**Instruction for the code to conduct iterative BTW analysis**

To conduct iterative BTW analysis, you can run the shell script “run\_iterated\_BTWest.sh”.

The code was confirmed to run on mac OS Sierra (10.12.5).

To run the code, please execute “./run\_iterated\_BTWest.sh” in your terminal.

**Data preparation**

Please prepare the following three files.

1. nucleotide sequence of all samples used in the analysis (sample\_seq.txt in EXAMPLE folder).
2. collapse pair sequences and single outgroup sequences (collapse.MFA.mfa in EXAMPLE folder).
3. fltrst file which is the output of BASEML analysis.

\* formats of sample\_seq.txt and collapse.MFA.mfa are the same to those of the input and output files of the code to generate collapse pair sequences (please check “instruction\_collapse\_verA.docx” in “\_\_make\_collapse\_seq\_for\_actual\_data\_analysis”).

**Parameter setting in the shell script**

Please set the values of 7+12 parameters.

**a**: total number of sequences in “collapse.MFA.mfa”. In the example file, a = 6.

**b**: number of internal nodes in the phylogeny. You can know this value by checking

fltrst file.



In this example, there are four internal nodes, so b = 4.

**c**: Order of the target node to which the ancestral state will be inferred.



If you want to estimate the SFS of mel population, the target node is the node9 (3rd letter of the states of ancestral nodes). So c = 3.

**d**: Number of the mutation categories you want to consider. This code allows you to categorize mutations. For example, if you want to consider

**E**: Order of the last sequence of group 1 in the sequence file. In the example file, D =

35

**F**: Number of species you want to use the single sequence. In the example file, we use

single sequences from *D. yakuba* and *D. erecta* as the outgroup. So, F = 2.

**G**: Order of the single sequence from outgroup 0 in the sequence file. In the example

file, single sequence from *D. yakuba* comes after *D. simulans*. So, this species is

named as single seq 0 and G = 36.

**H**: Order of the single sequence from outgroup 1 in the sequence file. In the example

file, single sequence from *D. erecta* comes after *D. yakuba*. So, this species is

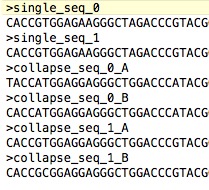
named as single seq 1 and H = 37.

**Dirpath**: path of the input folder.

**collapsedDirpath**: path of the output folder.

**Output**

The code outputs the file with collapse sequence in “collapsedDirpath” folder.



single\_seq\_X: shows the sequence from outgroup X which has a single sequence in the input file.

collapse\_seq\_X\_A and collapse\_seq\_X\_B: shows the two collapse sequences of group X which has polymorphic sequences in the input file.

The output file can be used as the input sequence file of BASEML software.