

## ORIGINAL ARTICLE

# Viral and microbial community dynamics in four aquatic environments

Beltran Rodriguez-Brito<sup>1,2,9</sup>, LinLin Li<sup>2,9,10</sup>, Linda Wegley<sup>2</sup>, Mike Furlan<sup>2</sup>, Florent Angly<sup>1,2</sup>, Mya Breitbart<sup>2,11</sup>, John Buchanan<sup>3</sup>, Christelle Desnues<sup>2,12</sup>, Elizabeth Dinsdale<sup>2</sup>, Robert Edwards<sup>1,2</sup>, Ben Felts<sup>4</sup>, Matthew Haynes<sup>2</sup>, Hong Liu<sup>2</sup>, David Lipson<sup>2</sup>, Joseph Mahaffy<sup>4</sup>, Anna Belen Martin-Cuadrado<sup>5</sup>, Alex Mira<sup>5,13</sup>, Jim Nulton<sup>4</sup>, Lejla Pašić<sup>5,6</sup>, Steve Rayhawk<sup>1,2</sup>, Jennifer Rodriguez-Mueller<sup>2</sup>, Francisco Rodriguez-Valera<sup>5</sup>, Peter Salamon<sup>4</sup>, Shailaja Srinagesh<sup>2</sup>, Tron Frede Thingstad<sup>8</sup>, Tuong Tran<sup>2</sup>, Rebecca Vega Thurber<sup>2,14</sup>, Dana Willner<sup>2</sup>, Merry Youle<sup>7</sup> and Forest Rohwer<sup>2</sup>

<sup>1</sup>Computational Science Research Center, San Diego State University, San Diego, CA, USA; <sup>2</sup>Department of Biology, San Diego State University, San Diego, CA, USA; <sup>3</sup>Aqua Bounty Technologies, San Diego, CA, USA; <sup>4</sup>Department of Mathematics, San Diego State University, San Diego, CA, USA; <sup>5</sup>Evolutionary Genomics Group, División de Microbiología, Universidad Miguel Hernández, San Juan de Alicante, Spain; <sup>6</sup>Department of Biology, Biotechnical Faculty, University of Ljubljana, Ljubljana, Slovenia; <sup>7</sup>Rainbow Rock, Ocean View, HI, USA and <sup>8</sup>Department of Biology, University of Bergen, Bergen, Norway

**The species composition and metabolic potential of microbial and viral communities are predictable and stable for most ecosystems. This apparent stability contradicts theoretical models as well as the viral–microbial dynamics observed in simple ecosystems, both of which show Kill-the-Winner behavior causing cycling of the dominant taxa. Microbial and viral metagenomes were obtained from four human-controlled aquatic environments at various time points separated by one day to >1 year. These environments were maintained within narrow geochemical bounds and had characteristic species composition and metabolic potentials at all time points. However, underlying this stability were rapid changes at the fine-grained level of viral genotypes and microbial strains. These results suggest a model wherein functionally redundant microbial and viral taxa are cycling at the level of viral genotypes and virus-sensitive microbial strains. Microbial taxa, viral taxa, and metabolic function persist over time in stable ecosystems and both communities fluctuate in a Kill-the-Winner manner at the level of viral genotypes and microbial strains.**

*The ISME Journal* (2010) 4, 739–751; doi:10.1038/ismej.2010.1; published online 11 February 2010

**Subject Category:** microbial population and community ecology

**Keywords:** viruses; community dynamics; Kill-the-Winner; metagenomics; aquatic environments

## Introduction

Viruses are ubiquitous in all ecosystems, highly abundant (Noble and Fuhrman, 1997, 1998, 2000),

and very diverse (Breitbart *et al.*, 2002). Archaea and Bacteria are the most abundant and diverse cellular organisms on the planet (Whitman *et al.*, 1998) and are the main prey of viruses. Microbial and viral communities have significant functions in geochemical cycling and food web dynamics (Sigeo, 2005).

Recently, it has become possible to look beyond the geochemical functions of viruses and microbes to determine whether the ecological laws that originated from the study of macroorganisms apply also to the microbial world (Horner-Devine *et al.*, 2004). Exploration of such questions had been previously hindered by the nonculturability of the majority of viruses and microbes combined with the difficulties in identifying them at the level of both species and strain. Metagenomics now provides a culture-independent method for characterizing the viral and microbial communities present in an ecosystem, including the metabolic capabilities

Correspondence: B Rodriguez-Brito, Computational Science Research Center, San Diego State University, 5500 Campanile Drive, San Diego, CA 92182-1245, USA.

E-mail: beltran.rodriguezbrito@gmail.com

<sup>9</sup>These authors contributed equally to this work.

<sup>10</sup>Current address: Delwart Lab, Blood System Research Institute, San Francisco, CA, USA.

<sup>11</sup>Current address: College of Marine Science, University of South Florida, FL, USA.

<sup>12</sup>Current address: URMITE, UMR CNRS-IRD 6236, Université de la Méditerranée, Marseille, France.

<sup>13</sup>Current address: Center for Advanced Research in Public Health, CSISP, Area de Genómica y Salud, Valencia, Spain.

<sup>14</sup>Current address: Biology Department, Florida International University, Miami, Florida, USA.

Received 15 September 2009; revised 22 December 2009; accepted 23 December 2009; published online 11 February 2010

encoded in their genomes and the halo of variation within each species (Breitbart *et al.*, 2002, 2003, 2004; Tyson *et al.*, 2004; Venter *et al.*, 2004; Breitbart and Rohwer, 2005; Tringe *et al.*, 2005; Angly *et al.*, 2006; Edwards *et al.*, 2006; Legault *et al.*, 2006; Cuadros-Orellana *et al.*, 2007; Desnues *et al.*, 2008; Dinsdale *et al.*, 2008a,b).

The study of predator–prey interactions is still underdeveloped in microbial ecology (Taylor, 1984; Hoffmann *et al.*, 2007). Macroecological models of these interactions predict a repetitive cycle in which an increase in prey population leads to an increase in the predator population that, in turn, decreases the prey population, thus causing its own subsequent decline. The observation of large numbers of microbes and their viral predators in the ocean led Thingstad (1997, 2000) to propose a Lotka-Volterra-type model for fluctuations of their populations in aquatic environments. This model, colloquially known as Kill-the-Winner, predicts that viruses will rapidly and drastically reduce the population of the most abundant microbial species, thus preventing the best microbial competitors from building up a high biomass. Kill-the-Winner is the current working paradigm for microbial–viral community dynamics. In contrast to this model, multiple lines of evidence indicate that the taxonomic composition and metabolic capabilities of microbial communities are relatively constant within a defined environment (Garland and Mills, 1991; Zehr *et al.*, 2003; Fuhrman *et al.*, 2006).

In support of the Kill-the-Winner model, numerous studies have reported viral–host dynamics characterized by dramatic changes in the relative concentrations of predators and prey (Wommack *et al.*, 1999; Fuhrman and Schwalbach, 2003; Harcombe and Bull, 2005). Most such studies have focused on artificial experimental systems that contain only one, or at most two, virus/host pairs, with a few recent reports exploring natural environments (Chen *et al.*, 2009; Short and Short, 2009; Winget and Wommack, 2009). However, to the best of our knowledge, no one has previously studied the dynamics of these interactions over time in complex communities at a depth comparable to that provided by this study.

In this study, we used metagenomics to monitor the viral and microbial communities in four distinctive, stable, human-controlled aquatic environments over time. The study environments included a freshwater aquaculture system and three solar saltern environments of differing salinity. Populations within the communities were monitored at both the coarse-grained level of species and the fine-grained level of strains. On the basis of our results, we propose a model for these stable aquatic environments that is consistent with a Kill-the-Winner cycling of predator and prey populations and in which sensitive microbial strains are killed off by viral predators and are replaced by phage-resistant strains of the same species.

## Materials and methods

### Water samples

Freshwater samples (80 l volume) were collected from two sites at the aquaculture facility of Kent SeaTech Corporation (Salton Sea, CA, USA). At the time of sampling, hybrid striped bass were being raised in the main ponds. The Prebead pond (FW2) received effluent from 60 nursery ponds and the Tilapia Channel (FW1) received the effluent from 96 production ponds. These sampling sites integrate the microbial and viral communities from either the nursery or the production ponds, thus avoiding pond-to-pond variation. Water samples were collected in November 2005, April 2006, and August 2006.

