

A novel freshwater PP-like cyanophage with an unknown source endonuclease VII gene

Chen Cheng

Hubei University of Technology

Cheng Kai

20121041@hbut.edu.cn

Hubei University of Technology

Short Report

Keywords: Blue-green algae virus, Cyanophage, Endonuclease VII, Horizontal gene transfer

Posted Date: October 16th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-8771459/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

In this study, a novel cyanophage, designated PP-DH22, infecting *Plectonema boryanum* IU594 and *Phormidium foveolarum* IU427, was successfully isolated from a water sample collected from the surface waters of East Lake in Wuhan. Subsequent genomic sequencing and comparative analysis revealed that PP-DH22 shares significant organizational similarities with the previously reported cyanophage PP, including an endonuclease VII gene. However, a pivotal and unexpected divergence was discerned specifically within its 39th open reading frame. Detailed BLAST analysis showed that this gene exhibits a relatively high nucleotide sequence identity (45.65%) to the endonuclease VII gene encoded by the Escherichia coli-infecting phage HZP2. This stands in stark contrast to the extremely low identity (30.66%) observed when the same gene is compared to its putative homolog in the cyanophage PP reference genome. Collectively, these findings indicate that PP-DH22 might possess a broader and potentially more diverse host range beyond the currently identified cyanobacterial strains, possibly encompassing other bacterial phyla.

INTRODUCTION

Cyanophages, which are widespread in aquatic ecosystems, are viruses that target and often lyse cyanobacteria, a group of photosynthetic microorganisms[1]. In certain marine regions, cyanophages can account for up to 21% of double-stranded DNA (dsDNA) viruses[2]. These viral populations exhibit seasonal abundance equilibria with cyanobacterial hosts, playing pivotal roles in: (i) modulating cyanobacterial population dynamics and biodiversity[3, 4], (ii) regulating aquatic primary productivity[5], and (iii) driving biogeochemical cycling processes[6–7]. Furthermore, cyanophages influence genomic diversity and evolutionary trajectories of both hosts and phages through horizontal gene transfer (HGT) mechanisms and antagonistic coevolution[8–9].

In order to deeply understand the specific mechanisms by which cyanophage regulate host populations, drive gene flow and evolution, it is particularly important to conduct detailed studies on specific groups of these cyanophages. Among them, a type of PP-like cyanophages that infect filamentous cyanobacteria (especially in the Oscillatoriales family) is of particular significance. This is because it has been successively discovered in various freshwater environments, shares a common host range and similar genomic characteristics, providing an ideal model for exploring the geographical distribution, host adaptation and co-evolutionary relationships of this type of cyanophages. The cyanophage PP was isolated from the surface waters of eutrophic water in Wuhan, Hubei Province, China in 2002 with a genomic size of 42,480 bp, and its hosts were *Plectonema boryanum* and *Phormidium foveolarum*[10]. Similar to PP, the cyanophage LPP was isolated from a waste pond in the United States in 1963 with a genomic size of 42,319 bp, and its hosts were genera *Lyngbya*, *Plectonema*, and *Phormidium*[11]. Also, the cyanophage Pf-WMP3 was isolated from a lake in Beijing with a genomic size of 43,249 bp, China in 2003, and its hosts were *P. boryanum* and *P. foveolarum*[12]. Thus, PP-like cyanophages, which includes PP, LPP and Pf-WMP3, shared common host range in Oscillatoriales and similar genomic size, and widely distributed in freshwaters[13].

Endonuclease VII, as a resolvase enzyme, plays a crucial role in genetic recombination by converting branched DNA intermediates into linear duplex structures[14]. This double-stranded endonuclease identifies DNA distortions and cleaves at these sites, generating staggered cuts typically spaced approximately 3–6 bp apart [15]. Uptodate, endonuclease VII was reported to be encoded by many virus, including bacteriophage T4[14] and cyanophage PP[10]. Previous studies have identified an endonuclease VII gene in the genome of cyanophage PP, which shows high similarity to that of vibriophage ICP2 2006 A[10]. However, in-depth studies on the endonuclease VII gene in PP-like cyanophages are still lacking.

To this end, a cyanophage, PP-DH22, capable of infecting *P. boryanum* and *P. foveolarum*, was isolated from a freshwater lake in Wuhan City, Hubei Province, China and the trace of a second endonuclease VII gene in its genome was studied in the present work.

1 Materials and methods

1.1 Cyanobacteria species and their culture method

P. boryanum IU594, *P. foveolarum* IU427, *Microcystis aeruginosa* FACHB-469, FACHB-905, FACHB-1027, FACHB-927, PCC-7806 and *Spirulina platensis* FACHB-439 were all sourced from the Freshwater Conservation Center of the Institute of Hydrobiology (Chinese Academy of Sciences Algae Culture Collection at the Institute of Hydrobiology, FACHB). All strains were maintained in BG11 medium under controlled conditions: 25°C incubation temperature, 2000 lx light intensity, and a 12 h:12 h light-dark photoperiod[13].

