HIV Recombination Overview

Presenter: Brian Foley

btf@lanl.gov

HIV Databases



Theoretical Biology and Biophysics,

Los Alamos National Laboratory





<u>www.hiv.lanl.gov</u> <u>seq-info@lanl.gov</u>



Presenter: Brian Foley

btf@lanl.gov

HIV Databases









Workshop Topics

HIV Intersubtype Recombination and Unique Circulating Recombinant Forms

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

Mechanism of Recombination: Template switching during reverse transcription, no DNA damage-repair enzymes needed.

Tools for detecting Intersubtype Recombination:

Overview of HIV-1 M group subtypes and CRFs:

Examples of what recombinants "look like":





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

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

Recombination occurs by template switching during reverse transcription of heterodimeric viruses

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



Virus Recombination Detection Tools:

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

- Pros: Adjustable window size, prebuilt test set, allows user input test set, adjustable significance threshold, contains consensus for each subtype.
- Cons: Does not output numerical breakpoint locations.
- jpHMMer: Gobics http://jphmm.gobics.de/submission_hiv.html
 - Pros: Statistical support of precise breakpoints, outputs table of breakpoints.
 - Cons: Does not have HMM models for CRFs, weak models for rare subtypes, does not include sububtypes (A3, A4, A6, F2), window size not adjustable.

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

- Pros: Accepts input of many sequences, phylogeny as well as other methods, CRFs updated reasonably often.
- Cons: Window size not adjustable,

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

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

SimPlot: Stuart Ray https://sray.med.som.jhmi.edu/SCRoftware/simplot/

Highlighter: HIV Databases https://www.hiv.lanl.gov/content/sequence/HIGHLIGHT/highlighter_top.html



Many more recombination detection tools:

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

Velcom	e to the comprehensive list of recombination analysis software maintained by the Robertson Lab.	
	Filter list by method or algorithm	
	Methods	
	Aigorithms	2
3seq • Des the • Imp • Ref	cription 3SEQ is a software program for identifying mosaic structure or recombination in nucleotide sequence data. 3SEQ takes as input a data set with a minimum of three aligned sequences, and it tests whether any sequence in demented algorithms # Recombination detection using hyper-geometric random walks # Recombination detection using hyper-geometric random walks	
4SIS		
Des valu Met Ade Ref	eription Two types of informative sites were distinguished, corresponding to the clustering of the putative recombinant with either of the parental representatives. The optimal breakpoint was located by maximizing a chi-square e. hods • Physicametrications itional notes temporarily unavailable. prence van Cuyck, H., Fan, J., Robertson, D.L. & Roques, P. (2005) [PubMed][doi]	
BARCE		
• Des • Met • Imp • Ado • Ref	cription BARCE is a C++ program for detecting recombination breakpoints in four-sequence alignments using hidden Markov models, Bayesian principles and Markov chain Monte Carlo sampling, • Phylogenetic match # Phylogenetic hidden Markov model with Bayesian interence • Hidden Markov models • Hidden Markov models • Hidden Markov models • Hidden Markov models • The Carlo Sample And Carlo Sa	
Bellerop	lon	
• Des • Imp • Ref	cription Bellerophon is a program for detecting chimeric sequences in multiple sequence datasets by an adaption of partial treeing analysis. emented algorithms • Datancematic collution across breakpoints prence Huber, T., Faulkner, G. & Hugenholtz, P. (2004) [PubMed][doi]	
BLAST G	enotyping	
• Des • Imp	cription This tool helps identify the genotype of a viral sequence. A window is slid along the query sequence and each window is compared by BLAST to each of the reference sequences for a particular virus. emented algorithms • Similaritydatarce joid	
cBrother		
• Des • Met • Imp • Ado • Ref	cription cBrother is software for inferring recombination when recombination is rare. This is a C version of the code originally written in Java available in DualBrothers. • Pryogenetic methods emented algorithms • Bayesian multiple change-point modelling • Bayesian multiple change-point modelling • Remoce Frang, F. (2005) [FoutMee][dos]	
DnaSP		
• Des bet • Met	cription DnaSP, DNA Sequence Polymorphism, is a software package for the analysis of nucleotide polymorphism from aligned DNA sequence data. DnaSP can estimate several measures of DNA sequence variation within and need populations (in noncoding, synonymous or nonsynonymous sites, or in various sorts of codon positions), as well as linkage disequilibrium, recombination, gene flow and gene conversion parameters. * Population genetics based renoce Linado, P. & Rozas, J. (2009) [PubMed][doi]	
DualBrot	hers	
• Des • Met	cription DualBrothers is a recombination detection software based on the dual Multiple Change-Point (MCP) model. This model allows for changes in topology and evolutionary rates across sites in a multiple sequence alignment. • Physicanetic methods	



How to decide "which tool"?

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

No single tool does everything very well.

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

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



HIV-1 subtypes and CRFs

- Subtypes and subsubtypes: A1 A6, B, C, D, F1 – F2, G, H, J, K
- Circulating Recombinant Forms: CRF01_AE CRF98_06B
- Both require a minimum of 2 complete genomes and at least one more partial genome with sequences from regions that can confirm the structure of the first 2.











The NCBI "Window BLAST" genotyping tool

CRF03_AB



CRF03_AB is recombinant between A6, the subsubtype of A found in the former Soviet Union region and B.

In this plot 2 reference genomes of CRF03_AB are included, along with A1, A2, B, C, F1, F2, and G genomes.

Window size 150 bases and step of 50 bases



CRF03_AB SimPlot



Los Alamos

CRF03_AB Maps



From publication describing CRF03_AB



From jpHMMer analysis at Gobics





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

••• •••



AliView (free for Mac) or BioEdit (free for Windows) multiple sequence alignment editor view of CRF03_AB plus reference genomes. <u>http://ormbunkar.se/aliview/</u> <u>http://www.mbio.ncsu.edu/BioEdit/bioedit.html</u>



