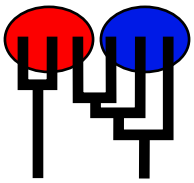
Example: sequence data set wit two loci [simula

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 14:04:02 2015
Program finished at Sun Feb 22 14:08:31 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	$M_{4\text{->}2}$	<displayed>
6	$\Delta_{4\text{->}2}$	<displayed>
7	$\sigma_{4\text{->}2}$	<displayed>
8	$\Delta_{4\text{->}3}$	<displayed>
9	$\sigma_{4\text{->}3}$	<displayed>
10	$\Delta_{1\text{->}4}$	<displayed>
11	$\sigma_{1\text{->}4}$	<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	0.050000	0.100000	0.010000	500
Theta	Uniform	0.000000	0.050000	0.100000	0.010000	500
Theta	Uniform	0.000000	0.050000	0.100000	0.010000	500
Theta	Uniform	0.000000	0.050000	0.100000	0.010000	500
M	Uniform	0.000000	10000.000000	20000.000000	2000.000000	500
M	Uniform	0.000000	10000.000000	20000.000000	2000.000000	500
M	Uniform	0.000000	10000.000000	20000.000000	2000.000000	500
M	Uniform	0.000000	10000.000000	20000.000000	2000.000000	500
M	Uniform	0.000000	10000.000000	20000.000000	2000.000000	500
M	Uniform	0.000000	10000.000000	20000.000000	2000.000000	500
M	Uniform	0.000000	10000.000000	20000.000000	2000.000000	500
M	Uniform	0.000000	10000.000000	20000.000000	2000.000000	500
M	Uniform	0.000000	10000.000000	20000.000000	2000.000000	500
M	Uniform	0.000000	10000.000000	20000.000000	2000.000000	500
M	Uniform	0.000000	10000.000000	20000.000000	2000.000000	500
M	Uniform	0.000000	10000.000000	20000.000000	2000.000000	500
Splittime mean	Uniform	0.000000	0.050000	0.100000	0.010000	1500
Splittime std	Uniform	0.000000	0.500000	1.000000	0.100000	1500
Splittime mean	Uniform	0.000000	0.050000	0.100000	0.010000	1500
Splittime std	Uniform	0.000000	0.500000	1.000000	0.100000	1500

Splittime mean	Uniform	0.000000	0.050000	0.100000	0.010000	1500
Splittime std	Uniform	0.000000	0.500000	1.000000	0.100000	1500

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]	2
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.seq
Haplotyping is turned on:	NO
Output file:	outfile-seq
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

