

Detection of bacteriophages and cyanophages for water quality control of a water supply system in Amazônia

Bruna Verônica Azevedo Gois

Universidade Federal do Pará: Universidade Federal do Para

Kenny da Costa Pinheiro

Universidade Federal do Pará: Universidade Federal do Para

Wylerson Guimarães Nogueira

Universidade Federal de Minas Gerais Instituto de Ciências Biológicas: Universidade Federal de Minas Gerais Instituto de Ciencias Biologicas

Andressa de Oliveira Aragão

Universidade Federal do Pará: Universidade Federal do Para

Ana Lúcia Cavalcante Queiroz

Universidade Federal do Pará: Universidade Federal do Para

Artur Silva

Universidade Federal do Para

Adriana Carneiro Folador

Universidade Federal do Pará: Universidade Federal do Para

Rommel Ramos (✉ rommelthiago@gmail.com)

Universidade Federal do Pará: Universidade Federal do Para <https://orcid.org/0000-0002-8032-1474>

Research article

Keywords: Amazonia, diversity, phages, metagenomics, virome

Posted Date: March 8th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-154634/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Despite the importance of understanding the ecology of freshwater viruses, there are not many studies on the issue when compared to marine viruses. The microbiological interactions that occur in these environments are still poorly known, especially between bacteriophages and their host bacteria, as well as between cyanophages and cyanobacteria. Lake Bologna, from Belém, capital of the Brazilian State of Pará, is a source of water that supplies the city and its metropolitan region, yet it remains unexplored regarding the contents of its virome and viral diversity composition. Therefore, this work's main aim is to clarify in terms of taxonomic diversity the species of DNA viruses that are present in this lake, especially bacteriophages and cyanophages, since they can act both as transducers of resistance genes and reporters of water quality for human consumption.

Results: For this work, we used the metagenomic sequencing data generated by Alves et al. (2020), and we analyzed it at the taxonomic level using the tools Kraken2, Bracken, and Pavian; later, the data was assembled using Genome Detective, which performs assembly of viruses. The results observed in this work suggest the existence of a widely diverse viral community and an established microbial phage regulated dynamics in the Lake Bolonha.

Conclusions: This work is the first-ever to describe the virome of Lake Bolonha using a metagenomic approach based on high-throughput sequencing, as it contributes to the understanding of water-related public health concerns regarding the spreading of antibiotic resistance genes and population control of native bacteria and cyanobacteria.

1. Introduction

Lakes close to urban areas are increasingly changing their ecosystem as human population expansion occurs and commercial, recreational, and residential uses increase [1]. The eutrophication process offers particular conditions for the replication of viruses, as environments of this type seem to provide high viral activation, as well as hypothetically controlling host abundance, respiration, and production [2].

In lake environments, there is little information about the ecology of the freshwater virus when compared to marine viruses [3]. Most of them are bacteriophages or human and other animal viruses, but plant viruses are also identified [4]. Bacteriophages can play an important function in the aquatic ecosystem as they can also contribute to the acquisition and spread of antibiotic resistance genes (ARGs) [5]. Some studies have shown ARGs-carrying phages to be abundant in many environments, especially those impacted by anthropogenic activities [6–10], which demonstrates that these types of viruses are relevant to local microbial ecology [6].

Like bacteriophages, another important category of viruses is cyanophages, which infects cyanobacteria and has a morphology similar to bacteriophages [11]. Being abundant in both fresh and saltwater, and they are relevant due to playing an important role in modulating cyanobacterial populations and preserving water quality [11, 12]. However, unlike bacteriophages, it has many genera of possible hosts;

therefore, freshwater cyanophages can be classified according to the taxonomy of their host organisms [13].

Given the need to study the diversity of different environments, including aquatic one, techniques have been developed, such as viral metagenomics, also known as virome. This technique allows the study of a variety of viruses from environmental samples [14]. In this way, the use of metagenomics and next-generation sequencing (NGS) to explore viral populations, both in aquatic environments and within the human microbiome, has demonstrated considerable genetic complexity as well as inter- and intra-species interaction [15, 16]. However, despite an increasing number of studies using the technique, there are still significant gaps in the virome databases. It has been estimated that 1,031 viral particles are infecting bacterial populations, but less than 2,200 double-chain DNA virus (dsDNA) and retrovirus genomes deposited at the National Center for Biotechnology Information (NCBI), compared to more than 45,000 bacterial genomes [17].

The diversity and abundance imbalance of important viruses, such as bacteriophages, can cause important changes in the aquatic ecosystem, including the transduction of resistance genes between bacteria that can be mediated by some of these phages, which can provide evolutionary advantages to microorganisms and affect the water quality [18]. As for cyanophages, they are abundant in aquatic environments and play a fundamental role in flowering dynamics, including growth regulation and photosynthesis of cyanobacteria [19].

Therefore, our objective was to characterize the diversity of DNA viruses in Lake Bolonha, especially those that have bacteriophagic and cyanophagic behavior, and thus can contribute to the transduction of genes associated with resistance to antimicrobials and water quality.

2. Materials And Methods

2.1. *Sample collection*

The samples of water were collected in January of 2017 at Lake Bolonha, Belém-PA, in three different points of the lake before arriving at the local water treatment station (Fig. 1), according to Alves et al. (2020) [20]. The sampling sites along the lake took place at the water catchment area from the lake to the water treatment station (P1), at the local evacuation of water to other treatment substations (P2), and at the channel connecting the lakes Água Preta and Bolonha (P3). Alves et al. also performed the steps of water quality assessment, DNA extraction and metagenomic sequencing [20].

Figure 1. Map of the lake area and identification of sampling points (P1, P2 and P3). Source: Alves et al., 2020 [20].

2.2. *Metagenomic dataset*

The metagenomic raw data used to this work is currently available at SRA/NCBI database under the access numbers SRR8893560 (P1), SRR8893561 (P2) and SRR8893559 (P3) (Table 1).

Table 1
General information about the metagenomic dataset used in this work.

