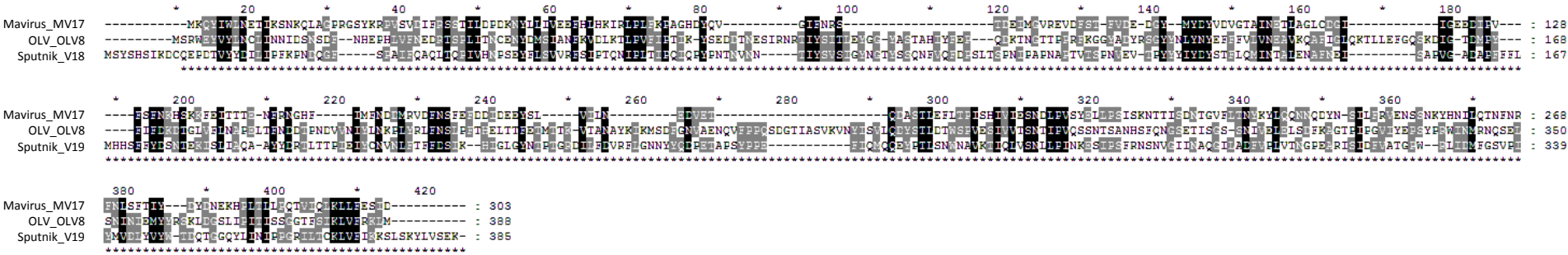
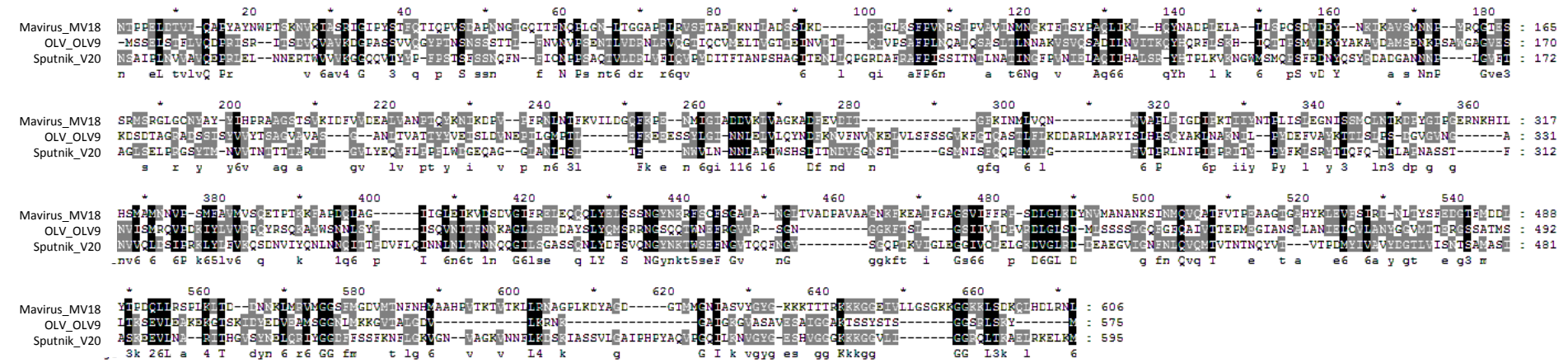
Minor capsid protein (mCP)



Major capsid protein (MCP)



Zn-ribbon domain (ZnR)