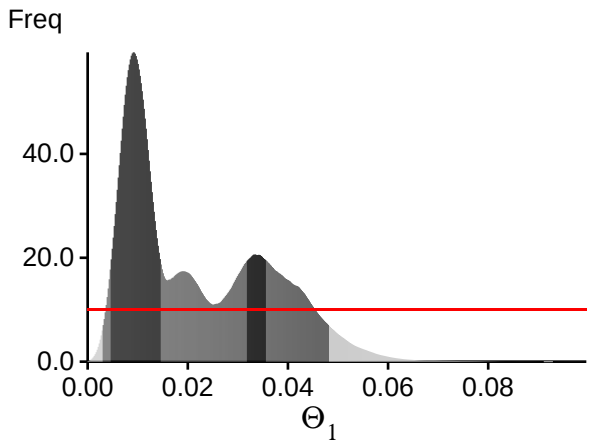
Locus	Parameter	2.5%	25.0%	Mode	75.0%	97.5%	Median	Mean
1	Θ_1	0.01600	0.02440	0.03010	0.03640	0.05120	0.03210	0.03293
1	Θ_2	0.00000	0.00180	0.00530	0.01460	0.06900	0.01330	0.01972
1	Θ_3	0.00040	0.00440	0.00790	0.01180	0.02620	0.01010	0.01152
1	Θ_4	0.00600	0.01280	0.01670	0.03080	0.05840	0.02690	0.02909
1	$M_{4->2}$	0.000	0.000	780.000	1800.000	11520.000	1820.000	3340.451
1	$D_{4->2}$	0.06933	0.07320	0.08003	0.08660	0.09193	0.04923	0.04948
1	$S_{4->2}$	0.40000	0.41867	0.44300	0.49467	0.50533	0.50233	0.50300
1	$D_{4->3}$	0.00000	0.00000	0.00203	0.00540	0.06853	0.00543	0.01861
1	$S_{4->3}$	0.00000	0.00000	0.00767	0.02667	0.10000	0.02700	0.18866
1	$D_{1->4}$	0.00000	0.00113	0.00250	0.00507	0.01073	0.00423	0.00487
1	$S_{1->4}$	0.00000	0.00200	0.00967	0.01733	0.03400	0.01500	0.01884
2	Θ_1	0.00320	0.00740	0.01030	0.01360	0.02140	0.01150	0.01190
2	Θ_2	0.00120	0.00180	0.00570	0.03780	0.09120	0.04270	0.04513
2	Θ_3	0.00000	0.00280	0.00570	0.00840	0.01700	0.00690	0.00761
2	Θ_4	0.00000	0.00340	0.00570	0.00800	0.01380	0.00650	0.00671
2	$M_{4->2}$	4000.000	5120.000	7140.000	12160.000	19680.000	10660.000	11190.433
2	$D_{4->2}$	0.00147	0.01640	0.02037	0.02560	0.04567	0.05023	0.05005
2	$S_{4->2}$	0.01667	0.27333	0.31567	0.34667	0.82067	0.49767	0.49691
2	$D_{4->3}$	0.00000	0.00013	0.00117	0.00213	0.02087	0.00197	0.00587
2	$S_{4->3}$	0.00000	0.00000	0.00300	0.01267	0.06133	0.01300	0.05338
2	$D_{1->4}$	0.00000	0.00053	0.00190	0.00320	0.02087	0.00277	0.00633
2	$S_{1->4}$	0.00000	0.00000	0.00567	0.01333	0.08067	0.01367	0.05032
All	Θ_1	0.00280	0.00440	0.00910	0.01460	0.04820	0.01790	0.02207
All	Θ_2	0.00000	0.00180	0.00530	0.01300	0.05940	0.01170	0.01690
All	Θ_3	0.00060	0.00420	0.00670	0.00900	0.01420	0.00730	0.00746
All	Θ_4	0.00000	0.00160	0.00570	0.01300	0.05240	0.01210	0.01904
All	$M_{4->2}$	0.000	0.000	740.000	2040.000	12800.000	2220.000	4750.852
All	$D_{4->2}$	0.00127	0.01507	0.01877	0.02207	0.04520	0.04963	0.04965
All	$S_{4->2}$	0.26400	0.27400	0.30367	0.33867	0.40067	0.50033	0.50053
All	$D_{4->3}$	0.00000	0.00027	0.00117	0.00200	0.00473	0.00170	0.00222
All	$S_{4->3}$	0.00000	0.00000	0.00300	0.01067	0.02467	0.01100	0.00610
All	$D_{1->4}$	0.00000	0.00067	0.00190	0.00300	0.00607	0.00243	0.00260
All	$S_{1->4}$	0.00000	0.00000	0.00633	0.01200	0.02667	0.01233	0.00700

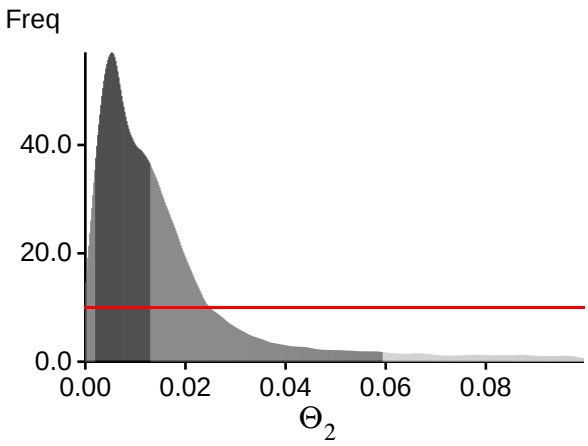
Citation suggestions:

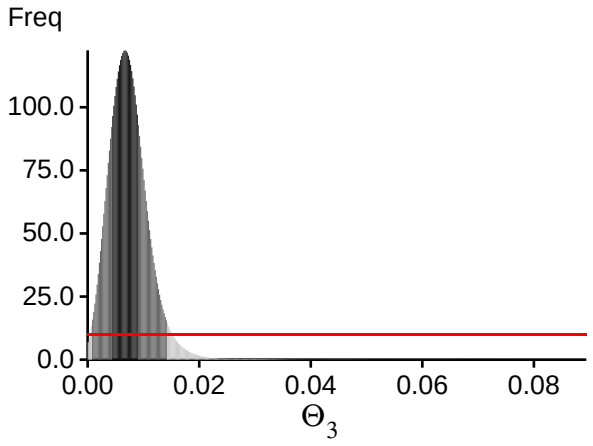
Beerli P., 2006. Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. Bioinformatics 22:341-345

