Supplementary Material

Combined antiretroviral therapy and immune pressure lead to *in vivo* HIV-1 recombination with ancestral viral genomes

Maria José Buzón[1], Terri Wrin[2], Francisco M. Codoñer[1], Judith Dalmau[1], Anna Bonjoch[3], Eoin Coakley[2], Bonaventura Clotet[1,3], and Javier Martinez-Picado[1,4*]

[1]Fundació IrsiCaixa, Hospital Universitari "Germans Trias i Pujol", Universitat Autònoma de Barcelona (UAB), Badalona, Spain
[2]Monogram Biosciences, South San Francisco, USA
[3]Fundació de Lluita contral la SIDA, Badalona, Spain
[4]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

Figure S1. Virological and immunological outcome of antiretroviral therapy. Red boxes represent the period with the protease inhibitor–sparing regimen. Grey lines represent the time points selected for phenotypic analysis at which plasma HIV-1 RNA was amplified, sequenced and cloned for further phenotypic assays. pVL, plasma viral load (HIV-1 RNA copies/ml); ABC, abacavir; AZT, zidovudine; 3TC, lamivudine; TDF, tenofovir; DDI, didanosine; D4T, stavudine; EFV, efavirenz; NVP, nevirapine; IDV, indinavir; LPV, lopinavir; RTV, ritonavir; SQV, saquinavir; APV, amprenavir; NFV, nelfinavir.

