

# HIV Recombination Overview

**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 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 and funding. Three navigation buttons are visible: "SEQUENCE DATABASE ▶", "IMMUNOLOGY DATABASE ▶", and "OTHER VIRUSES ▶". A "News:" section contains a link to "CATNAP: two new features" and a date of "20 February 2019". At the bottom, there is 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



# Workshop Topics

## HIV Intersubtype Recombination and Unique Circulating Recombinant Forms

*General introduction:*

*HIV, like all retroviruses, is effectively diploid, packages two copies of the viral genome per virion.*

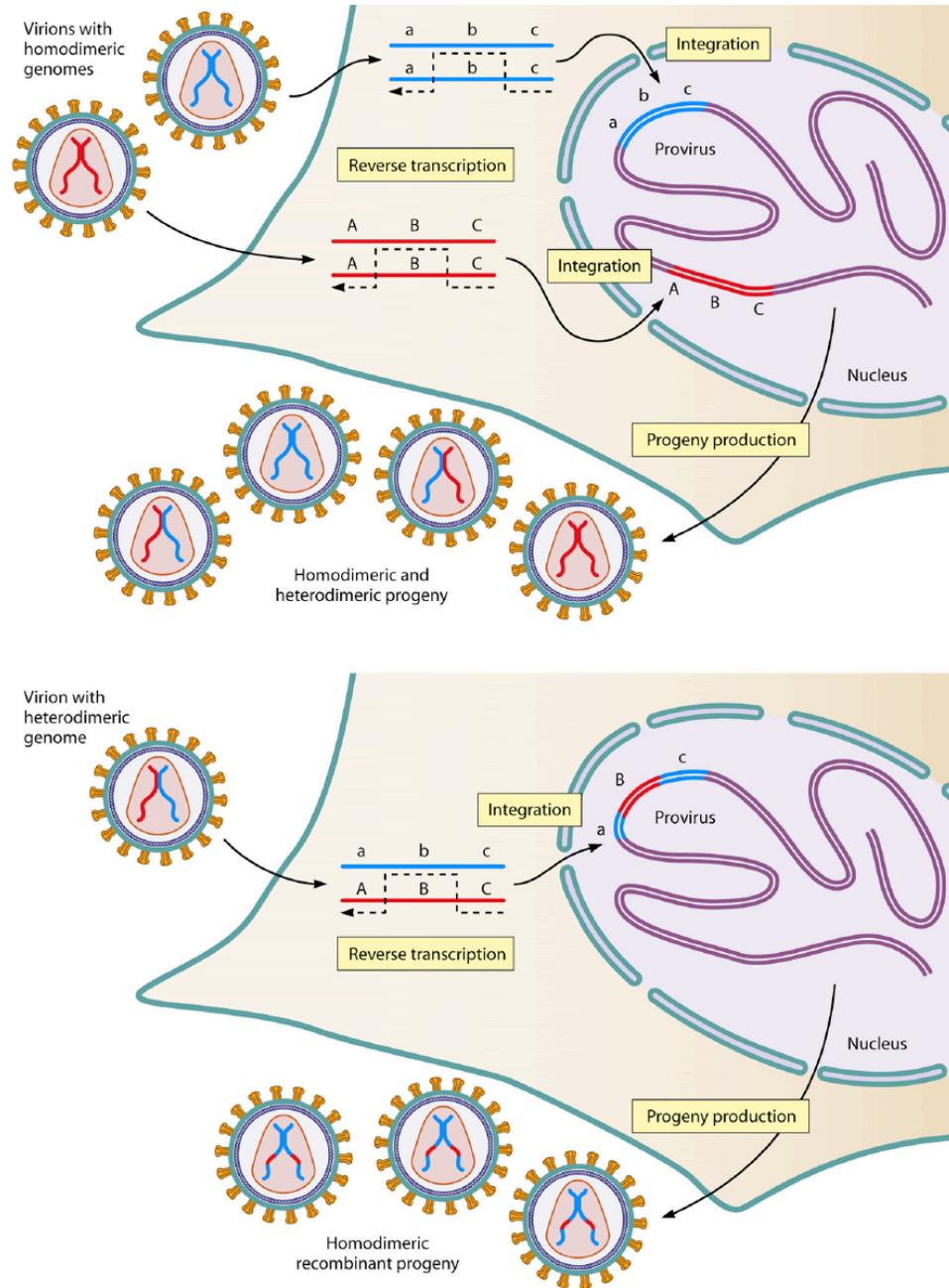
*Mechanism of Recombination:*

*Template switching during reverse transcription, no DNA damage-repair enzymes needed.*

*Tools for detecting Intersubtype Recombination:*

*Overview of HIV-1 M group subtypes and CRFs:*

*Examples of what recombinants “look like”:*



Recombinant viruses can be formed when one cell is infected with 2 viruses.

Distance or diversity between the two viruses can be large (intersubtype) or small (near identity intrapatent).

Recombination occurs by template switching during reverse transcription of heterodimeric viruses

FIG. 8. Requirements for generating recombinant genomes. (A) Coinfection with two genetically distinct viruses does not yield recombinants. However, a producer cell must be coinfecting with two genetically distinct viruses (shown here as viral particles with two blue or two red RNAs) to produce viral particles with heterodimeric gRNAs. (B) Recombination is observable in cells infected with heterodimeric virions (particle containing one red and one blue RNA strand). Template switching during reverse transcription can generate a recombinant provirus.

# Virus Recombination Detection Tools:

**RIP:** HIV-databases <https://www.hiv.lanl.gov/content/sequence/RIP/RIP.html>

Pros: Adjustable window size, prebuilt test set, allows user input test set, adjustable significance threshold, contains consensus for each subtype.

Cons: Does not output numerical breakpoint locations.

**jpHMMer:** Gobics [http://jphmm.gobics.de/submission\\_hiv.html](http://jphmm.gobics.de/submission_hiv.html)

Pros: Statistical support of precise breakpoints, outputs table of breakpoints.

Cons: Does not have HMM models for CRFs, weak models for rare subtypes, does not include subsubtypes (A3, A4, A6, F2), window size not adjustable.

**REGA Genotyping:** SANBI, REGA <http://regatools.med.kuleuven.be/typing/v1/subtyping.html>

Pros: Accepts input of many sequences, phylogeny as well as other methods, CRFs updated reasonably often.

Cons: Window size not adjustable,

**RDP3, RDP4:** <http://web.cbio.uct.ac.za/~darren/rdp.html>

**NCBI Genotyping:** NCBI <https://www.ncbi.nlm.nih.gov/projects/genotyping/formpagex.cgi>

**SimPlot:** Stuart Ray <https://sray.med.som.jhmi.edu/SCRsoftware/simplot/>

**Highlighter:** HIV Databases

[https://www.hiv.lanl.gov/content/sequence/HIGHLIGHT/highlighter\\_top.html](https://www.hiv.lanl.gov/content/sequence/HIGHLIGHT/highlighter_top.html)

# Many more recombination detection tools:

<http://bioinf.man.ac.uk/robertson/recombination/programs.shtml>

## Links to Recombinant Sequence Analysis/Detection Programs

Welcome to the comprehensive list of recombination analysis software maintained by the [Robertson Lab](#).

Filter list by method or algorithm

Methods

Algorithms

### 3seq

- **Description** 3SEQ is a software program for identifying mosaic structure or recombination in nucleotide sequence data. 3SEQ takes as input a data set with a minimum of three aligned sequences, and it tests whether any sequence in the data set is a recombinant or mosaic of any other two sequences in the data set.
- **Implemented algorithms**
  - Recombination detection using hyper-geometric random walks
- **Reference** Boni, M.F., Posada, D. & Feldman, M.W. (2007) [[PubMed](#)][[doi](#)]

### 4SIS

- **Description** Two types of informative sites were distinguished, corresponding to the clustering of the putative recombinant with either of the parental representatives. The optimal breakpoint was located by maximizing a chi-square value.
- **Methods**
  - Phylogenetic methods
- **Additional notes** Temporarily unavailable.
- **Reference** van Cuyck, H., Fan, J., Robertson, D.L. & Roques, P. (2005) [[PubMed](#)][[doi](#)]

### BARCE

- **Description** BARCE is a C++ program for detecting recombination breakpoints in four-sequence alignments using hidden Markov models, Bayesian principles and Markov chain Monte Carlo sampling.
- **Methods**
  - Phylogenetic methods
- **Implemented algorithms**
  - Phylogenetic hidden Markov model with Bayesian inference
  - Hidden Markov models
- **Additional notes** Implemented in TOPALI
- **Reference** McGuire, G., & Husmeier, D. (2003) [[PubMed](#)][[doi](#)]

### Bellerophon

- **Description** Bellerophon is a program for detecting chimeric sequences in multiple sequence datasets by an adaption of partial treeing analysis.
- **Implemented algorithms**
  - Distance-matrix calculation across breakpoints
- **Reference** Huber, T., Faulkner, G. & Hugenholz, P. (2004) [[PubMed](#)][[doi](#)]

### BLAST Genotyping

- **Description** This tool helps identify the genotype of a viral sequence. A window is slid along the query sequence and each window is compared by BLAST to each of the reference sequences for a particular virus.
- **Implemented algorithms**
  - Similarity/distance plot

### cBrother

- **Description** cBrother is software for inferring recombination when recombination is rare. This is a C version of the code originally written in Java available in DualBrothers.
- **Methods**
  - Phylogenetic methods
- **Implemented algorithms**
  - Bayesian multiple change-point modelling
- **Additional notes** Also implemented in DualBrothers.
- **Reference** Fang, F. (2005) [[PubMed](#)][[doi](#)]