1.2 Collection of water samples and enrichment of cyanophages

During October 2023, fourteen surface water (0.3m ~ 0.5m below water surface) samples were systematically collected from distinct aquatic ecosystems in Wuhan, including East Lake (114.37°E, 30.54°N) and Tangxun Lake (114.37° E, 30.45° N). Following field collection, samples were immediately filtered through 0.22 µm polyethersulfone (PES) membrane filters under sterile conditions. The resulting filtrates were preserved at 4°C during transport to the laboratory for subsequent analysis.

To initiate cyanophage propagation, 1 mL of each filtered sample was inoculated into 9 mL of IU594 cyanobacterial culture in logarithmic growth phase (cultivate for 7 days). The mixtures were maintained under controlled environmental conditions as previously described in Section 1.1. Sequential propagation cycles were implemented upon observable culture clarification (indicated by chromatic transition to yellow), with 1 mL aliquots from lysed cultures being serially transferred to fresh logarithmic-phase cyanobacterial suspensions. This iterative amplification protocol was rigorously performed through seven successive propagation cycles to ensure effective enrichment of cyanophage particles[16].

1.3 Purification of cyanophage

Following enrichment, chloroform (10% v/v) was introduced into the cyanophage lysate, followed by vigorous vortex mixing for 5 minutes. The mixture was allowed to settle at 4°C for 2 hours, after which the supernatant was collected by centrifugation and filtered through a 0.22 µm PES membrane[13]. The filtrate underwent serial logarithmic dilution in BG11 medium to achieve a final dilution factor of 10⁹.

For phage adsorption, 0.1 mL of the diluted cyanophage suspension was combined with 1.9 mL of logarithmic-phase IU594 culture. The mixture was incubated at 25°C for 20 minutes to facilitate phage-host interaction, followed by purification using the double-layer agar plaque assay[17]. Plates were incubated under standard conditions for 24–48 hours, after which the highest dilution gradient exhibiting discrete plaques was selected for further processing.

Individual plaques were excised and resuspended in 1 mL BG11 medium. Seven successive rounds of plaque purification were performed to obtain isolates with morphologically consistent plaque morphology. A single purified plaque was then amplified in 100 mL of logarithmic-phase IU594 culture under conditions specified in Section 1.1 until complete culture lysis.

The resultant lysate was treated with chloroform (1% v/v final concentration), filtered through a 0.22 µm PES membrane, and stored at 4°C as a stabilized phage stock for subsequent applications[13].

1.4 Determination of host range

A 1 mL aliquot of cyanophage lysate passed through 0.22 µm PES membrane filters under sterile conditions and was inoculated into 10 mL of different logarithmic-phase cyanobacteria cultures listed in Section 1.1. The suspensions were incubated under standardized culture conditions as detailed in Section 1.1, and the color change of the tested cyanobacterial solutions was observed to determine the occurrence of the viral lysis.

1.5 Genomic sequencing and annotation

The lysate passed through 0.22 µm PES membrane filters under sterile conditions was sent to Chengdu Xinhengji Biotechnology Co., Ltd. for whole genome second-generation sequencing and splicing: The concentrated phage sample was passed through a 0.22 µm filter. The filtrate was treated with 200 U Benzo DNase (NOVOPROTEIN SCIENTIFIC INC) and 0.1 mg/ ml RNase A followed by heat inactivation of DNases at 65 °C for 10 min. Phage genome DNA was extracted by Qiagen MinElute Virus Spin Kit following the manufacture's instruction. DNA concentration was quantified on a Qubit 3.0 Fluorometer (ThermoFisher Scientific) using Equalbit 1x dsDNA HS Assay Kit. The genome DNAsize was checked by 0.8% agarose gel electrophoresis as well. Shotgun libraries were constructed using VAHTS® Universal Plus DNA Library Prep Kit for Illumina. The libraries were quality-checked by an Agilent 4200 Bioanalyzer and sequenced on the DNBSEQ-T7 platform with 2 x 150bp pair-end reads. Qualified nucleic acid samples were used to construct DNA next-generation sequencing (NGS) libraries, obtaining NGS libraries with insert fragment sizes ranging from 200 to 500 bp. Illumina Novaseq 6000 platform was used for sequencing, and fastp (0.24.0) was used to filter and quality control the obtained original sequencing data[18] to obtain clean reads, which were assembled by de novo using metaSP Ades software (v4.0.0). At the same time, different k-mer lengths were selected for testing and optimal assembly results were obtained[19]. Functional annotation of predicted genes was performed using pharokka online platform

(run_pharokka_and_phold_and_phyteny.ipynb - Colab), eggNOG-mapper v5.0.2[20] and the ViPTree 4.0 online platform (accessed January 2024; <https://www.genome.jp/viptree/>)[21]. Genomic similarity analysis to assess the genomic relatedness between the two strains, the Average Nucleotide Identity (ANI) was calculated using the OrthoANlu algorithm provided by the EzBioCloud web server (<https://www.ezbiocloud.net/tools/ani>). The FASTA files of the respective genome sequences were uploaded to the platform, and the ANI values were computed to determine the degree of sequence identity[22].

