Trouble-shooting BEAST 2 Analyses

Tim Vaughan

Taming the BEAST Online 2021, Day 2

What could possibly go wrong?

Trouble shooting
Initialization Failure
One parameter does not mix
Many parameters do not mix
Problem solving in general

Many things! But we will focus on the following three possibilities:

What could possibly go wrong?

Trouble shooting
Initialization Failure
One parameter does not mix
Many parameters do not mix
Problem solving in general

Many things! But we will focus on the following three possibilities:

- MCMC chain fails to initialize.

What could possibly go wrong?

Trouble shooting
Initialization Failure
One parameter does not mix
Many parameters do not mix
Problem solving in general

Many things! But we will focus on the following three possibilities:

- MCMC chain fails to initialize.
- The chain runs, but one or more parameters mix slowly compared to the rest.

What could possibly go wrong?

Trouble shooting
Initialization Failure
One parameter does not mix
Many parameters do not mix
Problem solving in general

Many things! But we will focus on the following three possibilities:

- MCMC chain fails to initialize.
- The chain runs, but one or more parameters mix slowly compared to the rest.
- The chain runs, but many parameters fail to mix adequately.

Problem 1: MCMC chain fails to initialize


```
================================================================================
Start likelihood: -Infinity after 10 initialisation attempts
P(posterior) = -Infinity (was -Infinity)
        P(prior) = -Infinity (was -Infinity)
                P(CoalescentConstant.t:seqs) = -73.84438811380579 (was -73.84438811380579)
                P(ClockPrior.c:seqs) = 0.0 (was 0.0)
                P(PopSizePrior.t:seqs) = 1.2039728043259361 (was 1.2039728043259361)
                P(RateACPrior.s:seqs) = -Infinity (was -Infinity)
                P(RateAGPrior.s:seqs) = -3.1686658147294304 (was -3.1686658147294304)
                P(RateATPrior.s:seqs) = -3.184008455701433 (was -3.184008455701433)
                P(RateCGPrior.s:seqs) = -3.184008455701433 (was -3.184008455701433)
                P(RateGTPrior.s:seqs) = -3.184008455701433 (was -3.184008455701433)
        P(likelihood) = NaN (was NaN)  **
                P(treeLikelihood.seqs) = NaN (was NaN)  **
Fatal exception: Could not find a proper state to initialise. Perhaps try another seed.
See http://www.beast2.org/2018/07/04/fatal-errors.html for other possible solutions.
java.lang.RuntimeException: Could not find a proper state to initialise. Perhaps try another seed.
See http://www.beast2.org/2018/07/04/fatal-errors.html for other possible solutions.
        at beast.core.MCMC.run(Unknown Source)
        at beast.app.BeastMCMC.run(Unknown Source)
        at beast.app.beastapp.BeastMain.<init>(Unknown Source)
        at beast.app.beastapp.BeastMain.main(Unknown Source)
Fatal exception: Could not find a proper state to initialise. Perhaps try another seed.
See http://www.beast2.org/2018/07/04/fatal-errors.html for other possible solutions.
java.lang.RuntimeException: An error was encounted. Terminating BEAST
        at beast.app.util.ErrorLogHandler.publish(Unknown Source)
        at java.logging/java.util.logging.Logger.log(Logger.java:979)
        at java.logging/java.util.logging.Logger.doLog(Logger.java:1006)
        at java.logging/java.util.logging.Logger.log(Logger.java:1029)
        at java.logging/java.util.logging.Logger.severe(Logger.java:1776)
        at beast.app.beastapp.BeastMain.<init>(Unknown Source)
        at beast.app.beastapp.BeastMain.main(Unknown Source)
```

Trouble shooting
Initialization Failure
One parameter does not mix
Many parameters do not mix
Problem solving in general

Problem 1: MCMC chain fails to initialize

Trouble shooting
Initialization Failure
One parameter does not mix
Many parameters do not mix
Problem solving in general

```
================================================================================
Start likelihood: -Infinity after 10 initialisation attempts
P(posterior) = -Infinity (was -Infinity)
        P(prior) = -Infinity (was -Infinity)
                P(CoalescentConstant.t:seqs) = -73.84438811380579 (was -73.84438811380579)
                P(ClockPrior.c:seqs) = 0.0 (was 0.0)
                P(PopSizePrior.t:seqs) = 1.2039728043259361 (was 1.2039728043259361)
                P(RateACPrior.s:seqs) = -Infinity (was -Infinity)
                P(RateAGPrior.s:seqs) = -3.1686658147294304 (was -3.1686658147294304)
                P(RateATPrior.s:seqs) = -3.184008455701433 (was -3.184008455701433)
                P(RateCGPrior.s:seqs) = -3.184008455701433 (was -3.184008455701433)
                P(RateGTPrior.s:seqs) = -3.184008455701433 (was -3.184008455701433)
        P(likelihood) = NaN (was NaN)  **
                P(treeLikelihood.seqs) = NaN (was NaN)  **
Fatal exception: Could not find a proper state to initialise. Perhaps try another seed.
```

- Values quoted are $\log(P)$ values, so $-\infty$ corresponds to a probability of 0.

