



**Karolinska  
Institutet**

# **Web resources for analysis of HIV sequences**

**EDCTP ENNEA training on data management**

November 14-16 2011, Muhimbili hospital, Dar Es Salaam

Irene Bontell

([irene.bontell@ki.se](mailto:irene.bontell@ki.se))

# What information can we get from sequences?

Nucleic acid sequences can provide information on many things from functionality and structure of proteins to evolutionary histories of species.

For HIV common usages are:

- Subtype classification
  - Recombination analysis
  - Detection of drug resistance mutations
  - Selection of optimal antiretroviral therapy
  - Determination of co-receptor usage (V3-region of env-gene)
  - Phylogenetic relationship between strains, time of most recent common ancestor (tMRCA)
-

# How do I obtain sequences for analysis?

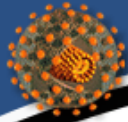
1. Sequence viral RNA or DNA from patient samples or cultured virus
2. Download from databases

All sequences from published studies **MUST** be submitted to GenBank. However, there are also many sequences in GenBank which have never been part of any publication. It is a good starting place to gain knowledge.

---

- Four HIV databases
  - Sequence
  - Immunology
  - Resistance (no longer active)
  - Vaccine trials
- The sequence database is extensive and very well maintained
  - User-friendly and very flexible search interface, plus premade alignments
  - Links to a large number of tools including external programs
  - Helpful tutorials and reference information (genome map, CRF maps, detailed base-by-base file of the reference strain HXB2, format converter, list of country codes, HIV-BLAST etc)

Start page LANL Sequence database:  
<http://www.hiv.lanl.gov/content/sequence/HIV/mainpage.html>



## HIV sequence database

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

## HIV Sequence Database

### Programs and Tools

[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

[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

### Information

[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

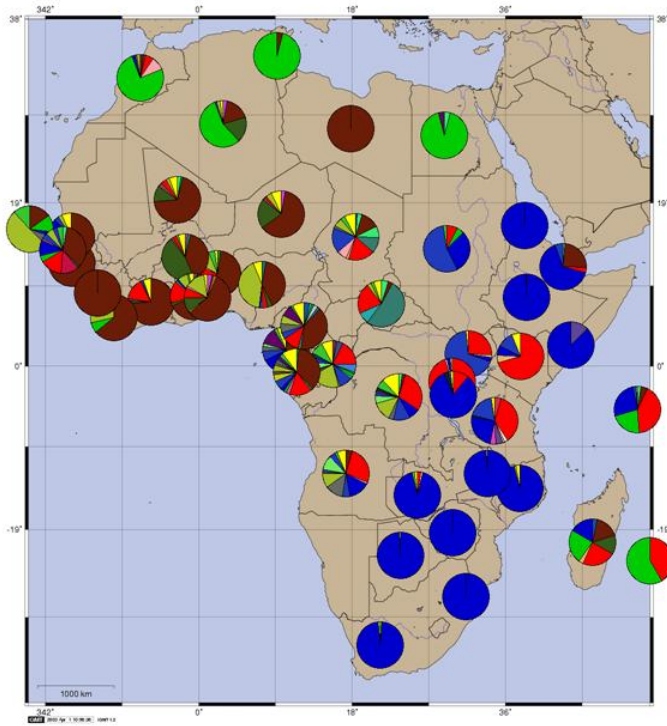
[FAQ](#) general information about this website

[Site Statistics](#) usage information for [www.hiv.lanl.gov](http://www.hiv.lanl.gov)

[How to Cite this Database](#)

[Editorial Board](#)

# LANL – sequence retrieval, Geography



Displaying 1 - 100 of 4246 sequences found:

Note: 110 [problematic](#) sequences were removed from this result. [Click here to repeat search to include problematic sequences.](#)

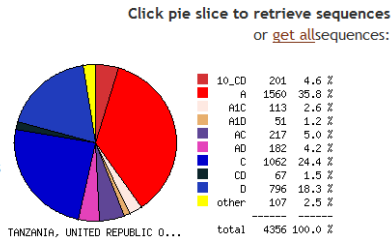
record  to 
  records per page

Click on field name to sort in ascending or descending order

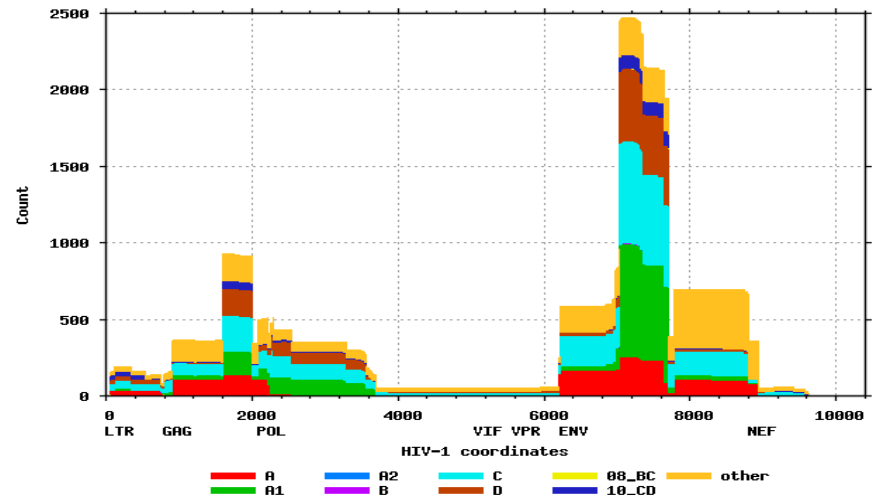
#	Select	Patient (id)	Code	Accession Name	Subtype	Country	Sampling Year	Georegion	Genomic Region	Sequence Length	Organism	
1	<input type="checkbox"/>	<a href="#">Blast 1(4019)</a>		AF038051	TZB0001	A	TZ	1996	AFR SSA		649	HIV-1
2	<input type="checkbox"/>	<a href="#">Blast 2(4020)</a>		AF038052	TZB0002	A	TZ	1996	AFR SSA		646	HIV-1
3	<input type="checkbox"/>	<a href="#">Blast 3(4021)</a>		AF038053	TZB0003	A	TZ	1996	AFR SSA		661	HIV-1
4	<input type="checkbox"/>	<a href="#">Blast 4(14296)</a>		AF038054	TZB0004	A	TZ	1996	AFR SSA		661	HIV-1
5	<input type="checkbox"/>	<a href="#">Blast 5(4022)</a>		AF038055	TZB0005	A	TZ	1996	AFR SSA		672	HIV-1
6	<input type="checkbox"/>	<a href="#">Blast 6(4023)</a>		AF038056	TZB0006	A	TZ	1996	AFR SSA		676	HIV-1
7	<input type="checkbox"/>	<a href="#">Blast 7(4024)</a>		AF038057	TZB0007	A	TZ	1996	AFR SSA		652	HIV-1

Distribution of all HIV-1 sequences: TANZANIA, UNITED REPUBLIC OF  
 Please note that this map only includes sequences for which the sampling country is known.

