Supplemental online content for Brumme *et al*, **"Extensive host immune adaptation in a concentrated North American HIV epidemic"**, AIDS 2018

Figure S1: Unrooted approximate maximum-likelihood phylogenies inferred from SK HIV Pol sequence alignments. ***Panel A:*** Whole SK dataset; terminal branches coloured by HIV subtype. ***Panel B:*** HIV subtype B only; terminal branches coloured by collection year.

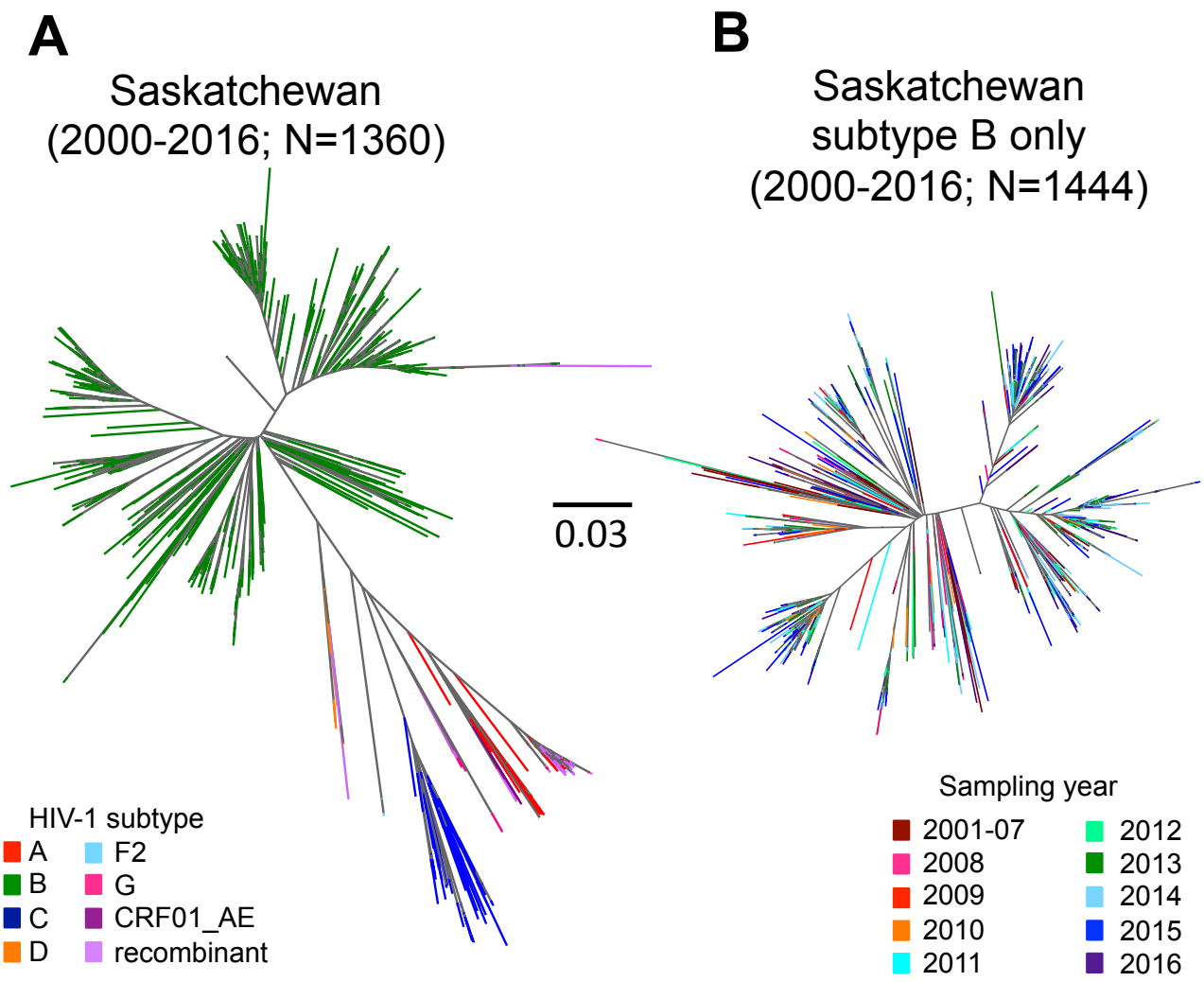


Figure S2: Rooted approximate maximum-likelihood phylogeny inferred from SK HIV Pol sequence alignments, coloured by cluster size. To facilitate visualization of clusters, branch colours are propagated to shared internal nodes.

