# Comparative Genomic and Phylogenomic Analyses Reveal a Conserved Core Genome Shared by Estuarine and Oceanic Cyanopodoviruses

**Sijun Huang[1]\*, Si Zhang[1], Nianzhi Jiao[2], Feng Chen[3]\***

**1** CAS Key Laboratory of Tropical Marine Bio-resources and Ecology, RNAM Center for Marine Microbiology, South China Sea Institute of Oceanology, Chinese Academy of Sciences, Guangzhou, Guangdong, China, **2** State Key Laboratory of Marine Environmental Science, Xiamen University, Xiamen, Fujian, China, **3** Institute of Marine and Environmental Technology, University of Maryland Center for Environmental Science, Baltimore, Maryland, United States of America

\* chenf@umces.edu (FC); huangsijun@scsio.ac.cn (SH)

## Abstract

Podoviruses are among the major viral groups that infect marine picocyanobacteria *Prochlorococcus* and *Synechococcus*. Here, we reported the genome sequences of five *Synechococcus* podoviruses isolated from the estuarine environment, and performed comparative genomic and phylogenomic analyses based on a total of 20 cyanopodovirus genomes. The genomes of all the known marine cyanopodoviruses are highly syntenic. A pan-genome of 349 clustered orthologous groups was determined, among which 15 were core genes. These core genes make up nearly half of each genome in length, reflecting the high level of genome conservation among this cyanophage type. The whole genome phylogenies based on concatenated core genes and gene content were highly consistent and confirmed the separation of two discrete marine cyanopodovirus clusters MPP-A and MPP-B. The genomes within cluster MPP-B grouped into subclusters mainly corresponding to *Prochlorococcus* or *Synechococcus* host types. Auxiliary metabolic genes tend to occur in a specific phylogenetic group of these cyanopodoviruses. All the MPP-B phages analyzed here encode the photosynthesis gene *psbA*, which are absent in all the MPP-A genomes thus far. Interestingly, all the MPP-B and two MPP-A *Synechococcus* podoviruses encode the thymidylate synthase gene *thyX*, while at the same genome locus all the MPP-B *Prochlorococcus* podoviruses encode the transaldolase gene *talC*. Both genes are hypothesized to have the potential to facilitate the biosynthesis of deoxynucleotide for phage replication. Inheritance of specific functional genes could be important to the evolution and ecological fitness of certain cyanophage genotypes. Our analyses demonstrate that cyanopodoviruses of estuarine and oceanic origins share a conserved core genome and suggest that accessory genes may be related to environmental adaptation.

## Introduction

Viruses are the most abundant biological entities in the ocean, and could affect the population structure and evolution of their hosts [1–3]. Cyanophage are a group of viruses that infect cyanobacteria. They have been recognized as an important biological factor that influences the abundance, diversity and productivity of picocyanobacteria *Synechococcus* and *Prochlorococcus* in the ocean [4–7]. In the past two decades, many cyanophages that infect marine *Synechococcus* and *Prochlorococcus* have been isolated and characterized, and known marine cyanophages are tailed double-stranded DNA viruses, belonging to three well-defined bacteriophage families: *Myoviridae*, *Podoviridae* and *Siphoviridae* [6, 8–16].

Cyanopodoviruses are highly host specific and have been extensively found in various marine habitats [8, 12–14], representing a ubiquitous and ecologically important viral fraction in the ocean. To date, a total of 12 complete cyanopodovirus genomes have been reported [17–20]. According to comparisons based on gene content and genome architecture, all known marine cyanopodoviruses are similar to archetypical coliphage T7, thus denoted as T7-like (the viral genus "T7-like viruses" has been renamed to "*T7likevirus*" from 2012 by International Committee on Taxonomy of Viruses) cyanophages. A few cyanopodovirus-encoded genes were found to be related to metabolic processes of hosts, such as photosynthesis, pentose phosphate pathway, phosphorus acquisition and carbon metabolism [19–22]. Recently, these phage-encoded host-like genes were delineated as auxiliary metabolic genes (AMGs) [23]. One of the AMGs, *psbA*, was shown to be expressed during infection and is thought to be able to confer fitness benefits to cyanophages [24–26].

The DNA polymerase gene (DNA *pol*) was used to investigate the genetic diversity of marine cyanopodoviruses and two marine picocyanobacterial podovirus clusters (MPP-A and MPP-B) were established [13]. This classification was also supported by a recent phylogenomic analysis mainly based on *Prochlorococcus* podoviruses [19]. Using this molecular marker, genetic diversity and temporal and spatial variations of marine cyanopodovirus community were described [14, 27, 28].

Despite the fact that a number of cyanopodovirus genomes have been delineated, however, within them, much fewer genomes were from *Synechococcus* podoviruses (n = 2) than from *Prochlorococcus* podoviruses (n = 10). Especially, only one genome of cyanopodovirus (*Synechococcus* phage P60) isolated from estuarine environment was described. Estuarine ecosystems such as the Chesapeake Bay harbor picocyanobacterial communities which are distinct from those in the open oceans [29, 30]. Although the MPP-B cluster contains the most numerically dominant cyanopodoviruses in the sea [14, 27, 28], no *Synechococcus* podovirus genome in this cluster has been described thus far. Currently, the MPP-A cluster only contains three cyanopodoviruses with known genomes. Therefore, additional genome sequences from *Synechococcus* podoviruses will deepen our understanding on the evolution of picocyanobacterial podoviruses and the relationship between genomes from MPP-A and MPP-B clusters.

In this study, we described four complete genome sequences of podoviruses that were isolated from the Chesapeake Bay using the estuarine *Synechococcus* strains, and one genome of podovirus infecting oceanic *Synechococcus*. Comparative genomic and phylogenomic analyses were performed based on the 20 known cyanopodovirus genomes including 9 *Synechococcus* podoviruses and 11 *Prochlorococcus* podoviruses. We classified the core- and pan-genomes and assessed the phylogenomic relationships among these genomes. Gene content variation among different clusters or subclusters was demonstrated and discussed.

## Materials and Methods

### Phage isolation and DNA extraction and sequencing

Five cyanopodoviruses (S-CBP1, S-CBP2, S-CBP3, S-CBP4 and S-CBP42) isolated from the Chesapeake Bay estuary [13, 27] were selected for genome sequencing. Phage propagation, harvesting and DNA preparation followed the methods described by Wang and Chen [13]. Genomes of S-CBP2, S-CBP3, S-CBP4 and S-CBP42 were sequenced and assembled using the 454 pyrosequencing platform at the Broad Institute [31]. Genome of S-CBP1 was sequenced at Majorbio Biotech (Shanghai, China) using ABI 3730XL DNA Analyzer and assembled using the Phred/Phrap package (http://www.phrap.org).

### Comparative genomics

Programs GeneMark [32] and Glimmer [33] were used to predict the open reading frames (ORFs). Protein sequences of ORFs were input to perform BLASTP comparisons against the NCBI nr protein database and potential functions were then assigned based on best hits. We performed an "all-to-all" BLASTP (-p blastp -W 3 -a 8 -e 0.001 -G 11 -E 1 -F F -U F -M BLO-SUM62) comparison of the 20 cyanopodovirus proteomes (Table 1). Orthologous relationship of any pairwise sequences was assigned when their reciprocal BLASTP hits met the cutoff e-value $\leq$ 1e-5 and alignment length covered at least 50% of the shorter sequence. For short sequences less than 100 amino acids, orthologous relationship was also determined when BLAST identity was $\geq$ 35% even if the e-value was not $\leq$ 1e-5. HMM profiles [34] were built for highly divergent genes (e.g. genes coding for a putative tail fiber and internal capsid proteins) by using HMMBUILD, and the resulting protein databases were searched by using HMMSEARCH and significant similarity was determined when $E$-value was $\leq$ 1e-5. A core gene represents an clustered orthologous group (COG) that is shared by all the 20 cyanopodoviruses. A pan-genome represents all the COGs (including singletons) found in a specific number of genomes. Pan- and core-genomes were plotted as a function of the number of genomes analyzed by using R scripts. Genome maps were created based on the outputs of genome annotations using Canvas v12. T-test was performed by using the SPSS software.