							AliView - *	03_PlusRefs_GeneCut	_May2019.FASTA				
	2920	2930	2940 2950	2960	2970 298	80 2990	3000	3010 30	20 3030	3040	3050	3060	3070
0206A_Z321_U76035	GAAATTTGCAGA	GAAATGGAAA	ACALGGARARTT	TCARARTTGGGCC	- GARARTCCATACAA	CACTCCANTATTT	GCALTARGANA	AGGACAGTACTAA.	TEGREALATTAG	TAGATTTCAGAG	GACTORATES	ALGAACTCAA	GATTCTCCCCACC:
A2.CM.2001.01CM-1445MV.GU201516	GALLT CTAR	GACATGGAL	ICCILCCULLITT.	TCALLATTOGOCC	TOTAL TUCKTICAL	ACTCCAGINITI		LCTLL.	LTGGLGLLLLTTLG	TAGATTT <u>C</u> AGA(GALCTGALTAS	GAGARCTCAR	GACTTCTGGGAAG
A2.CY.1994.94CY017-41.AF286237 A3.SN.1996.DDI360.AY521630	GAALT CTGTARA	GAGALGGAL	• ··· · <i>•</i> ·					ACTAL ACTAL		TAGATTT <mark>T</mark> AGA(GAACTGAATAA GAGCTGAATAA	GAGAACTCAA	GACTTCTGGGAAG
A3.SN.2001.DDI579.AY521629	GCARTTTGTARG	GAGATGGAA	AliView	Hiahliah	nt Differei	nces fro	m Maioi	rity	TGGAGGAAATTG G	TAGATTTCAGAG	GAG <mark>CTGAATAA</mark>	ALGARCACAG	GACTTCTGGGAAG
A3.SN.2001.DDJ369.AY521631 A1.AU.2003.PS1044-Day0.DQ676872	GAAATTTGTA GAA	GAGATGGAA	/ 11 1 10 10	i ngi mgi			in majoi	LC LC LL	LTGGLGGLLLTTLG LTGGLGLLL G TTLG	TAGATTTCAGA(GAGCTCAATAA GAGCTCAATAA	ALGALCTCLS	GACTTTTGGGAAG
A1.RW.1992.92RW008.AB253421 A1.UC 1992.92UC037-A40 AB253429	GARATTTGTTCA	GAGATGGAA						ACTAL.	TGGAGAAAATTAG	TAGATTTCAGAG	GAGCTGAATAA	AAGAACACAA	GACTTTTGGGAAG
A6.BY.1997.97BL006.AF193275	GACATTTGTA	GAGATIGAAA	AGGALGGLALLATT	TCARAATTNGGCC	T G A A A A T C C A T A C A A	TACTCCAGTATTT	GCTATAAAGAAAA	AGGACAG <mark>C</mark> ACTAA	GTGGAGGAAATTAG	TABATTTCASE	GACCTGAATAA	ALGANCTCAC	GACTTTTNNGARG
A6.GE.1999.99GEMZ011.DQ207944 A6.KZ.2002.02KZKAR300435.EF589042	GACATTTGTART GACATTTGTARG	GATATGGAAA GAGATGGAAA		T CAAAAATT G G G C C T CAAAAATT G G G C C	T G A A A A T C C A T A C A A T G A A A A T C C A T A C A A	TACTCCAGTATTT	GCTATAAAGAAAA	AGGACAGCACTAA	G T G G A G G A A A T T A G G T G G A G G A A A T T A G	TAGATTTCAGGO	GAACTGAATAA Ga gc tgaataa	ALGALCTCLC ALGALCTCLC	GACTTTTGGGAAG'
A6.KZ.2002.02KZPAV300480.EF589043	GACATTTGTARG	GAGATGGAAA	AGGALGGALLATT	TCARARTTEGECC	- GAAAAT CCATACAA	TACTCCAGTATTT	GCTATAAAGAAAA	AGGACAGCACTAA	. T.C.A.C.A.A.T.T.A.C	TAGATTTCAGG	GACCTORATES	ANGANCTCAC	GACTTTTGGGAAG
A6.KZ.2002.02KZPAV300502.EF589044	GAC ATTTGTA 6	GALLTGGALL	AGGLLGGLLLLTT	TCARAAATTGGGCC	TGARARTCCATACAA	TACTCCAGTATTT	GCTATALAGALAA	AGGACAGCACTAA	GTGGAGGAAATTAG	TAGATTTCAGO	GAG CTGAATAA	LIGILCTCLC	GACTTTTGGGAAG
A6.KZ.2002.02KZYUZ300413.EF589040 A6.KZ.2002.02KZYUZ300425.EF589041	GACATTTGTA C	GAGATGGAAA GAGATGGAAA		T CAARAATT G G G C C	T G A A A A T <mark>C C A T A C</mark> A A T G A A A A T <mark>C C A T A C</mark> A A	TACTCCAGTATTT	GTTATAAAGAAAA		G T G G A G G A A A T T A G G T G G A G G A A A T T A G	TAGATTTCAGG	GAGCTCAATAA GAGCTCAATAA	ALGANCTCAG	GACTTTTGGGAAG
A6.RU.2000.RU00051.EF545108	GACATTTGTARG	GAGATGGAAA	AGGAAGGAAAAATT	TCARAATTGGGCC	TGARAR <mark>C</mark> CCATACAA	TACTCCAGTATTT	GCTATABAGABAB	AGGACAGCACTAA	TGGAGGAAATTAG	TAGATTTCAGO	GAACTGAATAA	ALGARCTCAC	GACTTTTGGGAAG
A6.UA.2000.98UA0116.AF413987	GAGATTTGTALG	GAGATGGAAA	AGGLLGGLLLLTT	TCARAAATTGGGCC	TGALLATCCATACAA	TACTCCAGTATTT	GCTATALAGAAAA	AGGACAGCACTAA	GTGGAGGAAATTAG	TAGATTTCAGO	GAG <mark>CTGAATAA</mark>	ALGALCTCLC	GACTTTTGGGAAG
A6.UA.2001.01UADN121.DQ823358 A6.UA.2001.01UADN139.DQ823357	GACATTTGTA G	GAGTTGGAAA	AGGALGGALLATT	TCARAATTGGGCC	T G A A A A T C C A T A C A A	TACTCCA TATTT	GCTATAAAGAAAA	AGGACAGCACTAA	CTGGAGAAAATTAG	TAGATTTCAGG	GAGCTGAATAA	ALGARCTCAC	GACTTTTGGGAAG
A6.UA.2001.01UAKV254.DQ823361	GACATTTGTA	GAGATGGAAA	AGGALGGALLATT	TCABABATTGGGCC	FGAAAATCCATACAA	TACACCAGTATTT	GCTATALAGAAAA	AGGACAGCACTAA		TAGATTTCAGG	GAGCTGAATAA	ALGARCTCAC	GACTTTTGGGAAG
A6.UA.2001.01UAOD10.DQ823365 A6.UA.2001.01UAOD35.DQ823366	GACATTTGTART GACATTTGTGAT	GAGATGGAAA	GGGLLGGLLLLTT	TCARARATTGGGCC TCARARATTGGGCC	TGARARTCCATACAA TGARARTCCATACAA	TACTCCAGTATTT	GCTATAAAGAAAA	AGGACAGCACTAA	G T G G L G G L L L T T G G G T G G L G G L L L T T L G	TAGATTTCAGGO	GAGCTCAATAA GAGCTCAATAA	ALGALCTCLC	GACTTTTGGGAAG
A6.UA.2001.01UAOD89.DQ823367	GACATTTGTA & G	GAGATGGAAA	AGGALGGALLATT	TCARARTTGGGCC	- GAAAAT CCATACAA	TACTCCAGTATTT	GCTATAAAGAAAA	AGGACAGCACTAA	T G G A G G A A A T T A G	TAGATTTCAGO	GAGCTGAATAA	ALGANCTCAC	GACTTTTGGGAAG
A6.UA.2001.01UAPol303.DQ823360	GACATTTGTA	GAGATGGAAA	AGGALGGLLLLTT	TCARAATTGGGCC	T G A A A A T C C A T A C A A	TACTCCAGTATTT	GCTATAAAGAAAA	AGGACAGCACTAA		TAGATTTCAGG	GACCTCALTAR	ALGANCTCAC	GACTTTTGGGAAG
A6.UZ.2002.02UZ0659.AY829209 A6.UZ.2002.02UZ0663.AY829210	GALATTTGTA C	G A G A T G G A G A	AGGALGGLLLLLTT	T CAAAAATT G G G C C T CAAAAATT G G G C C	T G A A A A T C C A T A C A A T G A A A A T C C A T A C A A	TACTCCA <mark>R</mark> TATTT TACTCCAGTATTT	GTTLTLLLGLLLL GCTLTLLLGLLLL		G T G G A G G A A A T T A G G T G G A G A A A A T T A G	TAGATTTCAGGO	GAG <mark>CTGAATAA</mark> GAG <mark>CTGAATA</mark> A	ALGALCTCLG	GACTTTTGGGAAG
