

Simulated microsats

POPULATION SIZE, MIGRATION, DIVERGENCE, ASSIGNMENT, HISTORY

Bayesian inference using the structured coalescent

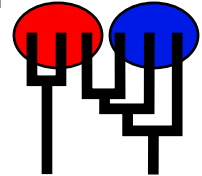
Migrate-n version 4.1.1a []

Compiled for PARALLEL computer architectures

One master and 8 compute nodes are available.

Program started at Fri Jan 30 18:08:21 2015

Program finished at Sat Jan 31 01:49:23 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:	(with internal timer) 230123139
Start parameters:	
Theta values were generated	RANDOM start value from the prior
M values were generated	RANDOM start value from 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_1	* *
2 Population_2	* *
Order of parameters:	
1	Θ_1 <displayed>

2	Θ_2	<displayed>
3	$M_{2 \rightarrow 1}$	<displayed>
4	$M_{1 \rightarrow 2}$	<displayed>

Mutation rate among loci: Varying ([crudely] estimated from data)

Rates per locus:	1.19149,	0.85106,	1.36170,	1.19149,	0.68085,
	1.02128,	1.02128,	0.68085		

Analysis strategy: Bayesian inference

Proposal distributions for parameter

Parameter	Proposal
Theta	Metropolis sampling
M	Metropolis sampling

Prior distribution for parameter

Parameter	Prior	Minimum	Mean*	Maximum	Delta	Bins
Theta	Uniform	0.000000	15.000000	30.000000	3.000000	1500
Theta	Uniform	0.000000	15.000000	30.000000	3.000000	1500
M	Uniform	0.000000	15.000000	30.000000	3.000000	1500
M	Uniform	0.000000	15.000000	30.000000	3.000000	1500

Markov chain settings: Long chain

Number of chains	1
Recorded steps [a]	1000
Increment (record every x step [b])	100
Number of concurrent chains (replicates) [c]	100
Visited (sampled) parameter values [a*b*c]	10000000
Number of discard trees per chain (burn-in)	1000

Multiple Markov chains:

Static heating scheme	4 chains with temperatures
	1000000.00 3.00 1.50 1.00
	Swapping interval is 5

Print options:

Data file:	infile
Haplotyping is turned on:	NO
Output file:	outfile
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
 Datatype: Microsatellite data [Brownian]
 [Data was used as repeat-length information]
 Number of loci: 8

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]

Population	Locus	Gene copies data	(missing)
1 Population_1	1	20	(0)
	2	20	(0)
	3	20	(0)
	4	20	(0)
	5	20	(0)
	6	20	(0)
	7	20	(0)
	8	20	(0)
2 Population_2	1	20	(0)
	2	20	(0)
	3	20	(0)
	4	20	(0)
	5	20	(0)
	6	20	(0)
	7	20	(0)
	8	20	(0)
Total of all populations	1	40	(0)
	2	40	(0)
	3	40	(0)
	4	40	(0)
	5	40	(0)
	6	40	(0)

7	40	(0)
8	40	(0)

Relative mutation rate among loci estimated from the data

Locus	Relative mutation rate	Number of alleles
1	1.19149	7
2	0.85106	5
3	1.36170	8
4	1.19149	7
5	0.68085	4
6	1.02128	6
7	1.02128	6
8	0.68085	4
All	1.00000	5.9

Allele frequency spectra

Locus 1

Allele	Pop1	Pop2	All
79	0.300	0.100	0.200
80	0.150	0.400	0.275
78	0.350	0.050	0.200
81	0.200	0.150	0.175
83	-	0.050	0.025
82	-	0.250	0.125
Alleles	4	6	6
Samplesize	20	20	40
H _{exp}	0.725	0.740	0.797

Locus 2

Allele	Pop1	Pop2	All
80	0.750	0.800	0.775
81	0.050	-	0.025
79	0.100	0.150	0.125
78	0.100	0.050	0.075
Alleles	4	3	4
Samplesize	20	20	40
H _{exp}	0.415	0.335	0.377

Locus 3

Allele	Pop1	Pop2	All
81	0.350	0.350	0.350
78	0.050	-	0.025
80	0.150	0.200	0.175
83	0.150	0.100	0.125
84	0.050	0.050	0.050
79	0.150	0.100	0.125
82	0.100	0.200	0.150
Alleles	7	6	7
Samplesize	20	20	40
H _{exp}	0.795	0.775	0.790

Locus 4

Allele	Pop1	Pop2	All
--------	------	------	-----

Allele	Pop1	Pop2	All
82	0.050	-	0.025
81	0.700	0.650	0.675
78	0.200	0.100	0.150
77	0.050	-	0.025
79	-	0.100	0.050
80	-	0.150	0.075
Alleles	4	4	6
Samplesize	20	20	40
H _{exp}	0.465	0.535	0.512
Locus 5			
Allele	Pop1	Pop2	All
80	0.750	0.650	0.700
79	0.200	0.250	0.225
81	0.050	0.100	0.075
Alleles	3	3	3
Samplesize	20	20	40
H _{exp}	0.395	0.505	0.454
Locus 6			
Allele	Pop1	Pop2	All
77	0.100	-	0.050
78	0.500	0.600	0.550
79	0.250	0.050	0.150
80	0.150	0.300	0.225
81	-	0.050	0.025
Alleles	4	4	5
Samplesize	20	20	40
H _{exp}	0.655	0.545	0.621
Locus 7			
Allele	Pop1	Pop2	All
78	0.150	0.250	0.200
79	0.200	0.050	0.125
80	0.500	0.450	0.475
81	0.150	0.200	0.175
82	-	0.050	0.025
Alleles	4	5	5
Samplesize	20	20	40

