# REVIEW

# Counts and sequences, observations that continue to change our understanding of viruses in nature[§]

## K. Eric Wommack[*], Daniel J. Nasko, Jessica Chopyk, and Eric G. Sakowski

*Delaware Biotechnology Institute, University of Delaware, Newark Delaware, USA*

**The discovery of abundant viruses in the oceans and on land has ushered in a quarter century of groundbreaking advancements in our understanding of viruses within ecosystems. Two types of observations from environmental samples – direct counts of viral particles and viral metagenomic sequences – have been critical to these discoveries. Accurate direct counts have established ecosystem-scale trends in the impacts of viral infection on microbial host populations and have shown that viral communities within aquatic and soil environments respond to both short term and seasonal environmental change. Direct counts have been critical for estimating viral production rate, a measurement essential to quantifying the implications of viral infection for the biogeochemical cycling of nutrients within ecosystems. While direct counts have defined the magnitude of viral processes; shotgun sequences of environmental viral DNA – virome sequences – have enabled researchers to estimate the diversity and composition of natural viral communities. Virome-enabled studies have found the virioplankton to contain thousands of viral genotypes in communities where the most dominant viral population accounts for a small fraction of total abundance followed by a long tail of diverse populations. Detailed examination of long virome sequences has led to new understanding of genotype-to-phenotype connections within marine viruses and revealed that viruses carry metabolic genes that are important to maintaining cellular energy during viral replication. Increased access to long virome sequences will undoubtedly reveal more genetic secrets of viruses and enable us to build a genomics rulebook for predicting key biological and ecological features of unknown viruses.**

*Keywords:* informational proteins, viral ecology, viromics

[*]For correspondence. E-mail: Wommack@dbi.udel.edu; Tel.: +1-302-831-4362; Fax: +1-302-831-3447

## Introduction

The abundance of free viral particles on Earth is truly staggering. Estimates from direct counts of viruses within natural water samples (Wilhelm and Suttle, 1999; Wommack and Colwell, 2000; Weinbauer, 2004) and soils (Srinivasiah *et al.*, 2008) indicate that the global abundance of free viruses is at least $10^{31}$ individuals (Whitman *et al.*, 1998; Suttle, 2005). This number places viruses within the realm of numbers typically reserved for astrophysics. Assuming an average length (or diameter) of 100 nm, lined up end-to-end the planet's viruses would extend $10^{24}$ meters, a distance equal to that of the sixty nearest galaxies (Suttle, 2005). By comparison, the known universe is $4.3 \times 10^{17}$ seconds old and contains $7 \times 10^{22}$ observable stars, numbers that are dwarfed by the number of viruses on Earth. This discovery of abundant viral particles within aquatic environments a quarter century ago (Proctor *et al.*, 1988; Bergh *et al.*, 1989) and more recently within soil environments (Williamson *et al.*, 2003) revealed our ignorance that viruses and the process of viral infection were critical components of natural ecosystems. This discovery also ignited a cascade of first order questions that researchers are still struggling to answer.

## Viral direct counts, the foundation of process-scale viral ecology

Our ability to address questions surrounding the role of viruses within microbial communities has been driven both by technology and the steady march of new discoveries surrounding viral biology. Addressing questions on the relationship between viral abundance and the physical and chemical factors that partition aquatic environments, such as depth in the ocean or the trophic status of ecosystems, required a means to reliably and accurately enumerate viruses from water samples. Early development of epifluorescence microscopy approaches answered this need enabling direct visualization of viral particles (Hara *et al.*, 1996). These approaches were substantially improved with the introduction of high porosity aluminum oxide filters and DNA-binding fluorescent stains with high quantum yield (Suttle and Fuhrman, 2010). More recently the use of flow cytometry (Brussaard *et al.*, 2010) and even wet-mount (Cunningham *et al.*, 2015) procedures have continued to advance researchers' ability to quantify viruses within environmental samples.

The ability to accurately enumerate viruses led to advance-

ments in understanding the distribution of virioplankton throughout the global ocean. We now know that changes in virioplankton abundance largely follow those occurring for bacterioplankton abundance. Within productive surface waters, virioplankton abundance typically exceeds bacterioplankton abundance by 10-fold; however, this ratio declines with depth and is closer to one in the deep sea (Wommack *et al.*, 2004). In productive estuarine environments, such as the Chesapeake Bay, the virus-to-prokaryote ratio (VPR) varies over an annual cycle with the highest values seen in summer months, indicating seasonal changes in the pressure of viral infection on co-occurring microbial host populations (Winget *et al.*, 2011). Viral abundance data collected from divergent aquatic environments, e.g. Canadian lakes, the Pacific and Arctic Oceans, have curiously shown ecosystem-scale differences in virioplankton abundance and VPR, with lakes and the Arctic demonstrating significantly lower viral abundance and VPR as compared to the Pacific (Clasen *et al.*, 2008). The existence of these large-scale differences hints at the existence of broad differences in the factors driving viral abundance across aquatic ecosystems, and regression analyses indicated, not surprisingly, that bacterial abundance was the strongest predictor of variability in viral abundance. However, chlorophyll a and cyanobacterial abundance were also predictive of viral abundance with a varying degree across these major aquatic biomes.

Ecosystem-scale differences in the interaction of viral communities with co-existing microbial host communities have also become apparent from direct count studies of soil ecosystems. Assessments of viral and prokaryotic cell abundances in soil samples have shown that moist, organic-rich soils demonstrate VPR values of around 10, similar to that seen in aquatic environments (Williamson *et al.*, 2003, 2005). However, in dryer soils with low organic matter content the abundance of viruses exceeds that of prokaryotic cells by more than 1,000 fold with the highest ratios seen in the extreme desert soils of the Antarctic Dry Valleys (Williamson *et al.*, 2007). When comparing moist and dry soils, most of the divergence in the virus to prokaryote ratio was a result of declines in prokaryotic cell abundance. The highest fold difference in viral abundance between dry Antarctic soils and moist temperate soils was ca. 20; whereas, prokaryote abundance showed a maximum difference of 9,600-fold (Srinivasiah *et al.*, 2008). The mechanisms contributing to the dramatically different virus to prokaryote ratios in soils are unknown. However, it is certain that cellular microbial populations respond quickly to the effects of soil drying (Van Gestel *et al.*, 2002); whereas, viral populations (as observed through direct epifluorescence microscopy) appear to be substantially less affected. The occurrence of abundant viral populations within dry soils raises intriguing questions about whether soil viral populations remain infective and capable of controlling co-existing microbial host populations during periods more favorable to microbial growth. A preliminary answer to this question came through microcosm experiments in which soil bacterial growth was stimulated through carbon amendment. Increases in soil bacterial abundance were accompanied by increases in viral abundance and a reduction in the VPR (Srinivasiah *et al.*, 2008, 2015). Accompanying data examining the composition of the soil bacterial community (fatty acid methyl-ester profiling) and viral community (randomly-amplified polymorphic DNA PCR) (Srinivasiah *et al.*, 2013) found that over time changes in the composition of soil bacterial communities were accompanied by changes in the composition of soil viral communities (Srinivasiah *et al.*, 2015). To a first approximation it appears that specific sub-populations of soil viruses respond to alterations in the growth of co-existing soil bacterial populations. However, this experiment could not establish whether the source of newly produced viruses was from new rounds of lytic infection (by virulent phages) or from the induction of temperate viruses existing as prophages within populations of host cells. Soils have been shown to harbor abundant lysogenic bacterial populations (Williamson *et al.*, 2007, 2008); thus, it is quite possible that at least some subset of new viral production within soils comes through prophage induction.