Water samples from the solar saltern of South Bay Salt Works (Chula Vista, CA, USA) were collected from three ponds with different salinities: low (6–8%), medium (12–14%), and high (27–30%). Salinity was measured with a hand refractometer. The time between samplings ranged from 1 day to ~1.33 years (Table 1). Sample volumes were 60, 20, and 5–6 l for the low, medium, and high salinity ponds, respectively.

To determine the number of microbes (Bacteria and Archaea) and virus-like particles (VLPs) present (Supplementary Table S6), samples from each environment were fixed in 2% paraformaldehyde, filtered onto a 0.02- $\mu\text{m}$  pore size Anodisc membrane filter (Whatman, Maidstone, Kent, England), stained with SYBR Gold (Molecular Probes, Inc., Eugene, Oregon, USA), and counted by epifluorescent microscopy (Noble and Fuhrman, 1998). Microbial generation time, viral production rate, and rate of cell lysis (Supplementary Table S6) were measured according to the methods of Noble and Fuhrman (2000).

### Preparation of DNA

Viral and microbial fractions were isolated from each water sample by passage through a 0.2- $\mu\text{m}$  tangential flow filter (TFF, Millipore, Westborough, MA, USA). The filtrate and the retentate were collected in separate tanks. The virus-containing filtrate was concentrated to a final volume of ~500 ml using a 100-kDa TFF filter. Polyethylene glycol (PEG 8000) was added to a final concentration of 10% (w/v) and the mixture was incubated for 12–18 h at 4 °C. The VLPs were pelleted by centrifuging for 2 h at 22 000 rpm in an SW41Ti swinging bucket rotor (RCF  $\text{avg} = 53\,000 \times g$ ). The VLP pellet was resuspended in filtered (100 kDa) saltern water to a final volume of ~50 ml. The virus concentrate was loaded onto a cesium chloride density gradient, centrifuged, and the 1.5 g ml<sup>-1</sup> fraction was collected. At each step, recovery of viral particles was verified by epifluorescence microscopy. The viral DNA was isolated by formamide lysis and cetyltrimethylammonium bromide extraction (Sambrook *et al.*, 1982).

The microbial fraction was collected from the 0.2- $\mu\text{m}$  TFF retentate by centrifugation at ~2000  $\times g$

**Table 1** Metagenomic libraries used in this study

Name	Date	# Reads	# Base pairs	Average read length (bp)	GC (%)
<i>Fresh-water viruses (tilapia channel)</i>					
A-V	November 2005	267 640	27 099 247	101.3	45.6
B-V	April 2006	63 978	6 420 157	100.4	46.2
C-V	August 2006	67 612	6 865 291	101.5	48.7
<i>Fresh-water microbes (tilapia channel)</i>					
A-M	November 2005	381 076	38 423 159	100.8	43.8
B-M	April 2006	63 978	6 420 157	100.4	44.0
C-M	August 2006	67 612	6 865 291	101.5	50.4
<i>Fresh-water viruses (prebead pond)</i>					
B2-V	April 2006	67 988	6 949 329	102.2	39.6
<i>Fresh-water microbes (prebead pond)</i>					
B2-M	April 2006	44 094	4 384 895	99.4	44.9
<i>Low salinity viruses</i>					
D-V	1 July 2004	268 534	27 785 763	103.5	46.5
E-V	10 November 2005	110 511	11 421 737	103.4	41.2
F-V	28 November 2005	62 685	6 448 398	102.9	47.7
<i>Low salinity microbes</i>					
D-M	1 July 2004	268 206	25 012 316	93.3	54.6
E-M	10 November 2005	49 074	4 583 126	93.4	53.8
F-M	28 November 2005	34 296	3 419 010	99.7	42.6
<i>Medium salinity viruses</i>					
G-V	10 November 2005	39 578	3 973 591	100.4	47.6
H-V	11 November 2005	35 053	5 703 709	99.2	49.1
I-V	16 November 2005	58 735	3 477 534	97.1	46.5
J-V	22 November 2005	55 903	5 578 541	99.8	48.5
K-V	28 November 2005	39 943	3 971 718	99.4	48.7
<i>Medium salinity microbes</i>					
G-M	10 November 2005	38 929	3 867 026	99.3	55.3
H-M	11 November 2005	23 261	2 299 980	99.9	60.9
I-M	16 November 2005	120 987	11 746 041	97.1	61.4
K-M	28 November 2005	8062	697 933	86.6	56.2
<i>High salinity viruses</i>					
L-V	16 November 2005	154 167	15 225 789	98.8	53.0
M-V	28 November 2005	4 645	460 564	99.2	47.7
N-V	7 December 2005	47 587	4 813 217	101.2	49.0
O-V	20 December 2005	303 274	28 008 560	92.4	53.9
<i>High salinity microbes</i>					
P-M	1 July 2004	582 681	58 877 500	101.1	63.6
M-M	28 November 2005	35 446	3 675 849	103.7	61.4

This table defines the time-point abbreviations used in the figures and tables.

for 10 min. Microbial DNA was extracted using the Ultra Clean Soil DNA Kit (Mo Bio Laboratories, CA, USA).

The microbial and viral DNA samples were amplified using the strand-displacement  $\Phi$ 29 DNA polymerase (GenomiPhi Amersham Biosciences, NJ, USA). The resulting metagenomic DNA was pyrosequenced (GS20 technology; 454 Life Sciences, CT, USA). Identical reads (that is reads with the same beginning, end, and base sequence) are a known artifact of the GS20. As the probability of two reads being identical is so small for the size of the libraries used in this study, any identical reads were assumed to be artifacts and were excluded from further analysis (<http://scums.sdsu.edu/>). Details of the

resultant viral and microbial libraries, including nomenclature, sampling date, number of reads, average read length, and average GC content, are shown in Table 1.

#### *Metagenome to metagenome comparisons*

The sequence similarities of all metagenomes were analyzed by pairwise comparisons using tBLASTx with a cut-off value of  $E < 0.001$  (Supplementary Table S1). The data shown for each pair of metagenomes is the average of the percentage of shared reads determined in each direction (that is metagenome X versus metagenome Y and metagenome Y versus metagenome X).

### Phage and microbial taxonomy

Sequences from each virome were compared with a phage genome database comprised of the complete genome sequences of 510 phages and prophages taken mainly from the Universal Virus Database of the International Committee on Taxonomy of Viruses (<http://www.ncbi.nlm.nih.gov/ICTVdb/>) using tBLASTx (Supplementary Table S4-B). The best similarities ( $E < 0.001$ ) were mapped onto an updated Phage Proteomic Tree containing 510 genomes (<http://www-rohan.sdsu.edu/~brodrigu/PPT/>). This tree was constructed using a previously demonstrated method (Rohwer and Edwards, 2002), but including an updated set of phage and prophage genomes. Phylogenetic tree graphics were created using the PHY-FI tool (Fredslund, 2006). Sequences from each virome were also compared with viral genomes in the NCBI complete genome database using tBLASTx with a cut-off value of  $E < 0.001$ . Taxonomies were assigned based on best similarities.

BLASTn with  $E < 10^{-5}$  was used to compare the sequences from each microbiome to the bacterial and archaeal 16S ribosomal RNA database of the Greengenes project (<http://greengenes.lbl.gov/cgi-bin/nph-index.cgi>) (DeSantis *et al.*, 2003, 2006) (Supplementary Table S4-A). Sequences from each microbiome were also compared with genomes in the NCBI complete genome database using tBLASTx with a cut-off value of  $E < 0.001$ . Taxonomies were assigned based on best similarities.

### Microbial metabolic potential

The metagenomes were compared with the NCBI nr/nt database using BLASTx to identify similar proteins in the SEED database (<http://theseed.uchicago.edu/FIG/index.cgi> and <http://seed.sdsu.edu/FIG/index.cgi>). The metabolic potential of each microbiome was determined by assigning functional annotations to its metagenome sequences and then assigning the sequences to subsystems. Pairwise comparisons between the microbiome subsystem compositions were made using XIPE (Rodríguez-Brito *et al.*, 2006), a nonparametric, difference of medians analysis method (Supplementary Table S5).

