

# simdata

POPULATION SIZE, MIGRATION, DIVERGENCE, ASSIGNMENT, HISTORY

Bayesian inference using the structured coalescent

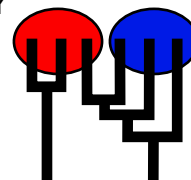
Migrate-n version 5.1.1(git:v5.0.2-50-gc469c7c) [March-31-2024]

Compiled for PARALLEL computer architectures

One master and 10 compute nodes are available.

Program started at Sun Mar 31 18:01:12 2024

Program finished at Sun Mar 31 18:02:48 2024 [Runtime:0000:00:01:36]



## Options

Inheritance multipliers in use for Thetas:

All loci use an inheritance multiplier of 1.0

Random number seed:

(from parmfile)

427203015

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 Romanshorn_0	*	0	0
2 Arbon_1	d	*	0
3 Kreuzlingen_2	0	d	*

Order of parameters:

1

$\Theta_1$

<displayed>

2	$\Theta_2$	<displayed>
3	$\Theta_3$	<displayed>
4	$\Delta_{1 \rightarrow 2}$	<displayed>
5	$\sigma_{1 \rightarrow 2}$	<displayed>
6	$\Delta_{2 \rightarrow 3}$	<displayed>
7	$\sigma_{2 \rightarrow 3}$	<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 (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	0.010000	1500	0.07143
2 Theta	* *	Uniform	0.000000	0.050	0.100	0.010000	1500	0.07143
3 Theta	* *	Uniform	0.000000	0.050	0.100	0.010000	1500	0.07143
4 Splittime mean	1 2	Gamma	0.000000	0.100	0.800	(a=1.50)	1500	0.07143
5 Splittime std	1 2	Gamma	0.000000	0.100	0.200	(a=1.50)	1500	0.07143
6 Splittime mean	2 3	Uniform	0.000000	0.005	0.010	0.000100	1500	0.07143
7 Splittime std	2 3	Uniform	0.000000	0.005	0.010	0.000100	1500	0.07143

[\* \* means priors were set globally]

Posterior distribution:

Parameter values were collected using MCMC, these values

were then used to generate the posterior histograms using BINNING

and subsequent MOVING AVERAGE SMOOTHING (window=41) for combination over loci

Markov chain settings:

Long chain

Number of chains

1

Recorded steps [a]	10000
Increment (record every x step [b])	100
Number of concurrent chains (replicates) [c]	1
Visited (sampled) parameter values [a*b*c]	1000000
Number of discard trees per chain (burn-in)	1000

## 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
	parmfile
Haplotyping is turned on:	NO
Output file:	outfile
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	500
2	500
3	500
4	500
5	500
6	500
7	500
8	500
9	500
10	500

