HIV Database Workshop

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

Presenters: Thomas Leitner & Bette Korber

Additional database staff: Werner Abfalterer, Gayathri Athreya, Marcus Daniels, Will Fischer, Brian Foley, Elisabeth Sharon Fung, Robert Funkhouser, Brian Gaschen, Carla Kuiken, Chien Chi Lo, Jenni Macke, Ina Maljkovic Berry, James Szinger, Jim Thurmond, Hyejin Yoon, Karina Yusim, Ming Zhang

Regular Contributors: John Mellors and Christian Brander
Project Officer: Geetha Bansal, NIAID, NIH

Theoretical Biology and Biophysics, T-10
Los Alamos National Laboratory

Workshop Topics

HIV Sequence Database and Immunology Database
Bette Korber / Thomas Leitner

General introduction

Session 1
Friday, March 28
12:00-1:30 PM
Beatty/Coleman

Sequence search interface – alignments and 'quick and dirty' trees
Geography search interface
BLAST - check for close sequences
HIV database alignments and subtype reference sequences
RIP and HIV Drawing Tool - recombination analysis
New jpHMM
Format Converter
HIV/SIV sequence locator tool
Genecutter - processing nucleotide sequences
FindModel, Branchlength, and Neighbor Treemaker
New sequence analysis tools soon to be released
Hypermut (if time permits)
Entropy (if time permits)

Session 2
Saturday, March 29
12:00-1:30 PM
Beatty/Coleman

Brief review of sequence search interface
Immunology database introduction
CTL search
Ab search
New Mosaic Vaccine Maker, Epicover, and Posicover
New Highlighter tool
New Protein Feature Accent
New HLA Frequency Tool
Epilign
Peptgen - list peptides for reagent development
Motifscan - find HLA anchor residues in a protein sequence
ELF

Bette Korber, Co-PI of the project, primary editor of the HIV Immunology Database, background in HIV evolution and immunology

Thomas Leitner, Co-PI of the project, responsible for HIV Sequence Database content and research, background in HIV evolution and molecular epidemiology

Workshop Goals

- Understanding the database content, how information was obtained, and what is available
- Database searching
- Tools for analyses
- Quality control tools

The HIV Databases

- HIV Sequence database – founded 1986, G. Myers
  - Relational database, data from GenBank with added fields from the literature
  - Alignments – align indels and reduce multiple sequences per person
  - Annual hard copy and reviews
  - Web search interfaces: subtype, phenotype, geographic, sampling year…
  - Analysis tools
- HIV Immunology database – founded 1995, B. Korber
  - Comprehensive HIV epitope database, > 300-400 new papers per year
  - Integrate HIV immunological and sequence data
  - Annual hard copy and reviews
  - Web search interfaces: epitope, protein, HLA type, immunogen, keywords
  - Analysis tools for immunologists
- HIV Drug Resistance database – founded 1997, J. Mellors
  - A searchable web listing of drug resistance mutations and literature links, updated annually by Dr. Mellors
- HIV Vaccine database – founded 2003, J. Mokili
  - A searchable relational database of published primate vaccine trials

HIV DATABASES

The **HIV databases** contain data on HIV genetic sequences, immunological epitopes, drug resistance-associated mutations, and vaccine trials. The website also gives access to a large number of tools that can be used to analyze these data. This project is funded by the Division of AIDS of the National Institute of Allergy and Infectious Diseases (NIAID), a part of the National Institutes of Health (NIH). Click on any of the links below to access a database.

SEQUENCE DATABASE ▸ | RESISTANCE DATABASE ▸

IMMUNOLOGY DATABASE ▸ | VACCINE TRIALS DATABASE ▸

News:

Archived News ▸

The 2006 HIV-1 web alignments are now available for download; the new HIV-2 and SIV alignments will follow shortly. We have also created a new download interface that gives access to all available alignments (web, subtype reference, and consensus/ancestral). This interface allows you to clip out a slice from the alignments if you provide the reference sequence coordinates. *25 February 2008*

We have added a new tool called ElimDupes, for Eliminate Duplicates. It will take your alignment and remove duplicate sequences. Several options can be set, and the sequences can be automatically divided into groups (e.g. if there are data from several patients with multiple sequences each). In the future we will add an option to also remove sequences that are more than x% similar where x<=100. *29 January 2008*

The Treemaker tool now lets you download your input sequences sorted in the order of the tree, which makes it much easier to select sequences in the alignment based on their phylogenetic behavior. *28 January 2008*

The 15th HIV Dynamics and Evolution meeting will be in Santa Fe, NM, in April 27-30, 2008. Abstract deadline is February 15. It is organized by Thomas Leitner and Carla Kuiken. There will also be a pre-meeting at the Santa Fe Institute on recent developments in phylogenetic methods, organized by Bette Korber and Tanmoy Bhattacharya on April 27.
The HIV database will have two workshops at the HIV Keystone meeting in Banff, Canada, in March 27 - April 1, 2008. *23 January 2008*

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

Operated by Los Alamos National Security, LLC, for the U.S. Department of Energy's National Nuclear Security Administration
Copyright © 2005-2006 LANSLLC All rights reserved | Disclaimer/Privacy

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

Los Alamos NATIONAL LABORATORY

HIV sequence database

DATABASES | SEARCH | ALIGNMENTS | TOOLS | PUBLICATIONS | GUIDES | Search Site

- Tutorials
- CRFs
- HIV-1 Gene Map
- 3D Structure
- How to cite this Database
- HIV Database News
- FAQs
- Links
- Site Map

HIV Sequence Database

Programs and tools

Search interface retrieves HIV and SIV sequences based on many search criteria. Sequences can be aligned, clipped, or used to build trees

Tools for working with sequences a complete list of our online tools, organized by function

About this website

FAQ Frequently-asked neral information about this site

Site Map provides direct links to most pages

How to Cite this Database

Alignments

Complete alignments all genes (nt and aa) and complete genomes (nt only)

Subtype reference alignments for use in trees, subtype comparisons and recombination research

Consensus and Ancestral sequences for M group and subtypes

Information

HIV Sequence Compendia print or order a copy of our annual compendium and review articles

Tutorials and other information unpublished web-based content posted on this site

Links to other HIV/AIDS sites for more information

News:

Archived News

MOSAIC is a suite of related tools that can help in designing vaccine reagents, a web implementation of our vaccine design strategy for desiging full length mosaic proteins that give optimal potential epitope coverage.It assesses sequence diversity in terms of fragments of a specified length (e.g.9-mers, the length of a typical CTL epitope). Independent overall summaries of the number of perfect or close matches with the target sequence population can be obtained using Epicover, and levels of similarity spanning an alignment can be obtained with Posicover. *20 March 2008*

The 14th International Bioinformatics Workshop on Virus Evolution and Molecular Epidemiology will be held in Cape Town, South Africa, September 1-5, 2008. This workshop includes lectures and hands-on molecular bioinformatics training, in phylogenetic analyses and other types of genetic/protein sequence analysis. *19 March 2008*

The 2006 HIV-1 web alignments are now available for download; the new HIV-2 and SIV alignments will follow shortly. We have also created a new download interface that gives access to all available alignments (web, subtype reference, and consensus/ancestral). This interface allows you to clip out a slice from the alignments if you provide the reference sequence coordinates. *25 February 2008*

We have added a new tool called ElimDupes, for Eliminate Duplicates. It will take your alignment and remove duplicate sequences. Several options can be set, and the sequences can be automatically divided into groups (e.g. if there are data from several patients with multiple sequences each). In the future we will add an option to also remove sequences that are more than x% similar where x<=100. *29 January 2008*

The Treemaker tool now lets you download your input sequences sorted in the order of the tree, which makes it much easier to select sequences in the alignment based on their phylogenetic behavior. *28 January 2008*

The 15th HIV Dynamics and Evolution meeting will be in Santa Fe, NM, in April 27-30, 2008. Abstract deadline is February 15. It is organized by Thomas Leitner and Carla Kuiken. There will also be a pre-meeting at the Santa Fe Institute on recent developments in phylogenetic methods, organized by Bette Korber and Tanmoy Bhattacharya on April 27.
The HIV database will have two workshops at the HIV Keystone meeting in Banff, Canada, in March 27 - April 1, 2008. *23 January 2008*

Los Alamos NATIONAL LABORATORY

Search Interface

- **Help**
  - Tips at the top of the page are often overlooked
    - Ranges, operators, wildcards, logical groupings
  - Mouse-over provides brief descriptions; click field names for details in Help file
- **Searches**
  - Searches are case-insensitive
  - Records are searchable through sequence, patient, genomic region, or publication information and can be matched to the genomic region of a user-provided alignment
  - First seven fields will appear in search results page by default
  - A "*" in a textbox will cause that field to be included in the results page
  - Patient information (Infection year, Infection country) is different than sequence information (Sampling year and Sampling country)
  - Problematic sequence filters (hypermutation, frequent ambiguities, potential contamination)
- **Analysis**
  - Build a tree with user alignment, search results and subtype reference sequences combined
- **Results**
  - Can download aligned or unaligned sequences
  - Alignments are based on multiple pair wise alignments – alignments are good, but need hand editing for an optimal alignment
  - Select all or a subset of sequences for download
  - Sequences can be re-ordered by clicking on fields at the top of the page

Tips

- Click or mouse over the field name for specific tips
- The *italicized fields* are listed in output by default
- To list fields that are not listed by default or included in the search, put an asterisk (*) in the input box
- For additional details about each field, see Help

Last GenBank update: Mar 11 2008

Advanced Search

Sequence Information

Accession number

Sequence name

Sampling country SE

Sampling city

Sampling year

Sequence length

Subtype Any, no subtype, A, A1, A2, B

Incl recombinants

Organism Any

Incl problematic seqs yes no

Patient Information

Patient code

Risk factor Any

Viral load

CD4 count

Infection country

Infection year

Incl only sequences from patients with HLA information yes no

CD8 count

Other Information

Coreceptor: list field in output, only CCR5, all CCR5, only CXCR4, all CXCR4, only R5X4

Author last name

Geographic region Any

Pubmed/Medline ID

Phenotype: list field in output, NSI, SI

Incl only drug naive sequences yes no

Days from seroconversion

Days from infection

Other fields

Find all sequences for a specific gene or region

Specific regions of HIV-1 and SIVcpz can be retrieved from the HIV database using the following options:

Include fragments of minimum length 100

Genomic region: Any, complete genome, 5' LTR, 5' LTR R, 5' LTR U3, 5' LTR U5, TAR, Gag-Pol, Gag, p17 (matrix), p24 (capsid), p7 (nucleocapsid), p6, Pol CDS

Or define start and end

To retrieve HIV-1 sequences form the database and integrate them into your own sequence alignment, paste your aligned nucleotide sequences into the submission box below.

Or upload your alignment file:

Choose File no file selected

Output

List 100 records per page

Search Reset

Show results selected

Advanced Search

Los Alamos NATIONAL LABORATORY

Please note that if a genomic region is specified, only HIV-1 sequences can be found

Make Tree | Download Sequences | Save Background Info | Clear

Displaying 1 - 16 of 16 sequences found:

Select all | Unselect all | Invert selection | One sequence/patient | Select record [ ] to [ ] | List [16] records per page

Click on field name to sort in ascending or descending order

#	Select	Patient Code (id)	Accession	Name	Subtype	Country	Sampling Year	Genomic Region	Sequence Length	Organism
1	☐ Blast	6165(5230)	AF061642	SE6165 G6165	G	SE	1993		9074	HIV-1
2	☐ Blast	SE8538(5242)	AF069669	SE8538	A1	SE	1995		8816	HIV-1
3	☐ Blast	SE7253(5238)	AF069670	SE7253	A1	SE	1994		8813	HIV-1
4	☐ Blast	SE7535(5240)	AF069671	SE7535	A1	SE	1994		8859	HIV-1
5	☐ Blast	(5239)	AF069672	SE6594	A1U	SE	1993		8791	HIV-1
6	☐ Blast	SE8891(5241)	AF069673	SE8891	A1	SE	1995		8785	HIV-1
7	☐ Blast	SE7108(5244)	AF071473	SE7108	A1D	SE	1994		8822	HIV-1
8	☐ Blast	SE9488(1002)	AF071474	SE9488	A1C	SE	1996		8813	HIV-1
9	☐ Blast	6954(10149539)	AF075701	SE6954	A1D	SE	1993		8879	HIV-1
10	☐ Blast	8603(10149538)	AF075702	SE8603	A1CD	SE	1995		8955	HIV-1
11	☐ Blast	SE7887(7259)	AF082394	SE7887 SE92809	J	SE	1993		8943	HIV-1
12	☐ Blast	SE7022(7525)	AF082395	SE7022 SE9173	J	SE	1994		8953	HIV-1
13	☐ Blast		AF107770	SE7812	02_AG	SE	1994		9063	HIV-1
14	☐ Blast	SE8131(7241)	AF107771	UGSE8131	A1	SE	1995		9016	HIV-1
15	☐ Blast	SE8646(10144930)	AY352654	SE8646	GKU	SE	1995		9074	HIV-1
16	☐ Blast	9010(10149540)	AY352655	SE9010	GKU	SE	1995		9066	HIV-1

The tab-delimited output is now a format option under **Save Background Info** at the top of the page.

Tree Builder

Purpose: This tool produces trees with a neighbor-joining algorithm, as implemented in PAUP* 4.0b.
Note: These "quick and dirty" trees are not intended for use in publications.
Learn how to make a Phylogenetic Tree.

Subtype reference sequences to include

None
All subtype reference sequences
All M Group (A-K) subtype reference sequences
All Non-Recombinant subtype reference sequences
A1.KE.94.Q23_17_AF004885 AF004885
A1.UG.92.92UG037_U51190 U51190
A1.SE.94.SE7253_AF069670 AF069670
A1.UG.98.98UG57136_AF484509 AF484509
A2.CD.97.97CDKTB48_AF286238 AF286238
A2.CY.94.94CY017_41_AF286237 AF286237

CAUTION: When including subtype reference sequences in alignments near the ends of the genome, especially the LTR, some reference sequences may be short and not cover your alignment. In that case, the tree may be meaningless. Check the resulting alignment on the next page.

Tree building parameters

Gap handling: (•) strip gaps before analysis ( ) treat as missing
Distance Model: Felsenstein 1984 (F84)
Site rates: (•) Equal ( ) Gamma Shape

Outgroup & sequence names

Use subtype reference as outgroup (•) N.CM.-.YBF106_AJ271370 AJ271370
Use selected search result as outgroup ( ) AF061642 G.SE.93.SE6165
Show names as: Subtype Country Year Name Accession

Run Reset

HIV BLAST

- Uses Basic Local Alignment Search Tool (BLAST) to compare your input sequence to all sequences in the HIV Databases.
- Good to make sure that your sequence is not identical to a “lab strain” or a sequence from another patient.
- Can be very useful for checking subtype of your sequence. But better tools, such as RIP or building a phylogenetic tree, are often needed.

DATABASES SEARCH ALIGNMENTS TOOLS PUBLICATIONS GUIDES Search Site

HIV BLAST Search

Purpose: HIV BLAST finds the database sequences most similar to your query.

Details: Although our DNA database contains mostly the same HIV sequences found in GenBank, doing the search here gives a more informative output. The output list will contain some of the fields we annotate, such as subtype, sampling country, isolation year, and the genomic location of the sequence alignment.

Note that results are listed by score, and score is not always correlated with percent identity. For example, if you BLAST a full-length sequence, the top scores will be other full-length sequences; shorter sequences of higher identity will be missed.

Input options: One nucleotide sequence or 1 amino acid sequence may be submitted per run. The sequence can be submitted in FastA format or as raw sequence. The sequence can be submitted by pasting into the text area, browsing for a file containing the sequence, or specifing a GenBank accession number.

Background options: The default is to BLAST your query against all sequences in the HIV Database. If you want to BLAST against your own submitted background set, browse for a file that contains those sequences in the space provided below.

Output options: You can download all or a selection of the sequences your BLAST search finds. If you choose the 'master-slave' output option the downloaded sequences will be aligned to one another. If you choose 'pairwise' the downloaded sequences will not necessarily be aligned.

Help files:
HIV BLAST Output Examples: examples of 'master-slave' and 'pairwise' output.
BLAST references

Input

Paste your input here [Sample Input]

or upload a file Choose File no file selected

or enter an accession number

Options

Enter a background set of sequences to BLAST against (optional) Choose File no file selected

Output style Pairwise with 'pairwise' (the default) downloaded sequences may not be aligned; with 'master-slave', they will be. Examples.

Number of BLAST matches to display 50

Submit Reset

Los Alamos NATIONAL LABORATORY

Recombination Analysis

- Many methods and programs exist to investigate potential recombination
  - http://www.bioinf.manchester.ac.uk/recombination/programs.shtml
- Investigating recombination requires many steps
- RIP 3.0
  - Automatic alignment
  - Selection of background sequences
  - Different window/gap handling options
  - Graphic & table output
- New recombination/classification tools
  - jpHMM (at GOBICS and soon at LANL)
  - Phyloplace (branching index)
  - wBLAST

Recombinant Identification Program: RIP 3.0

Purpose: RIP helps identify recombination in a **query** sequence by calculating its similarity to a **background** alignment of HIV-1 sequences of different subtypes in a sliding window.

Usage: Each RIP job requires two inputs from the user, a query sequence, and a background alignment. There are 3 options for providing the background alignment.

- The HIV database standard HIV-1 subtype consensus alignment is the RIP 3.0 default background. Under this option RIP aligns your query to this background automatically. The resulting alignment can be downloaded.
- You can construct a custom background from a list of representative subtype sequences we provide. Your query will be automatically aligned to this custom background.
- You supply the background. In this case the background must already be aligned to your query sequence, and in the same format. The background alignment must not contain the query sequence. Most standard formats are automatically recognized.

Window size and significance threshold are adjustable. Alignment gaps may be treated in four different ways. Multistate character codes (IUPAC) in your alignment can be scored as partial matches or as mismatches. When your analysis is complete you can create an illustration of the recombinant genome with our Recombinant HIV-1 Drawing Tool.

For additional details, see RIP Explanation.

Input Query

Upload your query sequence file [ ] Browse...

Or paste your query sequence [SAMPLE INPUT]

Or enter a GenBank Accession number. [ ]

Input Background

Use HIV database subtype consensus sequences as background (●)

Include 01_AE consensus [✓]

Select custom background from subtype consensuses and representative sequences ( )

Use your own alignment as background ( )

Upload your alignment here [ ] Browse...

Or paste your background here

Options

Window size 400

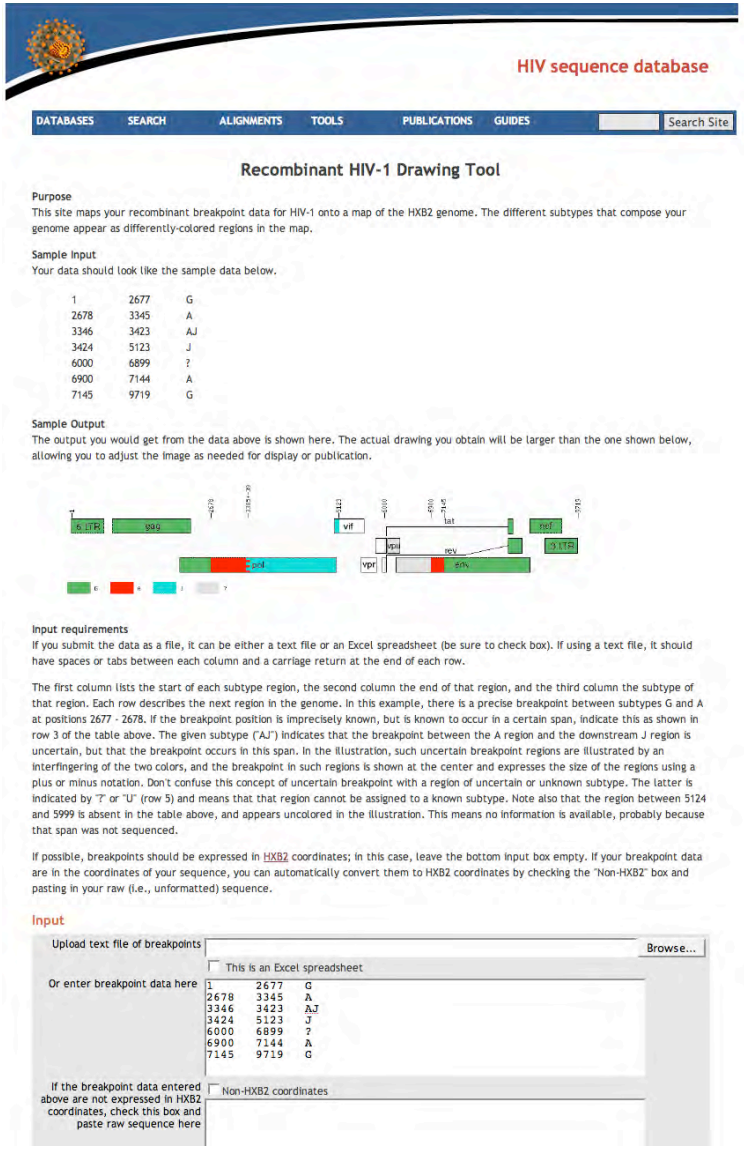
Significance threshold 90%

Gap handling
- (●) Treat gaps as characters; Plot all window values
- ( ) Treat gaps as characters; Don't plot window values for gaps in query.
- ( ) Strip all gaps; Plot all window values
- ( ) Strip all gaps; Plot blanks where gaps used to be.

Multistate characters match partially
- (●) True
- ( ) False

HIV Recombinant Drawing Tool

- Draws color figure of inter-subtype recombinant genomes.
- Takes simple input format.
- Illustrates “uncertain” regions and breakpoints.



HIV sequence database

DATABASES SEARCH ALIGNMENTS TOOLS PUBLICATIONS GUIDES Search Site

Recombinant HIV-1 Drawing Tool

Purpose

This site 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.

