

POPULATION SIZE, MIGRATION, DIVERGENCE, ASSIGNMENT, HISTORY

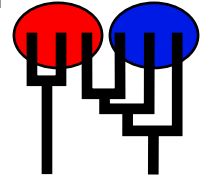
Bayesian inference using the structured coalescent

Migrate-n version 4.4.4(git:v4-series-30-gb190f74) [June-1-2019]

Compiled for a SYMMETRIC multiprocessors (Grandcentral)

Program started at Mon Nov 18 14:17:46 2019

Program finished at Mon Nov 18 14:22:56 2019 [Runtime:0000:00:05:10]



Options

Datatype:

DNA sequence data

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]

Random number seed:

(with internal timer)

225357753

Start parameters:

Theta values were generated

Using a percent value of the prior

M values were generated

Using a percent value of the prior

Connection matrix:

m = average (average over a group of Thetas or M,

s = symmetric migration M, S = symmetric 4Nm,

0 = zero, and not estimated,

* = migration free to vary, Thetas are on diagonal

d = row population split off column population, D = split and then migration

Population	1	2	3	4	5
1 Romanshorn_0	*	0	0	0	0
2 Arbon_1	*	*	0	0	0
3 Kreuzlingen_2	0	*	*	0	0
4 Frauenfeld_3	0	0	*	*	0
5 Guendelhart_4	0	0	0	*	*

Order of parameters:

1	Θ_1	<displayed>
2	Θ_2	<displayed>
3	Θ_3	<displayed>
4	Θ_4	<displayed>
5	Θ_5	<displayed>
6	$M_{1 \rightarrow 2}$	<displayed>
7	$M_{2 \rightarrow 3}$	<displayed>
8	$M_{3 \rightarrow 4}$	<displayed>
9	$M_{4 \rightarrow 5}$	<displayed>

Mutation rate among loci:

Mutation rate is constant for all loci

Analysis strategy:

Bayesian inference

-Population size estimation:

Exponential Distribution

-Geneflow estimation:

Exponential Distribution

Proposal distributions for parameter

Parameter	Proposal
Theta	Metropolis sampling
M	Metropolis sampling
Divergence	Metropolis sampling
Divergence Spread	Metropolis sampling
Genealogy	Metropolis-Hastings

Prior distribution for parameter

Parameter	Prior	Minimum	Mean	Maximum	Delta	Bins	UpdateFreq
1	Theta 00	Uniform	0.000000	0.010 0.100	0.010	1500	0.05556
2	Theta 11	Uniform	0.000000	0.010 0.100	0.010	1500	0.05556
3	Theta 22	Uniform	0.000000	0.010 0.100	0.010	1500	0.05556
4	Theta 33	Uniform	0.000000	0.010 0.100	0.010	1500	0.05556
5	Theta 44	Uniform	0.000000	0.010 0.100	0.010	1500	0.05556
6	M 30	Uniform	0.000000	100.0 1000.	100.0	1500	0.05556
7	M 12	Uniform	0.000000	100.0 1000.	100.0	1500	0.05556
8	M 32	Uniform	0.000000	100.0 1000.	100.0	1500	0.05556
9	M 34	Uniform	0.000000	100.0 1000.	100.0	1500	0.05556

[* * means priors were set globally]

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]	1
Visited (sampled) parameter values [a*b*c]	500000
Number of discard trees per chain (burn-in)	500

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: infile1
Haplotyping is turned on: NO
Output file: outfile_a
Posterior distribution raw histogram file: bayesfile
Raw data from the MCMC run: bayesallfile.gz
Print data: No
Print genealogies [only some for some data type]: None

Data summary

Data file: infile1
 Datatype: Sequence data
 Number of loci: 2

Mutationmodel:

Locus	Sublocus	Mutationmodel	Mutationmodel parameters
1	1	Felsenstein 84	[Bf:0.26 0.24 0.26 0.24, t/t ratio=2.000]
2	1	Felsenstein 84	[Bf:0.25 0.26 0.25 0.25, t/t ratio=2.000]

Sites per locus

Locus	Sites
1	1000
2	1000

Site rate variation and probabilities:

Locus	Sublocus	Region type	Rate of change	Probability	Patch size
-------	----------	-------------	----------------	-------------	------------

1	1	1	1.000	1.000	1.000
2	1	1	1.000	1.000	1.000

Population	Locus	Gene copies
1 Romanshorn_0	1	5
	2	5
2 Arbon_1	1	5
	2	5
3 Kreuzlingen_2	1	5
	2	5
4 Frauenfeld_3	1	5
	2	5
5 Guendelhart_4	1	5
	2	5
Total of all populations	1	25
	2	25

Bayesian Analysis: Posterior distribution table

Locus	Parameter	2.5%	25.0%	Mode	75.0%	97.5%	Median	Mean
1	Θ_1	0.01700	0.02307	0.02350	0.02360	0.03273	0.03643	0.04030
1	Θ_2	0.08447	0.08513	0.08523	0.08533	0.08547	0.06237	0.06205
1	Θ_3	0.04620	0.04620	0.04630	0.04633	0.04633	0.05270	0.05396
1	Θ_4	0.04487	0.04513	0.04523	0.04533	0.04640	0.05277	0.05493
1	Θ_5	0.02327	0.02353	0.02363	0.02373	0.02527	0.04643	0.04943
1	$M_{1 \rightarrow 2}$	10.7	30.0	34.3	42.7	72.7	39.0	41.6
1	$M_{2 \rightarrow 3}$	52.7	155.3	156.3	156.7	338.0	209.7	263.6
1	$M_{3 \rightarrow 4}$	9.3	33.3	35.7	44.7	128.0	57.0	69.7
1	$M_{4 \rightarrow 5}$	190.7	206.0	211.7	213.3	318.7	333.0	386.6
2	Θ_1	0.02940	0.02993	0.03017	0.03033	0.03513	0.04310	0.04591
2	Θ_2	0.02120	0.02133	0.02150	0.02153	0.02200	0.04903	0.05112
2	Θ_3	0.00967	0.01700	0.01757	0.01780	0.03147	0.02250	0.02561
2	Θ_4	0.01127	0.01127	0.01143	0.01147	0.01207	0.05203	0.05292
2	Θ_5	0.00327	0.00753	0.00797	0.00827	0.02200	0.01317	0.01641
2	$M_{1 \rightarrow 2}$	16.0	43.3	71.7	96.7	262.0	129.0	182.4
2	$M_{2 \rightarrow 3}$	9.3	43.3	53.0	53.3	126.7	58.3	69.8
2	$M_{3 \rightarrow 4}$	650.0	650.0	651.0	651.3	654.7	463.0	487.7
2	$M_{4 \rightarrow 5}$	0.0	31.3	34.3	35.3	124.7	54.3	67.6
All	Θ_1	0.01947	0.02793	0.03257	0.03887	0.04833	0.03317	0.03316
All	Θ_2	0.02073	0.03680	0.04023	0.04807	0.04993	0.03797	0.03635
All	Θ_3	0.00967	0.01473	0.02297	0.03127	0.04640	0.02530	0.02617
All	Θ_4	0.02007	0.03620	0.03877	0.04813	0.05000	0.03703	0.03595
All	Θ_5	0.00467	0.00647	0.01337	0.02067	0.03507	0.01570	0.01777
All	$M_{1 \rightarrow 2}$	6.0	33.3	46.3	61.3	117.3	53.0	70.6
All	$M_{2 \rightarrow 3}$	25.3	66.0	77.7	91.3	225.3	97.0	109.0
All	$M_{3 \rightarrow 4}$	22.0	41.3	75.7	118.0	229.3	93.0	116.3
All	$M_{4 \rightarrow 5}$	0.0	32.0	55.0	67.3	248.7	87.0	116.6

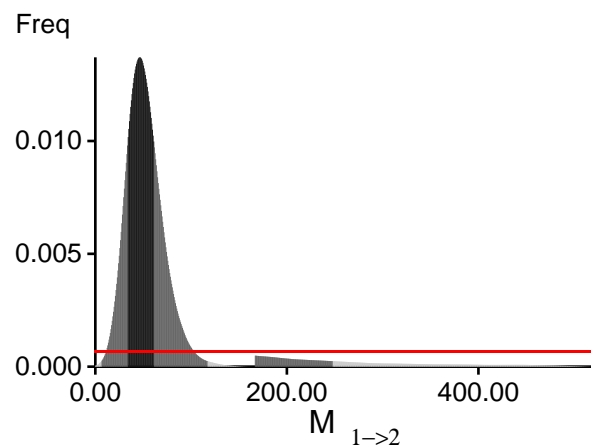
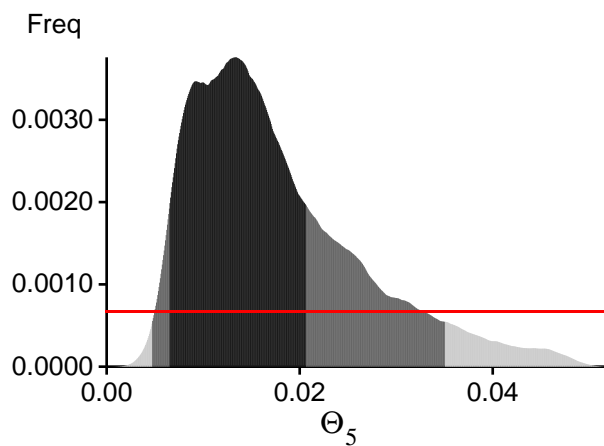
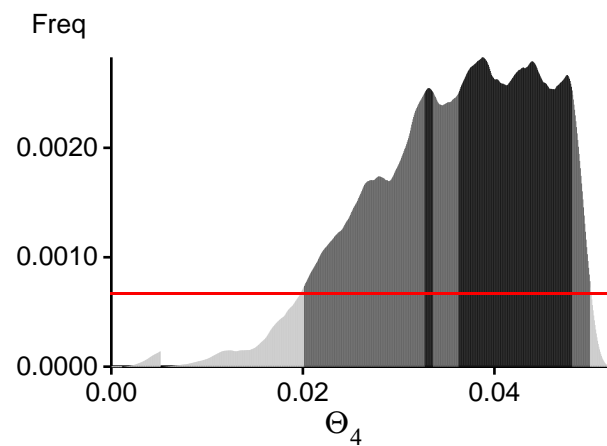
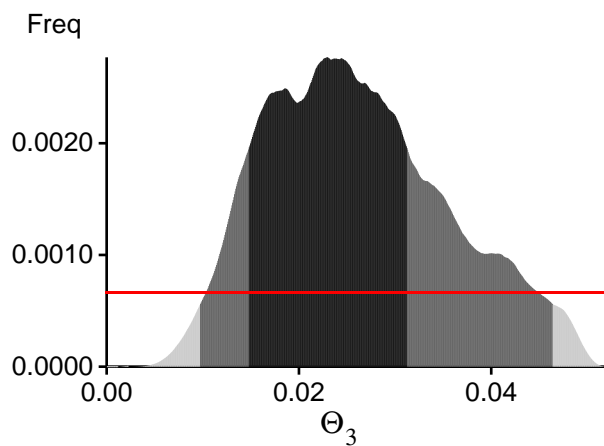
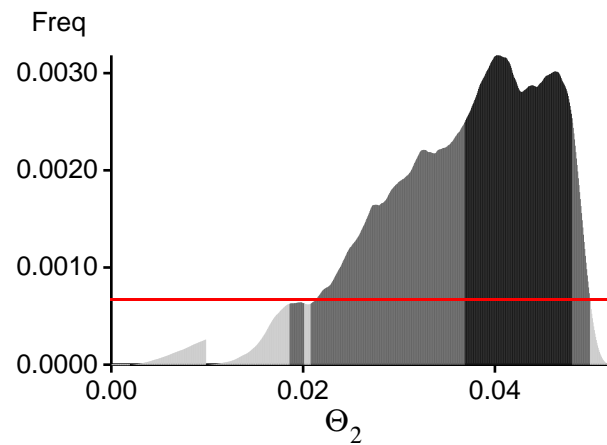
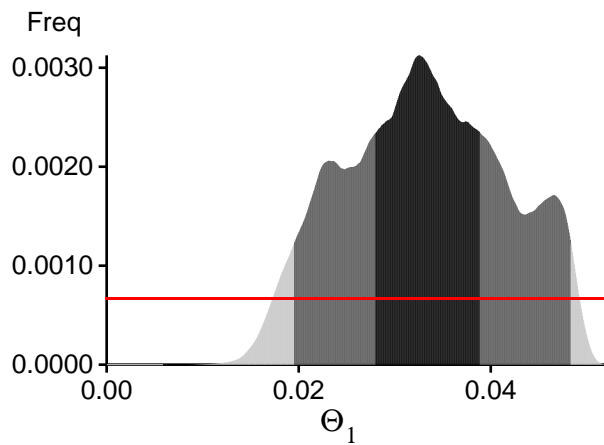
Citation suggestions:

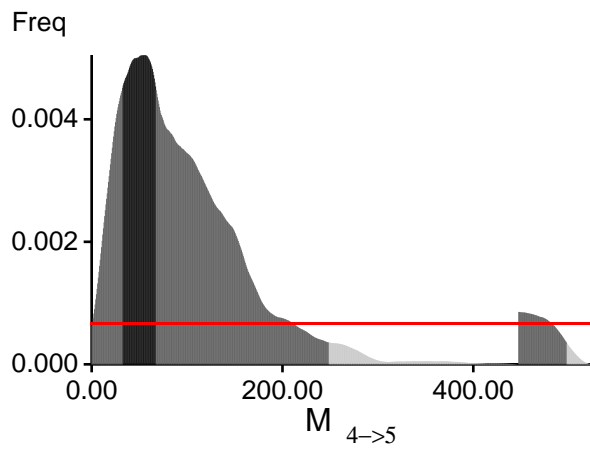
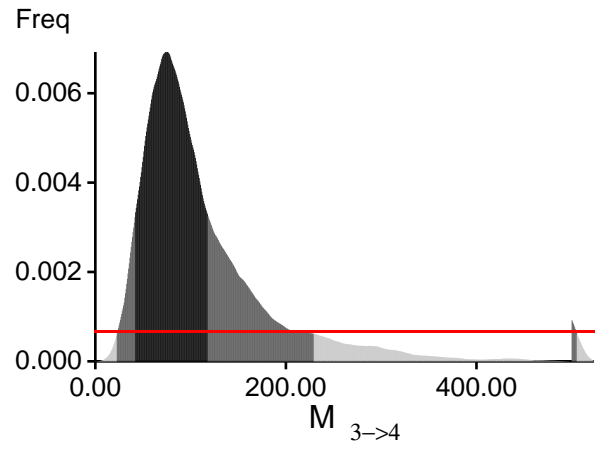
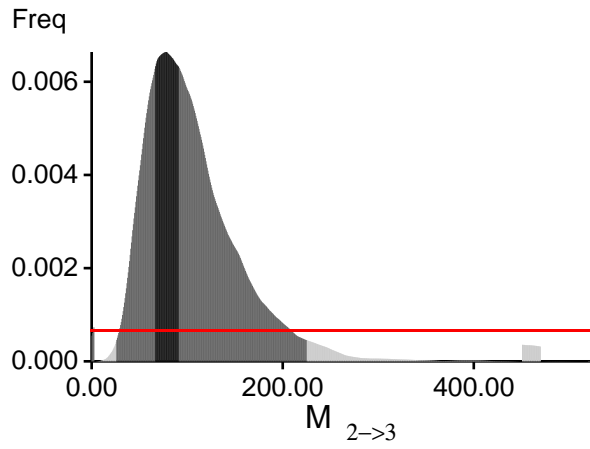
Beerli P., 2006. Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. *Bioinformatics* 22:341-345

Beerli P., 2007. Estimation of the population scaled mutation rate from microsatellite data, *Genetics*, 177:1967-1968.

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:

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

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	-14639.84	-11161.26	-10302.56
2	-15802.27	-10738.93	-9438.10
All	-30485.91	-21943.99	-19784.45

(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 = -43.792296]

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	34482/55353	0.62295
Θ_2	45684/55499	0.82315
Θ_3	37059/55792	0.66424
Θ_4	46695/55116	0.84721
Θ_5	37648/55586	0.67729
$M_{1 \rightarrow 2}$	31101/55677	0.55860
$M_{2 \rightarrow 3}$	24386/55555	0.43895
$M_{3 \rightarrow 4}$	31844/55656	0.57216
$M_{4 \rightarrow 5}$	30036/55174	0.54439
Genealogies	19592/500592	0.03914

MCMC-Autocorrelation and Effective MCMC Sample Size

Parameter	Autocorrelation	Effective Sampe Size
Θ_1	0.61473	2463.75
Θ_2	0.52168	3381.06
Θ_3	0.60739	2576.46
Θ_4	0.49647	3652.84
Θ_5	0.53569	3400.04
$M_{1 \rightarrow 2}$	0.60857	3023.69
$M_{2 \rightarrow 3}$	0.67585	2019.54
$M_{3 \rightarrow 4}$	0.65239	2106.32
$M_{4 \rightarrow 5}$	0.57301	3052.99
Genealogies	0.92439	393.05

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