

Example: sequence data set wit two loci [simula

POPULATION SIZE, MIGRATION, DIVERGENCE, ASSIGNMENT, HISTORY

Bayesian inference using the structured coalescent

Migrate-n version 4.1.3a [Feb-22-2015]

Using Intel AVX (Advanced Vector Extensions)

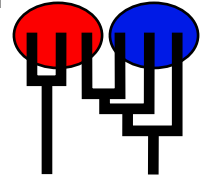
Compiled for PARALLEL computer architectures

One master and 7 compute nodes are available.

Compiled for a SYMMETRIC multiprocessors (Grandcentral)

Program started at Sun Feb 22 13:55:43 2015

Program finished at Sun Feb 22 13:56:24 2015



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:

(from parmfile)

310705631

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 Africa	*	0	0	0
2 Americas	0	*	0	D
3 Pacific	0	0	*	d
4 Asia	d	0	0	*

Order of parameters:

1	Θ_1	<displayed>
2	Θ_2	<displayed>
3	Θ_3	<displayed>
4	Θ_4	<displayed>
5	$M_{4 \rightarrow 2}$	<displayed>
6	$\Delta_{4 \rightarrow 2}$	<displayed>
7	$\sigma_{4 \rightarrow 2}$	<displayed>
8	$\Delta_{4 \rightarrow 3}$	<displayed>
9	$\sigma_{4 \rightarrow 3}$	<displayed>
10	$\Delta_{1 \rightarrow 4}$	<displayed>
11	$\sigma_{1 \rightarrow 4}$	<displayed>

Mutation rate among loci:

Mutation rate is constant for all loci

Analysis strategy:

Bayesian inference

Proposal distributions for parameter

Parameter	Proposal
Theta	Slice sampling
M	Slice sampling

Prior distribution for parameter

Parameter	Prior	Minimum	Mean*	Maximum	Delta	Bins
Theta	Uniform	0.000000	10.000000	20.000000	2.000000	500
Theta	Uniform	0.000000	10.000000	20.000000	2.000000	500
Theta	Uniform	0.000000	10.000000	20.000000	2.000000	500
Theta	Uniform	0.000000	10.000000	20.000000	2.000000	500
M	Uniform	0.000000	10.000000	20.000000	2.000000	500
M	Uniform	0.000000	10.000000	20.000000	2.000000	500
M	Uniform	0.000000	10.000000	20.000000	2.000000	500
M	Uniform	0.000000	10.000000	20.000000	2.000000	500
M	Uniform	0.000000	10.000000	20.000000	2.000000	500
M	Uniform	0.000000	10.000000	20.000000	2.000000	500
M	Uniform	0.000000	10.000000	20.000000	2.000000	500
M	Uniform	0.000000	10.000000	20.000000	2.000000	500
M	Uniform	0.000000	10.000000	20.000000	2.000000	500
M	Uniform	0.000000	10.000000	20.000000	2.000000	500
M	Uniform	0.000000	10.000000	20.000000	2.000000	500
M	Uniform	0.000000	10.000000	20.000000	2.000000	500
M	Uniform	0.000000	10.000000	20.000000	2.000000	500
M	Uniform	0.000000	10.000000	20.000000	2.000000	500
Splittime mean	Uniform	0.000000	50.000000	100.000000	10.000000	1500
Splittime std	Uniform	0.000000	50.000000	100.000000	10.000000	1500
Splittime mean	Uniform	0.000000	50.000000	100.000000	10.000000	1500
Splittime std	Uniform	0.000000	50.000000	100.000000	10.000000	1500

Splittime mean	Uniform	0.000000	50.000000	100.000000	10.000000	1500
Splittime std	Uniform	0.000000	50.000000	100.000000	10.000000	1500

Markov chain settings:

Number of chains	Long chain	1
Recorded steps [a]		500
Increment (record every x step [b])		1
Number of concurrent chains (replicates) [c]		2
Visited (sampled) parameter values [a*b*c]		1000
Number of discard trees per chain (burn-in)		10

