

# MIGRATE-N Analysis for the mtCR of *H. francisci*

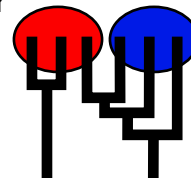
POPULATION SIZE, MIGRATION, DIVERGENCE, ASSIGNMENT, HISTORY

Bayesian inference using the structured coalescent

Migrate-n version 4.4.4(git:) [June-1-2019]

Program started at Sat Nov 27 13:45:03 2021

Program finished at Sat Nov 27 14:58:44 2021 [Runtime:0000:01:13:41]



## Options

Datatype:

DNA sequence data

Inheritance scalers in use for Thetas:

0.00

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

Data set was subsampled: used a random sample of size:

30 and seed 2542345

Random number seed:

(with internal timer) 2599591461

Start parameters:

Theta values were generated

Using a percent value of the prior

M values were generated

Using a percent value of the prior

Connection matrix:

m = average (average over a group of Thetas or M,

s = symmetric migration M, S = symmetric 4Nm,

0 = zero, and not estimated,

\* = migration free to vary, Thetas are on diagonal

d = row population split off column population, D = split and then migration

Population	1	1	1	2	3	4	5
1 Anacapa_Island	*	*	*	*	*	*	*
1 Santa_Cruz_Isla	*	*	*	*	*	*	*
1 Catalina_Island	*	*	*	*	*	*	*
2 California_Main	*	*	*	*	*	*	*
3 Bahia_Tortugas	*	*	*	*	*	*	*

4 LSIBM \* \* \* \* \* \* \*  
 5 Las\_Animas \* \* \* \* \* \* \*

## Order of parameters:

1	$\Theta_1$	<displayed>
2	$\Theta_2$	<displayed>
3	$\Theta_3$	<displayed>
4	$\Theta_4$	<displayed>
5	$\Theta_5$	<displayed>
6	M <sub>2→1</sub>	<displayed>
7	M <sub>3→1</sub>	<displayed>
8	M <sub>4→1</sub>	<displayed>
9	M <sub>5→1</sub>	<displayed>
10	M <sub>1→2</sub>	<displayed>
11	M <sub>3→2</sub>	<displayed>
12	M <sub>4→2</sub>	<displayed>
13	M <sub>5→2</sub>	<displayed>
14	M <sub>1→3</sub>	<displayed>
15	M <sub>2→3</sub>	<displayed>
16	M <sub>4→3</sub>	<displayed>
17	M <sub>5→3</sub>	<displayed>
18	M <sub>1→4</sub>	<displayed>
19	M <sub>2→4</sub>	<displayed>
20	M <sub>3→4</sub>	<displayed>
21	M <sub>5→4</sub>	<displayed>
22	M <sub>1→5</sub>	<displayed>
23	M <sub>2→5</sub>	<displayed>
24	M <sub>3→5</sub>	<displayed>
25	M <sub>4→5</sub>	<displayed>

Mutation rate among loci:

Mutation rate is constant

Analysis strategy:

Bayesian inference

-Population size estimation:

Exponential Distribution

-Geneflow estimation:

Exponential Distribution

## Proposal distributions for parameter

Parameter	Proposal
Theta	Metropolis sampling
M	Metropolis sampling
Divergence	Metropolis sampling
Divergence Spread	Metropolis sampling
Genealogy	Metropolis-Hastings

