

Example with tip dates

MIGRATION RATE AND POPULATION SIZE ESTIMATION
using the coalescent and maximum likelihood or Bayesian inference

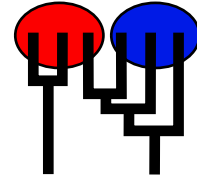
Migrate-n version 3.0.1

Compiled for a PARALLEL COMPUTER ARCHITECTURE

One master and 2 compute nodes are available.

Program started at Tue Oct 28 11:48:15 2008

Program finished at Tue Oct 28 11:48:39 2008



Options

Datatype: DNA sequence data
Random number seed: (with internal timer) 884378039

Start parameters:

Theta values were generated from the FST-calculation

M values were generated from the FST-calculation

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
1 Population1	*	*	*
2 Population2	*	*	*
3 Population3	*	*	*

Order of parameters:

1	Θ_1	<displayed>
2	Θ_2	<displayed>
3	Θ_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

Analysis strategy:

Bayesian inference

Proposal distributions for parameter

Parameter	Proposal
Theta	Slice sampling
M	Slice sampling

Prior distribution for parameter

Parameter	Prior	Minimum	Mean*	Maximum	Delta	Bins
Theta	Exponential	0.000000	0.010000	0.100000	-	200
M	Exponential	0.000000	10.000000	1000.000000	-	200

Markov chain settings:

Long chain

Number of chains	1
Recorded steps [a]	1600
Increment (record every x step [b])	100
Number of concurrent chains (replicates) [c]	2
Visited (sampled) parameter values [a*b*c]	320000
Number of discard trees per chain (burn-in)	130

Print options:

Data file:	infile.dated
Output file:	outfile.dated
Posterior distribution raw histogram file:	bayesfile.dated
Print data:	No
Print genealogies [only some for some data type]:	Yes, only the best
Histogram of the frequency of migration events	mighistfile.dated

Data summary

Datatype:

Sequence data

Number of loci:

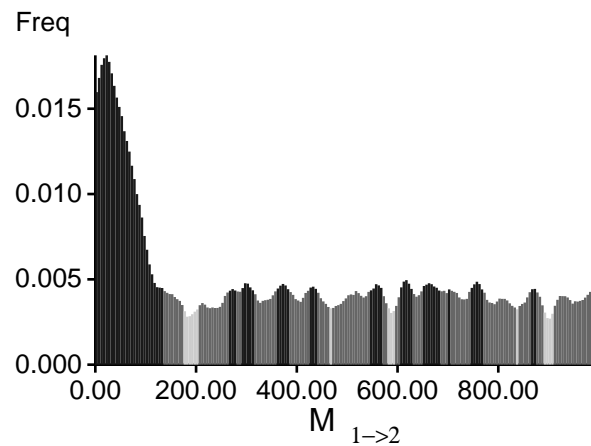
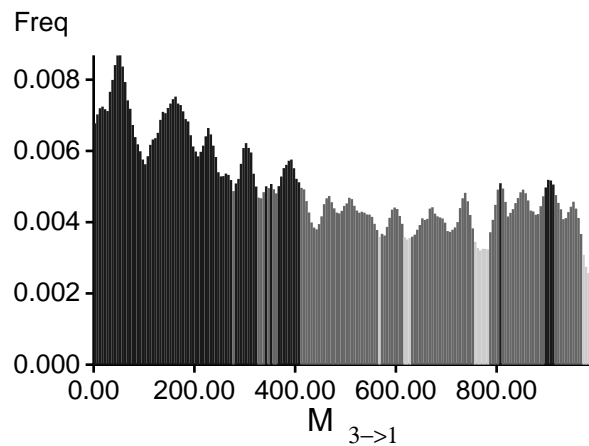
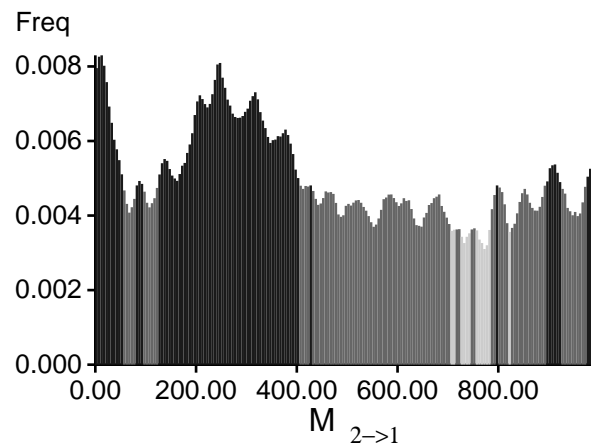
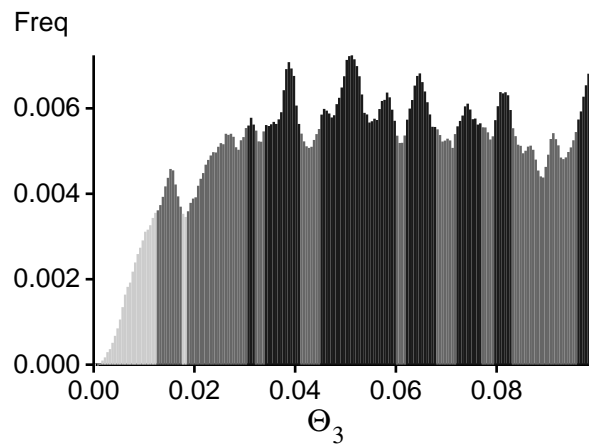
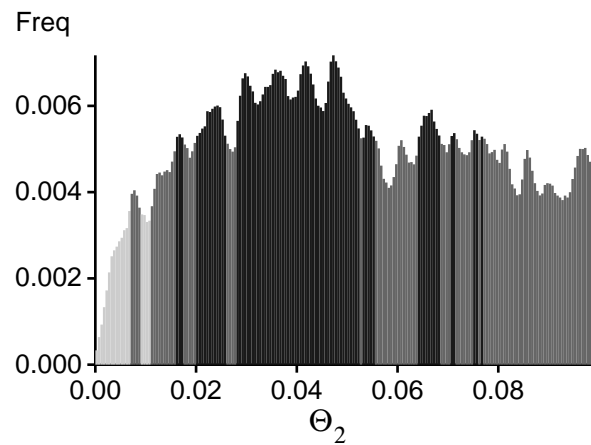
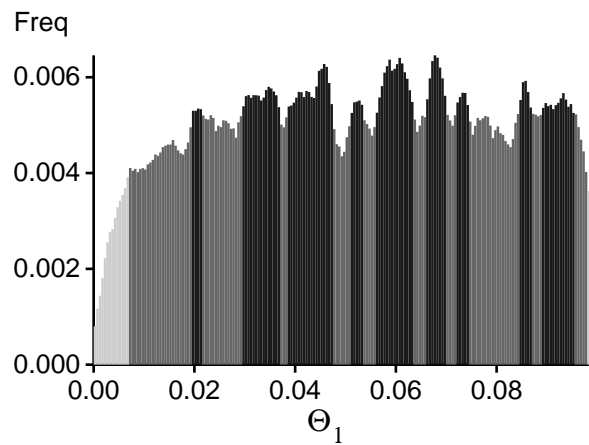
1

