

Book Chapter

Active Crossfire between Cyanobacteria and Cyanophages in Phototrophic Mat Communities within Hot Springs

Sergio Guajardo-Leiva¹, Carlos Pedrós-Alió², Oscar Salgado¹, Fabián Pinto¹ and Beatriz Díez^{1,3*}

¹Department of Molecular Genetics and Microbiology, Pontificia Universidad Católica de Chile, Chile

²Programa de Biología de Sistemas, Centro Nacional de Biotecnología – Consejo Superior de Investigaciones Científicas, Spain

³Center for Climate and Resilience Research, Chile

***Corresponding Author:** Beatriz Díez, Department of Molecular Genetics and Microbiology, Pontificia Universidad Católica de Chile, Santiago, Chile

Published **March 27, 2020**

This Book Chapter is a republication of an article published by Beatriz Díez, et al. at *Frontiers in Microbiology* in September 2018. (Guajardo-Leiva S, Pedrós-Alió C, Salgado O, Pinto F and Díez B (2018) Active Crossfire Between Cyanobacteria and Cyanophages in Phototrophic Mat Communities Within Hot Springs. *Front. Microbiol.* 9:2039. doi: 10.3389/fmicb.2018.02039)

How to cite this book chapter: Sergio Guajardo-Leiva, Carlos Pedrós-Alió, Oscar Salgado, Fabián Pinto, Beatriz Díez. Active Crossfire between Cyanobacteria and Cyanophages in Phototrophic Mat Communities within Hot Springs. In: Alexandre Morrot, editor. *Prime Archives in Microbiology*. Hyderabad, India: Vide Leaf. 2020.

© The Author(s) 2020. This article is distributed under the terms of the Creative Commons Attribution 4.0 International

License(<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Data Availability Statement: The datasets generated for this study can be found NCBI as follow: Access to raw data for metagenomes and metatranscriptomes is available through NCBI BioProject ID PRJNA382437.

<https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA382437>.

The genome of TC-CHP58 has the GenBank accession number KY888885. Contigs containing CRISPRs *loci* have been submitted to NCBI with GenBank accession numbers MG734911 to MG734917.

Author Contributions: SG-L and BD conceived and designed the experiments. SG-L, OS, and FP performed the experiments. SG-L, CP-A, OS, FP, and BD analyzed the data. SG-L, CP-A, and BD wrote the paper.

Funding: This work was financially supported by Ph.D. scholarships CONICYT N° 21130667 and 21172022, and CONICYT grant FONDECYT N°1150171. Sequencing was funded by Spanish grant CTM2013-48292-C3-1-R.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgments: We are grateful to Huinay Scientific Field Station for making our work in the Porcelana hot spring possible.

Abstract

Cyanophages are viruses with a wide distribution in aquatic ecosystems, that specifically infect Cyanobacteria. These viruses can be readily isolated from marine and fresh waters environments; however, their presence in cosmopolitan thermophilic phototrophic mats remains largely unknown. This study investigates the morphological diversity (TEM),

taxonomic composition (metagenomics), and active infectivity (metatranscriptomics) of viral communities over a thermal gradient in hot spring phototrophic mats from Northern Patagonia (Chile). The mats were dominated (up to 53%) by cosmopolitan thermophilic filamentous true-branching cyanobacteria from the genus *Mastigocladus*, the associated viral community was predominantly composed of Caudovirales (70%), with most of the active infections driven by cyanophages (up to 90% of Caudovirales transcripts). Metagenomic assembly lead to the first full genome description of a T7-like Thermophilic Cyanophage recovered from a hot spring (Porcelana Hot Spring, Chile), with a temperature of 58°C (TC-CHP58). This could potentially represent a world-wide thermophilic lineage of podoviruses that infect cyanobacteria. In the hot spring, TC-CHP58 was active over a temperature gradient from 48 to 66°C, showing a high population variability represented by 1979 single nucleotide variants (SNVs). TC-CHP58 was associated to the *Mastigocladus* spp. by CRISPR spacers. Marked differences in metagenomic CRISPR *loci* number and spacers diversity, as well as SNVs, in the TC-CHP58 proto-spacers at different temperatures, reinforce the theory of co-evolution between natural virus populations and cyanobacterial hosts. Considering the importance of cyanobacteria in hot spring biogeochemical cycles, the description of this new cyanopodovirus lineage may have global implications for the functioning of these extreme ecosystems.

Keywords

Hot-Springs; Cyanophages; Phototrophic Microbial Mat; CRISPR; Thermophilic Cyanobacteria

Introduction

Hot springs host microbial communities dominated by a limited variety of microorganisms that form well-defined mats [1,2]. Frequently, the uppermost layer of the mat is composed of photoautotrophs; such as oxygenic phototrophic cyanobacteria, including the unicellular cyanobacterium *Synechococcus* spp. [3-6], the filamentous non-heterocystous *Oscillatoria* spp., the

filamentous heterocystous *Mastigocladus* spp. [7-10], as well as filamentous anoxygenic phototrophs (FAPs), such as *Roseiflexus* sp. and *Chloroflexus* sp. [6,11,12]. These primary producers interact with heterotrophic prokaryotes through element and energy cycling [13]. Heterocystous cyanobacteria are a key component in hot springs, since these systems are commonly N-limited due to the rapid assimilation and turnover of inorganic nitrogen forms [10,14]. Thus, N₂-fixation by cyanobacteria is identified to be a key biological process in neutral hot spring microbial mats [10].

These simplified but highly cooperative communities have been historically used as models for understanding the composition, structure, and function of microbial consortia [2,6]. The role of a variety of abiotic factors, such as pH, sulfide concentration, and temperature, in determining microbial assemblages and life cycles in these ecosystems have been investigated [2,15]. However, there is a lack of investigation into biotic factors, such as viruses, on thermophilic photoautotrophic mats, with existing studies only reporting short or partial viral sequences [16,17]. Currently, viral communities from thermal mats have been characterized through indirect approaches, indicating the hypothetical presence of viruses [16,17]. Heidelberg et al. [16] used CRISPR spacer sequences extracted from the genomes of two thermophilic *Synechococcus* isolates, from a phototrophic mat in Octopus Spring. Subsequently, they searched for viral contigs from previously published water metaviromes from the Octopus and Bear Paw Springs in Yellowstone National Park (United States) [18]. Furthermore, Davison et al. used CRISPR spacers and nucleotide motive frequencies to link viral contigs to known hosts using a metavirome obtained by Multiple Displacement Amplification (MDA) of VLPs from a mat in Octopus Spring [17], as well as reference genomes from dominant species (*Synechococcus* sp., *Roseiflexus* sp., and *Chloroflexus* sp.) previously described in the same microbial mat. A key finding from these studies was the link between viruses and their hosts, indicating their co-evolution and an effective “arms race” within hot spring phototrophic mats.

Unlike thermophilic mat studies, most viral investigation carried out in hot springs occur within the source waters [18-23]. In these waters, virus abundances range between 10^4 and 10^9 virus like particles (VLPs) mL^{-1} [18,24,25]. They play an important role in both the structuring of host populations and as drivers of organic and inorganic nutrient recycling [24]. The majority of the viruses were dsDNA, with new and complex viral morphotypes, distinct to the typical head and tail morphologies [18,19,25-27]. Furthermore, the few metaviromes obtained in thermal waters indicate that natural thermophilic virus communities differ from those obtained in culture, given that there was only a 20–50% similarity between the sequences obtained and those in the databases [18,22,28,29]. Thus far, the genomes that have been isolated and sequenced from thermophilic viruses (57 genomes, of which 37 infected archaea and 20 infected Bacteria) generally yielded few significant matches to sequences in public databases [1]. More recently, a water metaviromic study from Brandvlei hot spring (BHS), South Africa [23] reported the presence of two partial genomes (10 kb and 27 kb), the first related to Podoviridae and the second to lambda-like Siphoviridae families. Both Caudovirales genomes did not have a confirmed host, but the presence of green microbial mat-patches around the contours of the hot spring, implied that filamentous Cyanobacteria and unclassified *Gemmata* species were the potential hosts, respectively. The last, based on the proximity of some viral predicted proteins with bacteria from well characterized microbial mats present in a nearby hot spring [30,31].

Given the lack of knowledge of viral communities within hot spring phototrophic microbial mats, the present study used the mats of Porcelana hot spring (Northern Patagonia, Chile), as a pH neutral model, to better understand the associated thermophilic viral communities within these mats. This pristine spring is covered by microbial mats that grow along a thermal gradient between 70 and 46°C, dominated by bacterial phototrophs, such as filamentous cyanobacteria from the genus *Mastigocladus* [9,10]. This is the dominant and most active cyanobacterial genus in the Porcelana mat environment, carrying out important biological processes such as carbon- and N_2 -

fixation [10,32]. Thus, this study proposes that the mats in Porcelana hot spring are dominated by viral communities of the Order Caudovirales, which is able to infect Cyanobacteria, preferably *Mastigocladus* spp.

The viral diversity in Porcelana was determined through the detection of viral signals in microbial mat omics data, and by TEM along the thermal gradient. The results demonstrate that the viral community was dominated by Caudovirales, which actively infect Cyanobacteria. Furthermore, the first complete genome description of a thermophilic cyanobacterial T7-like podovirus, Thermophilic Cyanophage Chile Porcelana 58°C (from now on TC-CHP58) is realized. The host is the dominant phototroph *Mastigocladus* spp, based on CRISPR spacers. Finally, the presence of different populations of this new podovirus are identified through single nucleotide variants (SNVs) analyses, and the co-evolution of *Mastigocladus* spp. and particular populations of TC-CHP58 at different temperatures is described through association of specific SNVs to different CRISPR spacers.

Materials and Methods

Sampling Site

Porcelana hot spring is located in Chilean Patagonia (42° 27' 29.1"S – 72° 27' 39.3"W). It has a neutral pH range between 7.1 and 6.8 and temperatures ranging from 70 to 46°C, when sampled on March 2013. Phototrophic microbial mats growing at 66, 58, and 48°C were sampled using a cork borer of 7 mm diameter. Cores of 1 cm thick were collected in triplicate at noon (12:00 PM), transported in liquid nitrogen and kept at -80°C until DNA and RNA extraction.

Transmission Electron Microscopy

Five liters of interstitial fluid was squeezed using 150 µm sterilized polyester net SEFAR PET 1000 (Sefar, Heiden, Switzerland) and filtered through 0.8 µm pore-size polycarbonate filters (Isopore ATTP, 47 mm diameter, Millipore, Millford, MA, United States) and 0.2 µm pore-size (Isopore

GTTP, 47 mm diameter, Millipore) using a Swinex filter holder (Millipore). Particles in the 0.2 μm filtrate were concentrated to a final volume of approximately 35 ml using a tangential-flow filtration cartridge (Vivaflow 200, 30 kDa pore size, Vivascience, Lincoln, United Kingdom). Viral concentrates (15 μL) were spotted onto Carbon Type-B, 200 mesh, Copper microscopy grids (Ted Pella, Redding, California, United States), stained with 1% uranyl acetate and imaged on an FEI Tecnai T12 electron microscope at 80 kV (FEI Corporate, Hillsboro, OR, United States) with attached Megaview G2 CCD camera (Olympus SIS, Münster, Germany). Imaging analysis was done at the Advanced Microscopy Unit, School of Biological Sciences at Pontificia Universidad Católica de Chile (Santiago, Chile).

