

Oct 30, 2015

The construction and analysis of marker gene libraries

DOI

<https://dx.doi.org/10.17504/protocols.io.dpi5kd>

Steven M. Short¹, Feng Chen¹, and Steven W. Wilhelm¹

¹Manual of Aquatic Viral Ecology

VERVE Net



Steven W W Wilhelm

The University of Tennessee, Knoxville

Create & collaborate more with a free account

Edit and publish protocols, collaborate in communities, share insights through comments, and track progress with run records.

Create free account

OPEN  ACCESS



DOI: <https://dx.doi.org/10.17504/protocols.io.dpi5kd>

Protocol Citation: Steven M. Short, Feng Chen, and Steven W. Wilhelm 2015. The construction and analysis of marker gene libraries. protocols.io <https://dx.doi.org/10.17504/protocols.io.dpi5kd>

License: This is an open access protocol distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Protocol status: Working



Created: August 27, 2015

Last Modified: November 09, 2017

Protocol Integer ID: 1482

Keywords: virus marker gene, analysis of marker gene libraries marker gene, analysis of marker gene library, marker gene libraries marker gene, marker gene library, manual of aquatic viral ecology, aquatic viral ecology, specific virus, infecting algae, related virus, heterotrophic flagellate, discussion on virus, virus, aquatic sample, bacteria, marker

Abstract

Marker genes for viruses are typically amplified from aquatic samples to determine whether specific viruses are present in the sample, or to examine the diversity of a group of related viruses. In this [protocol], we will provide an overview of common methods used to amplify, clone, sequence, and analyze virus marker genes, and will focus our discussion on viruses infecting algae, bacteria, and heterotrophic flagellates.

Short, S. M., F. Chen, and S. W. Wilhelm. 2010. The construction and analysis of marker gene libraries, p. 82–91. In S. W. Wilhelm, M. G. Weinbauer, and C. A. Suttle [eds.], *Manual of Aquatic Viral Ecology*. ASLO.

Guidelines

Viral gene markers

Viruses are probably the most diverse biological entity in the biosphere. Despite the fact that no universal gene marker (like the 16S and 18S ribosomal ribonucleic acid genes from prokaryotes and eukaryotes, respectively) is available for all viruses, many studies have demonstrated that certain genes are conserved among certain groups of viruses that infect closely related hosts. By designing oligonucleotide primers that hybridize to conserved regions of these marker genes, many researchers have used PCR to amplify virus marker genes from environmental samples to investigate the genetic diversity of specific groups of viruses in variety of aquatic environments (see **Table 1**). Currently, viral capsid related genes and virus-encoded deoxyribonucleic acid (DNA)/ribonucleic acid (RNA) polymerase gene are the most widely used genetic markers for aquatic viruses, and various PCR primer sets have been designed to target these genetic markers (Table 1). Studies of these virus marker genes have demonstrated that viruses in the marine environments are much more diverse than might be expected based on the limited numbers of cultivated viruses. With the recent rapid increase in the number of microbial genes and genomes available in public sequence databases, many viral signature genes (e.g., genes involved in photosynthesis or DNA replication) have been identified. By taking advantage of the plethora of information now available in sequence databases (e.g., NCBI's GenBank database at <http://www.ncbi.nlm.nih.gov/Genbank/>), polymerase chain reaction (PCR)-based methods can provide a rapid, sensitive, and economical approach to explore the diversity of viral genes or viral groups in nature and address important questions about the distribution, diversity, and even activity of virus in aquatic ecosystems.

Sample collection and preparation

The history and details of proper sample collection and processing before PCR amplification of virus genes are numerous. In some cases, virus markers can readily be amplified directly from unaltered whole water samples. In other cases, preconcentration of virus particles may be required; this process is thoroughly explained in another chapter (Wommack et al. 2010, this volume). For qualitative purposes, PCR amplification is often most successful from concentrated virus communities. However, the variety of steps involved in either ultrafiltration or ultracentrifugation increases the potential for particle loss, which can complicate quantitative analyses. Ultimately, the ambient abundance of viruses and the sensitivity of the particular assay will dictate the approach taken in preparing samples for analysis.