## Prior distribution for parameter

Parameter		Prior	Minimum	Mean	Maximum	Delta	Bins	UpdateFreq
1	Theta	** Uniform	0.000000	0.050	0.100	0.010	5000	0.02000
2	Theta	** Uniform	0.000000	0.050	0.100	0.010	5000	0.02000
3	Theta	** Uniform	0.000000	0.050	0.100	0.010	5000	0.02000
4	Theta	** Uniform	0.000000	0.050	0.100	0.010	5000	0.02000
5	Theta	** Uniform	0.000000	0.050	0.100	0.010	5000	0.02000
6	M	** Uniform	0.000000	500.0	1000.	100.0	1500	0.02000
7	M	** Uniform	0.000000	500.0	1000.	100.0	1500	0.02000
8	M	** Uniform	0.000000	500.0	1000.	100.0	1500	0.02000
9	M	** Uniform	0.000000	500.0	1000.	100.0	1500	0.02000
10	M	** Uniform	0.000000	500.0	1000.	100.0	1500	0.02000
11	M	** Uniform	0.000000	500.0	1000.	100.0	1500	0.02000
12	M	** Uniform	0.000000	500.0	1000.	100.0	1500	0.02000
13	M	** Uniform	0.000000	500.0	1000.	100.0	1500	0.02000
14	M	** Uniform	0.000000	500.0	1000.	100.0	1500	0.02000
15	M	** Uniform	0.000000	500.0	1000.	100.0	1500	0.02000
16	M	** Uniform	0.000000	500.0	1000.	100.0	1500	0.02000
17	M	** Uniform	0.000000	500.0	1000.	100.0	1500	0.02000
18	M	** Uniform	0.000000	500.0	1000.	100.0	1500	0.02000
19	M	** Uniform	0.000000	500.0	1000.	100.0	1500	0.02000
20	M	** Uniform	0.000000	500.0	1000.	100.0	1500	0.02000
21	M	** Uniform	0.000000	500.0	1000.	100.0	1500	0.02000
22	M	** Uniform	0.000000	500.0	1000.	100.0	1500	0.02000
23	M	** Uniform	0.000000	500.0	1000.	100.0	1500	0.02000
24	M	** Uniform	0.000000	500.0	1000.	100.0	1500	0.02000
25	M	** Uniform	0.000000	500.0	1000.	100.0	1500	0.02000

[\* \* means priors were set globally]

## Markov chain settings:

Number of chains	1	Long chain
Recorded steps [a]	10000	
Increment (record every x step [b])	1000	
Number of concurrent chains (replicates) [c]	1	
Visited (sampled) parameter values [a*b*c]	10000000	
Number of discard trees per chain (burn-in)	2500	

## 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:

migrate-n\_heterodontus\_input\_noGI.txt

Haplotyping is turned on:	NO
Output file:	migrate-n_het_fullmodel_noGI_out
Posterior distribution raw histogram file:	bayesfile
Raw data from the MCMC run:	bayesallfile.gz
Print data:	No
Print genealogies [only some for some data type]:	None

## Data summary

Data file: migrate-n\_heterodontus\_input\_noGI.txt  
 Datatype: Sequence data  
 Number of loci: 1

### Mutationmodel:

Locus	Sublocus	Mutationmodel	Mutationmodel parameters
1	1	HKY	[Bf:0.31 0.21 0.14 0.33, kappa=2.000]

Data set was subsampled: used a random sample of size: 30

### Sites per locus

Locus	Sites
1	724

### Site rate variation and probabilities:

Locus	Sublocus	Region type	Rate of change	Probability	Patch size
1	1	1	0.428	0.680	1.000
1	1	2	2.081	0.308	1.000
1	1	3	5.491	0.012	1.000

Population	Locus	Gene copies
1 Anacapa_Island	1	25
1 Santa_Cruz_Island	1	26
1 Catalina_Island	1	29
2 California_Mainland_Coast	1	131
3 Bahia_Tortugas	1	30
4 LSIBM	1	54
5 Las_Animas	1	12
Total of all populations	1	307

