Seasonal dynamics and metagenomic characterization of estuarine viriobenthos assemblages by Randomly Amplified Polymorphic DNA PCR (RAPD-PCR).

Rebekah R. Helton \(^1\) and K. Eric Wommack \(^{1,2*}\)

\(^1\)Department of Plant and Soil Sciences, Delaware Biotechnology Institute, University of Delaware, Newark, DE 19711 and \(^2\)College of Marine Studies, University of Delaware, Lewes, DE 19958

\(^*\)Corresponding author. Mailing address: Delaware Biotechnology Institute, 15 Innovation Way, Room 148, Newark, DE 19711. Phone: 302-831-4362. Fax: 302-831-3447. E-mail: wommack@dbi.udel.edu

Running title: Viriobenthos diversity
ABSTRACT

Direct enumeration and genetic analyses indicate that aquatic sediments harbor abundant and diverse viral communities. Thus far, synecological analysis of estuarine sediment viral diversity over an annual cycle has not been reported. This oversight is due in large part to a lack of molecular genetic approaches for assessing viral diversity within a large collection of environmental samples. Here, randomly amplified polymorphic DNA – polymerase chain reaction (RAPD-PCR) was used to examine viral genotypic diversity within Chesapeake Bay sediments. Using a single 10-mer oligonucleotide primer for all samples, RAPD-PCR analysis of sediment viral assemblages yielded unique banding patterns across spatial and temporal scales, with the occurrence of specific bands varying among the sample set. Cluster analysis of RAPD-PCR amplicon banding patterns indicated that sediment viral assemblages changed with season and to a lesser extent with geographic location. Sequence analysis of RAPD-PCR amplicons revealed that 76% of sediment viral sequences were not homologous to any sequence in the GenBank non-redundant protein database. Of the GenBank sequence homologs, the majority belonged to viruses within the Podoviridae (24%) and Myoviridae (22%) viral families, which agrees with the previously observed frequencies of these morphological families in Chesapeake Bay sediments. Furthermore, the majority of the sediment viral sequences homologous to GenBank non-redundant protein sequences were phage or prophage (57%). Hence, RAPD-PCR proved to be a reliable and useful approach for characterization of viral assemblages and the genetic diversity of viruses within aquatic sediments.
INTRODUCTION

Large numbers of viruses, an estimated abundance greater than $10^{31}$ viruses worldwide (11, 26), have been found in a variety of environments including marine (39) and freshwaters (19), sediments (25, 28) and soils (34). Viruses are not only abundant, but are also likely to significantly influence the population dynamics and genotypic composition of their bacterial host populations (29, 33). Process-level investigations of viral activity in sediments have shown that viruses are an active component of sediment microbial communities (23). Glud and Middelboe (23) found that bacterial growth rates and viral production increased in parallel with respiration, suggesting that viruses are active members of benthic microbial communities. Previous studies have shown that sediment viral abundance exceed co-existing bacterial abundance by 10 to 1,000 fold (15, 17, 25), creating the potential for viral processes to influence the microbial ecology of aquatic sediments. However, with the exception of small-scale metagenomic investigations (4, 8), there exists little information on the genetic content of viriobenthos assemblages or how the composition of these assemblages changes over ecological gradients.

Despite the high abundances of viruses in nature, the lack of a shared genetic marker creates a difficult problem when attempting to examine viral genetic diversity in environmental samples (31). Gene g20 encodes for a multifunctional protein within the collar between the capsid and tail in T4-like bacteriophage and has been of significant importance in examining the genetic diversity of cyanomyoviruses (22, 24, 32). As well, others have been able to evaluate the diversity of unidentified aquatic picorna-like viruses using the RNA-dependent RNA polymerase (RdRp) gene (13). Other studies have attempted to examine phage genetic diversity based on the DNA polymerase gene (6, 21). Unfortunately, not all known phage contain these specific genes,
hence their use as universal markers is markedly inadequate. Thus, molecular methods that do not rely on polymorphism analysis of a single gene product must be used to circumvent these limitations.  

