Supplementary information S1 (box). Supplementary Methods description.

Prokaryotic Genome Database

Archaeal and bacterial genome sequences were downloaded from the NCBI FTP site (ftp://ftp.ncbi.nlm.nih.gov/genomes/all/) in March 2016. For incompletely annotated genomes (coding density less than 0.6 CDS per kbp) the existing annotation was discarded and replaced with Meta-GeneMark [1] annotation. Altogether the database includes 4,961 completely sequenced and assembled genomes and 43,599 partially sequenced genomes coding for 1.85×10^8 proteins.

Clustering and Phylogenetic Analysis

To construct a non-redundant, representative sequence set, sequences were clustered using the NCBI BLASTCLUST program [2]. (ftp://ftp.ncbi.nih.gov/blast/documents/blastclust.html) with the sequence identity threshold of 90% and length coverage threshold of 0.9. The longest sequence was selected to represent each cluster. Permissive clustering of sequences was performed using UCLUST [3], with sequence similarity threshold of 0.3.

Multiple alignments of protein sequences were constructed using MUSCLE [4] and MAFFT [5] programs. Sites with the gap character fraction values >0.5 and homogeneity <0.1 [6] were removed from the alignment. Phylogenetic analysis was performed using the FastTree program [7], with the WAG evolutionary model and the discrete gamma model with 20 rate categories.

Relationships within diverse sequence families were established using the following procedure: initial sequence clusters were obtained using UCLUST [3] with the sequence similarity threshold of 0.5; sequences were aligned within clusters using MUSCLE [4]. Then cluster-to-cluster similarity scores were obtained using HHSEARCH [8] (including trivial clusters consisting of a single sequence each); a UPGMA dendrogram was constructed from the pairwise similarity scores. Highly similar clusters (pairwise score to self-score ratio >0.1) were aligned to each other using HHALIGN [8]; the procedure was repeated iteratively. At the last step, sequence-based trees were reconstructed from the cluster alignments using the FastTree program [7] as described above and rooted by mid-point; these trees were grafted to the tips of the profile similarity-based UPGMA dendrogram.

1

Pipeline for CRISPR-Cas loci annotation

The pipeline takes a list of locations (coordinates in the corresponding nucleotide sequence) of the seed features as input. Two types of seeds were used: locations of *cas1* genes in the NCBI WGS database and locations of CRISPR arrays in the WGS database and the prokaryotic genome database. TBLASTN searches [9] with the E-value cutoff of 0.01 and low complexity filtering turned off were run with the Cas1 profiles [10] as queries, resulting in the identification of 20,766 loci. The CRISPRfinder [11] and PILER-CR [12] programs were used with default parameters to identify CRISPR arrays in the WGS database (47,174 loci found) and in the prokaryotic genome database (45,373 loci found). Sequences including up to 10 kbp upstream and downstream of the seed features were extracted.

Open Reading Frame (ORF) annotation was performed using Meta-GeneMark [1] with the standard model MetaGeneMark_v1.mod (Heuristic model for genetic code 11 and GC 30). All ORFs were further annotated using RPS-BLAST [13] searches with 30,953 profiles (COG, pfam, cd) from the NCBI CDD database [14] and 217 custom Cas protein profiles [10]. The CRISPR-Cas system (sub)type identification for all loci was performed using the previously described procedures [10,15].

Among all permissive clusters constructed from proteins from the seed loci, potential candidates were selected using the size threshold (> 500aa), the distance to seed (genes closest to the seed were preferred); the selection of candidates was limited to those that were located within 4 genes from the seed; clusters that contained more homologs outside the seed loci than in those loci were discarded. Additional prediction of protein domains was performed using the CD-search [16] and HHpred [8].

The identified candidates were used as queries for a PSI-BLAST search against the NCBI NR and NCBI WGS databases for the Cas1 seeds, and NCBI WGS and prokaryotic databases for the CRISPR seeds in order to obtain additional loci that were added to the seed list. The evaluation procedure was then repeated until convergence.

Protospacer analysis

The initial pool of 488,437 spacers in the CRISPR arrays was reduced to 268,409 unique sequences. The MEGABLAST program [17] (word size 18) was used to search for protospacers in the virus subset of NR database (TaxID:10239) and the prokaryotic genome database. The maximum number of mismatches for a spacer with length l was limited to max($0, \sqrt{l-22}$). All MEGABLAST hits that target CRISPR arrays as well as all eukaryotic virus sequences were discarded. This procedure produced 63,939 hits to prokaryotic genomes and 5,095 hits to prokaryotic viruses. The 33480 ORFs that contained or intersected with the detected protospacers were used as BLASTP queries to search the virus database. All ORFs with strong hits (e-value $<10^{-6}$) were classified as originating from (pro)viruses.

Synteny analysis of subtype V-U loci

Protein sequences encoded by genes in the vicinity (±3 genes) of the Type V-U effector genes were extracted and clustered using UCLUST [3] with the similarity threshold of 0.3. Genes were annotated by the cluster IDs; each locus was represented as a set of genes and unordered gene pairs. Weighted Jackard similarity coefficient was calculated for all pairs of loci as previously described [15], a locus similarity graph was constructed with the similarity threshold of 0.61 ($e^{-0.5}$), and connected components (subsets of highly similar loci) were extracted.

Analysis of selection in the evolution of Class 2 effector genes

Nucleotide and protein sequences of the effector genes were collected; clusters of identical protein sequences were reduced to a single representative; remaining sequences were clustered using UCLUST [3] with the similarity threshold of 0.67. The sequences from each cluster were aligned, and a phylogenetic tree was constructed as described above and rooted using a modified midpoint procedure. Sub-alignments of protein sequences, corresponding to sub-trees with the mean depth <0.1, were extracted and converted to the nucleotide sequence alignments. Pairwise *dN, dS* and *dN/dS* values were obtained using the codeml program of the PAML package [18]. Sequence pairs with 0.0002≤ *dN* ≤1.0 and 0.0002≤ *dS* ≤1.0 were selected, and the *dN/dS* values were calculated.

References

1 Zhu, W., Lomsadze, A. & Borodovsky, M. Ab initio gene identification in metagenomic sequences. *Nucleic Acids Res* **38**, e132 (2010).

2 Wheeler, D. & Bhagwat, M. BLAST QuickStart: example-driven web-based BLAST tutorial. *Methods Mol Biol* **395**, 149-176 (2007).

3 Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460-2461 (2010).

4 Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-1797 (2004).

5 Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**, 772-780 (2013).

6 Yutin, N., Makarova, K. S., Mekhedov, S. L., Wolf, Y. I. & Koonin, E. V. The deep archaeal roots of eukaryotes. *Mol Biol Evol* **25**, 1619-1630 (2008).

7 Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* **5**, e9490 (2010).

8 Soding, J. Protein homology detection by HMM-HMM comparison. *Bioinformatics* **21**, 951-960 (2005).

9 Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389-3402 (1997).

10 Makarova, K. S. & Koonin, E. V. Annotation and Classification of CRISPR-Cas Systems. *Methods Mol Biol* **1311**, 47-75 (2015).

11 Grissa, I., Vergnaud, G. & Pourcel, C. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res* **35**, W52-57 (2007).

12 Edgar, R. C. PILER-CR: fast and accurate identification of CRISPR repeats. *BMC Bioinformatics* **8**, 18 (2007).

13 Marchler-Bauer, A. *et al.* CDD: a database of conserved domain alignments with links to domain three-dimensional structure. *Nucleic Acids Res* **30**, 281-283 (2002).

14 Marchler-Bauer, A. *et al.* CDD: conserved domains and protein three-dimensional structure. *Nucleic Acids Res* **41**, D348-352 (2013).

15 Makarova, K. S. *et al.* An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol* **13**, 722-736 (2015).

16   Marchler-Bauer, A. & Bryant, S. H. CD-Search: protein domain annotations on the fly. *Nucleic Acids Res* **32**, W327-331 (2004).

17   Zhang, Z., Schwartz, S., Wagner, L. & Miller, W. A greedy algorithm for aligning DNA sequences. *J Comput Biol* **7**, 203-214 (2000).

18   Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**, 1586-1591 (2007).

SUPPLEMENTARY INFORMATION

Supplementary information S2 (box). Supplementary Excel Files

located at the following FTP site:

ftp://ftp.ncbi.nlm.nih.gov/pub/wolf/_suppl/CRISPRclass2NRM/

Supplementary information S2 (box, part a) (MS Excel): Pipeline output for all protein families associated with CRISPR arrays. Protein clusters ±10 kb vicinity of CRISPR arrays, their annotation (if any) and representative sequences. All clusters are sorted by the relative frequency of genes in the CRISPR loci.

Supplementary information S2 (box, part b) (MS Excel): Pipeline output for all protein families associated with CRISPR arrays. Protein clusters ±10 kb vicinity of CRISPR arrays, their annotation (if any) and representative sequences. All clusters are sorted by the relative frequency of genes in the CRISPR loci.

Supplementary information S2 (box, part c): Class 2 loci. For each Class 2 effector gene the surrounding genomic locus is given. Protein-coding genes and CRISPR arrays are shown. Genes annotated in GenBank are identified with GenBank locus tags; genes annotated de novo are identified by contig IDs and gene numbers.

Supplementary information S2 (box, part d): TnpB family FastTree in the newick format. Complete tree used for the Figure 3A is provided. Sequences are denoted by a local GI number, species name and those that are located next to CRISPR array marked by "CRISPR" prefix. More details on the sequences could be found in supplementary information S2 (box, part g).

Supplementary information S2 (box, part e) (MS Excel): CRISPR array spacers. Unique spacers were retrieved from all CRISPR arrays in supplementary information S2 (box, part a). Similarity searches were performed using MEGABLAST (see supplementary information S1 (box)). Hits are annotated as follows: Spacer ID (column 1) includes contig ID where this spacer was found, CRISPR coordinates and position number of the spacer in CRISPR array separated by underscore. For the hits (column 2,3,4) coordinates of best spacer hit (contig ID, hit start position, hit end position) are shown. Brief information about hit identity and annotation is provided (column 5) as follows: "Intergenic" stands for hits that do not target ORFs or viruses. "ORF" for hits into ORF that do not have good hits in virus contigs, "Virus" for hits that targets viruses.

Supplementary information S2 (box, part f) (MS Excel): CRISPR-Cas systems and CRISPR arrays in the genomes with Type V-U system. For each complete genome that contains at least one V-U representative all CRISPR-Cas loci, CRISPR arrays and sequences or repeats are provided. Loci are annotated according to the CRISPR-Cas system classification. V-U genes are indicated.

Supplementary information S2 (box, part g) (MS Excel): HEPN domain proteins in the CRISPR vicinity. All protein containing HEPN domains from the known families located in the vicinity of CRISPR arrays are listed. The following information is provided: gene ID and location, HEPN family, CRISPR-Cas system type (if any), sequence cluster ID.

Supplementary information S2 (box, part h) (MS Excel): Sequences used for analysis of type V systems and TnpB family. For each sequence that was used for reconstruction of the phylogenetic tree (Figure 2 and supplementary information S1 (box)) and profile dendrogram (supplementary information S3) the following information is provided: TnpB sequence ID and its coordinated in the genome, cluster ID, subfamily description, Genome ID and species name and association with CRISPR array (if any).

Supplementary information S3 (figure). **Multiple alignment of representatives from five V-U families.**

V-U1 and V-U4 families are aligned with regular TnpB, whose sequence IDs are shown in bold font. RuvC catalytic residues are highlighted by red letters. For V-U5 with inactivated RuvC domain HHpred output demonstrating similarity with TnpB is shown. Additional information for respective sequences could be found using numeric IDs in supplementary information S2 (box, part h). Alignments are colored using http://www.bioinformatics.org/sms/multi_align.html server according to 80% consensus for the following groups of amino acids: ILMV,FWY,KRH,DE,GAS,C.

V-U1 family

```
1903609002   --------------------------------------------------------------MT    2
1024922355   --------------------------------------------------------------MT    2
1096423661   ---------------------------------------------------------------M    1
 118577413   MKRVTITIDGEQTKGIVIGTIAANHTAAEWLLTASVSAKSAKVRFDPEEAVAETS---SL       57
1003450287   --------------------------------------MPFGKKARHVKAY---QF          15
1507071745   ------------------------------------METAATKNYLALSF               14
 488601079   --------------------------------------------------------------     0
 297565028   --------------------------------------------------------------     0
 257060308   --------------------------------------------------------------     0

1903609002   TMTVHTMGVHY----------------------KW----QIPEVLRQQLWLAHNLRE        33
1024922355   RVTVQTAGVHY----------------------KW----QMPDQLTQQLRLAHDLRE        33
1096423661   AITVHTAGVHY----------------------RW--TDNPPEQLMRQLRLAHDLRE        34
 118577413   VMIAPTRTEKYLYLVPDEQVQPVTTIVRKYGLLSPLDWDCPDYPAGDAFEHLFLQNKLWN    117
1003450287   GADAPQEGMEAV-----------------------------LEQHRLRTDYYN         39
1507071745   GCLSPTRGEEYL-----------------------------LDQIKKKHDLWN         38
 488601079   --------------------------------------------------------------     0
 297565028   --------------------------------------------------------------     0
 257060308   --------------------------------------------------------------     0

1903609002   DLVSLQLAYDDDLKAIWSSYPDVAQAEDTMAAAEADAVALSERVKQARIEARSKKISTE-      92
1024922355   DLVTLEYEYEDAVKAVWSSYPAVAALEAQVAELDERASELASTVKEEKSRQRTKRPSHP-      92
1096423661   DLVTLQLDYETAKAGIWSSYPAVAAAETELADAESAAEQAAAAVSEERTKLRTKRITGP-      93
 118577413   DLVTIEREHRAKYRELIGFDEETAQMDTEIASIKDRLSVLDEGRKKLRVEHRKKKCPEID     177
1003450287   ALVEMELRQREE-RTALLANLAAES------GLESPNQVYERLKAA--GEKGIRKHPEYV      90
1507071745   KLVEKDREHREKVRQVMVFESETTK------KIKELEEELNSLREEIKNQRKTKRTGKVD      92
 488601079   --------------MLISYRFRIYPS-----------------------------------     12
 297565028   --------------MSLLVKCKLIPDA---------ST-----------------------     16
 257060308   --------------MLEQPITVACKL---------QV--------------T---------     15

1903609002   ---LTQQLRDAKKRLKDARQARRDAIAVVKDDAAERRKARSDQLAADQKALYGQYCRDGD     149
1024922355   ---AVAQLAETRAQLKAAKASRREAIASVRDEATERLRTISDERYAAQKQLYRDYCTDGL     149
1096423661   ---LAQKLTAARKRVREARSTRRAAISEVHEEAKGRLVDASDALKAQQKALYKTYCQDGD     150
 118577413   CLDENIKKL--KSELKAVASKAKETRAAAKDRIRAAGNDIENLEKDRQAAVIKAYN-NSG     234
1003450287   AARERQKALYGHPRLLELQSRQREERN--------------------ALRRSFG-AKG     127
1507071745   LTDQKARIEEIKPQLKQLKEKFKEERSFI---FEARKQELAQLEKER-WAVVKELGKGSG     148
 488601079   -----------------------------------------KTVQAKLNEQLELC--      26
 297565028   -----------------------------------------AEKLSRTNQFANAC--      31
 257060308   --T----------------------------------------ANTLAKEIDETMVFACAC--      34

1903609002   LYWASFNTVLDHHKTAVKRIAAQRASGKPATLRHHRFD-GSGTIAVQLQRQAGAPPRTPM     208
1024922355   LYWATFNAVLDHHKTAVKRIAAHRKQGRAAQLRHHRWD-GTGTISVQLQRQATDPARTPA     208
1096423661   LFTATFNDVLDHHKAAVKRIGQMRAAGQPAQLRHHRFD-GTGSIAVQLQRQAGQPQRTPE     209
 118577413   LWWGNYNAVLESYKKA--RIKA---LKDGAELKYHRFD-GSGRFTNQIQG--GMSVQD--     284
1003450287   LYSSNYLDVERAFDKA--RQSP--------ELRFRRYSPHEGRLAVLYTE--GLPMRE--     173
1507071745   LYWCNLEDVVNSYDIG--RKKA---KAAGGEMRFHRWD-GTGKVTVRFQK--GLPVNE--     198
 488601079   -RWLYNRLLSEVNKA--------RKEGRRIRRED--TQSLIVRIKR-----EEKPE      66
 297565028   -N------ALQVARR----------DNIWNF---A--LQRAVYADLR-----AEY--     61
 257060308   -D-------VNQNTP----------EKMTNT---A--MQSLVYQDVR-----VNF--     63

1903609002   VLADEA-GKYRNVLH------I---------------PGWTDPDVWEQMTSQCRQSGRV     246
```

```
1096423661 LIADVD-GKYGRVLS------V---------------P-WVQPDRWERIP SERRMIGRV 246
 118577413 LL-----EGNRNVAS------LRLVSSGELGDISGKKP--PSLDLQSVGS SDSREYGIL 331
1003450287 IG-----SDTRVQL----------------------P--LPDPIIYRD- STRRKHQRV 202
1507071745 MF-----SCTNNLLQ------I-------------DP--VDKDAWYNPV SIRRKKSRT 231
 488601079 LS-----KVYSKVLQMVNYQLRSNISSLNELRKKGVKVGWLRYKTSPNSF STLNFNQSGF 121
 297565028 ----------GLSANLAIRAI-----ARVGKRKGHKAGGFKATSVDYDQ SLSVNVDTE 105
 257060308 ----------GLSSNLAIQAIRRVCANRKTAKQKGKKVKEFKPTSISYDA SFSFRESDW 113


1903609002 TVRMRCGST-----DGQPQW---IDLP QVH NLPADADITGAE VVT V GIYRAK C  298
1024922355 VIRMRCGSSRNPDGTKTSEW---IDVP QQH MLPADADITAAQ TVR E ADLRAT G  303
1096423661 TVRMRAGQL-----SGEPQW---LDIP QQH MLPLDADITGAR TVT T GTLRAQ S  298
 118577413 AITLYTGTD-----EQSKKFRRTLSFP ILH PLPEGATLKSLS HRK V TDFVWS VF 386
1003450287 LIKFRVRSV-----ERQPLW---ITVP YLH CLPDG-VCREVS HWH V DRLRWT S  253
1507071745 RVRLRACSE-----NKKPLF---IELP VLH IPEDALIRTAS IRE V MRYRYK N  283
 488601079 KI-----D---FDRK-KLSLSKVGDIP RLH SI--GGKIKGVI KRT S ---KWYAI  167
 297565028 TV-----SLSTVDGRVKVPMRIAG-YQRHLL TA--KSIQGGQ- VRG D ---SWY H  153
 257060308 TV-----SVKLLNSRQRIKL-LIGNYQ GLL SK--NPTSAT-- VKR S ---NYY H  160


1903609002 TARIGDTEPVT----SGP-TVALHLGWRS-TEEGTAV TWRSDAPLDIPFGLRTVMRVDA 352
1024922355 TAKIPDQGEVD----EGP-TIAVHLGWRS-SDHGTVV TWRSTEPLDIPETLRGVITTQS 357
1096423661 TAKIPDPEPVT----DGP-DVAVHLGWRN-TDTGVRV RWRSTEPIEVPFDFRDTLTVDP 352
 118577413 TFTTDCP-TYDQRSSTGN-RCGLNLGWKKQAGGGLRV T-I----------------YD 426
1003450287 VVEVEGP-PVA--SPTGRGAVAVDLGWRR-VEGGLRA FWV---------------GE 292
1507071745 VLEILGE-NTNRILPALEGTAAIDLGWRT-VKDGLRV CLV---------------DD 324
 488601079 QAEVDKQ-PLP---PTGR-AIGIDVG ITH------FCV----------------DS 196
 297565028 WCEYDDP-PVL---DPQG-MLGVDLG VN-------I T------------------DS 182
 257060308 TLDEPTQ-PEA---KTDK-VLGVDLGRTD------I T------------------TS 189


1903609002 ASTS-GIIVVPAT ERRLT TE-NIASSRSLALDALRDK VGW SDNDA---PTYRDAPL 407
1024922355 AERTVGSIVVPHR EQRVH HA-TVASHRDLAVDSIRDT VAW TEHGPQPHP-YDGDPI 415
1096423661 GSRS-GEIFVPEA VPRRVE RAH-LIASHRADRMNELRAR VDY AETGPRPHPSREGEEL 410
 118577413 GSDA-RHITLPQA IDGLDYVNGDLQGRIDSAANENHAW LEQWGG---DELP--ESLQE 480
1003450287 DSAG-GEIALSEGDLKQFS VE-DLRSIRDQHLNALKEA AAW EA-PPAPLP--DWLAE 347
1507071745 KSHS-EELILDND LHEFN IK-DLQSIRDNLFNETKAK MEL KT---LELP--DEAKE 377
 488601079 DSNYFEH---PKY DRTLE K---------------IKK QKQ SRKQ---------K-- 226
 297565028 DGETYSG----KHLNSVRHR---------------HRR RKK QKKG------------210
 257060308 ESESWSG----KQ TAKRN I--------------YAK RTT QKKA---------SKG 220


1903609002 EAATVKQWK--SPQRFASLAHAWKDN----GTEISDILWAWFSLDR QWAQQENGRRKAL 461
1024922355 TAASVQRWK--APRRFAWLALQWRDTPPPEGADIAETLEAWRRADK LWLESEHGRGRAL 473
1096423661 GAGNVRMWK--SPNRFAWLARVWADDESV-STDIREALAQWRHQDWISWHHQEGGRRSA 467
 118577413 LRSMLRRSKRPHPAKFAKAVIAWRNYPEYLGDARDEA-EQRRKATK LTIEMAHKREKLL 539
1003450287 ETKTLPQWR--SPARFAALFRRWQSERVHADEAAYGLLEGWHKRDR LWQYEANLREQMI 405
1507071745 RTSHMANWR--SQQKMLRLHQYWRENRLPGDDEVWEVLEYWRKREI LYEWQENLRDQVL 435
 488601079 GSKN----------------------------------RE VRIGLAKLYEKLE 246
 297565028 TKGA----------------------------------KR I----LKKLSGKET 226
 257060308 TRSS----------------------------------RR CRQLLARLSGKER 240


1903609002 GHRDDLY RQI AVISDQA----GH VLVDDTS- AELSARA--MERT-ELPTEVQQKIDRR 513
1024922355 RHRTDLH RQV AYFAGVA----GR IVVDDSD- AQIAGTA---KHS-ELLTDVDRQIARR 524
1096423661 AQRLDVY RQV AVLVSQA----GR LVLDDTS-YADIAQRSATTKTE-ELPNETAARINRR 521
 118577413 RRRMDFY RNT AKQLTSVY----DV CLDKMD-LR---RLALLEKGDGTPNELTKIARKQ 590
1003450287 LRRREQY RVL ATLARQY----DA LIVEDFN-LR---AAAELDQG-G--SDLPDAARRY 453
1507071745 RRRKEIY RIF AAKITRKY----KT IVLREFT-LNK---TVQKPNPEEGPAGTLP--ANRN 485
 488601079 NQRNDFL IKL SRYYVNNY----DI IVVEDLN-VKE---MA---EN--GSSTT----LNRH 289
 297565028 RFSNHVN ITL SKRIVAKAQRTERA ALEDLQG IRE---RV---RL--RRPQ-------R 270
 257060308 RFQKHIN IEI SRQLVNNAVTNKQA AIHEDLTG IRE---RT---N---RKPRS----KKDK 287


1903609002 RDHA AP--GGLRAS VAAMTRDGVPVTI AADFTRTHSR HVN---PADDRYLSNPVR 568
1024922355 RAIA AP--GMLRAA VAAATRDEVPTTT HTGLSRVHAA HEN---PADDRYLMQPVL 579
1096423661 RAHA AP--GELRQT LVAAADRDAVPVDT HTGVSVVHAK HEN---PSDGRFMSVVA 576
 118577413 RQQA AI--SELREC LSKAAAKNGTQIEQ TA-SSATCSA GKM--EQVDGI----MWR 641
1003450287 RTIA SP--STLRDA LVNAFAQRGKPVRK LNPAHTTTDCHA GALVGDPAKEL----RLY 507
```