Nucleic Acid Extractions and High Throughput Sequencing

Nucleic acids (DNA and RNA) were extracted as previously described [10]. For RNA, Trizol (Invitrogen, Carlsbad, CA, United States) was added to the mat sample, and homogenized by bead beating, two pulses of 20 s. Quality and quantity of the extracted nucleic acids were checked and kept at -80°C .

Samples were sequenced by Illumina Hi-seq technology (Research and Testing Laboratory, Texas, United States). Briefly, for metagenomes, DNA was fragmented using NEBNext dsFragmentase (New England Biolabs, Ipswich, MA, United States), followed by DNA clean up using column purification, and a NEBUltra DNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, MA, United States) was used for library construction.

For metatranscriptomes, DNase treated total RNA was cleaned up of rRNA by a Ribo-Zero rRNA Removal Kit Bacteria (Illumina, San Diego, CA, United States), followed by purification using an Agencourt RNAClean XP Kit (Beckman Coulter, Indianapolis, IN, United States), and a NEXTFlex™ Illumina Small RNA Sequencing Kit v3 (Bio Scientific, Austin, TX, United States) was used for library construction.

For quality filtering, the following filters were applied using Cutadapt [33], leaving only mappable sequences longer than 30 bp (-m 30), with a 3' end trimming for bases with a quality below 28 (-q 28), a hard clipping of the first five leftmost bases (-u 5), and finally a perfect match of at least 10 bp (-O 10) against the standard Illumina adaptor. Finally, the removal of sequences representing simple repetitions that are usually due to sequencing errors was applied using PRINSEQ [34] DUST threshold 7 (-lc_method dust, -lc_threshold 7). Details of the number of sequences obtained are shown in Supplementary Table S1.

Identification of rRNA-Like Sequences and Viral Mining From Metagenomes and Metatranscriptomes

Metagenomic Illumina TAGs (miTAGs) [35] that are small subunit (SSU) 16S and 18S rRNA gene sequences in the metagenomes were identified and annotated using the Ribopicker tool [36] with the Silva 123 SSU database [37].

For viral mining, bacterial, archaeal and eukaryotic sequences were removed through end-to-end mapping, allowing a 5% of mismatch (-N 1 -L 20) against the NCBI non-redundant (NR) database (Nov-2015) using bowtie2 [38]. Viral sequences were then recruited against modified NCBI RefSeq (Release 75) viral proteins, where only amino acid sequences from viruses that do not infect animals (NAV) were considered to build the database, using the UBLAST algorithm (-strand both -accel 0.9) through the USEARCH sequence analysis tool [39]. Recruitment was made for sequences with over 65% of coverage and an E-value $< 1 \times 10^{-3}$ (-query_cov 0.65 -evalue 1e-3). For taxonomic assignment, recruited sequences were aligned against the NAV database using BLASTX [40] and parsed using the lowest common ancestor algorithm through MEGAN 6 [41] (LCA score = 30). The latter displays a graphical representation of abundance for each taxonomic group identified at the family and species levels. Species classification of viral reads was used to infer the phyla of the putative hosts based on viral RefSeq host information or through a manual search of the publication associated with each viral genome.

To extract putative viral genomes, all metagenomes (48, 58, and 66°C) were assembled using De Bruijn graphs as implemented in the Spades assembler [42], followed by gene prediction using Prodigal software [43] and the recovery of circular contigs over 5 kb using a Python script [44]. Only sequences over 5 kb were used in the subsequent analysis because all dsDNA viruses in the databases have genomes over that size. A homology search of the viral predicted proteins by Prokka [45] was done using BLASTX against the NAV protein database and NCBI nr as described before. Additionally, all contigs over 5 kb were analyzed using VirSorter [46] against the virome database option.

To quantify the abundance and activity of the retrieved viral genome, reads recruitment from each metagenome and metatranscriptome was performed using BWA-MEM (-M), resulting SAM file was parsed by BBmap pileup script (Bushnell B.)¹.

Phylogenetic Analysis

The protein inferred sequences of DNA polymerase and major capsid were aligned by Muscle [47] and MAFFT [48], respectively, using the amino acid substitution model determined by ProtTest 3 (Blosum62+G+F) [49] and modelFinder (LG+F+G4), respectively. The Bayesian Markov chain Monte Carlo method was implemented with MrBayes 3.6 [50] and MCMC results were summarized with Tracer 1.6². MrBayes was run using two independent runs, four chains, 1,500,000 generations and a sampling frequency of 100 with a burn-in value of 33% until the standard deviations of split frequencies remained below 0.01.

The maximum likelihood method was implemented with IQtree (-bb 10000 -nm 10000 -bcor 1 -numstop 1000) [51] using 100 standard bootstrap and 10,000 ultrafast bootstrap to evaluate branch supports. The details of the sequences used for phylogenetic analyses are listed in Supplementary Table S2.

CRISPR/Cas Virotopes

Assemblies for each temperature, were taxonomically grouped (bins) using the Expectation–Maximization (EM) algorithm implemented in MaxBin 2.0 [52]. In order to assess the completeness and contamination of each bin, CheckM [53] analyses were performed. Finally, the closest genome of each bin was searched using the Tetra Correlation Search (TCS) analysis implemented in Jspecies tool [54] with selection criteria of Z score greater than 0.999 and ANI over 95% [55].

CRISPR/Cas *loci* were identified in contigs assigned to *Mastigocladus* spp. from 48, 58, and 66°C assembled metagenomes using CRISPRFinder tool [56]. To quantify the activity of the CRISPR *loci*, reads recruitment from metatranscriptomes for the same temperatures was performed using BWA-MEM (-M), and the resulting SAM file was parsed by BBmap pileup script (Bushnell B.) (see footnote 1) and normalized by total number of reads and length of each *loci*.

Spacers from CRISPR containing contigs were mapped to viral contigs using bowtie2 [38] parameters (-end-to-end -very sensitive -N 1). Mapped spacers were manually annotated to the viral predicted proteins in viral contig.

Single Nucleotide Variants (SNVs)

To call variants occurring in TC-CHP58 populations at the three different metagenome temperatures, LoFreq method [57] was used. SNVs frequencies were quantified in ORFs from TC-CHP58 genome using Bedtools suite [58]. The alleles of SNVs present in proto-spacers were visualized in IGV tools for each virotope at each temperature.

Results

Morphological and Genetic Composition of VLPs

Transmission electron microscopy (TEM) was applied to identify the VLPs present in the interstitial fluid from microbial mats in Porcelana hot spring. Caudovirus-like particles

belonging to Myoviridae, Podoviridae and Siphoviridae families, typically infecting bacteria (Figures 1A–G) were identified. Additionally, filamentous and rod shaped VLPs were detected, that could be associated with Lipothrixviridae and Clavaviridae families, usually infecting archaea (Figures 1H–K). Viral read counts ranged between 0.47 and 0.78% of the total metagenome reads, and between 0.35 and 3.71% in the metatranscriptomes (Supplementary Table S1). At all temperatures, viral metagenomic sequences (Figure 2) revealed the dominance of the Order Caudovirales, followed by the Order Megavirales, with ~70% and ~23% of the total viral reads, respectively. Metatranscriptomic analysis results (Figure 2) showed a slightly different pattern, with a reduction in Caudovirales with increasing temperature (from ~78% at 48°C to ~57% at 66°C), whereas Megavirales did the opposite (from ~7% at 48°C to ~36% at 66°C).

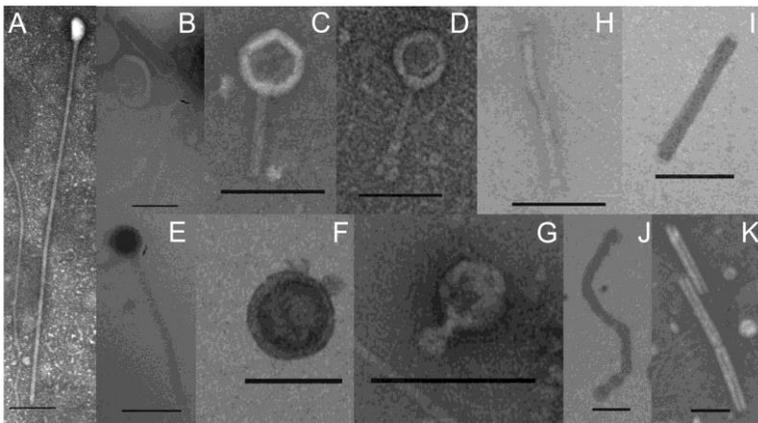


Figure 1: Transmission electronic micrographs of VLPs obtained from the interstitial fluid of phototrophic microbial mats growing between 62°C and 42°C in Porcelana hot spring. Scale bar: 100 nm. (A–G) Caudovirus-like particles belonging to Myoviridae, Podoviridae, and Siphoviridae families. (H–K) Filamentous and rod shaped VLPs that could be associated with Lipothrixviridae and Clavaviridae families.

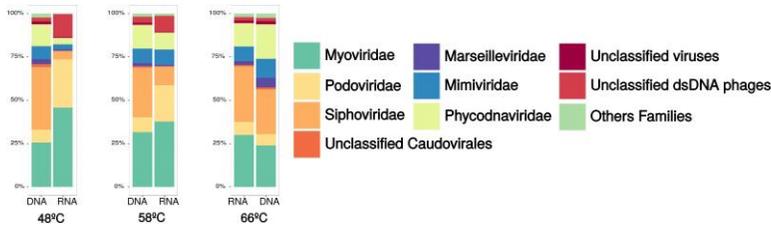


Figure 2: Relative abundances of viral Families in microbial mats from Porcelana hot spring; standardized by the total number metagenomic reads (DNA), and metatranscriptome (RNA) from each temperature samples.

In the metagenomes, Siphoviridae was the most abundant family of Caudovirales, with maximum abundance at 48°C. Myoviridae members were also well represented with a maximum of ~31% at 58°C and a minimum (~25%) at 48°C. Meanwhile, Podoviridae accounted for just ~8% at all temperatures (Figure 2). In metatranscriptomes, Siphoviridae increased sixfold with temperature, while Podoviridae and Myoviridae decreased with temperature (fivefold and twofold, respectively).

The Megavirales order was also present, however, at a lower abundance compared to Caudovirales. Megavirales were represented by Phycodnaviridae (~13%), Mimiviridae (~8%), and Marseilleviridae (~2%) families, remaining constant through all temperatures. Metatranscriptomics showed an increase in abundance of these three virus families with temperature.

