

Comparative genomics and evolutionary trajectories of viral ATP dependent DNA-packaging systems

A. Maxwell Burroughs^{+‡}, Lakshminarayan M. Iyer[‡] and L. Aravind^{‡*}

Supplementary Material

[‡] National Center for Biotechnology Information,
National Library of Medicine,
National Institutes of Health,
Bethesda, MD 20894, USA

+ Bioinformatics Program,
Boston University,
Boston, MA 02215, USA

*For correspondence: email aravind@ncbi.nlm.nih.gov; telephone (301) 594-2445;
fax: (301) 480-9241

Table of Contents

Figures Legend

Materials and Methods

References

Tables

Table 1. Viral packaging ATPase types, broken down by viral class and chromosome type

Table 2. Gene neighborhoods of TLS/portal proteins in completely-sequenced phage genomes

Table 3. Detailed classification of families comprising the Portal Protein superfamily

Supplementary figures

Fig. 1. ϕ 29-like ATPase alignment with FtsK/HerA superfamily sequences

Fig. 2. Terminase N-terminal ATPase alignment

Fig. 3. Adenovirus IVA alignment with ABC superfamily sequences

Fig. 4. Portal Protein superfamily alignment

Fig. 5. FIS-type HTH in terminase small subunit alignment

Fig. 6. MuF domain alignment

Legend for figures.

Proteins are represented by their gene name, species abbreviation, and gi number; separated by underscores. Amino acid residues are colored according to side chain properties and the extent of conservation in the multiple alignment. Coloring is indicative of the percent consensus, which is shown on the last line of the alignment and differs for each alignment. The consensus abbreviations and the general coloring scheme are as follows: h, hydrophobic residues (ACFILMVWY) shaded yellow; s, small residues (AGSVCDN) colored green; o, alcohol group containing residues (ST) colored aqua; b, big residues (EFHIKLMQRWY) shaded in light yellow. Critical residues are often shown in a distinct color and shading pattern. Secondary structure assignments are shown above the alignment where E represents a strand and H a helix.

Species abbreviations for the alignments are as follows:

Aper : *Aeropyrum pernix*; Asp. : *Acinetobacter* sp.; BAV2 : Bovine adenovirus 2; BAV3 : Bovine adenovirus 3; BAVA : Bovine adenovirus A; BAVB : Bovine adenovirus B; BP01 : Bacteriophage Felix 01; BP11 : *Staphylococcus* phage 11; BP11b : Bacteriophage 11b; BP186 : Enterobacteria phage 186; BP187 : Bacteriophage 187; BP2638A : Bacteriophage 2638A; BP315.6 : *Streptococcus pyogenes* phage 315.6; BP66 : Bacteriophage 66; BP85 : Bacteriophage 85; BP933W : Bacteriophage 933W; BPA118 : *Listeria* phage A118; BPA2 : *Lactobacillus casei* bacteriophage A2; BPAPSE-1 : *Acyrtosiphon pisum* bacteriophage APSE-1; BPAaphi23 : Bacteriophage Aaphi23; BPAeh1 : *Aeromonas* phage Aeh1; BPB103 : *Bacillus* phage B103; BPB3 : Bacteriophage B3; BPBCJA1c : *Bacillus clarkii* bacteriophage BCJA1c; BPBIP-1 : *Bordetella* phage BIP-1; BPBMP-1 : *Bordetella* phage BMP-1; BPBPP-1 : *Bordetella* phage BPP-1; BPBarnyard : *Mycobacterium* phage Barnyard; BPBcep1 : *Burkholderia cenocepacia* phage Bcep1; BPBcepB1A : *Burkholderia cenocepacia* phage BcepB1A; BPBcepC6B : *Burkholderia cepacia* complex phage BcepC6B; BPBcepMu : *Burkholderia cenocepacia* phage BcepMu; BPBcepNazgul : *Burkholderia cepacia* phage BcepNazgul; BPBxz2 : *Mycobacterium* phage Bxz2; BPC1 : *Streptococcus* phage C1; BPCp-1 : *Streptococcus* phage Cp-1; BPD3 : *Pseudomonas* phage D3; BPD3112 : Bacteriophage D3112; BPES18 : *Salmonella typhimurium* bacteriophage ES18; BPETA : Bacteriophage phi ETA; BPF116 : *Pseudomonas* phage F116; BPGA-1 : *Bacillus* phage GA-1; BPHF2 : *Halorubrum* phage HF2; BPHK022 : Enterobacteria phage HK022; BPHK620 : Enterobacteria phage HK620; BPHK97 : Enterobacteria phage HK97; BPI : Stx2 converting bacteriophage I; BPII : Stx2 converting bacteriophage II; BPJL001 : Bacteriophage phi JL001; BPKS7 : Bacteriophage KS7; BPKVP40 : *Vibrio* phage KVP40; BPL5 : *Mycobacterium* phage L5; BPLL-H : Bacteriophage LL-H; BPLj928 : *Lactobacillus johnsonii* prophage Lj928; BPLj965 : *Lactobacillus johnsonii* prophage Lj965; BPM13 : Enterobacteria phage M13; BPMB78 : Bacteriophage MB78; BPMu : Enterobacteria phage Mu; BPN15 : Enterobacteria phage N15; BPO1205 : *Streptococcus thermophilus* temperate bacteriophage O1205; BPP-SSM2 : Cyanophage P-SSM2; BPP-SSM4 : Cyanophage P-SSM4; BPP1 : Enterobacteria phage P1; BPP2 : Enterobacteria phage P2; BPP22 : Enterobacteria phage P22; BPP27 : Bacteriophage P27; BPPBC5 : *Sinorhizobium meliloti* phage PBC5; BPPaP2 : *Pseudomonas aeruginosa* bacteriophage PaP2; BPPaP3 : *Pseudomonas aeruginosa* phage PaP3; BPPsP3 : *Pseudomonas* phage PsP3; BPS-WHM1 : Cyanophage S-WHM1; BPSP6 : Enterobacteria phage SP6; BPSPBc2 : Bacteriophage SPBc2; BPSPP1 : *Bacillus* phage SPP1; BPST64B : *Salmonella typhimurium* phage ST64B; BPST64T : Enterobacteria phage ST64T; BPSf6 : Enterobacteria phage Sf6; BPSfV : *Shigella* phage SfV; BPSfi11 : *Streptococcus* phage Sfi11; BPSfi21 : *Streptococcus* phage Sfi21; BPT1 : Enterobacteria phage T1; BPT3 : Enterobacteria phage T3; BPT4 : Enterobacteria phage T4; BPT5 : Enterobacteria phage T5; BPT7 : Enterobacteria phage T7; BPTM4 : *Mycobacterium*

