

Setting priors and trouble shooting

Remco R. Bouckaert
remco@cs.auckland.ac.nz

Centre for Computational Evolution
Department of Computer Science, University of Auckland
Max Planck Institute for the Science of Human History

Seili, September 2017




Slides from Veronika Boskova, Chi Zhang, Joëlle Barido-Sottani, Denise Kühnert, Julija Pecerska, David Rasmussen


What is a prior?


- Using Bayes theorem, we can decompose the posterior:


$$P(\text{genetic sequences} | \text{genealogy}, \text{demographic model}, \text{substitution model}, \text{molecular clock model}) = \underbrace{P(\text{genealogy} | \text{genetic sequences})}_{P(\text{genetic sequences})} \underbrace{P(\text{demographic model} | \text{genealogy})}_{P(\text{genealogy})} \underbrace{P(\text{substitution model} | \text{demographic model})}_{P(\text{demographic model})} \underbrace{P(\text{molecular clock model} | \text{substitution model})}_{P(\text{substitution model})} P(\text{genetic sequences})$$

Tree prior
Prior information


 genetic sequences


 genealogy


 demographic model


 substitution model



 molecular clock model

Figure adapted from [du Plessis and Stadler, 2015]

Priors and starting values

Priors

Prior distribution

Tree prior

Substitution model prior

Clock prior

Parameter prior

Think twice

Starting values

References

- ▶ Allows us to include any information we have on the process, before looking at the data
 - ▶ Do not be afraid of using it in the inference
- ▶ Prior distribution does not have to, and is not expected to, be exactly the same as the posterior

Priors and starting values

Priors

Prior distribution

Tree prior

Substitution model prior

Clock prior

Parameter prior

Think twice

Starting values

References

- ▶ Should not be and is not universal for all the analyses you will ever do in your research
- ▶ Should incorporate prior (before looking at the data) knowledge about the parameter/underlying process
 - ▶ use results of previous independent experiments
 - ▶ use other independent evidence
- ▶ Should not be too restrictive if prior knowledge/assumptions are weak
 - ▶ One can use diffuse priors
- ▶ May not be adjusted after the run, to give higher and higher posterior support

Priors and starting values

Priors

Prior distribution

Tree prior

Substitution model prior

Clock prior

Parameter prior

Think twice

Starting values

References

- ▶ Is a choice of
 - ▶ model
 - ▶ tree-generating models, nucleotide/AA/codon substitution models, ...
- and of
 - ▶ distribution of plausible values for a parameter of interest
 - ▶ Uniform, Normal, Beta,...

Priors and starting values

Priors

Prior distribution

Tree prior

Substitution model prior

Clock prior

Parameter prior

Think twice

Starting values

References

- ▶ Have to pick one from Coalescent or Birth-death process framework
- ▶ Have to put priors on parameters of the chosen model
 - ▶ e.g. growth-rate of the population, R_0 , extinction rate, ...

Priors and starting values

Priors

Prior distribution

Tree prior

Substitution model prior

Clock prior

Parameter prior

Think twice

Starting values

References

- ▶ The selection is big: JC69, HKY85, ..., GTR
- ▶ Use model which has been previously identified to be best for your type of data
 - ▶ e.g. HKY85
 - ▶ Prior for transition/transversion rate ratio (κ)
 - ▶ Prior for base frequencies
- ▶ To choose the best model
 - ▶ Use model comparison to choose the one best fitting the data
 - ▶ Use rjMCMC directly in BEAST2 to sample from the posterior distribution including different substitution models. The model where rjMCMC spends the most time (samples the most from), is the best fitting model.

