

# HIV Database Workshop

[www.hiv.lanl.gov](http://www.hiv.lanl.gov)

[seq-info@lanl.gov](mailto:seq-info@lanl.gov)

Presenter: Brian Foley

Database PI: Brian Foley

Additional database staff: Werner Abfalterer,  
Katherine Belobrajdic, Elizabeth-Sharon Fung,  
Kumkum Ganguly, Bette Korber, Jennifer Macke,  
James Szinger, Hyejin Yoon



Contract Officer Representative: Anjali Singh, NIAID, NIH

*Theoretical Biology and Biophysics, T-6  
Los Alamos National Laboratory*

<https://tinyurl.com/HIVDB-IEDB-2021>



# Workshop Topics

*Introduction to HIV Sequence Database*

*HIV in comparison to other viruses*

*Exploration of HIV evolution, using one Gag epitope as an example.*

Entry page at <https://www.hiv.lanl.gov/>



The **HIV databases** contain comprehensive data on HIV genetic sequences and immunological epitopes. The website also gives access to a large number of tools that can be used to analyze and visualize these data. This project has been funded in whole or in part with Federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under Interagency Agreement No. AAI12007-001-00000. Our content is reviewed by an [Editorial Board](#).

[SEQUENCE DATABASE ▶](#)

[IMMUNOLOGY DATABASE ▶](#)

[OTHER VIRUSES ▶](#)

**News:**

[Archived News ▶](#)

**[Variant Visualizer](#)**

Variant Visualizer is a new tool that provides new options and output styles for visualizing variants in an alignment of nucleotide or protein sequences. The tool is similar to [Highlighter](#), but the Variant Visualizer has more options for refining the graphical output. The output can be refined interactively without rerunning the tool. *21 July 2021*

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

## HIV sequence database

DATABASES SEARCH ALIGNMENTS TOOLS PUBLICATIONS GUIDES search site Search

Sequence Databases  
Immunology DB  
HCV DB  
HFV DB  
COVID-19 Genome Analysis Pipeline  
Programs and Tools

### HIV Sequence Database

#### Information

[Search Interface](#) retrieves HIV and SIV sequences, which can then be aligned and used to build trees

[Geography Search Interface](#) retrieves HIV sequences based on geographical distribution

[Genome Browser](#) uses jBrowse to display diverse data about the HIV-1 genome and proteome

[Tools for working with sequences](#) lists all our online tools, organized by function

#### Alignments

[HIV Premade Alignments](#) includes Consensus and Ancestral Sequences, Subtype Reference Alignments, and Complete Alignments

[HIV Sequence Compendium](#) print or order our annual publication

[Tutorials and other information](#) unpublished web-based content

[Links](#) to other HIV/AIDS tools and information

#### About this website

[FAQ](#) general information about this website

[How to Cite this Database](#)

[Editorial Board](#)

Multiple paths to most tools

#### News:

Archived News ▶

#### [Variant Visualizer](#)

Variant Visualizer is a new tool that provides new options and output styles for visualizing variants in an alignment of nucleotide or protein sequences. The tool is similar to [Highlighter](#), but the Variant Visualizer has more options for refining the graphical output. The output can be refined interactively without rerunning the tool. 21 July 2021

last modified: Tue Sep 7 15:54 2021

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

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



HIV sequence database

All kinds of basic information about HIV and about our database

DATABASES SEARCH ALIGNMENTS TOOLS PUBLICATIONS GUIDES search site Search

Tutorials and Basic Information

Previous workshop presentations

Tutorials

- [HIV Database presentations](#) from conference workshops
- [FAQs](#) about the HIV Sequence Database
- [FAQs](#) about the HIV Immunology Database
- [Sequence quality control](#) explains several common problems with sets of viral sequences
- [How to make a phylogenetic tree](#) explains how to build a phylogenetic tree
- [HIV numbering](#) relative to reference strain HXB2
- [SIV numbering](#) relative to reference strain SIVmm239

Reference Information

- [Circulating recombinant CRFs](#) details about all documented CRFs
- [HIV-1 gene map](#) illustrates the organization of HIV-1, including HXB2 breakpoints
- [HXB2 annotated spreadsheet \(.xls\)](#) provides a fully-annotated sequence of HXB2 with base-by-base detail
- [HIV and SIV subtype nomenclature](#) gives an overview of HIV and SIV subtype nomenclature, particularly HIV-1 groups and subtypes
- [Primate immunodeficiency virus nomenclature](#) lists SIV species and nomenclature
- [How the HIV database classifies sequences](#) explains how recombinants are named and annotated
- [Common sequence formats for alignments](#) shows examples of common sequence formats for alignments
- [How to cite this Database](#) explains how to cite this website and the printed HIV compendia
- [Codes and symbols in sequences](#) decodes the symbols and IUPAC codes that appear in sequences and alignments
- [Codon table](#) gives the translation of nucleotides into amino acids
- [Links](#) HIV/AIDS resources and bioinformatics tools on other websites

Articles

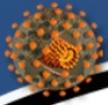
- [3D views of HIV macromolecular structures](#) provides links to 3D views of HIV proteins
- [Stalking the AIDS Virus \[PDF\]](#) article from LANL Research Quarterly (Fall 2003) about HIV Database research on the HIV-immune system interaction as a step toward an AIDS vaccine
- [Los Alamos research fundamental to first efficacy study for mosaic HIV-1 preventive vaccine](#) 2017 LANL news release about the HIV mosaic vaccine

Yes! We do respond to this e-mail address!

last modified: Wed Feb 28 14:38 2018

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





# HIV sequence database

DATABASES SEARCH ALIGNMENTS TOOLS PUBLICATIONS GUIDES search site Search

## Sequence Search Interface

### 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
- Use the + and - to see more or fewer search fields
- For other details about each field, see [Help](#) or [Data Dictionary](#)

Last [GenBank](#) update: 2021-09-20  
[Advanced Search](#)

### Sequence Information

[Upload accession file](#)  No file selected.

*Accession number*

*Sequence name*

*Sequence length*

exact  *Sampling year*

*Sampling country*

Virus

*Subtype*

- Any subtype
- No subtype
- A
- A1
- A2
- A3

Include [recombinants](#)

More sequence information

### Find all sequences for a specific gene or region (HIV-1, SIVcpz and SIVgor)

*Genomic region*

- Any
- complete genome
- 5' LTR
- 5' LTR R
- 5' LTR U3
- 5' LTR U5
- TAR

Or define *start*  and *end*

Include [fragments](#) of minimum length

### Combine database sequences with your own sequence alignment (HIV-1, SIVcpz and SIVgor)

### Publication Information

### Patient Information

### Geographical Information

### Amino Acid Motif Search

### Output

Include [problematic](#) sequences

[% of non-ACGT](#)

List  records per page

Show results selected  Show SQL

[Advanced Search](#)

last modified: Wed Aug 3 15:14 2016

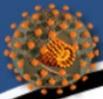
Questions or comments? Contact us at [seq-info@lanl.gov](mailto:seq-info@lanl.gov).

Almost all of our pages, tools, etc. have help files explaining the uses and providing help.

Sample data sets to test each tool are also provided.

Each of these sections of the search interface are expandable to provide additional search capabilities.





## HIV sequence database

DATABASES SEARCH ALIGNMENTS **TOOLS** PUBLICATIONS GUIDES

### HIV Database Tools

- Tools specific for HIV/SIV
- General use tools with some HIV/SIV-specific features
- General use tools

#### Analysis and Quality Control

- **Entropy** quantifies positional variation in an alignment using Shannon Entropy
- **GenSig** identifies genetic signatures. It can work on any phenotype file given in conjunction with a codon-aligned nucleotide alignment of a protein coding region
- **Glycan Shield Mapping** shows mapping absent hole-causing potential N-linked glycosylation sites (PNGS) on predicted glycan shields for an ENV sequence
- **HIV BLAST** finds sequences similar to yours in the HIV database
- **Hypermut** detects hypermutation
- **jpHMM at GOBICS** detects subtype recombination in HIV-1; hosted at GOBICS as a collaboration between the Department of Bioinformatics, University of Göttingen and the Los Alamos HIV Sequence Database
- **N-Glycosite** finds potential N-linked glycosylation sites
- **PCOORD** multidimensional analysis of sequence variation
- **Quality Control** runs several tools for quick troubleshooting of HIV-1 sequences; optional step prepares sequence submission for GenBank
- **RAPR** (Recombinant Analysis PRogram) uses the Wald-Wolfowitz Runs Test to check for recombination in every triplet in the alignment.
- **RIP** (Recombinant Identification Program) detects HIV-1 subtypes and recombination
- **SNAP** calculates synonymous/non-synonymous substitution rates
- **SUDI Subtyping** plots the distance of your sequence to established subtypes
- **Variable Region Characteristics** summarizes features of peptides from a protein alignment, reporting peptide characteristics such as length and N-linked glycosylation sites; particularly useful for characterizing variable regions of Env
- **VESPA** (Viral Epidemiology Signature Pattern Analysis) detects residues with different frequencies in two sequence sets

