

[illegible]

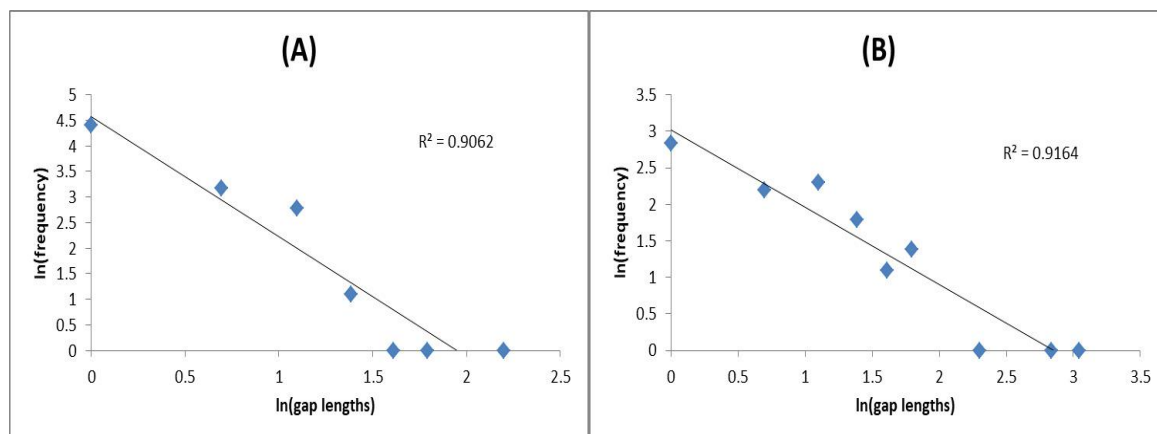


Fig. S7: gap lengths distribution of the third alignment. A) Using our method. B) Using EMBOSS Needle program.

NeseNPV	MENNILGLMKTIANITVPQ-----KHLFYSTDDIYVKTNEIT
BmNPV	--MIDNILQFFLKDVPEDKTYEINNLDQANHLIIRNTRTGTRRLFYVNNFQQFLNTIR
CpGV	--MSVDDIMMSY-ESVDRVE-----RVSCKYWYFTNGNK
AdorGV	--MSVEDIFKTY-ESVPEPK-----KNCNKYWFVDTSS
	: .:: : : :
NeseNPV	FKLHSDCIAT-----YLSDALHKSNDINDHNLTCINFCEM-LIHPFV-EKTFDQ
BmNPV	NNFNGPCAKHDMGSSCEDTEEATEKQAVQQTLDGHDWVLESNDFCI-FVKPFILKKHYEV
CpGV	K-----VRH---QCRNVEFFIK-LMTSVARRAHEWCMVPNYNS-KVVPFMPYDDYQR
AdorGV	KK----CVRR----PLGNTELFKTQLQSADSTLQHSWCMVPNYFLMTKIKPFMLCSDYEQ
	. * : * : **: . ::
NeseNPV	M--DALKNLLAADVSRYTTNQFQPKFDEHRDYFVCSQ---IKVTAFAWYFFMHKHFNLH
BmNPV	I---QKYINFENFFKSTDPGYINKCVQAGDYYPWPWPKGQAFSFGWQLFLNIKFGIV
CpGV	LNATHPHVFKFVDAFNNT---ISNTKRCGDYIVWPE---MSANFLGWVVYLHLNCQYQ
AdorGV	V--PDKSIFNFTDIKKRI---ICFPVKMGDYMVWQN---TNVTFLSWALYLYANTKYK
	: : : . ** : . .* .:: :
NeseNPV	I-LEVPLPKHVRLGNVSLFTLKNDEKNSNAQIFDIELQIVKQ-----NGVLSFINS
BmNPV	IEPTIPIIHNNKLGPDVDFV-----FDPKYFLNVELSLRTNHDPPQTLFVNGKTKFDES
CpGV	LPLEIPLIHHTRLGAFNLV-----SNAEYNLDVSCKVKEG-----SGERPFSNT
AdorGV	IDCLIPVAHNRTIGNFNLI-----NDSTNCLDLACKIYDI-----NNDLIFTNT
	: *: :: :* ..* . : : : : : : : : : *
NeseNPV	ESITEPSNMSKLIVFRHDDIMSVLLHGKHLILQNQYQDIFKFIKHNVSFYQRIVSTNKY-
BmNPV	HED--LF----ILKMADGTVVTCCKVNGELV-NSD--KNFFNYIRDDI-NLEECITVPKYK
CpGV	FQDTQLT----LFTVHSTDNINIMFSKECVYSHE--KKWFEYMKVTE-TISRCDFLPEYQ
AdorGV	DKTPKIE---MMTVQSKSIIKIYFGSNVYVTTN--KVWYNFLKQNE-NVASVTFLEPYN
	. : : . : . : : : : : : : : *
NeseNPV	FKLDANEIKYMNDNFENFIQDQEIANNDSPLCLFNDLH-GFSIHKQIETAIQNIIEFEN
BmNPV	HIVN-VNLKSLRVFEDNNFDKNDVD-----LSDTRSRK-----PRI-VPIISASSENAD
CpGV	FLCDHIDFNELRWLPEYYVKDSGMG-----LVVDNVKVGVSVGVERM-TPCSNYELDIN
AdorGV	FVCDKILFQNLN-YDDEKTPKIIRKEQ---CVTPSTSLQITDHV---SPCSNYEAEFNV
	. : : : : : : : : : : :
NeseNPV	KLIDKLN-----LKSDWSCEDTIYFKTAHFLVNNDFACFEYILSYVYQ---K
BmNPV	YIQTQINLGLIAIYENMVKVLATHEQANDPNLLQ--QYFEKSKFKNDFLIYVLWKILTK
CpGV	TVNECYNF----INETLMK--KCTFAHNETDLLS--WYLFSTGYSTFYILLISTWNYCKT
AdorGV	YINECLEY----INETMIS--KGSATNNIAKLLE--FYLHSSEYSTFYILIIALWQYCER
	: : : . : : : : * : : :

Fig. S8: Multiple sequence alignment [AdorGV, BmNPV, CpGV and NeseNPV] using EMBL-EBI MUSCLE program