Recently, metagenomic approaches (i.e., sequencing of random genomic DNA fragments from whole microbial assemblages) have been used to examine genetic diversity within viral (18) or prokaryotic (10) assemblages. For sediment environments, metagenomic analysis has revealed that the virobenthos is perhaps the most diverse of all viral assemblages and has been estimated to contain more than 10,000 genotypes per kg of sediment (4). Viral assemblages within a wide range of environments including marine (2, 8) and estuarine (3) waters, soils (20), stromatolites (16) and equine (9) and human feces (5, 41) have been examined. Overall, these studies have shown that a relatively low proportion (~30%) of viral metagenome sequences are similar to sequences found in the non-redundant GenBank database, but the probability of detecting significant BLAST homologs increases two-fold when including queries against other viral metagenome sequence libraries (3). Thus, the function of most viral genes is currently unknown; however, these genes are broadly distributed among viruses.  

While large-scale metagenomics offers unprecedented resolution of the diversity and composition of a viral assemblage, the significant costs and computational requirements preclude routine application in a large collection of environmental samples. Recently, Winget and Wommack (36) introduced a new, low-cost, high-throughput means for genetic analysis of viral diversity utilizing random amplified polymorphic DNA – polymerase chain reaction (RAPD-PCR). In this approach, a single 10 bp oligonucleotide serves as both the forward and reverse primer in a single thermocycler reaction. Target sequences in the template DNA are randomly selected, thus, development of a RAPD-PCR assay requires no prior information on the DNA
coding content within the sample or organism – a significant advantage considering the largely
unknown nature of most viral genes.

In this study, we assess the potential of RAPD-PCR as a tool to examine genotype-scale
compositional changes in the Chesapeake Bay viriobenthos and to explore the genetic diversity
of viruses within Chesapeake Bay sediments. To our knowledge, this is the first study to use
RAPD-PCR for the evaluation of sediment viral diversity and to document compositional
changes in viriobenthos assemblages over time and geographic location.

MATERIALS AND METHODS

Sampling sites and collection. Sediment samples were collected from the Chesapeake
Bay during nine cruises from April 2003 to October 2005 using a four-tube multi-corer device
(MC-400 Hedrick/Marrs, Ocean Instruments). Sediment samples were collected from surface
depths of 15 m, 25 m and 9 m for station 724 (37° 24’ N, 76° 05’ W), station 804 (38° 04’ N, 76°
13’ W) and station 908 (39° 08’ N, 76° 20’ W), respectively. Each 10 cm diameter core was sub-
sampled once with a sterile 60 ml cut-off syringe and the top 2 cm processed immediately
onboard.

Viral extractions. Sediments were processed for removal of viruses as in Helton et al.
(25). Briefly, 2 ml of surface sediment was placed inside sterile 50 ml centrifuge tubes to which
8 ml of 10 mM disodium pyrophosphate and 5 mM EDTA was added. Samples were vortexed at
high speed horizontally for 20 min. Larger particles were removed by centrifugation at 2,000 X
g for 25 min post-agitation. The supernatant was filtered through a 0.45 µm filter (Sterivex,
Millipore, Corp.). The filtrate was then passed through a 0.22 µm filter (Sterivex) to remove any
remaining particles and bacteria. Viral particles in the 0.22 µm filtrate (ca. 8 ml) were
concentrated using Centricon-YM30 filters (30,000 MWCO, Millipore), triple rinsed with sterile
TE (Tris HCl 100 mM, EDTA 10 mM) and filtered again with a 0.22 µm filter (Sterivex) prior to
storage at −20 °C. Several viral concentrate samples were treated with DNase alone, as well as
samples treated with heat + DNase as in Helton et al. (25) to test for the presence or effects of
any free DNA remaining in the concentrated filtrates. Treated samples from both DNase and the
heat + DNase treatments were used as template for testing RAPD-PCR amplification of viral
DNA.