Priors and starting values

Priors

Prior distribution

Tree prior

Substitution model prior

Clock prior

Parameter prior

Think twice

Starting values

References

- ▶ Strict clock: all branches have the same clock rate
- ▶ Relaxed clock
 - ▶ Uncorrelated: branches have independent clock rate distributions
 - ▶ Correlated: child branch has clock rate distribution correlated to distribution of the parent branch

Priors and starting values

Priors

Prior distribution

Tree prior

Substitution model prior

Clock prior

Parameter prior

Think twice

Starting values

References

- ▶ Can be fixed to a given value
(though this is generally not recommended)
- ▶ Can have upper and lower limits
 - ▶ If we know that any infected individual recovers after 5-10 days, we can set the distribution of infectious period to be e.g. min 4 days and max 11 days
- ▶ If specified by a parametric distribution, the parameters of this distribution can also be assigned a prior (hyperprior)
- ▶ You can visualise the distribution in BEAUti

Priors and starting values

Priors

Prior distribution

Tree prior

Substitution model prior

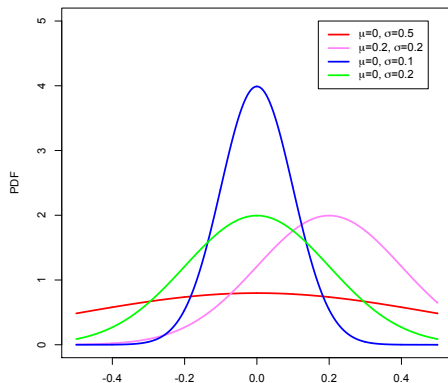
Clock prior

Parameter prior

Think twice

Starting values

References



- Parameters: mean $\mu \in \mathbb{R}$, standard deviation $\sigma > 0$
- Range of values: $(-\infty, \infty)$

Priors and starting values

Priors

Prior distribution

Tree prior

Substitution model prior

Clock prior

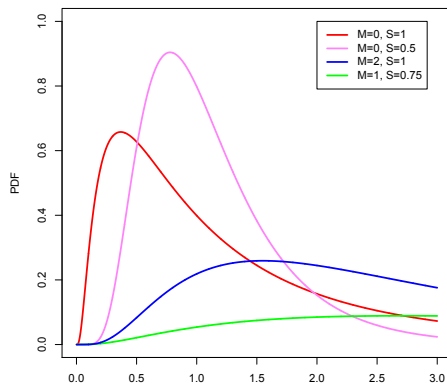
Parameter prior

Think twice

Starting values

References

Examples - LogNormal distribution



- Parameters: mean $M \in \mathbb{R}$, standard deviation $S > 0$
- Range of values: $[0, \infty)$
- Long tail, always positive

Priors and starting values

Priors

Prior distribution

Tree prior

Substitution model prior

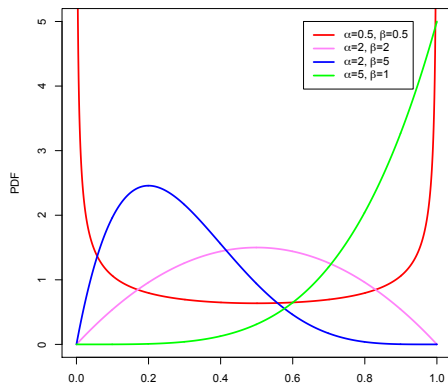
Clock prior

Parameter prior

Think twice

Starting values

References



- Parameters: shape $\alpha > 0$, shape $\beta > 0$
- Range of values: $[0,1]$
- Good for e.g. sampling probability prior

Priors and starting values

Priors

Prior distribution

Tree prior

Substitution model prior

Clock prior

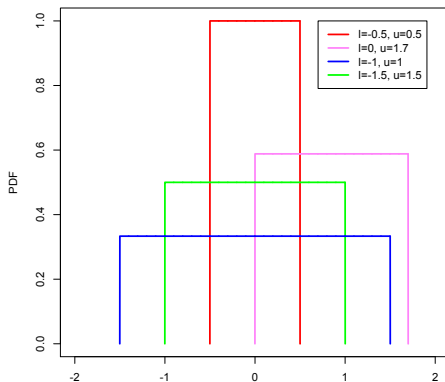
Parameter prior

Think twice

Starting values

References

Examples - Uniform distribution



- Parameters: lower, upper bound
- Range of values: $(-\infty, \infty)$

Priors and starting values

Priors

Prior distribution

Tree prior

Substitution model prior

Clock prior

Parameter prior

Think twice

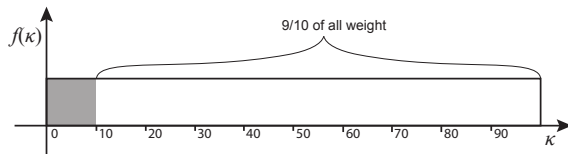
Starting values

References

Is uniform distribution a non-informative prior?

