INSTITUT PASTEUR

Research in Microbiology 159 (2008) 349–357

www.elsevier.com/locate/resmic

Phages across the biosphere: contrasts of viruses in soil and aquatic environments

Sharath Srinivasiah [a], Jaysheel Bhavsar [a], Kanika Thapar [a], Mark Liles [b], Tom Schoenfeld [c], K. Eric Wommack [a,*]

[a] *University of Delaware, Delaware Biotechnology Institute, 15 Innovation Way Newark, DE 19711, USA*
[b] *Department of Biology, Auburn University, Auburn, AL, USA*
[c] *Lucigen Corp., Middlefield, WI, USA*

Received 18 March 2008; accepted 14 April 2008
Available online 8 May 2008

Abstract

Despite the predominance of aquatic environments on the planet Earth, microbial abundance and diversity within soil environments exceed that of the aquatic realm. Most of what we know of viral ecology within natural systems has come through investigations of aquatic environments. However, the 'aquatic-bias' in viral ecology is beginning to change as the cultivation-independent approaches, which revealed the extraordinary abundance and diversity of viruses within aquatic systems, are now being applied to soils. This review briefly summarizes recent investigations of viral abundance and diversity in soil environments.

Keywords: Viral ecology; Virioplankton; Cultivation-independent approaches; Metagenome

1. Introduction

Double-stranded DNA viruses are, arguably, among the most pervasive biological entities on earth. All cellular life is affected by dsDNA viruses and because the most commonly used direct counting procedures preferentially detect these viruses, the global abundance of dsDNA viruses is estimated to be on the order of 10^{31} individuals [32,37]. The ubiquity and unrivalled abundance of this viral group is rivaled by its genetic diversity. Estimates based on assembly of viral metagenomic sequence data indicate that high density environments, such as marine sediments and productive coastal ecosystems, contain between one million and several hundred thousand viral genotypes [2,5]. Supporting these estimates is the fact that genome size among known and natural populations of dsDNA viruses varies ~250-fold with the largest dsDNA viral genome (Mimivirus, 1.18 Mb) as large as a small bacterial genome [29].

The most comprehensive view of the ecology of natural viral assemblages has come through studies of marine ecosystems. In productive coastal environments, dsDNA viral populations turn over once or twice per day and contribute significantly (20–100%) to the loss of bacterial standing stock [41]. The direct biogeochemical implications of viral lysis as an efficient means for the transfer of carbon (C) from bacterial biomass to the dissolved organic matter pool are easy to conceptualize. In fact, viral lysis may be a critical component for maintaining primary and secondary production through the release of micronutrients, especially Fe [28]. Less apparent are the connections between viral-mediated changes in host community composition and the efficiency of bacterially-driven biogeochemical cycles. Ultimately, it is the combined metabolic capabilities and physiological efficiency of bacterioplankton that drive these cycles in marine ecosystems. In short, increasing evidence is being found indicating that viral assemblages are a critical component within the sustainable

* Corresponding author. Tel.: +1 302 831 4362; fax: +1 302 831 3447.
E-mail address: wommack@dbi.udel.edu (K.E. Wommack).

0923-2508/$ - see front matter © 2008 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.resmic.2008.04.010

functioning of limnetic and marine ecosystems [33]; however, there is almost no data indicating the magnitude of functional contributions of viral assemblages in the soil ecosystems.

Our general lack of appreciation for viral-mediated processes in soil microbial communities highlights gaps in understanding the mechanistic details behind the microbial processes that sustain soil ecosystems. Development of new methodologies for the purification, extraction, and quantification [40] of viral assemblages from soils has opened up new avenues for exploring the ecological role of viruses and viral infection within soil microbial communities. This review will focus on recent advancements in our understanding of soil viral ecology based on cultivation-independent approaches and will contrast this knowledge with that of better studied aquatic ecosystems. For more expansive consideration of numerous autecological investigations of viruses within soil ecosystems, see a recent review by Kimura et al. [25].

2. Viral abundance in soil environments

Accurate inventories of individuals occurring within a given environment are essential to ecological investigations. Indeed, methodological breakthroughs in the accurate enumeration of bacteria [21] and later viruses [4] in marine environments ushered in the field of marine microbial ecology. Recent quantitative investigations of the role of viruses in soil ecosystems have only been possible through the careful precedent work of Williamson and co-workers [40] that yielded methods for the direct extraction and enumeration of viruses within soils. Subsequent application of these methods has revealed that viruses are extraordinarily abundant in a diverse range of soil types and locations, show less overall variability in abundance than co-occurring host bacterial assemblages, and can vastly exceed bacterial abundance depending on environmental conditions. As would be predicted, viruses and bacteria are abundant in moist soils rich in organic matter in comparison to dry and arid soils. For example, in cryptic Antarctic soils, viral abundance ranged between 23 and 64×10^7 viruses per gram dry weight of soils (gdw^{-1}), whereas, in a survey of Delaware soils which included wetlands and agricultural soils, viral abundance was $80-390 \times 10^7$ [38]. Surprisingly, despite the vast differences in composition and geographic location of these soil samples, there was only a 16-fold range in viral abundance. By comparison, across marine environments ranging from the deep sea to freshwater marshes, viral abundance changes over 2000-fold (Table 1). It is possible that with greater scrutiny of soil environments the recorded range in abundance will broaden; however, it is equally likely that current reports reflect the larger trend that viral abundance is a relatively stable parameter across soil environments.

