

# Multiple Sequence Alignment

**Presenter: Brian Foley**

[btf@lanl.gov](mailto:btf@lanl.gov)

**HIV Databases**

*Theoretical Biology and Biophysics,*

*Los Alamos National Laboratory*



[www.hiv.lanl.gov](http://www.hiv.lanl.gov)  
[seq-info@lanl.gov](mailto:seq-info@lanl.gov)



# Presenter: Brian Foley

[btf@lanl.gov](mailto:btf@lanl.gov)

## HIV Databases



The screenshot shows the HIV Databases website. At the top left is a logo of a virus particle. The title "HIV DATABASES" is in red. Below it is a paragraph of text about the database's content and funding. Three buttons are visible: "SEQUENCE DATABASE ►", "IMMUNOLOGY DATABASE ►", and "OTHER VIRUSES ►". A "News" section contains a headline "CATNAP: two new features" and a paragraph of text dated February 2019. At the bottom, there is a contact link "seq-info@lanl.gov" and a footer with logos for the Department of Health & Human Services, Los Alamos National Laboratory, and the National Institutes of Health.

**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 [Editorial Board](#).

**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<sub>50/80</sub>) or 20 (ID<sub>50/80</sub>) 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](mailto:seq-info@lanl.gov)

Operated by Triad National Security, LLC for the U.S. Department of Energy's National Nuclear Security Administration  
© Copyright Triad National Security, LLC. All Rights Reserved | [Disclaimer/Privacy](#)

DEPT OF HEALTH & HUMAN SERVICES    **Los Alamos** NATIONAL LABORATORY EST. 1943    NATIONAL INSTITUTES OF HEALTH



# 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.*

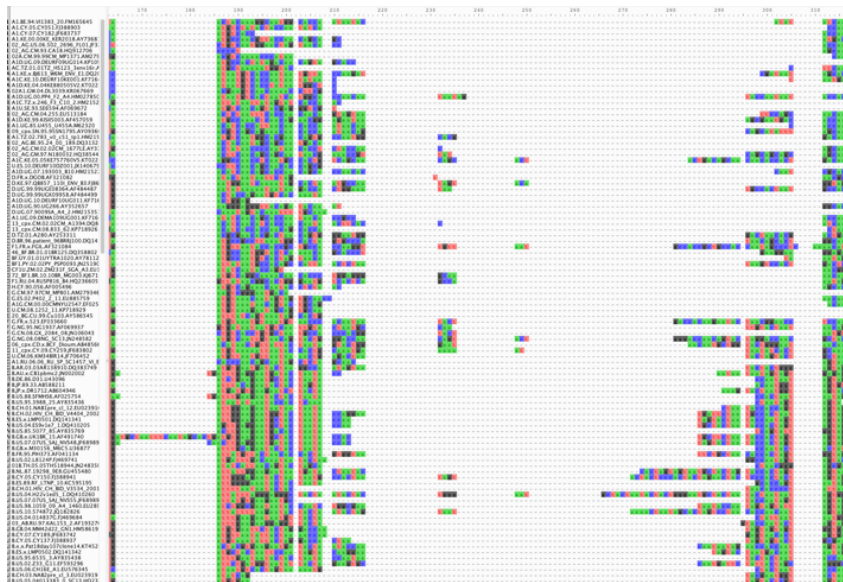
# Pairwise vs multiple alignment

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

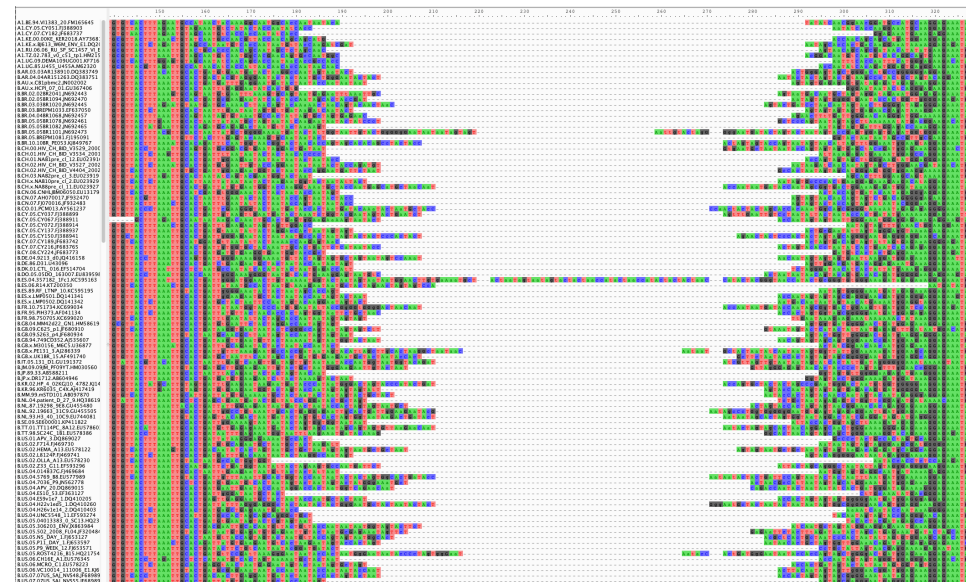
Multiple alignment is much more difficult.  
“Optimal” alignment is debatable.

BLAST  
Needle (EMBOSS)  
Water (EMBOSS)  
Align0

ClustalW ClustalOmega  
MUSCLE  
MAFFT  
GeneCutter (HIV specific)



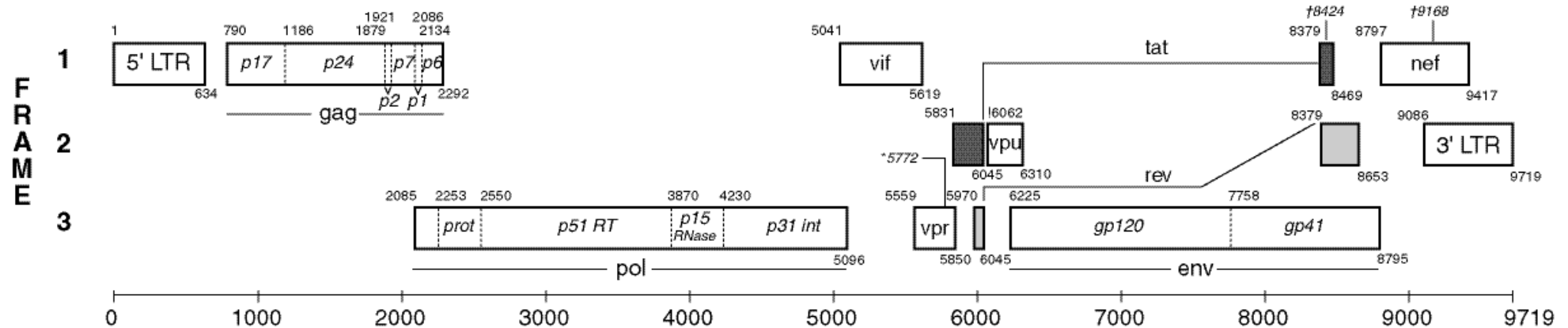
Typical alignment of HIV-1 env by MAFFT or similar tools. Cannot be translated to amino acid sequences.



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

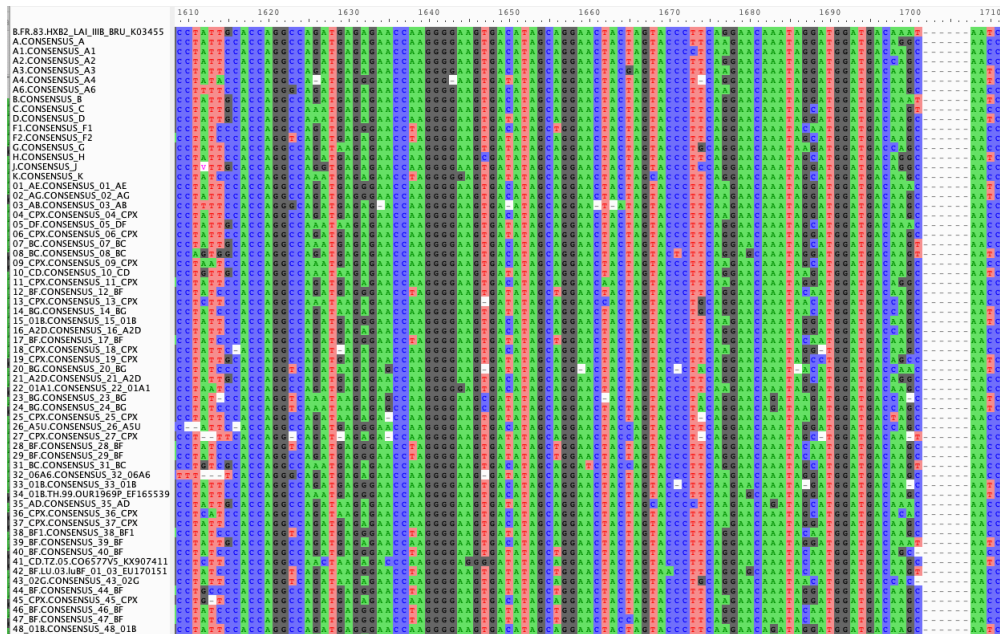


# HIV-1 Genome



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

Nine proteins/polypeptides 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.



Top example set to 25 codons,  
bottom zero codons.



