YAN ZHONG
Phylogenetic Diversity of Marine Cyanophages as Revealed by Sequences of Viral Capsid Assembly Protein Gene g20
(Under the Direction of Dr. ROBERT E. HODSON)

In order to characterize the phylogenetic diversity of marine cyanophages, oligonucleotide primers were designed to specifically amplify ca. 592 bp fragments of the viral capsid assembly gene g20. Phylogenetic analysis of isolates revealed that cyanophages were highly diverse and widely distributed without significant geographical segregation. Cloning and sequencing analysis of 6 natural virus concentrates revealed 9 phylogenetic groups in 114 different g20 homologs. The structure of cyanophage communities in the estuary and the open ocean samples were different from each other, with unique phylogenetic clusters found for each environment. Changes in clonal diversity were also observed from surface waters to deep chlorophyll maximum (DCM) layer in the open ocean. Only three clusters contained known cyanophage isolates, while the identity of the other six clusters remained unknown. The high cyanophage diversity revealed by the g20 sequences suggests that marine viruses can potentially play important roles in regulating microbial genetic diversity.

INDEX WORDS: Phylogenetic diversity, Cyanophage, Viral capsid assembly protein gene, g20,

PHYLOGENETIC DIVERSITY OF MARINE CYANOPHAGES AS REVEALED BY SEQUENCES OF VIRAL CAPSID ASSEMBLY PROTEIN GENE g20

by

YAN ZHONG

B.S. Wuhan University, P.R.China, 1991

M.S. Jinan Univerisity, P.R.China, 1994

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2001

PHYLOGENETIC DIVERSITY OF MARINE CYANOPHAGES AS REVEALED BY SEQUENCES OF VIRAL CAPSID ASSEMBLY PROTEIN GENE g20

by

YAN ZHONG

Approved:

Major Professor: Robert E. Hodson

Committee: Mary Ann Moran
Brian J. Binder

Electronic Version Approved:

Gordhan L. Patel
Dean of the Graduate School
The University of Georgia
August 2001

DEDICATION

Trusts in the LORD with all your heart

and do not lean on your own understanding

In all your ways acknowledge Him

and He will make your paths straight

Proverbs 3:5-6

ACKNOWLEDGEMENTS

I am greatly indebted to my direct mentor, Dr. Feng Chen, for supporting me throughout this project and for all the guidance, enthusiasm, and advice he gave me during my graduate education here.

I am grateful to my major professor, Dr. Robert E. Hodson for his advice and guidance in this project. I thank Dr. Mary Ann Moran and Dr. Brian J. Binder for serving on my committee and contributing their advice and comments to this project.

Thanks are due to Wendy Dustman and Kathryn Dudeck; I enjoyed working with you and won't forget the happy time we spent together. I sincerely thank the following individuals: Jingrang Lu, Edward Sheppard, Dave Bachoon, Jim Sullivan, Xiaozhen Mou, Eric Wommack, Wenying Ye, Alison Buchan, for their friendship and various help in my graduate research.

I would like to express my heartfelt thanks to my family in the Chinese Student Bible Fellowship: Susanne and Johnny, Alison and Jon, Darren and Heather, Ansley, Xiaojia, Hui, Lin Lin, Zhang Fan, Hua Jun, Yali, Jie and Haojie, Henry and Xiao. I cherish your love, care and encouragement, and am grateful for being in a family with you all.

Finally I would especially like to thank my husband and daughter, my enduring source of support and encouragement both throughout and beyond my graduate work presented here.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

The existence of viruses in the marine environment has been known for nearly half a century (e.g. Spencer 1955). However, it was not until the recent discovery of high viral abundance in the sea (Bergh et al. 1989) that interest has been focused on characterizing their role in marine ecology. Previous studies have demonstrated that viruses can be responsible for 10-50% of bacterial mortality (Fuhrman 1999, Fuhrman and Noble 1995, Binder 1999) and substantial reduction in marine primary productivity (Suttle et al. 1990, Suttle 1992). As viral lysis of sensitive cells short-circuits carbon flow to higher trophic levels, viruses appear to have the potential to significantly affect nutrient cycling and energy flow in marine ecosystems (Fuhrman 1999, Wilhelm and Suttle 1999, Thingstad et al. 1993). Furthermore, high phage-host interaction rates suggest that viruses may mediate genetic exchange among microorganisms in natural aquatic environments and may be an important factor in shaping the genetic structure and diversity of marine microbial communities (Bergh et al. 1989, Proctor and Fuhrman 1990, Fuhrman 1999, Wommack and Colwell 2000).

While these studies reveal some of the potential viral impacts on marine foodwebs, much remains to be elucidated. Furthermore, it is impossible to understand the specific role of individual viral types if they are only studied as a collective unit in complicated natural environments where diverse and dynamic host-phage systems co-exist. Marine cyanophages (viruses that infect cyanobacteria) and their host *Synechococcus* spp. may be good model systems to study the role of bacteriophages in the sea. The host *Synechococcus* cyanobacteria are important primary producers, and are reportedly responsible for 5 to 25% of carbon fixation in temperate and tropical oceans (Waterbury et al. 1986). They have been characterized extensively from perspectives of physiology,

ecology and genetics, and a large collection of axenic strains is available for phage isolation. *Synechococcus* cyanophages have been shown to be ubiquitous in the world's oceans, sometimes present at concentrations as high as 10^6-10^8 per liter (Suttle and Chan 1994, Waterbury and Valois 1993). It has been estimated that ca. 2-8% of *Synechococcus* cells are infected by cyanophages on a daily basis (Proctor and Fuhrman 1990, Suttle and Chan 1994, Garza and Suttle 1998). The fact that abundant and infective populations of cyanophages co-exist with their host populations suggests that they may exert considerable regulation on *Synechococcus* spp. populations. Thus, the study of cyanophages could prove to be insightful for characterizing the role of viruses in the marine microbial community.

Freshwater cyanophages have been extensively studied since the early 1960s with regard to the control of nuisance algal blooms and the development of genetic tools (e.g. Safferman and Morris 1963, Padan and Shilo 1973, Safferman 1973, Desjardins and Olson 1983). However, there has been little systematic effort to document the occurrence and importance of cyanophages in marine systems until very recently. Proctor and Fuhrman (1990) reported that 5% of marine *Synechococcus* spp. contained mature phages that might cause significant mortality to cyanobacteria populations. In the following sections literature relevant to marine cyanophage research will be reviewed, after which the purpose and approach of this study will be described briefly.

I. Literature in marine cyanophage research. Due to the current early stage of studies of marine cyanophages, only a few documents are available and these focus on the following aspects:

(i) Abundance and distribution. Since the first report that marine cyanophages could be easily isolated from seawater (Suttle et al. 1990), cyanophages have been isolated from various marine provinces including the Gulf Stream, the Sargasso Sea, the English Channel, the Yellow Sea, the Japanese Sea as well as Georgia river estuaries (Waterbury and Valois 1993, Suttle and Chan 1993, Wilson et al. 1993, Lu et al. 2001).

Cyanophages are present both inshore and offshore, in the surface and in deep waters of the euphotic zone in various marine habitats. Considering the ubiquitous distribution of host *Synechococcus* populations (Waterbury et al. 1986), it is reasonable to assume that *Synechococcus* cyanophages are also widely occurring in the world's oceans.

High abundance of cyanophage populations has been found in the marine environment by using classical culture-dependant most-probable-number techniques. Densities between 10^6 to 10^8 per liter are often observed, and higher densities (10^9 per liter) have also been reported during *Synechococcus* blooms (Suttle et al. 1996). The densities of cyanophages follow the general trend of overall marine viral abundance, typically one order of magnitude less than that of the host, and higher in the surface and coastal waters as compared to deep and offshore waters (Suttle and Chan 1993, 1994, Waterbury and Valois 1993). The abundance of cyanophages is tied to that of *Synechococcus* populations, with highest phage titers associated with greatest host concentrations (Suttle and Chan 1993; Wilson et al. 2000). Environmental factors such as temperature and salinity can also influence cyanophage abundance. There is a positive relationship between cyanophage concentrations and temperature (Suttle and Chan 1993) and between cyanophage density and salinity (Lu et al. 2001). Additionally, marine cyanophages infecting unicellular cyanobacteria are much more abundant than those found in freshwater systems (Waterbury and Valois 1993).

(ii) Morphology. Under transmission electron microscopy (TEM), different morphologies of cyanophages can be characterized. To date, all the known marine cyanophages have unique, structured tails and belong to the order *Caudovirales*. There is considerable variation among cyanophages in the size of the capsid and in the structure of the tail. Depending on tail length and contractility, marine cyanophages are assigned to any of three families: Myoviridae with long, contractile tails; Podoviridae with short tails; and Siphoviridae (formerly Styloviridae) with long, noncontractile tails (Francki et al. 1995). Besides these obvious differences, marine cyanophages also have a variety of

high-level morphological features, such as tail appendages and unusual bases, that are associated with structural variation within each family. Solely on the basis of observed morphological characteristics, marine cyanophages are considered very diverse. It has also been reported that the vast majority of marine cyanophages observed to date belong to the family Myoviridae (Waterbury and Valois 1993, Lu et al. 2001).

(iii) Host specificity and range. Like other viruses, marine cyanophages are highly host-specific. Cyanophages propagated on a *Synechococcus* strain can only infect that strain or genetically closely related strains (Lu et al. 2001). Marine cyanophages cannot readily infect host cells from freshwater systems (Suttle and Chan 1993). The high host specificity suggests low contact incidence between phages and suitable hosts, therefore high densities of cyanophage and cyanobacteria are required to maintain high cyanophage diversity.

Host range among marine cyanophages is complicated by unclear taxonomic relationships among marine *Synechococcus*. The current taxonomy of unicellular cyanobacteria has assigned most marine *Synechococcus* to two clusters: Marine Cluster A, whose members include both oceanic and coastal strains and use phycoerythrin as primary light harvesting pigment; and Marine Cluster B, which are exclusively coastal, phycocyanin-containing strains (Waterbury and Rippka 1989). Cyanophages infecting Marine Cluster A have a broader host range than those infecting members in Marine Cluster B (Waterbury and Valois 1993, Lu et al. 2001). It is relatively easier to isolate cyanophages that infect *Synechococcus* strains in Marine Cluster A than those in Marine Cluster B (Lu et al. 2001). Interestingly, some cyanophages which infect Marine Cluster A are also able to infect isolates of *Synechococcus* in Marine Cluster B (Suttle and Chan 1993). Cross infectivity of cyanophages may reflect the genetic relatedness of host *Synechococcus* spp. According to the phylogenetic relationship revealed by RuBisCO gene sequences, *Synechococcus* strains in Marine Cluster A are closely related to each other while members in Marine Cluster B appear to be very diverse with some of them

indicating a close relation to freshwater strains of *Synechococcus* (Chen et al., unpublished data). A recent study based on 16S rRNA gene sequence analysis has indicated that the genus of *Synechococcus* is so diverse that it may not be a natural taxon (Honda et al. 1999).

(iv) Infection and decay characteristics. Suttle and Chan (1993) studied the adsorption kinetics of an isolated cyanophage strain on host *Synechococcus* sp. The adsorption rate was slow compared to typical bacteriophages, and infection was light-dependent, with host photosynthesis continuing until the point of cell lysis. Burst size for specific cyanophage isolates varied from 92 to 234 (Suttle and Chan 1994), slightly higher than the value of 81 measured for natural cyanophage communities (Garza and Suttle 1998).

Among the factors responsible for virus decay, solar radiation has been reported to cause most loss of infectivity in natural cyanophage communities in surface seawater. The decay rate of cyanophage infectivity has both been measured in situ and inferred theoretically, and was found to range from 0.048 to 0.75 per day in the surface mixed layer in marine habitats (Suttle and Chan 1994, Garza and Chan 1998). Since the decay rate is a function of solar radiance, cyanophages in surface offshore water decay much faster than those in deeper offshore waters and inshore waters (Suttle and Chan 1994). Moreover, natural cyanophage communities have higher decay rates in summer and early fall, when solar radiation is stronger, than during other seasons (Garza and Suttle 1998). Additionally, cyanophage isolates are poor proxies for modeling UV impact on natural viral communities since their decay rates are substantially higher than those of natural communities (Weinbauer et al. 1999, Garza and Suttle 1998).

(**v) Cyanophage-caused host mortality.** Proctor and Fuhrman (1990) first assessed the mortality of *Synechococcus* populations due to viral infection in a study based on TEM observation of thin-sectioned cells. They observed that 0.8-2.8% of *Synechococcus* in natural marine populations contained mature viruses. Assuming that

assembled phages were visible during 10% of the latent period, they determined that an average of 15% of natural *Synechococcus* populations would be infected on a daily basis. However, this appeared to be an overestimate, since viruses are probably visible for 50% of the lytic cycle (Waterbury and Valois 1993), thus ca. 1.5-6% *Synechococcus* are infected daily (Suttle 1996). This result agrees with calculations via other approaches. Based on theoretical rates of adsorption and measured abundances of *Synechococcus* spp. and their phages, Waterbury and Valois (1993) reported that 0.005 to 3.2% of *Synechococcus* populations were lysed by phages per day. Using virus decay rate and assuming that viral production balances viral removal at steady state, Suttle and Chan (1994) as well as Garza and Suttle (1998) found that ca. 2 to 8% of *Synechococcus* community would have to be infected daily to balance the loss of viral particles. Overall, these estimates suggest that cyanophages infect several percent of natural cyanobacteria communities on a daily basis, therefore lytic cyanophages may not be significant agents of *Synechococcus* mortality.

(vi) Host resistance. The issue of host resistance is of interest because it directly affects the fate of host-virus interaction. Earlier studies with continuous cultures of heterotrophic bacteria and their phages predicted that resistant bacteria should dominate in a steady-state system (e.g. Lenski 1988). Waterbury and Valois (1993) isolated *Synechococcus* and cyanophage from a sample collected in nearshore water and studied cyanobacterial resistance to its co-occurring phages by serial dilution. Since the cyanobacteria in the most diluted treatment (i.e. those that were most abundant) were resistant to phage lysis, they suggested that marine *Synechococcus* communities were dominated by resistant cells. However, Suttle and Chan (1994), by comparing contact rates and rates of infection, found that different scenario of host resistance was present in nearshore and offshore waters. In offshore waters, while the calculated contact frequency between *Synechococcus* and its phages was only 5% per day, the infection rate was comparable (6.6%). This indicated that most contact resulted in infection, hence most

Synechococcus populations in offshore waters were susceptible to phage lysis. Conversely, in nearshore waters, compared to the high contact frequency (83%), only 0.5% of *Synechococcus* populations were infected, suggesting that most cyanobacteria were resistant.

One explanation for the observed difference in host resistance in marine habitats involves phage receptors at the surface of host cells that are essential proteins in nutrient acquisition. In nearshore (nutrient rich) waters where phage-host contact rate is high, viral receptors are only present on host cell surfaces at certain stage of cell cycle to minimize phage adsorption (Hennes et al. 1995, Suttle 2000). On the contrary, the low nutrient concentration in offshore waters results in selection for high levels of nutrient transport proteins. Since phage-host contact rates are lower, there is less selection pressure to avoid infection. Hence, most *Synechococcus* populations in offshore waters are susceptible to viral lysis. This presumption agrees with the report that most oceanic strains of *Synechococcus* are easily lysed, whereas most coastal strains are hard to crash by phages (Lu et al. 2001).

(vii) Genetic diversity. It has been suggested that viruses are in a unique position to shape the genetic diversity of microbial communities, mainly through virus-mediated genetic exchange such as transduction, transformation and lysogenic conversion (Fuhrman and Suttle 1993, Jiang et al. 1998, Fuhrman 1999). Previously, the diversity of isolated cyanophages was studied by TEM and restriction fragment length polymorphism, and a variety of cyanophage morphologies and genotypes were revealed (Waterbury and Valois 1993, Wilson et al. 1993, Lu et al. 2001, Chen et al. 1999). However, little is known about the genetic diversity and phylogenetic affiliation of cyanophages in natural marine environments. The recent discovery of a conserved functional gene g20 in cyanophages greatly facilitates such studies. G20 is an essential gene in coliphage T4 head morphogenesis and encodes capsid assembly protein gp20 that is responsible for the initiation of head assembly (Van Driel and Couture 1978, Marusich

and Mesyanzhinov 1989). Fuller et al. (1998) first reported the g20 sequences of three cyanomyoviurses, and designed a set of primers (CPS1/CPS2) that specifically amplified the g20 gene fragment from cyanomyoviruses in natural seawater. This primer set was subsequently applied in PCR-DGGE to study cyanophage diversity along a transect in the Atlantic Ocean (Wilson et al. 1999, 2000). High genetic diversity was observed in natural virus communities, with 4-5 and 7-10 distinct g20 DGGE fragments found in a single sample in surface and deep waters, respectively. Natural cyanophage community structure changed dramatically from surface to depth, and maximum diversity was always coincident with maximum *Synechococcus* spp. abundances. These results suggested that cyanophages might play an important role in affecting the *Synechococcus* clonal structure via virus selection pressure and genetic exchange (Wilson et al. 2000).