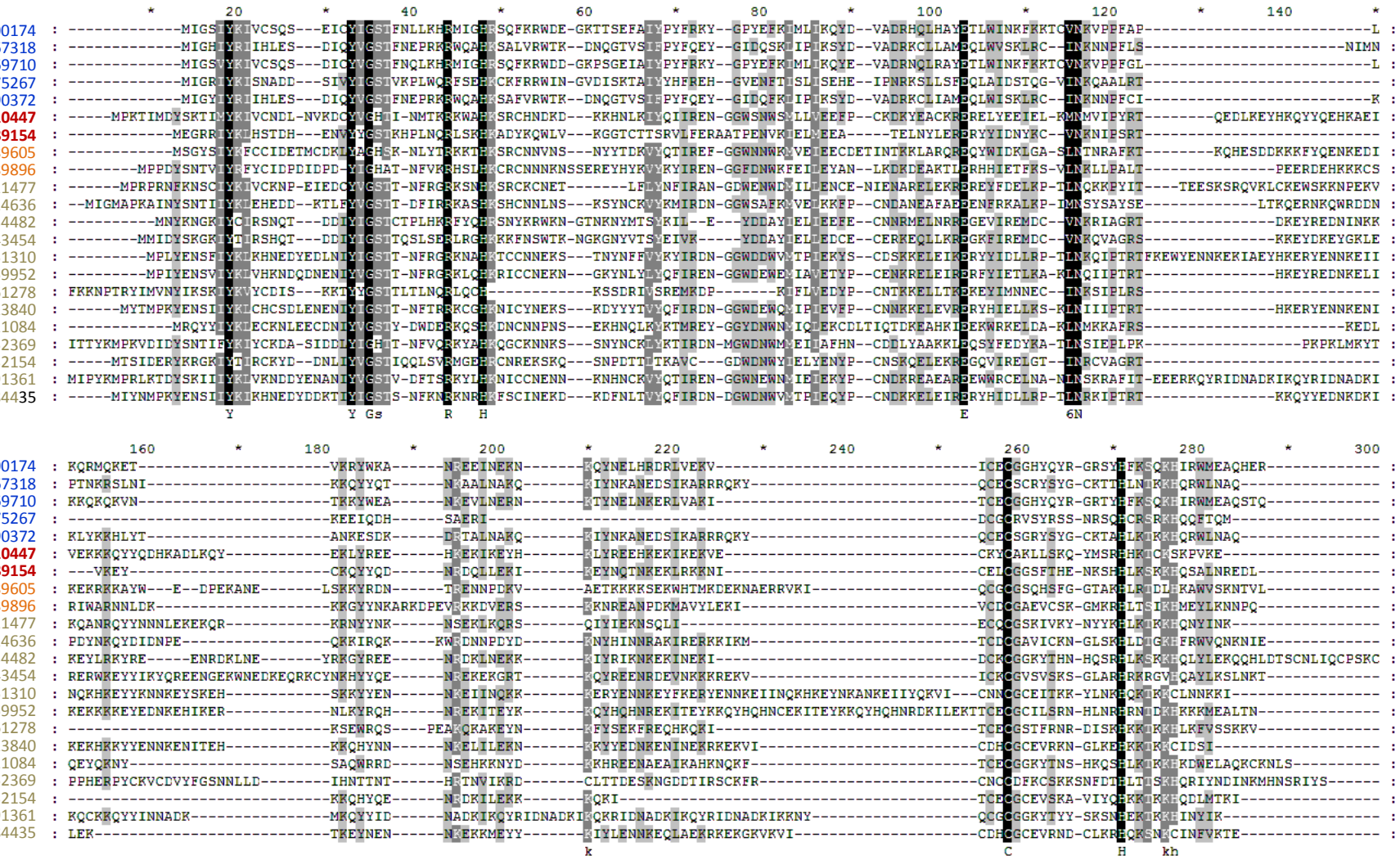
```
                                 *        20         *        40         *
env135528457 : SKAKQKYEENKDYLKKKITCE-CGSVIRITDKAKHEKTKKHIAYC-------
env136411477 : ------S--------QLIECQ-CGSKIVKYNYYKHLKTKKHQNYI-------
env136834482 : ----------KEKINEKIDCK-CGGKYTHNHQSRHLKSKKHQLYL-------
env136843454 : ------RDEVNKKKREKVICK-CGVSVSKSGLARHRKRGVHQAYL-------
env139737098 : ------RAVNN----FKIECS-CGASVSKPNIARHLKSKKHLDNI-------
env140745199 : ------KEKILEKGKEKVECP-CGSVVRKDNLPKHKKTQKHQNYL-------
env141683978 : ------REKIKEYRSQIIECE-CGSTIQLQEKARHLKSKKHINFT-------
env142032154 : ------RDKILEKKKQKITCE-CGCEVSKAVIYQHKKTKKHQDLM-------
Phyin301100174 : ------RDRLVE----KVICE-CGGHYQYRGRSYHFKSQKHIRWM-------
Physo348667318 : ------EDSIKARRRQKYQCE-CSCRYSYGCKTTHLNIKKHQRWL-------
Physo348669710 : ------KERLVA----KITCE-CGGHYQYRGRTYHFKSQKHIRWM-------
Physo348690299 : ------KERLVA----KITCE-CGGHYQYRGRTYHFKSQKHIRWM-------
Physo348690372 : ------EDSIKARRRQKYQCE-CSGRYSYGCKTAHLKIKKHQRWL-------