### DnaSP

- **Description** DnaSP, DNA Sequence Polymorphism, is a software package for the analysis of nucleotide polymorphism from aligned DNA sequence data. DnaSP can estimate several measures of DNA sequence variation within and between populations (in noncoding, synonymous or nonsynonymous sites, or in various sorts of codon positions), as well as linkage disequilibrium, recombination, gene flow and gene conversion parameters.
- **Methods**
  - Population genetics based
- **Reference** Librado, P. & Rozas, J. (2009) [[PubMed](#)][[doi](#)]

### DualBrothers

- **Description** DualBrothers is a recombination detection software based on the dual Multiple Change-Point (MCP) model. This model allows for changes in topology and evolutionary rates across sites in a multiple sequence alignment.
- **Methods**
  - Phylogenetic methods
- **Implemented algorithms**
  - Bayesian multiple change-point modelling
- **Additional notes** Also implemented in cBrother.
- **Reference** Minin, V., Dorman, K., Fang, F. and Suchard, M. (2005) [[PubMed](#)][[doi](#)]

### Frag-dist

- **Description** This program use the principle of "Recombinant has high similarity with its parental sequence" to visualize the location of breakpoints and gives out the potential parental sequences.
- **Methods**

# How to decide “which tool”?

The answer depends on many factors such as number of sequences which need to be screened, short or long sequences, diversity in the local population, etc...

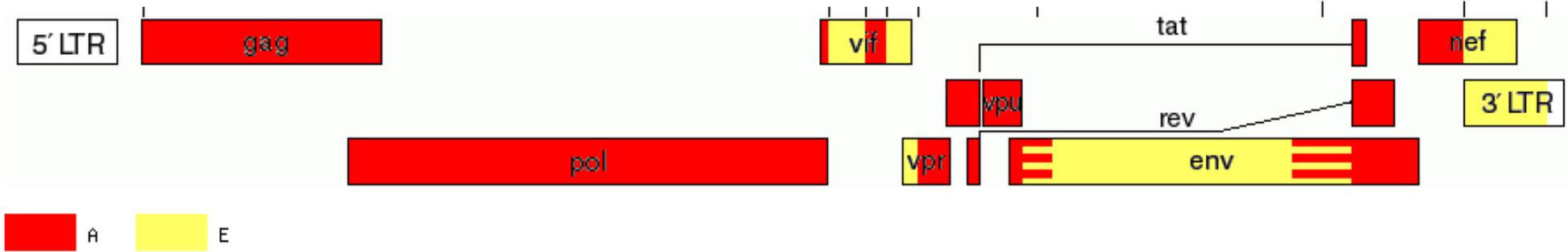
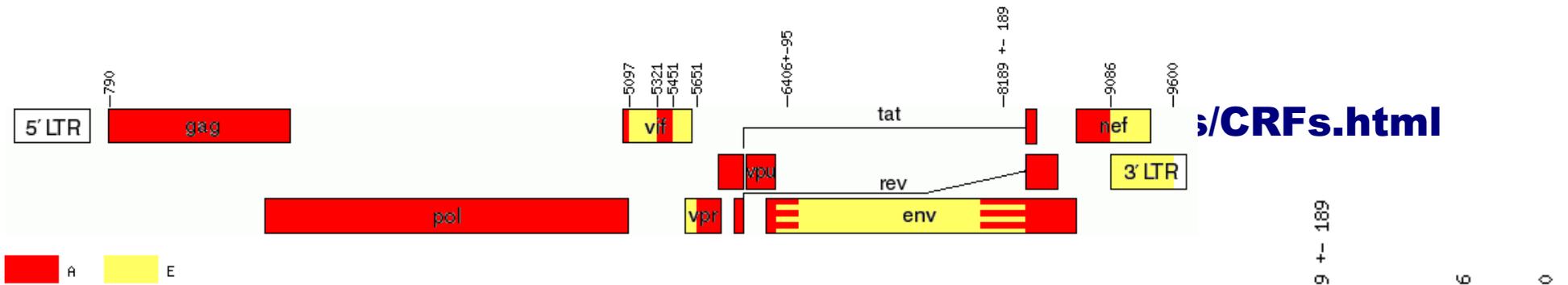
No single tool does everything very well.

There are trade-offs in speed vs accuracy, rate of false-positive vs false negative results, etc...

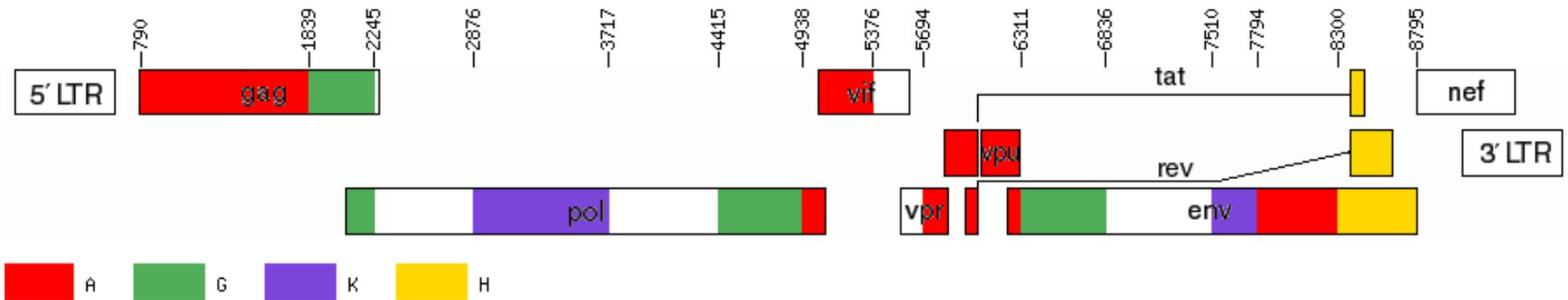
Very fast methods can be simple, such as BLAST of each sequence against a small subtype reference set local database. Parsing and interpreting the results may be the most difficult part.

# HIV-1 subtypes and CRFs

- Subtypes and subsubtypes: A1 – A6, B, C, D, F1 – F2, G, H, J, K
- Circulating Recombinant Forms: CRF01\_AE – CRF98\_06B
- Both require a minimum of 2 complete genomes and at least one more partial genome with sequences from regions that can confirm the structure of the first 2.

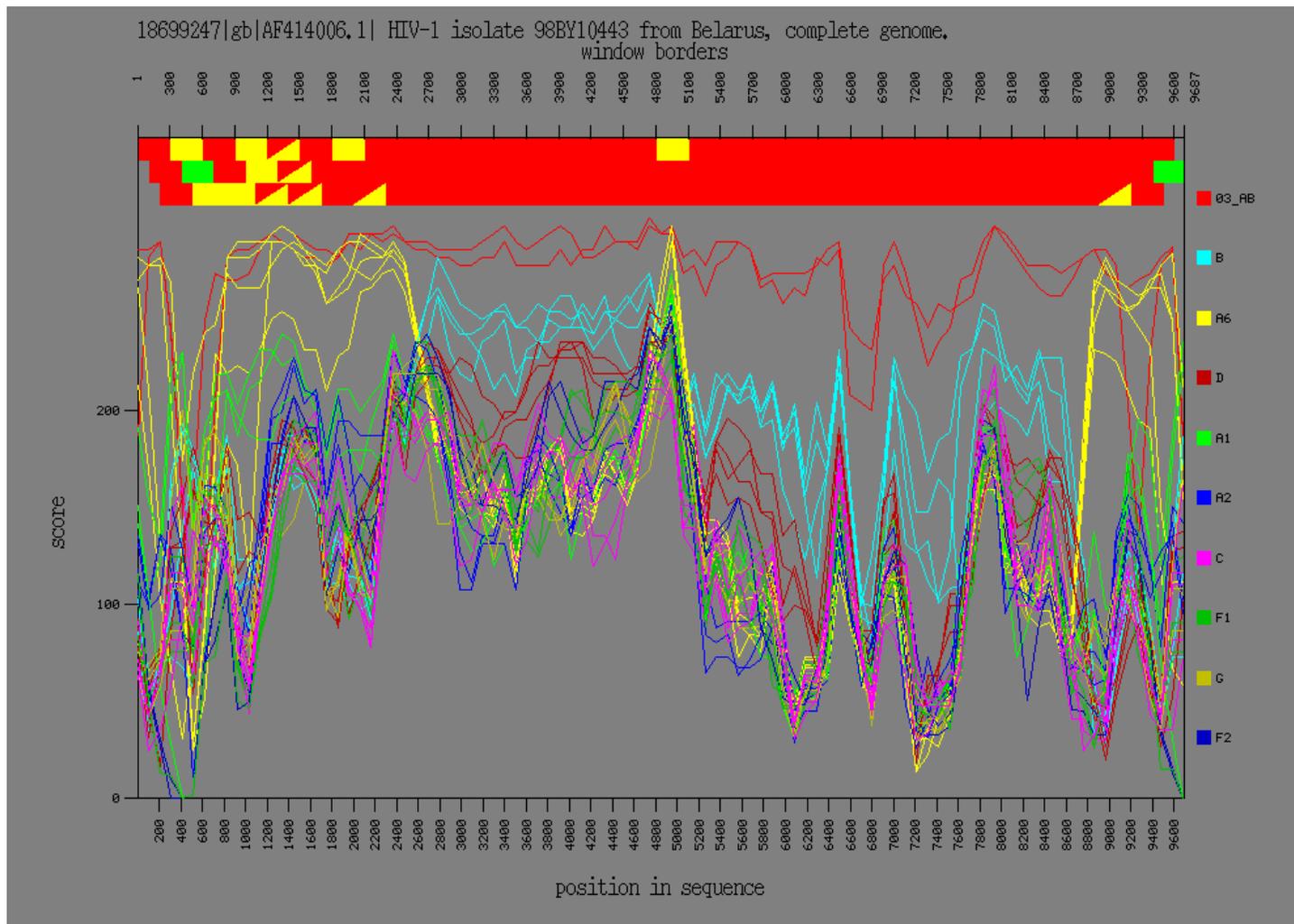


### CRF04\_cpx



# The NCBI “Window BLAST” genotyping tool

CRF03\_AB



CRF03\_AB is recombinant between **A6**, the subtype of A found in the former Soviet Union region and **B**.

