Streeck et al., http://www.jem.org/cgi/content/full/jem.20080281/DC1

SUPPLEMENTAL MATERIALS AND METHODS

Analysis of recombination using clonal gag sequences.

To validate the recombination in gag clonal sequences, we adopted a modified version of the Genetic Algorithm for Recombination Detection (1). A reference alignment of putatively nonrecombinant sequences and an associated phylogenetic tree is proposed. Putatively recombinant query sequences are tested for recombination by determining whether, when this sequence is added to the reference dataset, different segments of the query sequence cluster with different branches in the reference phylogeny. For a fixed number of breakpoints, a genetic algorithm is used to search for the best fitting location of the breakpoints in the sequence and the best placement of the segments in the phylogenetic tree. The number of breakpoints is estimated by determining the number associated with the lowest Bayesian Information Criterion. We assumed a general time-reversible (GTR) model of nucleotide substitution, and chose a reference set comprised of the 4 bulk gag sequences from time points 1–4 (T1–T4), and 16 clonal sequences from time point 4. 39 sequences from time points 5–9 were used as query sequences. Fig. S2 illustrates the 10 recombinants detected using this approach, with blue indicating regions that cluster with the superinfecting strain, and red indicating regions that cluster with the original strain. The pattern of recombination is highly similar among the 10 recombinants, suggesting that the replacement of one segment by another was a single evolutionary event, followed by outgrowth of this variant. Note that we also failed to detect these recombinant variants in time points 8 (8 sequences) and 9 (4 sequences), which is consistent with replacement of the original segment by one from the superinfecting strain.

REFERENCE

1. Kosakovsky Pond, S.L., D. Posada, M.B. Gravenor, C.H. Woelk, and S.D. Frost. 2006. Automated phylogenetic detection of recombination using a genetic algorithm. *Mol. Biol. Evol.* 23:1891–1901. PubMed doi:10.1093/molbev/msl051