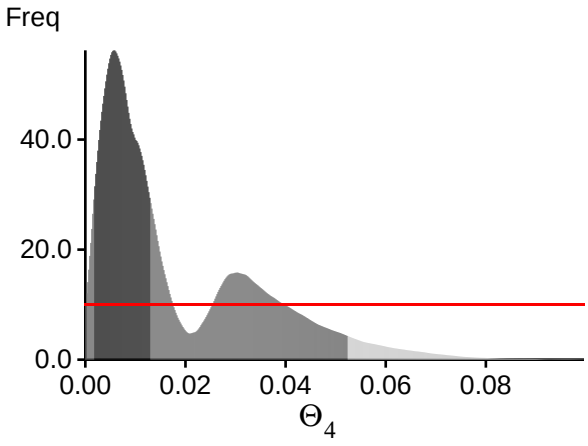
Beerli P., 2009. How to use MIGRATE or why are Markov chain Monte Carlo programs difficult to use? In Population Genetics for Animal Conservation, G. Bertorelle, M. W. Bruford, H. C. Hauffe, A. Rizzoli, and C. Vernesi, eds., vol. 17 of Conservation Biology, Cambridge University Press, Cambridge UK, pp. 42-79.

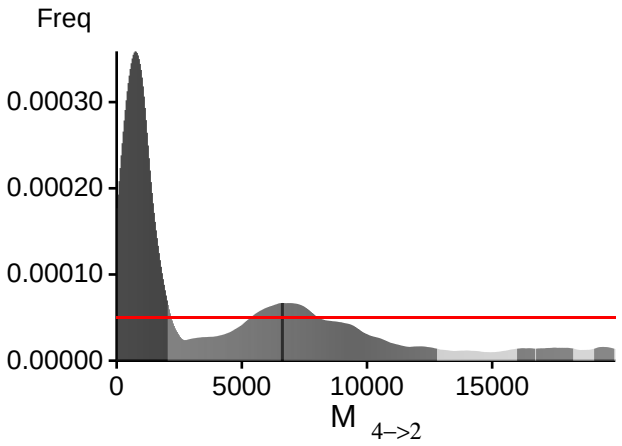
Bayesian Analysis: Posterior distribution over all loci

