

# *Prosopium williamsoni dataset*

POPULATION SIZE, MIGRATION, DIVERGENCE, ASSIGNMENT, HISTORY

Bayesian inference using the structured coalescent

Migrate-n version 4.4.4(git:v4-series-26-ge85c6ff) [June-1-2019]

Compiled for a SYMMETRIC multiprocessors (Grandcentral)

Program started at Sat May 22 15:33:55 2021

Program finished at Sun May 23 00:39:41 2021 [Runtime:0000:09:05:46]



## *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) 4120730003

Start parameters:

Theta values were generated RANDOM start value from the prior

M values were generated RANDOM start value from 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
1 Logan_River	*	*	*
2 Weber_River	*	*	*
3 Provo_River	*	*	*

Order of parameters:

1	$\Theta_1$	<displayed>
2	$\Theta_2$	<displayed>
3	$\Theta_3$	<displayed>

4	M	2->1	<displayed>
5	M	3->1	<displayed>
6	M	1->2	<displayed>
7	M	3->2	<displayed>
8	M	1->3	<displayed>
9	M	2->3	<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	0.010	0.100	0.010	1500	0.05556
2	Theta ** Uniform	0.000000	0.010	0.100	0.010	1500	0.05556
3	Theta ** Uniform	0.000000	0.010	0.100	0.010	1500	0.05556
4	M ** Uniform	0.000000	100.0	1000.	100.0	1500	0.05556
5	M ** Uniform	0.000000	100.0	1000.	100.0	1500	0.05556
6	M ** Uniform	0.000000	100.0	1000.	100.0	1500	0.05556
7	M ** Uniform	0.000000	100.0	1000.	100.0	1500	0.05556
8	M ** Uniform	0.000000	100.0	1000.	100.0	1500	0.05556
9	M ** Uniform	0.000000	100.0	1000.	100.0	1500	0.05556

[\* \* means priors were set globally]

Markov chain settings:

Long chain

Number of chains

1

Recorded steps [a]

10000

Increment (record every x step [b])

1000

Number of concurrent chains (replicates) [c]

2

Visited (sampled) parameter values [a\*b\*c]

20000000

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
Haplotyping is turned on:	NO
Output file:	outfile
Log file:	logfile
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: 2

### Mutationmodel:

Locus	Sublocus	Mutationmodel	Mutationmodel parameters
1	1	HKY	[Bf:0.23 0.29 0.18 0.30, kappa=3.010]
2	1	HKY	[Bf:0.25 0.31 0.17 0.27, kappa=3.010]

### Sites per locus

Locus	Sites
1	1161
2	1050

### 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 Logan_River			1	25	
			2	25	
2 Weber_River			1	28	
			2	35	
3 Provo_River			1	16	
			2	25	
Total of all populations			1	69	(0)
			2	85	(0)