1.6 Phylogenetic Analysis

To determine the taxonomic position and analyze genomic synteny, the complete genome sequence of phage PP-DH22 was analyzed using the ViPTree server (<http://www.genome.jp/viptree/>). ViPTree generated a viral proteomic tree based on genome-wide sequence similarities computed by tBLASTx. Closely related viral genomes identified by the proteomic tree were automatically selected for comparative alignment. Genomic organization and sequence homology were visualized in a linear alignment plot, where colored connecting lines indicate the tBLASTx percentage identity between homologous regions.

To comprehensively investigate the phylogenetic relationships of the target viral gene, we implemented a dual-track strategy combining broad-scale sequence retrieval with high-resolution Sequence Similarity Network (SSN) analysis. Initially, a global SSN was constructed using the EFI-Enzyme Similarity Tool (EFI database version: 2025_03 / 106) to survey the protein sequence space. A BLAST search (Option A) was performed against the UniProt database using the target sequence as a query ($E\text{-value} \leq 10^{-5}$), with a maximum retrieval limit of 1,000 sequences. To enhance structural clarity and computational efficiency, a Representative Node Network was generated at a 40% identity clustering level. This network was visualized in Cytoscape (v3.10.4), where the "First Neighbors" selection tool was applied to the query sequence under stringent alignment score filters. This down-sampling process yielded a refined subset of approximately 200 closely related sequences for subsequent multiple sequence alignment and tree reconstruction.

In parallel, to achieve a more granular resolution within specific viral clades, a secondary, targeted dataset was curated. This involved a BLASTP search against the NCBI Virus database to retrieve the top 100 hits, followed by a rigorous quality control step where sequences annotated as "hypothetical proteins" or originating from "unidentified viruses" were manually excluded. This core set was further augmented by integrating manually curated homologs from PP and PP-DH22, along with representative sequences from the PP and LPP lineages. The integrated dataset was then processed through the EFI-EST pipeline to generate a specialized, high-resolution SSN. Unlike the global survey, all sequences retrieved in this targeted stage were retained to capture the complete evolutionary trajectories and cluster bifurcations among closely related viral clades. Ultimately, two distinct SSN models were utilized: a pruned global network for key sequence identification and an inclusive targeted network for the detailed comparative analysis of specific viral lineages.

Related viral genomes were identified and analyzed using the ViPTree 4.0 server (<https://www.genome.jp/viptree/>). The dataset was established by querying the internal dsDNA prokaryotic virus database, specifically selecting all viral entries associated with the host phylum Cyanobacteriota. To serve as a reference, the genome of Enterobacteria phage T7 was manually incorporated into the analysis. Subsequently, a proteomic tree was generated based on genome-wide sequence similarities calculated by tBLASTx. Open reading frames (ORFs) obtained from the annotations were further analyzed using the BLASTx online platform and the HMMER sequence alignment tool (hmmsearch) (accessed January 2024; <https://www.ebi.ac.uk/Tools/hmmer/search/hmmsearch>). Functional annotations were assigned based on multiple standard databases, including COG, GO, KEGG, CAZy, BiGG, Pfam, and the NCBI non-redundant (nr) database.

Molecular evolutionary analyses were conducted using MEGA11[23]. Nucleotide sequences were aligned by codons using the ClustalW algorithm implemented in MEGA11 with default gap penalty parameters (gap opening penalty = 15.00). Phylogenetic trees were reconstructed using the maximum-likelihood (ML) method, and evolutionary distances were calculated using the maximum composite likelihood model. The pairwise deletion option was applied to handle gaps and missing data, and all codon positions (1st, 2nd, 3rd, and non-coding sites) were included in the analysis.

2 Results and analysis

2.1 Cyanophage and host range

The cyanophage isolated from the surface water of a sub lake of East Lake (114.37°E, 30.54°N) exhibited a narrow host range, demonstrating lytic activity exclusively against *P. boryanum* IU594 and *P. foveozarum* IU427, which could yellow the cyanobacterial fluid of IU594 within 24h (Supplementary Fig. 1a) and form circular plaque on the plate of IU594 (Supplementary Fig. 1b). The cyanophage was named after its two known hosts as cyanophage PP-DH22.

2.2 Genome structure analysis

After completing the nucleic acid extraction and sequencing, the final assembly was conducted with a 127-mer length, and there was only one contig. The cyanophage PP-DH22 was determined to possess a linear double-stranded DNA genome (GenBank accession number PP942118, the raw sequence data submission number SRR36145945) spanning 42,653 base pairs with a GC content of 46.36%, which is highly similar to known PP-like cyanophages (Table 1).

Table 1
Compares of the genome characters of PP-like cyanophages.

Cyanophage	Genome accession number in NCBI	Genomic size (bp)	GC Content	Number of ORFs	ANI with PP-DH22
PP-DH22	PP942118	42653	46.36%	42	-
PP	KF598865.1	42480	46.41%	41	99.97%
Pf-WMP3	NC_009551.1	43249	46.49%	41	95.42%
LPP-2	NC_111529.1	42319	-	38	94.29%
- Not available					

The genome of cyanophage PP-DH22 contains 42 closely packed ORFs, of which 18 ORFs are predicted to encode proteins with known functions, accounting for 43% of the total number of ORFs (Table 2).