phage TM4; BPTP901-1 : *Lactococcus* phage TP901-1; BPTuc2009 : *Lactococcus* phage Tuc2009; BPU2 : Mycobacteriophage U2; BPVHML : *Vibrio harveyi* bacteriophage VHML; BPVP16T : *Vibrio parahaemolyticus* phage VP16T; BPVP2 : *Vibriophage* VP2; BPVT2-Sa : Bacteriophage VT2-Sa; BPWocauB1 : Bacteriophage WOcauB1; BPXp15 : *Xanthomonas campestris* pv. *pelargonii* phage Xp15; BPbIL310 : Bacteriophage bIL310; BPbacteriophage : Stx1 converting bacteriophage; BPc2 : *Lactococcus* phage c2; BPepsilon15 : Enterobacteria phage epsilon15; BPfs2 : *Vibrio* phage fs2; BPgh-1 : *Pseudomonas* phage gh-1; BPlambda : Enterobacteria phage lambda; BPphBC6A51 : Bacteriophage phBC6A51; BPphi29 : *Bacillus* phage phi29; BPphiA1122 : *Yersinia pestis* phage phiA1122; BPphiAsp2 : *Actinoplanes* phage phiAsp2; BPphiE125 : Bacteriophage phiE125; BPphiHSIC : *Listonella pelagia* phage phiHSIC; BPphiJL-1 : *Lactobacillus plantarum* bacteriophage phiJL-1; BPphiP68 : *Staphylococcus aureus* phage phiP68; BPphiYeO3-12 : Bacteriophage phiYeO3-12; BPphig1e : Bacteriophage phig1e; BPpsiM100 : *Methanothermobacter wolfei* prophage psiM100; BPpsiM2 : *Methanobacterium* phage psiM2; BPr1t : *Lactococcus* phage r1t; BPsK1 : *Lactococcus* phage sk1; BPul36 : *Lactococcus* phage ul36; Bbro : *Bordetella bronchiseptica*; Bcer : *Bacillus cereus*; Bcla : *Bacillus clausii*; Bfun : *Burkholderia fungorum*; Bhen : *Bartonella henselae*; Blon : *Bifidobacterium longum*; Bper : *Bordetella pertussis*; Bvie : *Burkholderia vietnamiensis*; CAVt1 : Canine adenovirus type 1; Cjej : *Campylobacter jejuni*; DAVA : Duck adenovirus A; Ddes : *Desulfovibrio desulfuricans*; Dnod : *Dichelobacter nodosus*; Drad : *Deinococcus radiodurans*; Dvul : *Desulfovibrio vulgaris*; EHV1 : Equine herpesvirus 1; Ecol : *Escherichia coli*; Efae : *Enterococcus faecalis*; EsV1 : *Ectocarpus siliculosus* virus 1; FAV : Frog adenovirus; FAV1 : Fowl adenovirus 1; FAV2 : Fowl adenovirus 2; FAV8 : Fowl adenovirus 8; FAVA : Fowl adenovirus A; HAVt1 : Human adenovirus type 1; HAVt1 : Human adenovirus type 12; HHV7 : Human herpesvirus 7; HVH : Halovirus HF1; Hdud : *Haemophilus ducreyi*; Hinf : *Haemophilus influenzae*; Hpyl : *Helicobacter pylori*; Hsap : *Homo sapiens*; Hsom : *Haemophilus somnus*; Krad : *Kineococcus radiotolerans*; Linn : *Listeria innocua*; Llac : *Lactococcus lactis*; Lpla : *Lactobacillus plantarum*; Lpne : *Legionella pneumophila*; MAV1 : Murine adenovirus 1; MVP : Mycoplasma virus P1; Mace : *Methanosarcina acetivorans*; Mcap : *Methylococcus capsulatus*; Mjan : *Methanocaldococcus jannaschii*; Mmus : *Mus musculus*; Msp. : *Mesorhizobium* sp.; Msuc : *Mannheimia succiniciproducens*; Naro : *Novosphingobium aromaticivorans*; Nmen : *Neisseria meningitidis*; Npun : *Nostoc punctiforme*; OAV7 : Ovine adenovirus 7; OHV1 : Ostreid herpesvirus 1; PAV5 : Porcine adenovirus 5; PAVA : Porcine adenovirus A; PAVC : Porcine adenovirus C; PBCV : Paramecium bursaria Chlorella virus 1; Pflu : *Pseudomonas fluorescens*; Plum : *Photobacterium luminescens*; Psp. : *Polaromonas* sp.; Psp. : *Psychrobacter* sp.; Rpro : *Rickettsia prowazekii*; Rsph : *Rhodobacter sphaeroides*; SAV1 : Simian adenovirus 1; SAV2 : Simian adenovirus 25; SAVA : Simian adenovirus A; Sent : *Salmonella enterica*; Sepi : *Staphylococcus epidermidis*; Shae : *Staphylococcus haemolyticus*; Sisp : *Silicibacter* sp.; Sone : *Shewanella oneidensis*; Spyo : *Streptococcus pyogenes*; Ssp : *Synechocystis* sp.; Styp : *Salmonella typhimurium*; TAVA : Turkey adenovirus A; Tlit : *Thermococcus litoralis*; Tmar : *Thermotoga maritima*; Tpal : *Treponema pallidum*; TsAV : Tree shrew adenovirus; VP : Virus PhiCh1; Vvul : *Vibrio vulnificus*; Wsp. : *Wolbachia* sp.; Xfas : *Xylella fastidiosa*

Materials and Methods

Profile searches were conducted using the PSI-BLAST program with either a single sequence or alignment as query [1]. This were typically run with a PSSM inclusion expectation (*E*) value threshold of 0.01, and were iterated until convergence. Multiple alignments of protein sequence were constructed using the T_Coffee [2],

