Metagenomic exploration of viral diversity and virus-host interactions in a deep freshwater lake

Yusuke Okazaki[a,b,1], Yosuke Nishimura[c,d,e], Takashi Yoshida[c], Hiroyuki Ogata[d], Shin-ichi Nakano[a]

a. Center for Ecological Research, Kyoto University, 2-509-3 Hirano, Otsu, Shiga 520-2113, Japan

b. Bioproduction Research Institute, National Institute of Advanced Industrial Science and Technology, Central 6, Higashi 1-1-1, Tsukuba, Ibaraki 305–8566, Japan.

c. Graduate School of Agriculture, Kyoto University, Kitashirakawa-Oiwake, Sakyo-ku, Kyoto 606-8502, Japan

d. Institute for Chemical Research, Kyoto University, Uji, Kyoto 611-0011, Japan

e. Atmosphere and Ocean Research Institute, The University of Tokyo, Chiba 277-8564, Japan

[1]Corresponding author:

E-mail: okazaki.yusuke.e31@kyoto-u.jp

Phone: +81-29-861-6591, Fax: +81-29-861-6587

Keywords: freshwater lake, bacterioplankton, virioplankton, metagenomics

Abstract

The use of metagenomics has dramatically expanded the known virosphere, but freshwater viral diversity and the ecological relationships between viruses and hosts remain poorly understood. In this study, we conducted a large-scale metagenomic exploration of planktonic dsDNA prokaryotic viruses by sequencing both virion (<0.22 μm) and cellular (0.22–5.0 μm) fractions collected spatiotemporally from a deep freshwater lake (Lake Biwa, Japan). This simultaneously reconstructed 183 complete (i.e., circular) viral genomes and 57 bacterioplankton metagenome-assembled genomes. Spatiotemporal, intra- and extra-cellular dynamics of individual viruses assessed through metagenomic read coverage revealed a hypolimnion-specific viral community analogous to the vertically-stratified bacterioplankton community. The hypolimnetic community was generally stable during stratification, but occasionally shifted abruptly due to lysogenic induction, presumably reflecting low bacterial productivity in the hypolimnion. Genes involved in assimilatory sulfate reduction were encoded in 20 (10.9%) viral genomes, including those of dominant viruses, and may aid viral propagation in sulfur-limited freshwater systems. Hosts were predicted for 40 (21.9%) viral genomes, encompassing 10 phyla (or classes of Proteobacteria) including ubiquitous freshwater bacterioplankton lineages (e.g., LD12 and *Ca.* Nitrosoarchaeum). Comparison with published metagenomes revealed phylogeographic connectivity of viral communities in geographically isolated habitats, probably following the phylogeographic connectivity of their hosts. Notably, analogous to their hosts, actinobacterial viruses were among the most diverse, ubiquitous, and abundant viral groups in freshwater systems, with high lytic replication rates in surface waters. This result suggested that freshwater Actinobacteria are under high viral lytic pressure, which likely facilitated their genomic micro-diversification to elude viral recognition.

Significance Statement

Metagenomics allows for the reconstruction of bacterial and viral genomes without cultivation, contributing substantially to elucidation of bacterial and viral diversity and their key ecological roles. In this study, we report the largest-scale metagenomic exploration of freshwater prokaryotic viruses to date. We investigated the hypolimnion of a deep freshwater lake and tracked viruses throughout their life cycles, including both intra- and extra-cellular phases. This method allowed us to reconstruct both viral and host genomes and characterize numerous novel virus-host interactions, including those involving ubiquitous freshwater bacterial lineages. When the data were compared with published metagenomic datasets, we uncovered genome-resolved phylogeographic connectivity among viruses from geographically isolated habitats and identified ecologically important viral groups that are ubiquitous in aquatic habitats.

Introduction

Since the discovery three decades ago that a large number of viruses are present in aquatic systems (1), our understanding of viral ecology has grown in an unprecedented manner (2–5). Whereas viruses

can infect cells of all three domains of life, those infecting bacteria and archaea have been studied extensively, as they are numerically dominant and play important roles in their ecosystems. Such viruses rewire the microbial food web by releasing the host's cellular contents back into the environment (the viral shunt) (6, 7) and may carry host-derived metabolic genes (auxiliary metabolic genes; AMGs), which amend the host's metabolism to facilitate viral propagation (8, 9). They are key players in maintaining the diversity of the microbial community by selectively killing the dominant hosts ("Kill the Winner" hypothesis) (10) or promoting the genetic diversification of hosts (11, 12).

Although freshwater ecosystems occupy a small portion of the earth's surface, these areas cycle petagrams of carbon per year globally (13–15), and are of great importance to humans as drinking water sources. Furthermore, freshwater ecosystems are vulnerable to climate change (16, 17) and eutrophication (18). Microbial processes are key to understanding these ecological and biogeochemical mechanisms. Researchers have reached the consensus that surface freshwater systems are ubiquitously dominated by bacterioplankton lineages, as represented by the phyla Actinobacteria, Proteobacteria, and Bacteroidetes (19, 20), and their eco-physiological characterization has been a major research topic in recent years (21–23). In addition, we have previously unveiled the vertical stratification of bacterioplankton communities in deep freshwater lakes, where the oxygenated hypolimnion— i.e., the deep aerobic water layer below the thermocline, which generally constitutes a volumetrically significant part of deep temperate, oligo-mesotrophic lakes— was dominated by specific lineages, including those in Chloroflexi (24, 25), Planctomycetes, Nitrospira and Thaumarchaeota (26, 27). These lineages are likely responsible for important biogeochemical processes in the oxygenated hypolimnion, such as remineralization or conversion of semi-labile organic matter (28, 29), nitrification (30), and methane oxidation (31).

The recent development of sequencing technologies has facilitated viral metagenomics (viromics), opening the door to elucidating the vast diversity of uncultured viral genomes and genes (32–34) and challenging the classical taxonomy of viruses (35). In freshwater systems, early viromics investigations were performed using the 454 sequencing platform, which helped to clarify the uniqueness and diversity of the freshwater virosphere (36–40). Later, the Illumina sequencing platform allowed >10-fold increases in read throughput and library preparation without the biases of multiple displacement amplification, which facilitated the assembly of much longer viral genomic fragments, including complete viral genomes (41–45). Notably, recently published viromes from Lake Neagh, Northern Ireland (43), and Lake Soyang, South Korea (44), were based on deep (>20 Gb) metagenomic sequencing of the virion-sized fraction (<0.2 μm) and contributed substantially to the known freshwater viral sequence space.

Despite this remarkable progress, knowledge of viral diversity and ecology in freshwater systems is nowhere near comprehensive, remaining far behind that of their marine counterparts, which have been analyzed in global-scale investigations (46–51), including the sampling of dsDNA viral

genetic diversity in the surface ocean near saturation levels (33, 49, 52). Some major achievements in marine viromics include the identification of viruses infecting the dominant bacterioplankton lineages (53–56), which have eluded cultivation-based efforts (57–60). In freshwater systems, in addition to viruses that infect cyanobacteria (61–64), which are relatively amenable to cultivation, several viruses that infect ubiquitous freshwater bacterioplankton lineages have been identified in the last few years, including isolated viruses that infect LD28 (*Ca.* Methylopumilus) (44) and Comamonadaceae (65), and metagenomically-recovered viruses of Actinobacteria (66) and *Polynucleobacter* (67). Characterization of such virus-host interactions is key to further understanding the ecological and biogeochemical functions of the microbial community and overall ecosystem. However, viruses infecting other typical freshwater bacterioplankton lineages, such as LD12 (*Ca.* Fonsibacter) (68, 69) and *Limnohabitans* (70) have yet to be characterized. Moreover, despite the ecological and biogeochemical importance of the microbial community of the oxygenated hypolimnion, as described above, no viromics research has been performed in this realm, leaving hypolimnetic viral diversity and virus-host interactions totally unexplored.

To address these issues, the present study used metagenomics to investigate viral diversity and virus-host interactions in a deep freshwater lake. We first performed spatiotemporal sampling of both the cellular (size fraction = 0.22–5.0 μm) and virion (<0.22 μm) size fractions— totaling >50 Gb sequencing reads— including the oxygenated hypolimnion, making this the largest-scale exploration of freshwater viral metagenomes to date. This allowed simultaneous reconstruction of both viral and host (i.e., prokaryotic) genomes from the water column and spatiotemporal analysis of the lifecycle of individual viral lineages, including both intra- and extra-cellular stages. The novelty or ubiquity of the assembled viral genomes was assessed relative to those of published metagenomes and assembled viral genomes recovered from global aquatic habitats.

Results and Discussion

Overview of samples and assemblies

In total, 12 water samples were collected at a pelagic site on Lake Biwa, the largest freshwater lake in Japan. The epilimnion (5 m) was sampled over 3 months and the hypolimnion (65 m) was sampled over 9 months. Water temperature, dissolved oxygen concentration, and the abundances of bacteria and virus-like particles during the study period (Table S1) were within the range reported from previous studies of this lake (24, 71, 72). Notably, thermal stratification was observed from June to December and the hypolimnion was always oxygenated, with dissolved oxygen concentrations of >6.0 mg L^{-1}. Metagenomic sequencing with subsequent assembly (Table S2), decontamination, and dereplication steps eventually yielded 4,158 viral contigs of >10 kb (Lake Biwa Viral Contigs [LBVCs]), including 183 complete (i.e., circular) genomes (Lake Biwa Viral Genomes [LBVGs]; Fig. S1). The GC content of the LBVGs ranged from 29.1 to 66.7%, and all genome sizes were <120 kb except for LBVG_1 (318 kb) and LBVG_2 (190 kb; Fig. 1A), which were within the range of known

prokaryotic dsDNA viral genomes (73, 74). Functional annotation of open reading frames (ORFs) demonstrated that 177 of the 183 LBVGs contained a terminase large subunit gene, and 181 LBVGs had one or more tailed-virus hallmark genes (i.e., terminase, tail, capsid or portal), indicating that they were *bona fide* viral genomes. The two LBVGs lacking annotated hallmark genes (LBVG_174 and LBVG_183) were considered putative viral genomes and included in the analysis, as their close relatives were assembled in published viromes and their reads were represented in the virion fraction (SI Dataset S1) corresponding to encapsidated particles purified through cesium chloride (CsCl) ultracentrifugation (75).

