

Multiple Sequence Alignment

Presenter: Brian Foley

btf@lanl.gov

HIV Databases

Theoretical Biology and Biophysics,

Los Alamos National Laboratory



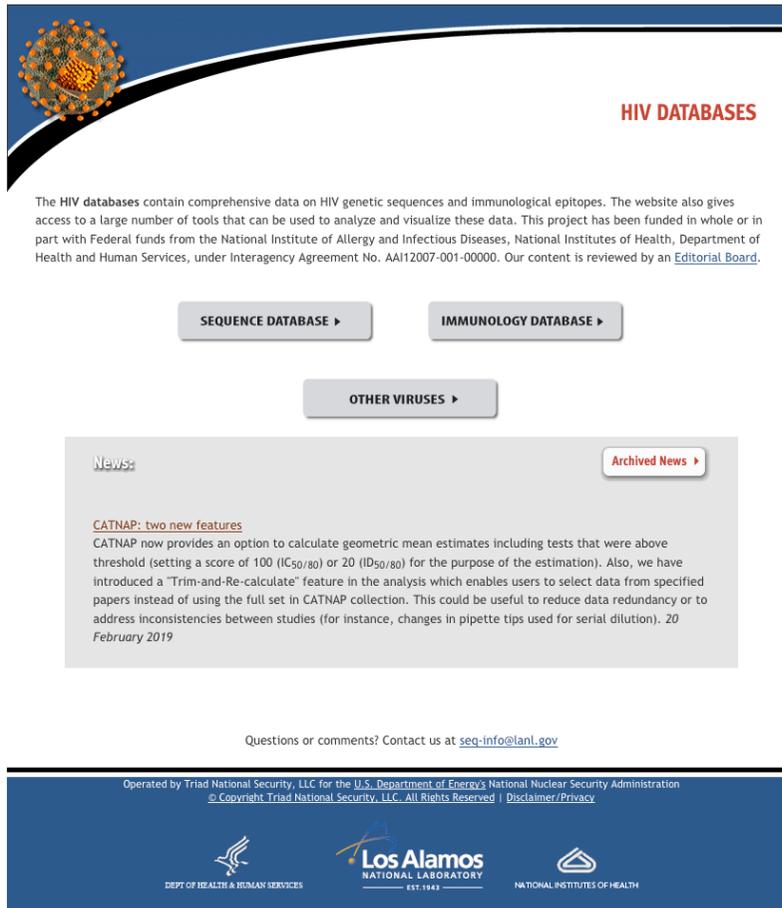
www.hiv.lanl.gov
seq-info@lanl.gov



Presenter: Brian Foley

btf@lanl.gov

HIV Databases



The screenshot shows the top section of the HIV Databases website. It features a blue header with a virus particle icon on the left and the text "HIV DATABASES" in red on the right. Below the header, a paragraph describes the site's content. Three navigation buttons are visible: "SEQUENCE DATABASE", "IMMUNOLOGY DATABASE", and "OTHER VIRUSES". A "News" section contains a recent update about CATNAP features. At the bottom, there is contact information and 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_{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

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](#)



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.

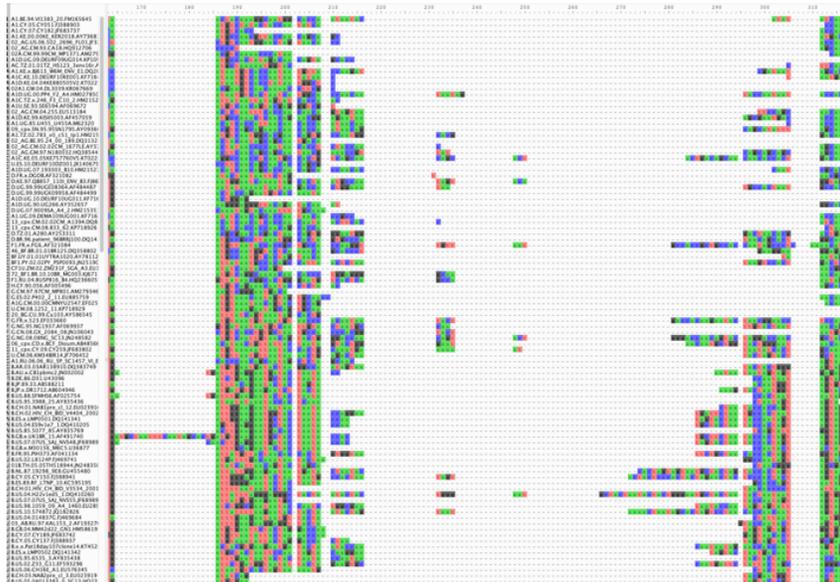
Pairwise vs multiple alignment

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

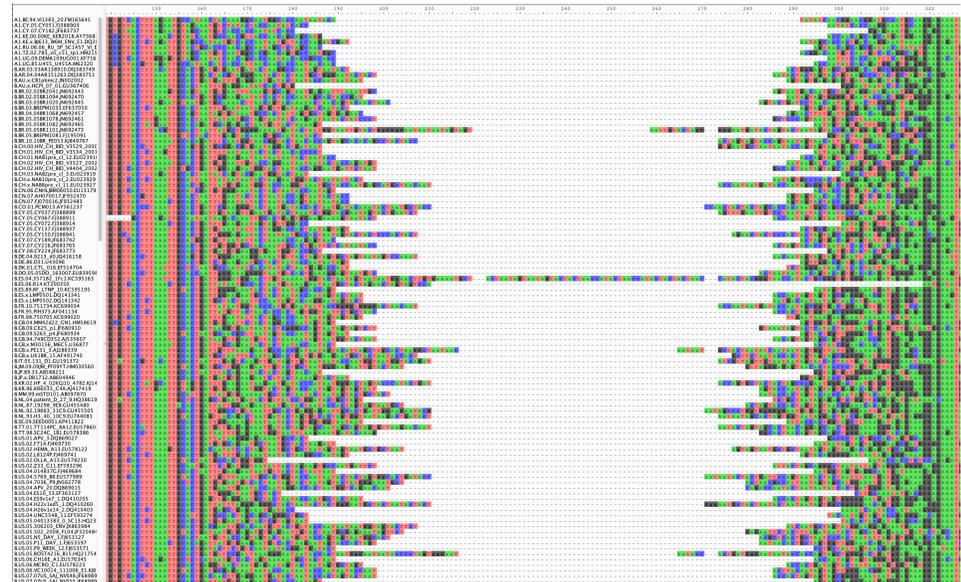
Multiple alignment is much more difficult.
“Optimal” alignment is debatable.

BLAST
Needle (EMBOSS)
Water (EMBOSS)
Align0

ClustalW ClustalOmega
MUSCLE
MAFFT
GeneCutter (HIV specific)

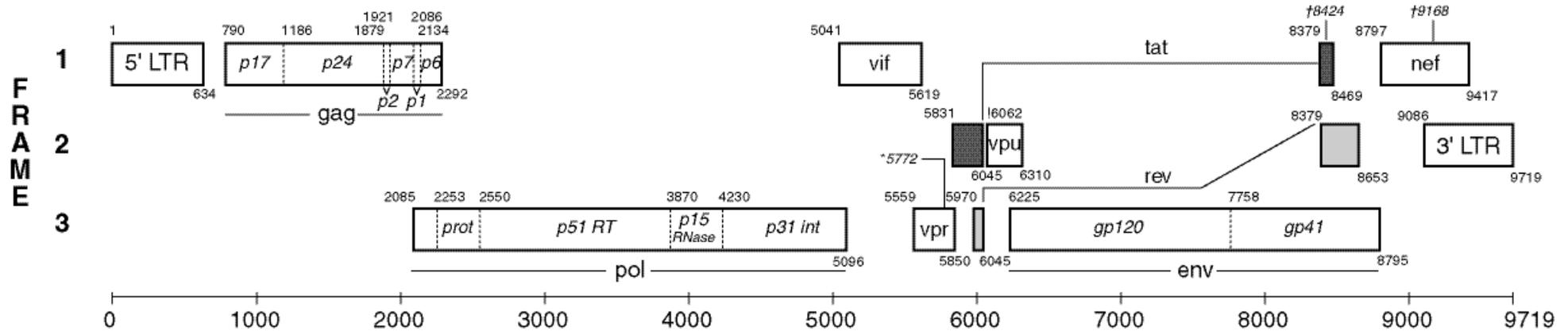


Typical alignment of HIV-1 env by MAFFT or similar tools. Cannot be translated to amino acid sequences.



