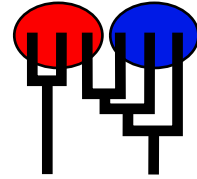


Example: sequence data set wit two loci [simulated data]

MIGRATION RATE AND POPULATION SIZE ESTIMATION
 using the coalescent and maximum likelihood or Bayesian inference
 Migrate-n version 3.5.2 [2152]
 Compiled for a SYMMETRIC MULTIPROCESSORS
 Program started at Wed Apr 24 15:38:15 2013
 Program finished at Wed Apr 24 15:38:16 2013



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]

Data set was subsampled: used a random sample of size: 5 and seed 13

Random number seed: (from parmfile) 310705631

Start parameters:

Theta values were generated RANDOM start value from U(min,msx)

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	*	0	*	*
2 Americas	*	*	*	*
3 Pacific	*	*	*	*
4 Asia	*	*	*	*

Order of parameters:

1	Θ_1	<displayed>
2	Θ_2	<displayed>
3	Θ_3	<displayed>
4	Θ_4	<displayed>
6	M _{3→1}	<displayed>
7	M _{4→1}	<displayed>
8	M _{1→2}	<displayed>
9	M _{3→2}	<displayed>
10	M _{4→2}	<displayed>
11	M _{1→3}	<displayed>
12	M _{2→3}	<displayed>
13	M _{4→3}	<displayed>
14	M _{1→4}	<displayed>
15	M _{2→4}	<displayed>
16	M _{3→4}	<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	10.000000	20.000000	2.000000	500
M	Uniform	0.000000	10.000000	20.000000	2.000000	500

Markov chain settings:

Long chain

Number of chains	1
Recorded steps [a]	500
Increment (record every x step [b])	1
Number of concurrent chains (replicates) [c]	2
Visited (sampled) parameter values [a*b*c]	1000
Number of discard trees per chain (burn-in)	100

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-bayes
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	5
	2	25
2 Americas	1	5
	2	25
3 Pacific	1	5
	2	25
4 Asia	1	5
	2	25
Total of all populations	1	20
	2	100

Subsampled dataset

Data set was subsampled randomly per population: 5 samples taken

Locus Population Individuals

1	Africa	0BAA 0BAN 0BAK 0BAT 0BAB
	Americas	1BAX 1BAK 1BBA 1BAU 1BAM
	Pacific	2BAP 2BAQ 2BAS 2BAV 2BAL
	Asia	3BAH 3BAR 3BAT 3BAS 3BAW
2	Africa	0BAA 0BAX 0BAC 0BAK 0BAM
	Americas	1BAC 1BAJ 1BAR 1BAB 1BAS
	Pacific	2BAM 2BAI 2BAP 2BAO 2BAQ
	Asia	3BAR 3BAC 3BAB 3BAM 3BAX