NeseNPV	MSQLHIISNQDEHKLFLNIFMQKLFSDNYD-CDNNVLKQII-TC--YIDL DYNLFAKFCQS
BmNPV	NENF--SYRETDIKLFLELLCESLFA----CDKEILNEALKRCEPYKKQEKVFNRTCNH
CpGV	AIKMHNQYSVHDILFFLKTLCDRAGNHDLYVDNLIYFSSNTAAENFMESLQFFVNPEEG
AdorGV	CINMENDYTIEDMLYFVKILCDKVDGGYKLFLENLIYFGNKNCCKAFMKSCLFFVDPCRG
	:: : *:: : : . ::: . . *
NeseNPV	FRFISDDSLCS-----YFALKIQGF-----VIEKHWKLSDYIVKKIDPEMLVSYC
BmNPV	W-FDFDDTKLCVSLGYYYGIHYMIYLTLSAKNETLDDDELWAYTYENVMALNLPDIVCK
CpGV	V-EEYFNAISS-----YFAIHLAVK-----NNDEKAPLNDYTVKMCGVDDNVKSM
AdorGV	K-EAFFESICC-----YFALHFSIY-----EKTGDWRITASTIKQSEVEDDVKSV
	:: . *::: . :
NeseNPV	GWFWVYK--TEDSIYCFSEYFKQVSKSRNDECSKIAYKSKLIYTPDICFNDSKFIYV
BmNPV	GFFRKLENVVTGVNLVFNNGNYQIVKKE---DDLFLKLVKSNC--YKLSNIKFNNWKYLYL
CpGV	GFFKKIK--VSKFDYIFNGNLYVYYKSKK--ENLLAAQFAKCPDVTVSSLIFNKTILNFYM
AdorGV	GFYKKIK--VGKFDYIFTGNIYENYKSKR--DHTLAFLMADCPEISVSGLTFNDFINFYH
	*:: : . *. . : ... : . :::: **. :*
NeseNPV	TENGVFNTITNKFVDFAPFVIANTMFLCALQ-NGQVYNISKYFFDDMLHNSKKEMLSMEL
BmNPV	TTYGVYNLFTNSFHSNCPFLGTTLTPQTLKKPTDEEY-LPEDAFNYMLSTSTDELSIYRI
CpGV	TEYGVFDVCKKIYREFCPFLVTSTLKQSFIT-KEQEY-LNKTVFDRLYSAIDKDLTLFKM
AdorGV	TQDGLFDVCRKTYKEPCPFVVMSTLRKNFIK-NNQVF-LEKSLFDRIYNITDLDLLKM
	* *::: : : . **:: .*: . : : : *: : :: :
NeseNPV	YHAAHIPRKMRR-----ICALIRLGLYTND--KLVNENASKIVDIFNHLHLCSHICLL
BmNPV	YHIAKMRDVKMLKTNMAIVNYMGNCNTCQA---DMRVALNNLFRDL--WNLD--DENLI
CpGV	YHARKFLDDYEM-----VCDNLRKCVLVGESMMEEKSKWEGRWRELLVWLLECKASDIV
AdorGV	YHARKFMTEYEQ-----TILCLKECKLANAS--NSVAILEEKIHEMVVWLLECKGSDIL
	** :: . : . ::: . ::
NeseNPV	AMIKFT-----APEITNLLCND-----GITLSDQTLQILVKYSLCSA
BmNPV	TLALYVNKNKVSDMLHNLKCKPCRSTVSGSRPKCKCYKKIKINRKALKVCLIADMFNGDA
CpGV	IMMIKR-----SGQLQNVLNNIVALS-----ECDLMGLQVAMICHLLWPNS
AdorGV	ILMMYL-----KDRLLQYINNIMKTAV-----DVDLMCMQSAIVAHLLWPNS
	: : : : :: .:
NeseNPV	HIKLLFLIVYLS--DNDNKLIEALGETVFDINANLLINRKNDVLKNL-TSLTQMKNIKQ
BmNPV	ELSKLIWMLIFTN-KTYVSTTLIRTNSEFVNQHGEFFFKHEHNKIIQYLYQTIHKIEYVDM
CpGV	KVELFFWALIFNAYQDYEDWLEEYNVGSLIHEN---VFNNKKKITEAMHRFLYNLSYDDE
AdorGV	KIEIFLWSLLYNNFFDYEDSLMDYECDELITKT---MFDNKKKIIIECVHAYLYQLDYSKC
	:: :: : . . .: ::: : : : :: .

Fig. S8 (Continued)

NeseNPV	L-ISHIYDSSLWSKNVLLIQNEIDPTRKKFKSD-VVDNPSVLLDQLVQ---QLTICNIWK
BmNPV	L-MDKFNDKRLFLTE--LRDDVAREPDVQFEES---DNICKFYTHHTDALIILKKYNVWW
CpGV	VSIERVQD-----LVDEVGDPDKRKYEQKRIVKNIGAERYKYNK---LRQQCNVWS
AdorGV	L---PFND-DLFEEIKTMVFNIPLADDKKFNKKRILNNVVELYEKYEN---YPDTYNVWS
	: . * : : : : . * : . * : *
NeseNPV	DKLILPVSNETTLQWLIRFYMRCIFPHLRGFVNYATNPHMINIAEAYCYMLIANNFAYDH
BmNPV	DKIILARSTDDLPTWLTRFYMRIIMSKMD--LKEYSYNYLKKIVEGYLYFKRFTNFNHAN
CpGV	DLLIVHQPGENMYDWLTRFYIRSFLNDYTGDMKE----LYNVVLGFSYFRVFTNFHSNN
AdorGV	DYLIVQNKNEDEMYTWLTRFYVRIFLNDLD--CNDANSRRLHNVVKGCLCYFRVFTNFHCNN
	* : * : : ** * : * : : : : . * : . * : *
NeseNPV	VLYLLTFAASLNIPQNYEKLLLLVLTGSSNAGKTSFISLLG--KIAVKMPSNTLFIDNRGN
BmNPV	AIMLMHFAASLAIPVDYGKKAIIYMPGEPGSGKSSFFELLD-YLVLMHKFDDDNHSGESNK
CpGV	SKALINFCASFAIPINNEKMCLVLSSKPNCCKSSWELFSKNKILVYKQDKEEY--KHSK
AdorGV	SKVLINFCASLAIPVDEKMCIVISSKPNCCKSSWELLS-NIILVHKQDKEVY--KHNN
	* : * . * : * : * : : . . . * : * : : . . : *
NeseNPV	GPSEMEFSRAISQLYEINEVHKTTAENIKNTADLSREVVRHAHSACQKIMLQYKPILCN
BmNPV	ETSDKEVSKLNSQLYVINELKQCSSEYFKKHADSSKSDSKSRKYQGLLKYEANYKMLIVN
CpGV	SERDEKVKLYESQLYVMNEAQKFSTFLKTIVDSTRMDSARCNYGIVEMFNITFKALVCN
AdorGV	NERDEKVKLYESQLYVMNEAQKFTKAYLKSIVDNNRTDSARCNYGVMETFKISFKSLVCN
	: : . . * : : * : : : : * . * . : : : * : : *
NeseNPV	N---RMLEIENCDAVENRLIIVYYDHVFE-KKVVFSGSIYEHYITKHYPDE---LINIG
BmNPV	N---KPLYVDDYDDGVQDRFLIVYTNHKFV-DSVKFAGSVYEHKSKQFPIMESMYYESLV
CpGV	NEDDKIL-VDGYDRACSNRIGQMYFDHDFD-DSNEFSGSVYEHHIRKRYCEVRDVTDYLK
AdorGV	NDDDKIVILDGYDKACSNRIGQIYFDHEFDNDIKEFCGSVYEHVKNKYSEVRDVVSKLT
	* . : : . * . : : * : * : * . * . * : * : *
NeseNPV	DSFRVLVSYMLIHYKDNKTNLISSRRMLT--PRMIKNIKTVMNMIFAEADYILDVQMT
BmNPV	TPVRLFLSHVLMYRRDPKTGFVVYKTLNNDPMHKNLMCMSTNNSPLYALIYILNIKT
CpGV	DPVRAFLANVLQHKCNPIDGQLYKCFLQGDNTYKHNNKCLYIYNNISIEALLYVMNVKEV
AdorGV	DSVEAFLANVLKHKCDKVDGKLYYKSILQNDTSYKHNNKCLYIYNVRAEALYVLNVKEC
	. . : : : * : : . : . : * : * : * * : * : : *
NeseNPV	DDPNLWIDVETLKDVFLYVPKILQTI-----CVSKKMITQEE-----CEEMF
BmNPV	RNATITIGEDKMEEMISIAVQHLKNFLHPSFVQYNYKKINASSSKSFVNEQVLLQQIK
CpGV	KTAQ-EFSHETLIDTIKLAEKYVLNVLH----YSKRNGVCAES-----LCGEFK
AdorGV	NRSP-EFSEEFVIDLIKAEKYVLQMLH----YNKRNTVNADT-----LLIDFK
	. . : : : : . : : . : : . : : *
NeseNPV	YSRHVDKYYENADSMGKNYYRLQINTNLDKIKSKKAPRFI-----
BmNPV	NKFK--NNYNKTTNV---FYNMTMALNRNDLNT-SVPNFVC-----
CpGV	KKYNVPRHFNPDTNM---YMGIIQITDEKYFQT-FTPKFKANVDVT
AdorGV	KRFNTTKFFNSHTKM---YRNLQIAKSENYFKH-CPPNFLANADIVL
	: . : : . : : : . . : : * . *