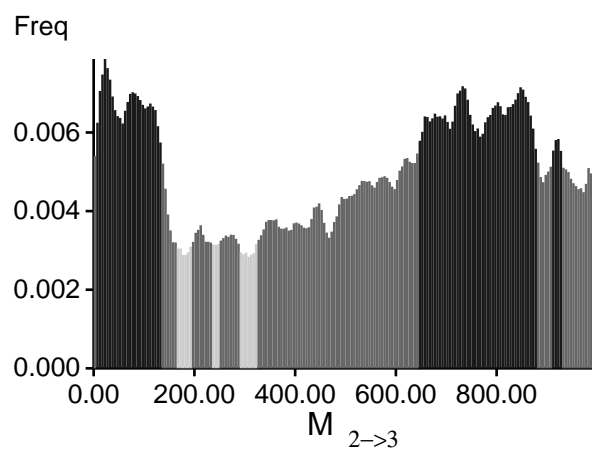
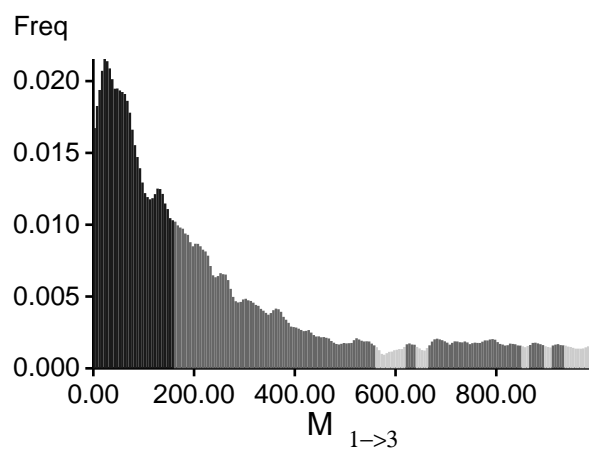
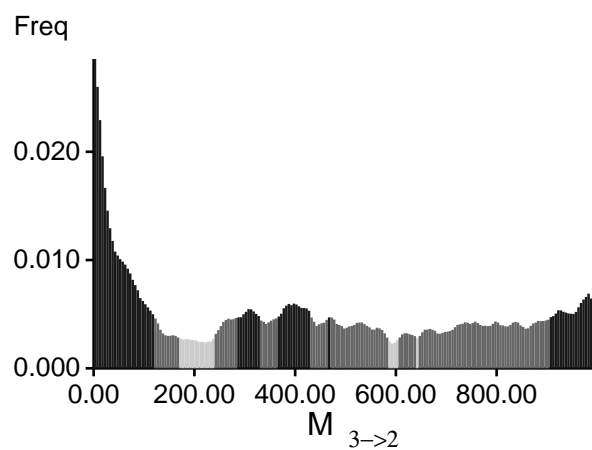
Population	Locus	Gene copies
1 Population1	1	3
2 Population2	1	3
3 Population3	1	3
Total of all populations	1	9

