

MCMC_BiMarkers

Description

Bayesian estimation of the posterior distribution of phylogenetic networks given bi-allelic genetic markers (SNPs, AFLPs, etc).

This method uses Jeigen package, if you encounter errors like:

```
java.lang.NoClassDefFoundError: Could not initialize class jeigen.JeigenJna$Jeigen
```

You need to follow this instruction to build Jeigen on your machine: <https://github.com/hughperkins/jeigen#how-to-build-linux> . After you build Jeigen, copy the library file to .jeigen folder in the home path.

Usage

<pre>MCMC_BiMarkers [-diploid] [-dominant dominantMarker] [-op] [-cl chainLength] [-bl burnInLength] [-sf sampleFrequency] [-sd seed] [-pl parallelThreads] [-mc3 temperatureList] [-mr maxReticulation] [-tm taxonMap] [-fixtheta theta] [-varytheta] [-esptheta] [-pp poissonParameter] [-dd] [-ee expPrior] [-snet startingNetwork] [-ptheta startingThetaPrior] [-pi0 PI0]</pre>		
MCMC Settings		
-cl <i>chainLength</i>	The length of the MCMC chain. Example: 500,000.	optional
-bl <i>burnInLength</i>	The number of iterations in burn-in period. Example: 200,000.	optional
-sf <i>sampleFrequency</i>	The sample frequency. The default value is 500.	optional
-sd <i>seed</i>	The random seed. The default seed is 12345678.	optional
-pl <i>parallelThreads</i>	The number of threads running in parallel. The default value is the number of threads in your machine.	optional
MC3 Settings		

-mc3 <i>temperatureList</i>	<p>The list of temperatures for the Metropolis-coupled MCMC chains. For example, <i>-mc3 (2.0, 3.0)</i> indicates two hot chains with temperatures 2.0 and 3.0 respectively will be run along with the cold chain with temperature 1.0. By default only the cold chain will be run. Note that</p> <ul style="list-style-type: none"> The temperatures should be DIFFERENT! For example, <i>-mc3 (2.0, 2.0, 3.0)</i> is invalid. The temperature of the cold chain should NOT be included. For example, <i>-mc3 (1.0, 2.0, 3.0)</i> is incorrect. Metropolis-coupled MCMC leads to faster convergence and better mixing, however, the running time increases linearly with the number of chains. We suggest you first run a standard MCMC chain (cold chain) without this command. If the trace plot indicates the chain is not mixed well (jagged, stuck in local maxima for a long time), then try this command. 	optional
Inference Settings		
-mr <i>maxReticulation</i>	The maximum number of reticulation nodes in the sampled phylogenetic networks. The default value is 4.	optional
-taxa <i>taxaList</i>	The taxa used for inference. For example, -taxa (a,b,c)	required
-tm <i>taxonMap</i>	<p>Gene tree / species tree taxa association. 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 <i>-tm <a:a1,a2; b:b1,b2;c:c></i>. Note that the taxa association should cover all species, e.g. <i>-tm <a:a1,a2; b:b1,b2></i> is incorrect because <i>c:c</i> is dropped out.</p> <p>If the set of taxa here is different than that in "-taxa taxaList", a common subset of taxa of both parameters will be used for inference.</p>	optional
-fixtheta <i>theta</i>	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
-varytheta	The population mutation rates across all branches may be different when estimating them. By default, we estimate a constant population size across all branches.	optional
-espthta	Estimate the mean value of prior of population mutation rates.	optional
Prior Settings		
-pp <i>poissonParameter</i>	The Poisson parameter in the prior on the number of reticulation nodes. The default value is 1.0.	optional
-dd	Disable the prior on the diameters of hybridizations. By default this prior on is exp(10).	optional
-ee	Enable the Exponential(10) prior on the divergence times of nodes in the phylogenetic network. By default we use Uniform prior.	optional
Starting State Settings		

-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.	o p t i o n a l
-ptheta startin gTheta Prior	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.	o p t i o n a l
Data related settings		
-diploid	Specify whether sequence sampled from diploids.	o p t i o n a l
-dominant dominantMarker	Specify which marker is dominant if the data is dominant. Either be '0' or '1'.	o p t i o n a l
-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.	o p t i o n a l

Example

Download: [example_bimarkers.nex](#)

```
#NEXUS
Begin data;
Dimensions ntax=5 nchar=10;
Format datatype=dna symbols="012" missing=? gap=-;
Matrix

a 1010101010
b 1100100110
c1 1010100011
c2 1001110100
d 1000011110

;End;

BEGIN PHYLONET;
MCMC_BiMarkers -cl 500000 -bl 200000 -sf 500 -diploid -dominant 1 -op -varytheta -pp 2.0 -ee 2.0 -mr 1 -pl 4 -esptheta -ptheta
0.3
-sd 12345678

-taxa (a,b,c1,c2,d)
-tm <A:a; B:b;C:c1,c2;D:d>;

END;
```

Note that an empty line should be left after "Matrix".

This command will run MCMC chain of 500000 iterations with 200000 burn-in iterations, and one sample will be collected every 500 iterations. The taxa are diploids and 1 is the dominant marker. Only polymorphic sites will be used. We will estimate population mutation rates for every branches, and they may be different. A Poisson prior of 2.0 will be adopted, and a Exponential(2.0) prior will be adopted. The number of reticulation nodes is limited to 1. We will sample the mean value of prior of population mutation rates, and the starting value of 0.3 is given. We use the random seed of 12345678. In the end, we indicate the mapping from taxa to species.