Example: Microsatellite data set

POPULATION SIZE, MIGRATION, DIVERGENCE, ASSIGNMENT, HISTORY
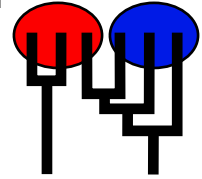Bayesian inference using the structured coalescent
Migrate-n version debug 4.2.5 [November-26-15]
  Using Intel AVX (Advanced Vector Extensions)
  Compiled for a SYMMETRIC multiprocessors (Grandcentral)
Program started at   Fri Nov 27 11:00:09 2015
Program finished at   Fri Nov 27 11:00:25 2015

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_numb   *   0
 2 population_numb   *   *

 Order of parameters:
  1   Θ_1 <displayed>

| 2 | Θ_2 | <displayed> |
| 3 | $M_{1->2}$ | <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]   1
   Number of concurrent chains (replicates) [c]   2
   Visited (sampled) parameter values [a*b*c]   10000
   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.msat
   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)

	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	0.000

Bayesian Analysis: Posterior distribution table

Locus	Parameter	2.5%	25.0%	Mode	75.0%	97.5%	Median	Mean
1	Θ_1	9.28000	11.64000	13.98000	15.04000	19.92000	14.34000	14.51016
1	Θ_2	7.08000	16.76000	19.18000	19.88000	20.00000	15.30000	14.63657
1	$M_{1->2}$	1.720	2.320	3.300	4.840	8.400	4.420	4.750
2	Θ_1	5.48000	10.40000	13.46000	15.92000	18.68000	12.50000	12.34998
2	Θ_2	1.16000	1.28000	2.50000	4.24000	4.92000	4.98000	7.59189
2	$M_{1->2}$	0.640	1.240	2.180	3.280	7.320	3.460	3.738
3	Θ_1	8.88000	11.28000	13.38000	14.96000	18.92000	13.62000	13.73221
3	Θ_2	4.08000	4.44000	6.18000	8.80000	19.16000	10.50000	11.11347
3	$M_{1->2}$	1.080	1.920	3.100	4.400	10.680	3.980	4.893
4	Θ_1	13.76000	17.48000	19.18000	19.72000	20.00000	17.82000	17.54064
4	Θ_2	12.36000	17.56000	19.22000	19.92000	20.00000	17.70000	17.21356
4	$M_{1->2}$	0.880	1.600	2.220	2.840	4.640	2.500	2.582
5	Θ_1	1.92000	2.60000	3.26000	4.08000	7.24000	4.34000	4.45623
5	Θ_2	2.96000	5.08000	5.86000	9.24000	14.52000	10.98000	11.34493
5	$M_{1->2}$	2.400	3.840	5.620	7.200	13.720	6.580	7.295
6	Θ_1	12.32000	16.00000	17.22000	19.04000	20.00000	16.74000	16.48717
6	Θ_2	3.20000	7.00000	8.78000	10.24000	13.92000	8.46000	8.64856
6	$M_{1->2}$	0.160	0.960	1.540	2.400	3.960	1.940	1.986
7	Θ_1	4.44000	5.68000	7.10000	8.72000	13.32000	7.98000	8.42471
7	Θ_2	3.76000	5.96000	7.58000	9.88000	16.00000	8.66000	9.22116
7	$M_{1->2}$	0.360	0.880	1.540	2.320	4.320	2.020	2.177
8	Θ_1	4.88000	9.44000	10.98000	12.32000	16.00000	10.74000	10.68421
8	Θ_2	6.16000	8.72000	10.38000	13.04000	19.96000	12.02000	12.40034
8	$M_{1->2}$	0.520	0.800	1.860	2.880	3.240	4.020	5.742
9	Θ_1	10.92000	12.40000	15.06000	16.56000	18.96000	14.02000	13.70250
9	Θ_2	10.56000	16.52000	19.18000	19.80000	20.00000	16.50000	16.06862
9	$M_{1->2}$	0.800	1.680	2.220	2.880	4.200	2.460	2.505

10	Θ_1	13.92000	16.92000	17.74000	19.32000	20.00000	17.50000	17.33407
10	Θ_2	9.16000	17.08000	19.18000	19.60000	20.00000	15.26000	15.07356
10	$M_{1->2}$	2.640	5.080	6.180	7.720	10.520	6.460	6.316
All	Θ_1	1.88000	2.68000	3.18000	3.60000	4.44000	3.22000	3.19120
All	Θ_2	1.04000	1.64000	2.06000	2.44000	3.08000	2.10000	2.07194
All	$M_{1->2}$	12.400	15.000	16.020	17.880	19.440	16.060	14.643