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>



## HIV sequence database

[DATABASES](#)[SEARCH](#)[ALIGNMENTS](#)[TOOLS](#)[PUBLICATIONS](#)[GUIDES](#)

### 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

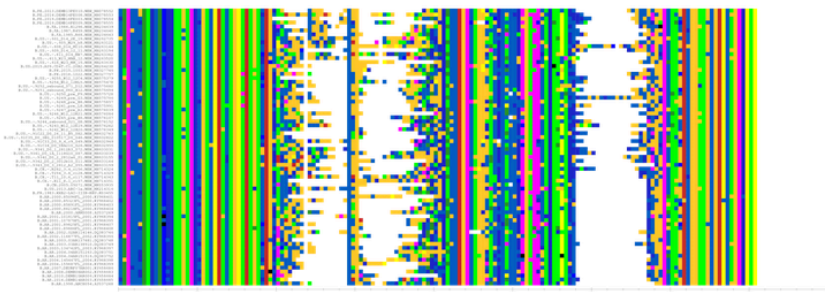
Scale bar size: 3

Margins: 0.5 inches

#### Results

[View large](#)

Download [\[EPS\]](#) [\[PDF\]](#) [\[PNG\]](#) [\[Legend\]](#) [\[All\]](#)

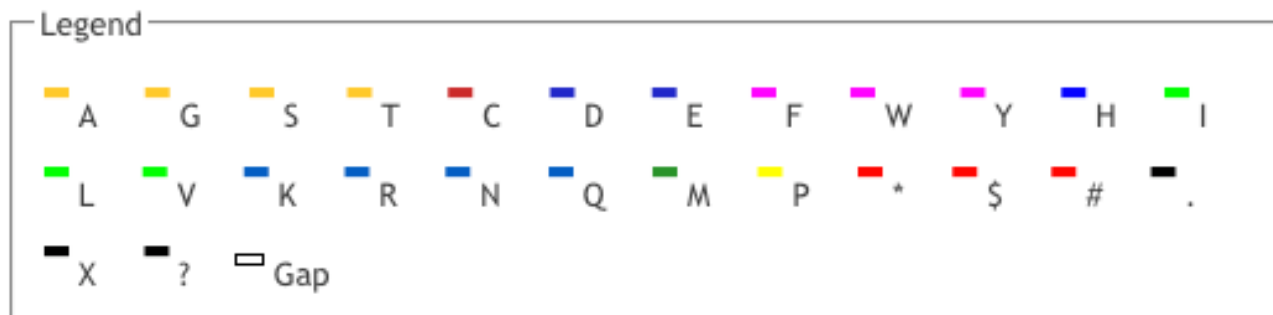
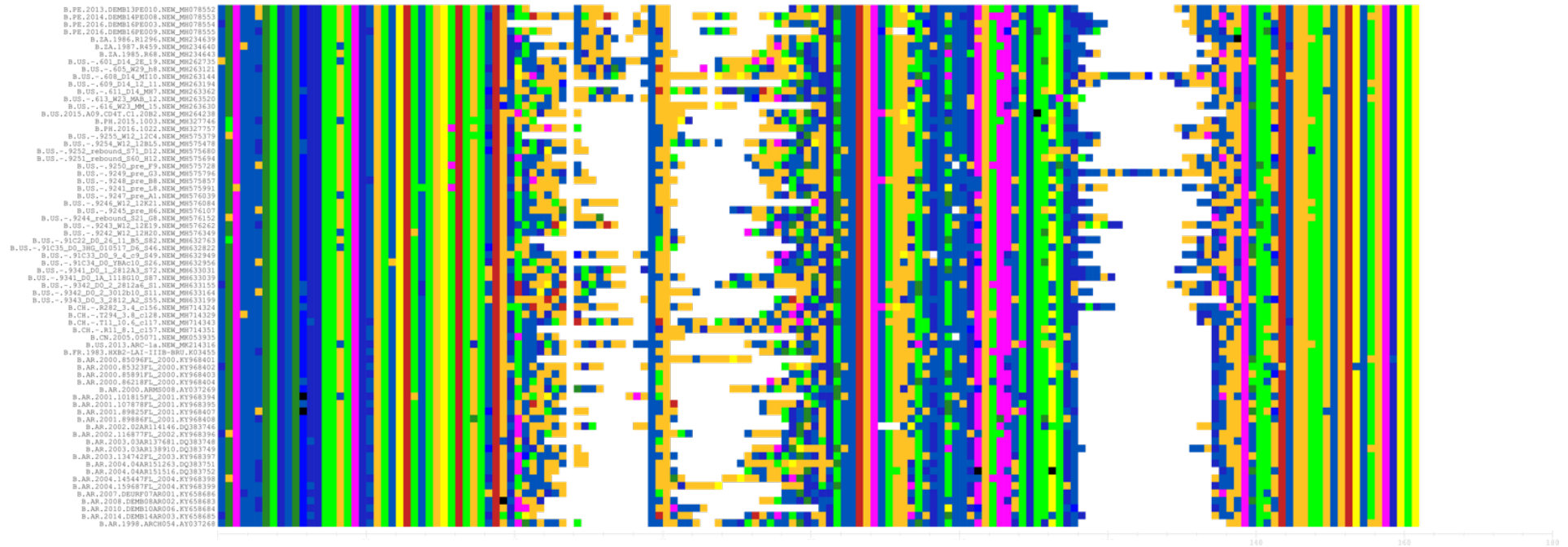


#### Legend

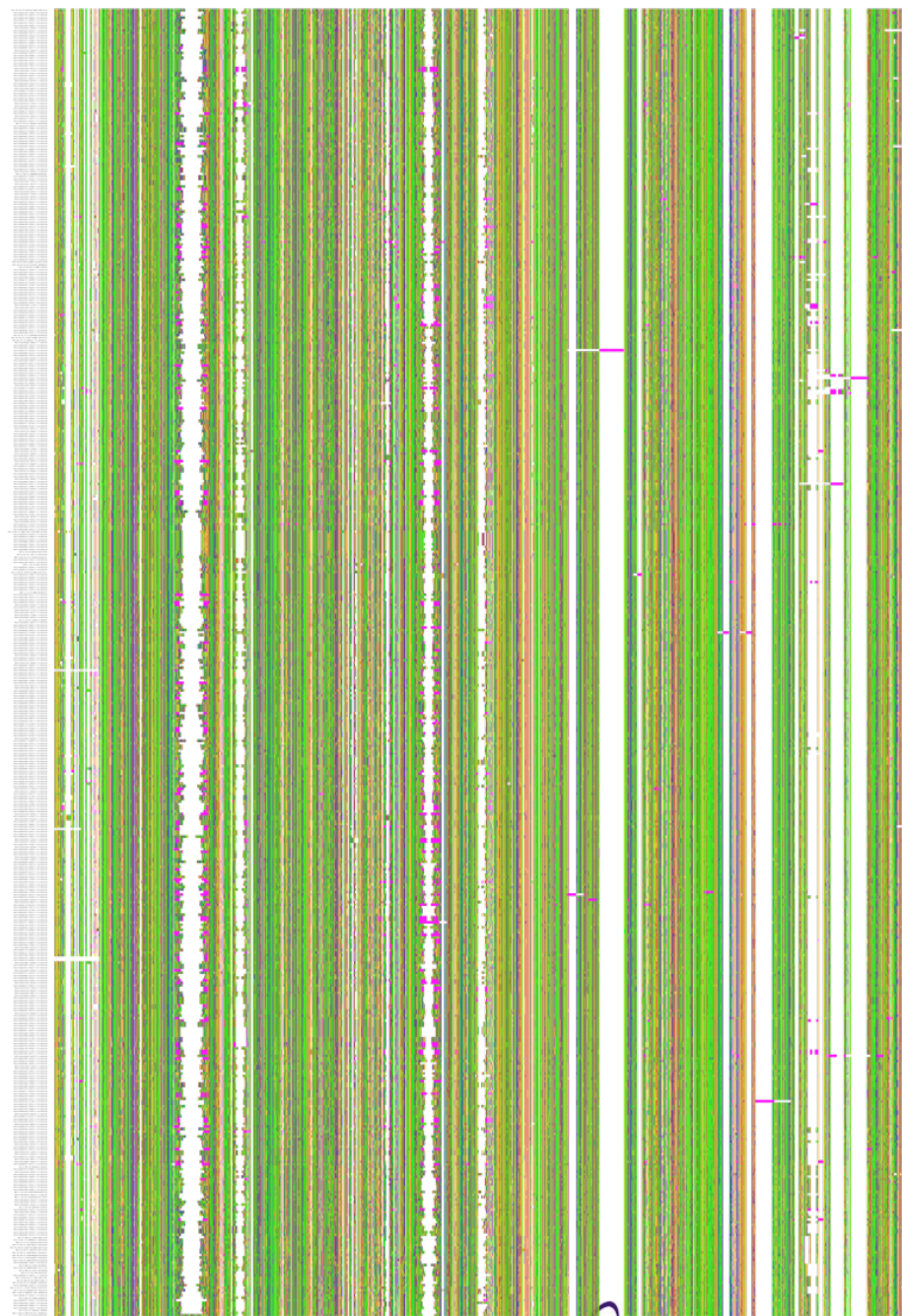
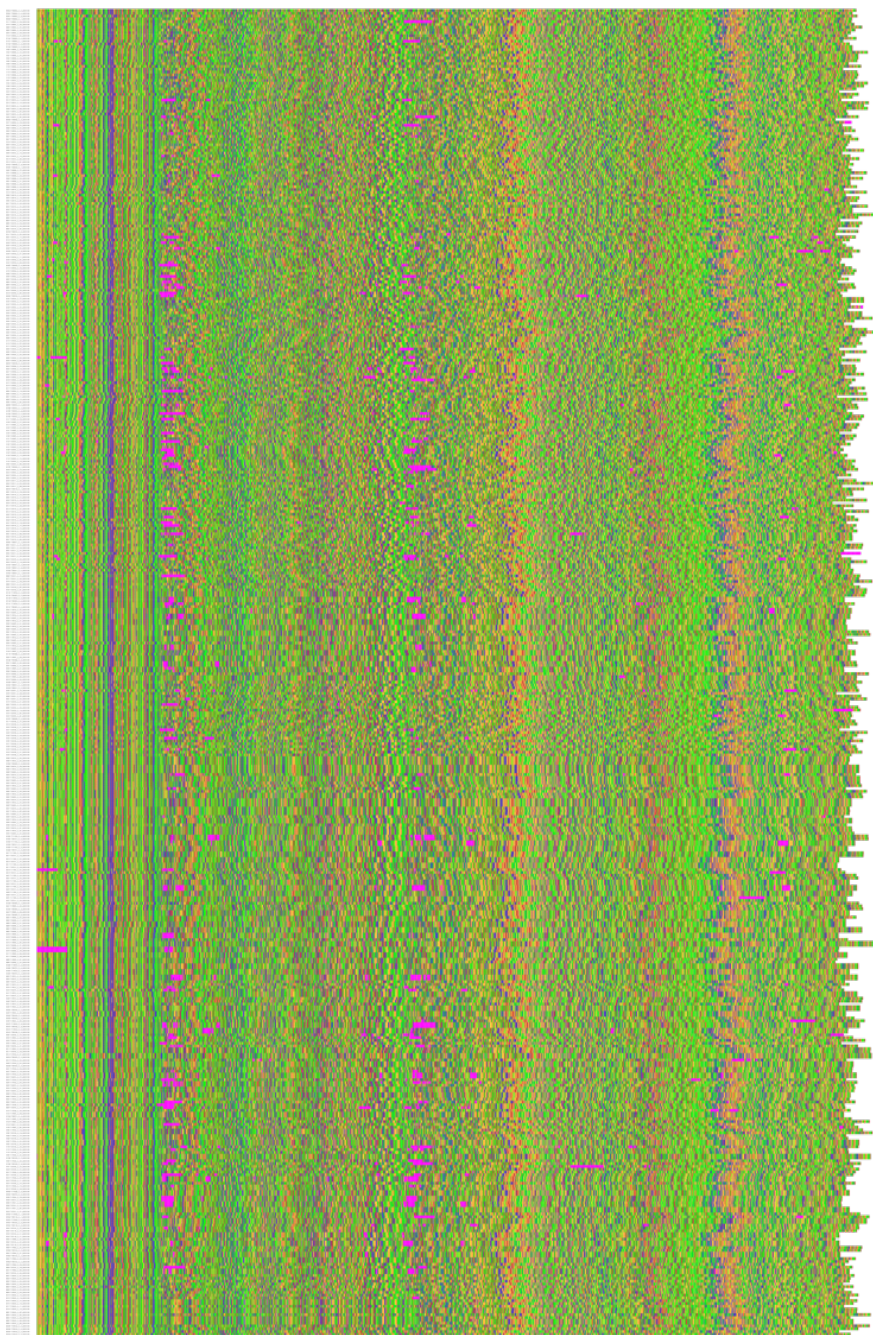
A	G	S	T	C	D	E	F	W	Y	H	I
L	V	K	R	N	Q	M	P	*	\$	#	.
X	?	Gap									

