

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 Tue Jun 1 13:45:41 2021

Program finished at Tue Jun 1 18:52:02 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)

2224364243

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

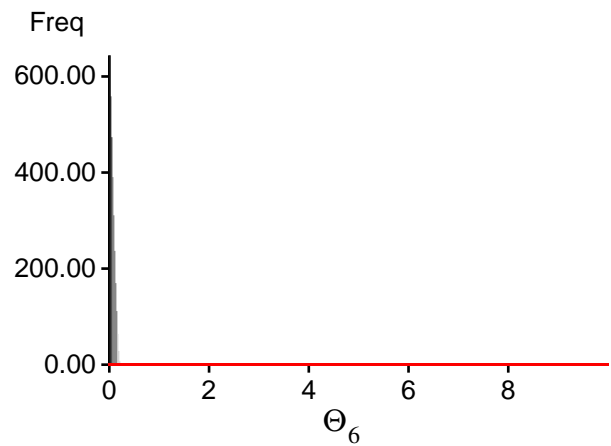
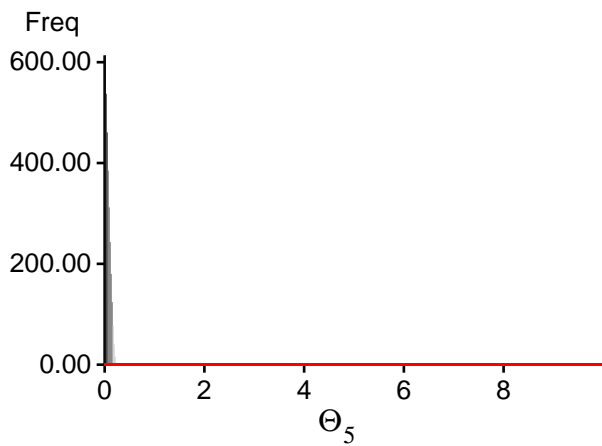
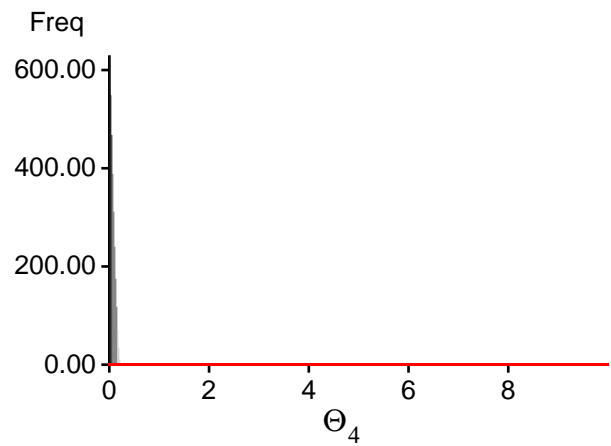
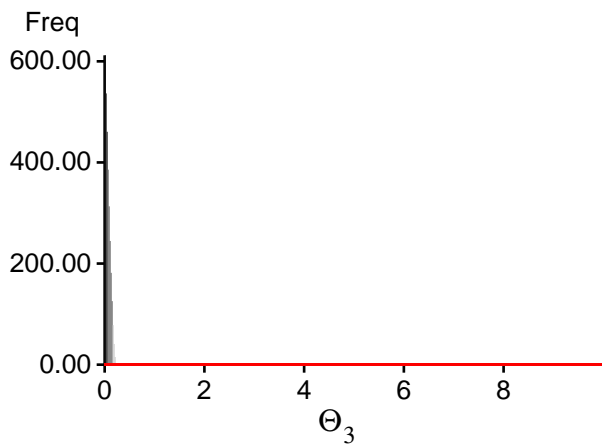
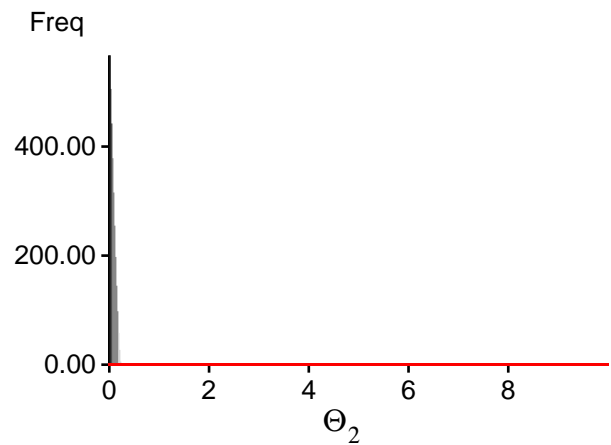
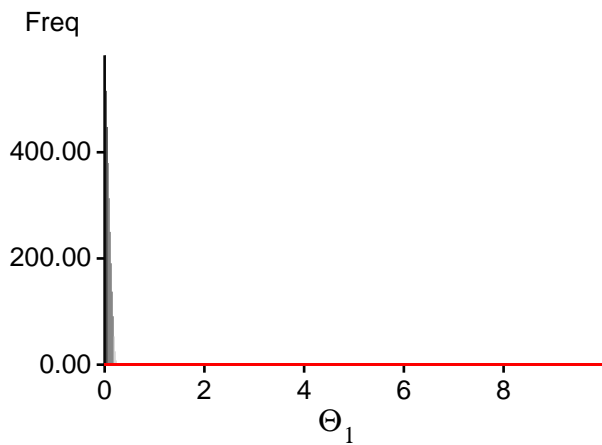
Bayesian Analysis: Posterior distribution table

