

ORIGINAL ARTICLE

Microbial and viral metagenomes of a subtropical freshwater reservoir subject to climatic disturbances

Ching-Hung Tseng^{1,2,3}, Pei-Wen Chiang³, Fuh-Kwo Shiah⁴, Yi-Lung Chen⁴, Jia-Rong Liou⁴, Ting-Chang Hsu⁴, Suhinthan Maheswararajah⁵, Isaam Saeed⁵, Saman Halgamuge⁵ and Sen-Lin Tang^{1,3}

¹Bioinformatics Program, Taiwan International Graduate Program, Academia Sinica, Taipei, Taiwan; ²Institute of Biomedical Informatics, National Yang-Ming University, Taipei, Taiwan; ³Biodiversity Research Center, Academia Sinica, Taipei, Taiwan; ⁴Research Center for Environmental Changes, Academia Sinica, Taipei, Taiwan and ⁵Optimisation and Pattern Recognition Research Group, Department of Mechanical Engineering, Melbourne School of Engineering, The University of Melbourne, Victoria, Australia

Extreme climatic activities, such as typhoons, are widely known to disrupt our natural environment. In particular, studies have revealed that typhoon-induced perturbations can result in several long-term effects on various ecosystems. In this study, we have conducted a 2-year metagenomic survey to investigate the microbial and viral community dynamics associated with environmental changes and seasonal variations in an enclosed freshwater reservoir subject to episodic typhoons. We found that the microbial community structure and the associated metagenomes continuously changed, where microbial richness increased after typhoon events and decreased during winter. Among the environmental factors that influenced changes in the microbial community, precipitation was considered to be the most significant. Similarly, the viral community regularly showed higher relative abundances and diversity during summer in comparison to winter, with major variations happening in several viral families including *Siphoviridae*, *Myoviridae*, *Podoviridae* and *Microviridae*. Interestingly, we also found that the precipitation level was associated with the terrestrial viral abundance in the reservoir. In contrast to the dynamic microbial community (L -divergence 0.73 ± 0.25), we found that microbial metabolic profiles were relatively less divergent (L -divergence 0.24 ± 0.04) at the finest metabolic resolution. This study provides for the first time a glimpse at the microbial and viral community dynamics of a subtropical freshwater ecosystem, adding a comprehensive set of new knowledge to aquatic environments.

The ISME Journal (2013) 7, 2374–2386; doi:10.1038/ismej.2013.118; published online 11 July 2013

Subject Category: Microbial ecology and functional diversity of natural habitats

Keywords: metagenome; metavirome; freshwater reservoir

Introduction

Natural disturbances cause short-term changes in environmental conditions, and usually result in long-lasting ecological effects that create temporal and spatial heterogeneity in ecosystems (Sousa, 1984). The island of Taiwan provides an ideal platform to study these effects. It is located in the Western Pacific on the boundary between tropics and subtropics, and typhoons (i.e., tropical cyclones or hurricanes) are among its various seasonal disturbances, with an average annual occurrence of 3.7 (Wu and Kuo, 1999). Strong winds and torrential rain are typical components of typhoons that disrupt

the natural environment. Strong wind-induced vertical mixing and upwelling pump nutrients (carbon, nitrogen, phosphorus and silicon) from deep water, and enhance chlorophyll *a* concentration (Vinayachandran and Mathew, 2003; Walker *et al.*, 2005; Byju and Kumar, 2011; Hung and Gong, 2011). Abrupt torrential rain contributes to the concentrated influx of organic carbon and nutrients to the ocean (West *et al.*, 2011), which is considered as the main cause of the post-storm increase in phytoplankton growth and primary production in coastal and estuary ecosystems (Chang *et al.*, 1996; Shiah *et al.*, 2000; Paerl *et al.*, 2006).

The impact of artificial disturbances to microbial community has also been studied in a temperate lake. According to Shade *et al.* (2011), supplementing oxygen was a more intense disturbance to the hypolimnion community when compared with water mixing, whereas adding nutrients was less intense than mixing in the epilimnion. Although the communities changed quickly upon disturbing, they demonstrated

Correspondence: S-L Tang, Biodiversity Research Center, Academia Sinica, 128 Academia Road, Section 2, Nankang, Taipei 11529, Taiwan.

E-mail: sltang@gate.sinica.edu.tw

Received 20 March 2013; revised 7 June 2013; accepted 10 June 2013; published online 11 July 2013

resilience 10 days after treatments. Shade *et al.* (2012) further extended their study for a larger disturbance, an artificial mixing that increased water temperature thus changed the limnological conditions. In both epilimnion and hypolimnion, the communities changed as the result of disturbance, and returned to the pre-manipulation state by 7 and 11 days, respectively, as some parameters (e.g., oxygen level) recovered. Both experiments consistently demonstrated that the lake microbial community was less resistant but highly resilient to artificial disturbances.

Several inland freshwater ecosystems in Taiwan have previously been used to monitor natural disturbances like typhoon. For instance, a study of a subalpine freshwater lake showed that typhoons disturbed the microbial community by inducing water-mixing (Jones *et al.*, 2008). The Feitsui Reservoir (FTR), an enclosed freshwater reservoir, is also a renowned study site located in North Taiwan (Figure 1). Existing evidence shows that typhoons disrupt the thermal stratification of the water column and elevates turbidity by initiating the suspension interflow (Chen *et al.*, 2006). Typhoons also change dissolved oxygen level by delivering oxygen-rich water to the reservoir (Fan and Kao, 2008). A 4-year survey further demonstrated the ecological interaction between typhoons and dissolved organic carbon dynamics. As phosphate is a limited nutrient in FTR, the typhoon-induced phosphate pulse relieves bacteria from phosphate starvation, and then leads to a high consumption rate of dissolved organic carbon (Tseng *et al.*, 2010). The advent of metagenomics has enabled us to ask the following questions to investigate these eminent typhoon-induced effects of reservoir ecology at a much finer scale: What is the microbial community composition and dynamics in FTR? What is the association between environmental changes and community dynamics? What are the metabolic traits of microbial community underpinning the reservoir?

In addition to microbes, environmental viruses are now adding a wealth of knowledge to ecological studies since the maturation of metagenomic technologies and approaches (Edwards and Rohwer,

2005). Studies on marine viruses (most of which are bacteriophages) have revealed their important roles in mediating microbial diversity and biogeochemical cycling (Breitbart, 2012). These significant discoveries have motivated us to also examine their freshwater counterparts. Viral communities and metagenomes in natural freshwater ecosystems have previously been described in various lakes and perennial ponds in Sahara. The study looking into freshwater RNA viruses suggested that the freshwater lake ecosystems might serve repositories of pathogenic and non-pathogenic RNA viruses because of their direct contact with humans, and domestic and wild animals (Djikeng *et al.*, 2009). In an Antarctic lake, Lopez-Bueno *et al.* (2009) found a large proportion of eukaryotic viruses and the seasonal shifting of viral types, which in ice-covered spring were more single-stranded DNA (ssDNA) viruses and in open-water summer the community was dominated by double-stranded DNA (dsDNA) viruses. The viral assemblages in two distant freshwater lakes in France differing by ecological parameters were genetically similar with each other but distinct from other aquatic ecosystems. The phylogenetic analysis on conserved proteins in major viral groups suggested that the two lakes probably composed of evolutionarily close virotypes and differed mainly in the relative abundance of viral species (Roux *et al.*, 2012). In the study of four perennial ponds in Sahara, tailed phages were found to be the most abundant in surveyed communities. The differences in viral community structure and temperate (i.e., lysogenic) virus abundance were hypothesized to result from the critically low water level in one pond (Fancello *et al.*, 2012). However, compared with marine environments, viral communities in freshwater are still much less studied. Our study is the first attempt to report the freshwater viral communities that are frequently disturbed by extreme climatic activities.

