

# *two (fake) Swiss towns*

POPULATION SIZE, MIGRATION, DIVERGENCE, ASSIGNMENT, HISTORY

Bayesian inference using the structured coalescent

Migrate-n version 5.0(git:v5.0-33-g6ee9998-dirty) [March-20-2021]

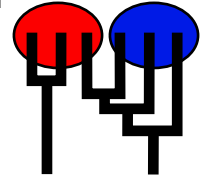
Compiled for PARALLEL computer architectures

One master and 3 compute nodes are available.

Compiled for a SYMMETRIC multiprocessors (Grandcentral)

Program started at Thu Nov 4 07:35:41 2021

Program finished at Thu Nov 4 07:39:33 2021 [Runtime:0000:00:03:52]



## *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) 3169851395

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
1 Aadorf	*	0
2 Bern	*	*

1 Aadorf growing/shrinking

2 Bern growing/shrinking

## Order of parameters:

1	$\Theta_1$	<displayed>
2	$\Theta_2$	<displayed>
3	$M_{1 \rightarrow 2}$	<displayed>
4	$g_1$	<displayed>
5	$g_2$	<displayed>

Mutation rate among loci:

Mutation rate is constant for all loci

## Analysis strategy:

Bayesian inference

-Population size estimation:

Exponential Distribution

-Population growth estimation: Growth [Exponential]

-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 **Exponent.	0.000000	0.010	0.100	-	1500	0.10000
2	Theta **Exponent.	0.000000	0.010	0.100	-	1500	0.10000
3	M ** Gamma	0.000000	100.0	10000	1000.	2000	0.10000
4	Growth ** Uniform	1000.000000	0.000	1000.	100.0	1500	0.10000
5	Growth ** Uniform	1000.000000	0.000	1000.	100.0	1500	0.10000

[\* \* means priors were set globally]

## Markov chain settings:

Long chain

Number of chains	1
Recorded steps [a]	10000
Increment (record every x step [b])	50
Number of concurrent chains (replicates) [c]	2
Visited (sampled) parameter values [a*b*c]	1000000
Number of discard trees per chain (burn-in)	5000

## Multiple Markov chains:

Adaptive\_standard heating scheme

4 chains with start values temperatures

1000000.00 3.00 1.50 1.00

Swapping interval is 1

Print options:

Data file:	twoswisstowns
	parmfile.twoswisstowns2
Haplotyping is turned on:	NO
Output file:	outfile-twoswisstowns2
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]:	Yes, only the best

## *Data summary*

Data file: twoswisstowns  
 Datatype: Haplotype data  
 Number of loci: 2

### Mutationmodel:

Locus	Sublocus	Mutationmodel	Mutationmodel parameters
1	1	Felsenstein 84	[Bf:0.25 0.27 0.22 0.26, t/t ratio=2.000]
2	1	Felsenstein 84	[Bf:0.26 0.24 0.24 0.26, 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 data	(missing)
1 Aadorf			1	10	
			2	10	
2 Bern			1	10	
			2	10	
Total of all populations			1	20	(0)
			2	20	(0)

