

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 00:17:04 2019

Program finished at Sun Jun 2 00:34:11 2019 [Runtime:0000:00:17:07]



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) 3355861582

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	D	*	0
3 Chur_3	0	D	*

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>
6	$\Delta_{1 \rightarrow 2}$	<displayed>
7	$\sigma_{1 \rightarrow 2}$	<displayed>
8	$\Delta_{2 \rightarrow 3}$	<displayed>
9	$\sigma_{2 \rightarrow 3}$	<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

-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 **	Uniform	0.000000	0.050	0.100	0.010	1500	0.05556
2 Theta **	Uniform	0.000000	0.050	0.100	0.010	1500	0.05556
3 Theta **	Uniform	0.000000	0.050	0.100	0.010	1500	0.05556
4 M **	Uniform	0.000000	500.0	1000.	100.0	1500	0.05556
5 M **	Uniform	0.000000	500.0	1000.	100.0	1500	0.05556
6 Splittime mean **	Uniform	0.000000	0.050	0.100	0.010	1500	0.05556
7 Splittime std **	Uniform	0.000000	0.050	0.100	0.100	1500	0.05556
8 Splittime mean **	Uniform	0.000000	0.050	0.100	0.010	1500	0.05556
9 Splittime std **	Uniform	0.000000	0.050	0.100	0.100	1500	0.05556

[* * 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_model2
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)

Bayesian Analysis: Posterior distribution table

Locus	Parameter	2.5%	25.0%	Mode	75.0%	97.5%	Median	Mean
1	Θ_1	0.00407	0.00747	0.00837	0.01007	0.01927	0.01097	0.01319
1	Θ_2	0.00093	0.00233	0.00270	0.00333	0.01593	0.00683	0.00828
1	Θ_3	0.00513	0.01153	0.01183	0.01273	0.02733	0.01523	0.01727
1	$M_{1 \rightarrow 2}$	5.333	20.000	44.333	55.333	137.333	130.333	207.357
1	$M_{2 \rightarrow 3}$	0.000	0.000	7.667	29.333	92.667	31.000	45.725
1	$D_{1 \rightarrow 2}$	0.00027	0.00027	0.00050	0.00073	0.00480	0.03837	0.04200
1	$S_{1 \rightarrow 2}$	0.00047	0.00047	0.00063	0.00067	0.00067	0.05610	0.05471
1	$D_{2 \rightarrow 3}$	0.00093	0.00113	0.00203	0.00213	0.00233	0.03990	0.04308
1	$S_{2 \rightarrow 3}$	0.07580	0.07660	0.07677	0.07680	0.07680	0.05603	0.05463
2	Θ_1	0.00253	0.00760	0.00777	0.00793	0.01647	0.00917	0.00963
2	Θ_2	0.00120	0.00200	0.00277	0.00413	0.00953	0.00437	0.00501
2	Θ_3	0.00147	0.00407	0.00450	0.00493	0.01233	0.00557	0.00670
2	$M_{1 \rightarrow 2}$	18.667	124.000	125.667	126.000	126.000	285.667	334.927
2	$M_{2 \rightarrow 3}$	0.000	0.000	0.333	17.333	119.333	39.000	67.871
2	$D_{1 \rightarrow 2}$	0.01153	0.01300	0.01330	0.01333	0.01333	0.03483	0.04037
2	$S_{1 \rightarrow 2}$	0.00020	0.00027	0.00043	0.00047	0.00080	0.05343	0.05248
2	$D_{2 \rightarrow 3}$	0.00580	0.00627	0.00643	0.00647	0.00693	0.04577	0.04746
2	$S_{2 \rightarrow 3}$	0.09760	0.09867	0.09883	0.09893	0.09933	0.05770	0.05548
3	Θ_1	0.00253	0.00520	0.00577	0.00613	0.01480	0.00823	0.00885
3	Θ_2	0.00247	0.00860	0.00943	0.00947	0.02620	0.01290	0.01972
3	Θ_3	0.00447	0.01107	0.01183	0.01200	0.01987	0.01277	0.01445
3	$M_{1 \rightarrow 2}$	2.000	66.000	67.667	68.000	236.667	301.667	353.449
3	$M_{2 \rightarrow 3}$	0.000	0.000	1.000	22.667	120.000	31.667	60.205
3	$D_{1 \rightarrow 2}$	0.00000	0.00000	0.00137	0.00213	0.01373	0.01037	0.01959
3	$S_{1 \rightarrow 2}$	0.00067	0.00080	0.00103	0.00107	0.00127	0.03177	0.03989
3	$D_{2 \rightarrow 3}$	0.04193	0.04220	0.04250	0.04253	0.04253	0.04643	0.04811
3	$S_{2 \rightarrow 3}$	0.07920	0.07980	0.07997	0.08027	0.08027	0.05623	0.05435
4	Θ_1	0.00207	0.00660	0.00670	0.00687	0.01240	0.00657	0.00833
4	Θ_2	0.00080	0.00227	0.00317	0.00593	0.01567	0.00690	0.00918
4	Θ_3	0.00580	0.01220	0.01237	0.01240	0.02660	0.01423	0.01636
4	$M_{1 \rightarrow 2}$	0.000	18.000	21.000	36.000	159.333	111.667	176.926
4	$M_{2 \rightarrow 3}$	0.000	0.000	0.333	23.333	112.667	21.667	44.796
4	$D_{1 \rightarrow 2}$	0.00000	0.00000	0.00037	0.00467	0.01133	0.01397	0.02307

