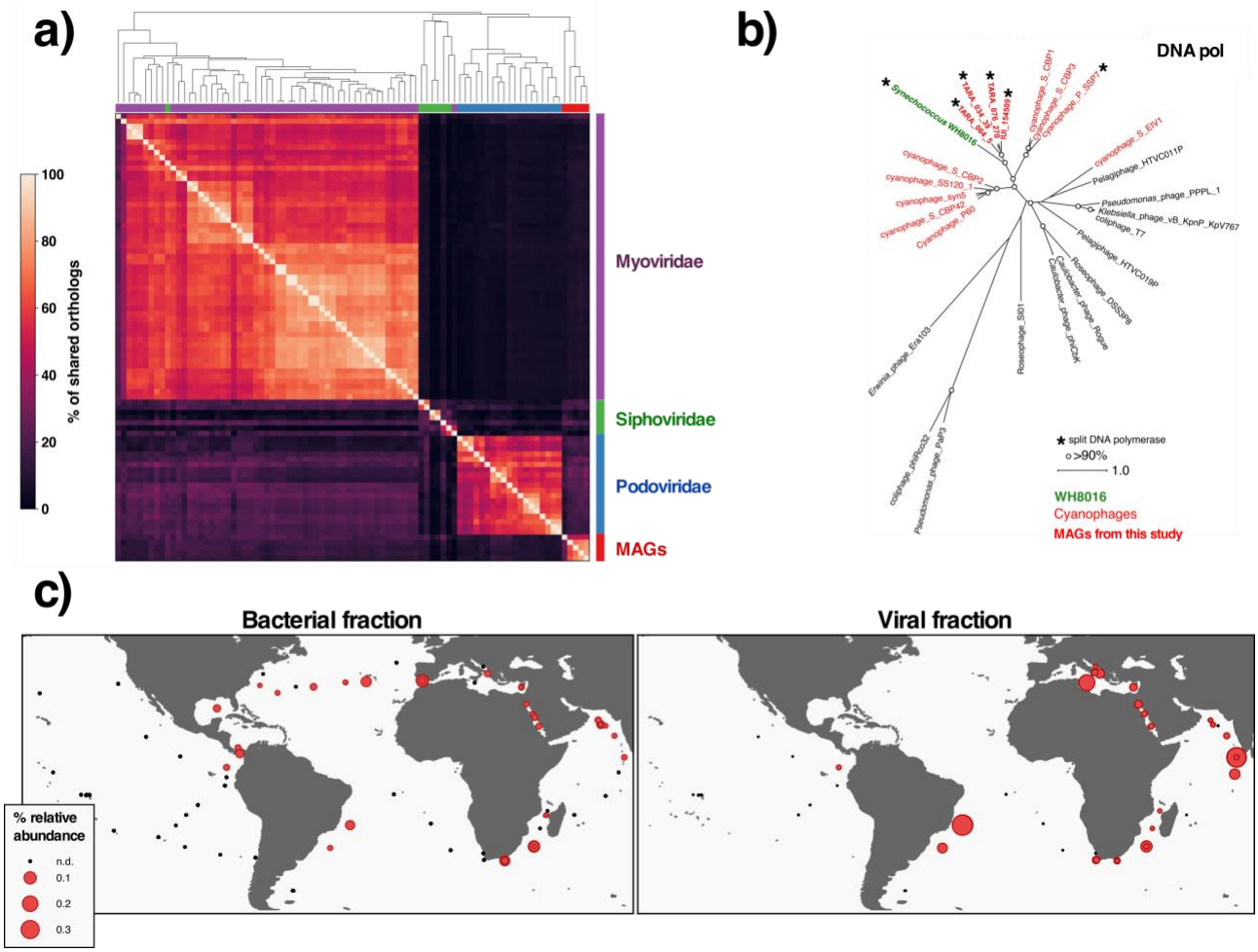

Figure 1. **a**) Comparison of Cyanophages genomes clustering using percentage of shared orthologs. Heat map representation of protein orthologs shared by cyanophages, scaled between 0 (darker) and 100% (lighter). Dendrogram tips are colored according to the taxonomic affiliation of the cyanophages using the following key: *Myoviridae* (purple), *Podoviridae* (blue), *Siphoviridae* (green), and the MAGs identified in this study (red). Clustering identifies four large groups corresponding to the taxonomic families from the cyanophages (Full details in the Supplemental information). **b**) Phylogenetic protein tree of DNA Polymerase. The newly identified MAGs and *Synechococcus* WH 8016 are shown in bold red. Cyanophage and cyanobacteria sequences are shown in red and green, respectively. Split DNA polymerases are marked with an asterisk. The MAGs form a monophyletic clade with a split DNA polymerase from a putative prophage in *Synechococcus* WH8016. **c**) Distribution and relative abundance of the newly identified cyanophages in Tara Oceans metagenomes. Each red circle is scaled to represent the