Bayesian Analysis: Posterior distribution table

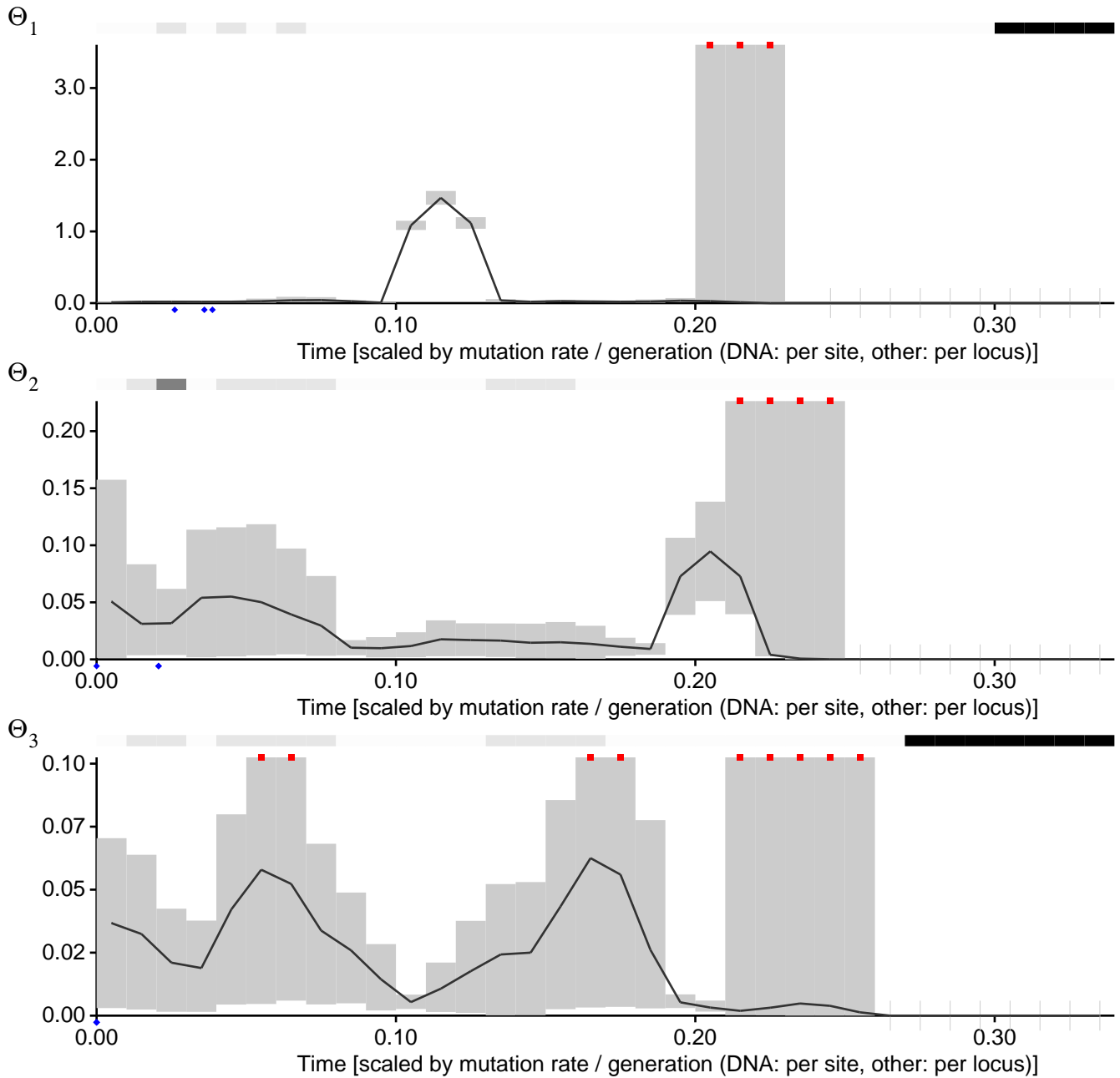
Locus	Parameter	2.5%	25.0%	Mode	75.0%	97.5%	Median	Mean
1	Θ_1	0.00650	0.06550	0.06775	0.07000	0.09800	0.05325	0.05255
1	Θ_2	0.01050	0.02750	0.04725	0.05250	0.09950	0.04925	0.05073
1	Θ_3	0.01800	0.04450	0.05125	0.06000	0.10000	0.05575	0.05523
1	$M_{2 \rightarrow 1}$	0.0	0.0	12.5	55.0	705.0	412.5	457.3
1	$M_{3 \rightarrow 1}$	0.0	0.0	52.5	275.0	565.0	402.5	443.0
1	$M_{1 \rightarrow 2}$	0.0	0.0	22.5	135.0	175.0	377.5	402.2
1	$M_{3 \rightarrow 2}$	0.0	0.0	2.5	120.0	170.0	402.5	429.4
1	$M_{1 \rightarrow 3}$	0.0	0.0	22.5	160.0	560.0	162.5	250.0
1	$M_{2 \rightarrow 3}$	0.0	0.0	22.5	135.0	165.0	572.5	518.7

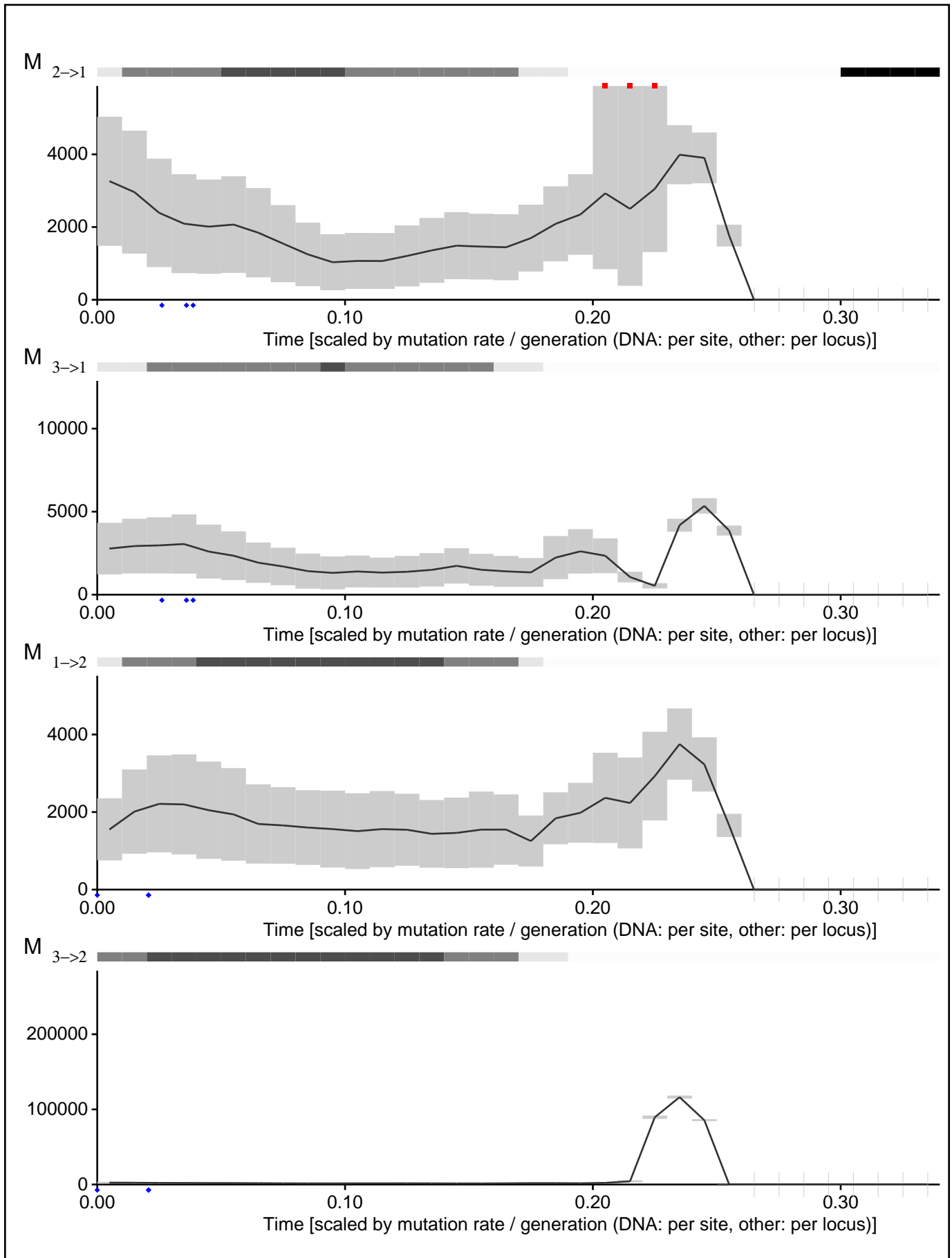
Bayesian Analysis: Posterior distribution over all loci

