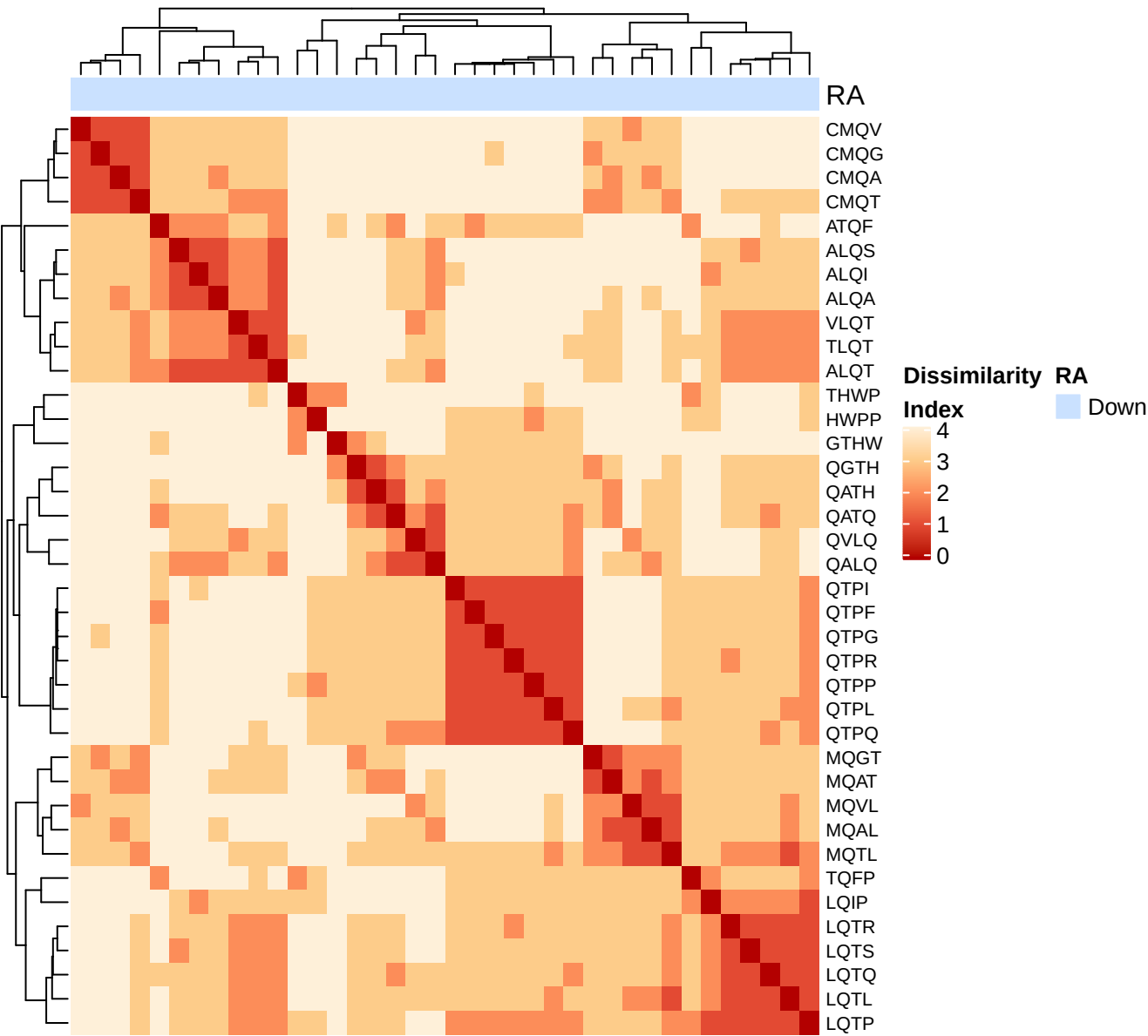
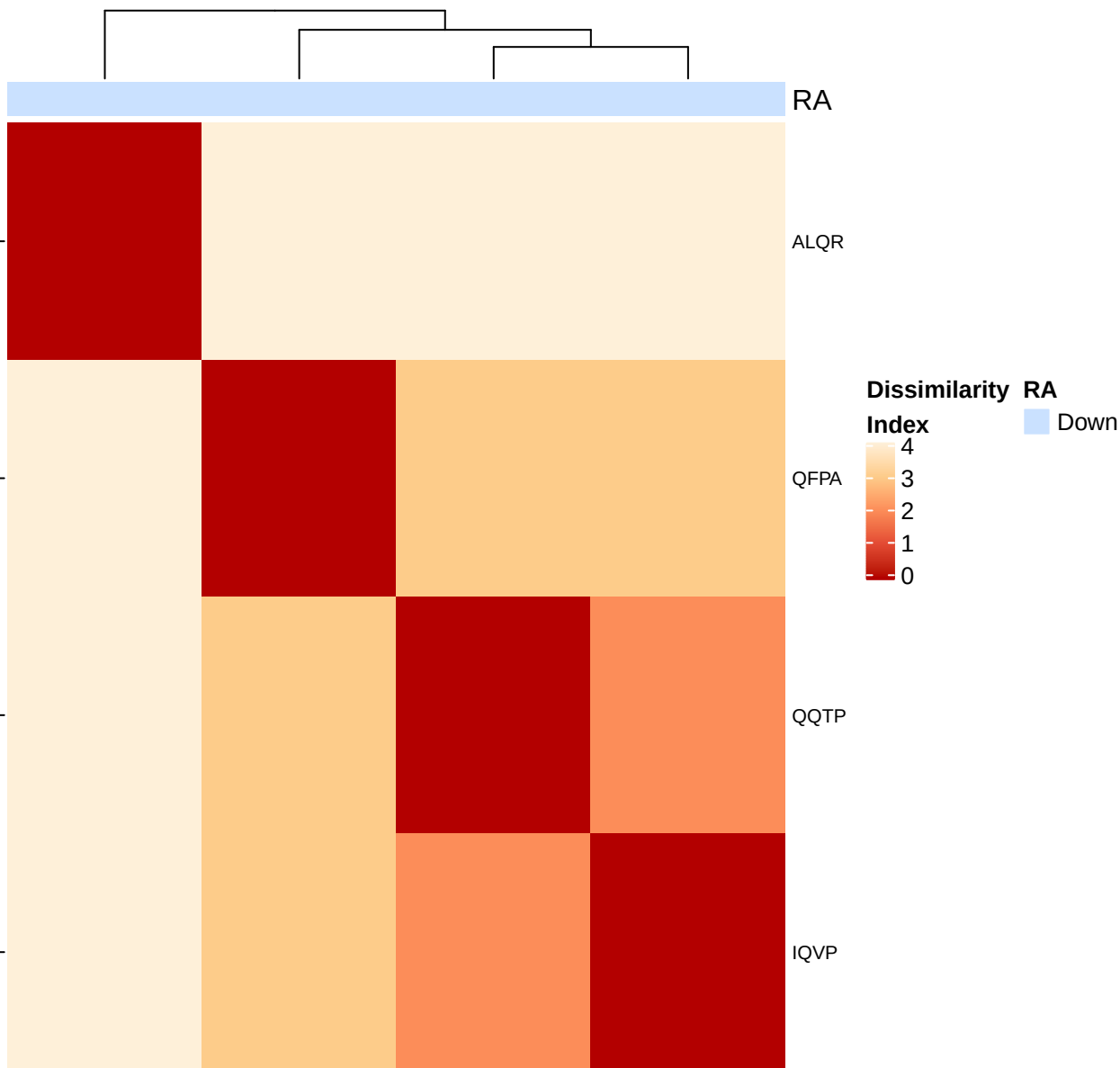


Fig S6. Pairwise dissimilarity matrix among the amino acid sequences of TRA/TRB/IGL/IGK k-mers associated with rheumatoid arthritis. K-mers over- and under-represented in rheumatoid arthritis are represented in dark and light blue, respectively. The dissimilarity index was computed using the Levenshtein distance measure so that the higher the Levenshtein distance, the higher the dissimilarity index between two k-mer sequences. Abbreviations: Cont, continuous model; Disc, discrete model; RA, rheumatoid arthritis.