II. Objective and approach of this study. In order to elucidate the impact of cyanophages on the genetic structure and diversity of *Synechococcus* populations, it is imperative to characterize the phylogenetic linkages among cyanophages in nature and the phylogenetic co-variation of cyanophage and *Synechococcus* populations in different marine environments. While some research has revealed diverse phyletic groups of natural marine *Synechococcus* populations (Palenik 1994, Ernst et al. 1995, Toledo and Palenik 1997), little is known about the phylogenetic relatedness and diversity within natural cyanophage communities. Our goal in this study is three-fold: (1) to develop primers that will specifically amplify the g20 gene fragment from cyanomyoviruses. (2) To study the phylogenetic relationship among widely collected marine cyanophage isolates. (3) To explore the phylogenetic diversity and affiliation of natural cyanophage communities in different marine environments. To accomplish these goals, we first designed cyanophage-specific primers according to the g20 gene sequences of three cyanophages available from GenBank (Fuller et al. 1998). Our main approach to studying the genetic diversity of natural cyanophages was via PCR, cloning and sequencing. Such strategies have been used to infer phyletic relatedness of marine algal viruses of the

Phycodnaviridae (Chen et al. 1996) and have proved to be powerful tools to investigate phylogenetic diversity of natural virus communities.

LITERATURE CITED

1. **Bergh, O, K. Y. Borsheim, G. Bratbak, and M. Heldal.** 1989. High abundance of viruses found in aquatic environments. Nature (London) **340:**467-468.
2. **Binder, B.** 1999. Reconsidering the relationship between virally induced bacterial mortality and frequency of infected cells. Aquat. Microb. Ecol. 18:207-215.
3. **Bradley, D. E.** 1967. Ultrastructure of bacteriophages and bacteriocins. Bact. Rev. **31:**230-314.
4. **Chen, F., C. A. Suttle and S. M. Short.** 1996. Genetic diversity in marine algal virus communities as revealed by sequence analysis of DNA polymerase genes. Appl. Environ. Microbiol. **62:**2869-2874.
5. **Chen, F, J. R. Lu, Y. Zhong, and R. E. Hodson.** 1999. Marine cyanophages reveal a high phenotypic and genotypic diversity. Abstract of the General Meeting of the American Society for Microbiology. **99:**
6. **Desjardins, P. R., and G. B. Olson.** 1983. Viral control of nuisance cyanobacteria (blue-green algae): II. Cyanophage strains, stability studies of phages and hosts, and effects of environmental factors on phage-host interactions. California Water Resources Center Contribution No. 185.
7. **Ernst, A., P. Marschall, and C. Postius.** 1995. Genetic diversity among *Synechococcus* spp. (cyanobacteria) isolated from the pelagial of Lake Constance. FEMS Microb. Ecol. **17:**197-204.
8. **Francki, R. I. B., C. M. Fauquet, D. L. Knudson, and F. Brown.** 1995. The Classification and Nomenclature of Viruses, Fifth Report of the International Committee on Taxonomy of Viruses, Archiv. Virol. (Suplementum 2). Springer-Verlag Wien New York.

9. **Fuhrman, J. A.** 1999. Marine viruses and their biogeochemical and ecological effects. Nature (London) **399:**541-547.

10. **Fuhrman, J. A., and C. A. Suttle.** 1993. Viruses in marine planktonic systems. Oceanogr. **6:**50-62.

11. **Fuhrman, J. A., and R. T. Noble.** 1995. Viruses and protists cause similar bacterial mortality in coastal seawater. Limnol. Oceanogr. **40:**1236-1242.

12. **Fuller, N. J., W. H. Wilson, I. R. Joint, and N. H. Mann.** 1998. Occurrence of a sequence in marine cyanophages similar to that of T4 gp20 and its application to PCR-based detection and quantification techniques. Appl. Environ. Microbiol. **64:**2051-2060.

13. **Garza, D. R., and C. A. Suttle.** 1998. The effect of cyanophages on the mortality of *Synechococcus* spp and selection for UV resistant viral communities. Microb. Ecol. **36:**281-292.

14. **Hennes, K. P., A. M. Chan, and C. A. Suttle.** 1995. Fluorescently labeled virus probes show that natural virus populations can control the structure of marine microbial communities. Appl. Environ. Microbiol. **61:**3623-3627.

15. **Jiang, S. C., C. K. Kellogg, and J. H. Paul.** 1998. Characterization of marine temperate phage-host systems isolated from Mamala Bay, Oahu, Hawaii. Appl. Environ. Microbiol. **64:**535-542.

16. **Lenski, R. E.** 1988. Dynamics of interactions between bacterial and virulent bacteriophage. Adv. Microb. Ecol. **10:**1-39.

17. **Lu, J., F. Chen, and R. E. Hodson.** 2001. Distribution, isolation, host specificity, and diversity of cyanophages infecting marine *Synechococcus* spp. in the Georgia river estuaries. Appl. Environ. Microbiol. In press.

18. **Marusich, E. I., and V. V. Mesyanzhinov.** 1989. Nucleotide and deduced amino acid sequences of bacteriophage T4 gene 20. Nucleic Acids Res. **17:**7514.

19. **Padan, E., and M. Shilo.** 1973. Cyanophages: viruses attacking blue-green algae. Bacteriol. Rev. **37:**343-370.

20. **Palenik, B.** 1994. Cyanobacterial community structure as seen from RNA polymerase gene sequence analysis. Appl. Environ. Microbiol. **90:**3212-3219.

21. **Proctor, L. M., and J. A. Fuhrman.** 1990. Viral mortality of marine bacteria and cyanobacteria.. Nature (London) **343:**60-62.

22. **Safferman, R. S.** 1973. The biology of the blue-green algae. Bot. Monogr. **9:**14-237.

23. **Safferman, R. S., and M. E. Morris.** 1963. Algal virus: isolation. Science 140:679.

24. Spencer, R. 1955. A marine bacteriophage. Nature (London) **175:**690-691.

25. **Suttle, C. A.** 1992. Inhibition of photosynthesis in phytoplankton by the submicron size fraction concentrated from seawater. Mar. Ecol. Prog. Ser. **87:**105-112.

26. **Suttle, C. A.** 1996. The significance of viruses to mortality in aquatic microbial communities. Microb. Ecol. **28:**237-243.

27. **Suttle, C. A.** 2000. Cyanophages and their role in the ecology of cyanobacteria, p. 563-589. *In* B. A. Whitton, and M. Potts (ed.), The ecology of cyanobacteria: their diversity in time and space. Kluwer Academic Publishers, Boston.

28. **Suttle, C. A., and A. M. Chan.** 1993. Marine cyanophages infecting oceanic and coastal strains of *Synechococcus*: abundance, morphology, cross-infectivity and growth characteristics. Mar. Ecol. Prog. Ser. **92:**99-109.

29. **Suttle, C. A., and A. M. Chan.** 1994. Dynamics and distribution of cyanophages and their effect on marine *Synechococcus* spp. Appl. Environ. Microbiol. **60:**3167-3174.

30. **Suttle, C. A., A. M. Chan, K. M. Rodda, M. G. Weinbauer, D. R. Garza, and S. W. Wilhelm.** 1996. The effect of cyanophages on *Synechococcus* spp. during a bloom in the western Gulf of Mexico. Eos **76**(suppl):OS207-OS208.

31. **Suttle, A. S., A. M. Chan, and M. T. Cottrell.** 1990. Infection of phytoplankton by viruses and reduction of primary productivity. Nature (London) **347:**467-469.

32. **Thingstad, T. F., M. Heldal, G. Bratbak, and I. Dundas.** 1993. Are viruses important partners in pelagic food webs? Trends Ecol. Evol. **8:**209-212.

33. **Toledo, G., and B. Palenik.** 1997. *Synechococcus* diversity in the California Current as seen by RNA polymerase (rpoC1) gene sequences of isolated strains. Appl. Environ. Microbiol. **63:**4298-4303.

34. **Van Driel, R., and E. Couture.** 1978. Assembly of the scaffolding core of bacteriophage T4 preheads. J. Mol. Biol. **123:**713-719.

35. **Weinbauer, M. G., S. W. Wilhelm, C. A. Suttle, R. J. pledger, and D. L. Mitchell.** 1999. Sunlight-induced DNA damage and resistance in natural viral communities. **17:**111-120.

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

37. **Waterbury, J. B., and R. Rippka.** 1989. The order *Chroococcales*, p. 1728-1746. *In* J. T. Staley, M. P. Bryant, N. Pfennig, and J. G. Holt (ed.), Bergey's manual of systematic bacteriology, vol. 3. Williams and Wilkins, Baltimore, Md.

38. **Waterbury, J. B., S. W. Watson, F. W. Valois, and D. G. Franks.** 1986. Biological and ecological characterization of the marine unicellular cyanobacterium *Synechococcus*. Can. Bull. Fish. Aquat. Sci. **214:**71-120.

39. **Wilhelm, S. W., and C. A. Suttle.** 1999. Viruses and nutrient cycles in the sea: viruses play critical roles in the structure and function of aquatic food webs. Biosci. **49:**781-788.

40. **Wilson, W. H., J. F. Nicholas, I. R. Joint, and N. H. Mann.** 1999. Analysis of cyanophage diversity and population structure in a south-north transect of the Atlantic Ocean. Bulletin de l'Institut Oceanographique (Monaco). **19:**209-216.

41. **Wilson, W. H., J. F. Nicholas, I. R. Joint, and N. H. Mann.** 2000. Analysis of cyanophage diversity in the marine environment using denaturing gradient gel

electrophoresis, p. 565-570. *In* C. R. Bell, M. Brylinsky, and P. Johnson-Green (ed.), Microbial Biosystems: New Frontier. Proceedings of the 8th International Symposium on Microbial Ecology. Halifax, Canada.

42. **Wilson, W. H., I. R. Joint, N. G Carr, and N. H. Mann.** 1993. Isolation and molecular characterization of five marine cyanophages propagated on *Synechococcus* sp. strain WH7803. Appl. Environ. Microbiol. **59:**3736-3743.
43. **Wommack, K. E., and R. R. Colwell.** 2000. Virioplankton: Viruses in Aquatic Ecosystems. Microbiol. Mol. Biol. Rev. **64:** 69-114.

CHAPTER 2

PHYLOGENETIC DIVERSITY OF MARINE CYANOPHAGES AS REVEALED BY SEQUENCES OF VIRAL CAPSID ASSEMBLY PROTEIN GENE g20[1]

[1] Zhong, Y., F. Chen and R.E. Hodson. 2001. Submitted to Applied and Environmental Microbiology, 7/18/01.

ABSTRACT

In order to characterize the genetic diversity and phylogenetic affiliations of marine cyanophage isolates and natural cyanophage assemblages, oligonucleotide primers CPS1/CPS8 were designed to specifically amplify ca. 592 bp fragments of the viral capsid assembly protein gene g20. Phylogenetic analysis of isolated cyanophages revealed that marine cyanophages were highly diverse yet more closely related to each other than to enteric coliphage T4. Genetically related marine cyanophage isolates were widely distributed without significant geographical segregation (i.e. no correlation between genetic variation and geographic distance). Cloning and sequencing analysis of 6 natural virus concentrates from estuarine and oligotrophic offshore environments revealed 9 phylogenetic groups in a total of 114 different g20 homologs, with up to 6 clusters and 29 genotypes encountered in a single sample. The composition and structure of natural cyanophage communities in the estuary and the open ocean samples were different from each other, with unique phylogenetic clusters found for each environment. Changes in clonal diversity were also observed from the surface waters to the deep chlorophyll maximum (DCM) layer in the open ocean. Only three clusters contained known cyanophage isolates, while the identity of the other six clusters remain unknown. Whether or not these unidentified groups are comprised of bacteriophages that infect different groups of *Synechococcus* or other closely related cyanobacteria remains to be answered. The high genetic diversity of marine cyanophage assemblages revealed by the g20 sequences suggests that marine viruses can potentially play important roles in regulating microbial genetic diversity.

INTRODUCTION

Viruses are now known to be an abundant and dynamic component in the marine microbial communities that can regulate biomass, production and species composition of bacteria and phytoplankton, influence biogeochemical cycling, and mediate gene transfer between microorganisms in aquatic ecosystems (Bergh et al. 1989, Proctor and Fuhrman 1990, Suttle et al. 1990, Thingstad et al. 1993, Fuhrman 1999, Wilhelm and Suttle 1999, Wommack 2000). Unicellular cyanobacteria of the genus *Synechococcus* are among the most abundant members of the marine picoplankton and reportedly contribute up to 25% of primary production in the open ocean (Waterbury et al. 1986). Viruses that infect specific strains of marine *Synechococcus* spp. (known as cyanophages) typically range from 10^3-10^5 ml^{-1} in nearshore and offshore waters, and sometimes can reach concentrations in excess of 10^6 ml^{-1} (Waterbury and Valois, 1993, Suttle and Chan 1994, Suttle 2000, Lu et al. 2001). Proctor and Fuhrman (1990) found that 1-3% of *Synechococcus* spp. from a variety of marine environments contained visible phage particles. Estimates made by using different approaches has indicated that viruses may infect 2-8% of the *Synechococcus* population daily (Suttle and Chan 1994, Garza and Suttle 1998).

Viruses also potentially influence the genetic diversity of microbial communities, mainly via virus selection pressure and virus-mediated genetic exchange such as transduction, transformation and lysogenic conversion (Waterbury and Valois 1993, Fuhrman 1999, Jiang et al. 1998). Cyanophage isolates are diverse in terms both of their morphology (Suttle and Chan 1993, Waterbury and Valois 1993) and their genetic fingerprints (Wilson et al. 1993, Lu et al. 2001). However, little is known about the genetic diversity and phyletic linkages of cyanophages in natural marine environments. Recently, the viral capsid assembly protein gene g20 was shown to be conserved among three marine cyanomyoviruses (Fuller et al. 1998). A set of PCR primers (CPS1/CPS2) was designed based on the conserved regions of these cyanophages and used to

specifically amplify a 165bp g20 fragment from cyanomyoviruses in natural seawater (Fuller et al. 1998). This primer set was subsequently applied in PCR-DGGE (denaturing gradient gel electrophoresis) to study cyanophage diversity along a transect in the Atlantic Ocean (Wilson et al. 1999, 2000). High genetic diversity of cyanophage g20 sequences was found in both surface waters and the deep layer of the euphotic zone. Natural cyanophage community structure changed significantly from surface to depth, and maximum diversity was always coincident with maximum *Synechococcus* spp. abundance (Wilson et al. 2000).

It is difficult to resolve the complex interactions between cyanophage and *Synechococcus* populations without knowing their phylogenetic distribution and phyletic variation in different environments. While natural marine *Synechococcus* populations have been well characterized phylogenetically (Palenik 1994, 1996, Toledo and Palenik 1997), little is known about the phyletic relatedness and diversity within natural cyanophage communities in different marine environments. Our earlier study has shown that the short g20 gene sequences flanked by the primers CPS1 and CPS2 (Fuller et al. 1998) is not adequate for phylogenetic inference compared to that flanked by the primers CPS1 and CPS8 (Zhong et al. 2000).

In this study, the primers CPS1/CPS8 were used to amplify the g20 gene fragment (*ca*. 592 bp) from cyanophage isolates and natural viral assemblages. Phylogenetic diversity of cyanophage communities in various marine environments was investigated based on the partial g20 gene sequences recovered from natural viral communities.

MATERIALS AND METHODS

Phage strains. Twenty-four strains of cyanophage isolates and seven strains of bacteriophages that infect heterotrophic bacteria were used in this study (Table 1).

DNA extraction. Phage DNA was extracted using the protocols described by Lu et al. (2001). Briefly, 500 ml fresh lysate from each cyanophage isolate was

centrifuged at 5, 500g and 4°C for 15 minutes to remove debris. The supernatant was filtered through 0.2 μm pore-size membrane filter, and centrifuged at 140,000g and 4°C for 6 hours to pellet viral particles. The subsequent extraction of phage DNA was performed by standard procedures (Sambrook et al. 1989).

