Multiple Sequence Alignment

Presenter: Brian Foley

btf@lanl.gov

HIV Databases



Theoretical Biology and Biophysics,

Los Alamos National Laboratory



<u>www.hiv.lanl.gov</u> <u>seq-info@lanl.gov</u>



Presenter: Brian Foley

<u>btf@lanl.gov</u>

HIV Databases

- 28 6 -	J
HIV DATABASES	
The HIV databases contain comprehensive data on HIV genetic sequences and immunological epitopes. The website also gives access to a large number of tools that can be used to analyze and visualize these data. This project has been funded in whole or in part with Federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under Interagency Agreement No. AAI12007-001-00000. Our content is reviewed by an <u>Editorial Board</u> .	
SEQUENCE DATABASE > IMMUNOLOGY DATABASE >	
OTHER VIRUSES ►	
News: Archived News >	
CATNAP: two new features CATNAP now provides an option to calculate geometric mean estimates including tests that were above threshold (setting a score of 100 (IC _{50/80}) or 20 (ID _{50/80}) for the purpose of the estimation). Also, we have introduced a "Trim-and-Re-calculate" feature in the analysis which enables users to select data from specified papers instead of using the full set in CATNAP collection. This could be useful to reduce data redundancy or to address inconsistencies between studies (for instance, changes in pipette tips used for serial dilution). 20 February 2019	
Questions or comments? Contact us at seq-info@lanl.gov	







Topics

Multiple Sequence Alignment, Primarily HIV-1

General introduction:

HIV, presents special challenges for alignment due to frequent insertions and deletions (in/dels) relative to most other organisms.

Multiple genes, and overlapping reading frames make it difficult to align the genome "in frame" for convenient translation to amino acid sequence.

Tools for aligning HIV genomes and genes.

Special problems and tips for dealing with them.

We will not cover alignment of short reads to genome; Bowtie etc.

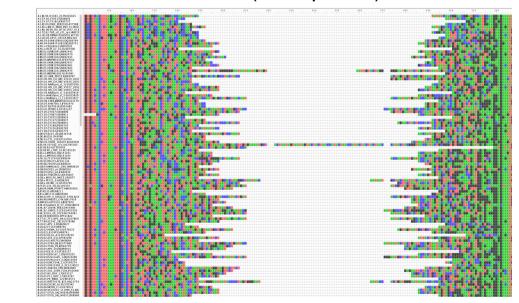


Pairwise vs multiple alignment

Pairwise alignment can be local or global. Simple, quick and fairly unambiguous. "Optimal" alignment is well defined.

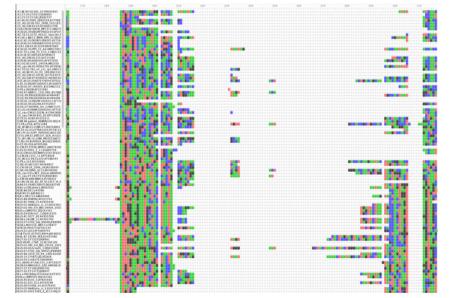
Multiple alignment is much more difficult. "Optimal" alignment is debatable. BLAST Needle (EMBOSS) Water (EMBOSS) Align0

ClustalW ClustalOmega MUSCLE MAFFT GeneCutter (HIV specific)



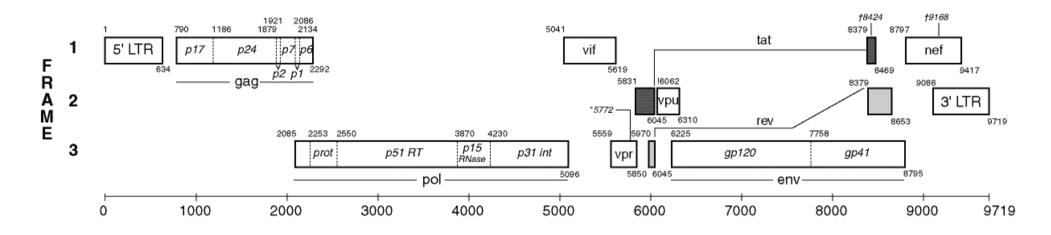
GeneCutter preferentially puts gaps in between codons, splits regions of uncertain alignment in the center.





Typical alignment of HIV-1 env by MAFFT or similar tools. Cannot be translated to amino acid sequences.

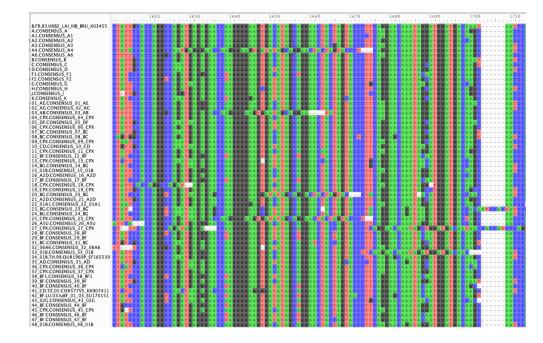
HIV-1 Genome

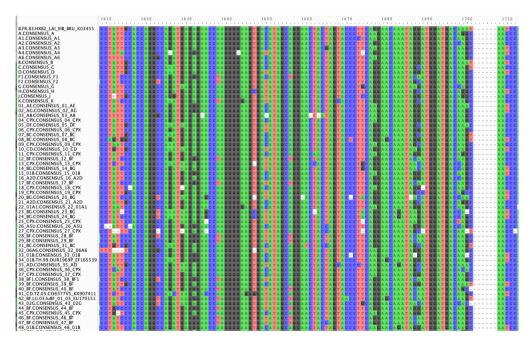


https://www.hiv.lanl.gov/content/sequence/HIV/MAP/landmark.html

Nine proteins/polyproteins produced, 7 regions with overlapping reading frame, one with triple overlap (tat/rev/gp41). Thus, no single genome alignment can be easily translated to the amino acid sequences for all proteins.







GeneCutter has an option for how many codons between compensating in/dels are allowed to be misaligned.

Top example set to 25 codons, bottom zero codons.



Gene Cutter

Sequence Alignment and Protein Extraction

Purpose: Gene Cutter is a sequence alignment and protein extraction tool for HIV-1, HIV-2 or SIV. This tool will:

Align your nucleotide sequences (if they aren't already aligned).
Clip coding regions from a nucleotide alignment.
Codon-align all coding regions.

Codon-augn au coding regions.
 Generate nucleotide and protein alignments of the cut regions.

Please read Gene Cutter Help, particularly before running large jobs.

Input	,
Organism	HIV1 (HXB2) 0
Paste your sequences [Sample Input]	
Or upload your file	Browse No file selected.
Check if appropriate	Sequences are unaligned 🙆
Options	
Region(s) to align and extract	All genes and complete sequence o
Reference options	□ Insert HXB2(K03455) for HIV-1 or SMM239(H33262, 100 INT+24 JIT INTO the results ®
	Remove HXB2(K03455) for HIV-1 or SW 39(M33262) for HIV-2/SIV from the results @
Check if appropriate	Codon align the region 🕖 Allow 15 codons to compensate frameshift
Output format	Translate to amino acids: ○ Codons containing an IUPAC character shown as 'X' @ ○ Codons containing an IUPAC character in a silent position translated; others shown as 'X' @ ⓒ Codons containing an IUPAC character translated @ @ codons containing an uLPAC character translated @ @ codons containing an uLPAC character translated @ Do not translate

Submit Reset



For many purposes, MAFFT or other multiple alignment tools are the best option, Factors to consider are speed, ability to add the alignment process to a "pipeline" or automated script or program, ease of use, etc. as well as overall "quality" or usefulness of the resulting alignment.

Often, a first pass "quick and dirty" alignment can be highly useful in checking the data, for quality, length variations, etc. before more time-consuming methods are applied.

Always use some multiple alignment viewing tool to at least spot-check the alignment before proceeding to further steps. Aliview (Mac) BioEdit (Windows), Pixel, etc..

https://ormbunkar.se/aliview/ http://www.mbio.ncsu.edu/BioEdit/bioedit.html https://www.hiv.lanl.gov/content/sequence/pixel/pixel.html





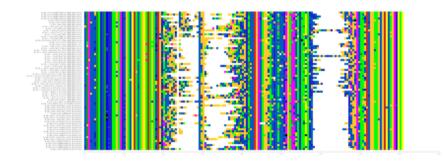
Pixel

Input & Options Alignment: GlycoExampleV12adj.FASTA Sequence type: AA Color scheme: Default Show differences: No Plot size: Automatically compute. Residue scale: 3 x 3 pixels Show sequence names: Yes Font size: 3 Show scale bar: Yes Scale bar size: 3 Margins: 0.5 inches

Results

[View large]

Download [EPS] [PDF] [PNG] [Legend] [All]

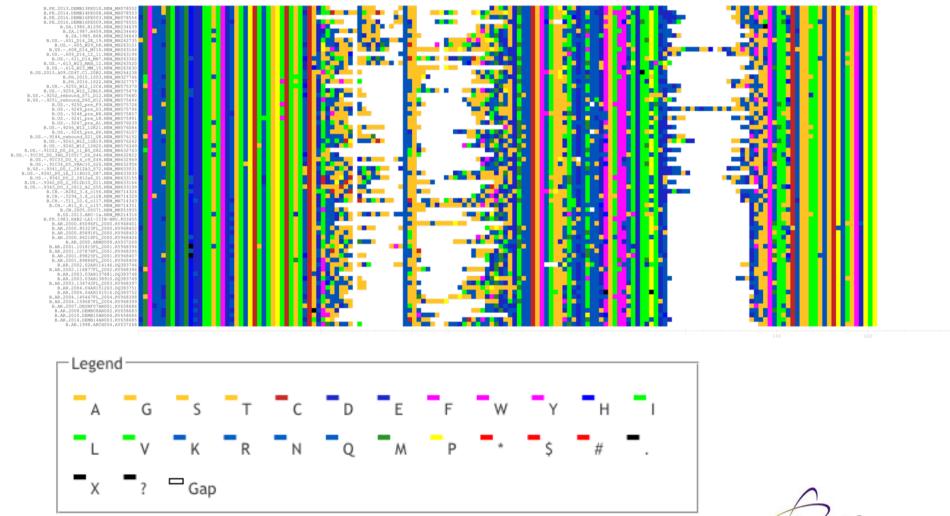




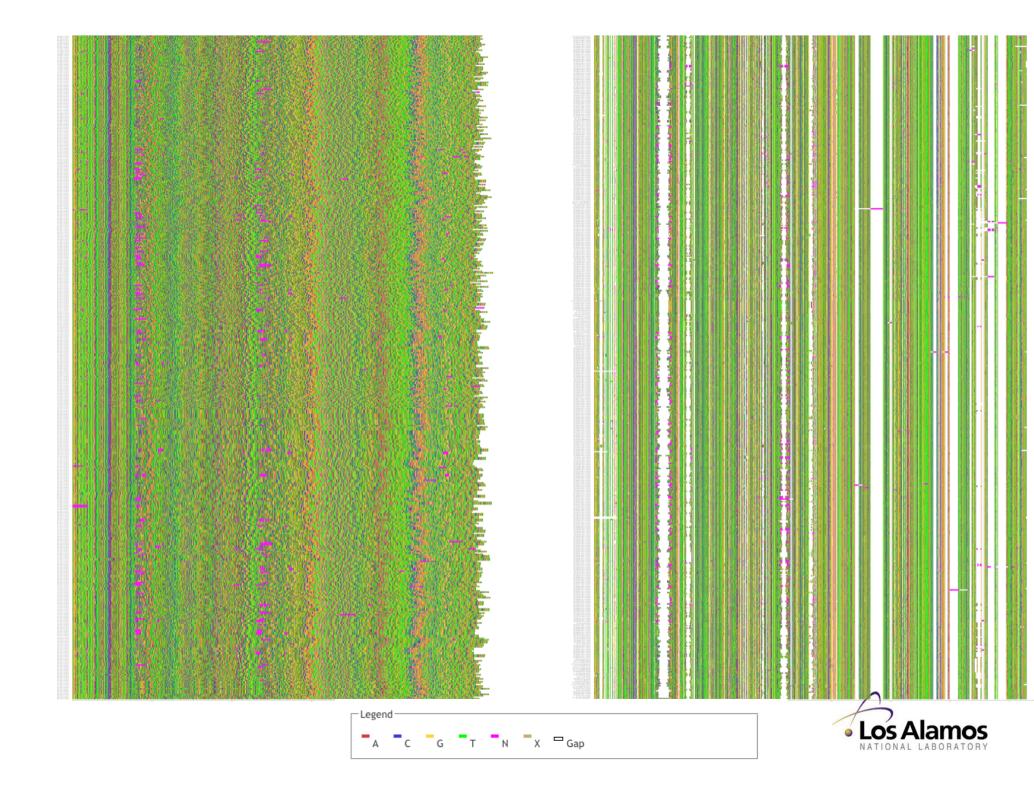
Questions or comments? Contact us at seq-info@lanl.gov.