PCMA [3] and MUSCLE [4] software packages, followed by manual correction based on the PSI-BLAST results. Searches of the PDB database with query structures were conducted using the DALI program [5]. Protein secondary structure was predicted using a multiple alignment as the input for the JPRED program, with information extracted from a PSSM, HMM and the seed alignment itself [6]. Similarity based clustering of proteins was carried out using the BLASTCLUST program (<ftp://ftp.ncbi.nih.gov/blast/documents/README.bcl>). Protein structures were visualized and manipulated using the Swiss-PDB viewer program [7]. Phylogenetic analysis was carried out using the maximum-likelihood, neighbor-joining and least squares methods. Briefly, this process involved the construction of a least squares tree using the FITCH program or a neighbor joining tree using the NEIGHBOR program (both from the Phylip package) [8], followed by local rearrangement using the Protml program of the Molphy package [9] to arrive at the maximum likelihood (ML) tree. The statistical significance of various nodes of this ML tree was assessed using the relative estimate of logarithmic likelihood bootstrap (Protml RELL-BP), with 10,000 replicates. Gene neighborhoods were analyzed using a custom script of the TASS package (V. Anantharaman, S. Balaji and L.A unpublished) that uses completely sequenced genomes or whole genome shot gun sequences to derive a table of gene neighbors centered on a query gene.

References

1. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;25:3389-3402.
2. Notredame C, Higgins DG, Heringa J: T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* 2000;302:205-217.
3. Pei J, Sadreyev R, Grishin NV: PCMA: fast and accurate multiple sequence alignment based on profile consistency. *Bioinformatics* 2003;19:427-428.
4. Edgar RC: MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 2004;5:113.
5. Holm L, Sander C: Dali: a network tool for protein structure comparison. *Trends Biochem Sci* 1995;20:478-480.
6. Cuff JA, Clamp ME, Siddiqui AS, Finlay M, Barton GJ: JPred: a consensus secondary structure prediction server. *Bioinformatics* 1998;14:892-893.
7. Guex N, Peitsch MC: SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* 1997;18:2714-2723.
8. Felsenstein J: *Inferring Phylogenies*. Sunderland, MA, Sinauer Associates, 2004.
9. Adachi J, Hasegawa M: *MOLPHY: Programs for Molecular Phylogenetics*. Tokyo, Institute of Statistical Mathematics, 1992.

Table 1. Viral packaging ATPase types, broken down by viral class and chromosome type

| virus type | packaging ATPase type | | | | | average size (kb) |
|---------------------------|-----------------------|-------------|--------|--------|---------------|-------------------|
| | terminase | phi-29 like | FtsK | none | other/unknown | |
| microviridae | 0/15 | 0/15 | 0/15 | 15/15 | 0/15 | 5085 |
| inoviridae | 0/22 | 0/22 | 20/22 | 1/22 | 1/22 | 7690 |
| caudovirales* | 186/205 | 9/205 | 3/205 | 6/205 | 2/205 | 56105 |
| globuloviridae | 0/2 | 0/2 | 0/2 | 0/2 | 2/2 | 24635 |
| lipothrixviridae | 0/1 | 0/1 | 0/1 | 0/1 | 1/1 | 40047 |
| rudiviridae | 0/3 | 0/3 | 1/3 | 0/3 | 2/3 | 28474 |
| tectiviridae | 0/3 | 0/3 | 3/3 | 0/3 | 0/3 | 14901 |
| corticoviridae | 0/1 | 0/1 | 1/1 | 0/1 | 0/1 | 10079 |
| fuselloviridae | 0/4 | 0/4 | 4/4 | 0/4 | 0/4 | 16030 |
| Plasmaviridae | 0/1 | 0/1 | 1/1 | 0/1 | 0/1 | 19343 |
| salterproviridae | 0/2 | 0/2 | 1/2 | 1/2 | 0/2 | 15265 |
| unclassified | 18/29 | 0/29 | 6/29 | 5/29 | 0/29 | 35908 |
| total fraction of ATPases | 204/288 | 9/288 | 40/288 | 28/288 | 8/288 | |

*the caudovirus Mycobacterium phage Omega contains both FtsK and terminase-like ATPases

| Type of ATPase | linear | linear % | circular | circular % | total |
|----------------|--------|----------|----------|------------|-------|
| Terminase | 174 | 84.9 | 31 | 15.1 | 205 |
| FtsK | 13 | 32.5 | 27 | 67.5 | 40 |