Sampling and ultrafiltration. Water samples (50 – 100 liters) were collected from the river estuary and the open ocean (Table 2). SE1 was a surface sample collected from the Pier of the Skidaway Institute of Oceanography in Savannah, Georgia. GS26 and GS27 were from the surface and deep chlorophyll maximum (DCM) of the Gulf Stream along the edge of the Sargasso Sea. SS48, SS47 and SS40 were from a surface station and two DCM stations in the Sargasso Sea. Viral particles in the water samples were concentrated following the protocols described by Chen et al. (1996). Briefly, 40 to 80 liters of water were filtered through 0.22 or 0.45 μm pore-size low-protein-binding Durapore membranes (Millipore). The filtrate was concentrated by ultrafiltration through a 30,000 MW cutoff Amicon S10Y30 spiral cartridge (Millipore) in a ProFlux M-12 system (Millipore), at 30% of maximum pump speed and 16-18 kpa of back pressure. Virus concentrates (VCs) were stored in the dark at 4°C. In order to further concentrate viral particles, an aliquot of 30 ml VC was centrifuged using an ultracentrifuge (Sorvall Discovery 100S) with a SURESPIN 630 swinging bucket rotor (Sorvall Inc., Newtown, CT) at 40,000g and 4°C for 3 hours. The virus pellet was resuspended in 200 μl distilled water and kept frozen at -20°C.

Primer design and PCR amplification. Several cyanophage-specific oligonucleotide primers were designed based on the g20 gene sequences of three cyanophages (Fig. 1). Either 1 μl of extracted DNA (0.1 μg) or VC was used as a DNA template for PCR amplification. The reaction mixture (total 25μl) contained 1 μl template, 20 pmole of CPS1 or CPS3, 16 pmole of CPS8 or CPS4, 1x PCR buffer (50mM Tris-HCl, 100 mM NaCl, 0.15mM $MgCl_2$), 250 μM of each deoxyribonucleotide triphosphate and 0.75 unit of Expand^{TM} High Fidelity DNA polymerase (Roche,

Indianapolis, IN). PCR amplification was carried out using a model PTC-200 DNA Engine thermocycler (MJ Research). Thermal cycling consisted of an initial denaturation step at 94°C for 3 min, followed by 35 cycles of denaturation at 94°C for 15s, annealing at 36°C for 15s, ramping at 0.3°C /s, and elongation at 73°C for 1 min, with a final elongation step at 73°C for 4 min. A 6 µl aliquot of PCR product was analyzed by electrophoresis in 1.5% agarose gel in 0.5x TBE buffer and stained with ethidium bromide for 15 min. The gel image was captured and analyzed using a gel documentation system (Alpha Innotech Corp., San Leandro, CA). Replicate PCR amplifications were combined to decrease the bias caused by PCR drift (Polz & Cavanaugh 1998, Wintzingerode et al. 1997).

Clone library construction. The PCR amplicon from each VC was purified using the Wizard PCR prep DNA purification system (Promega). The purified products were ligated into the pGEM-T Easy cloning vector (Promega) and then transformed into JM109 competent cells (Promega) as per the manufacturer's instructions. Positive clones (white-colonies) were picked randomly and transferred onto a new agar plate. Clones were numbered from 1 to 80 and assigned the prefix SE (for Skidaway Estuarine), GS (for Gulf Stream) or SS (for Sargasso Sea).

Sequencing. About 35 clones from each clone library were randomly picked and the plasmid inserts were PCR amplified with vector-specific primers T7 and SP6 along with Expand^{TM} High Fidelity DNA polymerase (Roche). PCR amplification involved a 3-min initial denaturation followed by 35 cycles of 94°C for 15s, 45°C for 15s, 73°C for 1 min, and a 4-min final extension at 73°C. The PCR products of the expected size were purified as described above and sequenced bidirectionally with primers T7 and SP6 using the BigDye Terminator Cycle Sequencing Kit (Applied Biosystems Inc.) on a model ABI310 or ABI3700 automated DNA sequencer (Applied Biosystems Inc.). For cyanophage isolates, the PCR products were purified and sequenced directly on both strands with CPS1 and CPS8, respectively.

Phylogenetic analysis. Pairs of sequences were aligned and corrected manually in Sequencher^{TM} 3.0 software (Gene Codes Corporation, Ann Arbor, MI) and the consensus sequences were further analyzed. Sequence alignment and similarity analysis were performed in the Genetic Computer Group package (Madison, WI) and checked visually for chimeric artifacts. Phylogenetic reconstruction was accomplished using the phylogeny inference package (PHYLIP version 3.5 [Felsenstein 1993]). Evolutionary distances were calculated by the Jukes-Cantor method in the DNADIST program. Phylogenetic trees were constructed from distance matrices by the neighbor-joining method, implemented through the NEIGHBOR program. The node reproducibility for tree topology was estimated by bootstrap analysis, which included 100 replicate data sets.

Nucleotide sequence accession numbers. The GenBank accession numbers for the nucleotide sequences determined in this study are from AY027938 to AY028078.

RESULTS

Primer specificity and PCR amplification. Two sets of primers were designed to detect cyanophage isolates and natural cyanophage concentrates. The primers CPS3/CPS4, designed within a conserved region of g20 gene, could amplify a ca. 860 bp fragment from some of the cyanophage isolates but not from the natural VCs. The primers CPS1/CPS8 efficiently amplified g20 gene fragments from both cyanophage isolates and VCs. Therefore, these were used in the investigation of cyanophage diversity and phylogenetic relationships. Among a wide range of cyanophages and other bacteriophages tested for primer specificity, the CPS1/CPS8 primers could only amplify cyanophages belonging to the family Myoviridae. No cyanophages of the family Podoviridae, or Siphoviridae, or other bacteriophages could be amplified (Table 1). For all the 21 strains of cyanomyovirus (cyanophage of Myoviridae) tested, an expected product of ca. 592 bp was amplified successfully (Fig. 2). Amplification with VCs yielded results similar to those of cyanophage isolates.

Phylogenetic diversity of cyanophage isolates. The sequence alignment of the g20 gene fragments from eleven cyanophage isolates demonstrated that the region amplified by primers CPS1/CPS8 was highly conserved (Fig. 3). The sequences between primers CPS1 and CPS8 provided adequate genetic information for phylogenetic inference which was supported by high bootstrap values, whereas the phyletic reconstruction based on the region between CPS1 and CPS2 was poorly supported (Fig. 4a and 4b). The nucleotide similarity analysis revealed that cyanophage isolates shared 64.7 to 96.1% sequence similarity between each other, while enteric coliphage T4 was only 50.4 to 55.2% similar to any of the cyanophage isolates (Table 3). Based on the phylogenetic affiliation among cyanophage isolates and clones from six natural virus communities (shown later, refer to Fig. 6), these isolates were assigned to 3 clusters, with most isolates falling in clusters I and III, and only one isolate (31B) in cluster II. Since S-BnM1 and P17 were distant to any isolates or clones, they were not grouped into any clusters (Fig. 4a).

Phylogenetic diversity of cyanophages in natural virus communities. Cyanophage diversity in estuarine, open ocean surface and the DCM were investigated by examining cyanophage g20 gene diversity in concentrated virus communities (Fig. 5a-f). For each community, about 35 clones were sequenced, among which identical clones were encountered, suggesting that the diversity of cyanophage populations had been adequately sampled (Palenik 1994). A comprehensive phylogenetic tree consisting of isolated cyanophages and representative clones from natural virus communities was constructed to show their genetic relatedness (Fig. 6).

(1) Skidaway Estuarine surface sample (SE1). A total of 29 different g20 sequences were assigned to 5 clusters, with the majority of sequences belonging to clusters A and I (Fig 5a). Nine different nucleotide sequences in cluster A shared similarity greater than 92.9%, while the similarity among the 15 different nucleotides in cluster I varied from 79.3 to 98.2%. According to the phylogenetic affiliation revealed in

Fig 6, three minor clusters (E, F and III) were assigned to this community, with only one or two clones in each cluster. Clone SE38 could not be assigned to any cluster due to the lack of any closely related neighbors.

(2) Gulf Stream surface sample (GS26). This community exhibited much less diversity in the cyanophage g20 gene pool than the Skidaway Estuarine surface community did. Of the 37 clones sequenced, 13 different g20 sequences were found and assigned to three clusters. GS2608 and GS2624 were in clusters III and B, respectively, while all other sequences fell into cluster E with pair-wise nucleotide similarity between 95.8 and 99.8% (Fig. 5b).

(3) Gulf stream DCM sample (GS27). The cyanophage community at the DCM of the Gulf Stream appeared to be more diverse than that at the surface (GS26). Among 33 clones analyzed, 23 different g20 gene sequences were encountered. The DCM layer contained the same three clusters as did the surface sample (cluster III, B and E), plus three other clusters (C, D and II) that were not present on the surface (Fig. 5c). However, unlike the surface sample (GS26) where cluster E was dominant, clusters B and E were the two major clusters at the DCM of the Gulf Stream.

(4) Sargasso Sea surface sample (SS48). The g20 gene diversity and phylogenetic structure of the Sargasso Sea surface sample (SS48) were similar to that of the surface sample in the Gulf Stream (GS26). Fifteen different g20 sequences were found in the 35 analyzed clones. Of the 3 phyletic clusters in this community, cluster B and E were both present, with E as the dominant cluster (Fig. 5d). The higher number of repeat sequences and fewer phyletic groups in GS26 and SS48 indicated a less diversified g20 gene pool in these oceanic surface waters.

(5) Sargasso Sea DCM sample (SS47 and SS40). Two samples in the DCM layer of the Sargasso Sea were analyzed and exhibited a similar phylogenetic pattern. Of 36 and 31 clones studied in samples SS47 and SS40, 24 and 26 different g20 sequences were identified, respectively. Four common phyletic clusters (II, C, E and B) were shared

and the tree topology within clusters II, E and B were very similar between the two communities (Fig.5e and 5f). Compared to the Sargasso surface sample SS48, the two DCM samples (SS47 and SS40) both contained more clusters. The dominant cluster shifted from cluster E in the surface to cluster B in DCM, as was observed in Gulf Stream. However, the two DCM samples in the Sargasso Sea also demonstrated difference from each other. For example, clone SS4716 found in the SS47 sample did not cluster with any other clones or cyanophage isolates, whereas SS40 had two more clusters than SS47.

Comprehensive results. Of 207 g20 clones randomly picked from 6 natural viral communities, 114 different g20 homologs were encountered. No clone showed an identical nucleotide sequence with any of the cyanophage isolates in this study, indicating that cyanophages are very diverse in natural environments. The distribution of clones and g20 sequences within each community was summarized in Table 4. Nine distinct clusters were revealed in the comprehensive phylogenetic tree in Fig 6 based on the amino acid sequence alignment of g20. Clusters I, II and III comprised all the 11 known isolates and 32 different natural clones (nucleotide sequences). Cluster I included 5 isolates and 15 different sequences that were exclusively from the Skidaway Estuarine sample. The pair-wise sequence similarity in cluster I ranged from 75.2 to 98.0%. Cluster II included an isolate from the Sargasso Sea and 8 different sequences from 3 oceanic DCM communities in the Gulf Stream and the Sargasso Sea. The sequence identity in this cluster ranged from 73.2 to 99.5%. Cluster III consisted of 6 isolates and 9 different clones from the Skidaway Estuary, Gulf Stream and Sargasso Sea, with 67.0 to 98.8% sequence similarity in this cluster.

Clusters A through F, however, contained only clones whose sequences did not match any of the known cyanophage isolates. Sequence similarity within these clusters was greater than 92.9% except in cluster E, in which the lowest similarity was only 79.3%. Cluster A and F contained clones exclusively from the Skidaway Estuary, while

clusters B, C and D were formed exclusively by the clones from the open ocean. Additionally, cluster D consisted only of the clones from the oceanic DCM communities as in cluster II, but there was no unique cluster for surface communities in the ocean samples. Clusters B and E were the two largest clusters in the oceanic samples while E is the only cluster that included clones from all 6 communities (Table 4).

DISCUSSION

Viruses are very diverse in the ocean. Since their genomes do not contain universal genetic markers such as rRNAs, it is necessary to develop virus-specific primers in order to interrogate their genetic diversity in nature. Chen et al. (1996) used algal virus-specific primers in PCR-cloning analysis of natural marine virus concentrates and identified a diverse community of genetically related viruses of the *Phycodnaviridae*. In this study, we employed the cyanophage capsid assembly protein gene g20 to investigate the genetic diversity of cyanophage isolates and natural cyanophage communities. Although the g20 gene has been used in DGGE analysis of natural cyanophage communities (Wilson et al. 1999, 2000), this is the first study to reveal phylogenetic affiliation among cyanophage isolates and natural cyanophage communities based on suitable regions in the g20 gene. The primers we designed, CPS1/CPS8, amplify a ca. 592bp conserved region that is adequate for phylogenetic inference. The target of these primers, cyanomyovirus, is the dominant form of cyanophages in marine environments (Waterbury & Valois 1993, Lu et al. 2001). Therefore, the application of primers CPS1/CPS8 allows us to investigate phylogenetic diversity of the vast majority of cyanophages in natural marine habitats.

Phylogenetic analysis with isolated cyanophages indicated that marine cyanophages were genetically divergent yet more closely related to each other than to bacteriophage T4 (Fig. 4). Similar results were reported for marine microalgal viruses, which formed a unique group distant from other double stranded DNA viruses infecting

eukaryotes (Chen and Suttle, 1996). Similarity analysis of g20 sequences further suggested that no correlation was present between genetic variation and geographic distance (Table 1 and 3). Isolates from the same or adjacent water bodies could share high g20 identity, e.g. 90.4% between 27A and 32A, and 96.1% between P79 and P81. However, isolates from different water bodies could also share a high similarity, e.g. 96.1% between S-WHM1 and P77. On the other hand, cyanophage isolates from the same region can be as diverse as those from different oceans. For example, there is 66.2% similarity between 31B and 44B, and 68.3% similarity between 31B and P17. Our results with cyanophage isolates support the suggestion that genetically related marine viruses are widely distributed in the ocean without significant geographic segregation (Kellogg et al. 1995).

Natural cyanophage communities from both estuarine and oceanic environments exhibited even higher genetic diversity than did the cyanophage isolates we studied. Among the 9 phylogenetic clusters identified from the g20 sequences, 3 clusters contained known cyanophage isolates, while the other 6 clusters contained only natural clones. Even within a single cluster (e.g. cluster I, II or III), clones demonstrated their divergent nature by the wide range in sequence similarity (73.2 to 99.5%). Furthermore, except for the two surface oceanic samples, all other communities contained 5 to 6 distinct genetic groups, with up to 29 different genotypes found in a single sample. The cyanophage diversity reported here is much higher than that reported from previous studies (Wilson et al. 2000, Lu et al. 2001). Wilson et al. (2000) used g20 gene-based cyanomyovirus-specific primers and DGGE to investigate cyanophage diversity in the Atlantic Ocean. Only between 2 and 10 different genotypes were found by DGGE analysis in each concentrated virus community. The difference between our results and those of Wilson et al. (2000) may result from the much higher sensitivity of our cloning-sequencing analysis and the degeneracy in our primers that was not included in the DGGE-oriented PCR primers. Many of the cyanophages used in this study were isolated

from Georgia estuaries and exhibited diverse TEM morphology and RFLP patterns (Chen et al. 1998, Lu et al. 2001). Our results not only revealed a much higher diversity of cyanophages in marine environments, but also demonstrated their phylogenetic affiliation via nucleic acid sequence analysis.

High phage diversity in natural marine environments could be due to either phage-host genetic exchange through transduction, or genetic exchange between co-infecting phages. The g20 gene is within a mobile genetic module that is exchangeable between viruses during co-infection of bacterial hosts (Monod et al. 1997). The observed diverse cyanophage assemblages and the dynamic phage-host interactions suggested that marine phages might be important in shaping the genetic diversity and composition of marine microbial communities.

It is likely that clusters I, II and III represent lytic phages that infect WH7803-like *Synechococcus* spp. The wide geographical origins of isolates in cluster I and III further suggest that cyanophages in these clusters may infect *Synechococcus* spp. that are adapted to widely differing marine environments, whereas cluster II includes cyanophages whose hosts are adapted solely to oligotrophic environments. The identity of clusters A through F was enigmatic due to the absence of any isolates in them. Possibly some clusters may represent cyanophages infecting *Synechococcus* cells that are genetically distant from the WH7803-like strains, on which most of our cyanophages were isolated. It is known that considerable genetic diversity exists within natural assemblages of marine *Synechococcus* (Palenik 1994, Toledo & Palenik 1997). Isolation of more cyanophages that infect different marine *Synechococcus* spp. will shed light on the origins of the unknown clusters. We also cannot rule out the possibility that some g20 clones in clusters A-F are other bacteriophages whose hosts are closely related to *Synechococcus*. For example, marine *Prochlorococcus* spp. are abundant (Chisholm et al. 1988, Partensky et al. 1999) and genetically similar to *Synechococcus* in 16S rRNA (Urbach et al. 1992) and RNA polymerase gene phylogeny (Palenik & Haselkorn 1992).