In format provided by Koonin *et al.* (doi:10.1038/nrmicro.2016.184)

```
1507071745  RFIAAI--SEFRNELANACRKNHVEFTYVPAENTTITCHKCGHKEKFDAAAQI----IHT  539
 488601079  ITDSAWSKFV--RLLCEKAERAARTVVKVNPKNTSKRCAMCGYIVNNLKLHDR----TFT  343
 297565028  ATLHSWAFFDLGQKLRYKAERAGVPLVFVDPRNTSRQCPACGHAERANRPTQA----LFR  326
 257060308  RLGNNWAFYQLRQFLTYKCILAGVKLILVNPAYTSLSCHKCLVIGDR---KGK----GFS  340

1903609002  CDGCGAMYDQDRSFVTLMLRAATAPSNP--------------------------------  596
1024922355  CDGCGRTYDTDLSATILMLQRASAATSN--------------------------------  607
1096423661  CDGCGEKYDQDESALTHMLTRAVQSAA---------------------------------  603
 118577413  CRECRALVDQDINAAANLFREVL-------------------------------------  664
1003450287  CPTCERFYDQDENAARNLLRRAQEVQAQV-------------------------------  536
1507071745  CSTCGELWDQDYNAAKNLLAFSQKGGVK--------------------------------  567
 488601079  CPICGWEADRDYNASLNILDVGMGRSRTPVEGEPLPCVISYR-EVIAGQVLSMKQEVPSV  402
 297565028  CVACGYSGAADYVAAVNIAVRGWAAVNRP--YLGEASRVSLH------------GSVPGS  372
 257060308  CNNCGNKCDADYNGAQNIKALGAIINRPG--GSGLSCKLKTNVQYIQLSLFEGLGLLKTS  398

1903609002  ---                                                          596
1024922355  ---                                                          607
1096423661  ---                                                          603
 118577413  ---                                                          664
1003450287  ---                                                          536
1507071745  ---                                                          567
 488601079  RAE                                                          405
 297565028  PRL                                                          375
 257060308  TSA                                                          401
```

V-U2 family

```
1046552329  -----------MAVKVFEFKIYPH--------KDFQEQFNRWAYGLKKFYNFCLQQFELL  41
1019491369  -----------MTVKVFEFKIYPK--------KEFQEQFNRWAYGLKKFYNFCLEQFELL  41
 428311397  ------------MLTLEFKADFS--------LEQQAKIDRWLEINRSLWNMGLAALEDF   39
 218248844  MFAIKSMSELVQHHITIQLKAYLS--------TTQTALFENWTDSLRPLYNLALGLLYEE  52
1030942420  ------------MKVLEFKIHPT--------EEQVSKIDQSLAACKLLWNLSIALKEES   39
1305565200  ------------MKTLEFKIYPT--------LAQSQTIDKWLDKLKWVWNTGLSLKLAG   39
 488601079  -----------MLISYRFRIYPS--------KTVQAKLNEQLELCRWLYNRLLSEVNKA   40
 297565028  -----------MSLLSVKCKLIPDAST----AEKLSRTVNQFANACNY-----ALQVARR  40
 257060308  -----------MLESQPITVACKLQVANTLAKEIDETMMVFACACDW------VNQNTP   42

1046552329  DEYTYWDKLSKTRVPCCPVPWSLKLIETLDPN--PYLPELKNKHYVS-YSNLIAPQDIPV  98
1019491369  DEYTYWDKLSKTRVPCCPIPWNLKLIETEEPN--PYLPELKNKHYVS-YSNLIAPKDIPV  98
 428311397  DDFYSYVKGQKEYAPCCPIQYEYRPLSEEEKACIPTHEKTSDRKYLAPFCRIISEKSRWY  99
 218248844  QQ-RRWRTNQKF-----------------LKNYLDKSSLQTYLNEIENK---PDIYP    88
1030942420  KQ-RYYRKKHKF-----------------------DEF--S---PEIWG          59
1305565200  RQ-KYYREKEIG-----------------------DQV--I---PDGVV          59
 488601079  R--KEGRRIRRE-----------------------------DTQSLIV            57
 297565028  D--NIWNKF-----------------------------ALQRAVY               54
 257060308  E--KMTNKT-----------------------------AMQSLVY               56

1046552329  ----IRNAPEARQYTLKKGETA--KDVFKRVKNPDQIVVTTAKPESSGLDKWPRGWLGGV  152
1019491369  ----IRNAPEARQYTLKEGETA--KDVFKRVKNPDKIVVTMAKPESSGLDKWPRGWLGGV  152
 428311397  ----VKKLP---IYKVPTPAEK--KDSWGWL-----------PSNHDED---RKYSNCT  135
 218248844  VEWHITKALPECDWLTKEENEVRKKDNTKSLA---CRTI----NRD--------GNFFT-  132
1030942420  LSY------SGHYDEKEFKTLKDKEKKLLIGNPCCKIAYFKKTSN--------GKEYTP  104
1305565200  LQWKWRKVVTE----DKKGKSTEKWEKVRLVG---TGVI---RPKN--------GYPYC-  100
 488601079  -------------RIKREEKPELSKVYSKVLQ---MVN-------------------    79
 297565028  -------------ADLRAEY------GLSAN---LAI-------------------    69
 257060308  -------------QDVRVNF------GLSSN---LAI-------------------    71

1046552329  GYSQPPKQ-------DYQPSLIKNFPNSEKTKDAGILVKKETLENSNLPDEIKKVILDVP  205
1019491369  GYSQPVKQ-------DYKPSLIKNFPNSEKTRDAGILVKKETLENSDLPDEIKKAILDVP  205
 428311397  GYSCPIPRYGSVENPSWYEPMIRN-PTYKGSGGLSLVSKTE-----NLPQWMKDS--DIP  187
 218248844  ----PIRPYWHLEEPQKLAKFKCF---TN----QWLISCNLLTNYHLQKLL-----NVN  175
1030942420  LNSIPIRRFMNAENIDKDAVNYLN---RK----K----LAFYFRENTAKFI----GEIE  148
1305565200  ----EIRQHLNIEDPDKYGQ-----------------CEFYRSDKIPDFM----ADVP  133
 488601079  ---YQLRS-----------NISS---LNELRKKG----------VKVGWLRYK--TSP  108
```

```
297565028   ---RAT--------------------ARVGKRKG-----------HKAGGFKAT--S D   92
257060308   ---QATRR-----------VCAN---RKTAKQKG-----------KKVKEFKPT--S S  100

1046552329  YKFRSGTLA-MLCTSVQEYLKSRSGQKDLKRGKPKYKRYDRIETIIHPNPNAGSSKPAS 264
1019491369  YKFRTGTLS-SLCTSVQEYMKSRTGQNDLKRGKPKYKRYDRIETIIHPNPNAGSSKPAS 264
 428311397  QRFRAGEMG-QLDTAWQEYLKSRYGQSEVKRGKPQYKRKRDKLQTLINTNPSANE-------241
 218248844  MKVRQSFISMNLMEAVKRY--Q-----KGDFRKLKFKSKRNPVISLCN----KQT-N-RI 222
1030942420  TEFKGFFKSVIKTAVDAA--K-----KGIRGIPRFKGRDKVETLVN----GQP-D-TI 195
1305565200  TKFKAGVID-SLKKSVKAYVTP-----KHPGRKPKFKGRNDKIKSLVNLAGGKS-K-EL 185
 488601079  NSFKTLNFN---QSGFKID-------------------------FD----R-----K- 128
 297565028  YDQRILSVN---VDTETVSLS----------------------TVD----G-----RV 116
 257060308  YDARIFSFR---ESDVTVSVK----------------------LLN----S-----RQ 124

1046552329  KDACRLEGDNILV-LPSFGKVKIKGDKRFRDNDGS-IPRVKVVK L--PSG YIQ TA 321
1019491369  KDACRLEGDNILV-LPSFGKIKIKGDRRFRDNDGS-IPRVKVVK L--PSA YVQ TA 321
 428311397  ----RLVGNNIFAGIPKLGKVRCKGDKRWRNPDGS-IPRVATYK C--PDA YIQ SG 295
 218248844  KFD--PEANCQLLGKEFGLIEFRGHNRH---QGQIQPRNG--S T--ADG YLN VF 274
1030942420  KIK-----SNGVIVSSKIGLLKVRGLDRL---QGK-APRMA--K T--ATG YLQ TV 242
1305565200  KPEKIPGSNNGYVQFPKLGKIRVKGFDRY---DWQ-EWGAA--R VI-EPSG YLH CV 238
 488601079  KLSLSKVG-DIPI----------R HRSI----G---GKIKGVI K T SGK Y I QA 169
 297565028  KVPMRIAG--YQR----------H LRTA----K---SIQGGQ- V G DSS YIH WC 155
 257060308  RIKL-LIG-NYQI----------G LKSK----N---PTSAT-- V R SGN YIH TL 162

1046552329  DINRSNKLFKKPLGAIGIDTGLKEDNWITTDRFAVTKP WYRESEE LA L KELDAKKL 381
1019491369  DINRSNKLFKKPLGAIGIDTGLKEENWITTDRFSVTKP WYRESEE LA L RELDAKKL 381
 428311397  EVQRSFSV-KATNASIGIDPGLQYELSL-SDGTRIQPQ FYRKSEE RA L QKLAKKLT 353
 218248844  QVEH--KPIPDSDLQVGIDPGLVTLLT-LSDGKCISNQ RFLKENER LTV KKLSRQT- 330
1030942420  ETDD--TIYKESDKCVGLDMGAVAIFT-DDLGRQSEAK YAKIQKK LN L RQASRQK- 298
1305565200  DVPD--EPLPKSDKSVGIDPGLLSVIT-TDQGREVPP LFRKQQA LR L RKASRQV- 294
 488601079  EVDK--QPLPPTGRAIGIDVGITHFCV-DSDGNYFEHP YLDRTLE IK V QKQLSRKQ- 225
 297565028  EYDD--PPVLDPQGMLGVDLGIVNIAT-DSDGETYSG- HLNSVRH HR L KKLQKKG- 210
 257060308  DEPT--QPEAKTDKVLGVDLGRTDIAT-TSEGESWSG- QITAKRN YA L TTIQKKA- 217

1046552329  QRVILWLNHPDNSIERIKTIFPSIAKEALEKVKACKRPQYLHELVKNNELSTSGLNQLKH 441
1019491369  QRVILWLNHPENSTERIKEVFPGISKESIEKVKGCKRPQDLQDLVSNNELSTSGLNQLKH 441
 428311397  ERLILWIHHPDRTIQEIRKNFFPISNESYEALRAAKTEAEVIK-----AIGASRLNTLKY 408
 218248844  -----------------------------P------------------GSKNWE-- 337
1030942420  -----------------------------D------------------GSNNQR-- 305
1305565200  -----------------------------K------------------GGCNQK-- 301
 488601079  -----------------------------K------------------GSKNRE-- 232
 297565028  -----------------------------------------------TKGAKR-- 216
 257060308  -----------------------------S------------------KG-TRSSRR-- 226

1046552329  FNFRDCEKVESCYLFDKLLSASNKEIELAQR RKLHE LKRRRRSHNQKQ TW TRKY-- 499
1019491369  FNFRDCEKVESCYLFDKLLSVSNKEIELAER GKLHE NKRRRRSHNQKQ TW TRKY-- 499
 428311397  NIVPDAPPT----MKDKSPFSGAKQKALEKA RKLDR ISLQRRNHDHKITTM VRNY-- 462
 218248844  -------KT------------------KKA AKIHKQTADHRKYYNHKV TH VNKY-- 369
1030942420  -------KT------------------YAK ARVHE IARQRKGRNAQL HK TSEY-- 337
1305565200  -------KT------------------YRK ALHHE IRRSRNAFNHKL TK VREY-- 333
 488601079  -------KV------------------RIG AKLYE LENQRNDFLHKL RYYVNNY-- 264
 297565028  -------R-------------------KKLSG ETRFSNHVNHTL KR VAKAQR 246
 257060308  -------RC------------------RQL ARLSG ERRFQKHINHEI RQ VNNAVT 260

1046552329  --SITRIEDGL-QKNV---GKSKAKVSEDNRSFERNAQNSRTGMNK VLDAAIGGFISL   553
1019491369  --SITRIEDGL-QKNV---GKSKAKISEDNRSFERNAQNSRAGMNK VLDAAIGGFISL   553
 428311397  --GFIAVEDGL-QDEKLR-KRTKPKEREDGQGYEQTGAKRKSGLSK LADASPGRKIAF   518
 218248844  --GAIAIEDTK-LTNM-N-KRPKAEKREDGKGYEHNGAKAKAGLNQ FHDAGLGQLRAF   424
1030942420  --QSVILEDLK-LKNM-T-AAAKPKEREDGDGYKQNGKKRKSGLNK LLDNAIGQLRTF   392
1305565200  --SGIVMEDIK-IQNL-N-RKPKAKKREDGNGYEQNGAKRKAGLNK FADSALGDLISK   388
 488601079  --DIMVVEDLN-VKEMAENGSSTTLN-----------RHITDSAW KF-----V--RL   302
 297565028  TERALALEDLQGIRERVRLRRPQ--------------RATLHSW FF-----DLGQK   285
 257060308  NKQAIALEDLTGIRERTN-RKPRSKK-----------DKRLGNNW FY-----QLRQF   302

1046552329  ESKSKE-WGRDFKRMKPGKGKAYSQR- DV I-HENKEQKDITNHQD-- N N NCG THR 608
```

```
1019491369 ED KSKE-WGRDFKR MKPGKGKAYSQR- PV H-HENKEQKDITNHQD-- N NCGFTHR 608
 428311397 KQ KADR-AGRVFSQ----HPAPYTTKE PV G-SMNEASYNVDDEGNRL L ICGWEC- 571
 218248844 ES KANSYENRHIER V---RANYTSQK SR G-HTDSENR-L---TQAS H KCGLEM- 474
1030942420 EN KANE-RGRKIIR V---NPKHTSQT PN G-NIDKANR-V---SQSK K SCGYEA- 441
1305565200 ET KCQD-TDREFVK V---AAHYTTVD SN GAKIKKA---L---SQRTHR ECGHTE- 436
 488601079 CE KAER-AARTVVK V---NPKNTSKR AM G-YIVNNLK-L---HDRT T ICGWEA- 351
 297565028 RY KAER-AGVPLVF V---DPRNTSRQ PA G-HAERANR-P---TQAL R ACGYSG- 334
 257060308 TY KCIL-AGVKLIL V---NPAYTSLS HK L-VIGDR-------KGKG S NCGNKC- 348

1046552329 SRDVIPGINMILD----E-----------FEEQQQ----------KNLLGIETE ---------637
1019491369 SRDIVPGVNMILD----E-----------FEAGD----------------------------627
 428311397 DRDVNSGVNIELA----Q-----------FGNNP----------------------------590
 218248844 PADLNAAINIEQT-------------------------------------------------487
1030942420 HADQNAANILIRGLRDEFLRAIGSLYKF------PVSMIGKYPGLAGEFTP--DLDAN 492
1305565200 GRDSNAAKNILLKGKKQ--LQTVYRAWAWEHGETRKPDSECETHCHQEGVQAPPEDEHSS 494
 488601079 DRDYNASLNILDVGM---------------GRSRTPVEG-----------EP---LPC 380
 297565028 AADYVAAVNIAVRGW---------------AAVNRP--Y-----------LG---EAS 361
 257060308 DADYNGAQNIKALGA---------------IINRPG--G-----------SG---LSC 375

1046552329 -----EQKIIWDDLSKECR--QAWRLREKWLSEN---------APGGGCQDEVNSDKPSN 681
1019491369 --------IQWDDLSKEAR--QAWRLREKWLSEN---------APGSGCQDEVTTDKPSN 668
 428311397 -----------HTVLSANAQ--RARFANSVW----------------EIAHPEASTKPRW 621
 218248844 --------AFGLDKS-----------------------------------------------494
1030942420 QESIGDAPIENADRSTSKQMKQE-GNRIPTQSENDSQSLIFLSAPPQPCEDSHGTNNPKA 551
1305565200 QLTLVRTPAKGTQRGRGDV--KT-STPAKTTSKLDTVSDPNLDNKPDTRAPCTSSISPNS 551
 488601079 VISYR-EVIAGQVLS----MKQE-------------------VPSVRAE ---------- 405
 297565028 RVSLH---------------GS-------------------VPGSPRL ---------- 375
 257060308 KLKTNVQYIQLSLFE----GLGL-------------------LKTSTSA ---------- 401

1046552329 --TRKSR-RRKK--------------------------- 690
1019491369 --ARRKR-NRKKKT------------------------- 679
 428311397 --KKTEK-RKKRK--------------------------- 631
 218248844 --------------------------------------- 494
1030942420 LSDKASK-RKSKKSRGAIPENPDQLTIWDL-----LD 582
1305565200 VENKIPKTKKNKRSAQSSSESFTQLTIWDTAGEISFE 588
 488601079 --------------------------------------- 405
 297565028 --------------------------------------- 375
 257060308 --------------------------------------- 401
```