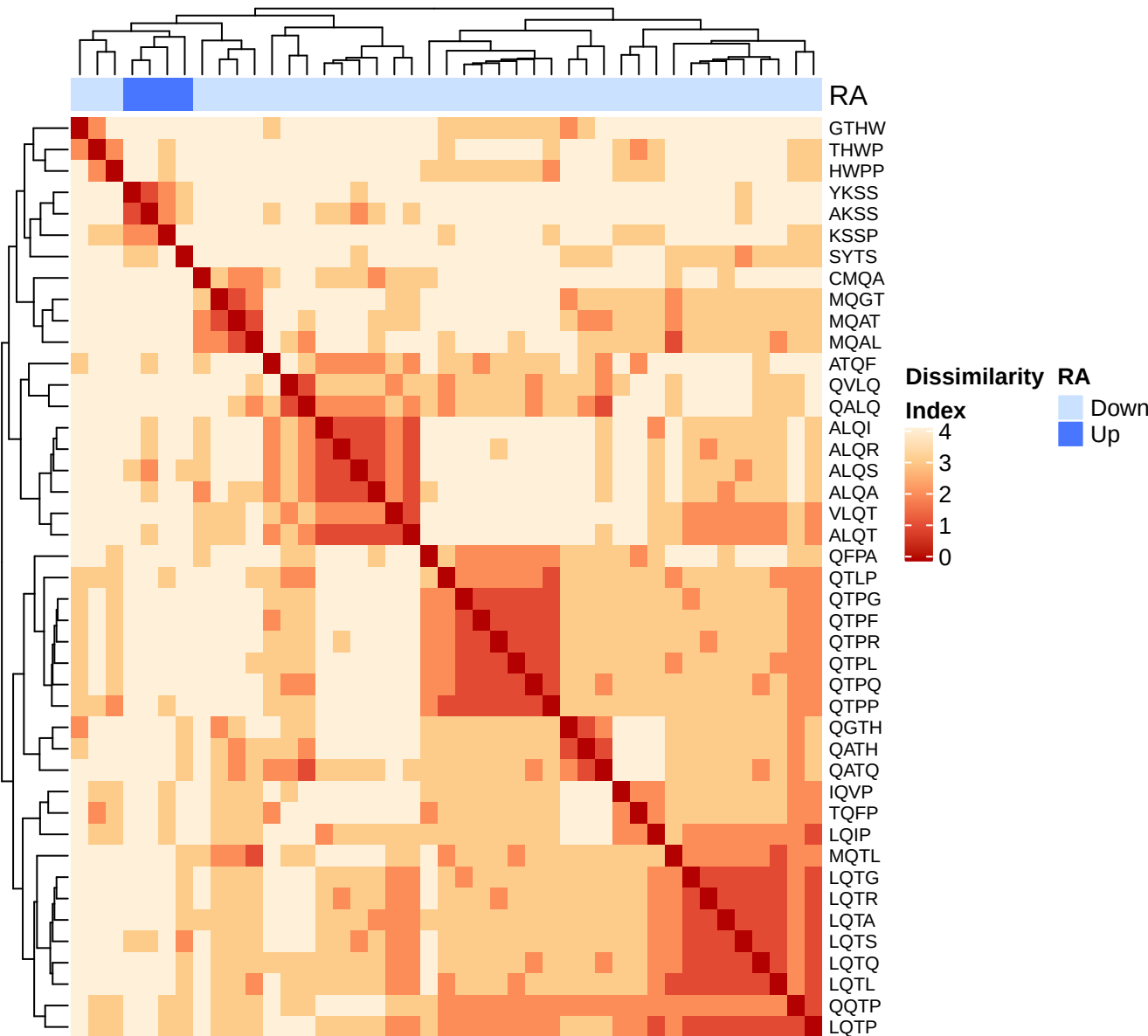
Kmers from IGK significant in Cont model



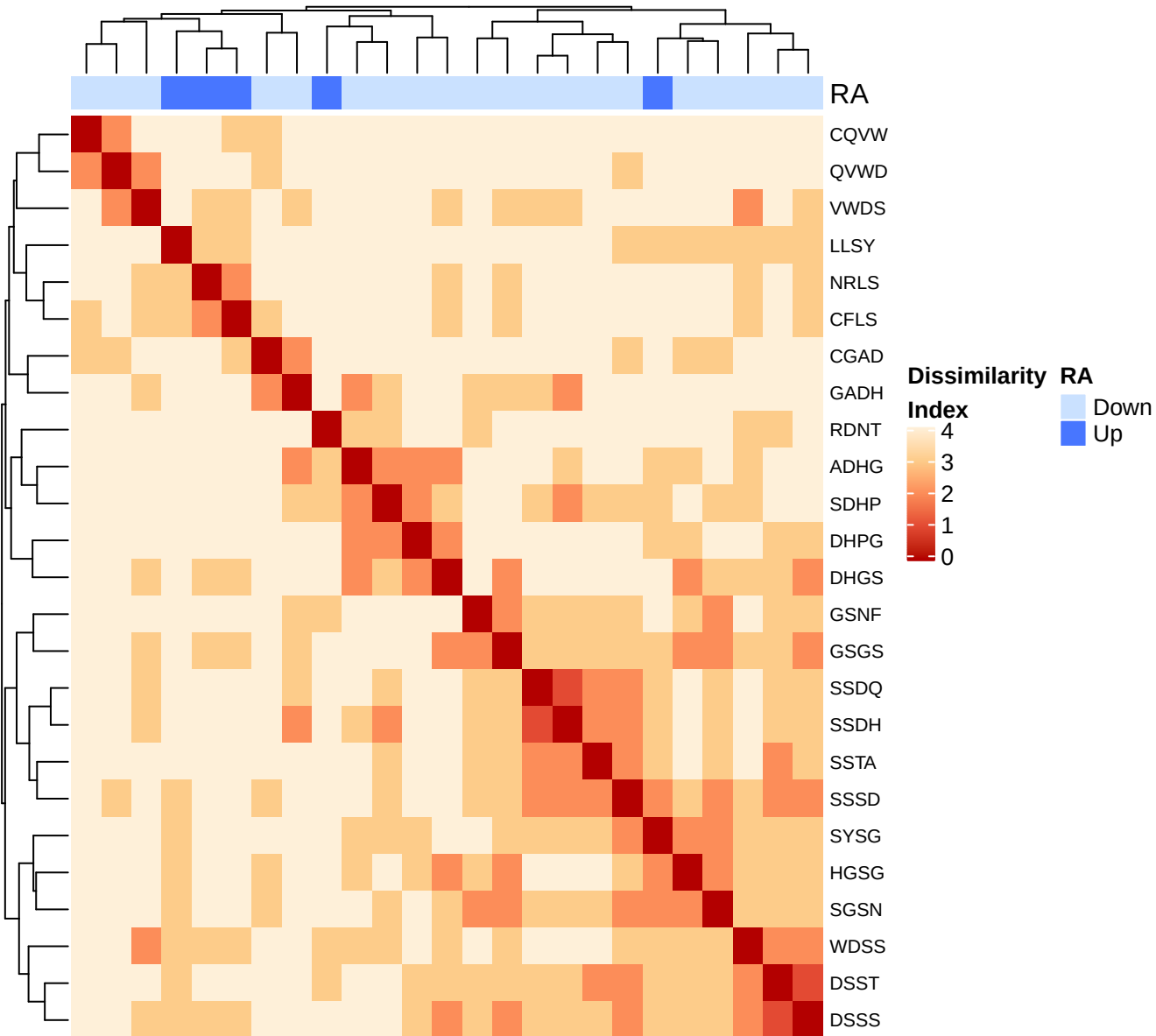