### 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 Romanshorn_0			1		10		
			2		10		
			3		10		
			4		10		
			5		10		
			6		10		
			7		10		
			8		10		
			9		10		
			10		10		
2 Arbon_1			1		10		
			2		10		
			3		10		
			4		10		
			5		10		
			6		10		
			7		10		
			8		10		
			9		10		
			10		10		
3 Kreuzlingen_2			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	$\Theta_1$	0.00447	0.00940	0.01317	0.01813	0.03387	0.01577	0.01742
1	$\Theta_2$	0.00147	0.00507	0.00757	0.01053	0.01927	0.00883	0.00961
1	$\Theta_3$	0.00000	0.00353	0.00610	0.00940	0.02093	0.00783	0.00912
1	$D_{1 \rightarrow 2}$	0.00107	0.07040	0.11973	0.14613	0.19680	0.10480	0.10287
1	$S_{1 \rightarrow 2}$	0.00240	0.01520	0.04767	0.07853	0.17160	0.07007	0.07780
1	$D_{2 \rightarrow 3}$	0.00009	0.00136	0.00334	0.00482	0.00897	0.00394	0.00433
1	$S_{2 \rightarrow 3}$	0.00027	0.00033	0.00081	0.00206	0.00657	0.00472	0.00485
2	$\Theta_1$	0.00553	0.01053	0.01483	0.01993	0.03680	0.01757	0.01936
2	$\Theta_2$	0.00000	0.00180	0.00337	0.00500	0.00873	0.00390	0.00396
2	$\Theta_3$	0.00173	0.00560	0.00863	0.01253	0.02580	0.01077	0.01212
2	$D_{1 \rightarrow 2}$	0.00053	0.06560	0.11440	0.14240	0.19520	0.10160	0.10051
2	$S_{1 \rightarrow 2}$	0.00413	0.01760	0.05447	0.08187	0.17187	0.07180	0.07912
2	$D_{2 \rightarrow 3}$	0.00066	0.00284	0.00491	0.00695	0.00980	0.00509	0.00511
2	$S_{2 \rightarrow 3}$	0.00151	0.00717	0.00786	0.00899	0.00993	0.00573	0.00572
3	$\Theta_1$	0.00733	0.01333	0.01910	0.02520	0.04687	0.02223	0.02444
3	$\Theta_2$	0.00000	0.00213	0.00377	0.00560	0.00993	0.00443	0.00459
3	$\Theta_3$	0.00000	0.00293	0.00530	0.00833	0.02000	0.00697	0.00834
3	$D_{1 \rightarrow 2}$	0.00000	0.05333	0.09520	0.12213	0.17867	0.08880	0.08961
3	$S_{1 \rightarrow 2}$	0.00200	0.01600	0.04140	0.07667	0.16920	0.06767	0.07575
3	$D_{2 \rightarrow 3}$	0.00009	0.00149	0.00322	0.00499	0.00813	0.00394	0.00431
3	$S_{2 \rightarrow 3}$	0.00067	0.00082	0.00114	0.00193	0.00978	0.00506	0.00512
4	$\Theta_1$	0.00153	0.00527	0.00790	0.01133	0.02140	0.00957	0.01047
4	$\Theta_2$	0.00073	0.00400	0.00617	0.00873	0.01527	0.00710	0.00758
4	$\Theta_3$	0.00207	0.00680	0.01077	0.01593	0.03733	0.01383	0.01637
4	$D_{1 \rightarrow 2}$	0.00213	0.07573	0.12133	0.15307	0.20107	0.10853	0.10619
4	$S_{1 \rightarrow 2}$	0.00280	0.01907	0.04540	0.08280	0.17173	0.07100	0.07844
4	$D_{2 \rightarrow 3}$	0.00061	0.00247	0.00484	0.00660	0.00967	0.00483	0.00490
4	$S_{2 \rightarrow 3}$	0.00161	0.00599	0.00634	0.00669	0.00994	0.00594	0.00586
5	$\Theta_1$	0.00507	0.01087	0.01497	0.02060	0.03820	0.01790	0.01960
5	$\Theta_2$	0.00093	0.00427	0.00657	0.00933	0.01713	0.00770	0.00839
5	$\Theta_3$	0.00000	0.00173	0.00337	0.00513	0.00993	0.00417	0.00431
5	$D_{1 \rightarrow 2}$	0.00000	0.05760	0.10053	0.12640	0.18080	0.09200	0.09175

