

Global distribution of nearly identical phage-encoded DNA sequences

Mya Breitbart, Jon H. Miyake, Forest Rohwer *

Department of Biology, LS301, Center for Microbial Sciences, San Diego State University, 5500 Campanile Drive, LS301, San Diego, CA 92182-4614, USA

Received 23 March 2004; received in revised form 26 May 2004; accepted 27 May 2004

First published online 9 June 2004

Abstract

Phages, the most abundant biological entities on the planet, play important roles in biogeochemical cycling, horizontal gene transfer, and defining microbial community composition. However, very little is known about phage diversity or biogeography, and there has not yet been a systematic effort to compare the phages found in different ecosystems. Here, we report that T7-like Podophage DNA polymerase sequences occur in every major biome investigated, including marine, freshwater, sediment, terrestrial, extreme, and metazoan-associated. The majority of these sequences belong to a unique clade that is only distantly related to cultured isolates. Some identical T7-like phage-encoded DNA polymerase genes from this clade were >99% conserved at the nucleotide level in multiple different environments, suggesting that these phages are moving between biomes in recent evolutionary time and that the global genomic pool for T7-like phages may be smaller than previously hypothesized.

© 2004 Federation of European Microbiological Societies. Published by Elsevier B.V. All rights reserved.

Keywords: Phage; DNA polymerase; Distribution; Ubiquitous sequence; Environment; T7-like Podophage

1. Introduction

Tailed, dsDNA phages are ubiquitous and abundant in culture collections and in the environment [1,2]. *Podoviridae* are one of the main groups of dsDNA phages that have been identified in the environment by electron microscopy [3], genomic analysis of uncultured marine viral communities [4], and pulse-field gel electrophoresis [5,6]. *Podoviridae* are classified by the International Committee on the Taxonomy of Viruses (ICTV) as short-tailed phages, with genomes of ~40 kb in length [7]. The Phage Proteomic Tree, a recently developed genome-based taxonomy for phages, splits the ICTV *Podoviridae* into two groups: T7-like and PZA-like Podophages [8].¹ All known T7-like Podophages have

genomes of ~40 kb and are short-tailed viruses with lytic life cycles. The T7-like Podophages also have conserved loci (e.g., DNA polymerase, primase, etc.) that can be used in a manner analogous to the 16S rDNA in prokaryotes [8]. These loci are not shared with the PZA-like Podophages. In this study, the DNA polymerase gene was used to study the global distribution and diversity of T7-like Podophages.

2. Materials and methods

2.1. Large-scale isolation of viral community DNA

Total viral communities were purified from 200 liters of water taken from Lake Hodges (San Diego, CA; 5/01), Mission Bay (San Diego, CA; 10/99 and 6/01), Torrey Pines Estuary (Del Mar, CA; 6/01), Scripps Pier (La Jolla, CA; 5/01 and 6/01), and the Salton Sea (Imperial Valley, CA; 6/01), as well as from ~2 liters of sediment from Mission Bay (San Diego, CA; 6/01) and

* Corresponding author. Tel.: +1-619-594-1336; fax: +1-619-594-5676.

E-mail address: forest@sunstroke.sdsu.edu (F. Rohwer).

¹ The suffix “-phage” is used when referring to the genome-based classification as described by the Phage Proteomic Tree, while the suffix “-viridae” is used when referring to the ICTV classification.

human fecal matter (7/02) using a combination of differential filtration and density-dependent gradient centrifugation. The water samples were initially filtered through a 0.16- μm Centramate tangential flow filter (TFF; Pall Filtron) to remove bacteria, eukaryotes, and large particles. Most of the water and approximately 90% of the viruses, as determined by epifluorescent microscopy [9], passed through the filter and were collected in a separate tank. Subsequently, the viruses in the filtrate were concentrated using a 100-kDa TFF filter until the final sample volume was <100 ml. After the TFF, 8.5 ml of the viral concentrate (at a density of 1.15 g/ml) was loaded onto a cesium chloride (CsCl) step gradient. The steps were 1 ml each of 1.7, 1.5, and 1.35 g/ml. The gradient was ultracentrifuged at 55,000g for 2 h, and the 1.35–1.5 g/ml fraction was collected. This fraction contains the majority of the viral DNA as previously determined by pulse field gel electrophoresis [6]. After CsCl purification, the viruses were lysed using a formamide extraction and the DNA was recovered by an isopropanol precipitation and a CTAB extraction [10]. For the fecal and sediment samples, sterile PBS or 100 kDa filtered seawater, respectively, was added to the sample. The samples were mixed vigorously, centrifuged at 8000g for 10 min to remove large particles, and the free viruses in the supernatant were concentrated in the same manner as the water samples.

