Cataloging Viral Diversity from Nonaxenic Terrestrial Cyanobacteria Cultures

Cassandra L. Ettinger[a,b#], Sudipto R. Paul[c], Neptali Flores[d], Ryan D. Ward[e], Nicole Pietrasiak[f], Jason E. Stajich[a,b,#]

[a]Department of Microbiology and Plant Pathology, University of California, Riverside, Riverside, California, USA.

[b]Institute for Integrative Genome Biology, University of California, Riverside, Riverside, CA, USA.

[c]Department of Biochemistry, University of California, Riverside, Riverside, California, USA.

[d]Department of Biological Sciences, San Jose State University, San Jose, California, USA.

[e]Laboratory of Genetics, University of Wisconsin, Madison, Madison, Wisconsin, USA.

[f]School of Life Sciences, University of Nevada, Las Vegas, Las Vegas, Nevada, USA.

Running Title: Terrestrial Cyanophage Survey

[#]**Address correspondence to**: Cassandra L. Ettinger, cassande@ucr.edu; Jason E. Stajich, jason.stajich@ucr.edu

Keywords: cyanophage, virus, phage, viral diversity, viral host, terrestrial, Cyanobacteria, desert ecosystem

Abstract

Viruses are exceedingly common, but little is known about their diversity let alone how they behave in extreme environments, and whether viruses facilitate adaptation of their hosts to harsh conditions. To set a foundation for understanding of these understudied viral-host interactions, we created a catalog of viruses through analysis of metagenomes from 50 unialgal but nonaxenic Cyanobacteria cultures with 47 cultures isolated from various terrestrial habitats, including desert soil and rock surfaces, tropical soil, and vernal pools. These cultures represent low diversity microbial consortia dominated by the terrestrial Cyanobacteria and its associated cyanosphere microbiome containing heterotrophic microbes. We identified viral sequences in metagenomes, grouped these into viral operational taxonomic units (vOTUs) and then placed vOTUs into viral clusters (VCs). We also calculated vOTU relative abundance and predicted possible bacterial hosts. In total we predicted 814 viral sequences representing 726 vOTUs. We assigned putative taxonomy to 72 of the 814 putative viral sequences; these were distributed into 15 VCs — mostly assigned to the recently abolished Caudovirales order (now Caudoviricetes class) of viruses. We found that nonaxenic cultures were dominated by unclassified and unclustered viral sequences. Furthermore, we predicted putative bacterial hosts for 211 vOTUs, with the majority of viruses predicted to infect a Proteobacteria (now Pseudomonadota) host. Overall, while limited, these results are consistent with the notion that both viruses and Cyanobacteria isolated from extreme environments are underrepresented in reference datasets. This work increases knowledge of viral diversity and sets a foundation for future exploration of viruses associated with terrestrial Cyanobacteria and their heterotroph associates, such as connecting specific viruses to critical cycling processes and investigating their metabolic functions.

Introduction

Viruses infect all domains of life and are thought to be the most abundant biological entities on Earth (Paez-Espino et al., 2016). Viruses rely on their hosts for replication, and thus are intricately entwined with their host communities. In the lytic life cycle viruses infect their hosts, hijack the cellular machinery to replicate, and induce host cell lysis for dispersal of the newly created viral particles (Guttman, Raya & Kutter, 2004). As a host defense mechanism, phage infections can be evidenced in host's genomes resulting from past infection events building up hosts' phage immunity (Barrangou et al., 2007; Hampton, Watson & Fineran, 2020). Frequently viruses can also integrate and persist in their host's genome as proviruses, with many bacteria harboring at least one prophage (Casjens, 2003). Prophages can have auxiliary metabolic genes (AMGs) or pathogenicity genes which they use to promote host growth (Bragg & Chisholm, 2008; Thompson et al., 2011; Fortier & Sekulovic, 2013; Hurwitz, Hallam & Sullivan, 2013). With such controls on host metabolism and fitness, viruses can have important and often underappreciated effects on bacterial population dynamics and turnover in microbial communities (Hurwitz et al., 2014; Hurwitz & U'Ren, 2016; Crummett et al., 2016).

Viral discovery has long been hindered by the lack of universal gene markers. However, modern machine-learning based and k-mer approaches have enabled researchers to robustly search metagenomes for viral sequences (Sullivan, 2015; Ren et al., 2017; Amgarten et al., 2018; Kieft, Zhou & Anantharaman, 2020; Guo et al., 2021). With these new methodologies, researchers can utilize available public sequencing data to discover and explore complex viral communities across previously uncharted ecosystems. However, virus discovery and potential host interactions and feedbacks to microbial communities in extreme environments, such as cave wall

biofilms, biofilms of ephemeral waterfalls, temporal vernal pools, and dryland soils, are still notably understudied.

Cyanobacteria are keystone taxa of such polyextreme terrestrial environments due to their important roles in primary production and microbial organic matter build up, nitrogen fixation, bioweathering, and biofilm, microbial mat or biological soil crust formation (Paerl, Pinckney & Steppe, 2000; Perera et al., 2018; Pombubpa et al., 2020; Velichko et al., 2021). As essential ecosystem engineers, Cyanobacteria interact with a broad variety of other microbes. For examples, filamentous biological soil crust cyanobacteria house their own set of associated microbes established around the polysaccharide rich sheath material forming the cyanosphere, a microbial community similar to the copiotrophic rhizosphere microbiomes around plant roots (Couradeau et al., 2019). Cyanobacteria also influence free-living co-occurring microbes as shown in another study in which cyanobacteria in hypersaline microbial mats generated fermentation byproducts that promoted the metabolism of Chloroflexi and Sulfate-reducing bacteria (Lee et al., 2014). However, our knowledge on Cyanobacteria phages, viruses that infect Cyanobacteria, is still largely based on prior investigations of samples collected from aquatic ecosystems (Patzelt et al., 2014; Pietrasiak et al., 2019, 2021; Ward et al., 2021; Baldarelli et al., 2022). Cyanophages in aquatic ecosystems have shown to carry AMGs related to photosynthesis, phosphorus utilization, and carbon metabolism which are thought to aid phage fitness (Bailey et al., 2004; Sullivan et al., 2005, 2010; Millard et al., 2009). Yet, if Cyanophages have similar dynamic roles in polyextreme terrestrial habitats, or how phages might impact Cyanobacteria associated microbiomes is still virtually unknown.

Previously, Ward et al. (2021) generated 50 metagenomes from nonaxenic Cyanobacterial cultures, which represented unialgal polycultures containing primarily Cyanobacteria and heterotrophic microbial associates living in the cyanosphere. The majority of the polycultures (47/50) were obtained from Cyanobacteria isolated from terrestrial habitats, with over half representing cultures from arid environments (28/50) (Ward et al., 2021). Due to the paucity of information on phage and associated host identity for Cyanobacteria-dominated microbial communities of polyextreme environments, investigating low diversity communities such as cyanosphere microbiomes of cultured Cyanobacteria hosts is a tangible starting point of discovery. In this study, we surveyed the viral diversity associated with these Cyanobacterial co-cultures as part of a remote undergraduate research experience. We sought to build foundational knowledge about these viruses by (i) creating a catalog of viral genomes present in these co-cultures, (ii) assessing the relative abundance of viral genomes across co-cultures, and (iii) beginning to explore possible bacterial hosts for these viruses.

Methods

Data collection

Publicly available metagenomic samples, obtained from whole community DNA of Cyanobacteria-dominant nonanexic unialgal cultures and previously reported on in Ward et al (2021), were downloaded from NCBI GenBank and imported to the KBase platform by remote undergraduate researchers using their respective SRA numbers (Table S1) (Arkin et al., 2018). Sample metadata associated with each metagenome was obtained from Table 1 of Ward et al (2021). Briefly, these metagenomes represent unialgal polycultures containing primarily Cyanobacteria and their cyanosphere heterotrophic microbial associates collected from three

aquatic and 47 terrestrial habitats across four broad biomes (arid: 28, tropical: 16, continental: 3, temperate: 3). The dominant Cyanobacterial members in these cultures are phylogenetically diverse spanning five taxonomic orders (Nostocales: 21, Pseudanabaenales: 12, Oscillatoriales: 7, Chroococcales: 5, Synechococcales: 3, Pleurocapsales: 2). These Cyanobacteria-dominant co-cultures were isolated by various phycologists between 1965-2015 and have been maintained in two culture collections at John Carroll University (JCU) and University of Nevada - Las Vegas (UNLV) on solid or liquid Z8 media (an oligotrophic freshwater algae medium (Carmichael, 1986)). Although dates of isolation span 50 years, close to identical culture conditions (use of Z8 medium, 16/8 hr light/dark photoperiod, low light intensity, grown in climate controlled chambers) have been maintained at JCU since 1992 which have been reproduced at UNLV. Collection years can be found in Supplemental Table S1 and Ward et al. (2021).

Identification of viral contigs

Initial viral analysis was performed following the KBase "Viral Analysis End-to-End" workflow (https://kbase.us/n/75811/85/) by remote undergraduate researchers during the COVID-19 pandemic. Sequence quality for each sample was assessed in KBase using FastQC v0.11.5 (Andrews). Afterwards, low-quality regions and residual Illumina adapters were trimmed using Trimmomatic v0.36 (Bolger, Lohse & Usadel, 2014). This step was followed by a reassessment of the quality of the sequence data again using FastQC. Next, the reads were assembled into contigs with MetaSPAdes v3.13.0 with k-mer sizes 21, 33, 55, 77, 99 and 127 (Nurk et al., 2017). Three samples failed to assemble at this stage using MetaSPAdes due to memory constraints in KBase; for these three samples, MEGAHIT v1.2.9 was utilized instead (Li et al.,

2015). Subsequently, all contigs shorter than 5000 base pairs were filtered out using the KBase application “Filter Assembled Contigs By Length - v1.1.2”.

