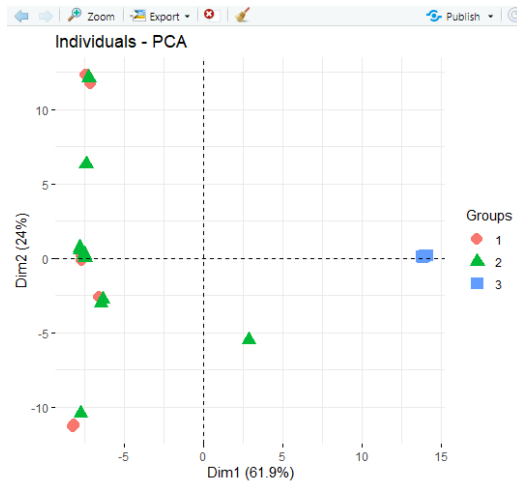
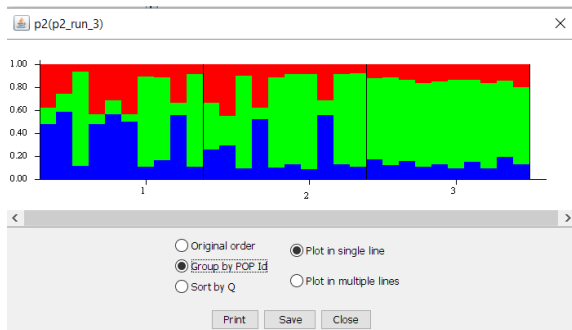


Computational Genetics: Structure Assignment

Section 6.1:



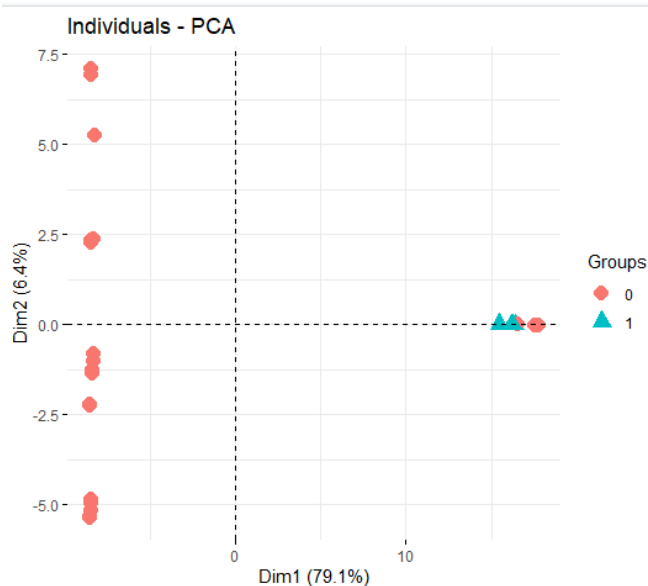
The PCA plot shows that there are clearly 3 sub populations. The green triangle population (2) overlaps with the red circle population (1) which signifies that there is some migration between population 2 and population one. It is unclear as to the extent of the migration but it's clear that population 3 has almost to no migration between the other two populations. It is a wild guess but since 4/6 green triangles (population 2) intersect with population 1, I would assume for the migration rate to be pretty high between the two. But I do not want to fall in and over interpret the results.



The Structure plot with the highest $\ln P(d)$ was $L=2$ which goes to the opposite of the Slim and pca assumption results. I believe there is something within Structure that I may have messed up. So I went back and did it again, with more 5 more runs for each K value. Now, I had on average the highest $\ln P(d)$ value and the largest $\ln P(d)$ value for $L=3$ which finally goes along with what was proposed in the PCA plot.

Surprisingly, the third population has some affect of migration as a 10% of color can be seen. This is what was actually was written in the Slim output but not shown in the PCA plot.

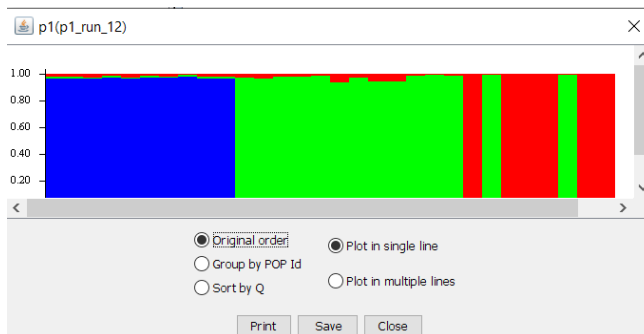
Section 6.2:



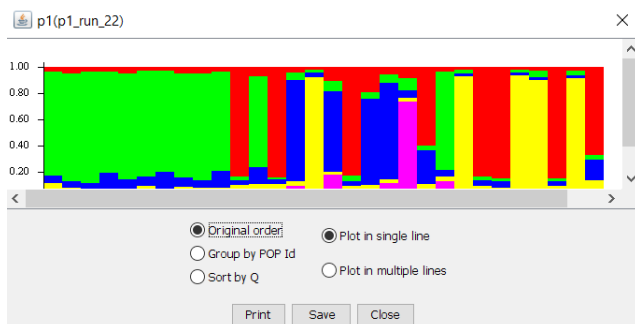
This model didn't depict what the Slim model was supposed to have, and I tried for a while to figure out where I made an error. What should have happened in my model is for the first population to not have migrated but have the second and third population migrate to the other populations. Based on this PCA plot, the population one has not migrated much compared to the first population. The Structure plot for this model was also completely off, but was very much related to this PCA model. The highest Ln Pd number indicated a subpopulation of $K=2$, but it should have been $K=3$ by my Slim model. There was also some indication of migration of the first population,

but absolutely nothing from the other population.

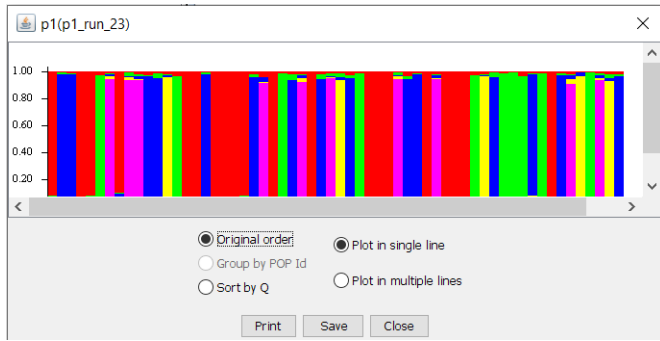
Section 6.3:



I did a range of $K=1$ to $K=6$ to make sure I wasn't missing any possibilities. The highest Ln Pd negative number was for $K=3$. Looking at the bar plot, I can tell that the first population did not merge at all with the second nor third population. The second population merged with the third by a large factor and that third population migrated into the first and third population by a little factor.



For Mystery 2, I also did a range of $K=1$ to $K=6$ to make sure I was not missing any possibilities. Surprisingly, I had a high Pd negative value for $K=5$. Looking at the bar plot, there is a large green area indicating that a large portion of a population did not migrate into another population. The other populations were very much merged into one another especially the red population.



to occur.

For Mystery 3, I did a range of $K = 1$, to $K = 6$ and found that the highest $\ln P_d$ value aligned with $K = 5$. Based on the bar plot, it looks that the 5 different populations migrated continuously between the populations. If anything the red population looks to have fully migrated throughout the other 4 populations. This probably means that the five populations must have been close to each other for this migration