Tohorā Cultural Shock code: reasoning behind each choice

By Marina Klemm, Casual Research Assistant

Contact: [marinaklemm@gmail.com](mailto:marinaklemm@gmail.com)

**• Bottleneck:**

Choice between models with: exponential growth, instantchange or event (resize).

Still need work: the popSize is not changing in the output even though it should be coded through the model I implemented; this is true for whatever type of model I chose.

**• Trajectory:**

beginFreq: attempted to use the mtDNA\_1.txt file with the initial frequencies, but it appears that I cannot use allele frequency = 0 (ValueError: Invalid frequency range 0.100000 - 0.000000).

For that reason, I need to use a pair of alleles, such as the SNP file.

I’m waiting for Bo Peng’s answer, but I cannot seem to be able to add a list of values to the simulation. For that reason, I chose two random values for this simulation.

endFreq: since trajectory requires a migrate\_to field that will separate the population in at least two subpopulations, but we still have a continuous geneflow between the two subpopulations described here (Auckland Islands and Campbell Island), we decided to add two subpopulations frequencies with the same value. This value was chosen randomly for this code to run, but it will then be adapted to Emma Carroll’s contemporary data.

PS: trajectory seems to be the most finicky part of this code. There are many ways for it to break:

(1) At first, I thought it was limited by my notebook’s computing capabilities;

(2) Then, I assumed that the bottleneck was so aggressive that it affected the population in a way that did not allow the code to run;

(3) Now, I think it’s just a matter of the allele frequencies not being able to fit the endFreq based on the beginFreq. This is because I am now able to run the trajectory starting with 30000 individuals, for 1000 generations, with no problem. The trick was to change beginFreq and endFreq until they were accepted. Therefore, this might break again when we add Emma’s contemporary frequencies, and so beginFreq will have to be readjusted.

**• Nitrogen and Carbon isotopic values:**

Initialized from the original code, using mean, variance and deviant proportion (10%) for N and C separately. Seem to be properly coded in each printed out population.

**• Haplotype frequencies:**

The last chromosome accounts for the different mitochondrial haplotypes; note that the last loci is always zero, since it is, by definition, haploid data. Regardless of the simulation done, it seems that there are 57 haplotypes in the final population.

I was able to print out a final freqmtDNA.txt file that contains 56 frequencies that add up to 1 at the end, which seems like the right output.

Still need work: how to print out the frequencies from both subpopulations, such was the freq.txt file for the other alleles.

• **Mating Scheme**

Within the pop.evolve function, a few mating schemes were added, and each one requires a weight to be added to them. From the [simuPOP manual](https://simupop.sourceforge.net/manual_release/build/userGuide_ch6_sec1.html?highlight=weight), the negative weights are processed before the positive, which justifies the choice of weight = -1 for the clone mating (since it refers to the mitochondrial haplotype).

In that same chunk of code, 6 virtual subpopulations were chosen out of the 8 present (adult male, juvenile female, etc) to account for the two subpopulations that were deceased and, therefore, not included.

**Note:**

Juhana Kammonen kindly sent me the code that his group used to simulate the Finnish human population evolution, which could have been useful to simulate its bottleneck. He asked us to cite his most recent work in a possible future publication using simuPOP. Reference:

Sundell et al (2012) - Retracing Prehistoric Population Events in Finland Using Simulation. *CAA2012 Proceedings of the 40th Conference in Computer Applications and Quantitative Methods in Archaeology, Southampton, United Kingdom, 26-30 March 2012.*

However, their simulation hard coded their bottleneck with an external .txt that had sample sizes, mutation rates, migration rates, among other mathematical functions, that were not super clear to me. Therefore, I did not end up using their code, although I appreciate their willingness to help us.