Fig. S8 (Continued)

```

NeseNPV      -MENNILGL-MKTIANT-----ITVPQKHLFYSTDDIYVKTNEIT-- 38
BmNPV        -MIDNILQFFLKDVPEDKTYEINNLDANHLIIRNTRTGTRRLFYVNNFQQFLNTRNN 59
CpGV         MSVDDIMM-----SYE-----SVDRVERVSCKYWYFTN----- 28
AdorGV       MSVEDIFK-----TYE-----SVPEPKKNCNKYWFVD----- 28
              ::*:                      .      .      :

NeseNPV      -----FKLHSDCIATYLSDA-LHKSNDINDHNLTCINF EK-MLIHPFVEKT-FDQMD 88
BmNPV        FNGPCAKHDMGSSCEDTEEATEKQAVQQTLDGHDWVLESNDFC-IFVKPFILKKHYEVIQ 118
CpGV         ---GNK--KVRHQCRNVEFFIKLMTS-VARRAHEWCMVPNYN-SKVVPFMPYDDYQRLN 81
AdorGV       ---TSSKKCVRRPLGNTLFTKTLQSDSTLQHSWCMVPNYFLMTKIKPFMLCSDEYQVP 85
              :      .      *. : *      : **:      :: :

NeseNPV      ALKNLLAADVSRYTTNQFQPKFDEHRDYFVCSQIK---VTAFAWYFFMHKHFNLHI-LE 143
BmNPV        KYINFENFF--KSTDPGYINKCVQAGDYWWPNWPKGQAFSFGWQLFLNIKFGIVIEPT 176
CpGV         ATHPHVFKF--VDAFNNTISNTKRCGDYIVWPEMS---ANFLGWVVYLHLNCQYQLPLE 135
AdorGV       D--KSIFNF--TDIKKRIICFPVKMGDYMVWQNTN---VTFLSWALYLYANTKYKIDCL 137
              .  **      :      .  .*  ... :      :

NeseNPV      VPLPKHVRLGNVSLFTLKNDEKNSNAQIFDIELQIVKQN--GVLSFINSESITEPSNMSK 201
BmNPV        IPIIHNNKLGPDVLFVFDPK-----YFLNVELSLRTNHDPPQTLFVNGKTKFDESHEDL 230
CpGV         IPLIHHTRLGAFNLVSN-AE-----YNLDVSCKVKEG-SGERPFSNT---FQDTQLTL 184
AdorGV       IPVAHNRTIGNFNLIND-ST-----NCLDLACKIYDIN--NDLIFTNT---DKTPKIEI 185
              **:  ::  :*  ..*.      :::  .:      * *      .  :

NeseNPV      LIVFRHDDIMSVLLHGKHLILQNQYQDIFKFIKHNVF-----SYQRIVSTNKYFKLD 253
BmNPV        FILKMADGTVVT-CKVNGEL-VNSDKNFFNYIRDDINLEECITVPKYKHIVNVNL-K--- 284
CpGV         FTVHSTDNINI--MFSKECV-YSHEKKWFYEMVKTETISRCDLPEYQFLCDHIDFN--- 238
AdorGV       MTVQSKKSIKI--YFGSNYV-YTTNKVWYNFLKQENNVASVTFLPEYNFVCDKILFQ--- 239
              : :      ..      .  :  .  :  ::::  .      .*:  : .

NeseNPV      ANEIKYMNDFENFNIQD-----QEIANNDSPCLFNDLHGFSIHKQQIETAI 299
BmNPV        --SLRV---FEDNNFDKNDVDL--SDTRS RKPRI VPIISASSENAD-----YIQTQI 329
CpGV         --ELRW---LPEYYVKDSGMGLVVDNVKGVSVGVERMTPCS NYEL-----DIVNTV 285
AdorGV       --NLRY---DDEKTPKIIRKEQCVTP--STSLQITDHSVPCS NYEA-----EFNVYI 284
              .::      :  .      .      :  .      :      :

NeseNPV      QNIIQEFENKLIDKLNLLKSDWSCEDTIYFKTAHFLVNNDFACFEYILSYVYQKM---SQ 356
BmNPV        NLGLIAIYENMVK---VLATHEQANDPN--LLQYFEKSKFKNFDLIYVLWKILTKNEN 384
CpGV         NECYNFINETLMK-----KCTFAHNETD--LLSWYLFSTGYSTFYILLISTWNYCKTAIK 338
AdorGV       NECLEYINETMIS-----KGSATNNIAK--LLEFYLSSEYSTFYILIIALWQYCERCIN 337
              :      :  ::::.      :      ::  .. :  *  ::  ::      :

```

Fig. S9: Multiple sequence alignment [AdorGV, BmNPV, CpGV and NeseNPV] using EMBL-EBI Clustal Omega program

