

POPULATION SIZE, MIGRATION, DIVERGENCE, ASSIGNMENT, HISTORY

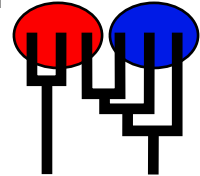
Bayesian inference using the structured coalescent

Migrate-n version 4.4.4(git:v4-series-30-gb190f74) [June-1-2019]

Compiled for a SYMMETRIC multiprocessors (Grandcentral)

Program started at Mon Nov 18 14:43:18 2019

Program finished at Mon Nov 18 14:48:59 2019 [Runtime:0000:00:05:41]



Options

Datatype:

DNA sequence data

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)

2844857774

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	4	5
1 Romanshorn_0	*	0	0	0	d
2 Arbon_1	*	*	0	0	0
3 Kreuzlingen_2	0	*	*	*	0
4 Frauenfeld_3	0	0	0	*	*
5 Guendelhart_4	0	0	0	0	*

Order of parameters:

1	Θ_1	<displayed>
2	Θ_2	<displayed>
3	Θ_3	<displayed>
4	Θ_4	<displayed>
5	Θ_5	<displayed>
6	$M_{1 \rightarrow 2}$	<displayed>
7	$M_{2 \rightarrow 3}$	<displayed>
8	$M_{4 \rightarrow 3}$	<displayed>
9	$M_{5 \rightarrow 4}$	<displayed>
10	$\Delta_{5 \rightarrow 1}$	<displayed>
11	$\sigma_{5 \rightarrow 1}$	<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	0.010	2000	0.04545
2	Theta	** Uniform	0.000000	0.050	0.100	0.010	2000	0.04545
3	Theta	** Uniform	0.000000	0.050	0.100	0.010	2000	0.04545
4	Theta	** Uniform	0.000000	0.050	0.100	0.010	2000	0.04545
5	Theta	** Uniform	0.000000	0.050	0.100	0.010	2000	0.04545
6	M	** Uniform	0.000000	500.0	1000.	100.0	1500	0.04545
7	M	** Uniform	0.000000	500.0	1000.	100.0	1500	0.04545
8	M	** Uniform	0.000000	500.0	1000.	100.0	1500	0.04545
9	M	** Uniform	0.000000	500.0	1000.	100.0	1500	0.04545
10	Splittime mean	** Uniform	0.000000	0.250	0.500	0.050	1500	0.04545
11	Splittime std	** Uniform	0.000000	0.250	0.500	0.050	1500	0.04545

[* * 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)		500
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:		infile1
Haplotyping is turned on:		NO
Output file:		outfile_d2
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: infile1
 Datatype: Sequence data
 Number of loci: 2

Mutationmodel:

Locus	Sublocus	Mutationmodel	Mutationmodel parameters
1	1	Felsenstein 84	[Bf:0.26 0.24 0.26 0.24, t/t ratio=2.000]
2	1	Felsenstein 84	[Bf:0.25 0.26 0.25 0.25, t/t ratio=2.000]

Sites per locus

Locus	Sites
1	1000
2	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

Population	Locus	Gene copies
1 Romanshorn_0	1	5
	2	5
2 Arbon_1	1	5
	2	5
3 Kreuzlingen_2	1	5
	2	5
4 Frauenfeld_3	1	5
	2	5
5 Guendelhart_4	1	5
	2	5
Total of all populations	1	25
	2	25