### Whole genome tree and tree comparison

Four methods were implemented to infer phage whole genome trees. *i)* A phylogenetic tree based on the concatenated core genes was built by PAUP* using the distance criterion. A heuristic search with 1000 bootstrap replications was conducted in this analysis. *ii)* The maximum likelihood (ML) trees for each of the core genes were also constructed by RAxML [35, 36] using the JTT protein substitution matrix and the GTRGAMMA+I model to estimate the proportion of invariable sites and the resulting trees were subsequently loaded to the CONSENSE program in PHYLIP package [37] to infer a consensus tree using the extended majority rule. *iii)* A dendrogram was built by SplitsTree4 [38] using ML distance measurement based on gene content. *iv)* Whole genome network was constructed with a ML distance estimator and represented as a neighbor net as implemented by SplitsTree4. For the methods *i)* and *ii)*, Clustal X2 [39] was used to align the sequences and the resulting alignments were trimmed to remove highly divergent regions by the program Gblocks [40]. The topological distances among phylogenetic trees for core genes were calculated based on the symmetric difference as implemented in TREEDIST in PHYLIP. The resulting distance matrix was loaded to PRIMER5 (http://www.primer-e.com/) to assess similarity relationships among phylogenetic trees using non-metric multidimensional scaling (NMDS).

**Table 1. Summary of marine cyanopodovirus genomes.**

| Group | Phage | Accession # | Original host | Host clade | Genome size (bp) | # ORFs | %G +C | Site of origin | Collection date | Depth (m) | Publication |
|---|---|---|---|---|---|---|---|---|---|---|---|
|  | P-RSP2 | HQ332139 | *Prochlorococcus* MIT9302 | HLII | 42257 | 48 | 34.0% | Red Sea | 14-Jul-96 | surface | [19] |
| MPP-A | P60 | AF338467 | *Synechococcus* WH7805 | 5.1-VI | 46675 | 55 | 53.3% | Satilla River Estuary | 16-Jul-88 | surface | [17] |
|  | Syn5 | EF372997 | *Synechococcus* WH8019 | 5.1-II | 46214 | 61 | 55.0% | Sargasso Sea | 30-Nov-86 | surface | [18] |
|  | P-SSP9 | HQ316584 | *Prochlorococcus* SS120 | LLII | 46997 | 54 | 40.5% | BATS | 31-Aug-95 | 100 | [19] |
|  | S-CBP2 | KC310806 | *Synechococcus* CB0208 | 5.2-CB4 | 46237 | 53 | 55.0% | Chesapeake Bay | 27-Sep-02 | surface | This study |
|  | S-CBP42 | KC310805 | *Synechococcus* WH7803 | 5.1-V | 45218 | 57 | 54.6% | Chesapeake Bay | 05-Jun-06 | surface | This study |
| MPP-B | P-SSP7 | AY939843 | *Prochlorococcus* MED4 | HLI | 44970 | 54 | 38.8% | Sargasso Sea | 01-Sep-99 | 100 | [20] |
|  | P-SSP5 | GU071100 | *Prochlorococcus* MIT9515 | HLI | 47055 | 55 | 39.2% | North Pacific gyre | Sep-99 | 120 | unpublished |
|  | P-RSP5 | GU071102 | *Prochlorococcus* NATL1A | LLI | 47741 | 68 | 38.7% | Red Sea | 13-Sep-00 | 130 | [19] |
|  | P-HP1 | GU071104 | *Prochlorococcus* NATL2A | LLI | 47536 | 66 | 39.9% | HOT | 8-Mar-06 | 25 | [19] |
|  | P-SSP2 | GU071107 | *Prochlorococcus* MIT9312 | HLII | 45890 | 59 | 37.9% | BATS | 31-Aug-95 | 120 | [19] |
|  | P-GSP1 | HQ332140 | *Prochlorococcus* MED4 | HLI | 44945 | 53 | 39.6% | Gulf Stream | Aug-95 | 80 | [19] |
|  | P-SSP3 | HQ332137 | *Prochlorococcus* MIT9312 | HLII | 46198 | 56 | 37.9% | BATS | 31-Aug-95 | 100 | [19] |
|  | P-SSP10 | HQ337022 | *Prochlorococcus* NATL2A | LLI | 47325 | 52 | 39.2% | BATS | 5-Jun-96 | 100 | [19] |
|  | P-SSP11 | HQ634152 | *Prochlorococcus* MIT9515 | HLI | 47039 | 54 | 39.2% | BATS | 1-Sep-99 | 100 | [19] |
|  | S-CBP1 | KC310802 | *Synechococcus* CB0101 | 5.2-CB4 | 46547 | 51 | 47.6% | Baltimore Inner Harbor | 16-Jul-02 | surface | This study |
|  | S-CBP3 | KC310803 | *Synechococcus* CB0101 | 5.2-CB4 | 45871 | 55 | 47.0% | Chesapeake Bay | 12-Jul-04 | surface | This study |
|  | S-CBP4 | KC310804 | *Synechococcus* CB0101 | 5.2-CB4 | 44147 | 49 | 44.4% | Chesapeake Bay | 15-Jul-07 | surface | This study |
|  | S-RIP1 | HQ317388 | *Synechococcus* WH8101 | 5.1-VIII | 44892 | 54 | 42.9% | Narragansett Bay | 26-Sep-07 | surface | unpublished |
|  | S-RIP2 | HQ317389 | *Synechococcus* WH7803 | 5.1-V | 45728 | 56 | 47.3% | Rhode Island Sound | 22-Oct-07 | surface | unpublished |

doi:10.1371/journal.pone.0142962.t001

## Phylogenies of the thymidylate synthase gene

Sequences of the thymidylate synthase gene *thyX* were retrieved from cyanobacterial and cyanophage genomes. The protein sequences were aligned using Clustal X2 [39] and ML trees were then built using MEGA6 [41] with the model JTT+GAMMA+I. Bootstrap tests were performed for 100 replicates.

## Nucleotide sequence accession number

The complete genome sequences of cyanopodoviruses S-CBP1, S-CBP2, S-CBP3, S-CBP4 and S-CBP42 have been deposited in the GenBank database under accession numbers KC310802, KC310806, KC310803, KC310804, and KC310805.

## Results and Discussion

### General features of cyanopodovirus genomes

Complete genome sequences of four *Synechococcus* podoviruses (S-CBP1, S-CBP2, S-CBP3, and S-CBP4) which infect Chesapeake Bay *Synechococcus* strains were obtained. S-CBP1, S-CBP3 and S-CBP4 were isolated from the Chesapeake Bay on *Synechococcus* strain CB0101,

while S-CBP2 was isolated from the Bay on *Synechococcus* strain CB0208 [13, 27]. In addition, we also sequenced the genome of S-CBP42, a podovirus which infects oceanic strain *Synechococcus* WH7803 [27] (Table 1). Previously, 12 complete genomes of marine cyanopodoviruses were reported [17–19], and three other genomes have been released in the GenBank (Table 1). Thus, among the 20 cyanopodoviruses with known genome, six (the five described above and *Synechococcus* podovirus P60) were isolated from estuarine waters and the others were from oceanic waters.

In general, marine cyanopodoviruses have a conserved genome size ranging from 42.3 to 47.7 kilo base pair (kbp), which is larger than the size of typical T7-like phages infecting heterotrophic bacteria (37.4 to 39.9 kbp) (data from NCBI GenBank) and freshwater cyanopodoviruses (40.9 to 43.2 kbp) [42, 43]. The *Prochlorococcus* podoviruses have a significant lower G+C content (34–40.5%, Mean = 38.6%, Standard Deviation (SD) = 1.7%, N = 11) than marine *Synechococcus* podoviruses (43–55%, Mean = 49.7%, SD = 4.8%, N = 9) (Table 1) (T-test, $P < 0.01$), reflecting the lower G+C content of *Prochlorococcus* than marine *Synechococcus* [44, 45]. Such a G+C distribution pattern suggests that podoviruses infecting marine *Synechococcus* and *Prochlorococcus* may follow different virus-host co-evolution paths. Generally, the genome sequences of these cyanopodoviruses are highly syntenic (Fig 1, homologous genes were connected by colored lines between genomes), suggesting that those genomes have very similar architectures. The homogeneity in genome organization and the high proportion of core genes (28% by gene number, 50% by genome size, see below) may reflect a constraint which could be an important force for marine cyanopodoviruses to maintain co-evolution with hosts.

## Pan- and core-genomes

A pan-genome of 349 COGs across all the 20 genomes was identified (Fig 2A, S1 File). This added additional 64 COGs into 285 COGs in the pan-genome of 12 marine cyanopodoviruses reported by Labrie and colleagues [19]. The gene accumulation curve was still far from being saturated (Fig 2A), suggesting the existence of vast unexplored genetic diversity of marine cyanopodoviruses. Similarly, the number of genes in the pan-genome of 28 cyanomyoviruses of the *T4likevirus* genus [46] and 12 *Prochlorococcus* [44] also appeared far from reaching a plateau. In contrast, the pan-genome size of *Streptococcus* was saturated for 26 genomes [47]. Pan-genome size depends on the level of genome sequence conservation and the number of genomes sampled. A larger number of cyanopodovirus genomes should be supplemented to estimate the pan-genome size of marine cyanopodoviruses.