Caudovirales Host Assignments

Porcelana mat communities based on miTAGs were dominated by bacteria (~96%), with low abundances of eukarya (~3%) and archaea (~1%) (Supplementary Table S1). At the phylum level (Figure 3A), bacterial communities were mostly composed of Cyanobacteria oxygenic phototrophs (33, 53, and 21% of total rRNA SSU sequences at 48, 58, and 66°C, respectively) and Chloroflexi anoxygenic phototrophs (higher than Cyanobacteria only at 66°C, with 35% of total rRNA SSU sequences). Other representative members of the community were Proteobacteria (5–11%), Deinococcus–Thermus (2–7%), Firmicutes (1–17%), and Bacteroidetes (4–8%) (Figure 3A).

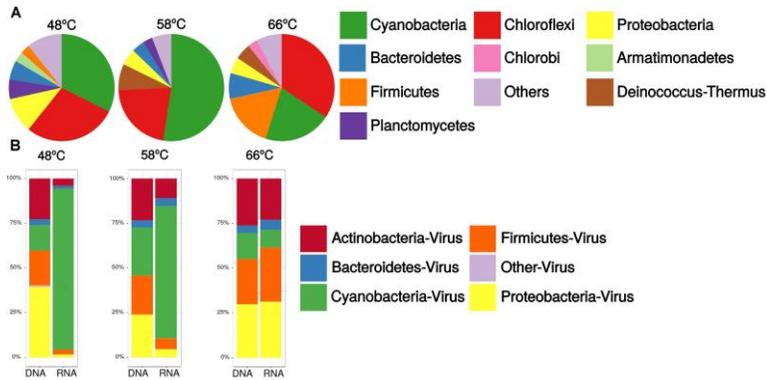


Figure 3: Relative abundances of (A) Bacterial community, to Phylum level, in the microbial mats obtained from 16S miTAGs, standardized by the total number of metagenomic reads (DNA) for each temperature sample, and (B) Caudovirales community at the host Phylum level, obtained from shotgun sequences in metagenomes (DNA) and metatranscriptomes (RNA), standardized by the total number of reads from each temperature sample.

The host assignment, based on taxonomy from viral reads of the most representative Caudovirales (Figure 3B), showed that viruses putatively infected members of the bacterial phyla Proteobacteria, Cyanobacteria, Actinobacteria, and Firmicutes. Metagenomic data showed that increases in temperature led to an increase in viruses from Actinobacteria and Firmicutes. Additionally, an increase in Cyanobacteria viruses was observed at 58°C. Viruses from Proteobacteria, Actinobacteria, and Firmicutes were represented by the three Caudovirales families, while viruses from Cyanobacteria were represented by Podoviridae and Myoviridae families only (Supplementary Table S3), where cyanopodovirus and cyanomyovirus reads increase from 31 to 50% at 48°C and from 30 to 45% at 58°C, then decrease to 23 to 28% at 66°C, respectively.

Metatranscriptomic sequences from Caudovirales potentially infecting Cyanobacteria, were predominant at 48°C and 58°C, with ~90% and ~74% of the total viral sequences, respectively. However, cyanophage transcripts abruptly decrease at 66°C. Cyanophages were exclusively related to the Myoviridae and Podoviridae families (Supplementary Table S3). Reads associated with cyanopodoviruses and cyanomyoviruses

gradually decreased with temperature; between 48 and 58°C, virus reads declined from 95% and 96% to 84% and 89%, respectively. On the other hand, at 66°C a more severe decline was observed, to 15% and 20%, respectively. Conversely, with the reduced representation of Cyanobacteria at 66°C, other caudovirales transcripts increased, including those that infect Proteobacteria (~31%), Firmicutes (~30%), and Actinobacteria (~23%).

Thermophilic Cyanophage Genome Recovery

The metagenome assembly recovered 3,912; 2,697; and 2,758 contigs, at 48°C, 58°C, and 66°C, respectively. A script search [44] resulted in 11 circular contigs, possibly indicating complete genomes. Subsequent BLASTP analysis [40] of predicted proteins indicated that only one circular contig had viral hallmark genes, meanwhile nine contigs had genes associated with bacterial mobile genetic elements and one contig remain completely unknown. These hallmark genes are shared by many viruses but are absent from cellular genomes [59]. VirSorter tool analysis [46] confirmed these results, obtaining the same complete putative viral contig from the 58°C assembly, 40,740 bp long and 43.9% of GC content. This contig, TC-CHP58 (Figure 4A), was associated with a Cyanobacterial host. TC-CHP58 was present (reads recruitment) over all temperatures in Porcelana hot spring (Figure 5 and Supplementary Figure S1). At 66°C, TC-CHP58 was sevenfold more abundant than their putative host (measured as *Mastigocladus* RUBISCO gene abundance); at 48°C, the virus-host ratio was 1:1, and at 58°C the host was fourfold more abundant than TC-CHP58. Metatranscriptomic reads also show that TC-CHP58 was active over all temperatures (Figure 5 and Supplementary Figure S2), but with lower transcription levels than the putative host (measured as *Mastigocladus* RUBISCO gene activity), ranging between 80- and 8-fold lower (Supplementary Table S4). TC-CHP58 viral DNA:RNA ratio indicated similar proportions (2.4) at 58°C, least similar (552.9) at 66°C; while at 48°C the ratio was 10.4 (Supplementary Table S4).

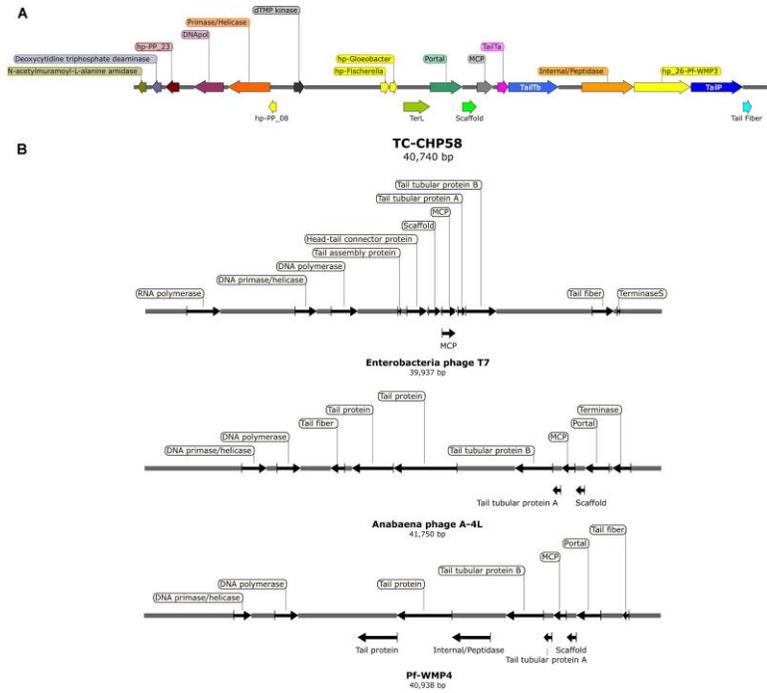


Figure 4: (A) Genomic organization of the thermophilic cyanophage CHP58 (TC-CHP58). Arrows indicate the size, position, and orientation of annotated ORFs, with predicted functions or homologs (e.g., DNApol, DNA polymerase; TailY a/b, tail tubular protein a/b; MCP, major capsid protein; TerL, large terminase subunit; TailP, Tail protein; hp-PP_08/23, homologous to hypothetical proteins 08/23 from cyanophage PP; hp-Fischerella/Gloebacter, homologous to hypothetical proteins in *Fischerella* sp. PCC 9605/*Gloebacter kilaueensis*). (B) Genomic organization of Enterobacteria phage T7, Anabaena phage A4L and Pf-WMP4. Arrows indicate the size, position, and orientation of viral core ORFs.

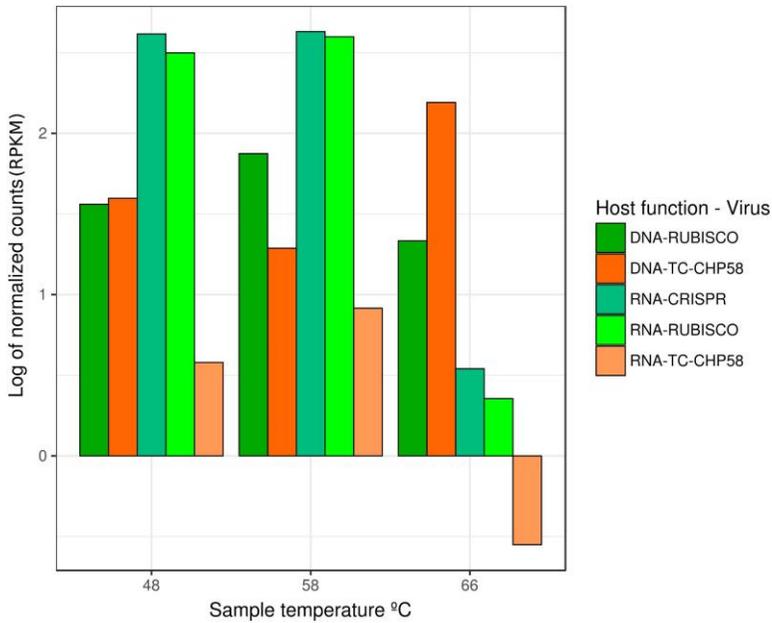


Figure 5: Relative abundance and transcriptomic expression of *Mastigocladus* spp. CRISPR systems and TC-CHP58. Abundance and expression of *Mastigocladus* RUBISCO was used as reference of the cyanobacterial presence and metabolic activity. Only specific CRISPR *loci* with proto-spacers in TC-CHP58 were fully quantified for each temperature. For improved visualization, counts are represented as Log of reads per kilobase million (RPKM).

Genomic Features and Organization of TC-CHP58

Complete protein prediction and annotation of TC-CHP58 using Prokka [45] and BLASTP revealed 39 putative ORFs, 10 of which were viral core proteins (i.e., capsid and tail-related proteins, DNA polymerase, Terminase, etc.), 22 had no significant similarities in NCBI nr database, and 4 were present in the database but with unknown function (Table 1).

Table 1: Blast analysis of predicted CDS from TC-CHP58 of known function against NCBI RefSeq (Release 75) and NR databases.