PIXEL view of HIV-1 Env V1-V2 region amino acid alignment





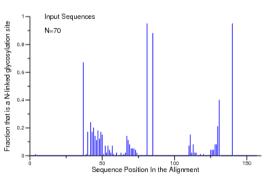


B.PE.2013.DEMB13PE010.NEW MH078552 B.PE.2014.DEMB14PE008.NEW MH078553 B.PE.2016.DEMB16PE003.NEW MH078554 B.PE.2016.DEMB16PE009.NEW MH078555 B.ZA. 1986. B1296. NEW MH234639 B.ZA.1987.R459.NEW MH234640 B.ZA.1985.R68.NEW_MH234643 B.US. -. 601 D14 2E 19.NEW MH262735 B.US. = . 605 W29 b8.NEW MH263121 B.US.-.608 D14 MI10.NEW MH263144 B.US.-.609_D14_12_11.NEW_MH263194 B.US.-.611_D14_MH7.NEW_MH263362 B.US.-.613 W23 MAB 12.NEW MH263520 B.US.-.616 W23 MM 15.NEW MH263630 B.US.2015.A09.CD4T.C1.20B2.NEW_MH264238 B.PH.2015.1003.NEW MH327746 B PH 2016 1022 NEW MH327757 B.US.-.9255 W12 12C4.NEW MH575379 B.US.-.9254 W12 12BL5.NEW MH575478 B.US.-.9252_rebound_S71_D12.NEW_MH575680 B.US.-.9251_rebound_S60_H12.NEW_MH575694 B.US.-.9250 pre F9.NEW_MH575728 B.US.-.9249 pre G3.NEW MH575796 B.US.-.9248_pre_B8.NEW_MH575857 B.US.-.9241 pre_L8.NEW_MH575591 B.US.-.9247 pre_A1.NEW_MH576039 B.US.-.9246_W12_12K21.NEW_MH576084 B.US.-.9245_pre_H6.NEW_MH576107 B.US.-.9244_rebound_S21_G8.NEW_MH576152 B.US.-.9243_W12_12E19.NEW_MH576262 B.US.-.9242_W12_12H20.NEW_MH576349 B.US.-.91C22 D0 26 11 B5 582.NEW MH632763 B.US.-91C22_00_20_11_0_5_02_1002_1002_2022 B.US.-91C35_00_94_c9_849.NEW_MH632822 B.US.-91C34_00_94_c9_849.NEW_MH632949 B.US.-91C34_D0_YBAC10_S26.NEW_MH632956 B.US.-9341_D0_1_2812A3_S72.NEW_MH633031 B.US.-.9341_D0_1A_1118G10_S87.NEW_MH633039 B.US.-.9342_D0_2_2812a6_S1.NEW_MH633155 B.US.-.9342_D0_2_3012b10_S11.NEW_MH633164 B.US.-.9343 D0 3 2812 A2 S55.NEW MH633199 B.CH.-.R282_3.4_c156.NEW_MH714324 B.CH.-.T294_3.8_c128.NEW_MH714329 B.CH.-.T11 10.6 cl17.NEW MH714343 B.CH.-.R11 8.1 c157.NEW MH714351 B.CN.2005.05071.NEW MK053935 B.US.2013.ARC-1a.NEW_MK214316 B.FR. 1983.HXB2-LAT-TTTB-BRU.K03455 B.AR.2000.85096FL 2000.KY968401 B.AR.2000.85323FL_2000.KY968402 B.AR.2000.85891FL_2000.KY968403 B.AR.2000.86218FL_2000.KY968404 B.AR.2000.ARMS008.AY037269 B.AR.2001.101815FL_2001.KY968394 B.AR.2001.107878FL_2001.KY968395 B.AR.2001.89825FL 2001.KY968407 B.AR.2001.89886FL 2001.KY968408 B.AR.2002.02AR114146.DQ383746 B.AR.2002.116877FL 2002.KY968396 B.AR.2003.03AR137681.DQ383748 B.AR.2003.03AR138910.DQ383749 B.AR.2003.134742FL 2003.KY968397 B.AR.2004.04AR151263.DQ383751 B.AR.2004.04AR151516.DQ383752 B.AR.2004.145447FL_2004.KY968398 B.AR.2004.159687FL 2004.KY968399 B.AR.2007.DEURF07AR001.KY658686 B.AR.2008.DEMB08AR002.KY658683 B AR 2010 DEMB10AR006 KV658684 B.AR.2014.DEMB14AR003.KY658685 B.AR.1998.ARCH054.AY037268 B.PE.2013.DEMB13PE010.NEW MH078552 B.PE.2014.DEMB14PE008.NEW MH078553 B.PE.2016.DEMB16PE003.NEW MH078554 B.PE.2016.DEMB16PE009.NEW_MH078555 B.ZA.1986.R1296.NEW MH234639 B.ZA.1987.R459.NEW MH234640 B.ZA.1985.R68.NEW_MH234643 B.US.-.601_D14_2E_19.NEW_MH262735 B.US.-.605 W29 h8.NEW MH263121 B.US.-.608 D14 MI10.NEW MH263144 B.US.-.609_D14_12_11.NEW_MH263194 B.US.-.611_D14_MH7.NEW_MH263362 B.US.-.613 W23 MAB 12.NEW MH263520 B.US.-.616 W23 MM 15.NEW MH263630 B.US.2015.A09.CD4T.C1.20B2.NEW MH264238

