Supplementary Information

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

Ching-Hung Tseng, Pei-Wen Chiang, Fuh-Kwo Shiah, Yi-Lung Chen, Jia-Rong Liou, Ting-Chang Hsu, Suhinthan Maheswararajah, Isaam Saeed, Saman Halgamuge, Sen-Lin Tang

Supplementary Materials and Methods

Supplementary Tables

Supplementary Figures

Supplementary Materials and Methods

Bacterial 16S rRNA gene clone library

The 16S ribosomal DNA (rDNA) was amplified by PCR using the bacterial universal primers 27F (5'-AGAGTTTGATCMTGGCTCAG-3') and 1492R (5'-TACGGYTACCTT-GTTACGACTT-3'). The quality-checked PCR products were purified (QIAquick gel extraction kit, Qiagen) and cloned into TOPO-TA cloning vector (TOP10 genotypes of *Escherichia coli*) by following the manufacturer's instructions (Invitrogen). Positive colonies were selected by blue-and-white screening. To avoid clone-picking bias, in each transformation plate with more than 1,000 positive colonies 200 were randomly numbered on the plate (50 colonies per one fourth area of plate) without preference for size. According to the random numbers generated by Microsoft Excel, 100 corresponding true-positive colonies (confirmed by M13 forward, 5'-GTAAAACGACGGCCAG-3', and reverse 5'-CAGGAAACAGCTATGAC-3', primers) were picked and sequenced. Putative chimeric sequences detected by at least two of Bellerophon (http://greengenes.lbl.gov/), ChimeraSlayer (http://sourceforge.net/projects/microbiomeutil/files/), and UCHIME (http://www.drive5.com/uchime/) were manually checked using the Ribosomal Database Project (RDP) Classifier (http://sourceforge.net/projects/rdp-classifier/), and excluded from further analyses if they appeared to be chimeric, leaving 86 and 96 clones in M1 and M2, respectively. These cloned sequences have been deposited to GeneBank under the accession numbers JN379110-JN379291.

Bacterial 16S rRNA gene V1-V2 hyper-variable region multiplex sequencing

Two bacterial universal primers, 27F (5'-AGAGTTTGATCMTGGCTCAG-3') and 341R (5'-TGCTGCCTCCCGTAG- G-3'), were used in the first round V1-V2 region PCR utilizing TaKaRa EX taq polymerase (Takara Bio). The PCR products were checked by polyacrylamide gel electrophoresis. The expected band (about 320 bp) was cut and extracted by electro-elution. Original primers were tagged with four extra nucleotides at 5' end in a second round PCR for distinguishing different samples. The tagged amplicons were purified (Qiagen PCR purification kit) and sent to pyrosequencing by the Roche 454 GS FLX Titanium System (Mission Biotech, Taipei). GS Run Processor (v2.5) with default setting was applied for reads quality control. Chimeric reads were identified by UCHIME and removed. Only reads without any ambiguous base (N) and ranging in size from 280 to 350 base pairs after tag removal were included for further analysis.

Foreign metagenomes used for BLAT analysis

Seven metagenomes were downloaded from CAMERA (http://camera.calit2.net/), including CAM_PROJ_AcidMine, CAM_PROJ_HumanGut, CAM_PROJ_FarmSoil, CAM_PROJ_WashingtonLake, CAM_PROJ_WhaleFall, JCVI_SMPL_1103283000026

(GS020), and CAMERA Viral Nucleotide Sequences.

Foreign COG profiles used for correspondence analysis (CA)

Fifty-four metagenomic COG annotation data (majorly are aquatic and soil environments) were downloaded from IMG/M (http://img.jgi.doe.gov/). Taxon object IDs (taxon_oid) are listed as below: 2100351014, 2084038011, 2140918017, 2081372007, 2061766008, 2100351015, 2140918027, 2084038019, 2014031002, 2013954000, 2016842008, 2013515002, 2119805007, 2100351008, 2014031005, 2014031006, 2015219000, 2014031004, 2016842003, 2013843003, 2014031003, 2015219002, 2015391001, 2015219001, 2010170004, 2016842005, 2014031007, 2013515001, 2013954001, 2077657023, 2077657024, 2140918001, 2016842004, 2013515000, 2199352004, 2199352003, 2199352005, 2088090031, 2199034001, 2199352001, 2199352000, 2199352002, 2222084007, 2001200001, 2124908043, 2124908044, 2124908038, 2124908040, 2140918006, 2124908041, 2119805008, 2162886008, 2088090029, and 2189573022. Seven global ocean sampling (GOS) metagenomes were annotated using CAMERA COG annotation workflow, including samples from Sargasso Sea (GS000A and GS000C), coastal ocean (GS008 and GS014), open ocean (GS018 and GS023), and freshwater Lake Gatun (GS020).

Foreign COG profiles used for functional enrichment and Poisson regression analysis

The COG annotation was obtained from CAMERA, using the "functional annotation with COG" workflow. Included metagenomes are: GS000A (Sargasso Sea; open ocean, 5m), GS000C (Sargasso Sea; open ocean, 5m), GS008 (Newport Harbor, Rhode Island, USA; coastal ocean, 1m), GS014 (South of Charleston, South Carolina; coastal ocean, 1m), GS018 (Rosario Bank; open ocean, 1.7m), GS020 (Lake Gatun, Panama; freshwater, 2m), and GS023 (30 miles from Cocos Island; open ocean, 2m). Four oceanic viral metagenomes include: V.ARC (Chukchi Sea (USA) & Canadian Arctic, 2m), V.BBC (Bay of British Columbia, 0m), V.GOM (Gulf of Mexico, 1m), and V.SAR (Sargasso Sea, Hydrostation S, 80m).