Questions or comments? Contact us at [seq-info@lanl.gov](mailto:seq-info@lanl.gov).

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









		10	20	30	40	50	60	70	80	90	100
B.PE.2013.DEMB13PE010.NEW_MH078552	NMWNNNSMVEQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	LSNSTT---	-----SN	GSESLRGEMK	NCSFNVITPDV	SDRKKREYAL		
B.PE.2014.DEMB14PE008.NEW_MH078553	NMWKNNMVDQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	VICITSSNSI	-----TTSPTITSP	NITTSPTEMK	NCSFNVITPDV	RDKTKEEYAF		
B.PE.2016.DEMB16PE003.NEW_MH078554	DMWKNNMVEQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTN	-----	-----STRGST	IGQNMGRGICQ	NCSFNVITTVV	HDVKKREYAL		
B.PE.2016.DEMB16PE009.NEW_MH078555	NMWKNNMVDQ	MHEDIISLWD	ESLKPCKVLT	PLCVTLNCTN	VITGTNNITD	-----GNDGT	RINVTDEIK	NCSFHT--HI	GDQKREYAI		
B.ZA.1986.R1296.NEW_MH234639	NMWKNNMVEQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	WRNTTNTNT	-----NKT	NTNSSIEGEMK	NCSFNVITTSI	RDVKKESAL		
B.ZA.1987.R459.NEW_MH234640	NMWNNMVEQ	MHEDIINLWD	QSLKPCVKLT	PLCVTLNCTD	DLKNTTNTNA	-----NASS	WGNIEKGEIK	NCSFNVITNR	RGMKREYAL		
B.ZA.1985.R68.NEW_MH234643	NMWKNNMVEQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	NLRNDTTTNN	-----SGIG	GIRMEKGEIK	NCSFNVITNI	RDKFKVYAL		
B.US.--601.D14.2E.19.NEW_MH262735	TMWKNNMVEQ	MHEDIINLWD	QSLKPCVKLT	PLCVTLNCTD	LNATSNAT	-----NTSSSYV	GPPIEBEIK	NCSFNVITSI	KNKVKREYAL		
B.US.--605.W29.H8.NEW_MH263121	NMWKNNMVDQ	MHEDIISLWD	QSLKPCVKLT	PLCVSLDCTD	PKNDTANGNE	-----FNVSIVRD	TANGNEGEMK	NCSFNVSSIV	RDVKVREYAL		
B.US.--608.D14.M110.NEW_MH263144	NMWKNNMVEQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	VIFVTATN	-----SAPSTPTST	TSTPTKEEVN	NCSFNVITSI	VDKVKREYAL		
B.US.--609.D14.12.11.NEW_MH263194	NMWKNNMVEQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	AMITSSB	-----QAS	TIRMEKEEIQ	NCSFNVITFG	RDKIOREYAL		
B.US.--611.D14.MH7.NEW_MH263362	NMWKNNMVDQ	MEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	ATRIIVNGTV	-----GTEEMKNC	IATNTEHEMK	NCSFNVITSDI	RDKIOREYAL		
B.US.--613.W23.MAB.12.NEW_MH263520	NMWKNNMVEQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	LNATSNAT	-----TTHSS	WERMCKEIK	NCSFNVITNI	GDVKREYAL		
B.US.--616.W23.MM.15.NEW_MH263630	NMWKNNMVDQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	VTLANNTATT	-----STTTPSS	-----SGEMK	NCSFNVITAP	RDKIOREYAL		
B.US.2015.A09.CD4T.C1.20B2.NEW_MH264238	NMWKNNMVEQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	YVGDTTTNS	-----TN	TNISMGEEMK	NCSFNVITPMI	RDKIOREYAL		
B.PH.2015.1003.NEW_MH327746	NVWKNNMVEQ	MHEDIINLWD	QSLKPCVKLT	PLCVTLNCTN	ATINSTTPE	-----T	ELRTLKKEEVN	NCSFNVITPSI	KNKMKREYAL		
B.PH.2016.1022.NEW_MH327757	NIWKNNMVEQ	MHEDIINLWD	QSLKPCVKLT	PFVCLDCTD	TGTGYSTIN	-----TT	ATIMQKGEIK	NCSFNVITSI	QDRKREYAL		
B.US.--9255.W12.12C4.NEW_MH575379	NMWKNNMVEQ	MHEDIISLWD	QSLKPCVKLT	PLCVLHCTD	VNRTSN	-----	-----EMTGELK	NCSFNVITKI	TNKVKREYAL		
B.US.--9254.W12.12BL5.NEW_MH575478	NMWKNNMVEQ	MEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	VEIDKTSANK	-----TINVTN	-----TNS	WERMDDPEIK	RDVKVREYAL		
B.US.--9252_rebound.S71.D12.NEW_MH575680	NMWKNNMVDQ	MEDIISLWD	QSLKPCVKLT	PLCVTLNCTN	LIDIPNNTAN	-----T	-----SIS	GSLEGGEMK	NCSFNVITSL	RDRKKREYAL	
B.US.--9251_rebound.S60.H12.NEW_MH575694	NMWKNNMVEQ	MEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	YONNANTTV	-----NAST	TNNNGKEIK	NCTFKVITGGI	RDKTOLEYAL		
B.US.--9250_pre.F9.NEW_MH575728	NMWKNNMVDQ	MEDIISLWD	QSLKPCVKLT	PFVCLDCTD	LQNSTSTNN	-----TNT	IDNNMKEIK	NCSFNVITTS	KEK--KGAF		
B.US.--9249_pre.G3.NEW_MH575796	NMWKNNMVEQ	MEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	VIFVTATN	-----	-----GN	GANITADMK	NCSFNVITSI	RDVKVREYAL	
B.US.--9248_pre.B8.NEW_MH575857	NMWKNNMVEQ	MEDIISLWD	ESLKPCKVLT	PLCVTLNCTN	ANFTTSTNT	-----NN	-----SEST	GEIKTTEIK	NCSFNVITFG	RDKKNNAL	
B.US.--9241_pre.L8.NEW_MH575991	NMCKNNMVEQ	MHEDIISLWD	QSLKPCVKLT	PFVCLNCTD	ADTGTN	-----	-----SS	STLEKEEIK	NCSFNVITSS	SDKPOEAL	
B.US.--9247_pre.A1.NEW_MH576039	NMWKNNMVEQ	MHEDVINLWD	QSLKPCVKLT	PLCVTLNCTN	LVNTN	-----SSS	KEGMAEEMK	NCSFNVITTEI	RNMKREYAL		
B.US.--9246.W12.12K21.NEW_MH576084	NMWKNNMVDQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	IRNSTVTST	-----TTSNTTT	-----WGMT	NCTFNITTSI	KDKMKREAL		
B.US.--9245_pre.H6.NEW_MH576107	NMWKNNMVEQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	VNATATMGND	-----TNDSSK	GKLMIEGEMK	NCSFNVITSI	GEKK--TENAL		
B.US.--9244_rebound.S21.G8.NEW_MH576152	NMWKNNMVDQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTE	LKNNTNSTT	-----MNT	-----SSI	EEREMTREM	NCSFNVITVI	RDVKQRYAL	
B.US.--9243.W12.12E19.NEW_MH576262	NMWKNNMVDQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	NLNLNCPHNN	-----TCS	-----NNTN	YMTKEGEIK	NCSFNVITTEI	RDVTKREHAL	
B.US.--9242.W12.12H20.NEW_MH576349	NMWKNNMVDQ	MHEDIINLWD	QSLKPCAKLT	PLCVTLNCTD	LRNDTVGNQT	-----N	-----YLN	TNTIQGREMT	NCSFNVITTEI	RDVKVREYAL	
B.US.--91C22.D0.26.11.B5.S82.NEW_MH632763	NIWKNNMVEQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTN	LVNTES	-----	-----DQMKEGGQ	NCSFNVITAGI	KDRMKREYAL		
B.US.--91C35.D0.3HG.010517.D6.S46.NEW_MH632822	NMWKNNMVEQ	MHEDVINLWD	QSLKPCVKLT	PLCVTLNCTN	TTKDNITI	-----	-----EGNEIK	NCSFNVITDL	RDVKQRYAL		
B.US.--91C33.D0.9.4.C9.S49.NEW_MH632949	NMWKNNMVEQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	LVNNTTHGK	-----SGS	-----SKNSSS	EMWVREEMK	NCSFNVITPSI	G--KRRREYAL	
B.US.--91C34.D0.YBac10.S26.NEW_MH632956	NMWKNNMVEQ	MHEDVINLWD	QSLKPCVKLT	PLCVTLNCTD	VIFVTATN	-----	-----ST	WERMKEGEIK	NCSFNVITSL	KAKVREYAL	
B.US.--9341.D0.1.2812A.S72.NEW_MH633031	NMWKNNMVEQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTN	VGTGVTFRIN	-----	-----DNAT	INMKTAEEMK	NCSFNVITSI	RDVKVREYAL	
B.US.--9341.D0.1A.1118G10.S87.NEW_MH633039	NMWKNNMVEQ	MHEDVINLWD	QSLKPCVKLT	PLCVTLNCTD	LKNAKTGVDV	-----	-----DNIGN	SSTEAMAEK	NCSFNVITTSI	KDKKREYAI	
B.US.--9342.D0.2.2812A.S11.NEW_MH633155	DMWKNNMVEQ	MEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	CDSSPNCIDT	-----KSTNAT	-----NSIKNSTD	ATTSITREIK	NCSFNVITTM	RDVQREYAL	
B.US.--9342.D0.2.3012B10.S11.NEW_MH633164	NMWKNNMVEQ	MEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	CDNPCTALR	-----NSTATN	-----	-----STEGEMK	NCSFNVITSI	RDVKVREYAL	
B.US.--9343.D0.3.2812.A2.S55.NEW_MH633199	NMWKNNMVDQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTN	ATTSKANON	-----	-----GNST	INCTNVEEMK	NCSFNVITRI	KNKVKREYAL	
B.CH.--R282.3.4.C156.NEW_MH714324	NMWKNNMVEQ	MEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	LCNTNCTNT	-----STNTSTN	-----ST	NSTWEDKGLR	NCSFYATPSI	RDVKKREYAI	
B.CH.--T294.3.8.C128.NEW_MH714329	NMWSNNMVEQ	MHEDVINLWD	QSLKPCVKLT	PLCVTLNCTD	LNSTDTTNT	-----T	-----IN	SNWEGKEVK	NCSFNVITSI	RDVKQRYAL	
B.CH.--T11.10.6.C117.NEW_MH714343	NMWKNNMVEQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	LKNDLTCAN	-----	-----SVGNDTNTN	NSSWIGEDMR	NCSFNVITTHI	VDKQOYAL	
B.CH.--R11.8.1.C157.NEW_MH714351	NMWKNNMVEQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	LKNDTSTNS	-----DTASTN	-----TNS	WGEMDKGEIK	NCSFNVITNI	LDRKREYAL	
B.CN.2005.05071.NEW_MK053935	NMWKNNMVEQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTN	YNTNSSTSS	-----	-----TEGEIK	NCSFNVITSI	NKV--RDYAL		
B.US.2013.ARC-1a.NEW_MK214316	NMWKNNMVEQ	MHEDIINLWD	ESLKPCKVLT	PLCVLHCTD	LRNIEQTSNE	-----TQTSNRN	-----QTSNRN	TEPOMGEIK	NCSFNVITVD	KDKKKREHAL	
B.FR.1983.HXB2-LAI-11TB-BRU.R03455	NMWKNNMVEQ	MHEDIISLWD	QSLKPCVKLT	PLCVSLCTD	LKNDTHNS	-----SSG	RMIMEKGEIK	NCSFNVITSI	RGKVREYAL		
B.AR.2000.85096FL.2000.KY968402	NMWKNNMVEQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	VTGVTFRIN	-----	-----TASPPS	-----	NCSFNVITSI	GKVREYAL	
B.AR.2000.85323FL.2000.KY968403	DMWKNNMVEQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTA	LKNTSSSDS	-----	-----SNG	TVSPAREEMT	NCSFNVITTHI	RNKIOREYAL	
B.AR.2000.85891FL.2000.KY968404	NMWKNNMVEQ	MHEDIISLWD	ESLKPCKVLT	PLCVTLNCTN	LNATNAT	-----	-----	-----SORIEGEMK	NCSFNVITSI	GNMKREYAL	
B.AR.2000.86218FL.2000.KY968404	NMWKNNMVEQ	MEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	ANVTATNI	-----	-----SW	GEALGKEIK	NCSFNVITTM	RDVKKAYAN	
B.AR.2000.ARM5008.AY037269	NMWKNNMVEQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	LNATEINMNN	-----ANA	-----TGTTNTN	ITTKDQNMK	NCSFNVITEG	RNRKKOHAI	
B.AR.2001.101815FL.2001.KY968394	NMWKNNMVEQ	MEDIISLWD	ESLKPCKVLT	PLCVTLNCTD	VTISPT	-----	-----AW	SNVITIREMT	NCSFNVITSI	RDVKTFEAL	
B.AR.2001.107878FL.2001.KY968395	NMWKNNMVEQ	MEDIISLWD	QSLKPCVKLT	PLCVTLNCTN	ANCTNANTC	-----	-----NA	TSSEALRDV	NCSFNVITNI	RDKMKREYAL	
B.AR.2001.89825FL.2001.KY968407	NMWKNNMVEQ	MEDIISLWD	ESLKPCKVLT	PLCVTLNCTN	ANGTVNNSN	-----RTT	-----NA	SLEKMEGEIK	NCSFNVITGR		
B.AR.2001.89886FL.2001.KY968408	NMWKNNMVEQ	MEDIISLWD	QSLKPCVKLT	PLCVMLNCTD	YTNNSNTSTN	-----	-----	-----LEQMKEIK	NCSFNVITNI		
B.AR.2002.02AR114146.DQ387346	NMWKNNMVEQ	MEDIISLWD	QSLKPCVKLT	PLCVTLNCTN	WNSTNTADTI	-----TTN	-----TTN	ATMDGEIK	NCSF--QPR		
B.AR.2002.116877FL.2002.KY968396	DAWKNNMVEQ	MEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	YNGVNTANS	-----SIN	-----NSSG	GREMEKGEIK	NCSFNVITNI		
B.AR.2003.03AR137681.DQ383748	NMWKNNMVEQ	MEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	TDQNT	-----	-----	-----TRGIEKLEIK	NCSFNVITSI		
B.AR.2003.03AR138910.DQ383749	NMWKNNMVEQ	MEDIISLWD	ESLKPCKVLT	PLCVTLNCTD	VNDTRANST	-----	-----TGSS	WDMFGGEMT	NCSFNVITSI		
B.AR.2003.134742FL.2003.KY968397	NMWKNNMVEQ	MEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	WVITATN	-----	-----SNEV	DMVNTVITSI	NCSFNVITSI		
B.AR.2004.04AR151263.DQ383751	NMWKNNMVEQ	MEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	YVGNITAGN	-----TT	-----NTST	VEPTASTGOK	NCSFNVITDL		
B.AR.2004.04AR151516.DQ383752	NMWKNNMVEQ	MEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	WKSNAITSTR	-----	-----INNS	SLGTMGQIK	NCSFNVITSI		
B.AR.2004.145447FL.2004.KY968398	NMWKNNMVEQ	MEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	LKNNTNSTM	-----	-----TTPSSRT	GEFMEGEMK	NCSFNVITDI		
B.AR.2004.159687FL.2004.KY968399	NMWKNNMVEQ	MEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	VTNKATNTT	-----SDATSTP	-----SSSEVQGM	NCSFNVITSI	NCSFNVITSI		
B.AR.2007.DEURF07AR001.KY658686	NMWKNNMVEQ	MEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	ARNTENTIT	-----	-----	-----RQMDKEIK	NCSFNVITML		
B.AR.2008.DEMB08AR002.KY658683	NMWKNNMVEQ	MEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	WTADNTN	-----	-----ST	IRREMERIK	NCSFNVITTEV		
B.AR.2010.DEMB10AR006.KY658684	NMWKNNMVEQ	MEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	YYSNATFNN	-----NT	-----	-----KEIK	NCSFNVITSL		
B.AR.2014.DEMB14AR003.KY658685	DMWKNNMVEQ	MEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	VPTNCTNG	-----	-----NCTN	WKIEGEGELK	NCSFNVITTR		
B.AR.1998.ARCH054.AY037268	NMWKNNMVEQ	MEDIISLWD	QSLKPCVKLT	PLCVTLNCTN	WRNTNTN	-----TSNNA	-----TNQDITRI	NNTMENGEMK	NCSFNVITTM		

