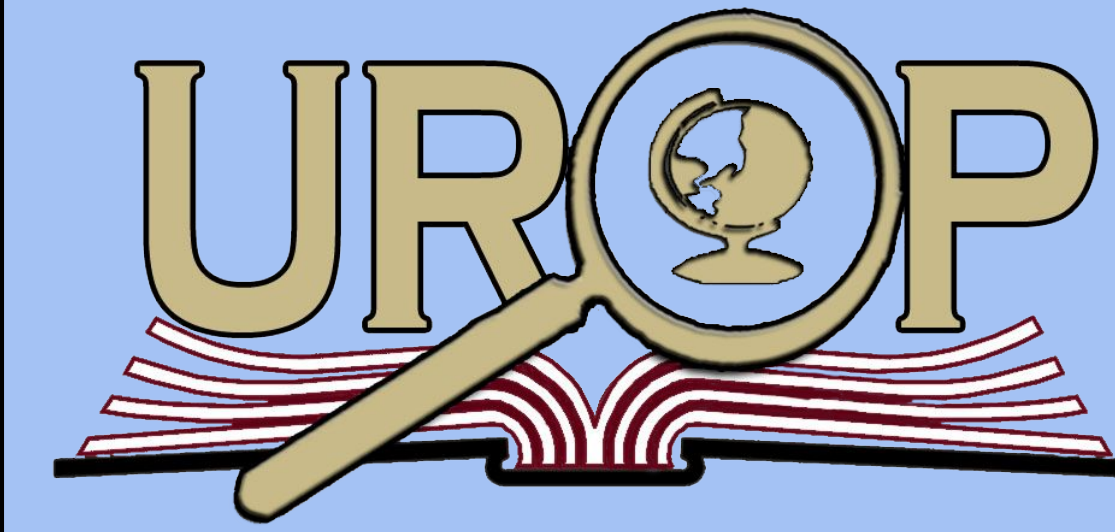




Estimating environmental heterogeneity in two populations exchanging migrants



Pacey Cheesebrough; Mentor: Dr. Peter Beerli

Introduction & Background

Coalescent Theory, primarily attributed John Kingman in 1982, is the process of tracing two alleles backwards to find the point of convergence, or most recent common ancestor. Initially, genetic polymorphism data was implemented via the fixation index (F_{ST}) to explore what made a population genetically different. Eventually, technology, namely the computer program MIGRATE, replaced (F_{ST}) with its ability to account for more parameters of a population, such as mutation rate or asymmetrical flow of genes (Beerli, 2001).

Recent research from 1999 onwards has used MIGRATE to explore how heterogeneity might affect a population's capacity to produce offspring. One study discovered that as an environment becomes increasingly variable, the time to the most recent common ancestor between two individuals becomes more recent (Mashayekhi & Beerli, 2019).

Many previous studies utilizing the coalescent approach have resulted in imprecise estimations of migration parameters and notable biases in the datasets, inflating results (Beerli, 1999). Further, previously unaccounted-for factors like life-history strategies, how organisms survive and how they have evolved, change how much of an effect environmental heterogeneity has on a population (Mashayekhi & Beerli, 2019). Hence, there is a need to rework the heterogeneity parameter used in previous simulations.

Now, this project, "Estimating environmental heterogeneity in two populations exchanging migrants", aims to use MIGRATE to simulate two geographically and genetically different populations with the intent of detecting heterogeneity between them with reworked parameters. We are ultimately observing whether or not this MIGRATE program can achieve its purpose of identifying heterogeneity.

Methods

- 1) This project begins by utilizing the program python simtree.py. We have 5 parameters used in this code: file name (-f), the number of samples in each population (-i), population size for each population (-t or θ), immigration rate into each population (-m or μ), and the heterogeneity parameter (-a or α) [1]. We use these values above to create a DNA dataset. The next step is to submit the dataset to the FSU Research Computing Center system where it is analyzed using Dr. Beerli's program MIGRATE. Once the analysis is complete, it will output an "outfile" with the new, altered parameters.
- 2) Next, this "outfile" will be divided into smaller files for each of the parameters. Here, the averages and standard deviations of the theta, alpha, and mutation rate values are obtained via a python script, averaged over the 100 loci analyzed.
- 3) Finally, another program is used to plot the true values of the parameters against the estimated values obtained from the previous step. Since each run yields two values of each parameter (α , θ , & μ), there will be two lines for the two values.