2.2. Small-scale viral community DNA isolation

Fifty milliliter water samples were collected from various locations and filtered through a 0.2- μm Acrodisc to remove bacteria, eukaryotes, and large particles. Polyethylene glycol (PEG 8000) was added to a final concentration of 10% and the samples were incubated for >12 h at 4 °C [10]. The samples were then centrifuged at 13,000g for 30 min to pellet the viral particles. DNA was recovered from the pellet using a CTAB extraction [10].

2.3. Total community DNA extraction from soil, sediment, and coral samples

Soil samples were collected using 50-ml tubes or syringe corers. Sediment and coral samples were collected on SCUBA. All samples were stored at either 4 or –20 °C until processing. Corals were processed as previously described [11]. DNA was extracted from 0.5 to 1.0 g sediment and soil samples using the UltraClean Soil DNA Kit (Mo Bio, California) according to the manufacturer's instructions.

2.4. Precautions taken to prevent contamination

One of the major concerns when working with any environmental sample is contamination. Sterile, virgin,

aerosol-resistant plasticware was used at all times. Sample DNA preparations and PCR assembly areas were physically separated from the thermocyclers and electrophoresis units. As described in Section 3, the widespread distribution of specific T7-like Podophage DNA polymerase sequences was confirmed by several independent investigators.

2.5. PCR amplification and cloning

All sequenced members of the T7-like Podophages encode a DNA polymerase, primase, and endonuclease (<http://salmonella.utm.edu/phage/tree/signature.html>). The DNA polymerase sequences of *Escherichia coli* ϕT7 , *Yersinia enterocolitica* $\phi\text{Ye03-12}$, and *Roseobacter SIO67* ϕSIO1 were aligned, and degenerate primers were designed to the conserved regions. Two of these primers, T7DPol230F (5' ARG ARM RIA AYG GIT 3') and T7DPol510R (5' GTR TGD ATR TCI CC 3'), were optimized to detect as few as 10^3 copies of *E. coli* ϕT7 by PCR. The PCR mixture (50 μl total volume) contained target DNA, 1 \times REDTaq Buffer (Sigma, Missouri), 1 U REDTaq (Sigma, Missouri), 200 μM dNTPs, 1 μM each primer, and 1 μl of 50 mM MgCl_2 . In some cases, an additional 1 U of REDTaq and 2 μl of BSA (100 $\mu\text{g}/\text{ml}$) were added to facilitate amplification of environmental samples. The thermocycler conditions were: 5 min at 94 °C, 30 cycles of [1 min at 94 °C, 1 min at 50–0.5 °C/cycle, and 2 min at 72 °C], and 10 min at 72 °C. After PCR, the products were cloned into pCR4-TOPO according to the manufacturer's instructions (Invitrogen, California). Transformants were selected on Luria–Bertani/ampicillin plates with X-gal (5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside). White colonies were randomly selected and screened by PCR with the M13F and M13R primers to identify clones with inserts between 500 and 1000 bp. Products of the correct size were isolated using a PCR Clean-up Kit (Mo Bio, California) and sequenced 3 \times using the M13F, M13R, and T7DPol510R primers. Consensus sequences were constructed using Sequencher 4.0 (Gene Codes, Michigan). The consensus sequences were compared against the GenBank database using TBLASTX [12,13] and those with similarity to known DNA polymerase sequences were retained for further analysis. GenBank Accession Nos. for the DNA polymerase sequences are AY599945–AY600060. Sequences showing >99% similarity on the nucleotide level were considered to represent the same sequence. Unique sequences were aligned and a Neighbor-Joining Tree was created using CLUSTALX [14,15]. The tree was bootstrapped 1000 times. No discernable effect on the tree structure was noticed when multiple substitutions were considered in the alignments [14,15].

2.6. Amplification of specific DNA polymerase sequences

Primers were designed to amplify two specific DNA polymerase sequences (HECTOR and PARIS – see Section 3) from the Polymerases from Uncultured Podophages (PUP) clade (Fig. 1). The HECTORPol29F (5' GCA AGC AAC TTT ACT GTG G 3') and HECTORPol711R (5' CGA GAG ATA CAC CAA CGA A 3'), as well as the PARISPol25F (5' ATA CTA CAC GCT ACT CTG G 3') and PARISPol701R (5' GAG TGG CAA GAG GAG TTA T 3') primer sets were used to amplify these specific DNA polymerases from phages present in a wide variety of environments. The reaction mixture for the HECTOR primers (50 μ l total volume) contained 1 μ l of target DNA, 1 \times Taq Buffer, 1 U Taq DNA Polymerase, 200 μ M dNTPs, 1 μ M each primer, and 1 μ l of 50 mM MgCl₂. The reaction mixture for the PARIS primers was identical except there was no additional MgCl₂. The thermocycler conditions for amplifying HECTOR were: 5 min at 94 °C, 35 cycles of [1 min at 94 °C, 1 min at 61–0.5 °C/cycle, and 1 min at 72 °C], and 10 min at 72 °C. The thermocycler conditions