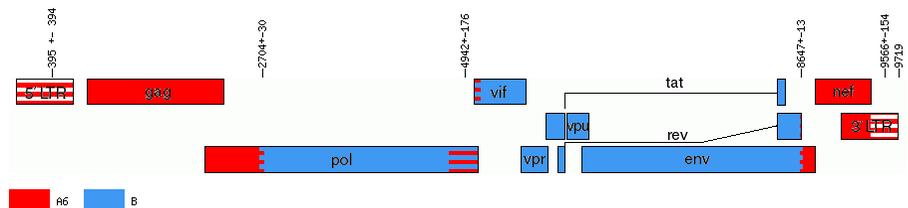
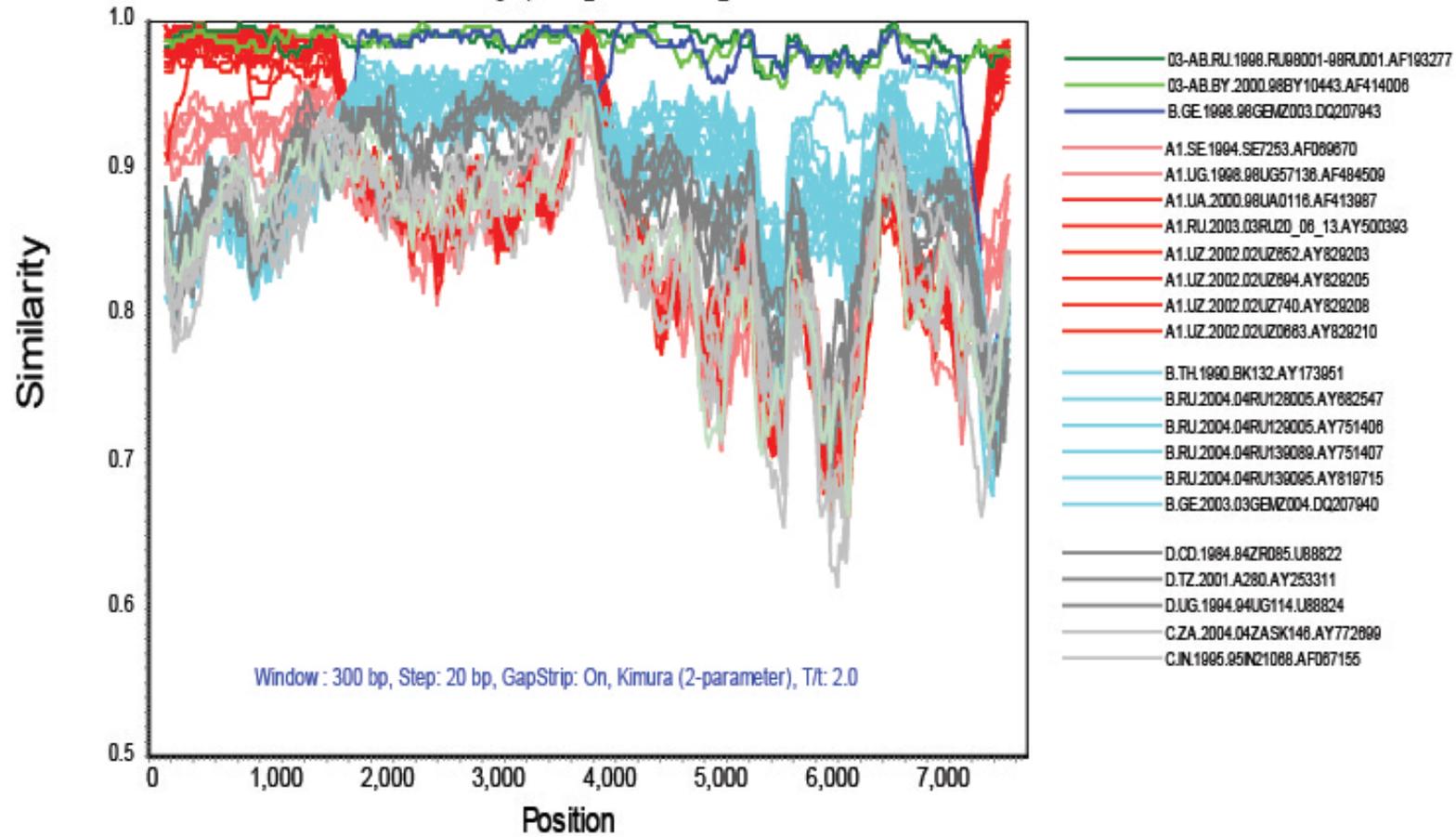
In this plot 2 reference genomes of **CRF03\_AB** are included, along with **A1**, **A2**, **B**, **C**, **F1**, **F2**, and **G** genomes.

Window size 150 bases and step of 50 bases

# CRF03\_AB SimPlot

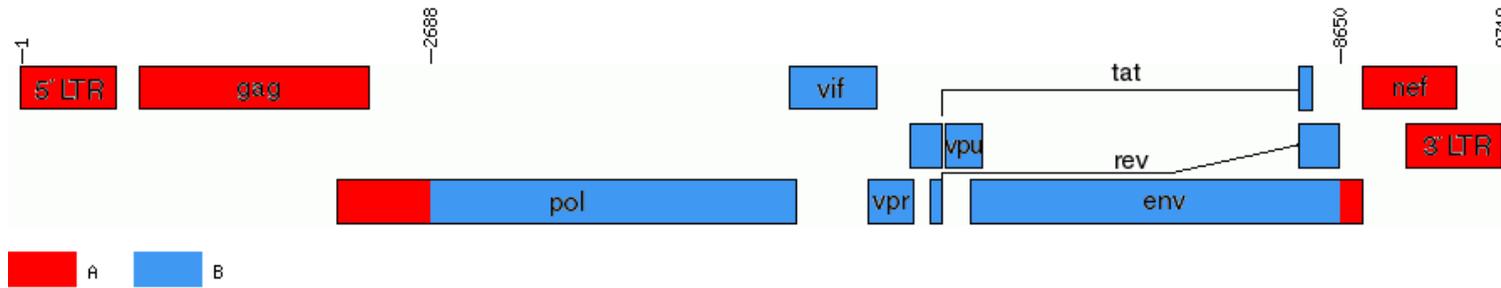
SimPlot - Query: 03-AB.RU.1997.KAL153-2.AF193276

FileName: Y:\Documents\SIMPLOT-Mgroup\CRF03\_AB Ukraine\CRF03\_PlusUkraineFixdSTRPD.FASTA