Allele	Pop1	Pop2	All
H_{exp}	0.665	0.690	0.688
Locus 8			
Allele	Pop1	Pop2	All
77	0.600	0.250	0.425
79	0.150	0.200	0.175
78	0.250	0.550	0.400
Alleles	3	3	3
Samplesize	20	20	40
H_{exp}	0.555	0.595	0.629
Average expected heterozygosity			
	Pop1	Pop2	All
H_{exp}	0.000	0.000	0.000

Bayesian Analysis: Posterior distribution table

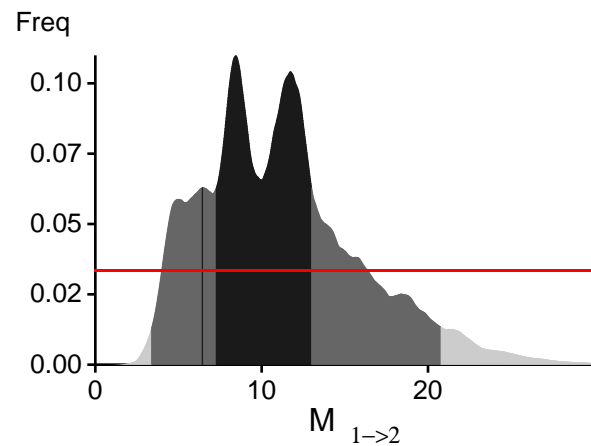
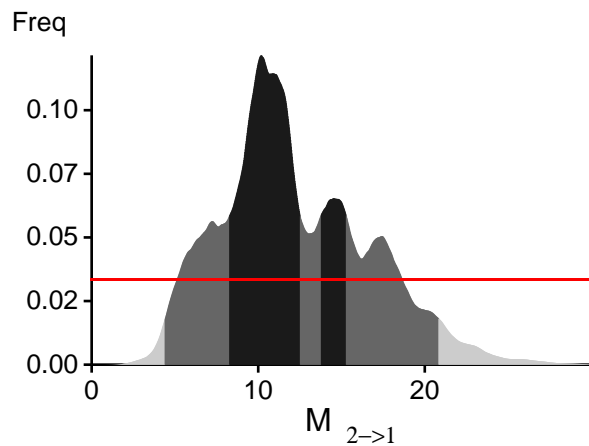
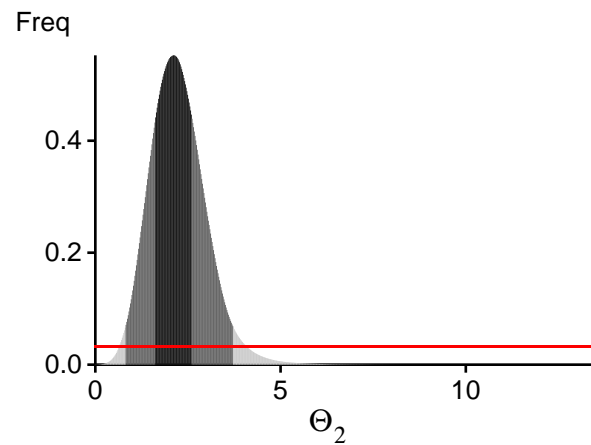
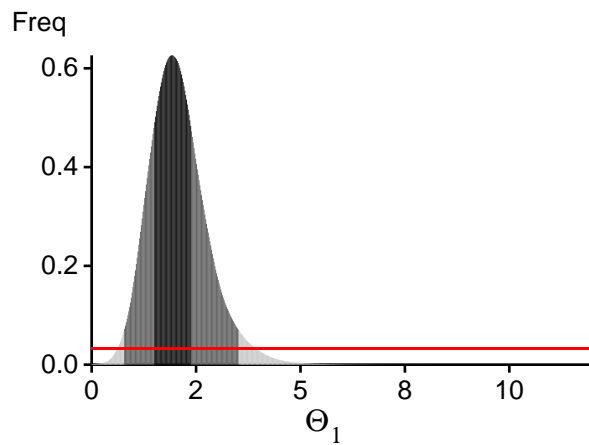
Locus	Parameter	2.5%	25.0%	Mode	75.0%	97.5%	Median	Mean
1	Θ_1	0.00000	0.20000	1.27000	5.46000	24.32000	5.29000	8.06065
1	Θ_2	1.06000	1.76000	3.85000	12.22000	26.70000	11.41000	12.89680
1	$M_{2 \rightarrow 1}$	0.140	1.900	5.110	12.820	24.760	11.530	12.520
1	$M_{1 \rightarrow 2}$	0.280	3.100	8.930	12.640	24.720	11.850	12.811
2	Θ_1	0.26000	0.42000	1.87000	7.40000	21.50000	7.17000	10.41996
2	Θ_2	0.00000	0.60000	1.49000	3.00000	19.40000	2.71000	5.54990
2	$M_{2 \rightarrow 1}$	5.100	15.900	19.730	27.100	29.960	18.330	17.754
2	$M_{1 \rightarrow 2}$	1.860	3.240	7.510	15.000	28.480	13.390	14.243
3	Θ_1	1.80000	3.52000	7.77000	13.26000	27.38000	11.81000	13.18833
3	Θ_2	1.28000	2.58000	5.29000	12.56000	27.06000	11.41000	12.79195
3	$M_{2 \rightarrow 1}$	1.820	6.800	9.630	16.060	26.980	13.030	13.707
3	$M_{1 \rightarrow 2}$	2.480	5.920	8.410	15.040	28.240	13.410	14.361
4	Θ_1	0.00000	0.52000	1.39000	2.76000	18.88000	2.49000	5.78786
4	Θ_2	0.00000	0.14000	1.55000	6.46000	22.14000	6.35000	9.78221
4	$M_{2 \rightarrow 1}$	2.220	4.980	10.310	16.260	28.760	14.110	14.661
4	$M_{1 \rightarrow 2}$	3.620	10.480	18.510	22.160	29.700	16.510	16.375
5	Θ_1	0.00000	0.74000	1.79000	4.14000	23.14000	3.79000	6.93835
5	Θ_2	0.14000	0.42000	1.95000	7.12000	26.38000	6.87000	10.12403
5	$M_{2 \rightarrow 1}$	1.320	1.860	4.590	12.620	27.340	12.310	13.598
5	$M_{1 \rightarrow 2}$	3.360	16.180	25.110	28.560	29.740	17.410	16.821
6	Θ_1	0.36000	0.54000	2.03000	9.26000	24.72000	8.97000	11.39981
6	Θ_2	0.00000	0.48000	1.31000	3.52000	20.96000	3.25000	5.77929
6	$M_{2 \rightarrow 1}$	3.640	8.780	11.110	20.660	29.640	15.670	15.766
6	$M_{1 \rightarrow 2}$	1.880	3.420	5.830	14.860	27.280	13.190	14.054
7	Θ_1	0.10000	0.54000	2.13000	7.64000	24.50000	7.31000	10.26064
7	Θ_2	0.16000	0.82000	2.55000	6.54000	25.76000	6.15000	9.17650
7	$M_{2 \rightarrow 1}$	3.100	9.660	17.070	21.220	29.440	15.530	15.488
7	$M_{1 \rightarrow 2}$	1.740	3.560	11.570	13.680	27.180	13.010	14.047
8	Θ_1	0.00000	0.26000	1.49000	4.00000	23.54000	3.81000	7.04321