**RAPD-PCR.** Primer OPA-9 [5’ – GGGTAACGCC – 3’] was used in all PCR reactions
of 25 µl contained 2.0 mM MgCl₂ (included in 10x buffer), 0.8 mM each deoxyribonucleoside
triphosphate (TaKaRa Bio Inc.), 4 µM primer and 2.5 U of TaKaRa ExTaq HotStart Version
(TaKaRa Bio Inc). Template concentrations were standardized by adding one microliter of a
sediment viral extract containing ca. 7 x 10⁷ virus particles to each RAPD-PCR assay. Reactions
were carried out in a MJ Research PTC-200 thermocycler using the following parameters: initial
denaturation for 10 min at 94 °C; 33 cycles of 3 min at 35 °C, 1 min at 72 °C, and 30 s at 94 °C;
with a final extension of 10 min at 72 °C. All RAPD-PCR products were visualized by gel
electrophoresis on a 1.8% Metaphor (FMC BioProducts) agarose gel in 0.5 x TBE buffer. Gels
were stained in a 1X SYBR Gold (Invitrogen) bath for 30 min and imaged with a Typhoon 8600
variable mode imager (Molecular Dynamics, Amersham Pharmacia Biotech) under 560 BP
30/Green 532 nm, focal plane +3 mm. Resulting band patterns were analyzed using GelCompar
II (v. 4.50, Applied Maths). The similarity of RAPD-PCR banding patterns was determined
using Jaccard’s coefficient and a dendogram depicting banding pattern similarity was generated
using the unweighted pair group method of averages (UPGMA) algorithm.
**Cloning and Sequencing.** Resulting RAPD-PCR amplicons from all three locations (908, 804 and 724) and from three seasons (autumn 2003, spring 2004 and autumn 2004) were used for genetic analyses. In addition, a recurring band of ~625 bp (in 12 of 20 samples) was selected from several samples, excised from the agarose gel and purified with the QIAquick Gel Purification Kit (Qiagen). Entire collections of RAPD-PCR amplicons were purified with the QIAquick PCR Purification Kit (Qiagen). Both purified products (gel excision and whole reactions) were cloned using the pCR8/GW/TOPO TA Cloning Kit (Invitrogen) as per the manufacturer's protocol. This vector was chosen as it includes terminator sequences on either side of the multiple cloning site to prevent the transcription of insert DNA. Cloned products were chemically transformed into One Shot Mach1-T1 (Invitrogen) chemically competent *Escherichia coli* cells as per manufacturer's instructions. Clones with plasmid DNA containing RAPD-PCR inserts were purified using the Qiagen DirectPrep 96 MiniPrep Kit (Qiagen).

A total of 518 clones were sequenced using an ABI Prism 3130XL Genetic Analyzer (Applied Biosystems). From that, 448 sequences were obtained and edited for contaminating vector and primer sequence using Sequencher 4.6 (Gene Codes Corp.) software. Sequence results for the ~625 bp band were aligned and compared using Sequencher 4.6 (Gene Codes Corp.) software. All sediment viral sequences were subjected to several BLAST analyses (1). A translated query versus protein (BLASTx) was used for homology searches against the GenBank non-redundant (nr) and Environmental non-redundant (env-nr) databases. A translated query versus the translated database (tBLASTx) was used for homology searches against the GenBank nucleotide (nt) and Environmental nucleotide (env-nt) databases as well as for three viral metagenomes: (i) Chesapeake Bay Vibrioplankton (CBV) (3), (ii) Delaware Soil Viruses (DSV) (37), and (iii) Other Viral databases (OV) (4, 5, 8). Only BLAST results showing expectation...
values of E < 0.001 were considered for phylogenetic and putative functional identification. All sequences were deposited in the GenBank database with accession numbers FJ640107 – FJ640554.

RESULTS

Banding Patterns and Sequences. Nine RAPD-PCR 10-mer primers were tested on sediment viral extract concentrates (Table. S1, supplemental material), and only primer OPA-9 yielded banding patterns for each location and sampling date (see Fig. S1, supplemental material). To determine the accuracy and fidelity of viral community RAPD-PCR, replicate samples from three stations were amplified in separate thermocycler reactions and the resulting banding patterns were compared. The level of similarity between replicate banding patterns ranged between 80% - 100% (see Fig. S2, supplemental material). In subsequent analyses of RAPD-PCR banding patterns, 80% or greater similarity was considered identical. Banding patterns from viral concentrates collected across the time series were compared using cluster analysis and showed that sample location was the strongest factor determining the clustering of RAPD-PCR fingerprints (Fig. 1). The maximum similarity among non-replicate sample banding patterns was 75% for station 804, between August 2003 and February 2004 samples. The average similarity across all banding patterns was 30% indicating that many of the sediment samples shared more than half of their total amplicons. Viral assemblages from mid-Bay station 804 showed the least change across the time series with the majority of samples having ~45% similarity in RAPD banding patterns. In contrast, sediments from upper bay station 908 had the greatest variability in viral assemblages (i.e., least overall similarity) with all samples showing
slightly less than 20% similarity. Viriobenthos assemblages within sediments from station 724 also had a high degree of variability with samples from 2003 scattered across several clades and separate from 2004 samples. Within each station, sediment viral assemblages from 2004 samples were typically more similar to one another than they were to 2003 samples, indicating a possible inter-annual change in Chesapeake Bay viriobenthos. Samples that did not amplify or showed less than 4 bands per lane were excluded from analyses due to probable degradation of DNA or incomplete amplifications (n = 5 of 25). DNase treated viral concentrates showed no loss of RAPD-PCR bands as compared to untreated controls (data not shown, (36)). Samples treated with heat + DNase prior to RAPD-PCR also showed no loss of bands. However a noticeable decrease in band intensity was observed.