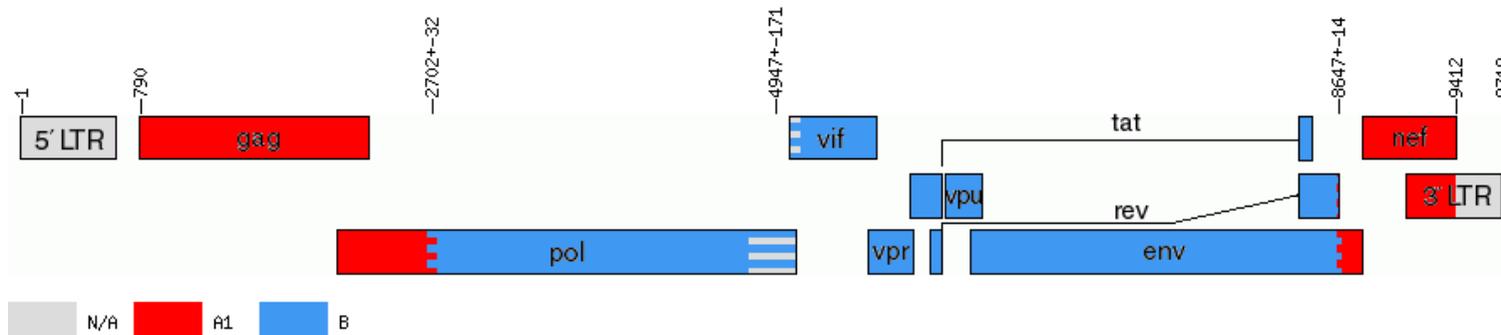


Map from jpHMMer

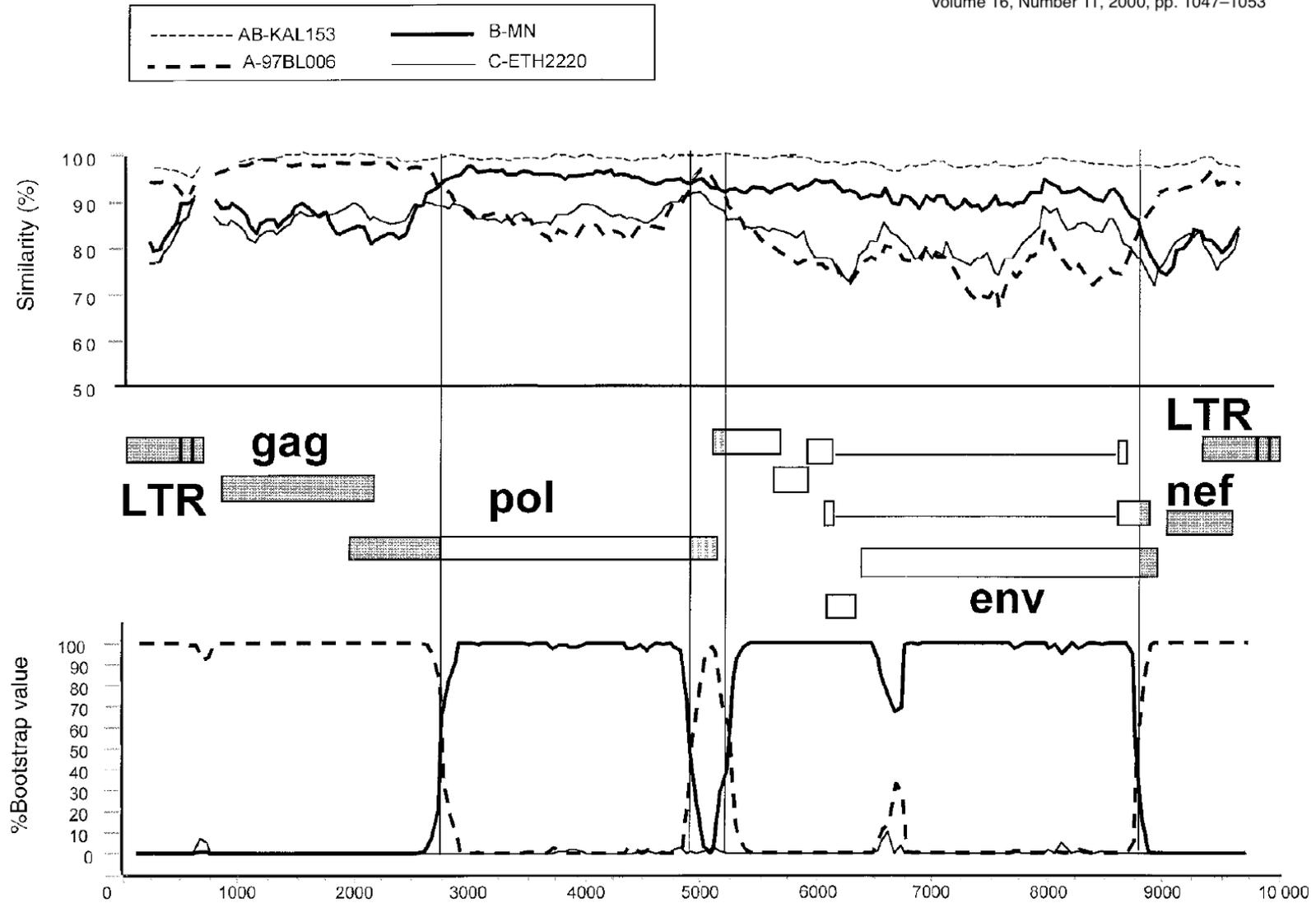
# CRF03\_AB Maps



From publication describing CRF03\_AB



From jpHMMer analysis at Gobics



**FIG. 2.** Recombinant structure of the Kaliningrad IDU-associated HIV-1 strain. Similarity (*top*) and bootscanning (*bottom*) analyses were used to map the complete genome sequence of the AB-98RU001 HIV-1 isolate. In both analyses a window of 400 bases and an increment of 50 bases were used. Gap regions in the alignment were excluded from the analyses. The Kimura 2-parameter model with 100 replicates was used as the algorithm for bootscanning. Subtype C isolate ETH2220 was used as an outgroup. Similarity and bootstrap value are shown on the y axis and positions on the full genome alignment are shown on the x axis. Vertical lines indicate the recombination points. The subtype origin of the different genome regions is indicated (*middle*) as gray (subtype A) and white (subtype B) regions. The small region in the *pol-vif* region, which seems to be derived from subtype A, could not be reliably verified by separate phylogenetic analysis and is therefore shown as uncertain (striped).



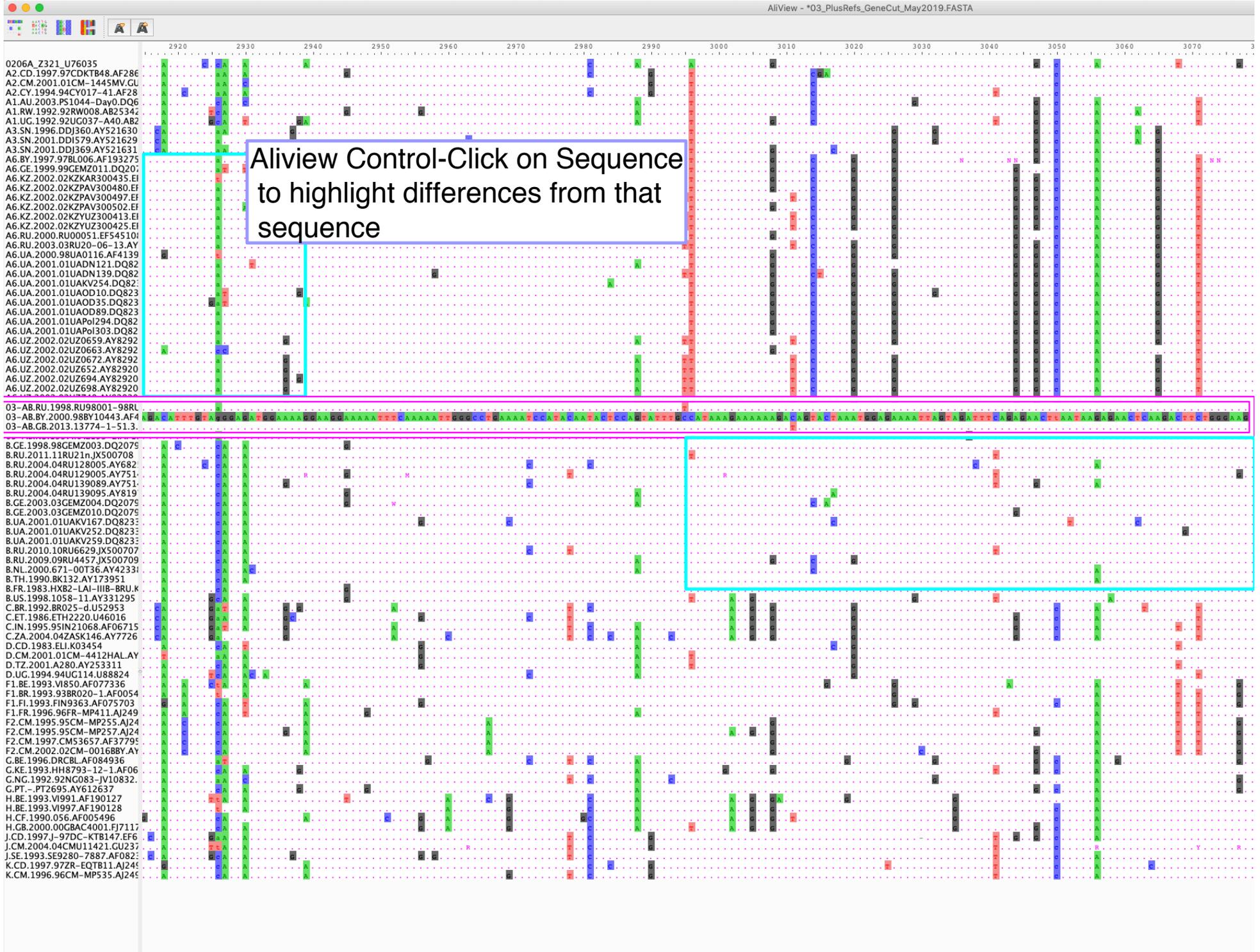
AliView (free for Mac) or BioEdit (free for Windows) multiple sequence alignment editor view of CRF03\_AB plus reference genomes.

