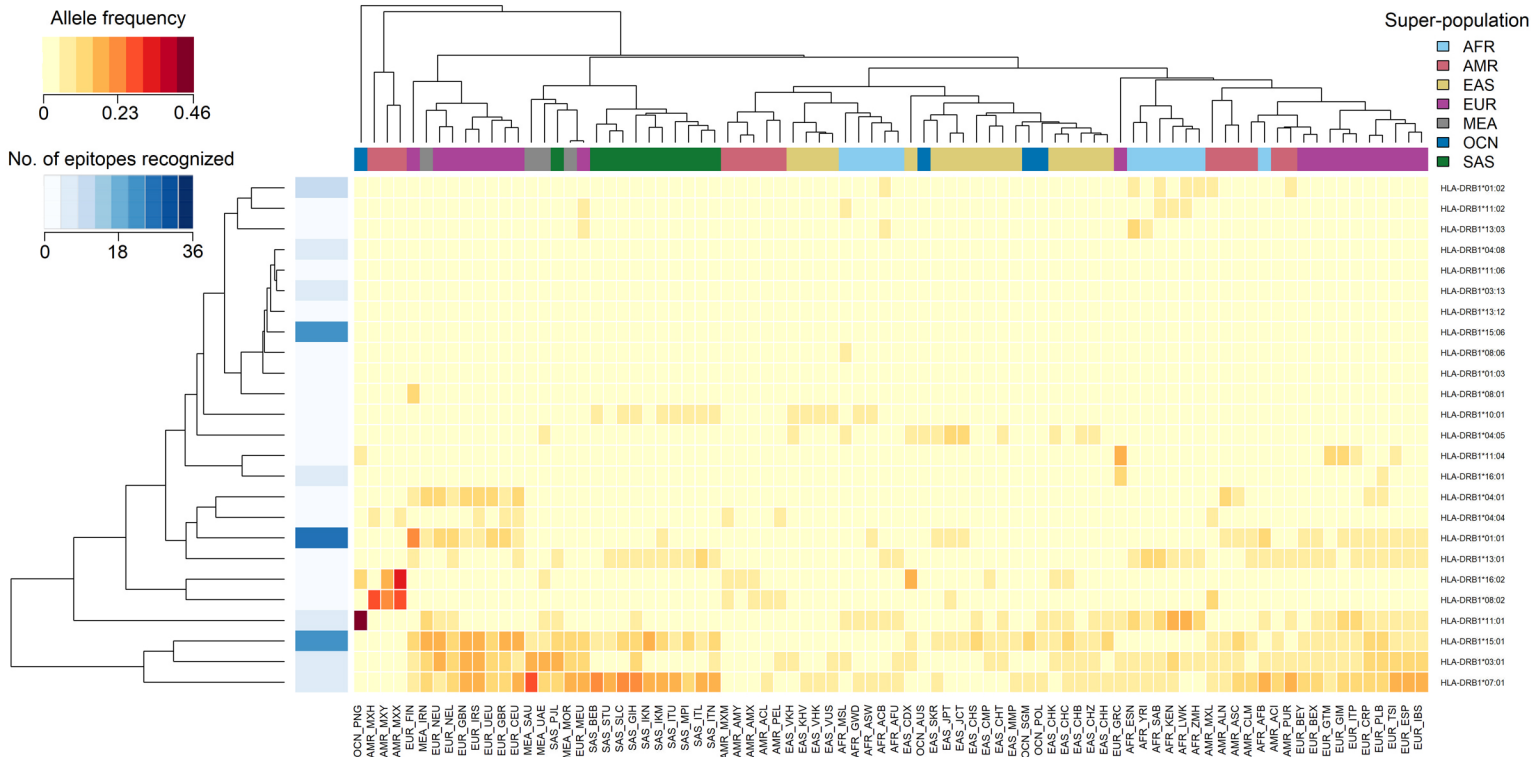
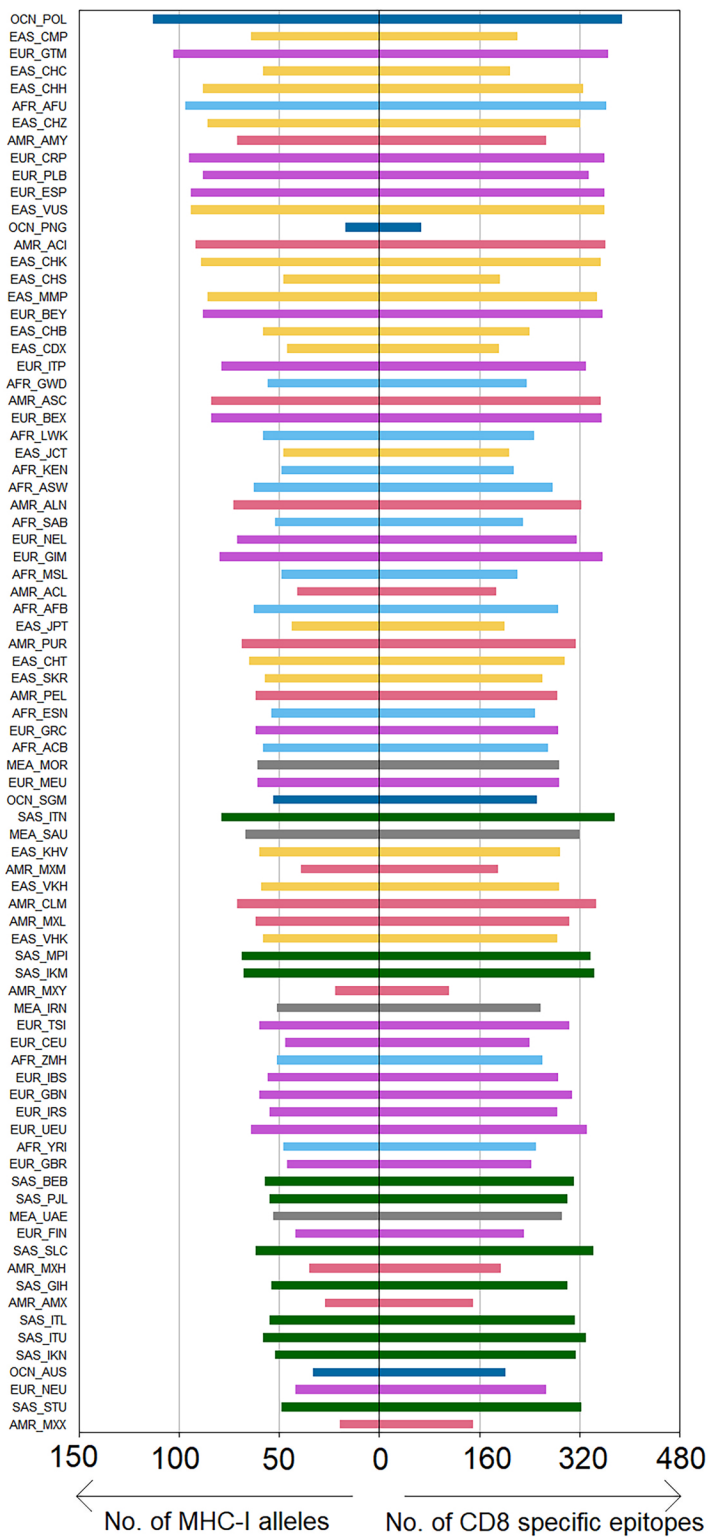


Supplementary Fig. 1: Distribution of MHC-I alleles among the ethnic groups. Heat-map depicting the distribution of MHC-I alleles among the 82 ethnic groups involved in this study. Intensity of the colour indicates frequency of a particular allele in an ethnic group. Both the ethnic groups and the MHC-I alleles have been hierarchically clustered. An additional colour-key along the vertical axis indicates the number of SARS-CoV2 epitopes recognized by the HLA alleles. Along the