Table 2
ORFs in the genome of PP-DH22 and their predictive functions annotated by pharokka

ORF	Position	Length	Predicted protein	Related phage(s) or microbes	E value ^b
1	1-453	151	-	-	
2	542-1102	187	-	-	
3	1099-1335	79	-	-	
4	1313-1516	68	-	-	
5	1523-1630	36	-	-	
6	1717-2970	418	-	-	
7	3060-3605	182	-	-	
8	3724-4392	223	-	-	
9	4392-4901	170	-	-	
10	5145-6404	420	-	-	
11	6535-8583	683	DNA primase/helicase	Cyanophage PP,LPP-2 st. SPI, Pf-WMP3	0
12	8644-8844	67	-	-	
13	8857-9063	69	-	-	
14	9067-10935	623	DNA polymerase I	Cyanophage PP,LPP-2 st. SPI, Pf-WMP3	0
15	10917-11072	52	-	-	
16	11047-11178	44	-	-	
17	11265-11510	82	-	-	
18	11511-12542	344	-	-	
19	12546-12668	41	-	-	
20	12847-13029	61	-	-	
21	13026-13268	8	-	-	
22	13276-13779	168	Erf-like ssDNA annealing protein	Cyanophage PP,LPP-2 st. SPI, Pf-WMP3	2.00E-120
23	13792-14001	70	-	-	
24	14015-14302	96	-	-	
25	14381-15241	287	Cas4-domain exonuclease	Cyanophage PP, Pf-WMP3	0
26	15305-16507	401	Tail fiber protein	Cyanophage PP,LPP-2 st. SPI, Pf-WMP3	0
27	16494-19955	1154	Tail fiber protein; host specificity	Cyanophage PP,LPP-2 st. SPI, Pf-WMP3	0
28	20053-25191	1713	ParB-like partition nuclease	Cyanophage PP,LPP-2 st. SPI, Pf-WMP3	0
29	25239-28223	995	Peptidase	Cyanophage PP,LPP-2 st. SPI, Pf-WMP3	0
30	28225-29415	397	Internal protein	Cyanophage PP,LPP-2 st. SPI, Pf-WMP3	0
31	29419-32502	1027	Tail protein	Cyanophage PP,LPP-2 st. SPI, Pf-WMP3	0
32	32495-32857	121	Endonuclease VII	Cyanophage PP	2.00E-82
33	32862-33527	222	Tail protein	Pf-WMP3	1.00E-161
- hypothetical protein					

ORF	Position	Length	Predicted protein	Related phage(s) or microbes	E value ^b
34	33970–35115	382	Major head protein	Cyanophage PP,LPP-2 st. SPI, Pf-WMP3	0
35	35138–35821	228	Head scaffolding protein	Cyanophage PP, Pf-WMP3	4.00E-143
36	35826–37781	652	Head-tail adaptor	Cyanophage PP,LPP-2 st. SPI, Pf-WMP3	0
37	37796–38221	142	HNH endonuclease	Cyanophage PP	2.00E-101
38	38258–39952	565	Terminase	Cyanophage PP,LPP-2 st. SPI, Pf-WMP3	0
39	39945–40358	138	Endonuclease VII	Escherichia phage HZP2	2.00E-28
40	40465–40965	167	-	-	
41	40971–41456	162	-	-	
42	41477–41731	85	-	-	
- hypothetical protein					

Since the genome of cyanophage PP-DH22 is most similar to that of cyanophage PP with an ANI of 99.97% (Table 1), the ORFs of cyanophage PP-DH22 and cyanophage PP are compared in detail in Fig. 1. In general, the ORFs of PP-DH22 and PP is very similar, while the only difference is that: The ORF 39 of PP-DH22 is a endonuclease VII, while the ortholog availability ORF in PP is a gene with unknown function.

Phylogenetic tree was constructed according to the complete genome (Supplementary Fig. 2, 3), indicating that cyanophage PP-DH22 is closely related to cyanophage PP and Pf-WMP3, as well as to cyanophage A-4L and Pf-WMP4.

2.3 A second endonuclease VII gene

Similar to cyanophage PP, cyanophage LPP and cyanophage Pf-WMP3, cyanophage PP-DH22 has an endonuclease VII gene (ORF 32) at 32495-32857bp. However, to our surprise, cyanophage PP-DH22 also has another endonuclease VII gene at ORF 39 (annotated by pharokka, eggNOG-mapper, blastx and Hmmer). Moreover, although the orthologs ORF (to the ORF 39 of PP-DH22) in cyanophage LPP and cyanophage Pf-WMP3 are similar in length (Table 2, Fig. 1), but their identity to the OFF 39 PP-DH22 is very low, and were all annotated as unknown genes by the four annotation tools (Table 3).

