

Preliminary migrate analysis of *M. californianus*

MIGRATION RATE AND POPULATION SIZE ESTIMATION

using the coalescent and maximum likelihood or Bayesian inference

Migrate-n version 3.7.2 [April-12-18]

Program started at Wed Jun 2 13:54:44 2021

Program finished at Wed Jun 2 21:23:17 2021



Options

Datatype:

DNA sequence data

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)

3051279299

Start parameters:

Theta values were generated

from guessed values

Theta = 0.01000

M values were generated

from guessed values

M-matrix:

100000.00 [all are the same]

Connection type matrix:

where m = average (average over a group of Thetas or M,

s = symmetric M, S = symmetric 4Nm, 0 = zero, and not estimated,

* = free to vary, Thetas are on diagonal

Population	1	2	3	4	5	6	7	8	9	10	11	12
1 ElfinCo	*	s	0	0	0	0	0	0	0	0	0	0
2 Bamfiel	s	*	s	0	0	0	0	0	0	0	0	0
3 PortRen	0	s	*	s	0	0	0	0	0	0	0	0
4 WalkOnB	0	0	s	*	s	0	0	0	0	0	0	0
5 BodegaH	0	0	0	s	*	s	0	0	0	0	0	0
6 Davenpo	0	0	0	0	s	*	s	0	0	0	0	0
7 VistaDe	0	0	0	0	0	s	*	s	0	0	0	0
8 HazardR	0	0	0	0	0	0	s	*	s	0	0	0
9 Refugio	0	0	0	0	0	0	0	s	*	s	0	0
10 Carpint	0	0	0	0	0	0	0	0	s	*	s	0

11 WhitePo	0	0	0	0	0	0	0	0	0	s	*	s
12 LaJolla	0	0	0	0	0	0	0	0	0	0	s	*

Order of parameters:

1	Θ_1		<displayed>
2	Θ_2		<displayed>
3	Θ_3		<displayed>
4	Θ_4		<displayed>
5	Θ_5		<displayed>
6	Θ_6		<displayed>
7	Θ_7		<displayed>
8	Θ_8		<displayed>
9	Θ_9		<displayed>
10	Θ_{10}		<displayed>
11	Θ_{11}		<displayed>
12	Θ_{12}		<displayed>
13	$M_{2 \rightarrow 1}$	= $M_{2 \rightarrow 1}$ [s]	<displayed>
24	$M_{1 \rightarrow 2}$	= $M_{2 \rightarrow 1}$ [s]	
25	$M_{3 \rightarrow 2}$	= $M_{3 \rightarrow 2}$ [s]	<displayed>
36	$M_{2 \rightarrow 3}$	= $M_{3 \rightarrow 2}$ [s]	
37	$M_{4 \rightarrow 3}$	= $M_{4 \rightarrow 3}$ [s]	<displayed>
48	$M_{3 \rightarrow 4}$	= $M_{4 \rightarrow 3}$ [s]	
49	$M_{5 \rightarrow 4}$	= $M_{5 \rightarrow 4}$ [s]	<displayed>
60	$M_{4 \rightarrow 5}$	= $M_{5 \rightarrow 4}$ [s]	
61	$M_{6 \rightarrow 5}$	= $M_{6 \rightarrow 5}$ [s]	<displayed>
72	$M_{5 \rightarrow 6}$	= $M_{6 \rightarrow 5}$ [s]	
73	$M_{7 \rightarrow 6}$	= $M_{7 \rightarrow 6}$ [s]	<displayed>
84	$M_{6 \rightarrow 7}$	= $M_{7 \rightarrow 6}$ [s]	
85	$M_{8 \rightarrow 7}$	= $M_{8 \rightarrow 7}$ [s]	<displayed>
96	$M_{7 \rightarrow 8}$	= $M_{8 \rightarrow 7}$ [s]	
97	$M_{9 \rightarrow 8}$	= $M_{9 \rightarrow 8}$ [s]	<displayed>
108	$M_{8 \rightarrow 9}$	= $M_{9 \rightarrow 8}$ [s]	
109	$M_{10 \rightarrow 9}$	= $M_{10 \rightarrow 9}$ [s]	<displayed>
120	$M_{9 \rightarrow 10}$	= $M_{10 \rightarrow 9}$ [s]	
121	$M_{11 \rightarrow 10}$	= $M_{11 \rightarrow 10}$ [s]	<displayed>
132	$M_{10 \rightarrow 11}$	= $M_{11 \rightarrow 10}$ [s]	
133	$M_{12 \rightarrow 11}$	= $M_{12 \rightarrow 11}$ [s]	<displayed>
144	$M_{11 \rightarrow 12}$	= $M_{12 \rightarrow 11}$ [s]	