horizontal axis ethnic groups have been tagged with different colours based on their affiliations to respective super-populations.

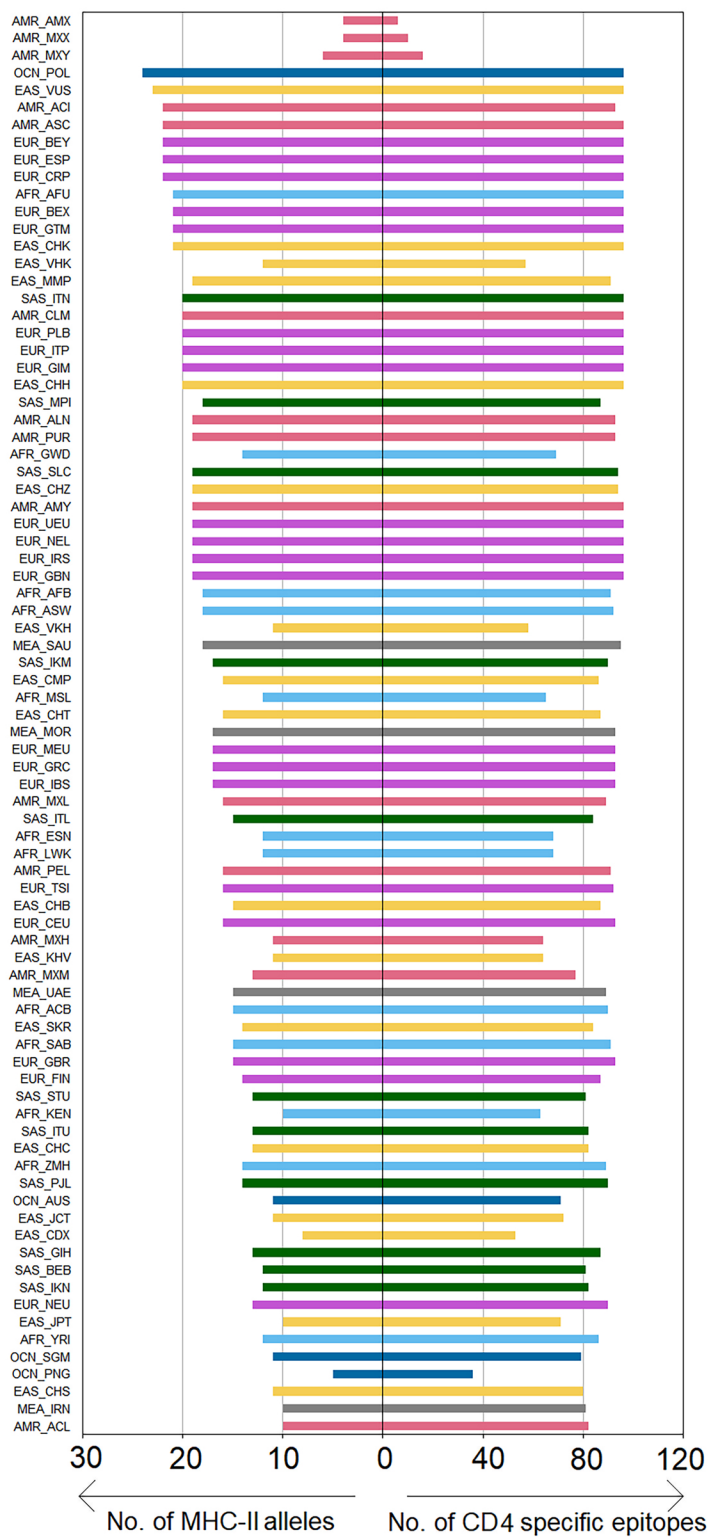


Supplementary Fig. 2: Distribution of MHC-II alleles among the ethnic groups. Heat-map depicting the distribution of MHC-II alleles among the 82 ethnic groups involved in this study. Intensity of the colour indicates frequency of a particular allele in an ethnic group. Both the ethnic groups and the MHC-II alleles have been hierarchically clustered. An additional colour-key along the vertical