

Practical on coalescence theory

You are interested in the size of two populations of Atlantic killer whales, one migrating towards a Northern geographical location and one towards a Southern location. These two populations share a recent common ancestor but their current population size is hypothesised to be different.



To test this hypothesis, you collected 10 (diploid) samples from the Northern population and 10 from the Southern population. For each sample, you obtained a genomic sequence of 50kbp. The data is stored in .csv files named `killer_whale_North.csv` and `killer_whale_South`. Each allele is encoded as 0 or 1 for the ancestral and derived state, respectively.

1. Estimate the effective population size for each population, using both the Watterson's and Tajima's estimator of "theta" assuming a mutation rate of 1×10^{-8} . Discuss the difference between values of "theta" using different estimators.
2. Calculate and plot the (unfolded) site frequency spectrum for each population and discuss it.