



A novel lineage of large aquatic bacteriophages identified through metagenomics

Carolina Perez-Hernandez¹ · Joud Aldaroub¹ · Zachary K. Barth¹ · Frank O. Aylward¹

Received: 6 June 2025 / Accepted: 15 September 2025 / Published online: 15 November 2025
© The Author(s) 2025

Abstract

“Jumbo phages” are tailed phages with genome sizes >200 kbp and physical dimensions reaching up to 0.45 μm . Although jumbo phages represent only a small fraction of the isolated phages to date, metagenomic surveys have shown that they are broadly distributed in a wide range of environments. In this study, we surveyed metagenomic data from aquatic systems and identified 25 genomes from a heretofore-undescribed lineage of jumbo phages with genomes reaching up to 307 kbp. We refer to these phages as “moraphages”, from the Gaelic word ‘móir’, for large. Moraphages represent a diverse lineage with inter-genome average amino acid identity (AAI) ranging from 39 to 95%, and our pan-genomic analysis identified only 26 viral orthologous groups (VOGs) found in at least 80% of the genomes. Our phylogenomic analysis suggests that moraphages are distant relatives of a recently described lineage of huge phages from marine sediment. Moraphages lack much of the genetic machinery found in other lineages of large phages, but they have a range of genes that may be used to take over host cellular machinery and subvert host defenses, such as glutamine synthetases, antitoxin genes, and chaperones. The predicted hosts of most moraphages are members of the phylum *Bacteroidota*, and some encode homologs of the chaperones DnaK and DnaJ that bear evidence of recent gene transfer from members of the order *Flavobacteriales*. Our work sheds light on the emerging diversity of large phages that are found across the biosphere.

Introduction

Bacteriophages, also called phages, are viruses that infect bacteria and are found in diverse environments. They are the most abundant and diverse biological entities on the planet, with estimates of 10^{31} phage particles existing on Earth [1]. Phages modulate the abundance and activities of microorganisms that make up a large component of the biomass on Earth [2] and have been used to advance our understanding of molecular, environmental, and evolutionary biology [3]. Phages have been isolated from soil, marine environments [4], wastewater [5], and even acidic hot springs [6], amongst other ecosystems, and have adapted to be the

ever-present predators of bacteria. Phages are a source of enormous genetic diversity but have gone relatively unnoticed when compared to other biological entities. Lytic phages control the population size of the bacterial host [7] and release valuable nutrients from within bacterial biomass into the environment that organisms at other trophic levels can then access [8, 9]. Obligate lytic phages are a potential tool that can be harnessed to combat the rising number of infections caused by bacteria that are resistant to commonly used antibiotics. This property and its applications extend to other realms as well, including biocontrol in agriculture and wastewater treatment.

Among the many phages that exist across all environments, those that have particularly large physical dimensions and genome lengths have remained largely unknown until recently. Referred to as “jumbo phages,” these predominantly lytic tailed phages have a genome size >200 kbp and have physical dimensions near 0.5 μm [10]. Isolation of large phages poses a number of challenges when using conventional methods. Somewhat counterintuitively, jumbo phages tend to form small plaques and are often overlooked in favor of more visible and easily isolated plaques. Plaque assays using higher percentages of

Handling editor Johannes Wittmann.

✉ Carolina Perez-Hernandez
caromex77@gmail.com

✉ Frank O. Aylward
faylward@vt.edu

¹ Department of Biological Sciences and Center for Emerging, Zoonotic, and Arthropod-borne Pathogens, Virginia Tech, 926 West Campus Drive, Blacksburg, VA, USA

agar prevent jumbo phage virions from diffusing through the agar matrix and forming visible plaques on the bacterial overlay [11]. Filters are commonly employed when preparing samples and may contribute to the bias against their detection and isolation. While removing bacteria and other organisms is necessary for all phage isolation, the use of filters may not be the optimal approach for jumbo phages [12].

When comparing jumbo phages to their small-genome counterparts, the makeup of their genome is quite different. Some jumbo phage lineages have been shown to have decreased dependence on host cellular machinery and often encode their own genes for genome replication and gene expression, including DNA polymerases and multi-subunit RNA polymerases (RNAPs) [13, 14]. There is a correlation between the presence of tRNAs and genome size, with the number of tRNAs encoded increasing as a genome increases in length [12]. The presence of tRNA genes in jumbo phage genomes might have arisen due to the host's own tRNA molecules becoming compromised during phage infection [15]. Other genes unique to jumbo phages include proteins that circumvent the bacterial CRISPR-Cas defense system. Certain jumbo phages have been found to form a selectively permeable "phage nucleus" within the bacterial cytosol, a structural feature thought to be associated with eukaryotic evolution [16]. This nucleus-like compartment acts as a physical shield that protects the phage genome from bacterial defenses when replication is underway. Some investigators have postulated that jumbo phages have a broader host range due to genes that reduce dependence on the host, allowing for efficient virion and genome production in many different strains or across species [17–19].

Due to the challenges of isolating jumbo phages from environmental samples, recent studies have used metagenomic methods for identifying jumbo phages. This approach is also challenging due to the difficulty of resolving large genomes, but continual advances in assembly algorithms over the last few decades have improved our ability to recover large phage genomes, and several approaches have been developed to make use of long-read sequencing, bin phage contigs, or iteratively resolve large scaffolds [20–22]. Metagenomic approaches have led to a dramatic increase in the number of complete jumbo phage genome sequences determined in recent years and have revealed some of the largest phage genomes ever identified [20, 23–26]. In this study, we sought to identify novel lineages of jumbo phages by mining metagenomic data, primarily from freshwater systems. We used a recently proposed phylogenomic framework for phage diversity [27] to demarcate a new group of 25 phage genomes. We present

a phylogenomic analysis of these phages and discuss the notable genes encoded in their genomes.

Methods

To search for large phage genomes across a range of different habitats, we surveyed 16,801 metagenomes from the SPIRE database [28]. We chose the metagenomes in this set to represent a wide range of environments, including freshwater ponds and lakes, marine habitats, soil and sediment samples, engineered systems such as wastewater treatment plants, and others. For this study, we downloaded assemblies that were already available in the SPIRE database and included only contigs >200 kbp in length in the analysis. These contigs were analyzed using geNomad v 1.8.0 with default parameters, and only contigs that were classified as belonging to members of the class *Caudoviricetes* were selected.

Large phage contigs were screened by iteratively constructing trees based on VOG profiles using VirTree [27] and visualizing the output in iTOL [29]. In these trees, we included a reference set of phages that was previously constructed from the Inphared database and various metagenomic datasets [27, 30]. While examining trees, we sought to find clades of phages with large genomes that did not include cultivated strains. This resulted in a set of 25 SPIRE contigs that formed a distinct clade with high bootstrap support and formed a divergent group with no obviously cultivated representatives. These contigs ranged in length from 200,148 to 306,874 base pairs, and geNomad analysis showed that four had direct terminal repeats (DTRs) indicative of complete phage genomes. The closest relatives of these phages in our database were three "ERM phages" that were previously reported in a metagenomic survey of large bacteriophages [26], and we therefore included these in our subsequent analyses as well. Because these phages represent a novel clade that, to our knowledge, has not been described previously in the literature, we will refer to them here as "moraphages", from the Gaelic word 'móir', for large.

Phylogenetic trees of marker genes and average amino acid identity

We constructed phylogenies of the moraphages together with divergent relatives, using three viral marker genes: the major capsid protein (MCP: VOG00292), terminase large subunit (TerL: VOG00012), and family B DNA polymerase (PolB: VOG00275). For each of these trees, reference sequences in the NCBI database were identified using BLASTp, with representative moraphage proteins as the

query. Amino acid sequences were aligned using Clustal Omega with default parameters [31], and trees were constructed using IQ-TREE multicore v. 2.3.4 [32, 33]. The treefile file of the output was visualized in iTOL [23]. This same approach was used to make trees of the viral DnaK homologs identified in some morophage genomes (see below). Average amino acid identity was calculated as described previously [34, 35], and figures were made using the pheatmap function in R.

Protein prediction and functional annotation

Proteins were predicted using Prodigal v. 2.6.3 with default parameters [36]. Functional annotation was done using the eggNOG-mapper v2 (emapper-2.1.12) [37] tool based on eggNOG v. 5.0 [38]. In addition, we compared all of the proteins to the VOG database v. 214, using hmmsearch in HMMER3 (e-value cutoff, 1e-5 [39]).

Orthogroup analysis

We used OrthoFinder v. 2.5.4 [40] to identify orthologous groups (OGs) in the moraphages (-og parameter). We selected an 80% cutoff, meaning that the core genome would consist of the OGs that are shared with 80% or more of the phages. The three ERM phages were counted as one occurrence, and this percentage was therefore calculated out of 26 rather than 28 phages. The core genome was identified based on occurrence rather than the number of hits to an OG a phage may have. The predicted functions of the core genome OGs were predicted manually using BLASTp and HHpred. From the OrthoFinder results, which compiled amino acid sequences into their respective OGs, sequences were input into each database, and the best annotation was chosen based on percent identity and e-values. Pan-genomic statistics were compiled using the micropan utility in R.

