

# MLE\_BiMarkers

## Description

Maximum likelihood estimation of phylogenetic networks given bi-allelic genetic markers (SNPs, AFLPs, etc).

## Usage

<b>MLE_BiMarkers</b> [-pseudo] [-sd seed] [-pl parallelThreads] [-mr maxReticulation] [-tm taxonMap] [-fixtheta theta] [-varytheta] [-esptheta] [-snet startingNetwork] [-ptheta startingTheta] [-pi0 PI0] [-diploid] [-dominant dominantMarker] [-op]		
<b>ML Settings</b>		
- mnr nu mR uns	The number of iterations of simulated annealing. The temperature of simulated annealing is reset in the beginning of each iteration, then the temperature reduces gradually as more states are examined. By doing this, the search can jump out of local optimum in the beginning of one iteration easily, then random walk in the space of phylogenetic networks is performed during each iteration. The default value is 100.	o p t i o n a l
- me c m ax Ex ami nati ons Co unt	The maximum allowed times of examining a state during one iteration. During one iteration of simulated annealing, each state is obtained by random walk in the space of phylogenetic networks. A state is proposed by randomly altering the topology or parameters in the previous state, then the new state is examined, and can be accepted or rejected. If the number of states examined exceeds this limit, the current iteration terminates, and a new iteration starts. The default value is 50,000.	o p t i o n a l
- mn o n um Opt imu ms	The number of optimal networks to output. The optimal networks are outputted after every iteration. The optimal networks outputted are the optimal networks in any state examined in any iteration. The default value is 10.	o p t i o n a l
- mf ma xFa ilur es	The maximum allowed times of failures to accept a new state during one iteration. If the number of times when new purposed states are continuously rejected exceeds this limit, the current iteration terminates, and a new iteration starts. The default value is 50.	o p t i o n a l
-pl par alle lTh rea ds	The number of threads running in parallel. The computation of pseudo-likelihood is parallelized since the likelihood of trinets can be computed independently. This number of threads indicates how many threads are used for computation of pseudo-likelihood. However, more threads don't necessarily mean faster computations usually. In practice, the user needs to figure out the best number of threads by experimenting on a smaller data set and see whether the inference is faster by increasing the number of threads. The default value is the number of threads in your machine.	o p t i o n a l
<b>Inference Settings</b>		
- pse udo	Use pseudolikelihood.	o p t i o n a l
- mr ma xR etic ulat ion	The maximum number of reticulation nodes in the sampled phylogenetic networks. This number is a bound on the number of reticulations that the method explores during the search. However, this does not mean that the inferred network has to have this number of reticulations. In theory, this number can be set to a very large value so as not to impose any real bound. However, in practice, the number of reticulations can affect the running time. Furthermore, in the absence of a real criterion for model selection, setting this parameter to a large value might result in overly complex networks. We recommend that the user sets the parameter at a value that is "reasonable" to them, based on knowledge of the data set. The default value is 4.	o p t i o n a l

-tm taxonMap	Gene tree / species tree <a href="#">taxa association</a> . By default, it is assumed that only one individual is sampled per species in gene trees. This option allows multiple alleles to be sampled. For example, the gene tree is $((a1,a2),(b1,b2)),c$ ; and the species tree is $((a,b),c)$ ; the command is <code>-tm &lt;a:a1,a2; b:b1,b2;c:c&gt;</code> . If the set of taxa appeared in this mapping is a subset of input data, the subset of input data will be used for the inference.	optional
-fixtheta	Fix the population mutation rates associated with all branches of the phylogenetic network to this given value (theta). By default, we estimate a constant population size across all branches.	optional
-esptheta	Estimate the mean value of prior of population mutation rates.	optional
<b>Starting State Settings</b>		
-snet	Specify the starting network. The input network should be ultrametric with divergence times in units of expected number of mutations per site, inheritance probabilities and population sizes in units of population mutation rate (optional). See example below. The default starting network is the MDC trees given starting gene trees.	optional
-ptheta startingThetaPrior	Specify the mean value of prior of population mutation rate (startingThetaPrior). The default value is 0.036. If -esptheta is used, startingThetaPrior will be treated as the starting value, otherwise startingThetaPrior will be treated as the fixed mean value of prior of population mutation rates.	optional
<b>Data related settings</b>		
-diploid	Specify whether sequence sampled from diploids. If the sequence is from diploid and there are not dominant markers, the characters in the sequence should be '0', '1' or '2'. '0' and '2' are the homozygotes and '1' is the heterozygote state.	optional
-dominant dominantMarker	Specify which marker is dominant if the data is dominant. The dominant marker can either be '0' or '1'. Only use when "-diploid" is specified. If this option is specified, the characters in the sequence should be '0' or '1'.	optional
-op	Specify whether or not to ignore all monomorphic sites. If this option is used, the data will be treated as containing only polymorphic sites, and all monomorphic sites are ignored. Then the frequencies of the monomorphic sites will be computed by the likelihood function.	optional
-pi0 value	Specify the stationary distribution of marker "0". Value should be between 0 and 1. If not specified, the stationary distribution will be calculated from input data.	optional

## Example

Download: [run\\_0.nex](#)

Please download the example instead of copying from this webpage and pasting into your local file!

```
#NEXUS
Begin data;
Dimensions ntax=5 nchar=100;
Format datatype=dna symbols="012" missing=? gap=-;
Matrix

A_0 1001011010101011001000010101010111001010011001100101111011000011111000001010001001100000110100001011
C_0 100111101110101100100101010101011101101001000110000111111001000111000001010011001100100100110001011
L_0 100101101010011100100001010101011100101001100110010111111001011110000001010001001100000110100001011
Q_0 100101101010101100100101010101011100101001100110010111111001011110000001010001001100000110100001011
R_0 100101101010101100110101000101011100111001100110010101111001011110000001010101001100000100100001001
;End;
BEGIN PHYLONET;
MLE_BiMarkers -pseudo -mnr 10 -mec 50000 -mno 20 -mf 100 -pi0 0.5 -dd -mr 1 -pl 8 -ptheta 0.006 -thetawindow 0.006 -sd 12345678 -tm <A:
A_0; C:C_0;L:L_0;Q:Q_0;R:R_0> ;
END;
```

This command will run **maximum pseudolikelihood estimation** of 10 iterations with 20 optimal networks printed. And after 100 times of failure to accept a new state, or after 50000 examinations of new states, it will start a new iteration. We will estimate population mutation rates for all branches, and they are the same across all branches. The number of reticulation nodes is limited to 1. The starting value of population mutation rate is given by 0.006. We use the random seed of 12345678. In the end, we indicate the mapping from taxa to species.