

# 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)

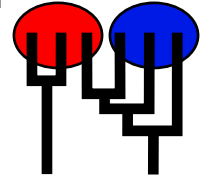
Compiled for PARALLEL computer architectures

One master and 8 compute nodes are available.

Compiled for a SYMMETRIC multiprocessors (Grandcentral)

Program started at Mon Jun 3 08:51:15 2019

Program finished at Mon Jun 3 08:53:19 2019 [Runtime:0000:00:02:04]



## 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) 2222964635

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	$\Theta_1$	<displayed>
2	$\Theta_2$	<displayed>
3	$\Theta_3$	<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	2500.	5000.	500.0	1500	0.05556
5	M **	Uniform	0.000000	2500.	5000.	500.0	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	1000000.00	4 chains with temperatures	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	$\Theta_1$	0.00800	0.00967	0.00977	0.00987	0.00987	0.00983	0.01115
1	$\Theta_2$	0.00607	0.00680	0.00690	0.00693	0.00707	0.00977	0.01051
1	$\Theta_3$	0.01580	0.01580	0.01590	0.01593	0.01620	0.01383	0.01646
1	$M_{1 \rightarrow 2}$	0.000	3.333	28.333	30.000	106.667	41.667	49.394
1	$M_{2 \rightarrow 3}$	0.000	0.000	8.333	13.333	50.000	128.333	226.599
1	$D_{1 \rightarrow 2}$	0.05733	0.05753	0.05763	0.05767	0.05767	0.04337	0.04545
1	$S_{1 \rightarrow 2}$	0.07607	0.07627	0.07637	0.07640	0.07653	0.06443	0.06181
1	$D_{2 \rightarrow 3}$	0.00180	0.00180	0.00190	0.00193	0.00207	0.03810	0.04344
1	$S_{2 \rightarrow 3}$	0.00693	0.00693	0.00703	0.00707	0.00707	0.05550	0.05400
2	$\Theta_1$	0.00640	0.01053	0.01063	0.01067	0.01147	0.00917	0.00979
2	$\Theta_2$	0.00107	0.00253	0.00263	0.00300	0.00633	0.00437	0.00508
2	$\Theta_3$	0.00207	0.00467	0.00503	0.00513	0.00827	0.00577	0.00686
2	$M_{1 \rightarrow 2}$	96.667	96.667	105.000	106.667	123.333	338.333	426.182
2	$M_{2 \rightarrow 3}$	0.000	0.000	18.333	40.000	143.333	48.333	91.730
2	$D_{1 \rightarrow 2}$	0.03753	0.03773	0.03783	0.03787	0.03813	0.04097	0.04387
2	$S_{1 \rightarrow 2}$	0.04060	0.04060	0.04070	0.04073	0.04073	0.05443	0.05377
2	$D_{2 \rightarrow 3}$	0.02033	0.02033	0.02043	0.02047	0.02073	0.04770	0.04930
2	$S_{2 \rightarrow 3}$	0.05187	0.05207	0.05217	0.05220	0.05227	0.05330	0.05286
3	$\Theta_1$	0.00280	0.00740	0.00763	0.00767	0.00833	0.00770	0.00831
3	$\Theta_2$	0.00420	0.00433	0.00457	0.00467	0.00467	0.01357	0.02035
3	$\Theta_3$	0.01093	0.01093	0.01103	0.01107	0.01107	0.01450	0.01636
3	$M_{1 \rightarrow 2}$	53.333	63.333	75.000	80.000	93.333	288.333	414.188
3	$M_{2 \rightarrow 3}$	0.000	0.000	1.667	20.000	76.667	21.667	33.678
3	$D_{1 \rightarrow 2}$	0.00000	0.00007	0.00037	0.00107	0.00127	0.00490	0.01276
3	$S_{1 \rightarrow 2}$	0.00060	0.00060	0.00090	0.00093	0.00093	0.01777	0.02787
3	$D_{2 \rightarrow 3}$	0.03433	0.03433	0.03450	0.03460	0.03460	0.04483	0.04684
3	$S_{2 \rightarrow 3}$	0.05993	0.06000	0.06010	0.06013	0.06013	0.05790	0.05592
4	$\Theta_1$	0.00487	0.00487	0.00503	0.00507	0.00513	0.00670	0.00701
4	$\Theta_2$	0.00107	0.00353	0.00383	0.00433	0.00473	0.00677	0.00856
4	$\Theta_3$	0.00953	0.01120	0.01130	0.01147	0.01187	0.01183	0.01346
4	$M_{1 \rightarrow 2}$	50.000	50.000	55.000	56.667	60.000	268.333	331.430
4	$M_{2 \rightarrow 3}$	0.000	0.000	5.000	13.333	26.667	15.000	31.462
4	$D_{1 \rightarrow 2}$	0.00200	0.00200	0.00223	0.00227	0.00240	0.01903	0.02847

