

# Example: sequence data set wit two loci [simula

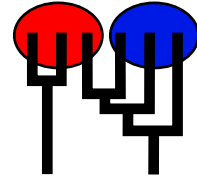
## MIGRATION RATE AND POPULATION SIZE ESTIMATION

using the coalescent and maximum likelihood or Bayesian inference

Migrate-n version 3.6.11 [June-18-15]

Program started at Wed Apr 6 20:18:16 2016

Program finished at Wed Apr 6 20:23:16 2016



## Options

Datatype:

DNA sequence data

Inheritance scalers in use for Thetas:

All loci use an inheritance scaler of 1.0

[The locus with a scaler of 1.0 used as reference]

Random number seed:

(from parmfile)

310705631

Start parameters:

Theta values were generated

from the FST-calculation

M values were generated

from the FST-calculation

Connection type matrix:

where m = average (average over a group of Thetas or M,  
s = symmetric M, S = symmetric 4Nm, 0 = zero, and not estimated,  
\* = free to vary, Thetas are on diagonal

Population	1	2	3	4
1 Africa	m	a	a	a
2 Americas	a	m	a	a
3 Pacific	a	a	m	a
4 Asia	a	a	a	m

Order of parameters:

1	$\Theta_1$	=	$\Theta_1$	[m]	<displayed>
2	$\Theta_2$	=	$\Theta_1$	[m]	
3	$\Theta_3$	=	$\Theta_1$	[m]	
4	$\Theta_4$	=	$\Theta_1$	[m]	

5	M	2→1	=	M	2→1	[a]	<displayed>
6	M	3→1	=	M	3→1	[a]	<displayed>
7	M	4→1	=	M	4→1	[a]	<displayed>
8	M	1→2	=	M	1→2	[a]	<displayed>
9	M	3→2	=	M	3→2	[a]	<displayed>
10	M	4→2	=	M	4→2	[a]	<displayed>
11	M	1→3	=	M	1→3	[a]	<displayed>
12	M	2→3	=	M	2→3	[a]	<displayed>
13	M	4→3	=	M	4→3	[a]	<displayed>
14	M	1→4	=	M	1→4	[a]	<displayed>
15	M	2→4	=	M	2→4	[a]	<displayed>
16	M	3→4	=	M	3→4	[a]	<displayed>

Mutation rate among loci:

Mutation rate is constant for all loci

Analysis strategy:

Bayesian inference

Proposal distributions for parameter

Parameter	Proposal
Theta	Slice sampling
M	Slice sampling

Prior distribution for parameter

Parameter	Prior	Minimum	Mean*	Maximum	Delta	Bins
Theta	Uniform	0.000000	0.050000	0.100000	0.010000	500
M	Uniform	0.000000	5000.000000	10000.000000	1000.000000	500

Markov chain settings:

Long chain

Number of chains	1
Recorded steps [a]	5000
Increment (record every x step [b])	10
Number of concurrent chains (replicates) [c]	2
Visited (sampled) parameter values [a*b*c]	100000
Number of discard trees per chain (burn-in)	5000

Multiple Markov chains:

Static heating scheme

1000000.00	4 chains with	temperatures
	3.00	1.50 1.00
	Swapping interval is 1	

Print options:

Data file:	infile.seq
Output file:	outfile-bayes1

Posterior distribution raw histogram file:

bayesfile

Print data:

No

Print genealogies [only some for some data type]:

None

*Data summary*

Datatype:	Sequence data
Number of loci:	2

Population	Locus	Gene copies
1 Africa	1	25
	2	25
2 Americas	1	25
	2	25
3 Pacific	1	25
	2	25
4 Asia	1	25
	2	25
Total of all populations	1	100
	2	100