A6.UZ.2002.02UZ0672.AY829212 A6.UZ.2002.02UZ652.AY829203	GACATTTGTARG	GAGATGGAGA	AGGAAGGAAAAATT	TCARARTTGGGCC	- GARARTCCATACAA	TACTOCALTATTT	GTTATAAAGAAAA	ANGATAGCACTAA	TECLECANTTAC	TAGATTTCAGO	GACCTCAATAA	ALGANCTCAC	GACTTTTGGGAAG
A6.UZ.2002.02UZ694.AY829205	GAC ATTTGTA & G	GAGATGGAGA	GGALGGLLLLTT	TCARAATTGGGCC	T G A A A A T C C A T A C A A	TACTOCALTATT	GTTATAAAGAAAA	ALGATIGC LCTLL	GTGGLGGLALTTLG	TAGATTTCAGG	GACCTCALTAR	ALGANCTCAC	GACTTTTGGGAAG
A6.UZ.2002.02UZ698.AY829206 A6.UZ.2002.02UZ740.AY829208	GACATTTGTA&G	GAGATGGAGA		T CARARATT G G G C C T CARARATT G G G C C	T G A A A A T <mark>C C A T A C</mark> A A T G A A A A T <mark>C C A T A C</mark> A A	TACTCCA ATATTT	GTTATAAAGAAAA		G T G G A G G A A A T T A G G T G G A G G A A A T T A G	TAGATTTCAGO	GAG <mark>CTGAATAA</mark> GAG <mark>CTGAATA</mark> A	ALGALCTCLC ALGALCTCLC	GACTTTTGGGAAG
03-AB.BY.2000.98BY10443.AF414006 03-AB.CR 2013 13774-1-51 3 ME109476	GACATTTGTAGG	GAGATGGAAA	AGGAAGGAAAAATT	TCARARTTGGGCC	- GAAAAT CCATACAA	TACTCCAGTATTT	GCCATARAGARA	ANGACAGTACTAN	TECLELLATIE	TAGATTTCAGAG	GAACTEAATAA	GAGAACTCAA	GACTTCTGGGAAG
03-AB.RU.1998.RU98001-98RU001.AF193277	GACATTTGTA	GAGATGGAAA	AGGALGGLLLLTT	TCARAATTGGGCC	T G A A A A T C C A T A C A A	TACTCCAGTATTT	GTCATAAAGAAAA	ALGACAGTACTAL.	TGGLGLLLTTLG	TAGATTTCAGA	GARCTLANTAS	GAGAACTCAA	GACTTCTGGGAAG
03-AB.RU.1997.KAL153-2.AF193276 B.GE.1998.98GEMZ003.DO207943	GAALCTTGTAC	G A G A T G G A A A	AGGALGGLALLATT	T CARARATT G G G C C T CARARATT G G G C C	T G A A A A T <mark>C C A T A C</mark> A A T G A A A A T <mark>C C A T A C</mark> A A	TACTCCAGTATTT TACTCCAGTATTT	GCCLTLLLGLLLL GCCLTLLLGLLLL		L T G G L G L L L T T L G L T G G L G L L L T T L G	TAGETTTCAGA(GAACTEAATAA GAACTEAATAA	GAGAACTCAA	GACTTCTGGGAAGT
B.RU.2011.11RU21n,JX500708	GAAATTTGTAGA	GALATGGAAA	AGGALGGALLATT	TCARARTTGGGCC	- GAAAATCCATACAA	TACTCCAGTATTT	GCTATAAAGAAAA	ALGACAGTACTAL.	TEGLELLITIE	TAGATTTEAGAC	GARCTLANTAS	GIGINCTCAN	GACTTCTGGGAAG
B.RU.2004.04RU129005.AY751406	GALATTTGTAGA	GALLTGGALL	ARGALGG CALLETT	TCEALATTGGGCC	T G A A A A T C C A T A T A A	TACTCCAGTATTT	GCCATAARGAAAA	ALGACAGTACTAL.	TGGLGLLLTTLG	TAGATTT		GAGAACTCAA	GACTTCTGGGAGG
B.RU.2004.04RU139089.AY751407 B.RU.2004.04RU139095.AY819715	GAAATTTGTA CA	GAAATGGAGA	AGGALGGLALLATT	T CAABAATT G G G C C T CAABAATT G G G C C	T G A A A A <mark>C</mark> C A T A C A A T G A A A A T C C A T A C A A	TACTCCAGTATTT TACTCCA TATTT	GCCATAAAGAAAA	AAGACAGTACTAA.	. T G G A G A A A A T T A G . T G G A G A A A A T T A G	TAGATTT <mark>H</mark> AGA(G A GC T E A A T A A G A A C T E A A T A A	GAGAACTCAA	GACTTCTGGGAAG
B.GE.2003.03GEMZ004.DQ207940 B.GE.2003.03GEMZ010.DQ207942	GARATTTGTACA	GAAATGGAAA	AGGALGGGALLATT	CARARTTEEECC	- GARRATCCATECES	TACTOCANTATTT	GCCATALAGAAAA	ALGACAG CALTAL	TGGLGLLLTTLG	TAGATTTCAGAG	GARCTERATES	GAGAACTCAA	GACTTCTGGGAAG
B.UA.2001.01UAKV167.DQ823362	GALATTTGTACA	GALLTGGALL	AGGALGGALLATT	TCARGANTTGGGCC	T G A C A A T C C A T A C A A	TACTCCAGTATTT	GCCATAAAGAAAA	ALGACAGTAC CAA	TGGLGLLLTTLG	TAGATTTCAGA		GAGAACCCAA	GACTTCTGGGAAG
B.UA.2001.01UAKV252.DQ823363 B.UA.2001.01UAKV259.DQ823364	GAAATTTGTA CA	GALLTGGALL	AGGALGGLLLLLTT	TCARARATTGGGCC TCARARATTGGGCC	TGARARTCCATACAA TGARARTCCATACAA	TACTCCAGTATTT TACTCCAGTATTT	GCCLTLLLGLLLL GCCLTLLLGLLLL	AAGACAGTACTAA. AAGACAGTACTAA.	LTGGLGLLLLTTLG LTGGLGLLLLTTLG	TAGATTTCAGA(TAGATTTCAGA(GAACTEAATAA GAACTEAATAA	GAGARCTCAN GAGARCTCAN	GACETCTGGGAAG
B.RU.2010.10RU6629,X500707 B.RU.2009.09RU4457,X500709	GARATTTGTACA	GALATGGALA	AGGALGGALLARTT	TCARARTTGGGCC	T G A A A A C C A T A T A A	TACTCCAGTATTT	GCCLTRALGARA	ALGACAGTACTAL.	TEGLERALTIE	TAGATTTEAGAG	GARCT <mark>t</mark> RATRA	GAGAACTCAA	GACTTCTGGGAAG
B.NL.2000.671-00T36.AY423387	GAAATTTGTAGA	GALCTGGALL	AGGALGGALLATT	TCARAATTGGGCC	T G A A A A T C C A T A C A A	TACTCCALTATTT	GCCATAAAGAAAA	ALGACAGCACTAL	TGGAGAAAATTAG	TAGATTTCAGA	GARCTERATES	ALGANCTCAL	GACTTCTGGGAAG
B.TH.1990.BK132.AY173951 B.FR.1983.HXB2-LAI-IIIB-BRU.K03455	GAAATTTGTA <mark>C</mark> A GAAATTTGTA <mark>C</mark> A	GAGATGGAAA	AGGALGGALLLATT	TCARARATTGGGCC TCARARATTGGGCC	TGARARTCCATACAA TGARARTCCATACAA	TACTCCAGTATTT TACTCCAGTATTT	GCCLTLLLGLLLL GCCLTLLLGLLLL	ALGACAGTACTAA. ALGACAGTACTAA.	LTGGLGLLLLTTLG LTGGLGLLLLTTLG	TAGATTTCAGA(TAGATTTCAGA(GAACTEAATAA GAACTEAATAA	GLGLLCTCLL	GACTTCTGGGAAG
