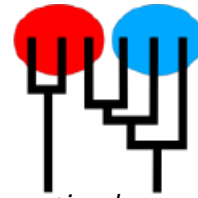


MIGRATE Tutorial 0



This tutorial discusses the data format of MIGRATE. Students will practice how a dataset is constructed. This tutorial can be finished in a mac, windows, or unix environment.

A reminder: if you want to keep results (they are written by default into a file called `outfile` and `outfile.pdf`), save it using either `cp outfile yourfavoredfilename` or `mv outfile yourfavoredfilename` or change the name of the output file in the menu.

Datasets for MIGRATE

Download the MIGRATE EPONGE tutorials from the website at <http://www.peterbeerli.com/tutorials/migratetutorial.zip>. For each sub-tutorial there are different datasets. Once you unpack the zip directory you will this structure

```
tutorial0
tutorial1
tutorial2
win
mac
```

The `tutorial0` directory contains these files

- `bindin.fasta`: 4 individual sequences from sea urchin locus *bindin*
- `tpsd.fasta`: 4 individual sequence from sea urchin locus `ebr1:tpsd`
- `infile.msat`: an example microsatellite file.
- `results`: contains migrate files that are result from this exercise.

For this tutorial, you will need to a text editor that can save the file in text (ascii) format. For example: on a mac: TextWrangler.app, TextEdit.app; on a Windows computer: notepad++, Visual Studio Code; unix: emacs, vim, there two general editing platforms for all systems: Light Table, Sublime, Blue Fish

Instructions

To construct a MIGRATE file you may need the help of the manual! I will guide you though the first conversion.

1. Change the FASTA file so that it contains for each individual only two lines, the header line with ">" and the data line with the sequence. This will look something like this, from:
>13f16_bin_a1

Peter Beerli, update September 2019

```
TACGTCAACACGATGGGCTACCCCAAGCCATGAGTCCGCAAATGGGAGG
TGTTAACTACGGACAACCAGCGCAGCAGGGTTATGGCGCTCAAGGAATGG
```

to:

```
>13f16_bin_a1
```

```
TACGTCAACACGATGGGCTACCCCAAGCCATGAGTCCGCAAATGGGAGGTGTTAACTACGGACAACCAGCGCAGCAGGGTTA
```

2. then change combine each line so that we have the individual name with precisely 10 characters and on position 11 or 12 (excess spaces do not matter) start the sequence

```
13f16a:1
```

```
TACGTCAACACGATGGGCTACCCCAAGCCATGAGTCCGCAAATGGGAGGTGTTAACTACGGACAACCAGCGCAGCAGGGTTA
```

3. Now add the header lines: number of populations, number of loci, and title and on the next line then number of sites (for bindin it is 252 and tpsd 456), we also need for each population a line with the number of individuals here this is 4.

```
1 1 examplelocus
```

```
(s252)
```

```
4 sea urchin population
```

```
13f16a:1 TACGTCAACACGATGGGCTACCCCAAGCCATGAGTCCGCAAATGGGAGGTGTTAACTACGGACAACCAGCGCA
```

```
....
```

4. NOW TRY THIS BY YOURSELF for the second file TPSD.fasta
5. Combining multiple loci is simple, there are two ways
 - (a) copy the second locus as a complete block without the names and paste the block so that each individual from locus 1 extends into the locus2, thus the data for each individual os on one line. The second line in the file then needs to change to (s256) (s456). This is a very versatile format as you will see on the last day.
 - (b) Copy paste the whole second locus file towards the end of the first locus file, remove the title and sites lines of the second locus and adjust the first two so that it now reads 1 2 two loci and (s256), (s456) [recognize the comma]
6. I will show more complex situations (with more loci) and also explain the microsatellite data format.