Problem 1: MCMC chain fails to initialize

Trouble shooting
Initialization Failure
One parameter does not mix
Many parameters do not mix
Problem solving in general

```
================================================================================
Start likelihood: -Infinity after 10 initialisation attempts
P(posterior) = -Infinity (was -Infinity)
        P(prior) = -Infinity (was -Infinity)
                P(CoalescentConstant.t:seqs) = -73.84438811380579 (was -73.84438811380579)
                P(ClockPrior.c:seqs) = 0.0 (was 0.0)
                P(PopSizePrior.t:seqs) = 1.2039728043259361 (was 1.2039728043259361)
                P(RateACPrior.s:seqs) = -Infinity (was -Infinity)
                P(RateAGPrior.s:seqs) = -3.1686658147294304 (was -3.1686658147294304)
                P(RateATPrior.s:seqs) = -3.184008455701433 (was -3.184008455701433)
                P(RateCGPrior.s:seqs) = -3.184008455701433 (was -3.184008455701433)
                P(RateGTPrior.s:seqs) = -3.184008455701433 (was -3.184008455701433)
        P(likelihood) = NaN (was NaN)  **
                P(treeLikelihood.seqs) = NaN (was NaN)  **
Fatal exception: Could not find a proper state to initialise. Perhaps try another seed.
```

- Values quoted are $\log(P)$ values, so $-\infty$ corresponds to a probability of 0.
- Note that posterior of starting state is 0.

Problem 1: MCMC chain fails to initialize


```
================================================================================
Start likelihood: -Infinity after 10 initialisation attempts
P(posterior) = -Infinity (was -Infinity)
        P(prior) = -Infinity (was -Infinity)
                P(CoalescentConstant.t:seqs) = -73.84438811380579 (was -73.84438811380579)
                P(ClockPrior.c:seqs) = 0.0 (was 0.0)
                P(PopSizePrior.t:seqs) = 1.2039728043259361 (was 1.2039728043259361)
                P(RateACPrior.s:seqs) = -Infinity (was -Infinity)
                P(RateAGPrior.s:seqs) = -3.1686658147294304 (was -3.1686658147294304)
                P(RateATPrior.s:seqs) = -3.184008455701433 (was -3.184008455701433)
                P(RateCGPrior.s:seqs) = -3.184008455701433 (was -3.184008455701433)
                P(RateGTPrior.s:seqs) = -3.184008455701433 (was -3.184008455701433)
        P(likelihood) = NaN (was NaN)  **
                P(treeLikelihood.seqs) = NaN (was NaN)  **
Fatal exception: Could not find a proper state to initialise. Perhaps try another seed.
```

- Values quoted are $\log(P)$ values, so $-\infty$ corresponds to a probability of 0.
- Note that posterior of starting state is 0.
- Likewise the prior of the starting state is 0.

Trouble shooting
Initialization Failure
One parameter does not mix
Many parameters do not mix
Problem solving in general

Problem 1: MCMC chain fails to initialize


```
================================================================================
Start likelihood: -Infinity after 10 initialisation attempts
P(posterior) = -Infinity (was -Infinity)
        P(prior) = -Infinity (was -Infinity)
                P(CoalescentConstant.t:seqs) = -73.84438811380579 (was -73.84438811380579)
                P(ClockPrior.c:seqs) = 0.0 (was 0.0)
                P(PopSizePrior.t:seqs) = 1.2039728043259361 (was 1.2039728043259361)
                P(RateACPrior.s:seqs) = -Infinity (was -Infinity)
                P(RateAGPrior.s:seqs) = -3.1686658147294304 (was -3.1686658147294304)
                P(RateATPrior.s:seqs) = -3.184008455701433 (was -3.184008455701433)
                P(RateCGPrior.s:seqs) = -3.184008455701433 (was -3.184008455701433)
                P(RateGTPrior.s:seqs) = -3.184008455701433 (was -3.184008455701433)
        P(likelihood) = NaN (was NaN)  **
                P(treeLikelihood.seqs) = NaN (was NaN)  **
Fatal exception: Could not find a proper state to initialise. Perhaps try another seed.
```

- Values quoted are $\log(P)$ values, so $-\infty$ corresponds to a probability of 0.
- Note that posterior of starting state is 0.
- Likewise the prior of the starting state is 0.
- The likelihood is quoted as having a value of "NaN" (not a number) because this was never computed by BEAST: it stopped when encountering a zero in the prior.

