

test snp data

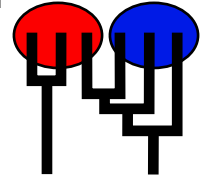
POPULATION SIZE, MIGRATION, DIVERGENCE, ASSIGNMENT, HISTORY

Bayesian inference using the structured coalescent

Migrate-n version 5.0.2(git:v5.0.2-0-gb89fa60-dirty) [December-1-2021]

Program started at Sat Feb 5 10:29:38 2022

Program finished at Sat Feb 5 10:31:35 2022 [Runtime:0000:00:01:57]



Options

Datatype: Single nucleotide polymorphism data

Inheritance multipliers in use for Thetas:

All loci use an inheritance multiplier of 1.0

Random number seed: (with internal timer) 3363151493

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
1 pop1	*	*
2 pop2	*	*

Order of parameters:

1	Θ_1	<displayed>
2	Θ_2	<displayed>
3	$M_{2 \rightarrow 1}$	<displayed>
4	$M_{1 \rightarrow 2}$	<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

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	M ** Uniform	0.000000	500.0	1000.	100.0	1500	0.12500
4	M ** Uniform	0.000000	500.0	1000.	100.0	1500	0.12500

[* * means priors were set globally]

Markov chain settings:

Long chain

Number of chains

1

Recorded steps [a]

10000

Increment (record every x step [b])

100

Number of concurrent chains (replicates) [c]

1

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.snpold

parmfile.snp

Haplotyping is turned on:

NO

Output file:

outfile_snp

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.snpold
 Datatype: SNP data
 Number of loci: 3

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]

Sites per locus

Locus	Sites
1	1
2	1
3	1

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

Population	Locus	Gene copies
1 pop1	1	2
	2	2
	3	2
2 pop2	1	3
	2	3
	3	3
Total of all populations	1	5
	2	5
	3	5

Allele frequency spectra

Locus 1

Allele	Pop1	Pop2	All
A	1.000	0.667	0.800
T	-	0.333	0.200
	-	-	0.000
	-	-	0.000
Alleles	0.000	0.444	0.320
Samplesize	2	3	5
H_{exp}			

Locus 2

Allele	Pop1	Pop2	All
A	0.500	0.667	0.600
W	0.500	-	0.200
T	-	0.333	0.200
	-	-	0.000
Alleles	2	2	4
Samplesize	0.500	0.444	0.560
H_{exp}			

Locus 3

Allele	Pop1	Pop2	All
A	1.000	0.333	0.600
R	-	0.667	0.400
	-	-	0.000
	-	-	0.000
Alleles	0.000	0.444	0.480
Samplesize	2	3	5
H_{exp}			

Average expected heterozygosity

	Pop1	Pop2	All
H_{exp}	0.167	0.444	0.453

Bayesian Analysis: Posterior distribution table

Locus	Parameter	2.5%	25.0%	Mode	75.0%	97.5%	Median	Mean
1	Θ_1	0.01120	0.01927	0.03530	0.06627	0.09893	0.05143	0.05230
1	Θ_2	0.01327	0.02227	0.03183	0.07953	0.09947	0.05383	0.05411
1	$M_{2 \rightarrow 1}$	136.7	187.3	348.3	868.7	889.3	507.7	516.2
1	$M_{1 \rightarrow 2}$	136.7	310.0	401.0	740.0	936.7	526.3	528.8
2	Θ_1	0.01220	0.02807	0.04537	0.05360	0.09880	0.05197	0.05288
2	Θ_2	0.01307	0.02007	0.03450	0.06847	0.09887	0.05217	0.05305
2	$M_{2 \rightarrow 1}$	317.3	438.0	475.0	554.7	648.0	501.0	512.7
2	$M_{1 \rightarrow 2}$	227.3	300.0	469.0	678.0	758.0	496.3	500.2
3	Θ_1	0.00513	0.01893	0.02350	0.03447	0.08433	0.04757	0.04891
3	Θ_2	0.00647	0.00720	0.03003	0.07300	0.08440	0.04703	0.04897
3	$M_{2 \rightarrow 1}$	152.0	247.3	461.0	779.3	974.7	529.7	534.4
3	$M_{1 \rightarrow 2}$	144.0	346.7	395.7	601.3	986.7	531.0	536.9
All	Θ_1	0.01267	0.01920	0.03283	0.05380	0.09807	0.04803	0.05018
All	Θ_2	0.01467	0.02233	0.03443	0.04747	0.09900	0.04863	0.05088
All	$M_{2 \rightarrow 1}$	256.7	434.7	471.7	555.3	673.3	493.0	502.7
All	$M_{1 \rightarrow 2}$	192.0	374.7	508.3	575.3	844.7	500.3	508.1

