Example: Microsatellite data set

MIGRATION RATE AND POPULATION SIZE ESTIMATION
using the coalescent and maximum likelihood or Bayesian inference
Migrate-n version 3.6.7 [March-23-15]
  Compiled for a PARALLEL COMPUTER ARCHITECTURE
  One master and 7 compute nodes are available.
Program started at   Thu Mar 26 15:19:52 2015
Program finished at   Thu Mar 26 15:19:59 2015
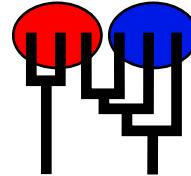
Options

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	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_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	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
   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	Θ_1	0.56000	0.96000	0.00000	1.88000	0.00000	0.00000	1.70807
1	Θ_2	0.28000	0.48000	0.00000	1.76000	0.00000	1.66000	3.04756
1	$M_{1->2}$	0.000	1.920	0.000	3.520	0.000	5.140	5.217
2	Θ_1	0.00000	0.04000	0.00000	0.52000	0.00000	0.00000	0.45217
2	Θ_2	0.00000	0.00000	0.00000	0.20000	0.00000	0.22000	0.11683
2	$M_{1->2}$	0.000	15.800	0.000	19.960	0.000	15.660	14.900
3	Θ_1	0.36000	0.60000	0.00000	1.32000	0.00000	0.00000	1.22085
3	Θ_2	0.00000	0.20000	0.00000	0.92000	0.00000	1.26000	1.38532
3	$M_{1->2}$	0.000	7.360	0.000	9.720	0.000	11.820	12.349
4	Θ_1	0.96000	1.48000	0.00000	2.24000	0.00000	0.00000	2.06922
4	Θ_2	0.12000	0.36000	0.00000	1.24000	0.00000	1.42000	1.59221
4	$M_{1->2}$	0.000	8.400	0.000	11.800	0.000	11.060	11.172
5	Θ_1	0.08000	0.76000	0.00000	1.48000	0.00000	0.00000	1.28955
5	Θ_2	0.00000	0.00000	0.00000	0.52000	0.00000	0.54000	0.55617
5	$M_{1->2}$	0.000	11.520	0.000	14.360	0.000	13.340	13.427
6	Θ_1	1.24000	1.72000	0.00000	2.84000	0.00000	0.00000	2.88401
6	Θ_2	0.00000	0.00000	0.00000	0.36000	0.00000	0.34000	0.27956
6	$M_{1->2}$	0.000	18.320	0.000	19.760	0.000	17.220	16.863
7	Θ_1	0.00000	0.00000	0.00000	0.20000	0.00000	0.00000	0.12122
7	Θ_2	0.00000	0.04000	0.00000	0.36000	0.00000	0.30000	0.24975
7	$M_{1->2}$	0.000	16.920	0.000	19.400	0.000	12.260	12.228
8	Θ_1	0.24000	1.04000	0.00000	2.48000	0.00000	0.00000	2.45958
8	Θ_2	0.00000	0.64000	0.00000	1.44000	0.00000	1.06000	1.04749
8	$M_{1->2}$	0.000	5.120	0.000	8.160	0.000	7.820	8.217
9	Θ_1	1.16000	1.76000	0.00000	2.48000	0.00000	0.00000	2.23824
9	Θ_2	0.68000	1.36000	0.00000	2.56000	0.00000	2.18000	2.38794
9	$M_{1->2}$	0.000	7.600	0.000	11.320	0.000	11.940	12.614

10	Θ_1	0.00000	0.36000	0.00000	0.96000	0.00000	0.00000	0.82314
10	Θ_2	0.00000	0.28000	0.00000	0.96000	0.00000	0.82000	0.91897
10	$M_{1->2}$	0.000	1.760	0.000	3.680	0.000	3.300	3.991
All	Θ_1	0.60000	1.40000	1.66000	2.00000	2.56000	1.70000	1.62365
All	Θ_2	0.00000	0.16000	0.38000	0.52000	0.72000	0.42000	0.37331
All	$M_{1->2}$	15.240	15.440	15.820	16.160	16.560	9.940	11.017