V-U3 family

```
1705264553 MPFILKQRGVRMITVRKLKIVCKDKE---FYDFFKWEQREQNKALNIAEGLIHSSTVLRS 57
1018643584 ----------MIIARKIKLIIIGENRDAQYKFIREERYKQNKALNVAMNHLYFLHVAKE 49
1077873159 ----------MITVRKIKLTIMGDKDTRNSQYKWIRDEQYNQYRALNMGMTYLAVNDILYM 51
1077873155 ----------MNKCIKVAIKNC-------------KELDYKVMSEKLRNIQY---LTC 32
1019720777 --------MSIAVKVMKYQIVCP-------------VNVEWKVFETYLRTLAY---QSR 35
1017485595 --------MVCNKVVKIALICD-QI-------DKDGKDVNYNDIYKLLWDLQK---QTR 40
 317133985 ----------MTKVVKLALISNVTD--------KDGNKVEYNELNRCLWDLQK---ETR 38
 488601079 ----------MLISYRFRIYPS-------------KTVQAKLNEQLELCRW---LYN 31
 297565028 ----------MSLLSVKCKLIPDAST----------AEKLSRTVNQFANACNY------ 33
 257060308 ----------MLESQPITVACKLQV------ANTLAKEIDETMMVFACACDW------ 36

1705264553 IDSGAEAQLKKSIGKLTQNIEKLGKELEKEKITDKKKEQLLKAINTNKELIASKEKELKA 117
1018643584 KIRLLDNKFLQDEKKLQESINKLYAEKKVIKDEKKRNELEKKIEKQTNELKKLRSKSNKE 109
1077873159 NESGLEIRTIKDLKDCEKDIDKNKKEIEKLTARLEKEQNKKNSSSEKLDEIKYKISLVEN 111
1077873155 KASNKVMQMYYMWENQKIDIKNKTGEYPDDKELFGK---------TYRNV--------- 73
1019720777 TIGNRTIQKIWEFDNLSLNHFKETGEYPSAQQLYGCTQK-------TISGY--------- 79
1017485595 EAKNKVIRLCWEWSGYSSEYFKTHEEYPKDKEIFGI---------SLRGY--------- 81
 317133985 DLKNAVVRECWEWYGFTNDYYKLNEEYPNERDYLKKAKSDGTIKDYALDGF--------- 89
 488601079 RLLSEVNKARKEGRRIRRE------------------------DTQSL--------- 55
 297565028 --ALQVARRDNIWNKF--------------------------ALQRA--------- 52
 257060308 ---VNQNTPEKMTNKT--------------------------AMQSL--------- 54

1705264553 GEEFRCGIDKKFNELYMNKTTIYHVIDSIC--------------DFKYKRTIELVRQKV 162
```

```
1018643584  ADKVLQEAIKINLSST-----TREV SKQF----------------ELIS-DTKDRITQKV  148
1077873159  KIEDYKLKIVELNKIIEETQKERMD QKEFKEKYVDDLYQVLDKIPFKHLDNK LVTQRI   171
1077873155  --------------------- EGE KTIM----------------NTINTSNV QTNAII    97
1019720777  --------------------- YDQ KEEY----------------QDINKANM TTIQKT   103
1017485595  --------------------- YDR KGDY----------------NLYS-GNL QSAEIA   104
 317133985  --------------------- YAKYSKKY----------------NLHS-GNY QTLRDA   112
 488601079  --------------------- VRIKREEK----------------PELS-KVY KVLQMV    78
 297565028  --------------------- YAD RAEY--------------------- LSANLA      68
 257060308  --------------------- YQD RVNF--------------------- LSSNLA      70

1705264553  KQDYSNSFTD- VT KVSLQNYKSTFPLMIDGSCISILKEVDELD VNGYKIKIMLGYE-   220
1018643584  SQDFKSDLKHG LS E VLRTYKKNNPLLIRG--RALNFYREGKD M----IKWYGGII-   201
1077873159  KADIKSDKSNG LK E SIRNYKRNFPLMTRGRDLKFKYD-DNDD E----IKWMEGIK-   225
1077873155  MKKWNTDKKE- LSYQ SLPNFKLNMPIYIKNKSFSIVKGTSGYE I----CSIFNKSQD   152
1019720777  LKNWNSRRKE- WR E KQNS------------------------             123
1017485595  YKEYKNSLKD- LR D SIINYRENQPLDIKNKAIQLLYENDNFF R----VALINKDKQ   159
 317133985  CGSFKNNLKE- LR D SILSYRADQPLDIKKTCIGLEYDKDTNTYY----VTLVLLNKN   167
 488601079  NYQLRSNISS- NELR KGVKVGWLRYKTSPNSFKTLNFNQSGFK D-------FDRK--   128
 297565028  IRAI-------ARV K KGHKAGGFKATSVDYDQRILSVNVDTET S----LSTVDGR--   115
 257060308  IQAIRRVCAN-RKT KQKGKKVKEFKPTSISYDARIFSFRESDWT S----VKLLNSR--   123

1705264553  ----------LDII GKRENENSLELQKT L CITGDYKICASS- QRD NNNVIFN TLD   270
1018643584  ----------FKCM GQHKN-NAPELKAT S VLEGSYKVCD---SSISVGKELILN SLD   248
1077873159  ----------FKVI GNRIK-NSLELRHT L VIEGKYKICDSS- QFD NNNLILN TLD   274
1077873155  LKR------LTFI DKLDG----NKKAT L IIDLTYKQGAGQ- IKD KGKWYFI SFG   201
1019720777  ------------------MKVI D IIDSTYAKGAC-- LHK KKKWYLS TYK          157
1017485595  KELNFKDCSVRFK LVKDD----STRTI L CFDEVYTITASKI YNK KKQWYIN GYK   215
 317133985  GVKHYNISDFRFK TVKDN----STRTI L CFDGVYGISASKL WNR KSQWFLN CYS   223
 488601079  ----------KLS SKVGD-----IPIR L SIGGKIKGVI--- KRT SGKWYAI QAE   170
 297565028  ---------VKVP RIAG------YQRH L TAKSIQGGQ---- VRG DSSWYIH WCE   156
 257060308  ---------QRIK LIGN-----YQIG L SKNPTSAT----- VKR SGNYYIH TLD   163

1705264553  I-PIEKDYKPVKG RV C GVDLGI KYPAYMCL---NEDTYKKEAVGSI----NNFLRI--   320
1018643584  IGEVDTNVSCKKG RV LGVDLGM KVPAYMSI---NDKPYIRKALGSL----DDFLKI--   299
1077873159  I-PIDIVNKKVSG RV VGVDLGL KIPAYCAL---NDVEYIKKSI RI----DDFLKV--   324
1077873155  F--ENKKRELDIN RI LGIDVGI TNLLTMQIWDCNLKEWDRLAWN CMVDGRELMHY--   257
1019720777  S-NIKEELKFDEDL IMGIDMG KINVLYFAF---N-KGLVRGAI G-----EEIEAF--   205
1017485595  F-TKEIDKTLDKD RILGDLGV INPLVASV---Y-GSYDRLII G-----GEIDKF--   263
 317133985  F-DKVEVKELDKK ILGVN LGVYYPLYASI---S-GEKDRLAI G-----DEIIEF--   271
 488601079  V---DKQPLPPTG A IGIDVGI THFC---------VDSDGNYFEHP----KYLDRT--LE  212
 297565028  Y---DDPPVLDPQG MLGVDLGI VNIA---------TDSDGETYSG-----KHLNSV--   197
 257060308  E---PTQPEAKTD KVLGVDLGR TDIA---------TTSEGESWSG-----KQITAK--   204

1705264553  Q MQERR KK D ELLLT---NG-- KG RTK TQA- EK RENEKN KTYNHAI R V F   374
1018643584  Q MQKRR RN LH TLVNV---KG-- KG REK LQA- DR KD KN TTYNHFL N V F   353
1077873159  Q MQSRR RR LD IAIQSA---KG-- KG RVN LQA- ERFAE KN KTYNHFL N V F   378
1077873155  K IEARR KS LL NSKISEKNTGKA HG SK IQA- DV RN KN DTFNHKY YA DF   316
1019720777  K IEHRR IS LL RQGKYCSGNRI-- KG REK IKP- DV ND AK NATNHKY NY VDQ 262
1017485595  R VEANKVQ ML KQGKYCGDGRI-- HG VNT NKP-AYN ED SR DTVNHKY KA VDY   320
 317133985  R EARRTA LK QAAVCGDGRI-- HG KT MKP- QN SD AN DTFNHKA K IDF      328
 488601079  K KKVQKQ LS RKQK--------- SKNRE VRIG AK YE ENQRNDFL L YY N      263
 297565028  RHRRL KK LQK-------KG--T KGAKR ---- KK SG TR NHV HL R V K        243
 257060308  HYAKLR TT IQ KKAS-----KG--T RSSRR CRQL LAR SG RR KHI HI RQ V N  257

1705264553  A RKNKCEY N LEK T-------------------------KDGFGDSI NWSYFE KM EY 412
1018643584  A KDNLAEQ N MEF A-----------LAGEDK---------NII NWSYYQ QF EY     392
1077873159  A VSNQAEQ N MEL S-----------LKETQN---------KSI NWSYYQ TM EY     417
1077873155  A IRNNCGI D MEN A-----------KFTEEVK--------EKM NWSYYD SK KY     356
1019720777  CLKYNCGT D LED K-----------G SKE----------QTF NWTYFD EK KN      300
1017485595  A VKNNCGT D MED K-----------G TQNKN--------ERY NWTYFD TK EY     360
 317133985  A IKNDCGI D LEN K-----------G TKNS---------EGF NWSFYD SK EN     367
 488601079  Y-----DI MV ED N-----------VKEMAENGSSTTLNRHITD-SA SK V--RL CE  304
 297565028  A DRTE-RA LA ED Q----------G RERVRLRRPQ----RAT S AFFD QK RY    287
 257060308  A VTNK-QAIA ED T----------G RERTN-RKPRSKKDKRLG-N WAFYQ QF TY   304
```

```
1705264553  KAKSKGIEVRYIDPCFTSQKCSKCGYIDKENRE---TQEDFICKKCGFKLNADHNASINI  469
1018643584  KAKREGIDVKYVDPYRTSQMCSKCGNYEPGQRE---SQEKFICKSCHLEINADYNASQNI  449
1077873159  KAQREGIKVKYIDPYHTSQTCSKCGNYEEGQRE---SQADFICKKCGYKVNADYNAARNI  474
1077873155  KAEEQGIKVNFIKPSYTSKRCSLCGAIDDRNRDCKNNQSKFDCVVDDHKEHADINAAKNI  416
1019720777  QANQYEVKVVKIDPFYTSQRCSECGYIHKNNRQ---DQSTFECQQCSFKVHADYNAAKNI  357
1017485595  KAKALGIEVKYKNPKYTSQRCSKCGHIAEENRP---EQKTFKCVKCGFKVNADYNASQNL  417
 317133985  KAKERGIKVVYIEPAYTSLRCSKCGCIHKDNHP---TREQFICQEDGYRVLHDYNASQNI  424
 488601079  KAERAARTVVKVNPKNTSKRCAMCGYIVNNLKL---HDRTFTCPIDGWEADRDYNASLNI  361
 297565028  KAERAGVPLVFVDPRNTSRQCPACGHAERANRP---TQALFRCVADGYSGAADYVAAVNI  344
 257060308  KCILAGVKLILVNPAYTSLSCHKCLVIGDR------KGKGFSCNNDGNKCDADYNGAQNI  358
```

```
1705264553  ARSKEFIK-----------------------------------------------  477
1018643584  AHSTKYITNKN--QSEYLKKLQQT--------------TKLEKYS           478
1077873159  AMSNKYITKKEESKYYKIKESMV--------------------            497
1077873155  ALPDIEELIES--KIG----------------------------           430
1019720777  SVYNIEKVIQK--QLELQEKLNLT-------KYKERYIEQMENIN           393
1017485595  AIKDIDKIIEQ--YYNKG--------------------------           433
 317133985  AVKDIDKIIKA--ELEKTEPEKKT--------------EEEKPEK          453
 488601079  LDVGMGRSRTPVEGEPLPCVISYR-EVIAGQVLSMKQEVPSVRAE           405
 297565028  AVRGWAAVNRP--YLGEASRVSLH-----------GSVPGSPRL           375
 257060308  KALGAIINRPG--GSGLSCKLKTNVQYIQLSLFEGLGLLKTSTSA           401
```

