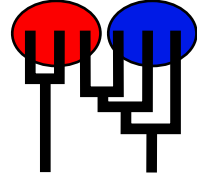


Example: Microsatellite data set

MIGRATION RATE AND POPULATION SIZE ESTIMATION
 using the coalescent and maximum likelihood or Bayesian inference
 Migrate-n version 3.6.8 [March-26-15]
 Compiled for a PARALLEL COMPUTER ARCHITECTURE
 One master and 10 compute nodes are available.
 Program started at Thu May 7 17:51:20 2015
 Program finished at Thu May 7 17:52:42 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]

Data set was subsampled: used a random sample of size: 5 and seed 13

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	1000000.00	4 chains with temperatures	3.00	1.50	1.00
		Swapping interval is 1			

Print options:

Data file:	infile
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
 [Data was used as repeat-length information]

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)

Subsampled dataset

Data set was subsampled randomly per population: 5 samples taken

Locus Population Individuals

1	population	0BAC_0BBU 0BBQ_0BBD 0BBO_0BBJ 0BAV_0BAO 0BAX_0BCA
	population	1BBN_1BAX 1BAQ_1BAN 1BAL_1BAV 1BBK_1BAS 1BAP_1BBP
2	population	0BAF_0BBV 0BAL_0BBL 0BAK_0BAD 0BBA_0BAI 0BAC_0BBU
	population	1BAH_1BAT 1BAE_1BAI 1BBF_1BAO 1BBM_1BAM 1BBJ_1BBC
3	population	0BAK_0BAD 0BBK_0BAH 0BBR_0BAW 0BBF_0BAE 0BAN_0BAT
	population	1BBE_1BAJ 1BAL_1BAV 1BAE_1BAI 1BBM_1BAM 1BBD_1BAG
4	population	0BBB_0BBX 0BAU_0BAP 0BAV_0BAO 0BAS_0BBT 0BBF_0BAE
	population	1BAE_1BAI 1BAQ_1BAN 1BBD_1BAG 1BAL_1BAV 1BBF_1BAO
5	population	0BAJ_0BAB 0BAU_0BAP 0BBF_0BAE 0BAC_0BBU 0BAG_0BBI
	population	1BBB_1BBR 1BAC_1BAB 1BAR_1BBA 1BBN_1BAX 1BBK_1BAS
6	population	0BAL_0BBL 0BCB_0BAM 0BAR_0BBW 0BBA_0BAI 0BAJ_0BAB
	population	1BBH_1BBO 1BAR_1BBA 1BBQ_1BBI 1BBL_1BAK 1BBG_1BAU
7	population	0BAU_0BAP 0BAV_0BAO 0BAG_0BBI 0BBF_0BAE 0BBB_0BBX
	population	1BBK_1BAS 1BBQ_1BBI 1BAA_1BAF 1BAE_1BAI 1BBL_1BAK
8	population	0BAF_0BBV 0BBO_0BBJ 0BBP_0BBC 0BCB_0BAM 0BAJ_0BAB
	population	1BBE_1BAJ 1BBH_1BBO 1BAP_1BBP 1BAD_1BAW 1BBL_1BAK
9	population	0BAU_0BAP 0BBR_0BAW 0BBH_0BAA 0BAG_0BBI 0BAS_0BBT
	population	1BAA_1BAF 1BAP_1BBP 1BAC_1BAB 1BBB_1BBR 1BAE_1BAI
10	population	0BBA_0BAI 0BAX_0BCA 0BCB_0BAM 0BAU_0BAP 0BBH_0BAA
	population	1BAP_1BBP 1BAR_1BBA 1BAQ_1BAN 1BBQ_1BBI 1BAD_1BAW

Bayesian Analysis: Posterior distribution table

Locus	Parameter	2.5%	25.0%	Mode	75.0%	97.5%	Median	Mean
1	Θ_1	9.08000	15.96000	0.00000	19.88000	0.00000	0.00000	15.17943
1	Θ_2	1.68000	9.16000	0.00000	11.24000	0.00000	10.42000	10.49413
1	$M_{1 \rightarrow 2}$	0.000	0.280	0.000	6.040	0.000	7.140	8.056
2	Θ_1	3.92000	6.12000	0.00000	12.28000	0.00000	0.00000	10.80187
2	Θ_2	1.96000	8.76000	0.00000	10.44000	0.00000	10.82000	10.79842
2	$M_{1 \rightarrow 2}$	0.000	1.800	0.000	8.040	0.000	8.540	9.186
3	Θ_1	4.12000	7.24000	0.00000	13.48000	0.00000	0.00000	11.41192
3	Θ_2	0.48000	0.64000	0.00000	7.60000	0.00000	9.18000	9.57587
3	$M_{1 \rightarrow 2}$	0.000	1.880	0.000	9.160	0.000	8.420	9.082
4	Θ_1	10.16000	16.48000	0.00000	19.88000	0.00000	0.00000	16.03611
4	Θ_2	2.40000	11.92000	0.00000	13.60000	0.00000	11.22000	10.97549
4	$M_{1 \rightarrow 2}$	0.000	0.040	0.000	3.200	0.000	3.140	5.468
5	Θ_1	0.80000	2.84000	0.00000	7.60000	0.00000	0.00000	7.20740
5	Θ_2	1.16000	2.12000	0.00000	5.12000	0.00000	10.10000	10.14515
5	$M_{1 \rightarrow 2}$	0.000	8.480	0.000	15.200	0.000	11.180	11.143
6	Θ_1	7.20000	12.88000	0.00000	18.28000	0.00000	0.00000	13.83971
6	Θ_2	1.68000	7.04000	0.00000	9.68000	0.00000	10.42000	10.42614
6	$M_{1 \rightarrow 2}$	0.000	5.400	0.000	12.360	0.000	10.380	10.449
7	Θ_1	0.60000	2.08000	0.00000	7.32000	0.00000	0.00000	7.36614
7	Θ_2	1.84000	19.44000	0.00000	19.84000	0.00000	10.18000	10.22056
7	$M_{1 \rightarrow 2}$	0.000	0.440	0.000	4.400	0.000	4.180	5.652
8	Θ_1	4.36000	6.72000	0.00000	13.04000	0.00000	0.00000	11.29828
8	Θ_2	0.16000	0.24000	0.00000	6.56000	0.00000	8.18000	8.74440
8	$M_{1 \rightarrow 2}$	0.000	8.200	0.000	11.480	0.000	10.740	10.736
9	Θ_1	5.72000	10.08000	0.00000	16.00000	0.00000	0.00000	12.88602
9	Θ_2	2.92000	11.60000	0.00000	13.80000	0.00000	11.26000	11.14590
9	$M_{1 \rightarrow 2}$	0.000	1.520	0.000	7.880	0.000	8.380	9.121

10	Θ_1	4.76000	8.00000	0.00000	13.44000	0.00000	0.00000	11.70211
10	Θ_2	2.52000	14.16000	0.00000	17.80000	0.00000	11.38000	11.27615
10	$M_{1 \rightarrow 2}$	0.000	0.560	0.000	6.120	0.000	6.580	7.711
<hr/>								
All	Θ_1	8.28000	10.56000	12.78000	13.32000	16.00000	12.18000	12.18479
All	Θ_2	2.92000	7.16000	9.58000	12.60000	17.88000	10.50000	10.60834
All	$M_{1 \rightarrow 2}$	1.080	2.040	3.580	4.280	8.120	3.860	4.259

Citation suggestions:

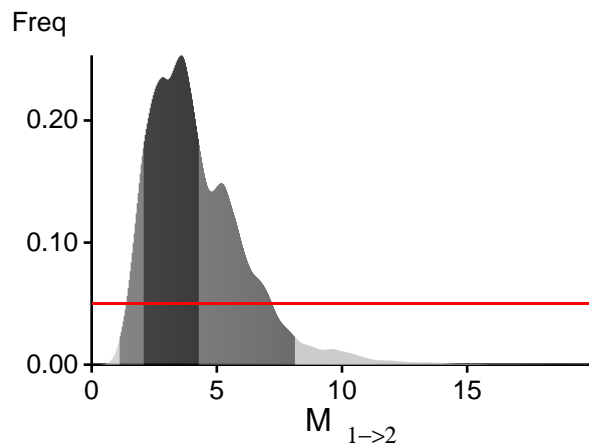
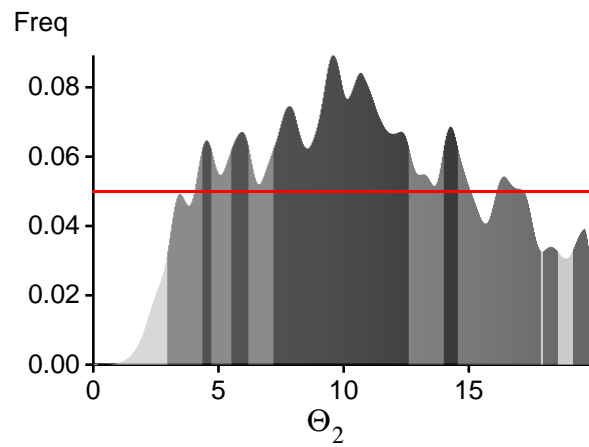
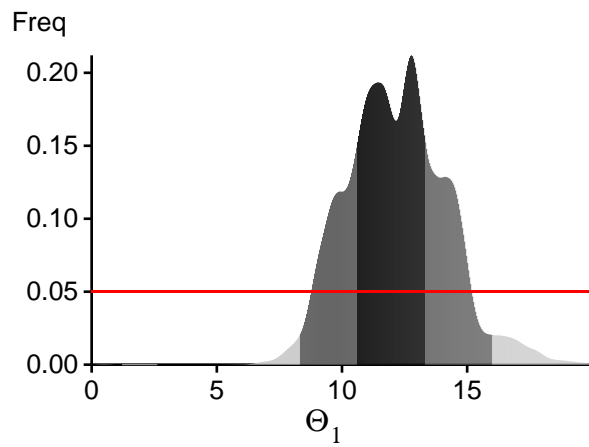
Beerli P., 2006. Comparison of Bayesian and maximum-likelihood inference of population genetic parameters.

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	-364.65	-91.09	-38.13
2	-81.48	-38.01	-30.62
3	-83.06	-39.99	-32.08
4	-574.64	-130.23	-44.05
5	-43.45	-27.54	-24.90
6	-285.64	-77.80	-36.51
7	-76.89	-36.41	-29.17
8	-205.38	-59.67	-33.97
9	-313.68	-80.15	-35.49
10	-167.09	-54.67	-32.90
All	-2193.62	-633.20	-335.47

(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 = 2.355785

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	1667191/1667191	1.00000
Θ_2	1667679/1667679	1.00000
$M_{1 \rightarrow 2}$	1666686/1666686	1.00000
Genealogies	2382663/4998444	0.47668

MCMC-Autocorrelation and Effective MCMC Sample Size

Parameter	Autocorrelation	Effective Sample Size
Θ_1	0.34993	48629.18
Θ_2	0.14565	74911.62
$M_{1 \rightarrow 2}$	0.57202	27754.57
$\text{Ln}[\text{Prob}(D G)]$	0.48364	35022.63

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