This study reports on the results of a long-term survey of FTR spanning from July 2007 to January 2009, comprising 12 microbial and viral metagenomes. We also identified several metabolic traits that explain

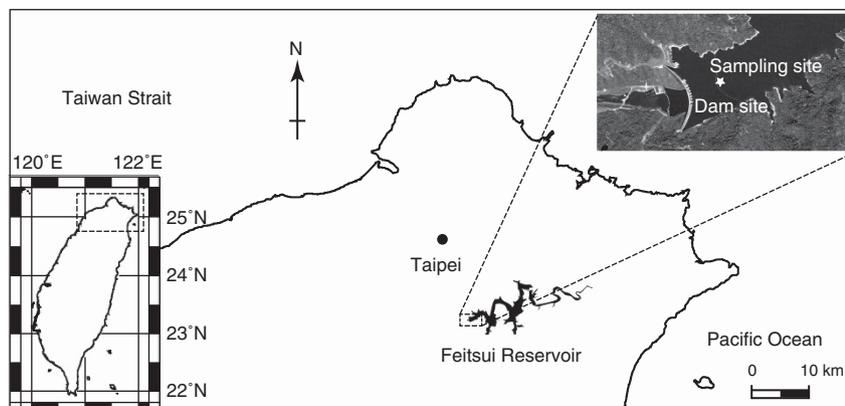


Figure 1 Map of Taiwan and the location of the sampling site at FTR.

environmental selection processes and functional preference using comparative approaches. This study has provided much needed insight into the interactions between microbes, viruses and environmental changes of a subtropical freshwater ecosystem.

Materials and methods

Sampling site, sample collection and preparation

The sampling site in this study was the FTR (24°54'36"N, 121°34'55"E) (Figure 1), which supplies water to over four million people living in metropolitan Taipei and surrounding suburbs. The deepest site of FTR is next to the dam, of 113.5-m depth. The reservoir is located downstream of the Kingkwa Creek, the Diyu Creek and the Peishih Creek, and has a watershed of 303 km² (Chen *et al.*, 2006; Fan and Kao, 2008). Freshwater samples were collected at the dam site (Figure 1), where all the water converged.

As disturbances to FTR are predominantly caused by typhoons during summer and deeper water mixing in winter (Tseng *et al.*, 2010), a total of six sampling events were conducted (four in summer and two in winter) on (1) 31 July 2007, (2) 28 August 2007, (3) 15 January 2008, (4) 22 July 2008, (5) 26 August 2008 and (6) 13 January 2009 (Table 1). Sampling was undertaken between 0900 and 1200 hours, water parameters were also measured as described previously (Tseng *et al.*, 2010), and daily precipitation data from 2007 to 2009 were obtained from the Taipei FTR Administration Bureau. During sampling, 150-l freshwater was pumped up from 5-m depth and directly filtered through a 10- μ m membrane into plastic carboy bottles. In the laboratory, the microbes and viruses contained in the freshwater samples were separated by 0.22- μ m membrane filters using a tangential-flow filtration system as described previously (Thurber *et al.*, 2009). The filtrate containing viruses was further concentrated using a 50-kDa filter. After being treated with DNase I (New England Biolabs UK Ltd, Herts, UK) to digest free DNAs, virus-like particles were then purified using discontinuous cesium chloride gradients by ultracentrifugation at 1.35–1.5 g ml⁻¹ density (Angly *et al.*,

2006). Virus-like particles were confirmed by transmission electron microscopy.

DNA extraction, sequencing, assembly and annotation

The 300-ml microbial retentate was firstly spun down by high-speed centrifugation and used for DNA extraction via cetyltrimethylammonium bromide (CTAB) method (Wilson, 2001). Viral DNAs were extracted from purified virus-like particles using formamide/cetyltrimethylammonium bromide method (Angly *et al.*, 2006). After DNA quality and quantity check by DNA agarose gel electrophoresis, viral DNAs were amplified with the Genomiphi kit (GE Healthcare Life Science, Piscataway, NJ, USA) to collect enough amounts of DNA for next-generation sequencing.

Both purified microbial and viral DNAs were subject to 454 pyrosequencing using the Roche GS FLX System (Roche 454 Life Sciences, Branford, CT, USA) at Mission Biotech (Taipei, Taiwan). The default settings in GS Run Processor (v2.5, Roche 454 Life Sciences) were used for read quality control. High-quality reads were subsequently assembled using GS *de novo* Assembler (v1.1.02, Roche 454 Life Sciences) with a 40-bp minimum overlap and 99% consensus. Contigs with any ambiguous base (N) were excluded from open reading frame (ORF) prediction by MetaGeneMark (<http://exon.gatech.edu>). Each putative ORF protein was searched against the National Center for Biotechnology Information (NCBI) non-redundant protein database using BLASTp. Matches with an e-value < 10⁻⁵ were grouped based on the identical description (i.e., annotation) of each hit. In the group with the highest totaled bit score, the best hit was chosen to annotate the query sequence. To infer Clusters of Orthologous Group (COG), each protein was assigned to the best match in eggNOG (v3.0) database (Powell *et al.*, 2012) using BLASTp with an e-value cutoff of 10⁻⁵. Clustered regularly interspaced short palindromic repeats were identified in contigs using PILER-CR (v1.06, <http://www.drive5.com/pilercr>) with default settings.

Table 1 Summary of limnological data and statistics of metagenomes

Sample	Date	Limnological characteristics ^a			Metagenomic data statistics			
		Temp (°C)	pH	O ₂ (p.p.m.)	No. of reads	Avg. read length (bp)	%GC	No. of contigs
M1/V1	31 Jul 2007	31.4	9.08	7.64	407 936/116 107	244/240	44.3/41.1	45 624/8250
M2/V2	28 Aug 2007	28.3	8.09	9.45	304 567/96 373	250/245	40.5/42.4	21 618/9800
M3/V3	15 Jan 2008	19.5	6.52	7.99	242 381/76 396	234/252	44.4/41.8	22 602/10 000
M4/V4	22 Jul 2008	30.1	8.78	7.35	315 807/68 647	240/254	55.2/46.7	40 311/6359
M5/V5	26 Aug 2008	30.5	9.13	6.92	259 455/62 588	241/256	56.2/42.8	33 926/6096
M6/V6	13 Jan 2009	19.0	6.96	6.96	210 276/60 313	239/263	55.3/38.5	24 607/5042

Abbreviation: %GC, percentage of guanine–cytosine content.

^aFull limnological characteristics are available in Supplementary Table S2.

Comparative genomics and statistical analyses

All microbial and viral metagenomic contigs were searched against seven foreign metagenomes using BLAT (v3.4) (Kent, 2002) with options '-t=DNA -q=DNA'. The average best-match scores of our metagenomes were analyzed by principal component analysis. Seven foreign metagenomes were downloaded from CAMERA (<http://camera.calit2.net>). For functional profile analysis, COG profiles of various environments were downloaded from IMG/M (<http://img.jgi.doe.gov>) and CAMERA for correspondence analysis. The downloaded functional assignments were filtered using an e-value cutoff of 10^{-5} . Every set of data was divided by total counts for normalization. Please refer to Supplementary Information for detailed lists of public data set. The association between community and environmental variables was verified by multivariate analysis of variance. All statistical and multivariate analyses were performed in R (<http://www.r-project.org>). The functional enrichment analysis and Poisson regression were performed using the R package ShotgunFunctionalizeR (Kristiansson *et al.*, 2009).

L-divergence and analysis of similarity

The difference of microbial functional profiles was quantified by L-divergence (Lin, 1991) and tested using analysis of similarity at the finest metabolic resolution level (i.e., COG families). Analysis of similarity was performed using the R package vegan (Oksanen *et al.*, 2007) with arbitrarily grouping six samples into two groups by all possible permutations: (1) one sample in one group and five samples in the other (6 permutations); (2) two samples in one group and four samples in the other (15 permutations); and (3) three samples in both groups (20 permutations).

Microbial and viral community analyses

Multiplex sequencing of the bacterial 16S V1–V2 variable regions was conducted as described previously (Hamp *et al.*, 2009; Chen *et al.*, 2011) for six microbial samples. The clone library of the first two samples was constructed as the control for validating community reproducibility. Details of both multiplex sequencing and clone library construction are available in Supplementary Information. In brief, after sequencing (Roche 454 GS FLX Titanium System; Mission Biotech), chimera checking and quality trimming, qualified reads were classified using RDP classifier (v2.3) (Wang *et al.*, 2007) with a bootstrap value of 0.8, followed by an analysis on unclassified reads using an in-house script called Greedy-Fetching (<http://tanglab.csie.org/gf>) (Chen *et al.*, 2011). Chloroplast reads were excluded from further analyses. Reads of each sample were aligned by MUSCLE (<http://www.drive5.com/muscle>) and further analyzed in MOTHUR (Schloss *et al.*, 2009) to generate operational taxonomic units (OTUs)

with a 3% cutoff value. OTUs are used for the estimation of evenness, richness, the Shannon–Weaver index, Chao1 estimator, the Simpson index and rarefaction curves. For community clustering, pairwise distances were calculated using FastUniFrac (Hamady *et al.*, 2010), and tested by a *P*-test on the FastUniFrac web interface (<http://bmf2.colorado.edu/fastunifrac>). The robustness of clustering was determined by the Jackknife Sample Clusters option.