V-U4 family

```
1040961339  MSYLCARGRFNLSGFSTGSMGVLPVRASLFALLFTTLHGARISAHWYGMSKENDHTVTCI  60
1405224212  --------------------------------MSSERAPKL---------RNVVTQQAY   18
1206162758  -----------------------------------MGTEG---------ALTH--RVV   12
1009979974  -----------------------------------MTSTTLAPE---------EPLMVFRGA   18
 451770599  ----------------MTTARANEIPTCQTQHLRSTHMTDV---------E--L--SAY   30
1305448890  -------------------------------------MAEV------------L--RAF    8
1706614983  -------------------------------MAEKTGTDA---------GTMN--RAY   16
1706842133  ----------------MPLNGWTMLAQVEVRAMTTTTGTDL---------APRL--RAF   32
 488601079  ------------------------------------------------------ML--ISY    5
 297565028  ------------------------------------------------------MSL--LSV    6
 257060308  ------------------------------------------------------ML--ESQ    5
```

```
1040961339  KICLEP---N-----KAQRAQFASFAGSARWAYNFALAIKIGYQKR-WFEARKQFIESGL  111
1405224212  KYALEP---T-----PRQQCAFSSHAGAARFAYNWGIARVADSLDA-----YAEQKAAGI   65
1206162758  RVCLDDAALT-----PDQRTLLDRHAGTARAVWNWGLAARNAQQDALMAHVRAVALEQAA   67
1009979974  RFRLDP---T-----GEQQGILSQQAGAARVAYNMMCTLNKDILEA-RSQLYSTLIKDGK   69
 451770599  RFALDT---T-----PAQLTMLRQHAGAARWAYNHALGVKFAALDE-RKTVIAGLVEQGL   81
1305448890  KFTLDP---T-----RAQVGALQQHAGAARWAFNWALGEKVAAHRE-WRRQVGALLAEGV   59
1706614983  KFRLDP---N-----QAQKAELMRCVGAARYTYNLLNAYNLQILRN-EQEYRNTRNAEGA   67
1706842133  KHRLDP---N-----PAQATLLAQYAGAARVAYNMLIAHNRAALAA-GAARRTELAETGL   83
 488601079  RFRIYPS--------KTVQAKLNEQLELCRWLYNRLLSEVNKARKE-GRRIRRE------   50
 297565028  KCKLIPDAST----AEKLSRTVNQFANACNY-----ALQVARRDNI-WNKF---------   47
 257060308  PITVACKLQVANTLAKEIDETMMVFACACDW------VNQNTPEKM-TNKT---------   49
```

```
1040961339  DEKAAGKKASEQVGRMPNYM-----SIATNEWTQLRDEV--------------------  145
1405224212  D------------EPDVKFPGHFDLC-KMWTAWKNTAE---WTDRHTGQTTTGV----  103
1206162758  SDETAAAELLDDRDWRTATIKAAPDELRRPLRAATLGRAF-------TAETRDPD----  115
1009979974  TKDKAKKELKAAAKEDPSLAIVWARDFDKNYITPERNRHK---HAAQRIAAGENPVDVWN  126
 451770599  DPKTSAAQAPK-IPTKPAIQ---------KALNTTKGDDR--------ISAAGD------  117
1305448890  AEEQARKQVRVPVPTKPTIQ--------KRLNSFKGDSR--------VQDLPDGVLG--  100
1706614983  DYETINGEIRKLRKKDPAYKFLGHAEYEKRYLTPEKQRHE---AIAQAITDGADPAVVWS  124
1706842133  AGPELAARMKAERAADPTLRVASYQSYATAHLTPLIRRHR---EAAAAIAAGADPAEAWT  140
 488601079  ----------------DTQ-----SL----------------------------IVR   58
 297565028  ----------------ALQ-----RA----------------------------VYA   55
 257060308  ----------------AMQ-----SL----------------------------VYQ   57
```

```
1040961339  ---------------CPWY----PEVPRRVFVGGFQRADAFKNWFDS--KSGRRS------  180
1405224212  --------------PWV----ASNFVGTYQAZLRDAAGWQRFFRA-----RKT------  134
1206162758  ---SRF------A--WWAVERHGVNRFAVSSXLQALDAFDRYYRD--TGGHRSARRARP  162
```

```
1009979974  PDEERF------NEPWL---HTANRRVLRSGQKQYEQALDNFFKS--QNGSRA------  168
 451770599  ------------CPWW---HTVSTYAFQSAFADADTAWKNWLAS--LTGKRS------  152
1305448890  ---PRR------PCPWW---WEVKTYCFQAAMADADTAWKNWLSS--LTGARA------  139
1706614983  -ETERF------AEPWL---HTIARRVLVSGIKNADKAWDNYNKS--RMKQRA------  165
1706842133  -D-ERY------AEPWM---HTVPRRVLVSGLQNAAKATENWMAS--ASGTRA------  180
 488601079  IKREEK-------------PELS-KVYSKVLQMVNYQLRSNISS-LNELR----------  93
 297565028  DLRAEY--------------------GLSANLAIRAI-------ARVGK----------  77
 257060308  DVRVNF--------------------GLSSNLAIQAIRRVCAN-RKTAK----------  85

1040961339  ---GAAMGWPKFKSKSKSRESFVIA-NDVQPAFVANLNRYI-----------KTGELA--  223
1405224212  ---GARAGRPRFKRGRARDSFQIH-------------------------GDGLR---  161
1206162758  RKDGRPAAWPRFKRGRATDAFALF--NLV---------------------VAGQDPWR  198
1009979974  ---GQKMGKPRFKKIRSTDSFTIDAVDVSSSTTLIRDIGPKDHARY-----KTGEASTG  220
 451770599  ---G-PIGAPRFKSKHRSRDSFRIH-HDVN---------------------NPTIRP--  183
1305448890  ---GQRVGYPRFKKGRARDSFRLH-HDVK---------------------KPGIR---  170
1706614983  ---GARMGIPRFKRKGVSRDSFTVP-HETTGAYGAYYHKKDPEYARRKVQLKRRGISAKP  221
1706842133  ---GARVGLPRFKKGRSRDSFTIPAPEVIGAAGTPY-------------KRGEPRRG  222
 488601079  -KKGVKVGWLRYKTS---PNSFKTLNFNQSGFKID--------FD-------RK------  128
 297565028  -RKGHKAGGFKATSV---DYDQRILSVNVDTETVS-----LSTVD------GR------  115
 257060308  -QKGKKVKEFKPTSI---SYDARIFSFRESDWTVS-----VKLLN------SR------  123

1040961339  -DMDYRHIKVPKCGEVRLTPGSAGQLGDLGRTML----------------------AE  258
1405224212  -IVDAKHVNLPKIGTVKTFEATRKLARLAKGSVPCPTCRATGKITDSASGKVKKCSDCK  220
1206162758  VIEGAHRIKVPSLGSLRVHENTKRLRRLIARG--------------------------230
1009979974  IIADYRHVRLSHLGTFRVFGSTKALVRQLDRG--------------------------252
 451770599  -DDGYRRIIVPRLGSLRVHDSTKRLKRAIDRG--------------------------214
1305448890  -LAGYRRLRLPTIGEVRLHDFGKRLARLIDRGR-------------------------202
1706614983  TITDYRHVRLASLGVIRTHNTTKPLVKAVRAG--------------------------253
1706842133  VITDHRHLRLASLGTIRTYDKTSRLVRACRRG--------------------------254
 488601079  ------KLSLSKVGD-----IPIRLHRSIGGK---------------------------149
 297565028  -----VKVPMRIAG------YQRHLLRTAKSI---------------------------136
 257060308  -----QRIKL-LIGN-----YQIGLLKSKNPT---------------------------144

1040961339  AKTGELITRITSGTISR-LEDVWYVSLVISG---PFVP-DA-----------ISVRQRR  302
1405224212  AAGSRPAARIVRGTVARDSAGVWYLATVEL---VREV-RT----------APVRQLA  265
1206162758  -------GRPTSARFTR-TGGVWYMSVVVAL---PATAASTVLSPAGAPRAAAPVAQTR  279
1009979974  -------GRIKSCTVSR-SADVWYVSFLVEL---PIEIARS----------TPVKQYK  290
 451770599  -------AVIQSVTISR-GEHVWYASILVKA---PAA--HA----------APVRQRQ  250
1305448890  -------AVVQSVTVAR-CEHVWYASVLCKV---DQSV-PQ----------RSVAQRR  239
1706614983  -------AEIKSFTVSR-AADVWYVSILVEL---TRP--ST----------APVAQRS  289
1706842133  -------AQIRSMTISQ-AEGVWYASILVAD---PTPI-RT----------GPSRRQRA  291
 488601079  -----IKGVI---IKRT-KSGKWYAIVQAEV---DKQP--L-----------PPIT------179
 297565028  -----QGGQ----LVRG-RDSSWYIHLWCEY---DDPP--V----------LDPV-------165
 257060308  -----SAT-----LVKR-KSGNVYIHITLDE---PTQP--E----------AKVT------172

1040961339  NGVVGVDLGSGRFYATTSG-----------GV-------LSIINPFVSKYEQEVANVG  343
1405224212  GGPVVGVDFGVV-RQVATLST-----------GV-----QLVDNPHLESHLRVKTAQQ  305
1206162758  ATLVGVDLGIV-KTLATMSD-----------GV-----TLVANAHGRAAARVLAVLQ  319
1009979974  ATAVGIDLGVV-KSLAALST-----------GV-----EIIPNPFLRTADKIKVLQ  330
 451770599  ASTVGVDLGVV-HHLAALST-----------GV-----DIIDNPHLAAGQKLTVAQ  290
1305448890  RGRVGVDLGVV-KHLAALSD----------PLHPYDRASLYVENPHLRRAAQLAVAQ  287
1706614983  AGAVGVDLGVV-RYLAALSDEQAPQRFAQYPSLEFTSDGAPTLANPWARAAEKLVVLQ  348
1706842133  NGAVGVDLGVV-KHLAALST-----------GV-----EVIDNGPGARQAALTVLQ  331
 488601079  GRAIGIDVGIV-THFCVDSD-----------GV-----NYFEHPYLDRTLEIKVVQ  219
 297565028  QEMLGVDLGIV-VNIATDSD-----------GV-----ETYSG-HLNSVRHHRVLR  204
 257060308  DKVLGVDLGRV-TDIATTSG-----------GV-----ESWSG-QITAKRNVYALRT  211

1040961339  ALAKTA---KG---GAARKALARLRVHARSALAVDGFSHQVSAWVSQF----AGVAV  393
1405224212  ALSRCP---PG---GRRAKAQQRLGVLHARVRHLVNSLQQATSAVHQH----SVVAV  355
1206162758  RAARQQGPRKGAAPGKGWVATQRTITVLQHDTAARVRGLVYEVTKTVAGYV---AAVAI  375
1009979974  KIARCQ---KG---GKNSIRLKRRVAVCHHELALQVAGYLNEVTSMVSSF----SAVAL  380
 451770599  ALSRTE---KG---GNRVRRAAARVGVRHHEITERVATTLHTVTKHVTNWV---ATVAI  340
1305448890  ALARTQ---KG---GKRVAVAVRRVGVLHHEVAVRVGSTLHQVTKRVTGFV---AEVAV  337
1706614983  ALSRAQ---KG---GKRVARIVQQVAVHHHLVALRVGSGLHQVSKRVTGYV---TLVGL  398
```

```
1706842133  AYARTQ---PG--SNRREVRRQ AALQHGIALR AGLLHQ STR ATDF----AV AL  381
 488601079  QLSRKQK------GSKNREVRIG A LYEKLENQ RIDFLHK SRYYVNNY----DI V   269
 297565028  K QK----KG-TKGA RR---- K LSGKETRFSNHVNHT SKR VAKAQRTERA AL  253
 257060308  T QKKAS--KG--TRSS RRCRQL A RLSGKERRFQKHINHE SRQ YNNAVTNKQA AI  267

1040961339  DKFDLASMLAS--AK TV KPGKNV----DVKARFNAH AD --V IASTIDK  YKGKR  444
1405224212  DGWDVQQTAQH----ASPKNLPKQI------RRNRNRA LD --T IGAARWQ  S  AW  402
1206162758  DDLNVAGMTATPAARPDP RPGVYLANGRAAKSGLNRA LD --V FGEFRRQ  K  PM  432
1009979974  DDLNVAGMTSS--AR TV NPGKNV----KQKAGLNRS LD --I PGRIRTL  K  TD  431
 451770599  DDLNVAGMTRS--AR TI NPGTNV----RAKAGLSRA LD --T PGELRRQ  K  GW  391
1305448890  DDLHVAGMTRS--AK TI DAPSRNV----RAKAGLNRS LD --T PGELRRQ  K  CW  388
1706614983  DDLAVAGMTAS--AA TI DAPGKNV----RQKAGLNRS LD --A FSTLRRQ  K  SW  449
1706842133  DDLNVAGMTRS--AR TL DAPGRNV----AAKSGLNRA LD --A LGMLRRQ  K  SW  432
 488601079  DDLN-----------VK M-----AEN-GSSTTLNRH TD SAW KFV--RL  E  ER  308
 297565028  DDLQ-----------IR R-----VRL-RRPQ----RATL-HSW FFDLGQK  K  ER  291
 257060308  DDLT-----------IR R-----TN--RKPRSKKDKRLG-NNW FYQLRQF  K  IL  308

1040961339  D CRVQV NTLDNS TT AKCGH--TCVCGPEQ T T DCGY-NAP QL ISA Y QL  501
1405224212  YG TT VVTDRHAPT RQ SACGTVKATPIPPTQDE R PACG T-SL R T  V AAV  461
1206162758  YG AQ LE ARFAPT KT STCGA-VRAKLRLDE T R HCGL-VI RD IAA NIAAL  490
1009979974  R VE QV IDRFFPS QL SSCGS--KTTIPLAQ IH DVCGG-VI RD IAA NIVYE  488
 451770599  YG ST AICDRWFPS DQ TE KV--RTKLRLSQ VT PACG YGPI RD H A NIAAY  449
1305448890  YG SR AV DRWWPS KT SA G R-QNPRLTLAD TH GCGL-RI RD IASRNIATH  446
1706614983  YG SQ QI IDRFFAS DT SA G ARAKTKLDLRV VE AACG V-RI RD IAA NIRAE  508
1706842133  A SQ KMI DRFAPS KA SR CG T-VKSTLSLAE TE ACHL-VI RD IAA NIRAW  490
 488601079  A RT VK NP-KNT KR AMCG Y-IVNNLKLHD TT PICGW-EA RD IASLNILDV  365
 297565028  A VPLVF VDP-RNT RQ PA CG H-AERANRPTQAL R VACG Y-SGAA D V A NIAVR  348
 257060308  B VK IL VNP-AYT LS HK LV-IGDR---KG G S NCGN-KC D IGA NIKAL  362

1040961339  TVG----------FDELGLDMTASLTPD--TGKRPIAFMTS--AH----------------  532
1405224212  AQH------HDAPSGGESKNARGENTRPT--APRRNGQFSAK--REPRSRPPGRGQTGTP  511
1206162758  HQA---LGTSPADAGDAKRRDHKEGDRERSVILVSQDLGPPR--SAERVTDSSPPTRAA-  544
1009979974  KRL-------VEQKCSEHSAPEGAEDKRPWSVHLPSTQYVDGLYTRKRQGPSGHSS------537
 451770599  AVAS-DTGETLTARRDTAEAPTRVGRRRGAVDAGRPHRETGA--ATPAEQPAGHPKHADQ  506
1305448890  ALA--DTAPPVAPDRGETQNARRAGTRPT---GPRAGRHPAT--KREDTAPAVPPQRSNP  499
1706614983  VRM---YEAQLAPGMGESLNGRGVTDSDAAVSVVLGDAALDA--SRPAAMGGGSP-------558
1706842133  VQEERGAGVELARGRRESRNGRGAAVSGP-PSGGAAGQGRGS--VKPAPQGVGMSSRATG  547
 488601079  MGRSRTP---VE-GEPL---PCVISYR-------EVIAG--QVLSMKQ---EVPSVRAE-  405
 297565028  WAAVNRP------YLGE---ASRVSLH-------------------G---SVPGSPRL-  375
 257060308  AIINRPG------GSGL---SCKLKTN------VQYIQL--SLFEGLG---LLKTSTSA-  401

1040961339  ----------------------  532
1405224212  GT--------------------  513
1206162758  ----------------------  544
1009979974  ----------------------  537
 451770599  RTLPLVS---------------  513
1305448890  LALPPPPTGYEQVTLF------  515
1706614983  ----------------------  558
1706842133  WSSQPPSTEGESAVRGASALAR  569
 488601079  ----------------------  405
 297565028  ----------------------  375
 257060308  ----------------------  401
```

V-U5 family

```
1002781834  ----------MST IT-IQCRLVAPE ATRQALWDLMAQKNTPLV SELLRQVAQHPDFET N  48
 257060041  ----------MST IT-IQCRLVAPE ATRQALWDLMAQKNTPLV SELLRQVAQHPDFET N  48
1405885249  ----------MSV IT-IQCRLVAED DSLRQLWELMTEKNTPFINEILLHLGKHPEFET N  48
 440680311  ----------MSV IT-IQCRLVAED DSLRQLWELMSEKNTPFINEILLQIGKHPEFET N  48
1111222001  ----------MTQ IT-VQCRLVATE TTRQELWDLMADKNTPLINEILKQVAEHSDFEA N  48
1016058386  ----------MSV IT-IQCRLVAED DTLRQVWELMTDKNTPLVNEILLAQVGKHPEFET N  48
1002781021  ----------MSQ IT-IQCRLVAKE PTRQTLWDLMAELNTPFINEILLDQKVAQNPDFEQ N  48
1002400088  ----------MTLK T-LECRLYAPSDTLRYLWLLMAEKNTPLINEIINHLSEHPDFDQ N  48
1002400093  ----------MTH ITVVQCRLIAPE STLQHIWK MMAQQQTPLINDLLHD INTHPDINT N  49
```

```
1000428792  --MCFVWYFIFMSQKT-IQCRLIASSSTRQKLWKLMAESNTPLINELLQQLSKLPDFEK  57
1004799941  -----------MSVIT-IQCRLIASSATRSYLWDLMAQKNTPLINELIEQLGIHPEIN  48
1004798910  -----------MSQIT-IQCRLVASSPSRHQLWKLMVDLNTPLINELIVQVAQLPEFEN  48
1002125473  -----------MSQIT-IQCRLIASSSTRQKLWKLMATLNTPLINELIEQLGKLPDFEN  48
1007062327  -----------MSVIT-IQCRLVADSTLRHLWFLMAEKNTPFANEILIQLAQLPFEN    48
1007109870  -----------MSQIT-IQCRLVASSTTRQQLWDLMAEKNTPLINELLSQIGKLPEFEN  48
 428305730  ----MSDQTSQNNEIT-IQSLISAPSTLCFLWDLMVNV-TPLINLLLIHLNTLPDFDIN  54
1030021203  MSRDQKKSTSPIHRT-IRCHLHASEDVLRKVWEEMTQKNTPLIVQLLKSVSEQPEFETN  59
  17230183  MSRDRKKGSKSPSLRT-IRCHLHTKEDVLRKVWEEMTQKNTPLIVELLKSVSEQPEFETN  59
1206817639  -----------MSIIT-IQCRLVANDRTLQHLWILMAEKITPLISLLLQLGKLPDFEN   48
1300380544  -----------MSTIT-IQCRLVAEATLRYFLWLMAEKITPLINELLCQLGQLPDFDTN  48
 220908491  -----------MSMYIT-IHCHLVASSPVRRHLWKLMAGSITPLVNILLLVSQLPDFEN  48
1085045623  -----------MSKIT-IQCRLVASSATRQYLWILMADIYTPFNEILRQLREDDNLEDN  48
1085056708  -----------MSQIT-IQCPLIASSSTRQKLWKLMAHLITPLINELLQQLSKLPDFEN  48
1406824382  -----------MSQIT-IQARLISFISNRQQLWKLMADLITPLINELLCQLGQHPDFEN  48
 428311644  -----------MSIIT-VQCRLKATKDSLRHLWGLMVEKITLVNELLCQLNTLPDLEN   48
1085030414  -----------MSQIT-IQCRLVAISARQQMWRLMAEIITPLINELLAQLGQLPDFEN   48
1300485477  -----------MSVIT-IQCKLVATLTRRALWILMAEKHTPLINELLCHLAQDSRIEN   48
 428312908  -----------MSQIT-IQCRLVANVSTRHQIWKLMAERITPLINELLCQLGQLPDFEN  48
1016001624  -----------MAHVT-IQCRLIASRDTRQFLWDLMAQKITPLINEILLRLKQHPDFLH  48
1507700957  -----------MSDNA-IQCRLIAPITTRRQQVDLMAEKITPLINELLCQLAELPELEN  48
 428306136  -----------MSQIT-VQCRLVASSSTRHHLWKLMADLITPLINELLARLIAQLPFEN  48
1019531886  -----------MSQIT-IQCRLVAKIATRQTLWDLMAELITPLNELLQQLAQYLDFEN   48
```

```
1002781834  RQQGKLEAGIIKKLCEPLKKDPRFYEQPARFYTSAISLVDYLYKSWLKVQRRLQNKLEGQ  108
 257060041  RQQGKLEAGIIKKLCEPLKKDPRINEQPARFYTSAIALVDYLYKSWLKLQQRLQRKLEGL  108
1405885249  LEKGRIPAESLKTLGNSLKTQEPFTGQPGRFYTSAIALVDYLYKSWFALQKRRKNQIEGK  108
 440680311  LEKGRIPAELLKTLGNSLKTQEPTTGQPGRFYTSAITLVDYLYKSWFALQKRRKQQLEGL  108
1111222001  RQKGKIPTGTIKQLCEPLKTDSRFIGQPGRFYTSAIALVSYLYKSWLALMKRLQYKLEGK  108
1016058386  LEKGKIPTEFLKTLVNSLKNQERFSDQPGRFYTSAIALVDYVYKSWFALQKRRKRQIEGK  108
1002781021  RQKGKLQKSVIKQLGDKLKKDPRYLGQPARFYTSGISLVEYIFKSWLKLQQRLQRKLDGK  108
1002400088  FKAKQIPKSAISDICNDLKSQENYQNQPGRFYSSAISLTHYMFKSWFAVHKQLQRRIEGK  108
1002400093  LTANQLPSKLVETLAQPLKTQSPYQGLPGRFITSAIILVKEMYASWFAIQTQKRLSLEGK  109
1000428792  RRNGKLPSTVVSQLCQPLKTDPSFTGQPSRFYISAIHIVDYLYKSWLTIQKRLQQQLDGK  117
1004799941  LKKGKLPDGVVKPLCDSLITQESFANQPKRFNKSAIEVVEYIYKSWLALQKERQQTIDRK  108
1004798910  RQKGKHPAKIVKELCEPLRTDPRFIGQPGRFYTSAIATVNYIYKSWFALMKRSQSQLEGK  108
1002125473  RQQGKLPTTVVSQLCQPLKTDPRFVGQPSRLYMSAIHIVDYLYKSWLAIQKRLQQQLDGK  108
1007062327  VKNSRVPATVIKELCDSLKNQELFAGQPGRFYTSATTLVTYIYKSWLAVNKRLQRKIEGK  108
1007109870  RQKGKHPTGIVKELCEPLKTDPRFIGQPARFYTSATASVNYIYESWFALMKRYQSQLDGK  108
 428305730  LTQCAIPASVINSFIKTATQSPYREMPIRFITSAQNLVIEIFKSWFELQRLKRLSLKII   114
1030021203  QEKGTISKKEITKLRKALTNDSDLQQQSGRLGSSADSLVTEVYTSWLTLSQKIKKQKEGK  119
  17230183  KENGKITKKEITNLRKSLTQDS------GRLYSSVDNLVQEVYSAWLTLYQKRKKQKEGK  113
1206817639  LKNGKVPKDTIKILCDSLKTQSRFAGQPGRFYTSAISQVKEIYKSWLTLQKRRQRQIEGK  108
1300380544  VQAGKMPEKTVENLCKSLEDREPFANQPGRFRTSAVALVKYIYKSWFALQKRRADRLEGK  108
 220908491  QRKEDISKSSVEALCEPLKET--VPGQPGRFYSSAILRVIYTIKSILAIQKNRRYRLEII  106
1085045623  RQSGKIPASVFEDYRKTLKTESRFQGMPGRWYYAGREEVKRIYKSWLALRRRLRNQLAGQ  108
1085056708  RKNGKLPSTVVNQLCQPLKTDPSFTGQPSRLYLSAIHVVDYLYKSWLAIQKRLQQQLDGK  108
1406824382  QQKGKLPSTVVSQLCQPLKTDPRFAGQPSRLYMSAIHIVDYLYKSWLAIQKRLQQQLDGK  108
 428311644  LKVENIKAEVIEGICDNLRTESRYQDMPGRFANIACKLVKDIYKSITAIQEERRFRLWRK  108
1085030414  RQKGKLPSTFISQLSQSFKSDPRFLGQPSRFYKSAFNAVEYIYKSWLALNKRLQQQLDRK  108
1300485477  SLTGKLPRLVVSEACNQLKQDPQFSGQPGRFYSSAISTVHRIFLSWLAIQTRLRNQISEQ  108
 428312908  RQKRKIPAGIIKQLCEPLRTDSRFIGQPGRFYASASALVDYYKSWLKLQQRLQRKLEGL  108
1016001624  RTKKRLPKDFLARQIAELKNNYPFEEQPSRFYASVNKVIDYLYKSWFEVQKALDWKLQGN  108
1507700957  KRKGKIPPGTVKNLCQPLRTCPQYINQPGRFYSSVISLAEYIYRSWLKLQRRLIFRLNGQ  108
 428306136  RKKGKLPDGIVKQLYQPLKTDPRFTNQPGRFYTSAITVVDYYKSWLTKLQQRLEQKLKLQ  108
1019531886  RQRGRLTAKVIEQLGNELKKDPRFLGQPARFYTSGISLVEYIFKSWLKLQQRLQRKLDGK  108
```

```
1002781834  NRWL-VMLKSDEELVQISDSSIETLQAKATDILSTLPE-K---PDKFPETSTTKGKKSK    163
 257060041  NRWL-AMLKSDDELIQISDTNIEILQAKAIEILSTLQPQDR--EQSSKKKAKKCKKSTNK  165
1405885249  QRWL-KMLKSDPELEQESDSSIEVIRTKAIELFSKFTPQSDSEALRRNQNDKSKKGKKTK  167
 440680311  QRWL-KMLKSDQELEQESDSSIEVIRNKAIELFSKFTPQSDSEALRRNQNDKQKKVKKTK  167
1111222001  IIFWL-EMLKSDEELVEISGVNLEALRSKAAEIIAQFSPQID---TNNSQNKRKTKSKNSK  164
1016058386  ERWL-IILKSDLQIEQESDCSINVIRTEAIEILAKFTPQSD------QNKNQRKSKRTR   160
1002781021  RRWL-TVLKSDEELIQQSQQNLEAIRHKASEILQTHQKTEK----------------------148
```

