

# Example: Microsatellite data set

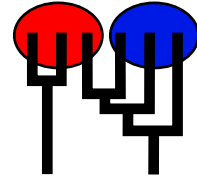
## MIGRATION RATE AND POPULATION SIZE ESTIMATION

using the coalescent and maximum likelihood or Bayesian inference

Migrate-n version 3.7.0 [April-06-16]

Program started at Fri Jun 24 09:29:10 2016

Program finished at Fri Jun 24 09:34:04 2016



## Options

Datatype: Microsatellite data [Brownian motion]  
Missing data: not included

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
1 population_num	*	0
2 population_num	*	*

Order of parameters:

1	$\Theta_1$	<displayed>
2	$\Theta_2$	<displayed>
4	$M_{1 \rightarrow 2}$	<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]	10000
Increment (record every x step [b])	10
Number of concurrent chains (replicates) [c]	2
Visited (sampled) parameter values [a*b*c]	200000
Number of discard trees per chain (burn-in)	10000

Multiple Markov chains:

Static heating scheme

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

Print options:

Data file: infile.msat  
 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: Microsatellite data  
[Data was used as repeat-length information]  
 Number of loci: 10

Population	Locus	Gene copies data	(missing)
1 population_number___0	1	50	(0)
	2	50	(0)
	3	50	(0)
	4	50	(0)
	5	50	(0)
	6	50	(0)
	7	50	(0)
	8	50	(0)
	9	50	(0)
	10	50	(0)
2 population_number___1	1	42	(0)
	2	42	(0)
	3	42	(0)
	4	42	(0)
	5	42	(0)
	6	42	(0)
	7	42	(0)
	8	42	(0)
	9	42	(0)
	10	42	(0)
Total of all populations	1	92	(0)
	2	92	(0)
	3	92	(0)
	4	92	(0)
	5	92	(0)
	6	92	(0)
	7	92	(0)
	8	92	(0)
	9	92	(0)
	10	92	(0)

## *Bayesian Analysis: Posterior distribution table*

Locus	Parameter	2.5%	25.0%	Mode	75.0%	97.5%	Median	Mean
1	$\Theta_1$	9.12000	12.16000	13.98000	16.40000	19.96000	14.50000	14.43303
1	$\Theta_2$	4.92000	5.36000	6.94000	10.76000	19.92000	10.94000	11.61535
1	$M_{1 \rightarrow 2}$	0.080	0.760	1.260	1.840	4.160	1.580	1.993
2	$\Theta_1$	2.88000	4.84000	6.82000	8.08000	13.76000	7.10000	7.63266
2	$\Theta_2$	2.36000	6.64000	9.22000	11.84000	19.44000	10.50000	10.66871
2	$M_{1 \rightarrow 2}$	1.000	1.240	1.580	1.920	2.400	10.900	10.662
3	$\Theta_1$	3.12000	5.48000	7.50000	8.92000	14.60000	7.94000	8.40332
3	$\Theta_2$	1.36000	1.60000	2.82000	7.48000	18.40000	8.82000	9.45046
3	$M_{1 \rightarrow 2}$	0.760	2.840	4.700	7.880	16.760	6.580	7.781
4	$\Theta_1$	13.12000	17.36000	19.62000	19.88000	20.00000	17.58000	17.19303
4	$\Theta_2$	4.36000	6.40000	6.94000	11.76000	19.96000	11.66000	11.72922
4	$M_{1 \rightarrow 2}$	8.760	12.880	13.540	14.520	20.000	14.740	14.532
5	$\Theta_1$	0.60000	1.92000	2.58000	3.64000	5.48000	2.94000	3.04717
5	$\Theta_2$	0.52000	0.60000	0.94000	1.28000	1.44000	9.82000	9.96516
5	$M_{1 \rightarrow 2}$	7.360	14.760	17.060	19.720	20.000	14.580	14.163
6	$\Theta_1$	2.92000	4.76000	5.54000	8.84000	15.48000	7.82000	8.80594
6	$\Theta_2$	0.00000	0.36000	0.70000	1.32000	3.92000	1.18000	1.48542
6	$M_{1 \rightarrow 2}$	1.160	3.280	5.460	9.520	16.960	8.100	8.990
7	$\Theta_1$	0.00000	0.00000	0.02000	0.24000	3.20000	0.26000	1.05976
7	$\Theta_2$	3.88000	4.92000	6.30000	9.08000	18.32000	10.46000	10.79200
7	$M_{1 \rightarrow 2}$	0.000	0.160	0.540	0.880	2.280	0.740	0.861
8	$\Theta_1$	4.32000	5.76000	7.42000	9.96000	17.00000	9.10000	9.79246
8	$\Theta_2$	2.64000	3.80000	8.38000	10.16000	19.32000	9.78000	10.30733
8	$M_{1 \rightarrow 2}$	5.920	13.360	16.340	19.720	20.000	13.660	13.319
9	$\Theta_1$	6.08000	8.04000	11.62000	13.36000	19.12000	11.74000	11.96815
9	$\Theta_2$	3.56000	3.80000	7.74000	9.08000	19.12000	11.10000	11.23704
9	$M_{1 \rightarrow 2}$	4.440	13.000	15.660	16.360	20.000	11.780	11.533

10	$\Theta_1$	7.36000	8.76000	9.78000	10.84000	20.00000	13.38000	13.37674
10	$\Theta_2$	3.20000	6.04000	9.14000	12.96000	19.08000	10.54000	10.68154
10	$M_{1 \rightarrow 2}$	0.360	0.800	1.340	3.400	9.600	3.300	3.953
<hr/>								
All	$\Theta_1$	0.00000	0.00000	0.02000	0.28000	0.52000	0.30000	0.90886
All	$\Theta_2$	3.56000	4.68000	5.82000	6.92000	9.32000	6.82000	9.07406
All	$M_{1 \rightarrow 2}$	18.800	19.000	19.460	19.960	20.000	14.060	13.595