SRA accession number	Sample Code	# of Spots	# of Bases	Size	Study	Sequencing Technology
SRR8893560	P1	16,671,734	2.3 Gb	1.7 Gb	Freshwater Metagenome	Ion Proton™
SRR8893561	P2	16,278,151	2.2 Gb	1.7 Gb	Freshwater Metagenome	Ion Proton™
SRR8893559	P3	12,236,522	1.7 Gb	1.3 Gb	Freshwater Metagenome	Ion Proton™

2.3. Treatment of raw data

The raw data was used to perform the taxonomic analysis through Kraken2 tool [21] with the parameter '-download-library viral' to download the complete viral sequences of RefSeq and classify the reads regards its taxonomy.

2.4. Assessment of Viral Diversity

The output generated by Kraken2 were submitted to the tool Bracken [22], using abundance and diversity to generate more accurate estimations on the viruses genus and species levels. The input parameters were “\${CLASSIFICATION_LVL} = 'S' (Species)” and “input data = kraken2 output (report)”. All other parameters were set as default. Later, the results were displayed with the Pavian tool [23], which allows comparing the taxonomic classifications obtained by Kraken2 and Bracken, as well as presenting abundance estimations in several samples.

2.5. Viral Metagenome Assembly

The Genome Detective software [24] was used assembling the sequencing data and classifying the contigs formed in their respective taxons, identified using metaSPAdes software for single-end reads [25].

3. Results

3.1. Virome Assembly

The viral portion of the metagenomic raw data was assembled in order to prepare the data for the classification of their respective taxon per each of the freshwater samples (Table 2). The species with the highest percentage of coverage after assembly were Cyanophage KBS-S-2A (P1) with 25.37% and 69.66 identity, *Cladosporium fulvum T-1 virus* (P2) with 23.14% coverage and 56.90% identity, and Cyanophage KBS-S-2A (P3) with 20.65% coverage and 71.18% identity.

Table 2

List of samples (P1, P2, and P3) and their respective number of reads obtained after the sequencing, after quality control, the number of reads associated with viruses, and the number of reads used for coverage depth.

Sample Code	Raw data	After control quality	Reads associated with viruses	Average depth of assembly
P1	16,671,734 reads	8,410,372 reads	202,318 reads	5,359 reads
P2	16,278,151 reads	8,339,898 reads	44,877 reads	4,067 reads
P3	12,236,522 reads	6,299,043 reads	131,370 reads	2,217 reads

3.2. Viral Diversity of Lake Bolonha

Taxonomic analysis performed by Kraken2 revealed that there might be more than 3.500 distinct species of viruses in Lake Bolonha (Supplementary Table 1, Supplementary Table 2, and Supplementary Table 3). The 20 species with a higher count of reads in all of the freshwater samples are described in Table 3, Table 4 and Table 5. The number of reads found for each of the viruses contained on this top selection was compared with their respective number of reads on the other two samples from this study.

In general, the sample P1 presented a prevalent abundance of *Synechococcus* phage, a cyanophage (Fig. 2). We obtained the highest amount of reads for a single virus in sample P2 for *Choristoneura fumiferana granulovirus* (Fig. 3). The sample P2 also presented a much larger amount of the genus *Pandoravirus* and *Mimivirus* when compared to other collection sites. The presence of *Haemophilus virus HP1* was only observed on sample P3 (Fig. 4).

Table 3

Diversity of viruses with higher abundance of reads in sample site P1 (most abundant), and the abundance of the same species in P2 and P3 samples.

Species	Most abundant in P1	P2	P3
<i>Synechococcus phage S-SM2</i>	629	61	288
<i>Cyanophage S-RIM50</i>	587	40	852
<i>Synechococcus phage S-SKS1</i>	578	66	330
<i>Yellowstone lake phycod0virus 3</i>	515	60	201
<i>Synechococcus phage S-PM2</i>	489	50	285
<i>Escherichia virus D3</i>	410	126	285
<i>Synechococcus phage S-CAM9</i>	371	40	215
<i>Yellowstone lake phycod0virus 2</i>	322	39	179
<i>Cyanophage S-RIM32</i>	314	22	141
<i>Agrobacterium virus Atupho07</i>	309	36	363
<i>Synechococcus phage S-SM1</i>	288	30	142
<i>Prochlorococcus phage P-TIM68</i>	279	27	138
<i>Synechococcus phage S-RSM4</i>	278	22	149
<i>Acidovorax virus ACP17</i>	270	30	337
<i>Synechococcus phage S-WAM2</i>	267	21	190
<i>Synechococcus phage S-CAM1</i>	261	26	149
<i>Prochlorococcus phage P-SSM2</i>	255	37	122
<i>Synechococcus phage S-RIM8 A.HR1</i>	250	48	218
<i>Mycobacterium phage Crossroads</i>	247	19	328
<i>Yellowstone lake phycod0virus 1</i>	238	25	110

Table 4

Diversity of viruses with higher abundance of reads in sample site P2 (most abundant), and the abundance of the same species in P1 and P3 samples.

Species	Most abundant in P2	P1	P3
<i>Choristoneura fumera</i> 0 granulovirus	2476	187	83
<i>Pandoravirus inopi</i> 0tum	435	185	138
<i>Pandoravirus neocaledonia</i>	317	121	143
<i>Halovirus HF1</i>	207	60	0
<i>Pandoravirus macleodensis</i>	179	98	122
<i>Pandoravirus salinus</i>	161	142	83
<i>Alcelaphine gammaherpesvirus 1</i>	140	8	3
<i>Escherichia virus DE3</i>	126	410	285
<i>Apis mellifera filamentous virus</i>	126	25	22
<i>Bacillus virus Bp8pC</i>	114	11	10
<i>Cafeteria roenbergensis virus</i>	110	44	27
<i>Mimivirus terra 2</i>	108	67	33
<i>Euproctis pseudoconspersa nucleopolyhedrovirus</i>	100	7	1
<i>Common bottlenose dolphin gammaherpesvirus 1 strain Sarasota</i>	98	1	3
<i>Cydia pomonella granulovirus</i>	94	6	4
<i>Synechococcus phage ACG-2014h</i>	88	100	50
<i>Pandoravirus dulcis</i>	86	102	69
<i>Wenzhou picor</i> 0-like virus 4	85	5	3
<i>Pandoravirus quercus</i>	85	107	99
<i>Enterobacteria phage Sf101</i>	82	1	1

Table 5

Diversity of viruses with higher abundance of reads in sample site P3 (most abundant), and the abundance of the same species in P1 and P2 samples.