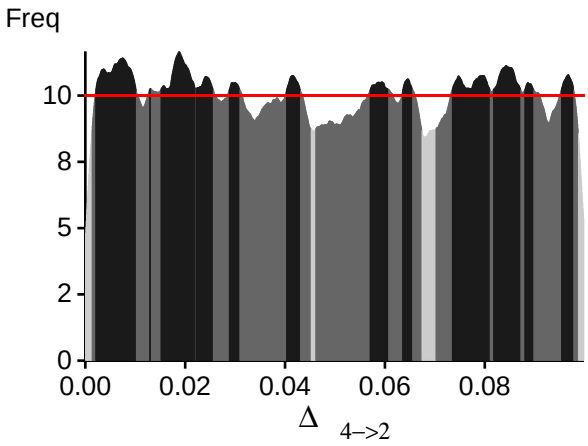


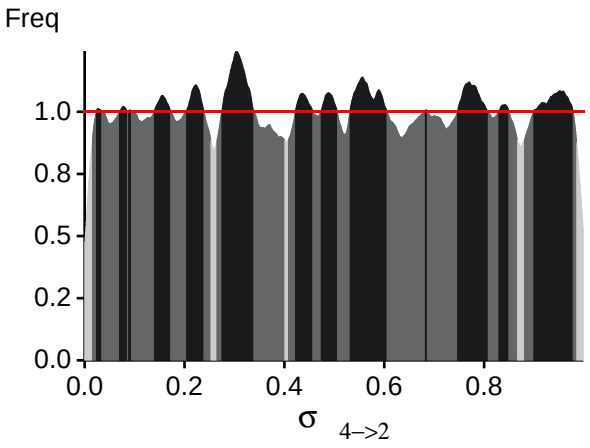


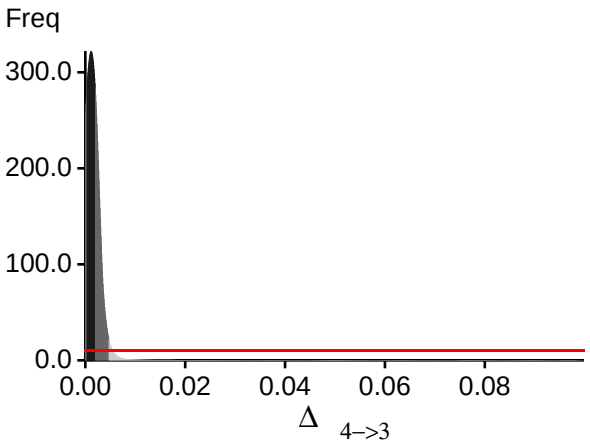


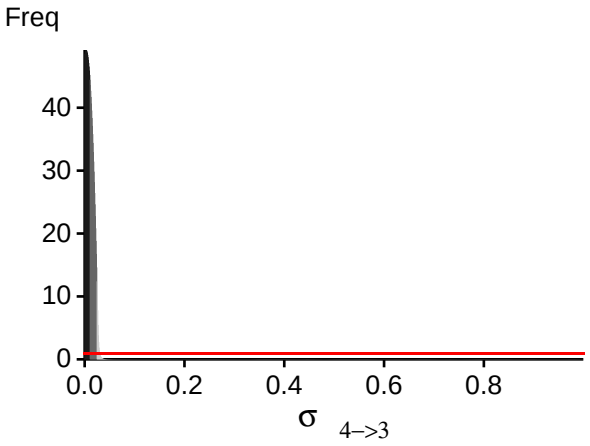


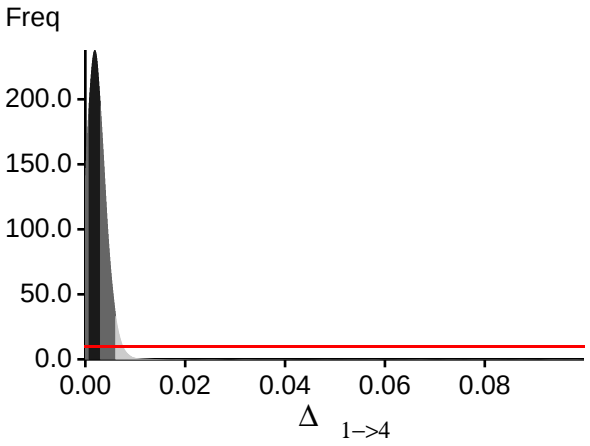


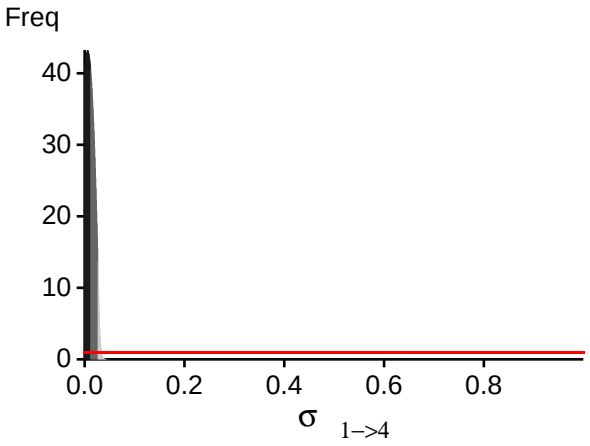



Freq
1.0
0.8
0.5
0.2
0.0
0.0
0.2
0.4
0.6
0.8
σ 4–>2


Freq
300.0
200.0
100.0
0.0
0.00
0.02
0.04
0.06
0.08
Δ 4–>3


Freq
40
30
20
10
0
0.0
0.2
0.4
0.6
0.8
σ 4–>3


Freq
200.0
150.0
100.0
50.0
0.0
0.00
0.02
0.04
0.06
0.08
Δ 1–>4


Freq
40
30
20
10
0
0.0
0.2
0.4
0.6
0.8
σ 1–>4

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	-970.92	-568.59	-451.37
2	-887.84	-733.75	-657.62
All	-1855.97	-1299.54	-1106.20

(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 = 2.791429
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	91122/91122	1.00000
Θ_2	91414/91414	1.00000
Θ_3	90868/90868	1.00000
Θ_4	90370/90370	1.00000
$M_{4->2}$	90470/90470	1.00000
$\Delta_{4->2}$	90744/90964	0.99758
$\sigma_{4->2}$	90499/91042	0.99404
$\Delta_{4->3}$	25702/90849	0.28291
$\sigma_{4->3}$	20197/91060	0.22180
$\Delta_{1->4}$	14563/90769	0.16044
$\sigma_{1->4}$	9768/90947	0.10740
Genealogies	168181/1000125	0.16816

MCMC-Autocorrelation and Effective MCMC Sample Size

Parameter	Autocorrelation	Effective Sampe Size
Θ_1	0.59491	5097.36
Θ_2	0.54448	6149.47
Θ_3	0.71852	2701.75
Θ_4	0.77297	2743.94
$M_{4->2}$	0.81479	2243.67
Ln[Prob(D|G)]	0.19277	13393.60

Potential Problems

This section reports potential problems with your run, but such reporting is often not very accurate. Whith many parameters in a multilocus analysi s, it is very common that some parameters for some loci will not be very informative, triggering suggestions (for example to increase the prior ran ge) that are not sensible. This suggestion tool will improve with time, therefore do not blindly follow its suggestions. If some parameters are fla gged, inspect the tables carefully and judge wether an action is required. For example, if you run a Bayesian inference with sequence data, for mac roscopic 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 rou tes 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