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. f 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 : *Acyrthosiphon 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 wolfeii* 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; Hduc : *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 : *Photorhabdus 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

	packaging ATPase type					
virus type	terminase	phi-29 like	FtsK	none	other/unknown	average size (kb)
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. Gene neighborhoods of LTS/portal proteins in completely-sequenced phage genomes with portal protein/small terminase family classifications and location of LTS in phages with linear chromosomes

organism	large terminase	portal	small terminase	MuF	Portal Protein Family	small terminase family	linear(l)/circular(c)	LTS linear gene position in genome
Streptococcus thermophilus bacteriophage Sfi19	9632897	9632899	9632896		T5-like	D3-like	l	4/45
Streptococcus thermophilus bacteriophage Sfi21	9632939	9632941	9632938		T5-like	D3-like	l	2/50
Salmonella typhimurium phage ST64B	23505447	23505449	23505446		T5-like	D3-like	c	
Lactococcus phage P335	30089892	30089894	30089891		T5-like	D3-like	l	31/49
Shigella phage SfV	19548992	19548993	19548991		T5-like	D3-like	l	2/53
Bacteriophage D3112	38229138	38229139	38229136	38229140	Mu-like	SPP1-like	l	26/55
Bacteriophage P27	18249900	18249902	18249899		T5-like	D3-like	l	36/58
Bacteriophage bIL286	13095784	13095786	13095783		T5-like	D3-like	l	41/61
Lactococcus phage BK5-T	14251127	14251130	14251126		T5-like	D3-like	c	
Lactobacillus bacteriophage phi adh	9633040	9633041	9633039		T5-like	D3-like	c	
Streptococcus thermophilus bacteriophage 7201	9634649	9634651	9634648		T5-like	D3-like	l	24/46
Bacteriophage phiE125	17975163	17975165	17975162		T5-like	D3-like	l	2/71
Bacteriophage PY54	33770510	33770511	62909765		T5-like	D3-like	l	2/67
Bacteriophage phiKO2	46402088	46402090	46402087		T5-like	D3-like	l	2/64
Bacteriophage phi1026b	38707892	38707893	38707891		T5-like	D3-like	l	2/83
Enterobacteria phage T4	9632591	9632594	9632590		T4-like	T4-like	l	163/278
Enterobacteria phage RB69	32453660	32453663	32453659		T4-like	T4-like	l	179/273
Aeromonas phage 44RR2.8t	37651634	37651637	37651633		T4-like	T4-like	l	153/252
Aeromonas phage Aeh1	38640149	38640152	38640148		T4-like	T4-like	l	227/352
Pseudomonas phage D3	9635588	9635589	9635587		T5-like	D3-like	l	2/95
Bacteriophage phi AT3	48697260	48697262	48697259		T5-like	D3-like	c	
Streptococcus pyogenes phage 315.2	28876230	28876227	28876231		T5-like	D3-like	l	35/60
Bacteriophage bIL309	13095844	13095846	13095843		T5-like	D3-like	l	39/56
Cyanophage P-SSM4	61806303	61806306	61806302		T4-like	solo	c	
Enterobacteria phage RB49	33620526	33620636	33620525		T4-like	T4-like	l	153/274
Bacteriophage phi-BT1	29366753	29366754	29366752		T5-like	D3-like	l	2/55
Bacteriophage phi3626	20065965	20065966	20065964		T5-like	D3-like	l	2/50
Lactobacillus casei bacteriophage A2	22296524	22296525	22296523		T5-like	D3-like	l	2/61
Staphylococcus aureus temperate phage phiSLT	12719430	12719431	12719429		T5-like	D3-like	l	39/61
Bacteriophage S-PM2	58532911	58532915	58532907		T4-like	T4-like	c	
Bacteriophage phi-C31	9631142	9631143	9631140		T5-like	D3-like	l	3/53
Cyanophage P-SSM2	61806000	61806003	61805996		T4-like	T4-like	c	
Mycobacteria phage D29	9630396	9630397		9630398	SPP1-like		l	10/79
Mycobacterium phage L5	9625443	9625444		9625445	SPP1-like		l	10/85
Vibrio phage KVP40	34419588	34419591	34419587		T4-like	T4-like	l	353/381
Bacteriophage phi-105	22854995	22854998	22854994		T5-like	D3-like	c	
Enterobacteria phage T7	9627482	9627467	9627479		T7/T3-like	T7-like	l	57/60
Staphylococcus aureus phage phi 13	29028696	29028698	29028695		T5-like	D3-like	l	29/49
Staphylococcus aureus bacteriophage PVL	9635167	9635169	9635166		T5-like	D3-like	l	2/62
Yersinia pestis phage phiA1122	30387500	30387485	30387497		T7/T3-like	T7-like	l	47/50
Bacteriophage PSP3	41057355,41057354	41057353	41057355%		T5-like	SPP1-like	l	3/42
Streptococcus pyogenes phage 315.1	28876175	28876176	28876174		T5-like	D3-like	l	30/56
Enterobacteria phage 186	9634058	9634056	9634058 %		T5-like	SPP1-like	l	3/46
Enterobacteria phage T3	17570838	17570823	17570835		T7/T3-like	T7-like	l	52/55
Mycobacterium phage Bxz2	29566031	29566032		29566033	SPP1-like		l	10/86
Enterobacteria phage P2	9630329	9630328	9630332 %		T5-like	SPP1-like	l	2/43
Bacteriophage phiYeO3-12	9634048	9634032	9634045		T7/T3-like	T7-like	l	56/59
Bacteriophage L-413C	30065706	30065705	30065706%		T5-like	SPP1-like	l	2/40
Pseudomonas phage gh-1	29366741	29366727	29366739		T7/T3-like	T7-like	l	40/42
Enterobacteria phage HK97	9634166	9634156	9634165		T5-like	solo	l	2/61
Cyanophage P-SSP7	61806451	61806424			T7/T3-like		c	
Mycobacterium phage Bxb1	12657864	12657865		12657866	SPP1-like		l	10/86
Enterobacteria phage HK022	9634121	9634122	9634120		T5-like	SPP1-like	l	2/57
Bacteriophage phi CTX	17313220	17313219	17313220%		T5-like	SPP1-like	c	
Staphylococcus phage phiN315	30043954	30043953	30043955		T5-like	phiN315-like	l	37/65
Bacteriophage 77	41189518	41189523	41189549		T5-like	phiN315-like	l	2/69
Bacteriophage phBC6A52	31415848	31415850	31415847		T5-like	D3-like	l	32/49
Bacteriophage phBC6A51	31415785	31415786	31415784	31415791	SPP1-like	SPP1-like	l	45/75
Streptococcus thermophilus bacteriophage DT1	9632420	9632422	9632419		T5-like	D3-like	l	4/45
Bacteriophage RM 378	30044056	30044028			T4-like		l	66/146
Haemophilus phage HP1	9628620	9628619	9628620 %		T5-like	SPP1-like	l	21/42
Staphylococcus aureus phage phi 12	29028647	29028648	29028646		T5-like	D3-like	l	30/49
Vibrio phage K139	17975126	17975125	17975126%		T5-like	SPP1-like	l	21/44
Pseudomonas aeruginosa bacteriophage PaP2	48697077	48697078	48697075			SPP1-like	l	7/58
Haemophilus phage HP2	17981830	17981829	17981830%		T5-like	SPP1-like	l	16/37
Streptococcus mitis phage SM1	32469469	32469465			SPP1-like		l	38/56
Bacteriophage r1t	23455748	23455746			SPP1-like		l	29/50
Bacteriophage phi LC3	45597419	45597416			SPP1-like		c	
Listeria phage 2389	17488505	17488506	17488504		T5-like	D3-like	l	2/59
Mycobacterium phage TM4	18496890	18496891		18496892	SPP1-like		l	4/89
Virus PhiCh1	22091106	22091110		22091111	T5-like		c	
Streptococcus pyogenes phage 315.3	28876285	28876291			T5-like		l	31/52
Vibriophage VpV262	21234401	21234402			T7/T3-like		c	
Mycobacterium phage Corndog	29566316	29566318			T5-like		l	32/122
Enterobacteria phage SP6	31711683	31711672	31711682?		T7/T3-like	solo	l	41/52
Mycobacterium phage Che9d	29566410	29566411			T7/T3-like		l	3/111
Halorubrum phage HF2	18138498	18138497			T5-like		l	108/114
Bacteriophage sk1	9629655	9629657	9629654		T5-like	bIL170-like	l	2/56
Lactococcus phage c2	9628694	9628686			T5-like?		l	34/39
Pseudomonas phage F116	56692921	56692922	56692920		SPP1-like (P22 subfamily)	SPP1-like	l	39/70
Bacteriophage bIL285	13095721	13095722	13095720		T5-like	D3-like	l	41/62
Bacteriophage bIL170	9630568	9630570	9630567		T5-like	bIL170-like	l	2/64
Enterobacteria phage Mu	9633519	9633520	9633518	9633521	Mu-like	Mu-like	l	28/55
Sinorhizobium meliloti phage PBC5	18071222	18071218	18071234		SPP1-like (P22 subfamily)	SPP1-like	l	46/83
Mycobacterium phage Rosebush	29566200	29566205?		29566206	lambda-like?		l	7/90
Bacteriophage Aaphi23	31544028	31544029	31544027	31544030	T1-like	SPP1-like	l	40/66
Enterobacteria phage T1	45686309	45686310	45686308	45686360	T1-like	T1/P22-like	l	26/78
bacteriophage phiKMV	33300854	33300841			T7/T3-like		l	43/48
Bacteriophage SPBc2	9630181	9630180			lambda-like?		l	56/185
Mycobacterium phage Cjw1	29565887	29565889	29565884		T5-like	D3-like	l	8/141
Burkholderia cepacia phage Bcep22	38640355	38640357	38640354		SPP1-like (P22 subfamily)	SPP1-like	l	48/78
Enterobacteria phage T5	46401884	46401881			T5-like		l	155/162
Mycobacterium phage Che8	29565768	29565769			SPP1-like		l	2/112
Mycobacterium phage Che9c	29566110	29566112			T5-like		l	2/84
Bacteriophage B3	56692599	56692600	56692598	56692601	Mu-like	Mu-like	l	31/59
Burkholderia cepacia phage Bcep781	23752329	23752328		23752327	T1-like		l	18/66
Burkholderia cepacia phage Bcep43	41057669	41057668		41057667	T1-like		l	18/65
Bacteriophage VWB	41057249	41057250	41057219	41057251	SPP1-like	lambda-like	c	
Bacillus clarkii bacteriophage BCJA1c	56694902	56694903	56694901	56694905	phig1e-like	SPP1-like	l	34/58
Lactococcus phage TP901-1	13786562	13786563	13786561	13786564	SPP1-like	SPP1-like	l	31/56
Burkholderia cenocepacia phage Bcep1	38638626	38638625		38638624	T1-like		l	19/71
Mycobacteriophage PG1	38638405	38638411?		38638412	lambda-like?		l	2/100
Lactobacillus plantarum bacteriophage phiJL-1	62327097	62327098	62327096	62327099	SPP1-like	SPP1-like	l	12/46
Mycobacterium phage Omega	29566748,29566740	29566750			T5-like		l	11/237
Cyanophage P60	18640524	18640510			T7/T3-like		l	55/80
Streptococcus pneumoniae bacteriophage MM1	15088772	26553451		15088775	phig1e-like		l	32/53
Burkholderia cenocepacia phage BcepMu	48696938	48696939	48696937	48696940	Mu-like	Mu-like	l	28/53
Bacteriophage phig1e	23455799	23455800	23455798	23455801	phig1e-like	SPP1-like	c	
Xanthomonas oryzae bacteriophage Xp10	32128418	32128419	32128417		T5-like	solo	l	6/60
Bacteriophage A118	16798785	16798786	16798784	16798787	phig1e-like	SPP1-like	l	2/72
Stx1 converting bacteriophage	32171001	32170835	32170999		SPP1-like (P22 subfamily)	SPP1-like	c	
Listonella pelagia phage phiHSIC	62362371	62362375	62362370	62362377	SPP1-like (KS7 subfamily)	SPP1-like	c	
Bacteriophage SPP1	22855048	22855055	22855046	22855056	SPP1-like	SPP1-like	c	
Roseophage SIO1	9964609 ,9964611	9964610, 9964612?			T7/T3-like		l	2/34
Enterobacteria phage Sf6	41057280	41057281	41057279		SPP1-like (P22 subfamily)	SPP1-like	c	
Enterobacteria phage HK620	13559866	13559867	13559865		SPP1-like (P22 subfamily)	SPP1-like	c	
Lactococcus phage u136	21716108	21716109	21716107	21716110	SPP1-like	SPP1-like	l	37/58
Bacteriophage Tuc2009	13487831	13487832	13487829	13487833	SPP1-like	SPP1-like	c	
Enterobacteria phage epsilon15	30387381	30387383	30387380		T7/T3-like	SPP1-like	l	2/49
Bacteriophage 11b	53793591	53793592	53793590	53793594	T5-like	SPP1-like	l	15/65
Bacteriophage Felix 01	38707825,38707714	38707826			Mu-like		l	105/247

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
Acyrthosiphon 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 Sfi11	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	phig1e-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		phig1e-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	20065798	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	82793123		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			SPP1-like		l	2/65
Bacteriophage phi ETA	17426268	17426269		17426270	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 1. φ29-like/FtsK ATPase combined alignment

```
                                                WA
SECONDARY STRUCTURE                 --EEEEE------HHHHHHHHHHHHHHH.....---EEEEE----------     -------EE--HHHHHHHH.....HHHHHHHHHHHHHHHHHHH--
P29_21_BPphi29_9626405           18 ILNFVIGA-RGIGKSYAMKVYPINRFIK.....YGEQFIYVRRY-KPELAK     VSNY-FNDVAQEFPDHELV.....FYIDGKLAGWAIPLSVWQSEK\phi-29 like clade
16_BPphi29_137944                18 ILNFVIGA-RGIGKSYAMKVYPINRFIK.....YGEQFIYVRRY-KPELAK     VSNY-FNDVAQEFPDHELV.....FYIDGKLAGWAIPLSVWQSEK|
B103_16_BPB103_22855164           7 ILNFVIGA-RGIGKSYAMKKHPIKRFIK.....HGEQFIYVRRY-KPELKK     IGNY-FNDIAQEFPNHEFK.....FCIDGKLTGWAIPLSAWQSEK|
gene_16_BPGA-1_12141293          18 ILSCVIGA-RGVGKSFAIKSKLVDDFLK.....NGRQFVYLRMY-KTEISK     NKSKFFDDIAINYPNTKFE.....FYINGKLAGYAIPLSGWQALK|
C1p06_BPC1_31072014              50 ILHYMSITSRGDGKSFNYIGALAWLSYH.....LNFGTMLLVRH-WSLMDK     MAEMVFEIIRTVGMFDIEN.....ITIEGREVFIITNLNNASDLK|
66ORF005_BP66_66395192           42 IDYYISISNRSDGKSFNYVSFFIYLAIK.....LDIKFTLLSRH-YTLRDA     YRDF-IEEIIDENPLFKSK.....IIYQDKEIGVITDLNSATDLK|
phiP68_09_BPphiP68_29565752       5 IDYYISISNRSDGKSFNYVSFFIYLAIK.....LDIKFTLLSRH-YTLRDA     YRDF-IEEIIDENPLFKSK.....IIYQDKEIGVITDLNSATDLK|
orf20_BPCp-1_1181972             18 YLNFVIGG-RGIGKTFALKKYLFKRFIE.....KGEQFIYLRRN-KSELDR     DKDKFFTTELLKQVFTNFE.....IILNGKIVCYLKSLSTWVDLK|
MpP1p03_MVP_9910039              59 SMRFIIKGKKNIGKTYALLER-MKDLAQ.....KGEKFLFLRIS-DSEVKF     -----LANEWTN--DVNNP.....VKQTKEHIGLISHIKNLQKLR/
gI_M13_17426226                   2 AVYFVTGK-LGSGKTLVSVGKIQ-DKIV.....AGCKIATNLD--LRLQNP     QVGR-FAKTPRVLRIPDKP.....--SISDLLAIGRGNDSYD---\SS DNA virus clade
Mp_fs2_9630769                    3 SVYFVTGK-LGSGKSLIAVSRIR-DALM.....RGVPVATNLN--INLKE-     MIGR-DKRNTRLYRLPDKP.....--TVEDIEILGYAISPYD---/
EsV_1_26_ESV_13242498            32 CNAVVIGK-RNTGKSVVIAELLYYLNK-.....-QKVPRACVFS-ATEESN     RFF--CRHIPDSFIFD---.....-EKNVELKLTEIVEQQKRLQL\A32 clade
A392R_PBCV_9631960               16 AIVGVVGR-RGSGKSVIIKDLLYYKR--.....-NALPIGLIMS-GTEAGN     GYF--SSFVPEIFVYDD--.....FNKDGLEKLLERQKKAAKKGN/
Npun6235_Npu_23129935           193 RSNGVFGK-SGTGKSFLTRLLLAGVIRK.....-AVNLIFDMHS-EYGWEA 137 CLESGKNVVIEFGSQSNML.....SYMLVTNMITRRIHEHYVKKA\HerA clade
sll0284_Syn_16331876            196 RSNGVFGK-SGTGKSFLTRLLISGIIHK.....-AVNLMFDMHS-EYGWEA 138 CLEAGKHVVIEFGSQSNLL.....SYMLATNMITRRIHSIYVKKA|
APE2093_Ape_14601838            214 GHIGVFGS-TGSGKTTTLATIACGAAEA.....-LPVVVADWHG-EYQRLV 104 PIGGRGLYIVDTGSIGNVQ.....LRKTYSILLAAYL--------/
SE1414_Sau_27468332             854 PHALIAGA-TGSGKSVCINSILMSLLYK.....-LRLLLIDPKMVELAPYN     LPHLVSPVITDVKAATQSL.....EEMEKRYKLFAQYHVRNITAF\FtsK Clade
TP0999_Tp_15639983              481 PHLLIAGA-TGSGKSVCVNALILSILYH.....-TKLLLIDPKIVELKLYN     IAHLLTPVITEPKRALQAL.....CEMERRYALLEQLECRDIKTY/
RP293_Rp_15604162               154 QHALLFAP-TGSGKGVGFVIPNLLFWT-.....-DSVVVHDIKLENHNLTS 158 FKKVKTTVYVGLTPDNIQ-.....RLQKLMQVFYQQATEFLSRKM\VirD4 clade
virD4_Wol_8885501               157 QHALLFAP-TGSGKGVGFVIPNLLFWT-.....-DSVIVHDIKLENYEITS 158 FKKKKITVYVGLTPDNLT-.....RLRPLMQVFYQQATEFLCRKL/
CagE_Hp_15611559                591 GHTLILGS-TGSGKTVFMSMTLNAMGQF.....KLTMVYMDKDYGAYGNIV 134 LDFSKTIIGVDGSSFLDN-.....NDVSPFICFYLFARIQEAMDG\VirB4 clade
lvhB4_Lpne_19919310             465 GHAAIFGG-NNAGKTTLVNFLDAQMGRF.....GGRSFFIDRDESSKIYIL 116 LNLDFDKVGFDVTYLMDQV.....SVIATPVYLYLLHRMRQCLDG/
Consensus/95%                       ...hl.h...s.GKsh.h..........     .....hh.....     .....hh...........
```

```
                                          WB                        Arg-finger     Sensor-1
SECONDARY STRUCTURE                 .....-------EEEEEE---HHHHHHHHHHHHHHHHHHHHHHHH-------.....EEEEE-------HHHHHHH-.....----EEEEEE-----------
P29_21_BPphi29_9626405              .....SNAYPNVSTIVFDEFIREKDNSNYIPNEVSALLNLMDTVFRNRERVR.....CICLSNAVS-VVNPYFLFFN.....LV--PDVNKRF-NVYDD--AL  186\phi-29 like clade
16_BPphi29_137944                   .....SNAYPNVSTIVFDEFIREKDNSNYIPNEVSALLNLMDTVFRNRERVR.....CICLSNAVS-VVNPYFLFFN.....LV--PDVNKRF-NVYDD--AL  186|
B103_16_BPB103_22855164             .....SNAYPMVTTIIFDEFIRERDNSGYIPNEVDALLNLMDTVFRTRENGR.....CICLSNAVS-IINPYFVYFG.....LV--PDINKRF-NAYKH--IL  175|
gene_16_BPGA-1_12141293             .....SSSYPNVYTIFFDEFVREKDLTGYPPNVVDSLLNVMDTVFRHRVN-D.....VICACNAVT-LVNPYFLYFG.....FY--PNPNQEF-SKNDS--IV  190|
C1p06_BPC1_31072014                 .....SAVLRNYPVVLYDEFLTLGED--YVTNELAKLQTIIKSIDRMGKR-P.....IIYLGNPVN-FDSPILPALN.....IF--YALQNQEINTIQQHGKT  235|
66ORF005_BP66_66395192              .....SNFLKHYPIIIYDEFLALEDD--YLIDEWDKLKTIYESIDRNHGNVD.....MFLLGNAVN-FSSPILSNLN.....IY--NLLQKHKMNTSRLYKNI  228|
phiP68_09_BPphiP68_29565752         .....SNFLKHYPIIIYDEFLALEDD--YLIDEWDKLKTIYESIDRNHGNVD.....MFLLGNAVN-FSSPILSNLN.....IY--NLLQKHKMNTSRLYKNI  191|
orf20_BPCp-1_1181972                .....GSEYDEVMSILYDEVLIVTSKKRYLDNEVEALLNFIFSVFRRRDGCH.....AYLLSNASN-FNNPYFAFLK.....FY--DDNGKRFYNL-KQY-AT  220|
MpP1p03_MVP_9910039                 .....SLQYNDYSHIFFDEFVA-FDEKSYGSSDSNASLTLIMDVQRAKEDIK.....IFCFGNNDI-AVDIFSKYFR.....LS--INEPYQVDNEAKI--TF  235/
gI_M13_17426226                     .....---ENKNGLLVLDECGTWFNT-RSWNDKERQPIIDWFLHARKLGWD-.....IIFLVQDLS---IVDKQARS.....LA--EHVVYCR-RLDRITLPF  156\SS DNA virus clade
Mp_fs2_9630769                      .....---TSKDGLIVLDECGTWFNS-RTWNDKNRQALLDRFLHIRKLGWD-.....VIFIVQNIS---MVDKQARE.....LA--EHVVHCK-RLDRMQIPY  155/
EsV_1_26_ESV_13242498               .....DRDTDLRVAIVLDDMGYNRKV------LTSQILTFIFMNGRHYDIT-.....LIVAIQHVM-QLTPALRSNT.....DY--VICLKEG-NKNVMRNLY  191\A32 clade
A392R_PBCV_9631960                  .....----MSKVFVVLDDLAFDTSI------MKQPVMRYIFMNGRHLNIF-.....LIFSSQYVADLGPPAIRANI.....DI--LLVCREA-IQANRWRLY  166/
Npun6235_Npu_23129935               .....PCDRPTPLMITIEEAHRFLDP----AIVQSTIFGTIARELRKYFVT-.....LLVVDQRPS-GIDNEVMSQI.....GT--RITALLN-DEKDIDAIF  502\HerA clade
sll0284_Syn_16331876                .....VGDRPRQLVITIEEAHRFLDS----AIVHQTIFGTIAREMRKYFVT-.....LLVVDQRPS-GIDNEVMSQI.....GT--RITCLLN-DEKDIEAIF  505|
APE2093_Ape_14601838                .....YSRELGPLLLVVDEAHNMFDG-------EESFPSIMMAESRKFGLY-.....IALATQNPH-LLPLRAVSNT.....NT--KIVHSLR-WWRDLESIA  474/
SE1414_Sau_27468332                 .....YEQRMPKIVIVIDELADLMMM---APQDVEQSIARIAQKARACGIH-.....MLVATQRPSVNVITGLIKA-.....NIPTRIAFMVS-SSVDSRTIL 1034\FtsK Clade
TP0999_Tp_15639983                  .....ATQPLPFIVIIIDEFADLMVA---SGKELETSVARLCAMSRAVGIH-.....LVLATQRPSIDVITGLIKA-.....NIPSRIAFMVS-SKMDSRIIL  665/
RP293_Rp_15604162                   .....LKEEPYGVMFLLDEFPTLGKM---------DTFKAGIAYFRGYRVR-.....LFLIIQDTQQLKGTYEDAGM.....NSTYRITFAAN-NYETANLIS  478\VirD4 clade
virD4_Wol_8885501                   .....SDDEPYGVLFLMDEFPTLGKM---------EQFQTGIAYFRGYRVR-.....LFLIVQDTEQLKGIYEEAGM.....NSTYRITFAAN-NIETANLIS  480/
CagE_Hp_15611559                    .....-----RRFVLDIDEAWKYLGD-----PKVAYFVRDMLKTARKRNAI-.....VRLATQSITDLLACPIADTI.....QCPTKIFLRND-GGNLSDYQR  900\VirB4 clade
lvhB4_Lpne_19919310                 .....-----RLTSFIIAEAWQLFAS-----PFWEKALREWLPTIRKKNGH-.....FIFDTQSPKTITDSPIKHIV.....NLATLIAFPNP-LADRETYME  745/
Consensus/95%                       .....hhb--h.....h....R.....hhh..p.....