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)  
 Region ▾ or  
 TANZANIA, UNITED REPUBLIC OF ▾  
 Show    sequences  
 Table ([html](#)) of the compiled subtype distribution.



Frequency of HIV-1 subtypes



# LANL – sequence retrieval, general search

## Sequence Information

exact

TZ

Include recombinants

Virus: HIV-1  
 Subtype: A2, B, C, **D**, F, F1

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

Genomic region: p8, Pol CDS, p51 (RT), p15 (RNase H), p31 (integrase), protease, Vif CDS

Or define start: 2253 and end: 3263

Include fragments of minimum length 100

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

Selects a specific genomic region based on your input alignment. Paste your aligned nucleotide sequences into the box:

Or upload your alignment file:

Use first sequence in alignment to define region  
 Use ragged ends to define region

## Publication Information

## Patient Information

## Geographical Information

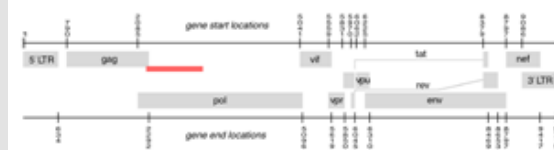
## Output

Include problematic sequences

List 100 records per page

Show results selected
  Show SQL

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



Download options:

Align (only HIV-1 and SIVcpz)
 Format: Fasta
 Gap handling: none

Sequence type: Nucleotides

Clip to selected region
  Include HXB2 Reference Sequence (K03455)

Show names as: Use-defined label or compose a label

Compose sequence labels:

Displaying 1 - 64 of 64 sequences found:

Note: 1 problematic sequences were removed from this result. [Click here to repeat search to include problematic sequences.](#)

Select records  to 

 records per page

Click on field name to sort in ascending or descending order

#	Select	Patient Code	Accession Name	Subtype	Country	Sampling Year	Genomic Region	Sequence Length	Organism
1	<input checked="" type="checkbox"/>	Blast 280(7087)	AY253311	A280	D	TZ	2001	8775	HIV-1
2	<input type="checkbox"/>	Blast	EU251720	TBK008	D	TZ	2005	1519	HIV-1
3	<input type="checkbox"/>	Blast	EU251723	TBK014	D	TZ	2005	1419	HIV-1
4	<input checked="" type="checkbox"/>	Blast	EU251727	TBK020	D	TZ	2005	1459	HIV-1
5	<input type="checkbox"/>	Blast	EU251728	TBK022	D	TZ	2005	1326	HIV-1
6	<input type="checkbox"/>	Blast	EU251731	TBK025	D	TZ	2005	1416	HIV-1
7	<input checked="" type="checkbox"/>	Blast	EU251732	TBK029	D	TZ	2005	1343	HIV-1
8	<input type="checkbox"/>	Blast	EU251740	TBK042	D	TZ	2005	1414	HIV-1
9	<input type="checkbox"/>	Blast	EU251752	TBK061	D	TZ	2005	1465	HIV-1

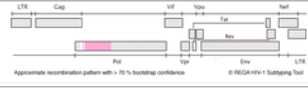

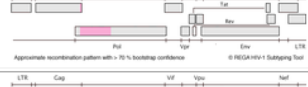


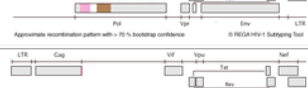

# Subtyping, REGA (can be found under external tools in LANL)

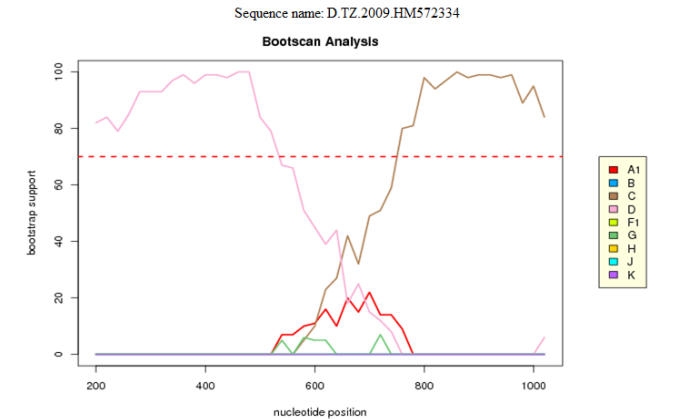
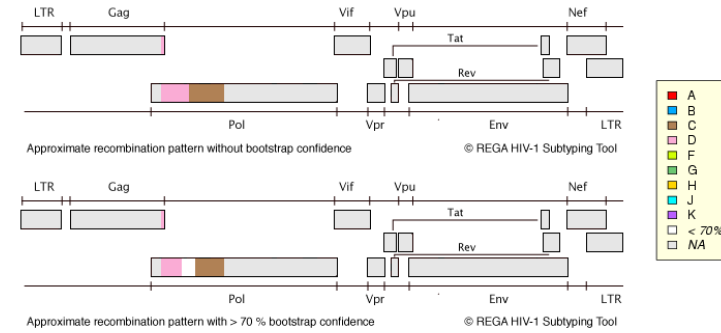


Sequence name : D.TZ.2009.HM572334, length: 1011 bps

Assignment: Check the bootscan, Bootstrap: NA

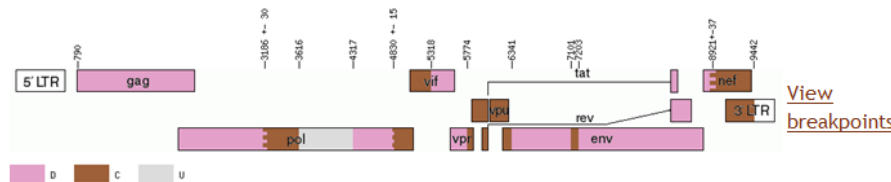
## HIV Genotyping Tool Results

Name	Length	Report	Assignment	Support	Genome
D.TZ.2001.AY253311	1011bp	<a href="#">Report</a>	HIV-1 Subtype D	100.0	
D.TZ.2005.EU251720	1011bp	<a href="#">Report</a>	HIV-1 Subtype D	100.0	
D.TZ.2005.EU251740	1011bp	<a href="#">Report</a>	HIV-1 Subtype D	100.0	
D.TZ.2005.EU251786	1011bp	<a href="#">Report</a>	HIV-1 Subtype D	97.0	
D.TZ.2005.EU251859	1011bp	<a href="#">Report</a>	HIV-1 Subtype D	100.0	
D.TZ.2009.HM572334	1011bp	<a href="#">Report</a>	Check the bootscan	NA	
D.TZ.2007.HM572341	1011bp	<a href="#">Report</a>	HIV-1 Subtype D	100.0	