Sample Input

Your data should look like the sample data below.

1	2677	G
2678	3345	A
3346	3423	AJ
3424	5123	J
6000	6899	?
6900	7144	A
7145	9719	G

Sample Output

The output you would get from the data above is shown here. The actual drawing you obtain will be larger than the one shown below, allowing you to adjust the image as needed for display or publication.

Input requirements

If you submit the data as a file, it can be either a text file or an Excel spreadsheet (be sure to check box). If using a text file, it should have spaces or tabs between each column and a carriage return at the end of each row.

The first column lists the start of each subtype region, the second column the end of that region, and the third column the subtype of that region. Each row describes the next region in the genome. In this example, there is a precise breakpoint between subtypes G and A at positions 2677 - 2678. If the breakpoint position is imprecisely known, but is known to occur in a certain span, indicate this as shown in row 3 of the table above. The given subtype ("AJ") indicates that the breakpoint between the A region and the downstream J region is uncertain, but that the breakpoint occurs in this span. In the illustration, such uncertain breakpoint regions are illustrated by an interfingering of the two colors, and the breakpoint in such regions is shown at the center and expresses the size of the regions using a plus or minus notation. Don't confuse this concept of uncertain breakpoint with a region of uncertain or unknown subtype. The latter is indicated by "?" or "U" (row 5) and means that that region cannot be assigned to a known subtype. Note also that the region between 5124 and 5999 is absent in the table above, and appears uncolored in the illustration. This means no information is available, probably because that span was not sequenced.

If possible, breakpoints should be expressed in HXB2 coordinates; in this case, leave the bottom input box empty. If your breakpoint data are in the coordinates of your sequence, you can automatically convert them to HXB2 coordinates by checking the "Non-HXB2" box and pasting in your raw (i.e., unformatted) sequence.

Input

Upload text file of breakpoints — Browse...

This is an Excel spreadsheet

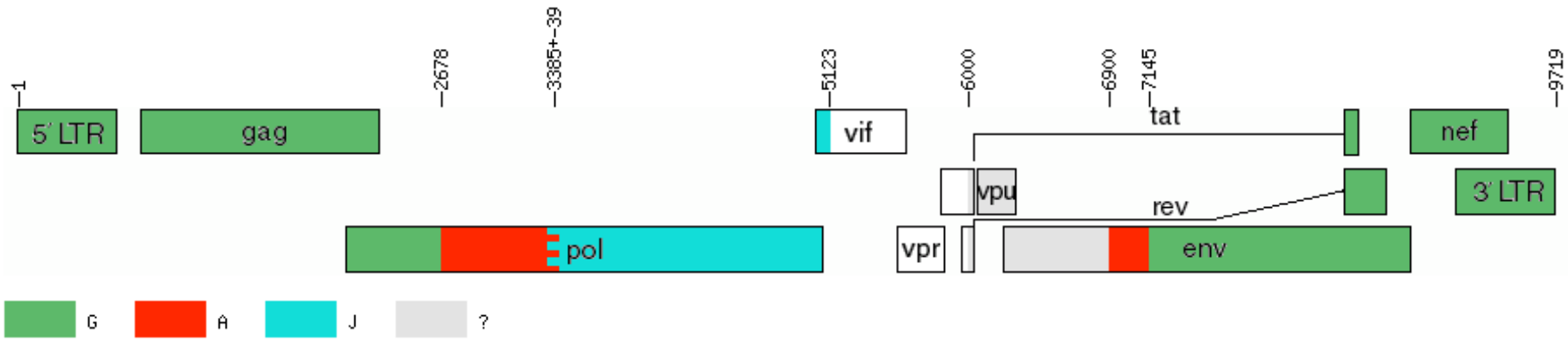
Or enter breakpoint data here

```
1       2677    G
2678    3345    A
3346    3423    AJ
3424    5123    J
6000    6899    ?
6900    7144    A
7145    9719    G
```

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

Los Alamos NATIONAL LABORATORY

Recombinant Drawing Tool



1	2677	G
2678	3345	A
3346	3423	AJ
3424	5123	J
6000	6899	?
6900	7144	A
7145	9719	G

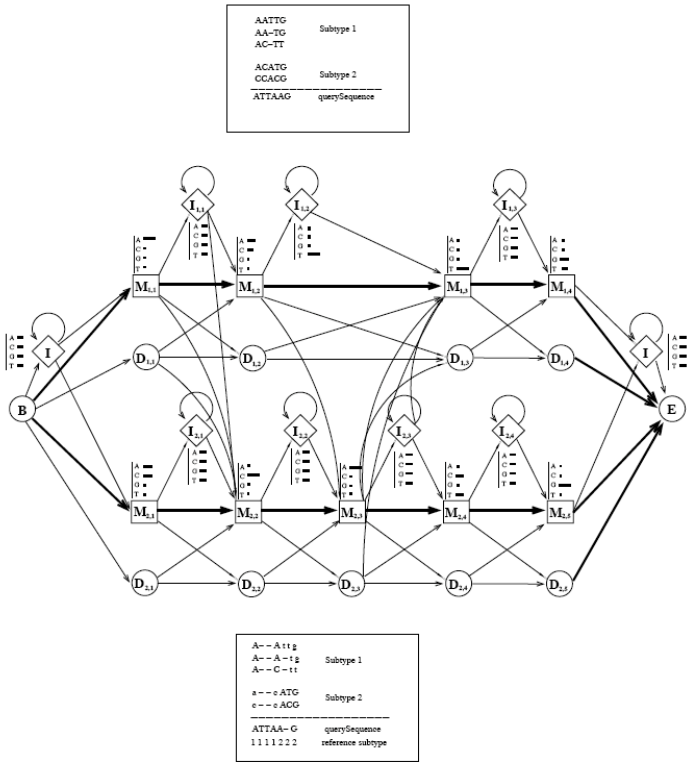
Los Alamos NATIONAL LABORATORY

jpHMM

- Automatically aligns query to reference sequences
- Builds a most probable path between subtypes

Schultz et al. BMC Bioinformatics 2006
Zhang et al. NAR 2006

BMC Bioinformatics 2006, **7**:265 http://www.biomedcentral.com/1471-2105/7/265



viterbi path (gives us the alignment of the query to the alignment):

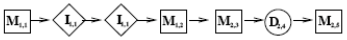


Figure 4
A toy example of a jumping profile Hidden Markov Model. This is build from a multiple sequence alignment of nucleotides with two subtypes. The first subtype consists of three sequences with four consensus columns, the second subtype consists of two sequences with five consensus columns. With each match and insert state a vector is associated for the emission probability values corresponding to the nucleotides A, C, G and T. For clarity, some transitions are omitted. Also, the figure does not show the delete states immediately after *B* from which each match state can be reached nor the delete state immediately before *E* that can be reached from each match state. Fat lines indicate high transition probabilities, thin lines correspond to low probabilities.

Los Alamos NATIONAL LABORATORY

University of Göttingen | Faculty of Biology | Institute of Microbiology and Genetics | Dept. of Bioinformatics

jpHMM-HIV [job submission]

Main page | Method | Online submission | Download | References | Program limitations

Please use the following form to submit your HIV-1 sequence(s): (up to 5 full-length HIV-1 genome sequences)
Please note that only **nucleotide** HIV-1 sequences in **FASTA** format are accepted.

Paste your query sequence(s) here: [Sample Input]

Or upload a sequence file here:

Choose File no file selected

Your email address for returning result:

Reset RUN jpHMM

Information in the result: [output sample]

- Sequence subtype;
- Predicted recombinant breakpoint position which is in two forms:
  1. based on input sequence's position number;
  2. based on HXB2 numbering.
- Genome map based on HXB2 numbering.
- **Important note regarding the result interpretation:**
  Sometimes this tool is not sensitive enough to detect HIV-1 subtype H, J, K, as only few full-length genome sequences of these subtypes are available worldwide. In this case, the jpHMM results should be compared with those of other HIV-1 subtyping tools, for example RIP.
  For further limitations of the program please see Program Limitations.

References

Please cite one of the following papers if you use this tool in your publication (a list of all references can be found here):

- M. Zhang, A.-K. Schultz, C. Calef, C. Kuiken, T. Leitner, B. Korber, B. Morgenstern, M. Stanke. jpHMM at GOBICS: a web server to detect genomic recombinations in HIV-1. *Nucleic Acids Res.*, 34:W463-5. 2006.
  (This paper describes the jpHMM web server and its usage.)
- A.-K. Schultz, M. Zhang, T. Leitner, C. Kuiken, B. Korber, B. Morgenstern, M. Stanke. A Jumping Profile Hidden Markov Model and Applications to Recombination Sites in HIV and HCV Genomes. *BMC Bioinformatics* 7:265. 2006.
  (This paper describes the jpHMM algorithm and the evaluation of the program.)

Questions or comments? Email contact

Copyright © 2005-2006 Dept. of Bioinformatics (IMG)

Los Alamos NATIONAL LABORATORY

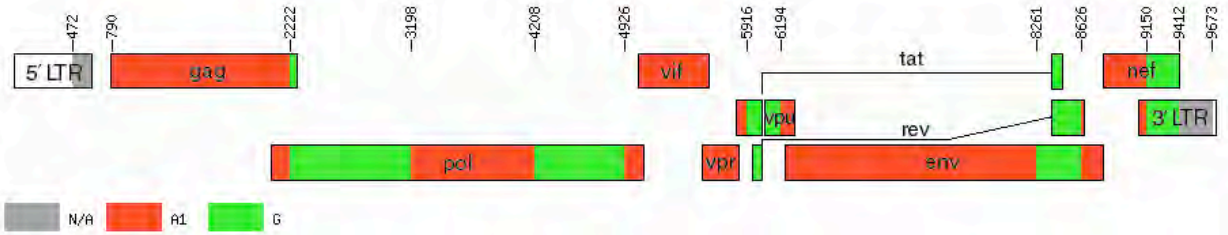
jpHMM result:

Sequence #1: 02_AG.NG.x.IBNG

This sequence is related to subtype: **A1 G** .

Fragment Start Position	Fragment End Position	Fragment Subtype
Position in the original sequence [text]		
1	329	N/A
330	1743	A1
1744	2722	G
2723	3732	A1
3733	4450	G
4451	5439	A1
5440	5726	G
5727	7766	A1
7767	8152	G
8153	8676	A1
8677	8938	G
8939	9201	N/A
Position based on HXB2 numbering [text]		
472	789	N/A
790	2221	A1
2222	3197	G
3198	4207	A1
4208	4925	G
4926	5915	A1
5916	6193	G
6194	8260	A1
8261	8625	G
8626	9149	A1
9150	9411	G
9412	9673	N/A

Genome map (based on HXB2 numbering)



Los Alamos NATIONAL LABORATORY

PhyloPlace

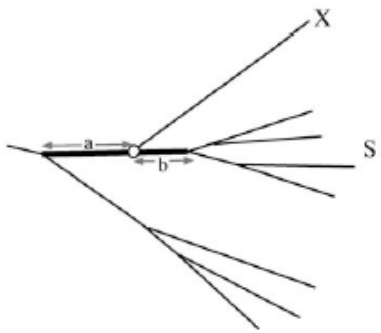
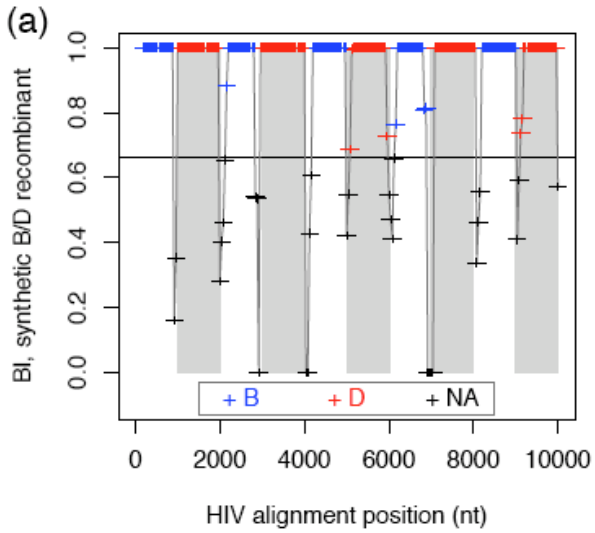


Fig. 5. Schematic picture of the branching index. Here the association of sequence X to the subtype cluster S is investigated. The letters a and b represent genetic distances that depend on the position of the node of sequence X (white circle) at the bold branch. The branching index is defined as $a/(a + b)$ and can take values between 0 and 1.



- Calculates the branching index across (informative) windows of the alignment
- Can assess if query sequence is
  - a known subtype,
  - recombinant of known subtypes,
  - <u>or is **not** composed of previously known genomes</u>

Wilbe et al, Virology 2003
Hraber et al, submitted 2007
Hraber et al, submitted 2008

Los Alamos NATIONAL LABORATORY

Sequence alignments

- Originally based on iterations of manual and HMM alignments
- Yearly updates using HMM and manual corrections
- Alignments are in reading frame (codon aligned)
- Alignments non-redundant (one sequence per patient)
- Compendium alignments show fewer sequences than web version
- Reference alignments contain up to four representatives of each subtype. One of each CRF.
- Protein alignments may contain frameshift compensations
- Subtype consensus with ties resolved, as well as maximum likelihood ancestors, are available for reagent production
  http://www.hiv.lanl.gov/content/hiv-db/CONSENSUS/M_GROUP/Consensus.html
- Special interest alignments will be added
  - Sequence sets of particular research interest
  - Suggestions welcome to tkl@lanl.gov

HIV sequence database

DATABASES SEARCH ALIGNMENTS TOOLS PUBLICATIONS GUIDES Search Site

HIV Sequence Alignments

- The **web alignments** provides nucleotide and protein alignments that represent the full spectrum of HIV and SIV sequences in the database.
- The **subtype reference alignments** contain approximately 4 representatives of each subtype, and are useful for classifying new sequences.
- The **consensus/ancestral sequences** of genetically associated subsets of HIV-1 sequences include a consensus of each subtype, an M-group consensus-of-consensuses, and some ancestral sequences.

Before use, please read the additional information below.

Options

Alignment type: ✓ All(complete)-Web alignments; Subtype Reference; Consensus/Ancestral

Year

Organism: HIV-1/SIVcpz

Region: Pre-defined region of the genome — Complete genome

User-defined range Start: End: (Coordinates: HIV1/SIV- HXB2, HIV2- MAC239)

Subtype: All

DNA/Protein: DNA

Format: Fasta

Get Alignment Reset

- Web alignments

What sequences are included

The alignments presented on the web differ from the ones that are printed in the compendia. The whole genome alignments are complete, meaning that they contain all complete genome sequences we have, including very similar ones.

The gene/protein alignments contain all complete gene sequences we have, with the important exception that very similar sequences (e.g. multiple clones from one isolate, multiple sequences from one person) have been deleted. The selection was made on the basis of phylogenetic trees: from tight clusters of sequences, one representative was retained and the others were removed from the alignment. An exception has been made for HXB2 and LAI, as these are important lab strains that are frequently used in experiments.

Los Alamos NATIONAL LABORATORY

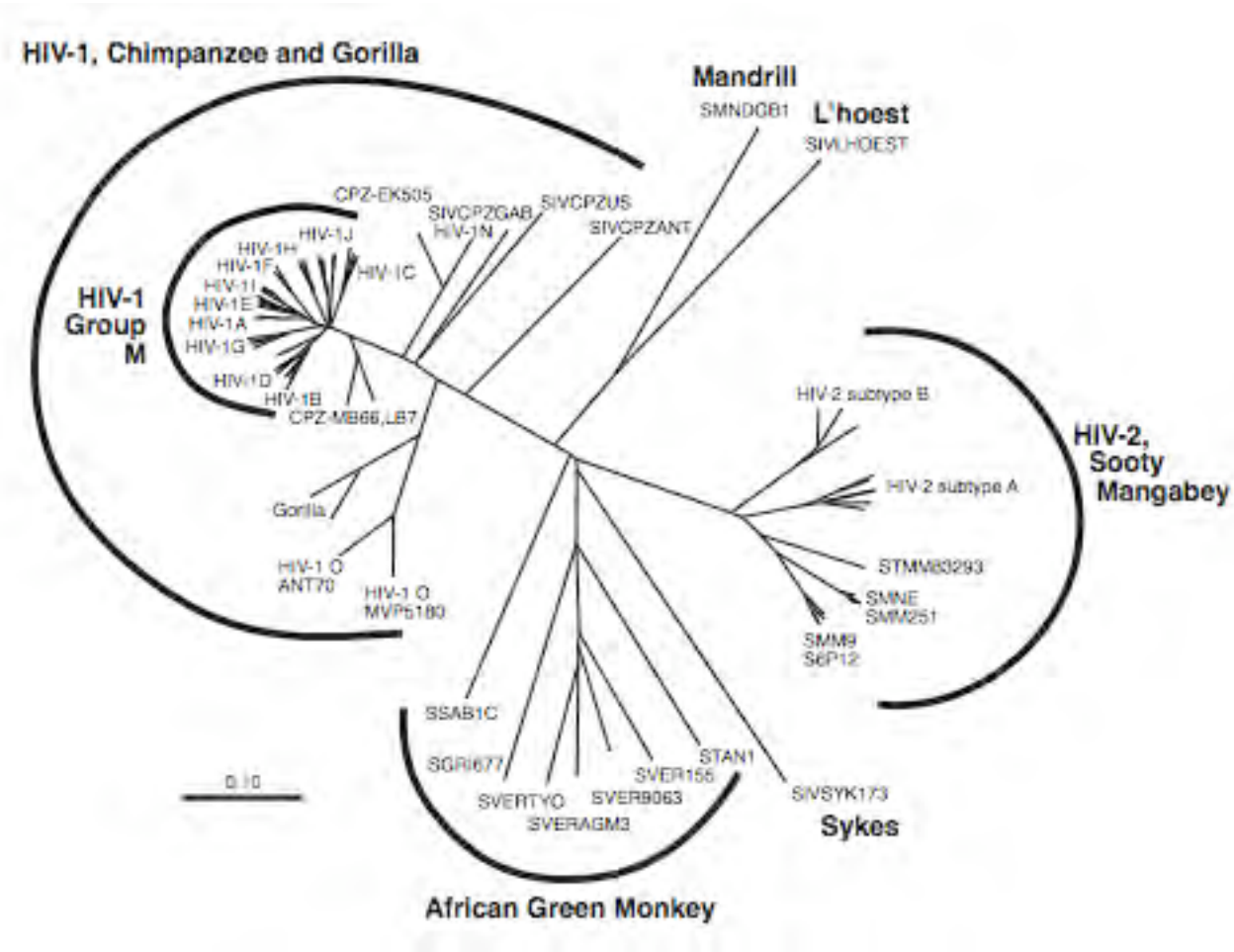
SIV/PLV Alignments

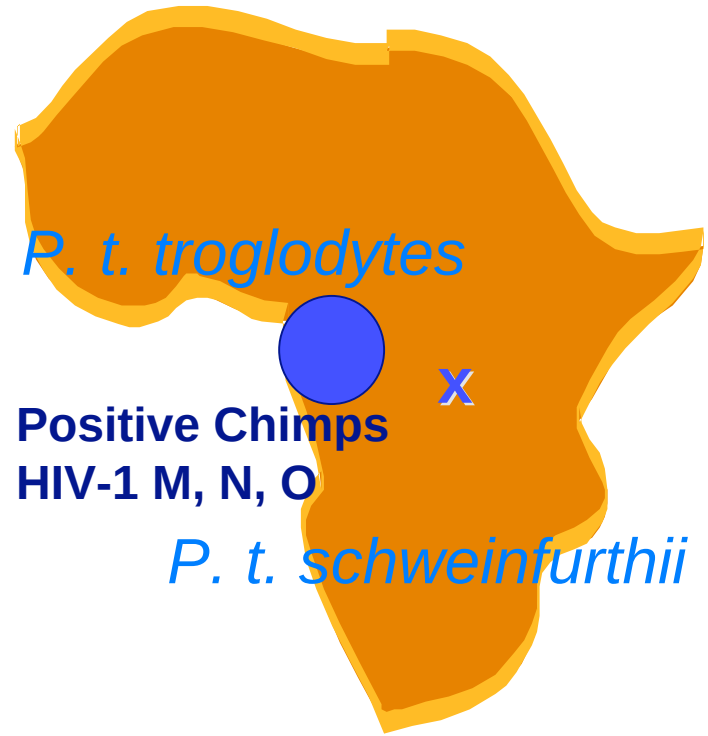
- Any non-human lentivirus is a SIV (=PLV), not just the SIV-SMM/SIV-MAC group from Sooty mangabeys.
- HIV-1s (M, N and O groups) are related to the SIV-CPZs from the chimps (*P. t. troglodytes)* and gorillas. We describe these alignments as HIV-1/CPZ.
- HIV-2s and SIV-MACs are related to SIV-SMMs from Sooty mangabeys. We describe these alignments as HIV-2/SMM.
- Dozens of other diverse non-human primates, such as African green monkeys, carry specific SIVs.
- Alignments of the diverse SIVs, plus HIVs, can help to identify highly conserved codons and other features. We describe these alignments as “other SIV” or HIV-1/HIV-2/SIV.