Species	Most abundant in P3	P1	P2
<i>Cyanophage S-RIM50</i>	852	587	40
<i>Haemophilus virus HP1</i>	418	0	0
<i>Agrobacterium virus Atuph07</i>	363	309	36
<i>Acidovorax virus ACP17</i>	337	270	30
<i>Synechococcus phage S-SKS1</i>	330	578	66
<i>Mycobacterium phage Crossroads</i>	328	247	19
<i>Synechococcus phage S-SM2</i>	288	629	61
<i>Synochococcus phage S-PM2</i>	285	489	50
<i>Escherichia virus DE3</i>	285	410	126
<i>Erwinia virus EaH2</i>	221	214	23
<i>Synechococcus phage S-RIM8 A.HR1</i>	218	250	48
<i>Synechococcus phage S-CAM9</i>	215	371	40
<i>Synechococcus phage ACG-2014e</i>	204	96	59
<i>Yellowstone lake phycod0virus 3</i>	201	515	60
<i>Salmonella virus SJ46</i>	192	54	5
<i>Synechococcus phage S-WAM2</i>	190	267	21
<i>Pseudomo0s phage phi3</i>	180	64	6
<i>Yellowstone lake phycod0virus 2</i>	179	322	39
<i>Synechococcus phage S-CAM3</i>	161	147	22
<i>Synechococcus phage S-CAM1</i>	149	261	26

Figure 4. Bar graph showing the distribution of reads for each top 20 viroplage found in sample P3, plotted against their respective representations on samples P1 and P2.

4. Discussion

When it comes to viral abundance, taxonomic analysis reveals that a plethora of more than 3.500 distinct species of viruses is present at freshwater Brazilian Lake Bolonha. It is possible to observe that the viruses with higher abundance of reads in P1 and P3 have a lower distribution among its P2, which can

be explained by the proximity between sampling sites P1 and P3 (Tables 3–5). The alpha diversity observed in the raw sequencing data of Table 1 of Alves et al. [20] also shows significant sample abundance in P1 and P3, what could also be related to this observation.

The sample P1 presented abundance of *Synechococcus* phage (Table 3), which is a type of phage that frequently infects cyanobacteria of the *Synechococcus* genus at diurnal rates of infection, due to the photosynthetic activities of its host [26]. This infection affects their population dynamics by killing part of this cyanobacteria population daily, estimated between 0.005% and 30% per day [19, 27, 28]. Additionally, it has been described that cyanophages play an important part in diversity and evolution of their host cyanobacteria [4, 17, 29–33].

Amongst the abundant species, the cyanophage S-RIM was found in the sample P1 and infects the cyanobacteria of the genus *Synechococcus* [34]. The abundance of cyanophage S-RIM50 has been reported in both fresh and seawater [12]. Cyanophages such as the cyanophage S-RIM50, as well as *Synechococcus* phage, are abundant in freshwater environments and have been isolated from a variety of freshwater reserves, including lakes, ponds, streams and sewage points [19, 26, 28]. They have an important contribution in the maintenance of the cyanobacterial community and in the preservation of water quality [11].

According to Alves et al. [20], the Amazonian vegetation on its shore characterizes The Lake Bolonha and the propagation of large plants under its surface, resulting in eutrophication. Also, it presents increased phosphorus and total nitrogen values in physical-chemical analysis and a high fecal coliform rate [20]. In this study, we noted the significant presence of cyanophages, mainly at P1 (Table 3) and P3 (Table 5). This finding contributes as an important indicator of the interference of these phages in the environment, known for their ability to perform photosynthesis by consuming oxygen and their potential for binding to nitrogen or production of toxins [31]. Nitrogen or phosphorus supplies, in addition to reduced growth rate and biomass, may naturally limit freshwater ecosystems, including those involving cyanobacteria accumulation [35].

Synechococcus phages, present in both samples P1 (Fig. 2) and P3 (Fig. 4), can be associated with the occurrence of health problems such as multiple sclerosis due to the protein expressed by the Epstein-Barr virus since it contains many short sequences identical to the products of 16 autoantigen related to multiple sclerosis susceptibility genes [36]. Other viruses associated with multiple sclerosis have also shown this behavior and the bacteriophage *Synechococcus* has been identified as a new and important contributor to this phenomenon. According to the distribution of multiple sclerosis, the cyanobacterial phage host prefers a temperate climate, and the ecology of bacteria and bacteriophages is consistent with the epidemiology of multiple sclerosis [34].

The presence of Yellowstone Lake *phycodnaviridae*, a double stranded DNA virus that infects algae, and *Escherichia* virus DE3 on the sample P1 were also observed (Fig. 2). A larger number of reads associated with *Shigella* phage SfIV was observed when compared to samples P2 and P3 (Table 3). It is important to mention a possible relationship with the presence of the *Shigella* bacterial host, responsible for causing

intestinal infection followed or not by fever, colic and diarrhea with blood and mucus [37]. This demonstrates the importance of this study in the characterization of the environmental conditions as a possible source of information for public health.

The P2 collection point presented a large amount of *Choristoneura fumiferana granulovirus* (Table 4), part of the *Baculoviridae* family [38]. It is interesting to observe a small number of reads (25 reads) associated with the species *Diplodia scrobiculata RNA virus 1* only in P2. The viral species has as preferable host the endophytic fungus *Diplodia scrobiculata*, which mostly affects the genus *Pinus* spp. among other conifers (Bihon et al., 2011). The presence of *Halovirus* HF1, responsible for infecting members of the *Halobacteriaceae* family [39], was also identified (Fig. 3).

