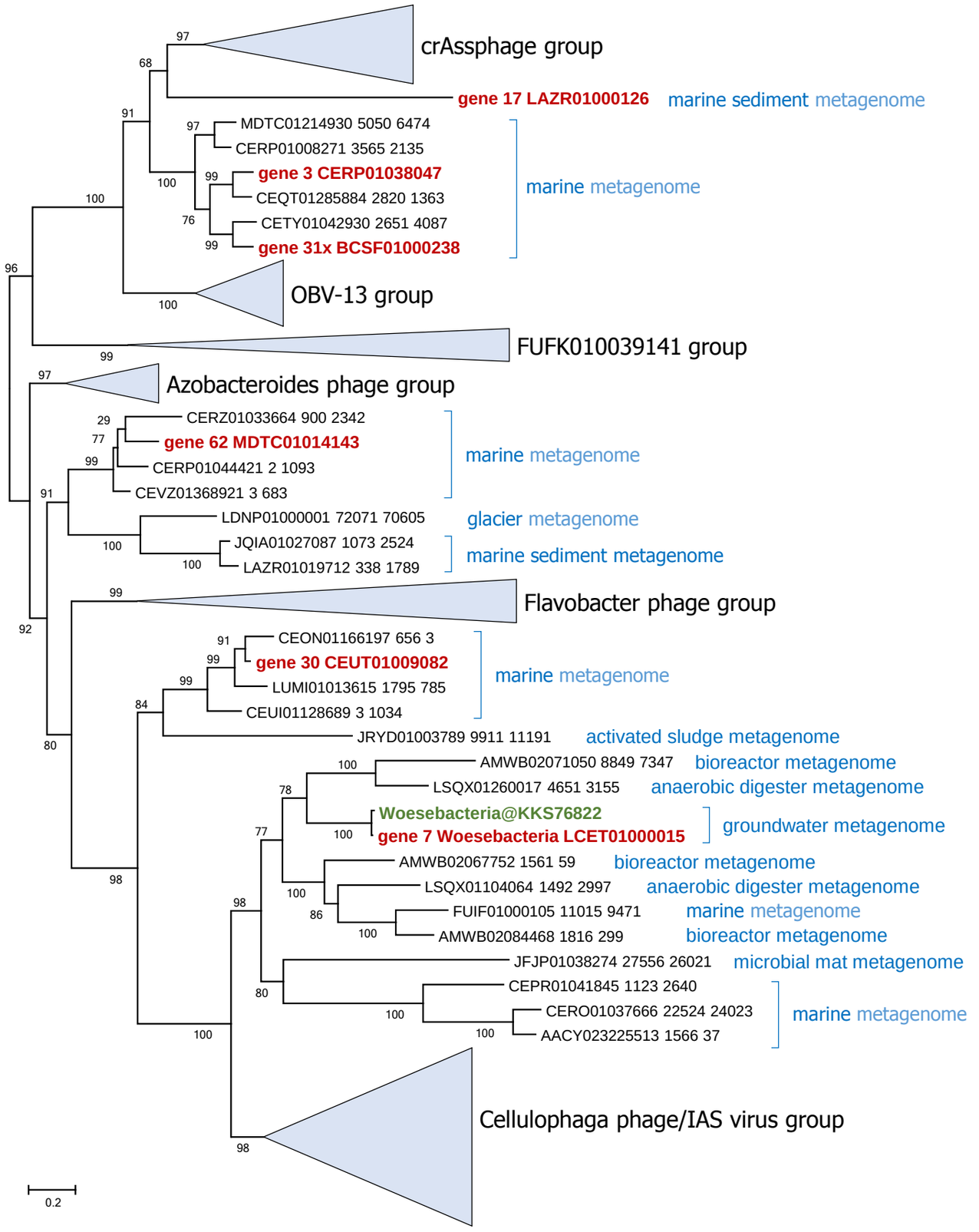
Supplementary Figures and Notes

Discovery of an expansive bacteriophage family that includes the most abundant viruses from the human gut

Natalya Yutin[1], Kira S. Makarova[1], Ayal B. Gussow[1], Mart Krupovic[2], Anca Segall[1,3], Robert A. Edwards[3], Eugene V. Koonin[1,*]

1, National Center for Biotechnology Information, National Library of Medicine, Bethesda, Maryland 20894, USA;
2, Institut Pasteur, Unité Biologie Moléculaire du Gène chez les Extrêmophiles, 25 rue du Docteur Roux, 75015 Paris, France;
3, Viral Information Institute, Department of Biology, San Diego State University, San Diego, California 92182, USA

Supplementary Figure 1

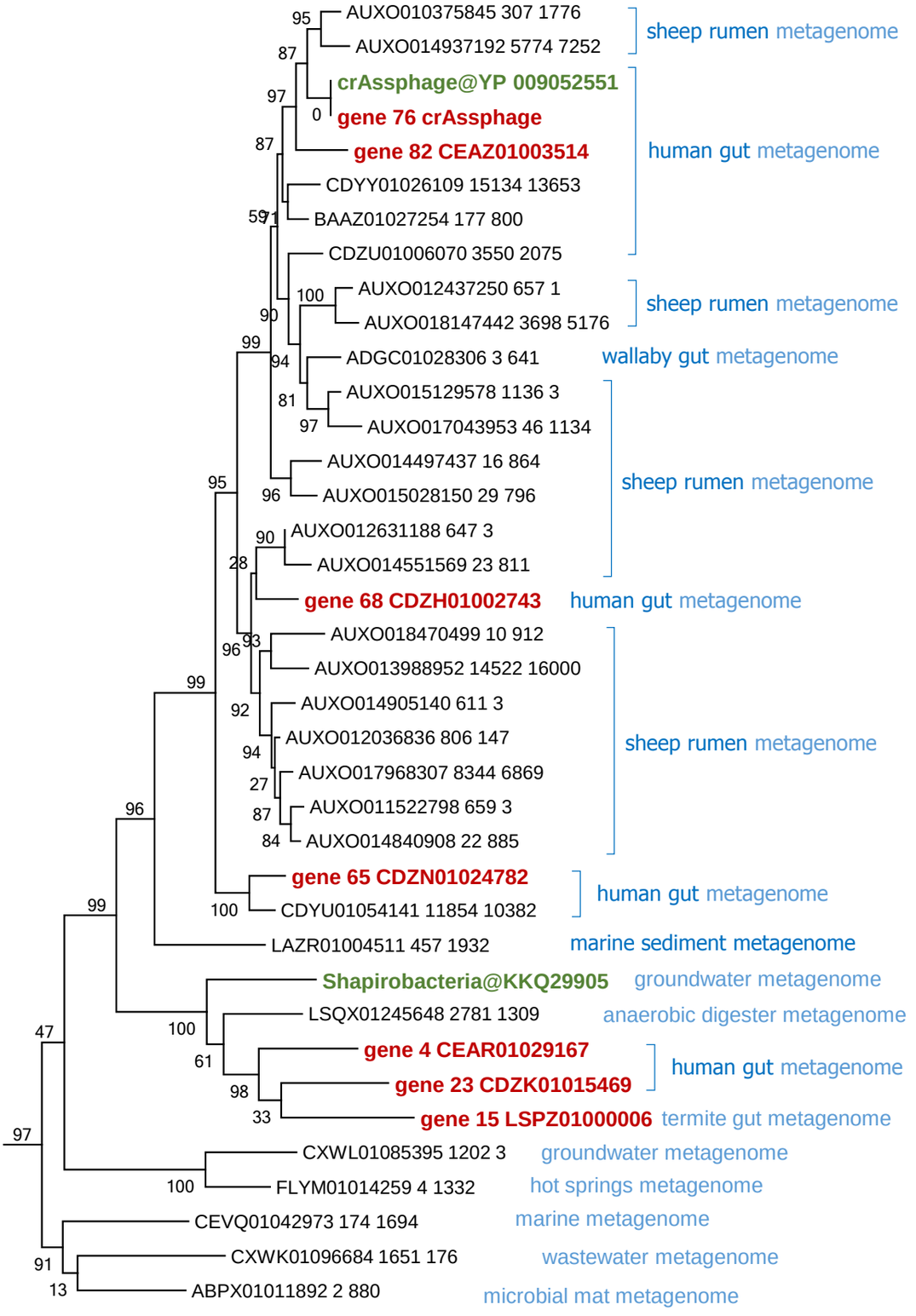


Phylogenetic tree of the MCP for all identified members of the crAss-like family.

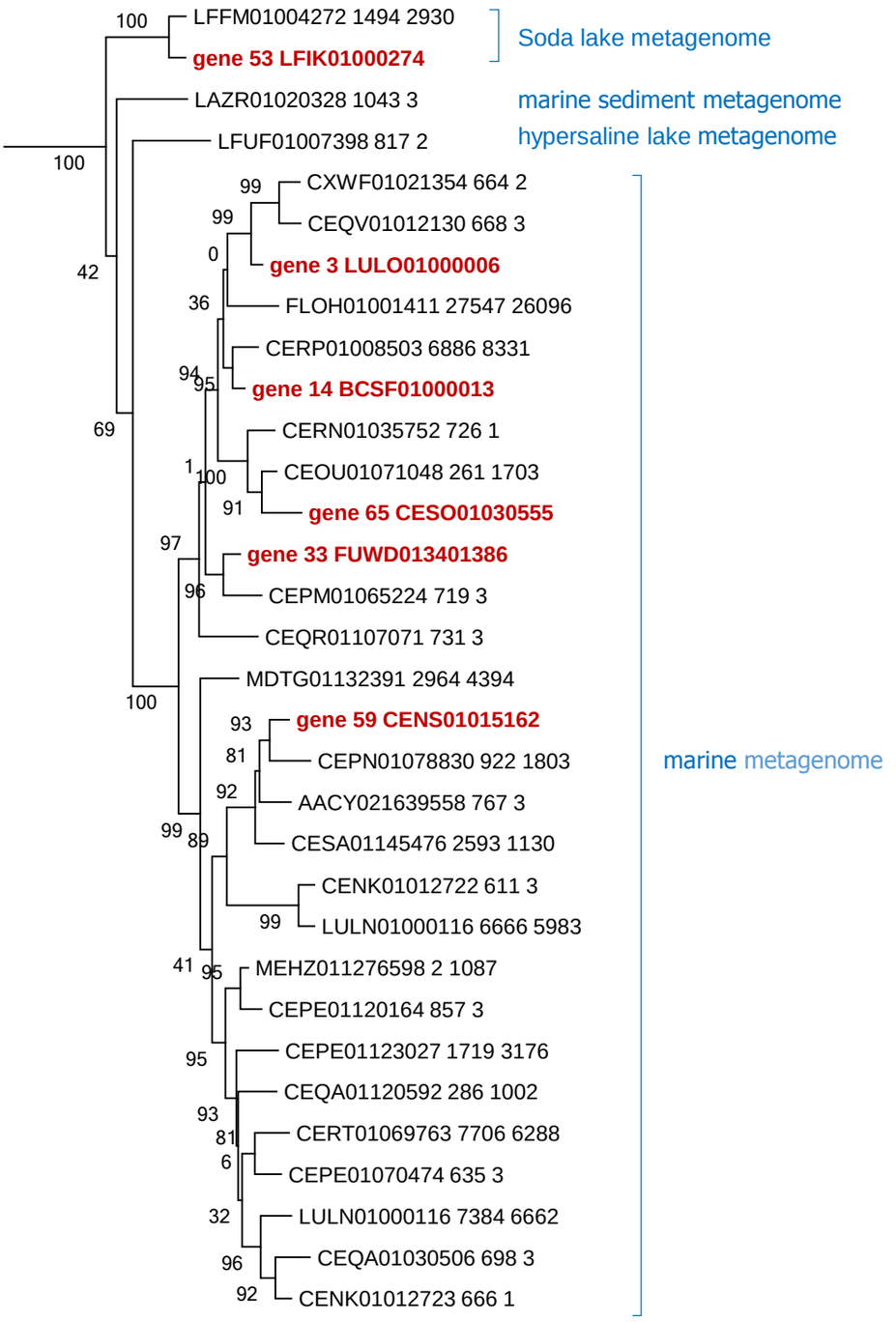
Translated wgs sequences denoted by three numbers: contig ID, orf start and orf end coordinates.

Representative sequences are shown in red, nr proteins (denoted by their source and protein ID) are shown in green. Branches corresponding to large groups are collapsed into triangles. The next 6 panels show the collapsed branches expanded. The tree was constructed using FastTree as described under Methods. Support values were obtained using 100 bootstrap replications; values greater than 50% are shown.

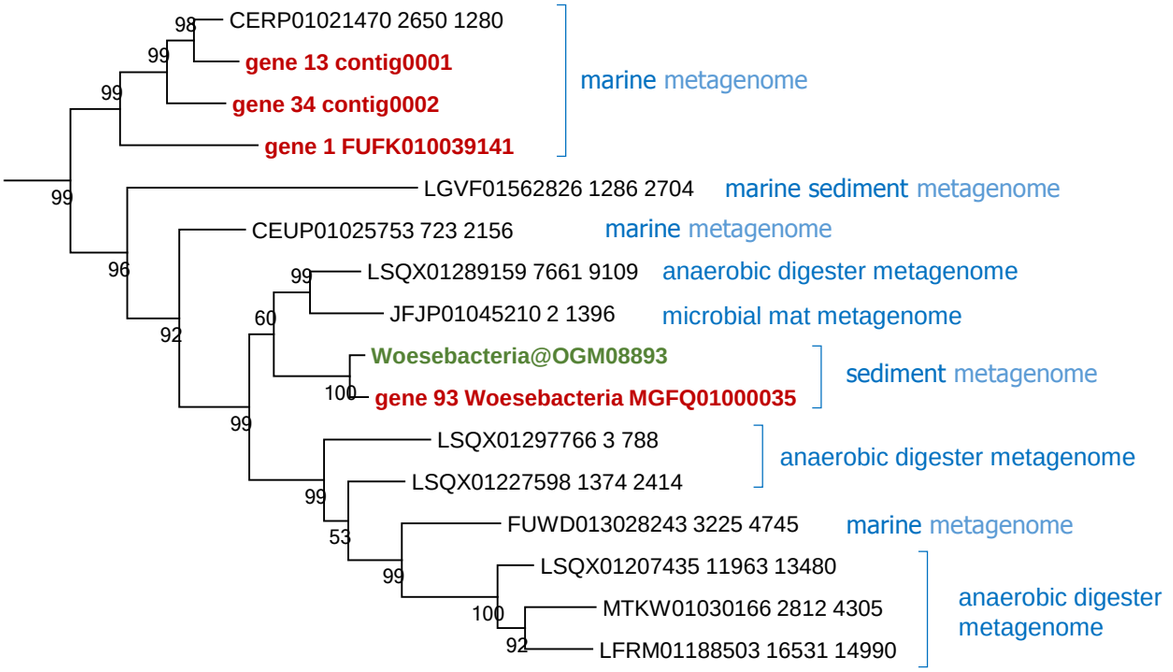
CrAssphage group



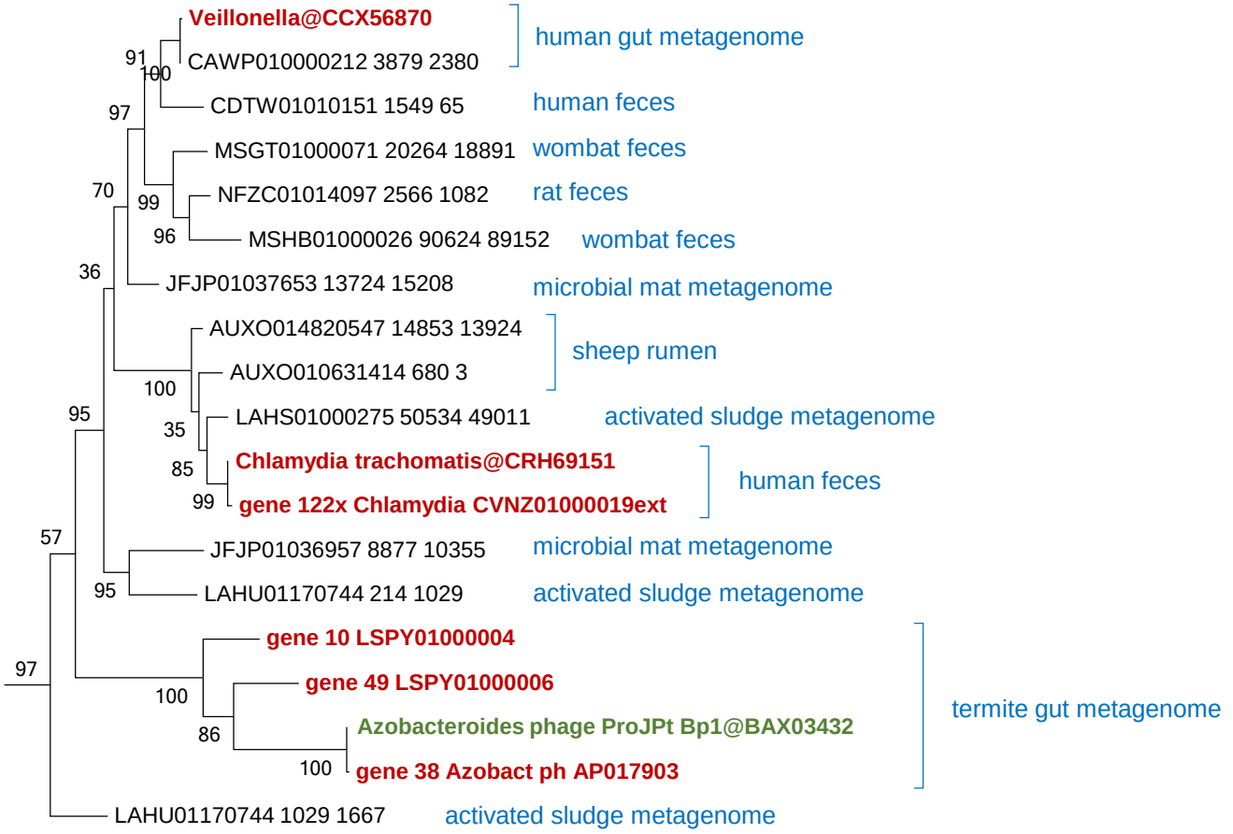
OBV-13 group


LFFM01004272 1494 2930
gene 53 LFIK01000274
Soda lake metagenome
LAZR01020328 1043 3
marine sediment metagenome
LFUF01007398 817 2
hypersaline lake metagenome
CXWF01021354 664 2
CEQV01012130 668 3
gene 3 LULO01000006
FLOH01001411 27547 26096
CERP01008503 6886 8331
gene 14 BCSF01000013
CERN01035752 726 1
CEOU01071048 261 1703
gene 65 CESO01030555
gene 33 FUWD013401386
CEPM01065224 719 3
CEQR01107071 731 3
MDTG01132391 2964 4394
gene 59 CENS01015162
CEPN01078830 922 1803
AACY021639558 767 3
CESA01145476 2593 1130
CENK01012722 611 3
LULN01000116 6666 5983
MEHZ011276598 2 1087
CEPE01120164 857 3
CEPE01123027 1719 3176
CEQA01120592 286 1002
CERT01069763 7706 6288
CEPE01070474 635 3
LULN01000116 7384 6662
CEQA01030506 698 3
CENK01012723 666 1
marine metagenome

