

Example: Microsatellite data set

MIGRATION RATE AND POPULATION SIZE ESTIMATION

using the coalescent and maximum likelihood or Bayesian inference

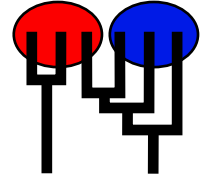
Migrate-n version 3.0

Compiled for a PARALLEL COMPUTER ARCHITECTURE

One master and 2 compute nodes are available.

Program started at Wed Jul 30 21:39:15 2008

Program finished at Wed Jul 30 21:40:28 2008



Options

Datatype:	Microsatellite data [Brownian motion]
Missing data:	not included
Random number seed:	(with internal timer) 822142067
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__num	*	*
2 population__num	*	*

Order of parameters:

1	Θ_1	<displayed>
2	Θ_2	<displayed>
3	$M_{2 \rightarrow 1}$	<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	Exponential	0.000000	1.000000	10.000000	-	200
M	Exponential	0.000000	10.000000	1000.000000	-	200

Markov chain settings:

Long chain

Number of chains	1
Recorded steps [a]	1600
Increment (record every x step [b])	2
Number of concurrent chains (replicates) [c]	2
Visited (sampled) parameter values [a*b*c]	6400
Number of discard trees per chain (burn-in)	130

Multiple Markov chains:

Static heating scheme

4 chains with temperatures

5.00 3.67 2.33 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
16	0.080	0.071	0.076
24	0.180	0.024	0.102

Allele	Pop1	Pop2	All
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

Allele	Pop1	Pop2	All
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

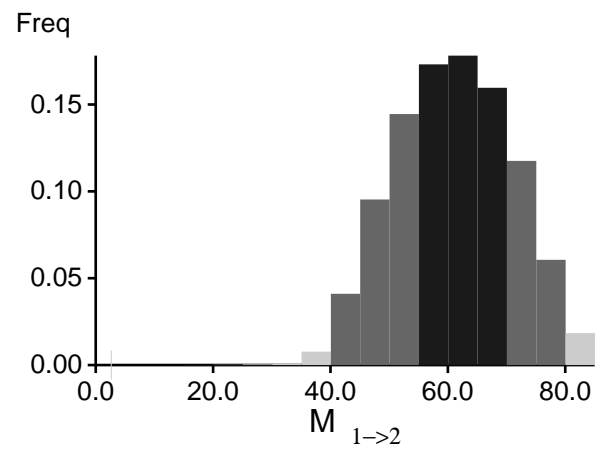
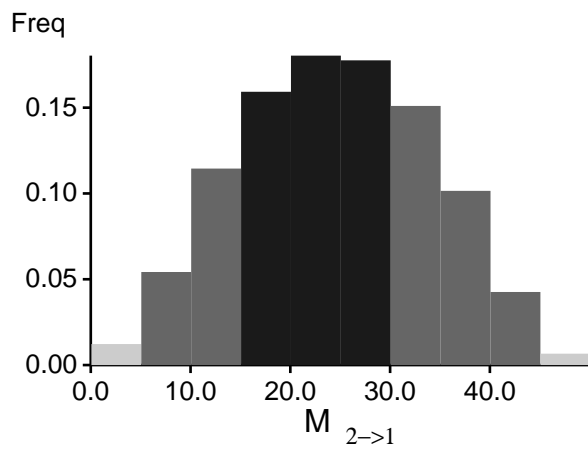
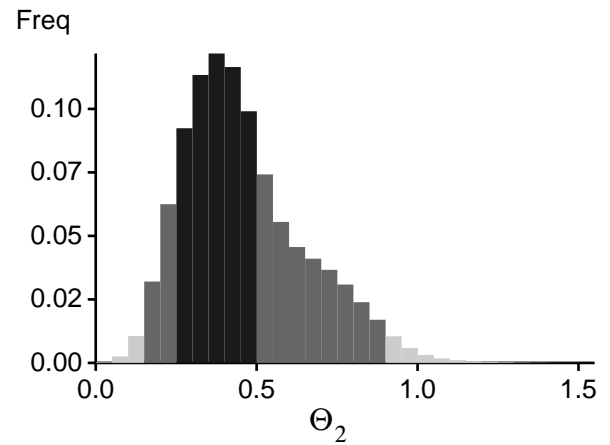
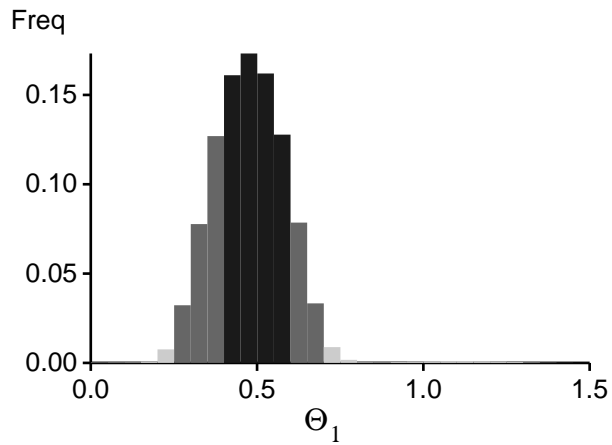
Locus 10

