2007

Exploring the Vast Diversity of Marine Viruses

Mya Breitbart
University of South Florida, mya@usf.edu

Luke Thompson
Massachusetts Institute of Technology

Curtis Suttle
University of British Columbia

Matthew Sullivan
Massachusetts Institute of Technology

Exploring the Vast Diversity of Marine Viruses

BY MYA BREITBART, LUKE R. THOMPSON, CURTIS A. SUTTLE, AND MATTHEW B. SULLIVAN

At abundances routinely greater than 10 million particles per milliliter, viruses are the most numerous biological entities[1] in the oceans. To put the sheer abundance of marine viruses in context, we note that they contain more carbon than 75 million blue whales and, if such viruses were joined end-to-end, they would stretch further than the nearest 60 galaxies (Suttle, 2005). While marine viruses were first described by Spencer (1955), they were largely ignored for three decades because of the relatively low abundances inferred using culture-based assays. However, since Bergh et al. (1989) recognized their numeric importance, they have been considered at least as abundant as marine microbes, and scientists have been characterizing them and trying to determine the extent of marine viral diversity. Extensive efforts have focused on understanding the role of viruses in horizontal gene transfer and microbial mortality, and on the consequent impacts on microbial abundance, diversity, and community structure.

Here, we review advances in understanding viral diversity and genome evolution, and discuss potentially fruitful areas for future research. Our emerging view of the virosphere, inferred from gigabases of sequence data ground truthed by model systems in culture, is one of immense but finely tuned genetic diversity, where viruses have seemingly endless genetic potential, yet clearly are maintaining key genetic elements to propagate their extraordinary success.

One focus area is the diversity of marine viruses and marine viral communities. Although viruses might defy traditional species concepts, it is clear that viral genetic diversity is extremely high. Mathematical modeling based on viral metagenomic data predicts that there are hundreds of thousands of viral genotypes in the world's ocean (Angly et al., 2006). This may not be surprising given that marine microbial prokaryotic and eukaryotic diversity is also enormous (e.g., Irigoien et al., 2004; Witman et al., 2004; Thompson et al., 2005; Worden, 2006), and there are likely to be multiple host-specific viruses infecting each marine organism (Moebus, 1991; Moebus, 1992; Waterbury and Valois, 1993; Wilson et al., 1993; Wichels et al., 1998; Sullivan et al., 2003). The diversity of marine viral morphologies ranges from a variety of icosahedral tailed phages (Figure 1) (Moebus, 1991; Moebus, 1992; Waterbury and

> ...65–95% of marine viral metagenomic sequences are not similar to previously described sequences, as opposed to ~ 10% for cellular metagenomic surveys.

MYA BREITBART *(mya@marine.usf.edu) is Assistant Professor, College of Marine Science, University of South Florida, St. Petersburg, FL, USA.* LUKE R. THOMPSON *is Ph.D. Candidate, Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA.* CURTIS A. SUTTLE *is Professor, Departments of Earth & Ocean Sciences, Botany, and Microbiology & Immunology, University of British Columbia, Vancouver, Canada.* MATTHEW B. SULLIVAN *is Postdoctoral Associate, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA.*

[1] Viruses themselves are nonmetabolic (outside of the infection process) and lack the standard genetic marker (ribosomal RNA) that allows routine genetic comparison of known and unknown life forms using the "Tree of Life," so they are often not considered "alive." The term "biological entities" is used to allow classification of viruses with other life forms.

