# *Richardsonius balteatus 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 Thu May 20 21:39:27 2021
Program finished at Thu May 20 23:22:41 2021 [Runtime:0000:01:43:14]

## *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) | 3941837968 |
| 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 Great_Salt_Lake | * | d | 0 |
| 2 Great_Salt_Lake | 0 | * | d |
| 3 Sevier_Desert | 0 | 0 | * |

Order of parameters:

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

| | | |
|---|---|---|
| 4 | $\Delta_{2->1}$ | <displayed> |
| 5 | $\sigma_{2->1}$ | <displayed> |
| 6 | $\Delta_{3->2}$ | <displayed> |
| 7 | $\sigma_{3->2}$ | <displayed> |

| | |
|---|---|
| Mutation rate among loci: | Mutation rate is constant |
| Analysis strategy: | Bayesian inference |
| -Population size estimation: | Exponential Distribution |
| -Geneflow estimation: | Exponential Distribution |
| -Divergence time estimation: | Normal Distribution Shortcut (mean and standard dev.) |

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 | 0.07143 |
| 2 | Theta | ** | Uniform | 0.000000 | 0.050 | 0.100 | 0.010 | 1500 | 0.07143 |
| 3 | Theta | ** | Uniform | 0.000000 | 0.050 | 0.100 | 0.010 | 1500 | 0.07143 |
| 4 | Splittime mean | ** | Uniform | 0.000000 | 0.250 | 0.500 | 0.050 | 1500 | 0.07143 |
| 5 | Splittime std | ** | Uniform | 0.000000 | 0.250 | 0.500 | 0.050 | 1500 | 0.07143 |
| 6 | Splittime mean | ** | Uniform | 0.000000 | 0.250 | 0.500 | 0.050 | 1500 | 0.07143 |
| 7 | Splittime std | ** | Uniform | 0.000000 | 0.250 | 0.500 | 0.050 | 1500 | 0.07143 |

[* * 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: | 1 |

Mutationmodel:

| Locus | Sublocus | Mutationmodel | Mutationmodel parameters |
|---|---|---|---|
| 1 | 1 | HKY | [Bf:0.32 0.22 0.15 0.31, kappa=6.700] |
| 1 | 2 | HKY | [Bf:0.26 0.28 0.17 0.29, kappa=6.700] |

Sites per locus

| Locus | Sites | |
|---|---|---|
| 1 | 959 | 1140 |

Site rate variation and probabilities:

| Locus | Sublocus | Region type | Rate of change | Probability | Patch size |
|---|---|---|---|---|---|
| 1 | 1 | 1 | 1.000 | 1.000 | 1.000 |
| 1 | 2 | 1 | 1.000 | 1.000 | 1.000 |

| Population | Locus | Gene copies data | (missing) |
|---|---|---|---|
| 1 Great_Salt_Lake | 1 | 13 | |
| 2 Great_Salt_Lake_Desert | 1 | 5 | |
| 3 Sevier_Desert | 1 | 10 | |
| Total of all populations | 1 | 28 | (0) |

