

# AUTO

POPULATION SIZE, MIGRATION, DIVERGENCE, ASSIGNMENT, HISTORY

Bayesian inference using the structured coalescent

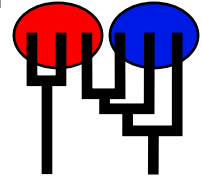
Migrate-n version debug 4.2.8 [April-1-2016]

Using Intel AVX (Advanced Vector Extensions)

Compiled for a SYMMETRIC multiprocessors (Grandcentral)

Program started at Sun Aug 7 16:57:37 2016

Program finished at Sun Aug 7 20:14:42 2016 [Runtime:0000:03:17:05]



## 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) 124052460

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
1 Romanshorn_0	*	*	0	t
2 Arbon_1	*	*	0	t
3 Kreuzlingen_2	0	0	*	0
4 ancestor	0	0	d	*

Order of parameters:

1  $\Theta_1$  <displayed>



Splittime std	Gamma	0.000000	1.000000	10.000000	1.000000	1500
Markov chain settings:						Long chain
Number of chains						1
Recorded steps [a]						10000
Increment (record every x step [b])						500
Number of concurrent chains (replicates) [c]						2
Visited (sampled) parameter values [a*b*c]						10000000
Number of discard trees per chain (burn-in)						10000
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_2_0.25_10
Haplotyping is turned on:						NO
Output file:						outfile_2_0.25_10
Posterior distribution raw histogram file:						bayesfile
Raw data from the MCMC run:						bayesallfile_2_0.25_10
Print data:						No
Print genealogies [only some for some data type]:						None

## Data summary

Data file: infile\_2\_0.25\_10  
 Datatype: Haplotype data  
 Number of loci: 2

### Mutationmodel:

Locus	Sublocus	Mutationmodel	Mutationmodel parameters
1	1	Felsenstein 84	[Bf:0.25 0.25 0.25 0.26, t/t ratio=2.000]
2	1	Felsenstein 84	[Bf:0.23 0.26 0.25 0.26, 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
					data (missing)
1 Romanshorn_0			1	20	
			2	20	
2 Arbon_1			1	20	
			2	20	
3 Kreuzlingen_2			1	1	
			2	1	
4 ancestor			1	0	
			2	0	
Total of all populations			1	41	(0)
			2	41	(0)

## *Bayesian Analysis: Posterior distribution table*

Locus	Parameter	2.5%	25.0%	Mode	75.0%	97.5%	Median	Mean
1	$\Theta_1$	0.00320	0.00773	0.01060	0.01347	0.02027	0.01127	0.01157
1	$\Theta_2$	0.00107	0.00520	0.00780	0.01027	0.01560	0.00820	0.00834
1	$\Theta_3$	0.00000	0.00067	0.00513	0.01400	0.04773	0.01327	0.01705
1	$\Theta_4$	0.00000	0.00000	0.00527	0.01533	0.05547	0.01540	0.02002
1	$M_{2 \rightarrow 1}$	0.000	30.667	64.333	110.000	158.667	118.333	147.124
1	$M_{1 \rightarrow 2}$	28.000	92.667	169.000	241.333	406.000	203.000	251.938
1	$D_{4 \rightarrow 1}$	0.00000	0.03000	0.03700	0.04533	0.13867	0.09100	0.10775
1	$S_{4 \rightarrow 1}$	0.03333	0.03333	0.20333	0.37333	0.37333	0.42333	0.56623
1	$D_{4 \rightarrow 2}$	0.00000	0.03000	0.03700	0.04533	0.13867	0.09100	0.10775
1	$S_{4 \rightarrow 2}$	0.03333	0.03333	0.20333	0.37333	0.37333	0.42333	0.56623
1	$D_{3 \rightarrow 4}$	0.00933	0.00933	0.02967	0.08400	0.08400	0.09967	0.11626
1	$S_{3 \rightarrow 4}$	0.00000	0.01333	0.21667	0.44667	1.85333	0.43000	0.58599
2	$\Theta_1$	0.00187	0.00573	0.00847	0.01093	0.01667	0.00900	0.00456
2	$\Theta_2$	0.00107	0.00493	0.00753	0.00987	0.01493	0.00793	0.00402
2	$\Theta_3$	0.00120	0.00120	0.00553	0.01307	0.01307	0.01340	0.00858
2	$\Theta_4$	0.00160	0.00160	0.00593	0.01427	0.01427	0.01553	0.00993
2	$M_{2 \rightarrow 1}$	12.667	12.667	61.000	115.333	115.333	111.667	70.852
2	$M_{1 \rightarrow 2}$	0.000	26.000	46.333	148.667	362.000	167.667	107.504
2	$D_{4 \rightarrow 1}$	0.00533	0.01133	0.03700	0.08800	0.11400	0.09100	0.05396
2	$S_{4 \rightarrow 1}$	0.00000	0.00000	0.00333	0.37333	1.75333	0.37667	0.28139
2	$D_{4 \rightarrow 2}$	0.00533	0.01133	0.03700	0.08800	0.11400	0.09100	0.05396
2	$S_{4 \rightarrow 2}$	0.00000	0.00000	0.00333	0.37333	1.75333	0.37667	0.28139
2	$D_{3 \rightarrow 4}$	0.01000	0.01000	0.03233	0.08000	0.08000	0.10100	0.05898
2	$S_{3 \rightarrow 4}$	0.00000	0.02000	0.22333	0.46667	1.85333	0.44333	0.29933
All	$\Theta_1$	0.00333	0.00707	0.00940	0.01173	0.01600	0.00967	0.00973
All	$\Theta_2$	0.00173	0.00520	0.00753	0.00947	0.01307	0.00767	0.00758
All	$\Theta_3$	0.00080	0.00107	0.00540	0.01253	0.01320	0.01180	0.01545
All	$\Theta_4$	0.00000	0.00067	0.00633	0.01627	0.05680	0.01553	0.02016
All	$M_{2 \rightarrow 1}$	11.333	11.333	63.000	128.000	128.000	87.000	101.288
All	$M_{1 \rightarrow 2}$	18.000	80.000	135.000	206.000	336.667	165.667	187.141
All	$D_{4 \rightarrow 1}$	0.01067	0.01067	0.04833	0.15467	0.15467	0.10500	0.11734
All	$S_{4 \rightarrow 1}$	0.00000	0.00000	0.15000	0.27333	1.18667	0.27000	0.35286
All	$D_{4 \rightarrow 2}$	0.01067	0.01067	0.04833	0.15467	0.15467	0.10500	0.11734
All	$S_{4 \rightarrow 2}$	0.00000	0.00000	0.15000	0.27333	1.18667	0.27000	0.35286