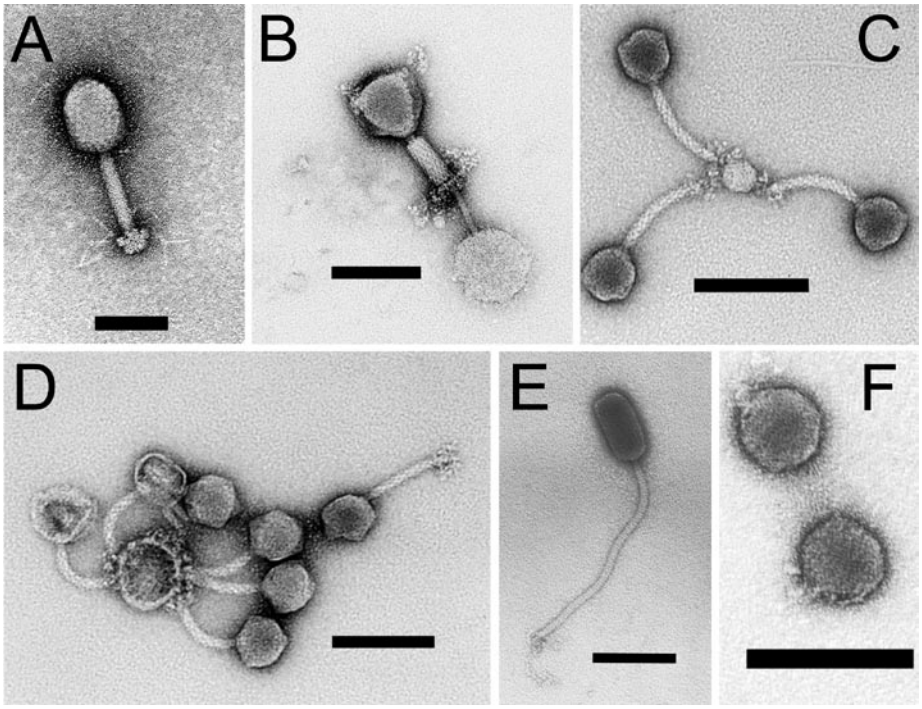

Figure 1. Electron micrographs of representative ocean cyanobacterial viruses that infect *Prochlorococcus* and *Synechococcus*. Panels A and B represent the noncontracted and contracted tails of myoviruses, respectively. Note that the tails are nonflexible and contain rather conspicuous baseplates and tail fibers. Panels C, D, and E represent siphoviruses that contain long, flexible, noncontractile tails. Note the variability in tail length, tail-terminus structures, and capsid morphology in C and D as compared to E. Panel F shows the icosahedral capsids of podoviruses that contain small, noncontractile tails. All black scale bars are 100 nm. *Photos by M.B. Sullivan, P. Weigele, and B. Ni. Images C and D were originally published in Sullivan et al. (2006)*

Valois, 1993; Proctor, 1997; Wichels et al., 1998; Sullivan et al., 2003) to long filamentous viruses (Middelboe et al., 2003) with particle diameters ranging from 25 nm (*Schizochytrium* single-stranded RNA virus SssRNAV) (Takao et al., 2005) up to ~ 300 nm for a virus that infects a marine phagotrophic protist (Garza and Suttle, 1995). Reported marine viral genome sizes range from 4.4 kilobases (kb) (Tomaru et al., 2004) to 630 kb (Ovreas et al., 2003), with representative genome sequences available from cultured isolates from nearly the extremes of the observed ranges (the 4.4 kb *Heterocapsa circularisquama* virus HcRNAV [Nagasaki et al., 2005a] and the 407 kb Coccolithovirus HeV-86 [Wilson et al., 2005]). Studies targeting genes conserved among members of a viral group (e.g., g20 and g23 of myophages [Fuller et al., 1998; Zhong et al., 2002; Marston and Sallee, 2003; Filee et al., 2005; Short and Suttle, 2005], the RNA polymerase of picorna viruses [Culley et al., 2003], or the DNA polymerase of algal viruses [Chen et al., 1996; Short and Suttle, 2002] and T7-like podophages [Breitbart and Rohwer, 2004]) demonstrate tremendous single-gene diversity even within these restricted groups of viruses. Thus, viral diversity in natural communities is enormous and dynamic as revealed at the levels of morphology, single genes, and whole genome sizes.

