

Example: Microsatellite data set

MIGRATION RATE AND POPULATION SIZE ESTIMATION

using the coalescent and maximum likelihood or Bayesian inference

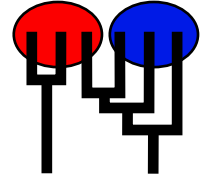
Migrate-n version debug 3.0

Compiled for a PARALLEL COMPUTER ARCHITECTURE

One master and 4 compute nodes are available.

Program started at Wed Oct 15 13:14:42 2008

Program finished at Wed Oct 15 13:15:36 2008



Options

Datatype: Microsatellite data [Brownian motion]
 Missing data: not included
 Random number seed: (with internal timer) 120227645
 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

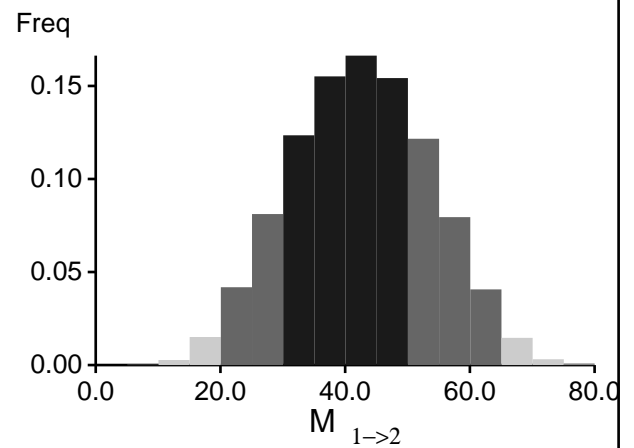
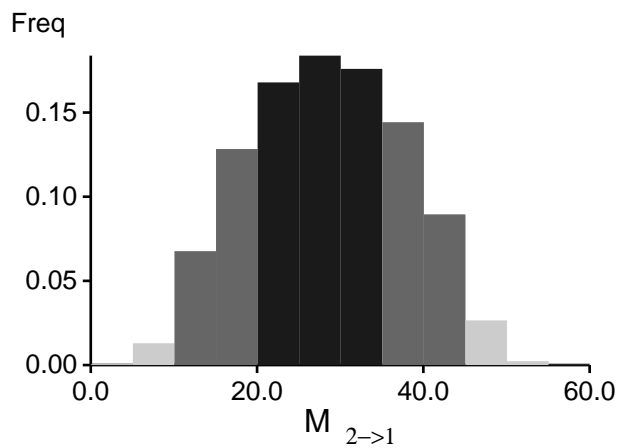
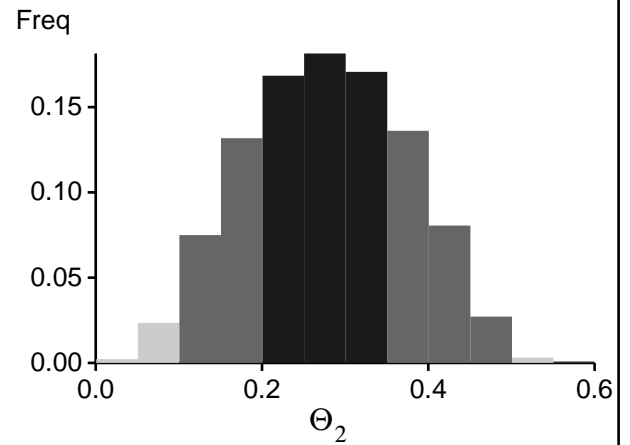
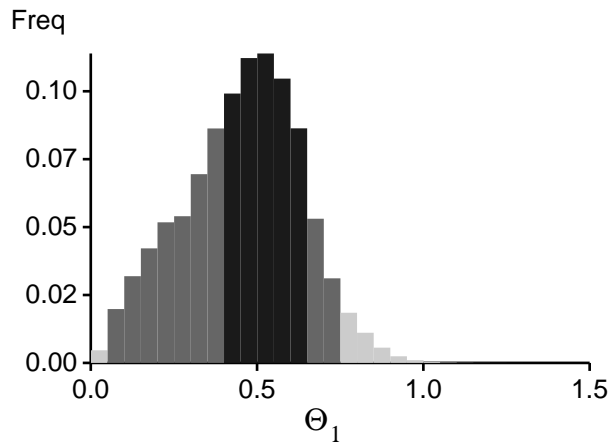
Allele	Pop1	Pop2	All
18	0.100	0.071	0.086
20	0.140	0.190	0.165
16	0.080	0.000	0.040
22	0.100	0.048	0.074
15	0.020	0.048	0.034
23	0.000	0.071	0.036
Locus 9			
Allele	Pop1	Pop2	All
24	0.080	0.024	0.052
19	0.300	0.429	0.364
20	0.300	0.167	0.233
23	0.180	0.143	0.161
22	0.080	0.024	0.052
18	0.020	0.071	0.046
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	0.15000	0.35000	0.57500	0.70000	3.00000	1.32500	1.56952
1	Θ_2	0.15000	0.30000	0.72500	1.35000	2.35000	4.17500	4.52974
1	$M_{2 \rightarrow 1}$	0.000	10.000	22.500	35.000	45.000	27.500	23.588
1	$M_{1 \rightarrow 2}$	0.000	5.000	22.500	30.000	50.000	27.500	23.391
2	Θ_1	0.00000	0.00000	0.02500	0.10000	0.25000	0.12500	0.06537
2	Θ_2	0.00000	0.00000	0.12500	0.20000	0.45000	0.22500	0.17390
2	$M_{2 \rightarrow 1}$	185.000	285.000	317.500	345.000	455.000	377.500	397.042
2	$M_{1 \rightarrow 2}$	215.000	245.000	297.500	325.000	395.000	282.500	268.183
3	Θ_1	0.00000	0.30000	0.52500	0.70000	1.30000	0.57500	0.59250
3	Θ_2	0.00000	0.10000	0.32500	0.65000	3.65000	1.82500	1.76407
3	$M_{2 \rightarrow 1}$	15.000	35.000	47.500	65.000	85.000	57.500	52.138
3	$M_{1 \rightarrow 2}$	35.000	55.000	67.500	80.000	100.000	72.500	70.238
4	Θ_1	0.00000	0.10000	0.27500	0.45000	3.15000	1.57500	1.67310
4	Θ_2	0.00000	0.10000	0.67500	1.00000	3.50000	0.97500	1.37376
4	$M_{2 \rightarrow 1}$	15.000	35.000	52.500	65.000	80.000	57.500	49.698
4	$M_{1 \rightarrow 2}$	20.000	35.000	52.500	65.000	125.000	72.500	72.446
5	Θ_1	0.05000	0.40000	0.67500	0.90000	1.30000	0.72500	0.70197
5	Θ_2	0.00000	0.05000	0.17500	0.30000	0.40000	0.22500	0.19397
5	$M_{2 \rightarrow 1}$	0.000	5.000	22.500	40.000	85.000	37.500	38.752
5	$M_{1 \rightarrow 2}$	35.000	60.000	82.500	100.000	175.000	102.500	104.871
6	Θ_1	0.10000	0.25000	0.52500	1.20000	2.35000	1.17500	1.94971
6	Θ_2	0.05000	0.20000	0.47500	0.95000	3.45000	1.97500	2.43047
6	$M_{2 \rightarrow 1}$	5.000	20.000	32.500	45.000	60.000	37.500	35.017
6	$M_{1 \rightarrow 2}$	15.000	25.000	42.500	50.000	70.000	47.500	43.720
7	Θ_1	0.00000	0.00000	0.02500	0.10000	0.20000	0.12500	0.02760
7	Θ_2	0.00000	0.00000	0.02500	0.10000	0.30000	0.12500	0.08648
7	$M_{2 \rightarrow 1}$	455.000	530.000	622.500	665.000	765.000	662.500	687.250
7	$M_{1 \rightarrow 2}$	335.000	365.000	392.500	430.000	840.000	597.500	576.271
8	Θ_1	0.00000	0.05000	0.27500	0.50000	0.60000	1.52500	1.71811
8	Θ_2	0.00000	0.05000	0.22500	0.35000	1.95000	0.87500	1.04200
8	$M_{2 \rightarrow 1}$	20.000	35.000	52.500	65.000	80.000	57.500	50.552
8	$M_{1 \rightarrow 2}$	105.000	115.000	132.500	150.000	165.000	77.500	86.850

9	Θ_1	0.30000	0.40000	0.67500	1.05000	1.20000	3.47500	3.47425
9	Θ_2	0.15000	0.30000	0.62500	1.00000	1.55000	3.87500	3.86017
9	$M_{2 \rightarrow 1}$	0.000	5.000	22.500	30.000	45.000	27.500	22.150
9	$M_{1 \rightarrow 2}$	0.000	10.000	27.500	35.000	50.000	32.500	25.440
10	Θ_1	0.00000	0.25000	0.57500	0.85000	2.25000	0.77500	1.58584
10	Θ_2	0.00000	0.45000	0.67500	0.85000	1.65000	0.72500	1.37334
10	$M_{2 \rightarrow 1}$	0.000	0.000	12.500	20.000	30.000	17.500	13.814
10	$M_{1 \rightarrow 2}$	5.000	20.000	32.500	45.000	75.000	42.500	38.820
All	Θ_1	0.00000	0.35000	0.52500	0.65000	0.75000	0.52500	0.45455
All	Θ_2	0.05000	0.15000	0.27500	0.35000	0.50000	0.32500	0.27738
All	$M_{2 \rightarrow 1}$	5.000	15.000	27.500	35.000	45.000	32.500	28.348
All	$M_{1 \rightarrow 2}$	15.000	25.000	42.500	50.000	65.000	47.500	42.440

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	-66540.109101	(1)
Harmonic mean	-2322299.934093	(2)

(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	23486/23486	1.00000
Θ_2	23574/23574	1.00000
$M_{2 \rightarrow 1}$	23512/23512	1.00000
$M_{1 \rightarrow 2}$	23166/23166	1.00000
Genealogies	26300/89196	0.29486

MCMC-Autocorrelation and Effective MCMC Sample Size

Parameter	Autocorrelation	Effective Sample Size
Θ_1	0.96086	1986.74
Θ_2	0.95498	2247.76
$M_{2 \rightarrow 1}$	0.92918	3269.16
$M_{1 \rightarrow 2}$	0.93001	3351.00
$\text{Ln}[\text{Prob}(D G)]$	0.98750	571.46