Example: Microsatellite data set

MIGRATION RATE AND POPULATION SIZE ESTIMATION
using the coalescent and maximum likelihood or Bayesian inference
Migrate-n version 3.5.0 [2129]
Compiled for a PARALLEL COMPUTER ARCHITECTURE
One master and 8 compute nodes are available.
Compiled for a SYMMETRIC MULTIPROCESSORS
Program started at Sun Mar 3 14:20:09 2013
Program finished at Sun Mar 3 14:20:17 2013

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_numb	*	0
2 population_numb	*	*

Order of parameters:

1	Θ_1	<displayed>
2	Θ_2	<displayed>

4	$M_{1\to 2}$	<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	5.000000	500
M	Uniform	0.000000	10.000000	20.000000	5.000000	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
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	0.220	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	0.757

Bayesian Analysis: Posterior distribution table

Locus	Parameter	2.5%	25.0%	Mode	75.0%	97.5%	Median	Mean
1	Θ_1	4.88000	5.44000	0.00000	8.88000	0.00000	0.00000	10.82314
1	Θ_2	11.48000	14.52000	0.00000	16.28000	0.00000	14.18000	13.60132
1	$M_{1\text{->}2}$	0.000	0.280	0.000	0.800	0.000	1.260	1.219
2	Θ_1	2.80000	3.56000	0.00000	4.84000	0.00000	0.00000	4.70100
2	Θ_2	0.44000	0.60000	0.00000	2.48000	0.00000	2.94000	5.27173
2	$M_{1\text{->}2}$	0.000	7.960	0.000	11.040	0.000	9.660	9.728
3	Θ_1	5.12000	6.08000	0.00000	7.92000	0.00000	0.00000	8.05833
3	Θ_2	2.16000	2.88000	0.00000	4.96000	0.00000	7.50000	8.53377
3	$M_{1\text{->}2}$	0.000	3.600	0.000	5.840	0.000	5.460	5.669
4	Θ_1	1.60000	1.80000	0.00000	4.44000	0.00000	0.00000	9.12468
4	Θ_2	0.12000	0.24000	0.00000	1.68000	0.00000	1.82000	8.22277
4	$M_{1\text{->}2}$	0.000	1.560	0.000	3.520	0.000	4.580	9.716
5	Θ_1	1.32000	2.28000	0.00000	3.36000	0.00000	0.00000	2.83023
5	Θ_2	5.80000	8.56000	0.00000	12.24000	0.00000	12.38000	12.70291
5	$M_{1\text{->}2}$	0.000	5.360	0.000	7.880	0.000	6.180	6.589
6	Θ_1	5.92000	7.24000	0.00000	9.24000	0.00000	0.00000	8.88443
6	Θ_2	0.28000	0.60000	0.00000	1.24000	0.00000	1.06000	1.05900
6	$M_{1\text{->}2}$	0.000	3.200	0.000	5.920	0.000	9.060	10.924
7	Θ_1	1.00000	1.20000	0.00000	2.32000	0.00000	0.00000	3.45695
7	Θ_2	2.24000	2.76000	0.00000	5.08000	0.00000	6.66000	7.15833
7	$M_{1\text{->}2}$	0.000	0.120	0.000	1.440	0.000	1.380	1.618
8	Θ_1	2.80000	5.32000	0.00000	7.20000	0.00000	0.00000	6.17402
8	Θ_2	4.72000	5.52000	0.00000	9.32000	0.00000	9.90000	10.69693
8	$M_{1\text{->}2}$	0.000	1.000	0.000	1.840	0.000	1.660	1.761
9	Θ_1	1.32000	1.64000	0.00000	3.80000	0.00000	0.00000	5.27911
9	Θ_2	1.88000	3.12000	0.00000	5.32000	0.00000	4.66000	4.97139
9	$M_{1\text{->}2}$	0.000	1.400	0.000	3.000	0.000	4.420	4.617