#### Alignment and sequence manipulation

- **Align Multi-tool** manipulates sequence alignments, including sorting, pruning, and renaming
- **Alignment Slicer** cuts vertical slices from sequence alignments
- **Analyze Align** shows weblogs, calculates frequency by position

#### Phylogenetics

- **AnnotateTree** creates a colored and weighted phylogenetic tree
- **Branchlength** calculates branch lengths between internal and end nodes; now included in the **TreeRate** tool
- **FindModel** finds which evolutionary model best fits your sequences
- **IQ-TREE** is a fast and effective stochastic algorithm for finding Maximum Likelihood trees, including site-specific rates of evolution at each alignment position
- **PhyloPlace** reports phylogenetic relatedness of an HIV-1 sequence with reference sequences
- **PhyML** generates much better trees than our simple TreeMaker tool
- **Poisson-Fitter** estimates time since MRCA and star-phylogeny. For use with acute (low diversity) samples
- **Rainbow Tree** Color code phylogenetic tree branches according to labels in the sequence names
- **TreeMaker** generates a Neighbor Joining phylogenetic tree
- **TreeRate** finds the phylogenetic root of a tree and calculates branch lengths and evolutionary rate

#### Immunology

- **CATNAP** (Compile, Analyze, Tally NAb Panels) provides meta-analysis of published neutralization panel data
- **CombiNAb** predicts and analyzes combination antibody neutralization scores using IC<sub>50</sub> and/or IC<sub>80</sub> for individual antibodies
- **ELF** (Epitope Location Finder) identifies known and potential epitopes within peptides
- **Epigraph Tool Suite** uses input of diverse sequences to generate Epigraph sequences for use in vaccine or reagent design
- **Epitign (QuickAlign)** aligns a protein sequence (e.g., epitope) to the appropriate protein alignment
- **Heatmap** displays a table of numbers by using colors to represent the numerical values
- **Hepitope** identifies potential epitopes based on HLA frequencies
- **Neutralization Index** computes a tier-like score for sera (using ID<sub>50</sub> titers) or antibodies (using IC<sub>50</sub> titers)
- **Mosaic Vaccine Tool Suite** designs and assesses polyvalent protein sequences for T-cell vaccines

Some of our tools are specific to HIV or lentiviruses, most can be used on DNA or Protein sequences from any organism.

Color-coded square indicates whether a given tool is specific to HIV or not.

Tools are listed alphabetically and organized into general categories.

Write to [seq-info@lanl.gov](mailto:seq-info@lanl.gov) for additional help, or to report problems or suggest new tools or features.

# All viruses evolve, but each has unique influences on its epidemiology and evolution

HIV, HCV and some other viruses persist in one host individual long after immune response is mature. Can be passed to recipient individual after significant evolution/selection in donor individual.

Influenza viruses, coronaviruses, and many other viruses typically infect a naïve host individual and are rapidly passed on to next recipient individual before much host immune selection pressure has taken place.

HIV has been infecting humans for 100 years or more, and has relatively slowly spread around the world. HCV has been infecting humans for more than 20,000 years and likewise has spread quite slowly.

Influenza viruses and coronaviruses can spread around the world in a matter of weeks.

Pending Filters

Reset Search

Filter Options ?

T Cell

Epitope ?

- Any
- Linear peptide
  - Length
  - Sequence
- Discontinuous
- Non-peptidic

3D structure available  
 Amino acid modification

Epitope Source ?

Organism  
**Human immunodeficiency 1**

Antigen  
Ex: core, capsid, myosin

Include related structure  
Select multiple options

TCR ?

Has TCR sequence

Type  
Any Type

Paired chains only

Chain  
Any Type

Sequence  
Exact Match

T Cell Assay ?

Outcome:  Positive

- Any
- Cytokine production
- MHC multimer
- In vivo

Ex: IL-2 release Find

Direct ex vivo detection

MHC Restriction ?

- Any
- Class I
- Class II
- Non-classical

Ex: HLA-A\*02:01 Find

Resolution

Current Filters:  Epitope Structure: Linear Sequence  Organism: Human immunodeficiency virus 1 (human immunodeficiency virus 1 HIV-1) (ID:11676, human immunodeficiency virus HIV-1)

Include Positive Assays  No B cell assays  No MHC assays  Host: Homo sapiens (human)

Please see [HIV Molecular Immunology Database](#) for more information.

Epitopes (257)		Antigens (7)		Assays (892)		Receptors (419)		References (89)	
Go To Records Starting At 1200									
257 Records Found									
Details	Epitope	Antigen	Organism	# References	# Assays				
59613	<a href="#">SLYNTVATL</a>	Gag polyprotein	Human immunodeficiency virus 1 (human immunodeficiency virus 1 HIV-1)	26	56				
27125	ILKEPVHGV	Gag-Pol polyprotein	Human immunodeficiency virus 1 (human immunodeficiency virus 1 HIV-1)	10	13				
29804	KAFSPEVIMPF	Gag polyprotein	Human immunodeficiency virus 1 (human immunodeficiency virus 1 HIV-1)	8	30				
33250	KRWILGLNK	Gag polyprotein	Human immunodeficiency virus 1 (human immunodeficiency virus 1 HIV-1)	7	15				
127246	TSTLQEQIGW	Gag-Pol polyprotein	Human immunodeficiency virus 1 (human immunodeficiency virus 1 HIV-1)	7	11				
12616	EIYKRWII	Gag polyprotein	Human immunodeficiency virus 1 (human immunodeficiency virus 1 HIV-1)	6	10				
102104	TAFTIPSI	Gag-Pol polyprotein	Human immunodeficiency virus 1 (human immunodeficiency virus 1 HIV-1)	6	20				
131070	<a href="#">SLFNTVATL</a>	Gag polyprotein	Human immunodeficiency virus 1 (human immunodeficiency virus 1 HIV-1)	6	17				
1401	AFSPEVIMFMSALSEGATPQ	Gag-Pol polyprotein	Human immunodeficiency virus 1 (human immunodeficiency virus 1 HIV-1)	5	45				
29352	IVLPEKDSW	Gag-Pol polyprotein	Human immunodeficiency virus 1 (human immunodeficiency virus 1 HIV-1)	4	6				
56620	RYPLFGWCF	Protein Nef	Human immunodeficiency virus 1 (human immunodeficiency virus 1 HIV-1)	4	19				
59612	<a href="#">SLYNTIATL</a>	Gag-Pol polyprotein	Human immunodeficiency virus 1 (human immunodeficiency virus 1 HIV-1)	4	10				
189277	<a href="#">SLFNTIATL</a>	Gag polyprotein	Human immunodeficiency virus 1 (human immunodeficiency virus 1 HIV-1)	4	16				
189288	<a href="#">SLYNTIATL</a>	Gag-Pol polyprotein	Human immunodeficiency virus 1 (human immunodeficiency virus 1 HIV-1)	4	14				
189293	<a href="#">SLYNTVAVL</a>	Gag-Pol polyprotein	Human immunodeficiency virus 1 (human immunodeficiency virus 1 HIV-1)	4	8				
21635	GPGHKARVL	Gag polyprotein	Human immunodeficiency virus 1 (human immunodeficiency virus 1 HIV-1)	3	7				
28657	ISPRTLNAW	Gag polyprotein	Human immunodeficiency virus 1 (human immunodeficiency virus 1 HIV-1)	3	3				
32201	KLTPLCVTL	Envelope glycoprotein gp160	Human immunodeficiency virus 1 (human immunodeficiency virus 1 HIV-1)	3	14				
34482	KYKLGHIWV	Gag polyprotein	Human immunodeficiency virus 1 (human immunodeficiency virus 1 HIV-1)	3	4				
101990	QASQEVKRW	Gag polyprotein	Human immunodeficiency virus 1 (human immunodeficiency virus 1 HIV-1)	3	6				
180191	<a href="#">SLFNTIATL</a>	Gag polyprotein	Human immunodeficiency virus 1 (human immunodeficiency virus 1 HIV-1)	3	8				
180236	<a href="#">SLFNTVAVL</a>	Gag polyprotein	Human immunodeficiency virus 1 (human immunodeficiency virus 1 HIV-1)	3	7				
1129	AEQASQDVKNW	Gag polyprotein	Human immunodeficiency virus 1 (human immunodeficiency virus 1 HIV-1)	2	4				
5295	AVDSLHFLK	Protein Nef	Human immunodeficiency virus 1 (human immunodeficiency virus 1 HIV-1)	2	7				
9974	DRFYKTLRA	Gag polyprotein	Human immunodeficiency virus 1 (human immunodeficiency virus 1 HIV-1)	2	2				

Searched the IEDB for CTL epitopes in HIV-1

Gag SLYNTVATL SLFNTVATL etc are found

How variable is this epitope in HIV-1?



