

Example: Microsatellite data set

MIGRATION RATE AND POPULATION SIZE ESTIMATION

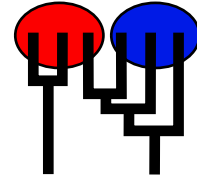
Maximum likelihood or Bayesian inference using the coalescent

Migrate-n version debug 4.0 [x]

Compiled for a SYMMETRIC MULTIPROCESSORS

Program started at Fri Dec 14 16:03:19 2012

Program finished at Fri Dec 14 16:03:58 2012



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

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
1 population_num	*	0
2 population_num	*	*

Order of parameters:

1	Θ_1	<displayed>
2	Θ_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
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]	5000
Increment (record every x step [b])	1
Number of concurrent chains (replicates) [c]	2
Visited (sampled) parameter values [a*b*c]	10000
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
Haplotyping is turned on:	NO
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)

Allele frequency spectra

Locus 1

Allele	Pop1	Pop2	All
16	0.208	0.167	0.189
18	0.062	0.119	0.089
15	0.229	0.024	0.133
21	0.021	0.167	0.089
23	0.021	0.119	0.067
17	0.292	0.095	0.200
22	0.062	0.119	0.089
19	0.021	0.071	0.044
25	0.062	0.024	0.044
24	0.021	-	0.011
26	-	0.024	0.011
27	-	0.048	0.022
29	-	0.024	0.011
Total	10	12	13
H _{exp}	0.806	0.883	0.874

Locus 2

Allele	Pop1	Pop2	All
16	0.520	0.571	0.543
19	0.040	-	0.022
18	0.220	0.119	0.174
17	0.160	0.167	0.163
15	0.020	-	0.011
21	0.020	0.071	0.043
20	0.020	0.024	0.022
22	-	0.048	0.022
Total	7	6	8
H _{exp}	0.653	0.624	0.644

Locus 3

Allele	Pop1	Pop2	All
16	0.020	-	0.011
19	0.240	0.262	0.250
18	0.080	0.095	0.087
20	0.260	0.476	0.359

Allele	Pop1	Pop2	All
21	0.280	0.119	0.207
22	0.120	0.048	0.087
Total	6	5	6
H _{exp}	0.775	0.679	0.751
Locus 4			
Allele	Pop1	Pop2	All
19	0.120	0.143	0.130
20	0.100	0.190	0.141
16	0.060	0.071	0.065
24	0.160	0.024	0.098
15	0.020	0.048	0.033
25	0.160	0.167	0.163
14	0.020	0.048	0.033
12	0.060	-	0.033
23	0.060	0.119	0.087
28	0.020	-	0.011
22	0.060	0.024	0.043
21	0.160	0.119	0.141
13	-	0.024	0.011
26	-	0.024	0.011
Total	12	12	14
H _{exp}	0.883	0.875	0.890
Locus 5			
Allele	Pop1	Pop2	All
16	0.020	-	0.011
24	0.020	-	0.011
20	0.380	0.524	0.446
21	0.400	0.357	0.380
19	0.180	0.119	0.152
Total	5	3	5
H _{exp}	0.662	0.584	0.633
Locus 6			
Allele	Pop1	Pop2	All
20	0.100	0.024	0.065
21	0.140	0.476	0.293
18	0.300	0.214	0.261
22	0.200	0.119	0.163

Allele	Pop1	Pop2	All
16	0.060	-	0.033
19	0.040	-	0.022
24	0.160	0.048	0.109
17	-	0.119	0.054
Total	7	6	8
H _{exp}	0.810	0.696	0.799
Locus 7			
Allele	Pop1	Pop2	All
19	0.040	0.095	0.065
20	0.660	0.143	0.424
21	0.100	0.333	0.207
22	0.180	0.190	0.185
23	0.020	0.238	0.120
Total	5	5	5
H _{exp}	0.520	0.766	0.725
Locus 8			
Allele	Pop1	Pop2	All
23	0.020	0.071	0.043
20	0.160	0.190	0.174
19	0.500	0.524	0.511
17	0.020	0.048	0.033
18	0.100	0.071	0.087
16	0.080	-	0.043
22	0.100	0.048	0.076
15	0.020	0.048	0.033
Total	8	7	8
H _{exp}	0.697	0.672	0.690
Locus 9			
Allele	Pop1	Pop2	All
19	0.300	0.429	0.359
17	0.020	-	0.011
24	0.060	0.024	0.043
20	0.300	0.167	0.239
23	0.180	0.143	0.163
22	0.080	0.024	0.054
18	0.020	0.071	0.043
21	0.040	0.095	0.065

