

**Table S1: Bayesian within-host phylogenetic inference details**

Participant	MCMC generations <sup>a</sup>	Phylogenies post burn-in	QC-passed phylogenies	$\Delta$ AIC <sup>b</sup> (median, IQR)	Phylogenetically estimated root date (95% HPD <sup>c</sup> )	Within-host HIV evolutionary rate <sup>d</sup> (mean, 95%HPD)
<i>gp120</i>						
1	100	15000	7218	488 (452 - 522)	Dec 1995 (Aug 1995 - Mar 1996)	$4.8 \times 10^{-5}$ ( $3.8 \times 10^{-5}$ - $6.0 \times 10^{-5}$ )
2	35	5250	3842	588 (550 - 627)	May 2003 (Jan - Sep 2003)	$5.9 \times 10^{-5}$ ( $4.2 \times 10^{-5}$ - $7.7 \times 10^{-5}$ )
3	30	4500	4500	196 (181 - 213)	Nov 2001 (Jul 2001 - Mar 2002)	$5.9 \times 10^{-5}$ ( $4.2 \times 10^{-5}$ - $7.6 \times 10^{-5}$ )
4	40	6000	6000	550 (525 - 574)	Feb 1995 (Oct 1994 - Jun 1995)	$6.7 \times 10^{-5}$ ( $5.5 \times 10^{-5}$ - $8.1 \times 10^{-5}$ )
5	10	1500	1500	40 (34 - 48)	Dec 2008 (Oct 2008 - Feb 2009)	$9.4 \times 10^{-5}$ ( $5.3 \times 10^{-5}$ - $1.4 \times 10^{-4}$ )
6	20	3000	2528	51 (45 - 60)	Jan 2007 (Sep 2006 - May 2007)	$8.1 \times 10^{-5}$ ( $6.1 \times 10^{-5}$ - $1 \times 10^{-4}$ )
7	10	1500	1278	300 (273 - 328)	May 1999 (Jul 1998 - Dec 1999)	$2.3 \times 10^{-5}$ ( $1.7 \times 10^{-5}$ - $2.9 \times 10^{-5}$ )
<i>gag</i>						
1	15	3000	2992	286 (254 - 317)	Feb 1995 (May - Nov 1995)	$2.6 \times 10^{-5}$ ( $1.9 \times 10^{-5}$ - $3.2 \times 10^{-5}$ )
3	10	1500	1500	261 (239 - 283)	Nov 2001 (May 2001 - May 2002)	$2.6 \times 10^{-5}$ ( $2 \times 10^{-5}$ - $3.5 \times 10^{-5}$ )
7	10	1500	1500	230 (209 - 254)	Feb 1998 (Dec 1996 - Apr 1999)	$1.3 \times 10^{-5}$ ( $9.0 \times 10^{-5}$ - $1.6 \times 10^{-5}$ )

<sup>a</sup> jModelTest returned GTR+I+G (General Time Reversible model with Invariable site plus Gamma distribution with four rate categories) as the best-fit nucleotide substitution model for all participants' *gp120* and *gag* nucleotide sequence alignments.

<sup>b</sup>  $\Delta$ AIC, delta Akaike information criterion. A phylogeny required a  $\Delta$ AIC of  $\geq 10$  (where its root date was prior to the first plasma sampling) was considered evidence of a molecular clock signal (see methods).

<sup>c</sup> 95% HPD, 95% highest posterior density.

<sup>d</sup> Expressed in estimated substitutions per nucleotide site per day.