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

HIV-1 Genome

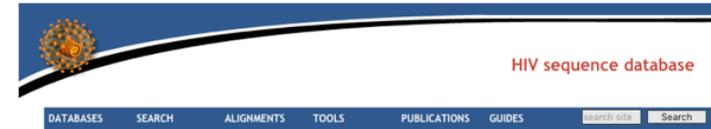
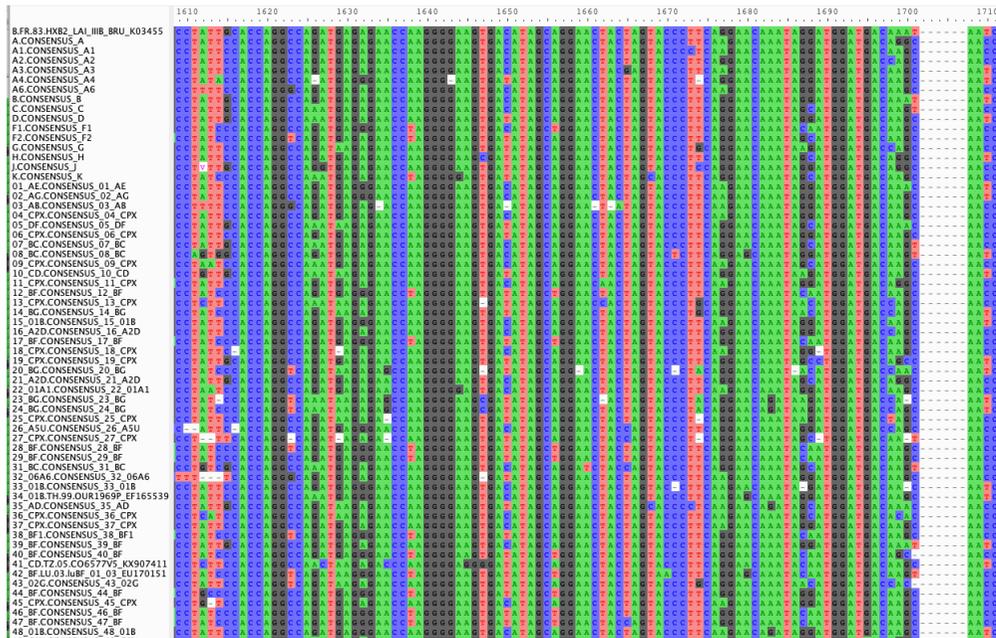
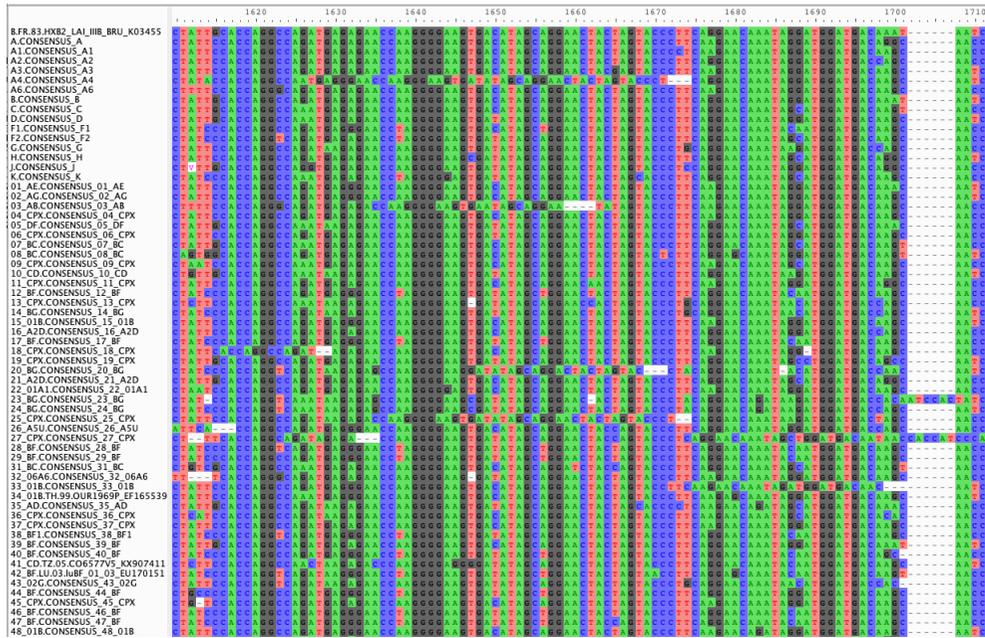


<https://www.hiv.lanl.gov/content/sequence/HIV/MAP/landmark.html>

Nine proteins/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.
- Generate nucleotide and protein alignments of the cut regions.

Please read [Gene Cutter Help](#), particularly before running [large jobs](#).

Input

Organism: HIV1 (HXB2)

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

Reference options: Insert HXB2(K03455) for HIV-1 or SW6239(M33262) for HIV-2/SIV from the results
 Remove HXB2(K03455) for HIV-1 or SW6239(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
 Return results as nucleotides:
 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

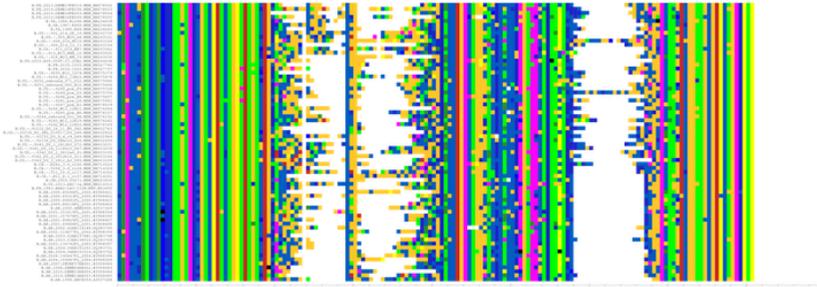
Input & Options

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

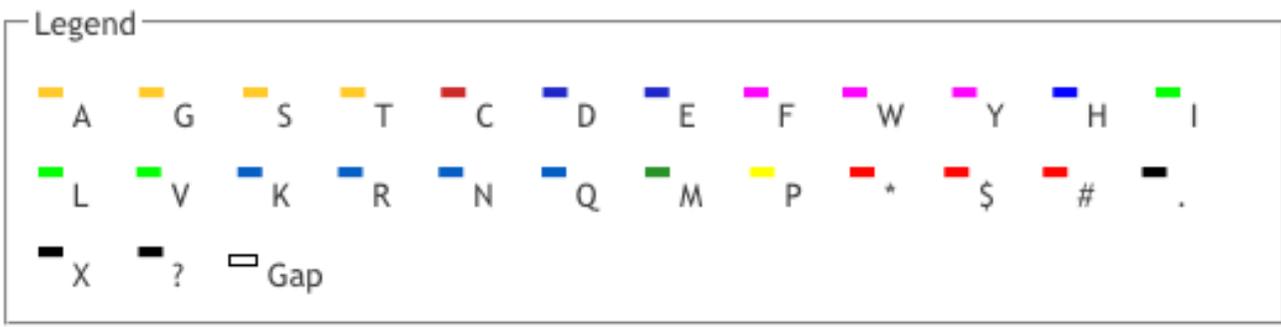
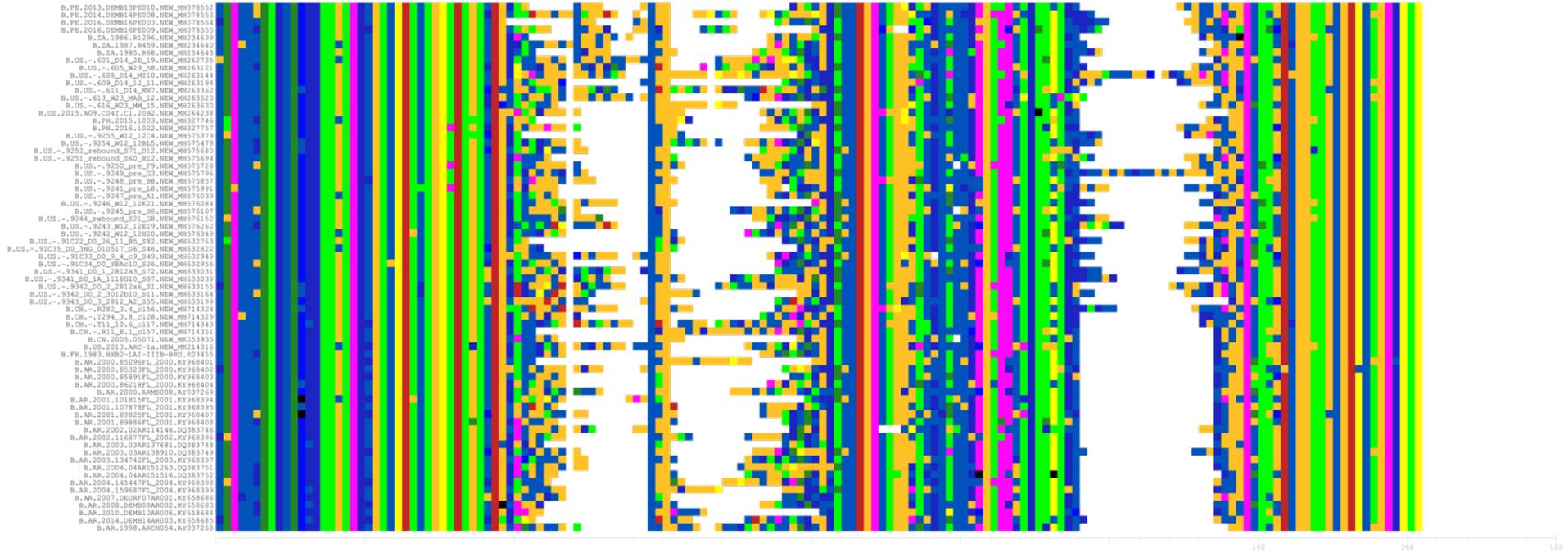


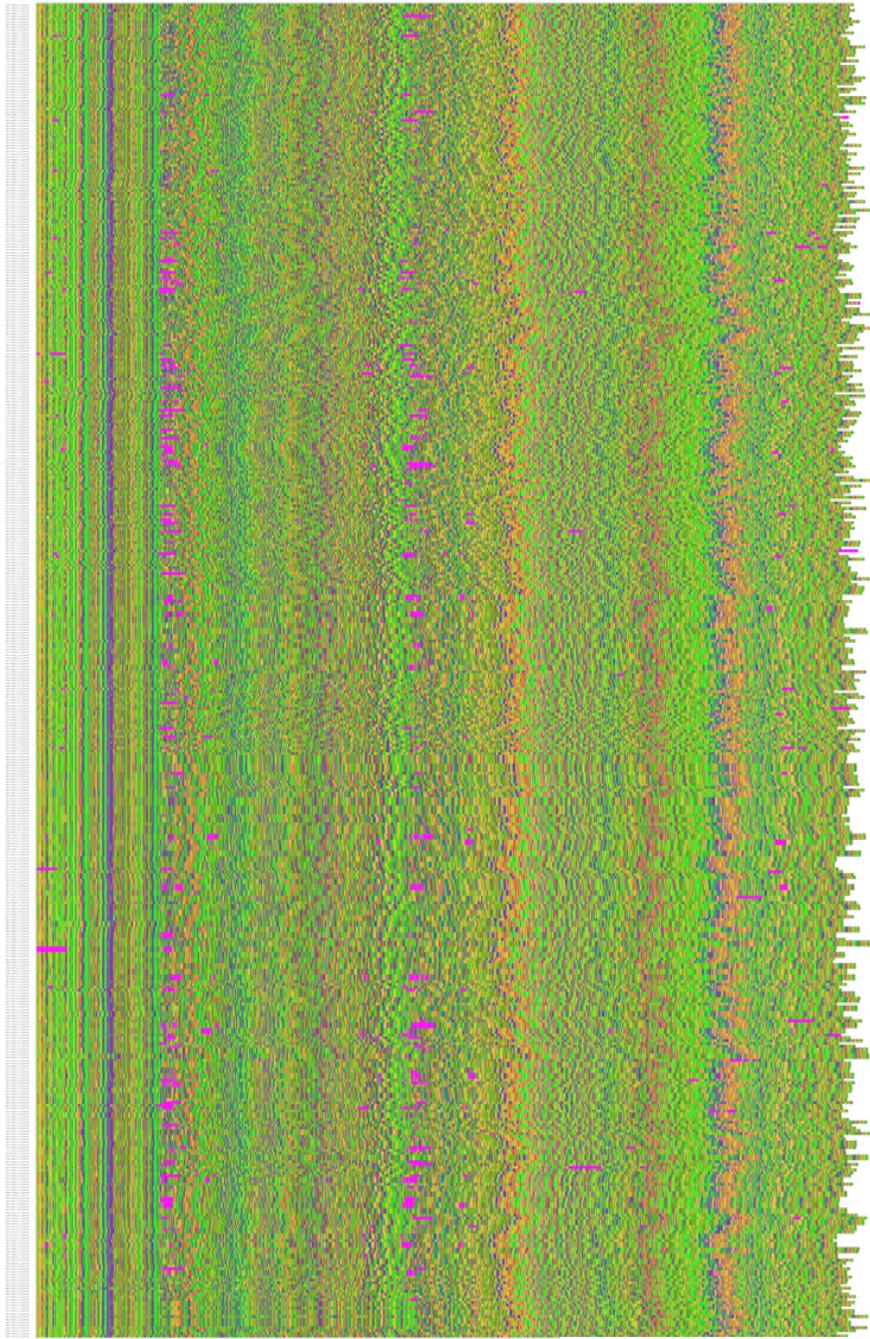
Legend

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.

