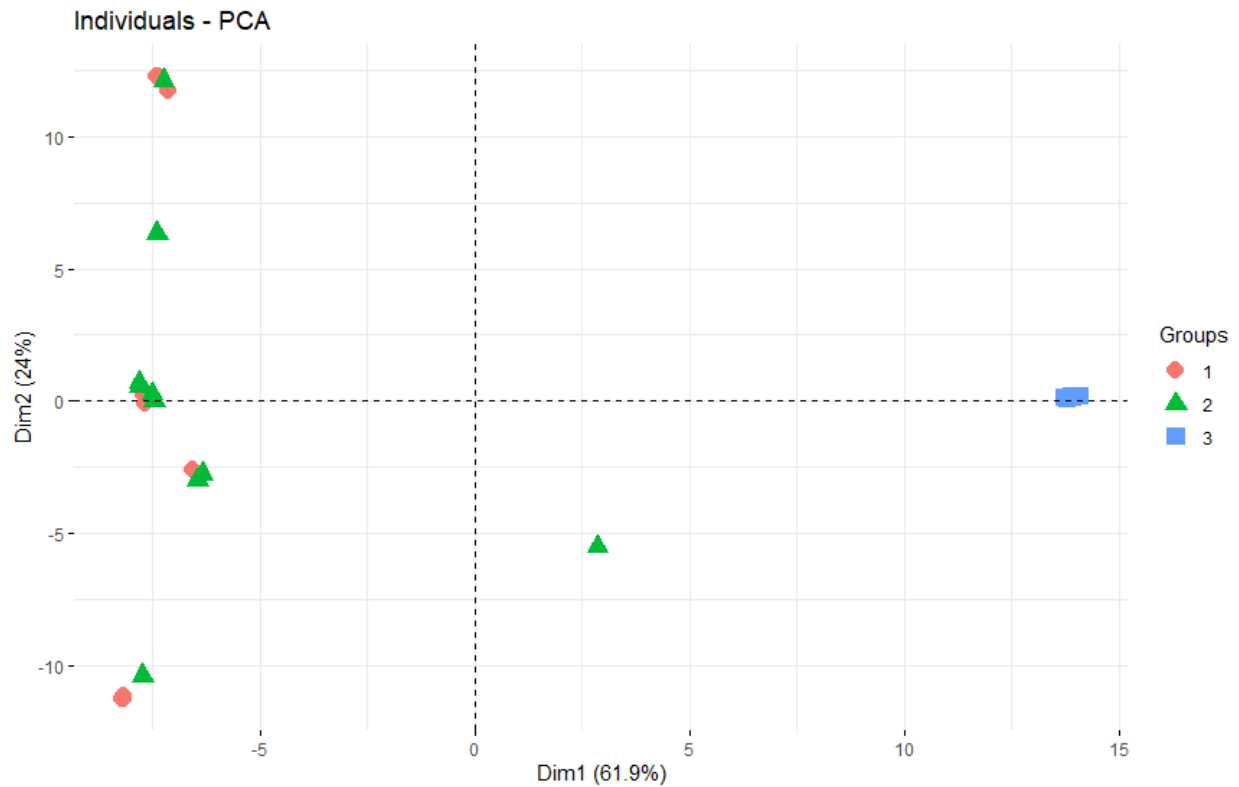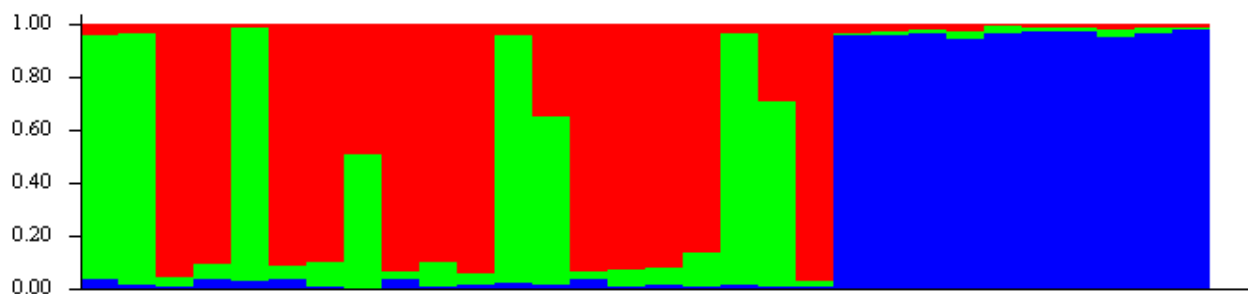# Three Pop Migration Analysis