Next, VirSorter v1.0.5 was used to predict the viral genomes in each sample and categorize them into six categories based on their predicted length and viral characteristics (Roux et al., 2015). Viral contigs from VirSorter categories 1-2 (high-confidence lytic viruses) and 4-5 (high-confidence proviruses) were merged into one file using the KBase application “Modify Bins in BinnedContigs - v1.0.2” and then extracted using the “Extract Bins as Assemblies from BinnedContigs - v1.0.2” application. The viral contigs for each sample were then exported from KBase. All predicted viral contigs were combined into a single file and then run through VIBRANT v1.2.1 in “virome” mode (Kieft, Zhou & Anantharaman, 2020). This second viral identification step with VIBRANT was performed to ensure proper trimming of proviruses. CheckV v0.8.1 was then run on the resulting file of predicted viral genomes using the “end_to_end” workflow to calculate viral genome quality, completeness and further predict and remove flanking host regions from integrated proviruses (Nayfach et al., 2021). When performing viral operational taxonomic units (vOTU) analyses, we removed any vOTUs that CheckV could not make quality predictions for (i.e. where quality was “not-determined”).

Clustering of viral contigs across all samples & relative abundance

Viral genomes were then clustered into 95% identity vOTUs using CD-HIT using the following options: -c 0.95 -aS 0.85 -M 0 -d 0 (Li & Godzik, 2006). In order to assign a putative taxonomy to vOTUs, we first used Prodigal v. 2.6.3 to predict open reading frames in vOTU representative genomes using the -p meta option (Hyatt et al., 2010). We then used VContact2 to cluster

vOTUs with the INPHARED August 2022 reference database into viral clusters (VCs) (Bin Jang et al., 2019; Cook et al., 2021) and assigned predicted taxonomy to viruses based on cluster membership as in Santos-Medellin et al (Santos-Medellin et al., 2021). For vOTUs in VCs of interest, ViPTree was further used to place vOTUs into proteomic trees with viral reference genomes using ViPTree database v. 3.6 with virus-host information from RefSeq release 218 (Nishimura et al., 2017).

To quantify the relative abundance of vOTUs across samples, we mapped reads from each metagenome to vOTUs using BBMap with a minid=0.90 (Bushnell, 2022). We used SAMtools to convert sam files to bam files and BEDTools genomecov to get coverage information for each vOTU across each sample from the bam files (Li et al., 2009; Quinlan & Hall, 2010). We then used bamM to parse the bam files to calculate the trimmed pileup coverage (tpmean), the average number of reads aligned to each base after removing the highest 10% and lowest 10% of positions, which we used as a proxy for viral relative abundance ("Ecogenomics/BamM: Metagenomics-focused BAM file manipulation"). We then filtered out vOTUs which displayed < 75% coverage over the length of the viral sequence in R v. 4.3.0 (R Core Team, 2021). Thresholds for analysis of relative abundance of vOTUs were based on previous community guidelines for identity (i.e. ≥ 95% identity) and detection (i.e. ≥ 75% of the viral genome length covered ≥ 1x by reads at ≥ 90% average nucleotide identity) (Roux et al., 2017, 2018). Viral relative abundance was visualized in R using tidyverse, phyloseq, GGally, network, patchwork, ggnewscale, ggforce and ggtext (Butts, C, 2008; McMurdie & Holmes, 2013; Wickham et al., 2019; Schloerke et al., 2021; Pedersen, 2022a,b; Wilke & Wiernik, 2022; Campitelli, 2023). Analysis scripts for the post-KBase portions of this work are available on GitHub

(https://github.com/stajichlab/Terrestrial_Cyanophage) and archived in Zenodo (Ettinger, Pietrasiak & Stajich).

Putative identification of hosts

In order to predict possible microbial hosts for the viral genomes identified, we used NCBI BLAST which has been previously reported to be sufficient for preliminary genus-level host identification (Edwards et al., 2016; Kothari et al., 2021). We downloaded all complete bacterial and archaeal genomes, as of April 7, 2022, from NCBI RefSeq using "ncbi-genome-download" ("kblin/ncbi-genome-download: Scripts to download genomes from the NCBI FTP servers"). In total this amounted to 25,984 bacterial and 416 archaeal genomes. We then used "makeblastdb" from NCBI BLAST 2.12.0+ to convert these complete genomes into a blast database and "blastn" to compare vOTUs to this reference (Camacho et al., 2009). We filtered the results of this analysis in R based on accepted thresholds from the literature as follows: matches had to share ≥ 2000 bp region with ≥ 70% sequence identity to the viral genome and possible host matches needed to have a bit score of ≥ 50 and minimum e value of 0.001 (Edwards et al., 2016; Roux et al., 2016; Kothari et al., 2021; Ettinger et al., 2023). As was done in Kothari et al. (2021), only the top five hits matching these thresholds were considered when predicting a host for each viral genome, with host predictions made at each taxonomic level only if all hits were in agreement. Discrepancies in taxonomy between hits resulted in a "mixed" assignment for that taxonomic level. Based on this method, the lowest level of host prediction was genus level, with some predictions only able to be made at higher taxonomic levels (e.g. class or order). For viruses with a predicted Cyanobacterial host and those that were in VCs with reference genomes reported to infect Cyanobacteria, Cenote-Taker 2 v. 2.1.5 (Tisza et al., 2021) was used to

annotate these genomes and Proksee (https://proksee.ca/) was used to visualize the resulting annotations (Stothard, Grant & Van Domselaar, 2019). We then used the comparative gene cluster analysis toolbox (CAGECAT) (van den Belt et al., 2023) to run clinker (Gilchrist & Chooi, 2021) to visualize homologous gene clusters between viral genomes and cyanophage reference genomes downloaded from NCBI GenBank (OM373202, BK059895, NC_015463, NC_029032, NC_016564) (Gao, Gui & Zhang, 2012; Huang et al., 2012; Voorhies et al., 2016; Laloum et al., 2022).

Results and Discussion

Successful viral identification from Cyanobacterial co-cultures

In total, 814 viral sequences were identified across the 50 Cyanobacterial co-cultures (Table S1, S2), with each metagenome having on average ~16 viral sequences (range: 1-54). Of the 814 viral sequences, CheckV predicted that 188 were integrated proviruses (Fig. 1A). CheckV assigned viral quality to 712 of the 814 viral genomes with 465 predicted to be low-quality, 208 to be medium-quality, 32 to be high-quality, and 7 to be complete (Fig. 1B). Based on MIUViG standards (Roux et al., 2018) the catalog represents 39 high-quality draft genomes (i.e., CheckV high-quality and complete genomes) and 775 genome fragments. The mean length of predicted viral sequences was 27,594 bp (range: 5,013–199,914 bp), with 652 viral sequences ≥ 10,000 bp, and with an average of ~35 predicted genes per viral sequence. We clustered these 814 viral sequences at 95% identity into 726 vOTUs, of which 585 vOTUs represented sequences ≥ 10,000 bp in length. These numbers are similar to previous metagenomic viral discovery efforts in cyanolichen metagenomes (Ponsero et al., 2021). While this work provides a catalog of the dominant DNA viruses that infect terrestrial Cyanobacteria and associated cyanosphere

heterotrophs, we explored relatively low diversity microbiomes which were maintained in lab culture conditions. It is known that detection of viruses from such low diversity metagenomes only captures a small amount of DNA virus diversity compared to viromes (Santos-Medellin et al., 2021), and we may have missed RNA viruses entirely which are unlikely to appear in DNA-based metagenomes. Further, VirSorter1, which we use for viral identification, was designed to detect DNA phage in the former order Caudovirales (now Caudoviricetes class), and thus this catalog may be biased towards detection of these phages as well (Roux et al., 2015). Finally, given that this catalog represents viruses identified from co-cultures of Cyanobacterial isolates with their associated cyanosphere microbiome, and not environmental samples directly, we are likely sampling a small subset of the microbiomes of the polyextreme terrestrial environments they originated from. Additionally, any reduction in viruses during passaging may lead to further underestimation compared to the original viral diversity occurring in these environments.