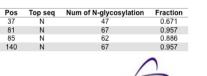
	10	20	30	40	50	60	70	80	90		10	00
				1			I		1			
ATM/ TRADE OF A	00	MURDER TO THE	OCT N DOLINE M	DT OURT DOON	TEL OPPORTO ON T	TTCTSNS	mmontrach	NTODODODAY	NOODNITOODO	DDI	12 13 13 12 1	
DMWKNSMVI	DQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLECSN PLCVTLECSN	VICTINSSNI	TTCTSNS	-TTSPNITSP	NITTSPTEMK	NCSFNITTPQ	RDKT	KEEY/	AF
NMWKNDMV		MHEDITSLWD	ESLKPCVKLT	PLCVTLNCTN	VTTTCNNTTD		GGNDT	RINVESDETK	NCSEHTHT	GDOK	OREV	AT
NMWKNNMV	EO	MHEDIISLWD	OSLKPCVKLT	PLCVTLNCTD	WRNTTNTNT-		NKTS	NTNSSIGEMK	NCSFNVTTSI	RDKV	KKES/	AL
NMWTNNMVI	EQ	MHDDIINLWD	QSLKPCVKLT	PLCVTLNCTD	DLKNTTNTNA	TNT	NASS	WGNIEKGEIK	NCSFNVTTNR	RGKM	QKEY/	AL
NMWKNNMV	EQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	NLRNDTTTNN	TNT	SGIG	GIKMEKGEIK	NCSFNITTNI	RDKF	QKVY/	AL
TMWKNNMVI	EQ	MHEDIINLWD	QSLKPCVKLT	PLCVTLNCTD	LNNATNANST	NGNNATNNAT	NTTSSYSW	GPPIEEGEIK	NCSFNITTSI	KNKV	QKEY	AL
						GEMKNCS						
NMWKNNMVI	DQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	PTFNTTSATN	TNVTNTTGTT	SAPSTPSNTT	TSPTIKEEVK	NCSFNITTNM	VDKW	QKEY/	AL
							000000000000000000000000000000000000000					
NMWKNNMV	EO	MHEDIISLWD	OSLEPCVELT	PLCVTLDCTN	LNATNSNTTN	SNATNSN	TTNHSS	WERMDKEETK	NCSENVTTNI	GDKM	OKAY	AT.
NMWKNNMV	DÕ	MHEDIISLWD	OSLKPCVKLT	PLCVTLECHN	VILNNINATI	EMKNCSIAIN SNATNSN TSTTPSS STT TT TINVNTN T		SGEMK	NCSFNVTAPI	RDKI	ÖKGY	GI
NMWKNNMV	EQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTA	YVGDTTTTNS	STT	TN	TNISMEGEMK	NCSFYTTPNI	RDKE	QKEY	AL
NVWKNDMV	EQ	MHEDIINLWD	QSLKPCVKLT	PLCVILNCTN	ATINSTTPE-		Т	ELRTLKEEVR	NCSFNVTPSI	KNKM	QKEY	AL
NIWKNNMV	EQ	MHEDIISLWD	QSLKPCVKLT	PFCVTLDCTD	TNGTYSTINN	TT	GG	ATIMQKGEMK	NCSFNITTSI	QDRR	QKEY	AL
NAWKNNMV	EQ	MHEDIISLWD	QSLKPCVKLT	PLCVILHCTD	VNRTSN			EMTGELK	NCSFNITTKI	TNKV	QKEY!	AL
NMWKNDMV	EQ	MOEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	VEIDKTSANK	TINVNTN	TTNS	WERMDPGETK CELECKCEME	NCSENTTENT	RDRV	KNEY	AL
NMWKNNMU	EO	MHEDVISLWD	OSLKPCVKLT	PICVTLNCTD	VONNANTTV-	T TNT ST NN	NAST	TVNNGKGETK	NCTERVTGGI	RDKT	OLEV	AL.
NMWKNGMV	DO	MHEDIIGLWD	OSLKPCVKLT	PFCVTLNCTD	LONSTSTNNN	TNT	NTT	IDNNMKGEIK	NCSENITTST	KEKK	-KGY/	AF
NMWKNNMV	ΕÕ	MHEDIISLWD	OSLKPCVKLT	PLCVTLNCTD	VETNSNEKAN	ST	NN	GANITAADMK	NCSFNITASI	RDKV	OKEY/	AL
NMWKNNMV	EQ	MHEDIISLWD	ESLKPCVKLT	PLCVTLNCTN	ANFTTTNSTT	NN	SSTT	GEIKTTGEIK	NCTFYTTDM	RDKK	QKNR/	AL
NMGKNNMV	EQ	MHEDIISLWD	QSLKPCVLLT	PFCVTLNCTD	ADTTGTN		SS	STLEEKGEIK	NCSFNLTSSI	SDKP	QEQR	AL
NMWKNNMV	EQ	MHEDVINLWD	QSLKPCVKLT	PLCVTLNCTN	LNVTNTN		SSS	KEGKMAEEMR	NCSFNITTEI	RNKM	QKEY	AL
NMWKNNMV	DQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	IRNSTVTSNT	TTSNTTT MNNT TCS N SGS DVAT	mupuccy	WGEMT	NCTFNITTSI	KDKM	KKEA	AL
NEWKINSMV	20	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD DI CUTI NOTE	VNATATMGND	MNINT	TNDNSSK	GREMEIGEMK	NCSENTITIST	BDEN	OFON	AL AL
NMWKNDMV	FO	MHEDITSLWD	OSLKPCVKLT	PLCVTLNCTD	NUNUNCENNIN	TCS	NNTN	VNMTEKCETK	NCSENVITTET	RDRV	TREN	AL.
NMWKNNMV	DO	MHEDIINLWD	OSLKPCAKLT	PLCVTLNCTD	LENDTVGNOT	N	YLNE	TNTIOGREMT	NCSFNITTEI	RDKV	RKEY	AL
NIWKNNMV	ΕÕ	MHEDIISLWD	OSLKPCVKLT	PLCVTLKCTN	LTNVTES			-DOMKEGGMO	NCSFNVTAGI	KDRM	OKEY/	AL
NMWKNNMV	EQ	MHEDVINLWD	QSLKPCVKLT	PLCVTLHCTN	TTKDNITI			EGNGEIK	NCSFNITTDL	RDKV	QKQR/	AL
NMWKNNMV	EQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTS	LTNNNTTHGK	SGS	SKNSSS	EWERVREEMT	NCSFNVTPSI	G-KK	RKEY	AL
NMWKNSMV	EQ	MHEDVISLWD	QSLTPCIKLT	PLCVTLDCTD	VNVTAKGN		ST	WEKMEKGEIK	NCSFNVTRSL	KAKV	RTEH	AL
						RGNNSNA KNSTNAT						
NMWKNSMV	EQ	MOEDIISLWD	OSLKPCVELT	PLCVTLNCTD	CDNPNCTALR	NSTTATN	NSIKNSID	ATTSTRGETK	RCSFNITTNM	RDOV	RIEI	AL AL
NMWKNNMV	DO	MHEDIISLWD	OSLKPCVKLT	PLCVTLTCTN	ATTSCANON-	NSTTATN	GNST	INCTNVEEMK	NCSENITERI	KNKV	KKEY	AL
NMWKNNMVI	EQ	MQEDIISLWD	QSLKPCVKLT	PLCVTLNCTI	LNCTTNCTTN	STTNSTN	ST	NSTWEDKGLR	NCSFYAPTSI	RDKV	KKEY	AI
NMWSNNMV	EQ	MHEDVISLWD	QSLKPCVKLT	PLCVTLNCTD	LTNSDTTDNT	T	IN	SNWEGKGEVK	NCSFNITTSI	RDKV	QKGY/	AT
NMWKNNMV	EQ	MHKDIISLWD	ESLKPCVKLT	PLCVTLNCTD	LKKDNLTCAN	LTKGSNCTNN	SSVGNDTNTN	NSSWIGEDMR	NCSFNATTNI	VDKV	QQQY	AL
NMWKNNMV	EQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	LKKNDTSTNS	DTTASTN	TNS	WGEMDKGEIK	NCSFNITTNI	LDKR	QKHY!	AL
NMWKNNMV	EQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTN	YNNTNSNTSS	T TQTSNRN	0000000	TEGGEIK	NCSFNITTSI	NTKV	-KDY	AL
NMWKNNMV	EQ	MHEDIINLWD	ESLKPCVKLT OCLKPCVKLT	PLCVELNCTD	LENEIGTSNE	TQTSNRN	QTSNRT	TEPOMOGETK	NCSENVITION	RDKK	KKEH/	AL
NMWKNNMV	EQ	MHEDIISLWD	ESLEPCVELT	PLCVTLNCTN	LKNNNTTTTN	TAASPSS		GGEIK	NCSENITTGI	GKKV	RSEV	AT
DMWKNNMV	EO	MHEDIISLWD	OSLKPCVKLT	PLCVTLNCTA	LKTNSTTSDS	T	SNG	TVSPAREEMT	NCSFNVTTNI	RNKI	OKEY/	AL
NMWKNNMV	EQ	MHEDIISLWD	ESLKPCVKLT	PLCVTLNCTN	LNNATNANI-			-SQRIEGEMK	NCSFNVTSNI	GNKM	QKEY/	AL
NMWKNNMV	EQ	MQEDIISLWD	QSLKPCVKLT	PLCVTLNCSD	ANVTRATNI-	T ANA RTT SIN NTN TT SDATSTP	SW	GEAIGKGEIK	NCSFNVTTNM	RDKV	KKAY	AN
NMWKNNMV	EQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	LNATEINMNN	ANA	TGTTNTN	ITTKDIQNMK	NCSFNITEGI	KNRR	KKQH/	AI
NMWKNNMV	EQ	MXEDIINLWD	ESLKPCAKLT	PLCVTLNCTD	VTISPNT		AW	SNVTIREEMT	NCSFNITTSI	RDKV	TKEF/	AL
NMWKNNMV	EQ	MQEDIISLWD	QSLKPCVKLT	PLCVTLNCTN DLCVTLNCTN	ANCTNANNTC	DWW	NA	TSSEALRDVK	NCSFKVTTNI	RDKM	KTEY/	ат.
NMWKNDMV	EQ	MHEDIISLWD	OSLEPCVELT	PLCVILICIN	VTNNSTNSTN	K11	INA	-LEOMKGEIK	NCSENTTENT	1		
NMWKNNMV	EO	MHDDTTSLWD	OSLKPCVKLT	PLCVTLNCTN	WNSTTNATDI		TTN	ATDMGEGEIK	KCSEOPR	1	1-	
DAWKNNMV	ΕÕ	MHEDIISLWD	OSLKPCVKLT	PLCVTLNCTD	YNGNVTANNS	SIN	NSSG	GREMEKGEIK	NCSFNVTTNI	1_£		
NMWKNNMV	EQ	MHKDIISLWD	QSLKPCVKLT	PLCVTLNCND	TDTQNI			TRGIEKLEIK	NCSFNITTSI	1.0	- 1	
NMWKNNMV	EQ	MHEDIISLWE	ESLKPCVKLT	PLCVTLHCTD	VNDTKANSNT		TGSS	WDMPGGGEMT	NCSFNITSSI	.ē		
NMWKNNMV	EQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTN	WHVNNNTTSN	NTN	SNEV	DLKMKEGEIK	NCSFNITTGI	1 E	0.07	
NMWKNDMV	EQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	YVGNTTNAGN	TT	NTTST	VSPTASTGMK	NCSFNVTTDL	18		
NAWKNDMV	EQ	MHEDVISLWD	QSLKPCVKLT	PLCVTLNCTD	WKSNATNSTR T KNNTNCCMT		INNS	SEGTMEGQIK	NCSENUTEDI	1 2		
NMWKNNMU	EQ	MHKDIISLWD	OSLKPCVKLT	PLCVALOCTN	VTNKNATNTT	SDATSTP		-SSSEEVOMK	NCSENVISEI	5	0.6	
NMWENNMV	EQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLSCTD	ANRTENTNIT	SDATSTP		-ROMDKEEIK	NCSFNVTTNL	linked		
NMWKNDMV	EQ	MHKDVISLWD	ESLKPCVQLT	PLCVTLQCXN	WTADNNTN		ST	IRREMERNIK	NCSFNITTEV	÷=	1	
NMWKNDMVI	DQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	YYSTNATFNN	NT		KEIK	NCSFNITSSL	2	0.4	
DMWKNNMVI	EQ	MHKDIISLWD	ESLKPCVKLT	PLCVTLDCID	VPTNCTNTG-		NCTN	WKIKEGGELK	NCSFNITTYR	2		
					WMRTNNTLNN	TSNNA	TNQDITRI	NNTMENGEVK		E.	- 1	
1	10	120	130	140	150	160	1/0	180	190	÷.		
LYKLDIVO		N	SNTNNS	NTSYRLISCN	TSVITQTCPK	TSFEPTP				Fraction that	0.2	
FYSODVVP	IE	EEK	SNRT	YRTYRLRSCN	TSVITQACPK	TSFEPIP				gc		
FYKLDIVP	IN	DDDNN	SSNAN	YSTYRLINCN	TSVITOACPK	VTFEPIP				L E		
FHTLDMVP	ΙK	NDNI	SNKN	DTLYRLMHCN	TSVIRQACPK TSAITQACPK	VSFEPIP					0+	_
FYKLDIVP	ID	N	DSN	NNXYRLINCN	TSAITQACPK	VSFEPIP				1	ó	
LYKLDVVP	IN	NR	T	NTSYRLINCD	TSVITQACPK	ISFEPIP				1		
					TSVITQACPK					1		
FIRLDVMP	10	NE	D	TTSTTLINCN NTSVDLISCN	TSVITQACPK TSVITQACPK	VIFEPIP						
LYKTDLVP	IN	DSSSLNNNCS	H-SDNSSVSM	YSRYRLISCS	TSVITQACPK	VSFEPTP				Se	quen	ce
					TSVITQACPK					1		
FYELDITP:	ID	N	S	NTEYRLITCN	TSVLRQACPK	VTFEPIP				P	os	Тор
					TSVITQACPK						37	
					TSVITQACPK					8		
					TSVITQACPK						35	
					TSTLTQACPK TSVITQACPK					1		
					TSVITQACPK							
					TSVITQACPK							
					TSDITQACPK							
	_	1 1				$\sim \sim$	OITE	- /	! .	- 1	- 1.	-

Envelope V1-V2 (and other regions) has highly variable N-linked glycosylation sites. The N-glycosite tool highlights N-x-(S/T) sites and computes statistics on them.

"Optimal" alignment may or may not involve adjusting these sites specifically, depending on what the purpose of the alignment is.



Sequence position(s) of High N-glycosylation Site (>60%)



https://www.hiv.lanl.gov/content/sequence/GLYCOSITE/glycosite.html B.US B.US B.US B. US

Los Ala

B.US.-.9247 pre A1.NEW MH576039 B.US.-.9246_W12_12K21.NEW_MH576084 B.US.-.9245_pre_H6.NEW_MH576107 B.US.-.9244_rebound_S21_G8.NEW_MH576152

B.PH.2015.1003.NEW_MH327746

B. PH. 2016, 1022, NEW MH327757 B.US.-.9255 W12 12C4.NEW MH575379 B.US.-.9254_W12_12BL5.NEW_MH575478 B.US.-.9252_rebound_S71_D12.NEW_MH575680

B.US

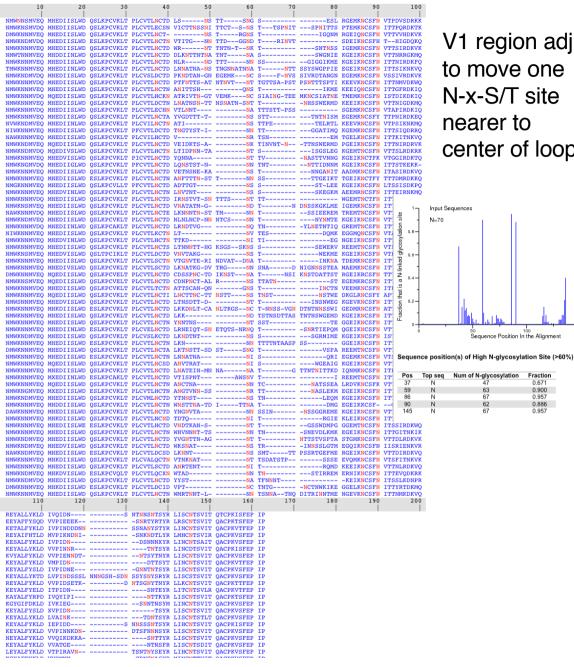
FYKLDVVPID D------ -----------------D NTSYTLINCN TSVITQACPK VSFEPIP FYKIDLVKID EEK----- -----NNS STRYRMINCN TSAITQACPK VSFEPIP FYTLDVAPLD R------ -----------------D NDSYRLISCN TSVITQACPK VSFEPIP FYRLDVVQIK DDNTS----- -----Y NTSYRLINCN SSTITQACPK VSFEPIP

Glycosite can output FASTA file with the N of N-x-(ST) recoded as "O"