Mutation rate among loci:

Mutation rate is constant

Analysis strategy:

Bayesian inference

Proposal distributions for parameter

Parameter	Proposal
Theta	Metropolis sampling
M	Slice sampling

Prior distribution for parameter

Parameter	Prior	Minimum	Mean*	Maximum	Delta	Bins
Theta	Exp window	0.000010	0.010000	10.000000	1.000000	500
M	Exp window	0.000100	100000.000000	1000000.000000	100000.000000	500

Markov chain settings:

Long chain

Number of chains	1
Recorded steps [a]	1000
Increment (record every x step [b])	100
Number of concurrent chains (replicates) [c]	3
Visited (sampled) parameter values [a*b*c]	300000
Number of discard trees per chain (burn-in)	1000

Multiple Markov chains:

Static heating scheme

4 chains with temperatures
 100000.00 3.00 1.50 1.00
 Swapping interval is 1

Print options:

Data file: ../../mcalifornianus_210528.mig
 Output file: outfile.txt
 Posterior distribution raw histogram file: bayesfile
 Print data: No
 Print genealogies [only some for some data type]: None

Data summary

Datatype: Sequence data
 Number of loci: 1

Population	Locus	Gene copies
1 ElfinCo	1	19
2 Bamfiel	1	23
3 PortRen	1	15
4 WalkOnB	1	16
5 BodegaH	1	7
6 Davenpo	1	17
7 VistaDe	1	19
8 HazardR	1	23
9 Refugio	1	16
10 Carpint	1	19
11 WhitePo	1	11
12 LaJolla	1	8
Total of all populations	1	193

