Prevalence of psbA-containing cyanobacterial podoviruses in the ocean

Qiang Zheng1, Nianzhi Jiao1, Rui Zhang1, Feng Chen2 & Curtis A. Suttle3

1State Key Laboratory for Marine Environmental Science, Institute of Marine Microbes and Ecospheres, Xiamen University, Xiamen 361005, People’s Republic of China, 2Institute of Marine and Environmental Technology, University of Maryland Center for Environmental Science, Baltimore, MD 21202, USA, 3Departments of Earth, Ocean and Atmospheric Sciences, Microbiology and Immunology, Botany, and the Biodiversity Research Centre, The University of British Columbia, Vancouver, British Columbia, Canada.

Podoviruses that infect marine picocyanobacteria are abundant and could play a significant role on regulating host populations due to their specific phage-host relationship. Genome sequencing of cyanophages has unveiled that many marine cyanophages encode certain photosynthetic genes like psbA. It appears that psbA is only present in certain groups of cyanopodovirus isolates. In order to better understand the prevalence of psbA in cyanobacterial podoviruses, we searched the marine metagenomic database (GOS, BATS, HOT and MarineVirome). Our study suggests that 89% of recruited cyanopodovirus scaffolds from the GOS database contained the psbA gene, supporting the ecological relevance of the photosynthesis gene for surface oceanic cyanophages. Diversification between Clade A and B are consistent with recent finding of two major groups of cyanopodoviruses. All the data also shows that Clade B cyanopodoviruses dominate the surface ocean water, while Clade A cyanopodoviruses become more important in the coastal and estuarine environments.

Viruses are abundant in the ocean and can influence population dynamics and genetic diversity of their hosts1–5. Cyanophage are a specific group of viruses which infect cyanobacteria mainly including Prochlorococcus and Synechococcus. Many cyanophages have been isolated, and all the known marine cyanophages belong to three phage families: Myoviridae, Siphoviridae and Podoviridae4–10. Recent studies showed that cyanopodoviruses might make up 50% of cyanophage community in the sea11–12, suggesting that cyanopodoviruses interact actively with cyanobacteria in the marine environment.

Currently, nearly 40 cyanophage genomes have been sequenced, and half of them are cyanomyoviruses. Cyanomyoviruses have a relatively large genome size and acquire many accessory metabolic genes via horizontal gene transfer (HGT), which constitute the large reservoir of genetic diversity pool13–19. Five genome sequences of cyanosiphoviruses have been reported with genome size ranging from 30,332 to 105,532 bps20–21. Compared to cyanomyoviruses and cyanosiphoviruses, cyanopodoviruses have a relatively conserved genome size ranging from 42,257 to 47,872 bps11,22–25.

Genome sequencing of marine cyanophages has shown that many marine cyanophages encode photosynthesis genes. All the isolated cyanomyoviruses and more than half of the isolated cyanopodoviruses were detected to contain the key photosystem II reaction centre gene psbA in their genomes11,13,17–19,23,26–29, while no psbA gene was found among the known cyanosiphoviruses20,30. Two recent studies showed that 24 of 39 marine cyanopodovirus isolates contained psbA12 and 8 of 12 sequenced cyanopodovirus genomes encoded psbA13. In these two studies, the frequency of psbA-containing podoviruses was estimated based on isolated cyanophages which could be biased by the host used for isolation. Is it possible to quantify the presence of psbA in cyanopodoviruses in the ocean using a culture-independent approach? The metagenomic database is a useful tool, however these datasets in the public domain are also limited and may not represent true community composition.