Similarity between viral genomes was evaluated using S_G, a similarity measure (ranging from 0 to 1) based on genome-wide tBLASTx scores (55). A previous study demonstrated that S_G >0.15 was the best threshold for genus-level clustering of viral genomes (55). Using this threshold, the 183 LBVGs were grouped into 28 genomic operational taxonomic units (gOTUs) containing 2–10 LBVGs each, leaving 90 LBVGs ungrouped (Fig. 2). To assess the novelty or ubiquity of LBVGs, published viral genomic sequences were compiled into two databases— the Reference Viral Genome (RVG) and Environmental Viral Genome (EVG). The RVG included 2,621 isolated prokaryotic dsDNA viral genomes. The EVG included 2,860 metagenomically recovered viral genomes (43, 44, 47, 50, 52, 54, 55) and manually collected viral genomes that had been reported to infect the dominant freshwater bacterioplankton lineages, including both isolated and metagenome-assembled genomes (44, 64–67) (Fig. S2). Using the threshold of S_G >0.15, 4 (2.2%) and 127 (69.4%) LBVGs were found to have genus-level relatives in the RVG and EVG, respectively. An S_G-based proteomic tree including all LBVGs and EVGs and alignments among genomes is available at https://www.genome.jp/viptree/u/LBV/retree/LBVGandEVG.

Metagenomic binning of the contigs assembled from the cellular fraction recovered 57 metagenome-assembled genomes of bacterioplankton (Lake Biwa metagenome-assembled genomes [LBMAGs]), including diverse bacterial and archaeal phyla (Figs. 3 and S3). Forty-five (78.9%) of these genomes were of high quality, with estimated completeness and contamination scores of >80% and <5%, respectively (Table S3). Consistent with previous observations (26, 27), coverage-based relative abundance estimates (fragments per kilobase per mapped million reads [FPKM]) indicated a depth-stratified bacterioplankton community— members of LD12, Actinobacteria, and Cyanobacteria were dominant in the epilimnion, whereas CL500-11 and *Ca.* Nitrosoarchaeum were dominant in the hypolimnion (Table S4). In several lineages (e.g., LD12, LD28, and OPB56), distinct genomes were recovered from the epilimnion and hypolimnion, indicating micro-diversification between the water layers. Two LBMAGs representing candidate phyla, *Ca*. Levybacteria and *Ca*. Saccharibacteria (76), had not been reported in a previous study conducted at Lake Biwa using 16S rRNA gene amplicon sequencing (27), as these phyla harbor 16S rRNA gene sequences that are undetectable with conventional universal primers.

Spatiotemporal dynamics of the viral community

Relative abundance estimates of LBVGs and LBVCs revealed that the epilimnion- and hypolimnion-specific viral lineages were clearly separated (Fig. 1B, C). Among the 183 LBVGs, 60 were epilimnion-specific and 62 were hypolimnion-specific, demonstrating epilimnion-hypolimnion preference (P_{epi}) values of >95% and <5%, respectively. The depth-stratified viral community is analogous to that reported from ocean (46, 52–54, 77) and likely reflects the depth-stratified community of bacterioplankton (27) (Table S4).

The Bray-Curtis dissimilarity of the viral community between consecutive months revealed that the hypolimnetic community was largely stable (0.21–0.34; except for two months discussed later) compared to the epilimnetic community (0.54–0.68) (Fig. 4). Similar patterns were reported for the bacterioplankton community in the lake (27). In line with previous studies reporting the rapid succession of the viral community in surface freshwater systems (43, 78), most LBVGs in the epilimnion were dominant (ranked within the 10 most abundant LBVGs) for no more than a month, aside from several actinoviruses that continuously dominated the community (Fig. 4). By contrast, in the hypolimnion, more viruses were persistently dominant for a month or more (Fig. 4). Notably, the decay rate of virioplankton ranges from 0.14 to 54% h^{-1} (half-life = 0.9 h to 20.6 days) (3), meaning that most virions in the water column turn over within a month. In addition, a recent study in a marine system reported that the majority of metagenomically-detected viruses were transcriptionally active within a period of 24 h (79). Consequently, the continued dominance of some viral lineages in the hypolimnion likely resulted from the continuous production of virions, rather than their carryover for more than a month. This pattern might be attributable to the lower production of bacterioplankton in the hypolimnion, resulting in a prolonged viral lytic cycle. Accordingly, a previous study in Lake Biwa estimated that bacterial production was 10-fold lower in the hypolimnion (0.4 ± 0.1 μg C l^{-1} d^{-1}) than in the epilimnion (4.2 ± 3.2 μg C l^{-1} d^{-1}), and the percentage of daily bacterial production lost to viruses was 4-fold lower in the hypolimnion (13.6 ± 5.2%) than in the epilimnion (52.7 ± 16.2%) (72).

In addition to the overall stability of the viral community in the hypolimnion, two dynamic community shifts were observed in the virion fraction, from September to October (Bray-Curtis dissimilarity = 0.54) and from December to January (dissimilarity = 0.50; Fig. 4). The first shift was attributed to several viruses switching from lysogeny to lysis (i.e., induction). In the cellular fraction, LBVG_87 was the most abundant LBVG in September but its abundance decreased abruptly in October, whereas its abundance in the virion fraction increased 120-fold over the same period, making LBVG_87 the most abundant virioplankton (Figs. 4, S4 and S5). Similarly, the relative abundances of LBVG_111 and LBVG_121 in the virion fraction increased >1800-fold and 38-fold from September to October, respectively, likely due to the induction of abundant lysogens in September (Figs. 4, S4 and S5). Unfortunately, the host for LBVG_87, LBVG_111, and LBVG_121 could not be identified. As proposed for the deep ocean (46, 54, 80, 81), the lysogenic strategy in deeper water likely results

from low bacterioplankton productivity, which is insufficient to support lytic viral propagation. Given that the hypolimnion during stratification is relatively stable in terms of lysis-inducing environmental stressors such as pH, temperature, nutrients, oxidative stress, and solar radiation (4, 82), the hypolimnetic induction events were likely trigged by the improved physiological condition of the host, with the virus acting as a "time bomb" and breaking the symbiotic relationship with their host once the host has successfully bloomed in the environment (81, 83). No difference in cellular-virion preference (P_{cell}) was observed between water layers (Fig. S6). Thus, lysogeny and induction may also be a common strategy in the epilimnion, although temporal trends were not clearly observed in that layer, presumably due to the rapid succession of the epilimnetic virion community, which could not be fully resolved through monthly monitoring.

The second shift in the virion community from December to January likely resulted from the onset of winter vertical mixing (Table S1), which intermingled the epilimnetic and hypolimnetic viral communities, as observed for bacterioplankton (24, 25, 27). Alternatively, environmental stresses associated with the mixing event (e.g., increased solar radiation) could have induced lysogens in the lytic cycle (84) and resulted in the observed viral community shift.

Possible metabolic reprogramming of hosts by viral AMGs

Functional annotation of viral genes (SI Dataset S1) revealed that LBVGs encoded AMGs that have previously been identified in marine viromes (8, 49, 77, 85, 86). For example, two cyanoviruses (LBVG_2 and LBVG_12) harbored the photosystem II D1 protein (*psbA*) gene, which accelerates the host's photosynthetic light reaction rates to facilitate viral propagation. Ribonucleotide reductases (RNRs) and thymidylate synthetase (ThyX), which are involved in biosynthesis of dNTPs required for viral replication, were encoded in four and seven LBVGs, respectively. LBVG_55, one of the dominant viruses in the hypolimnion with an unknown host (Fig. 4), encoded genes (*cobS* and *cobT*) involved in the biosynthesis of cobalamin, which is often absent from bacterial genomes, thus limiting their metabolic capability (87). Several LBVGs harbored a cluster of genes involved in the biosynthesis of lipopolysaccharides (LPS) and capsular polysaccharides. For example, LBVG_26 encoded genes involved in the biosynthesis of N-acetylneuraminic acid (*neuA*, *neuB*, and *neuC*), in addition to genes annotated as encoding putative sugar epimerase (*capD*), amidotransferase (*pseA*), aminotransferase (*pglC*), and acyltransferase (*pglD*). These genes may alter the cell surface properties of their hosts to prevent infection with other viruses (88) or to provide tolerance to temperature, oxidative stress, or antibiotics (55). Proteins in the adenylyl-sulfate kinase (CysC) and 3'-phosphoadenosine 5'-phosphosulfate (PAPS) reductase (CysH) families were encoded in 4 and 17 LBVGs, respectively, and LBVG_98 contained both (SI Dataset S1). These genes are involved in assimilatory sulfate reduction and likely facilitate the host's utilization of reduced sulfur compounds (89, 90). These genes are present in ubiquitous and abundant viral groups (e.g., gOTU_3 and gOTU_11), indicating that they are quantitatively significant in the ecosystem. By contrast, these

genes were not reported as often as AMGs in marine ecosystems (8, 77). Given that freshwater systems are sulfur-limited, and that the genomes of many predominant freshwater bacterioplankton lineages such as LD12 (68, 91), acI-B1 (23), and CL500-11 (92) lack the complete pathway for assimilatory sulfate reduction, these viral-encoded genes are likely key factors modulating microbial sulfur metabolism in the lake.