Recently, genomic sequencing of marine viral isolates and metagenomic sequencing of marine viral communities has revealed a plethora of previously unknown viruses. Among cultured marine phage genomes, typically between 60% and 80% of the open reading frames show no similarity to any sequences in GenBank (Paul and Sullivan, 2005), while some marine viruses infecting protists have almost no recognizable similarity to extant sequences (Nagasaki et al., 2005b). Furthermore, 65–95% of marine viral metagenomic sequences are not similar to previously described sequences (Breitbart et al., 2002, 2004; Angly et al., 2006; Culley et al., 2006), as opposed to ~ 10% for cellular metagenomic surveys (Tyson et al., 2004; Venter et al., 2004), suggesting that we have only begun to scratch the surface of marine viral sequence diversity.

One of the most striking features of this sequence diversity is an abundance of viral-encoded genes that were previously thought to be restricted to cellular genomes with metabolic capacity. For example, photosynthesis genes, which would seem of little use to something other than a photosynthetic cell, are now thought to be common in cyanophages (Mann et al., 2003; Lindell et al., 2004; Millard et al., 2004; Sullivan et al., 2006). Extensive sequencing efforts on these core photosystem II reaction-center genes show that cyanophages themselves act as genetic reservoirs for their hosts, generating diversity even at

the level of these globally distributed genes (Zeidner et al., 2005; Sullivan et al., 2006). Gene-expression studies on model phage-host pairs show that both messenger RNA (Lindell et al., 2005; Clokie et al., 2006) and protein (Lindell et al., 2005) are produced from viral photosynthesis genes during infection, which suggests that they are functional. Several other so-called "host genes," thought to be remnants of horizontal gene transfer, are present to varying degrees in cyanophages (Chen and Lu, 2002; Mann et al., 2005; Sullivan et al., 2005) and other marine phages (Rohwer et al., 2000; Miller et al., 2003; Lohr et al., 2005). Some of these genes are conserved across multiple phage lineages, such as the photosynthesis and carbon metabolism genes, which suggest that these genes play critical roles during infection, likely augmenting biochemical processes at key metabolic bottlenecks (Figure 2). For this reason, we suggest the term "auxiliary metabolic genes" (AMGs) rather than the potentially misleading term "host genes" when describing these genetic elements.

Traditionally, it was thought that the key role of viruses in microbial food webs was as agents of mortality (up to ~ 50% of prokaryotes are lysed per day by viruses; see reviews in Fuhrman [1999] and Weinbauer [2004]). However, the role of viruses in host metabolism is perhaps even more important. It is now recognized that marine viruses routinely procure AMGs to tap into critical, rate-limiting steps of host metabolism during infection (Sullivan et al., 2006; Angly et al., 2006). Such AMGs are not random evolutionary noise, but rather entrenched parts of viral genomes, akin to nucleotide-metabolism genes long known in coliphages (e.g., ribonucleotide reductase in T4-like phages), and are likely critical to the success of certain viruses in the marine environment. The impact of the role played by viruses is particularly important in environments where viral hosts have global-scale distributions (e.g., the ocean); here, viruses are likely modulating the biogeochemical cycles that run the planet.