| Table_2.txt | | | | | | | | |
|--|--------------------|--------------------|-----------|----------|----------------------------|----------------|---|---------|
| Streptococcus pyogenes phage 315.5 | 28876410 | 28876408 | 28876411 | 28876407 | SPP1-like | lambda-like | l | 28/55 |
| Streptococcus pyogenes phage 315.6 | 28876465 | 28876464 | 28876466 | 28876463 | SPP1-like | SPP1-like | l | 24/51 |
| Acyrtosiphon pisum bacteriophage APSE-1 | 9633565 | 9633566 | 9633564 | | SPP1-like (P22 subfamily) | SPP1-like | c | |
| Salmonella typhimurium bacteriophage ES18 | 62362215 | 62362218 | 62362214 | 62362219 | T1-like | SPP1-like | l | 2/79 |
| Pseudomonas aeruginosa phage PaP3 | 27476053 | 27476052 | 27476055 | | SPP1-like (P22 subfamily)? | SPP1-like | c | |
| Staphylococcus aureus phage phi 11 | 29028593 | 29028594 | 29028592 | 29028595 | SPP1-like | SPP1-like | l | 30/53 |
| Enterobacteria phage P1 | 46401730 | 46401636? | 46401729 | | T4-like | SPP1-like | l | 106/110 |
| Actinoplanes phage phiAsp2 | 48697461 | 48697405? | | 48697407 | lambda-like? | | l | 60/76 |
| Bacteriophage KS7 | 57472357 | 57472356 | 71801571 | 62327320 | SPP1-like (KS7 subfamily) | solo | l | 16/59 |
| Bacteriophage VT2-Sa | 9633446 | 9633447 | 9633445 | | SPP1-like (P22 subfamily) | SPP1-like | l | 52/83 |
| Bacteriophage phi JL001 | 62327245 | 62327246 | 62327244 | 62327248 | SPP1-like (KS7 subfamily) | SPP1-like | l | 59/90 |
| Bordetella phage BIP-1 | 45580769 | 45580765 | 45580770 | | T7/T3-like? | SPP1-like | l | 24/48 |
| Burkholderia cepacia complex phage BcepC6B | 48697193 | 48697195 | 48697192 | | T7/T3-like | SPP1-like | l | 3/46 |
| Burkholderia cepacia phage BcepNazgul | 34304547 | 34304545 | 34304548 | | lambda-like | solo | l | 44/75 |
| Lactobacillus casei bacteriophage A2 | 22217799 | 22217800 | 22217798 | | T5-like | D3-like | l | 2/61 |
| Methanobacterium phage psiM2 | 3249594 | 3249595 | 3249593 | 9630536 | T5-like | SPP1-like | l | 10/32 |
| Methanothermobacter wolfeii prophage psiM100 | 11692721 | 11692722 | 11692720 | 11863103 | T5-like | SPP1-like | l | 16/35 |
| Enterobacteria phage ST64T | 24250810 | 24250812 | 24250811 | | SPP1-like (P22 subfamily) | T1/P22-like | l | 52/65 |
| Streptococcus thermophilus bacteriophage Sfill | 7523569 | 3320433 | 7669463 | 9635012 | SPP1-like | SPP1-like | l | 4/53 |
| Vibrio harveyi bacteriophage VHML | 26891709 | 26891710 | 26891708 | | lambda-like | lambda-like | l | 22/57 |
| Vibriophage VP2 | 48696639 | 48696640 | | | T7/T3-like | | l | 3/47 |
| Enterobacteria phage N15 | 9630466 | 9630468 | 9630465 | | lambda-like | lambda-like | l | 2/60 |
| Enterobacteria phage lambda | 9626245 | 9626247 | 9626244 | | lambda-like | lambda-like | l | 2/71 |
| Staphylococcus phage K | 48696425 | 48696430 | | | T5-like | | l | 35/115 |
| Lactobacillus plantarum bacteriophage LP65 | 56693161 | 56693160 | | | T5-like | | c | |
| Halorubrum phage HF2 | 18138498 | 18138497 | | | T5-like | | l | 108/114 |
| Vibrio phage VP5 | 48696683 | 48696684 | | | T7/T3-like | | c | |
| Bacteriophage EJ-1 | 39653713 | 39653714 | 39653712 | 39653715 | SPP1-like | phiNIH1.1-like | l | 39/73 |
| Burkholderia cenocepacia phage BcepB1A | 48697520 | 48697517 | 48697530 | 48697516 | T1-like | SPP1-like | l | 33/73 |
| Halovirus HF1 | 32453919 | 32453918 | | | T5-like | | l | 95/102 |
| Bacteriophage 933W | 9632517 | 9632518 | 9632516 | | SPP1-like (P22 subfamily) | SPP1-like | l | 51/80 |
| Bacteriophage Mx8 | 15320614 | 15320615? | | | | | c | |
| Bacteriophage WPhi | 33438899 | 33438898 | 33438899% | | T5-like | SPP1-like | l | 2/44 |
| Bordetella phage BMP-1 | 45569518 | 45569514 | 45569519 | | T7/T3-like | SPP1-like | l | 24/47 |
| Bordetella phage BPP-1 | 41179386 | 41179382 | 41179387 | | T7/T3-like | SPP1-like | l | 25/49 |
| Enterobacteria phage P22 | 51236724 | 51236726 | 51236745 | | SPP1-like (P22 subfamily) | T1/P22-like | c | |
| Lactobacillus johnsonii prophage Lj928 | 41179314 | 41179315 | 41179313 | 41179316 | SPP1-like | SPP1-like | l | 27/50 |
| Lactobacillus johnsonii prophage Lj965 | 41179257 | 41179240 | 41179239 | 41179242 | SPP1-like | SPP1-like | l | 25/46 |
| Mycobacterium phage Barnyard | 29566981 | 29566982? | 29566993 | | | | l | 6/109 |
| Staphylococcus aureus prophage phiPV83 | 9635715 | 9635717 | 9635714 | | T5-like | D3-like | l | 40/65 |
| Streptococcus pyogenes phage 315.4 | 28876345 | 28876344 | 28876346 | 28876343 | phigle-like | phiNIH1.1-like | l | 36/64 |
| Streptococcus thermophilus temperate bacteriophage O1205 | 23455873 | 23455874 | 23455900 | 23455875 | SPP1-like | SPP1-like | l | 26/57 |
| Temperate phage phiNIH1.1 | 16271803 | 16271804 | 16271802 | | phigle-like | phiNIH1.1-like | l | 27/55 |
| Salmonella typhimurium bacteriophage ST104 | 46358697 | 46358698 | 46358696 | | SPP1-like (P22 subfamily) | T1/P22-like | l | 49/63 |
| Stx2 converting bacteriophage I | 20065962 | 20065978 | 20065960 | | SPP1-like (P22 subfamily) | SPP1-like | c | |
| Stx2 converting bacteriophage II | 32171173 | 32171003 | 32171171 | | SPP1-like (P22 subfamily) | SPP1-like | c | |
| Enterobacteria phage RB43 | 66391670 | 66391671 | 66391669 | | T4-like | T4-like | l | 220/292 |
| Aeromonas phage 31 | 66391955 | | | | | | l | 150/247 |
| Bacteriophage Lc-Nu | 78000026 | 78000027 | 77999989 | | T5-like | D3-like | l | 2/51 |
| Vibriophage VP4 | 68299748 | 68299738 | 68299739 | | T7/T3-like | T7-like | l | 31/31 |
| Bacteriophage 3A | 66395593 | 66395598 | 66395625 | | T5-like | D3-like | l | 2/67 |
| Bacteriophage 47 | 66395661 | 66395665 | 66395693 | | T5-like | D3-like | l | 2/72 |
| Bacteriophage 2638A | 66395455 | 66395461 | 66395478 | | T5-like | D3-like | l | 2/57 |
| Bacteriophage 42e | 66395511 | 66395516 | 66395551 | | T5-like | D3-like | l | 2/79 |
| Enterobacteria phage K1F | 77118209 | 77118196 | | | T7/T3-like | | l | 42/43 |
| Burkholderia cepacia phage Bcep176 | 77864705 | 77864704 | 77864706 | | T5-like | SPP1-like | l | 80/81 |
| Xanthomonas oryzae phage OP1 | 84662597 | 84662598 | | | T5-like | | l | 4/59 |
| Bacillus anthracis phage Gamma | 77020155 | 77020175 | 77020145 | | T5-like | D3-like | l | 2/53 |
| Bacteriophage 37 | 66395738 | 66395736 | 66395762 | | SPP1-like | SPP1-like | l | 2/77 |
| Bacillus anthracis phage Cherry | 77020208 | 77020228 | 77020199 | | T5-like | D3-like | l | 2/51 |
| Burkholderia pseudomallei phage phi52237 | 72537721 | 72537722 | 72537717 | | T5-like | SPP1-like | l | 45/48 |
| Bacteriophage 92 | 66396415 | 66396413 | 66396433 | | SPP1-like | SPP1-like | l | 3/74 |
| Bacteriophage EW | 66395816 | 66395815 | 66395841 | | SPP1-like | lambda-like | l | 2/77 |
| Bacteriophage X2 | 66394679 | 66395841 | 66394708 | | SPP1-like | lambda-like | l | 2/77 |
| Bacteriophage 71 | 66396048 | 66396046 | 66396079 | | SPP1-like | lambda-like | l | 2/72 |
| Bacteriophage ROSA | 66395973 | 66395972 | 66395998 | | SPP1-like | lambda-like | l | 2/74 |
| Xanthomonas campestris pv. pelargonii phage Xp15 | 66392134 | 66392058 | | | SPP1-like (KS7 subfamily) | | l | 7/84 |
| Bacteriophage RTP | 81343949 | 81343951 | 81343947 | | T1-like | T1/P22-like | l | 21/75 |
| Bacteriophage 29 | 66396200 | 66396197 | | | SPP1-like | | l | 2/75 |
| Bacteriophage 52A | 66396276 | 66396273 | | 17426270 | SPP1-like | | l | 2/65 |
| Bacteriophage phi ETA | 17426268 | 17426269 | | | SPP1-like | | c | |
| Bacteriophage 96 | 66395894 | 66395892 | | | SPP1-like | | l | 2/79 |
| Bacteriophage 85 | 66394878 | 66394876 | 66394897 | | SPP1-like | SPP1-like | l | 2/78 |
| Staphylococcus phage K | 48696425 | 48696430 | | | T5-like | | l | 34/115 |
| Xanthomonas oryzae phage OP2 | 84662666 | 84662665 | | | T1-like | | l | 13/62 |
| Enterobacteria phage K1E | 83571766, 83571768 | | | | | | l | 51/62 |
| Streptococcus thermophilus bacteriophage 2972 | 66391762, 66391763 | 66391764 | 66391761 | | SPP1-like | SPP1-like | l | 3/44 |
| Staphylococcus phage Twort | 66391300 | 66391267 | | | T5-like | | l | 171/195 |
| Bacteriophage JK06 | 71834119, 71834121 | 71834115, 71834116 | | | T1-like | | c | |
| Bacteriophage 187 | 66395220 | 66395218 | 66395243 | | SPP1-like | SPP1-like | l | 2/77 |
| Bacteriophage 53 | 66395379 | 66395376 | 66395396 | | SPP1-like | SPP1-like | l | 2/79 |
| Bacteriophage 88 | 66396341 | 66396340 | 66396360 | | SPP1-like | SPP1-like | l | 3/72 |
| Bacteriophage 69 | 66395301 | 66395299 | 66395317 | | SPP1-like | SPP1-like | l | 2/76 |
| Phage phiV10 | 89152423 | 89152428 | 89152470 | | T7/T3-like | SPP1-like | l | 2/56 |
| Clostridium botulinum phage C-St | 80159854 | 80159853 | | | lambda-like? | | l | 158/189 |

