**Additional file 1**

A: Phylogenetic tree for baseline sequences from the Tenofovir and Placebo groups.

B: Comparison of intra-group genetic diversity of gag sequences from the tenofovir and placebo arm. The genetic distance of each sequence from the reference was measured. The differences in medians between placebo and tenofovir were then compared to 1000 median differences obtained by chance from randomised arm samples, and obtained a probability of 0.02 that randomised samples will have a difference in medians as great at the difference in actual observed medians.

C: Twelve months post-infection transmission index. Comparison between gag sequences from the tenofovir (TFV) and those from the placebo recipients who had sequence data at baseline. This comprised of 58 participants (22 TFV and 36 Placebo recipients) with sequence data available at both baseline (enrolment) and 12 months post-infection.