## *Bayesian Analysis: Posterior distribution table*

Locus	Parameter	2.5%	25.0%	Mode	75.0%	97.5%	Median	Mean
1	$\Theta_1$	0.00013	0.00020	0.00037	0.00047	0.00073	0.00050	0.00043
1	$\Theta_2$	0.00013	0.00027	0.00043	0.00053	0.00080	0.00050	0.00047
1	$\Theta_3$	0.00000	0.00007	0.00023	0.00027	0.00053	0.00030	0.00027
1	$M_{2 \rightarrow 1}$	712.667	912.667	916.333	916.667	999.333	587.667	563.304
1	$M_{3 \rightarrow 1}$	287.333	289.333	290.333	290.667	290.667	519.667	511.429
1	$M_{1 \rightarrow 2}$	514.000	516.000	517.667	518.000	525.333	527.000	519.406
1	$M_{3 \rightarrow 2}$	58.667	71.333	73.000	73.333	74.667	500.333	499.870
1	$M_{1 \rightarrow 3}$	466.667	496.000	497.000	497.333	511.333	557.000	538.305
1	$M_{2 \rightarrow 3}$	870.667	885.333	887.000	888.000	964.667	592.333	565.886
2	$\Theta_1$	0.00027	0.00040	0.00057	0.00073	0.00113	0.00070	0.00068
2	$\Theta_2$	0.00040	0.00060	0.00083	0.00100	0.00153	0.00097	0.00095
2	$\Theta_3$	0.00027	0.00040	0.00063	0.00073	0.00120	0.00070	0.00070
2	$M_{2 \rightarrow 1}$	764.000	840.000	841.667	842.000	854.667	561.667	545.810
2	$M_{3 \rightarrow 1}$	497.333	873.333	875.000	875.333	999.333	633.000	600.932
2	$M_{1 \rightarrow 2}$	706.667	754.000	757.000	759.333	901.333	581.000	561.927
2	$M_{3 \rightarrow 2}$	864.000	870.000	871.000	873.333	880.667	525.000	517.441
2	$M_{1 \rightarrow 3}$	632.667	964.667	968.333	971.333	999.333	669.667	630.953
2	$M_{2 \rightarrow 3}$	434.000	958.667	988.333	999.333	999.333	759.000	715.309
All	$\Theta_1$	0.00000	0.00000	0.00057	0.00107	0.00227	0.00103	0.00104
All	$\Theta_2$	0.00000	0.00013	0.00070	0.00127	0.00233	0.00110	0.00110
All	$\Theta_3$	0.00000	0.00000	0.00050	0.00093	0.00220	0.00097	0.00099
All	$M_{2 \rightarrow 1}$	50.000	333.333	475.000	484.667	498.667	313.667	297.203
All	$M_{3 \rightarrow 1}$	59.333	390.667	475.667	486.667	498.667	313.667	296.458
All	$M_{1 \rightarrow 2}$	46.667	301.333	475.000	487.333	497.333	311.000	292.104
All	$M_{3 \rightarrow 2}$	56.000	264.000	415.667	481.333	486.667	266.333	259.380
All	$M_{1 \rightarrow 3}$	71.333	328.667	475.000	488.000	501.333	339.000	315.769
All	$M_{2 \rightarrow 3}$	145.333	380.667	475.667	489.333	504.667	388.333	361.413