NeseNPV	LHIISNQDEHKLFLNIFMQKLFSSNYDCDNNVLKQII-----TC--YIDL DYNLFAKFCQS	409
BmNPV	--FSYRETDIKLFLELLCESLFACDKE---ILNEAL----KRCEPYKKQEKVFNRTCNH	435
CpGV	MHNQYSVHDILFFLKTLCDRAGGNHDL---YVDNLIYFSSNTAAENFMESLQFFVNPEEG	395
AdorGV	MENDYTIEDMLYFVKILCDKVDGGYKL---FLENLIYFGNKNCGKAFMKSLCFFVDPCR	394
	: *:: : : . . ::: : . . .*	
NeseNPV	FRFIS----DDSLCSYFALKIQGF-----VIEKHWKLSDYIVKKIDPEMLVSYCG	455
BmNPV	WFD FDDTKLCVSLGYGGIHYMIYLTLSAKNETLDDDELWAYTYENVMALNLPDIVCKG	495
CpGV	V-----EEYFNAISSYFAIHLAVK-----NNDEKAPLNDYTVKMCGVDDNVKSMG	440
AdorGV	K-----EAF FESICCYFALHFSIY-----EKTGDWRITASTIKQSEVEDDVKSVG	439
	:: *:::: . : : *	
NeseNPV	WFVYK--TEDSIYCF SERYFKQVSKSRSNDECSKIAYKSKLIYTVPDICFND SKFIYVT	513
BmNPV	FFRKLENVVTGVNLVFNGKNYQIVKKE--DDL FKL VKSNC--YKLSNIKFN NWKYLYLT	550
CpGV	FFKKIK--VSKFDYIFNGNLYVYYSKK--ENLLAAQFAKCPDVTVSSLIFNKT LNFYMT	496
AdorGV	FYKKIK--VGKFDYIFTGNIYENYKSKR--DHTLAFLMADCPEISVSGLTFNDTINFYHT	495
	:: : . *. . : ... :. . :. : ** . : * *	
NeseNPV	ENGVFNTITNKFVDFAPFVIANTMFLCALQNGQVYNISKYFFDDMLHNSKKEMLSMELYH	573
BmNPV	TYGVYNLFTNSFHSNCPFL LGTTLPQTLKKPTDEEYLPEDAFNYMLSTSTDELSIYRIYH	610
CpGV	EYGVFDVCKKIYREFCPFLVTSTLKQSFIT-KEQEYLNKTVFDRLYSAIDKDLTLFKMYH	555
AdorGV	QDGLFDVCRKTYKEPCPFVVMSTLRKNFIK-NNQVFLEKSLFDRIYNITD TDL DLLKMYH	554
	*::: : : . .**:: *: : : : : : : : : : : : : : : *	
NeseNPV	AAHIPRKMRRICALIRLGLYTNNDKLVNEN-ASKIVDIFNHLH----LCS--HICLLAMI	626
BmNPV	IAKMC RDVKMLKTNMAIVNYMGNCNTCQADMRVALNNLFR--DLWNLDDE--NLITLAL-	665
CpGV	ARKFLDDYEMVCDNLRKCVLVGESMEE---KSKWEGRWRELLVWLLECKASDIVIMMI-	611
AdorGV	ARKFMTEYEQTILCLKECKLANASN--S---VAILEEKIHEMVVWLLECKGSDILILMM-	608
	:: . . : : : :	
NeseNPV	KFTAPE----ITNLLCND-----GITLSDQTLQILVKYSLC--SAHIK-	663
BmNPV	YVNKNKVS DMLHNLKCKPCRSTVSGSRPKCKCYKKIKINRKALKVCLIADMFGNDAELSK	725
CpGV	KR-SGQLQNV LNIV-----ALSECDLMGLQVAMICHLLWPNSKVEL	652
AdorGV	YL-KDRLLQYINNIMK-----TAVD VDL MCMQSAIVAHLLWPNSKIEI	650
	. : *: . : : : : : : : : : : : : : : :	
NeseNPV	LLFLIVYLS-DDNDNKL-----IEALGETVFDINANLLINRKNDVLKNLTS-LTQMKN	714
BmNPV	LIWMLIFTNKTYVSTTLIRTNSEFVNQHGEFF FKEHNK-----IIQYLYQTIHKIEY	777
CpGV	FFWALIFNAYQDYEDWLEEYNVGS L--IHENVFNKKK-----ITEAMHRFLYNLSY	702
AdorGV	FLWSLLYNNFFDYEDSLMDYECDEL--ITKTMFDNKKK-----IIECVHAYLYQLDY	700
	::: ::: . * : : *. : : : : : : : : :	

Fig. S9 (Continued)

NeseNPV	IKQLISHIYDSSLWSKNVLL----IQN-EIDPTRKKFKSDVVDNPSVLLDQLVQQLTICN	769
BmNPV	VDML---M---DKFNDKRLFLTELRDDVAREPDVQFEESDNICKFYTHHTDALIILKKYN	831
CpGV	DDEV---SI-----ERV-----QDLVDEVGDPDKRKYEQKRIVKNIGAERYKYNKLRQQCN	750
AdorGV	SKCL---PFNDLFEI-----KTMVFNIPLADDKKFNKKRILNNVVELYEKYENYPDTYN	753
	. : . : : : . *	
NeseNPV	IWKDKLILPVSNETTLQWLIRFYMRCIFPHLRGFVNYATNPHMINIAEAYCYMLIANNFA	829
BmNPV	VWWDKIILARSTDDLPTWLTRFYMRIIMSKMDLK--EYSYNYLKKIVEGYLYFKRFTNFN	889
CpGV	VWSDLLIVHQPGENMYDWLTRFYIRSFLNDYTG-----DMKELYNVVLGFSYFRVFTNFH	805
AdorGV	VWSDYLIVQNKNEDMYTWLTRFYVRIFLNDLDCN--DANSRRLHNVVKGLCYFRVFTNFH	811
	:* * :*: : ** :*: * : : : : : : : : : *	
NeseNPV	YDHVLYLLTFAASLNIPQNYEKLLLVLGTSSNAGKTSFISLLGKIAVKMPS--NTLFDIN	887
BmNPV	HANAIMLMHFAASLAIPVDYGGKAIYMPGEPGSGKSSFFELLDY-LVLMHKFDDDNHSGE	948
CpGV	SNNSKALINFCASFAIPINNEKMCLVLSSKPCGKSSLWELFSKNKILVYKQDKE--EYK	863
AdorGV	CNNSKVLINFCASLAIPVDNEKMCIVISSKPCGKSSLWELLS-NIILVHKQDKE--VYK	868
	: * : *.**: ** : * : : .. :*:*: :*: : : . :	
NeseNPV	RGNGPSEMEFSRAISQLYEINEVHKTATENIKNTADLSREVVVRHAHSACQKIMLQYKPI	947
BmNPV	SNKETSDKEVSKLNSQLYVINELKQCSSEYFKKHADSSKSDSKSRKYQGLLKYEANYKML	1008
CpGV	HSKSERDEKVKLYESQLYVMNEAQKFSKTLKTIVDSTRMDSARCNYGIVEMFNITFKAL	923
AdorGV	HNNNERDEKVKLYESQLYVMNEAQKFTKAYLKSIVDNNRTDSARCNYGVMETFKISFKSL	928
	: : : :. **** :*: : : :*. .* : : : : : *	
NeseNPV	LCNNR--MLEIENCDAVENRLIIVYYDHVFEK-KVVFSGSIYEHYITKHYPDELI---	1000
BmNPV	IVNNK--PLYVDDYDDGVQDRFLIVYTNHKFVD-SVKFAGSVYEHKSKQFPIESMYE	1064
CpGV	VCNNEDDKI-LVDGYDRACSNRIGQMYFDHDFDD-SNEFSGSVYEHHRKRYCEVRDVT	981
AdorGV	VCNNDDDKIVILDGYDKACSNRIGQIYFDHEFDNDIKEFCGSVYEHVNKKYSEVRDVVS	988
	: ** :.. * . :*: :* :* . *.**:*** *::	
NeseNPV	NIGDSFRVLVSYMLIHYKDNKTNLISSRRMLT--PRMIKNIKTVSVMNIFAEADYILDV	1058
BmNPV	SLVTPVRLFLSHVLMYRRDPKTGFVVYKTLNNDPMHKHNLMMSTNNSPLYALIYILNI	1124
CpGV	YLKDPVRAFLANVLQHKCNPIDGQLYYKCFLQGDNTYKHNNKCLYIYNNSIEALLYVMNV	1041
AdorGV	KLTDSEAFANVLKHKCDKVDGKLYYKSILQNDTSYKHNNKCLYIYNVRAEALIYVLNV	1048
	: .. :*: : : . : : :* :* : * * :* :*:	
NeseNPV	QMTDDPNLWIDVETLKDVFLYVPKILQTIC----V--S-KKMITQEECEEMFY-----S	1105
BmNPV	KTVRNATITIGEDKMEEMISIAVQHLKNFLHPSFVQYNYKKINASSSKSFVFNELVLLQ	1184
CpGV	KEVKTAQEFSE-ETLIDTIKLAEKYVLNVL-----HYSKRNGVCAE-----SLCG	1085
AdorGV	KECNRSPEFSE-EFVIDLIKAEKYVLQML-----HYNKRNTVNAD-----TLLI	1092
	: : : : . : : . : : : .	
NeseNPV	RHVDKY--YENADSMGKNYYRLQINTNLDKIKSKKAPRFI-----	1143
BmNPV	QIKNKF--KNNYNKTTNVFYNTMALNRNDLN-TSVPNFVC-----	1222
CpGV	EFKKKYNVPRHFNPDNMYMGIQITDEKYFQ-TFTPKFKANVDVT	1131
AdorGV	DFKKRFNTTKFFNSHTKMYRNLQIAKSENYFK-HCPPNFLANADIVL	1138
	:*: . : : : : : : : : *	