Results

In this study, we estimated the relative abundance and distribution of psbA-containing podoviruses based on the metagenomic data. Our approach is built on a conserved genomic structure of cyanopodoviruses. Cyanopodovirus genome organization can be divided into three parts: structural genes, nucleotide metabolism related genes and some hypothetical genes regions (Fig. 1)11,18,22–24. Both the composition and the arrangement of structural genes are conserved. One gene cluster, the “portal-capsid-tail/fiber”, existed in all cyanopodoviruses, as well as in other T7 phages31. Interestingly, the psbA gene was commonly located at a fixed position within the...
conserved gene cluster “portal-psbA-capsid”\textsuperscript{11}. Based on this conserved gene cluster, we searched (BLAST) the GOS scaffold database using portal, capsid assembly, psbA and major capsid protein (MCP) genes, and successfully retrieved 79 cyanopodoviral scaffolds from the GOS database.

Among the 79 cyanopodovirus scaffolds, 70 contain psbA and 9 have no psbA. All the MCP sequences (>200 aa) were used to construct the phylogenetic tree. The MCP based phylogeny separated cyanopodoviruses into two major clades (Clade A and B) (Fig. 2), which is consistent with the phylogenetic relationship based on the DNA polymerase gene\textsuperscript{10,12,21,32}. Nearly all cyanopodoviruses in Clade B carry the psbA gene whereas none of those in Clade A do (Fig. 2). A recent study also illustrated such psbA distribution pattern in cyanopodoviruses\textsuperscript{12}.

In the Bermuda (BATS) database, 58 Clade B MCP homologs were recruited, but no Clade A MCP was found (Fig. 3A). We recruited 17 Clade B homologs, but no Clade A homologs from the North Pacific (HOT) database (Fig. 3A). In the GOS database, 729 Clade B MCP homologs and 18 Clade A MCP homologs were found (Fig. 3A). Interestingly, 17 of 18 of reads were recruited from the coastal water. It is likely that most of Clade A like sequences are from the podoviruses infecting marine Synechococcus\textsuperscript{10,33,34}. In the MarineVirome database, 271 Clade B like MCP sequences and 4 Clade A like MCP sequences were detected (Fig. 3A).

**Discussion**

Podoviruses in Clade A could be a transitional group between Clade B and other T7-like non-cyanobacterial podoviruses (Fig. 2). Four scaffolds in Clade B do not contain psbA, and the psbA gene in these four scaffolds might be lost during the evolution. Interestingly, scaffold JCVI_SCAF_109662694693 (in Clade B) contains a high light-induced gene (hli), but no psbA.

Our analysis suggests that Clade A podoviruses only make up a very small proportion of cyanopodoviruses in the surface ocean. In the open ocean, Clade A podoviruses only account for 0.27% and 0.08% of total contigs and scaffolds, respectively.

---

**Figure 1** | The structure and organization of cyanopodoviruses and some scaffolds or contigs.

**Figure 2** | The neighbor-joining tree based on the MCP sequences. The sequences with red color mean scaffolds or cyanophage genomes without psbA genes. Values of >50% are shown, and indicate percentage bootstrap support based on 1000 replicates for distance, maximum parsimony (MP) and minimum evolution (ME) analyses in the order of NJ/MP/ME. Scale bar, 0.1 nucleotide substitution per site.
Based on the cyanopodovirus genomic conserved gene cluster “portal-psba-capsid”, we searched (BLAST) the GOS scaffold database using portal, capsid assembly, psba and major capsid protein (MCP) genes using a reciprocal best-hit BLAST strategy but no e-value cutoff limitation (Fig. 1)\(^1\). The structural genes (portal, MCP or capsid assembly gene) allowed the identification of cyanopodoviruses via searching against the NCBI non-redundant proteins database.

To analyze the occurrence frequency and geographic pattern of cyanopodoviruses in the ocean, we recruited reads from BATS, GOS, HOT and MarineVirome datasets using all MCP sequences from sequenced cyanopodoviral genomes as published in Labrie’s paper\(^2\)-\(^4\). Our approach is similar to the methods described by Zhao et al.\(^4\). Briefly, all homologous reads were recruited from binning by e-value cutoff to avoid potential bias, and then each putative hit was extracted and used as a query to search against the NCBI non-redundant proteins database\(^2\). Metagenomic sequences returned a best-hit which could be used to confirm the classification, and all identified reads are listed in Table S1. The number of recruited reads was not normalized, because the method for sampling is different among all the sites and doesn’t target the viruses. However, there should be no bias for cyanopodoviruses with or without psba gene using any methods for sampling.