Similarly, a debate continues as to whether nucleic acids need to be extracted from virus samples prior to PCR amplification, or whether viral genetic material can be directly amplified. Many of the early studies on virus diversity in aquatic systems employed virus concentrates (see Wommack et al. 2010, this volume) as starting material. More recently, researchers have directly amplified marker elements from unextracted virusbearing samples (Short and Short 2008; Wilhelm and Matteson 2008). Moreover, when comparing PCR amplification of unextracted virus concentrates and polyethylene glycol (PEG) precipitated virus concentrates to extracted viral DNA, the extracted DNA often produced poor PCR amplification yields (Chen et al. unpubl. results). While the approach of using unextracted virus DNA is often quite successful and requires only slight changes to the PCR protocol, its efficacy may depend on the capsid/membrane composition of the virus in question. Nonetheless, a simple freeze/heat treatment consisting of 3 repetitions of freezing virus samples until solid followed heating to

95°C for 2 min has been used to generate PCR-amplifiable virus DNA from a variety of aquatic samples (Chen et al. 1996; Short and Short 2008; Short and Suttle 2002).

Primer Design

In targeting a specific population or group of microorganisms in aquatic environments using PCR-based methods, primer design is often the most critical and challenging step. Thankfully, because PCR is a well established technique, many excellent volumes have been written on the optimization and application of PCR, and most include some discussion of the critical considerations for primer design (e.g., Altshuler 2006; Atlas 1993; Innis et al. 1990; Mcpherson and Moller 2006), and some focus entirely on primer design (Yuryev 2007). In addition, freely available software can be found on the Internet that can aid in primer design. For example, the program OligoAnalyzer 3.1 is available at the Integrated DNA Technologies Web site (<http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/>). This particular software allows the user to enter oligonucleotide (primer) sequences and provides general analytical information such as predicted melting temperatures for each primer, as well as more complicated but useful information such as the primer's potential for hairpin formation, self-dimer formation, and hetero-dimer formation. This software also allows the user to directly compare their primer sequences to sequences archived in the GenBank database.

As a general guideline, several criteria should be considered when designing PCR primers for the analysis of aquatic viruses:

- 1) the target gene should be evolutionarily conserved among the viruses of interest;
- 2) at least one region with a minimum of 6 consecutive amino acids (or >16 nucleotides) that is conserved only among the target organisms can be identified in multiple sequence alignments;
- 3) when multiple regions are available for primers, regions with the least degeneracy should be considered;
- 4) at sites of 4-fold degeneracy where G, A, T, and C should all be considered, the practical degeneracy of the primer can be reduced by using an inosine residue;
- 5) the desired size of PCR products may vary for different applications (e.g., shorter PCR amplicons ranging from 150-400 bp are ideal for DGGE applications and quantitative PCR (qPCR), whereas longer amplicons ranging from 500-800 bp are desirable for the phylogenetic analyses of clone libraries;
- 6) more than one set of primers should be designed and tested when multiple target regions are available;
- 7) because the design of specific PCR primers relies on the number of known target sequences, it is important to include as many related sequences as possible when creating sequence alignments for primer design;
- 8) PCR primers should be modified (redesigned) as more sequences belonging to the target organisms become available.

In some cases, PCR primers (e.g., primers that target the *g20* gene of cyanomyoviruses) were originally designed based on a limited number of gene sequences. This can result in poorly constrained sequence information since the specificity of primers was not well defined in the first place. Although it can easily be argued that poorly constrained sequence data are more valuable than no data at all, it is nonetheless important to use as much sequence data from representative groups of viruses when designing and redesigning PCR primers (**Fig. 1**). For example, using newly available sequence data, *g20* primers specific for cyanomyoviruses have been modified, and much higher PCR specificity has been achieved (Chen et al. unpubl. data; more details are described below; **Fig. 2**).

PCR Amplification

PCR is a widely used in vitro technique that generates millions or even billions of copies of specific gene fragments. There are numerous general and field-specific procedural references for PCR, and almost all of the major scientific vendors distribute PCR reagents and equipment. Therefore, this section will only provide a simple guide to help neophyte molecular biologists get started; obviously there are far too many options that could be considered for a particular PCR application to discuss them here. Whenever possible, the procedure outlined in published literature describing the use of a particular set of primers should be followed. However, researchers should not be surprised when they need to troubleshoot previously described conditions for a particular reaction. In our experience, different Taq DNA polymerases, thermal cyclers and reagents, and even different workers, can have a dramatic influence on PCR results.