Kmers from IGK significant in Disc model



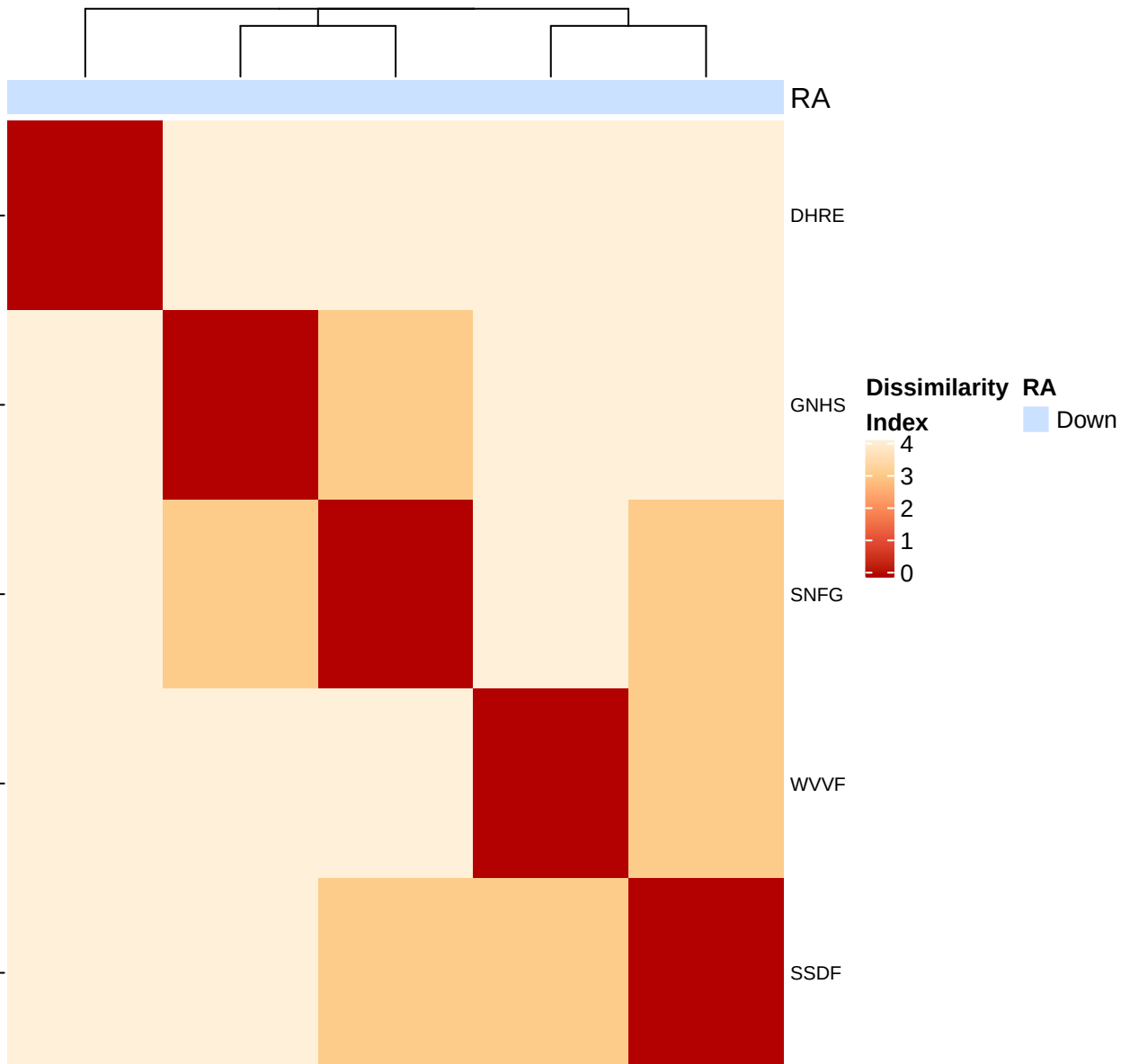