Viral community structure and diversity were modeled using PHACCS (v1.1.3, <http://sourceforge.net/projects/phaccs>) based on the contig spectra generated by Circonspect (v0.2.5, <http://sourceforge.net/projects/circonspect>) with default settings. For taxonomy identification, contigs in viral metagenomes were searched against the NCBI RefSeq viral protein collection (<ftp://ftp.ncbi.nlm.nih.gov/refseq/release/viral>) using BLASTx with an e-value cutoff of 10^{-5} (single best match). Owing to the fact that (1) the viral concentration in FTR (average 1.8×10^7 per ml; Supplementary Figure S15) is much higher than the number of sequenced reads (i.e., average 80 000 reads per sample; Table 1); and that (2) the length of viral genomes is relatively short, we therefore assumed that every read that assembled into a contig originated from different viral particles (of the same viral type). We were therefore able to calculate the abundance of viral communities using an equation weighting the number of reads by the alignment quality and contig length (Supplementary Material and methods).

Data deposition

Multiplex sequenced reads (bacterial 16S V1–V2 region) and both microbial and viral metagenomes have been deposited to the NCBI Sequence Read Archive under the accession number SRA048150.

Results

Sampling, typhoon history and metagenome overview

From July 2007 to January 2009, we collected samples at six predetermined intervals for both microbial (M) and viral (V) metagenome library construction. As water stratification is upset by typhoons and lower temperatures, we sampled M1 and M2 in the summer of 2007, before/after the occurrence of typhoons. We then sampled M3 in January 2008 during winter. In 2008, typhoon events occurred at similar times to 2007 (Figure 2), so we sampled M4 and M5 in July and August, respectively, to detect whether there was similar gene content in the reservoir. Finally, we sampled M6 in January 2009. Viral metagenomes were all copurified from the microbial samples, and denoted V1–V6.

We note that 12 typhoon alarms were sounded during the total period of collecting our samples: three occurred between M1 and M2, and two

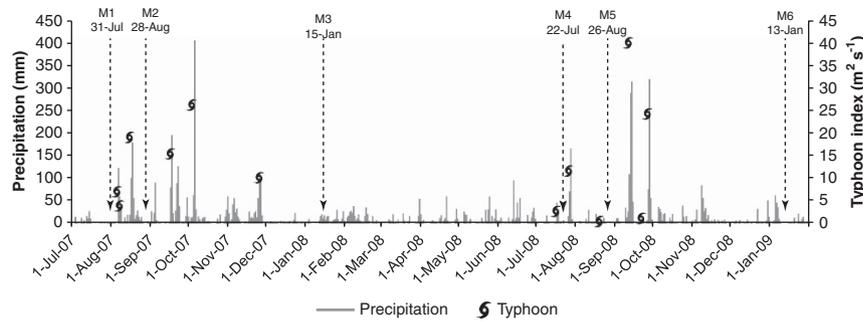


Figure 2 Sampling time point, typhoon and precipitation records. Typhoon index is defined as the product of precipitation (m) and wind speed (m s^{-1}).

between M4 and M5. Most typhoons were followed by peaks in precipitation levels (Figure 2), which demonstrated the fundamental role of a typhoon as the primary source of water for FTR in summer.

Microbial and viral metagenomes were sequenced using a whole-genome shotgun approach (Table 1 and Materials and methods). A total of 218 412 ORFs were identified, with an average of 28 458 ORFs in each microbial metagenome and an average of 7944 ORFs in each viral metagenome. An average of 57.3% microbial and 31.7% viral ORFs found hits in NCBI-nr database, and 41.0% microbial and 20.6% viral ORFs were assigned to known COGs. In the BLAT analysis, an average of 17.9% microbial and 12.6% viral contigs showed homology to compared foreign metagenomes (Supplementary Table S1).

Analysis of GC content (guanine–cytosine content) revealed different dynamic patterns in the metagenomes (Table 1). An annual %GC discrepancy was found between M1–M3 and M4–M6 metagenomes by 10%. Furthermore, M2 had distinctively lower %GC in comparison to M1 and M3, whereas M4, M5 and M6 showed little discrepancy. The viral metagenomes varied in %GC from between each sampling interval. When comparing the collected metagenomes with foreign metagenomes (see Materials and methods), we found that M4 to M6 clustered together, whereas the viral samples possessed a more scattered distribution (Figure 3). Similar environmental parameters that were observed when sampling M4 to M6 (Supplementary Table S2) suggested that the concurrent conditions have a more deterministic effect on microbial communities compared with the associated viral communities.

Microbial community in FTR

We cloned bacterial 16S rRNA genes in the M1 and M2 communities to validate the reproducibility of multiplex sequencing (Supplementary Materials and methods). High correlations indicated that the multiplex sequencing has enough resolution for microbial community analysis at the class level (Supplementary Figure S1). The comparison of microbial community compositions inferred from

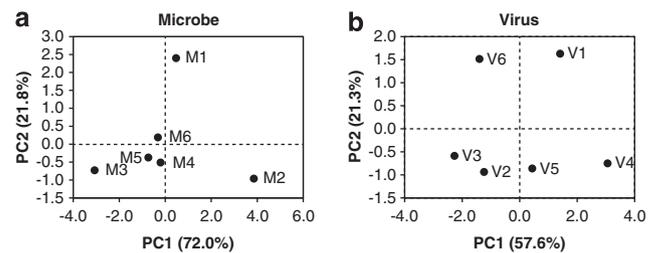


Figure 3 Principal component (PC) analysis of microbial and viral metagenomes. Ordinations of (a) microbial and (b) viral metagenomes were based on the similarity derived from BLAT analysis against foreign samples.

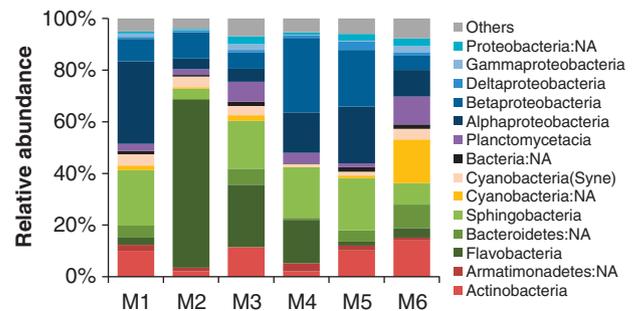


Figure 4 Microbial community structure in the FTR freshwater. Bacterial lineages without identifiable class names were indicated by phylum: NA (not applicable). Syne, *Synechococcus*.

multiplex sequencing and metagenomic data also show significant correlations (r 's) with a mean \pm s.d. of 0.756 ± 0.119 ($P < 0.05$) (Supplementary Table S3).

In FTR, microbial communities predominantly comprised of four phyla: *Bacteroidetes* (39%, averaged from M1 to M6) and *Proteobacteria* (33%), followed by *Actinobacteria* (8%) and *Cyanobacteria* (6%) (Figure 4). Varying levels of abundance of six major classes of bacteria, including *Actinobacteria*, *Flavobacteria*, *Sphingobacteria*, *Alphaproteobacteria*, *Betaproteobacteria* and *Cyanobacteria*, characterized M1 through to M6 (Supplementary Figure S2). The clustering results based on UniFrac distances suggested that M1 and M5 possessed similar community structure, whereas

M3 and M6 (winter samples) were also found to be similar (Figure 5). The rarefaction curves demonstrated that our samplings were close to saturation (Supplementary Figure S3). The Shannon index (H') ranging from 4.49 to 5.34 (Supplementary Table S4) quantitatively suggested a high diversity that is similar to the microbial communities in coastal reef water in southern Taiwan (Chen *et al.*, 2011) and coastal seawater in North Sea (Bolhuis and Stal, 2011). We observed a recurring pattern of an increase in species richness after typhoons (i.e., M2, M4 and M5) (Supplementary Table S4), which was probably due to the post-storm elevation of phosphate, a limited nutrient in FTR (Tseng *et al.*, 2010). Besides that, the runoff water could introduce exotic species to the reservoir, which might also contribute the increased richness after typhoons.