## *Subsampled dataset*

Data set was subsampled randomly per population: 30 samples taken

Locus	Population	Individuals
1	Anacapa_Is	A-159 A-164 A-157 A-175 A-196 A-171 A-160 A-167 A-174 A-156 A-193 A-161 A-162 A-173 A-194 A-163 A-197 A-192 A-158 A-166 A-172 A-199 A-165 A-198 A-176
	Santa_Cruz	SC-182 SC-185 SC-203 SC-206 SC-155 SC-62 SC-63 SC-205 SC-169 SC-150 SC-151 SC-186 SC-91 SC-152 SC-207 SC-184 SC-64 SC-154 SC-204 SC-90 SC-153 SC-209 SC-168 SC-187 SC-183 SC-188
	Catalina_I	C-F35 C-F30 C-F17 C-F25 C-F29 C-F44 C-F4 C-F10 C-F28 C-F45 C-F34 C-F198 C-F11 C-F53 C-F197 C-F48 C-F8 C-F43 C-F12 C-F177 C-F47 C-F14 C-F21 C-F13 C-F16 C-F27 C-F26 C-F32 C-F54
	California	MLCA-F230 MLCA-85 MLCA-F146 MLCA-F83 MLCA-F223 MLCA-F301 MLCA-F178 MLCA-F91 MLCA-F261 MLCA-F268 MLCA-F154 MLCA-F61 MLCA-F186 MLCA-F96 MLCA-75 MLCA-F131 MLCA-F77 MLCA-F218 MLCA-F69 MLCA-F86 MLCA-F176 MLCA-F147 MLCA-F293 MLCA-F241 MLCA-F294 MLCA-F152 MLCA-82 MLCA-F217 MLCA-68 MLCA-F94
	Bahia_Tort	BT-HF46 BT-HF41 BT-HF42 BT-HF36 BT-HF43 BT-HF24 BT-HF27 BT-HF30 BT-HF33 BT-HF39 BT-HF38 BT-HF31 BT-HF32 BT-HF20 BT-HF26 BT-HF22 BT-HF37 BT-HF35 BT-HF45 BT-HF23 BT-HF19 BT-HF29 BT-HF34 BT-HF28 BT-HF47 BT-HF44 BT-HF48 BT-HF25 BT-HF21 BT-HF40
	LSIBM	LSIBM-478 LSIBM-389 LSIBM-480 LSIBM-384 LSIBM-410 LSIBM-684 LSIBM-476 LSIBM-672 LSIBM-398 LSIBM-479 LSIBM-409 LSIBM-387 LSIBM-397 LSIBM-315 LSIBM-400 LSIBM-406 LSIBM-383 LSIBM-381 LSIBM-408 LSIBM-407 LSIBM-675 LSIBM-674 LSIBM-475 LSIBM-321 LSIBM-404 LSIBM-403 LSIBM-441 LSIBM-385 LSIBM-401 LSIBM-399
	Las_Animas	LA-655 LA-645 LA-651 LA-653 LA-646 LA-652 LA-650 LA-649 LA-647 LA-648 LA-654 LA-656

## *Bayesian Analysis: Posterior distribution table*

Locus	Parameter	2.5%	25.0%	Mode	75.0%	97.5%	Median	Mean
1	$\Theta_1$	0.00000	0.00000	0.00001	0.00092	0.00264	0.00093	0.00000
1	$\Theta_2$	0.00000	0.00000	0.00001	0.00092	0.00264	0.00093	0.00000
1	$\Theta_3$	0.00000	0.00000	0.00001	0.00092	0.00264	0.00093	0.00000
1	$\Theta_4$	0.00000	0.00000	0.00001	0.00092	0.00264	0.00093	0.00000
1	$\Theta_5$	0.00000	0.00000	0.00001	0.00092	0.00264	0.00093	0.00000
1	$M_{2 \rightarrow 1}$	0.0	166.7	291.0	404.0	500.7	268.3	526.7
1	$M_{3 \rightarrow 1}$	0.0	0.0	0.3	171.3	461.3	171.7	315.4
1	$M_{4 \rightarrow 1}$	0.0	0.0	0.3	204.7	480.7	205.0	348.2
1	$M_{5 \rightarrow 1}$	0.0	0.0	0.3	123.3	408.0	123.7	176.1
1	$M_{1 \rightarrow 2}$	81.3	294.7	389.0	465.3	576.0	357.7	633.3
1	$M_{3 \rightarrow 2}$	0.0	0.0	0.3	88.7	356.7	244.3	552.2
1	$M_{4 \rightarrow 2}$	0.0	0.0	0.3	209.3	487.3	209.7	484.4
1	$M_{5 \rightarrow 2}$	0.0	0.0	0.3	16.0	16.0	167.7	272.2
1	$M_{1 \rightarrow 3}$	0.0	0.0	0.3	158.7	453.3	159.0	275.2
1	$M_{2 \rightarrow 3}$	0.0	0.0	0.3	201.3	485.3	201.7	521.9
1	$M_{4 \rightarrow 3}$	0.0	236.7	384.3	461.3	592.0	318.3	672.9
1	$M_{5 \rightarrow 3}$	0.0	0.0	55.0	115.3	115.3	195.0	339.8
1	$M_{1 \rightarrow 4}$	0.0	129.3	227.0	368.0	503.3	258.3	424.2
1	$M_{2 \rightarrow 4}$	0.0	0.0	1.0	36.7	36.7	212.3	461.9
1	$M_{3 \rightarrow 4}$	0.0	156.7	247.7	397.3	458.0	259.7	511.8
1	$M_{5 \rightarrow 4}$	0.0	0.0	0.3	183.3	469.3	183.7	346.2
1	$M_{1 \rightarrow 5}$	0.0	0.0	0.3	209.3	486.7	209.7	419.6
1	$M_{2 \rightarrow 5}$	0.0	210.0	380.3	446.0	568.0	287.7	627.9
1	$M_{3 \rightarrow 5}$	0.0	196.0	379.7	443.3	486.0	275.7	609.1
1	$M_{4 \rightarrow 5}$	2.7	206.7	380.3	450.7	469.3	281.7	640.3

