

Example: Microsatellite data set

POPULATION SIZE, MIGRATION, DIVERGENCE, ASSIGNMENT, HISTORY

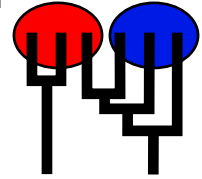
Bayesian inference using the structured coalescent

Migrate-n version 4.1.3a [Feb-22-2015]

Using Intel AVX (Advanced Vector Extensions)

Program started at Sun Feb 22 12:45:34 2015

Program finished at Sun Feb 22 12:58:07 2015



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

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	2
1 population_num	*	0
2 population_num	*	*

Order of parameters:

1	Θ_1	<displayed>
2	Θ_2	<displayed>

3 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
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])	50
Number of concurrent chains (replicates) [c]	2
Visited (sampled) parameter values [a*b*c]	500000
Number of discard trees per chain (burn-in)	5000

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.msaf
Haplotyping is turned on:	NO
Output file:	outfile-bayes
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: infile.msat
 Datatype: Microsatellite data [Brownian]
 [Data was used as repeat-length information]
 Number of loci: 10

Mutationmodel:

Locus	Sublocus	Mutationmodel	Mutationmodel parameters
1	1	Brownian Motion	[none]
2	1	Brownian Motion	[none]
3	1	Brownian Motion	[none]
4	1	Brownian Motion	[none]
5	1	Brownian Motion	[none]
6	1	Brownian Motion	[none]
7	1	Brownian Motion	[none]
8	1	Brownian Motion	[none]
9	1	Brownian Motion	[none]
10	1	Brownian Motion	[none]

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)

Total of all populations	10	42	(0)
	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	0.220	0.167	0.196
19	0.040	0.071	0.054
18	0.060	0.119	0.087
15	0.220	0.024	0.130
21	0.020	0.167	0.087
23	0.020	0.119	0.065
17	0.280	0.095	0.196
22	0.060	0.119	0.087
25	0.060	0.024	0.043
24	0.020	-	0.011
26	-	0.024	0.011
27	-	0.048	0.022
29	-	0.024	0.011
Alleles	10	12	13
Samplesize	50	42	92
H _{exp}	0.811	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
Alleles	7	6	8
Samplesize	50	42	92
H _{exp}	0.653	0.624	0.644

Locus 3

Allele	Pop1	Pop2	All
19	0.240	0.262	0.250
20	0.280	0.476	0.370

Allele	Pop1	Pop2	All
18	0.080	0.095	0.087
21	0.280	0.119	0.207
22	0.120	0.048	0.087
Alleles	5	5	5
Samplesize	50	42	92
H _{exp}	0.765	0.679	0.743
Locus 4			
Allele	Pop1	Pop2	All
16	0.080	0.071	0.076
24	0.180	0.024	0.109
15	0.020	0.048	0.033
25	0.160	0.167	0.163
14	0.020	0.048	0.033
19	0.100	0.143	0.120
12	0.060	-	0.033
20	0.080	0.190	0.130
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
Alleles	12	12	14
Samplesize	50	42	92
H _{exp}	0.882	0.875	0.892
Locus 5			
Allele	Pop1	Pop2	All
20	0.400	0.524	0.457
21	0.420	0.357	0.391
19	0.180	0.119	0.152
Alleles	3	3	3
Samplesize	50	42	92
H _{exp}	0.631	0.584	0.615
Locus 6			
Allele	Pop1	Pop2	All
19	0.060	-	0.033
20	0.100	0.024	0.065

Allele	Pop1	Pop2	All
18	0.300	0.214	0.261
22	0.200	0.119	0.163
21	0.120	0.476	0.283
16	0.060	-	0.033
24	0.160	0.048	0.109
17	-	0.119	0.054
Alleles	7	6	8
Samplesize	50	42	92
H _{exp}	0.813	0.696	0.804
Locus 7			
Allele	Pop1	Pop2	All
23	0.040	0.238	0.130
20	0.660	0.143	0.424
22	0.180	0.190	0.185
21	0.100	0.333	0.207
19	0.020	0.095	0.054
Alleles	5	5	5
Samplesize	50	42	92
H _{exp}	0.520	0.766	0.724
Locus 8			
Allele	Pop1	Pop2	All
19	0.520	0.524	0.522
17	0.040	0.048	0.043
18	0.100	0.071	0.087
20	0.140	0.190	0.163
16	0.080	-	0.043
22	0.100	0.048	0.076
15	0.020	0.048	0.033
23	-	0.071	0.033
Alleles	7	7	8
Samplesize	50	42	92
H _{exp}	0.682	0.672	0.682
Locus 9			
Allele	Pop1	Pop2	All
24	0.080	0.024	0.054
19	0.300	0.429	0.359
20	0.300	0.167	0.239