Host prediction based on homology search

The predicted proteins encoded by each phage genome were searched against the RefSeq v. 227 database using LAST (parameters -m 500), and the best hits were recorded. The phylum of each best match was identified, and the bacterial phylum *Bacteroidota* received the most best hits for almost every phage. To assess the confidence that members of the *Bacteroidota* were the hosts of the phages, we calculated a metric by dividing the number of best hits to the *Bacteroidota* by the number of hits to the next-most-prevalent phylum and expressed this value as a ratio. If the ratio exceeded 1.5, the most prevalent host phylum was chosen as the likely host.

Synteny plot

A synteny plot of the complete phage genomes was made using GenoPlotR [41] in the R programming environment. The genomes and gene predictions were used as input, together with pairwise BLASTp outputs (BLASTp version 2.16.0+, e-value threshold of 1e-10 [42]).

Results and discussion

Phylogenetic placement of the moraphages

Here, we examined a novel clade of large phages found in metagenomes generated from a wide range of environments. The genome length of the moraphages ranged in size from 200 to 307 kbp, and four contained DTRs indicative of potentially complete genomes. Moraphages were identified primarily from freshwater sources such as lakes and rivers, but some variety in their environment was found (Table 1). For example, moraphages were also found

Table 1 Moraphages and their environmental origins

Phage ID	Environment
SAMN06267920_k141_159460.fasta	Freshwater
SAMN12796234_k141_1275679.fasta	Freshwater
SAMEA8660007_k141_1141498.fasta	Freshwater
SAMEA6649935_k141_278539.fasta	Freshwater
SAMEA6265476_k141_1882880.fasta	Freshwater
SAMN11356859_k141_1778059.fasta	Freshwater
SAMEA6265477_k141_410594.fasta	Freshwater
SAMEA8660000_k141_1707698.fasta	Freshwater
SAMN12794754_k141_102644.fasta	Freshwater
SAMEA6380161_k141_158953.fasta	Freshwater
SAMEA6380159_k141_2950544.fasta	Freshwater
SAMN10386151_k119_19505.fasta	Freshwater
SAMN08331091_k141_1934352.fasta	Freshwater
SAMN10386151_k119_131298.fasta	Freshwater
SAMN10386137_k119_135130.fasta	Freshwater
SAMN11532505_k141_984480.fasta	Freshwater
SAMN11532503_k141_1797313.fasta	Freshwater
SAMN11533406_k141_1557757.fasta	Freshwater
SAMN11530718_k141_1870712.fasta	Freshwater
SAMN11532462_k141_687100.fasta	Freshwater
SAMN10386142_k119_302602.fasta	Reclaimed water/ recycled water
SAMN19228552_k141_217681.fasta	Estuary
SAMN09092241_k141_1008493.fasta	Coastal salt marsh
SAMN08778078_k141_2211054.fasta	Marine water column
SAMN10350962_k141_248555.fasta	Subsurface sediment (gas well)
ERM_PHAGE_CIR_34_117.fasta	Marine sediment
ERM_PHAGE_CIR_34_14.fasta	Marine sediment
ERM_PHAGE_CIR_75_34.fasta	Marine sediment

in human-made facilities, as one phage was obtained from reclaimed water used to irrigate crops, and another was obtained from sediment near a natural gas well. Three distinct natural environments apart from freshwater are also represented: a marine water column, a coastal salt marsh, and an estuary. Out of the four moraphages with complete sequences, three were of freshwater origin, and one was isolated from an estuary. Throughout our genomic and phylogenetic analyses, we observed that the moraphages are related to three “ERM phages” that were recently discovered in marine sediment metagenomes [26]. We therefore also included these genome sequences in most of our comparative analyses. Other reference phages that were included in our phylogenetic analysis were found by comparing morophage marker genes to entries in the NCBI databases, which showed similarities to smaller phages derived from various sources, including aquatic systems as well as the human gut and mouth.

The morophage sequences were compared with those of large phages of other lineages using VirTree to determine whether they share similarities with members of these groups (Fig. 1). Also included in this analysis were members of the T4 group, specifically the families *Straboviridae*, *Kyanoviridae*, and *Ackermannviridae*, as well as the Lak phages, which have been found in human and animal gut metagenomes [23, 43]. The resulting phylogeny showed that moraphages are a distinct lineage that

does not fall within any of these established groups of large phages.

To identify the possible hosts of moraphages, we performed a homology search of all proteins in the phage genomes against the NCBI RefSeq database to identify hits to host bacterial lineages as described previously [26]. This analysis was performed on the grounds that phages often acquire genes from their host and that homology searches often provide clues regarding the identity of the host. From this analysis, we determined that the likely hosts of most moraphages are members of the class *Flavobacteriia*, with the plurality of best hits for 23 phages falling within the phylum *Bacteroidota*, while the plurality of hits for two were to members of the phylum *Pseudomonadota* (Fig. 1). The prediction of a host within the *Bacteroidota* is also consistent with phylogenetic analysis of the heat shock proteins identified in the morophage genomes (see details below).

To determine the phylogenetic placement of this novel clade of jumbo phages relative to other phages, phylogenetic trees were constructed based on the terminase large subunit (TerL: VOG00012), major capsid protein (MCP: VOG00292), and family B DNA polymerase (PolB: VOG00275) sequences. When constructing these trees, we included a set of the best BLASTp hits to these proteins (Fig. 2, panels A-C). Some of the morophage sequences did not show any similarity to MCP, TerL, or PolB VOGs

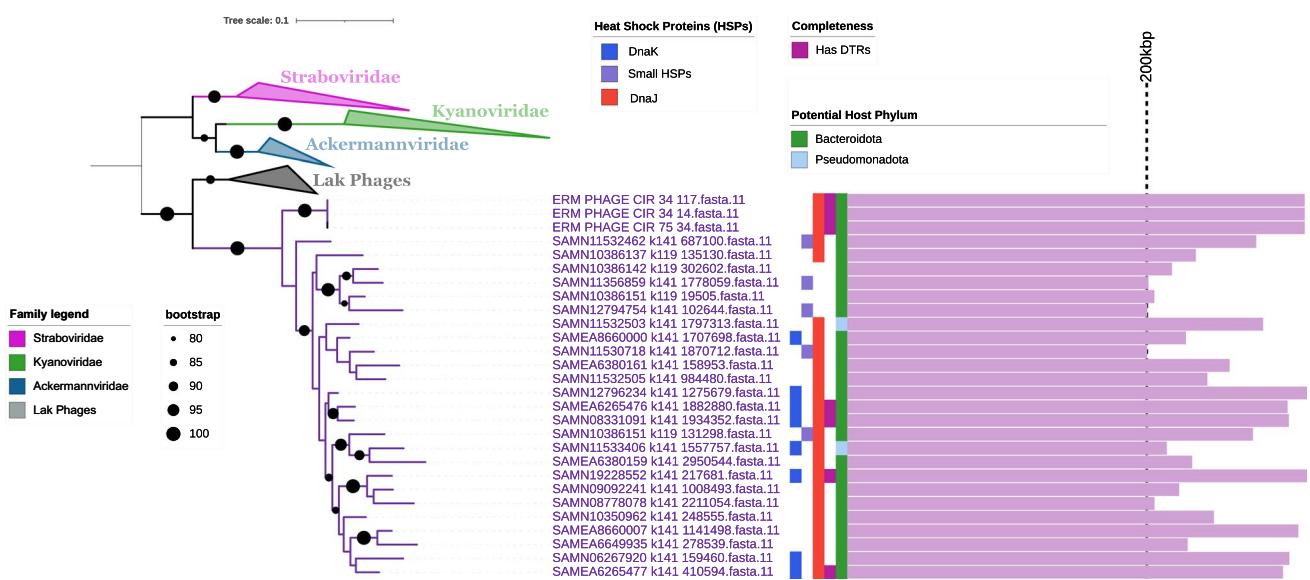


Fig. 1 Phylogenetic tree with host phylum prediction. The reference phages used in this tree are members of the T4 group and Lak phages, a group that includes phages with some of the largest genome sizes ever recorded. In this tree, the moraphages clustered with each other. Interestingly, the group 1 and Lak phages can be seen to share a more recent common ancestor than the one shared with the T4 phages. Their

hosts are predicted to belong to the same phylum – *Bacteroidota* – with the host of the Lak phages being narrowed down to members of the genus *Prevotella*. The presence of heat shock genes in all of the morophage genes is also indicated in the figure, as it was the basis for host prediction

and were therefore excluded from further analysis. In all of these trees, the moraphages formed a distinct clade that was divergent from any reference strains found in the NCBI database, confirming that they form a unique lineage of large phages. It should be noted that not all of the reference phages were jumbo phages, despite encoding proteins with sequence similarity to jumbo phage proteins. A common trend can be seen in all three trees, as the majority of the reference phages did not form a clade with the moraphages, resulting in two major nodes. A second trend was observed regarding the environments of the phages included in each tree. The reference phages, with a few exceptions, had been identified in metagenomic samples from the human gut or mouth. The trees suggest that the moraphages are a diverse lineage of phages that are distinct from previously characterized groups, and we therefore sought to quantify the genomic diversity within this lineage.

