Convert a Plink PED file to TreeMix:

1. Convert Genepop or whatever file to PED and MAP files for use in plink. These should have the same name, but different extensions (i.e. .ped and .map)
2. Convert this ped file to a binary ped file in Plink:
   ./plink --file InputFile --make-bed --out binaryOutput
3. Now make a frequency cluster file sorting data into clusters (probably your sampling sites):
   ./plink --noweb --bfile binaryOutput --freq --within clusters.clust --out SNPSforTreemix

   The clusters.clust file should be a modified version of the .fam file that is created when you make the binary ped file. Open this .fam file in a text editor, remove all columns except 1 and 2, and add a third column with the cluster name you want to assign for every individual. In the end, column 1 is the 'Family' column, 2 is the individual ID, and 3 is the cluster. There cannot be any trailing spaces, and these numbers/strings should be separated by spaces only, not tabs. Also, not sure if it needs to be .clust when you're putting it into Plink, .txt should be fine too.
4. Now gzip this created output file.
5. Run the python script that came with TreeMix. Put the Plink output gzipped file in the same directory as the python script, and in the terminal type:
   plink2treemix.py SNPSforTreemix.frq.gz treemixInput.frq.gz
6. This output gzipped file can be run directly in TreeMix now!

Using Treemix

Run from the terminal:

./Treemix -i inputfile.frq.gz -se -m K -root YourOutgroup -o outputtextK

Here, -se will provide standard error for the migration edges, and the 'K' for -m is the number of migration events you want to test. Start with 0, and sequentially run the program for each value of K you want to test. In the terminal, the log-likelihood for that migration event will be recorded. Depending on the number of loci, running migration events with -se will take several hours per value of K tested. If you don't have an outgroup to root your tree, don't put the -root option in.

Now the output needs to be plotted in R. There are two R plotting functions that come with Treemix, one to plot the tree and one to plot the residuals of how each population fits into the tree. Treemix output will make 6 g-zipped folders that R uses to do these plots, so don't change the name on any of them.

In RStudio:

```
>source("src/plotting funcs.R")
>plot_tree("outstem")
```

#where outstem is the extension of the output you provided in the Treemix command line

#And

```
>plot_resid("outstem", "poporder") #plot the residuals for a particular migration event. If
```
you test 10 migration events, you will want to make 10 residual plots.

"Poporder" is a text file that lists the order you want to populations to be listed in for the residual plot, followed by a number separated only by one space.

e.g.

PLB 1
SGB 2
KJI 3
.
.
.
.
TKT 10

In the .trees folder that is output from Treemix, the first line is a Newick code for building the tree, while the remaining lines are the migration events. You can use this Newick line to rebuild the tree in something like Figtree and midpoint root it. This is for if you didn't root the tree in Treemix (i.e. you don't have an outgroup).

Using Threepop

This is the three-population test developed by Reich et al. (2009). It uses the exact same input file as Treemix and is run from the same directory. It will provide an f3 statistic, that if negative, suggests admixture between the three populations it's testing. It runs in the format of A:B,C where A is trying to be fit to a 'tree' of B and C. If A has a negative f3 statistic, it suggests it is a hybrid of B and C. It will also provide the standard error of the f3 stat, and a Z-statistic for significance (Z< -2-3 is significant).

```
>./threepop -i input.gz -k 500
```

-k is the block of SNPs you want to test; in order to avoid linkage of any loci, threepop samples your SNPs in blocks of k size to generate the test statistics. For thousands of loci, use -k 500; for hundreds of loci, use -k 10-50

Threepop is very fast and produces a text file of output of all combinations of populations in your dataset.

*****Treemix has a simple manual containing most of the above steps, but isn't great at explaining everything. Read through the manual though if anything I wrote doesn't make sense!