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





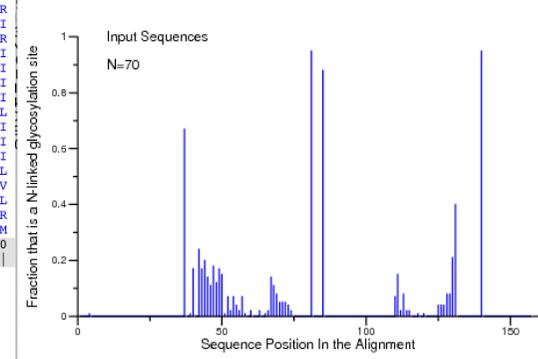
```

10      20      30      40      50      60      70      80      90      100
B.PE.2013.DEMB13PE010.NEW_MH078552  NMWNSNVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTD LSNSTT-----SN GSESLRGEEMK NCSFNVITPVD SDRKKREYAL
B.PE.2014.DEMB14PE008.NEW_MH078553  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTD VICTTNSNI TTCTNSNS---TTSFNITSP NITTSPTEMK NCSFNVITPQ RDKTKEEYAF
B.PE.2016.DEMB16PE003.NEW_MH078554  DMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTN -----STRGST IGQNRGRIEQ NCSFNVITVV HDVKVKREYAL
B.PE.2016.DEMB16PE009.NEW_MH078555  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTN VITGTNNITD -----GGNDT RINVTDEIK NCSFHT--HI GDQKREYAI
B.ZA.1986.R1296.NEW_MH234639  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTD WRNTTNTNT-----NKTSS NTNNSIIEGEMK NCSFNVITSI RDKVKKESAL
B.ZA.1987.R459.NEW_MH234640  NMWNTNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTD DLKNTNTNNA TMT-----NASS WGNIEKGEIK NCSFNVITNR RGMKREYAL
B.ZA.1985.R68.NEW_MH234643  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTD NLRNDTNTN S-----SGIG GIKMEKGEIK NCSFNVITNI RDKFKRYAL
B.US.--.601.D14.2E.19.NEW_MH262735  TMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTD LNNATNAST NGNNATNAST ---NTSSYSVW GPRKEEIK NCSFNVITSI KMKVREYAL
B.US.--.605.W29.H8.NEW_MH263121  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTD PKNDTANGNE GEMKRCSS---FNVSSIVRD TANGNEGEMK NCSFNVSSIV RDKVKREYAL
B.US.--.608.D14.M10.NEW_MH263144  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTD SAPSTPSNPT TSPTIKREVK NCSFNVITSI RDKVKREYAL
B.US.--.609.D14.12.11.NEW_MH263194  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTN AMITSHG-----QAS TIKMEKIEIQ NCSFNVITFG RDKLOREYAL
B.US.--.611.D14.MH7.NEW_MH263362  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTN ATRIVNGTIV EMKNCSLAIN ---GTEEMKCS IARTNTEMK NCSFNVITDI RDKLOREYAL
B.US.--.613.W23.MAB.12.NEW_MH263520  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTD LNAATNSNNT SNATNSN---TTHNSS WADNKEKIEK NCSFNVITNI GDMKREYAL
B.US.--.616.W23.MM.15.NEW_MH263630  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTN VTLNNTNAST TSTTSPSS---S-----SGEMK NCSFNVITAP RDKIORYGI
B.US.2015.A09.CD4T.C1.20B2.NEW_MH264238  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTD YVGTDTTNS STT-----TN TNISMEGEMK NCSFVITPNI RDKREYAL
B.PH.2015.1003.NEW_MH327746  NVWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTN ATINSITPE-----T ELRTLKKEVK NCSFNVITSI KMKKREYAL
B.PH.2016.1022.NEW_MH327757  NIWKNNSVQV MHEDIISLWD QSLKPCVKLL PFCVLTLDCT YVGTDTTNS STT-----GG ATIMQKGEIK NCSFNVITSI QDRKREYAL
B.US.--.9255.W12.12C4.NEW_MH575379  NAWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTD VNRST-----EMTGELEK NCSFNVITKI TNKVKREYAL
B.US.--.9254.W12.12BL5.NEW_MH575478  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTD VEIDKTSANK TINVNTN-----TNS WERMDDPEIK NCSFNVITNI RDRVKREYAL
B.US.--.9252_rebound_S71.D12.NEW_MH575680  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTN LTIDPNTAV T-----SIS GSLEKGEEMK NCSFNVITSL RDKKREYAL
B.US.--.9251_rebound_S60.H12.NEW_MH575694  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PFCVLTLDCT YNNANTVTN-----NAST TVNNGKIEK RDKTQLEYAL
B.US.--.9250_pre_F9.NEW_MH575728  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PFCVLTLDCT LQNSTTNNN TMT-----NPT IDNNMKEIK NCSFNVITSI KEK--KGYAF
B.US.--.9249_pre_G3.NEW_MH575746  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTN VTRAKGVTIN DVAT-----DADT WEFMEKIEK NCSFNVITSI KAKVREYAL
B.US.--.9248_pre_B8.NEW_MH575857  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTN ANFTSTST NA-----SSTT GEIKTGEIK NCSFVITTDG RDKKRNAL
B.US.--.9241_pre_L8.NEW_MH575991  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PFCVLTLDCT ADTGTN-----SSS STLEEKIEK NCSFNVITSI SDPKQREYAL
B.US.--.9247_pre_A1.NEW_MH576039  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTD LNVNTN-----SSS KEGMAEMK NCSFNVITSI RNKMKREYAL
B.US.--.9246.W12.12K21.NEW_MH576084  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTD IRNSTVSTN TTSNTTT-----WGMTM NCTFNITSI KDKMKREYAL
B.US.--.9245_pre_H6.NEW_MH576107  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTD NVNATATMDND ---TNDSSK GKLMIEGEMK NCSFVITSI GEKK--TENAL
B.US.--.9244_rebound_S21.G8.NEW_MH576152  NAWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTE LKNNNTST MNNT-----SSI EEREMTREM NCSFNVITSI RDKVQREYAL
B.US.--.9243.W12.12E19.NEW_MH576262  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTD NLNLCNPNNN TCS-----NNTN YNTEKIEK NCSFNVITSI RDRVTEHAL
B.US.--.9242.W12.12H20.NEW_MH576349  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTD LRNDTVGNQT N-----YLN E TMTQREMT NCSFNVITSI RDKVKREYAL
B.US.--.91C22.D0.26.11.B5.S82.NEW_MH632763  NIWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTN LNVNTE-----DQMEKGGMQ NCSFNVITAGI KDRMKREYAL
B.US.--.91C35.D0.3HG.010517.D6.S46.NEW_MH632822  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTN TTKDNITTI-----EGNIEK NCSFNVITDL RDKVQREYAL
B.US.--.91C33.D0.9.4.C9.S49.NEW_MH632949  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTN LTNNTTHGK SGS-----SKNSSS EWERKREIK NCSFNVITSI G--KRRREYAL
B.US.--.91C34.D0.YBAC10.S26.NEW_MH632956  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTN VTRAKGVTIN DVAT-----DADT WEFMEKIEK NCSFNVITSI KAKVREYAL
B.US.--.9341.D0.1.2812A3.S72.NEW_MH633031  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTN VTRAKGVTIN DVAT-----DADT WEFMEKIEK NCSFNVITSI KAKVREYAL
B.US.--.9341.D0.1A.1118G10.S87.NEW_MH633039  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTD LKNAKGTVDI RGNNSNA-----DIGNN SSTEAMEMK NCSFNVITSI KDKKREYAL
B.US.--.9342.D0.2.2812A6.S11.NEW_MH633155  DMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTD CDSSPNCIDI KNSTNAT---NSIKNSTD ATTSRTRIEK NCSFNVITNM RDVQREYAL
B.US.--.9342.D0.2.3012B10.S11.NEW_MH633164  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTD CNPNCTALR NSITNAT---S-----STEGEMK RCFNVITSI RDKVREYAL
B.US.--.9343.D0.3.2812.A2.S55.NEW_MH633199  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTN ATSCANQN-----GNST INCTNVEEMK NCSFNVITSI KMKVREYAL
B.CH.--.R282.3.4.C156.NEW_MH714324  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTN LNCITNCTN STNSTN-----ST NCTWEDKELK NCSFVITSI RDKVKREYAL
B.CH.--.T294.3.8.C128.NEW_MH714329  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTD LNSDSTDTT T-----IN SNWEGKIEK NCSFNVITSI RDKVQREYAL
B.CH.--.T11.10.6.C117.NEW_MH714343  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTD LKKNLTCAN LTRKSNCTNN SSVGNDTNN NSSWIGEDMR NCSFNVITSI RDKVQREYAL
B.CH.--.R11.8.1.C157.NEW_MH714351  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTD LKKNSTSTNS DTTASTN-----SIVGNDTNN NSSWIGEDMR NCSFNVITSI RDKVQREYAL
B.CN.2005.050711.NEW_MK053935  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTN YNNSTSTSS T-----TEGIEK NCSFNVITSI NTKV--KDAL
B.US.2013.ARC-1a.NEW_MK214316  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTD LRNEIOTSE TQTSNRN-----QTSNRN IEPOMGIEK NCSFNVITVD KDKKREYAL
B.FR.1983.HXB2-LA1-IT1B-BRU.K03455  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTD LKNDTHNS S-----SSG RMLMEKIEK NCSFNVITSI G--KRRREYAL
B.AR.2000.8509FPL.2000.KY968402  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTD LKNAKGTVDI RGNNSNA-----DIGNN SSTEAMEMK NCSFNVITSI KDKKREYAL
B.AR.2000.8532FPL.2000.KY968403  DMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTA LKNTSTSDS T-----SNG TVSPAREEMK NCSFNVITSI RMLKREYAL
B.AR.2000.8589FPL.2000.KY968404  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTN LNNATNANI-----S-----SORIEGEMK NCSFNVITSI GDMKREYAL
B.AR.2000.8621FPL.2000.KY968404  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTD ANVTRATNI-----SW GEALGKIEK NCSFNVITNM RDKVKREYAL
B.AR.2000.ARMS008.AY037269  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTD LNATEINMNN ANA-----TGTNTN ITTKDIONMK NCSFNVITSI KMKVREYAL
B.AR.2001.101815FPL.2001.KY968394  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTD VTIISPNT-----AW SNVITREEMT NCSFNVITSI RDKVTEHAL
B.AR.2001.107878FPL.2001.KY968395  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTN ANCTNANNTC -----NA TSEALRDVK NCSFNVITNI RDKKREYAL
B.AR.2001.89825FPL.2001.KY968407  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTN ANCTNANNTC -----NA TSEALRDVK NCSFNVITNI RDKKREYAL
B.AR.2001.89886FPL.2001.KY968408  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTN YNNSTSTSS T-----TEGIEK NCSFNVITSI NTKV--KDAL
B.AR.2002.02AR114146.DQ383746  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTN WNSTNATDI TTN-----TTN ATDMGKIEK NCSF--QPR
B.AR.2002.116877FPL.2002.KY968396  DAWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTD YNGNVTANN S-----NSSG GREMEKIEK NCSFNVITNI
B.AR.2003.03AR137681.DQ383748  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTD TDTQNT-----T-----TRGLEKIEK NCSFNVITSI
B.AR.2003.03AR138910.DQ383749  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTD VNDTRANSTP -----TSSG WDMFGGEMT NCSFNVITSI
B.AR.2003.134742FPL.2003.KY968397  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTD WNTNNTNNTN WTN-----SNEV DLKMEKIEK NCSFNVITDI
B.AR.2004.04AR151263.DQ383751  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTD YVGTDTTNS STT-----NPT IDNNMKEIK NCSFNVITSI KEK--KGYAF
B.AR.2004.04AR151516.DQ383752  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTD WKSNAKSTR -----INNS SLGTEGQIK NCSFNVITSI GDMKREYAL
B.AR.2004.145447FPL.2004.KY968398  NAWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTD LKNNNTSMT ---TTPSSRT GEFMERGIEK NCSFNVITDI RDKIORYGI
B.AR.2004.159687FPL.2004.KY968399  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTN VTNKATNTT SDATSTP-----SSSEVQMK NCSFNVITSI RDKVREYAL
B.AR.2007.DEURF07AR001.KY658686  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTD ARNTENTIT -----RQMKKIEK NCSFNVITML
B.AR.2008.DEMB08AR002.KY658683  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTN WTADNNTN---ST IRREMERIK NCSFNVITTEV
B.AR.2010.DEMB10AR006.KY658684  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTD YNSTATNPN NT-----KEIK NCSFNVITSL
B.AR.2014.DEMB14AR003.KY658685  DMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTD VPTNCTNTP -----NCTN WKIKGEGELK NCSFNVITTR
B.AR.1998.ARCH054.AY037268  NMWKNNSVQV MHEDIISLWD QSLKPCVKLL PLCVLNLCTN WRNTTNTNNA TSNNA-----TNDPITRI NNTMENGVEK NCSFNVITNM