B.US.1998.1058-11.AY331295 C BR 1992 BR025-d US2953	GAAATTTGTGTA	GALATGGALA	AGGALGGGLALLTT	TCARAATTGGGCC	GARARTCCATACAA	TACTCCAGTATTT	GCTATAAAAAAAA	ALGACAGTACTAL.	LTGGLGLLL <mark>G</mark> TTLG	TAGATTTEAGAG	GAACT <mark>E</mark> AATAA	GARANCTCAN	GACTTCTGGGAAG
C.ET.1986.ETH2220.U46016	GCAATTTGTGAA	GAATGGAGC	AGGALGGLALLATT	TCARGANTTGGGCC	T G A A A A C C A T A T A A	TACTCCAGTATTT	GCCATAAAAAAA	AGGACAGTACTAA		TAGATTTCAGO	GAACTGAATAA	ALGARCTCAR	GACTTTTGGGAAG
C.IN.1995.95IN21068.AF067155 C.ZA.2004.04ZASK146.AY772699	GCAATTTGTGAT GCAATTTGTGAG	GAGATGGAGA	AGGALGGALLART	ACAAAAAT <mark>C</mark> GGGCC	TGARARTCCATATAA TGARARTCCATATAA	CACCCCARTATT	GCCATABAAAAAA	AGGACAGTACTAA	G T G G L G L L L L T T L G G T G G L G L L L L T T L G	TAGATTTCAGGO TAGATTTCAGGO	GAACTCAATAA GaaCtcaataa	ALGARCTCAR	GACTTTTGGGAAG
D.CD.1983.ELI.K03454 D.CM 2001.01CM-4412HAL AV371157	GAAATTTGTAGA	GATATGGAAA	AGGALGGALLATT	TCARGARTTGGGCC	GARARTCCATACAA	TACTCCAATATTT	GCCATALAGAAAA	ALGACAGTACCAA	G T G G A G A A A A T T A G	TAGATTTCAGAG	GAACT SAATAA	GAGAACTCAA	GATTTCTGGGAAG
D.TZ.2001.A280.AY253311	GAAATTTGTACA	GALATGGALA	AGGALGGALLATT	TCARGANTTGGGCC	- GAAAATCCATACAA	TACTCCALTATTT	GCTATALAGAAAA	AAGACAGTACTAA	GTGGAGAAAATTAG	TAGATTTCAGAC	GAACTLAATAA	GAGAACTCAA	GATTTCTGGGAAG
F1.BE.1993.VI850.AF077336	GAAATATGT <mark>C</mark> A	GALLT GGALL	AGGILGGLLLLTT	TCARAATTGGGCC TCARAATTGGGCC	TGARAR <mark>O</mark> CCATACAA TGARARTCCATACAA	TACTCCARTATT	GCCLTLLGLLL	ALGACIGTICTIL	GTGGLGLGLLLTTLG LTGGLG G LLLTTLG	TAGATTTCAGA(GAACTEAATAA GAACTEAATAA	ALGARCTCAL	GATTTCTGGGAGG
F1.BR.1993.93BR020-1.AF005494	GAAATATGTALG	GAAATGGAAA	AGGALGGALLATT	TCARARTTGGGCC	- GAAAAT CCATACAA	TACTCCAGTATTT	GCCATALAGAAAA	ALGACAGTACTAL.	TGGAGGAAATTAG	TAGATTTCAGAG	GAACT <mark>E</mark> AATAA	ALGARCTCAR	GATTTTTGGGAGG
F1.FR.1996.96FR-MP411.AJ249238	GAAATATGTACA	GATATGGAAA		TCARARTTGGGCC	- GAAAATCCATACAA	TACTCCANTATTT	GCCATAAAGAAAA	ALGACAGTACTAL.	TEGLELLLTTLE	TAGATTTEAGAG	GARCTERATES	ALGANCTCAN	GATTTTTGGGAAG
F2.CM.1995.95CM-MP255.AJ249236 F2.CM.1995.95CM-MP257.AJ249237	GAAATCTGTAGA GAAATCTGTAGA	GAGATGGAAA	ALGALGGLLLLLTT	TCARARATTGGGCC TCARARATTGGGCC	1 G 1 1 1 1 7 7 7 7 1 7 1 7 1 1 1 G 1 1 1 1 7 7 7 7 7 1 7 1 1	TACTCCAGTATTT	GCCATAAAGAAAG	AGGACAGTACTAA.	LTGGLGLLLLTTLG LTGGLGLLLLTTLG	TAGATTTCAGA(TAGATTTCAGA(GAACTEAATAA GAGCTEAATAA	ALGARCTCAN ALGARCTCAN	GATTTTTGGGAGG GATTTTTGGGAGG
F2.CM.1997.CM53657.AF377956	GAAAT CTGTAGA	GAGATGGAAA	AGAAGGAAAATT	TCARAATTGGGCC	GAAAATCCATACAA	TACTCCAGTATTT	GCCATALAGAAAA	AGGACAGTACTAA.	TGGAGAAAATTAG	TAGATTTCAGAG	GAACTEAATAA	ALGARCTCAR	GATTTTTGGGAGG
G.BE.1996.DRCBLAF084936	GAAATTTGTA	GAGATGGAAA	AGGALGGALLATT	TCARAGATTGGGCC	T G A A A A C C C A T A T A A	CACTCCALTATT	GCCLTALLGLAS	LIGICIGTICTIC	TGGLGLLLTTGG	TAGATTTCAGAG	GAGCTGAATAA	ALGUACTCAL	GACTTCTGGGAGG
G.KE.1993.HH8793-12-1.AF061641 G.NG.1992.92NG083-JV10832.U88826	GALLTTTGTA CA	GALLTGGALL	BGGALGGALLALTT	TCARARTTGGGCC TCARARTTGGGCC	T G A A A A T C C A T A C A A T G A A A A T C C A T A T	CACTCCAATATTT	GCCLTLLGLLL	A G G A C A G T A C T A A.	LTGGLGLGLLLTTLG LTGGLGLLLLTT G G	TAGATTT T AGA(GARCTCAATAA GAACT <mark>b</mark> AATAA	ALGALCTCLL	GACTTCTGGGAAG
G.PTPT2695.AY612637 H RF 1993 VI991 AF190127	GALATTTGTAGA	GAATGGAAA	C C L L C C L L C L T T	TCARARTTGGGCC	GAAAAT CCATACAA	TACTCCANTATTT	GCCATAAAGAAAA	AAGACAGTACTAA.	TEGAGAAAATTAG	TAGATTTCAGAG	GACTORATES	AAGAACTCAA	GACTTCTGGGAGG
H.BE.1993.VI997.AF190128	GALATTTGTALG	GALLTGGALL	AGGALGGALLATT	TCARARATAGGGCC	FGAGAATCCATACAA	CACTOCALTATT	GCCLTLLL	ACGACAGTACTAL	LTGGLGLLLLTTLG	TGGATTTCAGAG	GALCTGLATAS	LIGILCTCLL	GACTTCTGGGALG
H.CF.1990.056.AF005496 H.GB.2000.00GBAC4001.FJ711703	GAAATTT GTA GA	GAGATGGAAA	A G G A A G G A A A A A T C	TCAAGAATAGGGCC TCAAGAATAGGGCC	TGAGAATCCATACAG	CACTCCAATATTT CACTCCAATATTT	GCCATAAAAAAA		LTGGLGLLLLTTLG LTGGLGLLLLTTLG	T G G A T T T C A G A G T G G A T T T C A G A G	GAACTGAATAA GA <mark>GCT</mark> GAATAA	ALGALCTCLL	GACTTCTGGGAAG'
J.CD.1997 J-97DC-KTB147.EF614151	CARATTTGTGAA	GALATGGALA	AGGALGGALLATT	TCARAATTGGGCC	GAAAATCCATATAA	CACTCCAGTGTTT	GCCATALAGAAAA	ALGACAGTACTAL	TEGLELLL TLC	TAGATTTTAGG	GACTORATES	ALGARCTCAL	GACTTCTGGGAAG
J.SE.1993.SE9280-7887.AF082394	CAATTTGT CT	GALLTGGALG	AGGALGGALLATT	TCALGAGTTGGGCC	- GAAAAT CCATATAA	CACTCCAGTATT	GCCLTLLGLLL	ALGACAGTACTAL.	TGGLGLLLTTLG	TAGATTTTAGAG	GARCTORATAS	ALGARCTCAL	GACTTCTGGGARG
K.CD.1997.97ZR-EQTB11.AJ249235 K.CM.1996.96CM-MP535.AJ249239	GAGATTTGTA GA	GALLTGGALL		TCARARATTGGGCC TCARARATTGGGCC	T G A A A A T C C A T A C A A T G A G A A T C C A T A T A A	CACCCCAGTGTTT CACTCCAGTGTTT	GCCATALAGAAAA	AAGACAGTACTAA.	LTGGL <mark>T</mark> AAAATTAG LTGGLGLAAATTAG	TAGATTTTAGAG	GAACTGAATAA GAACTGAATAA		GACTTCTGGGAAGT GACTTCTGGGAAGT
and a second													





