

Population Genetics Homework - Week 3

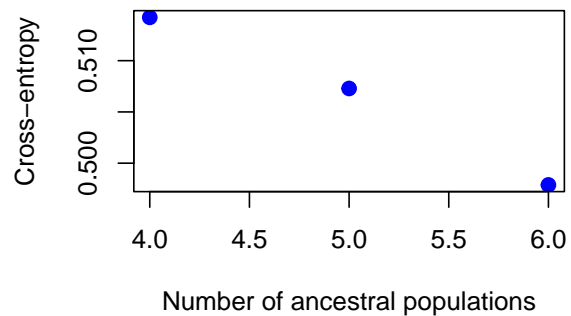
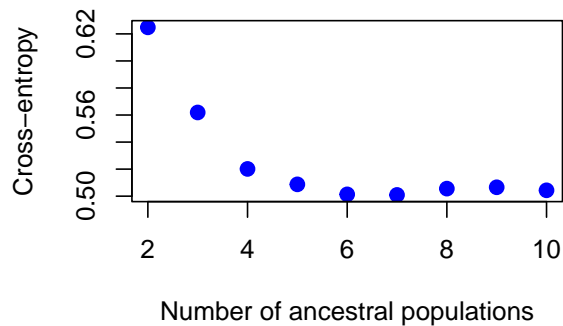
Christian Polania

9/17/2021

Question 2

```
download.file("http://darwin.eeb.uconn.edu/eeb348-resources/gila-trout.stru.geno",
              "c:/Users/cmpol/Desktop/gila-trout.stru.geno")
dat <- snmf("c:/Users/cmpol/Desktop/gila-trout.stru.geno", K = 2:10, repetitions = 10,
            project = "new",
            entropy = TRUE)

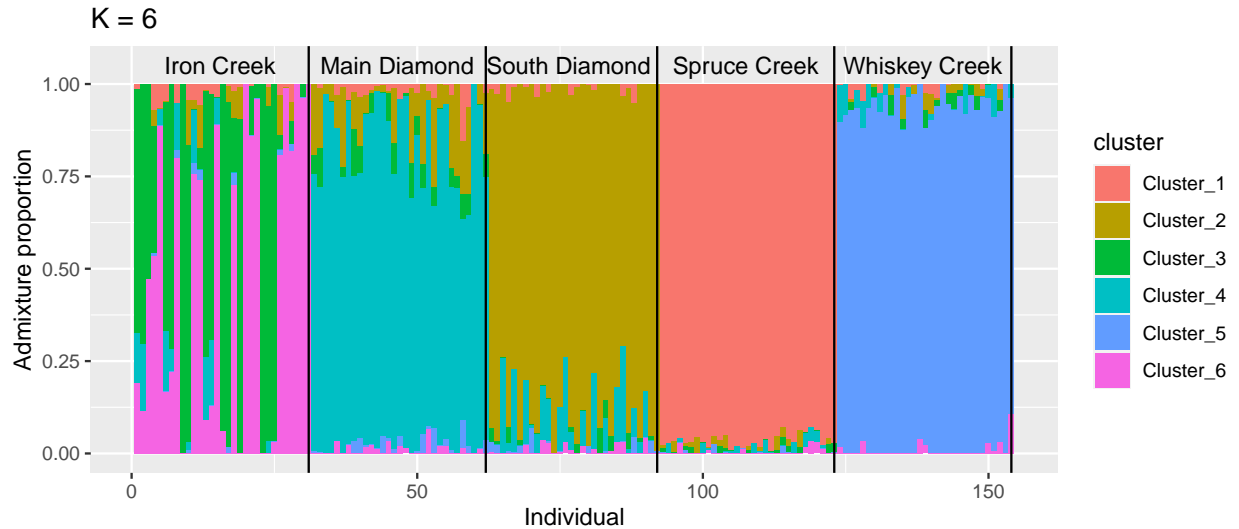
dat1 <- snmf("c:/Users/cmpol/Desktop/gila-trout.stru.geno", K = 4:6, repetitions = 10,
             project = "new",
             entropy = TRUE)
plot(dat, col = "blue", pch = 19, cex = 1.2)
plot(dat1, col = "blue", pch = 19, cex = 1.2)
```



After examining k near the “knee” of the curve, k=6 was found to yield the lowest cross-entropy.

Question 3

```
sites_table <- c(31, 31, 30, 31, 31)
names(sites_table) <- c("Iron Creek", "Main Diamond", "South Diamond",
                       "Spruce Creek", "Whiskey Creek")
plot_LEA(dat, 6, sites_table)
```



Question 4

From the graph above, admixture is clear in the populations sampled at Iron Creek, Main Diamond and South Diamond. Substantial admixture can be inferred from individuals that have large proportions of ancestry from more than one ancestral population. In the above figure, this would be represented by individuals with significant amounts of more than one color (not by the presence of individuals very different from each other).

Iron Creek shows a number of individuals with a large proportion (>0.25) of ancestry from both two clusters, neither of which is substantially represented in other populations. Main Diamond and South Diamond show consistent but lower (<0.25) shared ancestry with each other. This could be explained by more migration between Main Diamond and South Diamond than other pairs of populations.

One way to see the admixture seen at Iron Creek is resulting from two distinct ancestral populations that have interbred more than the others; considering those two populations as one seems acceptable, however, as $k=5$ had a cross-entropy value close to that of $k=6$. The graph for $k=5$ (below) suggests that the pair of clusters previously seen at Iron creek before are more closely related than any other pair.

```
plot_LEA(dat, 5, sites_table)
```

