

Example: Microsatellite data set

MIGRATION RATE AND POPULATION SIZE ESTIMATION

using the coalescent and maximum likelihood or Bayesian inference

Migrate-n version debug 3.2.13 [x]

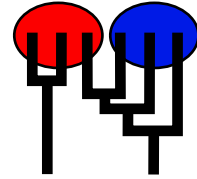
Compiled for a PARALLEL COMPUTER ARCHITECTURE

One master and 2 compute nodes are available.

Compiled for a SYMMETRIC MULTIPROCESSORS

Program started at Sat Jun 4 17:44:55 2011

Program finished at Sat Jun 4 21:14:45 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) inheritance 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_num * 0

2 population_num * *

Order of parameters:

1 Θ_1 <displayed>

2 Θ_2 <displayed>
 4 $M_{1 \rightarrow 2}$ <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])	100
Number of concurrent chains (replicates) [c]	2
Visited (sampled) parameter values [a*b*c]	1000000
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
 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

?	?	?	?
---	---	---	---

?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?

Locus 2

?	?	?	?
---	---	---	---

?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?

Locus 3

?	?	?	?
---	---	---	---

?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?

Locus 4

?	?	?	?
---	---	---	---

?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?
Locus 5			
?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?
Locus 6			
?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?
Locus 7			
?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?

Locus 8

?	?	?	?
---	---	---	---

?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?

Locus 9

?	?	?	?
---	---	---	---

?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?

Locus 10

?	?	?	?
---	---	---	---

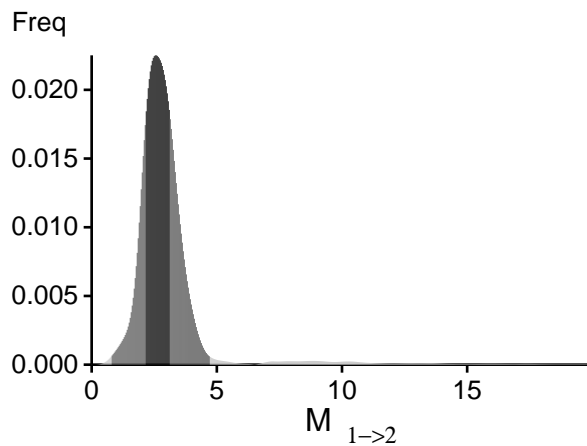
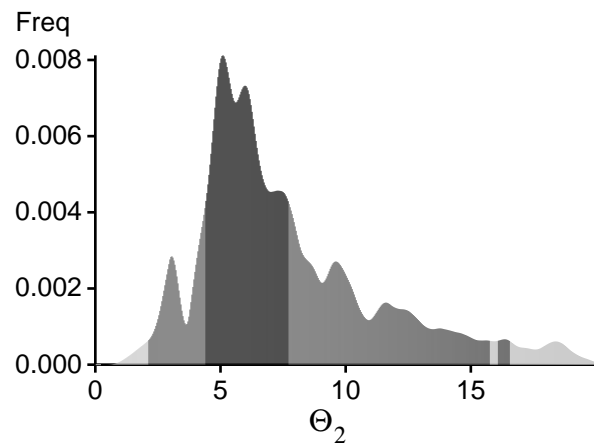
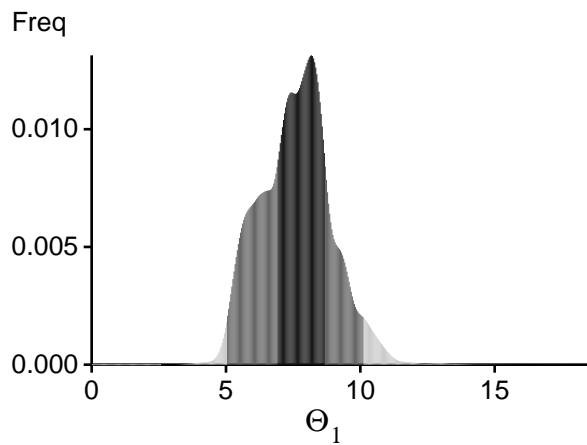
?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?

