



Article

Origin, Evolution and Diversity of ϕ 29-like Phages—Review and Bioinformatic Analysis

Peter Evseev ^{1,2,*} , Daria Gutnik ³ , Alena Evpak ¹, Anastasia Kasimova ⁴ and Konstantin Miroshnikov ^{1,*}

¹ Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Miklukho-Maklaya Street 16/10, 117997 Moscow, Russia

² Laboratory of Molecular Microbiology, Pirogov Russian National Research Medical University, Ostrovityanova Street 1, 117997 Moscow, Russia

³ Limnological Institute, Siberian Branch of the Russian Academy of Sciences, Ulan-Batorsakaya Street, 3, 664033 Irkutsk, Russia

⁴ Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, Leninsky Prospekt, 47, 119991 Moscow, Russia

* Correspondence: evseev@ibch.ru (P.E.); kmi@ibch.ru (K.M.)

Abstract: Phage ϕ 29 and related bacteriophages are currently the smallest known tailed viruses infecting various representatives of both Gram-positive and Gram-negative bacteria. They are characterised by genomic content features and distinctive properties that are unique among known tailed phages; their characteristics include protein primer-driven replication and a packaging process characteristic of this group. Searches conducted using public genomic databases revealed in excess of 2000 entries, including bacteriophages, phage plasmids and sequences identified as being archaeal that share the characteristic features of phage ϕ 29. An analysis of predicted proteins, however, indicated that the metagenomic sequences attributed as archaeal appear to be misclassified and belong to bacteriophages. An analysis of the translated polypeptides of major capsid proteins (MCPs) of ϕ 29-related phages indicated the dissimilarity of MCP sequences to those of almost all other known *Caudoviricetes* groups and a possible distant relationship to MCPs of T7-like (*Autographiviridae*) phages. Sequence searches conducted using HMM revealed the relatedness between the main structural proteins of ϕ 29-like phages and an unusual lactococcal phage, KSY1 (*Chopinivirus* KSY1), whose genome contains two genes of RNA polymerase that are similar to the RNA polymerases of phages of the *Autographiviridae* and *Schitoviridae* (N4-like) families. An analysis of the tail tube proteins of ϕ 29-like phages indicated their dissimilarity of the lower collar protein to tail proteins of all other viral groups, but revealed its possible distant relatedness with proteins of toxin translocation complexes. The combination of the unique features and distinctive origin of ϕ 29-related phages suggests the categorisation of this vast group in a new order or as a new taxon of a higher rank.

Keywords: phage ϕ 29; phage phi29; ϕ 29-like phages; ϕ 29 evolution; phage evolution; phage origin; phage diversity; bacteriophage biodiversity



Citation: Evseev, P.; Gutnik, D.; Evpak, A.; Kasimova, A.; Miroshnikov, K. Origin, Evolution and Diversity of ϕ 29-like Phages—Review and Bioinformatic Analysis. *Int. J. Mol. Sci.* **2024**, *25*, 10838. <https://doi.org/10.3390/ijms251910838>

Academic Editor: Alicja Węgrzyn

Received: 11 September 2024

Revised: 4 October 2024

Accepted: 7 October 2024

Published: 9 October 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Bacteriophage ϕ 29, infecting *Bacillus*, is one of the classical entities of phage biology. First described in the 1960s [1–4], ϕ 29 has now been the focus of more than 500 research articles, which have featured many aspects of its basic virology, genetics, protein and structural biology. Many features of its molecular organisation and infection cycle have been considered unique to ϕ 29, in contrast to other well-studied phages such as T4, T7 or λ . Subsequently, an accumulation of scientific data on phages sharing some keystone properties has resulted in the establishment of a group of ϕ 29-like phages that infect various Gram-positive and Gram-negative bacteria. Current global changes in the taxonomy of bacteriophages and viruses in general [5,6] require the listing of original genomic and

metabolic features that are common to all members of taxa at the level of the subfamily and above.

At present (August 2024), phage ϕ 29 and closely related phages are assigned to the class *Caudoviricetes*, family *Salasmaviridae*, subfamily *Picovirinae*, genus *Salasvirus* (<https://ictv.global/taxonomy>, accessed on 1 September 2024). Several other families approved by the International Committee on Taxonomy of Viruses (ICTV) families contain phages more distantly related to ϕ 29, infecting Bacillales and Clostridia, which are characterised by varying degrees of similarity of morphogenetic proteins to those of ϕ 29 and protein primer-driven replication. These are *Guelinviridae* (*Clostridium* phage CPV4 [7] and others), *Madridviridae* (e.g., *Streptococcus* phage CP-1 [8]) and *Rountreeviridae* (*Staphylococcus* virus Andhra [9] and others) families. Several actinophages, as were shown previously [10], also appear to be related to ϕ 29, which include the unclassified *Curtobacterium* phages Ayka and those related [10,11], and phages assigned by the ICTV to genera but not included in subfamilies or families; those are genera, namely, *Anjalivirus*, *Badaztecivirus* and *Dybvigvirus* [12]. In addition, several phages infecting Gram-negative bacteria appear also to possess morphogenesis proteins and DNA polymerase similar to ϕ 29 ones, they include *Salmonella* phages assigned to genus *Astrithrovirus*, *Pectobacterium* phage DU-PP-III, and *Acinetobacter* phage Aci01-2-Phanie (short name Phanie) [13].

The diversity of environmental niches and bacterial hosts of ϕ 29-related phages promotes the search of similar phages in metagenomic data, and their unusual features raise questions about the multiformity of this viral group and its origin. The evolutionary analysis of bacteriophages is a cumbersome task. The evolution of viruses is characterised by network character [14] and modular nature [15,16], and viral proteins tend to diverge fast. The last factor can be sometimes overcome by analysing protein folding and structural similarity, which are more conservative than amino acid sequence. This approach is, however, limited by a number of experimentally resolved structures which are much fewer than the available sequences.

ϕ 29-related phages are podoviruses, i.e., phages with short non-contractile tails. The major capsid proteins (MCPs) of these phages, as well as other *Duplodnaviria* viruses and shell proteins of encapsulins, have an HK97-like fold that reflects the common origin of the capsids of tailed viruses and encapsulins [17]. The details of the lineage and the functions of the last common ancestors of phage capsids and encapsulin, though, are still hypothesised. In our previous work [18], the structural models of MCPs and terminases and a phylogenetic analysis based on protein sequences were used to reveal the degree of relationship between 50 phage families classified by ICTV. The clustering that was performed has placed analysed families of ϕ 29-related phages into a clade distinct from other groups, but reliable conclusions about the closest relatives of ϕ 29-related phages can hardly be drawn due to the early divergence of this group, low statistical support (or its absence) and possibly inaccurate models. To the authors' knowledge, the origin and evolution of short podoviral tails, including those of ϕ 29-like phages, remain even less understood. Previously published results indicate that both the N-terminal domain of the tail knob protein of *Streptococcus* phage C1 (*Rountreeviridae* family) and, possibly, phage ϕ 29 tail knob protein, forming a hexameric ring, have a common protein fold usually found in the neck and tail tube proteins of siphoviruses and myoviruses. This allows to hypothesise that all three morphological groups of tailed viruses have a common origin [19]. Nevertheless, the authors mention [19] that these "N-terminal domains do not form a continuous β -barrel through tight interactions within a hexamer, as observed in other tube proteins".

Taxonomy and evolutionary history are tightly connected. The understanding of evolutionary relationships of viral proteins is important to plot a classifying scheme that reflects a "natural" evolutionary relationship between viral taxa. Currently, the classification of viruses is primarily based on the results of genomic analysis. In the current paper, modern genomic and structural prediction tools are used to provide an overview of the typical characteristics of ϕ 29-like phages, to search for representatives in databases including metagenomes, to suggest the evolutionary traits that led to the establishment of this

group of bacteriophages, and to reveal evolutionary relationship between the proteins of $\phi 29$ -related phages and other viruses/cellular organisms. First, the published data on the structure of the genome and virions of $\phi 29$ -like phages are reviewed for a better understanding of gene functions. Next, a thorough search for sequences similar to phage $\phi 29$ protein sequences is performed using different methods. Then, the found genomes are clustered and representative sequences are annotated and studied. Finally, we discuss the results and their implications for phage taxonomy, and propose hypotheses that may help in understanding the evolution of this group of phages.

2. Results

2.1. Brief Review of Published Studies on Phage $\phi 29$ Genomics and Structural Features

2.1.1. General Genomic Features of Phage $\phi 29$

The genome of *Bacillus* phage $\phi 29$ consists of a linear dsDNA molecule of 19,282 bp (NCBI Accession #EU771092.1) containing short inverted terminal repeats (5' AAAGTA) [20]. Terminal proteins (TPs) are covalently attached at 5'- termini of each strand of the genomic DNA [21]. The $\phi 29$ genome contains 27 protein-encoding genes with a coding capacity of 94.8%. There are no tRNA genes in the genome, but it encodes for 174 nt prohead RNA (pRNA, from 147 to 320 bp), which is essential for the packaging of the genome [22,23]. The functions of most proteins have been characterised to a greater or lesser extent.

Like many other tailed phages, the $\phi 29$ genome has a distinct modular structure (Figure 1). Three regions, containing blocks of genes responsible for different stages of infection, in order of transcription during the infection cycle, can be distinguished—these are the right and left early regions and a central late region [24,25]. Transcription is regulated by several promoters, including early promoters C1, C2, A1, A2a, A2b and A2c (promoters A2a and A2b are located close to A2c) and late promoters A3, B1 and B2 [26]. Compared with other $\phi 29$ promoters, promoters B1 and B2 resulted in minor amounts of RNA synthesis in vivo [27]. Two Rho-independent terminators for the early and late transcripts are positioned between late- and rightward-reading early regions. Another Rho-independent terminator, presumably facilitating the effective production of proteins p5 and p6, is located within gene 4 [28]. The left early region does not appear to contain Rho-independent transcription, meaning that transcription starting from the A1 and A2 promoters continues until the left end of the genome is reached [26]. Promoter A3, which is responsible for the formation of the transcript containing all the late region coding sequences, is regulated by protein p4 (a product of early gene 4); p4 binds upstream from RNA polymerase (RNAP) and the p4-RNAP complex interaction stabilises the polymerase at the promoter [29,30]. Four p4 binding sites are located between the A2c and A3 start sites [31]. The A1IV promoter, located in the coding region of the DNA polymerase, promotes the synthesis of the p1 protein and is weakly expressed in vivo [28,32].

Early regions also contain genes comprising the replication machinery of the phage. Like some other prokaryotic and eukaryotic viruses (adenoviruses [33], tectiviruses [34], *Sulfolobus ellipsoid virus 1* [35]) and linear plasmids [36,37]), phage $\phi 29$ features protein-primed replication [38]. The gene of terminal protein p3 is located in the right early region upstream (from the direction of transcription) of the DNA polymerase (DNAP) gene2. The left early region also encodes an unusual dsDNA binding protein, p6, with versatile functions, including participation in the genome organisation of genomic DNA, DNA replication and transcriptional regulation [39–41]. The latter function appears in the repression of the early C2 promoter and in the participation in the binding of the p4-RNAP complex upstream of the late promoter A3 and early promoter A2c. The low specificity of DNA-binding functionally relates p6 to prokaryotic histone-like proteins [40]; this relatedness is reflected in the fact that p6 is sometimes called a “viral histone-like protein”. The phage $\phi 29$ replication process initially uses p17, which supposedly recruits initiation factors at the replication origins in conditions of low concentrations of other replication proteins and becomes unnecessary after the establishment of the infection

process [42]. Protein p17 is encoded by gene17, which is located in the right early region, and is synthesised early at the beginning of infection.

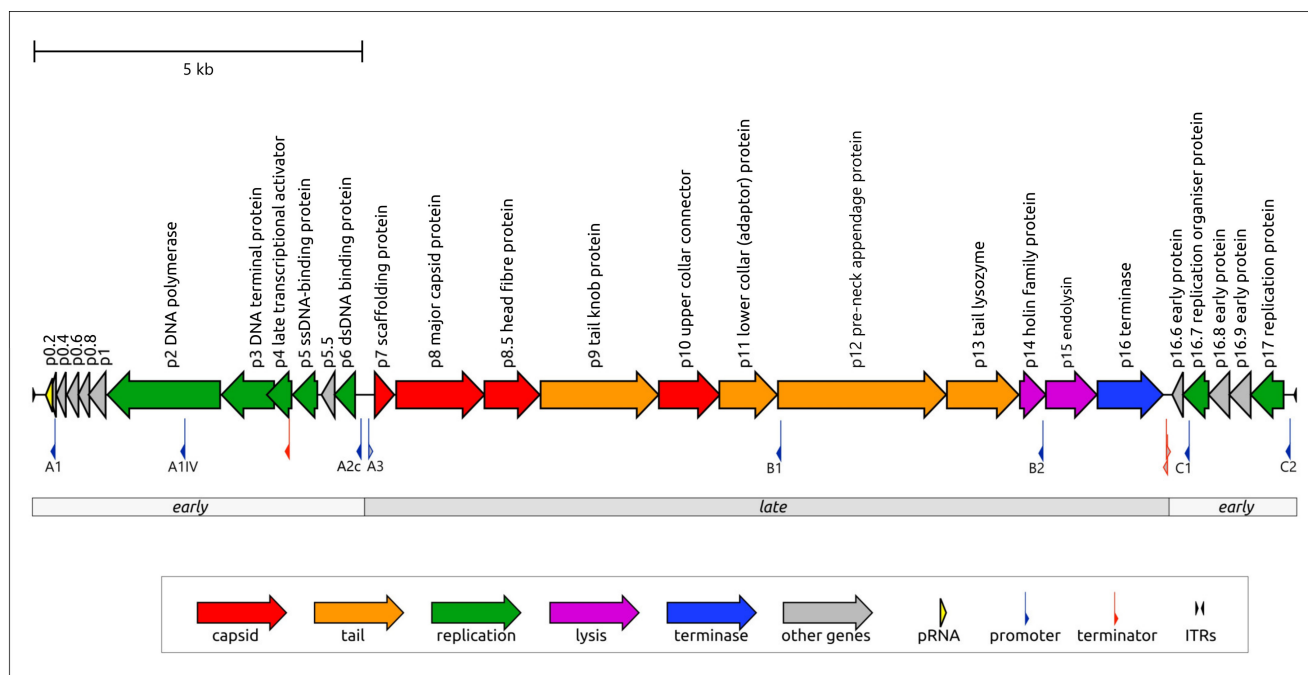


Figure 1. Genetic map of phage $\phi 29$ (based on genomic sequence under NCBI Accession #EU771092). Annotations are shown in labels and legends. Arrows indicate the direction of transcription. The scale bar indicates the length of the nucleotide sequence. The names of gene products are shown in accordance with experimentally assigned functions (see text). The grey bars at the bottom of the map indicate the location of the early and late regions.

The late region of the $\phi 29$ genome comprises a block of structural genes, a lysis module and a DNA encapsidation (packaging) protein (DNA encapsidation ATPase, further referred to as terminase or Ter). The structural module consists of eight genes, seven of which (the exception being the scaffolding protein) are present in the mature virion [43]. Four genes encode the capsid proteins (scaffolding protein p7, major capsid protein p8, head fibre protein p8.5 and upper collar protein p10) and the remaining four genes encode tail structural proteins, including tail protein p9, adaptor p11, pre-neck appendage (tail fibre) protein p12 and lysozyme domain-containing proteins p13, which are localised at the distal tip of the tail knob and participate in the degradation of the cell wall at the penetration stage [44]. The lysis module comprises genes of holin (p14) and endolysin (p15). The last gene of the late region encodes the terminase (p16).

2.1.2. General Characterisation of the $\phi 29$ Virion and Proteins

A mature virion of phage $\phi 29$ consists of a prolate head (capsid) and a short non-contractile tail (Figure 2a). The head is about 45×54 nm in size, with a $T = 3$, $Q = 5$ symmetry. As with other *Caudoviricetes* viruses, a scaffolding protein is required for correct capsid assembly. Experiments with a temperature-sensitive mutant scaffolding protein, however, showed that the capsid can spontaneously assemble into isometric particles ($T = 3$) 37 nm in diameter that lack a head–tail connector and also, to a lesser degree, into larger isometric particles ($T = 4$) 43 nm in diameter [45]. The capsid is decorated with 55 head fibres attached to quasi-3-fold symmetry positions [46]. Each fibre is composed of three monomers of p8.5; the structure of the fibre has a supercoiled triple repeating helix–turn–helix motif (Figure 2b, AF models are provided in Supplementary File S1). These fibres can facilitate the attachment of the phage onto the host–cell wall. The phage connector is located at the portal vertex and is assembled from 12 protomers of upper collar protein (p10) [47]

(Figure 2c). The connector contains a disordered region comprising 12 groups of 18-residue loops, N229–N246, which serves as a ‘clamp’, retaining the DNA within the capsid. The tail of phage ϕ 29 consists of a tail tube, a tail knob and 12 tail appendages (Figure 2c,d). The tail tube consists of 12 copies of lower collar protein p10 and the tail knob is built from six copies of tail knob protein p9 [43]. As well as typical phage tail fibres and tail spikes, ϕ 29 appendages consist of three monomers of protein p12; they participate in the recognition of *B. subtilis* cell-wall teichoic acid, which acts as the primary receptor [46].

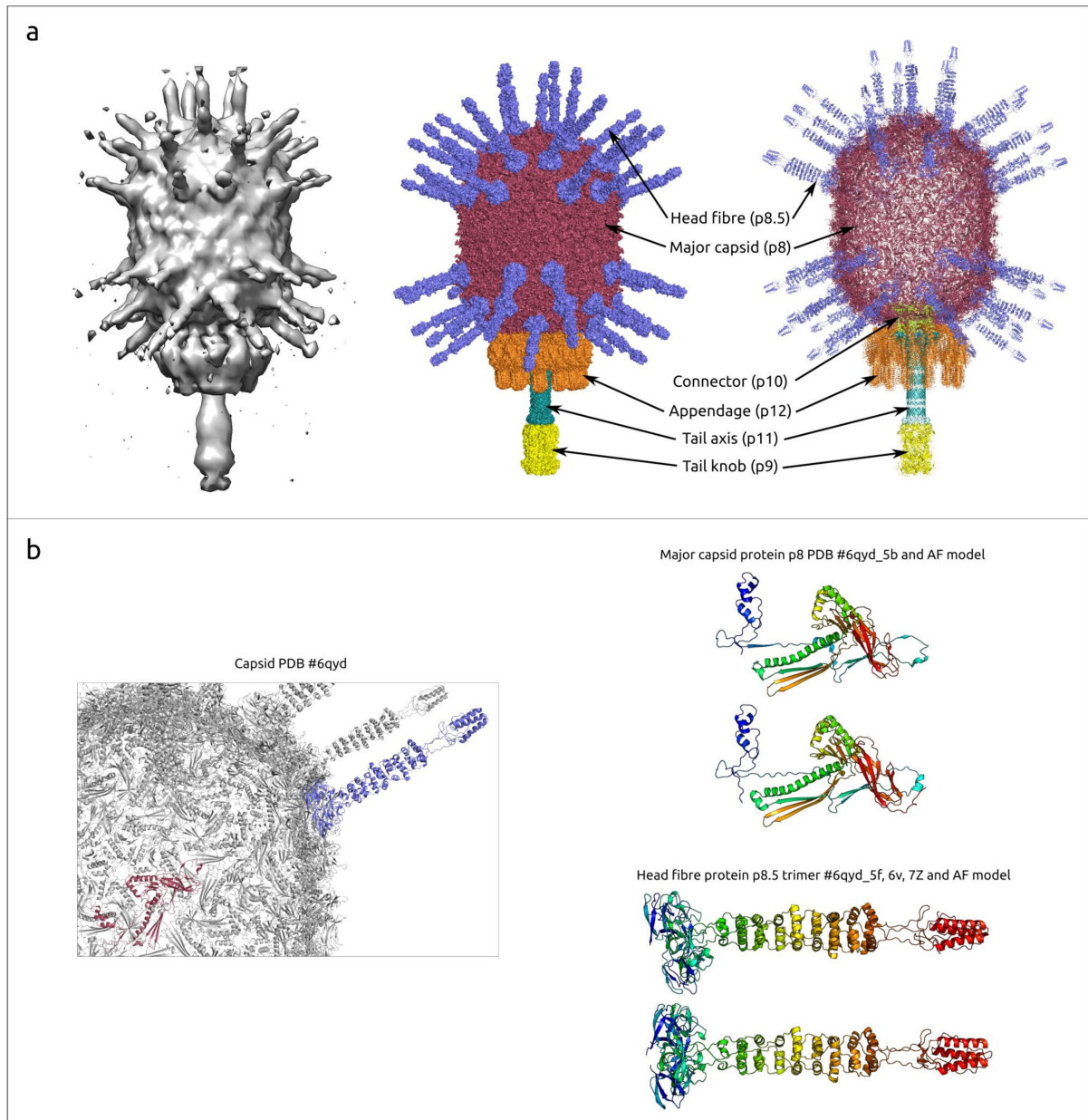


Figure 2. Cont.

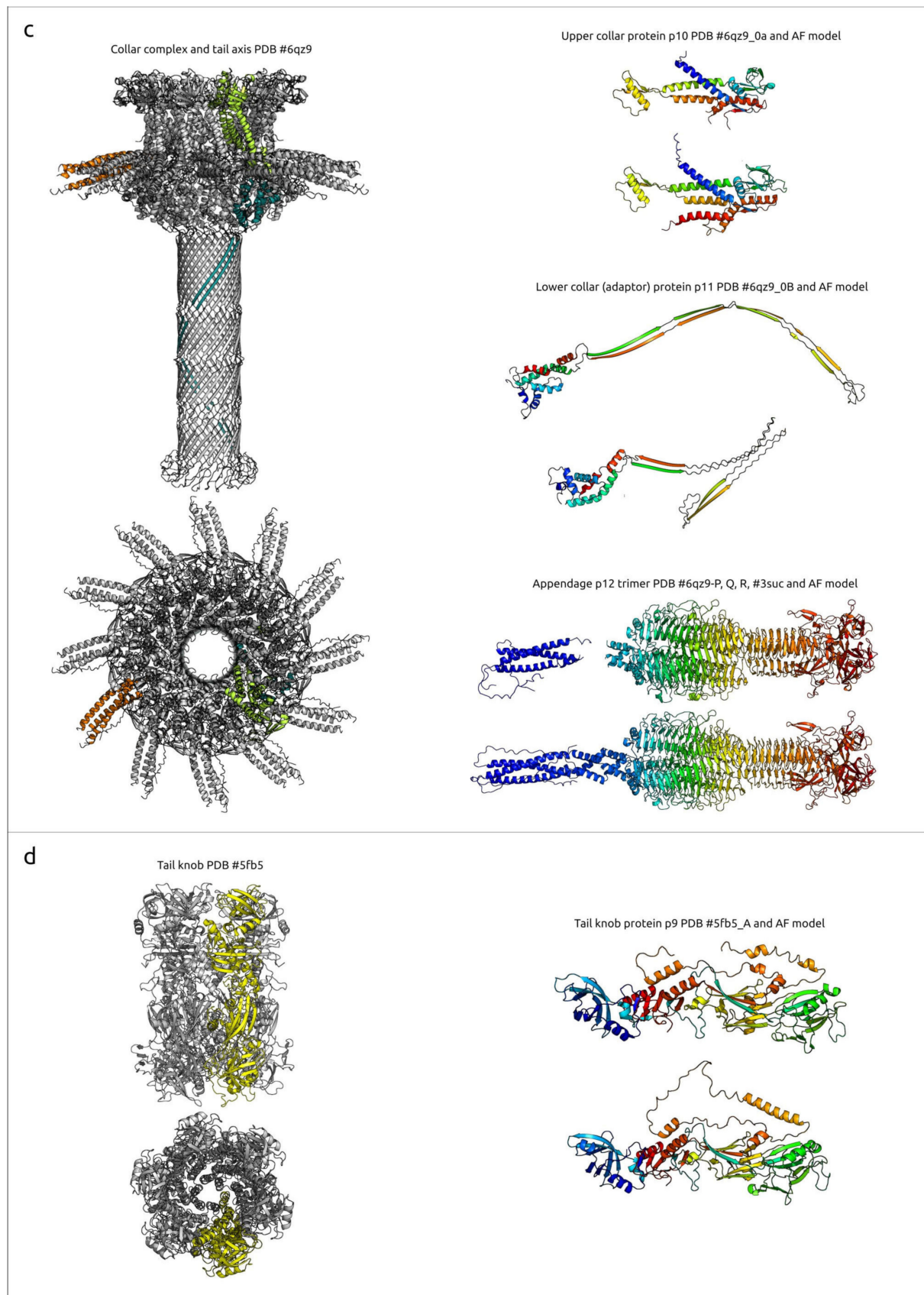


Figure 2. (a) A general view of the $\phi 29$ phage particle, based on cryoEM data (EMDB entry EMD-1502) [48], and a schematic view of bacteriophage $\phi 29$, including surface-rendered and ribbon presentations showing its structural components based on PDB structures 6qyd, 6qz9, 5fb5 and 3gq7 [43,49,50].