These same investigations by Williamson et al. [3,41] also revealed that co-occurring soil bacterial populations were significantly more variable in abundance, ranging from 0.035 to 330×10^6 cells (gdw^{-1}). Bacterial abundance was highest in temperate moist soils of Delaware; and only these soils showed a clear correlation between viral and bacterial abundance. One plausible explanation of this trend is the close coupling of viral and bacterial production and loss within temperate environments as opposed to a disconnection between these terms in the extremely arid and dry soil environments of Antarctica. In contrast to the nearly 10,000-fold variation in bacterial abundance from Antarctica to Delaware wetland soils, aquatic environments show a narrower range in bacterial abundance of 500–1000-fold. The degree to which changes in cell extraction efficiency from different soil types affects this variability is unknown.

For some time, the ratio of viral to bacterial abundance (i.e., the viral to bacterial ratio, VBR) has been used as a metric for the magnitude of viral influence within a given microbial ecosystem [44]. For most aquatic environments the VBR is around 10 but can vary from parity to ~70 (Table 1). The degree to which VBR actually reflects the level of viral production in aquatic systems is not well established; however, initial studies from soil environments indicate that the VBR strongly reflects ecosystem factors. Across Delaware and Antarctic environments VBR varied over 8000-fold, a range 100 times greater than that seen in aquatic ecosystems [38,39]. Within a variety of ecosystems in Delaware, forested soils tended to have VBRs similar to those of aquatic environments, ca. 10, while intensively managed agricultural soils showed values 100-fold higher [39]. Thus, across this ecosystem gradient VBR was inversely correlated with the soil organic matter (SOM) and water content of the soil sample. The cold desert soils of Antarctica demonstrated the highest reported VBRs (100–1000-fold more viruses than bacteria); yet even within this extreme soil environment the same inverse correlation of VBR to SOM and moisture was apparent [38].

3. Viral activity in soil environments

A parsimonious explanation for (1) the relatively small variation in viral abundance across widely divergent soil ecosystems, and (2) the high ratio of viral to bacterial abundance within SOM-depleted soils is that the loss and potentially the production of viruses within soils is a slow process as compared to aquatic environments. It is possible to envision a scenario in which viral production and loss are relatively invariant across soil ecosystems, while the loss of host cells is highly dynamic and linked to environmental conditions. Such a prediction is well supported by the high abundance of lysogenic bacterial hosts within soil environments [38]. To test these predictions, we conducted a soil microcosm experiment in which virus and host population dynamics were followed for 1 month following the addition of a C- and N-rich substrate. Bacterial response to the addition of yeast extract (5 mg C (gdw^{-1})) was dramatic, and within 3 days bacterial abundance (BA) increased 84-fold ($0.097-8.1 \times 10^8$ cells gdw^{-1}). After this initial period of growth, BA remained high for the rest of the observation period, with a slight upward trend (Fig. 1). During this growth phase, epifluorescence microscopic observations showed a predominance of large rod- or sausage-shaped cophiotrophic bacteria (data not included). The presence of large soil cophiotrophs was not dissimilar to reports from aquatic environments where

Table 1
Viral and bacterial abundance within selected environments as determined by epifluorescence microscopy

Environment	Location	Abundance		VBR[a]	Reference
		Viruses (10^7 ml^{-1} or g^{-1})	Bacteria (10^6 ml^{-1})		
Soils					
Agricultural	Matapeake silt loam, Delaware, USA	110 (SG_L)	0.396 (SG_L)	2750	[38,39]
	Evesboro loamy sand, Delaware, USA	87	0.267	3346	
Forested	Elkton silt loam, Delaware, USA	294	260.3	11	
	Woodstown Loamy sand, Delaware, USA	131	282.8	4.6	
	Piedmont wetland, Delaware, USA	417	337.6	12	
	Piedmont upland, Delaware, USA	148	130.1	11.3	
Antarctica	Eastern rim of Tom pond and Southern rim of Obelisk pond	23–64 (SG_L)	0.035–1.45 (SG_L)	170–8200	[38]
Marine					
Oligotrophic	Mediterranean Sea (West) Coast	1.60 (YP)	(D)	49.9	[17]
	Offshore	1.24		39.7	
	Pacific ocean pelagic water column	0.483 (30 m) (SG_L)	1.84 (30 m) (SG_L)	2.66	[46]
		0.33 (100 m)	0.563 (100 m)	5.85	
		0.161 (500 m)	0.485 (500 m)	3.33	
		0.253 (1500 m)	0.54 (1500 m)	4.65	
	Gulf of Mexico	0.03–0.82 (YP)	0.5–2.3 (D)		[36]
	Atlantic Ocean (north); subtropical	0.14 (SG_R)	0.02–0.29 (SG_R)	9–110	[27]
	Kuroshio area	0.25 (10 m) (D)	0.37 (10 m) (D)		[19]
	North Adriatic Sea	0.1–9.5 (D)			[34]
Mesotrophic and eutrophic					
	Sagami Bay, Japan	1.6 (0 m) (D)	2.9 (0 m) (D)		[19]
	Osaka Bay, Japan	4.0 (0 m) (D)	2.0 (0 m) (D)		
	Otsuchi Bay, Japan	0.28 (0 m)	0.34 (0 m)		
Fresh water					
Oligotrophic	Taylor Valley Lakes, Antarctica	0.42–4 (YP)	(D)	2.8–8.5	[24]
Mesotrophic	Lake Austin, Texas, USA	14.2 (YP)			[20]
	Barton springs spillway, Texas, USA	0.53			
Eutrophic	Lake Plußsee	1.3–4.3 (D)	4.6–7.7 (D)		[35]
	Fresh water marsh	70 (YP)			[20]
	Submerged cyanobacterial mat	96.1 (YP)			[20]
	Paddy field flood waters	0.56–120 (SG_R)	0.69–430 (SG_R)	0.11 to 72	[26]
Extreme					
Alkaline and hypersaline	Mono lake, California, USA	10–1000 (SG_R)	3–44 (SG_R)		[23]
Hot springs	California hot springs, USA	0.007–0.7 (SG_L)	0.059–1.2 (SG_L)	5.2	[8]
Deep-sea vents	Diffuse flow water, E. Pacific Rise	0.0313–0.148 (SG_L)	0.098–0.43 (SG_L)	1.19–3.45	[46]

Epifluorescence stains: YP, YoPro; D, DAPI; SG_R, SYBR Green I; SG_L, SYBR Gold.
[a] Virus to bacteria ratio.

larger and actively growing bacteria are believed to have larger burst sizes [16] and are more susceptible to viral infection [35,47] due to a higher density of cell surface receptors and higher rates of viral-cell contact.