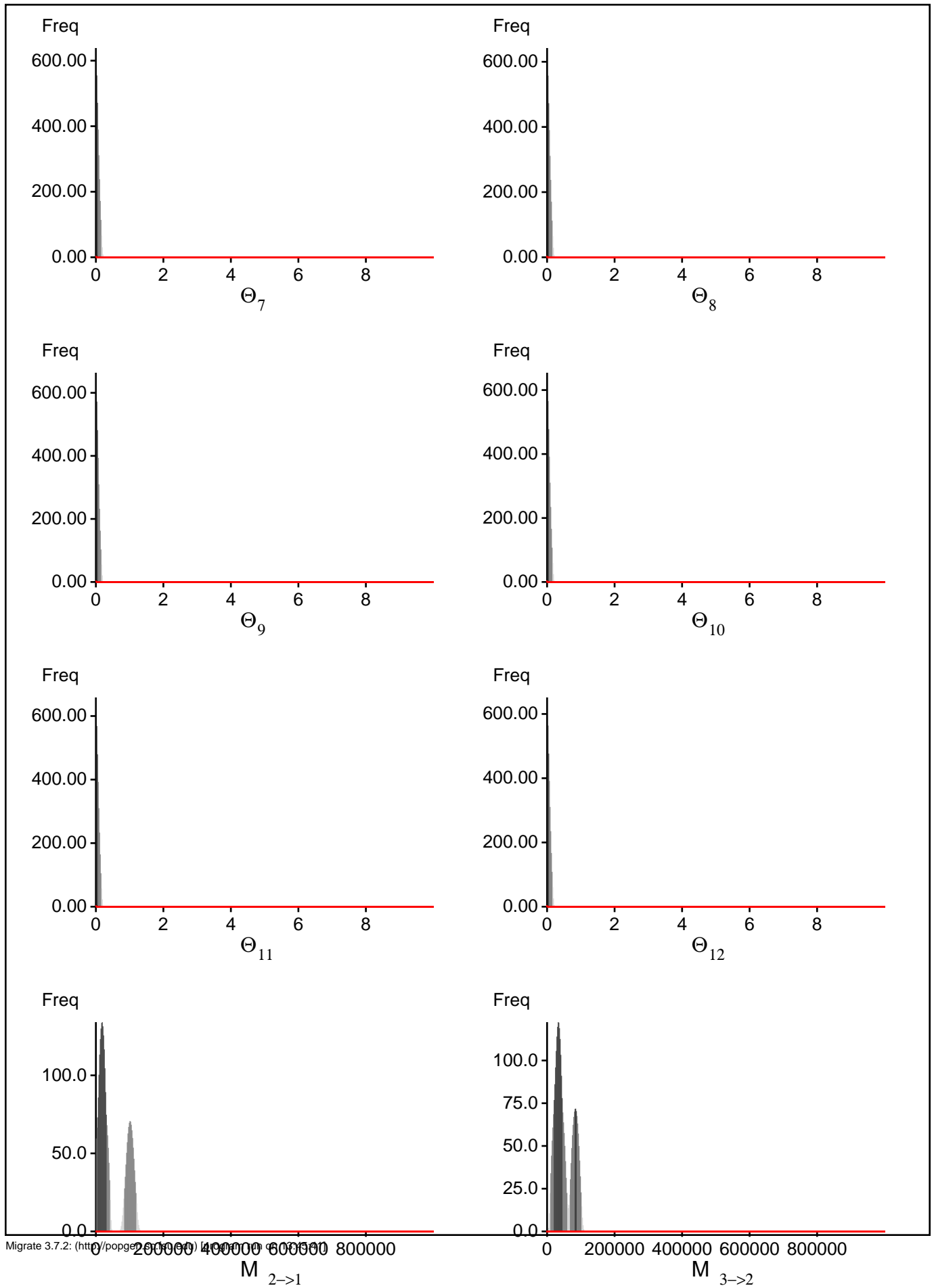
Locus	Parameter	2.5%	25.0%	Mode	75.0%	97.5%	Median	Mean
1	Θ_1	0.00001	0.00001	0.01001	0.06001	0.18001	0.07001	0.02338
1	Θ_2	0.00001	0.00001	0.01001	0.06001	0.18001	0.07001	0.02585
1	Θ_3	0.00001	0.00001	0.01001	0.06001	0.16001	0.07001	0.02190
1	Θ_4	0.00001	0.00001	0.01001	0.06001	0.16001	0.07001	0.01662
1	Θ_5	0.00001	0.00001	0.01001	0.06001	0.16001	0.07001	0.02267
1	Θ_6	0.00001	0.00001	0.01001	0.06001	0.16001	0.07001	0.01523
1	Θ_7	0.00001	0.00001	0.01001	0.06001	0.16001	0.07001	0.01695
1	Θ_8	0.00001	0.00001	0.01001	0.06001	0.16001	0.07001	0.01664
1	Θ_9	0.00001	0.00001	0.01001	0.06001	0.16001	0.07001	0.00606
1	Θ_{10}	0.00001	0.00001	0.01001	0.06001	0.16001	0.07001	0.01191
1	Θ_{11}	0.00001	0.00001	0.01001	0.06001	0.16001	0.07001	0.01002
1	Θ_{12}	0.00001	0.00001	0.01001	0.06001	0.16001	0.07001	0.01305
1	$M_{2 \rightarrow 1}$	0.0	2000.0	19000.0	32000.0	42000.0	31000.0	45987.0
1	$M_{1 \rightarrow 2}$	0.0	2000.0	19000.0	32000.0	42000.0	31000.0	45987.0
1	$M_{3 \rightarrow 2}$	8000.0	18000.0	33000.0	46000.0	60000.0	45000.0	50584.8
1	$M_{2 \rightarrow 3}$	8000.0	18000.0	33000.0	46000.0	60000.0	45000.0	50584.8
1	$M_{4 \rightarrow 3}$	20000.0	28000.0	41000.0	54000.0	62000.0	51000.0	59609.0
1	$M_{3 \rightarrow 4}$	20000.0	28000.0	41000.0	54000.0	62000.0	51000.0	59609.0