We also compared the moraphages to each other by examining their pairwise average amino acid identity (AAI), which is an index used to quantify whole-genome relatedness. We visualized these results using a heat map and a clustering dendrogram (Fig. 3). The values are mostly in the low percent identity range, with the majority of phages falling between 39 and 65%, suggesting that this lineage has broad diversity. There are some clear outliers from this average, namely three pairs of phages that have a high degree of amino acid sequence similarity. The pair with the second-highest identity value of 84.48% (SAMN12796234: SAMN08331091) and the pair with the highest identity value of 95.21% (SAME6649935: SAME8660007) originated from freshwater samples. Interestingly, the pair with the third-highest identity value, at 77.44% (SAMN09092241: SAMN08778078), are from different environments, from a coastal sea marsh and a marine water column, respectively. The phage SAMN09092241 contains a PolB gene but appears to lack an MCP or TerL VOG gene and, this phage was judged to have the most divergent PolB due to the length of its branch (Fig. 2 A-C). In sharp contrast, its paired phage SAMN08778078 had hits to all three hallmark proteins. The pair with the highest identity value were isolated from different freshwater locations but clustered together in all three trees. However, it should also be noted that the only phage genome out of these pairs that has DTRs was that of SAMN0833109, suggesting that, despite these phages being highly similar to each other, the AAI values of these three pairs may have been based on AAI measurements of incomplete genome comparisons.

Pan-genomics of the moraphages

The majority of the moraphages have low AAI percentages to each other, and we investigated if there were any genes that they shared as a group by assembling a pan-genome. The core genome consists of genes that are shared between all or the majority of the members of the clade, which offers a more focused view into the proteins that define the phage lineage. The pan-genome analysis also assesses the degree of gene diversity within the moraphage genomes by plotting a rarefaction curve (Fig. 4). While the moraphages may share a set of essential genes, the majority of their DNA encodes divergent and novel genes.

As all three ERM phages shared the same OGs, they were counted only once rather than three times for the core genome analysis. Many phages also had multiple hits to the same OG, but we identified the core genome based on occurrence. Our analysis showed that there are only 10 OGs that are shared between all moraphages, while 98 OGs are shared between 80% of them or more. To gain insight into the function of these core OGs, we manually investigated their predicted functions using BLASTp and HHpred (see Methods). There are some notable results from both BLASTp and HHpred analysis, namely that four separate OGs (0000011, 0000025, 0000026, and 0000028) are putative glycosyltransferases. The function of these proteins in phages has been reported to be glucosylation of phage DNA, which protects it from host restriction endonucleases as part of the phage replisome. Glycosyltransferases have also been observed in temperate phages, modifying host serotypes during lysogeny [44]. Genetic material is not the only target of glycosylation, the virion itself can be altered as well. Glycans have been observed at the C-termini of both the capsid and tail tube subunit of *Mycobacterium* phages, acting as a shield against anti-phage responses [45]. Due to their diverse roles, the acquisition of diverse glycosyltransferases has occurred many times independently in different phage lineages [46], and there is great diversity in their structure and function. Apart from this, structural proteins such as baseplate proteins, contractile tail proteins, and capsid assembly proteins also make up the core genome, which can be expected, as all of these components are necessary for the assembly of a virion. However, a gene encoding a tape measure protein (TMP), an essential tail assembly protein, was not identified within the genomes. The length of the TMP gene has been shown to be correlated with the length of the phage tail [47], but it might also have other important functions, such as peptidoglycan-degrading activity [48]. As phage

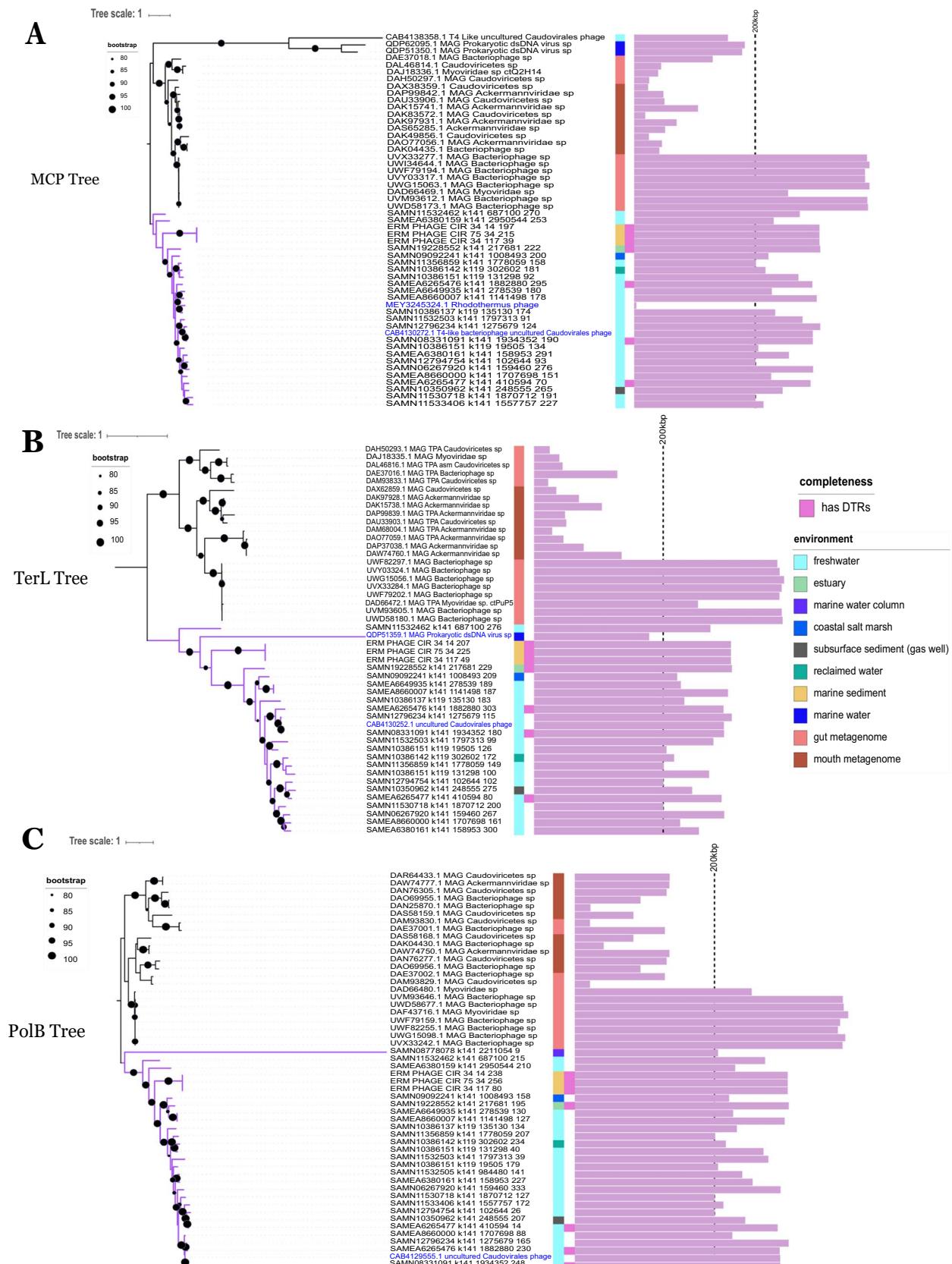


Fig. 2 Phylogenetic trees based on MCP, PolB, and TerL sequences. The group 1 phages formed a clade with each other (purple branches) based on shared VOGs in all trees. Most of the reference phages also formed a clade (black branches). The majority of the moraphages were collected from freshwater environments, but other environments such as marine water and sediments are represented. The three ERM phage sequences and four of the 25 moraphage sequences are considered complete due to the identification of DTRs. Bootstrap values range from 80 to 100 across the tree. (A) One *Caudovirales* phage and one phage infecting *Rhodothermus* (both labeled in blue) were not included in the group 1 clade, and both are of freshwater origin. The genome size of the *Rhodothermus* phage is small because the sequence is incomplete, but within this limited genome source, a match for the MCP was identified. (B) The TerL tree includes a *Caudovirales* jumbo phage and a MAG prokaryotic dsDNA virus (both labeled in blue) that clustered with group 1. The prokaryotic dsDNA virus is not a jumbo phage but does approach the 200 kb benchmark and is of marine origin. (C) A single reference jumbo *Caudovirales* phage (labeled in blue) was found to be closely related to the moraphages of freshwater origin. The longest branch on this tree belongs to one of the moraphages, suggesting that its PolB is more divergent from those of the other phages. This phage is also the only one that is of marine water column origin

tails are used for infection of the host and injection of the phage genome, the TMP is believed to contribute to this process [49]. Because of its importance for large tailed phages, we believe that the moraphages have a TMP gene within their genome that was not detected because of its divergence from its homologs in other phages.

