

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 Sat Jun 1 23:10:17 2019

Program finished at Sat Jun 1 23:12:23 2019 [Runtime:0000:00:02:06]



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) 938725822

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])	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_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	0.00813	0.00820	0.00830	0.00840	0.00840	0.01097	0.01290
1	Θ_2	0.00280	0.00393	0.00403	0.00407	0.00460	0.00430	0.00553
1	Θ_3	0.01667	0.01680	0.01690	0.01693	0.01733	0.01870	0.02001
1	$M_{1 \rightarrow 2}$	46.000	46.000	47.000	47.333	47.333	406.333	439.596
1	$M_{2 \rightarrow 3}$	0.000	1.333	2.333	4.000	4.000	19.000	33.803
1	$D_{1 \rightarrow 2}$	0.02453	0.02453	0.02470	0.02473	0.02473	0.03797	0.04300
1	$S_{1 \rightarrow 2}$	0.00020	0.00027	0.00037	0.00060	0.00060	0.04550	0.04694
1	$D_{2 \rightarrow 3}$	0.03873	0.03880	0.03890	0.03893	0.03920	0.05163	0.05254
1	$S_{2 \rightarrow 3}$	0.03273	0.03280	0.03290	0.03293	0.03293	0.05597	0.05515
2	Θ_1	0.00753	0.00880	0.00897	0.00900	0.00947	0.00917	0.01009
2	Θ_2	0.00160	0.00373	0.00397	0.00413	0.00700	0.00470	0.00629
2	Θ_3	0.00207	0.00340	0.00350	0.00353	0.00893	0.00530	0.00586
2	$M_{1 \rightarrow 2}$	51.333	51.333	52.333	52.667	53.333	250.333	301.550
2	$M_{2 \rightarrow 3}$	28.000	33.333	35.000	35.333	38.000	49.667	85.710
2	$D_{1 \rightarrow 2}$	0.01193	0.01260	0.01270	0.01273	0.01273	0.03770	0.04073
2	$S_{1 \rightarrow 2}$	0.00007	0.00007	0.00017	0.00020	0.00040	0.05223	0.05186
2	$D_{2 \rightarrow 3}$	0.03953	0.03967	0.03977	0.03980	0.03980	0.04910	0.04976
2	$S_{2 \rightarrow 3}$	0.02660	0.02660	0.02670	0.02673	0.02673	0.05623	0.05495
3	Θ_1	0.00420	0.00733	0.00770	0.00773	0.01133	0.00783	0.00820
3	Θ_2	0.00733	0.00733	0.00743	0.00747	0.00747	0.01703	0.02527
3	Θ_3	0.00853	0.00853	0.00870	0.00873	0.01033	0.01303	0.01493
3	$M_{1 \rightarrow 2}$	278.000	279.333	280.333	280.667	283.333	513.667	507.440
3	$M_{2 \rightarrow 3}$	0.000	13.333	16.333	16.667	26.667	19.667	27.209
3	$D_{1 \rightarrow 2}$	0.00000	0.00007	0.00017	0.00033	0.00447	0.01770	0.02458
3	$S_{1 \rightarrow 2}$	0.00073	0.00073	0.00090	0.00100	0.00107	0.03490	0.04035
3	$D_{2 \rightarrow 3}$	0.04060	0.04060	0.04070	0.04087	0.04113	0.04463	0.04663
3	$S_{2 \rightarrow 3}$	0.04733	0.04740	0.04750	0.04753	0.04773	0.06130	0.05903
4	Θ_1	0.00280	0.00413	0.00437	0.00440	0.00840	0.00650	0.00698
4	Θ_2	0.00253	0.00393	0.00403	0.00413	0.00447	0.00723	0.01001
4	Θ_3	0.00900	0.00920	0.00930	0.00933	0.00940	0.01617	0.01781
4	$M_{1 \rightarrow 2}$	22.000	22.000	23.000	25.333	25.333	248.333	297.450
4	$M_{2 \rightarrow 3}$	0.000	0.000	0.333	8.667	52.667	17.667	30.851
4	$D_{1 \rightarrow 2}$	0.00067	0.00067	0.00083	0.00087	0.00087	0.01610	0.02643

Locus	Parameter	2.5%	25.0%	Mode	75.0%	97.5%	Median	Mean
4	$S_{1 \rightarrow 2}$	0.00080	0.00080	0.00090	0.00093	0.00093	0.03843	0.04269
4	$D_{2 \rightarrow 3}$	0.05067	0.05067	0.05077	0.05087	0.05087	0.05563	0.05459
4	$S_{2 \rightarrow 3}$	0.04240	0.04240	0.04250	0.04253	0.04253	0.05977	0.05767
5	Θ_1	0.00167	0.00267	0.00330	0.00373	0.00667	0.00470	0.00630
5	Θ_2	0.00200	0.00200	0.00223	0.00227	0.00227	0.00770	0.00976
5	Θ_3	0.02253	0.02273	0.02297	0.02300	0.02333	0.02137	0.02361
5	$M_{1 \rightarrow 2}$	90.667	90.667	92.333	92.667	92.667	225.000	345.781
5	$M_{2 \rightarrow 3}$	0.000	0.000	0.333	0.667	70.000	30.333	39.421
5	$D_{1 \rightarrow 2}$	0.05660	0.05893	0.05903	0.05907	0.05933	0.05457	0.05496
5	$S_{1 \rightarrow 2}$	0.08520	0.08533	0.08543	0.08547	0.08547	0.06877	0.06705
5	$D_{2 \rightarrow 3}$	0.05140	0.05140	0.05157	0.05160	0.05180	0.04583	0.04762
5	$S_{2 \rightarrow 3}$	0.08953	0.08993	0.09003	0.09007	0.09013	0.05723	0.05587
6	Θ_1	0.00307	0.00333	0.00350	0.00367	0.00407	0.00577	0.00624
6	Θ_2	0.00180	0.00187	0.00197	0.00200	0.00213	0.00843	0.01449
6	Θ_3	0.03080	0.03087	0.03097	0.03120	0.03127	0.03690	0.03926
6	$M_{1 \rightarrow 2}$	576.667	578.667	580.333	580.667	580.667	669.667	643.075
6	$M_{2 \rightarrow 3}$	0.000	10.000	11.000	12.667	14.000	21.667	25.898
6	$D_{1 \rightarrow 2}$	0.00040	0.00053	0.00123	0.00133	0.00133	0.03790	0.04178
6	$S_{1 \rightarrow 2}$	0.00767	0.00780	0.00790	0.00793	0.00793	0.04870	0.05031
6	$D_{2 \rightarrow 3}$	0.03027	0.03033	0.03043	0.03053	0.03060	0.04803	0.04904
6	$S_{2 \rightarrow 3}$	0.00167	0.00167	0.00177	0.00180	0.00180	0.05283	0.05217
7	Θ_1	0.00980	0.01087	0.01097	0.01107	0.01373	0.01103	0.01138
7	Θ_2	0.00927	0.01013	0.01037	0.01040	0.01133	0.01403	0.02101
7	Θ_3	0.01113	0.01140	0.01177	0.01187	0.01320	0.01537	0.01766
7	$M_{1 \rightarrow 2}$	28.667	29.333	30.333	31.333	34.000	133.667	205.932
7	$M_{2 \rightarrow 3}$	0.000	0.000	1.667	2.000	72.000	45.667	70.495
7	$D_{1 \rightarrow 2}$	0.00053	0.00053	0.00077	0.00080	0.00200	0.00363	0.00748
7	$S_{1 \rightarrow 2}$	0.00040	0.00040	0.00050	0.00053	0.00053	0.01597	0.02309
7	$D_{2 \rightarrow 3}$	0.03307	0.03307	0.03317	0.03320	0.03327	0.04590	0.04718
7	$S_{2 \rightarrow 3}$	0.06127	0.06127	0.06137	0.06140	0.06140	0.05283	0.05259
8	Θ_1	0.00407	0.00647	0.00697	0.00700	0.00987	0.00757	0.00856
8	Θ_2	0.00553	0.00553	0.00563	0.00573	0.00573	0.01357	0.01579
8	Θ_3	0.00827	0.00867	0.00877	0.00887	0.00907	0.01170	0.01304
8	$M_{1 \rightarrow 2}$	0.000	28.667	31.000	32.000	76.667	56.333	108.628
8	$M_{2 \rightarrow 3}$	6.667	6.667	8.333	9.333	22.000	26.333	50.148
8	$D_{1 \rightarrow 2}$	0.00007	0.00140	0.00170	0.00187	0.00187	0.01243	0.02234
8	$S_{1 \rightarrow 2}$	0.00073	0.00073	0.00097	0.00100	0.00100	0.04490	0.04589

Locus	Parameter	2.5%	25.0%	Mode	75.0%	97.5%	Median	Mean
8	D _{2->3}	0.01460	0.01500	0.01510	0.01527	0.01527	0.03663	0.04094
8	S _{2->3}	0.02480	0.02480	0.02490	0.02493	0.02507	0.05763	0.05578
9	Θ_1	0.01093	0.01233	0.01243	0.01247	0.01387	0.01370	0.01498
9	Θ_2	0.00127	0.00300	0.00323	0.00327	0.00840	0.00710	0.00931
9	Θ_3	0.01767	0.01800	0.01810	0.01813	0.01853	0.02063	0.02378
9	M _{1->2}	35.333	36.000	37.000	38.000	38.000	61.667	145.245
9	M _{2->3}	0.000	0.000	3.667	5.333	62.000	87.000	116.705
9	D _{1->2}	0.04220	0.04240	0.04270	0.04273	0.04273	0.03983	0.04273
9	S _{1->2}	0.05153	0.05153	0.05163	0.05173	0.05173	0.06030	0.05915
9	D _{2->3}	0.01367	0.01387	0.01397	0.01400	0.01400	0.03530	0.04022
9	S _{2->3}	0.00467	0.00473	0.00483	0.00487	0.00487	0.04943	0.04942
10	Θ_1	0.00260	0.00540	0.00583	0.00587	0.01033	0.00657	0.00793
10	Θ_2	0.01293	0.01293	0.01303	0.01320	0.01387	0.01410	0.01550
10	Θ_3	0.01313	0.01413	0.01423	0.01440	0.01533	0.01997	0.02563
10	M _{1->2}	6.000	52.000	55.000	55.333	64.000	53.667	74.597
10	M _{2->3}	408.667	408.667	409.667	410.667	410.667	384.333	439.474
10	D _{1->2}	0.00593	0.00600	0.00617	0.00620	0.00627	0.01570	0.02318
10	S _{1->2}	0.02807	0.02813	0.02830	0.02833	0.02833	0.05217	0.05286
10	D _{2->3}	0.00000	0.00033	0.00050	0.00053	0.00113	0.02530	0.03378
10	S _{2->3}	0.00427	0.00427	0.00443	0.00467	0.00467	0.04010	0.04299
All	Θ_1	0.00513	0.00627	0.00697	0.00760	0.00893	0.00710	0.00703
All	Θ_2	0.00320	0.00420	0.00490	0.00573	0.00780	0.00523	0.00536
All	Θ_3	0.01027	0.01227	0.01357	0.01400	0.01767	0.01370	0.01395
All	M _{1->2}	19.333	40.000	51.667	62.000	146.667	54.333	59.103
All	M _{2->3}	0.000	3.333	8.333	13.333	24.000	11.000	10.955
All	D _{1->2}	0.00000	0.00080	0.00157	0.00220	0.00427	0.00177	0.00192
All	S _{1->2}	0.03200	0.03513	0.03783	0.04020	0.04573	0.02770	0.02732
All	D _{2->3}	0.01087	0.02493	0.02950	0.03267	0.04093	0.02523	0.02392
All	S _{2->3}	0.02260	0.04120	0.04323	0.04827	0.04907	0.04110	0.03807

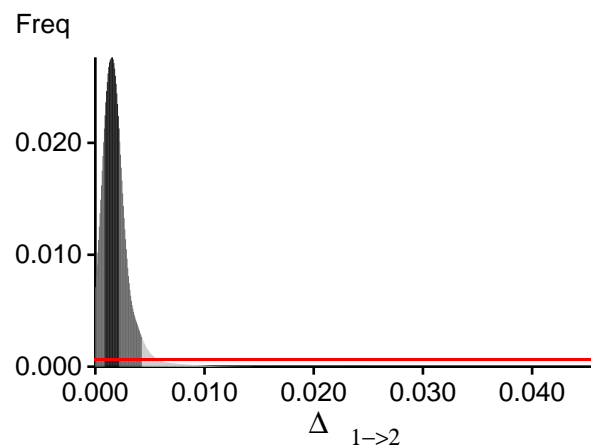
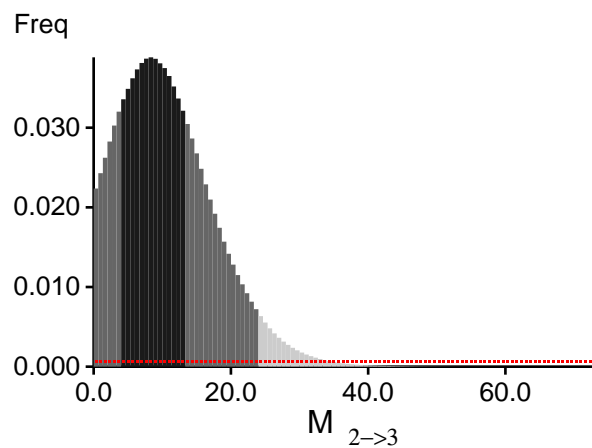
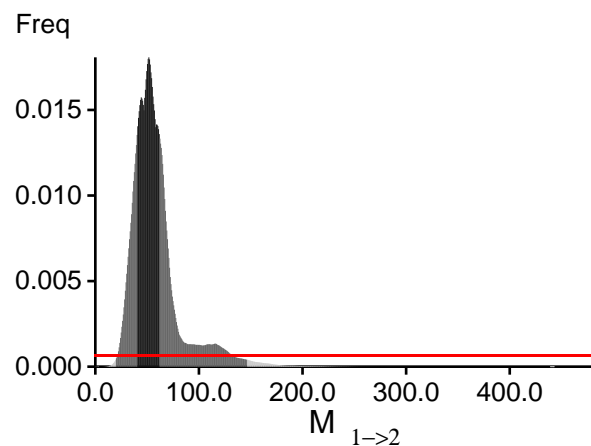
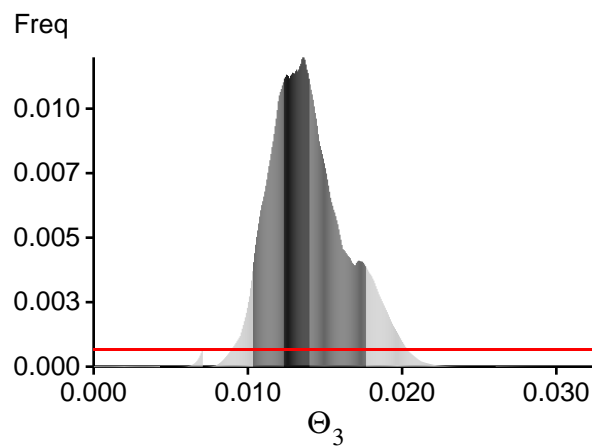
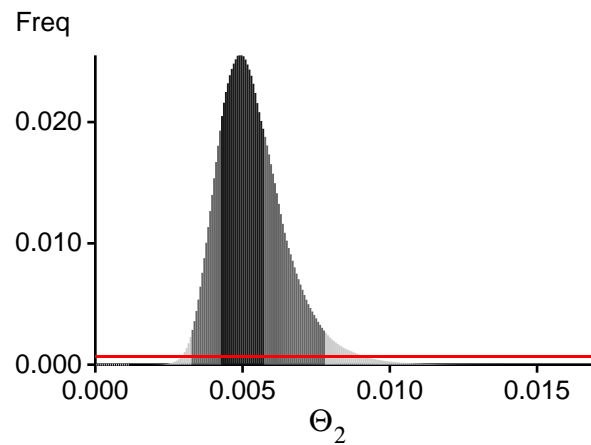
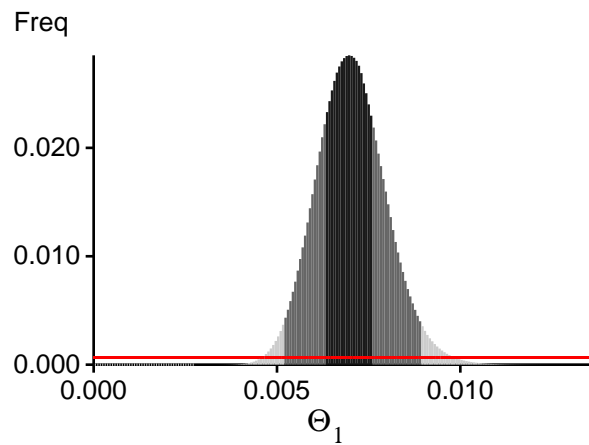
Citation suggestions:

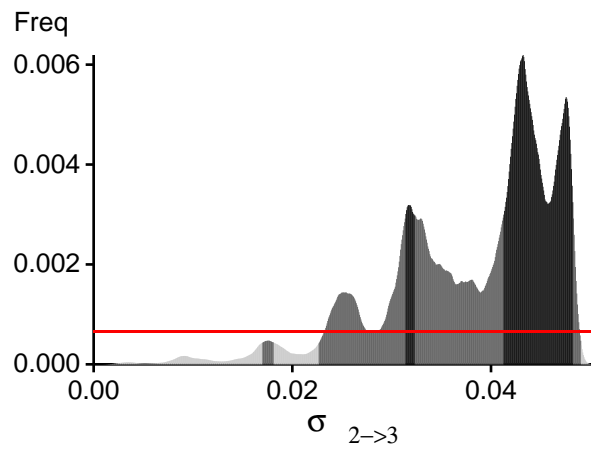
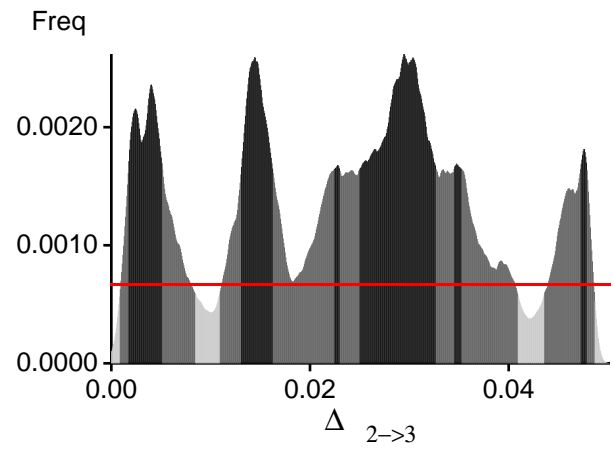
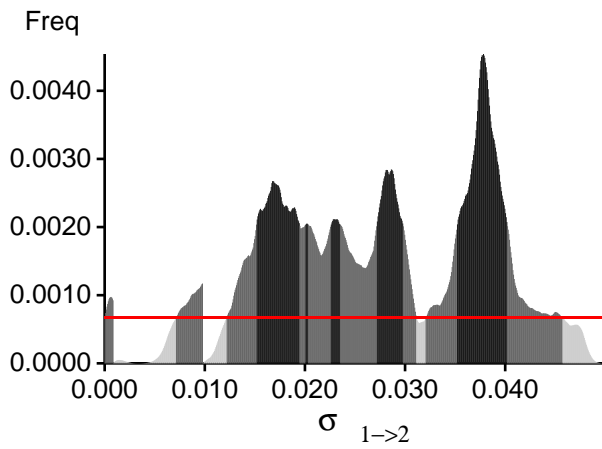
Beerli P., 2006. Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. *Bioinformatics* 22:341-345

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	-2942.17	-2387.88	-2294.54
2	-2494.51	-2038.41	-1959.69
3	-2690.40	-2194.29	-2113.67
4	-3057.58	-2360.25	-2238.50
5	-2771.11	-2266.37	-2180.98
6	-2787.65	-2369.88	-2299.71
7	-2758.31	-2282.16	-2203.03
8	-2833.71	-2315.43	-2229.25
9	-2869.00	-2358.58	-2281.10
10	-2672.20	-2229.00	-2153.60
All	-27835.24	-22760.84	-21912.69

(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 = 41.396917]

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	11902/28005	0.42500
Θ_2	10905/27877	0.39118
Θ_3	11057/28003	0.39485
$M_{1 \rightarrow 2}$	14598/27841	0.52433
$M_{2 \rightarrow 3}$	12771/27954	0.45686
$\Delta_{1 \rightarrow 2}$	19943/27730	0.71918
$\sigma_{1 \rightarrow 2}$	20709/27685	0.74802
$\Delta_{2 \rightarrow 3}$	23899/27741	0.86150
$\sigma_{2 \rightarrow 3}$	23636/27609	0.85610
Genealogies	38768/249555	0.15535

MCMC-Autocorrelation and Effective MCMC Sample Size

Parameter	Autocorrelation	Effective Sample Size
Θ_1	0.91989	3005.35
Θ_2	0.91443	2933.08
Θ_3	0.85439	5315.02
$M_{1 \rightarrow 2}$	0.91636	3185.18
$M_{2 \rightarrow 3}$	0.93108	2364.30
$\Delta_{1 \rightarrow 2}$	0.89463	4525.49
$\sigma_{1 \rightarrow 2}$	0.89909	4362.76
$\Delta_{2 \rightarrow 3}$	0.83932	6014.85
$\sigma_{2 \rightarrow 3}$	0.83519	6177.52
Genealogies	0.83519	6177.52

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 5): Upper prior boundary seems too low!