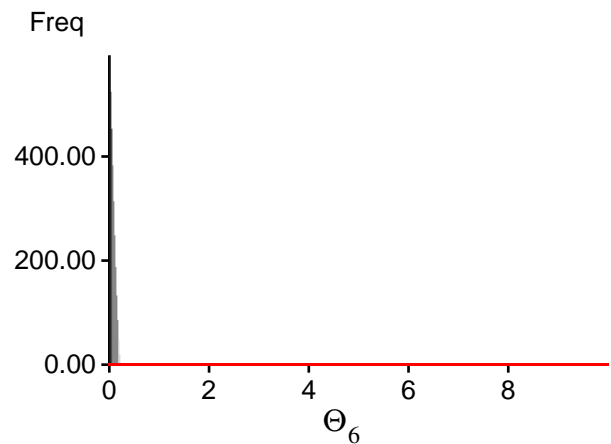
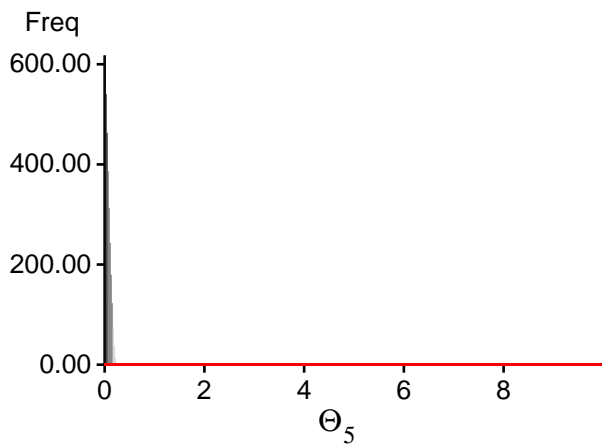
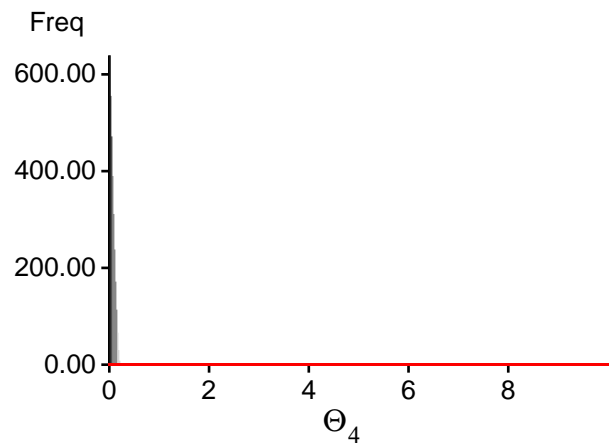
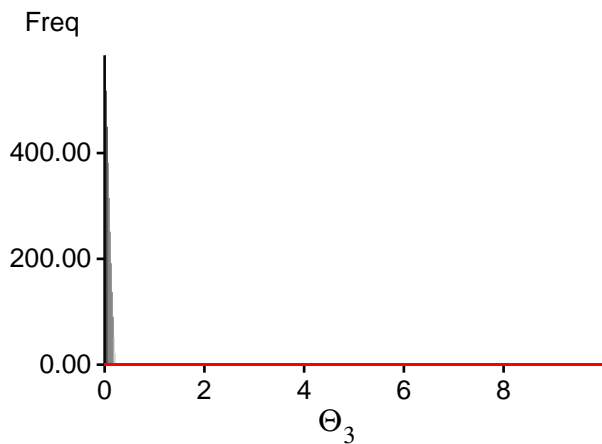
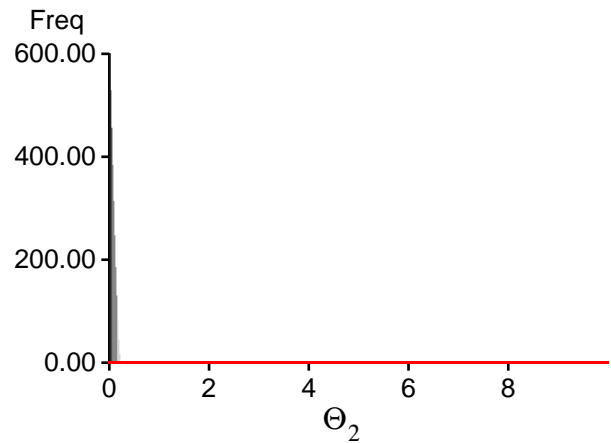
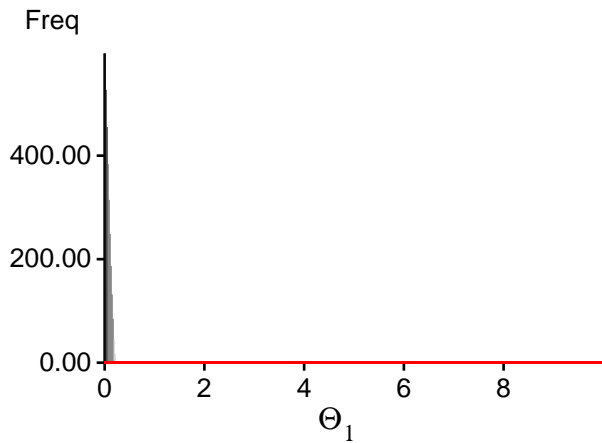
Bayesian Analysis: Posterior distribution table