Also of note, OG0000024 and OG0000031 are associated with RNA polymerase, serving as a sigma factor and a mediator of transcription, respectively, and OG0000043 was identified as a polymerase B subunit (see Data availability for raw data files). These results indicate that proteins associated with DNA replication, transcription, and virion structure are the most prevalent of the shared genes in the moraphage core genome. We also examined a set of 26 VOGs that were found in >80% of the moraphage genomes, which once again includes genes for a range of protein families predicted to be involved in DNA replication, transcription, and virion structure (Table 2).

We also constructed a synteny plot to visualize genome-level variability between the moraphages (Fig. 5). For this, we used the four moraphages that were predicted to be complete based on the presence of DTRs, as well as one representative ERM phage. This analysis confirmed the presence of conserved blocks of homologous proteins encoded by the phages, which in many cases correspond to genes involved in virion morphogenesis and genome replication. Moreover, it is apparent that several inversions and other genomic rearrangements have taken place in the evolution of this group,

suggesting that these may be important events that shape their genomic organization. Lastly, the synteny plots revealed several genomic regions with no or little homology to related phages, suggesting that they may represent genomic islands with variable composition across this lineage. If this is correct, an important avenue of future research would be to examine the functional repertoires of these islands and the dynamics of their gain and loss among the moraphages.

AMGs and other functional genes encoded by the moraphages

While the moraphages encode many unique proteins (discussed below), there are many that are missing that have been identified within other large phage lineages. For example, in a study by Al-Shayeb et al. [26], some clades of jumbo phage from many environments were found to encode tRNA synthetases that are distinct from those of their host, but we did not find any of these in the moraphage genomes. Another significant discovery was the “phage nucleus” that forms in the host during infection by phages of the family *Chimalliviridae* [16, 50]. Members of this family were found to encode a multi-subunit RNA polymerase for the transcription of viral genes within the “phage nucleus”, limiting the phage’s dependence on the host. We did not find any RNA polymerase gene homologs in the moraphages, suggesting that this clade of viruses is dependent on host transcriptional machinery for gene expression. This also suggests that the moraphages have a distinct infection program that differs from those of other lineages of large phages that have been characterized previously.

Of the 25 novel moraphage sequences, only four contained DTRs and are therefore presumed to represent complete genomes. We therefore performed detailed functional analysis using these four genome sequences. For this, we performed manual curation of protein annotations using BLASTp and HHpred, as well as cross-referencing with our automated annotations using EggNOG (see Methods). Using this approach, we found that these four moraphages were predicted to encode a histone H1-like protein (PF07432). Also referred to as a “linker histone”, this protein helps to stabilize the nucleosome by linking adjacent nucleosomes together [51], and homologs have been identified in the genomes of giant eukaryotic viruses such as medusavirus [52]. Medusavirus has been shown to encode a full set of histone homologs, but this H1-like protein is distinct in the timing

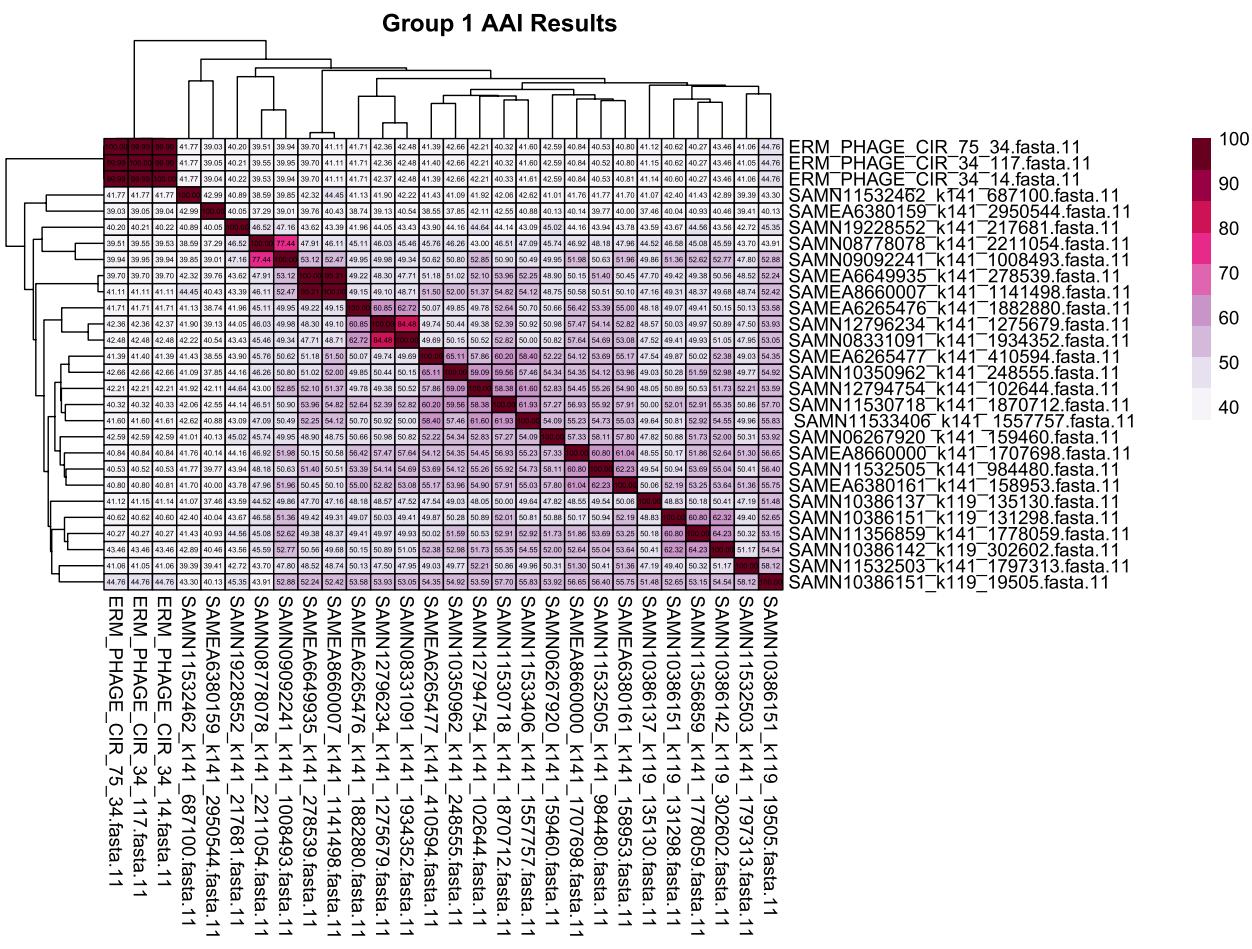


Fig. 3 Heat map displaying the amino acid identity between the members of group 1. The amino acid sequences of all group 1 phages were compared with each other using the AAI index. The majority of the phages have low identity values, ranging from 39% to 65%. The ERM phages are essentially identical in their sequences and have the highest

of its expression. In contrast to the other histones, H1 is expressed early in the infection cycle and may perform independent functions such as shutting down the host's genome or regulating the virus's genome for transcription [53]. A homolog of H1 has also been identified in *Chlamydia trachomatis* as Hc1, which binds to and densely compacts chromatin, regulating chlamydial gene expression [54, 55]. Hc1 has been expressed in *E. coli* to observe its effect on DNA structure and transcription/translation, and it was found that protein and DNA synthesis in the Hc1-expressing strain was considerably reduced [56, 57]. Although the function of these histone-like proteins in moraphages is still unknown, it is intriguing to consider that the phage may use it to compact its own genome during packaging in a manner analogous to that described for giant viruses [58, 59].

values assigned to them (100). Three pairs of phages were above the 65% threshold, with values of 77.44%, 84.48%, and 95.21%, respectively. The pair with the highest sequence similarity were from freshwater sources, and the pair with the third highest similarity were of coastal salt marsh and marine water column origin

Looking at the moraphages as a group, nine of them encode AMG glutamine synthetase (GS), whose function in bacteria is associated with nitrogen metabolism and ammonium reassimilation [60, 61]. This protein has been identified in oceanic viruses, suggesting that these viruses may manipulate nitrogen cycling in their environment [62]. GS components can also be used to activate alternative nitrogen-producing pathways under nitrogen-starved conditions [63]. Since the moraphages are of aquatic origin, nitrogen usage and potentially limited access to this element may have contributed to the phage's acquisition of a GS,