SUPPLEMENTARY INFORMATION

```
1002400088  RRWL-NLLKSDQELEQNCGQSLEIIIQKAEEILKLMDSEKS---QSSSKPKKPKKPKKKK 164
1002400093  KRFL-TILKSDKQLIQDSQTDFLTLCYKAQQLLKRTQNKLK-------LDEPQHSEKAHW 161
1000428792  LRWI-EMFNSDVELVEISGFSLEAIRTKASEILAITTPESD----PKTLLTKRGKTKQSK 172
1004799941  EHWL-KMLKSDELEQESKCTLDAIRSQATKILPKYLAQSE---QNNNQTQSQNKKKSK 163
1004798910  MRWW-EMLKSDAELVEVSGVTLESLRTKAAEILSQFAPQPD---TVEAQPAKGKKRKKTK 164
1002125473  MRWL-EMLNSDVELVETSGSSMGAIRTKASEILAKAMPTSD----SDSSQPKTKKGKEAK 163
1007062327  KQWL-DMLRSDTELEQESNSNLEKIRAKATEILDSFATRQINQVNSKSKTSKNNKNKQEK 167
1007109870  LRWL-EMFNSDAELVEHSGVSLDTLRATSAEILAQFAPQDT----NRDTSNKGKKSKMGK 163
 428305730  RHFF-TILKSDEELQRESQCTFDTLCQEAIILQQTQLQIEQKRLDANQPPPTLDKQAFW 173
1030021203  EYFLNNILKSDVELVQESNCDLQTIRCKAQDILSQPKEFLEQITNHDAVLNQTKSARKKV 179
  17230183  EYFLNNILKSDIELVEESNCDQILRAKAQEIISNPQDILRQITIDNPKDKPTKSIQKRV 173
1206817639  QRWL-GMLKSDVELQEESNCSLEKIRAKGTEILAEFVSKFTKDTTKKSKTKIKSTKKSNK 167
1300380544  ERWL-KMLKSDVELERESNCSLDIIRAKAGEILAKVTEGCA----PSNQTSSKRKKKKTK 163
 220908491  QRWL-HVVESDSELLHRSGSSLETLKYRAQDILYQLNVEQQTQKTSVTDISTITPEETPS 165
1085045623  NRWL-EVLQSDETLMEVSGLDLSALQAEASQLLNILGSKNKTSKNRSKKAKGKPKGKSAK 167
1085056708  IRWL-EMLNSDAELIEISGCSLEAIRTKAAEILAIATPDSD----VAAPLTKTGKAKKSK 163
1406824382  TRWL-EMLNSDAELVELSGDTLEAIRVKAAEILAIAMPASE----SDSASPKGKKGKKEK 163
 428311644  QRWF-SLLRSDLELEQESGLSLEKLRTEATKILIKAQLECSREAEPDQATTDNSSA------163
1085030414  RRWL-EILQSDTELAADSNCSLDAIRSKAAILSQAIQAPS----LDSSPPRGKKGKKSK 163
1300485477  TRWL-AMLQSDNELTIASQTDLNTLRLKASLLTHLNEPIS---ESDQPEVKKTRSKKKN 164
 428312908  MRWL-GMLKSDEELVNQSGCTLEVLRTKASLILVPLTSKNE------SPQPTQTKGKKSK 161
1016001624  LRWV-EMLLPDTELIKHFDNSLESLQQQATLILDSIDSTVS-----------------------148
1507700957  QRWL-QMLKSDEELVAESGRSLKEIEAKASEALDRLNREENPS--------------------150
 428306136  IRWL-GMLKSDEELAAESNTSLEVIRTNAAELLTSLSSEDGS-----------------------149
1019531886  RRWL-EVLKSDEQLIKDSQTDLETIRQKATEILQSYEGTER----------------------148


1002781834  KY--------KNNNS----------------LFTQLYNLYEKADDTLTHCAIRYLLKNGC 199
 257060041  NS----------------------------LFSQLQKLNEINNN.TCIRYLLKNGG 196
1405885249  KP------TKAKTSS----------------IFKILLNTYEEAEDPLTRCALAYLLKNNC 205
 440680311  KS------TKPKTSS----------------FKIFLSTLGEAEEPLTLCLLAYLLKNNC 205
1111222001  KS-----KSSDISKN----------------LSQQLFDTYRDTEDIVTHCAISYLLKNGS 203
1016058386  KS-----AKLQ-TPS----------------LFQNLLNTYEQTQETLTRCAIAYLLKNNC 198
1002781021  --------------------------------LFNHLFQLYQEEKNPFTHIALCYLLKNRC 177
1002400088  KS------SSEETIT----------------LFDRLFKAYNQGNDSLESYALAYLLKNNG 202
1002400093  S------------------------------IINALYPAYNNAKTPISRAAFALLIKNNG 191
1000428792  KS-----SASNPDRS----------------LSRKLFDAYQETDDILSRSAISYLLKNGC 211
1004799941  KS-----KTKNENST----------------LFDILFKAYDKAKNPLNRCTLAYLLKNNC 202
1004798910  KSDGDCAERTLRERS----------------ISDYLFEAYRDTEEILTRCAINYLLKNGC 208
1002125473  KS-----SSSSSDRS----------------LSNKLFEAYQETEDILSRSAISYLLKNGC 202
1007062327  EV--------KSLSI----------------QSNILFETYRQTEDNLTKCAIVYLLKNNC 203
1007109870  KS-----QKSDSEGN----------------LSKKLFDAYSSAEDNLTRCAISHLLKNGC 202
 428305730  E------------------------------VAHVLYKADQADTPLVSLIALLLKNQN 203
1030021203  QNSSNEINASKQREN--SDYQENVDKNIPETLTEILYKIHKITQDILTQCAVAYLIKNHN 237
  17230183  RKNINASNSDTAKNNILNSQEKTTEENISKSVIEILYEIHKTTQDTIIRCAVTYLIKNYT 233
1206817639  KT---KKDTEESNST----------------LFQALCDIYDKTEDTLSKCAIIYLLKNNC 208
1300380544  KS-----QATKDLPT----------------LFEIILKAYEQAEESLTRAALAYLLKNDC 202
 220908491  SK----------NES----------------LFPSLFRAQSEDDILSLCLIALLKNGG 199
1085045623  DPT----------------------------LYQALWELYRETEDIAKKCVIAYLLKHKC 199
1085056708  KS-----SASNPDRS----------------LSHKLFDAYQETDDILSRSAISYLLRNGC 202
1406824382  KP-----SSSSPKRS----------------LSKTLFDAYQETEDIKSRSAISYLLKNGC 202
 428311644  ------------------------------LWDNLFTADKFKSPRLKCVALLKNYC 192
1085030414  GR-----SSSSPVSS----------------LFANLFKAYQETDDIKCRCAISYLLKNNC 202
1300485477  QT------SNQAGAN----------------VSRTLFKLYDETEDPLTRCAIAYLLKNGC 202
 428312908  KP-----QDLGSNRS----------------ISKTLFQVQDIEDLLTNLICLLKNYC 200
1016001624  -----------HDR----------------ISTILFEKCGKTKKPEIKSAIIYLLKNGC 180
1507700957  ------------------------------ISNRLFDLYDETEDILIRSAIVYLLKNGC 179
 428306136  ------------------------------VSTRLWKTDETDDILTLCVLCLLKNLS 178
1019531886  ------------------------------LFNSLFQAYRDEQNILTQTALNYLLKNRC 177


1002781834  KLPQK-PEDPEKFAQRRRKVEIKLSLIEQLE-SSLPGRDLTGDSWLETLILNTATV 257
 257060041  KLPQR-PEDTEKFAQRRRKVEIKLSLIEQLE-SSLPGRDLTGDSWLETLILNTATV 254
1405885249  QLSEL-DENPEEFTRNRRRKEIELSLKDQLQ-SRIPGRDLTGEQWLETLILVKVPQ 263
 440680311  QLSEL-DENPEEFTRNRRRKEIELSLKDQLQ-SRIPGRDLTGEEWLETLILFNVPQ 263
1111222001  KLPNK-EESPEKFTQRRRKIEIQLSLREQLE-SRIPGRDLTDLNWLNTLILNQVPQ 261
```

```
1016058386 QISER-DEDPEEFNRNRRKKEIEIERLKDQLQ-SRIPKGRDLTGEEWLKTLEIATTNVPQ 256
1002781021 KLPQK-PEDPEKFTKRRRKVEIAIDRLQEQLE-GRLPQGRDLTNHNWETLSIACQTDPE 235
1002400088 QIPED-DEDLDKFALRKRKKEIEIERLQQQLE-NRIPLGRDLTGELWQEMLTIVNESIPQ 260
1002400093 QVPDT-PENPDYYQQRRKRKEIQRRLEEQLK-ASLPKGRILDSKHWENTLKLAQTPI-T 248
1000428792 KLNDK-EENPEKFAKRRRKVEIQIQRLTDKLT-SRIPKGRDLTYSKWLETLFTATTTVPE 269
1004799941 QVSQK-DEDPNQYALRSKKEKEIERLKKQLQ-SRKNGRDLTGREWQQTLIMATSSVPE 260
1004798910 KISNK-EENAEKRRRKLEIQIERLREKLE-ARIPKGRDLTDAKWLETLLLATLNVPE 266
1002125473 KLSDK-EEDSEKFRRRQVEIQIQRLTEKLI-SRMPKGRDLTNRKWLETLFTATTTFPE 260
1007062327 EVNDV-EEDIEEYEKNRKKEIQIKRLEDQLK-SRVPKGRDLTGEKWVEVLEKAVNSVPE 261
1007109870 KVSNK-EENSEKFTQRRRKLEIQIQRLTEKLA-ARIPKGRDLTDTQWLETLFTATYNVPE 260
 428305730 QVPEQ-FEDPKKYQQRRQAKQEEIERLEHQLQ-SKAPGRLTDQQELRILKKLQPI-T 260
1030021203 QVSDL-EEDIKKLKKRRTEKQVQIKRLEEQIQKNKLPNGRDITGERYNHAFDNLINQVPQ 296
   17230183 KISDT-EEDLNKLKERRAEKEIEIERLEKQIQDSRLPNGRDITGARYLEAFDKLINQVPK 292
1206817639 QVIDT-EENPDTFLKRKRAKEIEIKRLQDQIV-GRIPKGRDLTDKKWLDTIKLASSQVPQ 266
1300380544 EVSEV-DEDSEKFKKRRRKKEIEIERLRNQLK-SRIPKGRDLTGDKWLKTLEEATRNVPE 260
 220908491 KIPET-EEDQEKFTQRVNSKREEIQLEIELS-ARQPGRLTSEELCTAIPQQLSE 257
1085045623 QVPDK-PEDPKKFRHRREAEIRAERLNEQLIKTRLPKGRDLTNEQWLQVLEIATRQVPK 258
1085056708 KLNDK-EEDLEKFAKRRRKVEIQIQRLTDKLT-SRIPKGRDLTNAKWLETLFTATTTVPE 260
1406824382 KLTDK-EEDSEKFAKRRRQVEIQIQRLTEKLI-SRMPKGRDLTNAKWLETLLLTATTTVAE 260
 428311644 QVNKV-EEDPEAVQRRRRKKIQIERLKEQLK-SRLPIGRNLSEQELICLAAIQIGLI-I 249
1085030414 QLSDR-EEDPEKFAKRRRKVEIRIQRLTEKLN-SRMPNGRDLTNTRWLETLAIATTSVPQ 260
1300485477 KLPDQ-NENPEKFIKRRRKTEIRLERLMNTFQTTRIPRGRHLSWHSWIEALETATSHIPE 261
 428312908 KIPDQ-EEDKEKFAKRRRKTEIKISREEQLA-SRIPKGRNLTGEKWLETLIAATSTVPE 258
1016001624 TIPKK-PETTEKYQDLKRKVEIKITKHRQIE-SRIPLGRDLEDKKWLDTLITASTTAPI 238
1507700957 KIRQK-PEDPKKFARRRRKTEIRVKRLQEKLN-GKAPQGRDLTGEKWLNTLFTATSQVPQ 237
 428306136 KVPKKPEILEKFAKRRRKVGIKFERLRRQLE-SRIPQGRDLTSKNLCTAIPTTAPA 237
1019531886 QLPKK-PEDAKKFAKRRRKVEITIKRLQKQIN-GRLPQGRDLTNDNWLETLNLACDTDPK 235


1002781834 EASEIKSWQDKLLSQSKSIPYPVAYETNEDLTWSINEKGRLCVRFN--GLGKHT------ 309
 257060041 EAEDVKSWQDKLLSQSKSIPYPVAYETNEDLTVSIIEKGRLCVRFI--GLGKHT------ 306
1405885249 NENEAKAWQAALLRKTANVPFPVAYESNEDMTWLKNDKNRLFVRFN--GLGKLN------ 315
 440680311 NENEAKAWQAALLRKTANVPFPVAYESNEDMTWLKNDKNRLFVRFI--GLGKLT------ 315
1111222001 SETEAKSWQNNLLRKSHAVPFPVSYETNEDMTWFKNRKGRICVKFN--GLSEHT------ 313
1016058386 NENEAKAWQAALLRKPADVPFPVAYESNEDMTWLQNDKGRLFVRFN--GLGKLT------ 308
1002781021 DTAQARSWQDKLLIQSQSIPFPINYETNEDLTWHKNEKGRLCVKFN--GISDLS------ 287
1002400088 DENEASAWQAKLLKKSHNIPYPVAYETNTDLKWSKDSRGHLLVTFN--GLVESLKKLNLN 318
1002400093 TIEEITSLQTQLLQKYSHLPFPVFYGTNTDLTWFKNPQGRICVKFN--GLNQYP------ 300
1000428792 NNAEAKRWQDILLTRSSSIPFPVVFETNEDLVWSTNEKGRLCVHFN--GLSDLI------ 321
1004799941 SNDEANIWQKRLLKKDISLPFPIRFRTNEDLIWSKNEEGRICVSFSGEGLNDHI------ 314
1004798910 NEAEAKSWQDSLLKKSITVPFPVAYETNEDMTWFKNERGRICVKFS--GLSEHT------ 318
1002125473 DNAEAKRWQDILLTRPSSLPFPLVFETNEDMVWSKNQKGRLCVHFN--GLSDLS------ 312
1007062327 SENEAKSWQASLLRKSSQIPFPVVYETNEDIKWSINEKGRIFVSFN--GLGKLK------ 313
1007109870 DETEAKLWQNSLLRKFSSLPFPVAYETNEDLVWSKNRFGRICLTFP--TLREHI------ 312
 428305730 EVAEFRDIQAQLLRKFFAVVPVTFSTIDLQFSTIDQGRICIDIF--GMSKYT------ 312
1030021203 NNEEFAEWIASLLNKVSDLPYPIDY-LYSDLTWYKNEQRKICVYFN--GWAKFH------ 347
   17230183 NNEEFANWIADASRKISNLPYPIDY-LYSDLTWYKNQDGKIFVYFN--GWSKYH------ 343
1206817639 DENEAKSWQNQLLKTSSSVPYSVDYETNTDIKWVKHNNGSIFVNFN--GLGEHQ------ 318
1300380544 NEDEAKAWQAQLLREASSVPFPVAYETSEDMTWFTNEQGRIFVYFN--GSAKHK------ 312
 220908491 TVAQAREWNDKILTQPKFMPYPIIGSSTDVRRKTSKSRITIFI--SIDKYLKAADPE 315
1085045623 DEDEAAIWQSRLLTDAAKFPFPVAYETNEDLKWFLNGKGRLCVSFN--GLSEHT------ 310
1085056708 DNVEAKRWQDILLTRSSSVPFPLIFETNEDLVWSKNEKGRLCVHFN--GLSDLT------ 312
1406824382 DNAQAKRWQDILLTRSSSLPFPLVFETNEDMVWSKNQKGRLCVHFN--GLSDLI------ 312
 428311644 DDEHLRQVQASLTRKQSPVPFSISYETSTDLRFSSIEQGRICISIGK-ISKHT------ 303
1085030414 DEAQARQWQDVLLTKPKSLPFPLIFETNEDLFWSKNQQDRLCVHFP--GLRDLA------ 312
1300485477 NEEEAAGWQARLLTKPAILPFPVNYETNEDLRWSLNSQGRICVSFN--GLSEHF------ 313
 428312908 NESQARSWQDRLLTKPQSVPFPVTYESILDLTISKISKGRLCIFI--LSEHT------ 310
1016001624 DQTEANTWFSILKQNQSSIPYPILYETNEDLKWSLNEKNRLSIRFS--GLGEHS------ 290
1507700957 DEAQAKSWQDILLTKSKLVPYPIVYESNEDLTWSKNERGRLCVKFN--GLSDHT------ 289
 428306136 DEPEAQSWQDTLLTESKLVPFPVAYETNENLTISKICKGRLCVDIS--GLSKHI------ 289
1019531886 DVEQSRTWQDKLLKKSQSIPFPINYETNEDLTWSKNEKGRFCVQFN--GISDLK------ 287


1002781834 ------------FQIYCDQRQLKVFDRFYEDQQIKKDG-----KDHHYSAYTSLRSYRYV 352
 257060041 ------------FQIYCDQRQLKVFDRFYEDQQIKKDG-----KDHHYSAYTSLRSYRYV 349
1405885249 ------------FQIYCDKRHIHYFDRFLEDQEILKSS-----KRQHYSSYTTLRSYRYA 358
```

```
440680311   ------------FEIYCDKRH.H.FDRFLEDQEILKNS-----KRQHSSS.TLRS.R.A  358
1111222001  ------------FEIYCDNRQLHWFKRFLSDQQIKKNS-----KNQHSSSLFTLRSGRIS  356
1016058386  ------------FEIYCDNRHLHYFKRFLEDQELKRNS-----KNQHSSSLFTLRSGRIA  351
1002781021  ------------FEIYCDQRQLKWFQRFYEDQQAKKGS-----KNQHSSALFTLRSGRIL  330
1002400088  PE----------FEIRCDRRHLPWFQRFCKDQEI-KAN-----NDQHSSALFVLRSARLI  362
1002400093  ------------FQIACNKRQYPWFQRFFTDYQSYKSH-----KQQVPTGLMVLRSARLL  343
1000428792  ------------FEVYCDSRQLYWFKRFLEDQQTKRKS-----KNQHSSGLFTLRNGRLA  364
1004799941  ------------FEIYCGNRQHWFQRFLEDQNIKNDN-----NDQHSSALFTLRSAILA  357
1004798910  ------------FQVYCDSRQLQWFQRFLEDQQIKRNS-----KNQHSSSLFTLRSGRIA  361
1002125473  ------------FEVYCDNRQLHWFQRFLEDQQTKRQS-----KSQYSSGLFTLRNGHLV  355
1007062327  ------------FEIFCDKRHLHYFQRFLEDQDIKRQG-----KNQHSSSLFTLRSGRIS  356
1007109870  ------------FQIYCDSRQLHWFQRFLEDQEIKKNS-----KNQHSSALFTLRSGRIA  355
428305730   ------------FEIACDRRQLNFFCRFLADYQLYQH-----KTQIPTGLIPLRCR.V  355
1030021203  ------------FQICCNKRQLHFFKRFLEDYKALKESEKG--EIKLSGSLVTLRSVQLL  393
17230183    ------------FQICCNKRQRHFFERFLEDHAWKESEKG--EVKLSGSLVTLRCVQLL  389
1206817639  ------------FEVYCDSRQPYFQRFCEDMQIWHND-----EEKYSSALFMLRSARLV  361
1300380544  ------------FQVYCDRRQLHWFQRFVEDFQIKKNGDKKGSEKEYPAGLLTLCSTRLR  360
220908491   IKAWFKDHQEYPFRLHCDERQLPTFDRFLDNQFYQAN-----KETYPAGLLT.L.G  370
1085045623  ------------FEVYCGQRQLYWFNRFLEDQQIKKEN-----QGERSAGLFTLRSGRLV  353
1085056708  ------------FEVYCDRRQLHWFKRFLEDQQTKRKS-----KNQHSSGLFTLRNGRLA  355
1406824382  ------------FEVYCGNRQLHWFQRFLEDQQTKRKS-----KNQHSSGLFTLRNGHLV  355
428311644   ------------FLVFCDDRQLHFFRFYDYKIFTQN-----KDQVPAGLLT.RSAR.V  346
1085030414  ------------FQVYCDRRQLHWFHRFLEDQQTKHSS-----KNQHSSSLFTLRSAYLA  355
1300485477  ------------FEVYCDQRDLHWFNRFLEDQETKKAS-----KNQHSSSLFSLRSGQIA  356
428312908   ------------FQIYCDDRQLKJFDRFLDQQIKRES-----KDQHYSS.TTRSGQ.A  353
1016001624  ------------FQLCCDHRQLPYFQRFYEDQELKKAS-----KDQLSSALFTLRSAMIL  333
1507700957  ------------FQIYCDRRQLKIFNRFYEDQQIKKAS-----KNSHSSALFTLRSATIA  332
428306136   ------------FQIYCDDRQLKJFDRFYDQEIKKAN-----KDQYYSG.TTRSGR.A  332
1019531886  ------------FEIYCDQRQLKWIQRFYEDQQVKKDG-----KDQHSSGLFTLRSGRIL  330

1002781834  WQE---GLG--------------------------------------KGK---PWNIHRLTLHCS  373
257060041   WQE---GLG--------------------------------------KGK---PWN.H.L.  370
1405885249  WLP---GEE--------------------------------------KGE---HVKVNQLNFYCS  379
440680311   WLP---GEE--------------------------------------KGE---H.KNQ.FY  379
1111222001  WQE---EVG--------------------------------------KDR---PWNVHLKLDCT  377
1016058386  WSL---GEE--------------------------------------KGE---PWKVNKLHLYCT  372
1002781021  WQE---GNE--------------------------------------KGQ---PWNIHRLILDCT  351
1002400088  WRE---GQG--------------------------------------KED---PWKIHQLYLQCS  383
1002400093  WQP---TNG--------------------------------------QGE---PWNTHLSLHCA  364
1000428792  WQQ---GEG--------------------------------------KGE---PWNIGHLALYCC  385
1004799941  WQE---NKQ--------------------------------------HKENSLPWNTRRLTLYCT  381
1004798910  WQE---GEG--------------------------------------KSE---PWKVNRLILYCS  382
1002125473  WQE---GEG--------------------------------------KSE---PWNLNRLNLYCC  376
1007062327  WLE---QPG--------------------------------------KGK---PWNINRLLLFCS  377
1007109870  WQE---GEG--------------------------------------KGE---PWDIHILTLYCC  376
428305730   WTE---GQDDFALIVATWLLIAVIQHKFYHIAWLLLKNHRIISP---W.H.L.  409
1030021203  WQQ---GEG--------------------------------------AGE---PWKVNKLAHCT  414
17230183    WQQ---GEG--------------------------------------KGE---PWKVNKLSHCT  410
1206817639  WLE---KKG--------------------------------------RGK---PWNVNYLYLHCS  382
1300380544  WKE---SAE--------------------------------------KGD---PWNVHRLISCT  381
220908491   WRE---GEG--------------------------------------KGD---W.IN.L.  391
1085045623  WKPYSSDAS--------------------------------------RSD---PWMANQLTLQCS  377
1085056708  WQE---GEG--------------------------------------KGE---PWQINRLTLYCC  376
1406824382  WLE---GEG--------------------------------------KGE---PWNLHILTLYCC  376
428311644   WQE---GEG--------------------------------------EE---WQHL  367
1085030414  WQQ---GKE--------------------------------------KGE---PWNTHYLILYCC  376
1300485477  WQE---GKG--------------------------------------DAE---HVVHRLVLSCS  377
428312908   WAL---GKG--------------------------------------GD---A.IHL.  374
1016001624  WKE---DEG--------------------------------------KGE---LWDRHKLYLHCT  354
1507700957  WQE---GKG--------------------------------------KGE---PWNVNRLILYCT  353
428306136   WQE---GTD--------------------------------------GE---W.IHL.  353
1019531886  WQE---GKG--------------------------------------KGE---LWDIHRLTLDCT  351

1002781834  LDIRFWIEGTQQZQQEKSKKFQTNRLRMKPELTFSIFFRSQTLETYLQVWLVITAYRLQ  433
```