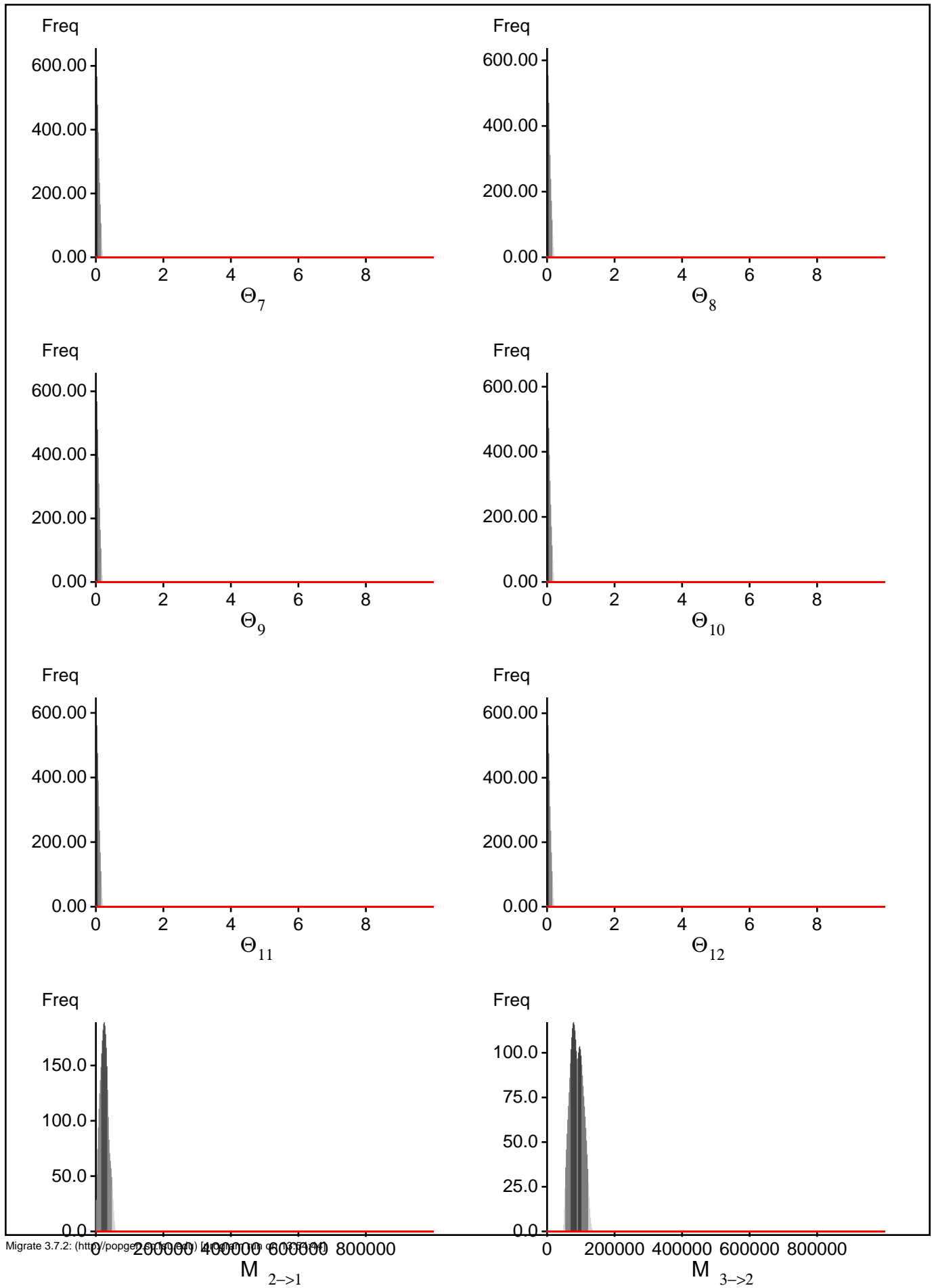
Locus	Parameter	2.5%	25.0%	Mode	75.0%	97.5%	Median	Mean
1	Θ_1	0.00001	0.00001	0.01001	0.06001	0.18001	0.07001	0.02278
1	Θ_2	0.00001	0.00001	0.01001	0.06001	0.16001	0.07001	0.02220
1	Θ_3	0.00001	0.00001	0.01001	0.06001	0.18001	0.07001	0.02510
1	Θ_4	0.00001	0.00001	0.01001	0.06001	0.16001	0.07001	0.01468
1	Θ_5	0.00001	0.00001	0.01001	0.06001	0.16001	0.07001	0.01752
1	Θ_6	0.00001	0.00001	0.01001	0.06001	0.18001	0.07001	0.02400
1	Θ_7	0.00001	0.00001	0.01001	0.06001	0.16001	0.07001	0.01006
1	Θ_8	0.00001	0.00001	0.01001	0.06001	0.16001	0.07001	0.01761
1	Θ_9	0.00001	0.00001	0.01001	0.06001	0.16001	0.07001	0.00980
1	Θ_{10}	0.00001	0.00001	0.01001	0.06001	0.16001	0.07001	0.01614
1	Θ_{11}	0.00001	0.00001	0.01001	0.06001	0.16001	0.07001	0.01399
1	Θ_{12}	0.00001	0.00001	0.01001	0.06001	0.16001	0.07001	0.01220
1	$M_{2 \rightarrow 1}$	0.0	14000.0	25000.0	34000.0	48000.0	27000.0	24571.2
1	$M_{1 \rightarrow 2}$	0.0	14000.0	25000.0	34000.0	48000.0	27000.0	24571.2
1	$M_{3 \rightarrow 2}$	52000.0	68000.0	79000.0	88000.0	122000.0	89000.0	87692.0
1	$M_{2 \rightarrow 3}$	52000.0	68000.0	79000.0	88000.0	122000.0	89000.0	87692.0
1	$M_{4 \rightarrow 3}$	64000.0	72000.0	87000.0	102000.0	110000.0	81000.0	71783.8
1	$M_{3 \rightarrow 4}$	64000.0	72000.0	87000.0	102000.0	110000.0	81000.0	71783.8