- Not really

- Imagine setting a $\text{Uniform}(0, 100)$ prior for the transition/transversion rate ratio (κ). You also know that the most likely values for κ are between 0 and 10. But you now put 9/10 of the weight to values > 10 .



- In fact there is nothing such as an non-informative prior
- If little or no information on the parameter is available, use diffuse priors
- Try to avoid $\text{Uniform}(-\infty, \infty)$ or $\text{Uniform}(0, \infty)$

Priors and starting values

Priors

Prior distribution

Tree prior

Substitution model prior

Clock prior

Parameter prior

Think twice

Starting values

References

- ▶ Sometimes the prior distribution is such that the sum or the integral of the prior values does not converge, this is called an IMPROPER prior
- ▶ Examples
 - ▶ $1/x$
 - ▶ $\text{Uniform}(-\infty, \infty)$

Priors and starting values

Priors

Prior distribution

Tree prior

Substitution model prior

Clock prior

Parameter prior

Think twice

Starting values

References

Are my priors what I set them to be?

- ▶ Not always
 - ▶ Induced priors may change the picture, i.e. if the parameters interact, the marginal prior distribution for each individual parameter may be different from the originally specified prior
- ▶ Use sampling from the prior, to see what your 'real' prior is

Priors and starting values

Priors

Prior distribution

Tree prior

Substitution model prior

Clock prior

Parameter prior

Think twice

Starting values

References

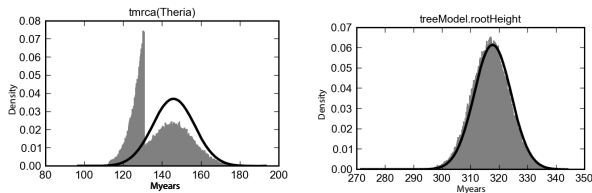


Figure adapted from [Heled and Drummond, 2012]

The marginal prior distributions that result from the multiplicative construction (gray) versus calibration densities (black line) specified for the calibrated nodes.

- ▶ Use all the prior knowledge you have to choose models and set appropriate parameter priors
- ▶ Sample from the prior distribution before using your data to check you really have the priors you want
- ▶ Check your posterior distribution against the prior

Priors and starting values

Priors

Prior distribution

Tree prior

Substitution model prior

Clock prior

Parameter prior

Think twice

Starting values

References

- ▶ In practice, it is important to evaluate the impact of the prior on the posterior in a Bayesian robustness analysis
- ▶ Ideally, the posterior should be dominated by your data, such that the choice of the prior has little influence on the result
- ▶ If this is not the case, the choice of prior is very important, and should be reported

Priors and starting values

Priors

Prior distribution

Tree prior

Substitution model prior

Clock prior

Parameter prior

Think twice

Starting values

References

- ▶ Are just starting values
- ▶ Have to be within the prior distribution, and its upper and lower limits, you chose for the parameter
- ▶ Use your best guess
 - ▶ BEAST2 attempts 10 times at most (can be changed) to initialize the run, but if the starting values are unreasonable, the runs may keep failing
- ▶ Start from different starting values to make sure the chains converge to the same distribution

Priors and starting values

Priors

Prior distribution

Tree prior

Substitution model prior

Clock prior

Parameter prior

Think twice

Starting values

References

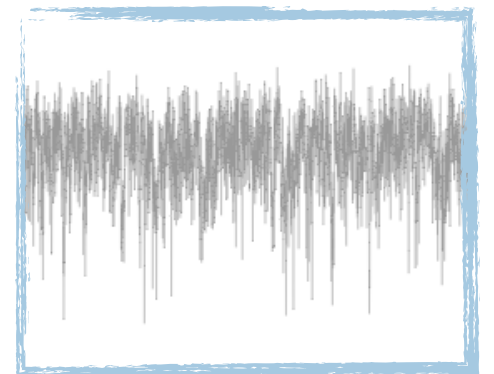
Possible scenarios

Beast analysis

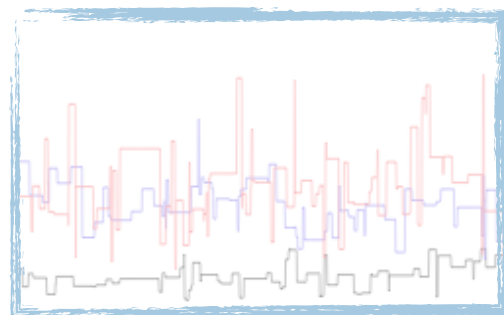
Initialisation
failed

Start likelihood: -Infinity after 1000
initialisation attempts
Fatal exception: Could not find a
proper state to initialise. Perhaps try
another seed.
P(posterior) = -Infinity (was -Infinity)

