**Complete genome sequence of the cyanophage S-PRM1 isolated from
Singapore coastal waters**

Caroline Chénard¹, Sandra Kolundžija¹ Federico M. Lauro^{1,2}

¹Asian School of the Environment, Nanyang Technological University, Singapore

²Singapore Centre for Environmental Life Sciences Engineering (SCELSE), Nanyang Technological University, Singapore

Corresponding author:

Federico M. Lauro
50 Nanyang Avenue, N2-01c-54, Singapore 639798
Tel: (65) 6790-9845
Email: flauro@ntu.edu.sg

Abstract

Here, we report the genomic sequence and comparative analysis of the cyanophage S-PRM1 which was recently isolated from Singapore coastal waters using *Synechococcus* sp. strain WH7803. The genome of 144,311 bp was sequenced using a combination of Illumina and Oxford Nanopore technologies. The calculated OrthoANI (Average Nucleotide Identity by Orthology) values of S-PRM1 were between 65.24% and 89.39% when compared with previously sequenced cyanomyoviruses with the highest identity with cyanophage S-CAM9. Auxiliary metabolic genes (AMGs) involved in photosynthesis, carbon metabolism, phosphate acquisition, DNA biosynthesis were found in the genome of S-PRM1.

Introduction

Cyanobacteria of the genera *Synechococcus* and *Prochlorococcus* are abundant and active members of the marine microbial communities. Cyanobacteria contribute to almost half of the photosynthesis in the oceans and as other members of the microbial communities they are targets of viral infection. Many cyanophages, viruses infecting the marine cyanobacteria *Synechococcus* and *Prochlorococcus*, have been sequenced (Chen and Lu, 2002; Mann *et al.*, 2005; Sullivan *et al.*, 2005, 2010; Millard *et al.*, 2009; Labrie *et al.*, 2013, Crummett *et al.*, 2016)). The majority are myoviruses and share similarity with the T4-like myoviruses (Millard *et al.*, 2009; Sullivan *et al.*, 2010). The cyanomyoviruses share a core set of genes including

those involved in virion structure, DNA replication and an array of auxiliary metabolic genes (AMGs) (Sullivan et al., 2010). The AMGs generally encode for proteins of key processes such as photosynthesis and carbon and phosphate metabolism and biosynthesis (Alperovitch-Lavy et al., 2011; Breitbart et al., 2007; Lindell et al., 2004; Millard et al., 2004; Sharon et al., 2009; Sullivan et al., 2006; Thompson et al., 2011). Despite their numerical abundance and fundamental role in cyanobacterial strain dynamics, there is little information about the cyanophages found in tropical coastal waters particularly in highly urbanized areas. Here, we present the genome of the first cyanophage isolated from Singapore coastal waters. S-PRM1 belongs to the *Myoviridae* family was isolated using *Synechococcus* sp. strain WH7803 and sequenced using a combination of Illumina and Oxford Nanopore technologies (Table 1).

Data Description

The cyanophage S-PRM1 was isolated from Singapore north coastal waters (01°22.883N 103°57.05E) using *Synechococcus* sp. strain WH7803 and purified by liquid serial dilution method (Chénard and Chan, 2018; Suttle and Chan, 1993). Once purified, an aliquot of the viral lysate was used to infect 1L of exponentially growing *Synechococcus* culture. After seven days, the lysed culture was filtered twice through a 47mm 0.22µm filter (PVDF, Merck, Germany) and concentrated using Amicon® Ultra 15 with a 30k-Da ultra-PL membrane (Merck, Germany).

The viral nucleic acids were extracted using the QIAMP MinElute Virus Spin Kit (Qiagen, Germany) following the manufacturer protocol. The nucleic acids were eluted into nuclease free water and further purified with the DNA Clean and Concentrator™-5 kit (Zymo Research, USA).

A DNA aliquot was subjected to paired-end sequencing (300bp) with Illumina TruSeq Nano DNA Library kit on a Illumina MiSeq sequencer. The remaining viral DNA was sequenced with Oxford Nanopore technology using a MinION R9.4/MinIONR9.5 flowcell, and 1D Ligation Sequencing Kit (SQK-LKS108) according to the manufacturer protocol.

The Nanopore reads were assembled using Canu v.1.6 (Koren et al. 2016) with the following parameters (CorOutCoverage= "2000", ContigFilter= "5 3000 1.0 1.0 2" and genomeSize= 200kb). The assembly generated a contig of ~137 kb (coverage 156x) which was further polished with Unicycler (Wick et al., 2017) using 340,642 Illumina reads. The final contig (length: 144,311 bp) was used for further analysis.