## HIV sequence database

DATABASES	SEARCH	ALIGNMENTS	TOOLS	PUBLICATIONS	GUIDES	search site	Search
			Index of all tools	Glycan Shield Mapping	Protein Feature Accent		
			Align Multi-tool	Heatmap	Quality Control		
		HIV	Alignment Slicer	Hepitope	QuickAlign		
			AnalyzeAlign	Highlighter	Rainbow Tree		
			AnnotateTree	HIV BLAST	RAPR		
			Branchlength	HIVAlign	Recombinant HIV-1 Drawing Tool		
			CATNAP	Hypermute	RIP	publication	
			Codon Alignment	IQ-TREE	SeqPublish	ed content	
			CombinAber	jpHMM at GOBICS	Sequence Locator		
			Consensus Maker	Mosaic Vaccine Tool Suite	SNAP		
			ELF	Motif Scan	SUDI Subtyping		
			ElimDupes	N-Glycosite	SynchAlign		
			Entropy	Neutralization Index	Translate		
			Epigraph	PCOORD	TreeMaker		
			FindModel	PepMap	TreeRate		
			Format Converter	PeptGen	Variable Region Characteristics		
			Gap Strip/Squeeze	PhyloPlace	Variant Visualizer		
			GenBank Entry Generation	PhyML	VESPA		
			Gene Cutter	Pixel	External Tools		
			Genome Browser	Poisson-Fitter			
			GenSig	PrimerDesign-M			

**Programs and Tools**

[Search Interface](#) retrieves HIV and SIV sequences, which can be aligned and used to build trees

[Geography Search Interface](#) retrieves HIV sequences based on geographical distribution

[Genome Browser](#) uses jBrowse to display diverse data about HIV-1 genome and proteome

[Tools for working with sequences](#) lists all our online tools, by function

**Alignments**

[HIV Premade Alignments](#) includes Consensus and Ancestral Sequences, Subtype Reference Alignments, and Complete Alignments

**News:**

No new news.

[Archived News](#) ▶

HIV Sequence Database  
Genome Browser

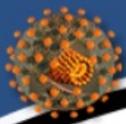
View genome or each  
protein to see features of  
interest

last modified: Tue Sep 7 15:54 2021

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

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





## HIV Genome Browser

**Purpose:** Interactive view of the HIV genome and proteome for juxtaposition and exploration of multiple types of data. [Help](#).

### Starting Views

**NOTE:** These are just starting points! Within the genome browser, you can move among any of these views. Please read the quick tips and Help file before you start!

#### HIV-1 protein-level views:

- [Env](#) • [Gag](#) • [Nef](#) • [Pol](#) • [Rev](#) • [Tat](#) • [Vif](#) • [Vpr](#) • [Vpu](#)

#### HIV-1 proteins, specific examples:

- [Env with CTL epitopes + entropy](#)
- [Pol with drug resistance sites + entropy](#)

#### Nucleotide-level views:

- [HIV-1 gene map](#)
- [SIV Mac239 gene map](#)
- [HIV-1 5' LTR](#)

### Quick Tips

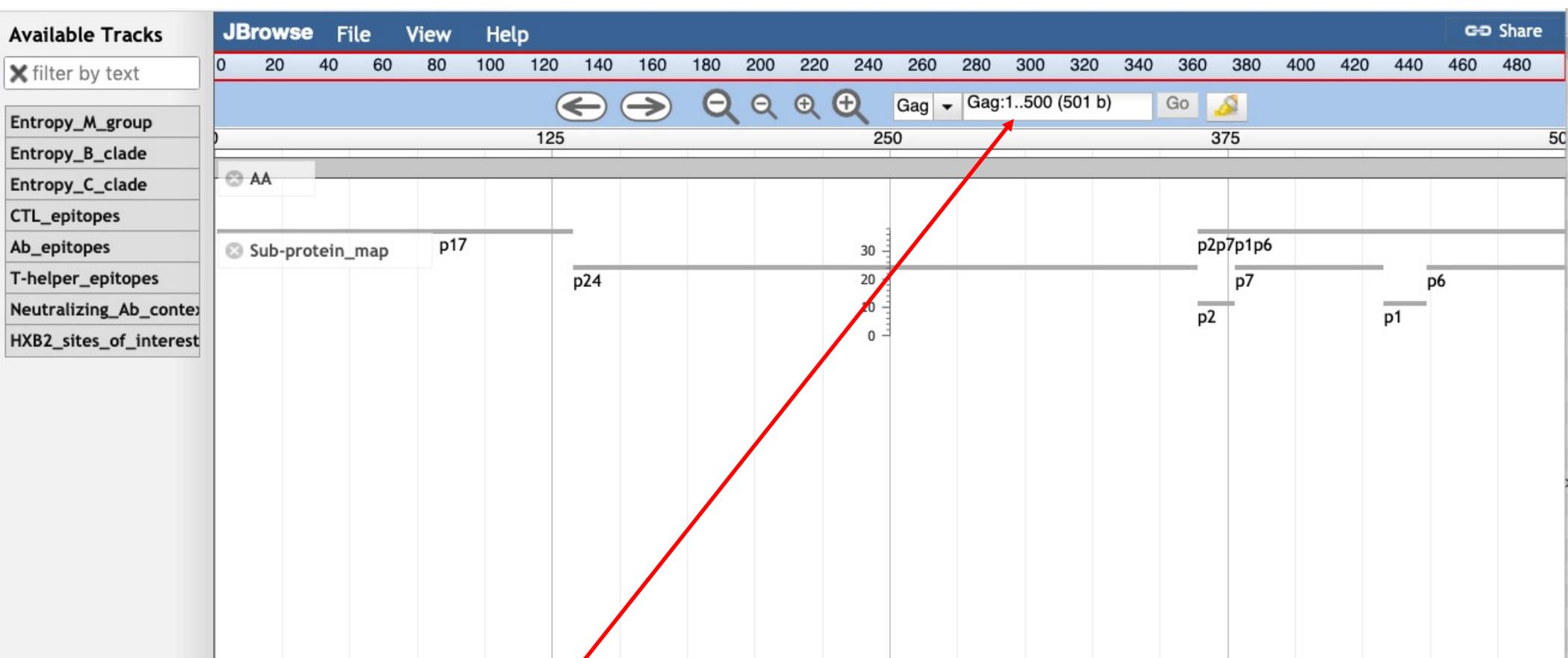
- **Mouseovers!** Look for mouseovers to guide you.
- **Click and right-click!** Features link to loads of information and analysis via click and right-click. If your mouse doesn't have right-click, use Ctrl-click.
- **Zoom!** There are several ways to zoom in/out. Some features can only be seen when zoomed-in or zoomed-out.
- For details about this interface, see [HIV Genome Browser Help](#).
- Watch the screencast video on the [JBrowse website](#).