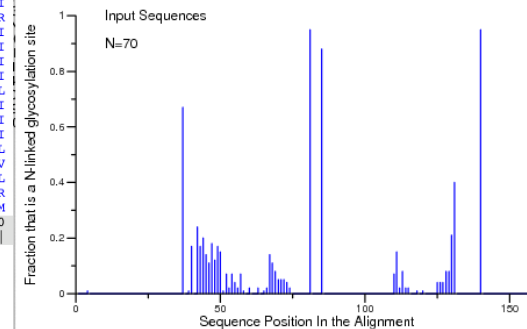
B.PE.2013.DEMB13PE010.NEW_MH078552	LYKLDIVQID	N-----	-----SMTNS	NTSVYRLSCN	TSVITQCPK	ISFEP
B.PE.2014.DEMB14PE008.NEW_MH078553	FYSQDVVP	EKE-----	-----SMT	VTYRLASCN	TSVITQCPK	TSFEP
B.PE.2016.DEMB16PE003.NEW_MH078554	FYKLDIVP	DDNN-----	-----SSMAN	YTSYRLSCN	TSVITQCPK	VSFEP
B.PE.2016.DEMB16PE009.NEW_MH078555	FHTLDMVP	IK NDI-----	-----SKM	DTLYRLMHCN	TSVITQCPK	VSFEP
B.ZA.1986.R1296.NEW_MH234639	FYKLDIVP	DI-----	-----DSN	NNKYRLSCN	TSVITQCPK	VSFEP
B.ZA.1987.R459.NEW_MH234640	LYKLDVVP	NR-----	-----T	NTSYRLSCN	TSVITQCPK	ISFEP
B.ZA.1985.R68.NEW_MH234643	FYKLDVVP	NE-----	-----NTS	YTYRLSCN	TSVITQCPK	VSFEP
B.US.--601.D14.2E.19.NEW_MH262735	FYKLDVMP	D-----	-----D	TTSYRLSCN	TSVITQCPK	VSFEP
B.US.--605.W29.H8.NEW_MH263121	FYSLDIVP	NE-----	-----GNT	NTSYRLSCN	TSVITQCPK	VSFEP
B.US.--608.D14.M110.NEW_MH263144	LYKTDLPV	IN DSSSLNNGS	H-SDSSSYN	TSYRLSCS	TSVITQCPK	VSFEP
B.US.--609.D14.12.11.NEW_MH263194	FYKLDVVP	SETK-----	-----DTSGN	YTYRLSCN	TSVITQCPK	VSFEP
B.US.--611.D14.MH7.NEW_MH263362	FYELDITP	N-----	-----S	TEYRLSCN	TSVITQCPK	VSFEP
B.US.--613.W23.MAB.12.NEW_MH263520	FYRFDIVQ	PI-----	-----N	TKYRLSCN	TSVITQCPK	VSFEP
B.US.--616.W23.MM.15.NEW_MH263630	FYKLDIVK	G-----	-----SNM	TSYRLSCN	TSVITQCPK	VSFEP
B.US.2015.A09.CD4T.C1.20B2.NEW_MH264238	FYSLDIVP	N-----	-----	TSYRLSCN	TSVITQCPK	VSFEP
B.PH.2015.1003.NEW_MH327746	LYKLDLVA	N-----	-----TD	NTSYRLSCN	TSVITQCPK	VSFEP
B.PH.2016.1022.NEW_MH327757	FYKLDIEP	D-----	-----SNSSS	NTSYRLSCN	TSVITQCPK	ISFEP
B.US.--9255.W12.12C4.NEW_MH575379	FYKLDVVP	NKDN-----	-----DTSFN	NTSYRLSCN	TSVITQCPK	VSFEP
B.US.--9254.W12.12BL5.NEW_MH575478	FYKLDVQIK	DKKA-----	-----S	NTYRLSCN	TSVITQCPK	VSFEP
B.US.--9252_rebound.S71.D12.NEW_MH575680	FYKLDVATG	E-----	-----N	TNSFRLSCN	TSVITQCPK	VSFEP