Trouble shooting
Initialization Failure
One parameter does not mix
Many parameters do not mix
Problem solving in general

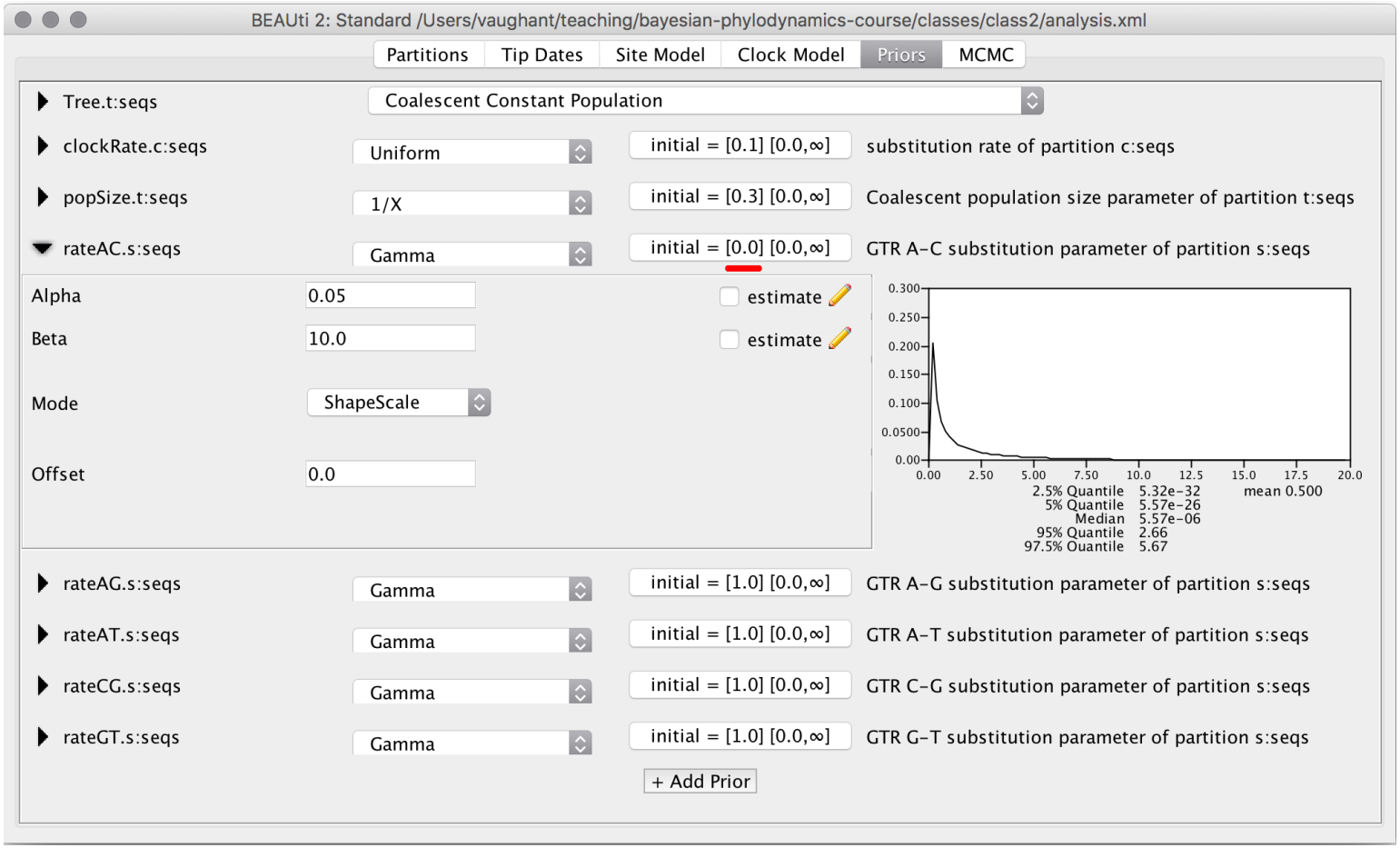
Problem 1: MCMC chain fails to initialize


```
================================================================================
Start likelihood: -Infinity after 10 initialisation attempts
P(posterior) = -Infinity (was -Infinity)
        P(prior) = -Infinity (was -Infinity)
                P(CoalescentConstant.t:seqs) = -73.84438811380579 (was -73.84438811380579)
                P(ClockPrior.c:seqs) = 0.0 (was 0.0)
                P(PopSizePrior.t:seqs) = 1.2039728043259361 (was 1.2039728043259361)
                P(RateACPrior.s:seqs) = -Infinity (was -Infinity)
                P(RateAGPrior.s:seqs) = -3.1686658147294304 (was -3.1686658147294304)
                P(RateATPrior.s:seqs) = -3.184008455701433 (was -3.184008455701433)
                P(RateCGPrior.s:seqs) = -3.184008455701433 (was -3.184008455701433)
                P(RateGTPrior.s:seqs) = -3.184008455701433 (was -3.184008455701433)
        P(likelihood) = NaN (was NaN)  **
                P(treeLikelihood.seqs) = NaN (was NaN)  **
Fatal exception: Could not find a proper state to initialise. Perhaps try another seed.
```