axis indicates the number of SARS-CoV2 epitopes recognized by the HLA alleles. Along the horizontal axis ethnic groups have been tagged with different colours based on their affiliations to respective super-populations.

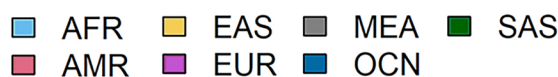
(A)



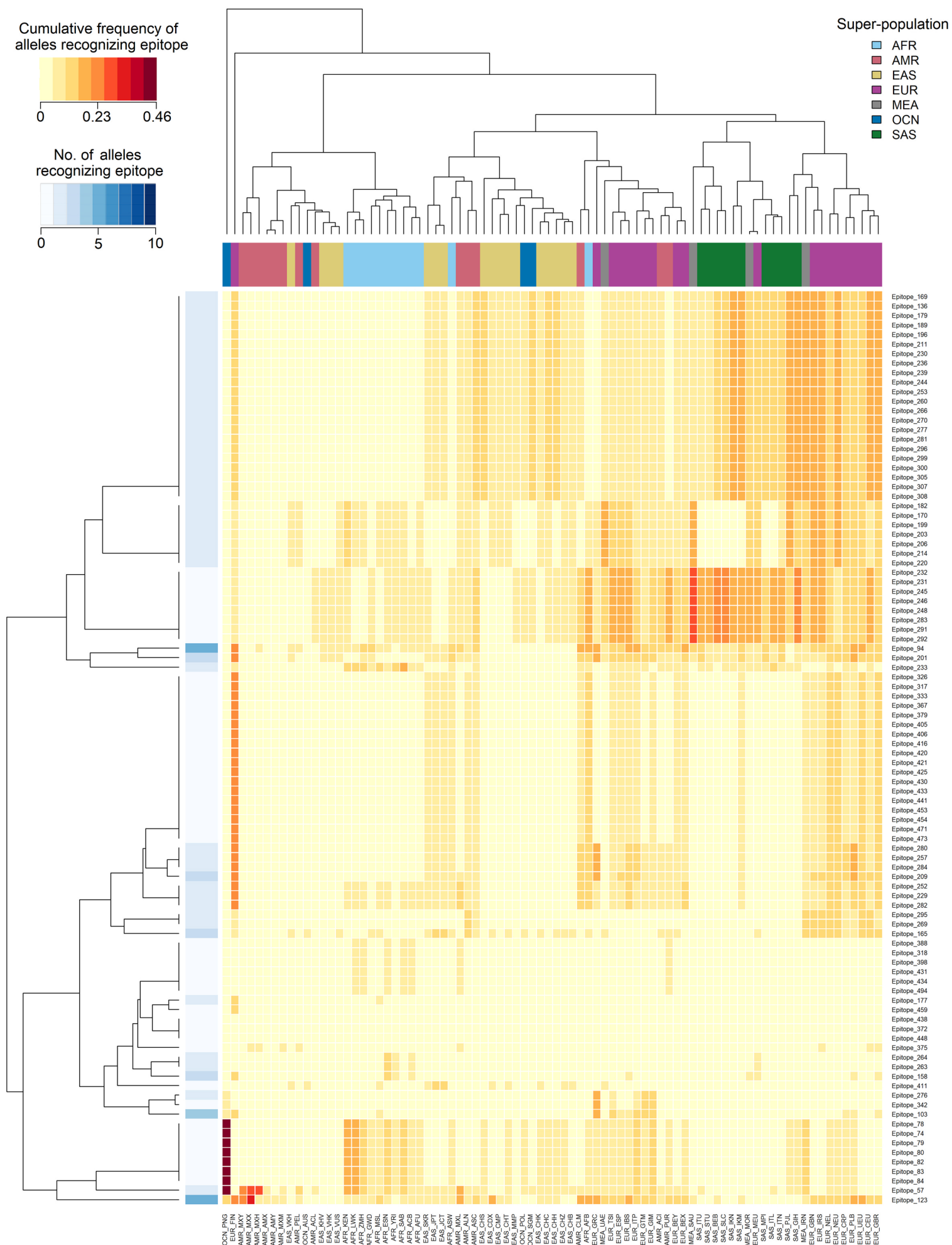
(B)



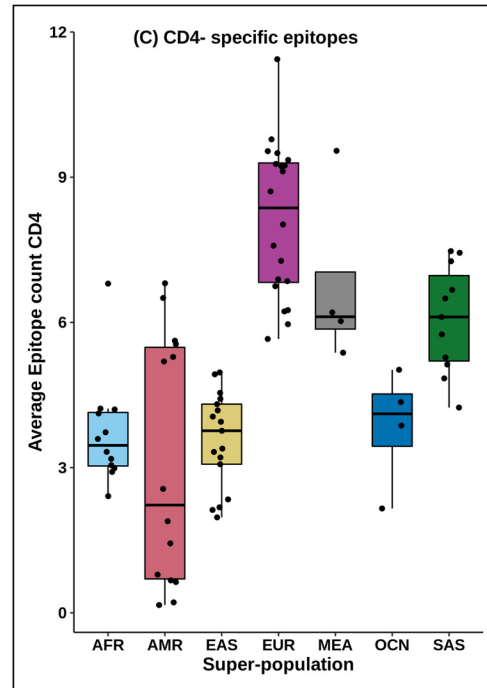
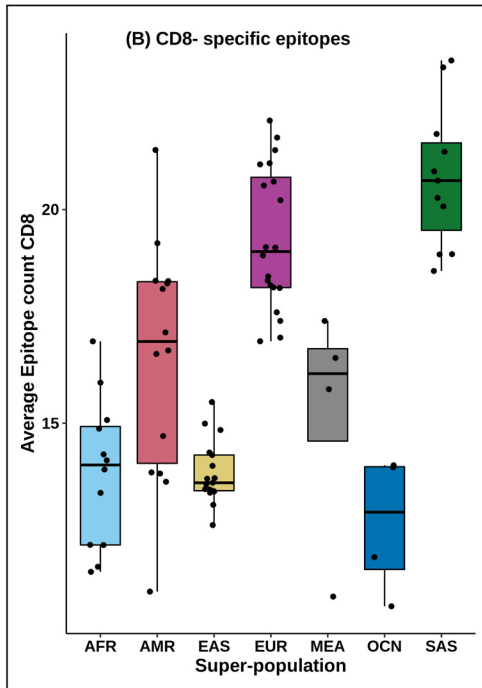
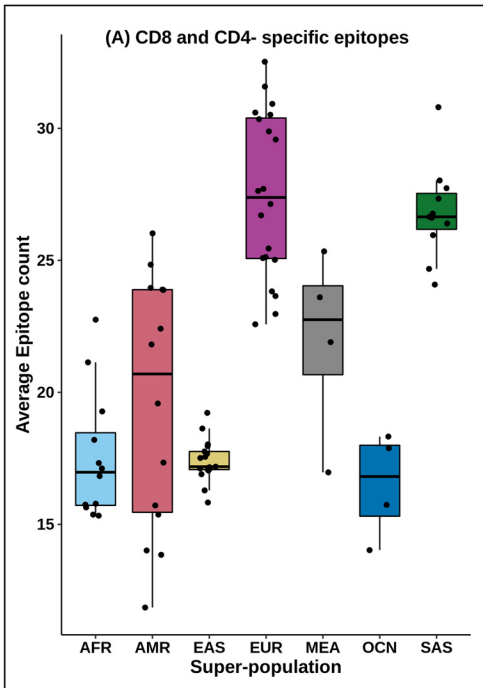
Super-population



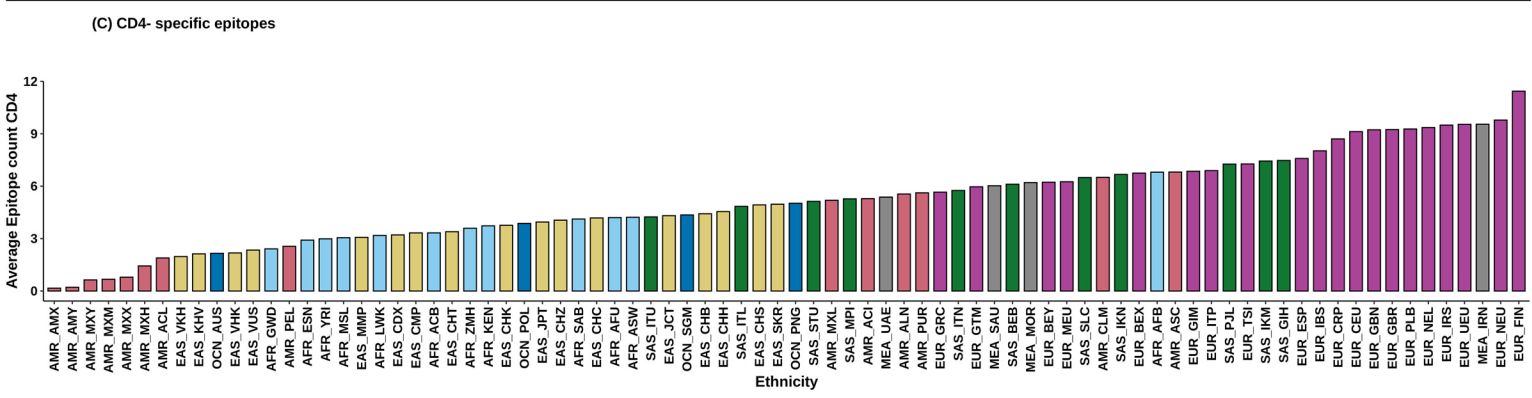
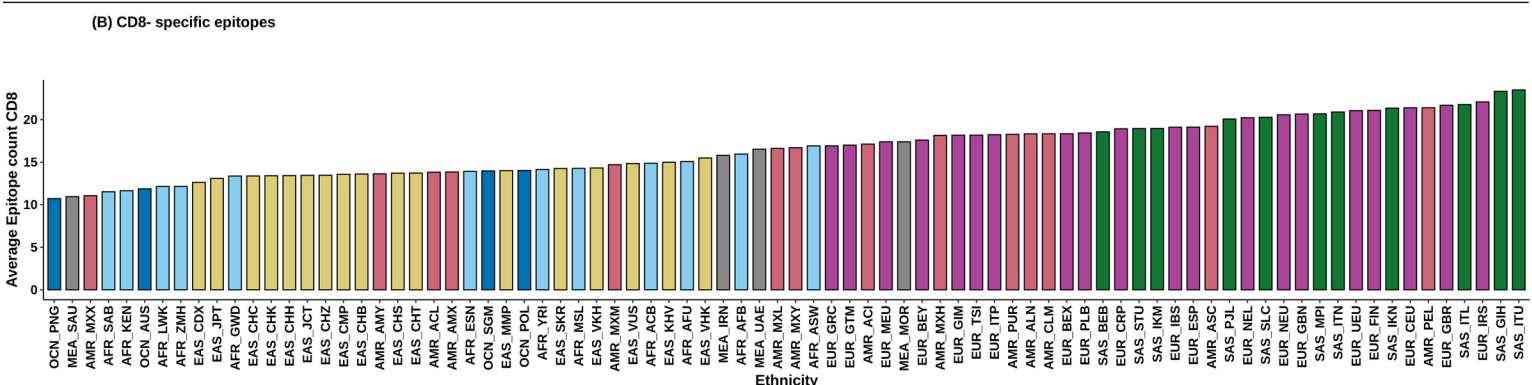
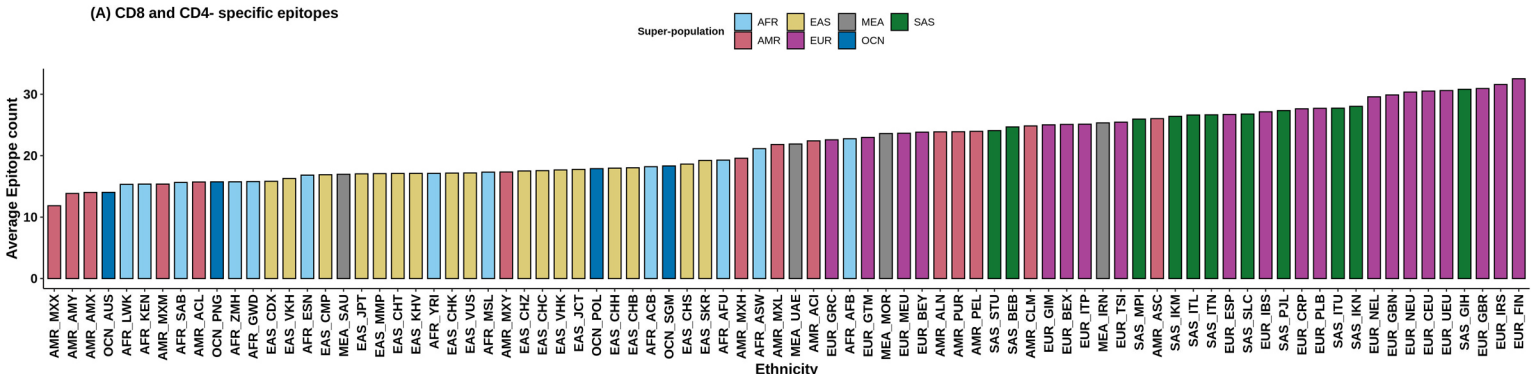
Supplementary Fig. 3: Comparative view of HLA allelic richness and cognate epitopes across populations. Mirror-plots indicating the richness of MHC alleles in different populations (along the left half of the plots) along with the richness of SARS-CoV2 epitopes predicted to be recognized by these HLA alleles (along the right half of the plots). (A) MHC-I alleles and predicted cognate CD8-specific epitopes and (B) MHC-II alleles and predicted cognate CD4-specific epitopes, across 82 different populations (ethnic groups). In each of the plots, the ethnicities were sorted along the vertical axis based on the ratio of the richness of MHC alleles to the richness of epitopes recognized. Super-population names are abbreviated as AFR- Africans; AMR- Amerindians; EAS- East Asians; EUR- Europeans; MEA- Middle East and Africans; OCN- Oceanians; SAS- South Asians, and prefixed to each of the population names in the plots. The super-populations are also represented with horizontal bars of specific colors. The abbreviations used for names of different populations/ ethnic groups are listed in Supplementary Table 2.