Table 3
Annotations of the ORF 39 in the genome of cyanophage PP-DH22 and its orthologs in the genomes of cyanophage PP, LPP and Pf-WMP3 by different annotation tools

cyanophage	amino acid length	eggNOG-mapper		blastx		Hmmer		VIPTree	
		Description	E-value	Description	E-value	Description	E-value	Description	E-value
PP-DH22	138	recombination endonuclease VII	1.37e-23	endonuclease VII	2e-28	endonuclease VII	1.1e-36	endonuclease VII	3.3e-23
PP	166	hypothetical protein	-	hypothetical protein	3e-62	hypothetical protein	8.2e-28	hypothetical protein	2e-57
LPP	166	hypothetical protein	-	hypothetical protein	1e-107	hypothetical protein	1.0e-26	hypothetical protein	-
Pf-WMP3	166	hypothetical protein	-	hypothetical protein	6e-111	hypothetical protein	9.9e-28	hypothetical protein	-
- Not available									

The evolutionary tree constructed for PP-DH22 ORF 39 was shown in Fig. 2. It was found that gene was more related to endonuclease VII gene of *Escherichia coli* phage HZP2 with an identity 45.65% (by blastX), and less related to its first endonuclease VII gene (at ORF 32) with an identity 30.66% (by blastX) and the endonuclease VII gene in cyanophage PP (at ORF 30) with an

identity 30.66% (by blastX). In another word, the second endonuclease VII gene of DH22 was relatively distantly related to its first endonuclease VII gene and the endonuclease VII gene in PP. Phylogenetic analysis via ERI-EST revealed that ORF39 of PP-DH22 exhibits a closer evolutionary affinity with the endonuclease VII of phages such as HZP2. Conversely, it is phylogenetically distant from the endonuclease VII and its homologs (encoded by ORF 38) in cyanophage PP (Fig. 3).

3 Discussion

The host range of cyanophage PP-DH22 closely matches cyanophage PP, demonstrating conserved infectivity characteristics. Among known cyanophages, PP is the closest relative to DH22. They exhibit extensive similarities ranging from their host specificity to their high ANI and conserved genomic organization

In this study, a second endonuclease VII gene (ORF 39) was located in PP-DH22, and its sequence was very different from the orthologs in other PP-like cyanophages. This result may be evidence of the HGT function of PP-like cyanophages, for two specific reasons:

Firstly, in terms of evolutionary relationship, ORF 39 is more closely related to genes of *E. coli* phage rather than its own endonuclease VII gene ORF 32 and other endonuclease VII genes in reported PP-like cyanophages, suggesting that the gene may originate from unknown hosts. Virus cross-species infection has been widely reported in the viral field. For instance, cyanophage Yong-L2-223 could infect 3 strains of cyanobacteria across 3 orders: Synechococcales, Chroococcales and Nostocales[16].

Secondly, endonuclease VII is a decomparting enzyme that converts branched DNA intermediates into double-stranded structures in the process of gene recombination. It is a double-stranded endonuclease enzyme that recognizes distortions in DNA and makes cuts in response to these perturbations, producing staggered cuts at approximately 3–6 bp intervals, and endonuclease VII showed broad specificity for DNA with secondary structure[15]. Thus, the recombinational function of endonuclease VII may lead to horizontal gene transmission.

In summary, although the origin of the ORF 39 in cyanophage PP-DH22 remains unclear, this gene exhibits a relatively high nucleotide sequence identity to the endonuclease VII gene encoded by the *Escherichia coli*-infecting phage HZP2, rather than the relative low identity to its putative homolog in the genomes of other PP-like cyanophages.

Declarations

Competing interests

The authors have no relevant financial or non-financial interests to disclose.

Author Contribution

Cheng Kai contributed to the study conception and design. Material preparation, data collection and analysis were performed by Chen Cheng. The manuscript was written by all authors, and all authors read and approved the final manuscript.

Acknowledgement

This study was supported by Major Project of Guangzhou National Laboratory, Grant No. GZNL2024A01007. This study thanks Li Yuena, Sun Fukang, Zhang Guang and Zhang Yongkang for their assistance and support in sample collection and technical discussion.

Data Availability

Whole genome sequences were submitted to the NCBI GenBank database as [PP942118] (<https://www.ncbi.nlm.nih.gov/nuccore/PP942118.1>). The raw sequence data reported in this paper have been deposited in the