Query sequence ID	Subject sequence ID	Identity %	E-value	Bit Score
TC-CHP58_sequence	KF598865.1 [Cyanophage PP]	93%	0.2	54.7
TC-CHP58_CDS1	YP_009042789.1 DNA polymerase [Anabaena phage A-4L]	29.03	4.00E-60	223
TC-CHP58_CDS3	YP_009042786.1 DNA primase/helicase [Anabaena phage A-4L]	25.21	2.00E-28	131
TC-CHP58_CDS4	YP_008766966.1 hypothetical protein PP_06 [Cyanophage PP]	29.7	0.0006	47
TC-CHP58_CDS7	WP_026824764.1 dTMP kinase [<i>Exiguobacterium marinum</i>]	30.61	4.00E-22	99.4
TC-CHP58_CDS13	WP_026731322.1 hypothetical protein [Fischerella sp. PCC 9605]	34.55	7.00E-12	69.3
TC-CHP58_CDS14	WP_023172199.1 hypothetical protein [Gloeobacter klauensis]	42.31	4E-05	48.1
TC-CHP58_CDS15	YP_008766995.1 terminase [Cyanophage PP]	44.39	2.00E-150	456
TC-CHP58_CDS16	YP_001285799.1 portal protein [Phormidium phage PF-WMP3]	42.96	0	551
TC-CHP58_CDS17	YP_009042804.1 scaffold protein [Anabaena phage A-4L]	30.69	1E-07	60.8
TC-CHP58_CDS18	YP_008766991.1 capsid protein [Cyanophage PP]	48.14	4.00E-109	335
TC-CHP58_CDS20	YP_009042802.1 tail tubular protein A [Anabaena phage A-4L]	29.52	3.00E-28	116
TC-CHP58_CDS21	YP_001285795.1 tail tubular protein B [Phormidium phage PF-WMP3]	36.49	0	630
TC-CHP58_CDS24	YP_009042798.1 internal protein [Anabaena phage A-4L]	29.86	5.00E-41	174
TC-CHP58_CDS25	YP_001285791.1 P1WMP3_26 [Phormidium phage PF-WMP3]	28.26	5.00E-23	117
TC-CHP58_CDS26	YP_009042796.1 tail protein [Anabaena phage A-4L]	24.8	4.00E-89	322
TC-CHP58_CDS32	WP_039085449.1 N-acetylmuramoyl-L-alanine amidase [<i>Tolypathrix butalis</i>]	46.29	3.00E-46	160
TC-CHP58_CDS35	WP_043587103.1 deoxycytidine triphosphate deaminase [<i>Diplospira collettarium</i>]	44.9	2.00E-48	167
TC-CHP58_CDS37	YP_008766981.1 hypothetical protein PP_23 [Cyanophage PP]	29.92	1.00E-21	101

Blast analysis of the viral genes in TC-CHP58, revealed 25 to 48% identity (amino acidic level) with proteins from Cyanophage PP, PF-WMP3 and Anabaena phage A-4L, that infect freshwater filamentous Cyanobacteria such as *Phormidium*, *Plectonema*, and *Anabaena* (Table 1). At the nucleotide level, there was almost no similarity to any known sequence except for a short segment of 40 nucleotides, which showed 93% similarity to a Portal protein gene sequence of *Plectonema* and *Phormidium* cyanopodoviruses (Cyanophage PP; NC_022751 and PF-WMP3; NC_009551).

Gene prediction by Prodigal indicated that the TC-CHP58 genome might be structured into two clusters, based on the transcriptional direction and putative gene functions (Figure 4A). The predicted ORFs (Table 1) in the sense strand encode proteins involved in DNA replication and modification, such as DNA polymerase and DNA primase/helicase. Conversely, the ORFs in the antisense strand (Table 1) encode proteins necessary for virion assembly, such as major capsid protein (MCP), tail fiber proteins, internal protein/peptidase, tail tubular proteins, scaffold protein, and portal protein. Moreover, two ORFs in the antisense strand had the best hits to the cyanobacterial hypothetical proteins found in the filamentous cyanobacterium *Fischerella* (WP_026731322. 1) and the unicellular *Gloeobacter* (WP_023172199.1).

Additionally, VIRFAM [60] was used to classify TC-CHP58 according to their neck organization (Supplementary Figure S3), being assigned to the Podoviridae Type 3 category with neck structural organization similar to the Enterobacteria phage P22 [60]. Hierarchical clustering of neck proteins grouped TC-CHP58 together with the freshwater cyanophages Pf-WMP3 and Pf-WMP4, separating them from marine cyanophages such as P60 and Syn5.

Even when a large number of viral reads were assigned to cyanophages of Myoviridae family, it was not possible to recover any genome of this type. Most of the Myoviridae related contigs only had non-structural genes or hypothetical proteins of unknown function which align with proteins of known cyanomyoviruses. Here, the absence of hallmark genes from Cyanobacteria related viruses makes their accurate classification as cyanomyoviruses impossible.

Phylogenetic Analysis of Phage TC-CHP58

To investigate the relationship of the phage TC-CHP58 within the Podoviridae family, the DNAPol gene was selected for comparison, using published viral genomes. The analysis included representatives of Picovirinae and Autographivirinae subfamilies, plus all the available DNAPol genes from known freshwater podoviruses (Pf-WMP3, PP, Pf-WMP4 and A-4L) infecting filamentous heterocystous cyanobacteria from the order Nostocales and non-heterocystous from order Oscillatoriales, plus those infecting marine *Synechococcus* spp. and *Prochlorococcus* spp. The DNAPol tree (Figure 6) showed the phage TC-CHP58 as part of a monophyletic clade with all cyanopodoviruses described as infecting freshwater filamentous cyanobacteria, and more distantly, with the marine cyanopodovirus clade that infects *Synechococcus* spp. and *Prochlorococcus* spp. Both cyanophage subgroups are closely related with podoviruses from the Autographivirinae subfamily, which includes all T7 relatives. Furthermore, the phylogeny of the MCP was constructed for freshwater and marine representatives of the Autographivirinae subfamily. The available MCP gene from BHS3 Cyanophage partial genome,

that is the only known thermophilic representative within the Podoviridae family, was also included [23]. The MCP tree (Supplementary Figure S4) showed similar results to the DNAPol tree (Figure 6), with a monophyletic origin for all freshwater cyanophages infecting filamentous cyanobacteria, emphasizing the division between freshwater and marine cyanobacterial viruses, and their affiliation with T7 phage. The thermophilic representatives of Podoviridae family were located in different branches inside the freshwater clade, with BHS3 more basal than TC-CHP58.

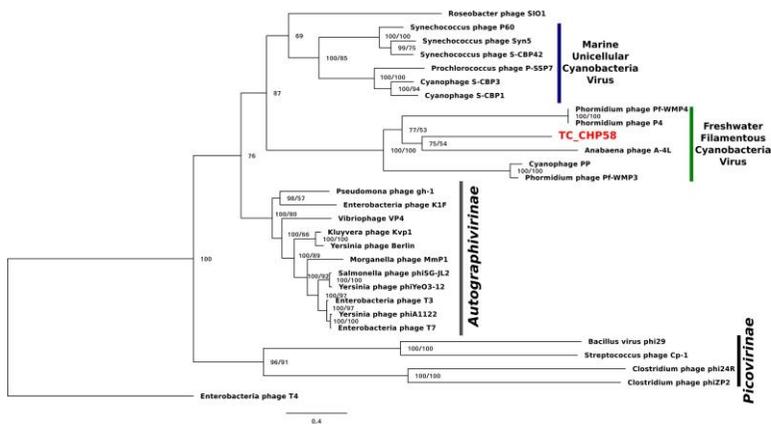


Figure 6: Bayesian inference phylogenetic reconstruction of DNA polymerase I protein of TC-CHP58. Numbers indicate Bayesian posterior probabilities as percentage/ultra-fast bootstrap values. Only UFBoot values over 80 and Bayesian PP over 50 are shown. The sequence characterized in the present study is reported in bold letters. Scale bar: 0.4 amino acid substitutions per site.

CRISPR Arrays on TC-CHP58 Host

Given the high abundance [9,10] and activity [10] of cyanobacteria, such as *Mastigocladus* spp., in Porcelana hot spring (Figure 3A), and in order to confirm the putative host of phage TC-CHP58, CRISPR spacer arrays were identified using the CRISPRFinder tool [56] for seven *Mastigocladus* spp. Contigs, obtained from metagenome assemblies at 48°C, 58°C, and 66°C. Three CRISPR *loci* were common between all temperatures (48_CRISPR_2, 58_CRISPR_5, and 66_CRISPR_2), while four *loci* were specific to higher

temperatures (58–66°C) (Table 2). In total, the seven CRISPR *loci* contain 562 spacers, of which 25 of them had a proto-spacer sequence in the TC-CHP58 genome (Table 2). From the 25 spacers, 19 have a target ORF of known function, such as DNA polymerase, dTMP, portal protein, M23-petidase, tail protein, tail fiber, and deoxycytidine triphosphate deaminase. In general, each CRISPR *loci* contained spacers against different ORFs on TC-CHP58, or even against different locations on the same ORF. For the 25 spacers, searching the nt/nr database, using BLASTN and BLASTX, showed no similarity to any known sequence. Finally, in order to check if CRISPR systems were active, expression of the seven *loci* was directly quantified in the three metatranscriptomes. For all temperatures, slightly lower transcript levels were found compared to the *Mastigocladus* RUBISCO gene (Figure 5).

Identifying Single Nucleotide Variants in TC-CHP58 Genome

To assess if mismatches between the CRISPR spacer and proto-spacer sequences in TC-CHP58 genome were concealing potential variations in TC-CHP58 populations, a SNV calling was conducted. For this task LoFreq tool was used, as it is high sensitivity and has low false positive rates, lower as <0.00005% [57] and higher as 8.3% [61]. This approach, together with the use of sequences with qualities over q28 (whose error probability in the base call is $\leq 1.58\%$), allow us to consider these SNVs as real mutations.

A different number of SNVs was found at each temperature. TC-CHP58 showed 1611, 930, and 671 variant sites at 48°C, 58°C, and 66°C, respectively, unevenly distributed throughout the viral genome (Supplementary Figure S5). Considering the three metagenomes, a total of 3212 variable sites were present in the TC-CHP58 genome, with 391 SNVs present over all temperatures (Supplementary Figure S5). Most of the SNVs (74% on average) were located at coding regions on the TC-CHP58 genome, with variable rates, ranging from 15 to 0 SNVs for each 100 bp (Supplementary Table S5) over different ORFs.

A detailed analysis of SNVs in CRISPRs proto-spacer sites revealed the presence of these polymorphisms in 14 of the 25 spacer targets, with 13 mismatches and 4 perfect matches (Table 2). The total number of polymorphic sites was 22, with 13 SNVs causing a synonymous substitution and 7 causing a non-synonymous substitution (Table 2).

Table 2:CRISPR *loci* at each temperature detailed information and SNVs analysis for TC-CHP58 proto-spacer including alleles frequency and SNV coding effect.