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.seq
Haplotyping is turned on:	NO
Output file:	outfile-bayes
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.seq
 Datatype: Haplotype data
 Number of loci: 2

Mutationmodel:

Locus	Sublocus	Mutationmodel	Mutationmodel parameters
1	1	Felsenstein 84	[Bf:0.31 0.19 0.28 0.22, t/t ratio=2.000]
2	1	Felsenstein 84	[Bf:0.28 0.22 0.21 0.29, t/t ratio=2.000]

Sites per locus

Locus	Sites
1	145
2	345

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 Africa	1	25	
	2	25	
2 Americas	1	25	
	2	25	
3 Pacific	1	25	
	2	25	
4 Asia	1	25	
	2	25	
Total of all populations	1	100	(0)
	2	100	(0)

Bayesian Analysis: Posterior distribution table

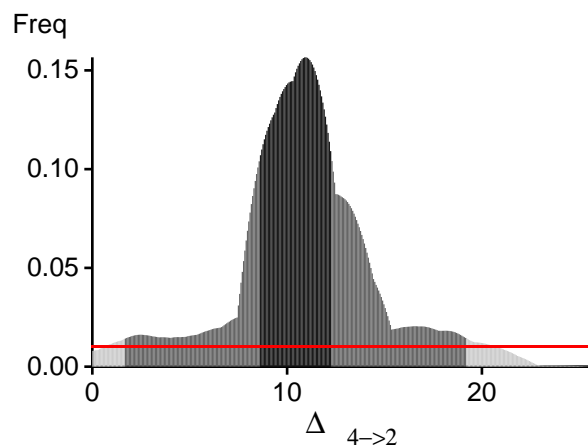
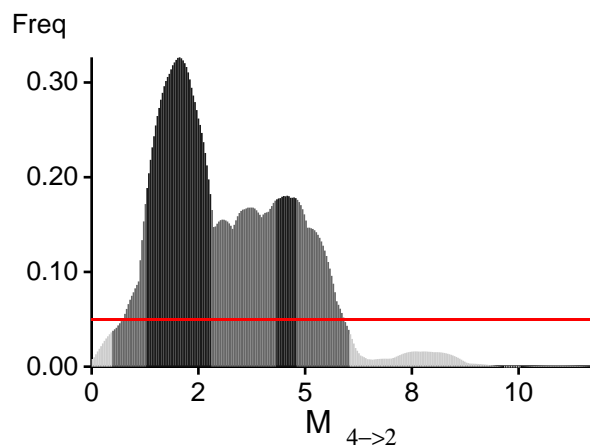
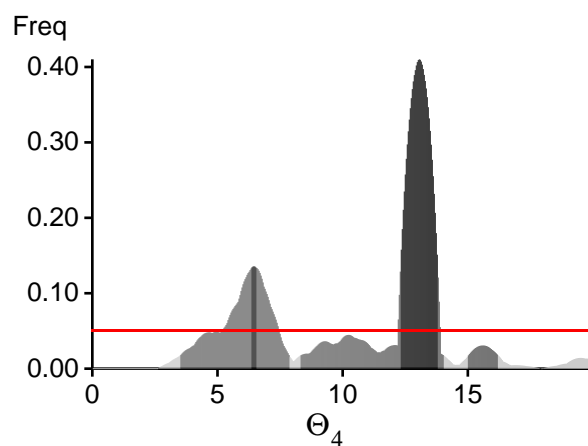
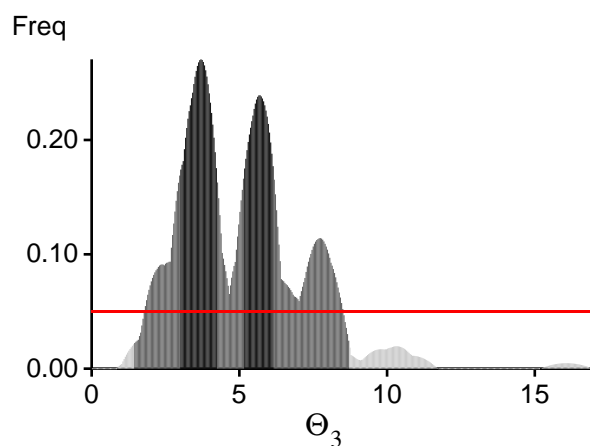
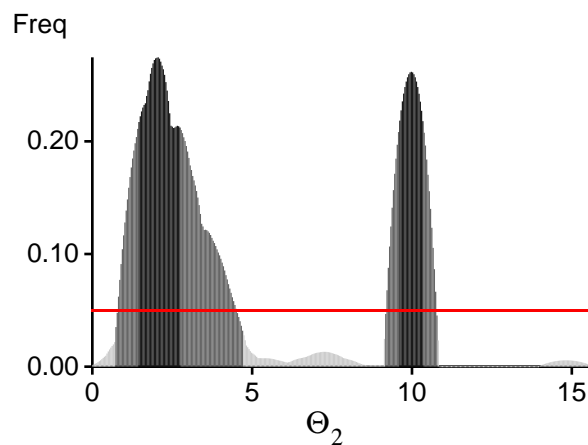
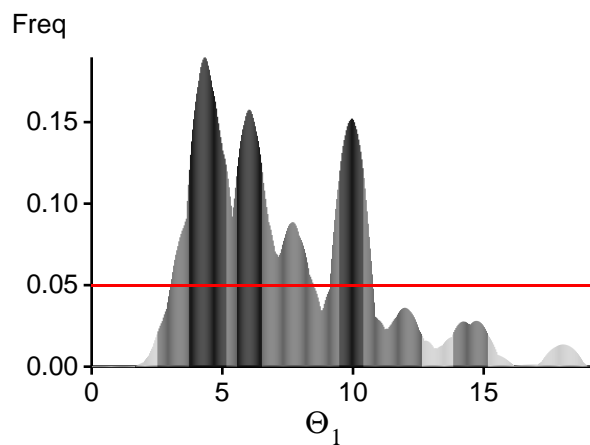
Locus	Parameter	2.5%	25.0%	Mode	75.0%	97.5%	Median	Mean
1	Θ_1	4.88000	5.48000	6.14000	6.96000	11.56000	7.98000	8.73234
1	Θ_2	1.20000	2.72000	3.58000	4.12000	5.60000	3.54000	3.90509
1	Θ_3	1.24000	5.00000	5.98000	6.96000	8.72000	5.82000	5.45310
1	Θ_4	5.08000	5.36000	6.18000	6.84000	13.84000	10.22000	10.30796
1	$M_{4 \rightarrow 2}$	0.360	2.920	3.620	3.880	6.080	3.460	3.503
1	$D_{4 \rightarrow 2}$	4.26667	5.86667	7.90000	9.66667	21.86667	12.76667	12.77253
1	$S_{4 \rightarrow 2}$	6.53333	14.93333	16.76667	18.53333	22.00000	15.50000	16.07695
1	$D_{4 \rightarrow 3}$	3.73333	9.20000	12.90000	15.20000	20.46667	12.83333	14.43618
1	$S_{4 \rightarrow 3}$	6.73333	11.40000	13.36667	15.66667	20.13333	13.43333	13.27158
1	$D_{1 \rightarrow 4}$	2.00000	8.26667	10.56667	14.40000	17.60000	10.50000	10.27715