References

1. Lin W, Li DF, Pan LT, Li MZ, Tong YG (2024) Cyanobacteria-cyanophage interactions between freshwater and marine ecosystems based on large-scale cyanophage genomic analysis. *Sci Total Environ* 950:175201. <https://doi.org/10.1016/j.scitotenv.2024.175201>
2. Carlson MCG, Ribalet F, Maidanik I, Durham BP, Hulata Y et al (2022) Viruses affect picocyanobacterial abundance and biogeography in the North Pacific Ocean. *Nat Microbiol* 7:570–580. <https://doi.org/10.1038/s41564-022-01088-x>
3. Konopka A, Lindemann S, Fredrickson J (2015) Dynamics in microbial communities: unraveling mechanisms to identify principles. *ISME J* 9:1488–1495. 10.1038/ismej.2014.251
4. Zimmerman AE, Howard-Varona C, Needham DM, John SG, Worden AZ et al (2020) Metabolic and biogeochemical consequences of viral infection in aquatic ecosystems. *Nat Rev Microbiol* 18:21–34. <https://doi.org/10.1038/s41579-019-0270-x>
5. Srivastava A, Gupta N, Mishra AK (2024) In: Cyanobacteria AK, Mishra (eds) Cyanophages: interacting mechanism and evolutionary significance. SS Sing. Academic, pp 255–282
6. Suttle CA (2007) Marine viruses—major players in the global ecosystem. *Nat Rev Microbiol* 5:801–812. <https://doi.org/10.1038/nrmicro1750>
7. Puxty RJ, Evans DJ, Millard AD, Scanlan DJ (2018) Energy limitation of cyanophage development: implications for marine carbon cycling. *ISME J* 12(5):1273–1286. <https://doi.org/10.1038/s41396-017-0043-3>
8. Gregory AC, Solonenko SA, Ignacio-Espinoza JC et al (2016) Genomic differentiation among wild cyanophages despite widespread horizontal gene transfer. *BMC Genomics* 17(1):930. <https://doi.org/10.1186/s12864-016-3286-x>
9. Xu Y, Zhang R, Wang N, Cai LL, Tong YG, Sun Q, Chen F, Jiao NZ (2018) Novel phage-host interactions and evolution as revealed by a cyanomyovirus isolated from an estuarine environment. *Environ Microbiol* 20(8):2974–2989. <https://doi.org/10.1111/1462-2920.14326>
10. Zhou YR, Lin J, Li N, Hu ZH, Deng F (2013) Characterization and genomic analysis of a plaque purified strain of cyanophage PP. *Virol Sin* 28(5):272–279. <https://doi.org/10.1007/s12250-013-3363-0>
11. Safferman RS, Morris ME (1963) Algal virus: isolation. *Science* 140(3567):679–680. <https://doi.org/10.1126/science.140.3567.679>
12. Kong SL, Liu XY, Fu LW, Yu XC, An CC (2012) I-PfoP3I: a novel nicking HNH homing endonuclease encoded in the group I intron of the DNA polymerase gene in *Phormidium foveolarum* phage Pf-WMP3. *PLoS ONE* 7(8):e43738. <https://doi.org/10.1371/journal.pone.0043738>
13. Zhao YJ, Cheng K, Shi ZL, Guo YX, Zhu HY, Zhang JH, Liu YD (2002) Isolation and identification of the first cyanophage (cyanobacteria virus) in China. *Prog Nat Sci* (09): 29–33
14. Mizuuchi K, Kemper B, Hays J, Weisberg RA (1982) T4 endonuclease VII cleaves holliday structures. *Cell* 29. 2357–365. [https://doi.org/10.1016/0092-8674\(82\)90152-0](https://doi.org/10.1016/0092-8674(82)90152-0)
15. Youil R, Kemper BW, Cotton RG (1995) Screening for mutations by enzyme mismatch cleavage with T4 endonuclease VII. *Proc Natl Acad Sci USA* 92(1):87–91. <https://doi.org/10.1073/pnas.92.1.87>
16. Sullivan MB, Lindell D, Lee JA, Thompson LR, Bielawski JP et al (2006) Prevalence and evolution of core photosystem II genes in marine cyanobacterial viruses and their hosts. *PLoS Biol* 4(8):e234. <https://doi.org/10.1371/journal.pbio.0040234>
17. Hu JQ, Pan LT, Zhou Q, Qian MH, Cai RQ, Wang Fei, Ren XQ, Lin W, Li DF, Tong YG (2024) Isolation and genome analysis of a novel *Synechococcus* cyanophage Yong-L2-223. *Acta Microbiol Sin* 64(01):189–207. <https://doi.org/10.13343/j.cnki.wsxb.20230365>
18. Suttle CA (1993) Enumeration and isolation of viruses. handbook of methods in aquatic microbial ecology, Boca Raton, FL
19. Chen SF, Zhou YQ, Chen YR, Gu J (2018) Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34(17):i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>

20. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA (2017) metaSPAdes: a new versatile metagenomic assembler. *Genome Res* 27(5):824–834. <https://doi.org/10.1101/gr.213959.116>

21. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J (2021) EggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol Biol Evol* 38(12):5825–5829. <https://doi.org/10.1093/molbev/msab293>

22. Zhang WJ, Ran Wang, Zou XX, Gu CW, Yang Q, He ML, Xiao WD, He LQ, Zhao MD, Yu ZH (2023) Comparative genomic analysis of alloherpesviruses: Exploring an available genus/species demarcation proposal and method. *Virus Res* Volume 33:199163. <https://doi.org/10.1016/j.virusres.2023.199163>

23. Potter SC, Luciani A, Eddy SR, Park YM, Lopez R, Finn RD (2018) Hmmer web server: 2018 update. *Nucleic Acids Res* 46(W1):w200–w204. <https://doi.org/10.1093/nar/gky448>

