The optimal number of subpopulations I got using the cross validation error method was 7 (unsure if the stacked barplot is correct or not). Beale et al. (2021) found that most sublineages are from Nichols sublineage, but most of the samples of assignment4.xlsx seem to show that they are from SS14 sublineage.

Chart, bar chart

Description automatically generated