Among the total 349 COGs, 15 were core genes that are shared by all the 20 cyanopodovirus genomes. These core genes are involved in virion structure and DNA replication and display remarkable synteny across the 20 genomes (Fig 1). Although an additional seven *Synechococcus* podoviruses were added into the analysis, the number of core genes has not decreased compared to the previous result [19]. It was also shown that the cumulative curve of core genes leveled off when 10 genomes were sampled (Fig 2B). Together, these results indicate that podoviruses infecting marine *Synechococcus* and *Prochlorococcus* share common conserved core genes, so do cyanopodoviruses isolated from brackish and oceanic waters. Our analysis suggests that the core gene set of marine cyanopodoviruses was well determined by known genomes.

Beside the 15 core genes, there were 99 accessory genes (shared by 2–14 genomes), and 235 unique genes (unique to a particular genome). On average, core, accessory and unique genes represented 28, 50, and 22% of total genes in each genome, respectively (Fig 2C). Due to relatively larger gene size of the core genes, they nearly made up 50% of each genome size (Fig 2D). Similarly, core genes make up 57% and 60% of the average genome sizes of marine
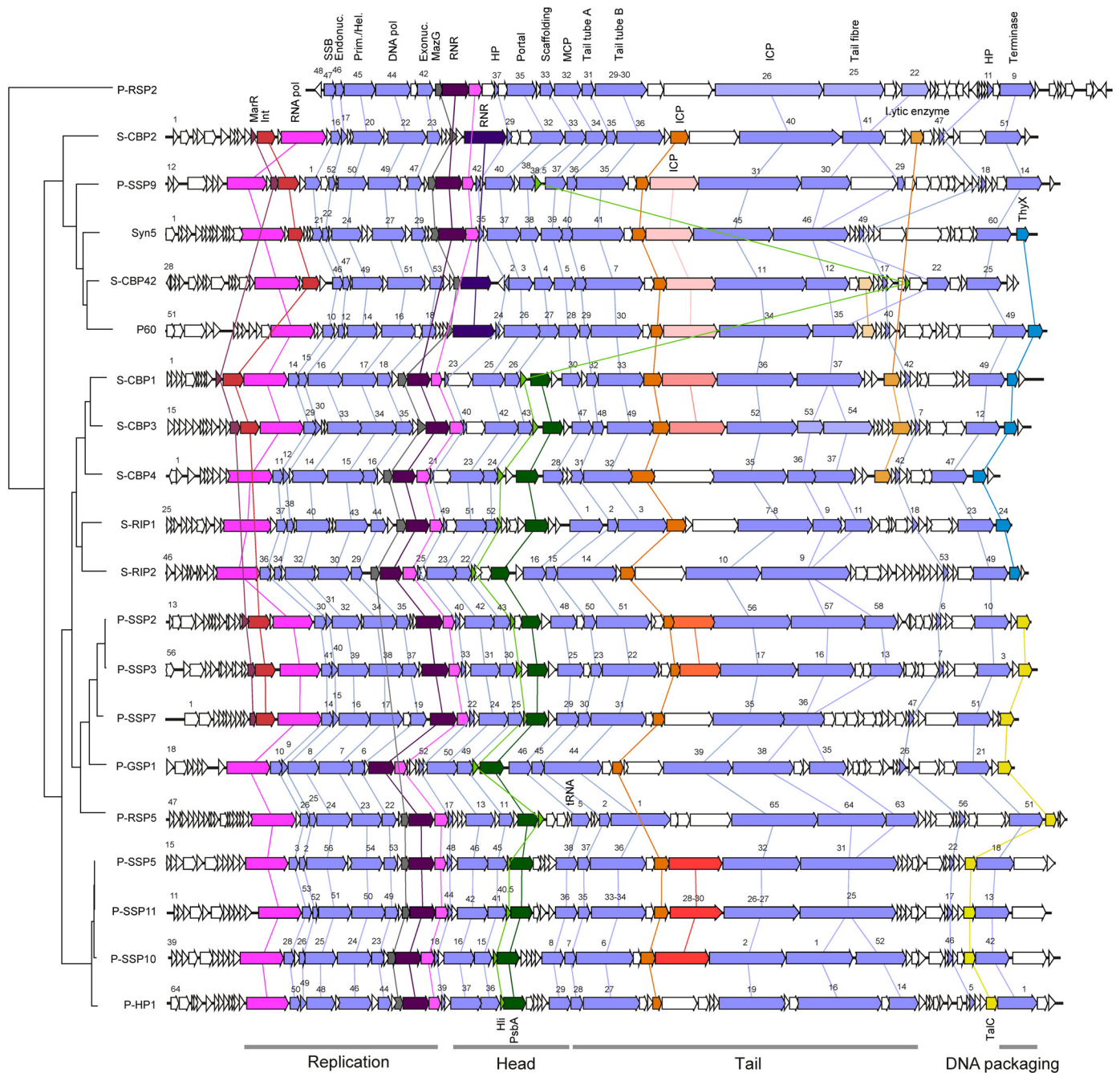
Fig 1. Alignment of the 20 marine cyanopodovirus genomes. Core genes are indicated by light blue arrows. The other arrows that are colored and linked by lines represent a few shared non-core genes with known or putative function. Abbreviation: MarR, MarR family transcriptional regulator; RNA pol, RNA polymerase; SSB, single-stranded DNA binding protein; endonuc., endonuclease; prim./hel., primase/helicase; DNA pol, DNA polymerase; exonuc., exonuclease; MazG, pyrophosphatase; RNR, ribonucleotide reductase; Hli, high light inducible protein; PsbA, photosystem II D1 protein; MCP, major capsid protein; ICP, internal core protein; TalC, transaldolase; ThyX, thymidylate synthase; HP, hypothetical protein.

doi:10.1371/journal.pone.0142962.g001

*Synechococcus* [45] and *Prochlorococcus* [44], respectively. In contrast, core genes only account for 26% of the size of each cyanomyovirus genome [48, 49], on average, while marine cyanosi-phoviruses comprise at least three distinct subtypes which do not share any core genes [50, 51]. The fraction of shared genes between two genomes showed a significant linear correlation to
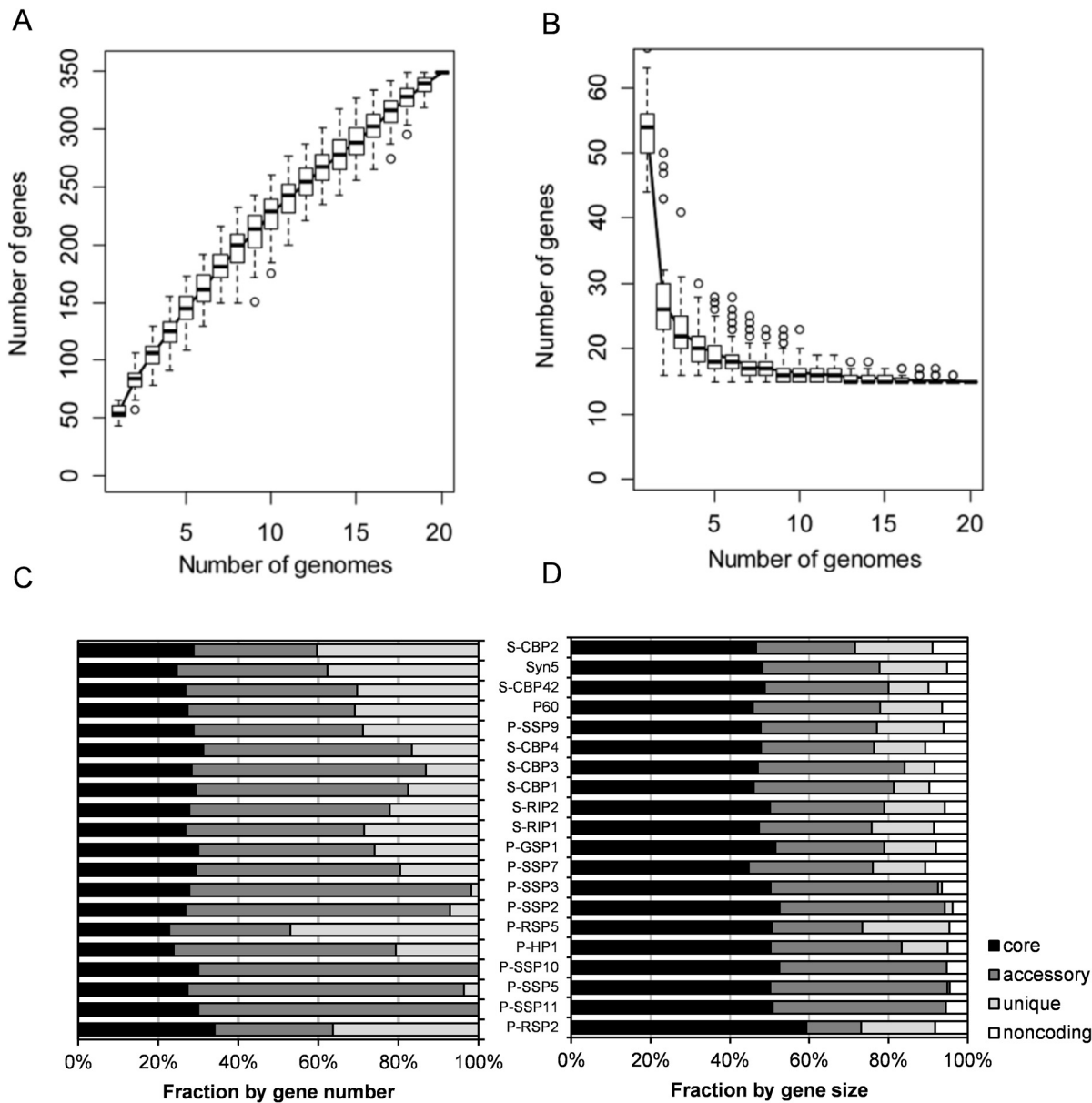
**Fig 2.** A and B. Pan- and core-genomes of cyanopodoviruses. The pan- (A) and core-genomes (B) were plotted as a function of the number of genomes analyzed. The pan-genome is the total number of genes of genomes in a subset sampled, while the core-genome is the genes shared by all genomes in the same subset. The line represents the average and the white box combing with dash lines represents estimated confidence interval. C and D. Fractions of core, accessory and unique genes of each genome.