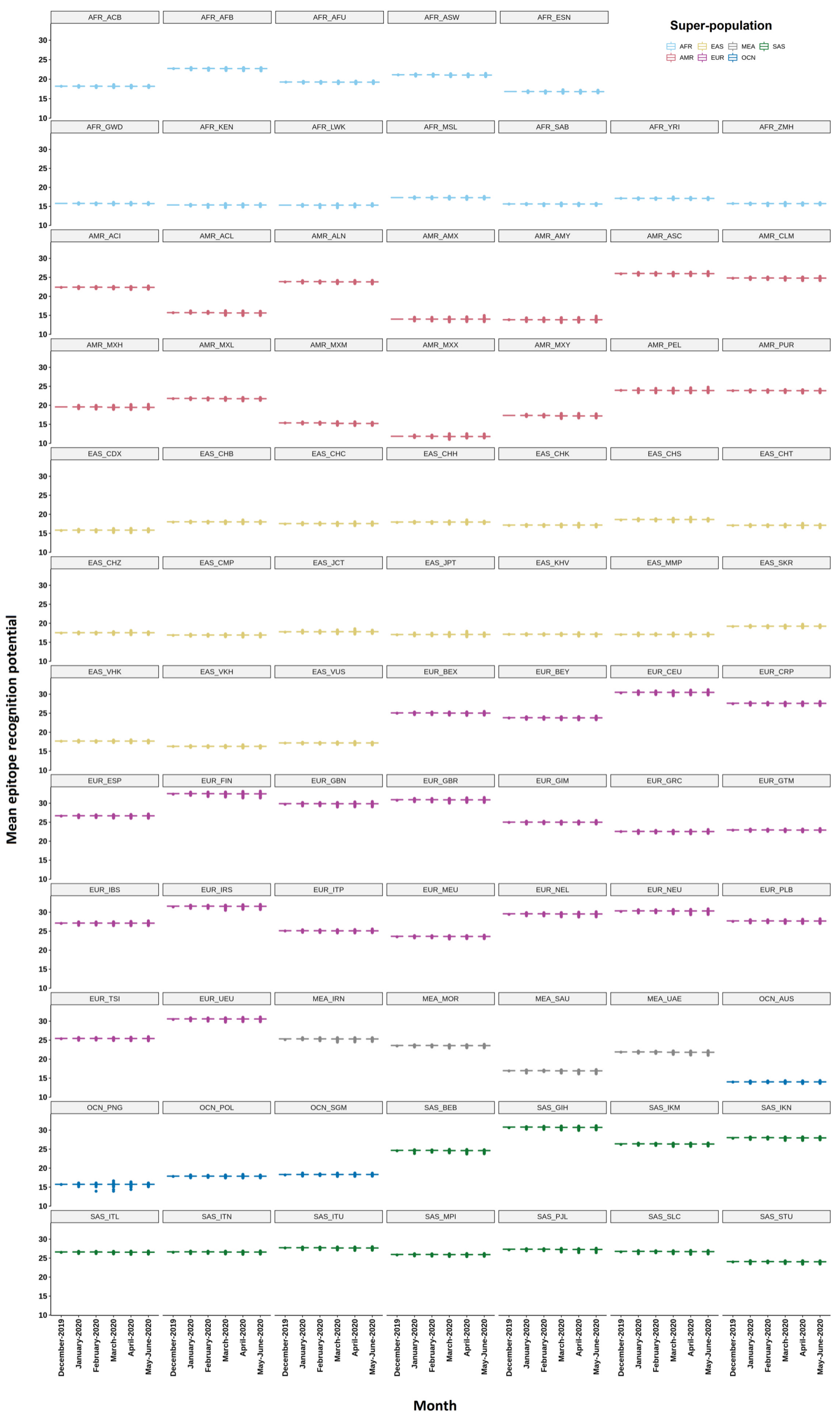
Supplementary Fig. 5: Distribution of CD4-specific epitopes recognized by the HLA-system among the ethnic groups. Heat-map depicting the distribution of CD4-specific epitopes that could be recognized by the HLA alleles prevalent among the 82 ethnic groups involved in this study. Both the ethnic groups and the CD4-specific epitopes have been hierarchically clustered. An additional colour-key along the vertical axis indicates the number of human HLA-types capable of recognizing the epitopes. Along the horizontal axis ethnic groups have been tagged with different colours based on their affiliations to respective super-populations.