### PHACCS and MaxiΦ predictions of genotypic dynamics

MaxiΦ was originally published in Angly *et al.* (2006) and determines the  $\beta$ -diversity between the two samples. MaxiΦ is based on PHACCS, an  $\alpha$ -diversity (within a sample) predictor. In PHACCS, a viral genotype (Breitbart *et al.*, 2002) is defined by the *in silico* assembly of metagenomic shotgun sequences under assembly conditions sufficiently stringent to discriminate between even closely related phages, such as the T3 and T7 coliphages (that is minimum overlap length of 35 bp with 98% sequence identity). These assemblies produce both contigs and singletons (unassembled reads), and it is an assumption of the method that each contig

originated from one particular viral genotype. In environments with high biodiversity, very few sequences will assemble. From the contig spectrum (that is the histogram of the contig count for singletons and for contigs composed of various numbers of reads), the number of genotypes present in each virome is estimated using a modified Lander–Waterman algorithm. To help automate this process, a program called CIRCONSPECT (<http://biome.sdsu.edu/circonspect/>) performs the assemblies and then PHACCS (<http://biome.sdsu.edu/phaccs/>) minimizes the Lander–Waterman algorithm. The output is a prediction of the community structure in terms of diversity, evenness, and richness (Breitbart *et al.*, 2002, 2004; Angly *et al.*, 2005, 2006). This is the  $\alpha$ -diversity of the sample.

MaxiΦ uses the same approach as PHACCS, but here to measure the  $\beta$ -diversity (that is inter-sample diversity). In this case, sequences from two different viromes are assembled together to identify those sequences that are shared by both samples. These are called cross-contigs. The underlying assumption is that the formation of more cross-contigs between a pair of viromes indicates that a greater abundance of viral genotypes are shared between the two communities. MaxiΦ itself is a Monte Carlo simulation used to model the likelihood that the observed cross-contig spectrum represents various percentages of genotypes shared or permuted. Details on MaxiΦ are presented in Angly *et al.* (2006).

### TaxiΦ

TaxiΦ is a new statistical analysis introduced in this manuscript to assess the similarities between two microbiomes by determining the percentage of known genomes shared between them and the amount of reshuffling of their relative abundances. This is essentially the same as MaxiΦ, only applied to taxa. It is described in detail in the Supplementary Methods.

## Results and discussion

The study sites included freshwater aquaculture ponds and a solar saltern system located in southern California, USA (Figure 1). Sampling locations and dates are shown in Table 1, along with the key characteristics of the 29 viral and microbial metagenomes, referred to respectively as viromes and microbiomes. The analyses of the 29 samples are presented below in two sections: (1) coarse-grained community dynamics and structure at the species level and (2) fine-grained community dynamics and structure at the level of viral genotypes and microbial strains. In this study, we defined species based on significant similarity (tBLASTx) to a complete genome sequence available in the NCBI complete genome database. Strains were distinguished based



**Figure 1** Environmental samples were collected from a solar saltern and a freshwater aquaculture facility, both located in California, USA. In the saltern, low, medium, and high salinity ponds were sampled. Satellite imagery: Google/Google Earth.

on assemblies for the viruses and on differences in *recA* for microbes.

*Coarse-grained taxonomic and metabolic analyses*  
*Comparisons between and within environments.*

Pairwise comparisons were made between all metagenomes using tBLASTx (Supplementary Table S1). This analysis showed greatest similarity between viromes or microbiomes from the same environment and less similarity between different environments. This supports the hypothesis that different environments have characteristic metagenomic signatures.

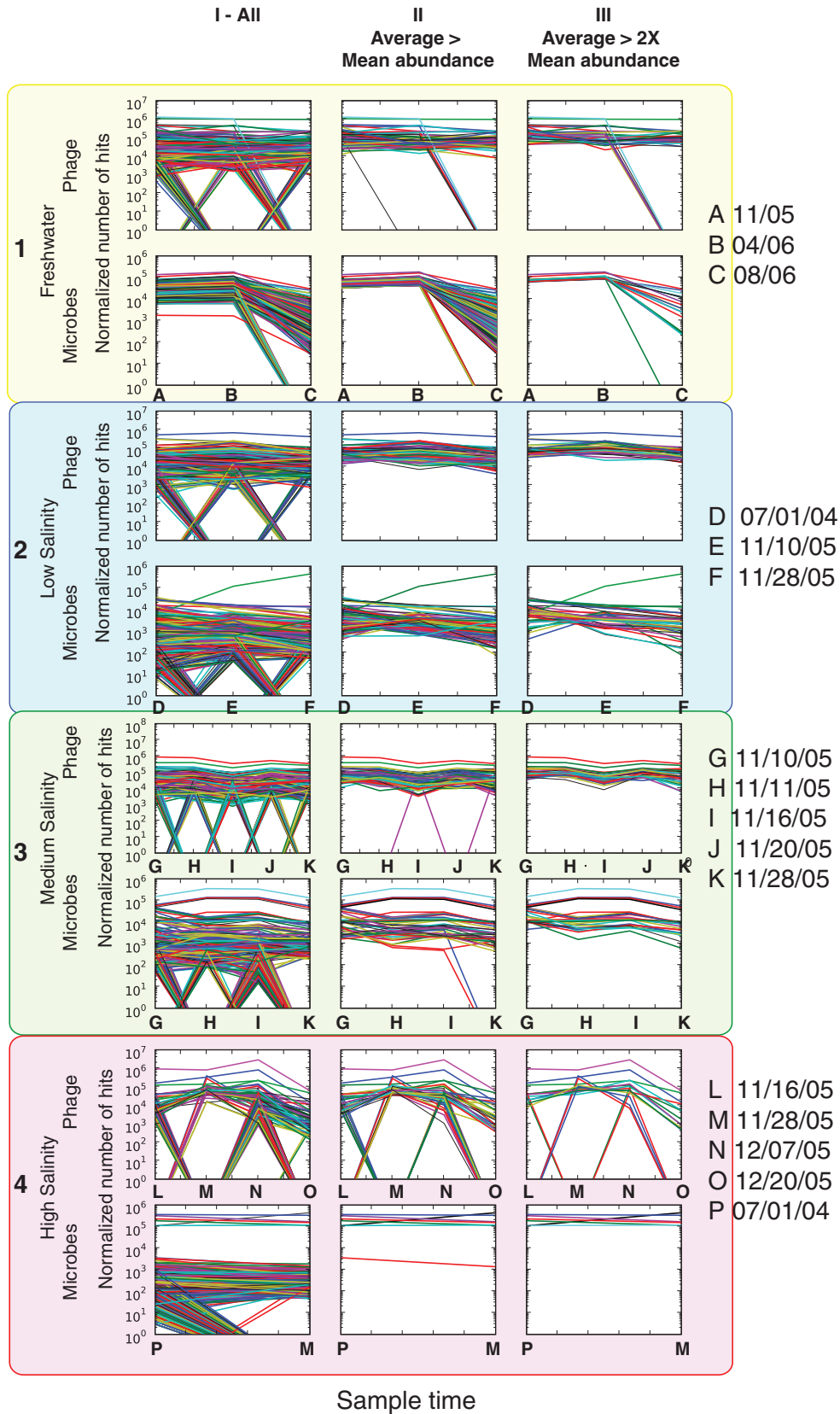
To investigate this hypothesis further, a more detailed survey of the 16S rDNA in the microbiomes was performed. Approximately 0.1% of the microbiome sequences had significant similarity ( $E < 0.001$ ) to 16S rDNA genes in the Ribosomal Database Project (Cole *et al.*, 2007). On the basis of these similarities, the microbial communities in the freshwater ponds exhibited typical freshwater bacterial groups as previously seen in other 16S rDNA surveys of freshwater environments (Arias *et al.*, 2004; Lindstrom *et al.*, 2005; Yannarell and

Triplett, 2005; Brie *et al.*, 2007). Corresponding analyses of microbiomes from the solar salterns showed the expected shift to an Archaea-dominated community in the high salinity pond (Rodriguez-Valera *et al.*, 1985; Benlloch *et al.*, 2002). Archaea represented 2.2%, 38%, and 54% of the significant similarities in the low, medium, and high salinity salterns, respectively. In all four environments, these 16S rDNA-based community profiles did not change over time, supporting the hypothesis that stable geochemistry leads to stable microbial communities (Supplementary Table S2).

