environmental microbiology

Environmental Microbiology (2011) doi:10.1111/j.1462-2920.2011.02667.x

Genome sequences of siphoviruses infecting marine *Synechococcus* unveil a diverse cyanophage group and extensive phage–host genetic exchanges

Sijun Huang,[1,2] Kui Wang,[2†] Nianzhi Jiao[1] and Feng Chen[2*]**

[1]*State Key Laboratory of Marine Environmental Science, Xiamen University, Xiamen 361005, China.*

[2]*Institute of Marine and Environmental Technology, University of Maryland Center for Environmental Science, Baltimore, MD 21202, USA.*

Summary

Investigating the interactions between marine cyanobacteria and their viruses (phages) is important towards understanding the dynamic of ocean's primary productivity. Genome sequencing of marine cyanophages has greatly advanced our understanding about their ecology and evolution. Among 24 reported genomes of cyanophages that infect marine picocyanobacteria, 17 are from cyanomyoviruses and six from cyanopodoviruses, and only one from cyanosiphovirus (*Prochlorococcus* phage P-SS2). Here we present four complete genome sequences of siphoviruses (S-CBS1, S-CBS2, S-CBS3 and S-CBS4) that infect four different marine *Synechococcus* strains. Three distinct subtypes were recognized among the five known marine siphoviruses (including P-SS2) in terms of morphology, genome architecture, gene content and sequence similarity. Our study revealed that cyanosiphoviruses are genetically diverse with polyphyletic origin. No core genes were found across these five cyanosiphovirus genomes, and this is in contrast to the fact that many core genes have been found in cyanomyovirus or cyanopodovirus genomes. Interestingly, genes encoding three structural proteins and a lysozyme of S-CBS1 and S-CBS3 showed homology to a prophage-like genetic element in two freshwater *Synechococcus elongatus* genomes. Re-annotation of the prophage-like genomic region suggests that *S. elongatus* may contain an intact prophage. Cyanosiphovirus genes involved in DNA metabolism and replication share high sequence homology with those in cyanobacteria, and further phylogenetic analysis based on these genes suggests that ancient and selective genetic exchanges occurred, possibly due to past prophage integration. Metagenomic analysis based on the Global Ocean Sampling database showed that cyanosiphoviruses are present in relatively low abundance in the ocean surface water compared to cyanomyoviruses and cyanopodoviruses.

Received 15 December, 2010; accepted 8 November, 2011. For correspondence. *E-mail chenf@umces.edu; Tel. (+1) 410 234 8866; Fax (+86) 410 234 8869. **E-mail jiao@xmu.edu.cn; Tel. (+86) 592 2187869; Fax (+86) 592 2185375. †Present address: Algenol Biofuels, Inc., 16121 Lee Road, Fort Myers, FL, 33928, USA; E-mail kwang@fgcu.edu. S. Huang and K. Wang contributed equally to this work.

Introduction

Synechococcus are a group of unicellular cyanobacteria that are responsible for a significant portion of ocean's primary production (Johnson and Sieburth, 1979; Waterbury *et al*., 1979). They are genetically diverse and widely distributed in various marine habitats (Scanlan and West, 2002). Viruses (phages) that infect marine *Synechococcus* are abundant in seawater and are able to influence the cyanobacterial biomass and community structure in the sea (Suttle, 2000; Mühling *et al*., 2005). *Synechococcus* viruses have been isolated from estuary, coastal waters and open oceans (Suttle and Chan, 1993; Waterbury and Valois, 1993; Wilson *et al*., 1993; Lu *et al*., 2001; Marston and Sallee, 2003; Sullivan *et al*., 2003; Wang and Chen, 2008), and they all belong to three double-stranded DNA tailed phage families based on phage tail morphology (*Myoviridae*, *Podoviridae* and *Siphoviridae*). Myoviruses and podoviruses are the dominant phage types among the cyanophages isolated from marine environments, while siphoviruses have been isolated in much lower frequency (Suttle and Chan, 1993; Waterbury and Valois, 1993; Wilson *et al*., 1993; Lu *et al*., 2001; Marston and Sallee, 2003; Sullivan *et al*., 2003). These three groups of cyanophages are not only distinguishable morphologically, but also exhibit various life cycles, such as broad host range (able to cross-infect) of myoviruses vs. high host specificity of podoviruses and siphoviruses, and shorter latent period of podoviruses than myoviruses and siphoviruses

(Sullivan *et al*., 2003; Wang and Chen, 2008). A great deal of genetic diversity of cyanomyoviruses and cyanopodoviruses have been uncovered by sequencing the genomes of cultivated phages (Chen and Lu, 2002; Sullivan *et al*., 2003; 2005; 2010; Mann *et al*., 2005; Pope *et al*., 2007; Weigele *et al*., 2007; Millard *et al*., 2009). With different gene markers, such as the *g20* gene encoding viral capsid assembly protein for cyanomyovirus, the viral DNA polymerase gene for cyanopodovirus and the core photosystem II *psbA* gene for both of them, vast genetic diversity of cyanobacterial myoviruses and podoviruses were also unveiled in various marine ecosystems (Fuller *et al*., 1998; Zhong *et al*., 2002; Marston and Sallee, 2003; Short and Suttle, 2005; Sullivan *et al*., 2006; 2008; Wilhelm *et al*., 2006; Chenard and Suttle, 2008; Chen *et al*., 2009; Huang *et al*., 2010). However, the genetic diversity of cyanosiphoviruses has not been explored.

To date, genome sequences of 24 cyanophages (17 myoviruses, six podoviruses and one siphovirus) isolated from marine ecosystems have been reported (Chen and Lu, 2002; Mann *et al*., 2005; Sullivan *et al*., 2005; 2009; 2010; Pope *et al*., 2007; Weigele *et al*., 2007; Millard *et al*., 2009; Thompson *et al*., 2011). Among those cyanophages reported, the existing genetic diversity is overshadowed by similarities in morphology and genomic structure that support classification into single podoviral (T7-like) and myoviral (T4-like) groups (Chen and Lu, 2002; Sullivan *et al*., 2005; Pope *et al*., 2007; Millard *et al*., 2009; Sullivan *et al*., 2010; Thompson *et al*., 2011). Cyanomyoviruses have both core genes shared by all of them and the genes interchangeable with hosts (Millard *et al*., 2009; Sullivan *et al*., 2010). Similarly, genomes of cyanopodoviruses contain both conserved genes (i.e. structural genes) which are crucial for their survival and variable regions which allow for flexibility for niche adapting (Sullivan *et al*., 2005; Pope *et al*., 2007; Liu *et al*., 2008). The presence of photosynthesis-related genes in cyanomyoviruses and some cyanopodoviruses has shed light on the ecological relevance of cyanophages that carry such a genomic property (Mann *et al*., 2003; Lindell *et al*., 2004; 2005; 2007; Millard *et al*., 2004; Zeidner *et al*., 2005; Clokie *et al*., 2006; Sullivan *et al*., 2006; Sharon *et al*., 2009; Thompson *et al*., 2011). In addition, genomes of four freshwater cyanophages (two myoviruses and two podoviruses) have been sequenced, among which one myovirus and two podoviruses show little genomic colinearity to the known marine cyanophage genomes (Liu *et al*., 2007; Liu *et al*., 2008; Yoshida *et al*., 2008) while another myovirus has nearly one-third of predicted genes homologous to marine cyanomyoviruses, including photosynthesis genes (Dreher *et al*., 2011).

Siphoviruses account for more than half of publically available bacteriophage genomes (261 out of 501). Currently, only 35 of them have been assigned into nine genera (Lambda-, L5-, c2-, N15-, ΦC31-, SPβ-, T1-, T5- and ΨM1-like) recognized by the International Committee on the Taxonomy of Viruses (9th versioin of ICTV Master Species List, 2009, available on http://www.ictvonline.org/). Phage family *Siphoviridae* often contains temperate members with capability of integrating into host genome, or entering a lysogenic lifestyle. Recently, Sullivan and colleagues (2009) reported the first genome sequence of cyanosiphovirus, P-SS2, an unclassified siphovirus infecting *Prochlorococcus* MIT9313. P-SS2 has a relatively large genome (108 kb) which was distantly related to known lambdoid phages, and contains several genes that may enable this phage to integrate into host genome, forming a lysogenic relationship. Occurrence of lysogeny has been reported in natural *Synechococcus* populations (McDaniel *et al*., 2002; McDaniel and Paul, 2005; Ortmann *et al*., 2002). However, the importance of lysogeny in marine picocyanobacteria is still largely unknown. Among 113 marine bacteria with known genome sequences, nearly 58% of them contain prophages identified by sequence similarity (Paul, 2008). In contrast, none of the 11 marine *Synechococcus* genomes and 12 *Prochlorococcus* genomes published contains identifiable complete prophage gene structures (Kettler *et al*., 2007; Dufresne *et al*., 2008). Indeed, prevalent mobile elements, usually containing a phage-like integrase gene (*int*) inside, in *Synechococcus* and *Prochlorococcus* genomes were thought acquired by phage-mediated horizontal gene transfer (Palenik *et al*., 2003; Coleman *et al*., 2006; Sullivan *et al*., 2009). It is anticipated that genome sequence of cyanobacterial siphoviruses may shed light on their relationship with their hosts. Several siphoviruses have been isolated from marine *Synechococcus* (Suttle and Chan, 1993; Waterbury and Valois, 1993; Wilson *et al*., 1993; Wang and Chen, 2008), but no genome sequence has been reported for the *Synechococcus* siphovirus.

In this study, we reported the genome sequences of four marine *Synechococcus* siphoviruses, S-CBS1, S-CBS2, S-CBS3 and S-CBS4. These phages were isolated from the Chesapeake Bay (Wang and Chen, 2008), infecting four different marine *Synechococcus* strains CB0201, CB0204, CB0202 and CB0101 respectively, which were also isolated from the Chesapeake Bay (Chen *et al*., 2006). Comparative genomics revealed a great genetic diversity among the siphoviruses infecting marine picocyanobacteria. In addition, the metagenomic survey showed that cyanosiphoviruses likely constitute a relatively small fraction of cyanophage community in marine environments compared to cyanomyoviruses and cyanopodoviruses.

Results and discussion

Brief description of Synechococcus *siphoviruses S-CBS1, S-CBS2, S-CBS3 and S-CBS4*

S-CBS2 has an elongated head (~50 × 90 nm) and a long flexible, non-contractile tail of *c.* 170 nm (Table 1, Fig. S1A). S-CBS2 is morphologically similar to but smaller than P-SS2, a siphovirus infecting marine *Prochlorococcus* MIT9313 (Sullivan *et al.*, 2009). P-SS2 also has an elongated head (~75 × 140 nm) and a flexible non-contractile tail (~325 nm). Currently, S-CBS2 and P-SS2 are the only two cyanosiphoviruses with elongated heads. S-CBS1 and S-CBS3 virions (isometric head with size ~55 nm) look quite alike and both have a non-contractile, flexible and relatively shorter tail (*c.* 80 nm) (Table 1, Fig. S1B and C). S-CBS4 virion has an isometric head (~ 72 nm) and a long flexible tail (~200 nm) (Table 1, Fig. S1D). Therefore, morphologically the known cyanosiphoviruses can be divided into three subtypes. As described previously, all the four *Synechococcus* phages are specific to their hosts, and do not cross-infect many marine and freshwater *Synechococcus* strains (Wang and Chen, 2008). The four phages have different latent periods (16 h for S-CBS1, 24 h for the other three), and their burst sizes also vary, which range from *c.* 60 (S-CBS2 and S-CBS4) to *c.* 200 (S-CBS1 and S-CBS3) (Wang and Chen, 2008).

