

Considerations when choosing priors

Nicola F. Müller

e-mail: Nicola.mueller@ucsf.edu

Priors denote our prior belief/expectations in certain parameter values

Posterior probability

Tree likelihood

Tree Prior describes how a tree was formed by a population model

Prior belief of what the parameters should be

$$P(E \text{ } \bullet \text{ } O \text{ } | \text{ } ACAC... \text{ } TCAC... \text{ } ACAG...) = \frac{P(ACAC... \text{ } TCAC... \text{ } ACAG... | E \text{ } \bullet \text{ } O) P(E | \bullet \text{ } O) P(\bullet \text{ } O) P(O)}{P(ACAC... \text{ } TCAC... \text{ } ACAG...)}$$

Practical example: Choosing priors on the evolutionary rates for SARS-CoV-2

- What are related viruses (preferably in the same host)
 - SARS
 - MERS
- What are their evolutionary rates

SARS evolutionary rate

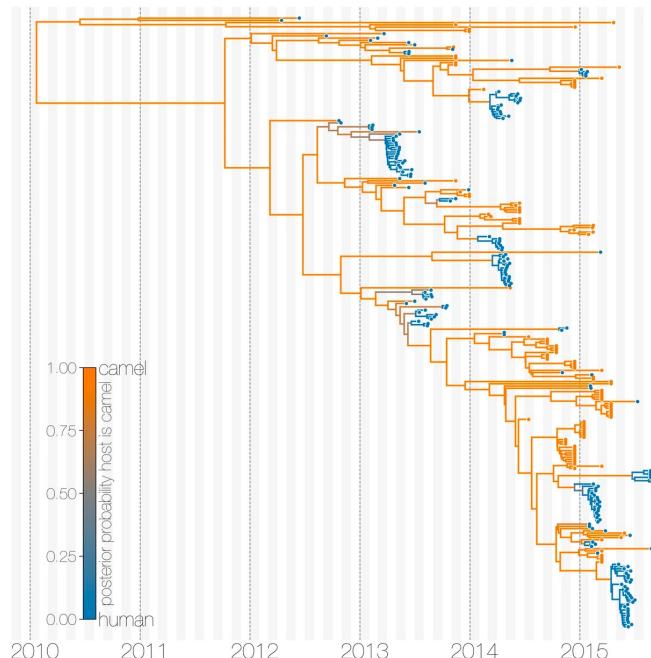
Table 2: Substitution rates ($\times 10^{-3}$ per site per year) and Ka/Ks ratio in the coding regions.

	Non-synonymous sites	Synonymous sites	Ka/Ks
Total	1.16–3.30	1.67–4.67	0.70
Nonstructural region	0.81–2.40	1.78–5.07	0.46
Structural region	2.03–5.53	1.40–3.69	1.47
Nsp1	1.05–3.13	0.85–2.60	1.22
S	1.11–3.02	3.22–8.50	0.35
EMN	3.35–9.22	0	» 1

The same divergence time as in Table 1 was used. Nonstructural region denotes the 5' two-thirds of the coding regions (sites 265 – 21485) and structural region denotes the 3' one-third of the coding regions (21492 – 29388). EMN denotes three genes E, M, and N.

MERS evolutionary rate

Bayesian structured coalescent estimate from primary analysis 9.57×10^{-4} (95% HPDs: $8.28 - 10.9 \times 10^{-4}$) subs/site/year.



SARS-CoV-2 has an evolutionary rate similar to MERS-CoV-2

- Duchene et al. (2019), Virus Evol. estimate the average rate to be 1.1×10^{-3} subs/site/year

Going wrong with the prior

- NL63: 1.3×10^{-4} (CI $1.1 - 1.5 \times 10^{-4}$)
- 229E: 2.5×10^{-4} (CI $2.2 - 2.7 \times 10^{-4}$)
- OC43: 2.1×10^{-4} (CI $1.9 - 2.3 \times 10^{-4}$)