Ecological role of viruses infecting the dominant bacterioplankton

Hosts were predicted for 40 LBVGs using a combination of *in silico* approaches (SI Materials and Methods and Table S5). The predicted hosts spanned 10 phyla (or Proteobacteria classes; Fig. 3), and the host of 6 LBVGs could be further resolved to the genus level (Table S5). Notably, we identified 13 LBVGs infecting Actinobacteria, which is one of the most diverse, abundant, and ubiquitous groups of freshwater bacterioplankton (20). Aside from LBVG_68 (genome size = 40.0 kb), all actinoviral LBVGs had small genomes (14.0–19.8 kb) and formed a monophyletic clade composed of four gOTUs (gOTU_8, gOTU_13, gOTU_14, and gOTU_28; Fig. 2). These LBVGs were closely related to viral genomes that originated from other freshwater systems (Fig. 5 and SI Alignment S1), including relatives of the G4 actinophage (uvFW-CGR-AMDFOS-S50-C341), one of four actinoviral groups revealed metagenomically in a Spanish reservoir (66). Relatives of the other three Spanish actinoviral groups (G1, G2, and G3) were absent from our LBVGs but were observed among LBVCs (Fig. S7) and EVGs recovered from other freshwater lakes (SI Alignment S2–4). Together, these results suggest that these actinoviral groups are ubiquitous in freshwater systems globally. Their broad range of GC contents (41.6–62.6%) indicates that their hosts include diverse members of Actinobacteria that also exhibit a wide range of GC contents (Table S3). Their predominance in both the cellular and virion fractions in the epilimnion (Figs. 4 and S8) suggests that they were actively replicating and lysing the host cells to produce virions. This finding is intriguing given the predominance of Actinobacteria in freshwater systems and evokes the "King of the Mountain" (KoM) hypothesis, which was proposed to explain the co-dominance of marine *Pelagibacter* and its viruses (93). In the KoM hypothesis, the predominance of the host species under high viral lytic pressure provides a greater chance for the host to acquire resistance against viral infection through genomic recombination, and this generates a positive feedback loop for host propagation. Supporting this hypothesis, single-cell amplified genomes of acI, the most representative freshwater Actinobacterial lineage, exhibited high intra-population diversification, suggesting that acI members have a high recombination rate (22). Another study revealed that members of the acI lineage commonly harbor genomic islands containing genes that may be involved in viral host recognition (e.g., extracellular structural genes), even among closely related strains (average nucleotide identity [ANI] >97%) inhabiting the same lake (23). Accordingly, the most closely related pair of actinoviral LBVGs in the present study (LVBG_169 and LBVG_172; $S_G = 0.91$, ANI = 96.4%) showed diversification in tail-related genes (e.g., tape measure protein), which are often involved in viral recognition of the host cell surface (94, 95) (SI Alignment S1). Moreover, the

streamlined acI genomes do not harbor the CRISPR-Cas system (23), and a recent meta-epigenomic analysis conducted at Lake Biwa demonstrated that the genomes of Actinobacteria lack both methylated DNA motifs and methyltransferase genes (96), implying that they lack the restriction-modification system for resisting viral infection. These findings further support the KoM hypothesis, which assumes that the host prefers a strategy against resource competition rather than act as a defense specialist (93). In light of this hypothesis, the high abundance and diversity of actinoviruses may result from an arms race with their hosts, leading to recombinational diversification of their genomes to avoid mortality induced by a high viral load. It should be noted that such microdiversification of viral genomes poses challenges for metagenomic assembly (97–99); indeed, 109 LBVCs encoded the characteristic actinoviral *whiB* gene (SI Dataset S1), suggesting that the majority of actinoviral genomes were not recovered as LBVGs (i.e., circularly assembled). Although further investigation is beyond the scope of the present study, our results revealed that, analogous to their hosts, actinoviruses are one of the most diverse, abundant, ubiquitous, and active viral groups in freshwater systems.

LD12 (*Ca*. Fonsibacter) is a freshwater division of the SAR11 lineage (Pelagibacteraceae), which is one of the most ubiquitous and abundant bacterioplankton in freshwater, as is the case for its marine counterparts (68, 69, 100). Indeed, LD12 was by far the most abundant LBMAG in the epilimnion throughout the study period (Table S4). Whereas viruses infecting marine SAR11 have been characterized with (59, 101) and without (50, 102) cultivation-dependent methods, those infecting LD12 have not been reported before the present study to our knowledge. Unlike actinoviruses, the three predicted LD12 viruses were not clustered with other LBVGs (Fig. 2) and thus their genomic diversification appears limited, possibly reflecting the fact that LD12 bacteria also show limited genomic diversification (22, 103). Among these LBVGs, LBVG_1 was the largest (317.9 kb) viral genome assembled in the present study. In addition to the characteristic genes of a tailed virus, LBVG_1 encoded many genes closely related to those of Pelagibacterales, including the 30S ribosomal protein S21, which is the most abundant virus-encoded ribosomal protein in aquatic systems and presumably facilitates takeover of the host's translational machinery (104) (SI Dataset S1). LBVG_1 was abundant in the cellular fraction in the epilimnion but rare in the virion fraction (Figs. 4 and S8), suggesting that lysogeny was their primary strategy during the study period. By contrast, the two other predicted LD12 viruses (LBVG_76 and LBVG_174) were abundant in both the cellular and virion fractions (Figs. 4 and S8), suggesting that these viruses were actively replicating and lysing their hosts. However, close relatives of these viruses have not been recovered in other freshwater systems (Fig. 2), and most of their genes were not functionally annotated, so further research is needed to elucidate their ecology. Overall, these results suggest that, in accordance with results in marine counterparts, viruses are an important driving factor for the population dynamics of LD12, which is one of the most quantitatively significant components of freshwater microbial ecosystems.

The present study recovered a LBMAG of *Ca*. Nitrosoarchaeum (Fig. 3), which is a member

of the ammonia-oxidizing, chemoautotrophic archaeal taxon Marine Group I (MGI). Members of MGI dominate the hypolimnion of deep freshwater lakes (26) (Table S4) and thus are important players in the nitrogen and carbon cycles of lakes. In marine systems, members of MGI are also dominant in the deep aphotic layer, and recent studies have reported the presence of several putative MGI viruses (105, 106), including those encoding AMGs such as *amoC* (107) and *cobS* (86). The present study predicted one *Ca*. Nitrosoarchaeum virus (LBVG_51)— the first report of a virus predicted to infect a freshwater planktonic MGI member. The virus did not show genomic similarities to known marine MGI viruses or to any genomes in the EVG database. This genome encoded replication protein A and minichromosome-maintenance helicase genes, which drive the archaeal DNA replication system (108), presumably to facilitate viral genomic replication in the host (109). LBVG_51 was mainly detected in the cellular fraction in the hypolimnion (Figs. 4 and S8), suggesting that lysogeny was their primary strategy during the study period. A previous study reported an abrupt drop in *Ca*. Nitrosoarchaeum abundance after the onset of the mixing period (27), but no evidence for mass induction was found in the present study. We further speculated that one incomplete contig, LVBC_1935 (16,737 bp), might also be derived from a *Ca*. Nitrosoarchaeum virus, as it encoded multiple archaeal genes and exhibited a GC content and spatiotemporal distribution corresponding to those of *Ca*. Nitrosoarchaeum (SI Dataset S1).

Viruses infecting other bacterioplankton taxa, namely Cyanobacteria, Bacteroidetes, Betaproteobacteria, Gammaproteobacteria, Nitrospira, Planctomycetes, and Verrucomicrobia were also predicted. Most of these virus-host interactions were newly discovered, and their characterization was facilitated by LBMAGs (SI Results and Discussion). Overall, the present study reports the largest-scale prediction of virus-host pairs in a freshwater system to date, as well as describes numerous key ecological groups of viruses that infect ubiquitous bacterioplankton lineages.

Viral genomic diversification among habitats

Most (116 of 183) of the LBVGs were closely related to genomes that were recovered from other freshwater habitats (Fig. 2), indicating that they belonged to ubiquitous freshwater viral lineages. Some of these lineages exhibited remarkably high inter-lake genomic similarities. For example, among actinoviruses (Fig. 5), LBVG_182 and 3300002835.a_100013722, a genome retrieved from Lake Mendota, USA (110), exhibited an S_G value of 0.76 and an ANI value of 84% (SI Alignment S1). Moreover, LBVG_144 and Soyang_Oct2014_scaffold_160, which has an unknown host, exhibited an S_G value of 0.94 and an ANI value of 99% (SI Alignment S5). Although nearly identical viral genomes have been reported from distant locations in marine environments, such as the Mediterranean Sea and Pacific Ocean (95), these results are unexpected in freshwater systems, as lakes and reservoirs are physically isolated from one another. Notably, the genomes of bacterioplankton inhabiting different freshwater systems are generally distinct even among those with nearly identical 16S rRNA gene sequences (22, 23, 111, 112), although highly similar genomes (ANI >95%) have occasionally been

reported from lakes on different continents (92). The migratory and evolutionary mechanisms behind these similarities are not well-understood, but our results revealed that common viral groups inhabit separate freshwater ecosystems, likely reflecting the phylogeographic connectivity of the host organisms.

Aside from this evidence of connectivity among freshwater habitats, we also observed 34 LBVGs that were similar (S_G >0.15) to marine viral genomes (Fig. 2). For example, gOTU_26 (LBVG_135 and LBVG_162)— one of the dominant viruses in the hypolimnion with an unidentified host (Fig. 4)— had a conserved genomic structure with viral genomes originating from a broad range of aquatic habitats— surface and deep waters of the Mediterranean Sea, Pacific Ocean, Atlantic Ocean, Arabian Sea, Osaka Bay, Lake Neagh and Lake Soyang (SI Alignment S6). As observed for cyanoviruses (62, 64), close phylogenetic relationships among marine and freshwater viruses suggest that their hosts are bacterioplankton belonging to taxa containing both marine and freshwater members, such as Pelagibacteriaceae (91), Methylophilaceae (113), and Flavobacteriaceae (114). Indeed, LBVG_76 and LBVG_174, which are putative LD12 viruses, share genomic structures with marine SAR11 phages recovered from Mediterranean viromes (50, 54) and the *Tara* Oceans Virome (47) (SI Alignments S7 and S8). A recent study (115) has challenged the idea that marine-freshwater transitions by bacterioplankton are infrequent (91, 116, 117), and viral diversification processes that allow crossing of this salt divide in association with their hosts are an intriguing research topic. Notably, the greatest similarity between marine and freshwater viral genomes was observed among the *Synechococcus* viruses S-EIV1, LBVG_9, and TARA_ERS488929_N000037 (collected at a pelagic site in the Arabian Sea), which exhibited higher similarity scores (S_G = 0.52) than among those collected in freshwater systems or even within the same lake (SI Alignment S9). Thus, in the diversification processes of at least some viral lineages, the transition between marine and freshwater habitats appears to be a relatively frequent event.