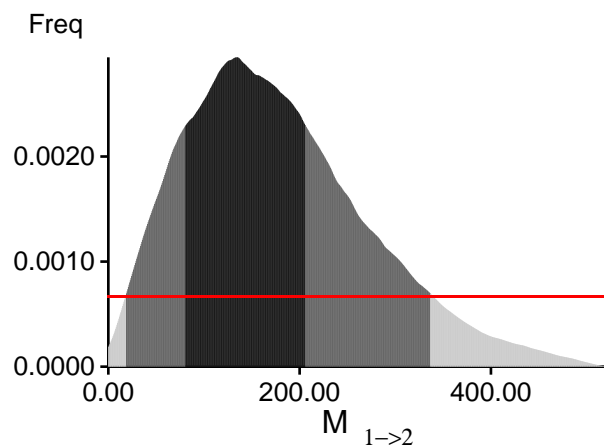
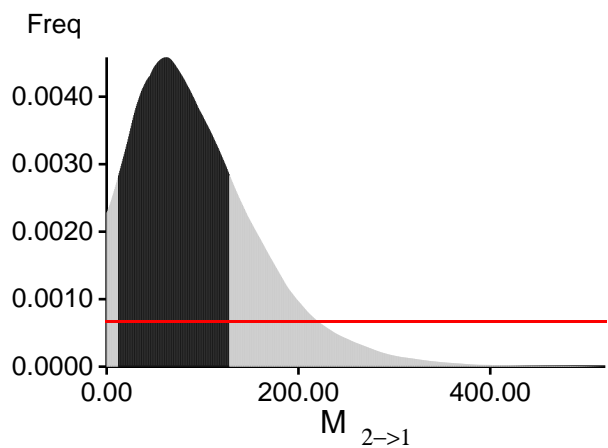
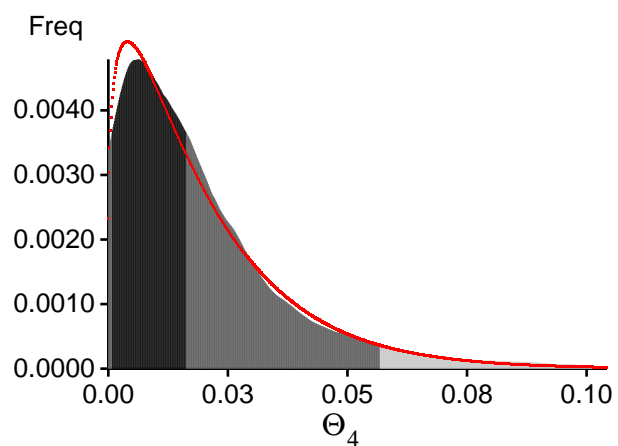
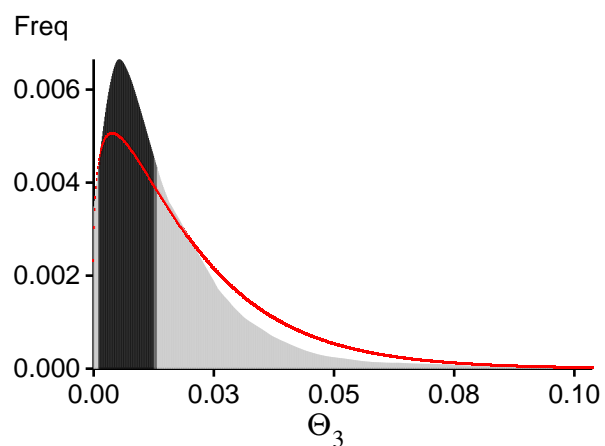
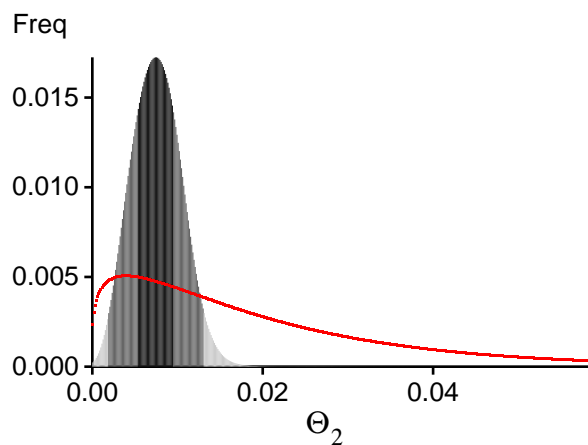
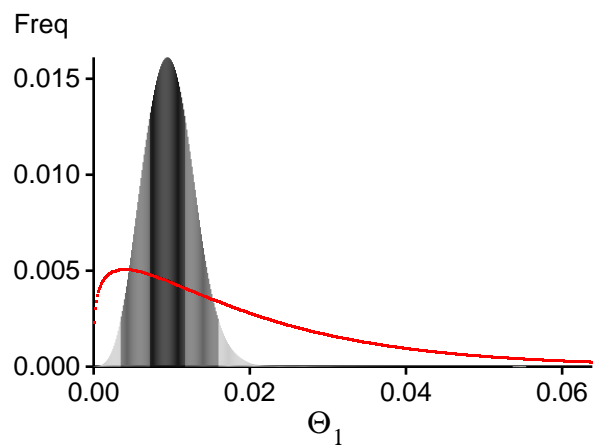
Locus	Parameter	2.5%	25.0%	Mode	75.0%	97.5%	Median	Mean
All	D <sub>3-&gt;4</sub>	0.00600	0.11533	0.17567	0.19067	0.21000	0.12967	0.13615
All	S <sub>3-&gt;4</sub>	0.00000	0.04667	0.20333	0.34667	1.24000	0.31000	0.40102