A considerable amount of *Bacillus virus Bp8pC* has also been observed on sample P2 (Table 4), which hosts the bacteria *Bacillus thuringiensis* and *Bacillus pumilus*. Both are of economic importance because they are used as pest control in agriculture bringing little harm to humans [40]. The presence of the genus *Bacillus* in lake water may indicate the contamination of the water environment by different types of residues coming from the watershed to the lake [41].

Exclusively on sample P2, a much larger amount of the genus *Pandoravirus* and *Mimivirus* was observed in comparison to the other collection points (Fig. 3). Previous studies suggest a potential role of *Mimivirus* in respiratory pathology, displayed during seroconversion in patients with pulmonary pneumonia. In addition, positive serology for *Mimivirus* is associated with increased duration of mechanical ventilation supported breathing and intensive care unit in patients with ventilator-associated pneumonia [42]. Both genera are constituted of very large viruses and have Amoeba as their common host [43–45]. *Pandoravirus* has a size of about 1 micron and may resemble some types of bacteria. Their genome contains more than 100 distinct genes and can be of order twice as large as the *Mimivirus* genome, besides the fact that their genome is quite different compared to other known organisms [46].

Haemophilus virus HP1 was only observed at sample P3 (Fig. 4). This bacteriophage infects the *Haemophilus influenzae* bacterium [47], and its sampling site is located at the starting point of the channel connecting both Lakes Bolonha and Água Preta. It is important to observe that what occurs at this site may in the future influence the environment of Lake Água Preta.

Overall, these results denote the presence of a diverse viral community and suggest the existence of an established regulation dynamics in the local microbial environment of Lake Bolonha, highly influenced by the bacteriophages and cyanophages that inhabit the location. The dispersion of those biological entities along the water distribution channels using Lake Bolonha as a water source, as well as general eutrophic activity, might contribute to the spread of minor genetic elements like ARGs and future unbalance microenvironments of close by freshwater sources, such as Lake Água Preta.

5. Conclusions

Given the importance of Lake Bolonha as a source of drinking water supply for the metropolitan region of Belém, as well as being an area of environmental preservation, studies to elucidate the viral diversity of this environment are relevant in order to provide a better understanding on how its exploration can affect it. The results observed in this work indicates the presence of a widely diverse viral community, specially bacteriophages and cyanophages. These findings also suggest the existence of an established micro-environmental dynamics in Lake Bolonha, possibly regulated by such phage entities. The dispersion of those viral being bare similarity along the course of the lake, apparently more related the deeper they are into the lake (P1, P3) and the further away they are from the water evacuation sites to other treatment substations (P2). This is the first ever work to describe the virome of Lake Bolonha and, as such, contributes to the understanding of water-related public health concerns regarding the spreading of antibiotic resistance genes and population control of native bacteria and cyanobacteria.

Declarations

Ethics approval and consent to participate:

Not applicable.

Consent for publication:

Not applicable.

Availability of data and materials:

Publicly available datasets were analyzed in this study. This data can be found here:

<https://www.ncbi.nlm.nih.gov>, <https://www.ncbi.nlm.nih.gov/sra/PRJNA506429>.

Competing interests:

The authors declare that they have no competing interests.

Funding:

We thank the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES and PROPESP/UFGA (Pró-Reitoria de Pesquisa e Pós-Graduação / Universidade Federal do Pará), and the funding agencies FAPESPA (Fundação Amazônia de Amparo à Estudos e Pesquisas) and FAPEMIG (Fundação de Amparo à Pesquisa do Estado de Minas Gerais) for financial support on this work.

Authors' contributions:

BVAG, KCP, and RTJR designed the study. BVAG and KCP compiled and curated the data and performed bioinformatic analysis. BVAG, KCP, and WGN interpreted the results. ACF and RTJR supervised and administered the project and provided funding. BVAG wrote the original draft and manuscript with input

from KCP, WGN, AOA, ALCQ, ACF, and RTJR. All authors critically reviewed the manuscript and approved the final version.

Acknowledgments:

Not applicable.

Abbreviations

ARGs: Antibiotic resistance genes;

NGS: Next-Generation Sequencing;

dsDNA: double-chain DNA;

NCBI: National Center for Biotechnology Information.

References

- [1] A.E. Putt, E.A. Maclsaac, H.E. Herunter, A.B. Cooper, D.T. Selbie, Eutrophication forcings on a peri-urban lake ecosystem: Context for integrated watershed to airshed management, *PLoS One*. 14 (2019) e0219241. <https://doi.org/10.1371/journal.pone.0219241>.
- [2] A. Cook, S. Rushton, J. Allan, A. Baxter, An evaluation of techniques to control problem bird species on landfill sites, *Environ. Manage.* 41 (2008) 834–843. <https://doi.org/10.1007/s00267-008-9077-7>.
- [3] C.N. Palermo, R.R. Fulthorpe, R. Saati, S.M. Short, Metagenomic analysis of virus diversity and relative abundance in a eutrophic freshwater harbour, *Viruses*. 11 (2019) 792. <https://doi.org/10.3390/v11090792>.
- [4] M. Mohiuddin, H.E. Schellhorn, Spatial and temporal dynamics of virus occurrence in two freshwater lakes captured through metagenomic analysis, *Front. Microbiol.* 6 (2015). <https://doi.org/10.3389/fmicb.2015.00960>.
- [5] B. Koskella, M.A. Brockhurst, Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities, *FEMS Microbiol. Rev.* 38 (2014) 916–931. <https://doi.org/10.1111/1574-6976.12072>.
- [6] W. Calero-Cáceres, A. Melgarejo, M. Colomer-Lluch, C. Stoll, F. Lucena, J. Jofre, M. Muniesa, Sludge as a potential important source of antibiotic resistance genes in both the bacterial and bacteriophage fractions, *Environ. Sci. Technol.* 48 (2014) 7602–7611. <https://doi.org/10.1021/es501851s>.
- [7] M. Colomer-Lluch, W. Calero-Cáceres, S. Jebri, F. Hmaied, M. Muniesa, J. Jofre, Antibiotic resistance genes in bacterial and bacteriophage fractions of Tunisian and Spanish wastewaters as markers to

compare the antibiotic resistance patterns in each population, *Environ. Int.* 73 (2014) 167–175. <https://doi.org/10.1016/j.envint.2014.07.003>.