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	0.25000	0.40000	0.72500	1.00000	3.50000	2.02500	2.25155
1	Θ_2	0.15000	0.40000	0.72500	1.40000	2.60000	6.82500	5.29596
1	$M_{2 \rightarrow 1}$	0.000	5.000	17.500	25.000	35.000	22.500	17.162
1	$M_{1 \rightarrow 2}$	0.000	5.000	22.500	35.000	65.000	32.500	29.070
2	Θ_1	0.00000	0.00000	0.02500	0.15000	0.50000	0.17500	0.16134
2	Θ_2	0.00000	0.00000	0.02500	0.20000	1.00000	0.67500	1.07072
2	$M_{2 \rightarrow 1}$	130.000	280.000	307.500	330.000	350.000	427.500	419.550
2	$M_{1 \rightarrow 2}$	45.000	55.000	77.500	90.000	100.000	187.500	250.049
3	Θ_1	0.05000	0.15000	0.37500	0.50000	1.15000	0.47500	0.52585
3	Θ_2	0.00000	0.05000	0.32500	0.60000	3.75000	1.77500	1.73347
3	$M_{2 \rightarrow 1}$	5.000	30.000	47.500	60.000	70.000	47.500	40.778
3	$M_{1 \rightarrow 2}$	0.000	10.000	27.500	40.000	85.000	47.500	44.831
4	Θ_1	0.00000	0.05000	0.32500	0.75000	4.20000	1.42500	1.70573
4	Θ_2	0.00000	0.05000	0.32500	1.60000	5.25000	1.57500	2.18912
4	$M_{2 \rightarrow 1}$	10.000	15.000	32.500	50.000	60.000	47.500	52.166
4	$M_{1 \rightarrow 2}$	5.000	15.000	32.500	45.000	85.000	42.500	43.774
5	Θ_1	0.70000	0.90000	1.12500	1.40000	1.65000	1.12500	1.00410
5	Θ_2	0.00000	0.00000	0.22500	0.35000	1.20000	0.32500	0.41303
5	$M_{2 \rightarrow 1}$	0.000	0.000	12.500	20.000	35.000	22.500	15.516
5	$M_{1 \rightarrow 2}$	35.000	100.000	127.500	150.000	185.000	132.500	122.937
6	Θ_1	0.05000	0.95000	1.27500	1.85000	3.00000	1.47500	1.55596
6	Θ_2	0.00000	1.00000	1.32500	1.65000	3.90000	1.42500	1.66826
6	$M_{2 \rightarrow 1}$	20.000	40.000	62.500	70.000	90.000	62.500	56.747
6	$M_{1 \rightarrow 2}$	30.000	50.000	62.500	75.000	85.000	67.500	63.652
7	Θ_1	0.00000	0.00000	0.02500	0.10000	0.20000	0.12500	0.02080
7	Θ_2	0.00000	0.00000	0.02500	0.20000	1.05000	0.22500	0.32897
7	$M_{2 \rightarrow 1}$	465.000	805.000	867.500	955.000	1000.000	787.500	743.893
7	$M_{1 \rightarrow 2}$	320.000	665.000	762.500	795.000	855.000	607.500	595.102
8	Θ_1	0.00000	0.35000	0.62500	1.00000	1.40000	0.87500	1.49873
8	Θ_2	0.00000	0.65000	0.92500	1.20000	4.15000	1.02500	1.54294
8	$M_{2 \rightarrow 1}$	5.000	30.000	52.500	65.000	80.000	52.500	47.920
8	$M_{1 \rightarrow 2}$	20.000	45.000	62.500	70.000	90.000	62.500	59.293