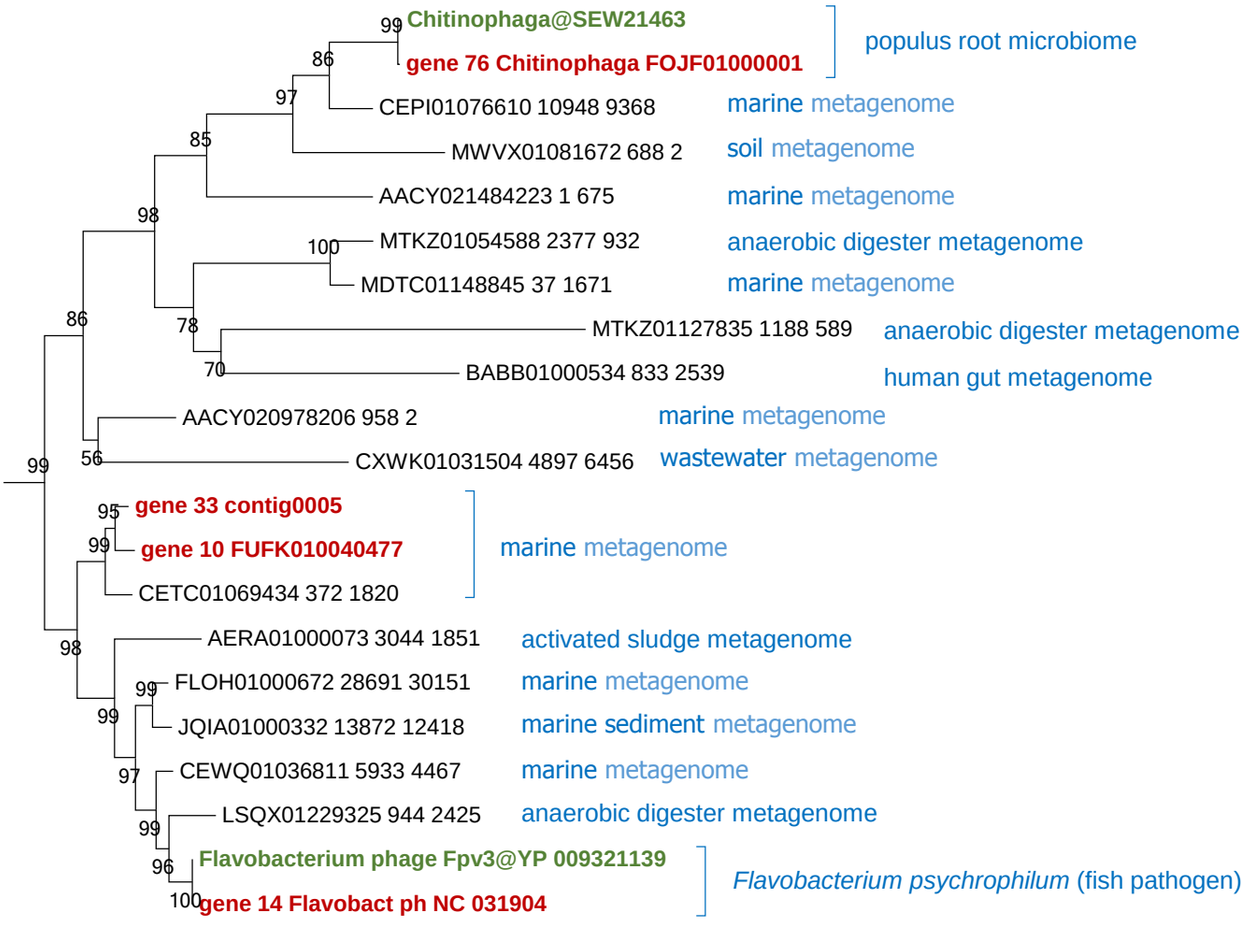
FUFK010039141 group



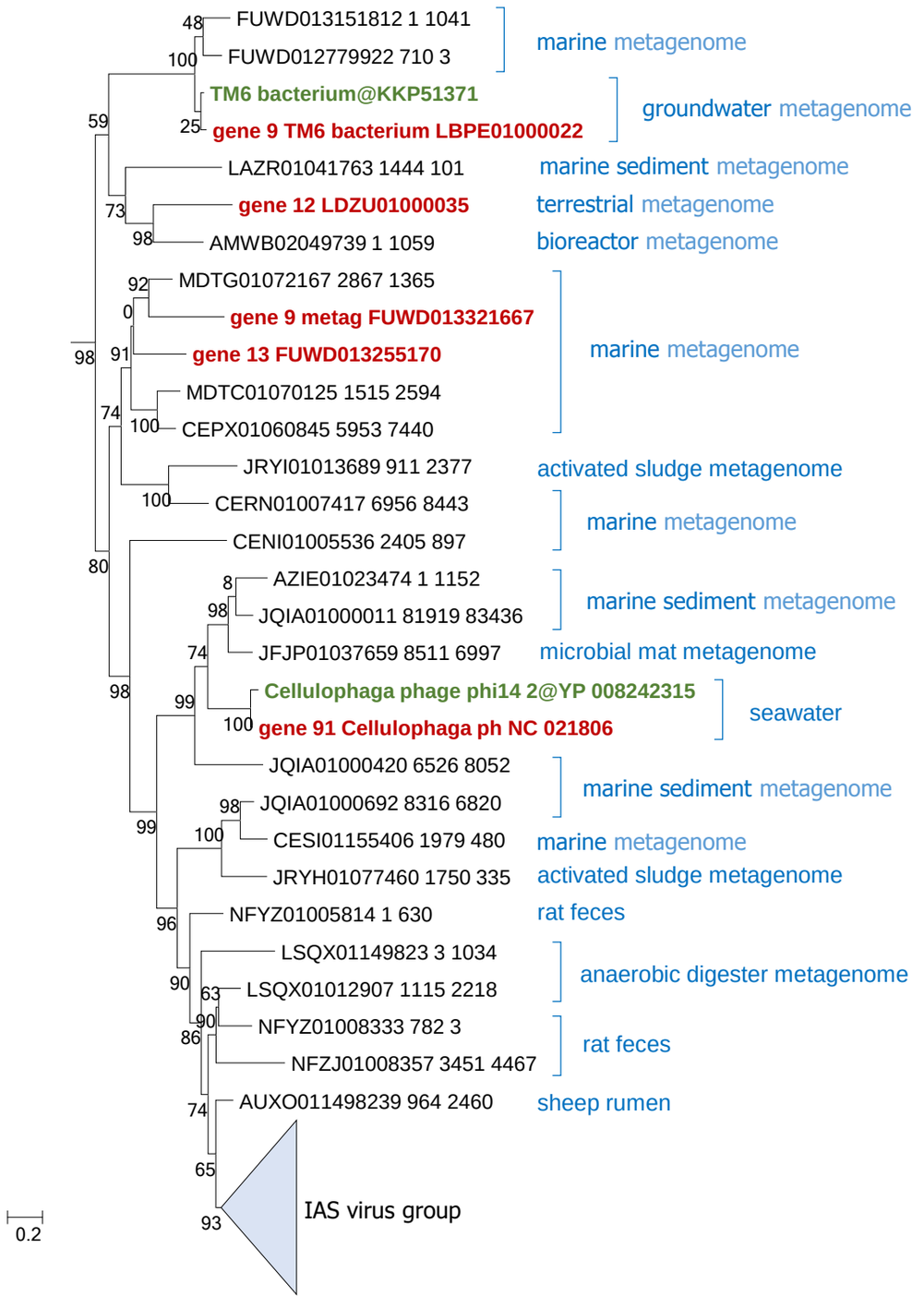
Azobacteroides phage group



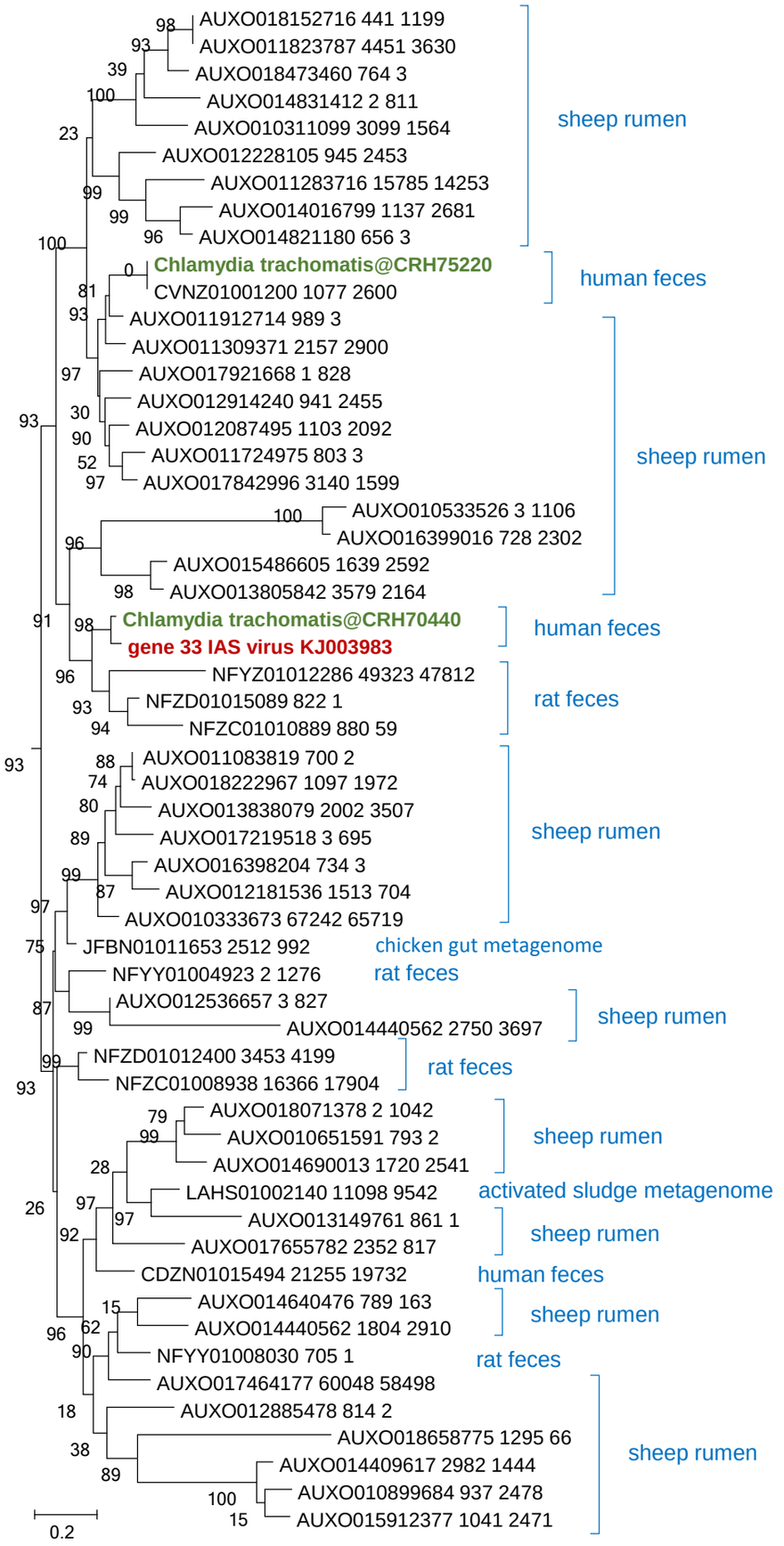
Flavobacter phage group



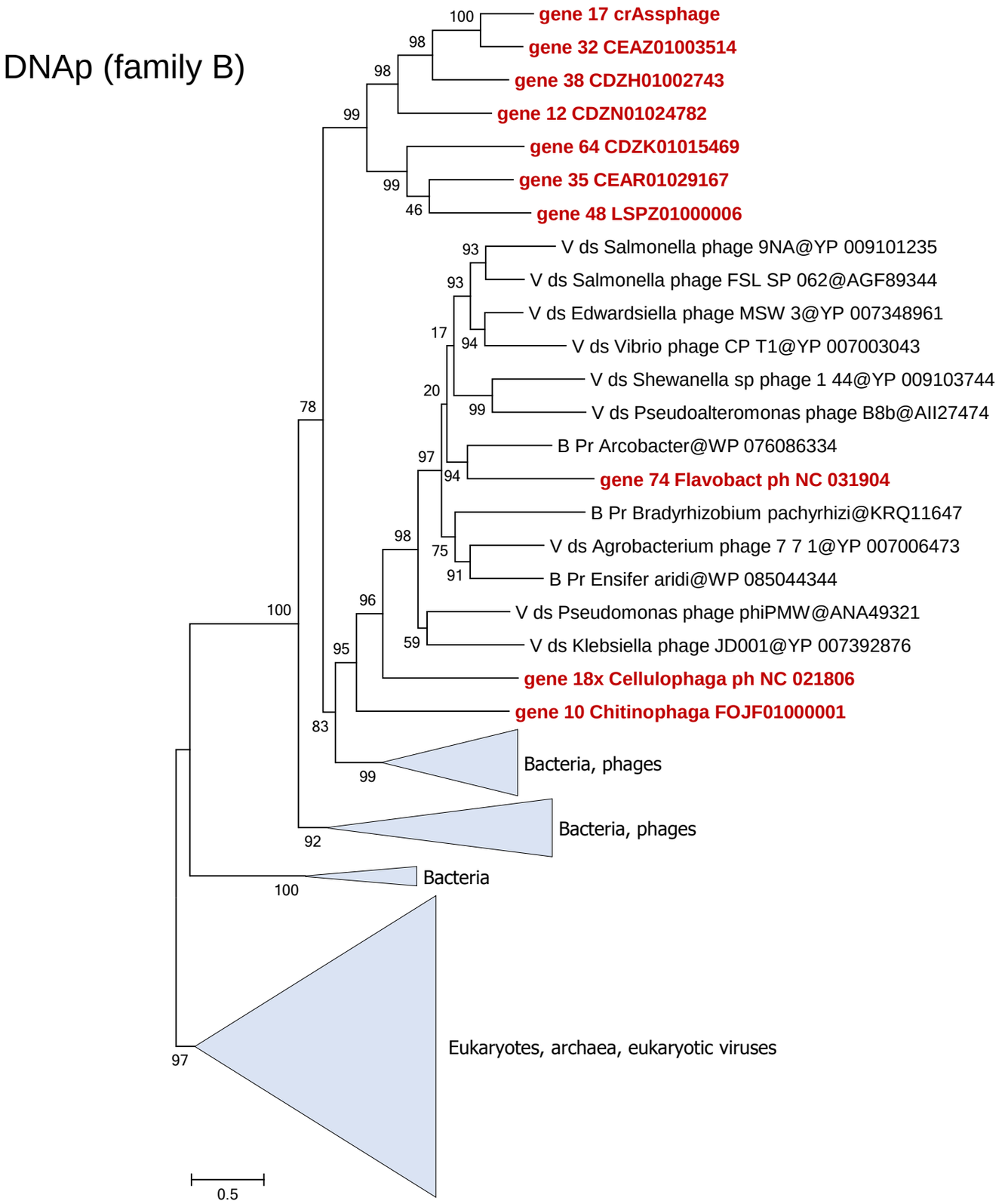
Cellulophaga phage/IAS virus group



IAS virus group

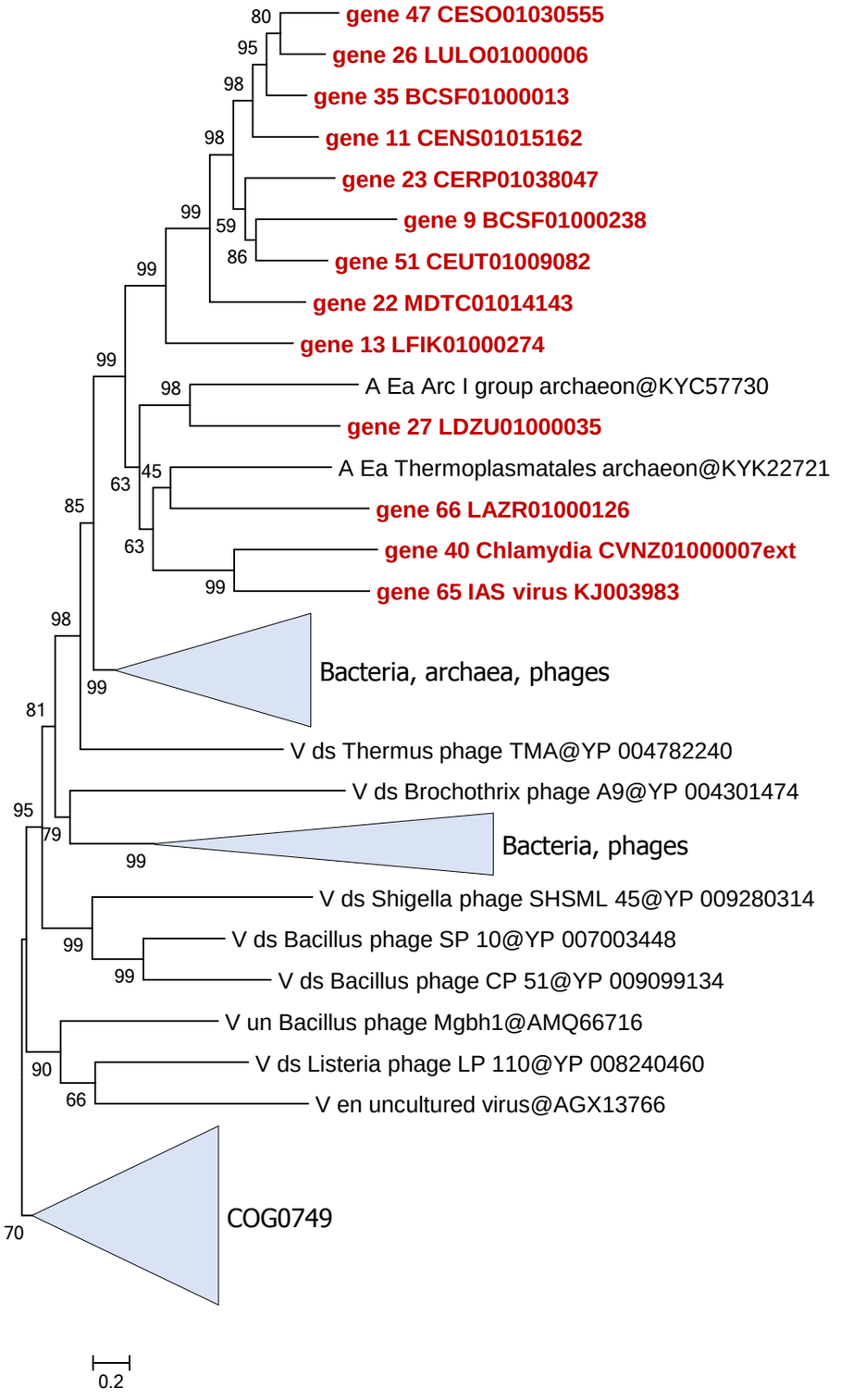


Supplementary Figure 2

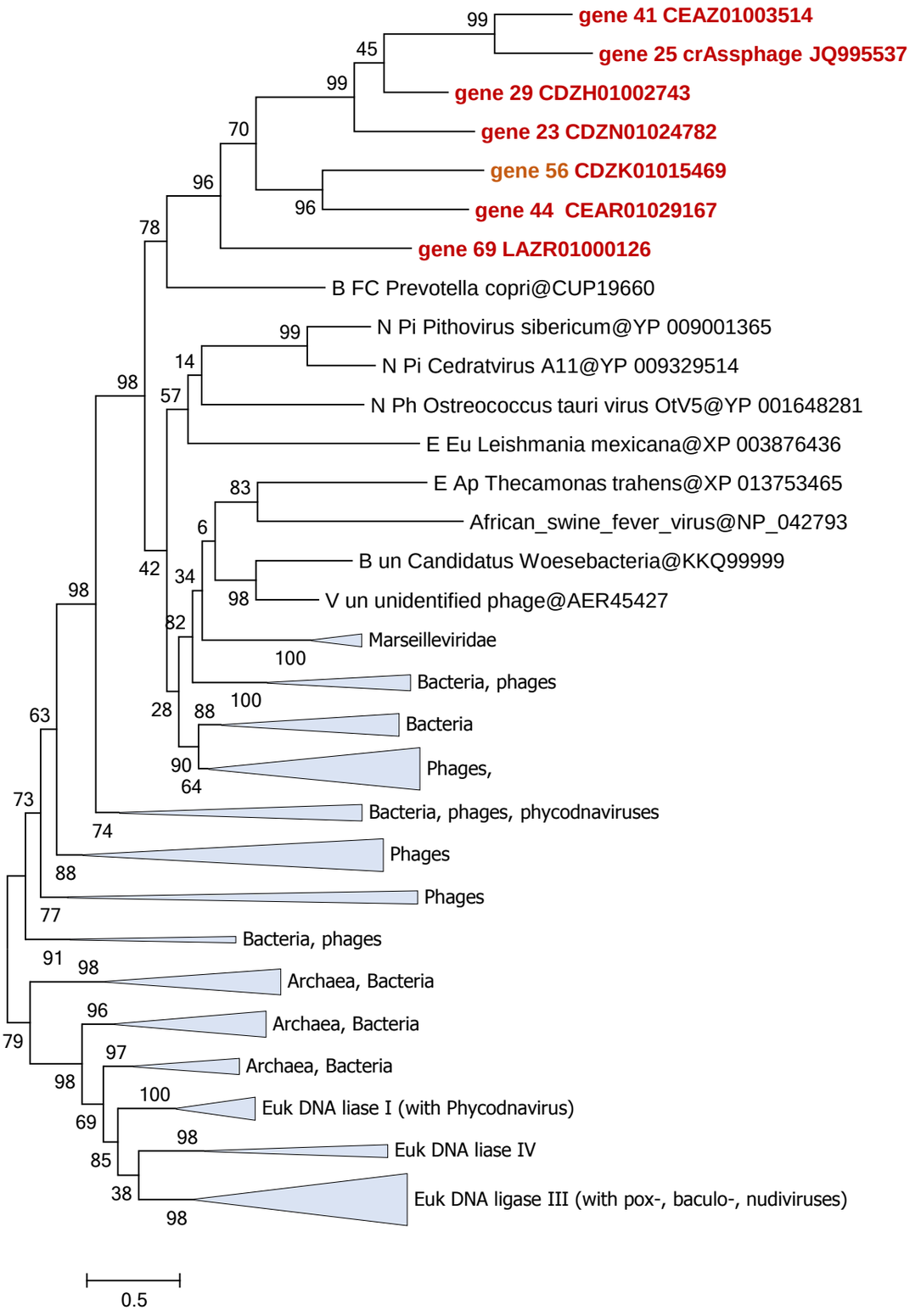


Phylogenetic trees for crAss-like family PolA, PolB, primase, and ligase.
Translated wgs sequences denoted by three numbers: contig ID, orf start and orf end coordinates.
Representative sequences are shown in red, nr proteins (denoted by their source and protein ID) are shown in green. Branches corresponding to large groups are collapsed into triangles. The next 6 panels show the collapsed branches expanded. The tree was constructed using FastTree as described under Methods. Support values were obtained using 100 bootstrap replications; values greater than 50% are shown.

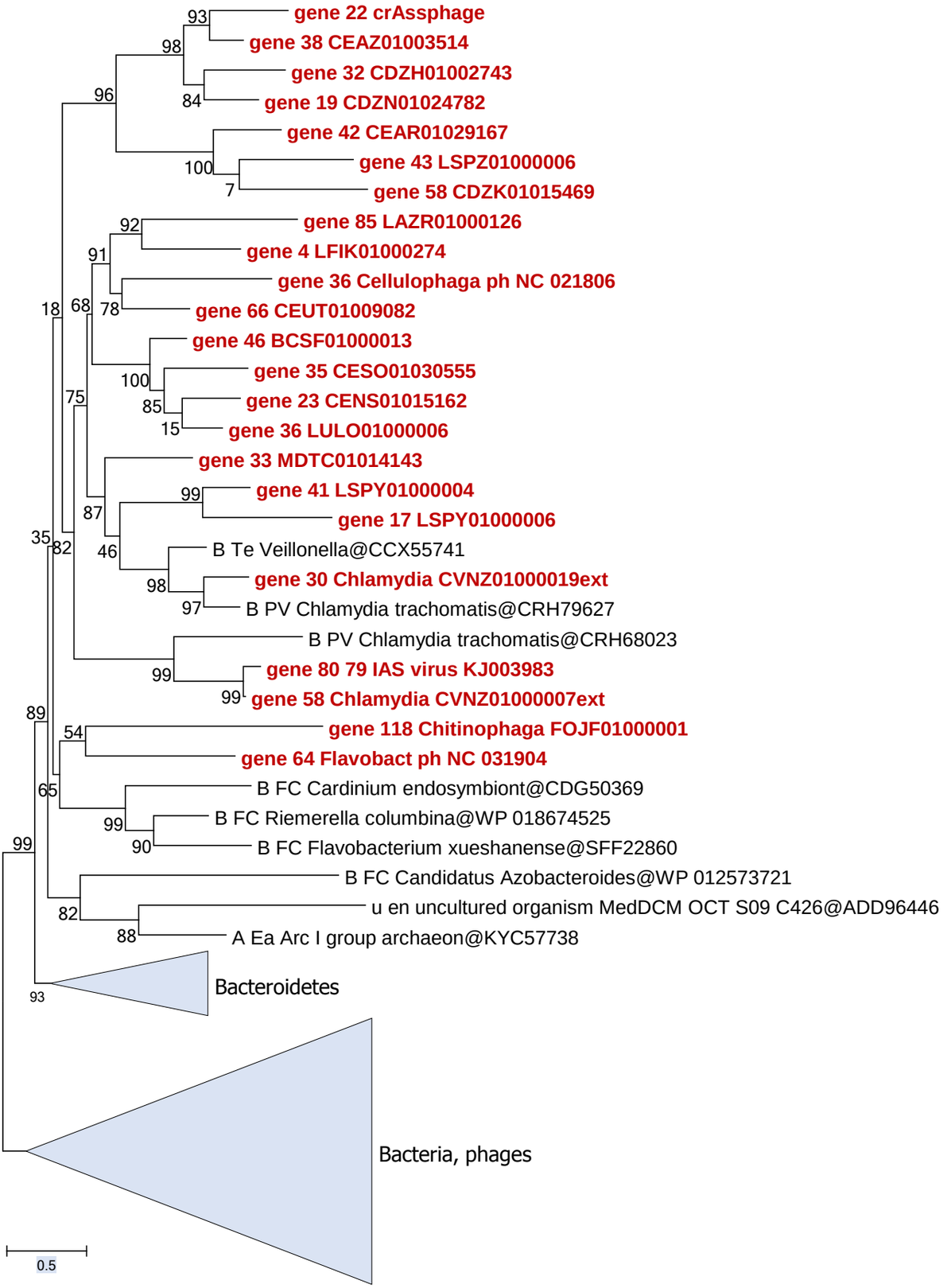
DNAp (family A)



ATP-dependent DNA ligase

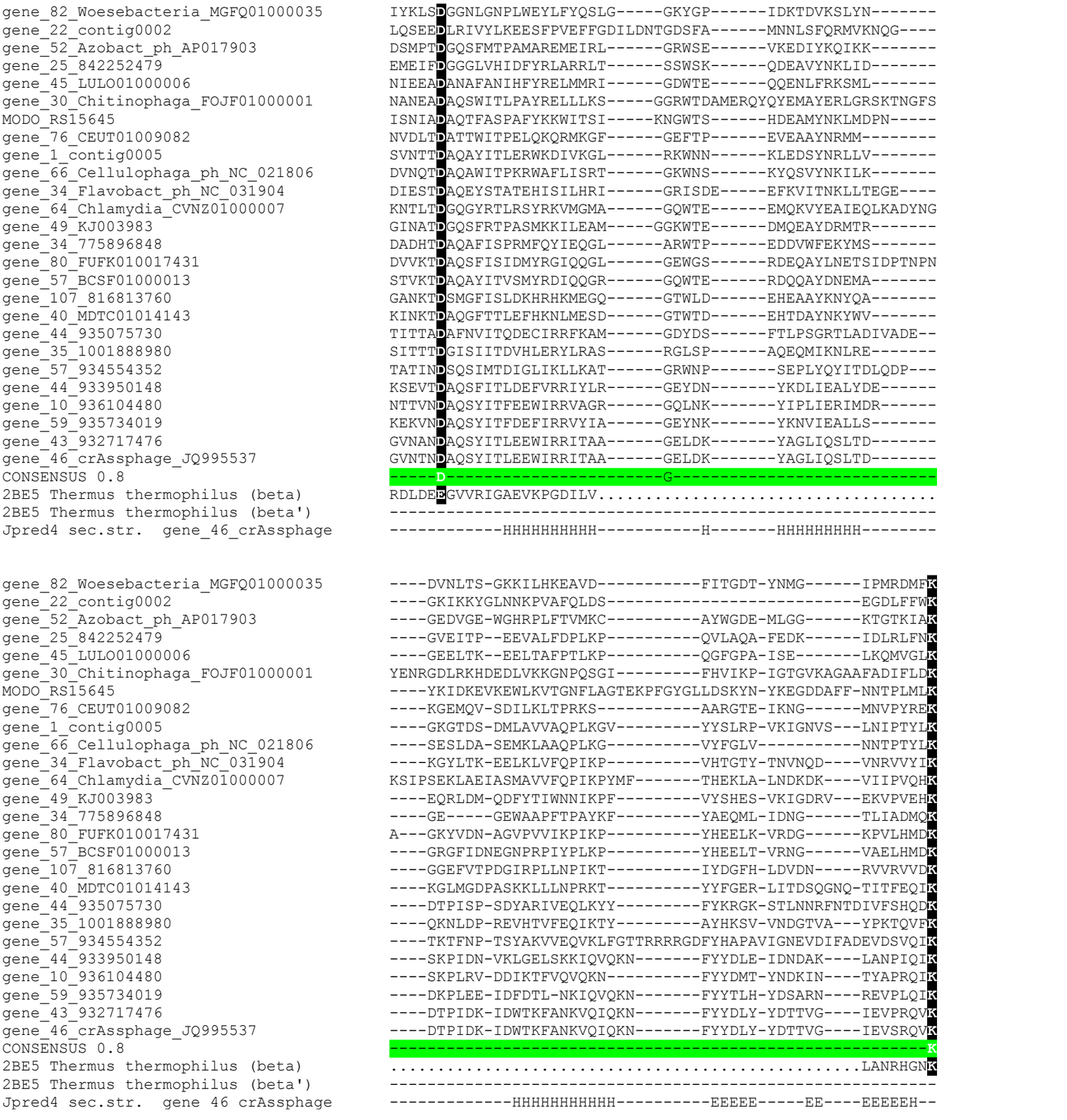


Primase


gene 22 crAssphage
gene 38 CEAZ01003514
gene 32 CDZH01002743
gene 19 CDZN01024782
gene 42 CEAR01029167
gene 43 LSPZ01000006
gene 58 CDZK01015469
gene 85 LAZR01000126
gene 4 LFIK01000274
gene 36 Cellulophaga ph NC 021806
gene 66 CEUT01009082
gene 46 BCSF01000013
gene 35 CESO01030555
gene 23 CENS01015162
gene 36 LULO01000006
gene 33 MDTC01014143
gene 41 LSPY01000004
gene 17 LSPY01000006
B Te Veillonella@CCX55741
gene 30 Chlamydia CVNZ01000019ext
B PV Chlamydia trachomatis@CRH79627
B PV Chlamydia trachomatis@CRH68023
gene 80 79 IAS virus KJ003983
gene 58 Chlamydia CVNZ01000007ext
gene 118 Chitinophaga FOJF01000001
gene 64 Flavobact ph NC 031904
B FC Cardinium endosymbiont@CDG50369
B FC Riemerella columbina@WP 018674525
B FC Flavobacterium xueshanense@SFF22860
B FC Candidatus Azobacteroides@WP 012573721
u en uncultured organism MedDCM OCT S09 C426@ADD96446
A Ea Arc I group archaeon@KYC57738
Bacteroidetes
Bacteria, phages
0.5

Supplementary Figure 3

Multiple alignment of conserved regions of β and β' RNA polymerase subunits of the crAss-like family.
The most conserved motifs of β and β' subunits are highlighted by reverse shading. High confidence alignment is highlighted by green along the Consensus line. HHpred output demonstrating similarity with catalytic region of β' subunit is shown below.



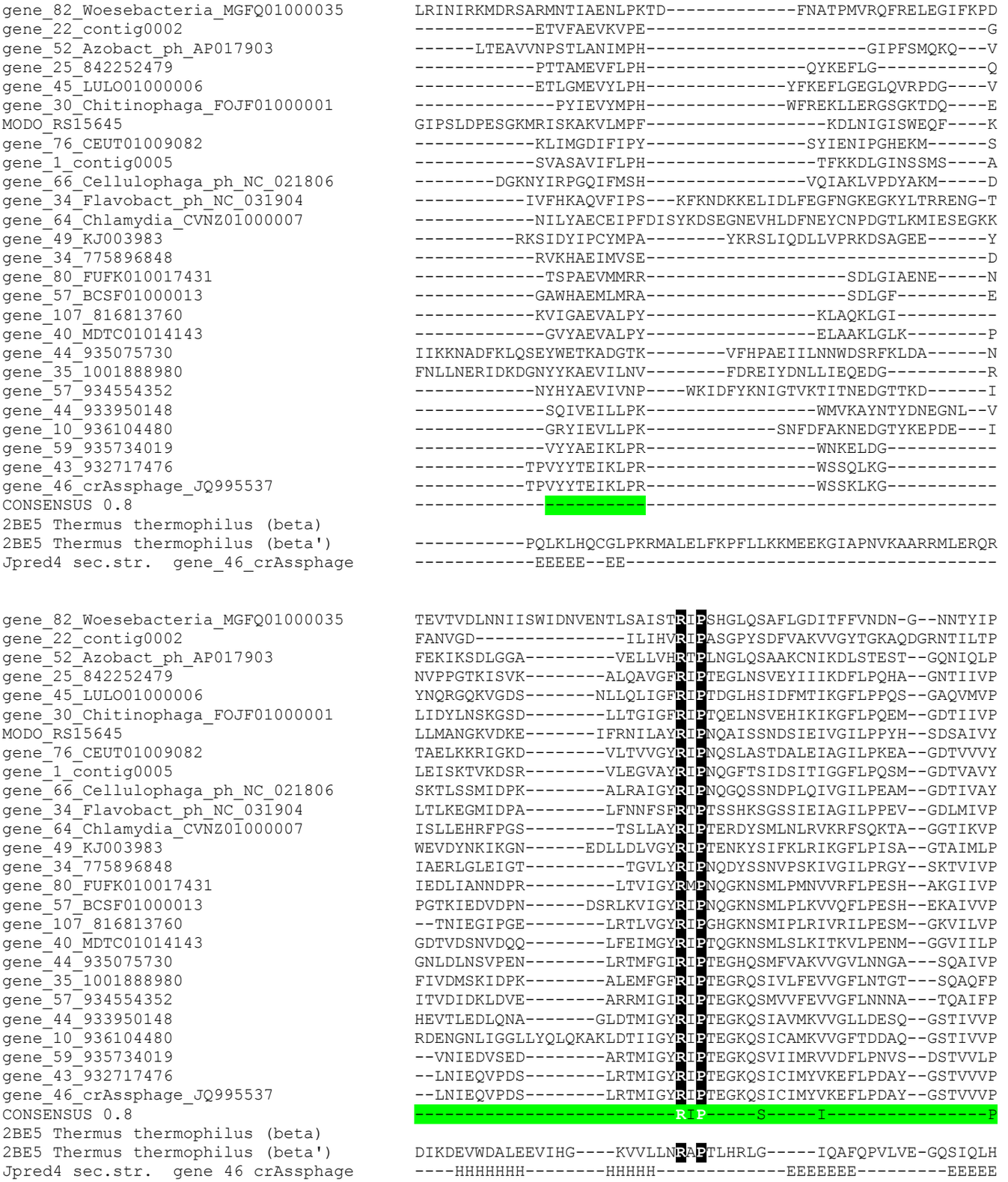
```
gene_82_Woesebacteria_MGFQ01000035      TMVDHTSMLMSKEL-----GRK--IDVFATWKMFYDY----------TGRNDFAKAIR
gene_22_contig0002                      TSTETITTDLRNT------SKA--YAFFGQLMDAFESKLGSE-T---KIVFAFESAMK
gene_52_Azobact_ph_AP017903             TSYKTLLPD---------EYPT--LQPLVKWMDKND----------IGVLHFGSAIK
gene_25_842252479                       FALFPIHPGLSKTVTDINDGNANVLDEMYADMNKHN----------LDYIVMESAVK
gene_45_LULO01000006                    LSLTPIFPQMTRI-----NGQDTALKGLLKDMYESG----------TDMVMHPTAAK
gene_30_Chitinophaga_FOJF01000001       TSAMPIYYRMS-------EGRG--LGKLYGKMSAAG----------ISYAIMESGRK
MODO_RS15645                            TSFYFLFPALM-------SGTR--GNEILKAMEGQG----------VELVSFESAVK
gene_76_CEUT01009082                    TAEVVLWPGL-V------KDAG--LKSLYDAMVQEEKDYKTMTGATLGMSVSVESAIK
gene_1_contig0005                       YSQAVLFPQVAN------ENSE--LRSIYNSMVDQG----------VDELVFDSGIK
gene_66_Cellulophaga_ph_NC_021806       YSQAVLLPQLV-------AGTQ--LQSLADAMNKQD----------IGESIVLDGVK
gene_34_Flavobact_ph_NC_031904          SSAFPLIPQLT-------AGTK--LDALRLKMEELETNTGR------FTRASFQTANK
gene_64_Chlamydia_CVNZ01000007          YAEAVLIPELLP------EGSK--LRDMAYWMDDNN----------VDLVGSTKIVK
gene_49_KJ003983                        NSEYMMTAFYNLIGSVLSKSPQ--LVALQEFMQDND----------IDVVHFESAVK
gene_34_775896848                       NSYVVLTKELA-------EGNE-LLTAMYNRMMDPNDP---------IDIINTESAKK
gene_80_FUFK010017431                   NSYVTVTKELSA------EYPV--LEKLRRVMQDQD----------IHVIHEENANK
gene_57_BCSF01000013                    NSYTVLTSELTA------NYPV--ADKLRLAMLRNN----------IGVANEEGATK
gene_107_816813760                      HVVFPLLREFTK------ASPA--FEKLRLAMEEQG----------IDEVNMESARK
gene_40_MDTC01014143                    HSTIPLIREFTEMFKGTEEQVS--LNNLRQRMEAVGKYKGLQP----IDMANFESALK
gene_44_935075730                       NSTLVIFKRMY-------KGTG--YETLYDWMKQEG----------IDSINFESGHK
gene_35_1001888980                      DSTIILFPQY-V------KGTD--MERIYNWMNSNK----------IDQLSFDSTHK
gene_57_934554352                       DSTVVLFESMT-------RGSA--MGQLYDWMIQNN----------IDQVSPISAVK
gene_44_933950148                       NAEFVLIPRFL-------GNSE--LGLLAKYMTDNN----------IGQVNFTTTEK
gene_10_936104480                       NAELVLVPRFI-------EGTD--LEKVYNLMKDHG----------IDQLNTEETSK
gene_59_935734019                       NAEYVLIPKLI-------KGTE--FEAIYNAMKAAG----------IDQLNTVETSK
gene_43_932717476                       NAEFVLIPKLI-------KGTE--LEKVYNIMTNRG----------IHQINTVETVK
gene_46_crAssphage_JQ995537             NAEFVLIPKLI-------KGTE--LEKVYNIMTNRG----------IHQINTVETVK
CONSENSUS 0.8                           -----L-------------------------M-------------------------K
2BE5 Thermus thermophilus (beta)        GVVAKILPVE
2BE5 Thermus thermophilus (beta')       ------------------------------------------------------------
Jpred4 sec.str.  gene_46_crAssphage     ----EEHHHHH------------HHHHHHHHH-----------------EEEEEEEE


gene_82_Woesebacteria_MGFQ01000035      QTVEYYIDDDEMRDS-FKSNLVWKLANRSS-----------------------------
gene_22_contig0002                      GSKKY-----------LDSGAYLDW----------------------------------
gene_52_Azobact_ph_AP017903             SWNTD------RAEL-FDNNGIFLG----------------------------------
gene_25_842252479                       LGAKKEKNTKKFSKL-IDDVGSYQP----------------------------------
gene_45_LULO01000006                    IGSRVDLVTGNAPQF-YNEDGT-------------------------------------
gene_30_Chitinophaga_FOJF01000001       VGAEF------LNPV-YQTNGDFN-----------------------------------
MODO_RS15645                            TNSVT------PTNL-VSVNQFNEYED--------------------------------
gene_76_CEUT01009082                    VGAGT------KTKI-DDDQGNI------------------------------------
gene_1_contig0005                       EGALS------PVSL-YSDAESTVLKT--------------------------------
gene_66_Cellulophaga_ph_NC_021806       VGATT------PNIV-TDENGDI------------------------------------
gene_34_Flavobact_ph_NC_031904          VGATI------KTVKPFDINSLST-----------------------------------
gene_64_Chlamydia_CVNZ01000007          VGGFG------STDISKASDAKSLGEA---------------------------------
gene_49_KJ003983                        VGGYG------QVNLNYDEERFEKLRDSNGNIKVGNEPIKATDYDTFMKNLRQAIVDE
gene_34_775896848                       GFKGQ------LFQVDQSAEGAM------------------------------------
gene_80_FUFK010017431                   GARKD------VRDA-YTDN---------------------------------------
gene_57_BCSF01000013                    GATGI------VLDL-YNDS---------------------------------------
gene_107_816813760                      VG--------------LDLSG-------------------------------------
gene_40_MDTC01014143                    VGASG------IENI-SN-----------------------------------------
gene_44_935075730                       VGGMP------KVQL-FDISRDIAVDSKGFPILDANGKYTYTNGTKATLNIQYNEATK
gene_35_1001888980                      IGGTY------PGSI-YDSEGNFIGDK--------------------------------
gene_57_934554352                       VSGIT------PVKI-HNDQGGLDIEA--------------------------------
gene_44_933950148                       ATTNR------VLEF-WDAHGKFPSKER-------------------------------
gene_10_936104480                       AGKAG------VLTL-FDEETGEVTDAH-------------------------------
gene_59_935734019                       AAKNR------LIEL-WDSSTGELTDAK-------------------------------
gene_43_932717476                       VAQHN------RMTL-WNNDGVLTDEA--------------------------------
gene_46_crAssphage_JQ995537             VAQHN------RMTL-WNNDGVLTDEA--------------------------------
CONSENSUS 0.8                           -----------------------------------------------------------
2BE5 Thermus thermophilus (beta)
2BE5 Thermus thermophilus (beta')       ------------------------------------------------------------
Jpred4 sec.str.  gene_46_crAssphage     E-----------EEE-E--------HH--------------------------------
```

```
gene_82_Woesebacteria_MGFQ01000035      ---VKGAISKINIYNPWNDIYDPLNIKEGKRGLDLLSDDI----------------RN
gene_22_contig0002                      ---KDTYNKSFESETDIDGV----------------------------------------
gene_52_Azobact_ph_AP017903             ---TDDLNDNYVST----------------------------------------------
gene_25_842252479                       ---LQNLTSVQ-------------------------------------------------
gene_45_LULO01000006                    ---INPIDSDL-----------------------------------------------H
gene_30_Chitinophaga_FOJF01000001       ---PEPFSGNV-----------------------------------------------
MODO_RS15645                            ---IMDYNEDPFSYSPGKFRL--------------------------------------NP
gene_76_CEUT01009082                    ---IEDVTLNPIVR---------------------------------------------
gene_1_contig0005                       ---GEDIKFNKMKL---------------------------------------------
gene_66_Cellulophaga_ph_NC_021806       ---LKSISLNPLTL---------------------------------------------
gene_34_Flavobact_ph_NC_031904          ---IKDYNADDVNA-------------------------------------------NM
gene_64_Chlamydia_CVNZ01000007          ---LNKAYVHQ------------------------------------------------
gene_49_KJ003983                        KITQEEYNKAIKDIDFENSEDGIDSCYSQLERAFGVLDENGEVVLDEDENLVLNEEKL
gene_34_775896848                       ---FDGLQAREMDA---------------------------------------------
gene_80_FUFK010017431                   ---LDNLKTFV------------------------------------------------
gene_57_BCSF01000013                    ---LENIQSNILKK---------------------------------------------
gene_107_816813760                      ---IKDINDDLSNW-------------------------------------------KT
gene_40_MDTC01014143                    ---LEGLNVEI------------------------------------------------
gene_44_935075730                       RLELKGYPKGVEDF---------------------------------------------K
gene_35_1001888980                      ---MSKQDRELSTT---------------------------------------------
gene_57_934554352                       ---LNRVDDRSILY---------------------------------------------
gene_44_933950148                       ---LKRFNEDIQTK---------------------------------------------Y
gene_10_936104480                       ---IQDFNNHVEDY---------------------------------------------K
gene_59_935734019                       ---LQRFIVDADKY---------------------------------------------K
gene_43_932717476                       ---LKEFDNNIFDN---------------------------------------------S
gene_46_crAssphage_JQ995537             ---LKKFDNNVFDN---------------------------------------------S
CONSENSUS 0.8                           ------------------------------------------------------------
2BE5 Thermus thermophilus (beta)
2BE5 Thermus thermophilus (beta')       ------------------------------------------------------------
Jpred4 sec.str.  gene_46_crAssphage     ---HHH---------------------------------------------------------


gene_82_Woesebacteria_MGFQ01000035      DKIKVVLNPYQDVFD-----PSRVT--------PMLQIMSYAGV---------------
gene_22_contig0002                      -QGYKKYKFNQDMLN---SLAASITEDVIPFSNIGMQGTVSKN---PNIGDI------
gene_52_Azobact_ph_AP017903             --HSWDNLRWQLPMD---IHPIRERG-------LSVQLIKTLS--DDVALHPI-----
gene_25_842252479                       -EYGLEYFGMQMDPK---GKFGQSVT-------LGTQSTSMLP--TNIFNNGV-----
gene_45_LULO01000006                    TVIDLSYFGIQVDIN---KEFKGKVT-------LGTQSRTHIL--ANLYEGGMPADYT
gene_30_Chitinophaga_FOJF01000001       -NVPFKYFGIQLETA---GTKHKQTR--------GTQLTKLAI--INLLASGIPQDIS
MODO_RS15645                            FIIPNDFIGRQQDTP---NKGAYEGD-------APKQAMKNIL--ANLDISSE-----
gene_76_CEUT01009082                    ---SHLLYGKQQETK---TKGIKSTT-------FASQLKLNII--ADIDDTTL-----
gene_1_contig0005                       ---ENRYWGLQQDLT---PHTTDTQL-------EGSQVTKNILWVDPQKFSKE-----
gene_66_Cellulophaga_ph_NC_021806       ---SNADWKLQQDLP---VKTIKPTL-------LGSQIQKNIY--SSLTDEAT-----
gene_34_Flavobact_ph_NC_031904          LVLNRDNFRIQQDVPFKSDKHQDDKV------SMGTQFFKLLF--GDGVIDKE-----
gene_64_Chlamydia_CVNZ01000007          --LSYGDYRIQTNVP---EHINSSQL-------FGTQVRKLIM--ANIMMDDYHYMNY
gene_49_KJ003983                        HVIPLSDYMVVQPSG---DHLTEDDL----MAIFGSQLRNILP--ADLPNDFS-----
gene_34_775896848                       -----SKLFMPQIIN---DKADMEAK-------MNRQIRKGIP--SMVENDVT-----
gene_80_FUFK010017431                   --LDSTKLKLPQIIP---VSKKDKIT-------FSRQIRKNII--SDWLLDPN-----
gene_57_BCSF01000013                    ---KSRNLRFPQTI----SRVEKPEI------VFARQVRKNIL--ANLTPEDK-----
gene_107_816813760                      VELETRFQRIPQIVP----ENDKPVM-------MSTQARKHII--SNMLSDEV-----
gene_40_MDTC01014143                    --LETKNLRSPQVIE---TKTKAPLD--------GTQMAKLIL--SNILDSYN-----
gene_44_935075730                       QTLSHSNLYIQQQVP---SHLMDEEN------KIGTQLQKRIL--DNLVFNGD-----
gene_35_1001888980                      -RIPITSYVIQQDIH---DHIMNAED------ATGSQVGRITL--SGLNSEEA-----
gene_57_934554352                       --MRDSDFVIQQDIK---ADLLDETT------ILGGQLVKQIM--EGLDWNNA-----
gene_44_933950148                       KTGWYSNLYTQQDIP---QHMDGENK-------AGLQIVKKLI--DNIGNTPE-----
gene_10_936104480                       EIYSYNFLYTQQETP---QHMNAENK-------AAIQIMKKIV--DNIPDTGT-----
gene_59_935734019                       EPYQYTYLYRQQEVP---SHLKDKEN------KIGIQIYKKLL--DNIPNDTT-----
gene_43_932717476                       ELFSYNYLYRQQEVP---QHMVDASN------KAAIQIMKKML--DNLPNRTE-----
gene_46_crAssphage_JQ995537             ELFSYNYLYRQQEVP---QHMVDASN------KAAIQIMKKML--DNLPNRTE-----
CONSENSUS 0.8                           ------------------------------------Q-----------------------
2BE5 Thermus thermophilus (beta)
2BE5 Thermus thermophilus (beta')       ------------------------------------------------------------
Jpred4 sec.str.  gene_46_crAssphage     HHH-HHHHHHHH------HHHH-----------HHHHHHHHHH--H------------
```

```
gene_82_Woesebacteria_MGFQ01000035      -----------GNEVNRRAGKEINRLFRTLFINAKREIDDKISEEI------------
gene_22_contig0002                      ----------------ALSNQKMMYTMNYSSDLSQKNKIEKLLTDLL-----------
gene_52_Azobact_ph_AP017903             ----------------------------------INKMSERKALDME-----------
gene_25_842252479                       ----VSDEYANTKFSDTETWEQAIDRYHNINKELIAREASTLASELGFSY--------
gene_45_LULO01000006                    GDIASWEKMTPEQKKKASSVHKDASDYLDAVNEITRRSRENLLKRFNLEKVV------
gene_30_Chitinophaga_FOJF01000001       MDAAAWNKLSETEKIQLSPSYKLISANIAILKAMADKGYQQVLDQLGISE--------
MODO_RS15645                            -------EKNYTYKGKNYTAKEMFDKVTSMGNTIINKQIDKFKRDMYT----------
gene_76_CEUT01009082                    --------------IGNRSGVEWREEVEANEIAISNLGREEFLTKYGI----------
gene_1_contig0005                       ----------------------IEEFHALTADKSRRGRNKLLRELGATFN-------
gene_66_Cellulophaga_ph_NC_021806       ----------YTIENEAFNGSGMFQAINDTVSAMSNLSIAGLSSELGK----------
gene_34_Flavobact_ph_NC_031904          ---------GFVIDGKELTGKELYQHFNKAFSDIVDSKKQELFLDLGL----------
gene_64_Chlamydia_CVNZ01000007          IGGDGTVNLGGNHGRVKLNGRNLVSFYNGLIVANILESFDNFKDGISD----------
gene_49_KJ003983                        ------VDVTFGKDKVTLDRDKYIELYNTLIVDQLLDSFSKVDKEFSS----------
gene_34_775896848                       ---------YTLPDGTTLTGEELLKQYHQTHSDIVDAQYDRLEKELGYD---------
gene_80_FUFK010017431                   ----------------------LNLEFQNLIAENIKEDTKKIEDELGLTALRNAKSL-
gene_57_BCSF01000013                    ---------------------PLYDTFHELVSENVNEDLKSLQDELGITE--------
gene_107_816813760                      GGKKGQWTQGYILGGKKIGGEELFDLYHNVTEQLVGRSLDDLNKQLGYTKEIVTETPF
gene_40_MDTC01014143                    ----------YNVFGGKYNGKQIKDLYNEIYAERIKRSHDALTNELGWDS--------
gene_44_935075730                       -YTIGSTVRKGKTGDYSYDGSGAFEYYQMLLSANANDEMYRLLADWGAI---------
gene_35_1001888980                      ---------IYELNGKKLTGNQIAVEFSNSLHSLVYDGLVRFFNKVGA----------
gene_57_934554352                       ---------IYELDGKNVTGKELFDEFQKTLATNIREDAMQLLYDIGGI---------
gene_44_933950148                       -------------------GQSLIKDFFDNFTANIQDSFKDAASRIGVSI--------
gene_10_936104480                       --------------------IGEVKKEFFKLYVANIKDSFNSLVKELNIPT--------
gene_59_935734019                       --------------------GRALKQRVFRNIVSNINTSFTNACTLLNIPL--------
gene_43_932717476                       --------------------LNDLKSKVFNNYVANIRNSFEKTCAELGIGL--------
gene_46_crAssphage_JQ995537             --------------------LNELKDKVFNNYVANIRNSFEKTCAELGIGL--------
CONSENSUS 0.8                           ------------------------------------------------------------
2BE5 Thermus thermophilus (beta)
2BE5 Thermus thermophilus (beta')       ------------------------------------------------------------
Jpred4 sec.str.  gene_46_crAssphage     --------------------HHHHHHHHHHHHHHHHHHHHHHHHHH------------

gene_82_Woesebacteria_MGFQ01000035      ------------------------------------------------------------
gene_22_contig0002                      ------------------------------------------------------------
gene_52_Azobact_ph_AP017903             ------------------------------------------------------------
gene_25_842252479                       ------------------------------------------------------------
gene_45_LULO01000006                    ------------------------------------------------------------
gene_30_Chitinophaga_FOJF01000001       ------------------------------------------------------------
MODO_RS15645                            ------------------------------------------------------------
gene_76_CEUT01009082                    ------------------------------------------------------------
gene_1_contig0005                       ------------------------------------------------------------
gene_66_Cellulophaga_ph_NC_021806       ------------------------------------------------------------
gene_34_Flavobact_ph_NC_031904          ------------------------------------------------------------
gene_64_Chlamydia_CVNZ01000007          ------------------------------------------------------------
gene_49_KJ003983                        ------------------------------------------------------------
gene_34_775896848                       ------------------------------------------------------------
gene_80_FUFK010017431                   ------------------------------------------------------------
gene_57_BCSF01000013                    ------------------------------------------------------------
gene_107_816813760                      PNSITQETYYHGSNSPDIKNFDLSLQGKAHEGRGINFALTREDAKVYGTHIYEVKLNI
gene_40_MDTC01014143                    ------------------------------------------------------------
gene_44_935075730                       ------------------------------------------------------------
gene_35_1001888980                      ------------------------------------------------------------
gene_57_934554352                       ------------------------------------------------------------
gene_44_933950148                       ------------------------------------------------------------
gene_10_936104480                       ------------------------------------------------------------
gene_59_935734019                       ------------------------------------------------------------
gene_43_932717476                       ------------------------------------------------------------
gene_46_crAssphage_JQ995537             ------------------------------------------------------------
CONSENSUS 0.8                           ------------------------------------------------------------
2BE5 Thermus thermophilus (beta)
2BE5 Thermus thermophilus (beta')       ------------------------------------------------------------
Jpred4 sec.str.  gene 46 crAssphage     ------------------------------------------------------------
```

```
gene_82_Woesebacteria_MGFQ01000035      ------------VEPG---LKG------EDRRIAQ----------YEKYLRGVGMD
gene_22_contig0002                      ------------------SIQL------DKSKAS-----------FSSTENIVRTF
gene_52_Azobact_ph_AP017903             ---------------------------------------------DERIPNKLVFY
gene_25_842252479                       ------------TPNG--QFQF------LGS--------------NRKMLDTIL-G
gene_45_LULO01000006                    ------------DSKG--EVSY------KLTGGN-----------IQKFADMIR-D
gene_30_Chitinophaga_FOJF01000001       ------------TENG-FHIA-----------------------DKKLVAAFISR
MODO_RS15645                            ------------DKYGVYSDRL------N----------------DAKIRNLVI-T
gene_76_CEUT01009082                    ------------DENF--DVN-----------------------EETLYKALI-D
gene_1_contig0005                       ------------EQSGVYRIND-----------------------QRKLIDKVL-T
gene_66_Cellulophaga_ph_NC_021806       ------------DSEG--KID-----------------------KRKLYDMLE-R
gene_34_Flavobact_ph_NC_031904          ------------DTNG--QIKN------EQNF-------------VKNLQDLLI-K
gene_64_Chlamydia_CVNZ01000007          ---------------------------------------------ANKLSEKLI-Q
gene_49_KJ003983                        ---------------------------------------------AENLQRALLNK
gene_34_775896848                       ------------------KLKK------DPSNPELRLNF-------LQRVRELIYDN
gene_80_FUFK010017431                   ------------EDKRDAKLDH-----------------------LQKLKERLL-S
gene_57_BCSF01000013                    ------------LNNA--KTDT------QRLKAKRKY---------LEKLRSALI-R
gene_107_816813760                      SKWVTTDAEAKTKTTYGIKDVDEIVVYHPDQVQIIKEEIKGLEEKEKFYTNLGRVIKR
gene_40_MDTC01014143                    ------------YQQAIQSTDL------ETRNKAQLDF--------LLKVKDVAI-N
gene_44_935075730                       ------------TNDG--NIKY------TSIETDGGLRNVIGVD--LDLVLADLR-R
gene_35_1001888980                      ------------ISNG--KINI------DNGTVVVDRQ------KVMRSVKNSLL-T
gene_57_934554352                       ------------DENG--EIRT------DARGAIQID---------INKLVNRFQ-E
gene_44_933950148                       ------------DARG--NVVY------EEGKVKID----------NNQFIALIK-D
gene_10_936104480                       ------------NEDG--SIKL------DANGNIEGLD--------MKLFFNKLR-K
gene_59_935734019                       ------------DSNG--NLQF------DEQGNILGLN--------YERLLKLAR-E
gene_43_932717476                       ------------DDNG--HIDL------NSNGTIKNLN--------RFVFFDRFK-E
gene_46_crAssphage_JQ995537             ------------DDNG--HIDL------NSNGTIKNLN--------RFVFFDRFK-E
CONSENSUS 0.8                           ---------------------------------------------------------
2BE5 Thermus thermophilus (beta)
2BE5 Thermus thermophilus (beta')       ---------------------------------------------------------
Jpred4 sec.str.  gene_46_crAssphage     --------------------E---------------HHH--------HHHHHHHHH-H


gene_82_Woesebacteria_MGFQ01000035      ALNEM--GQIGNYAE----LVLNEQ------------ISV-GIPPL-RQKLLQAYRSL
gene_22_contig0002                      GNQPL--GGAGVYGQNIALSALDNN-------------SLLNHPNV-SNLLQAQLSRM
gene_52_Azobact_ph_AP017903             GARPT--NATDNVLH------AAED--------GASVLAF-------KGLGKLRSSIS
gene_25_842252479                       EMTKR--DIPENVQA-----SIIDL--------FSREVTF-SGQLFEKNKIETLFNAI
gene_45_LULO01000006                    EIAKR--NYPENVAAGIESLLNQPD--------GDKKV---FDIIVNKGRLENLLFSL
gene_30_Chitinophaga_FOJF01000001       EAEAR--DLPDNFKS---AIGIDAE-------TGDFKISLEAISNY--RAVKSIIYSI
MODO_RS15645                            QLDPV--KDMDNIKL-----------------LRDPNVPIATMISS-TQWILPIITKK
gene_76_CEUT01009082                    TLSRQ--SFPDTALI----EGLENG------------LKFDSIFQS-RKKVMNLLANE
gene_1_contig0005                       SLKQTLSGLPENIEQGLKEVYTHIN-PVTKKEHTYLKNPLDSHPFY--KRIMPVLASM
gene_66_Cellulophaga_ph_NC_021806       EMLDK--GSAINLLK-----SIQKN------------LPIEAMPGI-KDKLYNIVFSK
gene_34_Flavobact_ph_NC_031904          EATSR--GYSVKSLAGLKIEKLAAA----AGFYYEFKTPL-WLSSD-SNRYESLLNSI
gene_64_Chlamydia_CVNZ01000007          GVINNSRESMDNVLA-----YAVDD--------GKFAMPL-FEGGL-EHDAAAMAFSM
gene_49_KJ003983                        IHNNL--KYGPEVEE---ALELNED-------KTGFKIPF-NNPNL-KNKIEELLLST
gene_34_775896848                       AARNN--KIDSNLDK---QMIIVED---ILQQRYDFNVPI-SFPVY-QREYQNLISAL
gene_80_FUFK010017431                   QIREK--GLPDSYLG---ALNIVPD----GQFDYRFAIPL-AFPNF-QAKFEQIYFSL
gene_57_BCSF01000013                    TLKDK--NLPDSYLG---ALDIVPN----GEGDVRFDIPL-AFPNY-QAKFEQIFFSM
gene_107_816813760                      SIEER--DLPENYERALNIIKTEND--------VTFELPL-AFPAY-QKRFEQIILSL
gene_40_MDTC01014143                    SLEER--DLPDNYYLAMKIEKLVDD-----VNAYGFEAPL-SFPPF-AKRFESILLSL
gene_44_935075730                       YFNET--EIDRNFIK----ATVVVN--------GKPFIPF-YHPTI-KSRIESVLLAR
gene_35_1001888980                      NVKSQ--NDIASISD--TVDYVMDELGNVVDTNNEMMFAL-SNPAI-RTKLASYMNSL
gene_57_934554352                       IISDD--VDAITIRK---ALEIKED--------GLPTMPL-SYPVI-KGKLEKILASM
gene_44_933950148                       ELTRR--GLDSNYRK---YAEINPE-------TGLPYMPA-WTNLV-RSKIENIVNSI
gene_10_936104480                       ECLRQ--GLDSNILE---FFTLNEDSPYTELGRANTVMPT-YMTNM-MSKAQNVCQSM
gene_59_935734019                       NAARN--GADKNTLD---FLTTDEN--------GNPRYPM-YLNSI-SNKIENLVNGI
gene_43_932717476                       NAQQQ--GVEKALLE---FFDLDSA--------GFNNLPL-FLSNI-NSKLESIANSY
gene_46_crAssphage_JQ995537             NAQQQ--GVEKALLE---FFDLDSA--------GFNNLPL-FLSNI-NSKLESIANSY
CONSENSUS 0.8                           -------==================================================
2BE5 Thermus thermophilus (beta)
2BE5 Thermus thermophilus (beta')       ---------------------------------------------RGAPVTNPGSD
Jpred4 sec.str.  gene 46 crAssphage     HHHH---HHHHHHHH---HHH-----------------E-E------HHHHHHHHH-
```

```
gene_82_Woesebacteria_MGFQ01000035      ITRKAIKPFFSGLRANQVTADYFDVFEKDGIPYSLKEVERDLKM--------EITREN
gene_22_contig0002                      LGNAGTKPRVSGAQLHMIPDIGSEE----------------------------NLKF--
gene_52_Azobact_ph_AP017903             GLAATVRTDLQGGAFVRTAAVGRQ-----------------------------DLRY--
gene_25_842252479                       VTNTVVRRKMNGNMVVLQSNFGYETMSKARKQDKNVPLER------------KLKSYG
gene_45_LULO01000006                    VGNNSITQKVKGENAVQVSSLGSENTVRIVRQGKVESSGRN-----------GLKF--
gene_30_Chitinophaga_FOJF01000001       VDKKLLSPKMPGGPKVQVSSALFERNNRQAVYKPAGAKTANWER--------VTDFDK
MODO_RS15645                            IDKQVAQVNTNKGSVVQVADIGLSSVTDIDKTEILQVGDSLELAPPIPVKFSMLSESD
gene_76_CEUT01009082                    LTKSTVTYSTNGANMVQISSFGLFQKPESITALEETNKS--------------SIRY--
gene_1_contig0005                       IKSSTIDLETKGGSYIQMSGLGFREKRTVDDKTK-------------------GIDY--
gene_66_Cellulophaga_ph_NC_021806       INSAAVKLKTNGGSFIQLSNFGLDKQTADAK----------------------GITW--
gene_34_Flavobact_ph_NC_031904          ITNKIMKHKMPGNGFVAGSESGFRMKENLEGVDKS------------------RIIY--
gene_64_Chlamydia_CVNZ01000007          FKKIVNKQQIKGGSAVQVSAMGIKGYEEDG-----------------------DLKY--
gene_49_KJ003983                        FKNAIQKQKINGGNIILVSNFGLSKDL--------------------------QTAY--
gene_34_775896848                       FKTNVHSVKLPGRELVQVAGPGKWKIGDEVR----------------------ELRH--
gene_80_FUFK010017431                   FNNGIFKQQQKGKELVQIAEVGGHIESS-------------------------ELKM--
gene_57_BCSF01000013                    FNNNVYTQKIKGKELVQVAEVGAYLDSQGQRK--------------------ELRM--
gene_107_816813760                      FKKNVINQTINGASLVQIAELGGVQEEKGTK----------------------ELGF--
gene_40_MDTC01014143                    FKNRVLKQRFNGMSVVQVAEFGYELDSQL------------------------EIRQ--
gene_44_935075730                       ITRRVTNLKLKGAHVTIQPDTFLQPAAVTLDKKGIIKGTQANVQRMYLEG--QIKFSD
gene_35_1001888980                      INKHTIRQEVAGITVPIMPNISWTYNTIENKARLKTMSAEEMKQHYNDG---AITY--
gene_57_934554352                       LSKQVINKYLPGFHAPIRADIFTASNELIKHNEFYSDKELYNKTIDELVANGSITYAS
gene_44_933950148                       FTNRVTRQVLPGFHASQVSDVGITALSGRTDLRDLMQSKVEEKHGYSLGR--KLTY--
gene_10_936104480                       FNNAITRQKLPGFHAAQVTNVGYSK----------------------------RLRY--
gene_59_935734019                       FNREITRQRMPGWHAAQVADFGFGGWKTTKDTATDD-----------------KLAY--
gene_43_932717476                       FNTNITRQLISGWHAAQLSDFGFKVDKQTQTDS--------------------KLQY--
gene_46_crAssphage_JQ995537             FNTNITRQLISGWHAAQLSDFGFKVDKQTQTDS--------------------KLQY--
CONSENSUS 0.8                           ----------G---------------------------------------------
2BE5 Thermus thermophilus (beta)
2BE5 Thermus thermophilus (beta')       RPLRSLTDILSGKQGRFRQNLLGKRVDYSGRS---------------------------
Jpred4 sec.str.  gene_46_crAssphage     --HHHHHHHH----HHHH-------------------------------------EE--


gene_82_Woesebacteria_MGFQ01000035      ISQVLSKGYTRRKLHGMRKGEDGTTPGEMVMPFSYMTRFGVKPSESVNDLFTIILNDG
gene_22_contig0002                      -------Y------------KDG-------------------------------------
gene_52_Azobact_ph_AP017903             ------------------MDEHG-------------------------------------
gene_25_842252479                       KDINPETLKGKSAEEIKRLKAAA---------------------------------------
gene_45_LULO01000006                    ------------YRKKDPNNPNS---------------------------------------
gene_30_Chitinophaga_FOJF01000001       LTDQEKSSVRLTSSDLKFYTRDA----------------------------------------
MODO_RS15645                            RNVLQSTFETEIEPNSTVYFNEE----------------------------------------
gene_76_CEUT01009082                    ---------LKDGKLKPPRIENG----------------------------------------
gene_1_contig0005                       -------IVDKEVLTGPKMNEDG----------------------------------------
gene_66_Cellulophaga_ph_NC_021806       -------LVEPSDLKPPVIEKDA----------------------------------------
gene_34_Flavobact_ph_NC_031904          ---LDSYNGKELQGTHTSTDENG----------------------------------------
gene_64_Chlamydia_CVNZ01000007          -----------------VTDGNG----------------------------------------
gene_49_KJ003983                        -----------------KVDKKG----------------------------------------
gene_34_775896848                       ---------------LDIDPKTG----------------------------------------
gene_80_FUFK010017431                   --------------------YDG----------------------------------------
gene_57_BCSF01000013                    ------------------YVEEG----------------------------------------
gene_107_816813760                      -----------------VRNKDG----------------------------------------
gene_40_MDTC01014143                    -------------------HANG----------------------------------------
gene_44_935075730                       DYWQSRAELNEDGTIKRDANGTP-----------------------------------------
gene_35_1001888980                      ---------TPAFLERAMNRKEG----------------------------------------
gene_57_934554352                       SFIEECKRTGRSLELQAEYREDG-----------------------------------------
gene_44_933950148                       -------------------HKDG----------------------------------------
gene_10_936104480                       -------------------HPDG----------------------------------------
gene_59_935734019                       -------------RKIGKYNDND----------------------------------------
gene_43_932717476                       ---------------KKIGEVDG----------------------------------------
gene_46_crAssphage_JQ995537             ---------------KKIGEVDG----------------------------------------
CONSENSUS 0.8                           ----------------------------------------------------------------
2BE5 Thermus thermophilus (beta)
2BE5 Thermus thermophilus (beta')       ------------------VIVVG----------------------------------------
Jpred4 sec.str.  gene 46 crAssphage     ---------------EE-----------------------------------------------
```


