Viral and microbial community dynamics in four aquatic environments

Beltran Rodriguez-Brito1,2,9, LinLin Li2,9,10, Linda Wegley2, Mike Furlan2, Florent Angly1,2, Mya Breitbart2,11, John Buchanan3, Christelle Desnues2,12, Elizabeth Dinsdale2, Robert Edwards1,2, Ben Felts4, Matthew Haynes2, Hong Liu2, David Lipson2, Joseph Mahaffy4, Anna Belen Martin-Cuadrado5, Alex Mira5,13, Jim Nulton4, Lejla Pasic’5,6, Steve Rayhawk1,2, Jennifer Rodriguez-Mueller2, Francisco Rodriguez-Valera5, Peter Salamon4, Shailaja Srinagesh2, Tron Frede Thingstad8, Tuong Tran2, Rebecca Vega Thurber2,14, Dana Willner2, Merry Youle7 and Forest Rohwer2

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

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 (Sigee, 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

Subject Category: microbial population and community ecology

Keywords: viruses; community dynamics; Kill-the-Winner; metagenomics; aquatic environments
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 distinct, 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-μm pore size Anodisc membrane filter (Whatman, Maidstone, Kent, England), stained with SYBR Gold (Molecular Probes, Inc., Eugene, Oregon, USA), and counted by epifluorescent microscopy. 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.

Preparation of DNA

Viral and microbial fractions were isolated from each water sample by passage through a 0.2-μm 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 avg = 53 000 × 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⁻¹ 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-μm TFF retentate by centrifugation at ~2000 × g
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.

### Table 1 Metagenomic libraries used in this study

<table>
<thead>
<tr>
<th>Name</th>
<th>Date</th>
<th># Reads</th>
<th># Base pairs</th>
<th>Average read length (bp)</th>
<th>GC (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fresh-water viruses (tilapia channel)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A-V November 2005</td>
<td>267640</td>
<td>27099247</td>
<td>101.3</td>
<td>45.6</td>
<td></td>
</tr>
<tr>
<td>B-V April 2006</td>
<td>63978</td>
<td>6420157</td>
<td>100.4</td>
<td>46.2</td>
<td></td>
</tr>
<tr>
<td>C-V August 2006</td>
<td>67612</td>
<td>6865291</td>
<td>101.5</td>
<td>48.7</td>
<td></td>
</tr>
</tbody>
</table>

| Fresh-water microbes (tilapia channel) |            |          |              |                          |        |
| A-M November 2005           | 381076     | 38423159 | 100.8        | 43.8                     |
| B-M April 2006              | 63978      | 6420157  | 100.4        | 44.0                     |
| C-M August 2006             | 67612      | 6865291  | 101.5        | 50.4                     |

| Fresh-water viruses (prebead pond) |            |          |              |                          |        |
| B2-V April 2006              | 67988      | 6949329  | 102.2        | 39.6                     |

| Fresh-water microbes (prebead pond) |            |          |              |                          |        |
| B2-M April 2006               | 44094      | 4384895  | 99.4         | 44.9                     |

| Low salinity viruses         |            |          |              |                          |        |
| D-V 1 July 2004              | 268531     | 27785763 | 103.5        | 46.5                     |
| E-V 10 November 2005         | 110511     | 11421737 | 103.4        | 41.2                     |
| F-V 28 November 2005         | 62685      | 6448398  | 102.9        | 47.7                     |

| Low salinity microbes        |            |          |              |                          |        |
| D-M 1 July 2004              | 268206     | 25012316 | 93.3         | 54.6                     |
| E-M 10 November 2005         | 49074      | 4583126  | 93.4         | 53.8                     |
| F-M 28 November 2005         | 34296      | 3419010  | 99.7         | 42.6                     |

| Medium salinity viruses      |            |          |              |                          |        |
| G-V 10 November 2005         | 39578      | 3973591  | 100.4        | 47.6                     |
| H-V 11 November 2005         | 35053      | 5703709  | 99.2         | 49.1                     |
| I-V 16 November 2005         | 58735      | 3477534  | 97.1         | 46.5                     |
| J-V 22 November 2005         | 55903      | 5578541  | 99.8         | 48.5                     |
| K-V 28 November 2005         | 39943      | 3971718  | 99.4         | 48.7                     |

| Medium salinity microbes     |            |          |              |                          |        |
| G-M 10 November 2005         | 38929      | 3867026  | 99.3         | 55.3                     |
| H-M 11 November 2005         | 23261      | 2299980  | 99.9         | 60.9                     |
| I-M 16 November 2005         | 120987     | 1174041  | 97.1         | 61.4                     |
| K-M 28 November 2005         | 8062       | 697933   | 86.6         | 56.2                     |

| High salinity viruses        |            |          |              |                          |        |
| L-V 16 November 2005         | 154167     | 15225789 | 98.8         | 53.0                     |
| M-V 28 November 2005         | 4645       | 460564   | 99.2         | 47.7                     |
| N-V 7 December 2005          | 47587      | 4813217  | 101.2        | 49.0                     |
| O-V 20 December 2005         | 303274     | 28008560 | 92.4         | 53.9                     |

| High salinity microbes       |            |          |              |                          |        |
| P-M 1 July 2004              | 582681     | 58877500 | 101.1        | 63.6                     |
| M-M 28 November 2005         | 35446      | 3675849  | 103.7        | 61.4                     |
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 (Robwer 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 (Rodriguez-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 β-diversity between the two samples. MaxiΦ is based on PHACCS, an α-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 α-diversity of the sample.

MaxiΦ uses the same approach as PHACCS, but here to measure the β-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...
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).
Table S4-B). Consistent with the 16S rDNA-based survey, these analyses showed that the more abundant microbial and viral taxa within each environment 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...
which pairwise comparisons identified significant differences (5) virulence; (6) stress response; (7) metabolism of aromatic compounds; and (8) phosphorous metabolism. Asterisks: subsystems for (1) cofactors, vitamins, prosthetic groups, pigments; (2) photosynthesis; (3) regulation and cell signaling; (4) secondary metabolism; salinity. Sequences for each microbiome were assigned to metabolic subsystems using the SEED database. The ‘Other’ category includes (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.

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 (Willingner 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
time in each environment were modeled using MaxiΦ. 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 Munoz 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 (Rodriguez-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
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**


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