In the present study, 56 of 183 LBVGs represented novel viral lineages that lack close relatives in the EVG database. These lineages included those that were dominant in the water column (e.g., LBVG_15, LBVG_165, and LBVG_168 in the epilimnion and LBVG_43, LBVG_111, and LBVG_156 in the hypolimnion; Fig. 4), indicating their quantitative significance in the ecosystem. Although these sequences might represent *bona fide* endemic viruses in Lake Biwa, this result was more likely due to the limited sequencing conducted for freshwater viral genomes. Current freshwater viral genome databases rely heavily on two deeply-sequenced viromes from Lake Neagh (43) and Lake Soyang (44). In fact, 787 (90.6%) of the 869 freshwater genomes identified as close relatives of LBVGs were derived from these two viromes (SI Dataset S1).

Conclusions

The present study reports the largest-scale metagenomic survey of freshwater prokaryotic viral diversity and includes the reconstruction of 183 complete viral genomes (LBVGs) and 57 draft host

genomes (LBMAGs). Notably, we obtained viromes from the oxygenated hypolimnion of a deep freshwater lake and revealed that the vertical stratification of the viral community is analogous to the distribution pattern of their hosts. Collection of samples over a broad spatiotemporal range from both the cellular and virion fractions allowed the capture of viruses during various stages of their lifecycles, including lysogeny (e.g., LBVG_51), induction (e.g., LVBG_87), and active replication-lysis (e.g., LBVG_179). Analysis of AMGs suggested that bacterial sulfate assimilation might often be assisted by virally encoded genes in sulfur-limited freshwater habitats. Host prediction substantially expanded our knowledge of freshwater virus-host interactions and identified potential key viral players in the ecosystem, which infect the dominant bacterioplankton lineages. Using published metagenomic datasets, we demonstrated the phylogeographic connectivity of viruses in different aquatic habitats and identified ubiquitous viral lineages. Overall, the present study provides a foundation for future investigations of freshwater viral ecology and evolution in relation to their hosts, while also revealing that our sampling was not saturated, motivating efforts toward further sample collection, deeper sequencing, and the introduction of novel technologies such as viral meta-transcriptomics (79), single-virus genomics (97) and long-read viral metagenomics (99).

Materials and Methods

Sample collection, sequencing, and assembly

Details of the sample collection, sequencing, and assembly procedures are available in SI Materials and Methods and Figure S1. Briefly, water samples were collected monthly at a pelagic station on Lake Biwa from June 2016 to February 2017 (Table S1). The cellular fraction (0.22–5.0 μm) was obtained shipboard through filtration. The virion fraction (<0.22 μm) was collected from filtered water by concentrating virions using iron(III) chloride flocculation (118), followed by purification using CsCl density gradient ultracentrifugation (75). After DNA extraction and library preparation for shotgun metagenomics, two sequencing runs (2 × 300 bp) were performed using the Illumina MiSeq system for each of the cellular and virion fractions. The sequence reads generated were (co-)assembled using metaSPAdes v. 3.10.1 (119) for the cellular fraction and SPAdes v. 3.9.0 (120) for the virion fraction (Table S2).

Construction of reference databases

The RVG and EVG databases were originally described by Nishimura et al. (55), and were updated in the present study (Fig. S2). The RVG was prepared by extracting 2,621 genomes of isolated dsDNA prokaryotic viruses from the Virus-Host Database (as of Feb. 6, 2018) (121). The EVG was built based on 1,811 genomes compiled for a previous study (55) and augmented with the following datasets: (i) 125,842 mVCs (metagenomic Viral Contigs), which were generated from metagenomes deposited in the IMG/M database (52); (ii) 28 complete viral genomes retrieved from a Mediterranean deep-ocean virome (54); (iii) contigs assembled from six viromes collected seasonally from Lake Soyang (PRJEB15535) (44); (iv) 488 contigs previously compiled from published freshwater viromes (43);

and (v) 313 complete viral genomes recovered from Lake Neagh (43). In addition, the genomes of recently reported freshwater viruses, including the *Synechococcus* virus S-EIV1 (64), eight actinoviruses (66), a LD28 virus (44), two Comamonadaceae viruses (65), and a *Polynucleobacter* virus (67) were compiled. Assembly of the Lake Soyang virome was performed using the same parameters used for the Lake Biwa virome. The contigs were inspected for circularity using ccfind v. 1.1 (https://github.com/yosuken/ccfind) (55). Only contigs that were >10 kb were retained and dereplicated using the BLASTn search against themselves with NCBI-BLAST+ v. 2.7.1 (122), where all >500 nt high-scoring segment pairs were considered and contigs with >99% nucleotide identity across >95% of their length were identified as duplicates. The resulting EVG database included 2,860 viral genomes.

Extraction of viral contigs

Viral contigs were extracted from 11 assemblies obtained from the viral fraction and 7 assemblies from the cellular fraction (Table S2). Only contigs longer than 10 kbp were considered for subsequent analysis. Identification of viral contigs was performed with VirSorter v. 1.0.3 (123) using Virome Decontamination mode with reference to the Virome Database, and all contigs assigned to categories 1–3 (virus and prophage) were extracted. The resulting 8,173 contigs were inspected for redundancy using BLASTn searches against themselves, where all >500 nt high-scoring segment pairs were considered, contigs with >95% nucleotide identity across >80% of the contig length (98) were identified as redundant, and the shorter contig in the pair was discarded. The resulting 4,315 contigs retained some apparent bacterial genomic fragments, presumably because they included category 3 (the least confident category) contigs, as assigned using VirSorter at the cost of an increased false-positive rate (123). Further decontamination was performed through inspection for viral hallmark genes. To this end, 134,388 open reading frames (ORFs) on 4,315 contigs were predicted with Prodigal v. 2.6.3 (124) and functionally annotated (see SI Materials and Methods for functional annotation workflow). Genes annotated as encoding terminase, tail, capsid or portal proteins were regarded as viral hallmark genes. Among contigs with no viral hallmark genes, those annotated as prophages using VirSorter (i.e., containing cellular hallmark genes) and those exclusively detected in the cellular fraction based on read coverage were regarded as bacterial genomic contaminants and were removed. Consequently, 4,158 contigs were retained, which were designated as LBVCs (Lake Biwa Viral Contigs). Further, 183 complete (i.e., circular) genomes were identified using ccfind v. 1.1 and designated as LBVGs (Lake Biwa Viral Genomes). Handling of FASTA files and calculations of genome size and GC contents were performed using SeqKit v. 0.5.0 (125).

Phylogenetic analysis of viral genomes

Calculation of S_G and construction of a S_G-based proteomic tree were performed using ViPTreeGen v. 1.1.1, a standalone version of ViPTree (126). The tree was drawn from the output newick file using the phytools package (127) and visually annotated with the circlize package (128) of R software v.

3.4.3 (https://www.R-project.org). LBVGs were searched against the RVG and EVG databases using the 2D mode of ViPTreeGen v. 1.1.1, and the number of hits with S_G >0.15 was determined for each LBVG. ANI was calculated using the ani.rb script in the enveomics collection (129).

Estimation of viral abundance based on read coverage

Calculation of FPKM was performed with the CountMappedReads v. 1.0 script (https://github.com/yosuken/CountMappedReads) using default parameters. From the cellular fraction, the quality-trimmed reads used for assembly were also employed for mapping. From the virion fraction, the raw reads were quality-trimmed in the same way as the cellular fraction and then employed for mapping. Habitat preference for the epilimnion over the hypolimnion (P_{epi}) was measured as the quotient of abundance (in FPKM) in the epilimnion versus the sum of abundance in the epilimnion and hypolimnion (i.e., epilimnion/[epilimnion + hypolimnion]), and habitat preference for the cellular fraction over the virion fraction (P_{cell}) was measured as the quotient of abundance in the cellular fraction versus the sum of abundance in the cellular and virion fractions (i.e., cellular/[cellular + virion]). Note that only samples for which both virion and cellular fractions were available (i.e., July–September in the epilimnion and September–December in the hypolimnion) were considered for this calculation.

Binning of prokaryotic genomes

Binning of LBMAGs was performed using MetaBAT v. 0.32.5 (130) (Fig. S1 and Table S2; see SI Materials and Methods for details). Phylogenetic assignment of LBMAGs was carried out by drawing a phylogenetic tree with published freshwater bacterioplankton genomes using PhyloPhlAn v. 0.99 (131) (Figs. 3 and S3) and the assignments were confirmed using GTDB-Tk v. 0.2.1 with reference to the release 86 database (132). The relative abundance (as FPKM) of LBMAGs in each sample was calculated using the CountMappedReads v. 1.0 script, as described above.

Host prediction

The host of each LBVG was predicted from combinations of *in silico* analyses, as described in SI Materials and Methods and Table S5.

Data availability

Raw reads from the cellular and virion fractions were deposited under accession numbers PRJDB6644 and PRJDB7309, respectively. LBMAGs were deposited under the BioSample identifiers SAMD00166046–00166102. The nucleotide sequences of the LBVGs, LBVCs, RVG, and EVG are available from https://doi.org/10.6084/m9.figshare.7934924.

Acknowledgments

This work was supported by JSPS KAKENHI (Nos. 15J00971, 16H06429, 16K21723, 16H06437, 17H03850, 17K19289, and 18J00300), the Environment Research and Technology Development Fund (No. 5-1607) of the Ministry of the Environment, Japan, and the Canon Foundation (No. 203143100025). Computation time was provided by the SuperComputer System, Institute for

Chemical Research, Kyoto University. Field sampling on Lake Biwa was conducted using the research vessel Hasu belonging to the Center for Ecological Research, Kyoto University, and supported by Y. Goda and T. Akatsuka. We are grateful to T. Narihiro and M. K. Nobu for their helpful advice about metagenomic analysis, and to H. Watai, N. Haruki, and M. Yamanaka for their assistance in sample preparation. We thank K. Arkhipova and K. Moon for providing the original datasets from their published works.