gene_82_Woesebacteria_MGFQ01000035
gene_22_contig0002
gene_52_Azobact_ph_AP017903
gene_25_842252479
gene_45_LULO01000006
gene_30_Chitinophaga_FOJF01000001
MODO_RS15645
gene_76_CEUT01009082
gene_1_contig0005
gene_66_Cellulophaga_ph_NC_021806
gene_34_Flavobact_ph_NC_031904
gene_64_Chlamydia_CVNZ01000007
gene_49_KJ003983
gene_34_775896848
gene_80_FUFK010017431
gene_57_BCSF01000013
gene_107_816813760
gene_40_MDTC01014143
gene_44_935075730
gene_35_1001888980
gene_57_934554352
gene_44_933950148
gene_10_936104480
gene_59_935734019
gene_43_932717476
gene_46_crAssphage_JQ995537
CONSENSUS 0.8
2BE5 Thermus thermophilus (beta)
2BE5 Thermus thermophilus (beta')
Jpred4 sec.str. gene_46_crAssphage
gene_82_Woesebacteria_MGFQ01000035
gene_22_contig0002
gene_52_Azobact_ph_AP017903
gene_25_842252479
gene_45_LULO01000006
gene_30_Chitinophaga_FOJF01000001
MODO_RS15645
gene_76_CEUT01009082
gene_1_contig0005
gene_66_Cellulophaga_ph_NC_021806
gene_34_Flavobact_ph_NC_031904
gene_64_Chlamydia_CVNZ01000007
gene_49_KJ003983
gene_34_775896848
gene_80_FUFK010017431
gene_57_BCSF01000013
gene_107_816813760
gene_40_MDTC01014143
gene_44_935075730
gene_35_1001888980
gene_57_934554352
gene_44_933950148
gene_10_936104480
gene_59_935734019
gene_43_932717476
gene_46_crAssphage_JQ995537
CONSENSUS 0.8
2BE5 Thermus thermophilus (beta)
2BE5 Thermus thermophilus (beta')
Jpred4 sec.str. gene 46 crAssphage

```
gene_82_Woesebacteria_MGFQ01000035      SEQTVYDDTDFDIDQRSVYFYA
gene_22_contig0002                      AAYIKASNADFDKDMLFTFKKF
gene_52_Azobact_ph_AP017903             HIMTSILGDDNDGDKLFFAMPY
gene_25_842252479                       SEIVAKSGADFDIDKLTLYFPN
gene_45_LULO01000006                    SELVVKAGSDFDIDKLTLYMPS
gene_30_Chitinophaga_FOJF01000001       SEIVLKAGSDFDIDKLNTYLKN
MODO_RS15645                            HDITNKTGSNFEIDKLYLMFPS
gene_76_CEUT01009082                    EEMTAKTGSDFDIDKLFMLIPN
gene_1_contig0005                       NDLTAKTGADFDIDKLYIMLAE
gene_66_Cellulophaga_ph_NC_021806       TEIPTKTGSDFDIDKMYVMLPN
gene_34_Flavobact_ph_NC_031904          KNFTKQKGLDYDIDKESAYQLN
gene_64_Chlamydia_CVNZ01000007          AQGTTIAGFDFDVDKLYFMRRE
gene_49_KJ003983                        SDIVEMSGTDFDIDKLFLMIKN
gene_34_775896848                       GNITIQTGSDFDIDKLFALFRN
gene_80_FUFK010017431                   GGITTQMGSDFDIDKMYVIQKE
gene_57_BCSF01000013                    GGITVQMGSDFDVDKMMVMQKE
gene_107_816813760                      GEITTQMGTDFDMDKVFLMMPN
gene_40_MDTC01014143                    AEITTMMGSDFDIDKMYLMFPE
gene_44_935075730                       EHLVTRTGWDYDIDSIYLSMKE
gene_35_1001888980                      YDLVLQTGWDFDIDKVYAYRKN
gene_57_934554352                       QSLITRTGWDFDIDSIYAYYRN
gene_44_933950148                       DEWVLQTGADFDIDSIYGIYHT
gene_10_936104480                       DDWVAQTGSDFDIDSVYGIQYN
gene_59_935734019                       ENWVHQSGSDYDVDSVYAMTFG
gene_43_932717476                       DEWVTQTGSDFDVDSVYGMSKT
gene_46_crAssphage_JQ995537             DEWVTQTGSDFDVDSVYGMSKT
CONSENSUS 0.8                           -------G-DFDID--------
2BE5 Thermus thermophilus (beta)
2BE5 Thermus thermophilus (beta')       PLVCEAFNADFDGDQMAVHVPL
Jpred4 sec.str.  gene_46_crAssphage     ---------------EEE----
```

HHpred output

>**pfam00623** RNA_pol_Rpb1_2 RNA polymerase Rpb1, domain 2. RNA polymerases catalyze the DNA dependent polymerization of RNA. Prokaryotes contain a single RNA polymerase compared to three in eukaryotes (not including mitochondrial. and chloroplast polymerases). This domain, domain 2, contains the active site. The invariant motif -NADFDGD- binds the active site magnesium ion.
Probab=46.30 E-value=37 Score=35.72 Aligned_cols=47 Identities=26% Similarity=0.279 Sum_probs=0.0

```
Q ss_pred            EcCCCCCCceeeEEEEEecCccCCCeEEechhHhcccCCCCchhheeEeec
Q AUXO017664536   631 RIPTEGKQSVIIMRVVDFLPNVSDSTVVLPENWVHQSGSDYDVDSVYAMTF  681 (697)
Q Consensus       631 RIPtq~~~S~~~i~IvgFlP~~~Gd~IiVP~eiv~q~GsDFDiDKly~~~~  681 (697)
                      |=||=-+.|+.+.++ .+.+.   .+|-++.-+.+-.+.|||.|.|.++.|
T Consensus       105 R~PsLh~~si~~~~~~i~~~~~~~i~~~~~c~~~naDFDGD~m~~~~  151 (166)
T pfam00623       105 RQPSLHRLSIMGHRV-RVLPG---KTFRLNLSVTTPYNADFDGDEMNLHVP  151 (166)
T ss_pred            CCccccccccccee-EEecC---CceeECcccCCCcCCCcCceeEeecC
```

Supplementary Note 1

Supporting evidence for remote sequence similarities used for crAssphage and IAS phage gene annotation

Query: gene_2_crAssphage_JQ995537

HHpred output:

```
>gene_2_crAssphage_JQ995537 Uncultured phage crAssphage, complete genome 249_aa|+|454|1203 crAssphage_JQ995537
Uncultured phage crAssphage, complete genome
MPDVKDLAQAAQGAAAPESVNVVEPTTSVNQPVQPTVDTDNQSSAQVETIDDVVRRICTDGHSYVMTTVITNIDCQERTGRNGKSYLNAFVTIASPVKGAQSMPDGTHRMGMLGAVQMPF
NQILLVMRKDKFYGRFVNYVGEAAEAGFASMYLTGVAVKVLCQFVPAGVQDHNPFTRKDNLYNVVDYDRYVYHIVGIEQPADPILVGAYNVLIKQIMEDA
RAAIAAKREAKAKAASFVATAMSDDDLPF (characteristic SSB motifs and features are highlighted)

>PRK07275 single-stranded DNA-binding protein; Provisional
  Probab=42.04  E-value=23  Score =30.23  Aligned_cols=7  Identities=100%  Similarity=1.659  Sum_probs=0.0

  Q ss_pred             ccCCCCC
  Q gene_2_crAssph  243 SDDDLPF  249 (249)
  Q Consensus       243 S~~D~PF  249 (249)
                        |||||||
  T Consensus       156 ~~~~~pf  162 (162)
  T PRK07275        156 SDDDLPF  162 (162)
  T ss_pred             ccCCCCC


////////////////////////////////////////////////////////////////////////////////////////////////////////////////////
```

Query: gene_15_crAssphage_JQ995537

CDD output:

```
pfam01443:Viral (Superfamily 1) RNA helicase; Helicase activity for this family has been demonstrated and NTPase
activity. This helicase has multiple roles at different stages of viral RNA replication, as dissected by mutational
analysis.

Pssm-ID: 307550  Cd Length: 226  Bit Score: 33.50  E-value: 0.18

                        10        20        30        40        50        60        70        80
                ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
3WRX_C       165 LVDGVPGCGKTKEIL----------------------SRVnFeedl--ilvpGRQaAEMIRRRAnas-giIVATKDNVR 218
query         46 alcGAGGTGKTFVIKyvinnckwsggvigcaapthkacRVLsNsi-------gGKEvNTIQSL-F------GFRLDVNIE 111

                        90       100       110       120       130       140       150       160
                ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
3WRX_C       219 TVDSFLMNYGKGarcQF------KRLFIDEGLMLHtgcvNFlv---emSLCDIAYVYGDTQQIPYInrv--tgfPYPAHF 287
query        112 NFDPENPAFNPV---GKdkldglKVLIIDEASMLNaklvKYisnkckkLQIK-VIMLGDSSQLPPVnektsqafLIASNT 187
```

```
                    170       180       190       200       210       220       230       240
                .....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
3WRX_C      288 AKLe--vDEVETR----RTTLRCPADVTHFLNQR----------YEGhVMCTSSEKksvsqemVs---gAASINPVsk-- 346
query       188 YYL----KEVVRQgdnnPISKLLKLLREDIDNKN----------Gwkfldyisknrqdyneetkgfyvcgqtefsdlidt 253

                    250       260       270       280       290       300       310       320
                .....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
3WRX_C      347 plKGKILTFTQSDKEaLLSRgy----------adVHTVHEVQGETYADVSLVRLtpTPvsIIARDSPHVLVSLSRHTKSL 416
query       254 cfndeeytknidlyriiaytnsrvaqwnnhvrhmiiqdadkslitrndlimsyttvvnvfndiiinnseeyivkdivdti 333


                ....
3WRX_C      417 KYYT 420
query       334 dndy 337

//////////////////////////////////////////////////////////////////////////////////////////////////////////////////////
```

Query: alignment of 21 closest homologs of gene_18_crAssphage_JQ995537
HHpred output:

```
5HZR_A SNF2-family ATP dependent chromatin remodeling; Swi2/Snf2, chromatin remodeling, TRANSCRIPTION; HET: SO4, KH2;
2.33A {Myceliophthora thermophila ATCC 42464}
Probab=100.00  E-value=6.6e-53  Score=401.36  Aligned_cols=470  Identities=8%  Similarity=-0.103  Sum_probs=0.0
Template_Neff=10.000

Q ss_pred              CchHHHHHHHHHHHH---cCCCeEEEeCCChhHHHHHHHHHHHHHHhCCCCCeEEEEeCcccHHHHHHHHHcCCCCcEE
Q gene_15_CDZN01     3 DVTDIYNEAANKWSD---NKGVGSVILSEPLSVMNFVTMVLDKMVAKTPNLTSLIITETMEDRANITYYLDNTSELKDIH   79 (488)
Q Consensus          3 ~L~~hQ~~Gv~~~~~~~~~~ggILADeMGLGKTiq~i~~i~~~~~~~~~~~~LIv~p~sl~~~w~~e~~~~~~~~~~~   79 (488)
                      +|+|||.+|+.||..   ++.+|||||+||+|||+|+++++..+.......+|+|||||.+++.+|..++.+|.+...++
T Consensus       109 ~L~~~Q~~~v~~l~~~~~~~~iLad~~GlGKT~~~i~~i~~l~~~~~~~~~~LIv~P~~~l~~w~~e~~~~~~~~~~~  188 (732)
T 5HZR_A          109 TLKEYQLKGLQWMLSLYNNNLNGILADEMGLGKTIQTISLITYLIEKKHQQGPYLVIVPLSTLTNWNLEFDKWAPSVAKV  188 (732)
T ss_dssp             CCCHHHHHHHHHHHHHHHTTCCEEECCCTTSCHHHHHHHHHHHHHHHHHCCCCCEEEECCHHHHHHHHHHHHHCTTSCEE
T ss_pred             cccHHHHHHHHHHHHHHhcCCCeEEEcCCCCCHHHHHHHHHHHHHHHHHhccCCCEEEEEeHHHHHHHHHHHHHCCCCcEE


Q ss_pred              EeCCCcHHHHHHhHhhhhcCCCCCCCCceeeeEeeccHhhhhcchhHhccchHHHHhccccccccccchhhHHHHhcccc
Q gene_15_CDZN01    80 KQLITDKKCLILTREYVERSPYKPSPSSHKDVLITINVKKFRKIAEKYSGNYFKFKLLATNAIDSVADNAVLMYKYAPKV  159 (488)
Q Consensus         80 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~  159 (488)
                      .+.+.....................++++.   ........+....+.++++|+.|++.+.................++++
T Consensus       189 ~~~~~~~~~~~~~~~~~~~~vvi~ty~~~~l~~~~~l~~~~~~iIiDEaH~lkn~~~~~~~~l~~~~~~~~llL  265 (732)
T 5HZR_A          189 VYKGPPNARKMQQEKIRQGKFQVLLTTYE---YIIKDRPLLSKIKWFHMIIDEGHRMKNANSKLSATIQQYYSTRFRLIL  265 (732)
T ss_dssp             ECCSCHHHHHHHHHHHHTTCCSEEEECHH---HHHHTHHHHTTSEEEEEEEECTTTC----CHHHHHHHHCEEEEEEEE
T ss_pred             EEeCCHHHHHHHHHHHcCCceEEEEeHH---HHHhhhHHHHhCCCcEEEEeCHHHhccccHHHHHHHHccCCeEEEe
```

```
Q ss_pred              ccccccccchhHHhcCChhhhcCCccccCCchhhccccchhcCcchhcccCcchHHHHHhhCHHHHHhHHhHHHhhhhcCC
Q gene_15_CDZN01  160 YEINYAHLINRSIHSPIKEYQKGVVLTDADRIYYDRCSQYINESVTIFGSFDKLEECRTGNPRLNIAAETCRLQVAESNG  239 (488)
Q Consensus       160 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~  239 (488)
                        +|||.+++..+++++..++.+..+.....+..+.................................+...++........
T Consensus       266 TgTP~~n~~~el~~ll~~l~p~~~~~~~~~f~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~l~~~l~~~~lrr~k~~~~~~~  345 (732)
T 5HZR_A          266 TGTPLQNNLAELWAMLNFVLPNIFKSAKTFDEWFNTPFANTGGQDKMELTEEEQILVIRRLHKVLRPFLLRRLKKDVEKD  345 (732)
T ss_dssp             CSSCCCSCHHHHHHHHHHHSTTTTCCCSCHHHHHTGGGC----CCCCCCCHHHHHHHHHHHHHHHGGGEECCCGGGSGGG
T ss_pred             cCCCcccCCHHHHHHHHHhhCCccccCCHHHHHHHhcChhHcCCCCcccccCCHHHHHHHHHHHHHHHHHHHhhchhcHHHHHhh


Q ss_pred              CcccccccHHHHHHHHHHhCchhHhhhhhhhhhhhhHHHHHHHHHHHHHHHHHHHHHhhcchhhchhccCcccccccchhh
Q gene_15_CDZN01  240 WSAKMDMTDAMCRKIDELYNPSSLIERVNQTYNIIRERTKIITDNIVKLDAILDIVKENIGKRILIISKNGVFASKVTEY  319 (488)
Q Consensus       240 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~  319 (488)
                        .+..............+...|.........   ....................++..|.|+............
T Consensus       346 lp~~~~~~~~v~~~~~~~~~q~~~y~~~~~~~~~~~~---~~~~~~~~~~~~~~~~~~~~~~~~~~~~l~~~~~~~p~l~~~~~~~~~~~~~------  416 (732)
T 5HZR_A          346 LPDKTEKVIKCKFSALQARLYKQMVTHQKI---AVSDANGGKTGARGLSNMIMQLRKLCNHPFVFDEVENQMNP------  416 (732)
T ss_dssp             CCSEEEEEEEECCCHHHHHHHHHHHHHHHH---C----------CTHHHHHHHHHHHHHCGGGSHHHHHHHCT------
T ss_pred             CCCCEEEEEECCCCHHHHHHHHHHHHHHHHHH---HHhccCCCCCcchhHHHHHHHHHHHhCChhHhcHHHHhcCc------


Q ss_pred              hcccccccccchhHHHHHHHHHHHHhcCCCcEEEEECCccHHHHHHhcchHhCCCCceEecCCccHHHHHHHHHHHhhcCC
Q gene_15_CDZN01  320 LNANIKYEGKSIMTNGEIFQTGISILQYDYCGNYHNDMEGIQAYDKNGKPKVYKSGAKVGQPVIIKAQAQRTRNLELFND  399 (488)
Q Consensus       320 ~~~~~~~~~~~~~~~s~k~~~~l~~~~l~~~~~~~~~~~~~~k~~ifs~~~~~~~l~~l~~~~l~~~~~~~~~~~~~~~~~~g~~~~~~r~~~~~~~~~~~~~~~  399 (488)
                        ...........++|...+..++..+...+++++||+++...++.+.+++...+..+.+++|++...++...+..|...++
T Consensus       417 ~~~~~~~~~~~~~~~s~K~~~L~~ll~~~~~~~~~~~~~kvlIF~~~~~~~~~~~~l~~~L~~~g~~~~~~l~g~~~~~~~R~~~i~~F~~~~~~  496 (732)
T 5HZR_A          417 ANVSNDLLWRTAGKFELLDRILPKYKATGHRVLMFFQMTAIMDIMEDFLRFRGLHYLRLDGTTKSEDRSELLRQFNQPDS  496 (732)
T ss_dssp             TCCCSTHHHHSCHHHHHHHHHHHHHHHHTTTCCEEEEESCHHHHHHHHHHHHHHHTCCEEEECTTSCHHHHHHHHHHHTSTTC
T ss_pred             ccCChhHHHhhhHHHHHHHHHHHHHHHHHHHHcCCcEEEEcCcHHHHHHHHHHHHHHHHHCCCCEEEEeCCCCHHHHHHHHHHHHcCCCC


Q ss_pred              CceEEEEeCCcCccCcc-cCCCEEEEECCCCCHHHHHH--HhhhhccCCCCCeEEEEEEeCCcHHHHHHHHHhcchHHHH
Q gene_15_CDZN01  400 DYMKVLSANNSIDTSFK-GVVDVVIFTSPLCSSIRDLK--YRIPNLSFSSVPNIIYKIYCRGTNEEKKLIETKGGKDYEI  476 (488)
Q Consensus       400 ~~~v~L~s~~agg~Gln-~~A~~vi~~D~~WNP~~~~q--~R~hRi~GQ~k~V~v~rli~~~TiEe~i~~~~~~~~K~~~~~~  476 (488)
                        ++.+||+|+++||+|+| .+|++||++|++|||+.+.|  +|+||+ ||+++|.||+|++.+|+||+|+.++..|+.+..
T Consensus       497 ~~~v~List~a~~~GlnL~~a~~VI~~d~~wn~~~~~Qa~gR~~R~-Gq~~~v~v~~l~~~~tiee~i~~~~~~~~K~~~~~~  575 (732)
T 5HZR_A          497 PYFMFLLSTRAGGLGLNLQTADTVIIYDSDWNPHQDLQAQDRAHRI-GQKNEVRILRLISSASVEEKILERARFKLDMDG  575 (732)
T ss_dssp             CCCEEEEESCC----CCCTTCCEEEECCCC---------------C-CSSSCEEEEEEEETTSHHHHHTC----------
T ss_pred             CcEEEEEeCccccCCCCccCCCEEEEeCCCCCHHHHHHHHHHHHHHh-CCCCcEEEEEEEeCCCHHHHHHHHHHHHHHHHHcc


Q ss_pred              Hhhccccc
Q gene_15_CDZN01  477 VKDSEIDFI  485 (488)
Q Consensus       477 ~~~~~~~~~~~  485 (488)
                        .....+...
T Consensus       576 ~~~~~~~~~~~  584 (732)
```

```
T 5HZR_A          576 KVIQAGRFD  584 (732)
T ss_dssp             ---------
T ss_pred             ccccCCCCC
```