### Citation suggestions:

Beerli P., 2006. Comparison of Bayesian and maximum-likelihood inference of population genetic parameters.

Bioinformatics 22:341-345

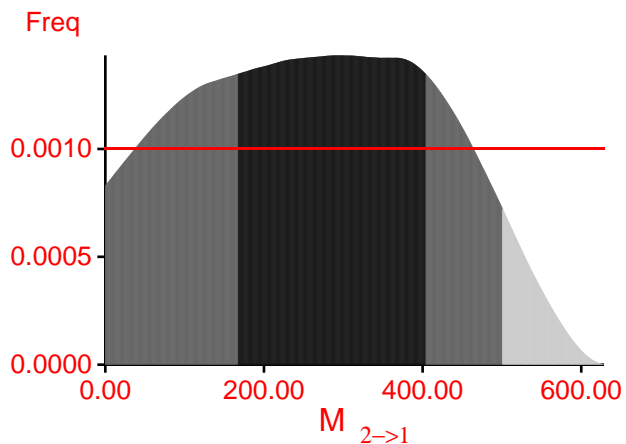
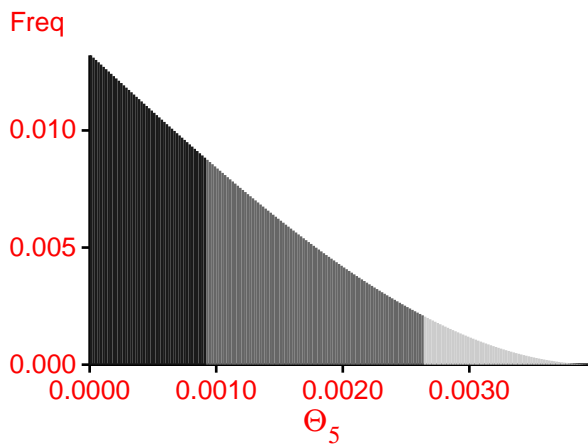
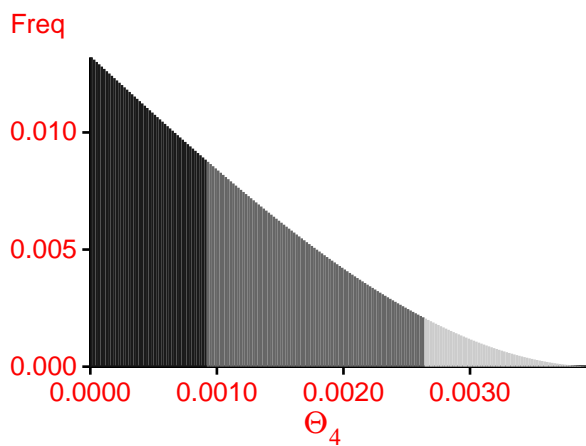
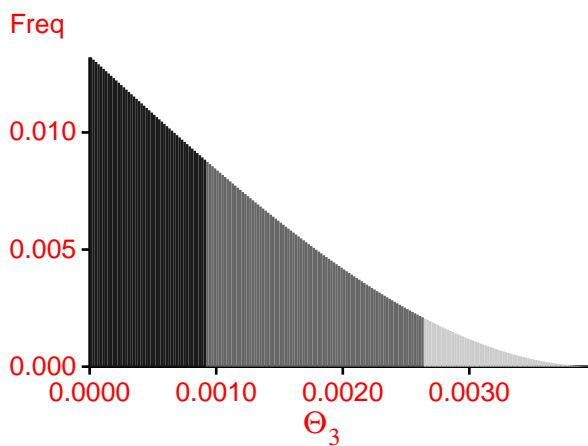
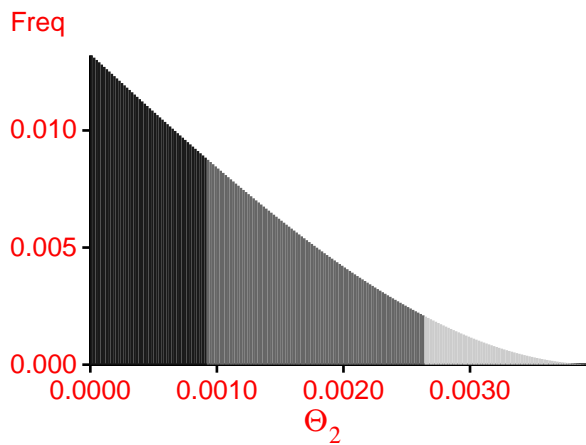
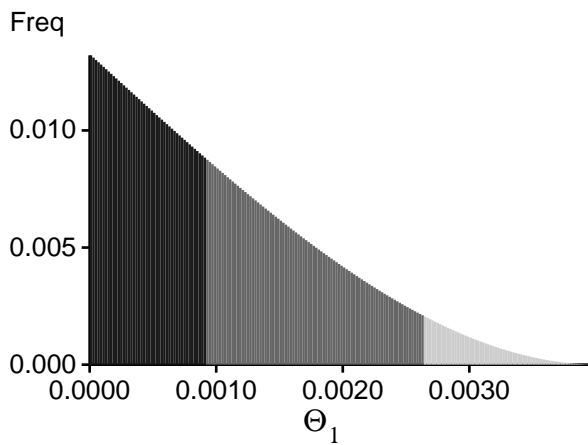
Beerli P., 2007. Estimation of the population scaled mutation rate from microsatellite data,