Bayesian Analysis: Posterior distribution table

Locus	Parameter	2.5%	25.0%	Mode	75.0%	97.5%	Median	Mean
1	Θ_1	0.00000	0.00000	0.02000	0.20000	0.68000	0.22000	0.26641
1	Θ_2	0.00000	0.00000	0.14000	0.44000	0.76000	0.46000	6.58620
1	Θ_3	0.00000	0.00000	0.02000	0.24000	1.44000	0.26000	5.44908
1	Θ_4	0.00000	0.00000	0.18000	0.32000	1.00000	0.34000	0.39966
1	$M_{3 \rightarrow 1}$	1.8	1.9	2.3	2.6	4.7	6.3	7.5
1	$M_{4 \rightarrow 1}$	0.0	0.9	1.2	1.5	1.8	2.5	6.7
1	$M_{1 \rightarrow 2}$	16.7	17.6	18.0	18.4	18.5	14.7	12.0
1	$M_{3 \rightarrow 2}$	5.2	5.4	5.8	6.1	6.2	6.9	7.7
1	$M_{4 \rightarrow 2}$	6.0	6.2	6.5	6.8	7.0	6.5	6.9
1	$M_{1 \rightarrow 3}$	1.4	1.5	2.0	2.4	3.1	4.8	5.9
1	$M_{2 \rightarrow 3}$	8.3	9.7	10.3	10.6	10.9	8.4	8.2
1	$M_{4 \rightarrow 3}$	13.8	13.9	14.3	14.6	15.5	14.0	13.8
1	$M_{1 \rightarrow 4}$	6.4	6.4	6.9	7.2	7.8	8.5	8.6
1	$M_{2 \rightarrow 4}$	0.0	0.0	0.3	0.8	2.7	1.3	2.2
1	$M_{3 \rightarrow 4}$	0.0	0.0	0.0	0.4	0.6	10.5	10.2
2	Θ_1	0.00000	0.00000	0.02000	0.16000	0.64000	0.18000	0.17854
2	Θ_2	0.00000	0.00000	0.02000	0.20000	0.72000	0.22000	1.05921
2	Θ_3	0.00000	0.00000	0.02000	0.16000	0.48000	0.18000	0.28333
2	Θ_4	0.00000	0.00000	0.14000	0.32000	1.12000	0.34000	0.34611
2	$M_{3 \rightarrow 1}$	6.8	7.1	7.6	8.1	8.2	7.2	7.7
2	$M_{4 \rightarrow 1}$	0.0	0.1	0.9	1.2	2.0	5.9	6.5
2	$M_{1 \rightarrow 2}$	0.0	0.6	1.1	1.6	2.7	2.1	4.5
2	$M_{3 \rightarrow 2}$	17.1	17.1	17.5	17.8	17.9	9.8	8.8
2	$M_{4 \rightarrow 2}$	0.0	0.0	0.3	1.0	1.3	5.0	6.9
2	$M_{1 \rightarrow 3}$	0.0	0.0	0.0	0.3	2.9	2.7	7.0
2	$M_{2 \rightarrow 3}$	0.0	0.8	1.3	1.8	1.9	3.8	4.8
2	$M_{4 \rightarrow 3}$	0.0	0.0	0.0	0.9	2.6	1.8	4.0
2	$M_{1 \rightarrow 4}$	0.0	0.0	0.5	0.8	2.8	2.4	5.3
2	$M_{2 \rightarrow 4}$	0.0	0.0	0.0	1.2	2.7	2.4	5.5
2	$M_{3 \rightarrow 4}$	16.3	16.5	17.0	17.4	17.5	10.3	10.0
All	Θ_1	0.00000	0.04000	0.22000	0.36000	0.56000	0.30000	0.21505
All	Θ_2	0.00000	0.04000	0.26000	0.40000	0.80000	0.34000	1.13590
All	Θ_3	0.00000	0.00000	0.18000	0.32000	0.72000	0.30000	0.97998
All	Θ_4	0.00000	0.08000	0.30000	0.44000	0.76000	0.38000	0.32552

Locus	Parameter	2.5%	25.0%	Mode	75.0%	97.5%	Median	Mean
All	$M_{3 \rightarrow 1}$	0.0	3.0	3.4	3.8	4.6	4.3	6.4
All	$M_{4 \rightarrow 1}$	0.0	0.1	1.0	1.4	3.6	2.1	4.3
All	$M_{1 \rightarrow 2}$	16.6	17.5	18.0	18.4	18.6	14.7	10.4
All	$M_{3 \rightarrow 2}$	0.0	0.0	0.3	0.6	4.4	5.9	7.3
All	$M_{4 \rightarrow 2}$	0.0	1.2	1.7	2.1	2.3	5.4	7.4
All	$M_{1 \rightarrow 3}$	0.0	1.5	1.9	2.4	2.9	2.5	5.0
All	$M_{2 \rightarrow 3}$	8.3	9.6	10.3	10.7	11.0	8.9	8.0
All	$M_{4 \rightarrow 3}$	0.0	0.0	0.4	0.9	2.7	1.6	4.4
All	$M_{1 \rightarrow 4}$	0.0	0.0	0.4	0.7	2.4	6.7	5.6
All	$M_{2 \rightarrow 4}$	0.0	0.0	0.4	0.6	2.7	0.6	2.0
All	$M_{3 \rightarrow 4}$	15.6	16.4	17.0	17.4	17.6	10.6	9.5