Richer taxonomical information than available from 16S rDNA alone is obtained from metagenomes by comparing the sequences obtained against whole genome databases (Edwards *et al.*, 2006; Huson *et al.*, 2007). To take advantage of this, taxa were also identified using tBLASTx ( $E < 0.001$ ) to bacterial, archaeal, and viral genomes in the NCBI complete genome database (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=genome>). The percentage of sequences in each microbiome with significant similarities to known genomes ranged between 16% and 41% (Supplementary Table S4-A). The percentages for viromes ranged between 1.0% and 5.4% (Supplementary

Table S4-B). Consistent with the 16S rDNA-based survey, these analyses showed that the more abundant microbial and viral taxa within each environ-

ment persist over time (Figure 2). To better visualize the more abundant taxa, three graphs are presented for each environment: one including all taxa, one

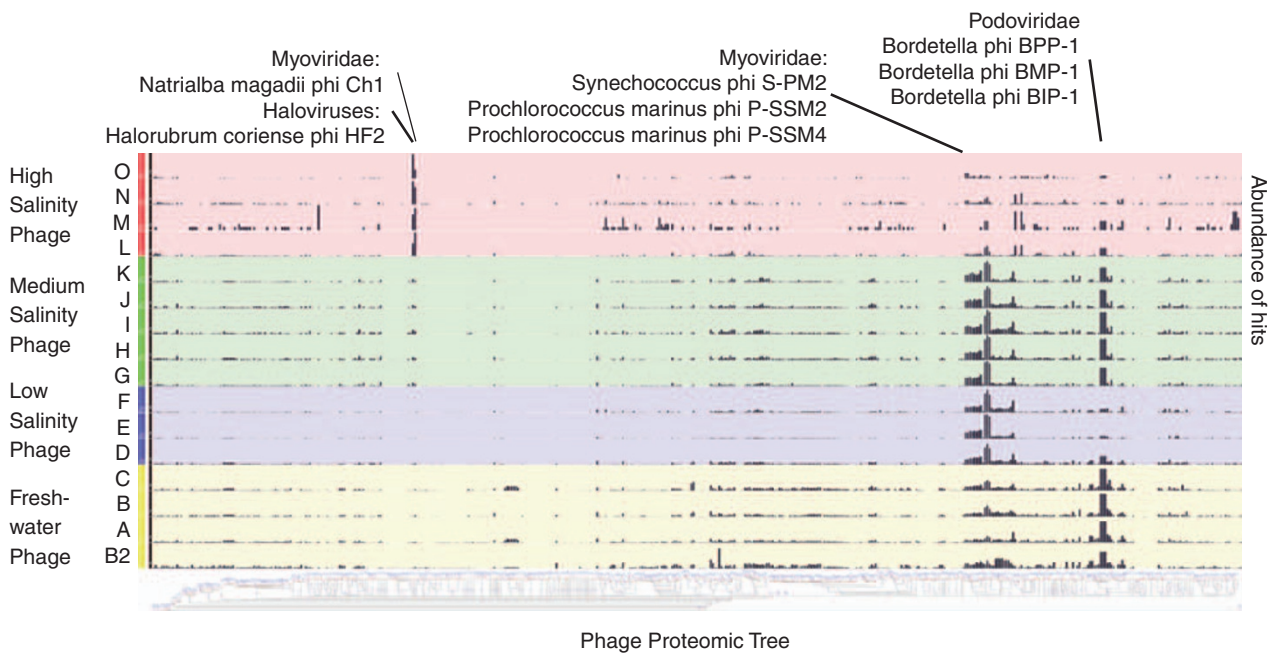


including those of at least average abundance, and one including only those of at least twice the average abundance. Although the more abundant taxa persist over time, many of the less abundant ones do not. These transient taxa may represent nonresident organisms that have been washed into the system or rare members of the community. There was some shuffling of the relative abundances of the most abundant taxa over time, but abundant taxa are persistent in all ecosystems.

The sequences from each virome were also compared with a database of 510 completely sequenced phage genomes using tBLASTx ( $E < 0.001$ ). Significant similarities were mapped to the Phage Proteomic Tree (Rohwer and Edwards, 2002; Edwards and Rohwer, 2005) (Figure 3). At this level of resolution, a unique phage taxonomic pattern characterizes each environment. Within an environment, the phage communities are very stable over time, that is, both the taxa present and their relative abundances are almost identical for every time sampled (Figure 3). The consistent patterns for each

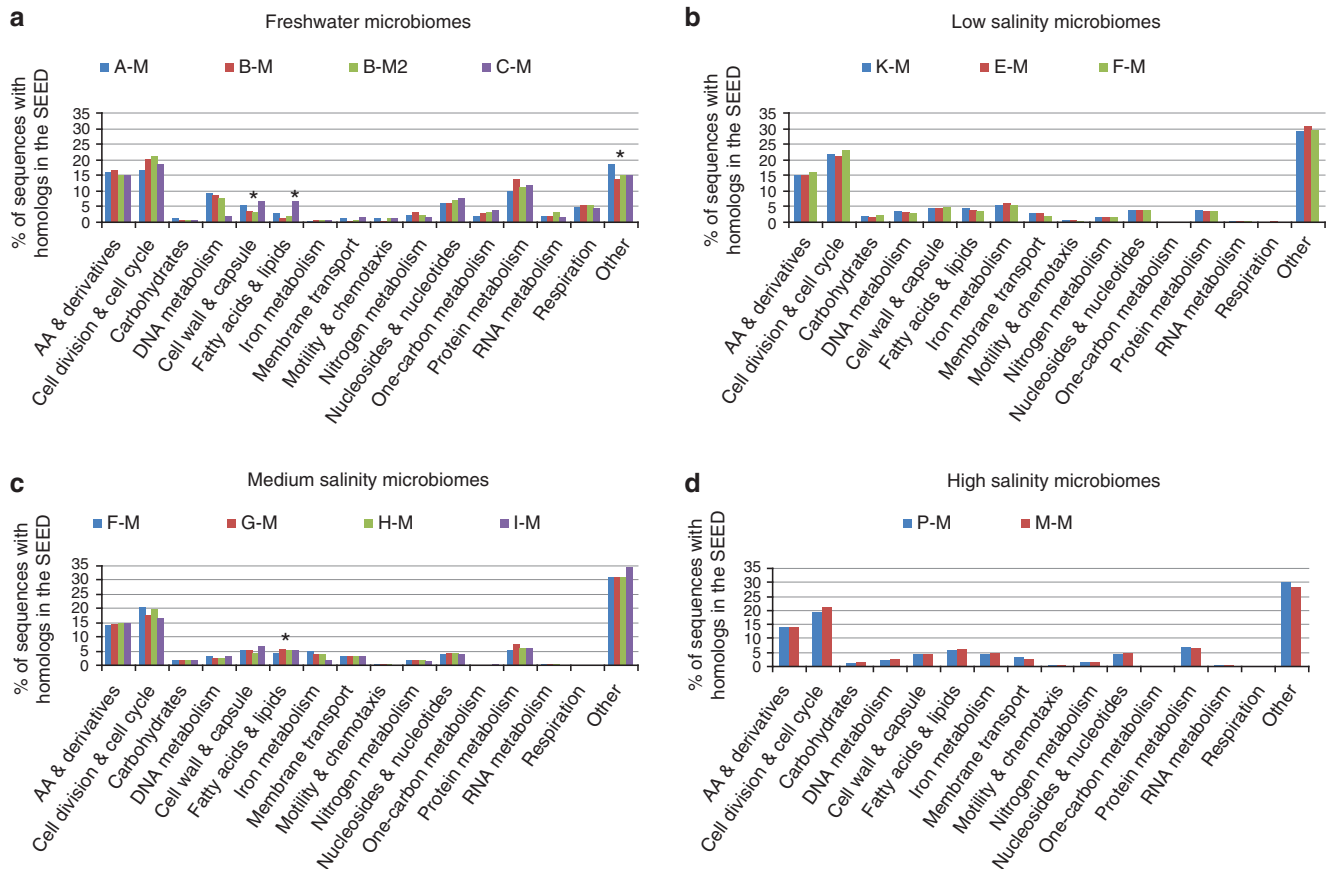
environment are quite remarkable for samples collected on different days by different people and sequenced at different facilities. The community stability is also evident in the lists of the 10 most abundant phage at each time point for each environment (Supplementary Table S3).