Open reading frames (ORFs) were predicted using GeneMark (Lukashin and Borodovsky, 1998) and the predicted ORFs were translated and assigned putative functions by comparing them with known protein sequences found in the GenBank (nr database) using the BLASTP program. The ORFs were considered to be homologous to a protein-encoding gene if the e-value was less than 10^{-4} . A total of 200 ORFs were identified and on those 81 % (163 ORFs) had similarity with known protein sequences. Most of them had similarity with known proteins in others cyanophages from the Myoviridae family (~95%). A total of 17 AMGs could be identified in the genome of cyanophage S-PRM1 and included genes involved in photosynthesis (eg. *psbA*, *psbD*, *hli*, *speD*, *petE*, *petF*, *pebS*, *ptoX*), carbon metabolism (*talC*, *zwf*, *gnd* and *CP12*), phosphate acquisition (*phoH*), DNA biosynthesis (*cobS*, *mazG*), heat shock protein (*hsp*) and thioredoxin (Figure 1). Identification of tRNA genes was performed using tRNAscan-SE (Lowe and Eddy, 1997) and demonstrate the presence of eight different tRNA type within the genome (e.g. Arg, Asn, Val, Gly, Met, Leu, Thr and Pro).

To compare the genome of cyanophage S-PRM1 with others cyanomyoviruses, average nucleotide identity were performed using OrthoANI (Average Nucleotide Identity by Orthology) (Lee et al., 2016). The OrthoANI values for all pairwise comparisons within the cyanomyovirus group varied from 62.91% to 89.39%. S-PRM1 showed between 65.24% and 89.39% OrthoANI identity having lower average nucleotide identity with S-CAM7 and the highest identity with S-CAM9. In addition, the gene encoding the major capsid protein (gp23) from cyanophage S-PRM1 was compared phylogenetically with the one from other cyanomyoviruses by aligning the inferred amino-acid sequences with ClustalW using default parameters in Geneious 9.1.6 followed by manually refining (Kearse et al. 2012). A maximum likelihood tree was constructed with RAxML rapid bootstrapping and ML search (100 replicates) (Stamatakis et al., 2008) assuming the James-Taylor Thornton model of substitution using empirical base frequencies and estimating the proportion of invariable sites from the data. The phylogenetic analysis also demonstrated that the cyanophage S-PRM1 is most closely related to cyanophage S-CAM9, isolated from Newport Beach in California (Crummett et al., 2016) and forms a distinct cluster together with S-CRM01, S-CAM7, P-SSM2 and S-SM2 within the cyanomyovirus clade. Interestingly, cyanophage S-PRM1 and S-CAM7 were closely related in term of gp23 but distant in term of OrthoANI. Cyanophage S-CAM7 was generally showing a lower average nucleotide identity with all other cyanophages as a result of the small numbers of AMGs in its genome (Figure 1, Crummett *et al.*, 2016).

Genome sequence accession number.

The complete genome of S-PRM1 has been deposited in GenBank under Bioproject accession PRJNA482903 and nucleotide accession MH629685.

Acknowledgements

The authors acknowledge financial support from Singapore's National Research Foundation under its Marine Science Research and Development Programme (Award No. MSRDP-P13). The authors would like to thank Daniela Drautz-Moses for technical assistance with Illumina sequencing .

References

- Alperovitch-Lavy, A., Sharon, I., Rohwer, F., Aro, E.M., Glaser, F., Milo, R., Nelson, N., Béjà, O., 2011. Reconstructing a puzzle: Existence of cyanophages containing both photosystem-I and photosystem-II gene suites inferred from oceanic metagenomic datasets. *Environ. Microbiol.* 13, 24–32.
<https://doi.org/10.1111/j.1462-2920.2010.02304.x>
- Breitbart, M., Thompson, L., Suttle, C., Sullivan, M., 2007. Exploring the Vast Diversity of Marine Viruses. *Oceanography* 20, 135–139.
<https://doi.org/10.5670/oceanog.2007.58>
- Chen, F., Lu, J., 2002. Genomic sequence and evolution of marine cyanophage P60 : a new insight on lytic and lysogenic phages. *Appl. Environ. Microbiol.* 68, 2589–2594.
- Chénard, C., Chan, A., 2018. Isolation and Purification of Viruses Infecting Cyanobacteria Using a Liquid Bioassay Approach. *Bio-Protocol* 8, 1–9.
<https://doi.org/10.21769/BioProtoc.2691>
- Crummett, L.T., Puxty, R.J., Weihe, C., Marston, M.F., Martiny, J.B.H., 2016. The genomic content and context of auxiliary metabolic genes in marine cyanomyoviruses. *Virology* 499, 219–229.
<https://doi.org/10.1016/j.virol.2016.09.016>
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Mentjies, P., & Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28(12), 1647-1649.
- Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., Phillippy, A.M., 2016. Canu : scalable and accurate long- - read assembly via adaptive k - - mer weighting and repeat separation. *Genome Res.* 27, 722–736.
- Labrie, S.J., Frois-Moniz, K., Osburne, M.S., Kelly, L., Roggensack, S., Sullivan, M.B., Gearin, G., Zeng, Q., Fitzgerald, M., Henn, M.R., Chisholm, S.W., 2013. Genomes of marine cyanopodoviruses reveal multiple origins of diversity. *Environ. Microbiol.* 15, 1356–1376.
- Lee, I., Kim, Y.O., Park, S.C., Chun, J., 2016. OrthoANI: An improved algorithm and software for calculating average nucleotide identity. *Int. J. Syst. Evol. Microbiol.* 66, 1100–1103.