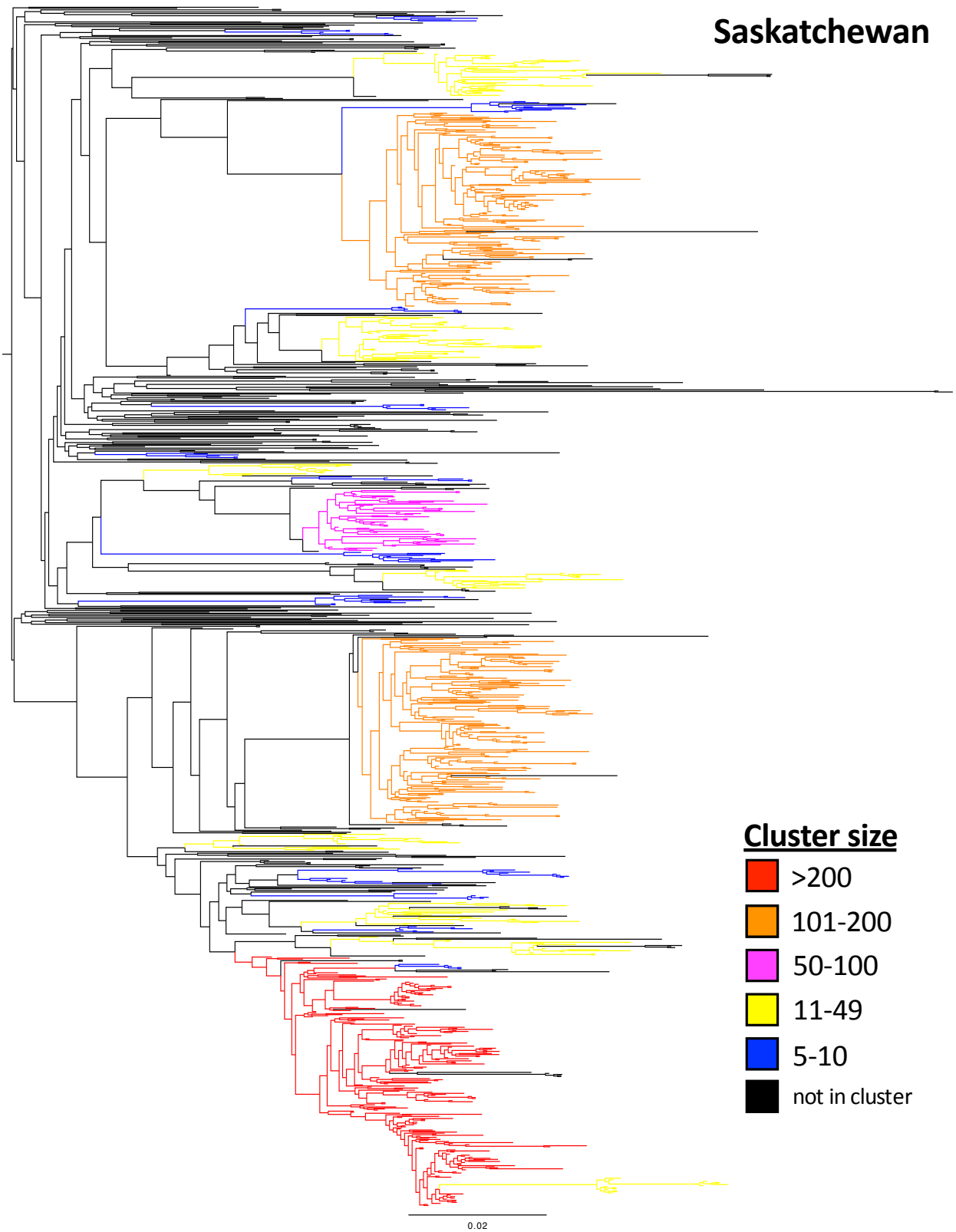


Figure S3: Rooted approximate maximum-likelihood phylogeny inferred from CA/US HIV Pol sequence alignments, coloured by cluster size. To facilitate visualization of clusters, branch colours are propagated to shared internal nodes.