Everything
mixed



Nothing
mixed



One parameter
did not mix

kappa.noncoding	13.143	191
kappa.1stpos	6.28	376
kappa.2ndpos	8.643	372
kappa.3rdpos	27.988	92
mutationRate.noncodi...	0.347	273
mutationRate.1stpos	0.459	238
mutationRate.2ndpos	0.185	237

Failed initialisation

Start likelihood: -Infinity after 1000 initialisation attempts

Fatal exception: Could not find a proper state to initialise. Perhaps try another seed.

P(posterior) = -Infinity (was -Infinity)

P(prior) = -Infinity (was -Infinity)

P(BDMM) = -Infinity (was -Infinity)

P(R0Prior) = -0.5586849541070393 (was -0.5586849541070393)

P(rPrior) = -11.46042136866474 (was -11.46042136866474)

P(rateMatrixPrior) = -0.14088025499381485 (was -0.14088025499381485)

P(samplingProportionPrior) = -10.049507225748343 (was -10.049507225748343)

P(becomeUninfectiousRatePrior) = -0.7811241751317991 (was -0.7811241751317991)

java.lang.RuntimeException: Could not find a proper state to initialise. Perhaps try another seed.

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)

at beast.app.beastapp.BeastLauncher.main(Unknown Source)

Fatal exception: Could not find a proper state to initialise. Perhaps try another seed.

BEAST has terminated with an error. Please select QUIT from the menu.

Parameter prior is -Infinity

Example: $P(\text{rateMatrixPrior}) = -\text{Infinity}$ (was -Infinity)

Possible solutions:

- Change seed;
- Adjust initial conditions;
- Use excludable/slicing priors;
- Check for silly/incompatible priors;
- Check for underflow (too low values);
- Talk to the BEAST2 support group!

Model prior is -Infinity

Example: $P(\text{BDMM}) = -\text{Infinity}$ (was -Infinity)

Possible solutions:

- Change seed;
- Talk to the BEAST2 support group!
- Rescale time (e.g. from year to milenia)

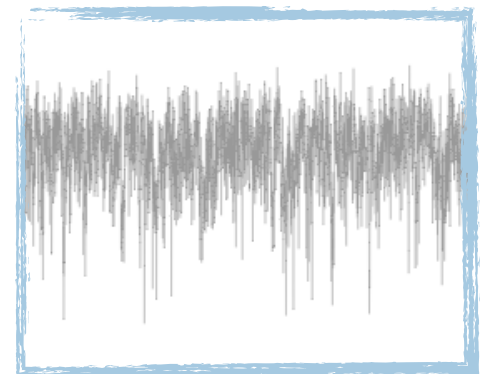
Possible scenarios

Beast analysis

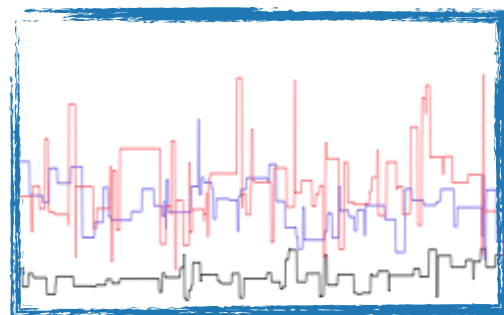
Initialisation
failed

Start likelihood: -Infinity after 1000
initialisation attempts
Fatal exception: Could not find a
proper state to initialise. Perhaps try
another seed.
P(posterior) = -Infinity (was -Infinity)