<http://ormbunkar.se/aliview/>

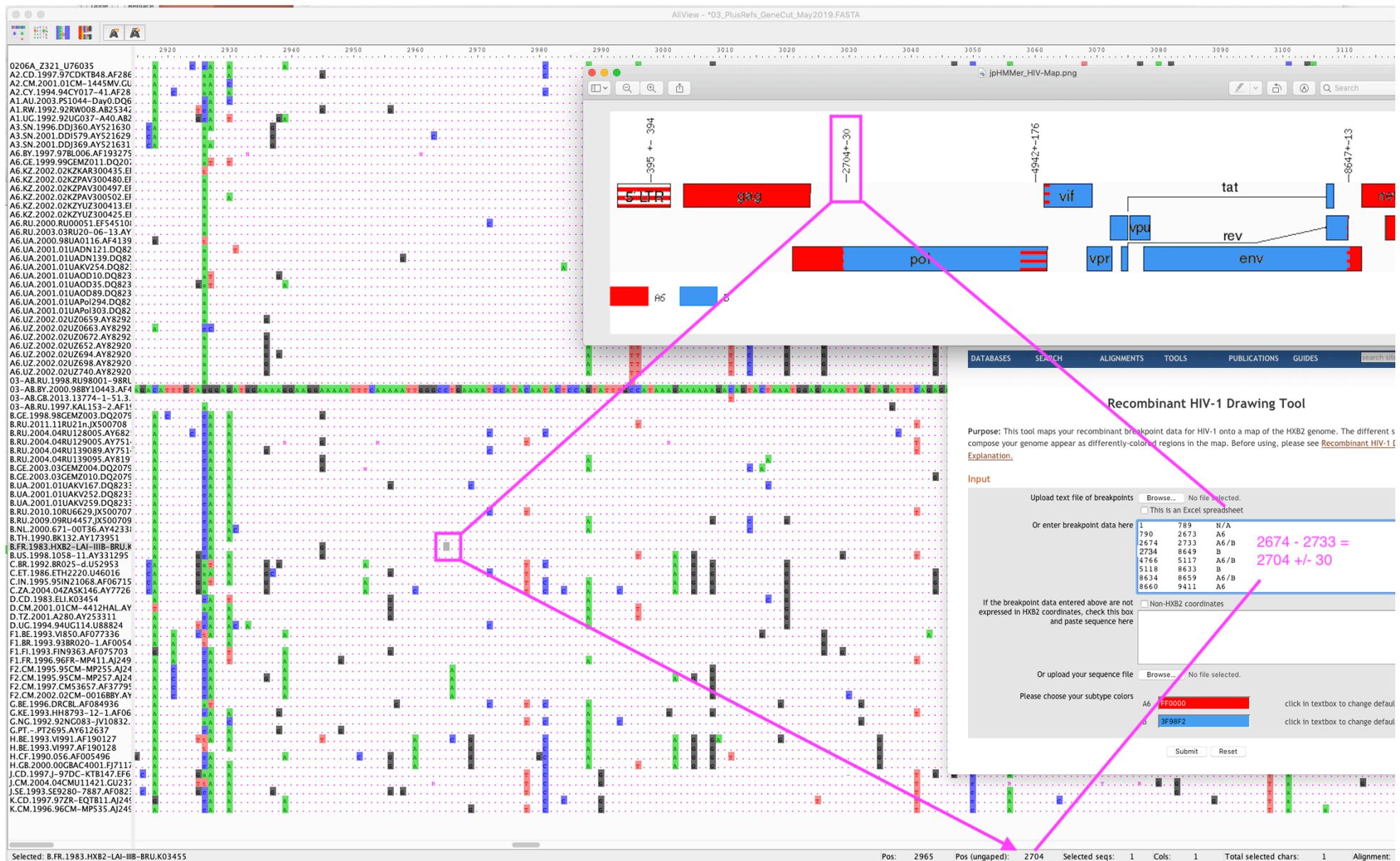
<http://www.mbio.ncsu.edu/BioEdit/bioedit.html>

0206A\_2321\_U76035  
A2.CD.1997.97CDK7848.AF286238  
A2.CM.2001.01CM-1445MV.GU201516  
A2.CV.1994.94CV017-41.AF286237  
A3.SN.1996.DD360.AY521630  
A3.SN.2001.DD0579.AY521629  
A3.SN.2001.DD369.AY521631  
A1.AU.2003.P51044-DAY0.DQ676872  
A1.RW.1992.92RW008.AB253421  
A1.LC.1992.92LC037-440.AB253429  
A6.BY.1997.97BY006.AF193275  
A6.GE.1999.99GEM2011.DQ207944  
A6.KZ.2002.02KZKAR300435.EF589042  
A6.KZ.2002.02KZPVA300480.EF589043  
A6.KZ.2002.02KZPVA300497.EF589039  
A6.KZ.2002.02KZPVA300502.EF589044  
A6.KZ.2002.02KZYU300413.EF589040  
A6.RU.2001.01RU0005.1.EF451108  
A6.RU.2003.03RU20-06-13.AY500393  
A6.UA.2000.98UA0116.AF413987  
A6.UA.2001.01UADN121.DQ823358  
A6.UA.2001.01UADN139.DQ823357  
A6.UA.2001.01UAKV294.DQ823361  
A6.UA.2001.01UAD010.DQ823365  
A6.UA.2001.01UAD035.DQ823366  
A6.UA.2001.01UAD089.DQ823367  
A6.UA.2001.01UAP0294.DQ823356  
A6.UA.2001.01UAP0303.DQ823360  
A6.UZ.2002.02UZ02659.AY829209  
A6.UZ.2002.02UZ20663.AY829210  
A6.UZ.2002.02UZ073.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.1998.98GEM2003.DQ207943  
B.RU.2011.11RU21n.JX500708  
B.RU.2004.04RU128005.AY682547  
B.RU.2004.04RU129005.AY751406  
B.RU.2004.04RU131905.DQ823357  
B.RU.2004.04RU139095.AY819715  
B.GE.2003.03GEM2004.DQ207940  
B.GE.2003.03GEM2010.DQ207942  
B.UA.2001.01UAKV167.DQ823362  
B.UA.2001.01UAKV259.DQ823363  
B.UA.2001.01UAKV259.DQ823364  
B.RU.2010.10RU6229.JX500707  
B.RU.2009.09RU4457.JX500709  
B.NL.2000.071-00736.AY423387  
B.TH.1990.BK132.AY173951  
B.FR.1983.HXB2-LAI-HIB-BR.K03455  
B.US.1998.1058-11.AY331295  
C.BR.1992.88025-d1US2953  
C.ET.1986.ETH220.U46016  
C.IN.1995.95IN21068.AF067155  
C.ZA.2004.04ZAK146.AY772699  
D.CD.1993.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.AF005494  
F1.FI.1993.FIN9363.AF075703  
F1.FR.1996.96FR-MP411.AJ249238  
F2.CM.1995.95CM-MP255.AJ249236  
F2.CM.1995.95CM-MP257.AJ249237  
F2.CM.1997.CM53657.AF377956  
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.AB1996  
H.GB.2000.00GBAC4001.FJ711703  
J.CD.1997J-97DC-KT8147.EF614151  
J.CM.2004.04CMU11421.GU237072  
J.SE.1997.97SE3080  
K.CD.1997.97ZB-EOT811.AJ249235  
K.CM.1996.96CM-MP355.AJ249239





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



Positions or sites in HIV-1 genomes are numbered using alignment to the HXB2 reference genome as the standard. The HIV Map drawing tool can be used to create maps of the genome colored by region.

[https://www.hiv.lanl.gov/content/sequence/DRAW\\_CRF/recom\\_mapper.html](https://www.hiv.lanl.gov/content/sequence/DRAW_CRF/recom_mapper.html)

## Input data:

- [sequence file](#)

[http://jpHMM.gobics.de/submission\\_hiv](http://jpHMM.gobics.de/submission_hiv)

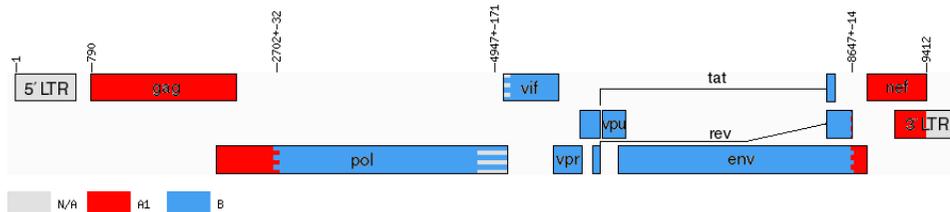
## jpHMM result:

Sequence #1: >03-AB\_BY\_2000\_98BY10443\_AF414006

This sequence is related to subtype(s): **A1 B**

Fragment Start Position	Uncertainty Region Start - End	Breakpoint Interval Start - End	Fragment End Position	Fragment Subtype
Position in the original sequence [ <a href="#">pred recombination</a> ], [ <a href="#">recombination incl UR and BPI</a> ], [ <a href="#">UR and BPI</a> ]				
1	-	-	798	N/A
799	-	2673 - 2736	2690	A1
2691	4779 - 5120	8587 - 8614	8603	B
8604	-	-	9368	A1
9369	-	-	9687	N/A
Position based on <a href="#">HXB2 numbering</a> [ <a href="#">pred recombination</a> ] [ <a href="#">recombination incl UR and BPI</a> ] [ <a href="#">UR and BPI</a> ]				
1	-	-	789	N/A
790	-	2670 - 2733	2687	A1
2688	4776 - 5117	8633 - 8660	8649	B
8650	-	-	9411	A1
9412	-	-	9719	N/A

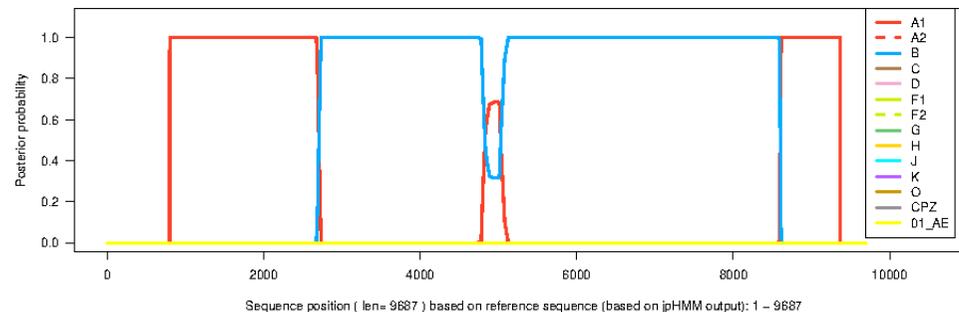
Genome map (based on [HXB2 numbering](#))



## Note:

- Numbers in the above figure denote intervals for recombination breakpoints based on HXB2 numbering.
- The uncolored regions denote missing information due to input fragment sequence.
- The gray regions denote missing information due to uninformative subtype models (subtype: N/A).
- The sequence regions of less than 10 nucleotides long are too short to be mapped onto the genome map.