## Linking viruses to carbon and nutrient cycles

Another significant outcome of the development of viral direct counting approaches has been the ability to address questions surrounding the influence of viral infection and lysis on the flow of nutrients and energy through ecosystems. The essential measurement needed to assess viral impacts on biogeochemical processes is the viral production rate. Several approaches have been developed for estimating viral production rate from natural water samples and, with the exception of lesser-used radioisotopic tracer approaches (Steward *et al.*, 1992), all of them require some form of viral direct count data. The most widely used approaches to estimate virioplankton production involve reducing the number of viruses within a water sample through dilution (Wilhelm *et al.*, 2002; Weinbauer *et al.*, 2010) or diafiltration (Helton *et al.*, 2005; Winget *et al.*, 2005) and then monitoring changes in viral abundance within the incubated water sample over time. The reduction in viral abundance at the outset of the experiment prevents new host infections. Thus, any increases in viral numbers are presumed to come from the lysis of host cells that were infected at the beginning of the experimental incubation. Linear regression of viral abundance plotted over time provides an estimate of the viral production rate. From the viral production rate it is possible to estimate the impact of viral lysis on bacterial cell mortality by assuming an average burst size for a viral infection event. Using other conversion factors it is possible to estimate viral release of trace elements and carbon from bacterioplankton (Winget *et al.*, 2011). From viral production estimates we know that the process of viral lysis has a significant impact on bacterioplankton mortality, accounting for the loss of between 20 and 50% of bacterial standing stock each day (Wilhelm and Suttle, 1999; Weinbauer, 2004). We also have learned that virioplankton production levels change throughout the day (Winget and Wommack, 2009) and that the timing between peaks in viral production changes seasonally with shorter peak intervals coinciding with longer summertime day length (Winget *et al.*, 2011). Application of a dilution-incubation approach to measuring viral production within marine sediments (Hewson and Fuhrman, 2003; Dell'Anno *et al.*, 2009)

indicated that viral impacts on sediment bacterial mortality can be substantial, accounting for all bacterial mortality within deep sea sediments and around 20% of mortality in productive coastal marine sediments (Danovaro *et al.*, 2008). Converting viral production estimates to levels of carbon released from lysed cells has indicated that the process of viral lysis can provide a significant input of available carbon for secondary bacterial production. In the Chesapeake Bay, we estimated that carbon released from viral lysis could support around 55% of daily bacterioplankton production (Winget *et al.*, 2011). Lytic release of carbon was estimated to account for 35% of carbon demand for benthic environments in the global ocean (Danovaro *et al.*, 2008). Regrettably, no methods have been developed for estimating viral production in soil environments; however, it is clear that viral lysis is a significant process in the biogeochemical cycling of nutrients within the water column and sediment environments of marine ecosystems.

Beyond its immediate impact on bacterioplankton mortality, the process of viral lysis is significant as it releases new viral particles along with the organelles, macromolecules, and metabolites that comprised the host cell. This transformation of biomass into dissolved organic matter (DOM) by viral lysis is unique from other biomass to DOM transformations (e.g., sloppy feeding by grazers) in that all cellular components are released into the DOM pool (Wommack and Colwell, 2000). Moreover, the composition of viral-released DOM is different from that released by non-infected cells. Elegant work with experimental model systems has shown that the release products of viral lysis are quickly remineralized into available nutrient ions which serve to stimulate the growth of uninfected host cells (Haaber and Middelboe, 2009). Recognizing that viral particles themselves are DOM, recent development of a biophysical scaling model predicted that within the oligotrophic sub-tropical gyres, typical of vast areas of the global ocean, viral particles account for ~5% of available dissolved organic phosphorous (DOP) and under specialized conditions, such as the spring phytoplankon bloom, viral particles may account for as much as 20% of DOP (Jover *et al.*, 2014).

The key issue recognized in this modeling effort was that viruses fundamentally transform the composition of DOM in ways that can have substantial influence on the productivity of marine ecosystems. The most direct illustration of this concept came through examinations of the quality and quantity of iron released through viral lysis. It is estimated that primary production within half of the world's oceans is limited by iron availability (Moore *et al.*, 2002), creating conditions where phytoplankton biomass (as inferred from chlorophyll a concentration) is lower than would be predicted from the available nutrient pools (i.e., high-nutrient, low-chlorophyll, HNLC). Given that the ratio of available Fe between the intracellular and environmental (i.e., the water column) pools is 19,000 to one (Behrenfeld *et al.*, 1996), it is clear that viral lysis could be a potent mechanism for controlling pools of available iron in oceanic ecosystems. Work with experimental model systems and bioreporters showed that the iron released through phage lysis was highly bioavailable and thus capable of supporting phytoplankton and bacterioplankton iron demand (Mioni *et al.*, 2005).

Combining viral production measurements with conversion factors for bacterial cellular iron content and the efficiency of viral-mediated iron release it was discovered that within the HNLC environment of the eastern sub-tropical Pacific as much as 90% of phytoplankton iron demand could be met through viral lysis of bacterioplankton (Poorvin *et al.*, 2004).

## Viral shotgun metagenomics, the window to viral diversity

In hindsight, it is amazing to think that these substantial advancements in understanding the role of viruses within natural ecosystems rested on the development of viral direct counting methods. However, questions surrounding the diversity and composition of viral communities required the development of approaches capable of defining and observing viral populations at high resolution, requirements beyond the capability of viral direct counting. Thus, a new approach that could access genetic information to define and observe viral populations was necessary. Marker gene approaches that utilize polymorphism within a single gene locus (typically the small subunit ribosomal RNA gene, SSU rRNA) to define and observe microbial populations have transformed scientific understanding of microbial communities. However, in the case of viral communities, the polyphyletic origins of viruses had confounded attempts to characterize entire viral communities based on a single gene marker. The polyphyletic evolution of viruses is clearly apparent in the fact that both RNA and DNA in either single or double stranded forms can serve as the genomic material for viruses. Indeed, there are phytoplankton viruses that demonstrate hybrid genomes of ssDNA and dsDNA (Tomaru *et al.*, 2011) and there is evidence of recombination between ssRNA and ssDNA viruses within a single genome of a virus within a hot-spring environment (Diemer and Stedman, 2012). A consequence of the polyphyletic origins of viruses is the impossibility of building a universal phylogenetic framework for viral evolution based on a single gene marker, an approach akin to the tree of cellular life built on the molecular phylogeny of the SSU rRNA gene (Woese and Fox, 1977; Woese *et al.*, 1990). What this has meant for viral ecology research is that approaches to characterizing the diversity of viral communities based on genetic polymorphism within a single gene locus are limited to the sub-sets of viral populations within the larger community carrying the marker gene (Adriaenssens and Cowan, 2014). Despite this limitation, single gene marker approaches continue to be an important tool in viral ecology research as these approaches can provide deep sampling of viral populations using information that is anchored within an evolutionary and biological context. However, informed estimates of the extent of viral diversity and the structure of entire viral communities within natural environments were simply unapproachable until technical advancements in DNA sequencing allowed for cost effective shotgun sampling of genomic DNA from viral communities.

Shotgun viral metagenomes, also known as viromes, provide a random sample of the genetic material of unknown

viral populations within natural samples. By and large, virome studies have focused on dsDNA viruses. Most bacteriophages and all cultivated archeal viruses have dsDNA genomes; and, direct counting has demonstrated that dsDNA viruses are highly abundant within soils and aquatic environments. Viromes of ssDNA and RNA viral populations have also been reported (Culley *et al.*, 2006; Labonté and Suttle, 2013), however, construction of shotgun metagenome libraries from ssDNA or RNA requires that the genomic material be converted to dsDNA as all DNA sequencing technology relies on dsDNA templates for sequencing.

Overlap-consensus assembly of virome sequence libraries has provided a means of estimating the genotypic diversity of viruses within a sample based on the read-scaffolding information for assembled contigs (Angly *et al.*, 2005, 2009). Our earliest virome-enabled views indicated that the genotypic diversity of dsDNA viruses in aquatic ecosystems is large with estuarine (Bench *et al.*, 2007) and oligotrophic oceanic environments (Angly *et al.*, 2006) containing thousands of viral genotypes. Assembly of viromes collected from different aquatic ecosystems provided estimates that global oceanic viral diversity exceeds 50,000 genotypes and could exceed 125,000 genotypes when considering samples collected over time at a single coastal location (Angly *et al.*, 2006). The accuracy of these estimates is difficult to confirm; however, assembly-based analyses from a number of different environments support the idea that viral diversity is high and parallels patterns in the diversity of microbial hosts. Using estimates of the average viral genome length, the average sequence read length within a virome library, and the minimum overlap of reads within the assembly, the Phage Communities from Contig Spectrum (PHACCS) tool (Angly *et al.*, 2005) enables estimation of the richness, evenness, and the percent abundance of the most abundant viral genotype within a viral community. Another tool, Genome relative Abundance and Average Size (GAAS) was later developed to enable more accurate estimation of average viral genome size based on sequence homology to known viruses (Angly *et al.*, 2009). PHACCS also tests the fit of various rank-abundance models to the contig spectral data to provide an estimate of community structure. Aquatic viral communities have typically demonstrated a power-law rank distribution of genotypes with a small number of abundant genotypes accounting for one to three percent of total abundance followed by a long tail of rare genotypes. Rank abundance curves of bacterioplankton populations based on 16S amplicon sequencing have shown a similar tendency towards a power-law distribution (Pedrós-Alió, 2012).