Everything
mixed



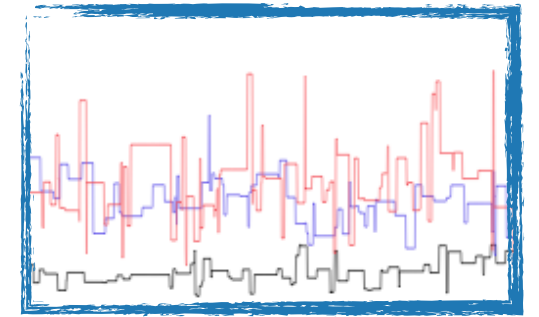
Nothing
mixed



One parameter
did not mix

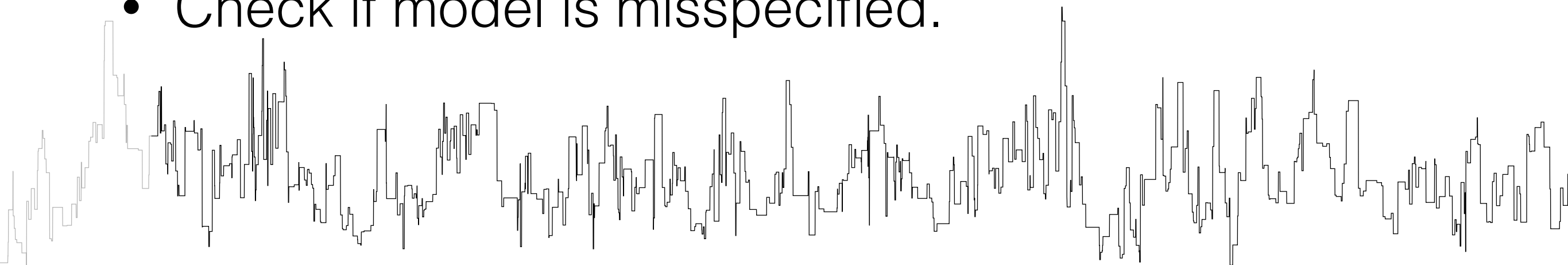
kappa.noncoding	13.143	191
kappa.1stpos	6.28	376
kappa.2ndpos	8.643	372
kappa.3rdpos	27.988	92
mutationRate.noncodi...	0.347	273
mutationRate.1stpos	0.459	238
mutationRate.2ndpos	0.185	237

Nothing mixed



Possible solutions:

- Increase chain length;
- Run multiple independent chains;
- Increase sampling frequency (if ACT permits);
- Check identifiability;
- Check if model is misspecified.



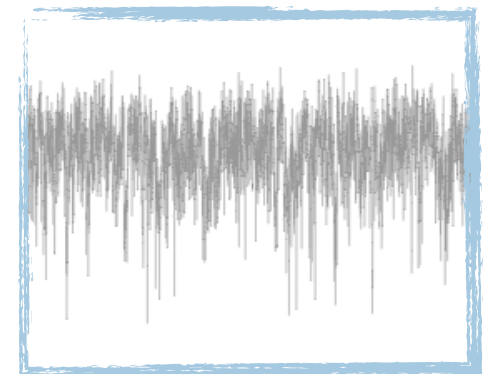
Possible scenarios

Beast analysis

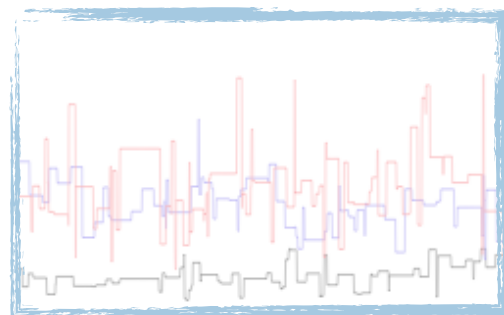
Initialisation
failed

Start likelihood: -Infinity after 1000
initialisation attempts
Fatal exception: Could not find a
proper state to initialise. Perhaps try
another seed.
P(posterior) = -Infinity (was -Infinity)

Everything
mixed



Nothing
mixed



One parameter
did not mix

kappa.noncoding	13.143	191
kappa.1stpos	6.28	376
kappa.2ndpos	8.643	372
kappa.3rdpos	27.988	92
mutationRate.noncodi...	0.347	273
mutationRate.1stpos	0.459	238
mutationRate.2ndpos	0.185	237