## *Bayesian Analysis: Posterior distribution table*

Locus	Parameter	2.5%	25.0%	Mode	75.0%	97.5%	Median	Mean
1	$\Theta_1$	0.00740	0.01000	0.01170	0.01320	0.01660	0.01210	0.01204
1	$\Theta_2$	0.00740	0.01000	0.01170	0.01320	0.01660	0.01210	0.01204
1	$\Theta_3$	0.00740	0.01000	0.01170	0.01320	0.01660	0.01210	0.01204
1	$\Theta_4$	0.00740	0.01000	0.01170	0.01320	0.01660	0.01210	0.01204
1	$M_{2 \rightarrow 1}$	0.0	20.0	110.0	180.0	280.0	150.0	119.2
1	$M_{3 \rightarrow 1}$	0.0	20.0	110.0	180.0	280.0	150.0	119.2
1	$M_{4 \rightarrow 1}$	0.0	20.0	110.0	180.0	280.0	150.0	119.2
1	$M_{1 \rightarrow 2}$	0.0	20.0	110.0	180.0	280.0	150.0	119.2
1	$M_{3 \rightarrow 2}$	0.0	20.0	110.0	180.0	280.0	150.0	119.2
1	$M_{4 \rightarrow 2}$	0.0	20.0	110.0	180.0	280.0	150.0	119.2
1	$M_{1 \rightarrow 3}$	0.0	20.0	110.0	180.0	280.0	150.0	119.2
1	$M_{2 \rightarrow 3}$	0.0	20.0	110.0	180.0	280.0	150.0	119.2
1	$M_{4 \rightarrow 3}$	0.0	20.0	110.0	180.0	280.0	150.0	119.2
1	$M_{1 \rightarrow 4}$	0.0	20.0	110.0	180.0	280.0	150.0	119.2
1	$M_{2 \rightarrow 4}$	0.0	20.0	110.0	180.0	280.0	150.0	119.2
1	$M_{3 \rightarrow 4}$	0.0	20.0	110.0	180.0	280.0	150.0	119.2
2	$\Theta_1$	0.00000	0.00100	0.00210	0.00280	0.00380	0.00230	0.00203
2	$\Theta_2$	0.00000	0.00100	0.00210	0.00280	0.00380	0.00230	0.00203
2	$\Theta_3$	0.00000	0.00100	0.00210	0.00280	0.00380	0.00230	0.00203
2	$\Theta_4$	0.00000	0.00100	0.00210	0.00280	0.00380	0.00230	0.00203
2	$M_{2 \rightarrow 1}$	460.0	720.0	910.0	1080.0	1380.0	930.0	915.9
2	$M_{3 \rightarrow 1}$	460.0	720.0	910.0	1080.0	1380.0	930.0	915.9
2	$M_{4 \rightarrow 1}$	460.0	720.0	910.0	1080.0	1380.0	930.0	915.9
2	$M_{1 \rightarrow 2}$	460.0	720.0	910.0	1080.0	1380.0	930.0	915.9
2	$M_{3 \rightarrow 2}$	460.0	720.0	910.0	1080.0	1380.0	930.0	915.9
2	$M_{4 \rightarrow 2}$	460.0	720.0	910.0	1080.0	1380.0	930.0	915.9
2	$M_{1 \rightarrow 3}$	460.0	720.0	910.0	1080.0	1380.0	930.0	915.9
2	$M_{2 \rightarrow 3}$	460.0	720.0	910.0	1080.0	1380.0	930.0	915.9
2	$M_{4 \rightarrow 3}$	460.0	720.0	910.0	1080.0	1380.0	930.0	915.9
2	$M_{1 \rightarrow 4}$	460.0	720.0	910.0	1080.0	1380.0	930.0	915.9
2	$M_{2 \rightarrow 4}$	460.0	720.0	910.0	1080.0	1380.0	930.0	915.9
2	$M_{3 \rightarrow 4}$	460.0	720.0	910.0	1080.0	1380.0	930.0	915.9
All	$\Theta_1$	0.00000	0.00000	0.00150	0.00300	0.00400	0.00330	0.00677
All	$\Theta_2$	0.00000	0.00000	0.00150	0.00300	0.00400	0.00330	0.00677