1	$S_{1 \rightarrow 4}$	10.93333	16.33333	18.36667	20.80000	27.66667	18.63333	18.67829
2	Θ_1	2.24000	3.08000	4.18000	5.80000	8.32000	5.42000	6.52924
2	Θ_2	0.16000	0.96000	1.58000	2.32000	3.92000	2.06000	3.22124
2	Θ_3	1.20000	2.80000	3.86000	4.64000	4.96000	5.98000	6.21836
2	Θ_4	3.40000	5.52000	6.70000	7.44000	7.88000	7.02000	8.46014
2	$M_{4 \rightarrow 2}$	0.120	1.800	2.620	3.480	6.280	3.020	3.498
2	$D_{4 \rightarrow 2}$	0.00000	0.93333	2.30000	3.86667	6.40000	9.23333	8.59434
2	$S_{4 \rightarrow 2}$	7.66667	8.73333	10.10000	11.53333	25.73333	14.56667	14.10327
2	$D_{4 \rightarrow 3}$	4.26667	5.80000	8.36667	11.20000	14.60000	8.43333	8.05747
2	$S_{4 \rightarrow 3}$	1.40000	7.60000	10.03333	13.60000	17.00000	10.03333	9.92163
2	$D_{1 \rightarrow 4}$	8.13333	17.46667	22.56667	23.86667	25.46667	19.10000	18.18640
2	$S_{1 \rightarrow 4}$	3.13333	11.20000	13.76667	15.80000	24.00000	13.70000	14.47894
All	Θ_1	2.48000	3.68000	4.34000	5.16000	12.64000	6.66000	7.48258
All	Θ_2	0.68000	1.40000	2.06000	2.76000	4.72000	3.18000	4.96637
All	Θ_3	1.40000	2.96000	3.70000	4.24000	8.72000	5.06000	5.13345
All	Θ_4	8.28000	12.28000	13.06000	13.80000	14.04000	12.62000	10.78857
All	$M_{4 \rightarrow 2}$	0.440	1.240	2.060	2.800	6.040	2.900	3.214
All	$D_{4 \rightarrow 2}$	1.60000	8.53333	10.96667	12.26667	19.20000	10.90000	10.84742
All	$S_{4 \rightarrow 2}$	7.00000	13.53333	16.83333	19.06667	26.13333	15.30000	14.98051
All	$D_{4 \rightarrow 3}$	4.40000	12.53333	13.83333	15.20000	16.46667	12.03333	11.78865
All	$S_{4 \rightarrow 3}$	7.20000	10.06667	11.23333	12.60000	16.06667	11.50000	11.38349
All	$D_{1 \rightarrow 4}$	6.60000	16.53333	18.16667	19.86667	24.33333	16.30000	14.57228
All	$S_{1 \rightarrow 4}$	14.86667	16.33333	17.50000	18.53333	23.80000	15.70000	14.35950