```

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 result in 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

```

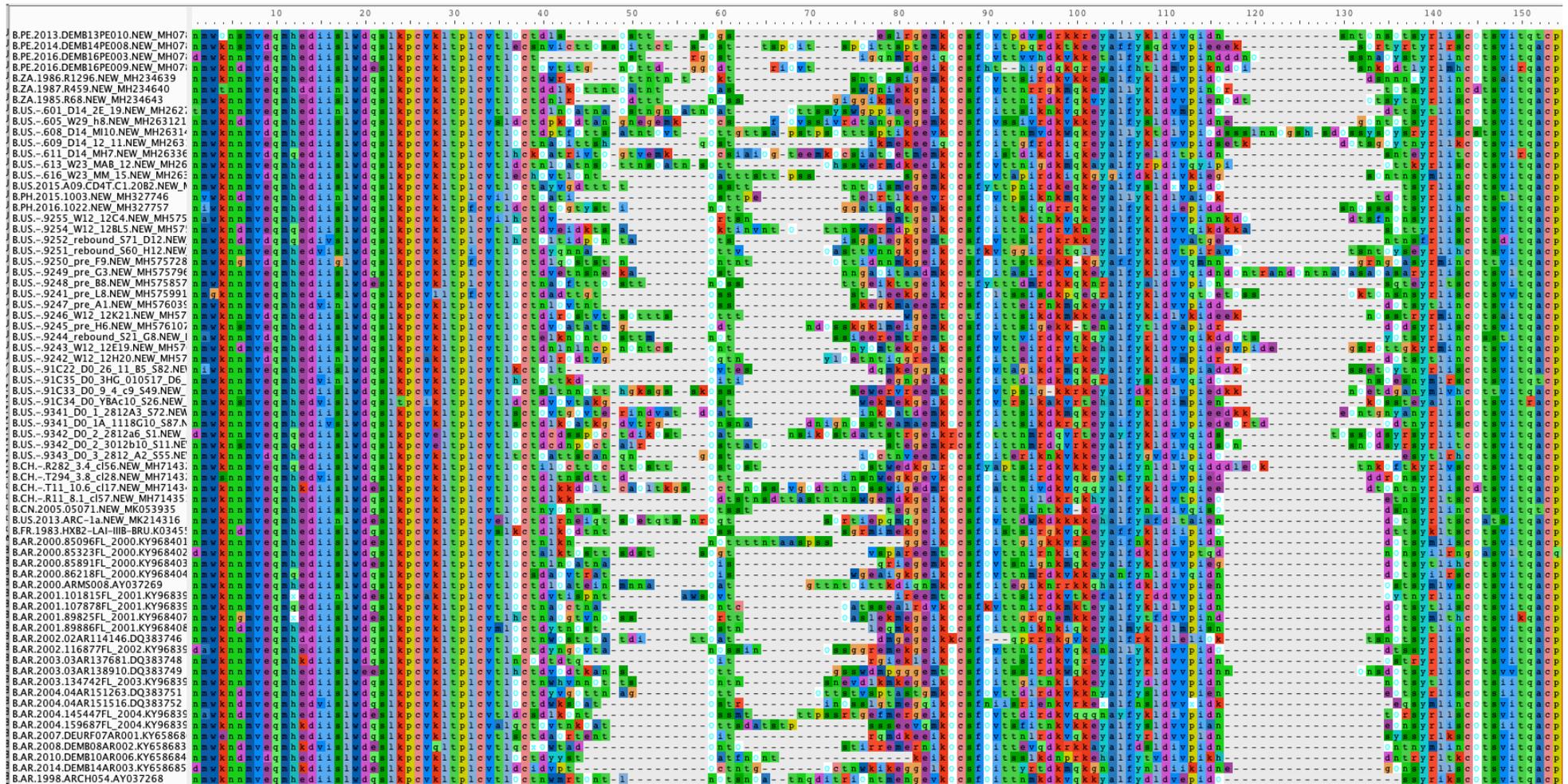
110      120      130      140      150      160      170      180      190
B.PE.2013.DEMB13PE010.NEW_MH078552  LYKLDIVQID N-----SMTNNS NTSYRLRINC TSVITQACPK ISFEPFP
B.PE.2014.DEMB14PE008.NEW_MH078553  FYSQDVVPIE EEK-----SMRT YRYRLRINC TSVITQACPK TSFEPFP
B.PE.2016.DEMB16PE003.NEW_MH078554  FYKLDIVPIN DDDNI-----SSMAN YSYRLRINC TSVITQACPK VTFEPFP
B.PE.2016.DEMB16PE009.NEW_MH078555  PHTLDMVPIK NDNI-----SKMN DTLYRLMHCN TSVIRQACPK VSFEPFP
B.ZA.1986.R1296.NEW_MH234639  FYKLDIVPID NR-----DSM NNXRYLRINC TSVITQACPK VSFEPFP
B.ZA.1987.R459.NEW_MH234640  LYKLDVVPIN NR-----T NTSYRLINC TSVITQACPK ISFEPFP
B.ZA.1985.R68.NEW_MH234643  FYKLDVVPID NNDT-----NTS YTYRYLRINC TSVITQACPK VSFEPFP
B.US.--.601.D14.2E.19.NEW_MH262735  FYKLDVMPID N-----T TTSYLRINC TSVITQACPK VTFEPFP
B.US.--.605.W29.H8.NEW_MH263121  FYSLDIVPID NE-----GNMT NTSYRLINC TSVITQACPK VSFEPFP
B.US.--.608.D14.M10.NEW_MH263144  LYKTDLVPIN DSSLSLNGS H-SNDSSYS YSYRLISCS TSVITQACPK VSFEPFP
B.US.--.609.D14.12.11.NEW_MH263194  FYKLDVVPID SETK-----DNTSGN YTYRYLRINC TSVITQACPK TTFEPFP
B.US.--.611.D14.MH7.NEW_MH263362  FYELDITPID N-----S NTEYRLITCN TSVIRQACPK VTFEPFP
B.US.--.613.W23.MAB.12.NEW_MH263520  FYRFDIVQVI PI-----N TTKYRLINC TSVITQACPK TSFEPFP
B.US.--.616.W23.MM.15.NEW_MH263630  FYRDLVQIE G-----SNM TSVYLRINC TSVITQACPK TTFEPFP
B.US.2015.A09.CD4T.C1.20B2.NEW_MH264238  FYSLDVVPIN N-----S NTSYRLINC TSVITQACPK VSFEPFP
B.PH.2015.1003.NEW_MH327746  LYKLDLVAIN K-----TD NTSYRLINC TSTLTQACPK VSFEPFP
B.PH.2016.1022.NEW_MH327757  FYKLDIEPID D-----SMNSS NTSYRLINC TSVITQACPK VSFEPFP
B.US.--.9255.W12.12C4.NEW_MH575379  FYKLDVVPIN NKDN-----DTSFP NNSYRLINC TSVITQACPK VSFEPFP
B.US.--.9254.W12.12BL5.NEW_MH575478  FYKLDVQVIK DKKA-----S MTYRYLRINC TSVITQACPK VSFEPFP
B.US.--.9252_rebound_S71.D12.NEW_MH575680  FYKLDVQVAT E-----N TNSFRLINC TSDITQACPK VSFEPFP
B.US.
B.US.
B.US.
B.US.
B.US.
B.US.
B.US.--.9247_pre_A1.NEW_MH576039  FYKLDVVPID D-----D NTSYRLINC TSVITQACPK VSFEPFP
B.US.--.9246.W12.12K21.NEW_MH576084  FYKLDLVDIK EEK-----NNS TSVYLRINC TSVITQACPK VSFEPFP
B.US.--.9245_pre_H6.NEW_MH576107  FYLTDVAPLD R-----NDSYRLINC TSVITQACPK VSFEPFP
B.US.--.9244_rebound_S21.G8.NEW_MH576152  FYRLDVQVIK DDDNTS-----Y NTSYRLINC SSTITQACPK VSFEPFP