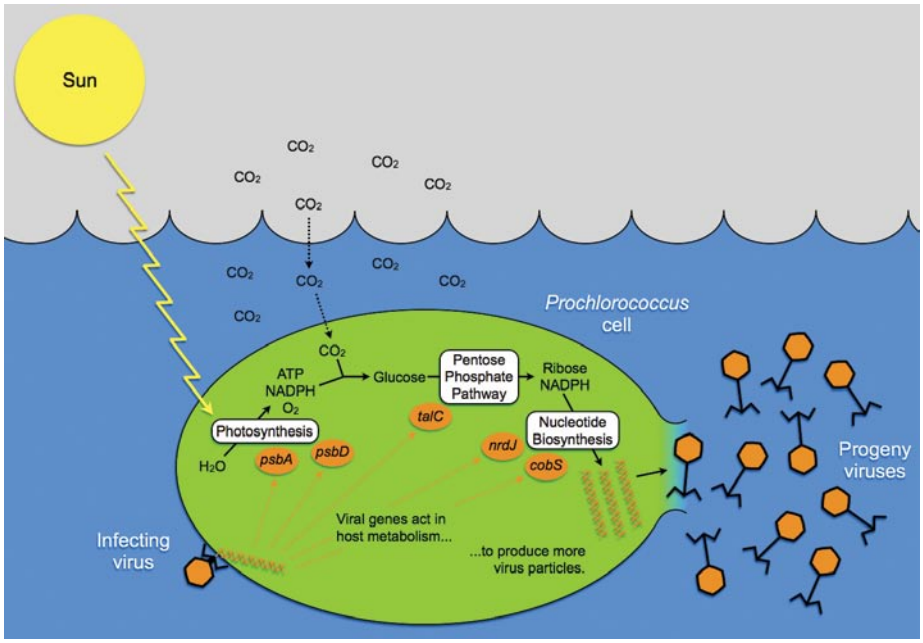

Figure 2. Schematic summarizing the potential roles of cyanophage-encoded "auxiliary metabolic genes" during infection of *Prochlorococcus*, a cyanobacterium. Three cellular metabolic pathways—photosynthesis, the pentose-phosphate pathway, and nucleotide biosynthesis—combine to make nucleotides, critical precursors for DNA replication in both cyanobacteria and their viruses. Infecting viruses often carry genes for photosystem II proteins (*psbA*, *psbD*), transaldolase (*talC*), ribonucleotide reductase (*nrdJ*), and biosynthetic enzymes for making B_{12} (*cobS*), a cofactor of ribonucleotide reductase. When expressed during infection, these genes may augment key steps in cellular metabolism, opening potential bottlenecks to increase nucleotide production, virus genomic DNA replication, and ultimately virus production.

As evidenced by work on photosynthesis genes in cyanophages, the approach of studying model systems in the laboratory is a powerful one. Model systems allow characterization of critical modeling parameters (e.g., extent and mechanisms of host range, burst size, lytic period length), complete genome sequencing to map the capacity to which a given virus might influence ecosystem processes, and, if genetic systems are available, functional assignments for unknown open reading frames. In particular, a synergistic, (meta)genomics-enabled, model-virus-host systems approach can be used to evaluate the ecological roles and the extent of marine-viral diversity. Undoubtedly, as long as the model systems approach relies upon the culturability of organisms (most marine microbes are resistant to culturing), then cautious extrapolation of laboratory results to natural

communities is warranted.

Deep exploration of the diversity and ecosystem function of marine viral communities is a daunting yet exciting task. Tremendous progress has been made using culture-based and signature-gene-based techniques, as well as through metagenomic surveys. Maximizing our interpretation of these rapidly growing metagenomic data sets will require an understanding of cloning and amplification biases of current techniques, and it will also require efforts to isolate and characterize representative viral community members. Future challenges include the development of genetic tools for tracking all major marine groups (e.g., in situ hybridization sequence-based assays using signature genes), the expansion of "snapshot" metagenomic characterizations to evaluate the temporal and spatial dynamics of natural communities, and the development of a robust theoretical framework to enhance our ability to model and predict the impacts of viruses on global ecosystem function.

For further reading on marine viruses, see the following comprehensive reviews: Dunigan et al., 2006; Fuhrman, 1999; Proctor, 1997; Suttle, 2005; Weinbauer, 2004; and Wommack and Colwell, 2000.

ACKNOWLEDGEMENTS

MBS and LRT are partially supported by the Gordon and Betty Moore Foundation and NSF grants to Sallie W. Chisholm.

It is now recognized that marine viruses routinely procure auxiliary metabolic genes to tap into critical, rate-limiting steps of host metabolism during infection.

REFERENCES

Angly, F.E., B. Felts, M. Breitbart, P. Salamon, R.A. Edwards, C. Carlson, A.M. Chan, M. Haynes, S. Kelley, H. Liu, and others. 2006. The marine viriomes of four oceanic regions. *PLoS Biology* 4:e368.