References

1. Bergh Ø, BØrsheim KY, Bratbak G, Heldal M (1989) High abundance of viruses found in aquatic environments. *Nature* 340(6233):467–468.
2. Suttle CA (2005) Viruses in the sea. *Nature* 437(7057):356–361.
3. Wommack KE, Colwell RR (2000) Virioplankton: Viruses in Aquatic Ecosystems. *Microbiol Mol Biol Rev* 64(1):69–114.
4. Weinbauer MG (2004) Ecology of prokaryotic viruses. *FEMS Microbiol Rev* 28(2):127–181.
5. Rohwer F, Thurber RV (2009) Viruses manipulate the marine environment. *Nature* 459(7244):207–212.
6. Wilhelm SW, Suttle CA (1999) Viruses and Nutrient Cycles in the Sea. *Bioscience* 49(10):781–788.
7. Fuhrman JA, Noble RT (1995) Viruses and protists cause similar bacterial mortality in coastal seawater. *Limnol Oceanogr* 40(7):1236–1242.
8. Hurwitz BL, U'Ren JM (2016) Viral metabolic reprogramming in marine ecosystems. *Curr Opin Microbiol* 31:161–168.
9. Breitbart M, Thompson L, Suttle C, Sullivan M (2007) Exploring the Vast Diversity of Marine Viruses. *Oceanography* 20(2):135–139.
10. Thingstad TF (2000) Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. *Limnol Oceanogr* 45(6):1320–1328.
11. Koskella B, Brockhurst MA (2014) Bacteria–phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS Microbiol Rev* 38(5):916–931.
12. Touchon M, Moura de Sousa JA, Rocha EP (2017) Embracing the enemy: the diversification of microbial gene repertoires by phage-mediated horizontal gene transfer. *Curr Opin Microbiol* 38:66–73.
13. Cole JJ, et al. (2007) Plumbing the global carbon cycle: Integrating inland waters into the terrestrial carbon budget. *Ecosystems* 10(1):171–184.
14. Tranvik LJ, et al. (2009) Lakes and reservoirs as regulators of carbon cycling and climate. *Limnol Oceanogr* 54(6part2):2298–2314.
15. Raymond PA, et al. (2013) Global carbon dioxide emissions from inland waters. *Nature*

503(7476):355–359.

16. Posch T, Köster O, Salcher MM, Pernthaler J (2012) Harmful filamentous cyanobacteria favoured by reduced water turnover with lake warming. *Nat Clim Chang* 2(11):809–813.
17. Woolway RI, Merchant CJ (2019) Worldwide alteration of lake mixing regimes in response to climate change. *Nat Geosci* 12(4):271–276.
18. Jenny J, Normandeau A, Francus P, Ecaterina Z, Gregory-eaves I (2016) Urban point sources of nutrients were the leading cause for the historical spread of hypoxia across European lakes. 113(45):1–6.
19. Zwart G, Crump BC, Kamst-van Agterveld MP, Hagen F, Han SK (2002) Typical freshwater bacteria: An analysis of available 16S rRNA gene sequences from plankton of lakes and rivers. *Aquat Microb Ecol* 28(2):141–155.
20. Newton RJ, Jones SE, Eiler A, McMahon KD, Bertilsson S (2011) A guide to the natural history of freshwater lake bacteria. *Microbiol Mol Biol Rev* 75(1):14–49.
21. Salcher MM (2013) Same same but different: ecological niche partitioning of planktonic freshwater prokaryotes. *J Limnol* 73(s1):74–87.
22. Garcia SL, et al. (2018) Contrasting patterns of genome-level diversity across distinct co-occurring bacterial populations. *ISME J* 12(3):742–755.
23. Neuenschwander SM, Ghai R, Pernthaler J, Salcher MM (2018) Microdiversification in genome-streamlined ubiquitous freshwater Actinobacteria. *ISME J* 12(1):185–198.
24. Okazaki Y, Hodoki Y, Nakano S (2013) Seasonal dominance of CL500-11 bacterioplankton (phylum Chloroflexi) in the oxygenated hypolimnion of Lake Biwa, Japan. *FEMS Microbiol Ecol* 83(1):82–92.
25. Okazaki Y, Salcher MM, Callieri C, Nakano S (2018) The Broad Habitat Spectrum of the CL500-11 Lineage (Phylum Chloroflexi), a Dominant Bacterioplankton in Oxygenated Hypolimnia of Deep Freshwater Lakes. *Front Microbiol* 9. doi:10.3389/fmicb.2018.02891.
26. Okazaki Y, et al. (2017) Ubiquity and quantitative significance of bacterioplankton lineages inhabiting the oxygenated hypolimnion of deep freshwater lakes. *ISME J* 11(10):2279–2293.
27. Okazaki Y, Nakano S-I (2016) Vertical partitioning of freshwater bacterioplankton community in a deep mesotrophic lake with a fully oxygenated hypolimnion (Lake Biwa, Japan). *Environ Microbiol Rep* 8(5):780–788.
28. Maki K, et al. (2010) Autochthonous origin of semi-labile dissolved organic carbon in a large monomictic lake (Lake Biwa): Carbon stable isotopic evidence. *Limnology* 11(2):143–153.
29. Thottathil SD, et al. (2013) Biogeochemical control on fluorescent dissolved organic matter dynamics in a large freshwater lake (Lake Biwa, Japan). *Limnol Oceanogr* 58(6):2262–2278.
30. Small GE, et al. (2013) Rates and controls of nitrification in a large oligotrophic lake. *Limnol Oceanogr* 58(1):276–286.

31. Murase J, Sugimoto A (2005) Inhibitory effect of light on methane oxidation in the pelagic water column of a mesotrophic lake (Lake Biwa, Japan). *Limnol Oceanogr* 50(4):1339–1343.
32. Brum JR, et al. (2016) Illuminating structural proteins in viral "dark matter" with metaproteomics. *Proc Natl Acad Sci* 113(9):2436–2441.
33. Cesar Ignacio-Espinoza J, Solonenko SA, Sullivan MB (2013) The global virome: not as big as we thought? *Curr Opin Virol* 3(5):566–571.
34. Brum JR, Sullivan MB (2015) Rising to the challenge: accelerated pace of discovery transforms marine virology. *Nat Rev Microbiol* 13(3):147–159.
35. Simmonds P, et al. (2017) Virus taxonomy in the age of metagenomics. *Nat Rev Microbiol* 15(3):161–168.
36. Roux S, et al. (2012) Assessing the diversity and specificity of two freshwater viral communities through metagenomics. *PLoS One* 7(3):e33641.
37. Green JC, Rahman F, Saxton MA, Williamson KE (2015) Metagenomic assessment of viral diversity in lake matoaka, a temperate, eutrophic freshwater lake in southeastern Virginia, USA. *Aquat Microb Ecol* 75(2):117–128.
38. Fancello L, et al. (2013) Viruses in the desert: a metagenomic survey of viral communities in four perennial ponds of the Mauritanian Sahara. *ISME J* 7(2):359–369.
39. Tseng C-H, et al. (2013) Microbial and viral metagenomes of a subtropical freshwater reservoir subject to climatic disturbances. *ISME J* 7(12):2374–2386.
40. Lopez-Bueno A, et al. (2009) High Diversity of the Viral Community from an Antarctic Lake. *Science (80-)* 326(5954):858–861.
41. Watkins SC, et al. (2016) Assessment of a metaviromic dataset generated from nearshore Lake Michigan. *Mar Freshw Res* 67(11):1700.
42. Skvortsov T, et al. (2016) Metagenomic characterisation of the viral community of lough neagh, the largest freshwater lake in Ireland. *PLoS One* 11(2):e0150361.
43. Arkhipova K, et al. (2018) Temporal dynamics of uncultured viruses: a new dimension in viral diversity. *ISME J* 12(1):199–211.
44. Moon K, Kang I, Kim S, Kim S-J, Cho J-C (2017) Genome characteristics and environmental distribution of the first phage that infects the LD28 clade, a freshwater methylotrophic bacterial group. *Environ Microbiol* 19(11):4714–4727.
45. Silva BS de O, et al. (2017) Virioplankton Assemblage Structure in the Lower River and Ocean Continuum of the Amazon. *mSphere* 2(5):1–17.
46. Luo E, Aylward FO, Mende DR, DeLong EF (2017) Bacteriophage Distributions and Temporal Variability in the Ocean's Interior. *MBio* 8(6):1–13.
47. Brum JR, et al. (2015) Patterns and ecological drivers of ocean viral communities. *Science (80-)* 348(6237):1261498–1261498.

48. Hurwitz BL, Sullivan MB (2013) The Pacific Ocean Virome (POV): A Marine Viral Metagenomic Dataset and Associated Protein Clusters for Quantitative Viral Ecology. *PLoS One* 8(2):e57355.
49. Roux S, et al. (2016) Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature* 537(7622):689–693.
50. Mizuno CM, Rodriguez-Valera F, Kimes NE, Ghai R (2013) Expanding the Marine Virosphere Using Metagenomics. *PLoS Genet* 9(12):e1003987.
51. Gregory AC, et al. (2019) Marine DNA Viral Macro- and Microdiversity from Pole to Pole. *Cell*:1–15.
52. Paez-Espino D, et al. (2016) Uncovering Earth's virome. *Nature* 536(7617):425–430.
53. López-Pérez M, Haro-Moreno JM, Gonzalez-Serrano R, Parras-Moltó M, Rodriguez-Valera F (2017) Genome diversity of marine phages recovered from Mediterranean metagenomes: Size matters. *PLOS Genet* 13(9):e1007018.
54. Mizuno CM, Ghai R, Saghaï A, López-García P, Rodriguez-Valera F (2016) Genomes of Abundant and Widespread Viruses from the Deep Ocean. *MBio* 7(4):e00805-16.
55. Nishimura Y, et al. (2017) Environmental Viral Genomes Shed New Light on Virus-Host Interactions in the Ocean. *mSphere* 2(2):e00359-16.
56. Philosof A, et al. (2017) Novel Abundant Oceanic Viruses of Uncultured Marine Group II Euryarchaeota. *Curr Biol* 27(9):1362–1368.
57. Holmfeldt K, et al. (2013) Twelve previously unknown phage genera are ubiquitous in global oceans. *Proc Natl Acad Sci* 110(31):12798–12803.
58. Kang I, Oh H-M, Kang D, Cho J-C (2013) Genome of a SAR116 bacteriophage shows the prevalence of this phage type in the oceans. *Proc Natl Acad Sci* 110(30):12343–12348.
59. Zhao Y, et al. (2013) Abundant SAR11 viruses in the ocean. *Nature* 494(7437):357–60.
60. Sullivan MB, Waterbury JB, Chisholm SW (2003) Cyanophages infecting the oceanic cyanobacterium Prochlorococcus. *Nature* 424(6952):1047–51.
61. Yoshida T, et al. (2008) Ma-LMM01 Infecting Toxic Microcystis aeruginosa Illuminates Diverse Cyanophage Genome Strategies. *J Bacteriol* 190(5):1762–1772.
62. Dreher TW, et al. (2011) A freshwater cyanophage whose genome indicates close relationships to photosynthetic marine cyanomyophages. *Environ Microbiol* 13(7):1858–1874.
63. Zhong KX, et al. (2018) A New Freshwater Cyanosiphovirus Harboring Integrase. *Front Microbiol* 9. doi:10.3389/fmicb.2018.02204.
64. Chénard C, Chan AM, Vincent WF, Suttle CA (2015) Polar freshwater cyanophage S-EIV1 represents a new widespread evolutionary lineage of phages. *ISME J* 9(9):2046–2058.
65. Moon K, Kang I, Kim S, Kim S, Cho J (2018) Genomic and ecological study of two distinctive freshwater bacteriophages infecting a Comamonadaceae bacterium. *Sci Rep* 8(1):7989.