Allele	Pop1	Pop2	All
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
25	-	0.048	0.022
Alleles	7	8	8
Samplesize	50	42	92
H_{exp}	0.773	0.751	0.775
Locus 10			
Allele	Pop1	Pop2	All
22	0.100	0.214	0.152
20	0.440	0.214	0.337
23	0.080	0.167	0.120
24	0.020	-	0.011
19	0.160	0.167	0.163
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
Alleles	9	8	10
Samplesize	50	42	92
H_{exp}	0.752	0.838	0.813
Average expected heterozygosity			
	Pop1	Pop2	All
H_{exp}	0.728	0.737	0.000

Bayesian Analysis: Posterior distribution table

Locus	Parameter	2.5%	25.0%	Mode	75.0%	97.5%	Median	Mean
1	Θ_1	8.84000	13.60000	15.58000	18.52000	20.00000	14.86000	14.61700
1	Θ_2	4.24000	14.28000	19.14000	19.80000	20.00000	14.54000	13.60393
1	$M_{1 \rightarrow 2}$	0.000	0.000	0.740	2.000	14.400	2.020	3.874
2	Θ_1	2.44000	4.56000	7.10000	8.36000	12.56000	6.98000	7.22636
2	Θ_2	1.92000	2.28000	3.58000	6.20000	19.40000	9.98000	10.26078
2	$M_{1 \rightarrow 2}$	7.320	14.720	19.140	19.560	20.000	15.140	14.630
3	Θ_1	2.76000	5.28000	7.90000	10.20000	17.04000	8.74000	9.33329
3	Θ_2	1.40000	1.80000	3.90000	9.56000	15.12000	8.82000	9.56871
3	$M_{1 \rightarrow 2}$	2.480	4.480	8.700	10.240	19.000	8.740	9.417
4	Θ_1	12.56000	17.12000	19.18000	19.76000	20.00000	17.42000	17.03526
4	Θ_2	5.36000	8.20000	11.66000	12.72000	19.88000	12.50000	12.53930
4	$M_{1 \rightarrow 2}$	0.720	1.520	2.900	6.600	11.360	6.060	7.999
5	Θ_1	0.04000	0.88000	1.50000	2.56000	5.20000	2.18000	2.43874
5	Θ_2	1.44000	3.88000	6.06000	6.96000	19.60000	10.42000	10.47412
5	$M_{1 \rightarrow 2}$	5.280	13.840	18.780	19.720	20.000	14.180	13.515
6	Θ_1	2.80000	4.76000	6.98000	9.44000	16.28000	8.26000	8.83058
6	Θ_2	0.00000	0.20000	0.82000	1.48000	5.00000	1.34000	1.70128
6	$M_{1 \rightarrow 2}$	1.120	2.640	4.900	9.120	15.440	7.820	8.894
7	Θ_1	0.88000	1.80000	2.74000	4.20000	8.84000	3.78000	4.31644
7	Θ_2	0.76000	1.04000	1.74000	8.80000	18.72000	8.58000	9.19534
7	$M_{1 \rightarrow 2}$	0.000	0.440	1.300	3.160	9.640	2.900	3.591
8	Θ_1	3.84000	6.08000	8.94000	10.88000	16.92000	9.42000	9.80119
8	Θ_2	2.80000	5.88000	9.38000	12.04000	19.84000	10.98000	11.10548
8	$M_{1 \rightarrow 2}$	5.000	14.720	19.180	19.880	20.000	14.940	13.979
9	Θ_1	4.64000	7.20000	10.22000	12.36000	18.04000	10.62000	10.91378
9	Θ_2	3.36000	5.08000	6.14000	12.08000	19.72000	11.14000	11.28512
9	$M_{1 \rightarrow 2}$	2.440	13.440	16.060	19.480	19.760	11.220	11.123