### References

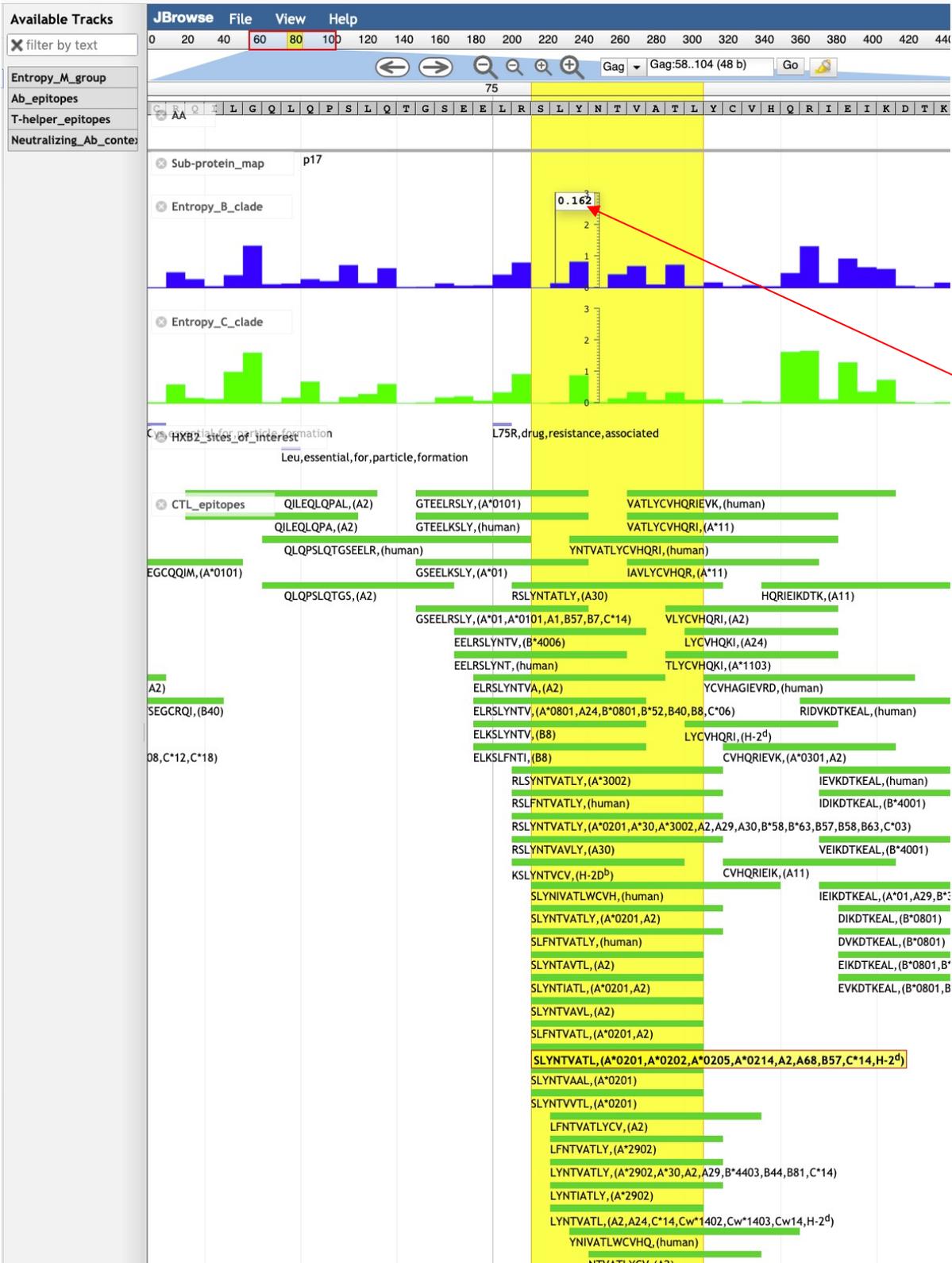
- Skinner ME, Holmes IH. Setting up the JBrowse genome browser. Curr Protoc Bioinformatics. 2010 Dec;Chapter 9:Unit 9.13. PMID: 21154710

HIV Genome Browser

Select Gag to see  
Gag protein view



Enter SLYNTVATL to zoom in to that site in the Gag protein.



HIV Genome Browser has tracks for entropy of each site

SLYNTVATL third amino acid, Y, has high entropy

Databases

Search

Tools

Products

Publications

search site

Search Site

## Search CTL/CD8+ T-Cell Epitope Database

<a href="#">HIV protein</a>	- ALL - Gag p17 p24 p2p7p1p6	
<a href="#">HXB2 protein location</a>	<input type="text"/> -- <input type="text"/>	Results overlap with query location ▾
<a href="#">HXB2 DNA location</a>	<input type="text"/> -- <input type="text"/>	Results overlap with query location
<a href="#">Epitope</a>	<input type="text" value="SLYNTVATL"/>	Results contain query sequence ▾
<a href="#">Epitope name</a>	<input type="text"/>	
<a href="#">Record number</a>	<input type="text"/>	
<a href="#">Subtype</a>	- ALL - ▾	
<a href="#">Immunogen</a>	- ALL - computer prediction engineered HIV-1 and HCV co-infection HIV-1 exposed seronegative HIV-1 infected monocyte-derived HIV-1 infection	
<a href="#">Vaccine details</a> if Immunogen is Vaccine	<a href="#">Vaccine type</a> - ALL - ▾ <a href="#">Vaccine strain</a> - ALL - ▾ <a href="#">Vaccine component</a> - ALL - ▾ <a href="#">Adjuvant</a> - ALL - ▾	
<a href="#">Species</a>	- ALL - ▾	
<a href="#">MHC/HLA</a>	- ALL - A*01 A*01:01 A*01:23 A*02 A*02:01 A*02:02	
<a href="#">Author</a>	<input type="text"/>	<input type="checkbox"/> First <input type="checkbox"/> Last
<a href="#">Country</a>	- ALL - ▾	
<a href="#">Keywords</a>	- ALL - acute/early infection adjuvant comparison antagonism antibody binding site antibody generation assay or method development	
<a href="#">Note</a>	<input type="text"/>	

  Click for [Search Help](#)
[Search CTL/CD8+ variants](#)

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

 HIV Immunology Database  
 CTL Search for SLYNTVATL

### Search CTL/CD8+ T-Cell Epitope Database

Found 301 matching records:

#### Displaying record number 57

Download this epitope [record as JSON](#).

<a href="#">HXB2 Location</a>	Gag(69-93) p17(69-93) DNA(994..1068)	<a href="#">Gag Epitope Map</a>
<a href="#">Author Location</a>	p17(69-93 BH10)	
<a href="#">Epitope</a>	QTGSEELRSLYNTVATLYCVHQRIE	<a href="#">Epitope Alignment</a>
<a href="#">Species (MHC/HLA)</a>	human(A2)	
<a href="#">Immunogen</a>	HIV-1 infection	
<a href="#">Experimental methods</a>		
<a href="#">Keywords</a>		

#### Notes

- Gag CTL response studied in three individuals.

#### References

Johnson 1991 R. P. Johnson, A. Trocha, L. Yang, G. P. Mazzara, D. L. Panicali, T. M. Buchanan, and B. D. Walker. HIV-1 Gag-Specific Cytotoxic T Lymphocytes Recognize Multiple Highly Conserved Epitopes. Fine Specificity of the Gag-Specific Response Defined by Using Unstimulated Peripheral Blood Mononuclear Cells and Cloned Effector Cells. *J. Immunol.*, 147:1512-1521, 1991. This study presented a detailed study of gag-specific CTL from HIV-1 seropositive individuals. Seven p24 and two p17 epitopes were described, that were recognized by class I-restricted CD3+CD8+ CTL. p17 epitopes: KIRLRPGGKKKYKLKHIVWASRELE and QTGSEELRSLYNTVATLYCVHQRIE; p24 epitopes: NPPPIPVGEIYKRWIILGLNKIV, VHQAI SPRTLNAWVKVVEEKAF, NAWVKVVEEKAFSPEVIPMFSA, SALSEGATPQDLNTMLNTYGGH, GHQAAMQLKETINEAAEWDR, and RAEQASQEVK. PubMed ID: [1715361](#). [Show all entries for this paper](#).

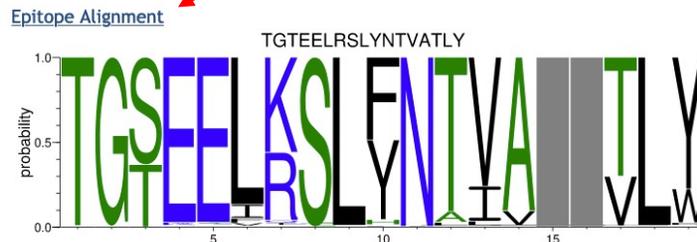
#### Displaying record number 58923

Download this epitope [record as JSON](#).

<a href="#">HXB2 Location</a>	Gag(70-86) p17(70-86) DNA(997..1047)	<a href="#">Gag Epitope Map</a>
<a href="#">Author Location</a>	Gag(70-)	

<a href="#">Epitope</a>	TGTEELRSLYNTVATLY
-------------------------	-------------------

<a href="#">Subtype</a>	B, C
<a href="#">Species (MHC/HLA)</a>	human(C*14)



301 Records found for SLYNTVATL in CTL epitopes

Each has a pre-done link to an analysis of diversity in sequence database alignment of one sequence per patient.



### AnalyzeAlign

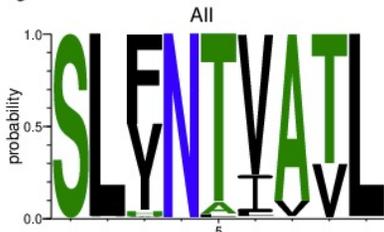
Input options

- Alignment: imported from QuickAlign (6144 amino acid sequence(s))
- Group the sequences: report results for all sequences as a single group
- Logo stack width: 22
- Stacks per line: 25
- Units: probability
- Show x-axis: yes
- Show y-axis: yes
- Color scheme: hydrophobicity
- Draw stacks with largest letters on top: yes
- Remove symbols from logo: show all
- Mark potential N-linked glycosylation sites: no
- Delete gaps and shift alignment to C-terminus: no
- Cut-off for calculating frequency by position: 95%
- Master sequence for finding variants: SLYNSVATL

Warning: Your sequences had to be padded - please make sure all sequences are of the same length.