Genetics, 177:1967-1968.

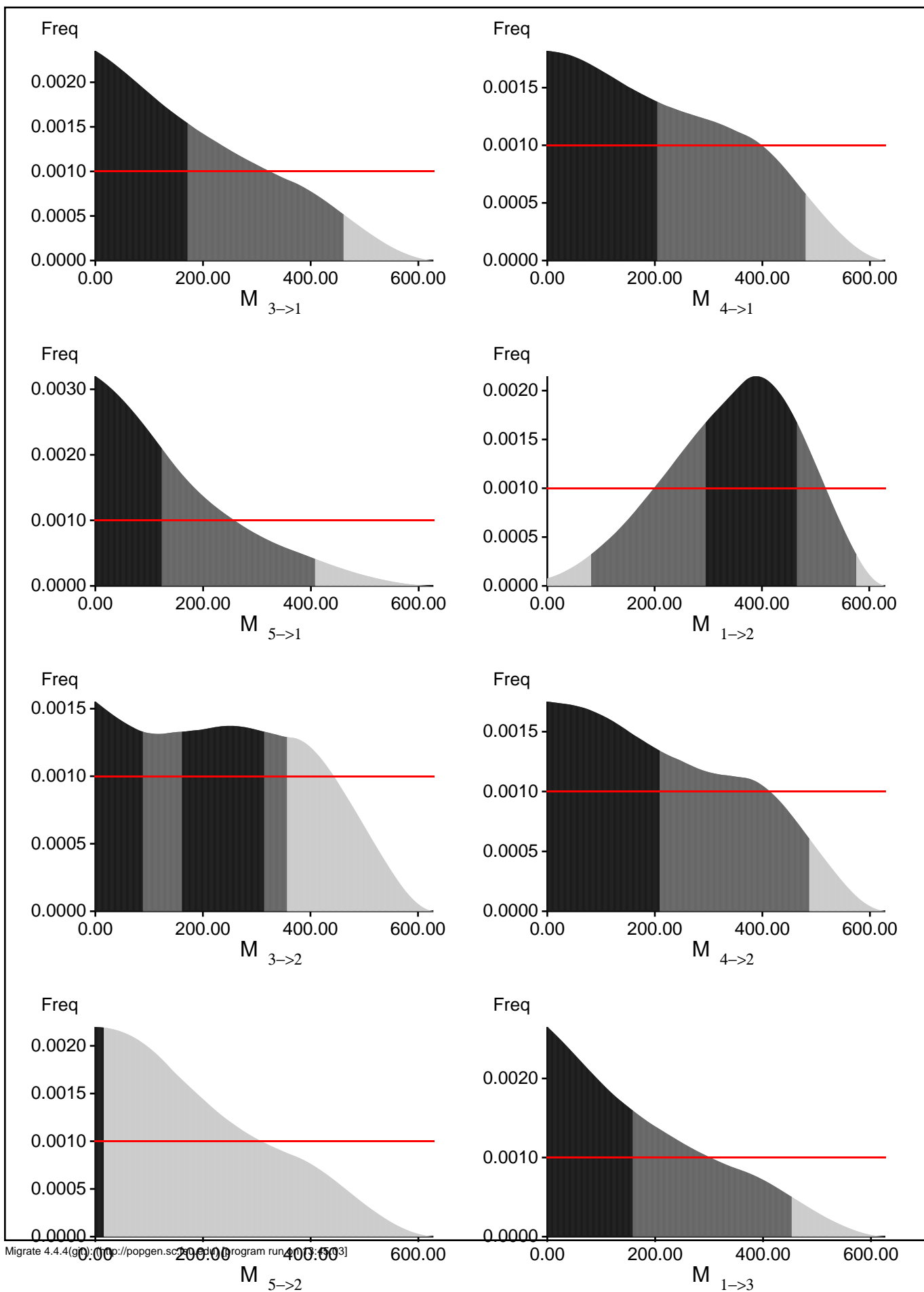
Beerli P., 2009. How to use MIGRATE or why are Markov chain Monte Carlo programs difficult to use?