Query: gene_18_933950148 close homolog of gene_21_crAssphage_JQ995537
HHpred output:

```
>TIGR00621 ssb single stranded DNA-binding protein (ssb). All proteins in this family for which functions are
  known are single-stranded DNA binding proteins that function in many processes including transcription, repair,
  replication and recombination. Members encoded between genes for ribosomal proteins S6 and S18 should be
  annotated as primosomal protein N (PriB). Forms in gamma-protoeobacteria are much shorter and poorly recognized by
this
  model. Additional members of this family include phage proteins. Eukaryotic members are organellar proteins.
  [DNA metabolism, DNA replication, recombination, and repair]
  Probab=65.26  E-value=46  Score=30.54  Aligned_cols=107  Identities=21%  Similarity=0.225  Sum_probs=0.0

  Q ss_pred             eeEehHHHHHHhhc---cCccccCCCCCCccеecccccC-cceeEEeecccCCccCCCccccCcceеEecccchhhc
  Q gene_18_933950  213 QRQFWFRLNRYYKN---KGDWAFSGQGSEEGDLVFPNIVG-QGIFEEKFMLDATHFKEPSLMFDITKERIAPMDGVQSKQ  288 (347)
  Q Consensus       213 pi~~WmKLlr~~kn---Kg~w~~~gs~~~~gdl~fp~FVg-~GviE~~~~~~ap~i~~~~l~id~~ke~i~~kd~~k~~k  288 (347)
                        .+.+|=+|......   ||++-++     .|.|....+.. +|.-.....-.+-+|.             ++...+++
  T Consensus        54 ~v~~wg~~Ae~i~~~l~KG~~V~V-----G~L~~~~~~~~~~G~~~~~~~~I~a~~i~------------~l~~~~~~~~---  113 (164)
  T TIGR00621        54 DIVIFGRLAEVAAQYLKKGSLVYV-----EGRLRTRKWEDQNGQKRSKTEIIADNVQ------------LLDLLGAR---  113 (164)
  T ss_pred             EEEEEhHHHHHHHHhccCCCEEEE-----EEEEEeeEEECCCCcEEEEEEEEEEEEE------------EccccCcc---


  Q ss_pred             cCCcCcccCCCCCccccccccCCCCCCCCcccccCCCcCCCCcccccCCCCCCCCCCCC
  Q gene_18_933950  289 KKAPNLAAAPGIGGIAMGAGIVNPSMPMGGFAGGVAGGFVSTESSAFAPDSEDNGGLPF  347 (347)
  Q Consensus       289 ~~~~~~~~~~~~m~g~~~~a~~~np~~P~gg~~gG~a~ap~~~~~~~a~~~~~n~~d~lpf  347 (347)
                              +-++..-.|..++.   +|+.+......|++++.....+....++.-..|++||
  T Consensus       114 -----~~~~~~~~~~~~~~~~~---~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~d~~pf  164 (164)
  T TIGR00621       114 -----AQGGEFSGGGGSGN---RPQGGQEQSATGKDPAQSAPQSGSADPIDGDDDEIPF  164 (164)
  T ss_pred             -----ccccccCCCCCcCC---CCCCCcccccCCCCcccccCcccccCCCCCCCCccCCCCC
```
//

Query: gene_19_crAssphage_JQ995537 (RecT)
HHpred output:

```
>TIGR01913 bet_lambda phage recombination protein Bet. This model represents the phage recombination protein Bet
  from a number of phage, including phage lambda. All members of this family are found in phage genomes or in
  putative prophage regions of bacterial genomes. [Mobile and extrachromosomal element functions, Prophage
```

```
  functions]
  Probab=98.11  E-value=2.4e-06  Score=76.45  Aligned_cols=38  Identities=21%  Similarity=0.205  Sum_probs=0.0

  Q ss_pred             cchhHHHHHHHHHHHHHHHHHHHHHHHHccCccchhH
  Q gene_16_933950  251 KDTYTKYMRTLIGHRAFTLAARDIAADVILGCMETTEA  288 (304)
  Q Consensus       251 Kd~W~~ypk~ml~~RA~~~aaRdiasD~L~Gi~~~eEl  288 (304)
                        +.+|+++|.+||+|+|.+.|+|..|||-|-|+|..||+
  T Consensus       142 ~~~W~~~P~~Ml~K~Ae~~ALR~AFp~~l~G~y~~eEm  179 (180)
  T TIGR01913       142 KSTWQSMPKTMLRKCAEAQALREAFPEGFGGIYTEEEM  179 (180)
  T ss_pred             cchhhhCcHHHHHHHHHHHHHHHhCccccccccchhhc

  No 2  HistogramsShow template alignmentPfamCDDPubMed
>pfam03837 RecT RecT family. The DNA single-strand annealing proteins (SSAPs), such as RecT, Red-beta, ERF and
  Rad52, function in RecA-dependent and RecA-independent DNA recombination pathways. This family includes proteins
  related to RecT.
  Probab=97.49  E-value=0.00017  Score=63.05  Aligned_cols=68  Identities=15%  Similarity=0.059  Sum_probs=0.0

  Q ss_pred             cCeEEEEEEEEeHHHHHHccCCCc----chhHHHHHHHHHHHHHHHHHHHHHHHHHccCccchhHHHhcCCCC
  Q gene_16_933950  228 DNRVITSIGHFSYSDAVTAGLASK----DTYTKYMRTLIGHRAFTLAARDIAADVILGCMETTEAKIVNNMSI  296 (304)
  Q Consensus       228 ~~~e~~~~G~FS~~dA~~Agl~gK----d~W~~ypk~ml~~RA~~~aaRdiasD~L~Gi~~~eEl~d~~~~~~  296 (304)
                        ++.+.. +..++..+.......++    .+|++||..|+.|+|+..|+|..+|+.+.|+|..+|+.+.....
  T Consensus       119 ~~~~~~~~~~~~~~e~~~~~~~~~~~~~~W~~~p~~M~~K~a~~~alRk~~p~~~~g~y~~dE~~~~~~~~~  190 (193)
  T pfam03837       119 DGGEPF-VVEMSLEEIEKRRSKSSKTNSGPWAKDPEEMAKKTALKRALRKAFPLSIEGAYTEDEMGDIEEEKE  190 (193)
  T ss_pred             CCCEEE-EEEeEHHHHHHHHHHhcCCCCCccCCHHHHHHHHHHHHHHHHHCCcchhhhhhhhhccccccccc

//////////////////////////////////////////////////////////////////////////////////////////////////////////////////////
```

Query: gene_34_crAssphage_JQ995537
HHpred output:

```
>pfam06067 DUF932 Domain of unknown function (DUF932). Family of prokaryotic proteins with unknown function.
  Contains a number of highly conserved polar residues that could suggest an enzymatic activity.
  Probab=100.00  E-value=7e-41  Score=302.88  Aligned_cols=192  Identities=33%  Similarity=0.522  Sum_probs=0.0

  Q ss_pred             EeEEEcCCeEEcChHHHHHHHHHhhcCCcEEEEEEEecCCcEEEEEEECCCceEecCCceEeEEEEEeCCCCCceeEEE
  Q Thu_Apr_27_21:   92 PLGVVKSKYTIVQNNDAFNFFDDAIGKNSAIWQTAGFWGNGERIFVSAKLPNNILVKGDPVENYLVFTNTHDGSGGVKIL  171 (353)
  Q Consensus        92 ~lgvvs~~Y~~iqn~e~~~~~~~l~~~~g~~~etag~l~~Gr~~f~~~~lp~~~~i~gD~i~~~l~l~NShDGs~a~~i~  171 (353)
                        +||+||++|+||||.+++++++.+++ .|+.++++|.+++|+++|+++++|+.+.++||.+.+||++|||||++++++.
  T Consensus         1 ~lg~vs~~Y~~i~~~~~~~l~~~~~~g~~~~t~g~l~~g~~~~~~~~~~~~~v~gd~~~~~l~~~nShdGs~a~~i~   79 (196)
  T pfam06067         1 ALGVVSDRYTVVQTIEVLEFLRALVG-SGFEYETAGRLRGGRRVWALHRLPDQITVKGDRVRPYLLLLNSHDGTSAYTAL   79 (196)
  T ss_pred             CCceeCCCceEECHHHHHHHHHHHhh-cCCeEEEEeecCCCceeEEEEcCCceeecCceEEEEEEEeccCCCcceEEE
```

```
Q ss_pred              EeceEEEecCceeEEeeCCCceEEEeeccchhHHHHHHHHHHHHhhHHhHHHHHHHHHHHHhCCCCCHHHHHHHHHHHccCc
Q Thu_Apr_27_21:  172 FTPIRVICQNTLNAAIRTSSNYVSFRHTTSVHNKISVAQEILGISKIKSEEFGQYCNLLANIKVTDEDVIQFIGENLLTS  251 (353)
Q Consensus       172 ~t~~RvVC~Ntl~~a~~~~~~~~~irHt~~~~~l~~a~~~l~~~~~~d~~~~~~~l~~~~lt~~e~~~~i~~l~~~~  251 (353)
                      ++++|+||+|||.++..+ ...++++|+++++.++++++|+++.+++++|.+.+++|++++|+++++.+|++
T Consensus        80 ~~~~R~VC~N~l~~~~~-~~~~~i~Ht~~~~~l~ea~~al~~~~~~~~~~~~~~~l~~~~lt~~~~~~~~~------  152 (196)
T pfam06067        80 ATPFRVVCANTLVVALGA-FGSVRVRHTGNAVAKVIEAAYALGLVLSYFDRFEEEAEALAAVTLTDREQEAFAE------  152 (196)
T ss_pred              eeeEEEEeeCccHHHhCC-CCEEEEeCCCchhhHHHHHHHHHHHHHHHHHHHHHHHHHHHhcCCCCHHHHHHHh------


Q ss_pred              hHHHHHhcCCCchhhhhHhccccccccchhHHHHHHHHHHHHHhhcCCCccCCcccHHHHHHHHHHHhccccccCChhHH
Q Thu_Apr_27_21:  252 DEIQRLKDTGHTIKDIAYRSGLALTDSKISSRKMNVISDTYSYYFDGPGQRDILGTAWGAVNAISGYYSNIDNIEGTKRF  331 (353)
Q Consensus       252 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~l~~l~~~~~~~~~~~~~~~T~w~~ynAvte~l~h~~~~~~~R~  331 (353)
                      ..                                                  +.+..++||+|++|||||||++|.++.++..++
T Consensus       153 ~~-------------------------------------------------~~~~~~gt~w~~~nav~e~~~h~~~~r~~~~~  187 (196)
T pfam06067       153 RG-----------------------------------------------ADLAGVRGTAWGALNAVTEYLDHGGRRRPVEDQ  187 (196)
T ss_pred              cC-----------------------------------------------ccccccccCchHHHHHHHHHHHHhcCcccCcchh


Q ss_pred              HHHhh
Q Thu_Apr_27_21:  332 DSICY  336 (353)
Q Consensus       332 ~~~l~  336 (353)
                      +..+.
T Consensus       188 ~~~~~  192 (196)
T pfam06067       188 DVKLN  192 (196)
T ss_pred              hHHHH
```

>pfam12987 DUF3871 Domain of unknown function, B. Theta Gene description (DUF3871). Based on Bacteroides thetaiotaomicron gene BT_2984, a putative uncharacterized protein As seen in gene expression experiments (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE2231). It appears to be upregulated in the presence of host or other bacterial species vs when in culture.
Probab=94.27 E-value=0.64 Score=46.11 Aligned_cols=163 Identities=20% Similarity=0.215 Sum_probs=0.0

```
Q ss_pred              cEEEEEEECCC-ceEecCCceEeEEEEEeCCC--------CCceeEEEEeceEEEecCceeEEeeCCCceEEEeeccchh
Q Thu_Apr_27_21:  133 ERIFVSAKLPN-NILVKGDPVENYLVFTNTHD--------GSGGVKILFTPIRVICQNTLNAAIRTSSNYVSFRHTTSVH  203 (353)
Q Consensus       133 r~~f~~~~lp~-~~~i~gD~i~~~l~l~NShD--------Gs~a~~i~~t~~RvVC~Ntl~~a~~~~~~~~~irHt~~~~  203 (353)
                      .|+-..+.+|. .-.|.|..+...+.-.=|++        +-..|++.+|.--.||+|.+....+- ...+++.-+....
T Consensus       108 eRmaF~ieIPsI~~~InGn~L~LtIGGVRaYn~eNLYskK~~EkFKvFIGF~n~VC~NlcvstDG~-~~~l~V~s~~eL~  186 (318)
T pfam12987       108 ERMAFAIEIPSIVETVNGNRLNLCIGGVRAYNEENLYSKKSPEKFKIFIGFRNRVCSNLCLTTDGF-KDDIEVMSTAELY  186 (318)
T ss_pred              HHHHHheecCceeEeecCCEEEEEEeeeeccchhhhhcccchhheeeeeeeehhhccceeeecCCc-cceEEECCHHHHH


Q ss_pred              HHHHHHHHHHHhhHHhHHHHHHHHHHHhCCCCCHHHHHHHHHH-----HccCchHHHHHhcCCChhhhhHhccccccccc
Q Thu_Apr_27_21:  204 NKISVAQEILGISKIKSEEFGQYCNLLANIKVTDEDVIQFIGE-----NLLTSDEIQRLKDTGHTIKDIAYRSGLALTDS  278 (353)
```

```
Q Consensus       204 ~~l~~a~~~l~~~~~~~d~~~~~~~~l~~~~lt~~e~~~~i~~-----l~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~  278 (353)
                      ..+-+.-....     .++-...|+.|.+..||+.+|-+++.+     .+|...- ...+             .-..+++
T Consensus       187 ~~~leLf~~yn-------~khl~~M~~l~~~~ltE~QFaQliGr~RlYq~LP~~~k-k~lP-------------~ll~~Ds  247 (318)
T pfam12987       187 RAALELFNNFN-----PAKELHLLQTLGNTSITEHQFCQIIGRMRLYQALPAGYQ-KRLP-------------RLLLTDS  247 (318)
T ss_pred             HHHHHHHHHcC-----HHHHHHHHHHhhcCcccHHHHHHHHHHHHHHHhhCCHHHH-hcCC-------------ccccChH


Q ss_pred             chhHHHHHHHHHHHHHhhcCCCccCCcccHHHHHHHHHHHHc
Q Thu_Apr_27_21:  279 KISSRKMNVISDTYSYYFDGPGQRDILGTAWGAVNAISGYYSN  321 (353)
Q Consensus       279 ~~~~~~~~~l~~l~~~~~~~~~~~~~~T~w~~ynAvte~l~h  321 (353)
                      ....-.++....      +.=......=|+|-+||-+|+.-+.
T Consensus       248 qiN~vak~Yy~D-------~nF~~~~~I~lW~~yNL~T~AnKs  284 (318)
T pfam12987       248 QINAVVKAYISD------ENFGSNGNDITLWKLYNLLTGANKS  284 (318)
T ss_pred             HHHHHHHHHhcC------ccccCCCCcccHHHHHHHhcccchh


////////////////////////////////////////////////////////////////////////////////////////////////////
```

Query: gene_46_934554352 same family as gene_55_KJ003983 IAS virus (ThyX Thymidylate synthase)

HHpred output:

```
>COG1351 ThyX Thymidylate synthase ThyX [Nucleotide transport and metabolism].
  Probab=99.97  E-value=1.2e-30  Score=237.61  Aligned_cols=163  Identities=20%  Similarity=0.215  Sum_probs=0.0

Q ss_pred             ceeCCcEEEccc--cHHHHHHHHHhhhhcCcccCCc--chHHHHHHHHHcC-CCCccccCeEEEEeccCCHHHHHHHh
Q Mon_May_01_23:    2 NVVEPAIEFCDY--SGLRKIELIGKVCTKQEDSIKP--DSAETFCRNRLIDG-HTAIFEHEYVYFNVTSIPNRIVREFVK   76 (279)
Q Consensus         2 kii~p~~ei~~~~~~~~~iE~agR~CYkSe~ki~~~~s~~~Fi~~l~~~g-H~SvlEH~~~~f~i~~~~~~~~~~~~~   76 (279)
                      ++..|-+.+..+  ..+..|+.|+|+||.|+.+-.+  ..++.|+++++..| |.|+|||+++||.|.|
T Consensus        17 ~~~~~~~~~~~~~~~~i~~a~r~~~~s~~~~~~~~~d~~~~~~~~~~h~s~~Eh~~~tF~I~~~~~~~~~~   85 (273)
T COG1351          17 KILDKGVVITIDSRGPLALIVQAARVSYPSGEKLEDGGQKDAELIRRIINEFGHESPLEHLVATFEIEG-----------   85 (273)
T ss_pred             ccCCCcceEeeccccchHHHHHHHHHhccccccccCCcccHHHHHHHHHccCcCChhhhceeEEEEec-----------


Q ss_pred             cCCeeeecccCCeeeeeHHHHHHHHHcChhhhHHHccCcCccccChhHHHHHHhchhccccccCccccchhhcccCcccce
Q Mon_May_01_23:   77 LSPYIRWSYLGNYIGFSYRVFLDIMSNSRKMKAIYNDIYHPSEVNDLFYNMLLLSKEFSHLLFDDKDITAKLEYGIDVSL  156 (279)
Q Consensus        77 ~~~~~~~~~~~lv~~nlR~f~e~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~  156 (279)

T Consensus        86 ----------------------------------------------------------------------------   85 (273)
T COG1351          86 ----------------------------------------------------------------------------   85 (273)
T ss_pred             ----------------------------------------------------------------------------


Q ss_pred             eeccchhHhhcCccccccEEEEEEechHHHHHHHHhccCCccccceeeeeeeccccccccCCC-----------CcceeecCC
Q Mon_May_01_23:  157 RIASDAEIRECAPEIYNVTYKITTDRGVTHEAVRHREMSFMQESTRWCNYAKGRLGYKYG-----------RNISVIEPP  225 (279)
```

```
Q Consensus       157 ~~~~~~~~~~~~~~~~~~t~~~~~sR~~~~qlvRHR~~Sfsq~SqRYv~~~~~~~~~~~~~~~~~~~~~~~~~~~~i~p~  225 (279)
                                        +||+++|||+|||++||+|+|||||-.....+.+...           ....+-.+.
T Consensus        86 ------------------------~Sr~~~~Ql~RHR~aSy~e~S~RYv~~~~e~~~~~~~~~~~~~~~~~~~~~~~~~  142 (273)
T COG1351          86 ------------------------VSRVAAHQLIRHRIASYSEKSQRYVLDGDEFYVPFEPLLEDRAAFNDLCRAVDDLYS  142 (273)
T ss_pred             ------------------------cCHHHHHHHHhccccCccccceeeecCCCceeecchhhcchhhhhHHHHhhccccc


Q ss_pred             CCCccHHHHHHHHHHHHHHHHHHHHCCCCHHHHHHHchhccccEEEEEEEcc
Q Mon_May_01_23:  226 FKNEDSLEKFYDVVAGNEAIYQELTNDGEPAQLARSVLPTATKSDIYISGTLD  278 (279)
Q Consensus       226 ~~~~~~~~~~~~~~~~~~Y~~l~~~G~~~E~AR~vLP~~~~T~iv~t~Nl~  278 (279)
                  ...+...+.+.++++.+.+.|..+++.|+++|+||+|||++|+|+++++|+|
T Consensus       143 ~~~e~~~~~~~~~~~~~~Y~~~l~~g~~~E~AR~vLP~at~T~~~vt~N~R  195 (273)
T COG1351         143 SKLEKVLDLLEKAYERSYELYKELLRSGIAREDARYVLPNATETRIVVTGNAR  195 (273)
T ss_pred             chhHHHHHHHHHHHHHhHHHHHHHHhCCCCHHHHHHhcchhheEEEEEEcHH
```

//

Query: alignment based on close homologs of gene_51_crAssphage (Phage stabilization protein)

HHpred output:

```
>pfam11134 Phage_stabilize Phage stabilization protein. Members of this family are phage proteins that are
  probably involved with stabilizing the condensed DNA within the capsid.
  Probab=86.29  E-value=31  Score=33.25  Aligned_cols=135  Identities=9%  Similarity=0.052  Sum_probs=0.0

Q ss_pred             Ccccccceecc--cceEEcCCCccccEEEEEeCCEEEEEeCCeeEEEccchheeccCCCcccccc-cceecCCeeeeecC
Q gene_5_9361044   22 SLANSWRHFRA--NNYKVLSKNKGNITNIIGVGTAFFVHTEHSLFYLNRDNLLKTSGNTAQLEMP-DLFEVEPIELFTSN   98 (177)
Q Consensus        22 s~in~~~~F~~--~~n~kdl~~~G~I~~L~~~~~~L~~~qe~k~~~~~~~~l~~~~~~~~~~~~-~~~~~~~~~~~~~~   98 (177)
                      |..++ ..|.+  +.|..-+..=-.|..+....+.|         .+.+..++..=..++..--+ ..|...+..+++
T Consensus       161 S~l~d-~s~~~~~l~fatAE~~PD~iv~~~~~~el----------~lfG~~TiEv~~ntG~~~f~~~f~r~~g~~i~--  228 (469)
T pfam11134       161 TDLED-ESHPDRYSAQYRAESQPDGIIGVGVWRDFI----------VCFGSSTIEYFTLTGSTTAGAALYVAQPSYMVQ--  228 (469)
T ss_pred             ccccc-cCccCccccceeeeccCCCceEEEEECCEE----------EEEcCCcEEEEEccCCCCCCcceeeccccEc--


Q ss_pred             CcccccCCCceeEEcCCceEEEECC---CCEEEEECCCcceechH-HHHHHHHhh----cCCceEEEEeCC-CCEEEEEe
Q gene_5_9361044   99 HGYGGLQHPQAWTVNSNGYWFVDAD---NKRIYNFDNNHLTDLTS-DILNWMNNV----QIADAHMVTDFA-NARVIMCL  169 (177)
Q Consensus        99 ~g~ggs~npeS~~~~~~~yF~D~~~~rg~V~~l~~~gi~~IS~~gm~~~F~~~~~~~~~~~~g~YD~~~~~~y~it~  169 (177)
                      -||--|.|++...+.+||+...   .+.|+|+++.+.+.||. .++.-++.+    +...+.-.|--. |.-|+|++
T Consensus       229 ----~G~~a~~s~~~~~s~~wlg~~~~g~~~V~~~~g~~~~~IST~~Ie~~l~~y~~~elsda~~~sy~~~GH~f~vl~~  305 (469)
T pfam11134       229 ----KGIAGTFCKCPYMDKYAIISHPATGAPSVYLIGSGQKSPIATASIEKIIRGYTADELAAGVMESVRFDSHELLIIHL  305 (469)
T ss_pred             ---cccccccchhhhCCEEEEEeCCCCCcEEEEccCCceeEeccHHHHHHHHHhccccccceEEEEEEEECCEEEEEEc


Q ss_pred             Cc
```

```
  Q gene_5_9361044  170 AY  171 (177)
  Q Consensus       170 ~d  171 (177)
                        .+
  T Consensus       306 P~  307 (469)
  T pfam11134       306 PR  307 (469)
  T ss_pred             CC
```

//

Query: gene_52_crAssphage_JQ995537 and gene_43_KJ003983 IAS virus (Tail tubular protein A)
HHpred output:

```
>3j4b_A Tail tubular protein A; bacteriophage, DNA ejection, tail complex, gatekeeper, viral; 12.00A
  {Enterobacteria phage T7}
  Probab=70.94  E-value=57  Score=29.68  Aligned_cols=70  Identities=10%  Similarity=0.041  Sum_probs=0.0

  Q ss_pred             EEE-eCCEEEEecCceEEEEEEeceecCCCccccCCCHHHHHHHHHHHHHHHHHhhhccCCChhhhHHHHHHHHHHHHH
  Q gene_26_Chitin  177 FDI-RDGKIFTGFTEGTLHIIFYKKEFDDSEFQLIPDNHRIIEYIKAFLKYKMYEQIFNEVADESFNQAERKYQLYKQEY  255 (291)
  Q Consensus       177 y~I-~~~~I~tnf~~G~v~l~y~~~~~d~~g~plIPD~~~~~eai~~yi~~k~l~~~~~~~~~y~~~~~w~~~~~A  255 (291)
                        |.+ .+++|.++    .|.|.|.... |.+-+     .+.+++|+.+.+..++-.+  .-|+....+..+++...+..+|
  T Consensus        92 yd~~~~~~i~t~~~~~v~v~yv~~~~d~~~~~~~~p~~f~~~i~~~lA~~~a~~~-~~~~~~~~~~~l~~~~~a~~~a  159 (184)
  T 3j4b_A           92 YDRTSQSDRFDS----GITVNIIRLR-DYDEM-----PECFRYWIVTKASRQFNNR--FFGAPEVEGVLQEEEDEARRLC  159 (184)
  T ss_pred             EECCCCceEeCC----cEEEEEEEcC-ChhhC-----CHHHHHHHHHHHHHHHHHHh--ccCCHHHHHHHHHHHHHHHHH


  Q ss_pred             HHH
  Q gene_26_Chitin  256 YDA  258 (291)
  Q Consensus       256 ~~~  258 (291)
                        ++.
  T Consensus       160 ~~~  162 (184)
  T 3j4b_A          160 MEY  162 (184)
  T ss_pred             HHH
```

//

Query: gene_53_Azobact_ph_AP017903, homolog of gene_85_crAssphage_JQ995537
HHpred output:

```
>3h4r_A Exodeoxyribonuclease 8; exonuclease, recombination, hydrolase; 2.80A {Escherichia coli}
  Probab=97.52  E-value=0.00049  Score=70.62  Aligned_cols=141  Identities=15%  Similarity=0.128  Sum_probs=0.0

  Q ss_pred             ccccccchhhHHHHHHHHHHHHccccCCchHHHH-------HHHHHHHHHcCc-cCCceecCChHHHHHHHHHHHHHHHH
  Q gene_55_100190  199 EATFSTETGTFTHGLMEVVHSLGKESTHKDRVD-------ELMRHWDKERGK-NDNPLYIEKDEDARRMCEKITDMYEEI  270 (1664)
```

```
  Q Consensus       199 k~tp~TeaG~~~H~lle~vh~l~~~~t~wd~~~~~~~~~ea~~~w~~e~g~~~hll~~~~d~~~r~i~~kit~~~e~i  270 (1664)
                        ..+++...|+.+|.+||..+.+.+.-...+..+      +++..|.++.-. ...++.-..-..+.+|+..+... +..
  T Consensus        39 ~~~~~~~~G~~~H~~le~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~  117 (265)
  T 3h4r_A           39 TKTKTLDLGTAFHCRVLELEEFSNRFIVAPEFNRRTNAGKEEEKAFLMECASTGKTVITAEEGRKIELMYQSVMAL-PLG  117 (265)
  T ss_dssp             CCCSCCTHHHHHHHHHSSHHHHTCC-----------------------------------CCCTTHHHHHHHHTTS-HHH
  T ss_pred             CCCcHHHHhHHHHHHHhCccccccceEEEccccccccchhhHHHHHHHHHHHHHhCCCcccCHHHHHHHHHHHHHHHhC-chH


  Q ss_pred             Hhhc-cEEeeeCCeEEEeCCCCeeeeeeeEEEEEeCCCCCeeEEEeeccccccCCchHHHHHHHhhcCcccCCCCcccch
  Q gene_55_100190  271 NKKY-EIVSMEAPVVLHSTEDGSPILGRCDIIALDKETGQGVLIDIKTHVENELPAEGEYNKFFQNYNLKPVGNGEHTAW  349 (1664)
  Q Consensus       271 ~~kf-~vis~E~Pl~~~~~Dg~~~~g~vDlIv~dK~tg~~giID~Kt~~e~~~~~~~f~kf~~n~~~K~vg~~eht~w  349 (1664)
                        ..-| ..+..|.|+.+....+|....|++|.|..    |...||||||..     ....|..-.....
  T Consensus       118 ~~~~~~~~~~E~~~~~~~~~~~~~G~iD~i~~~~~~~~~IiDyKT~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~  176 (265)
  T 3h4r_A          118 QWLVESAGHAESSIYWEDPETGILCRCRPDKIIP----EFHWIMDVKTTA-----DIQRFKTAYYDYR------------  176 (265)
  T ss_dssp             HHHHSSSCBSSCCEEEECTTTCCEEEECCSEEEG----GGTEEEEEEEES-----CHHHHHHHHHHT------------
  T ss_pred             HHHhccCCcceeEEEEEeCCCCeEEEEEeeeeec----CCCeEEEEeecC-----CHHHHHHHHHcC------------


  Q ss_pred             hhhcHHHHHHHHHH
  Q gene_55_100190  350 DHANVQTMGYAAAF  363 (1664)
  Q Consensus       350 d~~~~QT~aya~~~  363 (1664)
                          .+.|...|++++
  T Consensus       177 --~~~~Ql~~Y~~~~  188 (265)
  T 3h4r_A          177 --YHVQDAFYSDGY  188 (265)
  T ss_dssp             --TTHHHHHHHHHH
  T ss_pred             --hHHHHHHHHHHH
```

//

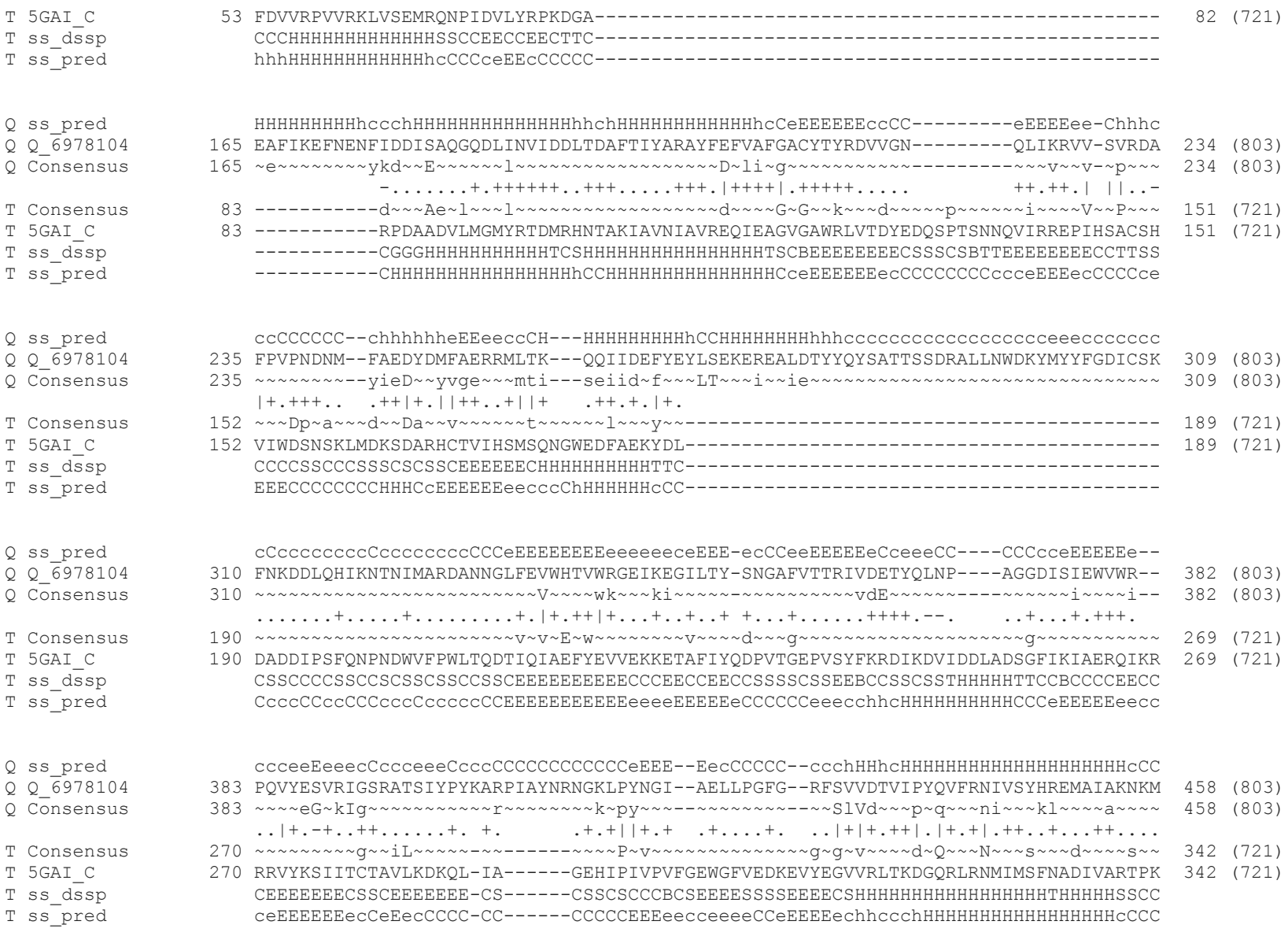
Query: Alignment homologs of gene_76_crAssphage_JQ995537 (major capsid protein)
PSI-BLAST output (first round):

```
>ANA48974.1 major capsid protein [Pseudomonas phage PaMx41]
Length=418

 Score = 198 bits (502),  Expect = 4e-58
 Identities = 82/488 (17%), Positives = 161/488 (33%), Gaps = 83/488 (17%)

Query  1    MGSMMQVYKGEYN-SGFTDENHLSNALLQEPEELSKIITHLYGSDDSRFPLTFLTEGMGN  59
            M     ++    N      ++     L + P   + ++
Sbjct  1    MSVYAGIFNTTLNPQELNMKSFAGTILRRVPNGSAPLLAM--------------------  40

Query  60   EETIGDTEYEWKVMGRSKRPSPIVTVNSSTTNPGVGGSTFEIEFEENWFAPGHVLVADDN  119
               +G T  +    G   +          T       +   +E   +    G +   +
```

```
Sbjct  41   TSVVGSTTAKASTHGYFSKTMVFA-SAVVTAEAAADATVLTVE-NSDGLTKGMIFYNEAT  98

Query  120  GYQVRVMNDPQPDGSGYVYELRLVNGDPSAFIDPSYLAAGSQWSMIGSAVEESSSEGGSS  179
            G  +R+           V  L L     +  I  + +AA ++  +IG+A EE  S+  ++
Sbjct  99   GENMRLEL---------VNGLNLTVKRQTGRISAAIIAANTKLIVIGTAFEEG-SQRPTA  148

Query  180  NYSTPGKMRNQLSTMRMSYSITGNAANDVLSYALKDSGGGTTNLWMDYEEWQFMLQWREE  239
                P  + N     R ++++T  A                    +       M     E
Sbjct  149  RSIQPVYVPNFTQIFRNAWALTDTARASYAEAG---------YSNITESRRDCMDFHATE  199

Query  240  KEQQLWYGQSNRDANGTIHMKDENGQPIPRGAGILE---QIEPSNYDTYSNLTEKKLEDI  296
            +E  +++GQ+             NGQP+    GI++   Q  P N +   N T    +D+
Sbjct  200  QETAIFFGQAFMGT--------YNGQPLHTTQGIVDAVRQYAPDNVNAMPNPTAVTYDDV  251

Query  297  LG---DLFYNRVDSGER-QIVLFTGEGGMREFHEALKNEAAGSGFMTALDSKFIEGSGQN  352
            +    D F   V+ G+  Q V+F    GMR   +                    +  + +
Sbjct  252  VDATIDAFKWSVNVGDNTQRVMFCDTVGMRTMQD----------IGRFFGE--VTVTQRE  299

Query  353  LSYGGYFTEYKAVNGHRITIKHLPLFDDGTMNGKMHPVTGLPMESYRMVFLDMGTSNGQS  412
             SYG  FTE+K   G R+ IK  PLF    ++     V  +P  + ++ ++D       +
Sbjct  300  TSYGMVFTEWKFFKG-RLIIKEHPLFSAIGISPGFAVVVDVP--AVKLAYMDGR----NA  352

Query  413  NIQMVRKKGREYLRWYVSGSVIPSGWGA-SNGNMMANDVDGYSVHVLSEAGIAIRNPTSC  471
             ++   + G E        +    G G  + G  + ++   +++ +L+  G A+
Sbjct  353  KVENYGQGGGENKSG---ATDYSYGHGVDAQGGSLTSE---WALELLNPQGCAVITGLQK  406

Query  472  GELECALN  479
             +    L
Sbjct  407  AKERVYLT  414
//////////////////////////////////////////////////////////////////////////////////////////////////////////////////
```