T°C Sample	Viropote Sequence	Viral target	CRISPR <i>loci</i>	Proto-spacer Start	Proto-spacer End	Mismatch position	SNV position	Alleles *	Frequency of In CRISPR allele	SNV effect	Codon change
48	ACCITTTGAGACCTAAGCT TAAAGTTACTATCACAGAT	Internal protein-M23-peptidase	48_CRISPR_2_NODE1_1554	26558	26594	26564; 26567; 26582	-	-	-	-	-
58	AGAAGTTTTTCTGGCA GATATATGGTGGTCTCA	DNA polymerase	58_CRISPR_10_NODE_13413	282	320	-	282; 292; 313	G/T; C/A; G/A	0.078; 0.552; 0.549	All silent	CGA/AGA; GGG/CGT; TTC/TTT
58	GTTGTTGGTGCCTTGGAGT ACCGTTGAGABGGT	Hypothetical protein	58_CRISPR_10_NODE_13413	35008	35942	-	35008	G/A	0.052	Skirt	GGC/GGT
58	AGTTGTGCCCTCGAGCTA GAGAATTGCTGCACCT	Internal protein-M23-peptidase	58_CRISPR_10_NODE_13413	24692	24727	-	-	-	-	-	-
58	TAACTCTGGTGGGATTTGG TAGATTCGATCGCTG	NC	58_CRISPR_10_NODE_13413	8740	8774	-	8753	C/G	0.53	-	-
58	ACTACTTGATCAACCGGG GCTACACGGTAATCGTAGA	Tail fiber protein	58_CRISPR_10_NODE_13413	30649	30688	30650	30675	C/T	0.522	AV	GCT/GTT
58	ACDITTTGACACTAAGCTCT AAAGTTACTATCACAGAT	Internal protein-M23-peptidase	58_CRISPR_8_NODE_1001	26558	26594	26565; 26567; 26582	-	-	-	-	-
58	CCCAACAAGCTTAAATAAA TCTTCTATGATATGG	Hypothetical protein	58_CRISPR_8_NODE_4438	24219	24254	24226; 24229; 24238	-	-	-	-	-
58	AATACGGTGTAGTACTCTT GAAGAGGTGTACCG	Hypothetical protein	58_CRISPR_8_NODE_4438	30971	31005	30972	30978	G/A	0.588	Silent	ACG/ACA
58	GAAGGGTAGAGTGTCAAA ATTGGGATATAGTGTAG	Internal protein-M23-peptidase	58_CRISPR_8_NODE_4438	27172	27210	-	27180; 27181; 27186	T/A; C/A; A/C	0.626; 0.613; 0.575	S/T; S/C; T/P	TCT/ACT; TCT/TAT; ACC/CCC
58	GGATTAATCGGGGTTAG GGTGATCCACTA	Tail protein	58_CRISPR_8_NODE_4438	34211	34243	34214; 34226; 34241	-	-	-	-	-
58	TAGCTTAACATTACACAG GGGATAAGCTGTGTATGTC	Deoxyribonucleic acid triphosphate deaminase	58_CRISPR_9_NODE_4711	38225	38264	38264	38225	C/T	0.057	All silent	CTG/CTA; GAC; GAT
58	GACTTGATCTTTCCCGCT TCTCTAGCGAGATCTT	DNA polymerase	58_CRISPR_9_NODE_4711	688	705	-	671;	C/T;	0.031;	R/K;	AGG/AGG
58	ACGAGGGTGTATCTCCCG CGAAGTGGTGTGACCGGAT	dTMP kinase	58_CRISPR_9_NODE_4711	6338	6376	6375	6350	G/C	0.575	K/N	AAG/AAC
58	AAATACATCCGCACTTTAG GAGGTAAACCCAC	Hypothetical protein	58_CRISPR_9_NODE_4711	36127	36159	-	-	-	-	-	-
58	ACAGCGAAAGCAATTGTC TCTGAGCTAACAAGTT	Internal protein-M23-peptidase	58_CRISPR_9_NODE_4711	25742	25777	25776	-	-	-	-	-
58	GTCATCTCAATATGCTCT TTGTAGTCTTGOA	Internal protein-M23-peptidase	58_CRISPR_9_NODE_4711	25917	25950	-	25946	G/A	0.041	Silent	ATC/ATA
58	CAATCACACATTAACCCAT AGBTGACGGCACAAACA	Porin protein	58_CRISPR_9_NODE_4711	15920	15954	-	15941; 15953	A/G; A/T	0.651; 0.055	All silent	GGG/GGG; ATA/ATT
58	TAGTGATTTGAAACAGAGA CGCTGGATATACAC	Tail protein	58_CRISPR_9_NODE_4711	33859	33893	-	-	-	-	-	-
66	ATCTGTGATGTAACCTTAG AGTTAGGCTTCAAAGGT	Internal protein-M23-peptidase	66_CRISPR_2_NODE_1045	26558	26594	26566; 26567; 26582	-	-	-	-	-
66	TTAGACCAAGCACCATATC TTGGCGAAGAAAACCTCT	DNA polymerase	66_CRISPR_3_NODE_1491	282	320	282	292; 313	C/A; G/A	0.437; 0.024	All silent	GGG/CGT; TTC/TTT
66	ACDITATCTGAACGGTACTCA AGAGSCACACAC	Hypothetical protein	66_CRISPR_3_NODE_1491	33008	33942	33908	-	-	-	-	-
66	AGGTGCAGCAAACTCTACG TCAAGGGGCACAACCT	Internal protein-M23-peptidase	66_CRISPR_3_NODE_1491	24692	24727	-	-	-	-	-	-
66	GATGTGATGAAATGTACA CAATCCGACAGGTTTA	NC	66_CRISPR_3_NODE_1491	8740	8774	-	8753	C/G	0.051	-	-
66	TTTAAAGATTACCGTGTAGC CCGGTGTGATCAGATAGT	Tail fiber protein	66_CRISPR_3_NODE_1491	30649	30688	30650	30675	C/T	0.056	AV	GCT/GTT

*Alleles of CRISPR alleles

Discussion

The study of viruses from thermophilic phototrophic microbial mat communities remains largely unexplored except for a few cases providing limited information on viral presence within these communities [16,17]. Thus far, no study has characterized

viral composition and activity, or the identity of any complete viral genome. Here, using metagenomic and metatranscriptomic approaches, the composition of the most abundant and active viruses associated with the dominant members of the thermophilic bacterial community have been characterized, describing for the first time a full genome from a thermophilic cyanopodovirus (TC-CHP58). Moreover, the active cross-fire between this new cyanophage and its host is demonstrated, through TC-CHP58 population diversification (SNV), and *Mastigocladus* spp. CRISPR heterogeneity, as a response to selective pressure from the host defense system and viral predation, respectively.

Active and Ubiquitous Cyanophage-Type Caudovirales in Phototrophic Microbial Mats

The taxonomic classification of small subunit rRNA (Supplementary Table S1) indicates that the phototrophic mats in Porcelana hot spring are dominated by Bacteria (96% on average) as commonly observed in other thermophilic phototrophic microbial mats [2,62].

Porcelana microbial mats are mainly built by filamentous representatives of two phototrophic phyla, Cyanobacteria (oxygenic) and Chloroflexi (anoxygenic), with *Mastigocladus*, *Chloroflexus*, and *Roseiflexus* as the main genera, respectively. This is verified by previous surveys carried out by the authors [9,10], as well as investigations from the White Creek, Mushroom, and Octopus hot springs in Yellowstone [2,13,62,63], presenting similar pH, thermal gradient and low sulfide concentrations.

Porcelana dominant viruses (~70% and ~68% of metagenomic and metatranscriptomic reads) are from the families Myoviridae, Podoviridae, and Siphoviridae within the Caudovirales Order (Figure 2), which typically infect Bacteria and some non-hyperthermophilic Archaea [64]. These results were also supported by TEM images (Figure 1). The small decrease in transcripts associated to caudovirales with the increase in temperature is due to the reduction of sequences related to

Podovirus and Myovirus families. A plausible explanation, is that at high temperatures some representatives of these families might have a lysogenic lifestyle, then a fraction of them will remain inactive as prophages.

Dominance by Caudovirales was only reported recently from the Brandvlei hot spring, South Africa, a slightly acidic (pH 5.7) hot spring with moderate temperature (60°C) and green microbial mat patches [23]. Previously, the presence of this viral order had only been suggested in moderate thermophilic phototrophic mats from Yellowstone hot springs, through indirect genomic approximations, such as spacers in CRISPR *loci*, from dominant bacterial members [16,17] or classifications based on nucleotide motives in metaviromic data [17,28].

Contributions from megavirus sequences were also identified in Porcelana hot spring (Figure 2), with an average of ~24% viral metagenomic reads, associated with unicellular eukaryotic hosts such as those from Phycodnaviridae and Mimiviridae families, and also the family Marseilleviridae, but to a lesser extent. The presence of VLPs from these three viral families could not be corroborated through TEM, using the limited available viral fraction (<0.2 µm) within the community, as it has been previously documented that nucleocytoplasmic large DNA viruses (NCLDV) particles are only found in larger viral fractions [65]. The ubiquity of NCLDVs in hot springs was previously described in a hydrothermal freshwater lake in Yellowstone, with assemblies of genomes from Phycodnaviridae and Mimiviridae [66].

Viral relative abundances and activity reported here can be affected by the lack of replicates at this highly local heterogeneity samples. However, the fact of having three different temperature sampling points for metagenomics and metatranscriptomics, partially compensates the replicate limitation.

Furthermore, many viruses in an environmental sample share a degree of similarity in their genomic sequence, and this intrinsic complexity of metagenomic/metatranscriptomic samples makes

difficult to accurately estimate the relative abundances or activity of specific phages at low ranks of taxonomy tree, such as the species level [67]. To avoid this problem, our strategy focused on the use of the LCA algorithm at higher taxonomic levels (Order and Family) to classify the viral reads, as well as for the inferred hosts, we use the phylum level.

Virus-host inference in Porcelana phototrophic mats (Figure 3B), demonstrated that the most frequent targets for viral infections were the most dominant and active components of the bacterial communities. Similarly, this is the case in other environments, such as in the human microbiome [68] and marine communities [69,70]. In Porcelana, it is demonstrated that within microbial mats at 48°C and 58°C, cyanophages were among the most active viruses (Figure 3B), as were Cyanobacteria, such as *Mastigocladus* spp., as exemplified in terms of primary production and nitrogen fixation [10]. The presence of cyanophages has been previously suggested in Yellowstone hot spring phototrophic mats [16,17], and more recently in the Brandvlei hot spring, South Africa [23]. Heidelberg et al. [16] found that CRISPR spacers in unicellular cyanobacteria *Synechococcus* isolates (Syn OS-A and Syn OS-B9) from Octopus Hot Spring, might have 23 known viral targets (lysozyme-related reads, PFAM DUF847) on an independently published metavirome from the same hot spring. More recently, 171 viral contigs associated with the host genus *Synechococcus*, based on tetranucleotide frequencies, were identified from a microbial mat (60°C) metavirome from Octopus Spring. The majority of the annotated ORFs on the viral contigs coded for glycoside hydrolases, with lysozyme activity, identifying six CRISPR proto-spacers in those genes [17]. Even though a taxonomic relationship with cyanophages was not confirmed for those proto-spacers containing contigs [16,17], it provides evidence toward the presence of cyanophages related sequences within these thermophilic mats. The work by Zablocki et al. [23] reconstructed a 10 kb partial genome of a new cyanophage (BHS3) from Brandvlei hot spring metavirome, stating that cyanophages appear to be the dominant viruses in the hot spring. The BHS3 contig (MF098555) contains nine ORFs, with the majority of the identified proteins having a close relation to the

Cyanophage PP and *Phormidium* phage Pf-WMP3, which infect freshwater filamentous cyanobacteria *Phormidium* and *Plectonema*.