One parameter not mixed

kappa.noncoding	13.143	191
kappa.1stpos	6.28	376
kappa.2ndpos	8.643	372
kappa.3rdpos	27.988	92
mutationRate.noncodi...	0.347	273
mutationRate.1stpos	0.459	238
mutationRate.2ndpos	0.185	237

Possible solutions:

- Tweak the operator weights:
 - Increase weight for low ESS parameters;

One parameter not mixed

kappa.noncoding	13.143	191
kappa.1stpos	6.28	376
kappa.2ndpos	8.643	372
kappa.3rdpos	27.988	92
mutationRate.noncodi...	0.347	273
mutationRate.1stpos	0.459	238
mutationRate.2ndpos	0.185	237

Possible solutions:

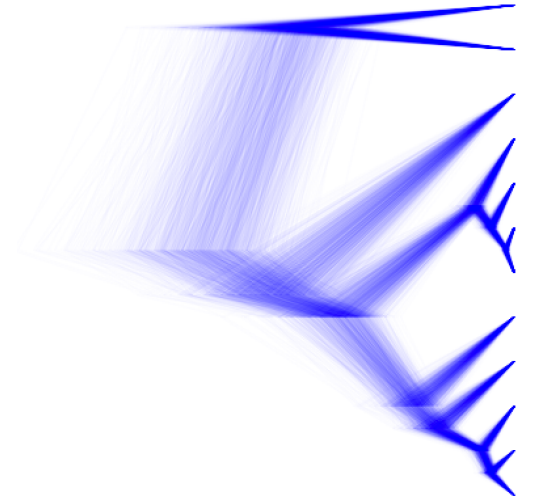
- Tweak the operator weights:
 - Increase weight for low ESS parameters;
 - Use updown operator for correlated parameters;
- Run longer (or combine several independent chains).

Only posterior not mixed

Possible reasons:

- Tree prior cares too much about a parameter the data says nothing about;
- Also one can always make a statistic that doesn't mix.

Tree space mixing



Bad news:

At the moment we can not directly examine the ESS of the trees;

Good news:

Good mixing of the continuous parameters and likelihoods is indicative;

Can also use AWTY to see if the clade probabilities stabilised.

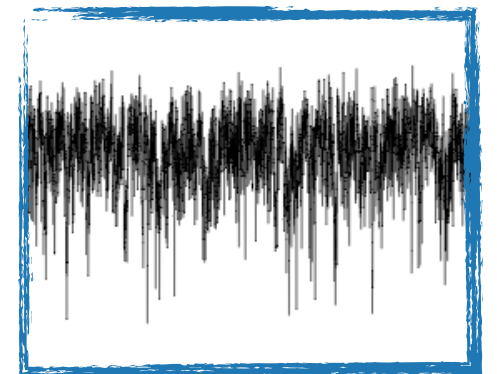
Possible scenarios

Beast analysis

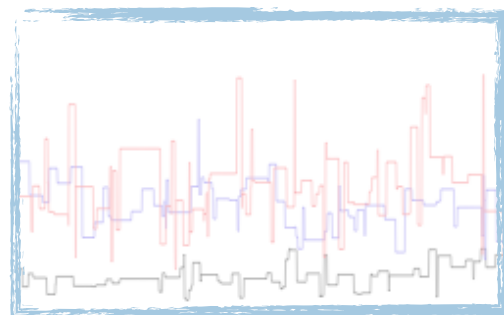
Initialisation
failed

Start likelihood: -Infinity after 1000
initialisation attempts
Fatal exception: Could not find a
proper state to initialise. Perhaps try
another seed.
P(posterior) = -Infinity (was -Infinity)

Everything
mixed



Nothing
mixed



One parameter
did not mix

kappa.noncoding	13.143	191
kappa.1stpos	6.28	376
kappa.2ndpos	8.643	372
kappa.3rdpos	27.988	92
mutationRate.noncodi...	0.347	273
mutationRate.1stpos	0.459	238
mutationRate.2ndpos	0.185	237

Everything mixed

Sanity check:

- Sampling from prior.

Good news:

Bayesian analysis always gives an answer!

Bad news:

The answer is how uncertain we are.

Troubleshooting time!



Roarrr!..