SUPPLEMENTARY INFORMATION

```
 257060041  LDTRFWTEEGTQQVQQEKSKKFQTNRLRMKPELTFSIFFRSQTLETYLQVWLVITAYRLQ 430
1405885249  LDTRMLTTEGTQQVVEEKVTAITEILTKTK------------------------------ 409
 440680311  LDTRMLTTEGTQQVVEEKVTAITEILNKTK------------------------------ 409
1111222001  VDTRIWTAEGTNQVREEKAEEIAKTITKSK------------------------------ 407
1016058386  LDTRMWTIEGTQQVVSEKTTKITETLNQAK------------------------------ 402
1002781021  LDTRIWTQEGTEEVKQEKAEEIAKVLTSMN------------------------------ 381
1002400088  VETQIWTEAGTKQVQSEKMVEFQLNQLRMKPELTFPIFFRSQSLPTYFNLWKVITSYRIL 443
1002400093  IDNDIWTISGIQQVKQQKILQTEQKIANFH------------------------------ 394
1000428792  VDNRIWTAEGTEQVRQEKAEEITKFITKMK------------------------------ 415
1004799941  LDTRIWTTDGTEKVKQEKVDEFTQQLANME------------------------------ 411
1004798910  VDTRIWTAEGTNLVREEKAEEIAKAIAQTK------------------------------ 412
1002125473  VDNRIWTADGTEQVRQEKAEEISKLITKMK------------------------------ 406
1007062327  IDTRMLTAEGTQQVIEEKIADTQNKIAKAQ------------------------------ 407
1007109870  VDTRIWTEEGTNLVKEEKAEEIAKTIITQTK----------------------------- 406
 428305730  VDHRIWTQEGKEIVKAEKIPQTEKLINDFQ------------------------------ 439
1030021203  YDARILTAEGTEEVRQEKTDTTQKQVTKAE------------------------------ 444
  17230183  YDTRIWTAEGTEEARKEKINKIQRQVEQAE------------------------------ 440
1206817639  LDTSIWTAEGTEQIRINKINETDEAIAKAKTK---------------------------- 414
1300380544  IDTRIWTLEGTEQVRAEKIAQVEKTISKRE------------------------------ 411
 220908491  FDTRIMTAEGTLDVQKEKSEKALKNLAKAK------------------------------ 421
1085045623  VDTRIWTAEGTEQVRQEKATSIAKVIAGTK------------------------------ 407
1085056708  VDNRIWSAEGTEQVRQEKEEEITKFITKMN------------------------------ 406
1406824382  VDNRIWTEEGTEIVRQEKADEITKFITNMK------------------------------ 406
 428311644  VETRIWTAQGTEEVRAEKIAQTQAAIDRQK------------------------------ 397
1085030414  VDTRIWTAEGTELVRQEKTAEIEKVINRTK------------------------------ 406
1300485477  IETDTWTQEGTEEIRQKKASDCAKVIASTK------------------------------ 407
 428312908  LDTRIWTAEGTEQVRQEKADDIAKTLARMK------------------------------ 404
1016001624  FETRCLTAEGTSTIVEEKQKEVTKIIDLMK------------------------------ 384
1507700957  FDNLILTTEGTEVVRQEKAEEIANTLTKIK------------------------------ 383
 428306136  VDTRIWTAEGTEQVCQEKAEDIAKTLTRMK------------------------------ 383
1019531886  LETRCWTHEGTEQVKQEKADEIAGILTRMN------------------------------ 381


1002781834  SFLDK-----GNVAKAHQEFQKAIKRNESSLQ-KIT--S--SYN-RPHKTLYQGKSHIFV 482
 257060041  SFLDK-----GNVAKAHQEFQKAIKRNESSLQ-KIT--S--SYN-RPHKTLYQGKSHIFV 479
1405885249  ---QK-----DDLNDKQQAF---ITRQQSTLS-RIN--N--PFP-RPSKPNYQGKSSILI 452
 440680311  ---QK-----DDLNDKQQAF---ITRQQSTLA-RIN--N--PFP-RPSKPNYQGKSSILI 452
1111222001  ---EK-----GELNDRQLAH---IKRKQSTLD-RIN--N--PYP-RPNKPLYKGLSNILV 450
1016058386  ---RK-----DVLNDKQQAF---VTRQQSTLD-RIN--N--PFP-RPSKPNYQGQPSILV 445
1002781021  ---EK-----GDLTKNQQAF---IKRKQSTLD-KLE--N--PFP-RPSQPLYRGQSNILV 424
1002400088  KFLEK-----GDFTKAQKNFQDAIKRTESCLE-NLQ--S--SYL-TSQKSLYQGNPEIIM 492
1002400093  ---SKA--LEKELTPNCQQR---LKASQTSLN-LLKTFDINEFF-PSKCSLYQGSPDIIL 444
1000428792  ---DK-----SDLSETQLAF---IKRKESTLT-RIN--N--SFD-RPSKPLYQGSHILV 458
1004799941  ---QK-----ENLNQNCQNY---VKRLQSTLN-KLN--N--AYP-RHNHDLYQGKPSILV 454
1004798910  ---AK-----GKLNDKQQAH---IKRKNSSLA-RIN--N--LFP-RPSKPLYKGQSHILV 455
1002125473  ---EK-----SDLKDTQKAF---IQRKESTLN-RMN--N--SFE-RPSQPLYQGQSHILV 449
1007062327  ---EK---CEGELNPNCQAH---INRKKSTLA-RIN--T--PFP-RPSKPLYQGKSHIVV 452
1007109870  ---AK-----GDLNDKQQAH---IKRKNSSLA-RIN--N--PFP-RPSQPLYKGQSHILL 449
 428305730  ---QKERIQENGLTTGCQQR---LKASETSLR-LLQ--NCDHFA-ASKRISYRGQPNRIL 489
1030021203  ---GN-----ENIDSDEQKN---INRNISSLS-RLN--N--SFA-RPSKPIYRGQSNIIV 487
  17230183  ---ET-----ENLDEQCQKQ---LKKNKSSLS-RLN--N--SFN-RPIQPIYQGLSNMIV 483
1206817639  ---DK-----QELTENQLAY---LQRQQSTRN-KLN--N--SFPGRPSKPIYKGNSHILV 458
1300380544  ---QE-----VNLSKTQLER---LQAKHSERE-RLN--N--IFPNRPSKPSYRGKSHIAI 455
 220908491  ---PD----------------PRNHSTLD-RLK--N----LPVRPSRTPYQGNPEILV 453
1085045623  ---AK-----GNLNQKCQDF---ITKREKTLE-LLH--N--PFP-RPSKPLYQGKPSIIA 450
1085056708  ---AK-----SDLSETQQAF---IKRKESTLT-RIN--N--SFE-RPSQFLYQGSHILV 449
1406824382  ---KK-----SDLSDTQQAL---IQRKQSTLT-RIN--N--SFE-RPSQPLYQGSHILV 449
 428311644  ---AK-----GTQ---------SKLNTSLE-RLK--TFQGFS-RPSRASYKGNCSIVI 435
1085030414  ---AK-----NDLTETQQAF---IQRQKSTLA-RIK--G--HFD-RPSQSIYQGSHILV 449
1300485477  ---AK-----ENRSQNQDAF---IRRRERMLE-LLE--N--QFP-RPSYPLYQGQPSILA 450
 428312908  ---EK-----GDLNDKQQAF---IKRKNSTLA-RLN--N--PFP-RPSQPLYQGRSHIAV 447
1016001624  ---AK-----EELSDSQQAF---IRRKNSTLA-KLN--N--TFP-RPSKPVYQGKPNVHL 427
1507700957  ---EK-----GDLNQKCQAF---IRRKETSLS-RIN--N--PFP-RPSRPLYKGKSNILL 426
 428306136  ---KK-----GDLNDRQQAF---IRRQQSTLA-RLN--N--PYP-RPSQPLYQGQPHILV 426
1019531886  ---EK-----GDLTKNQKAF---VRRKQSTLN-RLE--K--PFP-RPSQPLYQGKSNILV 424
```

```
1002781834  GVAMGLEKPATVAVVDGTTGKALAYRSLKQLGN-------------------------  516
 257060041  GVAMGLEKPATVAVVDGTTGKALAYRSLKQLGN-------------------------  513
1405885249  GVSFGLEKPVTVAVVDVVKNQVLAYRSVKQLGE-------------------------  486
 440680311  GVSFGLEKPVTVAVVDVVKNKVLAYRSVKQLGE-------------------------  486
1111222001  GVSLGLEKTAIVAILDASTDKVVTYRSIKQLGE-------------------------  484
1016058386  GVSFGLEKPVTLAVVDVIKNEVLAYRTVKQLGE-------------------------  479
1002781021  GVSMGVDKPATIAVVNVVTQEVLTIEYIKQLGN-------------------------  458
1002400088  GVAMGLSQPATIAVVNVVTQEVLTYRSLKQLGK-------------------------  526
1002400093  GVSIGLENPATIAIINISTQEILTYRTTKQLSRTRKVRNKKPN--SNNSNQSLSSAYKQ 502
1000428792  GVSLGLEKPATIAVVDAIAGKVLTYRSLKQLGD-------------------------  492
1004799941  GVSLGLEKPATIAIVDSSTNIVLAYRSIKQLGD-------------------------  488
1004798910  GVSLGLEKPTILAVVDGSIGKVLTYRNIKQLGD-------------------------  489
1002125473  GVSLGLEKPATVAVVDAIAGKVLAYRSIRQLGD-------------------------  483
1007062327  GVSLGLKATAIIAVFDAMNNQVLAYRSTKQLGD-------------------------  486
1007109870  GVSLGLEKPATVAVVDGTTGKVLTYRNIKQLGD-------------------------  483
 428305730  GVSIGLHEPVIIMLVNTTTGKTLASKNTKQLDKKRRVRDQQPELPKYEERQHFTNIYKE 549
1030021203  GVSFHPVELAILVVVDIITKEKLICKTVKQLGD-------------------------  521
  17230183  GVSFHPVELATVAVVDTTTQKVIAYKTINELLDN-------------------------  517
1206817639  GVSLGLEKPVTVAAVDVVSNKVLAYRSVKQLGQ-------------------------  492
1300380544  GVSFSLENPATVAVVDVATKKVLTYRSFKQLGD-------------------------  489
 220908491  GLRVGLTNPITAAVVNGRTGEVLTYRTPSTLLGD-------------------------  487
1085045623  AVSFGLEKPATLAIVDIVTDKAITYRSIRQLGQ-------------------------  484
1085056708  GVSLGLEKPATVAVVDAIAGKVLAYRSIKQLGD-------------------------  483
1406824382  GVSLGLEKPATVAVVDAIANKVLAYRSIKQLGD-------------------------  483
 428311644  GVSFGRAKPATVAVVNVETGEVLAYRDVKQINK--------------PIKEGKTKKK  479
1085030414  GVSLGLDKPATVAVVDAIAEKVLAYRNTRQLGD-------------------------  483
1300485477  GVSYGLDKPATLAIVNIQTGKAITYRSIRQILGK-------------------------  484
 428312908  GISLGLGKPATAIIVDGTTGEAIAYRSIRQLLGD-------------------------  481
1016001624  GIAMGLEQPVTIAIVDIETDKVITYRNTKQLLRE-------------------------  461
1507700957  GVAIRLDKPATVAIVDGATDKAIAVLSTKQLLGK-------------------------  460
 428306136  GLAFGLDKPATAVVDGTTGKAITYRSLKQLLGD-------------------------  460
1019531886  GVSMELKKPATIAVIDGVTRKVLTYRNIKQLLGK-------------------------  458

1002781834  --NYHLFNRDGKDQNTSHDRHKSQKHFADNQFFESDLGQTIDCLLAKAIISQAPTGG      574
 257060041  --NYHLFNRDGKDQNTSHDRHKSQKHFADNQFFESDLGQTIDCLLAKAIISQAPTGG      571
1405885249  --NYNLLNRDRDDDQRLSHERHKAQKQNAPNSFFESDLGQTVDRLADAIIAIAKKYQAG    544
 440680311  --NYNLLNRDRDDDQRLSHERHKAQKQNAPNSFFESDLGQTVDRLADAIIAIAKKYQAG    544
1111222001  --NYKLLNRDRDDKGELSHDRKIAQTQAASNQYFESDLGQTIDRLAKEIVAIAQKYSAG    542
1016058386  --NYNLLNRDRDDDQRLSHERHKVDKRNAPNSFFESDLGQTVDRLADAIIAIAKTYQAG    537
1002781021  --NYPLIQRDRDDKQHQSHDRNVAQRKEAFNQFVDSDLGETIDRLTKAIVTLAKKYKAG    516
1002400088  --NYNLLNRDRDDKQKLSHDRHKAQKKDAFNQYFESDLGQTVDRRIAKAIVQVAKEYQAD    584
1002400093  ISNYELFLQYQQQKHHNQHQRHNAQINDANNNYGEANLGLYLNRLLAKAILELAQQYDVS    562
1000428792  --NYELLNRQRRQQRSLSHERHKAQKSFSPNQFGASELGQYVDRLLAKEIVAIAQTYKAG    550
1004799941  --NYKLLNRQRQQQQRNSHERHKAQKSNMPNKLSESDLGKYIDNLLAQAIIALAKNYDAG    546
1004798910  --NYRLLNRQRQQKHTLSHQRQVAQILASPNQLGESELGQYVDRLLAKEIVAITQTYKAG    547
1002125473  --NYELLNRQRRQQRSSSHERHKAQKSFSPNQFGTSELGQYVDRLLAKEIIAIAQTYKAG    541
1007062327  --NYKLLNRQQQQKQRLSQQRHKSQKQFASNSFGESELGQYVDRLLAKEIVAVAKNFGAG    544
1007109870  --NYKLLNRQRQQKHLLSHQRHIAQRIAAPNNFGDSELGEYIDRLLAKEIIAIAQTYQAG    541
 428305730  ISDYELFLFYRDQKQQHQIDRHKAQIKATPDHSKIANLGLVINRLLAKIIEFIQHQV      609
1030021203  --AFSLLSRRRRQQVHFRKEREKAQKKDSPCNMGESQLGEYVDKLLAKRIVEVTKEYKAS    579
  17230183  --AFHLLSRMRRQQIHFRKERKKAQKKDSPCNLGESKLGEYVDKLLAKRIVEVAKEYQAI    575
1206817639  --NYKLLNRQRQQQKHLAQKRHESQKKQAPNQFGESELGLYVDRLLAKSIINFAKTYQAS    550
1300380544  --NYNLANRLRQQKQRLSHERHKAQKQGAPNSFGDSELGQYVDRLLAKSIVAIAKTYQAS    547
 220908491  --RYDLFNRRHDDEQNILERHKYDKRGVTYQPYESILGKIVDRLLAKIIEIADTGG       545
1085045623  --NYKLFTKHRLKQQQCAHQRHQNQVESAENRISEGGLGEHLDSLLAKAILETAAEYGAS    542
1085056708  --NYELLNRQRRQQQYLSHERHKAQKSFSPNQFGTSELGQYVDRLLAKEIIAIAQTHKAG    541
1406824382  --NYELLNRQRRQQQYLSHERHKAQKNFSPNQFGASELGQHIDRLLAKAIVALARTYKAG    541
 428311644  KTQYEYLKRRQEDDRLNSHDRHNAQKNGAPCNFFESKQGEIVDRLLAKIVEIAQIS       539
1085030414  --NYKLLNRQRRQQRSLSHKRHKAQKRADTNQFGESELGQYVERLLAKEIVAISQNYRAG    541
1300485477  --NYKLLNRYRLNQQRNAHKRHNNQRKGGSSQLRESNQGQYLDRLIAHEIVALADEYDVS    542
 428312908  --NYKLFNRDRDEKQRQSHDRHKAQKNATSNQFFESILGEVVDRLLAKSIVTIAPTGG     539
1016001624  --DYRLLRRRRIEKQKLSHQNIKARKRFNFQQKGESNLGEYLDRLIAKAILTVADEYDVS    519
1507700957  --NYHLLNRKRQQHILSHQRNVAQRHHANNKFGESELGQYIDRLLAKAIIQLAKDYRVG    518
```

```
428306136  --NYELLNKQRKRKQQQSHDRHKAQSNGRSNFGDSDLGEZV  RLLAKALVTLAQSYQAG  518
1019531886 --NYPLLNRQQRDKQRQSHDRNIAQRKEAFNQFGDSPLGQH  RLLAKAIISIAQKYQAG  516


1002781834 SIVVPKLKDMRELIQSEIQAKAEAKIPGYVEGQAKYAKSYRVQVHQWSHGRLIDNITSQA  634
257060041  SIVVPKLKDTRELIQSEIQAKAEAKIPGYVGDAKYAKSYRVQIHQWSHIRLIDNITQI    631
1405885249 SIVLPKLRDMREQISSEIQSRAENQCPGYKEGQQKYAKEYRINVHRWSYGRLIESIKSQA  604
440680311  SIVLPKLRDTREQISSEIQSRAENQCPGYKGDQKYAKEYRINIHIWSYIRLIESIQI     604
1111222001 SIVLPKLSDMREQINSEIQAKALEKCPDCIEAQKKYAKQYRSVNQWSYGRLIENIKSQA   602
1016058386 SIVIPKLRDMREQISSEIQSRAEKKCPGYKEVQQKYAKEYRMSVHRWYGGRLIESIKSQA  597
1002781021 SIVVPKLEDMREIVQTEIQTKAEERIPNCIEAQKNYAKCYRVQVHQWSYGRLIDNIEAQA  576
1002400088 SIAVPKIRQMREIIQSEVQARAERKIQGYKEGQKKYAQQYRENVHQWSYGRLIESIHQAS  644
1002400093 LTILPSLKNKRELIESEIRAKAELKYPGCKEKQDSYAKDYRTNVHQWSYQQLIKCIESKA  622
1000428792 SIVLPKLGDIREIVQSEIQAIAEAKCPSSSEIQQKYAKQYRVNVHQWSYGRLIQSIQSKA  610
1004799941 SIVLPTMKNVRESIQSEIEARAVKRCPNYKEGQQQYAKQYRQSIHRWSYNRLMQFIQSQA  606
1004798910 SIVLPKLGDMREQVQSEIQAKAEQK-SDLIEVQQKYSKQYRVSVHQWSYGRLIASIRSSA  606
1002125473 NIVLPKLGDMREIVQSEIQAIAEAKCPGSVEVQQKYAKQYRVNVHKWSYGRLIQSIQSKE  601
1007062327 SIVLPKLGDMREIIQSEVQAKAEKKIPGFIELQKNYAKEYRSAHNWSYGRLIENIQSQA   604
1007109870 SIVLPNLGDMREQIQSEIKAKAEQK-SDLVEVQKKYAKQYPNSVHQWSYGRLITNIDSQS  600
428305730  TIILPDLKNKRESIEBEAKALAKLKIPKDKSRDQFTNILSEYSQWSYKQLSDCINK      669
1030021203 CIVLPRLKDTREIRTSVIQAKAEAKFPGDVNAQKIYVKEYNRQIHNWSYGRLQESIKSKA  639
17230183   CIVLPKLKDMKEIRTSVIQAKAETKFPGNVNAQKLYVKEYNRQVHNWSYNRLQESIKSKA  635
1206817639 SIALPKLRDMREIIQSEIQAKAESKIPGYKEGQEKYAKEYRMSVHRWSYGRLIGNIQAQA  610
1300380544 SIVLPKLRYMREIIHNEVQAKAEKKIPGYKEGQKQYAKQYRISVHQWSYNRLSQILESQA  607
220908491  SIVIPCLTHIREIIASEIAARAEEKCPGSVLADNFAKFIRKIHIWSYNRLISAIK       605
1085045623 SIVLPELGNIREIIHAEIQAKAERKIPGLKEKQDEYAAKFRASVHRWSYGRLAQKVTTKA  602
1085056708 SIVLPKLGDMREIVQSEIQAIAEEKFPGYVEGQQKYAKQYRVNVHGWSHGRLIQSIQSKA  601
1406824382 SIVLPKLGDMREVVQSEIQAIAEQKFPGYIEGQQKYAKQYRVNVHRWSYGRLIQSIQSKA  601
428311644  SIVLPDLRNIEEAAESEVRARAEQKTPGNQKLDDSFAKDIRISIHIWSYIRLAQCILK    599
1085030414 SIVLPKLGDMQEILTSEIQARAEAKCPNYVEGQQKYAKQYRISIHKWSYGRLMQNIQSQA  601
1300485477 SLALPDLGDIREIVQSEVQARAEQKILGSIEQQRKYARQYRASVHRWRYAQLTQFIQSQA  602
428312908  SIVLPKLGDMRELVHSEIQTRAEQKIPSYVEGQQKYAKQYRVNVHQWSYGRLIENIDVQA  599
1016001624 TILIPRLRDMRSITEAEIQLRAEKKIPEYKEGQKKYAQDYRVQVHQWSYGRLIENVKLIC  579
1507700957 SIVVPYMEDTREIIQAEVQARAEAKIPGCIEKQKEYAKKYRTNIHKWSYGRLIDLIKAQA  578
428306136  SIVLPKLGDIRELIQSEIQAKAEQKIPGYIAGDKYAKDKISIHIWSYIRLIDNIIQ      578
1019531886 SIVVPKLEDIREATQSEIQAKAEAKIPNCIEAQAEYAKKYRMQVHEWSYGRLIDNIQAQA  576


1002781834 SKFNITVEEGEQPHQGNPQDKAKLLAIAAYHSRLCA-------              670
257060041  SKFNITVEEGEQPHQGNPQDKAKLLAIAAYHSRLCA-------              667
1405885249 AQAGIAIETGTQPIRASPQEKARDLALFAYQERQAALI-----              642
440680311  AQAGIAIETGKQSIRESPQEKARDLAVFTDERQAALI-----              642
1111222001 IKTGIVIEESKQPIRGSSQDKAKELATTAYKSRKKS-------              638
1016058386 AKAGIFTEIGTQPIRGSPQEKARDLAVFAYQERQAALI-----              635
1002781021 SKLGIVLEISQQPYKGTPHDKAIALALNAYQSRLSA-------              612
1002400088 AKFGIRVEIASQSYQGSFQEQAQNLAIAAYTNRLEAVG-----              682
1002400093 AQIGIDTATGKQMNLETSQDQARNLVLNFCQKFSPTQV-----              660
1000428792 AQIGIVIEEGKQPIRGSPQDKAKELAYSLRLARRS-----                648
1004799941 VKANISIEQGPQPIRGSSQEKARDLAIAAYYLRQNKS------              643
1004798910 AKVGIVIEESKQPIRGSPQEKARELAIAAYNSRRRT-------              642
1002125473 SQAGIVIEEGKQPVRGSPHEQAKELALSAYHDRLARRS-----              639
1007062327 TKEGIEIETGKQPTRGIPQEQARDLALFAYQCRIA--------              639
1007109870 KKAGIVIEEGKQQIRASPLEKAKELAINAYQSRKA--------              635
428305730  SQSGIAIEIQQISQGNPYQKARNLFITTFTKNSGNKCSAKIE              711
1030021203 AEFKISIEFGMQPSYDNLQEQAINLALSAYQCRINTIGR----              678
17230183   AELKISIEFGIQLSYDTLQAQARDLALSAYQCRIHTIDR----              674
1206817639 AQAGILIETSSGQIRGSPQEQAKHLAISAYIERQTILNK----              649
1300380544 TKAGISIERGSQVIQGSSQEQARDLALFAYNERQLSLG-----              645
220908491  EQLGIIVESGFDPYEGDSYQQAKDLAIAVFHSRDLALK-----              643
1085045623 SLHGLETESTRQSLQGTPQEKARNLAISAYESRKVAQRA----              641
1085056708 AQIGIVIEEGKQPIRGSPQDKAKELALSAYNLRLARRS-----              639
1406824382 AQTGIVIEEGKQPIRGSPHDKAKELALSAYNLRLTRRS-----              639
428311644  GKAGIATEKVHDPHGDTPQEKARDLFLAAYANKVSVS-----              637
1085030414 AQAEIVVEEGKQLIRGSPQEMAKELAIAAYQSRQPQ-------              637
1300485477 AQVGISIEITKQPLSGTPQEKARNLAIAAYQSRK---------              636
428312908  AKIGISIEQGQQPVRGSPQEKAKEMAIAAYHSRLNP-------              635
```

```
1016001624  EKVGIVVVEAKQPKQGTLTEKALQIVLSATEKNLKKK------           616
1507700957  AKAGIVIEESKQSIRGDPKKQAKEIAVCAYRDRIVPF------           615
 428306136  AKISIVIEEGQQPIRGSPQEKAKE AISAYDDRTKS-------           614
1019531886  SKLGIFIEESQQPLQGTPLQKAAELAFKAYRSRLSA-------           612
```