```

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

```
                                                                WA
SECONDARY STRUCTURE                     ---------HHHHHHHHH     H----EEEEEE-----HHHHHHHHHHHHHHH------------EEEEE------------HHHHHHHHHHHHHHHHHH---HH----------
EhV1gp47_EHV1_9626782              231  AFLKMVFNTPEFSDATIR     HFRQRATVFLVPRRHGKTWFLVPLIALALATFKG---------IKIGYTA-----------HIRKATEPVFDEIGARLRQ---WFGNSPVDHV---
OsHV1_gp101_OHV1_48696823          187  NYSKTILGLDNMSAEKIN  4  RLKKNISGYIIPRRCGKSSFSGCMIALVMAMCPSAG-------IKCLYTA-----------HKKNQCTDMYSSVEKHVVA---LIKEFNRINEAKF
U60_HHV7_51874288                  177  DYLKYKFDIEFINESSVN     ILKQKASVFLVPRRHGKTWFMIPVICFLLKNLEG---------ISIGYVA-----------HQKHVSHFVMKDVEFKCRR---FFPQKNITC----
17_BPP-SSM4_61806303                52  DEGVIPFDMYNFQEDMVT  1  FHQHRFNIAKLPRQSGKSTIVTAYLLWYVLFNAN---------VNVAILA-----------NKAPTAREMLGRLQLSYEN----LPKWMQQGILGW
17_BPP-SSM2_61806000                49  DEGLVPFNMYDFQEKLIT  1  FHENRFNICKMPRQTGKSTTCISYLLHYAVFNDN---------VNVAVLA-----------NKASTARDLLGRLQLAYEN----LPRWMQQGIISW
yonF_BPSPBc2_9630181                55  VKEYLGITLKLFQCILIY  1  MVHNHYFMYLASRGQGKTWLTSVYCCVQAILFPG---------TKIVIASG----------TKGQ-AREVIEKIDDLRKE----SPNLRREIEDLK
17_BPT4_9632591                    131  DYGVIKVQLRDYQRDMLK  1  MSSKRMTVCNLSRQLGKTTVVAIFLAHFVCFNKD---------KAVGILA-----------HKGSMSAEVLDRTKQAIEL----LPDFLQPGIVEW
Aeh1p227_BPAeh1_38640149           155  DWGVIKVQLRDYQKDMLR  1  MASERMSMHNLPRQLGKTTATAIFLTHFVVFNEA---------KAVGVLA-----------HKGDMSKEVLERTKQSIEL----LPDFLQPGIVEW
VP2p03_BPVP2_48696639               35  KKLPNNFIPRGYQEEAYD 27  PSWRKKAIQVLHRRAGKDIGALHLIAIASQLRVGN--------YKHILPY-----------KTQARDAIWDGIDALGNRFIRNAFPDEIVESINES
T5.155_BPT5_46401884                30  ILAQEGITPNGPQIAIIN  3  DPRHRFVTACVSRRVGKSFIAYTLGFLKLLEPN----------VKVLVVA-----------PNYSLANIGWSQIRGLIKK----YGLQTERENAKD
PhiCh1p03_VP_22091106               57  NYMAEEMEDDPYYLLDEH  6  KFDSGDRILLCHRDGLKTTITLAYLIAGLE-------------YKSGFRG-----------IWAMNNQIQVGKKADTEFWK-MVDRNPWLINLNAP
terL_BPAaphi23_31544028             40  TGFFNILIAQELQKFYQD  2  DGKQPRLMIYAPPRSGKSELFSRRFPAWVFGQNPELQI-----IACSYSA-----------DLASRMNLDVQRIIDDPIYH-SIFPNTALNIKNIA
T1p53_BPT1_45686309                 33  RIWFQLMQAQQFQPNWHH 11  AGRRGNTIFNVTPGSGKTEVFSIHLPVYAMLKCKKVRN-----LNVSFAD-----------SLVKRNSKRVREIISSNEFQ-ELWPCKFGTSKDEE
N15p02_BPN15_9630466                45  AYQEGRWETLPFQRAIMN  3  SDYIRIVNVIKSARVGYSKMLLGVIAYFIEHKQR---------NELLWLP-----------TDGDADNFMKSHVEPTIRD----VPSLLSLAPWYG
TLS_BPlambda_9626245                45  AYQEGRWETLPFQRAIMN  3  SDYIREVNVVKSARVGYSKMLLGVYAYFIEHKQR---------NTLIWLP-----------TDGDAENFMKTHVEPTIRD----IPSLLALAPWYG
JL001p59_BPJL001_62327245           11  GRKVGWAPLPGSQTMVLK     CTSIFELLYHGTRGPGKSDTLLMDFAQFVGRGWGLQ-------YKGIIF------------RQTYPQLEDLITKSKRWFPK--IFPGARYNGSAN-
T7p57_BPT7_9627482                  27  WKALNLPVPTKCQIDMAK  3  NGDNKKFILQAFRGIGKSFITCAFVVWSLWRDPQ---------LKILIV------------SASKERADANSIFIKNIID---LLPFLSELKPRPG
T3p53_BPT3_17570838                 27  WKALNLPVPTKCQIDMAK  3  NGDNKKFILQAFRGIGKSFITCAFVVWTLWRDPQ---------LKILIV------------SASKERADANSIFIKNIID---LLPFLAELKPRPG
L5p13_BPL5_9625443                  72  LDNENMFIPTDEQVRLVL  9  QYIYREGVIRRLKGWGKDPFTAALCLAELCGPVAFSH------FDADGNPVGKPRS-----AAWITVAAVSQDQTKNTFS---LFPVMISKKLKAE
P27p36_BPP27_18249900               70  AFKRMPITLEPWQLFCIC 11  LRRFREAYNEIPRKNGKSAMSAGVGLYMFACDD-EFG------AEVYSGA-----------TTEKQAWEVFRPARLMAQR---TEPLREAFGIE--
D3p02_BPD3_9635588                  71  SASKSKLVFEPWQVFIEA 11  KRRFREAYEEIPRKNGKSARLAARGIYLFAADG-ESG------AEVYSGA-----------TTEKQAFEVFRPAWMMAHK---LENLRNRFGIELS
SPP1p003_BPSPP1_22855048             7  SEKFTPHFLEVWRTVK--     AAQHLKYVLKGGRGSAKSTHIAMWIILLMMMMP----------ITFLVI------------RRVYNTVEQSVFEQLKEAID--MLEVGHLWKVSKS
TerlL_BPSfi11_7523569                4  SKLYTKRQLEVLNYIW--     NHDWFICGLHGAKRASKTVVNNDTFVTELSRVRKIADRLGVDEPIYILAG-----------TSSTAIQN-NVLQELYNKY-----GFEPKYDK---
gp2_BPST64T_24250810                49  THLLYEFTPYSKQREFID  1  GHDYPERCFMAGNQLGKSFTGAAEVAFHLTGRYPGT-------KGYPADG-----------KYGGEWKG-KRFYEPVVFW----VGGETNETVTKT
P22gp02_BPP22_51236724              31  THLLYEFAPYSKQREFID  1  GHDYPERCFMAGNQLGKSFTGAAEVAFHLTGRYPGT-------KGYPADG-----------KYGGEWKG-KRFYEPVVFW----IGGETNETVTKT
phig1ep37_BPphig1e_23455799         14  VIKISDLINPHFKRMW--     TTDKPYIVANGGRGSFKSSVISLKLVTMVKKAIMQHRKA----NVIAVLA-----------NKSDLHDTVYNQIQWALSM----LDMDNEFI----
A118p02_BPA118_16798785              1  MNATIHKQLIDYQIINVI 11  LSKHNHIIAKGGRSSMKSSVISLKLVEKKMANPMSNM------VCLRKVA-----------NTLYKSVYQQIKWALYEMG----VADQFK------
ORF9_BPpsiM100_11692721             51  LNPYIPVNPFHKQIKFL-     LSDEREVLYGGAAGGGKSVALLMGALQYVHYSD----------YAALIL------------RRTYPELSQEGGLIDMAND---WLGGTDAEWNEQ-
TerL_BPpsiM2_3249594                51  LNPYIPVNPFHKQIKFL-     LSDEREVLYGGAAGGGKSVALLMGALQYVHYSD----------YAALIL------------RRTYPELSQEGGLIDMAND---WLGGTDAEWNEQ-
Sf6p02_BPSf6_41057280                1  -MTSINPIFEPF------     IEAHRYKVAKGGRGSGKSWAIARLLVEAARRQ-----------PVRILCA-----------RELQNSISDSVIRLLEDTI---EREGYSAEFEIQ-
TerL_BPHK620_13559866                1  -MTSINPIFEPF------     IEAHRYKVAKGGRGSGKSWAIARLLVEAARRQ-----------PVRILCA-----------RELQNSISDSVIRLLEDTI---EREGYSAEFEIQ-
Orf12_BP186_9634058                137  LEELFFDQSFEYQLQWYR  1  GLAHRIRDILKSRQIGATFYFSREALLRALKTGH---------NQIFL-------------SASKTQAYVFREYIIQFAR---LVDVDL-------
P2p02_BPP2_9630329                 137  LEQIFFEQSFDYQLHWYR  1  GLEHRIRDILKSRQIGATFYFSREALLRALKTGH---------NQIFL-------------SASKTQAYVFREYIIAFAR---LVDVDL-------
SP6p41_BPSP6_31711683               42  NLITGNPDLNRVQADILK  2  FGGNKYRMVEAQRGQAKTTIAAIYAVFRIIHEPH---------KRIMIVS-----------QTAKRAEEIAGWVIKIFRG---LDFLEFMLPDIYA
HalHV1gp095_HVH_32453919            43  DEIEPLTLFRPYQPRIMH  3  YGDAKILNIYKGRRIGVSYIIGICILIEALLKP----------DTFYPIL-----------SKTKGQSNSRISDIKTLIK---NAKIDIP------
HF2p108_BPHF2_18138498              43  DEIEPLTLFRPYQPRIMH  3  YGDAKILNIYKGRRIGVSYIIGICILIEALLKP----------DTFYPIL-----------SKTKGQSNSRISDIKTLIK---NAKIDIP------
pacB_BPP1_46401730                  26  ADVLFGKTPTWQQDEIIE  2  QQDGSWTSVTSGHGTGKSDMTSIIAILFIMFFPG---------ARVILVANKRQQVLDGIFKYIKSNWATAVSRFPWLSK---YFILTETSFFEV-
r1tp29_BPr1t_23455748               31  IYEKSKHECYPWQKNLLK  9  LWTHQKFGYSIPRRNGKTEIVYILELWSLVQG-----------LSILHTA-----------HRISTSHSSYEKLKKYLED---SGYVEGEDFKSIK
TM4_04_BPTM4_18496890               32  TCGAMGLGFDLWQDDLGK  9  LYAADMFAMSIPRQTGKTYLLGALVFALCIKTPN---------TTVIWTA-----------HRTRTAAETFRSMQGLAKR---DKIAPHILNVHTG
Mup28_BPMu_9633519                  37  AQPRNEPVFLGYQRRWF-     EDESQICIAEKSRRTGLTWAEAGRNVMTAAKPKRRGG------RNVFYVG-----------SRQEMALEYIAACALFARA---FNQLAKADVWEQT
BcepMu28_BPBcepMu_48696938           7  RADRAPAVLLPYQQKWC-     ADTSPVKVCEKSRRVGLSWGEAADSALLAASQRG---------MDVWYVG-----------YNKDMA-QEFIRDCADWAK---FYSLAADEIEETE
B3ORF31_BPB3_56692599               10  IPANPDAIFLPYQSRWI-     TDPSRLKLMQKSRQIGLSWSTAYAAGERTAAESA---------RVDQWVS-----------SRDDLQARLFLEDCKMWAG---IMNQAAKDLGEIV
A2p02_BPA2_22296524                 57  PKSGKPQPLAPFQKFIIG 12  IRRFTDVFISMARKNGKSLLISGVILYEFLFGKNPANK-----RQLYTAA-----------NDRKQAGIVFGMVKDRLRA---LMRKDPGIKRMVK
Sfi21p02_BPSfi21_9632939            90  VDTGEPTQLMPWQKFILC 10  GKRFSRAIVSVGRGQGKTYLMAILTAYSYFIESFGLSN-----QDYLVTS-----------IKFKQTNKLLGYIKSMMKQ---IIQKEPFKSLANE
sk1p02_BPsk1_9629655                58  TGNLMKIELLPTQRWWYE 14  VNLINEIFLNLGRGSGKSSLMATRVLNWMILGGQYG-------GESLVIA-----------YDNTQARHVFDQVRNQTEA---SDTLRVYNENKIF
TerL_BPc2_9628694                   49  LPNNERFKIIPYHKAVLT  5  PYQIDEFVVIVGRSNAKSILDVMIALIELFLFPKPN-------SVIALMA-----------TKKDQAEKILMKHFRAMGN---CQGTIINKFKNQF
HK97p02_BPHK97_9634166              19  GKLIGQPMRLPFQKDFIL  5  PAGTDMAILSIARKNGKTGLIAGILLAHLVGPEAVQN------TQIVSGA-----------LSREQAAIVFNLAVKMVNL---NPKLQEIVHITPS
Consensus/80%                           ...............bp..hh...........b.hh..s+..GKo.hh.h.hhh.hh...............hhhh............p..................................


                                                                                             WB                                                   Sen-1
SECONDARY STRUCTURE                   ----EEEEEE----------EEEEEE-------------------E---EEEEHHHHHHHH----------HHHHHHHHHHH-----------------EEEEEE----       ----
EhV1gp47_EHV1_9626782                 KGENISFSFPDGS------KSTIVFASS--------HNTNGIRGQDFN---LLFVDEANFIRP----------EAVQTIIGFLNQ--------------TNCKIIFVSSTNT   1  KAST
OsHV1_gp101_OHV1_48696823          27 KDGALVVSFYKFTEAGRNIKTTDIPMAVNEFRSIVYKQKDTHRGATYN---LMFVDETNFLQP----------AIFNELFPMLNT--------------DKAKMICTSSQKN   1  QDAK
U60_HHV7_51874288                     QDNVITIEHETI-------KSTALFASC--------YNTHSIRGQSFN---LLIVDESHFIKK----------DAFSTILGFLPQ--------------SSTKIIFISSTNS   1  NHST
17_BPP-SSM4_61806303                  NKGSLELENGS----------KILASST----------SASAVRGMSFN---IIFLDEFAFVPNHIAEQF----FASVYPTISSGK--------------STKVIIISTPHG      ---M
17_BPP-SSM2_61806000                  NKGSLELENGS----------KISANST----------SSSAVRGGSYN---VIFLDEFAFIPNHIADDF----FASVYPTITSGQ--------------STKVIIVSTPRG      ---M
yonF_BPSPBc2_9630181               3  NDAKVEFHNGS----------WIKIVAS----------NDGARSKRAN---LLIVDEFRMV-DFEII------SKVLRKFLTAPRSPKYLEKEEYAHLKERNKEIYLSSCWY   5  FNRF
17_BPT4_9632591                       NKGSIELDNGS----------SIGAYAS----------SPDAVRGNSFA---MIYIDECAFIPNFHDS------WLAIQPVISSGR---------------RSKIIITTPNG      LNHF
Aeh1p227_BPAeh1_38640149              NKGNIELENGC----------SIGAYAS----------SPDAVRGNSFA---LIYVDECAFIEGFEDT------WKAILPVISSGR---------------QSRIILTSTPNG      INHW
VP2p03_BPVP2_48696639              2  LVRFTNGST------------YQLQGGD----------SDKLVGAGPV---GIVYSESAL-------------MSP-NVRTF-LR----------PMLDETGGWELHITTPR      --GK
T5.155_BPT5_46401884               1  EIELANGSL------------FKLASAA----------QADSAVGRSYD---FIIFDEAAISDV----------GGD-AFRVQ-LR----------PTLDKPNSKALFISTPR      --GG
PhiCh1p03_VP_22091106              1  EKEAVKAKVFAN---------GSILNAG---------WLGGGIEGDRAH---LLILDDIIKEKGDG-DT-----EDVLDWIEAVCV-PMVKDHGRTVVIGTRKRPDDIYTHFR      --TL
terL_BPAaphi23_31544028            2  SGKPLRNSEIFEIVGH-----LGAYRSA---------GVGGGITGMGAD---IAIIDDPVKDAKEANSQTVR--DSIWDWYTTTLY-TR-LSPKSGVLLGMTRWHEDDLAGRL   2  EAEN
T1p53_BPT1_45686309                6  DGKVWFELISAA---------AGGRITG---------SRGGYMTPGFSG---MVMLDDIDKPDDMFSKVKR---ERTHMLLKNTIR-SRRMHNETPIIAIQQRLHAQDSTWFM      MNGG
N15p02_BPN15_9630466               4  DNTLSMKRFTNG---------RGFWCLG-------GKAAKNYREKSVD---VVGYDELAAFDADI--------EKEGSPTFLGDK---------RIEGSVWPKSIRGSTPKL  12  SGHF
TLS_BPlambda_9626245               4  DNTLTMKRFTNG---------RGFWCLG-------GKAAKNYREKSVD---VAGYDELAAFDDDI--------EQEGSPTFLGDK---------RIEGSVWPKSIRGSTPKV  12  SPHF
JL001p59_BPJL001_62327245             KWVF-PDGEELL---------FRQFNKP---------IDYWKYHGQEFQ---FIGWDELTNWPT----------LEGYTSMMSCLR-----------TTHPDIPLRCRATTNP      -YGP
T7p57_BPT7_9627482                 1  RDSVISFDVGPAN-----PDHSPSVKSV---------GITGQLTGSRAD---IIIADDVEIPSNSATMGAREKLWTLVQEFAALLK------------PLPSSRVIYLGTPQT  11  RGYT
T3p53_BPT3_17570838                1  RDSVISFDVGPAK-----PDHSPSVKSV---------GITGQLTGSRAD---IIIADDVEIPSNSATQGAREKLWTLVQEFAALLK------------PLPTSRVIYLGTPQT  11  RGYT
L5p13_BPL5_9625443                 4  VNRFIIYSAA---------GGRIEAATS----------SPASMEGNRPT---FVVQNETQWWGQGPDGKVNEGHAMAEVIEGNMTK----------------VEGSRTLSICN   6  ETVA
P27p36_BPP27_18249900                 VHAQSMSRPED----------GARFEPL----------IGDPGDGASPS---CALIDEYHEHQTDA--------LYQTMQTGMGAR--------------RQPLIFGITTAGY   6  YDKR
D3p02_BPD3_9635588                 3  KNPGPMFVMED----------MSKFETV----------IGNPGDGASPH---AALVDEYHEHDTDA--------LVDTMQTGMGAR--------------EQPLLSIITTAGS   6  HEKR
SPP1p003_BPSPP1_22855048           2  RLTYIPRGNSII---------FRGGDDV------QKIKSIKASKFPVA---GMWIEELAEFKTEE---------EVSVIEKSV-LR----------AELPPGCRYIFFYSYNP      -PKR
TerlL_BPSfi11_7523569                 HGSFVFCGVKVV---------QVYTGSI---------SGLKRARGFTAF---GAYVNEASLANEFV-----------------FK---------EIISRCSGDGARVVWDSNP      -DNP
gp2_BPST64T_24250810                  TQRILCGRIEEN---------DEPGYGS-----IPKEDIISWKKSPFF---PNLVDHLLVKHHTPE-------GVEDGISICYFK-----------PYSQGRARWQGDTIHG   4  EEPP
P22gp02_BPP22_51236724                TQRILCGRIEEN---------DEPGYGS-----IPKEDIISWKKSPFF---PNLVDHLLVKHHTAD-------GVEDGISICYFK-----------PYSQGRARWQGDTIHG   4  EEPP
phig1ep37_BPphig1e_23455799           AYKSPLTIQ------------HKRTGSS--FYFYGADNPYKLKSNIVGDVVAVWYEEAANMKSS---------DVFDQANPTFIR----------QKPEWLDQVKVFYSYNP      -PKN
A118p02_BPA118_16798785               FGKSPMEIVHKT-------WGTGFYFSG-------CDDPAKLKSMKIPVGYVLWFEELAEFSGVT--------DIDVVE-DTFIR----------EDLPQGQEVTIYMSFNP      -PRN
ORF9_BPpsiM100_11692721               KKRWTFPSGAA----------LQFGHME------HEKDRYRYQGSSYH---YIAFDELTEFLE----------SQYRFMFRSLRK----------EADDPIPLRFRATSNPG      --GI
TerL_BPpsiM2_3249594                  KKRWTFPSGAA----------LQFGHME------HEKDRYRYQGSSYH---YIAFDELTEFME----------TQYRFMFRSLRK----------EVNDHIPLRVRATSNPG      --GI
Sf6p02_BPSf6_41057280                 RSMIRHLGTNA----------EFMFYGI------KNNPTKIKSLEGID---ICWVEEAEAV-----------------------TK--------------ESWDILIPTIRK      --PF
TerL_BPHK620_13559866                 RSMIRHLGTNA----------EFMFYGI------KNNPTKIKSLEGID---ICWVEEAEAV-----------------------TK--------------ESWDILIPTIRK      --PF
Orf12_BP186_9634058                   TGDPIVIGNNGA---------KLIFLGT---------NSNTAQSHNGD----LYVDEIFWIPNFQK-------LRKVASGMASQK--------------HLRSTYFSTPST   1  AHGA
P2p02_BPP2_9630329                    TGDPIVLGNNGA---------KLIFLGT---------NSNTAQSHNGD----LYVDEIFWIPNFQV-------LRKVASGMASQS--------------HLRSTYFSTPST   1  AHDA
SP6p41_BPSP6_31711683              9  EIHYTLRGSDKS---------PSVACYS---------IEAGMQGARAD---IILADDVESLQNSRT-------AAGRALLEDLTK-----EFESINQFGDIIYLGTPQSVNS   7  RGYQ
HalHV1gp095_HVH_32453919              LEKDNQDEIVLPN------GSRIKAYTG---------DPDSARGEDPPK--TVFIDEMAFLEDQS--------ATLDAYLPTISL--------------GSSQMVQVSTPKA      QNDE
HF2p108_BPHF2_18138498                LEKDNQDEIVLPN------GSRIKAYTG---------DPDSARGEDPPK--TVFIDEMAFLEDQS--------ATLDAYLPTISL--------------GSSQMVQVSTPKA      QNDE
pacB_BPP1_46401730                    TGKGVWTILI-----------KSCRPGN--------EEALAGEHADHLL---YIIDEASGVSDK---------AFSVITGALTGK-------------DNRILLLSQPTRPS      -GYF
r1tp29_BPr1t_23455748              2  GQERLELIESGG---------VIQFRTR---------TSSGGLGEGFD---ILVIDEAQEYTTEQ--------ESALKYTVTDSD--------------NPMTIMCGTPPT   4  GTVF
TM4_04_BPTM4_18496890              1  GKEAVLFKNGS----------RILFGAR---------ERGFGRGFAGVD--VLIFDEAQILTENA--------MDDMVPATNAAP--------------NPLILLAGTPPK   4  GEVF
Mup28_BPMu_9633519                 6  KEEILTYMIRFPN-----SGFKIQALSS---------RPSNLRGLQGD----VVIDEAAFHEALD--------ELLKAAFALNMW--------------GASVRIISTHNG      VDNL
BcepMu28_BPBcepMu_48696938         8  DKSILAFVIRFAS------GFRVTALSS---------RPSNLRGKQGR----VIIDEAAFHEQLG--------ELLKAAMALLMW--------------GGQVHIISTHDG      VDNA
B3ORF31_BPB3_56692599              3  KNKISAYVLEFAN------GRRIHSMSS---------NPDAQAGKRGG----RILDEFALHPDPR--------KLWSIAYPGITW--------------GGAMEIISTHRG      SQNF
A2p02_BPA2_22296524                2  RDELVNLDDGS----------TIRSFSR---------DTGLVDGYEPH---VAVVDEYANAKTTDM-------IETLASGQVLLP--------------SYLTFIISTAGF  19  GEEK
Sfi21p02_BPSfi21_9632939           6  SDQVIMKANNN----------ILRAISA---------ESGQYDSFHFT---TAIFDEMGEIETRNA-------VSKIVSGQVKVP--------------NRQFIQISTAYP  20  DNRE
sk1p02_BPsk1_9629655               3  KQGLEFASFKT----------TFKKQTN---------DTLRAQGGNSS---LNIFDEVHTYGED---------ITESVNKGSRQK-------------QDNWQSIYITSGGL  16  EEFY
TerL_BPc2_9628694                  4  EQILVKDNSILKS-----KGTEISIYAS---------NEDTLDGGREQ---LVIIDEFGAFKKNP--------LITIRQGLRKNK----------------GTLFISTTNNV  16  VKDD
HK97p02_BPHK97_9634166                GKKLIGLPCNV----------EYKALSA---------EGKTTHGLSPI---LAILDETGQVRGPQDD------FIDAITTAQGAH--------------ENPLLIVISTQAA   2  ADLL
Consensus/80%                       ..........................h.ss.............hpt.......hhhhDEh..h.......................b................b...h.*...........