Kmers from IGK significant in Hurdle model



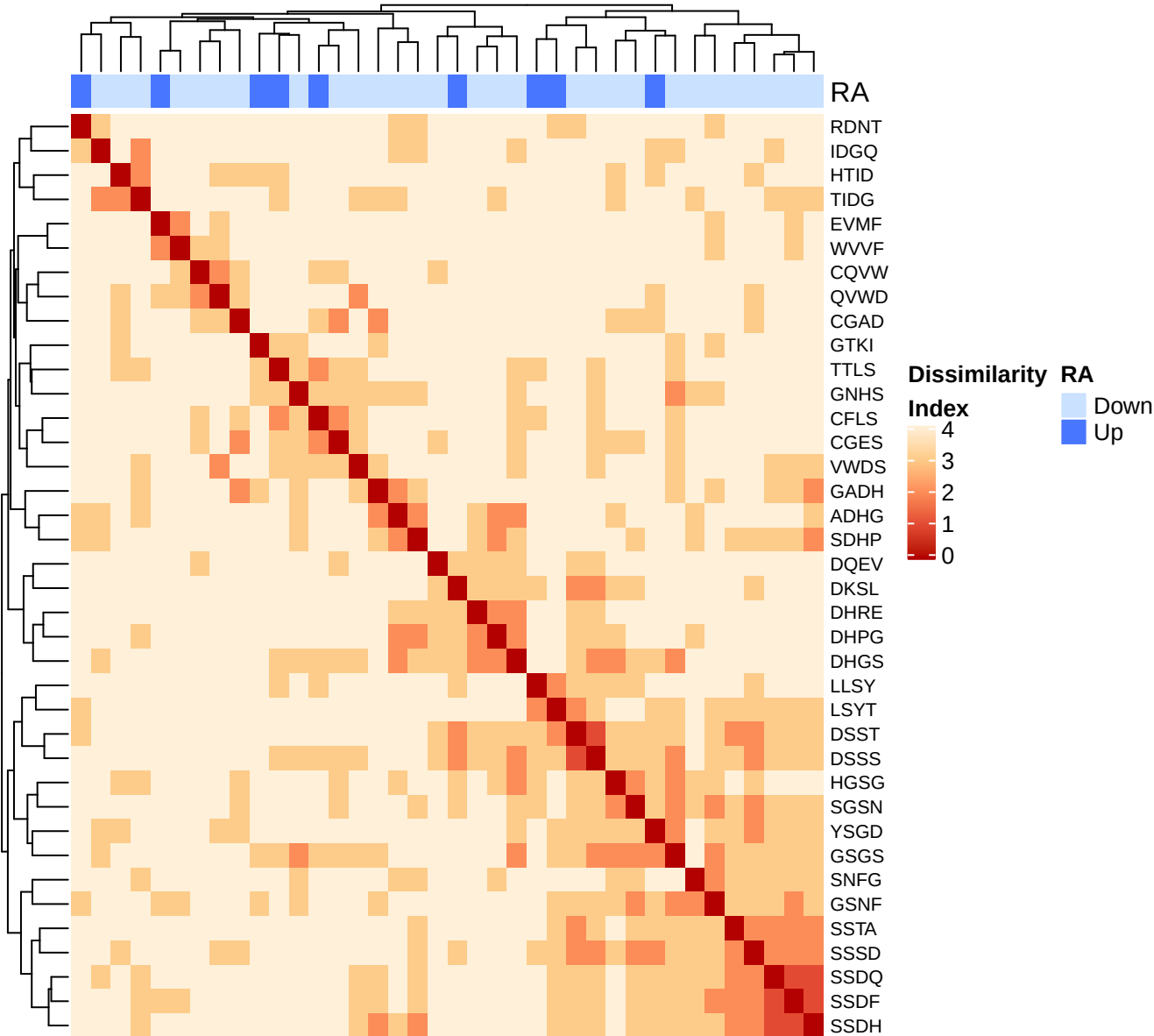
Kmers from IGL significant