To explain the dynamics in microbial community composition, multivariate analysis of variance was applied to identify the significant associations between environmental variables (Supplementary Table S2) and the relative abundance of the six identified major bacterial classes.

The abundance of *Flavobacteria* positively correlated with oxygen concentration in the reservoir (Pearson's correlation $r=0.952$; $P=0.003$), consistent with their aerobic lifestyle. The marginal significance of past 2-week (before sampling) precipitation levels ($r=0.815$; $P=0.048$) implied that the precipitation only had minor effects on *Flavobacteria* populations. On the other hand, *Sphingobacteria* were found to be negatively correlated with the precipitation levels up to 2 weeks before sampling ($r=-0.956$; $P=0.003$) as well as the precipitation brought by typhoons ($r=-0.831$; $P=0.040$), suggesting that the external water input reduced their relative abundance. The *Alphaproteobacteria* population was negatively correlated with the silicate concentration ($r=-0.854$; $P=0.003$). Finally, bacterial biomass—a measure the bacterial cell density in samples—followed a negative trend with the abundance of *Actinobacteria* ($r=-0.827$; $P=0.042$) and *Cyanobacteria*

($r=-0.862$; $P=0.027$), but followed a positive trend with *Betaproteobacteria* ($r=0.835$; $P=0.039$).

Microbial functional profiles in FTR

Functional gene annotation based on COG classes revealed that the metabolic profiles of M1–M6 metagenomes were very similar to each other (Supplementary Figure S4). As a COG class is a general level of functional categorization, we further quantified the difference of M1–M6 COG profiles at the finest metabolic resolution (e.g., COG family) using the L -divergence measure (Lin, 1991), which has a scale from 0 (identical) to 2 (highly divergent). Results showed a small divergence with a mean \pm s.d. of 0.24 ± 0.04 . The COG family profiles were also tested using analysis of similarity and no significant difference was reported from all possible permutations (see Materials and methods). With the distinct community structure of M2 (Figure 4), we also found its metagenome possessed more genes involved in cell wall/membrane/envelop biogenesis (M, $P=5.23 \times 10^{-5}$, Poisson model, comparing two groups), and less in amino-acid transport and metabolism (E, $P=0.0137$), and energy production and conversion (C, $P=0.0486$). We believe that M2 (post-typhoon) represents a special case among our samples. The higher oxygen and phosphate but lower nitrate concentration (Supplementary Table S2) in M2 might give rise to the observed differential metabolic preferences.

In correspondence analysis plot (Supplementary Figure S5), the functional profiles of M1–M6 clustered with other lake and separated from soil and ocean metagenomes, again highlighting the causal relationships of niches on metabolic potentials (Dinsdale *et al.*, 2008). As M1–M6 were collected from a subtropical freshwater reservoir, the enclosed nature of the reservoir caused them to be of different genomic make-up in comparison to the ocean and other lakes. We therefore applied a comparative approach to identify specific metabolisms that are inherent in the FTR microbial community.

The statistical comparison using ShotgunFunctionalizeR against ocean metagenomes (Sargasso Sea: GS000A and GS000C; coastal ocean: GS008 and GS014; open ocean: GS018 and GS023) and a tropical freshwater reservoir metagenome (GS020), respectively, revealed 625 and 607 OGs significantly enriched in FTR ($P<0.05$, Poisson model, comparing two groups).

Relative to the ocean metagenomes, FTR harbored more genes involved in replication, recombination and repair (category L, $P=5.08 \times 10^{-5}$, binomial test), carbohydrate transport and metabolism (G, $P=1.57 \times 10^{-4}$) and defense mechanisms (V, $P=0.0309$) (Supplementary Figure S6A). Among carbohydrate-related functions, in addition to genes for metabolites transporter (COG0477 and COG0697), polysaccharide degradation (COG0366,

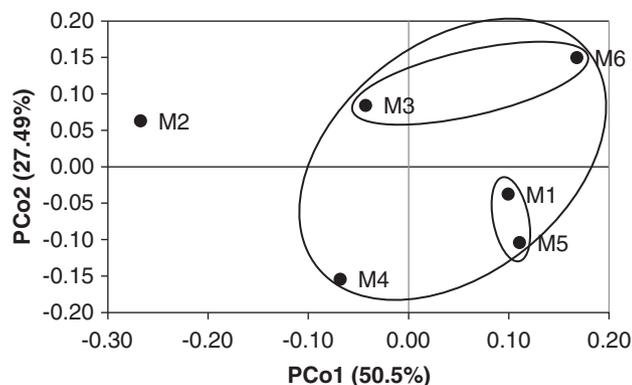


Figure 5 Principal coordinate (PCo) analysis of the microbial community. The clustering result is indicated by ellipses.

COG3250, COG3669 and COG4409) and glycolysis (COG1482), we also found several genes for the regulation of glycogen synthesis (COG0058, COG0296) (Supplementary Figure S7). Abundant functions for carbohydrate metabolism implied that microbes in FTR are more versatile in utilizing polysaccharides, which might result from the high dissolved organic carbon anomaly in the FTR (27–270 μM) (Tseng *et al.*, 2010) when compared with the tropical and subtropical surface oceans (60–80 μM) (Ogawa and Tanoue, 2003). Also, as suggested by the fact that bacteria usually perform glycogen synthesis under the limited growth condition with excessive carbon (Preiss, 1984) and based on the constantly low phosphate concentration in FTR (Supplementary Table S2), we suspect that FTR remains in an imbalanced trophic state with plentiful carbon but limited phosphate. In the DNA-processing COG category, we found that FTR microbial metagenomes were enriched with transposase (COG2801, COG3039 and COG5421) (Supplementary Figure S7), suggesting that bacterial genome recombination may be a frequent genomic activity in the local community and likely have a much more important role in microbial genome response to perturbation events.

In the comparison to Lake Gatun, not only did we again find overabundant genes for carbohydrate metabolism (G, $P = 1.94 \times 10^{-4}$) but we also found genes for defense mechanism (V, $P = 0.0074$), cell motility (N, $P = 0.0075$), signal-transduction mechanisms (T, $P = 0.0096$), secondary metabolites biosynthesis, transport and catabolism (Q, $P = 0.0414$) and function unknown (S, $P = 0.0439$) in FTR (Supplementary Figure S6B). The top-10 significant OGs of defense mechanism can be summarized into three major functions, including: (1) the restriction enzymes (COG1002, COG0286, COG0610, COG0732 and COG4096); (2) multidrug efflux pump and transporter (COG0841 and COG1132); and (3) bacteriocin and antibiotics resistance genes like antimicrobial peptide transporter, microcin resistance protein and β -lactamase (COG0577, COG1619 and COG1680) (Supplementary Figure S8). The enriched restriction enzyme and bacteriocin resistance genes suggested that the microbial community in FTR may be subject to constant attacks from not only viruses but also bacteria, which further implied the higher bacterial competitiveness in the reservoir with rich farm-discharging nutrient (Zehetner *et al.*, 2008) than Lake Gatun. As suggested by the enriched antibiotic resistance genes, we speculate that the antibiotics discharged from the tea plantation in the catchment area likely influenced metabolic selection within the community by retaining microbes that are more capable of dealing with antibiotics.

Following functional enrichment analysis, we used Poisson regression to associate environmental variables with COG categories, which revealed that the phosphate concentration correlated ($P < 0.05$)

with 19 OGs (0.43% of total OGs, the most), whereas bacterial biomass, pH, salinity and typhoon precipitation correlated with only one OGs (0.02%, the least). Among the top-10 phosphate-correlated OGs (Supplementary Figure S9), the abundance of three OGs positively correlated with phosphate availability, two of which were phage-related enzymes (ribonucleotide reductase, COG0208, $P = 0.0003$; muraminidase, COG3772, $P = 0.0014$) and one was without any recognizable function to phage physiology or metabolism (4-aminobutyrate aminotransferase, COG0160, $P = 0.0043$). According to the fact that 77% (30/39) COG0160 ORFs best matched to *Actinobacteria* homologs, the observed positive correlation might be contributed by the concurrent abundance of *Actinobacteria*, which also demonstrates the highest COG0160 copy number in all sequenced bacterial genomes (Supplementary Figure S10). The negative correlation between phosphate and alkaline phosphatase (COG1785, $P = 0.019$, ranked 13th) delineated the environmental selection of low phosphate availability on bacteria with alkaline phosphatase.