We also identified multiple antitoxin proteins in the four genomes with complete sequences. Antitoxins are typically components of toxin-antitoxin (TA) systems and are a common genetic module found in both bacterial

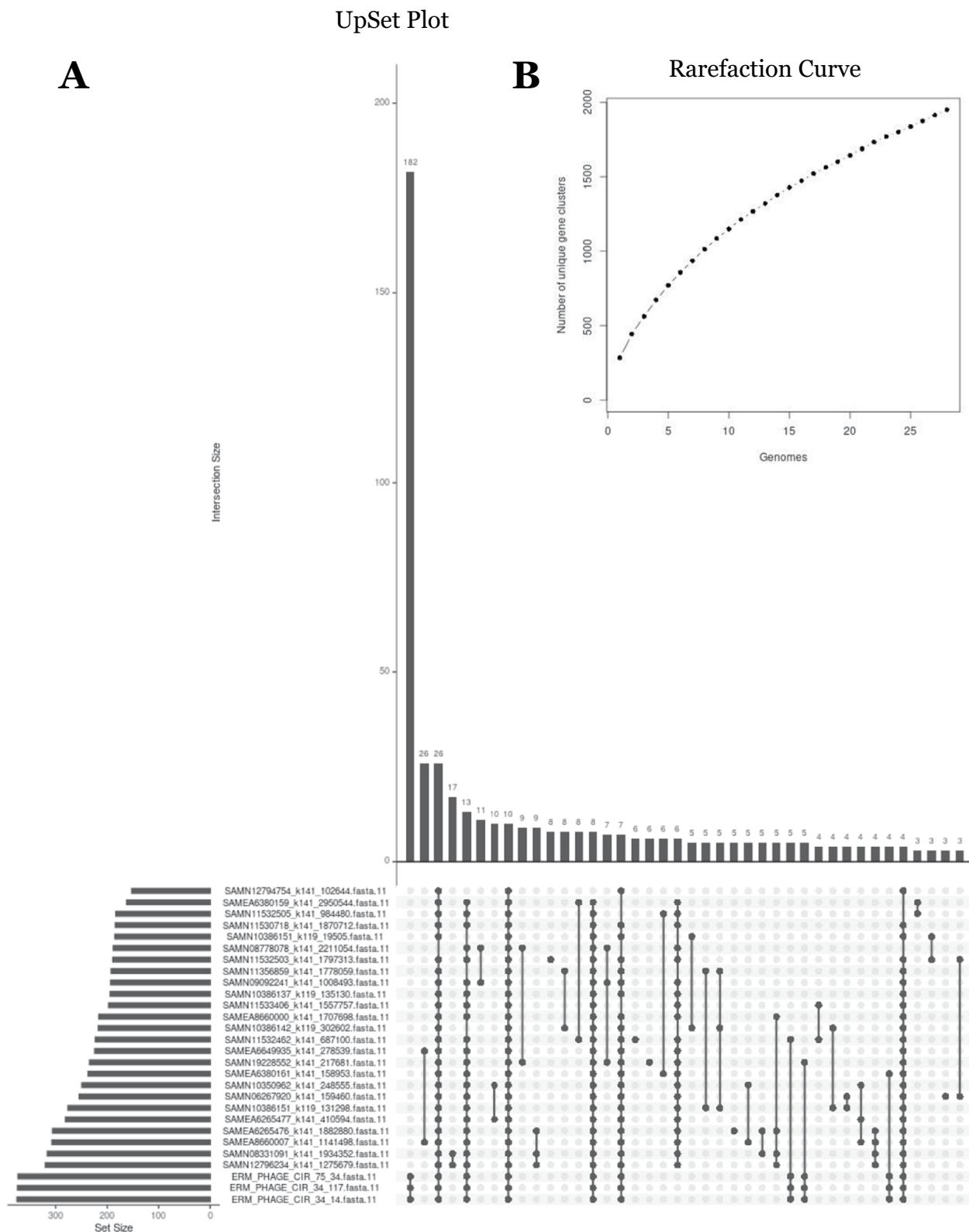


Fig. 4 UpSet plot and rarefaction curve generated from the pan-genomic analysis. Using the amino acid fasta files of each phage in group 1, a pangenomic analysis was conducted using OrthoFinder. The orthogroups generated were used to determine the core genome of the moraphages. (A) The ERM phages are an outlier when comparing them to each other, as they share many of the same genes, indicating that their genomes are essentially identical. The core genome consists of a relatively small subset of genes, and this can be attributed to the increasing number of unique gene clusters that appear as more genomes are added to the pan-genomic analysis, as seen in the rarefaction curve in panel B

cating that their genomes are essentially identical. The core genome consists of a relatively small subset of genes, and this can be attributed to the increasing number of unique gene clusters that appear as more genomes are added to the pan-genomic analysis, as seen in the rarefaction curve in panel B

Table 2 VOGs shared by moraphages

VOG	Occurrence	Annotation
VOG00031	24	DNA primase
VOG00042	24	Exonuclease subunit 2
VOG00043	23	Glutaredoxin
VOG00048	27	Deoxyuridine 5'-triphosphate nucleotidohydrolase
VOG00075	27	Tail tube protein gp19
VOG00084	27	Baseplate wedge protein gp25
VOG00090	24	RNA polymerase sigma-F factor
VOG00112	24	Terminase, large subunit
VOG00127	27	Tail sheath protein
VOG00275	27	DNA-directed DNA polymerase
VOG00292	26	Major capsid protein
VOG00369	27	Neck protein gp13
VOG00439	24	Lysine-tRNA ligase
VOG00674	25	Prohead core protein protease
VOG00681	25	Portal protein
VOG00710	25	DNA-directed DNA polymerase
VOG01084	27	Baseplate wedge protein gp6
VOG01986	28	DNA polymerase I
VOG02072	24	ATP-dependent Clp protease
VOG02083	24	ATP-binding subunit ClpA
VOG02246	23	Hypothetical protein
VOG02439	27	Single-stranded DNA-binding protein
VOG05220	24	PhoH-like protein
VOG06156	27	Hypothetical protein
VOG17084	25	Hypothetical protein
VOG18190	24	ATP-dependent DNA helicase uvsW
		Putative peptidase

and archaeal chromosomes and plasmids, with some species encoding multiple distinct systems [64]. TA systems have also been commonly identified in prophage genomes [65, 66]. These modules typically comprise a growth-inhibiting toxin and its antitoxin, allowing for reversible changes to the bacterial cell, including some in response to phage infection [67]. Recently, there has been renewed interest in the anti-phage activity of TA systems. Bacteria use the modules as a broad method of defense against phage infection, with various biochemical activities that can target different phage processes [68]. Some phages are known to use antitoxin mimics to abrogate restriction by anti-phage TA systems. This has led to speculation that freestanding antitoxins in phage genomes may act as a counter to host anti-phage immune systems. The abundance of antitoxin genes in moraphage genomes may reflect an investment in countering host cell defenses.

Intriguingly, we observed an association between CcdA/HicB-like antitoxins and a group of Pac-I-like putative homing endonucleases present in multiple copies in some of the moraphage genomes. Homing endonucleases are selfish genetic elements that can spread within phage genomes and facilitate allelic conversion during phage coinfection [69]. In addition to neutralizing toxin activity, CcdA and HicB are known to transcriptionally repress toxin expression [70]. The T4-encoded homing endonuclease Itev-I has been shown to repress its own promoter [71]. Pairing an antitoxin gene with a homing endonuclease in a single module may be an alternative strategy for regulating homing endonuclease activity. Alternatively, if the antitoxin provides a fitness benefit to the phage, pairing this gene with a homing endonuclease may be a way to offset any fitness cost imposed by the selfish endonuclease.

One of the most striking features of the moraphages is the presence of several heat shock proteins (HSPs), molecular chaperones that are widespread in plants, animals, and prokaryotes, regulating the folding of proteins and protection from improper folding [72]. Similar to antitoxin mimics, phage-encoded HSPs may be a means of overcoming anti-phage defenses. HSPs are known to promote antiviral responses in eukaryotes [73], and recent work has shown that DnaJ is required for the anti-phage activity of an NLR-like protein in *E. coli* [74]. Additionally, another bacterial chaperone, SecB, is required for the function of some antitoxins [75], and some HSPs are able to phenocopy SecB's canonical function for protein secretion [76]. As an alternative to a counterdefense function, phage-encoded DnaJ chaperones might recruit the host's DnaK and facilitate the assembly of viral proteins, as seen in eukaryotic viruses such as simian virus 40 [77] and crocodilepox virus [78]. In a 2012 study, a T4-related enterobacteriophage named RB43 was found to encode a putative functional J-domain protein with 63% amino acid sequence identity to known *Escherichia coli* DnaJs [79]. DnaJs also seem to be more common in the moraphage genomes, as can be seen in Figure 1. While these proteins are highly conserved across the domains of life, their presence in phage genomes and their overall role in phages is not well understood.