in Cont model



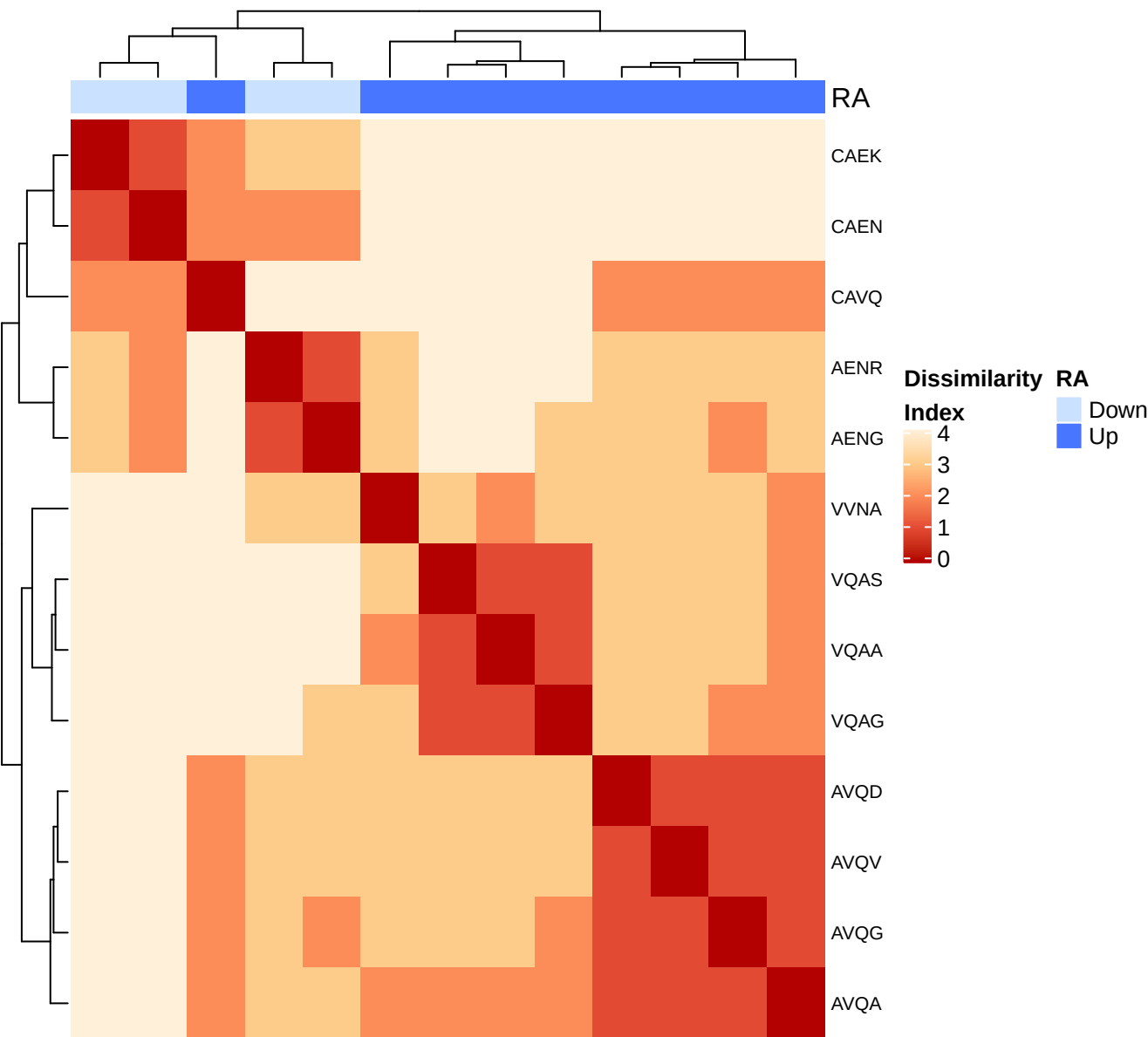
Kmers from IGL significant in Disc model