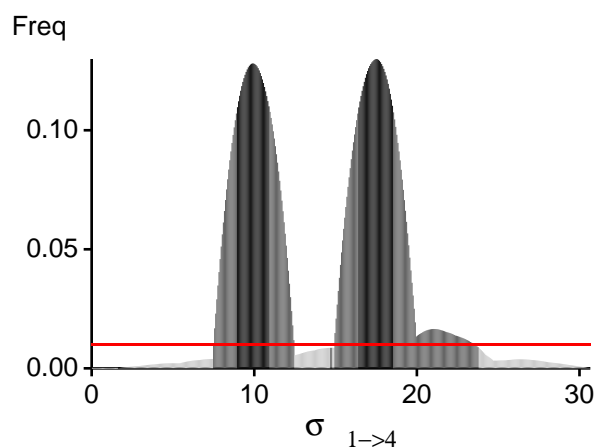
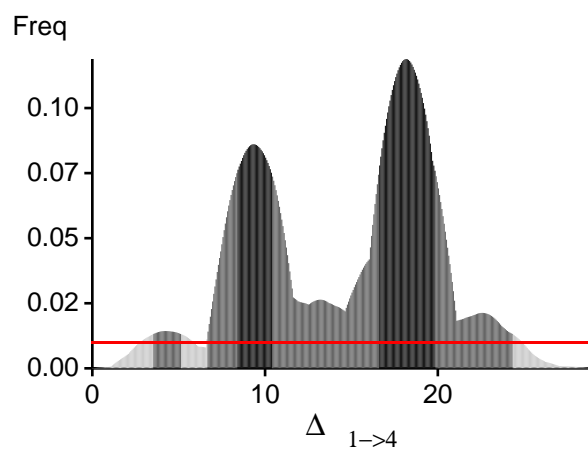
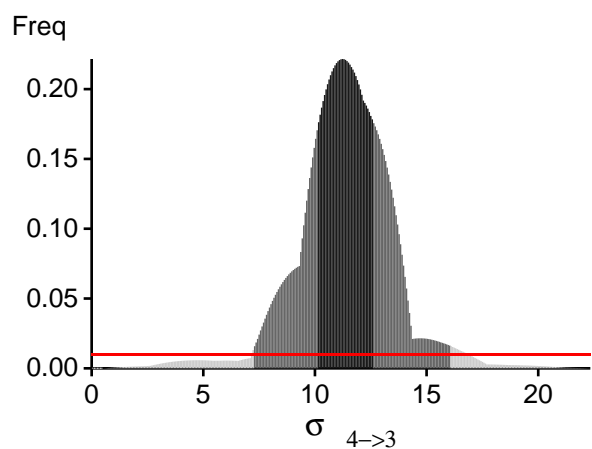
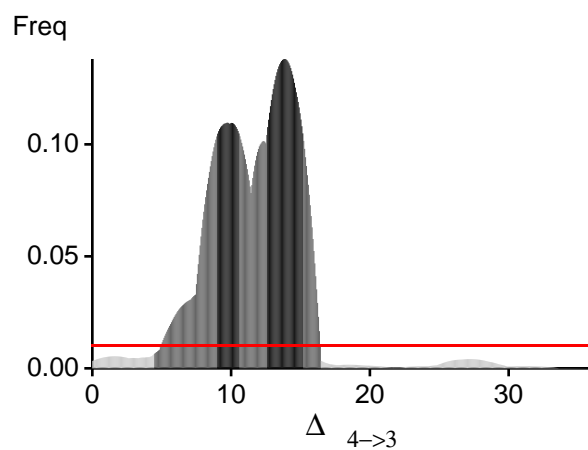
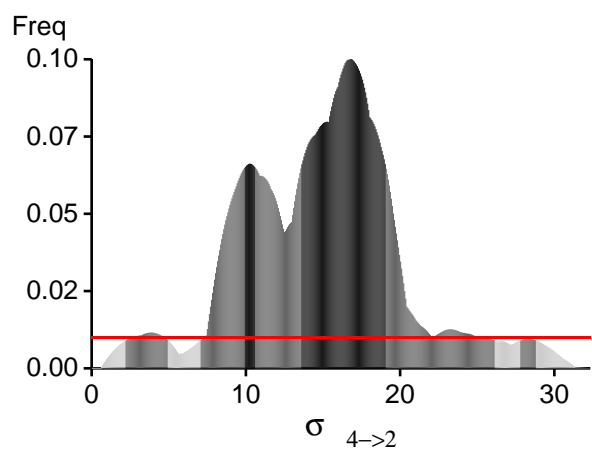
Citation suggestions:

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	-7974.50	-7674.00	-10133.41
2	-16546.93	-15783.51	-21657.35
All	-24514.61	-23450.68	-31783.93

(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 = 6.828643

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	105/105	1.00000
Θ_2	85/85	1.00000
Θ_3	66/66	1.00000
Θ_4	80/80	1.00000
$M_{4 \rightarrow 2}$	90/90	1.00000
$\Delta_{4 \rightarrow 2}$	94/95	0.98947
$\sigma_{4 \rightarrow 2}$	85/91	0.93407
$\Delta_{4 \rightarrow 3}$	76/93	0.81720
$\sigma_{4 \rightarrow 3}$	72/86	0.83721
$\Delta_{1 \rightarrow 4}$	94/111	0.84685
$\sigma_{1 \rightarrow 4}$	77/99	0.77778
Genealogies	247/999	0.24725

MCMC-Autocorrelation and Effective MCMC Sample Size

Parameter	Autocorrelation	Effective Sample Size
Θ_1	0.97066	27.08
Θ_2	0.94557	39.08
Θ_3	0.97335	23.24
Θ_4	0.97331	19.62
$M_{4 \rightarrow 2}$	0.94994	43.74
$\text{Ln}[\text{Prob}(D G)]$	0.98454	14.16

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 4: Effective sample size of run seems too short!