

# Current protocols example dataset

POPULATION SIZE, MIGRATION, DIVERGENCE, ASSIGNMENT, HISTORY

Bayesian inference using the structured coalescent

Migrate-n version 4.4.3(git:) [March-21-2019]

Using Intel AVX (Advanced Vector Extensions)

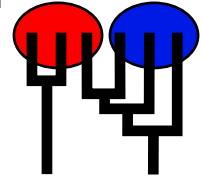
Compiled for PARALLEL computer architectures

One master and 8 compute nodes are available.

Compiled for a SYMMETRIC multiprocessors (Grandcentral)

Program started at Mon Jun 3 09:32:25 2019

Program finished at Mon Jun 3 09:46:13 2019 [Runtime:0000:00:13:48]



## 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) 3867647099

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	3
1 Arbon_1	*	0	0
2 Berg_2	*	*	0
3 Chur_3	0	*	*

Order of parameters:

1	$\Theta_1$	<displayed>
2	$\Theta_2$	<displayed>
3	$\Theta_3$	<displayed>
4	$M_{1 \rightarrow 2}$	<displayed>
5	$M_{2 \rightarrow 3}$	<displayed>

Mutation rate among loci:

Mutation rate is constant for all loci

Analysis strategy:

Bayesian inference

-Population size estimation:

Exponential Distribution

-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	** Uniform	0.000000	0.050	0.100	0.010	1500	0.10000
2	Theta	** Uniform	0.000000	0.050	0.100	0.010	1500	0.10000
3	Theta	** Uniform	0.000000	0.050	0.100	0.010	1500	0.10000
4	M	** Uniform	0.000000	2500.	5000.	500.0	1500	0.10000
5	M	** Uniform	0.000000	2500.	5000.	500.0	1500	0.10000

[\* \* means priors were set globally]

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]	1
Visited (sampled) parameter values [a*b*c]	500000
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_model1
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: 10

### Mutationmodel:

Locus	Sublocus	Mutationmodel	Mutationmodel parameters
1	1	Jukes-Cantor	[Basefreq: =0.25]
2	1	Jukes-Cantor	[Basefreq: =0.25]
3	1	Jukes-Cantor	[Basefreq: =0.25]
4	1	Jukes-Cantor	[Basefreq: =0.25]
5	1	Jukes-Cantor	[Basefreq: =0.25]
6	1	Jukes-Cantor	[Basefreq: =0.25]
7	1	Jukes-Cantor	[Basefreq: =0.25]
8	1	Jukes-Cantor	[Basefreq: =0.25]
9	1	Jukes-Cantor	[Basefreq: =0.25]
10	1	Jukes-Cantor	[Basefreq: =0.25]

### Sites per locus

Locus	Sites
1	1000
2	1000
3	1000
4	1000
5	1000
6	1000
7	1000
8	1000
9	1000
10	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
6	1	1	1.000	1.000	1.000

7	1	1	1.000	1.000	1.000		
8	1	1	1.000	1.000	1.000		
9	1	1	1.000	1.000	1.000		
10	1	1	1.000	1.000	1.000		
Population				Locus		Gene copies data	(missing)
1 Arbon_1				1		10	
				2		10	
				3		10	
				4		10	
				5		10	
				6		10	
				7		10	
				8		10	
				9		10	
				10		10	
2 Berg_2				1		10	
				2		10	
				3		10	
				4		10	
				5		10	
				6		10	
				7		10	
				8		10	
				9		10	
				10		10	
3 Chur_3				1		10	
				2		10	
				3		10	
				4		10	
				5		10	
				6		10	
				7		10	
				8		10	
				9		10	
				10		10	
Total of all populations				1		30	(0)
				2		30	(0)
				3		30	(0)
				4		30	(0)
				5		30	(0)
				6		30	(0)
				7		30	(0)
				8		30	(0)
				9		30	(0)
				10		30	(0)

10

30

(0)

## *Bayesian Analysis: Posterior distribution table*

Locus	Parameter	2.5%	25.0%	Mode	75.0%	97.5%	Median	Mean
1	$\Theta_1$	0.00307	0.00553	0.00710	0.00813	0.01813	0.00983	0.01094
1	$\Theta_2$	0.00073	0.00180	0.00263	0.00400	0.00933	0.00403	0.00489
1	$\Theta_3$	0.00373	0.00773	0.00910	0.00940	0.01947	0.01157	0.01328
1	$M_{1 \rightarrow 2}$	26.667	110.000	151.667	153.333	596.667	291.667	347.676
1	$M_{2 \rightarrow 3}$	0.000	6.667	55.000	106.667	330.000	98.333	132.793
2	$\Theta_1$	0.00167	0.00333	0.00417	0.00553	0.01447	0.00637	0.00896
2	$\Theta_2$	0.00067	0.00173	0.00303	0.00367	0.00853	0.00397	0.00468
2	$\Theta_3$	0.00093	0.00220	0.00270	0.00433	0.00947	0.00437	0.00498
2	$M_{1 \rightarrow 2}$	10.000	73.333	118.333	210.000	556.667	221.667	389.987
2	$M_{2 \rightarrow 3}$	0.000	113.333	148.333	170.000	556.667	191.667	272.450
3	$\Theta_1$	0.00220	0.00307	0.00530	0.00667	0.01253	0.00870	0.01021
3	$\Theta_2$	0.00087	0.00253	0.00437	0.00973	0.02847	0.01317	0.02248
3	$\Theta_3$	0.00447	0.00767	0.00790	0.01000	0.01940	0.01077	0.01232
3	$M_{1 \rightarrow 2}$	76.667	133.333	268.333	456.667	1263.333	495.000	900.946
3	$M_{2 \rightarrow 3}$	0.000	23.333	111.667	150.000	413.333	135.000	207.255
4	$\Theta_1$	0.00307	0.01060	0.01157	0.01200	0.02327	0.01357	0.01410
4	$\Theta_2$	0.00053	0.00173	0.00217	0.00387	0.01047	0.00417	0.00597
4	$\Theta_3$	0.00140	0.00607	0.00623	0.00660	0.01193	0.00650	0.00760
4	$M_{1 \rightarrow 2}$	43.333	186.667	191.667	243.333	1170.000	568.333	703.421
4	$M_{2 \rightarrow 3}$	0.000	100.000	158.333	223.333	726.667	278.333	439.139
5	$\Theta_1$	0.00113	0.00200	0.00277	0.00473	0.01473	0.00657	0.00762
5	$\Theta_2$	0.00073	0.00193	0.00297	0.00393	0.01527	0.00577	0.01003
5	$\Theta_3$	0.00587	0.01120	0.01150	0.01153	0.02453	0.01717	0.01926
5	$M_{1 \rightarrow 2}$	140.000	490.000	495.000	506.667	1063.333	651.667	829.156
5	$M_{2 \rightarrow 3}$	0.000	10.000	18.333	43.333	223.333	58.333	83.635
6	$\Theta_1$	0.00107	0.00200	0.00290	0.00347	0.00653	0.00337	0.00374
6	$\Theta_2$	0.00087	0.00473	0.00503	0.00573	0.01373	0.00730	0.01495
6	$\Theta_3$	0.00280	0.00973	0.01103	0.01113	0.02340	0.01283	0.01555
6	$M_{1 \rightarrow 2}$	170.000	313.333	335.000	463.333	1513.333	1145.000	1480.928
6	$M_{2 \rightarrow 3}$	0.000	26.667	71.667	150.000	706.667	181.667	331.056

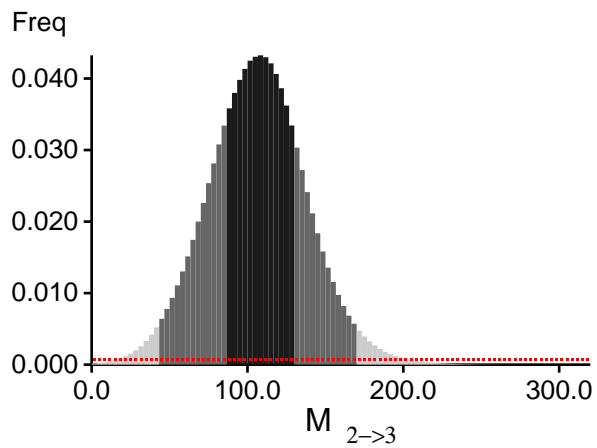
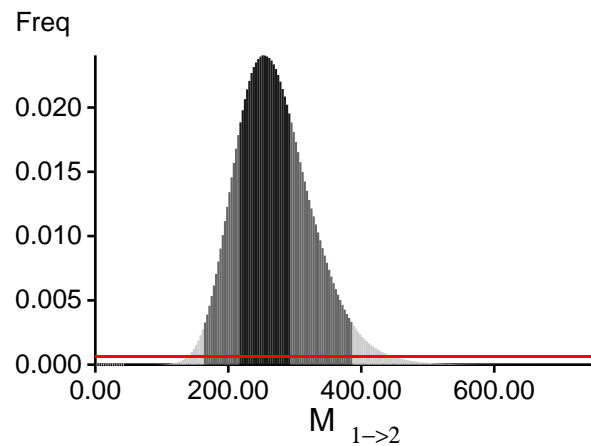
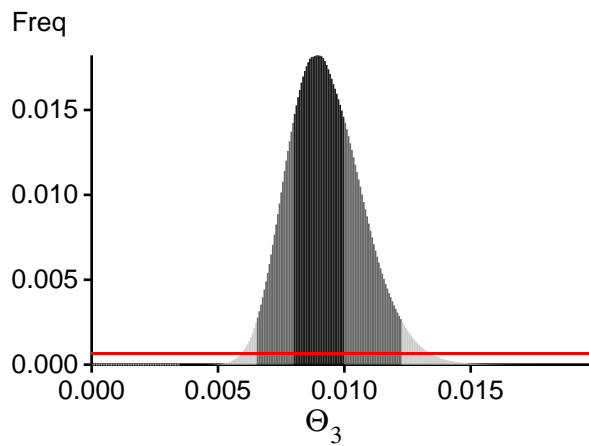
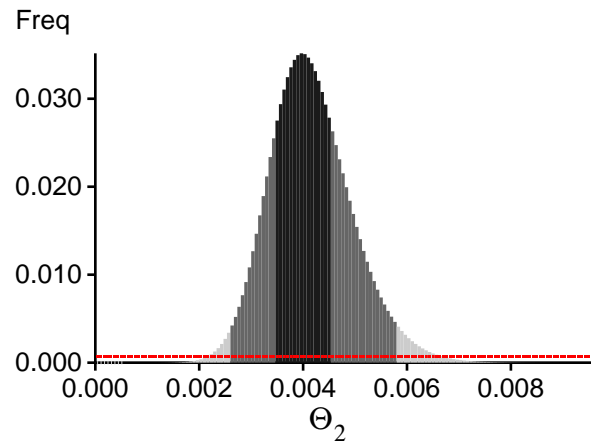
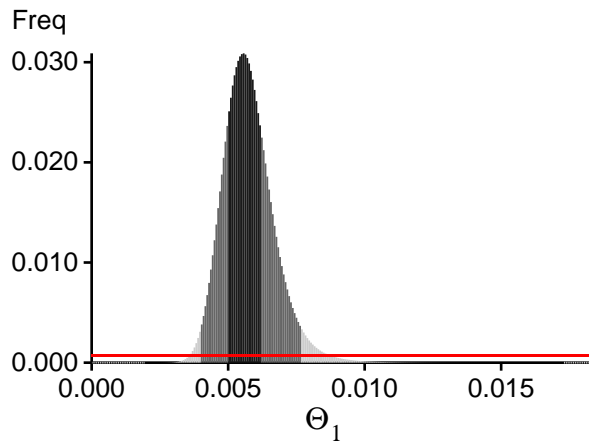
7	$\Theta_1$	0.00413	0.01140	0.01203	0.01220	0.02113	0.01270	0.01312
7	$\Theta_2$	0.00827	0.00827	0.00870	0.00887	0.02707	0.03117	0.03847
7	$\Theta_3$	0.00580	0.01020	0.01150	0.01160	0.02200	0.01417	0.01607
7	$M_{1 \rightarrow 2}$	120.000	203.333	268.333	433.333	693.333	1188.333	1386.303
7	$M_{2 \rightarrow 3}$	0.000	10.000	31.667	86.667	206.667	75.000	89.842
<hr/>								
8	$\Theta_1$	0.00280	0.00400	0.00517	0.00707	0.01507	0.00837	0.00947
8	$\Theta_2$	0.00147	0.00473	0.00550	0.00607	0.01227	0.00663	0.00791
8	$\Theta_3$	0.00247	0.00480	0.00670	0.00687	0.01433	0.00743	0.00836
8	$M_{1 \rightarrow 2}$	50.000	90.000	105.000	226.667	736.667	275.000	387.356
8	$M_{2 \rightarrow 3}$	0.000	30.000	118.333	160.000	440.000	168.333	211.529
<hr/>								
9	$\Theta_1$	0.00373	0.00607	0.00670	0.00700	0.02387	0.01470	0.01527
9	$\Theta_2$	0.00087	0.00147	0.00203	0.00307	0.00587	0.00283	0.00316
9	$\Theta_3$	0.04973	0.05213	0.05223	0.05227	0.05453	0.05657	0.05609
9	$M_{1 \rightarrow 2}$	33.333	83.333	161.667	216.667	633.333	278.333	339.051
9	$M_{2 \rightarrow 3}$	280.000	550.000	561.667	566.667	1276.667	748.333	813.296
<hr/>								
10	$\Theta_1$	0.00247	0.00373	0.00443	0.00727	0.01633	0.00810	0.00983
10	$\Theta_2$	0.00147	0.00493	0.00570	0.00740	0.01640	0.00850	0.01218
10	$\Theta_3$	0.00667	0.01253	0.01263	0.01367	0.02960	0.01657	0.01946
10	$M_{1 \rightarrow 2}$	46.667	240.000	281.667	310.000	893.333	441.667	667.265
10	$M_{2 \rightarrow 3}$	0.000	13.333	21.667	80.000	320.000	101.667	176.669
<hr/>								
All	$\Theta_1$	0.00393	0.00493	0.00557	0.00620	0.00767	0.00577	0.00577
All	$\Theta_2$	0.00253	0.00340	0.00397	0.00453	0.00580	0.00417	0.00416
All	$\Theta_3$	0.00647	0.00793	0.00890	0.01000	0.01227	0.00923	0.00928
All	$M_{1 \rightarrow 2}$	160.000	213.333	251.667	293.333	386.667	268.333	270.503
All	$M_{2 \rightarrow 3}$	40.000	83.333	108.333	130.000	170.000	108.333	106.480

## 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 = Exp[ ln(Prob(D | thisModel) - ln( Prob( D | otherModel)

or as LBF = 2 (ln(Prob(D | thisModel) - ln( Prob( D | otherModel))

shows the support for thisModel]

Locus	Raw thermodynamic score(1a)	Bezier approximation score(1b)	Harmonic mean(2)
1	-2960.67	-2406.68	-2310.49
2	-2458.77	-2035.10	-1962.41
3	-2633.68	-2187.60	-2118.03
4	-2968.59	-2374.12	-2261.29
5	-2843.78	-2294.35	-2197.53
6	-3003.20	-2549.51	-2392.59
7	-2789.64	-2289.48	-2204.74
8	-2954.45	-2360.37	-2249.33
9	-3010.16	-2412.23	-2294.13
10	-2640.75	-2225.42	-2161.37
All	-28205.85	-23077.01	-22094.07

(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 = 57.834983]

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
$\Theta_1$	275433/500739	0.55005
$\Theta_2$	224891/500012	0.44977
$\Theta_3$	233748/499696	0.46778
$M_{1 \rightarrow 2}$	212072/500835	0.42344
$M_{2 \rightarrow 3}$	231562/499110	0.46395
Genealogies	351732/2499608	0.14071

*MCMC-Autocorrelation and Effective MCMC Sample Size*

Parameter	Autocorrelation	Effective Sample Size
$\Theta_1$	0.70234	14577.91
$\Theta_2$	0.55301	21612.76
$\Theta_3$	0.38243	32826.16
$M_{1 \rightarrow 2}$	0.69847	12986.43
$M_{2 \rightarrow 3}$	0.59704	19450.12
Genealogies	0.71747	12375.29

## *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