**BLAST homology analysis of RAPD-PCR amplicon sequences.** Although the bay stations are quite different according to chemical and physical composition (25), one recurring band of ~ 625 bp was observed within twelve RAPD-PCR banding patterns from the total population of 20 samples (See Fig. S1, supplemental material). DNA within this band was cloned from the RAPD-PCR patterns of two samples (one each from stations 908 and 804). Subsequently, two clones from each of these bands were sequenced. Also included in this analysis was a single clone containing a sequence homologous to the ~625 bp band that occurred within a clone library of RAPD-PCR amplicons from another station 804 sample. Alignment of these five amplicon sequences showed a 1.6% overall divergence at the nucleotide level and each sequence gave the same top BLAST homolog when assessed against environmental sequence databases [CBV (E < 10^{-20}), env-nt (E < 10^{-9}), and DSV (E < 10^{-4})]. No significant BLAST homology was found in searches against GenBank nt/nr.
All RAPD-PCR sequences were compared to sequences in the GenBank non-redundant and environmental non-redundant databases as well as three databases of viral metagenomic sequences. In all cases only those homologs showing a BLAST expect value of E < 0.001 were considered significant. Median expectation values for all BLAST sequence homologs within each database, and according to taxonomic classification (i.e., virus, bacteria or environmental) are listed in Table 1. Of the 448 analyzed sediment viral sequences, 54% showed no significant homology to sequences within any of the subject BLAST databases (i.e., nr, nt, env-nr, env-nt, or viral metagenomes). The frequency of significant homology to sequences in the GenBank nr/nt and environmental databases was similar, 24 and 22%, respectively.

To determine whether sequence length significantly co-varied with BLAST expectation value, the length of RAPD-PCR amplicons was plotted against the expectation values (E-value) of their best BLAST alignment. Overall, the mean read length for RAPD-PCR amplicon sequences showing significant BLAST homology was 359 bp with an average 45% G+C content. However, when examined according to subject database, RAPD-PCR average sequence length varied slightly: 409 bp for homologs to nr/nt, 404 bp for env-nr/env-nt, 411 bp for OV, 390 bp for CBV and 388 bp for DSV metagenomes (Fig. 2). Longer sequence lengths did not appear to result in lower E-values as many of the longest RAPD-PCR sequences showed relatively high BLAST expectation values. Indeed the lowest E-scores were seen among sequences near the average sequence length for the entire collection of RAPD-PCR amplicons.

Median E-score for the collection of BLAST homologs to RAPD-PCR amplicons within each subject database or taxonomic domain provides a global indication of the genetic similarity between viroplankton assemblages and a given collection of sequences. By these criteria, sequences within the Chesapeake Bay Viroplankton (CBV) subject database showed the best
overall similarity (lowest log median E-score) to viriobenthos RAPD-PCR sequences as compared to any other environmental database (Table 1). Indeed, this median E-score (log -9.36) was nearly as low as that for GenBank nr/nt (log -9.48), a database containing nearly 1,000 fold more sequences. In contrast to the CBV database, the median E-score of BLAST homologs to metagenome sequences from Delaware soil viruses (DSV) was the highest of all subject databases (log -6.40), broadly indicating the larger genetic distance between viriobenthos assemblages and soil viral assemblages.