The presence of cyanophages related sequences in thermophilic phototrophic mats is significant, since these viruses are known to play an important role in the evolution of cyanobacteria [71]. Cyanophages affect the rate and direction of cyanobacterial evolutionary processes, through the regulation of abundance, population dynamics, and natural community structure. This has been extensively studied and demonstrated for marine environments [72,73]. These cyanophages are proven to play a relevant role in the marine biogeochemical cycles, through the infection and lysis of Cyanobacteria, affecting carbon and nitrogen fixation [74]. Moreover, cyanophages act as a global reservoir of genetic information, as they are vectors for gene transfer, meaning that cyanobacteria can acquire novel attributes within aquatic environments [75,76].

Caudoviruses were prevalent at 66°C in Porcelana, and potentially infecting Firmicutes, Proteobacteria, and Actinobacteria. These phyla have also been previously identified in other hot springs at temperatures above 76°C, such as in Octopus and Bear Paw [28]. At high temperatures in Porcelana also the phylum Chloroflexi was dominant in the phototrophic mat (Figure 3A). However, viral sequences related to this taxon could not be retrieved, as neither viruses nor viral sequences have been confirmed to infect members of this phylum in any environment. Davison et al. [17], described viral contigs associated with *Roseiflexus* sp. from a metavirome from Octopus Spring, but only raw reads are publicly available, without taxonomic assignment. Finally, the recently released IMG/VR database [77] contains three contigs associated by CRISPR spacers to *Chloroflexus* sp. Here, a BLASTP analyses against RefSeq viral proteins revealed that six of these proteins have a best hit in *Mycobacterium* phage proteins and one which best hit was a *Clavibacter* phage protein. These findings, suggest that some of the viral reads classified as Actinobacteria viruses could be instead from unknown Chloroflexi viruses.

Viral Mining Reveals a New Infective Thermophilic Cyanopodovirus Lineage

Metagenomic surveys of viral genomes are an effective way to detect unknown viruses [46,66,78,79]. In metagenomics, two key elements for virus detection are the presence of viral hallmark genes and the circularity of viral contigs [46,78]. Based on these two principles, a complete genome (TC-CHP58) was identified. The genome was represented by a viral contig of 50 kb, which is a typical size for Caudovirales members from the Podoviridae family. The genome size and viral core proteins affiliated with the Podovirus seems to make TC-CHP58 the first report of a full genome of a thermophilic cyanopodovirus. Moreover, the genome organization (Figure 4B) shows a consistent synteny with other cyanopodoviruses, which also lack RNA polymerase inside the T7 supergroup, as described for the viruses Pf-WMP4, Pf-WMP3, Cyanophage PP, Anabaena phage A-4L [80-83], and the recently reported partial genome of the thermophilic BHS3 cyanophage [23]. Initially, the presence of a single-subunit RNA polymerase that binds phage specific promoters was considered to be a major, and unique characteristic of the T7 supergroup [84]. However, more recently, it has been proposed that podoviruses that share extensive homology with T7, but lack the phage RNA polymerase, are still part of the T7 supergroup, as distant and probably ancient branches [85].

TC-CHP58 presented a genome organization that can be divided into two portions (Figure 4A); with ORFs in the sense strand related to DNA replication and modification, and genes encoded in the antisense strand related to virion assembly. This genome organization is also present in other freshwater T7-related podoviruses that infect filamentous cyanobacteria [80,82,83], including the thermophilic BHS3 cyanophage [23]. This setup is also similar to the class II and III organization genes in T7-like viruses, where class II genes are responsible for DNA replication and metabolism, and class III genes include structural and maturation genes [84]. The VIRFAM analysis of neck protein organization verifies the classification of TC-CHP58 within the Podoviridae family (Supplementary Figure S3), where the Type 3 podovirus encompasses T7-like phages from

Autographivirinae subfamilies and several other genera [60]. The T7-like classification for TC-CHP58, and other podoviruses that infect freshwater filamentous cyanobacteria, is supported by the organization of the genome into two portions as well as the organization of the neck proteins.

The phylogenetic position of TC-CHP58, based on DNA polymerase I (DNApol) (Figure 6) and MCP (Supplementary Figure S4) predicted proteins, confirm the affiliation of this new virus within the family Podoviridae. Both phylogenetic markers verify the separation between the marine from the freshwater cyanopodoviruses within the T7 family, as previously proposed [80,83]. These results also support the connection between the T7 phages and marine and freshwater cyanopodoviruses [80,83,85,86], including TC-CHP58 and BHS3 as representatives of a novel, and potentially globally distributed thermophilic cyanophage lineage. Moreover, this data demonstrates that marine and freshwater cyanopodoviruses, including the thermophilic TC-CHP58, are part of the Autographivirinae subfamily as previously suggested for Cyanophage P60 and Roseophage SIO1 [87], both included in this analysis.

In Porcelana, the virus host ratio relating to TC-CHP58 presence was lower than the typical values observed in freshwater environments [88], being more similar to other geothermal environments where viral density is typically lower, with 10- to 100-fold less viruses than host cells [89]. This is expected, considering that there are abundant cyanobacteria in phototrophic mats in Porcelana in comparison with the 10^4 mL^{-1} VLPs observed in the water of hot springs [24]. It is also demonstrated that TC-CHP58 presented higher infection efficiency, as revealed by the viral DNA to RNA ratios at lower temperatures (58°C, then 48°C) with cyanobacteria dominating, while at 66°C most of the TC-CHP58 remained inactive (Figure 5). Infection inefficiency is multidimensional, as it initiates from reduced phage adsorption, RNA, DNA, and protein production [90]. Thus, the high copy number of TC-CHP58 DNA at 66°C may be due to the persistence of viral DNA [91] encapsidated extracellularly and intermixed in the microbial mat were the host

(*Mastigocladus* spp.) has a low activity as evidenced by the low expression of the RUBISCO gene and the CRISPR *loci*. An alternative explanation is the absence, or the diminished presence, of the specific host due to intraspecific diversification as evidenced by the existence of different CRISPR *loci* at different temperatures. This theory has been proposed for other cyanobacteria, such as *Prochlorococcus* and *Phormidium*, where slight differences in fitness, niche, and selective phage predation, explain the coexistence of different populations [79,92]. The last explanation acquires special importance in light of recent evidence that variations in the structure and function of the heterocyst and differential CRISPR *loci* are fundamental to diversification of *Mastigocladus laminosus* (also known as *Fischerella thermalis*), a cosmopolitan thermophilic cyanobacterium, reinforcing the importance of viral predation [93].

CRISPR Spacers Assign *Mastigocladus* spp. as Putative Hosts for TC-CHP58

It was possible to verify *Mastigocladus* spp. as putative hosts for the new cyanopodovirus (TC-CHP58), via the analysis of CRISPR spacers found in the cyanobacteria, recovered from contigs obtained in the same metagenomic datasets. This methodology has been previously used for the identification of novel viruses in hot springs [16,17,94], as well as in other environments such as acid mines [95] the human microbiome [96], as well as sea ice and soils [97].

Observations from the CRISPR *loci* over all temperatures (Table 2) indicated that, in general, proto-spacers in the TC-CHP58 genome were distributed on coding, and therefore more conserved regions. The expression of seven CRISPR *loci* (Figure 5), demonstrated the activity of the *Mastigocladus* spp. defense system against TC-CHP58 over all temperatures. CRISPR arrays are transcribed into a long precursor, containing spacers and repeats, that are processed into small CRISPR RNAs (crRNAs) by dedicated CRISPR-associated (Cas) endoribonucleases [98]. Although it is not possible to measure mature crRNAs, as due to their small size they are likely to be filtered out in RNA-seq

libraries, this approximation has been validated using large datasets [99].

Despite variations in the number of CRISPR *loci* observed at each temperature, with 60% of the total CRISPR *loci* found in *Mastigocladus* contigs at 58°C, the abundance of reads agreed with the abundance of other genes required by these cyanobacteria, such as the RUBISCO gene (Figure 5). This further verified that the *loci* are from *Mastigocladus* populations. The different CRISPR *loci* found over the different temperatures in Porcelana (Table 2), also reinforces the notion that diversification of *Mastigocladus* is partly due to selective pressure exerted by the predation of viruses, such as TC-CHP58. This theory has been previously put forward for *Mastigocladus laminosus* in Yellowstone [93], and proposed for marine cyanobacteria [92,100].

Furthermore, each CRISPR *loci* contains spacers that corresponds to different proto-spacers in the TC-CHP58 genome. Increases in spacer number and diversity against the same virus may explain the increase in interference, whilst decreasing the selection of escape mutants [10]. Priming mechanisms are the most efficient form of obtaining new spacers [101], using a partial match between a pre-existing spacer and the genome of an invading phage to rapidly acquire a new “primed” spacer [102]. Then, over-representation of spacer sequences in some regions of the TC-CHP58 genome may be related to a site that has already been sampled by the CRISPR-Cas machinery or by other biases such as the secondary structure of phage ssDNA, GC content, and transcriptional patterns [103].

The selection pressure of multiple spacers in *Mastigocladus* CRISPR *loci* leads to the emergence of SNVs in the TC-CHP58 viral populations (Table 2), which cause mismatches between spacers and proto-spacers, resulting in the attenuation or evasion of the host immune response [104]. It is still possible to utilize mismatched spacers for interference and/or primed adaptation, however, the degree of tolerance to mismatches for interference among the CRISPR-Cas, varies substantially between different CRISPR-Cas type systems [104]. The variable frequency (0.6–

0.02) of the corresponding spacer SNVs alleles on TC-CHP58 proto-spacers, suggests that some variants are more prevalent throughout the population, regardless of whether the SNV causes a silent mutation. Based on this evidence, it has been proposed that, for other microbial communities, only the most recently acquired spacer can exactly match the virus. This suggests that community stability is driven by compensatory shifts in host resistance levels and virus population structure [95].

The present study describes the underlying viral community structure and activity of thermophilic phototrophic mats. Moreover, abundant virus populations are linked to dominant bacteria, demonstrating the effectiveness of omics approaches in estimating the importance and activity of a viral community, in this case with thermophilic cyanophages.