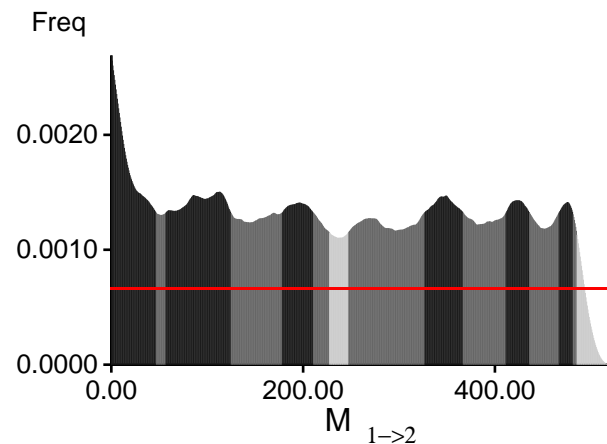
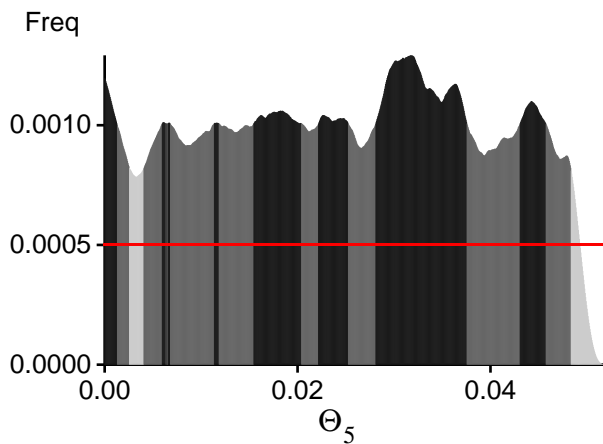
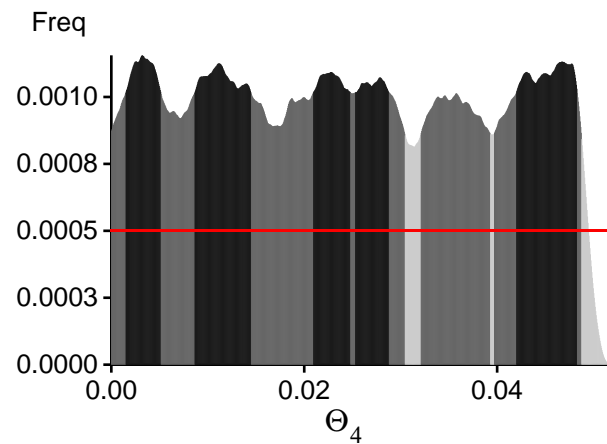
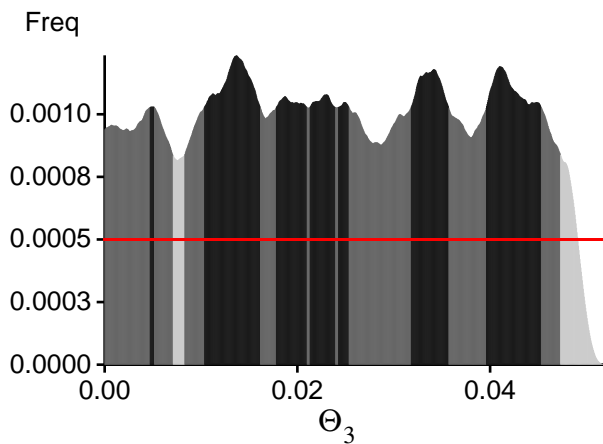
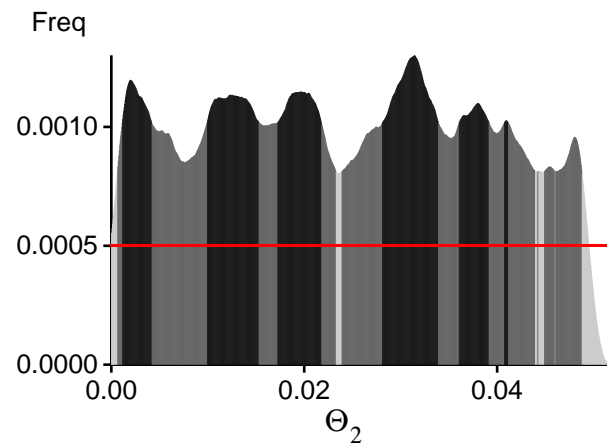
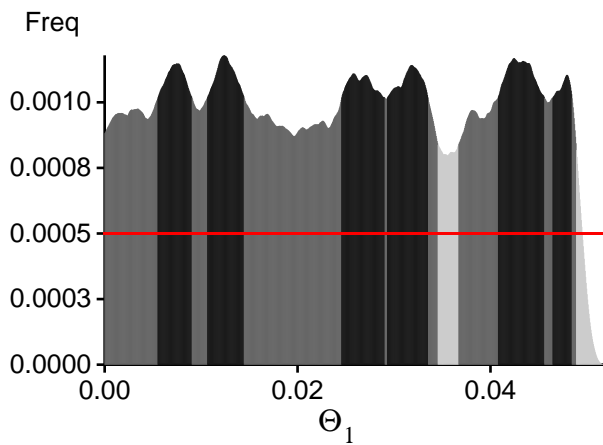
Bayesian Analysis: Posterior distribution table