**Taxonomic and functional characterization of RAPD-PCR amplicon sequences.** For each amplicon sequence, the top five hits were analyzed and the best hit was selected. Best hits were selected based on E value and most appropriate taxonomic identification. Analysis of significant homologs (E < 0.001) to sequences within GenBank (nr/nt) showed that 29% of the GenBank (nr/nt) sequence homologs were classified as viral and 70% bacterial. Among the RAPD-PCR amplicon sequences with a viral best hit, 46% belonged to bacteriophage within the order Caudovirales (*i.e.*, the tailed phages) and another 11% were of algal virus origin (*i.e.*, Phycodnaviridae). Analysis of viral homologous sequences showed that a large proportion of these belonged to prophage (34%), as well as the Podoviridae (24%) or Myoviridae (22%) viral families (see Fig. S3, supplemental material). Proteobacteria comprised the majority of bacterial best hits to RAPD-PCR amplicon sequences, of which most were from the class Alphaproteobacteria. Among the bacterial homologs, 57% were listed as prophage/phage-related and 23% were classified as hypothetical or conserved hypothetical proteins. Of the hypothetical proteins, 6% were located within 60 kb of phage-related genes (*e.g.*, terminase, integrase) and 4% near transposases within their host genomes (Fig. 3).
Gene homologs of RAPD viral sequences are listed in Tables 3 and 4. The majority of homologs were to hypothetical proteins of unknown function in each domain, 51% of the total from viruses, prophage and bacteria (Table 2). Replication, recombination and repair and structural protein functional groups accounted for 19% of the gene homologs within viruses and prophage. Of the bacterial homologs, cell wall and membrane genes were the most dominate functional class. For putative gene identifications, lowest expectation values for virus sequences were for terminase (E-value = -31.5) and putative carbohydrate kinase (E-value = -35.4) (see Table S2, supplemental material). Among prophage sequences, a late control gene and a hypothetical protein were the lowest expectation values, -16.6 and -16.2 respectively. With an E-value of -70.7, an acyl-CoA synthetase sequence homolog was the most significant among the bacterial sequences, however, this homolog also happened to be near (± 60kb) a phage integrase.

DISCUSSION

The objective of this study was to determine if RAPD-PCR could be an effective and useful method to evaluate the seasonal dynamics and preliminary metagenomic characterization of viruses in environmental samples, specifically in the eutrophic sediments of the Chesapeake Bay. The resulting banding patterns produced were robust and most importantly, reproducible. In effect, RAPD-PCR provided a two-pronged approach for assessment of viral diversity in estuarine sediments; i). at the community level, comparison of RAPD-PCR banding patterns allowed for inferences concerning the genetic similarity between sediment viral assemblages and ii). the level of individual viral genotypes, sequence analysis of RAPD-PCR amplicons provided a detailed view of the genetic composition of sediment viral assemblages.
RAPD-PCR profiling of viral communities in sediments. Overall, it is apparent that RAPD banding patterns do not remain absolutely consistent over time at any given location, indicating that viriobenthic assemblages are temporally and geographically dynamic. The high viral production rates and short turnover times reported for many sediment environments (14, 27) lend further support to the notion that the composition of viriobenthos assemblages can change quickly. In contrast to water column environments where PFGE and RAPD-PCR analysis indicated seasonally linked changes in Chesapeake virioplankton assemblages (36, 40), viriobenthos assemblages in the Bay tended to change with station location. The most likely explanation for the variability in sediments from different stations is the spatial heterogeneity of the bay sediments. Unlike the water column where a given volume of water is in constant motion, sediments are relatively stable and have limited movement as compared to the water column. Even though sediments shift due to movements by benthic organisms (e.g., worms, bivalves) as well as resuspension and sedimentation of particulates, these changes are less obtrusive in the sediment environment than in the water column.

RAPD-PCR amplicon sequence analysis. BLAST sequence homology searches for this study showed that over half of the sediment viral RAPD-PCR amplicon sequences were completely novel. This is a high proportion compared to the ca. 30% novel sequence typically reported for environmental dsDNA viral metagenome sequence libraries (3, 4). However, the 24% frequency of BLAST-positive RAPD-PCR amplicon sequences against sequences within GenBank nt/nr is similar to that seen in other viral metagenomic studies that have used traditional Sanger sequencing. In many cases, RAPD-PCR amplicons from sediment viruses displayed significant BLAST homology to sequences within more than one database. In those instances, it is clear that frequently the best homologs to sediment amplicon sequences occur in
the environmental and metagenomic databases and not in the known organismal databases.

Homologs to bacterial functional genes were also found among RAPD-PCR amplicons, a finding echoed in several other metagenomic studies [see review by Comeau et al. (12)]. A large metagenomic study of virioplankton assemblages found a far higher proportion (91%) of unknown sequence (2), a result which was largely attributable to the short read length of sequences (~100 bp) within those shotgun libraries. Recent *in silico* metagenome simulation experiments clearly demonstrated that short read sequencing technology performs especially poorly at characterizing the functional genetic diversity within virioplankton assemblages (38).