- Values quoted are $\log(P)$ values, so $-\infty$ corresponds to a probability of 0.
- Note that posterior of starting state is 0.
- Likewise the prior of the starting state is 0.
- The likelihood is quoted as having a value of "NaN" (not a number) because this was never computed by BEAST: it stopped when encountering a zero in the prior.
- The cause of the problem seems to be the 0 in the prior on the starting A→C transition rate.

Trouble shooting
Initialization Failure
One parameter does not mix
Many parameters do not mix
Problem solving in general

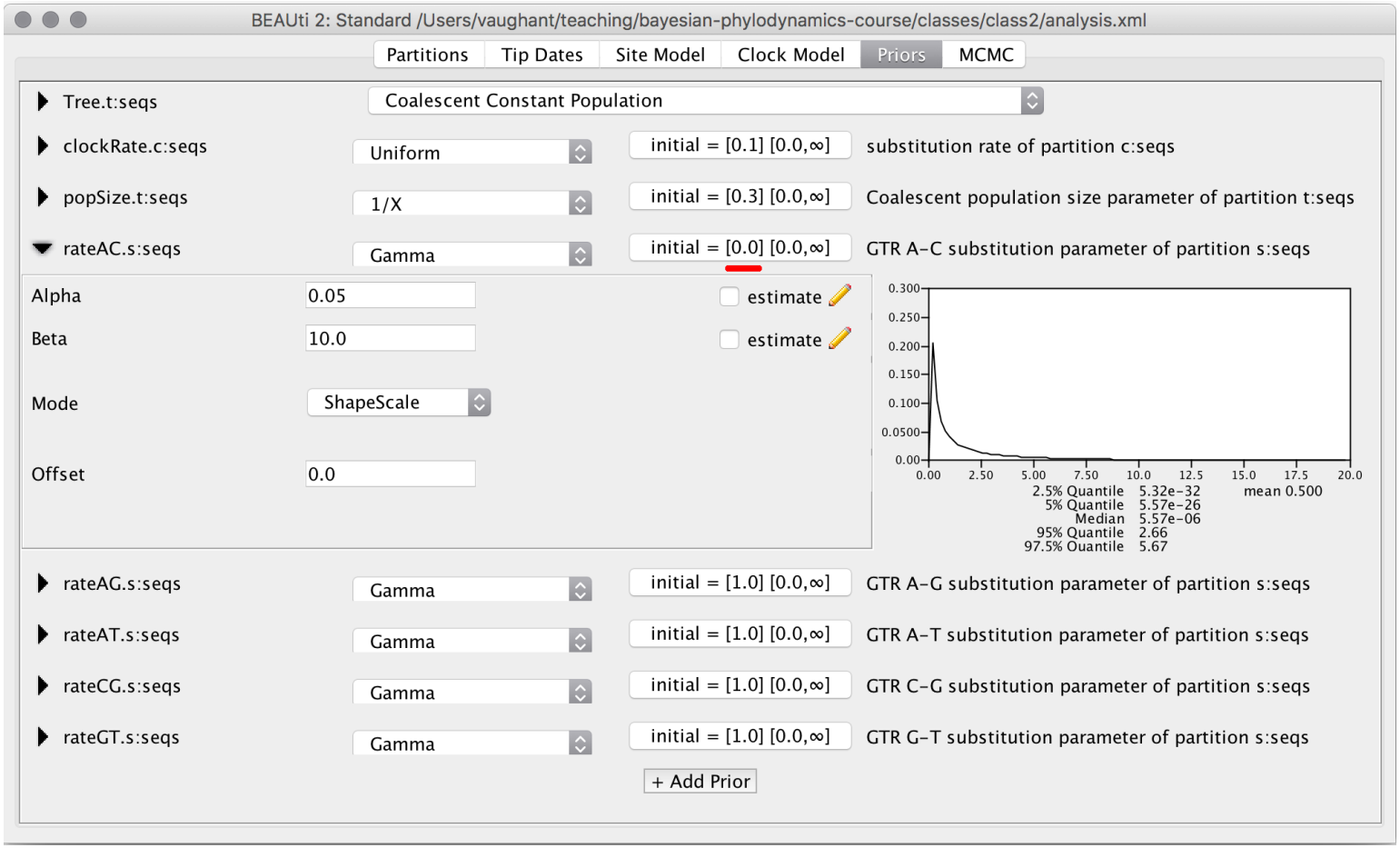
Problem 1: MCMC chain fails to initialize