Download all results: [ZIP](#)

Logo



Download: [PNG](#)

Frequency by position

[See full raw counts](#)

cutoff: 95%

	Percentage and raw count of non-gap	Non-gap/total (percentage)	Gap/total (percentage)
1	S: 99.87% (6136) other: 0.13% (8)	6144/6144 (100.00%)	0/6144 (0.00%)
2	L: 99.09% (6088) other: 0.91% (56)	6144/6144 (100.00%)	0/6144 (0.00%)
3	F: 48.76% (2996) Y: 46.88% (2880) other: 4.36% (268)	6144/6144 (100.00%)	0/6144 (0.00%)
4	N: 99.71% (6126) other: 0.29% (18)	6144/6144 (100.00%)	0/6144 (0.00%)
5	T: 90.89% (5584) A: 6.46% (397) other: 2.65% (163)	6144/6144 (100.00%)	0/6144 (0.00%)
6	V: 76.11% (4676) I: 20.51% (1260) other: 3.39% (208)	6144/6144 (100.00%)	0/6144 (0.00%)
7	A: 90.53% (5562) V: 8.61% (529) other: 0.86% (53)	6144/6144 (100.00%)	0/6144 (0.00%)
8	T: 70.17% (4311) V: 29.30% (1800) other: 0.54% (33)	6144/6144 (100.00%)	0/6144 (0.00%)
9	L: 99.06% (6086) other: 0.94% (58)	6144/6144 (100.00%)	0/6144 (0.00%)

Sequence variants

HIV QuickAlign or AnalyzeAlign tools show that the third site, Y, is often F

Is this due to some clades of virus having Y, others F, or is it F/Y within each clade?



### AnalyzeAlign

Input options

- Alignment: imported from QuickAlign (4661 amino acid sequence(s))
  - Major subtypes only
- Group the sequences: separately report results for subsets grouped by user's grouping choice
- Logo stack width: 22
- Stacks per line: 25
- Units: probability
- Show x-axis: yes
- Show y-axis: yes
- Color scheme: hydrophobicity
- Draw stacks with largest letters on top: yes
- Remove symbols from logo: show all
- Mark potential N-linked glycosylation sites: no
- Delete gaps and shift alignment to C-terminus: no
- Cut-off for calculating frequency by position: 95%
- Master sequence for finding variants: SLYNSVATL

Warning: Your sequences had to be padded - please make sure all sequences are of the same length.

Download all results: [ZIP](#)

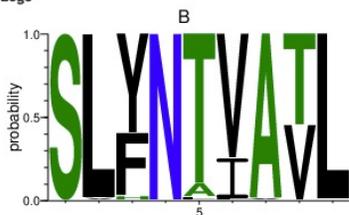
Groups

[B](#) [A1](#) [A2](#) [C](#) [D](#) [F1](#) [F2](#) [G](#) [O1\\_AE](#) [O2\\_AG](#)

Group B

[Go to top](#)

Logo



Download: [PNG](#)

Frequency by position

[See full raw counts](#)

cutoff: 95%

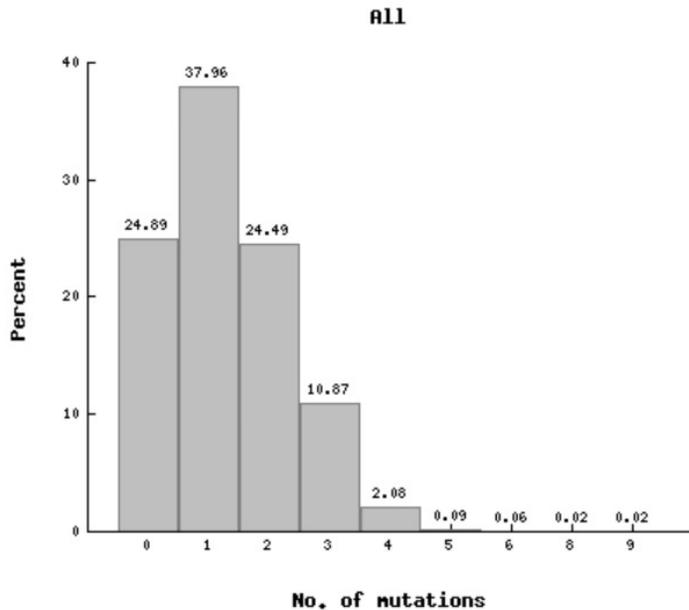
	Percentage and raw count of non-gap	Non-gap/total (percentage)	Gap/total (percentage)
1	S: 99.84% (1874) other: 0.16% (3)	1877/1877 (100.00%)	0/1877 (0.00%)
2	L: 98.03% (1840) other: 1.97% (37)	1877/1877 (100.00%)	0/1877 (0.00%)
3	Y: 59.51% (1117) F: 37.51% (704) other: 2.98% (56)	1877/1877 (100.00%)	0/1877 (0.00%)
4	N: 99.57% (1869) other: 0.43% (8)	1877/1877 (100.00%)	0/1877 (0.00%)
5	T: 89.50% (1680) A: 8.26% (155) other: 2.24% (42)	1877/1877 (100.00%)	0/1877 (0.00%)
6	V: 74.43% (1397) I: 24.29% (456) other: 1.28% (24)	1877/1877 (100.00%)	0/1877 (0.00%)
7	A: 98.40% (1847) other: 1.60% (30)	1877/1877 (100.00%)	0/1877 (0.00%)
8	T: 54.77% (1028) V: 44.91% (843) other: 0.32% (6)	1877/1877 (100.00%)	0/1877 (0.00%)
9	L: 99.09% (1860) other: 0.91% (17)	1877/1877 (100.00%)	0/1877 (0.00%)

Viewing each subtype separately, we see that the Y vs F at the third site is common in all clades or subtypes of HIV-1



Exploring that region of the HIV-1 sequence alignment, we see it is a TAC = Y vs TTC = F transversion in the genome. Nearby, ACC = T vs GTC = V is not co-varying with the Y vs F.

Sequence variants



Overall, 24.89% of HIV-1 have SLYNTVATL

22.51% are SLFNTVATL

6.45% are SLYNTVAVL

6.33% are SLFNTVAVL

Years ago, we noted that many sites “toggled” back and forth between two amino acids, and that the top few most common forms of each epitope thus made up the majority.

Variant	Count	Pct.	No. of mutations
SLYNTVATL			
-----	1529	24.89	0
--F-----	1383	22.51	1
-----V-	396	6.45	1
--F----V-	389	6.33	2
--F--I---	330	5.37	2
----I-V-	219	3.56	2
----I---	181	2.95	1
--F--I-V-	178	2.9	3
--F---V--	108	1.76	2
-----V--	107	1.74	1
--F-A--V-	100	1.63	3
--F--IV--	85	1.38	3
--H-----	81	1.32	1
----A--V-	56	0.91	2
--F-A----	56	0.91	2

This was the birth of our “Mosaic Vaccine Design” which has been proven useful not only for HIV vaccines but also for vaccines against several other viruses.

### The HIV-1 HLA-A2-SLYNTVATL is a help-independent CTL epitope.

Kan-Mitchell J, Bisikirska B, Wong-Staal F, Schaubert KL, Bajcz M, Bereta M.

J Immunol. 2004 May 1;172(9):5249-61. doi: 10.4049/jimmunol.172.9.5249.

PMID: 15100263 **Free article.**

The CTL response to the HLA-A\*0201-restricted, HIV-1 p17 Gag(77-85) epitope (SLYNTVATL; SL9) has been extensively studied in patients. ...In contrast to published reports for influenza and melanoma peptides and the HIV gag IV9 epitope studied here in p ...

### Characterization of cross-reactive CD8+ T-cell recognition of HLA-A2-restricted HIV-Gag (SLYNTVATL) and HCV-NS5b (ALYDVVSKL) epitopes in individuals infected with human immunodeficiency and hepatitis C viruses.

Vali B, Tohn R, Cohen MJ, Sakhdari A, Sheth PM, Yue FY, Wong D, Kovacs C, Kaul R, Ostrowski MA.

J Virol. 2011 Jan;85(1):254-63. doi: 10.1128/JVI.01743-10. Epub 2010 Oct 27.

PMID: 20980521 **Free PMC article.**

The immunologic mechanisms underlying the faster progression of hepatitis C virus (HCV) disease in the presence of human immunodeficiency virus (HIV) coinfection are not clearly understood. ...Our search for amino acid sequence homology between the HCV ...

### Persistent immunodominant anti-gag SLYNTVATL responses in HIV-patients with up to 7 years of HAART.

Zidovec Lepej S, Kosor E, Gagro A, Vince A, Remenar A, Poljak M.

Coll Antropol. 2006 Dec;30 Suppl 2:33-8.