Clustering of viral sequences with INPHARED reference data reveals possibly novel genera

To be able to assign putative taxonomy to vOTUs, we further used gene-sharing networks to cluster them with publicly available reference data to form VCs which approximate genus-level groupings. In total 305 vOTUs, representing 341 viral sequences, were placed in 111 VCs. Of these 111 VCs, 26 included representatives from the INPHARED reference dataset, while 79 represented putatively novel genera. Of the 26 VCs that included members that clustered with the reference dataset, 9 VCs included members that were identified as possible proviruses. One VC (VC 1442_0) included a reference viral genome for Cr□LKS3, a novel lambda phage in the former Siphoviridae (now unclassified), which was previously reported to infect the aquatic Cyanobacterial genus Cylindrospermopsis (Laloum et al., 2022). The remaining VCs included

references viruses reported to infect members of the phyla Proteobacteria (Pseudomonadota; 17 VCs), Spirochaetota (1 VC), Bacteroidota (1 VC), Bacillota (1 VC), Actinomycetota (1 VC), Cyanobacteria (1 VC), or were unreported (4 VCs). The majority of the 79 unique VCs originated from Cyanobacterial co-cultures isolated from arid environments, which is not unexpected as the majority of co-cultures were isolated from arid environments (56% of co-cultures) (Fig. 2A-B). This limited overlap with existing reference data is similar to recent work on cyanolichen metagenomes, where only 28 of 133 VCs included reference genomes (Ponsero et al., 2021). This is different from some metagenomic reports of phage from Cyanobacteria-dominated aquatic systems, including hot springs and lakes, where the majority of reported phage sequences were related to reference data (Skvortsov et al., 2016; Guajardo-Leiva et al., 2018a; Potapov et al., 2019). However, other efforts in freshwater and marine environments have reported a more similar percentage of unclassified viruses compared to reference databases (Sieradzki et al., 2019; Kavagutti et al., 2019). As viral exploration of Cyanobacteria-dominated and other extreme environments continues, we expect to see more sequences clustering with reference genomes over time.

The 26 VCs that clustered with the reference data represented 75 of the 814 predicted viral sequences in this dataset. Following new taxonomic recommendations, many viral reference genomes have not yet been reclassified resulting in the majority of the 26 VCs being unclassified under the new schema (Turner, Kropinski & Adriaenssens, 2021). Under the old schema, the majority of VCs (15) were clustered with members assigned to the Caudovirales order (now Caudoviricetes class) of tailed double-stranded DNA viruses, which have also been observed to be the dominant viruses in cyanolichens, biocrusts, hypoliths, and aquatic Cyanobacteria-

dominated metagenomes (Sabehi et al., 2012; Zablocki et al., 2014; Adriaenssens et al., 2015; Zablocki, Adriaenssens & Cowan, 2016; Skvortsov et al., 2016; Zeigler Allen et al., 2017; Guajardo-Leiva et al., 2018b; Van Goethem et al., 2019; Ponsero et al., 2021). Within this, VCs were further classified as belonging to the still accepted families Autographiviridae, Inoviridae, and Microviridae, as well as to the recently abolished families Myoviridae, Podoviridae, and Siphoviridae (now unclassified) (Fig. 2C-D). The former Myoviridae, Podoviridae and Siphoviridae represent the main viral families that have historically included cyanophages isolated from marine systems (Mann et al., 2005; Sullivan et al., 2005, 2009, 2010; Weigele et al., 2007) However, we cannot assume all members of VCs classified as belonging to these historical families here are cyanophage as these families also are known to contain members that do not infect Cyanobacteria.

Only 15 of the 39 sequences representing high-quality or complete viral genomes, which were made up of 14 vOTUs, came together to form 13 VCs. Of these, only five VCs included representative viruses from the INPHARED reference data, while 8 VCs were potentially novel genera. Of the VCs that clustered with reference genomes, one was clustered with the order Petitvirales, one with the order Tubulavirales, and three with members of the former order Caudovirales (now Caudoviricetes class). Additionally, VCs included references viruses reported to infect members of the phyla Proteobacteria (Pseudomonadota; 4 VCs), or were unreported (1 VC).

Most abundant vOTUs are unclassified or unclustered

While the number of viral sequences recovered from each Cyanobacterial co-culture varied (Fig. 3A), we were also curious about the distribution of viral sequences across co-cultures. Thus, we mapped metagenomic reads from each sample to the entire catalog of vOTUs to calculate their relative abundance across all co-cultures. Of 726 vOTUs, only 617 vOTUs had quality scores from CheckV and were detected above thresholds set for inclusion. Looking at the relative abundance of these remaining 617 vOTUs, we found that the majority of co-cultures were dominated by unclustered vOTUs or by vOTUs that formed VCs that did not cluster with INPHARED reference genomes (Fig. 3B). Only a small portion, on average 8.2% (range: 0 – 100%), of vOTU relative abundance in samples was from vOTUs in VCs that included reference genomes. The majority of vOTUs, 496 (80.4%), were singletons and only found in one sample. While a total of 121 vOTUs and 86 VCs (13 of which clustered with reference data) were detected in two or more samples. The most prevalent VC (714_0) did not cluster with INPHARED reference genomes and represented 25 vOTUS across 26 samples spanning different habitats and collection locations. Using ViPTree, we found that vOTUs in this VC form a unique monophyletic clade within the recently abolished family, Siphoviridae (now unclassified) (Fig. S1). Overall these results demonstrate how few viruses are present in reference databases relative to existing viral diversity in Cyanobacteria-dominant co-cultures and highlight the importance of viral discovery efforts in addressing this knowledge gap. Low diversity co-cultures from public or research culture collections, such as those investigated in our study, could function as manageable study systems to establish viral reference benchmarks in future comparative works.

Proteobacteria (Pseudomonadota), not Cyanobacteria, dominate host predictions

We explored possible host associations by using NCBI BLAST to compare vOTUs against a collection of complete bacteria and archaea genomes from RefSeq. This method has been reported to have high accuracy, but low recall as it is limited by existing databases (Coclet & Roux, 2021). The majority of viral sequences (65.8%) had no host prediction. However, we were able to predict possible bacterial hosts for 211 out of 617 vOTUs to at least the phylum-level (Fig. 3C). Of these 211 predictions, we were able to classify the hosts for 151 vOTUs to genus-level with the most common predicted host genera being *Brevundimonas* (15 vOTUs), *Sphingomonas* (13), *Pseudomonas* (11), *Bradyrhizobium* (11) and *Bosea* (10). Many of these genera have been reported in association with Cyanobacteria from both aquatic and terrestrial habitats (Bruno et al., 2006; Shi et al., 2009; Li et al., 2012; Kim et al., 2019). For those viruses with a host prediction, zooming out these were largely projected to infect Proteobacterial (now Pseudomonadotal) hosts (95.3%). Within the Proteobacteria (Pseudomonadota), almost half (47.3%) of the predicted hosts were members of the order Hyphomicrobiales. Further, 18 of the 25 vOTUs represented in the most shared VC (714_0) were predicted to infect members of the order Hyphomicrobiales.

Only three vOTUs (vOTU602, vOTU638, vOTU669), representing three unclassified viral sequences, had Cyanobacterial host predictions, and none of these belonged to the single VC (VC 1442_0) that included reference virus members reported to infect Cyanobacteria. Of these three vOTUS, two were predicted to be medium-quality (vOTU602, vOTU669) and one complete (vOTU638). Two of the three vOTUs formed separate VCs unique to this data (vOTU602 = VC 1342_0; vOTU638 = VC 1447_0). All three vOTUs were predicted to have hosts in the order Nostocales. However for two vOTUs we were able to only predict hosts to the

order-level (vOTU602, vOTU638), while the third vOTU was predicted to infect the genus *Calothrix* (vOTU669). These predictions match the taxonomy of the Cyanobacterial member of the cultures that they were obtained from. Two of these vOTUs (vOTU602, vOTU669) originated from cultures isolated from desert environments, while the third originated from a culture isolated from tropical soil (vOTU638). Each of these three vOTUs was only detected in the culture it originated from. Using ViPTree, we found that these vOTUs were placed in a clade where many viral sequences had aquatic Cyanobacterial host predictions (Fig. S2). However the vOTUs were on long branches with limited similarity to reference genomes and no reference viruses were close relatives to the three vOTUs here. While finding only three cyanophage vOTUs was surprising given the predominance of Cyanobacteria in these cultures, previous work from biocrusts and hypoliths have similarly reported a low abundance of cyanophage (Zablocki, Adriaenssens & Cowan, 2016; Van Goethem et al., 2019). An important caveat to remember is the reliance of host predictions on available genomes; as Cyanobacterial genome availability increases, we expect that ability to link viruses to their hosts in these ecosystems will increase as well.

Despite not having a Cyanobacterial host prediction using BLAST, we also used ViPTree to investigate the two vOTUs in VC 1442_0 (the VC that included the reference virus, Cr□LKS3, reported to infect aquatic Cyanobacteria). The vOTUs in VC 1442_0 represent low-quality (vOTU185) and medium-quality (vOTU15) viral genomes from cultures isolated from aquatic plant and cave rock wall environments respectively. Using ViPTree, we found that vOTU15 was placed on a long branch in a clade that included Cr□LKS3 and it's closest relative Cy-LDV1 (Laloum et al., 2022), but no other viruses that had Cyanobacterial host predictions (Fig. S3).

However, Cr□LKS3 was reported to have higher genomic similarity to heterotrophic bacteria phage than known cyanophage (Laloum et al., 2022), so it is possible these vOTU15 may just represent another member of this novel cyanophage group. Despite being in the same VC, vOTU185 was placed elsewhere in the tree. We believe this is an erroneous placement due to vOTU185's low-quality and small size.