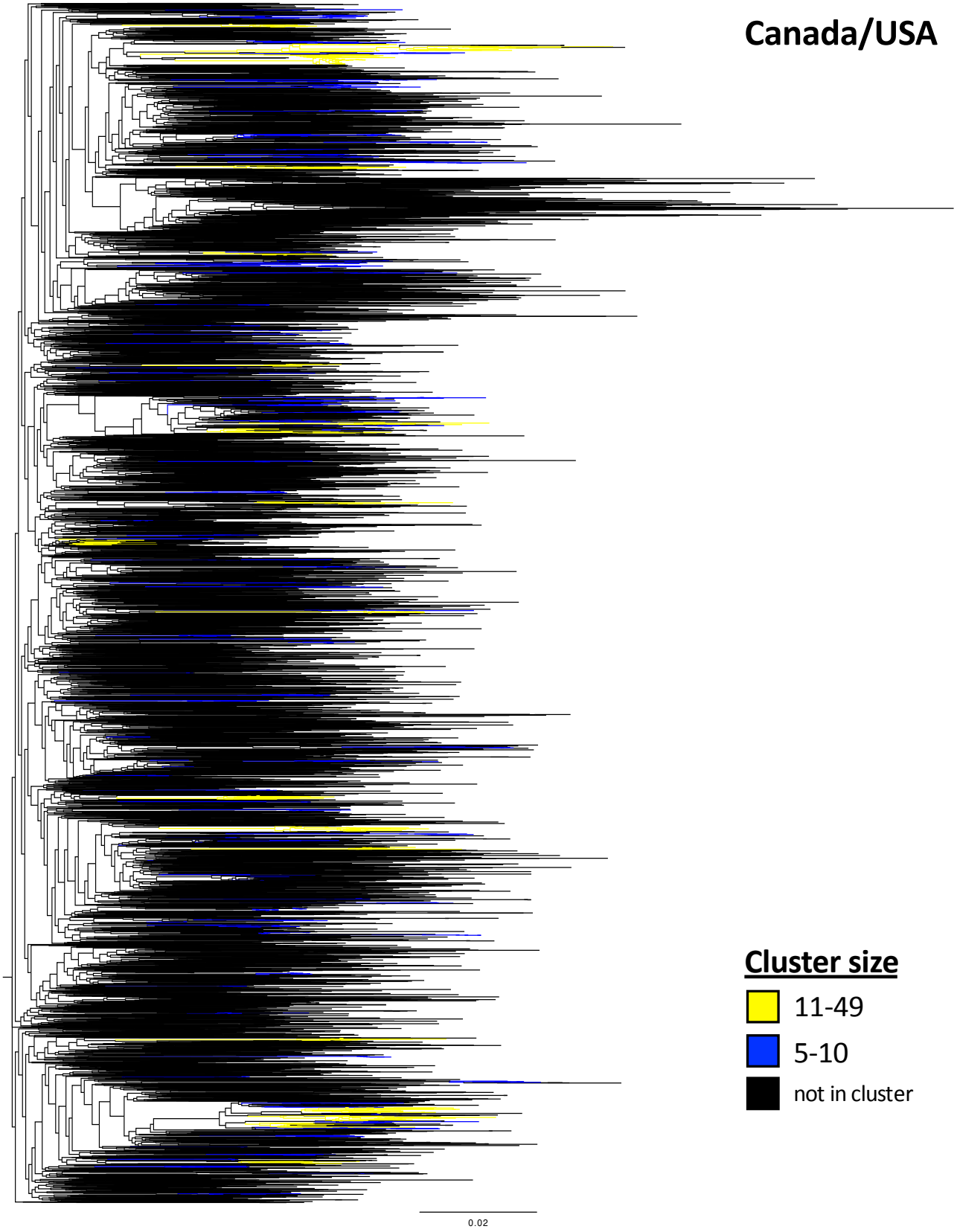


Figure S4: Combined SK/CA/USA phylogeny. Terminal branches are coloured by region (SK green; CA/US grey).