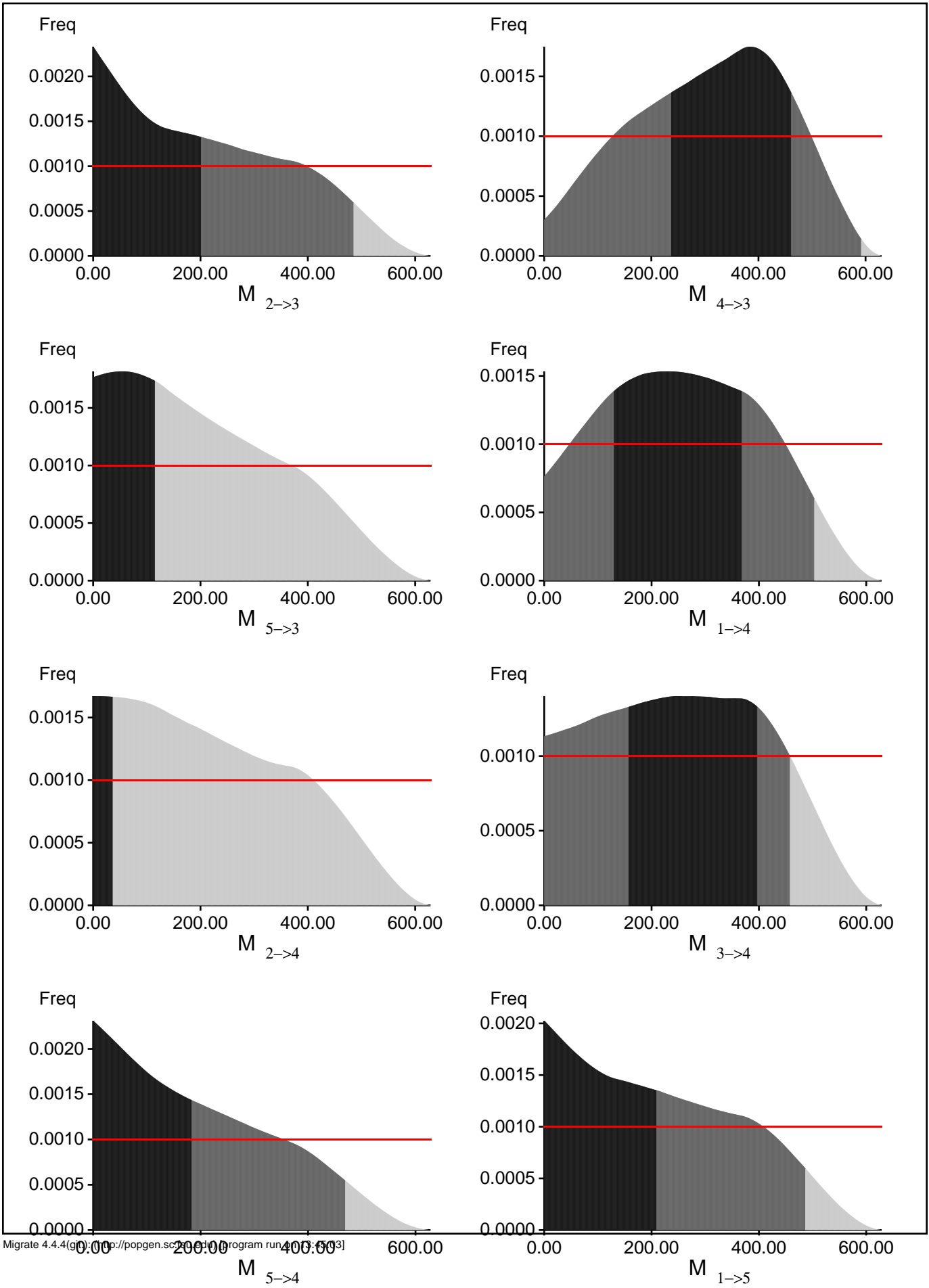
In Population Genetics for Animal Conservation, G. Bertorelle, M. W. Bruford, H. C. Hauffe, A. Rizzoli, and C. Vernesi, eds., vol. 17 of Conservation Biology, Cambridge University Press, Cambridge UK, pp. 42-79.