One of the most important considerations for PCR is lab hygiene. Because of its sensitivity, PCR reactions can easily be contaminated with amplifiable DNA. It is much easier to take proactive measures to prevent contamination than to have to track down the source of contamination after it has been detected. All reagents should be dispensed into small portions or working stocks before their use. This practice has the double benefit of preventing the loss of large stocks of reagents in the event that they become contaminated, and it also minimizes the number of freeze-thaw cycles that a reagent endures. The use of aerosol barrier tips for automatic pipettors, frequent sanitization of lab benches, and dedicated lab spaces or sterile hoods for setting up PCRs are also highly recommended. Although lab coats are generally recommended as essential personal protective equipment, they must be washed frequently if workers are to wear them when setting up PCR reactions; a dirty sleeve can be a major reservoir for contaminating nucleic acids! As a final comment, although it may seem obvious, it cannot be stressed enough that positive and negative controls must be included in every single PCR experiment.

PCR reactions are set up via the creation of a master mix that includes all reagents except the template nucleic acid. Generally, it is wise to prepare a slightly larger volume master mix that is absolutely necessary because the wasted reagents represent a trivial expense, and minor pipettor inaccuracies can lead to a short fall when dispensing the master mix into individual reaction tubes. The following reagents and concentrations are typical for many PCR reactions:

- PCR buffers are usually supplied at a 10x or 2x concentration with the polymerase enzyme. The buffers components are somewhat variable and are optimized by the manufacturer for use with a particular thermally-stable DNA polymerase enzyme.
- MgCl₂ is usually supplied in a 50 or 25 mM stock. The working concentration can vary between 1.5 to 4.0 mM depending on the primer sequences. For any particular PCR protocol, the optimal working concentration should be empirically determined as it can have a dramatic effect on the yield of PCR products and the stringency of the reaction.

- dNTPs can be purchased individually, or in mixtures of all four nucleotides. Generally, dNTPs are mixed and stored as stock solutions with each dNTP at a concentration of 10 mM, or 40 mM total for all dNTPs. For most PCR protocols, final concentrations of 0.2 mM of each dNTP is sufficient and provides ample product yield without negatively affecting the PCR specificity or fidelity.
- oligonucleotide primers can be ordered as lyophilized stocks and can be resuspended in sterile, pure water or TE buffer (10 mM Tris, 0.1 mM EDTA, pH 7.5) for long-term storage at a concentration of 100 μ M. Generally, aliquots of working stocks are made at 10 μ M and the final concentration of each primer in a PCR reaction can range from 0.1 to 1.0 μ M (i.e., a total of 10 to 100 pmol of primer in a final reaction volume of 50 μ L) depending on the primer. Generally, PCR with degenerate primers require slightly higher primer concentrations, but the optimal primer concentration should be determined empirically.
- thermally stable DNA polymerases are the key ingredient in PCR as these enzymes withstand the extreme temperature fluctuations of thermal cycling. For many years, *Taq* DNA polymerase was the standard enzyme using for PCR. However, many vendors now produce various enzymes or enzyme mixtures that are optimized for long PCR (amplification of fragments >10 Kb), or high fidelity amplification. Additionally, most manufacturers now produce reasonably priced hot-start enzymes that are not active until after the initial denaturation step. These hot-start enzymes are very useful as they prevent amplification artifacts produced by nonspecific primer annealing during the initial ramping up to the denaturation temperature.
- H₂O is added in sufficient volume to bring the total volume up to that desired for each reaction (the total volume for individual reactions is typically 25 or 50 μ L depending on the desired yield). Although it is often overlooked as a potential source of amplification difficulties, H₂O quality is critically important. When possible, certified nucleasefree water should be used, but good results can be obtained with pure water that has been ultrafiltered and is ion free (i.e., 18.3 M Ω -cm resistivity).

Cloning and Sequencing

By design, PCR methods for amplifying nucleic acids from aquatic viruses use universal primers that target related but different gene sequences. Because Sanger (dideoxy-based) sequencing reactions are confounded when more than one template is present, gene fragments from natural populations must be separated before sequencing reactions can be conducted. The most common approach to separate individual amplified gene fragments is to clone the PCR products into a plasmid vector, transform bacterial cells with the recombinant plasmids, and purify plasmids from individual isolated bacterial colonies; generally, each colony will contain only one type of recombinant plasmid. Purified plasmids can then be used as templates for sequencing reactions. Other methods like denaturing gradient gel electrophoresis (DGGE) can also be used to separate individual gene fragments from complex mixtures of PCR products. For these methods, individual bands that theoretically represent only a single DNA fragment are excised from the gel and are re-amplified with second round of PCR. After these second round PCR products are purified, they can then be used as templates for sequencing reactions.