Individuals - PCA

Based on the PCA results above, it seems Population 1 and Population 2 share a lot of mutations, indicating high migration between them. Population 3 on the other hand, is very isolated.

For the structure report, K = 3 had the highest Ln P(D), which make sense given we know there are three populations.


The bar plot from structure for K=3 reflects what the PCA graph demonstrated, that Populations 1 & 2 had high migration rates between them while Population 3 was mostly isolated.

Custom Slim Simulation

Individuals - PCA
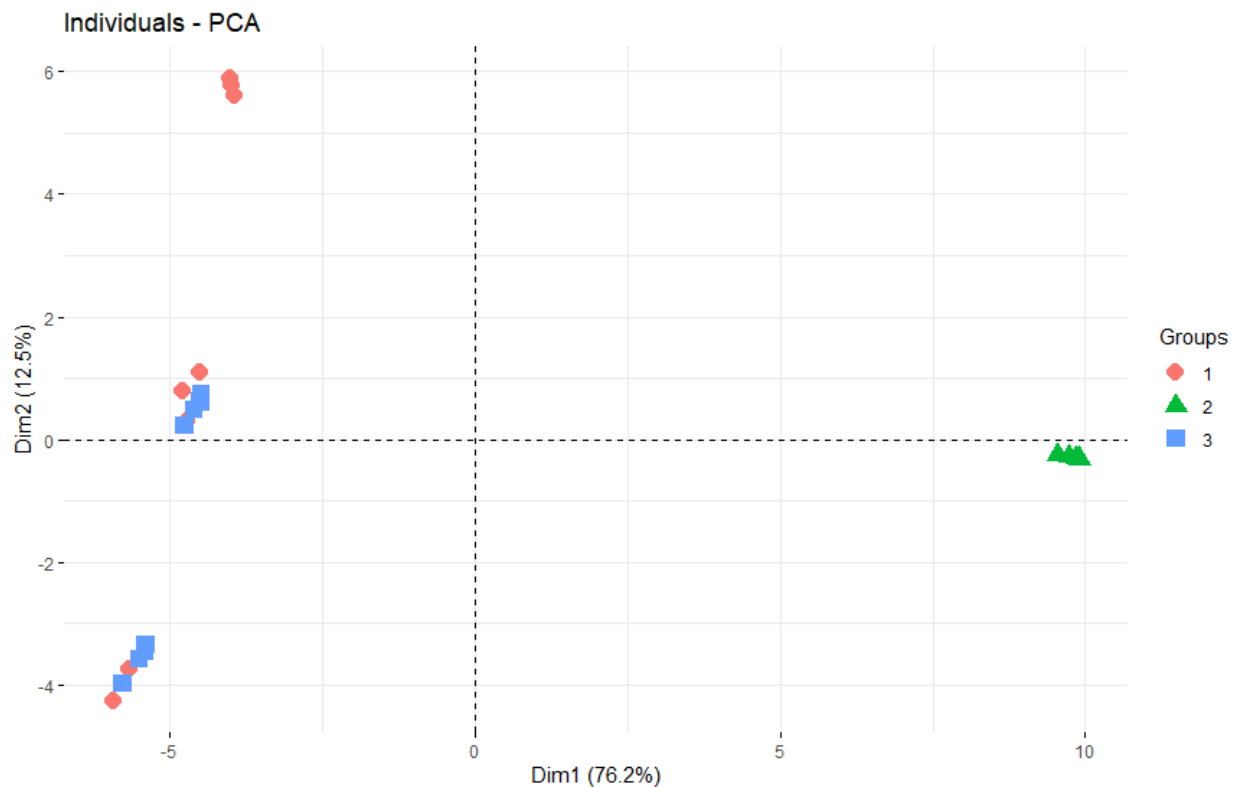
Based on the PCA results, Populations 1 and 3 are very similar and therefore had a lot of migration between the two populations, which is true to the actual simulation. Population 2 is isolated, which is also true to the actual simulation.