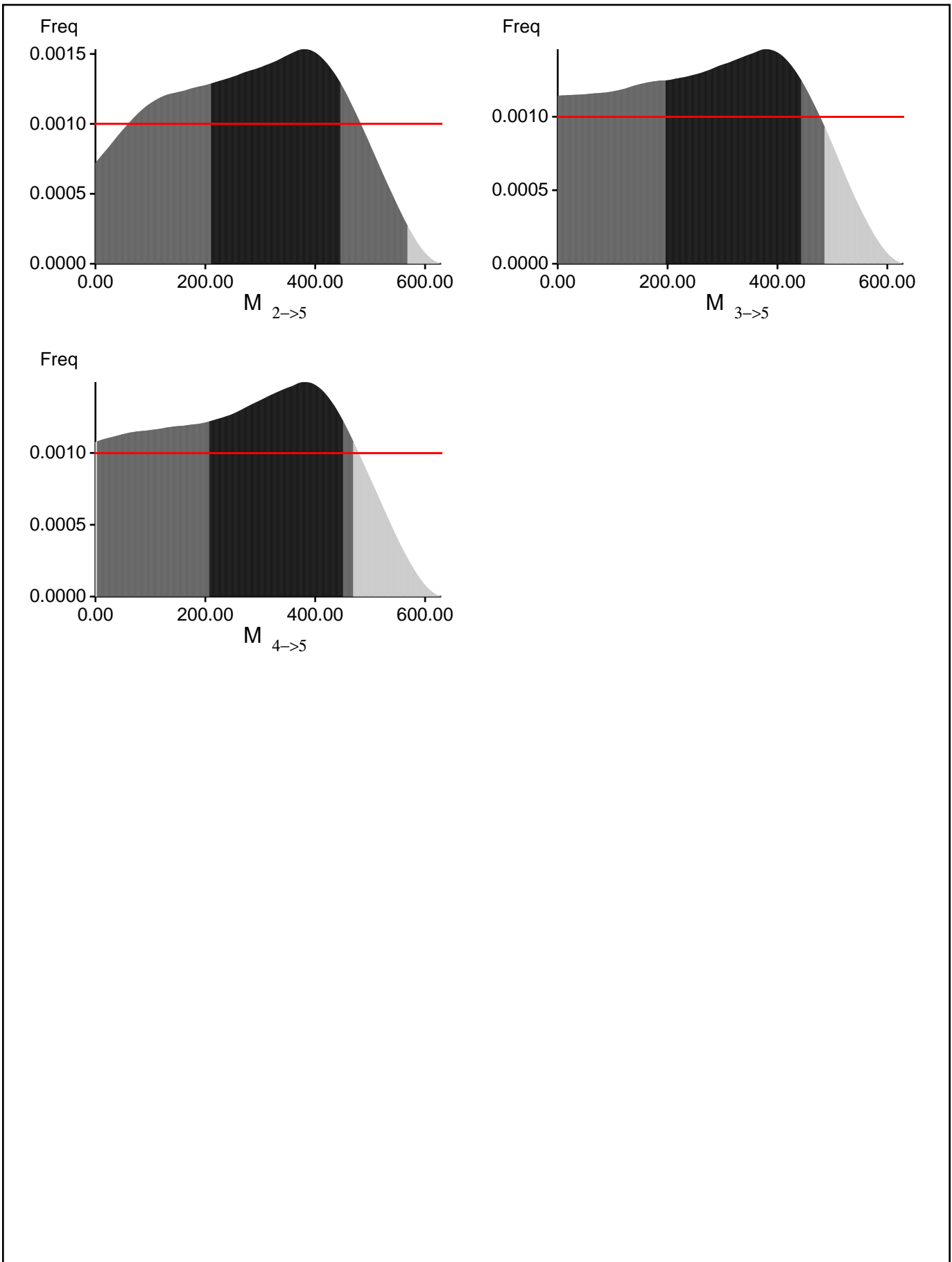
*Bayesian Analysis: Posterior distribution for locus 1*