In addition to DnaJ and DnaK, some moraphages encode the small heat shock protein HSP20. A subclass of HSPs, small heat shock proteins (sHSPs), with a few exceptions, are also ubiquitous in all forms of life but differ from HSPs. These proteins have a low molecular weight, ranging from 16 kDa to 42 kDa, and exhibit

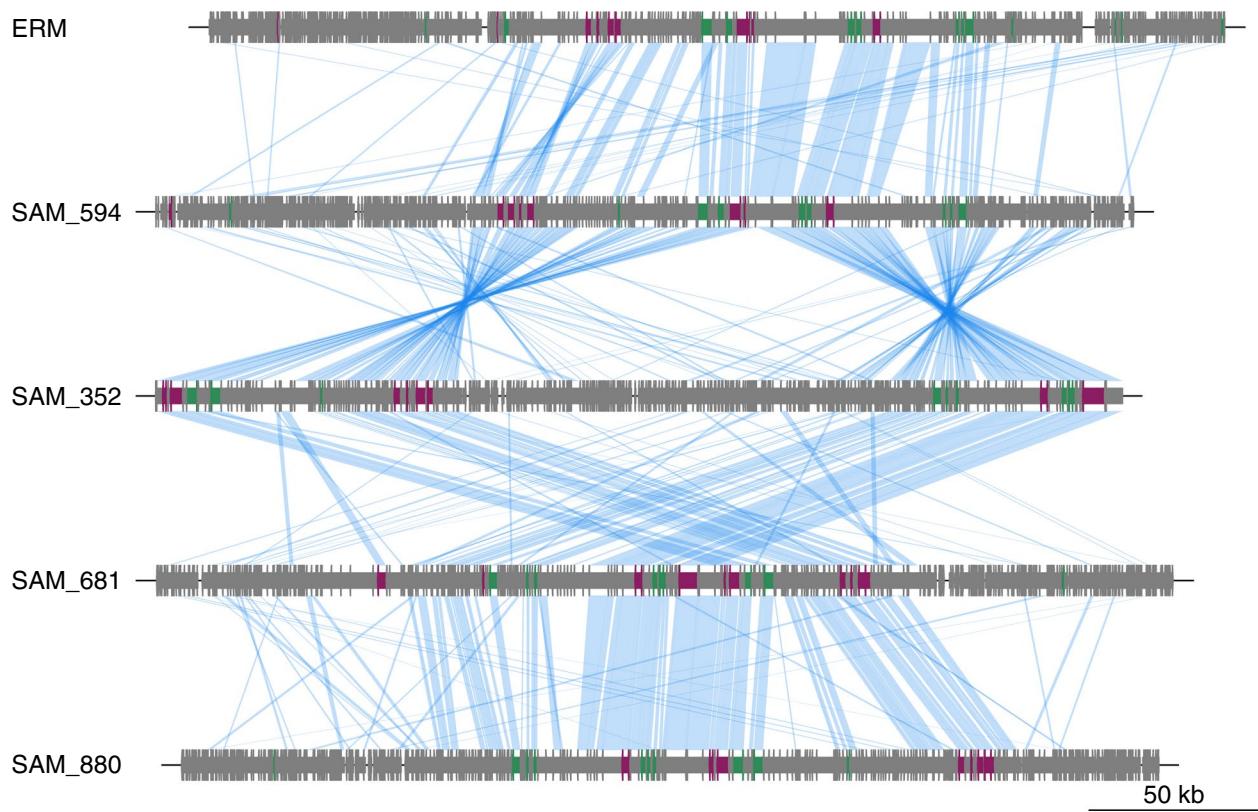


Fig. 5 Synteny plot of the complete genomes of four moraphages and one representative ERM phage. Blue lines denote the best BLASTp matches of the encoded proteins. Genes colored in green are those with matches to VOG families involved in DNA replication (polymerases, helicases, primases, etc.), while those colored in maroon are those with hits to VOG families involved in virion morphogenesis (major

capsid proteins, terminase subunits, tail fibers, etc.). The names are abbreviated for clarity: ERM, ERM_PHAGE_CIR_34_14; SAM_594, SAMEA6265477_k141_410594; SAM_352, SAMN08331091_k141_1934352; SAM_681, SAMN19228552_k141_217681; SAM_880, SAMEA6265476_k141_1882880

lower nucleotide and amino acid sequence conservation across species groups [80, 81]. Like other HSPs, these proteins have been found to participate in host protein folding and mediated disaggregation of proteins [82] and to respond to environmental changes [83]. The number of sHSPs in different organisms is also quite variable, with humans having 10 [84], *Drosophila* having four [85, 86], and some pathogenic bacteria such as *Helicobacter pylori* not possessing any [87].

Using HSP70 (DnaK) sequences from the eight moraphages encoding the protein and related sequences from bacteria identified by BLASTp searches of the NCBI RefSeq database, a phylogenetic tree was constructed. One particular moraphage (SAMN19228552) displayed a relatively long branch length and did not cluster with the other seven DnaK-encoding moraphages, suggesting that it might have acquired a DnaK gene independently

from the others. The other seven moraphage DnaK homologs formed a distinct clade within the family *Flavobacteriaceae* of the phylum *Bacteroidota*, consistent with the view that these phages likely infect these hosts and acquired HSP70 from a member of this bacterial lineage (Fig. 6).

Conclusions

Here, we present 25 novel phage genomes from a novel lineage of large bacteriophages that we refer to as "moraphages". These phages were identified primarily in freshwater systems, indicating that aquatic environments are home to an abundance of novel large viruses. These novel phages have genes encoding many unique proteins that are seldom seen in other jumbo phages, most notably

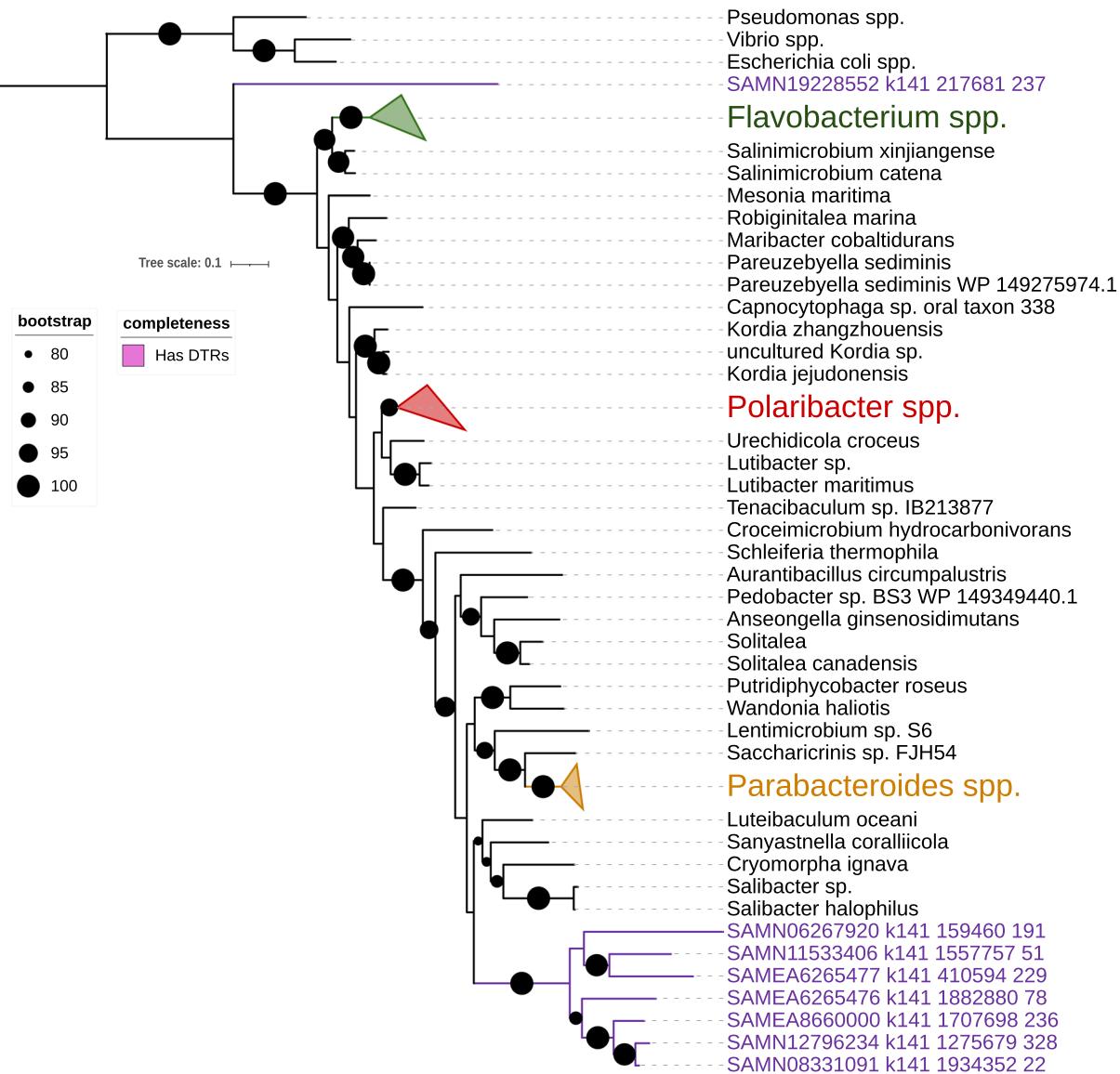


Fig. 6 DnaK tree with members of the phylum *Bacteroidota* as references. All morphages that encode a DnaK protein were included, together with representative best BLASTp hits of these proteins in the NCBI RefSeq database (searches performed in May 2025). All but one