Bergh, Ø., K.Y. Børsheim, G. Bratbak, and M. Heldal, 1989. High abundance of viruses found in aquatic environments. *Nature* 340:467–468.

Breitbart, M., B. Felts, S. Kelley, J.M. Mahaffy, J. Nulton, P. Salamon, and F. Rohwer. 2004. Diversity and population structure of a near-shore marine sediment viral community. *Proceedings of the Royal Society B* 271:565–574.

Breitbart, M., and F. Rohwer. 2004. Global distribution of nearly identical phage-encoded DNA sequences. *FEMS Microbiology Letters* 236: 245–252.

Breitbart, M., P. Salamon, B. Andresen, J.M. Mahaffy, A.M. Segall, D. Mead, F. Azam, and F. Rohwer. 2002. Genomic analysis of uncultured marine viral communities. *Proceedings of the National Academy of Sciences of the United States of America* 99:14,250–14,255.

Chen, F., and J. Lu. 2002. Genomic sequence and evolution of marine cyanophage P60: A new insight on lytic and lysogenic phages. *Applied and Environmental Microbiology* 68:2,589–2,594.

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. *Applied and Environmental Microbiology* 62:2,869–2,874.

Clokie, M.R.J., J. Shan, S. Bailey, Y. Jia, H.M. Krisch, S. West, and N.H. Mann. 2006. Transcription of a 'photosynthetic' T4-type phage during infection of a marine cyanobacterium. *Environmental Microbiology* 8:827–835.

Culley, A.I., A.S. Lang, and C.A. Suttle. 2003. High diversity of unknown picorna-like viruses in the sea. *Nature* 424:1,054–1,057.

Culley, A.I., A.S. Lang, and C.A. Suttle. 2006. Metagenomic analysis of coastal RNA virus communities. *Science* 312:1,795–1,798.

Dunigan, D.D., L.A. Fitzgerald, and J.L. Van Etten. 2006. Phycodnaviruses: A peek at genetic diversity. *Virus Research* 117:119–132.

Filee, J., F. Tetart, C.A. Suttle and H.M. Krisch. 2005. Marine T4-type bacteriophages, a ubiquitous component of the dark matter of the biosphere. *Proceedings of the National Academy of Sciences of the United States of America* 102:12,471–12,476.

Fuhrman, J.A. 1999. Marine viruses: Biogeochemical and ecological effects. *Nature* 399:541–548.

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 g20 and its application to PCR-based detection and quantification techniques. *Applied and Environmental Microbiology* 64:2,051–2,060.

Garza, D.R., and C.A. Suttle. 1995. Large double-stranded DNA viruses which cause the lysis of marine heterotrophic nanoflagellates (*Bodo* sp.) occur in natural marine virus communities. *Aquatic Microbial Ecology* 9:203–210.

Irigoien, X., J. Huisman, and R.P. Harris. 2004. Global biodiversity patterns of marine phytoplankton and zooplankton. *Nature* 429:863–867.

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

Lindell, D., M.B. Sullivan, Z.I. Johnson, A.C. Tolonen, F. Rohwer, and S.W. Chisholm. 2004. Photosynthesis genes in *Prochlorococcus* cyanophage. *Proceedings of the National Academy of Sciences of the United States of America* 101:11,013–11,018.

Lohr, J.E., F. Chen, and R.T. Hill. 2005. Genomic analysis of bacteriophage fJL001: Insights into its interaction with a sponge-associated alpha-proteobacterium. *Applied and Environmental Microbiology* 71:1,598–1,609.

Mann, N.H., M.R.J. Clokie, A. Millard, A. Cook, W.H. Wilson, P.J. Wheatley, A. Letarov, and H.M. Krisch. 2005. The genome of S-PM2, a "photosynthetic" T4-type bacteriophage that infects marine *Synechococcus* strains. *Journal of Bacteriology* 187:3,188–3,200.

