

Two fake Swiss 'towns'

POPULATION SIZE, MIGRATION, DIVERGENCE, ASSIGNMENT, HISTORY

Bayesian inference using the structured coalescent

Migrate-n version 4.2.7 [April-1-2016]

Using Intel AVX (Advanced Vector Extensions)

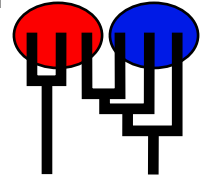
Compiled for PARALLEL computer architectures

One master and 4 compute nodes are available.

Compiled for a SYMMETRIC multiprocessors (Grandcentral)

Program started at Wed Jul 6 14:14:45 2016

Program finished at Wed Jul 6 14:56:49 2016



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) 1700814969

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 Ascona	*	*
2 Brissago	*	*

Order of parameters:

1 Θ_1 <displayed>

2 Θ_2 <displayed>
 3 $M_{2 \rightarrow 1}$ <displayed>
 4 $M_{1 \rightarrow 2}$ <displayed>

Mutation rate among loci:

Mutation rate is constant for all loci

Analysis strategy:

Bayesian inference

Proposal distributions for parameter

Parameter	Proposal
Theta	Metropolis sampling
M	Metropolis sampling

Prior distribution for parameter

Parameter	Prior	Minimum	Mean*	Maximum	Delta	Bins
Theta	Gamma	0.000000	0.010000	0.100000	0.010000	1500
Theta	Gamma	0.000000	0.010000	0.100000	0.010000	1500
M	Gamma	0.000000	500.000000	5000.000000	500.000000	1500
M	Gamma	0.000000	500.000000	5000.000000	500.000000	1500

Markov chain settings:

Long chain

Number of chains	1
Recorded steps [a]	5000
Increment (record every x step [b])	200
Number of concurrent chains (replicates) [c]	10
Visited (sampled) parameter values [a*b*c]	10000000
Number of discard trees per chain (burn-in)	5000

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
Haplotyping is turned on:	NO
Output file:	outfile_xxxx
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]:	None

Data summary

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

Mutationmodel:

Locus	Sublocus	Mutationmodel	Mutationmodel parameters
1	1	Felsenstein 84	[Bf:0.24 0.26 0.27 0.22, t/t ratio=2.000]
2	1	Felsenstein 84	[Bf:0.25 0.24 0.26 0.25, t/t ratio=2.000]
3	1	Felsenstein 84	[Bf:0.25 0.24 0.25 0.26, t/t ratio=2.000]
4	1	Felsenstein 84	[Bf:0.26 0.24 0.23 0.27, t/t ratio=2.000]
5	1	Felsenstein 84	[Bf:0.25 0.24 0.27 0.24, t/t ratio=2.000]

Sites per locus

Locus	Sites
1	1000
2	1000
3	1000
4	1000
5	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
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 Ascona	1	10
	2	10
	3	10
	4	10
	5	10
2 Brissago	1	10
	2	10
	3	10
	4	10

	5	10	
Total of all populations	1	20	(0)
	2	20	(0)
	3	20	(0)
	4	20	(0)
	5	20	(0)

Bayesian Analysis: Posterior distribution table

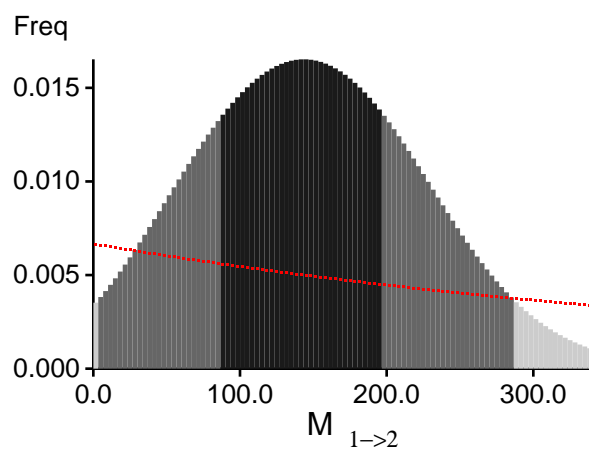
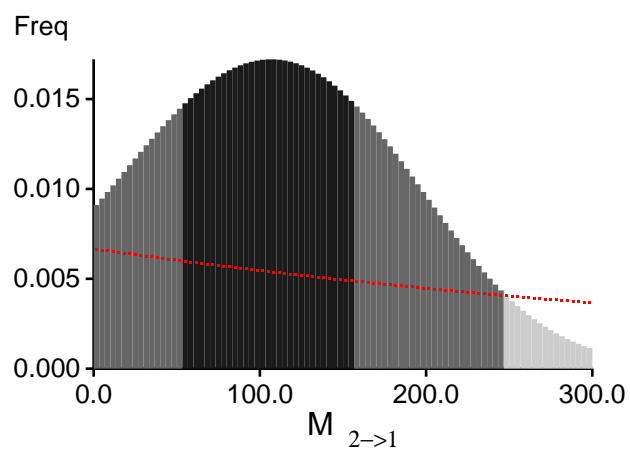
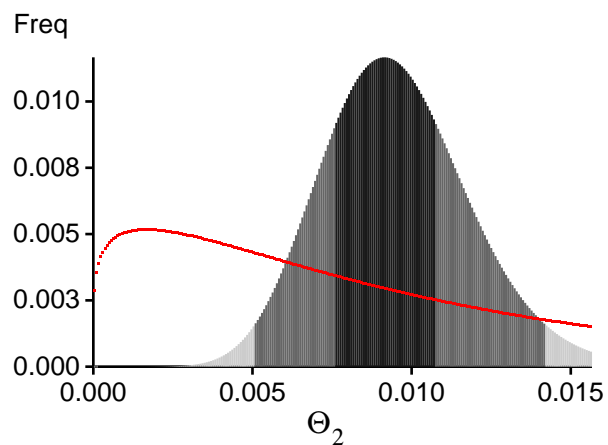
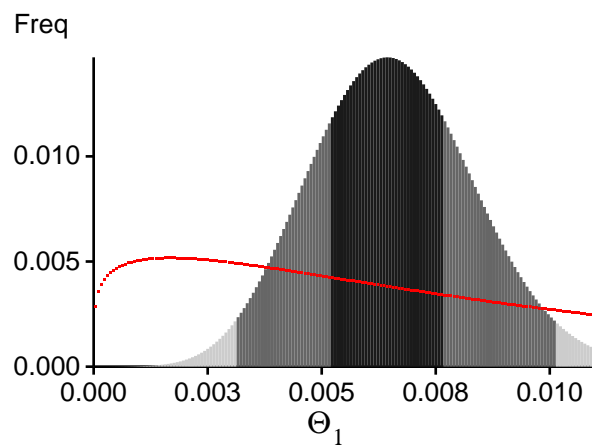
Locus	Parameter	2.5%	25.0%	Mode	75.0%	97.5%	Median	Mean
1	Θ_1	0.00280	0.00573	0.00677	0.00773	0.01193	0.00757	0.00801
1	Θ_2	0.00140	0.00300	0.00483	0.00673	0.00893	0.00537	0.00566
1	$M_{2 \rightarrow 1}$	0.000	0.000	1.667	100.000	336.667	101.667	104.941
1	$M_{1 \rightarrow 2}$	0.000	0.000	1.667	116.667	376.667	118.333	124.065
2	Θ_1	0.00193	0.00213	0.00563	0.01020	0.01047	0.00637	0.00675
2	Θ_2	0.00207	0.00207	0.00590	0.01093	0.01093	0.00670	0.00712
2	$M_{2 \rightarrow 1}$	0.000	0.000	41.667	136.667	410.000	138.333	142.674
2	$M_{1 \rightarrow 2}$	0.000	0.000	85.000	156.667	463.333	158.333	168.085
3	Θ_1	0.00053	0.00353	0.00543	0.00747	0.01260	0.00603	0.00635
3	Θ_2	0.00387	0.00833	0.01157	0.01540	0.02653	0.01323	0.01419
3	$M_{2 \rightarrow 1}$	13.333	13.333	85.000	136.667	136.667	171.667	196.390
3	$M_{1 \rightarrow 2}$	0.000	86.667	185.000	286.667	573.333	228.333	243.803
4	Θ_1	0.00000	0.00247	0.00443	0.00660	0.01260	0.00530	0.00577
4	Θ_2	0.00600	0.01313	0.01363	0.01407	0.02633	0.01530	0.01616
4	$M_{2 \rightarrow 1}$	0.000	0.000	78.333	136.667	400.000	138.333	139.034
4	$M_{1 \rightarrow 2}$	0.000	0.000	1.667	96.667	286.667	98.333	87.347
5	Θ_1	0.00267	0.00607	0.00743	0.00893	0.01440	0.00837	0.00887
5	Θ_2	0.00133	0.00500	0.00743	0.01033	0.01867	0.00863	0.00929
5	$M_{2 \rightarrow 1}$	0.000	70.000	178.333	303.333	673.333	248.333	273.149
5	$M_{1 \rightarrow 2}$	100.000	100.000	321.667	650.000	650.000	391.667	422.679
All	Θ_1	0.00307	0.00513	0.00643	0.00767	0.01013	0.00657	0.00658
All	Θ_2	0.00500	0.00753	0.00917	0.01073	0.01420	0.00943	0.00953
All	$M_{2 \rightarrow 1}$	0.000	50.000	108.333	156.667	246.667	118.333	113.168
All	$M_{1 \rightarrow 2}$	0.000	83.333	145.000	196.667	286.667	151.667	147.491

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	-1996.88	-1817.61	-1802.05
2	-1958.61	-1807.26	-1795.08
3	-2103.29	-1913.48	-1897.05
4	-2657.57	-2208.55	-2143.87
5	-2137.53	-1911.23	-1888.75
All	-10864.72	-9668.98	-9537.64

(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 = -10.842324]

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
Θ_1	2201781/6249439	0.35232
Θ_2	2332615/6253867	0.37299
$M_{2 \rightarrow 1}$	1815527/6251274	0.29043
$M_{1 \rightarrow 2}$	1828565/6246817	0.29272
Genealogies	3537069/24998603	0.14149

MCMC-Autocorrelation and Effective MCMC Sample Size

Parameter	Autocorrelation	Effective Sample Size
Θ_1	0.24193	222038.58
Θ_2	0.16625	250306.26
$M_{2 \rightarrow 1}$	0.44770	133249.94
$M_{1 \rightarrow 2}$	0.42095	148379.56
$\text{Ln}[\text{Prob}(D G)]$	0.34319	169692.09

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