Trouble shooting

- Initialization Failure
- One parameter does not mix
- Many parameters do not mix
- Problem solving in general

Problem 1: MCMC chain fails to initialize

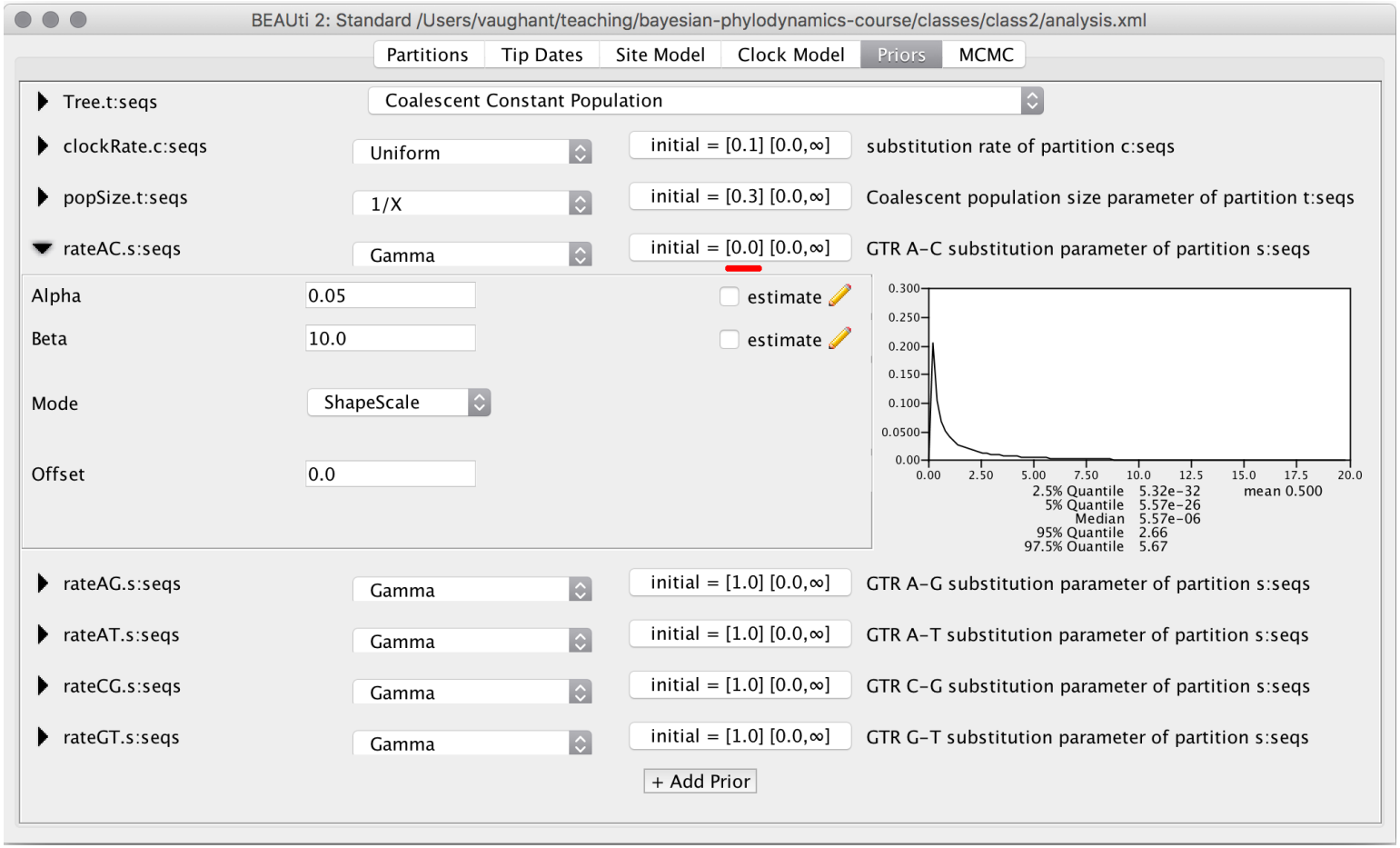


- Problem is that A→C rate is initially zero, while the prior on A→C excludes zero.

Trouble shooting
Initialization Failure
One parameter does not mix
Many parameters do not mix
Problem solving in general

Problem 1: MCMC chain fails to initialize

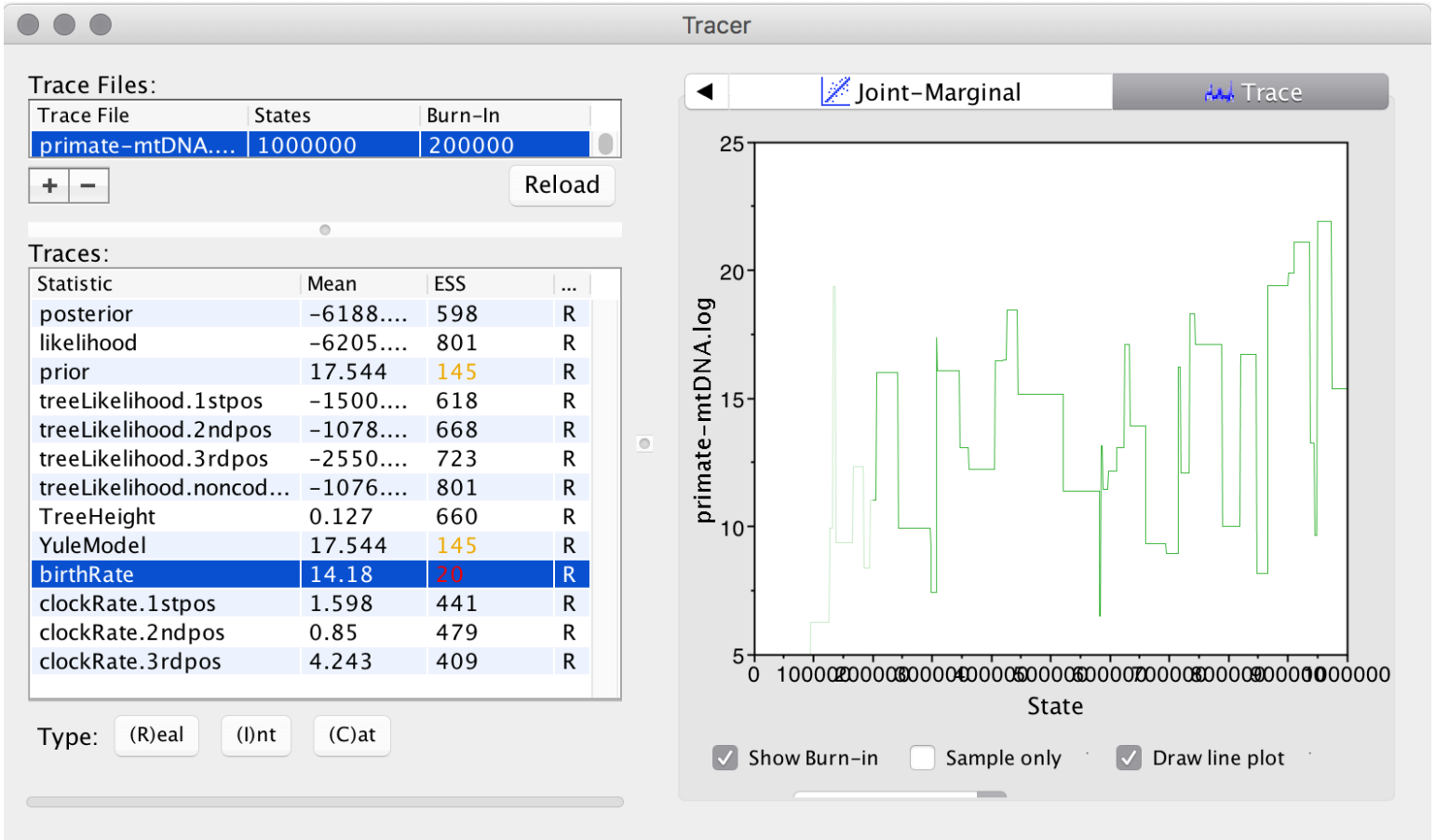
Trouble shooting
Initialization Failure
One parameter does not mix
Many parameters do not mix
Problem solving in general



- Problem is that A→C rate is initially zero, while the prior on A→C excludes zero.
- Fix is to change the starting value (proper solution in this particular situation) or to modify the prior to allow the starting value.

Problem 2: One parameter mixes slowly

Trouble shooting
Initialization Failure
One parameter does not mix
Many parameters do not mix
Problem solving in general



- Tracer output of run shows the birth rate parameter is mixing much slower than the others.

Why might MCMC mix poorly?

Trouble shooting
Initialization Failure
One parameter does not mix
Many parameters do not mix
Problem solving in general

- Mixing rate refers to how quickly (in terms of computation time) it takes the algorithm to explore the space of possible parameter values and trees.

Why might MCMC mix poorly?

Trouble shooting
Initialization Failure
One parameter does not mix
Many parameters do not mix
Problem solving in general

- Mixing rate refers to how quickly (in terms of computation time) it takes the algorithm to explore the space of possible parameter values and trees.
- ESS (effective sample size) provides an *estimate* of how thoroughly the state space has been explored. (I.e. large ESS = good.)

Why might MCMC mix poorly?

