SUPPLEMENTAL MATERIAL

Keele et al., http://www.jem.org/cgi/content/full/jem.20082831/DC1

Bayesian and Poisson models The Poisson model evolves a viral population forward in time, assuming no selection and using previously measured parameters for HIV viral replication dynamics. In contrast, the Bayesian model uses the entire sequence dataset for an analysis of the coalescent, starting with only the previously measured base substitution parameters as prior expectation. The use of summary statistics is expected to make the Poisson method more robust against effects like undetected recombination, whereas the Bayesian method incorporates the modeling uncertainties more completely in its estimates, but at the expense of statistical power. The Bayesian approach utilizes programs from the BEAST package that rely on a relaxed molecular clock and a distribution of evolutionary rates based on the site replacement rate value used for the Poisson model (mean = 1.44×10^{-5}/site/day; SD = 10^{-6}) determined from as many as 360,000,000 phylogenies that are sampled using a Monte Carlo-Markov Chain approach. Bayesian skyline models were used that sample over these phylogenies and do not require prior specification of a demographic model or effective population size. Four independent chains of 30–360 million iterations with 10% burn-in were analyzed, with each chain showing results consistent with the others. Results from each chain were combined using LogCombiner and the combined set of trees was analyzed with TreeAnnotator. Times to MRCA were derived and visualized using Tracer version 1.4.1. Because BEAST and Poisson methods both assume that mutations occur randomly across the genomes, and APOBEC-mediated G-to-A mutations violate this condition, we performed BEAST and Poisson calculations in two iterations. The first iteration included *env* sequences (n = 461) representing the major virus lineage from each animal stripped of all APOBEC signature positions (i.e., G within the context of GRD). The second iteration did not eliminate all APOBEC signature positions but did exclude 17 overtly hypermutated sequences. Results of the first analysis are displayed in Table II, although findings from the second analysis were not substantially different. 12 of the animals were sampled at two time points 1 wk apart when the viral load was increasing, and six of these were homogeneous infections where we could test sequences sequentially using the Poisson model. The model predicts that the Poisson parameter lambda describing sequence diversity should increase linearly with time with a coefficient of 2.38×10^{-5} per base position per day. After averaging of 6 animals, the Poisson parameter was seen to be changing at $1.5(\pm 0.5) \times 10^{-5}$/base/day, which is lower than but consistent with model predictions. Neither model adjusts for "purifying selection" (i.e., mutations that are selected against and go unobserved because they result in less fit viruses), and as a consequence, timing estimates tend to be biased toward a low estimate. Estimates of the time to a MRCA for sequences from each animal based on Poisson and Bayesian models were generally similar with confidence limits overlapping in every case but one.

The Journal of Experimental Medicine

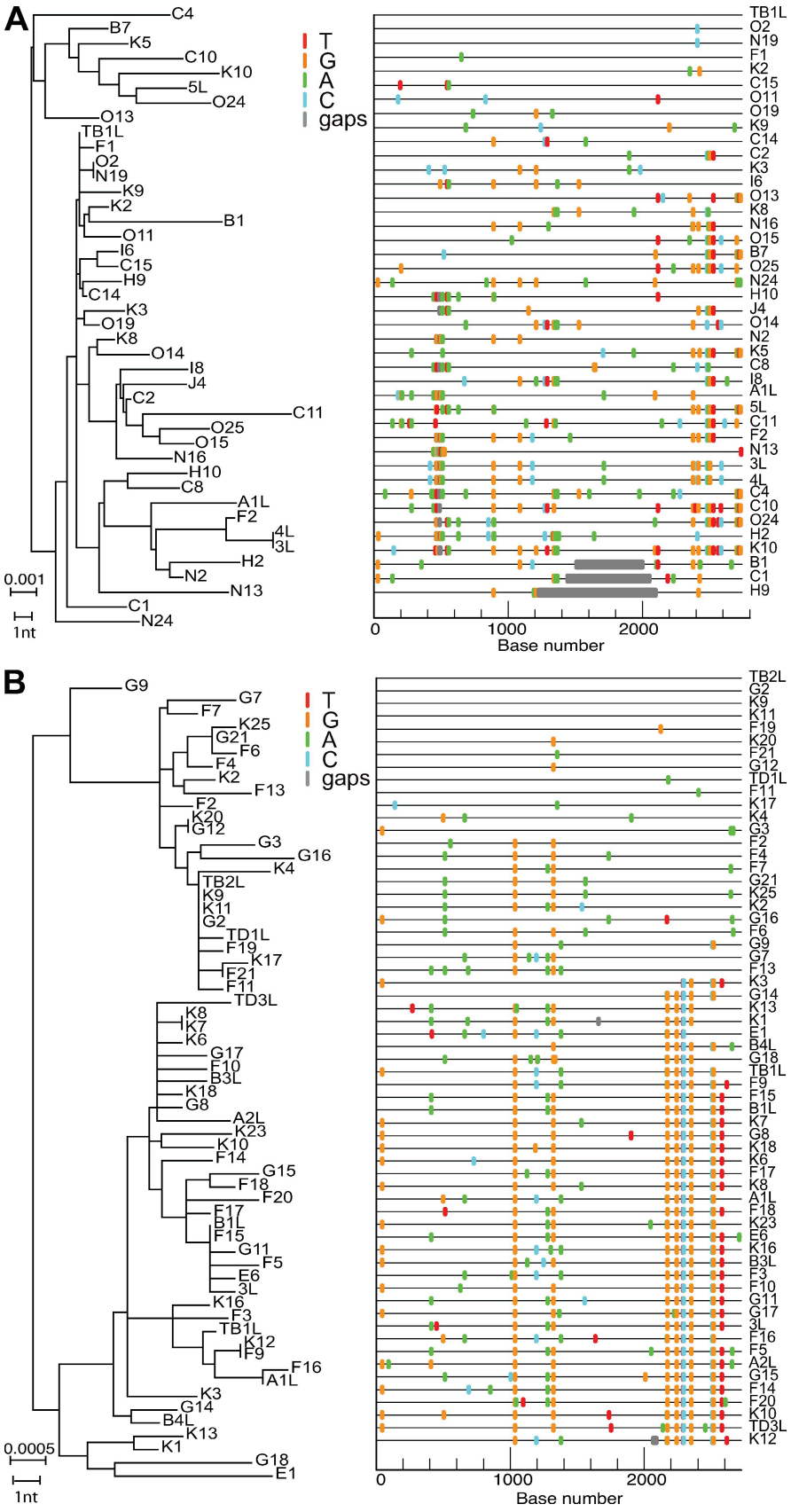

Figure S1. NJ and *Highlighter* plots of inoculum sequences. (A) SIVsmE660. (B) SIVmac251. nt, nucleotide.

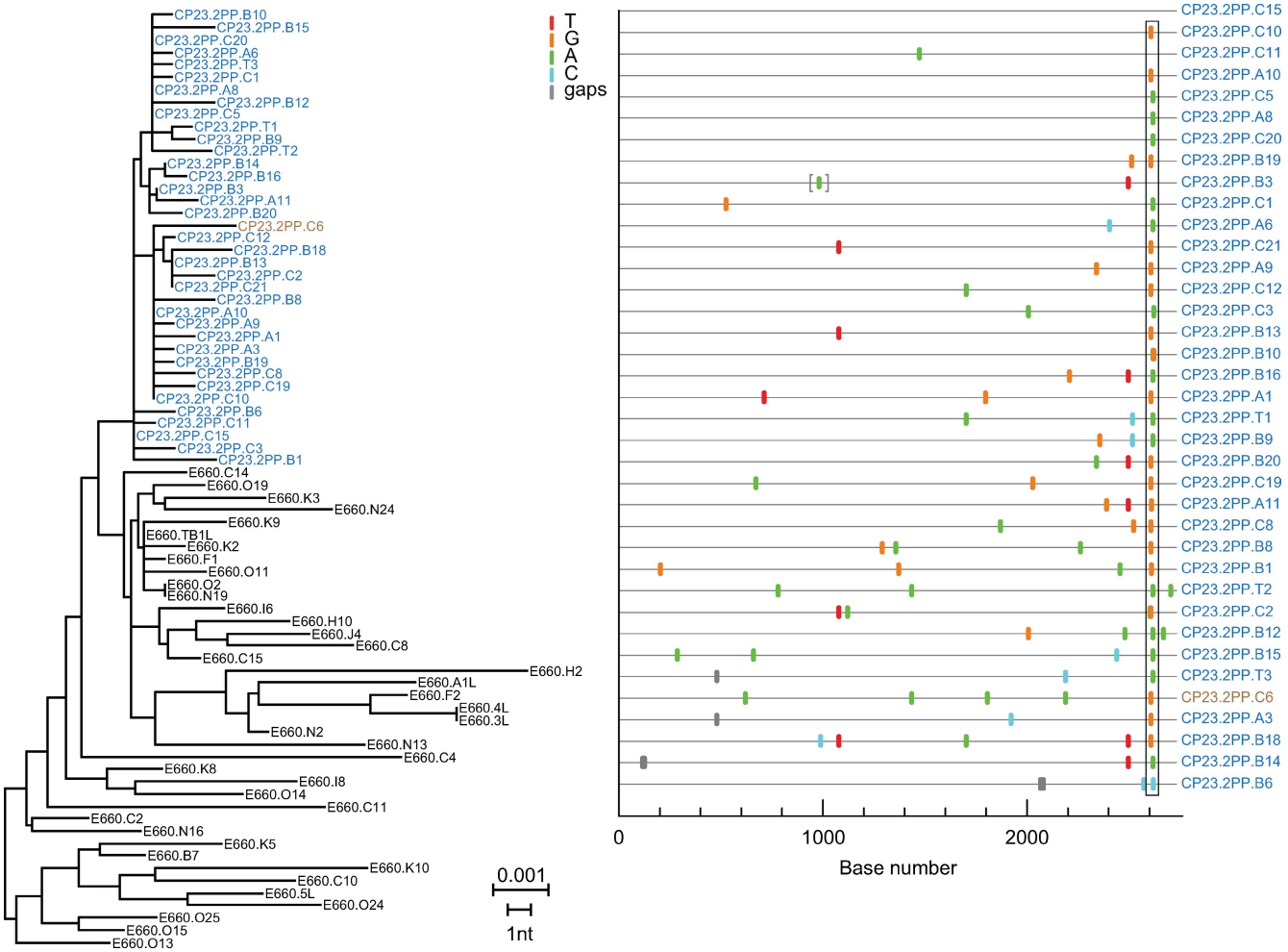

Figure S2. Viral sequence diversity 2 wk after peak viremia in i.r. infected animal CP23. NJ and *Highlighter* plots show concentrated mutations (boxed) in the *env/nef* overlap region of the sequences. Bracket indicates a single-nucleotide insertion. Brown label indicates sequence with three or more APOBEC-mediated G-to-A mutations compared with consensus.

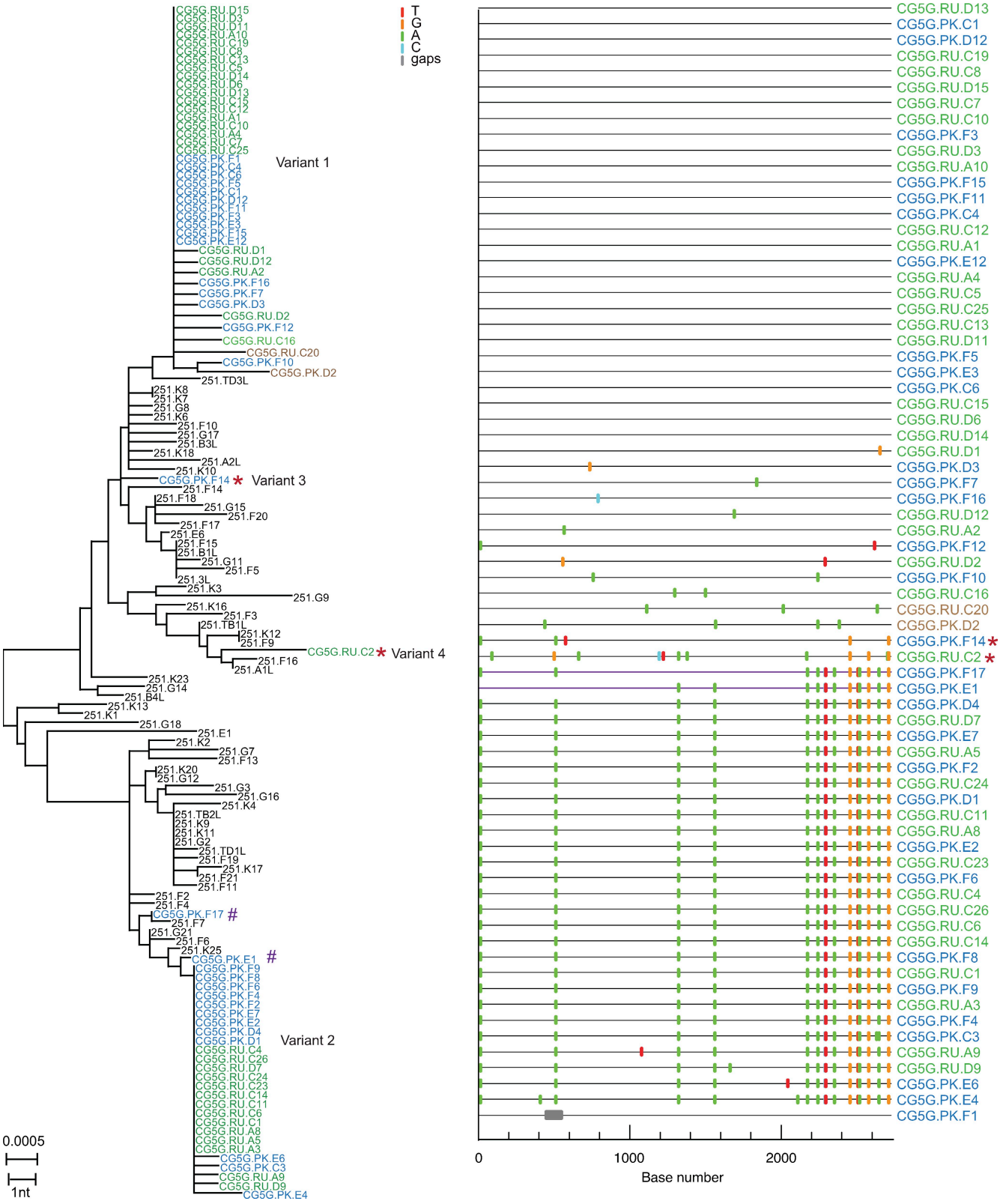

Figure S3. i.v. transmission of four viruses, followed by recombination. NJ and *Highlighter* plots from animal CG5G at ramp-up (green symbols) or peak (blue symbols) viremia, with recombinant sequences indicated by a purple symbol (#) and minor transmitted variants by a red asterisk. Brown labels indicate sequences with three or more APOBEC-mediated G-to-A mutations compared with consensus.

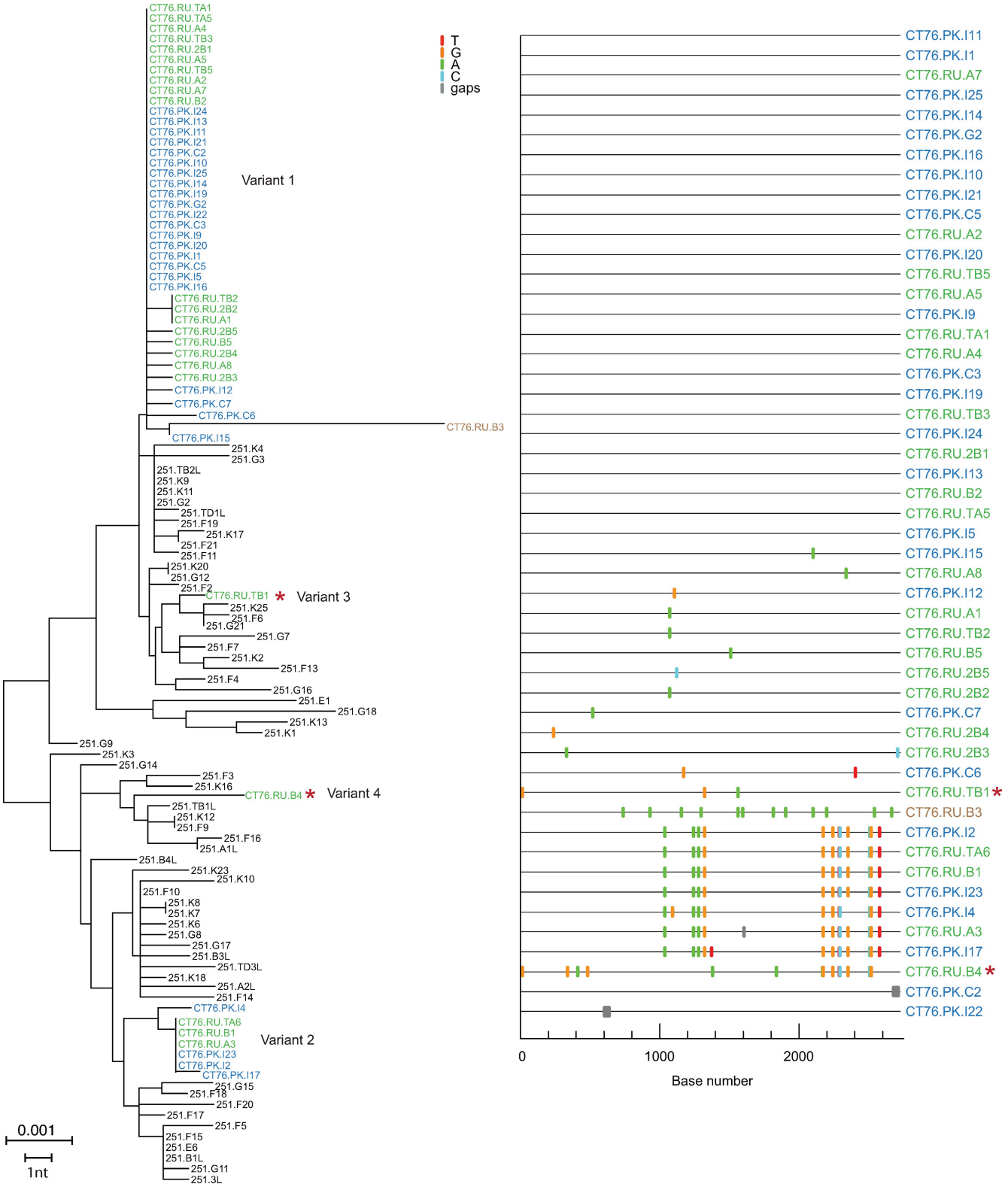

Figure S4. i.r. transmission of four viruses. NJ and *Highlighter* plots from animal CT76 at ramp-up (green symbols) or peak (blue symbols) viremia with an extensively APOBEC-mutated sequence (RU.B3) indicated by the brown label. Red asterisks indicate minor transmitted variants found in ramp-up sequences only.

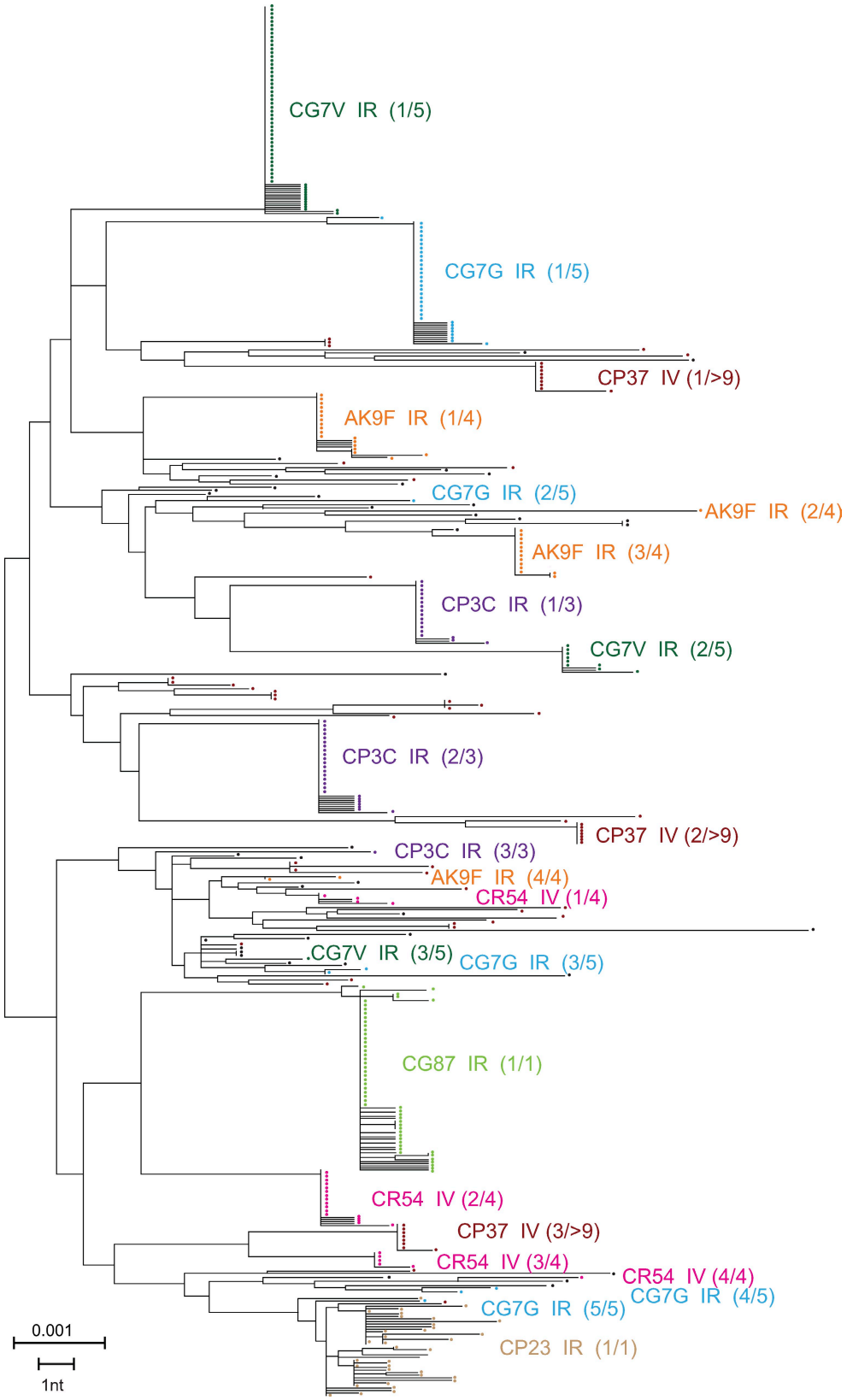

Figure S5. Composite SIVsmE660 tree from i.r. and i.v. infected animals. The NJ tree depicts numerous transmitted/founder virus lineages distributed throughout a diverse set of inoculum sequences (black symbols). Most i.r. or i.v. transmitted viruses are represented by discrete, low-diversity lineages. Animal designation, route of infection, and numbers of transmitted viruses are indicated.