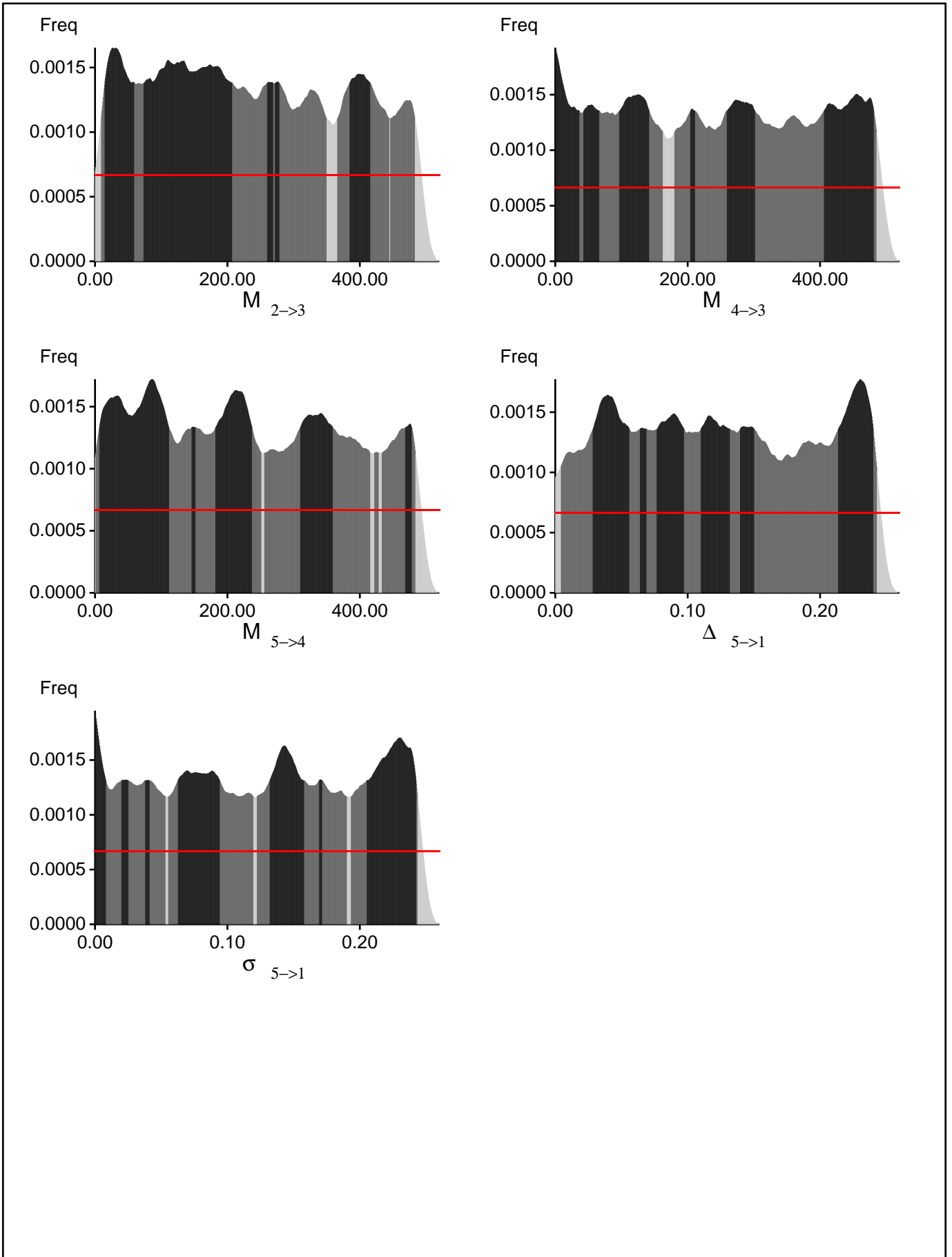
Locus	Parameter	2.5%	25.0%	Mode	75.0%	97.5%	Median	Mean
1	Θ_1	0.04275	0.04300	0.04313	0.04320	0.04330	0.04938	0.04972
1	Θ_2	0.07750	0.07765	0.07773	0.07775	0.07865	0.04723	0.04826
1	Θ_3	0.01065	0.01130	0.01143	0.01145	0.01205	0.04453	0.04699
1	Θ_4	0.00350	0.00350	0.00358	0.00360	0.00375	0.04973	0.05001
1	Θ_5	0.00985	0.01110	0.01118	0.01130	0.01335	0.04638	0.04802
1	$M_{1 \rightarrow 2}$	724.0	724.0	725.0	730.0	730.0	513.7	504.9
1	$M_{2 \rightarrow 3}$	987.3	987.3	989.0	989.3	990.0	483.7	486.8
1	$M_{4 \rightarrow 3}$	692.7	695.3	697.7	698.7	702.0	527.0	517.5
1	$M_{5 \rightarrow 4}$	955.3	958.7	961.0	962.0	974.7	483.7	490.6
1	$D_{5 \rightarrow 1}$	0.04967	0.05000	0.05083	0.05100	0.05100	0.26917	0.26058
1	$S_{5 \rightarrow 1}$	0.42000	0.42133	0.42283	0.42333	0.42567	0.25217	0.25334
2	Θ_1	0.00000	0.00010	0.00018	0.00025	0.00080	0.04993	0.05004
2	Θ_2	0.01320	0.01400	0.01408	0.01410	0.01525	0.04998	0.05014
2	Θ_3	0.07265	0.07345	0.07353	0.07360	0.07495	0.05058	0.05046
2	Θ_4	0.02815	0.02905	0.02923	0.02930	0.02995	0.05258	0.05138
2	Θ_5	0.03230	0.03285	0.03298	0.03305	0.03310	0.04893	0.04960
2	$M_{1 \rightarrow 2}$	187.3	189.3	191.0	192.7	195.3	501.0	495.9
2	$M_{2 \rightarrow 3}$	388.7	388.7	389.7	390.0	391.3	495.0	493.9
2	$M_{4 \rightarrow 3}$	873.3	883.3	884.3	884.7	888.0	498.3	497.2
2	$M_{5 \rightarrow 4}$	144.7	148.0	149.0	149.3	149.3	494.3	492.4
2	$D_{5 \rightarrow 1}$	0.03300	0.03767	0.03850	0.03867	0.03900	0.24550	0.24675
2	$S_{5 \rightarrow 1}$	0.28033	0.28033	0.28150	0.28167	0.28633	0.24583	0.24553
All	Θ_1	0.00000	0.01060	0.01243	0.01445	0.03455	0.02528	0.02494
All	Θ_2	0.02385	0.02805	0.03148	0.03390	0.04395	0.02463	0.02449
All	Θ_3	0.00825	0.01030	0.01368	0.01615	0.04730	0.02458	0.02473
All	Θ_4	0.00000	0.00145	0.00323	0.00515	0.03045	0.02453	0.02471
All	Θ_5	0.00400	0.02805	0.03173	0.03755	0.04835	0.02548	0.02498
All	$M_{1 \rightarrow 2}$	0.0	0.0	0.3	46.7	227.3	235.7	238.7
All	$M_{2 \rightarrow 3}$	8.7	14.0	26.3	59.3	350.0	231.7	239.1
All	$M_{4 \rightarrow 3}$	0.0	0.0	0.3	36.7	162.7	248.3	245.9
All	$M_{5 \rightarrow 4}$	0.7	6.0	86.3	112.0	251.3	230.3	238.2
All	$D_{5 \rightarrow 1}$	0.00400	0.21333	0.23017	0.24033	0.24300	0.12250	0.12442
All	$S_{5 \rightarrow 1}$	0.00000	0.00000	0.00017	0.00833	0.05333	0.12883	0.12613