the average protein sequence identity of core genes between these two genomes (Fig 3A). Such a correlation indicates that the rate of gene gain and loss is positively correlated to the evolution rate of broadly shared genes, and further suggests that the fraction of core genes in a genome reflects the level of genome conservation. Together, our results suggest that known cyanopodovirus genomes are highly conserved among the three cyanophage types, with respect to the core genome proportion.

Interestingly, genes coding for a tail fiber protein, an internal core protein (ICP), the major capsid protein (MCP) and two hypothetical proteins (represented by gp22 and gp47 in
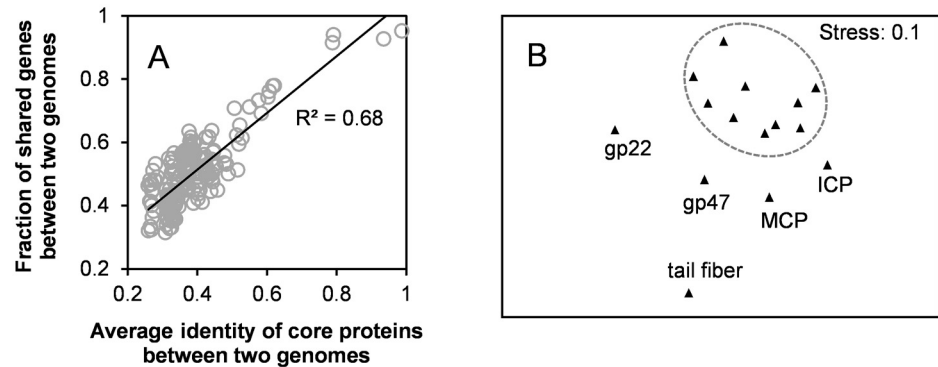
**Fig 3.** A. Linear relationship between the average protein sequence identity of core genes and the fraction of shared genes between two genomes. B. Multidimensional scaling showed the topological distances among the phylogenetic trees for core genes. The dash circle surrounds a relative more conserved core. Abbreviations refer to the legend of Fig 1.

P-SSP7) exhibited phylogenetic incongruence from the other 10 core genes (Fig 3B). It is possible that these five core genes are prone to more frequent genetic exchanges than the other 10 core genes. The genetic change on tail fiber gene may allow phages to adapt to rapidly changing host receptors [19]. In contrast, the *mcp* gene was thought to be among a more conserved gene regime, such as those of myoviruses [52] and cyanomyoviruses [53] of the *T4likevirus* genus. However, it is not clear why the *mcp* genes in cyanopodovirus are less conserved. The *mcp* genes of a few marine viruses have been used as molecular markers to explore the genetic diversity of specific viral groups, such as those of myoviruses [52] and cyanomyoviruses [54] of the *T4likevirus* genus. However, we suggest that the *mcp* gene of marine cyanopodovirus lacks enough conservation to serve as a molecular marker for diversity analysis.

## Whole genome phylogeny

We constructed phylogenies based on core gene alignments of cyanopodoviruses using three approaches (Fig 4A–4C) (see Materials and methods). Overall, significant congruence were observed among the tree constructed based on concatenated sequences of core genes (Fig 4A), the consensus tree of all core gene trees (Fig 4B) and the dendrogram based on gene content (Fig 4C). All these phylogenetic trees divided the 19 of the 20 cyanopodoviruses into two clusters, MPP-A and MPP-B, with the *Prochlorococcus* podovirus P-RSP2 as an outlier. This division agrees with the previous phylogenies built via a single DNA *pol* gene [13, 14] or the concatenated core genes of 12 genomes [19]. Most of MPP-A phages were isolated from *Synechococcus* while MPP-B phages from either *Synechococcus* or *Prochlorococcus* (Fig 4A–4C), in agreement with an observation based on more phage isolates [14].

In cluster MPP-B, *Prochlorococcus* and *Synechococcus* podoviruses were generally separated (Fig 4A–4C). The concatenated core gene phylogenies built by the distance method (Fig 4A) and the maximum likelihood method [55] are highly consistent, and both divided phages into four well supported subclusters (Fig 4A), two of which comprising *Prochlorococcus* podoviruses are identical to the subclusters (MPP-B1 and B2) defined previously [19]. The five *Synechococcus* podoviruses formed two independent subclusters (MPP-B3 and MPP-B4) in the MPP-B cluster (Fig 4A). Subcluster MPP-B3 consisted of three *Synechococcus* podoviruses (S-CBP1, S-CBP3 and S-CBP4) isolated from estuarine waters of the Chesapeake Bay and subcluster MPP-B4 contained two strains isolated from coastal waters (S-RIP1 and S-RIP2) (Fig 4A). The formation of four subclusters is also supported by the gene content dendrogram (Fig 4C).
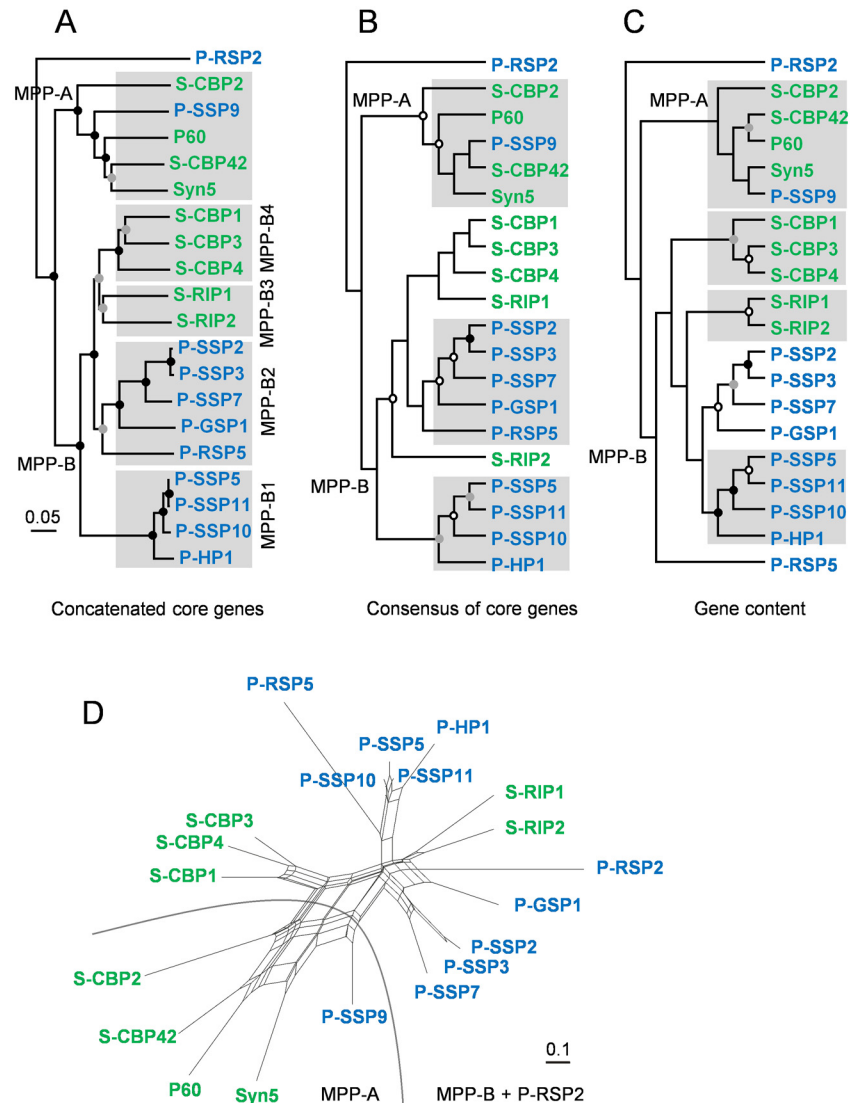
**Fig 4. Whole genome phylogenies and network of cyanopodoviruses.** A, a phylogenetic tree based on the concatenated core genes built by using the distance method; B, a consensus tree inferred from ML trees built for the 15 core genes; C, a dendrogram built by using ML distance measurement based on gene content; D, a whole genome network constructed based on gene content. *Synechococcus* podoviruses were shown in blue and *Prochlorococcus* podoviruses shown in green. Black, grey and open circles respectively represent bootstrap supports of 100%, 75–99% and 50–74%. The grey shading in panel A indicates cluster MPP-A and subclusters MPP-B1, B2, B3 and B4, and those cluster/subclusters that exist in panel B and C are also marked with shading.