66. Ghai R, Mehrshad M, Mizuno CM, Rodriguez-Valera F (2017) Metagenomic recovery of phage genomes of uncultured freshwater actinobacteria. *ISME J* 11(1):304–308.
67. Cabello-Yeves PJ, et al. (2017) Genomes of Novel Microbial Lineages Assembled from the Sub-Ice Waters of Lake Baikal. *Appl Environ Microbiol* 84(1):e02132-17.
68. Henson MW, Lanclos VC, Faircloth BC, Thrash JC (2018) Cultivation and genomics of the first freshwater SAR11 (LD12) isolate. *ISME J* 12(7):1846–1860.
69. Tsementzi D, et al. (2019) Ecogenomic characterization of widespread, closely-related SAR11 clades of the freshwater genus "Candidatus Fonsibacter" and proposal of Ca. Fonsibacter lacus sp. nov. *Syst Appl Microbiol*. doi:10.1016/j.syapm.2019.03.007.
70. Kasalický V, Jezbera J, Hahn MW, Šimek K (2013) The Diversity of the Limnohabitans Genus, an Important Group of Freshwater Bacterioplankton, by Characterization of 35 Isolated Strains. *PLoS One* 8(3):e58209.
71. Honjo M, Matsui K, Ishii N, Nakanishi M, Kawabata Z (2007) Viral abundance and its related factors in a stratified lake. *Fundam Appl Limnol / Arch für Hydrobiol* 168(1):105–112.
72. Pradeep Ram A, Nishimura Y, Tomaru Y, Nagasaki K, Nagata T (2010) Seasonal variation in viral-induced mortality of bacterioplankton in the water column of a large mesotrophic lake (Lake Biwa, Japan). *Aquat Microb Ecol* 58(3):249–259.
73. Mahmoudabadi G, Phillips R (2018) A comprehensive and quantitative exploration of thousands of viral genomes. *Elife* 7:e31955.
74. Perez Sepulveda B, et al. (2016) Marine phage genomics: The tip of the iceberg. *FEMS Microbiol Lett* 363(15):1–8.
75. Hurwitz BL, Deng L, Poulos BT, Sullivan MB (2013) Evaluation of methods to concentrate and purify ocean virus communities through comparative, replicated metagenomics. *Environ Microbiol* 15(5):1428–1440.
76. Brown CT, et al. (2015) Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature* 523(7559):208–211.
77. Hurwitz BL, Brum JR, Sullivan MB (2015) Depth-stratified functional and taxonomic niche specialization in the 'core' and 'flexible' Pacific Ocean Virome. *ISME J* 9(2):472–484.
78. Rodriguez-Brito B, et al. (2010) Viral and microbial community dynamics in four aquatic environments. *ISME J* 4(6):739–751.
79. Yoshida T, et al. (2018) Locality and diel cycling of viral production revealed by a 24 h time course cross-omics analysis in a coastal region of Japan. *ISME J* 12(5):1287–1295.
80. Weinbauer MG, Brettar I, Höfle MG (2003) Lysogeny and virus-induced mortality of bacterioplankton in surface, deep, and anoxic marine waters. *Limnol Oceanogr* 48(4):1457–1465.
81. Paul JH (2008) Prophages in marine bacteria: Dangerous molecular time bombs or the key to survival in the seas? *ISME J* 2(6):579–589.

82. Howard-Varona C, Hargreaves KR, Abedon ST, Sullivan MB (2017) Lysogeny in nature: mechanisms, impact and ecology of temperate phages. *ISME J* 11(7):1511–1520.
83. Brum JR, Hurwitz BL, Schofield O, Ducklow HW, Sullivan MB (2016) Seasonal time bombs: dominant temperate viruses affect Southern Ocean microbial dynamics. *ISME J* 10(2):437–449.
84. Winter C, et al. (2018) Mixing alters the lytic activity of viruses in the dark ocean. *Ecology* 99(3):700–713.
85. Sullivan MB, et al. (2010) Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments. *Environ Microbiol* 12(11):3035–3056.
86. López-Pérez M, Haro-Moreno JM, de la Torre JR, Rodriguez-Valera F (2018) Novel Caudovirales associated with Marine Group I Thaumarchaeota assembled from metagenomes. *Environ Microbiol*. doi:10.1111/1462-2920.14462.
87. Shelton AN, et al. (2019) Uneven distribution of cobamide biosynthesis and dependence in bacteria predicted by comparative genomics. *ISME J* 13(3):789–804.
88. Sullivan MB, Coleman ML, Weigele P, Rohwer F, Chisholm SW (2005) Three Prochlorococcus Cyanophage Genomes: Signature Features and Ecological Interpretations. *PLoS Biol* 3(5):e144.
89. Mizuno CM, Rodriguez-Valera F, Garcia-Heredia I, Martin-Cuadrado AB, Ghai R (2013) Reconstruction of novel cyanobacterial siphovirus genomes from mediterranean metagenomic fosmids. *Appl Environ Microbiol* 79(2):688–695.
90. Summer EJ, Gill JJ, Upton C, Gonzalez CF, Young R (2007) Role of phages in the pathogenesis of Burkholderia, or "Where are the toxin genes in Burkholderia phages?" *Curr Opin Microbiol* 10(4):410–417.
91. Eiler A, et al. (2016) Tuning fresh: radiation through rewiring of central metabolism in streamlined bacteria. *ISME J* 10(8):1902–1914.
92. Mehrshad M, et al. (2018) Hidden in plain sight—highly abundant and diverse planktonic freshwater Chloroflexi. *Microbiome* 6(1):176.
93. Giovannoni S, Temperton B, Zhao Y (2013) Giovannoni et al. reply. *Nature* 499(7459):E4–E5.
94. Mizuno CM, Ghai R, Rodriguez-Valera F (2014) Evidence for metaviromic islands in marine phages. *Front Microbiol* 5:1–10.
95. Rodriguez-Valera F, Mizuno CM, Ghai R (2014) Tales from a thousand and one phages. *Bacteriophage* 4(2):e28265.
96. Hiraoka S, et al. (2019) Metaepigenomic analysis reveals the unexplored diversity of DNA methylation in an environmental prokaryotic community. *Nat Commun* 10(1):159.
97. Martinez-Hernandez F, et al. (2017) Single-virus genomics reveals hidden cosmopolitan and abundant viruses. *Nat Commun* 8:15892.
98. Roux S, Emerson JB, Eloe-Fadrosh EA, Sullivan MB (2017) Benchmarking viromics: an in silico

evaluation of metagenome-enabled estimates of viral community composition and diversity. *PeerJ* 5:e3817.

99. Warwick-Dugdale J, et al. (2019) Long-read viral metagenomics captures abundant and microdiverse viral populations and their niche-defining genomic islands. *PeerJ* 7:e6800.

100. Salcher MM, Pernthaler J, Posch T (2011) Seasonal bloom dynamics and ecophysiology of the freshwater sister clade of SAR11 bacteria "that rule the waves" (LD12). *ISME J* 5(8):1242–1252.

101. Zhao Y, et al. (2018) Pelagiphages in the Podoviridae family integrate into host genomes. *Environ Microbiol*. doi:10.1111/1462-2920.14487.

102. Martinez-Hernandez F, et al. (2019) Single-cell genomics uncover Pelagibacter as the putative host of the extremely abundant uncultured 37-F6 viral population in the ocean. *ISME J* 13(1):232–236.

103. Zaremba-Niedzwiedzka K, et al. (2013) Single-cell genomics reveal low recombination frequencies in freshwater bacteria of the SAR11 clade. *Genome Biol* 14(11):R130.

104. Mizuno CM, et al. (2019) Numerous cultivated and uncultivated viruses encode ribosomal proteins. *Nat Commun* 10(1):752.

105. Labonté JM, et al. (2015) Single-cell genomics-based analysis of virus–host interactions in marine surface bacterioplankton. *ISME J* 9(11):2386–2399.

106. Chow C-ET, Winget DM, White RA, Hallam SJ, Suttle CA (2015) Combining genomic sequencing methods to explore viral diversity and reveal potential virus-host interactions. *Front Microbiol* 6:265.

107. Ahlgren NA, Fuchsman CA, Rocap G, Fuhrman JA (2019) Discovery of several novel, widespread, and ecologically distinct marine Thaumarchaeota viruses that encode amoC nitrification genes. *ISME J* 13(3):618–631.

108. Zatopek KM, Gardner AF, Kelman Z (2018) Archaeal DNA replication and repair: new genetic, biophysical and molecular tools for discovering and characterizing enzymes, pathways and mechanisms. *FEMS Microbiol Rev* 42(4):477–488.