2920	2930 2940 2950 2960 2970 2980 2990	3000 3010	3020	3030 3040	3050	3060 3070 3
0206A_Z321_U76035					· · · · G · · C · · · · · X ·	
A2.CD.1997.97CDKTB48.AF286	A A		C G A		• • • • • • • • • • • • • • • • • • • •	
A2.CY.1994.94CY017-41.AF28	A A	. <mark>T</mark>	c	· · · · · · · · · · · · · · · · · · ·	<u>.</u> <mark>c</mark> <u>.</u> .	· · · · · · · · · · · · · · · · · · ·
A1.RW.1992.92RW008.AB25342	Α	. <mark>.</mark> <u>.</u>	C	· · · 9 · · · · · · · · · · · · · · · · · · ·	GA.	· · · · · · · · · · · · · · · · · · ·
A1.UG.1992.92UG037-A40.AB2	A	. <mark>1</mark>	•	. G .		
A3.SN.2001.DDI579.AY521629a				<mark>G</mark>		<mark>A</mark> <mark>G</mark>
A6.BY.1997.97BL006.AF193275	Aliview Control-Click on Sequence	T	c c	N N	N G C A .	
A6.KZ.2002.02KZKAR300435.EI		T	C G		G G c	
A6.KZ.2002.02KZPAV300480.EF	to highlight differences from that	TTT	C G		GGCA. GGA.	
A6.KZ.2002.02KZYUZ300413.EL		T	C G		GG <mark>C</mark> A. GG <mark>C</mark> A.	
A6.KZ.2002.02KZYUZ300425.EI			<u>.</u>		GGA. G <u>.</u> A.	
A6.RU.2003.03RU20-06-13.AY A6.UA.2000.98UA0116.AF4139	· · · · · · · · · · · · · · · · · · ·	. <u>.</u>	C G		GG <mark>C</mark> A. GG <mark>C</mark> A.	
A6.UA.2001.01UADN121.DQ82 A6.UA.2001.01UADN139.DQ82		. T	C		GG <mark>C</mark> A. GGCA	
A6.UA.2001.01UAKV254.DQ82		. T	C G		G.G.C.A.	
A6.UA.2001.01UAOD35.DQ823	.	. <u>T</u>	c		GGA.	
A6.UA.2001.01UAPol294.DQ82	•••••••••••••••••••••••••••••••••••••••	TG	c G G		G	
A6.UZ.2002.02UZ0659.AY8292			c G G		G G c	
A6.UZ.2002.02UZ0672.AY8292	· · · · · · · · · · · · · · · · · · ·	T	c G K		GGA.	<mark>G</mark> <mark>T</mark>
A6.UZ.2002.02UZ694.AY82920		T	C G C		G G <mark>C</mark> A . G G <mark>C</mark> A	
A6.02.2002.0202698.A182920	······································	r T T	C G G		G G <mark>C</mark> <u>A</u> .	· · · · · · · · · · · · · · · · · · ·
03-AB.BY.2000.98BY10443.AF4	G G A <mark>G</mark> A ^T G G A A A A <mark>G G</mark> A A A A A TTT C A A A A A TT G G G C C T G A A A A T C C A T A C A A T A C T C C A <mark>G</mark> T A TTT G	с а <mark>т</mark> ааа <mark>с</mark> а алала са са с	TACTAAATGGAG	A A A T T A G T A G A T T T C A G	AGAACTTAATAAGA	λ G λ λ C T C λ λ G λ C T T C T G G G λ λ G
03-AB.GB.2013.13774-1-51.3.				<u> </u>		
B.GE.1998.98GEM2003.DQ2079	A	<u>.</u>		<u>.</u>		
	• • • • • • • • • • • • • • • • • • •	z		· · · · · · · · · · · · · <u>-</u> · · <mark>z</mark> · ·	<u>.</u> .	
B.RU.2004.04RU128005.AY682		2				
B.RU.2004.04RU128005.AY682 A B.RU.2004.04RU129005.AY751 A B.RU.2004.04RU129005.AY751 A B.RU.2004.04RU139095.AY819 A B.RU.2004.04RU139005.AY819 A B.RU.2004.04RU1390404 A B.RU.2004.04RU1390404 A B.RU.2004.04RU1390404 A B.RU.2004.04RU13904 A B.RU.2004.04RU1404 A B.RU.2004.04RU1404 A B.RU.2004.04R		R	A			
B.RU.2004.04RU128005.AY682. A B.RU.2004.04RU129005.AY751. A B.RU.2004.04RU139089.AY751. A B.RU.2004.04RU139095.AY819 A B.GE.2003.03GEMZ004.DQ2079 A G.G.Z003.03GEMZ004.DQ2079 A G.G.Z003.03GEMZ010.DQ2079 A G.G.Z003.03GEMZ0100000000000000000000000000000000000		R	a . a .			
B.RU.2004.04RU128005.AY682. A B.RU.2004.04RU129005.AY751. B B.RU.2004.04RU139089.AY751. A B.RU.2004.04RU139095.AY751. A B.RU.2004.04RU139095.AY751. A B.GE.2003.03GEMZ010.DQ2075 A B.GE.2003.03GEMZ010.DQ2075 A B.UA.2001.01UAKV167.DQ8233 A		R			a	
B.RU.2004.04RU128005.AY582. A B.RU.2004.04RU129005.AY751. B B.RU.2004.04RU139089.AY751. B B.RU.2004.04RU139095.AY751. B B.CE.2003.03CEM2004.DQ2075 A B.CE.2003.03CEM2010.DQ2075 A B.UA.2001.01UAKY167.DQ8235 A B.UA.2001.01UAKY252.DQ8235 A B.RU.2010.10RV6529.JX500707 A		1				
B. RU. 2004.04RU128005.AY582						
B. RU.2004.04RU128005.AY682 B. RU.2004.04RU129005.AY751 B. RU.2004.04RU139089.AY751. B. RU.2004.04RU139089.AY819 B. GE.2003.03GEMZ004.DQ2075 B. GE.2003.03GEMZ010.DQ2075 B. UA.2001.01UAKY167.DQ8233 B. UA.2001.01UAKY252.DQ8233 B. RU.2010.101UAKY259.DQ8233 B. RU.2010.101UAKY259.DQ8233 B. RU.2010.101UAKY259.DQ8233 B. RU.2010.101UAKY259.DQ8233 B. RU.2010.101UAKY253.DQ8233 B. RU.2010.101UAKY253.DQ8233 B. RU.2010.01UAKY253.DQ8233 B. RU.2010.01UAKY253.DQ8233 B. RU.2000.671-00736.AY42331 B. NL.2000.671-00736.AY42331 B. TH.1990.BK132.AY173951 B. FFL 1983.HX82-L41-IIIB-BRU.K			• • •			
B. RU.2004.04RU128005.AY582. B. RU.2004.04RU129005.AY751. B. RU.2004.04RU139089.AY751. B. RU.2004.04RU139095.AY819 B. GE.2003.03GEMZ004.DQ2075 B. GE.2003.03GEMZ010.DQ2075 B. UA.2001.01UAKY252.DQ8233 B. UA.2001.01UAKY252.DQ8233 B. RU.2010.01UAKY252.DQ8233 B. RU.2010.01UAKY259.DQ8233 B. RU.2010.01UAKY259.DQ8233 B. RU.2010.01UAKY259.DQ8233 B. RU.2010.01UAKY253.DQ8233 B. RU.2010.01UAKY253.DQ8233 B. RU.2010.09RU4457,1X500709 B. RU.2000.09RU4457,1X500709 B. RL 2000.09RU4457,1K500709 B. RL 2000.09RU4457,1K500709 B. RL 2000.8R1482-LA1-118-BRU.K B. FR. 1983.1K82-LA1-118-BRU.K B. FR. 1983.1K82-LA1-118-BRU.K B. G. B. 1992.BR025-d. US2953 A. B. A. B. B. A. B.			• • •			
B. RU.2004.04RU128005.AY582 B. RU.2004.04RU129005.AY751 B. RU.2004.04RU139005.AY751 B. RU.2004.04RU139005.AY519 B. GE.2003.03GEMZ004.DQ2075 B. GE.2003.03GEMZ010.DQ2075 B. GE.2003.03GEMZ010.DQ2075 B. UA.2001.01UAKV252.DQ8233 B. UA.2001.01UAKV252.DQ8233 B. RU.2010.01UAKV252.DQ8233 B. RU.2010.01UAKV252.DQ8233 B. RU.2010.01UAKV252.DQ8233 B. RU.2010.01UAKV253.DQ8233 B. RU.2010.01UAKV253.DQ8233 B. RU.2010.01UAKV253.DQ8233 B. RU.2010.01UAKV253.DQ8233 B. RU.2010.09RU4457.JX500709 B. RU.2000.09RU4457.JX500709 B. RU.2000.09RU4457.JX500709 B. RU.2000.08RU452.AV173951 B. FR.1983.HX82-LAI-IIIB-BRU.K B. US.1998.1058-11.AY331295 C. E. R1.992.BR025-du25953 C. ET.1986.ETH2220.U46016 C. IN.1995.95IN21068.AF06715 B. IN 2000.02120			• ** •			