1	$M_{5 \rightarrow 4}$	68000.0	82000.0	99000.0	114000.0	130000.0	89000.0	75313.9
1	$M_{4 \rightarrow 5}$	68000.0	82000.0	99000.0	114000.0	130000.0	89000.0	75313.9
1	$M_{6 \rightarrow 5}$	64000.0	74000.0	91000.0	102000.0	114000.0	83000.0	65887.3
1	$M_{5 \rightarrow 6}$	64000.0	74000.0	91000.0	102000.0	114000.0	83000.0	65887.3
1	$M_{7 \rightarrow 6}$	0.0	8000.0	19000.0	28000.0	40000.0	23000.0	18929.2
1	$M_{6 \rightarrow 7}$	0.0	8000.0	19000.0	28000.0	40000.0	23000.0	18929.2
1	$M_{8 \rightarrow 7}$	28000.0	42000.0	55000.0	64000.0	84000.0	57000.0	54807.8
1	$M_{7 \rightarrow 8}$	28000.0	42000.0	55000.0	64000.0	84000.0	57000.0	54807.8
1	$M_{9 \rightarrow 8}$	10000.0	56000.0	67000.0	76000.0	84000.0	69000.0	74063.9
1	$M_{8 \rightarrow 9}$	10000.0	56000.0	67000.0	76000.0	84000.0	69000.0	74063.9
1	$M_{10 \rightarrow 9}$	26000.0	42000.0	59000.0	74000.0	114000.0	69000.0	69229.4
1	$M_{9 \rightarrow 10}$	26000.0	42000.0	59000.0	74000.0	114000.0	69000.0	69229.4
1	$M_{11 \rightarrow 10}$	0.0	0.0	5000.0	14000.0	22000.0	99000.0	74384.8
1	$M_{10 \rightarrow 11}$	0.0	0.0	5000.0	14000.0	22000.0	99000.0	74384.8
1	$M_{12 \rightarrow 11}$	60000.0	74000.0	85000.0	92000.0	106000.0	87000.0	84160.8
1	$M_{11 \rightarrow 12}$	60000.0	74000.0	85000.0	92000.0	106000.0	87000.0	84160.8

