

two (fake) Swiss towns

POPULATION SIZE, MIGRATION, DIVERGENCE, ASSIGNMENT, HISTORY

Bayesian inference using the structured coalescent

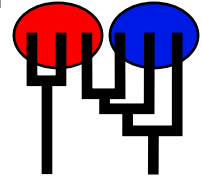
Migrate-n version 4.2.8 [June-24-2016]

Using Intel AVX (Advanced Vector Extensions)

Compiled for a SYMMETRIC multiprocessors (Grandcentral)

Program started at Tue Aug 2 10:56:58 2016

Program finished at Tue Aug 2 10:57:26 2016 [Runtime:0000:00:00:28]



Options

Inheritance scalers in use for Thetas:

1.00 0.25

[The locus with a scaler of 1.0 used as reference]

Random number seed:

(with internal timer)

1010124635

Start parameters:

Theta values were generated

Using a percent value of the prior

M values were generated

Using a percent value of 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 |
| 1 Aadorf | * | 0 |
| 2 Bern | * | * |

Order of parameters:

| | | |
|---|-----------------------|-------------|
| 1 | Θ_1 | <displayed> |
| 2 | Θ_2 | <displayed> |
| 3 | $M_{1 \rightarrow 2}$ | <displayed> |

Mutation rate among loci:

Mutation rate is constant for all loci

Analysis strategy:

Bayesian inference

Proposal distributions for parameter

| Parameter | Proposal |
|-----------|---------------------|
| Theta | Metropolis sampling |
| M | Metropolis sampling |

Prior distribution for parameter

| Parameter | Prior | Minimum | Mean* | Maximum | Delta | Bins |
|-----------|---------|----------|------------|-------------|------------|------|
| Theta | Uniform | 0.000000 | 0.050000 | 0.100000 | 0.010000 | 1500 |
| Theta | Uniform | 0.000000 | 0.050000 | 0.100000 | 0.010000 | 1500 |
| M | Uniform | 0.000000 | 500.000000 | 1000.000000 | 100.000000 | 1500 |
| M | Uniform | 0.000000 | 500.000000 | 1000.000000 | 100.000000 | 1500 |

Markov chain settings:

Long chain

| | |
|--|--------|
| Number of chains | 1 |
| Recorded steps [a] | 5000 |
| Increment (record every x step [b]) | 5 |
| Number of concurrent chains (replicates) [c] | 4 |
| Visited (sampled) parameter values [a*b*c] | 100000 |
| Number of discard trees per chain (burn-in) | 5000 |

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: | twoswisstowns |
| Haplotyping is turned on: | YES: report of haplotype probabilities |
| Output file: | outfile-twoswisstowns2 |
| 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: twoswisstowns
 Datatype: Haplotype data
 Number of loci: 2

Mutationmodel:

| Locus | Sublocus | Mutationmodel | Mutationmodel parameters |
|-------|----------|---------------|--------------------------|
| 1 | 1 | Jukes-Cantor | [Basefreq: =0.25] |
| 2 | 1 | Jukes-Cantor | [Basefreq: =0.25] |

Sites per locus

| Locus | Sites |
|-------|-------|
| 1 | 1000 |
| 2 | 1000 |

Site rate variation and probabilities:

| Locus | Sublocus | Region type | Rate of change | Probability | Patch size |
|-------|----------|-------------|----------------|-------------|------------|
| 1 | 1 | 1 | 0.349 | 0.511 | 1.000 |
| 1 | 1 | 2 | 1.438 | 0.431 | 1.000 |
| 1 | 1 | 3 | 3.425 | 0.057 | 1.000 |
| 1 | 1 | 4 | 6.788 | 0.001 | 1.000 |
| 2 | 1 | 1 | 0.349 | 0.511 | 1.000 |
| 2 | 1 | 2 | 1.438 | 0.431 | 1.000 |
| 2 | 1 | 3 | 3.425 | 0.057 | 1.000 |
| 2 | 1 | 4 | 6.788 | 0.001 | 1.000 |

| Population | Locus | Gene copies data | (missing) |
|--------------------------|-------|---------------------|-----------|
| 1 Aadorf | 1 | 10 | |
| | 2 | 10 | |
| 2 Bern | 1 | 10 | |
| | 2 | 10 | |
| Total of all populations | 1 | 20 | (0) |
| | 2 | 20 | (0) |