Locus	Parameter	2.5%	25.0%	Mode	75.0%	97.5%	Median	Mean
All	$\Theta_3$	0.00000	0.00000	0.00150	0.00300	0.00400	0.00330	0.00677
All	$\Theta_4$	0.00000	0.00000	0.00150	0.00300	0.00400	0.00330	0.00677
All	$M_{2 \rightarrow 1}$	480.0	700.0	910.0	1100.0	1420.0	910.0	861.0
All	$M_{3 \rightarrow 1}$	480.0	700.0	910.0	1100.0	1420.0	910.0	861.0
All	$M_{4 \rightarrow 1}$	480.0	700.0	910.0	1100.0	1420.0	910.0	861.0
All	$M_{1 \rightarrow 2}$	480.0	700.0	910.0	1100.0	1420.0	910.0	861.0
All	$M_{3 \rightarrow 2}$	480.0	700.0	910.0	1100.0	1420.0	910.0	861.0
All	$M_{4 \rightarrow 2}$	480.0	700.0	910.0	1100.0	1420.0	910.0	861.0
All	$M_{1 \rightarrow 3}$	480.0	700.0	910.0	1100.0	1420.0	910.0	861.0
All	$M_{2 \rightarrow 3}$	480.0	700.0	910.0	1100.0	1420.0	910.0	861.0
All	$M_{4 \rightarrow 3}$	480.0	700.0	910.0	1100.0	1420.0	910.0	861.0
All	$M_{1 \rightarrow 4}$	480.0	700.0	910.0	1100.0	1420.0	910.0	861.0
All	$M_{2 \rightarrow 4}$	480.0	700.0	910.0	1100.0	1420.0	910.0	861.0
All	$M_{3 \rightarrow 4}$	480.0	700.0	910.0	1100.0	1420.0	910.0	861.0