Allele	Pop1	Pop2	All
25	-	0.048	0.022
Total	8	8	9
H _{exp}	0.775	0.751	0.776
Locus 10			
Allele	Pop1	Pop2	All
24	0.040	-	0.022
19	0.180	0.167	0.174
22	0.080	0.214	0.141
20	0.420	0.214	0.326
23	0.080	0.167	0.120
21	0.060	0.048	0.054
18	0.080	-	0.043
15	0.020	0.071	0.043
17	0.040	0.048	0.043
25	-	0.071	0.033
Total	9	8	10
H _{exp}	0.765	0.838	0.819
Average expected heterozygosity			
	Pop1	Pop2	All
H _{exp}	0.735	0.737	0.760

Bayesian Analysis: Posterior distribution table

Locus	Parameter	2.5%	25.0%	Mode	75.0%	97.5%	Median	Mean
1	Θ_1	2.88000	3.48000	5.02000	6.32000	7.44000	7.42000	8.93092
1	Θ_2	1.36000	1.84000	2.22000	2.68000	5.40000	3.18000	3.27354
1	$M_{1 \rightarrow 2}$	0.000	0.080	0.540	1.000	1.240	1.860	2.327
2	Θ_1	1.56000	1.96000	2.62000	3.24000	3.96000	4.50000	5.22085
2	Θ_2	0.68000	1.32000	1.78000	3.04000	8.76000	2.78000	3.59957
2	$M_{1 \rightarrow 2}$	5.000	6.080	9.300	9.920	13.360	10.420	11.181
3	Θ_1	6.64000	6.84000	8.06000	10.52000	11.44000	12.50000	12.71881
3	Θ_2	4.48000	6.64000	8.66000	9.76000	17.64000	10.46000	10.89124
3	$M_{1 \rightarrow 2}$	0.520	0.920	1.380	1.960	3.840	1.780	1.966
4	Θ_1	11.60000	15.16000	15.90000	18.56000	20.00000	16.06000	15.84582
4	Θ_2	13.44000	15.48000	18.78000	19.40000	19.92000	14.54000	13.67096
4	$M_{1 \rightarrow 2}$	0.000	0.040	0.340	0.560	1.200	0.540	0.491
5	Θ_1	6.24000	7.80000	8.94000	10.36000	13.84000	9.58000	9.76017
5	Θ_2	0.48000	0.80000	1.34000	2.60000	7.68000	2.78000	3.37399
5	$M_{1 \rightarrow 2}$	0.000	0.000	0.020	0.840	5.560	1.540	2.203
6	Θ_1	7.32000	9.56000	10.50000	12.16000	16.68000	11.42000	11.65317
6	Θ_2	2.28000	2.92000	3.54000	5.00000	10.16000	4.70000	5.47695
6	$M_{1 \rightarrow 2}$	0.160	0.360	0.780	1.480	4.120	1.700	1.886
7	Θ_1	1.48000	2.04000	2.50000	3.08000	5.28000	2.82000	3.01102
7	Θ_2	0.56000	0.72000	1.18000	1.92000	5.20000	2.26000	2.52990
7	$M_{1 \rightarrow 2}$	0.000	0.200	0.580	1.760	4.280	1.660	1.814
8	Θ_1	6.36000	6.92000	9.74000	10.40000	13.92000	10.74000	11.56349
8	Θ_2	4.52000	6.08000	6.74000	8.40000	11.84000	12.02000	12.06711
8	$M_{1 \rightarrow 2}$	0.720	1.120	1.500	2.080	3.400	1.860	1.928
9	Θ_1	2.48000	2.72000	3.50000	4.52000	4.92000	5.02000	6.76348
9	Θ_2	0.72000	0.92000	1.82000	3.68000	12.36000	3.78000	6.09266
9	$M_{1 \rightarrow 2}$	10.000	18.520	19.620	19.920	20.000	13.900	13.502