%-small terminase in this instance is fused to large terminase

Table 3. Detailed classification/description of families comprising the Portal Protein superfamily

Pfam/COG/DUF identifier given after family name, where applicable.

I. T1, T5, λ clade

Conserved H in final β -strand of SH3 domain

A. T1-like family (COG3567)

Conserved HxSR in SH3 β -strand domain, R/K in N-terminal helical domain, scores of conserved C-terminal polar residues.

B. T5-like family (pfam04860, Phage_portal)

Conserved H and two charged residues in SH3 β -strand domain, two charged residues immediately downstream of second β -strand region.

C. Phage λ -like family (pfam05136, Phage_portal_2)

GEhhh first conserved glycine motif, RG second conserved glycine motif, GxD third conserved glycine motif; conserved H in SH3 domain; Dxxx-, -, Ghx[S/T] at C-terminus.

II. T4, phig1e, SPP1 clade

Two conserved aromatic residues in SH3 domain region

A. T4-like family (pfam07230,

Phage_T4_Gp20) Pair of N-terminal β -strands; conserved P-RR, RR-G, --W, RDE signatures; GKY third conserved glycine motif.

B. phig1e-like family (pfam05126, Phage_min_cap)

Pair of N-terminal β -strands, conserved E in

helix upstream of SH3 domain, conserved K/R in SH3 domain, conserved R immediately after second β -strand region, TATE signature upstream of third glycine motif

C. SPP1-like family (DUF1483 and pfam05133, Phage_prot_Gp6)

SPP1 classic subfamily (Conserved – residue upstream of first glycine motif, conserved S in C-terminal helical region)

P22 subfamily (extended SH3 region with Dx[6]DxxD signature, extended C-terminal helical region, R and Q N-terminal conserved residues, DNh signature)

KS7 subfamily (several conserved charged residues in SH3 domain)

III. Phage Mu clade (DUF935)

Lacks 1-2 β -strands in SH3 domain relative to other families in the superfamily, extended C-terminal helical region, conserved K, E upstream of first glycine motif, several conserved charged residues.

IV. T7, T3 clade

Several conserved polar N-terminal residues, aGRS second conserved glycine motif signature, conserved TAxE signature.