**Phylogenetic analyses.** All the MCP sequences (>200 aa) were used to construct the phylogenetic tree. Sequences were aligned using Clustal X and phylogenetic trees were constructed using the neighbour-joining, minimum-evolution and maximum-likelihood algorithms of MEGA software 3.0\(^5\). The phylogenetic trees were supported by bootstrap for re-sampling test with 1000 replicates.

Based on the cyanopodovirus genomic conserved gene cluster “portal-psba-capsid”, we searched (BLAST) the GOS scaffold database using portal, capsid assembly, psba and major capsid protein (MCP) genes using a reciprocal best-hit BLAST strategy but no e-value cutoff limitation (Fig. 1)\(^1\). The structural genes (portal, MCP or capsid assembly gene) allowed the identification of cyanopodoviruses via searching against the NCBI non-redundant proteins database.

To analyze the occurrence frequency and geographic pattern of cyanopodoviruses in the ocean, we recruited reads from BATS, GOS, HOT and MarineVirome datasets using all MCP sequences from sequenced cyanopodoviral genomes as published in Labrie’s paper\(^2\)-\(^4\). Our approach is similar to the methods described by Zhao et al.\(^4\). Briefly, all homologous reads were recruited from binning by e-value cutoff to avoid potential bias, and then each putative hit was extracted and used as a query to search against the NCBI non-redundant proteins database\(^2\). Metagenomic sequences returned a best-hit which could be used to confirm the classification, and all identified reads are listed in Table S1. The number of recruited reads was not normalized, because the method for sampling is different among all the sites and doesn’t target the viruses. However, there should be no bias for cyanopodoviruses with or without psba gene using any methods for sampling.

**Phylogenetic analyses.** All the MCP sequences (>200 aa) were used to construct the phylogenetic tree. Sequences were aligned using Clustal X and phylogenetic trees were constructed using the neighbour-joining, minimum-evolution and maximum-likelihood algorithms of MEGA software 3.0\(^5\). The phylogenetic trees were supported by bootstrap for re-sampling test with 1000 replicates.

