This directory contains data on four salmon populations from the Seeb et al (2007) reference, which can be consulted for more details.

Each line contains data on a single individual fish.

The columns are:

Individual: a unique id for that individual

Population: an identifier to say which population that fish actually came from

Ogo2: the first allele measured at the Ogo2 marker [treat this as a categorical variable]

Ogo2.1: the second allele measured at the Ogo2 marker

Ogo4: first allele at Ogo4 marker.

Ogo4| second allele at Ogo4 marker.

etc for other markers

12 markers(locus) in total

Note that there are some missing data, denoted "NA".

Note that there are two columns for each marker (eg Ogo2 and Ogo2.1 both refer to marker Ogo2). Thus there are two alleles (types) observed for each fish at each marker.

This is because fish, like humans, are ``diploid", which means they carry two copies of their genome, one inherited from their mother and the other from their father.

Note that the order of the two alleles at each marker is arbitrary. So for example, the first fish has alleles 220 and 222 at Ogo2 and Ogo2.1. It could just have well have been 222 at Ogo2 and 220 at Ogo2.1, and this would mean the same thing.

The data is in both csv format or rda format.

You can read it into R using

four_salmon_pops = read.csv("four_salmon_pops.csv")

or

load("four_salmon_pops.rda")