*Metabolic characterization of microbiomes.* The metabolic potential of the microbiomes was determined using the SEED database (Overbeek *et al.*, 2005; Aziz *et al.*, 2008). The SEED assigns genes to subsystems, each of which is composed of a group of functionally related proteins (for example the enzymes that make up a metabolic pathway, the structural proteins that form a cellular component such as a ribosome, or a class of proteins). Subsystem analysis has been previously used to show that environments have predictive metabolic profiles (Dinsdale *et al.*, 2008b). The analysis of the microbiomes showed that the metabolic potential of each environment was stable over time (Figure 4). Pairwise comparisons identified significant differences



**Figure 3** The relative abundance of various phages in each metagenome. Each horizontal row represents the metagenome from one sample; the metagenomes from each environment collected at different time points are grouped together and shown with a common background color. The 510 species nodes of the Phage Proteomic Tree (<http://www-rohan.sdsu.edu/~brodrigu/PPT/>) are arrayed along the X axis. The height of each bar is proportional to the percentage of sequences in that metagenome with tBLASTx similarities to the corresponding phage genome. Each environment shows a consistent distribution of phage species across time that is distinct from that of the other environments. Because UNIFRAC (Lozupone and Knight, 2005; Lozupone *et al.*, 2006) reported no significant statistical differences ( $P < 0.01$ ) among the samples collected from different environments at various time points, we developed other methods (e.g., *MaxiPhi*, Figure 5) to reveal the variations.

**Figure 2** Time series for viral and microbial taxa in each environment. Each plotted line corresponds to one taxon and shows its relative abundance at each time point. From left to right, the three columns contain all the species found, those whose average abundance is greater than the average for all species, and those whose average abundance is greater than twice the average abundance. X axis: the sampling date (legend to the right of the plots). Y axis: normalized tBLASTx similarities to each genome, calculated by dividing the number of similarities by the product of the genome size (Mbp) multiplied times the number of gigareads in that metagenomic library. (As zero cannot be represented on the logarithmic scale of the Y axis, the Y axis was truncated at  $10^0$  and  $10^{-1}$  was used to represent zero values. As a result, plots for taxa with zero similarities for a time point end at  $10^0$  between time points).



**Figure 4** Metabolic profiles for each environment. Panel **a** = freshwater. Panel **b** = low salinity. Panel **c** = medium salinity. Panel **d** = high salinity. Sequences for each microbiome were assigned to metabolic subsystems using the SEED database. The 'Other' category includes (1) cofactors, vitamins, prosthetic groups, pigments; (2) photosynthesis; (3) regulation and cell signaling; (4) secondary metabolism; (5) virulence; (6) stress response; (7) metabolism of aromatic compounds; and (8) phosphorous metabolism. Asterisks: subsystems for which pairwise comparisons identified significant differences ( $P < 0.01$ ) between microbiomes collected on different dates.

( $P < 0.01$ ) over time in only a few subsystems in only two environments (asterisks, Figure 4).

**Conclusions from coarse-grained analyses.** Distinctive taxonomic and metabolic profiles were obtained for all four environments. Coarse-grained community structure, as determined by tBLASTx similarity to known genomes, was stable for both microbes and viruses in all four environments over time periods ranging from 1 day to  $>1$  year. Although there was some shuffling of ranks, the top microbial and viral taxa persisted. Similarly, the metabolic profiles of the microbiomes of each environment were stable over time with only a few subsystems showing changes. Combined, these results strongly support the hypothesis that stable geochemistry leads to stable microbial and viral taxa, as well as stable metabolic potential.

**Caveats.** The methods used in the coarse-grained analyses reflect only the currently available sequenced genomes. As all taxa are not represented in the genome databases, it is possible that some dominant but unsequenced organisms may have been omitted from the lists of the most abundant

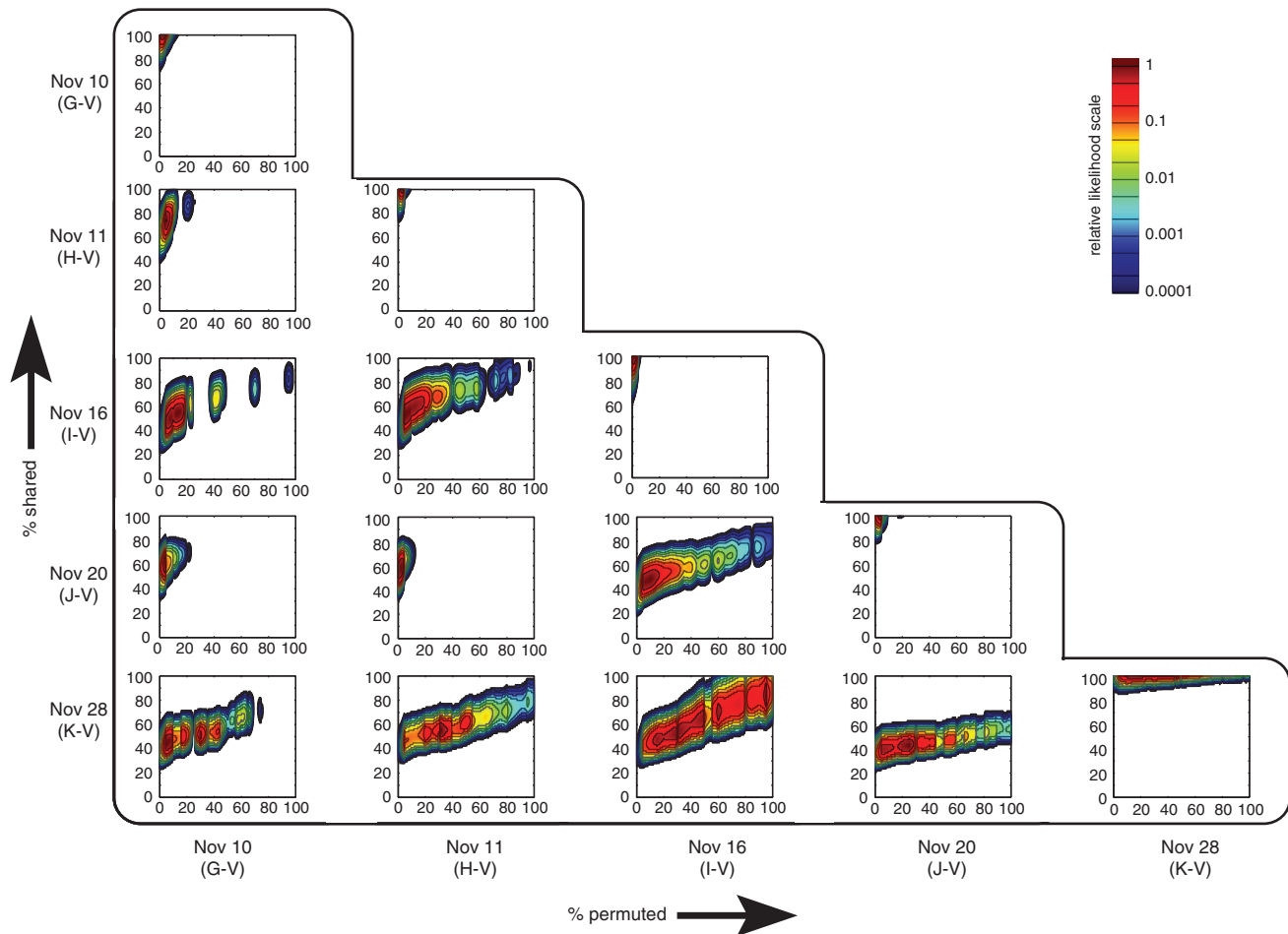
taxa. Being limited to only known genomes, any functional genomics study of microbes and viruses is open to alternative interpretations. However, Dinsdale *et al.* (2008b) showed characteristic metabolic profiles for nine biomes based on known genomes and metabolic characterizations from the SEED, whereas similar groupings were found using compositional metagenomic data (Willner *et al.*, 2009). This strongly suggests that the subset of known genomes are distinguished from the remainder of the population by our inability to identify the others, not by any fundamental differences.

#### *Fine-grained analyses of microbial strains and viral genotypes*

The results presented above support the hypothesis that geochemistry drives the species composition and metabolic potential of these four aquatic communities. We next examine these communities over time at the fine-grained level of viral genotypes and microbial strains.