```

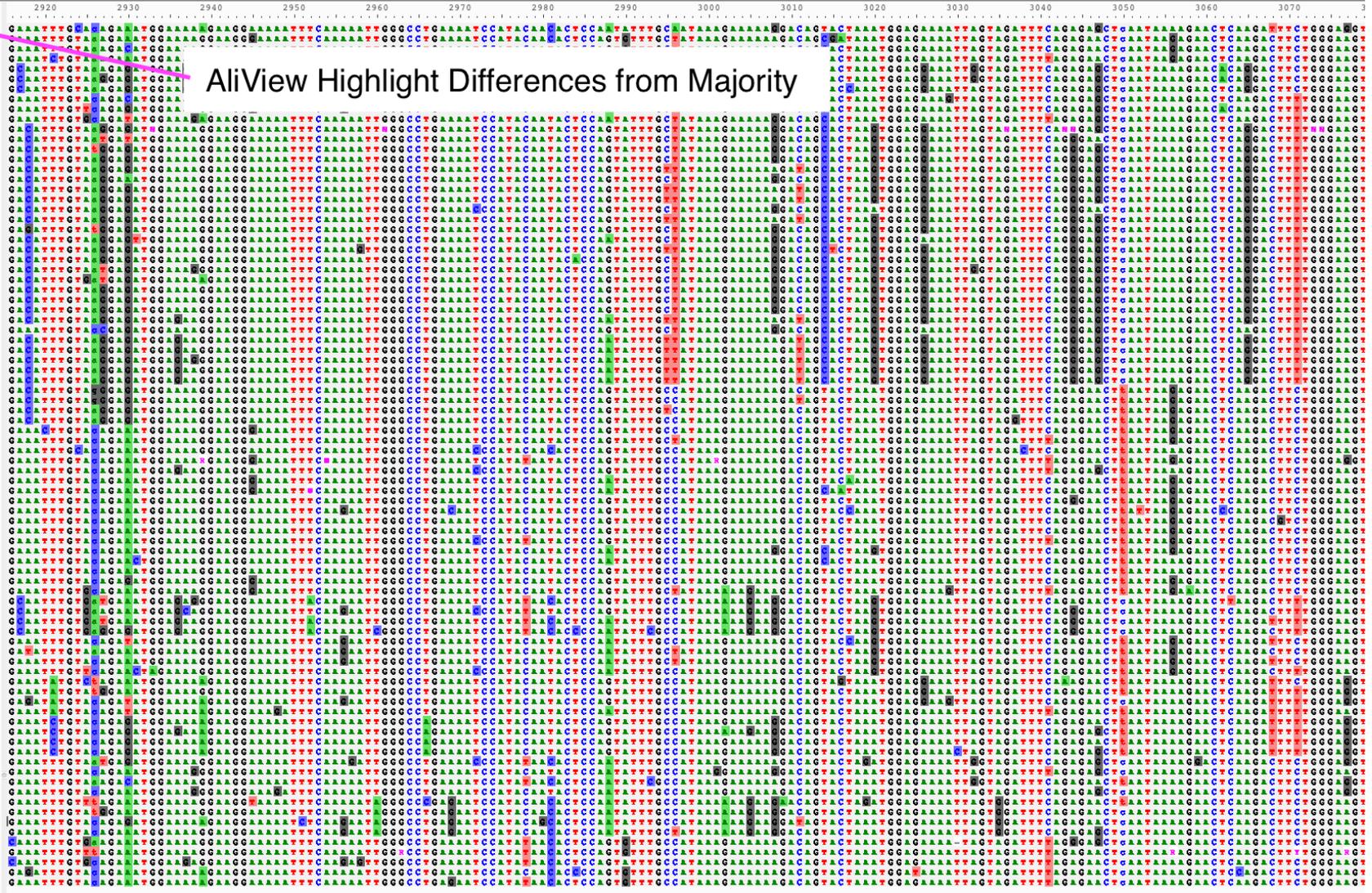
<https://www.hiv.lanl.gov/content/sequence/GLYCOSITE/glycosite.html>

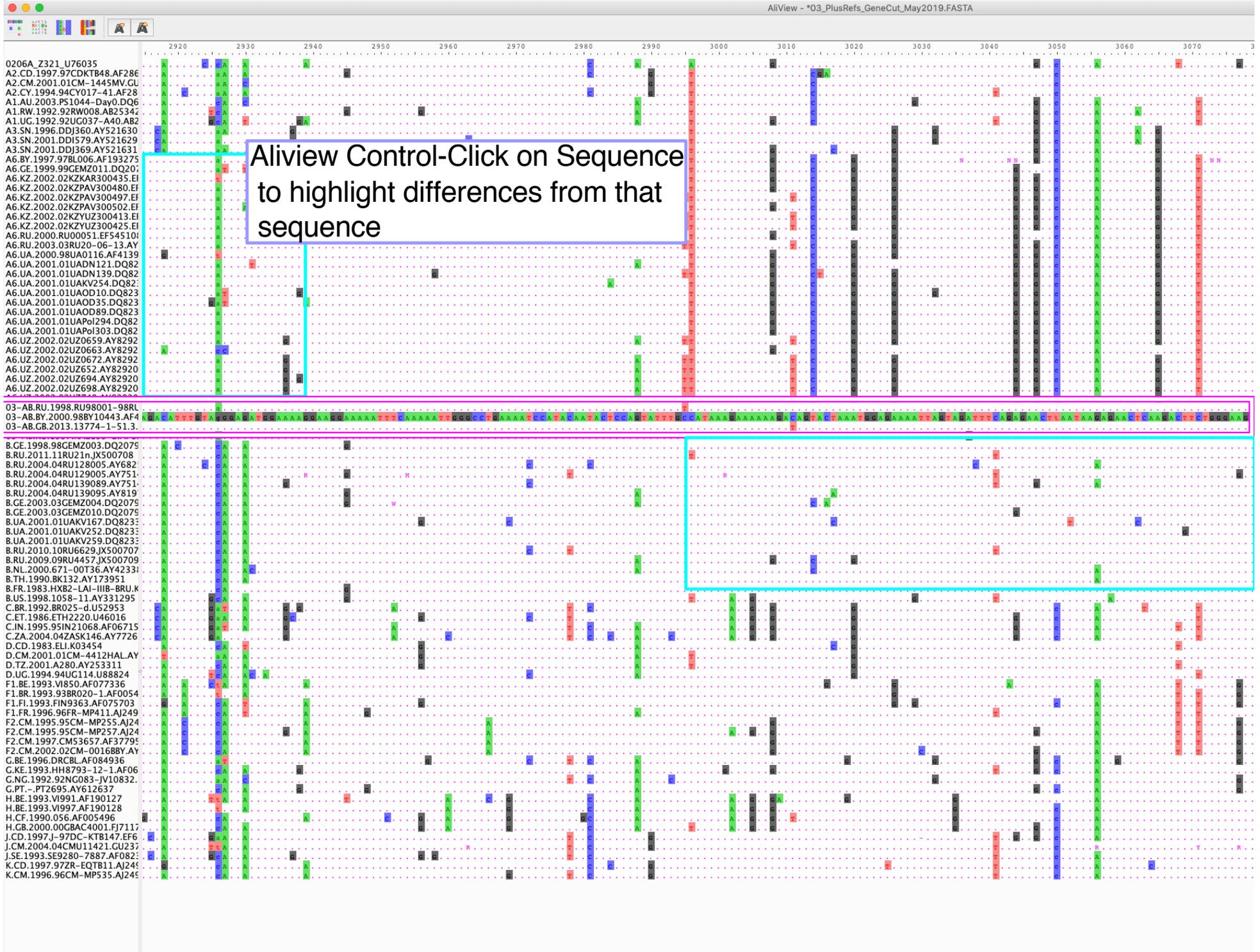


Glycosite can output FASTA file with the N of N-x-(ST) recoded as "O"