Posterior probabilities of the subtypes (based on [HXB2 numbering](#))



jpHMM-HIV at Gobics gives best recombination site location numbering.

vs input query sequence location

vs HXB2 standard sequence location

[https://www.hiv.lanl.gov/content/sequence/DRAW\\_CRF/recom\\_mapper.html](https://www.hiv.lanl.gov/content/sequence/DRAW_CRF/recom_mapper.html)

## Recombinant HIV-1 Drawing Tool

**Purpose:** This tool maps your recombinant breakpoint data for HIV-1 onto a map of the HXB2 genome. The different subtypes that compose your genome appear as differently-colored regions in the map. Before using, please see [Recombinant HIV-1 Drawing Tool Explanation](#).

### Input

Upload text file of breakpoints  No file selected.  
 This is an Excel spreadsheet

Or enter breakpoint data here

1	789	N/A
790	2673	A6
2674	2733	A6/B
2734	8649	B
4766	5117	A6/B
5118	8633	B
8634	8659	A6/B
8660	9411	A6
9412	9719	N/A

If the breakpoint data entered above are not expressed in HXB2 coordinates, check this box and paste sequence here  
 Non-HXB2 coordinates

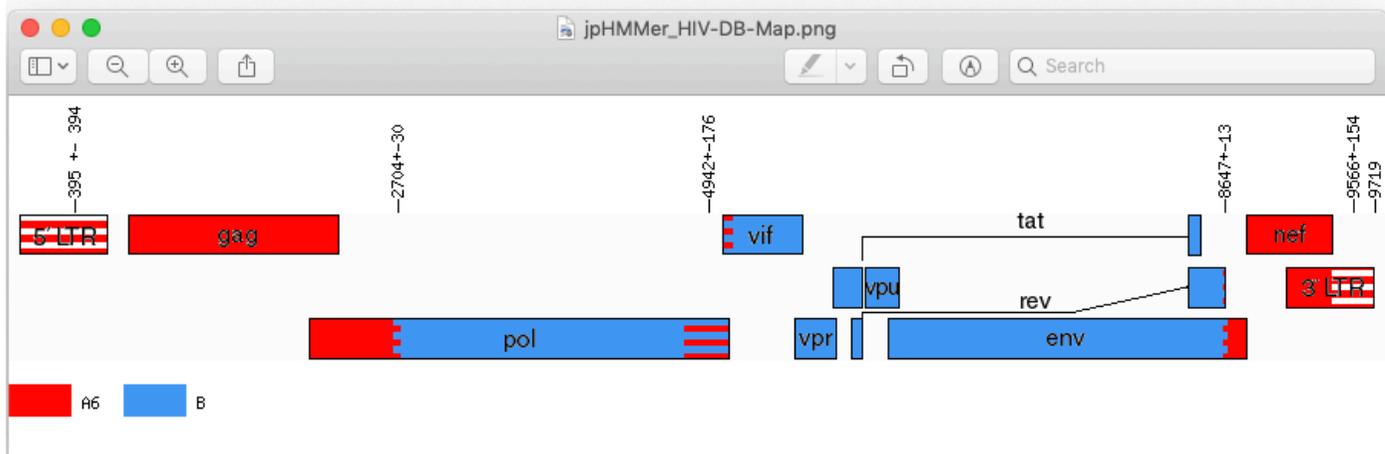
Or upload your sequence file  No file selected.

Please choose your subtype colors

A6  click in textbox to change default

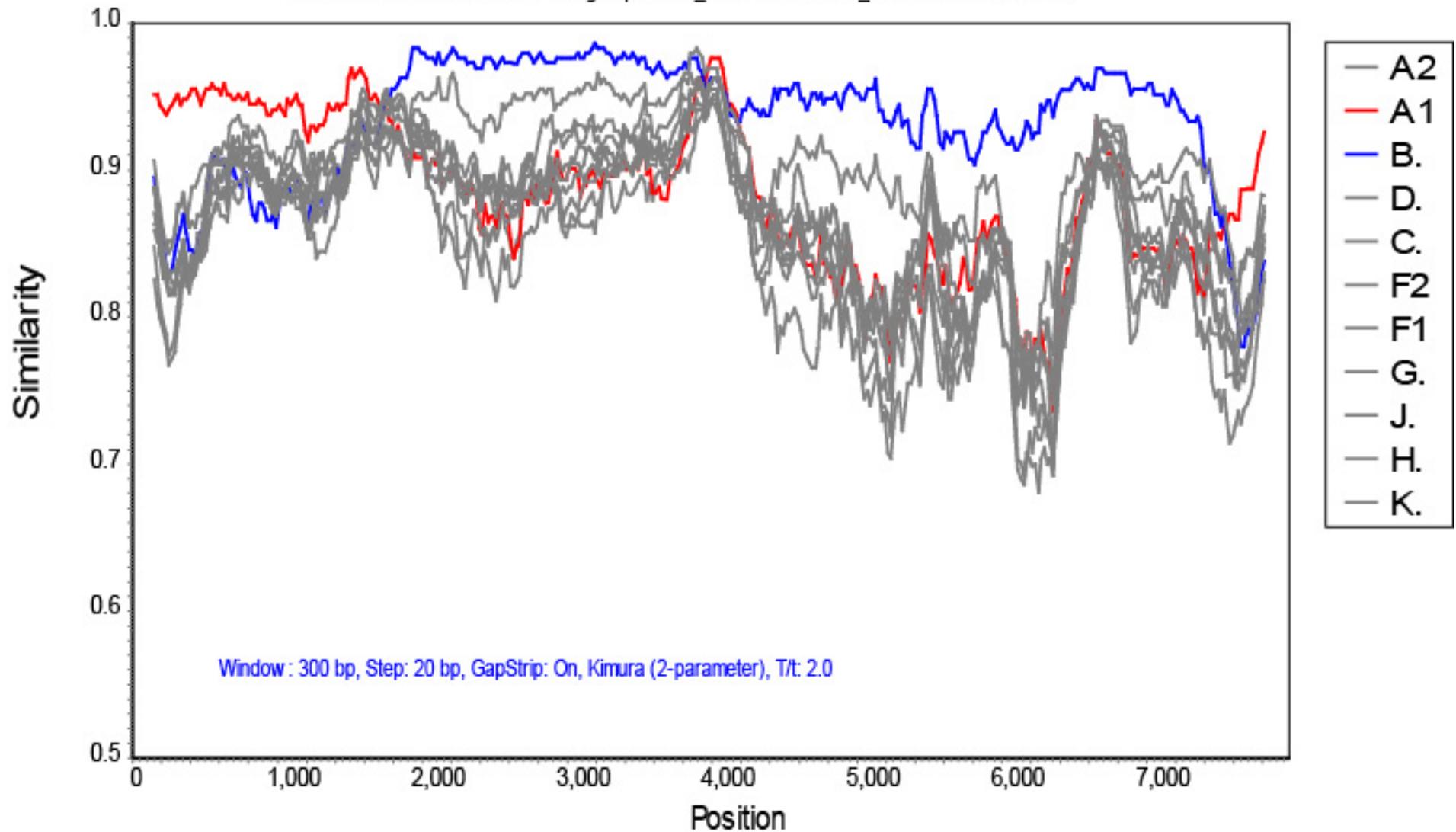
B  click in textbox to change default

HIV recombination Map Drawing tool at HIV-DB customizable to change colors used, and input your own numbers in cases where jpHMM (or another tool) did not give the correct sites.