Figure 2. Terminate N-terminal ATPase domain alignment- page 1

Table showing protein sequence alignment for the Terminate N-terminal ATPase domain. It includes columns for sequence identifiers (e.g., EHV1gp47_EHV1_9626782), residue numbers, and amino acid sequences. Consensus/80% is provided at the bottom.

Table showing protein sequence alignment for the WB and Sen-1 domains. It includes columns for sequence identifiers (e.g., EHV1gp47_EHV1_9626782), residue numbers, and amino acid sequences. Consensus/80% is provided at the bottom.

Figure 2. Terminase N-terminal ATPase domain alignment- page 2

| | | | | | |
|-----------------------------|--------------------------------|-----------------|--------------------|------------------------------------|-----------|
| SECONDARY STRUCTURE | -----EEEEEE-- | EEEE-- | -----H----- | HHHHHH-- | -----EE-- |
| EhV1gp47_EHV1_9626782 | SFLYNLKGAAADLL-----NVVTYICDE | 22 VFITMDGAMRNT | 1 ELFLPD----- | SFMQEIIT---GG--GNVSGAHRDEPVT | 483 |
| OsHV1_gp101_OHV1_48696823 | PFVDIRNTRQDGT-----CVVEYVCNP | 24 LHNAGAGVRKI | 1 AAFSVKTNKNMEVDD | 18 LTKNDIL---GM---NN--LQNMRLTSEL | 515 |
| U60_HHV7_51874288 | SFLTKLSNSPPEML-----TVVSYVCED | 22 KFISINADVKT | 1 DLFLG----- | AFKHEIM---GG--SLCN--VVNDTLITEQ | 425 |
| 17_BBP_BSM4_61806303 | NQFKYKLMHDAE-----RGANNVYAT | EVHNSQVGRDD | KWKQQTLEN-TSE-- | AQFRVFECEFLG--SVDTLITPSKLRIMPY | 287 |
| 17_BBP_BSM2_61806000 | NHFVRMWHHDE-----KKGSEYVAT | DVHNSVSPGRDE | EWKEQTIAN-TSE-- | QQKFLIECEFLG--SVNNTLINPAKLRNLYR | 284 |
| yonF_BPSBc2_9630181 | ITYYNAMMKGS-----KYFVCGLPY | QIAIKEGLDLDK | QVRDEMAEE-DFDP- | IGWSMEMEALWFG--ESEKAYFKFEDIEKRN | 313 |
| 17_BPT4_9632591 | YDIWTAAVEGK-----SGPEPYTAI | WNSVKERLYNDE | DIFDDGWQWSIQTN | 4 AQFRQEHHTAAFEG--TSGTLISGMKLAVMDF | 374 |
| Aeh1p227_BPAeh1_38640149 | YDLWEVLSKSD-----KGFKPYTTT | WITVKERLYDGS | DAYDDGFPEWASKQIN | 4 EAFQQEHLCRFMG--TSGTLINGFKLSKMTW | 398 |
| VP2p03_BPV2_48696639 | NWFYKMLAHAE-----KSEEWYKY | 1 TINDTWRWAYSS | 40 EKALIERGVV-KGYA | 19 FIVRQEHYCDWFG--ALQGSYGDLMITMYN | 357 |
| T5_155_BPT5_46401884 | NWFKEFYAYGFFD-----TLNPNWVSIH | 1 TYRDNPRADLND | IEEARRTVS-K---- | NYFRQYEAADFVS--FEGQIFDFTNAIDHVK | 264 |
| PhiCh1p03_VP_22091106 | EGYEFDEYPAILDYWD-----QQFSADDDY | 18 TGETLQVLWPEA | RGPRWLADK-RSKMA | 1 HRFWRWYSLVIMG--SSGDLIDAKDVRVPAE | 338 |
| terL_BPAphi23_31544028 | GGDQWRIVKFAIAE-----DDEEFKKEG | -----EPLHPERFDE | -----RLNKI-RQAVG | 1 QAWNALYQQRPN--KGGGILKGSWFGRYKV | 313 |
| T1p53_BPT1_45686309 | MGIEFDQISIPALVT-----EYGKTLPD | 18 DGVKHYSFWPSK | ESVHDLLAL-READQ | YTFDSQYQKPIA--LGGSVFNSEWWTYGS | 335 |
| N15p02_BPN15_9630466 | MRFHVACPHCG-----EQYLKFGD | RDPFPFGKWEPE | QAETVYYLCEHNACV | 6 DFNARYICELTGIWTRDGLRWLWSSNAEID | 317 |
| TL5_BPLambda_9626245 | MRFHVACPHCG-----EQYLKFGD | KETPFGLKWTDP | DPSSVYVYLCHEINACV | 6 DFTDARYICEKGTGIWTRDGLRWLWSSGEEIE | 317 |
| JL001p59_BPJL001_62327245 | GHNVVKHFPFLPAKD-----GIIRKTYE | FEDPETGEPPIK | 1 LCRVAIKGTIWENTI | LLKAQPNYIAQLKQADNPAKLAALWEGSMD | 258 |
| T7p57_BPT7_9627482 | TIIWPALYPRTR-----ENLYYSQR | 1 APMLRAEYDNP | 2 LAGTPTDPV-RFDRD | DLRERELLEYKAG--FTLQFMLNPNLSDAEK | 298 |
| T3p53_BPT3_17570838 | TIIWPALYPRSR-----EDLYYGER | 1 APMLREEFNDGF | 2 LQGQPTDPV-RFDME | DLRERELLEYKAG--FTLQFMLNPNLSDAEK | 298 |
| L5p13_BPL5_9625443 | EKAMDEYQKQAGSDVDTGMMYDLAEPAD | 19 IARGDSTVLPID | -----DIKSIIS-TKNPI | TESRRKHLNQVWA--AEDSWLSPQEWNRQCV | 383 |
| P27p36_BPP27_18249900 | REVIEMLGIVFND-----ELGFIYTI | 3 DDWTDPAVLRKA | 13 LSQQAAIN-NPRKV | GVFKTKHLNIWVA--AKDPAFNLNVWQKCED | 343 |
| D3p02_BPD3_9635588 | RDVTRILEGQTTIDE-----TIFGIYTI | 3 DPWDDPASLIKA | 13 LAQLQAKR-SASKQ | NAFRTKHLNQVWG--ARTVWMLNLAWRQQRK | 349 |