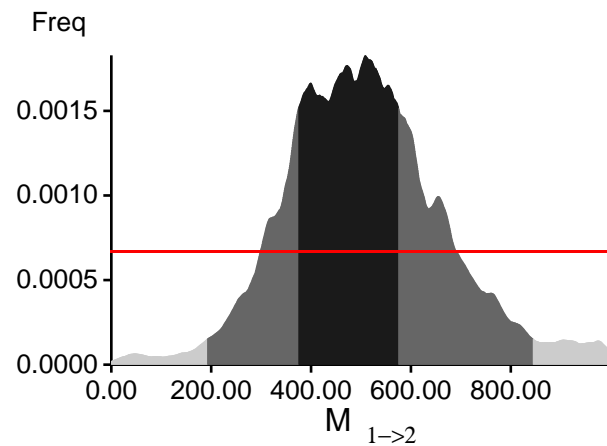
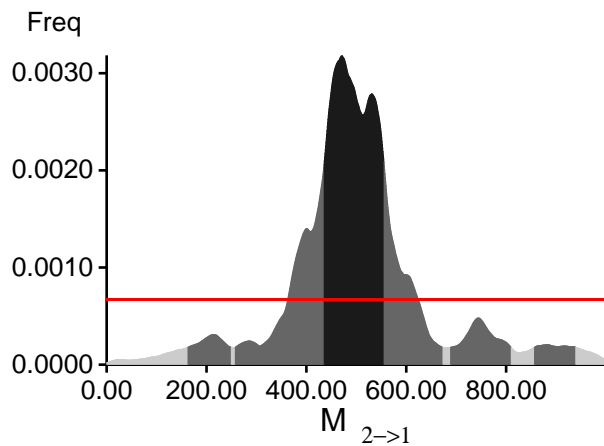
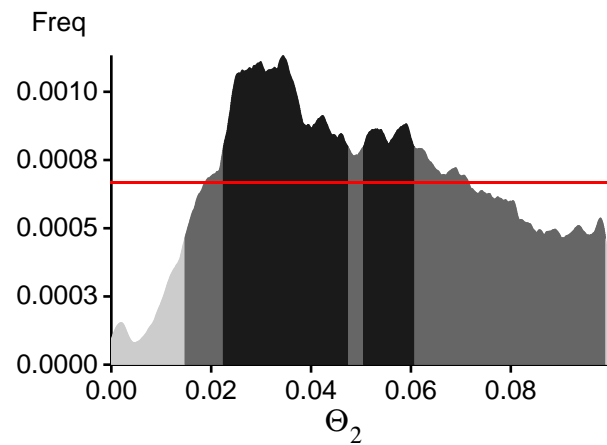
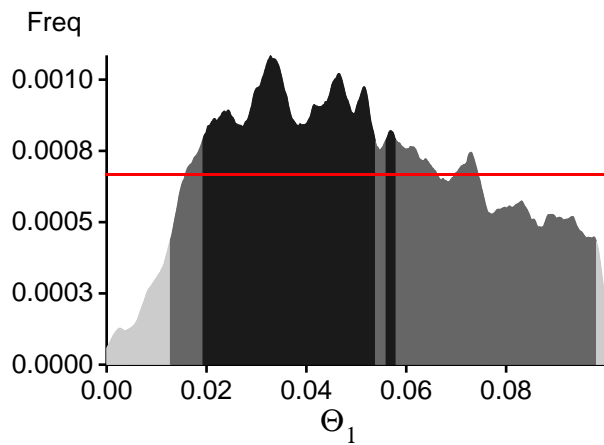
Citation suggestions:

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	TI(1a)	BTI(1b)	HS(3)
1	-5.12	-5.00	-5.10
2	-4.90	-4.78	-4.89
3	-0.09	-0.08	-0.09
All	-28.70	-28.47	-28.69

(1a) TI: Thermodynamic integration: $\log(\text{Prob}(D|\text{Model}))$: Good approximation with many temperatures

(1b) BTI: Bezier-approximated Thermodynamic integration: when using few temperatures USE THIS!

(2) SS: Steppingstone Sampling (Xie et al 2011)

(3) HS: Harmonic mean approximation: Overestimates the marginal likelihood, poor variance

[Scaling factor = -18.604912]

Citation suggestions:

Beerli P. and M. Palczewski, 2010. Unified framework to evaluate panmixia and migration direction among multiple sampling locations, *Genetics*, 185: 313-326.

Palczewski M. and P. Beerli, 2014. Population model comparison using multi-locus datasets.

In M.-H. Chen, L. Kuo, and P. O. Lewis, editors, *Bayesian Phylogenetics: Methods, Algorithms, and Applications*, pages 187-200. CRC Press, 2014.

Xie W., P. O. Lewis, Y. Fan, L. Kuo, and M.-H. Chen. 2011. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Systematic Biology*, 60(2):150â 160, 2011.

Acceptance ratios for all parameters and the genealogies

Parameter	Accepted changes	Ratio
Θ_1	273360/375011	0.72894
Θ_2	273824/374814	0.73056
$M_{2 \rightarrow 1}$	224316/375292	0.59771
$M_{1 \rightarrow 2}$	218623/374238	0.58418
Genealogies	1211850/1500645	0.80755

MCMC-Autocorrelation and Effective MCMC Sample Size

Parameter	Autocorrelation	Effective Sampe Size
Θ_1	0.10171	24478.73
Θ_2	0.08825	25150.44
$M_{2 \rightarrow 1}$	0.69326	6002.67
$M_{1 \rightarrow 2}$	0.69167	5993.65
Genealogies	0.10171	24478.73

Average temperatures during the run

Chain	Temperatures
-------	--------------

1	1.00000
2	1.50000
3	3.00000
4	1000000.00000

Adaptive heating often fails, if the average temperatures are very close together try to rerun using static heating! If you want to compare models using marginal likelihoods then you **MUST** use static heating

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