Figures

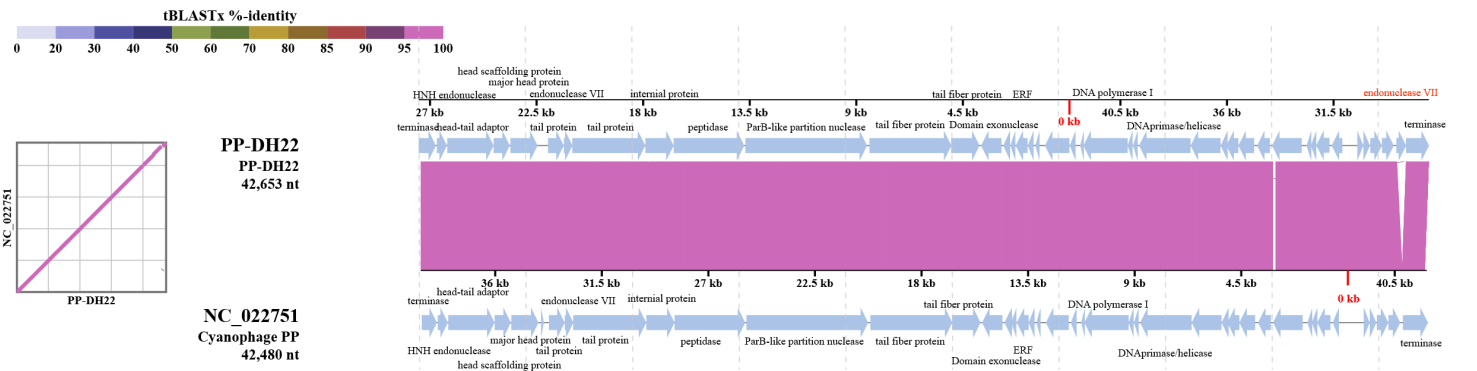


Figure 1

Genomic alignment and synteny analysis between cyanophage PP-DH22 and the reference strain Cyanophage PP.The comparative analysis was generated using the ViPTree server. (Left) A rectangular dot plot showing the collinearity and synteny between the two genomes. (Right) Linear genome alignment based on tBLASTx scores. The colored connecting blocks indicate regions of sequence similarity, with the color gradient representing the percentage of identity (purple indicates high identity, as shown in the scale bar). The arrows represent predicted open reading frames (ORFs) and indicate the direction of transcription.