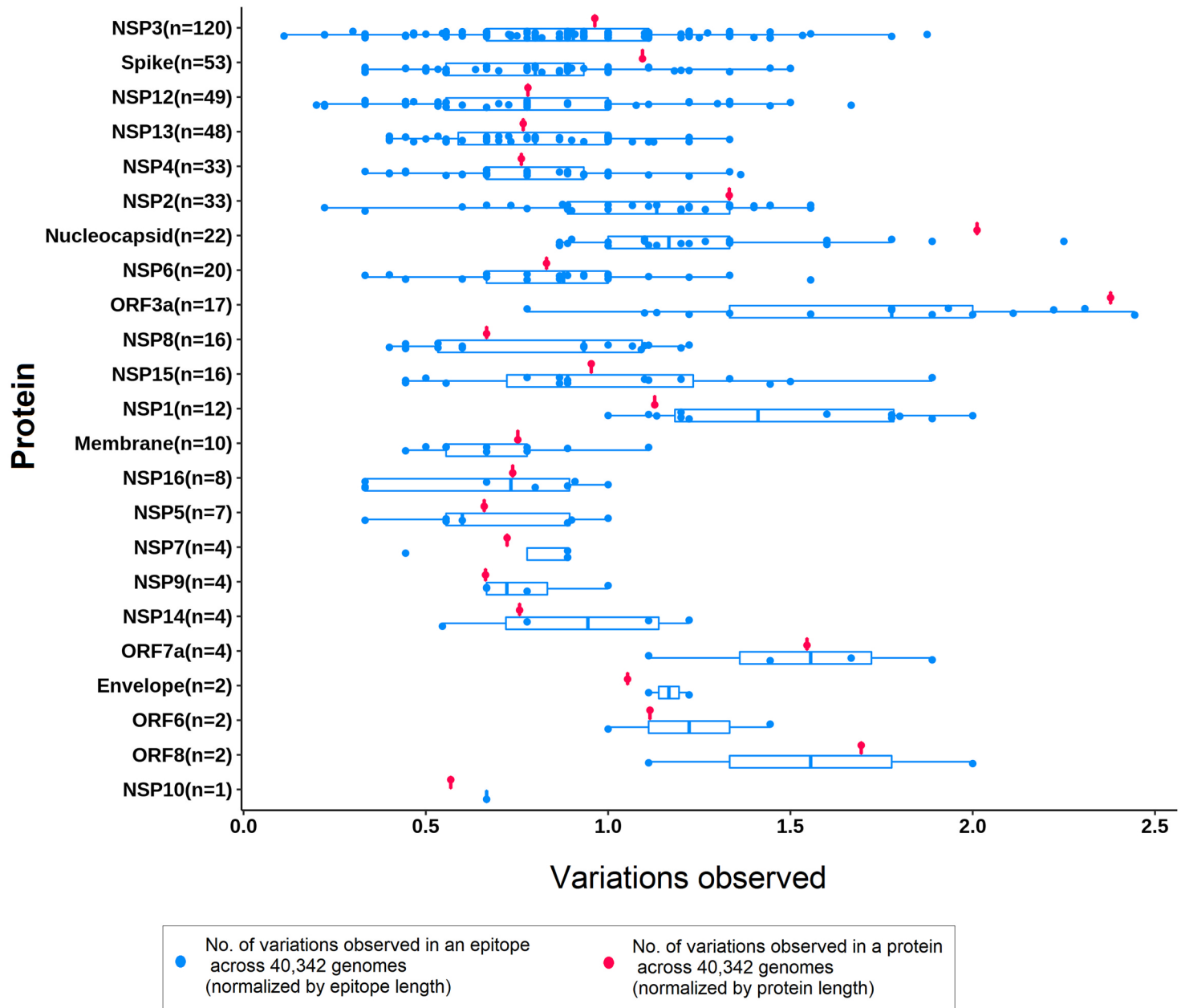
Supplementary Fig. 6: Average count of SARS-CoV2 epitopes recognized by each super-population. Box plot representing average count of the number of SARS-CoV2 epitopes identified by the ethnicities comprising each of the super-populations: (A) Total (CD8 and CD4-specific) epitopes, (B) CD8-specific epitopes and (C) CD4-specific epitopes. Super-population names are abbreviated as AFR- Africans; AMR- Amerindians; EAS- East Asians; EUR- Europeans; MEA- Middle East and Africans; OCN- Oceanians; SAS- South Asians.