Locus	Parameter	2.5%	25.0%	Mode	75.0%	97.5%	Median	Mean
4	$S_{1 \rightarrow 2}$	0.00053	0.00087	0.00123	0.00127	0.00127	0.04303	0.04588
4	$D_{2 \rightarrow 3}$	0.04613	0.04700	0.04750	0.04767	0.04800	0.04530	0.04680
4	$S_{2 \rightarrow 3}$	0.09120	0.09193	0.09223	0.09227	0.09247	0.05883	0.05755
5	Θ_1	0.00167	0.00260	0.00363	0.00487	0.01007	0.00523	0.00669
5	Θ_2	0.00120	0.00227	0.00237	0.00253	0.01440	0.00717	0.01066
5	Θ_3	0.01447	0.02253	0.02323	0.02333	0.03127	0.02417	0.02649
5	$M_{1 \rightarrow 2}$	199.333	199.333	200.333	205.333	210.667	514.333	512.509
5	$M_{2 \rightarrow 3}$	0.000	0.000	5.000	12.000	78.000	20.333	28.914
5	$D_{1 \rightarrow 2}$	0.00000	0.00260	0.00270	0.00293	0.00413	0.03230	0.03892
5	$S_{1 \rightarrow 2}$	0.01827	0.01847	0.01857	0.01867	0.01893	0.04730	0.04814
5	$D_{2 \rightarrow 3}$	0.01640	0.01653	0.01663	0.01667	0.01667	0.04983	0.05031
5	$S_{2 \rightarrow 3}$	0.04867	0.04867	0.04877	0.04880	0.04960	0.05537	0.05420
6	Θ_1	0.00233	0.00547	0.00557	0.00567	0.00907	0.00517	0.00552
6	Θ_2	0.00120	0.00347	0.00363	0.00513	0.01867	0.00763	0.01184
6	Θ_3	0.03113	0.03160	0.03250	0.03253	0.04660	0.03990	0.04282
6	$M_{1 \rightarrow 2}$	462.000	468.000	469.000	469.333	469.333	574.333	568.440
6	$M_{2 \rightarrow 3}$	0.000	3.333	11.000	15.333	53.333	19.667	23.513
6	$D_{1 \rightarrow 2}$	0.00000	0.00000	0.00163	0.00447	0.02060	0.02150	0.03067
6	$S_{1 \rightarrow 2}$	0.00680	0.00727	0.00763	0.00780	0.00780	0.04663	0.04777
6	$D_{2 \rightarrow 3}$	0.07427	0.07440	0.07457	0.07467	0.07560	0.05410	0.05336
6	$S_{2 \rightarrow 3}$	0.02967	0.02973	0.02990	0.03000	0.03113	0.05103	0.05110
7	Θ_1	0.00460	0.00840	0.00883	0.00900	0.01773	0.01103	0.01188
7	Θ_2	0.00247	0.00620	0.00630	0.00787	0.02567	0.01590	0.02420
7	Θ_3	0.00727	0.01107	0.01163	0.01167	0.02740	0.01703	0.01884
7	$M_{1 \rightarrow 2}$	7.333	39.333	41.667	42.000	104.000	283.667	340.700
7	$M_{2 \rightarrow 3}$	0.000	0.000	5.667	15.333	103.333	31.667	44.698
7	$D_{1 \rightarrow 2}$	0.00427	0.00920	0.01023	0.01100	0.01193	0.00750	0.00680
7	$S_{1 \rightarrow 2}$	0.00973	0.01100	0.01183	0.01220	0.01253	0.01370	0.01473
7	$D_{2 \rightarrow 3}$	0.00987	0.01033	0.01070	0.01073	0.01100	0.04810	0.04907
7	$S_{2 \rightarrow 3}$	0.06900	0.06940	0.06950	0.06953	0.07007	0.05383	0.05310
8	Θ_1	0.00307	0.00473	0.00763	0.00813	0.01520	0.00810	0.01054
8	Θ_2	0.00180	0.00500	0.00570	0.00633	0.02020	0.00990	0.01262
8	Θ_3	0.00327	0.00613	0.00623	0.00627	0.02013	0.01177	0.01347
8	$M_{1 \rightarrow 2}$	0.000	22.000	32.333	42.000	271.333	164.333	222.307
8	$M_{2 \rightarrow 3}$	0.000	0.000	0.333	25.333	88.667	26.333	45.349
8	$D_{1 \rightarrow 2}$	0.00000	0.00000	0.00063	0.00547	0.01433	0.01777	0.02623
8	$S_{1 \rightarrow 2}$	0.00107	0.00107	0.00117	0.00120	0.00120	0.04883	0.04902