B. RU.2004.04RU128005.AY582 B. RU.2004.04RU129005.AY751 B. RU.2004.04RU139089.AY751 B. RU.2004.04RU139089.AY751 B. GE.2003.03GEMZ004.DQ2075 B. GE.2003.03GEMZ010.DQ2075 B. GE.2003.03GEMZ010.DQ2075 B. UA.2001.01UAKV252.DQ8233 B. UA.2001.01UAKV252.DQ8233 B. RU.2010.01UAKV252.DQ8233 B. RU.2010.01UAKV252.DQ8233 B. RU.2010.01UAKV252.DQ8233 B. RU.2010.01UAKV252.DQ8233 B. RU.2010.01UAKV253.DQ8233 B. RU.2010.01UAKV253.DQ8233 B. RU.2010.01UAKV253.DQ8233 B. RU.2010.01UAKV253.DQ8233 B. RU.2010.01UAKV253.DQ8233 B. RU.2010.08RU32.AY13951 B. FR.1983.HX82-LAI-IIIB-BRU.K B. US.1998.1088-11.AY331295 C. E.R.1992.05811.AY331295 C. E.R.1995.51N21068.AF06715 C. ZA.2004.04ZAKS146.AY7726 B. C. DA1935.ELI.K03454 B. C. D.1985.ELI.K03454						
B. RU.2004.04RU128005.AY582 B. RU.2004.04RU129005.AY551 B. RU.2004.04RU139089.AY751 B. RU.2004.04RU139089.AY51 B. GE.2003.03GEMZ004.DQ2075 B. GE.2003.03GEMZ004.DQ2075 B. GE.2003.03GEMZ010.DQ2075 B. UA.2001.01UAKV252.DQ8233 B. UA.2001.01UAKV252.DQ8233 B. RU.2010.01UAKV252.DQ8233 B. RU.2010.01UAKV252.DQ8233 B. RU.2010.01UAKV252.DQ8233 B. RU.2010.01UAKV252.DQ8233 B. RU.2010.01UAKV252.DQ8233 B. RU.2010.01UAKV252.DQ8233 B. RU.2010.01UAKV253.DQ8233 B. RU.2010.01UAKV253.DQ8233 B. RU.2010.08RU342.Y173951 B. RT.1990.9RU342.Y173951 B. FT.1990.2RU25-4U52953 C. ET.1986.ETH2220.U46016 C. IN.1959.51N21068.AF06715 C. ZA.2004.04ZASK146.AY7726 D. CD.1983.ELLK03454 D. CD.1038.ELLK03454 D. CT.2001.02CM-4412HAL.AY						
B. RU.2004.04RU128005.AY581 B. RU.2004.04RU129005.AY551 B. RU.2004.04RU139089.AY551 B. RU.2004.04RU139089.AY51 B. GE.2003.03GEMZ004.DQ2075 B. GE.2003.03GEMZ014.DQ2075 B. GE.2003.03GEMZ014.DQ2075 G. GEMZ010.012.GEMZ						
B. RU.2004.04RU128005.AY581 B. RU.2004.04RU129005.AY551 B. RU.2004.04RU139089.AY751 B. RU.2004.04RU139089.AY51 B. GE.2003.03GEM2004.DQ2075 B. GE.2003.03GEM2004.DQ2075 B. GE.2003.03GEM2014.DQ2075 B. GE.2003.04044.25.DQ233 B. GE.2003.04044.25.DQ233 B. GE.1986.1058-11.AV331295 C. GE.1985.95.102.1068.AF06715 D. CD.1983.91.KIS0.AF07233 FI. BE.1993.93.8K020-1.AF0054 FI. BE.1993.93.8K020-1.AF0054						
B. RU.2004.04RU128005.AY581 B. RU.2004.04RU129005.AY551 B. RU.2004.04RU139089.AY751 B. RU.2004.04RU139089.AY51 B. GE.2003.03GEMZ014.0Q2075 B. GE.2003.09RU4457.1X500709 B. ST.1.990.9RU32.4Y173951 B. FT. 1990.9RU32.4Y173951 B. GE.11928.0058-11.4Y1312.95 C. GE.11986.1058-11.4Y1312.95 C. GE.11986.1058-11.4Y1312.95 C. GE.11986.1058-11.4Y1312.95 C. GE.11986.1578-11.4Y1312.95 C. GE.11986.1578-11.4Y1312.95 C. GE.11985.1058-11.4Y1312.95 C. GE.11985.1058-11.4Y1312.95 C. GE.11993.91850.4F06715 G. GE.11923.146.144 D. CD. 1983.91K150.4F77336 FI. BE.1993.91K802.0-1.4F0054 FI. BE.1993.91K930.4F075703 FI. FF.1993.95 C.ME.0755.01 FI. FF.1993.95 C.ME.0755.01 FI. FF.1993.95 C.ME.0755.01 FI. FF.1993.95 C.ME.0755.01 FI. FF.1993.95 C.ME.0755.01 FI. FF.1993.95 C.ME.0755.01 FI. FF.1995.95 C.ME.0555.01 FI. FF.1995.95 C.ME.0555.01 FI. FF.1995.95 C.ME.0555.01 FI. FF.1995.95 C.ME.0755.01 FI. FF.1995.95 C.ME.0555.01 FI. FF.1995.95 C.ME.0755.01 FI. FF.1995.95 C.ME.0555.01 FI. FF.1995.95 C.ME.0555.01 FI. FF.1995.95 C.ME.0755.01 FI. FF.1995.95 C.ME.0755.01 FI. FF.1995.95 C.ME.0755.01 FI. FF.1995.95 C.ME.0755.01 FI. FF.1995.95 C.ME.0555.01 FI. FF.1995.95 C.ME.0755.01 FI. FF.1995						
B. RU.2004.04RU128005.AY581 B. RU.2004.04RU129005.AY551 B. RU.2004.04RU139089.AY751 B. RU.2004.04RU139089.AY751 B. GE.2003.03GEMZ010.DQ2075 B. GE.2003.03GEMZ010.DQ2075 B. GE.2003.03GEMZ010.DQ2075 B. GE.2003.03GEMZ010.DQ2075 B. UA.2001.01UAKV252.DQ8233 B. UA.2001.01UAKV252.DQ8233 B. UA.2001.01UAKV252.DQ8233 B. RU.2010.01UAKV252.DQ8233 B. RU.2010.01UAKV252.DQ8233 B. RU.2010.01UAKV252.DQ8233 B. RU.2010.01UAKV252.DQ8233 B. RU.2010.09RU4457.JX500709 B. RL.2000.9RU432.Y173951 B. FTH.1990.9RU32.AY173951 B. FTH.1990.9RU32.AY173951 C. FT.1986. ICT521.046016 C. IN.1995.9SIC3-4U52953 C. Z. Z004.04ZASK146.AY7726 D.CD.1983.ELI.K03454 D.CM.2001.AI2KA142.AX124 D.CM.2013.VISD.AF077336 F1. BE.1993.YI850.AF077336 F1. BE.1993.SIR020-1.AF0054 F1. F1.9193.9SIK020-7.1AF0754 F1. F1.9193.9SIK020-7.1AF0754 F1. F1.9193.9SIK020-7.1AF0754 F1. F1.9193.9SIK020-7.1AF0754 F1. F1.9193.9SIK020-7.1AF0054 F1. F1.9193.9SIK020-7.1AF0754 F1. F1.9193.9SIK020-7.1AF0054 F1. F1.9193.9SIK020-7.1AF0054 F1. F1.9193.9SIK03.AF075703 F1. F1.9195.9SCM-WP255.A124 F2.CM.1995.9SCM-WP255.A124 F2.CM.1995.9SCM-WP255.A124 F2.CM.1995.9SCM-WP255.A124 F2.CM.1995.9SCM-WP255.A124 F2.CM.1995.9SCM-WP255.A124 F2.CM.1995.9SCM-WP255.A124 F2.CM.1995.9SCM-WP255.A124 F2.CM.1995.9SCM-WP255.A124 F2.CM.1995.9SCM-WP255.A124 F2.CM.1995.9SCM-WP255.A124 F2.CM.1995.9SCM-WP255.A124 F2.CM.1995.9SCM-WP255.A124 F2.CM.1995.9SCM-WP255.A124 F2.CM.1995.9SCM-WP255.A124 F2.CM.1995.9SCM-WP255.A124 F2.CM.1995.9SCM-WP255.A124 F2.CM.1995.9SCM-WP255.A124 F2.CM.1995.9SCM-WP255.A124 F3.CM.1995.9SCM-WP255.A124 F3.CM.1995.9SCM-WP255.A124 F3.CM.1995.9SCM-WP255.A124 F3.CM.1995.9SCM-WP255.A124 F3.CM.1995.9SCM-WP255.A124 F3.CM.1995.9SCM-WP255.A124 F3.CM.1995.9SCM-WP255.A124 F3.CM.1995.9SCM-WP255.A124 F3.CM.1995.9SCM-WP255.A124 F3.CM.1995.9SCM-WP255.A124 F3.CM.1995.9SCM-WP255.A124 F3.CM.1995.9SCM-WP255.A124 F3.CM.1995.9SCM-WP255.A124 F3.CM.1995.9SCM-WP255.A124 F3.CM.1995.9SCM-WP255.A124 F3.CM.1995.9SCM-P1255.A124 F3.CM.1995.9SCM-WP255.A124						