10	Θ_1	4.60000	5.64000	6.58000	7.84000	9.80000	10.02000	12.37632
10	Θ_2	3.56000	6.76000	7.74000	8.24000	12.00000	7.46000	7.86962
10	$M_{1 \rightarrow 2}$	0.000	0.240	0.500	0.680	1.280	0.580	0.589
<hr/>								
All	Θ_1	1.56000	1.72000	1.98000	2.20000	2.36000	2.10000	2.31938
All	Θ_2	1.04000	1.32000	1.50000	1.68000	1.96000	1.58000	1.55189
All	$M_{1 \rightarrow 2}$	0.320	0.560	0.780	0.920	1.200	0.820	0.771

Citation suggestions:

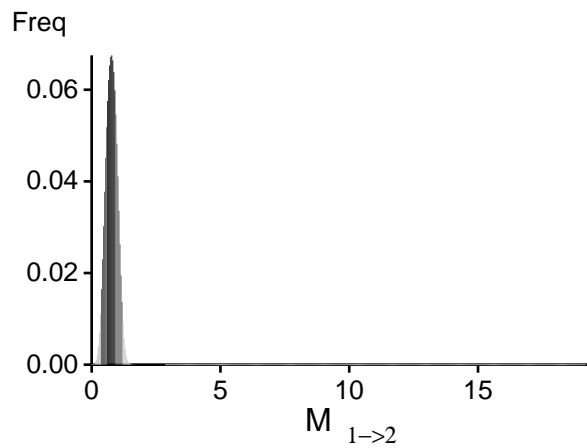
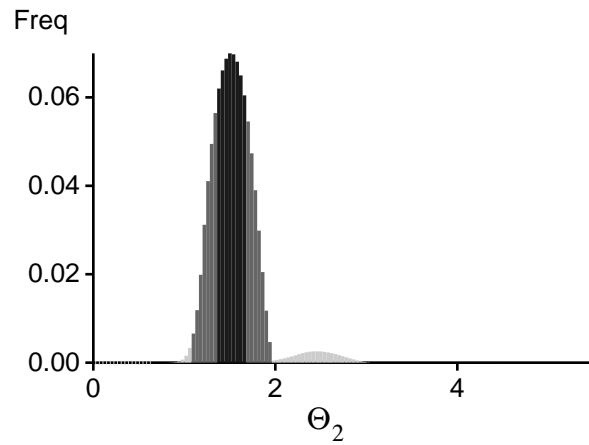
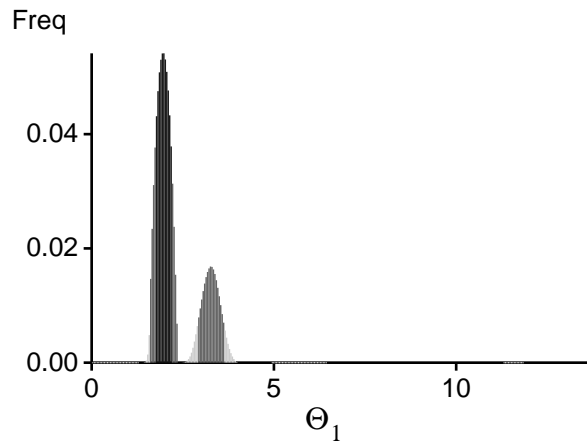
Beerli P., 2006. Comparison of Bayesian and maximum-likelihood inference of population genetic parameters.

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	-17576.66	-2941.41	-91.76
2	-5399.03	-947.30	-63.41
3	-562.96	-198.88	-107.13
4	-3139.41	-638.43	-119.79
5	-2174.25	-434.36	-76.87
6	-2099.25	-446.51	-88.03
7	-622.82	-192.11	-48.90
8	-773.03	-232.51	-94.21
9	-2979.74	-581.63	-66.65
10	-2857.57	-581.88	-111.74
All	-38243.36	-7253.64	-927.13

(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 = -58.623032

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	49821/49821	1.00000
Θ_2	49977/49977	1.00000
$M_{1 \rightarrow 2}$	50013/50013	1.00000
Genealogies	16221/50267	0.32270

MCMC-Autocorrelation and Effective MCMC Sample Size

Parameter	Autocorrelation	Effective Sample Size
Θ_1	0.92535	3897.93
Θ_2	0.93531	3366.65
$M_{1 \rightarrow 2}$	0.92008	4195.25
$\text{Ln}[\text{Prob}(D G)]$	0.99502	251.41

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

Param 4 (Locus 9): Upper prior boundary seems too low!