1. Proctor, L. M. & Fuhrman, J. A. Viral mortality of marine bacteria and cyanobacteria. *Nature* **343**, 60–62 (1990).
2. Fuhrman, J. A. Marine viruses and their biogeochemical and ecological effects. *Nature* **399**, 541–548 (1999).
3. Suttle, C. A. Viruses in the sea. *Nature* **437**, 356–361 (2005).
4. Suttle, C. A. & Chan, A. M. Dynamics and distribution of cyanophages and their effect on marine *Synechococcus* spp. *Appl Environ Microbiol* **60**, 3167–3174 (1994).
5. Waterbury, J. B. & Valois, F. W. Resistance to co-occurring phages enables marine *Synechococcus* communities to coexist with cyanophages abundant in seawater. *Appl Environ Microbiol* **59**, 3393–3399 (1993).
6. Wilson, W. H., Joint, I. R., Carr, N. G. & Mann, N. H. Isolation and molecular characterization of five marine cyanophages propagated on *Synechococcus* sp. strain WH7803. *Appl Environ Microbiol* **59**, 3736–3743 (1993).
7. Liu, J., Chen, F. & Hodson, R. E. Distribution, host specificity, and diversity of cyanophages infecting marine *Synechococcus* spp. in river estuaries. *Appl Environ Microbiol* **67**, 3285–3290 (2001).
8. Marston, M. F. & Salles, J. L. 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 (2003).
9. Sullivan, M. B., Waterbury, J. B. & Chisholm, S. W. Cyanophages infecting the oceanic cyanobacterium *Prochlorococcus*. *Nature* **424**, 1047–1051 (2003).
10. Wang, K. & Chen, F. Prevalence of highly host-specific cyanophages in the estuarine environment. *Environ Microbiol* **10**, 300–312 (2007).
11. Labrie, S. et al. Genomes of marine cyanopodoviruses reveal multiple origins of diversity. *Environ Microbiol* **15**, 1536–1576 (2013).
12. De Reuck-Bird, N. P. et al. Diversity and evolutionary relationships of T7-like podoviruses infecting marine cyanobacteria. *Environ Microbiol* **15**, 1476–1491 (2013).
13. Lindell, D., Sullivan, M. B., Johnson, Z. I., Tolonen, A. C., Rohwer, F. & Chisholm, S. W. Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. *Proc Natl Acad Sci USA* **101**, 11013–11018 (2004).
14. Clokie, M. R. J. et al. Transcription of a ‘photosynthetic’ T4-type phage during infection of a marine cyanobacterium. *Environ Microbiol* **8**, 827–835 (2005).
15. Clokie, M. R. J., Millard, A. D., Mehta, J. V. & Mann, N. H. Virus isolation studies suggest short-term variations in abundance in natural cyanophage populations of the Indian Ocean. *J MAR BIOL ASSOC UK* **86**, 499–506 (2006).
16. Millard, A., Clokie, M. R. J., Shah, D. A. & Mann, N. H. Genetic organization of the psba region in phages infecting marine *Synechococcus* strains. *Proc Natl Acad Sci USA* **101**, 11007–11012 (2004).
17. Millard, A. D., Zwiglmaier, K., Downey, M. J., Mann, N. H. & Scanlan, D. J. Comparative genomics of marine cyanomycoviruses reveals the widespread occurrence of *Synechococcus* host genes localized to a hyperplastic region: implications for mechanisms of cyanophage evolution. *Environ Microbiol* **11**, 2370–2387 (2009).
18. Sullivan, M. B., Coleman, M. L., Weigele, P., Rohwer, F. & Chisholm, S. W. Three *Prochlorococcus* cyanophage genomes: signature features and ecological interpretations. *PLoS Biol* 3, e14 (2005).
19. Sullivan, M. B. et al. Genomic analysis of oceanic cyanobacterial mycoviruses compared with T4-like mycoviruses from diverse hosts and environments. *Environ Microbiol* **12**, 3035–3056 (2010).
20. Sullivan, M. B. et al. The genome and structural proteome of an ocean siphovirus: a new window into the cyanobacterial ‘mobilome’. *Environ Microbiol* **11**, 2935–2951 (2009).

