



**Figure S2**

MAP topologies for HIV-1 dataset, arbitrarily rooted at subtype C sequence. Below each topology we have the first site of the region it comprises. The number at the tips correspond to the following sequences: (1) CH12, (2) 12\_BF.AR.99.ARMA159, (3) 12\_BF.UY.99.URTR35, (4) 12\_BF.AR.97.A32879, (5) BF.AR.99.ARMA029, (6) BF.AR.99.ARMA097, (7) BF.AR.99.A027 and (8) BF.AR.99.A047 .