B PE 2013. DEMB13PE010. NEW_MH07: B PE 2016. DEMB16PE003. NEW_MH2621: B US610. D14. 22 11. NEW_MH2631: B US610. D14. 12. 11. NEW_MH2631: B US610. D14. 12. 11. NEW_MH2631: B US610. D14. 12. 11. NEW_MH2633: B US610. NO3. D04. TC.1082. NEW_ B PH 2015. 1003. NEW_MH327737 B US8255_W12_1264. NEW_MH275 B US8255_W12_1264. NEW_MH375 B US8252_rebound_S71_D12. NEW_MH375 B US8252_rebound_S71_D12. NEW_MH375 B US8252_rebound_S71_D12. NEW_M1575 B US8252_rebound_S71_D12. NEW_M1575 B US8252_rebound_S71_D12. NEW_M1575 B US8244_rebound_S21_G8. NEW_H1575 B US8244_rebound_S21_G8. NEW_H1575 B US9244_rebound_S21_G8. NEW_H1575 B US9244_rebound_S21_G8. NEW_H1575 B US9244_rebound_S21_G8. NEW_H1575 B US9244_rebound_S21_G8. NEW_H1575 B US9244_rebound_S21_G8. NEW_H157 B US9244_rebound_S21_G8. NEW_H157 B US9244_rebound_S21_G8. NEW_H157 B US9244_rebound_S21_G8. NEW_H157 B US9244_rebound_S21_G8. NEW_H157 B US9244_rebound_S21_G8. NEW_H157 B US9245_D0_2_81128_S52. NEB B US9344_reb0_2_8126_S51. NEW_H157 B US9344_reb0_8	10	20	30	40	50 6	0 70	80	90 100	0 110	120	130 14	0 150
B.PE.2013.DEMB13PE010.NEW MH07;	n m w o n s m v e g m h e d i i		p c v k l t p l c v t l	0 c t d 1 s	- 0 s t t s 0 s		- e s 1 r g e m k 0 c s f		revallykldiv	σidn		liscotsvitater
B.PE.2014.DEMB14PE008.NEW_MH07	n m w <mark>k</mark> n s m v d q m h e d i i	s 1 w d q s 1 k	pcvkltplcvtl	ecsnvicttos	s 0 i t t c t s - 0 i	t t s p 0 i t s p 0 i	ltt s ptemkocsf	fOittpqrdktk	e e y a f f y <mark>s</mark> q d v v	pieeek	s 0 r t y r t y r	1 rscotsvitqacı
B.PE.2016.DEMB16PE003.NEW_MH07;	dmwknnmveqmhedii	s 1 w d q s 1 k j	pcvkltplcvtl		- <mark>0 5 t</mark> r g 0 t	t i g	q n m r g e i q O c s f	fovttvvhdkvk	k e t a l f y k l d i v	pindddn0	ssnaoystyr	lincotsvitqacp
B74 1986 R1296 NEW MH234639	n mwknamvaqmhedii	siwdesiki	pcvkltplcvtl	octovtitg				foweteirdkyk	reyalfhtidmv kogalfykidiv		Bnkodtlyr	lincotsvirgacp
B.ZA.1987.R459.NEW_MH234640	n m w t n n m v e q m h d d i i	nlwdqslk	pcvkltplcvtl	octddlkottn	t 0 a t n t 0	1 B B W G	niekgeikocsf	fovttnrrqkmq	k e y a l l y k l d v v	p i n o r		lincdtsvitgacr
B.ZA.1985.R68.NEW_MH234643	n mw <mark>k</mark> nnmveqmhedii	s 1 w d q s 1 k j	pcvkltplcvtl	0 c t d n l r	- 0 d t t t n (s s g i g g i	. <mark>k m</mark> e k g e i k o c s f	f 0 i t t n i r d k f q.	k v y a l f y k l d v v	pienodt	0 t s y t n y z	liscotsvitqacı
B.US601_D14_2E_19.NEW_MH262;	t mwknnmveqmhedii	nlwdqslkı	pcvkltplcvtl	o c t d l n o a t n a	- <mark>o s t n g n o a t n</mark> o a	t 0 t t s s y s w g p	<mark>opieegeik</mark> oc s f	f 0 i t t <mark>s</mark> i <mark>k n k v</mark> q	k e y a l f y <mark>k</mark> l d v m	p i d <u>o </u>	<mark>d</mark> tt s yt	lincotsvitqacp
BUS - 608 D14 MI10 NEW MH263121	n mwkndmvdqmhedii	siwd q si k j	pcvkltplcvsl	dctdpkodtan Octdptfotte	- g n e g e m K 0 c		ngnegemkocst ntikeevkocst	foittnwydkwa	keyalfysidiv keyallyktdiv		g ontotsyr	liscotsvitgacp
B.US609 D14 12 11.NEW MH263	n m w k n n m v e g m h e d i i	slwdqslk	pcvkltplcvtl	OctnaOittsh	q 0		kmekeeiq o c s f	foittqfrdkiq	revalfykldvv	pidsetk	dotsgoytnyr	llkcotsvitgacr
B.US611_D14_MH7.NEW_MH26336	n <mark>m w k</mark> n n m <mark>v d</mark> q m q e d i i	s 1 w d q s 1 k j	pcvkltplcvtl	h c k o a t r i v t o	- <mark>g t v e m k</mark> 0 d	<mark>s i a i 0 g - teemk 0 c s i</mark> a	t e t m e m k e c s f	f 0 i <mark>s</mark> t d i <mark>k</mark> d k i q	k e y a l f y e l d i t	p i d n	<mark>s n t e</mark> y r	litcotsvlrqacı
B.US613_W23_MAB_12.NEW_MH26	n m w k n n m v e q m h e d i i	s 1 w d q s 1 k j	pcvkltplcvtl	dctnloatns0	- <u>ttns</u> 0 <u>atn</u> - <u>s</u> 0 t	t 0 h s s w e	e <mark>rmdk</mark> eeikocsf	fovttnigdkmq	k a y a l f y r p d i v	qyipi	<mark>0</mark> tt <mark>ky</mark> r	liscotsvitqacp
BUS 2015 A09 CD4T C1 2082 NEW M	n mwknn mvaq mnedii	sindagsiki	pcvkitpicvti pcvkltplcvtl	echovtiont-			ismedemkocst	fyttppirdkiq.	k g y g 1 I d K I d 1 V			liscotsvitgacp
B.PH.2015.1003.NEW_MH327746	n v w k n d m v e q m h e d i i	nlwdqslk	pcvkltplcvil	octoati	0	ttpetel	rtlkeevrocsf	f v t p s i k n k m q	k e y a l l y k l d l v	a i 0 k	tdotsyr	liscotstltgacr
B.PH.2016.1022.NEW_MH327757	n i w <mark>k</mark> n n m v e q m h e d i i	s 1 w d q s 1 k j	pcvkltpfcvtl	d c t d t O g t y s t	- 🚺 n 🧧	t <u>t -</u> g g a t	i m q k g e m k o c s f	f 0 i t t <mark>s</mark> i q d r r q i	k e y a l f y k l d i e	pidd	s n o s s s o t s y r	li <mark>h</mark> cot sv itqacr
BUS9255_W12_12C4.NEW_MH575	n a w k n n m v e q m h e d i i	slwdqslk;	pcvkltplcvil	h c t d v	0	t s n	- emtgelkocsf	f 0 i t t <mark>k</mark> i t n k v q	k e y a l f y k l d v v	pinnkdo	dtsfnonsyr	liscotsvitgacp
BUS9252 rebound S71 D12.NEW	n m w k n d m v d g m g e d i v	s 1 w d g s 1 k	pcvkltplcvtl	hctoltidpon	- t a 0	Bisa	legkgemtocsf	f vttslrdkrk	keval fykldvv	at ge	ntpefr	liscotsditgacr
B.US9251_rebound_S60_H12.NEW	n m w k n n m v e q m h e d v i	s 1 w d q s 1 k	pcvkltpicvtl	octdyqnna	0	ttvoasttv	nngkgeik octf	fkvtggirdktg	leyalfyk ldvt	piravo	tsntoyseyr	lihcotsvitqacr
B.US9250_pre_F9.NEW_MH575728	n m w k n g m v d q m h e d i i	g l w d q s l k j	p c v k l t p f c v t l	octdlqostst	- <mark>n</mark> 0 i	1 t <u>n t</u> 0 t t i d	innmkgeikocsf	foittstkekk-	k g y a f f y k l d v v	q m n n	g r n g O a s y r	minc Ottvitq acr
BUS - 9249 pre_G3.NEW_MH57579t	n mwknn mveq mhedii	siwdqs1kj	pcvkitpicvtl	octovetnsne			iktt gaik oct f	fytttdmrdkka	keyaliykidiv koralfyalduv	qidhdontrandon pion	t n a o a s a o a s a r y r	liscotsvitgacp
B.US9241_pre_L8.NEW_MH575991	nmgknnmveqmhedii	slwdq slk	pcvlltpfcvtl	octdadttgt-	0	s	leekgeikocsf	f ltssisdkpq	eqralfykldvv	q t 0 e t 0 s s	0 k t 0 n s n s y r	liscotsvvtgacr
B.US9247_pre_A1.NEW_MH576039	n <mark>m w k</mark> n n m <mark>v e</mark> q m h e d v i	n 1 w d q s 1 k j	pcvkltplcvtl	octnlovtnt-	- <u></u> 0	<mark>s s</mark> s <mark>k</mark> e	eg <mark>kma</mark> eem <mark>r</mark> oc <mark>s</mark> f	f 0 i t t e i <mark>r n k m</mark> q	k e y a l f y k l d v v	p i d d	<mark>d 0</mark> t <mark>s y</mark> t	l i n c o t s v i t q a c r
B.US9246_W12_12K21.NEW_MH57	n m w k n n m v d q m h e d i i	slwdq slk	pcvkltplcvtl	octdirostvt	- <mark>5</mark> 0 <u>t t t 5</u> 0 1		wgemtoct1	f 0 i t t s i <mark>k d k m k</mark>	k e a a l f y k i d l v	kideek	n s s t r y r	minc ^o tsaitqacp
BUS - 9244 rebound S21 G8.NFW	n mwknsmveqmnedii		pcvkltplcvtl	octelknonto.	- g 0 g		rentrent ocsi	f ovttvirdkyg	tenaifytidva kavalfyrldvy			linconstitutor
B.US9243_W12_12E19.NEW_MH57	n m w k n d m v e q m h e d i i	slwdqslk	pcvkltplcvtl	octdnlnlncp	- n 0 n t c s 0 n	1t n y o	mtekgeikocsf	fovtteirdrvt	kehalfykldvv	pidegvpide		mincotsvitgacr
B.US9242_W12_12H20.NEW_MH57	n <mark>m w k n n m v d</mark> q m h <mark>e d</mark> i i	n 1 w d q s 1 k j	pcakltplcvtl	0 c t d 1 r 0 d t v g	0 (<u> t n y 1 0 e t n</u>	tigg rem toc s f	foitteirdkvr	k e y a l f y k l d v m	p i d r	<mark>d</mark> 0 t s y t	lincotsvitqac _r
B.US91C22_D0_26_11_B5_S82.NE	n i w k n n m v e q m h e d i i	s 1 w d q s 1 k j	pcvkltplcvtl	k c t 0 1 t	0	tesd	iq m k egg m q0c s f	f o v t a g i k d r m q	k e y a l f y k l d i v	piaddk0	sset0ytnyr	liscotsvitqacp
BUS - 91C33 D0 9 4 c9 S49 NFW	n mwknnm vegmnedvi		pcvkltplcvtl				- egngeikocsi	fortnairakvq	k q r a i r y s i d v v		nsoesnyn	lybcottyitgec
B.US91C34_D0_YBAc10_S26.NEW_	n m w k n s m v e q m h e d v i	slwdqslt	pcikltplcvtl	dctdvovtakg	0	t	a <mark>kmek</mark> geikocsf	fovtrslkakvr	tehalfnrldin	pien	<mark>k</mark> ssteya	lincotsvitracr
B.US9341_D0_1_2812A3_S72.NEW	n m w <mark>k</mark> n n m v e q m h e d i i	s 1 w d q s 1 k j	pcvkltplcvtl	sctovtgovte	- <mark>r i n d v a t</mark> d O a	1 t 1 r	n <mark>k o</mark> a t d e m k o c s f	f 0 i t t <mark>s i k d k m</mark> q i	k e y a l f y k l d i v	pieedkk	eontgnyanyr	liscotsvitqacr
BUS-9341_D0_1A_1118G10_587.N	n mwknnmveqmhediv	slwdqslk	pcvkltplcvtl	s c t d l k o a t k g	- dvtrg01	isna d n i g n o s s	teamaemkocsf	foittsikdkrq	reyaifykldiv	piedeortd		liscotsvitgacp
B.US9342 D0 2 3012b10 S11.NE	n mwknsmyeg mgedii	slwdgslki	pcveltplcvtl	octdcdnpoct	- a 1 r 0	ttat0	-stegemkrcsf	foittnmrdgvr	kevalfvkldiv	didson		litcotsvitgacr
B.US9343_D0_3_2812_A2_S55.NE	n m w <mark>k</mark> n n m v d q m h e d i i	s 1 w d q s 1 k	pcvkltplcvtl	tctoattscan	- <mark>q n g</mark> o i	t i (ctnveemkocsf	foiteriknkvk	k e y a l f y g v d v i	p i d 0	e s y t	l r h c o t s v i t q a c r
B.CHR282_3.4_cl56.NEW_MH7143	n m w <mark>k</mark> n n m v e q m q e d i i	s 1 w d q s 1 k j	pcvkltplcvtl	octilocttoc	- <mark>t t 0 s t t</mark> 0 s	t 0 s t 0 s	twedkglrocsf	fyaptsirdkvk	k e y a i f y n l d l v	qidddle k	<mark>t n k o f t k</mark> y r	1 v s c o t s v i t q a c p
BCH1294_3.8_Cl28.NEW_MH7143.	n mwsnnmveqmhedvi	s 1 w d q s 1 k j	pcvkltplcvtl	octdltnsdtt octdl t kdolt			wegkgevkocst	f 0 1 t t s 1 r d k v q	k g y a t f y n l d i v		ddronsyr	liscotsvitgacp
B.CHR11_8.1_cl57.NEW_MH71435	n m w k n n m v e q m h e d i i	slwdqslk	pcvkltplcvtl	octdlkk		itstnsdttastntnswg	e m d k g e i k c s f	foittnildkrq	k h y a l f y k l d v t	pie0	etnsyr	ltscotsvitgacr
B.CN.2005.05071.NEW_MK053935	n <mark>m w k</mark> n n <mark>m v e</mark> q m h e d i i	s 1 w d q s 1 k j	pcvkltplcvtl	octnyontns-	0	s s t	- <mark>teggeik</mark> oc s f	f 0 i t t <mark>s</mark> i n t k v -	k d y a l f y n l d i v	q i <mark>s 0</mark>	<mark>d s</mark> n s y r	lincotsvitqtcr
B.US.2013.ARC-1a.NEW_MK214316	nmwknnmveqmhedii	nlwdeslkj	pcvkltplcvel	octdlrneiqt	- s e t q t s - n r o d	1 t s 0 r t i e	epqmqgeikocsf	f O v t t d w k d k k k	kehalfyafdlt	a i e n	d <u>o t s y r</u>	ltscoatsitqacp
B.AR.2000.85096FL 2000.KY968401	n myknomyeg mhedii	sludes 1 k	pcvkltplcvil		<mark>.</mark>	ttttntaaspss		foittaiakkyr	sevaif nkldiv			liscotsvitgacr
B.AR.2000.85323FL_2000.KY968402	d m w k n n m v e q m h e d i i	s 1 w d q s 1 k	pcvkltplcvtl	octalktostt	- sdst s 🔾	It	spareemt ocsf	fovttnirnkiq	k e y a l f d k l d v v	p t q d	n 0 n s y i	lrng asvltqac
B.AR.2000.85891FL_2000.KY968403	n m w k n n m v e q m h e d i i	s 1 w d e s 1 k j	p c v k l t p l c v t l	octnlnoatna	0 :	s	q riegemk oc s f	f • v t s n i g n k m q	k e y a l f y k l d v v	p i e g	<mark>d</mark> 0 t s y t	lihcotsvitqac _r
BAR.2000.86218FL_2000.KY968404	n mwknn mveq mqedii		pcvkitpicvti	ocsdaovtrat			t k dig n m k c c s f	foitegiknerk	kayan fynidiv kghaifdkldvy			I R S C O L S V I L Q A C P
B.AR.2001.101815FL_2001.KY96839	n m w k n n m v e q m × e d i i	nlwdeslk	pcakltplcvtl	octdvtispnt	a w s 0	7 t	ireemtocsf	foittsirdkyt	kefalfyrddvv	p i d n	Oytnyr	lincotsvitgacr
B.AR.2001.107878FL_2001.KY96839	n <mark>m w k</mark> n n <mark>m v e</mark> q m q e d i i	slwdq slk	p c v k l t p l c v t l	octnaoctna-	0	1 t c 0 a t s	ssealrdvkocsf	f <mark>kvttnirdkm</mark> k	teyalfykldlv	p i d n	d 0 t s y t	liscotsvitqacı
B.AR.2001.89825FL_2001.KY968407	n m w k n g m v e q m × e d i i	s 1 w d e s 1 k j	p c v k l t p l c v t l	h c t n a o g t v n o	- 5 5 0	tt 0 a s 1	. e <mark>k m</mark> egeikocsf	foittgrgnemk	k e y a l f y t f d v v	p 1 n d	<mark>d</mark> 0 t s y m	lthcotsvikqacp
BAR.2002.02AR114146.D0383746	n mwkn n mye g m h ed 11		pcvkltplcvml		- t d i t t o	t 1	dmaegeikkogf	f gprrekav	keyalfrkldlm			liscotsvitgac
B.AR.2002.116877FL_2002.KY96839	dawknnmveqmhedii	s 1 w d q s 1 k	pcvkltplcvtl	o c t d y n g o v t a	n (ssin 0 ssggi	emekgeikocsf	fovttnirdkvq	k a n a l l y k l d v v	p i d 0	dtssyr	liscotsvitgacr
B.AR.2003.03AR137681.DQ383748	n m w k n n m v e q m h k d i i	s 1 w d q s 1 k	p c v k l t p l c v t l	ncodtdtg	0	E	giekleik ocsf	foittsirdkvq	r e y a l f y k l d v v	p i d n	 0 B t r y r	liscotsvitqac _r
BAR 2003 134742EL 2003 KV0682C	n m w k n n m v e q m h e d i i	siweesik	pcvkltplcvtl	h c t d v o d t k a n	0		1 m p g g g e m t o c s f	101tssirdkwq	keyalfyrldvv keyalfyrldvv	pidnn	Onsidotsyr	liscotsvvtqacp
B.AR.2004.04AR151263.DQ383751	n m w k n d m v e g m h e d i i	s 1 w d q s 1 k	pcvkltplcvtl	octdyygottn	- a g 0	t 0 t t s t v	ptastgmk ocsf	fovttdlrdkvk	knyalfysldvv	pien		liscotsiitgacr
B.AR.2004.04AR151516.DQ383752	n m w k n d m v e q m h e d v i	s 1 w d q s 1 k	p c v k l t p l c v t l	octdwksoat-	0	t z i n • s s 1	lgtmegqik ocsf	fniisrienkvr	k e x a l f n s l d v v	x i d k	e g s y m	lincotsvitqacı
B.AR.2004.145447FL_2004.KY96839	n a wkndm veqmhedii	slwdeslk	p c v k l t p l c v t l	d c s d l k o n t	0	smtttpssrtge	fmergeikocsf	fovttdi <mark>rdkv</mark> q	qqnayfykldiv	p i d n	<mark>t</mark> o <mark>r s y</mark> r	llscotsvitqac _r
BAR.2004.159687FL_2004.KY96839 BAR.2007.DEURE07AR001.KV65868	n mwann myaqmh kdii		pcvkitpicval	q c t o v t n K o a t		t s a t s t p	sseevqmk o csf		keyalfysidvv			liscotsitqacp
B.AR.2008.DEMB08AR002.KY658683	n m w k n d m v e q m h k d v i	slwdeslk	pcvqltplcvtl	q c x 0 w t a d	0	1 t 0 s t i r	remernikocsf	foittevqdkrk	kayalfdsldiv	pidn	0 n t n y n	lincotsvitgacr
B.AR.2010.DEMB10AR006.KY658684	n m w k n d m v d q m h e d i i	slwdq slk	p c v k l t p l c v t l	0 c t d y y s t	0	tfnont	k e i k o c s f	foitsslkdnpr	k e h a l f y t v d i v	p i k h	d n t r y r	ltkcotsvitqacı
B.AR.2014.DEMB14AR003.KY658685	dmwknnmveqmhkdii	s 1 w d e s 1 k j	pcvkltplcvtl	dcidvpt			ikeggelk ocsf	101ttyrtdkmq	k g n a l f y n l d i i	kidne	<mark>g 0 n s y r</mark>	llscotsvitqacp
1 3 3 3 3 5 3 A RCH 0 3 4 A 1 0 3 7 2 0 8			POT ALCPICATI		n	e mu o m e m d a r e t 1 o t		e c c c n m k u K v g	a a ya			