Viral abundance estimation

Since longer contigs tend to be assembled using more reads, and BLAST reports high-scoring segment pairs (HSP) only (i.e. local alignment), we calculated the abundance by dividing the number of reads (i.e. viral particle number) by the contig length (in bp) and multiplying by the number of aligned positions in each HSP (Equation S1).

$$\frac{\#\,\text{BLAST aligned position}}{\text{contig length}} \times \#\,\text{reads} \qquad \text{(Equation S1)}$$

Cyanobacteria cell number enumeration

Surface freshwater was collected and transferred back to the laboratory in 2 hours. The cyanobacteria number was enumerated using flow cytomerty CyFlow Space® (Partec, Germany) and analyzed by FloMax® software (Partec, Germany). Enumeration included cyanobacteria groups Pcy-A and Pcy-B (Crosbie et al 2003).

Supplementary Tables

Table S1. Summary of homology search results in NCBI-nr, COG, and BLAT analysis

	No. of ORF	% of ORF found hit		% of contig found hit
		NCBI-nr	COG	BLAT analysis
M1/V1	44 637/9 442	51.0/28.1	40.4/16.8	18.2/12.4
M2/V2	22 481/10 173	57.6/20.6	41.1/10.6	17.9/8.5
M3/V3	18 895/10 495	42.8/27.7	32.2/11.3	19.1/10.4
M4/V4	35 207/6 293	58.3/46.7	44.2/35.2	17.4/15.1
M5/V5	28 740/6 391	57.2/22.6	45.5/11.1	17.6/7.4
M6/V6	20 786/4 872	55.0/44.7	42.8/38.5	17.2/21.9
Average (M/V)	28 458/7 944	53.7/31.7	41.0/20.6	17.9/12.6

Table S2. Environmental and limnological characteristics at 5-m depth in FTR.

	M1	M2	M3	M4	M5	M6
Days since last typhoon	318	9	49	4	5	106
Typhoon precipitation (mm)	161.7	346.5	212.8	75.1	0.0	457.4
Typhoon wind speed (m s^{-1})	48.0	53.0	40.0	33.0	40.0	53.0
Past 1-week accumulated precipitation[a] (mm)	5.1	96.4	42.5	80.4	16.3	94.8
Past 2-week accumulated precipitation[a] (mm)	5.5	470.0	55.3	105.8	46.5	232.1
Temperature (°C)	31.4	28.3	19.5	30.1	30.5	19.0
Conductivity (mS/cm)	0.09	0.08	0.06	0.08	0.08	0.06
Salinity (psu)	0.04	0.04	0.03	0.04	0.04	0.04
O_2 (ppm)	7.64	9.45	7.99	7.35	6.92	6.99
pH	9.08	8.09	6.52	8.78	9.13	6.96
Chlorophyll-*a* (μg L^{-1})	1.97	2.41	3.33	2.74	3.09	1.74
NO_2^- (μM)	0.33	0.33	0.07	0.31	0.31	0.10
NO_3^- (μM)	13.75	17.58	60.97	40.21	35.79	37.61
PO_4^{3-} (μM)	0.08	0.09	0.08	0.01	0.01	0.04
Si (μM)	99.06	146.85	134.96	127.94	132.89	149.31
Total BB[b] (10^5 cells mL^{-1})	32.46	34.44	23.21	45.76	35.33	13.85
Cyanobacteria (10^5 cells mL^{-1})	1.12	1.81	2.31	1.02	1.09	1.11
BP[b] (mgC m^{-3} d^{-1})	7.29	17.43	2.73	2.17	4.97	2.27

[a]Data are available on http://w2.feitsui.gov.tw/opr/waterdata2.htm
[b]BB: Bacterial Biomass; BP: Bacterial Production.

Table S3. Pearson's correlation between microbial communities inferred from 16S multiplex sequencing and metagenomic data. Total 30 classes were involved in this comparison.

	r	r^2	*P*-value
M1	0.754	0.568	9.74E-07
M2	0.902	0.814	4.02E-12
M3	0.879	0.773	7.57E-11
M4	0.605	0.366	3.15E-04
M5	0.745	0.555	1.55E-06
M6	0.651	0.424	7.31E-05

Table S4. Sequence information and diversity estimates as represented in V1-V2 multiplex sequencing[a]

	M1	M2	M3	M4	M5	M6
S[b]	470	800	575	753	884	600
Singleton OTU	234	372	277	347	427	293
N[c]	2 246	8 885	3 981	7 552	8 184	3 534
Evenness[d]	0.85	0.67	0.82	0.76	0.79	0.81
Richness[e]	69.52	93.96	76.67	89.22	108.87	82.29
Shannon	5.23	4.49	5.21	5.03	5.34	5.18
Chao 1[f]	824 (719, 973)	1 352 (1 217, 1 530)	995 (879, 1 156)	1 245 (1 122, 1 409)	1 636 (1 459, 1 867)	1 000 (893, 1 146)
Simpson	0.011	0.046	0.013	0.015	0.014	0.015

[a]Calculations were based on OTUs formed at an evolutionary distance of 0.03 (or about 97% identity).
[b]*S* =the number of OTUs
[c]*N* =the number of sequences
[d]Evenness =Shannon/ln(the number of OTUs)
[e]Richness =(number of singleton OTUs-1)/$\log_{10}$N. The maximum value is (N-1)/$\log_{10}$N.
[f]95% Confidence intervals estimators are shown in parenthesis.

Table S5. CRISPRs predicted in FTR metagenomes.

Sample	# CRISPR	Repeats in CRISPR
M1	2	(1) TCGGCGCATCTCCGCGTGGGCGGAGGAACCG (2) GTTTTCCCCACCCGCGTGGGGATGGCCC
M2	1	(1) AAGTTGTGGTTTAGTAAAAGATTGGAATCATAAATACT
M3	1	(1) AATATCGTATTTTCTAATCTTTAATCAAACCACAAC
M5	2	(1) GGAGTCTCCTCGCCTCGGAAGTGGAATCAAGGCAAAGCCG (2) GCCGTTTCAGTCCCCCTGAGGGGCGGACGGGGTTCGCGGCG

Supplementary Figures

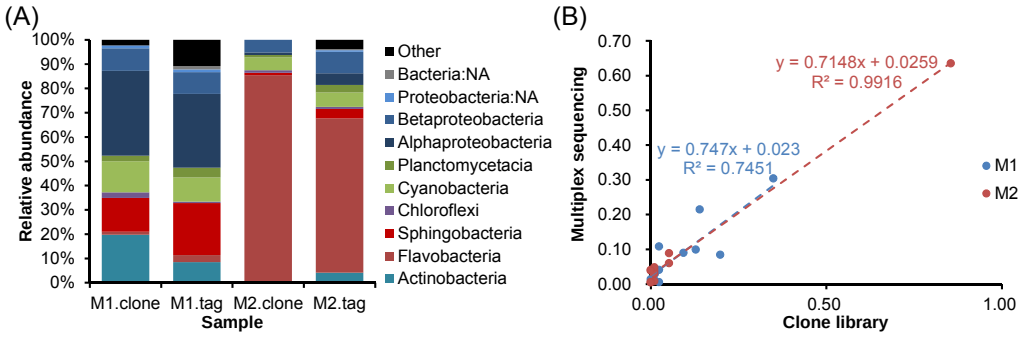

Figure S1. Reproducibility test for multiplex sequencing. (A) Microbial communities derived from multiplex sequencing (M1.tag and M2.tag) and clone library (M1.clone and M2.clone) were compared. (B) The data of relative abundance were fitted using linear regression.

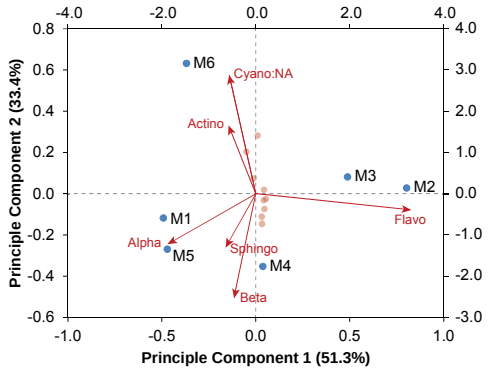

Figure S2. Principle component analysis of microbial community. Bacterial class name is labeled using abbreviations below; Alpha is short for *Alphaproteobacteria*, Beta for *Betaproteobacteria*, Sphingo for *Sphingobacteria*, Flavo for *Flavobacteria*, Actino for *Actinobacteria*, and Cyano for *Cyanobacteria*.

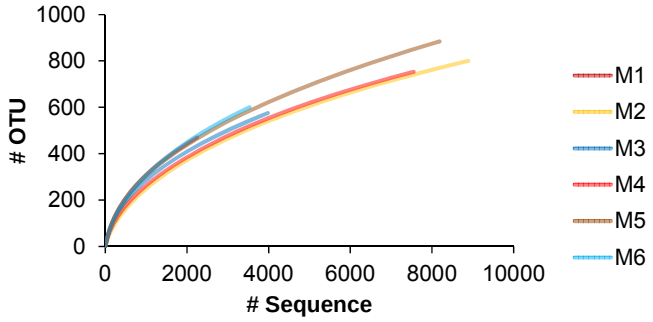

Figure S3. Rarefaction curve of microbial communities.

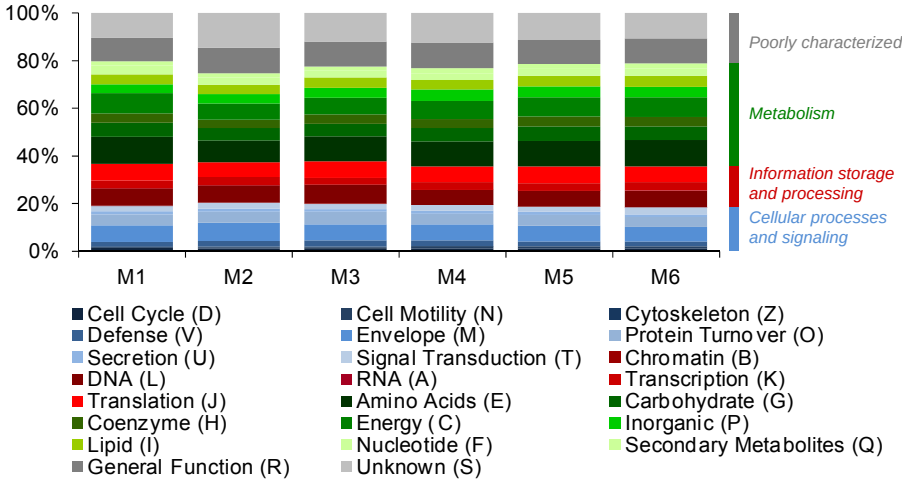

Figure S4. Microbial functional profiles based on COG categories. One-character abbreviation is denoted in the parenthesis.

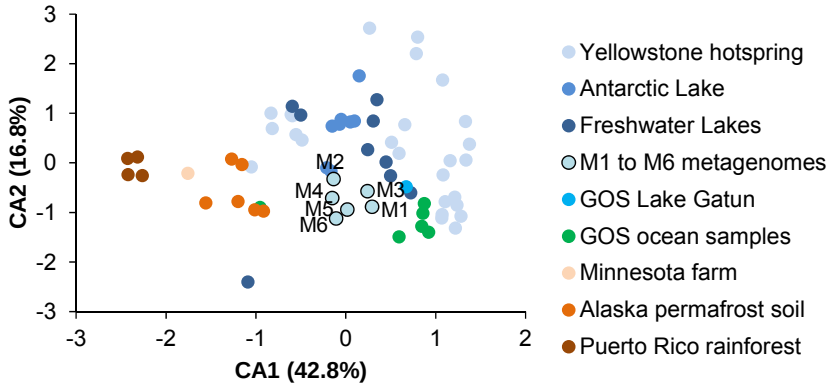

Figure S5. Correspondence analysis of COG class profiles. Metagenomes taken from lake, soil, and ocean are respectively colored with blue, brown, and green.

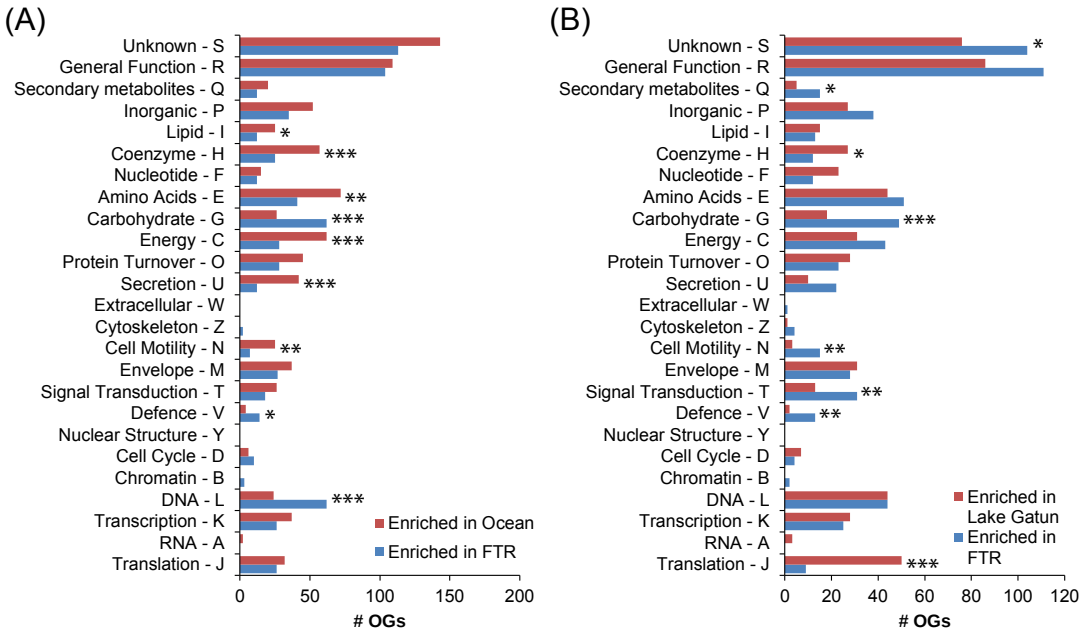

Figure S6. Enriched functional profile in comparisons of FTR against (A) oceanic metagenomes and (B) Lake Gatun. Two-tailed binominal test significance was denoted by asterisks: *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

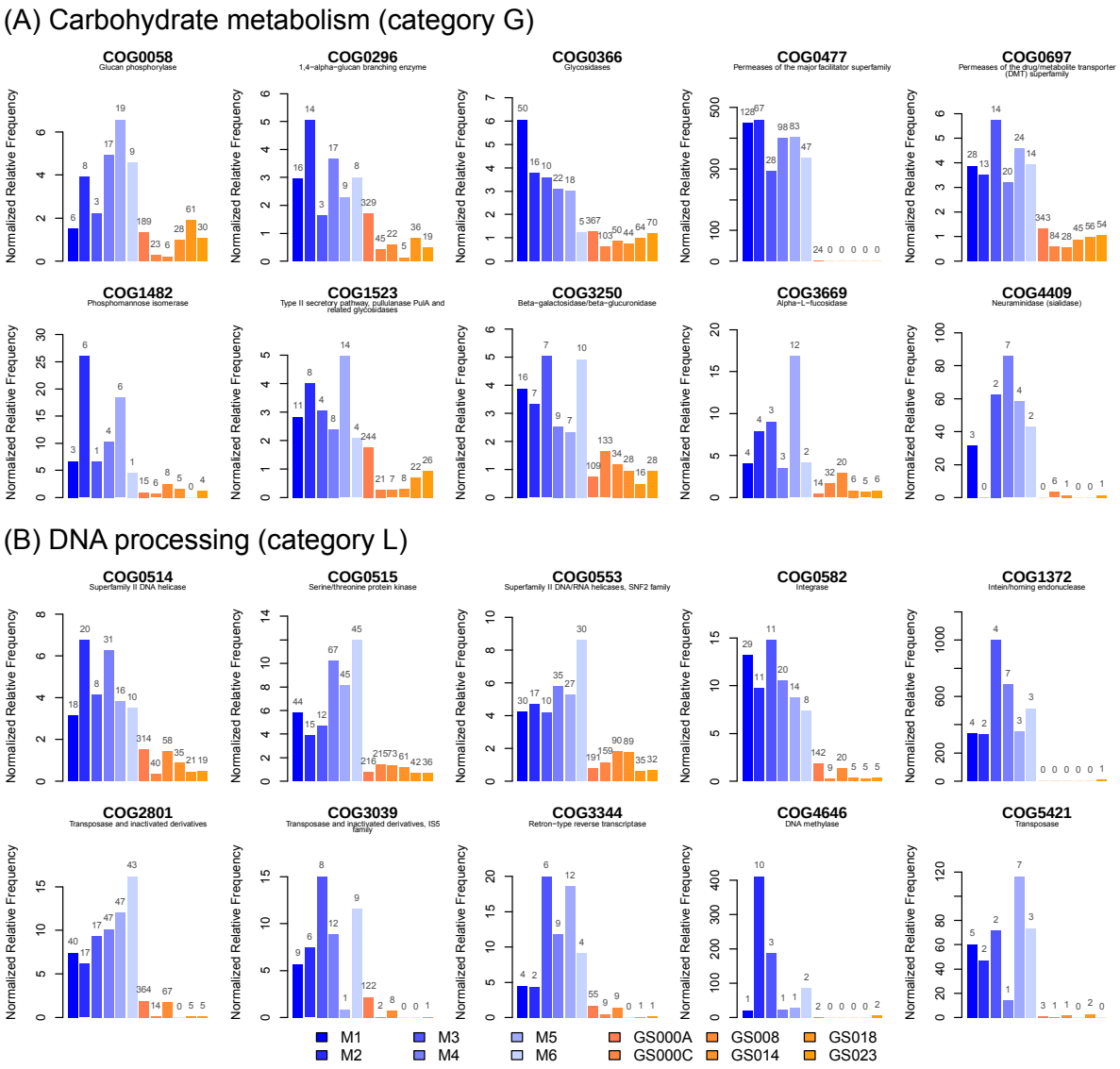

Figure S7. Top-10 enriched functions in microbial metagenomes of (A) carbohydrate metabolisms and (B) DNA processing in FTR, compared to oceanic metagenomes. The original counts are labeled on top of each bar.

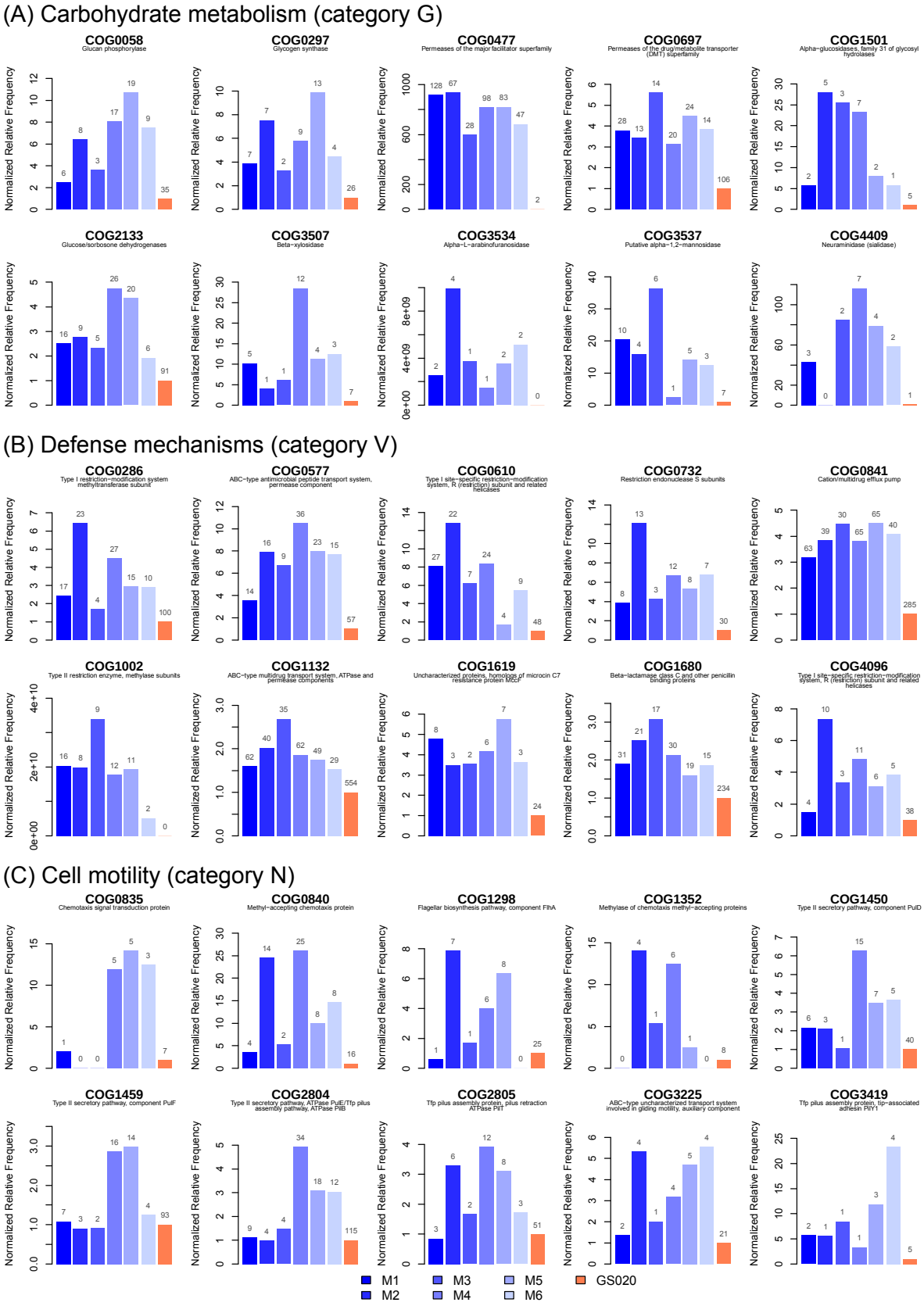

Figure S8. Top-10 enriched functions in microbial metagenomes of (A) carbohydrate metabolisms, (B) defense mechanisms, and (C) cell motility in FTR, compared to the Lake Gatun (GS020) metagenome. The original counts are labeled on top of each bar.

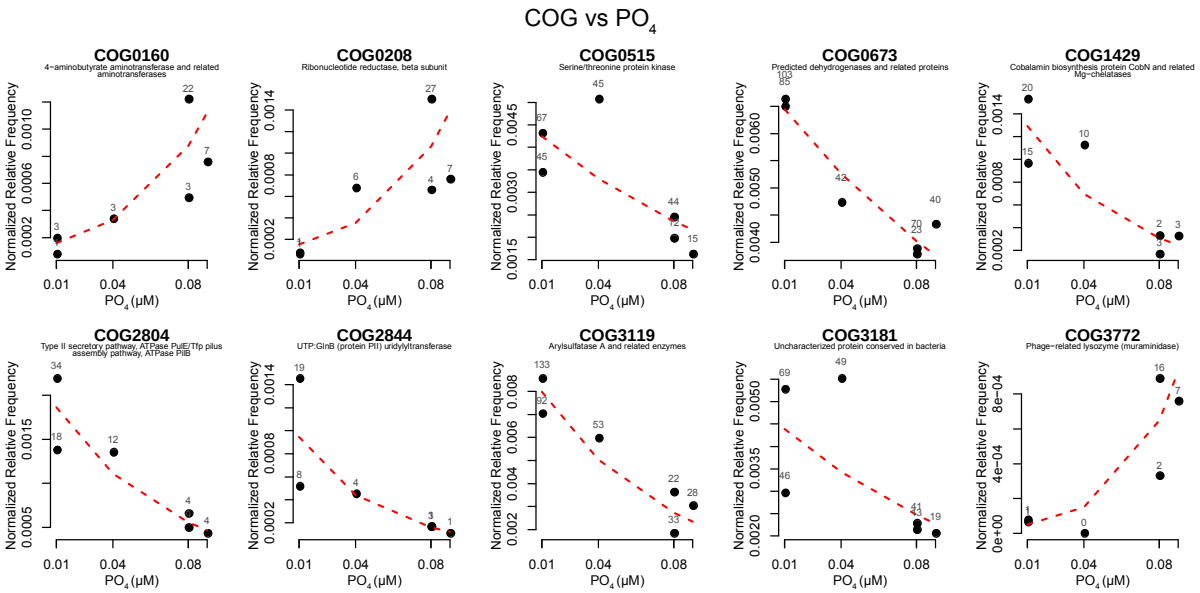

Figure S9. Top-10 OGs in microbial metagenomes correlated with phosphate concentration using Poisson regression (P<0.05). The original counts are labeled on each data point. Regression lines are shown as dashed lines in red.

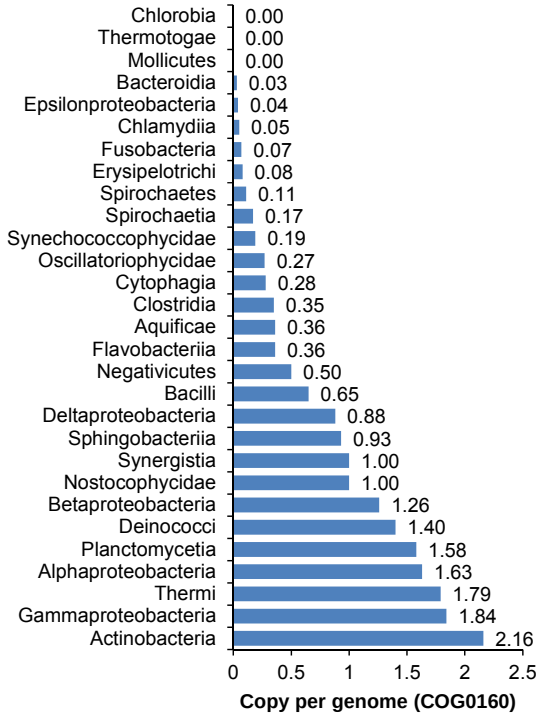

Figure S10. COG0160 gene copy number in bacteria. Data were downloaded from IMG database (http://img.jgi.doe.gov/). Bacterial classes with less 10 available genomes were excluded.

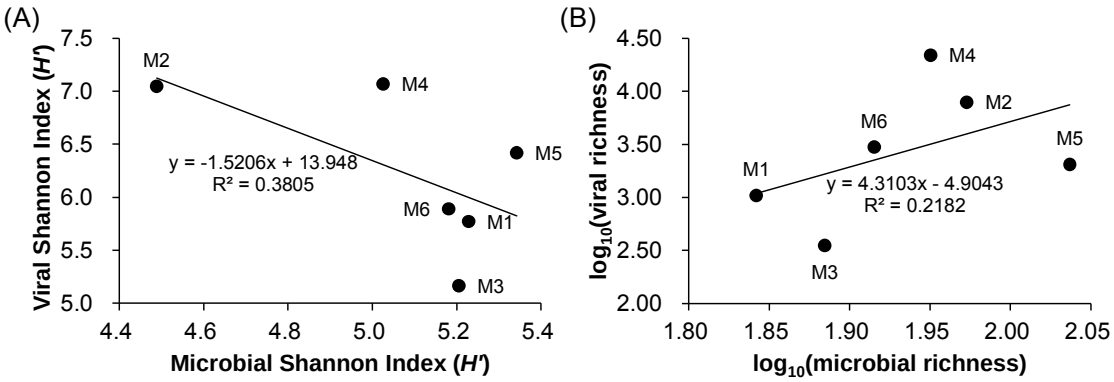

Figure S11. Correlation plots of (A) Shannon Index and (B) richness data in paired microbial and viral communities. Both of them have *P*-values >0.05 (0.190 for Shannon Index, 0.350 for richness) in correlation tests.

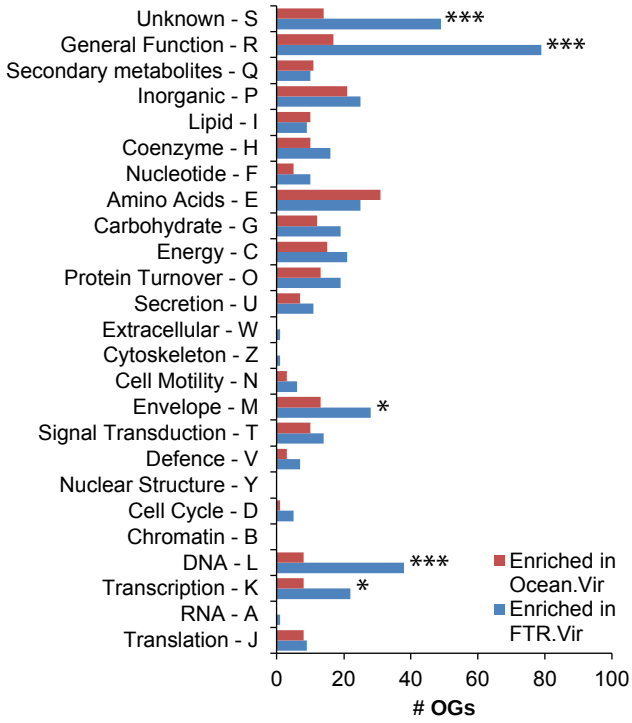

Figure S12. Enriched functional profile of viral metagenomes in comparison of FTR and oceans. Two-tailed binominal test significance was denoted by asterisks: *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

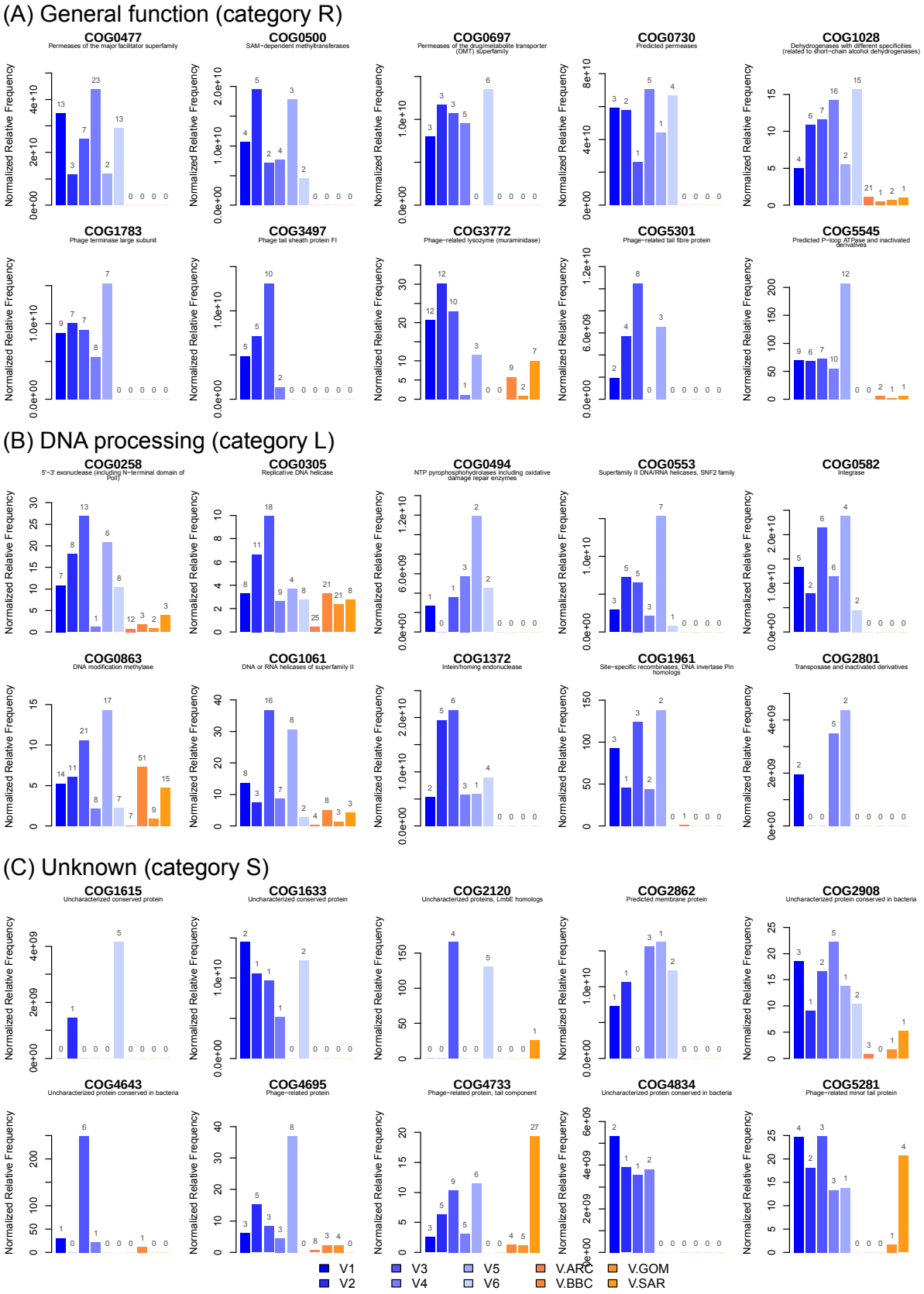

Figure S13. Top-10 enriched functions in viral metagenomes of (A) general function, (B) DNA processing, and (C) unknown in FTR, compared to oceanic viral metagenomes. The original counts are labeled on top of each bar.

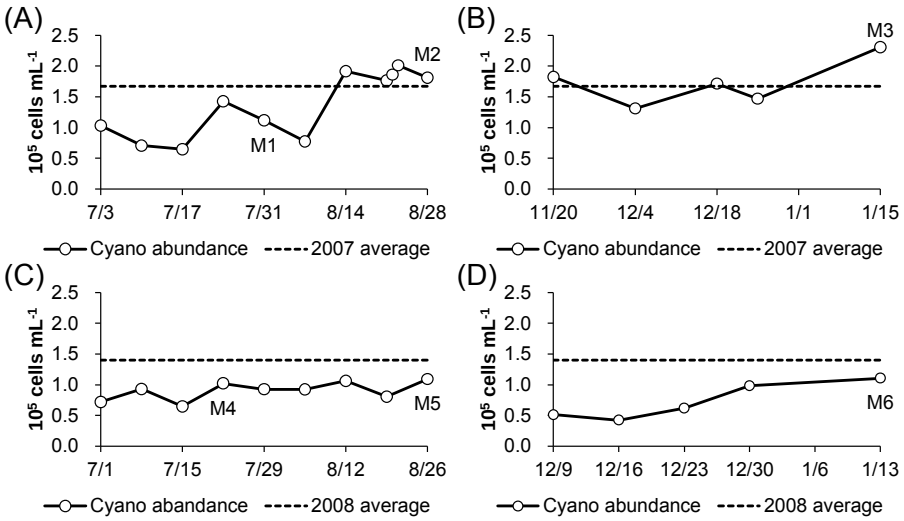

Figure S14. The cyanobacteria concentration before (A) M1 and M2, (B) M3, (C) M4 and M5, and (D) M6 in FTR.

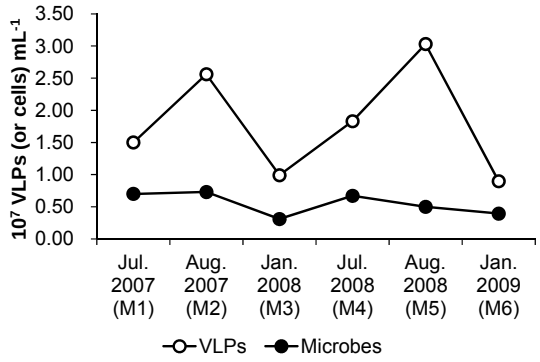

Figure S15. Concentration of microbes and virus-like particles (VLPs) in FTR. The number was determined by epifluorescence microscopy counting with SYBR Gold staining.

References

Crosbie ND, Teubner K, Weisse T (2003). Flow-cytometric mapping provides novel insights into the seasonal and vertical distributions of freshwater autotrophic picoplankton. *Aquatic Microbial Ecology* **33:** 53-66.