

# POPULATION SIZE, MIGRATION, DIVERGENCE, ASSIGNMENT, HISTORY

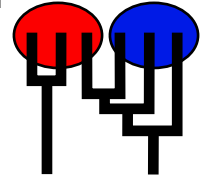
Bayesian inference using the structured coalescent

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 14:13:26 2019

Program finished at Mon Nov 18 14:17:46 2019 [Runtime:0000:00:04:20]



## 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)

1595789599

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	1	2	3	3
1 Romanshorn_0	*	*	0	c	c
1 Arbon_1	*	*	0	c	c
2 Kreuzlingen_2	*	*	*	*	*
3 Frauenfeld_3	c	c	0	*	*
3 Guendelhart_4	c	c	0	*	*

Order of parameters:

1	$\Theta_1$	<displayed>
2	$\Theta_2$	<displayed>
3	$\Theta_3$	<displayed>
4	$M_{1 \rightarrow 2}$	<displayed>
5	$M_{3 \rightarrow 2}$	<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	Update	Freq
1	Theta 00	Uniform	0.000000	0.010 0.100	0.010	1500	0.10000	
2	Theta 11	Uniform	0.000000	0.010 0.100	0.010	1500	0.10000	
3	Theta 22	Uniform	0.000000	0.010 0.100	0.010	1500	0.10000	
4	M 20	Uniform	0.000000	100.0 1000.	100.0	1500	0.10000	
5	M 02	Uniform	0.000000	100.0 1000.	100.0	1500	0.10000	

[\* \* 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

4 chains with temperatures

1000000.00 3.00 1.50 1.00

Swapping interval is 1

Print options:

Data file:

infile1

Haplotyping is turned on:	NO
Output file:	outfile_g
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
1 Arbon_1	1	5
	2	5
2 Kreuzlingen_2	1	5
	2	5
3 Frauenfeld_3	1	5
	2	5
3 Guendelhart_4	1	5
	2	5
Total of all populations	1	25
	2	25

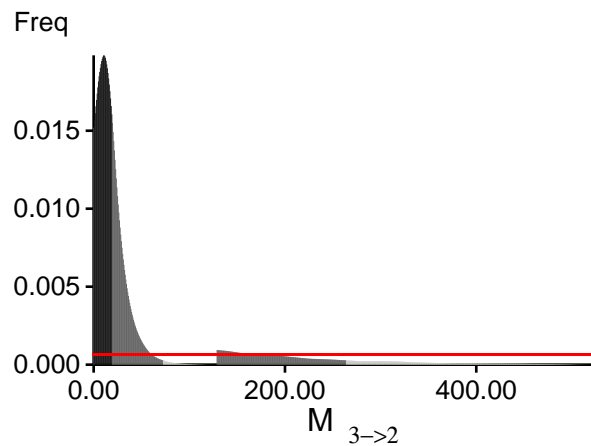
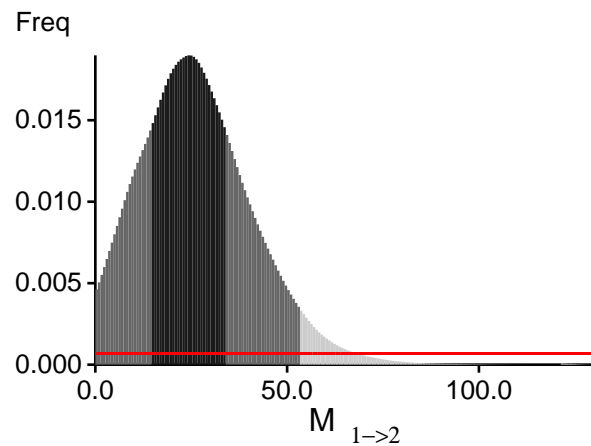
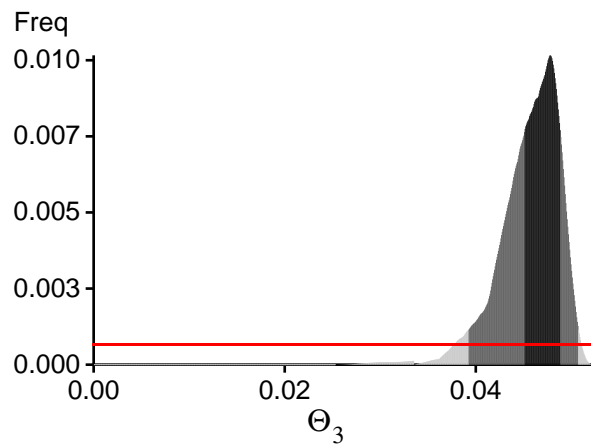
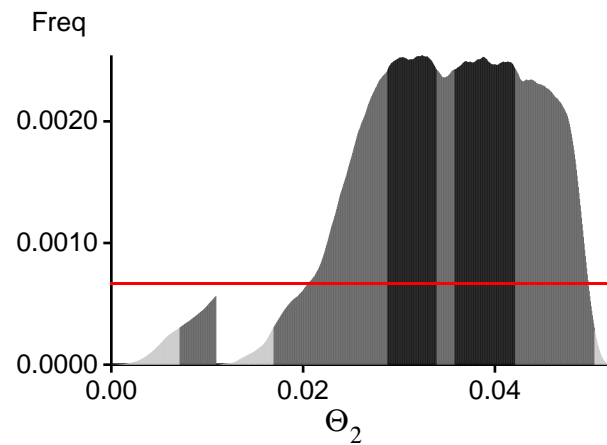
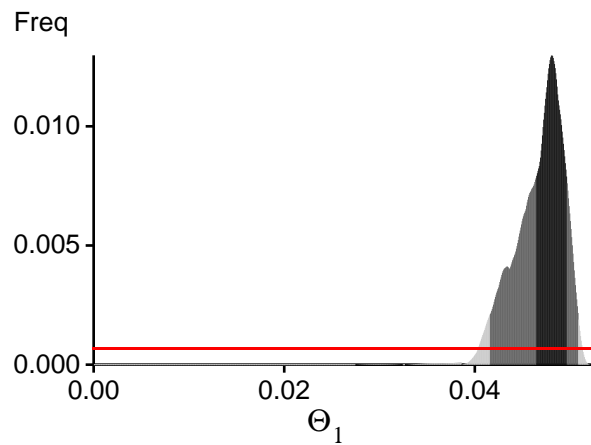
## *Bayesian Analysis: Posterior distribution table*

Locus	Parameter	2.5%	25.0%	Mode	75.0%	97.5%	Median	Mean
1	$\Theta_1$	0.08087	0.08973	0.08990	0.08993	0.09993	0.08183	0.08006
1	$\Theta_2$	0.03580	0.03873	0.03897	0.03900	0.04127	0.05563	0.05713
1	$\Theta_3$	0.07307	0.09533	0.09563	0.09567	0.09993	0.08623	0.08397
1	$M_{1 \rightarrow 2}$	0.0	0.0	4.3	26.7	34.0	11.0	13.4
1	$M_{3 \rightarrow 2}$	0.0	0.0	0.3	6.0	36.0	7.0	10.4
2	$\Theta_1$	0.07387	0.09553	0.09617	0.09660	0.09993	0.08810	0.08544
2	$\Theta_2$	0.01893	0.02753	0.02777	0.02780	0.03167	0.03770	0.04261
2	$\Theta_3$	0.04933	0.05320	0.05330	0.05340	0.05473	0.06897	0.06874
2	$M_{1 \rightarrow 2}$	33.3	45.3	55.7	58.7	75.3	57.7	59.1
2	$M_{3 \rightarrow 2}$	7.3	37.3	39.0	50.7	235.3	118.3	144.4
All	$\Theta_1$	0.04153	0.04640	0.04810	0.04967	0.05087	0.04730	0.04672
All	$\Theta_2$	0.01687	0.02873	0.03243	0.03393	0.05040	0.03517	0.03439
All	$\Theta_3$	0.03920	0.04507	0.04777	0.04887	0.05073	0.04623	0.04562
All	$M_{1 \rightarrow 2}$	0.0	14.0	24.3	34.0	53.3	26.3	26.9
All	$M_{3 \rightarrow 2}$	0.0	0.0	11.0	19.3	72.7	19.0	50.0

### 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	-14639.93	-10729.15	-9340.29
2	-16502.39	-10097.81	-8233.40
All	-31149.84	-20834.48	-17581.21

(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 = -7.518553]

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
$\Theta_1$	57551/99894	0.57612
$\Theta_2$	70765/100135	0.70670
$\Theta_3$	59192/100285	0.59024
$M_{1 \rightarrow 2}$	69772/100238	0.69606
$M_{3 \rightarrow 2}$	48201/99668	0.48362
Genealogies	16107/499780	0.03223



## *MCMC-Autocorrelation and Effective MCMC Sample Size*

Parameter	Autocorrelation	Effective Sample Size
$\Theta_1$	0.12346	7802.23
$\Theta_2$	0.07744	8617.03
$\Theta_3$	0.13926	7641.50
$M_{1 \rightarrow 2}$	0.56497	3838.77
$M_{3 \rightarrow 2}$	0.41860	4098.79
Genealogies	0.93087	358.14

## *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