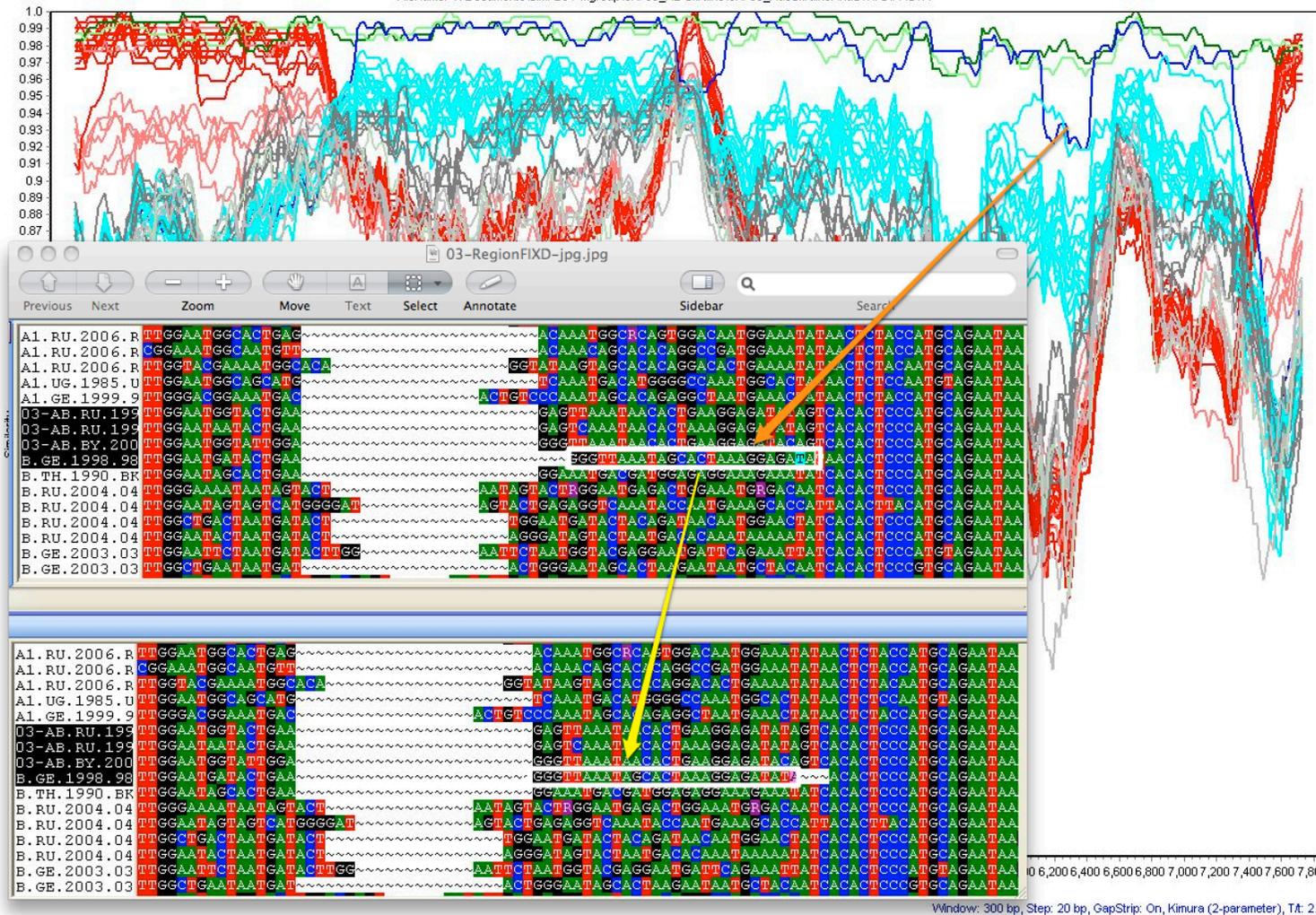


0206A_2321_U76035
A2.CD.1997.97CDK7848.AF286238
A2.CM.2001.01CM-1445MV.GU201516
A2.CV.1994.94CY017-41.AF286237
A3.SN.1996.DD360.AY521630
A3.SN.2001.D0579.AY521629
A3.SN.2001.DD369.AY521631
A1.AU.2003.P51044-DAY0.DQ676872
A1.RW.1992.92RW008.AB253421
A1.LC.1992.92UC037-440.AB253429
A6.BY.1997.97BL006.AF193275
A6.GE.1999.99GE2011.DQ207944
A6.KZ.2002.02KZAR300435.EF589042
A6.KZ.2002.02KZPAV300480.EF589043
A6.KZ.2002.02KZPAV300497.EF589039
A6.KZ.2002.02KZPAV300502.EF589044
A6.KZ.2002.02KZYU300413.EF589040
A6.RU.2001.01RU0005.1.EF545108
A6.RU.2003.03RU20-06-13.AY500393
A6.UA.2000.98UA0116.AF413987
A6.UA.2001.01UAADN121.DQ823358
A6.UA.2001.01UAADN139.DQ823357
A6.UA.2001.01UAAY294.DQ823361
A6.UA.2001.01UAOD10.DQ823365
A6.UA.2001.01UAOD35.DQ823366
A6.UA.2001.01UAOD89.DQ823367
A6.UA.2001.01UAP0294.DQ823356
A6.UA.2001.01UAP0303.DQ823360
A6.UZ.2002.02UZ0659.AY829209
A6.UZ.2002.02UZ20663.AY829210
A6.UZ.2002.02UZ0737.AY829212
A6.UZ.2002.02UZ2652.AY829203
A6.UZ.2002.02UZ2694.AY829205
A6.UZ.2002.02UZ2698.AY829206
A6.UZ.2002.02UZ740.AY829208
03-AB.BY.2001.01ABBY10443.AF414006
03-AB.GB.2013.13774-1-51.3.MF109476
03-AB.RU.1998.RU98001-98RU001.AF193277
03-AB.RU.1997.KAL153-2.AF193276
B.GE.1998.98GEMZ03.DQ207943
B.RU.2011.11RU21n.JX500708
B.RU.2004.04RU128005.AY682547
B.RU.2004.04RU129005.AY751406
B.RU.2004.04RU131905.AY751407
B.RU.2004.04RU139095.AY819715
B.GE.2003.03GEMZ004.DQ207940
B.GE.2003.03GEMZ010.DQ207942
B.UA.2001.01UAUKV167.DQ823362
B.UA.2001.01UAUKV259.DQ823363
B.UA.2001.01UAUKV259.DQ823364
B.RU.2010.10RU6629.JX500707
B.RU.2009.09RU4457.JX500709
B.NL.2000.071-00736.AY423387
B.TH.1990.BK132.AY173951
B.FR.1983.HXB2-LAI-HIB-RU.K03455
B.US.1998.1058-11.AY331295
C.BR.1992.88025-dUS2953
C.ET.1986.ETH2220.U46016
C.IN.1995.95IN21068.AF067155
C.ZA.2004.04ZAK146.AY772699
D.CD.1983.ELLK03454
D.CM.2001.01CM-4412HAL.AY371157
D.TZ.2001.A280.AY253311
D.UG.1994.94UG114.U88824
F1.BE.1993.VI850.AF077336
F1.BR.1993.93BR020-1.AF003494
F1.FI.1993.FIN9363.AF075703
F1.FR.1996.96FR-MP411.AJ249238
F2.CM.1995.95CM-MP255.AJ249236
F2.CM.1995.95CM-1390857.AJ249237
F2.CM.1997.CM53657.AJ377956
F2.CM.2002.02CM-00168BY.AY371158
G.BE.1996.DRCBL.AF084936
G.KE.1993.HH675-712-1.AF061641
G.NG.1992.92NG083-WI0832.U88826
G.PT.-PT2695.AY612637
H.BE.1993.VI991.AF190127
H.BE.1993.VI997.AF190128
H.CF.1990.05C.AB.1996
H.GB.2000.00GBAC4001.FJ711703
J.CD.1997J-97DC-KT8147.EF614151
J.CM.2004.04CMU11421.GU237072
J.SE.1997.97SE3930
K.CD.1997.97ZB-EOT811.AJ249235
K.CM.1996.96CM-MP535.AJ249239





Aliview Control-Click on Sequence to highlight differences from that sequence



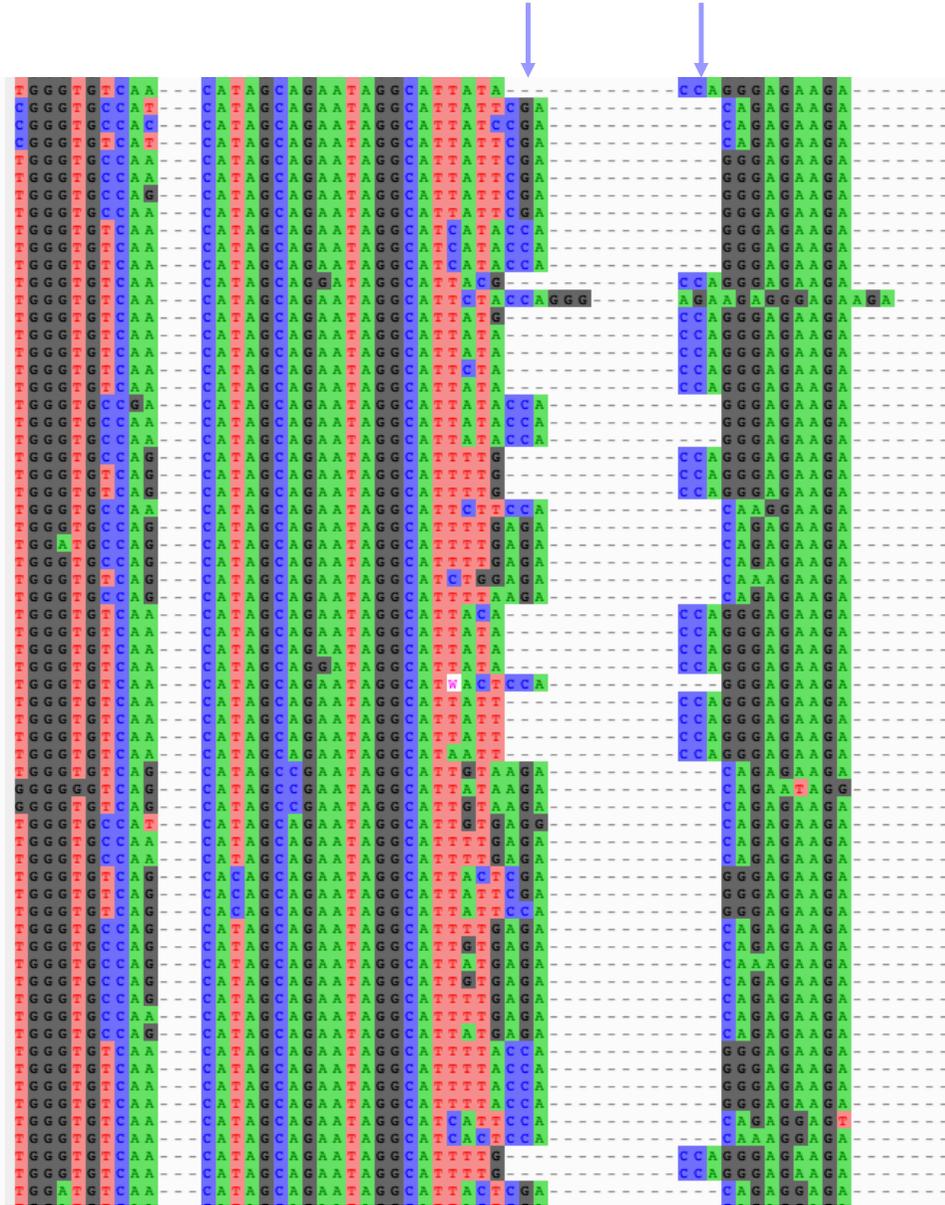
- A1.KE.1994.Q23-17.AF004885
- A1.UG.1992.92UG037.U51190
- A1.SE.1994.SET253.AF069670
- A1.UG.1998.98UG57136.AF484509
- A1.UA.2000.98UA0116.AF413987
- A1.RU.2003.03RU20_06_13.AY500393
- A1.UZ.2002.02UZ652.AY829203
- A1.UZ.2002.02UZ694.AY829205
- A1.UZ.2002.02UZ740.AY829208
- A1.UZ.2002.02UZ0683.AY829210
- A1.UZ.2002.02UZ0672.AY829212
- A1.UA.2001.01UADN139.DG823357
- A1.UA.2001.01UADN121.DG823358
- A1.UA.2001.01UAPoI293.DG823359
- A1.UA.2001.01UAPoI303.DG823360
- A1.UA.2001.01UAKV254.DG823361
- A1.UA.2001.01UAOD10.DG823365
- A1.UA.2001.01UAOD89.DG823367
- A1.KZ.2002.02KZPAV300497.EF589039
- A1.KZ.2002.02KZYUZ300413.EF589040
- A1.KZ.2002.02KZKAR300435.EF589042
- A1.RU.2007.Irkustk_5.JQ292891
- A1.RU.2002.RU01029.JQ292892
- A1.RU.2008.RU007.JQ292894
- A1.RU.2006.RU_915_10
- A1.RU.2006.RU_SP_B_049.JQ292900
- A1.UG.1985.U455-U455A.M62320
- A1.GE.1999.99GEMZ011.DQ207944
- 03-AB.RU.1998.RU98001-98RU001.AF193277
- 03-AB.BY.2000.98BY10443.AF414006
- B.GE.1998.98GEMZ003.DQ207943
- B.TH.1990.BK132.AY173951
- B.RU.2004.04RU128005.AY682547
- B.RU.2004.04RU129005.AY751406
- B.RU.2004.04RU139089.AY751407
- B.RU.2004.04RU139095.AY819715
- B.GE.2003.03GEMZ004.DQ207940
- B.GE.2003.03GEMZ010.DQ207942
- B.UA.2001.01UAKV167.DG823362
- B.UA.2001.01UAKV252.DG823363
- B.UA.2001.01UAKV259.DG823364
- B.US.1998.1058-11.AY331295
- B.US.1986.JRFL-JR-FL.U63632
- B.FR.1983.HXB2-LAI-HIB-BRU.K03455
- B.US.1990.WEAU160-GHOSH.U21135
- B.NL.2000.671-00T36.AY423387
- B.US.1983.RF-HA-T3.M17451
- D.CD.1983.NDK.M27323
- D.CD.1983.ELI.K03454
- D.CM.2001.01CM-4412HAL.AY371157
- D.CD.1984.84ZR065.U88822
- D.TZ.2001.A280.AY253311
- D.UG.1994.94UG114.U88824
- C.ZA.2004.04ZASK146.AY772699
- C.IN.1995.95IN21068.AF067155
- F1.FR.1996.96FR-MP411.AJ249238