Trouble shooting
Initialization Failure
One parameter does not mix
Many parameters do not mix
Problem solving in general

- Mixing rate refers to how quickly (in terms of computation time) it takes the algorithm to explore the space of possible parameter values and trees.
- ESS (effective sample size) provides an *estimate* of how thoroughly the state space has been explored. (I.e. large ESS = good.)
- Mixing rate is determined by the combination of the posterior distribution (function of the data) and the random walk strategy used to explore the state space.

Why might MCMC mix poorly?

Trouble shooting
Initialization Failure
One parameter does not mix
Many parameters do not mix
Problem solving in general

- Mixing rate refers to how quickly (in terms of computation time) it takes the algorithm to explore the space of possible parameter values and trees.
- ESS (effective sample size) provides an *estimate* of how thoroughly the state space has been explored. (I.e. large ESS = good.)
- Mixing rate is determined by the combination of the posterior distribution (function of the data) and the random walk strategy used to explore the state space.
- Random walk strategy defined by a series of "operators" that modify the state (parameter values or trees). Operator "weights" determine how frequently a given operator is used.

Why might MCMC mix poorly?

Trouble shooting
Initialization Failure
One parameter does not mix
Many parameters do not mix
Problem solving in general

- Mixing rate refers to how quickly (in terms of computation time) it takes the algorithm to explore the space of possible parameter values and trees.
- ESS (effective sample size) provides an *estimate* of how thoroughly the state space has been explored. (I.e. large ESS = good.)
- Mixing rate is determined by the combination of the posterior distribution (function of the data) and the random walk strategy used to explore the state space.
- Random walk strategy defined by a series of "operators" that modify the state (parameter values or trees). Operator "weights" determine how frequently a given operator is used.
- A single slow-mixing parameter may be the result of the operator responsible for adjusting that value having a weight which is too low.

Problem 2: One parameter mixes slowly

Trouble shooting
Initialization Failure
One parameter does not mix
Many parameters do not mix
Problem solving in general

BEAUti 2: Standard /Users/vaughant/Downloads/primates.xml

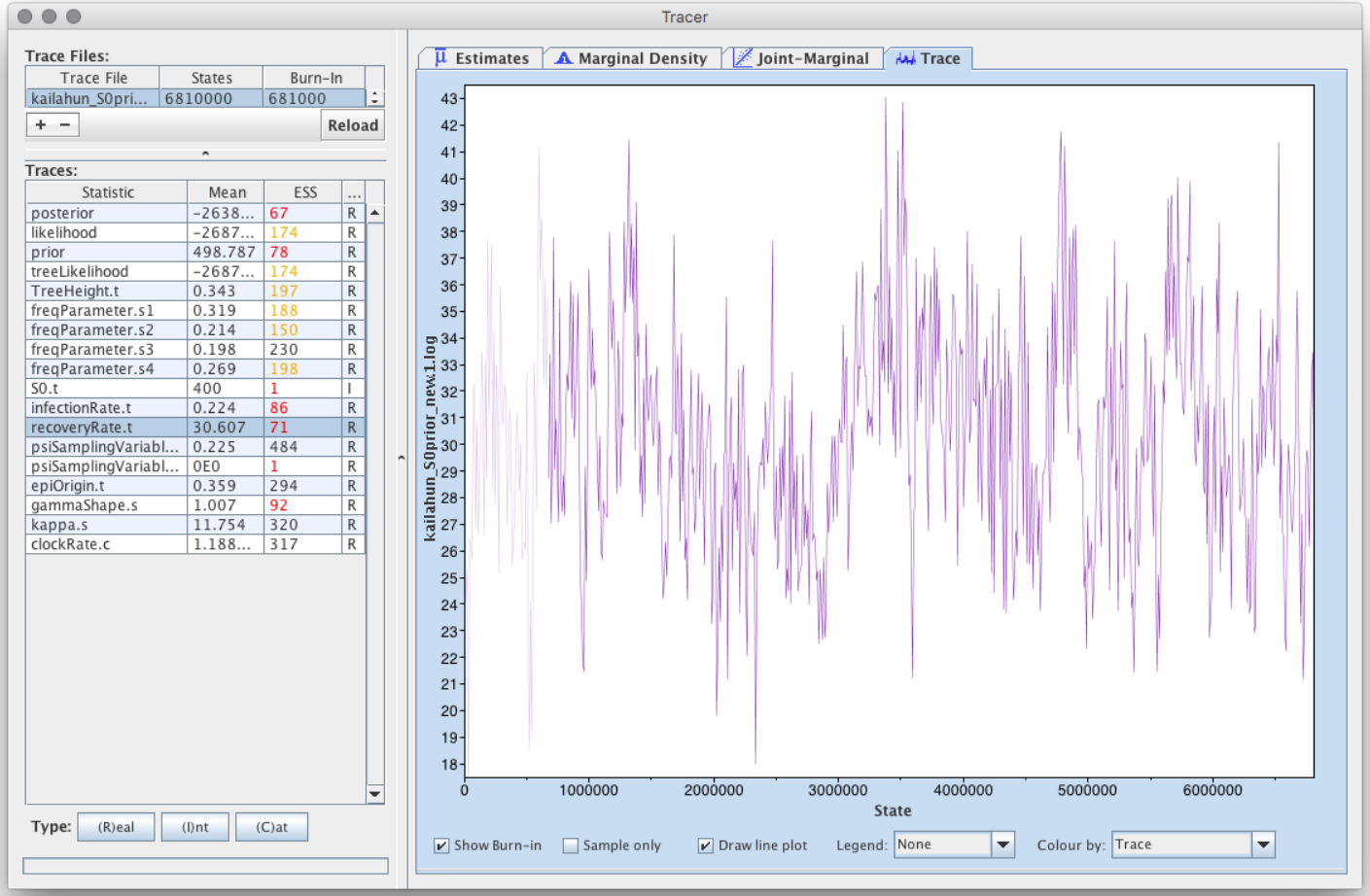
Partitions | Tip Dates | Site Model | Clock Model | Priors | Operators | MCMC