Locus	Parameter	2.5%	25.0%	Mode	75.0%	97.5%	Median	Mean
5	$S_{1 \rightarrow 2}$	0.00213	0.01507	0.04073	0.07653	0.16840	0.06753	0.07566
5	$D_{2 \rightarrow 3}$	0.00056	0.00215	0.00383	0.00603	0.00970	0.00464	0.00484
5	$S_{2 \rightarrow 3}$	0.00032	0.00044	0.00091	0.00421	0.00942	0.00485	0.00493
6	$\Theta_1$	0.00567	0.01080	0.01423	0.02053	0.03833	0.01803	0.01986
6	$\Theta_2$	0.00067	0.00400	0.00630	0.00893	0.01680	0.00743	0.00808
6	$\Theta_3$	0.00000	0.00147	0.00297	0.00460	0.00887	0.00363	0.00368
6	$D_{1 \rightarrow 2}$	0.00267	0.08693	0.15280	0.17920	0.22720	0.12400	0.12100
6	$S_{1 \rightarrow 2}$	0.00440	0.02600	0.04420	0.08080	0.17733	0.07833	0.08453
6	$D_{2 \rightarrow 3}$	0.00123	0.00363	0.00486	0.00758	0.00990	0.00545	0.00541
6	$S_{2 \rightarrow 3}$	0.00034	0.00066	0.00318	0.00371	0.00877	0.00476	0.00487
7	$\Theta_1$	0.00600	0.01127	0.01517	0.02087	0.03787	0.01837	0.02009
7	$\Theta_2$	0.00447	0.00907	0.01250	0.01687	0.02987	0.01463	0.01584
7	$\Theta_3$	0.00007	0.00293	0.00503	0.00740	0.01413	0.00610	0.00661
7	$D_{1 \rightarrow 2}$	0.00053	0.06613	0.11067	0.13600	0.18773	0.09840	0.09719
7	$S_{1 \rightarrow 2}$	0.00227	0.01373	0.03767	0.07480	0.16960	0.06673	0.07545
7	$D_{2 \rightarrow 3}$	0.00090	0.00297	0.00481	0.00694	0.00982	0.00510	0.00512
7	$S_{2 \rightarrow 3}$	0.00054	0.00273	0.00312	0.00463	0.00969	0.00490	0.00498
8	$\Theta_1$	0.00567	0.01107	0.01503	0.02093	0.03813	0.01830	0.02005
8	$\Theta_2$	0.00087	0.00400	0.00610	0.00847	0.01493	0.00697	0.00740
8	$\Theta_3$	0.00127	0.00547	0.00837	0.01227	0.02533	0.01043	0.01179
8	$D_{1 \rightarrow 2}$	0.00267	0.07253	0.12240	0.14827	0.19840	0.10587	0.10414
8	$S_{1 \rightarrow 2}$	0.00280	0.01653	0.03420	0.08200	0.17067	0.07220	0.07963
8	$D_{2 \rightarrow 3}$	0.00165	0.00421	0.00522	0.00760	0.00985	0.00538	0.00531
8	$S_{2 \rightarrow 3}$	0.00240	0.00713	0.00868	0.00965	0.00995	0.00635	0.00628
9	$\Theta_1$	0.00920	0.01673	0.02257	0.03053	0.05473	0.02683	0.02928
9	$\Theta_2$	0.00000	0.00273	0.00450	0.00647	0.01127	0.00517	0.00546
9	$\Theta_3$	0.00267	0.00793	0.01217	0.01880	0.04540	0.01650	0.01968
9	$D_{1 \rightarrow 2}$	0.00000	0.06080	0.10480	0.13387	0.18827	0.09627	0.09615
9	$S_{1 \rightarrow 2}$	0.00320	0.01800	0.04647	0.08013	0.17093	0.06913	0.07717
9	$D_{2 \rightarrow 3}$	0.00008	0.00173	0.00360	0.00523	0.00907	0.00408	0.00435
9	$S_{2 \rightarrow 3}$	0.00127	0.00414	0.00526	0.00747	0.00989	0.00552	0.00553
10	$\Theta_1$	0.01067	0.01800	0.02430	0.03260	0.05833	0.02877	0.03136
10	$\Theta_2$	0.00593	0.01113	0.01510	0.02007	0.03480	0.01750	0.01887
10	$\Theta_3$	0.00107	0.00500	0.00783	0.01207	0.02767	0.01037	0.01231
10	$D_{1 \rightarrow 2}$	0.00160	0.07360	0.12560	0.15360	0.20373	0.10800	0.10612
10	$S_{1 \rightarrow 2}$	0.00293	0.01653	0.03420	0.08067	0.16573	0.07073	0.07852



