Current protocols example dataset

POPULATION SIZE, MIGRATION, DIVERGENCE, ASSIGNMENT, HISTORY
Bayesian inference using the structured coalescent
Migrate-n version 4.4.3(git:) [March-21-2019]
Using Intel AVX (Advanced Vector Extensions)
Compiled for PARALLEL computer architectures
One master and 8 compute nodes are available.
Compiled for a SYMMETRIC multiprocessors (Grandcentral)
Program started at Mon Jun 3 08:53:19 2019
Program finished at Mon Jun 3 08:54:44 2019 [Runtime:0000:00:01:25]

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) 448135000
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	1	2
1 Arbon_1	*	*	0
1 Berg_2	*	*	0
2 Chur_3	d	d	*

Order of parameters:

1	Θ_1	<displayed>
2	Θ_2	<displayed>
3	$\Delta_{1\text{->}2}$	<displayed>
4	$\sigma_{1\text{->}2}$	<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 Shortcut (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	1500	0.12500
2	Theta	**	Uniform	0.000000	0.050	0.100	0.010	1500	0.12500
3	Splittime mean	**	Uniform	0.000000	0.050	0.100	0.010	1500	0.12500
4	Splittime std	**	Uniform	0.000000	0.050	0.100	0.100	1500	0.12500

[* * means priors were set globally]

Markov chain settings: Long chain

Number of chains	1
Recorded steps [a]	5000
Increment (record every x step [b]	10
Number of concurrent chains (replicates) [c]	1
Visited (sampled) parameter values [a*b*c]	50000
Number of discard trees per chain (burn-in)	5000

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
Haplotyping is turned on:	NO
Output file:	outfile_model3
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	1000
2	1000
3	1000
4	1000
5	1000
6	1000
7	1000
8	1000
9	1000
10	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
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 Arbon_1	1	10	
	2	10	
	3	10	
	4	10	
	5	10	
	6	10	
	7	10	
	8	10	
	9	10	
	10	10	
1 Berg_2	1	10	
	2	10	
	3	10	
	4	10	
	5	10	
	6	10	
	7	10	
	8	10	
	9	10	
	10	10	
2 Chur_3	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	Θ_1	0.00647	0.01033	0.01057	0.01067	0.01267	0.01030	0.01049
1	Θ_2	0.02153	0.02187	0.02210	0.02220	0.02247	0.02463	0.02691
1	$D_{1->2}$	0.03393	0.03460	0.03470	0.03473	0.03567	0.05403	0.05332
1	$S_{1->2}$	0.03307	0.03307	0.03317	0.03333	0.03333	0.06043	0.05914
2	Θ_1	0.00767	0.00987	0.01023	0.01047	0.01413	0.01170	0.01231
2	Θ_2	0.00193	0.00513	0.00550	0.00567	0.01027	0.00670	0.00781
2	$D_{1->2}$	0.01267	0.01267	0.01283	0.01347	0.01347	0.03717	0.04140
2	$S_{1->2}$	0.03327	0.03333	0.03343	0.03353	0.03393	0.06250	0.06035
3	Θ_1	0.00680	0.00780	0.00790	0.00793	0.01520	0.01123	0.01199
3	Θ_2	0.00713	0.01200	0.01217	0.01240	0.02713	0.01877	0.02121
3	$D_{1->2}$	0.03787	0.03833	0.03843	0.03847	0.03967	0.04563	0.04798
3	$S_{1->2}$	0.06280	0.06280	0.06290	0.06293	0.06367	0.06197	0.06100
4	Θ_1	0.00520	0.00620	0.00630	0.00633	0.01153	0.00823	0.00881
4	Θ_2	0.01767	0.01927	0.01937	0.01947	0.02247	0.02377	0.02627
4	$D_{1->2}$	0.07467	0.07513	0.07523	0.07527	0.07680	0.05437	0.05381
4	$S_{1->2}$	0.05000	0.05007	0.05017	0.05020	0.05233	0.06150	0.06013
5	Θ_1	0.00287	0.00467	0.00490	0.00527	0.00987	0.00630	0.00664
5	Θ_2	0.02453	0.02573	0.02590	0.02593	0.02807	0.03143	0.03445
5	$D_{1->2}$	0.04487	0.04487	0.04497	0.04500	0.04500	0.05030	0.05124
5	$S_{1->2}$	0.07060	0.07060	0.07070	0.07073	0.07120	0.05877	0.05747
6	Θ_1	0.00393	0.00600	0.00670	0.00687	0.01140	0.00763	0.00792
6	Θ_2	0.03927	0.04060	0.04097	0.04113	0.04140	0.04437	0.04718
6	$D_{1->2}$	0.03413	0.03433	0.03450	0.03453	0.03513	0.04417	0.04732
6	$S_{1->2}$	0.05467	0.05467	0.05477	0.05480	0.05507	0.05657	0.05491
7	Θ_1	0.00880	0.01180	0.01203	0.01207	0.01960	0.01423	0.01478
7	Θ_2	0.01960	0.02060	0.02070	0.02073	0.02073	0.01863	0.02029
7	$D_{1->2}$	0.07127	0.07140	0.07150	0.07153	0.07153	0.04103	0.04433
7	$S_{1->2}$	0.02147	0.02147	0.02157	0.02160	0.02167	0.06123	0.05940
8	Θ_1	0.00613	0.01073	0.01130	0.01167	0.01653	0.01110	0.01157