[8] I. Lekunberri, M. Villagrasa, J.L. Balcázar, C.M. Borrego, Contribution of bacteriophage and plasmid DNA to the mobilization of antibiotic resistance genes in a river receiving treated wastewater discharges, *Sci. Total Environ.* 601–602 (2017) 206–209. <https://doi.org/10.1016/j.scitotenv.2017.05.174>.

[9] E. Marti, E. Variatza, J.L. Balcázar, Bacteriophages as a reservoir of extended-spectrum β -lactamase and fluoroquinolone resistance genes in the environment, *Clin. Microbiol. Infect.* 20 (2014). <https://doi.org/10.1111/1469-0691.12446>.

[10] J. Subirats, A. Sànchez-Melsió, C.M. Borrego, J.L. Balcázar, P. Simonet, Metagenomic analysis reveals that bacteriophages are reservoirs of antibiotic resistance genes, *Int. J. Antimicrob. Agents.* 48 (2016) 163–167. <https://doi.org/10.1016/j.ijantimicag.2016.04.028>.

[11] H. Xia, T. Li, F. Deng, Z. Hu, Freshwater cyanophages, *Virol. Sin.* 28 (2013) 253–259. <https://doi.org/10.1007/s12250-013-3370-1>.

[12] M.S. Shane, Distribution of blue-green algal viruses in various types of natural waters, *Water Res.* 5 (1971) 711–716. [https://doi.org/10.1016/0043-1354\(71\)90093-5](https://doi.org/10.1016/0043-1354(71)90093-5).

[13] P. Singh, Structural, functional and molecular basis of cyanophage-cyanobacterial interactions and its significance, *AFRICAN J. Biotechnol.* 11 (2012) 2591–2608. <https://doi.org/10.5897/ajb10.790>.

[14] K.E. Wommack, D.J. Nasko, J. Chopyk, E.G. Sakowski, Counts and sequences, observations that continue to change our understanding of viruses in nature, *J. Microbiol.* 53 (2015) 181–192. <https://doi.org/10.1007/s12275-015-5068-6>.

[15] V. Aggarwala, G. Liang, F.D. Bushman, Viral communities of the human gut: Metagenomic analysis of composition and dynamics, *Mob. DNA.* 8 (2017) 1–10. <https://doi.org/10.1186/s13100-017-0095-y>.

[16] F.H. Coutinho, C.B. Silveira, G.B. Gregoracci, C.C. Thompson, R.A. Edwards, C.P.D. Brussaard, B.E. Dutilh, F.L. Thompson, Marine viruses discovered via metagenomics shed light on viral strategies throughout the oceans, *Nat. Commun.* 8 (2017) 1–12. <https://doi.org/10.1038/ncomms15955>.

[17] D. Paez-Espino, E.A. Eloë-Fadrosh, G.A. Pavlopoulos, A.D. Thomas, M. Huntemann, N. Mikhailova, E. Rubin, N.N. Ivanova, N.C. Kyrpides, Uncovering Earth's virome, *Nature.* 536 (2016) 425–430. <https://doi.org/10.1038/nature19094>.

[18] H.T.P. Williams, Phage-induced diversification improves host evolvability, *BMC Evol. Biol.* 13 (2013) 17. <https://doi.org/10.1186/1471-2148-13-17>.

[19] X. Jiang, C. Ha, S. Lee, J. Kwon, H. Cho, T. Gorham, J. Lee, Characterization of cyanophages in lake erie: Interaction mechanisms and structural damage of toxic cyanobacteria, *Toxins (Basel).* 11 (2019)

444. <https://doi.org/10.3390/toxins11080444>.

[20] J. Alves, L. Dias, J. Mateus, J. Marques, D. Graças, R. Ramos, L. Seldin, I. Henriques, A. Silva, A. Folador, Resistome in Lake Bolonha, Brazilian Amazon: Identification of Genes Related to Resistance to Broad-Spectrum Antibiotics, *Front. Microbiol.* 11 (2020). <https://doi.org/10.3389/fmicb.2020.00067>.

[21] D.E. Wood, J. Lu, B. Langmead, Improved metagenomic analysis with Kraken 2, *Genome Biol.* 20 (2019). <https://doi.org/10.1186/s13059-019-1891-0>.

[22] J. Lu, F.P. Breitwieser, P. Thielen, S.L. Salzberg, Bracken: Estimating species abundance in metagenomics data, *PeerJ Comput. Sci.* 2017 (2017) e104. <https://doi.org/10.7717/peerj-cs.104>.

[23] F.P. Breitwieser, S.L. Salzberg, Pavian: Interactive analysis of metagenomics data for microbiome studies and pathogen identification, *Bioinformatics.* 36 (2020) 1303–1304. <https://doi.org/10.1093/bioinformatics/btz715>.

[24] M. Vilsker, Y. Moosa, S. Nooij, V. Fonseca, Y. Ghysens, K. Dumon, R. Pauwels, L.C. Alcantara, E. Vanden Eynden, A.M. Vandamme, K. Deforche, T. De Oliveira, Genome Detective: An automated system for virus identification from high-throughput sequencing data, *Bioinformatics.* 35 (2019) 871–873. <https://doi.org/10.1093/bioinformatics/bty695>.

[25] S. Nurk, D. Meleshko, A. Korobeynikov, P.A. Pevzner, MetaSPAdes: A new versatile metagenomic assembler, *Genome Res.* 27 (2017) 824–834. <https://doi.org/10.1101/gr.213959.116>.

[26] R. Liu, Y. Liu, Y. Chen, Y. Zhan, Q. Zeng, Cyanobacterial viruses exhibit diurnal rhythms during infection, *Proc. Natl. Acad. Sci. U. S. A.* 116 (2019) 14077–14082. <https://doi.org/10.1073/pnas.1819689116>.