for amplifying PARIS were: 5 min at 94 °C, 35 cycles of [1 min at 94 °C, 1 min at 66–0.5 °C/cycle, and 1 min at 72 °C], and 10 min at 72 °C. HECTOR and PARIS products were sequenced from both directions and aligned using CLUSTALX in order to ensure that they represented the correct sequence.

2.7. Quantification of specific polymerase sequences

Real-time PCR and Taqman PCR were used to determine the concentrations of HECTOR and PARIS in a subset of the environmental samples. Primer and probe sequences were designed using the Beacon Designer (Premier Biosoft International, California) and Primer3 [16]. Primers for the HECTOR DNA polymerase sequence were HECTORPol563F (5' CTT CTC AGT TTT CTG TT 3') and HECTORPol800R (5' GCA AGC AAC TTT ACT GT 3'). For detection of the HECTOR sequence, a SYBR Green I real-time PCR assay was used. Taqman PCR was used for detection of PARIS. Primers for the PARIS DNA polymerase sequence were PARISPol480F (5' AAG TTG TGC TTC TGG TA 3') and PARISPol786R (5' ATA CTA CAC GCT ACT CT 3'). The Taqman probe for the PARIS DNA polymerase was 5'-FAM- TTA TTG AAG CTG AGA TGC AA-BHQ1-3'. All probes, templates, and primers were analyzed for secondary structure using the mFold program [17], and NCBI BLASTN [12,13] was used to determine specificity. Real-time amplification of HECTOR consisted of 1 cycle of 4 min at 95 °C and then 45 cycles of [10 s at 95 °C, 30 s at 60 °C, and 30 s at 72 °C]. Each 20 μ l reaction contained final concentrations of 1 \times Bio-Rad iQ SYBR Green Supermix (50 mM KCl, 20 mM Tris-HCl, pH 8.4, 0.2 mM dNTPs, 25 U/ml iTaq DNA polymerase, 3 mM MgCl₂, and SYBR Green I), 3 μ g/ μ l BSA, 200 nM of each primer, and 1 μ l sample DNA. During the optimization of the real-time PCR assays, it was discovered that some environmental samples contained inhibitors to this assay. The addition of BSA to the reaction mixture was able to decrease this inhibition by over 90%. The assays were performed in a Bio-Rad iCycler real-time PCR machine. Each sample was performed at two different concentrations and repeated at least three times. Analysis of the melt-curve of the PCR products indicated that there was only one species of product and this was also confirmed using an agarose gel. No primer-dimers were observed at any time. The Taqman assays to quantify PARIS were performed as described above except that they contained 50 nM PARIS Taqman probe and 1 \times iQ Supermix (Bio-Rad, California) which does not contain SYBR Green I. Absolute concentration values were derived using a plasmid containing one copy of the HECTOR or PARIS target sequence. Briefly, the HECTOR and PARIS T7-like DNA polymerase sequences were amplified by PCR and cloned into the pGEM-T vector

