D. Metzler WS 2013/2014

COMPUTATIONAL POPULATION GENETICS — EXERCISE SHEET 2

1. Two populations of size N exchange on average m migrants per generation in each direction. Assume that you randomly sample two alleles of an autosomal locus from the same island, and let T be the time to their most recent common ancestor, measured in units of $2N$ generations (neglect recombination). Let T' be the time to the most recent common ancestor for the case that the two alleles were sampled from different islands. Determine the expectation values $\mathbb{E}T$ and $\mathbb{E}T'$ (and if you are really ambitious also the standard deviations σ_T and $\sigma_{T'}$) by mathematical derivation or by computer simulation (e.g. with Hudson's ms) for all combinations of $N \in \{1000, 10^5, 10^8\}$ and $m \in \{0.1, 10, 1000\}$.

2. Simulate a dataset with 10 independent autosomal loci of lenght 1000 bp, 20 alleles sampled from each of two populations of $N_e = 10.000$ individuals that exchange 1 indivual in each direction every 5 generations. Use a DNA mutation model with double hits and $\theta = 10$ per locus. How accurately can LAMARC estimate θ and the migration rate? Explore also how long you should run LAMARC.

3. Simulate datasets of DNA loci sampled from three populations that exchange migrants at certain rates. Use pairwise migration rates lower than 1 migrant per generation. Explore how the accuracy of LAMARC estimates for the migration rates depends on the sample sizes, the number of available independent loci (assuming $\theta \approx 10$ per locus) and the LAMARC runtime. Use a mutation model that allows for double hits.

Due to the runtime of LAMARC, it may not be possible to present final results for exercises 2 and 3 within a week. However, you should perform preliminary test runs to estimate with which options and how long you need to run LAMARC, and start these final runs. Be prepared to present your approach and your initial findings next week.