Cloning PCR products has become relatively routine, and many manufacturers produce kits that can be used. Although the cost of the kits may exceed the cost of reagents prepared inhouse, the time savings and efficacy of the kits far exceeds the relatively minor increased cost of cloning. The same statement can be made for most kit-based molecular methods, and therefore we have included below, lists of some common kits that can be used for many of the steps involved in the creation of marker gene libraries. The list of kits that we have provided is not meant to indicate any preferences or be all inclusive. Rather, the lists that follow are included to simply suggest a few reliable sources for these kits; many other manufacturers produce similar kits that may be equally cost effective and efficient. Most cloning or DNA purification kits include detailed instructions and trouble-shooting guides, and generally the manufacturer's recommendations and protocols should be followed. The competent cells used for bacterial plasmid transformation are often included in cloning kits, or they can be purchased separately. Although competent cells prepared by individual labs are considerably less expensive than commercially prepared cells, the effort to produce them may not be worth the cost savings unless they will be used routinely. Most general molecular biology manuals provide a protocol for the preparation of competent cells (Ausubel et al. 2002; Sambrook et al. 1989). Two types of kits are available for cloning PCR products. Some are based on a TA-cloning method that takes advantage of the single deoxadenosine overhang left by *Taq* DNA polymerase and other non-proofreading polymerase enzymes, while others are designed to clone blunt-ended PCR products. In either case, the number of colonies that contain recombinant plasmids with the desired PCR fragment can be greatly enhanced by loading all of the PCR reaction in an agarose gel, excising the fragment of the appropriate size, and purifying the fragment using a commercial gel extraction kit. In our experience, this step greatly reduces the possibility of ligating primer-dimers or other PCR artifacts into the plasmid vector, thereby enhancing the recovery of clones containing the gene fragment of interest.

Like PCR product cloning, DNA sequencing has become routine despite the high cost of the instruments used for automated sequence analysis. Generally, because high throughput or multi-user sequencing facilities offer sequencing services at significantly reduced cost compared with sequencing within individual labs, they have become the most common option for nucleotide sequencing. Many academic institutions and private companies provide sequencing services at a reasonable cost, and a brief web search should reveal many options for sequencing services. Sequencing reagents are produced by several manufacturers and vary depending on the automated sequencing instrument used. Most, if not all, sequencing facilities will recommend specific reagent kits and protocols for their users. The most important consideration for obtaining good sequencing results is the purity of the sequencing template as a poor quality template DNA is the most common cause for failed sequencing reactions. Therefore, no matter if sequencing templates are purified plasmids or PCR products, we highly recommend the use of commercial DNA purification kits because of their ease of use and the consistent DNA purity that they provide.

Common UA- or TA-based PCR cloning kits:

- Fermentas InsTAclone™ PCR Cloning Kit (<http://www.fermentas.com/>)
- Invitrogen TOPO TA Cloning® Kit (<http://www.invitrogen.com/>)
- Promega pGEM-T and pGEM-T Easy Vector Systems (<http://www.promega.com/>)
- Stratagene StrataClone™ PCR Cloning Kit (<http://www.stratagene.com/>)

Common blunt-end PCR cloning kits:

- Clontech In-Fusion™ PCR Cloning Kits (<http://www.clontech.com/>). Note: although this kit does not require deoxyadenosine ("A") overhangs on PCR fragments to be cloned; blunt-end polishing is also not required.
- Fermentas CloneJET™ PCR Cloning Kit (<http://www.fermentas.com/>)
- Invitrogen Zero Blunt® TOPO® PCR Cloning Kit (<http://www.invitrogen.com/>)
- Stratagene StrataClone™ Blunt PCR Cloning Kit (<http://www.stratagene.com/>)

Common gel extraction kits:

- Fermentas DNA gel extraction kit (<http://www.fermentas.com/>)
- Invitrogen PureLink™ Gel Extraction Kit (<http://www.invitrogen.com/>)
- Promega Wizard® DNA Clean up system (<http://www.promega.com/>)
- Qiagen QIAquick Gel Extraction kit (<http://www.qiagen.com/>)
- Stratagene StrataPrep® DNA Gel Extraction Kit (<http://www.stratagene.com/>)

Common plasmid miniprep kits:

- Fermentas GeneJET™ Plasmid Miniprep Kit (<http://www.fermentas.com/>)
- Invitrogen ChargeSwitch® NoSpin Plasmid Micro Kit (<http://www.invitrogen.com/>)
- Promega Wizard® Plus Minipreps DNA purification system (<http://www.promega.com/>)
- Qiagen QIAprep Spin Miniprep Kit (<http://www.qiagen.com/>)
- Stratagene StrataPrep® Plasmid Miniprep Kit (<http://www.stratagene.com/>)

Common PCR cleanup kits:

- Applied Biosystems DNAClear™ kit (<http://www.appliedbiosystems.com/>)
- Fermentas DNA gel extraction kit (<http://www.fermentas.com/>)
- Invitrogen ChargeSwitch® PCR Clean-Up Kit (<http://www.invitrogen.com/>)
- Promega Wizard® DNA Clean up system (<http://www.promega.com/>)
- Qiagen QIAquick PCR purification kit (<http://www.qiagen.com/>)
- Stratagene StrataPrep® PCR Purification Kit (<http://www.stratagene.com/>)

Bioinformatic analysis

Once sequences have been obtained from a marker gene clone library, the steps involved in sequence analysis include 1) sequence editing, 2) sequence alignment, 3) phylogenetic inference, 4) drawing phylograms, and 5) calculating diversity indices (**Fig. 1**). Although the analysis of clone library sequences can seem daunting to the uninitiated, references such as Hall's book *Phylogenetic Trees Made Easy* (2008) offer excellent advice and background information that will walk beginners through the essential elements of sequence analysis; more in-depth discussions of phylogenetic inference can be found in advanced texts (Felsenstein 2004; Graur and Li 2000; Hillis et al. 1996).

By its very nature, bioinformatic analysis is computationally intensive and is conducted using a variety of software. In recent years, computer software and hardware has changed dramatically, and most of these changes have resulted in easy to use and widely available bioinformatic software. For example, Macintosh computers now use an Intel chip that allows them to use the Windows operating system, and there are Windows emulators available for both Linux and Unix operating systems. Therefore, to ensure that this discussion is useful to the

broadest possible audience, we have focused on the use of Windows-based software that is freely available on the World Wide Web (most of the software listed in the following paragraphs is also available in versions compatible with Unix or Macintosh operating systems). For the sake of brevity, we will not discuss the parameters that must be considered when analyzing genetic libraries. Instead, we will simply point readers to the excellent texts mentioned in the preceding paragraph, and provide a brief list of some of the available free software, noting their major functions and the Web site from which they can be downloaded:

- BioEdit (Hall 1999). This software can be used for sequence editing and much more. It is available free of charge at <http://www.mbio.ncsu.edu/BioEdit/BioEdit.html>. This software package can be used to view the chromatograms produced by several different types of automated sequencers, and it can also be used to analyze the physical properties of nucleic acid or amino acid sequences. Further, it can be used to translate DNA sequences, search sequences for defined motifs, conduct BLAST searches locally or to the GenBank database, align sequences using ClustalW, and it produces publication quality prints of sequence alignments. This is an extremely useful program that has far too many functions to list here.
- ClustalX (Thompson et al. 1997). This software is the most widely used sequence alignment software available. It can be used generate pairwise and multiple alignments of nucleotide and amino acid sequences, and a variety of parameters such as gap penalties and the substitution matrix can be set by the user. It is downloadable for free from <http://bips.u-strasbg.fr/fr/Documentation/ClustalX/>.
- Mega 4 (Tamura et al. 2007). This software can be used to align nucleic acid or amino acid sequences, estimate evolutionary distances using a variety of models, build phylogenetic trees via neighbor joining or maximum parsimony methods, and test phylogenetic tree reliability via interior branch tests or bootstrap analysis. In addition, Mega 4 has extensive tree viewing, manipulation, and editing tools that can be used to create publication quality trees in a variety of file formats. This software is free and can be downloaded from <http://www.megasoftware.net/>.
- MrBayes (Ronquist and Huelsenbeck 2003). This software is used for Bayesian phylogenetic inference. Bayesian inference of phylogeny has become very popular among molecular systematists and is based on the posterior probability distribution of trees using a Markov chain Monte Carlo simulation technique that approximates these posterior probabilities. Although this software is operated through command lines and is not as easy user friendly as other graphical interface programs, excellent documentation is provided with the software, and Hall (2008) provides a good tutorial to help beginning users get started. MrBayes is available for free download from <http://mrbayes.csit.fsu.edu/>.
- Phylogeny.fr: robust phylogenetic analysis for the nonspecialist (Dereeper et al. 2008). This free web service incorporates several alignment and phylogenetic tools into a user friendly website that can be used to reconstruct and analyze phylogenetic relationships between molecular sequences in a single-step or, for more experienced users, an "A la carte" menu can be used to tailor various aspects of the phylogenetic workflow. This site also includes extensive documentation. The site can be accessed at <http://www.phylogeny.fr/>.
- EstimateS: Statistical estimation of species richness and shared species from samples. Version 8.0.0, R. K. Colwell. 2006. This software can be used to calculate a variety of biodiversity functions, estimators, and indexes