Additionally, the first full genome of a new T7-related virus that infects thermophilic representatives of the cyanobacterium *Mastigocladus* spp. was here retrieved. This genome may represent a novel, globally present, freshwater thermophilic virus from a new lineage from the Podoviridae family. The latter was strongly suggested by the significant phylogenetic relationship and shared gene organization with the BHS3 cyanophage partial genome (South Africa). Even more, TC-CHP58 proteins also matches several contigs that include common viral hallmarks genes in the IMG/VR database. However, further work is necessary to fully understand the global representation and relevance of this virus, which complete genome is presented here as first reference available.

Finally, the evolutionary arms race between a specific cyanobacteria-cyanophage in the natural environment is exposed, where there exist a variety of potential scenarios. For instance, host resistance may increase over time forcing the decrease of viral populations, or a specific virus population may occasionally become extremely virulent and cause the crash of the host population as proposed by the “kill the winner” model [95]. Alternatively, if CRISPR systems and the diversification of the viral population remain in balance through time, a relatively stable virus and host community may result.

Footnotes

1. [^sourceforge.net/projects/bbmap/](https://sourceforge.net/projects/bbmap/)
2. [^http://beast.bio.ed.ac.uk/Tracer](http://beast.bio.ed.ac.uk/Tracer)

References

1. Uldahl K, Peng X. Biology, biodiversity and application of thermophilic viruses. In: T Satyanarayana, KY Littlechild Jennifer, editors. *Thermophilic Microbes in Environmental and Industrial Biotechnology*. Berlin: Springer. 2013; 271–306.
2. Inskeep WP, Jay ZJ, Tringe SG, Herrgård MJ, Rusch DB. The YNP metagenome project: environmental parameters responsible for microbial distribution in the yellowstone geothermal ecosystem. *Front. Microbiol.* 2013; 4: 67.
3. Steunou AS, Bhaya D, Bateson MM, Melendrez MC, Ward DM, et al. In situ analysis of nitrogen fixation and metabolic switching in unicellular thermophilic cyanobacteria inhabiting hot spring microbial mats. *Proc. Natl. Acad. Sci. U.S.A.* 2006; 103: 2398–2403.
4. Steunou AS, Jensen SI, Brecht E, Becraft ED, Bateson MM, et al. Regulation of *nif* gene expression and the energetics of N₂ fixation over the diel cycle in a hot spring microbial mat. *ISME J.* 2008; 2: 364–378.
5. Bhaya D, Grossman AR, Steunou AS, Khuri N, Cohan FM, et al. Population level functional diversity in a microbial community revealed by comparative genomic and metagenomic analyses. *ISME J.* 2007; 1: 703–713.
6. Klatt CG, Wood JM, Rusch DB, Bateson MM, Hamamura N, et al. Community ecology of hot spring cyanobacterial mats: predominant populations and their functional potential. *ISME J.* 2011; 5: 1262–1278.
7. Stewart W. Nitrogen fixation by blue-green algae in Yellowstone thermal areas. *Phycologia.* 1970; 9: 261–268.
8. Miller SR, Purugganan M, Curtis SE. Molecular population genetics and phenotypic diversification of two populations of the thermophilic cyanobacterium *Mastigocladus laminosus*. *Appl. Environ. Microbiol.* 2006; 72: 2793–2800.

9. Mackenzie R, Pedrós-Alió C, Díez B. Bacterial composition of microbial mats in hot springs in Northern Patagonia: variations with seasons and temperature. *Extremophiles*. 2013; 17: 123–136.
10. Alcamán M, Fernandez C, Delgado A, Bergman B, Díez B. The cyanobacterium *Mastigocladus* fulfills the nitrogen demand of a terrestrial hot spring microbial mat. *ISME J*. 2015; 9: 2290–2303.
11. Van der Meer MT, Klatt CG, Wood J, Bryant DA, Bateson MA, et al. Cultivation and genomic, nutritional, and lipid biomarker characterization of roseiflexus strains closely related to predominant in situ populations inhabiting yellowstone hot spring microbial mats. *J. Bacteriol.* 2010; 12: 3033–3042.
12. Liu Z, Klatt CG, Wood JM, Rusch DB, Ludwig M, et al. Metatranscriptomic analyses of chlorophototrophs of a hot-spring microbial mat. *ISME J*. 2011; 5: 1279–1290.
13. Klatt CG, Inskeep WP, Herrgard MJ, Jay ZJ, Rusch DB, et al. Community structure and function of high-temperature chlorophototrophic microbial mats inhabiting diverse geothermal environments. *Front. Microbiol.* 2013; 4: e106.
14. Lin KH, Liao BY, Chang HW, Huang SW, Chang TY, et al. Metabolic characteristics of dominant microbes and key rare species from an acidic hot spring in Taiwan revealed by metagenomics. *BMC Genomics*. 2015; 16: 1029.
15. Cole JK, Peacock JP, Dodsworth JA, Williams AJ, Thompson DB, et al. Sediment microbial communities in Great Boiling Spring are controlled by temperature and distinct from water communities. *ISME J*. 2013; 7: 718–729.
16. Heidelberg JF, Nelson WC, Schoenfeld T, Bhaya D. Germ warfare in a microbial mat community: CRISPRs provide insights into the co-evolution of host and viral genomes. *PLoS One*. 2009; 4: e4169.
17. Davison M, Treangen TJ, Koren S, Pop M, Bhaya D. Diversity in a polymicrobial community revealed by analysis of viromes, endolysins and CRISPR spacers. *PLoS One*. 2016; 11: e0160574.
18. Schoenfeld T, Patterson M, Richardson PM, Wommack KE, Young M, et al. Assembly of viral metagenomes from

- yellowstone hot springs. *Appl. Environ. Microbiol.* 2008; 74: 4164–4174.
19. Rachel R, Bettstetter M, Hedlund BP, Häring M, Kessler A, et al. Remarkable morphological diversity of viruses and virus-like particles in hot terrestrial environments. *Arch. Virol.* 2002; 147: 2419–2429.
 20. Yu MX, Slater MR, Ackermann HW. Isolation and characterization of *Thermus* bacteriophages. *Arch. Virol.* 2006; 151: 663–679.
 21. Bolduc B, Shaughnessy DP, Wolf YI, Koonin EV, Roberto FF, et al. Identification of novel positive-strand RNA viruses by metagenomic analysis of archaea-dominated Yellowstone hot springs. *J. Virol.* 2012; 86: 5562–5573.
 22. Bolduc B, Wirth JF, Mazurie A, Young MJ. Viral assemblage composition in Yellowstone acidic hot springs assessed by network analysis. *ISME J.* 2015; 9: 1–16.
 23. Zablocki O, van Zyl LJ, Kirby B, Trindade M. Diversity of dsDNA viruses in a South African hot spring assessed by metagenomics and microscopy. *Viruses.* 2017; 9: 348.
 24. Breitbart M, Wegley L, Leeds S, Rohwer F, Schoenfeld T. Phage community dynamics in hot springs. *Appl. Environ. Microbiol.* 2004; 70: 1633–1640.
 25. Redder P, Peng X, Brügger K, Shah SA, Roesch F, et al. Four newly isolated fuselloviruses from extreme geothermal environments reveal unusual morphologies and a possible interval recombination mechanism. *Environ. Microbiol.* 2009; 11: 2849–2862.
 26. Prangishvili D, Garrett RA. Exceptionally diverse morphotypes and genomes of crenarchaeal hyperthermophilic viruses. *Biochem. Soc. Trans.* 2004; 32: 204–208.
 27. Pawlowski A, Rissanen I, Bamford JKH, Krupovic M, Jalasvuori M. *Gammasphaerolipovirus*, a newly proposed bacteriophage genus, unifies viruses of halophilic archaea and thermophilic bacteria within the novel family *Sphaerolipoviridae*. *Arch. Virol.* 2014; 159: 1541–1554.
 28. Pride DT, Schoenfeld T. Genome signature analysis of thermal virus metagenomes reveals Archaea and thermophilic signatures. *BMC Genomics.* 2008; 9: 420.

29. Diemer GS, Stedman KM. A novel virus genome discovered in an extreme environment suggests recombination between unrelated groups of RNA and DNA viruses A novel virus genome discovered in an extreme environment suggests recombination between unrelated groups of RNA and DNA virus. *Biol. Direct.* 2012; 7: 1–14.
30. Tekere M, Lötter A, Olivier J, Jonker N, Venter S. Metagenomic analysis of bacterial diversity of Siloam hot water spring, Limpopo, South Africa. *Afr. J. Biotechnol.* 2011; 10: 18005–18012.
31. Jonker CZ, van Ginkel C, Olivier J. Association between physical and geochemical characteristics of thermal springs and algal diversity in Limpopo Province, South Africa. *Water SA.* 2013; 2013: 95–104.
32. Alcamán ME, Alcorta J, Bergman B, Vásquez M, Polz M, et al. Physiological and gene expression responses to nitrogen regimes and temperatures in *Mastigocladus* sp. strain CHP1, a predominant thermotolerant cyanobacterium of hot springs. *Syst. Appl. Microbiol.* 2017; 40: 102–113.
33. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* 2011; 17: 10.
34. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics.* 2011; 27: 863–864.
35. Logares R, Sunagawa S, Salazar G, Cornejo-Castillo FM, Ferrera I, et al. Metagenomic 16S rDNA Illumina tags are a powerful alternative to amplicon sequencing to explore diversity and structure of microbial communities. *Environ. Microbiol.* 2014; 16: 2659–2671.
36. Schmieder R, Lim YW, Edwards R. Identification and removal of ribosomal RNA sequences from metatranscriptomes. *Bioinformatics.* 2012; 28: 433–435.
37. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 2013; 41: 590–596.
38. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat. Methods.* 2012; 9: 357–359.
39. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics.* 2010; 26: 2460–2461.

40. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, et al. BLAST plus: architecture and applications. *BMC Bioinformatics*. 2009; 10: 1.
41. Huson DH, Beier S, Flade I, Górska A, El-Hadidi M, et al. MEGAN Community edition – Interactive exploration and analysis of large-scale microbiome sequencing data. *PLoS Comput. Biol.* 2016; 12: e4957.
42. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 2012; 19: 455–477.
43. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*. 2010; 11: 119.
44. Crits-Christoph A, Gelsinger DR, Ma B, Wierzos J, Ravel J, et al. Functional interactions of archaea, bacteria and viruses in a hypersaline endolithic community. *Environ. Microbiol.* 2016; 18: 2064–2077.
45. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014; 30: 2068–2069.
46. Roux S, Enault F, Hurwitz BL, Sullivan MB. VirSorter: mining viral signal from microbial genomic data. *PeerJ*. 2015; 3: e985.
47. Edgar RC. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004; 32: 1792–1797.
48. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 2002; 30: 3059–3066.
49. Darriba D, Taboada GL, Posada D. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics*. 2011; 27: 1164–1165.
50. Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, et al. MrBayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 2012; 61: 539–542.
51. Trifinopoulos J, Nguyen LT, von Haeseler A, Minh BQ. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* 2016; 44: 1–4.