For the structure report, K = 3 had the highest Ln P(D), which is the correct amount of populations in the simulation.
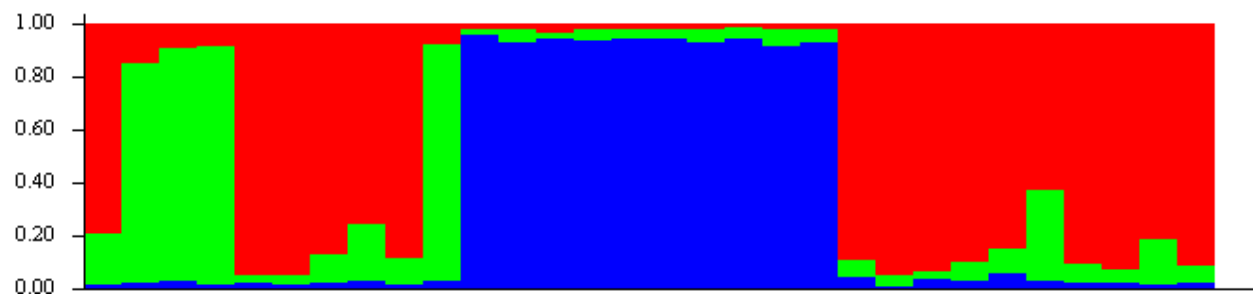

The bar plot from the structure with K=3 once again reflects the PCA graph, that Populations 1 & 3 showed high migration rates and Population 2 was isolated. However, there seems to be more migration from Population 3 to Population 1, although it should be the other way around. One possible explanation is that the program didn't quite interpret the data right given that there was migration between both populations just weaker in one way.

Mystery Data

Mystery 1

For the first mystery data set, the structure report indicated that K=3 had the highest Ln P(D), which is correct, given we know we have samples from 3 populations.
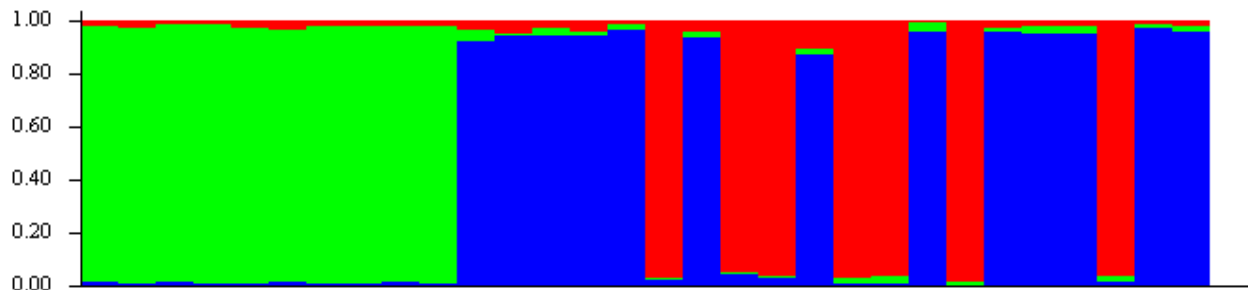
For the structure bar plot for K=3, it appears that Population 1 was isolated, while Populations 2 & 3 had migration between the two. Population 2 either seemed to have a bigger population size or had more migration to Population 3 than the other way around, given the large chunk of Population 2 that seemed to be isolated from Population 3.