Fig. S9 (Continued)

FrogTF	MDFSRLRVALCLSMALCLAIQKEKQVRWCVKSNSELKKCKDLVDTCKNKEIK--LSCVE
ChinTF	---CLATVYAAPVEGM-----VRWCVKSDKELQKCHDLAANVAE-----FSCVK
CowTF	MRPAVRALLACAVLGLCLA-DPERTVRWCTISTHEANKCASFRENVLRILESFPVSCVK
PossumTF	MRLTFSALLSFGIIGLCLA-STEQ-VRWCVISENELKKCNFKEAMNSINHPG--LACIK
	. . . : .: ****. * * :** .: .:***:
FrogTF	KSNTDECSLLFRKTMQMVFVWTGGDVYKGSQPYNLKPIMAENYGSHTETDTCYYAVAVV
ChinTF	RDDSFECIKAIKREEADAITLDGGDIYTAGLHNYNLQPIIAEDYG--EDSDTCYYAVAVA
CowTF	KTSHMDCIKAISNNEADAVTLDDGLVYEAGLKPNNLKPVVAEFHGTDKNPQTHYYAVAVV
PossumTF	KTSHLDCVKAIAENEADAVTIDGGLIYEAGLVPYNLKPVVAEDYGTENEPLSQYFAVAVG
	. . :* : .. ** :* ..* **:***:* :* .. :*:****
FrogTF	KKSSKFTFDELKDKKSCHTGIGKTAGWNIIIGLLERKLLKWAGPDSETWRNAVSKFFKA
ChinTF	KKGTFKGFNLHGKKSCHTGLGKSAGWNIPIGTLVTLDDQIQWAGIEDRPVESAVSDFNA
CowTF	KKDTEFKLNLERGKKSCHTGLGRSAGWNIPMAKLYKE---LPDPQ-ESIQRAAANFFSA
PossumTF	KKGTDFQIKDLKGKKSCHTGLGRSAGWVVPVGTLYELNLLDWEGPP-EPIDAVARFFSG
	..* : :*.***:*.*** : .. * . . *:***
FrogTF	SCVPGA--KEPKLSQLCAGIKEHKCSRSNNEPYNYAGAFKCLQDDQGDVAFVKQSTV-
ChinTF	SCVPGA--NTGSELQQLCMG---DCSRSHNEPYDYSGAFQCLKDGAGEVAFIKHLTV-
CowTF	SCVPCADQSSFPKLCQLCAGKGTDKCACSNHEPYFGYSGAFKCLMEGAGDVAFVKHSTVF
PossumTF	SCVPCA-LGKNPKLCSLCAGKGGDKCACSDREPYFGYSGAFQCLKDDVDGVSFVKHTTVL
	**** * . :*.** * .*: * .***:*.***:* ..*:***: *
FrogTF	---PEEFHKDYELLCPDNTRKSIKEYKNCNLAKVPAHAVLTRGRDDKSKDIEFLQEAQ
ChinTF	---PAAEKASYELLCKDNTRAPIDSYKTCHLARVPAHAVVSRKDTKLANHIYSKLMALK
CowTF	DNLNPNEDRKNYELLCGDNTRKSVDDYQECYLAMVPSHAVVARTVGGKEDVIWELLNHAQ
PossumTF	ENLQTESDKENYELLCPDNTRRPVDQFRDCNLAKIPAHAVVARSVNGKEDLIWNLLSTAQ
	. .***** **** :...: * ** :*:***:* . * . * :
FrogTF	K-----TQECKLFRLLPG-MGKGSNFQQRSESYSPPIFYQGFSVPRSRLFQCIQALKEG
ChinTF	D-----FNLFSSDGYAAKNLMFKDSTQNLVQLPMTTDSFLYLGAEYMSITIRSLTKA
CowTF	EHFGKDKPDNFQLFQSP--HGKDLLFKDSADGFLKIPSKMDFELYLGYEYVTALQNLRES
PossumTF	ENFGKGKAGKFRLFASD--HGKDLLFKDTANRLLRVPPKMDYELYLGYQYIQAIPKIGH
	. .** .*. *: . * . . :.:
FrogTF	VKED-DSAAQVKVRWCTQSKAEKTKCDDWTT--ISGGA--IECTEASTAEECIVQILKGD
ChinTF	QA---TGATSRAIKWCAVGHNEKVKCDAWTINSFTDGDRIECQDAPTVDCEIKKIMRKE
CowTF	KPPD-SSKDECMVKWCAIGHQERTKCDRWSG--FSGGA--IECETAENTEACIAKIMKGE
PossumTF	ALDKLHSLGDKKVKWCTVNHDEKKVCDEWSV--FSNGV--IACEVGENSEDCIVKIMKGE
	. . :.***: .: *. ** *: :..* * * . . :***:***: :

Fig. S10: Multiple sequence alignment [FrogTF, ChinTF, CowTF and PossumTF] using EMBL-EBI MUSCLE program