**Temporal dynamics of viral genotypes.** Fine-grained changes within the viral communities over





**Figure 5** Pairwise comparisons of viromes collected on different dates from the medium salinity environment. The similarity between each pair of viromes as modeled by *MaxiPhi* is represented by two percentages: the percentage of genotypes shared by the two viromes (Y axis) and the percentage of the most abundant genotypes whose ranks are permuted (X axis). (These two quantities vary independently between 0% and 100% and thus do not add together to 100%.) The colored areas on each graph show the relative likelihood (ranging between 1 and 0.0001) that the corresponding percentage of genotypes are shared and permuted between that pair of viromes. Comparison of each virome with itself (the five graphs on the diagonal) serve as controls and show the expected high percentages of genotypes shared and the low percentages of abundant genotypes permuted.

time in each environment were modeled using *MaxiPhi*. For this approach, a viral genotype (Breitbart *et al.*, 2002) is defined by the *in silico* assembly of metagenomic shotgun sequences under specific conditions (see Materials and methods). We modeled the similarities between each virome pair as the percentage of shared genotypes and the changes in the relative abundance of genotypes (that is the percent permuted) (Angly *et al.*, 2006).

For the medium salinity environment, the controls (each virome compared with itself) show the expected results: close to 100% of the genotypes are shared and the percentage of the abundant genotypes that are permuted is close to 0% (Figure 5). Pairwise comparisons between viromes from different sampling dates show significant variation even in samples collected just 1 day apart (Figure 5). Comparisons over longer time periods show greater differences. For example, samples collected 17 days apart (November 11 versus November 28) show

~40% genotypes shared and 15–60% of the top-ranked genotypes permuted. Similar results were obtained for viromes in the low and high salinity environments (Supplementary Figures S1 and S2).

The freshwater viral community exhibited even more dramatic changes (Supplementary Figure S3). Viromes collected 4 months apart (B-V versus C-V) were very different: <10% shared genotypes and an unresolved percentage of permuted genotypes ranging between 40% and 100%. In contrast, samples taken 5 months apart (A-V versus B-V) were quite similar: 80–100% shared genotypes and 20–30% of the most abundant genotypes permuted. This example shows that viral communities more closely related in time can differ more, and *vice versa*.

Combined, these results argue that viral genotypes were rapidly changing within all four communities even though the coarse-grained analyses (Figures 2 and 3) showed that viral taxa persisted over the same time periods.

*Temporal dynamics of microbial strains.* We developed a new method of analysis called *TaxiΦ* (Supplementary Methods) to study genomic variation of the microbial populations at the level of both species and strain. As different RecA proteins have been found to be associated with different strains of the same species (Zhou and Spratt, 1992; Parkinson *et al.*, 2009), variation in the populations of RecA proteins is used to indicate population changes at approximately the level of microbial strains. For analysis at the species level, we compare the known genomes present. In both methods, microbiomes are compared with respect to the percentage of the top 20 elements (genomes for coarse-level analysis and genes for fine-level analysis) that are shared between them and the percentage that are permuted.

For example, coarse-grained comparison of microbiomes from the medium salinity environment indicate that a high percentage of the most abundant species persist over time periods ranging from 1 to 18 days (Supplementary Figure S4). During these same intervals, their relative abundances are continually shifting. In contrast, the fine-grained plots for *recA* show a low percentage shared and a greater percentage permuted, thus indicating more rapid change at the fine-grained level. In summary, these results indicate that the most abundant microbial species persist but display shuffling of their relative abundances, while microbial strains come and go.

*Conclusions.* The *MaxiΦ* analyses of viromes from different time points show continuous variation of viruses and their relative abundances at the genotype level. Likewise, *recA*-based *TaxiΦ* analysis of the microbiomes showed a corresponding rapid change in the microbial strains present. These results are consistent with previous chemostat studies observing limited numbers of viral and microbial pairs (Bull *et al.*, 2006; Lennon and Martiny, 2008).

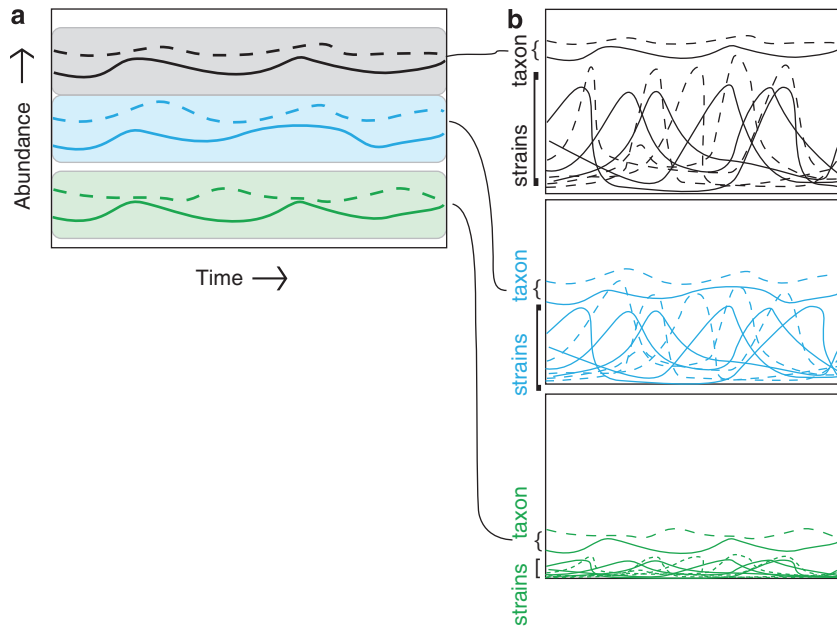
*Caveats.* Sequencing of these metagenomic samples was done using first generation pyrosequencing technology (that is GS20). Possible errors in the sequencing/base calling of homopolymers could lead to a spurious assembly or erroneous taxon assignment. However, other studies have reported good agreement between 16S rDNA sequences obtained from a pyrosequenced metagenomic library and a traditional 16S clone library (Edwards *et al.*, 2006) and likewise between the functional metabolic profiles obtained using pyrosequencing versus capillary sequencing (Turnbaugh *et al.*, 2006). Reproducibility of 454-based sequencing was verified by Stephan Schuster and Ed DeLong, who independently sequenced the same sample in their respective laboratories and found no important differences between the two data sets (personal communication). Likewise, Alejandro Muñoz and Jeff Gordon have resequenced multiple samples

and found no major differences in the predicted metabolic profile or taxa in the technical replicates (personal communication). Furthermore, the striking similarity between the different time points in this study strongly suggest that the sequencing reactions are not a major source of error. Therefore, we conclude that sequencing errors would create only minor variations in the results.

#### *Significance and model*

For this study, solar saltern and aquaculture pond environments were chosen because these systems are routinely monitored and maintained within tightly defined geochemical ranges and together they represent a wide range of environmental conditions (for example 0–30% salinity). The results reported here represent the first large-scale time series of microbial and viral community dynamics conducted at the DNA level over several environments. The metagenomic data were analyzed at two different levels: coarse grained at the species level and fine grained at the level of viral genotypes and microbial strains. Our analyses of these human-controlled ecosystems indicate that all four environments are biologically stable. First, each community has a characteristic profile of metabolic potential and that profile is strikingly stable over time (Figure 4). Second, the dominant microbial and viral taxa generally persist over time (Figures 2 and 3). If Kill-the-Winner dynamics were occurring at the species level, then temporal changes in the microbial and viral species would be expected, with the most abundant microbial taxa being sharply reduced or driven to local extinction by viral predation. This was not observed. Instead, we found an interplay between microbial species and viral predators that results in a reshuffling of the most successful species in each particular environment (Supplementary Tables S2 and S3). Viral genotypes were rapidly changing (Figure 5; Supplementary Figures S1–S3), thus suggesting that the microbial prey must also be changing. This was observed by the fine-grained *recA* analyses showing rapid change at the level of microbial strains. These conclusions are consistent with many other studies (Middelboe, 2000; Middelboe *et al.*, 2001; Zhong *et al.*, 2002; Bull *et al.*, 2006; Holmfeldt *et al.*, 2007; Stoddard *et al.*, 2007). Our conclusions are also supported by a recent study linking phage predation to the persistence of microbial diversity at the strain level (Rodríguez-Valera *et al.*, 2009).

