

Effect of Mutation Rate on Divergence

James O'Reilly

Student Number: r0773125

The goal is to understand the neutral population structure in stickleback. The data contains loci from four populations. The first step is to use the STRUCTURE program to infer the number of clusters present in the data. Figure 1 shows that the data is best split into two clusters. When several K values have similar $\ln(P(X|K))$ estimates it can be inferred with reasonable certainty that the smallest is the most appropriate for the data. One should choose the lowest value of K that captures the maximum degree of structure detected in the data. In our case $K = 2$ is clearly the most appropriate choice, as is clear in the Delta K plot (see Figure 1b).

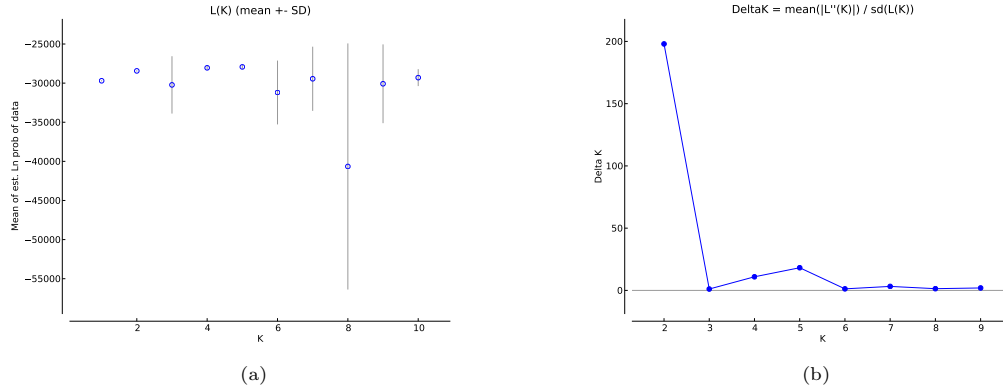


Figure 1: Graphical representation of the estimated probability of data for each K value.

The *K2.popfile* was then analysed using STRUCTURE.

How do you determine the proportion of immigrants in a population?

To answer this question, we can use the population Q-matrix, which presents the calculated proportions of membership of each pre-defined population in each of the K clusters. The population Q-matrix for $K = 2$ is given in Table 1.

| Population | Cluster A | Cluster B |
|------------|-----------|-----------|
| 1 | 0.026 | 0.974 |
| 2 | 0.022 | 0.978 |
| 3 | 0.857 | 0.143 |
| 4 | 0.939 | 0.061 |

Table 1: Membership proportions of each pre-defined populations in each of the K clusters

We can see that populations 1 and 2 have a high proportion of membership in cluster B, while populations 3 and 4 have a high proportion of membership in cluster A. Notably, population 3 has a significant number of individuals in cluster B.

Are there two or three populations in the area of study?

Two, based on the number of clusters and the membership proportions of these clusters.

How do you determine if an individual is a hybrid?

For any given individual, we can look at the individual Q-matrix to infer its ancestry components in each of the K clusters when using the admixture model. The model estimates the ancestry membership proportions of each individual. This can be viewed graphically with a bar-plot (see Figure 2). Any individual with estimated ancestry membership in both clusters can be considered a hybrid. We can also see clearly in the bar-plot that there are immigrants in populations three and four.

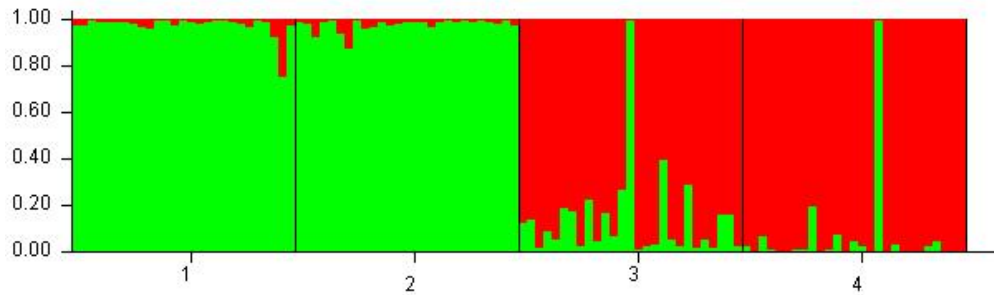


Figure 2: Bar-plot for the estimated ancestry membership proportion of each individual across the four populations. The two clusters are in red and green.

What does the alpha parameter tell us?

Alpha (α) is a Dirichlet parameter that reflects the relative admixture levels between populations. When $\alpha \gg 1$ each individual's alleles has their origin in all K populations in comparable proportions, so the individuals are highly admixed. For values of $\alpha \ll 1$ each individual has its origin mainly in one population (and each population is equally probable). When α is near 0, the models resemble the no admixture model. The STRUCTURE analysis returned an alpha value of 0.0943, which implies that the model has very little admixture. This result is consistent with the membership proportions discussed earlier.

Final Interpretation

The four populations are properly clustered into two clusters. Populations 1 and 2 share a cluster (cluster A), as do populations 3 and 4 (cluster B). There is a high level of gene flow between populations within clusters, and little gene flow between the two clusters. This lack of gene flow between clusters is consistent with a low α value, giving a model with little admixture. Populations 3 and 4 each contained one immigrant from cluster A, and also contained a large number of hybrids with ancestry membership proportions in cluster A.