Los Alamos
NATIONAL LABORATORY

Primate Lentiviruses

Alignments: http://www.hiv.lanl.gov/content/hiv-db/ALIGN_CURRENT/ALIGN-INDEX.html





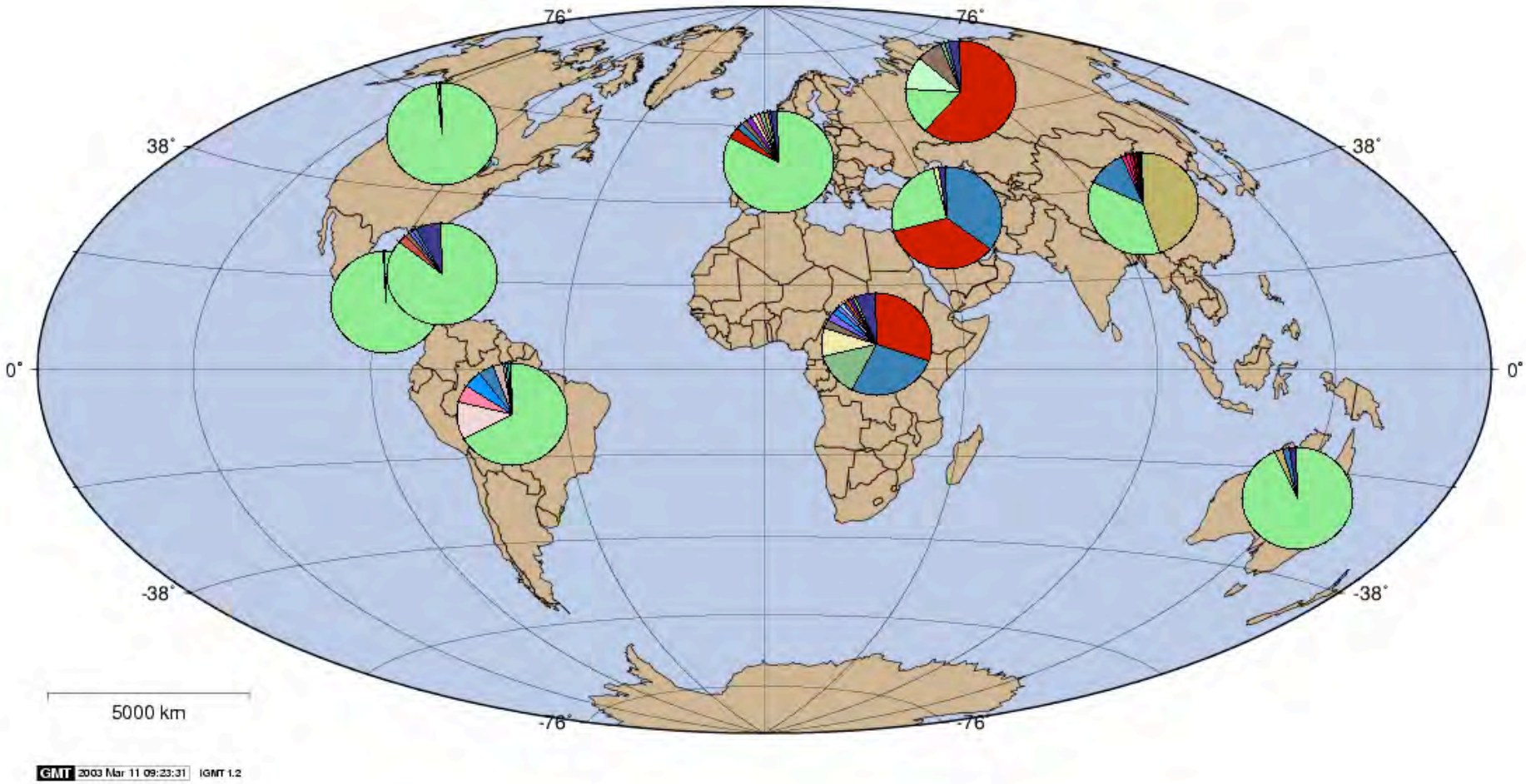
Van Heuverswyn, Nature 2006
Keele, *Science* 2006
Corbet, *J. Virol* 2000
Foley, HIV database

Los Alamos NATIONAL LABORATORY

Geography Tool

- Search/download sequences by geographic region or country
  - select all sequences from one region
  - select a specific subtype from a region
- Results are biased by sampling
  - Plots show the distributions of the samples in the database
  - The true regional subtype distribution may be much different
- Results can be selected for download (as in the search interface)

Distribution of HIV-1 sequences: World



Subtype distributions represent the frequency in the HIV Database and not the population.

About this geography site.

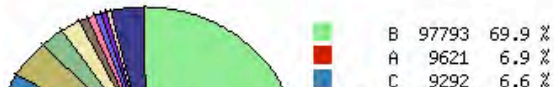
Select organism: HIV-1

Select (if a country is selected, it supersedes region)

World or Country Reset

Click pie slice to retrieve sequences

or get all sequences:



Format Converter

- Different software (including ours) may require different input formats for sequence data
- **Format converter** automatically recognizes input sequence format, and converts to any of 17 different formats

- Replaces Seq-Convert
- Requires ***plain text*** files (e.g., not .doc)
- Generates a file for copy or download

HIV sequence database

DATABASES SEARCH ALIGNMENTS TOOLS PUBLICATIONS GUIDES Search Site

Format Converter

Purpose: This program takes as input a sequence or sequences (e.g., an alignment) in an unspecified format and converts the sequence(s) to a different user-specified format.

Details: The format of the input file is automatically detected. The resulting converted file may be downloaded. Please write us if we are missing a format that you find useful, or if you find mistakes in our conversions.

For more details, see Format Converter Explanation. For descriptions of some common sequence formats, see Common Sequence Formats.

Input

Paste your sequences here [Sample Input]

Or upload your file: Choose File no file selected

Select output format ✓ table
fasta
Mase/IG
pretty print
output aligned
Clustal
Phylip interleaved
Phylip sequential
MSF
RSF
MEGA interleaved
MEGA sequential
SLX
GDE
GDEflat
Stockholm
Nexus interleaved
Nexus sequential
PIR/NBRF
GCG
raw

Output line width wide as possible

Submit Reset

last modified: Tue Dec 18 16:43 2007

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

Los Alamos NATIONAL LABORATORY

HIV/SIV Sequence Locator Tool

- Instantly computes position numbers of DNA or protein fragments relative to a reference strain (HXB2r for HIV-1, SMM239 for SIV)
  - Such numbers, often included in the literature, are frequently incorrect
- Shows the location of the sequence on an HIV map
- Presents protein translations of DNA sequences
- Can be used for input into the search interface, to align a new sequence you have generated with the database set
- Can also retrieve reference sequences
  - by coordinates (range of base or amino-acid positions)
  - by single position (retrieves flanking sequences)

HIV sequence database

DATABASES SEARCH ALIGNMENTS TOOLS PUBLICATIONS GUIDES Search Site

HIV/SIV Sequence Locator Tool

Purpose: This tool finds the position of your nucleotide or protein sequence(s) relative to the appropriate reference strain. For HIV-1, this will be the standard HXB2 strain; for HIV-2 or SIV it will be the standard SMM239 strain.

Alternative Tools:

- Retrieve from Coordinates This tool will retrieve the desired reference strain sequence (nucleotide or protein) from your beginning and ending position numbers.
- Retrieve from Point This tool will retrieve the 42 nucleotides immediately surrounding any given HXB2 coordinate.

Usage: Paste your sequence in the box below, or use the browse button to select a file that contains sequence(s) to upload. Most common sequence formats are accepted. You can specify that your input sequence(s) are HIV or SIV by selecting the appropriate choice from the list, or you can let the program decide.

Help: See Sequence Locator Explanation. See HIV-1 Genomic Map.

Useful Links: Old HXB2 Numbering Engine | HXB2 article | SMM239 article |

Input

Sequence type: Let program decide / HIV / SIV

Paste sequence [SAMPLE INPUT]

Or upload file Browse...

Examine reverse complement of sequence

Run Reset

last modified: Wed Dec 6 13:49 2006

Los Alamos NATIONAL LABORATORY

Sequence Locator: "find my sequence"

Organism: **HIV**

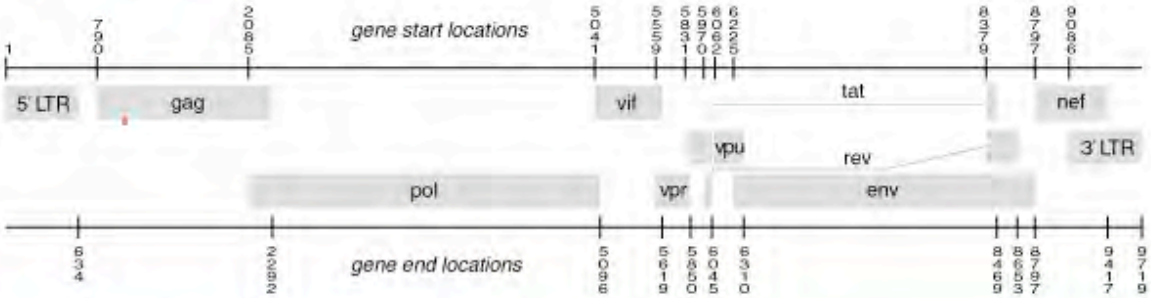


Table of protein regions touched by query sequence. AA = amino acid, NA = nucleic acid.					
CDS	**AA** position relative to protein start in HXB2	**AA** position relative to query sequence start	**AA** position relative to polyprotein start in HXB2	**NA** position relative to CDS start in HXB2	**NA** position relative to HXB2 genome start
Gag	77 → 85	1 → 9	77 -> 85	229 → 255	1018 → 1044
p17	77 → 85	1 → 9	77 -> 85	229 → 255	1018 → 1044

Alignment of the query sequence to HXB2 (Similarity 100.0%):

```
Query SLYNTVATL  9
      :::::::::
 HXB2 SLYNTVATL
```

Alignment of the protein and nucleotide equivalents of the query region in HXB2:

```
HXB2 DNA TCATTATATA ATACAGTAGC AACCCTC  1044
HXB2 PRO _S__L__Y__ N__T__V__A __T__L_
```


Los Alamos NATIONAL LABORATORY

Sequence Locator: “Retrieve from coordinates”

Reference Strain	*Type*	*Region*	*Start*	*End*
HXB2	nuc	Gag	229	255

Retrieved Sequence

TCATTATATAATACAGTAGCAACCCTC

Organism: **HIV**

LOCATION from start of HXB2 genome 1018 -> 1044 (shown as red bar in map between reading frames 1 and 2)

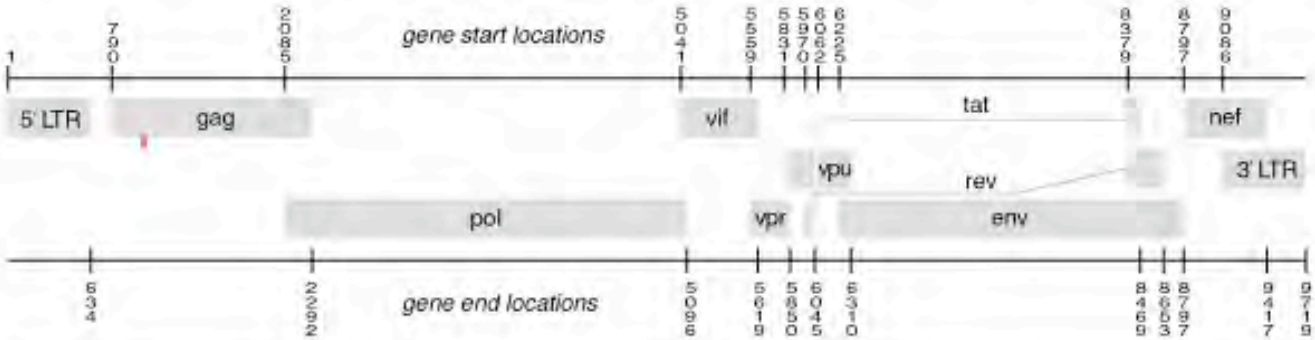
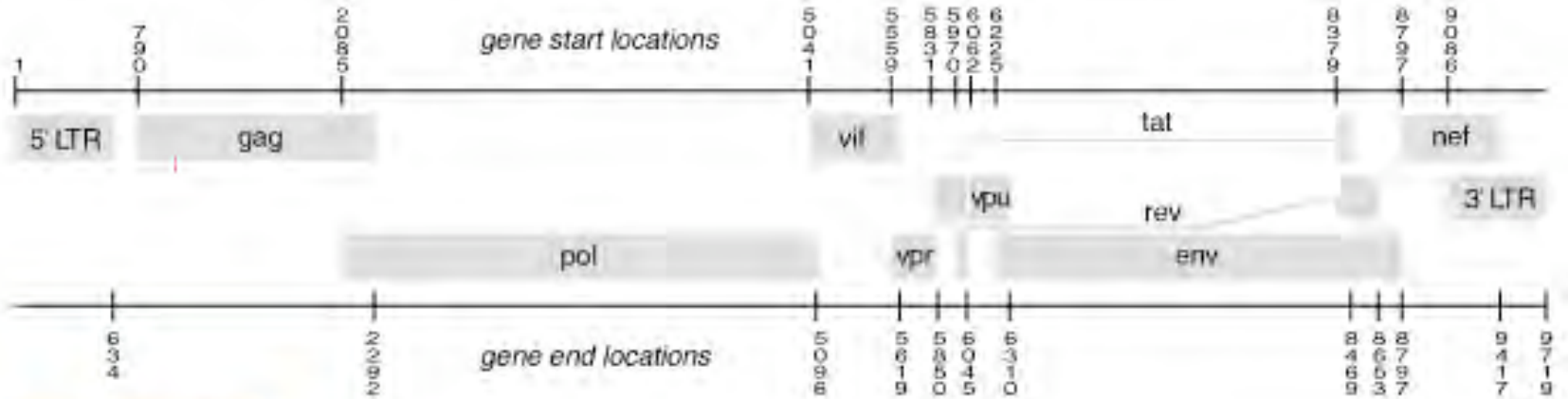


Table of genomic regions touched by query sequence. Query protein translation in blue.			
CDS	**NA** position relative to CDS start in HXB2	**NA** position relative to query sequence start	**AA** position relative to protein start in HXB2
Gag	229 -> 255	1 -> 27	77 -> 85
SLYNTVATL			
p17	229 -> 255	1 -> 27	77 -> 85
SLYNTVATL			

Los Alamos NATIONAL LABORATORY

Sequence Locator: "retrieve from point"

- Retrieves flanking sequence for a given location in HXB2



Download a simplified, text-only version of these results.

1009 1030 1050

GAACTTAGATCATTATATAATACAGTAGCAACCCTCTATTGT

Gag 74 81 87 ELRSLYNTVATLYC

p17 74 81 87 ELRSLYNTVATLYC

Los Alamos NATIONAL LABORATORY

Gene Cutter

- Unconventional Alignment/Homology program
- “Cuts out” specified genes and proteins from sets of DNA sequences
  - Aligns to HXB2 via HMMer
  - Splits input sequences into genes, and translates them
  - Aligns DNA sequences by codon
- Useful for processing new sequence data
  - annotating full length genomes
  - pulling out regions of interest from raw sequence data
- Input sequences may be aligned or unaligned
- Results may be better if the HXB2 sequence is included as a reference

Additional Features of GeneCutter

- HIV-2 and SIV also supported
  - must include SMM239 for SIV
- Recognizes IUPAC ambiguity/multi-state characters and translates them appropriately
- For each gene/region, maintains a list of anomalies
  - stop codons
  - codons containing multi-state characters
  - codons containing indels

Los Alamos NATIONAL LABORATORY

Gene Cutter: Sequence Alignment and Protein Extraction

Purpose: Gene Cutter is a sequence alignment and protein extraction tool. It can be used for any set of nucleotide sequences for HIV-1, HIV-2 or SIV.

Gene Cutter can:

- align your nucleotide sequences (if they aren't already aligned)
- clip pre-defined coding regions from a nucleotide alignment
- codon-align the coding regions
- generate nucleotide and protein alignments of the cut regions

Details: The reference sequence used by this tool is HXB2(Accession #K03455) for HIV-1 or SMM239(Accession #M33262) for HIV-2 or SIV. Gene coordinates are based on these reference sequences. This version of Gene Cutter doesn't require a reference sequence to be included in your input nucleotide alignment. Gene Cutter will also accept **unaligned sequence sets**. Gene Cutter uses Hmmer with a training set of the full-length genome alignment and will give a better multiple alignment than many computationally-based alignment programs. Mis-alignments at the ends of a coding region may result in a few amino acids/bases not appearing in the output for that coding region.

In some sequences, an insertion will be compensated within a short distance by a deletion, or vice versa. As these frameshifts may not inactivate the protein, if a compensating mutation is within 5 amino acids of an initial frameshift, the shifted reading frame is left intact. Otherwise, the frame shift is marked with the hash symbol (#), and the translation is continued in the correct reading frame beyond the offending codon. Stop codons are marked by a dollar sign ($).

The **best results** will be obtained if you submit an alignment that has been hand-aligned and contains the correct reference sequence. For more information, see Gene Cutter Explanation.

Input

Select the organism: HIV1 (HXB2)

Paste your sequences [Sample Input]

Or upload your file: Choose File no file selected

Check box if appropriate: ☑ Sequences are unaligned

Options

Region(s) to align and extract: All genes and complete sequence

☐ Insert HXB2(Accession #K03455) for HIV-1 or SMM239(Accession #M33262) for HIV-2 or SIV into the result set

☐ Remove HXB2(Accession #K03455) for HIV-1 or SMM239(Accession #M33262) for HIV-2 or SIV from the result set

☑ Codon align the region

Translation options

○ Codons containing an IUPAC character are shown as "X".

○ Codons containing an IUPAC character in a silent position are translated; others are shown as "X".

◉ Codons containing an IUPAC character are translated.

○ Do not translate to amino acids

Note: codons containing "-" are always translated to either "-" (gap) or "#" (partial codon)

Submit Reset

Please be patient. Your input file must download to our server, where the actual work is performed. This can take several minutes. Do not resubmit your sequences; you will not get a result any faster, and you will load up our server and make the process slower.

Los Alamos NATIONAL LABORATORY

GeneCutter Results

Gene Cutter Mailback Form

Please enter the email address to send the results set:

Submit email address

- Results are stored on our server
  - An HTML link is e-mailed to the user when the run is complete
  - For the workshop, use this URL:

Los Alamos
NATIONAL LABORATORY

FINDMODEL

- Determines the evolutionary model that best fits your sequence data
- Emulates Modeltest (Posada & Crandall, 1998), using the PAML program (Yang), implementation by Bill Bruno
- The best model is selected by AIC scores (takes into account the likelihood of the tree and number of model parameters)
- Model parameters are printed

FindModel Results

This program is **computationally intensive** and may take a while to run; please **don't resubmit** your request!

MODEL CONSIDERED:

Model name	AIC	LnL
JC : Jukes-Cantor (model 1)	3563.252246	-1781.626123
JC+G : Jukes-Cantor plus Gamma (model 3)	3504.693372	-1751.346686
F81 : Felsenstein 1981 (model 5)	3564.440496	-1779.220248
F81+G : Felsenstein 1981 plus Gamma (model 7)	3502.851558	-1747.425779
K80 : Kimura 2-parameter (model 9)	3499.844658	-1748.922329
K80+G : Kimura 2-parameter plus Gamma (model 11)	3423.016278	-1709.508139
HKY : Hasegawa-Kishino-Yano (model 13)	3499.375768	-1745.687884
HKY+G : Hasegawa-Kishino-Yano plus Gamma (model 15)	3412.457576	-1701.228788
TrN : Tamura-Nei (model 21)	3494.642212	-1742.321106
TrN+G : Tamura-Nei plus Gamma (model 23)	3413.187698	-1700.593849
GTR : General Time Reversible (model 53)	3487.768892	-1735.884446
GTR+G : General Time Reversible plus Gamma (model 55)	3411.658894	-1696.829447

AIC-SELECTED MODEL: **GTR+G : General Time Reversible plus Gamma** (model 55)
LnL = -1696.829447
AIC = 3411.658894

For more information about the parameters, click here.

These results are based on PAML likelihoods computed
on a weighbor reference tree; they could be different
from what would be obtained using PAUP.

Models with invariant sites are not supported in PAML
and not considered here.

This result was generated on: Thu 2/22/2007 14:59

General Time Reversible + γ

	T	C	A	G
T	f_T	a	b	c
C	a	f_C	d	e
A	b	d	f_A	f
G	c	e	f	f_G

(**Rate Parameters** with the same color are assumed to have the same value by the model.
Base Frequencies are all the same when they are showed as f_N and all different otherwise.)