based on a range of biological data. For example, EstimateS can be used to compute rarefaction and species accumulation curves, as well as a variety of different species richness estimators for data from marker gene libraries. EstimateS is a free software application that can be downloaded from <http://viceroy.eeb.uconn.edu/estimates>. Excellent supporting documentation for the software is also available at the same Web site.

•Rarefaction Calculator (<http://www2.biology.ualberta.ca/jbrzusto/rarefact.php>), Analytic Rarefaction (<http://www.uga.edu/strata/software/index.html>), and DOTUR (<http://schloss.micro.umass.edu/software/>) (Schloss and Handelsman 2005) are other free software applications that can be used to estimate rarefaction curves for data from marker gene libraries. We have included them because of their simplicity and ease of use.

References

Altshuler, M. L. 2006. PCR troubleshooting: The essential guide. Caister Academic Press.

Atlas, R. M. 1993. Detecting gene sequences using the polymerase chain reaction, p. 267-270. In P. F. Kemp, B. F. Sherr, E. B. Sherr, and J. J. Cole [eds.], Handbook of methods in aquatic microbial ecology. Lewis Publishers.

Ausubel, F. M., and others [eds.]. 2002. Short protocols in molecular biology: a compendium of methods from current protocols in molecular biology, 5th ed. Wiley.

Chen, F., 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.

Dereeper, A., and others. 2008. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res. 36:W465-W469.

Hall, B. G. 2008. Phylogenetic trees made easy: a how-to manual, 3rd ed. Sinauer Associates.

Hall, T. A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids. Symp. Ser. 41:95-98.

Innis, M. A., D. H. Geffland, J. J. Sninsky, and T. J. White. 1990. PCR protocols: A guide to methods and applications. Academic Press.

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. Appl. Environ. Microbiol. 69:4639-4647.

McPherson, M. J., and S. G. Miller. 2006. PCR, 2nd ed. Taylor & Francis.

Ronquist, F., and J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572-1574.

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

Schloss, P. D., and J. Handelsman. 2005. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl. Environ. Microbiol.* 71:1501-1506.

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

Short, S. M., and C. M. Short. 2008. Diversity of algal viruses in various North American freshwater environments. *Aquat. Microb. Ecol.* 51:13-21.

Tamura, K., J. Dudley, M. Nei, and S. Kumar. 2007. MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24:1596-1599.

Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25:4876-4882.

Wilhelm, S. W. and A. R. Matteson. 2008. Freshwater and marine viroplankton: a brief overview of commonalities and differences. *Freshw. Biol.* 53:1076-1089.

Wommack, K. E., T. Sime-Ngando, D. M. Winget, S. Jamindar, and R. R. Helton. 2010. Filtration-based methods for the collection of viral concentrates from large water samples, p. 110-117. In S. W. Wilhelm, M. G. Weinbauer, and C. A. Suttle [eds.], *Manual of Aquatic Viral Ecology*. ASLO.

Yuryev, A. [ed.]. 2007. *PCR primer design*. Humana Press.

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. *Appl. Environ. Microbiol.* 68:1576-1584.