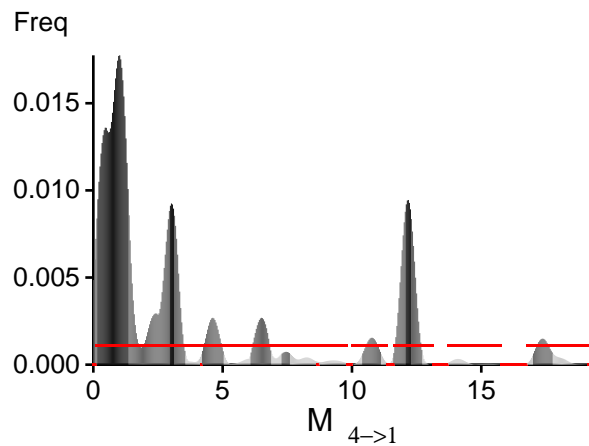
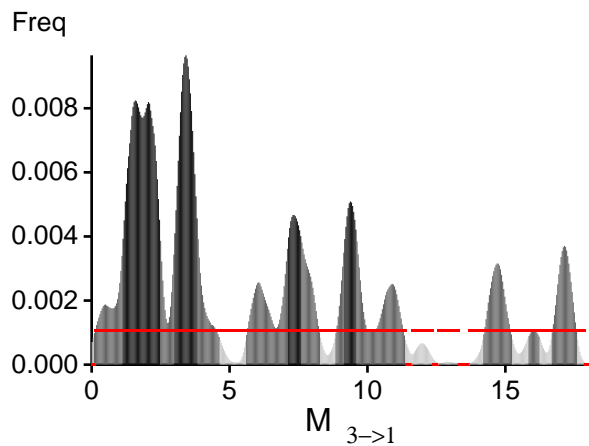
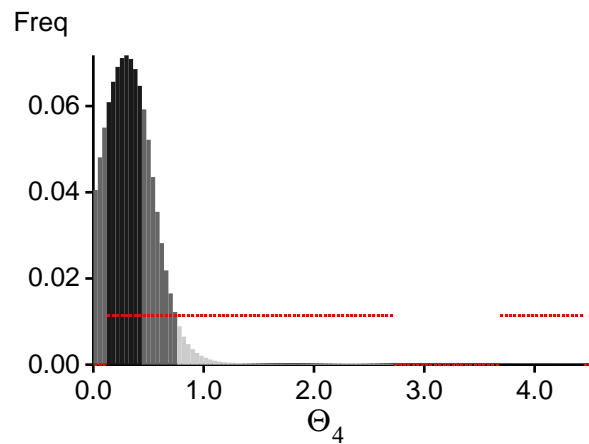
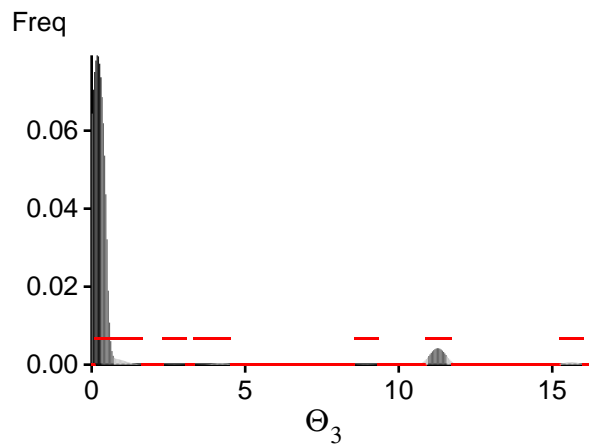
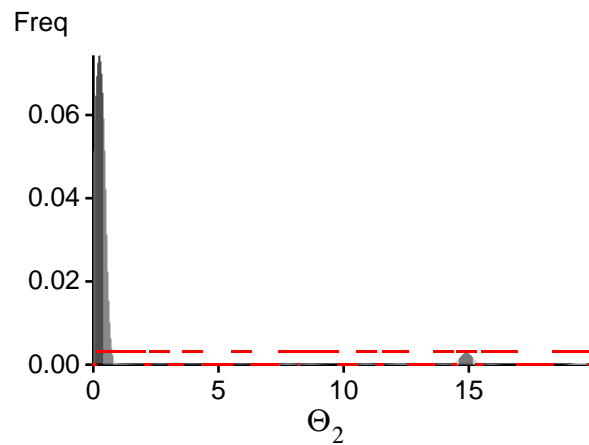
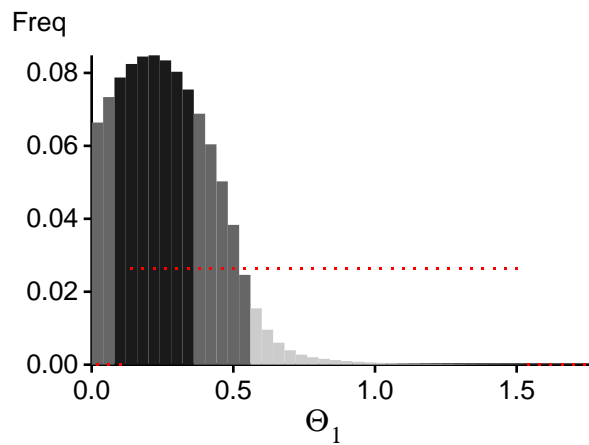
Citation suggestions:

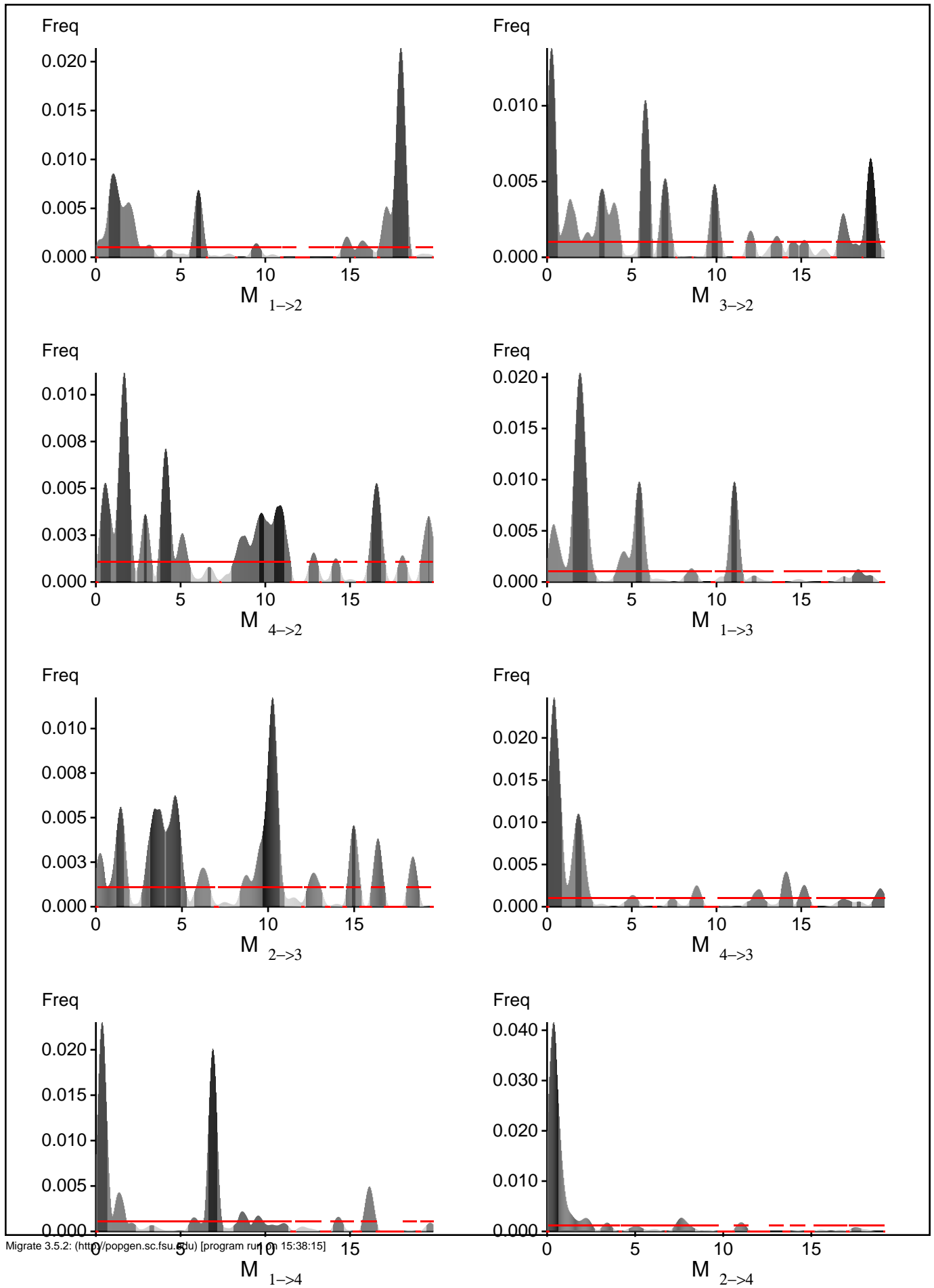
Beerli P., 2006. Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. *Bioinformatics* 22:341-345