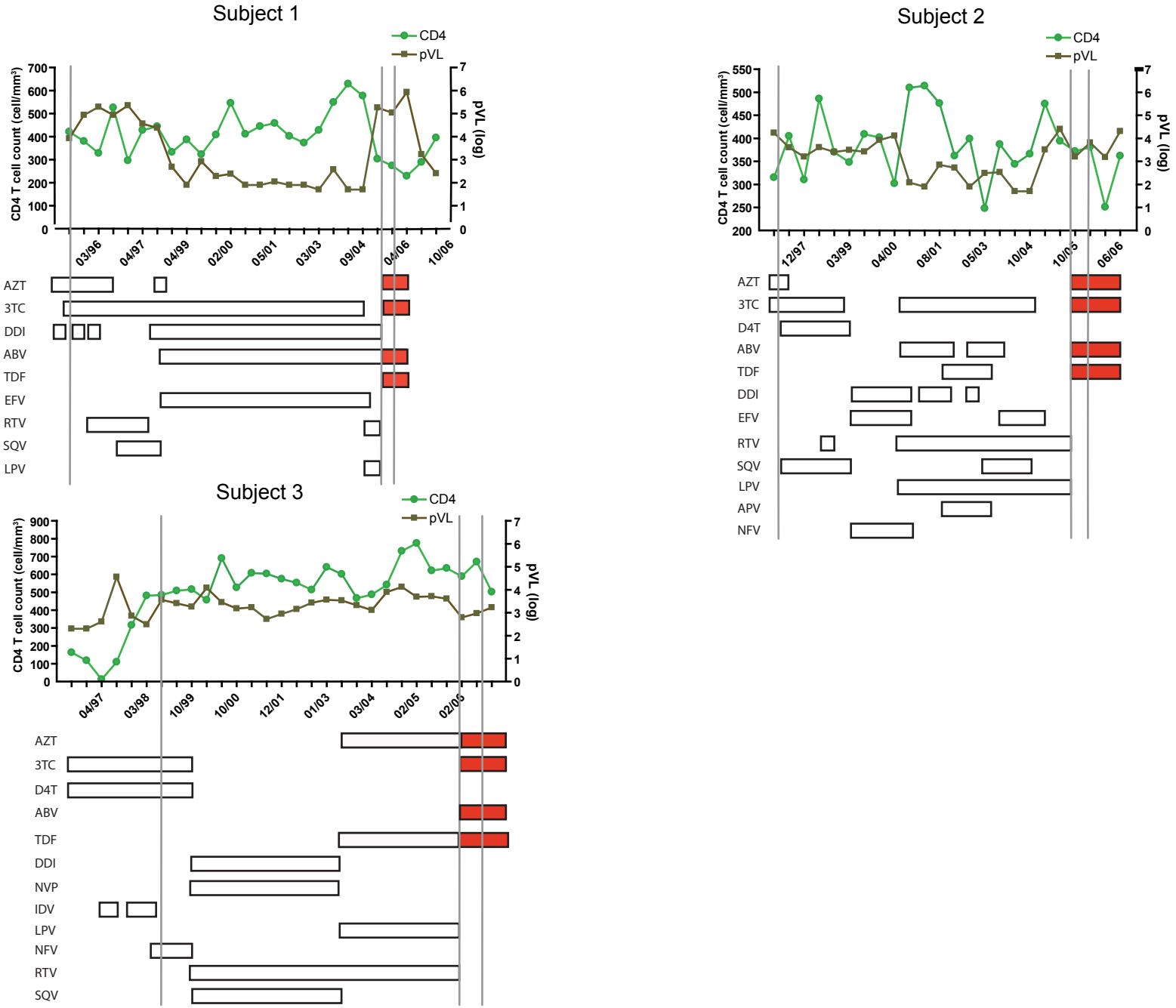
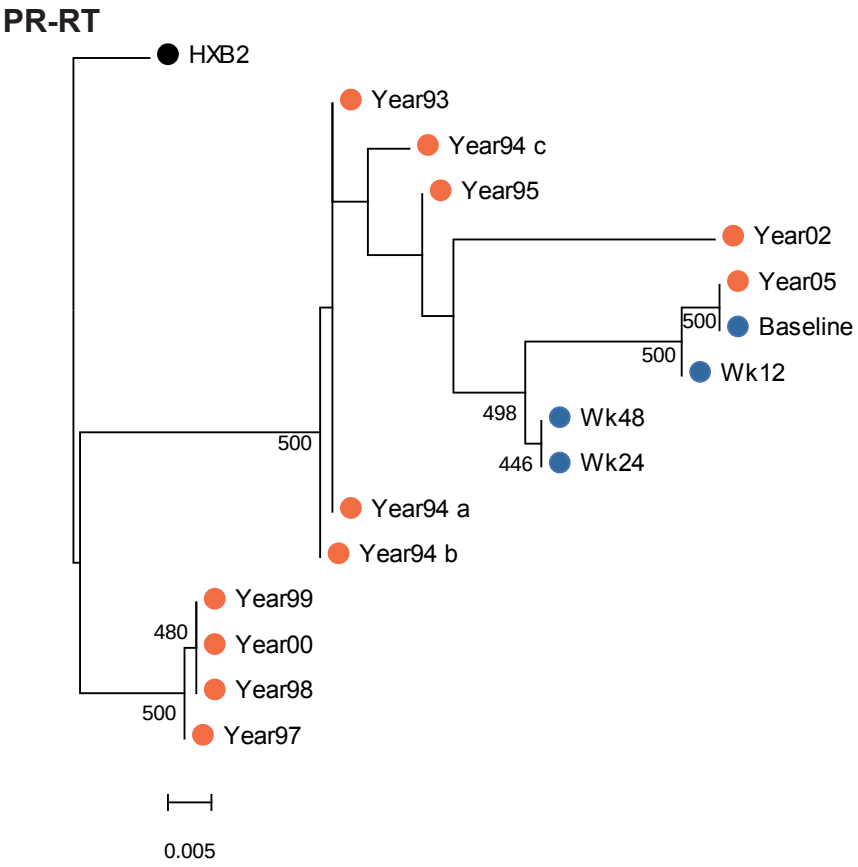