It is possible that our primers may amplify *Prochlorococcus* phages. This hypothesis remains to be tested when the phages infecting *Prochlorococcus* become available.

The genetic composition and structure of natural cyanophage communities in estuaries and the open ocean were different from each other. While 56% of estuarine g20 clones were within clusters I, II and III, the vast majority (87% average) of oceanic clones fell in clusters B-F. Moreover, unique phylogenetic clusters were found for each environment. For example, clusters A, F and I were amplified from the estuarine sample, whereas clusters B, C, D and II were only present in the oceanic samples (Table 4). We also observed consistently higher cyanophage diversity in the oceanic DCM samples than those in the surface samples, even though different waters were studied (i.e. Gulf Stream vs. Sargasso Sea). Finally, it is evident that the dominant g20 clusters shifted from cluster E at the oceanic surface to cluster B at the DCM (Table 4). The distinct cyanophage population structures in estuarine vs. offshore, and in surface vs. DCM suggested the presence of different host populations in response to different light, nutrient and other physical conditions, and a dynamic interaction between cyanophage and cyanobacterial populations. Currently, little is known about the influence of environmental conditions on the genetic structure of natural *Synechococcus* assemblages. Presumably, the increasing cyanophage diversity from the surface to the DCM layer is related to the more diverse host populations at depth in stratified water columns.

The viral capsid assembly protein gene g20 was successfully used in this study as a marker molecule to investigate genetic diversity and phylogenetic affiliation among cyanophage isolates and natural cyanophage assemblages in estuarine and oligotrophic environments. Phylogenetic analysis of cyanophage g20 gene sequences revealed that natural cyanophage populations are strikingly diverse, and that their genetic structures varied greatly in different marine environments. This study suggests that marine viruses (at least marine cyanophages), are under significant selection pressure and are able to maintain their diversity in response to host resistance. Further studies on the co-variation

of the genetic diversity of virus and host populations with changing environmental variables would provide new insights into the ecological roles of marine viruses.

ACKNOWLEDGEMENT

This work is supported by grants from The National Science Foundation (OCE-973060, OCE-0049098, OCE-9977040) and the Department of Energy (DE-FG02-97ER62451). We thank Jingrang Lu at the University of Georgia for providing the cyanophage isolates from Georgia estuaries, and Leo Poorvin at the University of Tennessee for providing cyanophage lysates from Sargasso Sea. Curtis Suttle at the university of British Columbia, Canada, John Waterbury at Woods Hole Oceanographic Institution, and Eric Wommack at the University of Georgia (currently at University of Delaware) kindly provided cyanophage and bacteriophage isolates for this study.

REFERENCES

1. **Bergh, O, K. Y. Borsheim, G. Bratbak, and M. Heldal.** 1989. High abundance of viruses found in aquatic environments. Nature (London) **340:**467-468.
2. **Chen, F., and C.A. Suttle.** 1996. Evolutionary relationships among large double-stranded DNA viruses that infect microalgae and other organisms as inferred from DNA polymerase genes. Virology **219:**170-178.
3. **Chen, F., C.A. Suttle and S.M. Short.** 1996. Genetic diversity in marine algal virus communities as revealed by sequence analysis of DNA polymerase genes. Appl. Environ. Microbiol. **62:**2869-2874.
4. **Chisholm, S. W., R. J. Olson, E. R. Zettler, R. Goericke, J. B. Waterbury, and N. A. Welschmeyer.** 1988. A novel free-living prochlorophyte abundant in the oceanic euphotic zone. Nature (London) **334:**340-343.
5. **Chisholm, S. W., S. L. Frankel, R. Goericke, R. J. Olson, B. Palenik, J. B. Waterbury, et al.** 1992. *Prochlorococcus marinus* nov. gen. Nov. sp.: an

oxyphototrophic marine prokaryote containing divinyl chlorophyll a and b. Atch. Microbiol. **157:**297-300.

6. **Felsenstein, J.** 1993. PHYLIP: phylogeny Inference Package (version 3.5). University of Washington, Seattle.
7. **Fuhrman, J. A.** 1999. Marine viruses and their biogeochemical and ecological effects. Nature (London) **399:**541-547.
8. **Fuhrman, J. A., and C. A. Suttle.** 1993. Viruses in marine planktonic systems. Oceanogr. **6:**50-62.
9. **Fuhrman, J. A., and R. T. Noble.** 1995. Viruses and protists cause similar bacterial mortality in coastal seawater. Limnol. Oceanogr. **40:**1236-1242.
10. **Fuller, N. J., W. H. Wilson, I. R. Joint, and N. H. Mann.** 1998. Occurrence of a sequence in marine cyanophages similar to that of T4 gp20 and its application to PCR-based detection and quantification techniques. Appl. Environ. Microbiol. **64:**2051-2060.
11. **Garza, D. R., and C. A. Suttle.** 1998. The effect of cyanophages on the mortality of *Synechococcus* spp and selection for UV resistant viral communities. Microb. Ecol. **36:**281-292.
12. **Jiang, S. C., C. K. Kellogg, and J. H. Paul.** 1998. Characterization of marine temperate phage-host systems isolated from Mamala Bay, Oahu, Hawaii. Appl. Environ. Microbiol. **64:**535-542.
13. **Kellogg, C. A., J. B. Rose, S. C. Jiang, J. M. Thurmond, and J. H. Paul.** 1995. Genetic diversity of related vibriophages isolated from marine environments around Florida and Hawaii, USA. Mar. Ecol. Prog. Ser. **120:**89-98.
14. **Lenski, R. E.** 1988. Dynamics of interactions between bacteria and virulent bacteriophage. Adv. Microb. Ecol. **10:**1-44.

15. **Lu, J., F. Chen, and R. E. Hodson.** 2001. Distribution, isolation, host specificity, and diversity of cyanophages infecting marine *Synechococcus* spp. in the Georgia river estuaries. Appl. Environ. Microbiol. In press.

16. **Monod, C., F. Repoila, M. Kutateladze, F. Tetart, and H. M. Krisch.** 1997. The genome of the pseudo T-even bacteriophages, a diverse group that resembles T4. J. Mol. Biol. **267:**237-249.

17. **Palenik, B.** 1994. Cyanobacterial community structure as seen from RNA polymerase gene sequence analysis. Appl. Environ. Microbiol. **90:**3212-3219.

18. **Palenik, B., and H. Swift.** 1996. Cyanobacterial evolution and prochlorophyte diversity as seen in DNA-dependent RNA polymerase gene sequences. J. Phycol. **32:**638-646.

19. **Palenik, B., and R. Haselkorn.** 1992. Multiple evolutionary origins of prochlorophytes, the chlorophyll *b*-containing prokaryotes. Nature (London) **355:**265-267.

20. **Partensky, F., W. R. Hess, and D. Vaulot**. 1999. *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. Microbiol. Mol. Biol. Rev. **63**(1):106-127.

21. **Polz, M. F., and C. M. Cavanaugh.** 1998. Bias in template-to-product ratios in multitemplate PCR. Appl. Environ. Microbiol. **64:**3724-3730.

22. **Proctor, L. M., and J. A. Fuhrman.** 1990. Viral mortality of marine bacteria and cyanobacteria. Nature (London) **343:**60-62.

23. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N. Y.

24. **Suttle, C. A.** 2000. Cyanophages and their role in the ecology of cyanobacteria, p. 563-589. *In* B. A. Whitton, and M. Potts (ed.), The ecology of cyanobacteria: their diversity in time and space. Kluwer Academic Publishers, Boston.

25. **Suttle, C.A., and A.M. Chan.** 1993. Marine cyanophages infecting oceanic and coastal strains of *Synechococcus*: abundance, morphology, cross-infectivity and growth characteristics. Mar. Ecol. Prog. Ser. **92:**99-109.

26. **Suttle, C.A., and A.M. Chan.** 1994. Dynamics and distribution of cyanophages and their effect on marine *Synechococcus* spp. Appl. Environ. Microbiol. **60:**3167-3174.

27. **Suttle, A. S., A. M. Chan, and M. T. Cottrell.** 1990. Infection of phytoplankton by viruses and reduction of primary productivity. Nature (London) **347:**467-469.

28. **Thingstad, T. F., M. Heldal, G. Bratbak, and I. Dundas.** 1993. Are viruses important partners in pelagic food webs? Trends Ecol. Evol. **8:**209-212.

29. **Toledo, G., and B. Palenik.** 1997. *Synechococcus* diversity in the California Current as seen by RNA polymerase (rpoC1) gene sequences of isolated strains. Appl. Environ. Microbiol. **63:**4298-4303.

30. **Urbach, E., D. L. Robertson, and S. W. Chisholm.** 1992. Multiple evolutionary origins of prochlorophytes within the cyanobacterial radiation. Nature (London) **355:**267-270.

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

32. **Waterbury, J. B., and R. Rippka.** 1989. The order *Chroococcales*, p. 1728-1746. *In* J. T. Staley, M. P. Bryant, N. Pfennig, and J. G. Holt (ed.), Bergey's manual of systematic bacteriology, vol. 3. Williams and Wilkins, Baltimore, Md.

33. **Waterbury, J. B., S. W. Watson, F. W. Valois, and D. G. Franks.** 1986. Biological and ecological characterization of the marine unicellular cyanobacterium *Synechococcus*. Can. Bull. Fish. Aquat. Sci. **214:**71-120.

34. **Wilhelm, S. W., and C. A. Suttle.** 1999. Viruses and nutrient cycles in the sea: viruses play critical roles in the structure and function of aquatic food webs. Biosci. **49:**781-788.

35. **Wilson, W. H., J. F. Nicholas, I. R. Joint, and N. H. Mann.** 1999. Analysis of cyanophage diversity and population structure in a south-north transect of the Atlantic ocean. Bulletin de l'Institut Oceanographique (Monaco). **19:**209-216.

36. **Wilson, W. H., J. F. Nicholas, I. R. Joint, and N. H. Mann.** 2000. Analysis of cyanophage diversity in the marine environment using denaturing gradient gel electrophoresis, p. 565-570. *In* C. R. Bell, M. Brylinsky, and P. Johnson-Green (ed.), Microbial Biosystems: New Frontier. Proceedings of the 8th International Symposium on Microbial Ecology. Halifax, Canada.

37. **Wilson, W.H., I.R. Joint, N.G Carr, and N.H. Mann.** 1993. Isolation and molecular characterization of five marine cyanophages propagated on *Synechococcus* sp. strain WH7803. Appl. Environ. Microbiol. **59:**3736-3743.

38. **Wintzingerode, F. V., U. B. Gobel, and E. Stackebrandt.** 1997. Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. FEMS Microbiol. Rev. **21:**213-229.

39. **Zhong, Y., F. Chen, and R. E. Hodson.** 2000. Phyletic diversity of marine cyanophages as revealed by the sequence of viral capsid assembly gene gp20. Abstracts of the General Meeting of the American Society for Microbiology. **100:**465.

Table 1. Viruses tested with primers CPS1 and CPS8, and reference cyanophages for primer design.

Virus	Family	Host(s)	Area of isolation	Reference
Cyanophages amplified by CPS1/CPS8 primers:				
P3	M[d]	WH7805	Sapelo Island, GA	15
P5	M	WH7803	Sapelo Island, GA	15
P6	M	WH7805	Dauphin Island, AL	15
P8	M	WH8101	Sapelo Island, GA	15
P12	M	WH8101	Sayll Estuary, AL	15
P16	M	WH7803	Savannah River Estuary, GA	15
P17	M	WH7803	Qingdao Coast, China	15
P39	M	WH7805	Savannah River Estuary, GA	15
P61	M	WH7803	Satilla River Estuary, GA	15
P66	M	WH7803	Satilla River Estuary, GA	15
P73	M	WH7805	Satilla River Estuary, GA	15
P76	M	WH8007	Altamaha River Estuary, GA	15
P77	M	WH8007	Altamaha River Estuary, GA	15
P79	M	WH7805	Satilla River Estuary, GA	15
P81	M	WH7805	Altamaha River Estuary, GA	15
ϕ9	M	WH7803	Woods Hole, MA	31
S-PWM1	M	WH7803	Gulf of Mexico	25
27A[a, b, c]		WH7803	Sargasso Sea	
31B[a, b, c]		WH7803	Sargasso Sea	
32A[a, b, c]		WH7803	Sargasso Sea	
44B[a, b, c]		WH7803	Sargasso Sea	
Cyanophages not amplified by CPS1/CPS8 primers:				
P1	S[e]	WH7803	Sayll Estuary, AL	15
SBP1	P[f]	WH7803	Bermuda	25
ϕ12	P	WH7805	Gulf Stream	31
Non-cyanophage viruses not amplified by CPS1/8 primers				
T4	M	*Escherichia coli*	Sigma	
T5	S	*Escherichia coli*	Sigma	
T7	P	*Escherichia coli*	Sigma	
CB38[b, c]		*Aeromonas* sp	Chesapeak Bay, MD	
CB908[b, c]		*Shewanella* sp	Chesapeak Bay, MD	
CB 45[b, c]		*Pseudomonas sp*	Chesapeak Bay, MD	
CB7[b, c]		*Aeromonas* sp	Chesapeak Bay, MD	
Reference cyanophages for primer design:				
S-BnM1	M	WH7803	Bergen, Norway	10
S-PM2	M	WH7803	ymouth, U.K.	10
S-WHM1	M	WH7803	Woods Hole, Mass.	10

[a] lysate only

[b] unpublished results

[c] morphology not known

Table 1. continued.

[d] Myoviridae

[e] Siphoviridae

[f] Podoviridae

Table 2. Stations where VCs were collected.

Name	Date	Concentr. Factor	Depth (m)	Latitude	Longitude	Surface temperature (°C)	Salinity (‰)
SE1	8/29/99	14, 630	surface	31°59′N	81°01′W	26.5	33.0
GS26	5/27/00	27, 077	surface	37°19′N	71°37′W	26.3	36.2
GS27	5/28/00	9, 908	83	36°24′N	71°20′W	21.9	36.4
SS40	6/5/00	11, 046	134	28°53′N	65°4′W	25.4	36.4
SS47	6/12/00	13, 831	100-148	34°43′N	68°7′W	23.9	36.7
SS48	6/13/00	21, 417	surface	36°47′N	71°3′W	23.8	36.7

Table 3. Percent (%) nucleotide sequence similarity between cyanophage isolates and phage T4.

	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. S-BnM1	70.9	69.8	69.1	69.1	68.3	67.8	70.4	69.6	69.1	68.0	69.5	68.5	69.3	55.1
2. 31B		71.4	69.7	68.2	68.4	68.2	65.0	64.7	68.4	66.6	66.2	66.2	68.3	51.6
3. 27A			90.4	79.6	80.1	76.8	69.4	70.0	70.2	68.6	70.2	68.9	69.9	53.7
4. 32A				81.2	80.5	75.0	68.5	68.3	68.2	66.0	68.3	68.0	69.7	52.6
5. P12					90.0	83.5	70.6	70.9	70.0	68.2	68.5	68.7	72.8	53.1
6. P6						84.0	69.8	69.8	69.8	68.0	68.5	67.8	71.5	53.9
7. S-PM2							67.2	67.6	68.5	66.7	67.0	67.4	70.8	53.5
8. S-WHM1								96.1	79.9	78.1	79.7	77.6	71.0	55.2
9. P77									81.5	79.8	81.9	78.9	71.5	54.6
10.P79										96.1	93.9	89.6	73.6	53.0
11.P81											90.2	87.9	70.5	52.2
12.44B												87.6	72.8	53.9
13.M1													71.9	51.9
14.P17														50.4
15.T4														

Table 4. Distribution of g20 gene clones and nucleotide sequences in the six virus communities.