(b) A ribbon representation of ϕ 29 capsid components, including major capsid proteins and head fibre trimers, and PDB structures and AF models of corresponding proteins based on PDB structure 6qyd. (c) A ribbon representation of ϕ 29 tail components, and PDB structures and AF models of corresponding proteins based on PDB structure 6qz9 [43]. (d) A ribbon representation of ϕ 29 tail knob, and PDB structures and AF models of corresponding proteins based on PDB structure 5fb5 [49]. Rainbow colouring uses a colour gradient where the N-terminal end is blue and the C-terminus is red.

Phage ϕ 29 and a number of related *Bacillus* phages possess a peculiar packaging apparatus that employs prohead RNA (pRNA) (Supplementary Figure S1) encoded in the 5' end part of the genome (Figure 1) [51]. According to current knowledge, this is a unique feature inherent in these phages. Five copies of pRNA, together with five copies of terminase p16 and a connector (consisting of 12 copies of p10), constitute the ϕ 29 packaging motor [23,43,52]. pRNA is present during phage DNA packaging but is missing in the mature virion. Full-sized pRNA has 174 bases, but shorter, 120-base pRNA is sufficient for efficient packaging [53]; purified proheads contain 120-base pRNA [54]. pRNA sequences were identified or predicted in other ϕ 29-like phages [55,56]. Using a BLAST search, similar sequences can be found in the genomes of other *Salasmaviridae* phages, including *Bacillus* phages BSP2, BSP4, PZA, Arbo1, vB_BsuP-Goe15 and others. The structure of terminase (protein p16), was determined experimentally. Its structural architecture was found to be similar to that of other dsDNA phages, including the two-domain structure comprising the N-terminal ATPase domain. The C-terminal endonuclease domain in the terminases of non- ϕ 29-like phages cleaves the concatemeric genome [57]. However, since the packaging substrate for ϕ 29 is unit-length DNA with covalently attached terminal proteins of ϕ 29, the endonuclease activity of ϕ 29 terminase C-terminal domain is unnecessary [58]. This domain of ϕ 29 Ter belongs to RNase H superfamily proteins and still retains its DNA binding capabilities [59].

Interestingly, the predicted structures of proteins p0.6 and p0.8 are similar, possibly indicating gene duplication events in the evolutionary history of phage ϕ 29. Tail-associated lysin p13 (TAL, tail lysozyme) is important for penetration into the host-cell. The TAL is, apparently, located at the distal end of the ϕ 29 tail knob. This protein has a two-domain structural architecture; the N-terminal domain can cleave the β (1–4) glycosidic bonds between N-acetylglucosamine and N-acetylmuramic acid residues and the C-terminal domain can cleave cross-links in peptidoglycan layers [48].

Experimentally derived and/or AF-predicted structures of other ϕ 29 proteins are shown in Supplementary Figure S2. Apparently, despite the similarity between the predicted and experimentally determined structure, the orientation of protein domains may be predicted incorrectly.

Major capsid protein p8, the main component of the ϕ 29 head, is predicted to be a 448 aa protein encoded by gene 8. The experimentally determined structures of the ϕ 29 MCP have unambiguously shown its attribution to HK97 fold proteins. An HK97 fold is characteristic of tailed phages, herpesviruses and encapsulins [17]. The structures of ϕ 29 proteins forming hexons and pentons of capsid are not identical, but very similar (RMSD about 3 Å), as well as being similar to AF models (RMSD about 1–2 Å for both hexons and pentons) (see Figure 3, Supplementary Figure S3). An interesting feature of the ϕ 29 MCP is the presence of an additional all β -strand immunoglobulin-like (Ig-like) domain [43].

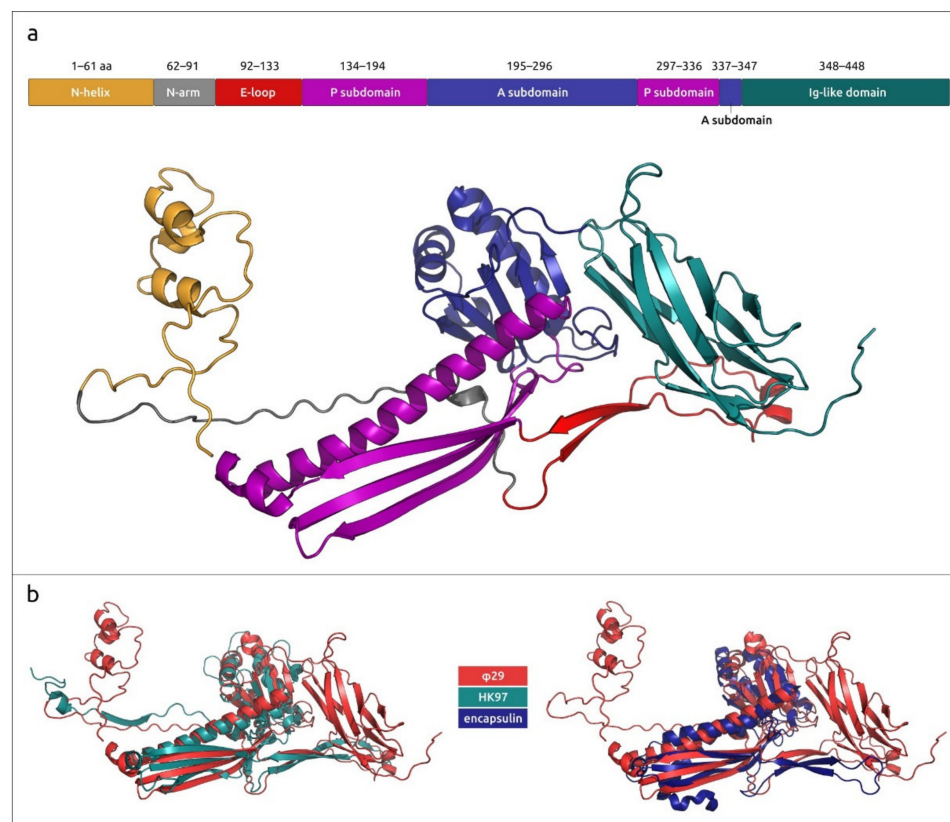


Figure 3. A scheme (a) and ribbon diagram showing the domainal structure of phage $\phi 29$ MCP (AF model), and the superimposition (b) of $\phi 29$ MCP (AF model) and PDB databank structures of phage HK97 MCP (PDB #1ohg [60]) and capsulase from *Thermotoga maritima* (PDB #7kq5 [61]).

Scaffolding protein p7, important for correct head assembly, is missing in the mature virus head. A cryo-EM analysis indicated that the scaffolding protein was organised inside the head as a series of concentric shells; it can bind DNA, and this binding can mediate the structural transition from prohead to mature capsid, liberating scaffolding protein [62]. According to experimental data, scaffolding protein has a helix–loop–helix motif and a disordered tail [43].

The remaining potential proteins encoded in the $\phi 29$ genome are non-structural proteins. Most of them take part in processes related to DNA replication and the regulation of protein synthesis. DNA replication is carried out by the phage’s own DNA polymerase (protein) p2, which belongs to the family B of DNA-dependent DNA polymerases and uses the protein primer for the initiation of replication [38,63]. This is a highly processive and accurate DNA polymerase, which implements both polymerisation and 3′–5′ exonuclease activities [64]. To initiate phage DNA replication, p2 interacts with a molecule of terminal protein, encoded by gene3, located in the left early region. During replication, the terminal protein is covalently bound to the phage DNA through a phosphoester bond between Ser232 and 5′-dAMP [65]. Another $\phi 29$ protein, p1, is also important for successful replication and was shown to be associated with *Bacillus subtilis* cell division protein FtsZ, colocalising with FtsZ in the medial region of *B. subtilis* [66]. Protein p4, as previously described, participates in the transcriptional regulation, and protein p5 apparently acts as a single-stranded DNA-binding protein, facilitating the DNA-replication-stimulating protein p3 [67]. Protein p6, sometimes referred to as a “histone-like protein”, is structurally completely different to cellular histone proteins. The left early region also encodes several small proteins with an unknown function.

It seems that the proteins of $\phi 29$ early genomic regions are responsible mainly for ensuring replication processes. The right early region encodes proteins p16.6, p16.7, p16.8, p16.9 and p17. Protein p16.6 is a small 54 aa protein, whose predicted structure contains

a short, double-stranded β -sheet. Protein p16.7 was found to be involved in the organization of membrane-associated viral DNA replication [26,68]. The structure of p16.7 was determined experimentally (PDB #1zae [69], 2bnk [69], 2c5r [70]).

Phage ϕ 29 has a comparatively simple two-component lysis machinery; protein p14 acts as a typical phage holin, accumulating in the cytoplasmic membrane and disrupting the host–cell membrane when it reaches a critical concentration [71,72]; protein p15 functions as a muramidase, destroying the cell-wall peptidoglycan [73].

2.2. Searches for Similar Proteins Using Sequences and Structures of ϕ 29 Proteins

2.2.1. Major Capsid Protein (p8)

A BLAST search, using sequences of proteins p8 to p11, which compose the virion, indicated a large number of similarities (E-value $< 1 \times 10^{-5}$, using the GenBank PHG database) between these proteins and corresponding proteins of phages belonging to the *Salasmaviridae*, *Guelinviridae* and *Madridviridae* families, genera *Anjalivirus*, *Badaztecivirus* and *Dybvigvirus*, and unclassified phages. All these phages have a small genome, which is comparable with ϕ 29 by size, encoding terminal proteins and DNA polymerase and terminase that are related to ϕ 29 counterparts, according to the results of HHpred and BLAST searches.

A search for PDB structures similar to the major capsid protein (p8) using DALI (the DALI search results for p2, p3, p6, p8, p9, p15 and p16 are provided in Supplementary File S2), indicated that several ϕ 29 virion proteins are structurally similar to their functional counterparts in other phages. Search results using experimentally obtained structures, where they have been published, and AF structures, were nearly identical. The AF-predicted structure of ϕ 29 MCP (p8) is most similar to the structures of the MCP of related *Staphylococcus* phage P68 (PDB #6iat [74], DALI Z-score 19.2) and is very similar (Z-score > 6) to that of other phages with podoviral, siphoviral and myoviral morphology.

When using the parts of MCP corresponding to different domains, BLAST and DALI searches found different related proteins. A BLAST search found homologues of Ig-like domain sequence among both *Salasmaviridae* phage capsid proteins and miscellaneous bacterial proteins not related to phages; these were annotated as S-layer homology domain-containing protein, Ig-like proteins, leucine-rich repeat protein, glycoside hydrolase domain-containing protein and others (E-value $< 10^{-5}$, using the nr NCBI database). A DALI search found the closest structure among some bacterial proteins, including β -1,3-glucanase from *Paenibacillus illinoisensis* (PDB #7r3t [75], Z-score 11.9), Ig-like protein from *Leptospira interrogans* (PDB #8gyr [76], Z-score 10.9), S-layer protein Sap from *Bacillus anthracis* (PDB #6hhu [77], Z-score 9.3), etc. and, interestingly, the C-terminal domain of phage tail tube protein from *Escherichia* phage T5 (PDB #5ngj [78], Z-score 11.3). Both BLAST and DALI searches using the domains common for all tailed HK97 proteins (excluding the Ig-like domain) failed to find similar proteins among the proteins belonging to *Bacillus* bacteria. A BLAST search detected homologues of phage ϕ 29 MCP (with the Ig-like domain excluded) among the phage capsid proteins and a DALI search revealed the most similar structures to be phage capsid proteins and encapsulins.

An HHpred search using the p8 sequence, with the Ig-like domain excluded, revealed an interesting feature of phage ϕ 29 MCP; all related sequences found belonged to small phages that featured protein-primed replication and had terminase and structural proteins related to those of ϕ 29. (Henceforth, these phages will be referred to as “ ϕ 29-related phages” or “ ϕ 29-related viruses”). Unlike searches using MCP sequences of phages of other classified families of bacteriophages and herpesviruses, the HHpred search found no similar sequences among HK97 viruses (*Heunggongvirae* viruses) that did not belong to ϕ 29-related viruses, even though the DALI search clearly showed structural similarities between the MCPs of phage ϕ 29 and other *Heunggongvirae* viruses. This might have been due to the early divergence of ϕ 29-related viruses. A DALI search indicated the relatedness of ϕ 29 MCP (AF model, with the Ig-like domain excluded) to MCPs of other *Duplodnaviria* viruses and encapsulins, and the most similar structures among the viral capsids, besides

other $\varphi 29$ phages, belonged to *Autographiviridae* capsid proteins (*Klebsiella* phage Kp9 and *Escherichia* phage T7, DALI Z-score 16.2–16.7). The clustering of experimentally determined structures of capsid proteins grouped $\varphi 29$ -like phages in a distinct cluster, separate from other encapsulins, herpesviruses and other phages, including those infecting Gram-positive bacteria (Figure 4, Supplementary Figure S4). Interestingly, this clustering placed almost all encapsulin major shell proteins, except the protein from *Synechococcus elongatus* PCC 7942 (PDB code 6x8m [79]), in a distinct group; the latter one was placed closer to the phage MCPs.

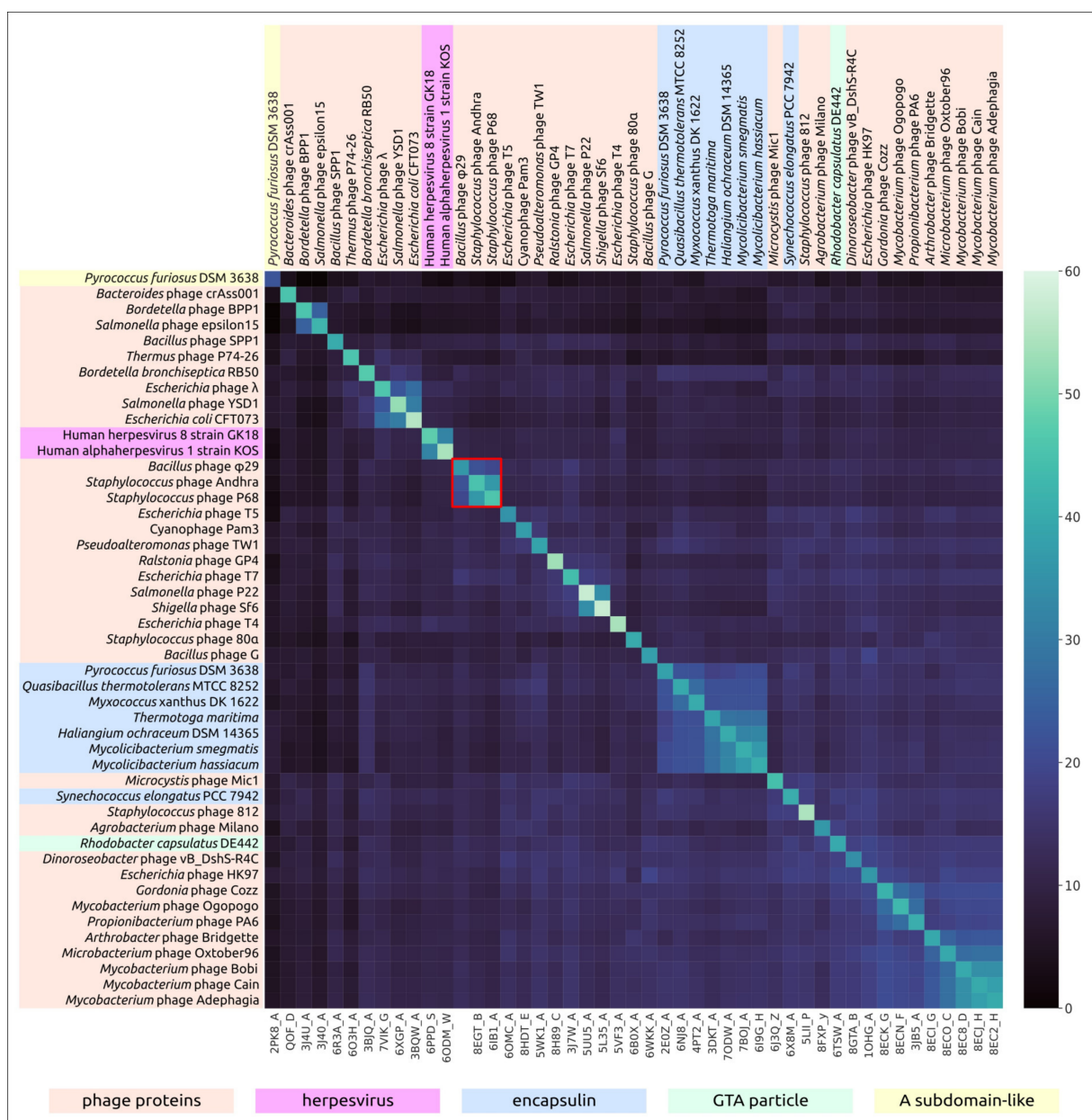


Figure 4. Heatmap based on the pairwise structural similarity of viral capsid proteins and encapsulin major shell proteins, as measured by DALI Z-scores. The cluster of $\varphi 29$ -related phages is outlined in red. The heatmap with numerical values is presented in Supplementary Figure S4.

A search for more distantly similar structures, apart from HK97 capsid proteins and encapsulins, using separated domains E and P of $\varphi 29$ MCP, revealed their vague resemblance ($Z\text{-score} \geq 4.0$) to the pilus-binding domain of an *Escherichia* filamentous phage Ike, eukaryotic DNAP η and bacterial dodecin (Figure 5d). The similarity between dodecin

and HK97 capsid proteins was previously noted by Koonin et al. [80]. The A subdomain of ϕ 29 and other HK97 proteins resembles the uncharacterised protein PF0899 from *Pyrococcus furiosus* (Figure 5b), which was also found previously [81]. As mentioned above, a DALI search also indicated a clear resemblance between the ϕ 29 Ig-like domain and domains of β -1,3-glucanase H from *Paenibacillus illinoisensis* IAM1165 and other carbohydrate metabolism proteins and phage tail tube proteins (Figure 5c). Apparently, the ϕ 29 Ig-like domain was obtained through horizontal gene transfer.

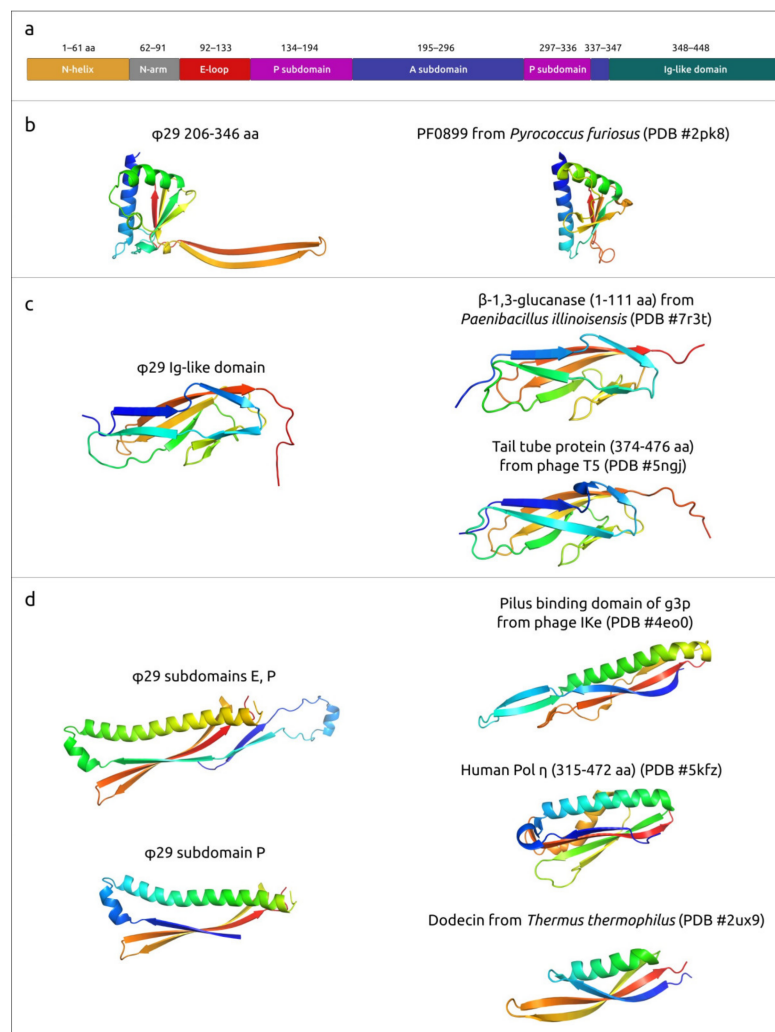


Figure 5. Scheme (a) and ribbon diagrams (b–d) indicating domainial structure similarities between different parts of phage ϕ 29 MCP and similar, experimentally derived non-capsid structures.

Since the amino acid sequences of the MCPs of phage ϕ 29 and related phages show no similarity to capsid proteins of other phages, a phylogenetic analysis was conducted using amino acid sequences of MCP encoded in the genomes of phages belonging to currently classified taxa containing ϕ 29-related, as well as recently isolated ϕ 29-related, phages infecting *Curtobacterium* [10,11]. In general, the tree groups ϕ 29-related phages in accordance with the current ICTV taxonomy, but some taxonomy of some phages seems to be needed for clarification (*Lactococcus* phage asccphi28, *Curtobacterium* and *Amedibacillus* phages), (see Figure 6, Supplementary Figure S5). Interestingly, the tree places phages infecting Gram-negative bacteria and Actinobacteria in two single clades, but the *Enterococcus* and *Streptococcus* phages are polyphyletic groups. The distinct, although partially similar, topologies of the phage and bacterial trees (Supplementary Figures S5 and S6), as well as the polyphyletic nature of ϕ 29-related phages in terms of their host taxonomy,

indicate that the evolution of ϕ 29-related phages was both coevolutionary and may have involved host switching. A phylogenetic analysis using MCP sequences conducted in a previous study [10] showed the close relatedness of ϕ 29-related phages infecting Actinobacteria and the unclassified, unpublished phage *Rhizobium* phage RHph_N3_8, but an analysis of predicted proteins indicated that this phage appears likely to infect an *Actinobacteria bacterium*.