Locus	Parameter	2.5%	25.0%	Mode	75.0%	97.5%	Median	Mean
8	Θ_2	0.00000	0.38000	1.63000	6.52000	26.56000	6.27000	9.53000
8	$M_{2 \rightarrow 1}$	0.340	1.260	2.690	11.980	24.940	12.230	13.205
8	$M_{1 \rightarrow 2}$	2.340	3.160	5.390	15.060	29.360	13.650	14.358
All	Θ_1	0.76000	1.48000	1.93000	2.38000	3.52000	2.03000	2.08307
All	Θ_2	0.80000	1.60000	2.13000	2.60000	3.72000	2.21000	2.24722
All	$M_{2 \rightarrow 1}$	4.380	8.260	10.190	12.520	20.840	11.450	12.216
All	$M_{1 \rightarrow 2}$	3.360	7.240	8.450	13.000	20.780	10.810	11.246

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	-391.50	-118.17	-85.83
2	-91.92	-44.89	-34.61
3	-729.70	-175.25	-70.21
4	-305.72	-89.45	-42.90
5	-75.02	-41.82	-39.46
6	-231.65	-79.89	-49.08
7	-244.19	-86.55	-72.97
8	-124.17	-56.71	-52.68
All	-2179.93	-678.81	-433.81

(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 = 13.924531

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	5123688/10001843	0.51227
Θ_2	5229854/9999871	0.52299
$M_{2 \rightarrow 1}$	4828787/9997327	0.48301
$M_{1 \rightarrow 2}$	4890560/10000640	0.48902
Genealogies	19028987/40000319	0.47572

MCMC-Autocorrelation and Effective MCMC Sample Size

Parameter	Autocorrelation	Effective Sample Size
Θ_1	0.76184	120770.74
Θ_2	0.75011	126869.67
$M_{2 \rightarrow 1}$	0.92905	31568.12
$M_{1 \rightarrow 2}$	0.93076	30716.78
$\text{Ln}[\text{Prob}(D G)]$	0.84642	70168.40

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

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