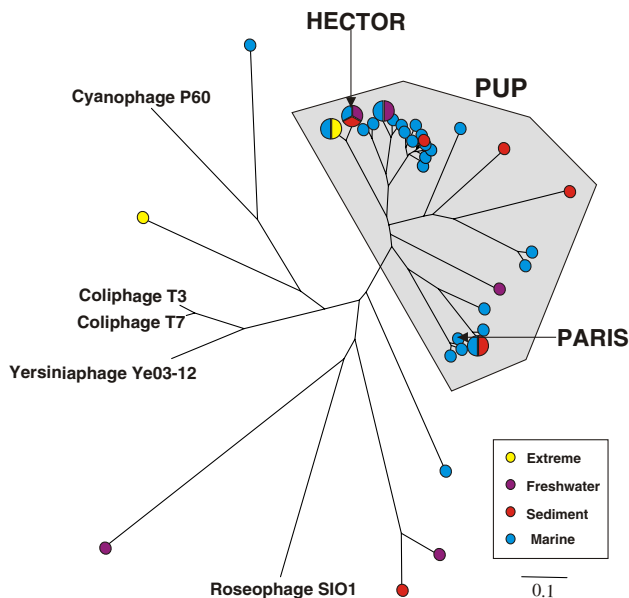


Fig. 1. Neighbor-Joining tree of T7-like DNA polymerases recovered from environmental samples using degenerate PCR primers. The circles represent unique polymerase sequences identified in this study. The polymerase sequences of phage isolates from this group are represented by the name of the isolate. Any sequences with $\geq 99\%$ identity at the nucleotide level were considered the same sequence. Using this criterion, a total of 34 unique T7-like DNA polymerase sequences were recovered. The environment(s) in which each sequence was found in more than one environment and are represented by slightly larger circles. Twenty-eight of the unique sequences belonged to the PUP clade, which is outlined in gray. The scale bar represents 0.1 changes per 100 bp. The tree was bootstrapped 1000 times. No discernable effect on the tree structure was noticed when multiple substitutions were considered in the alignments.

(Promega, Wisconsin) and then transformed into XL1-Blue competent cells. The plasmids were sequenced and concentrations were determined by spectrophotometry and confirmed by agarose gel analysis. Serial dilutions were used to produce final concentrations of 10^1 – 10^4 copies/ μ l. The iCycler software analysis program was used to calculate Ct values and determine sample concentrations based on the standards.

3. Results and discussion

To determine the diversity and biogeographical distribution of T7-like Podophages, 17 environmental viral communities (Table 1) were checked for the presence of T7-like Podophages using degenerate primers designed to conserved regions in the DNA polymerases of this group. The samples included representatives from marine, freshwater, estuarine, sediment, and hyper-saline environments. A group of DNA polymerase sequences similar to the T7-like Podophage DNA polymerases was recovered from each of these environments (Table 1). Since these sequences were obtained from purified viral communities, they are most likely phage-encoded.

The T7-like DNA polymerase sequences obtained with degenerate primers were grouped together by considering any sequences with $\geq 99\%$ identity at the nucleotide level as the same sequence. Using this criterion, a total of 34 unique sequences were obtained from the original 116 sequences (Table 1). Twenty-seven of the 34 unique T7-like DNA polymerase sequences were found in only one environmental sample. As many as

seven different sequences were recovered from a single sample, indicating that different Podophages can successfully co-occur. Surprisingly, seven of the T7-like DNA polymerase sequences were found in more than one sample and four of these sequences were recovered from more than one environment (Fig. 1).

The majority of the sequences recovered by degenerate PCR were not closely related to previously cultured members of the T7-like Podophages. Instead, the environmental sequences were much more closely related to each other and form a unique clade, which was named PUP (Fig. 1). Of the 34 unique T7-like DNA polymerase sequences recovered, 28 belong to the PUP clade. The sequences belonging to the PUP clade were >200 bp shorter than *E. coli* ϕ T7 DNA polymerase over the amplified region. The PUP sequences therefore appear to represent an abundant and diverse group of environmental phages only distantly related to the closest isolates.

There are currently 154 complete phage genomes and thousands of DNA polymerase genes in the GenBank database. Amongst these sequences there is not a single instance of a T7-like DNA polymerase in a non-T7-like genome. In environmental samples, we have determined that T7-like DNA polymerases are only associated with genomes of ~ 40 kb. This was accomplished by running 10 uncultured viral communities on pulse-field gels [5,6], cutting the gel into fragments, extracting DNA from these fragments, and looking for the presence of T7-like DNA polymerase sequences in each genome size range using PCR. These sequences were only found in the ~ 40 -kb fraction, suggesting that these sequences came

Table 1
Environmental viral communities analyzed with degenerate primers to the T7-like DNA polymerases

Location	Sample type	Date	CsCl?	# Sequences	# Unique sequences
Torrey Pines Estuary, California	Estuarine	06/01	Y	8	1
Salt Lake Marina, Utah	Hypersaline	05/02	N	1	1
Salton Sea, California	Hypersaline	06/01	Y	1	1
Hemet Well #27, California	Freshwater	07/02	N	5	1
Idaho Shallow Well	Freshwater	06/02	N	2	1
Lake Hodges, California	Freshwater	06/01	Y	3	3
Saddlebag Creek, California	Freshwater	07/01	N	5	1
Melbourne Beach, Florida	Marine	04/02	N	14	7
Melbourne Marina, Florida	Marine	04/02	N	16	3
Mission Bay, California	Marine	06/01	Y	5	4
Ocean Beach, California	Marine	06/02	N	4	2
Puerto Rico nearshore surface	Marine	02/02	N	7	2
Scripps Pier, California	Marine	05/01	Y	18	2
Scripps Pier, California	Marine	06/01	Y	3	3
Scripps Pier, California	Marine	04/02	N	14	7
Weimea Bay, Hawaii	Marine	02/02	N	1	1
Mission Bay, California	Sediment	06/01	Y	9	6
Total				116	34