Dicfa328871351 : ----MFRKKVLDYKNEKIECE-CGAITSRVNRYRHLKGKIHVNNM-------
Mavirus_MV06 : ------KEKLR----KKNICELCGGSFTHENKSHHLKSKKHQSAL-------
OLV_OLV1_322510447 : ----------KEKIKEKVECKYCAKLLSKQYMSRHHKTCKSKPVK-------
Sputnik_V14 : ------KPKIKKYYSKKIECPICGALYTRSNVTNHKKSQKHIKAINNDLESN
Sputnik_V4 : NGRTYGKKPAKAQPTDRLICEFCGKTYTRSNRSTHRKTEVCKAYQ-------
tv_L3_Megco374110339 : ------KPEDKAKPNDKLICNICGGGYTRANISKHKKTQKHQIFEQMN----
tv_L5_MouMo374110361 : ------NINEKAKPKDKLICDICGGSYTRSNVSKHKNIQKHQIFESMNDKIR
tv_L8_MouMo374110364 : ------KEERKEYYSKKVRCTICGAVYARSNVSNHKKTKKHNDAMDDLEIKY
tv_R5_Megco374110341 : --------------PNDKLKCLICNGSYIRQKKSIHEKTKKHCKAL-------
tv_R6_Mimle374110348 : --------------PSDKIKCLVCNKYYKRNAKSKHEKTKTHQRNL-------
tv_R7_MouMo374110363 : --------------PNDKLICLTCGGSYIRQKNLFMKKLINIQEH--------
zn_Phagl357289605 : EKWHTMKDEKNAERRVKIQCG-CGSQHSFGGTAKHLRTDLHKAWV-------
zn_Phagl357289896 : ------PDKMAVYL-EKIVCD-CGAEVCSKGMKRHLTSIKHMEYL-------
                                         C  Cg          h     h
```