Complementing these α-diversity (within a single sample) estimates have been β-diversity (between sample) estimates based on overlap consensus assembly of viromes from different samples. Cross-contig spectral analysis (MaxiΦ) (Angly *et al.*, 2006) has indicated that viral genotypes can occur over broad geographic distances in the ocean, albeit with changes in their relative rank abundance across environments (Angly *et al.*, 2006). MaxiΦ analyses across an 18-day period in a single environment demonstrated that with increasing time the proportion of shared genotypes fell and the permutation of genotypes (i.e., shifting rank abundance) increased (Rodriguez-Brito *et al.*, 2010). While this result could have

been predicted, the observation that the temporal stability of viral communities differed across ecosystems of differing salinities would not have been obvious without metagenomic data. Moreover, without viral metagenome data it would be impossible to approach estimates of the distribution and rank permutation of genotypes, estimates critical to understanding the population dynamics of viral communities.

The advantage of assembly-based approaches or k-mer based analyses (Pride *et al.*, 2006; Willner *et al.*, 2009) is that these approaches are agnostic to the genetic identity of metagenome sequences when comparing the composition and diversity of viral communities. This agnosticism can be an advantage as most viromes contain a high proportion of putative genes showing no significant homology (usually given as a BLAST expectation score, e-value, of $\leq 10^{-3}$) to a gene from a known organism (Wommack *et al.*, 2012). Typically between 60 to 70% of putative genes within a virome are unknown, a frequency little changed since early viral metagenome reports (Breitbart *et al.*, 2002; Hurwitz and Sullivan, 2013). However, a significant limitation of these sequence-agnostic analyses is that they neither provide insights on the biology of unknown viruses, nor details on the metabolic intricacies of phage-host interactions.

It is the genetic details that can truly advance scientific understanding of the viral processes that have substantial emergent impacts on ecosystems. A notable example of this fact was the discovery that some cyanophages infecting marine *Synechococcus* and *Prochlorococcus* hosts carry genes encoding photosystem proteins (Mann *et al.*, 2003; Lindell *et al.*, 2004; Sullivan *et al.*, 2005). Subsequent work demonstrated that cyanophage photosystem genes are expressed during infection to maintain the activity of the host photosystem (Lindell *et al.*, 2005) and that the host specificity of the cyanophage is connected with the number and type of photosystem genes it carries (Sullivan *et al.*, 2006). Considering the fact that unicellular cyanobacteria are responsible for 25% of global oxygen production (or photosynthesis) (Field *et al.*, 1998; Zhang *et al.*, 2008) and that viral lysis may account for around 5 to 15% of daily mortality of these unicellular cyanobacteria (Suttle and Chan, 1994) it is plausible that 1 to 5% of global oxygen production comes from cyanophage-infected *Synechococcus* and *Prochlorococcus* populations. Thus, perhaps the greatest contribution that metagenomic approaches will offer to scientific understanding of viruses in nature will be the ability to broadly observe the gene content of unknown viruses and make inferences on the biological and ecological features of viruses that influence ecosystem processes.

## Observing the ecology and biology of unknown viruses using informational proteins

Recent work focused on the diversity of informational proteins within virome sequence libraries has demonstrated the promise of metagenomics as a powerful observational tool for understanding unknown viruses. Informational proteins are the enzymes and structural proteins involved in the maintenance of genetic information. Many viruses carry these genes; for example, DNA polymerases, which are critical to

| Host Phyla or sub-phyla (line and text color) | Phage genomes | | Gene presence (ring color) | dsDNA Phage genomes | | Viral Family (circle color) | Phage genomes | |
|---|---|---|---|---|---|---|---|---|
| | Number | Frequency | | Number | Frequency | | Number | Frequency |
| Gammaproteobacteria | 419 | 35.4% | DNA polymerase A | 340 | 25.7% | Siphoviruses | 574 | 48.5% |
| Firmicutes | 285 | 24.1% | Ribonucleotide Reductase | 206 | 17.4% | Myoviruses | 226 | 19.1% |
| Actinobacteria | 278 | 23.5% | | | | Podoviruses | 151 | 12.8% |
| Betaproteobacteria | 43 | 3.6% | | | | Microvirus | 57 | 4.8% |
| Cyanobacteria | 37 | 3.1% | | | | Inovirus | 31 | 2.6% |
| Crenarchaeota | 31 | 2.6% | | | | Levivirus | 15 | 1.3% |
| Alphaproteobacteria | 18 | 1.5% | | | | Lipothrixiviru | 9 | 0.8% |
| Euryarchaeota | 16 | 1.4% | | | | Fusellovirus | 9 | 0.8% |
| Tenericutes | 11 | 0.9% | | | | Tectivirus | 9 | 0.8% |
| Bacteroidetes/Chlorobi | 9 | 0.8% | | | | Plectrovirus | 6 | 0.5% |
| Epsilonproteobacteria | 7 | 0.6% | | | | Rudivirus | 5 | 0.4% |
| Deinococcus-Thermus | 6 | 0.5% | | | | Halovirus | 5 | 0.4% |
| Chlamydiae/Verrucomicrobia | 6 | 0.5% | | | | Cystovirus | 5 | 0.4% |
| Deltaproteobacteria | 4 | 0.3% | | | | Viriophage | 4 | 0.3% |
| dsDNA Virus | 4 | 0.3% | | | | Bicaudavirus | 2 | 0.2% |
| Thaumarchaeota | 1 | 0.1% | | | | Globulovirus | 2 | 0.2% |
| Unknown | 9 | 0.8% | | | | Corticovirus | 1 | 0.1% |
| | | | | | | Plasmavirus | 1 | 0.1% |
| | | | | | | Ampullavirus | 1 | 0.1% |
| | | | | | | Guttavirus | 1 | 0.1% |
| | | | | | | Unknown | 70 | 5.9% |
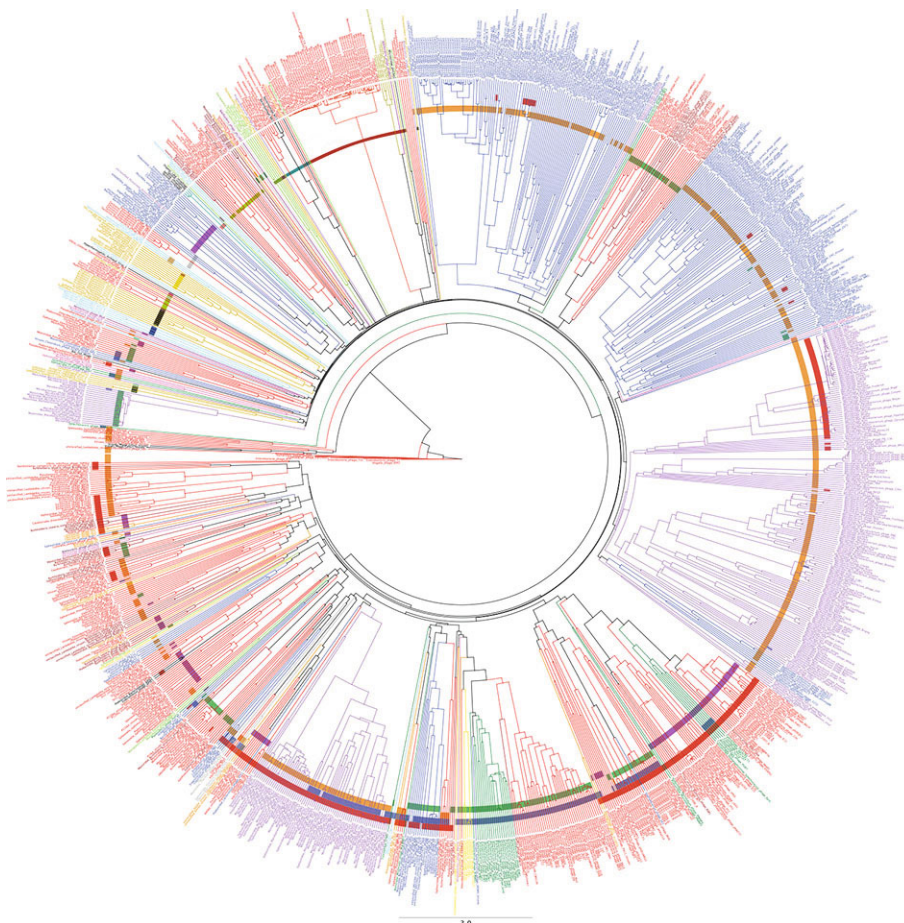


**Fig. 1. Phage proteomic tree 2013.** Names and line colors represent the host phyla of each virus. Moving inwards the six colored rings represent: viruses carrying the DNA *polA* gene; the large subunit RNR gene; siphoviruses; myoviruses; podoviruses; and other viral families. A full sized version of the tree is shown in Fig. S1.

viral replication and can play an outsized role in the evolutionary history and fitness of viruses (Gimenes *et al.*, 2011). One DNA polymerase that has been used for viral ecology research has been Pol I, a family A DNA polymerase encoded by the *polA* gene. Interestingly, Pol I is ubiquitous within bacteria functioning as a proofreading DNA polymerase with both 3′-5′ and 5′-3′ exonuclease activity (Kiefer *et al.*, 1998). Unlike bacteria, within phages, Pol I lacks the 5′-3′ exo domain and functions as the primary polymerase for phage genome replication (Doublié *et al.*, 1998). Because of its high catalytic efficiency and ability to incorporate dideoxy-nucleotides, the coliphage T7 Pol I was developed as the first DNA sequencing enzyme, known as Sequenase™ (Tabor and Richardson, 1989). Thus, the sequencing technology behind the genomics revolution began with, and continues to be based on, phage DNA polymerases such as the Family A and B DNA polymerases from bacteriophages (Chen, 2014).