B. RU.2004.04RU128005.AY581 B. RU.2004.04RU129005.AY551 B. RU.2004.04RU139089.AY751 B. RU.2004.04RU139089.AY751 B. GE 2003.03GEMZ010.DQ2075 B. GE 2003.09RU4457.1X500709 B. RU.2009.09RU4457.1X500709 B. RU.2009.09RU4457.1X500709 B. RU.2009.09RU4457.1X500709 B. TH. 1990.09RU452.1X500709 B. TH. 1990.09RU452.1X500709 B. TH. 1990.09RU452.1X500709 C. B. T1925.05E - 11.47331295 C. BE 1992.BR025-d.U52953 C. BE 1992.BR025-d.U52953 C. GE 1.986.1CH2220.U46016 C. IN.1995.95IN21068.AF06715 C. ZA. 2004.04ZASX146.4Y7726 D. CD.1983.ELI.K03454 D. CM.2001.01CH.44121HALAY D. TZ. 2001.A280.AY25311 D. UG.1994.94UG114.U88824 F1. BE.1993.918N930.AF075733 F1. BE.1993.918N930.AF075733 F1. BE.1993.918N930.AF075733 F1. FR.1995.95CM-MP255.AJ24 F2.CM.1995.95CM-MP255.AJ24 F2.CM.1995.95CM-MP255.AJ24 F2.CM.1995.95CM-MP255.AJ24 F2.CM.1995.95CM-MP255.AJ24 F2.CM.1995.95CM-MP255.AJ24 F2.CM.1995.95CM-MP255.AJ24 F2.CM.1995.PSCM-MP255.AJ24 F2.CM.1995.PSCM-MP255.AJ24 F2.CM.1995.PSCM-MP255.AJ24 F2.CM.1995.PSCM-MP255.AJ24 F2.CM.1995.PSCM-MP255.AJ24 F2.CM.1995.PSCM-MP255.AJ24 F2.CM.1995.PSCM-MP255.AJ24 F2.CM.1995.PSCM-MP255.AJ24 F2.CM.1995.PSCM-MP255.AJ24 F2.CM.1995.PSCM-MP255.AJ24 F2.CM.1995.PSCM-MP255.AJ24 F2.CM.1995.PSCM-MP255.AJ24 F2.CM.1995.PSCM-MP255.AJ24 F3.CM.1995.PSCM-MP255.AJ24 F3.CM.1995.PSCM-MP255.AJ24 F3.CM.1995.PSCM-MP255.AJ24 F3.CM.1995.PSCM-MP255.AJ24 F3.CM.1995.PSCM-MP255.AJ24 F3.CM.1995.PSCM-MP255.AJ24 F3.CM.1995.PSCM-MP255.AJ24 F3.CM.1995.PSCM-MP255.AJ24 F3.CM.1995.PSCM-MP255.AJ24 F3.CM.1995.PSCM-MP255.AJ24 F3.CM.1995.PSCM-MP255.AJ24 F3.CM.1995.PSCM-MP255.AJ24 F3.CM.1995.PSCM-MP255.AJ24 F3.CM.1995.PSCM-MP255.AJ24 F3.CM.1995.PSCM-MP255.AJ24 F3.CM.1995.PSCM-MP255.AJ24 F3.CM.1995.PSCM-MP255.AJ2						
B. RU.2004.04RU128005.AY581 B. RU.2004.04RU128005.AY551 B. RU.2004.04RU139089.AY751 B. RU.2004.04RU139089.AY751 B. GE 2003.03GEM2010.D02075 B. GE 2003.09RU4457.1X500709 B. RU.2009.09RU4457.1X500709 B. RU.2009.09RU452.1X500709 B. RU.2009.09RU452.1X500709 B. RU.2009.09RU452.1X500709 B. RU.2009.09RU452.1X500709 B. RU.2009.09RU452.1X500709 C. RE.1998.05S-11.AY331295 C. RE.1998.05S-11.AY331295 C. RE.1998.05S-11.AY331295 C. RE.1998.05S-11.AY331295 C. RE.1993.61CF1220.046016 C. IN.1995.95IN21068.AF06715 C. ZA.2004.04ZASL46A247726 D. CD.1983.ELIK03454 D. CM.2001.01CM-4412HALAY D. CM.2001.01CM-4412HALAY D. CM.2001.01CM-4412HALAY D. CM.2001.02CM-407336 F1. RE.1993.91RN936.AF075703 F1. RE.1993.91RN936.AF075703 F1. RE.1993.91RN9363.AF075703 F1. RE.1993.95CM-MP255.AJ24 F2.CM.1995.95CM-MP255.AJ24 F3.CM.1995.95CM-MP255.AJ24 F3.CM.1995.95CM-MP255.AJ24 F3.CM.1995.95CM-MP255.AJ24 F3.CM.1995.95CM-MP255.AJ24 F3.CM.1995.95CM-MP255.AJ24 F3.CM.1995.95CM-MP255.AJ24 F3.CM.1995.95CM-MP255.AJ24 F3.CM.1995.95CM-MP255.AJ24 F3.CM.1						
B. RU.2004.04RU128005.AY581 B. RU.2004.04RU128005.AY551 B. RU.2004.04RU139089.AY751 B. RU.2004.04RU139089.AY751 B. GE 2003.03GEM2010.D02075 B. GE 2003.09RU4457.1X500709 B. RU.2009.09RU4457.1X500709 B. RU.2009.09RU4457.1X500709 B. RU.2009.09RU4457.1X500709 B. RU.2009.09RU4457.1X500709 B. TH. 1990.09RU452.1X500709 B. TH. 1990.09RU452.1X500709 B. TH. 1990.09RU452.1X500709 C. B. 1985.10S8-11.A.Y1331295 C. B. 1992.8R025-d.U52953 C. B. 1992.8R025-d.U52953 C. C. L. 1985.6FTH2220.U46016 C. IN. 1995.9SIN21068.AF06715 C. ZA. 2004.04ZAS146A.Y7726 D. CD.2001.A280.AY253311 D. CL.2001.01CM-4412HALAY D. TZ.2001.A280.AY253311 D. CL.999.49LU514.U48824 F. I. B. 1993.9LN9363.AF077336 F. I. B. 1993.9LN9363.AF077336 F. I. B. 1993.9LN9363.AF077336 F. I. B. 1993.9LN9363.AF075703 F. I. F. 1995.9SCM-MP255.AJ24 F. 2. CM.1995.9SCM-MP255.AJ24 F. 2. CM.1995.9SCM-MP255.AJ24 F. 2. CM.1997.CM53657.AF37795 F. 2. CM.2002.02CM-0016B8Y.AY C. BE 1.996.DRCBLAF084936 C. KE. 1993.HH8793-12-1AF06 C. KC. 1993.2HR973-12-1AF06 C. KC. 1993.2HR073-12-1AF06 C. KC. 1993.2HR073-12-1AF0						
B. RU.2004.04RU128005.AY581 B. RU.2004.04RU128005.AY551 B. RU.2004.04RU139089.AY751 B. RU.2004.04RU139089.AY751 B. GE 2003.03CEM2004.DQ2075 B. GE 2003.03CEM2010.DQ2075 B. GE 2003.04CM4571.X500709 B. RU.2009.08K132.AY17351 B. GE 1998.K132.AY17351 B. GE 1998.K132.AY17351 B. TH. 1999.08K132.AY17351 B. TH. 1990.08K132.AY17351 C. BE. 1998.1058-11.AY1331295 C. BE. 1992.BR025-d.U52953 C. ET. 1986.CFTH2220.U46016 C. IN.1995.95IN21068.AF06715 C. ZA. 2004.04ZASK146.AY7726 D. CD. 101.01CM-4412HAL.AY D. TZ. 2001.A280.AY23311 D. GL.993.4UL0114.U88824 F1.BE. 1993.V1850.AF075703 F1.FR.1995.95CM-MP255.AJ24 F2.CM.1995.95CM						
B. RU.2004.04RU128005.AY581 B. RU.2004.04RU128005.AY551 B. RU.2004.04RU139089.AY751 B. RU.2004.04RU139089.AY751 B. G.E.2003.03CEMZ010.DQ2075 B. G.E.2003.09RU4457.1X500709 B. RU.2009.09RU4457.1X500709 B. RU.2009.09RU4457.1X500709 B. RU.2009.09RU4457.1X500709 B. RU.2009.09RU4457.1X500709 B. RU.2009.09RU452.1X500709 B. RU.2009.09RU452.1X500709 B. RU.2009.09RU452.1X500709 B. RU.2009.09RU452.1X500709 B. RU.2009.09RU452.1X500709 B. RU.2009.09RU452.1X500709 B. RU.2009.09RU452.1X500709 B. RU.2009.04RU452.1X500709 B. RU.2009.04RU452.1X500709 B. RU.2009.04RU452.1X500709 B. RU.2009.04RU452.1X500709 B. RU.2009.04RU452.1X500709 B. RU.2009.01CLM-44124ALAY D. CZ.2001.01CLM-44124ALAY D. CZ.2001.01CLM-44124ALAY D. CZ.2001.01CLM-44124ALAY D. CZ.201.028A.0727336 F. J. R. 1993.91N9363.AF075703 F. J. R. 1995.95CLM-MP255.AJ24 F.2.CM.1995.95CLM-MP255.AJ24 F						
B. RU.2004.04RU128005.AY581 B. RU.2004.04RU129005.AY551 B. RU.2004.04RU139089.AY751 B. RU.2004.04RU139089.AY751 B. RU.2004.04RU139089.AY751 B. G.E.2003.03CEMZ010.D02075 B. G.E.2003.03CEMZ010.D02075 B. G.E.2003.03CEMZ010.D02075 B. G.E.2003.03CEMZ010.D02075 B. G.E.2003.03CEMZ010.D02075 B. G.E.2003.03CEMZ010.D02075 B. G.E.2003.03CEMZ010.D02075 B. G.E.2003.03CEMZ010.D02075 B. G.E.2003.03CEMZ010.D02075 B. G.E.2003.09RU4457.1X500709 B. RU.2009.09RU4457.1X500709 B. RU.2009.09RU4457.1X500709 B. RU.2009.09RU4457.1X500709 B. RU.2009.09RU4457.1X500709 B. RU.2009.09RU452.1X500709 B. RU.2009.01CM-441244.1X424 F. LE.1993.V1850.AF077336 F. J. R.1995.95CM-MP255.A124 F2.CM.1995.						