Posterior is the product of Prior and Likelihood

Posterior probability

Tree likelihood

Tree Prior describes how a tree was formed by a population model

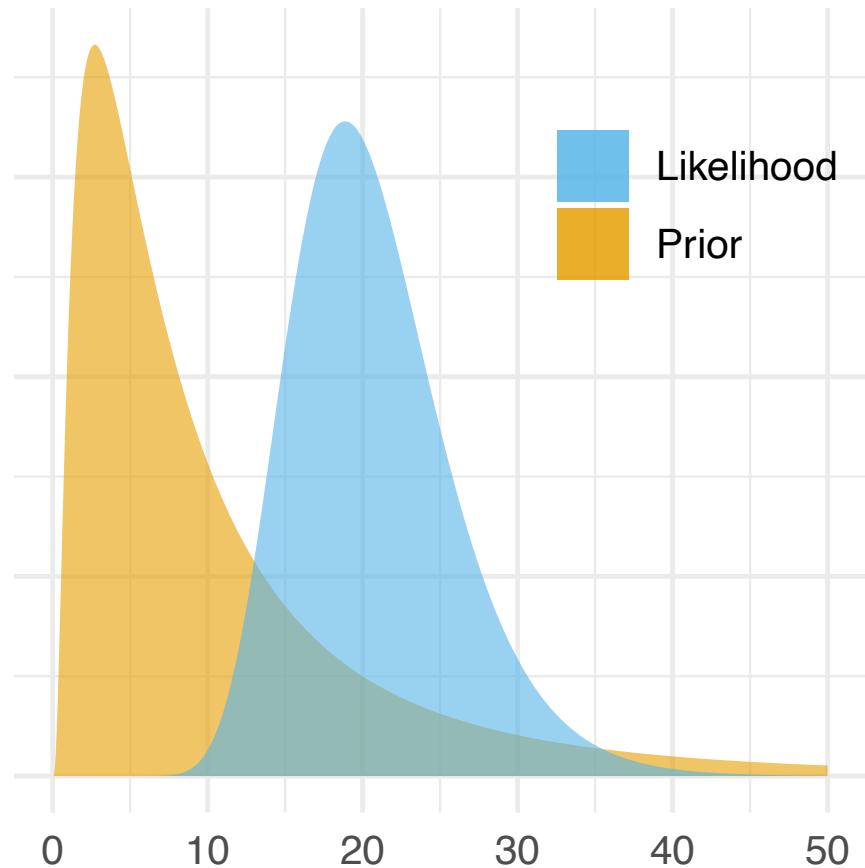
Prior belief of what the parameters should be

$$P(E \text{ } \square \text{ } \circ \text{ } \square \text{ } | \text{ } ACAC \dots) =$$

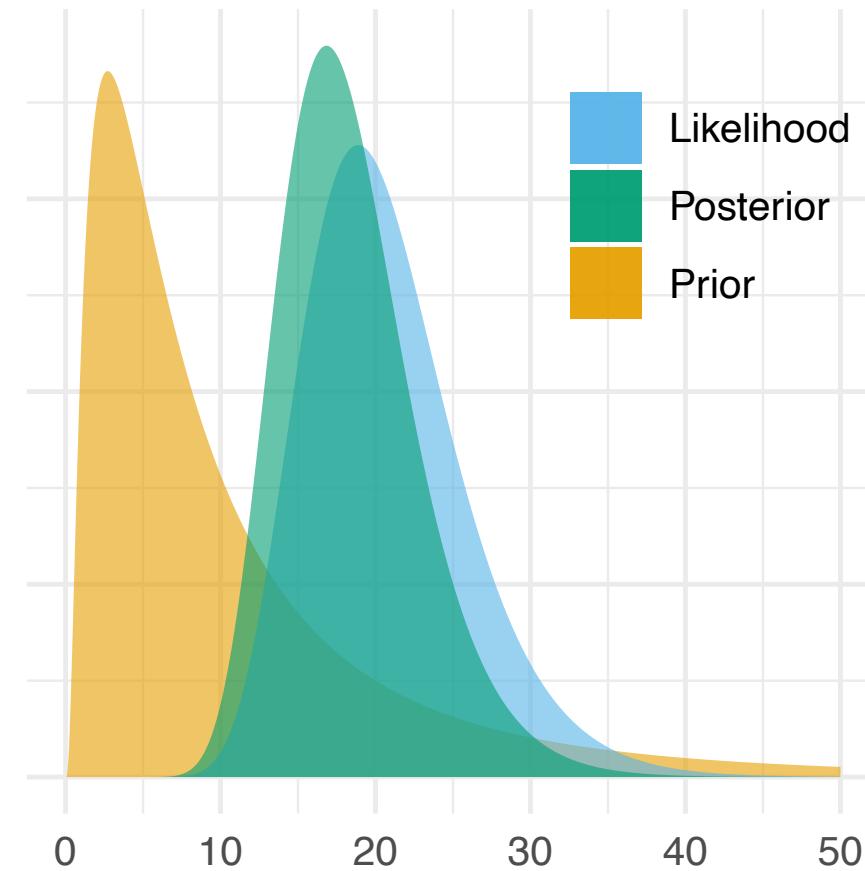
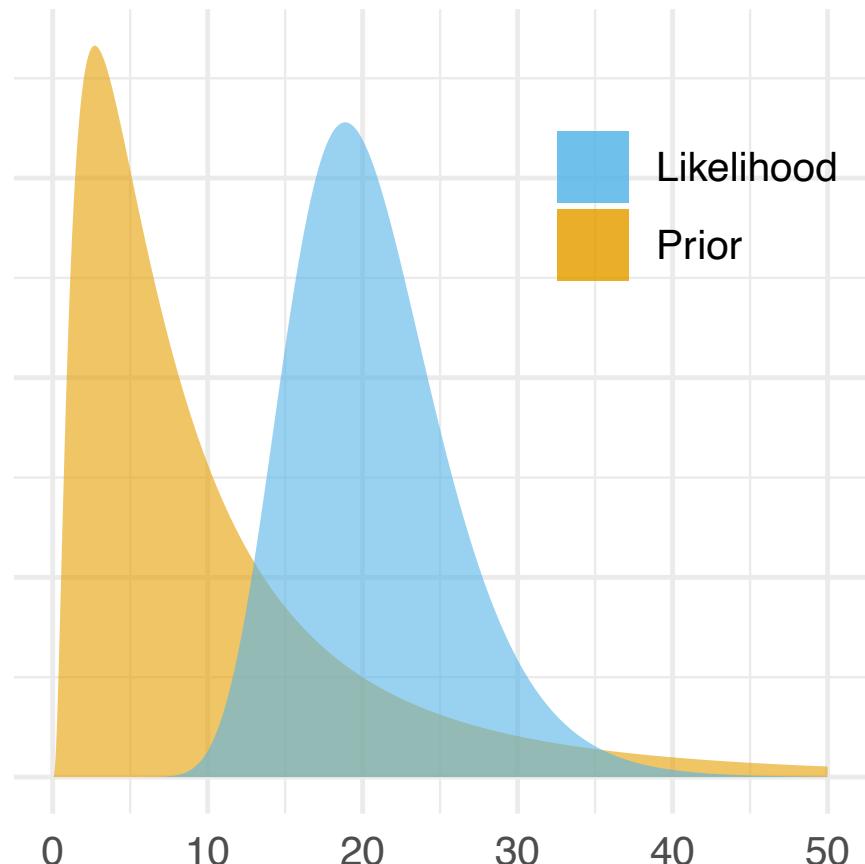
$$P(ACAC \dots | E \text{ } \square \text{ } \circ \text{ } \square) P(E | \square \text{ } \circ) P(\square \text{ } \circ) P(\square \text{ } \circ) P(\square)$$

