

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 Sun Jun 2 00:01:40 2019

Program finished at Sun Jun 2 00:17:03 2019 [Runtime:0000:00:15:23]



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) 2612977134

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	Θ_1	<displayed>
2	Θ_2	<displayed>
3	Θ_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	500.0	1000.	100.0	1500	0.10000
5	M **	Uniform	0.000000	500.0	1000.	100.0	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	Θ_1	0.00360	0.00713	0.00877	0.00993	0.01620	0.00937	0.01037
1	Θ_2	0.00140	0.00373	0.00437	0.00513	0.01013	0.00503	0.00565
1	Θ_3	0.00393	0.00820	0.00937	0.01067	0.02427	0.01363	0.01552
1	$M_{1 \rightarrow 2}$	11.333	88.667	91.000	95.333	204.667	107.000	143.048
1	$M_{2 \rightarrow 3}$	0.000	48.000	57.000	60.667	170.667	71.667	92.589
2	Θ_1	0.00247	0.00460	0.00550	0.00620	0.01127	0.00650	0.00702
2	Θ_2	0.00080	0.00220	0.00350	0.00440	0.00913	0.00430	0.00499
2	Θ_3	0.00133	0.00287	0.00323	0.00540	0.00960	0.00510	0.00578
2	$M_{1 \rightarrow 2}$	32.667	125.333	132.333	137.333	230.000	148.333	162.097
2	$M_{2 \rightarrow 3}$	4.000	38.667	42.333	48.000	208.667	95.667	130.782
3	Θ_1	0.00253	0.00513	0.00543	0.00553	0.01167	0.00670	0.00707
3	Θ_2	0.00147	0.00393	0.00643	0.00920	0.02460	0.01583	0.02423
3	Θ_3	0.00540	0.00940	0.00957	0.01020	0.02240	0.01323	0.01470
3	$M_{1 \rightarrow 2}$	337.333	412.000	413.000	413.333	462.667	499.000	519.034
3	$M_{2 \rightarrow 3}$	0.667	12.667	19.667	24.000	97.333	37.000	44.511
4	Θ_1	0.00187	0.00353	0.00523	0.00653	0.01247	0.00630	0.00795
4	Θ_2	0.00073	0.00173	0.00310	0.00473	0.01380	0.00537	0.00703
4	Θ_3	0.00347	0.01080	0.01130	0.01287	0.01900	0.01143	0.01282
4	$M_{1 \rightarrow 2}$	21.333	83.333	90.333	98.000	268.000	199.000	248.163
4	$M_{2 \rightarrow 3}$	0.000	8.667	16.333	40.667	158.000	55.667	83.091
5	Θ_1	0.00140	0.00227	0.00377	0.00553	0.01660	0.00743	0.00977
5	Θ_2	0.00067	0.00340	0.00377	0.00500	0.01147	0.00537	0.00723
5	Θ_3	0.00500	0.01267	0.01323	0.01413	0.02900	0.01763	0.01974
5	$M_{1 \rightarrow 2}$	132.000	133.333	135.000	142.667	150.000	409.667	449.100
5	$M_{2 \rightarrow 3}$	0.000	17.333	36.333	45.333	170.000	64.333	105.032
6	Θ_1	0.00113	0.00200	0.00250	0.00367	0.00593	0.00330	0.00348
6	Θ_2	0.00067	0.00167	0.00410	0.00540	0.01687	0.00577	0.01001
6	Θ_3	0.00520	0.01173	0.01197	0.01273	0.02447	0.01510	0.01705
6	$M_{1 \rightarrow 2}$	826.667	977.333	981.000	982.667	999.333	813.000	769.159
6	$M_{2 \rightarrow 3}$	12.000	64.000	65.000	70.000	204.000	100.333	116.489

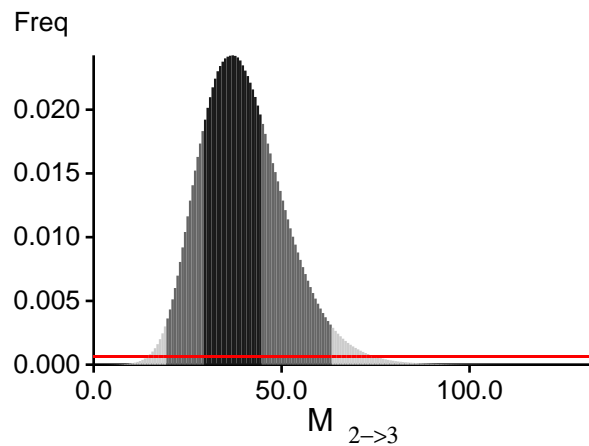
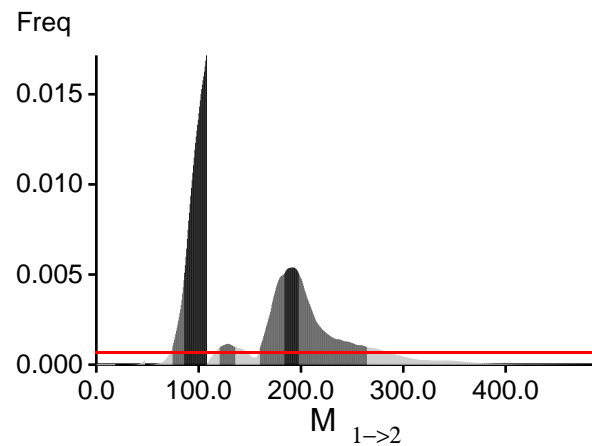
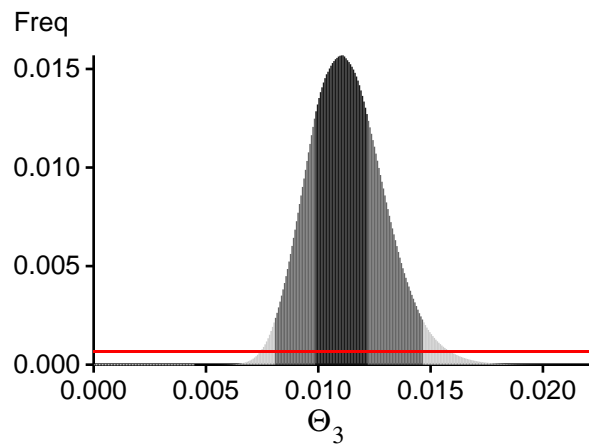
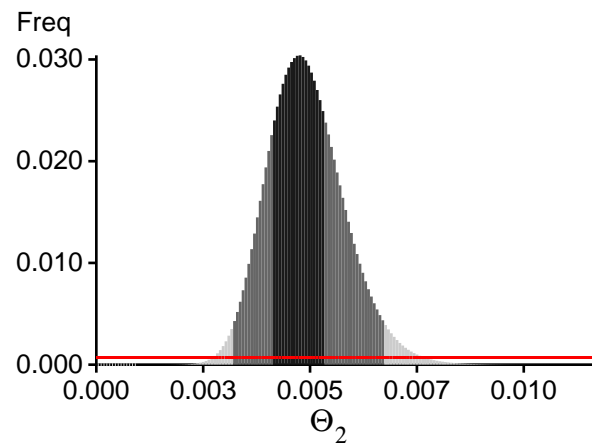
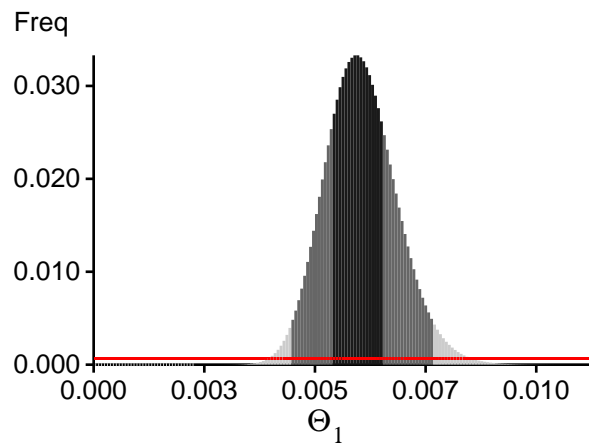
7	Θ_1	0.00427	0.00713	0.00803	0.00973	0.01547	0.00930	0.00973
7	Θ_2	0.00287	0.00553	0.00570	0.00607	0.02333	0.02663	0.03432
7	Θ_3	0.00653	0.01273	0.01317	0.01360	0.02753	0.01623	0.01794
7	$M_{1 \rightarrow 2}$	537.333	548.667	549.667	553.333	560.667	645.000	633.997
7	$M_{2 \rightarrow 3}$	0.000	17.333	21.000	27.333	102.667	38.333	50.415
<hr/>								
8	Θ_1	0.00287	0.00467	0.00563	0.00727	0.01513	0.00723	0.00889
8	Θ_2	0.00153	0.00493	0.00543	0.00560	0.01707	0.00910	0.01121
8	Θ_3	0.00293	0.00647	0.00710	0.00727	0.01573	0.00917	0.01032
8	$M_{1 \rightarrow 2}$	22.667	56.000	69.000	70.000	272.000	128.333	185.033
8	$M_{2 \rightarrow 3}$	1.333	53.333	56.333	71.333	203.333	114.333	147.983
<hr/>								
9	Θ_1	0.00367	0.00660	0.00830	0.00967	0.01867	0.01017	0.01136
9	Θ_2	0.00113	0.00253	0.00337	0.00440	0.00940	0.00470	0.00517
9	Θ_3	0.01440	0.01513	0.01530	0.01533	0.02167	0.04097	0.04578
9	$M_{1 \rightarrow 2}$	16.667	59.333	83.000	92.000	179.333	94.333	115.036
9	$M_{2 \rightarrow 3}$	618.667	625.333	626.333	626.667	642.667	473.667	483.016
<hr/>								
10	Θ_1	0.00267	0.00400	0.00523	0.00620	0.01087	0.00617	0.00658
10	Θ_2	0.00167	0.00620	0.00650	0.00793	0.02133	0.00963	0.01253
10	Θ_3	0.00687	0.01287	0.01437	0.01500	0.02807	0.01763	0.01965
10	$M_{1 \rightarrow 2}$	342.000	402.000	403.000	404.667	428.000	413.667	427.101
10	$M_{2 \rightarrow 3}$	0.000	15.333	20.333	32.000	96.667	37.000	50.401
<hr/>								
All	Θ_1	0.00440	0.00533	0.00597	0.00653	0.00767	0.00610	0.00605
All	Θ_2	0.00313	0.00407	0.00477	0.00533	0.00673	0.00490	0.00492
All	Θ_3	0.00800	0.00980	0.01110	0.01220	0.01467	0.01123	0.01128
All	$M_{1 \rightarrow 2}$	74.000	85.333	107.667	108.000	108.000	159.000	157.942
All	$M_{2 \rightarrow 3}$	18.667	28.667	37.000	44.667	63.333	39.667	39.941

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	-2786.78	-2367.22	-2295.30
2	-2320.86	-2010.66	-1958.44
3	-2496.88	-2163.85	-2115.76
4	-2844.37	-2324.81	-2239.87
5	-2622.60	-2292.59	-2181.90
6	-2862.65	-2526.67	-2374.74
7	-2678.95	-2265.09	-2201.38
8	-2808.78	-2319.77	-2229.52
9	-2762.79	-2341.81	-2275.50
10	-2572.29	-2213.13	-2152.15
All	-26716.72	-22785.37	-21984.34

(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 = 40.230092]

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	227084/500435	0.45377
Θ_2	269735/500584	0.53884
Θ_3	238606/499322	0.47786
$M_{1 \rightarrow 2}$	236063/498334	0.47370
$M_{2 \rightarrow 3}$	261122/500345	0.52188
Genealogies	370637/2500980	0.14820

MCMC-Autocorrelation and Effective MCMC Sample Size

Parameter	Autocorrelation	Effective Sampe Size
Θ_1	0.52454	25267.70
Θ_2	0.56819	22884.81
Θ_3	0.34132	35923.26
$M_{1 \rightarrow 2}$	0.40080	33138.37
$M_{2 \rightarrow 3}$	0.45100	28173.07
Genealogies	0.72194	12046.30

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.

Param 6 (Locus 6): Upper prior boundary seems too low!