```

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

```
SECONDARY STRUCTURE          ---------------------EEEEEEE--     EEEEE-------     -----H---------     HHHHHHH--------------------EE--
EhV1gp47_EHV1_9626782        SFLYNLKGAADDLL-------NVVTYICDE  22 VFITMDGAMRNT  1  ELFLPD---------     SFMQEII----GG--GNVSGAHRDEPVFTKT   483
OsHV1_gp101_OHV1_48696823    PFVDIRNTRQDGTTT------CVVEYVCPN  24 LHINAGAGVRKI  1  AAFSVKTNKNMEVDD  18 LTKNDIL----GM----NN--LQNMRLTSEL   515
U60_HHV7_51874288            SFLTKLSNSPFEML-------TVVSYVCED  22 KFISINADVKKT  1  DLFLEG---------     AFKHEIM----GG--SLCN--VVNDTLITEQ   425
17_BPP-SSM4_61806303         NQFYKLWHDAE----------RGANNYVAT     EVHWSQVPGRDD     KWKQQTIEN-TSE--     AQFRVEFECEFLG--SVDTLITPSKLRIMPY   287
17_BPP-SSM2_61806000         NHFYRMWHDSE----------KGKSEYVAT     DVHWSEVPGRDE     EWKEQTIAN-TSE--     QQFKIEFECEFLG--SVNTLINPAKLRNLVY   284
yonF_BPSPBc2_9630181         ITYYNAMMKGS----------KYFVCGLPY     QIAIKEGLLDKD     QVRDEMAEE-DFDP-     IGWSMEMEALWFG--ESEKAYFKFEDIEKNR   313
17_BPT4_9632591              YDIWTAAVEGK----------SGFEPYTAI     WNSVKERLYNDE     DIFDDGWQWSIQTIN  4  AQFRQEHTAAFEG--TSGTLISGMKLAVMDF   374
Aeh1p227_BPAeh1_38640149     YDLWEVSLKSD----------KGFKPYTTT     WITVKERLYDGS     DAYDDGFEWASKQIN  4  EAFQQEHLCRFMG--TSGTLINGFKLSKMTW   398
VP2p03_BPVP2_48696639        NWFYKLAMHAE----------KSEEWYYKY  1  TINDTWRWAYSS  40 EKAIERGVV-KGMYA  19 FIVRQEYYCDWDV--ALQGSYYGDLMITMYN   357
T5.155_BPT5_46401884         NWFKEFYAYGFDD--------TLPNWVSIH  1  TYRDNPRADLND     IEEARRTVS-K----     NYFRQEYEADFSV--FEGQIFDTFNAIDHVK   264
PhiCh1p03_VP_22091106        EGYEFDEYPAILDYWD-----QQFSADDDY  18 TGETLQVLWPEA     RGPRWLADK-RSKMA  1  HRFWREYSLVIMG--SSGDLIDAKDVRVPAE   338
terL_BPAaphi23_31544028      GGDQWRIVKFPAIAE------EDEEFRKEG     -EPLHPERFDLE     ----RLNKI-RQAVG  1  QAWNALYQQRPSN--KGGGIIKGSWFGRYKV   313
T1p53_BPT1_45686309          MGIEFDQISIPALVT------EEYGKTLPD  18 DGVKHYSFWPSK     ESVHDLLAL-READQ     YTFDSQYQQKPIA--LGGSVFNSEWWTYYGS   335
N15p02_BPN15_9630466         MRFHVACPHCG----------EEQYLKFGD     RDTPFGFKWEPE     QAETVYYLCEHNACV  6  DFSNARYICELTGIWTRDGLRWFSSSNAEID   317
TLS_BPlambda_9626245         MRFHVACPHCG----------EEQYLKFGD     KETPFGLKWTPD     DPSSVFYLCEHNACV  6  DFTDARYICEKTGIWTRDGILWFSSSGEEIE   317
JL001p59_BPJL001_62327245    GHNVVKFHFELPAKD------GIIRRKTYE     FEDPETGEPIRK  1  LCRVAIKGTIWENTI     LLKAQPNYIAQLKEQADNPAKLAAWLEGSWD   258
T7p57_BPT7_9627482           TIIWPALYPRTR---------EENLYYSQR  1  APMLRAEYDENP  2  LAGTPTDPV-RFDRD     DLRERELEYGKAG--FTLQFMLNPNLSDAEK   298
T3p53_BPT3_17570838          TIIWPALYPRSR---------EEDLYYGER  1  APMLREEFNDGF  2  LQGQPTDPV-RFDME     DLRERELEYGKAG--FTLQFMLNPNLSDAEK   298
L5p13_BPL5_9625443           EKAWDEYQKVQAGDSVDTGMMYDALEAPAD  19 IARGDSTWLPID     -DIIKSILS-TKNPI     TESRRKFLNQVNA--AEDSWLSPQEWNRCQV   383
P27p36_BPP27_18249900        REVIEMLEGIVPND-------ELFGIIYTI  3  DDWTDPAVLRKA  13 LSQQRAAIN-NPRKV     GVFKTKHLNIWVA--AKDAFFNLVNWQKCED   343
D3p02_BPD3_9635588           RDVIRILEGQTIDE-------TIFGIIYTI  3  DPWDDPASLIKA  13 LAQLQQAKR-SASKQ     NAFRTKHLNQWVG--ARTVWMNMLAWQRQKR   349
SPP1p003_BPSPP1_22855048     KQSWVNKVFNSS---------FLPANTFVD     HSTYLQNPFLSK  1  FIEEAEEVK-RR-NE     LKYRHEYLGEALG--SGVVPFENLQIEEGII   250
TerlL_BPSfi11_7523569        N-HWLNRDYIG----------KNDGKIIDF     SFKLDDNTFLSK  1  YIDSIKAVT-PK---     GKFYDRDILGHWT--VAEGAIYADYDSKIHV   236
gp2_BPST64T_24250810         YSIYGEGLTRTN---------KYGQFSILT  1  TPLMGMSDVVTK     FLKNPSKSQ-KVVNM     TIYDAEHYTDEQKEQIIASYPEHEREARARG   304
P22gp02_BPP22_51236724       YSIYGEGLTRTN---------KYGQFSILT  1  TPLMGMSDVVTK     FLKNPSKSQ-KVVNM     TIYDAEHYTDEQKEQIIASYPEHEREARARG   286
phig1ep37_BPphig1e_23455799  PYDWINEWIDKVSK-------DDNYLIDTS     DYRCDVRGFTSK  1  TLDLIEQYK-KN-DY     EYYRWLYLGEVIG--LGTSIYNPSLLKPLEV   262
A118p02_BPA118_16798785      PYEWVNEYVDSKRS-------DDDYLIHHT     TYLDDEKGFLSK  1  IIKKIEKYK-KN-DL     DYYRWMYLGEVIG--LGDNVYNMNLFQPLKA   261
ORF9_BPpsiM100_11692721      GHEWVKTRFITG---------EKTFIPSTW     RENPYLNRDEYE  2  LNMLDHVTR-RQ---     -LKEGDWDVSIQG------GVFRREWFEIID   282
TerL_BPpsiM2_3249594         GHEWVKTRFITG---------EKTFIPSTW     RENPYLNRDEYE  2  LNMLDHVTR-RQ---     -LKDGDWDVTLQG------GVFKREWFEVID   282
Sf6p02_BPSf6_41057280        SEIWVSFNPKNIL--------DDTYQRFVV     NPPDDICLLTVN  10 LRLEMEECK-RR-NP     TLYRHIWLGEPVS--ASDMAIIKREWLEAAT   225
TerL_BPHK620_13559866        SEIWVSFNPKNIL--------DDTYQRFVV     NPPDDICLLTVN  10 LRLEMEECK-RR-NP     TLYRHIWLGEPVS--ASDMAIIKREWLEAAT   225
Orf12_BP186_9634058          YPFWSGELFNKGRS-------RIADRIEID  2  HRALAGGQLCDD  17 TLFDLDQLK-RENSD     EDFKNLFMCEFVD--DKASVFPFEELQRCMV   389
P2p02_BPP2_9630329           YPFWSGELFNRGRA-------SAAERVEID  2  HNALAGGLLCAD  17 TLFDIEQLK-RENSA     DDFKNLFMCEFVD--DKASVFPFEELQRCMV   389
SP6p41_BPSP6_31711683        IRIWPGRYPTLEQ--------EACYGDFLA  1  MIRQDMIDDPSL  7  GTQGAPTCPEMYDDE     KLIEKEISQGTAK--FQLQFMLNTRLMDADR   319
HalHV1gp095_HVH_32453919     FMDANERGTPDGRN-------DFGILALKQ  2  FKNADEIQTDVS  11 DFDLMAAETQRASDP     NGFAQEYLCRPVS--DEYRFFSMPTIEDAMG   298
HF2p108_BPHF2_18138498       FMDANERGTPDGRN-------DFGILALKQ  2  FKNADEIQTDVS  11 DFDLMAAETQRASDP     NGFAQEYLCRPVS--DEYRFFSMPTIEDAMG   298
pacB_BPP1_46401730           YDSHHRLAIRPGN--------PDGLFTAII  1  NSEESPLV-DAK     -FIRAKLAEYGGRDN     PMYMIKVRGEFPK--SQDGFLLGRDEVERAT   276
r1tp29_BPr1t_23455748        TNYRDNTIAGKA---------KYSGWAEWS  1  EDVKDIHDVEAW  8  YHLNERKIE-AELGE  2  LDHNVQRLGYWPK--YNQKSVISEQEWNALK   293
TM4_04_BPTM4_18496890        TVMRLDALAGDV---------DDVGYVEIS  1  DEDADPDDRSQW  11 SARAILRMR-KALGD     ESFKREAMGIWPK--VSVHQPVVKSGRWHDL   296
Mup28_BPMu_9633519           FNQYIQD-AREGR--------KDYSVHRIT  2  DAIADGLYRRIC  15 RDGLYRNAPNKESAD     EEYGCIPKKSGGA--YLSRVLIEAAMTPARD   304
BcepMu28_BPBcepMu_48696938   FNELVTD-VRSGK--------KPYSLHRIT  2  DAVQDGLYQRIC  13 KWVKDIRASYGADAE     EELDCVPKNSGGA--WLSRALIESRMSADTP   269
B3ORF31_BPB3_56692599        FNQLVREIVEGGNP-------KNISLHTVT  2  DALNQGFLFKLQ  13 EAQYFDFIRAGCADE     ESFQQEYMCNPAD--DDVAFLEYDLIASAEY   270
A2p02_BPA2_22296524          AERYFAFIAEQDN--------VQEVDDPNS  6  LLDVDTLHSQIS  3  TTKLAQARA-DGSLN     AKLVKNFNIWRQA--TEDSYLDFDAWKAAEL   342
Sfi21p02_BPSfi21_9632939     ADTYLCLVWSQDN--------LEEVFQPET  6  LLDLEQERDNLM  3  LDKRDSDLL-SGNLS     DFQVKNMNCWLLA--DSNSFLDLKDIENAVI   378
sk1p02_BPsk1_9629655         NDRSFGLLYMLEN--------HEQVKDKKN  6  LIGSVPKWSGVI     -EEYELAQG-DPALQ     NKFLAFNMGLPMQ--DTAYYFTPQDTKLTDF   337
TerL_BPc2_9628694            DFSHWVFYYALDD--------YDEVKDSSK  2  KANPALGYTLSL     -EDIQKDFI-GAIGN  2  KMAKIITKRFNLS--MTDSIFSKQLVDKCLV   325
HK97p02_BPHK97_9634166       -SIWIDDAVKSK---------DPHIVCHVY  5  ADISKRESWLAA  13 ARQAEKAGR-MPSFE     NTFRNLNLNQRVS--TVSPFISRSVWELCGE   289
Consensus/80%                ...b................p.....................................................b..bb.....s......hb.......h........
```

Figure 3. Adenovirus IVA2/ABC ATPase combined alignment

```
                                                                    WA
Secondary Structure             -HHHHHEEEEEEEEE--    -EEEEE---EE-----EEEEEEE-----HHHHHHHHHH------------------------EEEEE----    -------EEEEEEHHHHHH-HHHHHHHHHHH    HHHH--------
IVa2_CAVt1_56158852         132 DIQDSITQCCIRASKYL    NKEGNCTSLNYYKQP-FIAAVYGPTGCGKSQLLRNLMSAQLI---------------VPTPETVFFITPQV    DMIPPQEIAAWETQICEGN-FLAGPENTVVP    QSGSIMPRF---\IVA2 ATPase
orf6_CAVt1_1526581          188 DIQDSITQCCIRASKYL    NKEGNCTSLNYYKQP-FIAAVYGPTGCGKSQLLRNLMSAQLI---------------VPTPETVFFITPQV    DMIPPQEIAAWETQICEGN-FLAGPENTVVP    QSGSIMPRF---|
TSAdVgp05_TsAV_51556698     132 EDNKNLIEKLNSTAPFL    KDDGSCKSLNYNTQP-VIGVIYGPTGCGKSQLLRNLISSHLL---------------SPAPETVFFIAPQT    DMIPPQEMTAWETQICEGN-YTPGPEGTIVP    LTGTLKPNF---|
IVa2_BAV3_56158806          131 DYNQAIRDAMNATMPYL    RPDGSCKSLNFTMQP-VIGVVYGPTGCGKSQLLRNLMSTGLI---------------QPPPETVFFIVPQV    DMIPPQELLAWETQICEGN-YTPGPTGTLVP    QSSTLKPAF---|
BAdVBgp06_BAVB_9630096       59 DYNQAIRDAMNATMPYL    RPDGSCKSLNFTMQP-VIGVVYGPTGCGKSQLLRNLMSTGLI---------------QPPPETVFFIVPQV    DMIPPQELLAWETQICEGN-YTPGPTGTLVP    QSSTLKPAF---|
BAdVAgp05_BAVA_52801684     132 EENVCVRELLNKCVPYL    NPQGACRSINFQMQP-VIGVVYGPTGCGKSQLLRNIISSQLI---------------TPPPETVFFIVPQI    DMIPPQEISAWETQICEGN-YSMGPQGTLVP    QSGSLMPKF---|
IVa2_BAV2_56158831          132 RDNVVIRDMLNKCSPYL    TQQGSCKSINFQMQP-VIGVVYGPTGCGKSQLLRNIISTQLL---------------TPPPETVFFIVPQI    DMIPPQEMSAWETQICEGN-YVIGEQGTLIP    QSGSLLPEF---|
IVa2_PAV5_56160569          131 LDSLEVKRMLNACSTFL    NRSGCCKSLNYQTQP-VIGVIYGPTGCGKSQLLRNIISTQLL---------------NPPPETVFFIVPQI    DMIPPQEISAWETQICEGN-YKMGPQGTLVP    QSGSLLPKF---|
PAdVCgp06_PAVC_13487884      59 LDSLEVKRMLNACSTFL    NRSGCCKSLNYQTQP-VIGVIYGPTGCGKSQLLRNIISTQLL---------------NPPPETVFFIVPQI    DMIPPQEISAWETQICEGN-YKMGPQGTLVP    QSGSLLPKF---|
PAdVBgp05_PAVA_46852407     132 QENRNLRDTLNACMPFL    TPAGTCRSLNYQVQP-VIAVIYGPTGSGKSQLLRNLISSQLL---------------SPPPETVFFITPQV    DMIPPQEVVAWETQICEGN-YRSGPDHTLIP    QTATLMPDF---|
IVa2_SAVA_51518820          133 RENLHVRELLNSVAPLL    KDDGSCSSLNYTLQP-VIAVIYGPTGCGKSQLLRNLLSTQLI---------------TPAPETIFFIAPQV    DMIPPSEIKAWEMQICEGN-YAPGPDGTIIP    QSGTLCPRF---|
IVa2_SAV1_61602088          132 SENKHVRSMMNEVAPLL    RPDGSCSSLNYQLQP-VIGVIYGPTGCGKSQLLRNLLSTQLI---------------NPPPETVFFIAPQV    DMIPPSEIKAWEMQICEGN-YAPGPDGTIIP    QSGTLLPRF---|
IVa2_HAVt1_33330444         136 RENMRVRDMLNEVAPLL    RDDGSCSSLNYQLQP-VIGVIYGPTGCGKSQLLRNLLSSQLI---------------SPTPETVFFIAPQV    DMIPPSELKAWEMQICEGN-YAPGPDGTIIP    QSGTLRPRF---|
IVa2_HAVt1_56160441         132 KQNIRVRQMMNEVAPLL    HEDGSCTSLNYHLQP-VIGVIYGPTGCGKSQLLRNLLSAQLV---------------TPAPETVFFIVPQV    DMIPPSEIKSWEMQICEGN-YVPGPEGTIIP    QSGTLCPKF---|
IVa2_SAV2_17105046          133 RENMHVREMMNEVAPLL    REDGSCLSLNYHLQP-VIGVIYGPTGCGKSQLLRNLLSAQLI---------------SPAPETVFFIAPQV    DMIPPSELKAWEMQICEGN-YAPGIEGTFVP    QSGTLRPKF---|
IVa2_FAV2_78068004           54 PRNTTSRGTARTGRQSV    PRRRTTTVAQHGQTA-RHRRHLRPDGIGQEPSAQSVDLLRYA---------------AADPGNGYLCDPGR    RIDPTGTKTAWNLQLLESN-YDCKEDGTFAP    RTCTFRPEF---|
FAdVAgp08_FAVA_9628843      114 EAHEEVKAQEQMDSRYL    Q-HGQLPSINMGKQP-IIGVIYGPTGSGKSHLLRALISCNML---------------DPIPETVIFITPEK    NMIPPIEQTSWNLQLVEAN-FDCREDGTIAP    KTSTFRPEF---|
IVa2_FAV1_56160741           65 EAHEEVKAQEQMDSRYL    Q-HGQLPSINMGKQP-IIGVIYGPTGSGKSHLLRALISCNML---------------DPIPETVIFITPEK    NMIPPIEQTSWNLQLVEAN-FDCREDGTIAP    KTSTFRPEF---|
IVa2_FAV8_78068044           67 EVQRAVERQEREDSRFL    V-DGQLPSLNMGRQP-VIGVIYGPTGSGKSHLLRALISCDML---------------QPIPETVIFVTPEK    NMIPPVEQTAWNLQLLESN-YSCRQDGTFAP    KTCTFRPDF---|
IVa2_TAVA_9630780            49 SNMRKYNEINNINKMYL    TKNSELPSINMNSDP-LISLVIGPTGCGKSQLIRNLLGFKKI---------------QPMPETIIFITPTK    GTISYDEVILWKTQLQEGN-YSAQ-DNTLYP    TTKVFTINF---|
FaV1gp03_FAV_9635798         49 EYMNHFYGIQKLNQKFL    SANGELSSLNYGLNP-FIGIISGPTGTGKSQLIRNLISNKLI---------------SPPPETVIFITPTK    GMLSPEETNLWKLQLQEGN-YEKNTDG-ICP    ITSVFSVDF---|
IVa2_DAVA_9629470            88 AAAEKHHKAVSEAHLYL    K-NDTIPTLNNGVQP-FIVTVYGPTGSGKSQFIRNILAGQLI---------------EPIPHTVFFITPEA    GTVSVEERLAWEAQCVEGN-Y----NTDIAP    LSGTFKPKF---|
IVa2_OAV7_7444361           131 NIAVNYQKTVNEAHNLL    K-DGCIPSLNCGLQP-FIVTVYGPTGSGKSQFIRNIISGKLI---------------DPIPETIFFITPEK    GTVPLEEKIAWKAQCVEGN-Y----DAFSNP    LTKTLNPKF---|
IVa2_MAV1_56160678          121 HDYEILCKSLNSAAPLL    SPEGNCESLNYSLQP-LIAIVYGPTGSGKSQLLRNLLSCHLI---------------DPSPETVFFVVPQI    DMIPPQEMSAWNAQLVEGN-YTCGPQQTIVP    KSGTLKPRL---/
ABCATPase_Tmar_23200171       1 MVSDIVLEVQSLHVYYG    AIHAIKGIDLKVPRG-QIVTLIGANGAGKTTTLSAIAGLVRA-----------------QKGKIIFNGQDI 4  AHVINRMGIALVPEGRRIFPELTVYENLMMG 11 RDLEWIFSL---\Classic ABC     \ABC ATPases
ABCATPase_Tmar_56966600      14 ----GAVVVKDLRKRIG    KKEILKGISFEIEEG-EIFGLIGPNGAGKTTTLRIISTLIKP-----------------SSGIVTVFGKNV 2  EPHEVRKLISYLPEEAGAYRNMQGIEYLRFV 12 EMVERATEI---|transporters    |
ABC_ATPase_Styp_6573453       5 ----NKLHVIDLHKRYG    GHEVLKGVSLQARAG-DVISIIGSSGSGKSTFLRCINFLEKP-----------------SEGAIIVNGQNI 16 QLRLLRTRLTMVFQHFNLWSHMTVLENVMEA 13 ERALKYLAK---|                |
Mj0796 _Mjan_15826057         1 -----MIKLKNVTKTYK 4  IIYALKNVNLNIKEG-EFVSIMGPSGSGKSTMLNIIGCLDKP-----------------TEGEVYIDNIKT 7  LTKIRRDKIGFVFQQFNLIPLLTALENVELP 15 KRALECLKM---|                |
Btuc_Ec_21465992              1 --MSIVMQLQDVAES--    --TRLGPLSGEVRAG-EILHLVGPNGAGKSTLLARMAGMTSG------------------KGSIQFAGQPL 3  SATKLALHRAYLSQQQTPPFATPVWHYLTLH 7  ELLNDVAGA---/                |
Mj1267_Ec_15826207            4 --TMEILRTENIVKYFG    EFKALDGVSISVNKG-DVTLIIGPNGSGKSTLINVITGFLKA-----------------DEGRVYFENKDI 4  PAELYHYGIVRTFQTPQPLKEMTVLENLLIG 25 EKAFKILEF--->His Perm binding|
Hlyb_Ec_33357348              1 -----DITFRNIRFRYK 2  SPVILDNINLSIKQG-EVIGIVGRSGSGKSTLTKLIQRFYIP-----------------ENGQVLIDGHDL 3  DPNWLRRQVGVVLQ-DNVLLNRSIIDNISLA 6  EKVIYAAKLAG->HB ATP-binding  |
Tap1_Hs_15988434             10 -HLEGLVQFQDVSFAYP 3  DVLVLQGLTFTLRPG-EVTALVGPNGSGKSTVAALLQNLYQP-----------------TGGQLLLDGKPL 3  EHRYLHRQVAAVGQEPQVF-GRSLQENIAYG 7  EEITAAAVKSGA>Tap1            |
Malk_Tlit_12084695            1 ---MAGVRLVDVWKVFG    EVTAVREMSLEVKDG-EFMILLGPSGCGKTTTLRMIAGLEEP-----------------SRGQIYIGDKLV 7  FVPPKDRDIAMVFQSYALYPHMTVYDNIAFP 12 QRVREVAEL--->MalK            |
Abcc7_Mmus_40889533           1 --STTGIIMENVTAFWE 28 GNPVLKNINLNIEKG-EMLAITGSTGSGKTSLLMLILGELEA-----------------SEGIIKHS----    ------GRVSFCSQFSWIM-PGTIKENIIFG 5  YRYKSVVKACQ->CFTR            |
Cftr_Hs_56967140              1 --STTEVVMENVTAFWE 21 GTPVLKDINFKIERG-QLLAVAGSTGAGKTSLLMMIMGELEP-----------------SEGKIKHS----    ------GRISFCSQFSWIM-PGTIKENIIAG 5  YRYRSVIKACQ->?               |
Lmra_Llac_39654151            1 -----MLSARHVDFAYD 1  SEQILRDISFEAQPN-SIIAFAGPSGGGKSTIFSLLERFYQP-----------------TAGEITIDGQPI 3  SLENWRSQIGFVSQDSAIM-AGTIRENLTYG 7  EDLWQVLDLAF->LmrA            |
Dsm3109_Tmar_13096783         1 ---MRLKKLY-LKGF--    -KSFGRPSLIGFSD--RVTAIVGPNGSGKSNIIDAIKWVFGEKFDMIFAGSENLPPAGSAYVELVFEENGE 33 GTGLGVDFYSIVGQGQIDR-IV--------- 4  RVNESFNRF--->Smc head domain |
Muts_Ec_28948844            571 FIDKPGIRITEGRHPVV 3  LNEPFIANPLNLSPQRRMLIITGPNMGGKSTYMRQTAL---------------------IALMAYIG----    --------SYVPAQKVEIG-PI---------    DRIFTRVG---->MutS            /
Consensus/95%                   ......................................h..hhGssGsGK*phh..h...........................h.h...................h..p.................................