HHpred output for sequence 1002781834

>COG0675 InsQ Transposase [Mobilome: prophages, transposons].
  Probab=98.99 E-value=3.4e-09 Score=106.41 Aligned_cols=142 Identities=20% Similarity=0.212 Sum_probs=0.0

```
                                      RuvC-I

Q ss_pred         CCCCCCCCcCCCCceEEEEcCCCCceEEEEEeCcccEEEecCHHHHhccHHHHHHHHHHHHHHHHHhHHHHhcccc
  Q Tue Jul 12 23:  465 SYNRPHKTLYQGKSH1PVGVAMGLEKPATVAVVDGTTGKAIAYRSLKQLLGNNYHLFNRQGKQKQNTSHQRHKSQKHFAD  544 (670)
  Q Consensus       465 ~f~RPskp~YqG~p~iivGVdlGL~~PATvAVvD~~t~Kvl~yrs~kQLLg~ny~Ll~R~R~qqq~lsh~Rhk~qk~~~~   544 (670)
                        ++.|..+. .-.+..+||||+|+...||++     .+........-........-..+.....+..++.+.++..+...
  T Consensus       162 ~~~~~~~~~~~~~~~~giDlg~~~~~~~-------~~~~~~~~~~~~~~~~~~~~~k~ls~~~~~~~~~~~~~~~~~        235 (364)
  T COG0675         162 LVEVEEEEP-VPKPGKWVGIDLGLKNLATVS-----DNGTVPIPRDGRKLLKRLKKAQKRLSRKKSRSKRRKKAKLKLAR  235 (364)
  T ss_pred         EEEecccc-cCCCCeEEEecCCceEEEEe-----CCccccCchhHHHHHHHHHHHHHHHHHhhhhhhhHHHHHHHHHHH
```

```
                                      RuvC-II

Q ss_pred         cchhHHHHhhhHHHHHHHHHHHHHHHcCCCEEEecccccHHHHHHHHHHHHHHhcCCChHHHHHHHHHHHHHHhHcCCCHH
  Q Tue_Jul_12_23:  545 NQFGESQLGQYIDCLIAKAIISVAQTYCAGSIVVPKLKDMRELIQSEIQAKAEAKIPGYVEGQAKYAKSYRVQVHQWSHG   624 (670)
  Q Consensus       545 ~q~~Es~LG~yvdrllAk~IV~lAq~yqAgsIVLpkL~diRe~i~SeiqArAe~K~pg~ke~QkkYaK~yR~~vH~WSYg   624 (670)
                        ......+.-... |.+|+.||+.|..++++.|++.+.+       ++.+|.|+|+
  T Consensus       236 ~~~~~~~~r~~~~~k~~~~~~v~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~i~~~~~          283 (364)
  T COG0675         236 LRERIRNRRKDF-HKLAKKLVELADLVGVETLVVEDLVK----------------------------RRSISDWAFG   283 (364)
  T ss_pred         HHHHHHHHHHH-HHHHHHHHHHhhheeeeeeeeccccc----------------------------ccchhhhhHH
```

```
  Q ss_pred         HHHHHHHHHHHHcCceEEEe
  Q Tue_Jul_12_23:  625 RLIDNITSQASKFNITVEEG   644 (670)
  Q Consensus       625 RL~q~I~~KA~GI~Ve~~    644 (670)
                        +|++.|+|||+.-||.|++.
  T Consensus       284 ~~~~~l~yka~~~~~~v~~~   303 (364)
  T COG0675         284 ELRRQLEYKAEWGGIVVKVV   303 (364)
  T ss_pred         HHHHHHHHHHhCCeeEEEc
```

Supplementary information S4 (figure). Strength of purifying selection for Class 2 effector protein families.