**Methods**

**Metagenomics.** Four metagenomic databases were used to search homologs in our study: three from the bacterial fraction: the Global Ocean Survey database (GOS)\(^6\), the Bermuda database (BATS)\(^7\)-\(^9\), the Hawaii Ocean Time-Series (HOT)\(^10\)-\(^12\), and one viral fraction database: the MarineVirome\(^13\). All databases were obtained from the CAMERA website (http://camera.calit2.net/index.shtml).
21. Huang, S., Wilhelm, S. W., Jiao, N. & Chen, F. Ubiquitous cyanobacterial podoviruses in the global oceans unveiled through viral DNA polymerase gene sequences. ISME J 4, 1243–1251 (2010).
22. Chen, F. & Lu, J. Genomic sequence and evolution of marine cyanophage P60: a new insight on lytic and lysogenic phages. Appl Environ Microbiol 68, 2589–2594 (2002).
23. Sullivan, M. B., Lindell, D., Lee, J. A., Thompson, L. R., Bielawski, J. P. & Chisholm, S. W. Prevalence and evolution of core photosystem II genes in marine cyanobacterial viruses and their hosts. PLoS Biol 4, e234 (2006).
24. Pope, W. H. et al. Genome sequence, structural proteins, and capsid organization of the cyanophage Syn5: a "horned" bacteriophage of marine Synechococcus. J Mol Biol 368, 966–981 (2007).
25. Raytcheva, D. A., Haase-Pettingell, C., Piret, J. M. & King, J. A. Intracellular assembly of cyanophage Syn5 proceeds through a scaffold-containing procapsid. J Virol 85, 2406–2411 (2011).
26. Mann, N. H., Cook, A., Millard, A., Bailey, S. & Clokie, M. Marine ecosystems: bacterial photosynthesis genes in a virus. Nature 424, 741–741 (2003).
27. Lindell, D., Jaffe, J. D., Johnson, Z. I., Church, G. M. & Chisholm, S. W. Photosynthesis genes in marine viruses yield proteins during host infection. Nature 438, 86–89 (2005).
28. Lindell, D. et al. Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. Nature 449, 83–86 (2007).
29. Chenard, C. & Suttle, C. A. Phylogenetic diversity of sequences of cyanophage photosynthetic gene psbA in marine and freshwaters. Appl Environ Microbiol 74, 5317–5324 (2008).
30. 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 14, 540–558 (2012).
31. Seed, K. D. et al. Evidence of a dominant lineage of Vibrio cholerae-specific lytic bacteriophages shed by cholera patients over a 10-year period in Dhaka, Bangladesh. MBio 2, e00334–10 (2011).
32. Chen, F. et al. Diverse and dynamic populations of cyanobacterial podoviruses in the Chesapeake Bay unveiled through DNA polymerase gene sequences. Environ Microbiol 11, 2884–2892 (2009).
33. Zwirglmaier, K. et al. Global phylogeography of marine Synechococcus and Prochlorococcus reveals a distinct partitioning of lineages among oceanic biomes. Environ Microbiol 10, 147–161 (2008).
34. Scanlan, D. J. et al. Ecological genomics of marine picocyanobacteria. Microbiol Mol Biol Rev 73, 249–299 (2009).
35. Rusch, D. B. et al. The Sorcerer II Global Ocean Sampling expedition: Northwest Atlantic through Eastern Tropical Pacific. PLoS Biol 5, 398–431 (2007).
36. Ghai, R. et al. Metagenome of the Mediterranean deep chlorophyll maximum studied by direct and fosmid library 454 pyrosequencing. ISME J 4, 1154–1166 (2010).
37. Frias-Lopez, J. et al. Microbial community gene expression in ocean surface waters. Proc Natl Acad Sci USA 105, 3805–3810 (2008).
38. Coleman, M. L. & Chisholm, S. W. Ecosystem-specific selection pressures revealed through comparative population genomics. Proc Natl Acad Sci USA 107, 18634–18639 (2010).
39. Angly, F. E. et al. The marine viromes of four oceanic regions. PLoS Biol 4, e368 (2006).
40. Wilhelm, L., Tripp, H. J., Givan, S., Smith, D. & Giovannoni, S. Natural variation in SAR11 marine bacterioplankton genomes inferred from metagenomic data. Bioil Direct 2, 27 (2007).
41. Zhao, Y. et al. Abundant SAR11 viruses in the ocean. Nature 494, 357–360 (2013).
42. Kumar, S., Tamura, K. & Nei, M. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform 5, 150–163 (2004).

Acknowledgments
This work was supported by the 973 Program (2013CB955700, 2011CB808800), the 863 Program (2012AA092003) and the NSFC (41376132). Both QZ and RZ were supported by the Fundamental Research Funds for the Central Universities (2013121051 and 2012121052 respectively). FC was supported by the Xiamen University 111 Program and Hanse-Wissenschaftskolleg Fellowship. Professor John Hodgkiss of The University of Hong Kong is thanked for polishing the English in this manuscript.

Author contributions
Q.Z., N.J. and C.A.S. conceived and designed the experiments; Q.Z., R.Z. and F.C. analyzed the data. All the authors helped in writing the manuscript, discussed the results and commented on the manuscript.

Additional information
Supplementary information accompanies this paper at http://www.nature.com/srep.

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Zheng, Q., Jiao, N., Zhang, R., Chen, F. & Suttle, C. Prevalence of psbA-containing cyanobacterial podoviruses in the ocean. Sci. Rep. 3, 3207; DOI:10.1038/srep03207 (2013).

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported license. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/3.0.