Locus	Parameter	2.5%	25.0%	Mode	75.0%	97.5%	Median	Mean
4	$S_{1 \rightarrow 2}$	0.00087	0.00087	0.00097	0.00107	0.00107	0.04637	0.04761
4	$D_{2 \rightarrow 3}$	0.00420	0.00467	0.00477	0.00487	0.00500	0.03777	0.04147
4	$S_{2 \rightarrow 3}$	0.06400	0.06420	0.06430	0.06447	0.06447	0.06003	0.05874
5	$\Theta_1$	0.00173	0.00460	0.00510	0.00540	0.00820	0.00510	0.00614
5	$\Theta_2$	0.00373	0.00413	0.00423	0.00433	0.00473	0.00717	0.01054
5	$\Theta_3$	0.01713	0.01920	0.01943	0.01947	0.01953	0.02250	0.02471
5	$M_{1 \rightarrow 2}$	40.000	40.000	45.000	46.667	50.000	578.333	843.785
5	$M_{2 \rightarrow 3}$	0.000	0.000	1.667	16.667	76.667	21.667	26.692
5	$D_{1 \rightarrow 2}$	0.00173	0.00173	0.00183	0.00213	0.00213	0.03697	0.04121
5	$S_{1 \rightarrow 2}$	0.01427	0.01433	0.01443	0.01453	0.01453	0.04890	0.04963
5	$D_{2 \rightarrow 3}$	0.03087	0.03147	0.03157	0.03160	0.03173	0.04757	0.04960
5	$S_{2 \rightarrow 3}$	0.06387	0.06387	0.06397	0.06400	0.06407	0.05423	0.05464
6	$\Theta_1$	0.00320	0.00320	0.00337	0.00340	0.00380	0.00837	0.00953
6	$\Theta_2$	0.00267	0.00373	0.00390	0.00400	0.00407	0.00903	0.01325
6	$\Theta_3$	0.01247	0.01247	0.01257	0.01267	0.01320	0.02617	0.02980
6	$M_{1 \rightarrow 2}$	50.000	66.667	91.667	93.333	183.333	515.000	792.282
6	$M_{2 \rightarrow 3}$	0.000	0.000	15.000	23.333	106.667	28.333	37.892
6	$D_{1 \rightarrow 2}$	0.00000	0.00047	0.00070	0.00073	0.00093	0.02290	0.03148
6	$S_{1 \rightarrow 2}$	0.00233	0.00233	0.00243	0.00247	0.00253	0.04630	0.04765
6	$D_{2 \rightarrow 3}$	0.03927	0.03927	0.03937	0.03940	0.03947	0.03617	0.04054
6	$S_{2 \rightarrow 3}$	0.05313	0.05367	0.05377	0.05380	0.05380	0.05477	0.05459
7	$\Theta_1$	0.00740	0.00953	0.00977	0.00980	0.01547	0.01210	0.01285
7	$\Theta_2$	0.00380	0.00440	0.00463	0.00500	0.00500	0.00970	0.01580
7	$\Theta_3$	0.01360	0.01400	0.01410	0.01420	0.01453	0.01770	0.01929
7	$M_{1 \rightarrow 2}$	110.000	130.000	135.000	150.000	150.000	368.333	500.989
7	$M_{2 \rightarrow 3}$	0.000	0.000	1.667	26.667	116.667	35.000	46.288
7	$D_{1 \rightarrow 2}$	0.00000	0.00060	0.00103	0.00140	0.00140	0.00883	0.02123
7	$S_{1 \rightarrow 2}$	0.00067	0.00067	0.00077	0.00080	0.00080	0.02843	0.03858
7	$D_{2 \rightarrow 3}$	0.00627	0.00627	0.00637	0.00640	0.00660	0.04117	0.04449
7	$S_{2 \rightarrow 3}$	0.03820	0.03820	0.03830	0.03833	0.03833	0.05617	0.05362
8	$\Theta_1$	0.00500	0.00607	0.00623	0.00627	0.00853	0.00770	0.00842
8	$\Theta_2$	0.00520	0.00520	0.00537	0.00540	0.00540	0.01057	0.01436
8	$\Theta_3$	0.00547	0.00547	0.00557	0.00560	0.00567	0.01403	0.01592
8	$M_{1 \rightarrow 2}$	0.000	0.000	18.333	56.667	213.333	221.667	305.112
8	$M_{2 \rightarrow 3}$	0.000	0.000	1.667	16.667	143.333	18.333	73.656
8	$D_{1 \rightarrow 2}$	0.00447	0.00493	0.00503	0.00507	0.00507	0.02483	0.03215
8	$S_{1 \rightarrow 2}$	0.00627	0.00627	0.00650	0.00660	0.00667	0.04757	0.04807



Locus	Parameter	2.5%	25.0%	Mode	75.0%	97.5%	Median	Mean
8	$D_{2 \rightarrow 3}$	0.01020	0.01027	0.01037	0.01040	0.01040	0.05070	0.05055
8	$S_{2 \rightarrow 3}$	0.03033	0.03040	0.03050	0.03060	0.03067	0.05770	0.05648
9	$\Theta_1$	0.01727	0.01747	0.01757	0.01760	0.01767	0.01383	0.01540
9	$\Theta_2$	0.00193	0.00233	0.00263	0.00280	0.00660	0.00577	0.00666
9	$\Theta_3$	0.02713	0.02733	0.02743	0.02747	0.02753	0.03263	0.04054
9	$M_{1 \rightarrow 2}$	0.000	13.333	48.333	53.333	146.667	55.000	92.664
9	$M_{2 \rightarrow 3}$	26.667	36.667	45.000	50.000	56.667	231.667	318.213
9	$D_{1 \rightarrow 2}$	0.00280	0.00287	0.00297	0.00300	0.00307	0.03650	0.04071
9	$S_{1 \rightarrow 2}$	0.00767	0.00767	0.00777	0.00780	0.00787	0.06077	0.05929
9	$D_{2 \rightarrow 3}$	0.00227	0.00233	0.00243	0.00267	0.00267	0.02477	0.03409
9	$S_{2 \rightarrow 3}$	0.00427	0.00440	0.00450	0.00467	0.00473	0.03837	0.04148
10	$\Theta_1$	0.00260	0.00367	0.00437	0.00440	0.00780	0.00763	0.00928
10	$\Theta_2$	0.00347	0.00693	0.00730	0.00733	0.01153	0.00950	0.01041
10	$\Theta_3$	0.01280	0.01287	0.01297	0.01300	0.01300	0.02090	0.02481
10	$M_{1 \rightarrow 2}$	20.000	23.333	31.667	33.333	40.000	358.333	563.840
10	$M_{2 \rightarrow 3}$	0.000	0.000	35.000	63.333	96.667	88.333	189.157
10	$D_{1 \rightarrow 2}$	0.00000	0.00020	0.00030	0.00047	0.00073	0.02783	0.03437
10	$S_{1 \rightarrow 2}$	0.00533	0.00533	0.00543	0.00553	0.00553	0.04410	0.04653
10	$D_{2 \rightarrow 3}$	0.01000	0.01000	0.01010	0.01013	0.01013	0.04543	0.04710
10	$S_{2 \rightarrow 3}$	0.07340	0.07340	0.07350	0.07353	0.07353	0.05783	0.05536
All	$\Theta_1$	0.00520	0.00647	0.00743	0.00820	0.00987	0.00757	0.00756
All	$\Theta_2$	0.00333	0.00440	0.00510	0.00593	0.00793	0.00543	0.00552
All	$\Theta_3$	0.01000	0.01093	0.01250	0.01360	0.01560	0.01283	0.01311
All	$M_{1 \rightarrow 2}$	10.000	53.333	78.333	103.333	143.333	81.667	79.241
All	$M_{2 \rightarrow 3}$	0.000	0.000	1.667	13.333	40.000	15.000	14.563
All	$D_{1 \rightarrow 2}$	0.00000	0.00073	0.00143	0.00213	0.00440	0.00177	0.00196
All	$S_{1 \rightarrow 2}$	0.00500	0.00633	0.00817	0.00933	0.01253	0.00897	0.01573
All	$D_{2 \rightarrow 3}$	0.00000	0.00000	0.00077	0.00760	0.03713	0.01150	0.01479
All	$S_{2 \rightarrow 3}$	0.02540	0.04520	0.04763	0.04847	0.04907	0.03823	0.03763

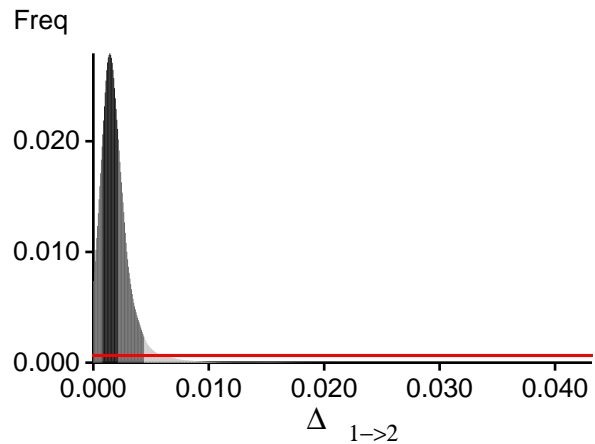
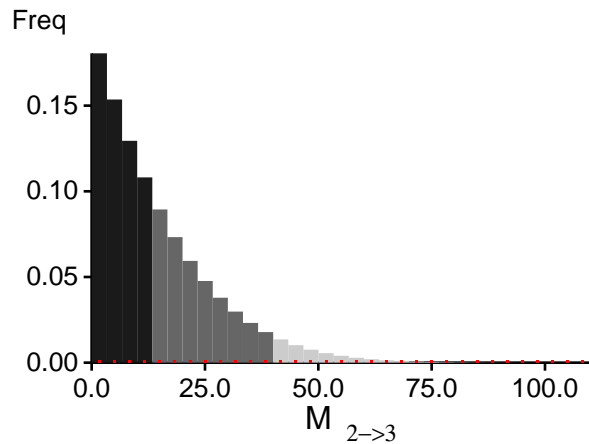
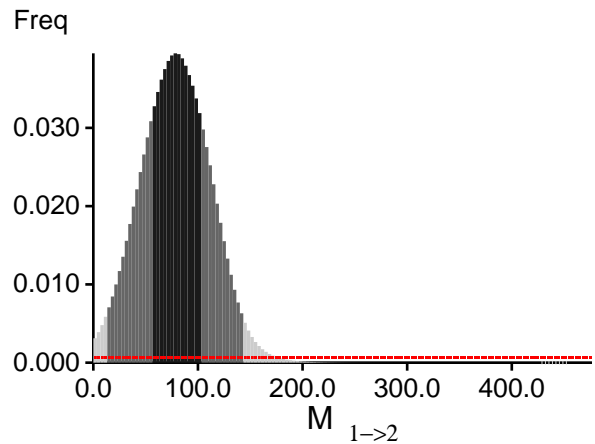
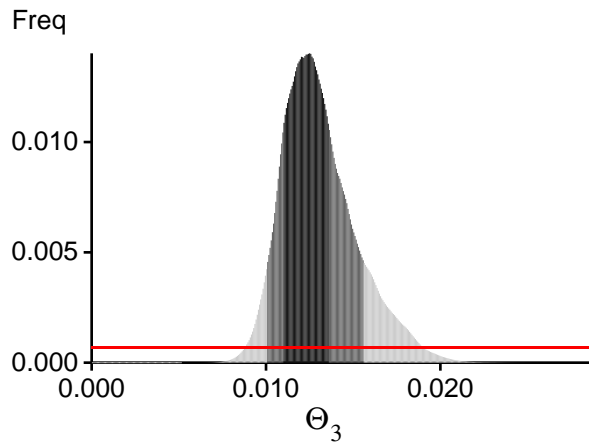
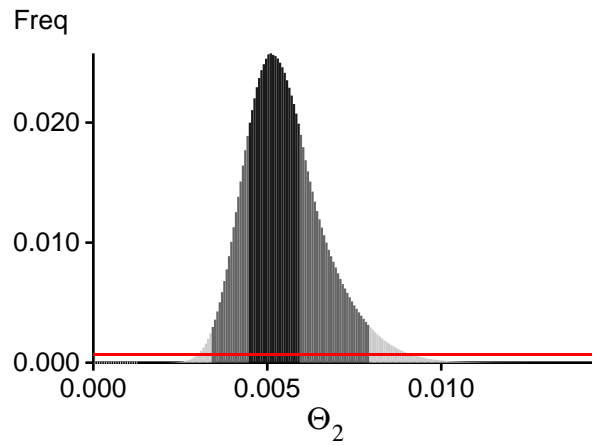
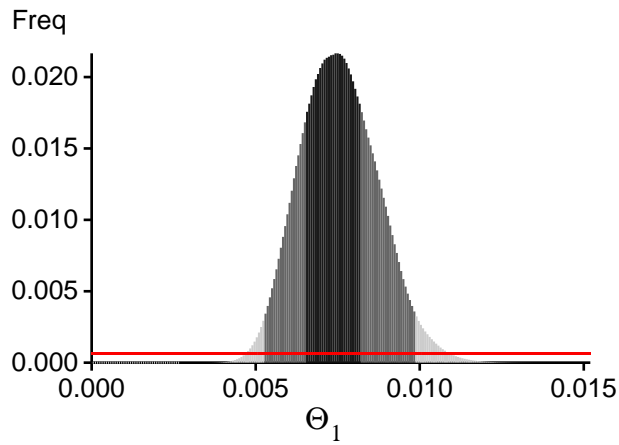
## Citation suggestions:

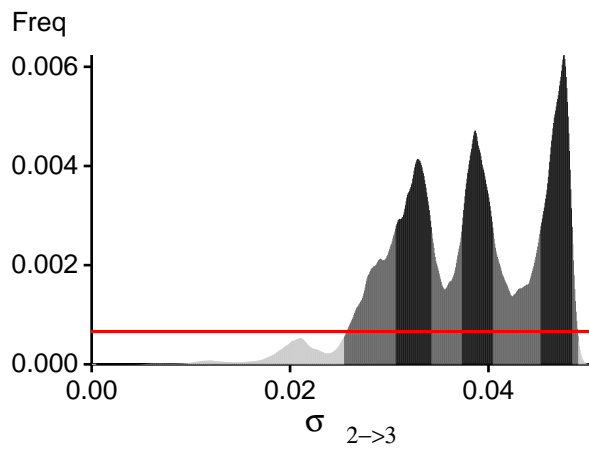
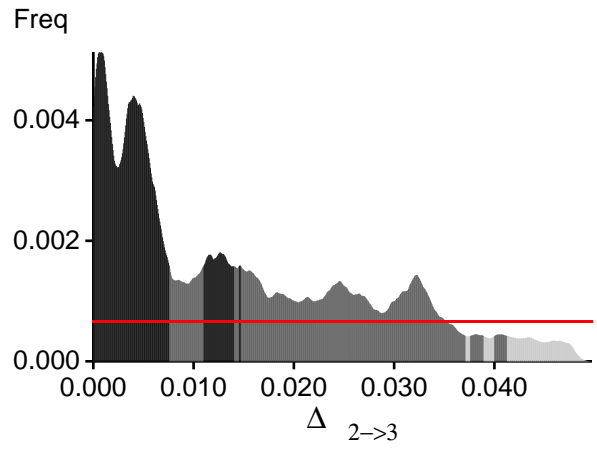
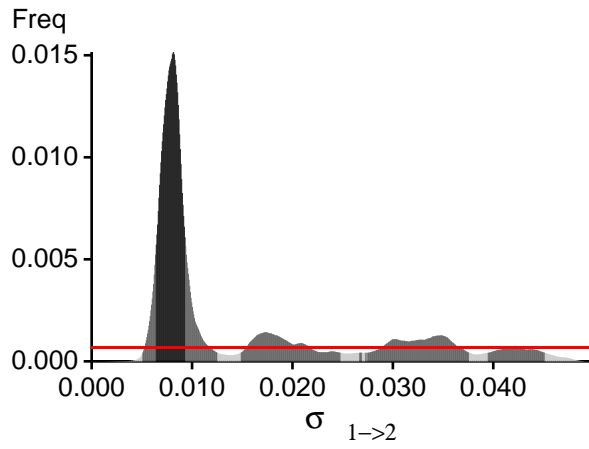
Beerli P., 2006. Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. *Bioinformatics* 22:341-345

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	-3227.99	-2436.44	-2294.64
2	-2652.05	-2065.89	-1962.14
3	-2812.02	-2213.92	-2116.81
4	-3163.16	-2373.49	-2235.77
5	-3009.80	-2302.46	-2178.33
6	-2934.93	-2393.07	-2302.65
7	-2852.94	-2293.68	-2202.00
8	-3211.20	-2375.83	-2233.85
9	-3241.03	-2414.88	-2275.83
10	-2832.77	-2254.70	-2153.21
All	-29866.18	-23052.66	-21883.51

(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 = 71.711405]

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
$\Theta_1$	11206/27918	0.40139
$\Theta_2$	12365/27781	0.44509
$\Theta_3$	11837/27841	0.42516
$M_{1 \rightarrow 2}$	10254/27672	0.37056
$M_{2 \rightarrow 3}$	11987/27656	0.43343
$\Delta_{1 \rightarrow 2}$	20467/27642	0.74043
$\sigma_{1 \rightarrow 2}$	21240/27662	0.76784
$\Delta_{2 \rightarrow 3}$	24024/27860	0.86231
$\sigma_{2 \rightarrow 3}$	23442/27543	0.85111
Genealogies	38782/250425	0.15486

## *MCMC-Autocorrelation and Effective MCMC Sample Size*

Parameter	Autocorrelation	Effective Sampe Size
$\Theta_1$	0.92183	2953.36
$\Theta_2$	0.91924	2834.93
$\Theta_3$	0.87922	4239.07
$M_{1 \rightarrow 2}$	0.93007	2388.87
$M_{2 \rightarrow 3}$	0.95129	1694.53
$\Delta_{1 \rightarrow 2}$	0.87672	4832.76
$\sigma_{1 \rightarrow 2}$	0.89643	4265.97
$\Delta_{2 \rightarrow 3}$	0.84319	5961.36
$\sigma_{2 \rightarrow 3}$	0.83000	6402.70
Genealogies	0.83000	6402.70

## *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