| SPP1p003_BPSPP1_22855048 | KQSVVNVKFNSS-----FLPANTFVD | HSTYLQNPFLSK | 1 FIEEAEVK-RR-NE | LKYRHEYLGEALG--SGVVPFENLQIEEGII | 250 |
| Ter1L_BPSfi11_7523569 | N-HWLNDRYIG-----KNDGKIIDF | SFKLDDNTFLSK | 1 YIDSIKAVT-PK--- | GKPYDRDILGHWT--VAEGAIYADYDSKTHV | 236 |
| gp2_BPST64T_24250810 | YSIYGEGLTRTN-----KYQFSILT | 1 TPLMGMSDVVTK | FLKNPSKSQ-KVVNM | TIYDAEHYTDQEKQIIASYPEHEREARARG | 304 |
| P22gp02_BPP22_51236724 | YSIYGEGLTRTN-----KYQFSILT | 1 TPLMGMSDVVTK | FLKNPSKSQ-KVVNM | TIYDAEHYTDQEKQIIASYPEHEREARARG | 286 |
| phig1ep37_BPphig1e_23455799 | PYDWINEWIDKVS-----DDNYLIDTS | DYRCDVRFPTSK | 1 TLDDLEQYK-KN-DY | EYRWLYLGEVIG--LGTSYNPNLLKPLEV | 262 |
| A118p02_BPA118_16798785 | PYEVWVNEYVDSKRS-----DDYLIHHT | 1 TYLDDKGFPLSK | 1 IKKIEKYK-KN-DY | DYRWMLYLGEVIG--LGDNVYNNMLFQPLKA | 261 |
| ORF9_BPpsiM100_11692721 | GHEWVKTRFITG-----EKTFFIPSTW | RENPLYNRDEYE | 2 LNMLDHVTR-RQ--- | -LKEGDWDVSIQG-----GVFRREWFEEIID | 282 |
| TerL_BPpsiM2_3249594 | GHEWVKTRFITG-----EKTFFIPSTW | RENPLYNRDEYE | 2 LNMLDHVTR-RQ--- | -LKEGDWDVTLQG-----GVFRREWFEEIID | 282 |
| Sf6p02_BPSf6_41057280 | SEIIVSFPNPKNIL-----DDTYQRFFV | NPPDDICLLTVN | 10 LRLEMEECK-RR-NP | TLYRHIWLGEPVS--ASDMAIKREWLEAAT | 225 |
| TerL_BPHK620_13559866 | SEIIVSFPNPKNIL-----DDTYQRFFV | NPPDDICLLTVN | 10 LRLEMEECK-RR-NP | TLYRHIWLGEPVS--ASDMAIKREWLEAAT | 225 |
| Orf12_BP186_9634058 | YFPWSGELFNKGRS-----RIADRIEID | 2 HRALAGGLCDD | 17 TLFDDQLK-RENSD | EDFKNLFMCEFVD--DKASVFPPELQRCMV | 389 |
| P2p02_BPP2_9630329 | YFPWSGELFNKGRS-----SAERVEID | 2 HNALAGGLCDD | 17 TLFDDQLK-RENSA | DDFKNLFMCEFVD--DKASVFPPELQRCMV | 389 |
| SP6p41_BPSp6_31711683 | IRIWPGRYPTLEQ-----BACYGDFLA | 1 MIRQDMIDPSL | 7 GQQGAPTCPEMYDDE | KLIEKEISQGTAK--FQLQFMLNTRLMADR | 319 |
| HalH1gp095_HVH_32453919 | FMDANERGTDPGRN-----DFGILALKQ | 2 FKNADEIQTQDVS | 11 DFDLMAAETQRASDP | NGFAQEVLCRPVS--DEYRFFSMPTIEDAMG | 298 |
| HF2p108_BPHF2_18138498 | FMDANERGTDPGRN-----DFGILALKQ | 2 FKNADEIQTQDVS | 11 DFDLMAAETQRASDP | NGFAQEVLCRPVS--DEYRFFSMPTIEDAMG | 298 |
| pacB_BPP1_46401730 | YDSSHRLAIRPGN-----PDGLPTAI I | 1 NSEESPLV-DAK | -FIRAKLAHYGGRDN | PMYMIKVRGEFPK--SQDGLLGRDVERAT | 276 |
| r1tp29_BPr1t_23455748 | TNYRDNTIAGKA-----KYSGWAESW | 1 EDVKDIHDVEAW | 8 YHLNERKIE-AELGE | 2 LDHNVQRLGYWPK--YNQKSVISQEWNALK | 293 |
| TM4_04_BPTM4_18496890 | TMVRLDALAGDV-----DDYGVVEIS | 1 DEDADPDRSQW | 11 SARAILRMR-KALGD | ESFKREAMGIWPK--VSVHQPVVKSGRWHDL | 296 |
| Map28_BPMu_9633519 | FNQYIQD-AREGR-----KDYSHRIT | 2 DAIADGLYRRIC | 15 RDGLYRNAPNKESAD | EYGCIPKKSOGGA--YLSRVLIEAAMTPARD | 304 |
| BcepMu28_BPBcepMu_48696938 | FNELVTD-VRSCK-----KPYSLHRIT | 2 DAVQDGLYQRIC | 13 KWVKDIRASYGADAE | EELDCVPKNSOGGA--WLSRALIESRMSADTP | 269 |
| B3ORF31_BPB3_56692599 | FNQLVREIVEGPN-----KNISLHTVT | 2 DALNQGFPLKQ | 13 EAQYDFIRAGCADE | ESFQOYMCNPAD--DDVAFLYDLIASAEY | 270 |
| A2p02_BPA2_22296524 | AERVFAPFAIEQDN-----VQEVDDPNS | 6 LLDVDTLHQSIS | 3 TTKLAQARA-DGSLN | AKLVKNFNIWQA--TEDSLDFDAWKAEL | 342 |
| Sf121p02_BPSf121_9632939 | ADTYLCLVWSQDN-----LEEVPQPET | 6 LLDLEQERDNL | 3 LDKRDSLAL-SGNLS | DFQVKNMNCWLLA--DSNSFLDLKDIENAVI | 378 |
| sk1p02_BPSk1_9629655 | NDRSFLGLYMLEN-----HEQVKDKKN | 6 LIGSVPKWSGI | -----EYELAQG-DPALQ | NKFLAFNMGLPMQ--DTAYYFTPDQTKLDF | 337 |
| TerL_BPc2_9628694 | DFSHWVFFYALDD-----YDEVKDSK | 2 KANPALGYTSL | -EDIQKDFI-GAIGN | 2 KMAKIIITKRFNLS--MTDSIFSKQLVVKCLV | 325 |
| HK97p02_BPHK97_9634166 | -SIWIDDVAVK-----DPHIVCHVY | 5 ADISKRESWLAA | 13 ARQAEKAGR-MPSFE | NTFRNLNLNQVRS--TVSFFISRSVWELCGE | 289 |
| Consensus/80% | ...b.....p..... | | | ...b..bb....s....hb.....h..... | |