This Zn-ribbon domain is present in **all three virophages** and **all transpovirons**. In Organic Lake virophage OLV1 and Mavirus protein MV06, ZnF is fused to GIY-YIG endonuclease. Closest full-length homologs are found in *Phytophthora infestans* T30-4, *Phytophtora sojae*, *Phaeocystis globosa* virus 12T, *Phaeocystis globosa* virus 14T, and numerous environmental sequences. Multiple paralogs in Phytophtoras (possible transposable elements).

'Stand-alone' ZnF domain is found in Sputnik V14, V4, *Dictyostelium*, and transpovirons.

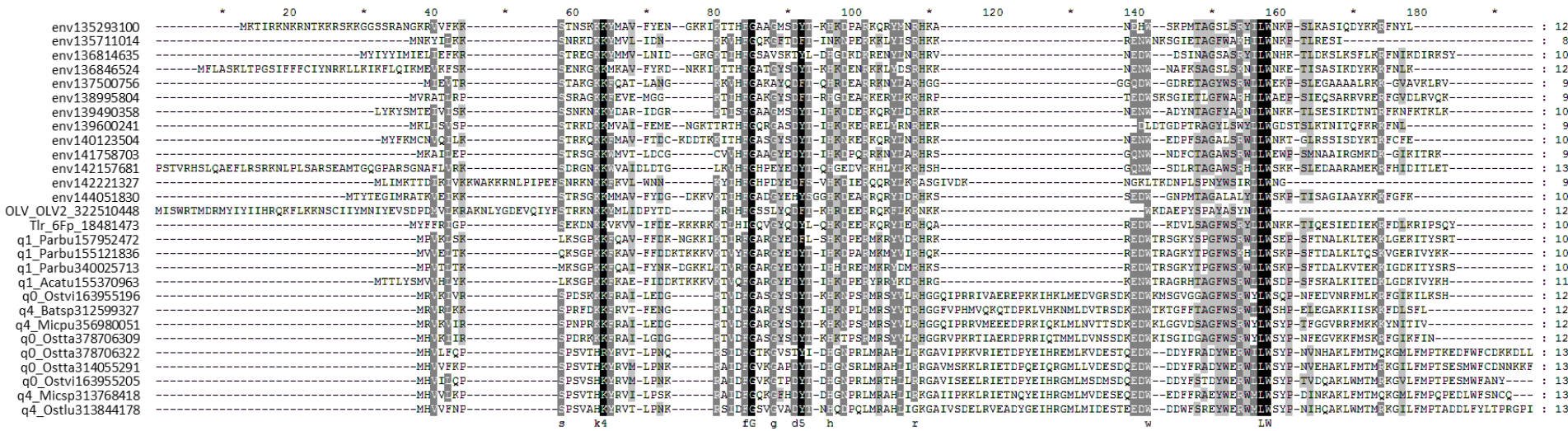
GIY-YIG endonuclease (GIY) and **Zn-ribbon domain** (ZnR)

Organic Lake virophage OLV1, MV06,
Phaeocystis globosa virus 12T (357289896, 357289605),
Phytophtoras, env seq – full length homologs