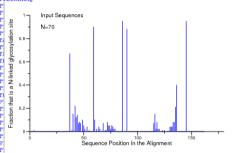


B.PE.2013.DEMB13PE010.NEW_MH078552 B.PE.2014.DEMB14PE008.NEW_MH078553 B.PE.2016.DEMB16PE003.NEW_MH078555 B.PE.2016.DEMB16PE003.NEW_MH0785554 B.PE.2016.DEMB16PE009.NEW_MH078555 B.ZA.1986.R1296.NEW MH234639 B.ZA.1987.R459.NEW_MH234640 B.ZA.1985.R68.NEW_MH234643 B.US.-.601_D14_2E_19.NEW_MH262735 B.US.-.605_W29_h8.NEW_MH263121 B.US.-.608_D14_MI10.NEW_MH263144 B.US.-.609_D14_12_11.NEW_MH263194 B.US.-.611 D14_MH7.NEW_MH263362 B.US.-.613_W23_MAB_12.NEW_MH263520 B.US.-.616_W23_MM_15.NEW_MH263630 B.US.2015.A09.CD4T.C1.20B2.NEW_MH264238 B.PH.2015.1003.NEW_MH327746 B.PH.2016.1022.NEW MH327757 B.US.-.9255_W12_12C4.NEW_MH575379 B.US.-.9254 W12_12BL5.NEW_MH575478 B.US.-.9252_rebound_S71_D12.NEW_MH575680 B.US.-.9251 rebound S60 H12.NEW MH575694 B.US.-.9250 pre F9.NEW MH575728 B.US.-.9249_pre_G3.NEW_MH575796 B.US.-.9248_pre_B8.NEW_MH575857 B.US.-.9241_pre_L8.NEW_MH575991 B.US.-.9247_pre_A1.NEW_MH576039 B.US.-.9246 W12 12K21.NEW MH576084 B.US.-.9245_pre_H6.NEW_MF576107 B.US.-.9244_rebound_S21_G8.NEW_MH576152 B.US.-.9244_W12_12E19.NEW_MH576262 B.US.-.9242_W12_12H20.NEW_MH576349 B.US.-.91C22_D0_26_11_B5_S82.NEW_MH632763 B.US.-.91C35_D0_3HG_010517_D6_S46.NEW_MH632822 B.US.-.91C33_D0_9_4_c9_S49.NEW_MH632949 B.US.-.91C34_D0_YBAc10_S26.NEW_MH632956 B.US.-.9341_D0_1_2812A3_S72.NEW_MH633031 B.US.-.9341_D0_1A_1118G10_S87.NEW_MH633039 B.US.-.9342_D0_2_2812a6_S1.NEW_MH633155 B.US.-.9342_D0_2_3012b10_S11.NEW_MH633164 B.US.-.9343_D0_3_2812_A2_S55.NEW_MH633199 B.CH.-.R282_3.4_c156.NEW_MH714324 B.CH.-.R28<u>3</u>, 4_C156.NEW_MH/1432 B.CH.-.T294_3.8_C128.NEW_MH714329 B.CH.-.T11_10.6_C117.NEW_MH714343 B.CH.-.R11_8.1_C157.NEW_MH714351 B.CM.2005.05071.NEW_MK053935 B.US.2013.ARC-1a.NEW_MK214316 B.FR.1983.HXB2-LAI-IIIB-BRU.K03455 B.AR.2000.85096FL_2000.KY968401 B.AR.2000.85323FL_2000.KY968402 B.AR.2000.85891FL_2000.KY968403 B.AR.2000.86218FL_2000.KY968404 B.AR.2000.ARMS008.AY037269 B.AR.2001.101815FL_2001.KY968394 B.AR.2001.107878FL_2001.KY968395 B.AR.2001.89825FL_2001.KY968407 B.AR.2001.89886FL_2001.KY968408 B.AR.2002.02AR114146.DQ383746 B.AR.2002.116877FL 2002.KY968396 B.AR.2003.03AR137681.DQ383748 B.AR.2003.03AR138910.DQ383749 B.AR.2003.134742FL 2003.KY968397 B.AR.2004.04AR151263.DQ383751 B.AR.2004.04AR151516.D0383752 B.AR.2004.145447FL_2004.KY968398 B.AR.2004.159687FL 2004.KY968399 B.AR.2007.DEURF07AR001.KY658686 B.AR.2008.DEMB08AR002.KY658683 B.AR.2010.DEMB10AR006.KY658684 B.AR.2014.DEMB14AR003.KY658685 B. AR. 1998. ARCH054. AV037268 B.PE.2013.DEMB13PE010.NEW MH078552 B.PE.2014.DEMB14PE008.NEW_MH078553 B.PE.2016.DEMB16PE003.NEW MH078554 B.PE.2016.DEMB16PE009.NEW MH078555 B.ZA.1986.R1296.NEW MH234639 B.ZA.1987.R459.NEW_MH234640 B.ZA.1985.R68.NEW_MH234643 B.US.-.601_D14_2E_19.NEW_MH262735 B.US.-.605_W29_h8.NEW_MH263121 B.US.-.608_D14_MI10.NEW_MH263144 B.US.-.609_D14_12_11.NEW_MH263194 B.US.-.611_D14_MH7.NEW_MH263362 B.US.-.613 W23 MAB 12.NEW MH263520 B.US.-.616_W23_MM_15.NEW_MH263630 B.US.2015.A09.CD4T.C1.20B2.NEW_MH264238 B.PH.2015.1003.NEW_MH327746 B.PH.2016.1022.NEW MH327757 B.US.-.9255 W12 12C4.NEW MH575379 B.US.-0254_W12_12L5.NEW_MH575478 B.US.-0254_W12_12L5.NEW_MH575478 B.US.-0252_rebound_S71_D12.NEW_MH575680 B.US.-0251_rebound_S60_H12.NEW_MH575694



V1 region adjusted to move one N-x-S/T site nearer to center of loop.