FrogTF	ADAVTLDDGGYMYTAGLCGLVPVMGEYYDQDDLTPCQRSCSQAKGV--YYAVAIIVKKG-TQ
ChinTF	ADAIADVGGEVFTAGKCGLVPVMVEQYD-----EVRCSAPGEASSYFAVAVAKKA-SG
CowTF	ADAMSLDGGYLYIAGKCGLVPVLAENYK-----TEGESCKNTPEKGYLAVAVVKTSDAN
PossumTF	ADAMSTDGGYVYIAGKCGLVPVLVENYAPKDAPGSHGPDCEKPAEGYYGVAVVKTG-TD
	.: *** :. ** **: * * . * .**:.*. . :
FrogTF	VSWSNLRGVKTCHTAVGRTAGWNIPVGLITSETANCDFASYVGESCAPGSDVKSNLCALC
ChinTF	LTWKTLLQGRRSCHTGLGRTAGWNIPMGLIHRRTMNCDFTTYFSKGCAPGFEVDSPFCAQC
CowTF	INWNNLKDKKSCHTAVDRTAGWNIPMGLLYSKINNCKFDEFFSAGCAPGSPRNSSLCALC
PossumTF	LTWDSLKGKKSCHTAVGRTAGWNIPMGLIYNKTQSCEFDKYFSESCAPGADPASSLCALC
	:.*.*. . :.***.:.*****:*. :*. * :. . ***** * :.*** *
FrogTF	IGDPEKLSEREKKCSPSASEAYYGYSAGFRCLVE-KGQVGFAKHTTVFENTDGKNPAGWA
ChinTF	RGSQSVGGDRACKASSEEQYYGYTGAFRCLVEDAGDVAFIKHTIVPEMTDGNGPV-WA
CowTF	IGSEKGTGKE--CVPNSNERYYGYTGAFRCLVE-KGDVAFVKDQTVIQNTDGNNNEAWA
PossumTF	IGKPGSPASD--KCASSTKETYYGYNGALRCLVENKGDVAFVKHTTVLENVEGKNTEPWA
	*. . * .:. * *****.***:***** *:*. * * * : .:*. . **
FrogTF	KDLKSEDFELLCPDGSRAPVTDYKRCNLAEVPAHAVVTLPDKREQVAKIVVNQQSLYGRK
ChinTF	QNLMSDFELLCDGTTQPVTKFRECHLAKVPAHAVITRPESRGEVVSILLEQQARFGSS
CowTF	KNLKKNFEVLCKDGRKPVTDACHLARGPNHAVVSRKDKATCVEKILNKQDDDFGKS
PossumTF	ANLKPKTLGSLCLDGSQKEVTRASECHLAKAPNHAVISRRDKADLVRQVLLDQQSKYGTY
	:* . : ** **: ** *:** * *****: :. * .: . ** :*
FrogTF	GFQ-KDIFQMFQSTGGKDLLFKDSTQCLLEIPSKTTMQEFLGDYHTAVTSLNKCSTSNE
ChinTF	GSD--SSFNMFKTDFGKNLLFKDSTKCLQEIPSGTKFQGFLGEEYMIAMQSLRECSNSTS
CowTF	VTDCSTNFCFLQSN-SKDLLFRDDTKCLASIAKKT-YDSYLGDDYVRAMTNLRQCSTSK-
PossumTF	GND-RKTFTLLPNT--KNLLFKDNTECLAKLPERITYEQFLGQEYVTAVGSLKKCSISS-
	: . * :. . *;***.*.*:** . . . : :***.* * : .*.*** *.
FrogTF	ASWLPAQFHSCMKIYIMVDCPLSII
ChinTF	DLEKACTFHSCQ-----
CowTF	-LLEACTFHKP-----
PossumTF	-LLPVCEFHKF-----
	. **.

Fig. S10 (Continued)

PossumTF	MRLTFSALLSFGIIGLCLAS-TE-QVRWCVISENELKKCNEFKEAMNSIN--HPGLACIK	56
CowTF	MRPAVRALLACAVLGLCLAD-PERTVRWCTISTHEANKCASFRENVLRILESGPFVSCVK	59
FrogTF	MDFSLRVALCLSLMALCLAIQKEKQVRWCVKSNSELKKCKDLVDTCKN---KEIKLSCVE	57
ChinTF	-----CLATVYAAPVEGMVRWCVKSDKELQKCHDLAANV-----AEFSCVK	41
	:. * * ****. * * :** .: .:***:	
PossumTF	KTSHLDCVKAIAENEADAVTIDGGLIYEAGLVPYNLKPVVAEDYGTENEPLSQYFAVAVG	116
CowTF	KTSHMDCIKAISNNEADAVTLDGGLVYEAGLKPNLKPVVAEFHGTDNPQTHYYAVAVV	119
FrogTF	KSNTDECSSLFRKTMQMVFVWTGGDVYKGSQPYNLKPIMAENYGSHTETDTCYYAVAVV	117
ChinTF	RDDSFECIKAIKREEADAITLDGGDIYTAGLHNYNLQPIIAEDYGED--SDTCYYAVAVA	99
	: . :* : . . ** :* ..* **:::** * : . : *::***	
PossumTF	KKGTDFQIKDLKGKKSCHTGLGRSAGWVVPVGTLYELNLLDWEGPP-EPIDAVARFFSG	175
CowTF	KKDTDKFLNELRGKKSCHTGLGRSAGWNIPMAKLYK---ELDPDQ-ESIQRAANFFSA	174
FrogTF	KKSSKFTFDELKDKKSCHTIGIGKTAGWNIIIGLLLERKLLKWAGPDSETWRNAVSKFFKA	177
ChinTF	KKGTKFGFLNLHGKKSCHTGLGKSAGWNIPIGTLVTLQIQWAGIEDRPVESAVSDFFNA	159
	..* : :*:***:***: : . * . . . *.: **..	
PossumTF	SCVPCALG-KNPKLCSLCAGKGGDKCACSDREPYFGYSGAFQCLKDDVGDVSFVKHTTVL	234
CowTF	SCVPCADQSSFPKLCQLCAGKGTDKCACSNHEPYFGYSGAFKCLMEGAGDVAFVKHSTVF	234
FrogTF	SCVPGAKE--P-KLSQLCAGIKEHKCSRSNNEPYNYAGAFKCLQDDQGDVAFVKQSTVP	234
ChinTF	SCVPGANT--GSELCQLCM----GDCSRSHNEPYDYSGAFQCLKDGAGEVAFIKHLTVP	213
	**** * :*.** .*: *.***:*.***:*** :. *::***: **	
PossumTF	ENLQTESDKENYELLCPDNTRRPVDQFRDCNLAKIPAHAVVARSVNGKEDIWNLLSTAQ	294
CowTF	DNLPNPEDRKNYELLCGDNTRKSVDDYQECYLAMVPSHAVVARTVGGKEDVIWELLNHAQ	294
FrogTF	E-----EFHKDYELLCPDNTRKSIKEYKNCNLAKVPAHAVLTRGRDDKSKDIEFLQEAQ	289
ChinTF	A-----AEKASYELLCKDNTRAPIDSYKTCHLARVPAHAVVSRKDTKLANHIYSKLMA--	266
	: .***** **** :...: * ** :*:***:~* . * . *	
PossumTF	ENFGKGKAGKFRLFASDH--GKDLLFKDTANRLLRVPPKMDYELYLGYQYIQAIPKIGH	352
CowTF	EHFGKDKPDNFQLFQSPH--GKDLLFKDSADGFLKIPSKMDFELYLGYEYVTALQNLRES	352
FrogTF	-----KTQECKLFRLPGMGKS-NFQGQRSESYSPPIFYGQFSVPRSRLFQCIQALKEG	342
ChinTF	-----LKDFNLFSSDGYAAKNLMFKDSTQNLVQLPMTTDSFLYLGAEYMSTIRSLTK-	318
	. ** . *:. . * . . . : : :	
PossumTF	ALDKLHSLGDKKVKWCTVNHDEKKVCDEWSVFS---NGVIACEVGENSEDCIVKIMKGE	408
CowTF	KPP-DSSKDECMVKWCAIGHQERTKCDRWSGFS---GGAIECETAENTEECIAKIMKGE	407
FrogTF	VKE-DDSAAQVKVRWCTQSKAEKTKCDDWTTIS---GGAIECTEASTAEECIVQILKGD	397
ChinTF	-AQ-ATGATSRAIKWCAVGHNEKVKCDAWTINSFTDGDSRIECQDAPTVDCEIKKIMRKE	376
	. . :***: .: * : ** * : * .. * * . . :*** :***: :	

Fig. S11: Multiple sequence alignment [FrogTF, ChinTF, CowTF and PossumTF] using EMBL-EBI Clustal Omega program