Viral community in FTR

As there is no universal gene in all viruses that can be used for phylogenetic or taxonomic identification (Rohwer and Edwards, 2002), we reconstructed the viral community through contig homology search (see Materials and methods). Typical bacteriophage families (i.e., *Siphoviridae*, *Myoviridae* and *Podoviridae*) and *Microviridae* were the dominant and constant viral members in FTR, accounting for

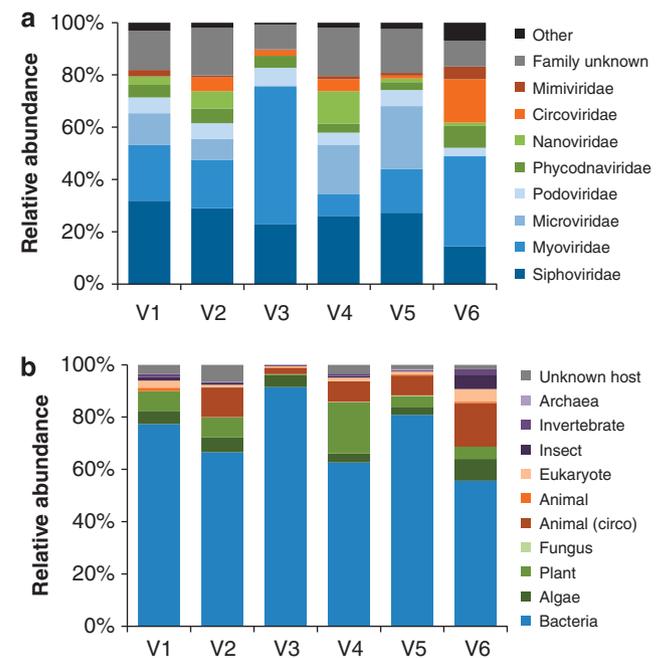


Figure 6 Viral community structure in FTR, classified by (a) families and (b) host types. Circo, circovirus.

66.7% on average (Figure 6a). The presence of typical phages and other viruses was confirmed by transmission electron microscopy (Figure 7). Among detected species, phages infecting cyanobacteria appeared with the top frequency of 29.2%, totaled from *Synechococcus* phages (19.3%), *Prochlorococcus* phages (7.6%) and cyanophages (2.3%), echoing the fact that *Synechococcus* is more detectable than *Prochlorococcus* in FTR. In the ocean, the cyanobacteria *Synechococcus* and *Prochlorococcus* are the major phytoplankton fixing carbons via photosynthesis (Waterbury *et al.*, 1986; Partensky *et al.*, 1999; DeLong and Karl, 2005; Jardillier *et al.*, 2010), and marine cyanophages reportedly

participate in the process by carrying the rate-limiting photosystem II protein D1 in their genomes (Sullivan *et al.*, 2006). Because cyanophage D1 protein were reportedly detectable in other freshwater lakes (Chenard and Suttle, 2008), we tried homology search and also detected D1 proteins in our metaviromes. However, any homology search is highly database-dependent. The detected prevalence of cyanophages in FTR could result from the extensive studies on them in the past decade. Still, average 68.3% ORFs (Supplementary Table S1) in our metaviromes remains unexplored.

Using host types to characterize the viral community provides another perspective to examine the

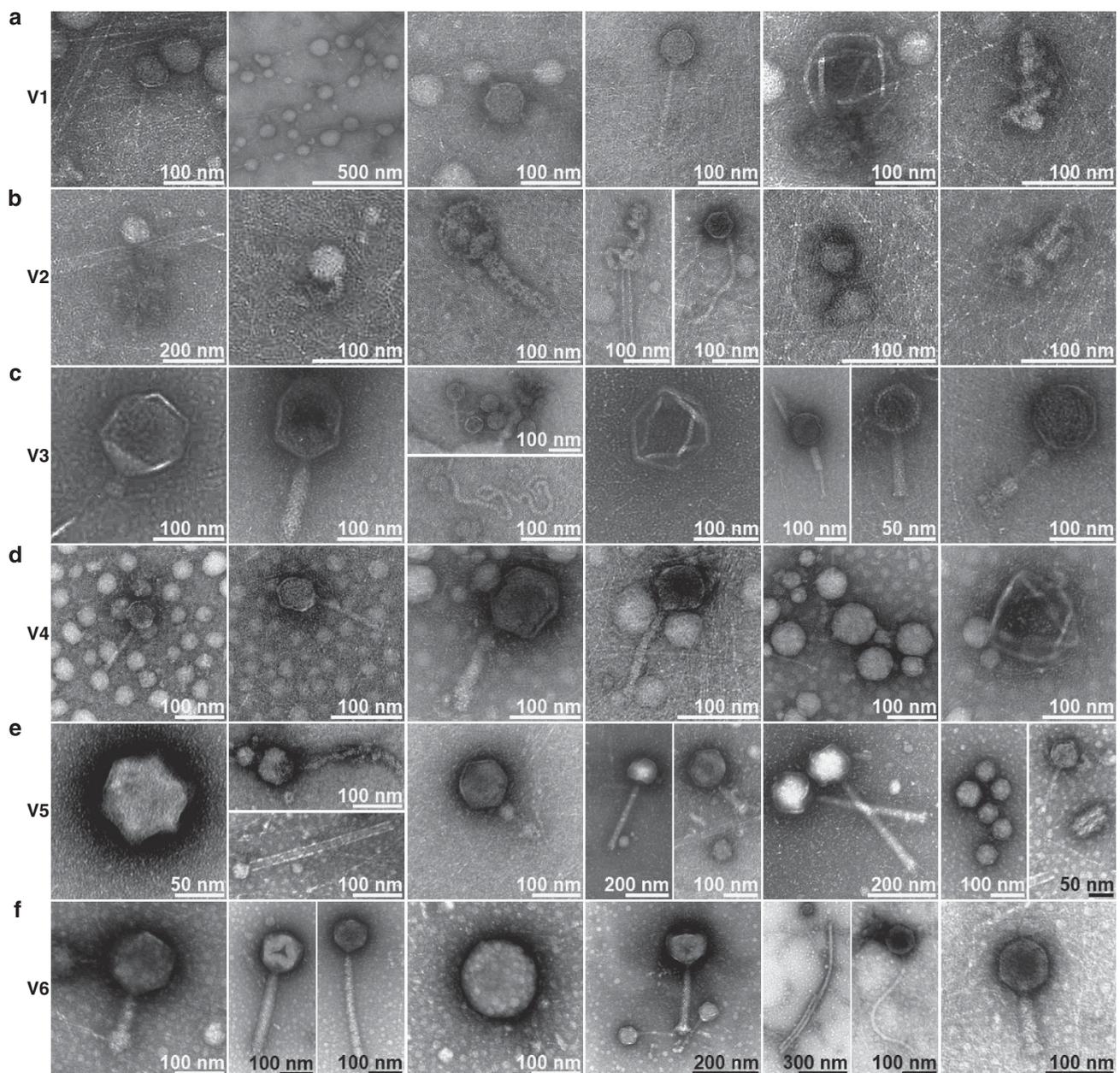


Figure 7 TEM observation of viral morphology in FTR. V1–V6 samples are shown in (a)–(f), respectively.

ecological implications of FTR. Bacteriophages were still the dominant type of virus in FTR, comprising 72.4% of samples on average (Figure 6b). Besides that, the relative abundance of viruses hosting in terrestrial plants (e.g., banana bunchy top, coconut foliar and subterranean clover) and animals (e.g., goose, starling and raven) seems to harmonize with the precipitation in the past 1 week before sampling ($r = 0.724$, $P = 0.104$) (Figure 8). Although the correlation is of a moderate significance in statistical test, the phenomenon that the runoff after precipitation transfers terrestrial viruses to FTR is intriguing and worth further studying.