However, the consensus tree of core genes ([Fig 4B](#)) shows different clustering within the MPP-B cluster ([Fig 4A and 4C](#)). This is not surprising because at least five out of the 15 core genes have diverged evolutionary trajectories ([Fig 3B](#)).

The separation of clusters MPP-A and MPP-B and the divergence of four subclusters within cluster MPP-B were well supported by phylogenies based on core genes and based on gene content. It appears that the gene content variation resulted from gene gain and loss is significantly constrained by the phylogenetic relationship. This inference is in keeping with the result

shown in Fig 3A. Such a pattern suggests that the horizontal gene transfer between the two cyanopodovirus clusters or among those subclusters is limited.

The relationship among phage isolates in the phylogenetic network constructed based on gene content is similar to those observed in Fig 4A and 4C, with notable exception of the positions of S-RIP1, S-RIP2 and P-RSP2, which are grouped more closely with MPP-B *Prochlorococcus* podoviruses (Fig 4D). Interestingly, in this network, phages S-CBP1, S-CBP3, S-CPB4 and P-SSP9 appear to occupy the intermediate positions connecting MPP-A and MPP-B clusters (Fig 4D). This pattern is corresponding to the observation that certain similarities in presence/absence of accessory genes existed between MPP-B4 *Synechococcus* phages and MPP-A *Synechococcus* phages, as well as between *Prochlorococcus* phage P-SSP9 and MPP-B *Prochlorococcus* phages (Fig 5G and 5H). Despite falling within MPP-A cluster, P-SSP9 still has host-like G+C content that differs greatly from other *Synechococcus* MPP-A phages. In addition, it is noticeable that MPP-B4 phages and three out of five MPP-A phages (S-CBP2, S-CBP42 and P60) were isolated from estuarine waters, while S-RIP1 and S-RIP2 were from coastal waters. It is plausible that such network pattern may be in part related to host population or to the origin of isolating environment.

## Accessory genes

No obvious diagnostic features on the content of accessory genes could be found to distinguish *Prochlorococcus* and *Synechococcus* podoviruses, similar to marine cyanomyoviruses [48] (Fig 5). Moreover, no COGs exclusively obtained by MPP-A phages were observed and only two such COGs (*psbA* and a gene without a known function) existed in these MPP-B phages (Fig 5). Despite this, blocks of genes were indeed enriched (Fig 5A–5F) or lost (Fig 5G and 5H) in some specific phage groups.

Phage AMGs such as those coding for photosystem II D1 protein (PsbA), high light inducible protein (Hli), pyrophosphatase (MazG) and transaldolase (TalC) were found in the accessory gene fraction (Figs 1 and 5). In contrast, among marine cyanomyoviruses, *psbA* and *hli* are within the core set [48, 56]. Certain AMGs likely appear in specific phylogenetic groups. The most striking example is that *psbA* was present in all the MPP-B phages but absent in all the MPP-A phages and the outlier P-RSP2 (Figs 1 and 5). Dekel-Bird and colleagues [14] also reported that all known MPP-A isolates do not encode *psbA* while nearly all MPP-B phages contain *psbA*. *hli* was not only present in all MPP-B phages but also in two MPP-A phages, P-SSP9 and S-CBP42 (Figs 1 and 5). Different from other AMGs that are highly syntenic, *hli* in S-CBP42 is located ~20 kbp downstream from the locus of other *hli* genes (Fig 1). It appears that S-CBP42 lost the *hli* at the common *hli* locus but acquired another one at a downstream site. This agrees with the inference that *hli* could be transferred to phages multiple times [21]. *talC* was only present in *Prochlorococcus* podoviruses in the MPP-B cluster but none of *Synechococcus* podoviruses contained *talC* (Figs 1 and 5). *mazG* was present in all the cyanopodoviruses except the group comprising P-SSP2, P-SSP3, P-SSP7 and P-GSP1 (Figs 1 and 5). This distribution pattern of AMGs in cyanopodoviruses suggests that their acquisitions or losses in specific groups likely occurred around or after the time of divergence of MPP-A and MPP-B clusters.

All the five *Synechococcus* podoviruses in cluster MPP-B and two in MPP-A encode a thymidylate synthase gene, *thyX*, which is located at the right end of each chromosome (Fig 1). Instead of encoding a *thyX*, all the *Prochlorococcus* podoviruses in cluster MPP-B have a *talC* at the same locus (Fig 1). It is likely that one of the two genes was replaced by another. The *thyX* genes are extremely divergent [57]. Thus, it is not surprised that the *thyX* sequences from cyanobacteria and cyanophages fell into three discrete clusters or versions (Fig 6). The sequences from
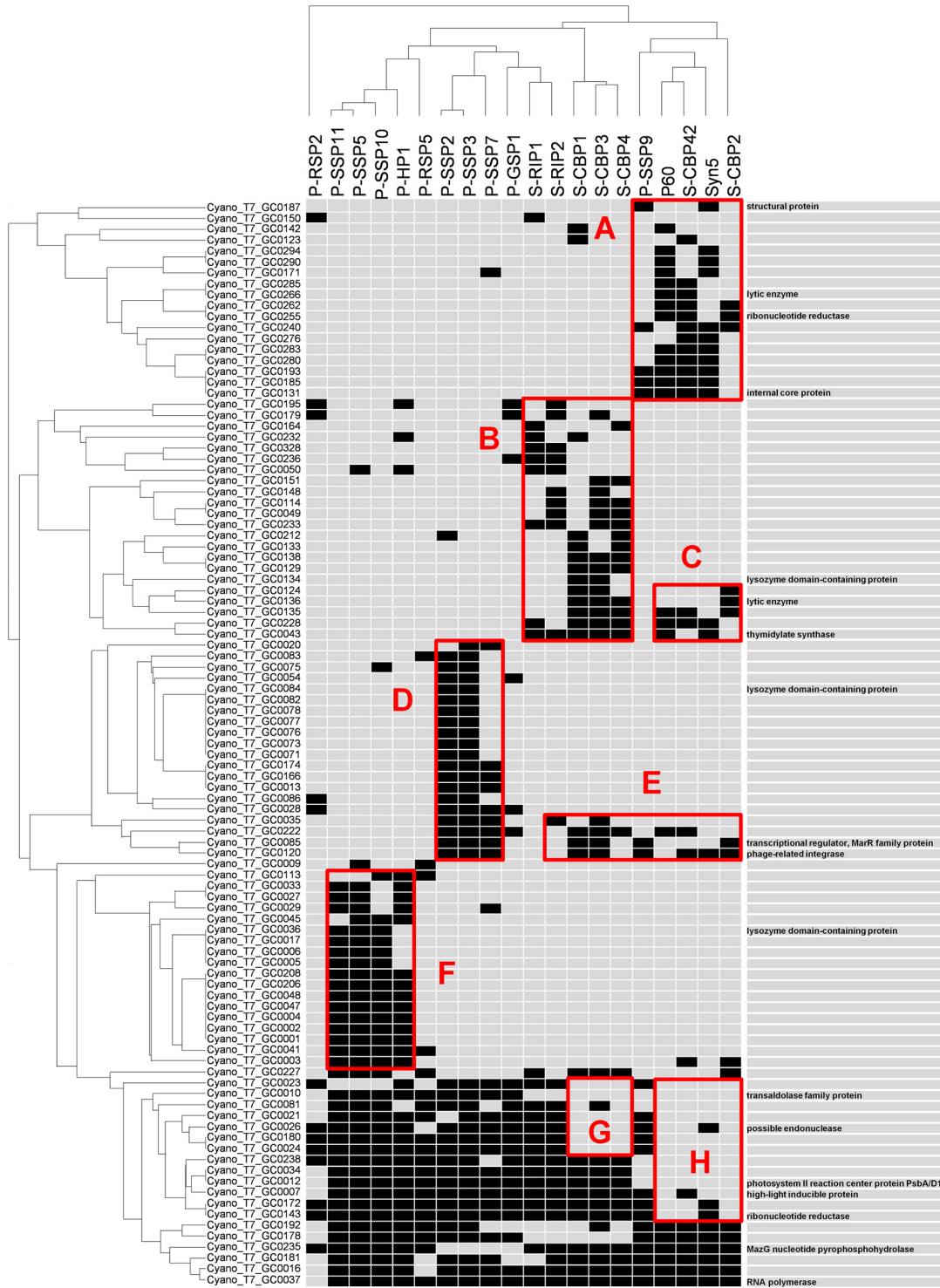
**Fig 5. Distribution pattern of accessory genes (n = 99) among the 20 marine cyanopodovirus genomes.** A black box represents a presence. Cyano_T7_GC stands for T7-like cyanophage gene cluster. The dendrograms were created based on the presence/absence matrix of accessory genes. The UPGMA and WPGMA methods were used to cluster the genes and the phages, respectively. The right column lists those genes with known/putative functions. Red boxes (A-H) indicate genes which were enriched or absent in certain phage groups. All the 349 COGs found among the 20 genomes were listed in S1 File.