Mystery 2

For the second mystery data set, the structure report indicated once again that K=2 had the highest Ln P(D), which is not correct in this case given there are only samples from 3 populations. I ran two additional parameter sets with a burn in rate of 3 and 5 and 5000 and 10000 mcmc reps, but both new runs still reported that K=2 had the highest Ln P(D).
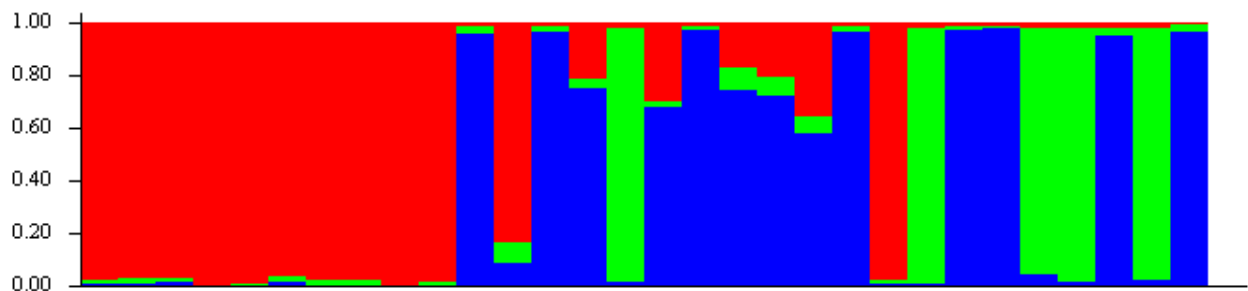

Looking at the K=3 bar plot for the original parameter set, it appears Population 1 was isolated from Populations 2 and 3, although it seems Population 2 did have Migration from Population 1. Population 2 and 3 seemed to have migration between the two populations, although there seemed to be a greater migration rate from Population 2 to Population 3 than the other way around. It could also indicate that the size of Population 2 was bigger than Population 3.

Mystery 3

In this population sample data set, we don't know the exact number of populations that are included in the sample. Once again, Ln P(D) was lowest for P=5, but since this was incorrect for the last mystery data set, I'm not inclined to necessarily believe it's correct. So, I created another parameter set with a burn in rate of 12 and 4000 MCMC iterations and ran a run with K from 2 to 10, since K could be larger than 5. This also reported that Ln P(D) was lowest for two. So, I decided to run one more test with K from 1 to 5

for the original parameter set. With this new run, Ln P(D) was highest for K=1, leading to the possibility that there was only one population.
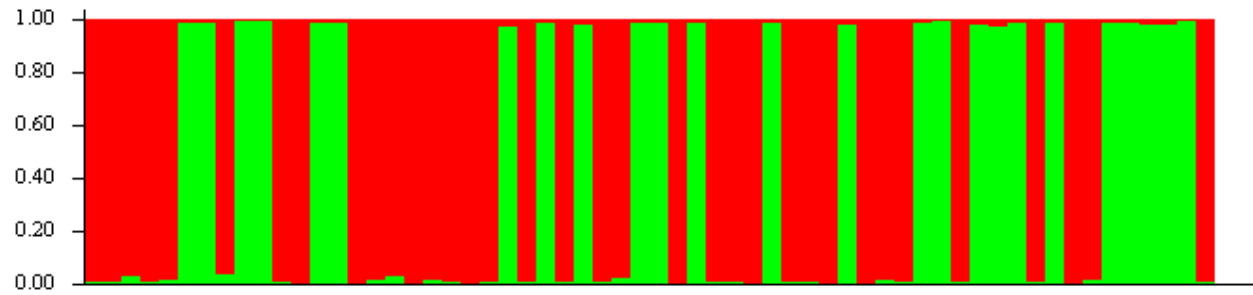
I'll assume K=2, since K=1 can't really be analyzed. Looking at the structure bar graph for K=2, it seems most individuals contain characteristics most common to only one population, suggesting there was little migration between the two populations. The number of bars of each color also seems to be roughly the same, suggesting equal population sizes, or that the number of individuals sampled from each population was the same regardless of the population size, which may be more likely given past data.