A total of 116 sequences were recovered from the 17 environmental samples. Sequences with $<99\%$ identity at the nucleotide level were considered unique. If a sequence was observed in more than one sample, then it was only counted once in the total number of unique sequences. Filtration (0.2 μ m) was used to remove non-viral particles from all of the samples. Where indicated, the viral community was purified by an additional density-dependent cesium chloride (CsCl) gradient step to ensure that the sequences were phage-encoded.

from T7-like Podophages and are not moving to phages with other sized genomes (data not shown). Therefore, we believe that these DNA polymerase sequences come from T7-like Podophages.

The T7-like DNA polymerase gene recovered most frequently using the degenerate primers was found in purified viral communities from three of the world's largest biomes – marine, subsurface terrestrial (represented by aquifer water), and sediments [18]. This sequence was designated HECTOR because it was found in samples from Hawaii, extreme environments, coral, terrestrial, ocean, and rumen. Conventional PCR primers designed to specifically amplify the HECTOR sequence resulted in the identification of this sequence in 49 out of 66 environmental samples originating from all over the globe and collected over a time period of 4 years. These primers detected the HECTOR DNA polymerase sequence in numerous marine (Atlantic, Pacific, and Antarctic), freshwater (ice, river, well, lake), estuarine, hot springs, sediment (Atlantic, Pacific), terrestrial (within the United States), and metazoan-associated (coral mucus, cow rumen, and human feces) samples.

Products from 18 samples positive for HECTOR representing each of the major biomes were sequenced: 11 were identical over the entire 533-bp region, two of the sequences differed by one nucleotide, three differed by two nucleotides, and two differed by three nucleotides. Therefore, the HECTOR sequences from these different environments differed by <0.6% at the DNA level, suggesting a recent origin.

To demonstrate that the HECTOR sequence was not unique in its widespread distribution, primers specific to PARIS (Panama, the Atlantic, rhizosphere, and sediment) were also constructed. PARIS was a T7-like DNA polymerase sequence that was recovered from only one marine sample in the original PCR with degenerate primers. PARIS was found in 27% of the samples screened using conventional PCR with specific primers. Although less abundant than HECTOR, PARIS was also widespread and present in each of the major biomes based on PCR data. Both HECTOR and PARIS were more abundant in the marine environment than in other environments.

Numerous precautions were taken to ensure that the results were not influenced by contamination. Evidence of contamination was never observed in 86 separate negative controls. Five independent investigators in six separate laboratories (in California, Delaware, and Florida), using completely new reagents each time, confirmed the widespread distribution of the HECTOR and PARIS sequences. Each investigator used their own DNA samples (i.e., the DNA had never been at San Diego State University) and used primers shipped directly from different vendors (Invitrogen, California or GenBase, California). Systematic contamination (e.g., in

one of the reagents) is not an explanation because many negative samples were observed. The samples analyzed in Florida were prepared by a sixth investigator using filtration and the reagents came from different manufacturers. We have also used our T7-like DNA polymerase primers in conjunction with vector primers to PCR amplify T7-like DNA polymerases and their flanking sequence out of shotgun libraries [4]. This would not be possible if these results were contamination from PCR products. Most of the shotgun libraries and many of the environmental DNA samples were prepared before the work described in this manuscript was started, so the PCR products did not even exist. Finally, we did not identify exactly the same sequences in all the samples, as would be expected with contamination. We observe 1–3 bp changes in some of the HECTOR sequences. These changes are different in different samples and they are repeatable within the same sample (i.e., we can go back to the same sample and find the same sequence).

The HECTOR and PARIS sequences were quantified using real-time PCR (Table 2). The real-time PCR supported the results from the conventional PCR and showed that both the HECTOR and PARIS sequences were widespread. Thirty-four percent of the samples were positive for HECTOR and 12% of the samples were positive for PARIS by real-time PCR (Table 2). These phage sequences never comprised >0.3% (and usually <0.1%) of the total number of phage particles in the sample.