100



Pos	Top seq	Num of N-glycosylation	Fraction
37	N	47	0.671
59	N	63	0.900
86	N	67	0.957
90	N	62	0.886
145	N	67	0.957

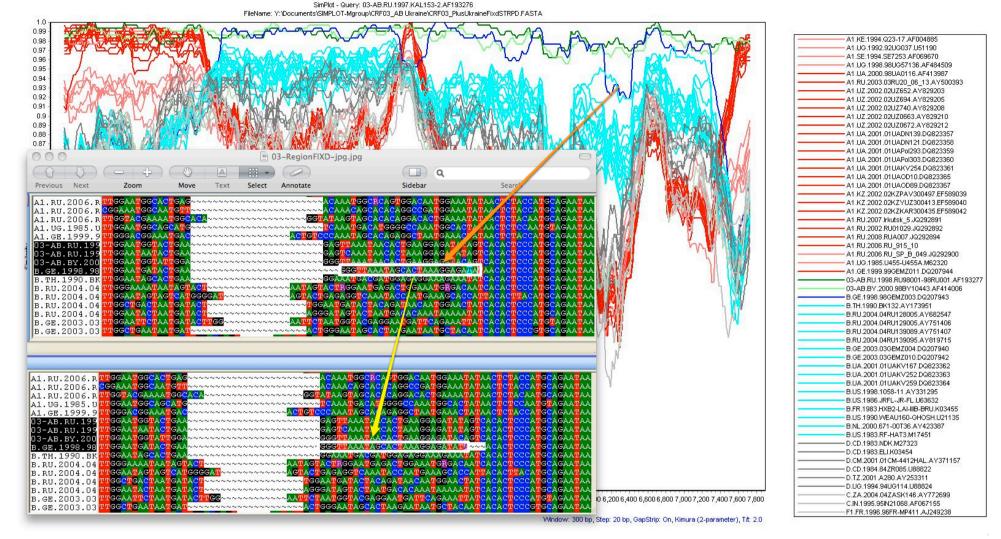


•	AliView - *03_PlusRefs_GeneCut_May2019.FASTA				
👫 📕 🚼 🛣					
	2920 2930 2940 2950 2960 2970 2980 2990 3000 3010 3020 3030 3040	3050 3060 3070			
5A_Z321_U76035	GARATTE CAR AT COARAA COARAA COARAATTICAARATTECCE CCTOARATCCATACAR A CTCCAR TATTICCA ATAACAAAA COACAATACAACAACAACAACAACAA	CTGAATAAAAGAACTCAAGA <mark>T</mark> TTCTG			
D.1997.97CDKTB48.AF286238 M.2001.01CM-1445MV.GU201516	AliView Highlight Differences from Majority	CTCAATAAGAGAACTCAAGACTTCTG CTCAATAAGAGAACTCAAGACTTCTG			
/.1994.94CY017-41.AF286237 I.1996.DDJ360.AY521630	GAATT TOTA AAATTGAA AATTGAAGAATTACAAGATTT AAAATTA AAAATTA AAAATTA AAAATTA AAAATTA AAAATTA AAAATTA AAAATTA AAAAT	CT GARTAR GARGARCT CARGACTTCTG			
.2001.DDI579.AY521629 .2001.DDJ369.AY521631	AliView Highlight Differences from Majority	CTGAATAAAAGAACACAGGACTTCTG			
.2003.PS1044-Day0.DQ676872		CTGAATAAAAGAACTCAAGACTTTTG			
7.1992.92RW008.AB253421 7.1992.92UG037-A40.AB253429	GAARTTGT BAGAGAATGAAAATGGAAAATTGGAGAAAATTGGAGCCTGAAAATCGCATACACCATACTCCATACGCAAAATGGAGAAAATGGAGGAAATTGGAGGAAATTGGAGGA	CTGAATAAAAGAACACAAGACTTTTG CTGAATAAAAGAACTCAAGACTTTTG			
1997.97BL006.AF193275 1999.99GEMZ011.DQ207944	GACATTE TAA GGACATA GANAAGGAAGGAAAAATTT CAAAAATT E GECT GANAAT CCATAGAATACT CCAGTATTT GCAAAAAAGGAACAA GACAAGGAAGAAATTAGTAATTT CAAATTT CAAATTT CAAATTT CAAGAAAAAGGAACGAAGGAAGAAATTT CAAATTT CAAATTT CAAATTT CAAGAATAA	CT CAATAAAAGAACT CAGGACTTTTH CT CAATAAAAGAACT CAGGACTTTTG			
2002.02KZKAR300435.EF589042 2002.02KZPAV300480.EF589043	G & C & T T T C T A 16 G & M A 7 G G A 16 A 8 G A 1 A 1 T T C A 1 A A 17 G G C C C A 1 A 1 A C C A C A C A 1 T C C C A C A 1 A 1 C C C A C A 1 A 1 C C C A C C A 1 A 1	CTGARTAARAGRACTCAGGACTTTTG			
2002.02KZPAV300497.EF589039 2002.02KZPAV300502.EF589044	G	CTGAATAAAAGAACTCAGGACTTTTG			
2002.02KZYUZ300413.EF589040	GA CATTE TA COACATA COACAT	CTGAATAAAAGAACTCAGGACTTTTG			
2002.02KZYUZ300425.EF589041 .2000.RU00051.EF545108	GA CATTTETA CGA CATEGAAAGGAAGGAAAAAATTTCAAAAATTCGCCTGAAAATCCCCAGAATTCGTACATAAGAAAAAGGAAGAAAAAGGAAGG	CTGAATAAAAGAACTCAGGACTTTTG			
.2003.03RU20-06-13.AY500393 .2000.98UA0116.AF413987	GA BATTE TAA GGABAT GGABAAGGABAAAATTT CAABAATT GGCCT GABAT CCATAGATACT CCAGTATTT GTAABAAAAAGABAAAGABCACAGTAATGGAAGAAATTAGTAGGATTT CAAGATTACTAGTAGTAGAAGAAAAAGABCAACAGTAATTGGAAGAAATTAGTAGTTT CCAGTTT CCAGTACTAATTC CTAGTATTC CTAGTAGTAATTGGAAGAAAAAGABCAACAGTAATTGGAAGAAAAATTGGAAGAAAATTGGAAGAATTTC CAAGTAATTGGAAGAAAATTGGAAGAAAATTGGAAGAAAATTGGAAGAA	CT CAATAAAAGAACT CAGGACTTTTG CT CAATAAAAGAACT CAGGACTTTTG			
.2001.01UADN121.DQ823358 .2001.01UADN139.DQ823357	GA CATTEGTA, GGA DET CGARANAGGARGGARAATTEGARGCEGARATEGGEGEGARATEGARGCEGARATEGARGARAARGCACAGGARATEGARGARAT	CTGARTAARAGRACTCAGGACTTTTG			
.2001.01UAKV254.DQ823361 .2001.01UA0D10.DQ823365	G 1 T T T T T T G G G A T G G A A A G G A A A A	CTGAATAAAAGAACTCAGGACTTTTG			
2001.01UAOD35.DQ823366 2001.01UAOD89.DQ823366	CAL ATTTC TER CALATACANAAA CAACAAAAAAAAAAAAAAAAAAAAAAAAAA	CTGAATAAAAGAACTCACGACTTTTG			
.2001.01UAPol294.DQ823356		CTGARTARAGAACTCAGGACTTTTG			
2001.01UAPol303.DQ823360 2002.02UZ0659.AY829209	GA BATTE TAR GGA BAT GGA BAAGGAAGAAAATTT CAAAAATT GG GC T GAAAAT CG TA CAACAT CT CCAG TATT CG TA TAAGAAAABGAACA G ACTAA T GGA GAAATTA CTAGT GT CT TT CAAGAAAA	CTCAATAAAAGAACTCAGGACTTTTG CTCAATAAAAGAACTCAGGACTTTTG			
2002.02UZ0663.AY829210 2002.02UZ0672.AY829212		CT GARTARARGARCT CARGACTTTTG			
2002.02UZ652.AY829203 2002.02UZ694.AY829205	G	CTGAATAAAAGAACTCAGGACTTTTG			
2002.02UZ698.AY829206		CTGAATAAAAGAACTCAGGACTTTTG			
2002.02UZ740.AY829208 .BY.2000.98BY10443.AF414006	GACATTEGTALGGADATGGADAAGGAAGGAAAAATTECAAAAATTGGGCCTGAAAATCCATACAATACTECAATAATTEGTALTAAAGAAAAAGATAGGACAAAATGGAGAAAATTAGTAGATTECAAGAAT GACATTEGTALGGADATGGADAAGGAAAGGAAAGATTECAAAATTEGGCCTGADAATCCATACAATACTECCAGTATTTGCCATAAAGAAAAAGATAGGACGAAAATTAGTAGATGAAGATTECAAGAA GACATTEGTALGGADATGGADAAGGAAGGAAAGATTECAAAATTEGGCCTGAAAATCCATACATCTECAGTATTTGCCATAAAGAAAAAGATAGGACGAAAATTAGTAGA	CT CARTARAGAACTCARGACTTTTG			
.GB.2013.13774-1-51.3.MF109476 .RU.1998.RU98001-98RU001.AF193277					
.RU.1997.KAL153-2.AF193276 998.98GEMZ003.DQ207943		CT LATTAGAGAACTCAAGACTTCTG			
011.11RU21n,JX500708 004.04RU128005.AY682547	G A A T T T C T A G A A T G G A A A G G A A G G A A G G A A A T T T C A A A T T G G C C T G A A A T T C C A T A C A T A C T C C A T A C A T T C C A T A C A C	CTEANTARGAGAACTCAAGACTTCTG			
004.04RU129005.AY751406	GALATTTGTAGAGAAAAGGAGAAAAAGGGAGAAAATTGGGGTGAAAATTGGGGTGAAAATTGGGGTGAAAATTGGGGGTGAAAATTGGGGGTGAAAATTGGGGGTGAAAATTGGGGGG	CTRATING CALCECTCALCACTTCTC			
004.04RU139089.AY751407 004.04RU139095.AY819715	GARTTEGTA BAGARTGGARGARAATTECARAATTEGGCCTGARAATCCCCATACATACCCCATATAGARAAAGACACAGAAAAAGACAGAAAATGGAGAAATTAGTAGGATTE GARATTEGTA BAGARAGGARAGGARAATTEGAAGTATAATTEGGCCTGARAATCCCCATACATACGACATTETGCCATAAAGAAAAAGACACAGAAATTAGTAGAAATTEG				
003.03GEMZ004.DQ207940 003.03GEMZ010.DQ207942	GAATTTE TA DAGAATE GAAAG GAAGAG MAAATTE CAAAATTE GECETGAAATE CEATAGATACTE CAATAATE CAAAAAAGACAE GAATAATE GAGAAATTE CAAGATTE CAAGAAAATG GAAGAATTE CAAGATTE CA	CTEALTARGAGARCTCARGACTTCTG CTEALTARGAGARCTCARGACTTCTG			
001.01UAKV167.D0823362 001.01UAKV252.D0823363		CTEATTARGAGARCCCAAGACTTCTG			
001.01UAKV259.DQ823364 010.10RU6629,JX500707	G A A T T T G T A A A G A A A T G G A A A A	CTEANTARGAGAACTCAAGACTTCTG			
009.09RU4457JX500709	GAAATTTGTA, AGAAAAGGAAAAAGGAAAAAGGAAAAATTTGTAGAATTCGATGATGAAAATTCGATGAAAAATTGGAGAAAATTGGAGAAAAATTGGAAGAA	CT LATAAGAGAACTCAAGACTTCTG			
000.671-00T36.AY423387 990.BK132.AY173951	GARTTEGTA <mark>BAGA K</mark> TEGARARGARGARARATTECARARATEGGECTGARATECGATACATACTECAR ^I TATTEGCATARGARARAGACAG <mark>B</mark> ACTARTEGAGARATTEGAGATTECGAGARATTEGAGATTECGAGARA	.CTEALTAAAGAACTCAAGACTTCTG .CTEALTAA <u>A</u> AGAACTCAAGACTTCTG			
983.HXB2-LAI-IIIB-BRU.K03455 998.1058-11.AY331295	GAATTTE TARGA AF GAAAA GGAAG GGAAAATTTCAAAATTE GGCCTGAAATCCCATACAATACTCCCATATATAAAAAAAA	CTEALTARCAGAACTCAAGACTTCTG CTEALTARCAACAACTCAAGACTTCTG			
992.BR025-d.U52953 986.ETH2220.U46016	G AATTET TE WEAATEGA AA GGAAGAAAATTA CAAAATTEGGCCTGAAAATCCATA AA AACACTCCATAAAAAACAA GAACAATAATGAAGAATTA CAAGAA	CT CARTARAGACTT AAGACTTTTG			
995.95IN21068.AF067155 004.04ZASK146.AY772699	G ALTTTET G A GALTEGA GALALATTEGA GALALATTEGA CCTALALATCCATA ALGALCTCCATALALALALALALALALA ALGALGALATATATAGA GALALATTEGA CCTALALATAGA GALALTTEGA CCTALALTGA CCTALALTG	CTGAATAAAAGAACTCAAGATTTTTG			
983.ELI.K03454		CTEALTAR GAGAACTCAAGATTTCTG			
2001.01CM-4412HAL.AY371157 001.A280.AY253311		CTEALTAAGAGAACTCAAGATTTCTG			
994.94UG114.U88824 993.VI850.AF077336	GARTIETE BAGA DIE GARARGGARAGENAATTICAARATIEGGCIEGARAGICCATACANTAICICCATATAGARARGGACAGTACTAATGARATAGGACAGTA	. C T L A T A A G A G A A C T C A A G A C T T T T G . C T L A T A A A A A G A A C T C A A G A <mark>T</mark> T T <u>C</u> T G			
993.93BR020-1.AF005494 993.FIN9363.AF075703	GANT TGTA MGA ATTGGA NA GGA NA ATTTGA NA ATTGGGCCTGA NA TGCATGA NA TCTCA GA TA TTGCA GA NA A A GA CA GTA CTA ATGGA GGA NA ATTGCA GGCTGA NA TGCA GGA GA ATTGCA GGA GA NA TTGCA GGA GA NA TTGCA GGCTGA NA GCA GGA GA NA TTGCA GGA NA TTGCA G	. С Т <mark>Е</mark> ААТ А А А А С А А С Т С А А С А Т Т Т Т Т Т С С Т С А А Т А А А А А С А А С Т С А А С А Т Т Т Т			
996.96FR-MP411.AJ249238 1995.95CM-MP255.AJ249236	GALT TETA TAS TAS GALAL GALGALA MATTCALALAT GEOCONTALAT CONTACALT CTCOLATATTCEALALAGEALALAGEACATATAATEGAGALATTEGAGALATTEGAGEATTTAGAGAL	CTEAATAAAAGAACTCAAGATTTTTG			
1995.95CM-MP257.AJ249237 1997.CM53657.AF377956		CTEALTAAAAGAACTCAAGATTTTTG			
2002.02CM-0016BBY.AY371158	GAAAA GAAAAA GAAAAA GAAAAAA GAAAAAA GAAAAAA	CTEASTASASGASCTCASGATTTTTG			
996.DRCBL.AF084936 993.HH8793-12-1.AF061641		CTCAATAAAAGMACTCAAGACTTCTG CT <u>C</u> AATAAAAGAACTCAAGACTTCTG			
1992.92NG083-JV10832.U88826 .PT2695.AY612637	GAATTEGTA HAG MATGGAAHAGGAAGGAATTECAAAATTEGGCCTGAAATCCATA AAMACCTCCA TATTEGCCATAAAGGAAGAATAGGAAGAATTEGTAGGAAGAATTEGTAGGAAGAATTEGTAGGAAGAATTEGTAGGAAGAATTEGTAGGAAGAATTEGTAGGAAGAATTEGTAGGAAGAATTEGTAGGAAGAATTEGTAGGAAGATTEGTAG	CTEALTAAAAGAACTCAAGACTTCTG			
993.VI991.AF190127 993.VI997.AF190128		CTEALTAAAAGAACTCAAGACTTCTG			
L990.056.AF005496		CTGARTARAGAACTCAAGACTTCTG			
2000.00GBAC4001.FJ711703 997.J-97DC-KTB147.EF614151	URATTTY TO TO A URAT TO GAARAA ATTTU A MAAT MUGUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU	CT CAATAAAA GAACTCAAGACTTCTC			
2004.04CMU11421.GU237072 993.SE9280-7887.AF082394	GARTTETTERAGARTEGARARGARGARARATTTCARRATTGG:CCTGARATCCATA AR ACTCCATATAGARARARGACAGTACTAATGGACARATTGAGARARTTGAGA GARTTETTECAATTGAARGARGARGARATTTCARATTTCARATTTGGACGCCTGARATCCATATAR	CTGAATAARAGAACTCAAGACTTITG CTGAATAAAAGAACTCA			
1997.97ZR-EQTB11.AJ249235 1996.96CM-MP535.AJ249239		CTGRATRARGARCTC CAGACTTCTG			