relative abundance of the MAGs in each station. Black points indicate stations where the MAGs could not be detected (<0.01%).

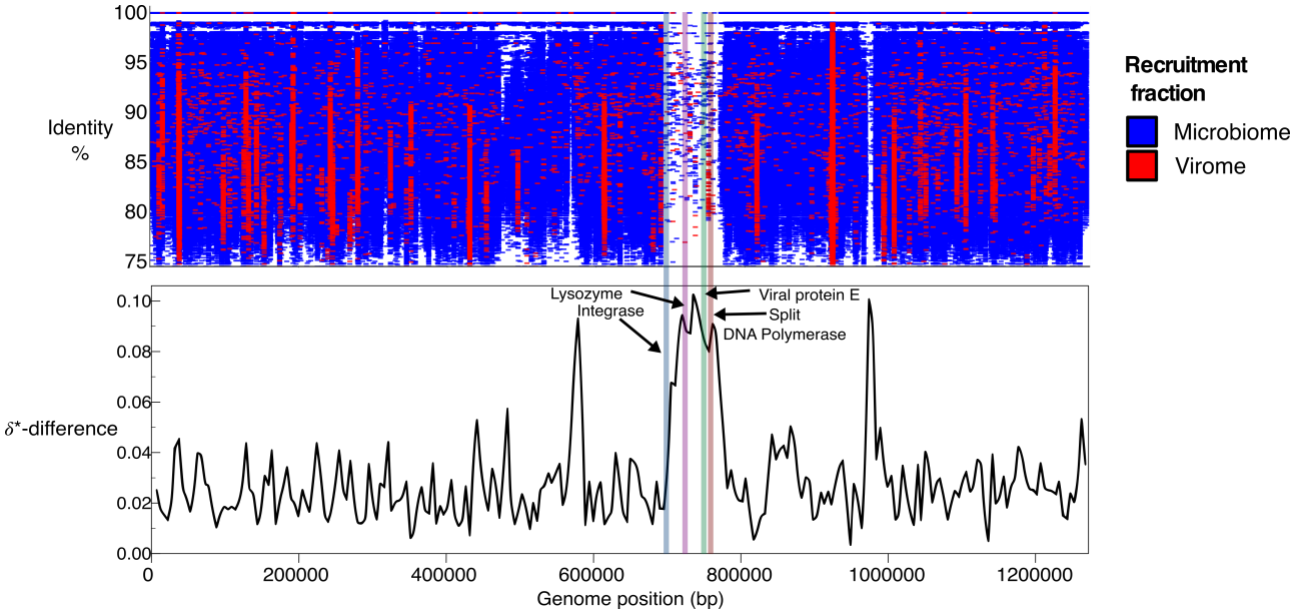

Figure 2. A putative prophage in the genome of *Synechococcus* WH8016. Top: Recruitment of *Tara* Oceans reads from 331 metagenomes onto *Synechococcus* WH8016 genome. Each recruited read is drawn according to the position of the genome where it aligned and the nucleotide identity to the recruitment area. Reads are colored according to the sample fraction of origin, red for virome (<0.2 μm) and blue for microbiome (0.2–3.0 μm). Bottom: Karlin signature (δ*) difference, the difference of the relative abundance of dinucleotides between a sliding window and the whole sequence (Karlin et al., 1998), was calculated using a window size of 10000 bp and a step size of 5000 bp. Colored bars spanning both panels indicate position of shared proteins in synteny with the newly identified cyanophages.

Materials and methods

A detailed methodology can be found in the Supplemental Information.

