

Table S2: Accession numbers of sequences making up the positive training set for sub-type H and J when the *gag-pol* region is used for classification. Since there are only 4 sequences of the *gag-pol* region for H and 3 for J, 3 of the 4 sequences of H and 2 of the 3 sequences of J were used to build the respective pHMMs which were then successful in classifying the remaining sequences.

H	J
AF005496	AF082394
AF190127	GU237072
AF190128	