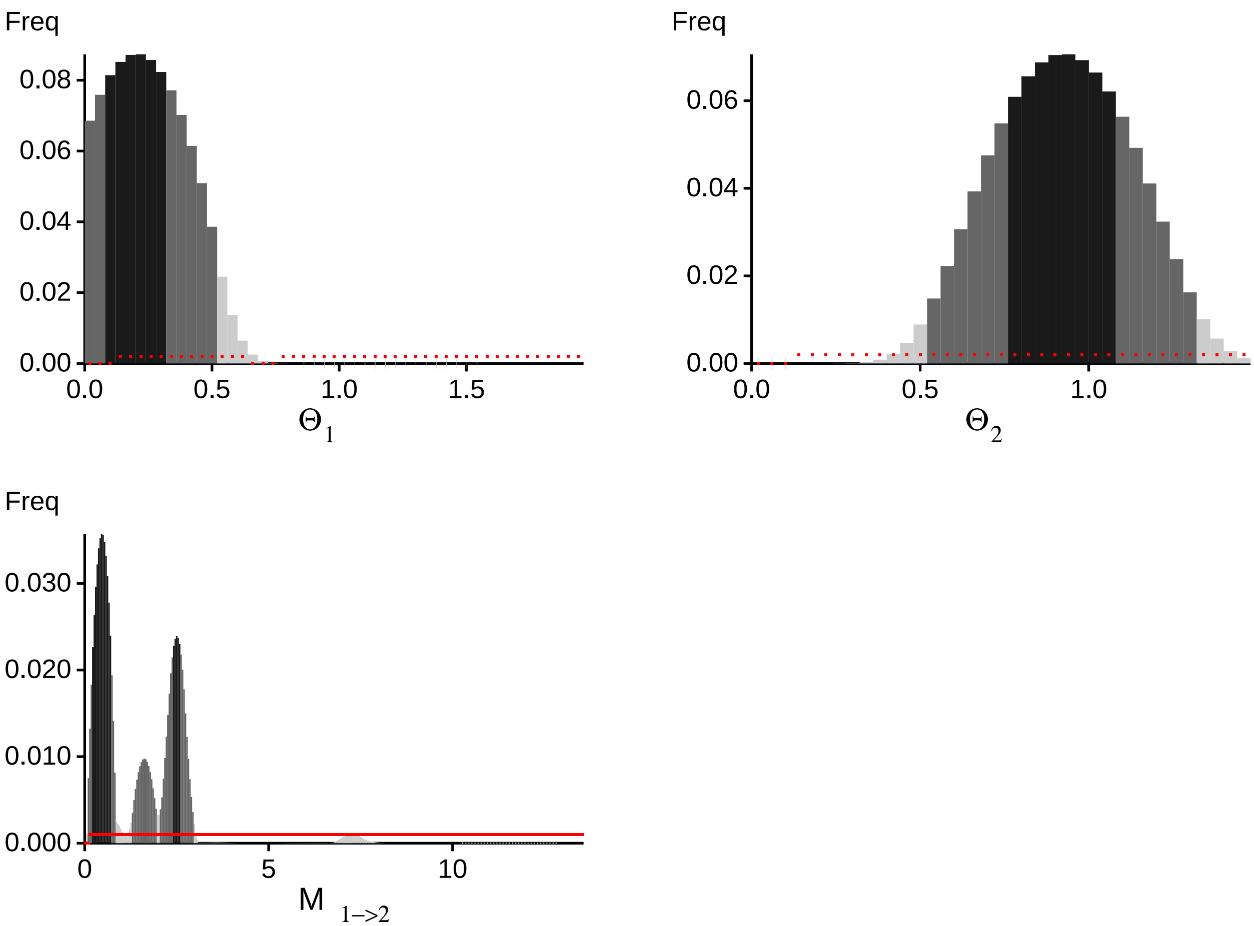
10	Θ_1	0.00000	0.00000	0.00000	0.56000	0.00000	0.00000	7.64207
10	Θ_2	0.12000	0.20000	0.00000	1.24000	0.00000	10.78000	8.47570
10	$M_{1->2}$	0.000	0.400	0.000	1.800	0.000	2.100	5.549
All	Θ_1	0.00000	0.04000	0.22000	0.32000	0.52000	0.30000	0.20407
All	Θ_2	0.48000	0.72000	0.94000	1.08000	1.32000	0.98000	0.92455
All	$M_{1->2}$	0.040	0.160	0.460	0.720	0.840	1.300	1.462

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 = Exp[ln(Prob(D | thisModel) - ln(Prob(D | otherModel)
or as LBF = 2 (ln(Prob(D | thisModel) - ln(Prob(D | otherModel))
shows the support for thisModel]

Locus	Raw thermodynamic score(1a)	Bezier approximation score(1b)	Harmonic mean(2)
1	-5883.42	-1068.00	-108.06
2	-828.99	-218.53	-46.36
3	-626.43	-199.04	-79.22
4	-4614.04	-854.34	-116.23
5	-1548.62	-321.00	-56.18
6	-9113.30	-1561.10	-58.56
7	-833.90	-222.87	-58.27
8	-694.38	-214.34	-75.01
9	-908.52	-247.97	-73.39
10	-40186.21	-6568.10	-184.60
All	-65064.46	-11301.94	-682.53

(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 = 173.345379
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	16696/16696	1.00000
Θ_2	16846/16846	1.00000
$M_{1->2}$	16633/16633	1.00000
Genealogies	15280/49825	0.30667

MCMC-Autocorrelation and Effective MCMC Sample Size

Parameter	Autocorrelation	Effective Sampe Size
Θ_1	0.93619	4731.39
Θ_2	0.91603	6325.69
$M_{1->2}$	0.91515	6631.69
Ln[Prob(D|G)]	0.99658	278.71

Potential Problems

This section reports potential problems with your run, but such reporting is often not very accurate. Whith 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 wether 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.

No warning was recorded during the run