PMID: 17508471

Gag SLYNTVATL-specific CD8+ T-cells were detectable in 18 of 26 treated patients (median 5.2 years of HAART) and in 10 of 14 untreated patients. Median percentage of Gag SLYNTVATL-specific CD8+ T-cells in treated patients was 0.10 (range 0.00-0.70%). . . .

### Inpatient escape in the A\*0201-restricted epitope SLYNTVATL drives evolution of human immunodeficiency virus type 1 at the population level.

Edwards CT, Pfafferoth KJ, Goulder PJ, Phillips RE, Holmes EC.

J Virol. 2005 Jul;79(14):9363-6. doi: 10.1128/JVI.79.14.9363-9366.2005.

PMID: 15994836 **Free PMC article.**

The hypothesis that the inpatient emergence of cytotoxic T-lymphocyte escape variants contributes to the evolution of human immunodeficiency virus type 1 at the population (interpatient) level was tested using the HLA-A\*0201-restricted gag p17 ...

A PubMed search for  
“SLYNTVATL” hits 55 papers.

HIV Immunology Database at  
LANL has 301 records for  
SLYNTVATL and 14 for  
SLFNTVATL

How do we explore the evolution  
of this epitope in the HIV  
Sequences Database?







# Searching Immunology Database for MHC/HLA A\*02 and proven diminished response in a Gag epitope, identifies SLFNTIATL as an escape variant.

Displaying record number 54556

<a href="#">HXB2 Location</a>	Gag(77-85)	<a href="#">Gag Epitope Map</a>
		<a href="#">View variants at this location</a>
<a href="#">Epitope</a>	SLYNTVATL	<a href="#">Epitope Alignment</a>
<a href="#">Variants</a>	SLfNTiATL	diminished response; escape documented in this paper
<a href="#">Epitope Name</a>	SL9	
<a href="#">Species (MHC/HLA)</a>	human(A*02)	

## Variant Details

Showing 1 of 4 variants.

Variant ID.	634
Epitope Seq.	SLYNTVATL
Variant Seq.	SLfNTiATL
Mutations	Y/F V/I
Epitope Location	Y3F V6I
HXB2 Location	Y79F V82I
Mutation Type	DR: diminished response E: escape documented in this paper
Method	CD8 T-cell Elispot - IFNy
Note	Variant detected by week 63. Lower magnitude responses than wt and V/I mutant. Positive selection. Seen with upstream changes E62G/V/A, where the 62A mutation persisted and is suggested to reduce fitness cost of the Y/F escape mutation.

## References

Karlsson2007 Annika C. Karlsson, Astrid K. N. Iversen, Joan M. Chapman, Tulio de Oliveira, Gerald Spotts, Andrew J. McMichael, Miles P. Davenport, Frederick M. Hecht, and Douglas F. Nixon. Sequential Broadening of CTL Responses in Early HIV-1 Infection Is Associated with Viral Escape. *PLoS ONE*, 2:e225, 2007. PubMed ID: [17311088](#). [Show all entries for this paper.](#)

Karlsson2007 paper  
has link to sequence data  
EF396480–EF396891

# Patient OP428 with SLF/YNTIATL epitope

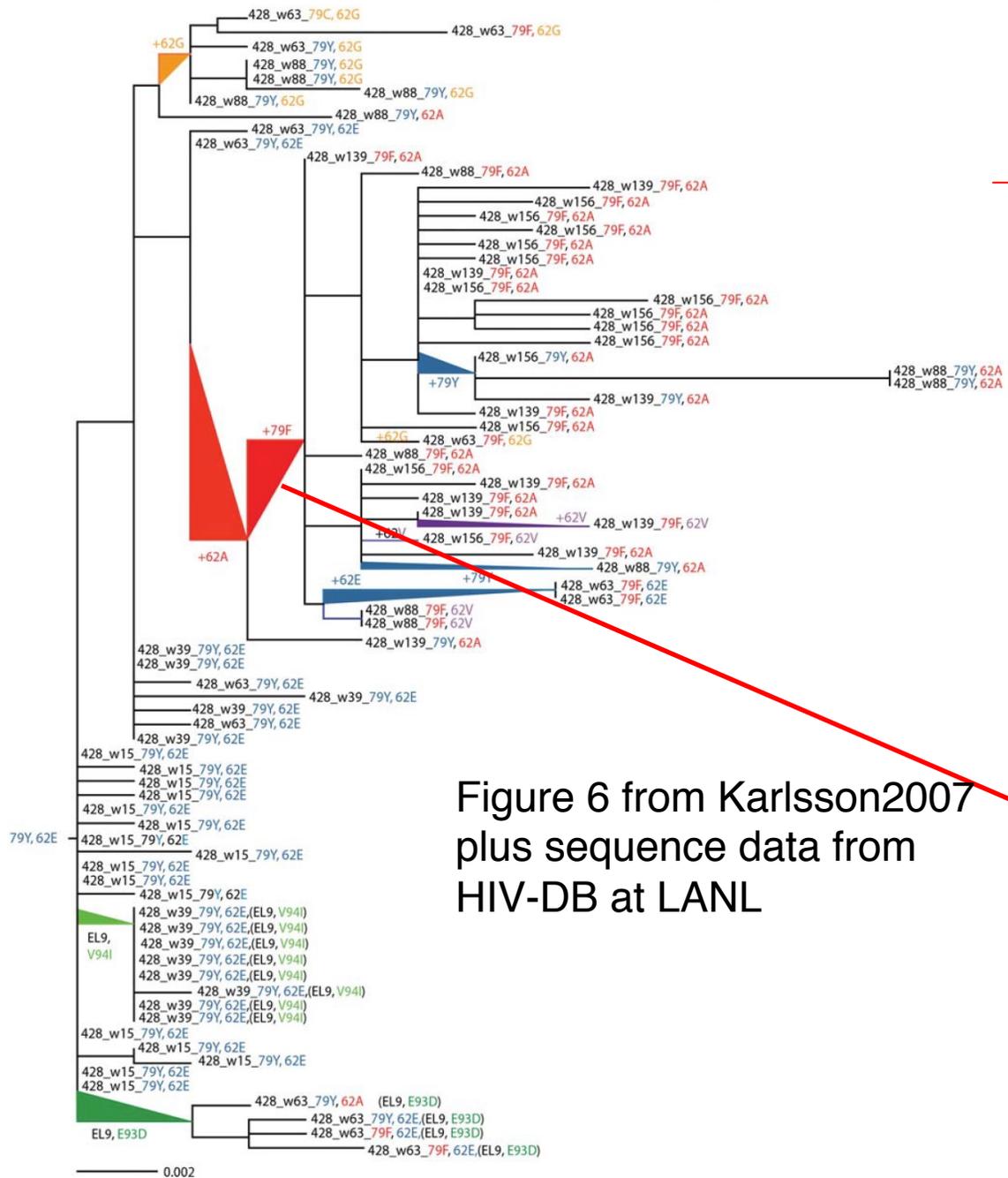
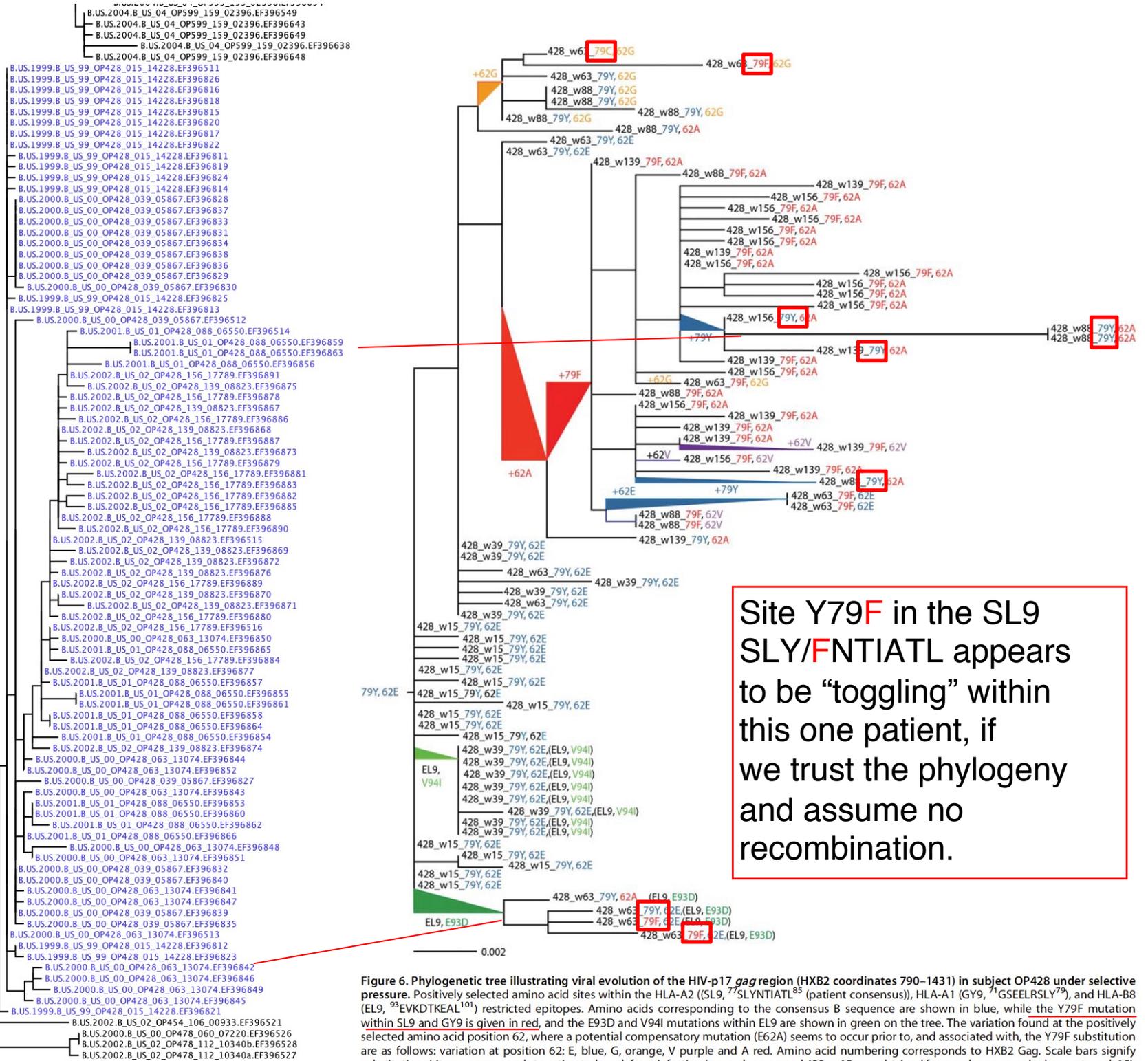


