

# AUTO

POPULATION SIZE, MIGRATION, DIVERGENCE, ASSIGNMENT, HISTORY

Bayesian inference using the structured coalescent

Migrate-n version 5.0.0a [May-20-2017]

Using Intel AVX (Advanced Vector Extensions)

Compiled for PARALLEL computer architectures

One master and 20 compute nodes are available.

Program started at Wed Oct 25 12:53:32 2017

Program finished at Wed Oct 25 14:38:23 2017 [Runtime:0000:01:44:51]



## 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)

3940788991

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

1 Romanshorn\_0

\* \*

1 Arbon\_1

\* \*

Order of parameters:

1

$\Theta_1$

<displayed>

Mutation rate among loci:

Mutation rate is constant for all loci

Analysis strategy:

Bayesian inference

-Population size 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 -11 Uniform	0.000000	0.050	0.100	0.010	1500	0.20000

[-1 -1 means priors were set globally]

Markov chain settings:

Long chain

Number of chains	1
Recorded steps [a]	50000
Increment (record every x step [b])	200
Number of concurrent chains (replicates) [c]	2
Visited (sampled) parameter values [a*b*c]	20000000
Number of discard trees per chain (burn-in)	10000

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:	infile.0.8
Haplotyping is turned on:	NO
Output file:	outfile_0.8_1.0
Posterior distribution raw histogram file:	bayesfile
Raw data from the MCMC run:	bayesallfile_0.8_1.0
Print data:	No
Print genealogies [only some for some data type]:	None

## Data summary

Data file: infile.0.8  
 Datatype: Sequence data  
 Number of loci: 10

### Mutationmodel:

Locus	Sublocus	Mutationmodel	Mutationmodel parameters
1	1	Jukes-Cantor	[Basefreq: =0.25]
2	1	Jukes-Cantor	[Basefreq: =0.25]
3	1	Jukes-Cantor	[Basefreq: =0.25]
4	1	Jukes-Cantor	[Basefreq: =0.25]
5	1	Jukes-Cantor	[Basefreq: =0.25]
6	1	Jukes-Cantor	[Basefreq: =0.25]
7	1	Jukes-Cantor	[Basefreq: =0.25]
8	1	Jukes-Cantor	[Basefreq: =0.25]
9	1	Jukes-Cantor	[Basefreq: =0.25]
10	1	Jukes-Cantor	[Basefreq: =0.25]

### Sites per locus

Locus	Sites
1	10000
2	10000
3	10000
4	10000
5	10000
6	10000
7	10000
8	10000
9	10000
10	10000

### 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
3	1	1	1.000	1.000	1.000
4	1	1	1.000	1.000	1.000
5	1	1	1.000	1.000	1.000
6	1	1	1.000	1.000	1.000

7	1	1	1.000	1.000	1.000	
8	1	1	1.000	1.000	1.000	
9	1	1	1.000	1.000	1.000	
10	1	1	1.000	1.000	1.000	
Population					Locus	Gene copies
1 Romanshorn_0					1	10
					2	10
					3	10
					4	10
					5	10
					6	10
					7	10
					8	10
					9	10
					10	10
1 Arbon_1					1	10
					2	10
					3	10
					4	10
					5	10
					6	10
					7	10
					8	10
					9	10
					10	10
Total of all populations					1	20
					2	20
					3	20
					4	20
					5	20
					6	20
					7	20
					8	20
					9	20
					10	20

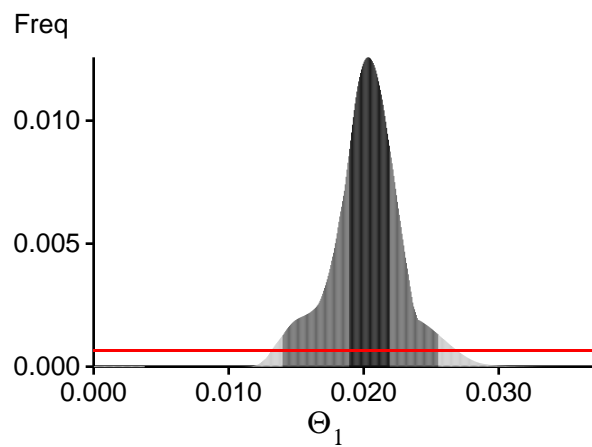
## *Bayesian Analysis: Posterior distribution table*

Locus	Parameter	2.5%	25.0%	Mode	75.0%	97.5%	Median	Mean
1	$\Theta_1$	0.02907	0.03913	0.04070	0.04313	0.05027	0.04023	0.04567
2	$\Theta_1$	0.02393	0.02887	0.03330	0.03793	0.04647	0.03463	0.03683
3	$\Theta_1$	0.01407	0.02033	0.02070	0.02100	0.03127	0.02230	0.02307
4	$\Theta_1$	0.00967	0.01467	0.01650	0.01853	0.02853	0.01770	0.01828
5	$\Theta_1$	0.00687	0.01000	0.01210	0.01447	0.02060	0.01297	0.01334
6	$\Theta_1$	0.01433	0.01873	0.02010	0.02173	0.02920	0.02177	0.02256
7	$\Theta_1$	0.01260	0.01680	0.01843	0.02020	0.02733	0.01983	0.02049
8	$\Theta_1$	0.01553	0.02100	0.02383	0.02720	0.03800	0.02570	0.02662
9	$\Theta_1$	0.01287	0.01300	0.01823	0.02600	0.02613	0.01963	0.02027
10	$\Theta_1$	0.02040	0.02953	0.03037	0.03113	0.04653	0.03217	0.03385
All	$\Theta_1$	0.01393	0.01887	0.02037	0.02193	0.02553	0.02037	0.02023

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	TI(1a)	BTI(1b)	SS(2)	HS(3)
1	-29180.27	-24806.67	-24188.61	-24257.19
2	-28847.76	-23357.16	-nan	-22574.94
3	-31009.31	-23845.30	-nan	-22746.40
4	-21122.99	-18864.26	-nan	-18652.46
5	-19139.92	-17650.31	-nan	-17571.83
6	-23102.29	-20619.07	-nan	-20384.19
7	-24783.02	-20970.57	-nan	-20483.18
8	-22211.53	-20252.95	-nan	-20117.53
9	-24019.24	-20315.17	-nan	-19841.06
10	-32885.24	-25344.19	-nan	-24190.86
All	-256303.85	-216027.93	-nan	-210821.92

(1a) TI: Thermodynamic integration:  $\log(\text{Prob}(D \mid \text{Model}))$ : Good approximation with many temperatures

(1b) BTI: Bezier-approximated Thermodynamic integration: when using few temperatures USE THIS!

(2) SS: Steppingstone Sampling (Xie et al 2011)

(3) HS: Harmonic mean approximation: Overestimates the marginal likelihood, poor variance

[Scaling factor = -2.276148]

Citation suggestions:

Beerli P. and M. Palczewski, 2010. Unified framework to evaluate panmixia and migration direction among multiple sampling locations, *Genetics*, 185: 313-326.

Palczewski M. and P. Beerli, 2014. Population model comparison using multi-locus datasets.

In M.-H. Chen, L. Kuo, and P. O. Lewis, editors, *Bayesian Phylogenetics: Methods, Algorithms, and Applications*, pages 187-200. CRC Press, 2014.

Xie W., P. O. Lewis, Y. Fan, L. Kuo, and M.-H. Chen. 2011. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Systematic Biology*, 60(2):150-160, 2011.

*Acceptance ratios for all parameters and the genealogies*

Parameter	Accepted changes	Ratio
$\Theta_1$	32296131/40005391	0.80729
Genealogies	3057505/159994609	0.01911



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

Parameter	Autocorrelation	Effective Sampe Size
$\Theta_1$	0.14260	762536.68
Genealogies	0.62655	229948.28

## *Average temperatures during the run*

Chain	Temperatures
-------	--------------

1	0.00000
2	0.00000
3	0.00000
4	0.00000

Adaptive heating often fails, if the average temperatures are very close together try to rerun using static heating! If you want to compare models using marginal likelihoods then you **MUST** use static heating

## *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