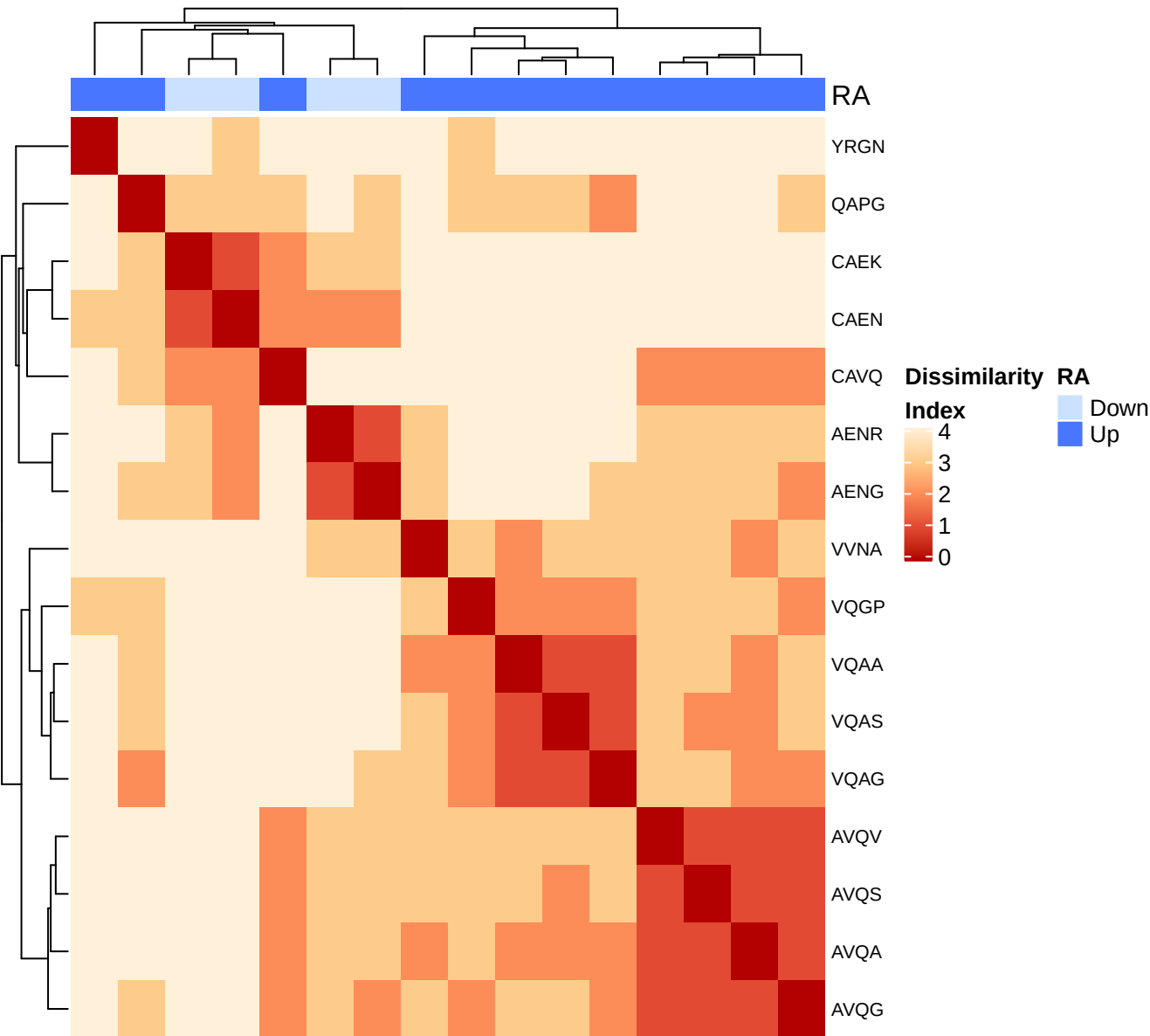
Kmers from IGL significant in Hurdle model



Kmers from TRA significant in Cont model



Kmers from TRA significant in Hurdle model



Kmers from TRB significant in Cont model