                                                            WB                                        Sen-1
Secondary Structure         -HHHHH    ---------------HHHHHHHHHHH-----------EEEE--------HHHHHHHHHHHHHHHHHHHH-----EEEEEEE--HHHHHHH---EEEEE---EEEE------------EEEEE------
IVa2_CAVt1_56158852         IQMSYA    QLTK----DENYDVTSPNNIFANAASKG-------PIAIVMDECMEDLGGNRGIAKFFHAFPSKLLDRFPKCTGYSVIVVLHNMNPRRDQGGNIANLKIQAKVHMISPKVHPSQLNRFINIYTKGQPTA 366\IVA2 ATPase
orf6_CAVt1_1526581          IQMSYA    QLTK----DENYDVTSPNNIFANAASKG-------PIAIVMDECMEDLGGNRGIAKFFHAFPSKLLDRFPKCTGYSVIVVLHNMNPRRDQGGNIANLKIQAKVHMISPKVHPSQLNRFINIYTKGQPTA 422|
TSAdVgp05_TsAV_51556698     IKLSYD    ELLA----DHNYDVTNPLNVFAKAASKG-------PIAIIMDECMEELGKHKGIAKFFHAFPSKLHDKYPQCTGYAVLVVLHNMNPRKDQGGNISNLKIQSKLHIISPKMHPSQLTRFINIYTKGLPTA 366|
IVa2_BAV3_56158806          KIMSYE    DLTM----DYNYDVAHPQNVFAQAAKTG-------PIAIIMDECMEDLGRHKAISKFFHAFPSKLHDKFPKCTGYSVFVVLHNMNPRKDLGGNIANLKIQAKLHLISPRMHPSQLNRFINTYTKSLPLP 365|
BAdVBgp06_BAVB_9630096      KIMSYE    DLTM----DYNYDVAHPQNVFAQAAKTG-------PIAIIMDECMEDLGRHKAISKFFHAFPSKLHDKFPKCTGYSVFVVLHNMNPRKDLGGNIANLKIQAKLHLISPRMHPSQLNRFINTYTKSLPLP 293|
BAdVAgp05_BAVA_52801684     IKMAYS    DLTQ----EYNYDLSDPRNIFAKAASKG-------PIAIILDECMEDLGAHKGISKFFHAFPSKLHDKFPKCTGYSVMVVLHNMNPRKDHAGNISNLKIQAKLHIISPKMQPSQLNRFINQYTKGLPLP 366|
IVa2_BAV2_56158831          IKMSYA    DLTQ----EFNYDLSDPRNVFAKAASKG-------PIAIILDECMEDMGNHKGIAKFFHAFPSKLHDRFPKCTGYAVLVVLHNMNPRKDHAGNISNLKIQAKLHIISPKMQPSQLNRFINSYTKGLPLP 366|
IVa2_PAV5_56160569          VKMSYS    DLTQ----DFNYDVTDPRNVFAKAASNG-------PIAIILDECMEDMGNHKGIAKFFHAFPSKLHDRYPKCTGYSVLVVLHNMNPRKDHAGNISNLKIQSKLHIISPKMQPPQLNRFINSYTKGLPPA 365|
PAdVCgp06_PAVC_13487884     VKMSYS    DLTQ----DFNYDVTDPRNVFAKAASNG-------PIAIILDECMEDMGNHKGIAKFFHAFPSKLHDRYPKCTGYSVLVVLHNMNPRKDHAGNISNLKIQSKLHIISPKMQPPQLNRFINSYTKGLPPA 293|
PAdVBgp05_PAVA_46852407     VRMSYQ    DLMQ----EHNYDVTDSRNVFARAAARG-------PIAIVLDECMEDLGGHKGVAKFFHAFPSKLHARFPRCTGYAVLVVLHNMNPRRDQSGNISNLKIQAKYHIMSPLMQPAQVNRFINAYTKGLPTA 366|
IVa2_SAVA_51518820          VKMSYD    ELTQ----DHNYDVTDPRNVFAQAASRG-------PIAIIMDECMENLGGHRGVSKFFHAFPSKLHDKFPRCTGYSVLVVLHNMNPRRDLGGNIANLKIQAKMHLISPRMHPSQLNRFVNTYTKGLPVA 367|
IVa2_SAV1_61602088          VKMAYD    ELTL----EQNYDVSNPQNVFAKAAARG-------PIAIIMDECMENLGGHKGVSKFFHAFPSKLHDKFPKCTGYTVLVVLHNMNPRRDLGGNIANLKIQAKMHIISPRMHPSQLNRFVNTYTKGLPLA 366|
IVa2_HAVt1_33330444         VKMAYD    DLIL----EHNYDVSDPRNIFAQAAARG-------PIAIIMDECMENLGGHKGVSKFFHAFPSKLHDKFPKCTGYTVLVVLHNMNPRRDMAGNIANLKIQSKMHLISPRMHPSQLNRFVNTYTKGLPLA 370|
IVa2_HAVt1_56160441         VKLSYD    DLTL----DHNYDVSDPENIFAQAAARG-------PIAIIMDECMENLGSHKGVSKFFHAFPSKLHDKFPKCTGYTVFVVLHNMNPRRDLAGNIANLKIQSKLHIMSPRMHPTQLNRFINTYTKGLPLA 366|
IVa2_SAV2_17105046          IKMAYD    DLTQ----DHNYDVSDPRNVFAQAAAHG-------PIAIIMDECMENLGGHKGVAKFFHAFPSKLHDKFPKCTGYTVLVVLHNMNPRRDLGGNIANLKIQAKMHLISPRMHPSQLNRFVNTYTKGLPVA 367|
IVa2_FAV2_78068004          VEMTYE    EATS----PENLNIEHPDNVYVNASRQG-------PLAIVMDECMDKLCSGSSVSVLFHALPSKLFARSANCTAFYVFVVLHNLAPRTAI-GNVPTLKVNAKIHILSCHIPQFQFSRFLYSYAHNISKE 289|
FAdVAgp08_FAVA_9628843      MEMTYE    EATA----PEHLNIDHPDNIYVKVSKRG-------PVAIIMDECMDKLCSGSSVSVLFHALPSKLFARSAHCTAFYIFVVLHNMAPRTAI-GNVPTLKVNAKMHILSCHIPQFQFARFLYAFAHNISKD 346|
IVa2_FAV1_56160741          MEMTYE    EATA----PEHLNIDHPDNIYVKVSKRG-------PVAIIMDECMDKLCSGSSVSVLFHALPSKLFARSAHCTAFYIFVVLHNMAPRTAI-GNVPTLKVNAKMHILSCHIPQFQFARFLYAFAHNISKD 297|
IVa2_FAV8_78068044          VEMTYE    EATS----PENLNIEHPDNVYVNASRQG-------PIAIVMDECMDKLCSGSSVSVLFHALPSKLFARSANCTAFYVFVVLHNLAPRTAI-GNVPTLKVNAKIHILSCHIPQFQFSRFLYSYAHNISKE 299|
IVa2_TAVA_9630780           LECAFD    DVIT----PENLDVNNENSIFNIHTKKG-------PVCVILDECMQKLIQKPNISPLYCSLPSKLSSRYG--HAFYMFVVLHNVNPLSGNGNNIMDLKTQAKLHILSTKNQPLQLSNFVHNRSGGMDSN 280|
FaV1gp03_FAV_9635798        VEIDFE    TAVS----PDNLDINNDKSCFVQAAKNG-------NVCIVIDECMKCLIDKRNISPLFCSLPSKISSKFK--TGFSMFVVLHNINPTSGNGNNIMDLKIQAKLHILSSKNAPFQLSRFVNTYGTGMSPQ 280|
IVa2_DAVA_9629470           ITLTFR    EAIS----DENLNIDNPNNIFKKAAMNG-------PVCIIMDECMNQLGSSHSISSFFHALPSKILGRFPKCTGYTVIVVLHNMNPRHDR-GNIKDLKIQAKCHVISPQLDSGQITRFIRNYSFGFPSS 316|
IVa2_OAV7_7444361           IQLAFK    DAIS----DTNLNIDNPDNIFCRAASEG-------SICIIIDECMNQLGSCHSISSFFHALPSKIFNRFAKCNGYTVVVVLHNMNPRHDR-GNIKDLKIQSKCHVISPQLDSSQINRFIKNYSFGFPSA 359|
IVa2_MAV1_56160678          ITLTYD    DLTA----DHNYDVTHPQNIFAQAAQRG-------PICIIVDECMEELGKHKSIAKFFHAFPSKLHDRFPQCTGYSVFVVLHNMNPRKDQAGNIATLKIQSKCHIISPKMQPSQVARFINTYTKAMPTA 355/
ABCATPase_Tmar_23200171     FPRLKE 2  KQLG----GTLSGGEQQMLAIGRALMSR-------PKLLMMDEPSLGLA-PILVSEVFE-VIQKINQ-E-GT---TILLVEQNALGALKVAHYGYVLETGQIVLEGKASELLDNEMVRKAYLGVA---- 240\Classic ABC     \ABC ATPases
ABCATPase_Tmar_56966600     AG-LGE 2  KDRV----STYSKGMVRKLLIARALMVN-------PRLAILDEPTSGLD-VLNAREVRK-ILKQASQ-E-GL---TILVSSHNMLEVEFLCDRIALIHNGTIVETGTVEELKERYKAQNIEEVFEEVVK 251|transporters    |
ABC_ATPase_Styp_6573453     LK-LSH 2  DRKA----GELSGGQMKLVEIGRALMTN-------PKMIVMDEPIAGVA-PGLAHDIFN-HVLELKA-K-GI---TFLIIEHRLDIVLNYIDHLYVMFNGQIIAEGRGEEEIKNVLSDPKVVEIYIGE- 257|                |
Mj0796 _Mjan_15826057       V-GIDE 3  GKYP----VHLSGGQQQRVSIARALAME-------PDVLLFDEPTSALD-PELVGEVLR-IMQQLAE-E-GK---TMVVVTHEMGFARHVSSHVIFLHQGKIEEEGDPEQVFGNPQSPRLQQFLKGSLK 258|                |
Btuc_Ec_21465992            AELEER 2  NHKP----NQLSGGQQQRVAIARALANN-------PPIILADEPTGALD-SKTGEKIMQ-LLKKLNE-EDGK---TVVVVTHDINVAR-FGERIIYLKDGEVEREEKLRGFDDR--------------- 235/                |
Mj1267_Ec_15826207          LA-LDD 2  GRST----NQLSGGEWQRVRLAAVVLQITPQANPAGQLLLLDEPMNSLD-VAQQSALDK-ILSALCQ-Q-GL---AIVMSSHDLNHTLRHAHRAWLLKGGKMLASGRREEVLTPPNLAQAYGMNFRRLD 238>His Perm binding|
Hlyb_Ec_33357348            AHDFIS 6  NTIVGEQGAGLSGGQRQRIAIARALVNN-------PKILIFDEATSALD-YESEHVIMR-NMHKICK---GR---TVIIIAHRLSTVKN-ADRIIVMEKGKIVEQGKHKELLSEPESLYSYLYQLQSD- 241>HB ATP-binding  |
Tap1_Hs_15988434            HSFISG 5  DTEVDEAGSQLSGGQRQAVALARALIRK-------PCVLILDDATSALD-ANSQLQVEQ-LLYESPE-RYSR---SVLLITQH-LSLVEQADHILFLEGGAIREGGTHQQLMEKKGCYWAMVQA----- 254>Tap1            |
Malk_Tlit_12084695          LG-LTE 2  NRKP----RELSGGQRQRVALGRAIVRK-------PQVFLMDEPLSNLD-AKLRVRMRA-ELKKLQR-QLGV---TTIYVTHDQVEAMTMGDRIAVMNRGVLQQVGSPDEVYDKPANTFVAGFIGSPPM 245>MalK            |
Abcc7_Mmus_40889533         LQQDIT 6  NTVLGEGGVTLSGGQRARISLARAVYKD-------ADLYLLDSPFGYLD-VFTEEQVFESCVCKLMA---NK---TRILVTSK-MEHLRKADKILILHQGSSYFYGTFSELQSLRPDFSSK-LMGYDTF 263>CFTR            |
Cftr_Hs_56967140            LEEDIS 6  NIVLGEGGITLSGGQRARISLARAVYKD-------ADLYLLDSPFGYLD-VLTEKEIFESCVCKLMA---NK---TRILVTSK-MEHLKKADKILILHEGSSYFYGTFSELQN-LQPDFSSKLMGCDSF 263>?      |
Lmra_Llac_39654151          ARSFVE 6  NTEVGERGVKISGGQRQRLAIARAFLRN-------PKILMLDEATASLD-SESESMVQK-ALDSLMK---GR---TTLVIAHR-LSTIVDADKIYFIEKGQITGSGKHNELVATHPLYAKYVSEQLTVG 242>LmrA            |
Dsm3109_Tmar_13096783       ISLLFF 16 DQKL----SLLSGGEKALVGLALLFALM---EIKPSPFYVLDEVDSPLD-DYNAERFKR-LLKENSK---HT---QFIVITHN-KIVMEAADLLHGVTMVNGVSAIVPVEVEKILEVGS---------- 316>Smc head domain |
Muts_Ec_28948844            ------    --------AGRSTFMVEMTETANILHNA-----TEYSLVLMDEIGAGTSTYDGLSLAWA-CAENLAN-KIKA---LTLFATHYFEL-TQLPEKMEGVANVHLDALEHGDTIAFMHSVQDGAASKSYGLA 774>MutS           /
Consensus/95%               ......................sh......bs.h...........s.hhhhDps...h........hb.....p.............hh..pp..........h..h.........p......................