Query: gene_78_crAssphage_JQ995537 (portal)
HHpred output:

```
>5GAI_C Portal protein, Peptidoglycan hydrolase gp4; virion, portal, tailspike, adhesin, VIRAL; 10.5A {Enterobacteria
phage P22}
Probab=99.56  E-value=3.1e-16  Score=192.06  Aligned_cols=590  Identities=11%  Similarity=-0.012  Sum_probs=0.0
Template_Neff=9.400

Q ss_pred            ccccchHHHHHHhcccccCCCcEEEEeCChHHHhHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHcCCChhhCCHHHhhcH
Q Q_6978104        85 YDMMKGIIRRYIGEYIKNPHDFIVGANNPEVVFARDAELGKQIMMLAEQAVAKKIQESYMQFVNEGNNPEQFNPEQAVDI  164 (803)
Q Consensus        85 ypIi~~~vd~l~ge~~~r~~~~v~a~d~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~p~~~~~~  164 (803)
                      +|++.++|+.++|.+...+..+.|...++.
T Consensus        53 ~N~i~~~v~~i~~l~~~p~~~v~p~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~  132 (721)
```



```
Q ss_pred            cEEEEehhhcCCCHHHHHHHHHHcCcEEecCCCccccccccceeeehhhhHHHHHHHHHHHHHHHHHHHHHcCCCHHHcc
Q Q_6978104      459 NVLMIAKSLLGKKPAETIYRMAADGVLYIDDEDDANLVKAQNVRYLESRMNNYITELGQLIQEIEQTAKMECDMTPQRYG  538 (803)
Q Consensus      459 ~~~~id~~~~~~~~~~~~~~~~Gv~~~~~~~~~~~~~~~~~~~~~~~~~~~~~i~~~i~l~~~~~q~i~~itGvn~ar~G  538 (803)
                    ..++++.+.+....+.|.. ....+++..++..... +..+...+..+....-.+.+..+++...+.+++++|+.....|
T Consensus      343 ~~~~v~~~~i~~~~~~~~~~-~~~~~~i~~~~~~~~g-~~~~~~~i~~~~~~~~~~~~~~l~~~~~~i~~vsg~G~~~~~  420 (721)
T 5GAI_C         343 KKPFFWPEQIAGFEHMYDG-NDDYPYYLLNRTDENS-GDLPTQPLAYYENPEVPQANAYMLEAATSAVKEVATLGVDTEA  420 (721)
T ss_dssp            SCEEECTTTSTTCSSSCSS-SSSCSSBCCCCCCSSS-CCCCSCCCEECCCCCCCCHHHHHHHHHHHHHHHHTTCCCCCSSS
T ss_pred            CCCccCHHHhcchHHHhhc-CCCCCeeeeccCCCCC-CCCCCCCCCceeCCCCCCHHHHHHHHHHHHHHHHHhcCCCcch


Q ss_pred            CCCccccchHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHhheeeecCcceEEEecCCceEeeEecC----------
Q Q_6978104      539 EIANSAGKGVTDEAVIRGSMGSVIIEFIFDKMRERDYQAEMDYTKLAWIDGLNTSYKTKDGDIRYLSLDV----------  608 (803)
Q Consensus      539 ~~~~~~~~~~q~a~~qSs~~t~~~~~~~~~~r~~~~l~~~k~~~~~~~~y~~~d~~~~l~i~~----------  608 (803)
                    ..+...+.++++....+++..+..++..+.....++.+.++..++..|..+....+...++...++.+++
T Consensus      421 ~~g~~~s~~~i~~l~~a~~~~~~~~~l~~~~~~~~~~~~li~~~y~~~r~iri~g~dg~~~~v~i~~~~d~~~g~~  500 (721)
T 5GAI_C         421 VNGGQVAFDTVNQLNMRADLETYVFQDNLATAMRRDGEIYQSIVNDIYDVPRNVTITLEDGSEKDVQLMAEVVDLATGEK  500 (721)
T ss_dssp            SCCTTHHHHHHHHHHHHHHHHHHHHHHHHHHTCCCCCCCSCCSSSCCCCSCSSSSCCSCSSCSSSSCSSSCCCSCCCCCCS
T ss_pred            hccchhhHHHHHHHHHHHhhhHHHHHHHHHHHHHHHHHHHHHHHHHHHHhcCCCceEEEEcCCCCeeEEEeccceeccccCce


Q ss_pred            ---CcchhhhceEEeeCCHHHHHHHHHHHHHHHHHHHhCCcchhHH------HhhcccHHHHHHHHHHHHHHHHH----
Q Q_6978104      609 ---NSHIFANYIVTCKTSVKEREKLEQYKQLAFSAAQNGNMDMANA------AIRGDNVAQISKLIDKYQNIQREH----  675 (803)
Q Consensus      609 ---~~~~~~~~dfgIfv~~~~~d~e~~~l~ql~q~alQng~i~~~d~------~~~~s~~~~~~le~~e~~~re~~~~~  675 (803)
                       .++ ..+|.|.|...+.-....+...+.+...++.....-...      +...-........++..++.....
T Consensus      501 ~~~~dl-~~~~dv~V~~g~s~~~~r~~~~~l~~l~~~~~~~p~~~~~~~~~~~~d~~~~~~~~~~~~~~~~~~~~~~~  579 (721)
T 5GAI_C         501 QVLNDI-RGRYECYTDVGPSFQSMKQQNRAEILELLGKTPQGTPEYQLLLLQYFTLLDGKGVEMMRDYANKQLIQMGVKK  579 (721)
T ss_dssp            SSSCCB-SCCCCEECCEESCCSSHHHHHHHHHHHHHHSSSCCTTHHHHHHHHHHHHHHSCCCSSHHHHHHHHHHHHHHTSSC
T ss_pred            EEeeec-ceeeEEEeeecCCcHHHHHHHHHHHHHHHHHhcCCCCCHHHHHHHHHHHhhhcCCCCHHHHHHHHHHHHHHcCCCC


Q ss_pred            -------HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHhhhhccCCCHHHHHHHHHHH
Q Q_6978104      676 -------ELDVERVSQQTEQLRQEFELAKIDRKAEQDRETIRVEKYLDGQIEAMKANANIMSFDNGLSDAEKNQAEERME  748 (803)
Q Consensus      676 -------~~~~~q~q~q~qq~~q~e~q~~~~~~~eke~~~~~~~e~~~~~~~~~~~~~~~~~d~~~~~~~~~e  748 (803)
                            ++.+++++++.++...+..+++.++........+......+.+.+...++.+.....   ......
T Consensus      580 p~~~e~~q~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~  656 (721)
T 5GAI_C         580 PETPEEQQWLVEAQQAKQGQQDPAMVQAQGVLLQGQAELAKAQNQTLSLQIDAAKVEAQNQLNAARIAEIF---NNMDLS  656 (721)
T ss_dssp            CCCSSCTTTTTTTTHHHHHHHHHHHHHHHHTTTTHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH---HHHHHH
T ss_pred            CCCHHHHHHHHHHHHHHhcCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH---hcCChh


Q ss_pred            HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHhhhhHHHHHH
Q Q_6978104      749 NARLNLERSKLSLDAQKTSVEAQLKEKELAVKLKESDDKVKI  790 (803)
Q Consensus      749 ~~kl~le~~k~~~~~k~~~d~~~k~k~l~~k~~e~~~k~ki  790 (803)
                    .....++..++-.......-+-.....++..+-.+..-+..+
T Consensus      657 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~  698 (721)
```

```
T 5GAI_C          657 KQSEFREFLKTVASFQQDRSEDARANAELLLKGDEQTHKQRM  698 (721)
T ss_dssp             HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
T ss_pred             HHHHHHHHHHHHHHhcccChhHHHHHHHHHhcCCCCCcCCcc
```

//

Query: gene_79_crAssphage_JQ995537 (terminase large subunit)

HHpred output:

```
1.2O0J_A DNA packaging protein Gp17; nucleotide-binding fold, HYDROLASE; HET: ADP; 1.8A {Enterobacteria phage T4}

Probability: 97.67   E-value: 1.6E-5   Score: 90.61   Aligned Cols: 269   Identities: 16%   Similarity: 0.14

Q  ss_pred                  hccCCchHHhhHHHHHcCCCCC-Ccccccccccceecc----cCCccccCcHHHcHHHHHHHHhCCccCCCCCChhhHHH
Q  Q_8181294     11       YIEEPDKEKKYKKASDCGWYDP-HNNFLIGDSGGFLLN----IRPGKFVNTELFNEAARTYQATGKYTQFKVDSIPHRQF
Q  Consensus     11       ~~~~~~~~~~~~~~~~~~~~~~-~~~~~~~~~~~~~~~----~~~~~~~~~~~~~~a~~f~~~~~~~~~~~~~s~~~~~f
                           .|..|+..+|+...-...+.+- +...-+..+..||--    .-+.+-+..+.|+.-...+++.-.|...|--...+.+.
T  Consensus     46       ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~k~~~~~~
T  2O0J_A        46       LIKHPSLAERKDEDGIHWIKSQWDGKWYPEKFSDYLRLHKIVKIPNNSDKPELFQTYKDKNNKRSRYMGLPNLKRANIKT
T  ss_dssp                  SSCCGGGSCEEEETTEEEEECTTTSCEEESBHHHHHHTSCCCBCCCBSSCTTSTTTBCCCSSTTSEETTEEEEEBTTCCC
T  ss_pred                  EecChhHhhcccccccceeeeeccCCccccccchHHHhhhcceEecCCCCCChHHHhhhhcccCCceeecCCccccccCCccc
Confidence                   3566776666554433222221 122222223333211    12334466677888888888888887776665555544


Q  ss_pred                  -----HHHHHHHHHcCcccCCccCCCCCcccеEeccccееEeeeeeccccCCCccccccccccceeecCCCccHHHHHHH
Q  Q_8181294     11       -----RRRECDRRRNGFSAPCWQNPDGSIEDVWITGGHYNFLNYTRMERTDESSVIVTEHGATAKKIYSFPSFIDAQFWT
Q  Consensus     11       -----~~~E~~r~~~G~~~~~~w~~~~g~~~~~~yITG~hYfyLn~~~I~~~~~~~~~~~~~~~~~~~~~k~~~~P~frd~~~~~
                                +..++.+|..-+                    .||.-+++.|...+.              ....+ .++..|..+
T  Consensus     46       ~~~~e~~~~~~k~~~D~--------------------~~f~e~~~~i~~~~g------------------~~~~~l~p~Q~ei
T  2O0J_A        46       QWTREMVEEWKKCRDDI--------------------VYFAETYCAITHIDY-------------GVIKV-QLRDYQRDM
T  ss_dssp                  CCCHHHHHHHHHHHHCH--------------------HHHHHHHCEEECSSS-------------CEEEC-CCCHHHHHH
T  ss_pred                  cCCHHHHHHHHHHHhhCH--------------------HHHHHHcCeeecCCC-------------Ceecc-CCCHHHHHH
Confidence                        233344443332                    223333444432110             00011 233344333


Q  ss_pred                  HHHHHHHHHcCCeEEEEcCCCCChhHHHHHHHHHHhhhcccCCcEEEEEecCChHHHhccchhHHHHhhHHhhhcCCCCCCC
Q  Q_8181294     11       WQIIEFCRRNGLHLIIDKTRRGGFSYIMAADSSNEVNLSKHKVVIHVAADNKYLIKQGGLSDFAVNNLKFFEEKTPFKRG
Q  Consensus     11       f~~~e~~~~~~~~~~~iiLK~Rr~G~S~~~~a~~~~~~~~~~~~~s~~~~~~~~~d~k~~~~d~~~~k~~~d~~~~~~~~~~~~~~~k~
                           +....    .....+++++||.|+|+.+++.+++.+.+.++..+++++.+.+...  ..+ ..+   ...+...+.+.++
T  Consensus     46       l~~l-----~~r~~ii~~gRq~GKTt~~a~~~l~~~~~~p~~~iliia~t~~qa~--~l~-~~i---~~~~i~~lp~~l~~
T  2O0J_A        46       LKIMS----SKRMTVCNLSRQLGKTTVVAIFLAHFVCFNKDKAVGILAHKGSMSA--EVL-DRT---KQAIELLPDFLQP
T  ss_dssp                  HHHHH----HSSEEEEEECSSSCHHHHHHHHHHHHHHHSSSSCEEEEEEESSHHHHH--HHH-HHH---HHHHHHSCTTTSC
T  ss_pred                  HHHHc----cCCeEEEEcCCCCChhHHHHHHHHHHHHHHhCCCceEEEEeCChhHHH--HHH-HHH---HHHHHHhCCHHhcC
Confidence                   22111    34567899999999999999999999999888888888888876544321  122 212   2222222222222
```

```
Q  ss_pred                 CCCCCCcceeeeeeeecCCCcccCcccceeEEEEcCCCccccCCceeEEEEeehhcCCCHHHHHHhhhhhhcCcceEEE
Q  Q_8181294     11       IYSPTTDSFKLGYRMKNGVEADDSWSSSLLSVSANNNPDCAIGKDAVTIKVEELSTMQNFDEFMNVTEPTMTVGTRTTGT
Q  Consensus     11       ~~~~~~~~~~~~~~~~~~~~~~~G~~S~I~~~s~~~n~~~~~G~~~~~l~~dE~Gk~~~~~~~~w~~~k~~l~~G~~~~Gk
                           .....+ ...        .....  .|.|...+.  ++.+.+|.++.++++||++.++...+.|..+.+++..+   ..|+
T  Consensus     46       ~~~~~~-~~~~~~~~~~~~~i~f~n--Gs~I~~~s~------~rG~~~~~viiDE~a~~~~~~~~~~~~~~l~p~l~~~~-~~~~
T  2O0J_A         46       GIVEWN-KGS--------IELDN--GSSIGAYAS--SPDAVRGNSFAMIYIEDCAFIPNFHDSWLAIQPVISSG--RRSK
T  ss_dssp                  CEEEEC-SSE--------EEETT--SCEEEEEEC--SHHHHHTSCCSEEEEESGGGSTTHHHHHHHHHHHHHST--TCCE
T  ss_pred                  CeeecC-CCE--------EEeCC--CCEEEEEeC--CCCcccccCCeeEEEEccHhhCCChHHHHHHHHHHhhcC--CCcE
Confidence                 111000 000        01111  366666554  467889999999999999987655789888888763  3478
```

```
Q  ss_pred                 EEEEEcCCcccchHHHHHhc
Q  Q_8181294     11       LMAWGTATAANMQIFEQNFY    340 (751)
Q  Consensus     11       ~i~~sT~n~~~~~~~~kem~~    340 (751)
                           +++.||.++.  +.|.++|.
T  Consensus     46       ii~iSTP~g~--~~f~~l~~    324 (385)
T  2O0J_A         46       IIITTTPNGL--NHFYDIWT    324 (385)
T  ss_dssp                  EEEEECCCSS--SHHHHHHH
T  ss_pred                  EEEECCCCCC--CHHHHHHH
Confidence                 9999998875  44555554
```

//

Query: alignment for close homologs of gene_56_crAssphage_JQ995537
HHpred output:

PF12571.7 ; DUF3751 ; Phage tail-collar fibre protein

```
Probability: 99.18   E-value: 7.8E-19   Score: 153.26   Aligned Cols: 146   Identities: 14%   Similarity: 0.202
Q  ss_pred   CCccceeecHHHHHHHHHHHcCCCcEEEEEEEecCCCC-CHHHHHHHhccccceeEecceEEEe--cCCCeEEEEEEEEeCCcccCCCEEEEEEEEEEEECCCCCeEEEEEEecCCCCcccCCcCCCcceeEEEEEEEEecCCceEEEEEECCcceecHHH
Q  Q_20456   AEYSKLYITNNGQALMAKMIAGSGNIDFTKVCSSSTQY-TESQLQALTALSNIKQTTLVSKVTR--TNEVAIKIDAAYSNVDLKEGYYMRTLGLYAVDPDKGEILYAVCIEKSNNCYMPPYNGVTVSAAYLQLYTTVGNADNVSLAVSPGAYATVGD    154
(600)
Q  Consens   a~~~~~~lT~~G~~~~a~a~a~~~~i~~t~~~~g~g~~-~~~~~~~~t~l~~~~~~~~i~~~~~-~~~~~~~~i~~~~~n~~~~~g~~~~eiGl~a~d~~~~eiLy~~~~~~~~~~~~p~~~~~~~~~~~~~~~~vsn~~~Vt~~id~~~~~t~~~    154
(600)
             ++|.+.+||++|+++++++++++.+|+||||++|+|.+ +..+...+|+|.+++++.+|+.+..  .+++++.+.+.+++.  .++|+++|+|||++++    +||++....  +.|+|...++....+.+++.+.++++++|++.+|++......+
T  Consens   ~~~~~~~iT~~G~~~~~~~~~~~~~l~~t~~~~G~g~~~~~~~~~~~~t~L~~~~~~~~i~~~~~~~~~~~~~~~~~~~~~~~-~~~~~~~~eiGl~~~~g----~l~a~~~~~-~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~v~~~id~~~~~~~~    150
(150)
T  PF12571   EQKYKTILTHHGERVIVEALANKIPVPLKEMAIGDGNGSSITPSASQTTLVREVYRAEITDLLEDPQNRHQMIAELLIPEN--VGGFIVREIGLFDEQG----GLVAVANCP--ENYKPVLEQGSGKVQYYRMILQVSSSDAVTLSINNNIVYATRT    150
(150)
T  ss_pred   CccccceEecHHHHHHHHHHHhCCCCcceEEEEEecCCCCCCCCChhhhcCcEEEEeecceeEECCCCCCEEEEEEEecCC--cCCEEEEEEEEEcCCC----CEEEEEECC--CCccCcccCCCccEEEEEEEEEEccCCCcEEEEEcCceeeeecC
```

//

Query: gene_39_crAssphage_JQ995537
CDD output:

```
Thioredoxin_like super family    cl00388
Protein Disulfide Oxidoreductases and Other Proteins with a Thioredoxin fold; The thioredoxin (TRX)-like superfamily is
a large, diverse group of proteins containing a TRX fold. Many members contain a classic TRX domain with a redox active
CXXC motif. They function as protein disulfide oxidoreductases (PDOs), altering the redox state of target proteins via
the reversible oxidation of their active site dithiol. The PDO members of this superfamily include the families of TRX,
protein disulfide isomerase (PDI), tlpA, glutaredoxin, NrdH redoxin, and bacterial Dsb proteins (DsbA, DsbC, DsbG, DsbE,
DsbDgamma). Members of the superfamily that do not function as PDOs but contain a TRX-fold domain include phosducins,
peroxiredoxins, glutathione (GSH) peroxidases, SCO proteins, GSH transferases (GST, N-terminal domain), arsenic
reductases, TRX-like ferredoxins and calsequestrin, among others.

The actual alignment was detected with superfamily member TIGR00411:

Pssm-ID: 320948  Cd Length: 82  Bit Score: 28.31  E-value: 0.26
                           10        20        30        40
                    ....*....|....*....|....*....|....*....|....
lcl|seqsig_MKKLI_ 44 TDTGRQEARSAGISDAPTAYCNGDIL-RGVQSDYTIRKYLRKLL 86
Cdd:TIGR00411     39 VMENPQKAMEYGIMAVPAIVINGDVEfIGAPTKEELVEAIKKRL 82

////////////////////////////////////////////////////////////////////////////////////////////////////////////////////
```

Query: gene_41_crAssphage_JQ995537
CDD output:

```
TRX_family    cd02947
TRX family; composed of two groups: Group I, which includes proteins that exclusively encode a TRX domain; and Group II,
which are composed of fusion proteins of TRX and additional domains. Group I TRX is a small ancient protein that alter
the redox state of target proteins via the reversible oxidation of an active site dithiol, present in a CXXC motif,
partially exposed at the protein's surface. TRX reduces protein disulfide bonds, resulting in a disulfide bond at its
active site. Oxidized TRX is converted to the active form by TRX reductase, using reducing equivalents derived from
either NADPH or ferredoxins. By altering their redox state, TRX regulates the functions of at least 30 target proteins,
some of which are enzymes and transcription factors. It also plays an important role in the defense against oxidative
stress by directly reducing hydrogen peroxide and certain radicals, and by serving as a reductant for peroxiredoxins. At
least two major types of functional TRXs have been reported in most organisms; in eukaryotes, they are located in the
cytoplasm and the mitochondria. Higher plants contain more types (at least 20 TRX genes have been detected in the genome
of Arabidopsis thaliana), two of which (types f amd m) are located in the same compartment, the chloroplast. Also
included in the alignment are TRX-like domains which show sequence homology to TRX but do not contain the redox active
CXXC motif. Group II proteins, in addition to either a redox active TRX or a TRX-like domain, also contain additional
domains, which may or may not possess homology to known proteins.

Pssm-ID: 239245  Cd Length: 93  Bit Score: 25.60  E-value: 3.0
                           10        20        30        40        50        60        70        80
                    ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*
```

```
lcl|seqsig_MIRID_  7 FTRDGCDACKiAIKNITDAINEANCDIT---LNIrnTNLDDILRK-EITKFPTTVITKvdndyKRKELARLEGSFPSDYIKDIIN 87
Cdd:cd02947       17 FWAPWCGPCK-AIAPVLEELAEEYPKVKfvkVDV--DENPELAEEyGVRSIPTFLFFK-----NGKEVDRVVGADPKEELEEFLE 93

////////////////////////////////////////////////////////////////////////////////////////////////////////////////
```

Query: gene_84_crAssphage_JQ995537
CDD output:

```
PG_binding_3 pfam09374
Predicted Peptidoglycan domain; This family contains a potential peptidoglycan binding domain.


Pssm-ID: 286462  Cd Length: 76  Bit Score: 39.38  E-value: 2.66e-05
                           10        20        30        40
                    ....*....|....*....|....*....|....*....|....*...
lcl|seqsig_MKLNK_  66 SLNAYKSDTTISADYFVAKYKLERIRYYNDIAGK-GNNIKFLRGWIRR 112
Cdd:pfam09374      27 TLAAVKQRASAGEDALIDAICLARRAFYLRLAAKrTTNARFLRGWVNR 74
////////////////////////////////////////////////////////////////////////////////////////////////////////////////
```

Query: gene_36_Cellulophaga_ph_NC_021806 (homolog of gene_22_crAssphage_JQ995537) primase
CDD output:

```
Toprim_N super family      cl26789
DNA primase catalytic core, N-terminal domain;

The actual alignment was detected with superfamily member TIGR01391:

Pssm-ID: 331610  Cd Length: 415  Bit Score: 36.82  E-value: 0.02
                         10        20        30        40        50        60        70        80
                  ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
gi 526178681   14 ITSEII---LERLNPIDVYKMYIK-----DEFVVNrpfsSPFRSDNIPSFSIYQDRRsdqiLFNDFVEG-GGNCIQFVKK 84
Cdd:TIGR01391   2 IPEEFIdelKERVDIVDVISEYVKlkkkgRNYVGL----CPFHHEKTPSFSVSPEKQ----FYHCFGCGaGGDAIKFLME 73
                         90       100       110       120       130       140       150       160
                  ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
gi 526178681   85 LFNSTWYQACSRIAIDFKISDDYTVDKMNNTPNVNLQNTLDREVKVKSKIVNLQVKIRNYNSDDIAYWDSYGISMETLKR 164
Cdd:TIGR01391  74 IEGISFVEAVEELAKRAGIDLPFEKDQQEKKEQKSKRKKLYELLELAAKFFKNQLKHTPENRAALDYLQSRGLSDETIDR 153

Toprim pfam01751
Toprim domain; This is a conserved region from DNA primase. This corresponds to the Toprim domain common to DnaG
primases, topoisomerases, OLD family nucleases and RecR proteins. Both DnaG motifs IV and V are present in the
alignment, the DxD (V) motif may be involved in Mg2+ binding and mutations to the conserved glutamate (IV) completely
abolish DnaG type primase activity. DNA primase EC:2.7.7.6 is a nucleotidyltransferase it synthesizes the
oligoribonucleotide primers required for DNA replication on the lagging strand of the replication fork; it can also
```

prime the leading stand and has been implicated in cell division. This family also includes the atypical archaeal A subunit from type II DNA topoisomerases. Type II DNA topoisomerases catalyze the relaxation of DNA supercoiling by causing transient double strand breaks.

```
Pssm-ID: 307732  Cd Length: 88  Bit Score: 24.60  E-value: 38
                         10        20        30        40        50
                  ....*....|....*....|....*....|....*....|....*....|....*...
gi 526178681  257 VSLQQENAVAKENVVDELKDRFDWIFVLYDNDYDkdtnwGEKFGKEFSEKHNLIFIMI 314
Cdd:pfam01751  31 LSLEKGPKKKALKALKELAKRAKEVILATDPDRE-----GEAIALKLLELLENAGGRV 83

////////////////////////////////////////////////////////////////////////////////////////////////////////////////////
```

Query: gene_43_crAssphage_JQ995537
HHpred output:

```
>PF13876 Phage_gp49_66:  Phage protein (N4 Gp49/phage Sf6 gene 66) family;  InterPro: IPR025915 This family of phage
proteins is functionally uncharacterised. The family includes bacteriophage Sf6 gene 66 (Q716B1 from SWISSPROT) as well
as phage N4 GP49 protein (A0MZD7 from SWISSPROT). Proteins in this family are typically between 87 and 154 amino acids
in length. There is a conserved NGF sequence motif.
Probab=100.00  E-value=1.2e-36  Score=215.95  Aligned_cols=76  Identities=42%  Similarity=0.649  Sum_probs=73.3

Q ss_pred             ccHHHHHHHHhceEEeeeC-----CeEEEEEEeCCCEEEEEEEEecCHHHcCHHHHHHHHHHHHHHHHHHHHHHHH
Q lcl|gene_43_cr   69 LKENDIKRFIKGIENVKVG-----TKTTNTTLTCLTGFEVHGQAACVKPENFDLNVGSNYARIKAEDKIWEGLGFVLQWA  143 (149)
Q Consensus        69 It~e~Id~~I~~~~~~~~g-----~~~Tvc~~~L~NGF~v~G~Sacvdp~nfd~eiG~~~A~~~A~~klW~Legy~L~~~  143 (149)
                      ||+++|+++|++++|++.+     +++|||+++|+|||+|+|+||||||+|||+|||++||++||++|||+|+|+++|+|+++
T Consensus         1 vT~~~Id~~I~~~~y~~~~g~~~~~Tvc~l~L~NGF~V~g~Sacv~p~Nfd~eiG~kiA~~na~~kiW~LeGy~L~~~   80 (81)
T PF13876_consen    1 VTPEDIDALIASEEYFTAGGGVPLTTLTVCVLTLRNGFTVTGESACVDPENFDEEIGRKIARENAVDKIWELEGYLLKQK   80 (81)
T ss_pred             CCHHHHHHHHheEEEEEecccCcCCceEEEEEECCCcEEEEeEccCChhhcCHHHHHHHHHHHHHHHHHHhhHHHHhhc
Confidence            6899999999999865     899999999999999999999999999999999999999999999999999997

////////////////////////////////////////////////////////////////////////////////////////////////////////////////////
```

Query: Alignment with 6 closest homologs of gene_45_crAssphage_JQ995537 (RepL)
PSI-BLAST output (first round):

```
>SFN62202.1 replication protein (RepL) [Izhakiella capsodis]
Length=134

 Score = 45.5 bits (106),  Expect = 2e-06
 Identities = 18/64 (28%), Positives = 35/64 (55%), Gaps = 1/64 (2%)
```

```
Query  102  SNVIDFSHNDITEYG-HINSRTFYKAIQELYTENIIRPTNKKNVYVVNHNYIFRGNINKFIQLY  164
            ++ I  S++ I      I+  T++K ++EL  +  I  T  +N Y +N +Y+F G+   F++ Y
Sbjct  62   TDTILMSYDIIDMKAVKISRTTYFKGMKELVEKQFIAETMIQNYYFINPDYMFNGDRLSFVKSY  125


>WP_074898747.1 hypothetical protein [Bacillus megaterium]
 SFH67410.1 replication protein (RepL) [Bacillus megaterium]
Length=156

 Score = 45.9 bits (107),  Expect = 2e-06
 Identities = 21/74 (28%), Positives = 38/74 (51%), Gaps = 2/74 (3%)

Query  90   IISYICDNLKYNSNVIDFSHNDITEYGHINSRTFYKAIQELYTENIIRPTNKKNVYVVNHNYIFRGNINKFIQL  163
            +++YI DNL +NSN++  +  +I +   I   T    I+ L  +  ++   K  V ++N + I  G+ NK   L
Sbjct  69   VVNYILDNLDWNSNILIKTQQEIAKEAGIGFNTVNTTIKMLVDKKFLK--VKTGVIMLNPDIIAYGSHNKRAHL  140

//////////////////////////////////////////////////////////////////////////////////////////////////////////////////
```

Query: Alignment of homologs of gene_54_crAssphage

HHpred output:

```
>1OWF_A Integration Host Factor Alpha-subunit/Integration Host; protein-DNA recognition, indirect readout, IHF; 1.95A
{Escherichia coli} SCOP: a.55.1.1
Probability: 31.26   E-value: 280.0   Score: 22.05   Aligned Cols: 37   Identities: 8%   Similarity: 0.191

Q  ss_pred                       cchHHHHHHHHHHHHhhCCCCCCCCccccHHHHHHHHHHHHHHHHHH
Q  gene_33_Azobac        1       --MTLNQIADDIILTAGYINK-EKAMTINRLQLLSWIHMYRNIIMLQ    60 (339)
Q  Consensus             1       ~mmTLnqIv~~I~~~~~~~~s~d~~~~rqIk~~I~~yRallIrQ    60 (339)
                                 |+||..+|+..|+....          +++.+++..+...-..+...
T  Consensus             1       ~~~~~~~l~~~ia~~~~~~~~~~~~s~~~v~~v~~~~~~i~~~    37 (99)
T  1OWF_A                1       MALTKAEMSEYLFDKLG----------LSKRDAKELVELFFEEIRRA    37 (99)
T  ss_dssp                       -CBCHHHHHHHHHHHHC----------CCHHHHHHHHHHHHHHHHHH
T  ss_pred                       CCCCHHHHHHHHHHHhC----------CCHHHHHHHHHHHHHHHHHH


//////////////////////////////////////////////////////////////////////////////////////////////////////////////////
```

Query: gene_57_crAssphage_JQ995537
HHpred output:

```
>4lin_A Tail needle protein GP26; helical, trimeric coiled-coil, viral fiber, IN-frame extensi elongated coiled coil
protein, protein engineering; 2.70A {Bacteriophage P22}
Probab=98.83  E-value=3.1e-06  Score=66.98  Aligned_cols=88  Identities=15%  Similarity=0.275  Sum_probs=50.3

Q ss_pred             HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHhHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHh
Q lcl|gene_57_cr   78 LLTEKFDLIHSDLATISDKLTTIDTNIKNMSSALQGKFNALIAAVNDMKASNDTKNDAIIAALQGLVTQVNRNTSNINSLDGRVDALEQ  166
(167)
Q Consensus       78 ~l~~~~~~l~~~i~~~~~~~~~~l~~~~~~~~~~~l~~~~~~~~~~~~~~~~~~~~~~~l~~~l~~l~~~v~~~~~~~i~~l~~rvd~le~  166
(167)
                     .|..+++.+++.++.+...+..+..++..++..+ ..++.-...++.++..++..++.+..++..++..+..+..++..++.||..|+.
T Consensus      111 ~Le~rl~~Le~~v~~Le~~v~~Le~rv~~le~~V-~~Le~~v~~Lesrls~Les~v~~Le~rv~~le~~Vs~L~~~vs~Le~rV~~Le~  198
(300)
T 4lin_A         111 ALNVRVTTAEGEIASLQTNVSALDGRVTTAENNI-SALQADVDDHESRITANTKAITALNVRVTTAEGEIASLQTNVSALDGRVTTAEN  198
(300)
T ss_dssp            HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH-HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHTTTHHHHHH
T ss_pred            HHHHHHHHHHHhhhhhhhHHHHHHHHHHHHHHHH-HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
Confidence         3444455555555555555555555555555554 33444444455555566666666666666666666666666666666653
```