Loci contributing to viral resistance are some of the most highly selected genes in microbial genomes (Tyson *et al.*, 2004; Venter *et al.*, 2004; Andersson and Banfield, 2008; Tyson and Banfield, 2008; Pasic *et al.*, 2009) and include targets for viral attachment (for example outer membrane proteins and O-antigen) (Cuadros-Orellana *et al.*, 2007), as well as cellular responses to viral invasion (for example CRISPRs) (Kunin *et al.*, 2008; Tyson and Banfield, 2008). Similarly, phage tails, which are important in



**Figure 6** Proposed model for taxonomically and metabolically stable ecosystems. Solid lines = microbes; dashed lines = viruses. (a) Microbial taxa and their co-occurring phage taxa coexist stably over time (Figures 2 and 3). (b) Three microbial taxa (*Black*, *Blue*, and *Green* spp.) and their phages are diagrammed separately. Underlying the relatively stable species populations is a dynamic cycling of microbial strains and viral genotypes (Figure 5 and Supplementary Figures S1–S4).

host recognition, also change rapidly (Liu *et al.*, 2002; Angly *et al.*, 2009). Rapid evolution of resistant microbial strains and their viral predators is part of the perpetual arms race referred to as Red Queen dynamics. These evolutionary changes are routinely observed in chemostat studies (Bohannan and Lenski, 2002; Lennon and Martiny, 2008; Middelboe *et al.*, 2009) and in environmental surveys (Vos *et al.*, 2009). Our results do not distinguish between variation due to selection acting on pre-existing diversity or *de novo* generation of mutants.

The highest level community model consistent with the great amount of data generated and analyzed in this study is one of metabolic and taxonomic stability, manifested as the stable coarse-grained structure observed for both the microbial and the viral populations. Underlying this stability are rapid population fluctuations at the fine-grained level of microbial strains and viral genotypes. In our proposed model, the abundances of both predator and prey taxa are relatively constant (Figure 6). Despite this stability, individual microbial strains increase markedly in abundance, encounter increased phage predation, and then decline—followed quickly by the decline of that particular phage genotype. This cycling of predator and prey populations is consistent with a Kill-the-Winner dynamic operating at the level of virus-sensitive microbial strains. Alternative interpretations for our data are possible, for example, attributing the observed changes to protist grazing, negative interactions between microbes, subtle shifts in environmental conditions, or stochastic processes. In

actuality, multiple factors are assuredly involved to varying extents. However, our data combined with the literature strongly suggest that viral predation is a major, and possibly the dominant, factor shaping microbial communities.

## Acknowledgements

We thank the Chula Vista Saltern Works and Kent Sea Tech for all their invaluable help and collaboration during this study. Funding for this study was received from the US National Science Foundation (DEB-BE04-21955), Gordon and Betty Moore Foundation, ARCS Fellowship, NSF Postdoctoral Fellowship #DBI-0511948, ATP/Kent Sea Tech and EMBO ASTF366-2007 short-term fellowship. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## References

- Andersson AF, Banfield JF. (2008). Virus population dynamics and acquired virus resistance in natural microbial communities. *Science* **320**: 1047–1050.
- Angly F, Youle M, Nosrat B, Srinagesh S, Rodriguez-Brito B, McNairnie P *et al.* (2009). Genomic analysis of multiple Roseophage SIO1 strains. *Environ Microb* **11**: 2863–2873.
- Angly F, Rodriguez-Brito B, Bangor D, McNairnie P, Breitbart M, Salamon P *et al.* (2005). PHACCS, an online tool for estimating the structure and diversity of uncultured viral communities using metagenomic information. *BMC Bioinformatics* **6**: 41.

- Angly FE, Felts B, Breitbart M, Salamon P, Edwards RA, Carlson C *et al.* (2006). The marine viromes of four oceanic regions. *PLoS Biol* **4**: e368.
- Arias CR, Welker TL, Shoemaker CA, Abernathy JW, Klesius PH. (2004). Genetic fingerprinting of Flavobacterium columnare isolates from cultured fish. *J Appl Microb* **97**: 421–428.
- Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA *et al.* (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* **9**: 75.
- Benlloch S, Lopez-Lopez A, Casamayor EO, Ovreas L, Goddard V, Daae FL *et al.* (2002). Prokaryotic genetic diversity throughout the salinity gradient of a coastal solar saltern. *Environ Microb* **4**: 349–360.
- Bohannon B, Lenski RE. (2002). Linking genetic change to community evolution: insights from studies of bacteria and bacteriophage. *Ecol Lett* **3**: 362–377.
- Breitbart M, Rohwer F. (2005). Method for discovering novel DNA viruses in blood using viral particle selection and shotgun sequencing. *Biotechniques* **39**: 729–736.
- Breitbart M, Hewson I, Felts B, Mahaffy JM, Nulton J, Salamon P *et al.* (2003). Metagenomic analyses of an uncultured viral community from human feces. *J Bacteriol* **185**: 6220–6223.
- Breitbart M, Felts B, Kelley S, Mahaffy JM, Nulton J, Salamon P *et al.* (2004). Diversity and population structure of a near-shore marine-sediment viral community. *Proc Biol Sci* **271**: 565–574.
- Breitbart M, Salamon P, Andresen B, Mahaffy JM, Segall AM, Mead D *et al.* (2002). Genomic analysis of uncultured marine viral communities. *Proc Natl Acad Sci USA* **99**: 14250–14255.
- Briece C, Moreira D, Lopez-Garcia P. (2007). Archaeal and bacterial community composition of sediment and plankton from a suboxic freshwater pond. *Res Microb* **158**: 213–227.
- Bull JJ, Millstein J, Orcutt J, Wichman HA. (2006). Evolutionary feedback mediated through population density, illustrated with viruses in chemostats. *Am Nat* **167**: E39–E51.
- Chen F, Wang K, Huang S, Cai H, Zhao M, Jiao N *et al.* (2009). Diverse and dynamic populations of cyanobacterial podoviruses in the Chesapeake Bay unveiled through DNA polymerase gene sequences. *Environ Microbiol* **11**: 2884–2892.
- Cole JR, Chai B, Farris RJ, Wang Q, Kulam-Syed-Mohideen AS, McGarrell DM *et al.* (2007). The ribosomal database project (RDP-II): introducing myRDP space and quality controlled public data. *Nucleic Acids Res* **35**: D169–D172.
- Cuadros-Orellana S, Martin-Cuadrado A-B, Legault BA, D'Auria G, Zhaxybayeva O, Papke RT *et al.* (2007). Genomic plasticity in prokaryotes: the case of the square haloarchaeon. *ISME J* **1**: 235–245.
- DeSantis TZ, Dubosarskiy I, Murray SR, Andersen GL. (2003). Comprehensive aligned sequence construction for automated design of effective probes (CASCADE-P) using 16S rDNA. *Bioinformatics* **19**: 1461–1468.
- DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K *et al.* (2006). Greengenes: a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *App Environ Microb* **72**: 5069–5072.
- Desnues C, Rodríguez-Brito B, Rayhawk S, Kelley S, Tran T, Haynes M *et al.* (2008). Biodiversity and biogeography of phages in modern stromatolites and thrombolites. *Nature* **452**: 340–343.
- Dinsdale EA, Pantos O, Smriga S, Edwards RA, Angly F, Wegley L *et al.* (2008a). Microbial ecology of four coral atolls in the Northern Line Islands. *PLoS ONE* **3**: e1584.
- Dinsdale EA, Edwards RA, Hall D, Angly F, Breitbart M, Brulc JM *et al.* (2008b). Functional metagenomic profiling of nine biomes. *Nature* **452**: 629–632.
- Edwards RA, Rohwer F. (2005). Viral metagenomics. *Nat Rev Microb* **3**: 504–510.
- Edwards RA, Rodríguez-Brito B, Wegley L, Haynes M, Breitbart M, Peterson DM *et al.* (2006). Using pyrosequencing to shed light on deep mine microbial ecology. *BMC Genomics* **7**: 57.
- Fredslund J. (2006). Phi-Fi: fast and easy online creation and manipulation of phylogeny color figures. *BMC Bioinformatics* **7**: 315.
- Fuhrman JA, Schwalbach MS. (2003). Viral influence on aquatic bacterial communities. *Biol Bull* **204**: 192–195.
- Fuhrman JA, Hewson I, Schwalbach MS, Steele JA, Brown V, Naeem S. (2006). Annually reoccurring bacterial communities are predictable from ocean conditions. *Proc Natl Acad Sci USA* **103**: 13104–13109.
- Garland JL, Mills AL. (1991). Classification and characterization of heterotrophic microbial communities on the basis of patterns of community-level sole-carbon-source utilization. *App Environ Microb* **57**: 2351–2359.
- Harcombe WR, Bull JJ. (2005). Impact of phages on two-species bacterial communities. *App Environ Microb* **71**: 5254–5259.
- Hoffmann KH, Rodríguez-Brito B, Breitbart M, Bangor D, Angly F, Felts B *et al.* (2007). Power law rank-abundance models for marine phage communities. *FEMS Microb Lett* **273**: 224–228.
- Holmfeldt K, Middelboe M, Nybroe O, Riemann L. (2007). Large variabilities in host strain susceptibility and phage host range govern interactions between lytic marine phages and their flavobacterium hosts. *App Environ Microb* **73**: 6730–6739.
- Horner-Devine MC, Lage M, Hughes JB, Bohannon B. (2004). A taxa-area relationship for bacteria. *Nature* **432**: 750–753.
- Huson DH, Auch AF, Qi J, Schuster SC. (2007). MEGAN analysis of metagenomic data. *Genome Res* **17**: 377–386.
- Kunin V, He S, Warnecke F, Peterson SB, Garcia Martin H, Haynes M *et al.* (2008). A bacterial metapopulation adapts locally to phage predation despite global dispersal. *Genome Res* **18**: 293–297.
- Legault BA, Lopez-Lopez A, Alba-Casado JC, Doolittle WF, Bolhuis H, Rodríguez-Valera F *et al.* (2006). Environmental genomics of 'Haloquadratum walbyi' in a saltern crystallizer indicates a large pool of accessory genes in an otherwise coherent species. *BMC Genomics* **7**: 171.
- Lennon JT, Martiny JBH. (2008). Rapid evolution buffers ecosystem impacts of viruses in a microbial food web. *Ecol Lett* **11**: 1178–1188.
- Lindstrom ES, Kamst-Van Agterveld MP, Zwart G. (2005). Distribution of typical freshwater bacterial groups is associated with pH, temperature, and lake water retention time. *App Environ Microb* **71**: 8201–8206.
- Liu M, Deora R, Doulatov SR, Gingery M, Eiserling FA, Preston A *et al.* (2002). Reverse transcriptase-mediated tropism switching in *Bordetella* bacteriophage. *Science* **295**: 2091–2094.
- Lozupone C, Knight R. (2005). UniFrac: a new phylogenetic method for comparing microbial communities. *App Environ Microbiol* **71**: 8228–8235.