```

Figure 4. Portal Protein superfamily alignment page 1

```
                                                     | predicted β-barrel-like region-------------------------------------------------------------------------------------------->|
SECONDARY STRUCTURE                            -HHHHHHHHHH----HHHHHHHHHHHHHH---EEEEEEE-       ------------EEE--------       ----------EEEEEEE--------EE-----EEEEEEEE-----------EEEEE--------       --EEEE------
gp2_BPWOcauB1_46804777                     99  WLRWTDEADSNGVSDFYGLQALVCRSMIEGGECFVRLRNR       KLEDGFSVPLQLQVLESEHLDN-       KTNQTLGNGNVIRNGI---------EFNRLGQREAYYLFKE--------HPGESMFGESVRVPA-       NDVLHIYRPLRP\phage
HalHV1gp094_HVH_32453918                   96  EEDIAPVALLLHNSEFWRVFNQWVEYAAQDGHSFMELVVE       --DEQFRPRILPTERMHKK----       --TDEYGFVTKYAL-----------EPPGGGGPDDDDATIY--------DPHEVAELWFTKDPTD  1   ------------|λ-like
B_BPlambda_9626247                        119  AEDDCCCIDVERKRTFTMMIREGVAMHAFNGELFVQATW-       DTSSSRLFRTQFRMVSPKRI---       SNPNNTGDSRNCRAGV---------QINDSGAALGYYVSED--------GYPGWMPQKWTWIPRE  5   ASFIHVFEPVED|
Bcep02005900_Bfun_48782926                134  WRLFSNDNDVSRQLTVSQQLRLALRHKLIDGEDLVINYWM   3   VGRGGARYATAYLVVDPDRL---       SNPNQMVDTRHMRNGV---------EVDENGVPLAYHIRKA--------HQNDWYNSVESMIWER  9   LRVIHDFERDRA|
42_BPBcepNazgul_34304545                  118  FELYAESIDNAAISTFTGLIRLGVVGYVKTGEVLATAEW-       DRAANRPYATCFQMVSTDRL---       SNPYQQLDTPTLRRGV---------QYDKRGRPQGYWIQVA--------HPGDLYQMAPDMYKWK  10  RQVIHILEPREP|
gp17_Bhen_49475135                        101  WWQWVDEADYDEDANFYGLQATIAREVFLTGECFVRLHYV       DLYGRSGVPFQLQIYPTEMLDLT       YNGPAETEGNYIRMGI---------EFNASGKRVAYHFWEH--------HPYDDCPANRAFKSQD  6   EMVLHIKERRIA|
orf23_BPVHML_26891710                     101  WGDWVNEADFDEVQSFYGLQALVVRTVINSGEAFVIKKPR       PLSEGLSVPLQLQIIEPDMLASD       IPDETLPSGGYVKGGI---------RFSNGGKRKAYCFYRN--------HPAESSLIGDPVDTV-  4   EHVLHV-TVLTV|
PFL_1995_Pflu_70729376                    101  WEDWVDESDADDLTDFYGQQALVARTVETSGECFVRLRPR       RTEDGLAVPLQVQLLAPEFVPH-       DKFETTRDGNVIRAGI---------EFNGVGKRVAYWMYRS--------HPGESSTLSGGYNQL-  5   TQVLHIFDPLEP|
MCA2660_Mcap_53803156                      95  WRDWTEEADAAGQTDFYGLQALACRAMLEGGECLIRLRPR       RLEDGLVVPLQLQLLEPEHLPI-       SLNTDLPSGNVVRSGI---------EFDSLGRRLAYHLYRS--------HPEDGRLAPMSGQGGM  6   REIIHLYRVLRP|
Rsph03000842_Rsph_46193344                 95  WQDWTDEADAEGVTDFYGLQRRIAREFFLTGECFVRLRAR       RPGDGLTVPLQLQCLPSEMLPI-       GRTEVLGGGRAIRQGI---------EFDAVGRRVAYHFHRR--------HPGDPTEPGLAGET--  5   EDVLHIVDPVES|
orf5T_BPVP16T_37626096                    116  FEAWGELCDFTKRYTFGQWQRVAKAEALIAGDVLVVE---       HHNSATELPTYELIDGDMVQTPY       DAMNKIGKNHRIDNGV---------EMTKDGEHVAFWVVQE--------GFKHKRIKATGANG--  3   RWMYYGTDKRHT/
20_BPT4_5354218                           124  NMMLDEFSDVLNHLSFQRKGSDHFRRWYVDSRIFFHKIID   7   IKELRRLDPRQVQYVREIITE--       TEAGTKIVKGYKEYFIYDTAHESYACDGRMYEAGTKIKIPK--------AAVVYAHSGLVDCCGK  2   IGYLHRAVKPAN\T4-
20v_BPKVP40_4760684                       116  AKINEEFDRVVSLLQMRKHGYKWFRKWYVDSRIYFHKILD   4   IIELRPLNPMKMELVREIQKE--       TIDGVEVVKGTLEYYVYKQSDKMPSWMSATNRAQTSFRIPK--------DAIVFAHSGLMRADDP  2   IGYLDRAIKPAN|like
20_BPAeh1_33414838                        121  EMVREEFRTILKLLKFEREGKRHFRRWYVDSRIYFHKMID   7   IKELRLLDPRNVEYYRVNLKS--       NENGNDVYKGVKEFFTYGATED--NRYNISGNSNNLVQIPI--------DAIVYSHSGKVDIDGK  2   VGYLHNVIKPAN|
20_BPP-SSM4_61806306                      104  KLIRSEFDEILRLLDFDNRAYEIFRRWYVDGRLFFHKVID   7   LVELRYVDPRKIRKVTEYEA---       KRPEALRTQDLNQQLTQQSASFLYNPKGLKNSTNQGMKIAP--------DSIAYCHSGIQDLNKN  2   LSHLHKAIKAVN|
homolog_BPS_WHM1_2353782                  103  KLMREEFDEVLRLLDFENRAYEIFRRWYVDGRLFYHKVID   7   LAELRYIDPRKIRKVTEYEQ---       KRPEQLRGVDINTQLTQKAAEFLYNPKGLKNSTNQGMKITT--------DSVCYCHSGIQDLNKN  2   LSHLHKAIKAVN/
Rorf508_BPphig1e_1926349                  101  NEADKFLNDVLEDNDFKNKFEEALEKGVALGGFAMRPYID   1   NHIKIAWVRADQFYPLQS-----       --NTNDISEAAIASRT---------QRTESNQTKYYTLLEFHQW----QDNGSYQITNELYKSDS  30  PLFAYFKTPGAN\phig1e-
lmo2299_Lmon1_6804338                      99  KAAEEFVLNVLKTNGFTKNMERYIEYGEAMGGFVIKVYHD   2   KNVKVSFATADCMYPLSN-----       --DSENVDECVIANSF-------------HKNDKYYTLLEWNEW-----QDDIYTVTTELYQSDD  28  PTFIYIKPNIAN|like
35_BPBCJA1c_56694903                      110  KTAHEFIQHVFQHNKFIKNLSDYLEPTFALGGLTVRPYVD   1   GEIEFSWALANAFYPLRS-----       --NSNGISEGVMKSVT---------TKVIGNKTVYYTLLEFHEWEKTEEGESLYVITNELYKSDN  28  PLFNYLKPSGFN|
g57_BPLL-H_623061                         100  ETANDFLDDVFQQNDFYTTFEEKLEEWIALGSGCVRPYVD   1   GKIKLAWATADQVYPLQA-----       --DTNQVNELAIASRT---------TEVENHRTIYYTLLEFHQW-----DHGDYVITNELYRSEA  30  PLFAFYRNKGAN/
T1p52_BPT1_45686310                        63  ----KEFKSLWDSYKLDSSLVDLLCWARLYGGAAMVAIKD   2   MLTSQAKPGAKLEGVRVY-----       --DRFAI-TVEKRVTN--------ARSPRYGEPEIYKVSPGDNM------QPYLIHHSRVFIADG      ERVAQQARKQNQ\T1-
STY1048_Sent_16502202                      84  AKDRGTLEAQLDQLQIWDALNDVLKWSRLYGGAVGFIMIE   6   PLRLETIGEGKFKGILPL-----       --DRWMINPVLTRRIK--------EMGPNLGKPELYDVVTTATG-----IPAWRIHHSRLIRFDG      VTLPFQQKMTEN|like
XF1676_Xf_15838277                         95  ADALEPMYRLEQALNVRAKVRDALAWARLYGGAVLFINVH   6   PFDPASVMPGSRLSLTVL-----       --DRWRVALGSGQMDQD-------PLSETYGQPRCYQIAG----------SVERVDRSRMIAFSG      AELPWEAFRGNG|
EF1456_Efae_29376023                       83  VK--ALYESKLRQLKAKERLQQLFTYERLYGDGFVSITIE   1   RSLSEPLDFENIKSVPYI-----       --NAFSGKKISNRIIED-------VFSPNYGQIESFEVNNRSNN-----SRIHLLNYSRATTIHR      SRVLHQLRFEDE|
ACIAD2174_Asp._50085285                    76  DGDLKKLQEACKKFRLNDILLSGLILSRLYGEAYVLLTTD   6   PLDITKLGTGRLEFFTVL-----       --KKKYL-RADKNTYL--------PLSESNGQPEFYDLRLNDGK------AAKKIHYTRLIAF--      -----RYADVVN|
CAD90807_BPAaphi23_31408073                71  SKQLDLFTKFERSLKLRETLTKALQWSSLYGSVGLLVVTD   2   NTSAPLKPTERLKRLIIL-----       --PKWKISPTGTKDDD--------VLSPNFGRYSEYSILGGS--------QSITVHHSRLIIL--      -NANDAPLSDND|
Hflu203000886_Hinf_46133462                71  SEQLDEFTKLERRLKLRETLTKALQWSSLYGAVGLLLVTD   2   NITSPLQPTERLKRLIIL-----       --PKWKISPTGQRDDD--------VFSPNFGRYSEYTITGGT--------QSVLVHHSRLLII--      -NANDAPLSDND|
plu3405_Plum_37527279                     130  DKKIKALNEAIEKFHVRERLKKAIEHDGFFGRGQIYIVRT   2   GVLAATDPVEITKGSLIG-----       --FNVIEPIWTYPGYN--------ASNPLSNKPSSWYVM-----------AKTVHDSRLITF--      ITRPLTFKPAYN/
T5.152_BPT5_46401881                       80  TLDTLLNVRPNPFMDISTFRRLVVTDLLFEGCAYIYWDG-       TSLYHVPAALMQVEADA------       --NKFIKKFIFNNQIN----------------YRVDEIIF----------IKDNSYVC-------      ------GTNSQI\T5-
ORF10_BPpsiM2_3249595                      77  DEGVVDEFIRACKPSFEYVLLRALEDLQVFNYCTLEVVRD   10  IPSHTIRVHKDGSRYRQTWDG--       VNITHFKDYRYEGEIN---------------PETGEDQDS----------VGANELVFIHI----      ------PSPVCS|like
psiM100p17_BPpsiM100_11863102              75  DDGGVEELLRACRPSFEFILLQALEDLQVFNYCTLEVVRD   10  IPAHTVRVHRDGSRYMQTWDG--       IHVTYFKDYRYEGEVN---------------PDNGEDQDG----------VGANEIIFIHL----      ------PSPICS|
2638AORF011_BP2638A_62635888               66  -LYYLLNVRPNRNQNAVDFWQKFIFKLIMDNEVLVVKN--       DEGHFFVADDFEKEDEL------       --GLYSHRFTNVLVN---------------DFEFKRVFTM----------DDVIYLKYNNQ----      ------KLDAFS|
D3p04_BPD3_9635589                         95  PLYDVVHNSPNDDMTAFQFWQAMVASMLLWGNAYAEIRRA   4   AALDFLLPSRVDLECDE------       --NGRLKYFYTTKKG-------------------ARREIE----------RTNMLHIPA------      ------FTLDGR|
4_BPphiE125_17484026                       81  PLYSILKYEPNSWQTPFEYQEQSQVAVGLRGNSYSFIDRD   3   VIQGLYPLDNEAVTVMR------       --GSDLKPVYRVR---------------------GSDPMP----------QRLVHHVRW------      ------MSINGY|
orf3_BPSfV_19483741                        96  PAFYLVHDEPNTWQTSYKWRELKQRHILGWGNGYTWVKRN   1   RGEVISLDCCMPWETTL------       --MNTGGRYTYGLYN-----------------EYGAFAIS----------PDDMIHIRA------      ------LGNNQK|
gene_3_BPHK97_1086132                      98  PLARLLRYSPNQYMTAQEFREAMTMQLCFYGNAYALVDRN   1   AGDVISLLPLQSANMDV------       --KLVGKKVVYRYQR-----------------DSEYADFS----------QKEIFHLKG------      ------FGFTGL|
sb4_BPST64B_23379196                       82  RLHKLISTHPNGYMTPQEFWELVVTCLCLRGNFYAYKVKA   1   -GEVAELLPVDPGCVVP------       --KLNSSWEPVYQVT---------------FPDGSTDVLS----------QEDIWHVRT------      ------LTLDGL/
gp29_BPMu_6010404                         102  KKDADMLNEYLHDAAWFEDALFDAGDAILKGYSMQEIEW-       -GWLGKMRVPVALHHR-------       DPALFCANPDNLNELRL--------------RDASYHGLEL---------QPFGWFMHRA-----      -----KSRTGYV\Mu-
Bcep1808DRAFT_4121_Bvie_67533455          103  KKQAAQLEEWFTDFANLDDVLFDQMDAVGHGFSAQEIEW-       -HQVEKVWLPKTLTHR-------       PQRWFRTPLYDGNDLRL--------------RDNSSDGAPL---------WPFGWLVHKH-----      -----RAKSGYL|like
D3112p27_BPD3112_38229139                 102  KADAEYLHELLLDLEGIEDLMLDCMDGVGHGYSAIELDW-       -SLQGREWLPQAFDHR-------       PQSWFQLNPDDQDELRL--------------RDNSIAGEVL---------QPFGWIMHKP-----      -----RSRSGYV|
BcepMu29_BPBcepMu_48696939                 93  SRVAKSIADVFADL-DLSRIVTEMLDAVLYGYQPMEITW-       -GKVGNYIVPIDVVGK-------       PADWFVYDPENQLRFRS--------------KDHWMQGEEL---------PARKFLVPRQ-----      -----EATYLNP|
plu3441_Plum_37527315                      93  VRVFNFIRDMLADL-DMSRIIGEMHDAVLYGYQPCELMW-       -NQNQS-WSVTDIVSK-------       PPDWFHYDRDNQLRFRA--------------RDAGVEGERV---------PLYKFVTPRQ-----      -----DATYDNP|
HD_0514_Hduc_33147938                     106  EKLVEEVDKLFYQIGNLEDLIMDCMDAVGHGFSALEIEW-       -AFNGKLWYPNTFTHR-------       PQSWFKWDKVDNLLLKT----------------PKNTGEPL---------REFGWVVHTH-----      -----KSRSTQA|
NMB1095_Nm_7226333                        101  EKLSDQAYEMMDSLPTLEDLIMDLMDAVGHGFSALEVEW-       -VFSDGLYLPRNFIHR-------       PQSWFKWDKDNGLLLRT---------------RENPEGEAL---------WPLGWVVHTQ-----      -----KSRSVQQ|
AAQ14780_BP01_33340429                     88  KMRADFFNSLMDDMEDWADFINSVMSFCTYGFCVNEK---       -VYKKRQKKGKYQSFDDG-----       LIGWAKLPIRNQSTLDK-----------------WYFDEKL---------PRAKFMLFKY-----      -----DDEYGNP/
CAA24425_BPT7_15602                       110  MVERIIMNYIESNSYRVTLFEALKQLVVAGNVLLYLPEPE   2   -NYNPMKLYRLSSYVVQR-----       -DAFGNVLQMVTRDQIAFGALPEDIRKAVEGQGGEKKADET---------IDVYTHIYLDEDSGE  19  PKEACPYIPIRM\T7/T3-
8_BPgh-1_29243586                         111  GMVERILMSYMEANSYRVTLFELIRQLALAGTALIYLPPP   3   SSYNPMKLYTLHNHVVQR-----       -DAFGNVLQIVTLDKVAYAALPEDVRNSLS-GGQEYKPEQE---------LEVYTHIYIDDESGD  19  PQDALPWIAVRW|like
8_BPphiYe03-12_6599023                    111  SMVERIIMNYIESNSYRVTLFECLKQLIVAGNALLYLPEP   1   GSYNPMKLYRLSSYVVQR-----       -DAYGNVLQIVTRDQIAFGALPEDVRSAVEKAGGEKKMDEM---------VDVYTHVYLDEESGD  19  PTDAMPYIPVRM|
VP2p04_BPVP2_48696640                     103  ENATHDVYSALQDSNFNLEANETYIDLCGYGNAIMVEEED   3   EGSVVFQSSPIQDSYFEE-----       -DSRGQVVNFYRVFRWTPAQIYDRFGDEGTPEAIIKKAKEA---------SNQLKQEVVVFTRYD  26  GYYEMPAYAIRW|
4_BPepsilon15_30266038                    103  EVVQRRMNEVFNKSNLYQSLPVMYASLGTFGTGAMAVMED   1   QDVIRTMPFPIGSYYLAN-----       -SPRGSVDTCIRQFSMTVRQMVQEFGLDNVSTSVKGMWENG---------TYETWVEVNITPNVN  26  GFDEFPILAPRW|
BIP-1p20_BPBIP-1_45580765                 104  ANVTRLMLMIFAKSNTYRALHSMYEELGAFGTASSIVLPD   1   DAVVYHHSLTAGEYAIAA-----       -DNQGRVNTLYREFQITVAQMVREFGKDKCSTTVQSLFDRG---------ALEQWVTVIIEPRAD  26  GYRSFRALCPRW/
187ORF003_BP187_62635570                  110  NQTNDKIIELNDLNDADEVNSDLALNLSIYGRAYEIVYRD   3   RDTFKVLDPKSTFVVYD------       -QTLDKKVVAGVRYFEK--------QDKDKVPVQHVEVYTT--------DKIYYIEIKGGTYHRV  7   NDVPIIEYLNDQ\SPP1   \SPP1-
gene_6_BPSPP1_2764847                     112  KTLLEYVNELAD-DDFDDILNETVKNMSNKGIEYWHPFVD   3   EFDYVIFPAEEMIVVYK------       -DNTRRDILFALRYYSY--------KGIMGEETQKAELYTD--------THVYYYEKIDGVYQMD  21  GRVPIIPFKNNE|classic|like
phiETA_41_BPETA_17426269                  117  DEVVKRIDEVLG-NRFDDKLHSVLTGASNKGIEWLHPYLD   3   EFKLFRVPAEQGIPIWT------       -DKEHEELEAFIRMY----------KLENETKVEYWDKVTV--------NYYVYENGSLIPDYSN  14  GKIPFIPFKNND|sub-    |
85ORF006_BP85_62635805                    118  KDVLEAIEAFNDLNDVESHNRSLGLDLSIYGKAYELMIRN   3   ETRLYKSDAMSTFVIYD------       -NTIERNSIAGVRYLRTKPI-----DKTDEDEVFTVDLFTS--------HGVYRYLTSRTNGLKL  12  ERMPITEFSNNE|family |
orf502_BPSfi11_3320433                    122  SQNDDAIKRIGRINDIDTHNRNLIRDLSQTGRAYEVIYRS   3   ETRIKRLSPLETFVIYD------       -NSLEDNSIAAVRYYN---------RGTLQNAKDVVEIYTN--------QHIYTLDASDSFNEI-  7   GTVPITEFLNNA|        |
SpyM3_1434_BP315.6_28876464                92  KDLQAAIDLMSVRNNEDYHNVKIKTDLSIYGRAYELLTVE   7   EVKLYQLPAEQTFVIYD------       -DTYQRNSLMAVHFYDI--------DYGSGKRKQIIKAYTS--------DTIYTYEDYNLETKGM  10  KGVPVNEYANNE|        |
ORF32_BPTP901-1_13661706                   92  KEILTKLQEFDNLNDMEDEESELAKMACIYGRAFEFLYQD   3   QTNVVYNSPENMFMVYD------       -DTVKQEPLFAVRYG------------VDEDKKLQGEVYTL--------LETIKISGENDEISFG  7   PDLPVVEFYFNE|        |
14_BPBxz2_29424757                         88  KA-DEQLWDWWQANDLDIESTLGHTDSLVHGRSYITISKP   10  EVPIIRVEPPTNLYAQI------       -DPRTRQVMRAIRAI----------EDEEGNEVIGATLYLP--------NNTVIWNREDGQWVQV  7   EMVPVIPIPNRT/        |
orf43_BPPBC5_17940324                     132  KSELLKYLADVNKTEFHV--SRSFADSAKVGVGWMEDGYQ   9   ARYESWRNMLWDSTATELDLD--   72  RQRIRIIECWFRIPDKT--------SRISGGTFAGEIYDPLSR------GHKVAIEEGDGEIIEK  15  LWLSPSPYRHNR\        |
VT2-Sap53_BPVT2-Sa_9633447                109  EAINAEFADACRLGNMNKARSDAYAEQIKAGLSWVEVRRN   9   VSTVSRNEVFWDWLSREADLS--   71  RRRVLLQVVYYRTFERL--------PVIELSNGRVVAFDKNNL------MQAVAVASGRVQVKVG  15  IVDRPCSAPQGM|P22     |
F116p40_BPF116_56692922                   109  DALNYRLNTAERQSGADRACSEAFRPQIACGIGWVEVSRE   9   CRPIRRDEIHWDMKCGDDW-E--   62  AWTVQEDHWYNPTSKEI--------CLVELWYRRWVQVHVLKS------PDGRVVEYDPNNLAHN  27  CLHDGPTPYTHR|sub-    |
1_BPHK620_13517603                        102  NKLNGLFRADYEETDGGEACDNAFDDAATGGFGCFRLTSM   9   DRQRIAIEPIYDPSRSVWFDP--   39  SMTSWEYDWFDADVIYI--------AKYYEVRKESVDVISYRH------PITGEIATYDSDQVED  31  GFLEKPRRIPGE|family |
Sf6p03_BPSf6_41057281                     102  NKLNGLFRADYEETDGGEACDNAFDDAATGGFGCFRLTSM   9   DRQRIAIEPIYDPSRSVWFDP--   39  SMTSWEYNWFGADVIYI--------AKYYEVRKESVDVISYRH------PITGEIATYDSDQVED  31  GFLEKPRRIPGE|        |
P19_BPAPSE-1_6118014                      102  NKLNGLFRADYEETDGGEACDNAFDDGSTGGFGCFRLTTN   9   ERQRICLEPIYDPARSVWFDP--   38  IERSWDYDWYDVDVVYI--------AKYYEVKKESVDVVSFQN------PLTSETVTYDSDQLEL  31  GFLEKAQRIPGE|        |
1_BPP22_51236726                           92  DVLMGMYRTDMRHNTAKIAVNIAVREQIEAGVGAWRLVTD   8   NNQVIRREPIHSACSHVIWDS--   42  NPNDWVFPWLTQDTIQI--------AEFYEVVEKKETAFIYQD------PVTGEPVSYFKRDIKD  31  AVLKDKQLIAGE/        |
ORF1_BPMB78_6179377                        92  ELEYLLKNADGSGVGLIQHAQDTLMEIDSVGRGGLLVDAP   15  NPTIAFYTTENIVNWRLTRVG--       SVNRVTMVVLRETWEYNEPGNE---FETKYGEQYRVLDIDS---------DGNYRQRLFRFDAEG      ----GAQEDVVE\KS7     |
orf15_BPKS7_57472356                       92  ELEYLLKNADGSGVGLIQHAQDTLMEIDSVGRGGLLVDAP   15  NPTIAFYTTENIVNWRLTRVG--       SVNRVTMVVLRETWEYHEPGNE---FETKYGEQYRVLDIDT---------DGNYRQRLFRFDAEG      ----GAQEEVVE|sub-    |
AAX84851_BPXp15_62529168                   85  AMSKYFE--DQSGIQFYEVFTRAVEETLLMGRVGVFIDRP   4   DPYISVYTTENILNWEEDEDG--       ---RLLMVVLREFYTVRDTADR---YVQNIRVRYRCLELV----------DGLLQITVHETQDGK      ----VWELAKTS|family |
JL001p60_BPJL001_62327246                  95  TMEPLREKATLRNESLEMLLRRINEEQLITGRLGLLTDMP   21  LPYIALYTAEAIINWDKGRSD--       GIQNLNLVVLDESEYERDKD-----FEWVWQEKYRVLEPQG---------QGVYSVGVFRDNQFT      ----FSQGEMVT/        /
Consensus/80%                                  ......b....p..ph......hh..h.h.G.hhhb.....................h..........................................................................................
