

# *two (fake) Swiss towns*

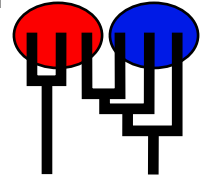
POPULATION SIZE, MIGRATION, DIVERGENCE, ASSIGNMENT, HISTORY

Bayesian inference using the structured coalescent

Migrate-n version 5.0.7 [May-01-2025]

Program started at Fri May 2 09:50:53 2025

Program finished at Fri May 2 09:52:17 2025 [Runtime:0000:00:01:24]



## *Options*

Inheritance multipliers in use for Thetas:

All loci use an inheritance multiplier of 1.0

Random number seed:

(with internal timer)

2251770202

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
Growth	Metropolis sampling
Genealogy	Metropolis-Hastings

Prior distribution for parameter

Parameter		Prior	Minimum	Mean*	Maximum	Delta	Bins	UpdateFreq
1 Theta	* *	Exponential	0.000000	0.010	0.100	-	1500	0.10000
2 Theta	* *	Exponential	0.000000	0.010	0.100	-	1500	0.10000
3 M	* *	Gamma	0.000000	100.0	10000	(a=2.000)	2000	0.10000
4 Growth	* *	Uniform	-1000.000000	0.000	1000.	100.00000	1500	0.10000
5 Growth	* *	Uniform	-1000.000000	0.000	1000.	100.00000	1500	0.10000

[\* \* means priors were set globally]

Posterior distribution:

Parameter values were collected using MCMC, these values

were then used to generate the posterior histograms using KERNEL SMOOTHING (window=41)

and subsequent SAVITZKY-GOLAY SMOOTHING (window=41) for combination over loci

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 replicate (burn-in * b)	250000

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.txt

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:	3

### Mutationmodel:

Locus	Sublocus	Mutationmodel	Mutationmodel parameters
1	1	Tamura-Nei	[Bf:0.30 0.25 0.24 0.22, k1=1.300, k2=0.800]
1	2	Felsenstein 84	[Bf:0.24 0.28 0.22 0.27, t/t ratio=2.000]
2	1	Tamura-Nei	[Bf:0.27 0.23 0.24 0.26, k1=1.300, k2=2.000]
3	1	Jukes-Cantor	[Basefreq: =0.25]

### Sites per locus

Locus	Sites
1	200      800
2	500
3	500

### Site rate variation and probabilities:

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

1	1	1	1.000	1.000	1.000
1	2	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

Population	Locus	Gene copies data	(missing)
1 Aadorf	1	10	
	2	10	
	3	10	
2 Bern	1	10	
	2	10	
	3	10	
Total of all populations	1	20	(0)
	2	20	(0)
	3	20	(0)

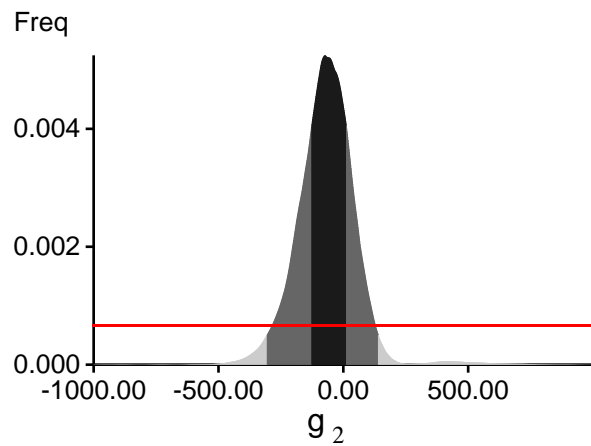
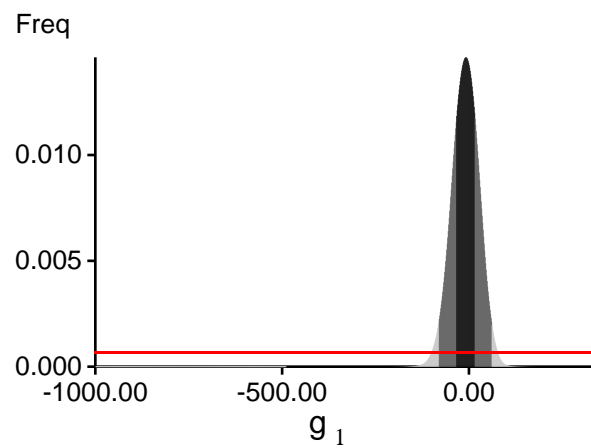
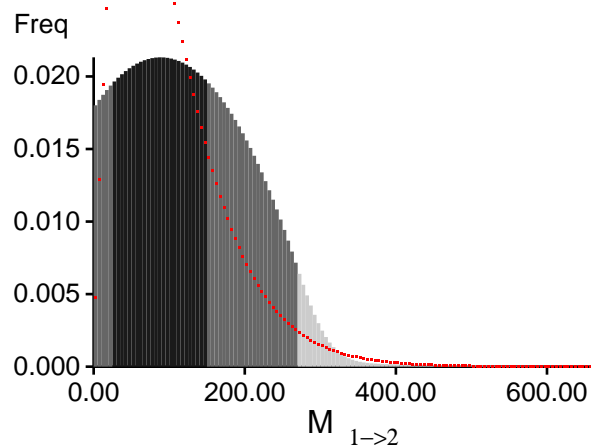
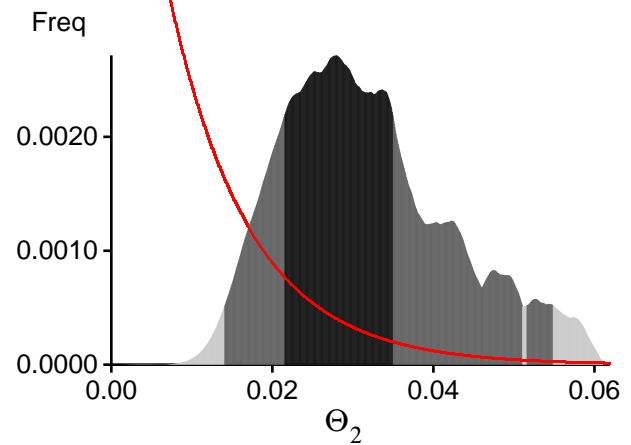
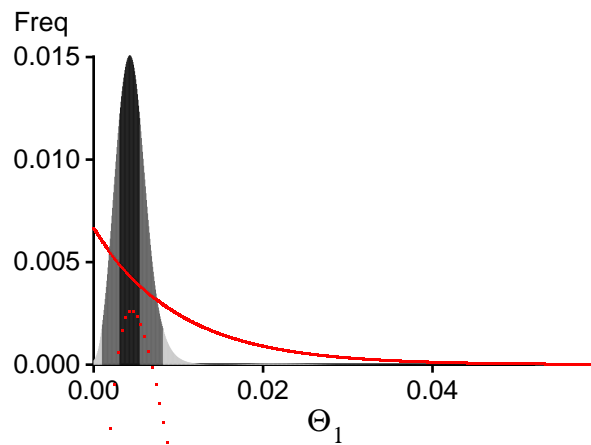
## *Bayesian Analysis: Posterior distribution table*

Locus	Parameter	2.5%	25.0%	Mode	75.0%	97.5%	Median	Mean
1	$\Theta_1$	0.00000	0.00147	0.00297	0.00440	0.00913	0.00357	0.00392
1	$\Theta_2$	0.00600	0.01147	0.01697	0.02227	0.04053	0.01950	0.02127
1	$M_{1 \rightarrow 2}$	0.000	10.000	72.500	130.000	255.000	117.500	70.827
1	$g_1$	-170.66667	-60.00000	-8.66667	40.00000	129.33333	-12.66667	-15.99144
1	$g_1$	-456.00000	-177.33333	-56.66667	25.33333	210.66667	-90.00000	-107.82904
2	$\Theta_1$	0.00020	0.00247	0.00397	0.00560	0.01013	0.00457	0.00491
2	$\Theta_2$	0.00493	0.01147	0.01637	0.02487	0.04833	0.02150	0.02387
2	$M_{1 \rightarrow 2}$	0.000	55.000	132.500	205.000	330.000	157.500	135.612
2	$g_1$	-157.33333	-44.00000	7.33333	52.00000	140.00000	0.66667	-3.22969
2	$g_1$	-718.66667	-173.33333	55.33333	236.00000	598.66667	14.00000	-8.46427
3	$\Theta_1$	0.00107	0.00367	0.00563	0.00827	0.01667	0.00703	0.00789
3	$\Theta_2$	0.00733	0.01427	0.02143	0.03000	0.05533	0.02617	0.02853
3	$M_{1 \rightarrow 2}$	0.000	15.000	82.500	140.000	265.000	122.500	81.717
3	$g_1$	-113.33333	-44.00000	-12.66667	16.00000	73.33333	-14.00000	-18.67454
3	$g_1$	-633.33333	-214.66667	-80.66667	45.33333	217.33333	-123.33333	-156.78599
All	$\Theta_1$	0.00093	0.00300	0.00430	0.00547	0.00820	0.00443	0.00462
All	$\Theta_2$	0.01400	0.02147	0.02797	0.03500	0.05107	0.03037	0.03179
All	$M_{1 \rightarrow 2}$	0.000	20.000	87.500	150.000	270.000	127.500	90.582
All	$g_1$	-81.33333	-34.66667	-7.33333	16.00000	61.33333	-7.33333	-9.51176
All	$g_1$	-306.66667	-128.00000	-74.00000	12.00000	140.00000	-64.66667	-66.72926

### Citation suggestions:

- Beerli P., 2006. Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. *Bioinformatics* 22:341-345
- Beerli, P., H. Ashki, S. Mashayekhi, and M. Palczewski, 2022. Population divergence time estimation using individual lineage label switching. *G3 Genes & Genomes & Genetics*, 12(4), 02 2022.
- 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.
- Beerli, P., S. Mashayekhi, M. Sadeghi, M. Khodaei, and K. Shaw, 2019. Population genetic inference with migrate. *Current Protocols in Bioinformatics*, 68(1):e87.



*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	-2372.26	-2266.66	-2248.05
2	-1299.38	-1228.60	-1199.00
3	-1409.86	-1314.02	-1280.90
All	-5088.44	-4816.22	-4734.88

(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!

(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 = -6.942602]

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.



## *Acceptance ratios for all parameters and the genealogies*

Parameter	Accepted changes	Ratio
$\Theta_1$	73002/300087	0.24327
$\Theta_2$	71039/299925	0.23686
$M_{1 \rightarrow 2}$	162079/300154	0.53999
$g_1$	153916/300284	0.51257
$g_2$	137497/299995	0.45833
Genealogies	261458/1499555	0.17436

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

Parameter	Autocorrelation	Effective Sample Size
$\Theta_1$	0.62313	14044.21
$\Theta_2$	0.63774	13375.29
$M_{1 \rightarrow 2}$	0.28395	33561.37
$g_1$	-0.00000	59994.00
$g_2$	-0.00000	59994.00
Genealogies	0.81838	5996.14

## *Average temperatures during the run*

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

1	1.00000
2	1.30975
3	2.98410
4	20004.58299

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 common that some parameters for some loci will not be informative. These parameters then trigger suggestions to increase the prior range that are not sensible. Do not blindly follow the suggestions given. If some parameters are flagged, inspect the tables carefully and judge whether an action is required. Suppose you run a Bayesian inference with sequence data for macroscopic species. In that case, there is rarely the need to increase the prior for Theta beyond 0.1. If you use microsatellites data, 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 prior range will not help in such situations, but reducing the number of parameters may help.

No warning was recorded during the run