<https://www.hiv.lanl.gov/content/sequence/GLYCOSITE/glycosite.html>

B.US.--9247_pre.A1.NEW_MH576039	FYKLDVVP	D-----	-----D	NTSYRLSCN	TSVITQCPK	VSFEP
B.US.--9246.W12.12K21.NEW_MH576084	FYKLDVVP	EKE-----	-----NNS	NTSYRLSCN	TSVITQCPK	VSFEP
B.US.--9245_pre.H6.NEW_MH576107	FYKLDVVP	R-----	-----D	NTSYRLSCN	TSVITQCPK	VSFEP
B.US.--9244_rebound.S21.G8.NEW_MH576152	FYKLDVQIK	DDNTS-----	-----Y	NTSYRLSCN	TSVITQCPK	VSFEP

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.

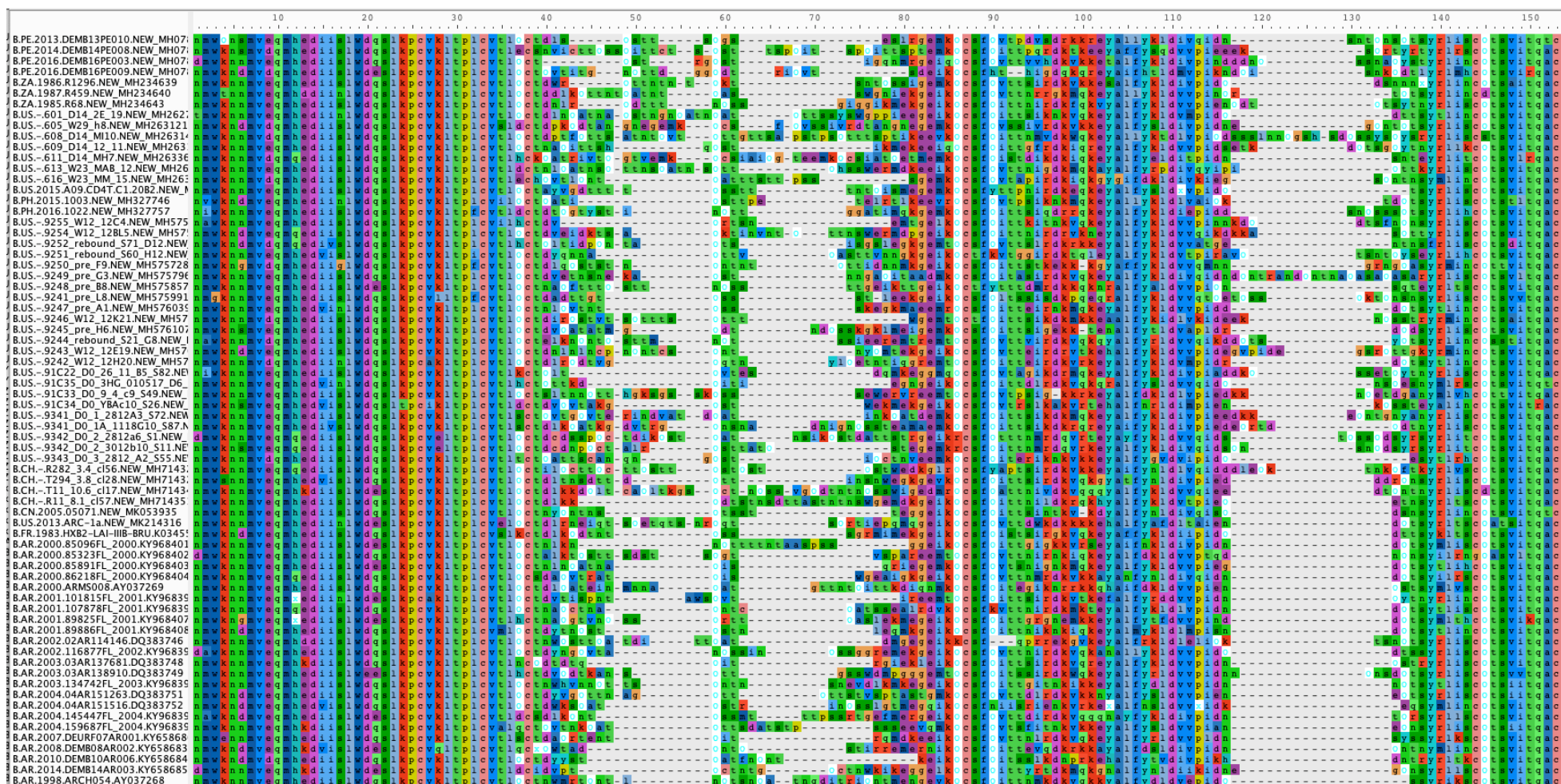


Sequence position(s) of High N-glycosylation site (>60%)

Pos	Top seq	Num of N-glycosylation	Fraction
37	N	47	0.671
81	N	67	0.957
85	N	62	0.886
140	N	67	0.957



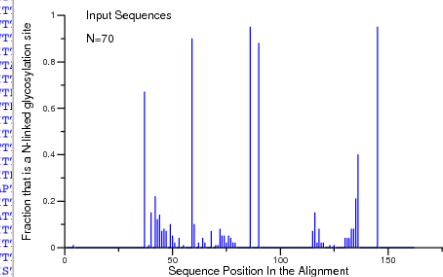
# Glycosite can output FASTA file with the N of N-x-(ST) recoded as "O"