Click one of the model names in the "Model Considered" table to show the parameter matrix for that model here

Questions or comments? Contact us at seq-info@t10.lanl.gov

Los Alamos NATIONAL LABORATORY

Treemaker

Check for phylogenetic relatives:

- TreeMaker produces a Neighbor Joining tree for a "quick-and-dirty" comparison
- TreeMaker uses PAUP* for its calculations; a few model options are available
- Reference sequences can be included, are aligned to the input automatically
- Trees are displayed using PHYLIP and ATV
- The alignment used for the tree can also be downloaded

Branchlength

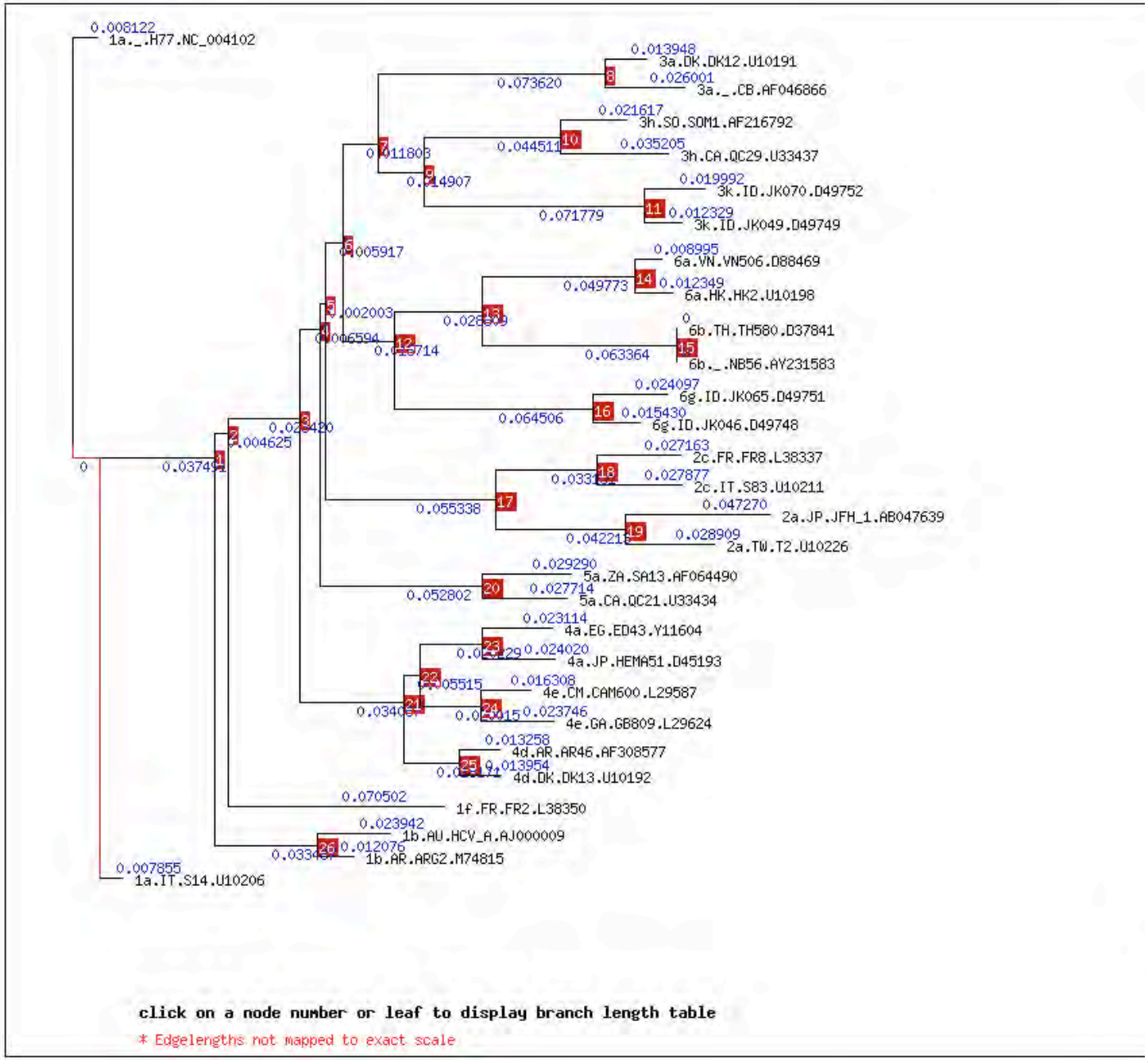
- Calculates total branch lengths in a tree
- Takes Newick file (standard treefile format) as input
- To be included in a new tool **TreeRate** that can reroot and infer the evolutionary rate between serial samples in a given tree

Single click on a node or leaf to get distances from that node or leaf
Double click to re-root the tree with that node or leaf

Sum of all branches = 1.449887

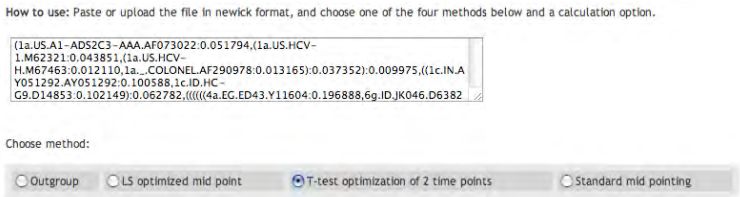
Download Newick File

Hide All | Show Nodes | Show EdgeLengths | Show All

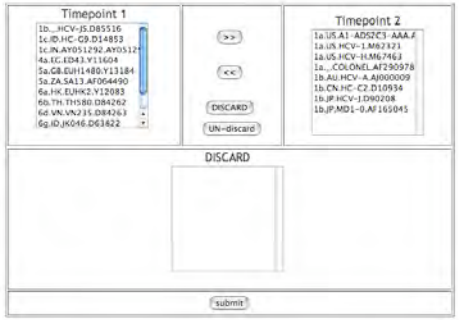


Los Alamos
NATIONAL LABORATORY

TreeRate



Several rooting options available



To calculate the evolutionary rate, sequences are divided into groups separated by time

Node Name	Avg Dist-disc	SOS disc	VAR-disc	Avg Dist-one	SumOfSquares one	variance-one	Avg Dist-two	SumOfSquares two	variance-two	t-value	p-value	delta 2 -1
1a.US.A1-AD52C3-AAA.AF073022	0		0	0.46356975	0.05842235249225	0.069774847200357	0.2352645	0.196429860858	0.156696306935582	-3.632474100795038	0.005685223776133128 8	-0.22830525
1a.US.HCV-1.M62321	0		0	0.46560175	0.05842235249225	0.069774847200357	0.231801	0.20833128872	0.161373514214694	-3.623709539803027	0.005875886486463294 4	-0.23380075
1a.US.HCV-H.M67463	0		0	0.47121275	0.05842235249225	0.069774847200357	0.22667125	0.2377410061155	0.172388009340666	-3.571578670116605 0	0.006620122902697898	-0.2445415
1a._.COLONEL.AF290978	0		0	0.47226775	0.05842235249225	0.069774847200357	0.2274625	[illegible]	[illegible]	-3.569284667185246	0.006650248697061706	[illegible]
2	0		0	0.45910275	0.05842235249225	0.069774847200357	0.21758875	0.2272743331155	0.168550561077196	-3.599852660536348 6	0.006260405271092838	-0.241514
1	0		0	0.42175075	0.05842235249225	0.069774847200357	0.19891275	0.1744026518195	0.147649353122313	-3.736386527156414	0.004637623491120 35	-0.222838
1c.IN.AY051292.AY051292	0		0	0.42390475	0.28549993991625	0.154245459121776	0.359789	0.005817427648	0.026966246605710 6	-1.346672103720555	0.202892433346097 02	-0.06411575
1c.ID.HC-G9.D14853	0		0	0.425205583333333	0.28751797146291 7	0.15478961998546	0.36135	0.005817427648	0.026966246605710 6	-1.336707622499505 6	0.206046095083274 2	-0.063855833333334
4	0		0	0.340081416666667	0.15204015033891 7	0.112561150172886	0.259201	0.005817427648	0.026966246605710 6	-2.282443807837335 0	0.040069515179762524	-0.080880416666666
4a.EG.ED43.Y11604	0		0	0.43881025	0.22083840362625	0.135658395619982	0.504798	0.000241209171999999	0.005491005964301 98	1.611219255908355 8	0.135287081755905 3	0.065987750000001
6g.ID.JK046.D63822	0		0	0.457878583333333	0.24280653468291 7	0.142245836108629	0.52768	0.000241209171999999	0.005491005964301 98	1.625597908638155 1	0.132186415077213071	0.069801416666668
10	0		0	0.274736916666667	0.01739042361296 17	0.038068385410246 3	0.30791	0.000241209172	0.005491005964301 99	2.844014944003855	0.015102792358846 27	0.033173083333334
6b.TH.TH580.D84262	0		0	[illegible]	0.273295858600917	0.150912739301259	[illegible]	0.000241209172	0.005491005964301 99	1.733782379318255 0	0.110740995540198 37	[illegible]
6a.HK.EUHK2.Y12083	0		0	0.44318875	0.22931942667225	0.138238750799794	0.515044	0.000241209171999999	0.005491005964301 98	1.721818333238923 6	0.112928445131973 7	0.07185525
11	0		0	0.29137625	0.02805613017225	0.048352981786933 3	0.332869	0.000241209172	0.005491005964301 99	2.817661083281394 0	0.016171077706391 14	0.04149275
9	0		0	0.263481583333333	0.01032134268291 67	0.029327664279938	0.291027	0.000241209172	0.005491005964301 99	[illegible]	0.010253082755808 22	0.027545416666667
6d.VN.VN235.D84263	0		0	0.446574583333333	0.22219923723091 7	0.136075725618408	0.498804	0.000241209171999999	0.005491005964301 98	1.271381627893650 28	0.229688034290872 67	0.052229416666666
8	0		0	0.25685125	0.00556956843225	0.021543692874888	0.271136	0.000241209172	0.005491005964301 99	2.094794696618522	0.056112458504580 49	0.014284750000001
6h.VN.VN004.D84265	0		0	0.448662083333333	0.23414558497891 7	0.139685833503532	0.500906	0.000241209171999999	0.005491005964301 98	1.238947940250741 95	0.241028823521203 56	0.052243916666667
6k.VN.VN405.D84264	0		0	0.449295416666667	0.23490808309891 7	0.139913092995532	0.501666	0.000241209171999999	0.005491005964301 98	1.239939384522285	0.240676203619747 01	0.052370583333334
12	0		0	0.283847916666667	0.02262980789691 67	0.043425998258451 8	0.303129	0.000241209172	0.005491005964301 99	1.454419335647784 36	0.172464758277734 30	0.019281083333333
7	0		0	0.254961916666667	0.00544260901091666	0.021296730991157 3	0.2598	0.000241209172	0.005491005964301 99	0.716936479692997 6	[illegible]	0.004833083333333 5
5a.ZA.SA13.AF064490	0		0	0.40297175	0.23695752031025	0.140522097998337	0.445166	0.000241209172	0.005491005964301 98	0.994682740510084	0.341168226557876 5	0.04219425
5a.GB.EUH1480.Y13184	0		0	0.455200916666667	[illegible]	[illegible]	0.507841	0.000241209171999999	0.005491005964301 98	1.072223288945441 3	0.306500175405715 45	0.052640083333334
13	0		0	0.33952175	0.13548520933025	0.106256454443267	0.369026	0.000241209172	0.005491005964301 98	0.919001560077 38	0.377646202555047 54	0.02950425
6	0		0	0.257310416666667	0.00794575303891666	0.025732199411950 6	0.245709	0.000241209172	0.005491005964301 99	1.444518371164806 4	0.173132518707097 45	-0.011601416666667
1b.CN.HC-C2.D10934	0		0	0.452605583333333	0.11849154705891 7	0.099369490228354 6	0.270164 25	0.137168659467 5	0.130943050344176	-3.153469690822449 8	0.008289034054644845	-0.182441333333333
1b._.HCV-JS.D85516	0		0	0.460679583333333	0.26276497225691 7	0.147976623226136	0.3126045	0.077057121214	0.098143467188346 3	-2.552008486467498	0.020032574743324035	-0.148075083333333
16	0		0	0.379824583333333	0.11849154705891 7	0.099369490228354 6	0.2155785	0.077057121214	0.098143467188346 3	-3.444530543985086 2	0.003600232674356865 6	-0.164246083333333
1b.AU.HCV-A.AJ000009	0		0	0.46424025	0.09198669163782 5	0.087553277293623 7	0.275697	0.142212022976	0.133328555235095	-3.314223856959541 7	0.007026615614162622	-0.18854325
15	0		0	0.35853625	0.09198669163782 5	0.087553277293623 7	0.196419	0.06474959488	0.089964989634857 4	-3.765981970021718 5	0.001949667859049316	-0.16211725
1b.JP.MD1-0.AF165045	0		0	0.46514825	0.09198669163782 5	0.087553277293623 7	0.27168125	0.148123699775 5	0.136071533439253	-3.346630702430149 8	0.006759238776166355 0	-0.193467
1b.JP.HCV-J.D90208	0		0	0.4742825	0.09198669163782 5	0.087553277293623 7	0.27853325	0.154800809327 5	0.139104641065413	-3.327303962414487 5	0.007126608840068 96	-0.195751
17	0		0	0.37732325	0.09198669163782 5	0.087553277293623 7	0.2058125	0.079678383838	0.099798787466331 5	-3.725240044803546 5	0.002401532535487175 5	-0.17151075
14	0		0	0.35853625	0.09198669163782 5	0.087553277293623 7	0.196419	0.06474959488	0.089964989634857 4	-3.765981970021718 5	0.001949667859049316	-0.16211725
5	0		0	0.281955416666667	0.03621450454489167	0.054935195236992 1	0.196419	0.000241209172	0.005491005964301 99	[illegible]	0.000300595270187379 4	-0.085536416666666
3	0		0	0.29822675	0.05842235249225	0.069774847200357	0.196419	0.005817427648	0.026966246605710 6	-4.355070671270703	0.000531231660281045 83	-0.10180775
0	0		0	0.41177575	0.05842235249225	0.069774847200357	0.196419	0.157077867332	0.140123993007978	-3.779028767091833	0.004156462618158448	-0.21535675

View Tree

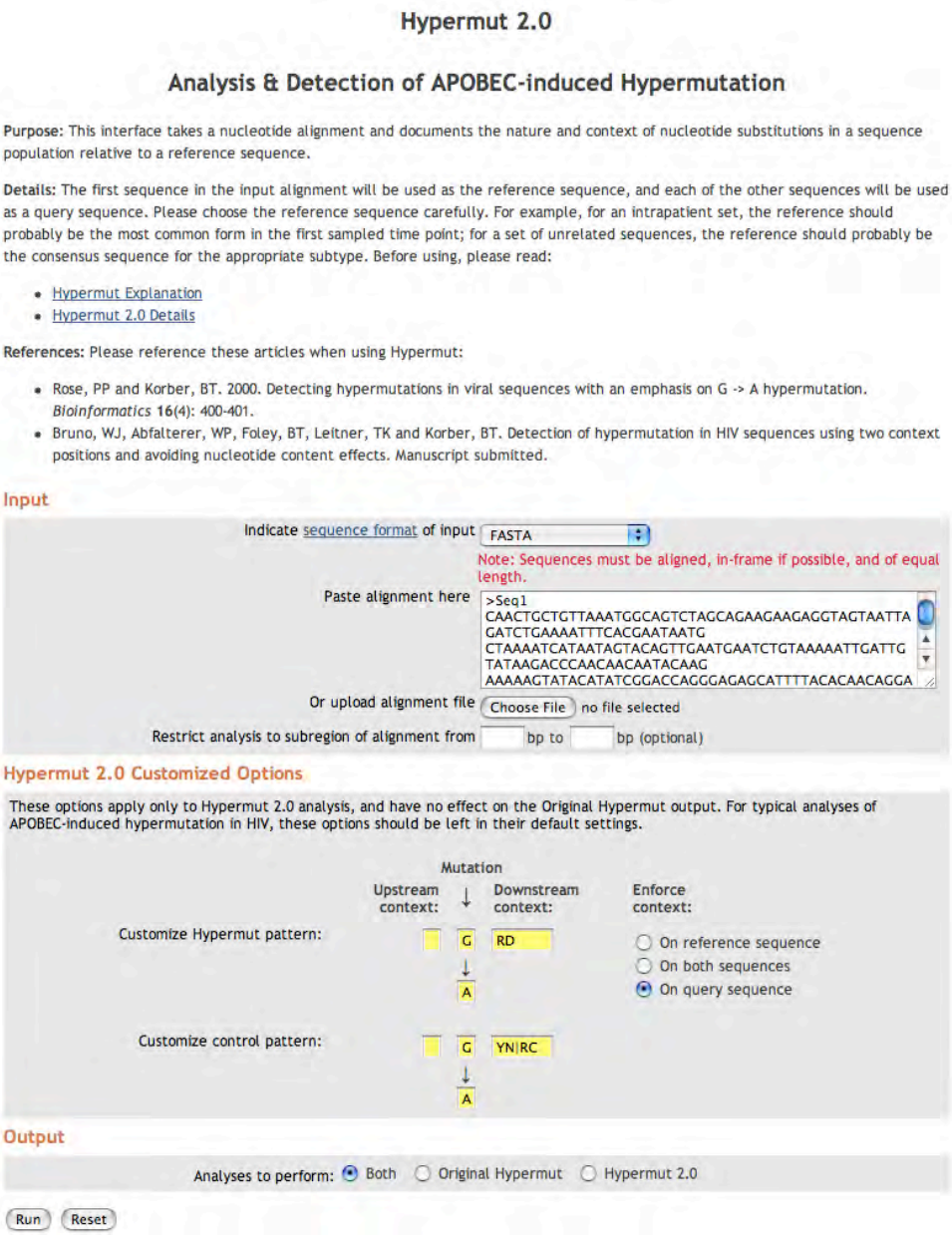
Output gives rate at all nodes and best rooting point

Several optimization criteria considered

Maljkovic Berry et al, J Virol 2007
Athreya et al, submitted 2008

Los Alamos NATIONAL LABORATORY

Hypermutation



Hypermut 2.0

Analysis & Detection of APOBEC-induced Hypermutation

Purpose: This interface takes a nucleotide alignment and documents the nature and context of nucleotide substitutions in a sequence population relative to a reference sequence.

Details: The first sequence in the input alignment will be used as the reference sequence, and each of the other sequences will be used as a query sequence. Please choose the reference sequence carefully. For example, for an intrapatient set, the reference should probably be the most common form in the first sampled time point; for a set of unrelated sequences, the reference should probably be the consensus sequence for the appropriate subtype. Before using, please read:

- Hypermut Explanation
- Hypermut 2.0 Details

References: Please reference these articles when using Hypermut:

- Rose, PP and Korber, BT. 2000. Detecting hypermutations in viral sequences with an emphasis on G -> A hypermutation. *Bioinformatics* **16**(4): 400-401.
- Bruno, WJ, Abfalterer, WP, Foley, BT, Leitner, TK and Korber, BT. Detection of hypermutation in HIV sequences using two context positions and avoiding nucleotide content effects. Manuscript submitted.

Input

Indicate sequence format of input: FASTA

Note: Sequences must be aligned, in-frame if possible, and of equal length.

Paste alignment here:
>Seq1
CAACTGCTGTTAAATGGCAGTCTAGCAGAAGAAGAGGTAGTAATTA
GATCTGAAAATTTCACGAATAATG
CTAAAATCATAATAGTACAGTTGAATGAATCTGTAAAAATTGATTG
TATAAGACCCAACAACAATACAAG
AAAAAGTATACATATCGGACCAGGGAGAGCATTTTACACAACAGGA

Or upload alignment file: Choose File no file selected

Restrict analysis to subregion of alignment from bp to bp (optional)

Hypermut 2.0 Customized Options

These options apply only to Hypermut 2.0 analysis, and have no effect on the Original Hypermut output. For typical analyses of APOBEC-induced hypermutation in HIV, these options should be left in their default settings.

	Mutation	Upstream context:	Downstream context:	Enforce context:
Customize Hypermut pattern:	G → A		RD	On reference sequence / On both sequences / On query sequence
Customize control pattern:	G → A		YN|RC	

Output

Analyses to perform: Both / Original Hypermut / Hypermut 2.0

Run Reset

- Detects APOBEC related A->G hypermutation as default
- Can be adapted to detect any fuzzy motif in relation to a control pattern

Los Alamos NATIONAL LABORATORY

Hypermut results

Hypermut 2.0

Your pattern definitions are as follows. Where there is no pattern (i.e., just '...') all sequences will match.

Pattern	Upstream	From →To	Downstream
'Mut'	...	G →A	RD ...
'Control'	...	G →A	YN|RC ...

Results

'Potential Mut' or 'Potential Control' means a match to the corresponding Upstream, From, and Downstream patterns above, while an actual 'Mut' matches those and the To pattern as well. We consider a P-value less than 0.05 to indicate a hypermutant when using the default patterns.

Sequence: (Select for graphing)	Muts: (Match Sites)	Out of: (Potential Mut Sites)	Controls: (Control Muts)	Out of: (Potential Controls)	Rate Ratio: (A/B)/(C/D)	Fisher Exact P-value: (=P(Muts,Poten.Muts-Muts, Cntrls,Poten.Cntrls-Cntrls))
☑ Seq2	0	71	0	54	undef	1
☑ Seq5	4	69	1	52	3.01	0.282669
☑ Seq7	26	71	1	54	19.77	5.35061e-07
☑ Seq14	48	71	9	54	4.06	8.26961e-09

View Sites Along Sequence

Type of graph:
- (•) Locations of Matches
- ( ) Cumulative Matches (try me!)

Graph Matches (opens in a new window)

Optional Controls:
Show region: From ____ to ____
Graph Title: Hypermut Custom Analysis
Access xmgrace compatible datafile.

Original Hypermut Output

The input file has 5 sequence(s)

Download the following info as text file

Sequence Length: 645
Compared to SEQ1, 264 As, 126 Gs, 92 Cs, 163 Ts

Graph	Table	Sequence names	Ratio	#diffs	perc_Gs	#A->G	#G->A	Dinuc Context: GG	GA	GC	GT	OBSERVED CHANGES
○	○	SEQ2	0/0	1	0.00	0	0	0	0	0	0	TC
○	○	SEQ5	5/2	33	3.97	2	5	2	3	0	0	GA TA GA CT GA GA AT CT AG A- T- A- G- T- A- C- T

Los Alamos NATIONAL LABORATORY

Entropy

- Calculates Shannon entropy, a measure of of diversity, for every position in an alignment
  - Entropy-One for one alignment:
    - Estimates the entropy for positions in one alignment
    - User specified parameters: all positions or a sub-region (window-sized)
    - Application example: compare the variability of sequence positions to immunologically important regions [Yusim *et al*.,2002]
  - Entropy-Two for two alignments:
    - Estimates if there is statistically supported evidence for different diversity in any positiion
    - User specified parameters:
      - Monte Carlo randomization number (A)
      - How many random samples can have a higher entropy difference than the actual data (B)
      - P-value: B/A
    - Application example: See if the observed entropy difference between a set of blood- vs. brain-derived HIV env sequences is likely to be due chance alone [Korber *et al.,* 1994]
- Tree based correction will soon be available

Entropy

Shannon Entropy-Two

Entropy-Two | Entropy-One | Entropy Readme | Entropy Options

Purpose: These tools apply Shannon Entropy as a measure of variation in DNA and protein sequence alignments. **ENTROPY-TWO compares two sets of aligned sequences** (named query and background sequences), and determine if there is greater variability in one set relative to the other. Each position with a significant difference in variability between these two sets will be highlighted against a query consensus.