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

Marston, M.F., and J.L. Sallee. 2003. Genetic diversity and temporal variation in the cyanophage community infecting marine *Synechococcus* species in Rhode Island's coastal waters. *Applied and Environmental Microbiology* 69:4,639–4,647.

Middelboe, M., R.N. Glud, and K. Finster. 2003. Distribution of viruses and bacteria in relation to diagenetic activity in an estuarine sediment. *Limnology and Oceanography* 48:1,447–1,456.

Millard, A., M.R.J. Clokie, D.A. Shub, and N.H. Mann. 2004. Genetic organization of the *psbAD* region in phages infecting marine *Synechococcus* strains. *Proceedings of the National Academy of Sciences of the United States of America* 101:11,007–11,012.

Miller, E.S., J.F. Heidelberg, J.A. Eisen, W.C. Nelson, A.S. Durkin, A. Ciecko, T.V. Feldblyum, O. White, I.T. Paulsen, W.C. Nierman, and others. 2003. Complete genome sequence of the broad-host-range vibriophage KVP40: Comparative genomics of a T4-related bacteriophage. *Journal of Bacteriology* 185:5,220–5,233.

Moebus, K. 1991. Preliminary observations on the concentration of marine bacteriophages in the water around Helgoland. *Helgolander Meeresunters* 45:411–422.

Moebus, K. 1992. Further investigations on the concentration of marine bacteriophages in the water around Helgoland, with reference to the phage-host systems encountered. *Helgolander Meeresunters* 46:275–292.

Nagasaki, K., Y. Shirai, Y. Takao, H. Mizumoto, K. Nishida, and Y. Tomaru. 2005a. Comparison of genome sequences of single-stranded RNA viruses infecting the bivalve-killing dinoflagellate *Heterocapsa circularisquama*. *Applied and Environmental Microbiology* 71:8,888–8,894.

Nagasaki, K., Y. Tomaru, Y. Takao, K. Nishida, Y. Shirai, H. Suzuki, and T. Nagumo. 2005b. Previously unknown virus infects marine diatom. *Applied and Environmental Microbiology* 71:3,528–3,535.

Ovreas, L., D. Bournde, R.A. Sandaa, E. Casamayor, S. Benlloch, V. Goddard, G. Smerdon, M. Heldal, and T.F. Thingstad. 2003. Response of bacterial and viral communities to nutrient manipulations in seawater mesocosms. *Aquatic Microbial Ecology* 31:109–121.

Paul, J.H., and M.B. Sullivan. 2005. Marine phage genomics: What have we learned? *Current Opinion in Biotechnology* 16:299–307.

Proctor, L.M. 1997. Advances in the study of marine viruses. *Microscopy Research and Technique* 37:136–161.

Rohwer, F., A.M. Segall, G. Steward, V. Seguritan, M. Breitbart, F. Wolven, and F. Azam. 2000. The complete genomic sequence of the marine phage Roseophage SIO1 shares homology with non-marine phages. *Limnology and Oceanography* 42:408–418.

Short, C.M., and C.A. Suttle. 2005. Nearly identical bacteriophage structural gene sequences are widely distributed in marine and freshwater environments. *Applied and Environmental Microbiology* 71:480–486.

Short, S.M., and C.A. Suttle. 2002. Sequence analysis of marine virus communities reveals that groups of related algal viruses are widely distributed in nature. *Applied and Environmental Microbiology* 68:1,290–1,296.

Spencer, R. 1955. A marine bacteriophage. *Nature* 175:690.

Sullivan, M.B., M.L. Coleman, P. Weigele, F. Rohwer, and S.W. Chisholm. 2005. Three *Prochlorococcus* cyanophage genomes: Signature features and ecological interpretations. *PLoS Biology* 3:790–806.

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

Sullivan, M.B., J.B. Waterbury, and S.W. Chisholm. 2003. Cyanophages infecting the oceanic cyanobacterium *Prochlorococcus*. *Nature* 424:1,047–1,051.