doi:10.1371/journal.pone.0142962.g005
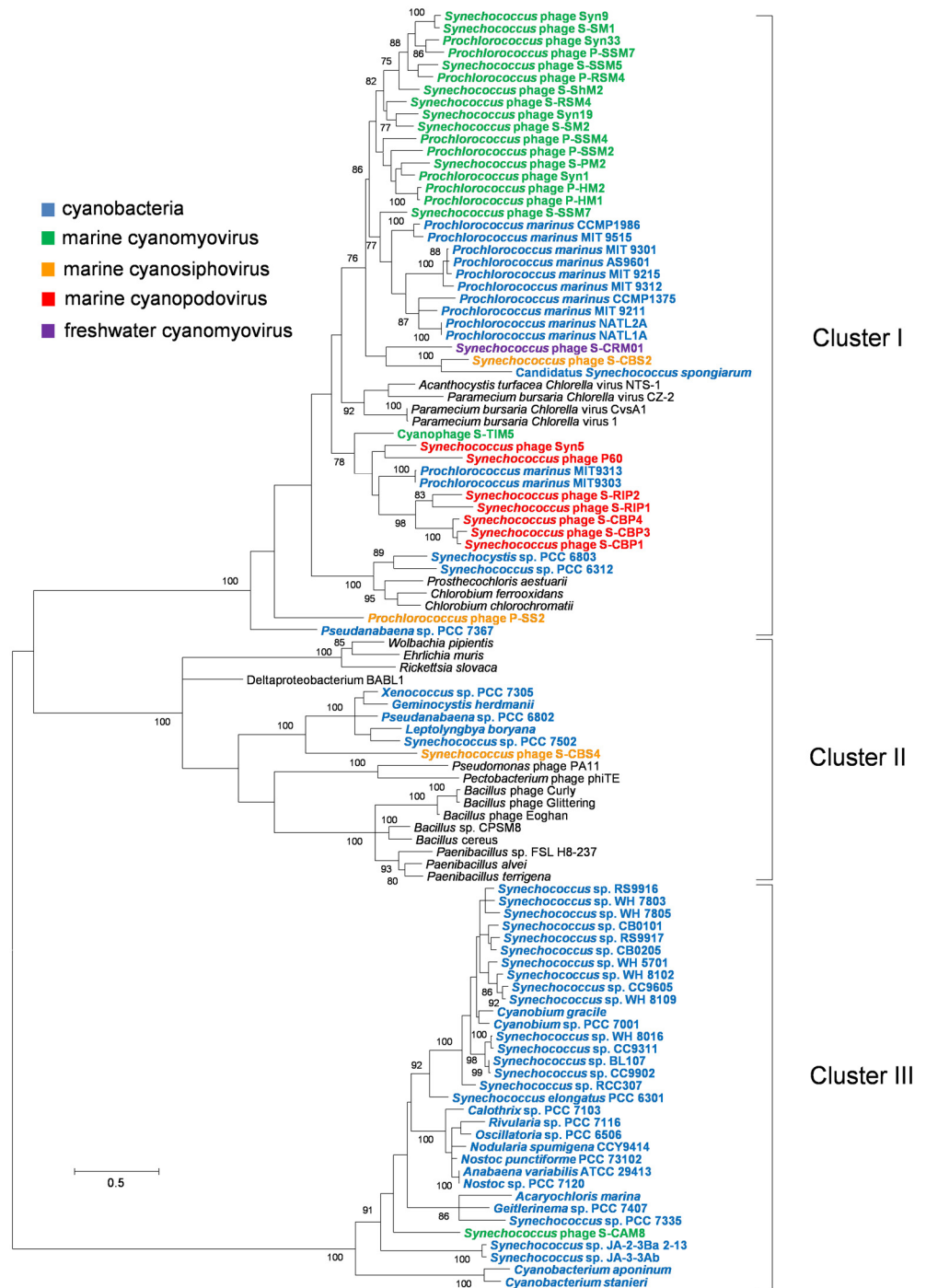
**Fig 6. Maximum likelihood phylogenetic analysis of thymidylate synthase gene *thyX* in cyanophages and cyanobacteria.** Cyanobacterial and cyanophage sequences were shown in color and other bacterial and viral sequences in black. Bootstrap test values higher than 75% were shown.

most of cyanobacteria (cluster III) were grouped together and likely follow a vertical descent in evolution, while those from *Synechococcus* podoviruses and most marine cyanomyoviruses were clustered with *Prochlorococcus* (cluster I). Moreover, two distant subclusters were emerged among these cyanophages and *Prochlorococcus*, one comprising marine cyanomyoviruses and

*Prochlorococcus*, and the other one comprising *Synechococcus* podoviruses, two low-light *Prochlorococcus* and one *Synechococcus* myovirus (Fig 6, cluster I). Together, these phylogenetic patterns strongly support the horizontal transfer of *thyX* between cyanophage and *Prochlorococcus* [53]. Furthermore, none of *talC* or *thyX* was found in T7-like heterotrophic bacteria phages [58] or freshwater T7-like cyanophages [42, 43]. *talC* is a typical bacteria gene and the cyanophage-encoded version is thought to be of bacteria origin [20, 56]. Thus, it is unlikely that cyanopodoviruses inherited *talC* and *thyX* from their T7-like phage ancestor but possibly acquired them elsewhere. ThyX is an alternative type of thymidylate synthase which synthesizes the essential DNA precursor, thymidylate (dTMP), from uridylate (dUMP) [57]. Interestingly, the product of cyanophage *talC* was found to be involved in redirection of host metabolism, which could increase deoxynucleotide biosynthesis [26]. Likely, the two functionally different genes at a same genome locus may lead to similar roles during phage replication. *Prochlorococcus* and *Synechococcus* podoviruses may employ different mechanisms to overcome the shortage of deoxynucleotide.

Cyanophage-encoded host-like genes can be expressed during the infection cycle and are thought to be beneficial to phage fitness [24–26]. The local phosphorus stress could affect the distribution of phosphorus metabolism related genes among cyanomyovirus isolates [48] and communities [46] from different oceans. We also observed different occurrence trends of AMGs between the two cyanopodovirus clusters, that is, MPP-B phages tend to obtain a few AMGs which are absent or only sporadically exist in MPP-A phages. MPP-A and MPP-B cyanopodoviruses likely have differentiation in ecological prevalence as revealed that MPP-B appears to be the dominant cluster in marine habitats [14, 27, 28]. Moreover, the relative abundances of subclusters within MPP-B are highly variable in different environments [55]. For instance, the Chesapeake Bay phages (MPP-B4) were found to be predominant in that estuary but quite rare in coastal and open ocean waters [14, 28, 55]. Such distribution preference of cyanopodovirus genotypes might be closely related to that of their hosts [55], reflecting adaptation to hosts as well as to environment. The cyanopodovirus genomes share a conserved core genome and both the phylogenies of core genes and the presence/absence pattern of non-core genes could distinguish the clusters and subclusters. It is likely that the majority of non-core genes co-evolved with the core, possibly both driven by adaptation to factors such as host and environment.