of the phage DnaKs formed a clade within the *Bacteroidota* lineage. Phage SAMN19228552 was the only exception; the DnaK homolog from this phage was highly divergent and might have been acquired independently

chaperones such as DnaJ and DnaK. They were also found to lack several genes for proteins that have been discovered recently in other large phages, such as those encoding tRNA synthetases and the machinery needed to form a phage nucleus, suggesting that these phages have an infection strategy that differs from those of other described jumbo phages. These results contribute to our

growing understanding of large phages in the biosphere and underscore the importance of large viruses to global viral diversity.

Acknowledgements This study was performed using computational resources at the Virginia Tech Advanced Research Computing Center. CPH was supported by an NIH-NIGMS training grant: Virginia Tech Postbaccalaureate Research Education Program grant, GM066534-20. This work was supported by NSF award #2141862 to FOA. We are

grateful to Anh D. Ha for assistance in processing the metagenomic datasets.

Funding National Science Foundation, #2141862, Frank O. Aylward

Data availability Supplementary data files, such as genome annotations, complete morophage amino acid and nucleotide sequences, and high-resolution figures, can be found in the following Zenodo repository: Morophage_Supplementary_Datasets_and_Figures (<https://doi.org/10.5281/zenodo.15586673>).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Mushegian AR (2020) Are there 1031 virus particles on Earth, or more, or fewer? *J Bacteriol* 202:10. <https://doi.org/10.1128/JB.00052-20>
2. Bar-On YM, Phillips R, Milo R (2018) The biomass distribution on Earth. *Proc Natl Acad Sci U S A* 115:6506–6511
3. Abedon ST (2009) Phage evolution and ecology. *Adv Appl Microbiol* 67:1–45
4. Suttle CA (2005) Viruses in the sea. *Nature* 437:356–361
5. Elahi Y, Nowroozi J, Fard RMN (2021) Isolation and characterization of bacteriophages from wastewater sources on *Enterococcus* spp. isolated from clinical samples. *Iran J Microbiol* 13:671–677
6. Bolduc B, Wirth JF, Mazurie A, Young MJ (2015) Viral assemblage composition in Yellowstone acidic hot springs assessed by network analysis. *ISME J* 9:2162–2177
7. CobíánGüemes AG, Youle M, Cantú VA et al (2016) Viruses as winners in the game of life. *Annu Rev Virol* 3:197–214
8. Wilhelm SW, Suttle CA (1999) Viruses and nutrient cycles in the sea. *Bioscience* 49:781–788
9. Braga LPP, Spor A, Kot W et al (2020) Impact of phages on soil bacterial communities and nitrogen availability under different assembly scenarios. *Microbiome* 8:52
10. Hendrix RW (2009) Jumbo bacteriophages. *Curr Top Microbiol Immunol* 328:229–240
11. Serwer P, Hayes SJ, Thomas JA, Hardies SC (2007) Propagating the missing bacteriophages: a large bacteriophage in a new class. *Virol J* 4:21
12. Kim JH, Son JS, Choi YJ et al (2012) Complete genome sequence and characterization of a broad-host range T4-like bacteriophage phiAS5 infecting *Aeromonas salmonicida* subsp. *salmonicida*. *Vet Microbiol* 157:164–171
13. Lavyshev D, Sokolova M, Minakhin L et al (2016) The genome of AR9, a giant transducing *Bacillus* phage encoding two multisubunit RNA polymerases. *Virology* 495:185–196
14. Antonova D, Belousova VV, Zhivkopoulos E et al (2023) The dynamics of synthesis and localization of jumbo phage RNA polymerases inside infected cells. *Viruses* 15:2096. <https://doi.org/10.3390/v15102096>
15. Nazir A, Ali A, Qing H, Tong Y (2021) Emerging aspects of jumbo bacteriophages. *Infect Drug Resist* 14:5041–5055
16. Laughlin TG, Deep A, Prichard AM et al (2022) Architecture and self-assembly of the jumbo bacteriophage nuclear shell. *Nature* 608:429–435
17. Matsuzaki S, Tanaka S, Koga T, Kawata T (1992) A broad-host-range vibriophage, KVP40, isolated from sea water. *Microbiol Immunol* 36:93–97
18. Ceyssens P-J, Minakhin L, Van den Bossche A et al (2014) Development of giant bacteriophage φKZ is independent of the host transcription apparatus. *J Virol* 88:10501–10510
19. Yuan Y, Gao M (2017) Jumbo bacteriophages: An overview. *Front Microbiol* 8:403
20. Weinheimer AR, Aylward FO (2022) Infection strategy and biogeography distinguish cosmopolitan groups of marine jumbo bacteriophages. *ISME J* 16:1657–1667
21. Chen L, Banfield JF (2024) COBRA improves the completeness and contiguity of viral genomes assembled from metagenomes. *Nat Microbiol* 9:737–750
22. Michniewski S, Rihtman B, Cook R et al (2021) A new family of “megaphages” abundant in the marine environment. *ISME Commun* 1:58
23. Devoto AE, Santini JM, Olm MR et al (2019) Megaphages infect *Prevotella* and variants are widespread in gut microbiomes. *Nat Microbiol* 4:693–700
24. Chen L-X, Méheust R, Crits-Christoph A et al (2020) Large freshwater phages with the potential to augment aerobic methane oxidation. *Nat Microbiol* 5:1504–1515
25. Li C, Liu K, Gu C et al (2024) Gastrointestinal jumbo phages possess independent synthesis and utilization systems of NAD. *Microbiome* 12:268
26. Al-Shayeb B, Sachdeva R, Chen L-X et al (2020) Clades of huge phages from across Earth’s ecosystems. *Nature* 578:425–431
27. Weinheimer AR, Ha AD, Aylward FO (2025) Towards a unifying phylogenomic framework for tailed phages. *PLoS Genet* 21:e1011595
28. Schmidt TSB, Fullam A, Ferretti P et al (2024) SPIRE: a Searchable, Planetary-scale mIcrobiome REsource. *Nucleic Acids Res* 52:D777–D783
29. Letunic I, Bork P (2024) Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res* 52:W78–W82
30. Cook R, Brown N, Redgwell T et al (2021) INfrastructure for a PHAge REference Database: Identification of Large-Scale Biases in the Current Collection of Cultured Phage Genomes. *Phage (New Rochelle)* 2:214–223
31. Madeira F, Madhusoodanan N, Lee J et al (2024) The EMBL-EBI Job Dispatcher sequence analysis tools framework in 2024. *Nucleic Acids Res* 52:W521–W525
32. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32:268–274
33. Minh BQ, Schmidt HA, Chernomor O et al (2020) IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* 37:1530–1534
34. Karki S, Moniruzzaman M, Aylward FO (2021) Comparative genomics and environmental distribution of large dsDNA viruses in the family Asfarviridae. *Front Microbiol* 12:657471
35. Moniruzzaman M, Martinez-Gutierrez CA, Weinheimer AR, Aylward FO (2020) Dynamic genome evolution and complex virocell metabolism of globally-distributed giant viruses. *Nat Commun* 11:1710

36. Hyatt D, Chen G-L, Locascio PF et al (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119

37. Cantalapiedra CP, Hernández-Plaza A, Letunic I et al (2021) Egg-NOG-mapper v2: Functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol Biol Evol* 38:5825–5829

38. Huerta-Cepas J, Szklarczyk D, Heller D et al (2019) eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res* 47:D309–D314

39. Eddy SR (2011) Accelerated Profile HMM Searches. *PLoS Comput Biol* 7:e1002195

40. Emms DM, Kelly S (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* 20:238

41. Guy L, Kultima JR, Andersson SGE (2010) genoPlotR: comparative gene and genome visualization in R. *Bioinformatics* 26:2334–2335

42. Camacho C, Coulouris G, Avagyan V et al (2009) BLAST+: architecture and applications. *BMC Bioinformatics* 10:421

43. Cook R, Crisci MA, Pye HV et al (2024) Decoding huge phage diversity: a taxonomic classification of Lak megaphages. *J Gen Virol* 105:001997. <https://doi.org/10.1099/jgv.0.001997>