Positions or sites in HIV-1 genomes are numbered using alignment to the HXB2 reference genome as the standard. The HIV Map drawing tool can be used to create maps of the genome colored by region. https://www.hiv.lanl.gov/content/sequence/DRAW_CRF/recom_mapper.html



jpHMM-HIV [Result]

Input data:

sequence file

http://jphmm.gobics.de/submission_hiv

jpHMM result:

Sequence #1: >03-AB_BY_2000_98BY10443_AF414006

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

Fragment Start Position	Uncertainty Region Start - End	Breakpoint Interval Start - End	Fragment End Position	Fragment Subtype
Position in the original seq	uence [pred_recombination	n], [recombination_incl_UR	and BPI], [UR and BPI]	
1	-		798	N/A
799	-	2673 - 2736	2690	A1
2691	4779 - 5120	8587 - 8614	8603	В
8604	-	-	9368	A1
9369	-		9687	N/A
Position based on HXB2 n	umbering [pred_recombina	ation] [recombination_incl_	UR_and_BPI] [UR_and_BF	<u>2]</u>
1	-		789	N/A
790	-	2670 - 2733	2687	A1
2688	4776 - 5117	8633 - 8660	8649	В
8650	-	-	9411	A1
9412	-	-	9719	N/A

Genome map (based on HXB2 numbering)



Note

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

Posterior probabilities of the subtypes (based on HXB2 numbering)



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

vs input query sequence location

vs HXB2 standard sequence location



DATABASES SEARCH ALIGNMENTS TOOLS GUIDES Search PUBLICATIONS earch site

https://www.hiv.lanl.gov/content/sequence/DRAW_CRF/recom_mapper.html

Recombinant HIV-1 Drawing Tool

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

Input

	Upload text file of breakpoints	Browse No file selected. This is an Excel spreadsheet	v to
	Or enter breakpoint data here	1 789 N/A 790 2673 A6 2674 2733 A6/B 2734 8649 B 4766 5117 A6/B 5118 8633 B 86634 8659 A6/B 8660 9411 A6 9412 9719 N/A	C
	If the breakpoint data entered above are not expressed in HXB2 coordinates, check this box and paste sequence here	Non-HXB2 coordinates	
	Or upload your sequence file Please choose your subtype colors	Browse No file selected. A6 FF0000 click in textbox to change default B 3F98F2 click in textbox to change default	
		Submit Reset	
		jpHMMer_HIV-DB-Map.png	
5″ 1	-395 +- 394 -2704+-30	921	-9566+-154
		pol vpr env	S LFR

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



-9566+-154 -9719



https://sray.med.som.jhmi.edu/SCRoftware/simplot/





Even a very small region of misalignment, hypermutation, or poor sequence quality can have a large impact on similarity plots, phylogenetic trees, and other analyses. Similarity plots can be quite useful for identifying sites in a multiple sequence alignment that should be scrutinized, and corrected if in error, as this example shows.





likely recombination and not just an aberrant region.

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







REGA HIV-1 & 2 Automated Subtyping Tool (Version 2.0)

This tool is designed to use phylogenetic methods in order to identify the subtype of a specific sequence. The sequence is analysed for recombination using bootscanning methods.

Note for batch analysis: The REGA subtype tool accepts up to 1000 sequences at a time.

Enter here your input data as FASTA format.

https://www.genomedetective.com/app/typingtool/hiv

Choose a mirror to subtype your sequences

>03-AB.RU.1998.RU98001-98RU001.AF193277	
aaatctctagcagtggcgcccgaaca-gggacttg- aaagcgaaagttccagagaagatctctcgacgc-aggactcggcttgctg-aggtgc- acacagcaagagggc-gagagcggcgactggtgagtacgcctaa-agat-tTTTGACTAGCGGAG GCTAGAAGGAGAGAG	1.
clear Run	

Submit sequences How to cite HIV Tutorials HIV Decision Trees HIV Subtyping Process HIV Example Sequences Contact us

Developed by: Tulio de Oliveira, Koen Deforche, Sharon Cassol, Andrew Rambaut and Anne-Mieke Vandamme.

Developed in cooperation with the <u>Evolutionary Biology Group</u> at University of Oxford, UK., the <u>HIV-1 Pathogenesis and Immunotherapeutics</u> <u>Program</u> at University of Pretoria, South Africa, and the <u>REGA Institute</u> at the Katholieke Universiteit Leuven, Belgium. Funded by the Marie Curie Fellowship, Flanders Bilateral Cooperation Program and the Wellcome Trust (grant # 061238).

For questions, suggestions or problems please contact: Dr. Tulio de Oliveira.



HighLighter for intra-patient recombination

https://www.hiv.lanl.gov/content/sequence/HIGHLIGHT/highlighter_top.html



Mismatches compared to master

Dual-infected patient.

Infected with two strains of the same subtype.



RAPR recombination analysis program

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

Base Number



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