# *Bayesian Analysis: Posterior distribution table*

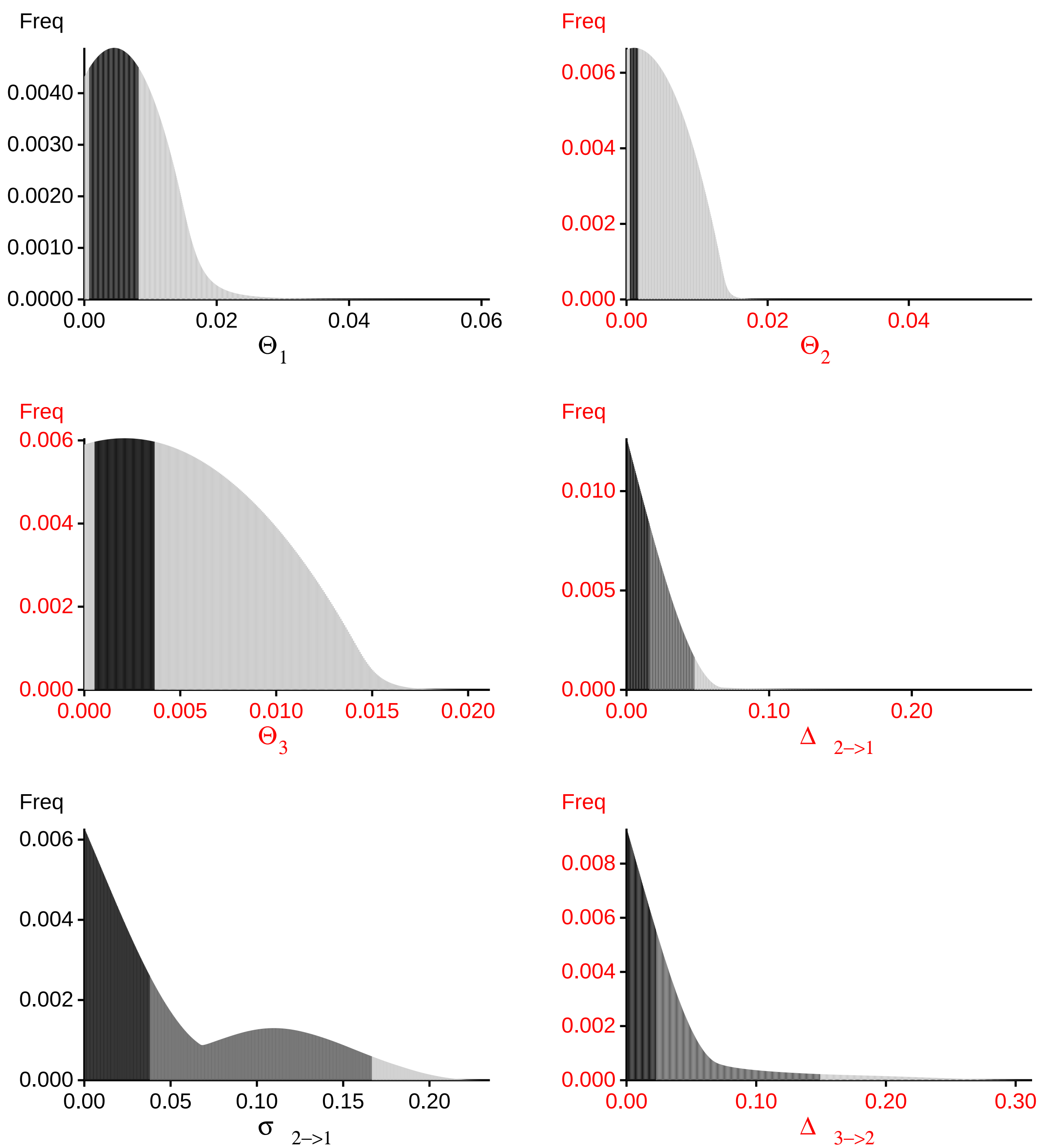
| Locus | Parameter | 2.5% | 25.0% | Mode | 75.0% | 97.5% | Median | Mean |
|---|---|---|---|---|---|---|---|---|
| 1 | $\Theta_1$ | 0.00067 | 0.00067 | 0.00450 | 0.00820 | 0.00820 | 0.00710 | 0.00471 |
| 1 | $\Theta_2$ | 0.00040 | 0.00040 | 0.00110 | 0.00167 | 0.00167 | 0.00517 | 0.00110 |
| 1 | $\Theta_3$ | 0.00047 | 0.00047 | 0.00210 | 0.00367 | 0.00367 | 0.00570 | 0.00210 |
| 1 | $D_{2->1}$ | 0.00000 | 0.00000 | 0.00017 | 0.01600 | 0.04767 | 0.01617 | 0.01640 |
| 1 | $S_{2->1}$ | 0.00000 | 0.00000 | 0.00017 | 0.03800 | 0.16667 | 0.03817 | 0.05582 |
| 1 | $D_{3->2}$ | 0.00000 | 0.00000 | 0.00017 | 0.02267 | 0.14933 | 0.02283 | 0.07020 |
| 1 | $S_{3->2}$ | 0.00467 | 0.00467 | 0.02183 | 0.03800 | 0.03800 | 0.05750 | 0.14191 |