CRF10\_CD Reference strain: TZBF061 Subtypes: C, D

check CRF  
breakpoints for all  
C-D recombinants!  
(LANL-Guides-  
CRFs)



790-3155 D  
3156-3216 D/C  
3217-3615 C

[View  
breakpoints](#)

# RIP (tools) and NCBI subtyping (external tools)

## Sequence

D.TZ.2001.AY253311  
 D.TZ.2005.EU251720  
 D.TZ.2005.EU251740  
 D.TZ.2005.EU251786  
 D.TZ.2005.EU251859  
 D.TZ.2009.HM572334  
 D.TZ.2007.HM572341  
 D.TZ.2008.HM572344  
 D.TZ.2009.HM572346  
 D.TZ.2008.HM572349  
 D.TZ.2007.HM572350  
 D.TZ.2009.HM572355

## Subtype

[CON\\_D](#)  
[CON\\_D](#)  
[CON\\_D](#)  
[CON\\_D](#)  
[CON\\_D](#)  
[CON\\_C,CON\\_D](#)  
[CON\\_D](#)  
[CON\\_D](#)  
[CON\\_D](#)  
[CON\\_D](#)  
[CON\\_D](#)  
[CON\\_D](#)

Select: Human immunodeficiency virus 1, (HIV-1) - 2004 "pure" & CRFs

[Add/remove reference](#)

[sequences as](#)

[Accession/subtype:](#)

U88822|D  
 U88824|D  
 R03454|D  
 AF289548|CRF10  
 AF289549|CRF10  
 AF289550|CRF10

[Enter local reference sequences as FASTA:](#)

[Enter your query sequence as Accession/GI or FASTA](#)

```
>D.TZ.2009.HM572334
CCTCAATCACTCTTTGGCAAGACCCCTTGTCCAGTAAAAATAGGGGACAGCTAAAGGAAGCTATTAGATACAGG
AGCAGATGATACAGTCTTAGAAGAATGAAITTTGCCAGGAAAAATGGAALCCAAAAATGATAGGGGAATGGAGGTTTAA
TCAAAGTTAGACAGTATGATCAAAATACCCCTAGAAATCTGGACATAAAGCTATAGGTACAGTATTAGTAGGACCTACA
CCTATAAACATAAATGGAGGAATTTGTTGACTCAGATTGGTGTACTTTGAAITTTCCAAATAGTCTCATTGAAACGGT
ACCAGTAAAAATAAAGCCAGGATGGATGGCCAAAAGTGAALCAATGGCCATTGACAGAGAAAAATAAAGCACTAA
```

Subtype Reset Clear

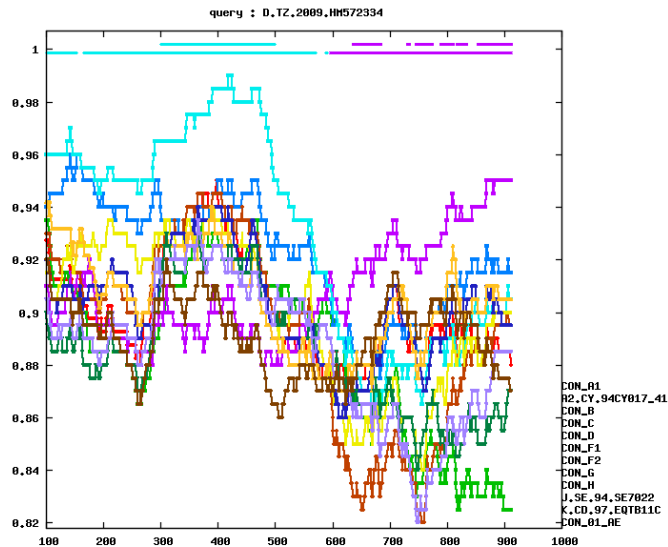
Enter window size: 300

Select background color: White

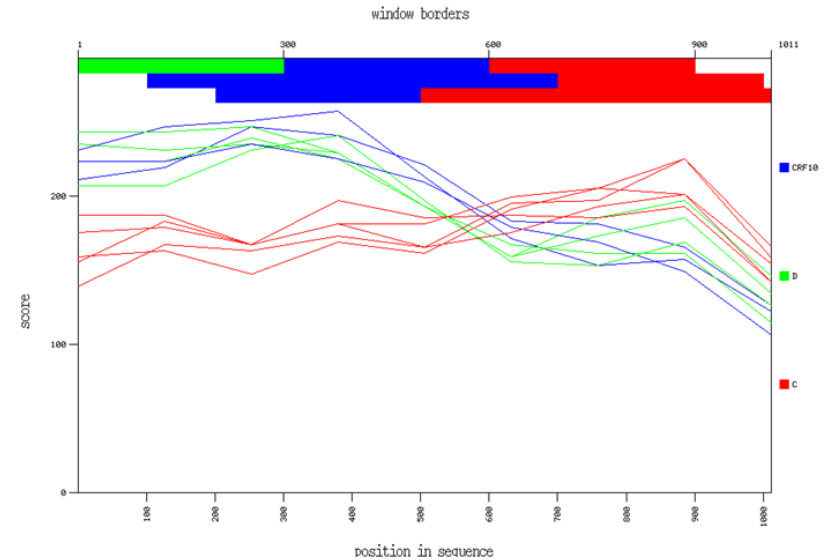
Enter increment step: 100

Similarity threshold (%): 30

Table output only:



D.TZ.2009.HM572334



# Stanford HIV drug resistance database



**STANFORD UNIVERSITY HIV DRUG RESISTANCE DATABASE**  
*A curated public database designed to represent, store, and analyze the divergent forms of data underlying HIV drug resistance.*

HOME GENOTYPE-RX GENOTYPE-PHENO GENOTYPE-CLINICAL HIVdb PROGRAM

**CPR**  
 Analyze sequence sets for proportions with Surveillance Drug Resistance Mutations (SDRMs)

**Calibrated Population Resistance (CPR) Tool Version 6.0**  
 New transparent sequence and codon specific quality control analysis, PDF report, phylogenetic analysis, and excel download

**HIVdb PROGRAM** Genotype Resistance Interpretation  
 Interprets user-entered mutations to infer the level of resistance to NRTIs, NNRTIs, PIs. Web Service is available.

**TCE**  
 Tools for predicting the virological response to ARV therapy.

**Treatment-Change Episode (TCE) Suite of Programs and Repository Version 1.0**  
 Designed for researchers studying the predictors of virological success with new antiretroviral (ARV) regimens. Contains a TCE Viewer, TCE Finder, and TCE Repository.