	10	20	30	40	50	60	70	80	90	100
B.PE.2013.DEMB13PE010.NEW_MH078552	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	LS-----NS	TT-----SNG	S-----	-----ESL	RGEKMKCSFN	VTPDVSRRKK
B.PE.2014.DEMB14PE008.NEW_MH078553	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	VICTTSSSI	TTCT--S-NS	T---TSPNIT	---SPNITTS	PTDMKCSFN	ITTPDRDKTK
B.PE.2016.DEMB16PE003.NEW_MH078554	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	-----NS	T-----RGS	T-----	---IQGM	KELIGKCSFN	VTTVHLDKTK
B.PE.2016.DEMB16PE009.NEW_MH078555	NMWNNSMVQ	MHEDIISLWD	ESLKPVCVKLT	PLCVTLNCTN	VTITG---NN	TTD---GGND	T---RINVT	-----	SDRIKCSFN	T--HIGDQKQ
B.2A.1986.R1296.NEW_MH234639	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	WR-----NT	TNTN-T--NK	T-----	---SNTNS	IGEMKCSFN	VTTSIKDKVK
B.2A.1987.R459.NEW_MH234640	NMWTNNMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	DLKNTTNTNA	TNT--NA	S-----	---SWGNE	KGEIKCSFN	VTTNRRGKMQ
B.2A.1985.R68.NEW_MH234643	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	NLR-----ND	TTT-----NS	S-----	---GIGIKME	KGEIKCSFN	ITTNIRDKFO
B.US--.601_D14_2E_19.NEW_MH262735	TMWKNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	LNNATNA-NS	TGNNATNA	T-----NTT	---SSYSWGPIE	EGEIKCSFN	ITTSIKNNVO
B.US--.605_W29_h8.NEW_MH263121	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVSLDCTD	PKNDTAN-GN	EGEMK--NC	S---P-NVS	SIIVRDGTANG	SSIVRDGTANG	VSVIRDRKVK
B.US--.608_D14_M110.NEW_MH263144	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	TFGTFTS--NT	NWVT--NT	TGTTTS--PST	PSMTTFTT	ITTSIRDRKVK	ITTSIRDRKVK
B.US--.609_D14_12_11.NEW_MH263194	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	ANITTSI--	-----ONS	T-----	---IKME	KELIGKCSFN	ITTFGRDKIO
B.US--.611_D14_MH7.NEW_MH263362	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	ATRIVTN-GT	VEMK-----NC	SIANG--TEE	MKNCSIATNE	TMEMKCSFN	ISTDIKDKIO
B.US--.613_W23_MAB_12.NEW_MH263520	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	LNATNSN-TT	NSNATH-SNT	T-----	---NHSWERM	KELIGKCSFN	VTTNIGDKMQ
B.US--.616_W23_MM_15.NEW_MH263630	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	VTLNNT--	-----NA	TTTSTT--PSS	-----	SGEMKCSFN	VTAPIKDKIO
B.US.2015.A09.CD4T.C1.20B2.NEW_MH264238	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTA	YVGDFTT-T-	-----NS	STT-----	---TNTNISM	EGBMKCSFN	ITPINIRDKVO
B.PH.2015.1003.NEW_MH327746	VYWKNDMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTN	ATI-----	-----NS	TTPE-----	---TELRLT	KELVRCNSFN	VTPSIRKNNVO
B.PH.2016.1022.NEW_MH327757	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	TFGTFTS-T-	-----NS	-----	---GGATLMO	KELVRCNSFN	VTPSIRKNNVO
B.US--.9255_W12_12C4.NEW_MH575379	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	V-----	-----NS	-----	---EM	TGELKCSFN	ITTTIKNNVO
B.US--.9254_W12_12B15.NEW_MH575478	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	VEIDKTS-A-	-----NK	TNTVNT--N-	---TTSNWERM	PGEIKCSFN	ITTNIRDRVK
B.US--.9252_rebound_S71_D12.NEW_MH575680	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	LTIDPNN-TA	-----NT	S-----	---ISGSLEG	KGEMTCSFN	VTTSLRDKRK
B.US--.9251_rebound_S60_H12.NEW_MH575694	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	YQNNNA--	-----NT	TV-----	---NASTVYNG	KGEIKCSFN	VTTGIRDKTK
B.US--.9250_pre_P9.NEW_MH575728	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	LQNSTST-N-	-----NN	TNT-----	---NTTIDNNM	KGEIKCSFN	ITTSSTKEK-
B.US--.9249_pre_G3.NEW_MH575796	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	VEPNSNE-KA	-----NS	T-----	---NNGANIT	AADMKCSFN	ITASIRDRVKQ
B.US--.9248_pre_B8.NEW_MH575857	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	ANFTTTS-ST	T-----NS	SS-----	---TTEIKRT	KGEIKCSFN	ITTTMRDRKVO
B.US--.9241_pre_L6.NEW_MH575991	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	LE-----	-----NS	-----	---TTEIKRT	KGEIKCSFN	ITTTMRDRKVO
B.US--.9247_pre_A1.NEW_MH576039	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTN	LNVNT--	-----NS	S-----	---SKEGKM	AEREMKCSFN	ITTEIRKNNVO
B.US--.9246_W12_12K21.NEW_MH576084	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	IRNSTVT-SN	TTTS--NTT	TT-----	---WGEMTCSFN	IT-----	IT-----
B.US--.9245_pre_H6.NEW_MH576107	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	VNATATM-G-	-----ND	T-----N	---DNSGKKLME	IGEMKCSFN	IT-----
B.US--.9244_rebound_S21_G8.NEW_MH576152	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTE	LKNNTNT-ST	TM-----NN	S-----	---SSIEEREM	TRMTCNSFN	IT-----
B.US--.9243_W12_12E19.NEW_MH576262	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	NLNLNCT-NN	NCTCS--NN	T-----	---NYNMTIE	KGEIKCSFN	IT-----
B.US--.9242_W12_12H20.NEW_MH576349	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	LRNDTVG--	-----NQ	TM-----	---YLNENTIQ	GREMTCSFN	IT-----
B.US--.91C22_D0_11_9_582.NEW_MH632763	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	-----	-----NS	-----	---DQML	EGBMKCSFN	IT-----
B.US--.91C35_D0_3HG_O10517_D6_S46.NEW_MH632822	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTN	TTKD-----	-----NI	TI-----	---EG	NGEIKCSFN	IT-----
B.US--.91C33_D0_9_4_c9_S49.NEW_MH632949	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	LTNNNTT-HG	KSQS--SKNS	S-----	---SEWERV	REEMTCNSFN	IT-----
B.US--.91C34_D0_YBAC10_S26.NEW_MH632956	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	VNVATG--	-----NS	T-----	---WEKME	KGEIKCSFN	IT-----
B.US--.9341_D0_1_2812A3_S72.NEW_MH633031	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTN	VTGNVTE-RI	NDVAT--DNA	T-----	---INKNA	TDEMTCNSFN	IT-----
B.US--.9341_D0_1A_1118G10_S87.NEW_MH633039	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	LKNATKG-DV	TRG--NN	SNA-----	---NIGNNSSTEA	MAEMKCSFN	IT-----
B.US--.9342_D0_2_2812A6_S1.NEW_MH633155	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	CDSSPNC-TD	IKNST--NA	T-----NSI	---KNSTDAITST	KGEIKCSFN	IT-----
B.US--.9342_D0_2012B10_S11.NEW_MH633164	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	-----	-----NS	-----	---ST	EGBMKCSFN	IT-----
B.US--.9343_D0_3_2812_A2_S55.NEW_MH633199	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTN	ATTSCTAN-GN	-----NS	-----	---INCTN	VEEMKCSFN	IT-----
B.CH--.R282_3_4_c156.NEW_MH714324	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTN	LNCTNCT-TT	NSTT--NS	TNST-----	---NSTWE	DKGLRCSFN	AP-----
B.CH--.T294_3_8_c128.NEW_MH714329	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	LTNSDIT-D-	-----NT	T-----	---INSNWEG	KGEIKCSFN	IT-----
B.CH--.T11_10_6_c117.NEW_MH714343	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	LKKNLCT-CA	NLTKGS--NC	T--NNSS-VGN	DTMTNNSWI	GEMKCSFN	AT-----
B.CH--.R11_8_1_c157.NEW_MH714351	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	LKK-----	-----ND	TSTNSDTTAS	TNTNSWGEDM	KGEIKCSFN	IT-----
B.CN.2005.05071.NEW_MK053935	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTN	YNNTNS--	-----NT	SST-----	---TE	GGEIKCSFN	IT-----
B.US.2013.ARC-1a.NEW_MK214316	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	LRNLTOT-SN	ETQTS--NRNQ	T-----	---SRTIEPQM	GGEIKCSFN	IT-----
B.FR.1983.HMK1-LAI-11B-BW0_K03455	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	LKNNTNT-S	-----NS	-----	---SGRMIE	KGEIKCSFN	IS-----
B.AR.2000.85096FL_2000.KY968401	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTN	LKN-----	-----NN	TTTTNTAASP	-----	GGEIKCSFN	IT-----
B.AR.2000.85323FL_2000.KY968402	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTA	LKNSTTT-SD	ST-----SNG	T-----	---VSPA	REEMTCNSFN	IT-----
B.AR.2000.85891FL_2000.KY968403	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTN	LNNATNA--	-----NI	S-----	---QRI	EGBMKCSFN	IT-----
B.AR.2000.86218FL_2000.KY968404	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	ANVTAT--	-----NI	S-----	---WGEAIG	KGEIKCSFN	IT-----
B.AR.2000.ARMS008.AY037269	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	LNATEIN-MN	NA-----NA	T-----G	---TMTNITTKD	IQNMKCSFN	IT-----
B.AR.2001.101815FL_2001.KY968394	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	VTISPT--	-----AMSNV	T-----	---I	REEMTCNSFN	IT-----
B.AR.2001.107478FL_2001.KY968397	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTN	WHVWNT-TE	-----NN	TN-----	---SNEVDLKM	EGBMKCSFN	ITTPGIRKIK
B.AR.2001.89825FL_2001.KY968407	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTN	AVGVNN-SS	-----NT	-----	---NATSEEA	LADVKCSFN	VT-----
B.AR.2001.89886FL_2001.KY968408	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	YTNST--	-----NS	TN-----	---LEQM	KGEIKCSFN	IT-----
B.AR.2002.02AR114146.DQ383746	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTN	WNSTNA-TD	I-----TTNA	T-----	---DMG	EGBMKCSFN	IT-----
B.AR.2002.116877FL_2002.KY968396	DAWKNNMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	YNGVTA--	-----NN	SSIN-----	---NSSGGREME	KGEIKCSFN	IT-----
B.AR.2003.03AR137681.DQ383748	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	TDQ-----	-----NI	T-----	---RGIE	KLEIKCSFN	IT-----
B.AR.2003.03AR138910.DQ383749	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	VNDTKAN-S-	-----NS	T-----	---GSSWDMFG	GEMTCNSFN	ITTSIRDRKVO
B.AR.2003.134742FL_2003.KY968397	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTN	WHVWNT-TE	-----NN	TN-----	---SNEVDLKM	EGBMKCSFN	ITTPGIRKIK
B.AR.2004.04AR151263.DQ383751	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	YVGNTH-AG	-----NT	-----	---NTTSTVSPIT	STGMKCSFN	ITVTDLRDKVK
B.AR.2004.04AR151516.DQ383752	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	WKSNA--	-----NS	TR-----	---INNSSLGTM	EQGKCSFN	ITISIRKNNVO
B.AR.2004.145447FL_2004.KY968398	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	LKNNT--	-----NS	SMT-----TT	---PSSRTGEFME	RGEIKCSFN	ITVTDLRDKVK
B.AR.2004.159687FL_2004.KY968399	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	VTNKNAT--	-----NT	TSDATSTP--	-----SSSE	EVQMKCSFN	ITVSPITNNVK
B.AR.2007.DEURF07AR001.KY658686	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	ANRTENT--	-----NI	T-----	---RQMD	KELIKCSFN	ITVTDLRDKVK
B.AR.2008.DEMB08AR002.KY658683	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	WTAD-----	-----NN	TN-----	---STIRREM	ERNIKCSFN	ITTEVODKRR
B.AR.2010.DEMB10AR006.KY658684	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	YST-----	-----NA	TNTNNT--	-----	KELIKCSFN	ITSSLDKNNVO
B.AR.2014.DEMB14AR003.KY658685	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTD	VPT-----	-----NC	TNTG-----	---NCTMKWIKI	GELIKCSFN	ITVTDLRDKVK
B.AR.1998.ARC0H54.AY037268	NMWNNSMVQ	MHEDIISLWD	QSLKPCVKLT	PLCVTLNCTN	WMRTNNT-L-	-----NN	TSNNA--TNG	DITRINTNTE	NGEVCNSFN	ITTMKDKVO