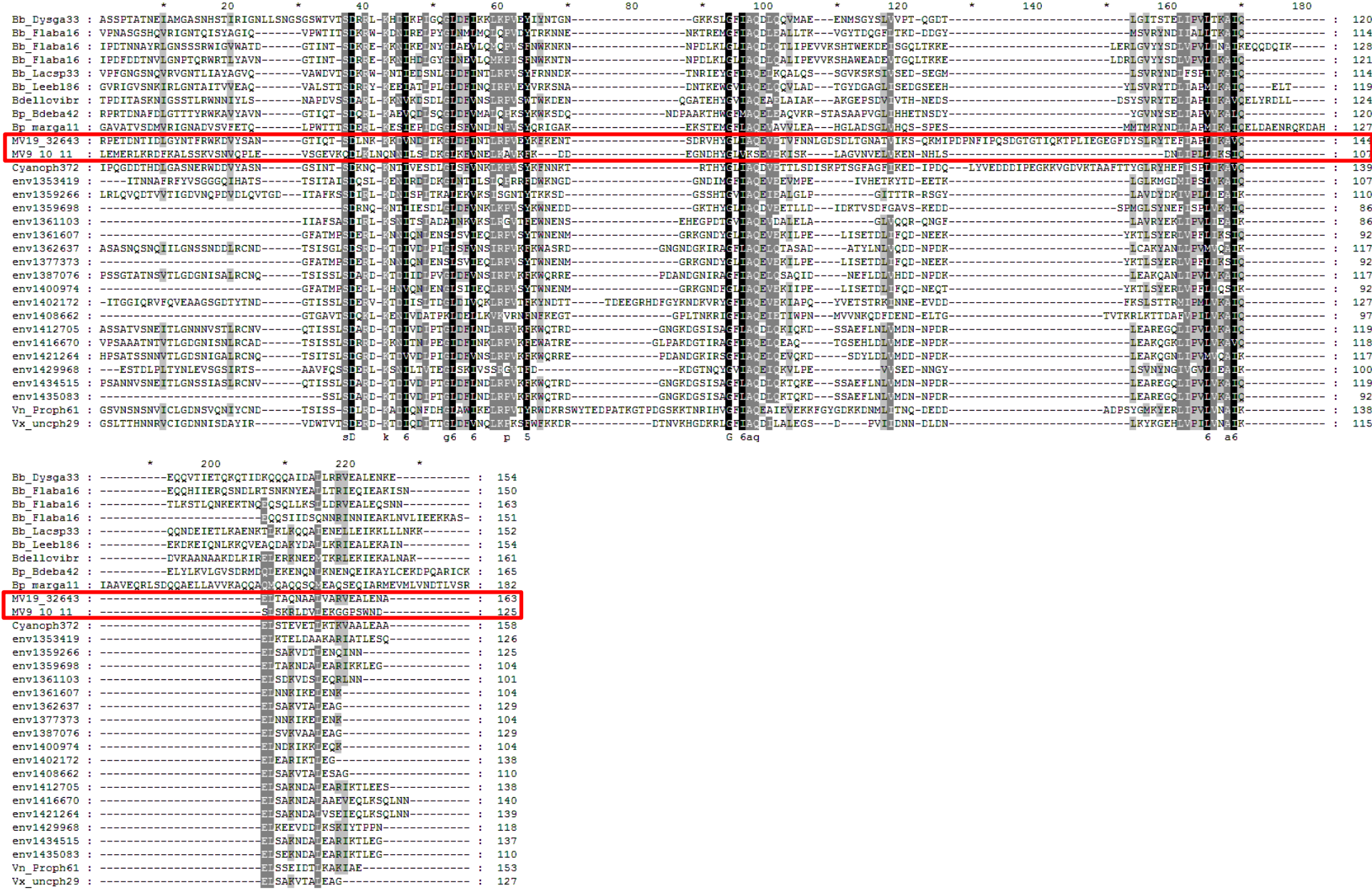
MV6_OLV1_full_len.afa

OLV2/Tlr 6F


* 20 * 40 * 60 * 80 * 100 * 120 * 140 * 160 * 180 *
env135293100
env135711014
env136814635
env136846524
env137500756
env138995804
env139490358
env139600241
env140123504
env141758703
env142157681
env142221327
env144051830
OLV_OLV2_322510448
Tlr_6Fp_18481473
q1_Parbu157952472
q1_Parbu155121836
q1_Parbu340025713
q1_Acatu155370963
q0_Ostvi163955196
q4_Batsp312599327
q4_Micpu356980051
q0_Ostta378706309
q0_Ostta378706322
q0_Ostta314055291
q0_Ostvi163955205
q4_Micsp313768418
q4_Ostlu313844178
s k4 fG g d5 h r w LW