In an attempt to gain further insight about the possible interactions of these five vOTUs with their possible Cyanobacteria hosts, we annotated these viral sequences using Cenote-Taker2 to look for putative candidate AMGs. Cyanophages in aquatic ecosystems often are reported to have AMGs related to photosynthesis, phosphorus utilization, and carbon metabolism (Bailey et al., 2004; Sullivan et al., 2005, 2010; Millard et al., 2009). However, in all five vOTUs, we found the majority of non-viral genes to be unannotated (i.e., hypothetical proteins) and observed no clear AMG candidates related to common cyanophage functions (Fig.S4). Further assessment of homologous gene clusters between vOTUs and cyanophage close relatives, based on ViPTree analyses, did not identify any homologs that shared > 50% protein similarity. Instead, we mainly observed lower similarity matches to viral hallmark genes or hypothetical proteins (Fig. S5). One reason for the lack of clear AMG candidates may relate to viral completeness. Four of the five vOTUs were predicted by CheckV to be low or medium-quality, thus the inability to detect AMGs may relate to these viral sequences being incomplete. Another possibility is that the lack of AMG candidates may be representative of phage lifestyle (lytic vs. temperate), with lytic phage more likely to focus on hijacking host replication machinery and less driven to promote long-term host fitness through AMGs (Breitbart et al., 2018; Warwick-Dugdale et al., 2019; Luo et al., 2022). Regardless, as more studies focus on viral discovery and characterization efforts,

the power to identify and annotate AMGs will increase leading to better overall comprehension of the role of viruses in Cyanobacteria-dominated ecosystems.

Conclusion

To expand knowledge of viruses infecting Cyanobacteria-dominated microbiomes, we employed viral discovery methods on existing metagenomes from nonaxenic Cyanobacterial co-cultures as part of remote undergraduate research experiences (Ward et al., 2021). We were able to identify 814 viral sequences including sequences that formed unique VC and which may represent novel viral genera. While this work provides a catalog of viruses that infect terrestrial Cyanobacteria and associated cyanosphere heterotrophs, due to methodological limitations, we are likely only sampling DNA phage and are undersampling viruses in these cultures. Future work should seek to profile terrestrial Cyanobacteria viruses using more comprehensive methods such as through viromics sequencing of environmental samples. While we observed no wide scale biome or Cyanobacterial host patterns driving viral diversity, this is likely due to the limited nature of this work including the incomplete sampling of the viral community, the mixed nature of the nonaxenic cultures, and the fact that these are isolated cultures, not true environmental samples. However despite study limitations, profiling the viral community of co-cultures resulted in our ability to identify viral diversity on par with environmental metagenomic studies, as well as enabled the identification of possible novel viruses, including several potential cyanophage. Overall, future studies should build on this work to include a more extensive investigation of viral-Cyanobacteria interactions in extreme environments and profile the auxiliary metabolic genes of viral genomes which may play roles in various cycling processes in terrestrial Cyanobacteria-dominated environments.

Acknowledgements

We thank Jeffrey R. Johansen for being a lifelong mentor, collaborator, and friend to N.P. Jeff initiated a culture collection of terrestrial cyanobacteria at JCU from which subcultures of at least half of the investigated strains were obtained and investigated in Ward et al. (2021).

References

Adriaenssens EM, Van Zyl L, De Maayer P, Rubagotti E, Rybicki E, Tuffin M, Cowan DA. 2015. Metagenomic analysis of the viral community in Namib Desert hypoliths. *Environmental Microbiology* 17:480–495. DOI: 10.1111/1462-2920.12528.

Amgarten D, Braga LPP, da Silva AM, Setubal JC. 2018. MARVEL, a Tool for Prediction of Bacteriophage Sequences in Metagenomic Bins. *Frontiers in Genetics* 9. DOI: 10.3389/fgene.2018.00304.

Andrews S. *FastQC A Quality Control tool for High Throughput Sequence Data*.

Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, Dehal P, Ware D, Perez F, Canon S, Sneddon MW, Henderson ML, Riehl WJ, Murphy-Olson D, Chan SY, Kamimura RT, Kumari S, Drake MM, Brettin TS, Glass EM, Chivian D, Gunter D, Weston DJ, Allen BH, Baumohl J, Best AA, Bowen B, Brenner SE, Bun CC, Chandonia J-M, Chia J-M, Colasanti R, Conrad N, Davis JJ, Davison BH, DeJongh M, Devoid S, Dietrich E, Dubchak I, Edirisinghe JN, Fang G, Faria JP, Frybarger PM, Gerlach W, Gerstein M, Greiner A, Gurtowski J, Haun HL, He F, Jain R, Joachimiak MP, Keegan KP, Kondo S, Kumar V, Land ML, Meyer F, Mills M, Novichkov PS, Oh T, Olsen GJ, Olson R, Parrello B, Pasternak S, Pearson E, Poon SS, Price GA, Ramakrishnan S, Ranjan P, Ronald PC, Schatz MC, Seaver SMD, Shukla M, Sutormin RA, Syed MH, Thomason J, Tintle NL,

Wang D, Xia F, Yoo H, Yoo S, Yu D. 2018. KBase: The United States Department of Energy Systems Biology Knowledgebase. *Nature biotechnology* 36:566–569. DOI: 10.1038/nbt.4163.

Bailey S, Clokie MRJ, Millard A, Mann NH. 2004. Cyanophage infection and photoinhibition in marine cyanobacteria. *Research in microbiology* 155:720–725. DOI: 10.1016/j.resmic.2004.06.002.

Baldarelli LM, Pietrasiak N, Osorio-Santos K, Johansen JR. 2022. *Mojavia aguilerae* and *M. dolomitestris* - two new Nostocaceae (Cyanobacteria) species from the Americas. *Journal of phycology* 58:502–516. DOI: 10.1111/jpy.13275.

Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315:1709–1712. DOI: 10.1126/science.1138140.

van den Belt M, Gilchrist C, Booth TJ, Chooi Y-H, Medema MH, Alanjary M. 2023. CAGECAT: The CompArative GEne Cluster Analysis Toolbox for rapid search and visualisation of homologous gene clusters. *BMC bioinformatics* 24:181. DOI: 10.1186/s12859-023-05311-2.

Bin Jang H, Bolduc B, Zablocki O, Kuhn JH, Roux S, Adriaenssens EM, Brister JR, Kropinski AM, Krupovic M, Lavigne R, Turner D, Sullivan MB. 2019. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nature biotechnology* 37:632–639. DOI: 10.1038/s41587-019-0100-8.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. DOI: 10.1093/bioinformatics/btu170.

Bragg JG, Chisholm SW. 2008. Modeling the fitness consequences of a cyanophage-encoded

photosynthesis gene. *PloS one* 3:e3550. DOI: 10.1371/journal.pone.0003550.

Breitbart M, Bonnain C, Malki K, Sawaya NA. 2018. Phage puppet masters of the marine microbial realm. *Nature microbiology* 3:754–766. DOI: 10.1038/s41564-018-0166-y.

Bruno L, Billi D, Albertano P, Urzí C. 2006. Genetic Characterization of Epilithic Cyanobacteria and Their Associated Bacteria. *Geomicrobiology Journal* 23:293–299. DOI: 10.1080/01490450600760732.

Bushnell B. 2022.BBMap. *Available at https://sourceforge.net/projects/bbmap/* (accessed September 20, 2022).

Butts, C. 2008. network: a Package for Managing Relational Data in R. *Journal of Statistical Software* 24. DOI: 10.18637/jss.v024.i02.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC bioinformatics* 10:1–9. DOI: 10.1186/1471-2105-10-421.

Campitelli E. 2023. *ggnewscale: Multiple Fill and Colour Scales in "ggplot2."* Comprehensive R Archive Network (CRAN).

Carmichael WW. 1986. Isolation, culture and toxicity testing of toxic freshwatercyanobacteria (blue-green algae). In: Shilov V ed. *Fundamental Research in Homogeneous Catalysis.* New York, NY: Gordon & Breach, 1249–1262.

Casjens S. 2003. Prophages and bacterial genomics: what have we learned so far? *Molecular microbiology* 49:277–300. DOI: 10.1046/j.1365-2958.2003.03580.x.

Coclet C, Roux S. 2021. Global overview and major challenges of host prediction methods for uncultivated phages. *Current opinion in virology* 49:117–126. DOI: 10.1016/j.coviro.2021.05.003.

Cook R, Brown N, Redgwell T, Rihtman B, Barnes M, Clokie M, Stekel DJ, Hobman J, Jones MA, Millard A. 2021. INfrastructure for a PHAge REference Database: Identification of Large-Scale Biases in the Current Collection of Cultured Phage Genomes. *PHAGE* 2:214–223. DOI: 10.1089/phage.2021.0007.

Couradeau E, Giraldo-Silva A, De Martini F, Garcia-Pichel F. 2019. Spatial segregation of the biological soil crust microbiome around its foundational cyanobacterium, Microcoleus vaginatus, and the formation of a nitrogen-fixing cyanosphere. *Microbiome* 7:55. DOI: 10.1186/s40168-019-0661-2.

Crummett LT, Puxty RJ, Weihe C, Marston MF, Martiny JBH. 2016. The genomic content and context of auxiliary metabolic genes in marine cyanomyoviruses. *Virology* 499:219–229. DOI: 10.1016/j.virol.2016.09.016.

Ecogenomics/BamM: Metagenomics-focused BAM file manipulation. *Available at https://github.com/Ecogenomics/BamM* (accessed September 20, 2022).

Edwards RA, McNair K, Faust K, Raes J, Dutilh BE. 2016. Computational approaches to predict bacteriophage-host relationships. *FEMS microbiology reviews* 40:258–272. DOI: 10.1093/femsre/fuv048.

Ettinger C, Pietrasiak N, Stajich J. *stajichlab/Terrestrial_Cyanophage: v1*. DOI: 10.5281/zenodo.7272374.

Ettinger CL, Saunders M, Selbmann L, Delgado-Baquerizo M, Donati C, Albanese D, Roux S, Tringe S, Pennacchio C, Del Rio TG, Stajich JE, Coleine C. 2023. Highly diverse and unknown viruses may enhance Antarctic endoliths' adaptability. *Microbiome* 11:103. DOI: 10.1186/s40168-023-01554-6.