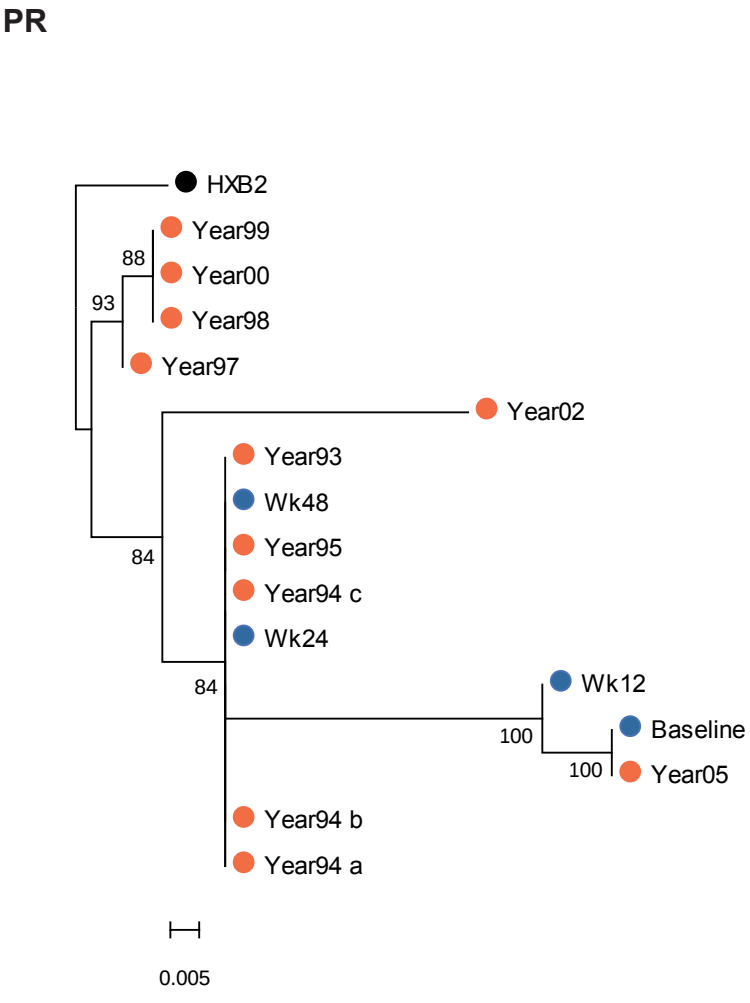