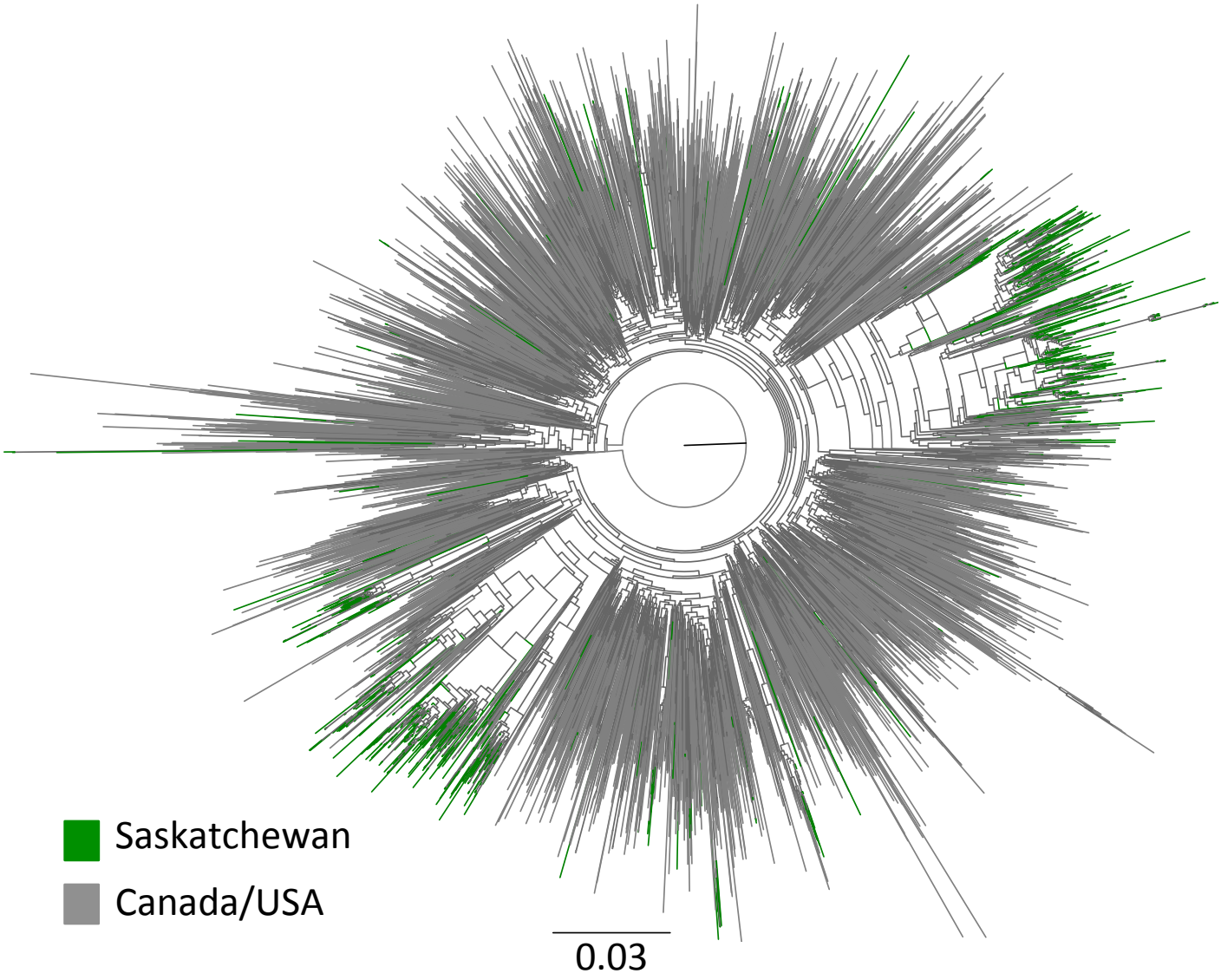


Figure S5: Comparison of phylogenies, and burden of HLA-associated HIV Pol adaptations in SK versus BC. ***Panel A:*** Unrooted Maximum-Likelihood phylogenies inferred from 1,144 and 6,525 HIV subtype B sequences collected between 2000-2016 from Saskatchewan (SK; green) and British Columbia ("BC"; blue) respectively, depicted on the same distance scale. ***Panel B:*** *top:* Percent of sequences in SK and BC datasets that fall within a cluster. *bottom:* cluster size distributions in SK and BC (expressed as a proportion of the total number of clusters in each dataset). ***Panels C, D, E:*** Frequency comparison of HLA-associated adapted variants in HIV Protease, RT codons 1-250 and RT codons 251-400, respectively, in SK versus BC, shown as linked pairs. Adapted variants with >10% *higher* frequency in SK compared to BC are labeled in green; the single variant with >10% *lower* frequency in SK is labeled in black. Horizontal dotted line indicates the 5% frequency threshold used to define common HLA-adapted variants. Overall, common HLA-adapted variants were significantly elevated in frequency in SK compared to BC (Mann-Whitney p=0.01).