Scale: birthRate.t:tree Scales birth rate of Yule prior for partition.t:tree	0.01
Scale: Tree.t:tree Scales all internal nodes for tree.t:tree	3.0
Scale: Tree.t:tree Scales root node for tree.t:tree	3.0
Uniform: Tree.t:tree Draws new internal node heights uniformally for tree.t:tree	30.0
Subtree Slide: Tree.t:tree Performs subtree slide rearrangement of tree.t:tree	15.0
Exchange: Tree.t:tree Narrow exchange performs local rearrangement of tree.t:tree	15.0
Exchange: Tree.t:tree Wide exchange performs global rearrangement of tree.t:tree	3.0
Wilson Balding: Tree.t:tree Performs Wilson–Balding global rearrangement of tree.t:tree	3.0
Scale: clockRate.c:3rdpos Scale substitution rate of partition c:3rdpos	3.0
Scale: clockRate.c:1stpos Scale substitution rate of partition c:1stpos	3.0
Scale: clockRate.c:2ndpos Scale substitution rate of partition c:2ndpos	3.0
Up Down: clockRate.c:2ndpos Tree.t:tree Scale up substitution rate c:2ndpos and scale down tree t:($n)	3.0
Up Down: clockRate.c:1stpos Tree.t:tree Scale up substitution rate c:1stpos and scale down tree t:($n)	3.0
Up Down: clockRate.c:3rdpos Tree.t:tree Scale up substitution rate c:3rdpos and scale down tree t:($n)	3.0

- Operator corresponding to birth rate parameter has a very low weight.
- Increasing this weight solves the mixing problem.

Problem 3: Many parameters mix slowly

Trouble shooting
Initialization Failure
One parameter does not mix
Many parameters do not mix
Problem solving in general



Problem 3: Many parameters mix slowly

Trouble shooting
Initialization Failure
One parameter does not mix
Many parameters do not mix
Problem solving in general

- Check burn-in percentage in tracer.

Problem 3: Many parameters mix slowly

Trouble shooting
Initialization Failure
One parameter does not mix
Many parameters do not mix
Problem solving in general

- Check burn-in percentage in tracer.
- In general, if an analysis is not mixing well there is not much you can do besides run the analysis longer.

Problem 3: Many parameters mix slowly

Trouble shooting
Initialization Failure
One parameter does not mix
Many parameters do not mix
Problem solving in general

- Check burn-in percentage in tracer.
- In general, if an analysis is not mixing well there is not much you can do besides run the analysis longer.
- A very good idea is to run multiple chains (i.e. multiple independent analyses).
  - This way, once you are happy that each chain is independently sampling the same posterior, you can combine the results to get sufficient ESS.
  - Results can be combined using the LogCombiner utility that comes with BEAST.
  - ONLY combine the results if each of the independent chains has "converged".

Problem 3: Many parameters mix slowly

Trouble shooting
Initialization Failure
One parameter does not mix
Many parameters do not mix
Problem solving in general

- Check burn-in percentage in tracer.
- In general, if an analysis is not mixing well there is not much you can do besides run the analysis longer.
- A very good idea is to run multiple chains (i.e. multiple independent analyses).
  - This way, once you are happy that each chain is independently sampling the same posterior, you can combine the results to get sufficient ESS.
  - Results can be combined using the LogCombiner utility that comes with BEAST.
  - ONLY combine the results if each of the independent chains has "converged".
- Note that you can resume completed chains to extend the analysis. (Select the "resume" option in the BEAST start dialog box.)

General problem-solving flow chart

Trouble shooting
Initialization Failure
One parameter does not mix
Many parameters do not mix
Problem solving in general