```

Figure 4. Portal Protein superfamily alignment page 2

```
SECONDARY STRUCTURE              --------HHH-----HHHHHHHHHHHHHHHHHHHHHHHHH----EEEE--------------------------HHHHHHHHHH----------------------EEEE---------------HHHHHHHHHHHHHHHHH----------------HHHH-
gp2_BPWOcauB1_46804777           GQI----RGEPWLSNILLKLYELDQYDDAELVREKTAAMF--AGFITRLDPEANILGE-----------SESNEQGVALSGLEP-------GTMQLLDP---GEDIKFSEPSD-------VGGSYEAFMKQQLRAIAIGTGITYEQLTGDLTGVNYSSIR\phage
HalHV1gp094_HVH_32453918         -------FGRSFVEPIQEQADMLRDMEFDYARFVATKAYP--PILWKL-GTDEEQWSE-----------DQIGGWLDTVEQIEP-------DSMLAAPH---DVDYDVVGVTSTSSSA--GAMKLEGTFEHLQNRIVTGLGIPAL-LMNMGGSTGEATAS|λ-like
B_BPlambda_9626247               GQT----RGANVFYSVMEQMKMLDTLQNTQLQSAIVKAMY--AATIESELDTQSAMDFILGANSQEQRERLTGWIGEIAAYYAAAPVRLGGAKVPHLMP---GDSLNLQTAQD-------TDNGYSVFEQSLLRYIAAGLGVSYEQLSRNYAQMSYSTAR|
Bcep02005900_Bfun_48782926       GQN----RGVGFIPVLAHAKMLARYYGIELQAAALAASI--GTYVTSPYDPSEVQDAVGGE-----DHERAEWNEERPAMFNG-------VRVPALAP---GEDIKAFTSDH-------PHNGFTEFVHEMQGCVASALGISLEQVTQDWSRSNYSNMR|
42_BPBcepNazgul_34304545         DQS----RGIADIVSGLKDMRMAKRFKEMSLQNAVINASY--AAAIESELPPEFIHSQMSGG-----SPNKAYVGGANNIQIDG-------AKIPHLFP---GTKLNLKPMGT-------PGGVGSEFEASLNRHLASAFGMSYEEFTRDFSKANYSSIQ|
gp17_Bhen_49475135               GQL----RGSPKITRSMTKIFQLESYDDAELDRKRTAALF--AVFITGKESHDAALEENREQ-----TRLKKPEEAADVDKIYP-------GSVNMVQD---GEKITFSTPVE-------VGGSYEAFQFRNILKICSALNMPYAVVTGDVTRGNFSNVR|
orf23_BPVHML_26891710            RSD----AGAPWF-QLLLRLNELDQYEDAELVRKKTAALF--AAFIQEATADSTGGPTIGQ---------PKRSKGGKRITGLNP-------GTLQYLQP---GQEVKFSNPAD-------VGTTYEPWLRYQLLSIAKGYGITYEMLTGDLRGVNYSSIR|
PFL_1995_Pflu_70729376           GQL----RGVPRLSPVLKRLRSLDNYDDAVLFRQEVANLF--AGFITRPTPEGGPPPLDPVT-----GQPLDRDGFTPMVALEP-------GTMQELGP---GEEVEFSKPPD-------AGNNYPDFMRQQLMAAAAGAGVPYEIVTGDMRDINDRVMR|
MCA2660_Mcap_53803156            GQI----RGEPWLSRALVKLNELDQYDDAELVRKKTAAMF--AGFVTRQNPEDNLMGE-----------GAADGDGIALAGLEP-------GTLQILEP---GEDIKFSDPAD-------VGGSYSEFLRTQFRAVAAAIGITYEQLTGDLTGVNYSSIR|
Rsph03000842_Rsph_46193344       GQL----RGVSRFAPAIVKLFLLDQYDDAELDRKKVAAMY--AMFITSNDPD------------------AAPLEGELGDQVAP-------GQIVRLDP---GEDMKVADPAD-------SGATYEPFQYRTLLQVSAALGIPYAHLSQDMVKANYSNAR|
orf5T_BPVP16T_37626096           EA-----RGEPLLTL-VMQNIAEIDKMRDATQRKATLGAQ-IVGFIQRQIGGKTPGTKPFTS-----GAQDGTVRQTNIAETPF-------GMMEGLAE---GEEIKAFTNDT-------TDEKFGEFEEAILRAVAWALGMPFSVLAMQFD-SSYSASR/
20_BPT4_5354218                  QL-----KLLED----AVVIYRITRAPDRRVWYVDTGNMP-ARKAAEHMQHVMNTMKNRVVYD-----ASTGKIKNQQHNMSMTE------DYWLQRRD---GKAVTEVDTL--------PGADNTGNMEDIRWFRQALYMALRVPLSRIPQDQQGGVMF\T4-
20v_BPKVP40_4760684              QL-----KMLED----ALVIYRLARAPERRVFYVDVGNLP-TQKAQQYVNGIMQNVKNRVVYD-----TQTGQVKNTTNAMSMLE------DYYLPRRE---GSKGTEVSTL--------PGGQSLGDIEDVLYFNRKLYKAMRIPTSRAASEDQGGINF|like
20_BPAeh1_33414838               QL-----KMLED----AMVIYRITRAPERRVFYIDVGTMP-NKKATQHLNNVMQGLKNRVVYD-----SSTGKVKNSSNNLAMTE------DYWLMRRD---GKATTEVSTL--------PGAQSMGEMDDVRWFNRKLYESMKIPLSRLPQEGAGVTFG|
20_BPP-SSM4_61806306             QL-----RMIED----SLVIYRLSRAPERRIFYIDVGNLP-KNKAEQYLREVMGRYRNKLVYD-----ANTGEIKDDKKFMSMLE------DFWLPRRE---GGRGTEISTL--------PGGQNLGELEDVKYFQKKLYKALNVPSSRLETETTFNIG-|
homolog_BPS_WHM1_2353782         QL-----RMIED----SLVIYRLSRAPERRIFYIDVGNLP-KNKAEQYLREVMSRYRNKMVYD-----ANTGEIKDDKKFMSMLE------DFWLPRRE---GGRGTEISTL--------PGGQNLGELEDVKYFQKKLYKALNVPGSRLETETTFNIG-/
Rorf508_BPphig1e_1926349         NINIESPLGLGVV---DNAKHVLDDINDTHDQFIWEIRLG-QKHIAVQPGML----------------RFDDEHKPTFDTEQN--------VYVGVLSD-DNNGLGVKDMTTP-------IRTVQYKDAIDHFIKEFEVQIGLSTGTFSYSNDGVKTAT-\phig1e-
lmo2299_Lmon1_6804338            NKNLTSPLGISVY---ANALDTLKTLDLMFDSFYQEFKLG-KKKVLVPSSFVKTAI------------GLDGSTTQYFDSTDE--------AFFLYQGDQDADGKSVKDISVE-------IRSTEFIESINAMLRIYAMQVGLSAGTFTFDENGLKTAT-|like
35_BPBCJA1c_56694903             NINPHSPLGLGIT---DNSVSTLKKINDTYDQFWWEIKMG-QRTVFVSDVMLRTVPD-----------ESGMPPPQVFDPDVNVY------KSIRMGTD----EEFVKDVTHD-------IRTEQYKEAINQALRTLEMELKLSVGTFSFDGRSMKTAT-|
g57_BPLL-H_623061                NKNFTSPMGMSLI---DNSYTVIDAINRTHDQFVDEVKKG-QRRLIVPAEWLKTGSSYGG--------QASETHPPMFDPDET--------VYQAMYGD--ASEVGFHDATSP-------IRVADYQATMDFFLREFENQTGLSQGTFTTSPSGIQTAT-/
T1p52_BPT1_45686310              GWG----ASVL-----KSLIDAICDYDYCESLATQILRRK-QQAVWKVKGLAEMCDD-----------DDAQYAARLRLAQVD------DNSGVGRAIG---IDAEEEYDVLNS------DISGVPEFLSSKMDRIVSLSGIHEIIIKNKNVGGVSA---\T1-
STY1048_Sent_16502202            EWG----MSVV-----ERIWDRLTAFDSATVGAAQLVYKA-HLRTYKVEKLRELIAL-----------GGPAFELLKNIDLIR------QFQSNEGMTL---MDAKDTFETHQY------SFSGLDDILSQFAEQISGAVGIPLVRLFGQSPKGFS----|like
XF1676_Xf_15838277               YWH----DSVL-----QAMYNALSRYDTATQGTASMFFEA-VVDVLRISGLSDTLSS-----------DQGTQEVHKRFQLAA------MMKSFNRMLL---LDAKDEYTQKTN------HFSGVKEVIEQFMMDISGAADIPATRLFGQSPKGMNA---|
EF1456_Efae_29376023             LEG----SSLL-----ENLYDILTVADTSVWSVGQILYDF-IFKVYKSADVRSLTP------------QDKLEIETKMDYQ----------FRTEAIAI---IDKEESLGKESS------SVAGIGELLDFVWDYLAGAARMPKTVLKGQEGGTVT----|
ACIAD2174_Asp._50085285          EEP----QSIL-----QEIYEDLLDHAAVKKGSASLVHES-KIDVIRTPNLVDKIK------------EDMKA-VAERFLSVG------LLKSLNGMLV---LDKDEEYDSKTY------NFAGLPDLMSEFSIQTAGAAEIPYTILFGQSPAGMNA---|
CAD90807_BPAaphi23_31408073      IWG----VSDL-----EKIIDVLKRFDSASVNVGDLIFES-KIDIFKIAGLSDKIA------------AGMENEVASVISAVQ------EIKSATNSLL---LDAENEYDRKEL------TFTGLKDLLTEFRNAVAGAADMPVTILFGQSVSGLA----|
Hflu203000886_Hinf_46133462      IWG----VSDL-----EKIIDVLKRFDSASANVGDLIFES-KIDIFKIAGLSDKIS------------AGLENDVAHVISAVQ------SIKSVTNSLL---LDAENEYDRKEL------SFGGLKDLLTEFRNAVAGAADMPVTILFGQSVSGLA----|
plu3405_Plum_37527279            FGG----LSLT-----QIAESYVQNWYRTRDSVSDLIHSF-SVSGFKTNLQSTLQG------------QTKNPDLIYRAELFN------KMRDNRGVMM---LDNEEEFFQFNT------PLSGLDSLQAQSQEHMSSVSSIPLVIFTGITPSGLNA---/
T5.152_BPT5_46401881             -------SGQSRV---ATVIDSLEKRSKMLNFKEKFLDNG-TVIGLILETDEILN-------------KKLRERKQEELQLDYNP---STGQSSVLILD---GGMKAKPYSQI-------SSFKDLDFKEDIEGFNKSICLAFGVPQVLLDGGNN-----\T5-
ORF10_BPpsiM2_3249595            Y------YGVPRY---VSAAPAILAMQKIDEYNYAFFDNYTIPSYVITVTGEFEDELEEDPDGN-PTGRTVIQALIEDNFKHLKE---APHTPLVFSIP---GGDTVKVTFTPLN-----TSQKELSFREYAAEKKYDIAAAHMIDPYRLGIADTGPLGG|like
psiM100p17_BPpsiM100_11863102    Y------YGVPRY---LSAAPSILAMQKIDEYNYAFFDNYTIPSYVITVTGEFEDEMELGSDGE-PTGRTVLQGLIEDNFKYLKE---APHTPLVFSIP---GGDTVEVTFTPLN-----TSQKELSFREYAAEKKHDIAAAHMIDPYRLGITDVGPLGG|
2638AORF011_BP2638A_62635888     -------LGLFE--------DYGEIFGRMIDLQMLNNQ---IRGILKVDATKFYNK------------EKQKELQAYIDTLFDAFQ--NNTIAVVPLTE---GLAYEEHSNRGAAQ----SAQQFSELNELKKTVLTDVARMIGVPPSLVLG-EM-----|
D3p04_BPD3_9635589               -------IGLSAI---RYGVDVFGSVMSAEDAANGTFKNG-LLPTVAFKVDRILQ-------------PAQREEFREYVK-SVSG---AMNSGRSPVLE---QGITPETIG---------INPVDAQLLETREHGVIEICRWFGVPPWMIGQTDKGS---|
4_BPphiE125_17484026             -------TGLSPV---LLHANAIGHAQAIQQYAGKSFMNG-TALSGVIERPKDAPALKD---------QASVDRITDGWNAKFGGS--GNAKKVALLQE---GMTFRPLS---------MTNVDAALIDALRLSALDIARIYKIPAHMVNELER-----|
orf3_BPSfV_19483741              -------MGLSPI---MQHAETIGMGMSGQKYTESFFSGN-ARPAGIVSVKSGLN-------------KESWGWLKDQWQKASQALRRQENKTMLLPAD---LDYKALT-----------VSPVDAQIIDMMKLNRSMIAGIFNIPAHMINDLEK-----|
gene_3_BPHK97_1086132            -------VGLSPI---AFACKSAGVAVAMEDQQRDFFANG-AKSPQILSTGEKVLT------------EQQRSQVEENFKEIAGG----PVKKRLWILE---AGFSTSAIG---------VTPQDAEMMASRKFQVSELARFFGVPPHLVGDVEKST---|
sb4_BPST64B_23379196             -------VGLNPI---AYAREAISLAAATEEHGARLFSNG-AVTSGVLRTEQTLS-------------DQAYERLKEDFEERHTG---LGNAHRPMILE---MGLDWKSMA---------LNAEDSQFLETRKFQLEEICRLFRVPLHMVQNTDR-----/
gp29_BPMu_6010404                GT-----NGLVRT---LIWPFIFKNYSVRDFAEFLEIYG--LPMRVGKYPTGS---------------TNREKATLMQAVMDIGR------RAGGIIPM---GMTLDFQSA---------ADGQSDPFMAMIGWAEKAISKAILGGTLTTEA-GDKGARS\Mu-
Bcep1808DRAFT_4121_Bvie_67533455  TR-----AGLHRV---LAWPYLFKTYAVSDLAEFLEIYG--LPLRVGKYPPGS---------------TKEEKATLLRAVAEIGH------NAAGIIPE---GMLIEFQEA---------ADGTKDPFEAMIDWCEKSVSKAILGGTLTSQADGKTSTNA|like
D3112p27_BPD3112_38229139        AR-----SGLFRV---LAWPYLFKTYSTADLAEMLELYG--LPIRLGKYPPGT---------------PDEEKVTLLRAVTGLGH------AAAGIIPE---SMSIDFQEA---------SKGSAEPFMAMMRWCDDSMSKAILGGTLTSQSESGGGAYA|
BcepMu29_BPBcepMu_48696939       YG-----FPDLSM---CFWPTTFKKGGLKFWVQFTEKYG--SPMLVGKHPRSA---------------SDGEKNLLLDCLEDMVQ------DAVAVVPD---DSSIEIKEAAG-------KTGSADVYERLLHFCRGEVSIALLGQNQTTEATSTRASAQ|
plu3441_Plum_37527315            YG-----FPDLSM---CFWPVTFKKGGMKFWVRFAEKYG--SPWVIGKHPRGT---------------PQSEIDLLLDSMEAMID------DAVAAIPD---DSSVEIKEAAG-------KSDSSDIYQNLITVARSEISIALLGQNQTTEANSNKASAQ|
HD_0514_Hduc_33147938            AR-----NNLFRT---LAWLYMFKHYSVHDFAEFLELYG--MPIRIGKYGAGA---------------TKDEKNTLKRALAEIGH------NAAGIMPE---SMSIELHNAANG-------VAAGNNPFLQMIDWCEKSIARLILGQTLTSGADGKTSTNA|
NMB1095_Nm_7226333               AR-----NGLFRT---LSWLYMFKHYAVHDFAEFLELYG--MPIRIGKYGAGA---------------TKEEKNTLLRAVAEIGH------NAAGIMPE---GMEIELHNAANG-------TTATSNPFLQMADWCEKSAARLILGQTLTSGADGKSSTNA|
AAQ14780_BP01_33340429           EG-----RSPLLN---AYVPWKYKIEEYEAVGVSRDLVG--MP-KIGDYLDEN---------------AEPEKKAFVQYCKTVND------RAGLIWPR---YIDPDTKEDIFEFSLVSRQGAKAYDTGSIIDRYSKQIMMAFMSDVLAMG-QSKYGSFS/
CAA24425_BPT7_15602              VRLDGESYGRSYI---EEYLGDLRSLENLQEAIVKMSMISSKVIGLVNPAGITQPRRLTKAQTG--DFVTGRPEDIFLQLEKQA----VDFTAKVSD---AIEARLSAFVFMLNSAV--QRTGERVTAEEIRYVASELEDTLGGVYSILSQELQLPLV-\T7/T3-
8_BPgh-1_29243586                TKRDGEHYGRSHV---EEYLGDLNSLESLNEAMIKFAMISSKVVGLVNPNGITQVRRLVKAQTG--DFVAGRKADIEFLQLEKTA----VDFTAKSVAD---AIEARLSYFVFMLNSAV--QRSGERVTAEEIRYVASELEDTLGGVYSILSQELQLPIV-|like
8_BPphiYeO3-12_6599023           VRIDGESYGRSYC---EEYLGDLRSLENLQEAIVKMSMISAKVIGLVNPAGITQPRRLTKAQTG--DFVPGRREDIDFLQLEKQA----VDFTAKAVSD---QIEARLSYAFMLNSAV--QRTGERVTAEEIRYVASELEDTLGGVYSILSQELQLPLV-|
VP2p04_BPVP2_48696640            RKSAGSQWGFGPS---HLALPDVLTANRYVELVLRSSEKVIDPAIMVTERGLISDIDLGAS-------GLTVVRDMESMKFESRA---VRFDSSIQLTD---LRSAVRRIYYVDQLQ---MKDSPAMTATEVQVRYELMQRLLGPTLGRLENDFLSPMI-|
4_BPepsilon15_30266038           EVNGEDVYASSPG---MLALGQVKALQVEQKRKAQLIDKATNPPMVAPTSLKNQRVSLLPGDVT----YLDVISGQDGFKYLVNP---DNTALLADIQD---TRQTINSAYFVDLFMMLQNINTRSMPVEAVIEMKEEKLLMLGPVLERLNDEALNPLI-|
BIP-1p20_BPBIP-1_45580765        ALVGGDIYGNSPA---MEALGDVRQLQHEQLRKAQAIDYKSNPPLQLPVSAKNQDISTVPGGLS----YVDAAAPNGGIRFEVNL---HDLSLLADIVD---VRERIKASFYADLFLMLANGTNPQMTATEVAERHEEKLLMLGPVLERMHNEILDPLI-/
187ORF003_BP187_62635570         FK-----QGDF-----ENVIALIDLYDSAQSDTANYMTDL-NDAMLAIIGNVDLDSE-----------DAKAFRDANMIHLEP--------GTNANGSE---GKAEVKYVYKQY------DVAGVEAYKKRLQNDIHKYTNTPDLNDEQFSGVQSGESM-\SPP1   \SPP1-
gene_6_BPSPP1_2764847            EM-----VSDL-----KFYKDLIDNYDSITSSTMDSFSDF-QQIVYVLKNYDGE--------------NPKEFTANLRYHSVI-----------KVSGD---GGVDTLRAEIP--------VDSAAKELERIQDELYKSAQAVDNSPETIGGGATGPAL-|classic|like
phiETA_41_BPETA_17426269         LE-----ISDI-----FMYKTLIDAYNRRLSDLSNTFKDS-NELTYVLTNYDDQ--------------ELPEFKRLLRYYGAI-----------KVSDN---GGVDTIQVEVP--------VENSKKYLDELYQKIMLFGQAVDFSSDKFGSAPSGVAL-|sub-   |
85ORF006_BP85_62635805           RR-----KGDY-----EKVITLIDLYDNAESDTANYMSDL-NDAMLLIKGNLNLDPV-----------EVRKQKEANVLFLEPTVYEN---RDTGIETE---GSVDGGYIYKQY------DVQGTEAYKDRLNSDIHMFTNTPNMKDDNFSGTQSGEAM-|family |
orf502_BPSfi11_3320433           DG-----IGDY-----ETELYLIDLYDSAESDTANHMSDM-ADAILAIYGDLALPQGM----------QASDMKRTRLMQLKPPK--------SADGKE---GTVKAEYLTKSY------DVSGAEAYKTRLNKDIHVFTNTPDMSDNHFSGNASGEAL-|       |
SpyM3_1434_BP315.6_28876464      ER-----TGAY-----ESVLDNIDAYDLSQSELANFQQDS-VNALLVIAGNAYTGADENDYLDD-GRLNPNGRLAISIGFKKAQVLIL---DDNPNPNG---VKPQAYFLKKEY------DTAGSEAYKNRLVADILRFTFTPDTQDMKFSGVQSGESM-|       |
ORF32_BPTP901-1_13661706         ER-----MSIF-----ESVISLVNAFNKAISEKANDVDYFSDQYLTFLGAAVEEE-------------DLKNIRSNRVINYY-------------ADGEG---KNVDVKFLEKPD------SDSQTENLLDRLTKLIFQTTMVANISDESFGS-SSGVSL-|       |
14_BPBxz2_29424757               RL-----SDLYG---TTEITPELRSVTDAAARTLMLMQAT--AELMGVPQRLLFG-------------VKGEELGVDPETGQTLFDAY---LARILAFE---DHESKAQQFSAA------ELRNFVDALDALDRKAAAYTGLPPYYLSFSSENPASAEAI/       |
orf43_BPPBC5_17940324            Y-PFTPIWNKRRG---RD--GMPYGMIRNIRDIQSDINKR-ASKALHILSSNK---------------VIMDDGAVEDIN-----------------------ELAEEIARPDAIIV---KQQGKEFKIDTDRELGQWHLELMSRNISMLQQVGGVTDEN\       |
VT2-Sap53_BPVT2-Sa_9633447       F-PLVPFWGYRKD---KT--GEPYGLISRAIPAQDEVNFR-RIKLTWLLQAKR---------------VIMDEDATQLSDNDL---------MEQIERP----DGIIKLNPVRKNQK---S-VADVFRVEQDFQVASQQFQVMQESEKLIQDTMGVYSAF|P22    |
F116p40_BPF116_56692922          HFPYVPFFGFRED---AT--GIPYGYVRGMKYAQDSLNSG-VSKLRWGMSVAR---------------VERTKGAVAMTDAQF---------RRQIARP----DADIVLDE---NHM---AKPGARFDVKRDYTLTDQHFQMLQDNRATIERVSNITAGF|sub-   |
1_BPHK620_13517603               HIPLIPVYGKRWF---IDDIERVEGHIAKAMDPQRLYNLQ-VSMLADTAAQDP---------------GQIPIVGMEQIRGLEKH-----WEARNKKRPA--FLPLREVRDKYGNII---AGATPAGYTQPAVMNQALAALLQQTSADIQEVTGGSQAM-|family |
Sf6p03_BPSf6_41057281            HIPLIPVYGKRWF---IDDIERVEGHIAKAMDPQRLYNLQ-VSMLADTAAQDP---------------GQIPIVGMEQIRGLEKH-----WEARNKKRPA--FLPLREVRDKSGNII---AGATPAGYTQPAVMNQALAALLQQTSADIQEVTGGSQAM-|       |
P19_BPAPSE-1_6118014             HIPLIPVYGKRWF---IDDIERVEGHIAKAMDAQRLYNLQ-VSMLADSATQDT---------------GSIPIVGKSQIKTLEKY-----WANRNKNRPA--FLPLNEIVDKQGNII---APPTPVGYTQPQPLNQAMAALLQQTGADIQEVTGGSSQAM-|       |
1_BPP22_51236726                 HIPIVPVFGEWGF---VEDKEVYEGVVRLTKDGQRLRNMI-MSFNADIVARTP---------------KKKPFFWPEQIAGFEHM-----YDGNDDY-------PYYLLNRTDENSG---DLPTQPLAYYENPEVPQANAYMLEAATSAVKEVATLGVDT/       |
ORF1_BPMB78_6179377              IYPD---LGESLR---GVIPFTFIGATNNDATIDDAPLLP-LAELNICHYRHSADN---------------EESSFVVGQP----------TLFIYPGE------NLTPQSFKE-------ANPNGIKFGSRCGHYLGFGGARITWPARICWDKEHRLSRL\KS7    |
orf15_BPKS7_57472356             IYPD---LGESLR---GVIPFTFIGATNNDATIDDAPLLP-LAELNIGHYRNSADN---------------EESSFVVGQP----------TLFIYPGD------NLTPQSFKE-------ANPNGIKFGSRCGHNLGYGGGENNLARQNMLDKEQQAIQI|sub-   |
AAX84851_BPXp15_62529168         TIQN---VGVTM----DYIPFFCITPSGLSMTPAKPPMID-IVDINYSHYRTSADL---------------EHGRHFTGLP----------TPWITGAE------SQSTMHIGS-------TKAWVIPEVAAKVGFLEFTGQGLQSLEKALSEKQAQLASL|family |
JL001p60_BPJL001_62327246        PSH----RGTTL----QMIPFTFINSCDIVPEPDQPPLLG-LSNLVLTIYRGEADY---------------RQSLFNQGQD----------TLVTIGAN------LQDEQARVG-------ANARLDLPTDADAKYIGVDSSGLSEQREALTNDYNRAAQK/       /
Consensus/80%                    ........t.............h..h..................hh.........................................................................s......p........h..h.................