Citation suggestions:

- Beerli P., 2006. Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. *Bioinformatics* 22:341-345
- Beerli P., 2007. Estimation of the population scaled mutation rate from microsatellite data, *Genetics*, 177:1967-1968.
- 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	-11265.83	-8391.51	-7885.24
2	-12231.14	-8065.29	-7298.35
All	-23541.99	-16501.82	-15228.61

(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 = -45.021911]

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	45072/45152	0.99823
Θ_2	45441/45460	0.99958
Θ_3	45630/45662	0.99930
Θ_4	45697/45738	0.99910
Θ_5	45516/45589	0.99840
$M_{1 \rightarrow 2}$	45507/45591	0.99816
$M_{2 \rightarrow 3}$	45580/45669	0.99805
$M_{4 \rightarrow 3}$	45544/45624	0.99825
$M_{5 \rightarrow 4}$	45338/45468	0.99714
$\Delta_{5 \rightarrow 1}$	45212/45270	0.99872
$\sigma_{5 \rightarrow 1}$	45830/45880	0.99891
Genealogies	26567/498897	0.05325

MCMC-Autocorrelation and Effective MCMC Sample Size

Parameter	Autocorrelation	Effective Sample Size
Θ_1	0.71166	1811.86
Θ_2	0.72891	1675.44
Θ_3	0.69751	1929.96
Θ_4	0.68984	1975.78
Θ_5	0.72080	1732.57
$M_{1 \rightarrow 2}$	0.72163	1722.96
$M_{2 \rightarrow 3}$	0.72133	1737.20
$M_{4 \rightarrow 3}$	0.69383	1952.75
$M_{5 \rightarrow 4}$	0.70614	1850.53
$\Delta_{5 \rightarrow 1}$	0.70922	1834.56
$\sigma_{5 \rightarrow 1}$	0.73527	1625.07
Genealogies	0.73527	1625.07

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 15 (Locus 1): Upper prior boundary seems too low!

Param 21 (Locus 1): Upper prior boundary seems too low!