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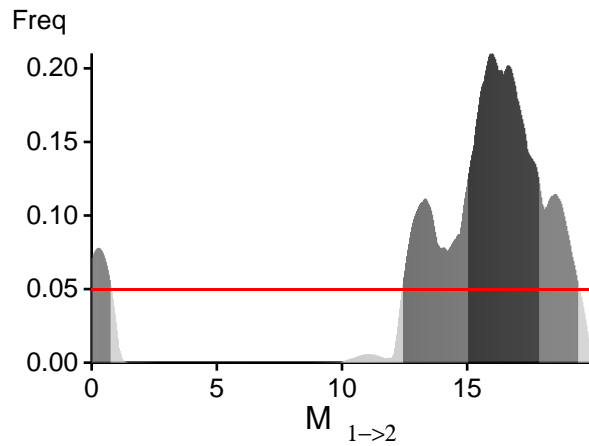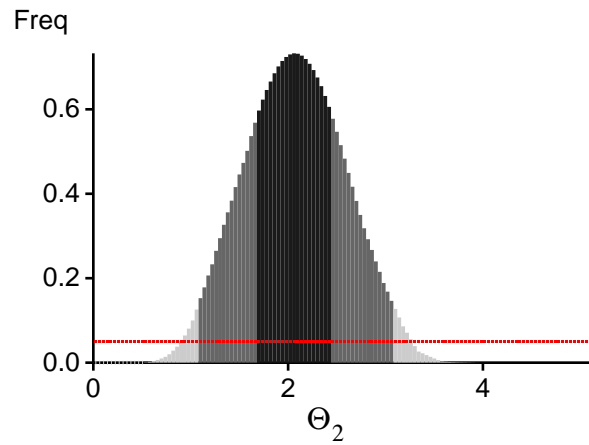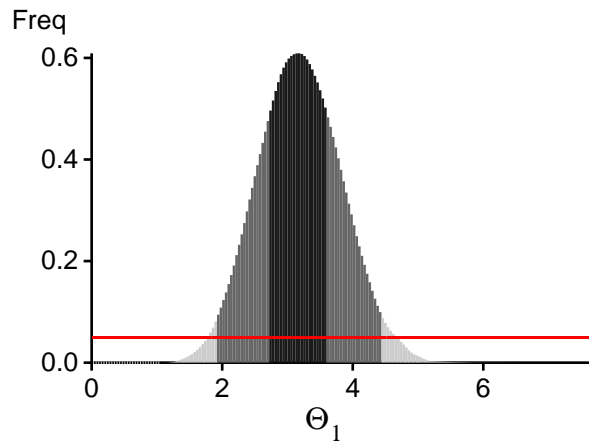
Log-Probability of the data given the model (marginal likelihood)

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	-3753.49	-729.30	-124.69
2	-733.06	-215.35	-91.29
3	-682.63	-211.56	-98.68
4	-5666.53	-1047.63	-127.28
5	-218.73	-110.96	-70.54
6	-3015.22	-595.52	-104.78
7	-650.85	-198.90	-85.76
8	-594.49	-199.68	-105.34
9	-5394.03	-973.49	-105.04
10	-3212.46	-640.85	-109.17
All	-23936.71	-4938.46	-1037.79

(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 = -15.214551]
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	16435/16435	1.00000
Θ_2	16804/16804	1.00000
$M_{1->2}$	16609/16609	1.00000
Genealogies	17143/50152	0.34182

MCMC-Autocorrelation and Effective MCMC Sample Size

Parameter	Autocorrelation	Effective Sampe Size
Θ_1	0.93637	3307.80
Θ_2	0.91193	4623.87
$M_{1 \rightarrow 2}$	0.91482	4461.34
Ln[Prob(D|G)]	0.99479	261.24

Potential Problems

This section reports potential problems with your run, but such reporting is often not very accurate. Whith many parameters in a multilocus analysi   s, it is very common that some parameters for some loci will not be very informative, triggering suggestions (for example to increase the prior ran   ge) that are not sensible. This suggestion tool will improve with time, therefore do not blindly follow its suggestions. If some parameters are fla gged, inspect the tables carefully and judge wether an action is required. For example, if you run a Bayesian inference with sequence data, for mac   roscopic 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 rou tes 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