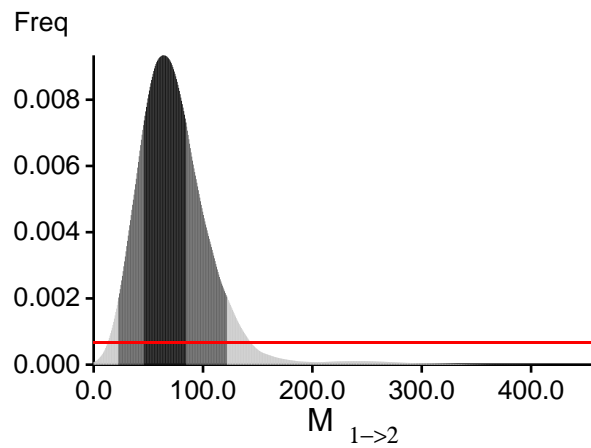
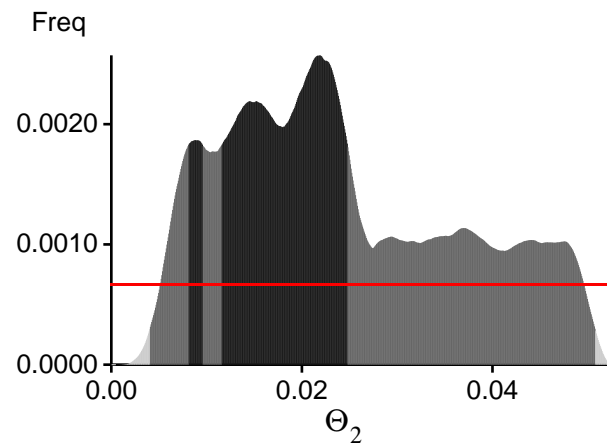
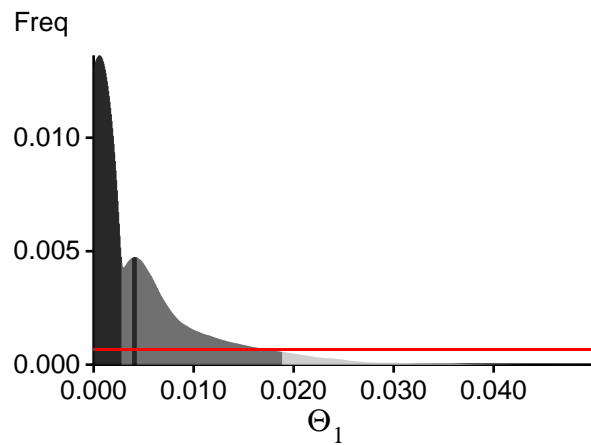
Bayesian Analysis: Posterior distribution table

| Locus | Parameter | 2.5% | 25.0% | Mode | 75.0% | 97.5% | Median | Mean |
|-------|-----------------------|---------|---------|---------|---------|---------|---------|---------|
| 1 | Θ_1 | 0.00000 | 0.00187 | 0.00410 | 0.00693 | 0.02120 | 0.00603 | 0.00803 |
| 1 | Θ_2 | 0.01547 | 0.03247 | 0.03697 | 0.03933 | 0.05060 | 0.03450 | 0.04567 |
| 1 | $M_{1 \rightarrow 2}$ | 3.333 | 24.000 | 52.333 | 93.333 | 242.000 | 82.333 | 102.047 |
| 2 | Θ_1 | 0.00000 | 0.00000 | 0.00030 | 0.00107 | 0.00240 | 0.00110 | 0.00032 |
| 2 | Θ_2 | 0.00640 | 0.01080 | 0.01343 | 0.01507 | 0.02313 | 0.01423 | 0.01450 |
| 2 | $M_{1 \rightarrow 2}$ | 16.000 | 48.667 | 71.000 | 95.333 | 153.333 | 79.000 | 82.524 |
| All | Θ_1 | 0.00000 | 0.00000 | 0.00057 | 0.00280 | 0.01887 | 0.00337 | 0.00419 |
| All | Θ_2 | 0.00400 | 0.01153 | 0.02190 | 0.02480 | 0.05080 | 0.02203 | 0.03327 |
| All | $M_{1 \rightarrow 2}$ | 22.000 | 45.333 | 63.667 | 84.667 | 122.000 | 70.333 | 75.228 |

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

Haplotype states and probabilities

Locus: 1

Locus: 2

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 | -2497.94 | -2271.62 | -2250.07 |
| 2 | -2885.99 | -2550.91 | -2485.31 |
| All | -5400.54 | -4839.14 | -4752.00 |

(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 = -16.610112]

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 | 11194/33112 | 0.33806 |
| Θ_2 | 20788/33396 | 0.62247 |
| $M_{1 \rightarrow 2}$ | 16839/33441 | 0.50354 |
| Genealogies | 17199/100051 | 0.17190 |

MCMC-Autocorrelation and Effective MCMC Sample Size

| Parameter | Autocorrelation | Effective Sample Size |
|-------------------------------|-----------------|-----------------------|
| Θ_1 | 0.89862 | 2163.41 |
| Θ_2 | 0.72510 | 6389.84 |
| $M_{1 \rightarrow 2}$ | 0.87487 | 2769.96 |
| $\text{Ln}[\text{Prob}(D G)]$ | 0.86244 | 3012.33 |

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