2220 0206A_Z321_U76035 A2.CD.1997.97CDKTB48_AF286 A2.CM.2001.01CM-1445MV.GU A2.CY.1994.94CY017-41.AF28 A1.AU.2003.P51044-Day0.DQ6 A1.RW.1992.92RW008.AB25342 A1.LW.1992.92RW008.AB25342 A3.SN.2001.DD1579.AY521629 A3.SN.2001.DD1579.AY521629 A3.SN.2001.DD1579.AY521629 A3.SN.2001.DD1579.AY521629 A3.SN.2001.DD1369.AY521631 A6.KZ.2002.02KZPAV300492.Ef A6.KZ.2002.02KZPAV300492.Ef A6.KZ.2002.02KZPAV300492.Ef A6.KZ.2002.02KZPAV300435.Ef A6.KZ.2002.02KZPAV300502.Ef A6.KZ.2000.98UA0116.AF4139 A6.KJ.2001.01UADN121.DQ82 A6.UA.2001.01UADN121.DQ82 A6.UA.2001.01UADN121.DQ82 A6.UA.2001.01UADN123.DQ82 A6.UA.2001.01UADN89.DQ823 A6.UA.2001.01UADN89.DQ823 A6.UA.2001.01UADN89.DQ823 A6.UZ.2002.02UZC55.AY8292 A6.UZ.2002.02UZC55.AY8292 A6.UZ.2002.02UZC653.AY8292 A6.UZ.2002.02UZC653.AY8292 A6.UZ.2002.02UZC653.AY8292 A6.UZ.2002.02UZC653.AY8292 A6.UZ.2002.02UZC653.AY8292	Aliview Control-Click on Sequence to highlight differences from that sequence	
03-AB.CB.2013.13774-1-51.3. B.CE.1998.98CEMZ003.DQ2075 B.RU.2011.11RU21nJX500708 A.RU.2004.04RU128005.AY682 B.RU.2004.04RU129005.AY751 B.RU.2004.04RU139095.AY819 B.RU.2003.03CEMZ010.DQ2075 B.CE.2003.03CEMZ010.DQ2075 B.UA.2001.01UAKV252.DQ8233 B.UA.2001.01UAKV252.DQ8233 B.UA.2001.01UAKV252.DQ8233 B.RU.2000.0671-0073.AY42331 B.RU.2000.0671-0073.AY42331 B.RU.2000.0671-0073.AY42331 B.RU.2000.671-0073.AY42331 B.RU.2000.671-0073.AY42331 B.RU.2000.671-0073.AY42331 B.RU.2009.09RU4457.JX500709 B.RU.2000.671-0073.AY42331 B.RU.2009.09RU455.AY42331 B.RU.2009.09RU455.AY42331 B.RU.2009.09RU455.AY42331 B.RU.2009.09RU455.AY42331 B.RU.2009.09RU455.AY42331 B.RU.2009.09RU455.AY42331 B.RU.2009.09RU455.AY42331 B.RU.2009.09RU455.AY42331 B.RU.2009.09RU455.AY42331 B.RU.2009.09RU455.AY42331 B.RU.2009.09RU55.AY42341 B.RU.2009.09RU55.AY42341 B.RU.2009.09RU55.AY42341 B.RU.2009.09RU55.AY44341 B.RU.2009.09RU55.A		
C.ET. 1986. ETH2220. U46016 C.IN. 1995. 59IN21068. AF667.15 D.C.A.2004.04ZASK146. AY7726 D.C.D. 1983. ELI.K03454 D.C.M.2001.01CM-4412.HAL.AY D.TZ.2001.A280. AY253311 D.UG. 1993. 938R020-1. AF0054 FI.BE. 1993. 038R020-1. AF0054 FI.BE. 1993. 038R020-1. AF0054 FI.FE. 1993. 038R020-1. AF0054 FI.FE. 1993. 038R020-1. AF0054 FI.FE. 1995. 95CM-MP255. AJ24 F2.CM. 1995. 95CM-MP255. AJ24 F2.CM. 1995. 95CM-MP255. AJ24 F2.CM. 1995. OSCM-MP255. AJ24 F2.CM. 1995. OSCM-MP255. AJ24 F2.CM. 1995. OSCM-MP257. AJ24 F2.CM. 1995. OSCM-MP257. AJ24 F2.CM. 1993. CMS 0567. AF37795 F2.CM. 2002. 02CM-00168BY.AY G.BE. 1993. M991. AF190127 H.BE. 1993. M991. AF190127 J.CD. 1997. J-PJC-KTB147.EF6 J.CM. 2004.04CMU11421. CU257 J.CM.2004.04CMU1421. CU257 J.CM.2004.04CMU1421. CU257 J.CM.2004.04CMU1421. CU257 J.CM.2004.97.972R-FC081.47.EF6 J.CM.2004.97.972R-FC081.47.EF6 J.CM.2004.97.972R-FC081.47.EF6 J.CM.2004.97.972R-FC081.47.EF6 J.CM.2004.97.972R-FC081.47.EF6 J.CM.2004.97.972R-FC081.47.EF6 J.CM.2004.97.972R-FC081.47.EF6 J.CM.2004.97.972R-FC181.47.EF6 J.CM.2004.97.972R-FC181.47.EF6 J.CM.2004.97.972R-FC181.47.EF6 J.CM.2004.97.972R-FC181.47.EF6 J.CM.2004.97.972R-FC1814.47.EF6 J.CM.2004.97.972R-FC1814.47.EF6 J.CM.2004.97.972R-FC1814.47.EF6 J.CM.2004.97.		



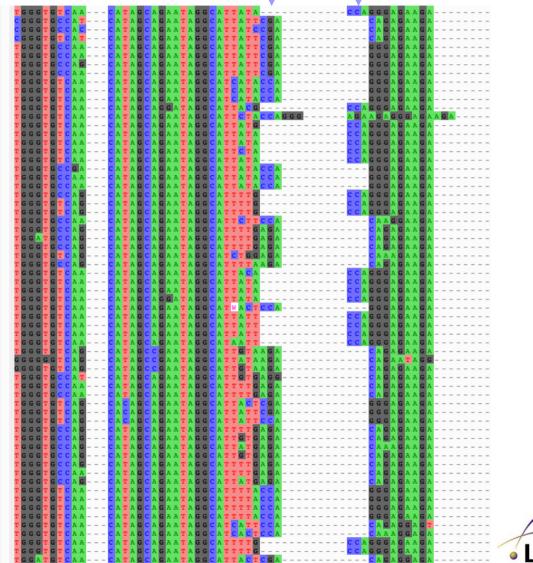
Even a very small region of misalignment, hypermutation, or poor sequence quality can have a large impact on similarity plots, phylogenetic trees, and other analyses. Similarity plots can be quite useful for identifying sites in a multiple sequence alignment that should be scrutinized, and corrected if in error, as this example shows.

https://sray.med.som.jhmi.edu/SCRoftware/simplot/



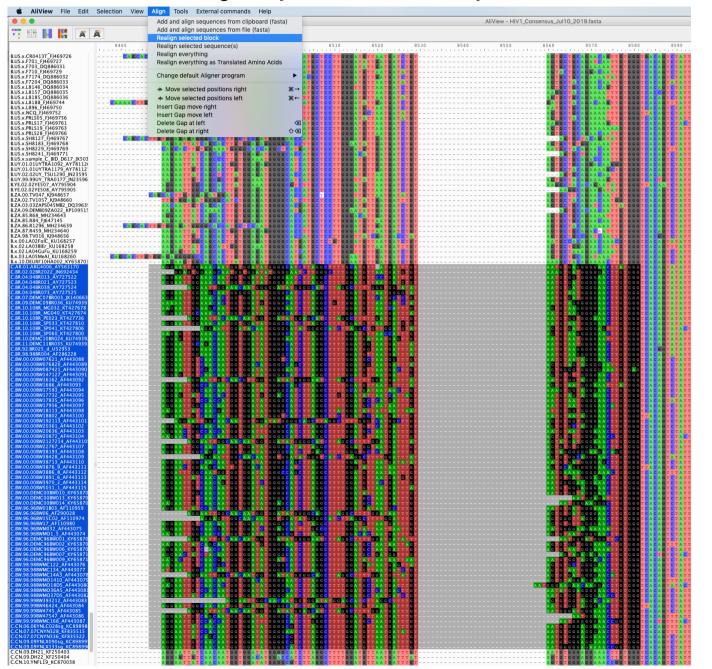
One codon being alternatively aligned in many sequences is a very common issue. Easiest to fix by sorting on one of these columns, to get all CCA-right or CCA-left together in one block, then adjusting that block.

48_01B.MY.07.07MYKT021_GQ175883 49_cpx.GM.02.N18380_HQ385477 49_cpx.GM.03.N26677_HQ385479 49_cpx.GM.97.N28353_HQ385478 50_A1D.GB.00.8179_JN417236 50_A1D.GB.03.33365_JN417239 50 A1D.GB.10.11762 JN417241 50_A1D.GB.10.12792_JN417240 51_01B.MN.12.12MNG12712_LC312714 51_01B.MY.11.11MYKL055_KJ485697 51 01B.SG.11.11SG HM021 JN029801 52_01B.MY.03.03MYKL018_1_DQ366664 52_01B.TH.00.00TH_R1741_AY945734 52 01B.TH.96.M043 DQ354113 52_018.MY.04.04MYKL016_1_DQ366663 53_018.MY.10.10MYKJ067_X390612 53_018.MY.10.10MYKJ079_X390611 53 01B.MY.11.11FIR164 JX390610 54_018.MY.07.07MYKL049_EU031915 54_018.MY.08.08MYKL044_X390977 54_018.MY.09.09MYSB023_X390976 55_01B.CN.08.08CYM047_JF340054 55_01B.CN.10.HNCS102056_JX574661 55_01B.CN.11.GDDG318_JX574662 56_cpx.FR.10.URF5_patient_A_JN882655 57_BC.CN.07.341_HM776939 57_BC.CN.09.09YNLX19sg_KC899008 57_BC.CN.09.1439_JX679207 57 BC.CN.09.YNFL37 KC870044 57_BC.CN.10.DH17_KF250400 58_01B.MY.09.09MYPR37_KC522031 58 01B.MY.10.10MYKJ036 KC522035 58_01B.MY.10.10MYPR87_KF425293 58_01B.MY.11.11MY1RJ704_KC522033 58_01B.MY.11.11MY1ZK731_KC522032 59_018.CN.07.GD070126_KF011494 59_018.CN.09.09LNA423_JX960635 59_018.CN.11.11CN.LNSY300876_KJ484434 59 01B.CN.12.12CN.YNKM200199 KJ484435 60 BC.GB.13.15228 1 80.3 MF109718 60_BC.IT.11.BAV499_KC899079 60_BC.IT.11.BAV636_KC899081 61 BC.CN.10JL100010 KC990124 62_BC.CN.10.YNFL13_KC870034 62 BC.CN.10.YNFL15_KC870035 63 02A.RU.10.10RU6637 JN230353 63_02A.RU.13.RU_8169_KJ197201 63_02A.RU.13.RU_8501_KJ197202 64_BC.CN.09.09YNLX047sg_KC898994 64_BC.CN.09.09YNLX219037sg_KC899009 64_BC.CN.09.YNFL31_KC870042 64_BC.CN.09.YNFL33_KC870043 64 BC.CN.10.YNFL10 1 KC870032 64_BC.CN.10.YNFL16_KC870036 64_BC.CN.10.YNFL22_KC870040 65 cpx.CN.10.YNFL01 KC870027 65 cpx.CN.10.YNFL02 KC870028 65_cpx.CN.11.ANHUI_HF104_KC183778 65_cpx.CN.x.JL15030_MH051841 66_BF1.PY.2002.JN251902 66_BF1.PY.2002,JN251903 67_01B.CN.11.ANHUI_HF115_KC183779 68_01B.CN.11.ANHUI_WH73_KC183782 69_01B,JP.03.03JP_5091K231_AB845344



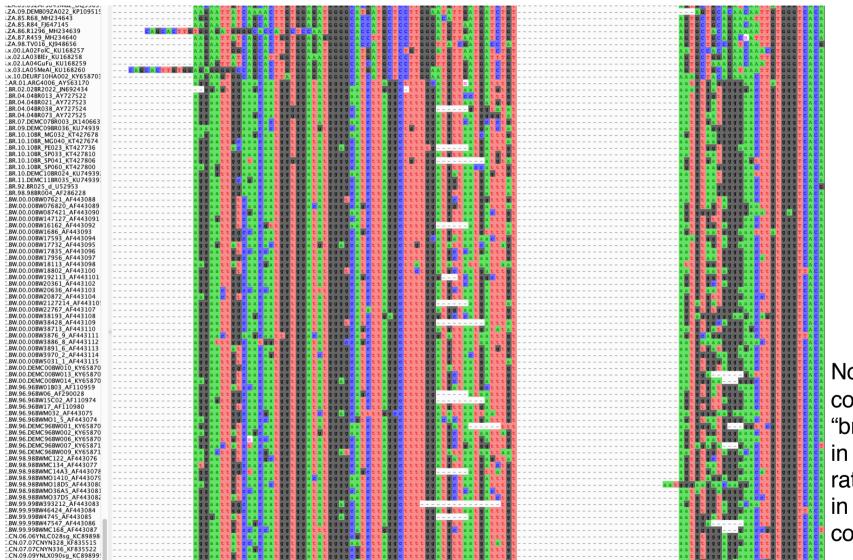


AliView has built-in alignment tools. Select a block and re-align only that block, is a very useful tool.





AliView has built-in alignment tools. Select a block and re-align only that block, is a very useful tool.

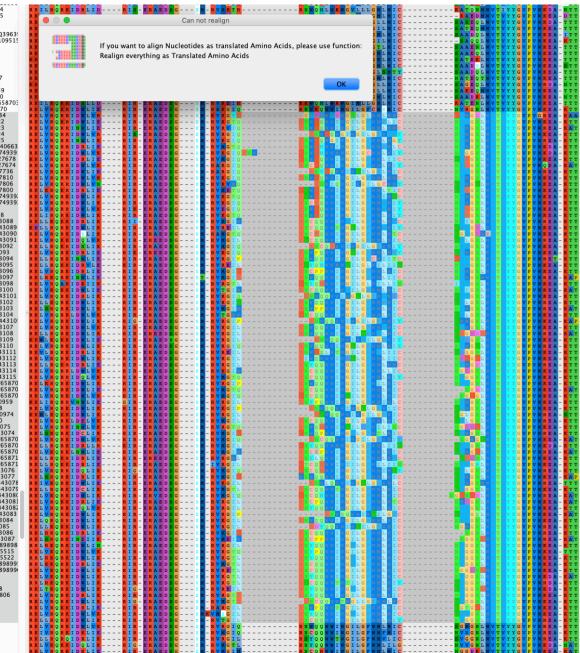


Note some codons now "broken"; gap in codon rather than in between codons.



AliView has built-in alignment tools. Select a block and re-align only that block, is a very useful tool.

B.YE.02.02YE507_AY795904 B.YE.02.02YE508_AY795905 B.ZA.00.TV047_KJ948657 B.ZA.02.TV1057_KJ948660 B.ZA.03.03ZAP5045MB2_DQ3963! B.ZA.09.DEMB09ZA022_KP109515 BZA.09.DEMB09ZA022. KP10 BZA.85.R68_MH234643 BZA.85.R84_FJ647145 BZA.85.R84_FJ647145 BZA.85.R459_MH234639 BZA.87.R459_MH234640 BZA.98.TV016_KJ948656 B.X.00_LA02FoIC_KU168257 B.X.02_LA03BIE_KU168258 B.x.02.LA03BitF_N0160238 B.x.02.LA04GuFu_KU168259 B.x.03.LA05MeAl_KU168260 B.x.10.DEURF10HA002_KY658703 C.AR.01.ARG4006_AY563170 C.BR.02.02BR20222_JN692434 C.BR.04.02BR20222_JN692434 C.BR.04.04BR013_AY727522 C.BR.04.04BR038_AY727523 C.BR.04.04BR038_AY727524 C.BR.04.04BR037_AY727525 L BR. 04-04 BR (021, A772725 C BR. 04-04 BR (021, A772755 C BR. 07 DEMC078 M037, X140663 C BR. 04-04 BR 053, X1747575 C BR. 07 DEMC078 M038, X140663 C BR. 010 BR, MC040, L7427674 C BR. 010 BR, MC040, L7427674 C BR. 010 BR, P503, K14277510 C BR. 010 BR, SP03, K14277510 C BR. 010 BR, SP03, K14277510 C BR. 010 BR, SP03, K14277810 C BR. 010 BR, SP03, K14277810 C BR. 010 BR, SP03, K14277810 C BR. 010 BR, SP03, K1427810 C BR. 010 BR, SP03, K1427810 C BR. 010 BR, SP03, K1427810 C BR. 010 BR, SP050, K1427800 C BR. 010 BR, SP050, K1427800 C BR. 010 BR, SP050, K1427800 C BR. 000 BR (0007621, A7443085 C BR. 000 008W07621, A7443085 C BR. 000 008W1752, A7443085 C BR. 000 008W17752, A7443095 C BR. 000 008W1755, A744305 C BR. 000 008W1755, A744305 C BR. 000 008W1752, A744305 C BR. 000 008W1752, A744305 C BR. 000 008W1752, A7443100 C BR. 000 008W1752, A7443102 C BR. 000 008W1752, A7443102 C BR. 000 008W1752, A7443105 C BR. 000 008W21721, A7443105 C BR. 000 008W3813, C BR. A7443105 C BR. 000 008W3813, C BR C.BW.00.DEMC00BW013_KY65870 C.BW.00.DEMC00BW014_KY65870 C.BW.96.96BW01B03_AF110959 C.BW.96.96BW150C_AF1290028 C.BW.96.96BW157_AF110974 C.BW.96.96BW17_AF110980__ C BW 36; SEBW17, AF110980 C BW 36; SEBW17, AF110980 C BW 36; SEBWM012, AF443075 C BW 36; SEBWM012, S, AF443075 C BW 36; DEMC96BW001, KY65870 C BW 36; DEMC96BW002, KY65871 C BW 36; DEMC96BW007, KY65871 C BW 36; DEMC96BW007, KY65871 C BW 38; SBBWMC134, AF443077 C BW 38; SBBWMC1340, AF443077 C BW 39; SBBWMC1340, AF443078 C BW 39; SBBWMC1340, AF443081 C BW 39; SBBWMC33521, AF443081 C BW 98:088WM03705_AF443083 C BW 99:998W193212_AF443083 C BW 99:998W1462_AF443083 C BW 99:998W4745_AF443085 C BW 99:998W475A_F443085 C BW 99:998W475A_F443085 C BW 99:998WW156A_F443087 C C N.06:06YNLC028*g_KC89899 C C N.07:07CNYN326_KF835515 C C N.09:07VNL2090_g_KC89899 C C N.09:07VNL2090_g_KC89899 C C N.09:07VNL2090_g_KC89899 C C N.09:07VL114_KF250404 C C N.09:07VL14_KF250404 C C N 09:07VL14_KF250404 C C N 09:07VL14_KF250404 C C N 09:07VL14_KF250404 C C N 09:07VL14_KF250404 C N 09:07VL14_KF25 C.CY.05.CY040_FJ388901 C.CY.05.CY069_FJ388913 C.CY.06.CY166_FJ388948 C.CY.06.CY176_FJ388952 C.CY.07.CY187_JF683740 C.CY.07.CY203_JF683755 C.CY.08.CY219_JF683768 C.CY.08.CY221_JF683770 C.CY.09.CY260_JF683803



Aliview does not currently support re-aligning a selected block as amino acids.

We try to work with software developers to ask for additions of such features, when reasonable.



"Perfecting" a large alignment can be a never-ending task. It is important to know when to call it "good enough".

Also consider methods such as gap-stripping which will often automatically remove regions of uncertain alignment such as the Env V1 and V2 hypervariable loop regions.

Feel free to write to me <u>btf@lanl.gov</u> or <u>seq-info@lanl.gov</u> for advice or help.