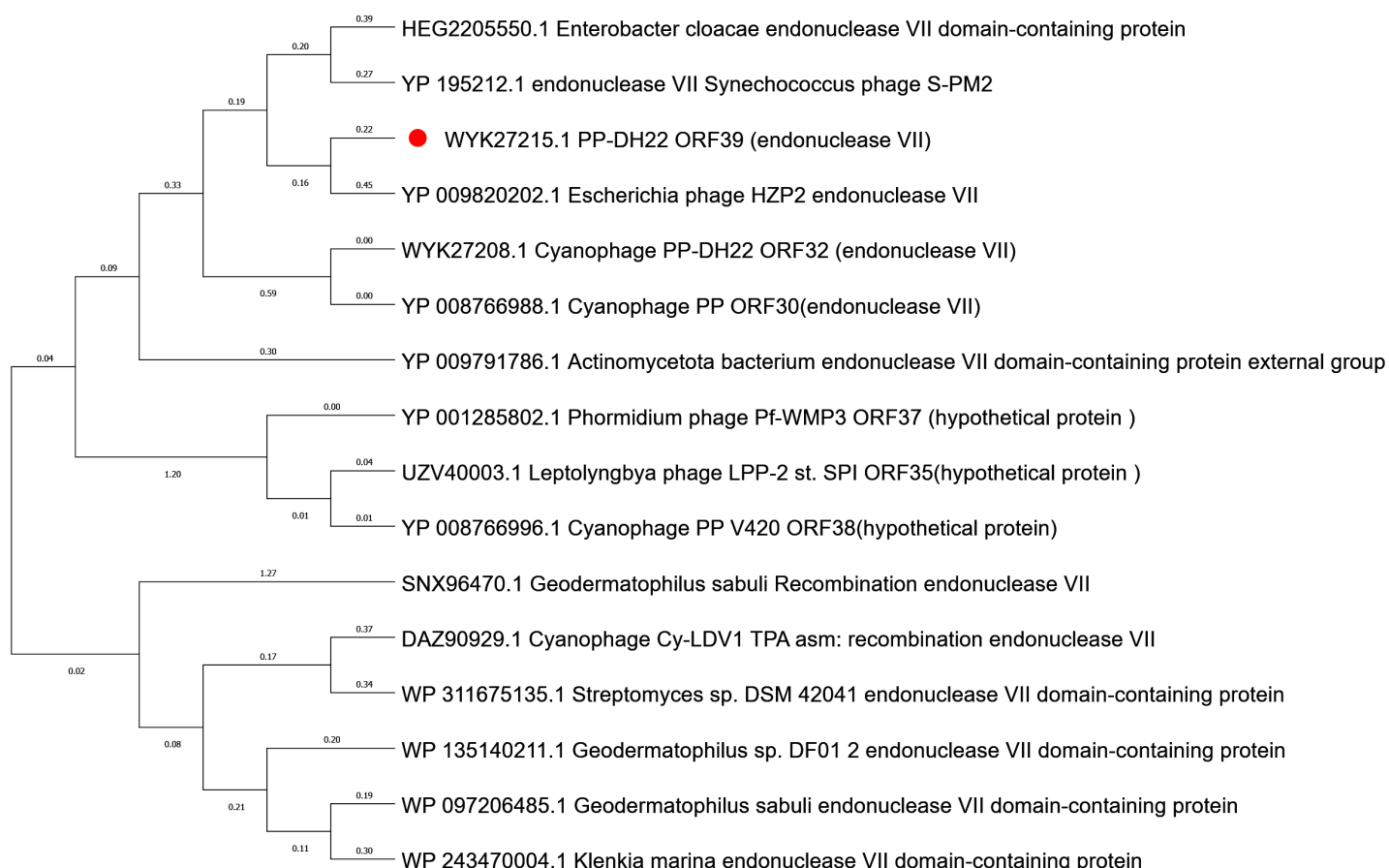


Figure 2

Phylogenetic analysis of the putative endonuclease VII protein (ORF39) from cyanophage PP-DH22. The phylogenetic tree was constructed using the Maximum Likelihood method implemented in MEGA11. The analysis involves the amino acid sequence of PP-DH22 ORF39 (indicated by a red dot) and its homologous sequences retrieved from the NCBI GenBank database (top BLAST hits and selected reference strains). The numbers on the branches indicate branch lengths, representing the number of substitutions per site.

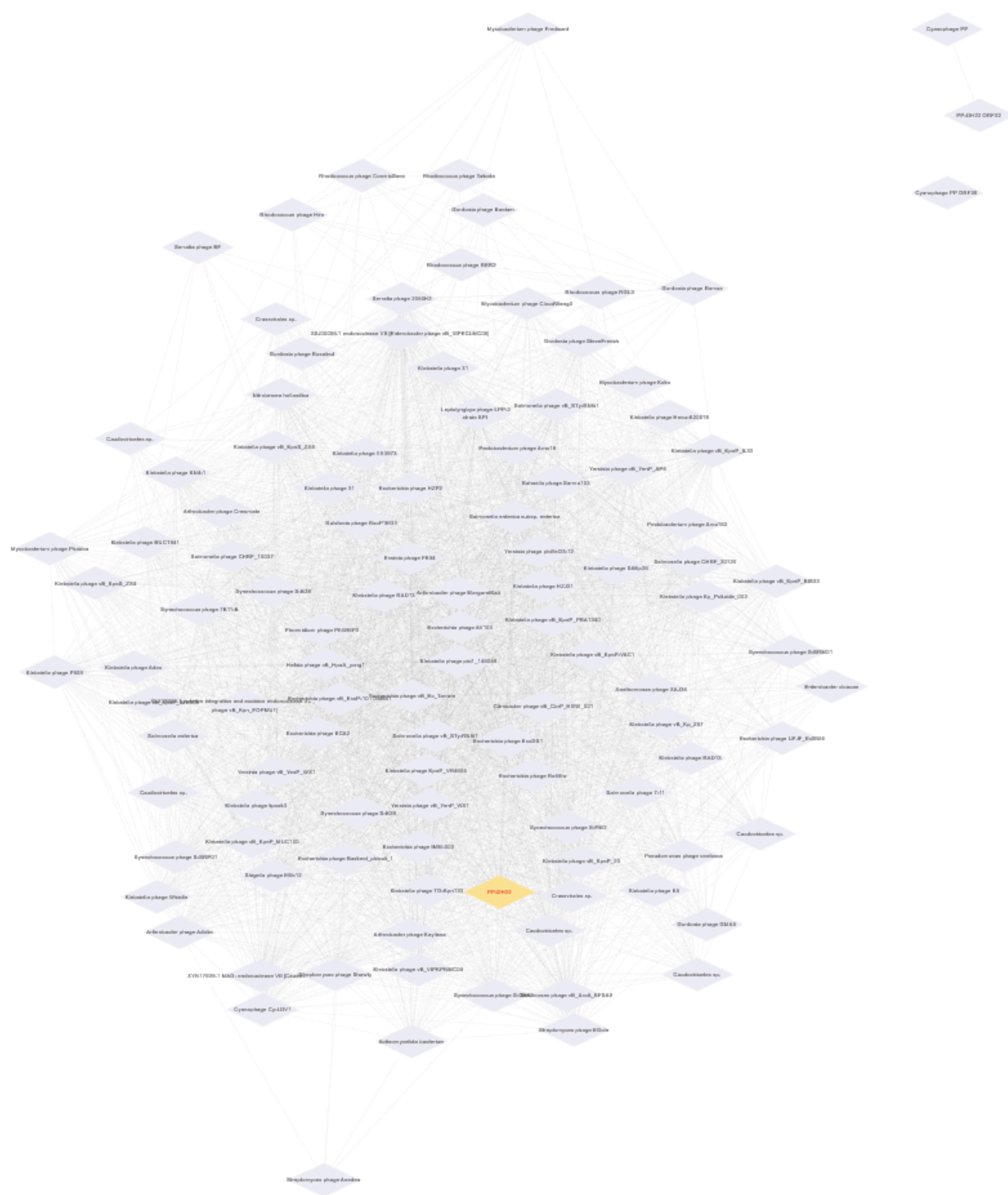


Figure 3

Sequence Similarity Network of the endonuclease VII from cyanophage PP-DH22 and its homologs. The network was constructed using the EFI-EST (Enzyme Function Initiative-Enzyme Similarity Tool) web server and visualized with Cytoscape. The dataset consists of the top 100 homologous sequences retrieved from NCBI BLAST, along with selected reference strains.

Supplementary Files

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

- [Supplementary.docx](#)