9	Θ_1	0.35000	0.95000	1.32500	1.85000	3.15000	1.57500	1.66080
9	Θ_2	0.20000	0.40000	1.22500	1.70000	2.90000	1.62500	2.83021
9	$M_{2 \rightarrow 1}$	0.000	5.000	17.500	30.000	45.000	27.500	20.138
9	$M_{1 \rightarrow 2}$	0.000	20.000	32.500	45.000	60.000	37.500	33.376
10	Θ_1	0.00000	0.35000	0.57500	0.85000	3.30000	1.42500	1.66067
10	Θ_2	0.00000	0.00000	0.22500	0.55000	3.15000	0.92500	1.12130
10	$M_{2 \rightarrow 1}$	0.000	0.000	17.500	25.000	45.000	22.500	19.433
10	$M_{1 \rightarrow 2}$	0.000	40.000	57.500	80.000	135.000	67.500	67.333
All	Θ_1	0.20000	0.35000	0.47500	0.60000	0.70000	0.52500	0.48034
All	Θ_2	0.10000	0.20000	0.37500	0.50000	0.90000	0.47500	0.46188
All	$M_{2 \rightarrow 1}$	0.000	10.000	22.500	30.000	45.000	27.500	24.424
All	$M_{1 \rightarrow 2}$	35.000	50.000	62.500	70.000	80.000	67.500	61.018

Bayesian Analysis: Posterior distribution over all loci

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

Use this value for Bayes factor calculations:

$BF = \text{Exp}[\ln(\text{Prob}(D \mid \text{thisModel})) - \ln(\text{Prob}(D \mid \text{otherModel}))]$
shows the support for thisModel]

Method	$\ln(\text{Prob}(D \mid \text{Model}))$	Notes
Thermodynamic integration	-150086.288686	(1)
Harmonic mean	-2496195.140726	(1)

(1 and 2) is an approximation to the marginal likelihood, make sure the program run long enough!

(1) and (2) should give a similar result, (2) is considered more crude than (1), but (1) needs heating with several well-spaced chains,

Acceptance ratios for all parameters and the genealogies

Parameter	Accepted changes	Ratio
Θ_1	42105/42105	1.00000
Θ_2	41965/41965	1.00000
$M_{2 \rightarrow 1}$	41020/41020	1.00000
$M_{1 \rightarrow 2}$	41425/41425	1.00000
Genealogies	47710/160320	0.29759

MCMC-Autocorrelation and Effective MCMC Sample Size

Parameter	Autocorrelation	Effective Sample Size
Θ_1	0.96220	3116.54
Θ_2	0.95940	3338.63
$M_{2 \rightarrow 1}$	0.93855	5150.78
$M_{1 \rightarrow 2}$	0.93812	5168.22
$\text{Ln}[\text{Prob}(D G)]$	0.98654	1087.98