**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. Mutations in one category will be summed up to calculate

SFS for the weighting. For example, if you want to put

T->C, T->G, A->C and A->G mutations into the first category,

C->T, T->G, C->A and G->A mutations into the second category and

A->T, T->A, G->C and C->G mutations into the third category,

you can set d = 3.

If you want to consider each of 12 mutations separately, you can set d = 12.

**e**: Order of the first sequence of the species you want to analyze in the sequence file

(sample\_seq.txt). In the example file, if you are estimating the SFS of *D.*

*melanogaster*, the sequences of this species come first and start from 1st line, so e =

1. If you want to analyze *D. simulans*, e = 15.

**f**: Order of the first sequence of the species you want to analyze in the sequence file

(sample\_seq.txt). In the example file, 1st ~ 14th sequences are from *D.*

*melanogaster*) and 15th ~ 35th sequences are from *D. simulans*. So f = 14 if you are

analyzing *D. melanogaster* and f = 35 if you are analyzing *D. simulans*. Note that e

and f refer the order of the “sequences” in the file, not the line number.

**mut\_TC~mut\_GA**: Category of the each of 12 mutations. If you set d = 3, you can

set 1, 2 or 3 for each mutation.

For example,

mut\_TC=1;

mut\_TA=3;

mut\_TG=1;

mut\_CT=2;

mut\_CA=2;

mut\_CG=3;

mut\_AT=3;

mut\_AC=1;

mut\_AG=1;

mut\_GT=2;

mut\_GC=3;

mut\_GA=2;

This makes

T->C, T->G, A->C and A->G mutations into the first category,

C->T, T->G, C->A and G->A mutations into the second category and

A->T, T->A, G->C and C->G mutations into the third category.

d = 12 and

mut\_TC=1;

mut\_TA=2;

mut\_TG=3;

mut\_CT=4;

mut\_CA=5;

mut\_CG=6;

mut\_AT=7;

mut\_AC=8;

mut\_AG=9;

mut\_GT=10;

mut\_GC=11;

mut\_GA=12;

This makes each of 12 mutations into different mutation category.

**iteration**: the number of iterations in the iterative BTWest analysis. Note that the first

round becomes BTWne, so if you set iteration = 6, initial BTWne analysis and

following BTWest analysis will be done.

**dirpath**: path of the input folder which contains the three input files.

**Output**

The code outputs two files.

**AWP\_substitution\_for\_post\_weighted\_collapse:**

The estimated number of nucleotide fixations estimated by AWP method (Matsumoto et al. 2015).

**estimated\_frequency\_spectrum:**

The estimated number of polymorphic mutations in each frequency class for 12 mutations.