Viral abundance (VA) within the C-amended microcosm showed a similarly dramatic increase, albeit with a time lag as compared to corresponding changes in bacterial abundance. By day 2, VA increased nearly 4-fold from initial levels ($0.5–1.7 \times 10^9$ viruses gdw^{-1}) and did not reach maximal levels until day 7 (3.5×10^9 viruses gdw^{-1}). At later time points (days 14 and 30) VA declined to the level achieved on day 2. The effect of yeast-extract amendment on changes in viral and bacterial abundance was significant as the control microcosm showed only slight and non-significant ($P > 0.05$) changes in both BA and VA across the month-long observation period (Fig. 1). Interestingly, viral and bacterial abundance did fluctuate within the control microcosm over the initial week of sampling, with increases in BA typically accompanied by decreases in VA. It is possible that these slight fluctuations in abundance were an effect of adding water to the control microcosm to ensure that moisture levels were constant across all microcosms. Comparisons of control and C-amended microcosms show substantially higher levels of viral and bacterial abundance within treatment microcosms and dramatic differences in the virus to bacteria ratio. VBR within treatment microcosms declined from an initial value of 50 and, with the exception of day 2, values were below 10. In contrast, VBR was highly variable within control microcosms (range 7–149) with a grand mean of 57. High VBR levels were seen despite the fact that all microcosm soils were maintained at field capacity (i.e., 20.7% moisture by dry weight basis (0.33 bar; pressure plate analyses)). These experiments indicate that VBR varies inversely with the physiological activity

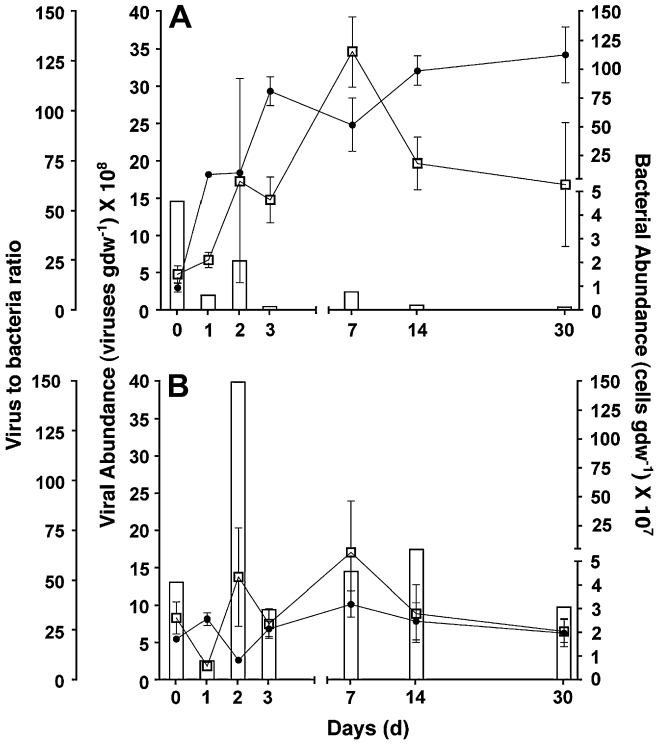

Fig. 1. Temporal variation in abundances of soil viral and bacterial assemblages. (A) Viral (panel A, open squares) and bacterial abundance (panel A, solid circles), and viral to bacterial ratio (panel A, bars) over 30 days in yeast-extract treated soil microcosms (epifluorescence direct counts of extractable bacteria and viruses), error bars represent standard error of mean ($N = 3$); (B) viral (panel B, open squares) and bacterial abundance (panel B, solid circles), and viral to bacterial ratio (panel B, bars) over 30 days in control microcosms; error bars represent standard error of mean ($N = 3$).

of bacteria and the replication activity of viruses in a given soil ecosystem. Overall, this straightforward soil microcosm experiment contradicts the idea that production and loss of autochthonous soil viruses is a slow and unchanging process. On the contrary, soil viral assemblages rapidly respond to changes in host growth and can be as rapidly lost from the system.

4. Cultivation-independent approaches to synecological investigations of viral diversity

Another essential component of ecological investigations is accurate assessment of the composition (viz., species richness and evenness) of biological communities. The well-known plate count anomaly [31] has forced microbial ecologists to rely on cultivation-independent approaches for encompassing estimates of community composition. These approaches have been largely based on DNA sequence polymorphism within the universal small subunit ribosomal RNA (SSU rRNA) gene. For autochthonous viral assemblages, the underrepresentation of viral diversity using cultivation approaches is likely to be more profound than that of prokaryotes, as each host species may have dozens of different viral predators. Thus, viral ecological investigations over the past two decades have largely avoided cultivation as a means for estimating the diversity of viruses within environmental samples.