There are an estimated 10^{30} prokaryotes in the biosphere, the vast majority of which are in the subsurface terrestrial, sediments, and the oceans [18]. Generally, there are ~1–10 phage particles per prokaryotic cell in environmental samples [19–22]. Therefore, most of the world's phages are probably found in these biomes as well. Signature genes for phage clades, as described on the Phage Proteomic Tree [8], or by other criteria [23–25] make it possible to begin examining phage diversity and distribution between different biomes. The PCR-based assays described here revealed that there is a large pool of environmental viruses carrying a DNA polymerase gene most closely related to that of the T7-like Podophages. As with 16S rDNA analyses of prokaryotes, characterizing a signature gene only allows speculation about the biology of the target. There is the possibility that this gene has moved laterally during viral evolution and may be encoded by unrelated phage types, though this is not supported by the known phage genomes [8]. Therefore, we surmise that the HECTOR and PARIS DNA polymerase sequences belong to a group of short-tailed, dsDNA phages with ~40 kb genomes and lytic lifestyles. The PUP clade is not closely related to cultured isolates and studies targeted at isolating phages from this group need to be performed. This approach of first identifying uncultured diversity and then

Table 2

Environmental viral communities analyzed by real-time PCR to detect HECTOR and by Taqman real-time PCR assay to detect PARIS

Location	Date	# Hector/10 ⁶ phage	# Paris/10 ⁶ phage
<i>Estuarine</i>			
Torrey Pines Estuary, California*	06/01	535	635
<i>Extreme</i>			
Salt Lake Marina, Utah*	05/02	–	–
Little Hot Creek Hot Springs, California*	07/02	637	–
<i>Metazon-associated</i>			
<i>Colpophyllia natans</i> , Panama	04/99	–	–
<i>Mohtastraea annularis</i> , Panama	06/00	–	–
<i>Mohtastraea franski</i> , Panama	06/00	–	–
<i>Porites astreoides</i> , Panama	06/00	–	–
<i>Porites furcata</i> , Panama	04/99	–	–
<i>Mohtastraea franski</i> , Bermuda	08/99	3070	630
Cow Rumen, Idaho	06/02	–	–
Human feces	07/02	–	–
<i>Fresh water</i>			
African Stream	07/02	–	–
Antarctic Ice	10/01	–	–
Colorado River, Colorado	07/02	–	–
Deer Creek, Utah	05/02	–	–
Hemet Well # 27, California*	07/02	370	–
Hemet Well # 29, California	07/02	–	–
Hemet Well # 34, California	07/02	–	–
Idaho Deep Well	06/02	–	–
Idaho Shallow Well*	06/02	–	–
Lake Havasu, Arizona	05/02	–	–
Lake Hodges, California*	06/01	473	–
Rio Grande, Arizona	07/02	–	–
Saddle Creek, California*	07/01	–	–
Utah Lake, Utah	05/02	–	–
<i>Marine</i>			
Antarctic 1, 500 m	10/01	10	–
Antarctic 15, 1 m	10/01	2	–
Antarctic 8, 1 m	10/01	–	–
Bermuda Atlantic Time Series, 100 m	09/99	–	–
Bermuda Atlantic Time Series, 15 m	09/99	–	31
Bermuda Atlantic Time Series, 3 m	09/99	–	–
Makapuu, Hawaii	02/02	55	–
Melbourne Beach, Florida	04/02	107	88
Mission Bay, California*	10/99	200	–
Mission Bay, California*	06/01	704	–
Newport Beach, California	06/02	712	761
Puerto Rico nearshore surface sample 1*	02/02	–	–
Puerto Rico nearshore surface sample 2	02/02	43	–
Puerto Rico, 30 ft deep	02/02	–	–
Scripps Pier, California*	05/01	636	–
Scripps Pier, California	04/02	689	–
Shark's Cove, Hawaii	02/02	–	–
Waianae, Hawaii	02/02	–	–
Weimea Bay, Hawaii*	02/02	–	–
<i>Sediment</i>			
Marsh Sediment, Georgia	06/02	–	0.08
Media Luna reef (Puerto Rico), 16 ft deep	04/01	–	1.62
Media Luna reef (Puerto Rico), 42 ft	04/01	–	–
Laurel reef (Puerto Rico), 56 ft deep	04/01	–	–
Laurel reef (Puerto Rico), 52 ft deep	04/01	–	–
Media Luna reef (Puerto Rico), 57 ft deep	04/01	–	–
Tarramote reef (Puerto Rico), 45 ft deep	04/01	–	–
La Parguera Mangroves, Puerto Rico	04/01	–	–
Media Luna reef (Puerto Rico), 18 ft deep	04/01	0.28	–
Mission Beach, California	07/03	0.41	n/a

Table 2 (continued)