**MARVEL**  
 Mutation ARV Evidence Listing

**ART-AIDE**  
 ARV Therapy - Acquisition & Display Engine

**GENOTYPE-TREATMENT CORRELATIONS**

- Retrieve sequences (and/or mutations) from persons receiving selected HIV drugs
- Retrieve sequences and treatments from viruses with specific mutations

**REFERENCES**

- Published drug resistance studies in HIVRT&PrDB
- Published studies by Stanford database group

**HIVseq Program**  
 Provides mutation frequencies by subtype and treatment

**Database Mirror**  
 Collaboration with Africa Centre for Health and Population Studies and South African MRC

**GENOTYPE-PHENOTYPE CORRELATIONS**

- Retrieve drug susceptibility data for isolates with selected mutations
- Download genotype-phenotype research datasets

**SURVEILLANCE MUTATIONS**

- World Health Organization 2009 Mutation List
- Mutation Prevalence
- Surveillance Drug Resistance Mutation (SDRM) Worksheet

**HIValg Program**  
 Compare HIVdb ANRS, Rega, or create your own algorithm

**HBVseq Program**  
 Provides HBV RT mutation frequencies by genotype and treatment.  
 HBVseq, ReleaseNotes

**GENOTYPE-CLINICAL CORRELATIONS**

- Summaries of genotype-clinical outcome studies
- Genotype-clinical outcome datasets (download)

**ANALYSIS TOOLS**

- REGA HIV-1 Subtyping Tool Version 2.0

**Drug Resistance Summaries**  
 Pls NRTIs NNRTIs PIs IIs

**Latest News**

- IAS 2011 (Rome) Oral Presentation: Comparative Analysis of Virological Efficacy of WHO 2010 TDF-Containing First Line Regimens
- Stanford Prosect Statement Supporting Expansion of HIVDB

**NEW SUBMISSIONS**

- Ndembu, et al. *Transmitted drug resistance in Kampala*
- Annaz, et al. *Drug resistance mutations among drug-naive patients in Morocco*
- Gulick et al. *ACTG A5095 Triple nucleoside vs efavirenz-containing regimens*
- Rey et al *Early virological failure with once-daily TDF/3TC/NVP*

Lots of tables and detailed information on genotype related to treatment, viral phenotype (resistance) and clinical data. Ex. Genotype-Pheno -> Patterns of Drug resistance mutations

## NRTI Mutation Pattern & Susceptibility

1058 patterns of major drug resistance mutations (defined [here](#)) from 56404 clinical isolates obtained from 50636 persons infected with HIV-1 Group M virus are arranged in decreasing prevalence in the table (updated on 2011-9-16). Sequences containing a mixture at one of the major drug resistance positions were excluded from the table. For each pattern, drug susceptibility data on isolates matching the specific mutation pattern is shown on the left side. The analysis was originally published in Antimicrobial Agents and Chemotherapy, Rhee et al. 2004.

The number of sequences with each pattern is indicated, as well as the median fold resistance, as assayed by the ViroLogic PhenoSense assay, to each of several individual members of the selected drug class. Subscripts indicate the number of sequences assayed for resistance to the designated drug. Interquartile ranges are available by clicking [here](#). Data on the 30 most common mutation patterns is also available in a [PDF format](#).

Fold Resistance		
AZT	4	20
D4T	1.5	3
TDF	1.5	3
ABC	2.5	5
DDI	1.5	3
DDC	1.5	3
3TC	4	20

Mutation Patterns	Number of Sequences	AZT fold <sub>n</sub>	D4T fold <sub>n</sub>	TDF fold <sub>n</sub>	ABC fold <sub>n</sub>	DDI fold <sub>n</sub>	DDC fold <sub>n</sub>	3TC fold <sub>n</sub>
None	35531	0.8 <sub>10</sub>	0.9 <sub>10</sub>	0.9 <sub>10</sub>	0.9 <sub>10</sub>	0.9 <sub>10</sub>	0.9 <sub>10</sub>	0.9 <sub>11</sub>
184V	4209	0.5 <sub>6</sub>	0.6 <sub>6</sub>	0.5 <sub>1</sub>	3.3 <sub>6</sub>	1.4 <sub>6</sub>	1.6 <sub>6</sub>	200 <sub>6</sub>
67N,70R,184V	811	2.8 <sub>7</sub>	1.3 <sub>7</sub>	0.9 <sub>1</sub>	4.2 <sub>4</sub>	1.5 <sub>7</sub>	1.9 <sub>7</sub>	200 <sub>7</sub>
41L,184V,210W,215Y	585	7.5 <sub>12</sub>	1.9 <sub>12</sub>	1.2 <sub>1</sub>	6.2 <sub>12</sub>	1.6 <sub>12</sub>	2.2 <sub>1</sub>	200 <sub>12</sub>
41L,184V,215Y	579	5.8 <sub>12</sub>	1.4 <sub>12</sub>	1.1 <sub>1</sub>	5.2 <sub>12</sub>	1.6 <sub>12</sub>	2.0 <sub>1</sub>	200 <sub>12</sub>
70R,184V	515	0.8 <sub>6</sub>	0.7 <sub>6</sub>	0.7 <sub>1</sub>	3.2 <sub>6</sub>	1.3 <sub>6</sub>	1.8 <sub>6</sub>	200 <sub>6</sub>
41L,67N,184V,210W,215Y	492	33 <sub>12</sub>	2.8 <sub>12</sub>	1.6 <sub>1</sub>	6.5 <sub>12</sub>	2.1 <sub>12</sub>	2.3 <sub>12</sub>	200 <sub>12</sub>
41L,67N,210W,215Y	462	52 <sub>12</sub>	4.8 <sub>12</sub>	4.6 <sub>1</sub>	6.6 <sub>12</sub>	1.8 <sub>12</sub>	1.8 <sub>1</sub>	5.0 <sub>12</sub>
41L,210W,215Y	455	108 <sub>12</sub>	2.5 <sub>12</sub>	2.6 <sub>1</sub>	3.1 <sub>12</sub>	1.5 <sub>12</sub>	1.7 <sub>12</sub>	3.0 <sub>12</sub>
67N,70R	434	13 <sub>12</sub>	1.6 <sub>12</sub>	1.7 <sub>1</sub>	1.5 <sub>12</sub>	1.1 <sub>12</sub>	1.1 <sub>12</sub>	1.8 <sub>12</sub>
70R	399	6.7 <sub>1</sub>	1.1 <sub>1</sub>	1.4 <sub>1</sub>	1.0 <sub>1</sub>	0.9 <sub>1</sub>	0.9 <sub>1</sub>	1.0 <sub>1</sub>
41L	373	1.1 <sub>1</sub>	0.8 <sub>1</sub>	1.3 <sub>1</sub>	0.8 <sub>1</sub>	0.8 <sub>1</sub>	0.7 <sub>1</sub>	1.3 <sub>1</sub>
41L,215Y	369	12 <sub>12</sub>	1.6 <sub>12</sub>	1.8 <sub>1</sub>	1.7 <sub>12</sub>	1.1 <sub>12</sub>	1.2 <sub>12</sub>	2.0 <sub>12</sub>
67N,70R,184V,215P	296	8.8 <sub>1</sub>	1.8 <sub>1</sub>	1.1 <sub>1</sub>	5.8 <sub>1</sub>	1.8 <sub>1</sub>	2.1 <sub>1</sub>	200 <sub>1</sub>
41L,184V	254	0.5 <sub>1</sub>	0.8 <sub>1</sub>	0.5 <sub>1</sub>	2.4 <sub>1</sub>	1.5 <sub>1</sub>	1.9 <sub>1</sub>	200 <sub>1</sub>