1	$M_{5 \rightarrow 4}$	26000.0	34000.0	49000.0	60000.0	68000.0	57000.0	77086.2
1	$M_{4 \rightarrow 5}$	26000.0	34000.0	49000.0	60000.0	68000.0	57000.0	77086.2
1	$M_{6 \rightarrow 5}$	56000.0	74000.0	85000.0	98000.0	118000.0	89000.0	87834.1
1	$M_{5 \rightarrow 6}$	56000.0	74000.0	85000.0	98000.0	118000.0	89000.0	87834.1
1	$M_{7 \rightarrow 6}$	44000.0	68000.0	89000.0	98000.0	108000.0	65000.0	60457.2
1	$M_{6 \rightarrow 7}$	44000.0	68000.0	89000.0	98000.0	108000.0	65000.0	60457.2
1	$M_{8 \rightarrow 7}$	100000.0	110000.0	125000.0	140000.0	148000.0	119000.0	97410.2
1	$M_{7 \rightarrow 8}$	100000.0	110000.0	125000.0	140000.0	148000.0	119000.0	97410.2
1	$M_{9 \rightarrow 8}$	10000.0	22000.0	43000.0	60000.0	72000.0	57000.0	63176.8
1	$M_{8 \rightarrow 9}$	10000.0	22000.0	43000.0	60000.0	72000.0	57000.0	63176.8
1	$M_{10 \rightarrow 9}$	50000.0	58000.0	71000.0	84000.0	92000.0	67000.0	51123.0
1	$M_{9 \rightarrow 10}$	50000.0	58000.0	71000.0	84000.0	92000.0	67000.0	51123.0
1	$M_{11 \rightarrow 10}$	36000.0	46000.0	59000.0	74000.0	110000.0	69000.0	71449.1
1	$M_{10 \rightarrow 11}$	36000.0	46000.0	59000.0	74000.0	110000.0	69000.0	71449.1
1	$M_{12 \rightarrow 11}$	12000.0	22000.0	39000.0	54000.0	102000.0	57000.0	57491.9
1	$M_{11 \rightarrow 12}$	12000.0	22000.0	39000.0	54000.0	102000.0	57000.0	57491.9

