TreeMix is another way to evaluate the admix group and plot a graph, which is similar to the admix graph in ADMIXTOOLS 2, but with different calculations.

We use a pipeline written by carolindahms with some modifications.

<https://github.com/carolindahms/TreeMix>

The original TreeMix program description is here:

<https://bitbucket.org/nygcresearch/treemix/wiki/Home>

If you want to use the original TreeMix script without carolindahms’ pipeline, you can follow the tutorial here:

<https://speciationgenomics.github.io/Treemix/>

To use carolindahms’ pipeline, you need some R packages:

plyr, dplyr, data.table, BITE, OptM

you also need a program: PHYLIP (you can install it with conda)

You might have a trouble installing BITE, and you can follow the instruction in BITE website:

<https://github.com/marcomilanesi/BITE>

To be more convenient, the environment setting is ready on h71 server. Just activate the conda environment “treemix”.

PHYLIP consensus program is at /home/hpc/crlee/miniconda3/envs/treemix/bin/consense

on h71 server.

IMPORTANT:

All the scripts in this pipeline are local scripts. If you want to run them on the server, please use “perl $create\_job -cj\_exc -cj\_conda treemix” in front of every command line.

Before running:

It’s better to convert the vcf format to the treemix format. Please use “vcf2treemix.sh” to convert it. Please also prepare a group information file as the following format (separated by tabs):

sample\_name sample\_name group

Usage: sh vcf2treemix.sh YOUR\_VCF\_FILE.vcf(.gz) GROUP\_INFO.txt

Output: YOUR\_VCF\_FILE\_NAME.treemix.frq.gz

Steps:

1. Put all files in a folder on the h71 server.
2. Run step 1~4.

eg. perl $create\_job -cj\_exc -cj\_qname Step1 -cj\_conda treemix sh Step1\_TreeMix.sh ……

IMPORTANT: Please follow the command line suggested by the official website: <https://github.com/carolindahms/TreeMix>

Use “sh” to run \*.sh scripts, and use “Rscript” to run \*.R scripts

1. For step 2 & 4, it is different from the instruction:
2. Check and modify the R script locations at line 12 & 13.
3. For step 4, modify the “nt.in” parameter as “number of independent runs (N)” in step 3 in the **Step2\_4\_TreeMix\_m.R** script(line 14).
4. For step 4, modify the “nboot.in” parameter as “bootstrap replicates” in step 3 in the **Step2\_4\_TreeMix\_m.R** script (line 15).
5. For step 4, modify the “admix.number” parameter as “number of migrations” in step 3 in the **Step2\_4\_TreeMix\_m.R** script (line 16).
6. For step 4, put col.txt and poporder.txt files in the “final\_runs” folder.

Script usage:

**Rscript Step2\_4\_TreeMix\_m.R step output\_folder\_name\_from\_previous\_step file\_name\_prefix newick\_file [legend\_position]**

Step 2:

Only first 2 arguments are required

eg. Rscript Step2\_4\_TreeMix\_m.R 2 test\_migrations

Step 4:

legend\_position is optional, default: right

eg. Rscript Step2\_4\_TreeMix\_m.R 4 final\_runs PREFIX\_1m\_finalrun\_ ../PREFIX\_finalconstree.newick right

PREFIX is the file name without path and extension, which is the same with “output file name” in step 1 or step 4. 1m means 1 mix group, 2m means 2 mix groups….

The “**file\_name\_prefix**” does not contain any characters after the pattern “finalrun\_”.

For plotting:

If you don’t have a good TreeMix figure, you can modify **Step2\_4\_TreeMix\_m.R** script. The main plotting function is “treemix.bootstrap.m”. Every parameter has its description, just follow the description and modify it.