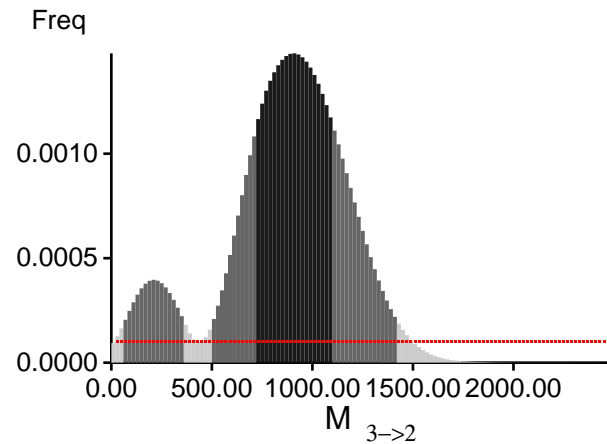
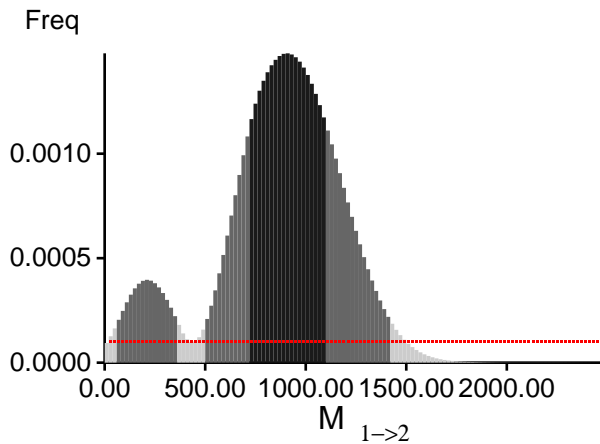
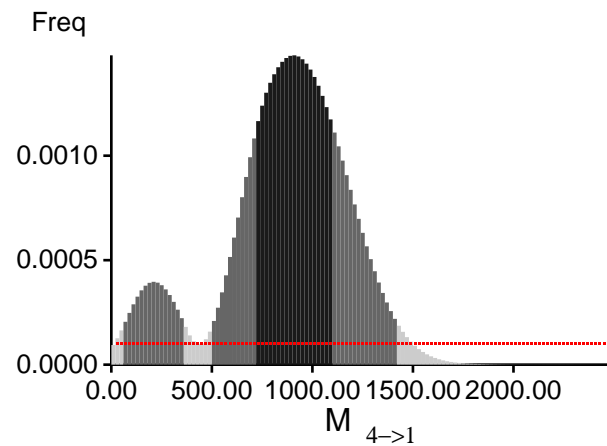
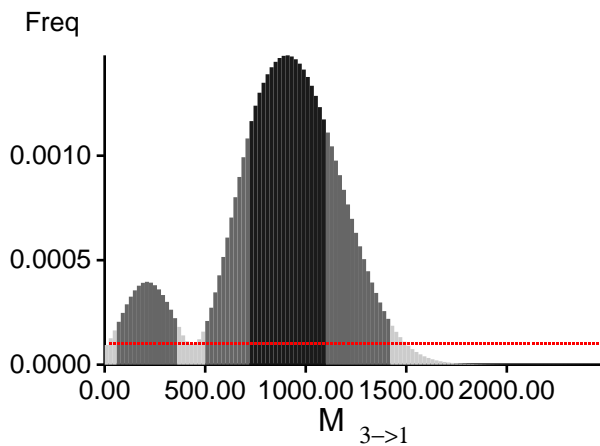
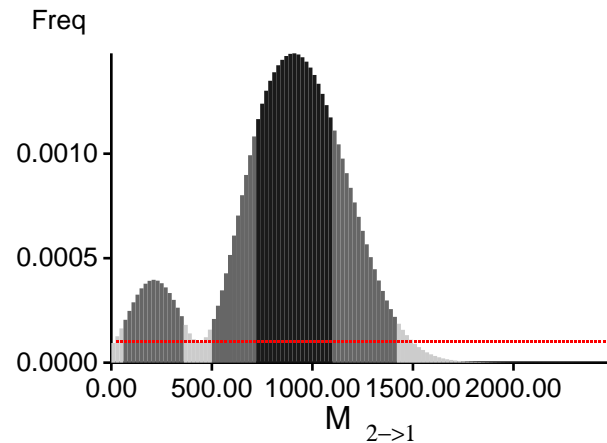
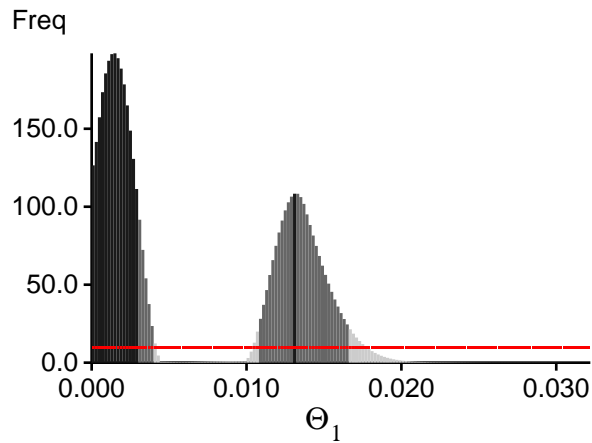
## Citation suggestions:

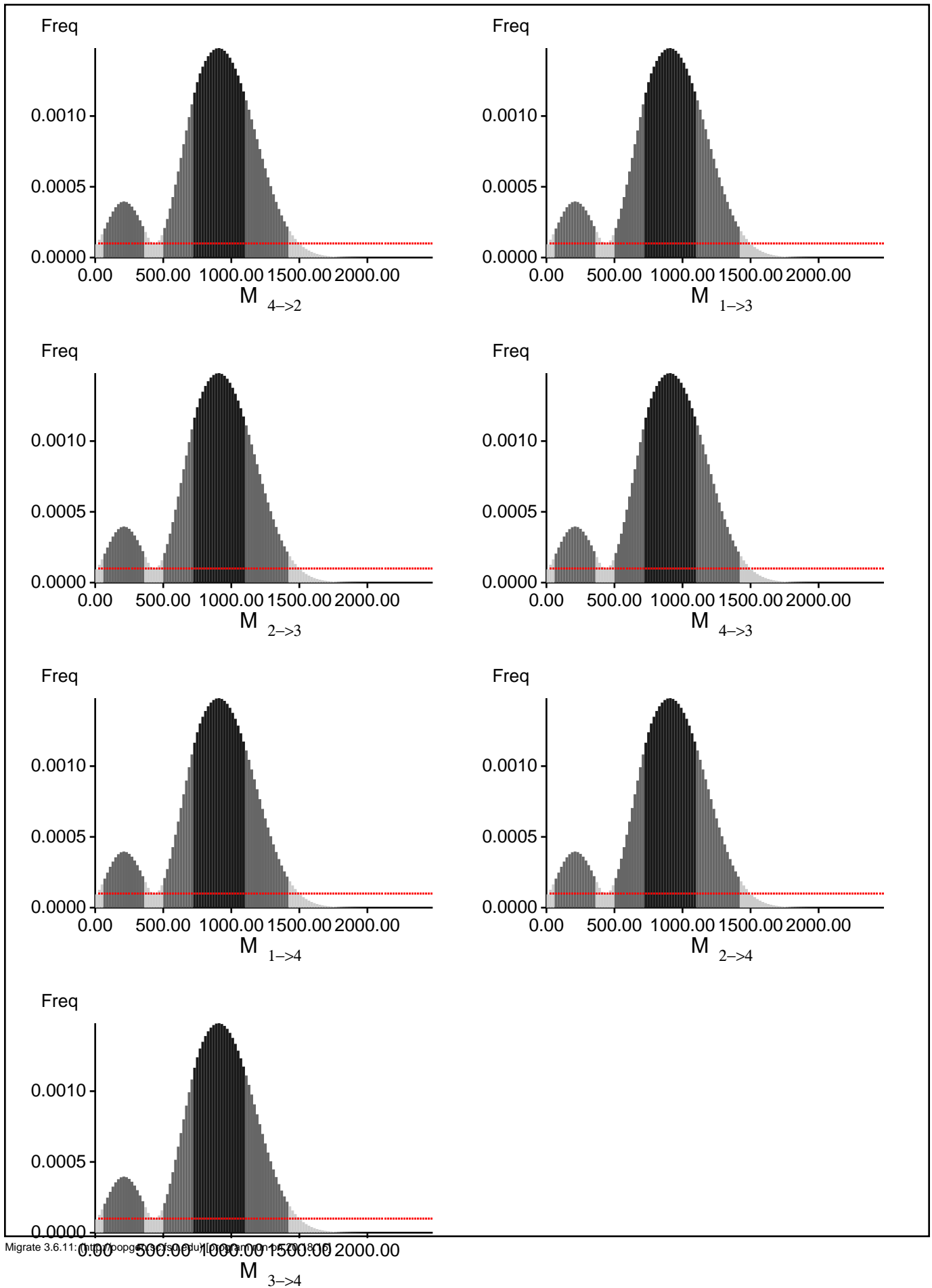
Beerli P., 2006. Comparison of Bayesian and maximum-likelihood inference of population genetic parameters.  
*Bioinformatics* 22:341-345

Beerli P., 2007. Estimation of the population scaled mutation rate from microsatellite data,  
*Genetics*, 177:1967-1968.

Beerli P., 2009. How to use MIGRATE or why are Markov chain Monte Carlo programs difficult to use?  
 In *Population Genetics for Animal Conservation*, G. Bertorelle, M. W. Bruford, H. C. Hauffe, A. Rizzoli,  
 and C. Vernesi, eds., vol. 17 of *Conservation Biology*, Cambridge University Press, Cambridge UK, pp. 42-79.

# *Bayesian Analysis: Posterior distribution over all loci*







## *Log-Probability of the data given the model (marginal likelihood)*

Use this value for Bayes factor calculations:

$BF = \text{Exp}[\ln(\text{Prob}(D \mid \text{thisModel})) - \ln(\text{Prob}(D \mid \text{otherModel}))]$

or as  $LBF = 2 (\ln(\text{Prob}(D \mid \text{thisModel})) - \ln(\text{Prob}(D \mid \text{otherModel})))$

shows the support for thisModel]

Locus	Raw thermodynamic score(1a)	Bezier approximation score(1b)	Harmonic mean(2)
1	-921.37	-533.88	-431.62
2	-847.12	-669.25	-589.41
All	-1767.65	-1202.28	-1020.18

(1a, 1b and 2) are approximations to the marginal likelihood, make sure that the program run long enough!