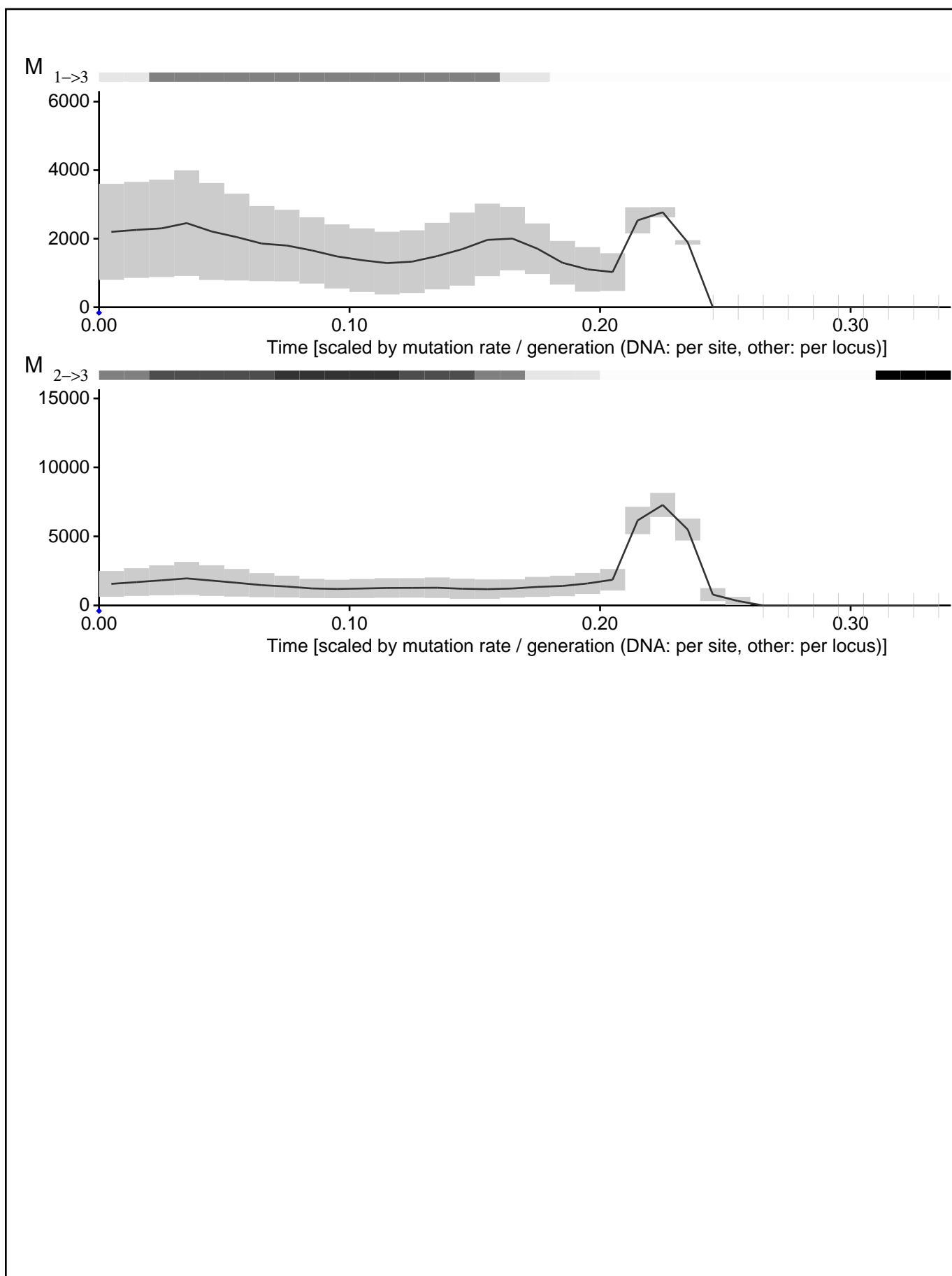




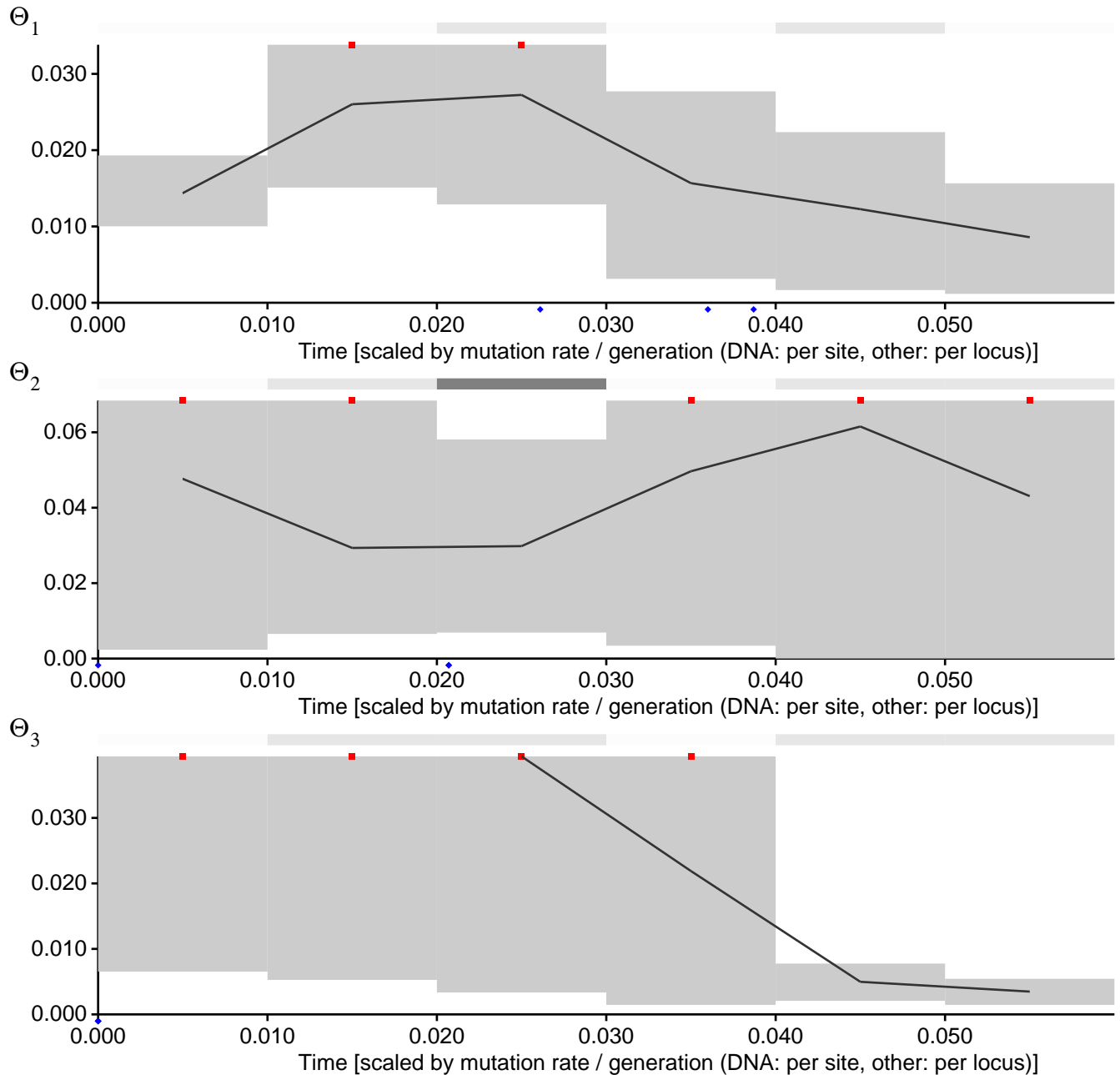
Parameter values through time

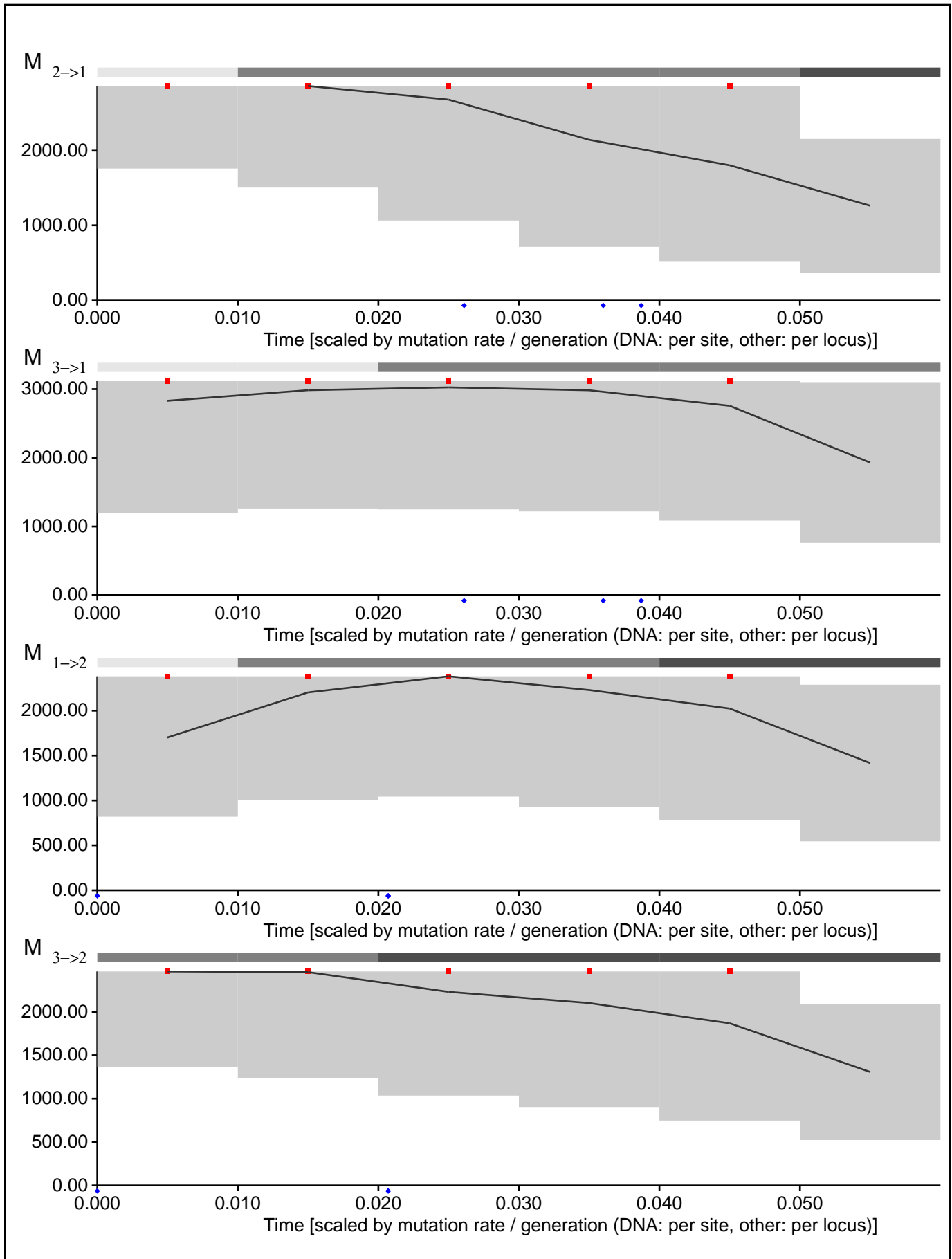


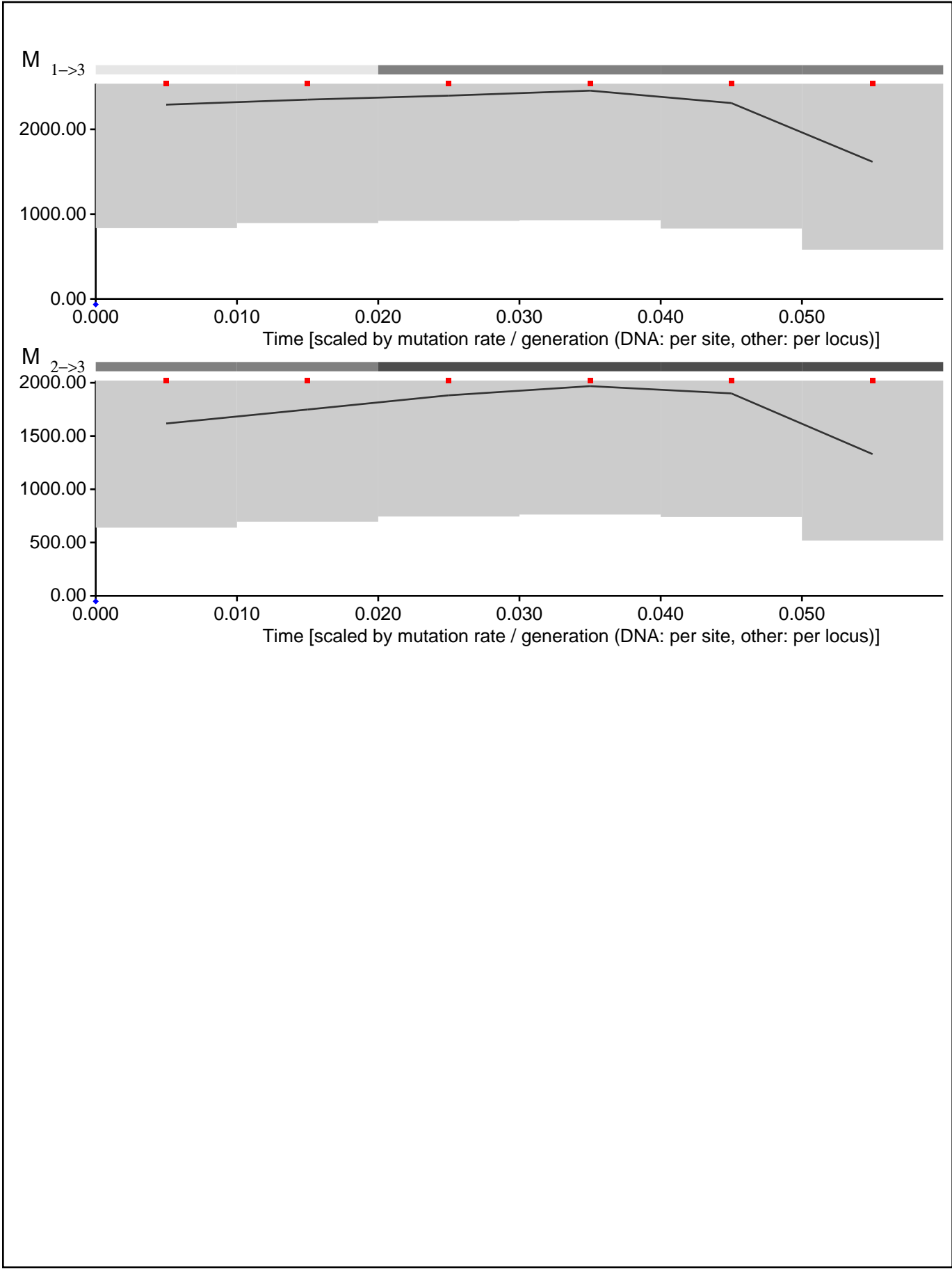




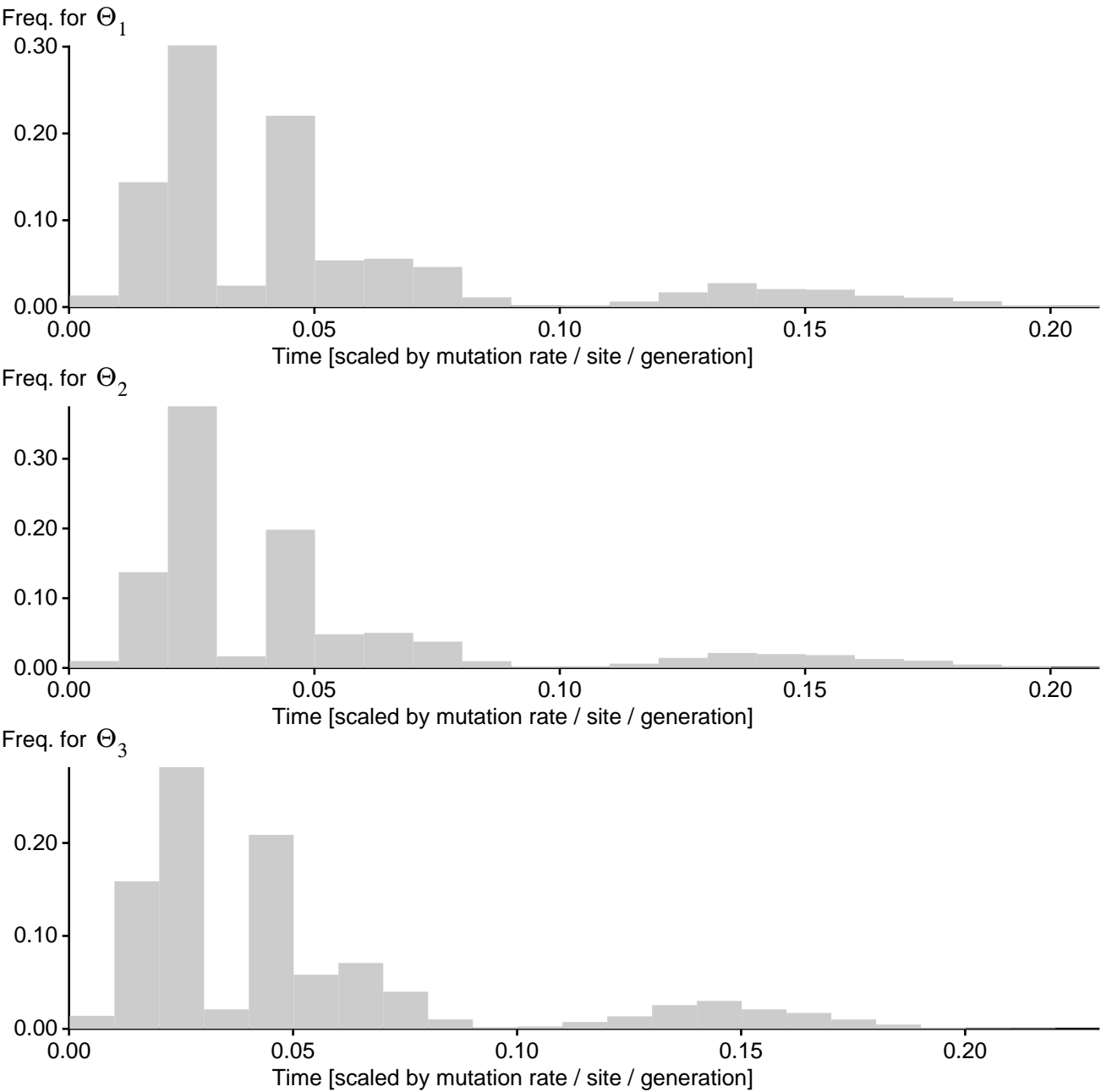
Parameter values through RECENT time

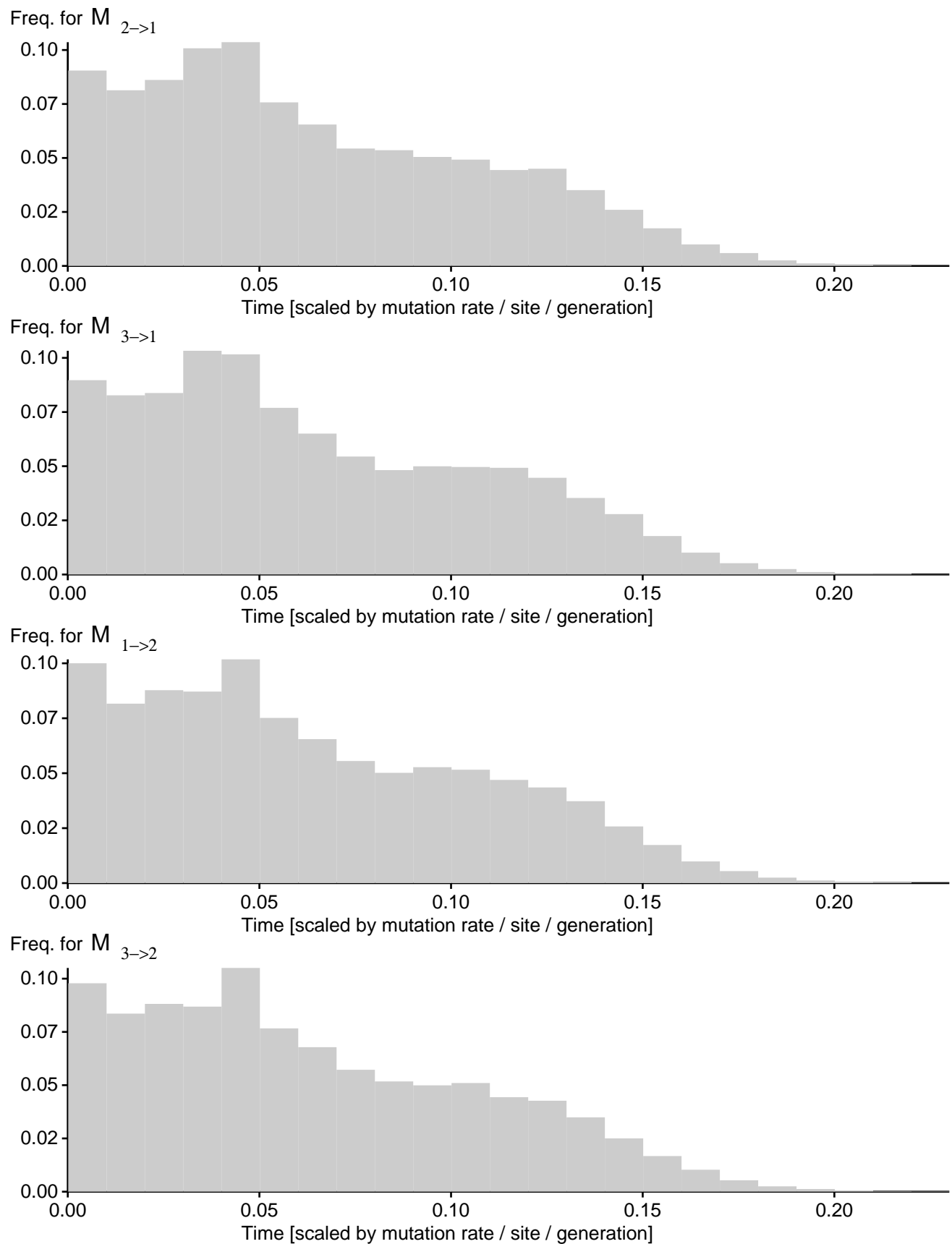


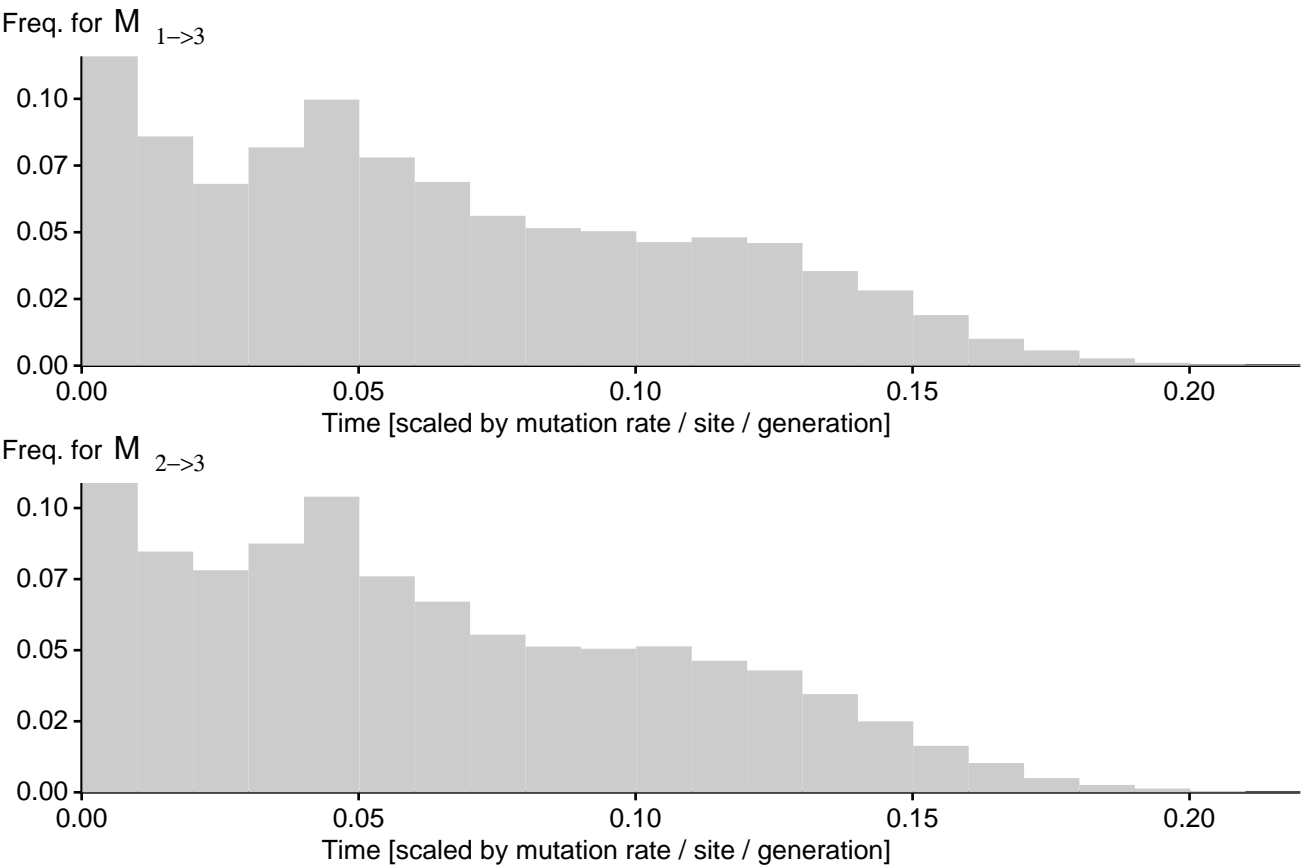




Events through time







Summary statistics of events through time

Locus 1					
Population		Time			Frequency
From	To	Average	Median	Std	
1	1	0.050697	0.045000	0.042013	0.006487
2	2	0.047779	0.025000	0.040524	0.011977
3	3	0.051424	0.045000	0.042599	0.010240
2	1	0.064603	0.055000	0.044301	0.141157
3	1	0.064823	0.055000	0.044402	0.124801
1	2	0.064602	0.055000	0.044646	0.172077
3	2	0.064060	0.055000	0.044284	0.204660
1	3	0.064593	0.055000	0.045257	0.090436
2	3	0.063778	0.055000	0.044554	0.238166