Figure S2. Phylogenetic trees of subject 1, 2 and 3. Phylogenetic analysis was carried out with PHYML v3.0. Four different trees were constructed for PR-RT, PR, RT, and env. Orange labels represent viral variants before the partial treatment interruption and blue labels represent viral variants during the study period. **A.** Subject1, **B.** Subject2. **C.** Subject3.

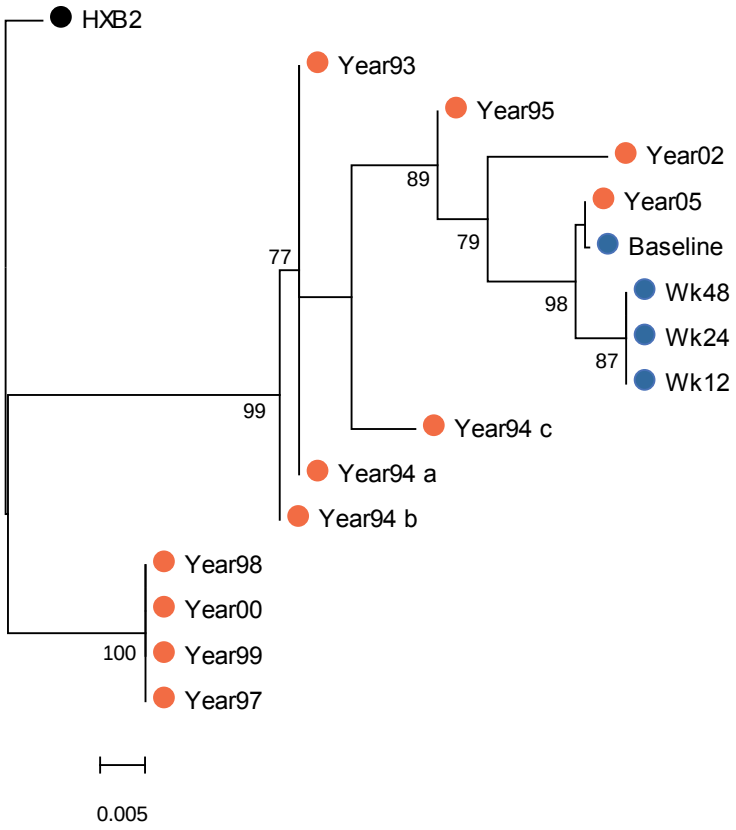
A





A

RT

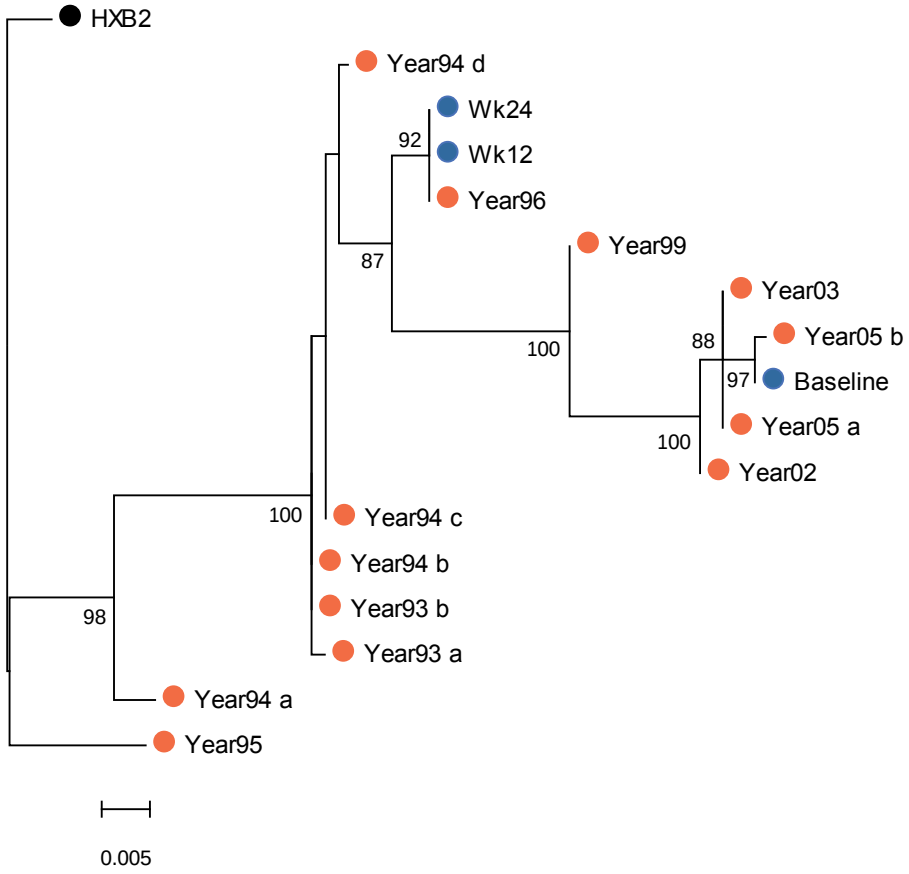

HXB2
Year93
Year95
Year02
Year05
Baseline
Wk48
Wk24
Wk12
Year94 c
Year94 a
Year94 b
Year98
Year00
Year99
Year97
89
79
77
98
87
99
100
0.005

env

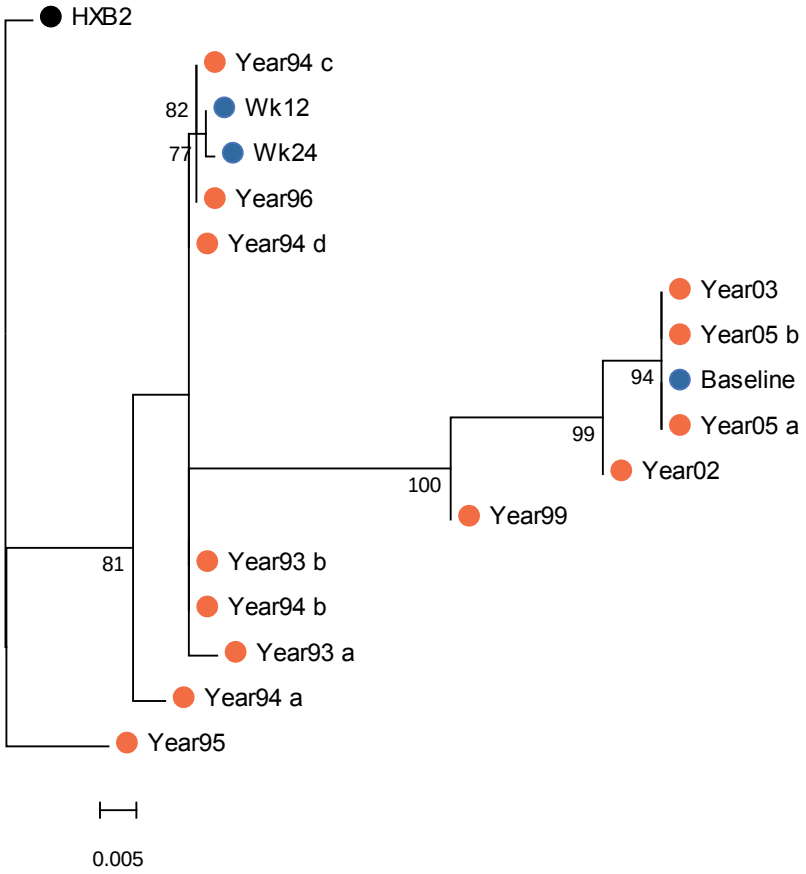

HXB2
Wk48
Wk24
Year05
Baseline
Wk12
Year94
Year95
Year96
77
99
0.02

B

PR-RT


HXB2
Year94 d
Wk24
92
Wk12
Year96
87
Year99
Year03
100
88
Year05 b
97
Baseline
Year05 a
100
Year02
100
Year94 c
Year94 b
98
Year93 b
Year93 a
Year94 a
Year95
0.005

PR

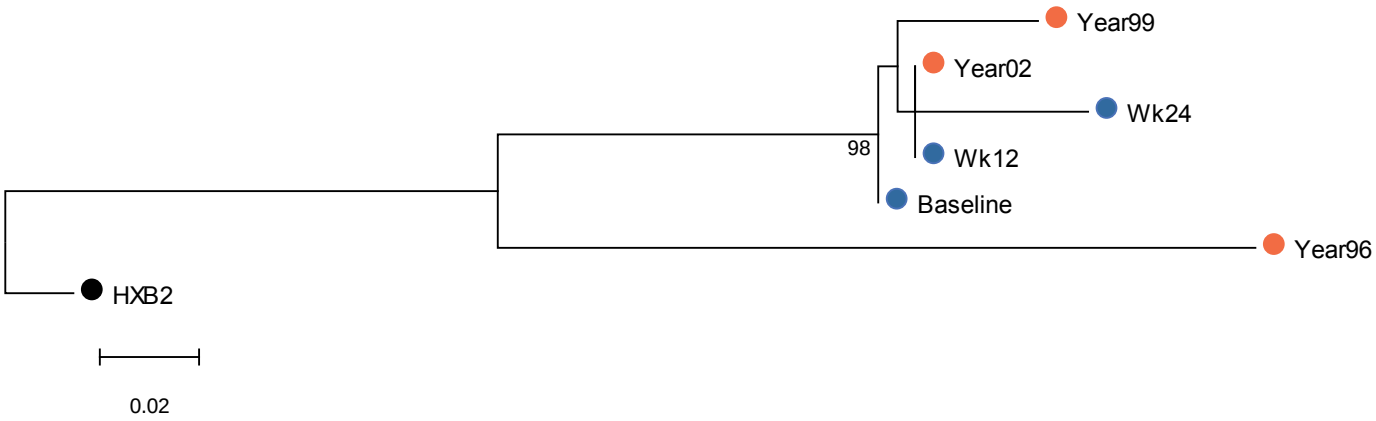

HXB2
Year94 c
82
Wk12
77
Wk24
Year96
Year94 d
Year03
Year05 b
94
Baseline
Year05 a
99
Year02
100
Year99
81
Year93 b
Year94 b
Year93 a
Year94 a
Year95
0.005

B

RT

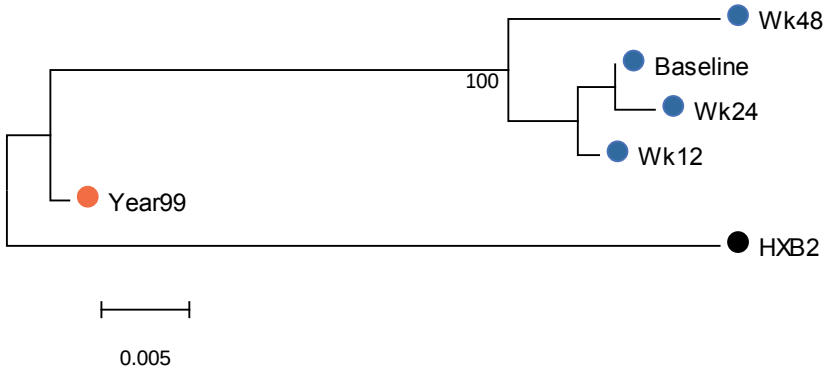

HXB2
Year94 d
Wk24
81
Wk12
Year96
88
Year99
Year02
96
Year05 b
96
Baseline
100
Year03
Year05 a
99
Year94 c
Year94 b
99
Year93 b
Year93 a
Year94 a
Year95
0.005

env

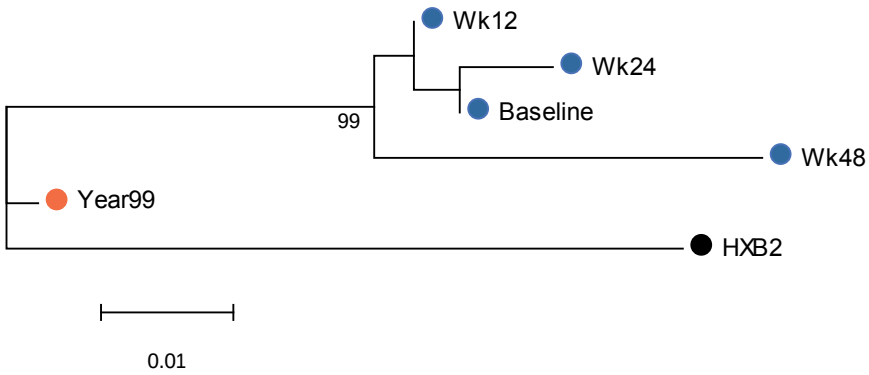

Year99
Year02
Wk24
98
Wk12
Baseline
Year96
HXB2
0.02

C

PR-RT

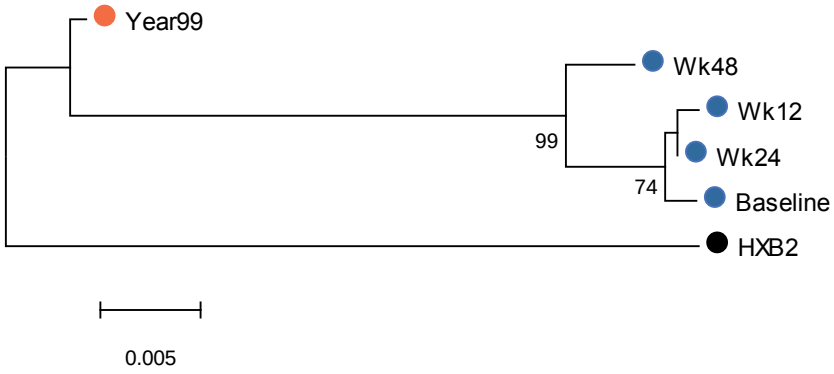

Wk48
Baseline
100
Wk24
Wk12
Year99
HXB2
0.005

PR


Wk12
Wk24
Baseline
99
Wk48
Year99
HXB2
0.01

C

RT


Year99
Wk48
Wk12
99
Wk24
74
Baseline
HXB2
0.005

env

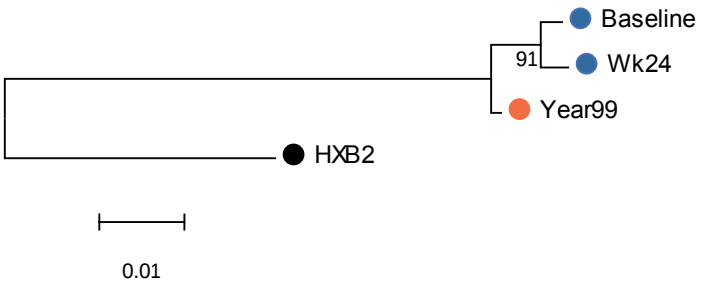

Baseline
91
Wk24
Year99
HXB2
0.01

Table S1. Recombination analysis.

Subject	Program	Breakpoint (nt)	Algorithm	P value
#1 (week 24-48)	RDP3	35-276	MaxChi	3.71×10^{-03}
			Chimera	3.71×10^{-03}
			SiScan	1.25×10^{-06}
			LARD	5.19×10^{-317}
			3Seq	4.02×10^{-04}
	GARD	289		2.00×10^{-04}
#3 (week 48)	RDP3	37-863	MaxChi	3.46×10^{-03}
			SiScan	9.67×10^{-16}
			LARD	3.46×10^{-03}
			3Seq	1.92×10^{-318}
	GARD	829		1.00×10^{-02}

Recombination Parameters

We tested different algorithms implemented in RDP3, [1,2] GENECONV, [3] BOOTSCAN, [4] MAXCHI, [5] CHIMERA, [6] SIS SCAN, [7] LARD, [8] and 3Seq. [9] The parameters were modified from the default for each program (in general for all programs p-value was set as 0.01 and Bonferroni correction was used to avoid significance for multiple testing):

1.- RDP: No Reference sequence was used, Windows Size was set to 20nt and 70-100% sequence identity. The rest by default.

2.- GENECONV. Default parameters.

3.- BOOTSCAN: Window size was set to 100nt and step size into 20nt, we performed 1,000 bootstrap replicates and the cut off percentage was set to 70. The binomial p-value was calculated and for the bootstrap Jukes and Cantor model from 1969 was used. All the rest by default.

4.- MAXCHI: variable sites per window = 20. We strip gaps. All the rest by default.

5.- CHIMERA: variable sites per window = 20. All the rest by default.

6.- SIS SCAN: window size was 100nt, step size 20nt, strip gaps, we used all codon positions, we used the nearest outlier for testing. We performed a slow scan. All the rest by default.

7.- PhyPro: window size was 100nt. We ignored gaps and we do not allow for self comparisons. All the rest by default.

8.- VisRD window size was 100nt. All the rest by default.

9.- LARD Parameters by default with the model HKY85.

10.- DSS: We use a window size of 100nt, a step size of 20nt and a smoothing window of 10nt. We constructed a Neighbour Joining (NJ) tree. The power was set to 2, the simulated data set was 1,000 and the p value was set to 0.01. All the rest by default.

To confirm the results from RDP3 package, we used an specific HIV program to detect recombination. We used GARD [10] program from the datamonkey.org [11] in order to test the detected breakpoint with RDP3 package. We used the HKY85 [12] as the evolutionary model that better explain our sequence alignment evolution and a discrete gamma distribution with 4 categories to take into account the distribution of mutations across the gene.

We test for positive selection using the programs SLAC and FEL [13] from datamonkey.org in order to detect sites under positive or negative selection.

For those subjects where recombination point were detected with RDP3 and it was confirmed with GARD, we used PARRIS program [14] implemented in the datamonkey.org to test the selective constraints acting over recombinant sequences.

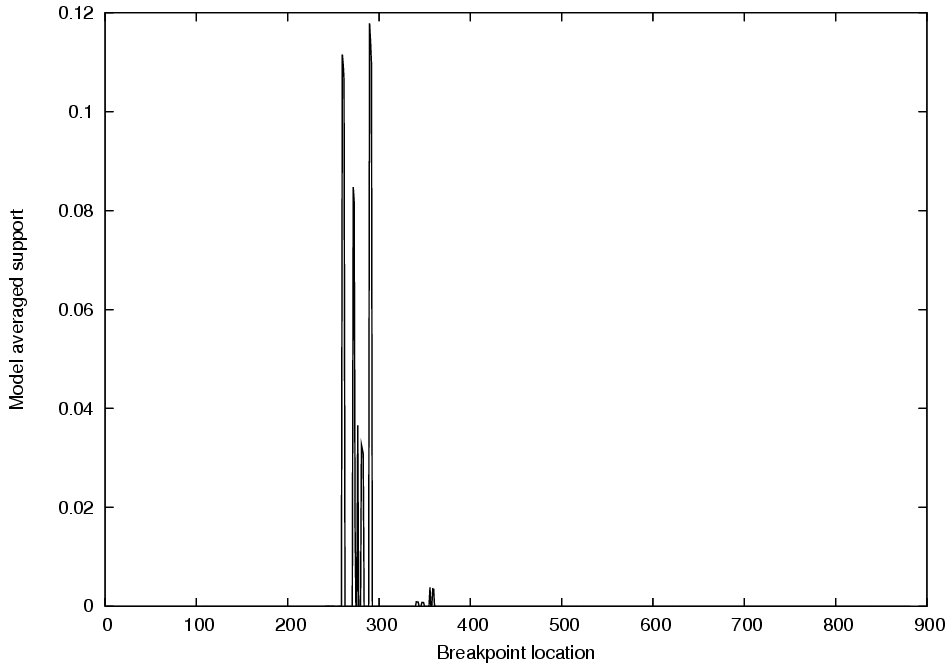
Subject 1

Polymerase coding region. GARD program detected one recombination breakpoint (nt 289) (supplementary figure 1) that is close to the recombination breakpoint detected in with RDP3 (nt 276). No evidence of positive selection was detected using SLAC ($\omega = 0.42$) and FEL. PARRIS did not detect any signal of positive selective pressure (p_value = 0.89), although FEL reported some sites to be under negative selection (codon 15, 64, 92, 99, 110, 140, 170, 178, 215, 221 and 233).

Envelope coding region. No recombination breakpoints were detected for the envelope gene. No evidence for positive or negative selection was found in the envelope gene using

SLAC (ω = 0.98). FEL detected no evidence of positive selection and just two codons were detected under negative selection (codon 53 and 105). We did not perform the PARRIS analysis since there was no evidence of recombination in this data.

Fig. S3. Recombination breakpoint of subject 1.



Subject 2

Polymerase coding region. RDP3 and GARD did not found any recombination events.

No evidence of positive selection was detected using SLAC (ω = 0.24) but two codons where detected to be under negative selection (codons 146 and 206). FEL detected 12 positions to be under negative selection (codons 14, 91, 102, 113, 137, 173, 181, 187, 206, 243, 254, 258). PARRIS was not performed since there were not detected recombination event detected in this subject.

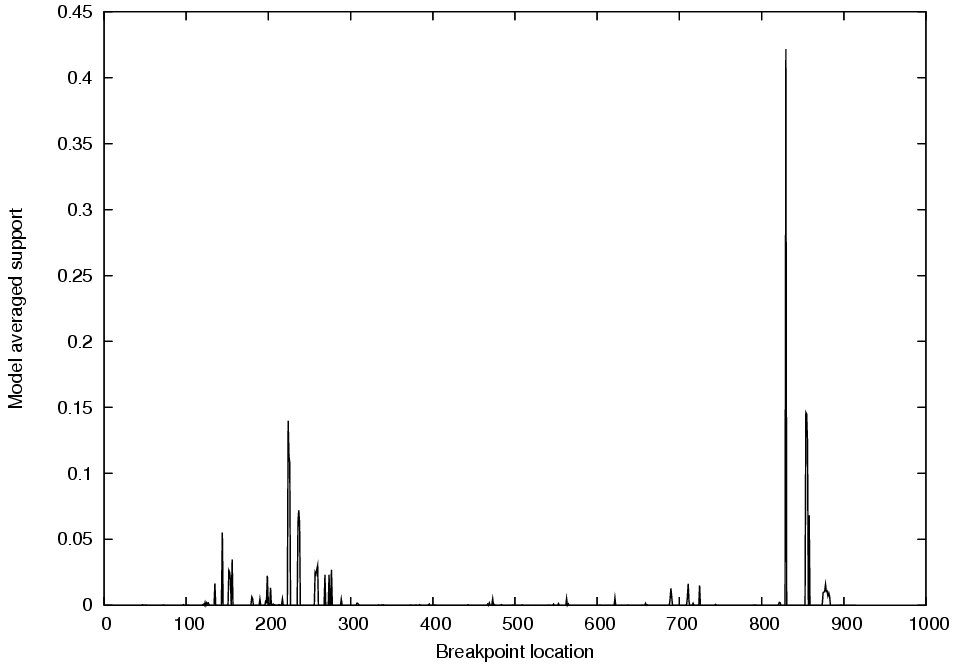
Envelope coding region. RPD3 and GARD did not found any recombination events. No evidence of positive selection was detected using SLAC (ω = 0.79), FEL detected 3 positions to be under negative selection (codon 33, 41 and 42). PARRIS was not performed since there has not been any recombination event detected in this subject.

Subject 3

Polymerase coding region. We detected one recombination breakpoint (nt 829) (supplementary figure 2) that is close to the recombination breakpoint detected in with RDP3 (nt 863). No evidence of positive selection was detected using SLAC. FEL detected one codon to be under positive selection (codon 7) and 9 under negative selection (codons 66, 72, 92, 96, 182, 218, 264, 274, 299). PARRIS did not detect any signal of positive selection (p_value = 0.77).

Envelope coding region. We were not able to detect any significant recombination breakpoint. No evidence for positive or negative selection was found using SLAC (ω = 0.79). FEL detected no evidence of positive selection and nine codons were under negative selection (codons 98, 176, 201, 216, 363, 372, 378, 487 and 626). We did not perform the PARRIS analysis since there was no evidence of recombination in this data.

Fig S4. Recombination breakpoint of subject 3.



The effect of negative selection detected by the different programs is an expected event, since it have been show that polymerase gene is under negative selection. This might be a clear indication that viruses from the different subjects tend to recombine in order to escape

from the antiretroviral pressure. For envelope, the omega values detected indicate that they are more close to neutrality than to negative selection.

REFERENCES

1. Martin D, Rybicki E. RDP: detection of recombination amongst aligned sequences. *Bioinformatics.* 2000;16:562-563.
2. Martin DP, Williamson C, Posada D. RDP2: recombination detection and analysis from sequence alignments. *Bioinformatics.* 2005;21:260-262.
3. Padidam M, Sawyer S, Fauquet CM. Possible emergence of new geminiviruses by frequent recombination. *Virology.* 1999;265:218-225.
4. Martin DP, Posada D, Crandall KA, Williamson C. A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. *AIDS Res Hum Retroviruses.* 2005;21:98-102.
5. Smith JM. Analyzing the mosaic structure of genes. *J Mol Evol.* 1992;34:126-129.
6. Posada D, Crandall KA. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proc Natl Acad Sci U S A.* 2001;98:13757-13762.
7. Gibbs MJ, Armstrong JS, Gibbs AJ. Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics.* 2000;16:573-582.
8. Holmes EC, Worobey M, Rambaut A. Phylogenetic evidence for recombination in dengue virus. *Mol Biol Evol.* 1999;16:405-409.
9. Boni MF, Posada D, Feldman MW. An exact nonparametric method for inferring mosaic structure in sequence triplets. *Genetics.* 2007;176:1035-1047.
10. Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD. Automated phylogenetic detection of recombination using a genetic algorithm. *Mol Biol Evol.* 2006;23:1891-1901.
11. Pond SL, Frost SD. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics.* 2005;21:2531-2533.
12. Hasegawa M, Kishino H, Yano T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol.* 1985;22:160-174.

13. Kosakovsky Pond SL, Frost SD. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol.* 2005;22:1208-1222.
14. Scheffler K, Martin DP, Seoighe C. Robust inference of positive selection from recombining coding sequences. *Bioinformatics.* 2006;22:2493-2499.