Diverse cyanosiphoviruses

Three subtypes of cyanosiphoviruses. In general, the four *Synechococcus* siphoviruses (S-CBS1 to S-CBS4) and *Prochlorococcus* siphovirus P-SS2 were highly variable in terms of genome size and gene content, reflecting their morphological differences. Nevertheless, significant genomic similarities were observed between certain cyanosiphoviruses within similar morphotype, further suggesting the grouping of three subtypes.

S-CBS2 is most similar to P-SS2 among known phages in terms of gene content, order and sequence homology (Fig. 1). S-CBS2 and P-SS2 represented a unique subtype of cyanosiphoviruses (see 'T*erL phylogeny*' below). The linearly assembled double-stranded DNA (dsDNA) genome of S-CBS2 is 72 332 bp in length, with a G + C content of 55% (Table 1). No tRNA sequence was identified in S-CBS2. Among the total 102 open reading frames (ORFs) predicted for S-CBS2 (Table S1), 42 (41%) were homologous to genes of P-SS2, which accounted for half of S-CBS2 genome (*c.* 38 kb). Interestingly, S-CBS2 has a smaller genome than P-SS2 (72 vs. 108 kb) and most genome reduction of S-CBS2 occurred among the structural genes that were predicted based on sequence homology. First, the total sequence length of structural genes in S-CBS2 is *c.* 30 kb, about half of that

Table 1. Genome features of five siphoviruses infecting marine *Synechococcus* or *Prochlorococcus*.

Cyanosiphovirus strain	Original host	Phage group[a]	Particle feature		Genome features			Number of homologues[b] to known proteins in			
			Capsid size (nm)	Tail length (nm)	Size (kb)	%G + C	#ORFs	GenBank	Non-cyanobacterial phages[c]	Cyanophages[d]	Cyanobacteria[e]
S-CBS1	CB0201	Non-classified	55	80	30.3	58.8	42	28	12	5	6
S-CBS2	CB0204	Non-classified	50 × 90	170	72.3	54.5	102	65	15	56	26
S-CBS3	CB0202	Non-classified	55	80	33.0	60.7	46	30	13	8	9
S-CBS4	CB0101	Non-classified	72	200	69.4	50.8	105	43	16	24	13
P-SS2[f]	MIT9313	Non-classified	75 × 140	325	107.6	52.3	131				

a. These five cyanophages belong to *Siphoviridae* family, *Caudovirales* order.
b. Homologue was defined as matched protein with *E*-value ≤ 0.001.
c. Non-cyanobacterial phages refer to phages that infect non-cyanobacterial bacteria.
d. Cyanophages involved in this analysis included 17 cyanomyoviruses and seven cyanopodoviruses and cyanosiphovirus P-SS2 listed in Table S5. Predicted genes can be homologous to genes both in cyanophages and non-cyanobacterial phages.
e. Cyanobacteria involved in this analysis included 30 strains listed in Table 2.
f. The data of cyanophage P-SS2 are from the previous study of Sullivan and colleagues (2009).

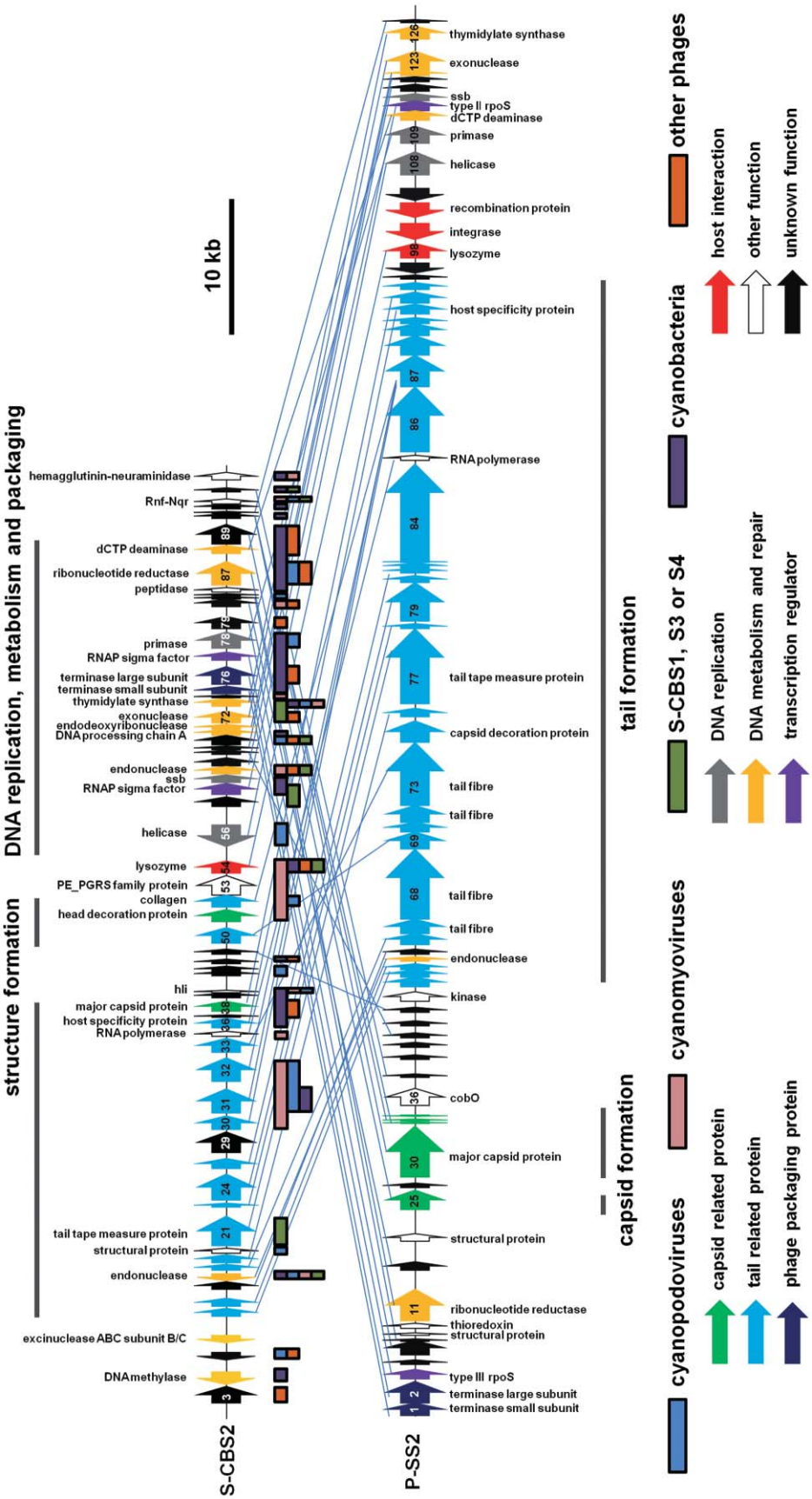

Fig. 1. Genome maps of *Synechococcus* siphovirus S-CBS2 and *Prochlorococcus* siphovirus P-SS2 (Sullivan *et al.*, 2009). Only ORFs with significant BLAST hits (*E*-values ≤ 0.001) in protein databases were shown. Thin blue lines connect the homologous ORFs between S-CBS2 and P-SS2. Functional modules were indicated by gray lines. ORFs that are homologous to those of cyanopodoviruses, cyanomyoviruses, cyanosiphoviruses (S-CBS1, S-CBS3 and S-CBS4), cyanobacteria and other phages were indicated by coloured boxes below the line of arrows. Detailed information of ORFs was provided in Table S1.

in P-SS2 (*c.* 60 kb). Second, S-CBS2 lacks 15 tail structure genes present in P-SS2. Furthermore, S-CBS2 contains genes of reduced size. For example, the possible tail tape measure protein (ORF 21) in S-CBS2 consists of 957 amino acids, whereas there are 1886 amino acids in its homologue in P-SS2 (ORF 77) (Table S1). The tail length of S-CBS2 is 170 nm, about half of the tail length of P-SS2. This is consistent with the observation that there is a close correspondence between the tail tape measure gene size and the phage tail length (Pedulla *et al.*, 2003). In general, S-CBS2 is similar to P-SS2, with reduced genome size and smaller morphological scale.

S-CBS1 and S-CBS3 belong to another subtype of cyanosiphoviruses. The linear dsDNA genomes of S-CBS1 and S-CBS3 have sizes of *c.* 30 and 33 kb, and G + C contents of *c.* 59 and 61% respectively (Table 1). No tRNA sequence was identified in both the genomes. S-CBS1 and S-CBS3 were 62% identical to each other based on nucleotide sequence and shared 29 (nearly 65% of total ORFs for each) homologous ORFs (Tables S2 and S3, Fig. 2), suggesting that they might evolve from a common ancestor. The structural genes on the left arm are conserved between S-CBS1 and S-CBS3, while the 'functional' genes on the right arm are more variable between the two phages (Fig. 2). S-CBS3 contains genes encoding endonuclease, single-strand binding protein (Ssb) and DNA methylase which are all absent in S-CBS1. It appears that the elements responsible for DNA metabolism and replication are vulnerable to genetic exchange, which may provide specific fitness to phages. There are only a few ORFs in S-CBS1 or S-CBS3 homologous to P-SS2 or S-CBS2 (up to five homologues).

S-CBS4 has a genome distinct to the other four cyanosiphoviruses (S-CBS1, S-CBS2, S-CBS3 and P-SS2) (Fig. 3). The linearly assembled genome of S-CBS4 comprises of *c.* 69 kb dsDNA with a G + C content of *c.* 51% (Table 1). A total of 105 ORFs and a tRNA-Thr gene were predicted from the genome. There are little genomic similarities between S-CBS4 and other four cyanosiphoviruses, such as that only 10 homologues in total were found between them (Fig. 3), reflecting its different morphology (isometric head and long tail) described above.

TerL phylogeny. The large terminase subunit (TerL), a protein responsible for phage DNA packaging, is essential for dsDNA phages (see Black, 1989 for a review). Casjens and colleagues (2005) proposed that different functional groups of phage-related terminases can be predicted from their amino acid sequences and that the phylogeny of TerL can resolve different phage groups. The TerL protein-based phylogeny showed that cyanosiphoviruses fell into three distantly phyletic groups (Fig. 4). S-CBS1 and S-CBS3 were most closely related to each other (in the lambda-like group), and S-CBS2 was most closely related to P-SS2. The phylogenetic kinship between S-CBS2 and P-SS2, together with their similar genomic architecture and the amount of homologous genes, suggested that these two siphoviruses with elongated head can also be classified into the same uncharacterized siphovirus subtype proposed by Sullivan and colleagues (2009). S-CBS4 was not closely related to any known cyanophages or other bacteriophages (Fig. 4), and may represent another uncharacterized subtype of *siphoviridae*. Unlike marine T7-like cyanopodoviruses and T4-like cyanomyoviruses (see Table S5, a summary of 29 cyanophage genomes) which appear to be monophyletic among their own groups, cyanosiphoviruses are polyphyletic based on the TerL phylogeny (Fig. 4). In general, the TerL phylogentic clustering of cyanophages not only reflects the relative genetic conservation among three cyanophage families, but also supports the separation of three subtypes of cyanosiphoviruses.

Divergent cyanosiphovirus genomes

Whole genome dot plots showed that there was no continuous colinearity across all the five cyanosiphovirus genomes, with the only observable colinearity between S-CBS1 and S-CBS3 (Fig. S2A). Surprisingly, S-CBS2 and P-SS2 did not exhibit significant genomic conservation on the map (Fig. S2A), despite the fact that they share a bulk of homologues. We consider this is due to the overall low sequence identities between those homologues (most BLASTP *E*-values $> 10^{-30}$, Table S1). This also suggests that the divergence between S-CBS2 and P-SS2 was not a recent event. In order to determine if there is any conservation among cyanosiphovirus genomes not detected by dot-plot mapping, a global searching for shared core genome by all the five cyanosiphoviruses was conducted. However, no such orthologous gene was found (even not all pairwise TerL proteins from cyanosiphoviruses have homology). In contrast, all the 17 cyanomyoviruses share certain genomic colinearity each other (Fig. S2B). It has also been reported that the marine cyanomyovirus genomes are colinear and share a large amount of core genes (approximately 63, one-third of the total ORFs) (Yoshida *et al.*, 2008; Millard *et al.*, 2009; Sullivan *et al.*, 2010). Moreover, among the seven cyanopodovirus genomes, five (P-SSP7, P-SSP5, P-RSP5, P-HP1 and P-SSP2) shared significant genomic colinearity and the other two (P60 and Syn5) showed some scattered genomic homology (Fig. S2C). In addition, 15 core genes were detected among the seven cyanopodovirus genomes, including six phage structure genes, seven genes related to DNA packaging or metabolism and two genes with unknown functions (Table S6). The genomic conservation among

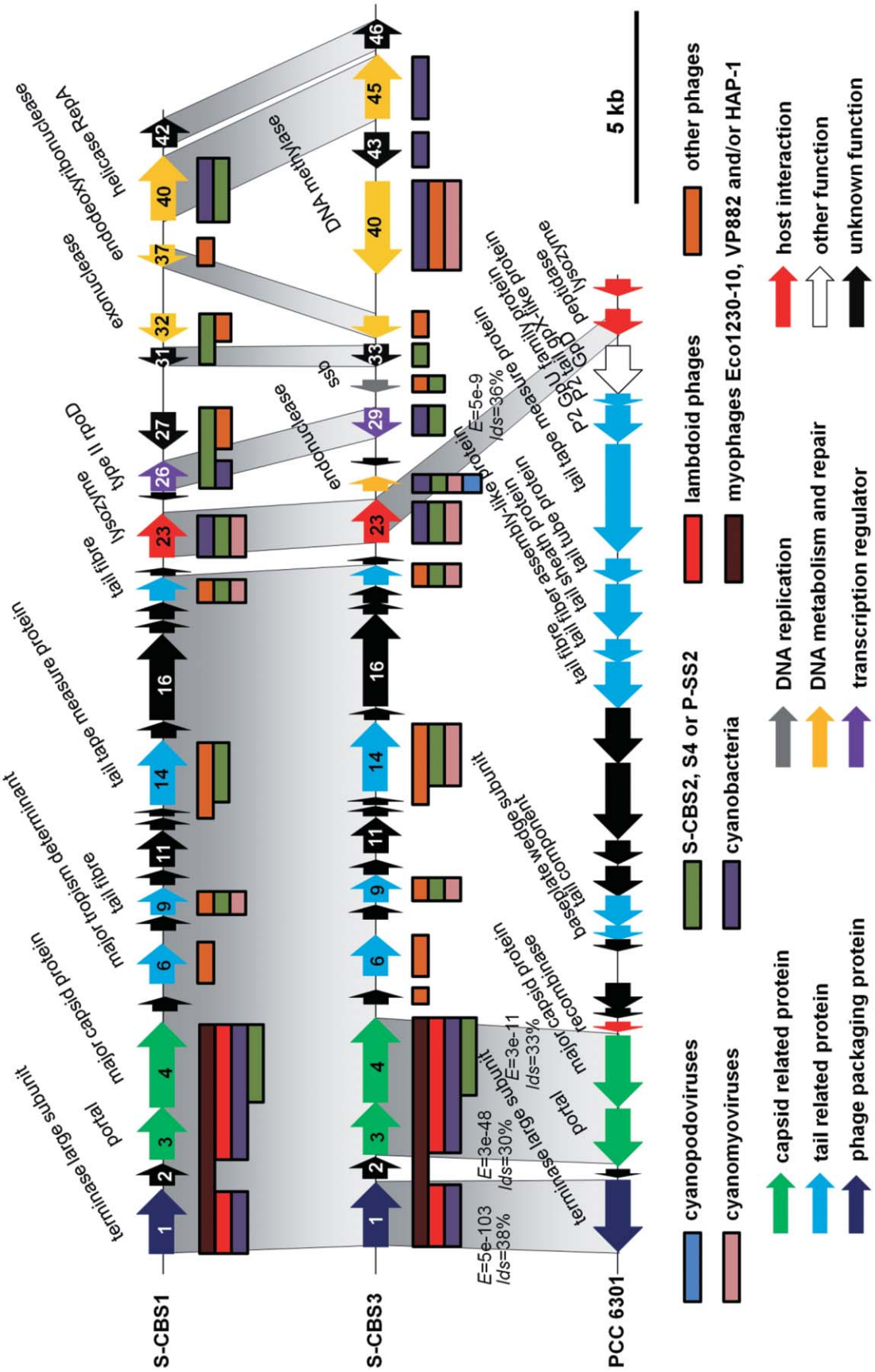

Fig. 2. Genome maps of *Synechococcus* siphoviruses S-CBS1 and S-CBS3, and a prophage-like region of *Synechococcus elongatus* PCC 6301 genome (Sugita *et al.*, 2007). The prophage-like sequences from *S. elongatus* PCC 6301 and PCC 7942 are completely identical, and therefore PCC 6301 was shown only. Only ORFs with significant BLAST hits (*E*-values ≤ 0.001) in protein databases were shown for S-CBS1 and S-CBS3. Homologous ORFs were connected by gray shadow. *E*-values (*E*) and identities (*Ids*) of the homologues between S-CBS3 and PCC 6301 were listed. ORFs which are homologous to those of cyanopodoviruses, cyanomyoviruses, cyanosiphoviruses (S-CBS2, S-CBS4 and P-SS2), cyanobacteria, lambdoid phages, myoviral *Enterobacteria* phage Eco1230-10, *Vibrio* phage VP882 and *Halomonas* phage ΦHAP-1 and other phages were indicated by coloured boxes below the line of arrows. Detailed information of ORFs for S-CBS1, S-CBS3 and PCC6301 were provided in Tables S2, S3 and S7 respectively.

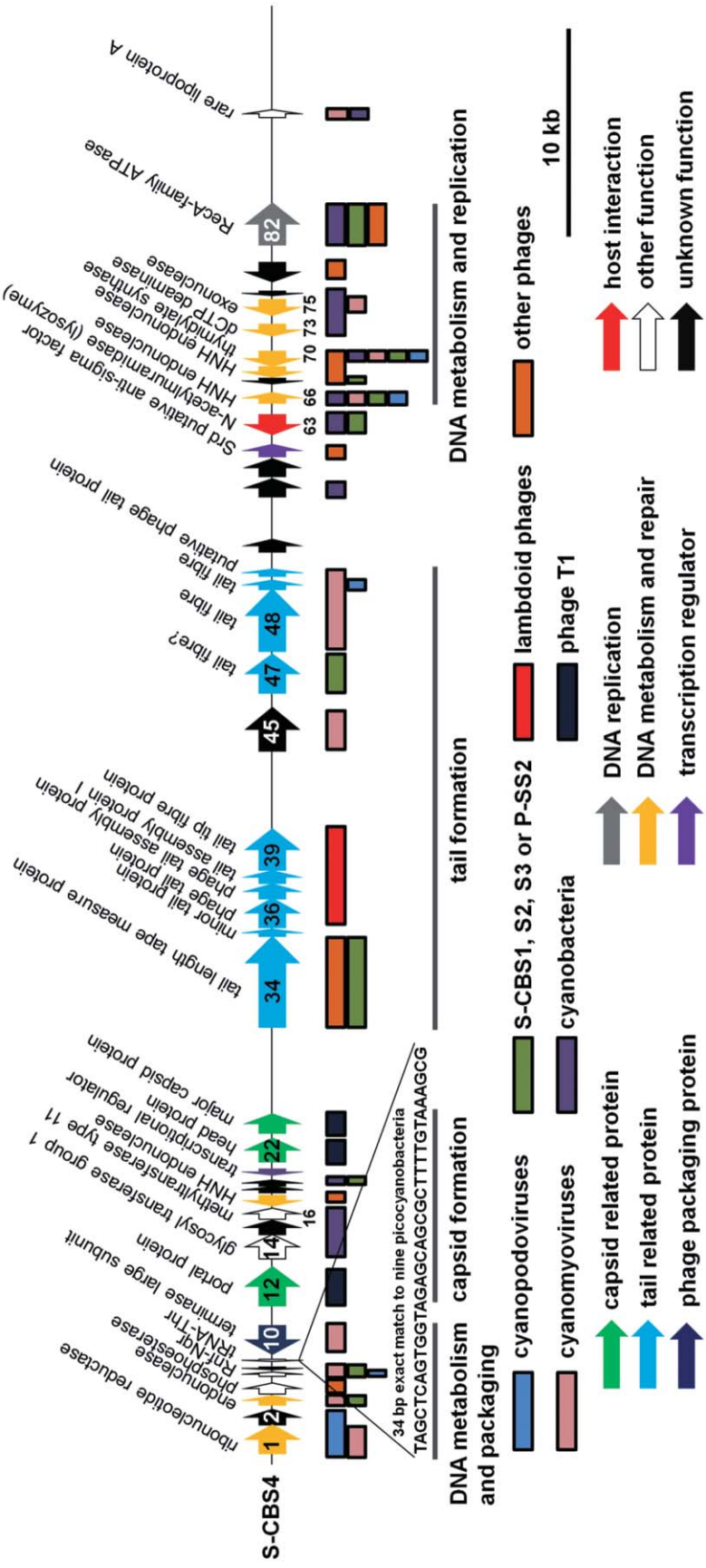

Fig. 3. Genome map of *Synechococcus* siphovirus S-CBS4. Only ORFs with significant BLAST hits (*E*-values $\leq$ 0.001) in protein databases were shown. ORFs that are homologous to those of cyanopodoviruses, cyanomyoviruses, cyanosiphoviruses (S-CBS1, S-CBS2, S-CBS3 and P-SS2), cyanobacteria, lambdoid phages, *Enterobacteria* phage T1 and other phages were indicated by coloured boxes below the line of arrows. Functional modules were indicated by gray lines. Detailed information of ORFs for S-CBS4 was provided in Table S4.

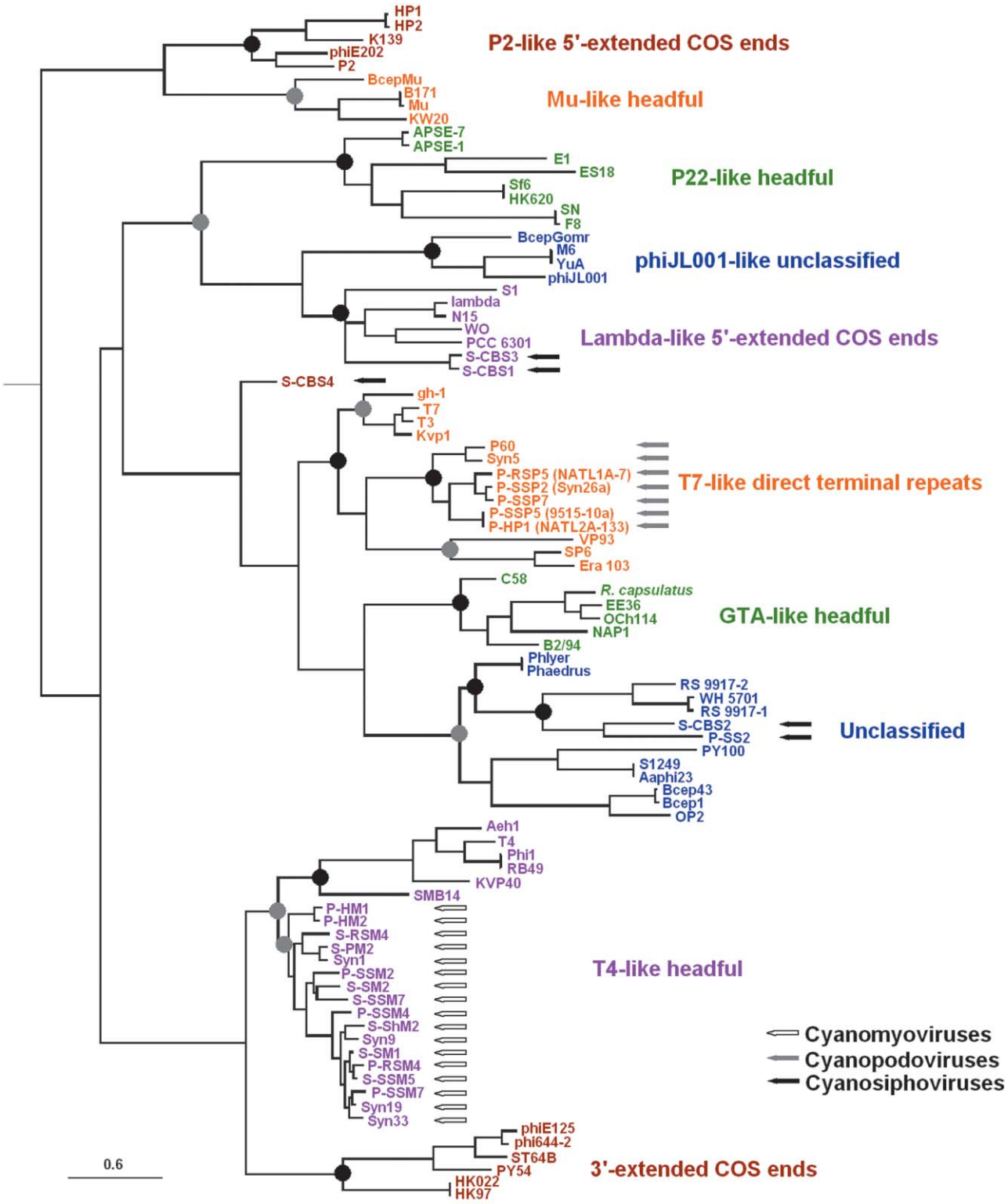

Fig. 4. Phylogenetic analysis based on TerL protein sequences showing the clustering of cyanosiphovirus subtypes. A maximum likelihood (ML) tree is shown. Distance, ML and maximum parsimony (MP) analyses were used to test the bootstrap supporting. Black dots at the node indicate bootstrap supporting by all the MP, ML and distance analyses with values > 75%, and gray dots indicate the supporting by at least two methods with values > 75%. Cyanosiphoviruses were indicated by black arrows, cyanopodoviruses by gray arrows and cyanomyoviruses by open arrows.

cyanomyoviruses or cyanopodoviruses has permitted the development of group-specific PCR primers to explore their genetic diversity in nature (Fuller *et al*., 1998; Zhong *et al*., 2002; Sullivan *et al*., 2008; Chen *et al*., 2009; Huang *et al*., 2010). However, the highly variable genomes of cyanosiphoviruses suggest that it is not possible to identify a specific gene marker for this group of cyanophage.

Without significant sequence similarity to known siphovirus types, cyanosiphoviurses so far characterized were not taxonomically identifiable (Table 1). P-SS2 genome was highlighted as its large and highly divergent genome compared to known siphovirus types (Sullivan *et al*., 2009). In the Phage Proteomic Tree (PPT) version 6 (http://www.phantome.org/PhageProteomicTree/latest/), P-SS2 alone represented a deep branch, which is consistent with its distinctness. PPT groups phages into taxa based on the overall similarity of entire predicted proteomes (Rohwer and Edwards, 2002). In the genomes of S-CBS1 and S-CBS3, the phage head module was most similar to unclassified myoviruses, while four tail-related genes respectively shared highest level of homology to *Bordetella* podoviruses, cyanomyoviruses, unclassified myoviruses and *Rhodococcus* siphoviruses (Fig. 2, Table S2). Similar mosaic nature was also observed for virion construction modules of S-CBS4, with the head-related genes most similar to Coliphage T1 and tail-related genes to lambdoid phages or cyanomyoviruses (Fig. 3). Indeed, siphoviruses are featured by their variable morphology (head shape and tail length) and highly divergent genomes and their taxonomic classification is most challenging due to intensive genetic recombination and genomic mosaicism (Hendrix *et al*., 1999; Juhala *et al*., 2000; Lawrence *et al*., 2002; Pedulla *et al*., 2003).

Potential lysogeny signatures in cyanosiphovirus genomes

Potential lysogeny in Synechococcus elongatus*.* Four genes (encoding terminase, capsid, portal and lysozyme) of S-CBS1 and S-CBS3 were homologous to predicted genes in the genomes of *Synechococcus elongatus* strains PCC 6301 and PCC 7942, two closely related freshwater picocyanobacteria with virtually identical genomes (Fig. 2). When the sequences in adjacent to these four genes of *S. elongatus* genomes (Holtman *et al*., 2005; Sugita *et al*., 2007) were carefully re-annotated, we identified a prophage-like structure in both genomes (Fig. 2, Table S7). This prophage-like element is *c.* 25 kb in length and contains 24 ORFs. The presence of prophage in microbial genomes can be misjudged particularly when the microbial genomes are not fully annotated (Zhao *et al*., 2010). Although most of these ORFs do not share homology to S-CBS1 or S-CBS3, our study demonstrates that new phage genome sequences could help uncover the unknown genetic features of microbial genomes. Although lysogeny has been reported in natural *Synechococcus* community (McDaniel *et al*., 2002; McDaniel and Paul, 2005; Ortmann *et al*., 2002), no intact prophage has been found among a dozen of known genomes of marine picocyanobacteria (Kettler *et al*., 2007; Dufresne *et al*., 2008). Moreover, unlike P-SS2, the other four cyanobacterial siphoviruses do not contain the gene encoding integrase or recombination protein. However, our result shows that a certain type of prophage may be present in freshwater *Synechococcus*. Whether the prophage element of *S. elongatus* is inducible requires further investigation.

Putative S-CBS4 integration site? A 34 bp sequence within the putative threonine tRNA gene of S-CBS4 is identical to part of tRNA-Thr sequence in non-coding intergenic region of its host genome (*Synechococcus* CB0101) and other *Synechococcus* and *Prochlorococcus* genomes (Figs 3 and S3). Phages commonly insert their DNA at specific sites on the host chromosome, such as tRNA (Campbell, 2003) and tmRNA (Williams, 2002). Such inferred phage–host site-specific attachment sites (*attP* for phages, *attB* for hosts), flanking an integrase gene, were identified in the host genomes of P-SSP7 (a cyanopodovirus) and P-SS2, with 42 bp and 36 bp (included in a 53 bp exact matching sequence) sequences perfectly matching parts of tRNA sequences in hosts respectively (Sullivan *et al*., 2006; Sullivan *et al*., 2009). Furthermore, the region nearby *attB* in P-SS2's host genome is seemingly located in a genomic island associated with phage insertion (Sullivan *et al*., 2009). However, an *int* gene was not found in S-CBS4 and therefore, it is unclear whether the S-CBS4 tRNA gene plays a role in integration into the host genome.

Genetic exchanges between cyanobacteria and cyanosiphoviruses

In order to understand what common genes are shared between cyanosiphoviruses and cyanobacteria, the entire proteomes of five cyanosiphoviruses were searched against 25 *Prochlorococcus* and *Synechococcus* genomes (Table 2). There were 57 shared genes and 40 of them with predicted functions could be roughly recognized as either host-related (8) or phage-related (32).

Cyanobacteria-related genes in cyanosiphoviruses. The cyanobacteria-related proteins found in cyanosiphoviruses were mainly associated with transcriptional regulation, photosynthesis or cobalamin synthesis, such as RNA polymerase (RNAP) sigma factor, high-light-inducible

Table 2. Summary of cyanosiphovirus proteins that have homologues in *Synechococcus* and *Prochlorococcus*.

			Synechococcus												
Gene product		Function[a]	CC9311	CC9605	WH 8102	CC9902	BL107	WH 7805	WH 7803	RS9917	RS9916	WH 5701	RCC307	PCC 6301	PCC 7002
S-CBS1_gp01	S-CBS3_gp01	Terminase large subunit												1	
S-CBS1_gp03	S-CBS3_gp03	Portal												1	
S-CBS1_gp04	S-CBS3_gp04	Major capsid protein			1					1				1	
S-CBS1_gp23	S-CBS3_gp23	Lysozyme								1				1	
S-CBS1_gp26	S-CBS3_gp29	Type II RNAP sigma factor	10	6	6	8	7	8	9	9	8	10	9	8	6
S-CBS1_gp40	S-CBS3_gp45	RepA helicase		1	1					1		1			1
	S-CBS3_gp24	Endonuclease			1	1									
	S-CBS3_gp40	DNA methylase										1			
	S-CBS3_gp43									1		1			
S-CBS4_gp014		Glycosyl transferase										1			
S-CBS4_gp015						1						1			
S-CBS4_gp016		Methyltransferase type II													
S-CBS4_gp019										1					
S-CBS4_gp063		N-acetylmuramidase										1			
S-CBS4_gp066		HNH endonuclease	2	2	2	2	1	1	2	1	1	1	2	1	1
S-CBS4_gp070		Thymidylate synthase	1	1	1	1	1	1	1	1	1	1	1	1	1
S-CBS4_gp073		dCTP deaminase	1	1	1	1	1	1	1	1	1	1	1	1	1
S-CBS4_gp075		Exonuclease		1			1								
S-CBS4_gp076								1		1	2			2	
S-CBS4_gp082		RecA-family ATPase		3	1								1		1
S-CBS4_gp093		Rare lipoprotein A	1	2	2	2	1	2	1	1	1	1	1	1	2
S-CBS2_gp031	P-SS2_gp087	Phage structural protein								1					
S-CBS2_gp036	P-SS2_gp091	Host specificity protein								1		1			
S-CBS2_gp038	P-SS2_gp092	Phage structural protein								1		1			
S-CBS2_gp054	P-SS2_gp098	Lysozyme								2		2			
S-CBS2_gp061	P-SS2_gp113	Type II RNAP sigma factor	8	6	6	6	6	7	9	8	8	8	9	7	6
S-CBS2_gp073	P-SS2_gp126	Thymidylate synthase	1	1	1	1	1	1	1		1	1	1	1	
S-CBS2_gp075	P-SS2_gp001	Terminase small subunit									1				
S-CBS2_gp076	P-SS2_gp002	Terminase large subunit								2	1	1			
S-CBS2_gp077	P-SS2_gp003	Type III RNAP sigma factor	3	2	3	3	2	2	3	5	3	4	2	3	2
S-CBS2_gp078	P-SS2_gp109	DNA primase	1	1	1	1	1	1	1	1	1	1	1	1	1
S-CBS2_gp087	P-SS2_gp011	Ribonucleotide reductase	1	1	1	1	1	1	1	1	1	1	1	1	
S-CBS2_gp088	P-SS2_gp111	dCTP deaminase						1	1			1		1	
S-CBS2_gp089	P-SS2_gp025	Phage structural protein								2		2			
S-CBS2_gp099	P-SS2_gp038									1	1		1		
S-CBS2_gp005		DNA methylase												1	2
S-CBS2_gp017		HNH endonuclease	2	2	2	2	1	1	2	2	1	3	2	1	2
S-CBS2_gp037					1										
S-CBS2_gp039				2	2			1			1		1		
S-CBS2_gp040		HLIP													
S-CBS2_gp047					1										

Table 2. *cont.*

Gene product		Function[a]	*Synechococcus*											
			CC9311	CC9605	WH 8102	CC9902	BL107	WH 7805	WH 7803	RS9917	RS9916	WH 5701	RCC307	PCC 6301	PCC 7002
S-CBS2_gp062		Ssb protein													
S-CBS2_gp070		DNA processing chain A		1	1		1				2	1	1	1	1
S-CBS2_gp086		Peptidase		1							1				
S-CBS2_gp092				1	1	1	1	1	1	1	1	1	1	1	
S-CBS2_gp095			1	1	1	1	1	1	1		1	1	1		
S-CBS2_gp102		Haemagglutinin-neuraminidase	1	1	2	2	2	1	1		1	2	2	1	
	P-SS2_gp014		1												
	P-SS2_gp028														
	P-SS2_gp036	CobO	3	3	3	3	3	3	3	3	3	4	3	3	
	P-SS2_gp049							1	3	1	1	1			
	P-SS2_gp053														
	P-SS2_gp097									1					
	P-SS2_gp100			1											
	P-SS2_gp101	Integrase	1	1	1	1	1	1	1	2	1	1	1	2	
	P-SS2_gp103		1					1	1	1	1	1	1		
	P-SS2_gp114	Ssb protein	1	1	1	1	1	1	1		1	1	1	1	

Gene product		Function[a]	*Prochlorococcus*											
			MED4	MIT 9515	MIT 9215	MIT 9312	AS9601	MIT 9301	NATL2A	NATL1A	SS120	MIT 9211	MIT 9313	MIT 9303
S-CBS1_gp01	S-CBS3_gp01	Terminase large subunit												
S-CBS1_gp03	S-CBS3_gp03	Portal												
S-CBS1_gp04	S-CBS3_gp04	Major capsid protein												
S-CBS1_gp23	S-CBS3_gp23	Lysozyme												
S-CBS1_gp26	S-CBS3_gp29	Type II RNAP sigma factor	5	5	5	5	5	5	5	5	5	5	8	9
S-CBS1_gp40	S-CBS3_gp45	RepA helicase												
	S-CBS3_gp24	Endonuclease	2	1	1	1	1	1	1	1	1		1	
	S-CBS3_gp40	DNA methylase												
	S-CBS3_gp43													
S-CBS4_gp014		Glycosyl transferase												
S-CBS4_gp015														
S-CBS4_gp016		Methyltransferase type II											1	
S-CBS4_gp019														
S-CBS4_gp063		N-acetylmuramidase												
S-CBS4_gp066		HNH endonuclease	4	3	3	3	3	3	2	2	2	2	2	2
S-CBS4_gp070		Thymidylate synthase	1	1	1	1	1	1	1	1	1	1	1	1
S-CBS4_gp073		dCTP deaminase	1	1	1	1	1	1	1	1	1	1	1	1
S-CBS4_gp075		Exonuclease												
S-CBS4_gp076														
S-CBS4_gp082		RecA-family ATPase	1		1	1		1						
S-CBS4_gp093		Rare lipoprotein A											1	1

Table 2. *cont.*

Gene product		Function[a]	*Prochlorococcus*											
			MED4	MIT 9515	MIT 9215	MIT 9312	AS9601	MIT 9301	NATL2A	NATL1A	SS120	MIT 9211	MIT 9313	MIT 9303
S-CBS2_gp031	P-SS2_gp087	Phage structural protein												
S-CBS2_gp036	P-SS2_gp091	Host specificity protein												
S-CBS2_gp038	P-SS2_gp092	Phage structural protein												
S-CBS2_gp054	P-SS2_gp098	Lysozyme												
S-CBS2_gp061	P-SS2_gp113	Type II RNAP sigma factor	5	5	5	5	5	5	5	5	5	5	8	9
S-CBS2_gp073	P-SS2_gp126	Thymidylate synthase	1	1	1	1	1	1	1	1	1	1	1	1
S-CBS2_gp075	P-SS2_gp001	Terminase small subunit												
S-CBS2_gp076	P-SS2_gp002	Terminase large subunit												
S-CBS2_gp077	P-SS2_gp003	Type III RNAP sigma factor	1	1	1	1	1	1	1	1	1	1	2	3
S-CBS2_gp078	P-SS2_gp109	DNA primase	1	1	1	1	1	1	1	1	1	1	1	1
S-CBS2_gp087	P-SS2_gp011	Ribonucleotide reductase	1	1	1	1	1	1	1	1	1	1	1	1
S-CBS2_gp088	P-SS2_gp111	dCTP deaminase												
S-CBS2_gp089	P-SS2_gp025	Phage structural protein												
S-CBS2_gp099	P-SS2_gp038													1
S-CBS2_gp005		DNA methylase								1				
S-CBS2_gp017		HNH endonuclease	4	3	3	2	3	3	2	2	2	2	2	2
S-CBS2_gp037														
S-CBS2_gp039			1		2	2	2	2	1	1	1	1	1	1
S-CBS2_gp040		HLIP	13	12	8	14	11	7	23	23	6	5	4	4
S-CBS2_gp047														
S-CBS2_gp062		Ssb protein	1	1	1	1	1				1	1		
S-CBS2_gp070		DNA processing chain A	1	1	1	1	1		1	1				
S-CBS2_gp086		Peptidase												
S-CBS2_gp092			1	1	1	1	1	1	1	1	1	1	1	2
S-CBS2_gp095			1	1	1	1	1	1	1	1	1			1
S-CBS2_gp102		Haemagglutinin-neuraminidase	2		2	2	2	2	2	2	1	2		
	P-SS2_gp014													
	P-SS2_gp028			1	1		1	1	1	1				
	P-SS2_gp036	CobO	3	3	3	3	3	3	3	3	3	3	3	3
	P-SS2_gp049													
	P-SS2_gp053												1	1
	P-SS2_gp097													
	P-SS2_gp100													
	P-SS2_gp101	Integrase	1	1	1	1	1	1	1	1	1	1	1	1
	P-SS2_gp103		1	1	1	1	1	1	1	1	1	1	1	1
	P-SS2_gp114	Ssb protein	1	1	1	1	1	1	1	1	1	1	1	1

a. Blank strands for unknown function.
A number in this table stands for the number of a cyanosiphovirus gene's homologues found in one *Synechococcus* or *Prochlorococcus* genome.

protein (HLIP), cobalamin synthesis component (CobO) (Table 2). Interestingly, all the sigma factors (type II and type III) found in cyanosiphoviruses (except for S-CBS4) likely contain host features as they all have homologues in cyanobacterial genomes (Table 2). In contrast, known cyanopodoviruses do not encode sigma factor and all cyanomyoviruses contain one for presumably T4-like late transcription without any homology to sigma factors in cyanobacterial hosts. RNAP sigma factors in P-SS2 were implicated to modulate host RNAP activity during infection (Sullivan *et al*., 2009). It is not clear whether the homology of sigma factors between phages and hosts enable phages to regulate host activities more efficiently.

None of the core photosystem reaction genes (*psbA* or *psbD*) were found in the S-CBS1, S-CBS2, S-CBS3 and S-CBS4 genomes, and this is consistent with the previous study based on PCR method (Wang and Chen, 2008). The two *Prochlorococcus* siphoviruses (P-SS1 and P-SS2) also lack the *psbA/D* genes (Sullivan *et al*., 2006). The *psbA* gene has been found commonly present in the cyanomyoviruses (Mann *et al*., 2003; Lindell *et al*., 2004; Millard *et al*., 2004; Sullivan *et al*., 2006) and in some cyanopodoviruses (Sullivan *et al*., 2006; Wang and Chen, 2008; Thompson *et al*., 2011). Interestingly, a HLIP-encoding gene (*hli*) was identified in S-CBS2, which was not found in other cyanosiphoviruses (Fig. 1). The *hli* genes have also been found in all the known cyanomyoviruses and some cyanopodoviruses. HLIPs in cyanobacterial cells are thought to protect the photosynthetic apparatus from photodamage (Havaux *et al*., 2003) and cyanophage-version HLIPs were found expressed during infection cycles (Lindell *et al*., 2005; Clokie *et al*., 2006) and in the natural environment (Sharon *et al*., 2007), suggesting a mutually beneficial relationship between host and virus. It appears the appearance of photosynthesis genes in the three phage families (myo-, podo- and siphoviruses) are not equal (Sullivan *et al*., 2006; Wang and Chen, 2008), that is, the *psbA/D* genes are more prevalent in cyanomyoviruses than cyanopodoviruses, and have not yet been detected in cyanosiphoviruses. The rare occurrence of these photosynthesis-related genes in cyanosiphoviruses may be related to the temperate lifestyle of siphovirus (Sullivan *et al*., 2009). Further study is needed to better understand if there is a link between the inheritance of photosynthesis genes and the lifestyle of cyanophages.

Cyanosiphovirus-related genes prevalently present in hosts. It is interesting that, among typical phage-like genes, those involved in DNA metabolism, replication and integration (DMRI) have homologues in almost all the *Synechococcus* and *Prochlorococcus* genomes analysed, whereas homologues of phage structural or packaging genes were only seen in a few genomes (Table 2). It is likely that cyanobacteria tended to share genes associated with cellular metabolism with cyanophages rather than typical virion components. More strikingly, most of these DMRI-related proteins have a bulk of top BLAST hits from cyanobacteria (Table 2), such as integrase, deoxycytidine triphosphate (dCTP) deaminase, thymidylate synthase (Td) and ribonucleotide reductase (RNR), suggesting the occurrence of phage–host genetic exchanges, possibly via prophage integration or homologous recombination. In order to answer what are the directionality of such genetic exchanges and when did they occur, we did phylogenetic analysis for the four proteins mentioned above.

The phylogeny of Td shows that *Prochlorococcus* and cyanophages from all the three families fell into a cluster while marine *Synechococcus* were clustered with all the other cyanobacteria (Fig. S4A), suggesting that *td* may be transferred to *Prochlorococcus* from cyanophage(s). This acquisition might occur before the descent of major *Prochlorococcus* lineages as the Td tree shows a same pattern to the 16S rRNA tree (Td co-evolved with 16S rRNA) (Fig. S4A).

Furthermore, the integrase phylogeny also agreed with that of 16S rRNA for both *Prochlorococcus* and *Synechococcus* (Fig. S4B). Note that almost every cyanobacterial genomes analysed in this study have a P-SS2-like *int* and most *Prochlorococcus* only have this one (Table 2). It is likely that this *int* represents a prophage integration into a cyanobacteria ancestor before the descent of major cyanobacterial lineages. The facts that no complete prophages were found in known marine picocyanobacterial genomes and that prophage signatures in these genomes were fragmentary and remnant (Sullivan *et al*., 2009) also support, at least in part, the hypothesis that some cyanophage integration were not recent events.

Similar to Td, the phylogeny of RNR also shows discrepancy (Fig. S4C), that is, (i) as basal branches, cyanophages P60, P-SS2 and S-CBS2 (Class II RNRs) were clustered with the cyanobacteria including marine picocyanobacteria and some freshwater species; (ii) also as a basal branch, freshwater cyanomyovirus Ma-LMM01 (Class I RNR) fell into the cluster mainly comprised of freshwater cyanobacteria including its host species, *Microcystis aeruginosa* (Yoshida *et al*., 2008). Since most cyanobacteria have a Class II RNR, this discrepancy suggests the occurrence of lateral gene transfer on Class I RNRs from cyanophages (such as Ma-LMM01) to cyanobacteria. However, although we infer that the RNR exchanges between P60, P-SS2 and S-CBS2 and hosts might also occur anciently [based on (i) cyanobacteria co-evolved with 16S rRNA; (ii) cyanophages are basal to cyanobacteria], we cannot determine the direction. It is the same case for dCTP deaminase between S-CBS4 and host (Fig. S4D).

Moreover, the clustering of RNRs from cyanosiphovirus S-CBS4 and six cyanopodoviruses (Fig. S4C) and the clustering of Tds from three cyanophage families (Fig. S4A) suggest direct and host-mediated phage-to-phage gene transfers respectively. This observation further supports the idea of genetic exchanges via a large phage gene pool (Hendrix, 1999; Pedulla *et al*., 2003).

Recruitment of cyanosiphoviruses from metagenomes

In order to explore a preliminary scenario of the relative abundances of the three cyanophage families in the sea, we recruited cyanophage-like sequences from the Global Ocean Sampling (GOS) Expedition database (Rusch *et al*., 2007) against 29 cyanophage genomes. Fragment recruitment of cyanosiphoviruses yielded much less sequences compared to cyanomyoviruses and cyanopodoviruses, suggesting that cyanosiphoviruses occur less frequently in the ocean surface water (Fig. 5A). The ratios of recruited sequences among the three cyanophage families were consistent for most of the marine habitats where GOS Expedition sampled (Fig. 5A). For example, in the vast open oceans, the hit counts ratio of cyanosiphovirus : cyanopodoviruse : cyanomyovirus is roughly 1:10:20 (Fig. 5A). The single protein-based (i.e. TerL) recruitment was consistent with the whole genome-based recruitment (Fig. 5B).

We also searched cyanosiphovirus homologues in a metagenome dataset from the deep chlorophyll maximum (DCM) depth of Mediterranean Sea (Ghai *et al*., 2010), and found four out of 197 fosmid clones contained at least 50% ORFs homologous to the cyanosiphoviruses described here (Table S8). Ghai and colleagues (2010) reported that 34 out of 197 fosmid clones were attributed to cyanophage, with 12 most closely related to cyanopodoviruses and nine to cyanomyoviruses but none to cyanosiphovirus. The newly sequenced genomes allow us to better estimate the contribution of cyanosiphoviruses in the metagenomic database.

Cyanophage sequences constitutes a proportion of sequences derived from microbial fraction-targeting metagenomes and were likely originated from phages that were replicating inside picocyanobacterial cells (DeLong *et al*., 2006; Williamson *et al*., 2008). Our study further suggests cyanosiphoviruses may cause less picocyanobacterial infection and contribute a smaller proportion of host lysis in the sea. However, the GOS samples were confined to microbial cells of 0.1–0.8 μm fraction and were not designed for viral metagenomics originally. Therefore, the GOS database is not perfect for searching all marine viruses, and could potentially bias our estimation of biogeographical patterns of major cyanophage types. Another limitation of GOS database is that only surface water samples were collected. The DCM metagenomic database from the Mediterranean Sea suggests that cyanosiphoviruses could be more important in the deeper euphotic zone.

Conclusions

Siphoviruses that infect marine picocyanobacteria have a wide genomic diversity compared to cyanomyoviruses and cyanopodoviruses. The genome sizes of five known cyanosiphoviruses vary from 30 to 108 kb. The four *Synechococcus* siphoviruses S-CBS1, S-CBS2, S-CBS3 and S-CBS4, plus *Prochlorococcus* siphovirus P-SS2, could be classified in three major subtypes based on morphology and comparison of genomic sequences. Comparison of cyanosiphovirus and host genomes suggests freshwater cyanobacteria may retain prophages while typical marine picocyanobacteria only exhibit past prophage integration signatures. Unlike cyanomyoviruses and some cyanopodoviruses, cyanosiphoviruses do not carry the photosynthesis genes *psbA* and *psbD* but share a bulk of genes involved in DNA metabolism and replication with hosts. It is likely that different virus–host lifestyles (i.e. broad host vs. narrow host, lytic vs. temperate) pose different selection pressures on gene acquisition between virus and host. Although diverse cyanosiphoviruses have been isolated in the Chesapeake estuary, they appear present in other oceanic regions but as a small portion of the cyanophage community compared to the other two groups of cyanophages. The genome sequences of these cyanosiphoviruses allow us to explore the genomic evolution and relative distribution of three cyanophage families in the ocean.

Experimental procedures

Cyanophage isolation, purification and DNA preparation for genomic sequencing

Phages S-CBS1, S-CBS2,S-CBS3 and S-CBS4 were isolated from the Chesapeake Bay water (Stn. 804, 38°04′N, 76°13′W) collected in September 2002 (S-CBS1 and S-CBS2), June 2003 (S-CBS3) and July 2004 (S-CBS4), on board the R/V *Cape Henlopen* (Wang and Chen, 2008). S-CBS1 was isolated from *Synechococcus* strain CB0201 which belongs to *Synechococcus* subcluster 5.1 (i.e. marine cluster A) (Chen *et al*., 2006). S-CBS2, S-CBS3 and S-CBS4 were isolated from *Synechococcus* stains CB0204, CB0202 and CB0101 respectively, all of which are the members of *Synechococcus* subcluster 5.2 (i.e. marine cluster B) (Chen *et al*., 2006). The sequential phage purification, scale-up and genomic DNA extraction were described previously (Wang and Chen, 2008). To prepare sufficient DNA templates for shotgun sequencing, purified phage DNA were amplified using Genomiphi V2 kit (GE Healthcare, Piscataway, NJ, USA) following the protocol provided by the manufacturer. For S-CBS1, S-CBS2 and S-CBS3, DNA was sheared by

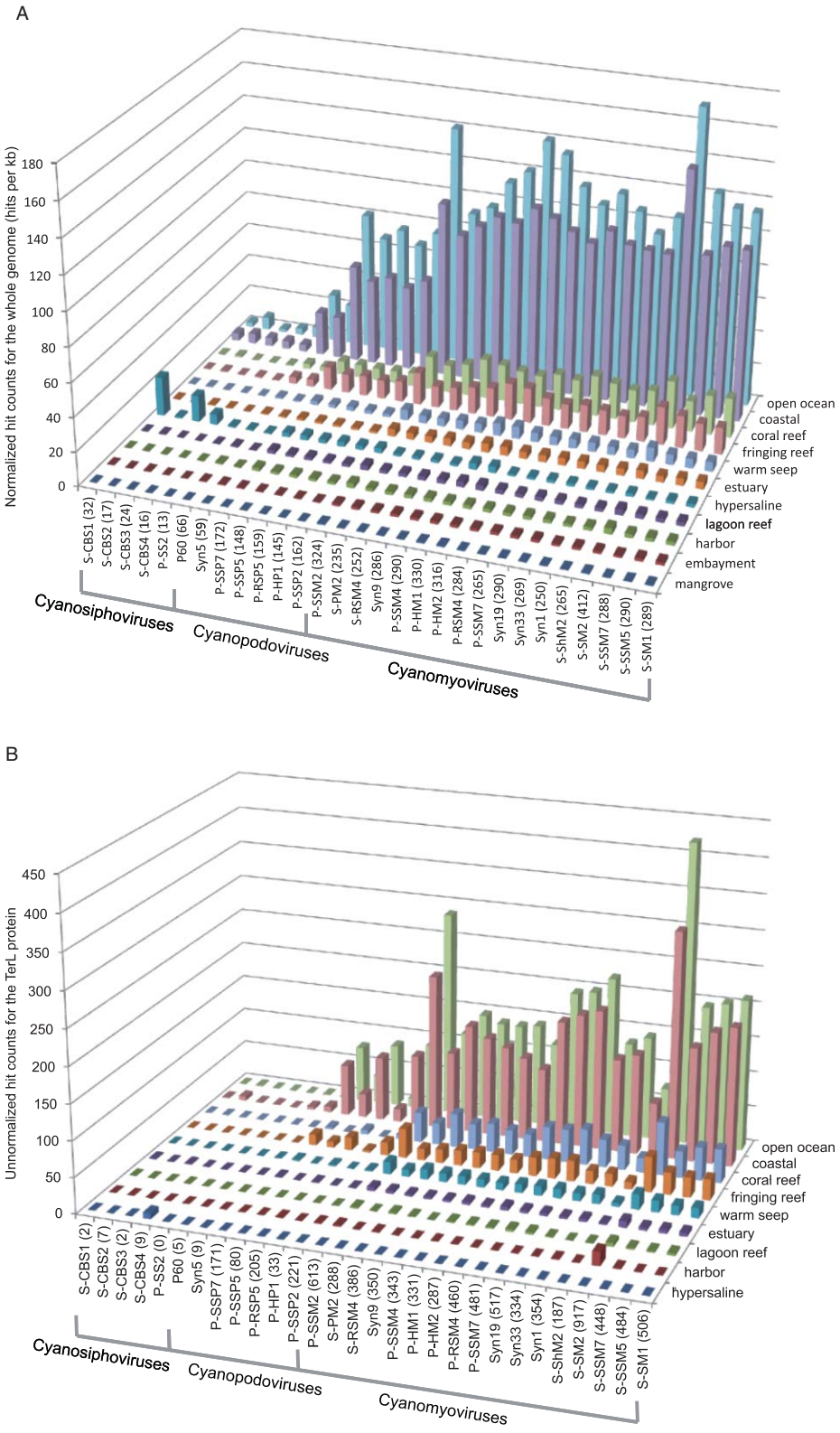

Fig. 5. BLASTP hits number of cyanophage whole genomes and TerL protein in GOS metagenomic library.
A. Normalized hit counts from the BLASTP of all the ORFs of 29 cyanophage genomes. The total number of hits per 1 kb genome for each cyanophage was shown in parentheses.
B. Unnormalized hit counts from the BLASTP of TerL protein in 29 cyanophages. The total hits number for each cyanophage was shown in parentheses. Red, blue and orange lines indicate cyanosiphoviruses, cyanopodoviruses and cyanomyoviruses respectively.

ultrasonication and 1.6–4 kb fragments were retrieved as inserts. Shotgun library was constructed using pUC19 vector and inserts were then sequenced using a commercial automated sequencer ABI3730x1 (Majorbio Biotech, Shanghai, China). Primer-walking was carried out for closing the gaps. Sequencing reads were assembled by using the Phred-Phrap software package (http://www.phrap.com). No detectable hosts or artificially chimera sequences were assembled although we used the genomic amplified DNA for sequencing. The coverages of the whole genome sequences for phages are *c.* 10-fold for S-CBS1, *c.* 12-fold for S-CBS2 and *c.* 26-fold for S-CBS3. The complete genome sequence of S-CBS4 was sequenced and assembled by Broad Institute, under the Marine Phage Sequencing Project (http://www.broadinstitute.org/annotation/viral/Phage/Home.html). The genome sequences were deposited in GenBank with accession number HM480106 for S-CBS1, GU936714 for S-CBS2, GU936715 for S-CBS3 and HQ698895 for S-CBS4.

Genome annotation, homologue searching and core genome determining

Open reading frames were predicted using Glimmer (Delcher *et al*., 1999) and GeneMark (Lukashin and Borodovsky, 1998). Translated ORFs were compared with known protein sequences in GenBank and Swiss-Prot databases using the BLASTP program. Generally, predicted ORFs were considered as hypothetical proteins and function annotation were assigned when BLASTP *E*-values were ≤ 0.001. The computer program tRNAscan-SE was used to identify tRNA sequence (Lowe and Eddy, 1997). Genome synteny was drawn using Microsoft PowerPoint. To better extract homology among cyanophages and between cyanosiphoviruses and cyanobacteria, two protein datasets were created, with one including 25 cyanophage genomes (17 cyanomyoviruses, seven cyanopodoviruses and cyanosiphovirus P-SS2, listed in Table S5) and the other one including 25 cyanobacterial genomes (listed in Table 2). Protein sets of S-CBS1 to S-CBS4 were BLASTP-searched against both the datasets. In addition, protein set of P-SS2 was also compared to the cyanobacterial protein dataset. An *E*-value cut-off ≤ 0.001 was set for homologue candidate. Core genome for five available cyanosiphoviruses was determined by local 'all against all' BLASTP comparison for all the cyanosiphovirus protein sequences. An orthologous gene was defined when one was harboured by all the cyanosiphoviruses and has an *E*-value lower than 0.001 between any pairwise amino acid sequences. Core genome of seven cyanopodoviruses (listed in Table S5) was also determined using the same method. Genomic dot plots for the three cyanophage families were created by using the Gepard program (Krumsiek *et al*., 2007), with the default 'DNA' matrix and word length (see the legend of Fig. S2).

Phylogenetic analyses

Amino acid sequences were aligned using Clustal X2 (Larkin *et al*., 2007). Neighbour-joining (NJ) and Maximum Parsimony (MP) analyses ware performed by using PAUP 4.0b10 software. Maximum likelihood (ML) analysis was carried out using the CIPRES web portal RAxML service (Stamatakis, 2006; Stamatakis *et al*., 2008). Bootstrap resamplings were conducted for 1000 replications in both the NJ and MP analyses and 100 replications in ML analysis.

Metagenomic analyses

In order to infer the relative abundance of different families of cyanophages, all the predicted protein sequences of the currently available 29 marine cyanophage genomes (Table S5) were used to recruit the GOS sequences in the CAMERA database (http://camera.calit2.net/) (Rusch *et al*., 2007). BLASTP program provided by the CAMERA interface was used to search the protein sequences in 'GOS All ORF Peptides' dataset, and a restricted *E*-value ($< 10^{-50}$) was set as the cut-off. Moreover, amino acid sequences of larger terminase subunit gene (*terL*) were retrieved from available cyanophage genomes and independently BLASTP-searched against the GOS metagenomic database. The optimized BLASTP *E*-value for TerL was set differently for different groups of cyanophages, which is $< 10^{-135}$ for cyanomyoviruses and cyanopodoviruses, $< 10^{-130}$ for S-CBS1 and S-CBS3, $< 10^{-90}$ for P-SS2, S-CBS2 and S-CBS4, based on the pairwise *E*-values among TerL from cyanophages and most closely related non-cyanophages (data not shown). In order to reduce the effects of uneven genome sizes, the raw hit counts from the whole genome recruitment were normalized by genome sizes. All the predicted proteins from the five cyanosiphovirus genomes were also BLASTP-searched against the fosmid library metagenome dataset of microbial community in the Mediterranean Deep Chlorophyll Maximum water (Ghai *et al*., 2010). A predicted gene from a fosmid clone was considered as cyanosiphovirus-related when the *E*-value was ≤ 0.001.

Acknowledgements

Sequencing of cyanophage genomes of S-CBS1, S-CBS2 and S-CBS3 was supported by the Xiamen University 111 Program to F.C. and the NSFC project 91028001 to N.J. Sequencing of cyanophage S-CBS4 genome was supported by the Gordon and Betty Moore Foundation under the Marine Phage Sequencing Project. We thank the Broad Institute Genome Sequencing Platform for their work on the genome of cyanophage S-CBS4. We also thank Wan-Hsin Chen for editing the manuscript. Finally, we thank four reviewers and our editor for their comments and suggestions that led to a great revision of our manuscript.

References

Black, L.W. (1989) DNA packaging in dsDNA bacteriophages. *Annu Rev Microbiol* **43:** 267–292.

Campbell, A. (2003) Prophage insertion sites. *Res Microbiol* **154:** 277–282.

Casjens, S., Gilcrease, E.B., Winn-Stapley, D.A., Schicklmaier, P., Schmieger, H., Pedulla, M.L., *et al*. (2005) The generalized transducing *Salmonella* bacteriophage ES18: complete genome sequence and DNA packaging strategy. *J Bacteriol* **187:** 1091–1104.

Chen, F., and Lu, J.R. (2002) Genomic sequence and evolution of marine cyanophage P60: a new insight on lytic and lysogenic phages. *Appl Environ Microbiol* **68:** 2589–2594.

Chen, F., Wang, K., Kan, J., Suzuki, M.T., and Wommack, K.E. (2006) Diverse and unique picocyanobacteria in Chesapeake Bay, revealed by 16S-23S rRNA internal transcribed spacer sequences. *Appl Environ Microbiol* **72:** 2239–2243.

Chen, F., Wang, K., Huang, S.J., Cai, H.Y., Zhao, M.R., Jiao, N.Z., *et al*. (2009) Diverse and dynamic populations of cyanobacterial podoviruses in the Chesapeake Bay unveiled through DNA polymerase gene sequences. *Environ Microbiol* **11:** 2884–2892.

Chenard, C., and Suttle, C.A. (2008) Phylogenetic diversity of sequences of cyanophage photosynthetic gene psbA in marine and freshwaters. *Appl Environ Microbiol* **74:** 5317–5324.

Clokie, M.R.J., Shan, J., Bailey, S., Jia, Y., Krisch, H.M., West, S., *et al*. (2006) Transcription of a 'photosynthetic' T4-type phage during infection of a marine cyanobacterium. *Environ Microbiol* **8:** 827–835.

Coleman, M.L., Sullivan, M.B., Martiny, A.C., Steglich, C., Barry, K., Delong, E.F., and Chisholm, S.W. (2006) Genomic islands and the ecology and evolution of *Prochlorococcus*. *Science* **311:** 1768–1770.

Delcher, A.L., Harmon, D., Kasif, S., White, O., and Salzberg, S.L. (1999) Improved microbial gene identification with GLIMMER. *Nucleic Acids Res* **27:** 4636–4641.

DeLong, E.F., Preston, C.M., Mincer, T., Rich, V., Hallam, S.J., *et al*. (2006) Community genomics among stratified microbial assemblages in the ocean's interior. *Science* **311:** 496–503.

Dreher, T.W., Brown, N., Bozarth, C.S., Schwartz, A.D., Riscoe, E., Thrash, C., *et al*. (2011) A freshwater cyanophage whose genome indicates close relationships to photosynthetic marine cyanomyophages. *Environ Microbiol* **13:** 1858–1874.

Dufresne, A., Ostrowski, M., Scanlan, D.J., Garczarek, L., Mazard, S., Palenik, B.P., *et al*. (2008) Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria. *Genome Biol* **9:** R90.

Fuller, N.J., Wilson, W.H., Joint, I.R., and Mann, N.H. (1998) 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* **64:** 2051–2060.

Ghai, R., Martin-Cuadrado, A.B., Molto, A.G., Heredia, I.G., Cabrera, R., Martin, J., *et al*. (2010) Metagenome of the Mediterranean deep chlorophyll maximum studied by direct and fosmid library 454 pyrosequencing. *ISME J* **4:** 1154–1166.

Havaux, M., Guedeney, G., He, Q., and Grossman, A.R. (2003) Elimination of high-light-inducible polypeptides related to eukaryotic chlorphyll a/b-binding proteins results in aberrant photoacclimation in *Synechocystis* PCC6803. *Biochim Biophys Acta* **1557:** 21–33.

Hendrix, R.W. (1999) Evolution: the long evolutionary reach of viruses. *Curr Biol* **9:** R914–R917.

Hendrix, R.W., Smith, M.C., Burns, R.N., Ford, M.E., and Hatfull, G.F. (1999) Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. *Proc Natl Acad Sci USA* **96:** 2192–2197.

Holtman, C.K., Chen, Y., Sandoval, P., Gonzales, A., Nalty, M.S., Thomas, T.L., *et al*. (2005) High-throughput functional analysis of the *Synechococcus elongatus* PCC 7942 genome. *DNA Res* **12:** 103–115.

Huang, S., Wilhelm, S.W., Jiao, N., and Chen, F. (2010) Ubiquitous cyanobacterial podoviruses in the global oceans unveiled through viral DNA polymerase gene sequences. *ISME J* **4:** 1243–1251.

Johnson, P.W., and Sieburth, J.M. (1979) Chroococcoid cyanobacteria in the sea: a ubiquitious and diverse phototrophic biomass. *Limnol Oceanogr* **24:** 928–935.

Juhala, R.J., Ford, M.E., Duda, R.L., Youlton, A., Hatfull, G.F., and Hendrix, R.W. (2000) Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. *J Mol Biol* **299:** 27–51.

Kettler, G.C., Martiny, A.C., Huang, K., Zucker, J., Coleman, M.L., Rodrigue, S., *et al*. (2007) Patterns and implications of gene gain and loss in the evolution of *Prochlorococcus*. *PLoS Genet* **3:** e231.

Krumsiek, J., Arnold, R., and Rattei, T. (2007) Gepard: a rapid and sensitive tool for creating dotplots on genome scale. *Bioinformatics* **23:** 1026–1028.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., *et al*. (2007) ClustalW2 and ClustalX version 2. *Bioinformatics* **23:** 2947–2948.

Lawrence, J.G., Hatfull, G.F., and Hendrix, R.W. (2002) Imbroglios of viral taxonomy: genetic exchange and failings of phenetic approaches. *J Bacteriol* **184:** 4891–4905.

Lindell, D., Sullivan, M.B., Johnson, Z.I., Tolonen, A.C., Rohwer, F., and Chisholm, S.W. (2004) Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. *Proc Natl Acad Sci USA* **101:** 11013–11018.

Lindell, D., Jaffe, J.D., Johnson, Z.I., Church, G.M., and Chisholm, S.W. (2005) Photosynthesis genes in marine viruses yield proteins during host infection. *Nature* **438:** 86–89.

Lindell, D., Jaffe, J.D., Coleman, M.L., Futschik, M.E., Axmann, I.M., Rector, T., *et al*. (2007) Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. *Nature* **449:** 83–86.

Liu, X., Shi, M., Kong, S., Gao, Y., and An, C. (2007) Cyanophage Pf-WMP4, a T7-like phage infecting the freshwater cyanobacterium *Phormidium foveolarum*: complete genome sequence and DNA translocation. *Virology* **366:** 28–39.

Liu, X., Kong, S., Shi, M., Fu, L., Gao, Y., and An, C. (2008) Genomic analysis of freshwater cyanophage Pf-WMP3 infecting cyanobacterium *Phormidium foveolarum*: the conserved elements for a phage. *Microb Ecol* **56:** 671–680.

Lowe, T.M., and Eddy, S.R. (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* **25:** 955–964.

Lu, J., Chen, F., and Hodson, R.E. (2001) Distribution, isolation, host specificity, and diversity of cyanophages infecting marine *Synechococcus* spp. in the Georgia river estuaries. *Appl Environ Microbiol* **67:** 3285–3290.

Lukashin, A., and Borodovsky, M. (1998) GeneMark.hmm: new solutions for gene finding. *Nucleic Acids Res* **26:** 1107–1115.

McDaniel, L., and Paul, J.H. (2005) Effect of nutrient addition and environmental factors on prophage induction in natural populations of marine *Synechococcus* species. *Appl Environ Microbiol* **71:** 842–850.