Location	Date	# Hector/10 ⁶ phage	# Paris/10 ⁶ phage
Mission Bay, California	07/03	0.18	n/a
<i>Terrestrial</i>			
Coastal Sage Scrub, California	05/02	–	–
Cultivated Land, Idaho	06/02	–	–
Desert Sand, New Mexico	07/02	–	–
Rhizosphere, Idaho	06/02	–	–
Sky Oaks Chapparal, California	02/02	–	–
Flynn Springs, California	07/03	0.3	n/a
Bahia Hotel, California	07/03	0.34	n/a
Vahala Sand Pit, California	07/03	0.085	n/a
Lake Murray, California	07/03	0.25	n/a
Total # of samples screened		64	57
Percent of positive samples		34%	12%

Both sequences were widespread in the environment, and the HECTOR sequence was found in all major biomes. HECTOR was found in 34% of the environmental samples tested and PARIS was found in 12% of these samples. All quantitative data are based on the number of positives per 10⁶ phage particles. n/a describes samples that were not assayed. Those samples marked with an asterisk represent purified, concentrated viral communities, treated with an additional density-dependent cesium chloride (CsCl) gradient step to ensure that the sequences were phage-encoded.

following up with targeted culturing efforts has been very successful with the marine bacterial clade SAR11 [26,27].

The presence of the HECTOR and PARIS sequences in extremely different global ecosystems demonstrates that phages must be moving around the biosphere in recent evolutionary time. There is no reason to believe that the HECTOR or PARIS sequences are unique in terms of their widespread distributions. Among the 116 sequences obtained with the degenerate primers, seven were recovered from two or more different samples and four of these sequences were recovered from multiple biomes (Fig. 1). In support of the argument of widely distributed phage sequences, we have recovered identical fragments of up to 405 bp of phage-encoded DNA in uncultured phage libraries from different sites and environments [unpublished results]. Hendrix and colleagues have observed an identical ~400 bp DNA fragment in two phage genomes, one of which was isolated in Hong Kong and the other in Pittsburgh [Hendrix, personal communication]. In addition, Short and Suttle [28] recovered nearly identical algal virus DNA polymerase sequences from marine samples from British Columbia and Antarctica. Together, these data suggest that there are many other widely distributed viral sequences on the planet, implying that the total phage genomic pool is relatively small.

The HECTOR DNA polymerase sequences found in the different environments were usually identical and never differed by more than 3 bp over a 533-bp fragment. Since we do not know whether the phages themselves or just this specific fragment of phage DNA sequence is moving between environments (e.g., as suggested by the mosaic model [29]), it is impossible to state at this time whether the bacterial host populations are as widespread as the phage sequence. Alternatively, these phages may have a broader host

range than is currently recognized for the T7-like Podophages. Either way, we do know that this phage sequence, or these phages, must be invasive. That is, the phages must be able to enter an ecosystem and establish themselves despite the fact that these environments contain native microbial communities with which they must compete. Recently, it has been shown that phage populations isolated from freshwater, sediment, and many marine locations can successfully replicate when added to seawater from another location, indicating that the ability of phage to move between environments may be a general phenomenon [Sano et al., in review].

The data presented here show that phage-encoded sequences are moving between biomes in recent evolutionary time. Genetic information could be carried by phages between environments and passed on to new hosts via transduction [30–33]. Therefore, movement of phages between environments may be responsible for the large amount of lateral gene transfer observed amongst microbes in different environments.

Acknowledgements

The authors thank Matthew Church, Hugh Ducklow, Matthew Erickson, David Kline, David Lipson, and Matthew Sullivan for collecting samples, as well as Ian Hewson from Jed Fuhrman's lab and Danielle Winget from Eric Wommack's lab for verifying the HECTOR PCR results. We also thank Anca Segall, Stan Maloy, Moselio Schaechter, and John Paul for helpful discussions. Funding was provided by the NSF GRANT OCE01-37748, NSF DEB 03-16518, the SDSU College of Sciences, and a Grant-in-Aid from the SDSU Foundation. Mya Breitbart is funded through a STAR fellowship from the Environmental Protection Agency.