Within the short history of viral ecology, direct observation of virus particles through transmission electron microscopy has been among the first approaches utilized to characterize viral assemblages. Early investigations that established the high abundance of viruses within aquatic environments utilized TEM observations to substantiate the idea that tailed bacteriophages predominate natural virioplankton assemblages [44]. More recent investigations have shown that the frequency distribution of virions across broad morphological classes can vary significantly between soil environments [39]. These TEM observations also served to substantiate that small fluorescent particles seen within soil extracts by epifluorescence microscopy were indeed viruses [40]. While virion morphology has been critical to taxonomic classification of viruses [10], the small set of distinguishing structural features among viruses and the substantial cost of this analysis mean that TEM-based characterization is inadequate for routine and high resolution examinations of viral assemblages. Within aquatic environments, the shortcomings of TEM analyses have been partly addressed by characterizing the distribution of viral genome sizes within whole virioplankton assemblages. Through pulsed-field gel electrophoresis (PFGE) it is possible to size-separate intact viral genomic DNA from a concentrated sample of virus particles and obtain a range of distinct DNA bands. In essence, the size distribution of DNA bands serves as a proxy measure of the composition of a viral assemblage. PFGE analysis has been important for demonstrating seasonally dynamic changes in virioplankton, e.g., the Chesapeake Bay estuary [45]; however, such PFGE approaches have proven intractable for soil viral assemblages (K.E. Williamson, personal communication).

The large variability in genome size among extant dsDNA viruses means that PFGE potentially has greater resolution for characterization of viral diversity patterns than TEM; however, this analysis falls well short of the resolution needed to discriminate between closely related viral strains or to elucidate evolutionary relationships within viral groups. Through approaches conceptually similar to those for assessing microbial diversity through SSU rRNA sequence, aquatic viral ecologists have utilized sequence data from a few genes that are believed to be broadly distributed among large groups of dsDNA viruses. Well-studied examples are: DNA polymerase of microalgal viruses [9,11] and T7-like phages [5]; g20, the vertex portal protein of bacteriophage T4 [15,30]; and most recently, g23, the major capsid protein of T4 [14,18].

A recurring outcome of aquatic viral diversity investigations using conserved marker genes is that natural viral assemblages contain unique viral groups that are evolutionarily distant from cultivated phages. In the case of T7-like DNA polymerase, unique and highly conserved sequences were broadly dispersed across a wide range of environments [6]. Using real-time PCR, Breitbart and co-workers [6] were able to detect HECTOR T7-like DNA polymerase clade sequences within whole soil DNA extracts, albeit at very low concentrations as compared to aquatic environments.

Rice paddies were recently shown to contain low abundances of unique g23 clades within surface soil and rice

straw samples [22]. Because rice paddy soils can be considered an aquatic environment for much of the growing season, it is difficult to credit the study by Jia and co-workers [22] as the first report of a well-known phage marker gene within a soil viral assemblage. Our attempts to detect g23, g20, and T7-like DNA polymerase sequences through PCR amplification of viral extracts from Delaware soils have been wholly unsuccessful. Control experiments in which T4 virions were added to a soil sample at 10% of total viral abundance and subsequently extracted along with the autochthonous viral assemblage did show positive amplification of g23 via PCR (data not shown). Thus, it is possible that soil viral assemblages are so substantially divergent from aquatic assemblages that the existing suite of PCR primers for amplifying marker genes will not work for soil viruses. The lack of a genetic marker sequence for a subpopulation of commonly occurring soil viruses has prevented more detailed investigations of viral biogeography or population dynamics within soil environments.

The corpus of data on the genotypic richness of viruses containing a shared genetic marker shows that this can be a robust approach for assessing the diversity of specific viral groups within whole viral assemblages. However, the degree to which diversity estimates based on single gene polymorphism within a viral group can be extrapolated to whole viral assemblages is not known. Fortunately, high throughput sequencing of genomic DNA within whole viral assemblages (i.e., viral metagenomics) offers a glimpse into the population structure and genotypic diversity of viruses in nature. The straightforward approach to estimation of viral diversity from shotgun metagenome sequences is to first assemble the metagenome library according to user-defined DNA match and sequence overlap parameters. The outcome of the assembly is then defined by the contig spectra, a string describing the number of assembled contigs containing one sequence (i.e., singletons), two sequences, three, four and so on. The experimental contig spectra along with an average assumed viral genome size, the average read length for the metagenome library, minimum base pair overlap of assembled sequences, an assumed model for the rank-abundance distribution of genotypes and the potential range of expected genotypes are entered into a web-based application known as the phage communities from contig spectrum (PHACCS , http://biome.sdsu.edu/phaccs/) [1]. Based on these inputs, PHACCS mathematically models the viral community to yield a predicted contig spectra and then produces an error estimate for the chosen rank abundance model by comparing the predicted contig spectra to the experimental one. In most cases, the power-law rank abundance distribution has produced the smallest error for experimental contig spectra of viral metagenomes.

Metagenome analyses support an emerging view that natural viral assemblages are comprised of a highly even and extremely rich assortment of viral genotypes, with the abundance of most genotypes being similar and small. Estimates from single time point samples of productive coastal waters have typically been in the thousands [7]. For example, a composite sample of viral assemblages from across the Chesapeake Bay collected in September 2002 was estimated to contain between $\sim$1000 and more than 4000 viral genotypes, with the most abundant genotype comprising less than 0.1% of the total assemblage [3]. The most expansive analysis of viral diversity based on contig spectra found that global oceanic viral diversity was in the range of a few hundred thousand species and that local diversity was nearly as high, indicating a significant influence of migration in these environments [2]. As might be expected from previous knowledge of the high diversity of bacteria in porous media environments [37], viral genotypic diversity within marine sediments and soils is 10–1000-fold higher than aquatic environments. A kilogram of coastal California sediment was estimated to contain $\sim 10^4$ viral genotypes [5] while rainforest soils likely contain more than 10^6 [13]. As might be expected, cross-assemblies of the three viral metagenome libraries from desert, rainforest, and prairie soils showed that soil viral assemblages are locally unique, with almost no overlap in viral genotypes. In contrast, a similar analysis of viral metagenome libraries from four oceanic biomes indicated that any pair of samples could share a majority of viral genotypes, albeit with differing positions along the rank abundance distribution [2]. These initial contrasts of soil and aquatic viral assemblages based purely on metagenome assembly data indicate that soils are likely to harbor the greatest diversity of dsDNA viruses in the biosphere.