Citation suggestions:

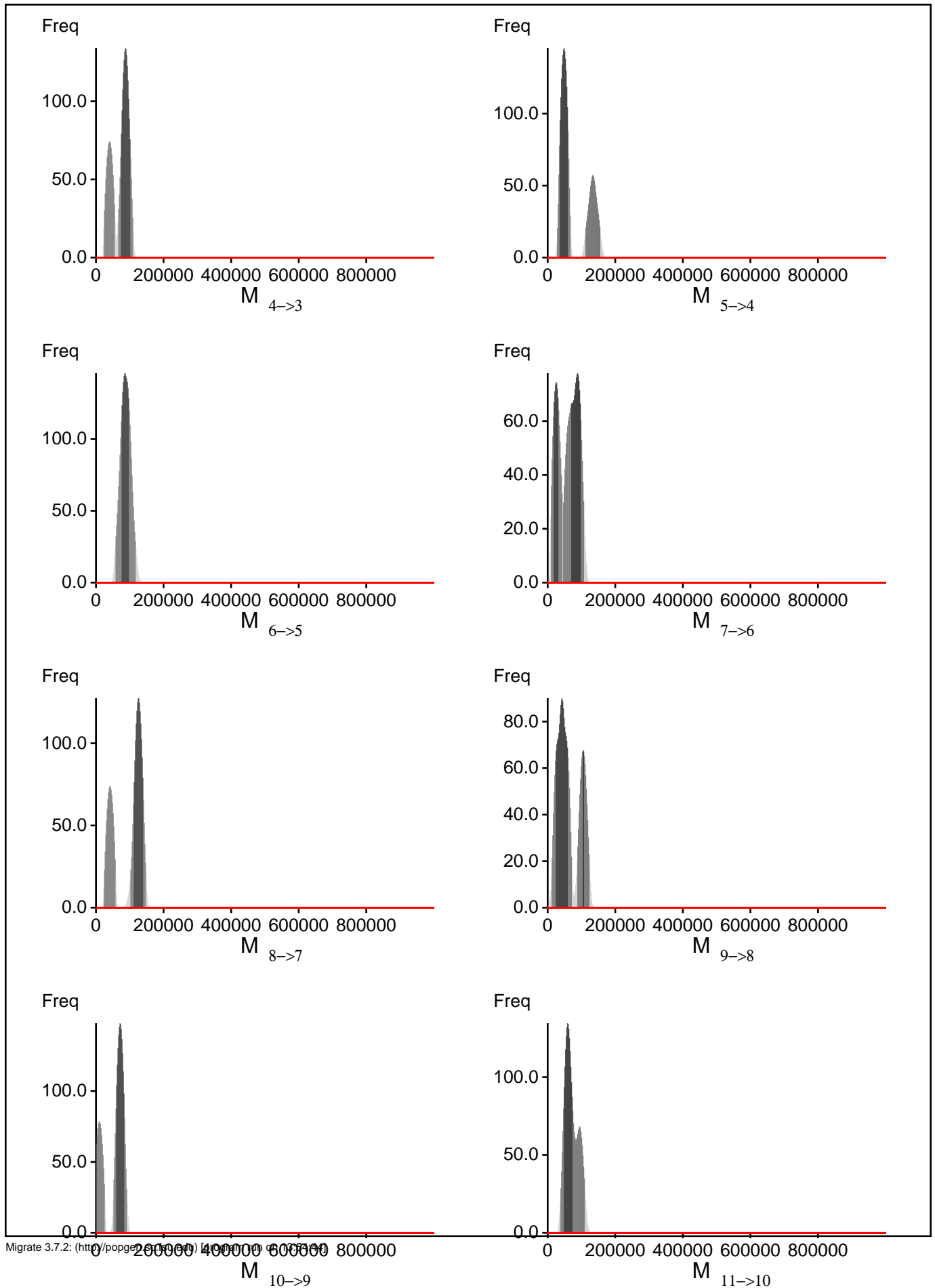
- Beerli P., 2006. Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. *Bioinformatics* 22:341-345
- Beerli P., 2007. Estimation of the population scaled mutation rate from microsatellite data, *Genetics*, 177:1967-1968.
- Beerli P., 2009. How to use MIGRATE or why 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.

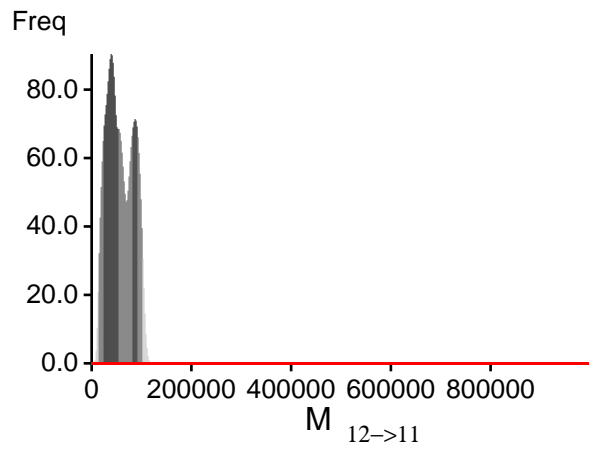
Bayesian Analysis: Posterior distribution over all loci





Preliminary migrate analysis of *M. californianus* CO1 haplotypes for Evolution 2 -- 9





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]

Method	$\ln(\text{Prob}(D \mid \text{Model}))$	Notes
Thermodynamic integration	-2270.138664	(1a)
	-2174.790707	(1b)
Harmonic mean	-1904.847514	(2)

