

Figure S1. (a) Terminase (all monomers are coloured grey) and pRNA (each monomer has a different colour) in the ϕ 29 packaging motor; ribbon and surface-rendered representations, including top view. Phage DNA is coloured orange. The image was made using PDB structure 7jqj [52]. (b) pRNA secondary structure, according to Ding et al. (2011) [150], and tertiary structure, based on PDB structure 7jqj [52].

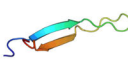
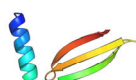
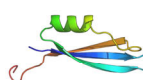
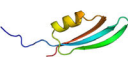
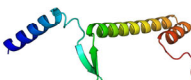
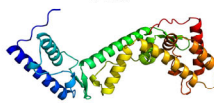
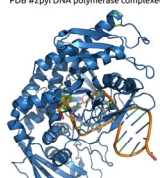
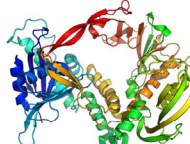
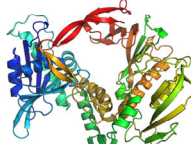
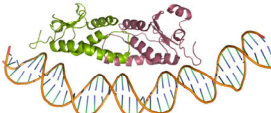


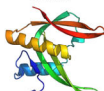
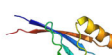
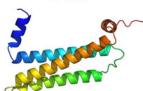

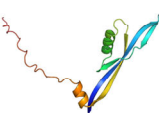


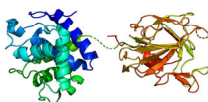
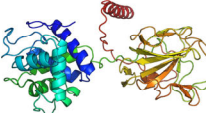
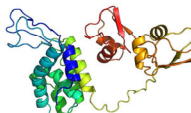
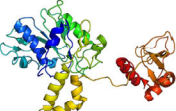
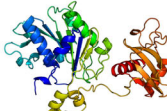

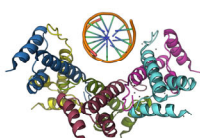
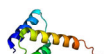

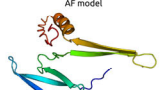
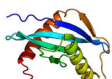
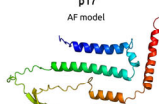
<p>p0.2 AF model</p> 	<p>p0.4 AF model</p> 	<p>p0.6 AF model</p> 
<p>p0.8 AF model</p> 	<p>p1 AF model</p> 	<p>p3 AF model</p> 
<p>p2 DNA polymerase</p> <p>PDB #2pyl DNA polymerase complexed with primer-template DNA and incoming nucleotide substrates</p>  <p>ternary complex</p>  <p>DNA polymerase monomer</p>  <p>AF model</p>		
<p>p4 transcription regulator</p> <p>PDB #2fio transcription regulator p4-DNA complex</p>  <p>ternary complex</p>  <p>DNA polymerase monomer</p>  <p>AF model</p>		
<p>p5 AF model</p> 	<p>p5.5 AF model</p> 	<p>p14 holin AF model</p> 
<p>p6 DNA-binding protein</p> <p>PDB #8pw4, p6 C-terminal delta20 truncated version</p>  <p>AF model</p> 		<p>p7 scaffolding protein</p> <p>PDB #9fzm, p7 scaffolding protein bound to intermediate-state MCP</p>  <p>AF model</p> 
<p>p13 tail lysin</p> <p>PDB #3cqs, p13 tail cell wall degrading enzyme</p>  <p>AF model</p> 		<p>p15 endolysin</p> <p>AF model</p> 
<p>p16 terminase</p> <p>PDB #7jqs, genome packaging motor assembly</p>  <p>AF model</p> 		<p>p16.6</p> <p>AF model</p> 
<p>p16.7 replication organiser protein</p> <p>PDB #2c5r replication organiser protein p16.7 in complex with dsDNA</p>  <p>ternary complex</p>  <p>p16.7 monomer</p>  <p>AF model</p>		
<p>p16.8 AF model</p> 	<p>p16.9 AF model</p> 	<p>p17 AF model</p> 

Figure S2. Experimentally derived structures and AF models of ϕ 29 proteins. In ternary complexes, different components are distinguished with different colours. Separate monomers are shown using a rainbow colour gradient, where the N-terminal end is blue and the C-terminus is red.