Fortier L-C, Sekulovic O. 2013. Importance of prophages to evolution and virulence of bacterial

pathogens. *Virulence* 4:354–365. DOI: 10.4161/viru.24498.

Gao E-B, Gui J-F, Zhang Q-Y. 2012. A novel cyanophage with a cyanobacterial nonbleaching protein A gene in the genome. *Journal of virology* 86:236–245. DOI: 10.1128/JVI.06282-11.

Gilchrist CLM, Chooi Y-H. 2021. clinker & clustermap.js: automatic generation of gene cluster comparison figures. *Bioinformatics* 37:2473–2475. DOI: 10.1093/bioinformatics/btab007.

Guajardo-Leiva S, Pedrós-Alió C, Salgado O, Pinto F, Díez B. 2018a. Active Crossfire Between Cyanobacteria and Cyanophages in Phototrophic Mat Communities Within Hot Springs. *Frontiers in microbiology* 9:2039. DOI: 10.3389/fmicb.2018.02039.

Guajardo-Leiva S, Pedrós-Alió C, Salgado O, Pinto F, Díez B. 2018b. Active Crossfire Between Cyanobacteria and Cyanophages in Phototrophic Mat Communities Within Hot Springs. *Frontiers in microbiology* 9:2039. DOI: 10.3389/fmicb.2018.02039.

Guo J, Bolduc B, Zayed AA, Varsani A, Dominguez-Huerta G, Delmont TO, Pratama AA, Gazitúa MC, Vik D, Sullivan MB, Roux S. 2021. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome* 9:37. DOI: 10.1186/s40168-020-00990-y.

Guttman B, Raya R, Kutter E. 2004. Basic Phage Biology. *Bacteriophages.* DOI: 10.1201/9780203491751.ch3.

Hampton HG, Watson BNJ, Fineran PC. 2020. The arms race between bacteria and their phage foes. *Nature* 577:327–336. DOI: 10.1038/s41586-019-1894-8.

Huang S, Wang K, Jiao N, Chen F. 2012. Genome sequences of siphoviruses infecting marine Synechococcus unveil a diverse cyanophage group and extensive phage-host genetic exchanges. *Environmental microbiology* 14:540–558. DOI: 10.1111/j.1462-

2920.2011.02667.x.

Hurwitz BL, Hallam SJ, Sullivan MB. 2013. Metabolic reprogramming by viruses in the sunlit and dark ocean. *Genome biology* 14:R123. DOI: 10.1186/gb-2013-14-11-r123.

Hurwitz BL, U'Ren JM. 2016. Viral metabolic reprogramming in marine ecosystems. *Current opinion in microbiology* 31:161–168. DOI: 10.1016/j.mib.2016.04.002.

Hurwitz BL, Westveld AH, Brum JR, Sullivan MB. 2014. Modeling ecological drivers in marine viral communities using comparative metagenomics and network analyses. *Proceedings of the National Academy of Sciences of the United States of America* 111:10714–10719. DOI: 10.1073/pnas.1319778111.

Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC bioinformatics* 11:119. DOI: 10.1186/1471-2105-11-119.

Kavagutti VS, Andrei A-Ş, Mehrshad M, Salcher MM, Ghai R. 2019. Phage-centric ecological interactions in aquatic ecosystems revealed through ultra-deep metagenomics. *Microbiome* 7:135. DOI: 10.1186/s40168-019-0752-0.

kblin/ncbi-genome-download: Scripts to download genomes from the NCBI FTP servers. *Available at https://github.com/kblin/ncbi-genome-download* (accessed September 20, 2022).

Kieft K, Zhou Z, Anantharaman K. 2020. VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome* 8:90. DOI: 10.1186/s40168-020-00867-0.

Kim M, Shin B, Lee J, Park HY, Park W. 2019. Culture-independent and culture-dependent analyses of the bacterial community in the phycosphere of cyanobloom-forming

Microcystis aeruginosa. *Scientific reports* 9:20416. DOI: 10.1038/s41598-019-56882-1.

Kothari A, Roux S, Zhang H, Prieto A, Soneja D, Chandonia J-M, Spencer S, Wu X, Altenburg S, Fields MW, Deutschbauer AM, Arkin AP, Alm EJ, Chakraborty R, Mukhopadhyay A. 2021. Ecogenomics of Groundwater Phages Suggests Niche Differentiation Linked to Specific Environmental Tolerance. *mSystems*:e0053721. DOI: 10.1128/mSystems.00537-21.

Laloum E, Cattan□Tsaushu E, Schwartz DA, Shaalan H, Enav H, Kolan D, Avrani S. 2022. Isolation and characterization of a novel Lambda□like phage infecting the bloom□forming cyanobacteria Cylindrospermopsis raciborskii. *Environmental microbiology* 24:2435. DOI: 10.1111/1462-2920.15908.

Lee JZ, Burow LC, Woebken D, Everroad RC, Kubo MD, Spormann AM, Weber PK, Pett-Ridge J, Bebout BM, Hoehler TM. 2014. Fermentation couples Chloroflexi and sulfate-reducing bacteria to Cyanobacteria in hypersaline microbial mats. *Frontiers in microbiology* 5:61. DOI: 10.3389/fmicb.2014.00061.

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. DOI: 10.1093/bioinformatics/btl158.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079. DOI: 10.1093/bioinformatics/btp352.

Li Z, Hobson P, An W, Burch MD, House J, Yang M. 2012. Earthy odor compounds production and loss in three cyanobacterial cultures. *Water research* 46:5165–5173. DOI: 10.1016/j.watres.2012.06.008.

Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31:1674–1676. DOI: 10.1093/bioinformatics/btv033.

Luo X-Q, Wang P, Li J-L, Ahmad M, Duan L, Yin L-Z, Deng Q-Q, Fang B-Z, Li S-H, Li W-J. 2022. Viral community-wide auxiliary metabolic genes differ by lifestyles, habitats, and hosts. *Microbiome* 10:190. DOI: 10.1186/s40168-022-01384-y.

Mann NH, Clokie MRJ, Millard A, Cook A, Wilson WH, Wheatley PJ, Letarov A, Krisch HM. 2005. The genome of S-PM2, a “photosynthetic” T4-type bacteriophage that infects marine Synechococcus strains. *Journal of bacteriology* 187:3188–3200. DOI: 10.1128/JB.187.9.3188-3200.2005.

McMurdie PJ, Holmes S. 2013. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PloS one* 8:e61217. DOI: 10.1371/journal.pone.0061217.

Millard AD, Zwirglmaier K, Downey MJ, Mann NH, Scanlan DJ. 2009. Comparative genomics of marine cyanomyoviruses reveals the widespread occurrence of *Synechococcus* host genes localized to a hyperplastic region: implications for mechanisms of cyanophage evolution. *Environmental microbiology* 11:2370–2387. DOI: 10.1111/j.1462-2920.2009.01966.x.

Nayfach S, Camargo AP, Schulz F, Eloe-Fadrosh E, Roux S, Kyrpides NC. 2021. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nature biotechnology* 39:578–585. DOI: 10.1038/s41587-020-00774-7.

Nishimura Y, Yoshida T, Kuronishi M, Uehara H, Ogata H, Goto S. 2017. ViPTree: the viral proteomic tree server. *Bioinformatics* 33:2379–2380. DOI: 10.1093/bioinformatics/btx157.

Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. 2017. metaSPAdes: a new versatile

metagenomic assembler. *Genome research* 27:824–834. DOI: 10.1101/gr.213959.116.

Paerl HW, Pinckney JL, Steppe TF. 2000. Cyanobacterial-bacterial mat consortia: examining the functional unit of microbial survival and growth in extreme environments. *Environmental microbiology* 2:11–26. DOI: 10.1046/j.1462-2920.2000.00071.x.

Paez-Espino D, Eloe-Fadrosh EA, Pavlopoulos GA, Thomas AD, Huntemann M, Mikhailova N, Rubin E, Ivanova NN, Kyrpides NC. 2016. Uncovering Earth's virome. *Nature* 536:425–430. DOI: 10.1038/nature19094.

Patzelt DJ, Hodač L, Friedl T, Pietrasiak N, Johansen JR. 2014. Biodiversity of soil cyanobacteria in the hyper-arid Atacama Desert, Chile. *Journal of phycology* 50:698–710. DOI: 10.1111/jpy.12196.

Pedersen TL. 2022a. *patchwork: The Composer of Plots*. Comprehensive R Archive Network (CRAN).

Pedersen TL. 2022b. *ggforce: Accelerating "ggplot2."* Comprehensive R Archive Network (CRAN).

Perera I, Subashchandrabose SR, Venkateswarlu K, Naidu R, Megharaj M. 2018. Consortia of cyanobacteria/microalgae and bacteria in desert soils: an underexplored microbiota. *Applied microbiology and biotechnology* 102:7351–7363. DOI: 10.1007/s00253-018-9192-1.

Pietrasiak N, Osorio-Santos K, Shalygin S, Martin MP, Johansen JR. 2019. When Is A Lineage A Species? A Case Study In *Myxacorys* gen. nov. (Synechococcales: Cyanobacteria) With The Description of Two New Species From The Americas. *Journal of phycology* 55:976–996. DOI: 10.1111/jpy.12897.