This research project is focusing on two phylogenetic trees representing two populations, both generated via computer simulation. The populations are of an unknown species as their identity is irrelevant and hypothetical. It is simulated such that these two populations are separated from one another but not without migration (-m or μ). The population size (-t or θ) will remain the same for both. However, the heterogeneity parameter (-a or α) will be altered to emulate environmental differences.

For the intents of this project, the computer program MIGRATE was utilized alongside Python scripts to test the program's ability to detect heterogeneity. Many existing mathematical formulas were used to design the code used in this simulation, including the Wright-Fischer Model and the Time to Coalescence equations.

[1] Heterogeneity is referring to the genetic differences between the two populations being examined, an amalgamation of mutation rate, environmental influences, etc.

Results

- During the completion of this project, an error involving the program MIGRATE. The program is meant to analyze the 6 parameters mentioned in the Introduction, however, due to a fault in the code, fails to recognize and analyze the alpha values for both populations.
- For this reason, the heterogeneity of the populations cannot be determined.

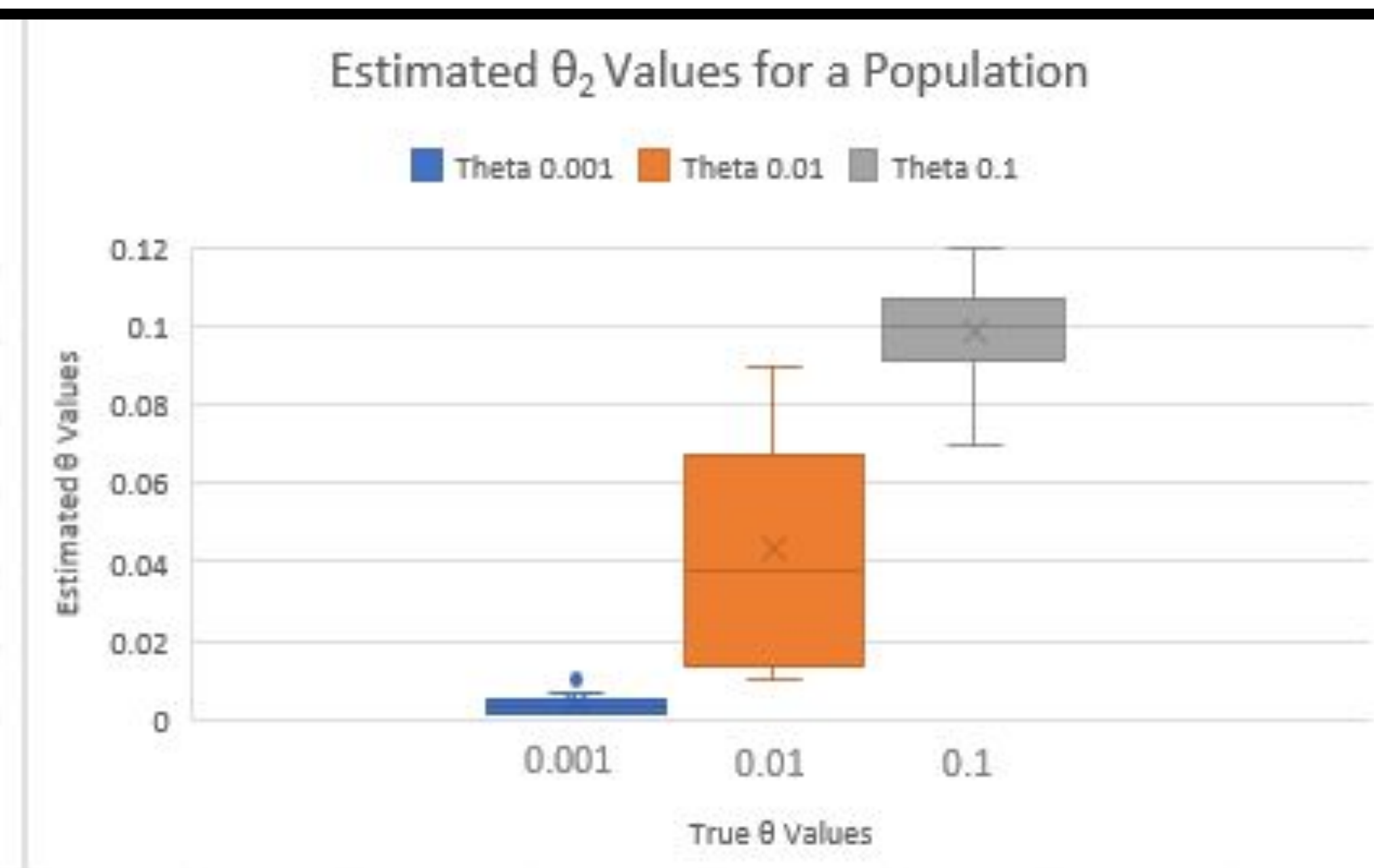
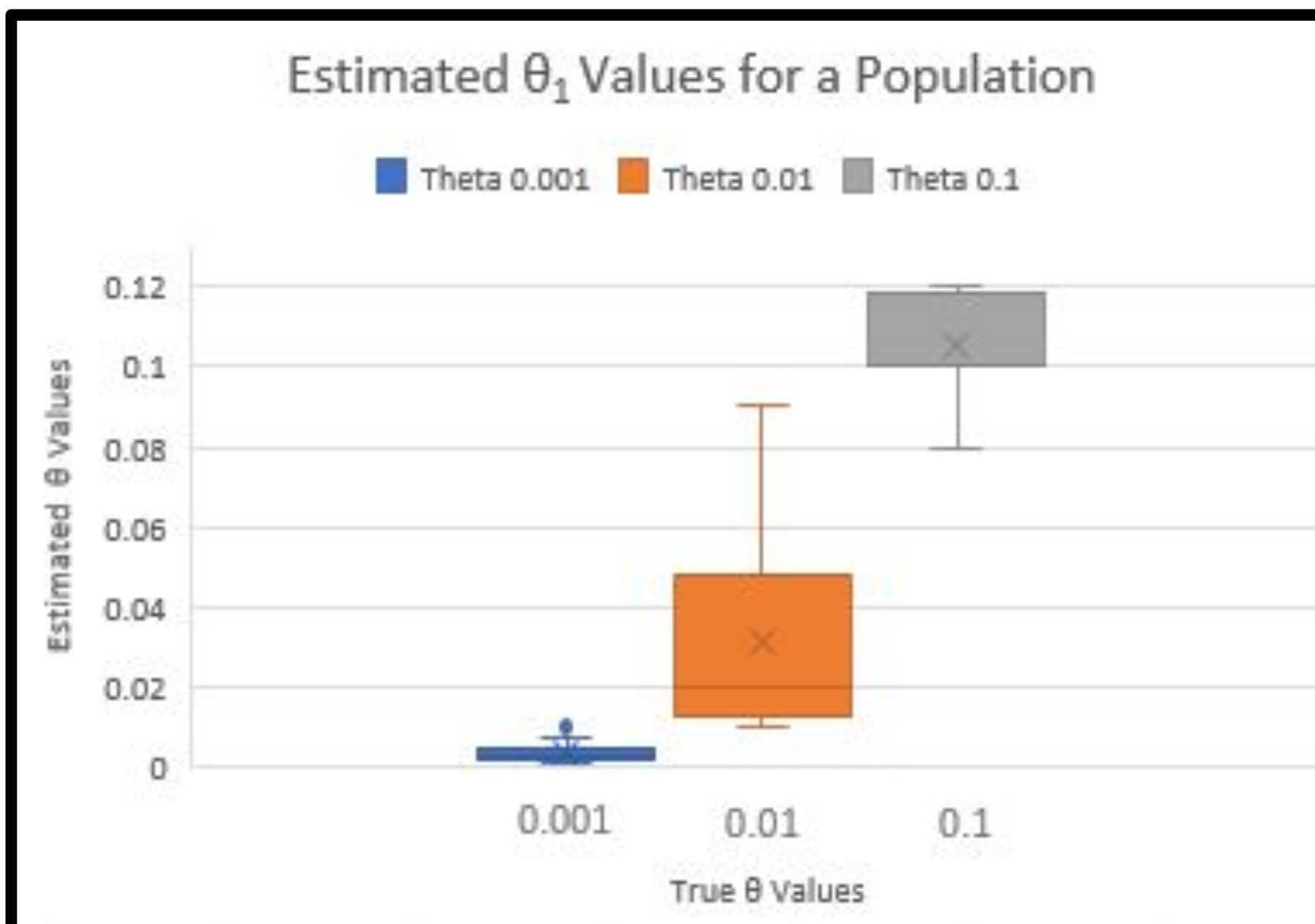
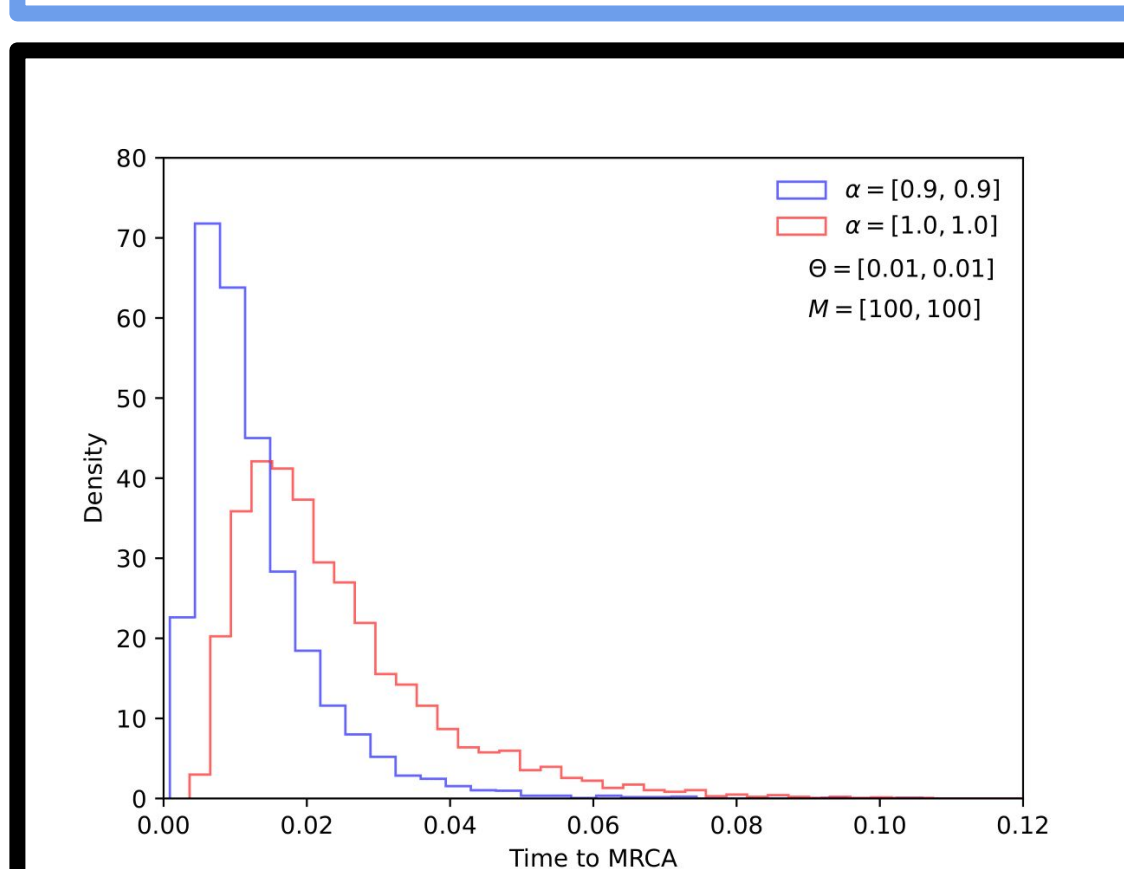
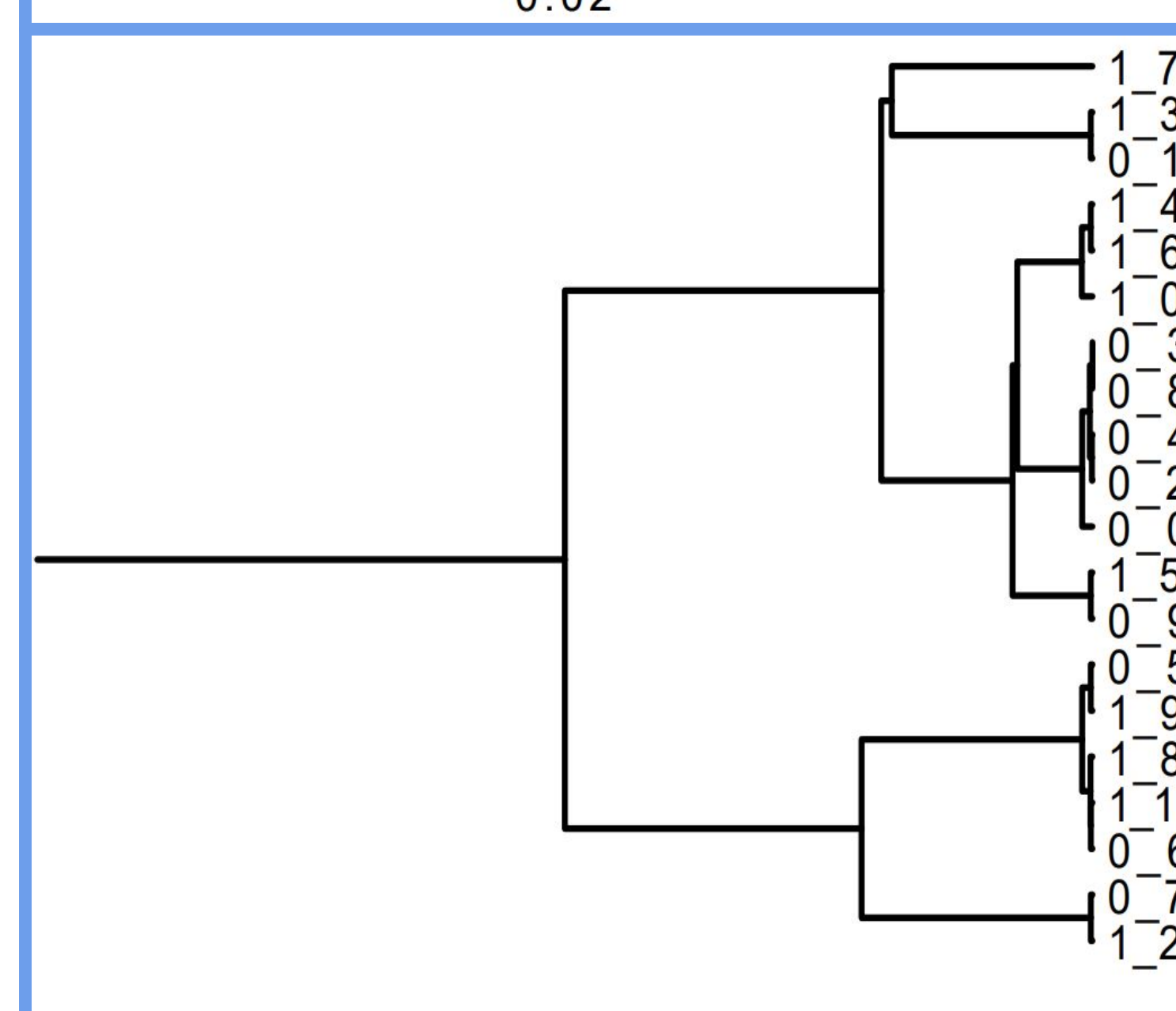
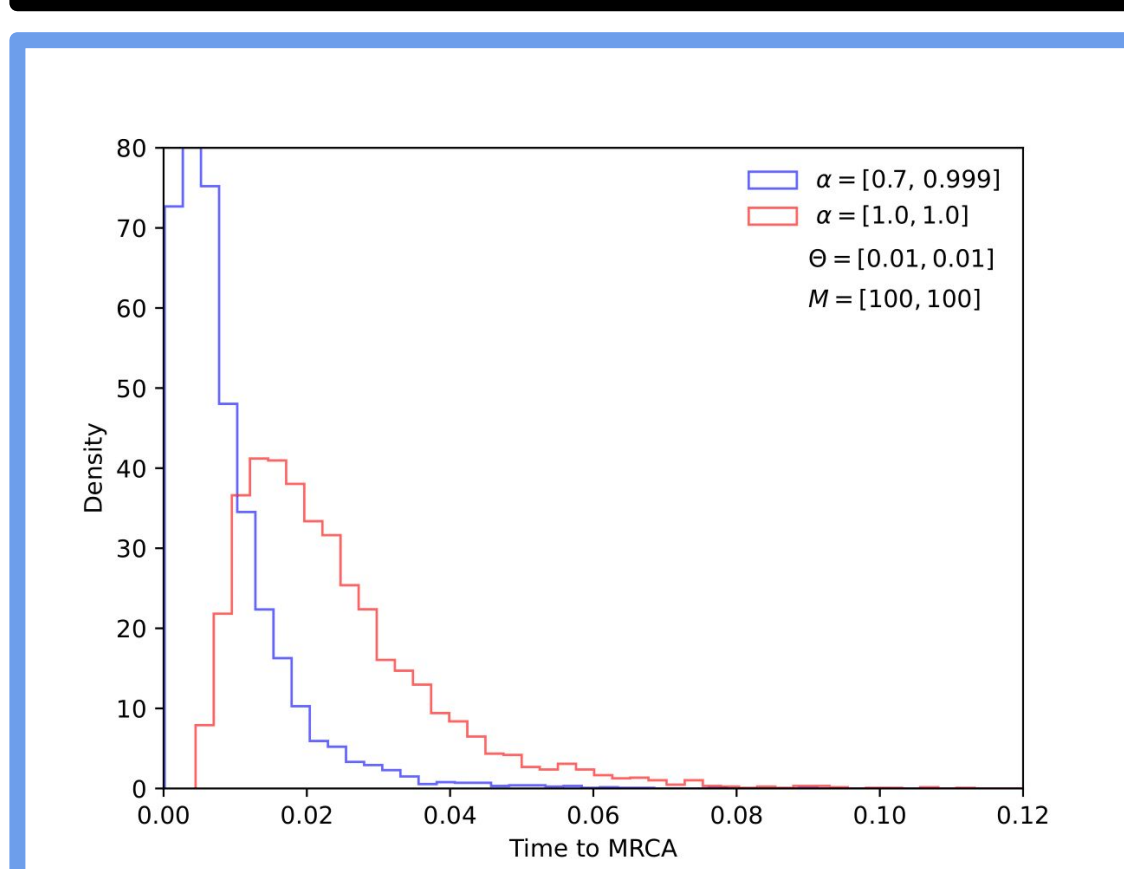
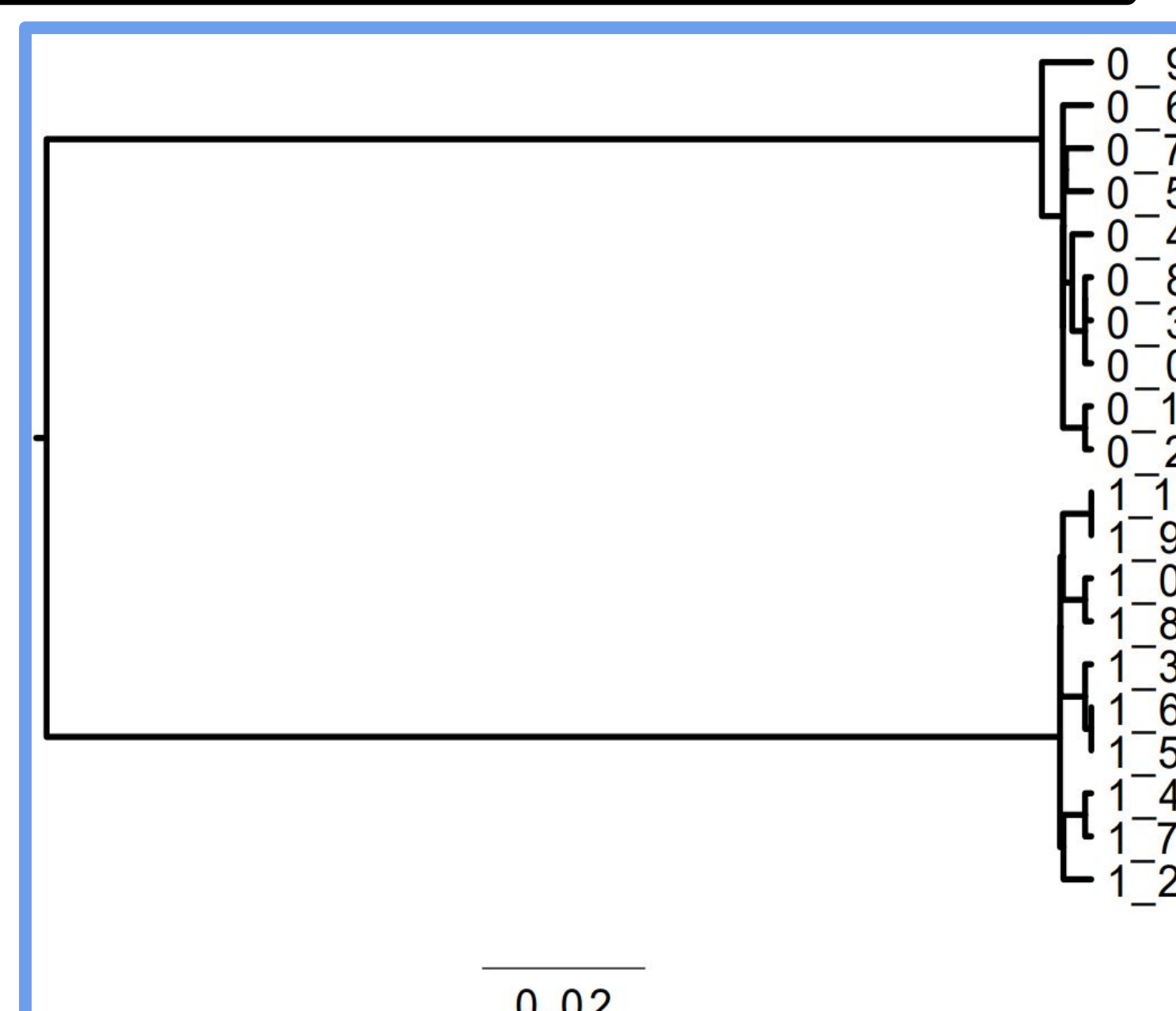
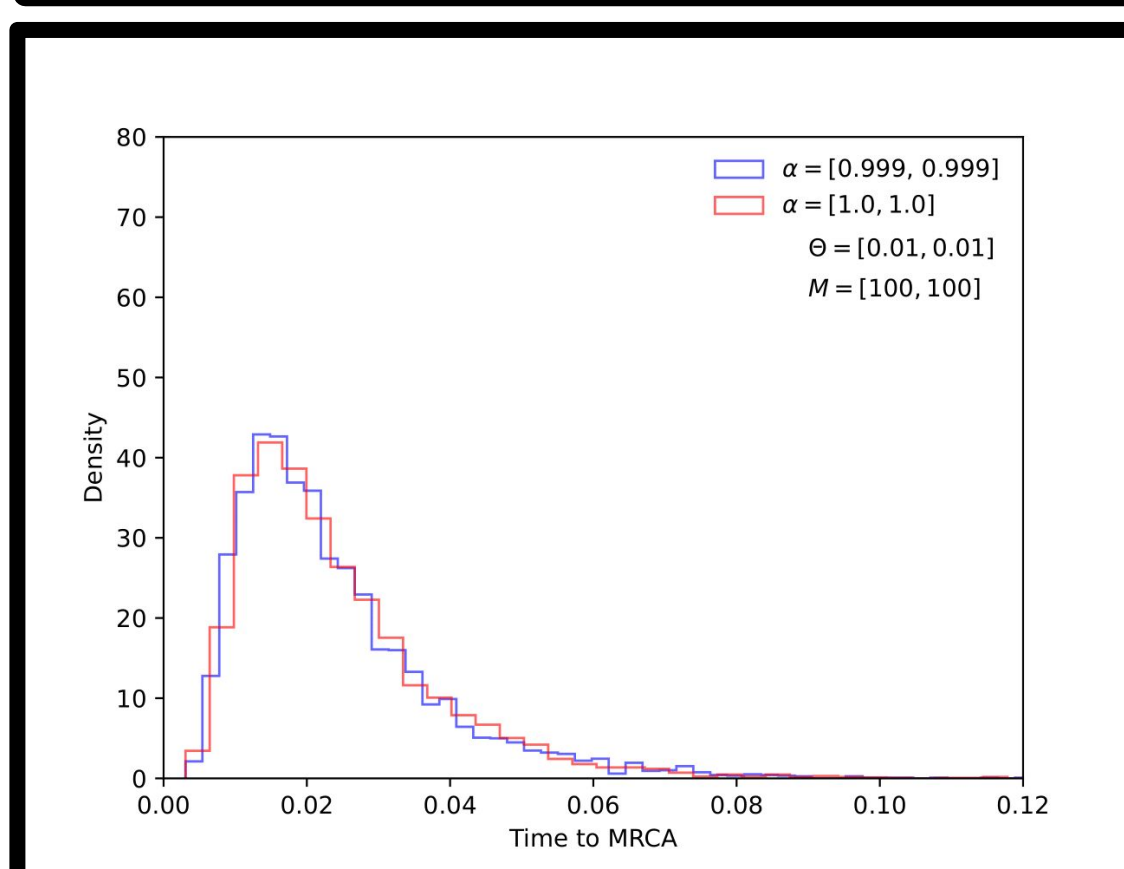
Discussion

Significance of Results

- Prior research indicates that as the α value for one population decreases from 1 to 0.5, the time to most recent common ancestor (TMRCA) becomes more recent.
- By this, this project hypothesized that the expected parameters for α outputted by MIGRATE would be relatively close to the true parameters known from previous studies. However, this was not true.
- Other literature suggest that this application of MIGRATE is possible with fixation of the code.

Further Applications

- i is the number of individuals in each population. This value could be altered to observe what changes occur in the other parameters when i proceeds to ∞ or 1.
- Other variables may also be altered, such as mutation rate, μ , can be altered to observe differences in heterogeneity.



The figures (left) depict theoretical distributions for the Theta values for two populations being analyzed. We would expect the whisker plots to display a relative, positive linear correlation, as depicted.

References

- Beerli, P., & Felsenstein, J. (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences*, 98(8), 4563-4568.
- Mashayekhi S. & Beerli, P. (2019) Fractional coalescent. *Proceedings of the National Academy of Sciences*, 116(13), 6244-6249.
- Beerli, P. & Felsenstein J. (1999) Maximum Likelihood Estimation of Migration Rates and Effective Population Numbers in Two Populations Using a Coalescent Approach. *Genetics*, 152(2), 763-773.