Before metagenomic datasets became available, researchers relied on known phage genome data and *a priori* approaches to develop genetic markers for assessing the population biology and evolutionary relationships of unknown viruses. Early work using PCR primer sets designed from *polA* genes within T7-like podoviruses found that nearly identical T7-like *polA* genes could be found in widely divergent biomes indicating broad movement of these phages (or at a minimum the *polA* gene) across the biosphere (Breitbart *et al.*, 2004; Breitbart and Rohwer, 2005). Later work, using a different primer set yielding a longer PCR amplicon, sampled a greater diversity of virioplankton *polA* genes with some alleles demonstrating environmental specificity not seen in the earlier study (Labonté *et al.*, 2009). The contrast between these findings illustrates the myopic nature of *a priori* PCR approaches based on known phages. The diversity of viral populations sampled in these studies was ultimately limited by breadth of the initial sequence dataset used to design the PCR primers. Indeed, from a survey of all known phage genome sequences we have found *polA* to be one of the most widely distributed genes among known phages. Among the 1,066 dsDNA phage genomes on the 2013 version of the phage proteomic tree (Rohwer and Edwards, 2002), a genome-based taxonomic system for phage classification, 25.7% carry the *polA* gene (Fig. 1 and S1). The gene is prevalent among tailed phages occurring mostly in podoviruses and siphoviruses, and less commonly in myoviruses. Because *polA* is an evolutionarily ancient protein-coding gene, there exists substantial divergence in the gene at the nucleotide level. Thus, the focus of these prior studies on only T7-like podoviruses was also an outcome of the technical impossibility of designing PCR primers capable of specifically amplifying the *polA* gene from across the divergent groups of phages that carry the gene.

Shotgun metagenomic sequencing of viral communities clearly offers a solution to these limitations in that the gene content of viruses is sampled randomly across viral populations. However, achieving random and quantitative metagenomic sampling of viral populations has been a long time coming. Only recently have we learned that the multiple-displacement amplification of environmental genomic DNA (from viruses or cellular microbes) causes substantial bias in the sampling of populations and should be avoided in all

instances where information on the frequency of populations or gene content is desired (Yilmaz *et al.*, 2010; Marine *et al.*, 2014). Moreover, the short read lengths (100 to 400 bp) of next-generation sequencing technologies have been an impediment to obtaining the full-length gene sequences from viral metagenomes that are necessary for successful primer design and robust phylogenetic reconstructions of viral evolutionary history (Wommack *et al.*, 2008). Using sequence read lengths averaging 650 bp from Sanger sequencing, we were able to selectively assemble nearly 100 full-length *polA* gene sequences from virioplankton metagenomes collected at three sites along the US east coast (Schmidt *et al.*, 2014). In the *polA* discovery pipeline, ORFs were predicted directly from the sequences (Noguchi *et al.*, 2008) and the library of predicted peptides were run through the Viral Informatics Resource for Metagenome Exploration (VIROME, virome. dbi.udel.edu) (Wommack *et al.*, 2012). The reads encoding ORFs with significant homology to known Pol I peptides were then collected from the library and assembled to obtain full-length *polA* gene sequences.

Molecular phylogenetic analysis of the virome *polA* sequences revealed an unexpectedly high diversity of this gene within the virioplankton. Three principal virioplankton Pol I clades were observed that were evolutionarily distant from one another. By comparison, a taxonomically diverse range of bacterial Pol I sequences formed a monophyletic clade on the tree with no overlapping virioplankton or known phage sequences. Digging deeper we discovered mutations in motif B of the enzyme, known to influence the fidelity and efficiency of Pol I (Loh and Loeb, 2005), differentiated the three principle virioplankton Pol I clades. Virioplankton Pol I sequences in clade II all contained a phenylalanine in position 762 (*E. coli* numbering) and were considered the wild type enzyme. All clade I sequences contained a Phe 762 Tyr mutation that is known to increase the efficiency of the enzyme at the cost of fidelity. Coliphage T7 was in this clade and a search of phage genomes indicated that all known phages having a Tyr762 *polA* gene were lytic as were all known phages having the wild type Phe762 *polA* gene. Thus, we predicted that the unknown viruses in clades I and II, carrying either a Phe 762 or a Tyr762 *polA* gene, had a virulent life cycle. Adding support to the idea that a single mutation in Pol I could be predictive of phage lifecycle, we found that most phages having a Phe762Leu mutation had a temperate life cycle. This mutation, found in all clade III virioplankton Pol I sequences, is known to increase the fidelity and lower the efficiency of the polymerase. Surprisingly, 75% of the virioplankton *polA* genes identified in the three viromes had the Phe762Leu mutation, indicating that a majority of the unknown *polA*-carrying phages in these environments could be temperate. This finding was paradoxical as it is a general consensus that virulent phage would predominate within the productive coastal environments sampled in the study (Maurice *et al.*, 2010; Payet and Suttle, 2013). This finding also illuminated the fact that earlier studies, using PCR primers designed from T7-like podoviral *polA* sequences having the Phe762Tyr mutation, missed the majority of virioplankton populations carrying the *polA* gene. These discoveries regarding the relative abundance and diversity of unknown *polA*-carrying virioplankton populations simply

would not have been possible without shotgun metagenomic sequence data.

Replication of dsDNA requires a steady supply deoxyribonucleotides. The ribonucleotide reductase holoenzyme is responsible for catalyzing the reduction of the 2′ hydroxide of ribose within di-phospho-nucleobase substrates to create deoxyribonucleotides for DNA synthesis (Kolberg *et al.*, 2004; Nordlund and Reichard, 2006). Because RNR activity regulates the amount of available substrate, it serves as a master controller of the overall rate of DNA synthesis. As such, RNR is a core housekeeping gene carried by all cells and occurs within the genomes of a diverse range of DNA viruses infecting hosts within all three kingdoms. RNRs have been among the most common genes observed within aquatic viral metagenome libraries (Angly *et al.*, 2006; Bench *et al.*, 2007). On the 2013 phage proteomic tree, RNRs were present within 17.4% of all dsDNA viral genomes on the tree (Fig. 1). Among these known phages, RNR occurs most commonly in Myoviruses many of which are T4-like phages. It also occurs in Siphoviruses, primarily ones infecting the actinobacteria *Myobacteriophage smegmatis*. On the tree, RNR is only rarely seen in Podoviruses, occurring only in marine cyanophages and two N4-like phages of marine α-proteobacteria (*Silicibacter* and *Sulfitobacter*) (Fig. 1 and S1). RNRs appear to only occur within virulent phages indicating the evolutionary pressure for these phages to tightly control rates of DNA synthesis.

The representation of RNRs among phage families and host phyla on the proteomic tree highlights the fact that known phages poorly represent the natural diversity of phages in marine environments. Only 16 of the over 1,000 dsDNA phages on the 2013 tree (Fig. 1 and S1) infect hosts within the α-protobacteria, yet bacteria within this sub-phyla are abundant within the global ocean (Wagner-Döbler and Biebl, 2006). Recent discovery of five phages infecting member strains within the α-protobacterial SAR 11 (Zhao *et al.*, 2013) and SAR 116 (Kang *et al.*, 2013) clusters showed that a surprisingly high proportion of marine virome sequence reads recruited to the genomes of these phages and in many instances relatives of these phages accounted for more than 10% of the dsDNA virioplankton. Our work examining the diversity of marine phages through phylogenetic analysis of the large sub-unit RNR genes from viral metagenome libraries indicated that this gene is very common among marine phages with greater than 90% of lytic marine phages carrying the gene (Sakowski *et al.*, 2014). The most predominant clade of large subunit RNRs (the class II RTPR clade) accounted for between 34 and 50% of all virioplankton RNRs, but contained only a couple of distantly related reference phage sequences, a finding that again indicates the poor representation of marine viruses among cultivated viruses.