//

Query: gene_65_crAssphage_JQ995537 (a common fragment with gene_69_crAssphage_JQ995537
PSI-BLAST output (against viruses subset of NR database, second iteration):

```
hypothetical protein UGP_077 [uncultured crAssphage] (gene_69_crAssphage_JQ995537)
Sequence ID: gi|674660388|YP_009052544.1

Length: 868Number of Matches: 1
Score  Expect Identities    Positives    Gaps
115 bits(287) 3e-30  26/60(43%)   41/60(68%)   0/60(0%)


Query  1    LTQDMVNEENTRYIIRYAFDLSGKTITMPVGCELVFEGGIIENGTINLNKCKLTGMVGEE  60
            LTQDM+NE NT YI++Y + L+G+TI +P    +++ GG I +G + LN+C+L     +E
Sbjct  274  LTQDMINEPNTIYILQYDYCLAGQTIELPDNSIILWRGGRIYDGAVKLNQCRLLSNYRQE  333



hypothetical protein UGP_073 [uncultured crAssphage] (gene_65_crAssphage_JQ995537)
Sequence ID: gi|674660384|YP_009052540.1

Length: 561Number of Matches: 1
Score  Expect Identities    Positives    Gaps
110 bits(275) 1e-28  60/60(100%)  60/60(100%)  0/60(0%)

Query  1    LTQDMVNEENTRYIIRYAFDLSGKTITMPVGCELVFEGGIIENGTINLNKCKLTGMVGEE  60
```

```
            LTQDMVNEENTRYIIRYAFDLSGKTITMPVGCELVFEGGIIENGTINLNKCKLTGMVGEE
Sbjct  351  LTQDMVNEENTRYIIRYAFDLSGKTITMPVGCELVFEGGIIENGTINLNKCKLTGMVGEE  410

Phage-related protein tail component [uncultured Mediterranean phage uvMED]
Sequence ID: gi|787039080|BAR14575.1

Length: 1128Number of Matches: 1
Score  Expect Identities   Positives    Gaps

29.2 bits(64) 5.8    12/42(29%)    20/42(47%)    0/42(0%)
Query  6    VNEENTRYIIRYAFDLSGKTITMPVGCELVFEGGIIENGTIN  47
            +N +    I R ++ L G  + +P    +   G I+ NGT N
Sbjct  275  INAQQFSQIPRRSYRLRGLKVQIPHNGIVQASGSIVYNGTFN  316

//////////////////////////////////////////////////////////////////////////////////////////////////////////////////////
```

Query: gene_67_crAssphage
PSI-BLAST output (against viruses subset of NR database):

```
minor tail protein [Mycobacterium phage BrownCNA]
Sequence ID: gi|971761687|YP_009214946.1
Length: 765Number of Matches: 1

Score  Expect Identities   Positives    Gaps
40.4 bits(93) 0.051  18/38(47%)    25/38(65%)    0/38(0%)

Query  237  LSNLTIEPIVNPKIWIGTATQYAAIAQKDNNTTYIVKS  274
            ++  T   + N  +W GTA QYAAIA K+ NT Y+VK+
Sbjct  728  VAGYTAAGLTNLNLWKGTAAQYAAIATKNANTIYVVKN  765

//////////////////////////////////////////////////////////////////////////////////////////////////////////////////////
```

Query: gene_63_crAssphage
PSI-BLAST output (against viruses subset of NR database):

```
Phage tail fibre adhesin Gp38 [uncultured Mediterranean phage uvMED]
Sequence ID: gi|775456205|BAQ90228.1
Length: 353Number of Matches: 1


Score  Expect Identities   Positives    Gaps
47.0 bits(110)      2e-04  32/90(36%)    43/90(47%)    10/90(11%)
Query  13   GYNGAGMNYDGGNRRDV-----NGKANAGLTLGIIGTALGAWALFGNRRSAGASILGGAG  67
            G  G  +N +GG R D      NG+ +    LG  GTA+G +        A  S   G G
Sbjct  210  GGRGYSVNANGGTRGDGYTITGNGQKSGTRGLGPSGTAIGGYGGSAGGGGANISTQAGGG  269
```

```
Query  68   GGGML-----GDGSTNINVFGATAGSGSGA  92
            GGGML     G+GS N +V    +G+ +GA
Sbjct  270  GGGMLITGTGGNGSANTSVGAGGSGNSAGA  299
```

//

Query: gene_64_crAssphage

PSI-BLAST output (against viruses subset of NR database):

```
tail sheath protein [Vibrio phage KVP40]
Sequence ID: gi|34419589|NP_899602.1
Length: 671Number of Matches: 1

Score  Expect Identities   Positives    Gaps
35.8 bits(81) 2.3   24/93(26%)   43/93(46%)   9/93(9%)
Query  59   DGQDLDTKSYYVNSPINSLNYVAVGWSWNLSQIPYLYWGGDNTNSGAECVMFNIESMIDL  118
            +GQ +    Y+ NS    +  +A GW        Y + GG + N+GA+  MF ++ + D
Sbjct  318  NGQSIFIDEYFENSGSAYITAIAEGWKTESGA--YNFGGGSDANAGADDWMFGLDMLSD-  374

Query  119  EDKMPDIMKMNLCANWYGSLSRGHVTVECTAYK  151
                P+++  NL     G+ +   V++  T  K
Sbjct  375  ----PEVLYTNLVI--AGNAAAEEVSIASTVQK  401
```

//

Query: gene_68_crAssphage_JQ995537

HHpred output (against VOG database):

```
>[VOG6373]|KU160660-ALY09872.1| 21688..23658 + 656 aa|tail protein
Probab=88.77  E-value=0.4  Score=56.56  Aligned_cols=155  Identities=21%  Similarity=0.308  Sum_probs=93.1

Q ss_pred            CCCCCccCHHHCCHhHHHHhcccccChHHhhhhhccCCCccCCCCCCcEEEEcCCCCCCCCcEEEEeCC------EEEEe
Q lcl|gene_68_cr  464 LDGDGKVPASQLPSYVDDVLEGYYVDETHFAEKYIEDAPVYYTPEKGKIYVDISESTDYSGKTYRWSGT------KYSVI  537 (696)
Q Consensus       464 LDasGKVPaaQLPSyVDDVLEgyYad~t~Fa~~~l~afP~~aTGEsGKIYVdld~gts~TNKtYRWSGS------tYV~I  537 (696)
                      -|-+|.||.+|||..  .|-|.|-+...        +..+..|...|.+-|..|     ++++|.-|--      -|.+|
T Consensus       373 AdvtGti~Ta~LPPL--AiNe~~~vAsQ--------aaMLALTAQRGDMAIRSD-----nG~tYvLSsDsPgTLadWKEi  437 (656)
T [VOG6373]|KU16  373 ADISGTVPTSALPPL--AVNDVFTVATQ--------AEMLALTAQRGDMAIRSD-----TGKSYALSTDSPGTLADWKEL  437 (656)
T ss_pred            hhccccccccCCcc--ccccccccchhH--------HHHHHHHhhcCceeEecc-----CCceEEEecCCCcchhhHHHH
Confidence            355899999999987  34443322211        224568889999999999     7888877632      34455


Q ss_pred            cCccccCccee-----eeeccCCcceeeechhcCCCCCCcceeE---EEeccCCceEEecceeE-ecccc--cccCCCCC
Q lcl|gene_68_cr  538 SETLALGEVTG-----TAYDGGKGKKTTDIVNSLPKYIPSTQIK---LFRSVNGNIVIGSHHYE-FNNTT--NVYESKPF  606 (696)
```

```
Q Consensus       538 ssslaLgeTa~-----Ta~rGDrGK~ayD~ss~~~~p~~st~i~------~a~TNgNi~Ig~~t~~-atnvt--s~~G~~~i  606 (696)
                      -..--++++++     ...+||.|...+|+.++..+|.+.++-.   ..+++.-.-.|.-.+.. .+-.+  -+|..--+
T Consensus       438 MAAGQVQSVAGKtGvV~LVK~DVGL~~VDNTsD~aKPvSTATQTALN~KANTsHThlwadltdkp~tfapsahvH~AAD~  517 (656)
T [VOG6373]|KU16  438 MAAGQVQSVAGKTGVVALVKGDVGLGSVDNTSDAAKPVSTATQTALNAKANTSHTHLWADLTDKPATFAPSAHVHSAADL  517 (656)
T ss_pred             Hhcccceeeeccccc EEEEEecccccccccccchhccCcchhhhHHhccccccccccccccccccccCccccccchhhh
Confidence            43333333333     567999999999999999998765422   12222222111111110 11111  12222233


Q ss_pred             CCC----CccccccCccccccccccceeccc
Q lcl|gene_68_cr  607 NDG----ITFPIVSKTESGVMSAADKVKLDE  633 (696)
Q Consensus       607 ~~g-----~tf~~vadt~~g~~s~~~k~~~~~  633 (696)
                      ++|    .-.|.-+++..|.||+.+|-..|.
T Consensus       518 tsGt~~~ARLP~ATtsa~GAmSAADKt~~D~  548 (656)
T [VOG6373]|KU16  518 TSGTVPVARLPLATTSANGAMSAADKTLFDS  548 (656)
T ss_pred             ccCccchhhccccccccchhhhhhhhhhe
Confidence            333    23677788899999999988765

////////////////////////////////////////////////////////////////////////////////////////////////////////////////////
```

Query: gene_73_crAssphage_JQ995537
HHpred output (against VOG database):

```
>[VOG2791]|NC_021560-YP_008130158.1| 7571..9049 + 492 aa|tail sheath protein
Probab=83.89  E-value=13  Score=39.74  Aligned_cols=117  Identities=32%  Similarity=0.387  Sum_probs=67.3

Q ss_pred             cceEEeceeeccccccCCeEEEEEecCccccccccceEEEEecCCCCCHHHHHHHHHHHHhcc------Ccccchhheeee
Q lcl|gene_73_cr   77 ATTFSANLTIGDVNAYSDYSIMIVKKGLKFNERNRWTATIHTGLNPTANDVAKKLANQINN------NTVGHGIKASVAE  150 (287)
Q Consensus        77 attfsanltigdvnaysdysimivkkglkfnernrwtatihtglnptandvakklanqinn------ntvghgikasvae  150 (287)
                      ++.-+.-+|++-.-...-- ..++--|    ||.+  .++-.|  -+++++|-||.+.||-      +. .--.-|++.-
T Consensus        91 g~Aat~~iTvtg~at~~g~~~~v~IaG----esv~--v~va~g--d~a~~iA~~i~~aINa~~~~~~~~~lPvtA~~~a  160 (492)
T [VOG2791]|NC_0   91 GTAEIRTMTVGALPAAGGQ-AVVQIAG----ESVS--IDIAAG--TSANDTAAALANAINAYFNPLSKK-SLPFTAVAAA  160 (492)
T ss_pred             ccceeEEEEecccccccce-EEEEECc----EEEE--EEEecC--CcHHHHHHHHHHHhccccccccC-CCCeEEEecc
Confidence            3334445555544333322 2222233    4443  344444  4678999999999997      32 3345677777


Q ss_pred             eEEEEEEeccC-----CCEEE--eccccee--eEEEEEEecCCcCcccHhHHHHHHhHHHHHHhCCcEE
Q lcl|gene_73_cr  151 AKITLTAESKG-----IDYEI--LGADELV--GISVTVTATGFPAYGDAAYITDLANKAAADAGIEYT  209 (287)
Q Consensus       151 akitltaeskg-----idyei--lgadelv--gisvtvtatgfpaygdaayitdlankaaadagieyt  209 (287)
                      +.+|||+.-||     ||.++  + ++|-+  |++.+++++ -+.-|+.    ||+.-.|+-.+.-|+
T Consensus       161 ~vVtLTak~KG~~gneldv~~~yy-~~~~~p~Gv~~t~~a~-T~G~g~p----dla~~lAamg~~~y~  222 (492)
T [VOG2791]|NC_0  161 NVVTITARHKGVYAAGLDIFIPVL-EGGNVFTGANLTMATT-TPGVGNP----DLSAILAAMGDDPFE  222 (492)
T ss_pred             ceEEEEEeeccccCcceEEEeec-CCccccccEEEEEecc-CCCCCCc----chHHHHHhCCCchh
Confidence            89999999998     45544  3 33333  888888764 2333332    455555555555444
```

//

Query: gene_87_Chitinophaga_FOJF01000001 (homolog of gene_53_crAssphage)
HHpred output:

```
>1HUU_A PROTEIN HU; DNA-BINDING PROTEIN, DNA SUPERCOILING, ALPHA/BETA; 2.0A {Geobacillus stearothermophilus} SCOP:
a.55.1.1
Probab=90.86  E-value=2.7  Score=28.86  Aligned_cols=57  Identities=11%  Similarity=0.217  Sum_probs=0.0
Template_Neff=9.900

Q ss_pred             CCHHHHHHHHHHCCCCCCCHHHHHHHHHHHHHHHHHHHHcCCceecCCCCeeEEEeecCCC
Q Q_6428000        10 TSKAAYNDFCSQHPNEQISFIQYKEIILGFNTLLADHVLETGERIKLPFGLGEISIAKFRPP   71 (163)
Q Consensus        10 ~s~~~Yk~fk~~~p~~~ld~~~f~~II~~~Nk~l~~~vie~g~~ikLP~~lG~I~I~k~~~~   71 (163)
                      +..++.+...+. ..  ++..+...++..+-+.+.+.+|. .|..|.|| +||.+.+.....+
T Consensus         2 ~~~~l~~~i~~~-~~~--~~~~~v~~v~~~~~~~i~~~L~~~g~~V~l~~~~G~f~~~~~~~   58 (90)
T 1HUU_A            2 NKTELINAVAET-SG--LSKKDATKAVDAVFDSITEALR-KGDKVQLI-GFGNFEVRERAAR   58 (90)
T ss_dssp             CHHHHHHHHHHH-HC--CCHHHHHHHHHHHHHHHHHHHH-TTCCEEET-TTEEEEEEEECCB
T ss_pred             CHHHHHHHHHHH-hC--CCHHHHHHHHHHHHHHHHHHHh-cCCeEEEc-CCEEEEEEEecCc
```

//

Query: gene_90_crAssphage_JQ995537
CDD output:

```
RepL,pfam05732 Firmicute plasmid replication protein (RepL)
Pssm-ID: 253356  Cd Length: 165  Bit Score: 31.78  E-value: 0.18


                        10        20        30        40        50        60        70
                ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*.
lcl|seqsig_MTFKE YIAENIKFNSNVIYISHDLIKGYGLVKPNYRDYYNAiayLEDENIIKRTNiRNIYVVNPIYIFRGDVNKLINIISE 171
Cdd:pfam05732    WLIDNLNSNNTLIMTQREIAEETGISLETVRQTMKA---LEEGNFLKKKT-SGVYMINPDLLFKGDDTKRLNLLLQ 134
```

//

Query: Alignment is shown for WP_052747845.1 (from Geoglobus ahangari), PSIBLAST hit to gene_3_IAS_virus_KJ003983
CDD output:

```
NT_Pol-beta-like super family    cl11966
Nucleotidyltransferase (NT) domain of DNA polymerase beta and similar proteins; This superfamily includes the NT domains
of DNA polymerase beta and other family X DNA polymerases, as well as the NT domains of Class I and Class II CCA-adding
enzymes, RelA- and SpoT-like ppGpp synthetases and hydrolases, 2'5'-oligoadenylate (2-5A)synthetases, Escherichia coli
adenylyltransferase (GlnE), Escherichia coli uridylyl transferase (GlnD), poly (A) polymerases, terminal uridylyl
```

transferases, and Staphylococcus aureus kanamycin nucleotidyltransferase, and similar proteins. The Escherichia coli CCA-adding enzyme belongs to this superfamily but is not included as this enzyme lacks the N-terminal helix conserved in the remainder of the superfamily. In the majority of the Pol beta-like superfamily NTs, two carboxylates, Dx[D/E], together with a third more distal carboxylate coordinate two divalent metal cations that are essential for catalysis. These divalent metal ions are involved in a two-metal ion mechanism of nucleotide addition. Two of the three catalytic carboxylates are found in Rel-Spo enzymes, with the second carboxylate of the DXD motif missing. Evidence supports a single-cation synthetase mechanism for Rel-Spo enzymes.

The actual alignment was detected with superfamily member cd07749:

```
Pssm-ID: 325146  Cd Length: 156  Bit Score: 109.19  E-value: 4.35e-31
                      10        20        30        40        50        60        70        80
                ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
gi 919192344    1 MRALEVLCERLGGKGIRWVVGGSLSLALQGVDVQPEDIDLITDEEGAYRIQEILSDFLVKPVRYSSIGAFSSHYSIFEVE 80
Cdd:cd07749     3 LRALRKFYERLKNINVNWALTGSLSFALQGVPVEPHDIDIQTDNEGAYEIERIFHEFVNTEVRFSESEGIRSHFGELQID 82
                      90       100       110       120       130       140       150
                ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....
gi 919192344   81 GVRVEVMGGLRVHAKGRAVDLTERLDRPV-YVKIGDLTVPLSRLDDHLESYRLLDRPKDREKIQKILEQLKNQS 153
Cdd:cd07749    83 GIKVEIMGDVQKRLEDGWWETPVDLNRYKrFVEVEGMKIPVLSLEYEYQAYLKLGRKKNQIEKAEMLKKYLQQR 156
```

Note conservation of DxD motif, signature of Nucleotidyltransferase family

//

Query: gene_12_IAS_virus_KJ003983
HHpred output:

```
>4PJ1_X mHsp60, mHsp10; Human, Mitochondrial, Chaperonin, Complex, Symmetric; HET: MG, ADP; 3.15A {Homo sapiens}
Probab=98.07  E-value=1.5e-07  Score=66.50  Aligned_cols=92  Identities=28%  Similarity=0.367  Sum_probs=0.0
Template_Neff=6.900

Q ss_pred              CccccccHHHhhhhhheeccccccccccceEEEeeCC-ccceeeeeeeccceee---------eceeccEEEeCCceeecc
Q Q_1786342          1 MLKVNSIKPLFNKIVTTCDTYDNDKTKGGIIIKTNG-TIKEYQRVEAVGSTVR---------DIKVGDLVMINPTRYIVP   70 (140)
Q Consensus          1 mlkinsi~plfnkivttcdty~~dktkggiiikt~g-tikeyqrveaigstvr---------dikvgdlv~inp~ryivp   70 (140)
                       +.+...++|+.++++...+.-+ ++++|||++-... .-....+|-|+|..+.         ++++||.|++++..
T Consensus          3 ~~~~~~i~Pl~drVLV~~~~~e-~~t~gGIilp~~~~~~~~G~VvavGpg~~~~~g~~~~~~vk~GD~Vl~~~~~~----   77 (114)
T 4PJ1_X             3 GQAFRKFLPLFDRVLVERSAAE-TVTKGGIMLPEKSQGKVLQATVVAVGSGSKGKGGEIQPVSVKVGDKVLLPEYG----   77 (114)
T ss_dssp              CCCBTTEEECTTEEEEEECCCC-CBCTTSCBCCTTTSCCCCEEEEEEECSCBSCSSSSCBCCSCCTTCEEECCSSC----
T ss_pred              ccccccccEecCCEEEEEecCCc-eeccceEEecCccCCCCeeeEEEEEecCCCCCCCCCcccceeecCCEEEecCCC----
```

```
Q ss_pred             cchhhcCcccccccceeeeeccccEEEeCCeEEEEEecccccEEEcccc
Q Q_1786342       71 KHKEKRGESLKGVIGDELTFGVNFPIVEYNDKRHLLINDQDIDYIIDGEE  120 (140)
Q Consensus       71 ~hkekr~eslkgiigdelt~~vnfp~vey~~krhlli~dqdidyiidgee  120 (140)
                                    |.+         ++++++.++++.+.||=-+++++.
T Consensus       78 --------------g~~~~~~~~~~~v~~~g~~~~ii~~~dIlAvi~~~~  104 (114)
T 4PJ1_X          78 --------------GTK---------VVLDDKDYFLFRDGDILGKYVDKL  104 (114)
T ss_dssp            --------------CEE---------EEETTEEEEEEETTTCCEEECC--
T ss_pred            --------------CEE---------EEECCEEEEEEEEhHHEEEEEcchh
```

//

Query: gene_17_IAS_virus_KJ003983
HHpred output:

```
>3ZK4_B DIPHOSPHONUCLEOTIDE PHOSPHATASE 1 (E.C.3.1.3.2); THREE-DOMAIN HEME-CU NITRITE REDUCTASE, ELECTRON; HET: FUC,
GOL, NAG, PO4; 1.65A {LUPINUS LUTEUS}
Probab=99.55  E-value=7.1e-15  Score=158.36  Aligned_cols=354  Identities=12%  Similarity=0.002  Sum_probs=0.0
Template_Neff=9.100

Q ss_pred             eeeeeeeCCCCCCcEEEEEeecCC---CceecCCeEEEEecC---CCCceeeEeEeecCCCCcHHHHH------HHHHHH
Q Q_4470287      658 TTHKCIVNGLTKGDYEYRVGRDND---PYYVSEPLKFKVLAN---SDVTSFTYAXVTDQQGFNWAEYQ------AWKKSA  725 (1147)
Q Consensus      658 ~~~~~~~~~~~~~~~~~~~~~~~~---~~~~~~~~~~~~~~~---~~~~~~~~~~~~~~~~~~~~~~~------~~~~~~  725 (1147)
                     ..|.....++.+...|.|......   ...+.....|+..+.   .....+....+...........      ......
T Consensus      198 ~~h~~~l~~L~p~t~Y~Y~vg~~~~~~~~~~s~~~~F~t~p~~~~~~~~~f~~~gD~~~~~~~~~~~~~~~~~~~~~~~~  277 (571)
T 3ZK4_B        198 YIHTSFLKELWPNREYTYKLGHRLFNGTTIWSKEYHFKASPYPGQSSVQRVVIFGDMGKAEADGSNEYNNFQPGSLNTTK  277 (571)
T ss_dssp            EEEEEEECSCCTTCEEEEEEEEEETTSCEEECCCEEEECCCCTTCCSCEEEEEESCCCCCCTTCCBCTTCCCTTHHHHHH
T ss_pred            eEEEEEEeCCCCCCEEEEEEeeeccCCCeeeceeEEEEcCCCCCCCCcEEEEEEecCCCCCCCCCccccCCCCCcHHHHH


Q ss_pred             HHHHhhCCCcceEeecCccccccCccchhhHhHhhhhHHHhcceeeeCCCCccccCCCCccccCCCCcccccceeeecccc
Q Q_4470287      726 MMLSREEPDIQFTVNTGDITQSGNRVSEWLDYYDGRQYLNNLVEMFTIGNNDLCGHNATELTNGEDATSKYSHINVLRYF  805 (1147)
Q Consensus      726 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~gnh~~~~~~~~~~~~~~~~~~~~~~~~~~~~~  805 (1147)
                     ............+..++..+..+....|..+...........+++...||||......           ...........
T Consensus      278 ~i~~~~~~~d~vl~~GDl~y~~g~~~~wd~f~~~~~~l~~~~P~~~~~GNHD~~~~~~-----------~~~~~~~~s~~  346 (571)
T 3ZK4_B        278 QIIQDLEDIDIVFHIGDLCYANGYISQWDQFTAQIEPIASTVPYMTASGNHERDWPGT-----------GSFYGNLDSGG  346 (571)
T ss_dssp            HHHHTGGGCCEEEEESCCSCCTTCTHHHHHHHHHHHHHTTSCEEEECCHHHHCCTTT-----------TCSSCCSTTTT
T ss_pred            HHHHhhccCCEEEEeccccCCCchHHHHHHHHHHHHHHcCCCEEEeccCCccCCCCC-----------CccccCCCCC


Q ss_pred             ccccccCCccccCCccCCCCceeEEecCceEEEEeecccccccccccccCCCccccccccchhHhhccccchHHHHHHH
Q Q_4470287      806 TFELDSDNTYEVEWNGVKYPIYSLYSFNYGDFHFVSLNSEIAIASSKMYKDWSSDTYAGDRTFAEAANAQIEAWFIKDLQ  885 (1147)
Q Consensus      806 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~  885 (1147)
                     .....+...+..+...   ....+|.+.++..+++.++++...                        .....|..
```

```
T Consensus        347 ~~~~~~~~~f~~p~~~---~~~~~ysf~~g~v~fi~ldt~~~~-----------------------------~~~~~Q~~  394 (571)
T 3ZK4_B           347 ECGVPAQTMFFVPAEN---REKFWYSTDYGMFRFCIAHTELDW-----------------------------RKGTEQYE  394 (571)
T ss_dssp              CTTHHHHHHSCCCCSS---TTTCCEEEEETTEEEEECCTTSCC-----------------------------STTSHHHH
T ss_pred              CCCCchhhcccCCCcc---CCCceEEEEeCCeEEEEEeCCCcc-----------------------------cCCcHHHH


Q ss_pred              HHhcCccc---cCCCeEEEEecCCcccccchhcccCCcCCCCcccccccccchHHHHHHHHHhCCcEEEeCCCceeeeee
Q Q_4470287         886 KWSGKDIP---TNCAKCIVYMHEMPFTIVTYDFMKGSXSRAGSHLNTLNTNGNYRFSRLFKKFGIRLVMGGHKHTYSISK  962 (1147)
Q Consensus         886 ~~~~~~~~---~~~~w~iv~~H~p~y~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~vd~~~~gH~H~Y~R~~  962 (1147)
                       |++..+..   ...+|+|+..|+|+|...........        .......+..+.+++.+++||++++||.|.|+|++
T Consensus        395 wL~~~L~~~~r~~~~w~Iv~~H~P~~~s~~~~~~~~~~--------~~~~~~~~~~l~~ll~~~~Vdlvl~GH~H~yer~~  466 (571)
T 3ZK4_B           395 FIEKCLASVDRQKQPWLIFLAHRVLGYSSAGFYVQEG--------SFEEPMGREDLQHLWQKYKVDIAMYGHVHNYERTC  466 (571)
T ss_dssp              HHHHHHHCCTTTTCCEEEEEESSCSSCCCBHHHHHTT--------CSSSTTSSTTHHHHHHHTTCSEEEECSSSSEEEEC
T ss_pred              HHHHHHHhcccccCCEEEEEEecCCcccCcccccccC--------CcccchhHHHHHHHHHHcCCCEEEECChhHHHhcc


Q ss_pred              eccCCCcccccccCC---CCceEEEEcCCCcCcccCCC--CccccccCCccCCEEEEEEEcceEEEEEEecCCCccEEEe
Q Q_4470287         963 PIYDAPLNYITTSNK---VNSTIDIMSSMSDDLSRKPV--IQVLNAKDVQTNNYARYEVVDKITAPTYVMSQATGYKLVS 1037 (1147)
Q Consensus         963 p~~~~~~~~~~~~~~---~~g~~y~~~g~~g~~~~~~~--~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ 1037 (1147)
                       |++++++........   .+|++||+.|+||..++...  ...+.......++|.++++.+..++....+...++. +.+
T Consensus        467 ~~~~~~~~~~~~~~~~~~~~g~~~iv~G~gG~~~~~~~~~~~~~~~~~~~~~~~~Gy~~l~v~~~~~l~~~~~~~~~g~-v~D  545 (571)
T 3ZK4_B           467 PIYQNVCTNKEKHNYKGNLNGTIHVVVGGGASLAEFAPINTTWSIFKDHDFGFVKLTAFDHSNLLLEYRKSSDGQ-VYD  545 (571)
T ss_dssp              SEETTEESCCCSSEEESCCSSBEEEEECCSSSCCCCBCSCCCTTEEEEECSCEEEEEEEEETTEEEEEEEETTTCC-EEE
T ss_pred              eeecCeeccccccccCCCCCceEEEEeCCCccccccCCCCCCCceeEecCceeEEEEEEEeCCEEEEEEEECCCCc-EEE


Q ss_pred              ccccCCCCCCCChhhceecccCCCCC
Q Q_4470287        1038 NKEQPSGNAYTIPWLMSYFKASTNAS 1063 (1147)
Q Consensus        1038 ~~~~~~~~~~~~~~~~~~~~~~~~~~ 1063 (1147)
                       .+.+.+.......-......+.....
T Consensus        546 ~f~i~~~~~~~~~~~~~~~~~~~~~~  571 (571)
T 3ZK4_B           546 SFTISRDYRDILACSVDSCPTTTLAS  571 (571)
T ss_dssp              EEEEECCHHHHHSCBTTBCCCCCSCC
T ss_pred              EEEEEeCccccccccccCCCccccCC
/////////////////////////////////////////////////////////////////////////////////////////////////////////////////
```

Query: gene_18_IAS_virus_KJ003983
CDD output:

LamG super family   cl22861   Laminin G domain;Laminin G-like domains are usually Ca++ mediated receptors that can have binding sites for steroids, beta1 integrins, heparin, sulfatides, fibulin-1, and alpha-dystroglycans. Proteins that contain LamG domains serve a variety of purposes including signal transduction via cell-surface steroid receptors, adhesion, migration and differentiation through mediation of cell adhesion molecules.

```
The actual alignment was detected with superfamily member pfam13385:

Pssm-ID: 328935  Cd Length: 152  Bit Score: 33.51  E-value: 0.15
                            10        20        30        40        50        60
                     ....*....|....*....|....*....|....*....|....*....|....*....|....*...
lcl|seqsig_MGNIK_583 ITYDIFIIKVYVDGVLTAVTKETS-FPSLSDRVYFGGRIAGDKSTYlcDCNIYNFQLYDTALTDFDIM 649
Cdd:pfam13385     87 VTYDGGTLRLYVNGVLVGSSTLTGgPPSGTGGPLYIGRSPGGDDYF--NGLIDEVRIYDRALTAAEIA 152


//////////////////////////////////////////////////////////////////////////////////////////////////////////////////
```

Query: Alignment is shown for CBL38474.1 CotH protein [Anaerostipes hadrus], the best hit to gene_19_IAS_virus_KJ003983
CDD output:

```
CotH protein; Members of this family include the spore coat protein H (cotH).

Pssm-ID: 331517  Cd Length: 319  Bit Score: 68.10  E-value: 1.01e-11
                          10        20        30        40        50        60        70        80
                    ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
gi 291559674   811 GTSSLQYAVKNYKIKLKnpdgsKYKYSPFKNgilEDTFCLKADYMESSHANNtgmaKFINDELYDTKVPPQQTNSkVRTA 890
Cdd:pfam08757    5 GNSSREFPKKSYRLKFD-----KGLEQRFPG---LRKLNLNAEFNDPSLLRN----KLAYDLFRDLGVPAPRARF-VELY 71
                          90       100       110       120       130       140       150       160
                    ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
gi 291559674   891 INGfpiqlyiakdsastpVYMGVFNF--NLDKGCNKSFGLDNEitgqENCMSFEVSSNSDTSagafKNDTDESlrtDFEL 968
Cdd:pfam08757   72 VNG---------------EYYGLYLLveSVDKHFLRAHGLDKD----GNLYKADDANFSLKL----SFGDPEK---YQLK 125
                         170       180       190       200       210       220       230       240
                    ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
gi 291559674   969 RYPDEDDCTSEQitekYNVLKRLVTWVKNADETTFKNELEQYFNKEYLLKYFLQVHLFGMVDNLGKNMMLTTWDGN---I 1045
Cdd:pfam08757  126 FEKELDEGGEED----WEDLSELIDFLNDTSEAEFEAELEKYIDVDSFLDWLAFNILIGNTDSFSHNYYLYRPDENgkwR 201
                         250       260       270       280       290       300       310       320
                    ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
gi 291559674  1046 WYPqfYDLDTQLGLDNTGYLKFYSDIDITEGVYNTSGSKLWTMV--ENVFADELSAMYKKLRTSKYRLDNILKYwYDGQV 1123
Cdd:pfam08757  202 FIP--WDLDLAFGRDWRGIGLTLDFDEPEEGIANPEENVLFRRLldNPEFRARYIARLEELLDGVFTEERLEAK-IDALH 278
                         330       340       350       360
                    ....*....|....*....|....*....|....*....|....*.
gi 291559674  1124 AQIGELqYNKDMEAKYIKFKNDYLFmlhgRRSEHMKKWVKERLLYL 1169
Cdd:pfam08757  279 ALIAPA-LERDPQKWGGLETATQYE----QEVEYLKDFIRQRRRYL 319

//////////////////////////////////////////////////////////////////////////////////////////////////////////////////
```

Query: gene_20_IAS_virus_KJ003983
CDD output:

```
LRR_3 super family  cl27891      Leucine Rich Repeat;
```

```
Pssm-ID: 332712  Cd Length: 1153  Bit Score: 33.31  E-value: 1.0

                            10        20        30        40        50        60        70        80
                        ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
lcl|seqsig_MNIKI_  130 DLTGCSKLRTIEINKCDSYKELRvsnlanletvkivscagiSSVIIINcpRLRTVNIEYCDKLTTIQIN-NCKSLvggts 208
Cdd:PLN03210       652 DLSMATNLETLKLSDCSSLVELP------------------SSIQYLN--KLEDLDMSRCENLEILPTGiNLKSL----- 706
                            90       100       110       120       130       140       150       160
                        ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
lcl|seqsig_MNIKI_  209 dNYIRVANCNNINS-LDLSNNsnlkkvsidgcdrvkIETLKIHQTNITDVSSSTNLNDNMKLDLTEFS--------QLKT 279
Cdd:PLN03210       707 -YRLNLSGCSRLKSfPDISTN---------------ISWLDLDETAIEEFPSNLRLENLDELILCEMKseklwervQPLT 770
                            170       180       190       200       210       220
                        ....*....|....*....|....*....|....*....|....*....|....*....|....*...
lcl|seqsig_MNIKI_  280 -FTCYYNKSVKYIAFANNQNApIPITSTFQECSNLER--IYGCVEL----SNTSYSG*YGL-FRGCSK 339
Cdd:PLN03210       771 pLMTMLSPSLTRLFLSDIPSL-VELPSSIQNLHKLEHleIENCINLetlpTGINLESLESLdLSGCSR 837

//////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////
```

Query: gene_29_IAS_virus_KJ003983
CDD output:

```
>HTH_17      pfam12728 Helix-turn-helix domain; This domain is a DNA-binding helix-turn-helix domain.

Pssm-ID: 315411  Cd Length: 51  Bit Score: 33.18  E-value: 2.53e-03
                          10        20
                      ....*....|....*....|....*..
lcl|seqsig_MXKVI   45 LSXHQACQKLNVSRATFDNLVREGKLP 71
Cdd:pfam12728       2 LTVEEAAELLGVSRSTVYRLIRSGELP 28

//////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////
```