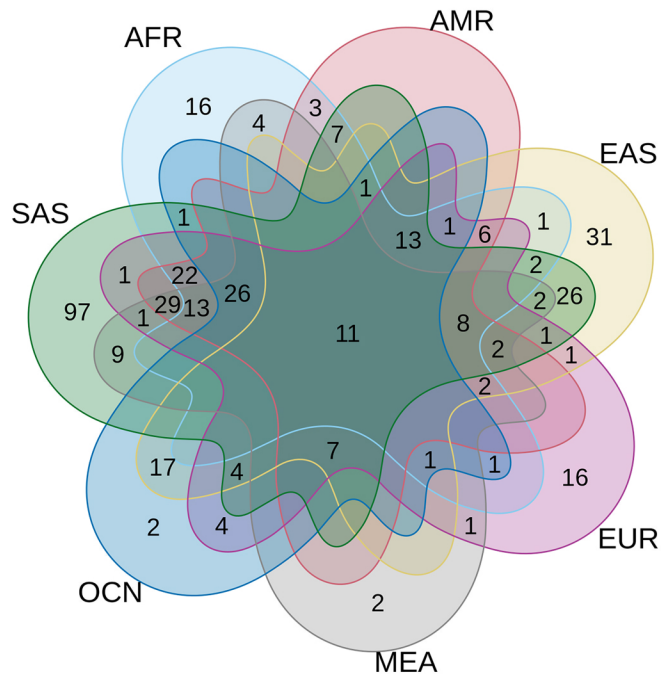
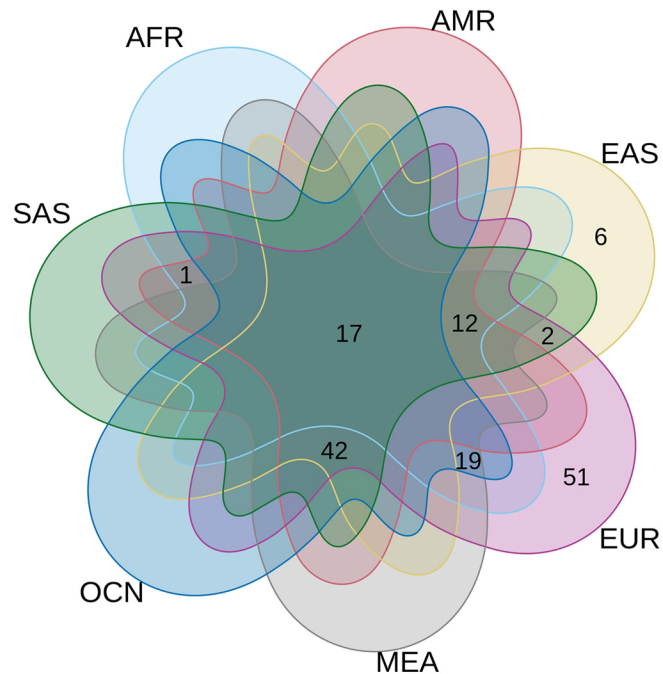
Supplementary Fig. 7: Average count of SARS-CoV2 epitopes recognized by each ethnic group. Bar plot representing average count of the number of SARS-CoV2 epitopes identified by the samples comprising each of the ethnic groups: (A) Total (CD8 and CD4-specific) epitopes, (B) CD8-specific epitopes and (C) CD4-specific epitopes. Super-population names are abbreviated as AFR- Africans; AMR- Amerindians; EAS- East Asians; EUR- Europeans; MEA- Middle East and Africans; OCN- Oceanians; SAS- South Asians, and prefixed to each of the population names in the plots. Abbreviations used for population names are reported in Supplementary Table 2.



Supplementary Fig. 8: SARS-CoV2 genome variations and epitope recognition abilities. Representation of the mean SARS-CoV2 epitope recognition potential among different ethnic groups with respect to SARS-CoV2 genomes sequenced till 11th June 2020, as obtained from the GISAID (<https://www.gisaid.org/>). Each box plot denotes an ethnic population; each data-point on the plot represents the average of the number of epitopes from SARS-CoV2 genomes (binned according to month of collection provided by GISAID), that could be identified by individuals in that ethnic population.



Supplementary Fig. 9: Variations observed in epitopes and corresponding proteins in SARS-CoV2 genomes. Plot representing number of variations in the amino acid sequence of the epitopes and corresponding proteins across 40,342 SARS-CoV2 genomes, normalized by the epitope and protein lengths respectively. Number of predicted epitopes identified in each individual protein is mentioned alongside the protein name.

(A) CD8**(B) CD4**

Supplementary Fig. 10: Distribution of epitopes shortlisted as vaccine peptides across super-populations. Euler representation of (A) CD8-specific and (B) CD4-specific SARS-CoV2 epitopes recognized by the HLA alleles in the seven super-populations. Epitopes observed in at least 10 SARS-CoV2 genomes and those potentially recognized by at least 5% of the individuals representing a super-population have been considered. The CD8-specific and CD4-specific epitopes at the intersection of the seven super-populations could serve as potential vaccine candidates. Super-population names are abbreviated as AFR- Africans; AMR- Amerindians; EAS- East Asians; EUR- Europeans; MEA- Middle East and Africans; OCN- Oceanians; SAS- South Asians.