Like Pol I, the biochemistry of RNRs can be predictive of the biological features of viruses that carry the gene. RNRs occur in three classes determined by $O_2$ reactivity (Nordlund and Reichard, 2006). Class I RNRs are dependent on $O_2$, whereas, class II RNRs are independent of $O_2$, but dependent on adenosylcobalamin (vitamin B12) as a co-factor. The class III RNRs are $O_2$ sensitive. Thus, the class of the RNR gene carried within a viral genome can be predictive of the physiological conditions under which viral replication

occurs. The ratio of class I to class II RNRs within viromes collected from marine environments across a latitudinal transect demonstrated a predictable decrease with decreasing $O_2$ concentration. Detailed phylogenetic analysis of virioplankton class I and class II RNRs indicated that for cyanophages, RNR class is predictive of phage morphology with class I RNRs (in the Cyano I sub-clade) belonging to cyanomyoviruses and class II RNRs (in the Cyano II subclade) belonging to cyanopodo- and cyanosiphoviruses (Sakowski *et al.*, 2014). The frequency of Cyano I and Cyano II RNRs within viromes showed an inverse temporal relationship in the prevalence of cyanomyoviral and cyanopodo/siphoviral populations over a 24 h period. This intriguing result indicates that Chesapeake Bay cyanophage populations may be infecting different host populations or have different sensitivities to the circadian rhythms of their hosts. Diel pulses in overall viral production have been documented in the past (Winget and Wommack, 2009); however, this work demonstrates the promise of metagenomic approaches for observing the dynamics of interacting viral – host pairs at the population-scale.

## Multi-locus inferences for systems-scale understanding of viral communities

In the dozen years since the first virome report (Breitbart *et al.*, 2002) the use of metagenomic approaches in viral ecology have steadily advanced beyond descriptive studies and towards more quantitative studies of viral populations and communities. Researchers have made advancements in the sample to sequence pipeline that have improved viral recovery from water samples (John *et al.*, 2010) and developed shotgun library construction techniques that circumvent the substantial bias introduced from multiple-displacement amplification of viral gDNA (Duhaime *et al.*, 2012; Duhaime and Sullivan, 2012; Hurwitz *et al.*, 2012; Marine *et al.*, 2014). However, extracting quantitative information on the taxonomic diversity, biological features, and hosts of unknown viruses from fragmentary virome data has proven to be more difficult. Advancements such as the discovery of genotype–phenotype connections for informational genes (Sakowski *et al.*, 2014; Schmidt *et al.*, 2014) and the isolation of phages more closely related to unknown marine viruses (Holmfeldt *et al.*, 2013; Kang *et al.*, 2013; Zhao *et al.*, 2013) will steadily improve our capacity to extract greater biological meaning from virome data. Methodological advancements such as viral-tagging have shown promise for characterizing the genotypic diversity of unknown phage populations capable of infecting (or at least attaching to) a specific cultivated host strain (Deng *et al.*, 2012, 2014). Combining virome sequencing with fluorescence-activated sorting has provided a means to sample populations of large algal viruses that have typically been missed within aquatic viromes because of pre-filtration steps (Martínez *et al.*, 2014). Physical fractionation of virioplankton using CsCl gradient centrifugation followed by anion-exchange chromatography has also provided a means to reduce the complexity of a viral community and link virome sequence data to viruses within a focused range of genome sizes (Brum *et al.*, 2013). However,

the most important technological advancement to advancing scientific understanding of viruses within ecosystems will be access to ever-longer genome sequence fragments from unknown viruses.

For some time we have known that for shotgun metagenome libraries longer sequence reads (i.e., 700 to 1,000 bp in length) perform better in homology searches and thus provide a greater amount of high confidence biological information (Wommack *et al.*, 2008). However, much of the technological advancement in high-throughput sequencing has been in providing a greater quantity of sequence data using individual sequence reads of less than 400 bp. While these advancements have enabled deeper sampling within metagenome libraries (i.e. more sequence reads), the short read lengths (and shear quantity of them) have required that sequences be computationally assembled into longer contiguous sequences (a.k.a., contigs) prior to homology searches. Assembly of metagenome sequence libraries faces additional computational challenges, such as the genotypic diversity of populations within a community and the presence of high similarity regions across populations, over and above that posed in the assembly of a single genome (de Cárcer *et al.*, 2014). Thus, the development and refinement of metagenomic assembly algorithms remains an important and fast-developing area of bioinformatics research.

The recent discovery that virioplankton populations within deep-sea hydrothermal vent plumes possess the genes necessary for sulfur oxidation is a noteworthy illustration of the ability of assembled short-read metagenome sequence libraries to reveal surprising new information on the biological capabilities of unknown viruses (Anantharaman *et al.*, 2014). In this study, contigs were assembled from microbial metagenome libraries collected at six different vent sites. Subsequently, the contigs were binned using an emergent self-organizing map approach (Dick *et al.*, 2009) based on genome signature data (i.e., tetramer frequency, GC%, and contig coverage). In many cases, contigs within bins belonged to particular phylogenetic groups of bacteria already known to be present within vent plume environments. Although, the libraries were constructed from microbial and not viral genomic DNA (i.e., the cellular fraction >0.2 μm), a number of viral contigs were also detected presumably resulting from the presence of infected cells within the original sample. Of particular interest were viral contigs containing complete or nearly complete operons of genes essential to dissimilatory sulfite reduction (DSR). All of these putative viral contigs occurred within a larger bin of contigs belonging to the SUP05 bacterial phyla, a group ubiquitous within hypoxic marine ecosystems and for which no known phages have previously been isolated. Thus, these data, along with phylogenetic evidence linking phage DSR genes and other genes to the SUP05 phyla, represent the first demonstration of phages capable of infecting an important group of chemolithoautotrophs in the global ocean (Anantharaman *et al.*, 2014).

Underlying these discoveries was the ability of long sequences (i.e., assembled contigs) to provide high confidence



**Fig. 2. Use of gene annotation data to predict the biological and ecological features of unknown viruses.** The three annotated contigs were from a deep-sea vent plume metagenomic dataset reported by Anantharaman *et al.* (2014).

information on both the putative function of predicted genes and contextual information linking gene functions to aspects of bacteriophage biology. As we refine our knowledge of the connections between particular genes and specific biological features of viruses it will be increasingly possible to build a picture of the ecology, and lifestyle of unknown viral populations from long virome sequences. Three viral contigs from the deep-sea vent plume study illustrate a multi-locus approach to typing the biological and ecological features of unknown viral populations (Fig. 2). The presence of RNR indicated that each of these unknown phage populations had a lytic lifecycle, a prediction supported by the presence of PolA having either the Phe762 or Tyr762 in motif B of the protein (Schmidt *et al.*, 2014). Two of the populations (Lau_87_Tahi_Moana_scaffold 69 and Lau_130_Kilo_Moana_ scaffold 893) would have been dependent on the presence of O$_2$ for the production of deoxynucleotides by class I RNRs (Sakowski *et al.*, 2014) and would likely have been capable of replicating under conditions of low phosphorous availability due to the presence of PhoH (Goldsmith *et al.*, 2011). The third (Lau_50_Tahi_Moana_scaffold 35), having a class II RNR, would not have required O$_2$ for ribonucleotide reduction, but would have needed to infect a host capable of supplying a B12 co-factor for RNR activity. Separate lines of

evidence predicted the morphological class of each population. The population represented by Lau_87_Tahi_Moana_ scaffold 69 had a siphoviral morphology owing to phylogenetic placement of its terminase gene (Anantharaman *et al.*, 2014) and the portal protein gene. The presence of a phage T4-like major capsid protein and other myoviral structural proteins predicted that Lau_50_Tahi_Moana_scaffold 35 would have been from a population of phages having a myoviral morphology. Moreover, this was a large contig (102.5 kb) and myoviruses tend to have larger genome sizes. Lastly, the phage population represented by Lau_130_Kilo_Moana_ scaffoled 893 likely had a podoviral morphology due to the presence of phage T7-like RNA polymerase and DNA polymerase A genes.

