

Inferring individual ancestry proportions using Frappe

We downloaded the SNP datafile from <http://gigadb.org/dataset/100008>

Frappe is a frequentist approach for estimating individual ancestry proportion

Downloaded frappe from <http://med.stanford.edu/tanglab/software/frappe.html> for Mac OS X

1.) Input file format

- 1a.) Parameter files: should be called the “parm.txt”
 - MaxIter: max iterations of EM to run
 - Li et al. used 10,000 iterations
 - K: number of ancestral populations
 - Li et al. ran the analysis with number of K clusters ranging from 2 to 5.
 - M: number of markers in the genotype data. I guess this means the number of SNPs?
 - I: number of individuals
 - Nout and step: how often you would like to output intermediate results, which gives records of ancestry estimates in the current EM iteration.
 - Genotype File: name of genotype data file
 - IndividualFile: name of optional input file

Format should be in a .ped and .map file

(See previous notes for scripts on converting Hapmap file to usable Hapmap and sorting the file)

1.) Convert hapmap to plink format

```
./run_pipeline.pl -h sortex.hmp.txt -export  
polarbearSNPclearnMar8subset1 -exportType Plink
```

This gives both .ped and .map files

- polarbearSNPclearnMar8subset1.plk.map
- polarbearSNPclearnMar8subset1.plk.ped

2.) Adjust the format of .ped file we need to change the A/T/G/C to 1/2/3/4 by recoding the dataset in plink

```
plink --ped polarbearSNPclearnMar8subset1copy.plk.ped --map  
polarbearSNPclearnMar8subset1copy.plk.map --recode12 --noweb
```

Output gives “plink.ped” as the recoded file.

3.) Run Frappe with differing K values in the parameter file

```
./frappe1.1_linux64 PolarBearParam.txt
```