[27] L. Bragg, G.W. Tyson, Metagenomics using next-generation sequencing, *Methods Mol. Biol.* 1096 (2014) 183–201. https://doi.org/10.1007/978-1-62703-712-9_15.

[28] D. Toyama, L.T. Kishi, C.D. Santos-Júnior, A. Soares-Costa, T.C. Souza de Oliveira, F.P. de Miranda, F. Henrique-Silva, Metagenomics analysis of microorganisms in freshwater lakes of the Amazon Basin, *Genome Announc.* 4 (2016) 1440–1456. <https://doi.org/10.1128/genomeA.01440-16>.

[29] J. Handelsman, Metagenomics: Application of Genomics to Uncultured Microorganisms, *Microbiol. Mol. Biol. Rev.* 68 (2004) 669–685. <https://doi.org/10.1128/mubr.68.4.669-685.2004>.

[30] P. Lorenz, J. Eck, Metagenomics and industrial applications, *Nat. Rev. Microbiol.* 3 (2005) 510–516. <https://doi.org/10.1038/nrmicro1161>.

[31] A. Jaskulska, J. Mankiewicz-Boczek, Cyanophages specific to cyanobacteria from the genus *Microcystis*, *Ecohydrol. Hydrobiol.* 20 (2020) 83–90. <https://doi.org/10.1016/j.ecohyd.2019.06.001>.

- [32] J.G. Tundisi, Água no século 21: enfrentando a escassez, 2003.
https://www.researchgate.net/publication/266878577_Agua_No_seculo_XXI_Enfrentando_a_Escassez
(accessed August 6, 2020).
- [33] A. Fedida, D. Lindell, Two synechococcus genes, two different effects on cyanophage infection, *Viruses*. 9 (2017) 136. <https://doi.org/10.3390/v9060136>.
- [34] L.I. Stoddard, J.B.H. Martiny, M.F. Marston, Selection and characterization of cyanophage resistance in marine *Synechococcus* strains, *Appl. Environ. Microbiol.* 73 (2007) 5516–5522.
<https://doi.org/10.1128/AEM.00356-07>.
- [35] K. Cheng, T. Frenken, C.P.D. Brussaard, D.B. Van De Waal, Cyanophage propagation in the freshwater cyanobacterium *phormidiumis* constrained by phosphorus limitation and enhanced by elevated pCO₂, *Front. Microbiol.* 10 (2019). <https://doi.org/10.3389/fmicb.2019.00617>.
- [36] C.J. Carter, Epstein-Barr and other viral mimicry of autoantigens, myelin and vitamin D-related proteins and of EIF2B, the cause of vanishing white matter disease: Massive mimicry of multiple sclerosis relevant proteins by the *Synechococcus* phage, *Immunopharmacol. Immunotoxicol.* 34 (2012) 21–35. <https://doi.org/10.3109/08923973.2011.572262>.
- [37] F.J. Martinez-Becerra, X. Chen, N.E. Dickenson, S.P. Choudhari, K. Harrison, J.D. Clements, W.D. Picking, L.L. Van De Verg, R.I. Walker, W.L. Picking, Characterization of a novel fusion protein from IpaB and IpaD of *Shigella* spp. and its potential as a Pan-*Shigella* vaccine, *Infect. Immun.* 81 (2013) 4470–4477. <https://doi.org/10.1128/IAI.00859-13>.
- [38] R.J. Clem, A.L. Passarelli, Baculoviruses: Sophisticated Pathogens of Insects, *PLoS Pathog.* 9 (2013) e1003729. <https://doi.org/10.1371/journal.ppat.1003729>.
- [39] A. Ventosa, *Halophilic Microorganisms*, Springer Berlin Heidelberg, 2004.
<https://doi.org/10.1007/978-3-662-07656-9>.
- [40] G. Stotzky, Persistence and Biological Activity in Soil of Insecticidal Proteins from *Bacillus thuringiensis* and of Bacterial DNA Bound on Clays and Humic Acids, *J. Environ. Qual.* 29 (2000) 691–705. <https://doi.org/10.2134/jeq2000.00472425002900030003x>.
- [41] S. Shafi, A.N. Kamili, M.A. Shah, S.A. Bandh, R. Dar, Dynamics of bacterial class Bacilli in the deepest valley lake of Kashmir-the Manasbal Lake, *Microb. Pathog.* 104 (2017) 78–83.
<https://doi.org/10.1016/j.micpath.2017.01.018>.
- [42] M.J. Vanspauwen, F.M.E. Franssen, D. Raoult, E.F.M. Wouters, C.A. Bruggeman, C.F.M. Linssen, Infections with mimivirus in patients with chronic obstructive pulmonary disease, *Respir. Med.* 106 (2012) 1690–1694. <https://doi.org/10.1016/j.rmed.2012.08.019>.

- [43] M. Legendre, E. Fabre, O. Poirot, S. Jeudy, A. Lartigue, J.M. Alempic, L. Beucher, N. Philippe, L. Bertaux, E. Christo-Foroux, K. Labadie, Y. Couté, C. Abergel, J.M. Claverie, Diversity and evolution of the emerging Pandoraviridae family, *Nat. Commun.* 9 (2018) 1–12. <https://doi.org/10.1038/s41467-018-04698-4>.
- [44] B. La Scola, S. Audic, C. Robert, L. Jungang, X. De Lamballerie, M. Drancourt, R. Birtles, J.M. Claverie, D. Raoult, A giant virus in amoebae, *Science* (80-.). 299 (2003) 2033. <https://doi.org/10.1126/science.1081867>.
- [45] D. Raoult, S. Audic, C. Robert, C. Abergel, P. Renesto, H. Ogata, B. La Scola, M. Suzan, J.M. Claverie, The 1.2-megabase genome sequence of Mimivirus, *Science* (80-.). 306 (2004) 1344–1350. <https://doi.org/10.1126/science.1101485>.
- [46] C. Abergel, M. Legendre, J.M. Claverie, The rapidly expanding universe of giant viruses: Mimivirus, Pandoravirus, Pithovirus and Mollivirus, *FEMS Microbiol. Rev.* 39 (2015) 779–796. <https://doi.org/10.1093/femsre/fuv037>.
- [47] J.M. Hakimi, J.J. Scocca, Purification and characterization of the integrase from the *Haemophilus influenzae* bacteriophage HP1; identification of a four-stranded intermediate and the order of strand exchange, *Mol. Microbiol.* 21 (1996) 147–158. <https://doi.org/10.1046/j.1365-2958.1996.6311351.x>.