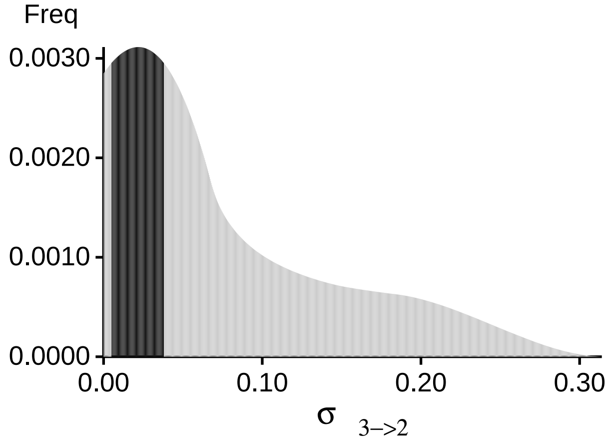
Citation suggestions:

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 for locus 1*




Freq
0.0030
0.0020
0.0010
0.0000
0.00
0.10
0.20
0.30
σ 3->2

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

Use this value for Bayes factor calculations:
BF = Exp[ ln(Prob(D | thisModel) - ln( Prob( D | otherModel)
or as LBF = 2 (ln(Prob(D | thisModel) - ln( Prob( D | otherModel))
shows the support for thisModel]

| Method | ln(Prob(D\|Model)) | Notes |
|---|---|---|
| Thermodynamic integration | -4089.591258 | (1a) |
| | -3457.477828 | (1b) |
| Harmonic mean | -3355.636258 | (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 |
|---|---|---|
| $\Theta_1$ | 411591/1429477 | 0.28793 |
| $\Theta_2$ | 630139/1428754 | 0.44104 |
| $\Theta_3$ | 951088/1428428 | 0.66583 |
| $\Delta_{2->1}$ | 802758/1427936 | 0.56218 |
| $\sigma_{2->1}$ | 894296/1427667 | 0.62640 |
| $\Delta_{3->2}$ | 643223/1427241 | 0.45068 |
| $\sigma_{3->2}$ | 671669/1427752 | 0.47044 |
| Genealogies | 2142872/10002745 | 0.21423 |

# *MCMC-Autocorrelation and Effective MCMC Sample Size*

| Parameter | Autocorrelation | Effective Sampe Size |
|---|---|---|
| $\Theta_1$ | 0.18865 | 13847.84 |
| $\Theta_2$ | 0.36130 | 10413.44 |
| $\Theta_3$ | 0.55077 | 7804.86 |
| $\Delta_{2->1}$ | 0.28793 | 11989.92 |
| $\sigma_{2->1}$ | 0.62839 | 5903.92 |
| $\Delta_{3->2}$ | 0.19108 | 13687.94 |
| $\sigma_{3->2}$ | 0.21579 | 12900.19 |
| Genealogies | 0.21579 | 12900.19 |

# *Potential Problems*

This section reports potential problems with your run, but such reporting is often not very accurate. Whith many parameters in a multilocus analysi s, it is very common that some parameters for some loci will not be very informative, triggering suggestions (for example to increase the prior ran ge) that are not sensible. This suggestion tool will improve with time, therefore do not blindly follow its suggestions. If some parameters are fla gged, inspect the tables carefully and judge wether an action is required. For example, if you run a Bayesian inference with sequence data, for mac roscopic 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 rou tes 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