Time and probability of location of most recent common ancestor

Locus 1 Population	Time			Frequency
	Average	Median	Std	
1	0.148419	0.145000	0.021639	0.232333
2	0.148448	0.145000	0.021448	0.382739
3	0.147933	0.145000	0.020349	0.384928

Legend for Skyline and Event plots

Skyline plots:

Skyline plots visualize the changes of population sizes and migration rates through time (today is on the left side and time is measured into the past. The time scale is in units of expected mutations per generation. To calculate the absolute time scale you must supply an mutation rate per year and the duration of a generation in years in the data option. You can calculate the absolute time by multiplying the scale by generation time times mutation rate per year (per site for DNA; per locus for all other datatypes).

With estimated mutation rate only the combined rate modifier is plotted.
[this will change to mutation rate plot].

The gray bars cover one approximate standard deviation up and down from the expected value. The bar with different shades of gray on top of each plot indicates the number of values that were used to calculate the expected value, white means there are very few and black means that there were man thousands of samples per bin.

On some plots one can see red squares below the grayscale bar, these suggest that either the upper quantile and/or the main value was higher than the visible part of the axis.

Event histograms:

All accepted events (migration events, coalescent events) are recorded and their frequency are shown as histograms over time with recent time on the left side. The frequency plots of populations with constant size and constant immigration