

COMPUTATIONAL POPULATION GENETICS — EXERCISE SHEET 3

1. Use software like Tracer to evaluate the results of the preliminary LAMARC runs from exercises 2 and 3 and decide which MCMC options may be appropriate for the full analysis. Start LAMARC runs with these parameters.
2. Proof that any distribution p that fulfills the reversibility condition for an irreducible aperiodic Markov chain X on a finite state space must be a stationary distribution of X .
3. (Mainly for Bioinformaticians) Write an efficient program that reads a tree in Newick format and simulates the evolution of DNA sequences along the tree according to the Jukes-Cantor model.
4. Analyze one or two of the datasets in `turtle.evolution:/home/dirk/Teach/WS1213/CompPopGen/DataSets`. The alignment programs `mafft`, `muscle` and `probcons` are installed on turtle. Read the pdfs contained in the same directories to decide which models for population structures could apply here. For next week, start with a preliminary study.
5. Simulate datasets for a population that consists of four sub-populations, with gene flow between the populations.
 - (a) How accurately can you estimate the rates and favored directions of migration between each pair of sub-populations and how does this depend on the number of loci and other properties of the dataset?
 - (b) Assume that each of the sub-populations has a substructure. How does this influence the results, especially if the sampling was limited to some of the sub-subpopulations. What if the sub-substructure is ignored or unknown?

Design the study and start the first preliminary test runs whose results we can discuss next week to decide about the program option settings for the rest of the study.