**Summary of Sequence Data Review**

**Sept 6, 2015**

**Drug Resistance Mutations – sequences\_drms\_6sep2016.txt**

All 4727 sequences were sent to Stanford’s HIVdb program via Sierra for drug resistance reports (includes 6 sequences that are Integrase only and 4721 sequences PRO/RT only)

From the Stanford reports (XML) files all major mutations were extracted using following lists <http://hivdb.stanford.edu/pages/download/resistanceMutations_handout.pdf>

Initial data file had 4727 sequences from 2915 individual CFAR\_PID . Only year and age at sample given. These three variables together do not uniquely identify sequences. There do not seem to be exact duplicates. Sequence identifiers created to uniquely identify sequences SEQ1 – SEQ4727

Fasta files

he 4721 PRO/RT sequences were aligned as these will be used for sequence/phylogenetic analysis