- Lozupone C, Hamady M, Knight R. (2006). UniFrac—an online tool for comparing microbial community diversity in a phylogenetic context. *BMC Bioinformatics* **7**: 371.
- Middelboe M. (2000). Bacterial growth rate and marine virus-host dynamics. *Microb Ecol* **40**: 114–124.
- Middelboe M, Holmfeldt K, Riemann L, Nybroe O, Haaber J. (2009). Bacteriophages drive strain diversification in a marine *Flavobacterium*: implications for phage resistance and physiological properties. *Environ Microb* **11**: 1971–1982.
- Middelboe M, Hagstrom A, Blackburn N, Sinn B, Fischer U, Borch NH *et al.* (2001). Effects of bacteriophages on the population dynamics of four strains of pelagic marine bacteria. *Microb Ecol* **42**: 395–406.
- Noble RT, Fuhrman JA. (1997). Virus decay and its causes in coastal waters. *Appl Environ Microb* **63**: 77–83.
- Noble RT, Fuhrman JA. (1998). Use of SYBR Green I for rapid epifluorescence counts of marine viruses and bacteria. *Aquat Microb Ecol* **14**: 113–118.
- Noble RT, Fuhrman JA. (2000). Rapid virus production and removal as measured with fluorescently labeled viruses as tracers. *Appl Environ Microb* **66**: 3790–3797.
- Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang HY, Cohoon M *et al.* (2005). The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res* **33**: 5691–5702.
- Parkinson N, Stead D, Bew J, Heeney J, Tsror LL, Elphinstone J. (2009). *Dickeya* species relatedness and clade structure determined by comparison of *recA* sequences. *Int J Syst Evol Microbiol* **59**: 2388–2393.
- Pasic L, Rodriguez-Mueller B, Martin-Cuadrado A-B, Mira A, Rohwer F, Rodriguez-Valera F. (2009). Metagenomic islands of hyperhalophiles: the case of *Salinibacter ruber*. *BMC Genomics* **10**: 570–580.
- Rodriguez-Brito B, Rohwer F, Edwards RA. (2006). An application of statistics to comparative metagenomics. *BMC Bioinformatics* **7**: 162.
- Rodriguez-Valera F, Ventosa A, Juez G, Imhoff JF. (1985). Variation of environmental features and microbial populations with salt concentrations in a multi-pond saltern. *Microb Ecol* **11**: 107–115.
- Rodriguez-Valera F, Martin-Cuadrado A-B, Rodriguez-Brito B, Pašić L, Thingstad TF, Rohwer F *et al.* (2009). Explaining microbial population genomics through phage predation. *Nat Rev Microbiol* **7**: 828–836.
- Rohwer F, Edwards R. (2002). The Phage Proteomic Tree: a genome-based taxonomy for phage. *J Bacteriol* **184**: 4529–4535.
- Sambrook J, Fritsch EF, Maniatis T. (1982). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press: Cold Spring Harbor, New York, USA.
- Short SM, Short CM. (2009). Quantitative PCR reveals transient and persistent algal viruses in Lake Ontario, Canada. *Environ Microbiol* **11**: 2639–2648.
- Sigee DC. (2005). Freshwater microbiology: biodiversity and dynamic interactions of microorganisms in the freshwater environment. In: Sigee DC (ed). *Freshwater Microbiology*. John Wiley & Sons, Ltd.: England, pp 105–180.
- Stoddard LI, Martiny JBH, Marston MF. (2007). Selection and characterization of cyanophage resistance in marine *Synechococcus* strains. *Appl Environ Microb* **73**: 5516–5522.
- Taylor R. (1984). *Predation* Chapman & Hall: New York, New York, USA.
- Thingstad F. (1997). A theoretical approach to structuring mechanisms in the pelagic food web. *Hydrobiologia* **363**: 59–72.
- Thingstad F. (2000). Elements of a theory for the mechanisms controlling abundance, diversity and the biogeochemical role of lytic bacterial viruses in aquatic systems. *Limnol Oceanogr* **45**: 1320–1328.
- Tringe SG, von Mering C, Kobayashi A, Salamov AA, Chen K, Chang HW *et al.* (2005). Comparative metagenomics of microbial communities. *Science* **308**: 554–557.
- Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JL. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**: 1027–1031.
- Tyson GW, Banfield JF. (2008). Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. *Environ Microb* **10**: 200–207.
- Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM *et al.* (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**: 37–43.
- Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA *et al.* (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**: 66–74.
- Vos M, Birkett PJ, Birch E, Griffiths RI, Buckling A. (2009). Local adaptation of bacteriophages to their bacterial hosts in soil. *Science* **325**: 833.
- Whitman WB, Coleman DC, Wiebe WJ. (1998). Prokaryotes: the unseen majority. *Proc Natl Acad Sci USA* **95**: 6578–6583.
- Willner D, Thurber RV, Rohwer F. (2009). Metagenomic signatures of 86 microbial and viral metagenomes. *Environ Microb* **11**: 1752–1766.
- Winget DM, Wommack KE. (2009). Diel and daily fluctuations in virioplankton production in coastal ecosystems. *Environ Microbiol* **11**: 2904–2914.
- Wommack KE, Ravel J, Hill RT, Chun J, Colwell RR. (1999). Population dynamics of Chesapeake Bay virioplankton: total-community analysis by pulsed-field gel electrophoresis. *Appl Environ Microb* **65**: 231–240.
- Yannarell AC, Triplett EW. (2005). Geographic and environmental sources of variation in lake bacterial community composition. *Appl Environ Microb* **71**: 227–239.
- Zehr JP, Jenkins BD, Short SM, Steward GF. (2003). Nitrogenase gene diversity and microbial community structure: a cross-system comparison. *Environ Microb* **5**: 539–554.
- Zhong Y, Chen F, Wilhelm SW, Poorvin L, Hodson RE. (2002). Phylogenetic diversity of marine cyanophage isolates and natural virus communities as revealed by sequences of viral capsid assembly protein gene g20. *Appl Environ Microb* **68**: 1576–1584.
- Zhou J, Spratt BG. (1992). Sequence diversity within the *argF*, *fbp* and *recA* genes of natural isolates of *Neisseria meningitidis*: interspecies recombination within the *argF* gene. *Mol Microbiol* **6**: 2135–2146.

Supplementary Information accompanies the paper on The ISME Journal website (<http://www.nature.com/ismej>)