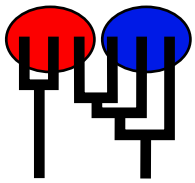
Example: Microsatellite data set

MIGRATION RATE AND POPULATION SIZE ESTIMATION
using the coalescent and maximum likelihood or Bayesian inference
Migrate-n version debug 3.2.17 [x]
Compiled for a SYMMETRIC MULTIPROCESSORS
Program started at Sun Oct 9 20:24:30 2011
Program finished at Sun Oct 9 20:25:18 2011



Options

Datatype: Microsatellite data [Brownian motion]
Missing data: not included

Inheritance scalers in use for Thetas: 1.00 1.00
1.00 1.00 1.00 1.00 1.00
1.00 1.00 1.00
[Each Theta uses the (true) ineritance scalar of the first locus as a 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	$M_{1->2}$	<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

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
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	0.193
19	0.040	0.071	0.056
18	0.060	0.119	0.090
15	0.220	0.024	0.122
21	0.020	0.167	0.093
23	0.020	0.119	0.070
17	0.280	0.095	0.188
22	0.060	0.119	0.090
25	0.060	0.024	0.042
24	0.020	0.000	0.010
26	0.000	0.024	0.012
27	0.000	0.048	0.024
29	0.000	0.024	0.012

Locus 2

Allele	Pop1	Pop2	All
16	0.520	0.571	0.546
19	0.040	0.000	0.020
18	0.220	0.119	0.170
17	0.160	0.167	0.163
15	0.020	0.000	0.010
21	0.020	0.071	0.046
20	0.020	0.024	0.022
22	0.000	0.048	0.024

Locus 3

Allele	Pop1	Pop2	All
19	0.240	0.262	0.251
20	0.280	0.476	0.378
18	0.080	0.095	0.088
21	0.280	0.119	0.200
22	0.120	0.048	0.084

Locus 4

Allele	Pop1	Pop2	All

Allele	Pop1	Pop2	All
16	0.080	0.071	0.076
24	0.180	0.024	0.102
15	0.020	0.048	0.034
25	0.160	0.167	0.163
14	0.020	0.048	0.034
19	0.100	0.143	0.121
12	0.060	0.000	0.030
20	0.080	0.190	0.135
23	0.060	0.119	0.090
28	0.020	0.000	0.010
22	0.060	0.024	0.042
21	0.160	0.119	0.140
13	0.000	0.024	0.012
26	0.000	0.024	0.012

Locus 5

Allele	Pop1	Pop2	All
20	0.400	0.524	0.462
21	0.420	0.357	0.389
19	0.180	0.119	0.150

Locus 6

Allele	Pop1	Pop2	All
19	0.060	0.000	0.030
20	0.100	0.024	0.062
18	0.300	0.214	0.257
22	0.200	0.119	0.160
21	0.120	0.476	0.298
16	0.060	0.000	0.030
24	0.160	0.048	0.104
17	0.000	0.119	0.060

Locus 7

Allele	Pop1	Pop2	All
23	0.040	0.238	0.139
20	0.660	0.143	0.401
22	0.180	0.190	0.185
21	0.100	0.333	0.217
19	0.020	0.095	0.058

Locus 8

Allele	Pop1	Pop2	All
19	0.520	0.524	0.522
17	0.040	0.048	0.044
18	0.100	0.071	0.086
20	0.140	0.190	0.165
16	0.080	0.000	0.040
22	0.100	0.048	0.074
15	0.020	0.048	0.034
23	0.000	0.071	0.036

Locus 9

Allele	Pop1	Pop2	All
24	0.080	0.024	0.052
19	0.300	0.429	0.364
20	0.300	0.167	0.233
23	0.180	0.143	0.161
22	0.080	0.024	0.052
18	0.020	0.071	0.046
21	0.040	0.095	0.068
25	0.000	0.048	0.024