Start page Stanford HIVDB: <http://hivdb.stanford.edu/index.html>

# Stanford HIV drug resistance database



## Mutation Prevalence According to Subtype and Treatment

Last updated on 09/2011

<b>ARV Drug Class</b> <input checked="" type="radio"/> PI-experienced vs PI-naive <input type="radio"/> NRTI-experienced vs RTI-naive <input type="radio"/> NNRTI-experienced(+/-NRTI) vs NNRTI-naive/NRTI-experienced <input type="radio"/> NNRTI-experienced(+/-NRTI) vs RTI-naive <input type="radio"/> RTI-experienced vs RTI-naive <input type="radio"/> <b>NEW!</b> INI-experienced vs INI-naive	<b>Position Mutation Count</b> <input checked="" type="radio"/> One mutation per person (>1 isolate per person when different isolates have a different mutation at the same position) <input type="radio"/> One isolate per person (last 'on-therapy' isolate)	<b>Mutation Threshold</b> <input checked="" type="radio"/> >= 1% AND >= 2 persons <input type="radio"/> >= 0.5% <input type="radio"/> All  <b>Mixture</b> <input checked="" type="radio"/> Exclude <input type="radio"/> Include
--	---	---

## Mutation Prevalence According to Subtype and Treatment

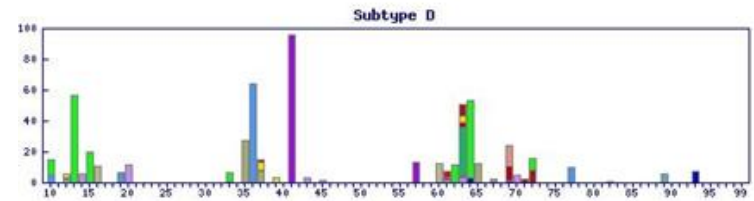
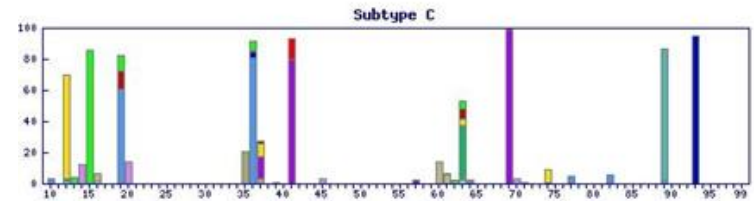
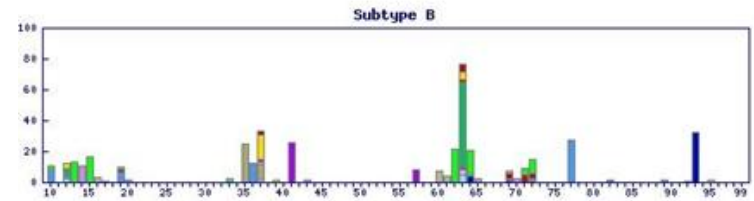
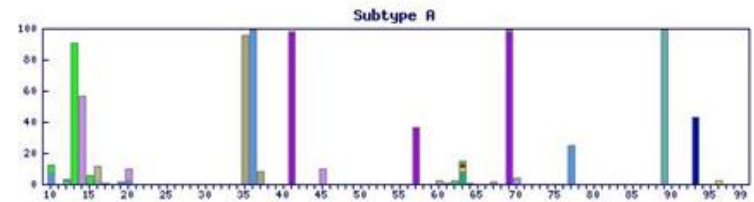
Last updated on 09/2011

ARV Drug Class: PI  
 IncludeMixtures: No  
 Mutations Threshold: >= 1% AND >= 2 persons  
 Position Mutation Count: Default

Column headers contain the number of persons with isolates according to subtype and drug class exposure. Table entries list the amino acid differences from consensus B and their frequency (%). Each amino acid is also a hyper link to a separate web page with detailed information on each isolate, including literature references with Medline abstracts, the GenBank accession number, and complete sequence and treatment records. Click [here](#) for a graphical summary of mutation frequency and variation across position and subtype in untreated persons

Pos	WT	PI Naive Persons								PI Treated Persons							
		A	B	C	D	F	G	AE	AG	A	B	D	F	G	AE	AG	
		2643	15474	5263	982	814	1051	2382	3090	180	7654	421	116	616	249	105	139
1	P																
2	Q																
3	I																V <sup>2.4</sup>
4	T																
5	L																
6	W																
7	Q																E <sup>1.8</sup>
8	R																
9	P																
10	L	I <sup>9.5</sup> V <sup>5.8</sup>	I <sup>7.2</sup> V <sup>2.9</sup>	I <sup>1.7</sup> V <sup>1.2</sup>	V <sup>6.8</sup> I <sup>2.4</sup>	V <sup>2.0</sup> I <sup>8.9</sup>	I <sup>9.7</sup> M <sup>1.2</sup>	I <sup>7.6</sup> V <sup>3.5</sup>	V <sup>8.2</sup> I <sup>5.5</sup>	I <sup>2.0</sup> V <sup>1.5</sup> E <sup>5.2</sup>	I <sup>3.9</sup> E <sup>1.2</sup> V <sup>6.7</sup>	I <sup>9.6</sup> E <sup>6.2</sup> V <sup>4.4</sup> R <sup>1.0</sup>	I <sup>2.3</sup> V <sup>2.1</sup> E <sup>6.3</sup>	I <sup>3.0</sup> V <sup>2.2</sup>	I <sup>1.7</sup> V <sup>3.4</sup>	I <sup>2.4</sup> V <sup>1.2</sup> E <sup>9.2</sup>	V <sup>2.5</sup> I <sup>1.8</sup> E <sup>1.5</sup>
11	V																I <sup>2.4</sup>
12	T	P <sup>1.5</sup> A <sup>1.2</sup>	S <sup>3.7</sup> P <sup>3.4</sup> A <sup>3.1</sup> N <sup>1.2</sup> I <sup>1.1</sup>	S <sup>8.4</sup> P <sup>2.3</sup> A <sup>1.4</sup>	S <sup>2.0</sup> A <sup>1.8</sup> P <sup>1.2</sup>	S <sup>5.3</sup> I <sup>4.0</sup> K <sup>3.6</sup> N <sup>3.3</sup> A <sup>1.9</sup> E <sup>1.9</sup> P <sup>1.5</sup> I <sup>1.2</sup>	A <sup>3.6</sup> K <sup>2.9</sup> P <sup>1.6</sup> S <sup>1.0</sup>	P <sup>1.7</sup>	A <sup>3.3</sup> K <sup>2.8</sup>	S <sup>1.1</sup> A <sup>1.1</sup>	P <sup>3.5</sup> S <sup>1.2</sup> A <sup>1.1</sup> N <sup>1.0</sup> K <sup>1.0</sup>	S <sup>5.5</sup> P <sup>1.7</sup> A <sup>1.2</sup>	S <sup>6.3</sup> P <sup>1.8</sup> K <sup>1.8</sup> E <sup>1.8</sup>	S <sup>3.4</sup> N <sup>2.9</sup> K <sup>2.2</sup> I <sup>2.2</sup> E <sup>2.1</sup> A <sup>1.6</sup>	A <sup>2.1</sup> K <sup>1.7</sup>	P <sup>2.9</sup> A <sup>1.9</sup>	A <sup>3.0</sup> K <sup>2.2</sup> S <sup>1.5</sup> V <sup>1.5</sup>