Data and code availability

The sequence of the MAGs can be found in File S1. The rest of the sequences are deposited at GenBank as follows: BAC21E04 [AY713439-AY713444], *Synechoccocus* WH8016 [AGIK01000001]. Table S2 contains the accessions for the collection of cyanophages. Supporting data can be retrieved from: https://osf.io/q3khc/. Jupyter Notebooks and scripts used are publicly available at https://github.com/BejaLab/TM_Cyanophages.

Author Contributions

J.F.-U. and O.B conceived the project. J.F.-U., A.P., I.S., and O.B. performed bioinformatic analyses. J.F.-U. and O.B. wrote the manuscript with contributions from all authors to data analysis, figure generation, and the final manuscript.

Acknowledgements

We thank Laurence Garczarek for sharing unpublished data for *Synechococcus* WH8016, and Curtis Suttle and Caroline Chénard for sharing DNA polymerase sequences. This work was funded by a European Commission ERC Advanced Grant (no. 321647), and the Louis and Lyra Richmond Memorial Chair in Life Sciences (to O.B.).

References

Breitbart, M. (2012) Marine Viruses: Truth or Dare. *Ann Rev Mar Sci* **4**: 425-448.

160 Brum, J., Ignacio-Espinoza, J.C., Roux, S., Doulcier, G., Acinas, S.G., Alberti, A. et al. (2015) Patterns and ecological drivers of ocean viral communities. *Science* **348**: 1261498.

Gao, E.B., Gui, J.F., and Zhang, Q.Y. (2012) A novel cyanophage with a cyanobacterial nonbleaching protein A gene in the genome. *J Virol* **86**: 236-245.

Howard-Varona, C., Hargreaves, K.R., Abedon, S.T., and Sullivan, M.B. (2017) Lysogeny in
165 nature: mechanisms, impact and ecology of temperate phages. *ISME J* **11**: 1511-1520.

Karlin, S., Campbell, A.M., and Mrazek, J. (1998) Comparative DNA analysis across diverse genomes. *Annu Rev Genet* **32**: 185-225.

Malmstrom, R.R., Rodrigue, S., Huang, K., Kelly, L., Kern, S., Thompson, A. et al. (2013) Ecology of uncultured *Prochlorococcus* clades revealed through single-cell genomics and
170 biogeographic analysis. *ISME J* **7**: 184-198.

McDaniel, L., Houchin, L.A., Williamson, S.J., and Paul, J.H. (2002) Lysogeny in marine *Synechococcus* populations. *Nature* **415**: 496.

Ortmann, A.C., Lawrence, J.E., and Suttle, C.A. (2002) Lysogeny and lytic viral production during a bloom of the cyanobacterium *Synechococcus* spp. *Microb Ecol* **43**: 225-231.

175 Paez-Espino, D., Eloe-Fadrosh, E.A., Pavlopoulos, G.A., Thomas, A.D., Huntemann, M., Mikhailova, N. et al. (2016) Uncovering Earth's Virome. *Nature* **536**: 425-430

Philosof, A., Yutin, N., Flores-Uribe, J., Sharon, I., Koonin, E.V., and Béjà, O. (2017) Novel abundant oceanic viruses of uncultured marine group II euryarchaeota. *Curr Biol* **27**: 1362-1368.

180 Sharon, I., Tzahor, S., Williamson, S., Shmoish, M., Man-Aharonovich, D., Rusch, D.B. et al. (2007) Viral photosynthetic reaction center genes and transcripts in the marine environment. *ISME J* **1**: 492-501.

Sullivan, M.B., Coleman, M.L., Weigele, P., Rohwer, F., and Chisholm, S.W. (2005) Three *Prochlorococcus* cyanophage genomes: Signature features and ecological interpretations.
185 *PLoS Biol* **3**: e144.

Sunagawa, S., Coelho, L.P., Chaffron, S., Kultima, J.R., Labadie, K., Salazar, G. et al. (2015) Structure and function of the global ocean microbiome. *Science* **348**: 1261359.

Zeidner, G., Bielawski, J.P., Shmoish, M., Scanlan, D.J., Sabehi, G., and Béjà, O. (2005) Potential photosynthesis gene recombination between *Prochlorococcus* & *Synechococcus* via viral
190 intermediates. *Environ Microbiol* **7**: 1505-1513.