## Citation suggestions:

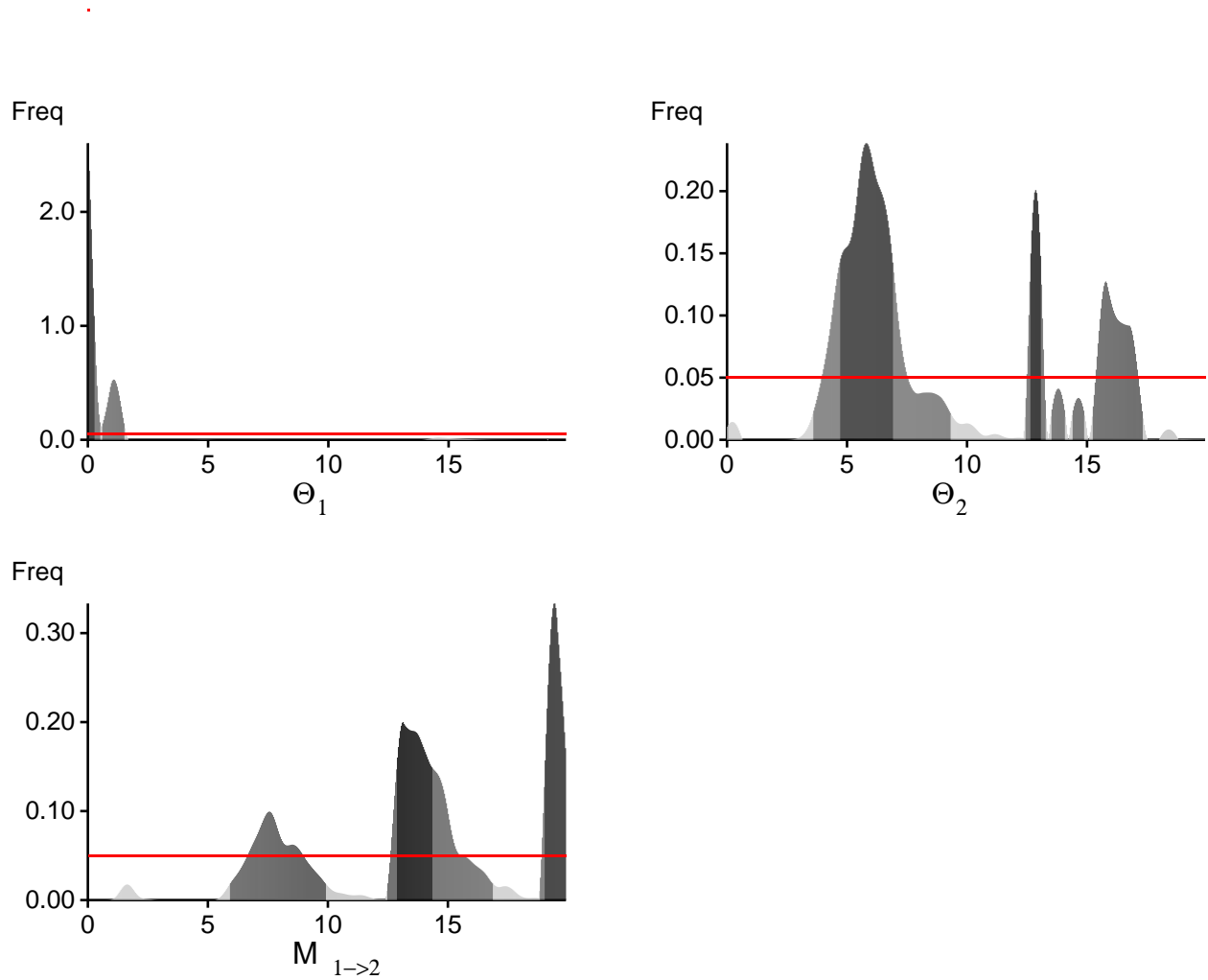
Beerli P., 2006. Comparison of Bayesian and maximum-likelihood inference of population genetic parameters.

Bioinformatics 22:341-345

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	-3153.33	-633.82	-124.61
2	-2095.58	-430.09	-91.78
3	-1057.85	-268.77	-97.98
4	-6844.32	-1227.40	-113.36
5	-1310.13	-282.61	-55.84
6	-3619.48	-685.63	-90.04
7	-863.34	-218.57	-63.65
8	-1360.05	-322.47	-97.24
9	-1972.10	-422.67	-99.44
10	-3911.44	-748.31	-104.36
All	-26191.33	-5244.03	-942.00

(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 = -3.698491

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
$\Theta_1$	333439/333439	1.00000
$\Theta_2$	332857/332857	1.00000
$M_{1 \rightarrow 2}$	333206/333206	1.00000
Genealogies	322871/1000498	0.32271



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

Parameter	Autocorrelation	Effective Sampe Size
$\Theta_1$	0.86226	15351.49
$\Theta_2$	0.74567	29830.89
$M_{1 \rightarrow 2}$	0.83610	18367.16
$\text{Ln}[\text{Prob}(D G)]$	0.99170	834.42

## *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 1 (Locus 4): Upper prior boundary seems too low!

Param 4 (Locus 5): Upper prior boundary seems too low!

Param 4 (Locus 8): Upper prior boundary seems too low!

Param 4 (all loci): Upper prior boundary seems too low!