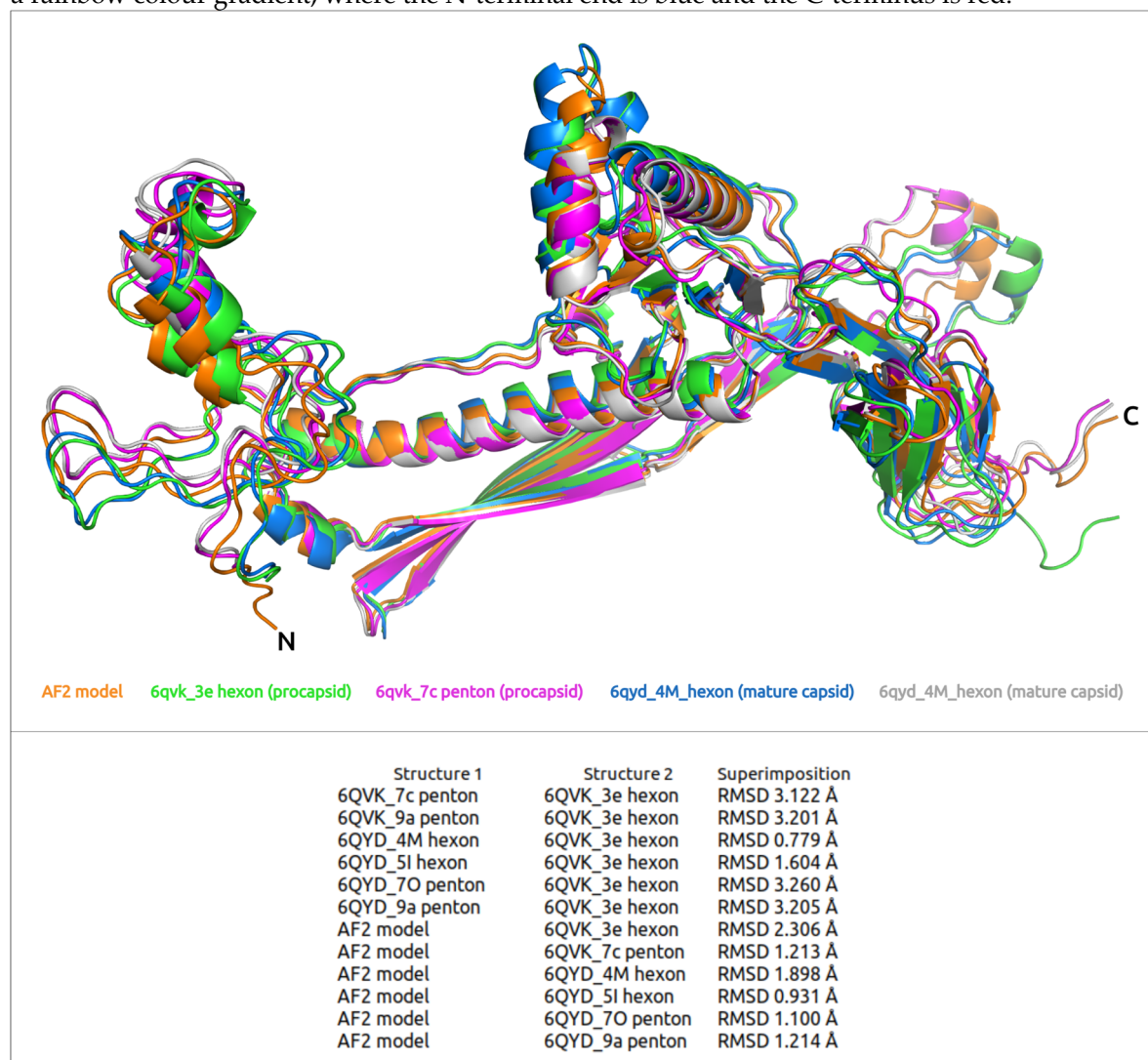


Figure S3. Superposition as predicted by AF and experimentally derived structures of major capsid protein. PDB codes, including chain numbers, are indicated in labels.

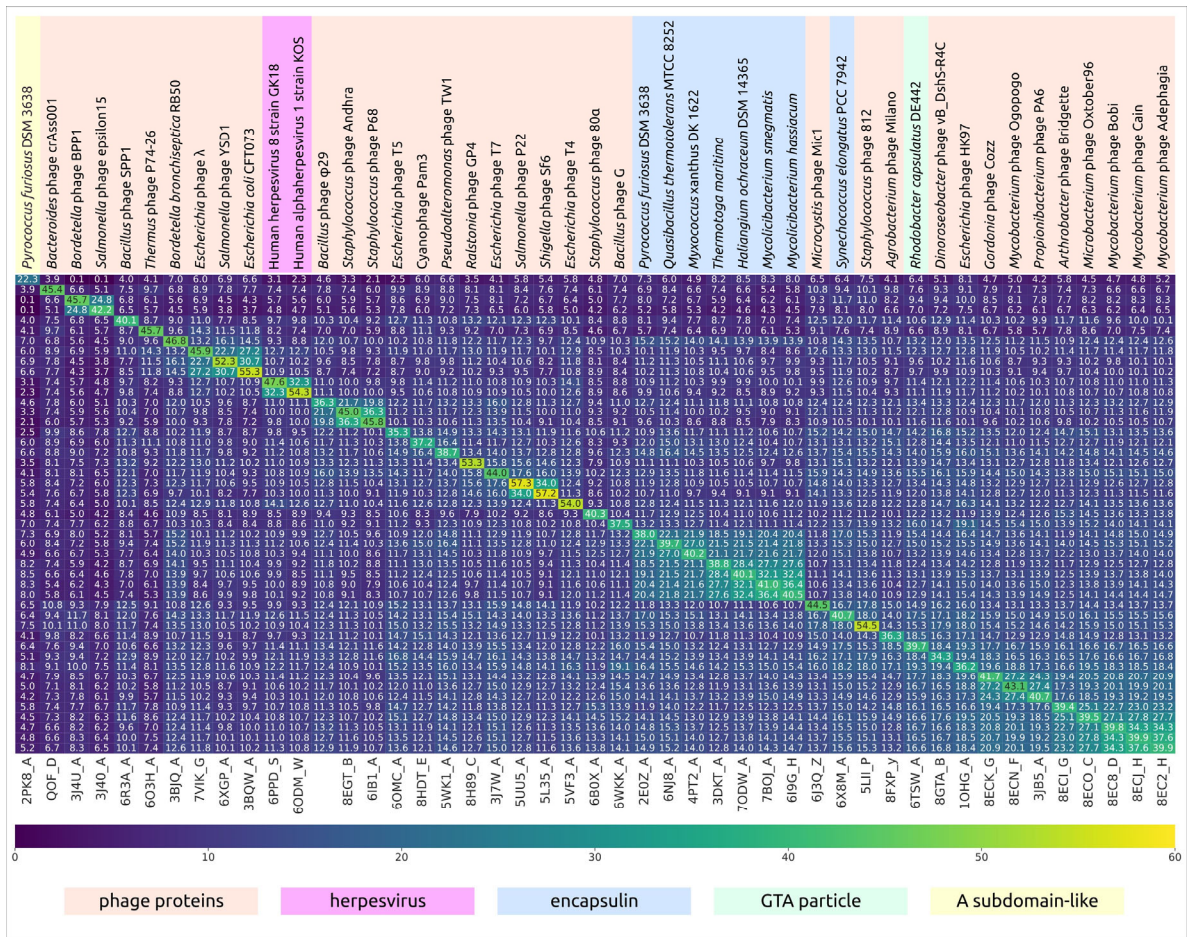


Figure S4. Heatmap with numerical values based on the pairwise structural similarity of viral capsid proteins and encapsulin major shell proteins, as measured by DALI Z-scores.



Figure S5. Maximum likelihood phylogenetic trees based on amino acid sequences of MCP predicted to be encoded in 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 legends. Branches with a bootstrap support lower than 50% have been deleted. Bootstrap values are shown near their branches. The scale bar shows one estimated substitution per site and the tree was rooted to the midpoint.

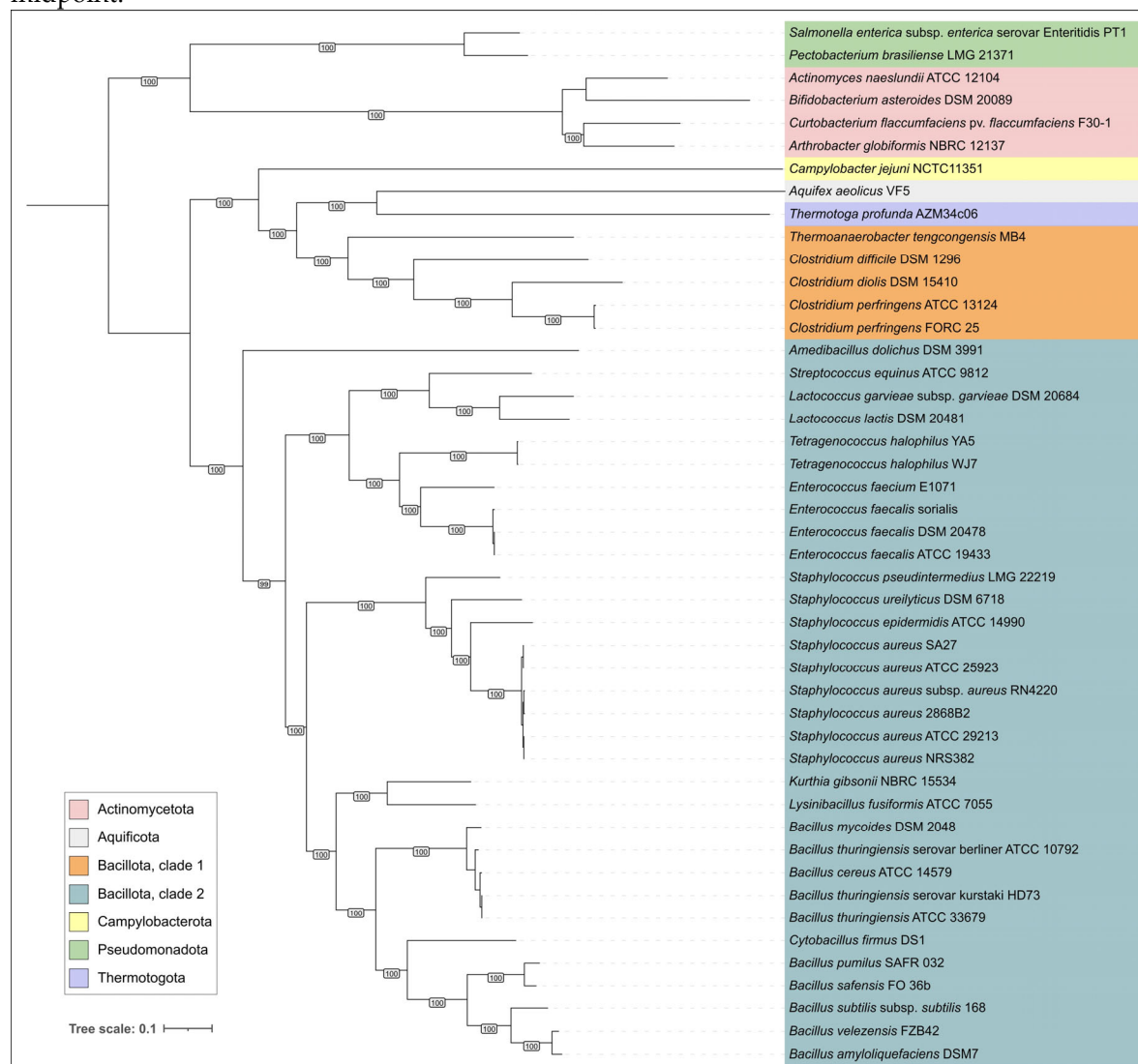
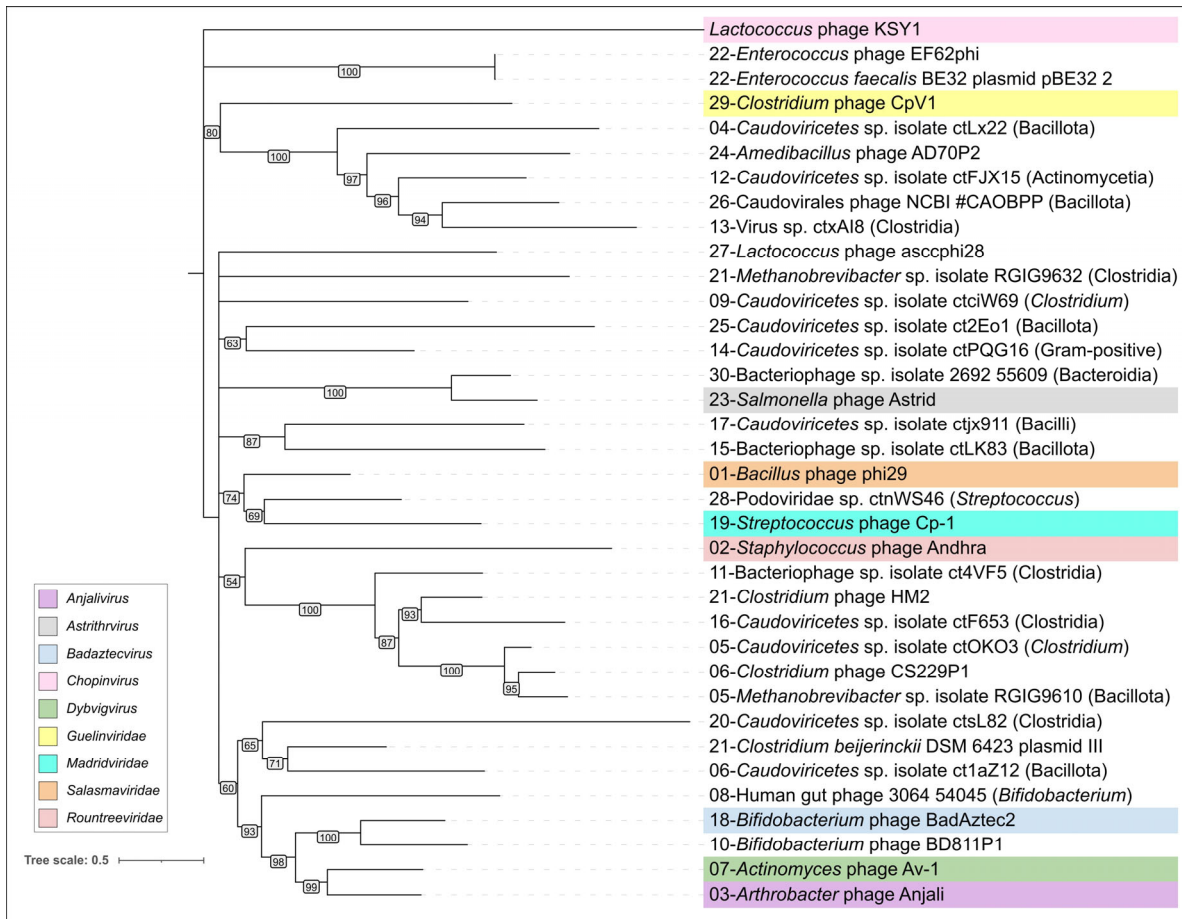
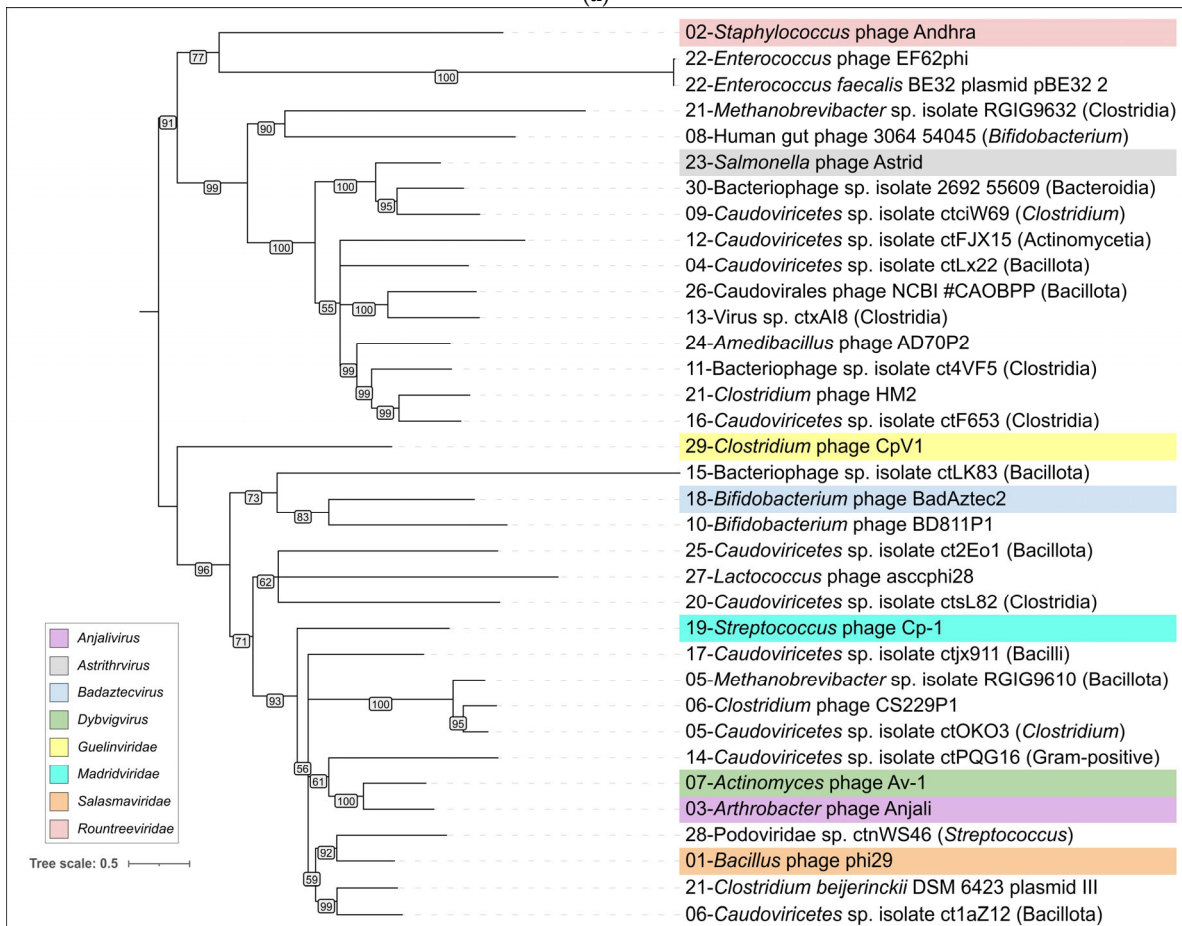


Figure S6. Maximum likelihood phylogenetic tree based on concatenated alignments of 81 bacterial core genes [143], extracted from the genomic sequences of bacterial hosts of ϕ 29-like and related phages, as well as other related bacteria. Bacterial taxonomy is indicated in the legend. Branches with a bootstrap support lower than 50% have been deleted. Bootstrap values are shown near their branches. The scale bar shows 0.1 estimated substitutions per site and the tree was rooted to the midpoint.

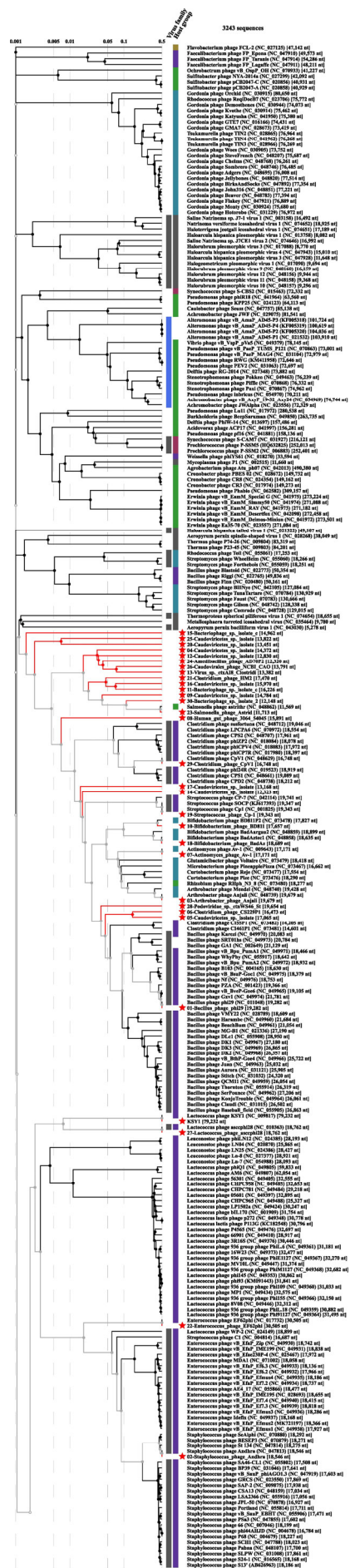


(a)



(b)

Figure S7. Maximum likelihood phylogenetic trees based on amino acid sequences of Ter (a) and DNAP (b), 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.



Aligned genome fraction

Genome length ratio

Intergenomic similarity

0 0.25 0.5 0.75 1

0 20 40 50 60 70 80 90 100

27-Lactococcus_phage_ascsp128
Lactococcus_phage_KSY1
09-Caudoviricetes_sp_isolate_ctcW69
23-Salmonella_phage_Astrid
30-Bacteriophage_sp_isolate_2692_55609
02-Staphylococcus_phage_Andhra
02-Staphylococcus_phage_S-CoN_Ph28
22-Enterococcus_faecalis_BE32_plasmid_pBE32_2
22-Enterococcus_phage_EF62phi
20-Caudoviricetes_sp_isolate_ctsL82
29-Clostridium_phage_CpV1
19-Streptococcus_phage_Cp-1_DNA
01-Bacillus_phage_phi29
28-Podoviridae_sp_ctnW546
12-Caudoviricetes_sp_isolate_ctfJX15
04-Caudoviricetes_sp_isolate_ctLX22
20-Caudovirales_phage_NCBI_#CAOBPP
14-Caudoviricetes_sp_isolate_ctPQG16
25-Caudoviricetes_sp_isolate_ct2Eo1
21-Methanobrevibacter_sp_isolate_RIGI9632
21-Clostridium_phage_HM2
24-Ameibacillus_phage_AD70P2
11-Bacteriophage_sp_isolate_ct4VF5
16-Caudoviricetes_sp_isolate_ctF653
06-Caudoviricetes_sp_isolate_ct1aZ12
21-Clostridium_beijerinckii_DSM_6423_plasmid_III
21-Bacteriophage_sp_isolate_ctCvG65
06-Clostridium_phage_CS229P1
05-Caudoviricetes_sp_isolate_ctOKO3
05-Methanobrevibacter_sp_isolate_RIGI9610
08-Human_gut_phage_3064_54045
03-Arthrobacter_phage_Anjali
07-Actinomyces_phage_Av-1
10-Bifidobacterium_phage_BD811P1
18-Bifidobacterium_phage_BadAztec2
17-Caudoviricetes_sp_isolate_ctpX911
13-Virus_sp_ctxA18
15-Bacteriophage_sp_isolate_ctLK83