rates show histograms that are similar to exponential distribution, if the populations come from a divergence model without migration then the frequency of migration events can show a peak in the past.

Log-Probability of the data given the model (marginal likelihood)

Use this value for Bayes factor calculations:

$BF = \text{Exp}[\ln(\text{Prob}(D \mid \text{thisModel})) - \ln(\text{Prob}(D \mid \text{otherModel}))]$

shows the support for thisModel]

Method	$\ln(\text{Prob}(D \mid \text{Model}))$	Notes
Thermodynamic integration	(not estimated [no heating])	(1)
Harmonic mean	-858.800441	(2)

(1 and 2) is an approximation to the marginal likelihood, make sure the program run long enough!

(1) and (2) should give a similar result, (2) is considered more crude than (1), but (1) needs heating with several well-spaced chains,

Acceptance ratios for all parameters and the genealogies

Parameter	Accepted changes	Ratio
Θ_1	17832/17832	1.00000
Θ_2	17862/17862	1.00000
Θ_3	17807/17807	1.00000
$M_{2 \rightarrow 1}$	18008/18008	1.00000
$M_{3 \rightarrow 1}$	18032/18032	1.00000
$M_{1 \rightarrow 2}$	17447/17447	1.00000
$M_{3 \rightarrow 2}$	17735/17735	1.00000
$M_{1 \rightarrow 3}$	17865/17865	1.00000
$M_{2 \rightarrow 3}$	17991/17991	1.00000
Genealogies	20301/159764	0.12707

MCMC-Autocorrelation and Effective MCMC Sample Size

Parameter	Autocorrelation	Effective Sample Size
Θ_1	0.33078	1877.15
Θ_2	0.39422	1388.40
Θ_3	0.35516	1667.21
$M_{2 \rightarrow 1}$	0.95186	175.07
$M_{3 \rightarrow 1}$	0.92450	140.28
$M_{1 \rightarrow 2}$	0.89893	104.57
$M_{3 \rightarrow 2}$	0.95920	65.37
$M_{1 \rightarrow 3}$	0.95801	89.46
$M_{2 \rightarrow 3}$	0.94742	193.51
$\text{Ln}[\text{Prob}(D G)]$	0.47943	1108.17