Figure 6 from Karlsson2007 plus sequence data from HIV-DB at LANL

Figure 6. Phylogenetic tree illustrating viral evolution of the HIV-p17 *gag* region (HXB2 coordinates 790–1431) in subject OP428 under selective pressure. Positively selected amino acid sites within the HLA-A2 (SL9, <sup>77</sup>SLYNTIATL<sup>85</sup> (patient consensus)), HLA-A1 (GY9, <sup>71</sup>GSEELRSLY<sup>79</sup>), and HLA-B8 (EL9, <sup>99</sup>EVKDKTEAL<sup>101</sup>) restricted epitopes. Amino acids corresponding to the consensus B sequence are shown in blue, while the Y79F mutation within SL9 and GY9 is given in red, and the E93D and V94I mutations within EL9 are shown in green on the tree. The variation found at the positively selected amino acid position 62, where a potential compensatory mutation (E62A) seems to occur prior to, and associated with, the Y79F substitution are as follows: variation at position 62: E, blue, G, orange, V purple and A red. Amino acid numbering corresponds to HXB2 Gag. Scale bars signify substitutions/site. w: corresponds to estimated week from infection (e.g. a clone named 428\_w15 was obtained from a plasma sample drawn at week 15). doi:10.1371/journal.pone.0000225.g006