PposumTF	ADAMSTDGGYVYIAGKCGLVPVLVENYAPKDAPGSHGPDCEKPAEGYGVAVVKTG-TD	467
CowTF	ADAMSLDGGYLYIAGKCGLVPVLAENYKT-----EGESCKNTPEKGYLAVAVVKTSDAN	461
FrogTF	ADAVTLDDGGYMYTAGLCGLVPVMGEYYDQDDLTPCQR--SCSQAKGVYVAIVKKG-TQ	454
ChinTF	ADAIADVDDGGEVFTAGKCGLVPVMVEQYDEVRCSE-----APGEASSYFAVAVAKKA-SG	428
	.: *** .: ** **: * *	
	* .**:.*. .:	
PposumTF	LTWDSLKGKKSCHTAVGRTAGWNIPMGLIYNKTSCEFDKYFSESCAPGADPASSLCALC	527
CowTF	INWNNLKDKKSCHTAVDRTAGWNIPMGLLYSKINNCKFDEFFSAGCAPGSPRNSSLCALC	521
FrogTF	VSWNLRGVKTCHTAVGRTAGWNIPVGLITSETANCDFASYVGESCAPGSDVKSNLALC	514
ChinTF	LTWKTQGRRSCHTGLGRTAGWNIPMGLIHRRTMNCDFTTYFSKGCAPGFEVDSPFCAQC	488
	:*.*. .:***.:*****:*. . .*. .:..**** * :** *	
PposumTF	IGKPGSPA--SDKCASSTKETYYGYNGALRCLVENKGDVAFVKHTTVLENVEGKNTPEWA	585
CowTF	IGSEKGT--GKECVPNSNERYGYTGAFRCLVE-KGDVAFVKDQTVIQNTDGNNEAWA	577
FrogTF	IGDPEKLSEREKKCSPSASEAYGYSGAFRCLVEK-GQVGFAKHTTVFENTDGKNPAGWA	573
ChinTF	RGSGQSVGGDRACKASSEEQYYGYTGAFRCLVEDAGDVAFIKHTIVPEMTDGNP-VWA	547
	. : .:.* ****.***.***. *.*. * . * : .:*. **	
PposumTF	ANLKPRTLGLSLCLDGSQKEVTRASECHLAKAPNHAIVSRDKADLVRQVLLDQSKYGT	645
CowTF	KNLKKENFEVLCKDGRKPVTDACHLARGPNHAVSRDKATCCEKILNKQDDFGKS	637
FrogTF	KDLKSEDFELLCPDGSRAPVTDYKRCNLAEVPAHAVVTLDPKREQVAKIVVNQSLYGRK	633
ChinTF	QNLMSDFELLCQDGTTPVTKFRECHLAKVPAHAVITRPESRGEVVSILLEQQARFGSS	607
	:* . : ** **: ** .*:*. * ***: :. * .:..** :*	
PposumTF	GNDRK--TFTLLPNTKNLLFKDNTECLAKLPERITYEQFLGQEYVTAVGSLKKCSISS--	701
CowTF	VTDCSTNFCLFQSNKDLLFRDDTKCLASIAK-KTYDSYLGDDYVRAMTNLRQCSTSK--	694
FrogTF	GFQKDIQMFQSTGGKDLLFKDSTQCLLEIPSKTTMQEFLGDKYHTAVTSLNKCSTNEA	693
ChinTF	GSD-SSFNMFKTDFGKNLLFKDSTKCLQEIPSGTKFQGFLEGEYMIAMQSLRECSNSTSD	666
	: *:.***.***.*** .: . . : :***.* *: .*.*** *	
PposumTF	LLPVCEFHKF-----	711
CowTF	LLEACTFHKP-----	704
FrogTF	SWLPAQFHSCMKIYIMVDCPLSII	717
ChinTF	LEKACTFHSCQ-----	677
	. **.	

Fig. S11 (Continued)

Table S1: Harmonic identical words, the orders of position in the two sequences, and their difference for each harmonic identical word in the 1st alignment [Chin.TF vs Cow TF]

Harmonic identical words	Position in 1 st sequence	Position in 2 nd sequence	Difference in positions	Harmonic identical words	Position in 1 st sequence	Position in 2 nd sequence	Difference in positions
VRWC	14	25	-11	KCD	341	376	-35
KC	26	37	-11	IEC	357	388	-31
NV	33	44	-11	ECI	367	398	-31
SCVK	38	56	-18	KIM	371	402	-31
CIKAI	48	66	-18	EADA	376	407	-31
EADA	56	74	-18	DGG	383	414	-31
TLDGG	61	79	-18	AGKCGLVPV	390	421	-31
AGL	70	88	-18	AVAV	419	451	-32
NL	76	94	-18	SCHT	439	472	-33
AE	82	100	-18	RTAGWNIP	446	479	-33
YYAVAV	93	113	-20	MGL	454	487	-33
KK	100	120	-20	NC	463	496	-33
LN	108	127	-19	FS	470	503	-33
GKKSCHTG	112	132	-20	GCAPG	473	506	-33
TGLG	120	140	-20	CA	485	518	-33
SAGWNIP	123	143	-20	GS	490	523	-33
ES	150	162	-12	YYGYTGAF	510	540	-30
FF	156	171	-15	RCLVE	518	548	-30
ASCV	159	174	-15	GDVAF	525	554	-29
LCQLC	171	188	-17	TDGN	539	568	-29
EPY	185	206	-21	WA	546	576	-30
YSGAF	190	211	-21	NL	549	579	-30
CL	196	217	-21	FE	555	585	-30
GAG	200	221	-21	LC	558	588	-30
VAF	204	225	-21	DGT	561	591	-30
KH	208	229	-21	PVT	566	596	-30
TV	211	232	-21	CHLA	573	603	-30
YELLC	220	246	-26	HAV	581	611	-30
DNTR	226	252	-26	IL	596	626	-30
LA	240	266	-26	QQ	600	630	-30
VP	243	269	-26	FG	604	634	-30
HAVV	246	272	-26	LLF	624	655	-31
NH	258	291	-33	TKCL	630	661	-31
KD	268	299	-31	LG	646	676	-30
LF	281	316	-35	AM	654	684	-30
FKDS	285	318	-33	LR	657	687	-30
LYLG	303	336	-33	STS	663	691	-28
EY	308	341	-33	LE	667	696	-29
KWCA	330	365	-35	ACTFH	670	698	-28
GH	335	370	-35				

Table S2: Harmonic identical words, the orders of position, and their difference for each harmonic identical word in the 3rd alignment [AdorGV vs BmNPV]

Harmonic identical words	Position in 1 st sequence	Position in 2 nd sequence	Difference in positions	Harmonic identical words	Position in 1 st sequence	Position in 2 nd sequence	Difference in positions
VPE	13	13	0	YNVW	752	830	-78
TE	42	75	-33	TWLTRFY	770	848	-78
TL	55	88	-33	RI	778	856	-78
KPF	73	106	-33	YF	803	881	-78
YE	81	114	-33	FTNF	807	885	-78
NF	91	125	-34	ASLAIPVD	823	901	-78
TD	93	130	-37	GKSS	845	923	-78
GDY	107	142	-35	ELL	851	929	-78
IP	138	177	-39	HK	860	938	-78
HN	142	181	-39	KL	879	960	-81
LR	241	286	-45	SQLYV	883	963	-80
IN	284	329	-45	NE	889	969	-80
AT	301	345	-44	DS	909	989	-80
LL	308	355	-47	NN	931	1011	-80
LI	324	371	-47	YD	943	1020	-77
LW	328	375	-47	GSVYEH	969	1045	-76
EN	339	383	-44	FL	997	1073	-76
LC	354	399	-45	VL	1001	1077	-76
IY	412	458	-46	YK	1015	1091	-76
GF	439	495	-56	ND	1021	1097	-76
FN	486	541	-55	KHN	1026	1102	-76
CPF	511	566	-55	ALIY	1041	1117	-76
TL	518	573	-55	LN	1046	1122	-76
TD	545	600	-55	LH	1077	1154	-77
YH	553	609	-56	YN	1079	1161	-82
KII	686	763	-77	NA	1086	1168	-82
FND	706	785	-79	KF	1102	1189	-87
LF	710	790	-80	PNF	1128	1218	-90

Table S3: Results of alignments using EMBOSS Needle program with different values of gap penalties.