(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

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	1045/4420	0.23643
Θ_2	1156/4411	0.26207
Θ_3	1044/4405	0.23700
Θ_4	970/4419	0.21951
Θ_5	858/4483	0.19139
Θ_6	1243/4363	0.28490
Θ_7	926/4331	0.21381
Θ_8	1699/4401	0.38605
Θ_9	1183/4383	0.26991
Θ_{10}	1588/4357	0.36447
Θ_{11}	1436/4366	0.32891
Θ_{12}	1081/4378	0.24692
$M_{2 \rightarrow 1}$	4287/4287	1.00000
$M_{1 \rightarrow 2}$	4287/4287	1.00000
$M_{3 \rightarrow 2}$	4476/4476	1.00000
$M_{2 \rightarrow 3}$	4476/4476	1.00000
$M_{4 \rightarrow 3}$	4372/4372	1.00000
$M_{3 \rightarrow 4}$	4372/4372	1.00000
$M_{5 \rightarrow 4}$	4526/4526	1.00000
$M_{4 \rightarrow 5}$	4526/4526	1.00000
$M_{6 \rightarrow 5}$	4485/4485	1.00000
$M_{5 \rightarrow 6}$	4485/4485	1.00000
$M_{7 \rightarrow 6}$	4411/4411	1.00000
$M_{6 \rightarrow 7}$	4411/4411	1.00000
$M_{8 \rightarrow 7}$	4398/4398	1.00000
$M_{7 \rightarrow 8}$	4398/4398	1.00000
$M_{9 \rightarrow 8}$	4456/4456	1.00000
$M_{8 \rightarrow 9}$	4456/4456	1.00000
$M_{10 \rightarrow 9}$	4447/4447	1.00000
$M_{9 \rightarrow 10}$	4447/4447	1.00000
$M_{11 \rightarrow 10}$	4368/4368	1.00000
$M_{10 \rightarrow 11}$	4368/4368	1.00000
$M_{12 \rightarrow 11}$	4494/4494	1.00000
$M_{11 \rightarrow 12}$	4494/4494	1.00000
Genealogies	37114/150167	0.24715

MCMC-Autocorrelation and Effective MCMC Sample Size

Parameter	Autocorrelation	Effective Sample Size
Θ_1	0.80947	315.59
Θ_2	0.76951	407.52
Θ_3	0.84164	257.93
Θ_4	0.84595	250.99
Θ_5	0.88473	184.05
Θ_6	0.79845	339.67
Θ_7	0.84017	272.13
Θ_8	0.77643	381.64
Θ_9	0.79346	384.73
Θ_{10}	0.76059	407.57
Θ_{11}	0.76448	414.08
Θ_{12}	0.81600	320.51
$M_{2 \rightarrow 1}$	0.81787	300.96
$M_{1 \rightarrow 2}$	0.81787	300.96
$M_{3 \rightarrow 2}$	0.79746	344.53
$M_{2 \rightarrow 3}$	0.79746	344.53
$M_{4 \rightarrow 3}$	0.78758	362.00
$M_{3 \rightarrow 4}$	0.78758	362.00
$M_{5 \rightarrow 4}$	0.85452	237.31
$M_{4 \rightarrow 5}$	0.85452	237.31
$M_{6 \rightarrow 5}$	0.90240	155.03
$M_{5 \rightarrow 6}$	0.90240	155.03
$M_{7 \rightarrow 6}$	0.86441	219.68
$M_{6 \rightarrow 7}$	0.86441	219.68
$M_{8 \rightarrow 7}$	0.80394	327.38
$M_{7 \rightarrow 8}$	0.80394	327.38
$M_{9 \rightarrow 8}$	0.82732	286.85
$M_{8 \rightarrow 9}$	0.82732	286.85
$M_{10 \rightarrow 9}$	0.85655	231.91
$M_{9 \rightarrow 10}$	0.85655	231.91
$M_{11 \rightarrow 10}$	0.83874	264.46
$M_{10 \rightarrow 11}$	0.83874	264.46
$M_{12 \rightarrow 11}$	0.85857	228.75
$M_{11 \rightarrow 12}$	0.85857	228.75
$\text{Ln}[\text{Prob(D G)}]$	0.97882	32.17

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