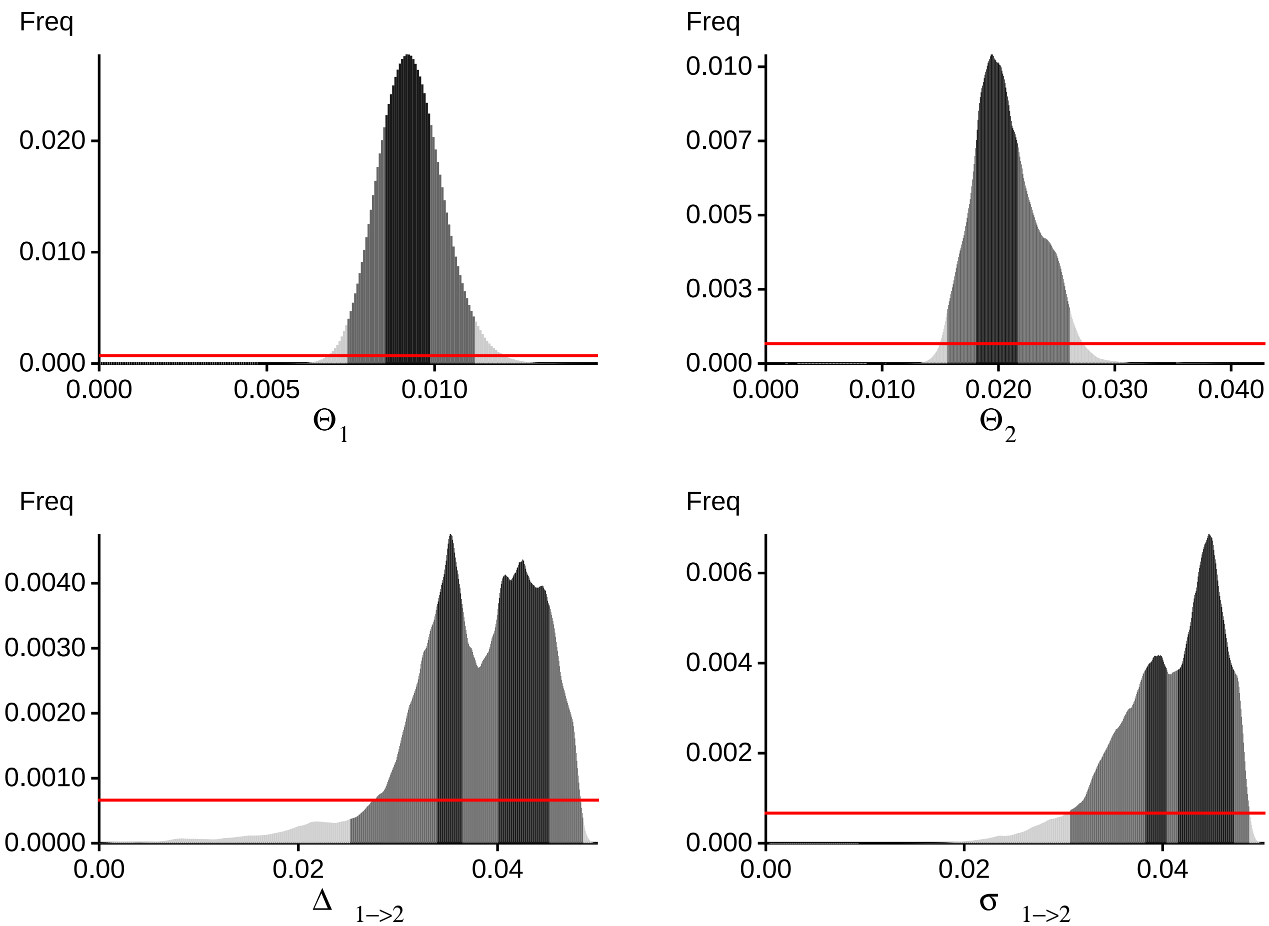
Locus	Parameter	2.5%	25.0%	Mode	75.0%	97.5%	Median	Mean
8	Θ_2	0.01393	0.01967	0.02003	0.02020	0.02353	0.02077	0.02342
8	$D_{1->2}$	0.04053	0.04073	0.04083	0.04087	0.04127	0.05457	0.05383
8	$S_{1->2}$	0.09380	0.09407	0.09423	0.09433	0.09433	0.06110	0.06032
9	Θ_1	0.00680	0.01120	0.01130	0.01133	0.01613	0.01270	0.01332
9	Θ_2	0.03027	0.03147	0.03157	0.03160	0.03213	0.02603	0.02842
9	$D_{1->2}$	0.04053	0.04053	0.04063	0.04067	0.04073	0.05277	0.05251
9	$S_{1->2}$	0.06340	0.06340	0.06370	0.06373	0.06373	0.06157	0.06053
10	Θ_1	0.00573	0.00813	0.00850	0.00933	0.01420	0.00970	0.01009
10	Θ_2	0.01880	0.01973	0.01983	0.01993	0.02513	0.02450	0.02720
10	$D_{1->2}$	0.03173	0.03227	0.03237	0.03240	0.03260	0.04503	0.04738
10	$S_{1->2}$	0.03640	0.03640	0.03650	0.03653	0.03653	0.05990	0.05889
All	Θ_1	0.00733	0.00847	0.00917	0.00987	0.01120	0.00930	0.00929
All	Θ_2	0.01553	0.01800	0.01943	0.02167	0.02613	0.02037	0.02067
All	$D_{1->2}$	0.02513	0.03387	0.03530	0.03647	0.04860	0.03857	0.03771
All	$S_{1->2}$	0.03060	0.04147	0.04470	0.04720	0.04873	0.04190	0.04076

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 = 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	-2506.18	-2314.99	-2295.89
2	-2119.16	-1981.24	-1964.93
3	-2301.34	-2128.28	-2114.57
4	-2498.34	-2265.25	-2233.61
5	-2367.72	-2197.47	-2175.63
6	-2550.28	-2329.28	-2298.39
7	-2445.66	-2225.63	-2198.12
8	-2530.41	-2263.81	-2230.64
9	-2512.08	-2297.33	-2270.43
10	-2370.38	-2175.63	-2153.37
All	-24193.48	-22170.83	-21927.50

(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 = 8.073502]
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	27257/62452	0.43645
Θ_2	27796/62747	0.44299
$\Delta_{1->2}$	55663/62737	0.88724
$\sigma_{1->2}$	52896/62735	0.84317
Genealogies	35164/249329	0.14103

MCMC-Autocorrelation and Effective MCMC Sample Size

Parameter	Autocorrelation	Effective Sampe Size
Θ_1	0.79500	7894.15
Θ_2	0.73910	9817.45
$\Delta_{1->2}$	0.62475	16182.59
$\sigma_{1->2}$	0.60649	17450.84
Genealogies	0.60649	17450.84

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.

Param 6 (Locus 8): Upper prior boundary seems too low!