

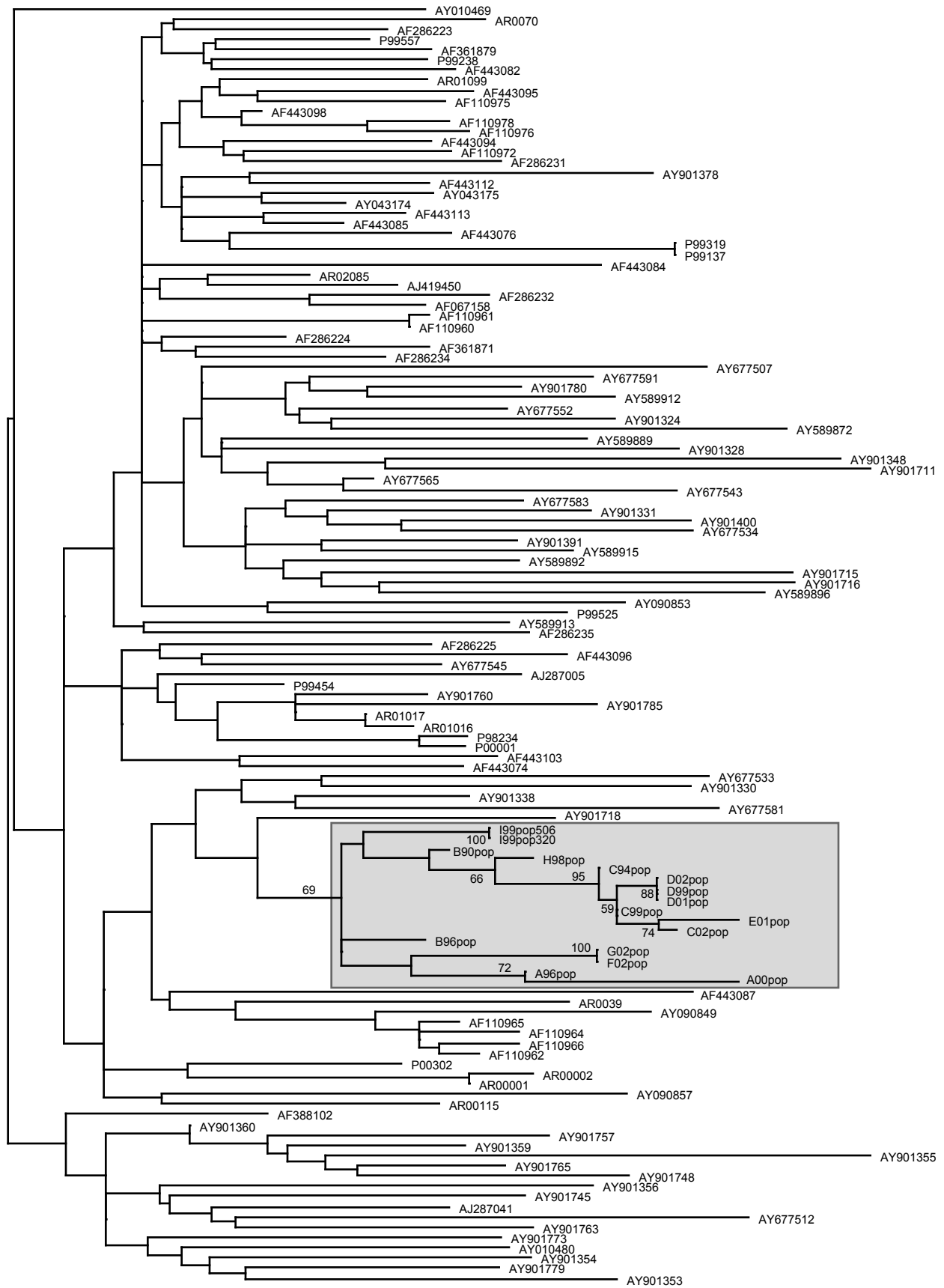
## SUPPLEMENTARY MATERIAL

**Figure 1.** Deduced amino acid alignment of the *pol* sequences obtained from the known transmission chain. Numbering according to HXB2 protease (PRO) and reverse transcriptase (RT) is shown above the alignment. The consensus sequence shows the most common amino acid in these sequences. Synonymous polymorphisms are indicated in blue. Allelic mixtures of synonymous and non-synonymous mutations are indicated with green and red respectively. The positions associated with drug resistance, removed to obtain a fully compatible tree, are indicated with a “\*” below the alignment.

**Figure 2.** Phylogenetic tree reconstructed based on a 591 basepair region in the RT gene. In addition to the sequences retrieved from Genbank and from a local database, this data set also includes additional strains obtained from the HIV Drug Resistance Database (<http://hivdb.stanford.edu/>). The sequences were retrieved based on the subtype C criterion and the presence of the T215Y mutation. This resulted in 52 additional sequences having drug resistance mutations on the following positions with decreasing frequency: 215 (100%), 41 (69%), 67 (48%), 184 (44%), 210 (31%), 118 (23%), 70 (19%), 219 (15%), 44 (15%), 69 (10%), 74 (6%), 62 (2%), 65(2%), 116(2%), 151(2%). Tree inference and bootstrapping was performed using PhyML (1). The transmission cluster is indicated with a shaded box in the tree. Bootstrap values (> 50%) are shown for the known transmission history.

1. **Guindon, S., and O. Gascuel.** 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**:696-704.





0.01