## *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]

Method	$\ln(\text{Prob}(D \mid \text{Model}))$	Notes
Thermodynamic integration	-1538.735632	(1a)
	-1318.051128	(1b)
Harmonic mean	-1181.923326	(2)

(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

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
$\Theta_1$	92530/200084	0.46246
$\Theta_2$	106503/199864	0.53288
$\Theta_3$	23027/199638	0.11534
$\Theta_4$	42395/200506	0.21144
$\Theta_5$	48338/199270	0.24258
M <sub>2→1</sub>	125608/199664	0.62910
M <sub>3→1</sub>	111477/199162	0.55973
M <sub>4→1</sub>	124819/200160	0.62360
M <sub>5→1</sub>	95160/199646	0.47664
M <sub>1→2</sub>	151184/200018	0.75585
M <sub>3→2</sub>	136463/199531	0.68392
M <sub>4→2</sub>	138840/199920	0.69448
M <sub>5→2</sub>	131380/199254	0.65936
M <sub>1→3</sub>	131375/199737	0.65774
M <sub>2→3</sub>	132529/200177	0.66206
M <sub>4→3</sub>	134197/199519	0.67260
M <sub>5→3</sub>	138632/200877	0.69013
M <sub>1→4</sub>	149932/199727	0.75068
M <sub>2→4</sub>	136217/200429	0.67963
M <sub>3→4</sub>	142218/199732	0.71204
M <sub>5→4</sub>	136059/200386	0.67898
M <sub>1→5</sub>	162113/200202	0.80975
M <sub>2→5</sub>	147548/199931	0.73799
M <sub>3→5</sub>	150029/200442	0.74849
M <sub>4→5</sub>	142707/199931	0.71378
Genealogies	1390032/5002193	0.27788

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

Parameter	Autocorrelation	Effective Sample Size
$\Theta_1$	0.53345	3042.19
$\Theta_2$	0.69627	1790.40
$\Theta_3$	0.62747	2288.78
$\Theta_4$	0.56854	2750.40
$\Theta_5$	0.65808	2061.91
M <sub>2→1</sub>	0.62559	2303.02
M <sub>3→1</sub>	0.62852	2280.87
M <sub>4→1</sub>	0.56700	2762.94
M <sub>5→1</sub>	0.35368	4774.10
M <sub>1→2</sub>	0.17370	7039.47
M <sub>3→2</sub>	0.54685	2929.17
M <sub>4→2</sub>	0.52764	3091.75
M <sub>5→2</sub>	0.20200	6638.26
M <sub>1→3</sub>	0.18125	6930.46
M <sub>2→3</sub>	0.58547	2614.26
M <sub>4→3</sub>	0.45476	3747.61
M <sub>5→3</sub>	0.33325	5000.42
M <sub>1→4</sub>	0.23586	6182.47
M <sub>2→4</sub>	0.53780	3005.25
M <sub>3→4</sub>	0.48154	3499.07
M <sub>5→4</sub>	0.38738	4415.22
M <sub>1→5</sub>	0.17691	6992.97
M <sub>2→5</sub>	0.37659	4528.25
M <sub>3→5</sub>	0.37434	4551.97
M <sub>4→5</sub>	0.42395	4045.00
Genealogies	0.83022	927.58

## *Potential Problems*

This section reports potential problems with your run, but such reporting is often not very accurate. With 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 whether 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