Locus 10

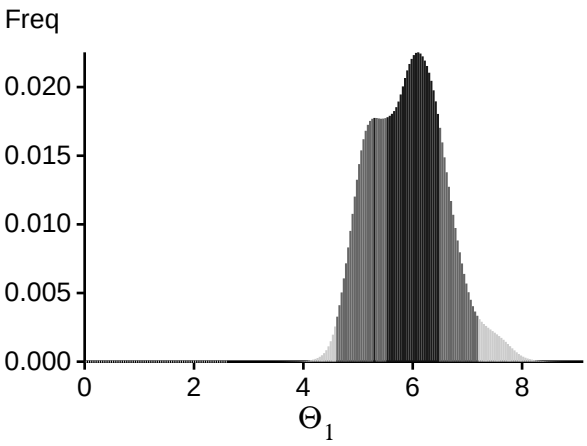
Allele	Pop1	Pop2	All
22	0.100	0.214	0.157
20	0.440	0.214	0.327
23	0.080	0.167	0.123
24	0.020	0.000	0.010
19	0.160	0.167	0.163
21	0.060	0.048	0.054
18	0.080	0.000	0.040
15	0.020	0.071	0.046
17	0.040	0.048	0.044
25	0.000	0.071	0.036

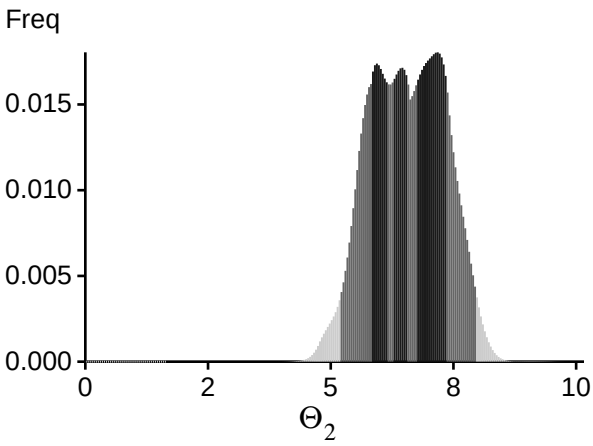
Bayesian Analysis: Posterior distribution table