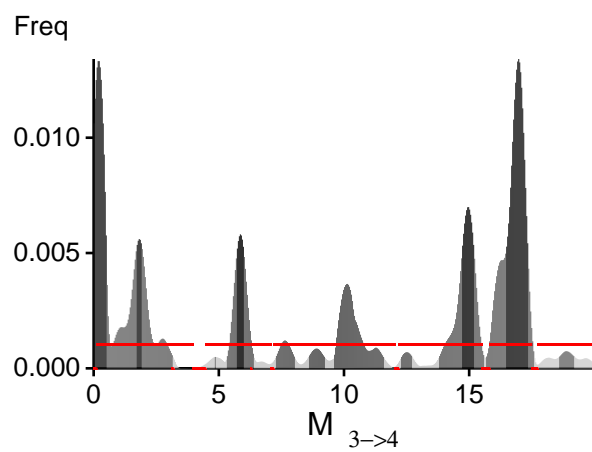
Beerli P., 2007. Estimation of the population scaled mutation rate from microsatellite data, *Genetics*, 177:1967-1968.

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	-667.88	-540.30	-603.55
2	-1775.81	-1317.35	-3516.81
All	-2353.48	-1767.44	-4030.15

(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 = 90.212376

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
Θ_1	76/76	1.00000
Θ_2	72/72	1.00000
Θ_3	53/53	1.00000
Θ_4	66/66	1.00000
M _{3→1}	86/86	1.00000
M _{4→1}	65/65	1.00000
M _{1→2}	71/71	1.00000
M _{3→2}	67/67	1.00000
M _{4→2}	64/64	1.00000
M _{1→3}	65/65	1.00000
M _{2→3}	61/61	1.00000
M _{4→3}	74/74	1.00000
M _{1→4}	73/73	1.00000
M _{2→4}	66/66	1.00000
M _{3→4}	59/59	1.00000
Genealogies	264/982	0.26884

MCMC-Autocorrelation and Effective MCMC Sample Size

Parameter	Autocorrelation	Effective Sample Size
Θ_1	0.98040	20.02
Θ_2	0.97869	21.50
Θ_3	0.96791	32.78
Θ_4	0.97359	26.73
$M_{3 \rightarrow 1}$	0.96986	30.63
$M_{4 \rightarrow 1}$	0.98153	18.63
$M_{1 \rightarrow 2}$	0.96394	36.72
$M_{3 \rightarrow 2}$	0.97655	23.77
$M_{4 \rightarrow 2}$	0.97601	24.27
$M_{1 \rightarrow 3}$	0.97256	27.86
$M_{2 \rightarrow 3}$	0.97718	23.12
$M_{4 \rightarrow 3}$	0.96679	33.78
$M_{1 \rightarrow 4}$	0.96585	34.99
$M_{2 \rightarrow 4}$	0.97404	26.27
$M_{3 \rightarrow 4}$	0.97274	27.72
$\text{Ln}[\text{Prob}(D G)]$	0.98512	14.99

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.

Param 8 (Locus 1): Upper prior boundary seems too low!

Param 8 (all loci): Upper prior boundary seems too low!

Param 7: Effective sample size of run seems too short!