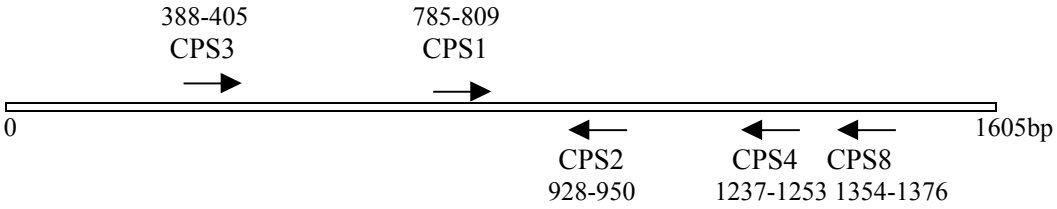
Community	Clone[a] /Nu[b] in each cluster										Total
	I	A	F	II	D	B	C	E	III	Ungrouped[c]	
SE1	18/15	12/9	2/2					1	1	1	35/29
GS26						1		34/11	2/1		37/13
SS48						6/4	1	28/10			35/15
GS27				1	4/2	12/10	7/2	8/7	1		33/23
SS47				4/3		15/13	9/3	7/4		1	36/24
SS40				5/4	2/2	10/8	1	5/5	8/6		31/26
Total	18/15	12/9	2/2	10/8	6/4	44/31	18/5	83/29	12/9	2/2	207/114[d]

[a] number of clones in the cluster

[b] number of different nucleotide sequences found in the cluster

[c] single clone that didn't cluster with others

[d] identical sequences found in more than one community were only counted once.



CPS1: 5'-GTAG[T/A]ATTTTCTACATTGA[C/T]GTTGG-3' (Fuller et al. 1998)

CPS2: 5'-GGTA[G/A]CCAGAAATC[C/T]TC[C/A]AGCAT-3' (Fuller et al. 1998)

CPS3: 5'-TGGTA[T/C]GT[T/C]GATGG[A/C]AGA-3'

CPS4: 5'-CAT[A/T]TC[A/T]TCCCA[A/T/C]TCTTC-3'

CPS8: 5'-AAATA[C/T]TT[G/A/T]CCAACA[A/T]ATGGA-3'

Figure 1. Relative positions and sequences of cyanophage-specific primers designed in this study and published previously.

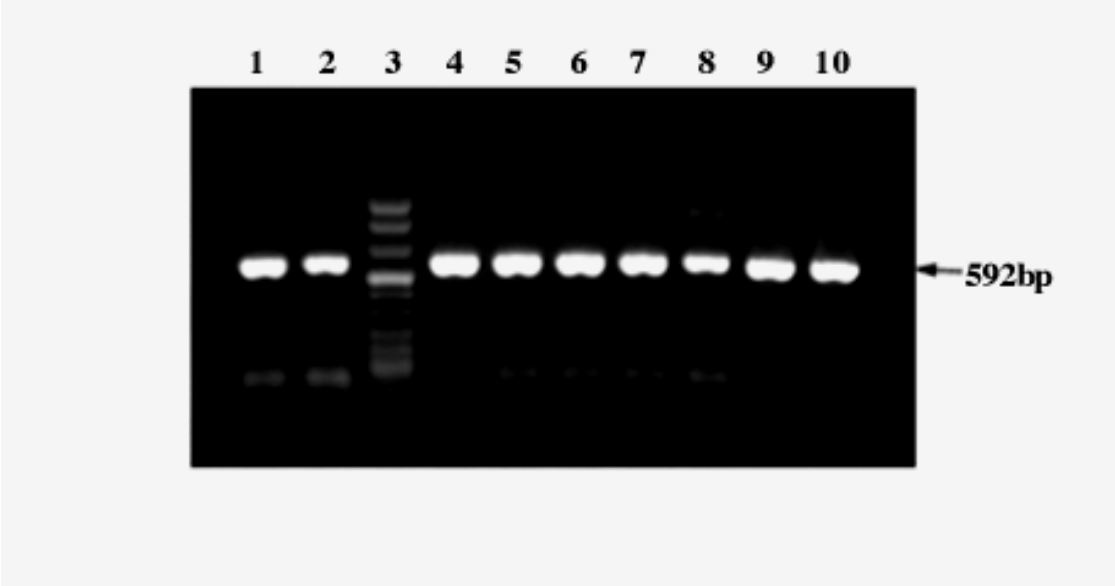


Figure 2. Gel electrophoresis of PCR products amplified with primers CPS1/CPS8. Lanes 1, 2, 4-10 corresponded to PCR products amplified from cyanophages 44B, 27A, P12, P6, 32A, 32B, P61, M1 and P79, respectively. Lane 3 was Molecular Weight Marker VIII which corresponded to 1114, 900, 692, 501, 404, 320, 242, 190, 147 and 124bp, respectively (from top to bottom).

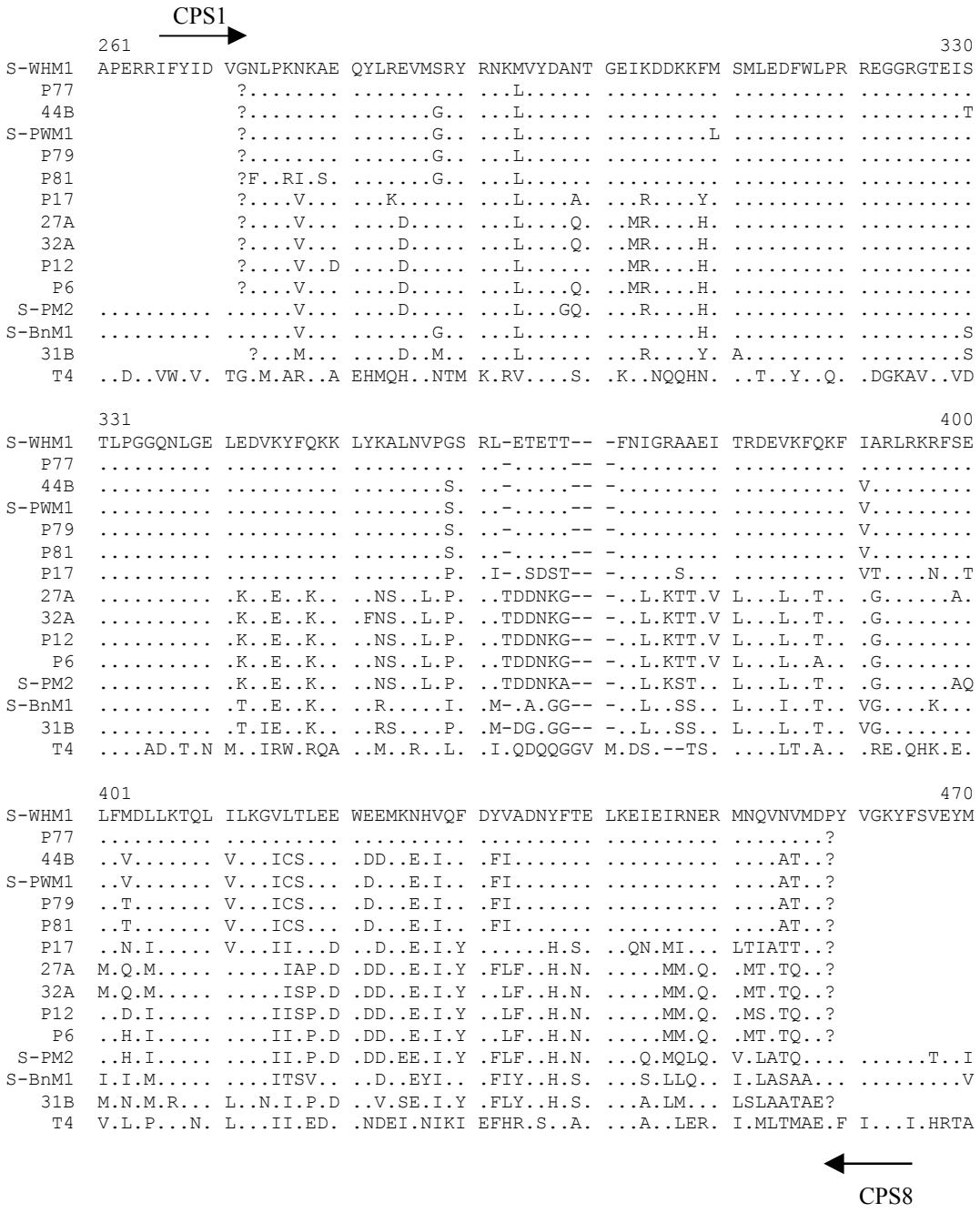
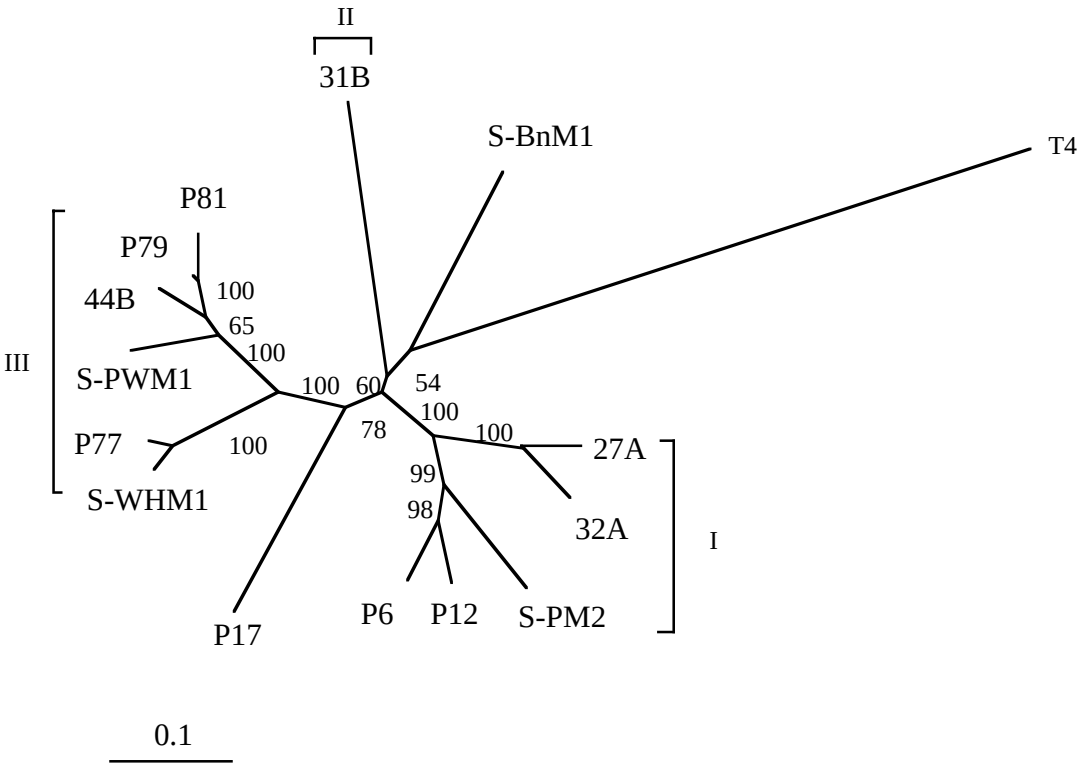
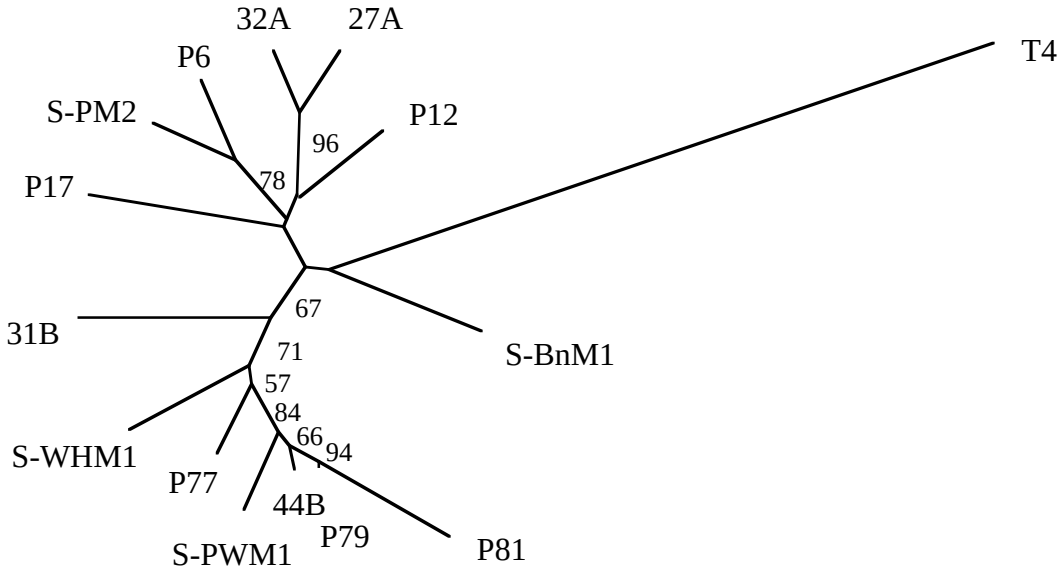


Figure 3. Manually refined GCG pileup of deduced amino acid sequences of g20 gene from 14 cyanophage isolates and coliphage T4. Numbers corresponded to amino acid positions in cyanophage S-WHM1 g20 sequence. Symbols: identical amino acids to that at the same position in the sequence of S-WHM1 were shown as ".", gap sites as "-" and undetermined sites as "?".

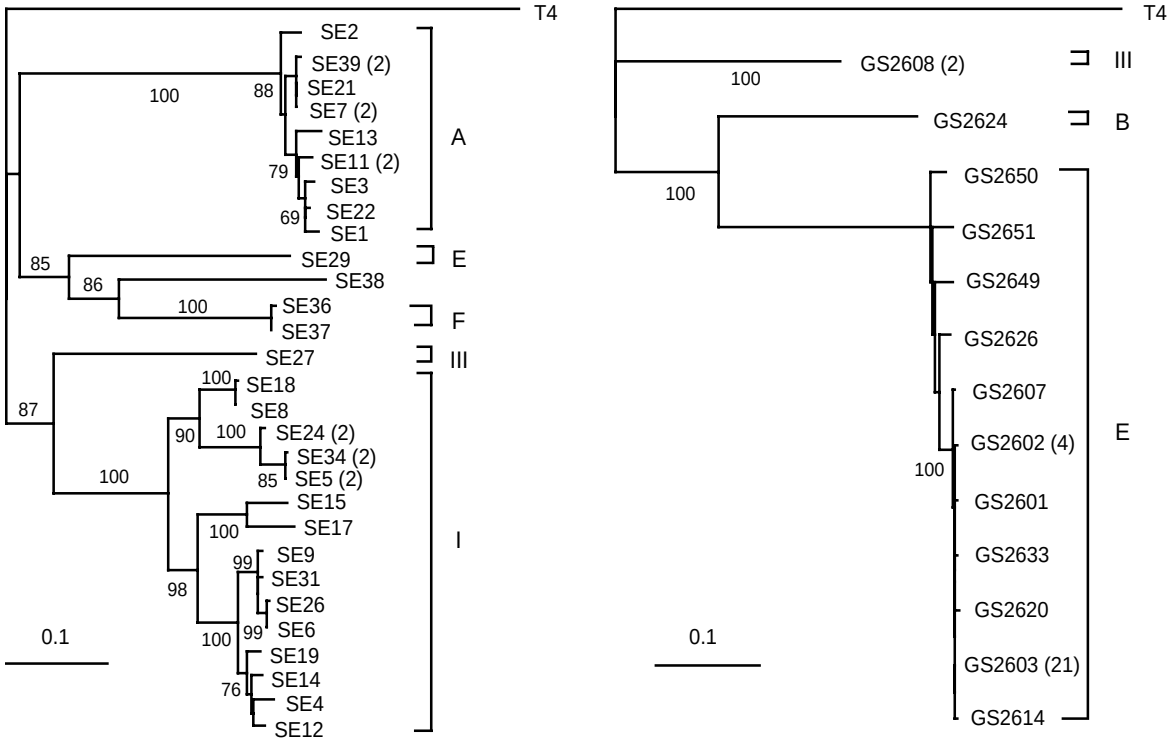


(a)

Figure 4. Phylogenetic affiliation of cyanophage isolates based on (a) 552 bp nucleotide sequences between primers CPS1 and CPS8 and (b) 114 bp sequences between CPS1 and CPS2. Coliphage T4 was used as an outgroup. Numbers at tree branches indicated bootstrap values with 100 replicates. The scale bar is equivalent to 0.1 replacement per site. Only bootstrap values larger than 50 are shown.