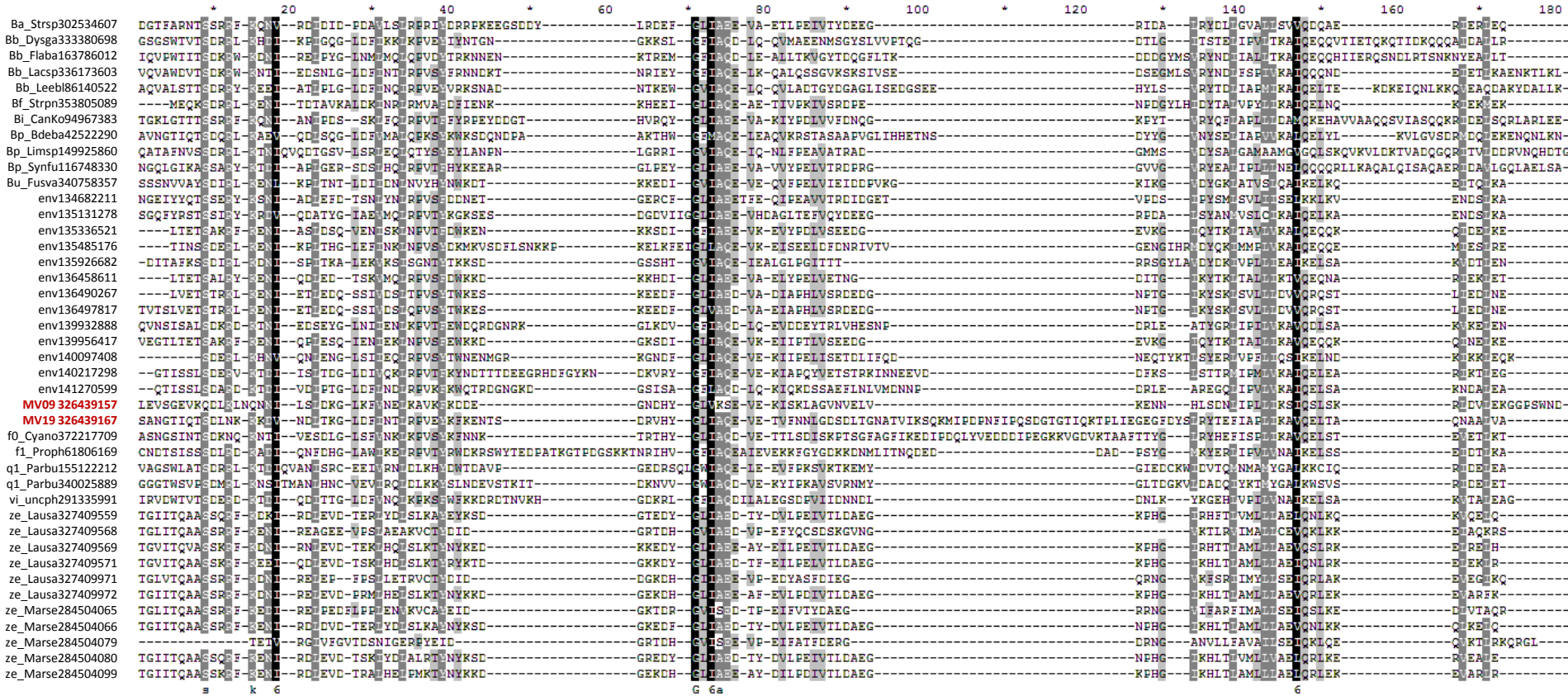
S74 superfamily Peptidase (Mavirus mv 19, mv9)

Phage (cyanophage?) origin; homologs in Marseillevirus, Lausannevirus, Paramecium virus



Peptidase (Mavirus mv 19, mv9)

With Marseillevirus, Lausannevirus and their best blast hits:



M19 N-terminal

```
                        *        20         *        40         *        60         *        80
MV19_32643 : ---MSISNLLYGNSYKLTCAG-LDAGKNKITNV-APPVDDYDCPNKAYVDSQIDSHSGTLSQVLAAGNDGNGDDMINIGEIQANTLS :  82
Polinton-1 : QTSSSACLKLDGTN---SMSANINLNSHKLINV-VDPTNAQDAATKNYVDRLIPYQHKY--KVLIIAGQSNTYAGREWGLSQSITNV :  81
El_Triad19 : HNEVIKVRGLKGNGYKINDDGDYDLESKRLTNLSSNVVDDSDATNKQYVDSRIDTKVSK--SGDVLYGNLDMNKNCIKNLHDIDPND :  85
El_Trica27 : LSASNAFFKSTPSPFVFTADGDLDLKHHRICNI-NDSTEDKDCANKVYIDLVSKQIYQK---------------------------- :  58
f2_Salph32 : MHVAEEARDLTADNLGVNNDGDLDARGRRIVNV-KDPTLPQDAATKNYVDLVSGGEGSAGKALEELAKPNGAYKIGLGDSTVGDIIG :  86
f2_Vibph32 : LHVAEEARDLAGDSIAVDDNGDLDARGRKIVNL-ADAEDDTDAVNYGMLKKFDNSTFNNAQKAEASANAAKSSENKAKDSELRSIEA :  86
f2_Yerph19 : IHVAEEARDLTADTIGVNNDGNLDARGRKIVNL-AFATSDYDAVPLKQITDRESSVWNAVTKASEQADRSNKEANRSRDEADRAKRE :  86
f2_Salph18 : LHVAEEARDLTADTIGVNNDGNLDARGRRIVNV-ADGVTDGDAISVGQVKRMNQNSWQARNESLQFRNEAEGFRNQASTSKQAAATS :  86
f2_Entph21 : MHVAEEARNLTTDTIGVNNDGHLDARGRRIVNL-ANAVDDRDAVSFGQLKTMNQNSWQARNETLQFRNEAETFRNQAEGFKDESSIN :  86
                        l          g  1    46 N6      d Da     6
```