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

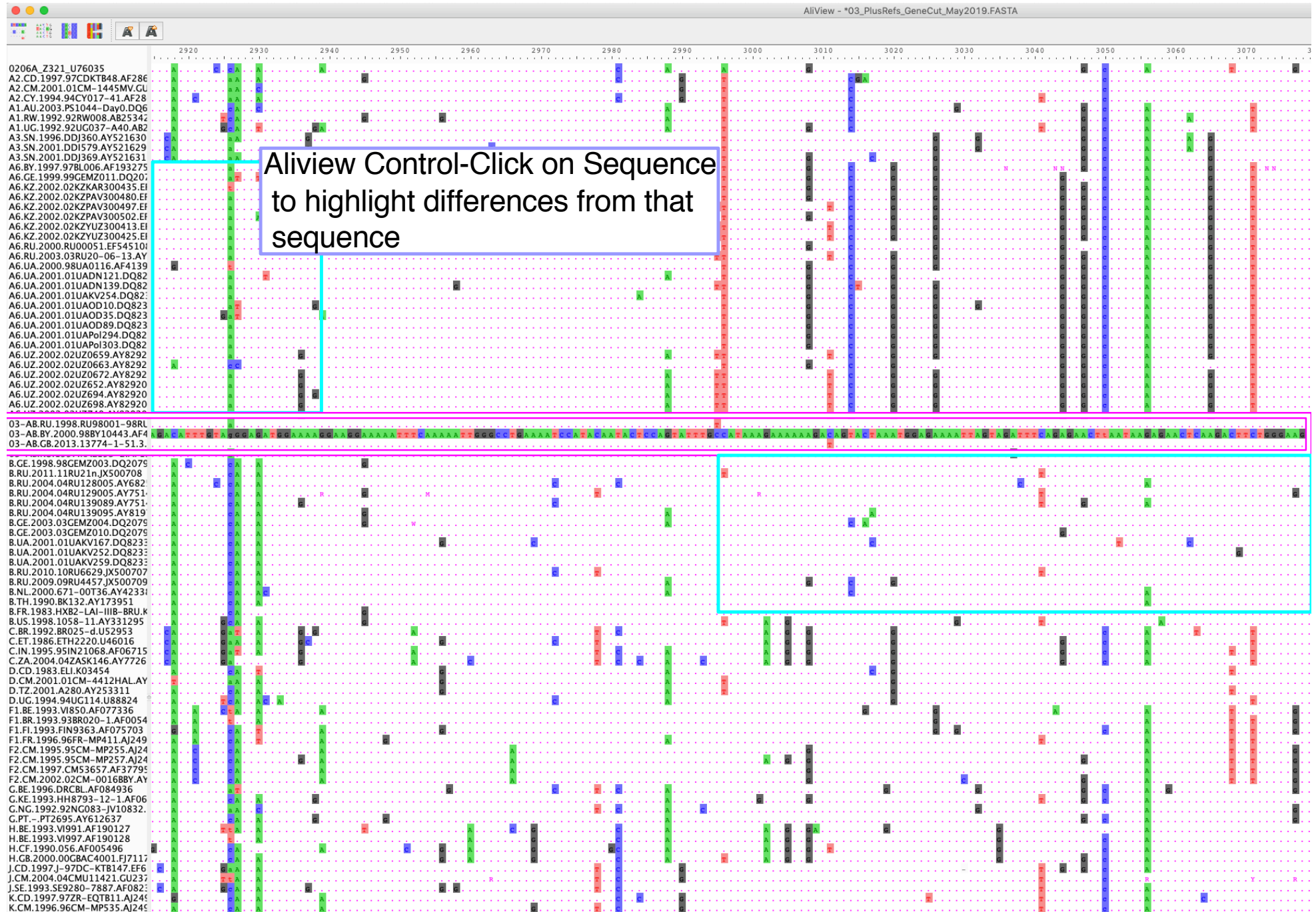


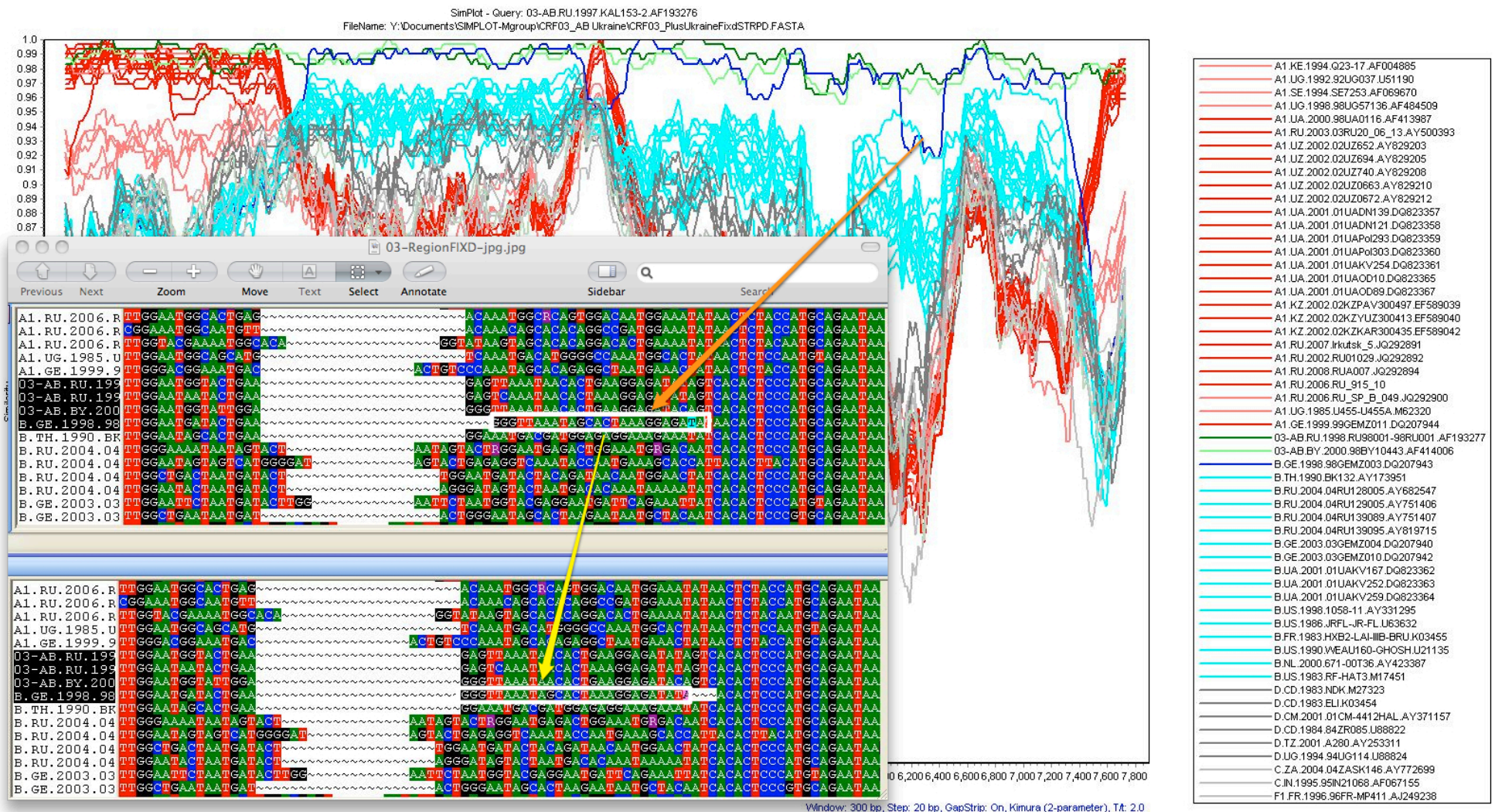
Sequence position(s) of High N-glycosylation Site (>60%)

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









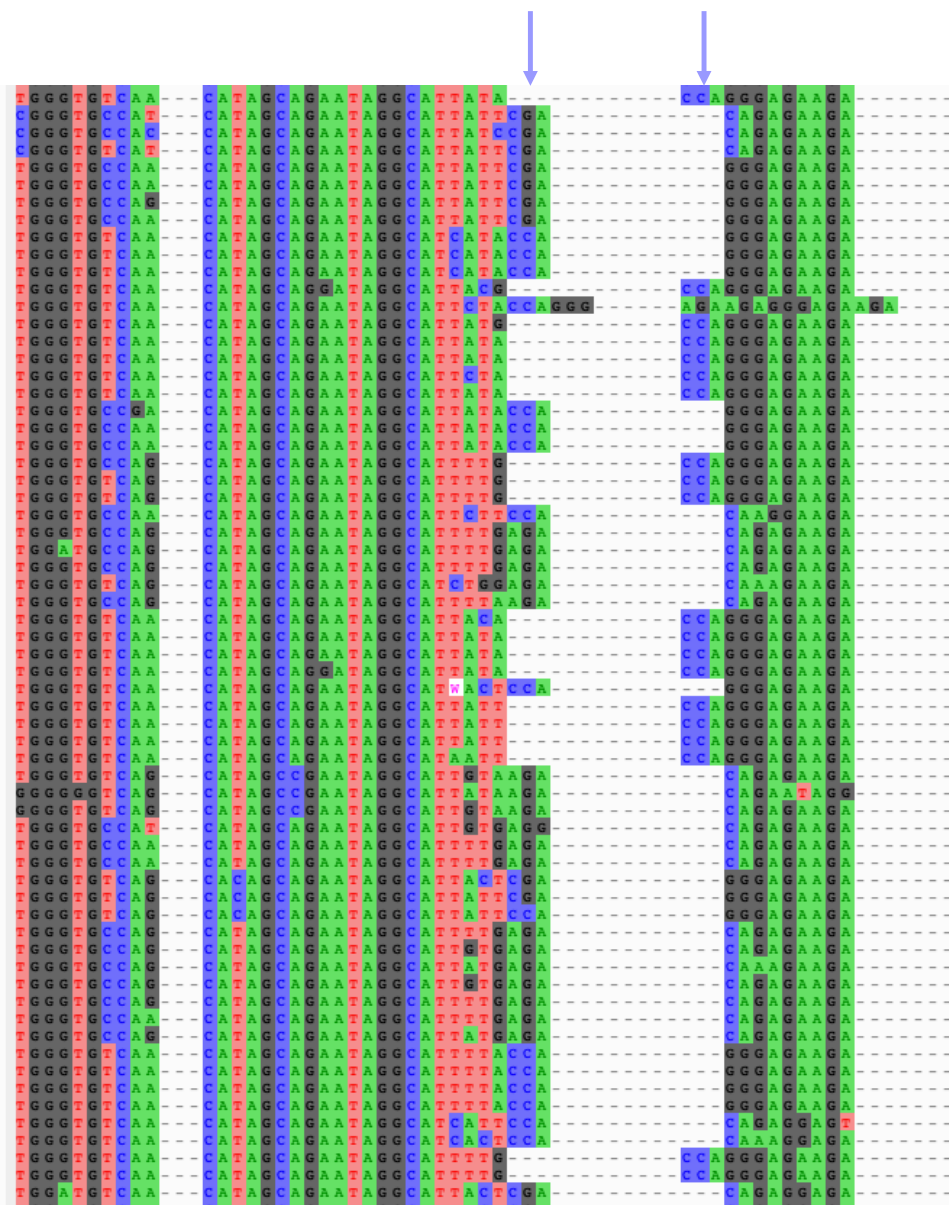
Even a very small region of misalignment, hypermutation, or poor sequence quality can have a large impact on similarity plots, phylogenetic trees, and other analyses. Similarity plots can be quite useful for identifying sites in a multiple sequence alignment that should be scrutinized, and corrected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