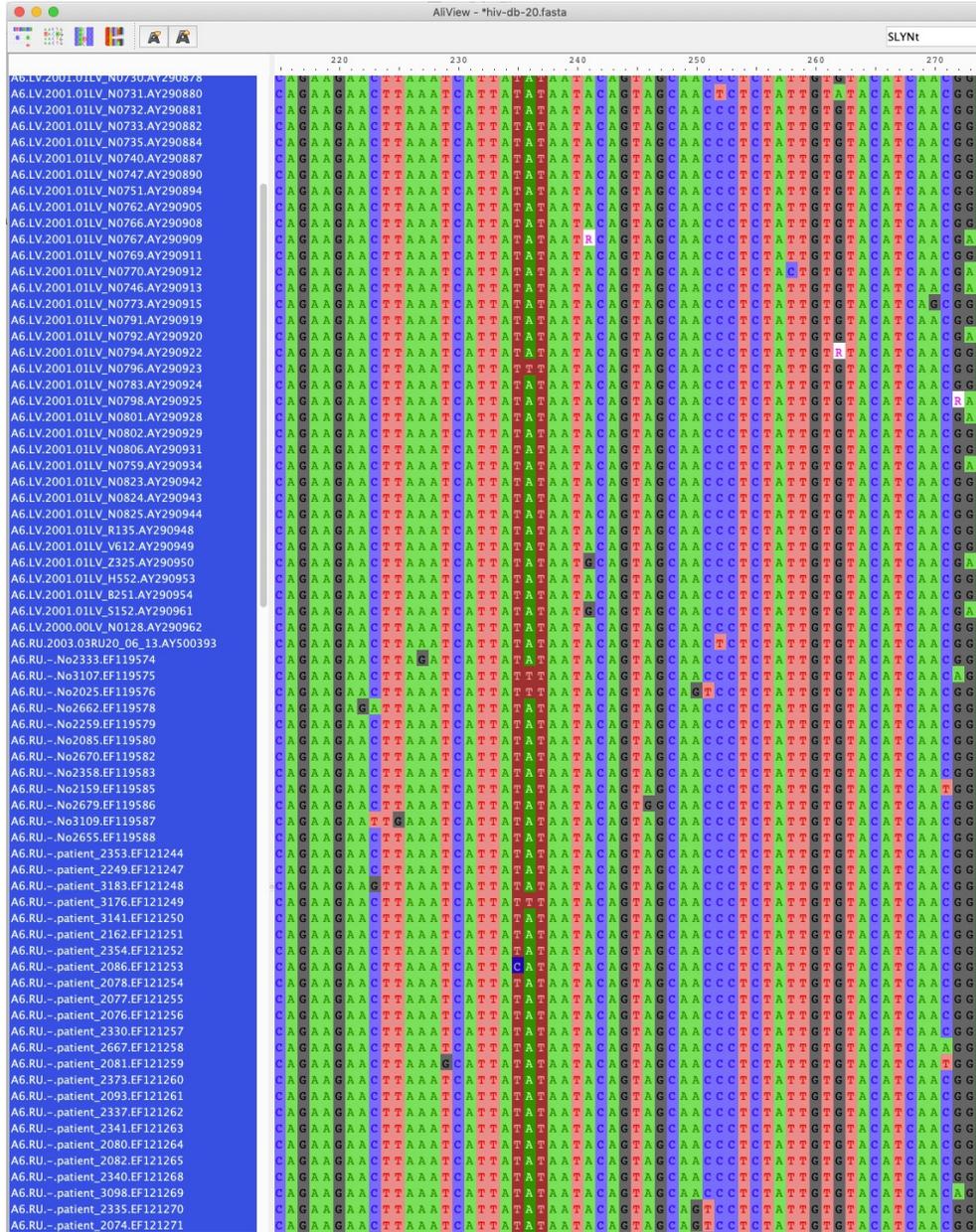


OP599  
OP428



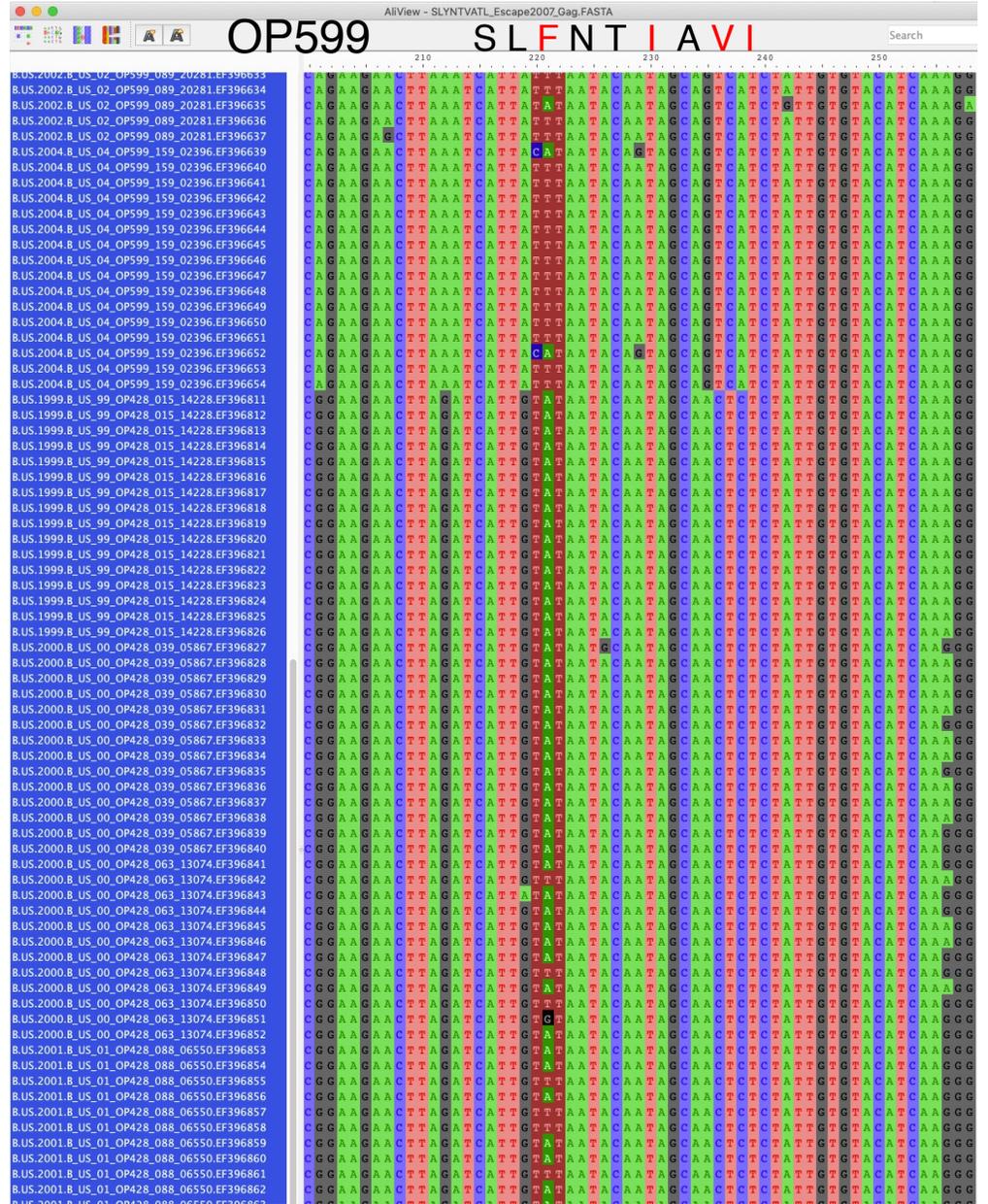
**Figure 6. Phylogenetic tree illustrating viral evolution of the HIV-p17 gag region (HXB2 coordinates 790–1431) in subject OP428 under selective pressure.** Positively selected amino acid sites within the HLA-A2 (SL9, <sup>77</sup>SLYNTIATL<sup>85</sup> (patient consensus)), HLA-A1 (GY9, <sup>71</sup>GSEELRSLY<sup>79</sup>), and HLA-B8 (EL9, <sup>99</sup>EVKDKTEAL<sup>101</sup>) restricted epitopes. Amino acids corresponding to the consensus B sequence are shown in blue, while the Y79F mutation within SL9 and GY9 is given in red, and the E93D and V94I mutations within EL9 are shown in green on the tree. The variation found at the positively selected amino acid position 62, where a potential compensatory mutation (E62A) seems to occur prior to, and associated with, the Y79F substitution are as follows: variation at position 62: E, blue, G, orange, V purple and A red. Amino acid numbering corresponds to HXB2 Gag. Scale bars signify substitutions/site. w: corresponds to estimated week from infection (e.g. a clone named 428\_w15 was obtained from a plasma sample drawn at week 15). doi:10.1371/journal.pone.0000225.g006

# Subsubtype A6 one per patient



SLYNT VATL R

# Subtype B two patients OP428 is HLA A02



OP428 SLYNT I A T L R  
 F G

## Protein Feature Accent

See [Protein Feature Accent Explanation](#) for details. Return to [input page](#).

Control Panel

Protein Data Bank structure ID: [1HIW](#)  
TRIMERIC HIV-1 MATRIX PROTEIN

re-center  spin

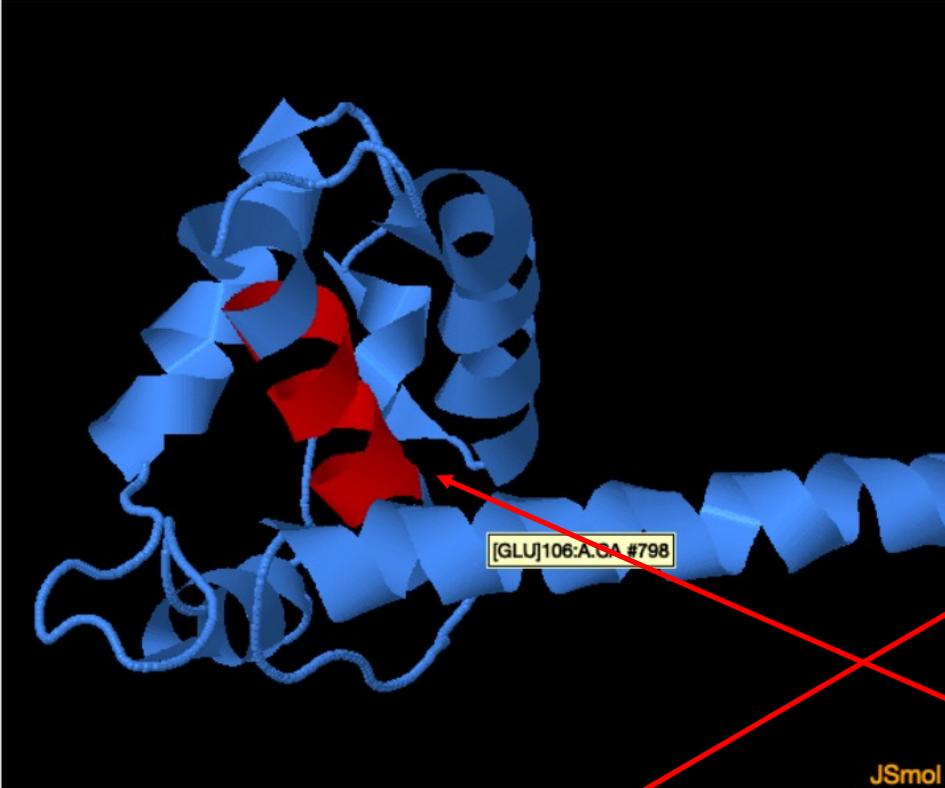
cartoon

pick color scheme

Background:

Jmol command script:  
help

download this view as a  
jpg  image



Sequence View

Select residues in the **top (PDB file) sequence** below to highlight them in the graphic above.

show PDB file annotation  show reference sequence  show reference sequence annotation

20 30 40 50 60 70 80 90 100 110 120

RLRPGGKKQYKLVASRELERFVAVNPGLLETSEGCRQILGQLQPSLQTGSEELRSLYNTIYAVLYCVHQRIDVKDTKEALDKIEEEQNKSCKKAQQAAD-----  
RLRPGGKKKYKLVASRELERFVAVNPGLLETSEGCRQILGQLQPSLQTGSEELRSLYNTVATLYCVHQRIDVKDTKEALDKIEEEQNKSCKKAQQAAD-----  
RLRPGGKKQYKLVASRELERFVAVNPGLLETSEGCRQILGQLQPSLQTGSEELRSLYNTIYAVLYCVHQRIDVKDTKEALDKIEEEQNKSCKKAQQAADTGN--  
RLRPGGKKYRLLKLVASRELERFALXPXLLSAEGCQIMEQLQPAAXTGXEEIKSLFNTVATLYCVHQRIDVKDTKEAXDKIEEIKNKSQRQTQQAADTGN--  
RLRPGGKKYRMLKLVASRELERFALNPGLLETTEGCQKIEQLQPSVKTGTGTEELKSLFNTVATLYCVHQRIDVKDTKEALDKIEEMQNKSQKQTQQAADIGNS--

Highlighting

high-contrast highlighting  make highlights persistent

Currently highlighted: 77-85

Highlight [residue numbers](#):  and atoms within  Å

Our Protein Feature Accent tool, loads a PDB 3D structure file and an alignment of protein sequences.

JMol is used to display the protein structure so you can rotate it, change the view from cartoon to spacefill, etc.

Selecting sites or regions in the sequence

highlights that site or region in the structure.



Gag from all primate lentiviruses.

Looking outside the HIV-1 M group (roughly 100 years of evolution in humans) to SIVs representing at least 100,000 years and more likely over 5 million years of evolution.

Can help determine which amino acids are invariant and thus likely critical for protein function.

# Thank you for attending!

**We are happy to help with research questions on the use of our tools and database.**

**We are thrilled to get ideas for further tool development!**

**Contact us: [seq-info@lanl.gov](mailto:seq-info@lanl.gov) or [immuno@lanl.gov](mailto:immuno@lanl.gov)**

# Datasets and programs

<https://tinyurl.com/HIVDB-IEDB-2021>

**Aliview**

**Bioedit**

**Figtree**

**slides**

**datafiles**