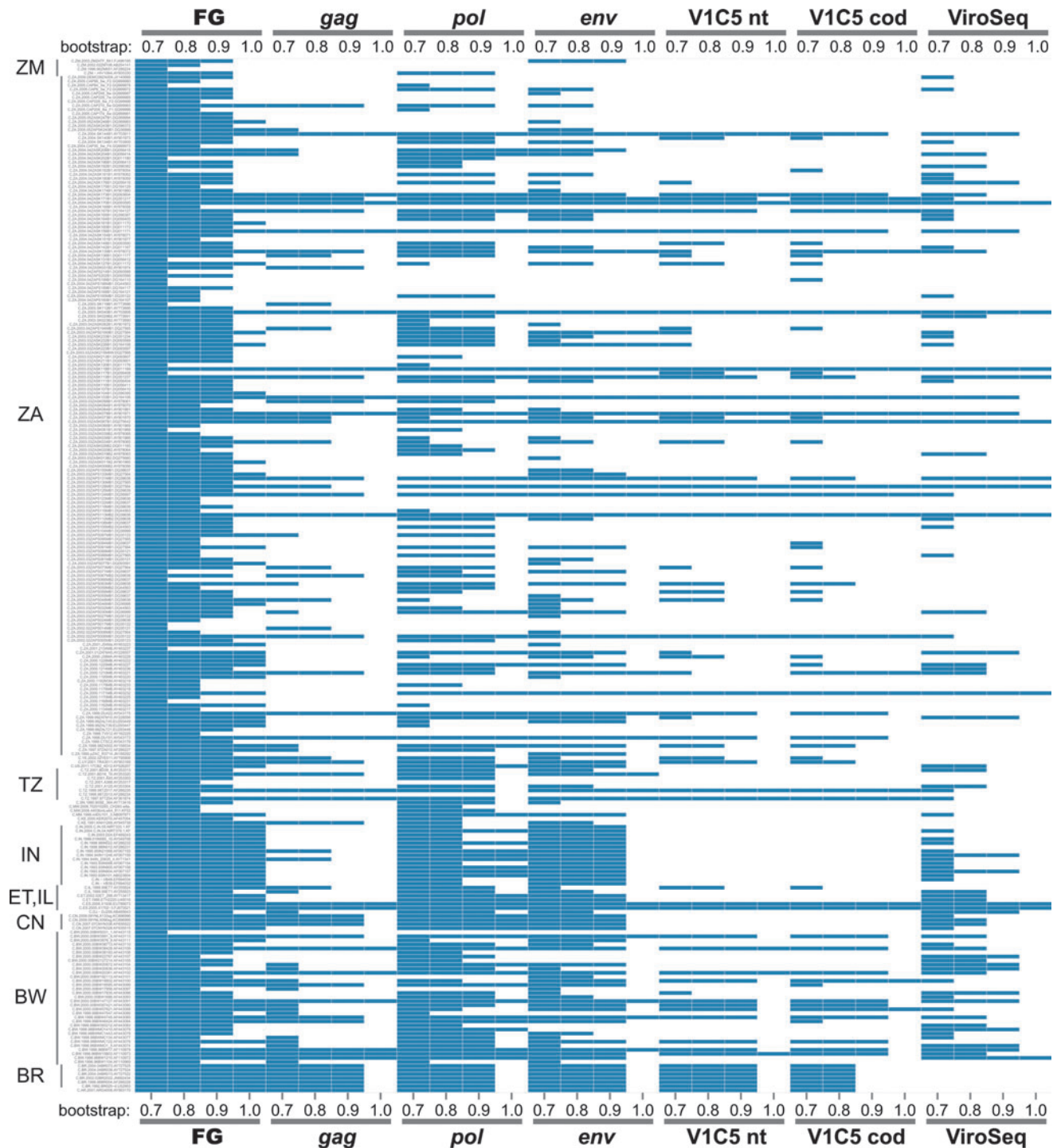


Supplementary Data



SUPPLEMENTARY FIG. S1. Clustering of HIV-1C sequences, $n=236$. A subset of 236 HIV-1C near full-length genome (FG) sequences found in clusters in bootstrapped $ML_{GTR+\Gamma}$ analysis by RAxML at a bootstrap support of 0.70 is shown. Clustering at four bootstrap thresholds, 0.7, 0.8, 0.9, and 1.0, is shown for FG, *gag*, *pol*, *env*, the V1C5 region of the gp120 nucleotide (V1C5 nt) and codon-based (V1C5 cod) alignments, and the partial *pol* region corresponding to ViroSeq. For clarity, analyzed regions and bootstrap thresholds are shown at the *top* and at the *bottom* of the graph. Clustering of the viral sequence at the specified bootstrap threshold is denoted by a *filled block*, while nonclustering is shown with a *blank block*. Viral sequences are listed on the *left* and grouped by country: BR, Brazil; BW, Botswana; CN, China; ET, Ethiopia; IL, Israel; IN, India; TZ, Tanzania; ZA, South Africa; and ZM, Zambia. Countries represented by small numbers of sequences are not delineated.