Details:
Entropy has the option of randomizing the combined sequence sets either with replacement or no replacement, and recalculating the entropies for two random data sets broken down into sets of the same size as the original two. If you wish to include sequences of variable length in the alignment, add asterisks (*) to compensate for shorter sequences. **Avoid** including columns that are all stars. For more details, see explanation of Entropy Options.

Related programs:

- Entropy-One calculates the entropy at each position in **one set of input sequences**. It can be used as a measure of the relative variation in different positions or regions of an aligned gene or protein.
- SNAP
- PeptGen
- HyPhy

References:

- Korber BT et al., J.Virol., 68(11):7467-81 (1994)

Input

Paste your aligned **background** sequences [Sample Input]

Or browse for sequence file [ ] Browse...

Paste your aligned **query** sequences

Or browse for sequence file [ ] Browse...

Options

Use amino acid class equivalents for the calculation [ ] (ONLY for protein sequences)

Calculate the frequency of the most common aa or nt in each position [ ]

Find the statistical confidence using randomization [✓]

(•) With replacement ( ) Without replacement

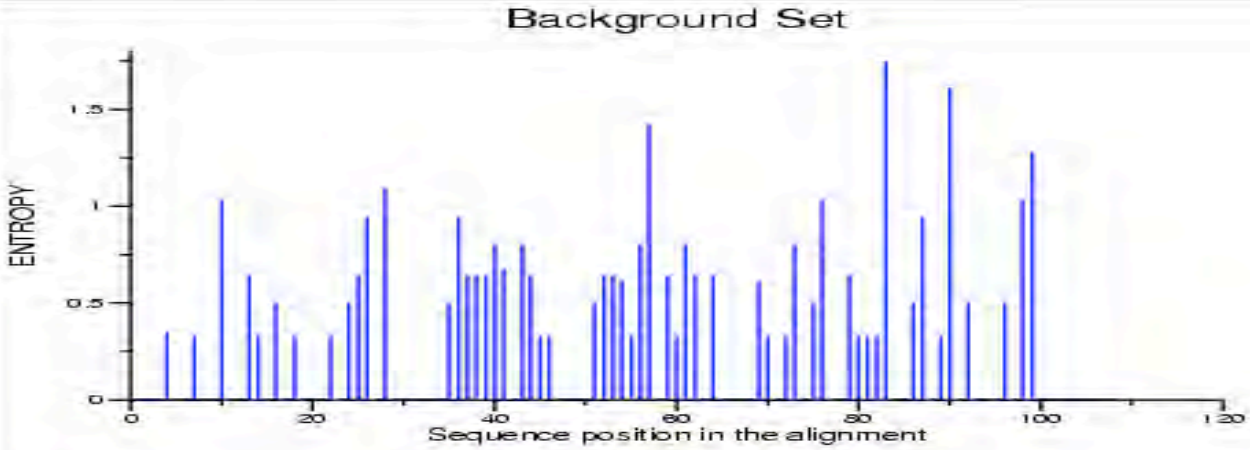
Number of randomizations [100]

Number of random samples that can have higher entropy difference than the actual data [5]

Run | Reset

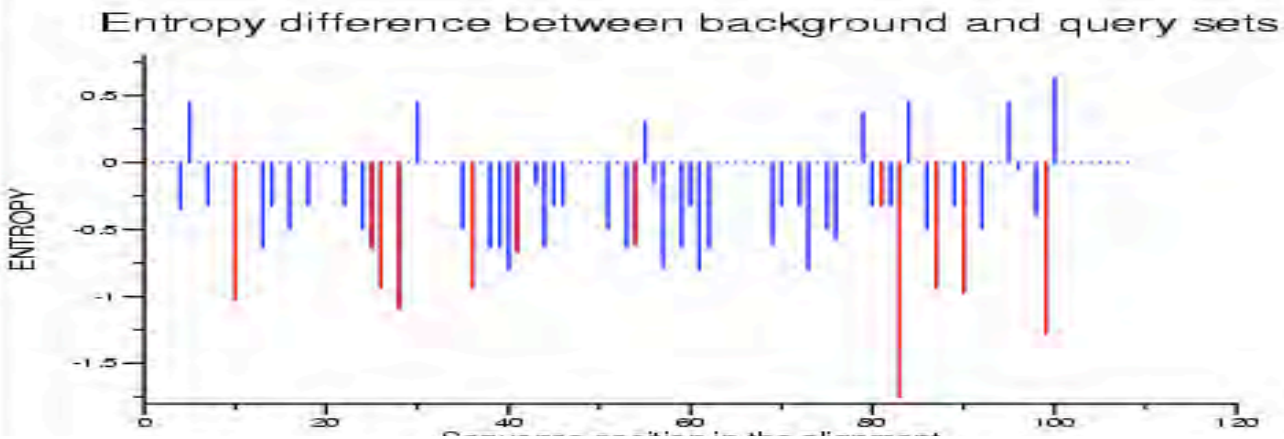
last modified: Tue Jan 16 11:36 2007

Los Alamos NATIONAL LABORATORY



[save] the plot

[save] the plot



Los Alamos
NATIONAL LABORATORY

Quality Control Tool

- Built from existing HIV database tools
- GeneCutter
- RIP
- Hypermut
- Neighbor-joining Trees
- Output is an email containing a link to a summary report
- http://www.hiv.lanl.gov/content/hiv-db/QC/index.html (beta version)

Los Alamos
NATIONAL LABORATORY

HIV sequence database
Development

DATABASES SEARCH ALIGNMENTS TOOLS PUBLICATIONS GUIDES Search Site

HIV Sequence Quality Control (QC)

Purpose: This tool has 2 objectives: (1) To examine sets of HIV-1 nucleotide sequences for quality control problems. (2) To prepare HIV-1 sequence sets, together with related data, for submission to GenBank.

Details for QC analysis: This tool will perform a set of tests that will help you easily find problems with your sequences. To run, enter a set of HIV-1 DNA sequences. The sequences should be a FastA file, aligned or unaligned, containing any number of sequences of any sequence length. Results will be returned to you by e-mail. The results will include the following analyses: subtyping (using RIP 3.0), Neighbor-joining tree of each sequence with subtype references, analysis of the number of stop codons and frameshifts (using GeneCutter), and analysis of hypermutation (using HyperMut). For the sequence set as a whole, the output also includes a Neighbor-joining tree of all sequences in the set and a complete GeneCutter alignment output. The results are given in a user-friendly table format, with links to more detail for each analysis. See Sample Output.

Details for preparing GenBank submissions: This tool can also be used to prepare sequences for GenBank submission. This step is *not* required if you only want to do the QC analysis. To prepare a GenBank submission, you must first run the QC tool as detailed above. From the output page you receive, there is a link labeled "Create GenBank Submission". After choosing this link, you will be taken to a form where you can fill in various data fields with information about the sequences (sampling year, clade, clinical parameters, etc.). From here, your data will be returned to you in a format ready for GenBank deposit.

Limitations: For very short sequences, some of the QC analyses may not work, but the tool can still be used to prepare the sequences and related data for GenBank entry. If your FastA file contains sequence fragments from different regions of the genome (for example, 20 gag sequences and 20 env sequences), this file will work for most of the QC analyses, but the tree intended to compare all sequences together will not be informative. In such cases, you should either run the tool separately for each genomic region (1 analysis for the gag sequences and 1 analysis for the env sequences), or alternatively run the set together here and perform separate tree analyses on your own (see Neighbor TreeMaker).

Input

Upload your sequence set [ ] Browse...
Enter your e-mail address [ ]

Submit

Related Links:
Sequence Quality Control Tutorial

Los Alamos NATIONAL LABORATORY

Quality Control Tool

- Summary of results from analysis programs
- Click on each result to obtain full analysis
- Useful for helping to determine subtype, hypermutation, mislabeling of samples

Immunology Database Overview

- HIV T-Cell (CTL, T-helper) and Antibodies (Ab)
- Types of data recorded
  - Epitope sequence and location: HXB2 numbering, subtype
  - Immunogen
  - Host HLA or MHC, and Ab isotype
  - Notes summarize main findings
- Contents: data from 1985 through 2006
  - 4807 CTL entries
  - 991 T-helper entries
  - 1425 Ab entries
  - 2611 published citations

HIV Molecular Immunology Database

The HIV Molecular Immunology Database is an annotated, searchable collection of HIV-1 cytotoxic and helper T-cell epitopes and antibody binding sites.

Search the Molecular Immunology Database

- CTL/CD8+ Search
- T Helper/CD4+ Search
- Antibody Search
- Search Help

Database Products

- All Database products and publications
- Epitope maps
- Epitope summary tables
- The HIV Molecular Immunology Compendium
- About the HIV Molecular Immunology database
- How to cite this database

Tools and Data Sets

- Tools for immunologists
- HIV "A list" Epitopes (PDF) review article summarizing the best-characterized HIV epitopes
- SIV Epitopes (PDF) review article summarizing known SIV epitopes
- Update interface to submit new epitopes to this database
- HLATEM HLA Typing and Epitope Mapping Data Sets
- Standardized Assessments of Neutralizing Antibodies for HIV/AIDS Vaccine Development Assay protocols from Duke Central Reference Laboratory

Immunology Database: Search

- T Cells
  - Cytotoxic T Lymphocytes (CTL)
  - Helper T Lymphocytes (T-helper)
  - Biological distinction between CTL and T-helper is not always obvious (CD8+ T-cells and CD4+ T-cells)
  - Organization is identical for CTL and T-helper
  - One reference per entry, epitope/HLA cobinations are often repeated
- B Cells (Antibodies)
  - One entry for each monoclonal antibodies
  - Many references per entry (well over 100 for well studied MAbs)

Immunology Database: Additional Information

- All entries for a reference
- Medline links to papers
- Epitope Tables
- Epitope Maps
  - Unique species/HLA for T cell epitopes
  - MAb name, species code for Ab
- Epitope Alignments
  - Extracted from HIV-sequence database, includes subtype, country and year of sampling

CTL/CD8+ T-cell Search

Databases Search Tools Products Publications Search Site

Search CTL/CD8+ T-Cell Epitope Database

HIV Protein	Proteins with defined epitopes: - ALL -, p17, p17-p24, p24, p24-p2p7p1p6	Proteins with undefined epitopes: - ALL -, p17, p24, Gag, Gag/Pol
Epitope		
Subtype	- ALL -	
Immunogen	- ALL -, computer prediction, HIV-1 and GBV-C co-infection, HIV-1 and HCV co-infection, HIV-1 exposed seronegative, HIV-1 infected monocyte-derived, HIV-1 infection	
Vaccine details	If Immunogen is Vaccine, additional search details	
	Vaccine type	- ALL -
	Vaccine strain	- ALL -
	Vaccine component	- ALL -
	Adjuvant	- ALL -
Species	- ALL -	
MHC/HLA	- ALL -, A*0101, A*02, A*0201, A*020101, A*0202, A*0205	
Author		☐ First ☐ Last
Country	- ALL -	
Keywords	- ALL -, acute/early infection, adjuvant comparison, antibody generation, assay standardization/improvement, binding affinity, characterizing CD8+ T cells	

Search Reset

Los Alamos NATIONAL LABORATORY

Found 186 matching records: Search SLYNTVATL

Displaying record number 54701

HXB2 Location	p17(70-86)	p17 Epitope Map
Author Location	p24	
Epitope	TGSEELRSLYNTVATLY	Epitope Alignment
Subtype	B	
Species (MHC/HLA)	human	
Immunogen	HIV-1 infection	
Country	Barbados, Haiti, United States	
Experimental methods	CD8 T-cell Elispot - IFNγ, Intracellular cytokine staining, Other	
Keywords	binding affinity, immunodominance	

Notes

- To identify HLA-independent immunodominant regions in HIV clade B across different ethnicities, the authors analyzed CTL responses to HIV in African-American (AA), Caucasian (C), Hispanic (H) and West Indian (WI) cohorts, and analyzed factors infuencing immune dominance. Host HLA-A11 and -A25 distributions differed among ethnicities significantly; but the virus regions most targeted by CTLs in breadth and magnitude were in Gag, Nef and Int proteins (Nef-76, most targeted). While Pro was the least targeted, 88% of peptides tested across the entire HIV sequence elicited CTL responses. 6 peptides tested were most differentially targeted between ethnicities; whilst 3 frequently targeted and 10 highly CTL-reactive regions across all 4 ethnicities were noted for vaccine design purposes.
- Suppressive treatment rather than treatment itself was found to decline CTL responses, whose breadth and magnitude correlates with CD4 count and not viral load. Factors found to determine peptide recognition by CTLs were (i) conserved peptides displaying low entropy (ii) low fraction of forbidden amino acids and (iii) high proteasomal cleavage score, confirming previous informatic studies [Yusim et al. J.Virol. 76:8757-68 (2002)]. These 3 variables did not correlate with each other but an inverse trend was seen between cleavage score and entropy.
- In addition, since all ethnicities were able to mount CTL responses to every HIV protein and only <12% of peptides were not recognized by any subject, the consensus sequence-based approach to vaccine design is substantiated.
- This immunodominant, frequently targeted overlapping peptide, TGSEELRSLYNTVATLY, had an overall frequency of recognition of 22% - 23.7% AA, 23.1% C, 22.7% H, 4.8% WI. This peptide is included in a 34 aa Gag-p17 highly reactive region to be used for vaccine design.

References

Frahm2004 Nicole Frahm, Philip J. R. Goulder, and Christian Brander. Broad HIV-1 Specific CTL Responses Reveal Extensive HLA Class I Binding Promiscuity of HIV-Derived, Optimally Defined CTL epitopes. In Bette Korber, Christian Brander, Barton Haynes, Richard Koup, Carla Kuiken, John P. Moore, Bruce D. Walker, and David I. Watkins, editors, *HIV Molecular Immunology 2003*. page 3 ff. Los Alamos National Laboratory, Theoretical Biology and Biophysics, Los Alamos, New Mexico, 2004. URL: http://hiv-web.lanl.gov/immunology. Show all entries for this paper.

Displaying record number 52778

HXB2 Location	p17(77-85)	p17 Epitope Map
Author Location	Gag(77-85)	
Epitope	SLYNTVATL	Epitope Alignment
Epitope Name	SL9	
Species (MHC/HLA)	human(A*0201)	
Immunogen	HIV-1 infection	
Experimental methods	Tetramer binding	
Keywords	genital and mucosal immunity	

Notes

- Lymphocytes from rectal biopsies were used to characterize the CD8+ T cell response to HIV in GALT, Gut-associated lymphoid tissues. Patients were selected on the basis of being HLA-A2+ and having detectable SLYNTVATL and ILKEPVHGV tetramer responses in PBMC. SLYNTVATL frequency was increased in GALT relative to PBMC in 6/7 patients studied, while a control response to a CMV-peptide was diminished in GALT. Only two patients had ILKEPVHGV CD8+ T cell responses, and both had slightly higher frequencies in GALT than PBMC.
- HIV may perturb lymphocyte retention in GALT, suggested by an overall reduction of GALT CD8+ cells expressing alphaEbeta7. GALT HIV-specific CD8+ T cells expressed alphaEbeta7, suggesting mucosal priming.

References

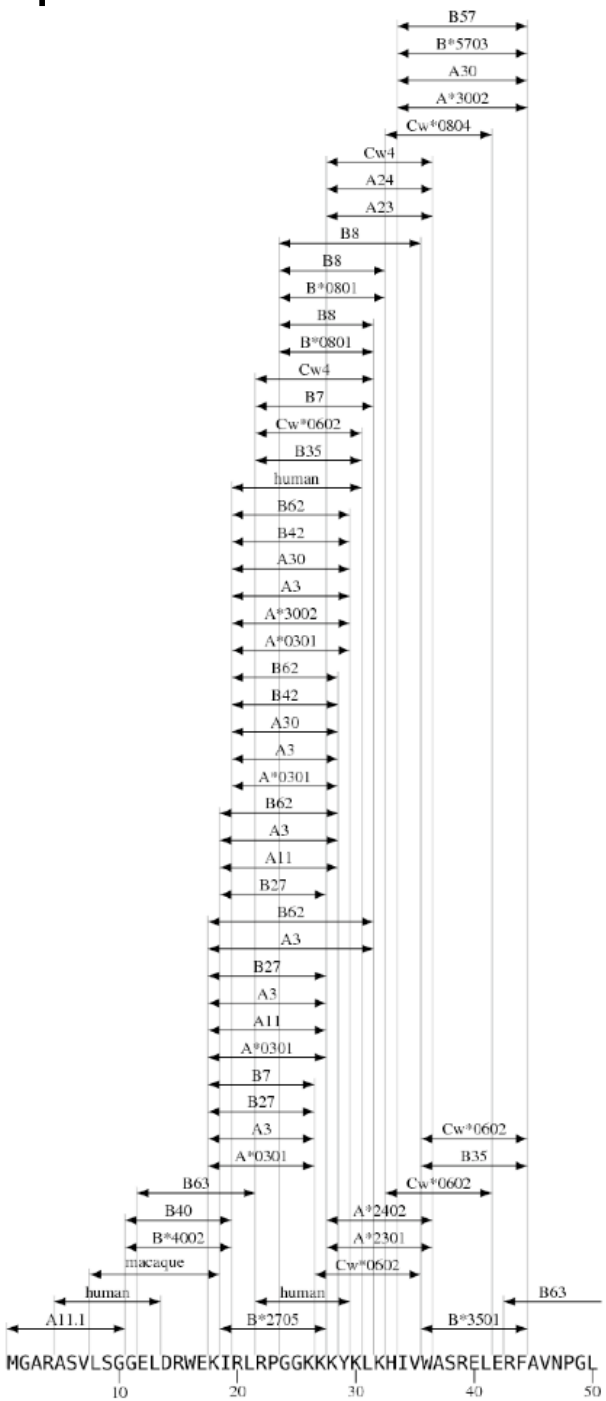
Shacklett2003 Barbara L. Shacklett, Catherine A. Cox, Johan K. Sandberg, Neil H. Stollman, Mark A. Jacobson, and Douglas F. Nixon. Trafficking of Human Immunodeficiency Virus Type 1-Specific CD8+ T Cells to Gut-Associated Lymphoid Tissue During Chronic Infection. *J. Virol.*, 77(10):5621-5631, May 2003. PubMed ID: 12719554. Show all entries for this paper.

Los Alamos NATIONAL LABORATORY

p17 CTL/CD8+ Epitope Map

p17 CTL/CD8+ Epitope Map

All HIV CTL, CD8+, epitopes mapped to within a region of 14 amino acids or less are indicated on the HIV protein epitope maps. The location and HLA restriction elements of CTL, CD8+, epitopes are indicated on protein sequences of HXB2. These maps are meant to provide the relative location of defined epitopes on a given protein, but the HXB2 sequence may not actually carry the epitope of interest, as it may vary relative to the sequence for which the epitope was defined. Epitopes with identical boundaries and HLA fields are included in the maps only once. If one laboratory determines HLA presenting molecules at the serotype level (example: A2) and another at the genotype level (example: A*0201) both will be included in the map. MHC specificities are indicative of the host species; when no MHC presenting molecule is defined, the host species is noted.
p17 CTL/CD8+ Map aa 1--50



Los Alamos NATIONAL LABORATORY

Antibody Search

Databases Search Tools Products Publications Search Site

Search Antibody Database

HIV Protein	Proteins with defined epitopes - ALL - p17 p17-p24 p24 p24-p2p7p1p6	Proteins with undefined epitopes - ALL - p17-p24 Gag RT Integrase
Epitope		
MAb ID	2F5 (List by name) (List by type)	
Subtype	- ALL -	
Immunogen	- ALL - anti-idiotype autoimmune disease HIV-1 exposed seronegative HIV-1 infection in vitro stimulation or selection SHIV infection	
Vaccine details	If Immunogen is Vaccine, additional search details Vaccine type - ALL - Vaccine strain - ALL - Vaccine component - ALL - Adjuvant - ALL -	
Ab Type	- ALL - C-domain C-HR C-term Env oligomer flap region gp120 adjacent to CD4BS	
Species	- ALL -	
Isotype	- ALL - IgA IgA1 IgA2 IgA22a IgE IgG	
Author	☐ First ☐ Last	
Country	- ALL -	
Keywords	- ALL - acute/early infection ADCC adjuvant comparison anti-idiotype antibody binding site definition and exposure antibody generation	

Search Reset

Los Alamos NATIONAL LABORATORY

Search Antibody Database

Found 1 matching record:

Displaying record number 815

MAb ID	2F5 (IAM 2F5, IAM-41-2F5, IAM2F5, c2F5)	
HXB2 Location	gp160(662-667)	gp160 Epitope Map
Author Location	gp41(662-667 BH10)	
Research Contact	Hermann Katinger, Institute of Applied Microbiology, Vienna, or Polymun Scientific Inc., Vienna, Austria	
Epitope	ELDKWA	Epitope Alignment
Ab Type	gp41 adjacent to cluster II, C-term, gp41 MPER (membrane proximal external region)	
Neutralizing	L P	
Species (Isotype)	human(IgG3κ)	
Immunogen	HIV-1 infection	
Keywords	acute/early infection, adjuvant comparison, antibody binding site definition and exposure, antibody sequence variable domain, assay standardization/improvement, binding affinity, co-receptor, escape, immunoprophylaxis, immunotherapy, kinetics, mimotopes, mother-to-infant transmission, neutralization, rate of progression, review, structure, subtype comparisons, vaccine antigen design, vaccine-specific epitope characteristics, variant cross-recognition or cross-neutralization, viral fitness and reversion	