Alignment sequences	values of gap parameter s				Results of the alignment				
	Gap open	Gap extended	End gap open	End gap extended	Common length including gaps	identity	similarity	Total length of gaps	score
First alignment: Chin.TF vs Cow TF	5	1	5	1	741	346/741 (46.7%)	459/741 (61.9%)	101/741 (13.6%)	1692
	5	0.5	5	0.5	745	347/745 (46.6%)	460/745 (61.7%)	109/745 (14.6%)	1713.5
	10	1	10	1	732	331/732 (45.2%)	444/732 (60.7%)	83/732 (11.3%)	1529
	10	0.5	10	0.5	737	334/737 (45.3%)	445/737 (60.4%)	93/737 (12.6%)	1550.5
Second alignment: CpGV vs NeseNPV	5	1	5	1	1299	328/1299 (25.3%)	560/1299 (43.1%)	324/1299 (43.1%)	952
	5	0.5	5	0.5	1381	324/1381 (24.8%)	571/1381 (41.3%)	488/1381 (35.3%)	1072
	10	1	10	1	1233	272/1233 (22.1%)	498/1233 (40.4%)	192/1233 (15.6%)	557
	10	0.5	10	0.5	1276	277/1276 (21.7%)	492/1276 (38.6%)	278/1276 (21.8%)	620
Third alignment: AdorGV vs BmNPV	5	1	5	1	1312	360/1312 (27.4%)	617/1312 (47.0%)	264/1312 (20.1%)	1313
	5	0.5	5	0.5	1369	370/1369 (27.0%)	624/1369 (45.6%)	378/1369 (27.6%)	1410
	10	1	10	1	1268	310/1268 (24.4%)	571/1268 (45.0%)	176/1268 (13.9%)	942
	10	0.5	10	0.5	1292	314/1292 (24.3%)	573/1292 (44.3%)	224/1292 (17.3%)	1002.5

Table S4: Different parts in the 1st alignment [Chin.TF vs Cow TF] using our method (left column) and EMBOSS Needle program (right column)

-CLATVYA--A-PVEGM----- : MRPA-VRALLACAVLGLCLADPERT Score* = 16, Gaps of total length 13	-----CLATVYAAPVEGM ... MRPAVRALLACAVLGLCL-----ADPERT Score* = 16, Gaps of total length 21 Comment: Gap attraction
---E---F Score = 11 Gaps of total length 6 RILESGPF	EF----- Score = 0, Gap of length 6 .. Comment: Gap attraction RILESGPF
E-DSD- Score = 5, Gaps of total length 2 . :. TKDNPQ	--EDSD Score = 2, Gap of length 2 :... Comment: Gap attraction TKDNPQ
FLN- Score = 10, Gaps of total length 2 -LNE	FLN Score = -3, No gaps ... Comment: Gap annihilation LNE
VTLDQIQWAG Score = 2, Gaps of total . : : length 7 Y---K-E---	VTLDQIQWAG Score = -1, Gap of length 7 :... Comment: Gap attraction -----YKE
RPVES--- Score = 14 . Gaps of total length 4 -PQESIQR	--RPVES Score = 3, Gap of length 2 ...: Comment: Gap annihilation PQESIQR
---D- Score = 6, Gaps of total length 4 KGTDK	----D Score = -1, Gap of length 4 . Comment: Gap attraction KGTDK
----K- Score = 5, Gaps of total length 5 TVGGKE	-----K Score = 1, Gap of length 5 : Comment: Gap attraction TVGGKE
MALKD---- Score = 11, Gaps of total . length 6 -G-KDKPDN	MALK--D Score = -2, Gap of length 2 ... : Comment: Gap annihilation GKDKPDN
-RA- Score = 0, Gaps of total length 2 .. DECM	R--A Score = -3, Gap of length 2 . . Comment: Gap attraction DECM
DSR Score =1, Gaps of total length 2 : -A-	DSR Score = -1, Gap of length 2 . Comment: Gap attraction --A
D-EVR- Score = 1, Gaps of total length 2 . .. KTEGES	--DEV R Score = -3, Gap of length 2 :... Comment: Gap attraction KTEGES
GDRA Score = 2, Gaps of total length 3 : --K-	GDRA Score = -1, Gap of length 3 . Comment: Gap attraction ---K
TDGNG- Score = 23, Gap of length 1 . TDGNNN	TDG-NG Score = 23, Gap of length 1 . Comment: Alignment using TDGNNN identical tetra word is more accurate than using identical triplet word.
-LEK Score = 9, Gaps of total length 2 LLE-	LEK Score = 2, No gaps .: Comment: Gap annihilation LLE

*The scores in the table were calculated using BLOSUM 62 score matrix and without gap penalty.

Table S5: Different parts in the 3rd alignment [AdorGV vs BmNPV] using our method (left column) and EMBOSS Needle program (right column)

[illegible]

Table S5 (continued)

-HNNNE-R Score = 11, Gaps of total length 2 .: . . . GESNKETS	HNNNER-- Score = 6, Gap of length 2 ... :. Comment: Gap attraction GESNKETS
-KLY Score = 9, Gaps of total length 2 SKL-	KLY Score = -3, No gaps ... Comment: Gap annihilation SKL
-ARCNY-GVMET Score = 23, Gaps of total length 4 : . . ::: KSR-KYQGLLK-	ARCNYGVMET Score = -5, No gaps Comment: Gap annihilation KSRKYQGLLK
HVNKKYSEVRDV- Score = 8, Gaps of total length 2 :..: -IKSKQFPIESMY	HVNKKYSEVRDV Score = -6, No gaps :.. Comment: Gap annihilation IKSKQFPIESMY
VLQ- Score = 5, Gaps of total length 2 : -LKN	VLQ Score = -1, No gaps :.. Comment: Gap annihilation LKN
-KRNT Score = 13, Gaps of total length 2 : YKKN-	KRNT Score = 0, No gaps :... Comment: Gap annihilation YKKN
DT--LLIDFKKR--FN--TTKF--- .: .: .: . . . SSSKSFV-FNEQVLLQIKNKFKNN Score = 20, Gaps of total length 10	DT-----LLIDFKKRENTTKF .: : ... SSSKSFVFNEQVLLQIKNKFKNN-- Score = 19, Gaps of total length 12 Comment: Gap attraction

*The scores in the table were calculated using BLOSUM 62 score matrix and without gap penalty.

Scripts:

Script1: the aim of this script is to find the identical words and their position orders in the two aligned sequences.

Script2: the aim of this script is to find the maximum score alignment of the two subsequences between each two successive HIWs.

Script3: the aim of this script is to evaluate the alignment quality by measuring the identity percent, gap percent and similarity.

All scripts S1, S2 and S3 are present at the link below:

https://colab.research.google.com/drive/1JzYn7Iz_kRk4JlFxlI9BvTz71C_3jINV#scrollTo=_ET_upz