Beyond providing a means to more accurately constrain the overall composition and structure of natural viral assemblages, metagenomic analyses have revealed that the majority of dsDNA viral genes are unknown. BLAST homology analyses against subject databases containing sequences from specific organisms (e.g., non-redundant GenBank) typically show that only $\sim$35% of viral metagenome sequences have significant homology ($E < 0.001$) to a 'known' gene (Table 2). Thus, viral assemblages contain a very large pool of unknown and novel genes. Exhaustive translated BLAST analyses against all publicly available sequence databases (GenBank nt/nr; GenBank Environmental nt/nr; and small viral metagenome data sets) indicate that an average of 36% of long read (>600 bp) sequences within libraries of environmental dsDNA have no significant BLAST homolog ($E < 10^{-3}$). No hit frequencies can be substantially higher for shorter length viral metagenome sequences reflecting the distant evolutionary connections between genes within natural viral assemblages and those sequences within large subject databases [43].

The proportion of unknown dsDNA viral sequences, i.e., BLAST homolog to only other environmental sequences, equals or exceeds that of novel sequences. Thus, on average, 2/3 of the genes within dsDNA viruses cannot be assigned a biological function or taxonomic affiliation according to BLAST homology. By comparison, ca. 85% of microbial metagenome sequences have significant BLAST homologs within GenBank nr (Table 2). This pool of novel metagenome sequences represents a much larger collection of novel genes. Genotypic richness estimates for Chesapeake virioplankton

Table 2
Frequency of known, unknown and novel sequences within environmental dsDNA virus sequence libraries from a range of microbial ecosystems

Library name	% G + C (±SD)	Percent of library		
		Known (hit GenBank nt or nr)[a]	Unknown (hit only environmental sequence)[a]	Novel (no BLAST hit)
dsDNA viruses				
Chesapeake Bay	46 (±6.7)	39	30	31
Delaware soil	56 (±7.4)	36	9	55
Wisconsin soil	55 (±6.4)	55	24	21
Octopus Spring	47 (±6.0)	32		
Bear Paw	43 (±8.3)	47	9	44
Hydrothermal vent	44 (±7.4)	26	41	33
Deep-sea	44 (±7.6)	25	44	31
Microbial (random 1000 read sub-samples)				
Sargasso Sea		85	ND	ND
Acid-mine drainage		84	ND	ND

[a] Significant BLAST homolog $E < 0.001$; ND, not determined.

indicated that in September 2002, the size of the Bay viral metagenome was ~205 Mb (4100 genotypes × 50 kb dsDNA per phage genome) (Bench et al. 2007). Assuming a 1 kb average gene length, rough estimates indicate at the time of sampling Bay virioplankton may have contained as many as 60,000 novel genes.

Looking across the metagenome sequence libraries of dsDNA viruses shown in Table 2, ecosystem trends in viral composition are broadly apparent. Viral assemblages within aquatic environments showed lower %G + C content, ca. 45%, as compared to the two soil libraries, ca. 55%, which reflect well-known trends for the major bacterial host groups in these environments (e.g., Cyanobacteria versus Actinobacteria). Marine environments tended to have the highest proportion of sequences with homology to only other environmental sequences (i.e., unknown sequences), while soil environments had some of the lowest frequencies in this category. This trend reflects the fact that environmental sequence databases are dominated by sequences of marine microorganisms. For Delaware soil sequences, the lack of homologs to environmental sequences meant that this library had the highest proportion of novel, i.e., no BLAST hit, sequences.

We have attempted to examine whether trends in environmental specificity within viral assemblages hold for the larger population of genes which show homology to only other unknown environmental sequences. Combined analysis of translated BLAST homolog frequency and alignment quality (as interpreted through median BLAST expectation (E) score) for environmental dsDNA viral sequences against a range of subject databases indicates that, in general, viral genes are distant relatives of known genes within GenBank nr (blue-shifted colors in Fig. 2). Not only did the aquatic libraries have a higher proportion of sequences with homology to environmental sequence databases (Table 2), but these homologs were evolutionarily closer as compared to homologs of GenBank nr sequences (red shifted colors, Fig. 2). In contrast, dsDNA viral sequences from soil were more distant from sequences within env-nr and dsDNA viral sequences from Chesapeake Bay (blue-shifted). Despite its small size, the Wisconsin soil viral metagenome library (1402 sequences) contained a large number of high quality homologs to dsDNA viral metagenome sequences within Delaware soil. With the exception of this library, all other viral metagenome libraries were evolutionarily distant from the Delaware soil library.

Metagenomics offers the most encompassing and potentially the most accurate means of characterizing viral diversity within environmental samples; yet, large improvements in the cost and yield of sequencing technologies notwithstanding, these analyses are still expensive and lack high enough throughput for routine analysis of soil viral assemblages. We recently began work to address the lack of a high-resolution, high-throughput assay for soil viral diversity. Through extensive work with Chesapeake Bay viral assemblages, Winget and Wommack [42] successfully developed a randomly amplified polymorphic DNA-PCR (RAPD-PCR) technique for analysis of virioplankton richness. RAPD-PCR uses a single decamer primer to amplify a collection of DNA fragments. Subsequent electrophoretic separation of RAPD-PCR fragments provides a banding pattern that can be used as a proxy fingerprint for the underlying complexity of the original DNA template. In the case of viral assemblages, the complexity of the RAPD fingerprint can indicate the richness of viral genotypes within a given sample. The approach is termed 'random' as the choice of a decamer primer need not be based on actual sequence information from the target DNA. The fact that successful RAPD-PCR amplification from a concentrate of autochthonous viruses can be obtained without the requirement of prior sequence information is important considering the novel and unknown nature of most viral sequences (Table 2).