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/SCRsoftware/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
53_01B.MY.04.04MYKL016_1_DQ366663
53_01B.MY.10.10MYKJ067_JX390612
53_01B.MY.10.10MYKJ079_JX390611
53_01B.MY.11.11FIR164_JX390610
54_01B.MY.07.07MYKLD49_EU031915
54_01B.MY.08.08MYKL044_JX390977
54_01B.MY.09.09MYSB023_JX390976
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.YNLF37_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.11.11MY1R1J704_KC522033
58_01B.MY.11.11MY1ZK731_KC522032
59_01B.CN.07.GD070126_KF011494
59_01B.CN.09.09LNA423_JX960635
59_01B.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.10.JL100010_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
69_01B.JP.05.05JP_1328M40_LC023100
```

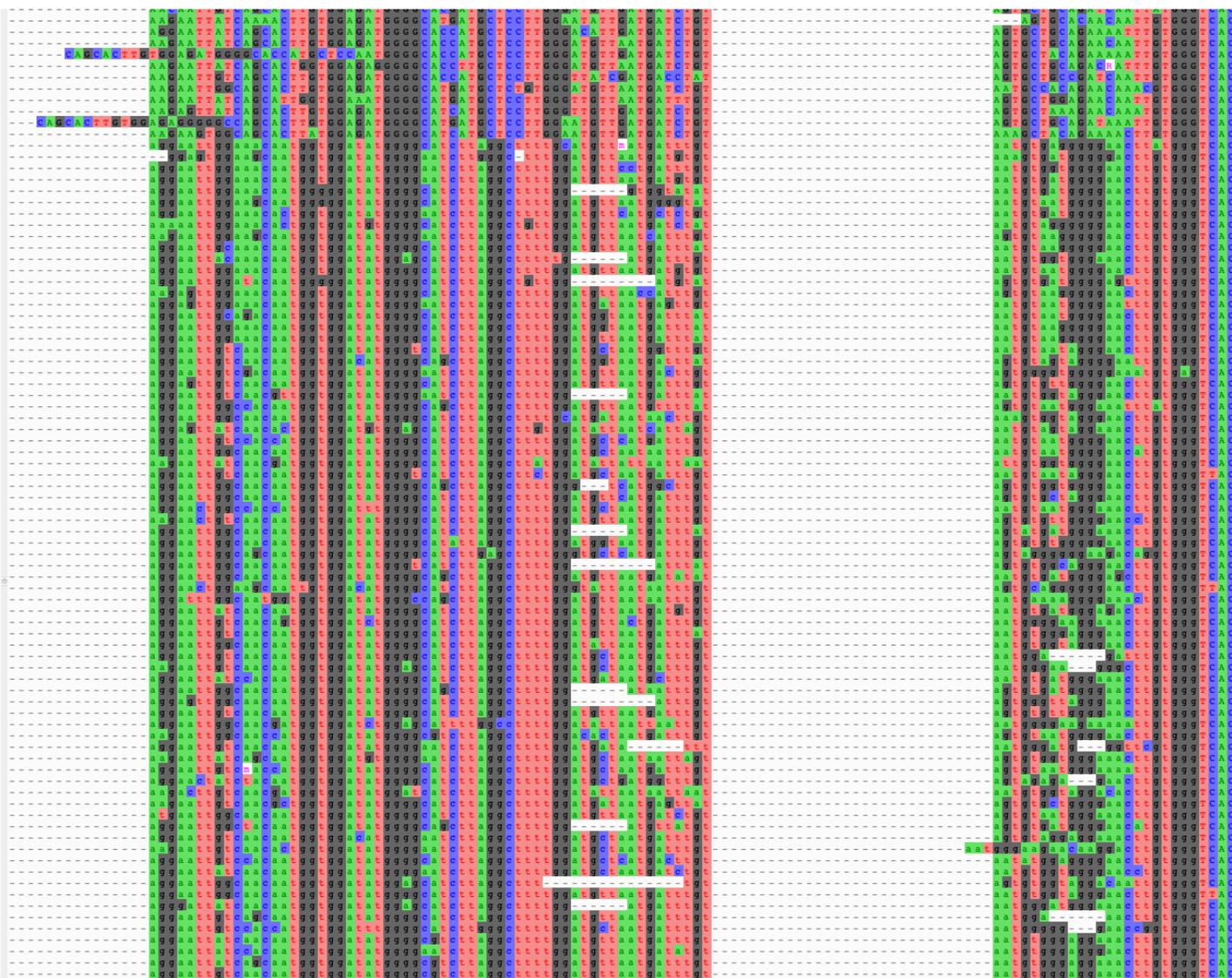


AliView has built-in alignment tools. Select a block and re-align only that block, is a very useful tool.

The screenshot displays the AliView application window. The top menu bar includes 'File', 'Edit', 'Selection', 'View', 'Align', 'Tools', 'External commands', and 'Help'. The 'Align' menu is open, showing options such as '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', and 'Delete Gap at right'. The main window shows a sequence alignment of HIV1_Consensus_Jul10_2019.fasta. The alignment is displayed as a grid of colored letters (A, C, G, T) with gaps represented by dashes. The alignment is currently in a 'realign' state, with a grey background behind the selected block. The sequence list on the left includes various identifiers such as B.US.x.CR0413T_FJ469726, B.US.x.F701_FJ469727, and B.US.x.F703_DQ886031.

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

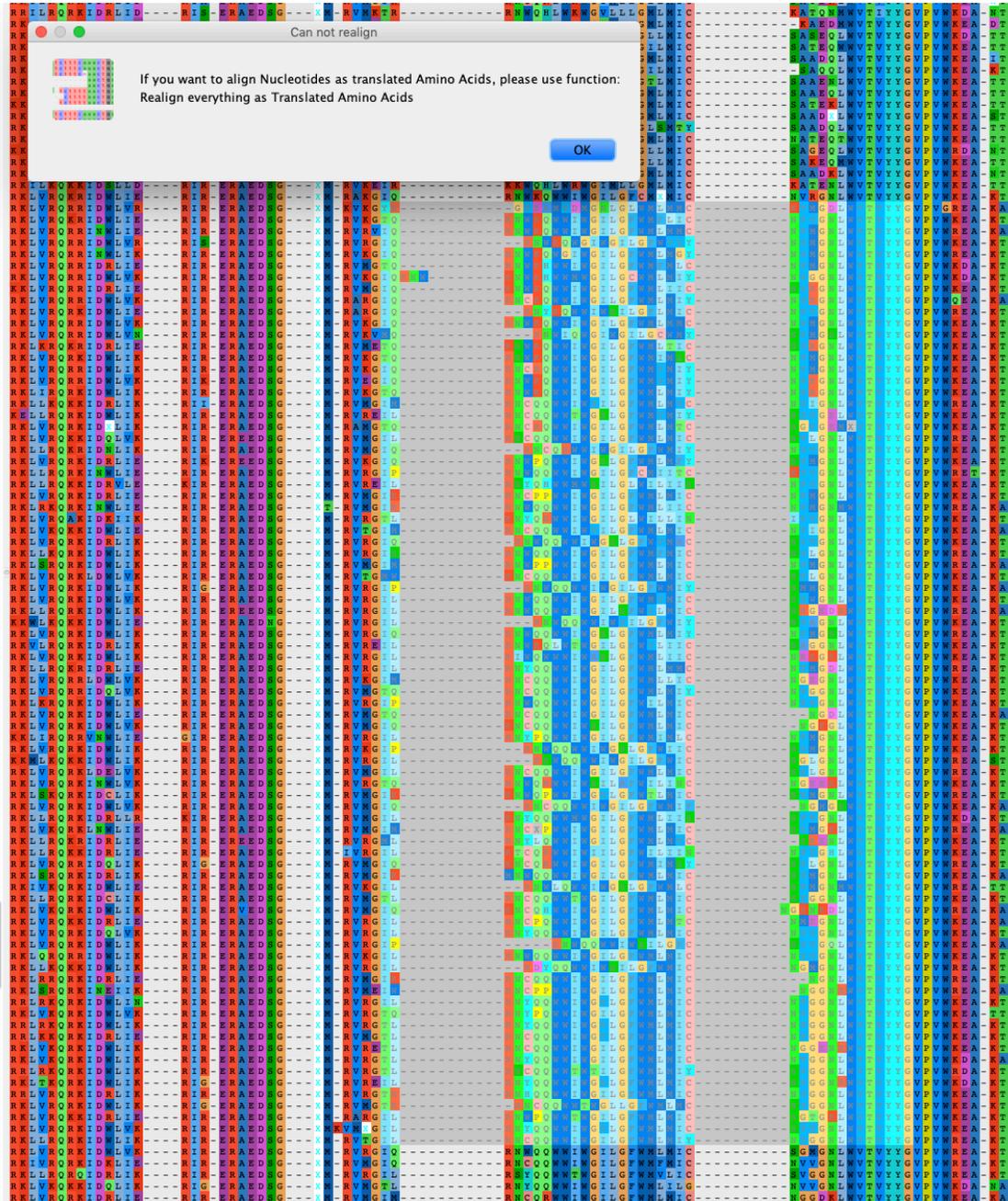
```
.ZA.09.DEMB092A02_KF109511
.ZA.85.R68_MH234643
.ZA.85.R84_FJ647145
.ZA.86.R1296_MH234639
.ZA.87.R459_MH234640
.ZA.98.TV016_Q6948656
.x.00.LA02FolC_KU168257
.x.02.LA038IEr_KU168258
.x.02.LA04GuFu_KU168259
.x.03.LA05MeAl_KU168260
.x.10.DEURF10HA002_KY658703
.AR.01.ARC4006_AY563170
.BR.02.02BR2022_IJ692434
.BR.04.04BR013_AY727522
.BR.04.04BR021_AY727523
.BR.04.04BR038_AY727524
.BR.04.04BR073_AY727525
.BR.07.DEMCO78R003_KJ140663
.BR.09.DEMCO98R036_KU74939
.BR.10.10BR_MG032_KT427678
.BR.10.10BR_MG040_KT427674
.BR.10.10BR_FE023_KT427736
.BR.10.10BR_SP033_KT427810
.BR.10.10BR_SP041_KT427806
.BR.10.10BR_SP060_KT427800
.BR.10.DEMC10BR024_KU749393
.BR.11.DEMC11BR035_KU749393
.BR.92.BR025_4_US2953
.BR.98.98BR004_AF286228
.BW.00.00BW07621_AF443088
.BW.00.00BW076820_AF443089
.BW.00.00BW087421_AF443090
.BW.00.00BW147127_AF443091
.BW.00.00BW16162_AF443092
.BW.00.00BW1686_AF443093
.BW.00.00BW17593_AF443094
.BW.00.00BW17732_AF443095
.BW.00.00BW17835_AF443096
.BW.00.00BW17936_AF443097
.BW.00.00BW18113_AF443098
.BW.00.00BW18802_AF443100
.BW.00.00BW192113_AF443101
.BW.00.00BW20361_AF443102
.BW.00.00BW20636_AF443103
.BW.00.00BW20872_AF443104
.BW.00.00BW2127214_AF443105
.BW.00.00BW22767_AF443107
.BW.00.00BW38193_AF443108
.BW.00.00BW38428_AF443109
.BW.00.00BW38713_AF443110
.BW.00.00BW3876_9_AF443111
.BW.00.00BW3886_8_AF443112
.BW.00.00BW3891_6_AF443113
.BW.00.00BW3970_2_AF443114
.BW.00.00BW50311_AF443115
.BW.00.DEMCO08W010_KY65870
.BW.00.DEMCO08W013_KY65870
.BW.00.DEMCO08W014_KY65870
.BW.96.96BW01803_AF110959
.BW.96.96BW06_AF290028
.BW.96.96BW15C02_AF110974
.BW.96.96BW17_AF110980
.BW.96.96BW032_AF443075
.BW.96.96BWMO1_5_AF443074
.BW.96.DEMC96BW001_KY65870