The viral genotype richness, evenness and diversity of metaviromes were modeled using PHACCS (Table 2). The V2 and V4 were the most genotype-rich (7868 and 21 832 genotypes) and diverse (H' of 7.045 and 7.066). However, the viral diversity in FTR was lower than the ocean (H' of 9.81) (Angly *et al.*, 2006). When combined with estimates of evenness (FTR, 0.797; ocean, 0.910), FTR was both lower in viral diversity and evenness than the ocean, implying the biased dominance of a few viruses in FTR. The simplest viral community was found in V3 (2008 winter) with only 352 predicted genotypes, which was less than the second winter sample (V6). By correlating paired microbial and viral community data, we found a negative correlation between them in Shannon index (H') ($r = -0.619$, $P = 0.190$), whereas a positive correlation in richness ($r = 0.467$, $P = 0.350$) (Supplementary Figure S11). As both of them have P -values above 0.05 in correlation tests, further investigation would be required to validate these observed correlations.

Viral functional profiles in FTR

Using the same comparative paradigm as microbes, we compared our viral metagenomes to four oceanic viral samples collected from the Arctic Ocean, the Bay of British Columbia, the Gulf of Mexico and the Sargasso Sea (Supplementary Materials and Materials and methods). Three hundred and eighty OGS enriched in FTR ($P < 0.05$, Poisson model, comparing two groups), especially in the categories of general function (R, $P = 9.94 \times 10^{-11}$, binomial test), DNA processing (category L, $P = 9.25 \times 10^{-6}$), function unknown (S, $P = 1.11 \times 10^{-5}$), transcription (K, $P = 0.0161$) and envelope (M, $P = 0.0275$) (Supplementary Figure S12).

In the top-10 significantly enriched functions of DNA processing (Supplementary Figure S13B), essential functions for viral DNA replication and infection were detected in all compared metaviromes, but were on average more abundant in FTR. These included DNA polymerase I (COG0258), replicative DNA helicase (COG0305), DNA modification methylase (COG0863), site-specific recombinase (COG1961) and integrase (COG0582). Over 70% (15/21) genes assigned to intein/homing endonuclease (COG1372) matched eukaryotic

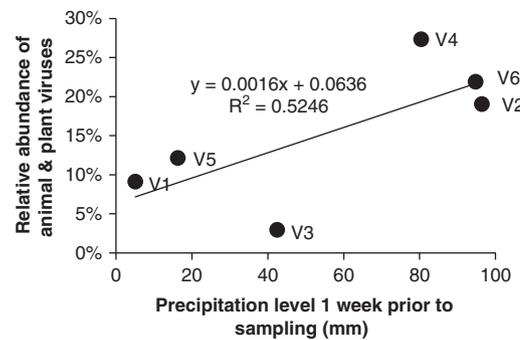


Figure 8 Correlation plot of relative abundance of plant and animal viruses versus precipitation 1 week before sampling in FTR.

Table 2 Viral community structure predicted from assembly of metagenomic sequences

Sample	Richness	Evenness	Most abundant (%)	Shannon (H')
V1	1040 genotypes	0.831	8.438	5.772
V2	7868 genotypes	0.785	7.251	7.045
V3	352 genotypes	0.881	7.969	5.167
V4	21 832 genotypes	0.707	9.499	7.066
V5	2052 genotypes	0.841	6.586	6.416
V6	2971 genotypes	0.737	11.517	5.890

homologs, probably brought by eukaryotic viruses in FTR. In the general function (category R), several typical virus structural proteins including sheath (COG3497) and tail fiber (COG5301) were marginally detectable in ocean viral metagenomes (Supplementary Figure S13A), suggesting morphological differences between all the viral assemblages that were compared.

Discussion

In this long-term study, we applied metagenomics to monitor the microbial and viral communities in a subtropical freshwater reservoir and to validate metabolic characteristics underpinning local system. In FTR, microbial communities were found to be dynamic throughout the entire survey period, which differs from the seasonal patterns found in other marine microbial communities (Gilbert *et al.*, 2012). As our survey was conducted over only 2 years, with only six sampling instances, a more intensive (i.e., monthly) and longer sampling design could help to more clearly define underlying succession patterns in the FTR microbial community. The variation of viral communities majorly occurred within different viral families (*Siphoviridae*, *Myoviridae*, *Podoviridae* and *Microviridae*), and several terrestrial eukaryotic viruses were detected intermittently. As multiple

displacement amplification reportedly prefers single-stranded DNA viruses (Kim and Bae, 2011), the proportion of *Microviridae*, a single-stranded DNA viral family, might be overestimated by the preference of multiple displacement amplification. For microbial metagenomes, the embedded metabolic profiles were with relatively smaller *L*-divergence (0.24 ± 0.04) than the community genera composition (0.73 ± 0.25). Enriched genes in carbohydrate metabolism and glycogen regulation suggested that FTR is in an imbalanced trophic state, with abundant carbon but limited concentrations of phosphorus. The abundant antibiotic resistance genes reflected the deterministic outcome of exotic antibiotics shaping local metabolism. For genes recovered in viral metagenomes, the enriched functions were mostly happening in those benefiting host DNA integrity and viral infection. Several common viral genes detected in FTR but not in other oceanic samples exhibited possible niche-specificity of local viral communities.

Freshwater microbial communities have been found to be resilient to both natural and artificial disturbance (Jones *et al.*, 2008; Shade *et al.*, 2011, 2012). Usually the community composition returns to the control state within weeks or months, depending on the intensity of the disturbance. In our survey, M1 and M5 possessed similar community structure (Figures 4 and 5), which were two communities that were sampled 1-year apart. Although the sampling schedule for M5 was formulated to observe the effects on the community post-typhoon, the closest typhoon, *Nuri* at 2008, did not land and brought little precipitation to FTR. In addition, the typhoon before *Nuri* occurred 1 month earlier to M5 sampling (Figure 2), suggesting that the resilience of the local community could occur when no severe disturbance like typhoon takes place within a period of time long enough in summer, for example, 1 month. The 6-year survey in surface ocean bacterial community suggested that the most common and abundant bacterial OTUs have temporal niches defined by broad seasonal changes, and the most variable OTUs can be temporally defined as well as by nutrient pulses and changes in currents (Gilbert *et al.*, 2012). Thus, the similar community composition shared by M1 and M5 may represent a typical community structure having temporal niche defined by constant summer conditions, whereas typhoons in summer may result in a short-term niche by changing limnological parameters and introducing exotic species that disrupt community compositions, which are plausibly explaining the exceptional community of M2. For metagenomic contents, typhoons can perform similar disruption by introducing foreign DNAs to the FTR freshwater. After the passage of typhoons, the homogenization of metagenomes could happen in the enclosed ecosystem if no severe disturbance occurred in a period of time, which thus accounts for the similar genomic make-up shared by M4 and

M5 (based on the %GC and principal component analysis). However, how the metagenome was homogenized (or became similar) after typhoon is still elusive. We hypothesized that it would be partly contributed by the constrained environmental natures and constant seasonal condition.

FTR comprises *Actinobacteria*, *Alphaproteobacteria*, *Betaproteobacteria*, *Cyanobacteria*, *Flavobacteria* and *Sphingobacteria* as major lineages accounting for 77.9% abundance on average. The exceptional dominance of *Flavobacteria* in M2 can be explained by the potential association between them and cyanobacteria. *Flavobacteria* reportedly thrive after a cyanobacteria bloom (Grossart *et al.*, 2005; Berg *et al.*, 2009) because they are proficient in degrading polysaccharides released by cyanobacteria (Kirchman, 2002). Suggested by the enumeration data (Supplementary Materials and methods), we found that a cyanobacteria bloom occurred before M2 but not before other samplings (Supplementary Figure S14); thus, the successive pattern 'cyanobacteria first, then flavobacteria' is likely applicable in FTR. Besides that, the typhoon has been reported to deliver oxygen-rich water in FTR (Fan and Kao, 2008) and a higher oxygen concentration was found in M2 (Supplementary Table S2). Thus, we speculated that the oxygen also in part contributed to the dominance of *Flavobacteria* at that time.