52. Wu YW, Simmons BA, Singer SW. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics*. 2016; 32: 605–607.
53. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from. *Cold Spring Harb. Lab. Press Method*. 2015; 1: 1–31.
54. Richter M, Rosselló-Móra R, Glöckner F, Peplies J. JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics*. 2016; 32: 929–931.
55. Konstantinidis KT, Rosselló-Móra R, Amann R. Uncultivated microbes in need of their own taxonomy. *ISME J*. 2017; 11: 2399–2406.
56. Grissa I, Vergnaud G, Pourcel C. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res*. 2007; 35: 52–57.
57. Wilm A, Aw PPK, Bertrand D, Yeo GHT, Ong SH, et al. LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Res*. 2012; 40: 11189–11201.
58. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010; 26: 841–842.
59. Koonin EV, Senkevich TG, Dolja VV. The ancient virus world and evolution of cells. *Biol. Direct*. 2006; 1: 27.
60. Lopes A, Tavares P, Petit MA, Guérois R, Zinn-Justin S. Automated classification of tailed bacteriophages according to their neck organization. *BMC Genomics*. 2014; 15: 1027.
61. Huang HW, Mullikin JC, Hansen NF. Evaluation of variant detection software for pooled next-generation sequence data. *BMC Bioinformatics*. 2015; 16: 235.
62. Bolhuis H, Cretoiu MS, Stal LJ. Molecular ecology of microbial mats. *FEMS Microbiol. Ecol*. 2014; 90: 335–350.
63. Miller SR, Strong AL, Jones KL, Ungerer MC. Bar-coded pyrosequencing reveals shared bacterial community properties along the temperature gradients of two alkaline

- hot springs in Yellowstone National Park. *Appl. Environ. Microbiol.* 2009; 75: 4565–4572.
64. Maniloff J, Ackermann HW. Taxonomy of bacterial viruses: establishment of tailed virus genera and the order Caudovirales. *Arch. Virol.* 1998; 143: 2051–2063.
 65. Pesant S, Not F, Picheral M, Kandels-Lewis S, Le Bescot N, et al. Open science resources for the discovery and analysis of Tara Oceans data. *Sci. Data.* 2015; 2: 150023.
 66. Zhang W, Zhou J, Wang Y. Four novel algal virus genomes discovered from Yellowstone Lake metagenomes. *Sci. Rep.* 2015; 5: 15131.
 67. Sohn MB, An L, Pookhao N, Li Q. Accurate genome relative abundance estimation for closely related species in a metagenomic sample. *BMC Bioinformatics.* 2014; 15: 242.
 68. Macklaim JM, Fernandes AD, Di Bella JM, Hammond JA, Reid G, et al. Comparative meta-RNA-seq of the vaginal microbiota and differential expression by *Lactobacillus iners* in health and dysbiosis. *Microbiome.* 2013; 1: 12.
 69. Thingstad TF, Vage S, Storesund JE, Sandaa RA, Giske J. A theoretical analysis of how strain-specific viruses can control microbial species diversity. *Proc. Natl. Acad. Sci. U.S.A.* 2014; 111: 7813–7818.
 70. Zeigler-Allen L, McCrow JP, Ininbergs K, Dupont CL, Badger JH, et al. The baltic sea virome: diversity and transcriptional activity of DNA and RNA viruses. *mSystems.* 2017; 2: e125-16.
 71. Shestakova SV, Karbysheva EA. The role of viruses in the evolution of Cyanobacteria. *Biol. Bull. Rev.* 2015; 5: 527–537.
 72. Weinbauer MG, Rassoulzadegan F. Are viruses driving microbial diversification and diversity? *Environ. Microbiol.* 2004; 6: 1–11.
 73. Avrani S, Wurtzel O, Sharon I, Sorek R, Lindell D. Genomic island variability facilitates *Prochlorococcus*-virus coexistence. *Nature.* 2011; 474: 604–608.
 74. Suttle CA. Cyanophages and their role in the ecology of cyanobacteria. In: BA Whitton, M Potts, editors. *The Ecology of Cyanobacteria.* Dordrecht: Springer. 2000.

75. Kristensen DM, Mushegian AR, Dolja VV, Koonin EV. New dimensions of the virus world discovered through metagenomics. *Trends Microbiol.* 2010; 18: 11–19.
76. Chénard C, Wirth JF, Suttle CA. Viruses infecting a freshwater filamentous cyanobacterium (*Nostoc* sp.) encode a functional CRISPR array and a proteobacterial DNA polymerase B. *mBio.* 2016; 7: e00667–00683.
77. Paez-Espino D, Eloë-Fadrosh EA, Pavlopoulos GA, Thomas AD, Huntemann M, et al. Uncovering Earth's virome. *Nature.* 2016; 536: 425–430.
78. Roux S, Hallam SJ, Woyke T, Sullivan MB. Viral dark matter and virus-host interactions resolved from publicly available microbial genomes. *Elife.* 2015; 4: e08490.
79. Voorhies AA, Eisenlord SD, Marcus DN, Duhaime MB, Biddanda BA, et al. Ecological and genetic interactions between cyanobacteria and viruses in a low-oxygen mat community inferred through metagenomics and metatranscriptomics. *Environ. Microbiol.* 2016; 18: 358–371.
80. Liu X, Shi M, Kong S, Gao Y, An C. Cyanophage Pf-WMP4, a T7-like phage infecting the freshwater cyanobacterium *Phormidium foveolarum*: complete genome sequence and DNA translocation. *Virology.* 2007; 366: 28–39.
81. Liu X, Kong S, Shi M, Fu L, Gao Y, et al. Genomic analysis of freshwater cyanophage Pf-WMP3 infecting cyanobacterium *Phormidium foveolarum*: the conserved elements for a phage. *Microb. Ecol.* 2008; 56: 671–680.
82. Zhou Y, Lin J, Li N, Hu Z, Deng F. Characterization and genomic analysis of a plaque purified strain of cyanophage PP. *Virol. Sin.* 2013; 28: 272–279.
83. Ou T, Liao XY, Gao XC, Xu XD, Zhang QY. Unraveling the genome structure of cyanobacterial podovirus A-4L with long direct terminal repeats. *Virus Res.* 2015; 203: 4–9.
84. Dunn JJ, Studier FW, Gottesman M. Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. *J. Mol. Biol.* 1983; 166: 477–535.
85. Hardies SC, Comeau AM, Serwer P, Suttle CA. The complete sequence of marine bacteriophage VpV262 infecting *Vibrio parahaemolyticus* indicates that an ancestral

- component of a T7 viral supergroup is widespread in the marine environment. *Virology*. 2003; 310: 359–371.
86. Chen F, Lu J. Genomic sequence and evolution of marine cyanophage P60: a new insight on lytic and lysogenic phages genomic sequence and evolution of marine cyanophage P60: a new insight on lytic and lysogenic phages. *Appl. Environ. Microbiol.* 2002; 68: 2589–2594.
 87. Labonté JM, Reid KE, Suttle CA, Labont JM, Reid KE, et al. Phylogenetic analysis indicates evolutionary diversity and environmental segregation of marine podovirus DNA polymerase gene sequences. *Appl. Environ. Microbiol.* 2009; 75: 3634–3640.
 88. Maranger R, Bird DF. Viral abundance in aquatic systems: a comparison between marine and fresh waters. *Mar. Ecol. Prog. Ser.* 1995; 121: 217–226.
 89. López-López O, Cerdán M, González-Siso M. Hot spring metagenomics. *Life*. 2013; 3: 308–320.
 90. Howard-Varona C, Roux S, Dore H, Solonenko NE, Holmfeldt K, et al. Regulation of infection efficiency in a globally abundant marine Bacteriodes virus. *ISME J.* 2017; 11: 284–295.
 91. Mengoni A, Tatti E, Decorosi F, Viti C, Bazzicalupo M, et al. Comparison of 16S rRNA and 16S rDNA T-RFLP approaches to study bacterial communities in soil microcosms treated with chromate as perturbing agent. *Microb. Ecol.* 2005; 50: 375–384.
 92. Kashtan N, Roggensack SE, Rodrigue S, Thompson JW, Biller SJ, et al. Single-cell genomics reveals hundreds of coexisting subpopulations in wild *Prochlorococcus*. *Science* (80-). 2014; 344: 416–420.
 93. Sano EB, Wall CA, Hutchins PR, Miller SR. Ancient balancing selection on heterocyst function in a cosmopolitan cyanobacterium. *Nat. Ecol. Evol.* 2018; 2: 510–519.
 94. Snyder JC, Bateson MM, Lavin M, Young MJ. Use of cellular CRISPR (clusters of regularly interspaced short palindromic repeats) spacer-based microarrays for detection of viruses in environmental samples. *Appl. Environ. Microbiol.* 2010; 76: 7251–7258.

95. Andersson AF, Banfield JF. Virus population dynamics and acquired virus resistance in natural microbial communities. *Science* (80-). 2008; 320: 1047–1050.
96. Stern A, Mick E, Tirosh I, Sagy O, Sorek R. CRISPR targeting reveals a reservoir of common phages associated with the human gut microbiome. *Genome Res.* 2012; 22: 1985–1994.
97. Sanguino L, Franqueville L, Vogel TM, Larose C. Linking environmental prokaryotic viruses and their host through CRISPRs. *FEMS Microbiol. Ecol.* 2015; 91: 1–9.
98. Brouns SJJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJH, et al. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* (80-). 2008; 321: 960–964.
99. Ye Y, Zhang Q. Characterization of CRISPR RNA transcription by exploiting stranded metatranscriptomic data. *RNA.* 2016; 22: 945–956.
100. Rodriguez-Valera F, Martin-Cuadrado AB, Rodriguez-Brito B, Pasi  L, Thingstad TF, et al. Explaining microbial population genomics through phage predation. *Nat. Rev. Microbiol.* 2009; 7: 828–836.
101. Staals RHJ, Jackson SA, Biswas A, Brouns SJJ, Brown CM, et al. Interference-driven spacer acquisition is dominant over naive and primed adaptation in a native CRISPR-Cas system. *Nat. Commun.* 2016; 7: 12853.
102. Westra ER, Dowling AJ, Broniewski JM, van Houte S. Evolution and Ecology of CRISPR. *Ann. Rev. Ecol. Evol. Syst.* 2016; 47: 307–331.
103. Paez-Espino D, Morovic W, Sun CL, Thomas BC, Ueda KI, et al. Strong bias in the bacterial CRISPR elements that confer immunity to phage. *Nat. Commun.* 2013; 4: 1430–1437.
104. Shmakov SA, Sitnik V, Makarova KS, Wolf YI, Severinov KV, et al. The CRISPR spacer space is dominated by cross the CRISPR spacer space is dominated. *mBio.* 2017; 8: 1–18.

Supplementary Material

The Supplementary Material can be accessed online at:
https://videleaf.com/wp-content/uploads/2020/03/PAMICR-19-08_Supplementary-Material.pdf