Pietrasiak N, Reeve S, Osorio-Santos K, Lipson DA, Johansen JR. 2021. *Trichotorquatus* gen. nov. - a new genus of soil cyanobacteria discovered from American drylands. *Journal of*

phycology 57:886–902. DOI: 10.1111/jpy.13147.

Pombubpa N, Pietrasiak N, De Ley P, Stajich JE. 2020. Insights into dryland biocrust microbiome: geography, soil depth and crust type affect biocrust microbial communities and networks in Mojave Desert, USA. *FEMS microbiology ecology* 96. DOI: 10.1093/femsec/fiaa125.

Ponsero AJ, Hurwitz BL, Magain N, Miadlikowska J, Lutzoni F, U'Ren JM. 2021. Cyanolichen microbiome contains novel viruses that encode genes to promote microbial metabolism. *ISME Communications* 1:1–4. DOI: 10.1038/s43705-021-00060-w.

Potapov SA, Tikhonova IV, Krasnopeev AY, Kabilov MR, Tupikin AE, Chebunina NS, Zhuchenko NA, Belykh OI. 2019. Metagenomic Analysis of Virioplankton from the Pelagic Zone of Lake Baikal. *Viruses* 11. DOI: 10.3390/v11110991.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841–842. DOI: 10.1093/bioinformatics/btq033.

R Core Team. 2021. *R: A Language and Environment for Statistical Computing*.

Ren J, Ahlgren NA, Lu YY, Fuhrman JA, Sun F. 2017. VirFinder: a novel k-mer based tool for identifying viral sequences from assembled metagenomic data. *Microbiome* 5:69. DOI: 10.1186/s40168-017-0283-5.

Roux S, Adriaenssens EM, Dutilh BE, Koonin EV, Kropinski AM, Krupovic M, Kuhn JH, Lavigne R, Brister JR, Varsani A, Amid C, Aziz RK, Bordenstein SR, Bork P, Breitbart M, Cochrane GR, Daly RA, Desnues C, Duhaime MB, Emerson JB, Enault F, Fuhrman JA, Hingamp P, Hugenholtz P, Hurwitz BL, Ivanova NN, Labonté JM, Lee K-B, Malmstrom RR, Martinez-Garcia M, Mizrachi IK, Ogata H, Páez-Espino D, Petit M-A, Putonti C, Rattei T, Reyes A, Rodriguez-Valera F, Rosario K, Schriml L, Schulz F, Steward GF,

Sullivan MB, Sunagawa S, Suttle CA, Temperton B, Tringe SG, Thurber RV, Webster NS, Whiteson KL, Wilhelm SW, Wommack KE, Woyke T, Wrighton KC, Yilmaz P, Yoshida T, Young MJ, Yutin N, Allen LZ, Kyrpides NC, Eloe-Fadrosh EA. 2018. Minimum Information about an Uncultivated Virus Genome (MIUViG). *Nature biotechnology* 37:29–37. DOI: 10.1038/nbt.4306.

Roux S, Brum JR, Dutilh BE, Sunagawa S, Duhaime MB, Loy A, Poulos BT, Solonenko N, Lara E, Poulain J, Pesant S, Kandels-Lewis S, Dimier C, Picheral M, Searson S, Cruaud C, Alberti A, Duarte CM, Gasol JM, Vaqué D, Tara Oceans Coordinators, Bork P, Acinas SG, Wincker P, Sullivan MB. 2016. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature* 537:689–693. DOI: 10.1038/nature19366.

Roux S, Emerson JB, Eloe-Fadrosh EA, Sullivan MB. 2017. Benchmarking viromics: an evaluation of metagenome-enabled estimates of viral community composition and diversity. *PeerJ* 5:e3817. DOI: 10.7717/peerj.3817.

Roux S, Enault F, Hurwitz BL, Sullivan MB. 2015. VirSorter: mining viral signal from microbial genomic data. *PeerJ* 3:e985. DOI: 10.7717/peerj.985.

Sabehi G, Shaulov L, Silver DH, Yanai I, Harel A, Lindell D. 2012. A novel lineage of myoviruses infecting cyanobacteria is widespread in the oceans. *Proceedings of the National Academy of Sciences of the United States of America* 109:2037–2042. DOI: 10.1073/pnas.1115467109.

Santos-Medellin C, Zinke LA, Ter Horst AM, Gelardi DL, Parikh SJ, Emerson JB. 2021. Viromes outperform total metagenomes in revealing the spatiotemporal patterns of agricultural soil viral communities. *The ISME journal* 15:1956–1970. DOI: 10.1038/s41396-021-00897-y.

Schloerke B, Cook D, Larmarange J, Briatte F, Marbach M, Thoen E, Elberg A, Crowley J. 2021. *GGally: Extension to "ggplot2."* Comprehensive R Archive Network (CRAN).

Shi L, Cai Y, Yang H, Xing P, Li P, Kong L, Kong F. 2009. Phylogenetic diversity and specificity of bacteria associated with Microcystis aeruginosa and other cyanobacteria. *Journal of environmental sciences* 21:1581–1590. DOI: 10.1016/s1001-0742(08)62459-6.

Sieradzki ET, Ignacio-Espinoza JC, Needham DM, Fichot EB, Fuhrman JA. 2019. Dynamic marine viral infections and major contribution to photosynthetic processes shown by spatiotemporal picoplankton metatranscriptomes. *Nature communications* 10:1169. DOI: 10.1038/s41467-019-09106-z.

Skvortsov T, de Leeuwe C, Quinn JP, McGrath JW, Allen CCR, McElarney Y, Watson C, Arkhipova K, Lavigne R, Kulakov LA. 2016. Metagenomic Characterisation of the Viral Community of Lough Neagh, the Largest Freshwater Lake in Ireland. *PloS one* 11:e0150361. DOI: 10.1371/journal.pone.0150361.

Stothard P, Grant JR, Van Domselaar G. 2019. Visualizing and comparing circular genomes using the CGView family of tools. *Briefings in bioinformatics* 20:1576–1582. DOI: 10.1093/bib/bbx081.

Sullivan MB. 2015. Viromes, Not Gene Markers, for Studying Double-Stranded DNA Virus Communities. *Journal of Virology* 89:2459–2461. DOI: 10.1128/jvi.03289-14.

Sullivan MB, Coleman ML, Weigele P, Rohwer F, Chisholm SW. 2005. Three *Prochlorococcus* cyanophage genomes: signature features and ecological interpretations. *PLoS biology* 3:e144. DOI: 10.1371/journal.pbio.0030144.

Sullivan MB, Huang KH, Ignacio-Espinoza JC, Berlin AM, Kelly L, Weigele PR, DeFrancesco AS, Kern SE, Thompson LR, Young S, Yandava C, Fu R, Krastins B, Chase M, Sarracino

D, Osburne MS, Henn MR, Chisholm SW. 2010. Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments. *Environmental microbiology* 12:3035–3056. DOI: 10.1111/j.1462-2920.2010.02280.x.

Sullivan MB, Krastins B, Hughes JL, Kelly L, Chase M, Sarracino D, Chisholm SW. 2009. The genome and structural proteome of an ocean siphovirus: a new window into the cyanobacterial "mobilome." *Environmental microbiology* 11:2935–2951. DOI: 10.1111/j.1462-2920.2009.02081.x.

Thompson LR, Zeng Q, Kelly L, Huang KH, Singer AU, Stubbe J, Chisholm SW. 2011. Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism. *Proceedings of the National Academy of Sciences of the United States of America* 108:E757–64. DOI: 10.1073/pnas.1102164108.

Tisza MJ, Belford AK, Domínguez-Huerta G, Bolduc B, Buck CB. 2021. Cenote-Taker 2 democratizes virus discovery and sequence annotation. *Virus evolution* 7:veaa100. DOI: 10.1093/ve/veaa100.

Turner D, Kropinski AM, Adriaenssens EM. 2021. A Roadmap for Genome-Based Phage Taxonomy. *Viruses* 13. DOI: 10.3390/v13030506.

Van Goethem MW, Swenson TL, Trubl G, Roux S, Northen TR. 2019. Characteristics of Wetting-Induced Bacteriophage Blooms in Biological Soil Crust. *mBio* 10. DOI: 10.1128/mBio.02287-19.

Velichko N, Smirnova S, Averina S, Pinevich A. 2021. A survey of Antarctic cyanobacteria. *Hydrobiologia* 848:2627–2652. DOI: 10.1007/s10750-021-04588-9.

Voorhies AA, Eisenlord SD, Marcus DN, Duhaime MB, Biddanda BA, Cavalcoli JD, Dick GJ.

2016. Ecological and genetic interactions between cyanobacteria and viruses in a low-oxygen mat community inferred through metagenomics and metatranscriptomics. *Environmental microbiology* 18. DOI: 10.1111/1462-2920.12756.

Ward RD, Stajich JE, Johansen JR, Huntemann M, Clum A, Foster B, Foster B, Roux S, Palaniappan K, Varghese N, Mukherjee S, Reddy TBK, Daum C, Copeland A, Chen I-MA, Ivanova NN, Kyrpides NC, Shapiro N, Eloe-Fadrosh EA, Pietrasiak N. 2021. Metagenome Sequencing to Explore Phylogenomics of Terrestrial Cyanobacteria. *Microbiology resource announcements* 10:e0025821. DOI: 10.1128/MRA.00258-21.

Warwick-Dugdale J, Buchholz HH, Allen MJ, Temperton B. 2019. Host-hijacking and planktonic piracy: how phages command the microbial high seas. *Virology journal* 16:15. DOI: 10.1186/s12985-019-1120-1.