109. Krupovic M, Cvirkaite-Krupovic V, Iranzo J, Prangishvili D, Koonin E V. (2018) Viruses of archaea: Structural, functional, environmental and evolutionary genomics. *Virus Res* 244(July 2017):181–193.

110. Roux S, et al. (2017) Ecogenomics of virophages and their giant virus hosts assessed through time series metagenomics. *Nat Commun* 8(1):858.

111. Kang I, Kim S, Islam MR, Cho J-C (2017) The first complete genome sequences of the acI lineage, the most abundant freshwater Actinobacteria, obtained by whole-genome-amplification of dilution-to-extinction cultures. *Sci Rep* 7(1):42252.

112. Hahn MW, Jezberová J, Koll U, Saueressig-Beck T, Schmidt J (2016) Complete ecological isolation and cryptic diversity in Polynucleobacter bacteria not resolved by 16S rRNA gene

sequences. *ISME J* 10(7):1642–1655.

113. Salcher MM, Neuenschwander SM, Posch T, Pernthaler J (2015) The ecology of pelagic freshwater methylotrophs assessed by a high-resolution monitoring and isolation campaign. *ISME J* 9(11):2442–2453.

114. Eiler A, Bertilsson S (2007) Flavobacteria Blooms in Four Eutrophic Lakes: Linking Population Dynamics of Freshwater Bacterioplankton to Resource Availability. *Appl Environ Microbiol* 73(11):3511–3518.

115. Paver SF, Muratore D, Newton RJ, Coleman ML (2018) Reevaluating the Salty Divide: Phylogenetic Specificity of Transitions between Marine and Freshwater Systems. *mSystems* 3(6):347021.

116. Logares R, Brate J, Heinrich F, Shalchian-Tabrizi K, Bertilsson S (2010) Infrequent Transitions between Saline and Fresh Waters in One of the Most Abundant Microbial Lineages (SAR11). *Mol Biol Evol* 27(2):347–357.

117. Logares R, et al. (2009) Infrequent marine-freshwater transitions in the microbial world. *Trends Microbiol* 17(9):414–422.

118. John SG, et al. (2011) A simple and efficient method for concentration of ocean viruses by chemical flocculation. *Environ Microbiol Rep* 3(2):195–202.

119. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA (2017) metaSPAdes: a new versatile metagenomic assembler. *Genome Res* 27(5):824–834.

120. Bankevich A, et al. (2012) SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J Comput Biol* 19(5):455–477.

121. Mihara T, et al. (2016) Linking Virus Genomes with Host Taxonomy. *Viruses* 8(3):66.

122. Camacho C, et al. (2009) BLAST plus: architecture and applications. *BMC Bioinformatics* 10(421):1.

123. Roux S, Enault F, Hurwitz BL, Sullivan MB (2015) VirSorter: mining viral signal from microbial genomic data. *PeerJ* 3:e985.

124. Hyatt D, et al. (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11(1):119.

125. Shen W, Le S, Li Y, Hu F (2016) SeqKit: A cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS One* 11(10):1–10.

126. Nishimura Y, et al. (2017) ViPTree: the viral proteomic tree server. *Bioinformatics* 33(15):2379–2380.

127. Revell LJ (2012) phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol Evol* 3(2):217–223.

128. Gu Z, Gu L, Eils R, Schlesner M, Brors B (2014) circlize implements and enhances circular visualization in R. *Bioinformatics* 30(19):2811–2812.

129. Rodriguez-R LM, Konstantinidis KT (2016) The enveomics collection : a toolbox for specialized analyses of microbial genomes and metagenomes. *Peer J Prepr* 4:e1900v1.

130. Kang DD, Froula J, Egan R, Wang Z (2015) MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* 3:e1165.

131. Segata N, Börnigen D, Morgan XC, Huttenhower C (2013) PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nat Commun* 4:2304.

132. Parks DH, et al. (2018) A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol*. doi:10.1038/nbt.4229.

Figures and legends

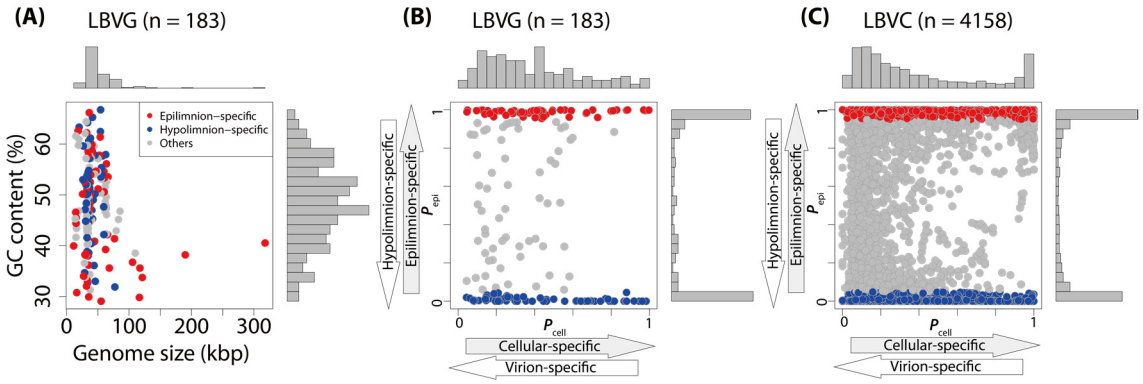

Figure 1.

Profiles of assembled viral genomes and contigs. (A) Distribution of genome size and GC content (%) of Lake Biwa Viral Genomes (LBVGs). Habitat preferences of (B) LBVGs and (C) Lake Biwa Viral Contigs (LBVCs). The vertical and horizontal axes indicate the average P_{epi} and P_{cell} for each LBVG, respectively (see Materials and Methods for calculation details). Viruses with average P_{epi} >95% and <5% were designated as epilimnion- and hypolimnion-specific viruses and are indicated with red and blue points, respectively.

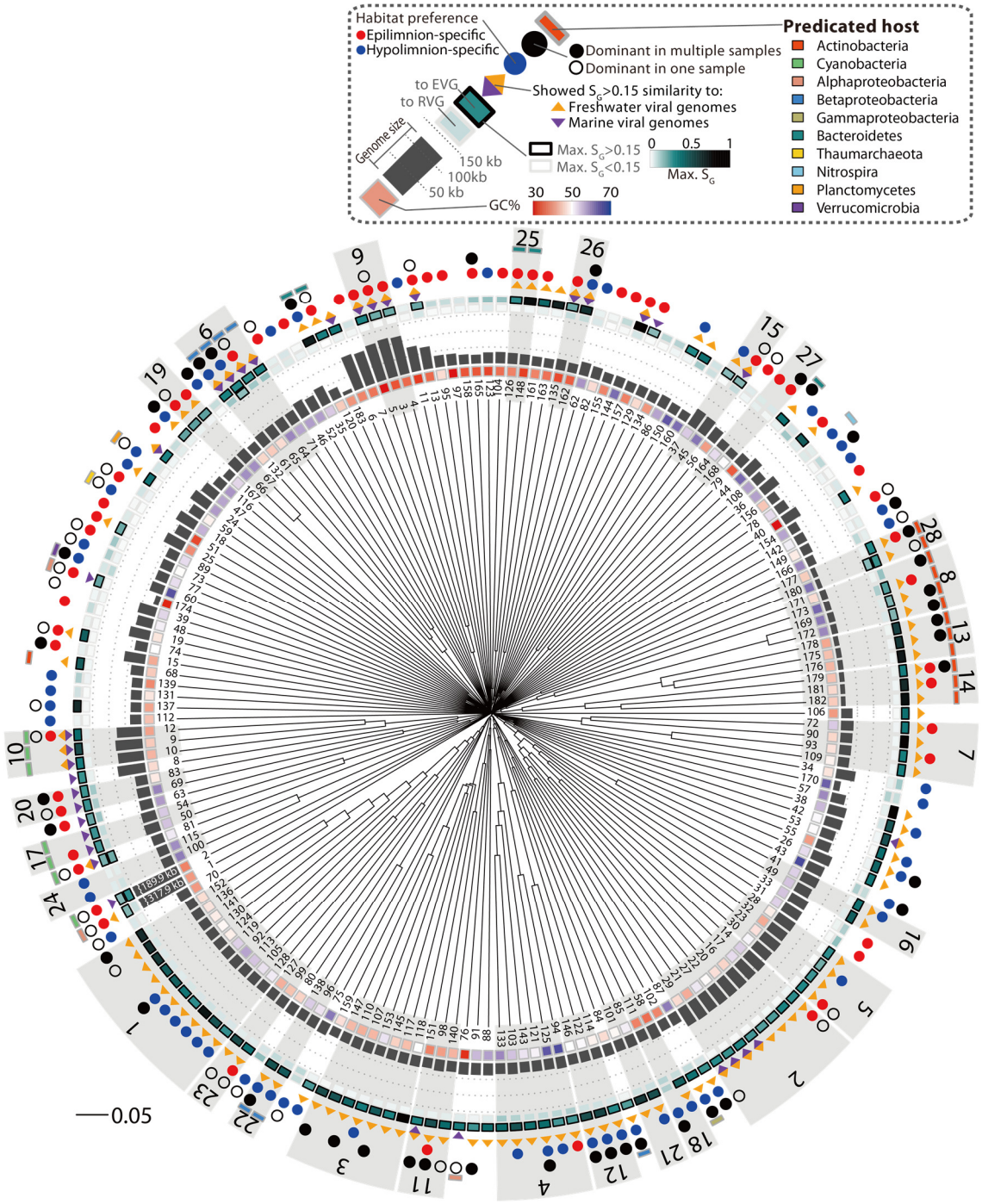

Figure 2.

Proteomic tree of 183 LBVGs with summarized analytical results. The gray shaded boxes indicate individual gOTUs with corresponding gOTU numbers. Annotation of epilimnion- and hypolimnion-specific viruses corresponds with the definition in Figure 1. "Dominant" viruses indicated by the black and white circles are "ranked among the 10 most abundant viruses," as described in Figure 4.

