

# ombined

## POPULATION SIZE, MIGRATION, DIVERGENCE, ASSIGNMENT, HISTORY

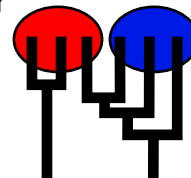
Bayesian inference using the structured coalescent

Migrate-n version 4.2.8 [June-24-2016]

Compiled for a SYMMETRIC multiprocessors (Grandcentral)

Program started at Tue Jul 26 16:44:33 2016

Program finished at Tue Jul 26 16:51:40 2016 [Runtime:0000:00:07:07]



## Options

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) 3921600642

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 urchins \*

Order of parameters:

1  $\Theta_1$  <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	0.050000	0.100000	0.010000	1500
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				4 chains with temperatures		
				100000.00	3.00	1.50 1.00
				Swapping interval is 1		
Print options:						
Data file:				infile.gap		
Haplotyping is turned on:				YES: NO report of haplotype probabilities		
Output file:				outfile-gap0		
Posterior distribution raw histogram file:				bayesfile		
Raw data from the MCMC run:				bayesallfile-gap0		
Print data:				No		
Print genealogies [only some for some data type]:				None		

## Data summary

Data file: infile.gap  
 Datatype: Haplotype data  
 Number of loci: 5

### Mutationmodel:

Locus	Sublocus	Mutationmodel	Mutationmodel parameters
1	1	Felsenstein 84	[Bf:0.25 0.25 0.32 0.18, t/t ratio=2.000]
2	1	Felsenstein 84	[Bf:0.29 0.17 0.21 0.34, t/t ratio=2.000]
3	1	Felsenstein 84	[Bf:0.30 0.21 0.19 0.30, t/t ratio=2.000]
4	1	Felsenstein 84	[Bf:0.29 0.22 0.20 0.29, t/t ratio=2.000]
5	1	Felsenstein 84	[Bf:0.31 0.18 0.19 0.31, t/t ratio=2.000]

### Sites per locus

Locus	Sites
1	252
2	921
3	425
4	459
5	713

### 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

Population	Locus	Gene copies data	(missing)
1 urchins	1	24	
	2	24	
	3	24	
	4	24	
	5	24	
Total of all populations	1	24	(0)
	2	24	(0)
	3	24	(0)
	4	24	(0)

5 24 (0)

## *Bayesian Analysis: Posterior distribution table*

Locus	Parameter	2.5%	25.0%	Mode	75.0%	97.5%	Median	Mean
1	$\Theta_1$	0.00520	0.00987	0.01383	0.01780	0.03047	0.01543	0.01685
2	$\Theta_1$	0.01653	0.02233	0.02550	0.03173	0.04280	0.02850	0.01498
3	$\Theta_1$	0.03067	0.04107	0.04763	0.04900	0.05127	0.04270	0.01870
4	$\Theta_1$	0.01087	0.01613	0.01990	0.02540	0.03680	0.02243	0.00592
5	$\Theta_1$	0.04120	0.04640	0.04810	0.04973	0.05140	0.04750	0.01612
All	$\Theta_1$	0.01047	0.01907	0.01983	0.02093	0.03047	0.02070	0.02630

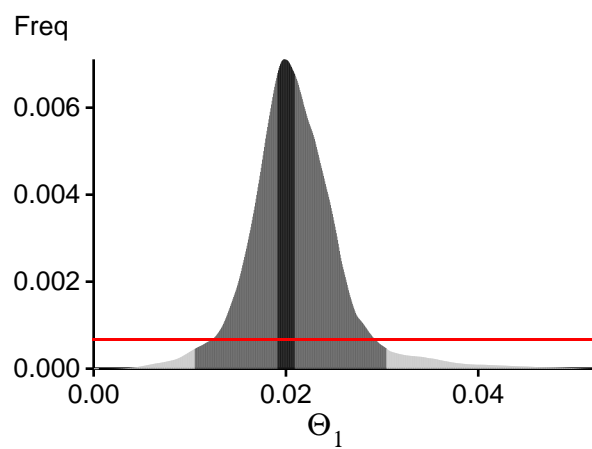
Citation suggestions:

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:

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

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	-465.39	-442.69	-427.58
2	-1684.80	-1618.02	-1608.97
3	-1037.09	-967.00	-954.86
4	-906.45	-874.43	-866.92
5	-3079.07	-2481.09	-2378.01
All	-7180.44	-6390.86	-6243.98

(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 = -7.640135]

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
$\Theta_1$	2498777/2498777	1.00000
Genealogies	346202/2501223	0.13841



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

Parameter	Autocorrelation	Effective Sampe Size
$\Theta_1$	0.24203	30942.42
Ln[Prob(D G)]	0.66082	10729.37

## 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