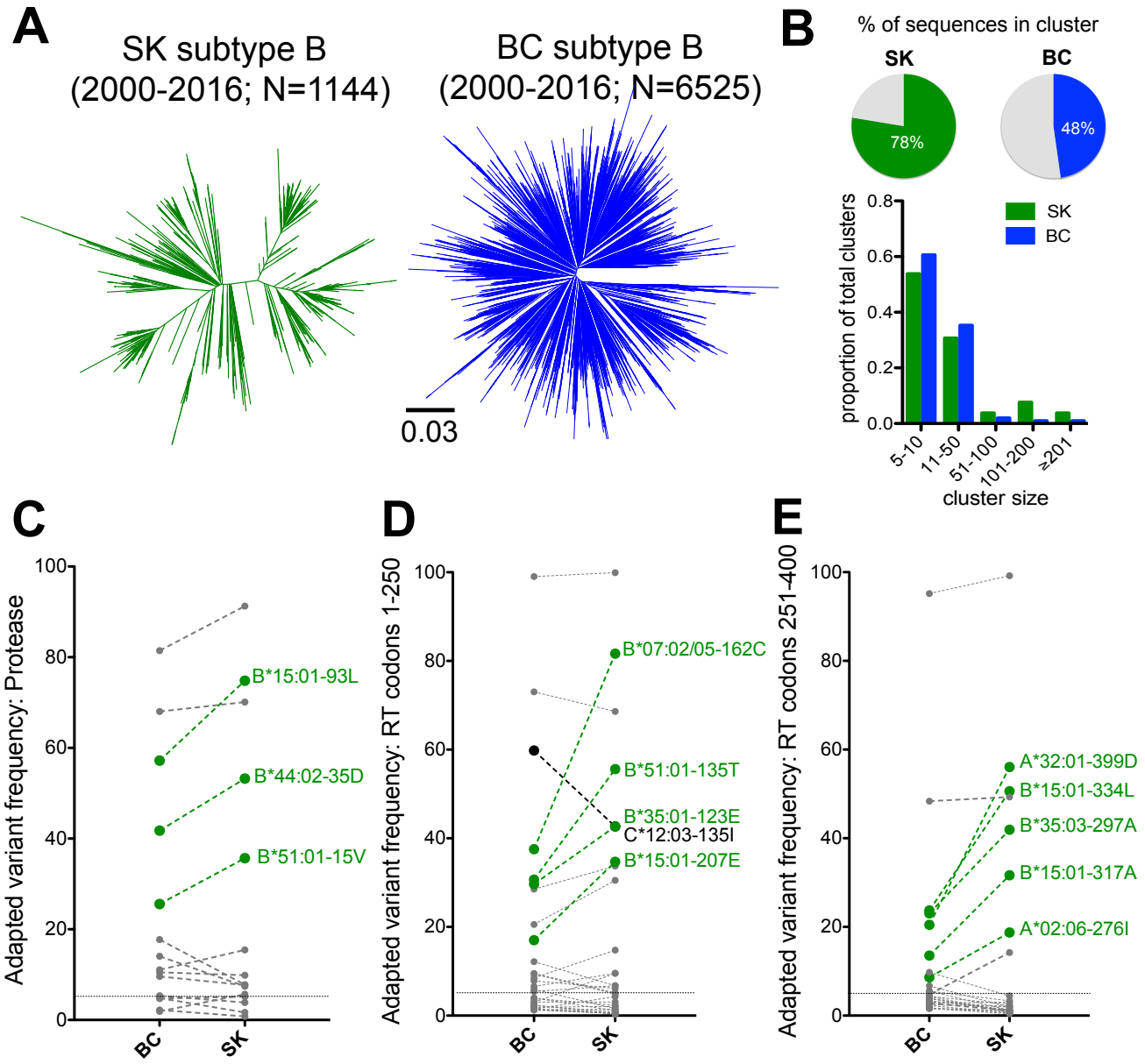


Figure S6: Rooted approximate maximum-likelihood phylogeny inferred from BC HIV Pol sequence alignments, coloured by cluster size. To facilitate visualization of clusters, branch colours are propagated to shared internal nodes.