32A
27A
P6
T4
S-PM2
P12
96
78
P17
67
31B
71
S-BnM1
57
84
66
94
S-WHM1
P77
44B
P79
P81
S-PWM1
0.1

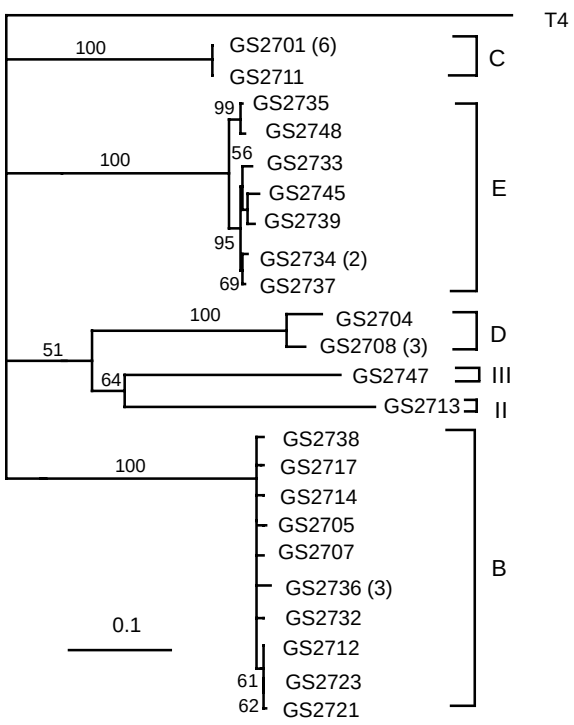
(b)



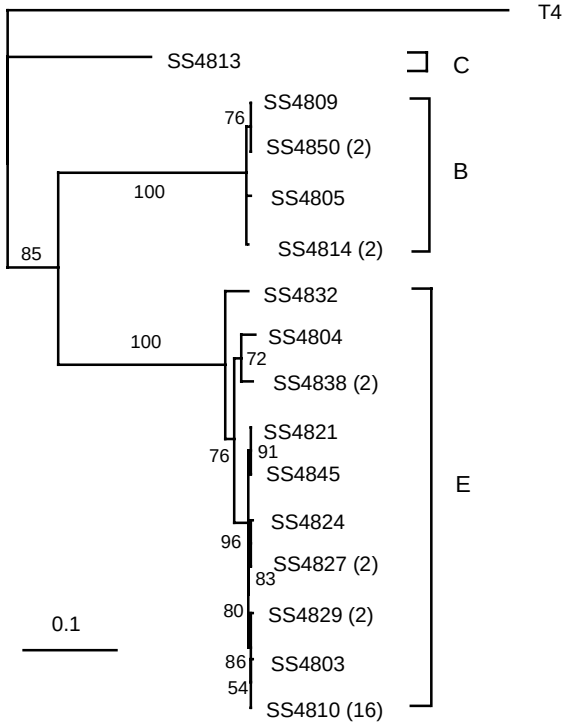
a. Skidaway Esturine surface sample

b. Gulf Stream surface sample

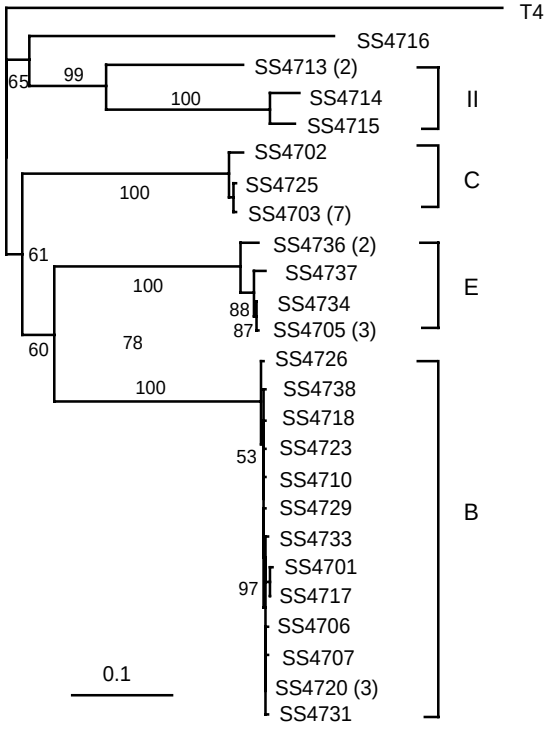
Figure 5. Phylogenetic relationship among the clones amplified with primers CPS1/CPS8 from concentrated natural virus communities at: (a) Skidaway Estuarine surface SE1, (b) Gulf Stream surface GS26, (c) Gulf Stream DCM GS27, (d) Sargasso Sea surface SS48, (e) Sargasso Sea DCM SS47 and (f) Sargasso Sea DCM SS40. The neighbor-joining tree was constructed based on 552 bp nucleotide sequence alignment with T4 as an outgroup. Only bootstrap values greater than 50 are shown. Numbers in parenthesis indicate number of clones bearing identical nucleotide sequences in a community. The clusters were assigned according to Fig.6.



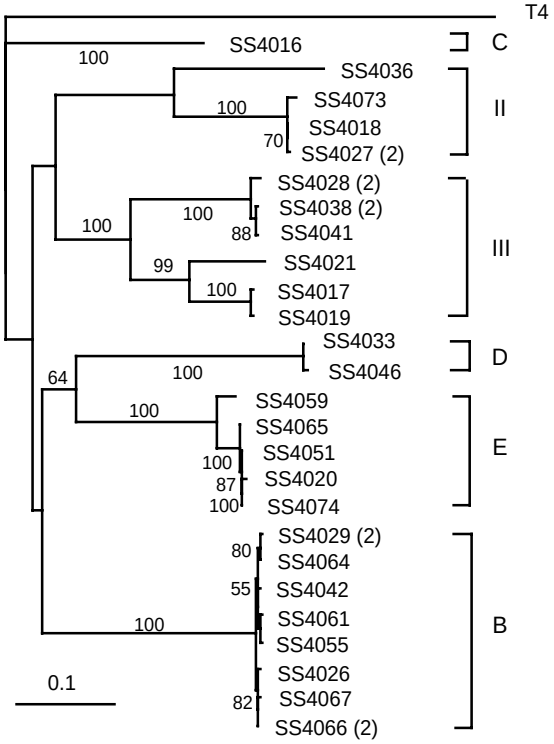
c. Gulf Stream DCM sample



d. Sargasso Sea surface sample



e. Sargasso Sea DCM sample



f. Sargasso Sea DCM sample

Figure 5. Continued.

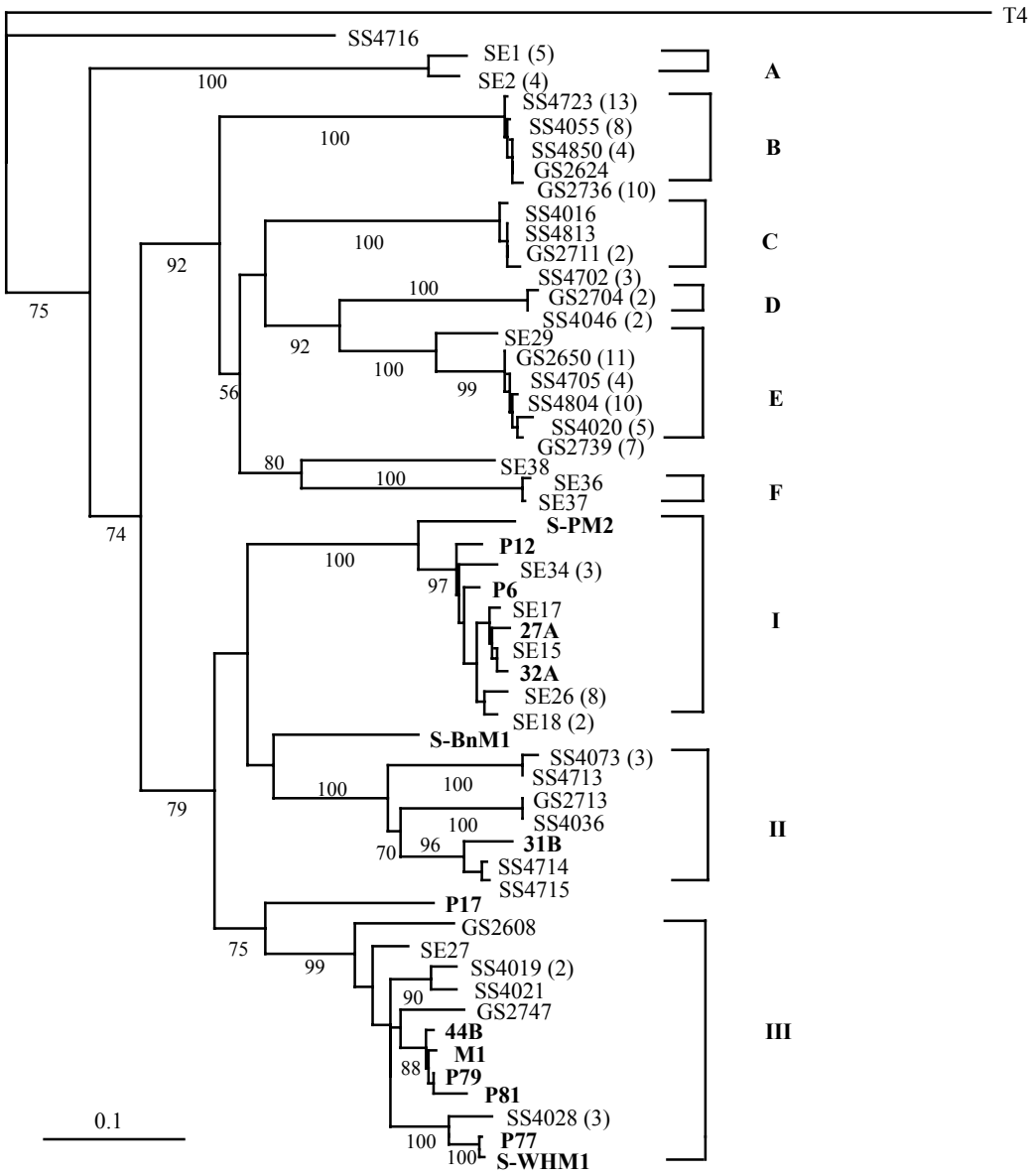


Figure 6. A neighbor-joining tree showing the phylogenetic affiliation of cyanophage isolates (boldtype) and representative clones from all the 6 natural virus communities. The tree was constructed based on 176 amino acid sequence alignment, with T4 as an outgroup. Numbers in parenthesis indicated the number of different nucleotide sequences in the same cluster and same community as the representative clone. Clusters A through F and I through III were assigned according to phylogenetic relatedness. Bootstrap values less than 50 were not shown. The scale bar indicated 0.1 substitution per site.

CHAPTER 3

SUMMARY

The objectives of this study were threefold: (1) to develop primers with high specificity for marine cyanomyoviruses that amplify suitable g20 gene fragments for phylogenetic inferences, (2) to study the phylogenetic diversity of marine cyanophage isolates originated from a wide variety of geographical locations, (3) to assess the phylogenetic relationship and diversity of cyanophage communities in different marine environments. These objectives have been achieved successfully.

Although the database of the g20 gene sequences was very small, we were able to design degenerate oligonucleotide primers CPS1/CPS8 based on a conserved region in the g20 sequences of three cyanomyoviruses. These primers were shown to amplify ca. 592bp fragments efficiently from both cyanophage isolates and natural virus concentrates. The target organism, cyanomyovirus, is the dominant form of cyanophage in marine environments. Therefore, this primer set allows us to investigate genetic diversity of a vast majority of marine cyanophages. We also provided evidence that the sequences flanked by our primers are suitable for phylogenetic inference among marine cyanophages. The primers CPS1/CPS8 are only used in PCR-cloning analysis in this study, yet they are potentially applicable in techniques such as T-RFLP (terminal restriction fragment length polymorphism) following PCR to obtain rapid genetic fingerprints of natural cyanophage communities.

Although marine viruses are now known to be a ubiquitous, abundant, diverse and dynamic component of microbial consortia, little is known about their phylogenetic diversity. This study is the first to employ g20 gene sequences to infer phylogenetic relatedness among cyanophage isolates. Our data demonstrated that geographically separated marine cyanophages were highly diverse yet more closely related to each other

than to other bacteriophages such as enteric coliphage T4. There was no correlation between genetic variation and geographic distance between sampling sites, suggesting that phylogenetically related marine viruses are widely distributed without significant geographical segregation.

This is also the first study to compare phylogenetic structure of natural virus communities in different marine environments. Concentrated natural cyanophage communities from either estuarine or oceanic environments each exhibited strikingly high phylogenetic diversity. The genetic composition and community structure of natural cyanophages in estuarine and open ocean environments are substantially different from each other, with unique phylogenetic clusters found for each environment. It is also shown that natural cyanophage communities in the oceanic DCM are more diverse than those from surface samples, with obvious clonal shifts from surface to the deeper layer. Our data not only revealed a much higher diversity of marine cyanophages than those found by morphology and DGGE studies, but also demonstrated their phylogenetic affiliation via nucleic acid sequence analysis.

A number of new cyanophage genotypes are identified in this study. Of the 114 different g20 sequences we encountered, none of them was identical to the sequence of any of the cyanophage isolates we studied. Nine phylogenetic clusters were found for natural cyanophage communities, with only three clusters containing known cyanophage isolates. The other six unidentified groups may be comprised of bacteriophages that infect different groups of *Synechococcus* or other closely related cyanobacteria. Isolation of more cyanophages that infect different marine *Synechococcus* spp. and other cyanobacteria will shed light on the origins of these unknown clusters.

Viruses are suggested to be able to influence microbial diversity though genetic exchange. The high genetic diversity of marine cyanophage assemblages revealed by this study implies that marine viruses can play an important role in the regulation of microbial genetic diversity. Further studies on the co-variation of the phylogenetic structure of

cyanophages (e.g. by g20 gene phylogeny) and host populations (e.g. by RuBisCO gene analysis) in different environmental conditions will provide new insights into the roles of marine viruses.

APPENDIX

Table A1. GenBank nucleotide sequence accession numbers for the 11 cyanophage isolates and 114 different clones sequenced in this study.

Isolates or clones	Accession number	Isolates or clones	Accession number	Isolates or clones	Accession number
GS2601	AY027938	GS2732	AY027963	SE4	AY027988
GS2602	AY027939	GS2733	AY027964	SE5	AY027989
GS2603	AY027940	GS2734	AY027965	SE6	AY027990
GS2607	AY027941	GS2735	AY027966	SE7	AY027991
GS2608	AY027942	GS2736	AY027967	SE8	AY027992
GS2614	AY027943	GS2737	AY027968	SE11	AY027993
GS2620	AY027944	GS2738	AY027969	SE12	AY027994
GS2624	AY027945	GS2739	AY027970	SE9	AY027995
GS2626	AY027946	GS2745	AY027971	SE13	AY027996
GS2633	AY027947	GS2747	AY027972	SE14	AY027997
GS2649	AY027948	GS2748	AY027973	SE15	AY027998
GS2650	AY027949	27A	AY027974	SE17	AY027999
GS2651	AY027950	31B	AY027975	SE18	AY028000
GS2701	AY027951	32A	AY027976	SE19	AY028001
GS2704	AY027952	44B	AY027977	SE21	AY028002
GS2705	AY027953	S-PWM1	AY027978	SE22	AY028003
GS2707	AY027954	P6	AY027979	SE24	AY028004
GS2708	AY027955	P12	AY027980	SE26	AY028005
GS2711	AY027956	P17	AY027981	SE27	AY028006
GS2712	AY027957	P77	AY027982	SE29	AY028007
GS2713	AY027958	P79	AY027983	SE31	AY028008
GS2714	AY027959	P81	AY027984	SE34	AY028009
GS2717	AY027960	SE1	AY027985	SE36	AY028010
GS2721	AY027961	SE2	AY027986	SE37	AY028011
GS2723	AY027962	SE3	AY027987	SE38	AY028012

Table A1. Continued.