Figures

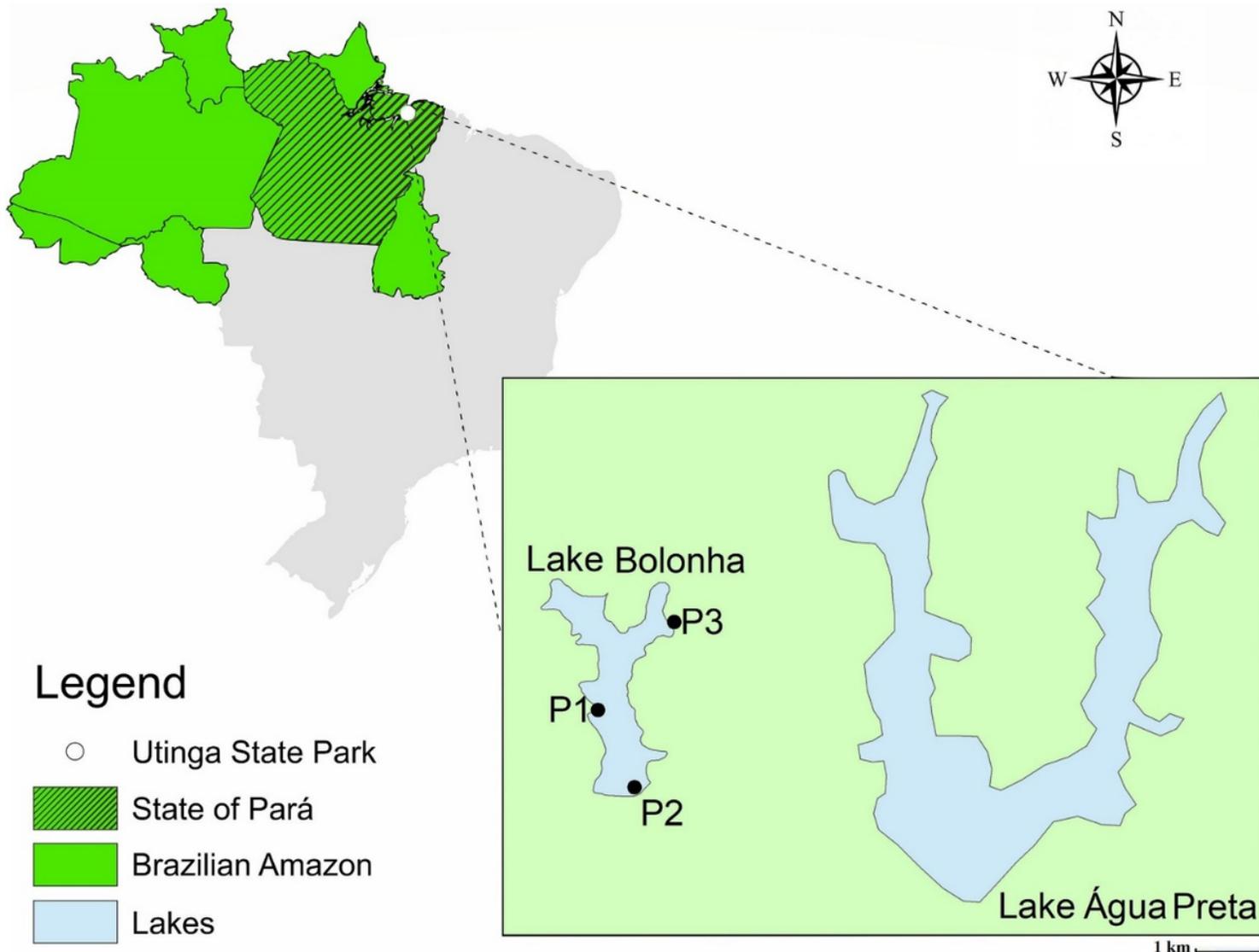


Figure 1

Map of the lake area and identification of sampling points (P1, P2 and P3). Source: Alves et al., 2020 [20].

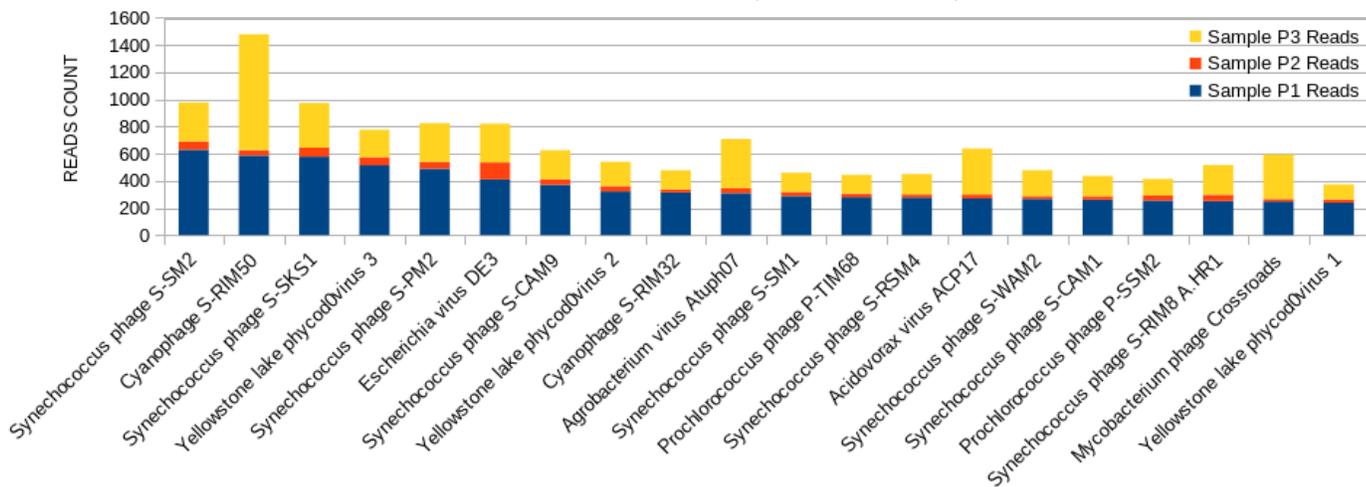


Figure 2

Bar graph showing the distribution of reads for each top 20 virophage found in sample P1, plotted against their respective representations on samples P2 and P3.