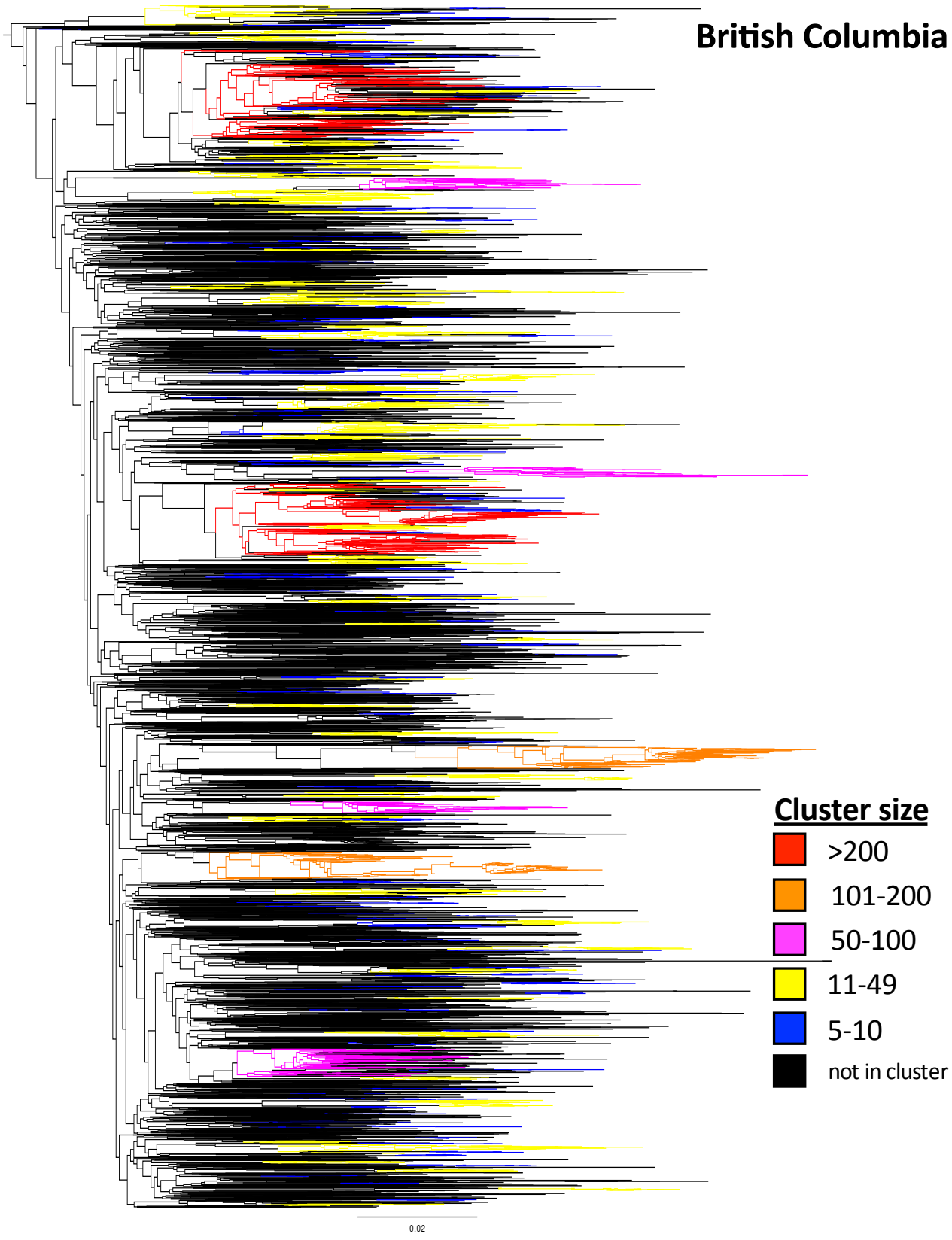


Figure S7: HLA frequency distributions and allele coverage of adaptation analysis, in relevant populations. ***Panel A:*** Estimated population frequencies of HLA class I alleles among HIV-infected populations in Canada and the USA [38]. Only HLA alleles observed in >1% of persons are shown; those associated with at least one Pol adaptation analyzed in the present study are coloured. ***Panel B:*** Population HLA-A and HLA-B serotype frequencies among Indigenous persons in Saskatchewan, published as part of a study of diabetes risk (see table 1 in [21]), presented here with kind permission from Dr. Roland Dyck. Frequency summaries represent those from the combined Indigenous cohort (*i.e.* participants with and without diabetic end-stage renal disease, as there were no significant HLA frequency differences between these groups) [21]. HLA serotypes associated with at least one Pol adaptation analyzed in the present study are coloured; A02 is cross-hatched as our analysis includes A*02:06, but not the more common A*02:01.