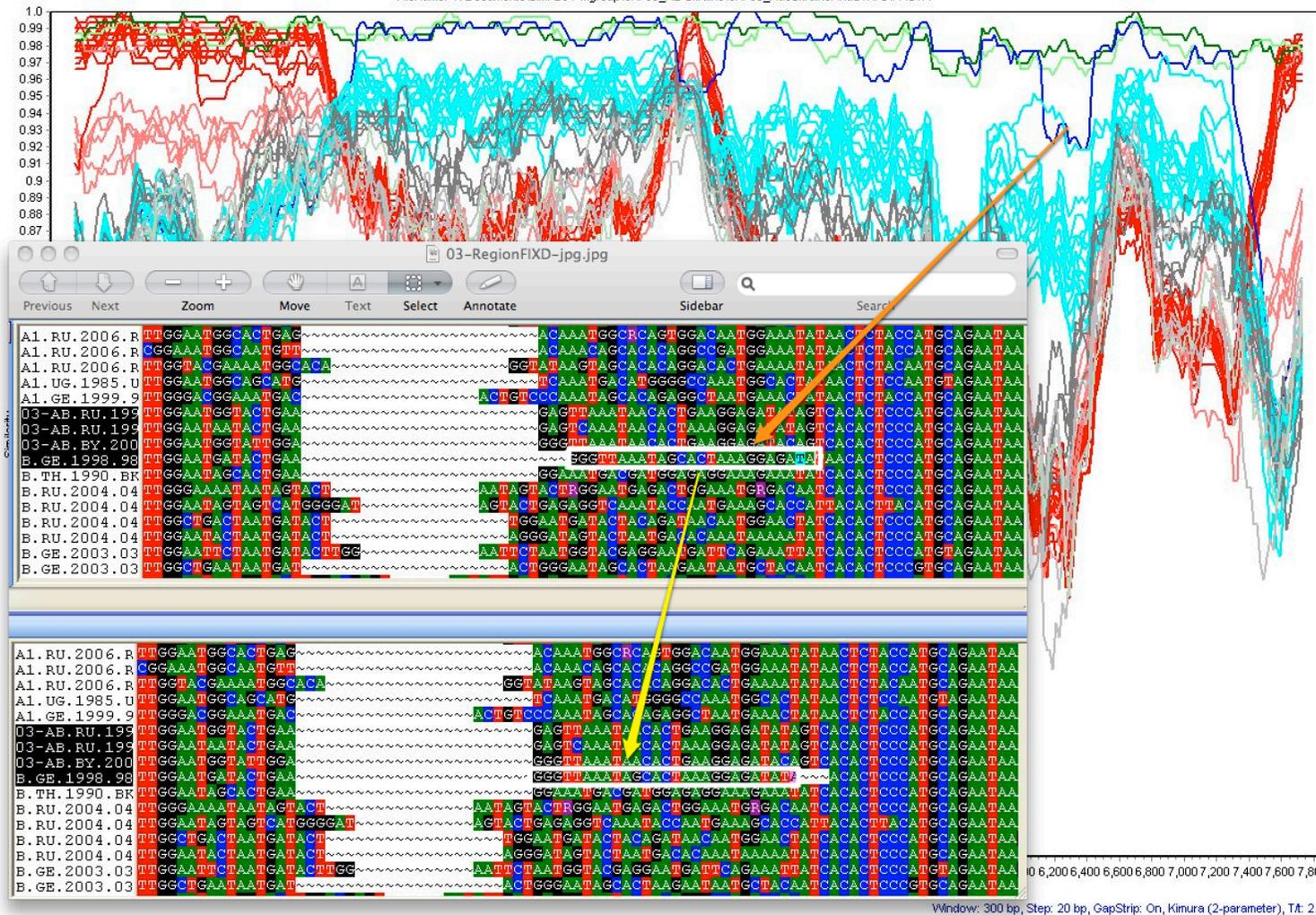


# SimPlot - Query: CRF03\_A1B

FileName: Y:\Documents\SIMPLOT-Mgroup\CRF03\_AB Ukraine\CRF03\_PlusRefsSTRPD.FASTA

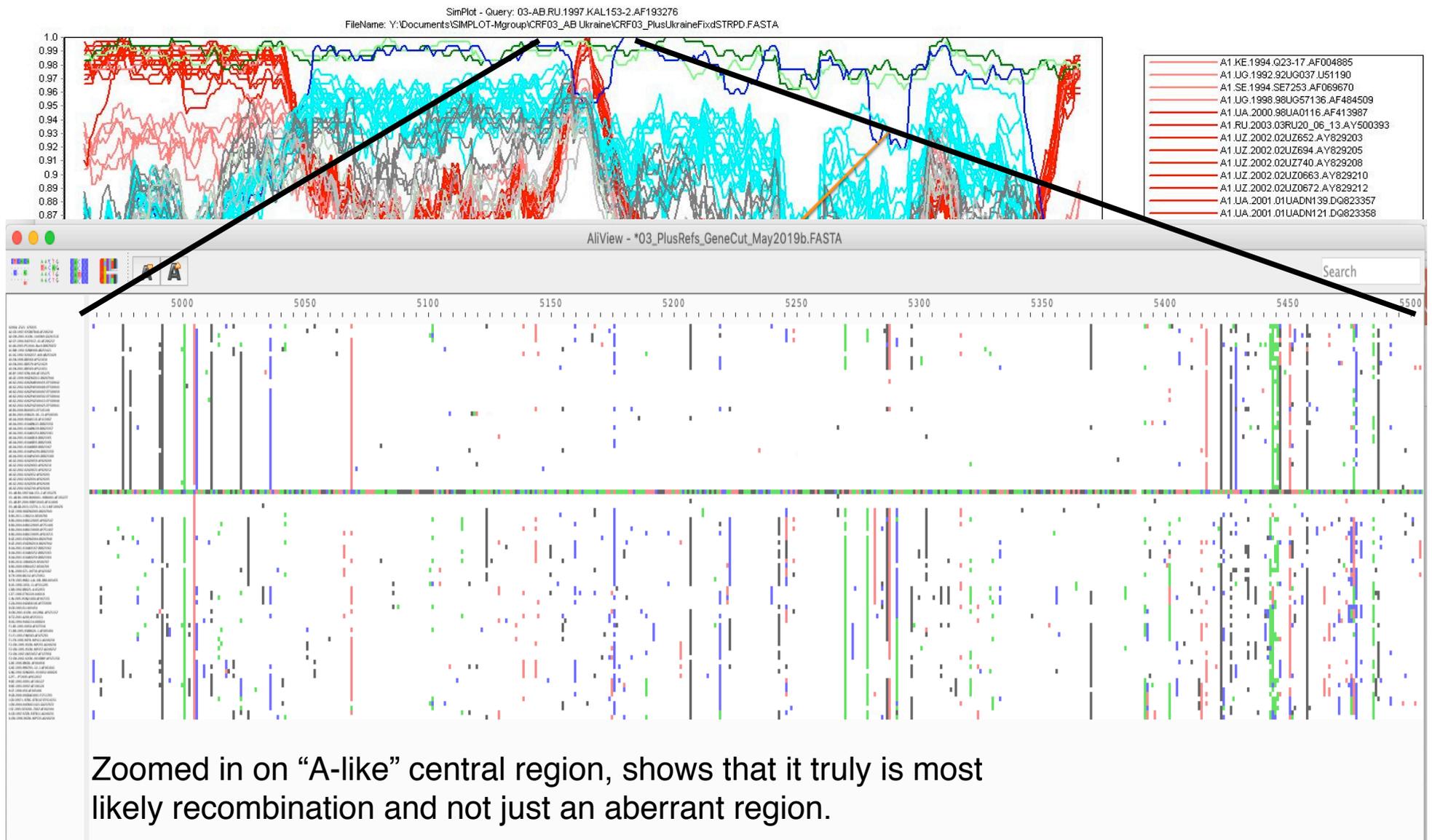


<https://sray.med.som.jhmi.edu/SCSoftware/simplot/>



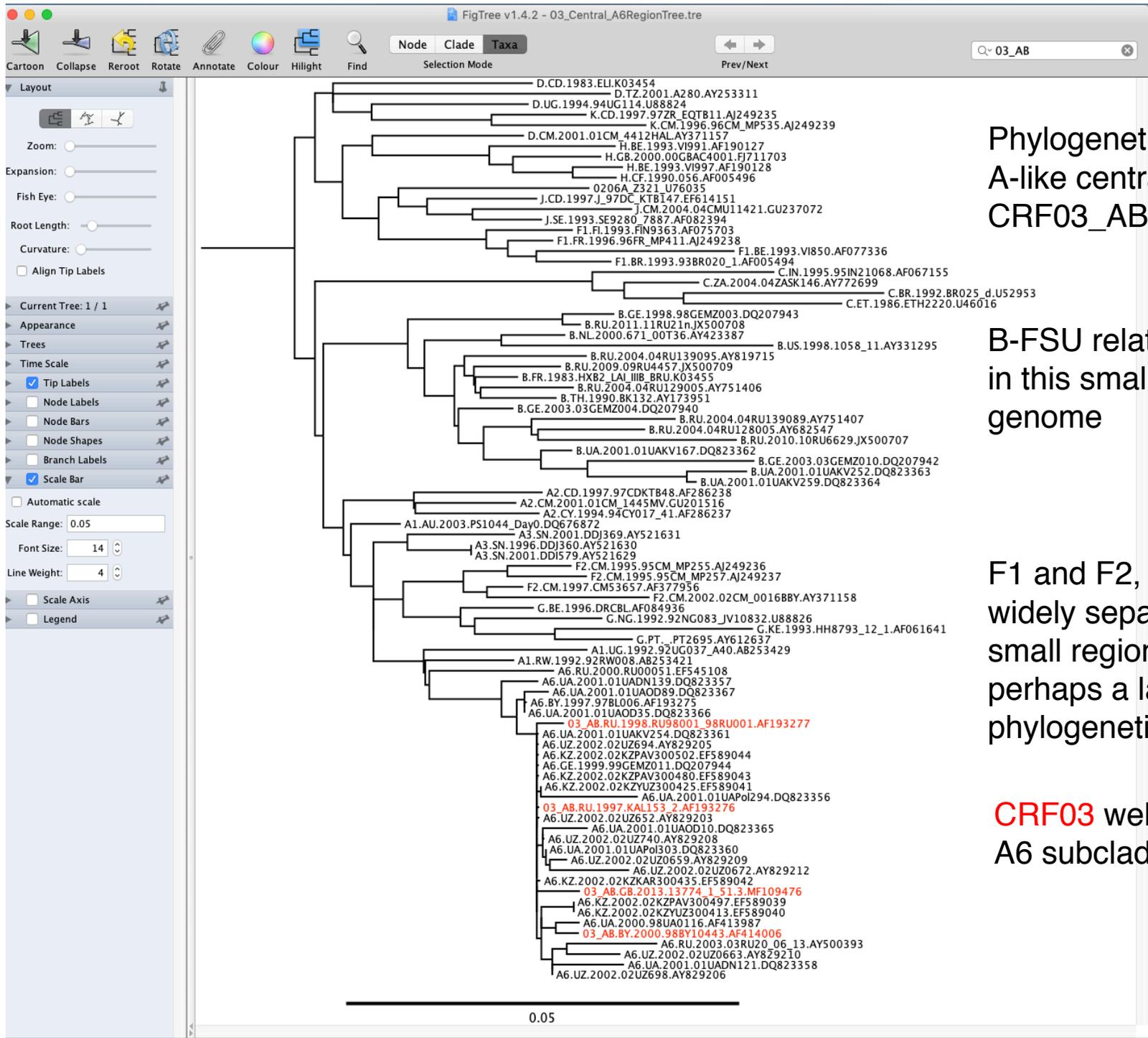
- A1.KE.1994.Q23-17.AF004885
- A1.UG.1992.92UG037.U51190
- A1.SE.1994.SE7253.AF069670
- A1.UG.1998.98UG57136.AF484509
- A1.UA.2000.98UA0116.AF413987
- A1.RU.2003.03RU20\_06\_13.AY500393
- A1.UZ.2002.02UZ652.AY829203
- A1.UZ.2002.02UZ694.AY829205
- A1.UZ.2002.02UZ740.AY829208
- A1.UZ.2002.02UZ0683.AY829210
- 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.kkustk\_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-LAIIB-BRU.K03455
- B.US.1990.WEALU160-GHOSH.U21135
- B.NL.2000.671-00T36.AY423387
- B.US.1983.RF-HA73.M17451
- D.CD.1983.NDK.M27323
- D.CD.1983.ELI.K03454
- D.CM.2001.01CM-4412.HAL.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.