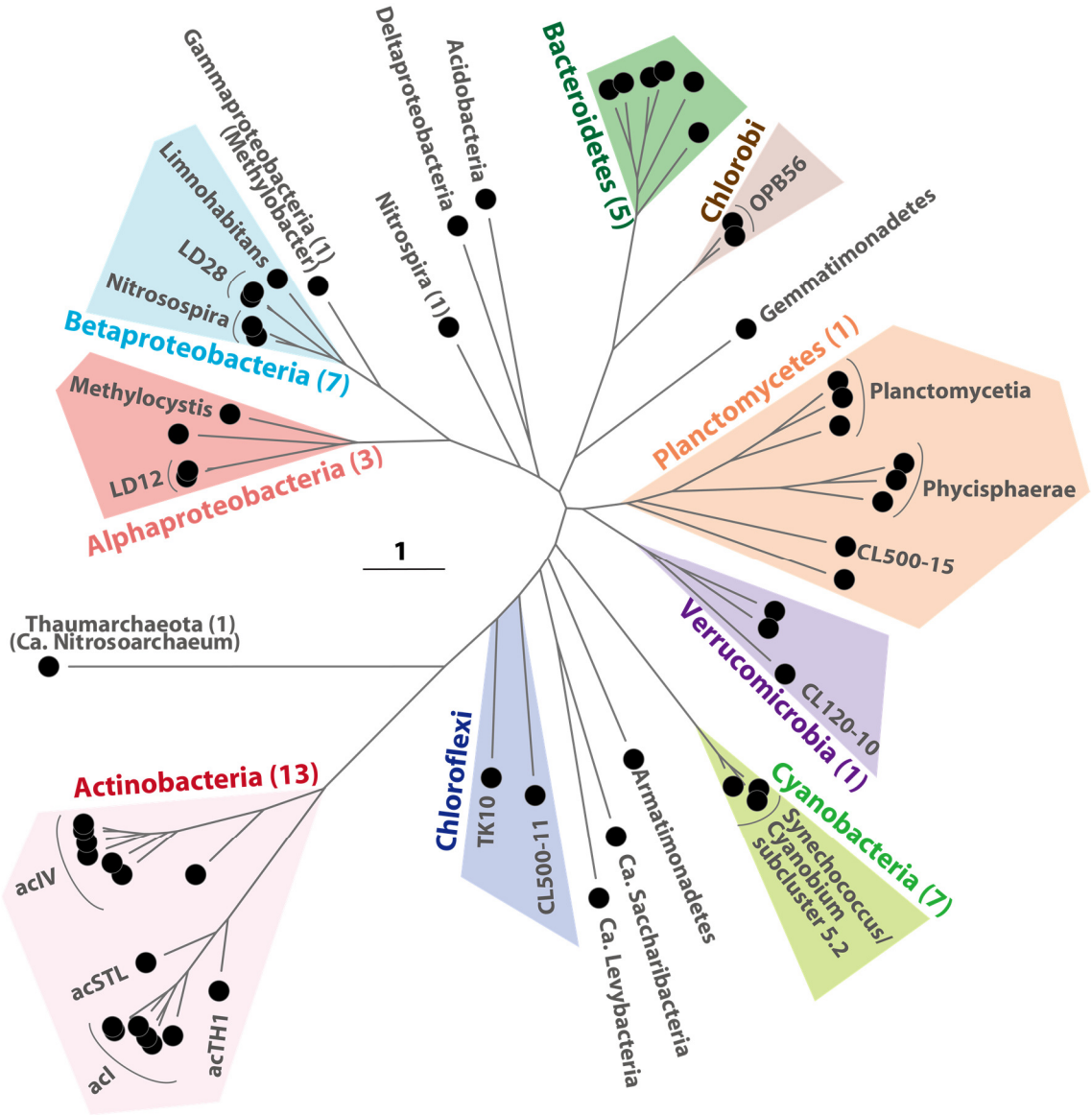

Figure 3.

A genome-resolved phylogenetic tree of Lake Biwa metagenome-assembled genomes (LBMAGs) based on conserved single-copy genes selected using PhyloPhlAn software. Members of the same phylum (or Proteobacteria class) were grouped by the same color shade. Numbers in parentheses indicate the number of LBVGs predicted to infect each taxonomic group. Note that Chloroflexi did not form a monophyletic clade here but did in a tree that included reference genomes (Fig. S3).

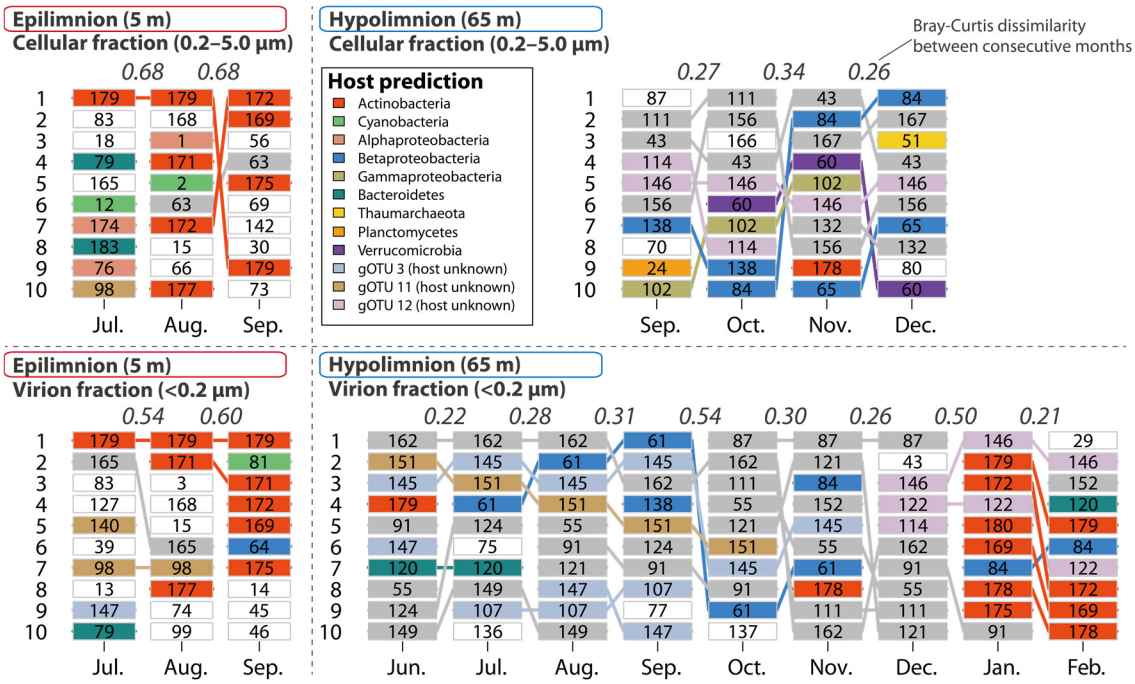

Figure 4.

Succession of the 10 most abundant viral genomes in each sample fraction. The number in each box corresponds to the number of LBVGs. The color of each box indicates the phylum of the predicted host. The abundant gOTUs (gOTU_3, gOTU_11, and gOTU_12) are also shaded, although their hosts remain unknown. Otherwise, members that were dominant over multiple months are in gray boxes and those that were dominant only within one month are in white boxes. Italicized numbers at the top of each panel indicate the extent of the community composition shift in consecutive months, based on the Bray-Curtis dissimilarity of the whole community. Quantitative data for this figure are available in Figure S4.

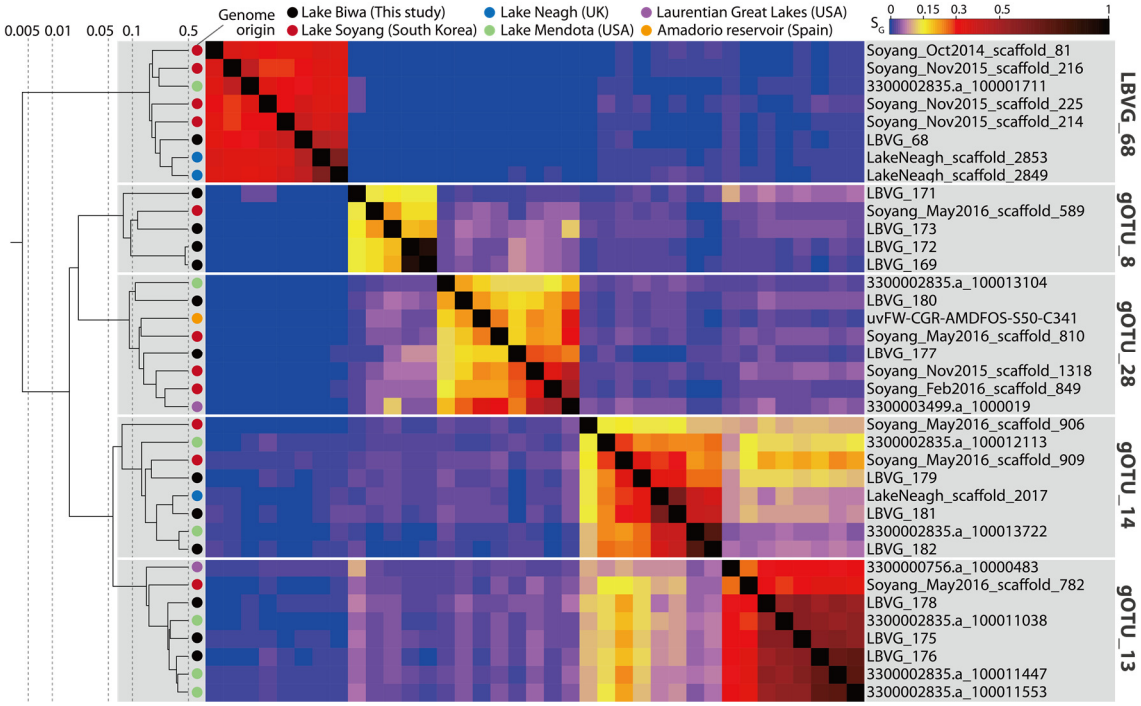

Figure 5.

Proteomic tree and SG-based pairwise distance matrix among five actinoviral groups identified in the present study and their close relatives in the Environmental Viral Genome (EVG) database. Alignments between individual genomes can be found in SI Alignment S1.