Locus	Parameter	2.5%	25.0%	Mode	75.0%	97.5%	Median	Mean
8	D _{2->3}	0.00300	0.00360	0.00390	0.00400	0.00413	0.04377	0.04544
8	S _{2->3}	0.04420	0.04427	0.04443	0.04460	0.04500	0.05737	0.05613
9	Θ_1	0.00413	0.01107	0.01150	0.01153	0.01927	0.01190	0.01308
9	Θ_2	0.00127	0.00213	0.00257	0.00287	0.01367	0.00657	0.00745
9	Θ_3	0.00993	0.01380	0.01423	0.01467	0.03087	0.02757	0.03501
9	M _{1->2}	0.000	8.000	13.667	14.667	82.000	27.000	49.012
9	M _{2->3}	0.000	0.667	2.333	8.667	200.000	172.333	205.610
9	D _{1->2}	0.00000	0.00080	0.00130	0.00247	0.01727	0.02117	0.02800
9	S _{1->2}	0.04480	0.04647	0.04657	0.04660	0.04673	0.06090	0.06026
9	D _{2->3}	0.00000	0.00000	0.00270	0.00613	0.00720	0.01257	0.02508
9	S _{2->3}	0.00073	0.00227	0.00370	0.00407	0.01600	0.02903	0.03718
10	Θ_1	0.00253	0.00420	0.00463	0.00613	0.01133	0.00623	0.00680
10	Θ_2	0.00253	0.00967	0.01070	0.01093	0.02273	0.01290	0.01578
10	Θ_3	0.00713	0.01513	0.01570	0.01613	0.02840	0.01857	0.02186
10	M _{1->2}	0.000	8.667	17.000	24.667	189.333	97.000	176.765
10	M _{2->3}	0.000	0.000	2.333	19.333	181.333	43.667	98.194
10	D _{1->2}	0.00000	0.00000	0.00150	0.00293	0.01513	0.00537	0.01214
10	S _{1->2}	0.00027	0.00047	0.00077	0.00107	0.00120	0.02163	0.03086
10	D _{2->3}	0.00813	0.01027	0.01057	0.01073	0.01160	0.04197	0.04462
10	S _{2->3}	0.04540	0.04553	0.04563	0.04573	0.04780	0.05710	0.05584
All	Θ_1	0.00480	0.00580	0.00650	0.00720	0.00873	0.00677	0.00674
All	Θ_2	0.00340	0.00447	0.00523	0.00600	0.00787	0.00550	0.00556
All	Θ_3	0.01060	0.01313	0.01450	0.01613	0.01947	0.01483	0.01489
All	M _{1->2}	19.333	32.000	41.667	52.667	86.000	47.000	49.607
All	M _{2->3}	0.000	0.000	0.333	4.667	16.000	5.000	5.711
All	D _{1->2}	0.00000	0.00000	0.00003	0.00113	0.00433	0.00117	0.00149
All	S _{1->2}	0.00787	0.00933	0.01330	0.01587	0.02247	0.01510	0.01679
All	D _{2->3}	0.00000	0.00000	0.00003	0.00747	0.01560	0.00990	0.01534
All	S _{2->3}	0.01820	0.04027	0.04717	0.04813	0.04887	0.03790	0.03625