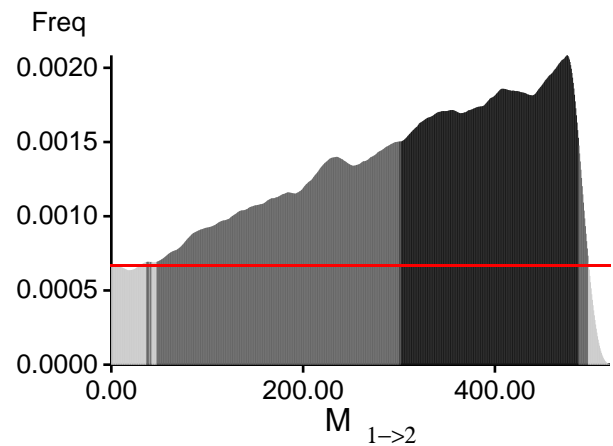
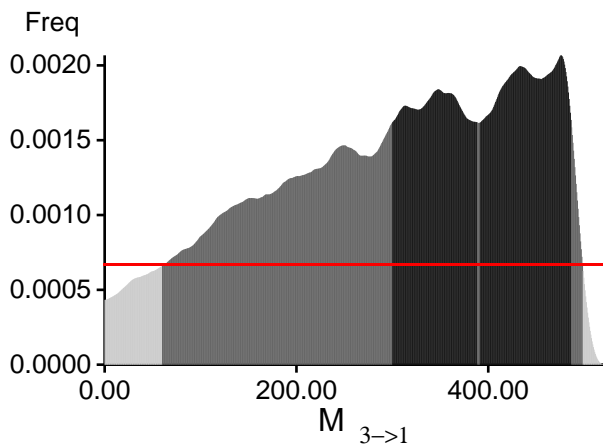
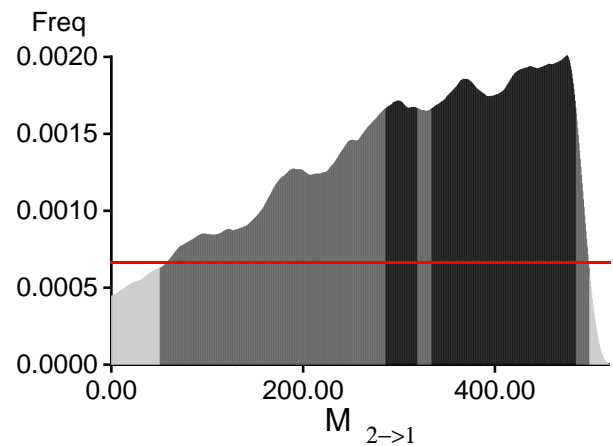
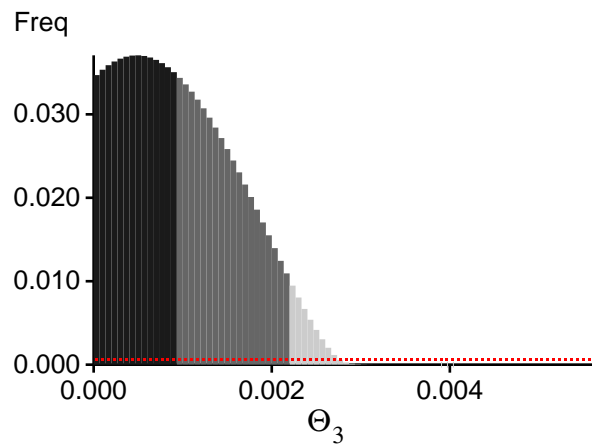
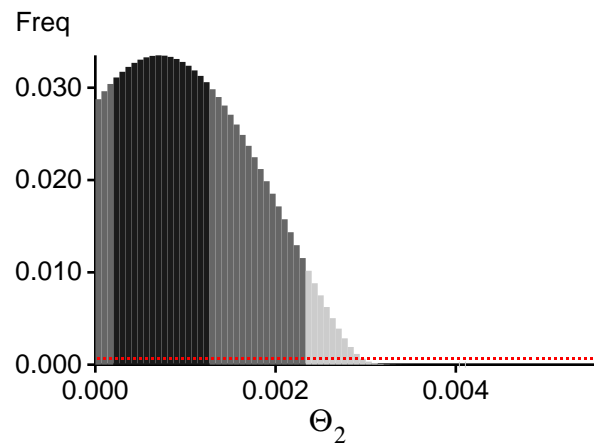
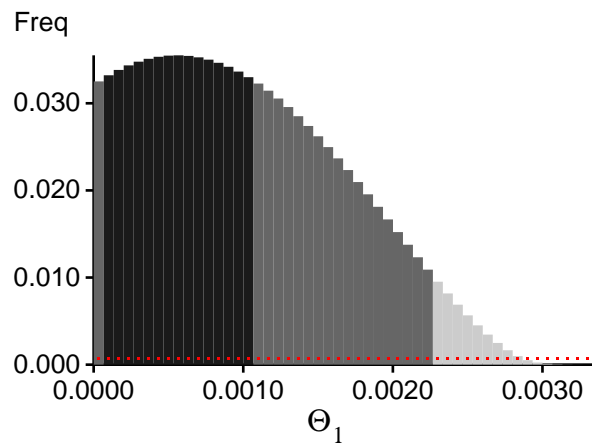
Citation suggestions:

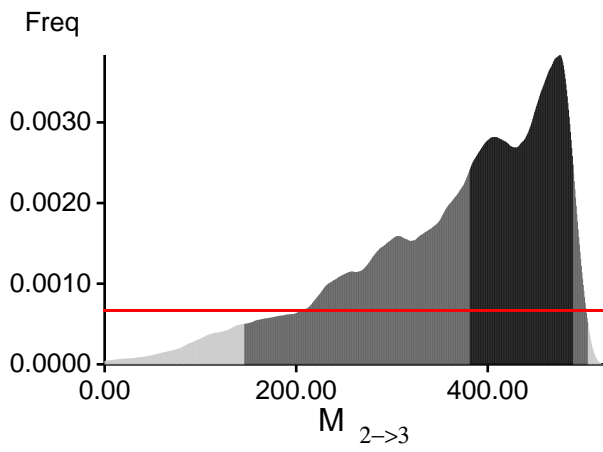
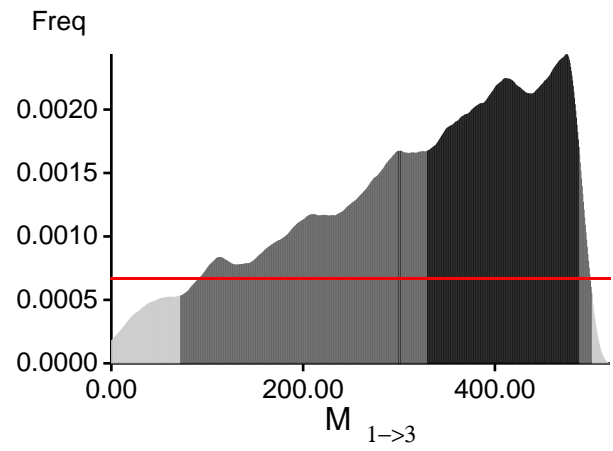
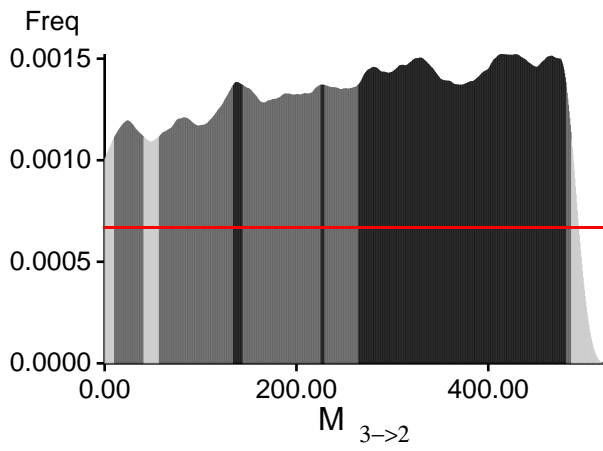
Beerli P., 2006. Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. *Bioinformatics* 22:341-345

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	-1687.81	-1658.70	-1672.85
2	-1622.53	-1549.80	-1527.00
All	-3326.56	-3224.72	-3216.07

(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 = -16.220510]

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
$\Theta_1$	942201/2221429	0.42414
$\Theta_2$	1117549/2221764	0.50300
$\Theta_3$	915119/2221207	0.41199
$M_{2 \rightarrow 1}$	1926656/2223810	0.86638
$M_{3 \rightarrow 1}$	1922633/2223226	0.86479
$M_{1 \rightarrow 2}$	1910856/2221456	0.86018
$M_{3 \rightarrow 2}$	1912873/2222529	0.86067
$M_{1 \rightarrow 3}$	1891384/2221519	0.85139
$M_{2 \rightarrow 3}$	1779411/2221833	0.80088
Genealogies	8588861/20001227	0.42942

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

Parameter	Autocorrelation	Effective Sample Size
$\Theta_1$	0.55441	12114.68
$\Theta_2$	0.50076	14514.04
$\Theta_3$	0.56839	12055.91
$M_{2 \rightarrow 1}$	0.05996	35498.86
$M_{3 \rightarrow 1}$	0.05899	35660.55
$M_{1 \rightarrow 2}$	0.07997	34195.63
$M_{3 \rightarrow 2}$	0.09806	33156.24
$M_{1 \rightarrow 3}$	0.07789	34313.81
$M_{2 \rightarrow 3}$	0.05366	35960.52
Genealogies	0.32157	21250.11

## *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 (Locus 1): Upper prior boundary seems too low!

Param 8 (Locus 2): Upper prior boundary seems too low!

Param 9 (Locus 2): Upper prior boundary seems too low!