Locus	Parameter	2.5%	25.0%	Mode	75.0%	97.5%	Median	Mean
10	$D_{2 \rightarrow 3}$	0.00094	0.00313	0.00490	0.00666	0.00986	0.00511	0.00521
10	$S_{2 \rightarrow 3}$	0.00018	0.00026	0.00069	0.00314	0.00904	0.00421	0.00452
All	$\Theta_1$	0.01160	0.01427	0.01590	0.01760	0.02113	0.01623	0.01628
All	$\Theta_2$	0.00367	0.00547	0.00663	0.00773	0.00967	0.00670	0.00668
All	$\Theta_3$	0.00333	0.00527	0.00650	0.00760	0.00973	0.00657	0.00654
All	$D_{1 \rightarrow 2}$	0.09600	0.11840	0.13040	0.14187	0.16320	0.13093	0.13017
All	$S_{1 \rightarrow 2}$	0.01587	0.04960	0.06820	0.08600	0.14253	0.07273	0.07576
All	$D_{2 \rightarrow 3}$	0.00232	0.00347	0.00375	0.00468	0.00629	0.00418	0.00424
All	$S_{2 \rightarrow 3}$	0.00251	0.00321	0.00457	0.00541	0.00964	0.00576	0.00582

## Citation suggestions:

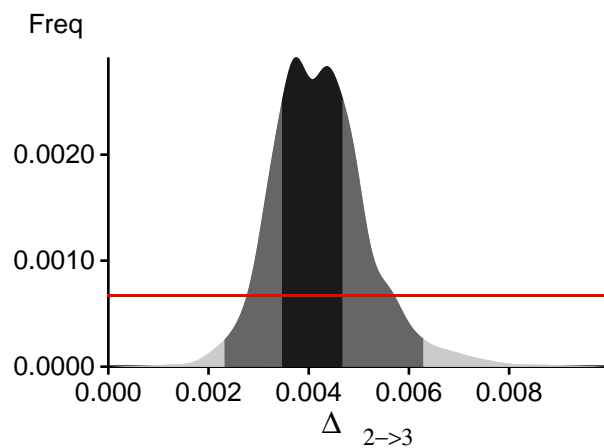
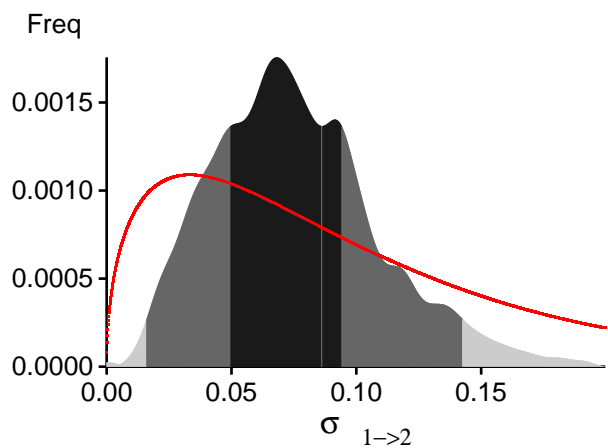
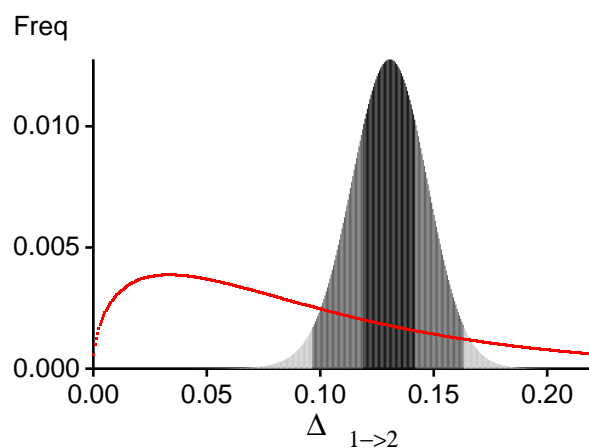
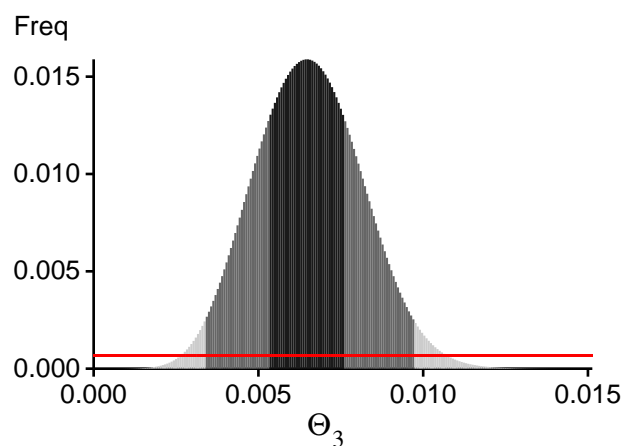
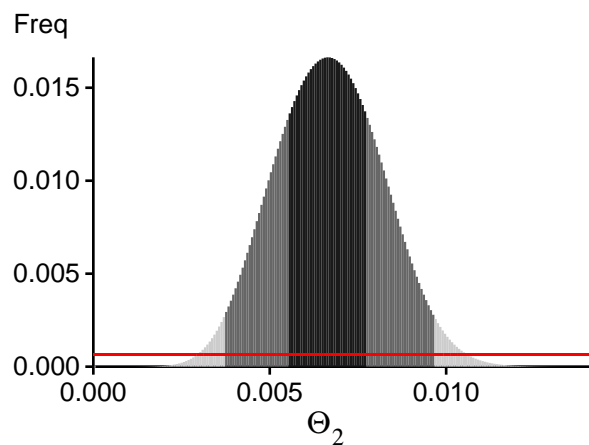
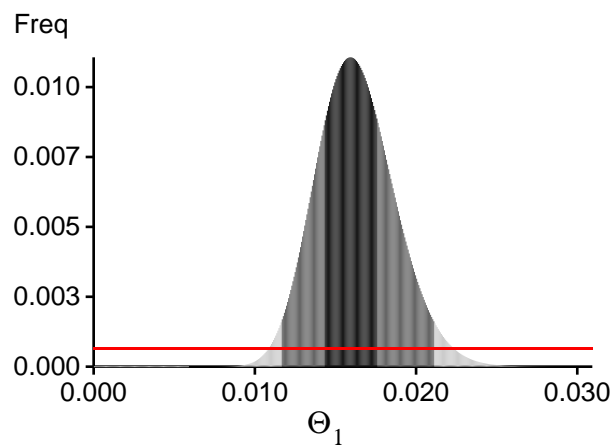
Beerli P., 2006. Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. *Bioinformatics* 22:341-345

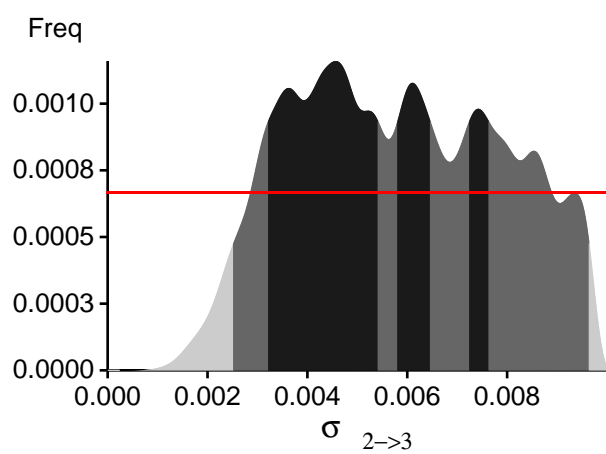
Beerli, P., H. Ashki, S. Mashayekhi, and M. Palczewski, 2022. Population divergence time estimation using individual lineage label switching. *G3 Genes & Genomes & Genetics*, 12(4), 02 2022.

