Root-to-tip				LSD	
Clade	Rate $\times 10^{-3}$	tMRCA	R^2	Rate $\times 10^{-3}$	tMRCA
AE	2.49 (2.31,2.67)	1971.4	0.51	1.87 (1.85, 2.10)	1968.6 (1965.5, 1974.6)
AG	2.21 (1.75,2.67)	1957.4	0.35	2.27 (2.10, 2.68)	1961.9 (1957.5, 1969.0)
A1	2.69 (2.17,3.21)	1932.0	0.26	2.45 (2.32, 2.66)	1966.3 (1964.0, 1969.5)
B	2.43 (2.32,2.54)	1941.2	0.34	1.47 (1.46, 1.57)	1951.7 (1951.2, 1954.8)
C	1.99 (1.75,2.23)	1926.9	0.13	1.80 (1.78, 1.96)	1939.8 (1937.5, 1946.7)
D	1.90 (1.47,2.32)	1944.5	0.33	1.88 (1.68, 2.11)	1957.8 (1952.7, 1962.9)
F1	2.33 (1.85,2.81)	1970.4	0.57	1.67 (1.34, 2.03)	1956.2 (1943.9, 1965.2)

Table 1: Summary of the evolutionary rate estimates, times to most recent common ancestor (tMRCAs), and R^2 values generated by applying root-to-tip and least-squares dating models to our seven clade-specific trees. The 95% confidence intervals for the evolutionary rates of both models and for the tMRCAs estimates of the LSD model are enclosed in brackets. Both models are shown to illustrate the differences between fitting strict (root-to-tip) and relaxed clock models to our sequence data.

indel rates for each variable loop using a binomial-Poisson model, where the probability of detecting an indel event in a cherry increased exponentially with the divergence time. The indel rate estimates across the five variable loops and seven HIV-1 clades in this study ranged between 3.0×10^{-5} to 1.5×10^{-3} indels/nt/year (Figure 2). We could not obtain an indel rate estimate for V3 in F1 due due to low sample size for this sub-subtype, such that no cherries had discordant sequence lengths in V3. Similarly, we observed wide confidence intervals for the rate estimates for indels within V1 in AG and F1, and for V5 in F1. The frequency of indels was significantly lower in subtype B than the other clades in our data (binomial GLM, $p < 2 \times 10^{-16}$; Supplementary Table S1). In addition, indels were significantly less frequent in V3 irrespective of clade. Estimated interaction effects in the model also indicated that indels were significantly less frequent than expected in V2 within clades B and C.

Under the assumption that differences in sequence lengths of variable loops was caused by a single fixed indel (*i.e.*, no multiple hits), we examined the distribution of indel lengths among variable loops and clades. Cherries with putative indels in the HIV-1 subtype C phylogeny tended to contain significantly longer indels than expected (Figure 3). Conversely, the variable loops V1, V2 and V4 tended to contain longer indels than expected irrespective of clade, whereas V3 and V5 tended to contain shorter indels.

Next, we examined the frequencies of nucleotides in indel- and non-indel regions of sequences in cherries with putative indels (Figure 4). Because these frequencies measured for different clades tended to cluster by variable loop, we treated the clades as rudimentary replicates for this com-