The three quartiles of the distribution of the *dN/dS* ratio, estimated for sequence pairs with 0.0002<dN<1.0 and 0.0002<*dS*<1.0 (see supplementary information S1 (box) for details) are given. The background color highlights the scale from low (blue) to high (red) values.

system	gene	no. of sequence pairs	dN/dS		
			1st quartile	median	3rd quartile
II-A	cas9	2239	0.12	0.19	0.25
II-B	cas9	67	0.21	0.32	0.64
II-C	cas9	2756	0.08	0.12	0.19
V-A	cas12a	48	0.04	0.13	0.21
V-B	cas12b	4	0.11	0.17	0.25
V-U1	c2c4	4	0.14	0.22	0.44
V-U2	c2c8	3	0.08	0.25	0.30
V-U3	c2c10	14	0.03	0.04	0.12
V-U4	c2c9	11	0.07	0.15	0.36
V-U5	c2c5	16	0.15	0.16	0.19
VI-A	cas13a1	8	0.27	0.39	0.41
VI-B	cas13b	515	0.34	0.39	0.46
VI-C	cas13a2	3	0.28	0.28	0.31

Supplementary information S5 (figure). UPGMA dendrogram of protein family profile similarity

Profiles were built for distinct subfamilies of type V systems (red) and TnpB family (blue). The profiles correspond to the clusters, information for which is provided in the supplementary information S2 (box, part h). The profile dendrogram was built on the basis of similarity score matrix obtained using the HHalign program (see details in the Supplementary Methods). The dotted line indicates the arbitrary sumilarity cutoff ~2 (in distance units shown by the scale bar below the tree) which, empirically, corresponds to the limit of confident identification of relationships between groups of sequences (i.e. the groups to the right of the line are considered to be confidently identified.)
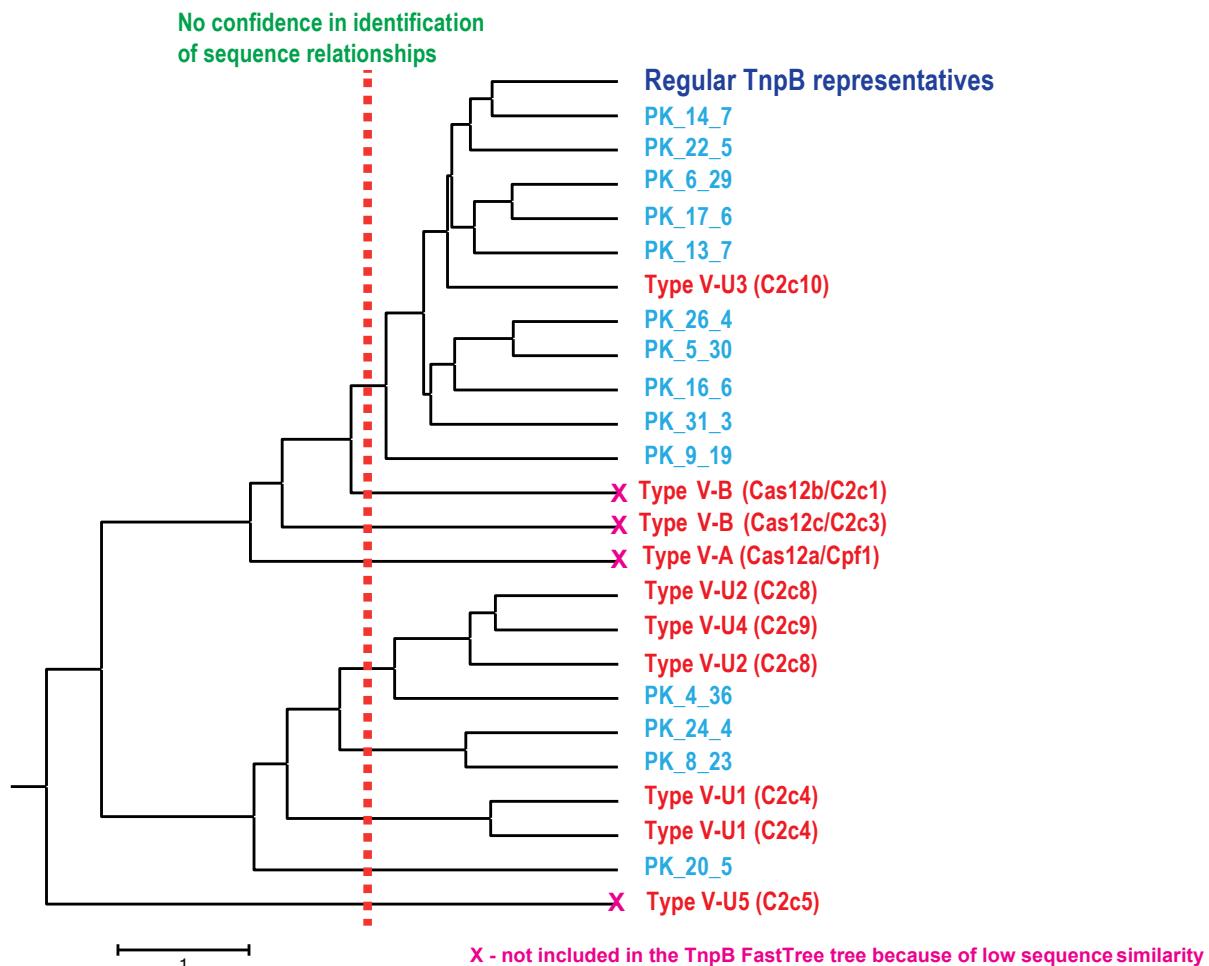
Supplementary information S6 (figure).   Membrane proteins associated with Cas13b genes

Multiple alignment of representatives of predicted membrane proteins associated with Cas13b genes. Numeric sequence IDs are from Genbank. Secondary structure was predicted using the Jpred4 server. Transmembrane domain were predicted using the TMHMM server.

Variant 1.

TMHMM output

```
# WEBSEQUENCE Length: 181
# WEBSEQUENCE Number of predicted TMHs:  1 (shown in cyan)
# WEBSEQUENCE Exp number of AAs in TMHs: 21.52507

# WEBSEQUENCE Exp number, first 60 AAs:  21.51976
# WEBSEQUENCE Total prob of N-in        0.65484
# WEBSEQUENCE POSSIBLE N-term signal sequence
WEBSEQUENCE        TMHMM2.0    inside      1     6
WEBSEQUENCE        TMHMM2.0    TMhelix     7    29
WEBSEQUENCE        TMHMM2.0    outside    30   181
```

HEPN domain motifs (E…RxxxxH) are shown by reverse shading.

```
                                                                       α1
739007499|Bacte|Prevotella_aurantiaca           MDCTLISKEVATALLSTISSFIPIVIAAYLTYRYAIKKLHKESIENIERAKYEAILKAHQSIYKLLRFTT
339613715|Bacte|Prevotella_pallens_ATCC_700821  MDCTSILQDMIIALITSLFSSATIIFAGYRTYRNAIKKLREESIENIKRAHYEAILKAHQSIYKLLRFTT
496453857|Bacte|Prevotella_saccharolytica       -----MECNDCSQLITTLVSLLTAALSGFLSYYFAVRKYRKESQENVERWKYEAIWHAHQSFYKLLRFMT
916920636|Bacte|Prevotella_falsenii             ---------MIIALIPALINSAAIIVVVYLTYFYAVKKLREESFENIERAKYEAILKAHQSVYKLLRFIT
490473138|Bacte|Prevotella_buccae               --------MDYMELAKEAFSIICTFIAAYVAYYYAIKQLHQKSVENIEYAKYQAVLQAHKSLYRLLRFTT
771514465|Bacte|Prevotella_intermedia           MDCTLISKEVATAMFSTISSLIPIVIAAYLTYRYAIKKLRKESFENIERAKYEAVLKAHQSIYKLLRFIT
640560407|Bacte|Prevotella_pleuritidis          MDCTLISKEVATALFSTLSSLVPIVIAAYLTYRYAIKKLHKESFESIECAKYEAILKAHQSIYKLLRFIT
696216473|Bacte|Porphyromonas_sp_COT-052_OH4946 ---MMTMTICWTPICVQLIELSGIFIAAYLAYRYAVRKLSKESIENIERCKYQAVLEAHRSFYKLLRFTT
333804400|Bacte|Porphyromonas_gingivalis_TDC60  ---MMTMTICWTPICVQLIKLSGIFIAAYLAYRYAVRKLSLESIENIERCKYQAVLEAHRSFYKLLRFTT
700239026|Bacte|Porphyromonas_gulae             ---MMTMTICWTPICVQLIELSGIFIAAYLAYRYAVRKLSKESIENIERCKYQAVLEAHRSFYKLLRFTT
491058122|Bacte|Riemerella_anatipestifer        ------MYWILEVLKIVTPTIAVIVSAIVSTIVLSRKIRQELKGNIERQKYEDILHAHKQMYRLLAYMT
CONSENSUS                                        ----------------------I-----Y--A--K---E---NIE--KYE-IL-AH-S-YKLLRF-T
Jnet 491058122                                   :-------HHHHHHHHHHHHHHHHHHHHHH----HHHHHHHHHHHHHHHHHHHHHHHHH-

                                                         α2                 α2.1
739007499|Bacte|Prevotella_aurantiaca           DTENDDCILLWEQSKGSK-EKVYYFRQANIRKFIKELTEEIYNKGNGIYLSKEVILLVFKYRTLVHKLLL
339613715|Bacte|Prevotella_pallens_ATCC_700821  DTENDDCILVWEQPKDGG-EKTYYFRQANIRQFFKELTEEVYNKGNGIYLSKEVMSLIFKYRTLVHKLLL
496453857|Bacte|Prevotella_saccharolytica       DNENADSILVFRKQEGSK-QPQPIFRKDKAREFLAELTDEFYKKGNGLFLSKEVADSLFKYRHIVFGLLQ
916920636|Bacte|Prevotella_falsenii             DTENDDCILVWEEPKGGG-EKAYYFRQANIRKFIRELTEEIYNKGNGIFLSKKTMSLIFEYRTLVYGLLL
490473138|Bacte|Prevotella_buccae               DTENEDSILIWERPKDGGLEATYYFRKGNIRKFIKELSKEIYNEGCGIFMSKEALSLISEYRNIVYGFML
771514465|Bacte|Prevotella_intermedia           DTENDDCIMVWEQPKGGG-EKVYYFRQANIRKFIKELTEEIYNKGNGIFLSKKVMPLIFEYRSLVYGLLL
640560407|Bacte|Prevotella_pleuritidis          DTENDDCILVWEQPKGGG-EKTYYFRQANIRKFIKELTEEIYNKGNGIYLSKEVMSLIFKYRTLVHKLLL
696216473|Bacte|Porphyromonas_sp_COT-052_OH4946 DTENADSILVWQKAKGSG-AKTYYFRPACIRGFLSELTDEFYKNGNGVFLSKEIISRIFEYRSIVYGLLL
333804400|Bacte|Porphyromonas_gingivalis_TDC60  DTENADSILVWQKAKGGG-AKTYYFRPACIRGFLSELTDEFYKNGNGIFLSKEIISRIFEYRSIVYGLLL
700239026|Bacte|Porphyromonas_gulae             DTENADSILVWQKAKGGG-AKTYYFRPACIRGFLSELTDEFYKNGNGIFLSKEIISRIFEYRSIVYGLLL
491058122|Bacte|Riemerella_anatipestifer        DQDNPKNLLKWEVPKGQK-DKIHYINRANAQAFLRELPELFYGEGCGLFLSEEVTKKFFEYRSIVHKLLL
CONSENSUS                                        D-EN-D-IL-W---KG-------YFR----R-F--EL----Y--G-GIFLSKEV----F-YR-IV--LLL
Jnet 491058122                                   -------EEEEE--------------HHHHHHHHHHHHH----HHHHHHHHHHHHHHHHHHHHH

                                                         α3            α4
739007499|Bacte|Prevotella_aurantiaca           AEKNNPDEKIMISKQEIAKSMIEIHQSLSIQIRNDINLKKRDLCLNSENT------------
339613715|Bacte|Prevotella_pallens_ATCC_700821  AEKNNPDEKIMIHNREVVKEMIEINQSLSIQIRKDIDLQQRNLYNHNKKGKKYWWKRKKQNI
496453857|Bacte|Prevotella_saccharolytica       VFSNYDENEVQLKNTQAVENMKKNFQQLSPEIRKSLHLHRRDLLFS---------------
916920636|Bacte|Prevotella_falsenii             TAKDNPAEAIEITNKKLAKRMIEIHQSLSIQIRKDRNLQQ--------------------
490473138|Bacte|Prevotella_buccae               SAQNNPQETIRITNRESVERIKKIHQNLSIEIRQAINLKKRDLRF----------------
771514465|Bacte|Prevotella_intermedia           TAKDKPDETIEIKNEKLAKRMIEIHQSLSIQIRKDLQRDLQFDS---------------
640560407|Bacte|Prevotella_pleuritidis          AKKNNPDEKIMIDKRELAKRMIEIHQNLSIKIREAINLQQRDLHFK---------------
696216473|Bacte|Porphyromonas_sp_COT-052_OH4946 SERDSSDERIVMNKPETAERMIRIHQELTQTVREAIALKGRTLNF----------------
333804400|Bacte|Porphyromonas_gingivalis_TDC60  SERQNSDERIVMNKPETAERMISIHQELTQTVREAIALKKRTLNF----------------
700239026|Bacte|Porphyromonas_gulae             SERDSSDERIVMNKPETAERMIRIHQELTQTVREAIALKGRTLNF----------------
491058122|Bacte|Riemerella_anatipestifer        AEQNSTEAEFRLKNEEAATRMKVLHHKLSQSIRQCLKIEQRDLKAL---------------
CONSENSUS                                        ---------I-----E----M---Q-LS--IR----L--R-L------------------
Jnet 491058122                                   H-------EEEE--HHHHHHHHHHH--HHHHHHHHHHHH------------------
```

Variant 2.

TMHMM output

```
# WEBSEQUENCE Length: 219
# WEBSEQUENCE Number of predicted TMHs:  4 (shown in cyan)
# WEBSEQUENCE Exp number of AAs in TMHs: 88.84536
# WEBSEQUENCE Exp number, first 60 AAs: 37.17311
# WEBSEQUENCE Total prob of N-in:        0.99188
# WEBSEQUENCE POSSIBLE N-term signal sequence
WEBSEQUENCE          TMHMM2.0  inside        1     8
WEBSEQUENCE          TMHMM2.0  TMhelix       9    31
WEBSEQUENCE          TMHMM2.0  outside      32    45
WEBSEQUENCE          TMHMM2.0  TMhelix      46    68
WEBSEQUENCE          TMHMM2.0  inside       69   140
WEBSEQUENCE          TMHMM2.0  TMhelix     141   163
WEBSEQUENCE          TMHMM2.0  outside     164   177
WEBSEQUENCE          TMHMM2.0  TMhelix     178   200
WEBSEQUENCE          TMHMM2.0  inside      201   219
```

```
501433496|Bacte|Porphyromonas gingivalis          -----MEKFSLYDFLAIILPGIAFIVVFRIIFSSLHLSLPVDIPLGLESTIVYALI-CGA
545407573|Bacte|Bacteroides pyogenes              -----MNKFSFYDFLSILLPGVIFLVAIRVAQLFWRFNTGLYFPQGWEFSLVYSLVIGAS
818455803|Bacte|Wolfebacteria bacterium GW2011    MDTKTFPKTSLFDQISFLMVGAIATIVIVFNMVYF-FTMPIPT-----VDLGSALVWLVI
754594446|Bacte|Capnocytophaga canimorsus         -----MDRLSIYELLSFVVPGVIMIELINFSAEYV-FGKDRLI---TAGNLSDGLLFFAI
405580122|Bacte|Bergeyella zoohelcum ATCC 43767   -----MDRLSIYELLSFVVPGVIMIELINFSAEYV-FGKAPLI---TAGNLSDGLLFFAI
754599666|Bacte|Capnocytophaga cynodegmi          -----MERLSIYELLSFVVLGVIMIELMNFSAEYV-FGKAPLI---TAGNLSDSLLFFAI
493307991|Bacte|Myroides odoratimimus             -----MDKLSLYELFSFVIPGGIALHLLNWCSVNV-LSTGTLF---NLSDLSNSLIALVF
503850671|Bacte|Flavobacterium branchiophilum     -----MEKLSLYEILSFVVPGYVLLELSKWFGSFF-I--DANF--NLDESLNSLLQFII
CONSENSUS 0.8                                      -----------YE--S---PG----------------------------L-----

501433496|Bacte|Porphyromonas gingivalis          VLYVLSFSL----VKLFPRL---FGL-YRHVADLYQKMK-ALHPIMND-TLNRQAEQWGL
545407573|Bacte|Bacteroides pyogenes              L-YVLGFSVKKNYSGFFRC---LGL-YEHVTILYYRFE-TLHPFMNG-ALNKYAEEWNG
818455803|Bacte|Wolfebacteria bacterium GW2011    S-YFLGHVIHGISNAIRDIPIIPLLD-WESKLAYLPHEALLLKE-----AEAYFEQPYQE
754594446|Bacte|Capnocytophaga canimorsus         A-LFIGCLIHIITFRLIKRK---WYKKLAYKPINEIENNAYIKQ-----IFTTLKEEYSK
405580122|Bacte|Bergeyella zoohelcum ATCC 43767   A-LFIGCLIHIITFWLIKRK---WYKKLACTPMDKMKNDEYITE-----ILPDLKEIYRT
754599666|Bacte|Capnocytophaga cynodegmi          A-LFIGCLIHIITFRLIKRK---WYQKLACTPMDKMKNDEYITE-----ILPHLKEIYRA
493307991|Bacte|Myroides odoratimimus             A-LLIGVTLHIITFNILLK----CGS--YRQIIYKSVQEIKLDDYIQQ-VIPFLNQEYFH
503850671|Bacte|Flavobacterium branchiophilum     A-LFLGIIIHVLTFKIFMK----IKW--YKDLIYKSVQKISVENDFIQKAIPFLNEHYIA
CONSENSUS 0.8                                      -----G-----------------------------------------------------

501433496|Bacte|Porphyromonas gingivalis          GKIYLSEEEF--CQSEEKEKIRMLQSDFYDRMWYRLDFRGKLGNAKSFQCYYFFFRHSF-
545407573|Bacte|Bacteroides pyogenes              TKPYCTVEQYDAMDASAQKEIEDAQDIFYDHMYYRLDCKGKLEGAKAFQSYYLCFLHSFL
818455803|Bacte|Wolfebacteria bacterium GW2011    KKP--------------------DHLWSLCYIFATAKDTTSQVETFSAYYNLYRG---
754594446|Bacte|Capnocytophaga canimorsus         IHN------------INSIDEISKLNIFESGYYYLEAQGKISQAKNFQSIYFLFRNIVT
405580122|Bacte|Bergeyella zoohelcum ATCC 43767   NKG------------ISVTENISNGSVFDTAYYYLEAQGKISQAKNFQSIYFLFRNIVT
754599666|Bacte|Capnocytophaga cynodegmi          NKG------------ISATENISNGSVFDTAYYYLEAQGKISQAKNFQSIYFLFRNIVT
493307991|Bacte|Myroides odoratimimus             NKK---------HEVAANTNNAVPAENLFDYAYYYLEVNGKNAQAKNFQSLYFFFRNMFT
503850671|Bacte|Flavobacterium branchiophilum     HR----------KHTEKPVNENEAENNLFDYAYYYLEINDKIAAAKSFQSLYFLFRNFFT
CONSENSUS 0.8                                      -----------------------------Y--L----K---------Y--------

501433496|Bacte|Porphyromonas gingivalis          ---WGLVLISLILLSYKLLAY--IPACDMEDIGWREYSDIAVPIMILSALFVFL--AQWF
545407573|Bacte|Bacteroides pyogenes              GLLIFGVYLLICILLSYFMDV--LLADT--------WQVAFLFLMNLCVMYLFMRLARWF
818455803|Bacte|Wolfebacteria bacterium GW2011    ---WFTTLLIETLFLSYFLVF--SYTHA--------MLAVWILSI-AIAVLMYR-RAKRF
754594446|Bacte|Capnocytophaga canimorsus         LSLFVLPVSVIFLLISFFMKD-CTLSGK--------IITIIIGTL-VIGGLSSV-IAQWF
405580122|Bacte|Bergeyella zoohelcum ATCC 43767   LSLFVLPVSVIFLLASFFMND-CSLSEK--------IITIIIGTF-VLGGLSSV-IAQWF
754599666|Bacte|Capnocytophaga cynodegmi          LSLFVLPVSVIFLLASFFMND-CKLSGK--------IITIVIGTL-VIGGISSV-IAQWF
493307991|Bacte|Myroides odoratimimus             LGIVSIVILIIALVYSTITSV-GKDVLS--------EIVLKIAFFAVIIGIAVP-VANWL
503850671|Bacte|Flavobacterium branchiophilum     ITLFFLPISVITILILYTQNYCCCLQRT--------ALTAFILLI-IIAFILVY-IGRWL
CONSENSUS 0.8                                      --------------------------------------------------------

501433496|Bacte|Porphyromonas     gingivalis      RIKMVEKMYWTFYISLIEQENSNL--
545407573|Bacte|Bacteroides pyogenes              RQRMVLKMYWAFYESLIE--------
818455803|Bacte|Wolfebacteria bacterium GW2011    WQYLHDKVFGIFLITLKFPKKETV--
754594446|Bacte|Capnocytophaga    canimorsus      RVKMTDRIFGLYYAELTHNKK-----
405580122|Bacte|Bergeyella zoohelcum ATCC 43767   RVKMTDRIFGLYYAELTHNKK-----
754599666|Bacte|Capnocytophaga    cynodegmi       RVKMTDRIFGLYYAELTHHKK-----
493307991|Bacte|Myroides odoratimimus             RKKMIITVFGCYYADRVHQTNK----
503850671|Bacte|Flavobacterium  branchiophilum    REKTVEKVIWSYYVAITHEENNNKTK
CONSENSUS 0.8                                      -----------YY-------------
```