We envision a future where detailed understanding of genotype to phenotype connections in viruses has advanced to a degree sufficient to construct a "Viral Genomics Rulebook" for predicting the ecologically important features of unknown viral populations from long virome sequences (Fig. 3). Some aspects of the rulebook have already come into focus such as the ability of informational proteins to predict the type of viral lifecycle and the environmental conditions (niche) favorable to viral replication. The presence of metabolic genes on long virome contigs such as those encoding
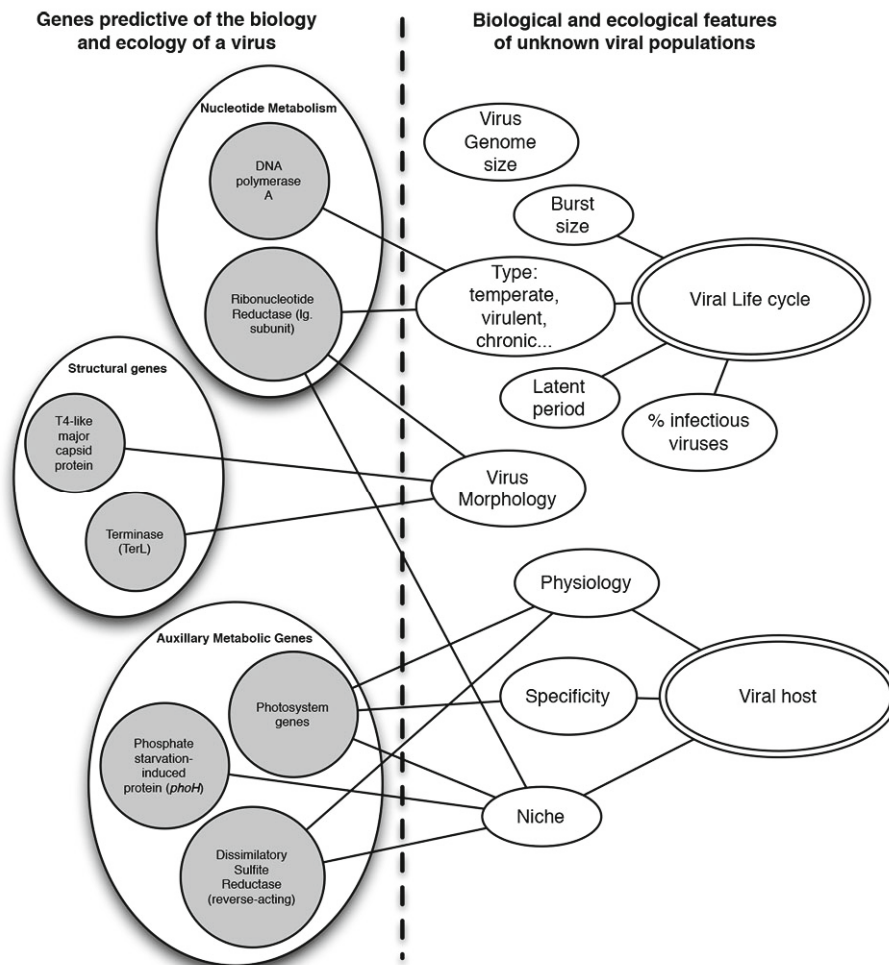


**Fig. 3.** Conceptual diagram of the "Viral Genomics Rulebook", a multi-locus approach to inferring the biological and ecological features of unknown viral populations.

photosystem proteins (Lindell *et al.*, 2005; Sullivan *et al.*, 2006) or proteins involved in dissimilatory sulfite reduction (Anantharaman *et al.*, 2014) can connect unknown viral populations with host groups and a particular environmental niche where the host thrives. However, there are several viral life cycle features where genotypic connections are not known, but would be useful. For instance, the ability to predict the burst size and latent period of an unknown virus based on genetic information from long virome contigs would be immensely helpful in building better models for predicting the ecosystem scale biogeochemical impacts of viral infection (Weitz *et al.*, 2015). Integrating the Viral Genomics Rulebook with existing bioinformatics pipelines for the analysis of virome sequence data (Wommack *et al.*, 2012) would provide a means to compare viral communities using criteria that are anchored within an ecological and evolutionary framework. Creation of just such an informational infrastructure will pave the way for a 'systems-scale' approach to understanding the influence of viruses and viral processes on microbial host communities and the critical ecosystems services these communities maintain.

## Acknowledgements

## References

**Adriaenssens, E.M. and Cowan, D.A.** 2014. Using signature genes as tools to assess environmental viral ecology and diversity. *Appl. Environ. Microbiol.* **80**, 4470–4480.

**Anantharaman, K., Duhaime, M.B., Breier, J.A., Wendt, K.A., Toner, B.M., and Dick, G.J.** 2014. Sulfur oxidation genes in diverse deep-sea viruses. *Science (New York, NY)* **344**, 757–760.

**Angly, F.E., Felts, B., Breitbart, M., Salamon, P., Edwards, R.A., Carlson, C., Chan, A.M., Haynes, M., Kelley, S., Liu, H., *et al.*** 2006. The marine viromes of four oceanic regions. *PLoS Biol.* DOI 10.1371/journal.pbio.0040368.

**Angly, F., Rodriguez-Brito, B., Bangor, D., McNairnie, P., Breitbart, M., Salamon, P., Felts, B., Nulton, J., Mahaffy, J., and Rohwer, F.** 2005. PHACCS, an online tool for estimating the structure and diversity of uncultured viral communities using metagenomic information. *BMC Bioinformatics* **6**, 41.

**Angly, F.E., Willner, D., Prieto-Davó, A., Edwards, R.A., Schmieder, R., Vega-Thurber, R., Antonopoulos, D.A., Barott, K., Cottrell, M.T., Desnues, C., *et al.*** 2009. The GAAS metagenomic tool and its estimations of viral and microbial average genome size in four major biomes. *PLoS Comput. Biol.* **5**, e1000593.

**Behrenfeld, M.J., Bale, A.J., Kolber, Z.S., Aiken, J., and Falkowski, P.G.** 1996. Confirmation of iron limitation of phytoplankton photosynthesis in the equatorial Pacific Ocean. *Nature* **383**, 508–511.

**Bench, S.R., Hanson, T.E., Williamson, K.E., Ghosh, D., Radosovich, M., Wang, K., and Wommack, K.E.** 2007. Metagenomic characterization of Chesapeake Bay virioplankton. *Appl. Environ. Microbiol.* **73**, 7629–7641.

**Bergh, O., Borsheim, K.Y., Bratbak, G., and Heldal, M.** 1989. High abundance of viruses found in aquatic environments. *Nature (London)* **340**, 467–468.

**Breitbart, M., Miyake, J.H., and Rohwer, F.** 2004. Global distribution of nearly identical phage-encoded DNA sequences. *FEMS Microbiol. Lett.* **236**, 249–256.

**Breitbart, M. and Rohwer, F.** 2005. Here a virus, there a virus, everywhere the same virus? *Trends Microbiol.* **13**, 278–284.

**Breitbart, M., Salamon, P., Andresen, B., Mahaffy, J.M., Segall, A.M., Mead, D., Azam, F., and Rohwer, F.** 2002. Genomic analysis of uncultured marine viral communities. *Proc. Natl. Acad. Sci. USA* **99**, 14250–14255.

**Brum, J.R., Culley, A.I., and Steward, G.F.** 2013. Assembly of a marine viral metagenome after physical fractionation. *PLoS One* **8**, e60604.

**Brussaard, C.P.D., Payet, J.P., Winter, C., and Weinbauer, M.G.** 2010. Quantification of aquatic viruses by flow cytometry. In. American Society of Limnology and Oceanography, pp. 102–109.

**Chen, C.Y.** 2014. DNA polymerases drive DNA sequencing-by-synthesis technologies: both past and present. *Front. Microbiol.* **5**, 305.

**Clasen, J.L., Brigden, S.M., Payet, J.P., and Suttle, C.A.** 2008. Evidence that viral abundance across oceans and lakes is driven by different biological factors. *Freshw. Biol.* **53**, 1090–1100.

**Culley, A.I., Lang, A.S., and Suttle, C.A.** 2006. Metagenomic analysis of coastal RNA virus communities. *Science* **312**, 1795–1798.

**Cunningham, B.R., Brum, J.R., Schwenck, S.M., Sullivan, M.B., and John, S.G.** 2015. An inexpensive, accurate and precise wet-mount method for enumerating aquatic viruses. *Appl. Environ. Microbiol.* in revision.

**Danovaro, R., Dell'Anno, A., Corinaldesi, C., Magagnini, M., Noble, R., Tamburini, C., and Weinbauer, M.** 2008. Major viral impact on the functioning of benthic deep-sea ecosystems. *Nature* **454**, 1084–1087.

**de Cárcer, D.A., Angly, F.E., and Alcamí, A.** 2014. Evaluation of viral genome assembly and diversity estimation in deep metagenomes. *BMC Genomics* **15**, 989–989.

**Dell'Anno, A., Corinaldesi, C., Magagnini, M., and Danovaro, R.** 2009. Determination of viral production in aquatic sediments using the dilution-based approach. *Nature Protocols* **4**, 1013–1022.

**Deng, L., Gregory, A., Yilmaz, S., Poulos, B.T., Hugenholtz, P., and Sullivan, M.B.** 2012. Contrasting life strategies of viruses that infect photo- and heterotrophic bacteria, as revealed by viral tagging. *mBio* **3**, pii: e00373-12.

**Deng, L., Ignacio-Espinoza, J.C., Gregory, A.C., Poulos, B.T., Weitz, J.S., Hugenholtz, P., and Sullivan, M.B.** 2014. Viral tagging reveals discrete populations in *Synechococcus* viral genome sequence space. *Nature* **513**, 242–245.

**Dick, G.J., Andersson, A.F., Baker, B.J., Simmons, S.L., Thomas, B.C., Yelton, A.P., and Banfield, J.F.** 2009. Community-wide analysis of microbial genome sequence signatures. *Genome Biol.* **10**, R85.

**Diemer, G.S. and Stedman, K.M.** 2012. A novel virus genome discovered in an extreme environment suggests recombination between unrelated groups of RNA and DNA viruses. *Biol. Direct* **7**, 13.

**Doublié, S., Tabor, S., Long, A.M., Richardson, C.C., and Ellenberger, T.** 1998. Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 A resolution. *Nature* **391**, 251–258.

**Duhaime, M.B., Deng, L., Poulos, B.T., and Sullivan, M.B.** 2012. Towards quantitative metagenomics of wild viruses and other ultra-low concentration DNA samples: a rigorous assessment and optimization of the linker amplification method. *Environ. Microbiol.* **14**, 2526–2537.

**Duhaime, M.B. and Sullivan, M.B.** 2012. Ocean viruses: rigorously evaluating the metagenomic sample-to-sequence pipeline. *Virology* **434**, 181–186.

**Field, C., Behrenfeld, M., Randerson, J., and Falkowski, P.** 1998. Primary production of the biosphere: integrating terrestrial and oceanic components. *Science (New York, NY)* **281**, 237–240.

**Gimenes, M.V., Zanotto, P.M.d.A., Suttle, C.A., da Cunha, H.B., and Mehnert, D.U.** 2011. Phylodynamics and movement of Phycodnaviruses among aquatic environments. *ISME J.* **6**, 237–247.

**Goldsmith, D.B., Crosti, G., Dwivedi, B., McDaniel, L.D., Varsani, A., Suttle, C.A., Weinbauer, M.G., Sandaa, R.A., and Breitbart, M.** 2011. Development of phoH as a novel signature gene for assessing marine phage diversity. *Appl. Environ. Microbiol.* **77**, 7730–7739.

**Haaber, J. and Middelboe, M.** 2009. Viral lysis of *Phaeocystis pouchetii*: implications for algal population dynamics and heterotrophic C, N and P cycling. *ISME J.* **3**, 430–441.

**Hara, S., Koike, I., Terauchi, K., Kamiya, H., and Tanoue, E.** 1996. Abundance of viruses in deep oceanic waters. *Mar. Ecol. Prog. Ser.* **145**, 269–277.

**Helton, R.R., Cottrell, M.T., Kirchman, D.L., and Wommack, K.E.** 2005. Evaluation of incubation-based methods for estimating virioplankton production in estuaries. *Aquat. Microb. Ecol.* **41**, 209–219.

**Hewson, I. and Fuhrman, J.A.** 2003. Viriobenthos production and virioplankton sorptive scavenging by suspended sediment particles in coastal and pelagic waters. *Microb. Ecol.* **46**, 337–347.

**Holmfeldt, K., Solonenko, N., Shah, M., Corrier, K., Riemann, L., Verberkmoes, N.C., and Sullivan, M.B.** 2013. Twelve previously unknown phage genera are ubiquitous in global oceans. *Proc. Natl. Acad. Sci. USA* **110**, 12798–12803.

**Hurwitz, B.L., Deng, L., Poulos, B.T., and Sullivan, M.B.** 2012. Evaluation of methods to concentrate and purify ocean virus communities through comparative, replicated metagenomics. *Environ. Microbiol.* **15**, 1428–1440.

**Hurwitz, B.L. and Sullivan, M.B.** 2013. The Pacific Ocean virome (POV): a marine viral metagenomic dataset and associated protein clusters for quantitative viral ecology. *PLoS One* **8**, e57355.

**John, S., Mendez, C., Deng, L., Poulos, B., Kauffman, A., Kern, S., Brum, J., Polz, M., Boyle, E., and Sullivan, M.** 2010. A simple and efficient method for concentration of ocean viruses by chemical flocculation. *Environ. Microbiol. Rep.* **3**, 195–202.

**Jover, L.F., Effler, T.C., Buchan, A., Wilhelm, S.W., and Weitz, J.S.** 2014. The elemental composition of virus particles: implications for marine biogeochemical cycles. *Nat. Rev. Microbiol.* **12**, 519–528.

**Kang, I., Oh, H.M., Kang, D., and Cho, J.C.** 2013. Genome of a SAR116 bacteriophage shows the prevalence of this phage type in the oceans. *Proc. Natl. Acad. Sci. USA* **110**, 12343–12348.

**Kiefer, J.R., Mao, C., Braman, J.C., and Beese, L.S.** 1998. Visualizing DNA replication in a catalytically active *Bacillus* DNA polymerase crystal. *Nature* **391**, 304–307.

**Kolberg, M., Strand, K.R., Graff, P., and Andersson, K.K.** 2004. Structure, function, and mechanism of ribonucleotide reductases. *Biochim. Biophys. Acta* **1699**, 1–34.

**Labonté, J.M., Reid, K.E., and Suttle, C.A.** 2009. Phylogenetic analysis indicates evolutionary diversity and environmental segregation of marine podovirus DNA polymerase gene sequences. *Appl. Environ. Microbiol.* **75**, 3634–3640.

**Labonté, J.M. and Suttle, C.A.** 2013. Previously unknown and highly divergent ssDNA viruses populate the oceans. *ISME J.* **7**, 2169–2177.

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

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

**Loh, E. and Loeb, L.A.** 2005. Mutability of DNA polymerase I: implications for the creation of mutant DNA polymerases. *DNA Repair* **4**, 1390–1398.

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

**Marine, R., McCarren, C., Vorrasane, V., Nasko, D., Crowgey, E., Polson, S.W., and Wommack, K.E.** 2014. Caught in the middle with multiple displacement amplification: the myth of pooling for avoiding multiple displacement amplification bias in a metagenome. *Microbiome* **2**, 3.

**Martínez, J.M., Swan, B.K., and Wilson, W.H.** 2014. Marine viruses, a genetic reservoir revealed by targeted viromics. *ISME J.* **8**, 1079–1088.

**Maurice, C.F., Mouillot, D., Bettarel, Y., De Wit, R., Sarmento, H., and Bouvier, T.** 2010. Disentangling the relative influence of bacterioplankton phylogeny and metabolism on lysogeny in reservoirs and lagoons. *ISME J.* **5**, 831–842.

**Mioni, C.E., Poorvin, L., and Wilhelm, S.W.** 2005. Virus and siderophore-mediated transfer of available Fe between heterotrophic bacteria: characterization using an Fe-specific bioreporter. *Aquat. Microb. Ecol.* **41**, 233–245.

**Moore, J.K., Doney, S.C., Glover, D.M., and Fung, I.Y.** 2002. Iron cycling and nutrient-limitation patterns in surface waters of the World Ocean. *Deep-Sea Res. Part II-Top. Stud. Oceanogr.* **49**, 463–507.

**Noguchi, H., Taniguchi, T., and Itoh, T.** 2008. MetaGeneAnnotator: detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes. *DNA Res.* **15**, 387–396.

**Nordlund, P. and Reichard, P.** 2006. Ribonucleotide reductases. *Annu. Rev. Biochem.* **75**, 681–706.

**Payet, J.P. and Suttle, C.A.** 2013. To kill or not to kill: The balance between lytic and lysogenic viral infection is driven by trophic status. *Limnol. Oceanogr.* **58**, 465–474.

**Pedrós-Alió, C.** 2012. The rare bacterial biosphere. *Ann. Rev. Mar. Sci.* **4**, 449–466.

**Poorvin, L., Rinta-Kanto, J.M., Hutchins, D.A., and Wilhelm, S.W.** 2004. Viral release of iron and its bioavailability to marine plankton. *Limnol. Oceanogr.* **49**, 1734–1741.

**Pride, D., Wassenaar, T., Ghose, C., and Blaser, M.** 2006. Evidence of host-virus co-evolution in tetranucleotide usage patterns of bacteriophages and eukaryotic viruses. *BMC Genomics* **7**, 8.