- Lindell, D., Sullivan, M.B., Johnson, Z.I., Tolonen, A.C., Rohwer, F., Chisholm, S.W., 2004. Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. *Proc. Natl. Acad. Sci. U. S. A.* 101, 11013–8.
- Lowe, T.M., Eddy, S.R., 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955–64.
- Lukashin, A. V., Borodovsky, M., 1998. GeneMark . hmm : new solutions for gene finding. *DNA Seq.* 26, 1107–1115.
- Mann, N.H., Clokie, M.R.J., Millard, A., Cook, A., Wilson, W.H., Wheatley, P.J., Letarov, A., Krisch, H.M., 2005. The Genome of S-PM2 , a “ photosynthetic ” T4-type bacteriophage that infects marine *Synechococcus* strains. *J. Bacteriol.* 187, 3188–3200.
- Millard, A., Clokie, M.R.J., Shub, D.A., Mann, N.H., 2004. Genetic organization of the psbAD region in phages infecting marine *Synechococcus* strains. *Proc. Natl. Acad. Sci. U. S. A.* 101, 11007–12.
- Millard, A.D., Zwirgmaier, K., Downey, M.J., Mann, N.H., Scanlan, D.J., 2009. Comparative genomics of marine cyanomyoviruses reveals the widespread occurrence of *Synechococcus* host genes localized to a hyperplastic region : implications for mechanisms of cyanophage evolution. *Environ. Microbiol.* 11, 2370–2387.
- Sharon, I., Alperovitch, A., Rohwer, F., Haynes, M., Glaser, F., Atamna-Ismaeel, N., Pinter, R.Y., Partensky, F., Koonin, E. V., Wolf, Y.I., Nelson, N., Béj, O., 2009. Photosystem I gene cassettes are present in marine virus genomes. *Nature* 461, 258–262.
- Stamatakis, A., Hoover, P., Rougemont, J., 2008. A rapid bootstrap algorithm for the RAXML Web servers. *Syst. Biol.* 57, 758–71.
- Sullivan, M.B., Huang, K.H., Ignacio-Espinoza, J.C., Berlin, A.M., Kelly, L., Weigele, P.R., Defrancesco, A.S., Kern, S.E., Thompson, L.R., Young, S., Yandava, C., Fu, R., Krastins, B., Chase, M., Sarracino, D., Osburne, M.S., Henn, M.R., Chisholm, S.W., 2010. Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments. *Environ. Microbiol.* 12, 3035–3056.
- Sullivan, M.B., Lindell, D., Lee, J.A., Thompson, L.R., Bielawski, J.P., Chisholm, S.W., 2006. Prevalence and evolution of core photosystem II genes in marine cyanobacterial viruses and their hosts. *PLoS Biol.* 4, 1344–1357.
<https://doi.org/10.1371/journal.pbio.0040234>
- Sullivan, M.B., Waterbury, J.B., Chisholm, S.W., 2003. Cyanophages infecting the oceanic cyanobacterium *Prochlorococcus*. *Nature* 424, 1047–51.
- Suttle, C.A., Chan, A.M., 1993. Marine cyanophages infecting oceanic and coastal strains of *Synechococcus*: abundance, morphology, cross-infectivity and growth characteristics. *Mar. Ecol. Prog. Ser.* 92, 99–109.
- Thompson, L.R., Zeng, Q., Kelly, L., Huang, K.H., Singer, A.U., Stubbe, J., Chisholm, S.W., 2011. Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism. *Proc. Natl. Acad. Sci. U. S. A.* 108, E757-64.

Figure 1. A) Maximum likelihood amino-acid tree of major capsid protein of cyanomyoviruses (*gp23*) Bootstrap values are indicated as black (90 to 100%) or grey (75 to 89%) dots at the nodes (100 replicates). B) Diagram showing presence (blue square) or absence (grey square) of the auxiliary metabolic genes in different cyanophage genomes.

Table 1. Classification, general features, and genome sequencing information for cyanophage S-PRM1 according to the MlxS recommendations.

Item	Description
Classification	Domain: unassigned (ds DNA viruses)
	Order <i>Caudovirales</i>
	Family <i>Myoviridae</i>
MlxS data	
Submitted_to_insd	MH629685 (Genbank)
Investigation_type	Virus
Geo_loc_name	Singapore
Lat_lon	(01°22.883N 103°57.05E)
Depth	0.1 m
Collection_date	10-2016
Env_biome	Marine biome (ENVO:00000447)
Env_feature	Coastal water body (ENVO_02000049)
Env_material	Sea water (ENVO_00002150)
Seq_meth	Illumina MiSeq/Nanopore MinION
Number of contigs	1
Assembly method	Canu and Unicycler
Finishing_strategy	Complete; Approx. coverage >150x MinION; >700x MiSeq

213

214

215

216

217