.BW.96.DEMC96BW002_KY65870
.BW.96.DEMC96BW006_KY65870
.BW.96.DEMC96BW007_KY65871
.BW.96.DEMC96BW009_KY65871
.BW.98.98BWMC122_AF443076
.BW.98.98BWMC134_AF443077
.BW.98.98BWMC14A3_AF443078
.BW.98.98BWM01410_AF443075
.BW.98.98BWM018D5_AF443088
.BW.98.98BWM036A5_AF443081
.BW.98.98BWM037D5_AF443082
.BW.99.99BW393212_AF443083
.BW.99.99BW46424_AF443084
.BW.99.99BW4745_AF443085
.BW.99.99BW47547_AF443086
.BW.99.99BWMC168_AF443087
.CN.06.06YNLCO285g_KC89898
.CN.07.07CNYN328_KF835515
.CN.07.07CNYN336_KF835522
.CN.09.09YNLX090sg_KC89899
```



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_K948657
B.ZA.02.TV1057_K948660
B.ZA.03.03ZAP5045M82_DQ39631
B.ZA.09.DEMB09ZAO22_KP109511
B.ZA.85.R68_MH234643
B.ZA.85.R84_FJ647145
B.ZA.86.R1296_MH234639
B.ZA.87.R459_MH234640
B.ZA.98.TV016_K948656
B.x.00.LA02FoIC_KU168257
B.x.02.LA03BIEF_KU168258
B.x.02.LA04GufU_KU168259
B.x.03.LA05MeAL_KU168260
B.x.10.DEURF10HA002_KY658703
C.AR.01.ARG4006_AY563170
C.BR.02.02BR2022_NB92494
C.BR.04.04BR013_AY727522
C.BR.04.04BR021_AY727523
C.BR.04.04BR038_AY727524
C.BR.04.04BR073_AY727525
C.BR.07.DEMC07BR003_KJ140663
C.BR.09.DEMC09BR036_KU74939
C.BR.10.10BR_MG032_KT427678
C.BR.10.10BR_MG040_KT427674
C.BR.10.10BR_PEG13_KT427676
C.BR.10.10BR_SFD33_KT427810
C.BR.10.10BR_SP041_KT427806
C.BR.10.10BR_SP060_KT427800
C.BR.10.DEMC10BR024_KU749399
C.BR.11.DEMC11BR035_KU749399
C.BR.92.BR025_d_US2953
C.BR.98.98BR004_AF286228
C.BW.00.00BW07621_AF443088
C.BW.00.00BW076820_AF443089
C.BW.00.00BW087421_AF443090
C.BW.00.00BW147127_AF443091
C.BW.00.00BW16162_AF443092
C.BW.00.00BW1686_AF443093
C.BW.00.00BW17593_AF443094
C.BW.00.00BW17732_AF443095
C.BW.00.00BW17835_AF443096
C.BW.00.00BW17956_AF443097
C.BW.00.00BW18113_AF443098
C.BW.00.00BW18802_AF443100
C.BW.00.00BW192113_AF443101
C.BW.00.00BW20361_AF443102
C.BW.00.00BW20636_AF443103
C.BW.00.00BW20872_AF443104
C.BW.00.00BW212714_AF443110
C.BW.00.00BW22767_AF443107
C.BW.00.00BW38193_AF443108
C.BW.00.00BW38428_AF443109
C.BW.00.00BW38713_AF443110
C.BW.00.00BW3876_9_AF443111
C.BW.00.00BW3886_8_AF443112
C.BW.00.00BW3891_6_AF443113
C.BW.00.00BW3970_2_AF443114
C.BW.00.00BW5031_1_AF443115
C.BW.00.DEMC00BW010_KY65870
C.BW.00.DEMC00BW013_KY65870
C.BW.00.DEMC00BW014_KY65870
C.BW.96.96BW1803_AF110959
C.BW.96.96BW06_AF290028
C.BW.96.96BW15C02_AF110974
C.BW.96.96BW17_AF110980
C.BW.96.96BW032_AF443075
C.BW.96.96BWM015_AF443074
C.BW.96.DEMC96BW001_KY65870
C.BW.96.DEMC96BW002_KY65870
C.BW.96.DEMC96BW006_KY65870
C.BW.96.DEMC96BW007_KY65871
C.BW.96.DEMC96BW009_KY65871
C.BW.98.98BWM0122_AF443076
C.BW.98.98BWM0134_AF443077
C.BW.98.98BWM01443_AF443078
C.BW.98.98BWM01410_AF443079
C.BW.98.98BWM018D5_AF443081
C.BW.98.98BWM036A5_AF443081
C.BW.98.98BWM037D5_AF443082
C.BW.99.99BW393212_AF443083
C.BW.99.99BW46424_AF443084
C.BW.99.99BW4745_AF443085
C.BW.99.99BW47547_AF443086
C.BW.99.99BWM168_AF443087
C.CN.06.06YNL1C0289g_KC89898
C.CN.07.07CNYN328_KF835515
C.CN.07.07CNYN336_KF835522
C.CN.09.09YNLX0909g_KC89899
C.CN.09.09YNLX1339g_KC89899
C.CN.09.DH21_KF250403
C.CN.09.DH22_KF250404
C.CN.10.YNFL19_KC870038
C.CN.98.YNRL9840_AY967806
C.CY.05.CY040_FJ388901
C.CY.05.CY069_FJ388913
C.CY.06.CY166_FJ388948
C.CY.06.CY176_FJ388952
C.CY.07.CY187_FJ683740
C.CY.07.CY203_FJ683755
C.CY.08.CY219_FJ683768
C.CY.08.CY221_FJ683770
C.CY.09.CY260_FJ683803
```



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 or seq-info@lanl.gov for advice or help.