

# two (fake) Swiss towns

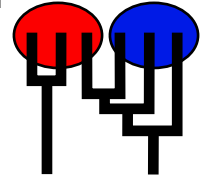
POPULATION SIZE, MIGRATION, DIVERGENCE, ASSIGNMENT, HISTORY

Bayesian inference using the structured coalescent

Migrate-n version 5.0.2(git:v5.0.2-0-gb89fa60-dirty) [December-1-2021]

Program started at Sat Feb 5 16:19:13 2022

Program finished at Sat Feb 5 16:23:41 2022 [Runtime:0000:00:04:28]



## Options

Inheritance multipliers in use for Thetas:

All loci use an inheritance multiplier of 1.0

Random number seed: (with internal timer) 26507500

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	*	*

Order of parameters:

1	$\Theta_1$	<displayed>
2	$\Theta_2$	<displayed>
3	$M_{1 \rightarrow 2}$	<displayed>

Mutation rate among loci: Mutation rate is constant for all loci

## Analysis strategy:

- Population size estimation:
- Geneflow estimation:

Bayesian inference  
Exponential Distribution  
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	0.13889
2	Theta **Exponent.	0.000000	0.010	0.100	-	1500	0.13889
3	M ** Gamma	0.000000	100.0	10000	1000.	2000	0.13889

[\* \* 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]	1
Visited (sampled) parameter values [a*b*c]	500000
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.twoswisstowns  
Haplotyping is turned on: NO  
Output file: outfile-twoswisstowns  
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: 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*

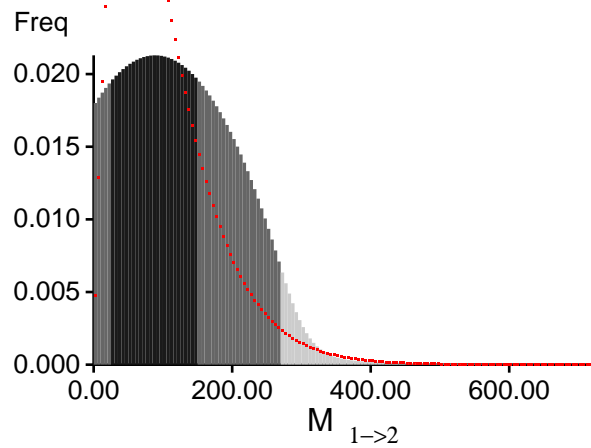
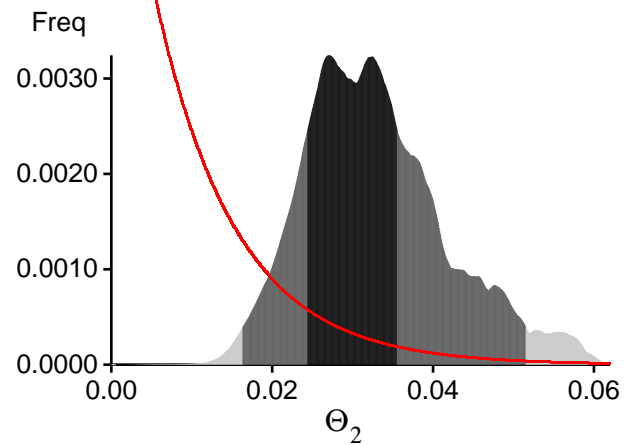
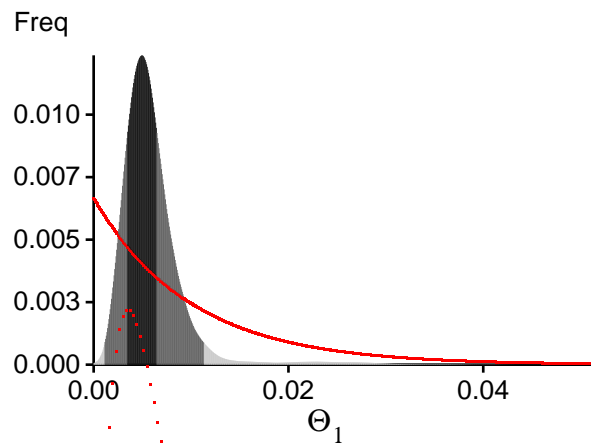
Locus	Parameter	2.5%	25.0%	Mode	75.0%	97.5%	Median	Mean
1	$\Theta_1$	0.00000	0.00200	0.00377	0.00620	0.01380	0.00517	0.00597
1	$\Theta_2$	0.00727	0.01440	0.01810	0.02573	0.04173	0.02210	0.02336
1	$M_{1 \rightarrow 2}$	0.000	10.000	72.500	130.000	255.000	117.500	70.640
2	$\Theta_1$	0.00033	0.00247	0.00410	0.00600	0.01173	0.00503	0.00549
2	$\Theta_2$	0.00627	0.01313	0.01723	0.02480	0.04300	0.02130	0.02303
2	$M_{1 \rightarrow 2}$	0.000	55.000	132.500	205.000	330.000	157.500	134.912
3	$\Theta_1$	0.00113	0.00373	0.00603	0.00960	0.02167	0.00837	0.00971
3	$\Theta_2$	0.01047	0.02213	0.02883	0.03567	0.06307	0.03297	0.03514
3	$M_{1 \rightarrow 2}$	0.000	15.000	82.500	140.000	270.000	122.500	82.742
All	$\Theta_1$	0.00107	0.00340	0.00497	0.00647	0.01133	0.00550	0.00605
All	$\Theta_2$	0.01627	0.02433	0.02710	0.03553	0.05153	0.03197	0.03292
All	$M_{1 \rightarrow 2}$	0.000	20.000	87.500	150.000	270.000	127.500	90.245

### 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	-2392.52	-2268.84	-2247.32
2	-1301.83	-1228.53	-1202.94
3	-1415.86	-1317.33	-1279.10
All	-5124.28	-4828.77	-4743.44

(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

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

[Scaling factor = -14.076171]

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
$\Theta_1$	53448/208504	0.25634
$\Theta_2$	42306/208767	0.20265
$M_{1 \rightarrow 2}$	112727/208368	0.54100
Genealogies	104992/874361	0.12008

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

Parameter	Autocorrelation	Effective Sample Size
$\Theta_1$	0.69207	5475.57
$\Theta_2$	0.52702	9652.55
$M_{1 \rightarrow 2}$	0.29279	16616.51
Genealogies	0.69207	5475.57



## *Average temperatures during the run*

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

1	1.00000
2	1.31151
3	3.34763
4	20005.58782

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

## *Summary Assignment of Individuals to Populations*

Individual	Population	
	1	2
?BAH0	0.996	0.004
?BAF0	0.937	0.063
?BAG1	0.004	0.996
?BAJ1	0.009	0.991
?BAH1	0.000	1.000
?BAI1	0.000	1.000
?BAF1	0.000	1.000

## *Detailed Assignment of Individuals to Populations*

Individual	Locus	Population	
		1	2
?BAH0	1	0.940	0.060
?BAH0	2	0.805	0.195
?BAH0	3	0.795	0.205
?BAH0	All	0.996	0.004
?BAF0	1	0.690	0.310
?BAF0	2	0.698	0.302
?BAF0	3	0.742	0.258
?BAF0	All	0.937	0.063
?BAG1	1	0.114	0.886
?BAG1	2	0.190	0.810
?BAG1	3	0.122	0.878
?BAG1	All	0.004	0.996
?BAJ1	1	0.278	0.722
?BAJ1	2	0.125	0.875
?BAJ1	3	0.138	0.862
?BAJ1	All	0.009	0.991
?BAH1	1	0.021	0.979
?BAH1	2	0.007	0.993
?BAH1	3	0.310	0.690
?BAH1	All	0.000	1.000
?BAI1	1	0.011	0.989
?BAI1	2	0.009	0.991
?BAI1	3	0.312	0.688
?BAI1	All	0.000	1.000
?BAF1	1	0.036	0.964
?BAF1	2	0.010	0.990
?BAF1	3	0.140	0.860
?BAF1	All	0.000	1.000