10	Θ_1	8.60000	14.84000	18.66000	19.44000	20.00000	15.38000	14.97991
10	Θ_2	0.84000	1.40000	2.94000	7.56000	17.48000	7.06000	8.19014
10	$M_{1 \rightarrow 2}$	0.480	1.480	4.820	7.280	16.720	7.260	7.990
<hr/>								
All	Θ_1	0.04000	0.72000	1.50000	2.36000	10.16000	4.50000	4.96088
All	Θ_2	5.16000	9.24000	11.26000	14.76000	19.68000	11.62000	11.04517
All	$M_{1 \rightarrow 2}$	0.000	1.720	2.700	3.960	9.720	4.300	5.825

Citation suggestions:

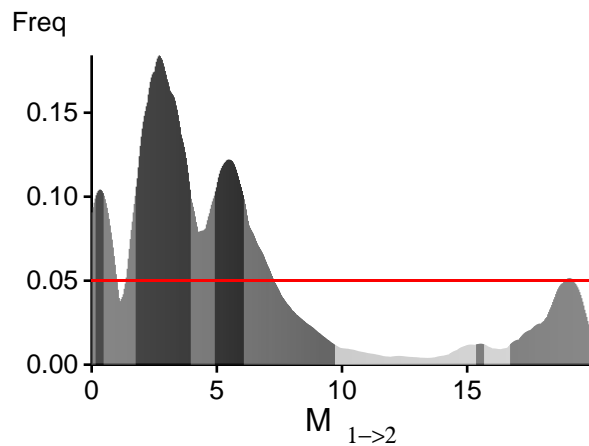
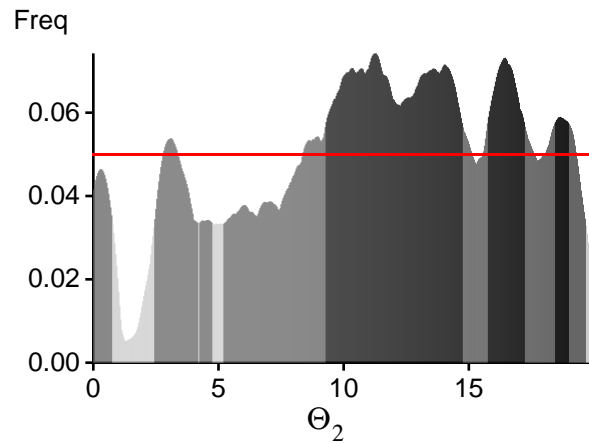
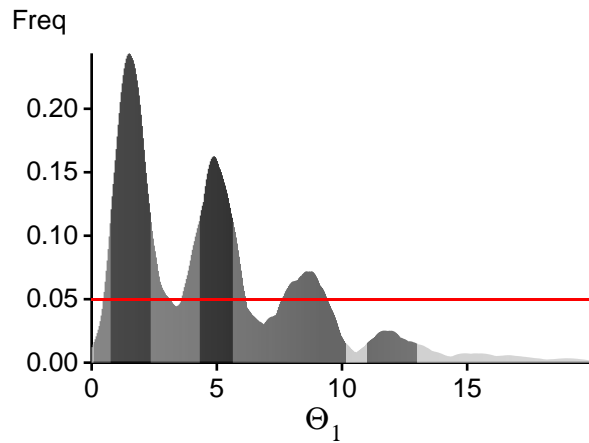
Beerli P., 2006. Comparison of Bayesian and maximum-likelihood inference of population genetic parameters.

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	-11224.80	-1926.70	-125.35
2	-1394.99	-315.24	-91.21
3	-2748.49	-539.43	-99.85
4	-7648.82	-1358.22	-117.22
5	-1002.82	-232.57	-65.48
6	-9972.00	-1702.10	-75.87
7	-751.08	-211.03	-85.50
8	-2443.76	-494.06	-91.75
9	-2315.88	-477.72	-100.19
10	-8335.71	-1458.77	-108.78
All	-47839.07	-8716.57	-961.94

(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 = -0.720889

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	832079/832079	1.00000
Θ_2	833455/833455	1.00000
$M_{1 \rightarrow 2}$	832531/832531	1.00000
Genealogies	792401/2501935	0.31672

MCMC-Autocorrelation and Effective MCMC Sample Size

Parameter	Autocorrelation	Effective Sampe Size
Θ_1	0.84631	8652.46
Θ_2	0.67816	19866.10
$M_{1 \rightarrow 2}$	0.85783	7783.23
$\text{Ln}[\text{Prob}(D G)]$	0.97390	1323.03

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 4 (Locus 2): 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 (Locus 9): Upper prior boundary seems too low!