Notes

- 2F5: Viruses from 304 days and at 643 days (time of death) post-infection of a macaque infected with SHIV SF162P4 were resistant to contemporaneous serum that had broadly reactive NAbs. While resistance to anti-V3, b12, and anti-V1 MAbs developed over time, viruses remained senstitive to 2F5 and 2G12. *Kraft2007* (**neutralization, escape**)
- 2F5: Binding of 2F5 to gp41 was not significanly affected by the small molecule HIV-1 entry inhibitor IC9564. IC9564 induces conformational change of gp120 to allow CD4i antibody 17b to bind, but inhibits CD4-induced gp41 conformational changes. *Huang2007*
- 2F5: Using sychronously infected cell cultures, the binding of b12, 2F5 and 2G12 to the cell-free virus interferes with a step of infection subsequent to cell attachment. HIV escape from b12 ocurred 30 and 10 min before escape from 2F5 for IIIB infection of HeLa cells and JRFL infection of Cf2Th-CD4/CCR5 cells, respectivelly, indicating that neutralization efficiency is determined by the time frames during which Ab can bind to the receptor-activated envelope proteins during the entry phase. 2F5 neutralization was enhanced by a decreasing the rate of coreceptor CXCR4 engagement, presumably by increasing the time the CD4 bound Env was available and slowing viral entry kinetics. *Haim2007* (**kinetics**)
- 2F5: Kinetics experiments of 2F5 binding to MPER region during viral fusion showed that the 2F5 kinetics resembled those of the six-helix bundle formation and fusion blocker C34, indicating that the function of MPER in the fusion cascade is still in effect at a late stage in the fusion reaction. Binding of 2F5 was shown to decrease upon triggering HIV-1 Env-expressing cells with appropriate target cells and addition of C34 did not counteract this loss, suggesting that changes in exposure of MPER occur independently of the six-helix bundle formation. *Dimitrov2007* (**antibody binding site definition and exposure, neutralization, kinetics, binding affinity**)
- 2F5: Chimeric SIV viruses containing 2F5 and 4E10 epitopes were not neutralized by broadly neutralizing sera from two clade B and one clade A infected asymptomatic individuals, indicating that MPER NAb epitopes did not account for the broad neutralizing activity observed. *Dhillon2007* (**antibody binding site definition and exposure, neutralization**)
- 2F5: SOSIP Env proteins are modified by the introduction of a disulfide bond between gp120 and gp41 (SOS), and an I559P (IP) substitution in gp41, and form trimers. The KNH1144 subtype A virus formed more stable trimers than did the prototype subtype B SOSIP Env, JRFL. The stability of gp140 trimers was increased for JR-FL and Ba-L SOSIP proteins by substituting the five amino acid residues in the N-terminal region of gp41 with corresponding residues from KNH1144 virus. b12, 2G12, 2F5, 4E10 and CD4-IgG2 all bound similarly to the WT and to the stabilized JRFL SOSIP timers, suggesting that the trimer-stabilizing substitutions do not impair the overall antigenic structure of gp140 trimers. *Dey2007* (**vaccine antigen design**)

Los Alamos NATIONAL LABORATORY

Antibody Search, 2

Databases Search Tools Products Publications Search Site

Search Antibody Database

HIV Protein	Proteins with defined epitopes: - ALL -, p17, p17-p24, p24, p24-p2p7p1p6 Proteins with undefined epitopes: - ALL -, p17-p24, Gag, RT, Integrase
Epitope	
MAb ID	polyclonal (List by name) (List by type)
Subtype	- ALL -
Immunogen	autoimmune disease HIV-1 exposed seronegative HIV-1 infection in vitro stimulation or selection SHIV infection vaccine virus
Vaccine details	If Immunogen is Vaccine, additional search details Vaccine type: - ALL - Vaccine strain: - ALL - Vaccine component: - ALL - Adjuvant: - ALL -
Ab Type	gp41 cluster I gp41 cluster II gp41 cluster III gp41 cytoplasmic domain gp41 internal trimeric coiled-coil of N-helices gp41 MPER (membrane proximal external region) gp41 six-helix bundle and the internal trimeric coiled-coil of N-helices
Species	- ALL -
Isotype	- ALL -, IgA, IgA1, IgA2, IgA22a, IgE, IgG
Author	First Last
Country	- ALL -
Keywords	- ALL - acute/early infection ADCC adjuvant comparison anti-idiotype antibody binding site definition and exposure antibody generation

Search Reset

Los Alamos NATIONAL LABORATORY

Search Antibody Database

Found 2 matching records:

Displaying record number 813

MAb ID	polyclonal	
HXB2 Location	gp160(662-667)	gp160 Epitope Map
Author Location	gp41(669-674)	
Epitope	ELDKWA	Epitope Alignment
Ab Type	C-domain, gp41 MPER (membrane proximal external region)	
Neutralizing		
Species (Isotype)	mouse, rabbit	
Immunogen	vaccine	
Keywords		

Vaccine Details

Vaccine type	peptide
Vaccine component	Env
Adjuvant	BSA

Notes

- Strong epitope-specific neutralizing antibody responses were induced using a Env peptide bound to BSA, C(ELDKWAG)_4-BSA, but not full gp160. *Xiao2000*

References

Xiao2000 Y. Xiao, M. Liao, Y. Lu, M. P. Dierich, and Y. H. Chen. Epitope-vaccines: a new strategy to induce high levels of neutralizing antibodies against HIV-1. *Immunobiology*, 201:323-31, 2000. PubMed ID: 10776789. Show all entries for this paper.

Displaying record number 818

MAb ID	polyclonal	
HXB2 Location	gp160(662-667)	gp160 Epitope Map
Author Location	gp41(662-667)	
Epitope	ELDKWA	Epitope Alignment
Ab Type	gp41 MPER (membrane proximal external region)	
Neutralizing	no	
Species (Isotype)	guinea pig	
Immunogen	vaccine	
Keywords		

Vaccine Details

Vaccine component	gp41

Notes

- 2F5: DP178 is a peptide derived from the C-term heptad repeat of gp41 that is a potent inhibitor of viral-mediated fusion -- it contains ELDKWA but fails to stimulate 2F5-like NAbs upon immunization -- the peptide was extended to force an increase in helicity, and the modified peptide had a increase in affinity for 2F5, but upon guinea pig immunization although high peptide-specific Ab titers were achieved the sera were incapable of viral neutralization -- the authors propose that 2F5 may be a low affinity maturation intermediate, which may account for its breadth and why it is hard to recreate the NAb response, but also suggests that the high concentrations required for neutralization are not relevant *in vivo*. *Joyce2002*

References

Joyce2002 Joseph G. Joyce, William M. Hurni, Michael J. Bogusky, Victor M. Garsky, Xiaoping. Liang, Michael P. Citron, Renee C. Danzeisen, Michael D. Miller, John W. Shiver, and Paul M. Keller. Enhancement of Alpha -Helicity in the HIV-1 Inhibitory Peptide DP178 Leads to an Increased Affinity for Human Monoclonal Antibody 2F5 but Does Not Elicit Neutralizing Responses in Vitro: Implications for Vaccine Design. *J. Biol. Chem.*, 277(48):45811-45820, 29 Nov 2002. PubMed ID: 12237296. Show all entries for this paper.

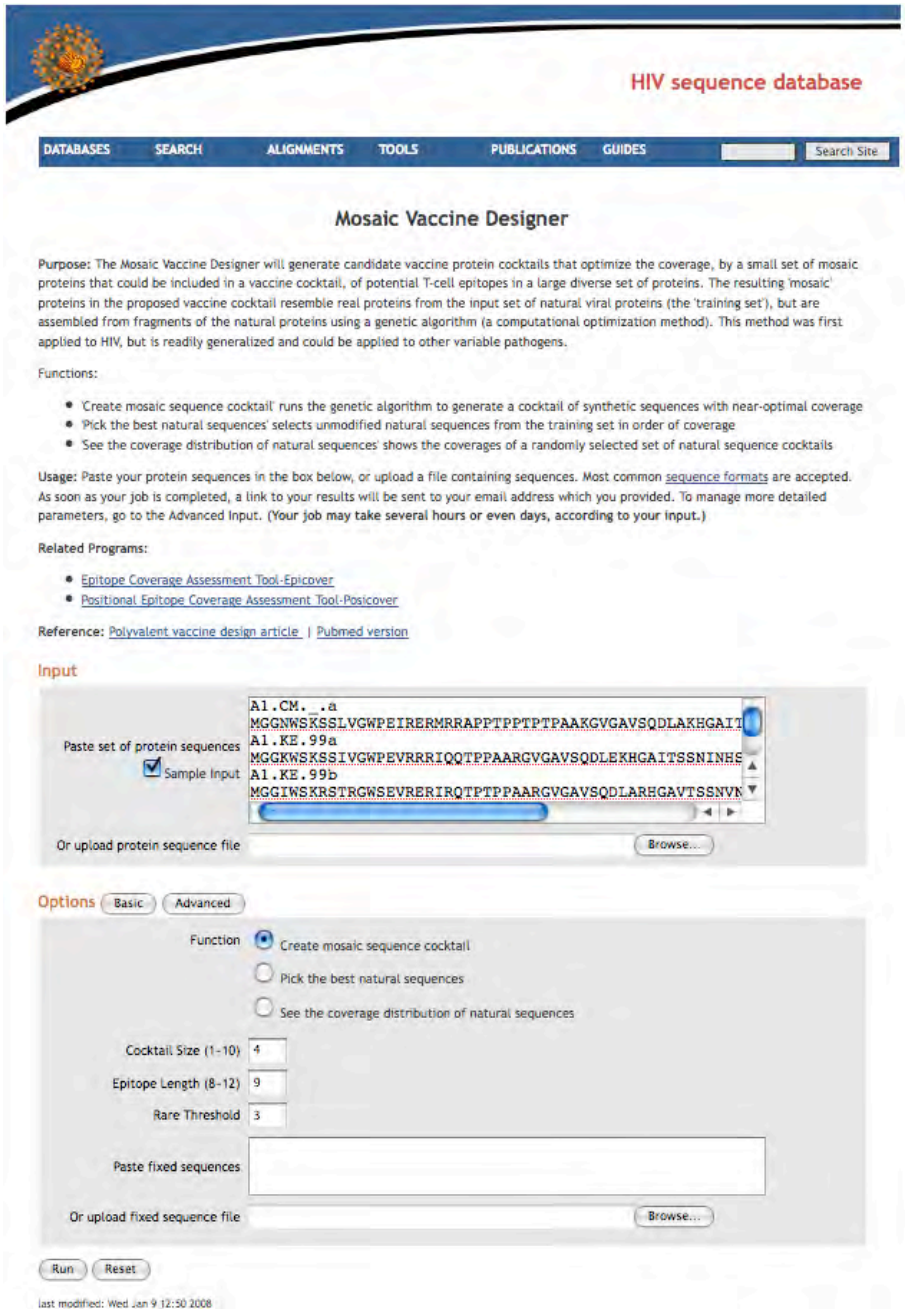
Los Alamos NATIONAL LABORATORY

Mosaic Vaccine Maker, Epicover, Posicover

- The mosaic vaccine maker is a specialized tool for people interested in designing a set of vaccine antigens using the strategy described in Fischer et al, Nat. Med. Dec. 2007.
  - People most likely to be interested:
    - Those interested in the method, who would like to apply it to a particular population of sequences (HIV or other) -- it yields intact proteins, maximizes coverage, minimizes unusual 9mers and excludes unnatural junctions.
    - Those who would like to compare their own vaccine candidate to this strategy
- Epicover and Posicover are general tools that can be used to assess coverage of N-mers (typically N=9) in a population of sequences relative to a proposed vaccine or peptide reagent set.
- Epicover does not require that the input sequences be aligned, and provides a overall summary, following a concept developed by Steve Self and colleagues to design test peptides based on coverage (PTE peptides, Li et al. Vaccine 24:6893 (2006))
- Posicover does require that the sequence population to be compared be aligned, but not the putative vaccine or peptides.

Los Alamos
NATIONAL LABORATORY

Mosaic Vaccine Maker: Input



Input: protein sequence set for a target population, does not need to be aligned.

Number of mosaic proteins in the set.

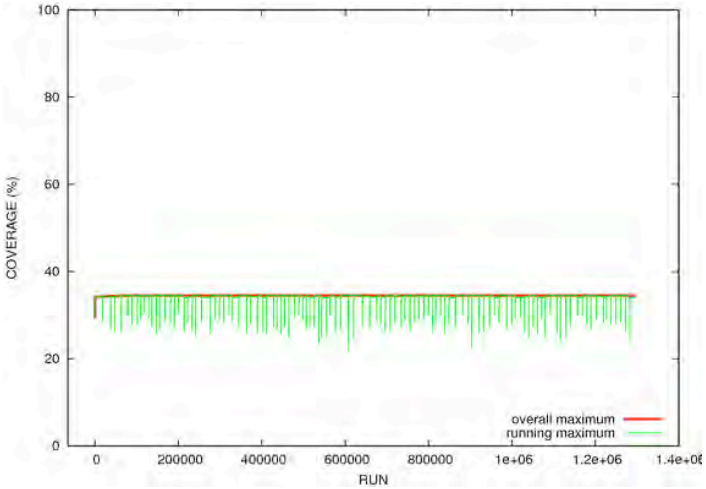
Epitope length.

Los Alamos NATIONAL LABORATORY

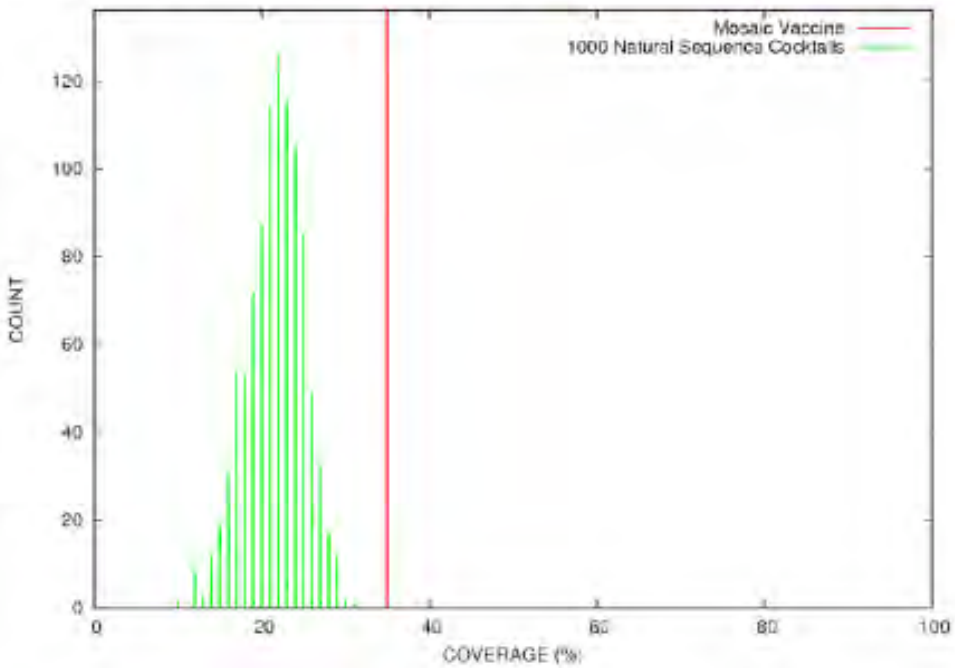
Mosaic Vaccine Maker: Output

Mosaic Vaccine Sequences: See table format
Coverage: 0.345111

>seq1
MGGKWSKSSIVGWPAIRERMRRAEPAAEGVGAVSQDLARHGAITSSNTAA
NNADCAWLQAQEEEE
>seq2
MGGKWSKSSVVGWPAVRERMRRAEPAADGVGAVSRDLERHGAITSSNTAA
TNADCAWLEAQEEDE
>seq3
MGSKWSKSSIVGWPAIRERMRRTEPAAEGVGAASRDLEKHGAITSSNTAA
NNPDCAWLEAQEEEE
>seq4
MGGKWSKSSIVGWPAVRERIRRTEPAAEGVGAASQDLDKHGAITSSNTAT
NNADCAWLEAQEDEE



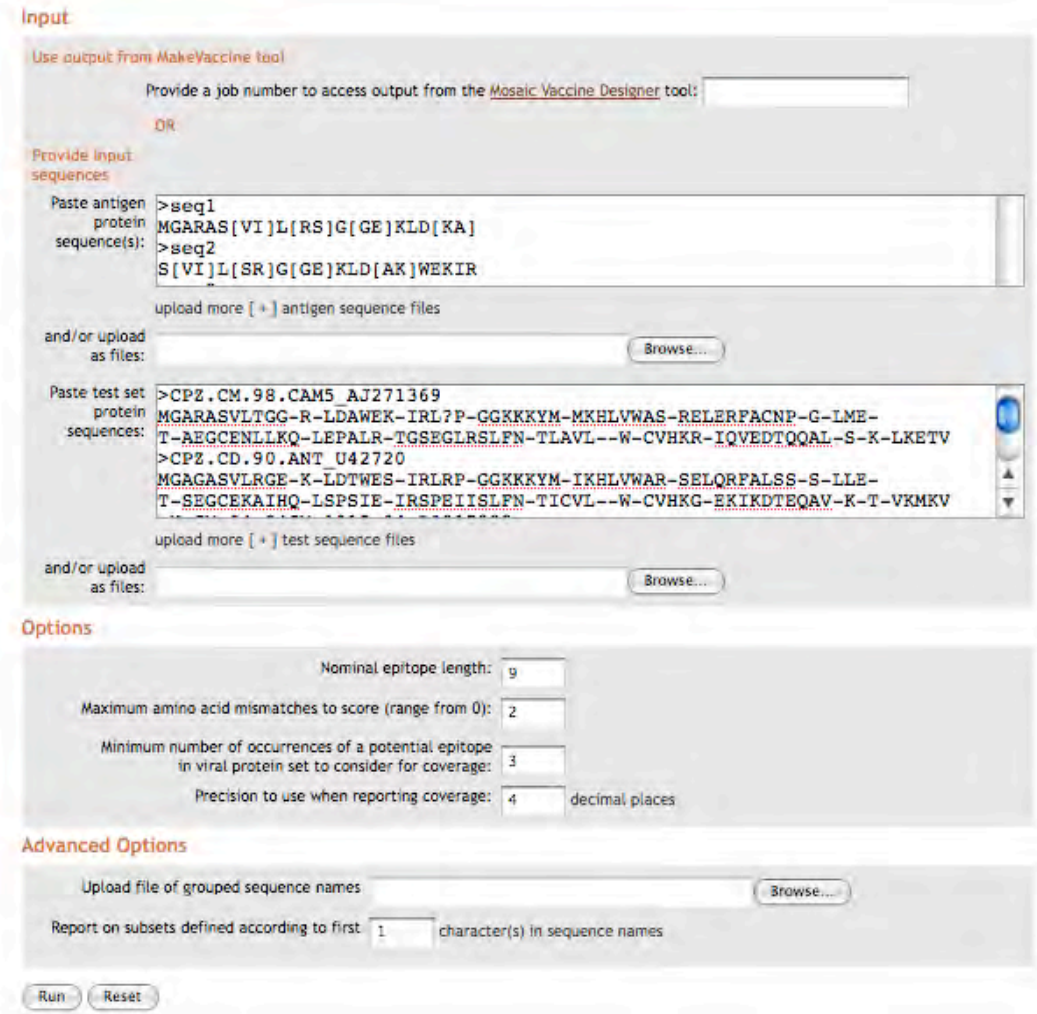
Compare with Natural Sequence Cocktails: See the raw data(coverages of natural sequences Cocktails)