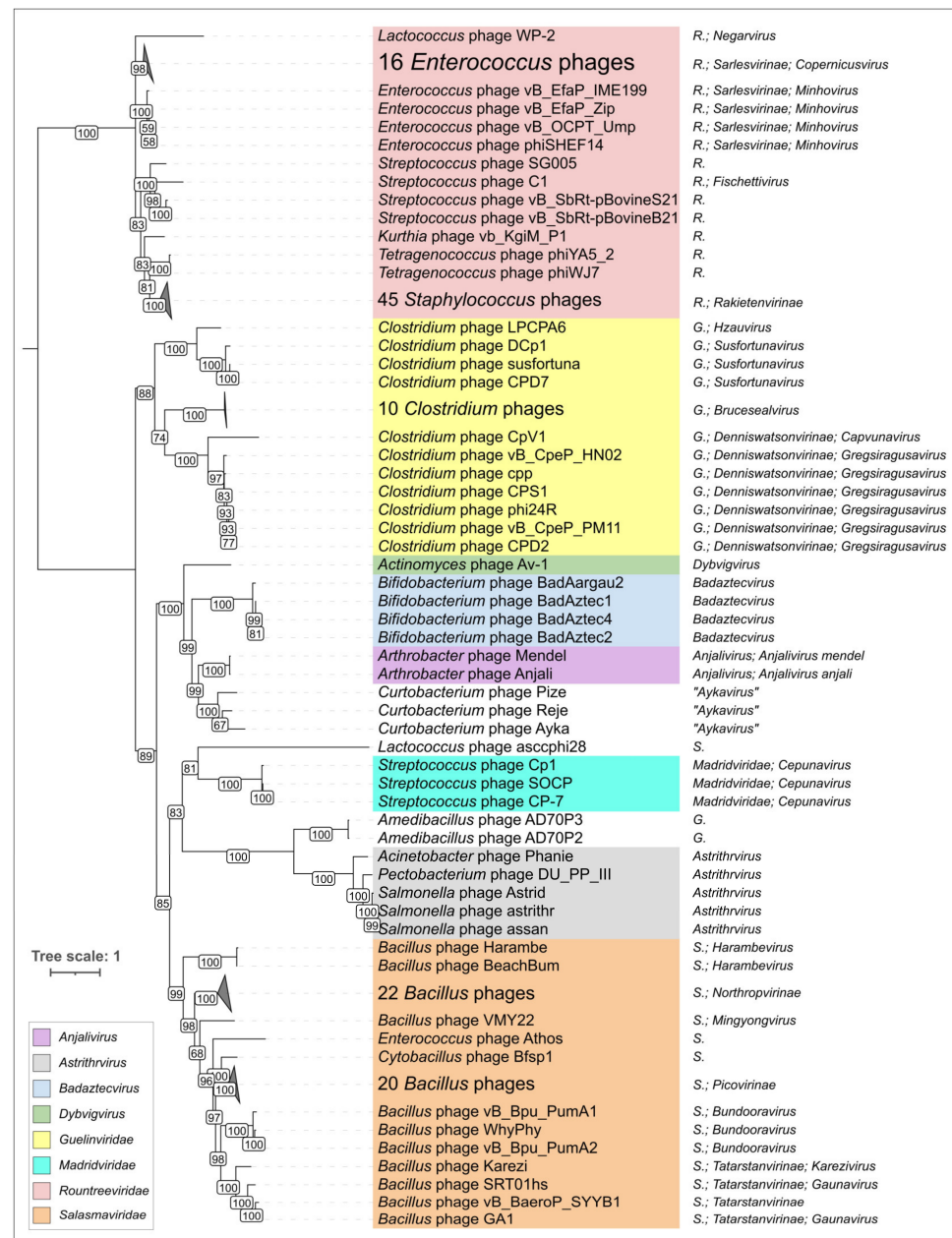


Figure 6. Maximum likelihood phylogenetic trees based on the amino acid sequences of MCP predicted to be encoded in the genomes of isolated ϕ 29-like and related phages. Taxonomy, obtained from the ICTV classification and based on the results of this phylogenetic analysis, is indicated in the legend. Branches with a bootstrap support lower than 50% have been deleted. Some clades have been collapsed, and a tree with all clades expanded is shown in Supplementary Figure S5. Bootstrap values are shown near their branches. The scale bar shows one estimated substitution per site and the tree was rooted to the midpoint.

2.2.2. Head Fibre Protein (p8.5)

A DALI search revealed that the N-terminal part of head fibre protein p8.5 is reminiscent of minor capsid proteins of several other phages of different morphology, e.g., phage YSD1 (PDB #6xgq [82], Z-score 8.1). The similarity of the N-terminal part of p8.5 to the N-terminal parts of minor capsid proteins of other phages was also indicated by HHpred (e.g., siphovirus *Pseudomonas* phage TW1 HHpred probability 97.45% [83]). Also, the HHpred found similarity between the 83 aa C-terminal fragment of p8.5 and a fragment of the head fibre protein gp21 of *Bacteroides* phage crAss001 [84] (HHpred probability 99.3%). A DALI search did not detect significant structural similarities between the C-terminal part of p8.5 and other proteins (the Z-scores for all matches found, except ϕ 29, were 2.9 or lower).

2.2.3. Tail Knob Protein (p9), Upper Collar Protein (p10) and Lower Collar Protein (p11)

The upper collar protein p10 is structurally similar to portal proteins from many other phages, including those distantly related (*Bacillus* phage GA1, PDB #7pv2 [85], Z-score 22.2; *Escherichia* phage T7, PDB #6tjp [86], Z-score 9.8; *Escherichia* phage λ #8k38 [87], Z-score 7.8; *Escherichia* phage T4 #3ja7 [88], Z-score 6.0) and, remarkably, the counterpart from the GTA particle *Rhodobacter capsulatus* (PDB #6toa [89], Z-score 7.4). The upper collar protein has a structure similar to that of other phages of different morphotypes, including podoviruses, siphoviruses and myoviruses. Interestingly, the structure of p10 has similarities with 30S ribosomal protein (r-protein) S15 (PDB #5lmv [90], Z-score 6.2, and other PDB entries), as 30S r-protein S15 is the primary binding protein that orchestrates the assembly of ribosomal proteins S6, S11, S18 and S21 with the central domain of 16S ribosomal RNA to form the 30S subunit platform [91].

The lower collar protein p11 has no structural similarity to proteins from non- ϕ 29-like phages but has a fairly obvious relationship to analogues from related phages. The lower collar protein shows a similarity to the counterpart of *Staphylococcus* phage Andhra (PDB #8egr [9], Z-score 11.5) and *Staphylococcus* phage P68 (PDB #6iac [74], Z-score 10.3). It is noteworthy that the lower collar protein p11 has a detectable level of structural similarity with the protective antigen of the anthrax toxin translocation complex of *Bacillus anthracis* (PDB #7kxr [92], DALI Z-score 4.4; see Figure 7) and with the tube-forming proteins of toxin translocase complexes of other Gram-positive bacteria, e.g., *Clostridium difficile* (PDB #6uwr [93], DALI Z-score 4.2). The overall architecture of these toxin translocation complex tubes and part of the phage tail formed by the ϕ 29 protein p11 is also similar, but differs in the number of monomers that make up the tube. A structural alignment using extended β -sheet domains of the proteins that make up the tube and lower collar protein ϕ 29 and similar structures from translocase complexes yields RMSD of 5.7 Å and 6.6 Å for 7kxr and 6uwr, respectively.

The tail knob protein p9 is structurally close to its counterpart from *Staphylococcus* phage Andhra (PDB #8egs [9], Z-score 19.6) but, surprisingly, its N-fragment has a structural similarity to bovine cationic trypsin (PDB code 7al8 [94], Z-score 6.3). Another interesting case of structural similarity of apparently evolutionarily distant proteins is that between the N-terminal domain of ϕ 29 tail knob protein p9 and the stopper protein of the GTA particle of *Rhodobacter capsulatus* (PDB #6te9 [89], Z-score 4.4). The N-terminal domain of phage ϕ 29 has the same fold as major tube proteins of myoviruses and siphoviruses [19]. A DALI search conducted using the separated structure of the N-fragment only indicated (PDB #5fb5 chain A) its strong resemblance to tail knob proteins of ϕ 29-related phages (Z-score ≥ 10.2) and a significantly weaker resemblance to the tail tip protein of phage λ (PDB #8xcg [95], Z-score 5.8), flavodoxin and other proteins (Z-score 5.4 and lower).

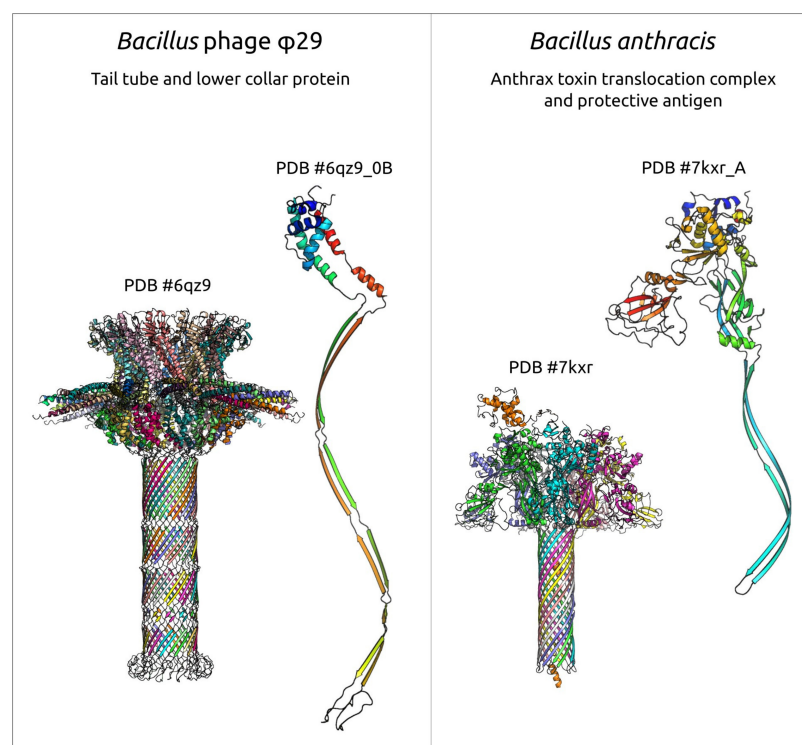


Figure 7. The tail tube of phage $\phi 29$ and lower collar protein p11, and anthrax toxin translocation complex and protective antigen protein from *Bacillus anthracis*. In ternary complexes, different chains have a different colour. Separate monomers are shown using rainbow colouring, with a colour gradient where the N-terminal end is blue and the C-terminus is red.

2.2.4. Appendage Protein (p12)

Protein p12, which makes up $\phi 29$ appendages, has a strong structural similarity to its counterpart of related *Staphylococcus* phage Andhra (PDB #8egr [9], Z-score 20.5) and tail spikes of other phages (podovirus *Salmonella* phage Sf6, PDB #2vbm [96], Z-score 33.5; myovirus *Escherichia* phage CBA120, PDB #6w4q [97], z-score 31.6). Interestingly, an HHpred search pointed to tail fibres of a $\tau 7$ -like marine podophage carin-1 [98] as to the closest non- $\phi 29$ -like phage sequence (HHpred probability 100%, e-value 5.7×10^{-54}). The appendage protein p12 has a noticeable sequence similarity with other *salasmaviridae* phages and many phages belonging to viral groups of very different genome architecture and lifestyle, including the *spbetavirus* genus (genome size of about 130 kb), the *autographiviridae* and *kyanoviridae* families, and other taxa. Also, a blast search found close homologs of p12 in among multiple bacillales proteins (an e-value of about 10^{-51} and lower). This should be a consequence of the evolutionary history of phage receptor-binding proteins (RBPs) accompanied by lateral transfers of RBP domains [99].

2.2.5. DNA Polymerase (p2), Terminal Protein (p3) and “Histone-like Protein” (p6)

There are several experimentally determined structures of $\phi 29$ DNA polymerase including PDB #1xhz [100], 2pzs [101] and others. The structures are very similar to each other and to the AF-predicted structure (the superimposition of 1xhz and AF structure onto 2pzs gives RMSD 0.646 Å and 0.880 Å, respectively). However, the search results are not identical, even though the searches give similar Z-score and RMSD values and lists of found PDB entries. Using the 1xhz structure, DALI points to Human DNA polymerase α (Pol α) catalytic subunit (PDB #5iud [102], Z-score 14.2) as the most similar structure, as well as finding numerous eukaryotic polymerases, and archaeal, bacterial and viral B-family polymerases.

An HHpred search indicated the resemblance of $\phi 29$ DNA polymerase to the DNAP of a $\phi 29$ -like phage *Streptococcus* Cp-1 (HHpred probability 100%, E-value 4.9×10^{-54})

and viruses, not belonging to tailed phages, including *Enterobacteria* phage PRD1 (HHpred probability 100%, E-value 6.2×10^{-42}), *Salterprovirus* His1 (HHpred probability 100%, E-value 7×10^{-41}), adenoviruses, among archaeal, bacterial and eukaryotic polymerases. BLAST searches showed the clear resemblance of $\phi 29$ DNA polymerase to counterparts of phages belonging to the family *Salasmaviridae*, family *Guelinviridae*, family *Madridviridae*, family *Rountreeviridae*, genera *Anjalivirus*, *Astrithrovirus*, *Badaztecivirus* and *Dybvigvirus*, unclassified small-tailed phages and, interestingly, phages of the *Tectiviridae* family infecting both Gram-positive and Gram-negative bacteria. Tectiviruses show a distinct homology between their DNA polymerase and the DNAP of phage $\phi 29$ (E-value of 10^{-11} – 10^{-7} , using the GenBank PHG database). In addition, homologous sequences were found in genomes identified as being Firmicutes plasmids, containing phage structural genes and apparently belonging to *Tectiviridae* or *Salasmaviridae*-related phage groups. Phylogenetic analysis using DNAP sequences placed $\phi 29$ -like phages in a distinct clade, together with bacteria, not far from *Autolykiviridae* and *Tectiviridae* phages and distant from archaea, archaeal and eukaryotic viruses and mitochondria (Figure 8).



Figure 8. Maximum likelihood phylogenetic trees based on amino acid sequences of DNA polymerases related to $\phi 29$ DNAP, according to the results of BLAST and HHpred searches. Taxonomy is indicated in the legend. Branches with a bootstrap support lower than 50% have been deleted. Bootstrap values are shown near their branches. The scale bar shows 0.2 estimated substitutions per site and the tree was rooted to the midpoint.

A BLAST search using a sequence of the terminal protein (TP) p3 found homologous sequences belonging to ϕ 29-like phages infecting Bacillaceae and various mainly metagenomic bacterial sequences. An HHpred search found meaningful matches only between the TPs of phage ϕ 29 and closely related (family *Salasmaviridae*) *Bacillus* phage Nf (HHpred probability 100%). A DALI search using the p3 PDB structure 2ex3 (chain J) did not find similar structures among experimentally determined phage proteins but indicated some level of similarity of the C-terminal coiled-coil-like domain of p3 to various structures, including an amino-terminal coiled-coil domain of the ebolavirus polymerase cofactor VP35. VP35 binds blunt ends of dsRNA ebolavirus genome and is important for a number of crucial processes including viral replication, being a part of viral RNA-dependent RNA polymerase complex [103]. N-terminal coiled-coil is essential for RNA polymerase function [104], but the dsRNA-binding function is associated with the C-terminal domain [105].

The ϕ 29 “histone-like” protein p6, according to the BLAST search results, has close homologs among the proteins encoded in genomes of *Salasmaviridae* phages and *Enterococcus faecium* strains. Neither HHpred nor DALI indicated strong matches to known protein sequences and determined structures.

2.2.6. Terminase (p16)

Like the large subunit of terminase found in un ϕ 29-like phages, the ϕ 29 terminase has a two-domain structure contained the N-terminal ATPase domain (approx. 1–230 aa) and the C-terminal domain. A BLAST search found viral homologues of ϕ 29 terminase in phages belonging to the *Salasmaviridae* family, *Guelinviridae* family and *Madridviridae* family, genera *Anjalivirus*, *Astrithrovirus*, *Badaztecivirus* and *Dybvigvirus*, and unclassified phages, but not amongst the un ϕ 29-like phages. The use of HHpred allowed us to identify remote homology between the N-terminal part of ϕ 29 terminase, and ATPase domains of various tailed phages with different morphology and herpesviruses (a HHpred probability of 97.5% or higher and an E-value of $0.0069\text{--}2.3 \times 10^{-9}$), and the most similar terminases were found in phages *Bacillus* SPP1 and *Shigella* Sf6 (a P22-like podophage of the *Lederbergvirus* genus). At the same time, HHpred failed to detect any significant similarity between the C-terminal domain of ϕ 29 Ter and the endonuclease domains of terminases that do not belong to ϕ 29-like phages. A DALI search using the ATPase domain of ϕ 29 terminase (extracted from PDB structure #7jqj [52]) found similarities between the ϕ 29 N-terminal ATPase domain terminase and the terminases of other dsDNA viruses (including phages of different taxa and morphotypes and herpesviruses) (DALI Z-score 8.3 and higher). Both HHpred and DALI show a similarity of the ATPase domain of ϕ 29 terminase and of DEAD/DEAH-box RNA helicases and other proteins characterised by ASCE fold, which has been noted previously [106].

A DALI search using the C-terminal domain of ϕ 29 terminase (PDB structure #7cnb [107]) found strong similarity to ϕ 29-like phage terminases (DALI Z-score 8.9 and higher) and slight resemblance to only one non- ϕ 29-like phage terminase, *Escherichia* phage P22 (DALI Z-score 3.8, PDB #4dkw [108]). However, when using another experimentally derived structure of the C-terminal domain extracted from structure PDB #7jqj [52], DALI did not detect its resemblance to non- ϕ 29-like phage terminases, possibly due to the relatively low degree of structural similarity. A DALI search using the C-terminal domain of ϕ 29 terminase AF model indicated some similarity to the endonuclease domains of phages HK97, SPP1 and P22 (DALI Z-score 3.0–4.1).

2.2.7. Lysozyme (Endolysin) (p15)

ϕ 29 endolysin was predicted to have a two-domain structure, typical for many phage endolysins [109], where an N-terminal globular domain is an enzymatically active domain (EAD) linked to a C-terminal domain that may be a cell-wall-binding domain (CBD) (Supplementary Figure S2). The functional roles of ϕ 29 endolysin domains are supported by the results of the BLAST, HHpred and DALI analyses. HHpred indicated the relatedness of the EAD to muramidase domains of a prophage endolysins from *Acinetobacter bauman-*

nii [110] (HHpred probability 99.49%) and different phages. A DALI search indicated a high similarity of the ϕ 29 endolysin EAD to the muramidase domain of the phage origin of SpmX from *Asticacaulis excentricus* (DALI Z-score 23.0, PDB #6h9d [111]). In turn, the CBD resembles several sugar-binding proteins (HHpred probability 87.14–95.02%, DALI Z-score up to 9.3). Interestingly, a BLAST search using both the EAD and CBD sequences found highly homologous (E-value $< 10^{-57}$ using the GB PHG database) sequences among endolysins of a dozen (but not all) closely related ϕ 29-like phages of the *Salasmaviridae* family which were considerably homologous, belonging to different evolutionary distant phages (E-value $< 10^{-9}$). Also, BLAST searches indicated that the ϕ 29 endolysin has a high homology to cell-wall hydrolases from different bacilli (E-value $0.0069\text{--}2.3 \times 10^{-9}$). This corresponds to previous findings that ϕ 29 endolysin has a noticeable level of homology to lysozymes from such evolutionary distant phages as *Escherichia* phage T4 and *Salmonella* phage P22 [112]. The obtained results indicate a complex evolutionary history of endolysins, including horizontal transfer events, which may complicate the reconstruction of evolutionary history based on vertical inheritance.

2.3. Search for ϕ 29-Related Viruses

A search for ϕ 29-related viruses was performed using BLAST (tblastn algorithm), amino acid sequences of ϕ 29 MCP and terminase and three GenBank databases: nt, Genome and GB PHG. In addition, the proteins predicted to be encoded in all the GB PHG genomes were checked for relatedness to ϕ 29 MCPs using HHblits. Genomes with an arbitrary chosen cutoff of a genome length of more than 7500 bp were taken for testing and analysis. Combined searches produced 2115 genomic sequences that contained both MCP and Ter genes (Supplementary Table S1). Interestingly, a BLAST search using MCP sequences and the NCBI nt database found phages classified as belonging to the families *Salasmaviridae*, *Gueliniviridae* and *Madridviridae*, and genera *Anjalivirus*, *Astrithrovirus*, *Badaztecivirus* and *Dybvigvirus*, while an HHblits search using MCP sequences and the GB PHG database added representatives of the *Rountreeviridae* family and *Lactococcus* phage KSY1 (*Chopinivirus KSY1*) to the previous results. The latter phage has a much larger genome size, evolutionarily distant DNA polymerase and terminase, and will be discussed separately (see Section 2.7). A phylogenetic analysis of MCPs of ϕ 29-related viruses indicated the separate position of *Rountreeviridae* phages from *Salasmaviridae* and other ϕ 29-related phages (Figure 5, Supplementary Figure S5), which explains why they were not found using a BLAST search only.

Surprisingly, 48 NCBI sequences were identified as archaeal metagenomic sequences from the gastrointestinal tract, including 45 Euryarchaeota, 1 *Candidatus* Thermoplasmata and 2 unclassified Archaea obtained from horse faeces from the endurance horse Fontainebleau. Twenty-four of the sequences were defined as being *Enterococcus faecalis* plasmid and one sequence was identified as a *Clostridium beijerinckii* plasmid. The remaining 2042 sequences were listed as viruses, but only 169 of them were attributed to specific viral taxa at the family level and below, (the same taxa specified as being related in the previous section, in the families *Gueliniviridae*, *Madridviridae*, *Rountreeviridae* and *Salasmaviridae*, and in the genera *Anjalivirus*, *Astrithrovirus*, *Badaztecivirus* and *Dybvigvirus*). The length of the sequences found ranged from 7914 bp (MAG, Euryarchaeota archaeon isolate USDE_maxbin.043 k141_1210644) to 34,630 bp (*Enterococcus faecalis* strain BE54 plasmid pBE54_1); GC content ranged from 22.8% (MAG, *Caudoviricetes* sp. isolate ctjXJ9) to 65.9% (MAG, *Caudoviricetes* sp. isolate ctfy85). According to GenBank attributes, isolated viruses infected the Gram-negative bacteria of genera *Acinetobacter*, *Pectobacterium*, *Rhizobium* (questionable) and *Salmonella*, and Gram-positive bacteria of genera *Actinomyces*, *Amedibacillus*, *Arthrobacter*, *Bacillus*, *Bifidobacterium*, *Clostridium*, *Curtobacterium*, *Cytobacillus*, *Eggerthella*, *Enterococcus*, *Glutamicibacter*, *Hungatella*, *Kurthia*, *Lactococcus*, *Microbacterium*, *Staphylococcus*, *Streptococcus*, *Tetragenococcus* and *Weissella*.

2.4. Clustering of ϕ 29-Related Viruses

To cluster the 2115 sequences found, two separate networks were constructed, using Ter and MCP sequence alignments based on identity scores from DIAMOND BLASTP. Using the Modularity clustering algorithm in Gephi, 30 clusters among MCP sequences, and 24 clusters among Ter sequences were identified (Figure 9, Supplementary Table S1).

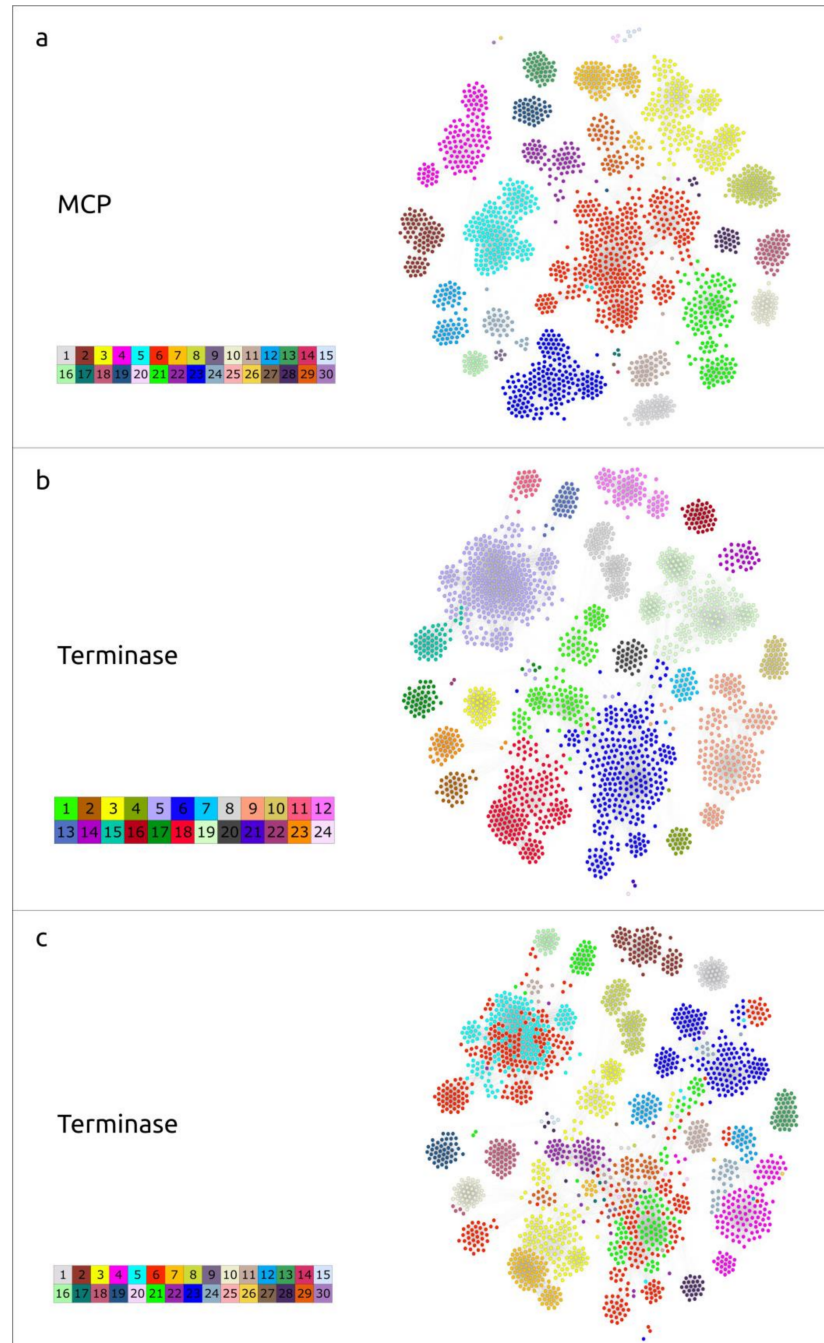


Figure 9. Networks showing the distribution of 2115 MCP sequences generated using the Modularity clustering algorithm in Gephi. Each node corresponds to a distinct sequence of the capsid protein (left panel) and terminase (right panel). All nodes are coloured in different colours and numbered according to their membership in clusters and legends (Supplementary Table S1). (a,c) Nodes are coloured according to the results of MCP clustering; (b) nodes are coloured according to the results of terminase clustering.

Despite the difference in the number of clusters in these two groups, the general pattern of MCP and Ter distribution was similar; most members of classified phage groups were grouped together into the same cluster, for both the major capsid protein and terminase. For example, 56 (out of 60 in total) of the phages belonging to the *Salasmaviridae* family, according to NCBI Taxonomy attributes, were placed in MCP cluster 1, together with two sequences classified as *Enterococcus faecium* and 12 unclassified bacteriophages. In the case of the Ter clustering, the situation was as follows: the same 56 *Salasmaviridae* phages and two sequences from *Enterococcus faecium* were arranged together in the same cluster, which also contained three metagenomic sequences attributed as *Methanobacteriaceae* and 85 unclassified bacteriophages. The clustering of the remaining representatives of the *Salasmaviridae* phages, with both MCP and Ter, was similar: those of *Lactococcus* phage asccphi28 were placed in distinct clusters and three others were arranged together in other distinct clusters.

A noticeable difference can be observed when analysing the clustering of the *Guelinviridae* phages. According to the results of the MCP clustering, the representatives of the family *Guelinviridae* were arranged into three distinct clusters: *Clostridium* (twenty-three phages placed in cluster 29), *Amedibacillus* (two phages placed in cluster 24) and *Eggerthella* phage (one phage in cluster 6). In the case of terminase, *Clostridium* phages were split almost equally (by 11 and 12 representatives) between two clusters: 6 and 18, respectively. Cluster 6 also included the *Eggerthella* phage, with *Amedibacillus* phages allocated to cluster 9. Cluster 18, according to the Ter grouping, was a mixture of representatives of different taxa; it included species belonging to *Guelinviridae*, *Anjalivirus* and *Dyboivirus*.

The smaller number of clusters for the terminase sequences could, presumably, be explained by its more conserved nature, compared with the major capsid protein. Differences in clusterisation may also suggest recombination events in evolutionary history leading to the emergence of mosaic genomes. Among the Ter clusters, there was only one 'singleton cluster'—cluster 24—compared with five singleton clusters in the MCPs group. This CANXRI000000000 sequence fell far out of the coherent picture of the network constituted by terminase. Given that MCP is a hallmark protein characteristic of most known viruses, further genomic analyses were conducted based on the results of MCP clustering.

2.5. Genome Organisation of ϕ 29-Related Viruses: Main Structural Proteins

Genome organisation within the MCP clusters was similar, although some rearrangements can be observed. The general features of arbitrarily chosen representatives from each cluster are shown in Table 1. Only 14 clusters contained isolated phages with a known host. Putative hosts for other phages and possible viruses attributed as Archaea were suggested based on the similarity of predicted proteins, primarily cell-wall-degrading enzymes, to proteins encoded in cellular and viral genomes found in NCBI databases. Apparently, putatively archaeal sequences were misclassified, and belong to bacteriophages. HHblits/HHpred and DALI searches using AF models of predicted proteins indicated that all representative sequences of ϕ 29-related viruses contain proteins related to ϕ 29 proteins p8 (major capsid protein), p9 (tail knob protein), p10 (upper collar protein), p11 (lower collar protein), p2 (DNA polymerase) and p16 (terminase) (Figure 10). These searches indicated that the most similar sequences show either similarity to proteins of ϕ 29-related viruses only, or they are more similar to their counterparts from ϕ 29-related viruses than to proteins from other viruses or cellular organisms. It appears that the 'main' morphogenetic proteins (p8–p11), as well as DNAP and Ter of ϕ 29-related viruses, share a common origin and are a 'hallmark' of the whole group. The BLAST and HHblits/HHpred searches showed that cell-wall-degrading enzymes or protein domains of endolysins and receptor-binding proteins are often more similar to their counterparts from phages that are evolutionarily distant from ϕ 29-related viruses, or from bacteria. This indicates horizontal transfers of cell-wall-degrading enzymes between phages, with the possible involvement of their hosts. It is often not easy to identify the protein cell-wall-degrading domain to which they belong; genomes of ϕ 29-related phages can encode different proteins, including tail tip, tail or head

spike proteins and endolysins, that possess peptidoglycan-degrading activity. Interestingly, some sequences contain characteristic plasmid genes, including all three components of a typical plasmid segregation system (a ParA-type ATPase gene, a gene of a DNA-binding adapter protein with the RHH fold, and a region of DNA with repeats, to which the adapter protein probably binds) as well as a toxin–antitoxin system, suggesting that such sequences belong to phage plasmids, which is in agreement with NCBI GenBank annotations (i.e., cluster 22, *Enterococcus faecalis* BE32 plasmid pBE32_2) [113]. Most ϕ 29-related phages, including all classified phages with genome sequences published by early 2024, do not appear to contain tRNA genes, but some sequences belonging to clusters 04–06, 08, 11, and 21 have been predicted to encode tRNAs.

Table 1. General genome features of sequences representing 30 groups of ϕ 29-related viruses obtained by MCP clustering, and *Lactococcus* phage KSY1.

Cluster	Name	Sequence Length, bp	% GC	Host	NCBI Taxonomy	Isolation Source, NCBI
01	<i>Bacillus</i> phage phi29	19,282	40.0%	<i>Bacillus subtilis</i>	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes; Salasmaviridae; Picovirinae; Salasvirus; Salasvirus phi29	
02	<i>Staphylococcus</i> phage Andhra	18,546	29.8%	<i>Staphylococcus epidermidis</i>	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes; Rountreeviridae; Rakietenvirinae; Andhravirus; Andhravirus andhra	
03	<i>Arthrobacter</i> phage Anjali	19,679	59.4%	<i>Arthrobacter globiformis</i>	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes; Anjalivirus; Anjalivirus anjali	Soil
04	<i>Caudoviricetes</i> sp. isolate ctLx22	14,372	35.7%	Bacillota (suggested)	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes	Human Metagenome
05	<i>Caudoviricetes</i> sp. isolate ctOKO3	17,065	35.7%	<i>Clostridium</i> (suggested)	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes	Human Metagenome
05	<i>Methanobrevibacter</i> sp. isolate RGIG9610	16,264	34.6%	Bacillota (suggested)	Archaea; Euryarchaeota; Methanomada group; Methanobacteria; Methanobacteriales; Methanobacteriaceae; Methanobrevibacter	Ruminant gastrointestinal tract
06	<i>Caudoviricetes</i> sp. isolate ct1aZ12	15,307	42.6%	Bacillota (suggested)	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes	Human Metagenome
06	<i>Clostridium</i> phage CS229P1	16,473	37.8%	<i>Clostridium symbiosum</i>	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes	
07	<i>Actinomyces</i> phage Av-1	17,171	49.5%	<i>Actinomyces naeslundii</i>	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes; Dybvigvirus; Dybvigvirus Av1	
08	Human gut phage 3064_54045	15,891	51.2%	<i>Bifidobacterium</i> (suggested)	Viruses	Faeces
09	<i>Caudoviricetes</i> sp. isolate ctiW69	14,784	41.5%	<i>Clostridium</i> (suggested)	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes	Human Metagenome
10	<i>Bifidobacterium</i> phage BD811P1	17,657	54.8%	<i>Bifidobacterium dentium</i>	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes	Municipal sewage

Table 1. Cont.

Cluster	Name	Sequence Length, bp	% GC	Host	NCBI Taxonomy	Isolation Source, NCBI
11	Bacteriophage sp. isolate ct4VF5	16,226	38.7%	Clostridia (suggested)	Viruses	Human Metagenome
12	Caudoviricetes sp. isolate ctFJX15	12,830	38.7%	Actinomycetia (suggested)	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes	Human Metagenome
13	Virus sp. ctxAI8	13,382	49.8%	Clostridia (suggested)	Viruses	Human Metagenome
14	Caudoviricetes sp. isolate ctPQG16	12,323	53.5%	Gram-positive (suggested)	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes	Human Metagenome
15	Bacteriophage sp. isolate ctLK83	14,962	40.1%	Bacillota (suggested)	Viruses	Human Metagenome
16	Caudoviricetes sp. isolate ctF653	15,970	41.3%	Clostridia (suggested)	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes	Human Metagenome
17	Caudoviricetes sp. isolate ctjx911	13,168	37.8%	Bacilli (suggested)	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes	Human Metagenome
18	Bifidobacterium phage BadAztec2	18,689	48.7%	Bifidobacterium asteroides	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes; Badaztecovirus; Badaztecovirus badaztec1	Honeybee gut
19	Streptococcus phage Cp-1	19,343	38.8%	Streptococcus pneumoniae	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes; Madridoviridae; Cepunavirus; Cepunavirus Cp1	
20	Caudoviricetes sp. isolate ctsL82	13,451	32.4%	Clostridia (suggested)	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes	Human Metagenome
21	Clostridium beijerinckii DSM 6423 plasmid III	16,762	28.4%	Clostridium beijerinckii	Bacteria; Bacillota; Clostridia; Eubacteriales; Clostridiaceae; Clostridium	
21	Clostridium phage HM2	17,470	29.4%	Clostridium saccharoperbutyl-lactonicum	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes	
21	Methanobrevibacter sp. isolate RGIG9632	16,579	34.2%	Clostridia (suggested)	Archaea; Euryarchaeota; Methanomada group; Methanobacteria; Methanobacteriales; Methanobacteriaceae; Methanobrevibacter	Gut metagenome
22	Enterococcus faecalis BE32 plasmid pBE32_2	29,025	32.5%	Enterococcus faecalis	Bacteria; Bacillota; Bacilli; Lactobacillales; Enterococcaceae; Enterococcus	
22	Enterococcus phage EF62phi	30,505	32.7%	Enterococcus faecalis	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes	
23	Salmonella phage Astrid	11,713	39.8%	Salmonella enterica	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes; Astrithrovirus; Astrithrovirus astrithr	Wastewater
24	Amedibacillus phage AD70P2	12,320	26.7%	Amedibacillus sp.	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes; Gueliniviridae	Municipal sewage
25	Caudoviricetes sp. isolate ct2Eo1	13,822	45.1%	Bacillota (suggested)	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes	Human Metagenome

Table 1. Cont.

Cluster	Name	Sequence Length, bp	% GC	Host	NCBI Taxonomy	Isolation Source, NCBI
26	Caudovirales phage NCBI #CAOBPP	13,791	37.5%	Bacillota (suggested)	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes; environmental samples	Human faeces
27	<i>Lactococcus</i> phage asccphi28	18,762	33.7%	<i>Lactococcus lactis</i>	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes; Salasmaviridae	
28	Podoviridae sp. ctnWS46	19,654	32.3%	<i>Streptococcus</i> (suggested)	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes	Human Metagenome
29	<i>Clostridium</i> phage CpV1	16,748	30.5%	<i>Clostridium perfringens</i>	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes; Guelinviridae; Denniswatsonvirinae; Capounavirus; Capounavirus CpV1	
30	Bacteriophage sp. isolate 2692_55609	12,148	42.5%	Bacteroidales (suggested)	Viruses	Faeces

A phylogenetic analysis conducted with amino acid sequences of MCP, Ter and DNAP provided mismatched topologies of trees (Figure 11, Supplementary Figure S7). This can be a result of the complex evolutionary history, accompanied by exchanges of individual genes and genomic modules. The composition of the DNAP tree clades is less similar to that of the MCP and Ter trees; the latter two have more in common, here, suggesting that the lateral exchange of the DNAP gene could occur more often. Some differences in tree topology may be due to the high level of divergence within the large group of ϕ 29-related phages, making it difficult to identify the true evolutionary history.

Genomes of some ϕ 29-related phages and *Lactococcus* phage KSY1 contain two copies of genes encoding the proteins related to ϕ 29 MCP. The corresponding phylogenetic tree indicates close relatedness, and the possible origin of capsid proteins of *Lactococcus* phage KSY1 by gene duplication, but places two predicted capsid proteins of *Clostridium* phage CpV1 (family *Guelinviridae*) in two clades containing other phages possessing only one MCP. This can indicate a duplication of the MCP gene and its subsequent loss during the evolution of ϕ 29-related phages.

The whole-genome-based phylogenomic tree generated by VICTOR (Figure 11b), in some cases, shows patterns of phage distribution similar to trees based on the MCP and Ter sequences. For example, it places the same phages infecting the same Actinobacteria in a distinct clade and indicates relatedness between *Salmonella* phage Astrid and an unclassified phage found in metagenomic sequences, supposedly infecting a Bacteroidia bacterium. The topology of this tree does, however, have its own peculiarities, which can be explained both by the specificity of phage evolution, distinguished by the different evolutionary history of different genomic modules, and by the high level of sequence divergence. According to the results of phylogenetic analysis, cluster c28 is most closely related to cluster c01, which includes phage ϕ 29. A representative of cluster c28, phage Podoviridae sp. ctnWS46, with a similar genome length of 19,654 bp, has very little (1.2%) intergenomic similarity to phage ϕ 29. Comparisons of morphogenesis and replication proteins indicate a clear relationship between these proteins, but the order of a significant proportion of genes differs from ϕ 29, which indicates frequent genetic rearrangements accompanying the evolution of ϕ 29-like phages.

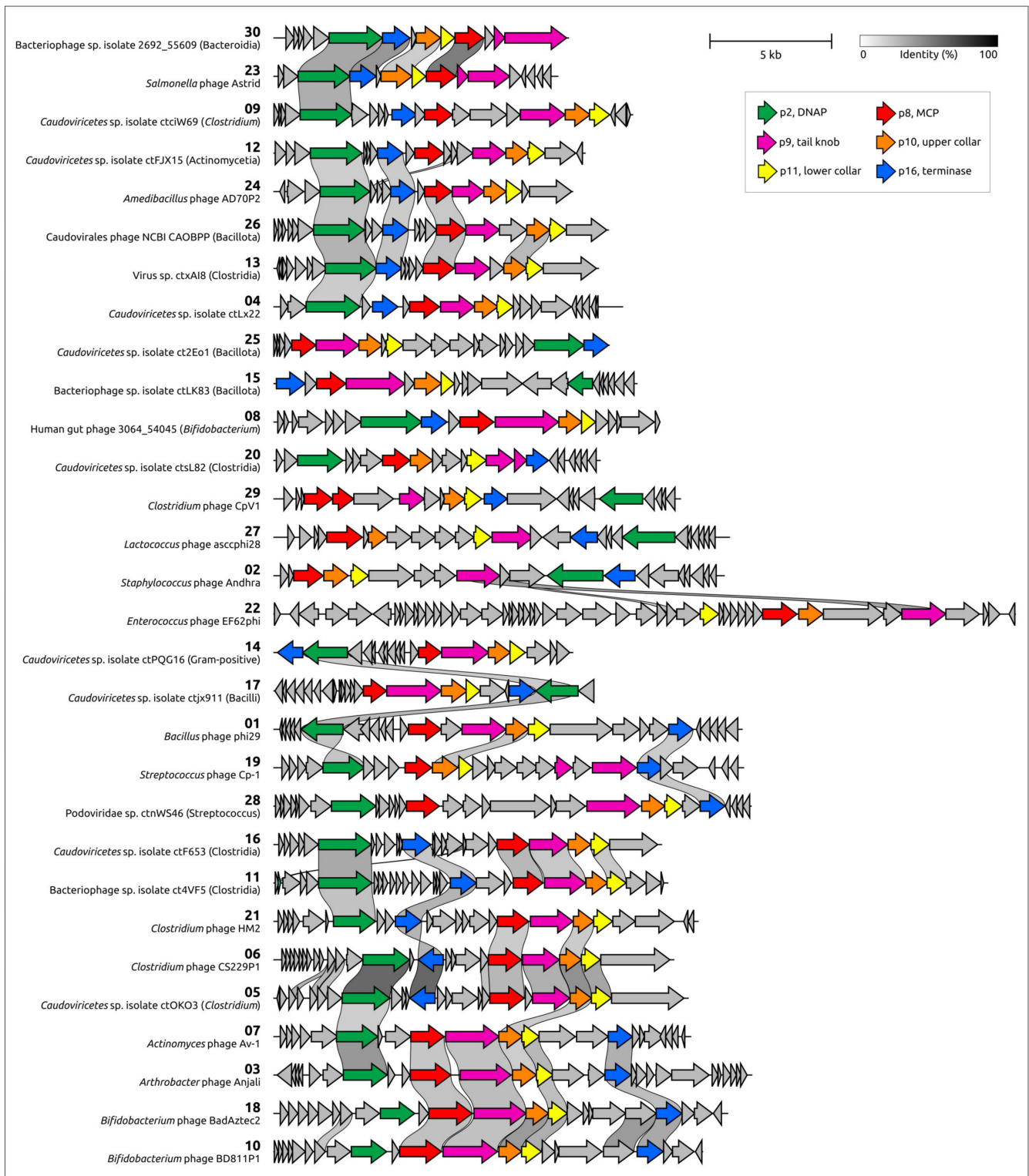
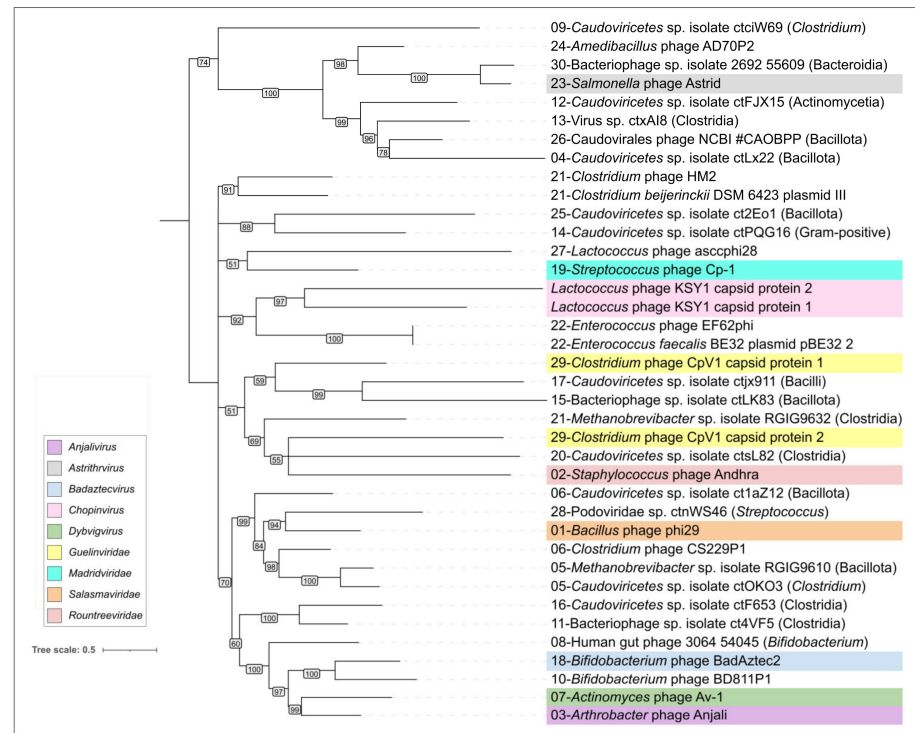
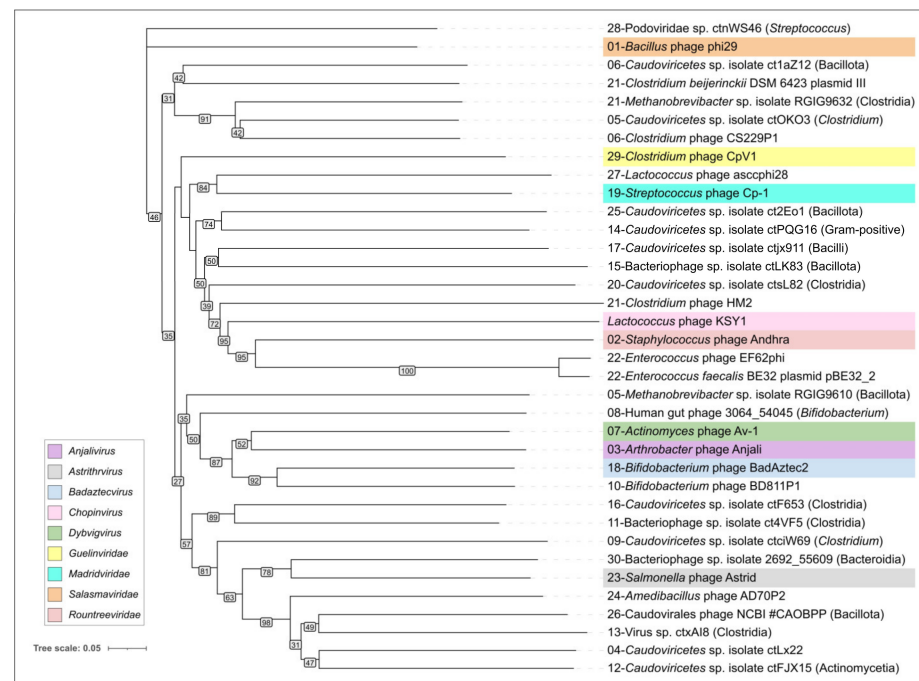


Figure 10. Genetic map of $\phi 29$ -related viruses. Annotations are shown in labels. Arrows indicate the direction of transcription. The scale bar indicates the level of identity between genes according to a black white gradient, the legend for which is shown in the upper-right of the figure. The number of clusters and the name of sequences are shown on the left. Genes encoding proteins similar to $\phi 29$ proteins p2, p8–p11 and p16 are coloured according to the legend.



(a)



(b)

Figure 11. (a) Maximum likelihood phylogenetic trees based on amino acid sequences of MCP, predicted in sequences representing clusters of ϕ 29-like and related phages. The taxonomy of isolated phages is indicated in legends. Branches with a bootstrap support lower than 50% have been deleted. Bootstrap values are shown near their branches. The scale bar shows 0.5 estimated substitutions per site and the trees were rooted to the midpoint. (b) Whole-genome-based phylogenomic tree generated by VICTOR, using translated ORFs, predicted in the genomic sequences of phages representing 30 groups, clustered using amino acid sequences of MCP. VICTOR's pseudo-bootstrap values are indicated near to corresponding branches. The scale bar shows 0.05 estimated substitutions per site.

2.6. Variations in the Structural Architecture of Morphogenetic Proteins of ϕ 29-Related Phages Infecting Gram-Negative Bacteria

While most of the sequences of known ϕ 29-related viruses found apparently belong to phages infecting Gram-positive bacteria, some of them were experimentally found, or predicted, to infect Gram-negative bacteria. Such isolated phages are related to *Salmonella* phage astrithr of the genus *Astrithrovirus* (e.g., *Acinetobacter* phage Phanie [13], *Pectobacterium* phage DU_PP_III, *Salmonella* phages Astrid, astrithr and assan [114,115]) and were assigned to cluster 23, based on the similarity of MCPs. Phages Astrid, astrithr and assan are very close (intergenomic similarity >95%) [13] and represent one species, *Astrithrovirus astrithr*, classified by the ICTV [12]. Phages DU_PP_III and Phanie have a significant level of intergenomic similarity to phages *Astrithrovirus astrithr* (about 30–40%, [13]), but this value is still less than the genus demarcation threshold. In addition, a search for homologous proteins strongly suggested that singleton cluster 30 contains a putative phage infecting a Bacteroidales bacterium, a metagenome-assembled genome annotated as “Bacteriophage sp. isolate 2692_55609” and further referred to as “2692_55609”, found in human faeces. Genomes of phages assigned to clusters 23 and 30 were significantly smaller than the ϕ 29 genomes, being typically about 11–14 kbp. The structural architecture of morphogenetic proteins of such phages, at least in some cases, can be different from phages that infect Gram-positive bacteria.

The genomes of the *Salmonella* phages Astrid, *Pectobacterium* phage DU_PP_III, *Acinetobacter* phage Phanie and phage 2692_55609 have common features that unite them together and distinguish them from the phage ϕ 29 (Figure 12). HMM searches clearly indicate the relationship of the morphogenesis and replication proteins to the proteins of the phage ϕ 29. A list of these proteins includes phage ϕ 29-like DNA polymerase (ϕ 29 p2), terminal protein (counterpart of ϕ 29 p3-like), ssDNA-binding protein (ϕ 29 p5), counterparts of ϕ 29 main virion proteins (MCP, upper and lower collar proteins, tail knob protein, but encoded by two genes, unlike in ϕ 29). The lower collar proteins of these phages are smaller than that of ϕ 29. HHpred and BLAST searches did not reveal genes that encoded tail fibre and tail spike proteins present in some ϕ 29-like phages [9,10]. Interestingly, putative endolysins show resemblance to CHAP domain-containing endolysins and tail-associated lysins of various phages, e.g., an HHpred search revealed clear similarities between the Phanie gp11 and the tail tip lysin of ϕ 29-like phage Andhra, which infects *Staphylococcus epidermidis* [9] (HHpred probability 94%).

A comprehensive study by Pourcel et al. [13] showed that *Acinetobacter* phage Phanie (genome size 11,885 bp) depends on a helper myophage, and that Phanie virions are non-infectious unless they bind to the contractile tail of the unrelated phage Aci01-1, forming chimeric myoviruses. Perhaps other ϕ 29-related phages infecting Gram-negative bacteria may also follow the strategy of multiplication similar to phage Phanie, since they share some its genomic features, including the structure of tail knob protein. Structure predictions, made in [13], indicated a similarity between the phage Phanie gene product 9 (gp9) and the N-terminal part of tail knob protein p11 of phage ϕ 29 and a similarity between the C-terminal part of Phanie gp10 corresponded to the C-terminal part of ϕ 29 tail knob protein p9; AlphaFold 2 modelling of Phanie gp10 predicted a multidomain structural architecture (Figure 13) for this protein, where the C-terminal part of Phanie gp10 corresponded to the C-terminal part of ϕ 29 tail knob protein p9. AlphaFold modelling conducted in present study indicates that, counting from the N-terminus, the predicted structure of gp10 contained six domains, followed by the C-terminal part. Neither the BLAST nor the HHpred searches revealed the significant similarity of these domains to any known proteins with an experimentally determined function; the DALI search also found no similar viral structures, but revealed a moderate level of structural similarity (Z-score 5.1–6.1) between all domains of gp10 and an Ig-like non-catalytic domain of L,D-transpeptidase from *Mycobacterium tuberculosis* (PDB code 4Z7A [116]) involved in the biosynthesis of peptidoglycan (PG) [116]. A comparison of predicted structures of putative proteins composing the tail knobs of phages Astrid, DU_PP_III and 2692_55609

indicated the same general plan of their structural architecture, including an obvious similarity of structures of proteins, similar to the N-terminal part of $\phi 29$ tail knob proteins, but the predicted proteins, containing the counterpart of the C-terminal part of $\phi 29$ tail knob proteins, have a different number of additional Ig-like domains (Figure 13). As was mentioned in the Introduction section, the N-terminal part of $\phi 29$ tail knob protein p9 has a fold found in the neck and tail tube proteins of siphoviruses [19]; the predicted proteins of phages Astrid, DU_PP_III, Phanie and 2692_55609, which are similar to the N-terminal part of $\phi 29$ p9, contain the similarly folded domain as well.

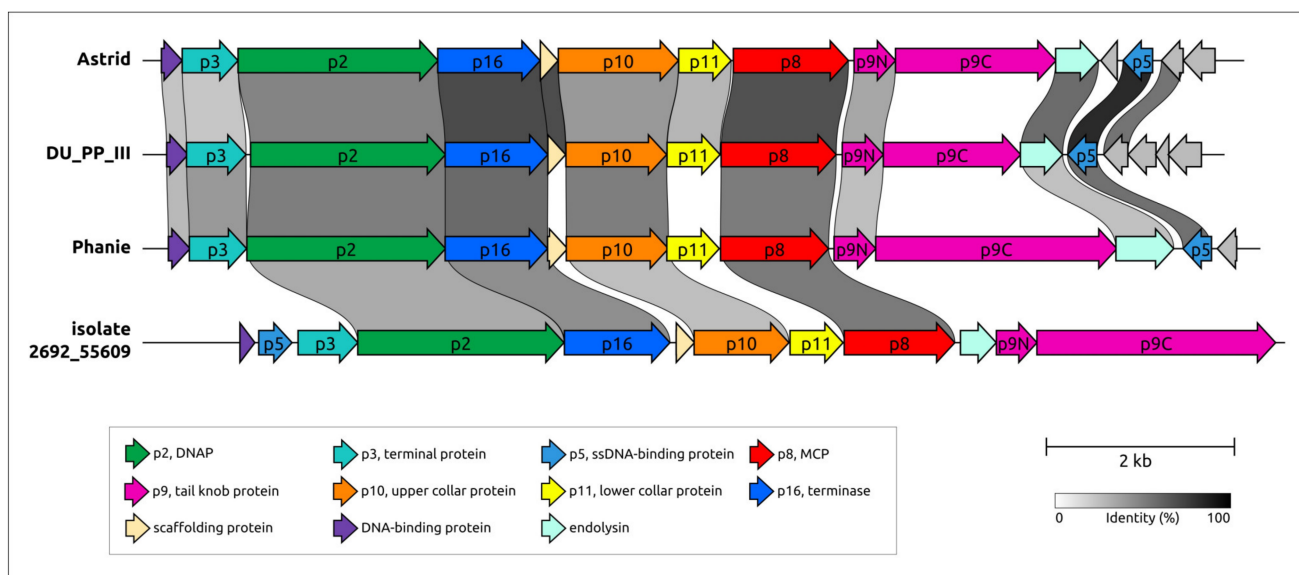


Figure 12. Genetic map of phages belonging to clusters 23 and 30. Phages' names are as follows: Astrid, *Salmonella* phages Astrid; DU_PP_III, *Pectobacterium* phage DU_PP_III; Phanie, *Acinetobacter* phage Phanie; isolate 2692_55609, Bacteriophage sp. isolate 2692_55609. Annotations based on the results of HMM searches, AF predictions and an earlier-published genomic analysis of phage Phanie [13] are shown in labels. Arrows indicate the direction of transcription. The scale bar indicates the level of identity between genes according to a black–white gradient, the legend for which is shown at the bottom of the figure. The number of clusters and the name of sequences are shown on the left. Genes encoding proteins similar to $\phi 29$ (p2, p5, p6, p8–p11, p16) and other proteins are coloured according to the legend.

The AlphaFold modelling complexes that uses six copies of each protein, including both supposedly composing the N-part (minor part) of the tail knob of the phages Astrid, DU_PP_II and the C-part (major part), can offer an interesting suggestion for the interactions of the proteins forming the tail knob (Figure 14); both predicted structures contain β -sheets formed by two different monomers. Additional Ig-like domains are located laterally in both structures. The ipTM score of the DU_PP_III model is 0.69, the pTM score is 0.70 and, for the complex containing the Astrid proteins, the scores are lower (0.51 and 0.52, respectively).

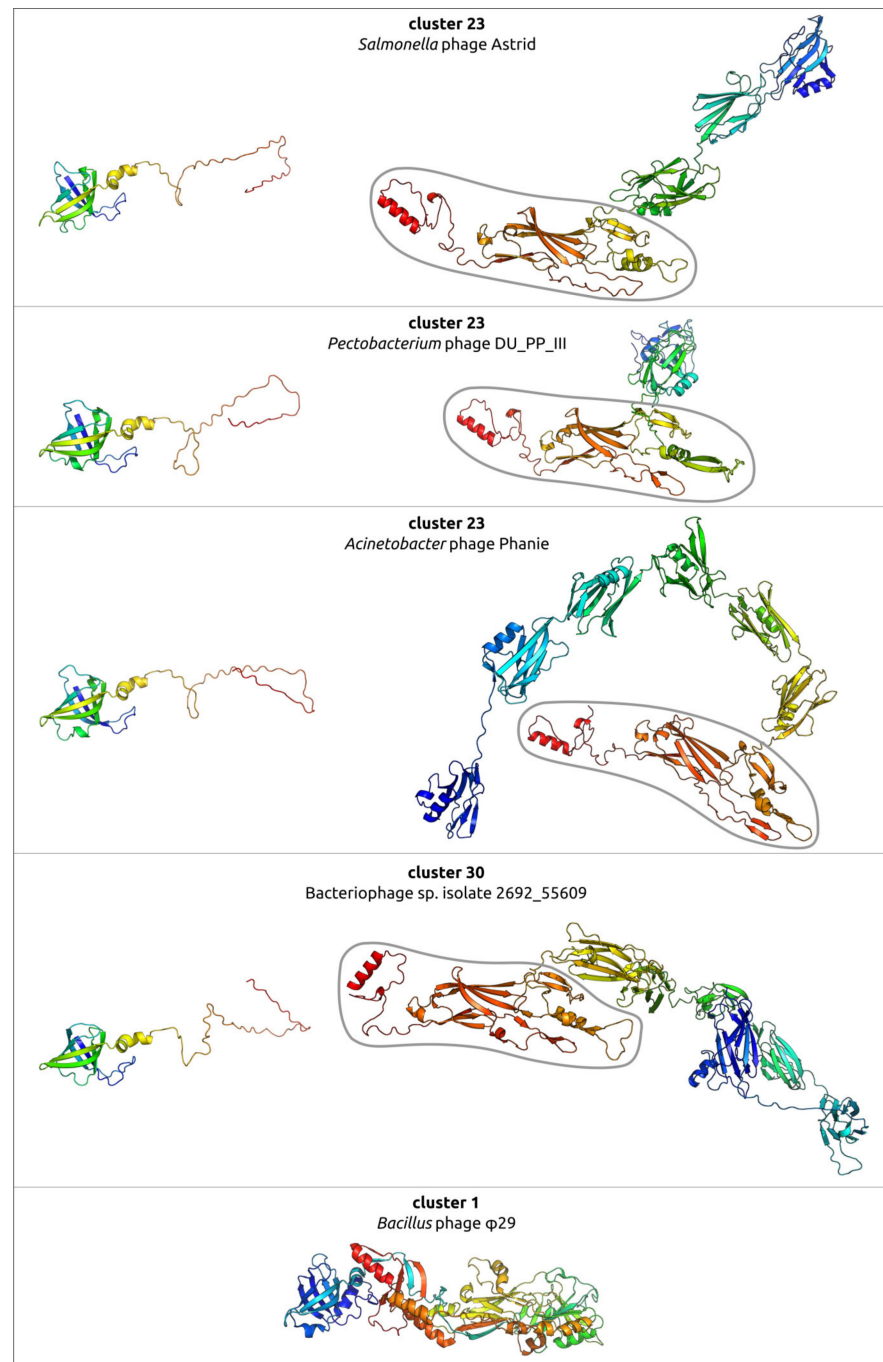


Figure 13. A ribbon diagram of the predicted proteins similar to the N-terminal part of the ϕ 29 tail knob protein and the C-terminal part (major part), coloured using a rainbow gradient scheme, where the N-terminus of the polypeptide chain is coloured blue and the C-terminus is coloured red. The C-terminal fragment of the major part protein, similar to ϕ 29 p9, according to the results of an HHpred search, is contained within a grey lozenge. The image at the bottom of the figure shows the experimentally determined structure of tail knob protein p9 of *Bacillus* phage ϕ 29 (PDB code 5fb5 [49]).

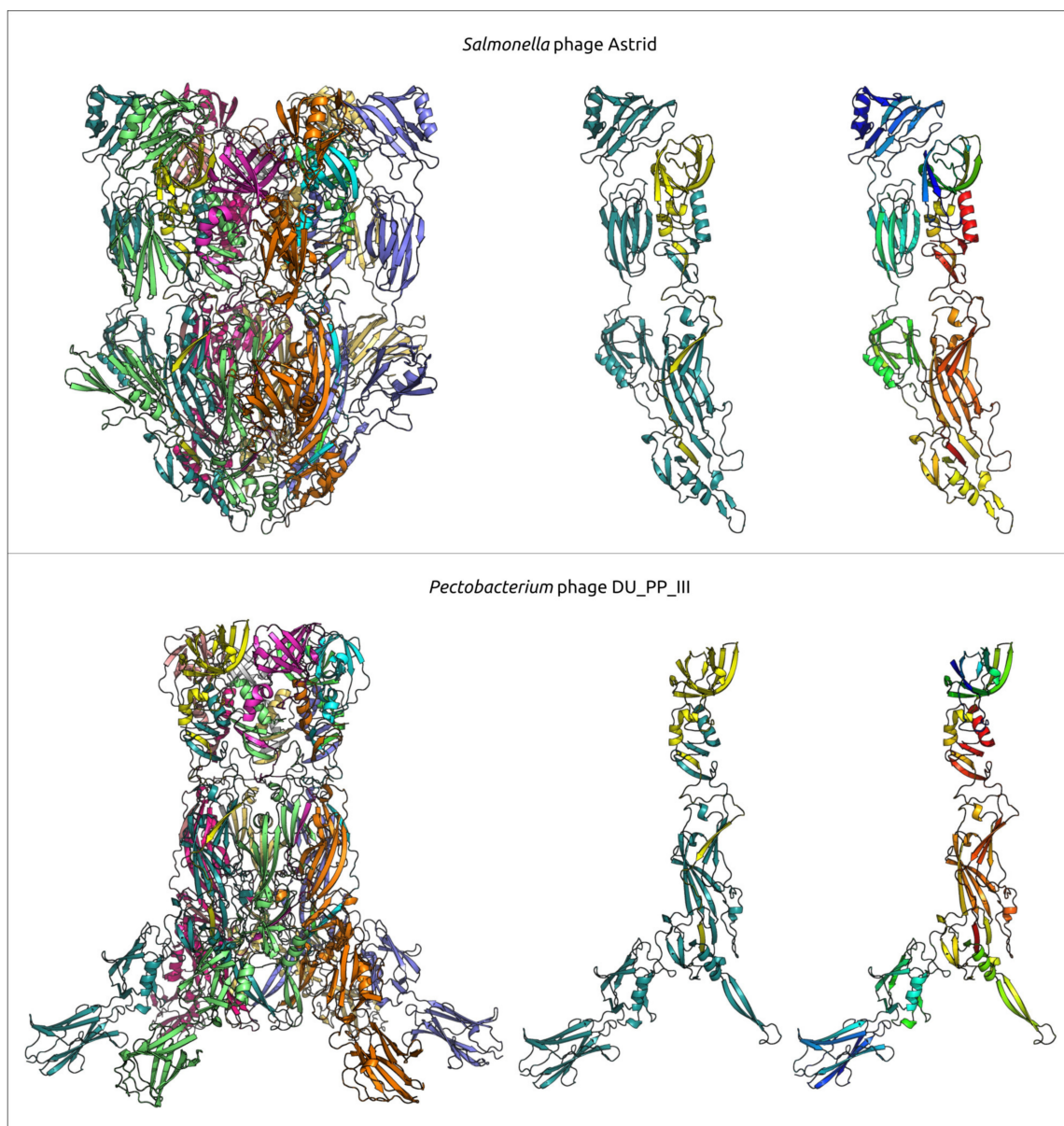


Figure 14. A ribbon diagram of the protein complexes, modelled by AlphaFold3, supposedly corresponding to tail knobs of *Salmonella* phages Astrid and *Pectobacterium* phage DU_PP_III. Left depictions show the full-sized model comprising 12 monomers (each monomer has a different colour). The central depiction shows two closest different monomers, the protein similar to the N-terminal part of the ϕ 29 tail knob protein is coloured yellow, and the protein similar to the C-terminal part is coloured cyan. The right-side depictions show the same monomers as the central one using a rainbow colour gradient, where the N-terminal end of each monomer is blue and the C-terminus is red.

2.7. *Lactococcus* Phage KSY1: General Genomic Features and Relatedness to ϕ 29-like Phages

Lactococcus phage KSY1, representing a singleton genus *Chopinivirus*, is a virulent phage infecting *Lactococcus lactis*. Phage KSY1 possesses a unique complex of morphological and genomic features; the phage was reported in 2007 [117], and no closely related phages have been isolated since then. Phage KSY1 is a podovirus of the C3 morphotype and it has a very characteristic appearance, with an elongated capsid (223 nm long, 45 nm wide) and a short tail (32 nm) [117]. HHblits/HHpred searches have clearly indicated the relatedness of the main virion protein to their counterparts in phage ϕ 29. At the same time, the KSY1 genome contains two genes encoding for phage T7-like DNA-dependent RNA polymerase (RNAP) (Figure 15).

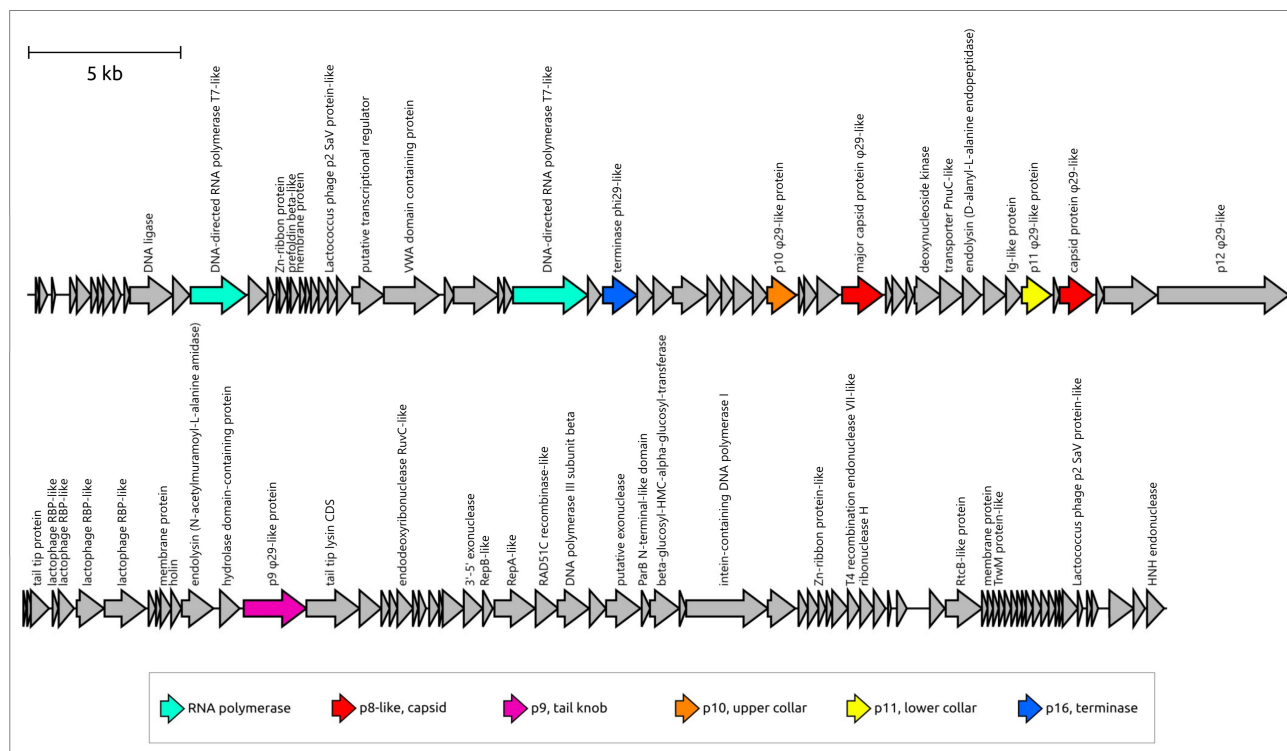


Figure 15. A genetic map of *Lactococcus* phage KSY1. Arrows indicate the direction of transcription. The scale bar indicates the length of the nucleotide sequence. Gene functions are shown in labels and the legend. Genes encoding RNA polymerase and proteins similar to $\phi 29$ proteins p8–p11 and p16 are coloured according to the legend.

According to [117], phage KSY1 has a linear dsDNA genome that is 79,232 bp in length, containing 131 closely packed, predicted open-reading-frames (ORFs) and three tRNA genes. The G + C genome content (35.1%) is close to that of the host (35.3%), indicating the possible long-term coexistence of the phage and its bacterial host. Apparently, the KSY1 genome is blunt-ended and does not contain terminal repeats or redundancy at the genome ends [117].

A BLAST search using the GenBank phage database (GBPHG) indicated that 83 proteins out of 131 have no homologues among phage proteins (E-value < 0.05). The sequences of predicted hydrolases, endolysin gp053a, baseplate proteins, RNA-splicing ligase and proteins with unknown functions have similarities with their counterparts from *Lactococcus* phages, which have a dissimilar main virion, and other proteins. This can indicate that corresponding genes have been acquired via gene exchange. Some other proteins, including endolysin gp073, ribonuclease H and nucleic-acid-processing proteins are more similar to proteins of phages infecting other Gram-positive bacteria.

BLAST and HHpred searches could not find homologues for $\phi 29$ DNAP, but found a homologue of DNA polymerase III β subunit (β -clamp) among predicted proteins. It seems that the replication process of KSY1 is quite different to that of phage $\phi 29$. At the same time, an HHpred search indicated that ATPase domain of KSY1 terminase, upper and lower collar proteins and a tail knob-like protein are similar to corresponding proteins of phage $\phi 29$ (p16, p10, p11 and p9). Interestingly, a BLAST search could not find similarities in amino acid sequences between the main virion proteins and terminase of $\phi 29$ -like phages and phage KSY1, except for the tail knob proteins of *Enterococcus*, *Lactococcus* and other phages belonging to the *Rountreeviridae* family and the tail knob-like protein of phage KSY1. This can indicate a lower rate of evolution of this protein, or it might be the result of gene transfer.

The original paper [117] mentions only one gene of T7-like RNAP predicted in the phage KSY1 genome, but an HHpred search, using the latest databases, indicated that

the KSY1 genome contains two genes encoding T7-like RNAP (KSY1 gp014 and gp033). A BLAST search indicated the clear resemblance of KSY1 gp014 to *Autographiviridae* (T7-like) and *Schitoviridae* (N4-like) phages, (the smallest E-value was 3.59×10^{-7} for T7-like *Synechococcus* phage S-SRP01), but showed only minor traces of similarity between gp033 and proteins from other phages, (the smallest E-value was 3.17×10^{-1} for N4-like phage vB_EamP_Gutmeister infecting *Erwinia*). A BLAST search using the GenBank nr database found homologues of RNAP gp014 in evolutionarily distant groups of cellular organisms and viruses. Phylogenetic analysis placed phage KSY1 gp014 close to a protein from *Lactococcus lactis* and distant from *Autographiviridae* and *Schitoviridae* RNAPs, which, in turn formed distinct clades (Figure 16).

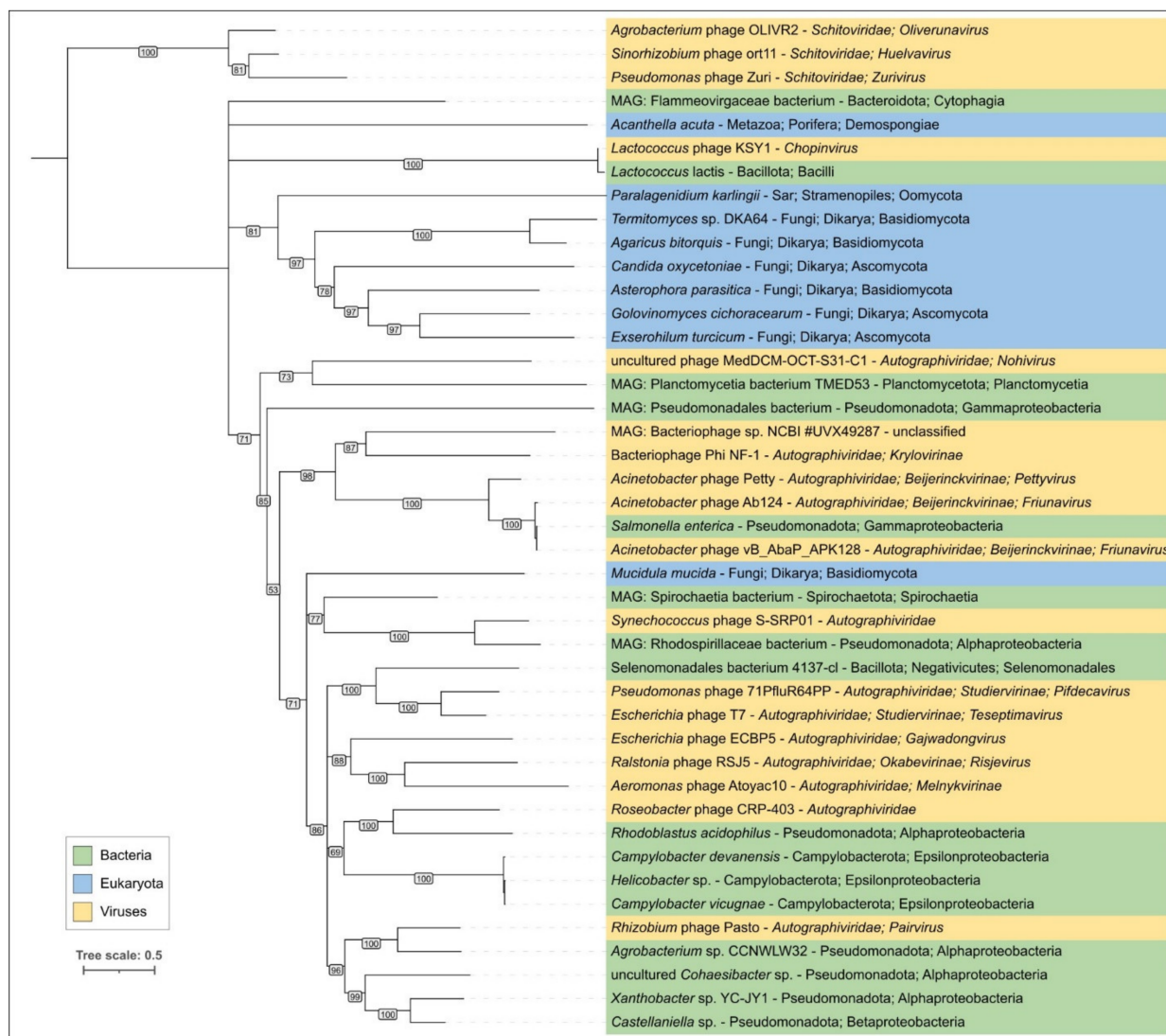


Figure 16. A maximum likelihood phylogenetic tree based on the amino acid sequences of proteins similar to *Lactococcus* phage KSY1 RNA polymerase gp014. NCBI taxonomy is indicated in labels and in the legend. Branches with a bootstrap support lower than 50% have been deleted. Bootstrap values are shown near their branches. The scale bar shows 0.5 estimated substitutions per site and the tree was rooted to the midpoint.

2.8. ViPtree Proteome Phylogeny and VIRIDIC Clustering

A phylogenetic tree based on the proteome similarity was obtained using genomic sequences of representatives of each cluster and *Lactococcus* phage KSY1. The tree was obtained by the ViP server that predicts proteins, calculates the level of similarity between

proteins, and clusters them using a neighbour-joining algorithm. This approach cannot resolve problems associated with the conflicting evolutionary history of individual proteins obtained by genetic exchanges, but, together with other clustering tools, it is recommended by ICTV [6]. The tree (Supplementary Figure S8) places all ϕ 29-like phages except representatives of cluster 02 (e.g., *Staphylococcus* phage Andhra of the *Rountreeviridae* family), cluster 22 (e.g., unclassified phage plasmids *Enterococcus*) and cluster 27 (*Lactococcus* phage asccphi28) in a distinct clade containing only ϕ 29-like phages. Representatives of phages of clusters 02, 22 and 27 were placed in another big clade, while *Lactococcus* phage asccphi28 (cluster 27) was grouped together with *Lactococcus* phage asccphi28, and representatives of clusters 02 and 22 were placed in a branch containing the ϕ 29-like phages infecting *Enterococcus* and *Staphylococcus*. Importantly, the topology of the tree, constructed by the ViPtree, in the part of deep-rooted clades, was different for different sets of phages, and some relationships were inferred from just a few or one protein, e.g., the ViPtree placed *Lactococcus* phage asccphi28 and *Lactococcus* phage KSY1 in a separate clade, based on the similarity of a fragment of endolysin corresponding to the region 48,427–47,858 bp.

Pairwise intergenomic similarity calculations with VIRIDIC (Supplementary Figure S9) indicated low levels of intergenomic similarity between members of different ϕ 29-like phage clusters. These values were 0–5%, which are well below the genus-delimiting threshold of 70%.

3. Discussion

The application of the definition of taxonomic species to the description of evolutionary history of biological objects with chimeric genomes is a difficult task. The limited period of the existence of species and the lability of viral genomes can also make attempts to reconstruct the evolutionary history almost pointless. The processes of genetic exchange seem to be especially pronounced in temperate phages, where the phylogenetic analysis of even the most-conserved proteins (e.g., MCP or terminase) shows a strikingly different topology of the trees [118]. As can be seen from the results of clustering, genomic comparisons and phylogenetic analysis, genetic rearrangements between evolutionarily close groups of phages can also affect ϕ 29-like phages. However, we can try to plot the evolutionary traits of genes encoding separate essential proteins and their stable complexes.

The group of ϕ 29-like phages possesses pronounced unique features in packaging and replication, which are interrelated processes. Phages of this group have a similar architecture of their genomes and similar virion proteins (e.g., capsid and portal proteins). Structural interactions between morphogenesis proteins drives the co-evolution of these proteins and may be one of the reasons for the stability of ϕ 29-like genomes. HMM and BLAST searches using the sequences of the major capsid protein and terminase carried out in this study allowed us to identify relative phages. The results indicate that the set of main virion proteins (p6-p11) is the hallmark feature of ϕ 29-like group, though their sequences may significantly differ from the sequences of other phage proteins. The proteins comprising replication complex and terminase are also characteristic for these phages.

The present research indicates that the evolution of ϕ 29-like viruses was accompanied by genome rearrangements, gene duplication and loss, and the acquisition of domains of apparently cellular origin. The latter applies even to such a conserved protein as the major capsid protein. Notably, a comparison of the structures of the ϕ 29 major capsid protein (excluding the Ig-like domain presumably acquired by horizontal transfer) showed that the ϕ 29 MCP is more similar to its counterpart from *Autographiviridae* phages than to its counterparts from other groups of *Caudoviricetes* phages (Figures 4 and 17, Supplementary Figure S5). This could be an argument for the evolutionary relatedness of ϕ 29-like and T7-like phages. The unlikeness of amino acid sequences of MCP of ϕ 29-like phages to sequences of MCPs of all other *Duplodnaviria* viruses (except phage KSY1), including T7-like phages, however, which has been confirmed by HMM searches, is a strong indication of the ancient divergence of the ϕ 29-like phage group.

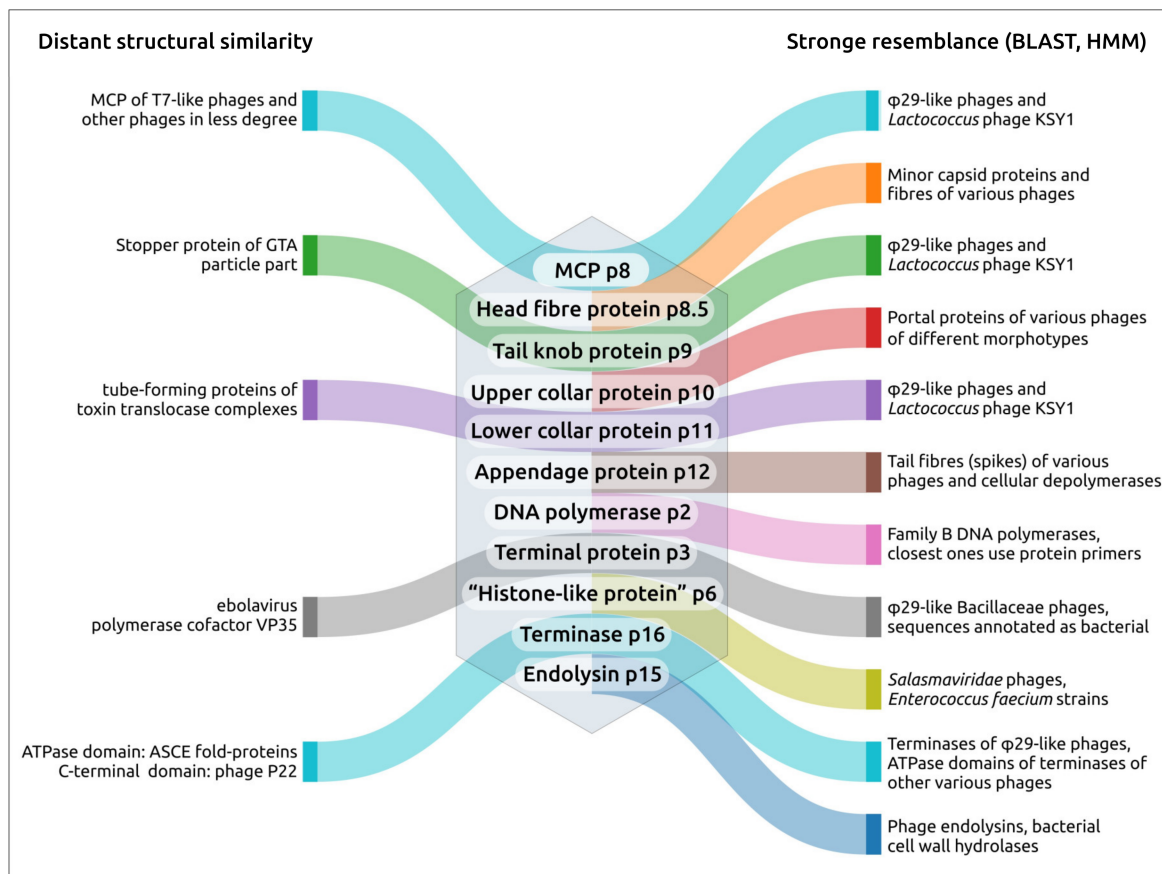


Figure 17. Diagram indicating the close and distant relatedness of $\phi 29$ proteins, based on the conducted analyses and earlier-published data.

The evolution of $\phi 29$ -like phages may involve both coevolution with the host and host switching, based on an analysis and comparisons of phylogenetic trees relating to phage and bacterial sequences. Members of $\phi 29$ -like groups of phages are able to infect a wide range of hosts, and the analysis of metagenomic sequences indicated that, in addition to bacteria of the Bacillales, Clostridia, Enterobacterales and Pseudomonadales orders, these phages might be able to infect representatives of the Bacteroidales order. Perhaps the phage Bacteriophage sp. isolate 2692_55609, representing cluster c30, is the earliest diverged $\phi 29$ -like phage (among the analysed representatives) which infects Gram-negative bacteria. A phylogenetic analysis of MCP and DNAP indicates that this phage is related to *Amedibacillus* phage AD70P2 (cluster c24). *Amedibacillus* bacteria are phylogenetically distant from Gram-negative bacteria. The origin and evolution of $\phi 29$ -like phages infecting Gram-negative bacteria is probably linked to host switching followed by modification of some proteins. Particularly, Gram-negative-specific phages have shortened lower collar (tail tube) protein, which may be due to the thinner cell wall of the bacterial host. Also, the recent publication on *Acinetobacter* phage Phanie [13] and the results of this work suggest the distinct composition of the tail knob in phages infecting Gram-negative bacteria. The tail knob of analysed $\phi 29$ -like phage infecting Gram-negative bacteria appears to be a hexamer, each unit of which consists of two proteins. One of them is located closer to the phage head and corresponds to N-terminal part of $\phi 29$ tail knob protein p9, and the second distal protein is analogous to the carboxy-terminal part of $\phi 29$ p9. This split of the gene encoding the tail knob protein probably occurred in the ancestor of $\phi 29$ -like phages infecting Gram-negative bacteria due to genetic exchange, and resulted in the insertion of the fragment encoding Ig-like domains. It has been shown that Ig-like domains can participate in the interaction with carbohydrates on the surface of the bacterial cell [119] and participate in the formation of phage tail appendages and sheaths [120,121]. The results

of structural predictions suggest that Ig-like domains are located lateral to the tail tube. Particularly, the AF model of the protein complex of a tail knob of *Pectobacterium* phage DU_PP_III, predicted with a decent quality (ipTM score 0.69, pTM score 0.70), depicts the Ig-like domains down to the phage head as resembling short tail fibres, though with architecture fundamentally different from common trimeric architecture. A less reliable model of the tail knob of *Salmonella* phages Astrid also places the Ig-like domains lateral to the tail tube, but adjacent to the tail tube. *Acinetobacter* phage Phanie contains six such domains in gp11. They resemble the Ig-like non-catalytic domain of L,D-transpeptidase from *Mycobacterium tuberculosis*; hypothetically, these domains are able to exhibit carbohydrate-binding properties. The formation of chimeric virions of phage Phanie and helper myovirus *Acinetobacter* phage Aci01-1 was experimentally demonstrated [13]. Based on TEM images [13], myoviral tail tube is attached distally to the tail of phage Phanie. In a previous work [121], it was suggested that the additional Ig-like domains of a myoviral tail sheath protein can facilitate increasing the stability of the sheath due to protein–protein interactions involving additional domains. It is probable that these domains play an important role in the formation of chimeric virions.

Unexpectedly, HMM searches revealed that a peculiar phage, *Lactococcus* phage KSY1 (*Chopinivirus* KSY1) [117], which is very much not like ϕ 29 in genome size and content, has ϕ 29-like terminase and the virion proteins, including homologues of ϕ 29 MCP and non-RBP tail proteins. Perhaps ϕ 29-like phages are just a part of an even larger clade of evolutionarily related bacteriophages. Seemingly, *Lactococcus* phage KSY1 acquired genes of adsorption apparatus from other phages infecting *Lactococcus* or related bacteria, but the source of the genes involved in transcription and replication requires additional research. However, a sequence search has clearly indicated the relatedness of *Chopinivirus*, *Autographiviridae* and *Schitoviridae* RNA polymerases, and phylogenetic analysis has indicated that the RNAPs of all three taxa belong to different clades. It will be interesting to see if the *Lactococcus* phage KSY1 can help find the link between the T7- and ϕ 29-like viruses. In spite of having transcription apparatus resembling T7-like phages [117], the genome of phage KSY1 contains neither T7-like DNA polymerase nor ϕ 29-like DNAP. According to the results of phylogenetic analysis, the latter is related to the DNAP of tectiviruses.

The initial steps of phage infection include the attachment and injection of phage DNA into the bacterial host [122]. Bacterial cells typically contain peptidoglycan, which acts as an obstacle to phage infection; phage ϕ 29 must penetrate an approximately 250 Å thick outer peptidoglycan cell wall [48]. Different phages overcome this obstacle in different ways. Tectiviruses (class *Tectiviricetes*) use a tubular structure made of a viral membrane that is formed after adsorption to the host–cell to deliver the phage genome into the host–cell [123]. Tailed phages (class *Caudoviricetes*) can use their tails, which are made of proteins, as a genome delivery device. *Caudoviricetes* tails can be one of three types, namely short noncontractile (podoviruses), long noncontractile (siphoviruses) and long contractile (myoviruses). A structural analysis of tail proteins reveals the evolutionary relatedness of the long tails of tailed phages, myoviruses and siphoviruses, as well as their common evolutionary origin with secretion systems of type VI (T6SS) and bacteriocins [124–126]. It is difficult to say which functions were performed by the last common ancestor of these nanomachines, and the details of their origin need clarification, but some observations indicate that T6SSs might represent an early diverged branch, whereas bacteriocins could appear multiple times in different organisms and have proteins similar to the tail proteins of phages that infect these organisms [121].

Although some evolutionary relatives of long-tailed phages are known, the details of the evolution of short-tailed phages are less understood. Actually, some domains of phage structural proteins, especially those participating in adsorption and cell-wall degradation, can have a resemblance with proteins or protein domains of cellular origin, as well as with proteins of viral origin, even those encoded in genomes of phages belonging to distant viral groups, but these cases are apparently the result of horizontal transfer [99]. From this point of view, the results of the current study, which indicate the resemblance of the ϕ 29

tail tube formed by lower collar protein p11 and tail tubes of toxin translocase complexes of Gram-positive bacteria, could offer a partial explanation of the origin of the tail of $\phi 29$ -like phages. A search for similar structures using non-RBP tail proteins of $\phi 29$, excluding upper collar (portal) protein, which is common for all tailed phages and herpesviruses, and which could be considered to be part of the tail head (capsid), did not find similar proteins among podoviruses belonging to well-studied groups of T7-like phages (family *Autographiviridae*) and N4-like phages (family *Schitoviridae*). A previous publication [19] indicates that the amino-terminal domain of the tail knob proteins of a $\phi 29$ -like phage C1 infecting *Streptococcus*, and apparently of $\phi 29$ itself, has a fold found in tube-forming proteins of siphoviruses and myoviruses phages, indicating evolutionary ties between long-tailed and short-tailed phages. A similar fold was found in phage and herpesvirus proteases [127], as well as in the tail stopper protein of the GTA of *Rhodobacter capsulatus*, structurally resembling p11 according to DALI (Section 2.2.3). However, this fold seems to be missing the protein structures of another podoviral tail tube, namely of phage T7 (see, for example, PDB #6R21 [128]). Perhaps the morphological classification of podoviruses needs some clarification, since their tails may have different origin and architecture.

Koonin et al. [80] suggested the importance of capturing host genes by viruses during their evolution. This suggestion is consistent with the assumption that the ancestor of tailed phages adapted different bacterial transport systems to deliver its genome into the host-cell. Future myoviruses and siphoviruses could adapt an ancestor of T6SS, and the ancestor(s) of modern $\phi 29$ -like phages could utilise parts of bacterial toxin translocase complexes. It is also possible that, during evolution, phages might change their morphology. The results of recent studies by Pourcel et al. [13], for example, which show the formation of a chimeric virion, may indicate the possibility of the emergence of new chimeric viruses. The first step of such a transformation could be changes in protein structures, which lead to the appearance of stable chimeric virions, with subsequent steps involving genetic exchanges that enable the formation of a chimeric genome. Similarly, one could imagine that the ancestor of the T6SS or pore-forming translocase complex and a tailless phage gave rise to tailed phages. Moreover, hypothetically, viral ancestors of tailed phages might perform a successful penetration and infection function using mainly cellular machinery.

In addition, the cellular origin of phage proteins may help to shed light on the origin of the packaging complex of $\phi 29$ -like phages that contain unusual packaging RNA. The origin of the C-terminal domain of phage $\phi 29$ terminase (CTD) resembling an RNase H nuclease fold may be linked to the loss of the nuclease function of a large subunit of phage terminase that resulted in the formation of the simplified vestigial nuclease fold CTD [59]. The authors also suggest the further simplification of CTD, resulting in an experimentally shown loss of ability to bind the connector and the replacement of this function by pRNA [59]. Alternatively, the authors propose the origin of pRNA based on the RNA-world hypothesis, where RNA was used as both genetic material and functional macromolecules. In this case, pRNA is a relict that may be left conserved in $\phi 29$ -like phages as descendants of ancient phages, while other phages have transferred this function to proteins [59]. We would also propose a third hypothesis, in which pRNA is a vestige left over from a cellular complex adapted by phages for their purposes and subsequently lost. The similarity between the ATPase domain of $\phi 29$ terminase and the DEAD/DEAH-box RNA helicases and other proteins characterised by the ASCE fold may indicate a cellular origin for terminase. HMM and DALI searches have indicated a clear similarity between the $\phi 29$ terminase and RNA-dependent nucleoside triphosphatases, which can employ RNA-stimulated nucleoside triphosphatase activity to remodel RNA or ribonucleoprotein complexes [129]. Cellular proteins with ATPase and nuclease domains are also known, e.g., DNA2 nuclease/helicase [130].

Summarising the aforesaid, the group of $\phi 29$ -like phage clearly differs from other phage groups. It has much in common, though it has variations in genomes, bacterial hosts and infection mechanisms. A search for $\phi 29$ -like and related viruses conducted for the present study could not find many members of this group, since it employed the

HMM-based method on a limited database. The search did, however, identify more than 2000 sequences, indicating the great abundance of these viruses. The prevalence of ϕ 29-related viruses, the complex of unique features, including the characteristic morphology and replication processes, and their distinctive origin, including the early divergence of conserved phage proteins, suggest the possibility of delineating this group as a separate taxon. Perhaps it should be ranked at the level of an order or higher.

4. Materials and Methods

4.1. Data Collection and Database Construction

Phage genomic sequences were downloaded from NCBI GenBank PHG and NCBI Genome databases (<https://www.ncbi.nlm.nih.gov>, accessed on 10 January 2024). Downloaded sequences were checked for the presence of duplicates and small fragments that did not contain viral structural, replication or lysis genes, using the Geneious Prime version 2023.2.1 (Biomatters, Inc., Auckland, New Zealand) tools and a BLAST search [131] of the NCBI nt database. Protein structures were downloaded from the Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB) (<https://www.rcsb.org>, accessed on 27 October 2023). Virion structure was downloaded from the Electron Microscopy Data Bank (EMDB) (<https://www.ebi.ac.uk/emdb/>, accessed on 1 March 2024). BLAST databases were constructed using the “makeblastdb” BLAST command.

4.2. Protein Structure Modelling

Protein structures were predicted using AlphaFold (AF); protein complexes were modelled using AlphaFold version 3 [132] (<https://alphafoldserver.com/>, accessed on 1 September 2024), and other protein structures were predicted with AlphaFold version 2.3.2 [133] using full databases and the command line parameters “--monomer” (for monomeric protein) or “--multimer” (for protein complexes). The best-ranked structures were selected for further study. Protein structures were aligned and visualised using Pymol version 2.5.4 (Schrödinger Inc., New York, NY, USA).

4.3. Functional Annotation

Phage genomes were partially reannotated with the assistance of Glimmer 3.0.2 [134], which was used for open-reading-frame (ORF) detection. The functions of the major capsid protein were predicted using a BLAST homology search and an HHpred search using PDB70_mmcif_2023-06-18, PfamA-v35, UniProt-SwissProt-viral70_3_Nov_2021 and NCBI_Conserved_Domains(CD)_v3.19 databases [135]. The functions of tail proteins' domains were predicted using an HHpred search and a DALI search [136].

4.4. Phylogenetic Analysis, Structural Comparison and VIRIDIC Analysis

Multiple alignments of primary amino acid sequences were obtained with MAFFT version 7.48 [40], with default settings, and using the L-INS-i algorithm. A phylogenetic analysis based on the sequence alignments of viral sequences was performed using IQ-TREE version 2.2.5 [137] and the “--alrt 10000 -B 5000” command line parameters. These parameters assume the use of ModelFinder [138] for finding the best substitution model. The resulting consensus trees with bootstrap support values (10,000 replicas) were visualised using iTOL version 6 [139]. Protein structural similarity was assessed using the DALI Z-score [140]. A heatmap based on the pairwise structural similarity of viral capsid proteins and encapsulin major shell proteins, measured by DALI Z-score, was generated using DALI DaliLite.v5. Intergenomic comparison was performed using VIRIDIC version 1.1 [141], applying the default settings. A VICTOR whole-genome phylogenetic tree was generated by running VICTOR on amino acid data, using the D6 formula [142]. A phylogenetic analysis based on the concatenated alignments of essential bacterial genes was performed using UBCG2 [143]. Bacterial genomic sequences were downloaded from NCBI Genome databases (<https://www.ncbi.nlm.nih.gov>, accessed on 10 January 2024). Genome comparison and genetic maps were made with clinker [144].

4.5. Similarity Clustering Analysis

Based on checked and revised annotations of phage genomes, two separate databases were constructed, using major capsid protein (MCPs) and terminase (Ter) predicted sequences. DIAMOND version 2.1.9 [145] was employed as a sequence aligner for proteins, to obtain sequence identity percentages, after which Gephi version 0.10.1 [146] was applied to construct networks based on the identity measure. The networks were visualised using the Force Atlas 2 layout algorithm (tolerance = 0.6, approximate repulsion = true, approximation = 0.8, scaling = 4.0, stronger gravity = true, gravity = 0.04, prevent overlap = true, normalised edge weights = true) [147] and clustered using the community detection algorithm—Modularity [148] (resolution = 2.0 [149]).

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms251910838/s1>. Ref. [150] is cited in the Supplementary Materials.

Author Contributions: Conceptualisation, P.E. and K.M.; methodology, P.E.; software, P.E., D.G., A.E. and A.K.; validation, P.E. and D.G.; formal analysis, P.E.; investigation, P.E., D.G., A.E. and A.K.; resources, K.M.; data curation, P.E. and D.G.; writing—original draft preparation, P.E. and K.M.; writing—review and editing, K.M.; visualisation, P.E., D.G. and A.E.; supervision, P.E. and K.M.; project administration, P.E. and K.M.; funding acquisition, P.E. and K.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in the study are included in the article/Supplementary Materials; further inquiries can be directed to the corresponding authors.

Acknowledgments: The authors thank Andrey Shadrin and Emma Pilgrimova for the fruitful discussions of plasmids' genomics.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Reilly, B.E. *A Study of the Bacteriophages of Bacillus subtilis and Their Infectious Nucleic Acids*; Case Western Reserve University: Cleveland, OH, USA, 1965.
2. Anderson, D.L.; Hickman, D.D.; Reilly, B.E. Structure of *Bacillus subtilis* Bacteriophage Φ 29 and the Length of Φ 29 Deoxyribonucleic Acid. *J. Bacteriol.* **1966**, *91*, 2081–2089. [[CrossRef](#)] [[PubMed](#)]
3. Anderson, D.L.; Reilly, B.E. Analysis of Bacteriophage Φ 29 Gene Function: Protein Synthesis in Suppressor-Sensitive Mutant Infection of *Bacillus subtilis*. *J. Virol.* **1974**, *13*, 211–221. [[CrossRef](#)] [[PubMed](#)]
4. Reilly, B.E.; Nelson, R.A.; Anderson, D.L. Morphogenesis of Bacteriophage Φ 29 of *Bacillus subtilis*: Mapping and Functional Analysis of the Head Fiber Gene. *J. Virol.* **1977**, *24*, 363–377. [[CrossRef](#)] [[PubMed](#)]
5. Koonin, E.V.; Dolja, V.V.; Krupovic, M.; Varsani, A.; Wolf, Y.I.; Yutin, N.; Zerbini, F.M.; Kuhn, J.H. Global Organization and Proposed Megataxonomy of the Virus World. *Microbiol. Mol. Biol. Rev.* **2020**, *84*, e00061-19. [[CrossRef](#)]
6. Turner, D.; Kropinski, A.M.; Adriaenssens, E.M. A Roadmap for Genome-Based Phage Taxonomy. *Viruses* **2021**, *13*, 506. [[CrossRef](#)]
7. Volozhantsev, N.V.; Oakley, B.B.; Morales, C.A.; Verevkin, V.V.; Bannov, V.A.; Krasilnikova, V.M.; Popova, A.V.; Zhilenkov, E.L.; Garrish, J.K.; Schegg, K.M.; et al. Molecular Characterization of Podoviral Bacteriophages Virulent for *Clostridium Perfringens* and Their Comparison with Members of the Picovirinae. *PLoS ONE* **2012**, *7*, e38283. [[CrossRef](#)]
8. Martín, A.C.; López, R.; García, P. Pneumococcal Bacteriophage Cp-1 Encodes Its Own Protease Essential for Phage Maturation. *J. Virol.* **1998**, *72*, 3491–3494. [[CrossRef](#)]
9. Hawkins, N.C.; Kizziah, J.L.; Hatoum-Aslan, A.; Dokland, T. Structure and Host Specificity of Staphylococcus Epidermidis Bacteriophage Andhra. *Sci. Adv.* **2022**, *8*, eade0459. [[CrossRef](#)] [[PubMed](#)]
10. Tarakanov, R.I.; Lukianova, A.A.; Evseev, P.V.; Pilik, R.I.; Tokmakova, A.D.; Kulikov, E.E.; Toshchakov, S.V.; Ignatov, A.N.; Dzhalilov, F.S.-U.; Miroshnikov, K.A. Ayka, a Novel Curtobacterium Bacteriophage, Provides Protection against Soybean Bacterial Wilt and Tan Spot. *Int. J. Mol. Sci.* **2022**, *23*, 10913. [[CrossRef](#)] [[PubMed](#)]
11. Alanin, K.W.S.; Olsen, N.S.; Djurhuus, A.M.; Carstens, A.B.; Nielsen, T.K.; Rothgardt, M.M.; Russel, A.M.; Wagner, N.; Lametsch, R.; Bak, F.; et al. Four Novel Curtobacterium Phages Isolated from Environmental Samples. *Arch. Virol.* **2023**, *168*, 89. [[CrossRef](#)]

12. Current ICTV Taxonomy Release | ICTV. Available online: <https://ictv.global/taxonomy> (accessed on 9 November 2022).
13. Pourcel, C.; Essoh, C.; Ouldali, M.; Tavares, P. Acinetobacter Baumannii Satellite Phage Aci01-2-Phanie Depends on a Helper Myophage for Its Multiplication. *J. Virol.* **2024**, *98*, e00667-24. [[CrossRef](#)] [[PubMed](#)]
14. Koonin, E.V.; Dolja, V.V. Virus World as an Evolutionary Network of Viruses and Capsidless Selfish Elements. *Microbiol. Mol. Biol. Rev. MMBR* **2014**, *78*, 278–303. [[CrossRef](#)] [[PubMed](#)]
15. Botstein, D. A Theory of Modular Evolution for Bacteriophages. *Ann. N. Y. Acad. Sci.* **1980**, *354*, 484–490. [[CrossRef](#)] [[PubMed](#)]
16. Hendrix, R.W.; Smith, M.C.M.; Burns, R.N.; Ford, M.E.; Hatfull, G.F. Evolutionary Relationships among Diverse Bacteriophages and Prophages: All the World's a Phage. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 2192–2197. [[CrossRef](#)] [[PubMed](#)]
17. Krupovic, M.; Koonin, E.V. Multiple Origins of Viral Capsid Proteins from Cellular Ancestors. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E2401–E2410. [[CrossRef](#)]
18. Evseev, P.; Gutnik, D.; Shneider, M.; Miroshnikov, K. Use of an Integrated Approach Involving AlphaFold Predictions for the Evolutionary Taxonomy of Duplodnaviria Viruses. *Biomolecules* **2023**, *13*, 110. [[CrossRef](#)]
19. Aksyuk, A.A.; Bowman, V.D.; Kaufmann, B.; Fields, C.; Klose, T.; Holdaway, H.A.; Fischetti, V.A.; Rossmann, M.G. Structural Investigations of a Podoviridae Streptococcus Phage C1, Implications for the Mechanism of Viral Entry. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 14001–14006. [[CrossRef](#)]
20. Yoshikawa, H.; Ito, J. Terminal Proteins and Short Inverted Terminal Repeats of the Small Bacillus Bacteriophage Genomes. *Proc. Natl. Acad. Sci. USA* **1981**, *78*, 2596–2600. [[CrossRef](#)]
21. Ito, J. Bacteriophage Phi29 Terminal Protein: Its Association with the 5' Termini of the Phi29 Genome. *J. Virol.* **1978**, *28*, 895–904. [[CrossRef](#)]
22. Guo, P.X.; Erickson, S.; Anderson, D. A Small Viral RNA Is Required for in Vitro Packaging of Bacteriophage Phi 29 DNA. *Science* **1987**, *236*, 690–694. [[CrossRef](#)]
23. Hill, A.C.; Bartley, L.E.; Schroeder, S.J. Prohead RNA: A Noncoding Viral RNA of Novel Structure and Function. *Wiley Interdiscip. Rev. RNA* **2016**, *7*, 428–437. [[CrossRef](#)] [[PubMed](#)]
24. Escarmis, C.; Salas, M. Nucleotide Sequence of the Early Genes 3 and 4 of Bacteriophage Phi 29. *Nucleic Acids Res.* **1982**, *10*, 5785–5798. [[CrossRef](#)] [[PubMed](#)]
25. Kohm, K.; Basu, S.; Nawaz, M.M.; Hertel, R. Chances and Limitations When Uncovering Essential and Non-Essential Genes of *Bacillus subtilis* Phages with CRISPR-Cas9. *Environ. Microbiol. Rep.* **2021**, *13*, 934–944. [[CrossRef](#)] [[PubMed](#)]
26. Meijer, W.J.J.; Horcajadas, J.A.; Salas, M. Φ 29 Family of Phages. *Microbiol. Mol. Biol. Rev.* **2001**, *65*, 261–287. [[CrossRef](#)]
27. Monsalve, M.; Mencía, M.; Rojo, F.; Salas, M. Transcription Regulation in *Bacillus subtilis* Phage Phi 29: Expression of the Viral Promoters throughout the Infection Cycle. *Virology* **1995**, *207*, 23–31. [[CrossRef](#)]
28. Barthelemy, I.; Salas, M.; Mellado, R.P. In Vivo Transcription of Bacteriophage Phi 29 DNA: Transcription Termination. *J. Virol.* **1987**, *61*, 1751–1755. [[CrossRef](#)]
29. Badia, D.; Camacho, A.; Pérez-Lago, L.; Escandón, C.; Salas, M.; Coll, M. The Structure of Phage Φ 29 Transcription Regulator P4-DNA Complex Reveals an N-Hook Motif for DNA Binding. *Mol. Cell* **2006**, *22*, 73–81. [[CrossRef](#)]
30. Monsalve, M.; Mencia, M.; Rojo, F.; Salas, M. Activation and Repression of Transcription at Two Different Phage Phi29 Promoters Are Mediated by Interaction of the Same Residues of Regulatory Protein P4 with RNA Polymerase. *EMBO J.* **1996**, *15*, 383–391. [[CrossRef](#)]
31. Camacho, A.; Salas, M. Molecular Interplay between RNA Polymerase and Two Transcriptional Regulators in Promoter Switch. *J. Mol. Biol.* **2004**, *336*, 357–368. [[CrossRef](#)]
32. Bravo, A.; Illana, B.; Salas, M. Compartmentalization of Phage Φ 29 DNA Replication: Interaction between the Primer Terminal Protein and the Membrane-Associated Protein P1. *EMBO J.* **2000**, *19*, 5575–5584. [[CrossRef](#)]
33. de Jong, R.N.; van der Vliet, P.C.; Brenkman, A.B. Adenovirus DNA Replication: Protein Priming, Jumping Back and the Role of the DNA Binding Protein DBP. *Curr. Top. Microbiol. Immunol.* **2003**, *272*, 187–211. [[CrossRef](#)] [[PubMed](#)]
34. Ravantti, J.J.; Gaidelyte, A.; Bamford, D.H.; Bamford, J.K.H. Comparative Analysis of Bacterial Viruses Bam35, Infecting a Gram-Positive Host, and PRD1, Infecting Gram-Negative Hosts, Demonstrates a Viral Lineage. *Virology* **2003**, *313*, 401–414. [[CrossRef](#)] [[PubMed](#)]
35. Wang, H.; Guo, Z.; Feng, H.; Chen, Y.; Chen, X.; Li, Z.; Hernández-Ascencio, W.; Dai, X.; Zhang, Z.; Zheng, X.; et al. Novel Sulfolobus Virus with an Exceptional Capsid Architecture. *J. Virol.* **2018**, *92*, e01727-17. [[CrossRef](#)] [[PubMed](#)]
36. Bao, K.; Cohen, S.N. Terminal Proteins Essential for the Replication of Linear Plasmids and Chromosomes in Streptomyces. *Genes Dev.* **2001**, *15*, 1518–1527. [[CrossRef](#)] [[PubMed](#)]
37. Klassen, R.; Meinhardt, F. Linear Protein-Primed Replicating Plasmids in Eukaryotic Microbes. In *Microbial Linear Plasmids*; Meinhardt, F., Klassen, R., Eds.; Microbiology Monographs; Springer: Berlin/Heidelberg, Germany, 2007; pp. 187–226. ISBN 978-3-540-72025-6.
38. Mendez, J.; Blanco, L.; Salas, M. Protein-Primed DNA Replication: A Transition between Two Modes of Priming by a Unique DNA Polymerase. *EMBO J.* **1997**, *16*, 2519–2527. [[CrossRef](#)]
39. Calles, B.; Salas, M.; Rojo, F. The Φ 29 Transcriptional Regulator Contacts the Nucleoid Protein P6 to Organize a Repression Complex. *EMBO J.* **2002**, *21*, 6185–6194. [[CrossRef](#)] [[PubMed](#)]

40. Camacho, A.; Salas, M. Mechanism for the Switch of Φ 29 DNA Early to Late Transcription by Regulatory Protein P4 and Histone-like Protein P6. *EMBO J.* **2001**, *20*, 6060–6070. [[CrossRef](#)]
41. González-Huici, V.; Alcorlo, M.; Salas, M.; Hermoso, J.M. Binding of Phage Φ 29 Architectural Protein P6 to the Viral Genome: Evidence for Topological Restriction of the Phage Linear DNA. *Nucleic Acids Res.* **2004**, *32*, 3493–3502. [[CrossRef](#)]
42. Crucitti, P.; Lázaro, J.M.; Benes, V.; Salas, M. Bacteriophage Φ 29 Early Protein P17 Is Conditionally Required for the First Rounds of Viral DNA Replication. *Gene* **1998**, *223*, 135–142. [[CrossRef](#)]
43. Xu, J.; Wang, D.; Gui, M.; Xiang, Y. Structural Assembly of the Tailed Bacteriophage Φ 29. *Nat. Commun.* **2019**, *10*, 2366. [[CrossRef](#)]
44. Cohen, D.N.; Erickson, S.E.; Xiang, Y.; Rossmann, M.G.; Anderson, D.L. Multifunctional Roles of a Bacteriophage Φ 29 Morphogenetic Factor in Assembly and Infection. *J. Mol. Biol.* **2008**, *378*, 802–815. [[CrossRef](#)] [[PubMed](#)]
45. Choi, K.H.; Morais, M.C.; Anderson, D.L.; Rossmann, M.G. Determinants of Bacteriophage Φ 29 Head Morphology. *Structure* **2006**, *14*, 1723–1727. [[CrossRef](#)]
46. Xiang, Y.; Rossmann, M.G. Structure of Bacteriophage Φ 29 Head Fibers Has a Supercoiled Triple Repeating Helix-Turn-Helix Motif. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 4806–4810. [[CrossRef](#)] [[PubMed](#)]
47. Jing, P.; Burris, B.; Cortes, M. The PLB Measurement for the Connector in Φ 29 Bacteriophage Reveals the Function of Its Channel Loop. *Biophys. J.* **2021**, *120*, 1650–1664. [[CrossRef](#)]
48. Xiang, Y.; Morais, M.C.; Cohen, D.N.; Bowman, V.D.; Anderson, D.L.; Rossmann, M.G. Crystal and cryoEM Structural Studies of a Cell Wall Degrading Enzyme in the Bacteriophage Φ 29 Tail. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 9552–9557. [[CrossRef](#)]
49. Xu, J.; Gui, M.; Wang, D.; Xiang, Y. The Bacteriophage Φ 29 Tail Possesses a Pore-Forming Loop for Cell Membrane Penetration. *Nature* **2016**, *534*, 544–547. [[CrossRef](#)]
50. Xiang, Y.; Leiman, P.G.; Li, L.; Grimes, S.; Anderson, D.L.; Rossmann, M.G. Crystallographic Insights into the Autocatalytic Assembly Mechanism of a Bacteriophage Tail Spike. *Mol. Cell* **2009**, *34*, 375–386. [[CrossRef](#)]
51. Bourassa, N.; Major, F. Implication of the Prohead RNA in Phage Φ 29 DNA Packaging. *Biochimie* **2002**, *84*, 945–951. [[CrossRef](#)]
52. Woodson, M.; Pajak, J.; Mahler, B.P.; Zhao, W.; Zhang, W.; Arya, G.; White, M.A.; Jardine, P.J.; Morais, M.C. A Viral Genome Packaging Motor Transitions between Cyclic and Helical Symmetry to Translocate dsDNA. *Sci. Adv.* **2021**, *7*, eabc1955. [[CrossRef](#)]
53. Grimes, S.; Anderson, D. Cleaving the Prohead RNA of Bacteriophage Φ 29 Alters the *in Vitro* Packaging of Restriction Fragments of DNA-Gp3. *J. Mol. Biol.* **1989**, *209*, 101–108. [[CrossRef](#)]
54. Wichitwechkarn, J.; Bailey, S.; Bodley, J.W.; Anderson, D. Prohead RNA of Bacteriophage Φ 29: Size, Stoichiometry and Biological Activity. *Nucleic Acids Res.* **1989**, *17*, 3459–3468. [[CrossRef](#)] [[PubMed](#)]
55. Ismail, A.; Saini, T.; Al Qaffas, A.; Erill, I.; Caruso, S.M.; Temple, L.; Johnson, A.A. Evidence of a Set of Core-Function Genes in 16 *Bacillus* Podoviral Genomes with Considerable Genomic Diversity. *Viruses* **2023**, *15*, 276. [[CrossRef](#)] [[PubMed](#)]
56. Bailey, S.; Wichitwechkarn, J.; Johnson, D.; Reilly, B.E.; Anderson, D.L.; Bodley, J.W. Phylogenetic Analysis and Secondary Structure of the *Bacillus subtilis* Bacteriophage RNA Required for DNA Packaging. *J. Biol. Chem.* **1990**, *265*, 22365–22370. [[CrossRef](#)] [[PubMed](#)]
57. Hilbert, B.J.; Hayes, J.A.; Stone, N.P.; Xu, R.-G.; Kelch, B.A. The Large Terminase DNA Packaging Motor Grips DNA with Its ATPase Domain for Cleavage by the Flexible Nuclease Domain. *Nucleic Acids Res.* **2017**, *45*, 3591–3605. [[CrossRef](#)]
58. Zhao, W.; Jardine, P.J.; Grimes, S. An RNA Domain Imparts Specificity and Selectivity to a Viral DNA Packaging Motor. *J. Virol.* **2015**, *89*, 12457–12466. [[CrossRef](#)]
59. Mahler, B.P.; Bujalowski, P.J.; Mao, H.; Dill, E.A.; Jardine, P.J.; Choi, K.H.; Morais, M.C. NMR Structure of a Vestigial Nuclease Provides Insight into the Evolution of Functional Transitions in Viral dsDNA Packaging Motors. *Nucleic Acids Res.* **2020**, *48*, 11737. [[CrossRef](#)]
60. Helgstrand, C.; Wikoff, W.R.; Duda, R.L.; Hendrix, R.W.; Johnson, J.E.; Liljas, L. The Refined Structure of a Protein Catenane: The HK97 Bacteriophage Capsid at 3.44 Å Resolution. *J. Mol. Biol.* **2003**, *334*, 885–899. [[CrossRef](#)]
61. Wiryaman, T.; Toor, N. Cryo-EM Structure of a Thermostable Bacterial Nanocompartment. *IUCr* **2021**, *8*, 342–350. [[CrossRef](#)]
62. Morais, M.C.; Kanamaru, S.; Badasso, M.O.; Koti, J.S.; Owen, B.A.L.; McMurray, C.T.; Anderson, D.L.; Rossmann, M.G. Bacteriophage Φ 29 Scaffolding Protein Gp7 before and after Prohead Assembly. *Nat. Struct. Biol.* **2003**, *10*, 572–576. [[CrossRef](#)]
63. del Prado, A.; Rodríguez, I.; Lázaro, J.M.; Moreno-Morcillo, M.; de Vega, M.; Salas, M. New Insights into the Coordination between the Polymerization and 3′-5′ Exonuclease Activities in Φ 29 DNA Polymerase. *Sci. Rep.* **2019**, *9*, 923. [[CrossRef](#)]
64. de Vega, M.; Lázaro, J.M.; Mencía, M.; Blanco, L.; Salas, M. Improvement of Φ 29 DNA Polymerase Amplification Performance by Fusion of DNA Binding Motifs. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 16506–16511. [[CrossRef](#)] [[PubMed](#)]
65. Hermoso, J.M.; Salas, M. Protein P3 Is Linked to the DNA of Phage Φ 29 through a Phosphoester Bond between Serine and 5′-dAMP. *Proc. Natl. Acad. Sci. USA* **1980**, *77*, 6425–6428. [[CrossRef](#)] [[PubMed](#)]
66. Ballesteros-Plaza, D.; Holguera, I.; Scheffers, D.-J.; Salas, M.; Muñoz-Espín, D. Phage Φ 29 Protein P1 Promotes Replication by Associating with the FtsZ Ring of the Divisome in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 12313–12318. [[CrossRef](#)] [[PubMed](#)]
67. Martín, G.; Lázaro, J.M.; Méndez, E.; Salas, M. Characterization of the Phage Φ 29 Protein P5 as a Single-Stranded DNA Binding Protein. Function in Φ 29 DNA-Protein P3 Replication. *Nucleic Acids Res.* **1989**, *17*, 3663–3672. [[CrossRef](#)] [[PubMed](#)]

68. Meijer, W.J.J.; Serna-Rico, A.; Salas, M. Characterization of the Bacteriophage Φ 29-Encoded Protein P16.7: A Membrane Protein Involved in Phage DNA Replication. *Mol. Microbiol.* **2001**, *39*, 731–746. [[CrossRef](#)]
69. Asensio, J.L.; Albert, A.; Muñoz-Espín, D.; Gonzalez, C.; Hermoso, J.; Villar, L.; Jiménez-Barbero, J.; Salas, M.; Meijer, W.J.J. Structure of the Functional Domain of Phi29 Replication Organizer: Insights into Oligomerization and Dna Binding. *J. Biol. Chem.* **2005**, *280*, 20730–20739. [[CrossRef](#)] [[PubMed](#)]
70. Albert, A.; Muñoz-Espín, D.; Jiménez, M.; Asensio, J.L.; Hermoso, J.A.; Salas, M.; Meijer, W.J.J. Structural Basis for Membrane Anchorage of Viral Phi29 DNA during Replication. *J. Biol. Chem.* **2005**, *280*, 42486–42488. [[CrossRef](#)] [[PubMed](#)]
71. Young, R. Phage Lysis: Three Steps, Three Choices, One Outcome. *J. Microbiol. Seoul Korea* **2014**, *52*, 243–258. [[CrossRef](#)] [[PubMed](#)]
72. Steiner, M.; Lubitz, W.; Bläsi, U. The Missing Link in Phage Lysis of Gram-Positive Bacteria: Gene 14 of *Bacillus subtilis* Phage Phi 29 Encodes the Functional Homolog of Lambda S Protein. *J. Bacteriol.* **1993**, *175*, 1038–1042. [[CrossRef](#)]
73. Saedi, M.S.; Garvey, K.J.; Ito, J. Cloning and Purification of a Unique Lysozyme Produced by Bacillus Phage Phi 29. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 955–958. [[CrossRef](#)]
74. Hřebík, D.; Štveráková, D.; Škubník, K.; Füzik, T.; Pantůček, R.; Plevka, P. Structure and Genome Ejection Mechanism of Staphylococcus Aureus Phage P68. *Sci. Adv.* **2019**, *5*, eaaw7414. [[CrossRef](#)] [[PubMed](#)]
75. RCSB Protein Data Bank. RCSB PDB-7R3T: Crystal Structure of the Dimeric C-Terminal Big_2-CBM56 Domains from Paenibacillus Illinoisensis (*Bacillus circulans* IAM1165) Beta-1,3-Glucanase H. Available online: <https://www.rcsb.org/structure/7R3T> (accessed on 25 August 2024).
76. Kumar, P.; Vyas, P.; Faisal, S.M.; Chang, Y.-F.; Akif, M. Crystal Structure of a Variable Region Segment of Leptospira Host-Interacting Outer Surface Protein, LigA, Reveals the Orientation of Ig-like Domains. *Int. J. Biol. Macromol.* **2023**, *244*, 125445. [[CrossRef](#)] [[PubMed](#)]
77. Fioravanti, A.; Van Hauwermeiren, F.; Van der Verren, S.E.; Jonckheere, W.; Goncalves, A.; Pardon, E.; Steyaert, J.; De Greve, H.; Lamkanfi, M.; Remaut, H. Structure of S-Layer Protein Sap Reveals a Mechanism for Therapeutic Intervention in Anthrax. *Nat. Microbiol.* **2019**, *4*, 1805–1814. [[CrossRef](#)] [[PubMed](#)]
78. Arnaud, C.-A.; Effantin, G.; Vivès, C.; Engilberge, S.; Bacia, M.; Boulanger, P.; Girard, E.; Schoehn, G.; Breyton, C. Bacteriophage T5 Tail Tube Structure Suggests a Trigger Mechanism for Siphoviridae DNA Ejection. *Nat. Commun.* **2017**, *8*, 1953. [[CrossRef](#)]
79. Nichols, R.J.; LaFrance, B.; Phillips, N.R.; Radford, D.R.; Oltrogge, L.M.; Valentin-Alvarado, L.E.; Bischoff, A.J.; Nogales, E.; Savage, D.F. Discovery and Characterization of a Novel Family of Prokaryotic Nanocompartments Involved in Sulfur Metabolism. *eLife* **2021**, *10*, e59288. [[CrossRef](#)] [[PubMed](#)]
80. Koonin, E.V.; Dolja, V.V.; Krupovic, M. The Logic of Virus Evolution. *Cell Host Microbe* **2022**, *30*, 917–929. [[CrossRef](#)]
81. Clancy Kelley, L.-L.; Dillard, B.D.; Tempel, W.; Chen, L.; Shaw, N.; Lee, D.; Newton, M.G.; Sugar, F.J.; Jenney, F.E.; Lee, H.S.; et al. Structure of the Hypothetical Protein PF0899 from Pyrococcus Furiosus at 1.85 Å Resolution. *Acta Crystallogr. Sect. F* **2007**, *63*, 549–552. [[CrossRef](#)]
82. Hardy, J.M.; Dunstan, R.A.; Grinter, R.; Belousoff, M.J.; Wang, J.; Pickard, D.; Venugopal, H.; Dougan, G.; Lithgow, T.; Coulibaly, F. The Architecture and Stabilisation of Flagellotropic Tailed Bacteriophages. *Nat. Commun.* **2020**, *11*, 3748. [[CrossRef](#)]
83. Wang, Z.; Hardies, S.C.; Fokine, A.; Klose, T.; Jiang, W.; Cho, B.C.; Rossmann, M.G. Structure of the Marine Siphovirus TW1: Evolution of Capsid-Stabilizing Proteins and Tail Spikes. *Structure* **2018**, *26*, 238–248.e3. [[CrossRef](#)]
84. Bayfield, O.W.; Shkoporov, A.N.; Yutin, N.; Khokhlova, E.V.; Smith, J.L.R.; Hawkins, D.E.D.P.; Koonin, E.V.; Hill, C.; Antson, A.A. Structural Atlas of a Human Gut Crassvirus. *Nature* **2023**, *617*, 409–416. [[CrossRef](#)]
85. Javed, A.; Villanueva, H.; Shataer, S.; Vasciaveo, S.; Savva, R.; Orlova, E.V. Cryo-EM Structures of Two Bacteriophage Portal Proteins Provide Insights for Antimicrobial Phage Engineering. *Viruses* **2021**, *13*, 2532. [[CrossRef](#)] [[PubMed](#)]
86. Fàbrega-Ferrer, M.; Cuervo, A.; Fernández, F.J.; Machón, C.; Pérez-Luque, R.; Pous, J.; Vega, M.C.; Carrascosa, J.L.; Coll, M. Using a Partial Atomic Model from Medium-Resolution Cryo-EM to Solve a Large Crystal Structure. *Acta Crystallogr. Sect. Struct. Biol.* **2021**, *77*, 11–18. [[CrossRef](#)] [[PubMed](#)]
87. Xiao, H.; Tan, L.; Tan, Z.; Zhang, Y.; Chen, W.; Li, X.; Song, J.; Cheng, L.; Liu, H. Structure of the Siphophage Neck–Tail Complex Suggests That Conserved Tail Tip Proteins Facilitate Receptor Binding and Tail Assembly. *PLOS Biol.* **2023**, *21*, e3002441. [[CrossRef](#)] [[PubMed](#)]
88. Sun, L.; Zhang, X.; Gao, S.; Rao, P.A.; Padilla-Sanchez, V.; Chen, Z.; Sun, S.; Xiang, Y.; Subramaniam, S.; Rao, V.B.; et al. Cryo-EM Structure of the Bacteriophage T4 Portal Protein Assembly at near-Atomic Resolution. *Nat. Commun.* **2015**, *6*, 7548. [[CrossRef](#)]
89. Bárdy, P.; Füzik, T.; Hřebík, D.; Pantůček, R.; Thomas Beatty, J.; Plevka, P. Structure and Mechanism of DNA Delivery of a Gene Transfer Agent. *Nat. Commun.* **2020**, *11*, 3034. [[CrossRef](#)]
90. Hussain, T.; Llácer, J.L.; Wimberly, B.T.; Kieft, J.S.; Ramakrishnan, V. Large-Scale Movements of IF3 and tRNA during Bacterial Translation Initiation. *Cell* **2016**, *167*, 133–144.e13. [[CrossRef](#)]
91. Bubunencko, M.; Korepanov, A.; Court, D.L.; Jagannathan, I.; Dickinson, D.; Chaudhuri, B.R.; Garber, M.B.; Culver, G.M. 30S Ribosomal Subunits Can Be Assembled in Vivo without Primary Binding Ribosomal Protein S15. *RNA* **2006**, *12*, 1229–1239. [[CrossRef](#)]
92. Machen, A.J.; Fisher, M.T.; Freudenthal, B.D. Anthrax Toxin Translocation Complex Reveals Insight into the Lethal Factor Unfolding and Refolding Mechanism. *Sci. Rep.* **2021**, *11*, 13038. [[CrossRef](#)]

93. Xu, X.; Godoy-Ruiz, R.; Adipietro, K.A.; Peralta, C.; Ben-Hail, D.; Varney, K.M.; Cook, M.E.; Roth, B.M.; Wilder, P.T.; Cleveland, T.; et al. Structure of the Cell-Binding Component of the Clostridium Difficile Binary Toxin Reveals a Di-Heptamer Macromolecular Assembly. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 1049–1058. [\[CrossRef\]](#)
94. RCSB Protein Data Bank. RCSB PDB-7AL8: Structure of ATSI with Bovine Trypsin. Available online: <https://www.rcsb.org/structure/7AL8> (accessed on 25 August 2024).
95. Ge, X.; Wang, J. Structural Mechanism of Bacteriophage Lambda Tail's Interaction with the Bacterial Receptor. *Nat. Commun.* **2024**, *15*, 4185. [\[CrossRef\]](#)
96. Müller, J.J.; Barbirz, S.; Heinle, K.; Freiberg, A.; Seckler, R.; Heinemann, U. An Intersubunit Active Site between Supercoiled Parallel Beta Helices in the Trimeric Tailspike Endorhamnosidase of Shigella Flexneri Phage Sf6. *Structure* **2008**, *16*, 766–775. [\[CrossRef\]](#)
97. Greenfield, J.; Shang, X.; Luo, H.; Zhou, Y.; Linden, S.B.; Heselpoth, R.D.; Leiman, P.G.; Nelson, D.C.; Herzberg, O. Structure and Function of Bacteriophage CBA120 ORF211 (TSP2), the Determinant of Phage Specificity towards *E. coli* O157:H7. *Sci. Rep.* **2020**, *10*, 15402. [\[CrossRef\]](#) [\[PubMed\]](#)
98. d'Acapito, A.; Roret, T.; Zarkadas, E.; Mocaër, P.-Y.; Lechat, F.; Baudoux, A.-C.; Schoehn, G.; Neumann, E. Structural Study of the Cobetia Marina Bacteriophage 1 (Carin-1) by Cryo-EM. *J. Virol.* **2023**, *97*, e00248-23. [\[CrossRef\]](#) [\[PubMed\]](#)
99. Pas, C.; Latka, A.; Fieseler, L.; Briers, Y. Phage Tailspike Modularity and Horizontal Gene Transfer Reveals Specificity towards *E. Coli* O-Antigen Serogroups. *Virol. J.* **2023**, *20*, 174. [\[CrossRef\]](#) [\[PubMed\]](#)
100. Kamtekar, S.; Berman, A.J.; Wang, J.; Lázaro, J.M.; de Vega, M.; Blanco, L.; Salas, M.; Steitz, T.A. Insights into Strand Displacement and Processivity from the Crystal Structure of the Protein-Primed DNA Polymerase of Bacteriophage Phi29. *Mol. Cell* **2004**, *16*, 609–618. [\[CrossRef\]](#) [\[PubMed\]](#)
101. Berman, A.J.; Kamtekar, S.; Goodman, J.L.; Lázaro, J.M.; de Vega, M.; Blanco, L.; Salas, M.; Steitz, T.A. Structures of Phi29 DNA Polymerase Complexed with Substrate: The Mechanism of Translocation in B-Family Polymerases. *EMBO J.* **2007**, *26*, 3494–3505. [\[CrossRef\]](#)
102. Coloma, J.; Johnson, R.E.; Prakash, L.; Prakash, S.; Aggarwal, A.K. Human DNA Polymerase α in Binary Complex with a DNA:DNA Template-Primer. *Sci. Rep.* **2016**, *6*, 23784. [\[CrossRef\]](#)
103. Reid, S.P.; Cárdenas, W.B.; Basler, C.F. Homo-Oligomerization Facilitates the Interferon-Antagonist Activity of the Ebolavirus VP35 Protein. *Virology* **2005**, *341*, 179–189. [\[CrossRef\]](#)
104. Bruhn, J.F.; Kirchdoerfer, R.N.; Urata, S.M.; Li, S.; Tickle, I.J.; Bricogne, G.; Saphire, E.O. Crystal Structure of the Marburg Virus VP35 Oligomerization Domain. *J. Virol.* **2017**, *91*, e01085-16. [\[CrossRef\]](#)
105. Dilley, K.A.; Voorhies, A.A.; Luthra, P.; Puri, V.; Stockwell, T.B.; Lorenzi, H.; Basler, C.F.; Shabman, R.S. The Ebola Virus VP35 Protein Binds Viral Immunostimulatory and Host RNAs Identified through Deep Sequencing. *PLoS ONE* **2017**, *12*, e0178717. [\[CrossRef\]](#)
106. Zhao, H.; Christensen, T.E.; Kamau, Y.N.; Tang, L. Structures of the Phage Sf6 Large Terminase Provide New Insights into DNA Translocation and Cleavage. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 8075–8080. [\[CrossRef\]](#) [\[PubMed\]](#)
107. Saeed, A.F.U.H.; Chan, C.; Guan, H.; Gong, B.; Guo, P.; Cheng, X.; Ouyang, S. Structural Insights into Gp16 ATPase in the Bacteriophage Φ 29 DNA Packaging Motor. *Biochemistry* **2021**, *60*, 886–897. [\[CrossRef\]](#) [\[PubMed\]](#)
108. Roy, A.; Cingolani, G. Structure of P22 Headful Packaging Nuclease. *J. Biol. Chem.* **2012**, *287*, 28196–28205. [\[CrossRef\]](#) [\[PubMed\]](#)
109. Cernooka, E.; Rumnieks, J.; Zrelavs, N.; Tars, K.; Kazaks, A. Diversity of the Lysozyme Fold: Structure of the Catalytic Domain from an Unusual Endolysin Encoded by Phage Enc34. *Sci. Rep.* **2022**, *12*, 5005. [\[CrossRef\]](#) [\[PubMed\]](#)
110. Sykilinda, N.N.; Nikolaeva, A.Y.; Shneider, M.M.; Mishkin, D.V.; Patutin, A.A.; Popov, V.O.; Boyko, K.M.; Klyachko, N.L.; Miroshnikov, K.A. Structure of an Acinetobacter Broad-Range Prophage Endolysin Reveals a C-Terminal α -Helix with the Proposed Role in Activity against Live Bacterial Cells. *Viruses* **2018**, *10*, 309. [\[CrossRef\]](#)
111. Randich, A.M.; Kysela, D.T.; Morlot, C.; Brun, Y.V. Origin of a Core Bacterial Gene via Co-Option and Detoxification of a Phage Lysin. *Curr. Biol. CB* **2019**, *29*, 1634–1646.e6. [\[CrossRef\]](#)
112. Garvey, K.J.; Saedi, M.S.; Ito, J. Nucleotide Sequence of Bacillus Phage Phi 29 Genes 14 and 15: Homology of Gene 15 with Other Phage Lysozymes. *Nucleic Acids Res.* **1986**, *14*, 10001–10008. [\[CrossRef\]](#)
113. Pilgrimova, E.G.; Kazantseva, O.A.; Kazantsev, A.N.; Nikulin, N.A.; Skorynina, A.V.; Koposova, O.N.; Shadrin, A.M. Putative Plasmid Prophages of Bacillus Cereus Ssensu Lato May Hold the Key to Undiscovered Phage Diversity. *Sci. Rep.* **2021**, *11*, 7611. [\[CrossRef\]](#)
114. Olsen, N.S.; Hendriksen, N.B.; Hansen, L.H.; Kot, W. A New High-Throughput Screening Method for Phages: Enabling Crude Isolation and Fast Identification of Diverse Phages with Therapeutic Potential. *Phage* **2020**, *1*, 137–148. [\[CrossRef\]](#)
115. Miroshnikov, K.A.; Evseev, P.V.; Lukianova, A.A.; Ignatov, A.N. Tailed Lytic Bacteriophages of Soft Rot Pectobacteriaceae. *Microorganisms* **2021**, *9*, 1819. [\[CrossRef\]](#)
116. Brammer Basta, L.A.; Ghosh, A.; Pan, Y.; Jakoncic, J.; Lloyd, E.P.; Townsend, C.A.; Lamichhane, G.; Bianchet, M.A. Loss of a Functionally and Structurally Distinct Ld-Transpeptidase, LdtMt5, Compromises Cell Wall Integrity in *Mycobacterium tuberculosis*. *J. Biol. Chem.* **2015**, *290*, 25670–25685. [\[CrossRef\]](#) [\[PubMed\]](#)
117. Chopin, A.; Deveau, H.; Ehrlich, S.D.; Moineau, S.; Chopin, M.-C. KSY1, a Lactococcal Phage with a T7-like Transcription. *Virology* **2007**, *365*, 1–9. [\[CrossRef\]](#) [\[PubMed\]](#)

118. Evseev, P.; Lukianova, A.; Sykilinda, N.; Gorshkova, A.; Bondar, A.; Shneider, M.; Kabilov, M.; Drucker, V.; Miroshnikov, K. Pseudomonas Phage MD8: Genetic Mosaicism and Challenges of Taxonomic Classification of Lambdoid Bacteriophages. *Int. J. Mol. Sci.* **2021**, *22*, 10350. [\[CrossRef\]](#) [\[PubMed\]](#)
119. Fraser, J.S.; Yu, Z.; Maxwell, K.L.; Davidson, A.R. Ig-Like Domains on Bacteriophages: A Tale of Promiscuity and Deceit. *J. Mol. Biol.* **2006**, *359*, 496–507. [\[CrossRef\]](#) [\[PubMed\]](#)
120. Yuan, Y.; Gao, M. Jumbo Bacteriophages: An Overview. *Front. Microbiol.* **2017**, *8*, 403. [\[CrossRef\]](#) [\[PubMed\]](#)
121. Evseev, P.; Shneider, M.; Miroshnikov, K. Evolution of Phage Tail Sheath Protein. *Viruses* **2022**, *14*, 1148. [\[CrossRef\]](#)
122. Sulakvelidze, A.; Alavidze, Z.; Morris, J.G. Bacteriophage Therapy. *Antimicrob. Agents Chemother.* **2001**, *45*, 649–659. [\[CrossRef\]](#)
123. Peralta, B.; Gil-Carton, D.; Castaño-Díez, D.; Bertin, A.; Boulogne, C.; Oksanen, H.M.; Bamford, D.H.; Abrescia, N.G.A. Mechanism of Membranous Tunnelling Nanotube Formation in Viral Genome Delivery. *PLoS Biol.* **2013**, *11*, e1001667. [\[CrossRef\]](#)
124. Leiman, P.G.; Basler, M.; Ramagopal, U.A.; Bonanno, J.B.; Sauder, J.M.; Pukatzki, S.; Burley, S.K.; Almo, S.C.; Mekalanos, J.J. Type VI Secretion Apparatus and Phage Tail-Associated Protein Complexes Share a Common Evolutionary Origin. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 4154–4159. [\[CrossRef\]](#)
125. Veessler, D.; Cambillau, C. A Common Evolutionary Origin for Tailed-Bacteriophage Functional Modules and Bacterial Machineries. *Microbiol. Mol. Biol. Rev. MMBR* **2011**, *75*, 423–433. [\[CrossRef\]](#)
126. Zoued, A.; Brunet, Y.R.; Durand, E.; Aschtgen, M.-S.; Logger, L.; Douzi, B.; Journet, L.; Cambillau, C.; Cascales, E. Architecture and Assembly of the Type VI Secretion System. *Biochim. Biophys. Acta* **2014**, *1843*, 1664–1673. [\[CrossRef\]](#) [\[PubMed\]](#)
127. Fokine, A.; Rossmann, M.G. Common Evolutionary Origin of Procapsid Proteases, Phage Tail Tubes, and Tubes of Bacterial Type VI Secretion Systems. *Structure* **2016**, *24*, 1928–1935. [\[CrossRef\]](#) [\[PubMed\]](#)
128. Cuervo, A.; Fàbrega-Ferrer, M.; Machón, C.; Conesa, J.J.; Fernández, F.J.; Pérez-Luque, R.; Pérez-Ruiz, M.; Pous, J.; Vega, M.C.; Carrascosa, J.L.; et al. Structures of T7 Bacteriophage Portal and Tail Suggest a Viral DNA Retention and Ejection Mechanism. *Nat. Commun.* **2019**, *10*, 3746. [\[CrossRef\]](#) [\[PubMed\]](#)
129. Pietrzyk-Brzezinska, A.J.; Absmeier, E.; Klauck, E.; Wen, Y.; Antelmann, H.; Wahl, M.C. Crystal Structure of the Escherichia Coli DEXH-Box NTPase HrpB. *Structure* **2018**, *26*, 1462–1473.e4. [\[CrossRef\]](#) [\[PubMed\]](#)
130. Zheng, L.; Meng, Y.; Campbell, J.L.; Shen, B. Multiple Roles of DNA2 Nuclease/Helicase in DNA Metabolism, Genome Stability and Human Diseases. *Nucleic Acids Res.* **2020**, *48*, 16–35. [\[CrossRef\]](#)
131. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic Local Alignment Search Tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [\[CrossRef\]](#)
132. Abramson, J.; Adler, J.; Dunger, J.; Evans, R.; Green, T.; Pritzel, A.; Ronneberger, O.; Willmore, L.; Ballard, A.J.; Bambrick, J.; et al. Accurate Structure Prediction of Biomolecular Interactions with AlphaFold 3. *Nature* **2024**, *630*, 493–500. [\[CrossRef\]](#)
133. AlQuraishi, M. AlphaFold at CASP13. *Bioinforma. Oxf. Engl.* **2019**, *35*, 4862–4865. [\[CrossRef\]](#) [\[PubMed\]](#)
134. Delcher, A.L.; Bratke, K.A.; Powers, E.C.; Salzberg, S.L. Identifying Bacterial Genes and Endosymbiont DNA with Glimmer. *Bioinforma. Oxf. Engl.* **2007**, *23*, 673–679. [\[CrossRef\]](#) [\[PubMed\]](#)
135. Zimmermann, L.; Stephens, A.; Nam, S.-Z.; Rau, D.; Kübler, J.; Lozajic, M.; Gabler, F.; Söding, J.; Lupas, A.N.; Alva, V. A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at Its Core. *J. Mol. Biol.* **2018**, *430*, 2237–2243. [\[CrossRef\]](#)
136. Holm, L. Dali Server: Structural Unification of Protein Families. *Nucleic Acids Res.* **2022**, *50*, W210–W215. [\[CrossRef\]](#) [\[PubMed\]](#)
137. Nguyen, L.-T.; Schmidt, H.A.; von Haeseler, A.; Minh, B.Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* **2015**, *32*, 268–274. [\[CrossRef\]](#)
138. Kalyaanamoorthy, S.; Minh, B.Q.; Wong, T.K.F.; von Haeseler, A.; Jermini, L.S. ModelFinder: Fast Model Selection for Accurate Phylogenetic Estimates. *Nat. Methods* **2017**, *14*, 587–589. [\[CrossRef\]](#)
139. Letunic, I.; Bork, P. Interactive Tree of Life (iTOL) v5: An Online Tool for Phylogenetic Tree Display and Annotation. *Nucleic Acids Res.* **2021**, *49*, W293–W296. [\[CrossRef\]](#)
140. Holm, L.; Kääriäinen, S.; Rosenström, P.; Schenkel, A. Searching Protein Structure Databases with DaliLite v.3. *Bioinformatics* **2008**, *24*, 2780–2781. [\[CrossRef\]](#)
141. Moraru, C.; Varsani, A.; Kropinski, A.M. VIRIDIC—A Novel Tool to Calculate the Intergenomic Similarities of Prokaryote-Infecting Viruses. *Viruses* **2020**, *12*, 1268. [\[CrossRef\]](#)
142. Meier-Kolthoff, J.P.; Göker, M. VICTOR: Genome-Based Phylogeny and Classification of Prokaryotic Viruses. *Bioinformatics* **2017**, *33*, 3396–3404. [\[CrossRef\]](#)
143. Kim, J.; Na, S.-I.; Kim, D.; Chun, J. UBCG2: Up-to-Date Bacterial Core Genes and Pipeline for Phylogenomic Analysis. *J. Microbiol. Seoul Korea* **2021**, *59*, 609–615. [\[CrossRef\]](#)
144. Gilchrist, C.L.M.; Chooi, Y.-H. Clinker & Clustermap.js: Automatic Generation of Gene Cluster Comparison Figures. *Bioinformatics* **2021**, *37*, 2473–2475. [\[CrossRef\]](#)
145. Buchfink, B.; Reuter, K.; Drost, H.-G. Sensitive Protein Alignments at Tree-of-Life Scale Using DIAMOND. *Nat. Methods* **2021**, *18*, 366–368. [\[CrossRef\]](#)
146. Bastian, M.; Heymann, S.; Jacomy, M. Gephi: An Open Source Software for Exploring and Manipulating Networks. *Proc. Int. AAAI Conf. Web Soc. Media* **2009**, *3*, 361–362. [\[CrossRef\]](#)
147. Jacomy, M.; Venturini, T.; Heymann, S.; Bastian, M. ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software. *PLoS ONE* **2014**, *9*, e98679. [\[CrossRef\]](#) [\[PubMed\]](#)

148. Blondel, V.D.; Guillaume, J.-L.; Lambiotte, R.; Lefebvre, E. Fast Unfolding of Communities in Large Networks. *J. Stat. Mech. Theory Exp.* **2008**, *2008*, P10008. [[CrossRef](#)]
149. Lambiotte, R. Multi-Scale Modularity and Dynamics in Complex Networks. In *Dynamics On and Of Complex Networks*; Mukherjee, A., Choudhury, M., Peruani, F., Ganguly, N., Mitra, B., Eds.; Modeling and Simulation in Science, Engineering and Technology; Springer: New York, NY, USA, 2013; Volume 2, pp. 125–141. ISBN 978-1-4614-6728-1.
150. Ding, F.; Lu, C.; Zhao, W.; Rajashankar, K.R.; Anderson, D.L.; Jardine, P.J.; Grimes, S.; Ke, A. Structure and Assembly of the Essential RNA Ring Component of a Viral DNA Packaging Motor. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 7357–7362. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.