Bayesian Analysis: Posterior distribution table

Locus	Parameter	2.5%	25.0%	Mode	75.0%	97.5%	Median	Mean
1	Θ_1	7.60000	12.68000	0.00000	17.72000	0.00000	0.00000	14.08060
1	Θ_2	8.32000	16.04000	0.00000	19.88000	0.00000	15.66000	15.02241
1	$M_{1 \rightarrow 2}$	0.000	0.240	0.000	1.200	0.000	1.060	1.193
2	Θ_1	3.16000	4.92000	0.00000	8.88000	0.00000	0.00000	8.44529
2	Θ_2	1.24000	1.60000	0.00000	8.44000	0.00000	8.58000	9.31743
2	$M_{1 \rightarrow 2}$	0.000	16.720	0.000	19.720	0.000	12.420	11.879
3	Θ_1	1.32000	3.40000	0.00000	7.76000	0.00000	0.00000	7.26715
3	Θ_2	1.84000	2.28000	0.00000	8.76000	0.00000	9.02000	9.65277
3	$M_{1 \rightarrow 2}$	0.000	3.360	0.000	10.360	0.000	9.620	10.073
4	Θ_1	11.32000	17.20000	0.00000	19.88000	0.00000	0.00000	16.67780
4	Θ_2	2.96000	13.76000	0.00000	19.64000	0.00000	12.02000	11.67922
4	$M_{1 \rightarrow 2}$	0.000	1.200	0.000	8.120	0.000	7.980	8.845
5	Θ_1	0.60000	1.40000	0.00000	3.36000	0.00000	0.00000	3.24399
5	Θ_2	2.32000	2.52000	0.00000	7.20000	0.00000	9.98000	10.12323
5	$M_{1 \rightarrow 2}$	0.000	14.960	0.000	19.960	0.000	13.900	13.243
6	Θ_1	3.28000	5.32000	0.00000	10.44000	0.00000	0.00000	9.67723
6	Θ_2	0.00000	0.40000	0.00000	2.40000	0.00000	2.26000	3.27446
6	$M_{1 \rightarrow 2}$	0.000	0.800	0.000	7.080	0.000	6.980	7.814
7	Θ_1	0.84000	2.08000	0.00000	4.52000	0.00000	0.00000	4.36873
7	Θ_2	0.16000	2.84000	0.00000	9.72000	0.00000	8.34000	9.00202
7	$M_{1 \rightarrow 2}$	0.000	0.600	0.000	3.360	0.000	3.100	4.257
8	Θ_1	3.40000	5.52000	0.00000	9.04000	0.00000	0.00000	9.34483
8	Θ_2	2.12000	2.76000	0.00000	6.48000	0.00000	9.90000	10.17295
8	$M_{1 \rightarrow 2}$	0.000	14.720	0.000	19.720	0.000	12.380	11.787
9	Θ_1	3.36000	5.76000	0.00000	11.24000	0.00000	0.00000	10.25692
9	Θ_2	0.64000	0.72000	0.00000	2.16000	0.00000	9.94000	9.96279
9	$M_{1 \rightarrow 2}$	0.000	17.680	0.000	19.800	0.000	10.500	10.603

10	Θ_1	8.80000	15.24000	0.00000	19.76000	0.00000	0.00000	14.94378
10	Θ_2	0.84000	2.08000	0.00000	8.00000	0.00000	8.50000	9.05914
10	$M_{1 \rightarrow 2}$	0.000	0.240	0.000	4.680	0.000	7.100	8.047
All	Θ_1	5.00000	6.88000	8.18000	8.68000	10.12000	7.70000	7.60998
All	Θ_2	2.08000	4.36000	5.10000	7.72000	15.76000	6.66000	7.68575
All	$M_{1 \rightarrow 2}$	0.760	2.120	2.580	3.120	4.720	2.780	3.059

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	-8835.91	-1544.44	-129.55
2	-1349.13	-309.95	-91.26
3	-1018.92	-261.72	-88.67
4	-7712.43	-1367.30	-120.82
5	-362.47	-131.11	-70.56
6	-7453.36	-1298.67	-83.26
7	-1045.98	-258.07	-86.32
8	-1608.54	-359.53	-90.52
9	-2500.67	-505.98	-97.81
10	-2151.22	-468.34	-119.26
All	-34098.20	-6564.69	-1037.59

(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 = -59.573040

Acceptance ratios for all parameters and the genealogies

Parameter	Accepted changes	Ratio
Θ_1	8502495/8502495	1.00000
Θ_2	8499625/8499625	1.00000
$M_{1 \rightarrow 2}$	8503140/8503140	1.00000
Genealogies	7910705/25004515	0.31637

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.

No warning was recorded during the run