Based on the results of multivariate analysis of variance, the precipitation at FTR had an important role in the community dynamics, in agreement with the suggestion that the extreme precipitation is one of the proposed disturbances shifting microbial community (Shade *et al.*, 2012). This is a reasonable observation since the runoff from catchments brings terrestrial bacteria along with various kinds of nutrients to the reservoir. Both direct (i.e., new bacteria) and indirect (i.e., nutrients) disturbances contribute to the alternations in the FTR microbial community. According to the geographical location of Taiwan, seasonal precipitations (i.e., plum rain in late spring, typhoons in summer and autumn, and northeastern monsoons in winter) thus make FTR a frequently disturbed ecosystem. As suggested by the lack of consistent community structure over this 2-year survey, more intensive sampling is required to fully capture the rapidly changing dynamics of the community.

Similar to the decoupling between microbial community composition and metabolism found in the open ocean (Hewson *et al.*, 2010) and bacteria on green macroalga (Burke *et al.*, 2011), we found that M1 to M6 COG family profiles are less divergent (*L*-divergence 0.24 ± 0.04) than the genera composition (0.73 ± 0.25). Because genomic diversification is important for microbes to adapt to different environments (Sanchez-Perez *et al.*, 2008; Wiedenbeck and Cohan, 2011; Wu *et al.*, 2011), the enriched transposase in microbial metagenomes not only indicates the high genomic recombination activities

in local community, but also suggests the important role of genomic recombination for microbes surviving in such an ecosystem with frequent disturbances. Moreover, the sparsely detectable clustered regularly interspaced short palindromic repeats (Supplementary Table S5)—which are genetic loci that immunizes prokaryotes against exogenous genetic elements (Marraffini and Sontheimer, 2010)—may also serve as another way of improving microbial fitness in FTR by lowering resistance to foreign DNAs.

Gene enrichment analysis demonstrated that there were niche-specific relationships between environmental conditions and gene function by comparing FTR with other ecosystems. For example, in FTR, enriched antibiotic genes and transporters reflected the need to cope with antibiotics in the surrounding water. Furthermore, Poisson regression revealed dynamic associations between gene abundance and environmental changes. A good example is that the copy number of alkaline phosphatase increased when phosphate concentration decreased, delineating the selection of phosphate availability on microbes that are more efficient in phosphate scavenging.

In this 2-year survey, viral abundance followed seasonal variations, increasing in summer and decreasing in winter (Supplementary Figure S15). This is suspected to be caused by the low host activity (i.e., productivity) and abundance in winter (Supplementary Figure S15). The enrichment of viral integrase in the FTR metaviromes could result from the comparison against marine viral communities, in which the composed phages seemingly lack recognizable integrase in their genomes (Paul and Sullivan, 2005) and may involve the higher diversity in their lysogeny mechanisms (Breitbart, 2012). In addition to bacteriophages, we also detected terrestrial eukaryotic (i.e., plant and animal) viruses in FTR, varying from 2.9 to 27.9% in proportion, which is higher than desert ponds (Fancello *et al.*, 2012). One of the major sources of terrestrial plant and animal viruses is the catchment area. In support of this statement, we found that the total abundance of animal and plant viruses confluctuated with precipitation levels at FTR 1 week before sampling (Figure 8), although with moderate significance, suggesting that seasonal climatic activities with continuous or heavy precipitations (e.g., plum rain, typhoons) can disturb viral communities by bringing exotic viruses to FTR.

Conclusions

The outcome of metagenomic studies has successfully revealed tremendous valuable knowledge about the genomics and ecology of uncultured microbes. However, to the best of our knowledge, there is still limited long-term and time-series metagenomic data that simultaneously probes the

microbial community, metabolic profiles and related environmental changes (brought by extreme climatic activities). The data of environmental viruses is even more limited at the same experimental scale. This study is the first to analyze a subtropical freshwater ecosystem. The insights provided by this study on a freshwater ecosystem add much needed knowledge to the area, which currently lags behind in ocean and temperate ecosystem research.

Here, the microbial and viral metagenomes in a subtropical reservoir in Taiwan were studied in two consecutive summers and winters. The increase in microbial richness of all post-typhoon samples (i.e., M2, M4 and M5) could be one of the regular ecological successions associated with typhoon events. Among various environmental variables, precipitation was found to be the major factor affecting microbial community. The reason of smaller divergence in functional profiles compared with microbial community is still unknown. We speculate that both abiotic (i.e., constrained environmental conditions) and biotic factors (i.e., frequent microbial genome recombination) could be both the selection forces contributing to this phenomenon. In contrast to microbes, viral communities showed little association with typhoon events. However, viral abundance and genotype diversity was lower in winter (i.e., V3 and V6). The intermittent addition of terrestrial viruses into the community was plausibly associated with precipitation 1 week before sampling. As such, the viral community is presumably diversified by climatic activities that bring continuous or torrential rains. It is not clear that the terrestrial viruses are actually part of the aquatic community or whether they are simply a transiently passing signal of freshwater input. To better understand this phenomenon further experiments are required to justify their roles in the local viral communities. These results indicate that climatic activities cause temporary changes to an environment through a multitude of complex and possibly inter-related factors. Therefore, future work will involve intensive sampling and research to untangle the ecological interaction among microbes, viruses and environmental changes.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

We thank Taipei FTR Administration Bureau for their help in sample collection. This research was supported by the National Science Council (NSC) in Taiwan through Grant NSC 97-2621-B-001-004-MY2 and the NSC T-WEBS project. CHT was supported in part by the Student Exchange Program of National Yang-Ming University, and Australian Research Council (Grant Number DP1096296).