Weigele PR, Pope WH, Pedulla ML, Houtz JM, Smith AL, Conway JF, King J, Hatfull GF, Lawrence JG, Hendrix RW. 2007. Genomic and structural analysis of Syn9, a cyanophage infecting marine Prochlorococcus and Synechococcus. *Environmental microbiology* 9:1675–1695. DOI: 10.1111/j.1462-2920.2007.01285.x.

Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H. 2019. Welcome to the Tidyverse. *Journal of Open Source Software* 4:1686. DOI: 10.21105/joss.01686.

Wilke C, Wiernik B. 2022. *ggtext: Improved Text Rendering Support for "ggplot2."* Comprehensive R Archive Network (CRAN).

Zablocki O, Adriaenssens EM, Cowan D. 2016. Diversity and Ecology of Viruses in Hyperarid

Desert Soils. *Applied and environmental microbiology* 82:770–777. DOI: 10.1128/AEM.02651-15.

Zablocki O, van Zyl L, Adriaenssens EM, Rubagotti E, Tuffin M, Cary SC, Cowan D. 2014. High-level diversity of tailed phages, eukaryote-associated viruses, and virophage-like elements in the metaviromes of antarctic soils. *Applied and environmental microbiology* 80:6888–6897. DOI: 10.1128/AEM.01525-14.

Zeigler Allen L, McCrow JP, Ininbergs K, Dupont CL, Badger JH, Hoffman JM, Ekman M, Allen AE, Bergman B, Venter JC. 2017. The Baltic Sea Virome: Diversity and Transcriptional Activity of DNA and RNA Viruses. *mSystems* 2. DOI: 10.1128/mSystems.00125-16.

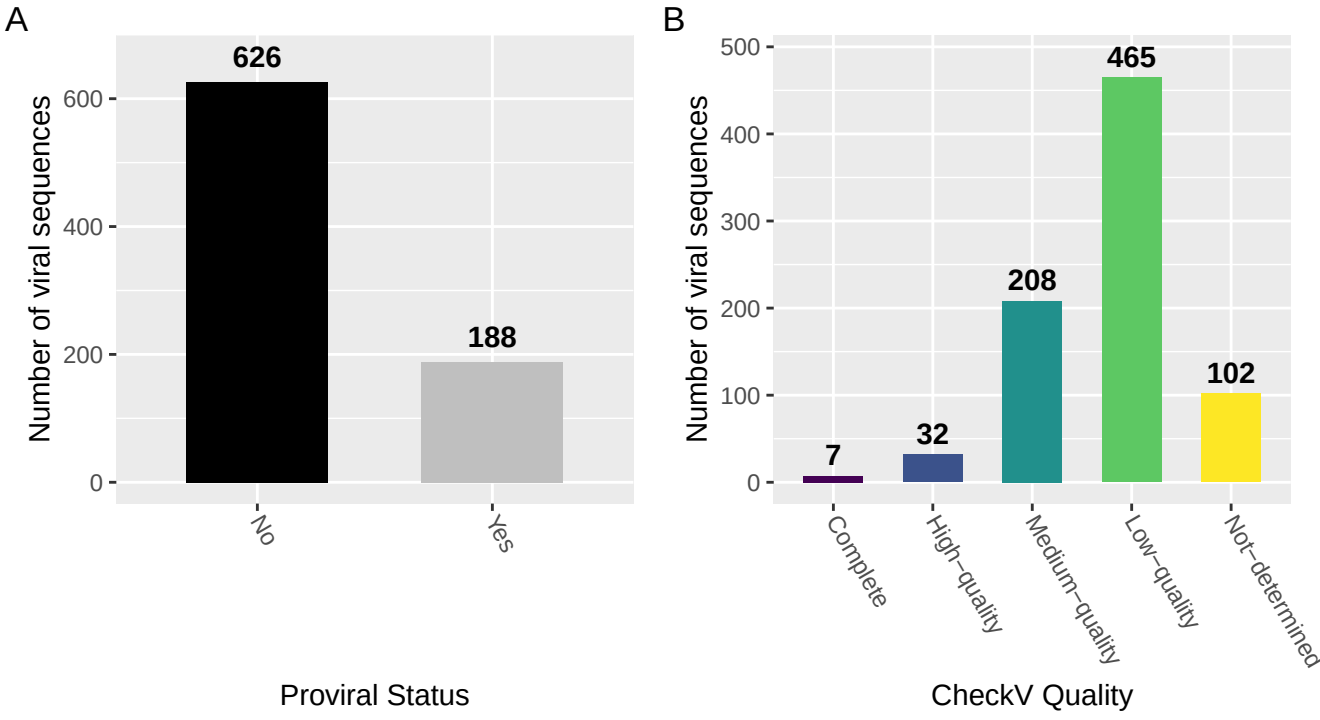
Figures and Tables:

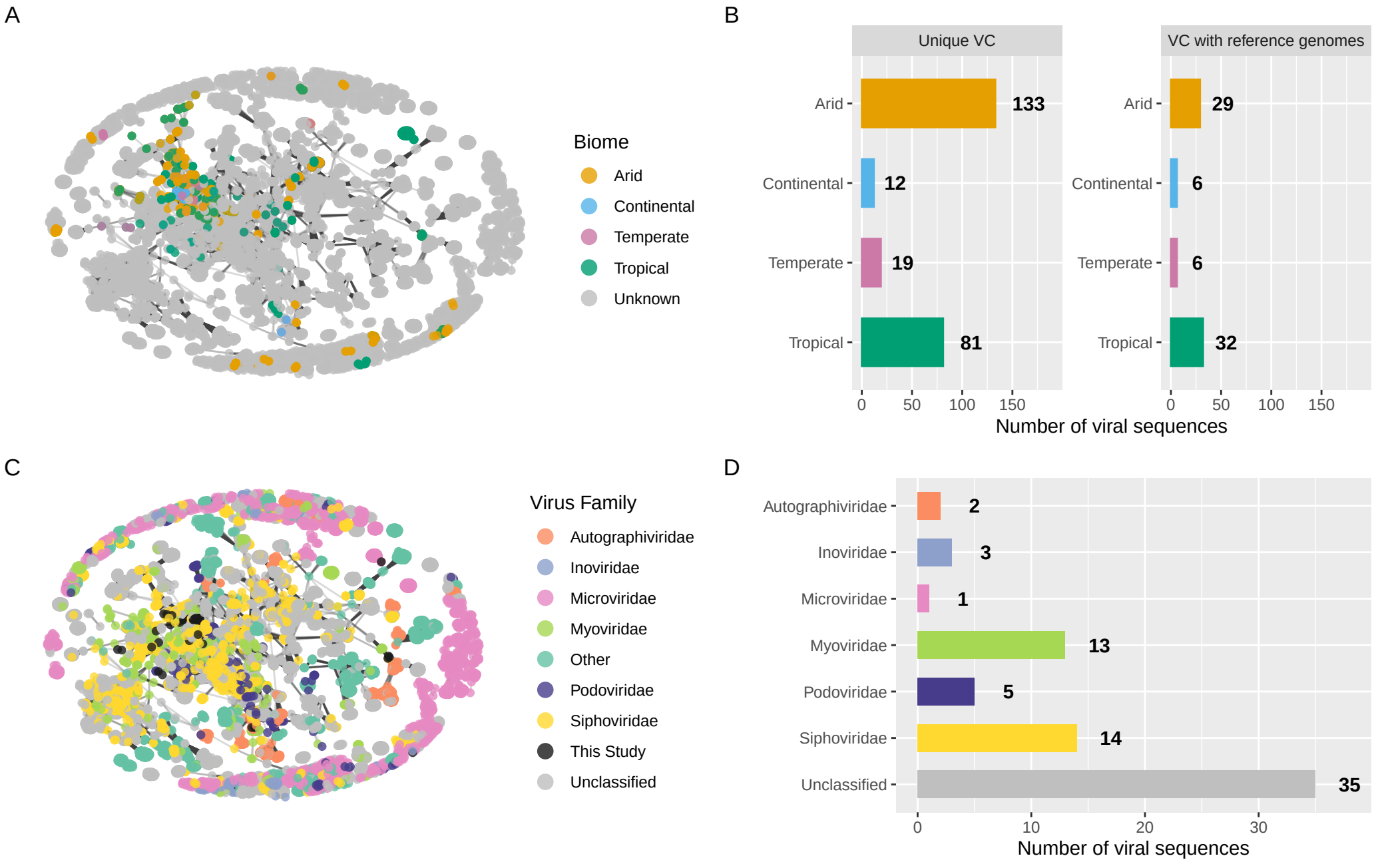
Figure 1: Viral sequences identified from Cyanobacterial co-cultures. (A) The proviral status of predicted viral genomes is shown as a bar graph, with the number above the bar representing the total number of sequences. (B) CheckV quality predictions for viral sequences are shown as a bar graph with bars colored by quality and the number above the bar representing the total number of sequences in each bar.