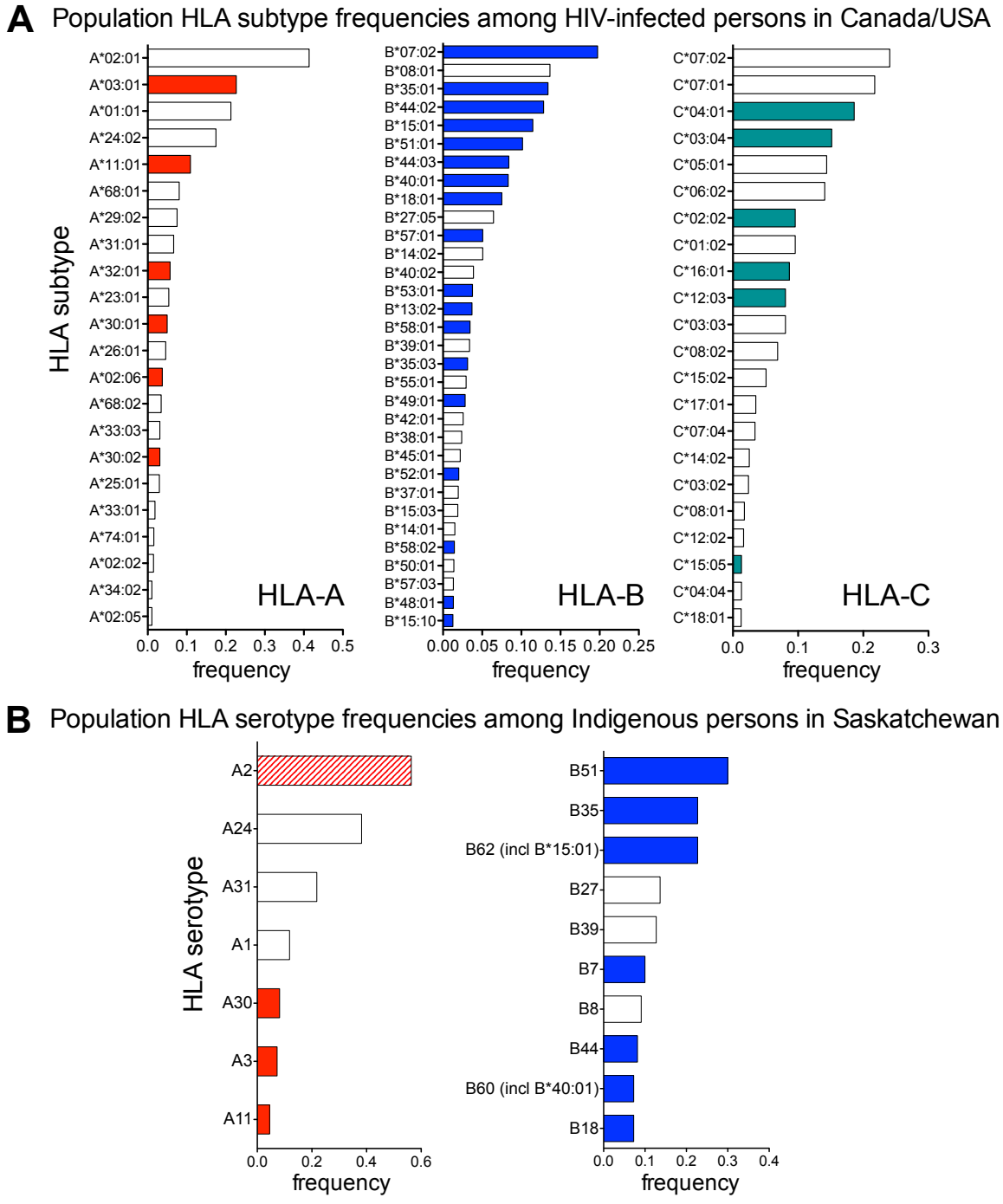


Figure S8: Higher resolution map of the 70 HLA-associated polymorphisms investigated in the present study (same as Figure 2A). Legend on following page.