Figure 3. Adenovirus IVA2/ABC ATPase combined alignment

Table showing the alignment of Adenovirus IVA2/ABC ATPase. Columns include Secondary Structure, sequence positions (1-571), conserved motifs (WA, EEEEE, EEEEE, EEEEE, EEEEE), and functional annotations (IVA2 ATPase, Classic ABC transporters, His Perm binding, HB ATP-binding, Tap1, MalK, MutS, etc.).

Table showing the alignment of Adenovirus IVA2/ABC ATPase. Columns include Secondary Structure, sequence positions (1-571), conserved motifs (WB, Sen-1), and functional annotations (IVA2 ATPase, Classic ABC transporters, His Perm binding, HB ATP-binding, Tap1, MalK, MutS, etc.).

SECONDARY STRUCTURE
gp2_BFWocauB1_46804777
HalH1vpg094_HVH_32453918
B_Bp1ambda_9626247
Bcep02005900_Bfun_48782926
42_BPBcepNazqul_34304454
... (alignment continues with various protein sequences and their structural annotations)

Consensus/80%

Figure 5. FIS-type HTH in terminase small subunit alignment

SECONDARY STRUCTURE
Tuc2009_28_BPTuc2009_13487829
PBC5p48_BPPBC5_18071234
D3112p24_BPD3112_38229136
HK022p01_BPHK022_9634120
paca_BBP1_46401729
JL001p58_BPJL001_62327244
phiA1122p14_BpPhiA1122_30387467
orf5_BPPaP2_48697075
ORF8_BPpsIM2_3249593
ORF8_BPpsIM100_11692720
ORF148_BpPhiJL_1_62327096
terS_BPTP901_1_13786561
ul36_36_BpUl36_21716107
gp1_BPES18_62362214
PaP3p01_BPPaP3_27476055
Ljo_0307_BPLJ965_41179239
epsilon15p01_BpEpsilon15_30387380
gp33_BPBcJalC_56694901
BcepB1A_gene45_BPBcepB1A_48697530
Ljo_1440_BPLJ928_41179313
g02_BPBcepC6B_48697192
F116p38_BPF116_56692920
phi11_29_BP11_29028592
gp172_BPSF11_7669463
O1205p25_BP01205_23455900
terS_BPAaphi23_31544027
bIL310p27_BpPhiL310_13095888
A118p01_BPA118_16798784
SPP1p001_BPSP1_22855046
SpyM3_1436_BP315_6_28876466
phiglep38_BpPhigle_23455798
bbp26_BPBPP-1_41179387
BMP-1p25_BPBMP-1_45569519
BIP-1p25_BPBIP-1_45580770
ORF7_BpPhiH3IC_62362370
BC1890_BPphBC6A51_31415784
APSE-1_17_BPAPE-1_9633564
W_BP186_9634057
PSP3p02_BPPsP3_41057354
933wp50_BP933wp_9632516
VT2-Sap51_BPVT-2-Sa_9633445
Stx2Iip165_BPI_20065960
Stx1p165_BPBacteriophage_32170999
Stx2Iip168_BPII_32171171
PHG11b_14_BP11b_53793590
gp3_BPHK620_13559865
Sf6p01_BPSF6_41057279
Consensus/80%

SECONDARY STRUCTURE
Tuc2009_28_BPTuc2009_13487829
PBC5p48_BPPBC5_18071234
D3112p24_BPD3112_38229136
HK022p01_BPHK022_9634120
paca_BBP1_46401729
JL001p58_BPJL001_62327244
phiA1122p14_BpPhiA1122_30387467
orf5_BPPaP2_48697075
ORF8_BPpsIM2_3249593
ORF8_BPpsIM100_11692720
ORF148_BpPhiJL_1_62327096
terS_BPTP901_1_13786561
ul36_36_BpUl36_21716107
gp1_BPES18_62362214
PaP3p01_BPPaP3_27476055
Ljo_0307_BPLJ965_41179239
epsilon15p01_BpEpsilon15_30387380
gp33_BPBcJalC_56694901
BcepB1A_gene45_BPBcepB1A_48697530
Ljo_1440_BPLJ928_41179313
g02_BPBcepC6B_48697192
F116p38_BPF116_56692920
phi11_29_BP11_29028592
gp172_BPSF11_7669463
O1205p25_BP01205_23455900
terS_BPAaphi23_31544027
bIL310p27_BpPhiL310_13095888
A118p01_BPA118_16798784
SPP1p001_BPSP1_22855046
SpyM3_1436_BP315_6_28876466
phiglep38_BpPhigle_23455798
bbp26_BPBPP-1_41179387
BMP-1p25_BPBMP-1_45569519
BIP-1p25_BPBIP-1_45580770
ORF7_BpPhiH3IC_62362370
BC1890_BPphBC6A51_31415784
APSE-1_17_BPAPE-1_9633564
W_BP186_9634057
PSP3p02_BPPsP3_41057354
933wp50_BP933wp_9632516
VT2-Sap51_BPVT-2-Sa_9633445
Stx2Iip165_BPI_20065960
Stx1p165_BPBacteriophage_32170999
Stx2Iip168_BPII_32171171
PHG11b_14_BP11b_53793590
gp3_BPHK620_13559865
Sf6p01_BPSF6_41057279
Consensus/80%