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
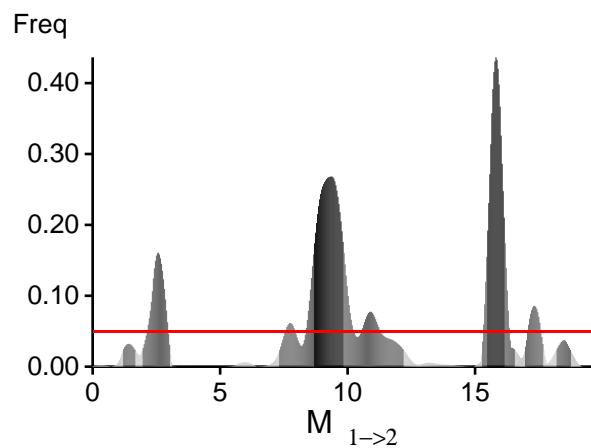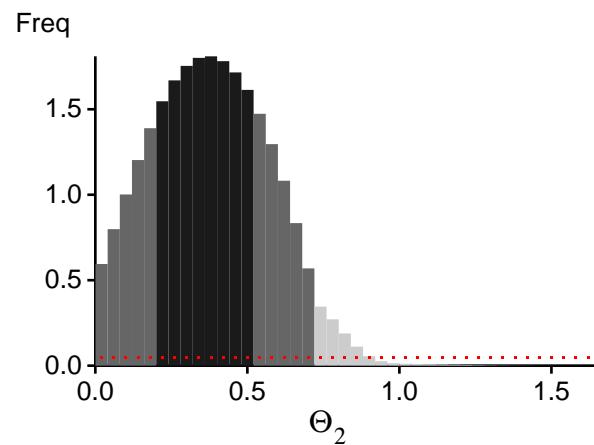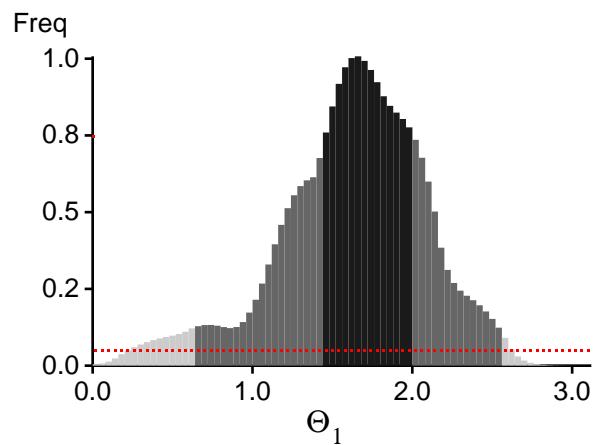
Log-Probability of the data given the model (marginal likelihood)

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	-13807.95	-2309.56	-72.88
2	-36095.54	-5895.21	-245.29
3	-2332.48	-438.36	-31.53
4	-47954.98	-7771.11	-115.36
5	-3488.38	-604.73	-18.15
6	-12584.94	-2085.05	-50.66
7	-27734.49	-4472.42	-67.72
8	-22885.28	-3721.65	-48.04
9	-20261.21	-3320.62	-48.40
10	-47551.06	-7681.86	-82.68
All	-234672.32	-38276.56	-756.71

(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 = 24.004377
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	16952/16952	1.00000
Θ_2	16645/16645	1.00000
$M_{1->2}$	16575/16575	1.00000
Genealogies	13291/49828	0.26674

MCMC-Autocorrelation and Effective MCMC Sample Size

Parameter	Autocorrelation	Effective Sampe Size
Θ_1	0.90694	7405.06
Θ_2	0.91983	6855.26
$M_{1->2}$	0.88859	9662.09
Ln[Prob(D|G)]	0.97789	1624.44

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