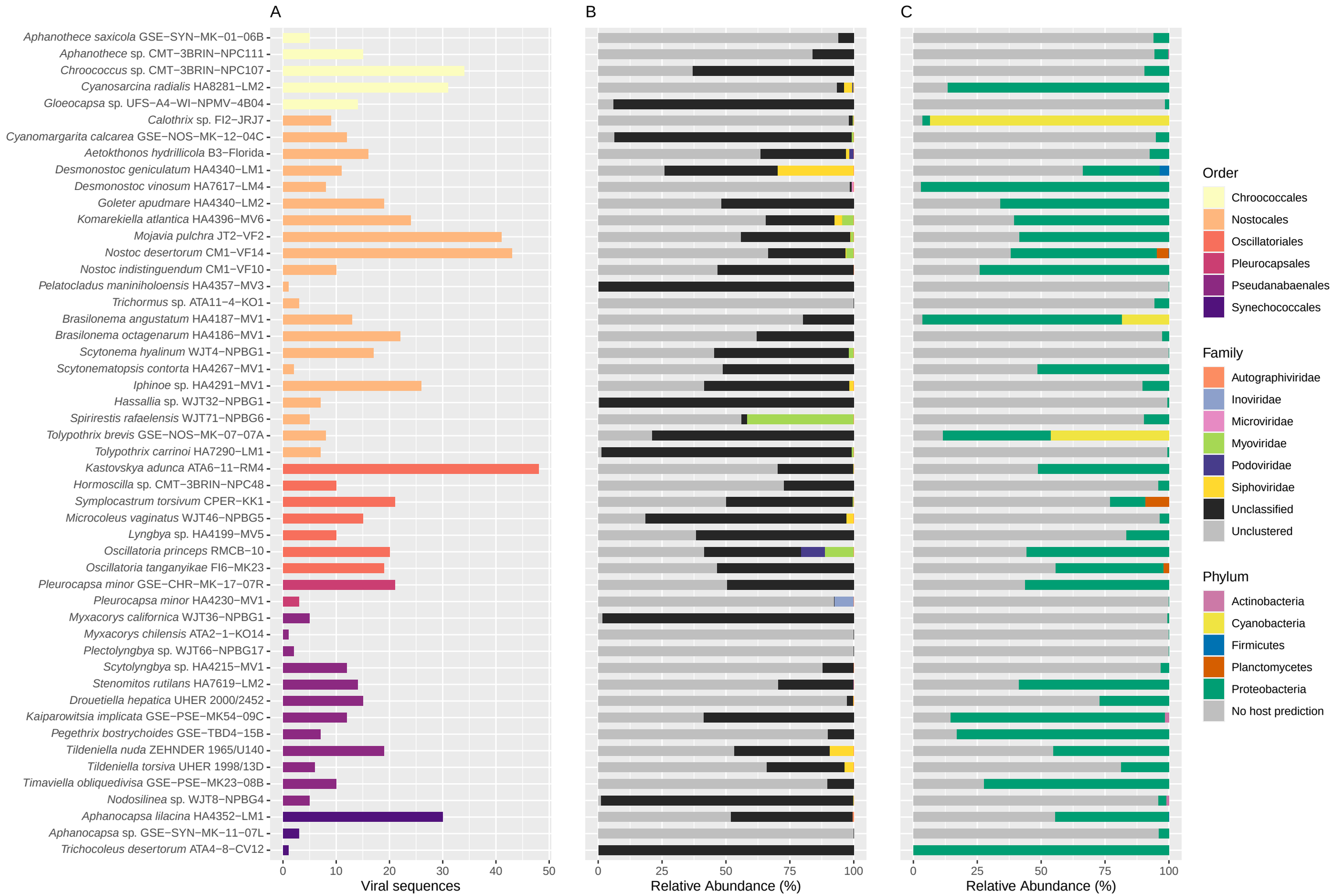
Figure 2: Viral catalog comparison to reference databases (A) Gene-sharing network of vOTUs that formed viral clusters (VCs), including VCs unique to this dataset. Each node is a viral sequence and is colored by the broad biome of the Cyanobacterial co-culture (sample sizes: arid: 28, tropical: 16, continental: 3, temperate: 3). Edges connect sequences with overlap in predicted protein content. (B) Bar charts displaying the number of viral sequences in VCs from this study colored by biome and shown facetted by whether the VC clustered with reference data. Bars are colored by biome, as in A, and the number next to the bar represents the total number of sequences it represents. (C) Subnetwork of VCs with members that include reference genomes obtained from INPHARED. Each node is a viral sequence and is colored by the predicted historical viral taxonomic family for the VC, with the exception of sequences identified in this study which are colored in black. Edges connect sequences with overlap in predicted protein content. (D) Bar chart displaying the number of viral sequences from this study assigned to historical viral families based on VC membership with reference genomes from INPHARED. Bars are colored by historical viral families, as in C, and the number above the bar represents the

total number of sequences it represents. For A-D, only viral sequences generated in this study where CheckV was able to predict quality are shown.

Figure 3: Viral communities across Cyanobacterial co-cultures. (A) The number of viral sequences predicted from each co-culture metagenome. Bars are arranged and colored based on the taxonomic order the Cyanobacteria in the co-culture is taxonomically classified as. (B) Stacked bar charts displaying the taxonomic relative abundance of vOTUs in each co-culture. Bars are colored based on predicted historical viral taxonomic families. Viral sequences that clustered into VCs, but that lack taxonomy (i.e., unique VCs that lack reference data, or VCs where reference data taxonomy is unclear) are labeled “Unassigned”, while viral sequences that did not cluster into VCs are labeled “Unclustered”. (C) Stacked bar charts of the relative abundance of vOTUs based on predicted hosts in each co-culture. Bars are colored based on the predicted host taxonomy, at the phylum level, of each vOTU . For A-C, only viral sequences where CheckV was able to predict quality are shown.


A
Number of viral sequences
0
200
400
600
626
188
No
Yes
Proviral Status
B
Number of viral sequences
0
100
200
300
400
500
7
32
208
465
102
Complete
High−quality
Medium−quality
Low−quality
Not−determined
CheckV Quality


A
Biome
Arid
Continental
Temperate
Tropical
Unknown
B
Unique VC
VC with reference genomes
Arid
Continental
Temperate
Tropical
133
12
19
81
29
6
6
32
0
50
100
150
Number of viral sequences
C
Virus Family
Autographiviridae
Inoviridae
Microviridae
Myoviridae
Other
Podoviridae
Siphoviridae
This Study
Unclassified
D
Autographiviridae
Inoviridae
Microviridae
Myoviridae
Podoviridae
Siphoviridae
Unclassified
2
3
1
13
5
14
35
0
10
20
30
Number of viral sequences


A
B
C
Aphanothece saxicola GSE−SYN−MK−01−06B
Aphanothece sp. CMT−3BRIN−NPC111
Chroococcus sp. CMT−3BRIN−NPC107
Cyanosarcina radialis HA8281−LM2
Gloeocapsa sp. UFS−A4−WI−NPMV−4B04
Calothrix sp. FI2−JRJ7
Cyanomargarita calcarea GSE−NOS−MK−12−04C
Aetokthonos hydrillicola B3−Florida
Desmonostoc geniculatum HA4340−LM1
Desmonostoc vinosum HA7617−LM4
Goleter apudmare HA4340−LM2
Komarekiella atlantica HA4396−MV6
Mojavia pulchra JT2−VF2
Nostoc desertorum CM1−VF14
Nostoc indistinguendum CM1−VF10
Pelatocladus maniniholoensis HA4357−MV3
Trichormus sp. ATA11−4−KO1
Brasilonema angustatum HA4187−MV1
Brasilonema octagenarum HA4186−MV1
Scytonema hyalinum WJT4−NPBG1
Scytonematopsis contorta HA4267−MV1
Iphinoe sp. HA4291−MV1
Hassallia sp. WJT32−NPBG1
Spirirestis rafaelensis WJT71−NPBG6
Tolypothrix brevis GSE−NOS−MK−07−07A
Tolypothrix carrinoi HA7290−LM1
Kastovskya adunca ATA6−11−RM4
Hormoscilla sp. CMT−3BRIN−NPC48
Symplocastrum torsivum CPER−KK1
Microcoleus vaginatus WJT46−NPBG5
Lyngbya sp. HA4199−MV5
Oscillatoria princeps RMCB−10
Oscillatoria tanganyikae FI6−MK23
Pleurocapsa minor GSE−CHR−MK−17−07R
Pleurocapsa minor HA4230−MV1
Myxacorys californica WJT36−NPBG1
Myxacorys chilensis ATA2−1−KO14
Plectolyngbya sp. WJT66−NPBG17
Scytolyngbya sp. HA4215−MV1
Stenomitos rutilans HA7619−LM2
Drouetiella hepatica UHER 2000/2452
Kaiparowitsia implicata GSE−PSE−MK54−09C
Pegethrix bostrychoides GSE−TBD4−15B
Tildeniella nuda ZEHNDER 1965/U140
Tildeniella torsiva UHER 1998/13D
Timaviella obliquedivisa GSE−PSE−MK23−08B
Nodosilinea sp. WJT8−NPBG4
Aphanocapsa lilacina HA4352−LM1
Aphanocapsa sp. GSE−SYN−MK−11−07L
Trichocoleus desertorum ATA4−8−CV12
0
10
20
30
40
50
Viral sequences
0
25
50
75
100
Relative Abundance (%)
0
25
50
75
100
Relative Abundance (%)
Order
Chroococcales
Nostocales
Oscillatoriales
Pleurocapsales
Pseudanabaenales
Synechococcales
Family
Autographiviridae
Inoviridae
Microviridae
Myoviridae
Podoviridae
Siphoviridae
Unclassified
Unclustered
Phylum
Actinobacteria
Cyanobacteria
Firmicutes
Planctomycetes
Proteobacteria
No host prediction