## Frequency of protease mutation at positions 10 to 99 from PI-naive persons



# Stanford HIV drug resistance database

## CPR- calibrated population resistance tool



STANFORD UNIVERSITY  
**HIV DRUG RESISTANCE DATABASE**

http://cpr.stanford.edu email: hivdb@stanford.edu

Code Version 6.0 » [ReleaseNotes](#)

Summary
Methods
QA Details
Complete Mutation List
Genetic Diversity

New analysis

Generated 2011-11-7 ( CPR report for download: [in pdf](#) and [in excel](#) )

### Number of input sequences

	No. submitted	No. filtered (QA)	No. analyzed for SDRMs
Sequences	12		
Sequences containing either RT or PR	12	0	12
Sequences containing RT	12	0	12
Sequences containing PR	12	0	12
Sequences containing both RT and PR	12	0	12

### Proportion of sequences with SDRMs

Resistance category	No. analyzed	No. containing SDRM	%
Sequences with any SDRM	12	3	25.0
PR Sequences with any PI SDRM	12	1	8.3
RT Sequences with any NRTI SDRM	12	2	16.7
RT Sequences with any NNRTI SDRM	12	1	8.3
RT Sequences with any NRTI + any NNRTI SDRM	12	1	8.3
PRRT Sequences with any NRTI + any NNRTI + PI SDRM	12	0	0.0

### Sequences with SDRMs

SequenceID	NRTI SDRMs	NNRTI SDRMs	PI SDRMs
D.TZ.2009.HM572334	<b>M184V</b>	<b>K103N</b>	None
D.TZ.2008.HM572349	None	None	<b>V82A</b>
D.TZ.2007.HM572350	<b>M41L</b>	None	None

Seq ID: D.TZ.2009.HM572334

### Summary Data

Sequence includes PR: codons: 1 - 99  
Sequence includes RT: codons: 1 - 238

There are no insertions or deletions

Subtype and % similarity to closest reference isolate:

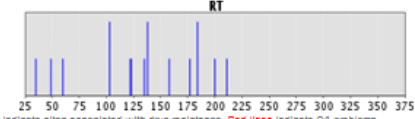
1. PR: D (95.3%)
2. RT: D (91.9%)

### Sequence Quality Assessment

Gene	QA Problem	Codons
PR	Stop Codons, Frame Shifts:	None
PR	Ambiguous Positions:	None
PR	Unusual Residues:	None



Gene	QA Problem	Codons
RT	Stop Codons, Frame Shifts:	None
RT	Ambiguous Positions:	None
RT	Unusual Residues:	None



Blue lines indicate differences from consensus B; tall blue lines indicate sites associated with drug resistance. Red lines indicate QA problems.

### Drug Resistance Interpretation: RT

NRTI Resistance Mutations: M184V  
NNRTI Resistance Mutations: K103N, E138Q  
Other Mutations: V35T, K49R, V60I, K122E, D123S, I135T, A158S, D177E, T200A, R211K

Nucleoside RTI		Non-Nucleoside RTI	
lamivudine (3TC)	High-level resistance	efavirenz (EFV)	High-level resistance
abacavir (ABC)	Potential low-level resistance	etravirine (ETR)	Low-level resistance
zidovudine (AZT)	Susceptible	nevirapine (NVP)	High-level resistance
stavudine (D4T)	Susceptible	rilpivirine (RPV)	Low-level resistance
didanosine (DDI)	Susceptible		
emtricitabine (FTC)	High-level resistance		
tenofovir (TDF)	Susceptible		

### RT Comments

- NRTI**
- M184V/I cause high-level in vitro resistance to 3TC and FTC and low-level in vitro resistance to ddI and ABC. M184V/I increase susceptibility to AZT, TDF, and d4T.
- NNRTI**
- K103N causes high-level resistance to NVP, and EFV. By itself it has no effect on ETR susceptibility. However, it may have a synergistic effect with L100I and possibly K101P on ETR susceptibility.
  - E138Q is a non-polymorphic mutation that emerges during ETR therapy. It reduces ETR and RPV susceptibility by ~2-3 fold.



Integration of viral genomics with clinical data to predict response to anti-HIV therapy

home / data analysis / documentation / links / database

HIV-1 Sequence Mutation List XML file

## HIV-1 sequence (mandatory)

Choose one of the two methods

### Text/fasta input

Paste sequence text in the text below

[Insert sample sequence](#)

```
>D.TZ.2009.HM572334
CCTCAAATCACTCTTTGGCAACGACCCCTTGTCCACAGTAAAAATAGGGGACAGCTAAAGGAAGCTCTATTAGATACAGG
AGCAGATGATACAGTCTTAGAAGAAATGAATTTGCCAGGAAATGGAAACCAAAAATGATAGGGGAATTGGAGGTTTTA
TCAAAGTTAGACAGTATGATCAAATAACCCTAGAAATCTGTGGACATAAAGCTATAGGTACAGTATTAGTAGGACCTACA
```

### Text/fasta file Upload

Choose a file to upload from your computer using the file selection box below

Browse...