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.

Beerli, P., S. Mashayekhi, M. Sadeghi, M. Khodaei, and K. Shaw, 2019. Population genetic inference with migrate. *Current Protocols in Bioinformatics*, 68(1):e87.

## *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	TI(1a)	BTI(1b)	HS(3)
1	-1585.76	-1453.94	-1435.89
2	-1637.74	-1459.90	-1434.32
3	-1503.82	-1389.95	-1370.74
4	-1549.64	-1414.31	-1396.20
5	-1511.06	-1386.76	-1371.15
6	-1638.30	-1474.70	-1452.92
7	-1766.69	-1549.36	-1516.53
8	-1702.31	-1527.14	-1500.76
9	-1686.84	-1520.97	-1497.88
10	-2031.14	-1741.16	-1695.38
All	-16632.63	-14937.51	-14691.10

(1a) TI: Thermodynamic integration:  $\log(\text{Prob}(D|\text{Model}))$ : Good approximation with many temperatures

(1b) BTI: Bezier-approximated Thermodynamic integration: when using few temperatures USE THIS!

(2) SS: Steppingstone Sampling (Xie et al 2011)

(3) HS: Harmonic mean approximation: Overestimates the marginal likelihood, poor variance

[Scaling factor = -19.323628]

Citation suggestions:

Beerli P. and M. Palczewski, 2010. Unified framework to evaluate panmixia and migration direction among multiple sampling locations, *Genetics*, 185: 313-326.

Palczewski M. and P. Beerli, 2014. Population model comparison using multi-locus datasets.

In M.-H. Chen, L. Kuo, and P. O. Lewis, editors, *Bayesian Phylogenetics: Methods, Algorithms, and Applications*, pages 187-200. CRC Press, 2014.

Xie W., P. O. Lewis, Y. Fan, L. Kuo, and M.-H. Chen. 2011. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Systematic Biology*, 60(2):150â 160, 2011.

## *Acceptance ratios for all parameters and the genealogies*

Parameter	Accepted changes	Ratio
$\Theta_1$	329347/714044	0.46124
$\Theta_2$	349350/714176	0.48917
$\Theta_3$	345162/714990	0.48275
$\Delta_{1 \rightarrow 2}$	474578/714180	0.66451
$\sigma_{1 \rightarrow 2}$	488259/713891	0.68394
$\Delta_{2 \rightarrow 3}$	556584/713444	0.78014
$\sigma_{2 \rightarrow 3}$	544684/715739	0.76101
Genealogies	1162109/4999536	0.23244

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

Parameter	Autocorrelation	Effective Sample Size
$\Theta_1$	0.40940	42797.20
$\Theta_2$	0.50608	33653.94
$\Theta_3$	0.52572	32793.44
$\Delta_{1 \rightarrow 2}$	0.05596	89434.58
$\sigma_{1 \rightarrow 2}$	0.08476	84408.27
$\Delta_{2 \rightarrow 3}$	0.17867	69781.22
$\sigma_{2 \rightarrow 3}$	0.23977	61731.56
Genealogies	0.40940	42797.20

## *Average temperatures during the run*

Chain	Temperatures
-------	--------------

1	0.00000
2	0.00000
3	0.00000
4	0.00000

Adaptive heating often fails, if the average temperatures are very close together try to rerun using static heating! If you want to compare models using marginal likelihoods then you **MUST** use static heating

## *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 common that some parameters for some loci will not be informative. These parameters then trigger suggestions to increase the prior range that are not sensible. Do not blindly follow the suggestions given. If some parameters are flagged, inspect the tables carefully and judge whether an action is required. Suppose you run a Bayesian inference with sequence data for macroscopic species. In that case, there is rarely the need to increase the prior for Theta beyond 0.1. If you use microsatellites data, 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 prior range will not help in such situations, but reducing the number of parameters may help.

Param 13 (Locus 8): Upper prior boundary seems too low!