## *Bayesian Analysis: Posterior distribution table*

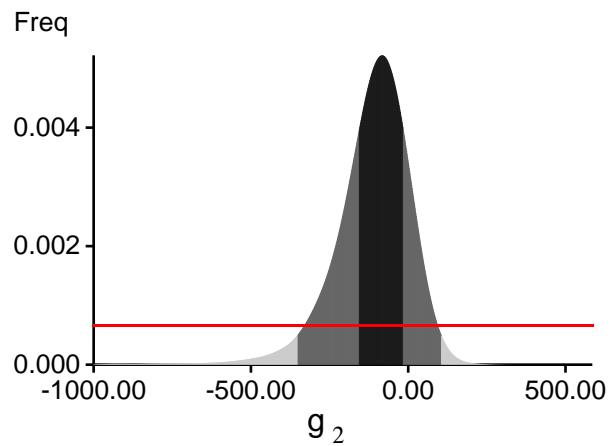
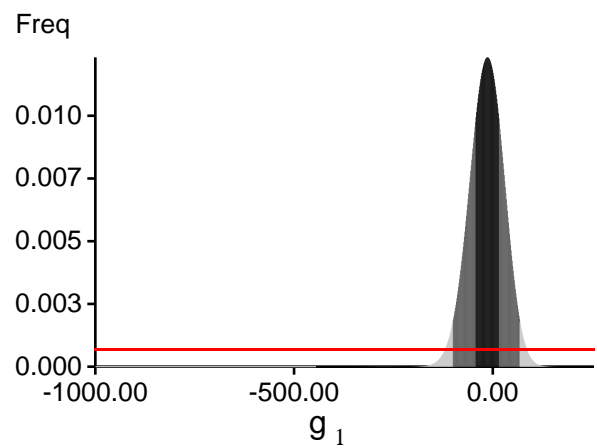
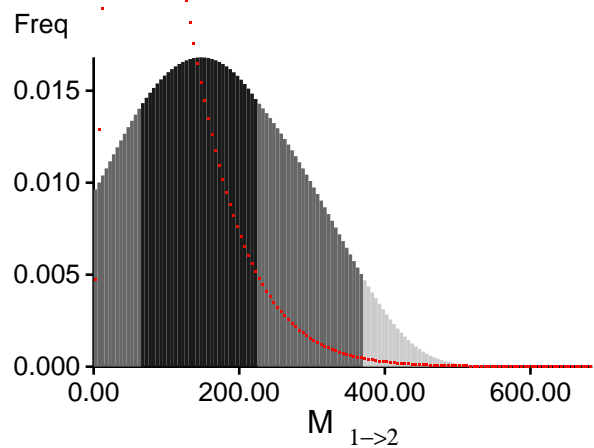
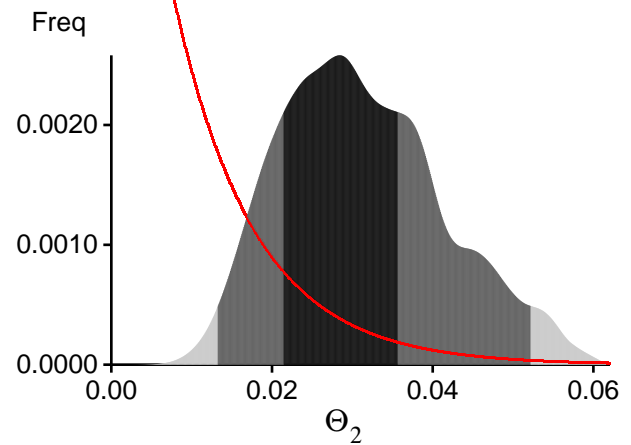
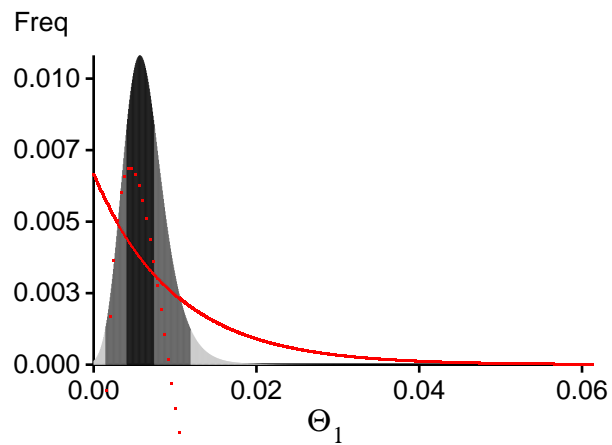
Locus	Parameter	2.5%	25.0%	Mode	75.0%	97.5%	Median	Mean
1	$\Theta_1$	0.00000	0.00213	0.00397	0.00600	0.01167	0.00483	0.00521
1	$\Theta_2$	0.00513	0.01213	0.01703	0.02467	0.04380	0.02110	0.02288
1	$M_{1 \rightarrow 2}$	0.000	40.000	117.500	190.000	345.000	152.500	71.767
1	$g_1$	-178.66667	-80.00000	-28.66667	17.33333	100.00000	-32.66667	-36.43153
1	$g_2$	-780.00000	-192.00000	-75.33333	22.66667	173.33333	-123.33333	-183.02025
2	$\Theta_1$	0.00100	0.00440	0.00670	0.00947	0.01860	0.00797	0.00881
2	$\Theta_2$	0.00740	0.01620	0.02350	0.03273	0.05747	0.02830	0.03039
2	$M_{1 \rightarrow 2}$	0.000	45.000	127.500	200.000	360.000	162.500	84.377
2	$g_1$	-109.33333	-41.33333	-7.33333	26.66667	88.00000	-7.33333	-9.03696
2	$g_2$	-637.33333	-221.33333	-94.00000	17.33333	229.33333	-132.66667	-166.74443
All	$\Theta_1$	0.00140	0.00400	0.00570	0.00747	0.01193	0.00617	0.00643
All	$\Theta_2$	0.01320	0.02140	0.02843	0.03567	0.05220	0.03030	0.03125
All	$M_{1 \rightarrow 2}$	0.000	60.000	147.500	225.000	370.000	177.500	148.075
All	$g_1$	-101.33333	-44.00000	-12.66667	16.00000	68.00000	-12.66667	-15.27542
All	$g_2$	-352.00000	-157.33333	-83.33333	-16.00000	105.33333	-98.00000	-113.54034

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	TI(1a)	BTI(1b)	HS(3)
1	-2378.30	-2267.02	-2254.61
2	-2655.21	-2492.61	-2478.29
All	-5045.39	-4771.51	-4744.79

(1a) TI: Thermodynamic integration:  $\log(\text{Prob}(D|\text{Model}))$ : 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

Adaptive heating was ON, therefore the values of (1) may be incorrect),

[Scaling factor = -11.887137]

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):150â 160, 2011.

*Acceptance ratios for all parameters and the genealogies*

Parameter	Accepted changes	Ratio
$\Theta_1$	52866/199858	0.26452
$\Theta_2$	41593/199839	0.20813
$M_{1 \rightarrow 2}$	103902/199362	0.52117
$g_1$	64913/199957	0.32463
$g_2$	92628/200285	0.46248
Genealogies	146868/1000699	0.14677



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

Parameter	Autocorrelation	Effective Sample Size
$\Theta_1$	0.60611	12138.69
$\Theta_2$	0.72003	8935.12
$M_{1 \rightarrow 2}$	0.35133	21217.06
$g_1$	-0.00000	49995.00
$g_2$	-0.00000	49995.00
Genealogies	0.81396	5402.75

## *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