## Conclusions

Podoviruses which infect marine *Synechococcus* and *Prochlorococcus* share a highly conserved genomic structure, despite differences in host systems and origins of habitat (estuarine or oceanic waters). Core genes make up half of genome length of marine cyanopodoviruses. Our whole genome phylogenetic analyses confirmed the divergence of two discrete clusters of marine cyanopodoviruses, MPP-A and MPP-B. MPP-B phages encode several accessory genes (i.e. *psbA*, *talC* and *thyX*), which can potentially provide phages with selection advantage for inhabiting nutrient poor marine environments. Future studies are needed to explore the role of phage-encoded auxiliary metabolic genes in the ecological distribution of cyanobacterial podoviruses.

## Supporting Information

**S1 File. All the COGs identified based on the 20 cyanopodovirus genomes analyzed in this study.**
(XLSX)

## Acknowledgments

## Author Contributions

Conceived and designed the experiments: SH FC. Performed the experiments: SH. Analyzed the data: SH. Contributed reagents/materials/analysis tools: SZ NJ FC. Wrote the paper: SH SZ NJ FC.

## References

1.  Suttle CA. Viruses in the sea. Nature. 2005; 437(7057):356–61. doi: 10.1038/nature04160 PMID: 16163346.

2.  Suttle CA. Marine viruses—major players in the global ecosystem. Nat Rev Microbiol. 2007; 5(10):801–12. doi: 10.1038/nrmicro1750 WOS:000249525500025. PMID: 17853907

3.  Breitbart M. Marine viruses: truth or dare. Ann Rev Mar Sci. 2012; 4:425–48. doi: 10.1146/annurev-marine-120709-142805 PMID: 22457982.

4.  Suttle CA, Chan AM, Cottrell MT. Infection of phytoplankton by viruses and reduction of primary productivity. Nature. 1990; 347(6292):467–9. doi: 10.1038/347467a0 WOS:A1990EB44600068.

5.  Suttle CA, Chan AM. Dynamics and distribution of cyanophages and their effect on marine *Synechococcus* spp. Appl Environ Microbiol. 1994; 60(9):3167–74. Epub 1994/09/01. PMID: 16349372; PubMed Central PMCID: PMC201785.

6.  Waterbury JB, Valois FW. Resistance to co-occurring phages enables marine *Synechococcus* communities to coexist with cyanophages abundant in seawater. Appl Environ Microbiol. 1993; 59(10):3393–9. Epub 1993/10/01. PMID: 16349072; PubMed Central PMCID: PMC182464.

7.  Wang K, Wommack KE, Chen F. Abundance and distribution of *Synechococcus* spp. and cyanophages in the Chesapeake Bay. Appl Environ Microbiol. 2011; 77(21):7459–68. Epub 2011/08/09. doi: 10.1128/AEM.00267-11 PMID: 21821760; PubMed Central PMCID: PMC3209163.

8.  Suttle CA, Chan AM. Marine cyanophages infecting oceanic and coastal strains of *Synechococcus*: abundance, morphology, cross-infectivity and growth characteristics. Mar Ecol Prog Ser. 1993; 92(1–2):99–109. WOS:A1993KP58100010.

9.  Wilson WH, Joint IR, Carr NG, Mann NH. Isolation and molecular characterization of five marine cyanophages propagated on *Synechococcus* sp. strain WH7803. Appl Environ Microbiol. 1993; 59(11):3736–43. Epub 1993/11/01. PMID: 16349088; PubMed Central PMCID: PMC182525.

10. Lu J, Chen F, Hodson RE. Distribution, isolation, host specificity, and diversity of cyanophages infecting marine *Synechococcus* spp. in river estuaries. Appl Environ Microbiol. 2001; 67(7):3285–90. Epub 2001/06/27. doi: 10.1128/AEM.67.7.3285-3290.2001 PMID: 11425754; PubMed Central PMCID: PMC93013.

11. Marston MF, Sallee JL. Genetic diversity and temporal variation in the cyanophage community infecting marine *Synechococcus* species in Rhode Island's coastal waters. Appl Environ Microbiol. 2003; 69(8):4639–47. Epub 2003/08/07. PMID: 12902252; PubMed Central PMCID: PMC169111.

12. Sullivan MB, Waterbury JB, Chisholm SW. Cyanophages infecting the oceanic cyanobacterium *Prochlorococcus*. Nature. 2003; 424(6952):1047–51. Epub 2003/08/29. doi: 10.1038/nature01929 nature01929 [pii]. PMID: 12944965.

13. Wang K, Chen F. Prevalence of highly host-specific cyanophages in the estuarine environment. Environ Microbiol. 2008; 10(2):300–12. Epub 2007/09/29. doi: 10.1111/j.1462-2920.2007.01452.x PMID: 17900294.

14. Dekel-Bird NP, Avrani S, Sabehi G, Pekarsky I, Marston MF, Kirzner S, et al. Diversity and evolutionary relationships of T7-like podoviruses infecting marine cyanobacteria. Environ Microbiol. 2013; 15(5):1476–91. doi: 10.1111/1462-2920.12103 WOS:000318041800018. PMID: 23461565

15. Fuller NJ, Wilson WH, Joint IR, Mann NH. Occurrence of a sequence in marine cyanophages similar to that of T4 *g20* and its application to PCR-based detection and quantification techniques. Appl Environ

Microbiol. 1998; 64(6):2051–60. Epub 1998/06/03. PMID: 9603813; PubMed Central PMCID: PMC106277.

16. Millard AD, Mann NH. A temporal and spatial investigation of cyanophage abundance in the Gulf of Aqaba, Red Sea. J Mar Biol Assoc UK. 2006; 86(3):507–15. WOS:000237149800008.

17. Chen F, Lu J. Genomic sequence and evolution of marine cyanophage P60: a new insight on lytic and lysogenic phages. Appl Environ Microbiol. 2002; 68(5):2589–94. Epub 2002/04/27. PMID: 11976141; PubMed Central PMCID: PMC127578.

18. Pope WH, Weigele PR, Chang J, Pedulla ML, Ford ME, Houtz JM, et al. Genome sequence, structural proteins, and capsid organization of the cyanophage Syn5: a "horned" bacteriophage of marine Syne-chococcus. J Mol Biol. 2007; 368(4):966–81. doi: 10.1016/j.jmb.2007.02.046 PMID: 17383677; PubMed Central PMCID: PMC2971696.

19. Labrie SJ, Frois-Moniz K, Osburne MS, Kelly L, Roggensack SE, Sullivan MB, et al. Genomes of marine cyanopodoviruses reveal multiple origins of diversity. Environ Microbiol. 2013; 15(5):1356–76. doi: 10.1111/1462-2920.12053 PMID: 23320838.

20. Sullivan MB, Coleman ML, Weigele P, Rohwer F, Chisholm SW. Three Prochlorococcus cyanophage genomes: signature features and ecological interpretations. PLoS Biol. 2005; 3(5):e144. doi: 10.1371/journal.pbio.0030144 PMID: 15828858; PubMed Central PMCID: PMC1079782.

21. Lindell D, Sullivan MB, Johnson ZI, Tolonen AC, Rohwer F, Chisholm SW. Transfer of photosynthesis genes to and from Prochlorococcus viruses. Proc Natl Acad Sci U S A. 2004; 101(30):11013–8. Epub 2004/07/17. doi: 10.1073/pnas.0401526101 0401526101 [pii]. PMID: 15256601; PubMed Central PMCID: PMC503735.

22. Sullivan MB, Lindell D, Lee JA, Thompson LR, Bielawski JP, Chisholm SW. Prevalence and evolution of core photosystem II genes in marine cyanobacterial viruses and their hosts. PLoS Biol. 2006; 4(8): e234. Epub 2006/06/29. 06-PLBI-RA-0262R2 [pii] doi: 10.1371/journal.pbio.0040234 PMID: 16802857; PubMed Central PMCID: PMC1484495.

23. Breitbart M, Thompson LR, Suttle CA, Sullivan MB. Exploring the vast diversity of marine viruses. Oceanography. 2007; 20(2):135–9. WOS:000261638100023.

24. Lindell D, Jaffe JD, Johnson ZI, Church GM, Chisholm SW. Photosynthesis genes in marine viruses yield proteins during host infection. Nature. 2005; 438(7064):86–9. doi: 10.1038/nature04111 PMID: 16222247.

25. Lindell D, Jaffe JD, Coleman ML, Futschik ME, Axmann IM, Rector T, et al. Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. Nature. 2007; 449(7158):83–6. Epub 2007/09/07. nature06130 [pii] doi: 10.1038/nature06130 PMID: 17805294.

26. Thompson LR, Zeng Q, Kelly L, Huang KH, Singer AU, Stubbe J, et al. Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism. Proc Natl Acad Sci U S A. 2011; 108(39):E757–64. doi: 10.1073/pnas.1102164108 PMID: 21844365; PubMed Central PMCID: PMC3182688.