### 10 BEST DRUG COMBINATIONS REGARDLESS OF YOUR SELECTION [1]

Rank	Regimen	Success probability	Range	Graphic bar
1	AZT ETV DRV/rv	98.87 %	97.53 % - 100.0%	
2	AZT TDF DRV/rv	98.55 %	97.1 % - 99.9%	
3	TDF ETV DRV/rv	98.3 %	95.75 % - 100.0%	
4	ABC ETV DRV/rv	98.16 %	95.47 % - 100.0%	
5	AZT DDI DRV/rv	97.74 %	94.94 % - 99.8%	
6	D4T ETV DRV/rv	97.66 %	93.75 % - 100.0%	
7	DDI ETV DRV/rv	97.28 %	92.65 % - 100.0%	
8	AZT NVP DRV/rv	96.62 %	91.71 % - 100.0%	
9	AZT EFV DRV/rv	96.44 %	90.21 % - 100.0%	
10	AZT 3TC DRV/rv	95.91 %	88.88 % - 100.0%	

### Summary of input data

#### GENOTYPIC DATA ENTERED

Individual mutations found/selected:

- Reverse transcriptase: V35T K49R V60I K103N K122E D123S I135T E138Q A158S D177E M184V T200A R211K
- Protease: I13V R41K L63T I64L V82I

### HIV mutations and subtype

- Sequence includes PR codons: 1 - 99
- Sequence includes RT codons: 1 - 238
- There are no insertions or deletions
- Subtype and % similarity to closest reference isolate: C (91.0%)
- NRTI Resistance Mutations: M184V
- NNRTI Resistance Mutations: K103N
- PI Major Resistance Mutations: I13V I64L V82IV

Start page EuResist prediction engine:  
[http://engine.euresist.org/data\\_analysis/viral\\_sequence/new](http://engine.euresist.org/data_analysis/viral_sequence/new)

## Previously used treatments

### Nucleoside RT inhibitors

- Abacavir
- Emtricitabine
- Didanosine
- Lamivudine
- Stavudine
- Tenofovir
- Zalcitabine
- Zidovudine

### Non-Nucleoside RT inhibitors

- Efavirenz
- Nevirapine

### Protease inhibitors

- Atazanavir
- Amprenavir/fosamprenavir
- Indinavir
- Lopinavir/Ritonavir
- Nelfinavir
- Ritonavir (boosting dose)
- Ritonavir (full dose)
- Saquinavir

## Drugs to be used to build the treatment regimen (optional)

### Nucleoside RT inhibitors

- Abacavir
- Emtricitabine
- Didanosine
- Lamivudine
- Stavudine
- Tenofovir
- Zidovudine

### Non-Nucleoside RT inhibitors

- Efavirenz
- Nevirapine
- Etravirine

### Protease inhibitors

- Atazanavir
- Atazanavir/Ritonavir
- FosAmprenavir/Ritonavir
- Indinavir/Ritonavir
- Lopinavir/Ritonavir
- Nelfinavir
- Saquinavir/Ritonavir
- Tipranavir
- Darunavir

## Baseline markers

HIV-1 RNA

copies/ml:

CD4+ T-cells

cells/microlitre:

## Demographic data

Gender:

Age:

Mode of hiv transmission:

## 10 BEST DRUG COMBINATIONS REGARDLESS OF YOUR SELECTION [1]

Rank	Regimen	Success probability	Range	Graphic bar
1	<b>D4T ETV</b> DRV/rv	99.47 %	99.24 % - 99.7%	
2	<b>TDF ETV</b> DRV/rv	99.42 %	99.15 % - 99.6%	
3	<b>DDI ETV</b> DRV/rv	99.45 %	99.17 % - 99.7%	
4	<b>AZT ETV</b> DRV/rv	99.46 %	99.08 % - 99.7%	
5	<b>ABC ETV</b> DRV/rv	99.35 %	99.01 % - 99.6%	
6	<b>FTC ETV</b> DRV/rv	99.04 %	98.4 % - 99.6%	
7	<b>3TC ETV</b> DRV/rv	99.15 %	98.4 % - 99.9%	
8	<b>AZT 3TC</b> DRV/rv	97.17 %	95.3 % - 98.85%	
9	<b>D4T 3TC</b> DRV/rv	96.06 %	93.59 % - 98.78%	
10	<b>DDI 3TC</b> DRV/rv	95.53 %	92.47 % - 98.52%	

## YOUR FILTERED LIST

Rank	Regimen	Success probability	Range	Graphic bar
1	<b>AZT 3TC</b> DRV/rv	97.17 %	95.3 % - 98.85%	
2	<b>D4T 3TC</b> DRV/rv	96.06 %	93.59 % - 98.78%	
3	<b>DDI 3TC</b> DRV/rv	95.53 %	92.47 % - 98.52%	
4	<b>ABC 3TC</b> DRV/rv	95.27 %	92.21 % - 98.49%	
5	<b>AZT DDI</b> DRV/rv	95.36 %	89.1 % - 98.52%	
6	<b>AZT TDF</b> DRV/rv	95.25 %	88.4 % - 98.68%	
7	<b>TDF FTC</b> DRV/rv	93.16 %	87.8 % - 98.46%	
8	<b>AZT 3TC</b> LPV/rv	80.28 %	75.4 % - 87.01%	
9	<b>D4T 3TC</b> LPV/rv	76.49 %	70.4 % - 85.19%	
10	<b>DDI 3TC</b> LPV/rv	74.11 %	69.4 % - 82.52%	

# Geno2Pheno - resistance

Page: [Input](#) [Results](#) [Help](#) [References](#) [Contact](#) [Disclaimer](#)

Sequence:

Feature: [Alignment](#) [Prediction](#) [Subtype](#) [Scored Mut.](#) [THEO](#)

[PDF](#)

## General Information

Sequence identifier: (No identifier given)  
 FASTA header: D.TZ.2009.HM572334  
 Date: November 8, 2011

## Phenotype Prediction

Drug	RF	Z-Score	Prob.Score
ZDV	2.613	1.106	0.05
ddl	2.058	2.460	0.75
d4T	0.989	-0.945	0.001
3TC	38.165	12.651	1
ABC	3.077	5.937	1
TDF	0.827	-1.410	0.002
NVP	61.812	4.246	1
EFV	30.405	6.163	1
SQV/r	0.924	-0.531	0
IDV/r	1.283	-0.185	0
NFV	0.757	-1.367	0
APV/r	1.087	0.027	0.001
LPV/r	1.080	0.168	0
ATV/r	1.026	-0.673	0

Subtype: [10\\_CD](#) 100% (0.999999984481)

Drug	RF(*)	z-score	Scored Mutations(**)
ZDV	2.613	1.106	49R 135T 158S 35T 60I 103N 123S
ddl	2.058	2.460	184V 123S 158S 177E
d4T	0.989	-0.945	158S 184V 35T 123S 211K 103N
3TC	38.165	12.651	184V
ABC	3.077	5.937	184V 35T 49R 214F 123S
TDF	0.827	-1.410	184V 60I 177E 135T 214F 49R
NVP	61.812	4.246	103N 135T 60I 158S
EFV	30.405	6.163	103N 158S 200A 135T 123S
SQV	0.924	-0.531	41K 82I 37N 64L
IDV	1.283	-0.185	82I 83T 41K
NFV	0.757	-1.367	82I 3I 41K 64L
APV	1.087	0.027	41K
LPV	1.080	0.168	82I 41K 64L
ATV	1.026	-0.673	41K 82I 3I 64L