RAPD-PCR amplification of Chesapeake Bay viral concentrates shows that banding patterns were reproducible across a 1000-fold change (10^5–10^8) in the amount of virus particles added to a single assay (Fig. 3). The lowest titer reaction (10^5 viruses) produced the weakest banding. For soil assemblages, even lower viral titers (10^5 and 10^4 viruses per reaction) were necessary to obtain a RAPD-PCR banding pattern. Titers above 10^5 soil viruses resulted in a complete loss of banding,

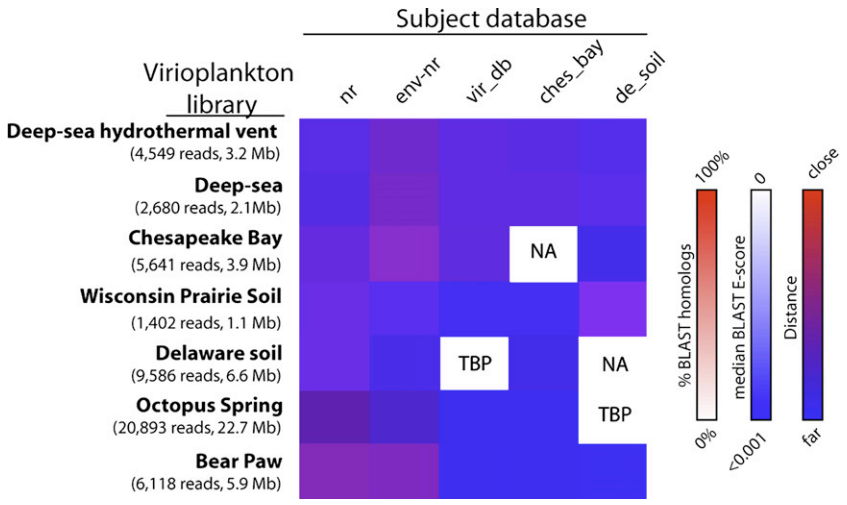

Fig. 2. Heat diagram depicting the frequency and quality of BLAST homologs to dsDNA viral metagenome sequences from eight different environments. Sequences within each library were compared to five subject databases using translated BLAST analysis. Subject databases were: nr, GenBank non-redundant protein; env-nr, GenBank environmental non-redundant protein; vir–db, viral metagenome libraries of California coastal seawater and sediments (see Edwards et al. [12] for review); ches_bay, metagenome sequences from Chesapeake Bay virioplankton [3]; and de_soil, metagenome sequences from viruses within Delaware agricultural soil. Query dsDNA viral metagenome libraries were from the following environments: a composite sample of Chesapeake Bay virioplankton assemblages; deep-sea (2500 m depth); diffuse-flow water from a deep-sea hydrothermal vent; Yellowstone hot springs (Octopus Spring; and Bear Paw); Delaware agricultural soil; and Wisconsin dairy farm soil. Numbers in parentheses are the number of sequences and the total base pairs of DNA within each query library. Percent BLAST homologs is the proportion of each query library with a significant BLAST homolog ($E < 0.001$). Median BLAST E-score is the median BLAST expectation score (E-score) of all the BLAST-positive query sequences. Distance is the combined color of each of these two categories. Blue colors indicate that query sequences within a given library were only distantly related to sequences in the subject library. NA, not applicable; TBP, to be performed.

which was likely due to template dilution in the more genotypically rich soil assemblage. Because of this requirement for lower titers, RAPD-PCR assays of soils required more cycles than Chesapeake assays, 35 versus 30, respectively. Thus, the behavior of natural soil viral assemblages in RAPD-PCR analyses compared to that of virioplankton likely reflects the significantly higher viral richness within soils. Initial indications are that RAPD-PCR may offer a viable means for high-throughput synecological analyses of soil viral assemblages. In this preliminary investigation RAPD-PCR clearly demonstrated genotypic differences between viral assemblages within agricultural soils located in the northern (Newark) and southern (Lewes) regions of Delaware (Fig. 3).

The coming years hold great promise for significant advancement in our understanding of the ecological implications of viral infection within soil microbial communities. This promise stems largely from the oncoming flood of microbial and viral metagenome sequence data from soil environments.