References

- 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.
- Berg KA, Lyra C, Sivonen K, Paulin L, Suomalainen S, Tuomi P *et al.* (2009). High diversity of cultivable heterotrophic bacteria in association with cyanobacterial water blooms. *ISME J* **3**: 314–325.
- Bolhuis H, Stal LJ. (2011). Analysis of bacterial and archaeal diversity in coastal microbial mats using massive parallel 16S rRNA gene tag sequencing. *ISME J* **5**: 1701–1712.
- Breitbart M. (2012). Marine viruses: truth or dare. *Ann Rev Mar Sci* **4**: 425–448.
- Burke C, Steinberg P, Rusch D, Kjelleberg S, Thomas T. (2011). Bacterial community assembly based on functional genes rather than species. *Proc Natl Acad Sci USA* **108**: 14288–14293.
- Byju P, Kumar SP. (2011). Physical and biological response of the Arabian Sea to tropical cyclone Phyan and its implications. *Mar Environ Res* **71**: 325–330.
- Chang J, Chung CC, Gong GC. (1996). Influences of cyclones on chlorophyll a concentration and *Synechococcus* abundance in a subtropical western Pacific coastal ecosystem. *Mar Ecol Progr Ser* **140**: 199–205.
- Chen CP, Tseng CH, Chen CA, Tang SL. (2011). The dynamics of microbial partnerships in the coral *Isopora palifera*. *ISME J* **5**: 728–740.
- Chen YJC, Wu SC, Lee BS, Hung CC. (2006). Behavior of storm-induced suspension interflow in subtropical Feitsui Reservoir, Taiwan. *Limnol Oceanogr* **51**: 1125–1133.
- Chenard C, Suttle CA. (2008). Phylogenetic diversity of sequences of cyanophage photosynthetic gene *psbA* in marine and freshwaters. *Appl Environ Microbiol* **74**: 5317–5324.
- DeLong EF, Karl DM. (2005). Genomic perspectives in microbial oceanography. *Nature* **437**: 336–342.
- Dinsdale EA, Edwards RA, Hall D, Angly F, Breitbart M, Brulc JM *et al.* (2008). Functional metagenomic profiling of nine biomes. *Nature* **452**: 629–632.
- Djikeng A, Kuzmickas R, Anderson NG, Spiro DJ. (2009). Metagenomic analysis of RNA viruses in a fresh water lake. *PLoS One* **4**: e7264.
- Edwards RA, Rohwer F. (2005). Viral metagenomics. *Nat Rev Microbiol* **3**: 504–510.
- Fan CW, Kao SJ. (2008). Effects of climate events driven hydrodynamics on dissolved oxygen in a subtropical deep reservoir in Taiwan. *Sci Tot Environ* **393**: 326–332.
- Fancello L, Trape S, Robert C, Boyer M, Popgeorgiev N, Raoult D *et al.* (2012). Viruses in the desert: a metagenomic survey of viral communities in four perennial ponds of the Mauritanian Sahara. *ISME J* **7**: 359–369.
- Gilbert JA, Steele JA, Caporaso JG, Steinbruck L, Reeder J, Temperton B *et al.* (2012). Defining seasonal marine microbial community dynamics. *ISME J* **6**: 298–308.
- Grossart HP, Levold F, Allgaier M, Simon M, Brinkhoff T. (2005). Marine diatom species harbour distinct bacterial communities. *Environ Microbiol* **7**: 860–873.
- Hamady M, Lozupone C, Knight R. (2010). Fast UniFrac: facilitating high-throughput phylogenetic analyses of microbial communities including analysis of pyrosequencing and PhyloChip data. *ISME J* **4**: 17–27.
- Hamp TJ, Jones WJ, Fodor AA. (2009). Effects of experimental choices and analysis noise on surveys of the ‘rare biosphere’. *Appl Environ Microbiol* **75**: 3263–3270.
- Hewson I, Poretsky RS, Tripp HJ, Montoya JP, Zehr JP. (2010). Spatial patterns and light-driven variation of microbial population gene expression in surface waters of the oligotrophic open ocean. *Environ Microbiol* **12**: 1940–1956.
- Hung CC, Gong GC. (2011). Biogeochemical responses in the Southern East China Sea after typhoons. *Oceanography* **24**: 42–51.
- Jardillier L, Zubkov MV, Pearman J, Scanlan DJ. (2010). Significant CO₂ fixation by small prymnesiophytes in the subtropical and tropical northeast Atlantic Ocean. *ISME J* **4**: 1180–1192.
- Jones SE, Chiu CY, Kratz TK, Wu JT, Shade A, McMahon KD. (2008). Typhoons initiate predictable change in aquatic bacterial communities. *Limnol Oceanogr* **53**: 1319–1326.
- Kent WJ. (2002). BLAT—the BLAST-like alignment tool. *Genome Res* **12**: 656–664.
- Kim KH, Bae JW. (2011). Amplification methods bias metagenomic libraries of uncultured single-stranded and double-stranded DNA viruses. *Appl Environ Microbiol* **77**: 7663–7668.
- Kirchman DL. (2002). The ecology of Cytophaga–Flavobacteria in aquatic environments. *FEMS Microbiol Ecol* **39**: 91–100.
- Kristiansson E, Hugenholtz P, Dalevi D. (2009). Shotgun-FunctionalizeR: an R-package for functional comparison of metagenomes. *Bioinformatics* **25**: 2737–2738.
- Lin JH. (1991). Divergence measures based on the Shannon entropy. *IEEE Trans Inform Theory* **37**: 145–151.
- Lopez-Bueno A, Tamames J, Velazquez D, Moya A, Quesada A, Alcamí A. (2009). High diversity of the viral community from an Antarctic Lake. *Science* **326**: 858–861.
- Marraffini LA, Sontheimer EJ. (2010). CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat Rev Genet* **11**: 181–190.
- Ogawa H, Tanoue E. (2003). Dissolved organic matter in oceanic waters. *J Oceanogr* **59**: 129–147.
- Oksanen J, Kindt R, Legendre P, O’Hara B, Stevens MHH, Oksanen MJ *et al.* (2007). The vegan package. *Commun Ecol Package Version 1*: 8–5.
- Paerl HW, Valdes LM, Peierls BL, Adolf JE, Harding LW. (2006). Anthropogenic and climatic influences on the eutrophication of large estuarine ecosystems. *Limnol Oceanogr* **51**: 448–462.
- Partensky F, Hess WR, Vaulot D. (1999). *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. *Microbiol Mol Biol Rev* **63**: 106.
- Paul JH, Sullivan MB. (2005). Marine phage genomics: what have we learned? *Curr Opin Biotechnol* **16**: 299–307.
- Powell S, Szklarczyk D, Trachana K, Roth A, Kuhn M, Muller J *et al.* (2012). eggNOG v3.0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges. *Nucleic Acids Res* **40**: D284–D289.
- Preiss J. (1984). Bacterial glycogen synthesis and its regulation. *Annu Rev Microbiol* **38**: 419–458.
- Rohwer F, Edwards R. (2002). The Phage Proteomic Tree: a genome-based taxonomy for phage. *J Bacteriol* **184**: 4529–4535.

- Roux S, Enault F, Robin A, Ravet V, Personnic S, Theil S *et al.* (2012). Assessing the diversity and specificity of two freshwater viral communities through metagenomics. *PLoS One* **7**: e33641.
- Sanchez-Perez G, Mira A, Nyiro G, Pasic L, Rodriguez-Valera F. (2008). Adapting to environmental changes using specialized paralogs. *Trends Genet* **24**: 154–158.
- Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB *et al.* (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75**: 7537–7541.
- Shade A, Read JS, Welkie DG, Kratz TK, Wu CH, McMahon KD. (2011). Resistance, resilience and recovery: aquatic bacterial dynamics after water column disturbance. *Environ Microbiol* **13**: 2752–2767.
- Shade A, Read JS, Youngblut ND, Fierer N, Knight R, Kratz TK *et al.* (2012). Lake microbial communities are resilient after a whole-ecosystem disturbance. *ISME J* **6**: 2153–2167.
- Shiah FK, Chung SW, Kao SJ, Gong GC, Liu KK. (2000). Biological and hydrographical responses to tropical cyclones (typhoons) in the continental shelf of the Taiwan Strait. *Cont Shelf Res* **20**: 2029–2044.
- Sousa WP. (1984). The role of disturbance in natural communities. *Annu Rev Ecol System* **15**: 353–391.
- Sullivan MB, Lindell D, Lee JA, Thompson LR, Bielawski JP, Chisholm SW. (2006). Prevalence and evolution of core photosystem II genes in marine cyanobacterial viruses and their hosts. *PLoS Biol* **4**: e234.
- Thurber RV, Haynes M, Breitbart M, Wegley L, Rohwer F. (2009). Laboratory procedures to generate viral metagenomes. *Nat Protoc* **4**: 470–483.
- Tseng YF, Hsu TC, Chen YL, Kao SJ, Wu JT, Lu JC *et al.* (2010). Typhoon effects on DOC dynamics in a phosphate-limited reservoir. *Aquat Microb Ecol* **60**: 247–260.
- Vinayachandran PN, Mathew S. (2003). Phytoplankton bloom in the Bay of Bengal during the northeast monsoon and its intensification by cyclones. *Geophys Res Lett* **30**: doi:10.1029/2002gl016717.
- Walker ND, Leben RR, Balasubramanian S. (2005). Hurricane-forced upwelling and chlorophyll a enhancement within cold-core cyclones in the Gulf of Mexico. *Geophys Res Lett* **32**: doi:10.1029/2005gl023716.
- Wang Q, Garrity GM, Tiedje JM, Cole JR. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* **73**: 5261–5267.
- Waterbury JB, Watson SW, Valois FW, Franks DG. (1986). Biological and ecological characterization of the marine unicellular cyano bacterium *Synechococcus*. *Can J Fish Aquat Sci* **214**: 71–120.
- West AJ, Lin CW, Lin TC, Hilton RG, Liu SH, Chang CT *et al.* (2011). Mobilization and transport of coarse woody debris to the oceans triggered by an extreme tropical storm. *Limnol Oceanogr* **56**: 77–85.
- Wiedenbeck J, Cohan FM. (2011). Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. *FEMS Microbiol Rev* **35**: 957–976.
- Wilson K. (2001). Preparation of genomic DNA from bacteria. *Curr Protoc Mol Biol* **Chapter 2**: Unit 2.4; doi:10.1002/0471142727.mb0204s56.
- Wu CC, Kuo YH. (1999). Typhoons affecting Taiwan: current understanding and future challenges. *Bull Am Meteorol Soc* **80**: 67–80.
- Wu XA, Monchy S, Taghavi S, Zhu W, Ramos J, van der Lelie D. (2011). Comparative genomics and functional analysis of niche-specific adaptation in *Pseudomonas putida*. *FEMS Microbiol Rev* **35**: 299–323.
- Zehetner F, Vemuri NL, Huh CA, Kao SJ, Hsu SC, Huang JC *et al.* (2008). Soil and phosphorus redistribution along a steep tea plantation in the Feitsui reservoir catchment of northern Taiwan. *Soil Sci Plant Nutr* **54**: 618–626.

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