Zoomed in on “A-like” central region, shows that it truly is most likely recombination and not just an aberrant region.

Other factors should also be considered too. In this case for example we know that the recombinant was formed in a person dual-infected with A6 and B viruses, so a region of A6 is not at all unexpected.



Phylogenetic tree of A-like central region of CRF03\_AB

B-FSU relatively diverse in this small region of the genome

F1 and F2, B and D widely separated in this small region, indicates perhaps a lack of solid phylogenetic information

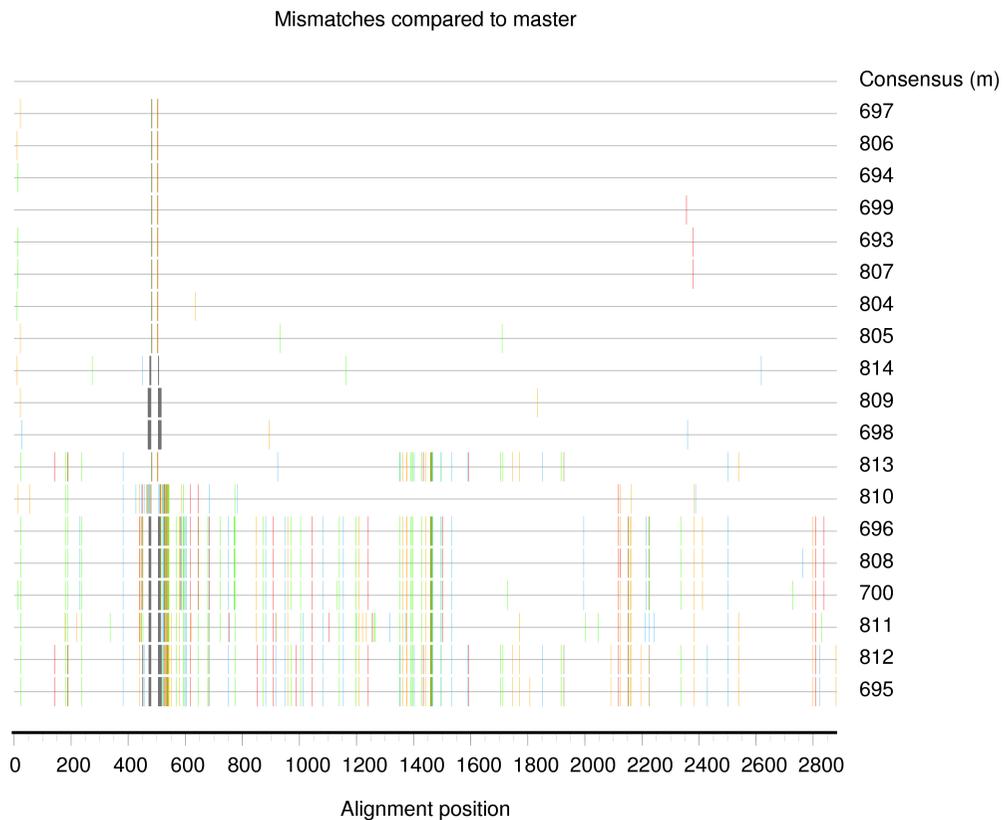
CRF03 well within the A6 subclade





# HighLighter for intra-patient recombination

[https://www.hiv.lanl.gov/content/sequence/HIGHLIGHT/highlighter\\_top.html](https://www.hiv.lanl.gov/content/sequence/HIGHLIGHT/highlighter_top.html)

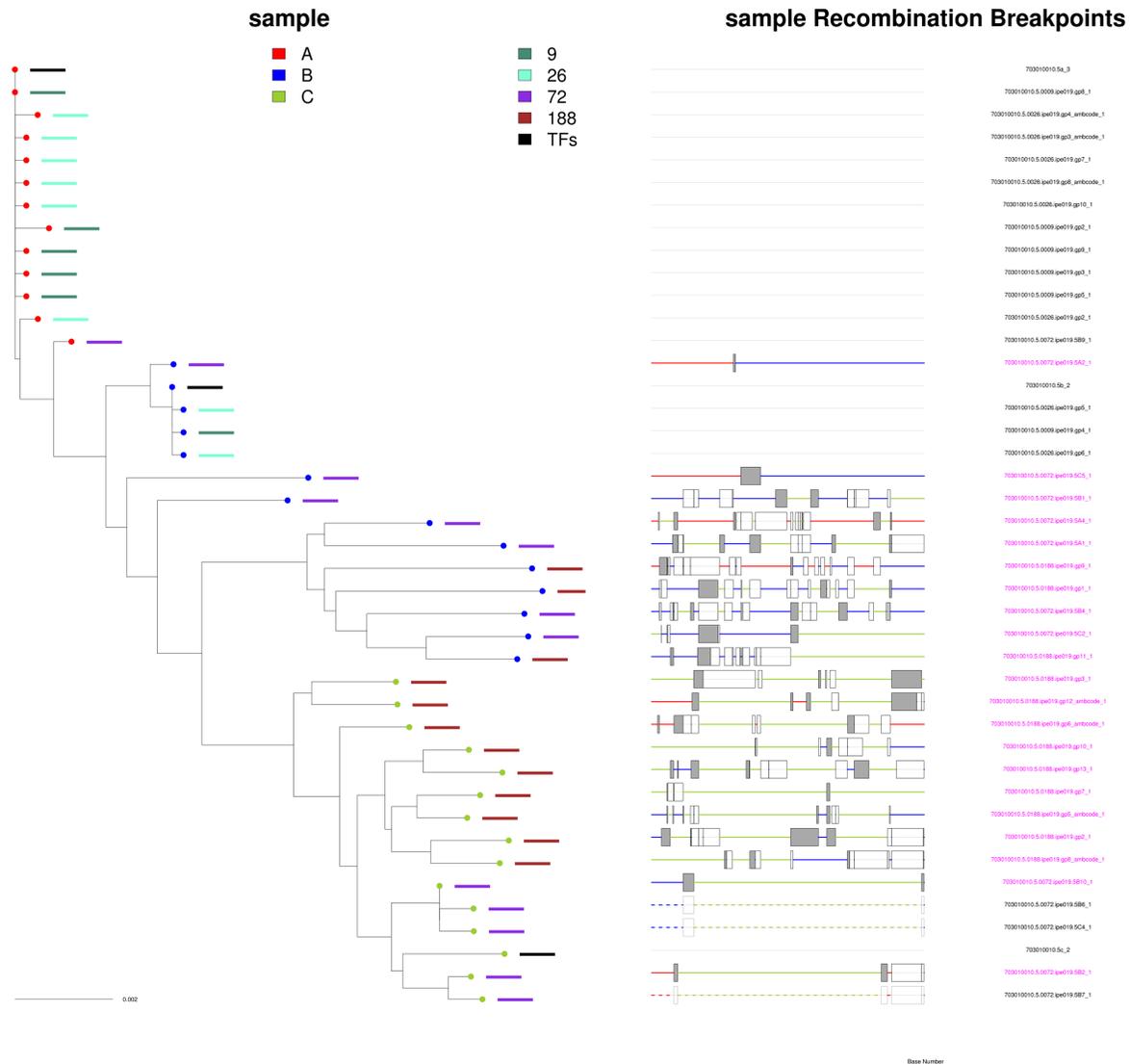


Dual-infected patient.

Infected with two strains  
of the same subtype.

# RAPR recombination analysis program

<https://www.hiv.lanl.gov/content/sequence/RAP2017/>



Song H, Giorgi EE, et al.  
Tracking HIV-1 recombination to resolve  
its contribution to HIV-1 evolution in natural  
infection.  
Nat Commun. 2018  
PubMed: 29765018