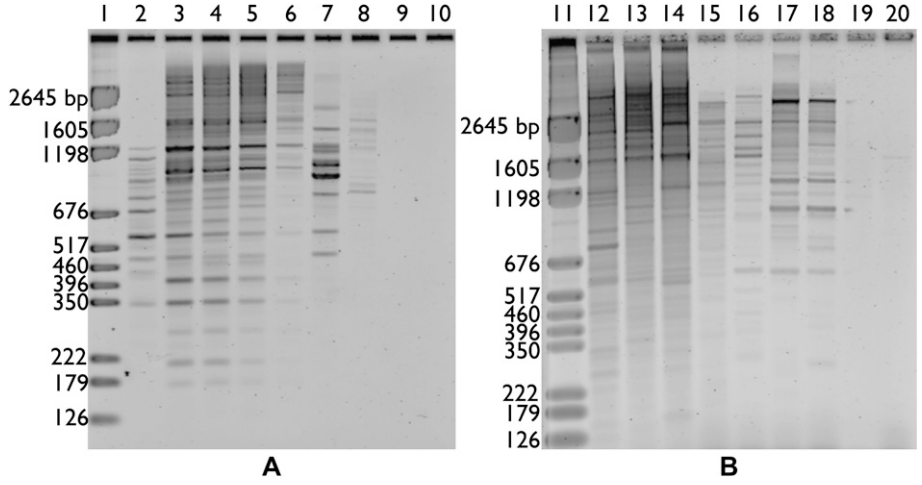

Fig. 3. Agarose gel electrophoresis of aquatic (A) and soil (B) viral assemblages amplified with a single 10 mer primer using randomly amplified polymorphic DNA-PCR (RAPD-PCR). Lanes 1 and 11, molecular weight marker (pGEM, Promega, USA); lanes 2–7, Chesapeake Bay RAPD-PCR amplicons; lanes 3–6, decreasing amounts of viral template 10^8, 10^7, 10^6, and 10^5 viruses per RAPD-PCR, respectively; lanes 15–18, soil RAPD-PCR amplicons; lanes 15 and 16, Newark, Delaware soil viral template 10^5 and 10^4, viruses per RAPD-PCR, respectively; lanes 17 and 18, Lewes, Delaware soil viral template 10^5 and 10^4, viruses per RAPD-PCR, respectively. Lanes 10, 19 and 20, no template controls.

Decreasing sequencing costs and increasing data throughput will soon provide an encompassing view of extant genetic diversity and enable high precision analyses of the diversity, composition, and possibly the activity of viruses within all realms of the biosphere.

Acknowledgements

The authors are grateful for the support of the National Research Initiative of the USDA Cooperative State Research, Education and Extension Service grant 2005-35107-15214 to K.E. Wommack and the support from Auburn University's Department of Biological Sciences to M.R. Liles. The authors also acknowledge David Mead and Melodee Patterson of Lucigen Corp for their assistance with the construction and analysis of the Yellowstone hot springs viral libraries.

References

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

[2] Angly, F.E., Felts, B., Breitbart, M., Salamon, P., Edwards, R.A., Carlson, C., Chan, A.M., Haynes, M., et al. (2006) The marine viromes of four oceanic regions. PLoS Biol. 4, 2121–2131.

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

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

[5] Breitbart, M., Felts, B., Kelley, S., Mahaffy, J.M., Nulton, J., Salamon, P., Rohwer, F. (2004) Diversity and population structure of a near-shore marine-sediment viral community. Proc. R. Soc. Lond., B, Biol. Sci. 271, 565–574.

[6] Breitbart, M., Miyake, J.H., Rohwer, F. (2004) Global distribution of nearly identical phage-encoded DNA sequences. FEMS Microbiol. Lett. 236, 249–256.

[7] Breitbart, M., Salamon, P., Andresen, B., Mahaffy, J.M., Segall, A.M., Mead, D., Azam, F., Rohwer, F. (2002) Genomic analysis of uncultured marine viral communities. Proc. Natl. Acad. Sci. U.S.A. 99, 14250–14255.

[8] Breitbart, M., Wegley, L., Leeds, S., Schoenfeld, T., Rohwer, F. (2004) Phage community dynamics in hot springs. Appl. Environ. Microbiol. 70, 1633–1640.

[9] Brussaard, C.P.D. (2004) Viral control of phytoplankton populations – a review. J. Eukaryot. Microbiol. 51, 125–138.

[10] Buchen-Osmond, C. (2003) The universal virus database ICTVdB. Computing in Science & Engineering 5, 16–25.

[11] Chen, F., Suttle, C.A., Short, S.M. (1996) Genetic diversity in marine algal virus communities as revealed by sequence analysis of DNA polymerase genes. Appl. Environ. Microbiol. 62, 2869–2874.

[12] Edwards, R.A., Rodriguez-Brito, B., Wegley, L., Haynes, M., Breitbart, M., Peterson, D.M., Saar, M.O., Alexander, S., et al. (2006) Using pyrosequencing to shed light on deep mine microbial ecology. BMC Genomics 7, 57.

[13] Fierer, N., Breitbart, M., Nulton, J., Salamon, P., Lozupone, C., Jones, R., Robeson, M., Edwards, R.A., et al. (2007) Metagenomic and small-subunit rRNA analyses reveal the genetic diversity of bacteria, archaea, fungi, and viruses in soil. Appl. Environ. Microbiol. 73, 7059–7066.

[14] Filee, J., Tetart, F., Suttle, C.A., Krisch, H.M. (2005) Marine T4-type bacteriophages, a ubiquitous component of the dark matter of the biosphere. Proc. Natl. Acad. Sci. U.S.A. 102, 12471–12476.

[15] Fuller, N.J., Wilson, W.H., Joint, I.R., Mann, N.H. (1998) Occurrence of a sequence in marine cyanophages similar to that of T4 g20 and its application to PCR-based detection and quantification techniques. Appl. Environ. Microbiol. 64, 2051–2060.

[16] Gons, H.J., Hoogveld, H.L., Simis, S.G.H., Tijdens, M. (2006) Dynamic modelling of viral impact on cyanobacterial populations in shallow lakes: implications of burst size. J. Mar. Biolog. Assoc. U.K. 86, 537–542.

[17] Guixa-Boixereu, N., Vaque, D., Gasol, J.M., Pedros-Alio, C. (1999) Distribution of viruses and their potential effect on bacterioplankton in an oligotrophic marine system. Aquat. Microb. Ecol. 19, 205–213.

[18] Hambly, E., Tetart, F., Desplats, C., Wilson, W.H., Krisch, H.M., Mann, N.H. (2001) A conserved genetic module that encodes the major virion components in both the coliphage T4 and the marine cyanophage S-PM2. Proc. Natl. Acad. Sci. U.S.A. 98, 11411–11416.

[19] Hara, S., Terauchi, K., Koike, I. (1991) Abundance of viruses in marine waters – assessment by epifluorescence and transmission electron-microscopy. Appl. Environ. Microbiol. 57, 2731–2734.