Query: gene_42_IAS_virus_KJ003983
HHpred output:

```
>3CT0_A Morphogenesis protein 1; Cell wall, phi29, hydrolase, infection; HET: NAG; 1.77A {Bacteriophage phi-29}
Probab=99.47  E-value=1.8e-15  Score=138.45  Aligned_cols=136  Identities=22%  Similarity=0.313  Sum_probs=0.0
Template_Neff=8.600

Q ss_pred             ccccccChHHHHHHHHHHHHCCCCHHHHHHHHHHHHHHHcCCCCCcccC-----CCCCcccceeeecccHHHHHHHHhC--CCCC
Q Q_1802265       219 KINKVNPNAVRALNYFMNKGLTREQSAGLVGNLMAETGMNIRAVN-----PYSGAYGIAQWLGSRKTALFNKYG--NNPT  291 (472)
Q Consensus       219 ~~S~~~knAk~I~~aLkk~G~S~a~AAGILGNIq~ESGgNP~AvN------~GG~A~GL~QWt~~Rf~aL~A~aG--ni~N  291 (472)
```

```
                     +.+...++++.|+++|+.+|+++.++|+||+|+++|||++|.+.+     .+++++||+||++.+...-++...  ++.+
T Consensus         10 ~~~~~~~~a~~i~~~l~~~G~s~~~aagIlgn~~~ESg~~p~~~~~~~~~~~~~g~Gl~Qw~~~~~~~~~a~~~g~~~~~   89 (159)
T 3CT0_A            10 TMSEMKVNAQYILNYLSSNGWTKQAICGMLGNMQSESTINPGLWQNLDEGNTSLGFGLVQWTPASNYINWANSQGLPYKN   89 (159)
T ss_dssp              CHHHHHHHHHHHHHHHHHHTTCCHHHHHHHHHHHHHHHHHSSCTTCBGGGCTTCTTSCBTTTTBSSHHHHHHHHHHHTCCSSS
T ss_pred              ChhHHhHHHHHHHHHHHHhCCCCHHHHHHHHHHHHHHHHhcCCCccccccccCCCCCCceeeeeecCcchHHHHHHHcCCCCC


Q ss_pred              HHHHHHHHHHHHhccCcH-------HHHHHHHcCCCHHHHHHHHHHHhccccCChhHHHHHHHHHcccccccC
Q Q_1802265       292 LDQQLDFIWHELNSSHS-------RGLRMLRQSNNPSDAAANAFGYYEFSAGPLQAVRAMNAAGKNTKWK  354 (472)
Q Consensus       292 ~~dQL~fainEl~~~y~~~~~~~~s~lk~Lk~a~d~~~AA~af~~~yErpgg~~~~~R~~~A~~~~~k~~  354 (472)
                     ...|+.++++++.....       .....+....++.+++..|+..||+|+|+.....+|+.+|+.++.+++
T Consensus         90 ~~~Q~~~~~~e~~~~~~~~~~~~~~~~~~~~~~~~~~~~aa~~~~~~yEr~~~~~~~R~~~A~~~~~~~~  159 (159)
T 3CT0_A            90 MDSELKRIIWEVNNNAQWINLRDMTFKEYIKSTKTPRELAMIFLASYERPANPNQPERGDQAEYWYKNLS  159 (159)
T ss_dssp              HHHHHHHHHHHHHHHTCSCCCSSSCCHHHHHTCCSCHHHHHHHHCCCCCCCSCCCCTHHHHHHHHHHHCC
T ss_pred              HHHHHHHHHHHHHhcCCccccccCCccHHHHHhhCCCHHHHHHHHHHHcCCCCCCCchHHHHHHHHHHhcC


>3CB7_A Lys-rich lysozyme 2 (E.C.3.2.1.17); Digestive lysozyme 2, Musca domestica; HET: ACY; 1.9A {Musca domestica}
Probab=94.85  E-value=0.024  Score=51.02  Aligned_cols=51  Identities=24%  Similarity=0.143  Sum_probs=0.0
Template_Neff=8.300

Q ss_pred              HHHHHHHHCCCCHHHHHHHHHHHHHHcCCCCCccCC----CCCccceeeeccHH
Q Q_1802265       229 RALNYFMNKGLTREQSAGLVGNLMAETGMNIRAVNP----YSGAYGIAQWLGSRK  279 (472)
Q Consensus       229 ~I~~aLkk~G~S~a~AAGILGNIq~ESGgNP~AvN~----~~GG~A~GL~QWt~~Rf  279 (472)
                     .|.+++.+.|++...++.++..+++||+|+|.+++.    ++.++||+|..+.++
T Consensus         11 ~l~~~~~~g~~~~l~~~~~ia~~ES~f~~~~~~~~~g~~~Gl~Qi~~~~~   65 (126)
T 3CB7_A            11 SLAREMYKLGVPKNQLARWTCIAEHESSYNTKAVGSLNSNGSRDYGIFQINNYYW   65 (126)
T ss_dssp              HHHHHHHHTTCCGGGHHHHHHHHHHHHSSBTTCBCCCCTTSCCEETTTTEETTTT
T ss_pred              HHHHHHHHhCCCHHHHHHHHHHHHHhhCCCccccccCCCCCCceEEEEEecccee


////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////
```

Query: gene_50_IAS_virus_KJ003983
HHpred output:

```
>3UQZ_A; SAM and Rossmann Fold, DNA; HET: SO4; 2.7A {Streptococcus pneumoniae}
Probab=100.00  E-value=4.4e-37  Score=323.75  Aligned_cols=268  Identities=12%  Similarity=-0.037  Sum_probs=0.0
Template_Neff=8.400

Q ss_pred              CCCCcccCCCCCchhHHHHHHcCccccccccccccCChhhHHHHHHhhHHHHHHHHHHhccCchhH-hhCCCcccccCCCC
Q Q_5475696       134 XXXLAILRGLDNSLPISTQQFYDPKNGKTWEKGRWNKEDIEKFKETIDAEFEAIKEALNSGKYDR-IVLPPIESLFHEKS  212 (837)
Q Consensus       134 ~~~Ls~L~GIg~~~~~L~~f~~~~~~a~e~~~~~e~L~ki~~~~~~~~e~~k~~~~~~~~~~-~~l~~i~i~~~De~  212 (837)
                     +..|.. +|+|...+.++.+++...      .......++..+...........+.......... ....+..+++.|++
```

```
T Consensus         8 ~~~~~~~~gig~~~~~~~~~~~~~~~~~~~~~~~~l~~~~~~~l~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~it~~d~~   80 (288)
T 3UQZ_A            8 IYKLKK-SGLTNQQILKVLEYGENV------DQELLLGDIADISGCRNPAVFMERYFQIDDAHLSKEFQKFPSFSILDDC   80 (288)
T ss_dssp             HHHHHH-TTCCHHHHHHHHHHHGGG------TTBCCHHHHHHHTTCSCHHHHHHHHHHCCHHHHHHHHHSSCEEETTSTT
T ss_pred             HHHHHh-CCCChHHHHHHHHHhhcC------ChHHhhcCHHHHhcCCChHHHHHHHhcCChHHHhHHHcCCCEeecCCCC


Q ss_pred             ccHHHhcCCCCCeEEEEecCHHHc----eEEEeCCCCCHHHHHHHHHHHHHHHhhCCCeEEEcCCCCCHHHHHHHHHHcC
Q Q_5475696       213 LPAKGKKGDKTSISDISKERVPKL----YEYLKNKYEEVFGELEVEDKKETKNTTSVTKIISGGQTGVDTIGLQVAKELD  288 (837)
Q Consensus       213 YP~~LkeI~dpP~VLY~kGnl~~l----IAIVGTRk~S~YG~~~A~~ia~~LA~~Git~IVSGlA~GIDtaAH~gALeaG  288 (837)
                      ||.+|+++++||++|||+||++.+    |||||||+++.+|...++.+++.|+ .|++ ||||+|.|||.+||++|+++|
T Consensus        81 YP~~L~~~~~pP~~Lf~~G~~~ll~~~~iaIvGsR~~~~~~~~~~~l~~~l~-~~~~~-ivsG~a~GiD~~a~~~al~~~  158 (288)
T 3UQZ_A           81 YPWDLSEIYDAPVLLFYKGNLDLLKFPKVAVVGSRACSKQGAKSVEKVIQGLE-NELV-IVSGLAKGIDTAAHMAALQNG  158 (288)
T ss_dssp             SCHHHHTSTTCCSEEEEEECGGGGGSCEEEEEECTTCCHHHHHHHHHHHHTTT-TCSE-EEECCCTTHHHHHHHHHHHHT
T ss_pred             ccHHHHhcCCCCEEEEEEcChhhhcCCcEEEEecCCCCHHHHHHHHHHHHHHH-CCCE-EEcCCCCcHHHHHHHHHHHCC


Q ss_pred             CCEEEEecCCCcCccccchhhHHHHHHHHhCCEEEEcCCCCCCCCccccHHHHHHHHccCCCEEEEEeCCCCcHHHHHHHH
Q Q_5475696       289 VETGGTAPKAFLREEGIDKEDVRSYGLTEITDEEQEEYTARTGKKDNYTGRTDLNVKNSDGTVYFNYGNDSTGLKATRRS  368 (837)
Q Consensus       289 G~TIAVLg~Glld~~yP~~n~~L~~~I~e~G~lISE~pPgt~p~~~~Fp~RNRIIAGLSdgtVVVEAg~kSGGSLiTA~~  368 (837)
                      ++||+|+|+|+| .+..||..++.+++.+.++|+++||++|+..+.+++|+.|||+++++|+++||++++..+| ++.|+++
T Consensus       159 g~~I~Vlp~g-l~~~~p~~~~~~~~~i~~~~lviSe~~~~~~~~~~~~~RNr~i~~lsd~vivv~~~~~~G-T~~~~~  236 (288)
T 3UQZ_A          159 GKTIAVIGTG-LDVFYPKANKRLQDYIGNDHLVLSEYGPGEQPLKFHFPARNRIIAGLCRGVIVAEAKMRSG-SLITCER  236 (288)
T ss_dssp             CCEEEECSSC-TTCCSSGGGHHHHHHHHHHSEEEESSCTTCCCCTTHHHHHHHHHHHHCSEEEEESCCTTCH-HHHHHHH
T ss_pred             CcEEEEECCC-hhhhcCHHhHHHHHHHHcCCEEEEcCCCCCCCchhhHHHHHHHHHhhcCCEEEEEEeCCCCH-HHHHHHH


Q ss_pred             HHHhCCCEEEeCCCCccHhhcHhccchhccccCccccccchHHHHHHHHHHH
Q Q_5475696       369 AEEHNKPFLLNPTAKELRQWIKDNNIKTLNVAGNRGSKLAKNNNVAKTLRDAL  421 (837)
Q Consensus       369 AleqGR~VfAVPG~i~~~~s~G~N~I~ILNvAGpreS~~agI~e~a~dIle~L  421 (837)
                      |+++||||+++|+...+..+.|+|.+.         ..++.......++.+.+
T Consensus       237 A~~~gk~V~~~~~~~~~~~g~~~Li----------~~Ga~~v~~~~~~~~~  280 (288)
T 3UQZ_A          237 AMEEGRDVFAIPGSILDGLSDGCHHLI---------QEGAKLVTSGQDVLAEF  280 (288)
T ss_dssp             HHHTTCEEEECCCSSSSTTHHHHHHH---------HTTCEECSSHHHHHHHC
T ss_pred             HHHcCCCEEEeCCCcCCcCCHHHHHH---------HCCCEEeCCHHHHHHHH
```

//

Query: gene_52_IAS_virus_KJ003983
CDD output:

vWFA cd00198 Von Willebrand factor type A (vWA) domain was originally found in the blood coagulation protein von Willebrand factor (vWF). Typically, the vWA domain is made up of approximately 200 amino acid residues folded into a classic a/b para-rossmann type of fold. The vWA domain, since its discovery, has drawn great interest because of its widespread occurrence and its involvement in a wide variety of important cellular functions. These include basal

membrane formation, cell migration, cell differentiation, adhesion, haemostasis, signaling, chromosomal stability, malignant transformation and in immune defenses In integrins these domains form heterodimers while in vWF it forms multimers. There are different interaction surfaces of this domain as seen by the various molecules it complexes with. Ligand binding in most cases is mediated by the presence of a metal ion dependent adhesion site termed as the MIDAS motif that is a characteristic feature of most, if not all A domains.

```
Pssm-ID: 238119  Cd Length: 161  Bit Score: 54.88  E-value: 1.94e-09
                          10        20        30        40        50        60        70        80
                   ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
lcl|seqsig_MRTNL  23 LDMVIAFDTTGSMA-SYIKAVRKHVIELIPKLFAANPKLKISIVAFGDycdmnsksdfgNAYQVIDLTD--NKDELIKFV 99
Cdd:cd00198        1 ADIVFLLDVSGSMGgEKLDKAKEALKALVSSLSASPPGDRVGLVTFGS-----------NARVVLPLTTdtDKADLLEAI 69
                          90       100       110       120       130       140       150       160
                   ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
lcl|seqsig_MRTNL 100 KNARNTSGGDGDefYELVIKKIVEETSW--REGSTKSVLLIADAYPHEvgysyrasisesyiienNQIDWREEAKKAAAK 177
Cdd:cd00198       70 DALKKGLGGGTN--IGAALRLALELLKSakRPNARRVIILLTDGEPND-----------------GPELLAEAARELRKL 130
                         170       180
                   ....*....|....*....|....*....
lcl|seqsig_MRTNL 178 GIKIDTMQCSNTHNSIWYKELSDITNGIN 206
Cdd:cd00198      131 GITVYTIGIGDDANEDELKEIADKTTGGA 159
```

//

Query: gene_55_IAS_virus_KJ003983
CDD output:

```
Thy1    pfam02511 Thymidylate synthase complementing protein; Thymidylate synthase complementing protein (Thy1)
complements the thy

Pssm-ID: 308231  Cd Length: 186  Bit Score: 66.88  E-value: 2.16e-13
                           10        20        30        40        50        60        70        80
                    ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
lcl|seqsig_MKLIK_ 137 YTVHFI-TSRVTMDSFRTHITLSHLGESTRYCNYNKDkfdnqlTFIIPDrydieecdkytATDFYNSEDLTDEspkynWL 215
Cdd:pfam02511     52 FTFAIEgVSRAVLRQLVRHRIASFSQQSQRYVKLDDE------DFVIPP-----------EIAKAQSPELLEL-----YE 109
                           90       100       110       120       130       140       150
                    ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....
lcl|seqsig_MKLIK_ 216 NAMINAELTYMRLLKLKESPQYARGVLPLDVKSELISCGfkDA--WDNFFDKRCANDAHPMAREIATAVRNKLQ 287
Cdd:pfam02511    110 EAMEEAYEAYEELLEKGVAREDARYVLPNATETRIVVTM--NArsLLHFLELRCCPRAQWEIRELAEAMLEELK 181
```

//

Query: gene_60_IAS_virus_KJ003983

CDD output:

```
HicA_toxin super family    cl00752      HicA toxin of bacterial toxin-antitoxin,; HicA_toxin is a bacterial family of
toxins that act as mRNA interferases. The antitoxin that neutralizes this is family HicB, pfam15919.

Pssm-ID: 321147  Cd Length: 66  Bit Score: 33.51  E-value: 7.84e-04
                            10        20        30        40
                     ....*....|....*....|....*....|....*....|....*...
lcl|seqsig_MKQYT_   8 EFIKVVRRNGFFYNRHSGDHAIYVNNRGRHISIPHKLESVIARRLIKE 55
Cdd:COG1724        11 EVIKALEKDGFQLVRQKGSHRQYKHPDGGRVTVPFHPGEDLPPGTLRS 58
```

//

Query: gene_61_IAS_virus_KJ003983
CDD output:

```
tatD_link_rSAM   TIGR04038 radical SAM protein, TatD family-associated; Members of this family are radical SAM proteins
found in about 5 percent of microbial genomes. A portion occur as gene fusions with, or adjacent to, members of the TatD
family of hydrolases (pfam01026). The TatD family may have several paralogs per genome, including TatD itself from E.
coli (a soluble protein not actually part of the twin-arginine translocation complex), which appears to act in quality
control for TAT, directing turnover of misfolded TAT substrates. The functions of TatD family hydrolases in general
(other than TatD itself, which may be exceptional within its larger family), and of this radical SAM domain protein
modeled here, are unknown.

Pssm-ID: 274936  Cd Length: 191  Bit Score: 36.44  E-value: 3.77e-03
                            10        20        30        40        50        60        70        80
                     ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
lcl|seqsig_MEENK_   4 NKKLRLLVTTLCPNKCPLCCNNSwDF---------SKLPVV--------NRWNYDEIMFTG-GEPLLFLDEVVTLAKSIK 65
Cdd:TIGR04038      7 RNSLYLNITNRCTLRCQFCPKHR-DFqvkgydlwlDKEPSAeeviaaigDPKKYDEVVFCGyGEPLLRLDVVKEVAKWIK 85
```

//

Query: gene_62_IAS_virus_KJ003983
CDD output:

```
Adenylation_DNA_ligase_like super family  cl12015  Adenylation domain of proteins similar to ATP-dependent
polynucleotide ligases; ATP-dependent polynucleotide ligases catalyze the phosphodiester bond formation of nicked
nucleic acid substrates using ATP as a cofactor in a three step reaction mechanism. This family includes ATP-dependent
DNA and RNA ligases. DNA ligases play a vital role in the diverse processes of DNA replication, recombination and
repair. ATP-dependent DNA ligases have a highly modular architecture, consisting of a unique arrangement of two or more
discrete domains, including a DNA-binding domain, an adenylation or nucleotidyltransferase (NTase) domain, and an
oligonucleotide/oligosaccharide binding (OB)-fold domain. The adenylation domain binds ATP and contains many active site
residues. Together with the C-terminal OB-fold domain, it comprises a catalytic core unit that is common to most members
```

of the ATP-dependent DNA ligase family. The catalytic core contains six conserved sequence motifs (I, III, IIIa, IV, V and VI) that define this family of related nucleotidyltransferases including eukaryotic GRP-dependent mRNA-capping enzymes. The catalytic core contains both the active site as well as many DNA-binding residues. The RNA circularization protein from archaea and bacteria contains the minimal catalytic unit, the adenylation domain, but does not contain an OB-fold domain. This family also includes the m3G-cap binding domain of snurportin, a nuclear import adaptor that binds m3G-capped spliceosomal U small nucleoproteins (snRNPs), but doesn't have enzymatic activity.

Pssm-ID: 325160 Cd Length: 325 Bit Score: 37.88 E-value: 3.66e-03

```
                            10        20        30        40        50        60        70        80
                         ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
lcl|seqsig_MEYQK_  36 WECTEKIDGTNIRIYVTMEA-----------GEGENPWLYGVTIKGRTNRaelpsklVKKLENIFlkvdWAKvfpALTPE 104
Cdd:TIGR02307     27 WVAREKIHGTNFSIIIERDFkvtcakrtgiiLPNEDFFGYHILIKNYTAS-------VKAIQDIL----ETK---AIIVV 92
                            90       100       110       120       130       140       150       160
                         ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
lcl|seqsig_MEYQK_ 105 DTVCIYGEGYGAGIQKvgsKYIKNDVNFILFDVKFN--------DWWLKREDCEDIAkkcNVDIVSLIGYMTIPQATEFV 176
Cdd:TIGR02307     93 VSVQVFGELAGPGYQK---PVVYSDKDFYAFDIKYTetsddvtlVDDYMMESFCNVP---KLKYAPLLGRGTLDELLAFD 166
                           170       180       190       200       210
                         ....*....|....*....|....*....|....*....|....*....|....*..
lcl|seqsig_MEYQK_ 177 KNGFKSRISEDKDLD--------AEGLVLRTT-CGLRFRNGERIITKIKHCDFEKFK 224
Cdd:TIGR02307    167 VENFTTDHPALVDAGnyplegntAEGYVVKHCrPGKWLRNGNRTIIKCKNSKFSEKK 223
```

//

Query: gene_64_IAS_virus_KJ003983
CDD output:

DUF3310 super family cl13237 Protein of unknwon function (DUF3310); This is a family of conserved bacteriophage proteins of unknown function.

```
Pssm-ID: 314594  Cd Length: 58  Bit Score: 53.05  E-value: 2.93e-11
                            10        20        30        40        50        60        70
                         ....*....|....*....|....*....|....*....|....*....|....*....|....*....|
lcl|seqsig_MEKKD_   8 VEHPSHYiwLKGicGIEVIDITRHM-------NFNLGNVIKYVLRSGHKSeqgmsdkqKQIEDLKKARFY 70
Cdd:pfam11753      1 VNHPEHY--GAG--GIECIDVIRAQltgeefkGFCLGNAIKYLSRAGKKN--------GEEEDLKKAKWY 58
```

See https://www.ncbi.nlm.nih.gov/pubmed/20497505 for protein experimental characterization

//

Query: gene_65_IAS_virus_KJ003983
CDD output:

DNA_pol_A super family cl02626 Family A polymerase primarily fills DNA gaps that arise during DNA repair, recombination and replication; DNA polymerase family A, 5'-3' polymerase domain. Family A polymerase functions primarily

to fill DNA gaps that arise during DNA repair, recombination and replication. DNA-dependent DNA polymerases can be classified into six main groups based upon phylogenetic relationships with E. coli polymerase I (classA), E. coli polymerase II (class B), E.coli polymerase III (class C), euryarchaeota polymerase II (class D), human polymerase beta (class X), E. coli UmuC/DinB and eukaryotic RAP 30/Xeroderma pigmentosum variant (class Y). Family A polymerases are found primarily in organisms related to prokaryotes and include prokaryotic DNA polymerase I, mitochondrial polymerase gamma, and several bacteriophage polymerases including those from odd-numbered phage (T3, T5, and T7). Prokaryotic polymerase I (pol I) has two functional domains located on the same polypeptide; a 5'-3' polymerase and a 5'-3' exonuclease. Pol I uses its 5' nuclease activity to remove the ribonucleotide portion of newly synthesized Okazaki fragments and the DNA polymerase activity to fill in the resulting gap. The structure of these polymerases resembles in overall morphology a cupped human right hand, with fingers (which bind an incoming nucleotide and interact with the single-stranded template), palm (which harbors the catalytic amino acid residues and also binds an incoming dNTP) and thumb (which binds double-stranded DNA) subdomains.

Pssm-ID: 322025 Cd Length: 593 Bit Score: 196.33 E-value: 1.66e-54

```
                          10        20        30        40        50        60        70        80
                     ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
lcl|seqsig_MIYLV_  16 NYKVIGVEESL----KLLDPLTIVGLDTETTGLDPWTKELKSIQLGN------YDFQVVIDTTTINPTLyKEYLES-NRL 84
Cdd:COG0749        2 PYGTITDLAVLnawlTKLNAAANIAFDTETDGLDPHGADLVGLSVASeeeaayIPLLHGPEQLNVLAAL-KPLLEDeGIK 80
                          90       100       110       120       130       140       150       160
                     ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
lcl|seqsig_MIYLV_  85 FVGWNLKFDLKFLFRQNIVLKNVWDGYLAEKLmwLGYPPGIHSLSLKAagENYLDIEL--DKSVRGKIIYAGLTEDV--- 159
Cdd:COG0749       81 KVGQNLKYDYKVLANLGIEPGVAFDTMLASYL--LNPGAGAHNLDDLA--KRYLGLETitFEDIAGKGKKQLTFADVkle 156
                         170       180       190       200       210       220       230       240
                     ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
lcl|seqsig_MIYLV_ 160 --IAYSANDVKYLEKIMKLQRIELAKKGLEKAIIYEN--KFVLPLAYCEYCGIKLDADKWKAKMQKDKQRVTTaldncnk 235
Cdd:COG0749      157 kaTEYAAEDADATLRLESILEPELLKTPVLLELYEEIemPLVRVLARMERNGIKVDVQYLKELSKELGCELAE------- 229
                         250       260       270       280       290       300       310       320
                     ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
lcl|seqsig_MIYLV_ 236 wlLENEpnseyifidrqgdifnGFNLEPQvKLNWNSAKQL--IpLFKKYGVNVTTeDKVKGGTKdSIDAKSLKPQKDKCS 313
Cdd:COG0749      230 --LEEE----------------IYELAGE-EFNINSPKQLgeI-LFEKLGLPPGL-KKTKTGNY-STDAEVLEKLADDHP 287
                         330       340       350       360       370       380       390       400
                     ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
lcl|seqsig_MIYLV_ 314 LIPLYLEYKEAIKVTSTYGENFLKQINPVSGRIHTNYQQMGADTTRLTSggKDKnakveyvNLLNLPANAET----RACF 389
Cdd:COG0749      288 LPKLILEYRQLAKLKSTYTDGLPKLINPDTGRIHTSFNQTGTATGRLSS--SDP-------NLQNIPIRSEEgrkiRKAF 358
                         410       420       430       440       450       460       470       480
                     ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
lcl|seqsig_MIYLV_ 390 VAENGNKWISIDYSGQETYLMASIANDEAIIKELTEGSgDIHSLTAyMSYHEIPRDtnikDIKKkyhDLRQDAKGIEFAI 469
Cdd:COG0749      359 VAEKGYTLISADYSQIELRILAHLSQDEGLLRAFTEGE-DIHTATA-AEVFGVPIE----EVTS---EQRRKAKAINFGL 429
                         490       500       510       520       530       540       550       560
                     ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
lcl|seqsig_MIYLV_ 470 NYGGDANTISKNKGIPIEEAKKIYNAYMAGFKGLKRYQDFRRKDWFNKGYIllNPLTGHKAYIYDykelLEDKKwmatld 549
Cdd:COG0749      430 IYGMSAFGLAQQLGIPRKEAKEYIDRYFERYPGVKEYMERTKEEAREDGYV--ETLFGRRRYLPD----INSSN------ 497
                         570       580       590       600       610       620       630       640
                     ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
lcl|seqsig_MIYLV_ 550 wdyyremkiacpecetvqrvrhffKRKSASEKQSINYPIQATGSMCLRVSMINFFEYLRSNNLLFKVLICVtpYDEINCE 629
```

```
Cdd:COG0749       498 ------------------------RVVRAAAERAAINAPIQGTAADIIKLAMIKVDKALKEEKLKARLLLQV--HDELVFE 552
                                650       660
                           ....*....|....*....|
lcl|seqsig_MIYLV_ 630 APAEIADTVVKVLYNCMVKA 649
Cdd:COG0749       553 VPKEELEEVKKLLKAIMENA 572
```

//

Query: gene_66_IAS_virus_KJ003983
CDD output:

pepsin_retropepsin_like super family cl11403 Cellular and retroviral pepsin-like aspartate proteases; This family includes both cellular and retroviral pepsin-like aspartate proteases. The cellular pepsin and pepsin-like enzymes are twice as long as their retroviral counterparts. The cellular pepsin-like aspartic proteases are found in mammals, plants, fungi and bacteria. These well known and extensively characterized enzymes include pepsins, chymosin, rennin, cathepsins, and fungal aspartic proteases. Several have long been known to be medically (rennin, cathepsin D and E, pepsin) or commercially (chymosin) important. The eukaryotic pepsin-like proteases contain two domains possessing similar topological features. The N- and C-terminal domains, although structurally related by a 2-fold axis, have only limited sequence homology except in the vicinity of the active site. This suggests that the enzymes evolved by an ancient duplication event. The eukaryotic pepsin-like proteases have two active site ASP residues with each N- and C-terminal lobe contributing one residue. While the fungal and mammalian pepsins are bilobal proteins, retropepsins function as dimers and the monomer resembles structure of the N- or C-terminal domains of eukaryotic enzyme. The active site motif (Asp-Thr/Ser-Gly-Ser) is conserved between the retroviral and eukaryotic proteases and between the N-and C-terminal of eukaryotic pepsin-like proteases. The retropepsin-like family includes pepsin-like aspartate proteases from retroviruses, retrotransposons and retroelements; as well as eukaryotic DNA-damage-inducible proteins (DDIs), and bacterial aspartate peptidases. Retropepsin is synthesized as part of the POL polyprotein that contains an aspartyl-protease, a reverse transcriptase, RNase H, and an integrase. The POL polyprotein undergoes specific enzymatic cleavage to yield the mature proteins. This family of aspartate proteases is classified by MEROPS as the peptidase family A1 (pepsin A) and A2 (retropepsin family).

```
Pssm-ID: 325019  Cd Length: 90  Bit Score: 31.51  E-value: 0.06
                            10        20        30        40        50        60        70        80
                     ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
lcl|seqsig_MIIET_  51 NGIALNFLIDTGSNISHIIPSVLKDIKGNKIQNNNEIAGIGAitNESSSIEVKVKDTLS----KEYDITLVISKGlensa 126
Cdd:pfam13650      6 NGKPVRFLVDTGASGTLISPSLAERLGLKPDGLAYTVRVSTA--GGTVEAALVRLDSLRlgglTLKNVPALVLDL----- 78

                            90
                     ....*....|....*....
lcl|seqsig_MIIET_ 127 kyieqstGVKIHGLLGTDF 145
Cdd:pfam13650      79 -------GDLIDGLLGMDF 90
```

//

Query: gene_67_IAS_virus_KJ003983
CDD output:

```
trimeric_dUTPase super family     cl00493
Trimeric dUTP diphosphatases; Trimeric dUTP diphosphatases, or dUTPases, are the most common family of dUTPase, found in
bacteria, eukaryotes, and archaea. They catalyze the hydrolysis of the dUTP-Mg complex (dUTP-Mg) into dUMP and
pyrophosphate. This reaction is crucial for the preservation of chromosomal integrity as it removes dUTP and therefore
reduces the cellular dUTP/dTTP ratio, and prevents dUTP from being incorporated into DNA. It also provides dUMP as the
precursor for dTTP synthesis via the thymidylate synthase pathway. dUTPases are homotrimeric, except some monomeric
viral dUTPases, which have been shown to mimic a trimer. Active sites are located at the subunit interface.

Pssm-ID: 294336  Cd Length: 142  Bit Score: 70.73  E-value: 5.22e-16
                            10        20        30        40        50        60        70        80
                          ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
lcl|seqsig_MKIKV_  35 DLRAAEDYEFEApqasilhqkdgiktgdvkFDTKVISLGLAIQLPKGLVGRIVERSS-----GVVKLNikkmGGGYIDNC 109
Cdd:TIGR00576     25 DLRAAEDVTIPP------------------GERALVPTGIAIELPDGYYGRVAPRSGlalkhGVTIDN----SPGVIDAD 82
                            90       100       110       120       130       140       150
                          ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*..
lcl|seqsig_MKIKV_ 110 YRGDndiwkIPV----TSMVKQTIHKGDRICQFHIELsqfatpwqklkwLFSKPKFIFVDSFNENNNR-GGFGSTGV 181
Cdd:TIGR00576     83 YRGE-----IKVilinLGKEDFTVKKGDRIAQLVVEK------------IVTEVEFEEVEELDETERGeGGFGSTGV 142

////////////////////////////////////////////////////////////////////////////////////////////////////////////////////
```

Query: gene_69_IAS_virus_KJ003983
CDD output:

```
RNR_activ_nrdG3  TIGR02826 anaerobic ribonucleoside-triphosphate reductase activating protein; Members of this family
represent a set of radical SAM enzymes related to, yet architecturally different from, the activating protein for the
glycine radical-containing, oxygen-sensitive ribonucleoside-triphosphate reductase (RNR) as described in model
TIGR02491. Members of this family are found paired with members of a similarly divergent set of anaerobic
ribonucleoside-triphosphate reductases. Identification of this protein as an RNR activitating protein is partly from
pairing with a candidate RNR. It is further supported by our finding that upstream of these operons are examples of a
conserved regulatory element (described Rodionov and Gelfand) that is found in nearly all bacteria and that occurs
specifically upstream of operons for all three classes of RNR genes. [Purines, pyrimidines, nucleosides, and
nucleotides, 2'-Deoxyribonucleotide metabolism]

Pssm-ID: 274317  Cd Length: 147  Bit Score: 150.19  E-value: 2.51e-47
                            10        20        30        40        50        60        70        80
                          ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
lcl|seqsig_MLKYV_  3 KYVDSKVVFAEIPDEITLAINISNCPCHCEGCHSSYL-AEDIGEPLDLQHLTNLIDSNKG-ITCVCIMGGDANPSEVDDI 80
Cdd:TIGR02826     1 KYVNEDIVFQEVPDEVSLAFYISGCPLGCPGCHSPELwHEDEGTPLTPEVLAQLLDKYRSlITCVLFLGGEWEPEALLSL 80
                            90       100       110       120       130
                          ....*....|....*....|....*....|....*....|....*....|....*...
lcl|seqsig_MLKYV_ 81 AQDIKEyYPNLKVGWYSGRDYISKDIDMSNFNYIKYGHYDKDKGPLNSKTTNQVMLEI 138
```

```
Cdd:TIGR02826     81 LKYVKE-HAGLKVCLYTGREPKDPLELVQHLDYLKTGPWIETLGGLDSPTTNQRFYDI 137
//////////////////////////////////////////////////////////////////////////////////////////////////////////////////
```

Query: gene_70_IAS_virus_KJ003983
CDD output:

```
NRDD super family        cl26133  Anaerobic ribonucleoside-triphosphate reductase;

Pssm-ID: 330954  Cd Length: 623  Bit Score: 612.78  E-value: 0e+00
                           10        20        30        40        50        60        70        80
                       ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
lcl|seqsig_MKVLK_  63 IQMEVEKILMDLAPYNVAKAYIIYRNKHEESRFIRERIDYMSNYADSDdNAASSSETDPNANVTQKNVANLDGEVYKTKN 142
Cdd:PRK08271       1 MKVQVSFEPGFEELYSKLKSKYGYEMLRLEGIQIQLKVGFISEYFFAK-NAADGSKIDANANVRHKNIATYEAELMKDFF 79
                           90       100       110       120       130       140       150       160
                       ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
lcl|seqsig_MKVLK_ 143 RIIQRQRMKDELNVLY-PEVAKQYEIDVENHIIYPHDEASvptLKFYCQADSLYPLMTEGVGNIDGVTpSPPNDLQSFSG 221
Cdd:PRK08271      80 KLINRYLVWNKIKELFgKELADEYLRQIENHEIYVHDETS---LKPYCFSYTMYPILEDGLTKIGGES-KAPKHLSSFCG 155
                           170       180       190       200       210       220       230       240
                       ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
lcl|seqsig_MKVLK_ 222 QITNLTFLLSSQCKGAVAFGEYFIALNYYIIAEFGDKWYEkldcvvtnshckvqrTVRDFIEKAFKQFIYGINQPAGNRS 301
Cdd:PRK08271     156 SFINFVFAVSSQFAGAVATVEFLVYFDYFARKDYGDDYLP---------------THRKEIENHLQIVVYSLNQPAAARG 220
                           250       260       270       280       290       300       310       320
                       ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
lcl|seqsig_MKVLK_ 302 YQSPFTNVSYYDHTYFSSLFGEFYYPDGTKPEWAAVNVLQKMFMKFFNKLRTKQILTFPVETLAMVHDGKDIIDKEYKDF 381
Cdd:PRK08271     221 YQSVFWNISYYDRNYFKAMFGGFVYPDGSTPNWEDIIALQKFFMEWFNKEREKAMLTFPVVTAALLTDDGKCKDEDFADF 300
                           330       340       350       360       370       380       390       400
                       ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
lcl|seqsig_MKVLK_ 382 CAEMYAEGHSFFTYISDSADSLASCCRLRNELAENTFNPTSGLTGVMTGSCHVITLNINRIVQDCNkayglkrnggwkeN 461
Cdd:PRK08271     301 IAKENSKGNSFFIYISDSADSLASCCRLRNEISDNGFSYSLGAGGVNTGSINVITINLPRIAQEAR------------D 367
                           410       420       430       440       450       460       470       480
                       ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
lcl|seqsig_MKVLK_ 462 TSFIRDYLISILDRVYKYHIAYKTMLYEQEEKGMFAACNGGYIHMSKLYSTIGINGLNEAARFLGLKVSNNPEYIKFLQL 541
Cdd:PRK08271     368 RDDFLEILRERVDKIHKYQLAYREIMEERIAAGMLPLYDAGFISLDKQFLTIGINGMVEAAEFMGLTVGYNEEYKDFVQE 447
                           490       500       510       520       530       540       550       560
                       ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
lcl|seqsig_MKVLK_ 542 ILGTIKEANKKHSihdSKRPFLFNSEVVPAESLGGKNYRWDKKDGYWVPEDenLYNSYFFDAHD-DTSVLDKMILHGRQT 620
Cdd:PRK08271     448 VLKVIYEANEKAS---KEYGFTFNTEFVPAENLGVKLAKWDREDGYGVPRQ--CYNSYSYVVEDaNTDALDKFKLHGKEL 522
                           570       580       590       600       610       620       630       640
                       ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
lcl|seqsig_MKVLK_ 621 AQYCDGGSACHINLEDHLSKEQYLKLIEFAVKEGTNYFTFNIPNSKCDDCGYITKHPITECPKCHSHNITWYTRVIGYLR 700
Cdd:PRK08271     523 DKYLSGGSALHLNLDERLSEEGYRKLLNIAAKTGCNYFAFNVKITICNDCHHIDKRTGKRCPICGSENIDYYTRVIGYLK 602
                           650
                       ....*....|....*...
lcl|seqsig_MKVLK_ 701 PIKAFGIDRFIEAGKRVY 718
Cdd:PRK08271     603 RVSAFSKVRQKEYPRRHY 620
```

