***run\_make\_BASEML\_input\_collapse\_mstyeo\_100\_reps.sh***

This code reads the output of the simulation and make two collapse pair sequences for each of m ,s, t, y, e and o species.

The sequence will be converted from 0,1,2,3 to T,C,A,G.

Site with more than two states at least in one species will be automatically filtered.

The code first sample 10 sequences and make two collapse pair sequence.

Output file “BASEML\_input.MFA” becomes the input sequence file of BASEML analysis.

Output file “sample\_seq\_X.txt” is the 10 sequences sample from species X.

Output file “site\_not\_used\_for\_collapse\_method.txt” list the position of sites which have >2 states and filtered from the analysis.

Output file “site\_ used\_for\_collapse\_method.txt” list the position of sites which have <3 states and used to generate the collapse pair sequences.

***run\_make\_BASEML\_input\_6\_ancestors\_100\_reps.sh***

This code reads the output of the simulation and make the ancestral sequence of six species from the simulation record.