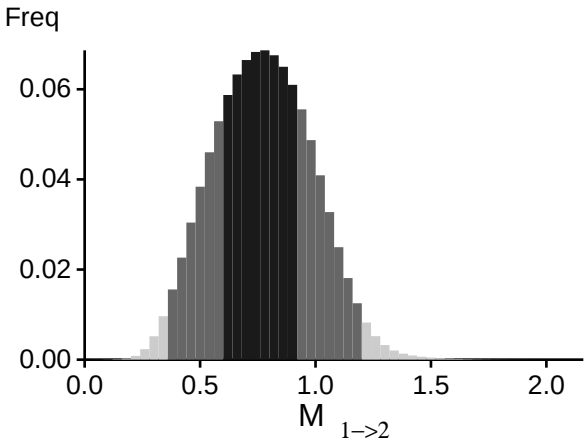
Locus	Parameter	2.5%	25.0%	Mode	75.0%	97.5%	Median	Mean
1	Θ_1	0.40000	0.56000	1.22000	2.08000	2.32000	2.74000	4.45556
1	Θ_2	0.52000	0.72000	1.18000	2.32000	2.92000	3.58000	4.73761
1	$M_{1->2}$	0.000	0.000	0.300	0.760	0.920	2.420	2.561
2	Θ_1	2.80000	3.96000	4.50000	5.64000	8.96000	5.22000	2.72766
2	Θ_2	4.56000	6.08000	7.02000	8.68000	12.92000	8.02000	4.22616
2	$M_{1->2}$	0.000	0.240	0.460	0.680	1.280	0.580	0.282
3	Θ_1	2.40000	3.60000	4.66000	5.32000	7.32000	4.70000	1.59238
3	Θ_2	1.44000	2.08000	2.66000	3.36000	8.08000	3.10000	1.25279
3	$M_{1->2}$	0.000	0.120	0.620	1.320	5.800	2.180	0.806
4	Θ_1	4.60000	5.72000	6.98000	8.08000	11.88000	7.62000	1.98870
4	Θ_2	3.36000	4.60000	5.50000	6.68000	10.12000	6.02000	1.62657
4	$M_{1->2}$	0.120	0.640	1.020	1.400	2.120	1.140	0.283
5	Θ_1	2.36000	2.68000	3.30000	4.40000	11.92000	6.06000	1.34937
5	Θ_2	1.72000	6.00000	6.46000	8.84000	13.88000	7.02000	1.46795
5	$M_{1->2}$	0.240	0.680	1.020	1.520	3.680	1.860	0.376
6	Θ_1	1.44000	3.04000	3.82000	4.64000	5.88000	3.78000	0.62760
6	Θ_2	0.80000	1.20000	1.54000	1.84000	2.68000	1.70000	0.28551
6	$M_{1->2}$	2.160	4.480	5.500	6.040	10.680	6.540	1.108
7	Θ_1	2.28000	3.92000	4.66000	5.68000	7.56000	4.90000	0.70744
7	Θ_2	1.96000	2.76000	3.98000	5.44000	9.40000	5.18000	0.80050
7	$M_{1->2}$	0.200	0.520	0.780	1.880	3.720	1.700	0.255
8	Θ_1	2.44000	2.72000	3.86000	5.04000	12.44000	5.78000	0.83500
8	Θ_2	2.00000	2.40000	3.10000	5.44000	12.04000	5.66000	0.94295
8	$M_{1->2}$	0.200	0.680	1.020	1.280	1.960	1.100	0.139
9	Θ_1	5.24000	6.36000	7.10000	8.36000	11.16000	7.78000	0.88399
9	Θ_2	4.04000	5.28000	5.90000	7.16000	10.04000	6.54000	0.75126
9	$M_{1->2}$	0.240	0.640	0.940	1.320	2.080	1.100	0.125

10	Θ_1	0.20000	0.28000	0.78000	1.48000	1.76000	4.34000	0.46388
10	Θ_2	0.00000	0.16000	0.62000	1.04000	1.52000	3.66000	0.37738
10	$M_{1->2}$	0.080	0.320	1.060	2.280	2.760	3.500	0.577
All	Θ_1	4.56000	5.48000	6.10000	6.48000	7.20000	5.94000	5.91176
All	Θ_2	5.16000	6.72000	7.18000	7.36000	7.96000	6.62000	6.56064
All	$M_{1->2}$	0.320	0.560	0.780	0.920	1.200	0.820	0.774

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	-28361.13	-4631.15	-84.72
2	-818.49	-214.46	-71.72
3	-741.76	-194.19	-50.64
4	-4314.87	-789.95	-86.77
5	-535.14	-156.61	-76.49
6	-3340.26	-595.30	-46.93
7	-722.15	-189.26	-71.86
8	-1427.39	-316.14	-81.59
9	-1119.87	-275.76	-88.54
10	-6074.74	-1050.17	-157.09
All	-47504.49	-8461.67	-865.03

(1a, 1b and 2) is an approximation to the marginal likelihood, make sure the program run long enough!
(1a, 1b) and (2) should give a similar result, (2) is considered more
crude than (1), but (1) needs heating with several well-spaced chains,
(1b) is using a Bezier-curve to get better approximations for runs with low number
of heated chains
[Scaling factor = -48.672332

Acceptance ratios for all parameters and the genealogies

Parameter	Accepted changes	Ratio
Θ_1	16707/16707	1.00000
Θ_2	16672/16672	1.00000
$M_{1->2}$	16825/16825	1.00000
Genealogies	14890/49816	0.29890

MCMC-Autocorrelation and Effective MCMC Sample Size

Parameter	Autocorrelation	Effective Sampe Size
Θ_1	0.92931	3619.95
Θ_2	0.89824	4375.93
$M_{1->2}$	0.91471	4134.21
Ln[Prob(D|G)]	0.99589	285.94

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