//

Query: gene_73_IAS_virus_KJ003983
CDD output:

```
DUF1064 super family     cl05706  Protein of unknown function (DUF1064); This family consists of several phage and
bacterial proteins of unknown function.


Pssm-ID: 283910  Cd Length: 117  Bit Score: 33.99  E-value: 0.01
                           10        20        30        40
                    ....*....|....*....|....*....|....*....|....*.
lcl|seqsig_MPNNK_  91 GIRYTPDFYFRY-GKLDVYIEAKGIENDVFYIKKKLFR-KFLDDKL 134
Cdd:pfam06356      61 KIKYIADFLIYHnDGLEEVIDVKGMATKDANIKRKLFDyKYRQVKL 106
```

//

Query: gene_75_IAS_virus_KJ003983
CDD output:

```
UDG_like super family    cl00483
Uracil-DNA glycosylases (UDG) and related enzymes; Uracil-DNA glycosylases (UDG) catalyzes the removal of uracil from
DNA, which initiates the DNA base excision repair pathway. Uracil in DNA can arise as a result of mis-incorporation of
dUMP residues by DNA polymerase or via deamination of cytosine. Uracil in DNA mispaired with guanine is one of the major
pro-mutagenic events, causing G:C->A:T mutations. Thus, UDG is an essential enzyme for maintaining the integrity of
genetic information. At least five UDG families have been characterized so far; these families share similar overall
folds and common active site motifs. They demonstrate different substrate specificities, but often the function of one
enzyme can be complemented by the other. Family 1 enzymes are active against uracil in both ssDNA and dsDNA, and
recognize uracil explicitly in an extrahelical conformation via a combination of protein and bound-water interactions.
Family 2 enzymes are mismatch specific and explicitly recognize the widowed guanine on the complementary strand, rather
than the extrahelical scissile pyrimidine. This allows a broader specificity so that some Family 2 enzymes can excise
uracil as well as 3, N(4)-ethenocytosine from mismatches with guanine. A Family 3 UDG from human was first characterized
to remove Uracil from ssDNA, hence the name hSMUG (single-strand-selective monofunctional uracil-DNA glycosylase).
However, subsequent research has shown that hSMUG1 and its rat ortholog can remove uracil and its oxidized pyrimidine
derivatives from both, ssDNA and dsDNA. Enzymes in Families 4 and 5 are both thermostable. Family 4 enzymes specifically
recognize uracil in a manner similar to human UDG (Family 1), rather than guanine in the complementary strand DNA, as
does E. coli MUG (Family 2). These results suggest that the mechanism by which Family 4 UDGs remove uracils from DNA is
similar to that of Family 1 enzyme. Although Family 5 enzymes are close relatives of Family 4, they show different
substrate specificities.


Pssm-ID: 320999  Cd Length: 201  Bit Score: 134.17  E-value: 2.52e-39
                           10        20        30        40        50        60        70        80
                    ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
```

```
lcl|seqsig_MTIDE_  33 TLCPSIKDIFKAFKLCPYNNCRVIFMGQDPYCQRGVAQGVLFgnsSDTPENKLSPSLQVI----KESVINFEIPHNLitf 108
Cdd:cd10027        18 TVYPPKEDIFRAFKLTPPDEVKVVILGQDPYHGPGQAHGLAF---SVPKGVKLPPSLRNIykelKKDLGIFPPPHGD--- 91
                            90       100       110       120       130       140       150       160
                       ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
lcl|seqsig_MTIDE_ 109 dptLESWARQGILMLNSSLTTEVNKIGVHIMK-WRPFMVAFLKQMSLINPGIIYVLFGNQAQILEPYI-GKNNYVIKIEH 186
Cdd:cd10027        92 ---LTSWAKQGVLLLNTVLTVEAGKANSHANKgWETFTDAVIKALSERNKPVVFLLWGNHAQKKKKLIdIKKHLILKSSH 168
                          170       180       190
                       ....*....|....*....|....*....|...
lcl|seqsig_MTIDE_ 187 PAYFARTHKKMPYHIWHDINNKLYELYGERIEW 219
Cdd:cd10027       169 PSPLSAYRGFFGSKHFSKANEYLKEHGKKPIDW 201
```

//

Query: gene_76_IAS_virus_KJ003983
HHBLITS output:

```
>tr|E5EQ37|E5EQ37_9CAUD RIIB protector from prophage-induced early lysis OS=Acinetobacter phage Acj9 GN=rIIB PE=4 SV=1
Probab=98.99  E-value=1.2e-12  Score=128.76  Aligned_cols=162  Identities=31%  Similarity=0.619  Sum_probs=135.9
Template_Neff=2.253

Q 6452161          74 SVRDDVVYWEDVSCLSIPXKLAEAIVSAEEKHDELKISTYRNFWTLMSLNPDERCRKNLFWFLQKYGMTISRCGFFVGYR  153 (320)
Q Consensus        74 svrddvvywedvsclsip~klaeaivsaeekhdelkistyrnfwtlmslnpdercrknlfwflqkygmtisrcgffvgyr  153 (320)
                      .+..+-.|+.++.   |-.-|+..|..+-++.....  -|--|..-+-|||.++.-+.||.||+.-.+.|+.=|.|+.+.
T Consensus       263 ~I~nG~L~Y~~ie---iK~g~~~RIi~~M~~GeDFE--F~iPFLEnLMlNPS~~av~RlFDFL~ANDIeI~~DG~f~AWK  337 (430)
T tr|E5EQ37|E5EQ  263 KIQNGQLYYKGIE---IKSGLTQRIIDSMEKGEDFE--FFIPFLENLMLNPSNRAVNRLFDFLEANDIEITKDGYFIAWK  337 (430)
Confidence            3455666776653   55678888888877776543  4566766667999999999999999999999999999998


Q 6452161         154 NVDKTAEEGVFTDYHSHSFKIRIGEMVTMDRNKCDKVQEHTCSTGLHIGGRGWLKKSYYGDTG----LACLVNPADVVAV  229 (320)
Q Consensus       154 nvdktaeegvftdyhshsfkirigemvtmdrnkcdkvqehtcstglhiggrgwlkksyygdtg----laclvnpadvvav  229 (320)
                      .|..     -|-|.|+..|.-+.|..++|-||..|.-.+.|||.|||.-...+|.  .||.+.    .-|-|||+|||++
T Consensus       338 KVr~-----d~~DI~tGt~dN~Pg~~~m~Rn~vdeDD~~TCS~GLHvCSkSYL~--hyG~~snnrVV~vkv~P~DVVsi  410 (430)
T tr|E5EQ37|E5EQ  338 KVRH-----DFKDIHTGTMDNSPGQTVMMPRNQVDEDDTQTCSRGLHVCSKSYLQ--HYGTCSNNRVVKVKVNPRDVVSI  410 (430)
Confidence            7753     48899999999999999999999999999999998877763  467665    5688999999999


Q 6452161         230 PPFDSYGKLRTCAYLPME  247 (320)
Q Consensus       230 ppfdsygklrtcaylpme  247 (320)
                      |---...|+|||.|+-.|
T Consensus       411 P~DY~nAKMRtC~Y~V~~  428 (430)
T tr|E5EQ37|E5EQ  411 PVDYKNAKMRTCEYVVVE  428 (430)
Confidence            998888999999998654
```

//

Query: gene_77_IAS_virus_KJ003983
HHBLITS output:

```
>tr|A0A0A0RQP8|A0A0A0RQP8_9CAUD RIIA-like protector protein OS=Streptomyces phage Jay2Jay GN=223 PE=4 SV=1
Probab=99.40  E-value=4.7e-16  Score=166.63  Aligned_cols=262  Identities=28%  Similarity=0.512  Sum_probs=200.1
Template_Neff=1.660

Q gene_77_KJ0039   20 KTSIDPKNLEFITTLLSSNLYSDPEQSFIREIVSNAWDSHVEAGTTDVPVIVRFRKNDIHWEVTIRDFGTGLSPKRFQEV   99 (732)
Q Consensus        20 ktsidpknlefittllssnlysdpeqsfireivsnawdshveagttdvpvivrfrkndihwevtirdfgtglspkrfqev   99 (732)
                      ..+.|...+.-|..+|. .+||||-...|||--.||.|||+|||..+ ||-|.. .|..|--..|||.|.|||.+....|
T Consensus        20 ~m~vd~~s~~~~ms~lt-d~ysd~~~a~ireystna~ds~~~ag~~~-p~~~~~-p~a~~p~~~~rd~g~gls~~di~~~   96 (637)
T tr|A0A0A0RQP8|   20 AMSVDAASMAHVMSILT-DMYSDPVMAVIREYSTNAFDSHVEAGVQR-PIEVSI-PNAMSPFFKVRDYGVGLSVEDIRNV   96 (637)
Confidence            3567888888877774 79999999999999999999865 665543 57788899999999999999999


Q gene_77_KJ0039  100 YCNIGSSTKRESNDFIGGFGIGKYSSLACSNTVYITSYYNGTAFLYVMVKSGN-TITTNLVMEKPTEEKNGVEVTIKSIR  178 (732)
Q Consensus       100 ycnigsstkresndfiggfgigkysslacsntvyitsyyngtaflyvmvksgn-tittnlvmekpteekngvevtiksir  178 (732)
                      |..-|.||||++|||.|-.|+|--+.|.-....-+++--.|..-+....++.. .-+...|.-.|+.+.|||+|.|---|
T Consensus        97 ys~ygastkr~~n~~~g~lg~gck~alty~~qft~~~~kdg~~~~vaisr~~dg~g~m~~v~~~~~~~~pngv~~~~pv~~  176 (637)
T tr|A0A0A0RQP8|   97 YSKYGASTKRSTNEQVGMLGLGCKSALTYTQQFTVRSVKDGQMAHVAISRTEDGSGVMQVVHTQEVDEPNGVEISVPVHR  176 (637)
Confidence            99999999999999999999999988888888898888776666555432 33456777899999999999886544


Q gene_77_KJ0039  179 NIE-PYRRALRYIVFFP-NVYIDGIDSENNNTKLKRFNNFAVASCMIDPKILLGNVLYPCNRGM-LARESRAFLDRIEHS  255 (732)
Q Consensus       179 nie-pyrralryivffp-nvyidgidsennntklkrfnnfavascmidpkillgnvlypcnrgm-laresrafldriehs  255 (732)
                      .-. ....+.-+-.|-| .|.|||...|.-. -++--.||.+.-..--|.+.+|||-||....- +-|....+. .=..-
T Consensus       177 ~~d~~~~~~~~~~f~~~~~g~vl~~g~~~~~~~d-g~~~~d~~~~~~~~s~t~~~vmg~v~y~~~~e~~~~~~~~~~~~~-~g~~~  254 (637)
T tr|A0A0A0RQP8|  177 QNDFEWKAKEFFKFWEPGTVLIDGVEPDRFD-GLKLSDNITLVEGSNTDYVVMGNVGYRVSNENTLYRPQNGYY-HGRNF  254 (637)
Confidence            322 2233333333333 4788998776322 233446777777777899999999999975433 223111111 01235


Q gene_77_KJ0039  256 GIVIKFDVGEISITPNRENIIYSNDTIAKID  286 (732)
Q Consensus       256 givikfdvgeisitpnreniiysndtiakid  286 (732)
                      |++-+...|++..+|.||...|...|++.+.
T Consensus       255 ~~i~~~~~~g~v~f~psre~l~yt~~t~~tl~  285 (637)
T tr|A0A0A0RQP8|  255 GIIAKVAIGSVNFTPSREDLHYTEHTLKTLE  285 (637)
Confidence            8888999999999999999999998763
```

//

Query: gene_79_IAS_virus_KJ003983
CDD output:

```
TOPRIM_primases  cd01029  TOPRIM_primases: The topoisomerase-primase (TORPIM) nucleotidyl transferase/hydrolase domain
found in the active site regions of bacterial DnaG-type primases and their homologs. Primases synthesize RNA primers for
the initiation of DNA replication. DnaG type primases are often closely associated with DNA helicases in primosome
assemblies. The TOPRIM domain has two conserved motifs, one of which centers at a conserved glutamate and the other one
at two conserved aspartates (DxD). This glutamate and two aspartates, cluster together to form a highly acid surface
patch. The conserved glutamate may act as a general base in nucleotide polymerization by primases. The DXD motif may co-
ordinate Mg2+, a cofactor required for full catalytic function. The prototypical bacterial primase. Escherichia coli
DnaG is a single subunit enzyme.

Pssm-ID: 173779  Cd Length: 79  Bit Score: 48.81  E-value: 1.33e-08
                          10        20        30        40        50        60        70        80
                   ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
lcl|seqsig_LCKSK  41 NRICICSSLKDALCLWANTGIPSLAIQGEGyrMSDTAINELKKRFNRIYICLDNDEAGLK---DAIQLASKTGFiNVVLP 117
Cdd:cd01029        1 DEVIIVEGYMDVLALHQAGIKNVVAALGTA--NTEEQLRLLKRFARTVILAFDNDEAGKKaaaRALELLLALGG-RVRVP 77

                   ..
lcl|seqsig_LCKSK 118 QF 119
Cdd:cd01029       78 PL 79

//////////////////////////////////////////////////////////////////////////////////////////////////////////////////
```

Query: gene_83_IAS_virus_KJ003983
CDD output:

```
AAA_24   pfam13479 AAA domain; This AAA domain is found in a wide variety of presumed phage proteins.

Pssm-ID: 316040  Cd Length: 195  Bit Score: 74.22  E-value: 3.70e-16
                          10        20        30        40        50        60        70        80
                   ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
lcl|seqsig_MIQLP  15 NPKLMVIFGKPKSGKSSFVAAIDDNLIIDLEDGYRALAVMKVQarTARDLQEIRDAIVTKGRELHkaPYKFITIDNATRL 94
Cdd:pfam13479      1 KKLKILIYGPSGIGKTTFAKTLPKPLFIDTEGGTKSLDGDRFP--YIRSWQDFLDIIVELAAELA--DYKTIVIDTIDWA 76
                          90       100       110       120       130       140       150       160
                   ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
lcl|seqsig_MIQLP  95 EEMsvalaaelyratpmgaSWGYMTDAKGmvIKNPKTGKPmvdpkadvrqlaNGAGWLYMRKAIRQLVDMFKNLCETLIL 174
Cdd:pfam13479     77 ERL-----------------CLAYVCRQNG--KGSGIEDGG------------YGKGYGELAEEFRRLLDKLQELGKNVIF 126
                         170       180       190       200
                   ....*....|....*....|....*....|....*....|....*.
lcl|seqsig_MIQLP 175 VCHVKDKQIKKNGEEM-SEM*VDLA**TGDIICGEADAVGYLYRDG 219
Cdd:pfam13479    127 TAHAKTRKDEDPDGGKyTRYEPKLGKKTANLLKEWVDLVLFANYKT 172

//////////////////////////////////////////////////////////////////////////////////////////////////////////////////
```

Query: Alignment of 12 closest homologs of gene_85_IAS_virus_KJ003983
HHpred output:

```
>3BGW_F DNAB-Like Replicative Helicase; ATPase, REPLICATION; 3.91A {Bacillus phage SPP1}
Probab=99.85  E-value=4.4e-21  Score=204.97  Aligned_cols=287  Identities=13%  Similarity=0.061  Sum_probs=0.0
Template_Neff=10.500

Q ss_pred                CHHHHHHHHHHHHHHHhcCCCCCCCcccCCcHHHHHHhCCCCCCcEEEEECCCCCCHHHHHHHHHHHHHHHHHHHcchhH
Q gene_186_Cellu    4 KLFKDFTKEAFLELNKLQKGEKRLPKTGEDF--IDSHIGCILPGDVILISALSGQGKTETL-------------------   62 (368)
Q Consensus        16 ~~~~~~~~~~~~~~~~~~~~~~~giptg~~~~~LD~~lgGl~~G~l~vI~G~~g~GKS~~~~~la~~~~~~~~~~~~~~~~   95 (519)
                     ++.+...+.......  ..+...++|||  ||.||.+++|+.+|++++|+|.||+|||+|+++++     ..++..
T Consensus       158 ~~~~~~~~~~~~i~~--~~~~~~~i~tg--~~~Ld~~~~g~~~G~l~~i~g~~g~GKS~l~~~l~-----~~~~~~----  224 (444)
T 3BGW_F         158 SIDEALVTVYEEIES--ADGNITGVPSG--FTELDRMTYGYKRRNFVLIAARPSMGKTAFALKQA-----KNMSDN----  224 (444)
T ss_dssp                  CSSSSGGGTTTTTSS--CCSSCSSBCCS--CHHHHHHSSBCSSCEEEEEEECSSSSHHHHHHHHH-----HHHHHT----
T ss_pred                CHHHHHHHHHHHHHh--cCCCCcCcCCC--chHHHHHhhcccCCeEEEEEeCCCCCHHHHHHHHH-----HHHHcC----


Q ss_pred                HHHHhCCCChHHHHHHHHHHHHcCCEEEEEeCCCCHHHHHHHHHHHHCCCccCCcccccccccccCcchhhcccccccCC
Q gene_186_Cellu   63 -------------------QRMKKKILSLTVNPDAENYVFLDVSLEM---KVFN-------------------------I   95 (368)
Q Consensus        96 ~~~~~~~~~~~~~~~~~~~~~~g~~Vly~SlE~s~~e~~~Rl~a~l~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~i  175 (519)
                                           |.+|+|+++|++..++..|+.+...                               ++
T Consensus       225 ----------------------~~~vl~~s~E~~~~~~~~~~r~~~~~~~-------------------------------~~  251 (444)
T 3BGW_F         225 ----------------------DDVVNLHSLEMGKKENIKRLIVTAG-------------------------------SI  251 (444)
T ss_dssp                  ----------------------TCEEEEECSSSCHHHHHHHHHHHHS-------------------------------CC
T ss_pred                ----------------------CCeEEEEeCCCCHHHHHHHHHHHHc-------------------------------CC


Q ss_pred                CHHHhccCCCCCCCChHHHHhCCHHHHHHHHHHHHHHHHHhcCCCEEEEcCCCCHHHHHHHHHHHHHhcCCeeeeecccc
Q gene_186_Cellu   96 LLRGIAGTIKKKKSD-ILFKEFTEEEKEEVNKYYK---------------------GVNDKRHYLSQSPTTVD-------  146 (368)
Q Consensus       176 ~~~~i~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~l~~~d~~~~~~~i~~~i~~l~~~~g~~~~~~~~~~~~  255 (519)
                     +...+...          ...+.++......+.+.   ........+.-....++..+...++.+....+.
T Consensus       252 ~~~~~~~~----------~~~~~~~~~~~~~~~~~~~~~---~~~~~~~~~~~~~~~~~~~~~i~~~i~~~~~~~~~~~---------  309 (444)
T 3BGW_F         252 NAQKIKAA----------RRDFASEDWGKLSMAIG---EISNSNINIFDKAGQSVNYIWSKTRQTKRKNPG---------  309 (444)
T ss_dssp                  CHHHHHTG----------GGGTCCSCHHHHHHHHH---HHHTSCEEEECCSSCBHHHHHHHHHHHHHHSCS---------
T ss_pred                CHHHHHhc----------ccCCCHHHHHHHHHHHH---HHHcCCcEEEcCCCCCHHHHHHHHHHHHHHCCC---------


Q ss_pred                CCCCCCCHHHHHHHHHHHHHHHhcCCCCcEEEEecHHhcCCCCCchhHHHHHHHHHHHHHHHHHhCCcEEEEeecccccc
Q gene_186_Cellu  147 --------QFEEKVREFLTEHKDKEAVFVALDHVLLFKGA----EKQSLLEKLCEAVNQ-LKLEFSNVYFILISQNNRAL  213 (368)
Q Consensus       256 ~~~~~~~~~~~~~~~~~~~~~i~~~i~~~~~~~~~~~~~~~~~~lVIID~l~~l~~~~~~~~~~~~~e~i~~i~~~Lk~~~lA~~~gi~VI~vsQlnr~~  335 (519)
                                             .+++||||++..+.......+.++.+..++..|+  .+|+++|+++++++|.++..
T Consensus       310 ------------------------~~~lvvID~l~~~~~~~~~~~~~~~~~~~~~~~~~~~~~L~--~la~~~~~~~~vi~~~q~~~~~~~  363 (444)
T 3BGW_F         310 ------------------------KRVIVMIDYLQLLEPAKANDSRTNQISQISRDLK--KMARELDVVVIALSQLSRQV  363 (444)
T ss_dssp                  ------------------------SCEEEEEEECSSSSBCSCSSSCHHHHHHHHHHHH--HHHHHHTCEEEEEEECCGGG
T ss_pred                ------------------------CCEEEEEeCHHHcCCccccCCCHHHHHHHHHHHH--HHHHHhCCEEEEEecCCcch
```

```
Q ss_pred              cCcccCCCCCCCCCCCHHHcCCCchHHHhCCEEEEEECCcchhhccccccCCCcccchhhhhhhcCCCcCCCcccCCCCcE
Q gene_186_Cellu  214 VGRIAEKSNS--ASPNAGDVFGSSFLDQLCSFNIILYDAF--KMGISQYMKVNPDRYDYLSEYFGDEDSKGKVSFETENK  289 (368)
Q Consensus       336 ~~~~~~~~~~g~~p~l~di~gS~~I~~~AD~vi~L~r~~~~~~~~~~~~~~~~~~~~~~~~~~g~~~~~~~~~~~~~  415 (519)
                      .....       ..|.+.+++||+.+.+.||.|+.+++..  .......              ..+             .
T Consensus       364 ~~~~~~~~~~~~~~~~~~gs~~~~~ad~vi~l~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~  407 (444)
T 3BGW_F          364 EQRQD-------KRPMLSDLRESGQLEQDADIIEFLYRDD--YYDKESE--------------SKN-------------I  407 (444)
T ss_dssp             GGSSC-------CCCCGGGCCSCSHHHHHCSEEEECCBGG--GTCSSCS--------------STT-------------E
T ss_pred             hhCCC-------CCCChHHHhccCcHHHhCCEEEEEECHH--HcCCccc--------------CCc-------------e


Q ss_pred              EEEEEEecCCCCCCCCCCeeeeeeeeeEecccCEEEeCCCccccc
Q gene_186_Cellu  290 IFVHLIKTRESDTPYKDIFVIEKDVSEESKARVSFKEKTPSGPKFT  335 (368)
Q Consensus       416 i~l~i~KnR~G~~~~g~~v~l~~~~~fd~~~~f~e~~~~~~~~  461 (519)
                      ..+.+.|+|.|..   + .+.+.     |++++++|.+++.+..+.
T Consensus       408 ~~l~i~K~R~g~~~~~~~~~~~~~~~~~f~~~~~~~~~  444 (444)
T 3BGW_F          408 VEVIIAKHRDGPV---G-TVSLA-----FIKEYGNFVNLERRFDDR  444 (444)
T ss_dssp             EEEEECCCSSSCC---E-EEEEE-----EETTTTEEECCC------
T ss_pred             EEEEEEecCCCCC---e-EEEEE-----EeccceeEEeccccccCC
```

//

Query: alignment of close homologs of gene_86_IAS_virus_KJ003983
HHpred output:

```
>1LJ9_A transcriptional regulator SlyA; HTH DNA binding protein, structural; 1.6A {Enterococcus faecalis} SCOP:
a.4.5.28; Related PDB entries: 1LJ9_B
Probab=92.87  E-value=0.51  Score=36.89  Aligned_cols=71  Identities=17%  Similarity=0.188  Sum_probs=0.0
Template_Neff=9.600

Q ss_pred              ccccCCCHHHeehhhhhhcccch----------------HHHHHHHHHcCceeeecCCCC---ceEEEeHHHHHHHHHHh
Q gene_86_IAS_vi   10 XEQHNITLEEFLVLYLGAKNADI----------------KSISQEVIRKGLATRDLFSDN---RYIVVSNKVKDLIASII   70 (198)
Q Consensus        10 l~~~~is~~e~L~l~lv~k~~di----------~~~l~~L~~KG~~~~~~~~~~~~~~~vt~Kf~DLf~~~~   70 (198)
                      ++..|+|+.|+.++..+..+..+               ...+..|+++|||.+.....+   ..+.+|.+..++
T Consensus        22 ~~~~~lt~~~~~l~~i~~~~~~~~la~~~~i~~~v~~~l~~L~~~gli~~~~~~d~r~~~~~lT~~G~~~~~~~   96 (144)
T 1LJ9_A           22 FKELSLTRGQYLYLVRVCENPGIIQEKIAELIKVDRTTAARAIKRLEEQGFIYRQEDASNKKIKRIYATEKGKNV-----   96 (144)
T ss_dssp             TGGGTCTTTHHHHHHHHHHHSTTEEHHHHHHHHTCCHHHHHHHHHHHHHHTTSEEEEECSSCTTCEEEEECHHHHHH-----
T ss_pred             hHhcCCChHHHHHHHHHhhCCCccHHHHHHHhCCCHHHHHHHHHHHHHHCCCeeEecCCCCCceeeEEeCHHHHHH-----


Q ss_pred              cCCCCcccCCcHHHHHHHHHHHHC
Q gene_86_IAS_vi   71 VNSDKNIVDKDEEYTRLANKLRELY   95 (198)
Q Consensus        71 ~~~~~~~~~~~~~~~~~eL~e~Y   95 (198)
```

```
                              .......+.++....
T Consensus        97 ----------~~~~~~~~~~~~~~~~  111 (144)
T 1LJ9_A           97 ----------YPIIVRENQHSNQVA  111 (144)
T ss_dssp             ----------HHHHHHHHHHHHHHH
T ss_pred             ----------HHHHHHHHHHHHHHH


>PF06926.10 ; Rep_Org_C ; Putative replisome organiser protein C-terminus
Probab=80.61  E-value=8.8  Score=28.15  Aligned_cols=71  Identities=6%  Similarity=-0.017  Sum_probs=0.0
Template_Neff=9.200

Q ss_pred             HHHHHHHHHHHHCc--ccCCCCCeeeccccchHHHHHHHHHHHHHHh-CCccCHHHHHHHHHHHHhhc------CcchHHH
Q gene_86_IAS_vi   83 EYTRLANKLRELYP--AGRKEGTTYMWRGTVSEVARKLKTLVVKY-GYTINHDEVIKATKEYVNSFN------GNYRYMQ  153 (198)
Q Consensus        83 ~~~~~~~eL~e~YP--v~~~~Gt~~~lr~~~~~~~~klk~~y~k~~g~~~~H~~vi~~lk~~l~~~~~~~~~gn~~Ym~  153 (198)
                      .+...|.+|+..||  +++               ..-.+.+ ... ......+.++.++..+.....      ...+|+.
T Consensus         2 ~~~~~F~~~w~~yp~k~~k----------------~~a~~~~~~~~~~~~~i~~~~~~~~~~~~~~~~~~~~   65 (92)
T Q9CGS4_LACLA/1    2 SLLSEYLDSFIEFSSKNIA---------------KRAMAQV-EFMKLSSEEKKQAVIGAKNYFEWYKQENPEDKTKKFSI   65 (92)
T ss_pred             chhHHHHHHHHHhcccccH---------------HHHHHHH-HHhcCChhHHHHHHHHHHHHHHHHHhCCcccCCCccc


Q ss_pred             HHHHHHHhhcccCCCCcchHHHHHH
Q gene_86_IAS_vi  154 LLKYFILKSVKDADGNVEVKSELMS  178 (198)
Q Consensus       154 ~le~~I~~~~w~~~~~~~~e~y~~  178 (198)
                      ...+||++..|         +++..
T Consensus        66 ~~~~wl~~~~w----------~d~~~   81 (92)
T Q9CGS4_LACLA/1   66 NSYAFLESATF---------KSFQQ   81 (92)
T ss_pred             ChHHHhccccc---------cchhh
```

//

Query: gene_88_IAS_virus_KJ003983
CDD output:

```
Glucosaminidase  pfam01832       Mannosyl-glycoprotein endo-beta-N-acetylglucosaminidase; This family includes
Mannosyl-glycoprotein endo-beta-N-acetylglucosaminidase EC:3.2.1.96. As well as the flagellar protein J that has been
shown to hydrolyse peptidoglycan.
Pssm-ID: 307787  Cd Length: 88  Bit Score: 57.96  E-value: 9.47e-12

                           10        20        30        40        50        60        70
                      ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*..
lcl|seqsig_MRNXW_  67 EEALSYYGLKHkDIVYAQAILETGNFKSKVCKENNNLFG----------LYNSKKKRYCKFKHWAESVVAYkeWIQK 133
Cdd:pfam01832       5 IEAAKKYGIPA-SVALAQAILESGWGTSRLAKESNNLFGikgswkaydtDEVTVAARFRKYDSVEESIRDY--YAEK 78
```

//

Query: gene_89_IAS_virus_KJ003983
CDD output:

SPFH_prohibitin cd03401 Prohibitin family; SPFH (stomatin, prohibitin, flotillin, and HflK/C) superfamily; This model characterizes proteins similar to prohibitin (a lipid raft-associated integral membrane protein). Individual proteins of the SPFH (band 7) domain superfamily may cluster to form membrane microdomains which may in turn recruit multiprotein complexes. These microdomains, in addition to being stable scaffolds, may also be dynamic units with their own regulatory functions. Prohibitin is a mitochondrial inner-membrane protein which may act as a chaperone for the stabilization of mitochondrial proteins. Human prohibitin forms a hetero-oligomeric complex with Bap-37 (prohibitin 2, an SPFH domain carrying homolog). This complex may protect non-assembled membrane proteins against proteolysis by the m-AAA protease. Prohibitin and Bap-37 yeast homologs have been implicated in yeast longevity and in the maintenance of mitochondrial morphology.

```
Pssm-ID: 259799  Cd Length: 195  Bit Score: 120.31  E-value: 9.13e-34
                           10        20        30        40        50        60        70        80
                     ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
lcl|seqsig_MSSCG    9 VDAGCEGIKVNLYgseKGVDDVSLVTGAVWYNPFTEQVYEYPTYVQTVDYPaFTINAKDGSEFSIDPTISLKIADGKSPQ 88
Cdd:cd03401         4 VDAGEVGVVFRRG---KGVKDEVLGEGLHFKIPWIQVVIIYDVRTQPREIT-LTVLSKDGQTVNIDLSVLYRPDPEKLPE 79
                           90       100       110       120       130       140       150       160
                     ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
lcl|seqsig_MSSCG   89 VFKKYRKELADVIngtLFNYVKDAFRIQLNKYTTDEIVSNRDMVEKAIEAHLSKALLKENFQLE--QLTSgLKYPQSIVN 166
Cdd:cd03401        80 LYQNLGPDYEERV---LPPIVREVLKAVVAQYTAEELYTKREEVSAEIREALTERLAPFGIIVDdvLITN-IDFPDEYEK 155
                          170       180       190       200
                     ....*....|....*....|....*....|....*....|
lcl|seqsig_MSSCG 167 AVNAKNAAIQRAQKAQNELAVVKAEAEKKVVAAQAEAEAN 206
Cdd:cd03401      156 AIEAKQVAEQEAERAKFELEKAEQEAERKVIEAEGEAEAQ 195
```

Supplementary Note 2
CRISPR spacers matching crAss-like family genomes

spacer host genome (genbank ID)	viral contig with protospacer	CRISPR array position	spacer start	spacer end	e-value	spacer sequence (black font) protospacer sequence (red font)
JIAF01000004.1	CDZH01002743	87522	79007	79036	1.28E-04	AAACTAAAGAAGAT-AAGAAAAAGCAATTA
852462_854654_14_spacer_853503_30						AAACTAAAGAAGATGAAGAATAAGCAGTTA
GG775004.1	crAssphage_JQ995537	97065	46335	46364	1.28E-04	TAT-TTTATCCAATACCTTTTTACCATTAT
67033_67769_1_spacer_67080_30						TATCTTTATCAAATACTTTTTTACCATTAT
BAKP01000029.1	IAS_virus_KJ003983	99915	98754	98721	2.78E-07	AGCAGACTCATAATCAGAAGGTCATAGGTTCAAG
40807_41491_6_spacer_41202_34						AGCTGACTCATAATCAGAAGACCATAGGTTCAAG
JUJT01000004.1	Chlamydia_CVNZ01000007ext	82794	80442	80410	6.81E-08	GGA-TAGAGCGACAGCCTTCTAAGCTGTAGGTT
169511_170499_3_spacer_169659_32						GGATTAGAGCAACAGCCTTCTAAGCTGTAGGTT
JUJT01000004.1	CENS01015162	79855	11070	11101	6.81E-08	GGATAGAGCGACAGCCTTCTAAGCTGTAGGTT
169511_170499_3_spacer_169659_32						GGATAGTGCAACAGCCTTCTAAGCTGTAGGTT
KL544021.1	crAssphage_JQ995537	97065	84581	84553	1.17E-09	GTTATGAAGATAGAGGTTATTCTAATGAA
123220_124110_2_spacer_123344_29						GTTATGAAGATAGAGGTTATTCTAATGAA

Bacterial genomes with spacers matching crAss-like phages:

1. (sheep rumen) Prevotella sp. HUN102 P150DRAFT_scf7180000000012_quiver.4_C, whole genome shotgun sequence
2,328,889 bp linear DNA
JIAF01000004.1 GI:607832995

2. (Human oral cavity) Prevotella scopos JCM 17725 DNA, contig: JCM17725.contig00029, whole genome shotgun sequence
41,494 bp linear DNA
BAKP01000029.1 GI:602603948

3. Parabacteroides sp. 20_3 genomic scaffold supercont1.36, whole genome shotgun sequence
343,425 bp linear DNA
GG775004.1 GI:300829907

4. (isolation_source="infected leaves") Pectobacterium carotovorum subsp. carotovorum strain BC D6 B6.scaffold4, whole genome shotgun sequence
486,112 bp linear DNA
JUJT01000004.1 GI:741149445

5. (This is a reference genome for the Human Microbiome Project) Porphyromonas sp. 31_2 genomic scaffold acTiZ-supercont2.1, whole genome shotgun sequence
1,861,588 bp linear DNA
KL544021.1 GI:659424147