**Proctor, L.M., Fuhrman, J.A., and Ledbetter, M.C.** 1988. Marine bacteriophages and bacterial mortality. *Eos* **69**, 1111–1112.

**Rodriguez-Brito, B., Li, L., Wegley, L., Furlan, M., Angly, F., Breitbart, M., Buchanan, J., Desnues, C., Dinsdale, E., Edwards, R., *et al.*** 2010. Viral and microbial community dynamics in four aquatic environments. *ISME J.* **4**, 739–751.

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

**Sakowski, E.G., Munsell, E.V., Hyatt, M., Kress, W., Williamson, S.J., Nasko, D.J., Polson, S.W., and Wommack, K.E.** 2014. Ribonucleotide reductases reveal novel viral diversity and predict biological and ecological features of unknown marine viruses. *Proc. Natl. Acad. Sci. USA* **111**, 15786–15791.

**Schmidt, H.F., Sakowski, E.G., Williamson, S.J., Polson, S.W., and Wommack, K.E.** 2014. Shotgun metagenomics indicates novel family A DNA polymerases predominate within marine virioplankton. *ISME J.* **8**, 103–114.

**Srinivasiah, S., Bhavsar, J., Thapar, K., Liles, M., Schoenfeld, T., and Wommack, K.E.** 2008. Phages across the biosphere: contrasts of viruses in soil and aquatic environments. *Res. Microbiol.* **159**, 349–357.

**Srinivasiah, S., Lovett, J., Ghosh, D., Roy, K., Fuhrmann, J.J., Rado-**

sevich, M., and Wommack, K.E. 2015. Dynamics of autochthonous soil viral communities parallels dynamics of host communities under nutrient stimulation. Submitted.

Srinivasiah, S., Lovett, J., Polson, S., Bhavsar, J., Ghosh, D., Roy, K., Fuhrmann, J.J., Radosevich, M., and Wommack, K.E. 2013. Direct assessment of viral diversity in soils using RAPD-PCR. *Appl. Environ. Microbiol.* **79**, 5450–5457.

Steward, G.F., Wikner, J., Cochlan, W.P., Smith, D.C., and Azam, F. 1992. Estimation of virus production in the sea: I. method development. *Mar. Microb. Food Webs* **6**, 57–78.

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

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

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

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

Suttle, C.A. and Fuhrman, J.A. 2010. MAVE: Enumeration of virus particles in aquatic or sediment samples by epifluorescence microscopy. *In* Wilhelm, S.W., Weinbauer, M.G., and Suttle, C.A. (eds.), pp. 145–153. Manual of Aquatic Viral Ecology. ASLO, Waco, Tex.

Tabor, S. and Richardson, C.C. 1989. Selective inactivation of the exonuclease activity of bacteriophage T7 DNA polymerase by *in vitro* mutagenesis. *J. Biol. Chem.* **264**, 6447–6458.

Tomaru, Y., Takao, Y., Suzuki, H., Nagumo, T., Koike, K., and Nagasaki, K. 2011. Isolation and characterization of a single-stranded DNA virus infecting *Chaetoceros lorenzianus* Grunow. *Appl. Environ. Microbiol.* **77**, 5285–5293.

Van Gestel, M., Merckx, R., and Vlassak, K. 2002. Microbial biomass responses to soil drying and rewetting: The fate of fast- and slow-growing microorganisms in soils from different climates. *Soil Biol. Biochem.* **25**, 109–123.

Wagner-Döbler, I. and Biebl, H. 2006. Environmental biology of the marine *Roseobacter lineage*. *Microbiol.* **60**, 255–280.

Weinbauer, M.G. 2004. Ecology of prokaryotic viruses. *FEMS Microbiol. Rev.* **28**, 127–181.

Weinbauer, M., Rowe, J., and Wilhelm, S. 2010. Determining rates of virus production in aquatic systems by the virus reduction approach, pp. 1–8. Manual of Aquatic Viral Ecology. American Society of Limnology and Oceanography.

Weitz, J.S., Stock, C.A., Wilhelm, S.W., Bourouiba, L., Coleman, M.L., Buchan, A., Follows, M.J., Fuhrman, J.A., Jover, L.F., Lennon, J.T., *et al.* 2015. A multitrophic model to quantify the effects of marine viruses on microbial food webs and ecosystem processes. *ISME J.* doi:10.1038/ismej.2014.220.

Whitman, W.B., Coleman, D.C., and Wiebe, W.J. 1998. Prokaryotes: the unseen majority. *Proc. Natl. Acad. Sci. USA* **95**, 6578–6583.

Wilhelm, S.W., Brigden, S.M., and Suttle, C.A. 2002. A dilution technique for the direct measurement of viral production: A comparison in stratified and tidally mixed coastal waters. *Microb. Ecol.* **43**, 168–173.

Wilhelm, S.W. and Suttle, C.A. 1999. Viruses and nutrient cycles in the sea. *Bioscience* **49**, 781–788.

Williamson, K.E., Radosevich, M., Smith, D.W., and Wommack, K.E. 2007. Incidence of lysogeny within temperate and extreme soil environments. *Environ. Microbiol.* **9**, 2563–2574.

Williamson, K.E., Radosevich, M., and Wommack, K.E. 2005. Abundance and diversity of viruses in six Delaware soils. *Appl. Environ. Microbiol.* **71**, 3119–3125.

Williamson, K.E., Schnitker, J.B., Radosevich, M., Smith, D.W., and Wommack, K.E. 2008. Cultivation-based assessment of lysogeny among soil bacteria. *Microb. Ecol.* **56**, 437–447.

Williamson, K.E., Wommack, K.E., and Radosevich, M. 2003. Sampling natural viral communities from soil for culture-independent analyses. *Appl. Environ. Microbiol.* **69**, 6628–6633.

Willner, D., Thurber, R.V., and Rohwer, F. 2009. Metagenomic signatures of 86 microbial and viral metagenomes. *Environ. Microbiol.* **11**, 1752–1766.

Winget, D.M., Helton, R.R., Williamson, K.E., Bench, S.R., Williamson, S.J., and Wommack, K.E. 2011. Repeating patterns of virioplankton production within an estuarine ecosystem. *Proc. Natl. Acad. Sci. USA* **108**, 11506–11511.

Winget, D.M., Williamson, K.E., Helton, R.R., and Wommack, K.E. 2005. Tangential flow diafiltration: An improved technique for virioplankton production. *Aquat. Microb. Ecol.* **41**, 221–232.

Winget, D.M. and Wommack, K.E. 2009. Diel and daily fluctuations in virioplankton production in coastal ecosystems. *Environ. Microbiol.* **11**, 2904–2914.

Woese, C.R. and Fox, G.E. 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl. Acad. Sci. USA* **74**, 5088–5090.

Woese, C.R., Kandler, O., and Wheelis, M.L. 1990. Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. USA* **87**, 4576–4579.

Wommack, K.E., Bhavsar, J., Polson, S.W., Chen, J., Dumas, M., Srinivasiah, S., Furman, M., Jamindar, S., and Nasko, D.J. 2012. VIROME: a standard operating procedure for analysis of viral metagenome sequences. *Std. Genom. Sci.* **6**, 427–439.

Wommack, K.E., Bhavsar, J., and Ravel, J. 2008. Metagenomics: read length matters. *Appl. Environ. Microbiol.* **74**, 1453–1463.

Wommack, K.E. and Colwell, R.R. 2000. Virioplankton: Viruses in aquatic ecosystems. *Microbiol. Mol. Biol. Rev.* **64**, 69–114.

Wommack, K.E., Williamson, S.J., Sundbergh, A., Helton, R.R., Glazer, B.T., Portune, K., and Cary, S.C. 2004. An instrument for collecting discrete large-volume water samples suitable for ecological studies of microorganisms. *Deep-Sea Res. Part I-Oceanogr. Res. Pap.* **51**, 1781–1792.

Yilmaz, S., Allgaier, M., and Hugenholtz, P. 2010. Multiple displacement amplification compromises quantitative analysis of metagenomes. *Nat. Methods* **7**, 943–944.

Zhang, Y., Jiao, N., and Hong, N. 2008. Comparative study of picoplankton biomass and community structure in different provinces from subarctic to subtropical oceans. *Deep-Sea Res. Part II-Top. Stud. Oceanogr.* **55**, 1605–1614.

Zhao, Y.Y., Temperton, B.B., Thrash, J.C.J., Schwalbach, M.S.M., Vergin, K.L.K., Landry, Z.C.Z., Ellisman, M.M., Deerinck, T.T., Sullivan, M.B.M., and Giovannoni, S.J.S. 2013. Abundant SAR11 viruses in the ocean. *Nature* **494**, 357–360.