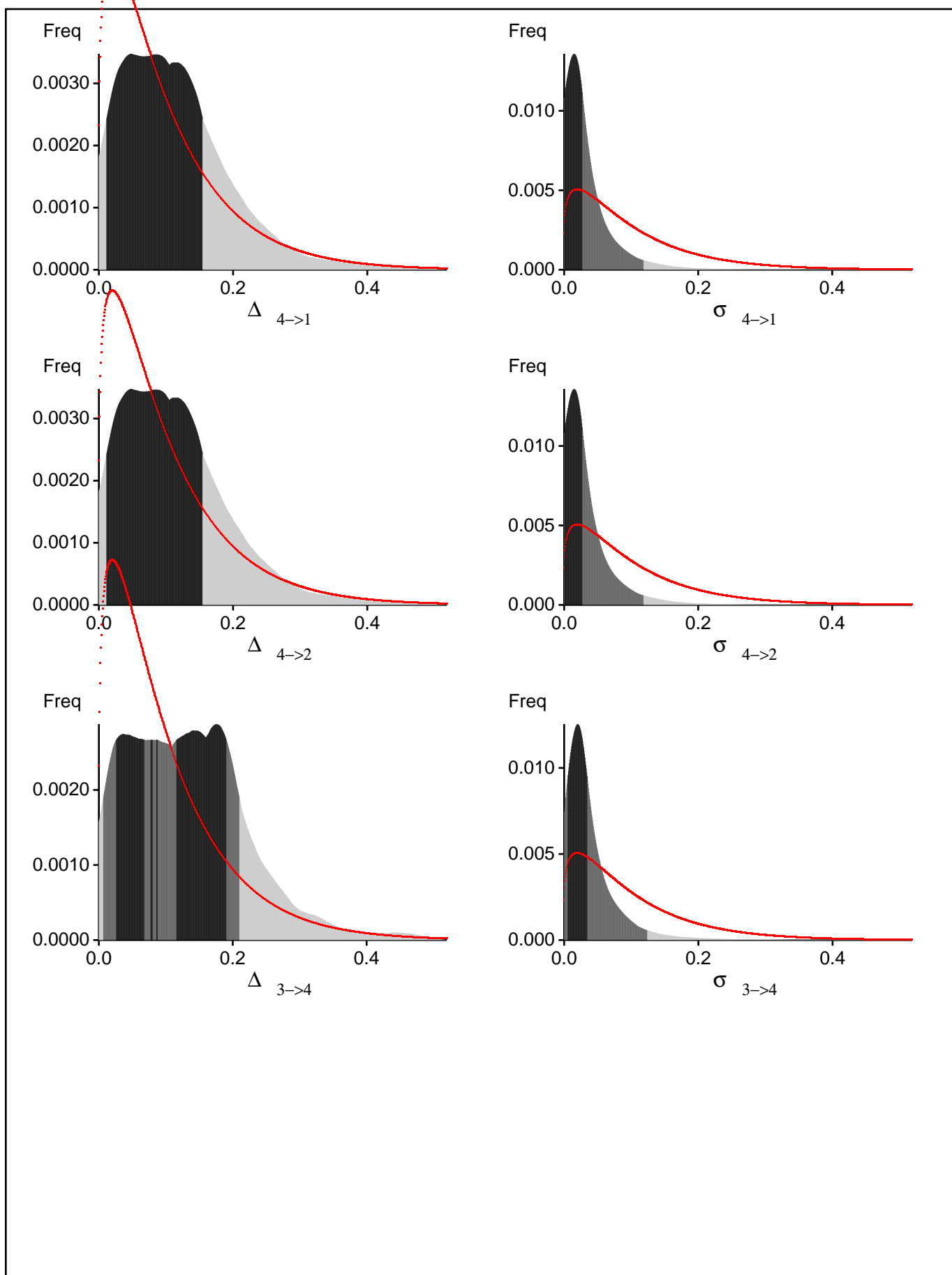
Citation suggestions:

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	-3299.90	-2911.72	-2866.94
2	-3130.06	-2830.94	-2793.67
All	-6462.50	-5775.19	-5693.15

(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 = -32.538802]

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
$\Theta_1$	191504/832900	0.22992
$\Theta_2$	173930/832966	0.20881
$\Theta_3$	546557/833007	0.65613
$\Theta_4$	825746/833677	0.99049
$M_{2 \rightarrow 1}$	423949/834172	0.50823
$M_{1 \rightarrow 2}$	284577/832614	0.34179
$\Delta_{4 \rightarrow 1}$	709362/833543	0.85102
$\sigma_{4 \rightarrow 1}$	503439/832408	0.60480
$\Delta_{4 \rightarrow 2}$	708441/832654	0.85082
$\sigma_{4 \rightarrow 2}$	503119/832966	0.60401
$\Delta_{3 \rightarrow 4}$	703425/833031	0.84442
$\sigma_{3 \rightarrow 4}$	513278/833620	0.61572
Genealogies	1206141/10002442	0.12058

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

Parameter	Autocorrelation	Effective Sampe Size
$\Theta_1$	0.33061	20156.58
$\Theta_2$	0.33287	20028.68
$\Theta_3$	0.13101	30732.09
$\Theta_4$	0.00146	39891.63
$M_{2 \rightarrow 1}$	0.78367	5009.96
$M_{1 \rightarrow 2}$	0.67915	7650.69
$\Delta_{4 \rightarrow 1}$	0.01039	39177.16
$\sigma_{4 \rightarrow 1}$	0.00271	39796.12
$\Delta_{4 \rightarrow 2}$	0.01039	39177.16
$\sigma_{4 \rightarrow 2}$	0.00271	39796.12
$\Delta_{3 \rightarrow 4}$	0.01593	38745.27
$\sigma_{3 \rightarrow 4}$	0.00665	39476.19
$\text{Ln}[\text{Prob}(\mathbf{D} \mathbf{G})]$	0.48377	13989.05

## *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