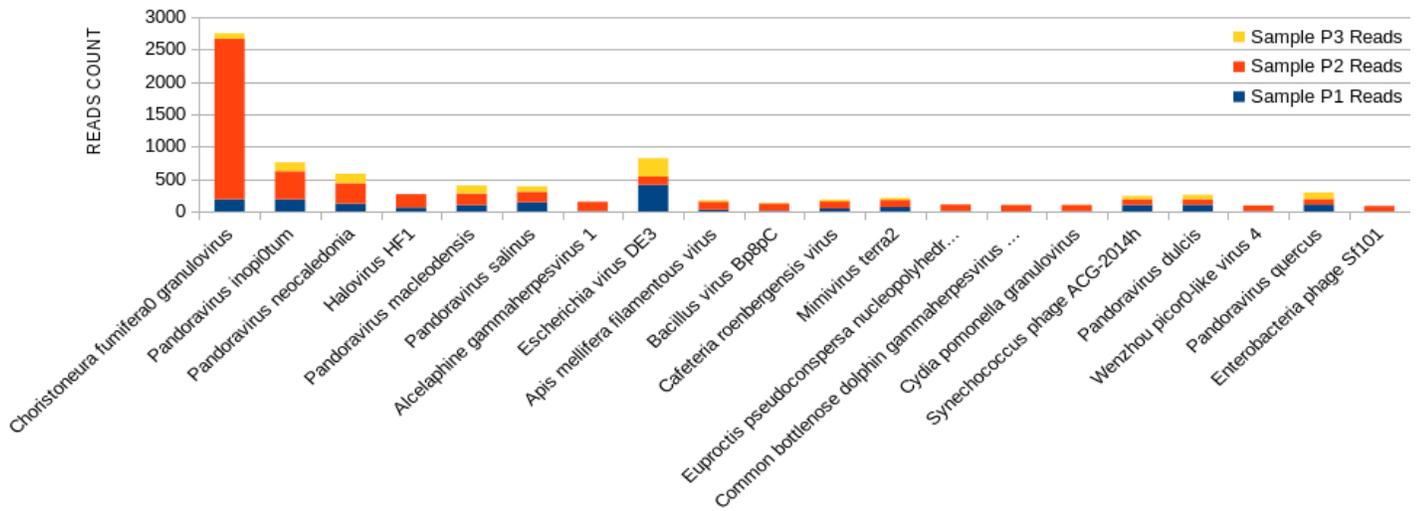


Figure 3

Bar graph showing the distribution of reads for each top 20 virophage found in sample P2, plotted against their respective representations on samples P1 and P3.

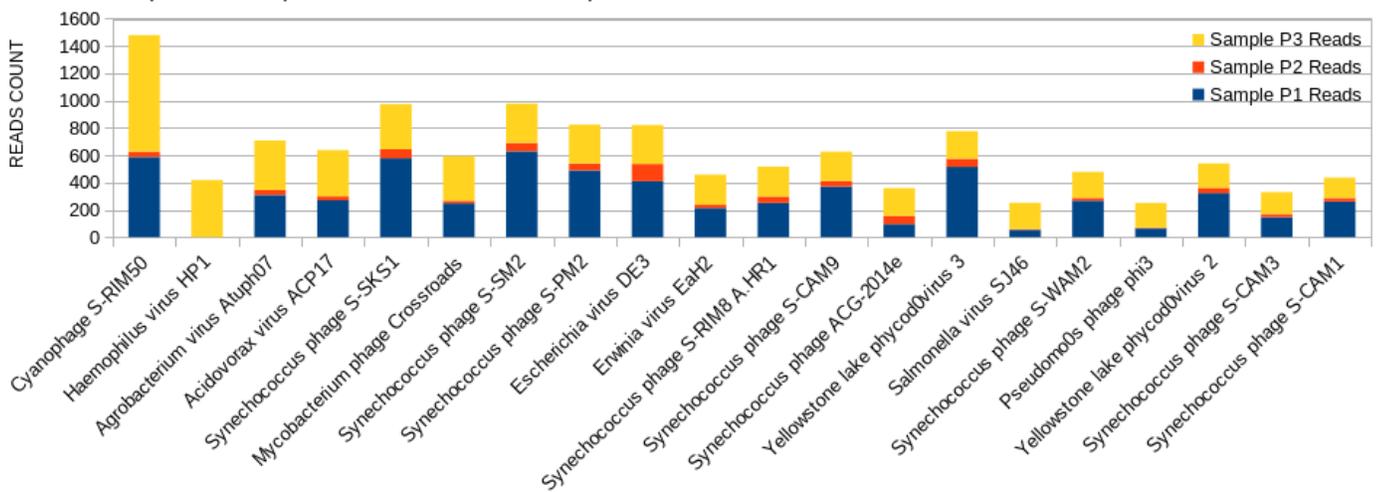


Figure 4

Bar graph showing the distribution of reads for each top 20 virophage found in sample P3, plotted against their respective representations on samples P1 and P2.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplementarymaterialTable1.xlsx](#)
- [supplementarymaterialTable2.xlsx](#)

- [supplementarymaterialTable3.xlsx](#)