References

- [1] Ackermann, H.W. (2001) Frequency of morphological phage descriptions in the year 2000. *Arch. Virol.* 146, 843–857.
- [2] Wommack, K. and Colwell, R. (2000) Virioplankton: viruses in aquatic ecosystems. *Microbiol. Mol. Biol. Rev.* 64, 69–114.
- [3] Borsheim, K.Y. (1993) Native marine bacteriophage. *FEMS Microbiol. Ecol.* 102, 141–159.
- [4] 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.
- [5] Wommack, K.E., Ravel, J., Hill, R.T., Chun, J.S. and Colwell, R.R. (1999) Population dynamics of Chesapeake Bay virioplankton: total community analysis by pulsed-field gel electrophoresis. *Appl. Environ. Microbiol.* 65, 231–240.
- [6] Steward, G., Montiel, J. and Azam, F. (2000) Genome size distributions indicate variability and similarities among marine viral assemblages from diverse environments. *Limnol. Oceanogr.* 45, 1697–1706.
- [7] In: *Virus Taxonomy: Sixth Report of the International Committee on Taxonomy of Viruses* (Murphy, F.A., Fauquet, C.M., Bishop, D.H.L., Ghabrial, S.A., Jarvis, A.W., Martelli, G.P., Mayo, M.A. and Summers, M.D., Eds.), Vol. Suppl. 10. Springer, New York.
- [8] Rohwer, F. and Edwards, R. (2002) The phage proteomic tree: a genome based taxonomy for phage. *J. Bacteriol.* 184, 4529–4535.
- [9] Noble, R.T. and Fuhrman, J.A. (1998) Use of SYBR Green I for rapid epifluorescence counts of marine viruses and bacteria. *Aquat. Microb. Ecol.* 14, 113–118.
- [10] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Edn. Cold Spring Harbor Laboratory Press.
- [11] Rohwer, F., Breitbart, M., Jara, J., Azam, F. and Knowlton, N. (2001) Diversity of bacteria associated with the Caribbean coral *Montastraea franksi*. *Coral. Reefs* 20, 85–95.
- [12] Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.
- [13] Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402.
- [14] Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. (1997) The Clustal X Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 24, 4876–4882.
- [15] Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680.
- [16] Rozen, S. and Skaltsky, H. (2000) In: *Bioinformatics Methods and Protocols: Methods in Molecular Biology* (Krawetz, S. and Misener, S., Eds.), pp. 365–386. Humana Press, New Jersey.
- [17] Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31, 3406–3415.
- [18] Whitman, W., Coleman, D. and Wiebe, W. (1998) Prokaryotes: the unseen majority. *Proc. Natl. Acad. Sci. USA* 95, 6578–6583.
- [19] Yin, X., Zeph, L. and Stotzky, G. (1997) A simple method for enumerating bacteriophages in soil. *Can. J. Microbiol.* 43, 461–466.
- [20] Bergh, O., Borsheim, K.Y., Bratbak, G. and Heldal, M. (1989) High abundance of viruses found in aquatic environments. *Nature* 340, 467–468.
- [21] Danovaro, R., Dell'Anno, A., Trucco, A., Serresi, M. and Vanucci, S. (2001) Determination of virus abundance in marine sediments. *Appl. Environ. Microbiol.* 67, 1384–1387.
- [22] Hewson, I., O'Neil, J.M., Fuhrman, J.A. and Dennison, W.C. (2001) Virus-like particle distribution and abundance in sediments and overlying waters along eutrophication gradients in two subtropical estuaries. *Limnol. Oceanogr.* 46, 1734–1746.
- [23] Hambly, E., Tetart, F., Desplats, C., Wilson, W.H., Krisch, H.M. and 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. USA* 98, 11411–11416.
- [24] Chen, F., Suttle, C.A. and 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.
- [25] Chen, F. and Suttle, C.A. (1996) Evolutionary relationships among large double-stranded DNA viruses that infect microalgae and other organisms as inferred from DNA polymerase genes. *Virology* 219, 170–178.
- [26] Connon, S. and Giovannoni, S. (2002) High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates. *Appl. Environ. Microbiol.* 68, 3878–3885.
- [27] Rappe, M., Connon, S., Vergin, K. and Giovannoni, S. (2002) Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. *Nature* 418, 630–633.
- [28] Short, S.M. and Suttle, C.A. (2002) Sequence analysis of marine virus communities reveals that groups of related algal viruses are widely distributed in nature. *Appl. Environ. Microbiol.* 68, 1290–1296.
- [29] Hendrix, R., Smith, M., Burns, R., Ford, M. and Hatfull, G. (1999) Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. *Proc. Natl. Acad. Sci. USA* 96, 2192–2197.
- [30] Goodman, A.E., Marshall, K.C. and Hermansson, M. (1994) Gene transfer among bacteria under conditions of nutrient depletion in simulated and natural aquatic environments. *FEMS Microbiol. Ecol.* 15, 55–60.
- [31] Jiang, S. and Paul, J.H. (1998) Gene transfer by transduction in the marine environment. *Appl. Environ. Microbiol.* 64, 2780.
- [32] Zeph, L.R., Onaga, M.A. and Stotzky, G. (1987) Transduction in soil. *Abstr. Ann. Meet. Am. Soc. Microbiol.* 87, 303.
- [33] Paul, J.H. (1999) Microbial gene transfer: an ecological perspective. *J. Mol. Microbiol. Biotechnol.* 1, 45–50.