With more polyntons:

```
                        *        20         *        40         *        60         *        80
Polinton-3 : IHDFNQ-----------ILNDKKLDTIGNDVSEIKKSIIELYDKIKV--IEFFINENSSTTTTTTRAH--------------------- :  55
Polinton-7 : LEGSQQVSRGPPGIGFNFTSDGDFDLEXKRLCNLGEP-RXPNDAITLHSLKVILQTEXNYIASKLXGIGEVIEEYKQQVEKHQLEV--- :  85
Polinton-6 : LKKNEG-DRGPPGFGFKITKDGQYDLDNKRLCNVAAP-HKLNEAVDLGTLHTIHQMEHKKVRDVTDKLREELKDLDSLIEAHRDEM--- :  84
Polinton-4 : SNRSKG-RRGPPGVGYKLTADGHFDIGSRRLCNVAEP-LELNDAVNLAVLQRSINSEIRSVHEITSDLRQSIDAVELAVQIAKDEN--- :  84
Polinton-1 : QTSSSA-CLKLDGTN---SMSANINLNSHKLINVVDP-TNAQDAATKNYVDRLIPYQHKYKVLIIAGQSNTYAGREWGLSQSITNV--- :  81
El_Triad19 : HNEVIK-VRGLKGNGYKINDDGDYDLESKRLTNLSSNVVDDSDATNKQYVDSRIDTKVSKSGDVLYGNLDMNKNCIKNLHDIDPND--- :  85
MV19_32643 : MSISNL----LYGNSYKLTCAG-LDAGKNKITNVAPP-VDDYDCPNKAYVDSQIDSHSGTLSQVLAAGNDGNGDDMINIGEIQANTLS- :  82
El_Trica27 : LSASNA-FFKSTPSPFVFTADGDLDLKHHRICNINDS-TEDKDCANKVYIDLVSKQIYQK---------------------------- :  58
f2_Salph32 : MHVAEE-ARDLTADNLGVNNDGDLDARGRRIVNVKDP-TLPQDAATKNYVDLVSGGE-GSAGKALEELAKPNGAYKIGLGDSTVGDIIG :  86
f2_Vibph32 : LHVAEE-ARDLAGDSIAVDDNGDLDARGRKIVNLADA-EDDTDAVNYGMLKKFDNSTFNNAQKAEASANAAKSSENKAKDSELRSIEA- :  86
f2_Salph18 : LHVAEE-ARDLTADTIGVNNDGNLDARGRRIVNVADG-VTDGDAISVGQVKRMNQNSWQARNESLQFRNEAEGFRNQASTSKQAAATS- :  86
f2_Entph21 : MHVAEE-ARNLTTDTIGVNNDGHLDARGRRIVNLANA-VDDRDAVSFGQLKTMNQNSWQARNETLQFRNEAETFRNQAEGFKDESSIN- :  86
f2_Yerph19 : IHVAEE-ARDLTADTIGVNNDGNLDARGRKIVNLAFA-TSDYDAVPLKQITDRESSVWNAVTKASEQADRSNKEANRSRDEADRAKRE- :  86
                                 g  1     6 n6        da     6
```