**Instruction for the code to make collapse sequences**

To make the collapse sequences, you can run the shell script “run\_make\_collapse\_seqs.sh”.

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

To run the code, please execute “./run\_make\_collapse\_seqs.sh” in your terminal.

**Data preparation**

Please prepare nucleotide sequence files as in the “EXAMPLE” folder.

\* The first letter of each sample name has to be “>”.

\* Samples from one species have to be listed continuously. Our example file lists

sequences form *Drosophila melanogaster* in 1st ~ 28th line and from *D. simulans* in

29th ~ 70th line. You can not list these sequence, for example, like 1st ~14th and 31st

~ 44th lines are from *D. melanogaster*.

**Parameter setting in the shell script**

Please set the values of eight parameters A ~ H in the shell script.

**A**: Number of species you want to make the collapse sequences. In the example file,

we have polymorphic sequence from *D. melanogaster* and *D. simulans*. So, A = 2.

**B**: Order of the first sequence of group 0 in the sequence file. In the example file, *D.*

*melanogaster* sequences come first and start from 1st line, so this species is named

as group 0 and B = 1.

**C**: Order of the last sequence of group 0 in the sequence file. In the example file, 1st

~ 14th sequences are from group 1 (*D. melanogaster*), so C = 14. Note that B and C

refer the order or the “sequences” in the file. So C is not 28.

**D**: Order of the first sequence of group 1 in the sequence file. In the example file, *D.*

*simulans* sequences come just after *D. melanogaster*. So, *D. simulans* is named as

group 1 and D = 15.

**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

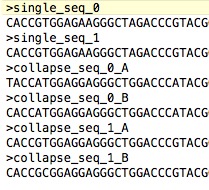
named as single seq 1 and H = 37.

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

**collapsedDirpath**: name 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.