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

Mann, N.H., Cook, A., Millard, A., Bailey, S., and Clokie, M. (2003) Bacterial photosynthesis genes in a virus. *Nature* **424:** 741.

Mann, N.H., Clokie, M.R.J., Millard, A., Cook, A., Wilson, W.H., Wheatley, P.J., *et al*. (2005) The genome of S-PM2, a 'Photosynthetic' T4-type bacteriophage that infects marine *Synechococcus* strains. *J Bacteriol* **187:** 3188–3200.

Marston, M.F., and Sallee, J.L. (2003) Genetic diversity and temporal variation in the cyanophage community infecting marine *Synechococcus* species in Rhode Island's coastal waters. *Appl Environ Microbiol* **69:** 4639–4647.

Millard, A.D., Clokie, M.R., Shub, D.A., and Mann, N.H. (2004) Genetic organization of the *psb*AD region in phages infecting marine *Synechococcus* strains. *Proc Natl Acad Sci USA* **101:** 11007–11012.

Millard, A.D., Zwirglmaier, K., Downey, M.J., Mann, N.H., and 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.

Mühling, M., Fuller, N.J., Millard, A., Somerfield, P.J., Marie, D., Wilson, W.H., *et al*. (2005) Genetic diversity of marine *Synechococcus* and co-occurring cyanophage communities: evidence for viral control of phytoplankton. *Environ Microbiol* **7:** 499–508.

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.

Palenik, B., Brahamsha, B., McCarren, J., Waterbury, J., Allen, E., Webb, E.A., *et al*. (2003) The genome of a motile marine *Synechococcus*. *Nature* **424:** 1037–1041.

Paul, J.H. (2008) Prophages in marine bacteria: dangerous molecular time bombs or the key to survival in the seas? *ISME J* **2:** 579–589.

Pedulla, M.L., Ford, M.E., Houtz, J.M., Karthikeyan, T., Wadsworth, C., Lewis, J.A., *et al*. (2003) Origins of highly mosaic mycobacteriophage genomes. *Cell* **113:** 171–182.

Pope, W.H., Weigele, P.R., Chang, J., Pedulla, M.L., Ford, M.E., Houtz, J.M., *et al*. (2007) Genome sequence, structural proteins, and capsid organization of the cyanophage Syn5: a 'horned' bacteriophage of marine *Synechococcus*. *J Mol Biol* **368:** 966–981.

Rohwer, F., and Edwards, R. (2002) The phage proteomic tree: a genome-based taxonomy for phage. *J Bacteriol* **184:** 4529–4535.

Rusch, D.B., Halpern, A.L., Sutton, G., Heidelberg, K.B., Williamson, S., Yooseph, S., *et al*. (2007) The sorcerer II global ocean sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol* **5:** e77.

Scanlan, D.J., and West, N.J. (2002) Molecular ecology of the marine cyanobacterial genera *Prochlorococcus* and *Synechococcus*. *FEMS Microbiol Ecol* **40:** 1–12.

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.

Sharon, I., Alperovitch, A., Rohwer, F., Haynes, M., Glaser, F., Atamna-Ismaeel, N., *et al*. (2009) Photosystem I gene cassettes are present in marine virus genomes. *Nature* **461:** 258–262.

Short, C.M., and Suttle, C.A. (2005) Nearly identical bacteriophage structural gene sequences are widely distributed in both marine and freshwater environments. *Appl Environ Microbiol* **71:** 480–486.

Stamatakis, A. (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22:** 2688–2690.

Stamatakis, A., Hoover, P., and Rougemont, J. (2008) A rapid bootstrap algorithm for the RAxML web-servers. *Syst Biol* **57:** 758–771.

Sugita, C., Ogata, K., Shikata, M., Jikuya, H., Takano, J., Furumichi, M., *et al*. (2007) Complete nucleotide sequence of the freshwater unicellular cyanobacterium *Synechococcus elongatus* PCC 6301 chromosome: gene content and organization. *Photosynth Res* **93:** 55–67.

Sullivan, M.B., Waterbury, J.B., and Chisholm, S.W. (2003) Cyanophages infecting the oceanic cyanobacterium *Prochlorococcus*. *Nature* **424:** 1047–1051.

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

Sullivan, M.B., Lindell, D., Lee, J.A., Thompson, L.R., Bielawski, J.P., and Chisholm, S.W. (2006) Prevalence and evolution of core photosystem II genes in marine cyanobacterial viruses and their hosts. *PLoS Biol* **4:** e234.

Sullivan, M.B., Coleman, M.L., Quinlivan, V., Rosenkrantz, J.E., DeFrancesco, A.S., Tan, G., *et al*. (2008) Portal protein diversity and phage ecology. *Environ Microbiol* **10:** 2810–2823.

Sullivan, M.B., Krastins, B., Hughes, J.L., Kelly, L., Chase, M., Sarracino, D., *et al*. (2009) The genome and structural proteome of an ocean siphovirus: a new window into the cyanobacterial 'mobilome. *Environ Microbiol* **11:** 2935–2951.

Sullivan, M.B., Huang, K.H., Ignacio-Espinoza, J.C., Berlin, A.M., Kelly, L., Weigele, P.R., *et al*. (2010) Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments. *Environ Microbiol* **12:** 300–312.

Suttle, C.A. (2000) Cyanophages and their role in the ecology of cyanobacteria. In *The Ecology of Cyanobacteria: Their Diversity in Time and Space*. Whitton, B.A., and Potts, M. (eds). Boston, MA, USA: Kluwer Academic Publishers, pp. 563–589.

Suttle, C.A., and Chan, A.M. (1993) Marine cyanophages infecting oceanic and coastal strains of *Synechococcus*: abundance, morphology, cross-reactivity 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. *et al*. (2011) Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism. *Proc Natl Acad Sci USA* **108:** E757–E764.

Wang, K., and Chen, F. (2008) Prevalence of highly host-specific cyanophages in the estuarine environment. *Environ Microbiol* **10:** 300–312.

Waterbury, J.B., and Valois, F.W. (1993) Resistance to co-occuring phages enables marine *Synechococcus* communities to coexist wth cyanophages abundant in seawater. *Appl Environ Microbiol* **59:** 3393–3399.

Waterbury, J.B., Watson, S.W., Guillard, R.R.L., and Brand, L.E. (1979) Widerspread occurrence of a unicellular, marine, planktonic cyanobacterium. *Nature* **277:** 293–294.

Weigele, P.R., Pope, W.H., Pedulla, M.L., Houtz, J.M., Smith, A.L., Conway, J.F., *et al*. (2007) Genomic and structural analysis of Syn9, a cyanophage infecting marine *Prochlorococcus* and *Synechococcus*. *Environ Microbiol* **9:** 1675–1695.

Wilhelm, S.W., Carberry, M.J., Eldridge, M.L., Poorvin, L., Saxton, M.A., and Doblin, M.A. (2006) Marine and freshwater cyanophages in a Laurentian Great Lake: evidence from infectivity assays and molecular analyses of *g20* genes. *Appl Environ Microbiol* **72:** 4957–4963.

Williams, K.P. (2002) Integration sites for genetic elements in prokaryotic tRNA and tmRNA genes: sublocation preference of integrase subfamilies. *Nucleic Acids Res* **30:** 866–875.

Williamson, S.J., Rusch, D.B., Yooseph, S., Halpern, A.L., Heidelberg, K.B., Glass, J.I. *et al*. (2008) The Sorcerer II Global Ocean Sampling Expedition: metagenomic characterization of viruses within aquatic microbial samples. *PLoS One* **3:** e1456.

Wilson, W.H., Joint, I.R., Carr, N.G., and Mann, N.H. (1993) Isolation and molecular characterization of five marine cyanophages propagated on *Synechococcus* sp. strain WH7803. *Appl Environ Microbiol* **59:** 3736–3742.

Yoshida, T., Nagasaki, K., Takashima, Y., Shirai, Y., Tomaru, Y., Takao, Y., *et al*. (2008) Ma-LMM01 infecting toxic *Microcystis aeruginosa* illuminates diverse cyanophage genome strategies. *J Bacteriol* **190:** 1762–1772.

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

Zhao, Y.L., Wang, K., Ackermann, H.W., Halden, R.U., Jiao, N.Z., and Chen, F. (2010) Searching for a 'hidden' prophage in a marine bacterium. *Appl Environ Microbiol* **76:** 589–595.

Zhong, Y., Chen, F., Wilhelm, S.W., Poorvin, L., and Hodson, R.E. (2002) Phylogenetic diversity of marine cyanophage isolates and natural virus communities as revealed by sequences of viral capsid assembly protein gene *g20*. *Appl Environ Microbiol* **68:** 1576–1584.

Supplemental information

Additional Supporting Information may be found in the online version of this article:

Fig. S1. Transmission electron micrographs of *Synechococcus* siphovirus S-CBS2 (A), S-CBS1 (B), S-CBS3 (C) and S-CBS4 (D). The bar length is equivalent to 100 nm.

Fig. S2. Genomic dot plots for marine cyanophage whole genomes. (A) Dot-plot map for all the currently genome-sequenced marine cyanosiphoviruses; (B) for cyanomyoviruses; (C) for cyanopodoviruses. The program Gepard (http://mips.gsf.de/services/analysis/gepard) was used to generate dot plots between two DNA sequences. All the genome sequences of a cyanophage family were concatenated and the concatenated sequence was compared to itself ('self-plot') with the default 'DNA' matrix and word length of 10 bp.

Fig. S3. Alignment of 34 bp possible integration region within the putative tRNA-Thr genes of S-CBS4 (shaded in yellow), its host *Synechococcus* CB0101 (boxed in red) and other cyanophages and picocyanobacteria. S-CBS4 sequence was set as reference and highlighted by colourful background and nucleotides different with S-CBS4 in other source were also coloured. The base-pair numbers in the end of sequence labels indicate the base pair identical to S-CBS4.

Fig. S4. Phylogenetic analyses showing the relationships of four phage proteins associated with DNA metabolism, replication and integration in cyanobacteria and cyanophages: (A) thymidylate synthase, (B) integrase, (C) ribonucleotide reductase and (D) dCTP deaminase. Sequences were aligned using Clustal X2 and phylogenetic analyses were performed using MEGA 5.02. Neighbour-joining trees were constructed with Poisson model, uniform rates among sites and 1000-replication bootstrap test. Maximum likelihood analyses, using WAG + F model and Gamma distribution rates among sites, were complemented to test the clustering, with bootstrap of 100 replications (data not shown). Marine picocyanobacteria *Prochlorococcus* and *Synechococcus* were labelled as their affiliation determined by 16S rDNA similarity (fig. 1 in the reference Scanlan *et al*., 2009), with six *Prochlorococcus* clades HLI, HLII and LLI-LLIV and three *Synechococcus* subclusters 5.1, 5.2 and 5.3 and 10 clades I–X in subcluster 5.1.

Table S1. All the predicted ORFs of *Synechococcus* phage S-CBS2.

Table S2. All the predicted ORFs of *Synechococcus* phage S-CBS1.

Table S3. All the predicted ORFs of *Synechococcus* phage S-CBS3.

Table S4. All the predicted ORFs of *Synechococcus* phage S-CBS4.

Table S5. Summary of 29 cyanophage genomes.

Table S6. Core genome shared by seven cyanopodoviruses.

Table S7. Re-annotation result of partial genome of *Synechococcus* sp. PCC 6301 from ORF *syc0777_c* to ORF *syc0800_c*.

Table S8. BLASTP results of five cyanosiphovirus genomes (all the ORFs) against a Mediterranean Deep Chlorophyll Maximum metagenomic fosmid clone library dataset constructed by Ghai and co-workers (Ghai *et al*., 2010).

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.