Suttle, C.A. 2005. Viruses in the sea. *Nature* 437:356–361.

Takao, Y., K. Nagasaki, K. Mise, T. Okuno, and D. Honda. 2005. Isolation and characterization of a novel single-stranded RNA virus infectious to a marine fungoid protist, *Schizochytrium sp.* (Thraustochytriaceae, Labyrinthulea). *Applied and Environmental Microbiology* 71:4,516–4,522.

Thompson, J.R., S. Pacocha, C. Pharino, V. Klepac-Ceraj, D.E. Hunt, J. Benoit, R. Sarma-Rupavtarm, D.L. Distel, and M.F. Polz. 2005. Genotypic diversity within a natural coastal bacterioplankton community. *Science* 307:1,311–1,313.

Tomaru, Y., N. Katanozaka, K. Nishida, Y. Shirai, K. Tarutani, M. Yamaguchi, and K. Nagasaki. 2004. Isolation and characterization of two distinct types of HcRNAV, a single-stranded RNA virus infecting the bivalve-killing microalga *Heterocapsa circularisquama*. *Aquatic Microbial Ecology* 34:207–218.

Tyson, G.W., J. Chapman, P. Hugenholtz, E.E. Allen, R.J. Bram, P.M. Richardson, V.V. Solovyev, E.M. Rubin, D.S. Rokhsar, and J.F. Banfield. 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* 428:37–43.

Venter, J.C., K. Remington, J.F. Heidelberg, A.L. Halpern, D. Rusch, J.A. Eisen, D. Wu, I. Paulsen, K.E. Nelson, W. Nelson, and others. 2004. Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304:66–74.

Waterbury, J.B., and F.W. Valois. 1993. Resistance to co-occurring phages enables marine *Synechococcus* communities to coexist with cyanophages abundant in seawater. *Applied and Environmental Microbiology* 59:3,393–3,399.

Weinbauer, M.G. 2004. Ecology of prokaryotic viruses. *FEMS Microbiology Reviews* 28:127–181.

Wichels, A., S.S. Biel, H.R. Gelderblom, T. Brinkhoff, G. Muyzer, and C. Schuett. 1998. Bacteriophage diversity in the North Sea. *Applied and Environmental Microbiology* 64:4,128–4,133.

Wilson, W., D.C. Schroeder, M.J. Allen, M.T.G. Holden, J. Parkhill, B.G. Barrell, C. Churcher, N. Hamlin, K. Mungall, H. Norbertczak, and others. 2005. Complete genome sequence and lytic phase transcription profile of a coccolithovirus. *Science* 5737:1,090–1,092.

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. *Applied and Environmental Microbiology* 59:3,736–3,743.

Witman, J.D., R.J. Etter, and F. Smith. 2004. The relationship between regional and local species diversity in marine benthic communities: A global perspective. *Proceedings of the National Academy of Sciences of the United States of America* 101:15,664–15,669.

Wommack, K.E., and R.R. Colwell. 2000. Virioplankton: Viruses in aquatic ecosystems. *Microbiology and Molecular Biology Reviews* 64:69–114.

Worden, A.Z. 2006. Picoeukaryote diversity in coastal waters of the Pacific Ocean. *Aquatic Microbial Ecology* 43:165–175.

Zeidner, G., J.P. Bielawski, M. Shmoish, D.J. Scanlan, G. Sabehi, and O. Beja. 2005. Potential photosynthesis gene recombination between *Prochlorococcus* and *Synechococcus* via viral intermediates. *Environmental Microbiology* 7:1,505–1,513.

Zhong, Y., F. Chen, S.W. Wilhelm, L. Poorvin, and R.E. Hodson. 2002. Phylogenetic diversity of marine cyanophage isolates and natural virus communities as revealed by sequences of viral capsid assembly protein gene g20. *Applied and Environmental Microbiology* 68:1,576–1,584.