$BlockFe_{ST}$: Input formats

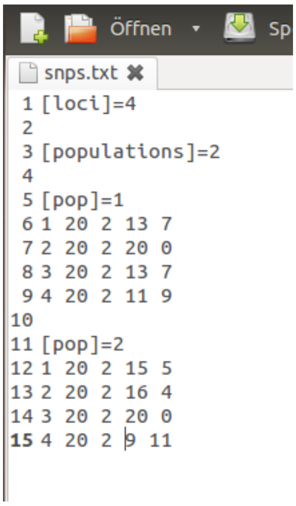
Bastian Pfeifer

May 6, 2018

The function $BlockFe_{ST}$ requires two text files to perform the simulations (see the *exdata* sub-directory of the BlockFeST R package).

1 File 1: SNP data (sample sizes, populations and allele frequencies)

The main input format of $BlockFe_{ST}$ is a space-seperated text file (snps.txt).



```
 1 [loci]=4
 2
 3 [populations]=2
 4
 5 [pop]=1
 6 1 20 2 13 7
 7 2 20 2 20 0
 8 3 20 2 13 7
 9 4 20 2 11 9
10
11 [pop]=2
12 1 20 2 15 5
13 2 20 2 16 4
14 3 20 2 20 0
15 4 20 2 9 11
```

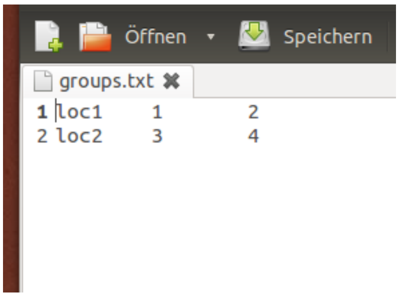
The first row ([loci]=4) defines the total number of SNPs observed from the data where the next line ([populations]=2) defines the total number of populations. From line 5 to 9 (population 1) and from line 11 to 15 (population 2) the following population specific informations are required:

column:
1) numerical id of the SNPs (1,. . .,n.SNPs)
2) sample size of the population.
3) The number of alleles observed in one SNP (default=2).
4) Counts of the major allele.
5) Counts of the minor allele.

The values again have to be seperated by a space " " shift.

2 File 2: Define the regions by creating groups of SNPs

A second input file (groups.txt) is required to set the grouping of the SNPs.



```
1 loc1      1       2
2 loc2      3       4
```

The text file contains three tab-seperated columns, where the first column defines the name of the gene, the second column the start positions (SNP position) and the third column the corresponding end position (SNP position). The total number of rows is limited by the total number of genes (in our example [loci]=4).

The function call would than be:

```
results <- BlockFeST(input="snps.txt",GROUP="groups.txt")
```

3 Using the R-package PopGenome to generate the input

BlockFe$_{ST}$ is fully integrated with the R-package PopGenome which can read in a wide range of commonly used formats storing SNP, as well as alignment data (VCF, FASTA ...). Doing so PopGenome creates an object of class `GENOME` which is the input object of the interface function (also provided by PopGenome) `getBayes()`. See the `Usage` section for more details.

```
install.packages("PopGenome")
library(PopGenome)


GENOME.class <- readData("Alignments")
GENOME.class <- set.populations(GENOME.class,list(c(...),c(...)))
input <- getBayes(GENOME.class, snps=TRUE)
results <- BlockFeST(input)
```