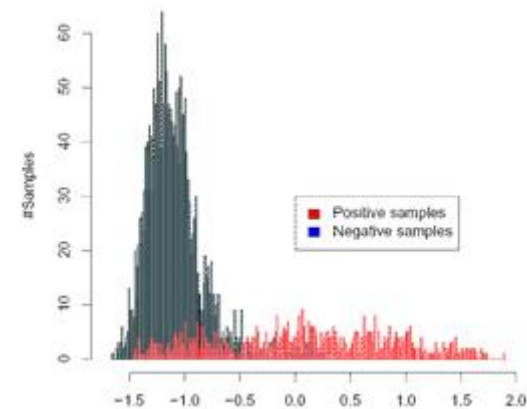
Start page Geno2Pheno: <http://www.geno2pheno.org>

# Geno2Pheno - coreceptor

## Geno2pheno [coreceptor] 1.2

Page: [Input](#) [Results](#) [Help](#) [Guidelines](#) [Celsentri](#) [Team](#) [Links](#) [Changes](#) [References](#) [Contact](#)

Nr	header	V3 loop	V3 subtype	FPR	percentile
0	D.TZ.1996.AF038099	CTRPSNNTRQSVHLGPGQALYTLKIIGDIRQAHC	D	54.4%	78.93%
1	D.TZ.1996.AF038104	CIRPYNNTNRQSTHIGPGRAYYTSNIIGDIRQAHC	D	3.3%	79.29%
2	D.TZ.1997.AF106359	CTRPSNNTRQSVHLGPGQALYTLKIIGDIRQAHC	D	54.4%	78.93%
3	D.TZ.2000.AY123853	CARPSNNTRRSIRIGPGQAFYTTGDIIGNIRQAYC	A or AG	16.5%	42.66%
4	D.TZ.2001.AY253311	CTRPNNNTRKGIRIGPGQTFFTAETGDIRKAYC	D	17.7%	76.66%
5	D.TZ.1995.DQ198968	CTRPYNNTNRQGTHIGPGRALFTTKIVGDIRQAHC	D	2.6%	87.1%
6	D.TZ.2005.EU263374	CTRPYNNTRTSIHMGPGRAWYTTDIIGDIRQAHC	B	5%	69.79%
7	D.TZ.2006.HM215284	CTRPYNNTNRQSIHIGPGQAIYATDARKIVGDIRQAYC	D	3.7%	79.8%
8	D.TZ.2004.HM215327	CTRPYNNTNRQRIHIGPGQAIWTNDARVIVGDIRQAHC	D	0.6%	92.55%
9	D.TZ.2008.HQ615975	CTRPYKNVRHRVHMGPGRAFVTRRITGDIRQAHC	D	0%	95.2%
10	D.TZ.2000.HQ659623	CTRPYNNTRTSEHMGPGRAFFTKITGDIRQAHC	B	1.7%	90.2%
11	D.TZ.1988.L01298	CTRPYSRVRQGAHIGPGRAYYTDDIVGDIRQAHC	D	1.7%	88.98%
12	D.TZ.1994.U61880	CTRPNNNTRQSTHIGPGQALYTTKIIGDIRQAHC	D	14.5%	73.64%
13	D.TZ.1987.U65075	CTRPYIKYRQKTSIGQGQALYTRLEPKPTGKIEPAYC	D	0.3%	98.79%





### 1. General information

Sequence identifier: No sequence identifier was provided.

FASTA header: D.TZ.1996.AF038099

Date: November 8, 2011

Selected false positive rate: Recommendations from the European Consensus Group on clinical management of HIV-1 tropism testing

Predicted V3-subtype: A or AG

### 2. Additional clinical parameters

Viral load 1000-10000 CCR5-genotype not determined

CD4-cell percentages not determined CD4-cell counts 51-100

CD8-cell counts not determined

### 3. Aligned V3 region

```

TGT ACA AGA CCC AAC AAC AAT ACA AGA AAA AGT ATA CAT ATA GGA CCA GGG AGA GCA TTT TAT ACA ACA GGA GAA ATA ATA GGA GAT ATA AGA CAA GCA CAT TGT
Consensus B: C T R P N N N T R K S I H I G P G R A F Y T T G E I I G D I R Q A H C
Query: C T R P S N N T R Q S V H L G P G Q A L Y - T - - - - - - - - - - -
TGC ACA AGG CCC AGC AAC AAT ACA AGA CAA AGT GTG CAT CTA GGA CCA GGG CAA GCA TTA TAT - ACA - - - - - - - - - - -

```

### 4. Predicted phenotype

[\[Help\]](#)

Model	Prediction	FPR	Remarks
Clonal	CCR5-antagonists like Maraviroc or Vicriviroc are likely to be effective.	99.2%	

This prediction is based on clonal training data and V3-sequences alone.

```

A GAA ATA ATA GGA GAT ATA AGA CAA GCA CAT TGT
E I I G D I R Q A H C
K P T G K I E P A Y C
AAA CCA ACA GGA AAG ATA GAA CCA GCA TAT TGC

```

### 4. Predicted phenotype

[\[Help\]](#)

Model	Prediction	FPR	Remarks
Clonal	CCR5-antagonists like Maraviroc or Vicriviroc are not likely to be effective.	0.2%	The 11/25 rule would predict this sequence as an X4-virus.

This prediction is based on clonal training data and V3-sequences alone.

# Sequence analysis

**Today we have downloaded sequences from a public database and performed:**

- Subtype classification
- Recombination analysis
- Detection of drug resistance mutations
- Selection of optimal antiretroviral therapy
- Determination of co-receptor usage (V3-region of env-gene)

Practise: Download a sequence set of your interest and analyse using the tools mentioned in this session

## **Tomorrow**

- Editing and alignment of sequences
  - Phylogenetic relationship between strains, time of most recent common ancestor (tMRCA)
-

# Please download and install the following programs before the alignment and phylogeny sessions tomorrow

BioEdit	<a href="http://www.mbio.ncsu.edu/bioedit/bioedit.html">http://www.mbio.ncsu.edu/bioedit/bioedit.html</a>
ClustalX	<a href="http://www.clustal.org">http:// www.clustal.org</a>
jModelTest	<a href="http://darwin.uvigo.es/software/jmodeltest.html">http://darwin.uvigo.es/software/jmodeltest.html</a>
BEAST, Tracer, FigTree	<a href="http://tree.bio.ed.ac.uk/software/">http://tree.bio.ed.ac.uk/software/</a>

---