```

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

```
SECONDARY STRUCTURE                    --HHHHHHHHHHHHH-----HHHHHHHHH     -----HH-HHHHHHH--------------HHHHHH-----HHHHHHHHHHH--------HHHHHHHHH----EEEEE----------
Tuc2009_28_BPTuc2009_13487829       9  WILEEGLLKIQGWARDGL-TEEQIAHNMG     IAVSTLG-NWKKSHL--------------EILEAL-----KKGKEVVDIQV--------ENALLKRALGYEFVEVTKEL--------
PBC5p48_BPPBC5_18071234            42  KGATKQGSNVKAKVKKAP-AKPRTARKAA  23 FSDKDIADSFGIDVETLEWWK---EEHPDFGAAFR-----PDPRNYGGRPTSWDDRNITANTLATLGATDLEIAQAFEVSIRTIHRW
D3112p24_BPD3112_38229136           2  AHPKETRDALRRAYVLDR-QSLEVAAAMF  1 VSYGTAR-RWKQQ-----------AEAEGDDWDKA-----QSAQLLAGGGL--------EDVARQVLAGLVTQFQATMEAIQVDADI
HK022p01_BPHK022_9634120           12  AAAVQAMKNAAAVDTIDPP-SHAGLEKKAE  6 IRSKALD-SWTPADLLAAVELANNQLYITVLRKDL-----RKEERIRGEER--------DEGLIKDLRKQIVELQRTIL--------
pacA_BPP1_46401729                 14  RDGGYTIAQYAAEFNLNP-NTARRYLRAF  23 INDQHAGDHIAAEIAEKQRVNAVVSAAVENAKRQN-----KRINDRSDDHDVITRAHRTRDRLERDTLDDDGERFEFEVGDYLIDNV
JL001p58_BPJL001_62327244          14  SKNQPRAETTTDNLPATT-GSVAHVELEV      -TDQDLP-DKRTKKFMDKPVVEQRDRAAAKFLAYY-----AQTGRKGASARAA------GVSYNVVRHWEVHDDDFGTLVQEAHNDW
phiA1122p14_BPphiA1122_30387467     8  IEPTDSCIEWTGAFHKNGYGVISLNRKVA      -AKLGIP-RVQFVHR---------ASYIQHKGPIG-----AGLVVRHLCHN--------RKCFNPAHLEVGTQYDNFEDSVKVGRNK
orf5_BPPaP2_48697075               20  ADLVAVSPSYISQLLTSD-SDFRDYVTYL  4 AEETGQP-TEAEALELKYEAAEHTLKALLERLPMM-----EDKNLLTAVQV----------LSNRRQQEAKAKLLAKGIGYG-----
ORF8_BPpsiM2_3249593                1  MKETPTQRKAFEYYYSLG-DKRNLKKVAE  3 YSYDTIR-GWSSKLRWRERI----YQYEKKQLQEI-----RRAREELSEES--------KEYYADLWSKYLKACSLTLEGYIESI--
ORF8_BPpsiM100_11692720            1  MKETPTQRKAFEYYYSLG-DKRNLKKVAE  3 YSYDTIR-GWSSKLRWRERI----YQYEKKQLQEI-----RRAREELSEES--------KEYYADLWSKYLKACSLTLEGYIESI--
ORF148_BPphiJL-1_62327096          13  FYSLPENQQKAIMLLFSG-KQTQGRIAEN  2 VARVTLT-LWRQKDK---------FRKAQDEYNRF-----MLRDLTNEAILTM------RDLLNSRSEMVRF----------------
terS_BPTP901-1_13786561             3  VLENARHEKFVQCLISGM-SQRKAYREAF  5 WKDSTVD-VKASELFGKV------LVRYKELQEEA-----QDAAIMTRKER--------MVTLSEIAKNA----------------
ul36_36_BPul36_21716107             3  VLENARHEKFVQCLISGM-SQRKAYREAF  5 WKDSTVD-VKASELFGKV------LVRYKELQEEA-----QDAAIMTRKER--------MVTLSEIAKNA----------------
gp1_BPES18_62362214                 2  TGLTIKQEAFCQAYIETG-NASEAYRTAY  4 MKPEVVH-VQACKLQDNPKI----ALRIKELRGEI-----KQRHNVTVDSL--------LAELEEARQKALS----------------
PaP3p01_BPPaP3_27476055            14  LSAKEKLDLYCEGLADGL-NKTQAYVAAG     FSPNHAQRNVAAY-----------HRKHSEYINAF----ISERIGSHVPMA--------LRVIVSIAEDP----------------
Ljo_0307_BPLj965_41179239          78  DELTDKQRAFCLYYLQRY-NATWAYQKAY  1 GNYETAL-RAGPRLLGNVGV----KNYLAELKKQQ-----SQDLYATANDI--------LLRYLKQATSDVTDVLSFKTKKRLVYYK
epsilon15p01_BPepsilon15_30387380  16  KPLTAMQEAYAQEYTKCPENQTQAAINAG     FSPNTAA-VKASVMMRDERI----QKRIAELMEER-----NKRLRVSADYV--------LLRLVEIDQMDVIDILNDDMSIKPVSEW
gp33_BPBCJA1c_56694901             65  TNLNDRQRMFCLYYIKYF-NATKAYQKAY  1 CSYATAN-VEGYRTLVKPSI----KQEIERMKVEQ-----HSGLLLDAQEV--------LQKYIDIAFADITDFVEFGTKESAEVNV
BcepB1A_gene45_BPBcepB1A_48697530   1  MKITPRVLSFIELVVEGV-NNSDAYRRCY  4 LKAATIH-RAAHDLRHDPDVTAEIKRLAERRAASI-----RNAADVANEFL--------RIGFADPGEIVQHRRLCCRFCHGDSHA-
Ljo_1440_BPLj928_41179313           2  KELTPKQKKFCQEYLKDF-NASRAYKAVY  2 KSDNQAR-ANGSRLIANDNV----SEYLSETMHQT-----KVNDILDINGV--------LDNLSQLAIGKPREKVFKRISYKGK---
g02_BPBcepC6B_48697192              2  SKLTAKQQLFVDEYLVDL-NASAAARRAG     YSAKTAG-QMGDENLKKPQI----AAKIEAAMKAR-----GERTRMTADQVLLLAE---SMLLADVNQLISYQHRCCRHCWGIGHA-
F116p38_BPF116_56692920            12  AKLTSQQAAFVQEYLVDL-NATQAAIRAK     YSAKTAR-SIGSQLLTKLNI----QEAIAKARLEQ-----QERTQITADRVLREAW---NQVTADARELTQLRIGCCRYCHGFEHRY
phi11_29_BP11_29028592              1  --MNEKQKRFADEYIMNGCNGKKAAITAG     YSKKTAE-SLASRLLRNVNV----SEYIKERLEQI-----QEERLMSITEA--------LALSASIARGEPQEAYSKKYDHLN----
gp172_BPSfi11_7669463               9  IGLNQRQKLFVSEYIKLG-NGTQAAINAG     YSPKTSG-AQAEKLLKKVEI----KNSINAEIEKM-----HNENIMDAKEA--------LSILSDIARGKRDEEIVMMNPVSGEVER
O1205p25_BPO1205_23455900           1  MGLNQRQKLFVSEYIKLG-NGTQAAINAG     YSPKTSG-AQAEKLLKKVEI----KNSINAEIEKM-----HNENIMDAKEA--------LSILSDIARGKRDEEIVMMNPVSGEVER
terS_BPAaphi23_31544027            16  SKLTNKQKQFIEEYLIDL-NATQAAIRSG     YSKDTAK-EIGYENLTKPYI----QQAIAEAQNKR-----SERTQIKQDDVI-------RMLLENIEIASGKKAVIKTEIR------
bIL310p27_BPbIL310_13095888         1  --MTPKQRKFCDEYIKTG-NATQSAINAG     YSQKTAK-SIGAENLTKPDL----KKYIDSKLKDI-----SNNAIATAEET--------LTILTQIVRGEHTEQVITAEGDVIDK--
A118p01_BPA118_16798784             1  MKLTEKQKRFADEYIKCG-NATEAARLAG     YSLKTAN-RIATENLSKPVI----KDYIANALEKL-----EADRVMDYTEA--------MQLLTEIARGEMTEKVVVTYGDNYDVI-
SPP1p001_BPSPP1_22855046            6  GKWTPKLERFVDEYFING-MNATKAAIAA  1 YSKKSAS-TIAAENMQKPHV----RARIEERLAQM-----DKKRIMQAEEV--------LEHLTRIALGQEKEQVLMGIGKGAETKT
SpyM3_1436_BP315.6_28876466         2  SKLTLKQKRFADEYIISA-NATAAAIKAG     YSKKTAR-SIGQENLTKPDI----KAYIDERLEKL-----ESEKIATQEEV--------LQYLTSIMRGDQQEKTLISVG-------
phig1ep38_BPphig1e_23455798         3  RKLTPKQQKFADEYIKSG-NAADAARKAG     YSKRSAR-SVGQENLTKPDI----KQYIDERMDEI-----ASKRIMDATEA--------VELLTRIARGEEKETVYIGTADGV----
bbp26_BPBPP-1_41179387             16  ARKTARMKNASTYTEAIA-TEICDRIADG  6 CRDEHMP-AWRTV-----------YHWI-EARPEF-----AERMERAR-----------RAGFDAIAEEALEIANTPQTGEET----
BMP-1p25_BPBMP-1_45569519          16  ARKTARMKNASTYTEAIA-TEICDRIADG  6 CRDEHMP-AWRTV-----------YHWI-EARPEF-----AERMERAR-----------RAGFDAIAEEALEIANTPQTGEET----
BIP-1p25_BPBIP-1_45580770          16  ARKTARMKNASTYTEAIA-TEICDRIADG  6 CRDEHMP-AWRTV-----------YHWI-EARPEF-----AERMERAR-----------RAGFDAIAEEALEIANTPQTGEET----
ORF7_BPphiHSIC_62362370            21  GAPTKYKPEYCELLMDMA-KEGASFTEFR     AAIGAVS-RQTLH------------NWK-EAHPDF-----LDAYTRAEAYG--------EAYWDKQMRKDLMYN-------------
BC1890_BPphBC6A51_31415784         10  SEMMAKLDELKQKLTAKQIQAAYLLVENE     -LMESNNEEKRTQDEMANELGINRWEWRTKNQDFI-----AFKSEVA------------DSFLAEKREQVYSKLMQLILGPQ-----
APSE-1_17_BPAPSE-1_9633564         12  SKLTVSLEKAKAYLMGEY-KTVGDVVPNI  8 ISRSTI------------------YEWSSSSNVEF-----SDI----------------VEGILALQENKLLNS-------------
W_BP186_9634057                    17  ILHNDSFDYQATWLRVGK-LNIDRSITKS  8 FSREALLDALTTGDN---------HVWFAHTIEHA-----RVALMYMSNLSA-------RVGVSLTSNGHSLQLDDGAVISFVGEES
PSP3p02_BPPsP3_41057354            17  ILHNDSFDYQATWLRVGK-LNIDRSITKS  8 FSREALLDALTTGDN---------QIWFAHTVEHA-----RVALMYMVNLST-------RVGVRLASNGYSVQLDSGATISLVGEES
933Wp50_BP933W_9632516             18  AETGITLLDWCRKKKINY-NTARTRIKMG  23 CNDAGSGDEKCAKNSEKNCANSAETKRIRGSRLLPPSNAFSQRNTHAVRHRGYAKYLEADNLMDDASDMVLFDELVFTRARALSVTK
VT2-Sap51_BPVT2-Sa_9633445         18  AETGITLLDWCRKKKINY-NTARTRIKMG  23 CNDAGSGDEKCAKNSEKNCANSAETKRIRGSRLLPPSNAFSQRNTHAVRHRGYAKYLEADNLMDDASDMVLFDELVFTRARALSVTK
Stx2Ip165_BPI_20065960             18  AETGITLLDWCRKKKINY-NTARTRIKMG  23 CNDAGSGDEKCAKNSEKNCANSAETKRIRGSRLLPPSNAFSQRNTHAVRHRGYAKYLEADNLMDDASDMVLFDELVFTRARALSVTK
Stx1p165_BPbacteriophage_32170999  18  AETGITLLDWCRKKKINY-NTARTRIKMG  23 CNDAGSGDEKCAKNSEKNCANSAETKRIRGSRLLPPSNAFSQRNTHAVRHRGYAKYLEADNLMDDASDMVLFDELVFTRARALSVTK
Stx2IIp168_BPII_32171171           18  AETGITLLDWCRKKKINY-NTARTRIKMG  23 CNDAGSGDEKCAKNSEKNCANSAETKRIRGSRLLPPSNAFSQRNTHAVRHRGYAKYLEADNLMDDASDMVLFDELVFTRARALSVTK
PHG11b_14_BP11b_53793590           16  VGVFLYLCFMAYDIEKTF-SSILSFIEDG  6 LKREGMP-DSTTF-----------YKWI-DADENK-----SIQYTRACETRA-------DVKFESIEKDYLEE--------------
gp3_BPHK620_13559865                1  MATEPKAGRPSDYMPEVA-DDICSLLSSG  6 CKRPGMP-DKSTV-----------FRWL-AKHEDF-----RDKYAKATEARA-------DSIFEEIFE-------------------
Sf6p01_BPSf6_41057279               1  MATEPKAGRPSDYMPEVA-DDICSLLSSG  6 CKRPGMP-DKSTV-----------FRWL-AKHEDF-----RDKYAKATEARA-------DSIFEEIFE-------------------
Consensus/80%                         .......s.pb..h..bbb..h.s..p..b..h.....hp..sh...b...............bb...b.........p..................hb.ph.p...................
```

```
SECONDARY STRUCTURE                    ----------EEE---EEEEEE------HHHHHHHH-----------------------------------HHHHHHHHH--
Tuc2009_28_BPTuc2009_13487829          ------TDAGLLV---TKEVTKQQAPDTTAAIFWLKNRK---------PNEWRDKSQTELSGGIKVSNPYEGLTEEELRKLANL   153
PBC5p48_BPPBC5_18071234                KLEYPEEIGKEAADKKVEDSLYKRATGYSFDSEKIVEGEVHR------AEIIEHVPPDTKAAMWWLQNRRPGTWRDVRHIKHDV   263
D3112p24_BPD3112_38229136              ------KPAVKVQLLASLADAYNKTVSASKRVLPETSALATA------MEVLQRLASFIRERFPQHAQAFAEVLEPFGEVIAKE   164
HK022p01_BPHK022_9634120               ------AQRRDLQIHSHATNGESRDQKKRNQNDRDA----------------------------RNTKNEHQDQDDNLIAFPKHG   161
pacA_BPP1_46401729                     EARKAARSGADVLETTLLEKSLSHLLMLENARDTCIRLVQEMNSCSAQISSLINTIYSIRNNYRKESREAEGGKVPPLMLEQIK   298
JL001p58_BPJL001_62327244              LE----LLERELFRRAVEGVLEPVVAGKDPEIVTYVRKY---------SDKLLELALKKADPSGYGNREAVNVNVNTGVLVAPE   191
phiA1122p14_BPphiA1122_30387467        R-----KLLDEDVKSIRESTMSNRKLAKLYGVSSTT------------------------IFHIKHNNKWRNLNGTSTIQTA--   152
orf5_BPPaP2_48697075                   ------GSATQVVVDGKSVKVTHLQIPAVCAPDFTY------------SDTKEIVAIGDRSTNPMPTPQLKGLLESYLGKADPS   186
ORF8_BPpsiM2_3249593                   ------ENGGEGLTLQTSKDLHNLGLAIALGLGDPT----------------------EIHEHRESGGELEDLLEDLVSATKNL   154
ORF8_BPpsiM100_11692720                ------ENGGEGLTLQTSKDLHNLGLAIALGLGDPT----------------------EIHEHRESGGELEDLLEDLVSATKNL   154
ORF148_BPphiJL-1_62327096              ------NAAKDIL---DRSMSYEQTRKIKADAELAE---------------------------LRVKQADGSADSQVVVNVHL   140
terS_BPTP901-1_13786561                ------EKEADMIKAIDTLNKMDGDYTSKVELSGSV----------------------------KTNPFVDLSTEELRKLASR   134
ul36_36_BPul36_21716107                ------EKEADMIKAIDTLNKMDGDYTSKVELSGSV----------------------------KTNPFVDLSTEELRKLASR   134
gp1_BPES18_62362214                    ------AETPQSSAAVAATMGKAKLVGLDKQIIDHT----------------------------SSDGTMATKPTTIRLVGVD   136
PaP3p01_BPPaP3_27476055                ------NEKGGIRLKAAQDILDRGGFGAKQKVELTT----------------KNVEDMSAEDLMNEVRRILDEEPDLAKVLSFP   150
Ljo_0307_BPLj965_41179239              IH----DKDGPYE---DAGGKFRYEPKIDPETGEQAYYY-----------EHLITLKDSEDIDTSNIKSIRIDKGEPVVEMEDR   241
epsilon15p01_BPepsilon15_30387380      P-----KVWRQYLTGFELADMFEGRGDEKELVGILK--------KIKWPDKVKNLELIGKHVDVNAFKERLEVSGTVTIADRMA   184
gp33_BPBCJA1c_56694901                 ITGKPLDKDGEQVMCSYSYVDFKNHDEVDGTIITEVKKGKDG-VSVKLADKMKALEFLAKYTDLLSDNDKKRLQEEKLKAETEK   248
BcepB1A_gene45_BPBcepB1A_48697530      ------YQWRDEHEFNEAMRAWENDEQARLAAKGPKRKGRPQCDGEGTPDVFIADTTKMSPEQRRLIERVKVTKEGIEIRFRNQ   205
Ljo_1440_BPLj928_41179313              ------KPKVEYDNVTTVTPEDQDQLKALELLGKYY---------------------------KIFTDKVETQTDITVNIDPG   147
g02_BPBcepC6B_48697192                 ------YQWKNERELAIAINSHARAIAAAKKAKVPVDAWPAPCFGEGVQHEVIADTRKLPAAVRALYTGLKRTKEGFEIKTQSK   202
F116p38_BPF116_56692920                QRTRN-EFDQDYAEWEADSGTDKGDFNVKGGPGFDPNAEPHPCAGDGVPRVVLMDTRNLSDGARSLFAGVKQTKFGIEIQMHSK   201
phi11_29_BP11_29028592                 ------DEVEKEVTYTITPTFEERQRSIDHILKVHG---------------------------AYIDKKEITQKNIEINIGEY   142
gp172_BPSfi11_7669463                  VTRK--ADNNTVIKAITEILKRYPTTKQAEKLQLEI---------------------------EKLKSQIGGDEGQDEKIAGFL   160
O1205p25_BPO1205_23455900              VTKK--ADNNTVIKAITEILKRYPTTKQAEKLQLEI---------------------------EKLKSQIGGNEGQDEKIAGFL   152
terS_BPAaphi23_31544027                ------KSEDGQLVGDDIAQFVYESSSVNRSLELLG-------------KHFGMFSEKVEVSGDLHIEQRNELNLSGLDINELE   172
bIL310p27_BPbIL310_13095888            ------HPDTNQVIRASAEILKRYPLTQDINIKGNL---------------------------KVSNPFENLTEEELRTLASR   143
A118p01_BPA118_16798784                ------DKEPEISQRINALKEIVKRHAAGGRDKLQEELI-------------------QAQIDKLRADTKQESNQGITTIIMSN   155
SPP1p001_BPSPP1_22855046               HVE---VSAKDRIKALELLGKLTQCLRTSRRSRLTKSSS----------LTIAAMQNEESKAFREVYTAFLRGVENGQSRSTPQ   174
SpyM3_1436_BP315.6_28876466            ------EFGQKIV---DIDVGAKDRIKAAELLGKRY----------------------------RLFTDKVEMDVSSDVTINVG   138
phig1ep38_BPphig1e_23455798            ------YEKHKEADLKTRISATKEILKRYPDNNKLVE---------------------QQIRKLKADADIAESKAKEVRSGESN   152
bbp26_BPBPP-1_41179387                 ------EESDDGYKVKRGDMLGHRKLQVETRLKLLAKWH-------------------PTKYGDKTSMELTGANGGPVQINDTE   162
BMP-1p25_BPBMP-1_45569519              ------EESDDGYKVKRGDMLGHRKLQVETRLKLLAKWH-------------------PTKYGDKTSMELTGANGGPVQINDTE   162
BIP-1p25_BPBIP-1_45580770              ------EESDDGYKVKRGDMLGHRKLQVETRLKLLAKWH-------------------PTKYGDKTSMELTGANGGPVQINDTE   162
ORF7_BPphiHSIC_62362370                ------KEVNAPLVKLYFANRFGWSDKTETKNDHTS----------------------------SDGSMTPSKIELVVATNDN   144
BC1890_BPphBC6A51_31415784             ------PSVKAMQLYMQRFGLLTDKKVIEGDLGNAT----------------------------RTNAEIEEQLQKLKKLTGE-   155
APSE-1_17_BPAPSE-1_9633564             ------GLKGEFNPTITKLMLSKHGYADKQETELSG----------------------------KAGGAIKTDNKISIEFIGIP   132
W_BP186_9634057                        HCAA--LAGNVYL---DEFGWFNNPLRAAKVAAGIACHKRHS-------LTMFTSPSDNYDAFRVWNGTTRRHRPSPLINTGDS   190
PSP3p02_BPPsP3_41057354                HCAA--LAGNVYL---DDFGWFNNPLRAAKVAAAIACHKRHS-------LTMFTSPSDNYDAFRVWNGTFRRHRPTPLINTGDS   190
933Wp50_BP933W_9632516                 ALKGMFADLEEATDVETRVALYDKILKAEQALDRNIARIESIERSLLTLDVLAETAPKLRADRERINAARDKLRAETDILTNQR   239
VT2-Sap51_BPVT2-Sa_9633445             ALKGMFADLEEATDVETRVALYDKILKAEQALDRNIARIESIERSLLTLDVLAETAPKLRADRERINAARDKLRAETDILTNQR   239
Stx2Ip165_BPI_20065960                 ALKGMFADLEEATDVETRVALYDKILKAEQALDRNIARIESIERSLLTLDVLAETAPKLRADRERINAARDKLRAETDILTNQR   239
Stx1p165_BPbacteriophage_32170999      ALKGMFADLEEATDVETRVALYDKILKAEQALDRNIARIESIERSLLTLDVLAETAPKLRADRERINAARDKLRAETDILTNQR   239
Stx2IIp168_BPII_32171171               ALKGMFADLEEATDVETRVALYDKILKAEQALDRNIARIESIERSLLTLDVLAETAPKLRADRERINAARDKLRAETDILTNQR   239
PHG11b_14_BP11b_53793590               ------PQRDPVSGKIDAGWVQLQRLKIDSKKWELGKLQ------------------PKKYGDKIQNEISGAITTNIISLGSG   156
gp3_BPHK620_13559865                   ------IADNAIP---DAAEVAKARLRVDTRKWALARMN------------------PRKYGDKVTNELVGKDGGAIQIETSP   133
Sf6p01_BPSf6_41057279                  ------IADNAIP---DAAEVAKARLRVDTRKWALARMN------------------PRKYGDKVTNELVGKDGGAIQIETSP   133
Consensus/80%                          ......................p............h.................................b...p...............