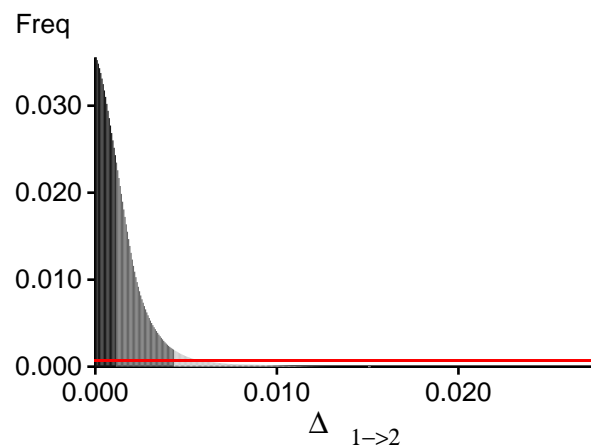
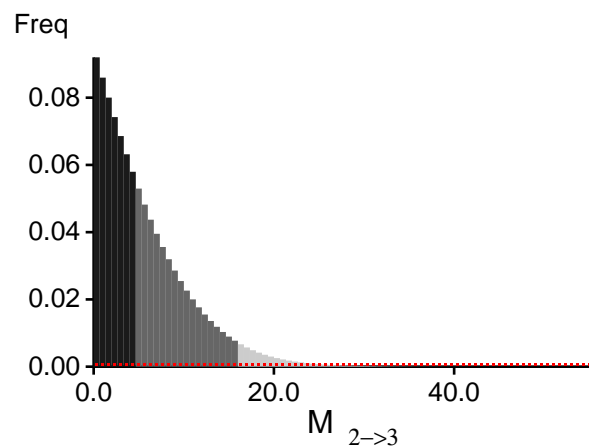
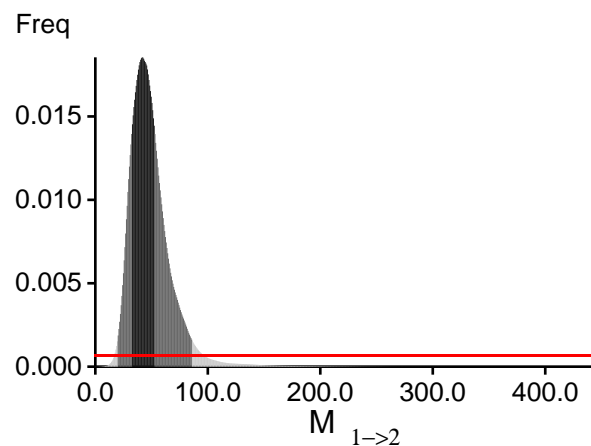
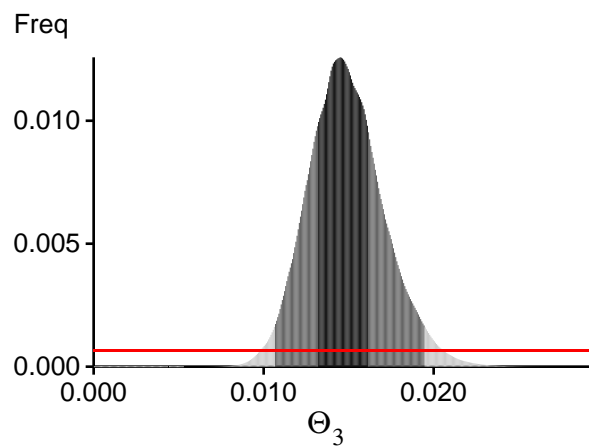
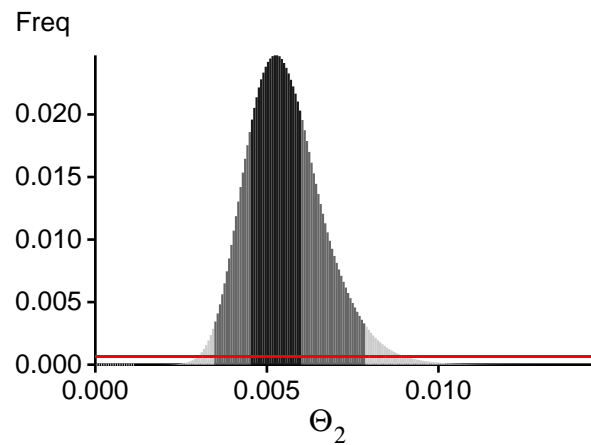
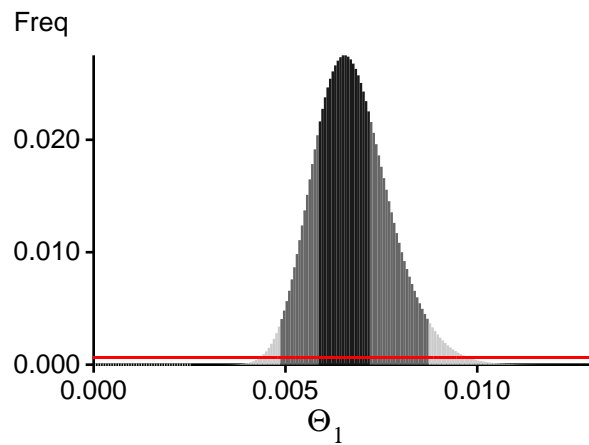
Citation suggestions:

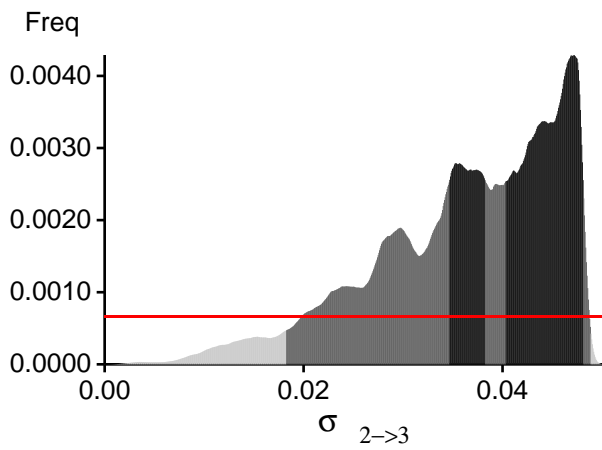
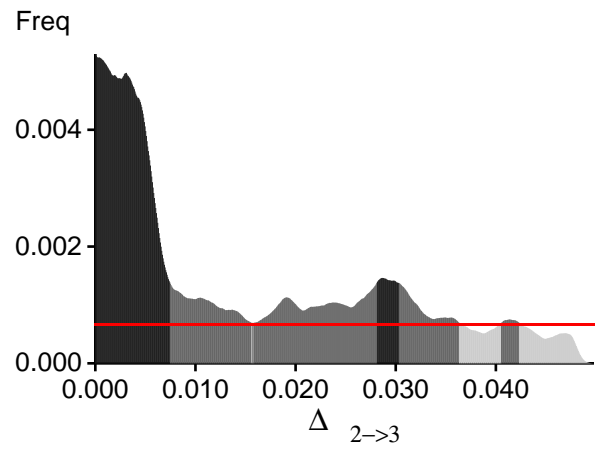
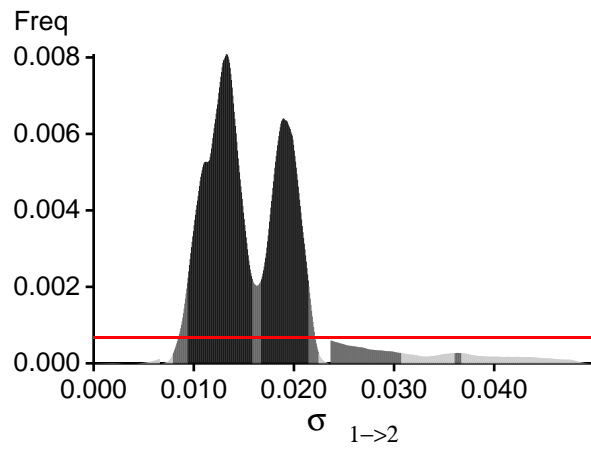
Beerli P., 2006. Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. *Bioinformatics* 22:341-345