27. Chen F, Wang K, Huang S, Cai H, Zhao M, Jiao N, et al. Diverse and dynamic populations of cyanobac-terial podoviruses in the Chesapeake Bay unveiled through DNA polymerase gene sequences. Environ Microbiol. 2009; 11(11):2884–92. Epub 2009/08/26. doi: 10.1111/j.1462-2920.2009.02033.x PMID: 19703219.

28. Huang S, Wilhelm SW, Jiao N, Chen F. Ubiquitous cyanobacterial podoviruses in the global oceans unveiled through viral DNA polymerase gene sequences. ISME J. 2010; 4(10):1243–51. Epub 2010/05/14. doi: 10.1038/ismej.2010.56 PMID: 20463765.

29. Xu YL, Jiao NZ, Chen F. Novel psychrotolerant picocyanobacteria isolated from Chesapeake Bay in the winter. J Phycol. 2015; 51(4):782–90. doi: 10.1111/jpy.12318 WOS:000359918800016.

30. Chen F, Wang K, Kan J, Suzuki MT, Wommack KE. Diverse and unique picocyanobacteria in Chesa-peake Bay, revealed by 16S-23S rRNA internal transcribed spacer sequences. Appl Environ Microbiol. 2006; 72(3):2239–43. Epub 2006/03/07. doi: 10.1128/AEM.72.3.2239-2243.2006 PMID: 16517680; PubMed Central PMCID: PMC1393199.

31. Henn MR, Sullivan MB, Stange-Thomann N, Osburne MS, Berlin AM, Kelly L, et al. Analysis of high-throughput sequencing and annotation strategies for phage genomes. PLoS One. 2010; 5(2):e9083. Epub 2010/02/09. doi: 10.1371/journal.pone.0009083 PMID: 20140207; PubMed Central PMCID: PMC2816706.

32. Lukashin AV, Borodovsky M. GeneMark.hmm: new solutions for gene finding. Nucleic Acids Res. 1998; 26(4):1107–15. PMID: 9461475; PubMed Central PMCID: PMC147337.

33. Delcher AL, Harmon D, Kasif S, White O, Salzberg SL. Improved microbial gene identification with GLIMMER. Nucleic Acids Res. 1999; 27(23):4636–41. PMID: 10556321; PubMed Central PMCID: PMC148753.

34. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 2011; 39(Web Server issue):W29–37. doi: 10.1093/nar/gkr367 PMID: 21593126; PubMed Central PMCID: PMC3125773.

35. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 2006; 22(21):2688–90. doi: 10.1093/bioinformatics/btl446 PMID: 16928733.

36. Stamatakis A, Hoover P, Rougemont J. A rapid bootstrap algorithm for the RAxML Web servers. Syst Biol. 2008; 57(5):758–71. doi: 10.1080/10635150802429642 PMID: 18853362.

37. Felsenstein J. PHYLIP-phylogeny inference package (version 3.2). Cladistics. 1989; 5:164–6.

38. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol. 2006; 23(2):254–67. doi: 10.1093/molbev/msj030 PMID: 16221896.

39. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. Bioinformatics. 2007; 23(21):2947–8. doi: 10.1093/bioinformatics/btm404 PMID: 17846036.

40. Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst Biol. 2007; 56(4):564–77. doi: 10.1080/10635150701472164 PMID: 17654362.

41. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. 2013; 30(12):2725–9. doi: 10.1093/molbev/mst197 PMID: 24132122; PubMed Central PMCID: PMC3840312.

42. Liu X, Shi M, Kong S, Gao Y, An C. Cyanophage Pf-WMP4, a T7-like phage infecting the freshwater cyanobacterium *Phormidium foveolarum*: complete genome sequence and DNA translocation. Virology. 2007; 366(1):28–39. Epub 2007/05/15. doi: 10.1016/j.virol.2007.04.019 PMID: 17499329.

43. Liu X, Kong S, Shi M, Fu L, Gao Y, An C. Genomic analysis of freshwater cyanophage Pf-WMP3 Infecting cyanobacterium Phormidium foveolarum: the conserved elements for a phage. Microb Ecol. 2008; 56(4):671–80. Epub 2008/04/30. doi: 10.1007/s00248-008-9386-7 PMID: 18443848.

44. Kettler GC, Martiny AC, Huang K, Zucker J, Coleman ML, Rodrigue S, et al. Patterns and implications of gene gain and loss in the evolution of *Prochlorococcus*. PLoS Genet. 2007; 3(12):e231. doi: 10.1371/journal.pgen.0030231 PMID: 18159947; PubMed Central PMCID: PMC2151091.

45. Dufresne A, Ostrowski M, Scanlan DJ, Garczarek L, Mazard S, Palenik BP, et al. Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria. Genome Biol. 2008; 9(5):R90. Epub 2008/05/30. doi: 10.1186/gb-2008-9-5-r90 PMID: 18507822; PubMed Central PMCID: PMC2441476.

46. Kelly L, Ding HM, Huang KH, Osburne MS, Chisholm SW. Genetic diversity in cultured and wild marine cyanomyoviruses reveals phosphorus stress as a strong selective agent. ISME J. 2013; 7(9):1827–41. doi: 10.1038/ismej.2013.58 WOS:000323385600013. PMID: 23657361

47. Lefebure T, Stanhope MJ. Evolution of the core and pan-genome of *Streptococcus*: positive selection, recombination, and genome composition. Genome Biol. 2007; 8(5):R71. doi: 10.1186/gb-2007-8-5-r71 PMID: 17475002; PubMed Central PMCID: PMC1929146.

48. Sullivan MB, Huang KH, Ignacio-Espinoza JC, Berlin AM, Kelly L, Weigele PR, et al. Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments. Environ Microbiol. 2010; 12(11):3035–56. Epub 2010/07/29. doi: 10.1111/j.1462-2920.2010.02280.x PMID: 20662890; PubMed Central PMCID: PMC3037559.

49. Millard AD, Zwirglmaier K, Downey MJ, Mann NH, Scanlan DJ. 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. 2009; 11 (9):2370–87. Epub 2009/06/11. doi: 10.1111/j.1462-2920.2009.01966.x PMID: 19508343.

50. Sullivan MB, Krastins B, Hughes JL, Kelly L, Chase M, Sarracino D, et al. The genome and structural proteome of an ocean siphovirus: a new window into the cyanobacterial 'mobilome'. Environ Microbiol. 2009; 11(11):2935–51. Epub 2009/10/21. doi: 10.1111/j.1462-2920.2009.02081.x PMID: 19840100; PubMed Central PMCID: PMC2784084.

51. Huang S, Wang K, Jiao N, Chen F. Genome sequences of siphoviruses infecting marine *Synechococcus* unveil a diverse cyanophage group and extensive phage-host genetic exchanges. Environ Microbiol. 2012; 14(2):540–58. doi: 10.1111/j.1462-2920.2011.02667.x PMID: 22188618.

52. Filee J, Tetart F, Suttle CA, Krisch HM. Marine T4-type bacteriophages, a ubiquitous component of the dark matter of the biosphere. Proc Natl Acad Sci U S A. 2005; 102(35):12471–6. doi: 10.1073/pnas.0503404102 WOS:000231675900035. PMID: 16116082

53. Ignacio-Espinoza JC, Sullivan MB. Phylogenomics of T4 cyanophages: lateral gene transfer in the 'core' and origins of host genes. Environ Microbiol. 2012; 14(8):2113–26. doi: 10.1111/j.1462-2920.2012.02704.x WOS:000306899900024. PMID: 22348436

54. Marston MF, Taylor S, Sme N, Parsons RJ, Noyes TJ, Martiny JB. Marine cyanophages exhibit local and regional biogeography. Environ Microbiol. 2013; 15(5):1452–63. doi: 10.1111/1462-2920.12062 PMID: 23279166.

55. Huang S, Zhang S, Jiao N, Chen F. Marine cyanophages demonstrate biogeographic patterns throughout the global ocean. Appl Environ Microbiol. 2015; 81(1):441–52. doi: 10.1128/AEM.02483-14 PMID: 25362060; PubMed Central PMCID: PMC4272729.

56. Sabehi G, Shaulov L, Silver DH, Yanai I, Harel A, Lindell D. A novel lineage of myoviruses infecting cyanobacteria is widespread in the oceans. Proc Natl Acad Sci U S A. 2012; 109(6):2037–42. doi: 10.1073/pnas.1115467109 WOS:000299925000051. PMID: 22308387

57. Myllykallio H, Lipowski G, Leduc D, Filee J, Forterre P, Liebl U. An alternative flavin-dependent mechanism for thymidylate synthesis. Science. 2002; 297(5578):105–7. doi: 10.1126/science.1072113 PMID: 12029065.

58. Dunn JJ, Studier FW. Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. J Mol Biol. 1983; 166(4):477–535. PMID: 6864790.