Isolates or clones	Accession number	Isolates or clones	Accession number	Isolates or clones	Accession number
SE39	AY028013	SS4065	AY028035	SS4729	AY028057
SS4016	AY028014	SS4066	AY028036	SS4731	AY028058
SS4017	AY028015	SS4067	AY028037	SS4733	AY028059
SS4018	AY028016	SS4073	AY028038	SS4734	AY028060
SS4019	AY028017	SS4074	AY028039	SS4736	AY028061
SS4020	AY028018	SS4701	AY028040	SS4737	AY028062
SS4021	AY028019	SS4702	AY028041	SS4738	AY028063
SS4026	AY028020	SS4703	AY028042	SS4803	AY028064
SS4027	AY028021	SS4705	AY028043	SS4804	AY028065
SS4028	AY028022	SS4706	AY028044	SS4805	AY028066
SS4029	AY028023	SS4707	AY028045	SS4809	AY028067
SS4033	AY028024	SS4710	AY028046	SS4810	AY028068
SS4036	AY028025	SS4713	AY028047	SS4813	AY028069
SS4038	AY028026	SS4714	AY028048	SS4814	AY028070
SS4041	AY028027	SS4715	AY028049	SS4821	AY028071
SS4042	AY028028	SS4716	AY028050	SS4824	AY028072
SS4046	AY028029	SS4717	AY028051	SS4827	AY028073
SS4051	AY028030	SS4718	AY028052	SS4829	AY028074
SS4055	AY028031	SS4720	AY028053	SS4832	AY028075
SS4059	AY028032	SS4723	AY028054	SS4838	AY028076
SS4061	AY028033	SS4725	AY028055	SS4845	AY028077
SS4064	AY028034	SS4726	AY028056	SS4850	AY028078

```
        251                                                        300
   SE1  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~SIFYID VGNLPKGKAE
   SE2  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~VVIFYID VGNLPKGKAE
S-WHM1  MTLSHLHKAI KAVNQLRMIE DSLVIYRLSR APERRIFYID VGNLPKNKAE
   P77  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~RIFYID VGNLPKNKAE
SS4028  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~SIFYID VGNLPKNKAE
SS4019  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~RIFYID VGNLPKNKAE
SS4021  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~RIFYID VGNLPKNKAE
   44B  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~SIFYID VGNLPKNKAE
 SPWM1  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~IFYID VGNLPKNKAE
   P79  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~SIFYID VGNLPKNKAE
   P81  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~FLPRIKSE
GS2747  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~SIFYID VGNLPKNKAE
  SE27  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~SIFYID VGNLPKNKAE
 S2608  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~RIFYID VGNLPKQKAE
   P17  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~XIFYID VGNLPKVKAE
  SE15  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~SIFYID VGNLPKVKAE
   32A  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~RIFYID VGNLPKVKAE
  SE17  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~RIFYID VGNLPKVKAE
   27A  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~RIFYID VGNLPKVKAE
  SE26  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~SIFYID VGNLPKVKAE
    P6  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~*IFYID VGNLPKVKAE
  SE18  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~SIFYID VGNLPKVKAE
   P12  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~FYID VGNLPKVKAD
  SE34  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~RIFYID VGNLPKVKAE
 S-PM2  MTLSFLHKAI KSLNQLRMIE DSLVIYRLSR APERRIFYID VGNLPKVKAE
SS4073  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~SIFYID VGNLPKVKAE
SS4713  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~SIFYID VGNLPKVKAE
GS2713  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~SIFYID VGNLPKVKAE
SS4036  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~RIFYID VGNLPKVKAE
SS4714  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~RIFYID VGNLPKMKAE
SS4715  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~RIFYID VGNLPKMKAE
   31B  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~LPKMKAE
S-BnM1  MVLSYLHKAI KALNQLRMIE DSLVIYRLSR APERRIFYID VGNLPKVKAE
GS2624  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~RIFYID VGNLPKMKAE
SS4850  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~RIFYID VGNLPKMKAE
SS4055  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~RIFYID VGNLPKMKAE
GS2736  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~SIFYID VGNLPKMKAE
SS4723  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~SIFYID VGNLPKVKAE
GS2711  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~SIFYID VGNLPKQKAE
SS4813  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~RIFYID VGNLPKQKAE
SS4016  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~SIFYID VGNLPKQKAE
SS4702  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~SIFYID VGNLPKQKAE
GS2704  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~RIFYID VGNLPKMKAE
SS4046  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~RIFYID VGNLPKMKAE
GS2650  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~SIFYID VGNLPKMKAE
SS4705  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~RIFYID VGNLPKMKAE
SS4804  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~SIFYID VGNLPKMKAE
SS4020  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~SIFYID VGNLPKMKAE
GS2739  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~RIFYID VGNLPKMKAE
```

Figure A1. Alignment of amino acid sequences of cyanophage isolates and representative clones from 6 natural virus communities.

```
  SE29  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~SIFYID VGNLPKMKAE
  SE36  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~RIFYID VGNLPKAKAE
  SE37  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~RIFYID VGNLPKAKAE
  SE38  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~RIFYID VGNLPKAKAE
SS4716  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~SIFYID VGNLPKQKAE
    T4  NIIGYLHRAV KPANQLKLLE DAVVIYRITR APDRRVWYVD TGNMPARKAA

        301                                                 350
   SE1  EYLKSVMTRH RNKLVYDAET GQIRDDRKHM SMLEDFWLPR REGGRGTEIS
   SE2  EYLKNVMTRY RNKLVYDAET GQIRDDRKHM SMLEDFWLPR REGGRGTEIS
S-WHM1  QYLREVMSRY RNKMVYDANT GEIKDDKKFM SMLEDFWLPR REGGRGTEIS
   P77  QYLREVMSRY RNKLVYDANT GEIKDDKKFM SMLEDFWLPR REGGRGTEIS
SS4028  QYLREVMSRY RNKMVYDANT GEIKDDKKFM SMLEDFWLPR REGGRGTEIS
SS4019  QYLREVMGRY RNKLVYDANT GEIKDDKKFM SMLEDFWLPR REGGRGTEIS
SS4021  QYLREVMGRY RNKLVYDANT GEIKDDKKFM SMLEDFWLPR REGGRGTEIS
   44B  QYLREVMGRY RNKLVYDANT GEIKDDKKFM SMLEDFWLPR REGGRGTEIT
 SPWM1  QYLREVMGRY RNKLVYDANT GEIKDDKKFL SMLEDFWLPR REGGRGTEIT
   P79  QYLREVMGRY RNKLVYDANT GEIKDDKKFM SMLEDFWLPR REGGRGTEIT
   P81  QYLREVMGRY RNKLVYDANT GEIKDDKKFM SMLEDFWLPR REGGRGTEIT
GS2747  QYLREVMGRY RNKLVYDANT GEIKDEKKFL SMLEDFWLPR REGGRGTEIT
  SE27  QYLREVMGRY RNKLVYDANT GEIKDDKKFM SMLEDFWLPR REGGRGTEIT
GS2608  QYLREVMGRY RNKLVYDAAT GEIRDDKKFM SMLEDFWLPR REGGRGTEIT
   P17  QYLKEVMSRY RNKLVYDAAT GEIRDDRKYM SMLEDFWLPR REGGRGTEIT
  SE15  QYLRDVMSRY RNKLVYDAQT GEMRDDKKHM SMLEDFWLPR REGGRGTEIT
   32A  QYLRDVMSRY RNKLVYDAQT GEMRDDKKHM SMLEDFWLPR REGGRGTEIT
  SE17  QYLRDVMSRY RNKLVYDAQT GEMRDDKKHM SMLEDFWLPR REGGRGTEIT
   27A  QYLRDVMSRY RNKLVYDAQT GEMRDDKKHM SMLEDFWLPR REGGRGTEIT
  SE26  QYLRDVMSRY RNKLVYDAQT GEMRDDKKHM SMLEDFWLPR REGGRATEIT
    P6  QYLRDVMSRY RNKLVYDAQT GEMRDDKKHM SMLEDFWLPR REGGRGTEIT
  SE18  QYLRDVMSRY RNKLVYDAQT GEMRDDKKHM SMLEDFWLPR REGGRGTEIT
   P12  QYLRDVMSRY RNKLVYDANT GEMRDDKKHM SMLEDFWLPR REGGRGTEIT
  SE34  QYLRDVMARY RNKLVYDAST GEMRDDKKHM SMLEDFWLPR REGGRGTEIT
 S-PM2  QYLRDVMSRY RNKLVYDGQT GEIRDDKKHM SMLEDFWLPR REGGRGTEIT
SS4073  QYLRDVMNRY RNKLVYDAGT GEIRDDKKYM SMLEDFWLPR REGGRGTEIT
SS4713  QYLRDVMNRY RNKLVYDAGT GEIRDDKKYM SMLEDFWLPR REGGRGTEIT
GS2713  QYLRDVMQRY RNKLVYDAST GEIRDDKKFM SMMEDFWLPR REGGRGTEIS
SS4036  QYLRDVMQRY RNKLVYDAST GEIRDDKKFM SMMEDFWLPR REGGRGTEIS
SS4714  QYLRDVMMRY RNKLVYDANT GEIRDDKKYM AMLEDFWLPR REGGRGTEIS
SS4715  QYLRDVMMRY RNKLVYDANT GEIRDDKKYM AMLEDFWLPR REGGRGTEIS
   31B  QYLRDVMMRY RNKLVYDANT GEIRDDKKYM AMLEDFWLPR REGGRGTEIS
S-BnM1  QYLREVMGRY RNKLVYDANT GEIKDDKKHM SMLEDFWLPR REGGRGTEIS
GS2624  QYLRDMMVKH KNRLIYDAET GQIRDDRKFM TMLEDYWLPR REGGRGTEIS
SS4850  QYLRDMMVKH KNRLIYDAET GQIRDDRKFM TMLEDYWLPR REGGRGTEIS
SS4055  QYLRDMMVKH KNRLIYDAET GQIRDDRKFM TMLEDYWLPR REGGRGTEIS
GS2736  QYLRDMMVKH KNRLIYGAET GQIRDDRKFM TMLEDYWLPR REGGRGTEIS
SS4723  QYLRDMMVKH KNRLIYDAET GQIRDDRKFM TMLEDYWLPR REGGRGTEIS
GS2711  QYLRDMMVKH KNRLVYDAST GEVRDDRKFM TMLEDFWLPR REGGRGTEIT
SS4813  QYLRDMMVKH KNRLVYDAST GEVRDDRKFM TMLEDFWLPR REGGRGTEIT
SS4016  QYLRDMMVKH KNRLVYDAST GEVRDDRKFM TMLEDFWLPR REGGRGTEIT
SS4702  QYLRDMMVKH KNRLVYDAST GEVRDDRKFM TMLEDFWLPR REDGRGTEIT
GS2704  QYLRDMMVKH KNKLTYDAQT GEVRDDRKFM TMLEDFWLPR REGGRGTEIT
SS4046  QYLRDMMVKH KNKLTYDAQT GEVRDDRKFM TMLEDFWLPR REGGRGTEIT
GS2650  QYLRDMMVKH KNKLVYDAST GEVRDDRKFM TMLEDFWLPR RDGGRGTEIT
SS4705  QYLRDMMVKH KNKLVYDAST GEVRDDRKFM TMLEDFWLPR RDGGRGTEIT
```

Figure A1. Continued.

```
SS4804  QYLRDMMVKH KNKLVYDAST GEVRDDRKFM TMLEDFWLPR RDGGRGTEIT
SS4020  QYLRDMMVKH KNKLVYDAST GEVRDDRKFM TMLEDFWLPR RDGGRGTEIT
GS2739  QYLRDMMVKH KNKLVYDAST GEVRDDRKFM TVLEDFWLPR RDGGRGTEIT
  SE29  QYLRDMMVKH KNKLVYDAAT GEVRDDRKFM TMLEDFWLPR REGGRGTEIT
  SE36  QYLRDMMVKH KNKLVYDAQS GEVRDDRRHM TMLEDFWLPR REGGRGTEIT
  SE37  QYLRDMMVKH KNKLVYDAQS GEVRDDRRHM TMLEDFWLPR REGGRGTEIT
  SE38  QYLRDMMVKH KNKLVYDANT GEIRDDRRHM TMLEDFWLPR REGGRGTEIT
SS4716  QYLRSLMNRY RSKLVYDQST GEIRDDRRHM SMLEDYWLPR REGGRGTEIS
    T4  EHMQHVMNTM KNRVVYDAST GKIKNQQHNM SMTEDYWLQR RDGKAVTEVD

        351                                                 400
   SE1  TLPGGENLGQ IDDVVYFQKR LYRSLNVPIN RL..EQEAQG ALGRSTEISR
   SE2  TLPGGENLGQ IDDVVYFQKR LYRSLNVPIN RL..EQEAQF SLGRSTEISR
S-WHM1  TLPGGQNLGE LEDVKYFQKK LYKALNVPGS RL..ETETTF NIGRAAEITR
   P77  TLPGGQNLGE LEDVKYFQKK LYKALNVPGS RL..ETETTF NIGRAAEITR
SS4028  TLPGGQNLGE LEDVKYFQKK LYKALNVPGS RL..ETETTF NVGRAAEITR
SS4019  TLPGGQNLGE LEDVKYFQKK LYKSLNVPNS RL..ETETTF NIGRAAEITR
SS4021  TLPGGQNLGE LEDVKYFQKK LYKSLNVPNS RL..ETETTF NIGRAAEITR
   44B  TLPGGQNLGE LEDVKYFQKK LYKALNVPSS RL..ETETTF NIGRAAEITR
 SPWM1  TLPGGQNLGE LEDVKYFQKK LYKALNVPSS RL..ETETTF NIGRAAEITR
   P79  TLPGGQNLGE LEDVKYFQKK LYKALNVPSS RL..ETETTF NIGRAAEITR
   P81  TLPGGQNLGE LEDVKYFQKK LYKALNVPSS RL..ETETTF NIGRAAEITR
GS2747  TLPGGQNLGE LEDVKYFQKK LYKSLNVPNS RL..ETETTF NIGRAAEITR
  SE27  TLPGGQNLGE LEDVKYFQKK LYKALNVPSS RL..ETETTF NIGRAAEITR
GS2608  TLPGGQNLGE LEDVKYFQKK LYKALNVPTS RL..ETETTF NIGRAAEITR
   P17  TLPGGQNLGE LEDVKYFQKK LYKALNVPPS RI..ESDSTF NIGRSAEITR
  SE15  TLPGGQNLGE LKDVEYFKKK LYNSLNLPPS RLT.DDNKGF NLGKTTEVLR
   32A  TLPGGQNLGE LKDVEYFKKK LFNSLNLPPS RLT.DDNKGF NLGKTTEVLR
  SE17  TLPGGQNLGE LKDVEYFKKK LFNSLNLPPS RLT.DDNKGF NLGKTTEVLR
   27A  TLPGGQNLGE LKDVEYFKKK LYNSLNLPPS RLT.DDNKGF NLGKTTEVLR
  SE26  TLPGGQNLGE LKDVEYFKKK LYNSLNLPPS RLT.DDNKGF NLGKTTEVLR
    P6  TLPGGQNLGE LKDVEYFKKK LYNSLNLPPS RLT.DDNKGF NLGKTTEVLR
  SE18  TLPGGQNLGE LKDVEYFKKK LYNSLNLPPS RLT.DDNKGF NLGKTTEVLR
   P12  TLPGGQNLGE LKDVEYFKKK LYNSLNLPPS RLT.DDNKGF NLGKTTEVLR
  SE34  TLPGGQNLGE LKDVEYFKKK LYNSLNLPPS RLT.DDNKGF NLGKTTEVLR
 S-PM2  TLPGGQNLGE LKDVEYFKKK LYNSLNLPPS RLT.DDNKAF NLGKSTEILR
SS4073  TLPGGQNLGE ITDIKYFQDK LYRSLNVPVT RAP.SNESGF DLGRSSEILR
SS4713  TLPGGQNLGE ITDIKYFQDK LYRSLNVPVT RAP.SNESGF NLGRSSEILR
GS2713  TLPGGQNLGE LADIKYFQEK LYKSLNVPQS RI..AGDGGF NLGRSSEILR
SS4036  TLPGGQNLGE LADIKYFQEK LYKSLNVPQS RI..AGDGGF NLGRSSEILR
SS4714  TLPGGQNLGE ITDIEYFKKK LYRSLNVPPS RM..DGEGGF NLGRSSEILR
SS4715  TLPGGQNLGE ITDIEYFKKK LYRSLNVPPS RM..DGEGGF NLGRSSEILR
   31B  TLPGGQNLGE ITDIEYFKKK LYRSLNVPPS RM..DGEGGF NLGRSSEILR
S-BnM1  TLPGGQNLGE LTDVEYFKKK LYRALNVPIS RME.A.EGGF NLGRSSEILR
GS2624  TLPGGQNLGE MDDVLYFQKK MYKSLNVPVS RL..EPETGM TLGRATEINR
SS4850  TLPGGQNLGE MDDVLYFQKK MYKSLNVPVS RL..EPETGM TLGRATEINR
SS4055  TLPGGQNLGE MDDVLYFQKK MYKSLNVPVS RL..EPETGM TLGRATEINR
GS2736  TLPGGQNLGE MDDVLYFQKK MYKSLNVPVS RL..EPETGM TLGRATEINR
SS4723  TLPGGQNLGE MDDVLYFQKK MYKSLNVPVS RL..EPETGM TLGRATEINR
GS2711  TLPGGQNLGE MDDVEYFRKR LYKALHVPVT RM..EAENQF NLGRATEISR
SS4813  TLPGGQNLGE MDDVEYFRKR LYKALHVPVT RM..EAENQF NLGRATEISR
SS4016  TLPGGQNLGE MDDVEYFRKR LYKALHVPVT RM..EAENQF NLGRATEISR
SS4702  TLPGGQNLGE MDDVEYFRKR LYKALHVPVT RM..EAENQF NLGRATEISR
GS2704  TLPGGQNLGE MDDVDYFRRK LYKSLNVPTS RM..EAENQF NLGRASEITR
```