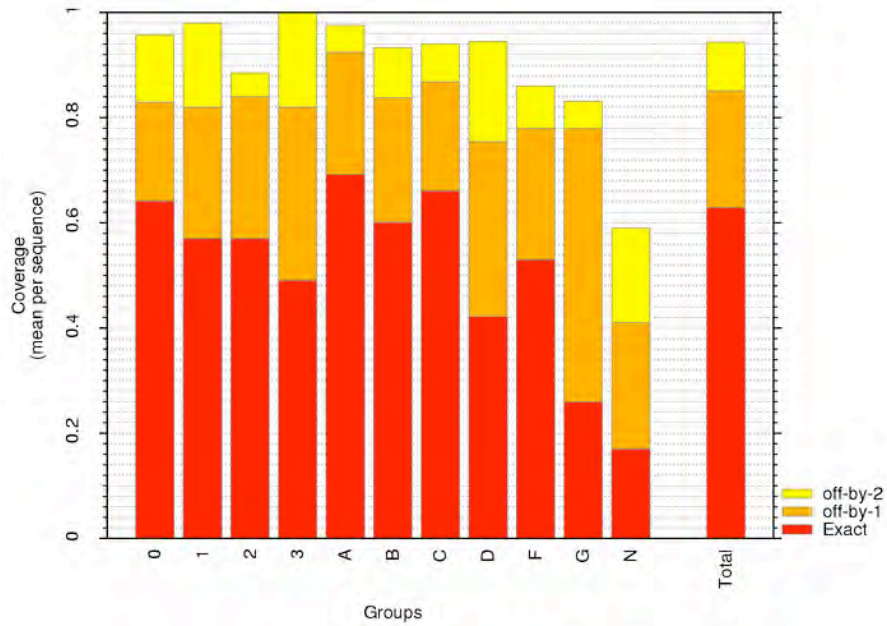


Mosaic coverage of 9-mers relative to combinations of natural strains

Los Alamos NATIONAL LABORATORY

Epicover: 9 mer coverage of vaccines or peptides

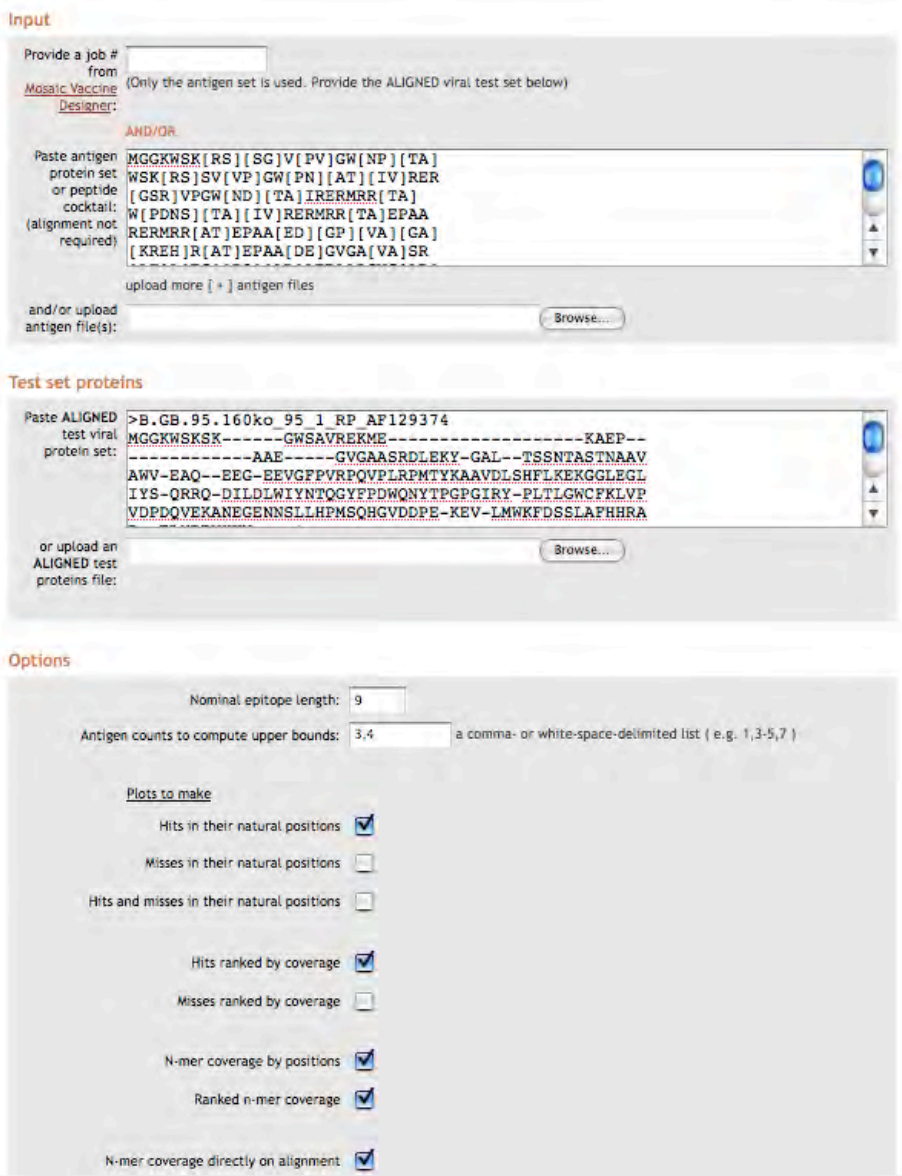




- Input 1: peptides or vaccines
- Input 2: sequences representing a population of interest (alignment independent)
- Output: above graphic and the numbers to generate it

Los Alamos NATIONAL LABORATORY

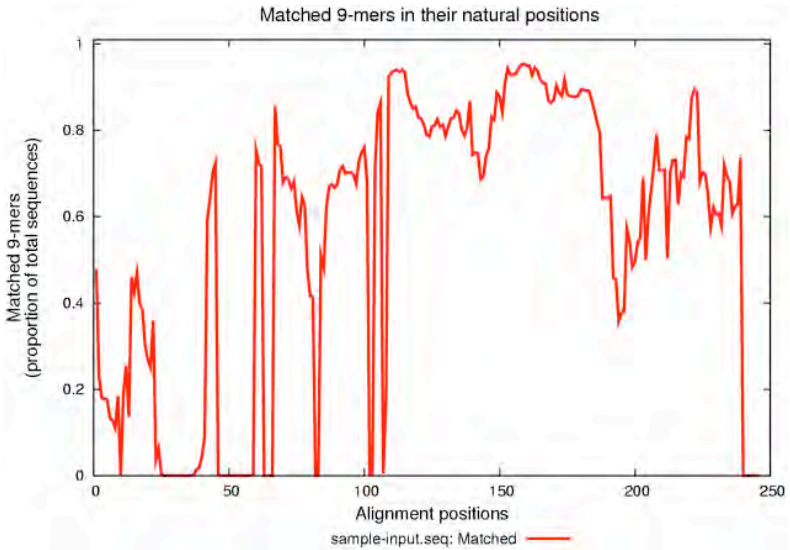
Posicover: 9 mer coverage of vaccines or peptides

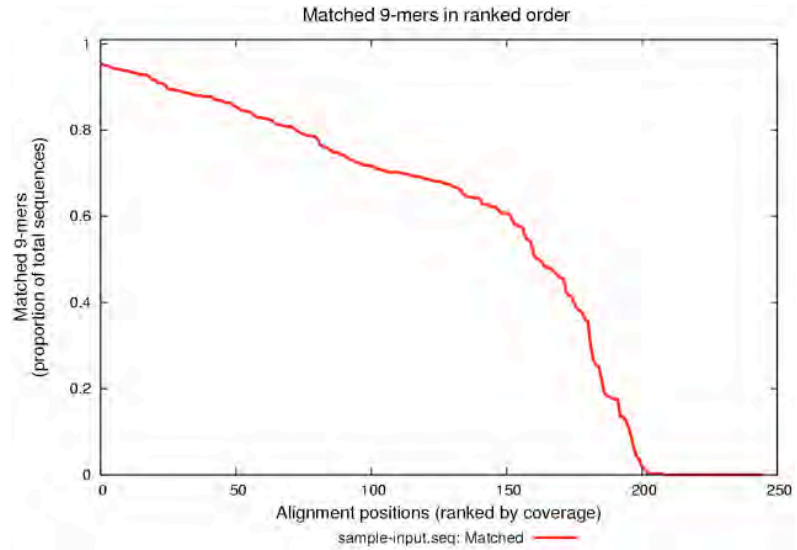


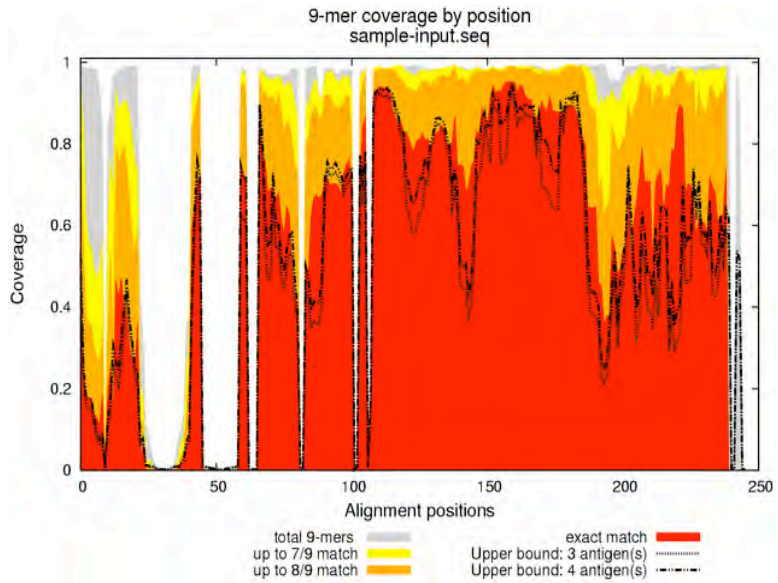
- Input 1: peptides or vaccines; toggling allowed.
- Input 2: sequences representing population of interest (alignment required)
- Output: a series of graphics

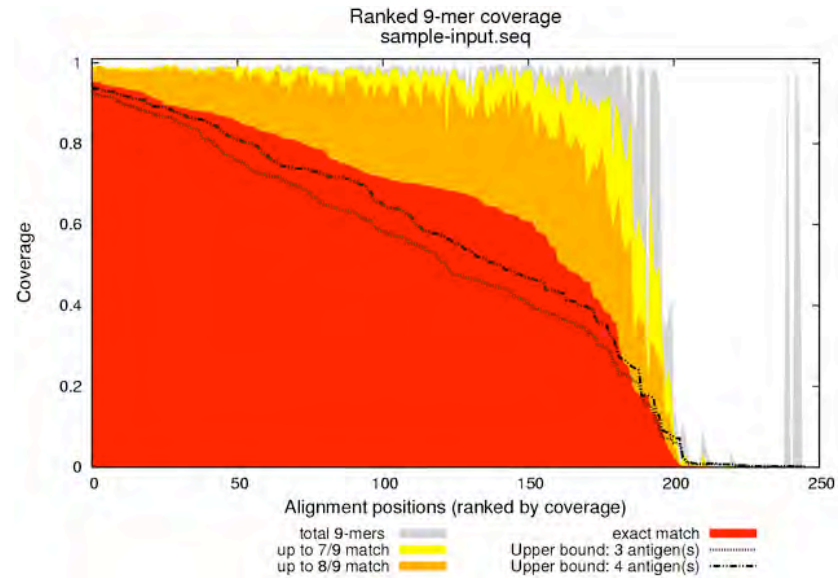
Los Alamos NATIONAL LABORATORY

Posicover output: I









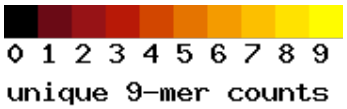
Los Alamos
NATIONAL LABORATORY

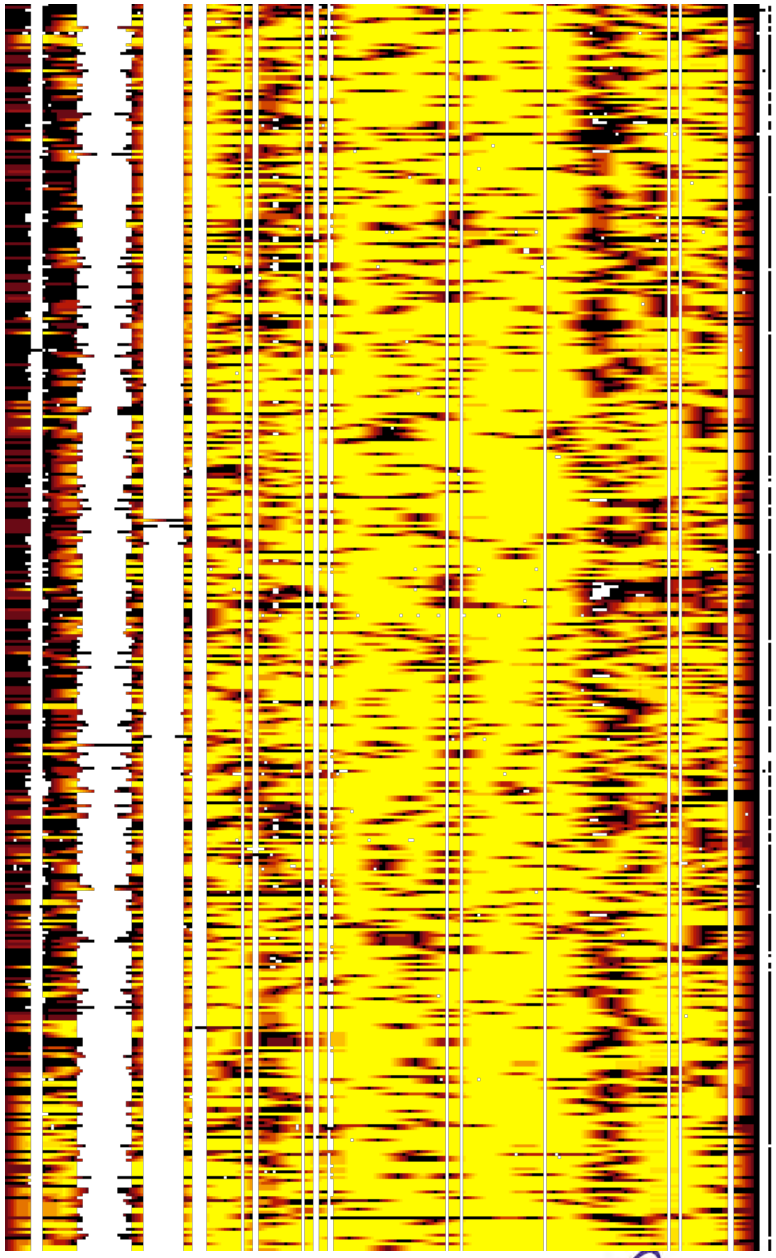
Posicover output: II

Each row of pixels is a sequence in the input alignment, each column a position.

White is a gappy region.

Yellow is good, the position is embedded in matched 9-mers.





Los Alamos NATIONAL LABORATORY

Highlighter

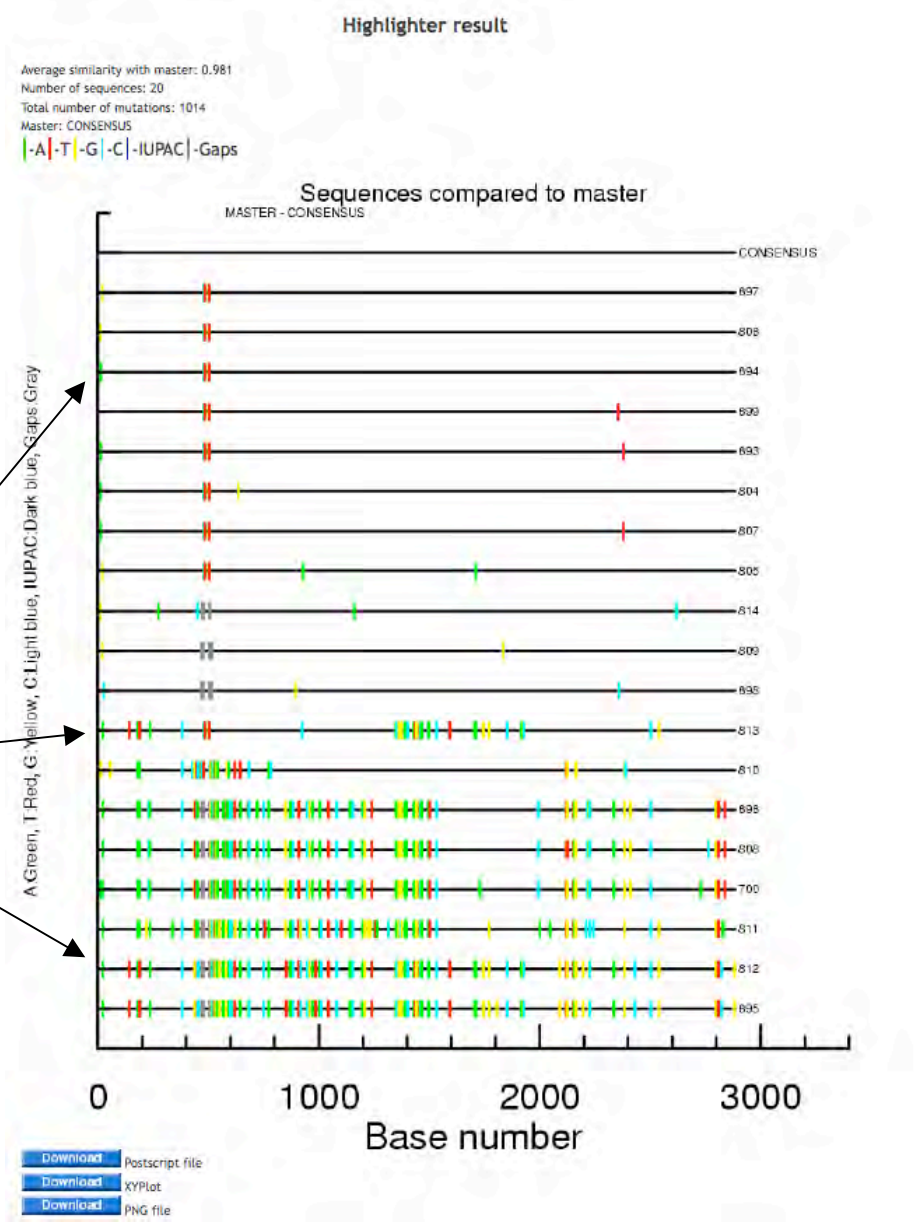
- Useful for comparing within-patient single or longitudinal samples
- The top of the alignment should be a consensus, MRCA, or most common form in the individual -- some good reference strain.
- Useful for:
  - tracking variation overtime in large sequence alignments
  - Identifying within patient recombination (a useful statistic will be added soon, currently it is visual).
  - Spotting apobec 3 signatures
  - Tracking locations of synonymous/non-synomous changes
  - Coming soon: ability to collapse all variation in a sample onto a single sequence for comparisons to the consensus.

Highlighter: input/output 1

Input: a within patient alignment

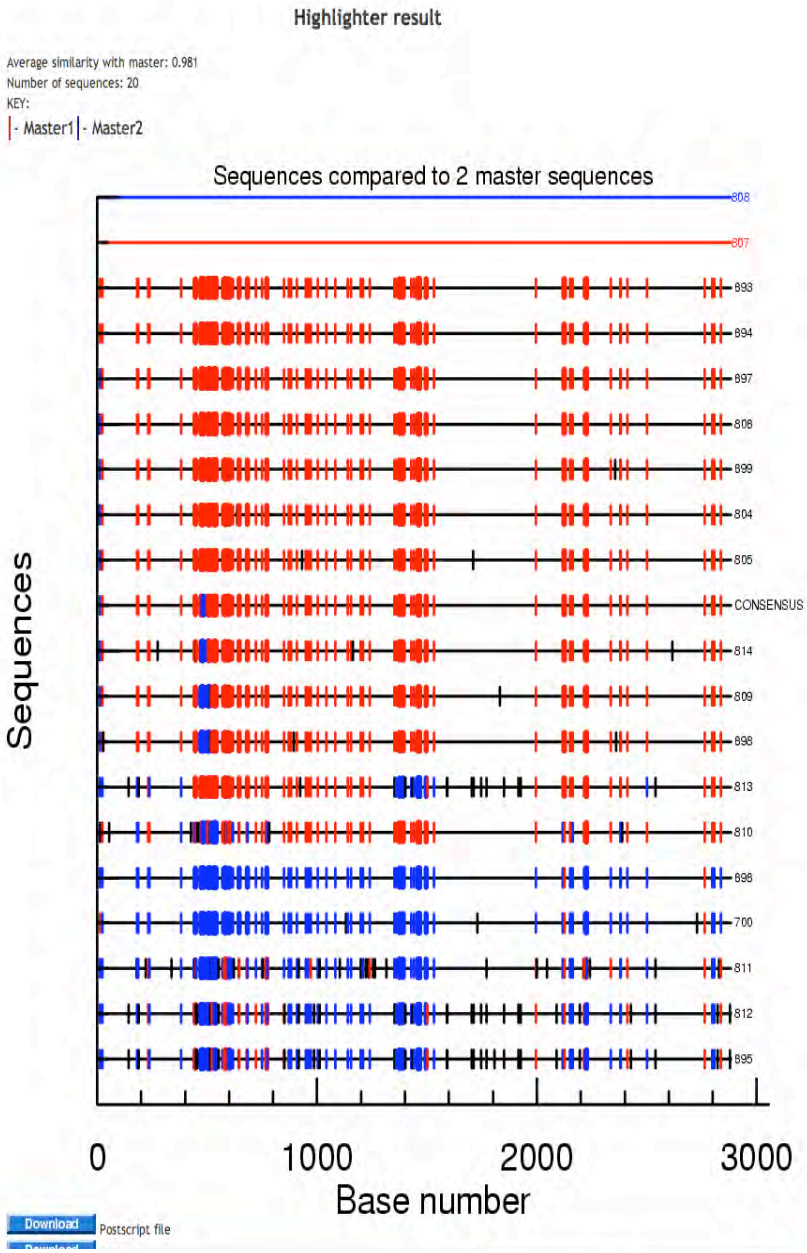
Output: default values, all changes
Relative to the “Master sequence”

Recombination:



Los Alamos NATIONAL LABORATORY

Highlighter: input/output II
Illustrating recombination



Select:
Change masters

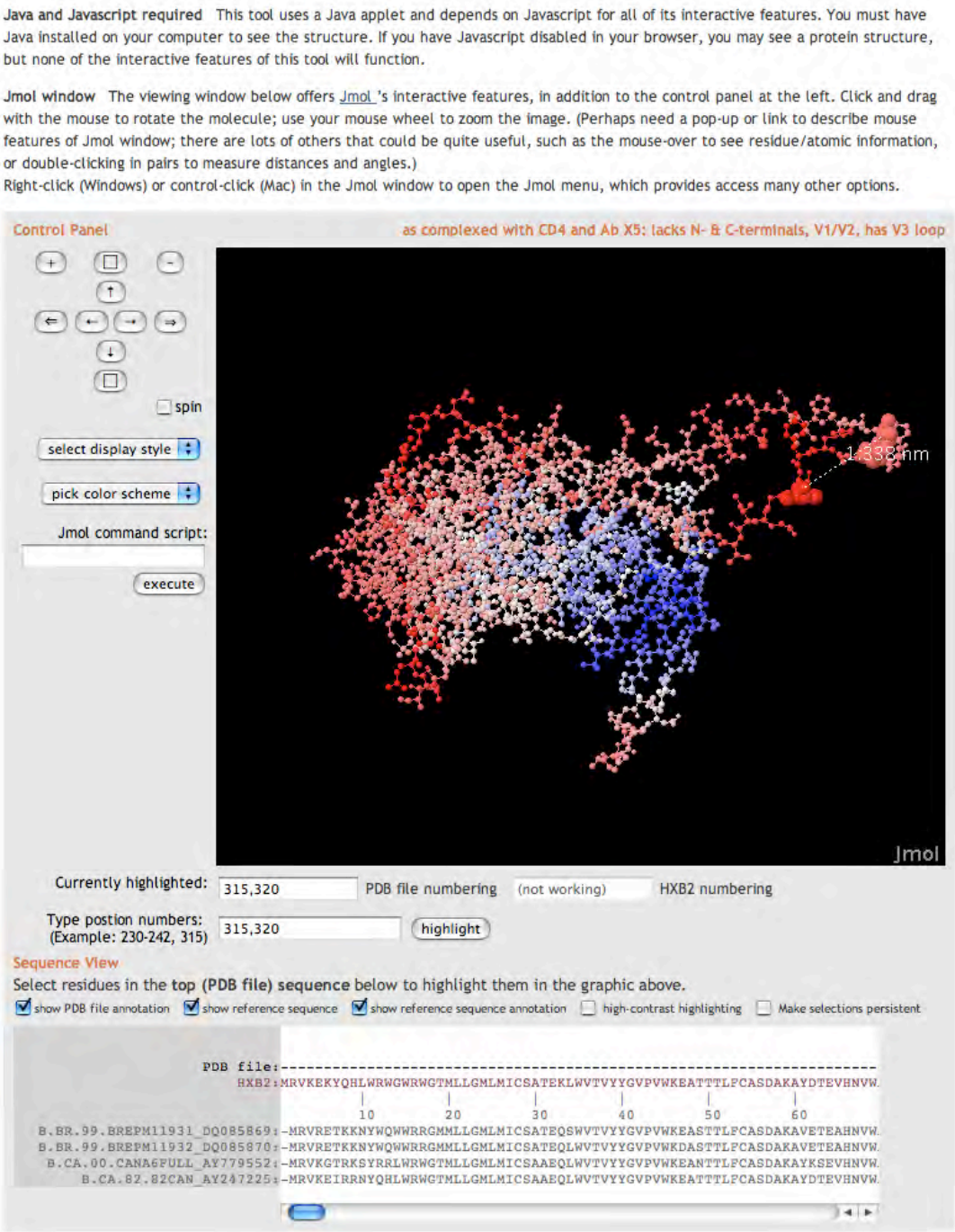
Pick two sequences:
3 and 14 are a good example

Los Alamos NATIONAL LABORATORY

Protein feature accent

- Brand new, still developing (expect changes weekly)
- Allows the user to *easily* highlight a region(s) of interest on an HIV protein -- so far, Env only.
- Automatically aligns your alignment to HXB2 and to the sections of the protein that are in a given structure
- In the works
  - Entropy mapping from an input alignment
  - Identifying all positions within a given distance of an amino acid of interest on the protein sequence
  - Selected structures representing other HIV proteins
  - Suggestions???