<https://sray.med.som.jhmi.edu/SCSoftware/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\_O1B.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\_O1B.MN.12.12MNG12712\_LC312714  
 51\_O1B.MY.11.11MYKL055\_KJ485697  
 51\_O1B.SG.11.11SG\_HM021\_JN029801  
 52\_O1B.MY.03.03MYKL018\_1\_DQ366664  
 52\_O1B.TH.00.00TH\_R1741\_AY945734  
 52\_O1B.TH.96.M043\_DQ354113  
 53\_O1B.MY.04.04MYKL016\_1\_DQ366663  
 53\_O1B.MY.10.10MYKJ067\_JX390612  
 53\_O1B.MY.10.10MYKJ079\_JX390611  
 53\_O1B.MY.11.11FIR164\_JX390610  
 54\_O1B.MY.07.07MYKLD49\_EU031915  
 54\_O1B.MY.08.08MYKL044\_JX390977  
 54\_O1B.MY.09.09MYSB023\_JX390976  
 55\_O1B.CN.08.08CYM047\_JF340054  
 55\_O1B.CN.10.HNCS102056\_JX574661  
 55\_O1B.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.YNFI37\_KC870044  
 57\_BC.CN.10.DH17\_KF250400  
 58\_O1B.MY.09.09MYPR37\_KC522031  
 58\_O1B.MY.10.10MYKJ036\_KC522035  
 58\_O1B.MY.10.10MYPR87\_KF425293  
 58\_O1B.MY.11.11MY1R1J704\_KC522033  
 58\_O1B.MY.11.11MY1ZK731\_KC522032  
 59\_O1B.CN.07.GD070126\_KF011494  
 59\_O1B.CN.09.09LNA423\_JX960635  
 59\_O1B.CN.11.11CN.LNSY300876\_KJ484434  
 59\_O1B.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.10.JL100010\_KC990124  
 62\_BC.CN.10.YNFI13\_KC870034  
 62\_BC.CN.10.YNFI15\_KC870035  
 63\_O2A.RU.10.10RU6637\_JN230353  
 63\_O2A.RU.13.RU\_8169\_KJ197201  
 63\_O2A.RU.13.RU\_8501\_KJ197202  
 64\_BC.CN.09.09YNLX047sg\_KC898994  
 64\_BC.CN.09.09YNLX219037sg\_KC899009  
 64\_BC.CN.09.YNFI31\_KC870042  
 64\_BC.CN.09.YNFI33\_KC870043  
 64\_BC.CN.10.YNFI10\_1\_KC870032  
 64\_BC.CN.10.YNFI16\_KC870036  
 64\_BC.CN.10.YNFI22\_KC870040  
 65\_cpx.CN.10.YNFI01\_KC870027  
 65\_cpx.CN.10.YNFI02\_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\_O1B.CN.11.ANHUI\_HF115\_KC183779  
 68\_O1B.CN.11.ANHUI\_WH73\_KC183782  
 69\_O1B.JP.03.JP\_5091K231\_AB845344  
 69\_O1B.JP.05.05JN132850\_LC327100



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

The screenshot displays the AliView application window. The 'Align' menu is open, showing options for re-aligning sequences. The main window shows a large alignment of sequences, with a specific block highlighted in blue. The sequences are color-coded by amino acid type (e.g., red for basic, green for acidic, blue for hydrophobic). The interface includes a menu bar (File, Edit, Selection, View, Align, Tools, External commands, Help) and a toolbar with icons for various alignment and editing functions.

Menu options visible in the 'Align' menu:

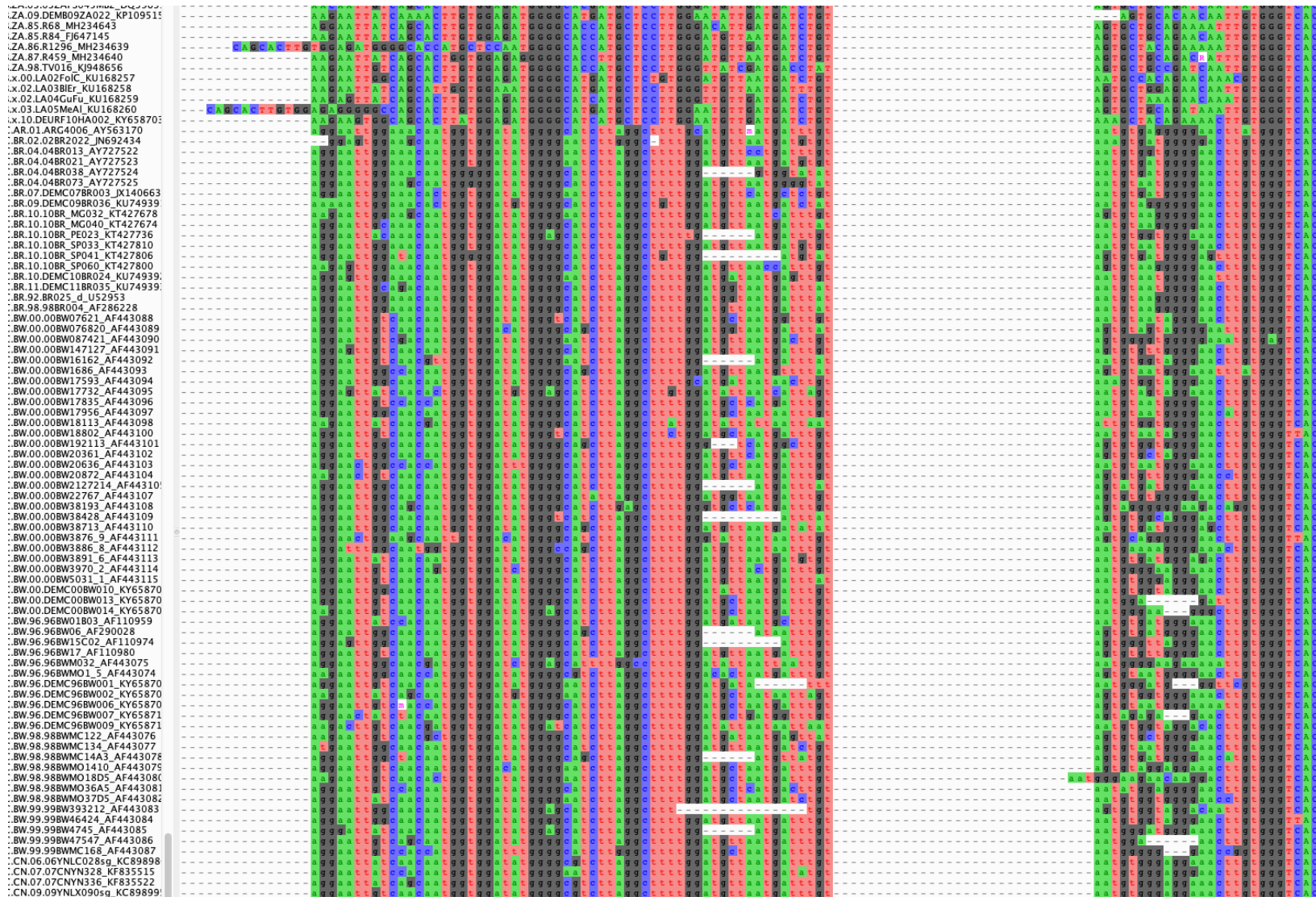
- Add and align sequences from clipboard (fasta)
- Add and align sequences from file (fasta)
- Realign selected block**
- Realign selected sequence(s)
- Realign everything
- Realign everything as Translated Amino Acids
- Change default Aligner program
- Move selected positions right
- Move selected positions left
- Insert Gap move right
- Insert Gap move left
- Delete Gap at left
- Delete Gap at right

Sequences listed on the left include:

- BUS.KC0413T\_FJ469726
- BUS.KF703\_DQ886031
- BUS.KF710\_FJ469729
- BUS.KF714\_DQ886032
- BUS.KF7204\_DQ886033
- BUS.KL8146\_DQ886034
- BUS.KL8157\_DQ886035
- BUS.KL8185\_DQ886036
- BUS.KL8188\_FJ469744
- BUS.KL896\_FJ469750
- BUS.KNCQ\_FJ469752
- BUS.KPRL505\_FJ469756
- BUS.KPRL517\_FJ469761
- BUS.KPRL519\_FJ469763
- BUS.KPRL528\_FJ469766
- BUS.KSH8127\_FJ469767
- BUS.KSH8183\_FJ469768
- BUS.KSH8229\_FJ469769
- BUS.KSH8241\_FJ469771
- BUS.sample\_C\_BID\_D617\_K503
- BUY.O1.01UYTRA1092\_AY781121
- BUY.O1.01UYTRA1179\_AY781122
- BUY.O2.02UY\_TSU1290\_NJ23595
- BUY.09.09UY\_TBA0177\_NJ23596
- BYE.O2.02YES07\_AY795904
- BYE.O2.02YES08\_AY795905
- BZA.O0.TV047\_K048617
- BZA.O2.TV1057\_K048660
- BZA.O3.03ZAP0545MB2\_DQ39631
- BZA.O9.DEMB05Z022\_KP109515
- BZA.S5.R68\_MH234643
- BZA.S5.R84\_FJ647145
- BZA.S6.R1296\_MH234639
- BZA.S7.R455\_MH234640
- BZA.S8.TV016\_K048656
- B.x.00.LA02FoIC\_KU168257
- B.x.02.LA03BIEF\_KU168258
- B.x.02.LA04Gufu\_KU168259
- B.x.03.LA05MeAl\_KU168260
- B.x.10.DEUR10HA02\_KY658702



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.

Can not realign

If you want to align Nucleotides as translated Amino Acids, please use function:  
Realign everything as Translated Amino Acids

OK

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 [btf@lanl.gov](mailto:btf@lanl.gov) or [seq-info@lanl.gov](mailto:seq-info@lanl.gov) for advice or help.