Moreover, the low homolog detection frequency of short viral metagenome reads is not alleviated by the increases in sequence coverage newer technologies can provide.

It is likely that the high proportion of unknown sequences seen for this collection of RAPD-PCR amplicons can be explained, in part, by the relatively short average read length of these sequences. Nevertheless examination of the distribution of the sequence length versus significant BLAST expectation values, indicated that longer sequence lengths did not necessarily yield lower expectation values (Fig. 2). Thus even short reads can produce good quality sequence matches. Other studies which have examined the functional and taxonomic diversity of RAPD-PCR amplicons from viral assemblages though BLAST homology searches reported similar BLAST-positive and novel sequence frequencies to those seen here (35, 36). In the case of viral assemblages at deep-sea hydrothermal vents (35), the frequencies of BLAST-positive to novel sequence was similar for RAPD-PCR amplicon libraries and a small random shotgun sequence library which did not involve a selective PCR step. Overall, the similar behavior of sediment RAPD-PCR sequences in BLAST analysis indicate that these amplicons were derived from viral assemblages and are a reasonable means to superficially survey the functional and
taxonomic diversity within sediment viral assemblages. These findings also reinforce the idea
that the majority of extant viral diversity in the biosphere is poorly characterized.

A preponderance of sequences with homology to genes belonging to members of the
*Podoviridae* and *Myoviridae* families was seen within a metagenome library of Chesapeake Bay
virioplankton (3) as well as in this study. This finding contrasts with previous viral metagenome
studies from marine waters (8), coastal sediments (4) and equine feces (9) which found that the
majority of viral BLAST hits belonged to phages within the *Siphoviridae* family (18). Phage
within the *Podoviridae* and *Myoviridae* families are often considered to be virulent, while phages
within the *Siphoviridae* family have been considered to be more often temperate (4). Thus, the
observation of frequent homologs to sequences belonging to members of the *Podoviridae* and a
low frequency of hits to *Siphoviridae* sequences indicates that the sediments of the Chesapeake
Bay do not contain many phage families known to be temperate.

Persistence of identical RAPD-PCR bands indicates that the same viral genes can be
found and are maintained within different sediment environments over time. Conserved genes
exist among viral groups and extensive gene transfer can occur among dissimilar viruses in the
environment (26, 30). Whether this ~ 625 bp sequence occurred within only one type of virus or
within several different viruses cannot be unambiguously ascertained from this study.

However, the possibility of finding a common viral gene within a large collection of viral
assemblages is not unexpected as previously reported data has shown that nearly identical T7-
like DNA polymerase sequences can be detected in widely divergent environments (6).

Moreover, diversity analyses based on metagenomic sequence data have reported that local scale
viral diversity is high, but is more limited on a global scale (7). By extension, this result predicts
that specific sequences can be found across a broad range of environments. Thus, there is a good
likelihood that this ~625 bp sequence is distributed across several different viruses and not limited to a single commonly occurring virus.

In the present study, RAPD-PCR has been successfully used to examine the composition of viriobenthos assemblages in sediment samples from the Chesapeake Bay. Diversity measurements based solely on morphological classification of viral particles or analysis of viral metagenome sequences, lack the sensitivity, resolution and/or sample throughput necessary for determining short term changes in the composition and diversity of viriobenthos assemblages. As the data shows, RAPD-PCR banding patterns can address these morphological shortcomings and thus provide data needed to assess fine-scale synecological effects of viruses within sediment microbial environments.

ACKNOWLEDGEMENTS

This research was supported in part by an NSF Microbial Observatories grant (MCB-0132070) awarded to K. E. W. We would also like to express our sincerest gratitude to S. Polson and J. Bhavsar for computational assistance as well as D. M. Winget and K. E. Williamson for helpful comments, and thank the captain and crew of the R/V Cape Henlopen, as well as Old Dominion University for use of the multi-corer device and the staff of the DNA Sequencing & Genotyping Center at the Charles C. Allen, Jr., Biotechnology Laboratory at the University of Delaware.
Table 1. Log transformed median BLAST expectation values for RAPD-PCR sequences with significant BLAST homology (E < 0.001). *Sequence was similar to at least one environmental database sequence (Env nt, Env nr, or viral metagenomes), but showed no similarity to nt or nr databases. ^For each environmental database the # of sequences includes all sequences that had a significant BLAST homolog to that database regardless of hits to other databases.