```

Figure 6. MuF alignment

```
SECONDARY STRUCTURE                     --HHHHHHHHHH---HH----HHHHHHH---HHHHHHHHHHHHHHHHH-----HHHHHHHHHH-----------HHHHHHHHHHHHHH
VV1_0114_Vvul_27359714          114     IDAAVLATRTMQQKDGL----QLSDRLWRV-HRNAKQELTNAVERAVI-LGQSASEAAQDYQRRMQPSHIAQQMGINQIKRVMRTEINRA
MA_3823_Mace_19917898            85     SAAAAKPLPITEEAV------RQASSL----SKEVTESVIRILKDDGL-YLEHPNKLEKRVRDIWG-------NQKHKAIRFTRTFADV
18_BPBarnyard_29425718          127     ESQRQAARNGVAHAISSKSR-EPLSARVYRTRSLANNWIQRLVTSSIV-RGDSAQDIAKAVRAHIRPGTPG--GTAYAALRLGRTELNNA
PHG11b_18_BP11b_53793594         92     LREIVLYLQSNGGVKI-----TSVSDTLAN-DIIKAIRAQ--LGENAT-VIQLQQAILKVVNNTQT-------FYRYQALRIARTETTTA
lp_0663_Lpla_28377535           142     RRALYRVTAQNEPEN------TSLDLIFKH-ANKLAIDLDNIIKFQMQ-NHVNPNSISKIVAKELGV------AGKPDMERILVTESKAT
ABC2837_Bcla_56964601            99     LHREALEQIVNDTMDDIKAAVEVAKGYVKADVEYVRQTFKDDLAKGSI-QGDHNRVISARLTERFKGIVGHRTTIDFYASTVVRTKLRDA
SPy0975_Spyo_13622132           172     SVKSLADAVINANFKGA----KWSDNIWDR-QDKLRSIISQSVQSAIL-KGKNGLTIARDIRREFD-------VSASYAKRLAITEHARV
BC_1897_Bcer_29895584           108     VNKQIIQGVIKTGEDGL----VLSDRVWNL-SGDMRDRLSSVIRPSVL-KGESITMISQKIKEVHD-------NEKWKVERVAISESTNT
gene27_BPBcepB1A_48697516       138     PLHDMIDASVKENI-------ALIKRV----PSKYFDQMQGDVMRSIQ-MGNGLQDLIPAFEKREV-------QVRNWAQNVALDQTRKA
pas6_BPphiAsp2_48697407         103     PAVVRYLTEVR----------NRLVRI----PDEVYDLVAGELAAGIN-LGDTIPELAARVDSVLSRTASE--RWPNRATVIARTETIGA
M6_Spy0044_Spyo_50913390        162     INGPVLEELVRTPFDGY----NYSEQLWGN-TDNLVKNLQKRLKQGFV-RGEHPRAMARDLAKRFN-------VANHRAETLIRTDGTMV
RoseDRAFT_2913_Ssp._69298219    110     PAADQILAAVNRPFQGK----LLKEVYSEL-SASSFRKVRDTIRAGFV-EGRTTDEIVRDLRGTKAFKDGVLDTNRRATETVVRTAVNHT
BP3382_Bper_33594267             96     TAGQVYAGAMAQPFQGR----LLREWMAGL-EAGRASRIRDAVRMGYV-EGQTIQQVVQRVRGTRAYADGLLEIDRRNAEAVVRTAISHT
orf13_BPKS7_57472354            106     AAEQVWAAATFAPLELSEKPVDFVSLMSGW-RQTEVNRLVMGVKSGFV-QGMTTRRIVKNVVGPGGLAD----ISERNAATVIRTALAHV
PFL_3764_Pflu_68345423          112     PGAAAIRAAILSNPLSVVDGGKLLKSFIDGFTTTERQRLTGAIRQGFF-EGQTNFQIIKNIRGTKAYNDGILATTNRNAGAVVRTAVQHV
JL001p62_BPJL001_62327248       121     PAADLLQSIVTRPFEGQ----TMRQWASNI-RNSDLRRIERAIRIGVT-QGEPTNAIARRVIGTVRGRDGVTQITRRNAEAITRTAINGI
6_BPES18_62362219               122     GRESLAAILASDPYQQRLVHARVFEEMKGL-TADVKRDMARVLTDGVG-RGLNPSDIARNLTAQAG------IEKRRANRIARTEVTTA
EF2015_Efae_29376530            166     LTKNELENIMNYPWSGK----TFSSRLWGN-ISTLEQRLSNSIINSLA-SGEGVVEALRTMKNDGVSGMFKLEQGKFSIENLVRTEYSHF
MJ0329_Mjan_15668503            481     YDDLDQQTIDILKQHWE----QAFYNI----YSSYSQQFLDVLTEGIQ-KGLGEEEIAKNLKKVA--------KDVKGGSRLQMRA-REEM
MesoDRAFT_2473_Msp._68191497    163     GGVIGLTTAQSEYVARAELLSGEPDQLHYLSRGRRDKRFDRVVTAAIK-EGKPPAETVNRIVGRYSDR-----LLELRGEMLARTETMTA
EF1457_Efae_29376024            115     QTEKWLDSYVQKIAENI----SYVTNI----RDDYSKKFEQVIYRGIT-EGKSSNEIREELVHQAG-------MSSDKAAFIARDQTGTI
lin1730_Linn_16800798           116     QSEPWLSEFMDASIENV----SYIGSI----ADEFNAKTEQIILRGVK-SGQSLKDMRDELVTQVG-------TSKNKAEFIANDQTGTI
Saro02002195_Naro_48848936      129     DVRAPLGTVIERNVELV----RSV-------SDQARARISDTVFRGLA-EKKPARAVAAEIREAVA-------MGRRRALNIASDQLSKA
EF0335_Efae_29374974            116     INHDYIMNLVNAPVAGK----RLSKRLYKY-RDELAQNVTNNIITGLF-EGKSYAEIARWINEETE-------ASYKQALRIARTEAGRT
SH2356_Shae_68448081            129     PTQAIIQKAIEQPVDLI----KLVPTLERH-RNNVLKKIRIEITKGIV-NGKGYTHIAKALRDDLG-------MTKAQAQRVARTEAGRA
Bcep1-17_BPBcep1_38638624       138     PSQRLVMKAKVENV-------ALIRSI----SQEYHTAVEGEVLRHFI-AGRDLKGLQDKLVERGK-------VTTNRAAFIARDQCNKA
plu3404_Plum_37527278            98     EMNNALQTVIGENV-------GLIKSI----PRQYLSQVETLVMQSVA-RGRDLGTLTKALQERYD-------ITKRRAVLIARDQNNKA
DRA0097_Drad_15807765           111     TVPPTLQLEWMDDAER-----RMREFWGNE-PARFRQEIHDALVDGLK-RGQGIDQMSKRIQDRVG-------VSRSRASLIARNEVGNA
MS0100_Msuc_52306207             23     YLDVLGE-EHDYFF-------MVAGAN----RNEVLTALREAVDAAVL-KGETLDGFRRRFDDIIANWEYN-GGRNWRTRIIYDTNVYGA
psiM2p12_BPpsiM2_9630536         26     DVLTRLKVTDVESFL------SDFHLL----TKAEFMRRYGVSEREYN-ELQAALQVVEDLNLPTFEDEDLI-IWFMGLAVVYLARDIERD
Psyc03001496_Psp._52853963       28     YYNVMTPIQRQQAV-------SIAGLG----QTEQIKHVMGLVNNQLD-SGGTFADFQNAVDIDINL-------PRHRLDNIFRTNIQGA
CJE0251_Cj_57237261              29     YDELSYSTHKKVFT-------I-AKLM----DESLLKDMQDTLVNAIK-NGDKFSTWSKIAEEKLKAKKTH-FNSARLKK-IFEENSRKA
ORF240'_Dnod_924617              18     FYALQAKLKQEAWT-------M-SKIA----QLKQIQGVLDDLAQNIQ-RGGTFEEWQKNY-------------PRHYLETVYRTACLRA
B3ORF33_BPB3_56692601            26     DYNAIRGAANDRSF-------VSAGAN----RADLVADLQAIIDQAIR-DGMTLAEFRKDYDAVLDHWEPE-GGRAWRARVIYETNLRTS
HD_1557_Hduc_33148823            28     WTDIYNAEHDYAAV-------V-AGAN----RREIIEDFAKSIQDFIE-NGKTLEDFRKDFDKIVAKWQYN-GGRNWRSRIIYETNLRSS
NMA1850_Nm_7380487               29     WDVKWNE-AQTKAR-------TIAGIH----RQDIVGEFHAAMYEAAK-SGKSFEAWRDEVQGRLKATGEV-LGRGITKHTIFRTQMQSA
DVU1119_Dvul_46579532            31     VFKSLEAGARARAF-------AVVGLA----RLDQVSALQTALADALE-QGETLASFQKRAAAIIEE-------AGWKGRNIFRTNVQQA
Ddes02000523_Ddes_53692070       27     VFKNLSGQARARAF-------AVSGLS----RQDQIAAVQQAVHEAMA-NGETLKDFKGRMDAVLEGARL----PRWRLENIYRTNVQSA
BB3630_Bbro_33602605             32     ALDAWDA-ARERAF-------TVTGIA----RLDVLQDIKGELQRSLN-DGKTYASFKAGVEDTLRTEVAA-NLPGYRLKNIFRTNMQSA
SO2681_Sone_24374222             27     YRDVWKE-EQANAF-------TVAKML----NADLLVEVKQLVEDAIG-QGQTFEQFRDILKPLLVKKLVQ-LGSEGRIKTIYKTNMRTA
NMB1096_Nm_7226334               29     SYRNLTASEIAKVY-------TIARMT----DLDMLNDIKTSMVESAK-SGQSFDDWRKGILNLLSNTGEV-FGSPRRLETIYRTNMQTA
HD_0515_Hduc_33147939            30     LKKELHESALARAT-------TIARLT----SLEMTKDIYASMETARR-EGKSFNTWKKGLDEMERKTGEY-FGTPRRLNTIYRVNMQAA
Haso02001213_Hsom_32029860       29     DEKGLLESARAKAT-------RIANIT----SLEMTKDIYQSLIIAQE-QGQSFGEWKKGIFEHFKKTGEF-FGTPRRLETIYRTNMQAA
BproDRAFT_1408_Psp._67909833     28     WQDTFNE-EHNRAF-------TVAKMA----DIDLLTDMRKAVDKAIS-DGLTLEQFKKLIEPTLQETQAE-LGSPRRLRTIYRTNLKTA
SO0668_Sone_24372259             37     WQDVWTR-AHARAF-------TVAKAT----QMDVLTAIRNEVDAALS-QGLTAKQFQANLKPQLEKREVQ-LGSPYRLNTIYRQNLQTA
Bcep1808DRAFT_4120_Bvie_67545322 30     WQDVAAE-AHAKAF-------TVAGVM----KVDVLQDIRQALTTSLK-KGTTFDEFKRQLSPVLEKTGEI-EGKRLTPRTIFQTNMQSA
DVU2707_Dvul_46581111            32     WHDLWQE-AHARAF-------TVAGVA----RLDVLEDIREALTTHLK-EGKGEYRFHKQLEPTLRSRVVR-MGSPARLRLIFRQNIQTA
t1371_Sent_29141825              33     WEEMWQE-AHARAF-------TVAKVT----RLDILEDIRGALQQAVD-EGKTDRWFRQELEPVLKRVTIQ-QGSPWRLDTIFRTNMSVL
12_BPU2_40769326                 60     ARRFYDSQRAQHHP-------DLPRND----RPLEGTTFEK-FVENM--DPAR-ERM-QQSDT-----RGD-ALTHLTLRAVREVE--NA
Blon03000705_Blon_23335596       82     DAVRSAWAEA-AGVDLP----AYTPSR----VSADRAFWQK-FVDVIN-HHSR-AGL-TMDDLWAMK-TDG-YGQDEWMNLAADVG--VT
TM4_06_BPTM4_18496892            70     ARQQPAGKSANAQL-------AGDFRG----GPANTQGGKL-FVPEPAPLPDP-DRL-GANARWALL-QND-PVVALQGSATRAVM--DS
KradDRAFT_2833_Krad_67987746     71     EAFRAAELDALADIDIDPGERPRPRRP----ATAADEDLPE-FLRDDVDDGVR-RRI-ERGEEAPRA-LEV-AGVEVAGAASRQTL--YG
Consensus/90%                           ....................................h...h...h.........h...h................h..hh.s....h

SECONDARY STRUCTURE             HHHHHHHHHH---------EEEEEEE-------HHHHH     H----EE----------     ------     ---EEEEE----
VV1_0114_Vvul_27359714          YGMAFQNSAFE----DEFVVGTKFKLSPN-HPKRDICD  7  YGLGRGVYPRGKS----     PWPAHP     NTLSYEQVVFVD   293
MA_3823_Mace_19917898           ATNTELWRYQD-----SGIDDLQFYAKID-DKTSPQCR     MLHGTIFRADSPEVRRY     RPPLHF     HGRSDLIPVPVT   223
18_BPBarnyard_29425718          FHATSIEISKD----RPWVEGMRWNNSIH-EPDMEICT     RLNGQIYTVETV-----     PKKPHP     QGRGFVTPNLIP   280
PHG11b_18_BP11b_53793594        SGFAAIQRAKE----SNLILEKKWIAVND-SRTRLDIT     LEQGQKVDLDKPFIMAD  10 NGGASQ  2  NGRGTISFAGKR   245
lp_0663_Lpla_28377535           QTRECAKQYSN-----LGFTKLKIVTRDN-PHVCKGCE     GHDGTIVEIKDAVVGMN     VPPLHP     RGHGNVIPVQMD   305
ABC2837_Bcla_56964601           NTKGAVNRYTE-----NGVDLVKIS--GQ-GQTCQVCR     -YQNMVISLSGKHEGFA  8  LPPFHP     NGRGTCRPYVIE   269
SPy0975_Spyo_13622132           QMEVGRLSMAE-----NGFAMFDIL--PE-PKACDVCK     -DIAK-HGPYHLDKWRI  3  SPPFHP     YGRGAIVGVDES   314
BC_1897_Bcer_29895584           YRAATIQNGLE----SEIVTGYQIIDNHR-HRYKHMCK  8  LGAGKYPKNIPESLMNQ     LISPHP     QGSSRLNYLISE   266
gene27_BPBcepB1A_48697516       FNGINAGRMQA-----LGMDEYEWIHGGS-NHPREYFD     ELNGNIYKLSEPPIIDK  6  KPGQLP     FGRGTMRPVFRL   284
pas6_BPphiAsp2_48697407         LNFGRGEAYATLAEEDPDLELELWLATED-HRTRATFV     AADGQRVPVKGRFIVGG  3  AFPGDP  8  QGRGTMLLVEPG   260
M6_Spy0044_Spyo_50913390        INNATARRYLN-----AGLKYYRDLVRLD-DRTTEICR  2  AKENKRKLLSELKPGIN     VAPYHF     NGRGTIIPDEDE   307
RoseDRAFT_2913_Ssp._69298219    ANTAREYTYERN---ADLVKGVRWNSTLD-GRTSAVCR     ARDGKVYPDGNGP----     RPPAHF     NGRGSTSPVLAS   262
BP3382_Bper_33594267            AGFARDRWYDAN---DDIVGALAWVSTLD-SRTSQMCR     LRDGLRYEPDSHKPIGH  7  PGRLHW     QGRGTSVPILKG   258
orf13_BPKS7_57472354            SNEARQQVYAQN---GDIITKYEWVSTLD-SRTSAVCR     SRDSMQYEIGKGP----     LPPAHP     NGRGSTAPVISP   254
PFL_3764_Pflu_68345423          ATQARMETLKEN---SDVVQAVEWVSTLD-SKTTPQCR     TLDKHRFKLTEGP----     RPPIHI     NGRGTVVAVTRF   270
JL001p62_BPJL001_62327248       ANQAKREFYKAN---ATLFEEEIYVATLD-SRTTPICR     SLDGQRFAIGEGP----     IPPLHF     NGRGLRVAAIDG   273
6_BPES18_62362219               LRRAKWDEDQEA-NDLFGLKTLLVHISLS-PTTRHTFA     VRHAHLYTNEEVREWYA     KDANSI     NGKGSQQSVLVD   277
EF2015_Efae_29376530            AVEGVRKSLKD-----IGVKQTSWSA-ED-ERVCSICG     GRHGKEIK-DDW-----     HPPYHG     RGRGTEIPIIPE   314
MJ0329_Mjan_15668503            NKTYNLTRARR-----FWNDKVIYVTMKD-ERVRPSIR     KLHGLIFVPAERPEL--     VPPLGY     GGRGTITPVRD-   616
MesoDRAFT_2473_Msp._68191497    LGKSRDDAIRQQIAAGKIMVEDLWRSAGD-SRVRHTFR     ILSGKSAPMDGFFQSPS  4  RYPGDP  8  GGRGWVEYKIDH   338
EF1457_Efae_29376024            LGQMNSERQKR-----AGFQAFRWSDSGD-ERVRDSIR     ERNGKIYFYADNPL---     LPGEEY     NGRGVAEPVDDE   254
lin1730_Linn_16800798           FGQMTAERHKK-----AGIPGFTWRDSGD-AAVRPSIH     ARNGKVYSYDDPTAP--     IPGTDY     RGRGTAGPEFDE   256
Saro02002195_Naro_48848936      AETLNEERRRQ-----AGLMAWEWVSSHK-VNYRPEIA     ARDGKRYS-DDPADGA-     PPPQDR  5  FGGGTSRAVLSL   269
EF0335_Efae_29374974            QSVTTQKGYEEA-KELGINIKKKWLATID-KHTRRTIQ     ELDGKEVDVDEEFTIRG  10 VASEDV     NGRGTTIEVVDG   273
SH2356_Shae_68448081            MSQAGLDSAKVA-KDNGLSMHKRWLATKD-TRTRHSIR     HLDGQSLRLHEVFHSSG  11 DASAKE  2  NGRGKLLYYIDE   289
Bcep1-17_BPBcep1_38638624       TAQFNSARQRE-----LGLHWATWQHSAG-KEPRPNIM     RA-GREKWIFRTNQVGID  8  LPGEAI     NGRGSSRTIIPG   285
plu3404_Plum_37527278           TFVMQSARQRS-----MGFKTGIWRHSAA-KKFRQSIL     KADGKEFDLSKGLYIDG  3  MPGEEI     NGGGGWQVVIPK   240
DRA0097_Drad_15807765           AAYATQKSQAQ-----AGCAEYIWRSAKD-RRVRPEIA     KRDGKRFSWDDP-----     PPDGHP  4  NGRGVALAVIPG   252
MS0100_Msuc_52306207            YNRGRLQQHLDM---AEDMPYWEYQHNDN-AHPREQIM     AWDGLVLRYDDP-----     WWRYHY  5  GGHGTVVAHDEA   170
psiM2p12_BPpsiM2_9630536        TEIQAYCQHSEA---HGGELYLNWVTMGD-DHVCETCQ  1  YEDGSPYRVEEF-----     PHIPHP     FGRGHPEPCDEN   172
Psyc03001496_Psp._52853963      YGRGRWYQQQQN---KDERPYLMRDGIND-IRQRPAIK     VLDGVVRHIDDPFWATH     YAPSGY     RGRGIMRSLTES   169
CJE0251_Cj_57237261             KAKAIY--ENQM---KSTKPYFKYCTQKD-SLVRDKIR     AFDGIVLHKDDPFWDSH     YTMHDY     GGRGYVLEVGES   191
ORF240'_Dnod_924617             YNAGKWVQFRAQ---KENRPILRYSAIND-SRTRPHIK     ALHGFMAPVDDAVWRTL     AAPNGF     NGRGTLMSLSAR   156
B3ORF33_BPB3_56692601           YAAGRYAQLQAV---KGERPFWMYNHSAV-EHPRELIL     AWDGLVIHADNPWWRTH     YPPNGW     GGQGSVSAYAED   175
HD_1557_Hduc_33148823           YQAGRYAQLQEL---KETMPYWEYVHNAV-TYPRVEIM     HWDGLILRHDDPWWKTH     FPINAW     GGQGTVIARSQA   174
NMA1850_Nm_7380487              YMAGHWQAFEEG---RDDAPWLQYSAILD-SRTRQSIA     AAHGAVYHIDDPFWDYF     YPPNGF     NGRGTVRALSDS   192
DVU1119_Dvul_46579532           YQAGRYVQMRRA---SKARPYWQYVAVQD-RRTRPAIA     VLDGKVYPADHEFWNAH     YPPNGF     QGRGTVRSLSPR   174
Ddes02000523_Ddes_53692070      YMAGRYAQMQRT---TALRPYWRYVAVAD-KRTRPDIL     ALHGLVYPHDHKFWSTY     YPPNGF     AGRGTVQTLSER   169
BB3630_Bbro_33602605            LMAGKYRQLSQQ---VDIAPFWQYVAVMD-GKTRPGIA     AMHGKVFRYDDPVWNSH     YPPCDY     QGRGGIRALSQA   194
SO2681_Sone_24374222            RSAGQWERIQRT---KRTMPYLLYQLGPS-EQHRIEIL     KWNNTLLPADDPWWDVH     MTPNGW     GGKGWIRQVSQY   189
NMB1096_Nm_7226334              YNAGQYQGYMAN---IDARPYWMYDAVGD-SRTRPAIS     AIDGLVYRYDDPFWATF     YPPNGY     NGRGSVIALSER   193
HD_0515_Hduc_33147939           YSGARYQRMMDN---VDNRPYWKYSAVSD-SRTRPAIL     ALNGRVYRYDAPFWATF     YPPNGF     NGRGTVIALNAR   200
Haso02001213_Hsom_32029860      YSAQRYQQMRDN---VDSRPYWQYSAVND-DRTRPSIS     AMNGLVYAYDDPFWHTF     YPPNGF     NGRGTVIALAER   192
BproDRAFT_1408_Psp._67909833    YAAGQWAQIEET---KAEAPFLLYDAVDD-DRTRPEIA     EWDNTILPIDDAWWSFH     TPPNGW     EGRGGVIQLSKA   190
SO0668_Sone_24372259            YMAGRYRRMLSR---TKTHPYWQYVAIDD-GQTRPAIA     RLRGKVFRFDDPIWDII     YPPNGW     GGRGRVRALTEA   193
Bcep1808DRAFT_4120_Bvie_67545322 YMAGRYAAQLEQ---VDTHPYWEYVAVLD-SRTRPAIR     ALAGAIYRYDDPFWQTF     YPPNGY     KGRGRVRTRTRA   191
DVU2707_Dvul_46581111           YMAGRYRQMHEN---AEARPFWQYVAVLD-SKTRPGIK     MLHGRVFRHDDPFWHSH     YPPNGW     GGRGRVRALSPH   193
t1371_Sent_29141825             YSAGRWAEQMEN---VDDRPYWMYTGIND-SHTRRSIL     ALHGLVLRWDDPFWQAF     YPPNGW     RGRGSVIALSAT   195
12_BPU2_40769326                GRQQIIHAVEND---PEPRVLRGWARVATGRETCAWCL  17 DLDTEHALELF--ENKD  11 IKQWHP     GGDGKVIPVFRN   221
Blon03000705_Blon_23335596      ARLTAKFNGEHD---PS---QPRYARVPVGP-TCAFCI  2  ASRGFVYWSEEKAGGRD     -NRYHK     NDDGRIVSSWGE   236
TM4_06_BPTM4_18496892           SRRTVLDNAKRE---GV-----RWVRY-ASVTACGFCR     MLATRGAVYKSADT---     ALRSHD     HGVGLAVPD-RN   201
KradDRAFT_2833_Krad_67987746    GRKAMLELVRED---RA---ALGYIRV-TDFQPCYFCA     MLASRGPVYKSDSFAES  8  KQKVHD     HGGGTIESVYSV   260
Consensus/90%                   ....................b.......psp..p.       ..ps..h..........     ......      .Cps........
```