[20] Hennes, K.P., Suttle, C.A. (1995) Direct counts of viruses in natural-waters and laboratory cultures by epifluorescence microscopy. Limnol. Oceanogr. 40, 1050–1055.

[21] Hobbie, J.E., Daley, R.J., Jasper, S. (1977) Use of nucleopore filters for counting bacteria by fluorescence microscopy. Appl. Environ. Microbiol. 33, 1225–1228.

[22] Jia, Z.J., Ishihara, R., Nakajima, Y., Asakawa, S., Kimura, M. (2007) Molecular characterization of T4-type bacteriophages in a rice field. Environ. Microbiol. 9, 1091–1096.

[23] Jiang, S., Steward, G., Jellison, R., Chu, W., Choi, S. (2004) Abundance, distribution, and diversity of viruses in alkaline, hypersaline Mono Lake, California. Microb. Ecol. 47, 9–17.

[24] Kepner, R.L., Wharton, R.A., Suttle, C.A. (1998) Viruses in Antarctic lakes. Limnology Oceanography 43, 1754–1761.

[25] Kimura, M., Jia, Z.J., Nakayama, N., Asakawa, S. (2008) Ecology of viruses in soils: Past, present and future perspectives. Soil Sci. Plant Nutr. 54, 1–32.

[26] Nakayama, N., Okumura, M., Inoue, K., Asakawa, S., Kimura, M. (2007) Seasonal variations in the abundance of virus-like particles and bacteria in the floodwater of a Japanese paddy field. Soil Sci. Plant Nutr. 53, 420–429.

[27] Parada, V., Sintes, E., van Aken, H.M., Weinbauer, M.G., Herndl, G.J. (2007) Viral abundance, decay, and diversity in the meso- and bathypelagic waters of the North Atlantic. Appl. Environ. Microbiol. 73, 4429–4438.

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

[29] Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., La Scola, B., Suzan, M., Claverie, J.M. (2004) The 1.2-megabase genome sequence of Mimivirus. Science 306, 1344–1350.

[30] Short, C.M., Suttle, C.A. (2005) Nearly identical bacteriophage structural gene sequences are widely distributed in both marine and freshwater environments. Appl. Environ. Microbiol. 71, 480–486.

[31] Staley, J.T., Konopka, A. (1985) Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. Annu. Rev. Microbiol. 39, 321–346.

[32] Suttle, C. (2005) Crystal ball. The viriosphere: the greatest biological diversity on Earth and driver of global processes. Environ. Microbiol. 7, 481–482.

[33] Suttle, C.A. (2007) Marine viruses – major players in the global ecosystem. Nat. Rev. Microbiol. 5, 801–812.

[34] Weinbauer, M.G., Fuks, D., Puskaric, S., Peduzzi, P. (1995) Diel, seasonal, and depth-related variability of viruses and dissolved DNA in the northern Adriatic Sea. Microb. Ecol. 30, 25–41.

[35] Weinbauer, M.G., Hofle, M.G. (1998) Size-specific mortality of lake bacterioplankton by natural virus communities. Aquat. Microb. Ecol. 15, 103–113.
[36] Weinbauer, M.G., Suttle, C.A. (1997) Comparison of epifluorescence and transmission electron microscopy for counting viruses in natural marine waters. Aquat. Microb. Ecol. 13, 225–232.
[37] Whitman, W.B., Coleman, D.C., Wiebe, W.J. (1998) Prokaryotes: the unseen majority. Proc. Natl. Acad. Sci. U.S.A. 95, 6578–6583.
[38] Williamson, K.E., Radosevich, M., Smith, D.W., Wommack, K.E. (2007) Incidence of lysogeny within temperate and extreme soil environments. Environ. Microbiol. 9, 2563–2574.
[39] Williamson, K.E., Radosevich, M., Wommack, K.E. (2005) Abundance and diversity of viruses in six Delaware soils. Appl. Environ. Microbiol. 71, 3119–3125.
[40] Williamson, K.E., Wommack, K.E., Radosevich, M. (2003) Sampling natural viral communities from soil for culture-independent analyses. Appl. Environ. Microbiol. 69, 6628–6633.
[41] Winget, D.M., Williamson, K.E., Helton, R.R., Wommack, K.E. (2005) Tangential flow diafiltration: an improved technique for estimation of virioplankton production. Aquat. Microb. Ecol. 41, 221–232.
[42] Winget, D.M., Wommack, K.E. (2008) Randomly-amplified polymorphic DNA (RAPD) PCR as a tool for assessment of marine viral richness. Appl. Environ. Microbiol. 74, 2612–2618.
[43] Wommack, K.E., Bhavsar, J., Ravel, J. (2008) Metagenomics: read length matters. Appl. Environ. Microbiol. 74, 1453–1463.
[44] Wommack, K.E., Hill, R.T., Kessel, M., Russekcohen, E., Colwell, R.R. (1992) Distribution of viruses in the Chesapeake Bay. Appl. Environ. Microbiol. 58, 2965–2970.
[45] Wommack, K.E., Ravel, J., Hill, R.T., Chun, J.S., Colwell, R.R. (1999) Population dynamics of Chesapeake Bay virioplankton: total-community analysis by pulsed-field gel electrophoresis. Appl. Environ. Microbiol. 65, 231–240.
[46] Wommack, K.E., Williamson, S.J., Sundbergh, A., Helton, R.R., Glazer, B.T., Portune, K., 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.
[47] Yager, P.L., Connelly, T.L., Mortazavi, B., Wommack, K.E., Bano, N., Bauer, J.E., Opsahl, S., Hollibaugh, J.T. (2001) Dynamic bacterial and viral response to an algal bloom at subzero temperatures. Limnol. Oceanogr. 46, 790–801.