<table>
<thead>
<tr>
<th>Database</th>
<th>Median</th>
<th># of sequences</th>
<th>% of total</th>
</tr>
</thead>
<tbody>
<tr>
<td>GenBank nt/nr</td>
<td>- 9.48</td>
<td>107</td>
<td>24</td>
</tr>
<tr>
<td>Bacteria</td>
<td>- 6.78</td>
<td>75</td>
<td>17</td>
</tr>
<tr>
<td>Virus</td>
<td>- 7.66</td>
<td>31</td>
<td>7</td>
</tr>
<tr>
<td>Environmental*</td>
<td>- 8.00</td>
<td>97</td>
<td>22</td>
</tr>
<tr>
<td>Env nt/nr^</td>
<td>- 8.01</td>
<td>147</td>
<td>32</td>
</tr>
<tr>
<td>CBV</td>
<td>- 9.36</td>
<td>122</td>
<td>27</td>
</tr>
<tr>
<td>DSV</td>
<td>- 6.40</td>
<td>33</td>
<td>7</td>
</tr>
<tr>
<td>OV</td>
<td>- 7.54</td>
<td>86</td>
<td>19</td>
</tr>
<tr>
<td>No homology</td>
<td>-</td>
<td>244</td>
<td>54</td>
</tr>
<tr>
<td>Total</td>
<td>N/A</td>
<td>448</td>
<td>100</td>
</tr>
</tbody>
</table>
Table 2. Distribution of homologs to RAPD-PCR amplicon sequences according to protein functional groups and taxonomic domain. The value in parentheses (#) is the % of homologs within that domain. All sequences had expectation values E < 0.001.

<table>
<thead>
<tr>
<th>Domain</th>
<th>Protein function</th>
<th># of homologs</th>
<th>(% of domain)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Virus</td>
<td>Nucleotide transport and metabolism</td>
<td>5 (16)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Replication, recombination and repair</td>
<td>6 (19)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Structural</td>
<td>6 (19)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Unknown function</td>
<td>13 (45)</td>
<td></td>
</tr>
<tr>
<td>Prophage</td>
<td>Replication, recombination and repair</td>
<td>4 (19)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Structural</td>
<td>6 (29)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Unknown function</td>
<td>10 (52)</td>
<td></td>
</tr>
<tr>
<td>Bacteria</td>
<td>Energy production and conversion</td>
<td>1 (2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cell wall / membrane / envelope biogenesis</td>
<td>7 (13)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Posttranslational modification, protein turnover, chaperones</td>
<td>1 (2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Amino acid transport and metabolism</td>
<td>2 (4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cell motility</td>
<td>1 (2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Carbohydrate transport and metabolism</td>
<td>1 (2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cell envelope biogenesis, outer membrane</td>
<td>1 (2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Replication, recombination and repair</td>
<td>2 (4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Inorganic ion transport and metabolism</td>
<td>5 (9)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Intracellular trafficking and secretion / Extracellular structures</td>
<td>4 (7)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Translation, ribosomal structure and biogenesis</td>
<td>1 (2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Unknown function</td>
<td>28 (52)</td>
<td></td>
</tr>
</tbody>
</table>
**Figure 1.** UPGMA tree of viriobenthos RAPD-PCR banding patterns shown in Fig. 1. Jaccard's similarity scale is listed along top.

**Figure 2.** Sequence length distribution of RAPD-PCR amplicons versus their log transformed expectation values less than $10^{-3}$. Overall average sequence length was 359 bp. A) All databases results combined; B) GenBank non-redundant database (nr); C) GenBank translated database (nt); D) Environmental non-redundant database (nr); E) Environmental translated database (nt); F) Chesapeake Bay Virioplankton (CBV); G) Other Viral metagenomes (OV); and H) Delaware Soils Viruses (DSV).

**Figure 3.** Comparison of protein homologs to sediment viral concentrate RAPD-PCR sequences, n = 112 of a total n = 448 sequenced. Phage, phage-related, or near-phage sequences comprised 69% of the total. Values in parentheses = (median log transformed expectation values). Results based on tBLASTx expectation values ($E < 0.001$). Abbreviations are as follows: Bact-bacteria; Hyp-hypothetical gene; Fxn-gene with defined function; +P within 60 kb of phage-related gene(s); -P no close proximity to phage-related gene(s); and +T within 60 kb of transposon-related gene(s).
REFERENCES:


