Comprehensive insightsinto composition, metabolic potentials and interactions of prokaryotic and viral assemblages in meromictic Lake Shunet in Siberia

Yu-Ting Wu[1,2], Cheng-Yu Yang[2], Pei-Wen Chiang[2], Ching-Hung Tseng[2,3],Hsiu-Hui Chiu[2],Isaam Saeed[4], Bayanmunkh Baatar[2,5,6], Denis Rogozin[7,8], Andrei Degermendzhi Georgievich[7], Sen-Lin Tang[2,3,5,6]

[1]Department of Forestry, National Pingtung University of Science and Technology, Taiwan

[2]Biodiversity Research Center, Academia Sinica, Taiwan

[3]Bioinformatics Program, Taiwan International Graduate Program (TIGP), Academia Sinica, Taiwan

[4]Department ofMechanical Engineering, Melbourne School of Engineering, The University ofMelbourne, VIC 3010, Parkville, Australia

[5]Molecular and Biological Agricultural Sciences, Taiwan International Graduate Program (TIGP), Academia Sinica, Taiwan

[6]Graduate Institute of Biotechnology and Department of Life Sciences, National Chung-Hsing University, Taiwan

[7]Institute of Biophysics, Siberian Division of Russian Academy of Sciences, Russia

[8]Siberian Federal University, Russia

Correspondence:

Sen-Lin Tang

Biodiversity Research Center, Academia Sinica,Taipei, 115,Taiwan

TEL: +886-2-27893863; FAX: +886-2-27890844; E-mail: sltang@gate.sinica.edu.tw

Conflict of Interest:The authors declare no conflict of interest.

Abstract

Meromictic lakes commonly have stratified biogeochemical characteristics at specific depths which are thought to be mainly shaped by microorganisms. However, microbial assemblages and their potential metabolic roles in meromictic lakesare poorly characterized. Therefore, we analyzed amplicon and shotgun metagenomic data to determinecomposition and capacity of prokaryotes and viruses along stratification of the water column in the landlocked meromicticLake Shunet(54° 25'N, 90° 13'E, ~ 6m deep)in southern Siberia.This lake hasa sharp gradient of physicochemical parameterswithmarked salinity(~ 71g l^{-1}) and H_2S (~ 400mg l^{-1}) inthemonimolimnion.Deep sequencing 265 gigabases, followed by high-quality assembly, identified a small group of *Flavobacteria* (*Nonlabens* nov.sp.)with a nearly complete genome and capability toproduce H_2S in a virus fraction (< 0.22 μm), 38 prokaryotic bins (0.2 – 5.3 megabases), and several viral contigs with similar maximal size at 5.0 m. Surface water had the most diverse prokaryotes and viruses, followed by the bottom water, and then the transition zone. Bacterial and archaeal communities were dominated by *Thiocapsa* and *Methanococcoides*, respectively, whereas *Siphoviridae* was the most abundant viral family.Variations in prokaryotic assemblages were consistent with physicochemical parameters, including salinity and availability of H_2S and NH_4^+. Reconstructed nutrient cycles of the three layers were apparently discriminated, corresponding to biogeochemical properties. Notably, several predominant unknown and uncultured bacteria were capable of energy generation, consistent with detection of highly

abundant untapped prokaryotes in rRNA pyrotags. Viruses at 3.0 m had specificity to prokaryotes at the same depth, whereas no pattern of interaction was apparent at 5.0 or 5.5 m, perhaps due to highly diverse viruses in mixolimnion. In conclusion,prokaryotic community structures, including unknowns, were explained by environmental variables and also mediated nutrient cycling.Furthermore, viral and prokaryotic assemblages interacted, revealing a pattern associated with dissolved oxygen concentrations in this specific lake ecosystem.

Introduction

Meromictic lakes have a mixed upper oxic mixolimnion, an interface chemocline, and a lower stagnant anoxic monimolimnion, which never mixes with upper layers (Bowman *et al.,* 2000). Their specific zones with stratified and unique biogeochemical characteristics provide novel opportunities to compare microbial loops and biogeochemical processes among distinct zones (Bowman *et al*. 2000; Comeau *et al.,* 2012; Lauro *et al.,* 2011).

Lake Shunet (54° 25'N, 90° 13'E), located in the central Altai-Sayan Mountain region in southern Siberia, is one of the only three meromictic lakes in the entire Asian part of Russia. This lake hassharp gradients ofsalinity, dissolved oxygen (D.O.) and hydrogen sulfide (H_2S), and a dense population of anoxygenic phototrophic (*i.e.*, purple sulfur) bacteria in the chemocline zone, second only to Lake Mahoney (Canada) of all known lakes(Rogozin *et al.,* 2012).With its stable stratification, Lake Shunet provides a good opportunity for comparative studies on biogeochemistry (Kallistova *et al.,* 2006), physicochemical parameters (Degermendzhy *et al.,* 2010), and microbial communities(Lunina *et al.,* 2007; Rogozin *et al.,* 2009; Rogozin *et al.,* 2010; Rogozin *et al.,* 2012). Previous studies to characterize microbes in Lake Shunet have used microscopic techniques, pigment analysis, culture, and PCR-DGGE, which have not provided a fine-scale comprehensive insight into microbial communityassemblages and their potential functions. Previous studies that included both microbial populations (metagenetics) and functional microbial processes (metagenomics) in meromictic salt lakes with whole genome

shot-gun approaches have only been done in Antarctica, i.e.,ACE Lake (Lauro *et al.,* 2011; Ng *et al.,* 2010) and Organic Lake (Yau *et al.,* 2011).

Similar toACE Lake(68° S) in Antarctica (Rankin *et al.,* 1999), Lake Shunet (54° N) is also a saline, high-altitude meromictic lake. The ACE Lake is one of the most investigated lakes in Antarctica,in terms of chemical-physical profiles, communitystructure, and functional dynamics (Laybourn-Parry and Bell, 2014). Some characteristics of both lakes are similar, including permanent water stratification into three main zones, rich salinity and H_2S in monimolimnion and a gradient of dissolved oxygen (Supplementary Table S1; Lauro *et al*., 2011; Laybourn-Parry and Bell, 2014).

Metagenetics and metagenomics have uncovered an unprecedented magnitude of microbial species, including dominant untapped ones and can provide a good assessment of microorganisms (Pedrós-Alió, 2012), community structure and ecological importance of environmental viruses (Edwards and Rohwer, 2005), in addition togene content,enabling prediction of functional roles in a variety of ecosystems, including deep sea (Sogin *et al.,* 2006), global oceans (Rusch *et al.,* 2007), meromictic lakes (Comeau *et al*., 2012; Lauro *et al*., 2011; Yau *et al.,* 2013) and underwater cave systems(Tetu *et al.,* 2013). Studies on marine viruses have demonstrated their contributions to themicrobial community and regulation of biogeochemical cycling (Breitbart, 2012; Suttle, 2007). However, there has been limited use of metagenomics to assess viral assemblages in lake ecosystems, let alone in meromictic lakes (Lauro*et al*., 2011).

Ecosystems of salt lakes have received less attention thanother aquatic ecosystems, such as oceans and freshwater lakes(Degermendzhy *et al.,* 2010). Therefore, we chose to study Lake Shunet,a typical high-latitudemeromictic salt lake. Objectives were to: (1) describe microbial (bacteria, archaea, and viruses) communities; (2) explore how gene contents corresponded to microbial community structures and what limnological parameters significantly explained microbial community structures over a depth profile; (3) reconstruct nutrient cycles through the depths; (4) investigate associations between gene contents involved in energy metabolism and abundant unclassified prokaryotic groups; and (5) unravel the history of potential interactions between viral and prokaryotic assemblages, based on detection of clustered regularly interspaced short palindromic repeats (CRISPR).

Materials and methods

Geographic description of study site and sampling procedures

Sampling was conducted in Lake Shunet (54° 25'N, 90° 13'E), 1.2 x 0.4 km, area 0.47 km^2and6.2 m deep, a hypersaline meromictic lake in The Republic of Khakasia in Siberia(Parnachev and Degermendzhy, 2002), in the north polar region. Lake Shunet is permanently stratified into an upper oxygenic layer (mixolimnion, 0.0-4.0 m), transition layer (chemocline, 5.0 ±0.2 m)between oxic and anoxic layers and a lower anoxic layer (monimolimnion, 5.5-6.0 m; Degermendzhy *et al.,* 2010, Kallistova *et al.,* 2006).

Sampling was done on July 21th, 2010. Approximately 20 L water was pumped from a specific depth of each layer (i.e., 3.0, 5.0, and 5.5 m), collectedin sterile containers and directly transported to ourlaboratory. Water samples were initially filtered through a10-μm plankton net and subsequently submitted to a Millipore-Pellicon TFF system (0.22-μm filter membrane) to separate microbes and viruses (Tseng *et al*., 2013).Microbes were retainedon 0.22-μm polycarbonate membrane filters and viruses were concentrated from the filtrate water using a 50-kDa filter.The viral fraction was treated with DNase I (New England Biolabs UK Ltd., Herts, UK) to digest DNAs, and virus-like particles were purified in cesium chloride gradients by ultra centrifugation at 1.35–1.50 g mL^{-1} density (Angly *et al.,* 2006; Tseng *et al*., 2013). Virus-like particles were observed using transmission electron microscopy (TEM).

Analysis of hydroparameters

The vertical profiles of temperature (Temp), dissolved oxygen (O_2), and pH were measured using Hydrolab Data-Sonde 4a (Hydrolab, Austin, TX, USA) and YSI 6600 (Yellow Springs,OH, USA) submersible profilers. Conductivity readings at *in situ* temperatures (C_t) were standardized to specific conductance at 25°C using $K_{25} = C_t \times (1+0.0204\times(T-25))^{-1}$, where T is the *in-situ* temperature in degrees Centigrade (Hydrolab, YSI). Salinity was calculated based on the relationship between salinity and conductivity, S (g L^{-1}) = 1.117 K_{25}- 7.9716. The formula for conductivity and salinity (determined as ash content) was derived from Lake Shira water, another

meromictic salt lake located 8 km northwest of LakeShunet. Conductivity sensors were calibrated against 0.2 3M KCl (Hydrolab, YSI) before each sampling. To determine sulphide concentrations, subsamples were fixed with zinc acetate and measured by the colourmetric method (Volkov and Zhabina, 1990). Total carbon (C_{tot}) and nitrogen (N_{tot}) were determined by CN–elemental analyzer FlashEA 1112 NC Soil/MAS 200 (Neolab LLC, USA). Ferric (Fe) and sodium (Na)were detected by inductively coupled plasma (Inductively Coupled Argon Plasma Optical Emission Spectrometer (ICP-OES).Phosphorus (P), NH_4^+, NO_2^-, NO_3^-, SO_4^{2-}, and total minerals (Tot_{miner}) were all measured as described (Kalachevaet *al.,* 2002).Hydroparameters were checkedfor collinearity using Spearman's rank correlation and a set of non-collinear parameters were maintained for further analyses (Supplementary Figure S1).

DNA extraction, establishment of archaeal/bacterial amplicon libraries, sequencing

The membrane filter with microbial retentate was rinsed withsterile double-distilled water and subsequently 10 ml Milli-Q water. The membrane filter was then transferred into a 50 ml Falcon tube for microbial DNA extraction using the cetyltrimethylammoniumbromide method(Wilson, 2001). ViralDNAswere extracted from purified virus-like particles using the formamide/cetyltrimethylammonium bromidemethod(Angly *et al.,* 2006) and checked by agarose gel electrophoresis. To collect enough DNA for sequencing, viral DNAs were amplified using aGenomiphi kit (GE Healthcare Life Science, Piscataway, NJ, USA).

A portion of microbial DNAs was used to analyze bacterial and archaeal communities, based on16S rRNA V4and V1-V2 hypervariable regions. Amplicon libraries were generated using the universal primer pair of 16S rRNA gene, 571F/ 751R for archaea and 27F/ 341R for bacteria,respectively. The PCR reaction was carried out in a total volume of 50 µlcontaining 2.5U *Superrun EX taq*™ HS, 5 µl of 10X*EX taq* Buffer, 200 µM of dNTPs, 0.2 µM of each primer, and 2-5 µg diluted template DNA (final concentration 100ng). The PCR program included an initial denaturation at 95°C for 5 min, followed by 30 cycles of 95°C for 20 s, 54°C for 10 s for archaea/52°C for 20 s for bacteria and 72°C for 20 s, with the final step of 72°C for 5 min and cooling at 4°C.Thereafter,PCR products were separated by 1% agarose gel electrophoresis with 1X TE buffer and SYBR® Green I. Bands of expected sizes (~180bp for archaea/315bp for bacteria) were cut from the gel and purified using a QIAEX II Gel Extraction Kit (QIAGEN Inc., Valencia, USA). Purified DNA fragments were quantified using a NanoDrop spectrophotometer (Thermo Scientific, Vantaa, Finland). In the second round of PCR, individual tags were added to 5´ends of 571F/751R and 27F/341R primers for each sample. The PCR mixture contained 2.5U *TaKaRa EX taq*™ HS (TaKaRa Bio, Otsu, Japan),5 µl of 10X*EX taq* Buffer, 200 µM of dNTPs, 0.4 µM of each tagged primer, and 100ng V4/ V1-V2 amplicon in a final volume of 50 µl.The PCRprogram for tag addition had an initial denaturation at 95°C for 5 min, followed by five cycles of 95°C for 20 s, 56°C for 10 s, and 72°C for 20 s, with the final step of 72°C for 5 min, and then cooling at 10°C.The PCR products were purified and a 200ng mixture of tagged V4/V1-V2

regions was subject to 454 pyrosequencing using the Roche GS454 FLX Titanium System (Roche 454 Life Sciences, Branford, CT, USA) at Mission Biotech (Taipei, Taiwan).

After quality trimming of sequences including chimera checking and removal of ambiguous nucleotide (N), mismatched primers, and incomplete barcodes, a total of 44,281/13,894 qualified bacterial/archaeal reads were retained. Thesequalified reads were sorted into subgroups according to specific barcodes, and then classifiedusing a Ribosomal Database Project classifier (v2.3) of 2015 with abootstrap value of 0.8 for taxonomic assignment of sequences(Wang *et al*., 2007). Chloroplast reads were removed from subsequent analyses. Reads of each sample were aligned with MUSCLE (http://www.drive5.com/muscle). The distance matrix was calculated using PHYLIP package (v3.69) and clustered inMOTHUR (v.1.14.0) (http://www.mothur.org/)to assign operational taxonomic unit (OTU) with a threshold of 97%sequence similarity(Schloss*et al*., 2009). Singletons were only involved in richness and diversity estimation; otherwise, they were excluded.

Sequencing, assembly and annotation of metagenomes

Both purified microbial and viral DNAs (i.e., metagenomes) were sequenced with anIlluminaHiSeq 2000 sequencing system **(**San Diego, CA, USA) atYourgene Bioscience (Taipei, Taiwan).Metagenomic reads with ambiguous nucleotides (>2), and short lengths (<35bp) were removed. Reads were subsequentlyassembled using *de novo*assembly algorithm of CLC

Genomics Workbench under criteria of 40-bp minimumoverlap and 99% consensus. Openreading frames (ORFs) were predicted from contigs by MetaGeneMark(http://exon.gatech.edu), and searched against the EggNOG (http://eggnog.embl.de) and Kyoto Encyclopedia of Genes and Genomes protein(Kanehisa *et al.,* 2012)databases using BLASTp (e-value $\leq 10^{-5}$and bits ≥ 100) for functional annotation according to the best match. The abundance of ORFs was initially calculated based on read count of coverage and then normalized by total sample size as relative abundance for further analyses. Data were log(x+1) transformed andCOG functional annotation wasevaluated using STAMP (Parks and Beiko, 2010).

Comparative metagenomic analyses

For comparative analysis, metagenomic reads of Ace Lake (filtered by 0.8 and 3.0 μm) were assembled using the assembly workflow on the CAMERA portal (https://portal.camera.calit2.net) with default settings. Then, ORF prediction and functional annotation were performed using the same strategy that was applied to Lake Shunet metagenomes (see above). The ORF abundance was normalized by total sample size as relative abundance for the analysis.Non-metric multidimensional scaling (NMDS) analysiswith Bray-Curtis distance metrics was usedto explore the similarity in COG profiles between ACE Lake and Lake Shunet.Statistical analyses were performed usingthe Vegan package (Oksanen*et al.,*2011)in R (http://www.r-projec-t.org).

Metagenomic binning and reconstruction of carbon, nitrogen, and sulfur cycles

Metagenomic contigs (>1000 bp) were involved in binning processes according to nucleotide compositions (Saeed *et al.,* 2012). The ORFs predicted from the contigs of microbial bins and Flavobacteria from the viral fraction were annotated using a BLASTp search against the KEGG proteindatabase (e-value $\leq 10^{-5}$) to reconstruct carbon (C), nitrogen (N), and sulfur (S) cycles in Lake Shunet. Relative abundance of each bin was initially calculated based on the read counts of coverage of ORFs and subsequentlynormalized by total sample size.

Bioinformatics analyses of viralcommunities

Viral richness estimates and Shannon-Wiener Index (*H'*) were determined using PHACCS (v1.1.3,http://sourceforce.net/ projects/phaccs), based on the contig spectragenerated by Circonspect (v0.2.5, http://sourceforge.net/projects/circonspect) with Cap3 assembler.Viral community composition was defined according to the homology search. Viral ORFs were searched against the NCBI RefSeq viral andmicrobial proteincollection (ftp://ftp.ncbi.nlm.nih.gov/refseq/release) using BLASTp(e-value $\leq 10^{-5}$, single best match) to assign taxonomic information. Nearly 1.06, 0.88 and 0.82% of ORFs in the respective viral metagenomes were assigned to viruses (Supplementary FigureS2).The final taxonomy of each contig was determined by voting of constituted ORFs. Relative abundance of each viral taxa was measured as described(Tseng *et al.,* 2013). Function of viral ORFs was annotated using

GhostKOALA of KEGG database.

Measurement of diameter of virus-like particles and analysis ofFlavobacterial genomes in viral fraction

The diameter of virus-like particle was measured from TEM pictures using NIS-Elements AR 4.0 of Nikon microscopy and calculated as follows:

$$\text{Diameter of virus-like particle} = \frac{\text{nm (scale bar)}}{\text{pixels (scale bar)}} \times \text{pixels of virus-like particles}$$

Several contigs derived from viral metagenomes (the largest was ~4.1 Mb) were hit to Flavobacteria. Analyses including 16S rRNA, reciprocal BLAST and genome completeness were processed to confirm characteristics of Flavobacteria. Furthermore, a reciprocal BLAST best hit was used to examine the similarity of Flavobacteria from microbial and viral fractions, with cutoff of e-value of 10^{-5} and bits ≥100. Hidden Markov Models (HMMs) was applied to estimate genome completeness using TIGRFAMs and Pfam libraries (Finn *et al.,* 2010, Haft *et al.,* 2003), whereas fragment recruitment analysis was doneusing MUMmer (http://mummer.sourceforge.net/manual/) withdefault settings. Ribosomal RNA genes were predicted withWebMGA(Huang *et al.,* 2009) and BLASTn search against the NCBI nucleotide databasewas performed to identify taxonomically related species. Based onBLASTn results,*Nonlabensdokdonensis* DSW-6 was the best hit(97% similarity), followed by*Psychroflexustorquis* ATCC 700755 (92% similarity) and Flavobacteria bacterium BBFL7

(92% similarity). The CLC Genomics Workbench (CLC bio) software was used to map contigs assigned to Flavobacteria into a draft genome. Contigs>10 kb were used for single-copy gene analysis to estimate genome completeness. A genome map was drawn withCircos(Krzywinski *et al.,* 2009), and annotated according to reference genomes (*Nonlabensdokdonensis* DSW-6, *Psychroflexustorquis* ATCC 700755 and Flavobacteria bacterium BBFL7).

Analyses of interactions between microbial and viral assemblages

Congruence between ordinations of viral and bacterial/ archaeal communities was quantified using the procrustean superimposition method andexamined for statistical significance using a Monte Carlo procedure (999 permutations).The m_{12} values, the goodness of fit statistic that measures congruence between two ordination configurations, were transformed to Procrustes correlation (r) by calculating the square root of their complements using the Vegan package(Oksanen *et al.,* 2011).Pearson correlation coefficients between viral and bacterial/archaealdiversity indices and richness estimate were calculatedwith the Psych package (Revelle, 2011).CRISPRon contigswere predictedby PILER-CR (v1.06, http://www.drive5.com/pilercr) with default settings. To predict potential interactions between host and virus, reciprocal BLAST analysis between microbial and viral metagenomicswas initially done, followed by ORF clustering, generated using CD-HIT with a threshold of 90%sequence similarity(Li and Godzik, 2006). Moreover, contigsof microbial bins were searched

against viral contigs to deduce potential host-virus interactions (BLASTn, e-value $\leq 8\times10^{-21}$, bits ≥100).

Statistical analyses

All of thefollowing statisticalanalyses were done with R software (Team RDC, 2011). Functional diversity was calculated using the R packageShotgunFunctionalizeR (Kristianssonet *al.,* 2009). Count data (including community structure and functional profiles) were transformed into relative abundance by dividing sample size. Hydroparameters were log (x+1) transformed for all analyses. The clustering heatmap wasgenerated usinga gplots package (Bolker*et al*., 2012).Differencesamong the three depths ingene contents involved in energy metabolism was examined based on a Wilcoxon test (Bauer, 1972), whereasa Student's *t*-test was used for testing differences of viral diversity, richness estimates and viral-like particle sizes. Nonmetric Multidimensional Scaling (NMDS) based on Bray-Curtis distances was performed using the Vegan packageto visualizedistribution of bacterial/archaeal community composition or the most abundant 50 gene functions involved in metabolism, followed by post-hoc regression of individual explanatory variables on the ordination scores. Goodness-of-fit values and their significancewere calculated using999 random permutations. Correspondence analysis (CA) (Perrière and Thioulouse, 2003) usingthe Vegan package was applied to compare variances of gene function involved in energy metabolism and the 10 most abundant unclassified

bacterial/archaeal OTUs. Only the top 50 genes based on the relative abundance were shown on the ordination plot.

Phylogenetic analysis

Neighbor-joining trees of functional gene *psbA* wereconstructed from 500 bootstrap iterations usingMEGA 6.0 (Tamura *et al.,* 2013). In addition to the four partial *psbA* genes obtained from virome data in this study, we also downloaded genes of various *Synechococcus* and *Prochlorococcus* phages from UniProt and NCBI databases. Phylogenetic trees were constructed separately, due to no common regions among the four partial *psbA* genes.

Results

Physicochemical profiles of lake water

The gradient of each hydro-parameter varied according todepth (Table 1). BothpH (8.1-6.7) and temperature (15.5-7.5°C) gradually decreased with depth. Furthermore, O_2, H_2S and salinity sharply changed across depth, in which O_2 was reduced from 10.9 mg $l^{-1}$to nearly zero in the chemocline, whereas salinity (~ 71g l^{-1}) and H_2S (~ 400 mgl^{-1}) were both much more markedin themonimolimnion (i.e., bottom water).

Prokaryotic community structure and diversity at three sampling depths

Three representative depths for mixolimnion, chemocline and monimolimnion layers were selected for a microbial community survey, according to hydrographical data. Massively parallel pyrosequencing of archaeal and bacterial 16S ribosomal RNA genes was used to characterize community structures. After tagging, quality and mosaic sequence trimming, qualified sequences were submitted to the RDA pipeline to define operational taxonomic unit (OTU) at 98% nucleotide identity. Following removal of singletons, a total of293 archaeal and 125bacterial OTUs wereidentified. Based on the Shannon's diversity index (*H'*), the 5.5 and 3.0 m depths had the most diverse bacteria (*H'*= 5.46) and archaea (*H'*= 6.24), respectively (Supplementary Table S2).Archaeal and bacterial communities were characterized by rarefaction curves, whereas bacterial communities had steeper asymptotes (Supplementary Figure S3).

Archaeal and bacterial community structures apparently differedthrough the three sampling

depths, comprised of3 phyla,5 classes and5 genera, and 9 phyla, 15 classes and35 genera (relative abundance > 0.1%), respectively.The archaeal communities were dominated by *Methanomicrobia* and by *Methanococcoides* at the class and genus levels, respectively. Among detected genera, *Nitrososphaera* and *Methanospirillum*were only present specifically in the mixolimnion (Figure 1a).Regardingbacterial composition,*Gammaproteobacteria* and *Thiocapsa* were the most predominant class and genus. The 3.0, 5.0 and 5.5 m-depthsweredominated by *Cyanobacteria,Thiocapsa* and *Halomonas*. Furthermore, some specific depth-dependent bacterial genera were also identified, including*Loktanella* in mixolimnion (3.0 m), *Enterobacter* in chemocline (5.0 m) and Candidatus*Cloacamonas* in monimolimnion (5.5 m; Figure 1b). Notably, predominant unclassified bacteria and archaea as well as those at unknown class and genus levels, were observed across the three depths (Figure 1a and b).

Functional profiles varied among depths

Using a deep sequencing strategy to uncover and comprehend potential metabolisms of microbial communities in Lake Shunet, between 26 and 72 Gb of sequencing sizes for each of six samples were generated (Supplementary TableS3), as well as a total of ~132 Gb of microbial and ~133 Gb of viral sequences. Consequently, between 121 Mbp and 14 Gbp of contigs were assembled in each of the six metagenomes. The largest microbial and viral contigswere0.47 and 0.09 Mbp, respectively (Supplementary Table S3). A total of 1,302,473 microbial and

1,379,639viral ORFs were predicted from the contigs. Approximately 36-40% of microbial ORFs and 3-21% of viral ORFs were annotated by either COGs or KEGG (Supplementary Table S3). There were 4356 COGs families, of which 71.9% were shared by the three depths, and 9.1, 3.7, and 2.5% were specific to respective depths (Supplementary Figure S4). Functional diversity index of metagenomicdata (Shannon index = 7.67-7.81), calculated using ShotgunFunctionalize R, was higher than the North Pacific Ocean (Shannon index = 6.46-6.93; DeLong *et al.,* 2006; Kristiansson *et al.,* 2009).

Based on clustering analysis, COGs profile of the 5.0-m metagenome was distinct from that at 3.0 and 5.5 m (Supplementary Figure S5),whereasthe 5.0 and 5.5 m were clustered based on KO profile of energy metabolism (Figure 2). According to COGs functional profiles, numerous genes were detected in specific categories, including replication, recombination and repair, amino acid transport and metabolism, cell wall/membrane/envelope biogenesis, and energy production and conversion (Supplementary Figure S5). In addition, STAMP analysis revealed that several COG categories including genefamilies and abundance of energy production and conversion differed ($P < 0.05$) among the three depths(Supplementary FigureS6).Furthermore, gene contents involved in energy metabolic subsystems differed among the three depths ($P< 0.05$; Figure 2). Based on their relative abundance,KO profiles involved in oxidative phosphorylation were the best-represented at both 3.0 and 5.0m-depths, followed by nitrogen metabolism, whereas KO profiles involved in methane metabolism were the best-represented at 5.5 m-depth, followed by carbon fixation

pathways in prokaryotes (Figure 2).

Relationships between metabolic functions and archaeal/bacterial community structure and hydroparameters

NMDS, based on Bray-Curtis distances with post hoc regression, were used to explore the correlation between the 50 most abundant genes involved in KO profiles of metabolism and prokaryotic assemblages. Five archaeal groups, including *Methanococcoides*, *Methanosphaera*, and *Methanolobus*, and bacterial groups including *Desulfobacterium*, *Desulfuromonadaceae*:NA, and *Thiocapsa*hadsignificant goodness of fit ($P \leq 0.05$) with regards to relative importance of the 50 most abundant genes involved in metabolism(Figures 3a and 3b). Specifically, *Methanococcoides*werecorrelatedwith K00163 (pyruvate dehydrogenase E1 component), K00366 (ferredoxin-nitrite reductase), K00266 (glutamate synthase (NADPH/NADH) small chain), and K03891 (ubiquinol-cytochrome c reductase cytochrome b subunit), whereas K10203 (elongation of very long chain fatty acids protein 6) was enriched by *Thiocapsa*.

In NMDS based on bacterial and archaeal community structures, followed by post hoc regression of individual environmental variables, parameters including H_2S, O_2, NH_4^+, NO_2^- and 11others,significantly explained the variance of both community structures along depths (Figures 3c and 3d). Specifically, *Methanococcoides* and *Methanolobus* were positively inclined with increases in N_{tot}(total nitrogen) and Temp (temperature), respectively (Figure 3c). The bacterial

clusters including *Desulfonatronum*, *Halothiobacillus*, *Desulfobacteraceae*:NA, *Desulfobacterium* and *Candidatus*_Cloacamonaswere enriched as increased in H_2S, whereas *Cyanobacteria*, *Marinobacter*, *Alcaligenaceae*:NA and *Planctomycetaceae*:NA correlated with changes in O_2 (Figure 3d).

Reconstruction of nutrient cycles through the depths

Many long-sized contigs obtained by deep sequencing greatly facilitated binning metagenomic data that enabled prediction ofpotential metabolic networks among microbial groups. A total of 14, 6, and 18 microbial bins (namelytaxonomically related groups) were identified in respective metagenomes of the three depths (Table 2). Metabolic pathways involved in nutrient cycling, includingcarbon, nitrogen, and sulfur cycles of individual microbial bins, were annotated to predict the nutrient cycles of the three layers (Figure 4). Each bin was assigned to serial numbers in the format of "depth-number" in the following descriptions.

Carbon cycling (respiration, aerobic and anaerobic carbon fixation) was reconstructed based on carbon fixation pathways, including reductive citric acid, hydroxypropionate-hydroxybutyrate cycles and Wood−Ljungdahl pathway. Groups of *Bacteroidetes*(3-1), *Flavobacteria* (3-2), *Pseudoalteromonas* (3-3), unknown bacterium (3-6), *Gammaproteobacteria* (3-8), *Rhodobacteraceae* (3-10), *Alcanivorax* (3-13), *Flavobacteria* (3-Flavo-in-vir), *Bacteroidetes* (5-4), and unculturedbacterium (5.5-3) carried genes that encoded enzymes to oxidize organic

carbon. Carbon fixation using oxygen as an electron acceptor was only detected in the bin of *Chroococcales* (3-12), the most abundant group in mixolimnion. *Hyphomonas neptunium*-like bacteria (3-11), *Enterobacteriaceae* (5-5), and *Halomonas* (5.5-17) were non-phototrophic groups thatharbored *prkB* or *rbcS* genes for autotrophic carbon fixation atrespective depths. Genes for anaerobic carbon fixation were identified in a variety of microbes, including*Halomonas* (3-9), *Flavobacteria* (3-Flavo-in-vir), *Clostridia* (5-1), *Bacteroidetes* (5-4), *Enterobacteriaceae* (5-5), uncultured bacterium (5.5-2),*Halanaerobium* (5.5-4), uncultured bacterium (5.5-6), *Deltaproteobacteria* (5.5-9), uncultured candidate division OP1 bacterium (5.5-10), *Marinilabiaceae* (5.5-13), and *Clostridiaceae* (5.5-16) through the three sampling depths, but particularly in anoxic monimolimnion. In addition, carbon monoxide (CO) was an intermediate in the Wood−Ljungdahl pathway of acetyl-CoA biosynthesis, possibly produced by the abundant unknown bacteria (3-4 and 3-6),*Rhodobacteraceae* (3-10), and uncultured bacterium (5.5-6).

Nitrogen cycling including denitrification, nitrogen fixation, and reduction of nitrate and nitrite were generally characterized at all sampling depths (Figure 4). In the 3.0-m sample, genes involved in denitrificationfor energy production were completely defined in the bins of*Pseudomonas stutzeri*-like bacteria(3-7), *Gammaproteobacteria*(3-8), *H. neptunium*-like bacteria (3-11), *Alcanivorax*(3-13), unknown bacterium (3-4) and*Flavobacteria* (3-Flavo-in-vir).*Bacteroidetes* (3-1), *Flavobacteria* (3-2), *Gammaproteobacteria* (3-8), *Halomonas* (3-9), *Alcanivorax* (3-13), and *Flavobacteria* (3-Flavo-in-vir) harbored genes for

nitrite reduction. *P.stutzeri*-like bacteria (3-7), *Gammaproteobacteria* (3-8), *H. neptunium*-like bacteria (3-11) and *Alcanivorax* (3-13) were potentially capable of reducing nitrate. Moreover, nitrogenase (EC 1.18.6.1 that processes nitrogen fixation, was only identified in bins of an unknown bacterium (3-6) and *Rhodobacteraceae*(3-10). In the 5.0 m sample, *Enterobacteriaceae* (5-5)as the secondly abundant group after *Thiocapsa* in the chemocline had genes for both nitrate reduction and nitrite oxidation, but*Bacteroidetes* (5-4) only contained genes involved in nitrite reduction.*Clostridia* (5-1) had genes involved in an alternative pathway to produce ammonium from hydroxylamine (NH_2OH)in nitrogen cycling. In the 5.5-m sample, onlyunculturedbacteria (5.5-3and 5.5-5) were detected with genes involved in nitrogen cycle, mediating nitrite reduction and nitrogen fixation, respectively.

Genes involved in the key processes in sulfur cycling including sulfite (SO_3^{2-}), thiosulfate ($S_2O_3^{2-}$) oxidation, sulfate (SO_4^{2-}) reduction, and sulfur disproportionation ($S_2O_3^{2-} \rightarrow H_2S + SO_4^{2-}$) were detected through depths (Figure 4). Pathways of thiosulfate oxidized to sulfate and sulfite were attributed to*Rhodobacteraceae* (3-10) and *Flavobacteria* (3-Flavo-in-vir) at 3.0 m, and *Thiocapsa*(5-6 and 5.5-18) at 5.0 and 5.5 m. Genes that oxidized sulfite to sulfate were identified inbins that included*Bacteroidetes* (3-1), *Deltaproteobacteria* (5.5-9), *Desulfobacteraceae* (5.5-12 and 5.5-15),*Chthoniobacterflavuz*-like bacteria (3-14), and *Flavobacteria* (3-Flavo-in-vir). Other bins such as *H. neptunium*-like bacteria (3-11), *Staphylococcus* (5-2), *Enterobacteriaceae* (5-5), *Thiocapsa*(5-6), unculturedbacterium (5.5-3), *Deltaproteobacteria* (5.5-9), and

Desulfobacteraceae (5.5-12 and 5.5-15) had capability to utilize sulfate as terminal electron acceptor to produce hydrogen sulfide (H_2S). *Bacteroidetes* (5-3) had genes to consume 3-phosphoadenylyl sulfate (PAPS) and sulfur disproportionation to produce sulfate.

By comparing pathways annotated from bins and total metagenomic data, pathways predicted from the 3.0-m metagenome were exactly defined in the 3.0-m bin sequences. However,a few pathways related to nitrogen fixation, denitrification and nitrate reduction at 5.0 and 5.5 m discovered in metagenomic data were absent incharacterized bins.

Virus-like particle size, viral genome size, diversity and function annotation

Sizes of virus-like particles was measured byTEM (Figure 5a) and significantly decreased across depths, with anaverage of ~32, 30, and 28 nm at each sampling depth (P<0.05; Figure 5b). A similar pattern was also observed in viral contig length that at 3.0 m was longer (28.9 kb on average) than 5.0 or 5.5 m(24.4 and 19.9 kbp, respectively; Supplementary FigureS7).

The viral community structure differed among the three sampling depths (Figure 6). A total of 14 abundant viral families (≥0.1% relative abundance) were detected through the water column, including several typical bacteriophage groups, such as *Siphoviridae*, *Microviridae*, *Myoviridae*, *Podoviridae*and three viral groups (*Nanoviridae*, *Circoviridae* and *Phycodnaviridae*). The 3.0-m viral assemblage was dominated by *Siphoviridae*(35.6%), followed by *Myoviridae,* whereasthe 5.0- and 5.5-m viral assemblages were dominated by *Microviridae*(34.4%) and

Nanoviridae(27.4%) and followed by*Nanoviridae*and*Siphoviridae*, respectively. Furthermore, many unknown viruses were detected through the three depths, with the greatest enrichment of 24.4% at 5.5 m (Figure 6).

Viral diversity determined with PHAACSdiffered significantly among the three depths (Table 3). The layer of 3.0m had the most genotype-rich (15877 genotypes) and diverse (H' = 9.15) viral assemblage, followed by 5.5 m (10000 genotypes and H'= 8.06) and 5.0 m (2498 genotypes and H'=6.70). The highest evenness estimate at 3.0 m indicated that all viral genotypes were close to being equally abundant, whereas some viral genotypes at 5.0 and 5.5 m were predominant. The lowest evenness occurred at 5.0 m viral community, and the most abundant genotype constituted 5.51% of the community.

Viral ORFs were mostly annotated to carbohydrate metabolism at 3.0 m and genetic information processing at both 5.0 and 5.5 m, based on the KEGG database (Supplementary FigureS8). Additionally, approximately4.0, 3.0, and 2.0% of viral ORFs were annotated in energy metabolism, respectively, including key genes involved in nitrogen metabolism, sulfur metabolism, photosynthesis and oxidative phosphorylation (Supplementary TableS4). Specifically, cyanophage-encoded auxiliary metabolic genesincluding four partial*psbA*genes were detected. In the phylogenetic analyses, only the gene_id_217671_5.0m clustered in an individual branch, close to D1_protein_Synechococcus_phage_S-SSM5 (Supplementary FigureS9).

Potential interactions between viral and prokaryotic assemblages

Viral diversity was positively correlated with both bacterial (r^2=0.558, P=0.46) and archaeal diversity (r^2=0.999, P<0.05) and richness estimates(bacteria: r^2=0.848, P=0.25; archaea: r^2=0.989, P=0.06; Supplementary FigureS10).

Procrustesanalyses revealed congruence among viral and bacterial (m_{12}=0.99, P<0.05) and archaealassemblages (m_{12}=0.99, P<0.01). Approximately 1094 (3.0 m), 1129 (5.0 m), and 2886 (5.5 m) CRISPR arrays were predicted from respective microbial metagenomes and the spacers were hit mostly to the viral assemblage from the same depth (Figure 7). In addition, reciprocal BLAST analysis revealed approximately74,777 (3.0 m), 3,761 (5.0 m), and 26,972 (5.5 m) hits between viral and microbial ORFs at each depth.

A total of 9, 22 and 13 unique viral sequences were detected in microbial bin sequences respectively along the depths(Supplementary Table S5). Among detected viral taxa, approximately half including *Cronobacter* phages Vb_CsaM_GAP31, *Pseudomonas* phage YuA and *Prochlorococcus* phage P-SSM7, did not have specificity to hosts.

Microbial groups at 3.0 m seemed to be attacked only by 3.0-m viruses, except *Flavobacteria* (3-2),which wereattacked by Cronobacter phage from 5.0 m.Unlike the 3.0 m, microbial groupsat both 5.0 and 5.5 m seemed to be infected randomly by viruses from both depths.

Flavobacteriumin the 3.0 m viral fraction

A group of *Flavobacteria* in the 3.0-m viral fraction was verified distinctfrom the *Flavobacteria*detected in the 3.0-m microbial fraction, according to two findings: (1) reciprocal BLAST analysis indicated that the*Flavobacteria*from viral fraction (total of 160,318 ORFs) and the one from microbial fraction (a total of 78,840 ORFs) only shared 2043 ORFs (0.01 and 0.03%, respectively); and (2) 16S rRNA genes predicted from the both Flavobacteria metagenomes hadonly 94% shared identity.Over 90% of the genome completeness of the flavobacterium in viral fraction was estimated using bacterial single-copy genes(Supplementary Table S6). Fragment recruitment demonstrated high similarity and coverage toward the genome of *Nonlabensdokdonensis*DSW-6 (Supplementary FigureS11),suggesting that the identified flavobacterium in the viral fraction were closely related to the *Nonlabens* genus. The draft genome sizeof this flavobacteriumwas ~4.2 Mb, comprised of 113 contigs (Figure 8), with complete gene suites for assimilatory sulfate reduction, thiosulfate oxidation,and genes for denitrification (e.g., *nirK* and *nosZ*), nitrite reduction (e.g., *nirBD*), organic carbon oxidation,and carbon fixation.

Discussion

We applied both metagenetic and metagenomic approaches to deeply explore microbial communities (bacteria, archaea and viruses) and potential metabolic capacity in Lake Shunet. Remarkably, a total of~132 Gb of microbial and 133 Gb of viral metagenomics data were

generated which provided much higher resolution for microbial community structure than similar studies conducted in meromictic lakes (Lauro *et al*., 2011; Yau *et al*., 2013). These sequence data facilitated clustering sequences into individual microbial bins, reconstruction of a draft genome of *Flavobacteria* discovered in the viral fraction, and generation ofmany nearly completed viral genomes.

Prokaryotic diversity was high in Lake Shunet

Despite deep sequencing, rarefaction curves of sequence reads against OTU numbers still has not reached a clear asymptote, suggesting that additionalsequencing efforts would detect even greaterdiversity in this lake.Regardless, the observed OTU richness and diversity estimates were generally higher than other aquatic ecosystems, including Yellowstone Lake (Clingenpeel *et al.,* 2011; Kan *et al.,* 2011), freshwater reservoir (Tseng *et al.,* 2013), and marine and mesohaline (Hugoni *et al.,* 2015), consistent withhigh microbial diversity in Lake Shunet.

Variations of prokaryotic communities were significantly explained by physicochemical parameters

Prokaryotic community compositions were discriminated amongthe three depths and their variations were significantly explained by indigenous environmental parameters. The sharp salinity gradient betweenmixolimnion and monimolimnion caused a lack of homogeneity of

hydroparameters through stratification and likely created various niches with distinct abiotic features that shapedspecial prokaryotic assemblages and metabolic capabilities(Lindstrom, 2001;Degermendzhy *et al.,* 2010, Wear *et al.,* 2013).

Many bacterial genera, including *Cyanobacteria,Pseudoalteromonas,Alcanivorax*and*Rhodobacteraceae*:NA were enrichedbyO_2, temperature, pH and NH_4^+ in the mixolimnion, where key genes involved in oxygenicphotosynthesis, nitrogen fixation, and nitrite reduction were detected in bin sequences of exactly those bacteria. In contrast, at 5.5 m, H_2S apparentlyenriched not only*Desulfonatronum*,*Desulfobacteraceae*:NA, *Desulfobacterium*, and *Halothiobacillus*which are either sulfate reducing or sulfur oxidizing bacteria (Castro *et al.,* 2000, Shi *et al.,* 2011)but also Candidatus*Cloacamonas*which is probably a syntrophic bacterium involved in anaerobic digestion (Pelletier *et al.,* 2008), producing H_2S during breakdown of organic matter.

In archaeal community composition, *Nitrososphaera* and *Methanospirillum* were significantly explained by NO_3^-. The former is a typical ammonia-oxidizing archaea which mediatesnitrification(Spang *et al.,* 2012), whereas the latter is a common syntrophic benzoate-degrading partner in benzoate degradation(Dolfing and Tiedje, 1988)in which the genes involved were also detected at 3.0 m.Perhaps degradation of benzoate occurred and the availability of nitrate used as an electron acceptor facilitatedmineralization of benzoate (Berry *et al.,* 1987). This syntrophic relationshipmayhave enriched both archaea at 3.0 m.

Energy metabolism was significantly explained by prokaryotic assemblages

The KO profiles of energy metabolism as well as prokaryotic community compositions significantly differed among the three sampling depths, asmicrobial functional profiles are usually associated with the community (Gilbert *et al.,* 2010) or specific microbial groups (Debroas *et al.,* 2009). For example, in the NMDS ordination, relative abundance of *Thiocapsa* (a bacterial genus)and *Methanococcoides* (an archaeal genus)significantly explained the relative importance of KO involved in metabolism, specifically the former enriched K00395 (adenylylsulfate reductase, subunit B)involved in sulfur metabolism; the latter enriched genes including K00266 (glutamate synthase (NADPH/NADH) small chain)were involved in nitrogen metabolism. *Thiocapsa* dominated the bacterial community structure in Lake Shunet; they usedhydrogen sulfide, thiosulfate, and elemental sulfur as electron donors in the anoxic chemocline zone (Caumette *et al.,* 2004; Rogozin *et al.,* 2012; Baatar *et al.,* 2016). Moreover, reconstruction of nutrient cycling through Lake Shunet also indicated animportant role of *Thiocapsa* in sulfate reduction, consistent with a previous observation (Caumette *et al.,* 2004). Allen *et al*. (2009)discovered that *M. burtonii*isolated from ACE Lake were able to assimilate ammonia usinga two-step glutamine synthetase (Mbur_1975) and glutamate synthase (Mbur_0092) pathway.

Unclassified prokaryotes have an important role in nutrient cycling in Lake Shunet

There were abundant untapped prokaryotes detected across depths, specifically at 5.5 m, indicating that the lake harbored highly diverse prokaryotes. According to the reconstructed nutrient cycling of the lake based onbinning results, several unknown or uncultured bacteria in highly relative abundance (based on coverage) contributed to energy generation at both 3.0 and 5.5 m. Apparently, underexplored microbial groups were not only predominant but also hadimportant roles in energy metabolism in this lake ecosystem.

To speculate aboutpotential roles of unclassified bacterial and archaeal OTUs(Supplementary Table S7), correspondence analysis was used to visualize relationships between the 10 most abundant bacterial and archaeal OTUsand genes involved in energy metabolism, respectively.For example, arcOTU_282and bacOTU_26 were enriched in genes of K02123 (V-type H^+-transporting ATPase subunit I) andK02117/K02118 (V-type H^+-transporting ATPase subunit A/B) involved in oxidative phosphorylation, respectively (Supplementary FigureS12).Interestingly, K00266 (glutamate synthase small chain) and K00262 (glutamate dehydrogenase)were potentially enriched by the specific unclassified archaea (arcOTU_31). Thisunknown archaea might also assimilate ammonia usingpathwayssimilarto those in*M. burtonii*(Allen *et al.,* 2009).

Reconstruction of nutrient cycles in Lake Shunet

Nutrient cycles apparently differed among the three sampling depths. For example, aerobic carbon fixation was only annotated in oxicmixolimnion, and no genes involved in sulfite oxidation were annotated in chemocline, whereasnitrite oxidation was only detectedin the chemocline layer. These differences were attributed to variations inmicrobial community structure inthe three layers, as individual microbial groups evolved to addressdisparate core redox reactions (Falkowski *et al.,* 2008).

Based on 16S rRNA genes and read counts of ORFs, *Chroococcales*was the most abundant bacterial group in oxicmixolimnionandthe only biotic source of O_2 production. Similar to ACE Lake, both mixolimnion and monimolimnion of Lake Shunetwere involved in production of carbon monoxide (CO), due to incomplete oxidation of organic compounds, indicating that CO oxidation might be an important pathway for *Rhodobacteraceae* and unknown aerobic/anaerobic groups to generate energy, as numerous microbes are able to utilize CO for growth (Mörsdorf *et al.,* 1992; Oelgeschlager and Rother, 2008).Both upper and lower layers harbored highly diverse and abundant prokaryotes involved in the microbial loop of this specific lake, competing for organic carbon. However, some bacteria obtained energy from CO.

Lake Shunethad methane concentrations of up to 32 μM (Kallistova *et al.,* 2006); however,no genes involved in methane oxidation ormethanogenesis were annotated in the current study, in contrast to previous research suggesting methane production commonly occurs in lake systems(Bastviken *et al.,* 2004; Pasche *et al.,* 2011). Our results also contrasted with those

reported forACE Lake, where methane concentrations were up to 22.4 μM and genes for methane production weredetected (Lauro *et al.,* 2011). We also observed prevalentmethanogenic archaea in Lake Shunet; therefore, we speculated that either methane was not an important carbon source in anoxic zone, or conditionsat sampling were not favorable for methane production. Moreover, no genes involved in methane oxidation were annotated, perhaps due to zero consumption from methanotrophic bacteria, which were not identified in the lake.

Many bins in high relative abundance potentially supportednitrate reduction and nitrogen fixation, leading toproduction ofammonium, consistent with high ammonium concentrations detected in Lake Shunet, specifically in mixolimnion. Perhaps the N_2 source for nitrogen fixation was from the oxic zone.

Genes involved in the complete process of denitrification were annotated through the depths, including the oxic layer. In a previous study,abundance ofdenitrifying bacteria were at equal levels(3.2-5.3× 10^5 copy ml^{-1}) from surface to bottom water in Arabian Sea, but denitrification activity was only detectable by stable isotope tracer method in the middle water, which was the oxygen minimum zone(Ward *et al.,* 2009). Therefore, we inferred thatthe presence of denitrifying bacteria did not always activate genes involved in denitrification.

In sulfur cycling, *Thiocapsa*(bin 5-6 and 5.5-18) was the dominant purple sulfur bacteria in the lake that also utilized thiosulfate ($S_2O_3^{2-}$) instead of H_2S as an electron donor and CO_2 as a carbon source under anoxic conditions, as reported (Caumette *et al.,* 2004).Annotation of genes revealed

that not only *Thiocapsa*(bin 5-6 and 5.5-18) but also other bacteria throughout the lake,including *Hyphomonas neptunium*-like bacteria (bin 3-11), *Enterobacteriaceae* (bin 5-2), and *Desulfobacteraceae* (bin 5.5-12) harbored genes involved in reducing sulfate to H_2S as a terminal product (Figure 4). According to the relative abundance of bins, H_2S seemed to be produced mainly in chemocline. TheH_2S concentration was nearly zero at both mixolimnion and chemocline, but increased sharply to 400 mg l^{-1}at monimolimnion.PerhapssomeH_2Sleached from the sediment and generated by *Thiocapsa*and *Desulfobacteraceae*;in addition,H_2Saccumulated at the bottom of the lake from upper layers, due to gravity and increasing salinity.

Three potential chemolithoautotrophic groups, including *Hyphomonas neptunium*-like bacteria (bin 3-11), *Enterobacteriaceae*(bin 5-5) and *Halomonas* (bin 5.5-17) were identified and potentially capable of derivingcarbon from carbon fixation and energy from nitrate and sulfate reduction for growth. Specifically, bin 5-5 was the only bin that harbored genes for the second step of nitrification, not detected in ACE Lake metagenomes. The three potential chemolithoautotrophs all belonged to non-phototrophic proteobacteria, consistent with discovery ofchemolithoautotrophyamong ubiquitous bacterial lineages(Swan *et al*. 2011).

Nonlabens*sp.–** ***a tiny Flavobacteria discovered in the viral fraction

When analyzing viral metagenomic data, numerous flavobacterialcontigs were unexpectedly detected at 3.0 m.*Flavobacteria*are ubiquitous strictly aerobic bacteria,comprised of diverse

members (Alonso *et al.,* 2007).According to acomparison of 16S rRNA gene sequences,*Flavobacteria*are phylogenetically close to *Nonlabensdokdonensis* (97% identity), of which cells were previously identified as *Donghaeanadokdonensis* and followed by *Psychroflexustorquis* ATCC 700755 (92% identity; Yi and Chun, 2012; Yoon *et al.,* 2006).

Why was *Flavobacteria* detected in the viral fraction? The cell morphology is generally rod-shaped or filamentous with a size range of 1.2-50 μm ×0.2-1.5 μm (Thomas-Jinu and Goodwin 2004; Yoon *et al.,* 2006; Yoshizawa *et al.,* 2012); *Psychroflexus* are those that are0.2μm in width (Bowman *et al.,* 1998). However, *Flavobacteria* detected in the viral fraction (< 0.22 μm) in this study was *Nonlabens*sp., which has apparently not been described.We speculated that this *Nonlabens*sp. likely was 0.2μm wide and thereforecould have passed through the 0.22 μm filter. *N. dokdonensis* are strictly aerobic, corresponding to the exclusive detection in the oxic 3.0 m. In addition, some phenotypic characteristics of *N. dokdonensis,* including the requirement of NaCl for growth, the capability of producing H_2S and performing nitrate reduction (Yoon *et al.,* 2006),were consistent with environmental conditions of surface water of Lake Shunet.

The draft genome of the tiny *Flavobacteria*had more than 70% of genes in common with the reference genome of *Nonlabensdokdonensis* DSW-6 and harbored ~3% specific genes, suggesting the bacterium *Flavobacteria*was a yet to be identified *Nonlabens* sp. with small size (< 0.2 μmwide) and also apparently the first one detected in an inland lake ecosystem.

Viral communities were highly diverse and discriminated along depths

In Lake Shunet, the viral assemblage included 14 major viral families (≥0.1% relative abundance), exhibiting a greater diversity (*H'*: 6.70~9.15) and genotype richness (from 2498 to 15877) than other aquatic ecosystems, most of which had only three to 12 viral families (Fancello *et al.,* 2012; Lopez-Bueno *et al.,* 2009; Tseng *et al.,* 2013).

Viral communities including *Siphoviridae, Microviridae*, *Myoviridae*, *Podoviridae*, *Plasmaviridae*, and *Inoviridae*were typical bacteriophages. Viral communities had different distribution patterns along depth. For example, *Siphoviridae*were the most abundant at 3.0 m, whereas*Microviridae* and *Nanoviridae*were most abundant at 5.0 and 5.5 m, respectively. *Siphoviridae*, *Myoviridae*, and *Podoviridae*were distributed ubiquitously, whereas*Microviridae*and *Inoviridae*were detected in a wide range of environments, including reclaimed water (Rosario *et al.,* 2009), the Sargasso Sea (Angly *et al.,* 2006) and marine sediment (Breitbart *et al.,* 2004).

Prior to the current study,*Plasmaviridae*have apparently only have been reported in Antarctic lakes (Rosario and Breitbart, 2011). Perhaps certain viral typesare enriched by local environmental condition through selective pressure (Angly *et al.,* 2006; Winter *et al.,* 2013). With the exception of bacteriophages, terrestrial eukaryotic viruses were detected in Lake Shunet, specifically abundant *Nanoviridae* at 5.5 m. The latterare DNA viral pathogens of plants (Lukert*et al.,* 1995). Furthermore, blast results indicated that those plants were mostly faba bean necrotic

yellows, subterranean clover stunt and milk vetch dwarf. As Lake Shunet has an inflow from a stream (Degermendzhy *et al.,* 2010), this would be a potential source of terrestrial viruses, flowing into lake during thesnow melt and subsequentlyaccumulatingat the lake bottom, due to its small size and increasing water salinity.

Viral and prokaryotic assemblages intimately interacted

Viral richness and diversity of Lake Shunet significantly differedamong depths and were associated with bacterial and archaeal richness and diversity. Therefore, weconcluded that microbial diversity was largely influenced by local viral assemblages, as viruses have been deemed important vehicles, shuttling genetic materials and subsequently influencing speciation of microorganisms in aquatic ecosystems (Weinbauer and Rassoulzadegan, 2003). In addition, our results indicated that some viral taxa had specificity to certainmicrobial bins, consistent with the previous study that viruses are thought to be strain-specific predators. Consequently, is likely that high prokaryotic diversity results in high viral diversity (Rohwer *et al.,* 2009). Based on Procrustes correlation analysis, there was significant congruence between viral and prokaryotic communities in Lake Shunet, consistent withinteractionsbetween prokaryotes and viral populations.

Unlike 3.0 m, microbial bins at both 5.0 and 5.5 m seemed to be randomly infected by viruses from both depths. We speculated that this might be either due to the similar environmental

690 conditions of both layers which harbored similarviral assemblages, or perhapsviruses of the two bottom layers were more generalists that infected diverse taxonomic microbial groups.

Clustered regularly interspaced short palindromic repeats (CRISPR) is an antiviral defense system common in microbial genomes (Marraffini and Sontheimer, 2010). Importantly,"infection traces", namely that viral sequences incorporated in the CRISPR arrays were detected in the microbial metagenomes of Lake Shunet, suggesting previous interactions between viral and microbial communities (Berg Miller *et al.,* 2012). For example, CRISPRs detected at both 5.0 and 5.5 m were randomly matched to viral sequences of both depths, whereas viralsequences at 3.0 m seemed to specifically match CRISPRs at 3.0 m only. Infection patterns were in accordance with patterns characterized in interactions between bins and viruses, supporting the assertion that both

700 chemocline and monimolimnion might harbor similar viral assemblages, whereas potentially infected diverse taxonomic groups and viruses at3.0 m weremore strain-specific (Atanasova *et al.,* 2012; Rohwer *et al.,* 2009).

Lake Shunet is a unique meromictic Lake

Lake Shunet has several lake characteristics and patterns of physicochemical profile that resemble ACE Lakein Antarctica, which has been wellinvestigated (Laybourn-Parry and Bell Supplementary 2014). To determine specificity of Lake Shunet in terms of microbial communities and functions, NMDSanalysis based on embedded COG profiles obtained from

microbial metagenomes of both lakeswas initially performed (Supplementary Figure 13). In the NMDS ordination plot, Lake Shunet and ACE Lake were obviously discriminated, indicating that there were distinct biogeochemical processes characterizing local microbial communities.Furthermore, initial NMDS results suggested that Lake Shunetwas a unique meromictic lake in terms of microbial community and potential functions, making it worthwhile to investigate. Since only a few studies to characterize microbial diversity in meromictic lakes have beenconducted withpyrosequecing, our efforts contributed to baselines of microbial community composition and functional processes in the meromictic lakeecosystems, specifically the unique Lake Shunet.

Acknowledgements

We thank our Russian colleagues in the Institute of Biophysics SB RAS for assistance with providing samples and hydro-parameterdata.This research was supported by the Ministry of Science and Technology in Taiwan through Taiwan-Russian Joint ProjectGrant NSC 99-2923-B-001-001-MY3and NSC 102-2923-B-001-004 and the Russian Foundation forBasic Research, Grant No. 16-05-00091-a 14-04-01060-a. The research was partially supported by the Council on grants from the President of the Russian Federation for support of leading scientific schools (grant NSh-9249.2016.5)

References

Allen MA, Lauro FM, Williams TJ, Burg D, Siddiqui KS, De Francisci D *et al.* (2009). The genome sequence of the psychrophilic archaeon, *Methanococcoides burtonii*: the role of genome evolution in cold adaptation. ISME J 3:1012-1035.

Alonso C, Warnecke F, Amann R, Pernthaler J. (2007). High local and global diversity of Flavobacteria in marine plankton. Environ Microbiol 9:1253-1266.

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.

Atanasova NS, Roine E, Oren A, Bamford DH, Oksanen HM. (2012). Global network of specific virus-host interactions in hypersaline environments. Environ Microbiol 14: 426-440.

Baatar B, Chiang PW, Rogozin DY, Wu YT, Tseng CH, Yang CY*et al* (2016). Bacterial Communities of Three Saline Meromictic Lakes in Central Asia. PLoS One **11:** e0150847.

Bastviken D, Cole J, Pace M, Tranvik L. (2004). Methane emissions from lakes: Dependence of lake characteristics, two regional assessments, and a global estimate. Global Biogeochem Cy 18:GB4009.

Bauer DF.(1972). Constructing confidence sets using rank statistics. J Amer Statist Assoc67:687–690.

Berg Miller ME, Yeoman CJ, Chia N, Tringe SG, Angly FE, Edwards RA *et al.* (2012). Phage-bacteria relationships and CRISPR elements revealed by a metagenomic survey of the rumen microbiome. Environ Microbiol 14:207-227.

Berry DF, Francis AJ, Boliag JM. (1987). Microbial metabolism of homocyclic and heterocyclic aromatic compounds under anaerobic conditions. Microbiol Rev 51:43-59.

Bolker B, Bonebakker L, Gentleman R, Huber W, Liaw A, *et al*. (2012) gplots: Various R programming tools for plotting data. R package version 2.11.0. Available: http://CRAN.R-project.org/package=gplots.

Bowman JP, McCammon SA, Lewis T, Skerratt JH, Brown JL, Nichols DS *et al.* (1998). *Psychroflexus torquis* gen. nov., sp. nov. a psychrophilic species from Antarctic sea ice, and reclassification of *Flavobacterium gondwanense* (Dobson *et al.*. 1993) as *Psychroflexus*

gondwanense gen. nov., comb. nov.. Microbiology 144:1601-1609.

Bowman JP, Rea SM, McCammon SA, McMeekin TA. (2000). Diversity and community structure within anoxic sediment from marine salinity meromictic lakes and a coastal meromictic marine basin, Vestfold Hills, Eastern Antarctica. Environ Microbiol 2:227-237.

Breitbart M, Felts B, Kelley S, Mahaffy JM, Nulton J, Salamon P*et al.* (2004). Diversity and population structure of a near–shore marine–sediment viral community. Proc R SocLond B BiolSci271:565-574.

Breitbart M. (2012). Marine viruses: truth or dare. Annu Rev Mar Sci 4:425-448.

Castro HF, Williams NH, Ogram A.(2000). Phylogeny of sulfate-reducing bacteria. FEMS MicrobEcol 31:1-9.

Caumette P, Guyoneaud R, Imhoff JF, Suling J, Gorlenko V. (2004). *Thiocapsa marina* sp. nov., a novel, okenone-containing, purple sulfur bacterium isolated from brackish coastal and marine environments. Int J Syst Evol Microbiol 54:1031-1036.

Clingenpeel S, Macur RE, Kan J, Inskeep WP, Lovalvo D, Varley J *et al.* (2011). Yellowstone Lake: high-energy geochemistry and rich bacterial diversity. Environ Microbiol 13:2172-2185.

Comeau AM, Harding T, Galand PE, Vincent WF, Lovejoy C. (2012). Vertical distribution of microbial communities in a perennially stratified Arctic lake with saline, anoxic bottom waters. Sci Rep 2:604.

Debroas D, Humbert JF, Enault F, Bronner G, Faubladier M, Cornillot E. (2009). Metagenomic approach studying the taxonomic and functional diversity of the bacterial community in a mesotrophic lake (Lac du Bourget--France). Environ Microbiol 11:2412-2424.

Degermendzhy AG, Zadereev ES, Rogozin DY, Prokopkin IG, Barkhatov YV, Tolomeev AP *et al.* (2010). Vertical stratification of physical, chemical and biological components in two saline lakes Shira and Shunet (South Siberia, Russia). Aquatic Ecology 44:619-632.

DeLong EF, Preston CM, Mincer T, Rich V, Hallam SJ, Frigaard N-U *et al.* (2006). Community Genomics Among Stratified Microbial Assemblages in the Ocean's Interior. Science 311:496-503.

Dolfing J, Tiedje JM. (1988). Acetate Inhibition of Methanogenic, Syntrophic Benzoate

Degradation. Appl Environ Microbiol 54:1871-1873.

Edwards RA, Rohwer F. (2005). Viral metagenomics. Nat Rev Micro 3:504-510.

Falkowski PG, Fenchel T, Delong EF. (2008). The microbial engines that drive Earth's biogeochemical cycles. Science 320:1034-1039.

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.

Finn RD, Mistry J, Tate J, Coggill P, Heger A, Pollington JE *et al.* (2010). The Pfam protein families database. Nucleic Acids Res 38:D211-222.

Gilbert JA, Field D, Swift P, Thomas S, Cummings D, Temperton B *et al.* (2010). The taxonomic and functional diversity of microbes at a temperate coastal site: a 'multi-omic' study of seasonal and diel temporal variation. PLoS ONE 5:e15545.

Haft DH, Selengut JD, White O. (2003). The TIGRFAMs database of protein families. Nucleic Acids Res 31:371-373.

Huang Y, Gilna P, Li W. (2009). Identification of ribosomal RNA genes in metagenomic fragments. Bioinformatics 25:1338-1340.

Hugoni M, Agogue H, Taib N, Domaizon I, Mone A, Galand PE *et al.* (2015). Temporal dynamics of active prokaryotic nitrifiers and archaeal communities from river to sea. Microb Ecol 70:473-483.

Kalacheva GS, Gubanov VG, Gribovskaya IV, Gladchenko IA,Zinenko GK, Savitsky SV. (2002). Chemical analysis ofLake Shira water (1997–2000). Aquat Ecol 36:123–141.

Kallistova AY, Kevbrina MV, Pimenov NV, Rusanov II, Rogozin DY, Wehrli B *et al.* (2006). Sulfate reduction and methanogenesis in the Shira and Shunet meromictic lakes (Khakasia, Russia). Microbiology 75:720-726.

Kan J, Clingenpeel S, Macur RE, Inskeep WP, Lovalvo D, Varley J *et al.* (2011). Archaea in Yellowstone Lake. ISME J 5:1784-1795.

Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res 40:D109-114.

Kristiansson E, Hugenholtz P, Dalevi D. (2009). ShotgunFunctionalizeR: an R-package for functional comparison of metagenomes. Bioinformatics 25:2737-2738.

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D *et al.* (2009). Circos: an information aesthetic for comparative genomics. Genome Res 19:1639-1645.

Lauro FM, DeMaere MZ, Yau S, Brown MV, Ng C, Wilkins D *et al.* (2011). An integrative study of a meromictic lake ecosystem in Antarctica. ISME J 5:879-895.

Laybourn-Parry J, Bell E. (2014). Ace Lake: three decades of research on a meromictic, Antarctic lake. Polar Biol 37:1685-1699.

Li W, Godzik A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22:1658-1659.

Lindstrom ES. (2001). Investigating Influential Factors on Bacterioplankton Community Composition: Results from a Field Study of Five Mesotrophic Lakes. Microb Ecol 42:598-605.

Lopez-Bueno A, Tamames J, Velazquez D, Moya A, Quesada A, Alcami A. (2009). High diversity of the viral community from an Antarctic lake. Science 326:858-861.

Lunina ON, Bryantseva IA, Akimov VN, Rusanov II, Rogozin DY, Barinova ES *et al.* (2007). Seasonal changes in the structure of the anoxygenic photosynthetic bacterial community in Lake Shunet, Khakassia. Microbiology 76:368-379.

Lukert PD, de Boer GF, Dale JL, Keese P, McNulty MS, Randles JW *et al*. (1995). The Circoviridae. In: Murphy FA, Fauquet CM, Bishop DHL, Ghabrial SA, Jarvis AW, Martelli GP *et al.* (ed). Virus Taxonomy, Classification and Nomenclature of Viruses. Sixth Report of the International Committee on Taxonomy of Viruses. Springer Press: Wien, New York, pp 166–168.

Marraffini LA, Sontheimer EJ. (2010). CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. Nat Rev Genet 11:181-190.

Mörsdorf G, Frunzke K, Gadkari D, Meyer O. (1992). Microbial growth on carbon monoxide. Biodegradation 3:61-82.

Ng C, DeMaere MZ, Williams TJ, Lauro FM, Raftery M, Gibson JA *et al.* (2010). Metaproteogenomic analysis of a dominant green sulfur bacterium from Ace Lake, Antarctica. ISME J 4:1002-1019.

Oelgeschlager E, Rother M. (2008). Carbon monoxide-dependent energy metabolism in anaerobic bacteria and archaea. Arch Microbiol 190:257-269.

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB *et al.* (2011). vegan: Community Ecology Package. R package version 20-1http://CRAN.R-project.org/package=vegan.

Parks DH, Beiko RG. (2010). Identifying biologically relevant differences between metagenomic communities. Bioinformatics 26:715-721.

Parnachev VP, Degermendzhy AG. (2002). Geographical, geological and hydrochemical distribution of saline lakes in Khakasia, Southern Siberia. Aquat Ecol 36:107-122.

Pasche N, Schmid M, Vazquez F, Schubert CJ, Wüest A, Kessler JD *et al.* (2011). Methane sources and sinks in Lake Kivu. J Geophys Res 116:G03066.

Pedrós-Alió C. (2012). The rare bacterial biosphere. Annu rev mar sci 4:449-466.

Pelletier E, Kreimeyer A, Bocs S, Rouy Z, Gyapay G, Chouari R *et al.* (2008). "Candidatus *Cloacamonas acidaminovorans*":genome sequence reconstruction provides a first glimpse of a new bacterial division. J Bacteriol 190:2572-2579.

Perrière G, Thioulouse J (2003). Use of correspondence discriminant analysis to predict the subcellular location of bacterial proteins. Comput Methods Programs Biomed70: 99-105.

Rankin LM, Gibson JAE, Franzmann PD, Burton HR. (1999). The chemical stratification and microbial communities of ACE Lake, Antarctica: a review of the characteristics of a marine-derived meromictic lake. Polarforschung 66:33-52.

Revelle W. (2011). psych: Procedures for Psychological, Psychometric, andPersonality Research. R package version 1.01.9.http://personality-project.org/r/psych.manual.pdf.

Rogozin DY, Zykov VV, Chernetsky MY, Degermendzhy AG, Gulati RD. (2009). Effect of winter

conditions on distributions of anoxic phototrophic bacteria in two meromictic lakes in Siberia, Russia. Aquat Ecol 43:661-672.

Rogozin DY, Trusova MY, Khromechek EB, Degermendzhy AG. (2010). Microbial community of the chemocline of the meromictic Lake Shunet (Khakassia, Russia) during summer stratification. Microbiology 79:253-261.

Rogozin DY, Zykov VV, Degermendzhi AG. (2012). Ecology of purple sulfur bacteria in the highly stratified meromictic Lake Shunet (Siberia, Khakassia) in 2002–2009. Microbiology 81:727-735.

Rohwer F, Prangishvili D, Lindell D. (2009). Roles of viruses in the environment. Environ Microbiol 11:2771-2774.

Rosario K, Nilsson C, Lim YW, Ruan Y, Breitbart M. (2009). Metagenomic analysis of viruses in reclaimed water. Environ Microbiol 11:2806-2820.

Rosario K, Breitbart M. (2011). Exploring the viral world through metagenomics. Curr opin virology 1:289-297.

Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooseph S *et al.* (2007). The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. PLoS Biol 5:e77.

Saeed I, Tang SL, Halgamuge SK. (2012). Unsupervised discovery of microbial population structure within metagenomes using nucleotide base composition. Nucleic Acids Res 40:e34.

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 comparingmicrobial communities. Appl Environ Microbiol 75:7537–7541.

Shi X, Lin H, Yuan X, Zhao Y. (2011). Isolation and characterization of a novel sulfuroxidizing chemolithoautotroph Halothiobacillus from Pb polluted paddy soil. Afr J Biotechnol 10:4121-4126.

Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, Neal PR *et al.* (2006). Microbial diversity in the deep sea and the underexplored "rare biosphere". Proc Natl Acad Sci USA 103:12115-12120.

Spang A, Poehlein A, Offre P, Zumbragel S, Haider S, Rychlik N *et al.* (2012). The genome of the ammonia-oxidizing Candidatus Nitrososphaera gargensis: insights into metabolic versatility and environmental adaptations. Environ Microbiol 14:3122-3145.

Suttle CA. (2007). Marine viruses--major players in the global ecosystem. Nat Rev Microbiol 5:801-812.

Swan BK, Martinez-Garcia M, Preston CM, Sczyrba A, Woyke T, Lamy D *et al.* (2011). Potential for chemolithoautotrophy among ubiquitous bacteria lineages in the dark ocean. Science 333:1296-1300.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. MolBiolEvol 30:2725-2729.

Team RDC. (2011) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. http:// www.R-project.org/.

Tetu SG, Breakwell K, Elbourne LD, Holmes AJ, Gillings MR, Paulsen IT. (2013). Life in the dark: metagenomic evidence that a microbial slime community is driven by inorganic nitrogen metabolism. ISME J 7:1227-1236.

Thomas-Jinu S, Goodwin AE. (2004). Morphological and genetic characteristics of Flavobacterium columnare isolates: correlations with virulence in fish. JFish Dis 27:29-35.

Tseng CH, Chiang PW, Shiah FK, Chen YL, Liou JR, Hsu TC *et al.* (2013). Microbial and viral metagenomes of a subtropical freshwater reservoir subject to climatic disturbances. ISME J 7:2374-2386.

Volkov II, Zhabina NN.(1990). Method of determination of reduced sulfur species in sea water. Okeanologiya 30:778-782.

Ward BB, Devol AH, Rich JJ, Chang BX, Bulow SE, Naik H *et al.* (2009). Denitrification as the dominant nitrogen loss process in the Arabian Sea. Nature 461:78-81.

Wang Q, Garrity GM, Tiedje JM, Cole JR. (2007). NaïveBayesian classifier for rapid assignment of rRNAsequences into the new bacterial taxonomy. ApplEnviron Microbiol 73:5261–5267.

Wear EK, Koepfler ET, Smith EM. (2013). Spatiotemporal Variability in Dissolved Organic Matter Composition is More Strongly Related to Bacterioplankton Community Composition than to Metabolic Capability in a Blackwater Estuarine System. Estuaries and Coasts 37:119-133.

Weinbauer MG, Rassoulzadegan F. (2003). Are viruses driving microbial diversification and diversity? Environ Microbiol 6:1-11.

Wilson K. (2001). Preparation of genomic DNA frombacteria. Curr Protoc Mol Biol Chapter 2: Unit 2.4;doi:10.1002/0471142727.mb0204s56.

Winter C, Matthews B, Suttle CA. (2013). Effects of environmental variation and spatial distance on Bacteria, Archaea and viruses in sub-polar and arctic waters. ISME J 7:1507-1518.

Yau S, Lauro FM, DeMaere MZ, Brown MV, Thomas T, Raftery MJ *et al.* (2011). Virophage control of antarctic algal host-virus dynamics. Proc Natl Acad Sci USA 108:6163-6168.

Yau S, Lauro FM, Williams TJ, Demaere MZ, Brown MV, Rich J *et al.* (2013). Metagenomic insights into strategies of carbon conservation and unusual sulfur biogeochemistry in a hypersaline Antarctic lake. ISME J 7:1944-1961.

Yi H, Chun J. (2012). Unification of the genera *Nonlabens*, *Persicivirga*, *Sandarakinotalea* and *Stenothermobacter* into a single emended genus, *Nonlabens*, and description of *Nonlabens agnitus* sp. nov. Syst Appl Microbiol 35:150-155.

Yoon JH, Kang SJ, Lee CH, Oh TK. (2006). *Donghaeana dokdonensis* gen. nov., sp. nov., isolated from sea water. Int J Syst Evol Microbiol 56:187-191.

Yoshizawa S, Kawanabe A, Ito H, Kandori H, Kogure K. (2012). Diversity and functional analysis of proteorhodopsin in marine Flavobacteria. Environ Microbiol 14:1240-1248.

Figure legends

Figure 1. Community structure of (a) archaea and (b) bacteria across three depths of Lake Shunet

at phylum:class and genus levels. Underlined legends had more than one class or genus in the taxonomic unit.

Figure 2. Heatmap and functional clustering of predicted ORFs from microbial metagenomic reads based on KEGG Orthology (KO) of energy metabolism at three depths. The heat scale is the percentage of ORFs assigned to each individual KO of energy metabolic category.
[a-c]Depths without a common letter differed ($P< 0.05$).

Figure 3. Non-metric multidimensional scaling (NMDS) ordination of three sampling depths based on (a and b) the 50 most abundant KOinvolved in metabolic pathways, (c) bacterial and (d) archaeal community structure at genus level.In each diagram, archaeal, bacterial groups and hydroparameterswith a significant goodness of fit based on post-hoc correlations ($P\leq0.05$) are represented as vectors. In (b) G1* includes Planctomycetaceae:NA and Marinobacter; G2* includes Spartobacteria:NA,Roseivirga, Loktanella, Rhodobacteraceae:NA, Oceanospirillales:NA,Psychroserpens, Luteolibacter, Burkholderiales:NA, Pseudoalteromonas, Persicivirga, Alcanivorax and Flavobacteriales:NA; G3* includes Pantoea, Desulfotignum,Bacillus and Clostridiales:NA. Cyano: Cyanobacteria, Alcali: Alcaligenaceae, Pseudo: Pseudomonas, Desulfuro: Desulfuromonadaceae, Exiguo: Exiguobacterium, Burkholde:Burkholderiales, Flavo: Flavobacteriales, Saccharo: Saccharospirillum, Rosei: Roseivirga, Luteoli: Luteolibacter, Psychro: Psychroserpens, Rhodo: Rhodospirillaceae, Pseudoalte: Pseudoalteromonas, Hypho: Hyphomonas, Rhodospi: Rhodospirillaceae:NA, Sparto: Spartobacteria, Alcani: Alcanivorax, Persici: Persicivirga, Oceanospi: Oceanospirillales. The 50 genes include**K01633**dihydroneopterin aldolase, **K10211**4,4'-diaponeurosporenoate glycosyltransferase,**K03186** 3-octaprenyl-4-hydroxybenzoate carboxy-lyaseUbiX, **K01915**glutamine synthetase,**K00948** ribose-phosphate pyrophosphokinase,**K02689**photosystem I P700 chlorophyll a apoprotein A1, **K00265** glutamate synthase (NADPH/NADH) large chain, **K02121** V-type H+-transporting ATPase subunit E, **K00548**methyltetrahydrofolate--homocysteine methyltransferase,**K01674** carbonic anhydrase, **K01955** carbamoyl-phosphate synthase large subunit, **K01652**acetolactate synthase I/II/III large subunit, **K01624** fructose-bisphosphate aldolase, class II, **K02293**15-cis-phytoene desaturase, **K02302**uroporphyrin-III C-methyltransferase / precorrin-2dehydrogenase/sirohydrochlorinferrochelatase, **K04042** bifunctional UDP-N-acetylglucosaminepyrophosphorylase /Glucosamine-1-phosphate N-acetyltransferase, **K02299** cytochrome o ubiquinol oxidase subunit III, **K00266**glutamate synthase (NADPH/NADH) small chain, **K03891**ubiquinol-cytochrome c reductase cytochrome b subunit,**K15666**fengycin family lipopeptidesynthetase C, **K03464**muconolactone D-isomerase, **K00382**dihydrolipoamide dehydrogenase, **K00104**glycolate oxidase, **K01190** beta-galactosidase, **K01649**2-isopropylmalate synthase, **K00075** UDP-N-acetylmuramate dehydrogenase.**K16794**

platelet-activating factor acetylhydrolase IB subunit alpha, **K01803**triosephosphate isomerase (TIM), **K07806** UDP-4-amino-4-deoxy-L-arabinose-oxoglutarate aminotransferase, **K01535** H+-transporting ATPase,**K01953** asparagine synthase (glutamine-hydrolysing), **K00395** adenylylsulfate reductase, subunit B, **K01903** succinyl-CoA synthetase beta subunit, **K02285** phycocyanin beta chain,**K00058** D-3-phosphoglycerate dehydrogenase, **K05577** NAD(P)H-quinone oxidoreductase subunit 5,**K01710** dTDP-glucose 4,6-dehydratase, **K00363** nitrite reductase (NAD(P)H) small subunit, **K03635** molybdopterin synthase catalytic subunit, **K11731** citronellyl-CoA dehydrogenase, **K00485** dimethylaniline monooxygenase (N-oxide forming),**K10203** elongation of very long chain fatty acids protein 6, **K16047** 3-hydroxy-9,10-secoandrosta-1,3,5(10)-triene-9,17-dione monooxygenase subunit HsaA, **K00163** pyruvate dehydrogenase E1 component, **K00366** ferredoxin-nitrite reductase, **K15234** citryl-CoA lyase, **K11311** anthranilate dioxygenase reductase, **K01179** endoglucanase, **K00368** ; nitrite reductase (NO-forming), and**K15655**surfactin family lipopeptidesynthetase B.

Figure 4. Reconstruction of C, N and S cycles in Lake Shunet. The energy flux is presented in each depth based on the KEGG annotation of bins and metagenomes (inlet square at right-bottom corner). Numbers denote the bins. Please refer to Table 1 for information regardingeach bin. Arrows in bold indicate the presence of the pathways in the corresponding metagenomes, but not presentin bins. In the C cycle, black arrow: respiration, dark gray arrow: aerobic carbon fixation, and light gray arrow: anaerobic carbon fixation. PAPS: 3'-Phosphoadenylyl sulfate, PAS: Adenylyl sulfate, C_{org}: Organic matter. Numbers in bold with underlines represent unknown bacteria. 3-1:*Bacteroidetes*, 3-2: *Flavobacteria*, 3-3: *Pseudoalteromonas*, 3-4: unknown bacterium, 3-6: unknown bacterium, 3-7:*Pseudomonas stutzeri*-like bacteria, 3-8: *Gammaproteobacteria*, 3-9:*Halomonas*, 3-10: *Rhodobacteraceae*, 3-11: *Hyphomonas neptunium*-like bacteria, 3-12: *Chroococcales*, 3-13: *Alcanivorax*, 3-14: *Verrucomicrobia*/*Chthoniobacterflavus*, 5-1: *Firmicutes*/*Clostridia*, 5-2: *Staphylococcus*, 5-3: *Bacteroidetes*, 5-4: *Bacteroidetes*, 5-5: *Enterobacteriaceae*, 5-6: *Thiocapsa*, 5.5-2: Uncultured bacterium, 5.5-3: Uncultured bacterium, 5.5-4:*Halanaerobium*, 5.5-5:Uncultured bacterium, 5.5-6: Uncultured bacterium, 5.5-9: *Deltaproteobacteria*, 5.5-10: Uncultured candidate division OP1 bacterium, 5.5-12: *Desulfobacteraceae*, 5.5-13:*Bacteroidetes*/*Marinilabiaceae*, 5.5-15: *Desulfobacteraceae*, 5.5-16:*Clostridiaceae*, 5.5-17: *Halomonas*, 5.5-18: *Thiocapsa*

Figure 5. Measurement of diameter of viral particles. (a) TEM photos ofviral fractions at each sampling depth (left to right: 3.0, 5.0, and 5.5m, respectively).(b) Boxplot of measured diameters.[a-c]Diameters without a common superscript differed(P<0.05).

Figure6.Viral community structures at family level in Lake Shunet. Virus:NA represents unknown virus.

Figure7.Predicted CRISPR of microbial metagenomes were blasted against viral metagenomesin the corresponding depth. (a) At 3.0m: 1094 CRISPR (327hits matched); (b) At 5.0m: 1129 CRISPR (220hits matched); and (c) At 5.5m: 2886 CRISPR (591hits matched)

Figure 8. Draft genome organization of *Nonlabens* sp. and gene conservation among phylogenetically related Flavobacterial reference genomes. The outermost concentric circle representedthe assembled genome of *Nonlabens* sp. Green concentric circle depicted*Nonlabensdokdonensis* DSW-6, followed by red and blue ones that denotetFlavobacteria bacterium BBFL7 and *Psychroflexustorquis* ATCC 700755, respectively. The 5S and 23S rRNAwere located between 2.1 and 2.2 Kb.

1120

Table 1.Physicochemical parameters of three sampling depths in Lake Shunet.
Temp: temperature, N_{tot}: total nitrogen, C_{tot}: total carbon, $Tot_{minerals:}$ Tot minerals

Parameters	3.0 m	5.0 m	5.5 m
Salinity (g l^{-1})	26	40	71
pH	8.1	7.6	6.7
Temp (^{o}C)	15.5	9.5	7.5
O_2 (mg l^{-1})	10.9	0	0
N_{tot} (mg l^{-1})	2.9	4.6	5
C_{tot} (mg l^{-1})	100.7	160	245
Na (g l^{-1})	4.8	8.3	11.2
P (mg l^{-1})	0.2	0.5	1.1
H_2S (mg l^{-1})	0	0	400
NH_4^+ (µg l^{-1})	516.7	490	410
NO_2^- (µg l^{-1})	19.2	34	49
NO_3^- (µg l^{-1})	63.8	24	21
SO_4^{2-} (g l^{-1})	10.4	22	38
Fe (mg l^{-1})	2.2	4.2	8
Tot_{miner} (g l^{-1})	28.1	54	80

Table2. Results summary of the two-tiered binning approach applied to microbial metagenomes. Taxonomic assignments were predicted according to BLASTp results (*e*-value $<10^{-5}$) using MEGAN, with a minimum support of 20 and a minimum bit score of 100 in alignment.

Depth	Bin	Length (bp)	GC (%)	Assignment (BLAST/MEGAN)	Relative abundance (%)
3.0m	3-1	5,320,493	42.62	*Bacteroidetes* (with hits to *Cyclobacterium marinum* DSM 745)	10.7
	3-2	4,794,946	38.62	*Flavobacteria*	15.7
	3-3	1,367,999	40.41	*Pseudoalteromonas*	1.7
	3-4	1,817,139	49.62	Unknown bacterium	3.6
	3-5	320,919	48.99	*Gammaproteobacteria* /*Rheinheimera*	0.3
	3-6	2,895,826	57.63	Unknown bacterium	9.9
	3-7	616,028	57.66	*Pseudomonas stutzeri* -like bacteria	3.0
	3-8	3,571,362	56.35	*Gammaproteobacteria*	11.7
	3-9	498,520	57.30	*Halomonas*	4.6
	3-10	829,727	65.51	*Rhodobacteraceae*	2.1
	3-11	991,406	64.10	*Hyphomonas neptunium* -like bacteria	3.7
	3-12	591,555	66.11	*Chroococcales*	22.5
	3-13	1,175,729	65.17	*Gammaproteobacteria* /*Alcanivorax*	4.4
	3-14	506,831	63.41	*Verrucomicrobia* /*Chthoniobacter flavus*	2.2
5.0m	5-1	1,745,731	34.09	*Firmicutes* /*Clostridia*	0.5
	5-2	207,265	33.78	*Staphylococcus*	0.0
	5-3	1,149,980	44.71	*Bacteroidetes* (with hits to *Anaerophaga thermohalophila* DSM 12881)	0.6
	5-4	1,769,280	43.68	*Bacteroidetes* (with hits to *Anaerophaga thermohalophila* DSM 12881)	0.4
	5-5	4,105,682	55.09	*Enterobacteriaceae*	3.0
	5-6	1,626,616	64.30	*Thiocapsa*	21.6
5.5m	5.5-1	233,710	28.74	Uncultured bacterium (with hits to *Halanaerobiales*)	0.1
	5.5-2	1,215,938	31.03	Uncultured bacterium (with hits to *Halanaerobium*)	0.6
	5.5-3	1,005,967	35.40	Uncultured bacterium (with hits to *Bacteroidetes* , *Firmicutes* and *Proteobacteria*)	0.4
	5.5-4	451,924	35.33	*Halanaerobium*	0.1
	5.5-5	435,003	36.34	Uncultured bacterium (with hits to *Firmicutes*)	0.1
	5.5-6	1,084,706	37.25	Uncultured bacterium (Candidatus *Cloacamonas acidaminovorans* str. Evry)	1.7
	5.5-7	503,708	39.05	Uncultured *Desulfobacterium* sp. (with hits to delta proteobacterium NaphS2)	0.3
	5.5-8	361,955	38.51	*Bacteroidales* /*Marinilabiaceae*	0.1
	5.5-9	1,017,801	44.56	*Deltaproteobacteria*	0.3
	5.5-10	1,029,185	47.56	Uncultured candidate division OP1 bacterium	1.5
	5.5-11	605,053	45.43	*Sphaerochaeta pleomorpha* -like bacteria	0.2
	5.5-12	1,051,753	47.58	*Desulfobacteraceae*	0.5
	5.5-13	1,216,430	47.17	*Bacteroidetes* /*Marinilabiaceae*	0.3
	5.5-14	271,285	53.82	Unknown bacterium	0.2
	5.5-15	300,891	53.73	*Desulfobacteraceae*	0.1
	5.5-16	419,569	54.53	*Clostridiaceae*	0.1
	5.5-17	1,317,006	56.70	*Halomonas*	0.7
	5.5-18	1,692,070	64.65	*Thiocapsa*	3.6

Table3. Diversity and richness estimates from the viral metagenomes were determined using PHAACS. [a-c]Numbers without a common superscript differed (P<0.05).

	3.0 m-V	5.0 m-V	5.5m-V
Best model	Power	Power	Power
Richness	15877[a]	2498[b]	10000[c]
Evenness	0.94[a]	0.86[b]	0.87[c]
Most abundant genotype (%)	0.90	5.51	3.28
Shannon-Wiener Index (H')	9.15[a]	6.70[b]	8.06[c]