

Assignment 5, 1-2 cr

The data: *Streptococcus pneumoniae* bacterium, 7 gene sequences, concatenated into one sequence stretch (length 3182bp) from two population samples, Norway and USA, and from two time points, before vaccination and after vaccination. Vaccination can be expected to represent a selection pressure towards an organism.

The 7 genes are markers from the bacterial genome: the so called MLST-approach, multiple locus sequence typing.

Questions:

Are there differences in terms of nucleotide diversity, Tajima D, between populations (USA vs. Norway), analyse the two timepoints separately.

Are there differences between the time points?

This makes 1 cr.

Another 1 cr:

Pay additional attention (nucleotide diversity and Tajima D) to the different genes, i.e. can you notice (for example) one of them to be especially deviating between the populations, and/or between the time points.

The genes, and the nucleotides are:

aroE 1-405

gdh 406-864

gki 865-1347

recP 1348-1795

spi 1796-2266

xpt 2267-2751

ddl 2752-3192

By using the DnaSP-option, specify these sequence stretches.

It is useful to look at the differences (between populations, between timepoints), along the total sequence stretch, by using the sliding window option! Note that the default window is 100/25, but it is reasonable to use smaller window (why?).

Please, ask for more instructions if you don't know what to do 😊