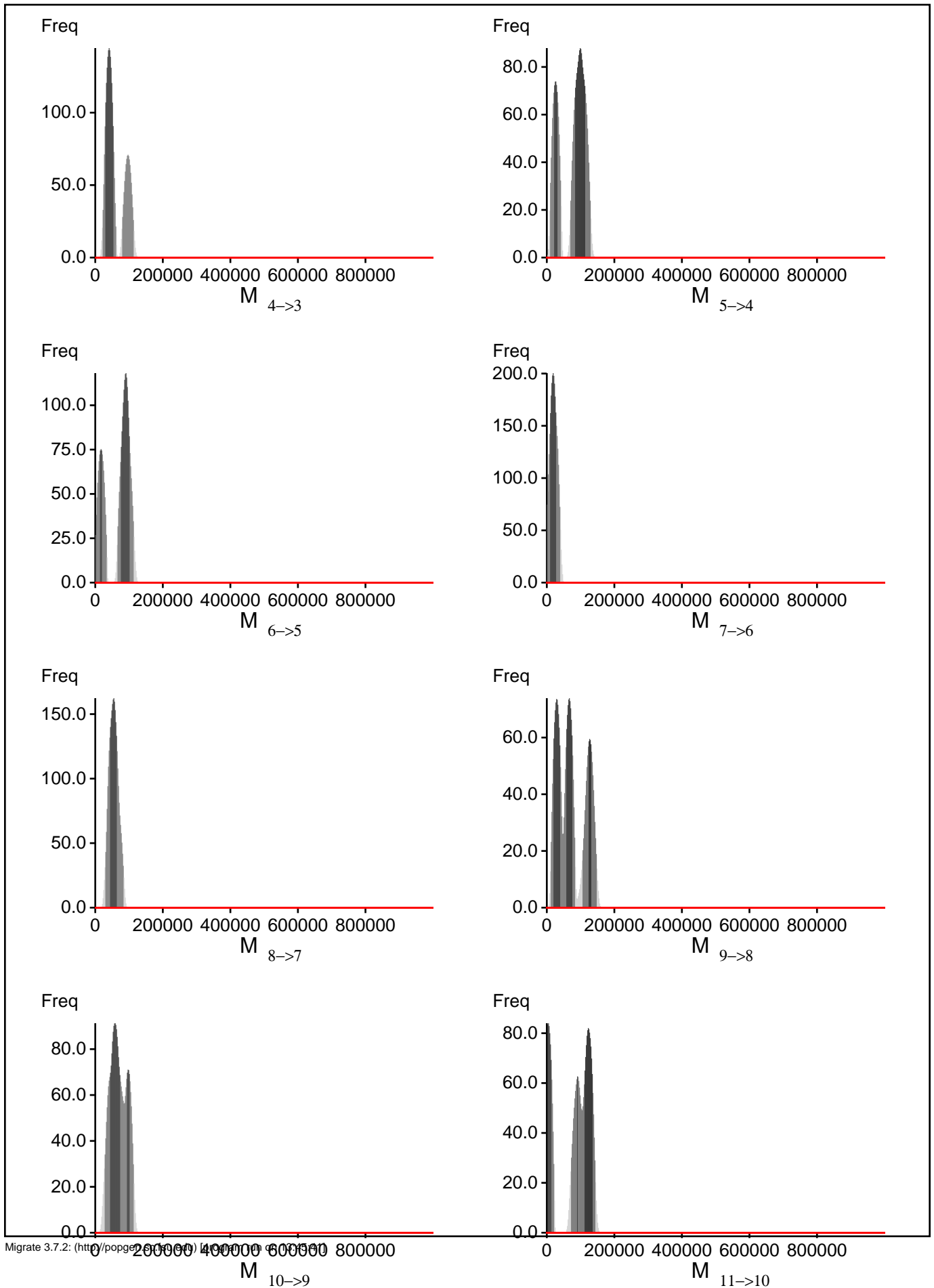
Citation suggestions:

- 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.

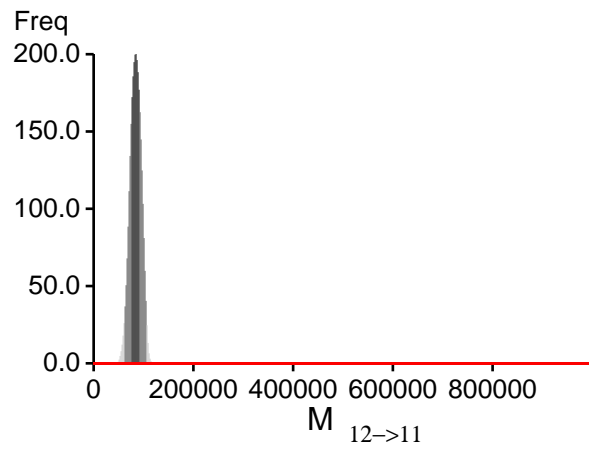
Bayesian Analysis: Posterior distribution over all loci







Preliminary migrate analysis of *M. californianus* CO1 haplotypes for Evolution 2 -- 10



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	-2256.241802	(1a)
	-2163.411531	(1b)
Harmonic mean	-1878.055351	(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	421/4451	0.09459
Θ_2	984/4348	0.22631
Θ_3	1196/4379	0.27312
Θ_4	767/4415	0.17373
Θ_5	1197/4427	0.27039
Θ_6	1234/4526	0.27265
Θ_7	1435/4405	0.32577
Θ_8	1591/4466	0.35625
Θ_9	463/4370	0.10595
Θ_{10}	1206/4359	0.27667
Θ_{11}	1309/4359	0.30030
Θ_{12}	1486/4442	0.33453
$M_{2 \rightarrow 1}$	4570/4570	1.00000
$M_{1 \rightarrow 2}$	4570/4570	1.00000
$M_{3 \rightarrow 2}$	4505/4505	1.00000
$M_{2 \rightarrow 3}$	4505/4505	1.00000
$M_{4 \rightarrow 3}$	4294/4294	1.00000
$M_{3 \rightarrow 4}$	4294/4294	1.00000
$M_{5 \rightarrow 4}$	4330/4330	1.00000
$M_{4 \rightarrow 5}$	4330/4330	1.00000
$M_{6 \rightarrow 5}$	4354/4354	1.00000
$M_{5 \rightarrow 6}$	4354/4354	1.00000
$M_{7 \rightarrow 6}$	4415/4415	1.00000
$M_{6 \rightarrow 7}$	4415/4415	1.00000
$M_{8 \rightarrow 7}$	4426/4426	1.00000
$M_{7 \rightarrow 8}$	4426/4426	1.00000
$M_{9 \rightarrow 8}$	4431/4431	1.00000
$M_{8 \rightarrow 9}$	4431/4431	1.00000
$M_{10 \rightarrow 9}$	4532/4532	1.00000
$M_{9 \rightarrow 10}$	4532/4532	1.00000
$M_{11 \rightarrow 10}$	4526/4526	1.00000
$M_{10 \rightarrow 11}$	4526/4526	1.00000
$M_{12 \rightarrow 11}$	4368/4368	1.00000
$M_{11 \rightarrow 12}$	4368/4368	1.00000
Genealogies	33522/149723	0.22389

MCMC-Autocorrelation and Effective MCMC Sample Size

Parameter	Autocorrelation	Effective Sample Size
Θ_1	0.94272	89.66
Θ_2	0.83773	269.78
Θ_3	0.80733	320.34
Θ_4	0.87618	203.95
Θ_5	0.86034	229.06
Θ_6	0.79091	351.83
Θ_7	0.77311	392.51
Θ_8	0.80880	319.15
Θ_9	0.91744	136.58
Θ_{10}	0.80385	336.20
Θ_{11}	0.77485	388.08
Θ_{12}	0.81159	315.54
$M_{2 \rightarrow 1}$	0.81039	314.87
$M_{1 \rightarrow 2}$	0.81039	314.87
$M_{3 \rightarrow 2}$	0.80807	320.82
$M_{2 \rightarrow 3}$	0.80807	320.82
$M_{4 \rightarrow 3}$	0.82166	297.15
$M_{3 \rightarrow 4}$	0.82166	297.15
$M_{5 \rightarrow 4}$	0.84048	260.84
$M_{4 \rightarrow 5}$	0.84048	260.84
$M_{6 \rightarrow 5}$	0.77833	376.83
$M_{5 \rightarrow 6}$	0.77833	376.83
$M_{7 \rightarrow 6}$	0.84533	251.43
$M_{6 \rightarrow 7}$	0.84533	251.43
$M_{8 \rightarrow 7}$	0.84217	257.35
$M_{7 \rightarrow 8}$	0.84217	257.35
$M_{9 \rightarrow 8}$	0.83155	277.72
$M_{8 \rightarrow 9}$	0.83155	277.72
$M_{10 \rightarrow 9}$	0.84466	256.10
$M_{9 \rightarrow 10}$	0.84466	256.10
$M_{11 \rightarrow 10}$	0.85773	233.30
$M_{10 \rightarrow 11}$	0.85773	233.30
$M_{12 \rightarrow 11}$	0.83267	274.23
$M_{11 \rightarrow 12}$	0.83267	274.23
$\text{Ln}[\text{Prob}(D G)]$	0.97888	32.03

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