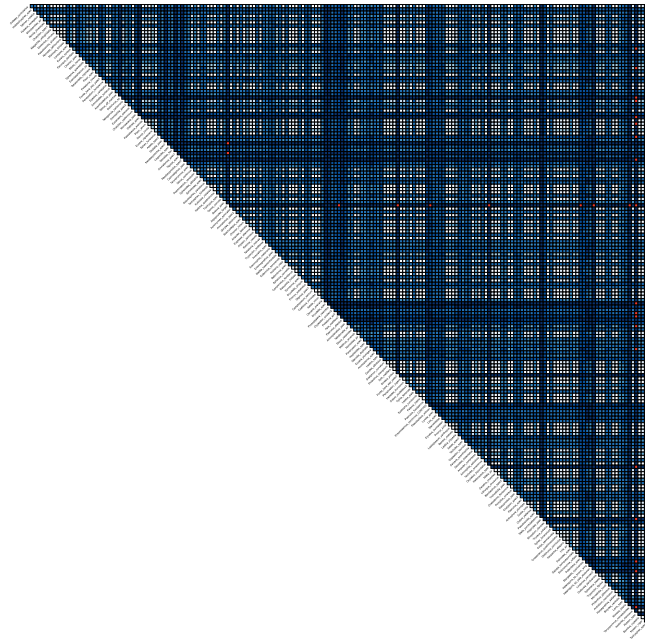
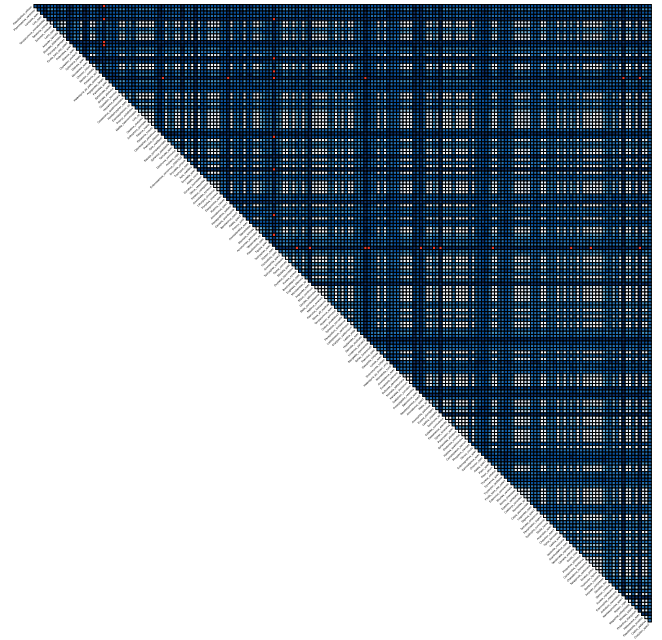


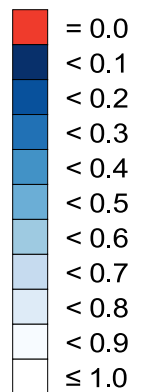
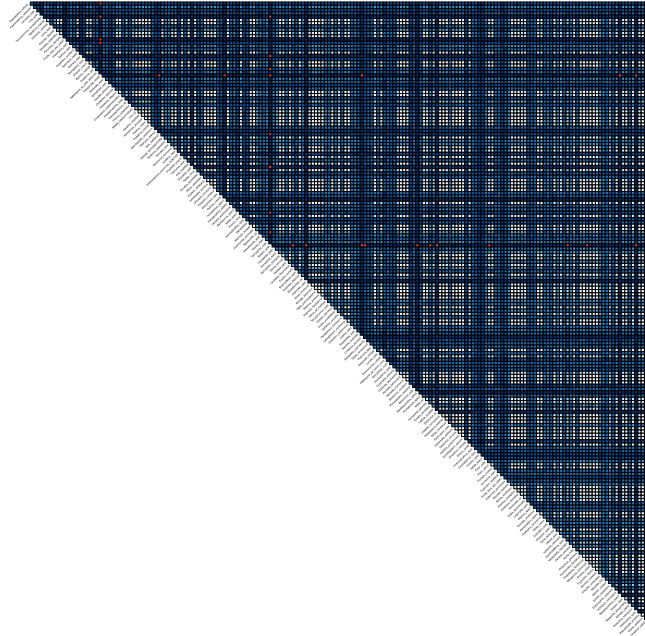
(a) sm-aa



(b) sm-nt12



(c) sm-nt123



**Figure S2:** AliStat heat maps of pairwise sequence comparison when analysing the datasets (a) sm-aa, (b) sm-nt12, and (c) sm-nt123. The graph shows the distribution of missing data, with dark blue meaning low shared site coverage. The overall completeness alignment score inferred with AliStat (Ca) is 52.3%, the maximum C-score for individual sequences (Cr\_max) is 97.2%, and the minimum C-score for individual sequences (Cr\_min) is 0.11%.