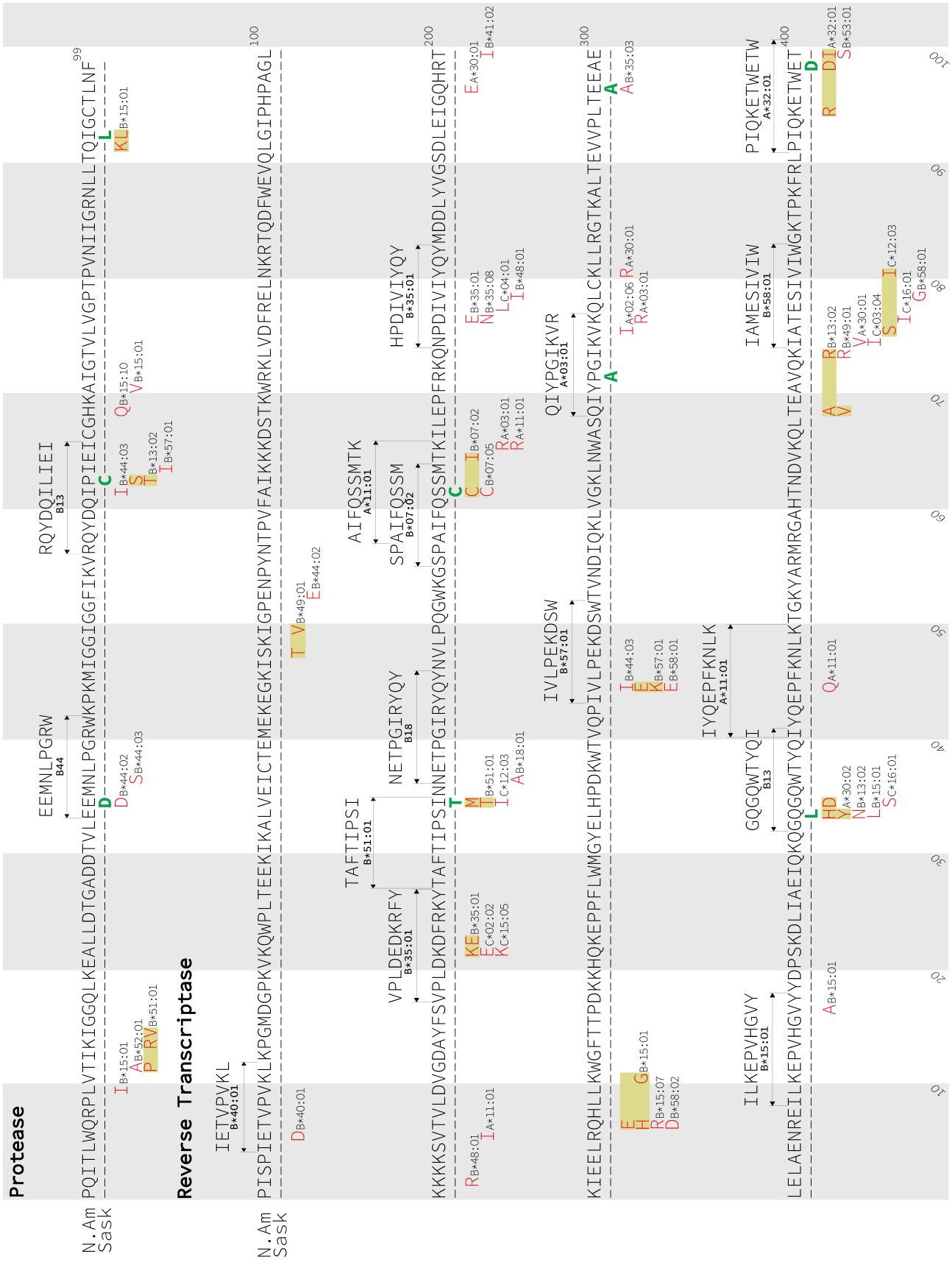


Figure S8 (previous page): Higher resolution map of the 70 HLA-associated polymorphisms investigated in the present study (same as Figure 2A).

This higher-resolution version of Figure 2A depicts the 70 HLA-associated adaptations in HIV subtype B Protease and the first 400 codons of RT, defined at $q<0.05$ in an independent population-based study [38], that were investigated in the present analysis. Specific adaptations are shown in red, along with their restricting HLA, below the CA/US and Saskatchewan consensus Pol amino acid sequences. Nearby adaptations restricted by the same HLA allele are boxed together in yellow. For HLA-adapted variants that occur within an optimally-described HIV CTL epitope restricted by that allele [47], the published epitope sequence and HLA restriction are shown. In total, the 70 adaptations are restricted by 34 HLA class I alleles (6 HLA-A, 23 HLA-B and 5 HLA-C), occur at 52 Protease and RT codons, and include numerous experimentally verified immune escape mutations in optimally-described CTL epitopes (*e.g.*, B*51:01-RTI135T that occurs at the C-terminus of the B*51:01-restricted TI8 epitope and abrogates its ability to bind this allele [5,12,38]). The list includes HIV codons where the same adapted form is selected by distinct HLA alleles (e.g., both B*07:02 and B*07:05 select RT-S162C), codons where two or more adapted forms are described for the same allele at a given position (*e.g.*, B*51:01 normally selects RT-I135T but may also select I135M) and codons where distinct alleles select opposing mutations (*e.g.*, B*51:01 selects I135T, while C*12:03 induces pressure to maintain the global consensus I at this site).