Figure A1. Continued.

```
SS4046  TLPGGQNLGE MDDVDYFRRK LYKSLNVPTS RM..EAENQF NLGRASEITR
GS2650  TLPGGQNLGE MEDVDYFRRK LYKSLNVPVT RM..EADNQF NLGRASEITR
SS4705  TLPGGQNLGE MEDVDYFRRK LYKSLNVPVT RM..EADNQF NLGRASEITR
SS4804  TLPGGQNLGE MEDVDYFRRK LYKSLNVPVT RM..EADNQF NLGRASEMTR
SS4020  TLPGGQNLGE MEDVDYFRRK LYKSLSVPVT RM..EADNQF NLGRASEITR
GS2739  TLPGGQNLGE MEDVDYFRRK LYKSLNVPVT RM..EADNQF NLGRASEITR
  SE29  TLPGGQNLGE MEDVDYFRRK LYKSLNVPIT RM..EADNQF NLGRASEITR
  SE36  TLPGGQNLGE MDDVMYFQKK LYQSLNVPIT RV..EPEATY TIGRATEISR
  SE37  TLPGGQNLGE MDDVMYFQKK LYQSLNVPIT RV..EPEATY TIGRATEISR
  SE38  TLPGGQNLGE LDDVLYFQKK LYKSLNVPVS RL..EAEVNF NIGRSTEISR
SS4716  TLPGGQNLGE MTDVEYLLRK VYNALNVPIT RMM.PQD.GF NLGRSAEITR
    T4  TLPGADNTGN MEDIRWFRQA LYMALRVPLS RIPQDQQGGV MFDSGTSITR

        401                                                 450
   SE1  DELKFQKFID RLRMRFSHLF YELLKRQLIL KNVITEEDWN EYIRDLRIEY
   SE2  DELKFQKFID RLRMRFSHLF YELLKRQLIL KNIITEEDWN EYIRNLRIEY
S-WHM1  DEVKFQKFIA RLRKRFSELF MDLLKTQLIL KGVLTLEEWE EMKNHVQFDY
   P77  DEVKFQKFIA RLRKRFSELF MDLLKTQLIL KGVLTLEEWE EMKNHVQFDY
SS4028  DEVKFQKFIA RLRKRFGELF MDLLKAQVVL KGIITLEEWD EMKTHIQFDY
SS4019  DEVKFQKFVA RLRKRFGELF VDLLKTQLVL KGVVTIEEWE EMKEHIQFDY
SS4021  DEVKFQKFVA RLRKRFGELF VDLLKTQLVL KGICSIEEWE DMKEHIQFDF
   44B  DEVKFQKFVA RLRKRFSELF VDLLKTQLVL KGICSLEEWD DMKEHIQFDF
 SPWM1  DEVKFQKFVA RLRKRFSELF VDLLKTQLVL KGICSLEEWD EMKEHIQFDF
   P79  DEVKFQKFVA RLRKRFSELF TDLLKTQLVL KGICSLEEWD EMKEHIQFDF
   P81  DEVKFQKFVA RLRKRFSELF TDLLKTQLVL KGICSLEEWD EMKEHIQFDF
GS2747  DEVKFQKFVA RLRKRFSEMF IDLLKTQLVL KGIVTLEEWD TMKEHVQFDY
  SE27  DEVKFQKFVA RLRKRFSELF TDLLKTQLVL KGIVSLEEWE DMKEHIQYDY
GS2608  DEVKFQKFIA RLRKRFGEMF LDLLKTQLVL KGIISIEEWD DMKENITIDY
   P17  DEVKFQKFVT RLRKNFSTLF NDILKTQLVL KGIITLEDWE DMKEHIQYDY
  SE15  DELKFTKFIG RLRKRFSEMF QDMLKTQLIL KGVISPEDWD DMKEHIQYDY
   32A  DELKFTKFIG RLRKRFSEMF QDMLKTQLIL KGVISPEDWD DMKEHIQYDY
  SE17  DELKFTKFIG RLRKRFSELF QDMLKTQLIL KGVISPEDWD DMKEHIQYDF
   27A  DELKFTKFIG RLRKRFAEMF QDMLKTQLIL KGVIAPEDWD DMKEHIQYDF
  SE26  DELKFTKFIG RLRKRFSELF HDILKTQLIL KGVISPEDWD DMKEHIQYDY
    P6  DELKFAKFIG RLRKRFSELF HDILKTQLIL KGIITPEDWD DMKEHIQYDY
  SE18  DELKFTKCIG RLRKRFSELF HDILKTQLIL KGVISPEDWD DMKEHIQYDY
   P12  DELKFTKFIG RLRKRFSELF DDILKTQLIL KGIISPEDWD DMKEHIQYDY
  SE34  DELKFAKFIG RLRKRFSELF HDILKTQLIL KGVISPEDWD DMKEHIQYDY
 S-PM2  DELKFTKFIG RLRKRFAQLF HDILKTQLIL KGIITPEDWD DMEEHIQYDF
SS4073  DEVKFSKFVG RLRKRFSNLF NDMLRTQLLL KNIVTPEDWE LMSEHIQYDF
SS4713  DEVKFSKFVG RLRKRFSNLF NDMLRTQLLL KNIVTPEDWE LMSEHIQYDF
GS2713  DELKFTKFVG RLRKRFSNIF NDMLRTQLLL KNVVTPEDWE IMSEHIQYDF
SS4036  DELKFTKFVG RLRKRFSNIF NDMLRTQLLL KNVVTPEDWE IMSEHIQYDF
SS4714  DELKFTKFVG RLRKRFSNMF NDMLKTQLIL KNIITPEDWE TMSEHIQYDF
SS4715  DELKFTKFVG RLRKRFSNMF NDMLKTQLIL KNVITPEDWE SMSEHIQYDF
   31B  DELKFTKFVG RLRKRFSNMF NDMLRTQLLL KNVITPEDWE VMSEHIQYDF
S-BnM1  DEIKFTKFVG RLRKKFSEIF IDMLKTQLIL KGITSVEEWE DMKEYIQFDF
GS2624  DEVKFQKFIQ RLRMRFSMLF DAALEKQLVL KGHMTPEEYA DIRRNIKYDF
SS4850  DEVKFQKFIQ RLRMRFSMLF DAALEKQLVL KGHMTPEEYA DIRRNIKYDF
SS4055  DEVKFQKFIQ RLRMRFSMLF DAALEKQLVL KGHMTPEEYA DIRRNIEYDF
GS2736  DEVKFQKFIQ RLRMRFSMLF DAALEKQLVL KGHMTPEEYA DIRRNIKYDF
SS4723  DEVKFQKFIQ RLRMRFSMLF DAALEKQLVL KGHMTPEEYA DIRRNIKYDF
GS2711  DEVKFAKFVG RLRLRFSRLF QEILYTQLTL KGILKPSEWD LIKDQINYDY
SS4813  DEVKFAKFVG RLRLRFSRLF QEILYTQLTL KGILKPSEWD LIKDQINYDY
```

Figure A1. Continued.

```
SS4016  DEVKFAKFVG RLRLRFSRLF QEILYTQLTL KGILKPSEWD LIKDLINYDY
SS4702  DEVKFAKFVG RLRLRFSRLF QEILYTQLTL KGILKPSEWD LIKDQINYDY
GS2704  DGLKFNKFIQ RLRNRFNTLF NDLLEIHLAL MGVTTRKEWQ EIKQHVYYDY
SS4046  DELKFNKFIQ RLRNRFNTLF NDLLEIHLAL MGVTTRKEWQ EIKQHVYYDY
GS2650  DEIKFNKFIK RLRNRFSHLF NGLLEIQLVL KGVLSRSDWD EMRNNIHYNF
SS4705  DEIKFNKFIK RLRNRFSHLF NGLLEIQLVL KGVLSRSDWE EMRNNIHYNF
SS4804  DEIKFNKFIK RLRNRFSHLF NGLLEIQLVL KGVLSRSDWE EMRNNIHYNF
SS4020  DEIKFNKFIK RLRNRFSHLF NGLLEIQLVP KGVLSRSDWE EMRNNIHYNF
GS2739  DEIKFNKFIK RLRNRFSHLF NGLLEIQLVL KGVLSRSDWE EMRNNIHYNF
  SE29  DEIKFNKFVQ RLRNRFTHLF DGLLEIQLVL KGVLSRADWE QMRNTIHYDF
  SE36  DEVKFSKFIN RLRNRFSVLF DNLLEIQLVL KGTSTRGEWE DMKDRISYHF
  SE37  DEVKFSKFIN RLRNRFSVLF DNLLEIQLVL KGISTRGEWE DMKDQISYHF
  SE38  DELKFQKFIN RIRNKFAVLF DSLLEVQLVL RGIMTKPEWD QVRNSITYNF
SS4716  DEVTFFKFIE RLRMKFSLMF LQLLRVQLIL KGVMTEEEWN SISSDVFFMF
    T4  DELTFAKFIR ELQHKFEEVF LDPLKTNLLL KGIITEDEWN DEINNIKIEF

        451                                                 500
   SE1  IRDNHFTELK DAELLRERLQ SMEQVAPYVG KY~~~~~~~~ ~~~~~~~~~~
   SE2  VRDNHFSELK DAELLREGLQ SMEQVAPFVG KY~~~~~~~~ ~~~~~~~~~~
S-WHM1  VADNYFTELK EIEIRNERMN QVNVMDPYVG KYFSVEYMRR QVLKQTEQEI
   P77  VADNYFTELK EIEIRNERMN QVNVMDPYVG KY~~~~~~~~ ~~~~~~~~~~
SS4028  VADNYFTELK EIEIRNERMN QVNVMDPYVG KY~~~~~~~~ ~~~~~~~~~~
SS4019  IADNYFTELK DIEIRNERMN EVAQMDPYVG KY~~~~~~~~ ~~~~~~~~~~
SS4021  IADNYFSELK DIEIRNERMN EVNQMDPFVG KY~~~~~~~~ ~~~~~~~~~~
   44B  IADNYFTELK EIEIRNERMN QVATMDPFVX KX~~~~~~~~ ~~~~~~~~~~
 SPWM1  IADNYFTELK EIEIRNERMN QVATMD~~~~ ~~~~~~~~~~ ~~~~~~~~~~
   P79  IADNYFTELK EIEIRNERMN QVATMDPFVG KYL~~~~~~~ ~~~~~~~~~~
   P81  IADNYFTELK EIEIRNERMN QVATMD~~~~ ~~~~~~~~~~ ~~~~~~~~~~
GS2747  VADNYFTELK EIEIRNERMN QVATMDPFVG KY~~~~~~~~ ~~~~~~~~~~
  SE27  IADNYFTELK ETEIRNERLN LVSAMDPFVG KY~~~~~~~~ ~~~~~~~~~~
GS2608  IADSYFNELK ETEIRNERMN LVNVMDPYVG KY~~~~~~~~ ~~~~~~~~~~
   P17  VADNHFSELK QNEMINERLT IATTMDPYVG K~~~~~~~~~ ~~~~~~~~~~
  SE15  LFDNHFNELK EIEMMNQRMM TVTQMDPYVG KY~~~~~~~~ ~~~~~~~~~~
   32A  LFDNHFNELK EIEMMNQRMM TVTQMDPYV~ ~~~~~~~~~~ ~~~~~~~~~~
  SE17  LFDNHFNELK EIEMMNQRMM TVTQMDPFVG KY~~~~~~~~ ~~~~~~~~~~
   27A  LFDNHFNELK EIEMMNQRMM TVTQMDPF~~ ~~~~~~~~~~ ~~~~~~~~~~
  SE26  LFDNHFNELK KIEMMNQRMM TVTQMDPYVG KY~~~~~~~~ ~~~~~~~~~~
    P6  LFDNHFNELK EIEMMNQRMM TVTQMDPYVG ~~~~~~~~~~ ~~~~~~~~~~
  SE18  LFDNHFNELK EIEMMNQRMM SVTQMDPFVG KY~~~~~~~~ ~~~~~~~~~~
   P12  LFDNHFNELK EIEMMNQRMM SVTQMDPYVG K~~~~~~~~~ ~~~~~~~~~~
  SE34  LFDNHFNELK EIEMMNQRIS AVTQMDPYVG KY~~~~~~~~ ~~~~~~~~~~
 S-PM2  LFDNHFNELK EQEMQLQRVN LATQMDPFVG KYFSTEYIRR KILMQTENEF
SS4073  LYDNHFAELK ESELFNERLT MVAAAEPYVG KY~~~~~~~~ ~~~~~~~~~~
SS4713  LYDNHFAELK ESELFNERLT MVAAADPYVG KY~~~~~~~~ ~~~~~~~~~~
GS2713  LYDNHFSELK NAELMNERLG LAATIDPFVG KY~~~~~~~~ ~~~~~~~~~~
SS4036  LYDNHFSELK NAELMNERLG LAATIDPYVG KY~~~~~~~~ ~~~~~~~~~~
SS4714  LYDNHFSELK EAELLNERLT LAQAADPYVG KY~~~~~~~~ ~~~~~~~~~~
SS4715  LYDNHFSELK EAELLNERLT LAQAADPFVG KY~~~~~~~~ ~~~~~~~~~~
   31B  LYDNHFSELK EAELMNERLS LAATAEPYVX KXF~~~~~~~ ~~~~~~~~~~
S-BnM1  IYDNHFSELK ESELLQERIN LASAADPYVG KYFSVEYVRS KILHQTDTEM
GS2624  KQDNYFTELK ENEVITERIN TVNAVDPFVG KY~~~~~~~~ ~~~~~~~~~~
SS4850  KQDNYFTELK ENEVITERIN TVNAVDPYVG KY~~~~~~~~ ~~~~~~~~~~
SS4055  KQDNYFTELK ENEVITERIN TVNAVDPFVG KY~~~~~~~~ ~~~~~~~~~~
GS2736  KQDNYFTELK ENEVITERIN TVNAVDPFVG KY~~~~~~~~ ~~~~~~~~~~
```

Figure A1. Continued.

```
SS4723  KQDNYFTELK ENEVITERIN TVNAVDPFVG K~~~~~~~~~ ~~~~~~~~~~
GS2711  LHDNHFSELK FTEVLENRIR ILNELDPYVG KY~~~~~~~~ ~~~~~~~~~~
SS4813  LHDNHFSELK FTEVLENRIR ILNELDPYVG KY~~~~~~~~ ~~~~~~~~~~
SS4016  LHDNHFSELK FTEVLQNRIR ILNELDPYVG KY~~~~~~~~ ~~~~~~~~~~
SS4702  LHDNHFSELK FTEVLENRIR ILNELDPFVG KY~~~~~~~~ ~~~~~~~~~~
GS2704  VEDNHFSELK QTEVLTERLR LLGDIDPYVG KY~~~~~~~~ ~~~~~~~~~~
SS4046  VEDNHFSELK QTEVLTERLR LLGDIDPSVG KY~~~~~~~~ ~~~~~~~~~~
GS2650  KEDNYFSELK NSEIMTERLR IAGEVDPFVG KY~~~~~~~~ ~~~~~~~~~~
SS4705  KEDNYFSELK NSEIMTERLR IAGEVDPFVG KY~~~~~~~~ ~~~~~~~~~~
SS4804  KEDNYFSELK NSEIMTERLR IAGEVDPYVG KY~~~~~~~~ ~~~~~~~~~~
SS4020  KEDNYFSELK NSEIMTERLR IAGEVDPFVG KY~~~~~~~~ ~~~~~~~~~~
GS2739  KEDNYFSELK NSEIMTERLR IAGEVDPSVG KY~~~~~~~~ ~~~~~~~~~~
  SE29  KQDNYFSELK NAEIMTERLR LAGEIDPFVG KY~~~~~~~~ ~~~~~~~~~~
  SE36  ATDQHFEELK NAEIMSERLR LLNDLDPYVG KY~~~~~~~~ ~~~~~~~~~~
  SE37  ATDQHFEELK NAEIMSERLR LLNDLDPYVG KY~~~~~~~~ ~~~~~~~~~~
  SE38  VNDNHFEELK EGEIMAERLR ILGEVDPFVG KY~~~~~~~~ ~~~~~~~~~~
SS4716  AKDSYFSELK EAEILRNRID MAAALDPFVG KY~~~~~~~~ ~~~~~~~~~~
    T4  HRDSYFAELK EAEILERRIN MLTMAEPFIG KYISHRTAMK DILQMTDEEI
```

Figure A1. Continued.