(1a, 1b) and (2) should give similar results, in principle.

But (2) is overestimating the likelihood, it is presented for historical reasons and should not be used

(1a, 1b) needs heating with chains that span a temperature range of 1.0 to at least 100,000.

(1b) is using a Bezier-curve to get better approximations for runs with low number of heated chains

[Scaling factor = 0.847480

Citation suggestions:

Beerli P. and M. Palczewski, 2010. Unified framework to evaluate panmixia and migration direction among multiple sampling locations, *Genetics*, 185: 313-326.

## *Acceptance ratios for all parameters and the genealogies*

Parameter	Accepted changes	Ratio
$\Theta_1$	6393/6393	1.00000
$\Theta_2$	6393/6393	1.00000
$\Theta_3$	6393/6393	1.00000
$\Theta_4$	6393/6393	1.00000
M <sub>2→1</sub>	6394/6394	1.00000
M <sub>3→1</sub>	6239/6239	1.00000
M <sub>4→1</sub>	6140/6140	1.00000
M <sub>1→2</sub>	6247/6247	1.00000
M <sub>3→2</sub>	6318/6318	1.00000
M <sub>4→2</sub>	6177/6177	1.00000
M <sub>1→3</sub>	6124/6124	1.00000
M <sub>2→3</sub>	6253/6253	1.00000
M <sub>4→3</sub>	6186/6186	1.00000
M <sub>1→4</sub>	6221/6221	1.00000
M <sub>2→4</sub>	6240/6240	1.00000
M <sub>3→4</sub>	6275/6275	1.00000
Genealogies	17893/100113	0.17873

## *MCMC-Autocorrelation and Effective MCMC Sample Size*

Parameter	Autocorrelation	Effective Sample Size
$\Theta_1$	0.88543	1215.96
$\Theta_2$	0.88543	1215.96
$\Theta_3$	0.88543	1215.96
$\Theta_4$	0.88543	1215.96
$M_{2 \rightarrow 1}$	0.60724	5163.47
$M_{3 \rightarrow 1}$	0.60724	5163.47
$M_{4 \rightarrow 1}$	0.60724	5163.47
$M_{1 \rightarrow 2}$	0.60724	5163.47
$M_{3 \rightarrow 2}$	0.60724	5163.47
$M_{4 \rightarrow 2}$	0.60724	5163.47
$M_{1 \rightarrow 3}$	0.60724	5163.47
$M_{2 \rightarrow 3}$	0.60724	5163.47
$M_{4 \rightarrow 3}$	0.60724	5163.47
$M_{1 \rightarrow 4}$	0.60724	5163.47
$M_{2 \rightarrow 4}$	0.60724	5163.47
$M_{3 \rightarrow 4}$	0.60724	5163.47
$\text{Ln}[\text{Prob}(D G)]$	0.84962	1641.47

## *Potential Problems*

This section reports potential problems with your run, but such reporting is often not very accurate. Whith many parameters in a multilocus analysis, it is very common that some parameters for some loci will not be very informative, triggering suggestions (for example to increase the prior range) that are not sensible. This suggestion tool will improve with time, therefore do not blindly follow its suggestions. If some parameters are flagged, inspect the tables carefully and judge wether an action is required. For example, if you run a Bayesian inference with sequence data, for macroscopic species there is rarely the need to increase the prior for Theta beyond 0.1; but if you use microsatellites it is rather common that your prior distribution for Theta should have a range from 0.0 to 100 or more. With many populations ( $>3$ ) it is also very common that some migration routes are estimated poorly because the data contains little or no information for that route. Increasing the range will not help in such situations, reducing number of parameters may help in such situations.

No warning was recorded during the run