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	-2955.31	-2391.13	-2292.61
2	-2467.57	-2036.62	-1961.93
3	-2654.22	-2190.63	-2115.74
4	-2995.03	-2349.66	-2241.24
5	-2776.29	-2266.38	-2181.94
6	-2803.93	-2372.26	-2300.20
7	-2760.83	-2281.16	-2202.68
8	-2967.75	-2336.23	-2229.22
9	-2981.54	-2374.81	-2278.22
10	-2701.09	-2234.87	-2156.36
All	-28015.96	-22786.15	-21912.53

(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 = 47.608852]

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	125446/277247	0.45247
Θ_2	129594/278118	0.46597
Θ_3	114514/277667	0.41241
$M_{1 \rightarrow 2}$	129549/277653	0.46659
$M_{2 \rightarrow 3}$	164587/277794	0.59248
$\Delta_{1 \rightarrow 2}$	193619/278054	0.69634
$\sigma_{1 \rightarrow 2}$	205567/277048	0.74199
$\Delta_{2 \rightarrow 3}$	234581/277060	0.84668
$\sigma_{2 \rightarrow 3}$	235030/278472	0.84400
Genealogies	386801/2500887	0.15467

MCMC-Autocorrelation and Effective MCMC Sample Size

Parameter	Autocorrelation	Effective Sample Size
Θ_1	0.65755	14777.38
Θ_2	0.59290	18439.58
Θ_3	0.41245	28144.76
$M_{1 \rightarrow 2}$	0.57603	19172.42
$M_{2 \rightarrow 3}$	0.72118	11009.40
$\Delta_{1 \rightarrow 2}$	0.44298	29890.23
$\sigma_{1 \rightarrow 2}$	0.49067	27109.85
$\Delta_{2 \rightarrow 3}$	0.26641	40227.31
$\sigma_{2 \rightarrow 3}$	0.24681	41958.11
Genealogies	0.24681	41958.11

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 13 (Locus 2): Upper prior boundary seems too low!

Param 13 (Locus 4): Upper prior boundary seems too low!