Tutorial Spring 2018 Smithsonian Institution

Download the Tutorial directory

The tutorial can be reached here: [Smithsonian tutorial 2018](#). Once downloaded, unpack and then move the directory smi2018 to a good location, for example your Documents or your Desktop folder.

Download migrate 4.2.14 from the migrate website

Download the version of migrate for your operating system: [Open the download site Migrate](#) To install a working copuy of **migrate-n** copy the executable into the data directory or then into a directory that is on the search path; I can give little hints for windows but on macs and unix you can see the search path by executing this command on the commandline (in a terminal or shell window).

```
set | grep ^PATH
```

then copy the migrate-n executable to one of the directories on the list, I assume that most directories in that list are NOT writable, except */usr/local/bin*, or *~/bin*; if you cannot see these then copy migrate-n into the tutorial directory.

If the above instructions sound foreign to you, copy the executable *migrate-n* into the tutorial directory.

Detect boundaries within a species

We have a transect of 5 populations that look like this

$$a \longleftrightarrow b \longleftrightarrow c \longleftrightarrow d \longleftrightarrow e$$

Preliminary analyses have shown that **a** and **e** are different: for example (I can reject the hyptohesis **a=e**, and migration from $a \to e$ and $e \to a$ is small. But where is the boundary? There are several questions a program like migrate cannot answer: what is a species? Are these different species?

But we can attempt to investigate were there may be a break in the geneflow pattern, or whether we may have a cline between the extremes. There are many hypotheses to test, and we will not consider all all possible ones, we will not even consider all usefule ones.

We will use very simple models that assume the populations existed for a long time and that (almost) all observed allele patterns are caused by migration:

(0) All sampling locations belong to the same population.

(a) Gene flow is moving alleles from left to right.

(b) Gene flow is moving alleles from right to left.

(c.) Gene flow is moving alleles left and right.

(d) Geneflow from $1 \to 2 \to 3 \leftarrow 4 \leftarrow 5$ with a *fixed* low immigration rate from $1 \leftrightarrow 5$}

(d2) Geneflow from $1 \to 2 \to 3 \leftarrow 4 \leftarrow 5$ with a *divergence event* $1 \leftrightarrow 5$}

(g) Locations 1,2 are a single population A, location 4 and 5 are one single population B, location 3 is an admixture of population A and B.

Hands-on part

We will run the tutorial in a director/orchestra way: I show you parts of the interaction with the program and we all run the program together! We will start with the simple models (0), (a), and (b). I will explain how to setup the parameter file and how to run, and once the runs a are done we will look at the output and I will explain some more.

Results (stop reading here – spoiler alert)

This 7 models compare using marginal likelihoods like this:

```
arzak:data2>grep "All    " out* | sort -n -k 4,4 | bf.py
Model                       Log(mL)    LBF     Model-probability
-------------------------------------------------------------
1:outfile_0:                -19156.12 -2367.50        0.0000
2:outfile_d:                -18123.36 -1334.74        0.0000
3:outfile_b:                -18053.53 -1264.91        0.0000
4:outfile_a:                -17920.43 -1131.81        0.0000
5:outfile_c:                -17778.12  -989.50        0.0000
6:outfile_g:                -17508.34  -719.72        0.0000
7:outfile_d2:               -16788.62     0.00        1.0000
```

A Moment of truth

The data was simulated using model _____

Simulated data was generated with this command

We use this MS commandline:

```
ms 25 2 -t 50 -r 0. 1000 \
    -I 5 5 5 5 5 5 0.0 \
    -m 1 2 1.0 -m 2 3 1.0 \
    -m 3 4 1.0 -m 4 5 1.0
    -M -q 1000 -T -ej 20.0 5 1 > tree; \
cat tree | migdata
```

The ms simulation program by Hudson was modified to generate trees that are compatible with the simulator in our lab, I will need to generate a distributable version of hte modified ms, and more instructions about how to use the simulator [ask if you need this]