44. Markine-Goriaynoff N, Gillet L, Van Etten JL et al (2004) Glycosyltransferases encoded by viruses. *J Gen Virol* 85:2741–2754

45. Freeman KG, Robotham AC, Parks OB et al (2023) Virion glycosylation influences mycobacteriophage immune recognition. *Cell Host Microbe* 31:1216–1231.e6

46. Pyle JD, Lund SR, O'Toole KH, Saleh L (2024) Virus-encoded glycosyltransferases hypermodify DNA with diverse glycans. *Cell Rep* 43:114631

47. Katsura I, Hendrix RW (1984) Length determination in bacteriophage lambda tails. *Cell* 39:691–698

48. Piuri M, Hatfull GF (2006) A peptidoglycan hydrolase motif within the mycobacteriophage TM4 tape measure protein promotes efficient infection of stationary phase cells. *Mol Microbiol* 62:1569–1585

49. Cumby N, Reimer K, Mengin-Lecreux D et al (2015) The phage tail tape measure protein, an inner membrane protein and a periplasmic chaperone play connected roles in the genome injection process of *E. coli* phage HK97: Phage and host protein requirements for HK97 genome injection. *Mol Microbiol* 96:437–447

50. Chaikeratisak V, Nguyen K, Khanna K et al (2017) Assembly of a nucleus-like structure during viral replication in bacteria. *Science* 355:194–197

51. Talbert PB, Armache K-J, Henikoff S (2022) Viral histones: pick-pocket's prize or primordial progenitor? *Epigenetics Chromatin* 15:21

52. Yoshikawa G, Blanc-Mathieu R, Song C et al (2019) Medusavirus, a novel large DNA virus discovered from hot spring water. *J Virol* 93:10. <https://doi.org/10.1128/JVI.02130-18>

53. Zhang R, Endo H, Takemura M, Ogata H (2021) RNA sequencing of medusavirus suggests remodeling of the host nuclear environment at an early infection stage. *Microbiol Spectr* 9:e0006421

54. Hackstadt T, Baehr W, Ying Y (1991) *Chlamydia trachomatis* developmentally regulated protein is homologous to eukaryotic histone H1. *Proc Natl Acad Sci U S A* 88:3937–3941

55. Perara E, Ganem D, Engel JN (1992) A developmentally regulated chlamydial gene with apparent homology to eukaryotic histone H1. *Proc Natl Acad Sci U S A* 89:2125–2129

56. Barry CE III, Brickman TJ, Hackstadt T (1993) Hc1-mediated effects on DNA structure: a potential regulator of chlamydial development. *Mol Microbiol* 9:273–283

57. Pedersen LB, Birkeland S, Christiansen G (1994) Interaction of the *Chlamydia trachomatis* histone H1-like protein (Hc1) with DNA and RNA causes repression of transcription and translation in vitro. *Mol Microbiol* 11:1085–1098

58. Liu Y, Bisio H, Toner CM et al (2021) Virus-encoded histone doublets are essential and form nucleosome-like structures. *Cell* 184:4237–4250.e19

59. Valencia-Sánchez MI, Abini-Agbomson S, Wang M et al (2021) The structure of a virus-encoded nucleosome. *Nat Struct Mol Biol* 28:413–417

60. Bolay P, Muro-Pastor MI, Florencio FJ, Klähn S (2018) The distinctive regulation of cyanobacterial glutamine synthetase. *Life (Basel)* 8:52

61. Prasse D, Förstner KU, Jäger D et al (2017) sRNA154 a newly identified regulator of nitrogen fixation in *Methanoscincus mazei* strain Gö1. *RNA Biol* 14:1544–1558

62. Roux S, Brum JR, Dutilh BE et al (2016) Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature* 537:689–693

63. Huergo LF, Chandra G, Merrick M (2013) P(II) signal transduction proteins: nitrogen regulation and beyond. *FEMS Microbiol Rev* 37:251–283

64. Harms A, Brodersen DE, Mitarai N, Gerdes K (2018) Toxins, targets, and triggers: An overview of toxin-antitoxin biology. *Mol Cell* 70:768–784

65. Makarova KS, Wolf YI, Koonin EV (2009) Comprehensive comparative-genomic analysis of type 2 toxin-antitoxin systems and related mobile stress response systems in prokaryotes. *Biol Direct* 4:19

66. Encina-Robles J, Pérez-Villalobos V, Bustamante P (2024) The HicAB system: Characteristics and biological roles of an underappreciated toxin-antitoxin system. *Int J Mol Sci* 25:12165

67. Yamaguchi Y, Park J-H, Inouye M (2011) Toxin-antitoxin systems in bacteria and archaea. *Annu Rev Genet* 45:61–79

68. LeRoux M, Laub MT (2022) Toxin-antitoxin systems as phage defense elements. *Annu Rev Microbiol* 76:21–43

69. Barth ZK, Dunham DT, Seed KD (2023) Nuclease genes occupy boundaries of genetic exchange between bacteriophages. *NAR Genom Bioinform* 5:Iqad076

70. Chan WT, Espinosa M, Yeo CC (2016) Keeping the wolves at bay: Antitoxins of prokaryotic type II toxin-antitoxin systems. *Front Mol Biosci* 3:9

71. Edgell DR, Derbyshire V, Van Roey P et al (2004) Intron-encoded homing endonuclease I-TevI also functions as a transcriptional autorepressor. *Nat Struct Mol Biol* 11:936–944

72. Singh MK, Shin Y, Ju S et al (2024) Heat shock response and heat shock proteins: Current understanding and future opportunities in human diseases. *Int J Mol Sci* 25:4209. <https://doi.org/10.3390/ijms25084209>

73. Wan Q, Song D, Li H, He M-L (2020) Stress proteins: the biological functions in virus infection, present and challenges for target-based antiviral drug development. *Signal Transduct Target Ther* 5:125

74. Conte AN, Ruchel ME, Ridgeway SM et al (2025) DnaJ mediates phage sensing by the bacterial NLR-related protein bNACHT25. *PLoS Biol* 23:e3003203

75. Guillet V, Bordes P, Bon C et al (2019) Structural insights into chaperone addiction of toxin-antitoxin systems. *Nat Commun* 10:782

76. Altman E, Kumamoto CA, Emr SD (1991) Heat-shock proteins can substitute for SecB function during protein export in *Escherichia coli*. *EMBO J* 10:239–245

77. Wright CM, Seguin SP, Fewell SW et al (2009) Inhibition of Simian Virus 40 replication by targeting the molecular chaperone function and ATPase activity of T antigen. *Virus Res* 141:71–80

78. Afonso CL, Tulman ER, Delhon G et al (2006) Genome of crocodilepox virus. *J Virol* 80:4978–4991

79. Perrody E, Cirinesi A-M, Desplats C et al (2012) A bacteriophage-encoded J-domain protein interacts with the DnaK/Hsp70 chaperone and stabilizes the heat-shock factor σ32 of *Escherichia coli*. *PLoS Genet* 8:e1003037
80. Stromer T, Ehrnsperger M, Gaestel M, Buchner J (2003) Analysis of the interaction of small heat shock proteins with unfolding proteins. *J Biol Chem* 278:18015–18021
81. Scheibel T, Weikl T, Buchner J (1998) Two chaperone sites in Hsp90 differing in substrate specificity and ATP dependence. *Proc Natl Acad Sci U S A* 95:1495–1499
82. Ratajczak E, Zietkiewicz S, Liberek K (2009) Distinct activities of *Escherichia coli* small heat shock proteins IbpA and IbpB promote efficient protein disaggregation. *J Mol Biol* 386:178–189
83. Guzzo J, Jobin M-P, Delmas F et al (2000) Regulation of stress response in *Oenococcus oeni* as a function of environmental changes and growth phase. *Int J Food Microbiol* 55:27–31
84. Kappé G, Franck E, Verschueren P, et al (2003) The human genome encodes 10 alpha-crystallin-related small heat shock proteins: HspB1-10. *Cell stress & chaperones*. 8:53. [https://doi.org/10.1379/1466-1268\(2003\)8%3C;53:THGEC%3E;2.0.CO;2](https://doi.org/10.1379/1466-1268(2003)8%3C;53:THGEC%3E;2.0.CO;2)
85. Corces V, Holmgren R, Freund R et al (1980) Four heat shock proteins of *Drosophila melanogaster* coded within a 12-kilobase region in chromosome subdivision 67B. *Proc Natl Acad Sci U S A* 77:5390–5393
86. Ingolia TD, Craig EA (1982) Four small *Drosophila* heat shock proteins are related to each other and to mammalian alpha-crystallin. *Proc Natl Acad Sci U S A* 79:2360–2364
87. Narberhaus F (2002) Alpha-crystallin-type heat shock proteins: socializing minichaperones in the context of a multichaperone network. *Microbiol Mol Biol Rev*. 66:64–93 (table of contents)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.