Protein feature accent



Los Alamos NATIONAL LABORATORY

HLA Comparison

- Statistically and graphically compare HLA frequencies in 2 populations
- Identify HLA alleles that are in linkage disequilibrium
- Input can be grouped as 2 digit typing

HLA Comparison: Compare 2 populations

HLA Comparison

NOTICE: Experimental software

This software is still under development and has not been fully tested. Please let us know what you think at immuno@lanl.gov.

This tool takes two series of tab-delimited HLA types and compares the HLA frequencies in the two sets. For more information, see details below, and try *Sample Input*.

Input

Population 1

Paste your HLA data here

Sample Input shown:

```
13W038510	A1101	A3401	B1521	B1801	C0403	C0704
13W038511	A0201	A1101	-	-	C0403	C0801
13W038507	A0101	A1101	B3502	B5106	-	-
13W038508	-	-	-	-	C0403	C1502
13W038509	A1101	A1101	-	-	-	-
13W038504	A020301	A2410	B1803	B3802	-	-
13W038505	A0201	A020301	-	-	C0303	C0703
13W038506	A1101	A2402	B3802	B4006	-	-
13W038501	A0207	A2402	B1801	B4601	C0102	C0704
13W038502	A1104	A2402	-	-	-	-
13W038503	A2402	A3401	-	-	C0403	C0801
```

Population 2

Paste your HLA data here

```
13W041495	A0201	A0301	B1402	B2705	C0202	C0802
13W041496	A0201	A0301	B4002	B5701	C0202	C06020101
13W041492	A0201	A3303	B0702	B380101	C0702	C1203
13W041493	A0101	A0201	B0702	B4402	C0501	C0702
13W041494	A0101	A0201	B0801	B2705	C0102	C0701
13W041489	A3002	A320101	B1302	B4402	C0501	C06020101
13W041490	A0201	A2601	B4001	B5501	C0303	C0303
13W041491	A0101	A320101	B0801	B4402	C0501	C0701
13W041486	A0101	A2501	B5101	B5501	C0303	C1502
13W041487	A0101	A0201	B0801	B3901	C0701	C0702
13W041488	A0101	A68020101	B0801	B3501	C0401	C0701
```

Options

Compare what class of HLA types? 4-digit HLA comparison

Sort the output by Sort by p-value

Submit Reset

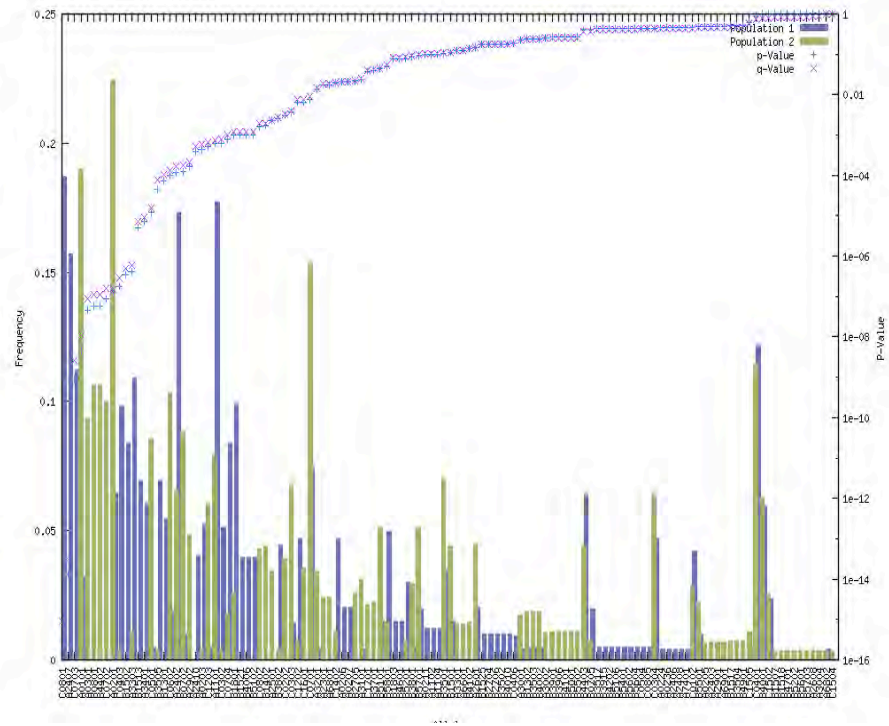
HLA Comparison

HLA Summary

	Population 1			Population 2		
	Distinct Alleles	Total Alleles	Population	Distinct Alleles	Total Alleles	Population
HLA-A	24	248	124	25	290	145
HLA-B	38	202	101	38	272	136
HLA-Cw	20	214	107	21	280	140
Total			131			145

Saw K=122 distinct alleles.

HLA Frequency Graph



Los Alamos NATIONAL LABORATORY

EPILIGN

- Generates an alignment of your HIV-1 amino acid sequence against our web alignments
- Can be used to align epitopes, functional domains, or any protein region of interest
- Sequence names include subtype, country and year of sampling

Example of output:

Query:	SLYNTVATL
Query Length:	9
HXB2 Location:	Gag 77-85 = p17 77-85
Alignment:	GAG, 458 sequences

Summarize

```
Query                    SLYNTVATL
A1.KE.86.ML170           --F------
A1.KE.94.Q23             --F------
A1.SE.94.SE7253          --F----V-
A1.SE.94.SE7535          ---------
A1.SE.95.SE8538          ---------
A1.SE.95.SE8891          ---------
A1.SE.95.UGSE8131        ---------
A1.TZ.97.97TZ03          --F----V-
```


Los Alamos NATIONAL LABORATORY

PeptGen

- Creates maps of overlapping peptides on proteins to aid in peptide design for mapping epitopes
- Consensus sequences for all HIV subtypes for all proteins are available
- Use alignments to design comparable sets of peptides (for example, to compare clades)
- INPUT
  - Query sequence or aligned sequences
  - Desired length of peptides, peptide overlap,
  - Forbidden C- and N-terminal amino acids (suggested by Dr. Philip Goulder, Oxford and Dr. Otto Yang, UCLA)
- OUTPUT
  - Maps of overlapping peptides (forbidden AAs are taken into account)
  - Simplified output for ordering
  - Highlighted forbidden amino acids
  - Hydropathicity scores for the peptides are available (suggested by Dr. Andrew Bradbury (LANL)

PeptGen: output

PeptGen Results

Download a copy of these results in format: Text PostScript PDF

```
                    Word length: 18
 Overlap consecutive peptides by: 10
                     Shorten by:
                    Lengthen by:
    Forbidden C-term amino acids:
    Forbidden N-term amino acids:
    Number of peptides generated: 9
                 Sequence names: CON_B
                                 CON_C
                                 CON_G


HIVWASRELERFAVNPGLLETSEGCRQILGQLQPSLQTGSEELRSLYNTVATLYCVHQRIEVKDTKEALEKIEEEQNKSK
HIVWASRELERFAVNPGL (18)
-L-----------L----
-L-----------L--D-
       LERFAVNPGLLETSEGCR (18)
       -----L-----------K
       -----L--D----A---Q
              GLLETSEGCRQILGQLQP (18)
              ---------K--IK----
              D----A---Q--M-----
                     CRQILGQLQPSLQTGSEE (18)
                     -K--IK----A----T--
                     -Q--M-----A----T--
                            QPSLQTGSEELRSLYNTV (18)
                            --A----T----------
                            --A----T------F---
                                   EELRSLYNTVATLYCVHQ (18)
                                   -----------------E
                                   ------F-----------
                                          TVATLYCVHQRIEVKDTK (18)
                                          ---------EK---R---
                                          ------------------
                                                 HQRIEVKDTKEALEKIEE (18)
                                                 -EK---R------D----
                                                 --------------EV-K
                                                        TKEALEKIEEEQNKSK (16)
                                                        -----D---------Q
                                                        ------EV-KI-K--Q
```

Duplicate

Simple output, “new”

PeptGen Results

Download a copy of these results in format:

HIVWASRELERFAVNPGL	1_1.1
HLVWASRELERFALNPGL	1_2.1
HLVWASRELERFALNPDL	1_3.1
LERFAVNPGLLETSEGCR	2_1.1
LERFALNPGLLETSEGCK	2_2.1
LERFALNPDLLETAEGCQ	2_3.1
GLLETSEGCRQILGQLQP	3_1.1
GLLETSEGCKQIIKQLQP	3_2.1
DLLETAEGCQQIMGQLQP	3_3.1
CRQILGQLQPSLQTGSEE	4_1.1
CKQIIKQLQPALQTGTEE	4_2.1
CQQIMGQLQPALQTGTEE	4_3.1
QPSLQTGSEELRSLYNTV	5_1.1
QPALQTGTEELRSLYNTV	5_2.1
QPALQTGTEELRSLFNTV	5_3.1
EELRSLYNTVATLYCVHQ	6_1.1
EELRSLYNTVATLYCVHE	6_2.1
EELRSLFNTVATLYCVHQ	6_3.1
TVATLYCVHQRIEVKDTK	7_1&3.2
TVATLYCVHEKIEVRDTK	7_2.1
HQRIEVKDTKEALEKIEE	8_1.1
HEKIEVRDTKEALDKIEE	8_2.1
HQRIEVKDTKEALEEVEK	8_3.1
TKEALEKIEEEQNKSK	9_1.1
TKEALDKIEEEQNKSQ	9_2.1
TKEALEEVEKIQKKSQ	9_3.1

Los Alamos NATIONAL LABORATORY

HLA Binding Motif Scanner: MotifScan

- Finds HLA anchor motifs within protein sequences for specified HLA genotypes, serotypes, or supertypes
- HLA anchor motif dictionaries are available on line
- Main motif and supermotif sources:
  - SYPHEITHI Database, Rammensee *et al*. www.syfpeithi.de
  - HLA Facts Book, Marsh *et al*. 2000
  - Sette & Sidney, *Immunogenetics* **50**:201-212, 1999
  - Isobella Honeyborne and Philip Goulder, Yusim *et al* 2003, (HLA-C motifs predicted by genetic similarities in HLA molecules)
- INPUT:
  - User defined query sequence or aligned sequences, or reference set
  - Selected HLA anchor motifs, or user defined motif
  - The user defined motif function could be used to search for other patterns of interest in sequences
- OUTPUT:
  - Anchor residue positions are highlighted in the query sequence
  - Potential epitopes and positions are listed
  - Output can be downloaded as text, convenient for further analysis

- **Can be used to finds all motifs in a sequence**

HIV HLA Anchor Residue Motifs (Motif Scan)

This page serves three functions: (1) find HLA anchor residue motifs within protein sequences for specified HLA genotypes, serotypes, or supertypes, (2) search any single protein for all known HLA anchor residue motifs, and (3) view motif libraries. For details, refer to the Motif Scan Help page.

Search Proteins for Specified Motifs

Select the motif type(s) of interest and submit. On the next page, you may choose HIV-1 consensus proteins to search, and/or enter your own protein sequences.

	Genotype	Serotype	Supertype
HLAs	A*01 A*0101 A*0201 A*0202 A*0204 A*0205 A*0206	A1 A2 A3 A11 A24(9) A25(10) A26(10)	A1 A2 A24 A3 B27 B44 B58
Motif Source	☑ Marsh2000 ☑ SYFPEITHI ☑ Others		
Motif Length	☐ 8 ☑ 9 ☐ 10 ☐ 11		
Custom Motif			

Search Reset

Search a Protein for All Motifs

Paste one protein sequence in raw format

Showing sample input

PIVQNIQGQMVHQAISPRTLNAWVKVVEEKAFSPEVIPMFSALSEGATPQ

Scan All Reset

View or Download Data Dictionaries

View or download the HLA genotype/serotype dictionary.

View or download the HLA genotype/motif dictionary.

Los Alamos NATIONAL LABORATORY

HIV HLA Anchor Residue Motifs (Motif Scan)

Binding Motif Scan

Your Motifs

Serotype	Genotype	Motif	Source
A2	A*0201	x-[L(M)]-x-x-x-x-x-x-[V(L)]	Marsh2000
A2	A*0201	x-[LM]-x-x-x-x-x-x-[VL]	SYFPEITHI

Search Patterns

x-[LM]-x-x-x-x-x-x-[VL]

Genotype	Serotype	Supertype	Motif	Source
A*0201	A2		x-[LM]-x-x-x-x-x-x-[VL]	SYFPEITHI

Matches (Download as FASTA)

```
>HXB2-Gag x-[LM]-x-x-x-x-x-x-[VL] A*0201
mgarasvlsg geldrwekir lrpggkkkyk lkhivwasre lerfavnpgl   50
letsegcrqi Lgqlqpslqt gseeLrsLyn tVatLycvhq rieikdtkea  100
ldkieeeqnk skkkaqqaaa dtghsnqvsq nypivqniqg qmvhqaispr  150
tLnawvkvVe ekafspevip mfsalsegat pqdLntmlnt Vgghqaamqm  200
lketineeaa ewdrvhpvha gpiapgqmre prgsdiagtt stlqeqigwm  250
tnnppipvge iykrwiilgl nkivrMyspt siLdirqgpk epfrdyvdrf  300
yktlraeqas qevknwmtet llvqnanpdc ktilkalgpa atleeMmtac  350
qgVggpghka rvLaeamsqV tnsatimmqr gnfrnqrkiv kcfncgkegh  400
tarncraprk kgcwkcgkeg hqmkdcterq anflgkiwps ykgrpgnflq  450
srpeptappe esfrsgvett tppqkqepid kelypltslr slfgndpssq  500
```


Los Alamos NATIONAL LABORATORY

HLA Binding Motif Scanner - Microsoft Internet Explorer

File Edit View Favorites Tools Links Address http://www.hiv.lanl.gov/content/immunology/motif_scan/motif_scan

Genotype	Serotype	Supertype	Motif	Source
A*0202	A2		x-[L]-x-x-x-x-x-[L]	Marsh2000

Matches (Download as FASTA)

```
                    10         20         30         40         50
                     |          |          |          |          |
Test.5B   ktiifkpssg gdpeivthsf ncggeffycn ttklfnstwn stwd----ln
Test.6B   ktiafkkssg gdpeivlhsf ncggeffhcn stqlfnstwn gnd-----td
Test.7B   ktivfnqssg gdpeivrhtf ncggeffycn tsqlfngtwn ----------

                    60         70         80         90        100
                     |          |          |          |          |
Test.5B   qtanh-egnd --titlpcri kqivnm-wqe vgkamyappi egqiacLshi
Test.6B   tketn-dteg --tiilpcri kqivnl-wqe vgkamyappi rgqircssni
Test.7B   ---nt-iens --titlpcri rqivnm-wqe vgkamyappi rgqircssni

                   110        120        130        140        150
                     |          |          |          |          |
Test.5B   tglLltrdgg -nd------- tnnetfrpgg gnmkdnwrse lykykvvkik
Test.6B   tgllltrdgg -ni--t---- neteifrpgg gdmrdnwrse lykykvvkie
Test.7B   tgllltrdgg -nng-s---- nttetfrpgg gnmrdnwrse lyrykvvkie

                   160        170        180        190        200
                     |          |          |          |          |
Test.5B   plgiaptkak rrvvqrekra v-gtlgamfl g-flgaagst mgaasvtl
Test.6B   plgvaptkak rrvvqrekra v-g-lgalfl g-flgaagst mgaasmtl
Test.7B   plgvaptkak rrvvqrekra v-gtigamfl g-flgaagst mgaasitl
```

List of potential epitopes (Download as TSV)

Protein	Seq. Pos.	Aln. Pos.	Sequence	Anchors
Test.5B	88-96	96-104	CLSHITGLL	.L.....L

References

Marsh2000 Steven G. E. Marsh, Peter Parham, and Linda D. Barber. *The HLA FactsBook*. Academic Press, San Diego, 2000.

SYFPEITHI The SYFPEITHI Database of MHC Ligands, Peptide Motifs and Epitope Prediction. Jan 2003. URL: http://syfpeithi.bmi-heidelberg.com/

Internet

ELF

- If you have a peptide that reacts with CD8+ T cells from a person with known HLA type, enter:
  - The peptide that reacts with CD8+ T-cells
  - The HLA type of the person with the reactive CD8+ T cells

- ELF will help identify the possibly reactive epitope by
  - Highlighting appropriate HLA anchor motifs in the peptide
  - Listing all known epitopes embedded in the peptide from the database, with links
  - Showing maps of known epitopes in the protein that carries the peptide, highlighting the input HLAs, to allow ready identification of options

ELF: Epitope Location Finder

This is a beta-test version of the HIV Molecular Immunology Database "**ELF**" site. Please e-mail comments, suggestions, and problems to the address at the bottom of the page.

Purpose:
ELF searches a submitted protein sequence for the occurrence of any epitopes known from our immunology database. Those epitopes whose HLA agrees with the submitted HLA(s) are flagged. Short peptides within the protein, "potential epitopes," whose sequence agrees with the binding motifs of the submitted HLAs are also displayed. Maps can be prepared that highlight every known epitope of the submitted HLA across the HIV proteome.

For details, see ELF Explanation file.

Usage:
Select one or more genotypes and/or serotypes from the lists. Then paste your HIV protein sequence (in raw format, i.e., only the amino acids themselves) into the text box. When "Show all epitopes" is checked, ELF will find all known epitopes in our database within the bounds of your submitted protein regardless of their HLA. (Any epitopes found which *do* agree with your submitted HLAs will be flagged.) If you uncheck this box, the program will find only epitopes whose serotype/genotype agrees with your submitted HLAs. If you want ELF to display maps that highlight every known epitope of the submitted HLAs across all HIV proteins, select the "All proteins" button. Similar maps can be displayed for only the protein containing you peptide sequence if you select the "Only this protein" button. The default is to draw no maps.

Input

Choose HLA

[Sample Input]

by genotype

A*01
A*0101
A*0201
A*0202
A*0204
A*0205
A*0206

by serotype

A1
A2
A3
A11
A24(9)
A25(10)
A26(10)

Paste protein sequence

Options

Show all known epitopes ☑

Draw maps? ◉ No ○ For only this protein ○ For all proteins

Run Reset

ELF Results

View Genomic location of your peptide.

View HLAs associated with your submitted HLA.

View all anchor motifs used in this analysis.

View potential "epitopes" ordered by HLAs.

View database records for known CTL epitopes in this region, regardless of HLA.

Epitopes from our database aligned to your query sequence

Bold letters colored **red** indicate residues in known epitopes which differ from the equivalent residues in the query sequence. The symbol ◄ means the epitope's HLA matches one of your submitted HLAs.

Download this alignment in format table fasta mase pretty

```
DTVLEDMNLPGRWKPKMIG
DTVLEEMNL A*6802  align
DTVLEDINL A*6802  align
DTVLEEWNL A*6802  align
DTVLEEMNL A68  align
    EEMNLPGRW B*44  align ◄
    EEMNLPGRW B44  align
```

Potential "epitopes" in your input sequence

These peptides have C-term anchor residues, highlighted in **blue**, and internal anchors highlighted in **magenta**. These anchor residues match one or more motifs associated with the submitted HLA, but are *not* found in our database.

Download this alignment in format table fasta mase pretty

```
DTVLEDMNLPGRWKPKMIG
DTVLEDMNL (A*0205 ........[L])
 TVLEDMNLP (A*0206 .[VQ].......)
     DMNLPGRWK (A*0301 .[LVMIAST]......[KYFR])
     DMNLPGRWK (B*2703 .[RM].......)
       NLPGRWKPK (A*0301 .[LVMIAST]......[KYFR])
          GRWKPKMIG (B*2703 .[RM].......)
          GRWKPKMIG (B*2705 .[RK].......)
```

Maps

View maps of all HIV-1 epitopes.

Please leave any comments or suggestions with us: