

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 11:12:13 2019

Program finished at Sun Jun 2 11:13:38 2019 [Runtime:0000:00:01:25]



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) 2065230715

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	1	2
1 Arbon_1	*	*	0
1 Berg_2	*	*	0
2 Chur_3	d	d	*

Order of parameters:

1	Θ_1	<displayed>
2	Θ_2	<displayed>
3	$\Delta_{1 \rightarrow 2}$	<displayed>
4	$\sigma_{1 \rightarrow 2}$	<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

-Divergence time estimation:

Normal Distribution Shortcut (mean and standard dev.)

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 *Exponential	0.000000	0.060	0.100	-	1500	0.12500
2	Theta *Exponential	0.000000	0.060	0.100	-	1500	0.12500
3	Splittime mean *Exponential	0.000000	0.080	0.100	-	1500	0.12500
4	Splittime std *Exponential	0.000000	0.080	0.100	-	1500	0.12500

[* * means priors were set globally]

Markov chain settings:

Long chain

Number of chains

1

Recorded steps [a]

5000

Increment (record every x step [b])

10

Number of concurrent chains (replicates) [c]

1

Visited (sampled) parameter values [a*b*c]

50000

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_prior
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		
1 Berg_2			1		10		
			2		10		
			3		10		
			4		10		
			5		10		
			6		10		
			7		10		
			8		10		
			9		10		
			10		10		
2 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)

Bayesian Analysis: Posterior distribution table

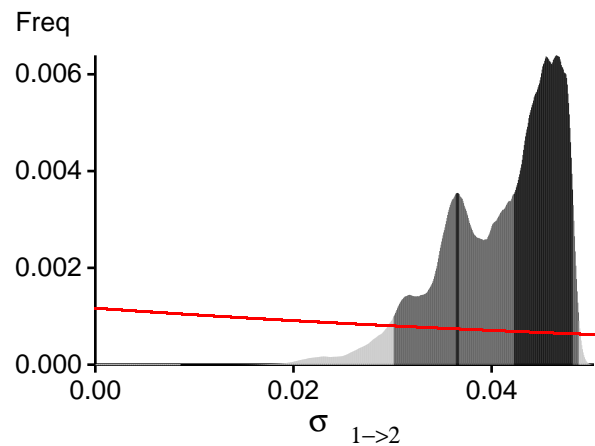
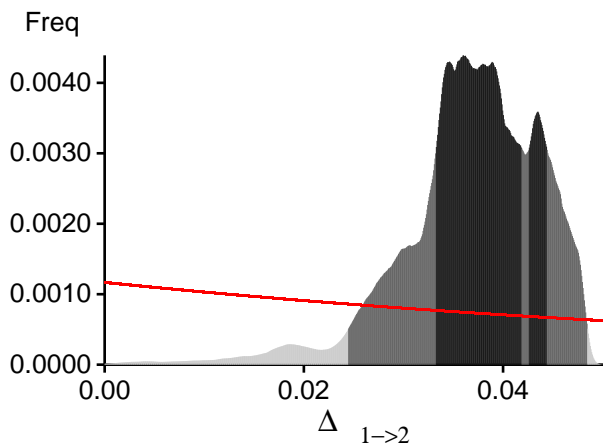
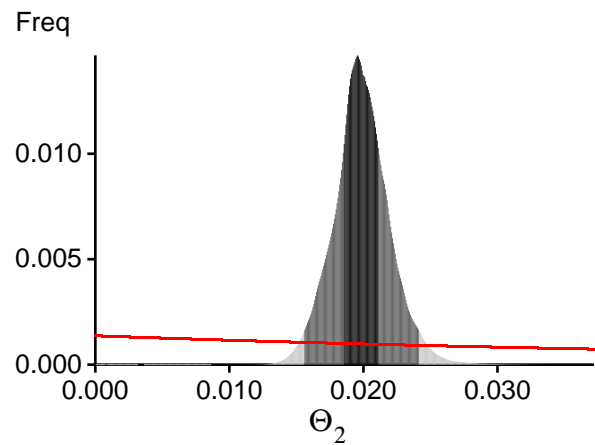
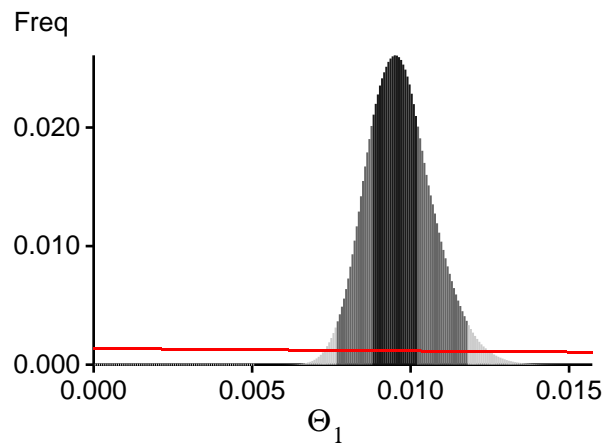
Locus	Parameter	2.5%	25.0%	Mode	75.0%	97.5%	Median	Mean
1	Θ_1	0.00960	0.01180	0.01203	0.01213	0.01353	0.01190	0.01244
1	Θ_2	0.01660	0.01733	0.01783	0.01787	0.01940	0.02097	0.02347
1	$D_{1 \rightarrow 2}$	0.04227	0.04313	0.04337	0.04340	0.04360	0.04323	0.04419
1	$S_{1 \rightarrow 2}$	0.03853	0.03860	0.03870	0.03873	0.03873	0.04977	0.05172
2	Θ_1	0.00680	0.00693	0.00703	0.00720	0.00720	0.01063	0.01126
2	Θ_2	0.00280	0.00567	0.00577	0.00587	0.00753	0.00683	0.00842
2	$D_{1 \rightarrow 2}$	0.01600	0.01687	0.01703	0.01720	0.01733	0.02850	0.03338
2	$S_{1 \rightarrow 2}$	0.02227	0.02240	0.02257	0.02260	0.02273	0.04983	0.05136
3	Θ_1	0.00947	0.00973	0.00983	0.00987	0.00987	0.01243	0.01267
3	Θ_2	0.01027	0.01260	0.01290	0.01300	0.01587	0.01483	0.01683
3	$D_{1 \rightarrow 2}$	0.01940	0.01940	0.01983	0.01987	0.02193	0.03063	0.03542
3	$S_{1 \rightarrow 2}$	0.01947	0.01947	0.01957	0.01960	0.01960	0.05117	0.05253
4	Θ_1	0.00567	0.00820	0.00830	0.00840	0.00973	0.00837	0.00882
4	Θ_2	0.02073	0.02220	0.02250	0.02253	0.02287	0.02257	0.02460
4	$D_{1 \rightarrow 2}$	0.05013	0.05087	0.05097	0.05100	0.05100	0.04717	0.04700
4	$S_{1 \rightarrow 2}$	0.03180	0.03180	0.03197	0.03200	0.03220	0.05157	0.05363
5	Θ_1	0.00427	0.00747	0.00763	0.00767	0.00840	0.00610	0.00659
5	Θ_2	0.02960	0.03107	0.03117	0.03127	0.03400	0.02910	0.03187
5	$D_{1 \rightarrow 2}$	0.03367	0.03367	0.03377	0.03380	0.03447	0.04037	0.04224
5	$S_{1 \rightarrow 2}$	0.06847	0.06913	0.06923	0.06933	0.06947	0.04883	0.05053
6	Θ_1	0.00580	0.00667	0.00677	0.00693	0.00953	0.00857	0.00899
6	Θ_2	0.03500	0.03560	0.03570	0.03580	0.03613	0.04003	0.04244
6	$D_{1 \rightarrow 2}$	0.02747	0.02760	0.02783	0.02800	0.02847	0.03237	0.03518
6	$S_{1 \rightarrow 2}$	0.01600	0.01613	0.01623	0.01627	0.01627	0.04177	0.04490
7	Θ_1	0.01333	0.01340	0.01350	0.01353	0.01793	0.01503	0.01570
7	Θ_2	0.00907	0.01373	0.01383	0.01387	0.01847	0.01590	0.01809
7	$D_{1 \rightarrow 2}$	0.01807	0.02387	0.02410	0.02413	0.02427	0.02963	0.03380
7	$S_{1 \rightarrow 2}$	0.06593	0.06633	0.06650	0.06653	0.06667	0.04963	0.05079
8	Θ_1	0.00667	0.00773	0.00783	0.00787	0.00980	0.01097	0.01170

Locus	Parameter	2.5%	25.0%	Mode	75.0%	97.5%	Median	Mean
8	Θ_2	0.01947	0.01967	0.01990	0.01993	0.02140	0.01937	0.02068
8	$D_{1 \rightarrow 2}$	0.06193	0.06193	0.06203	0.06220	0.06220	0.04317	0.04452
8	$S_{1 \rightarrow 2}$	0.04840	0.04847	0.04863	0.04880	0.05033	0.05170	0.05339
9	Θ_1	0.01113	0.01147	0.01177	0.01180	0.01233	0.01257	0.01338
9	Θ_2	0.02027	0.02247	0.02257	0.02273	0.02313	0.02257	0.02472
9	$D_{1 \rightarrow 2}$	0.03647	0.03880	0.03890	0.03907	0.04060	0.03970	0.04229
9	$S_{1 \rightarrow 2}$	0.02773	0.02880	0.02890	0.02900	0.02913	0.05050	0.05265
10	Θ_1	0.00627	0.00773	0.00797	0.00807	0.00807	0.00883	0.00932
10	Θ_2	0.01640	0.01720	0.01743	0.01747	0.01773	0.02210	0.02403
10	$D_{1 \rightarrow 2}$	0.00367	0.00427	0.00437	0.00440	0.00440	0.03537	0.03852
10	$S_{1 \rightarrow 2}$	0.02067	0.02073	0.02083	0.02087	0.02113	0.05057	0.05172
All	Θ_1	0.00760	0.00873	0.00950	0.01020	0.01180	0.00970	0.00967
All	Θ_2	0.01553	0.01847	0.01957	0.02113	0.02413	0.01990	0.01989
All	$D_{1 \rightarrow 2}$	0.02440	0.03320	0.03603	0.04187	0.04847	0.03757	0.03682
All	$S_{1 \rightarrow 2}$	0.03007	0.04220	0.04657	0.04820	0.04880	0.04250	0.04091

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	-2529.67	-2317.39	-2292.53
2	-2118.88	-1978.29	-1959.75
3	-2304.40	-2128.08	-2111.65
4	-2538.69	-2268.96	-2236.78
5	-2380.99	-2198.28	-2174.73
6	-2552.82	-2327.41	-2299.31
7	-2461.12	-2227.49	-2201.30
8	-2532.69	-2260.29	-2227.64
9	-2527.16	-2298.13	-2273.36
10	-2389.59	-2180.58	-2155.14
All	-24335.80	-22184.69	-21931.97

(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 = 0.218941]

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	8687/62233	0.13959
Θ_2	20214/62313	0.32439
$\Delta_{1 \rightarrow 2}$	51121/62301	0.82055
$\sigma_{1 \rightarrow 2}$	43528/62121	0.70070
Genealogies	35226/251032	0.14032

MCMC-Autocorrelation and Effective MCMC Sample Size

Parameter	Autocorrelation	Effective Sampe Size
Θ_1	0.81203	7349.90
Θ_2	0.64454	14544.28
$\Delta_{1 \rightarrow 2}$	0.33854	34780.97
$\sigma_{1 \rightarrow 2}$	0.45250	26725.51
Genealogies	0.45250	26725.51

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