$P(ACAC \dots | \square \text{ } \circ \text{ } \square)$

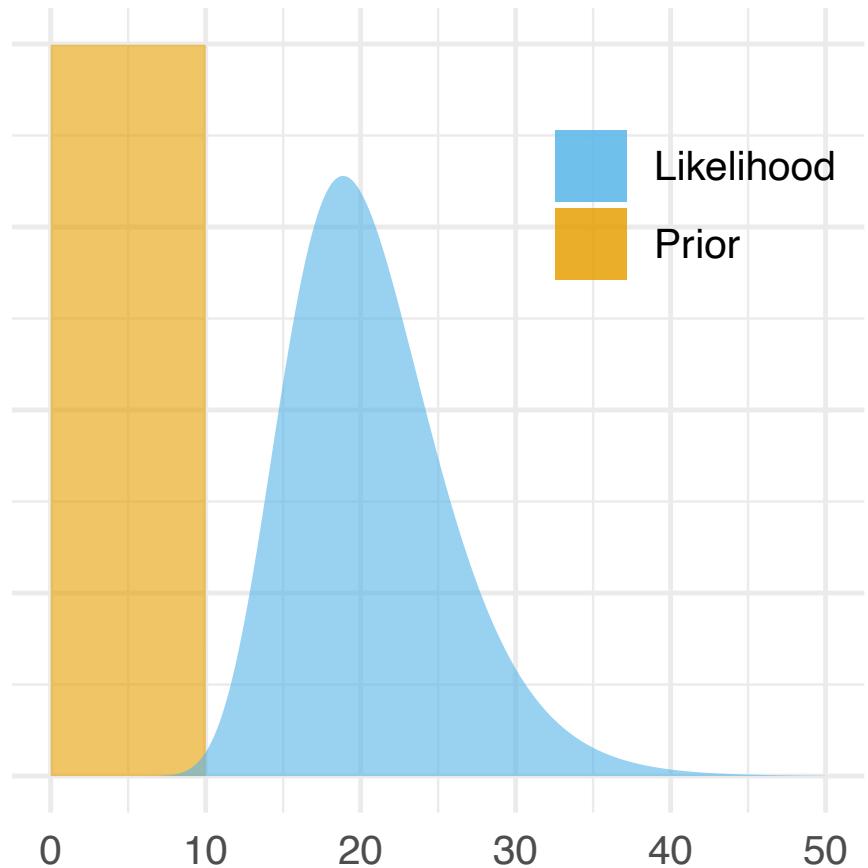
Posterior is the product of Prior and Likelihood



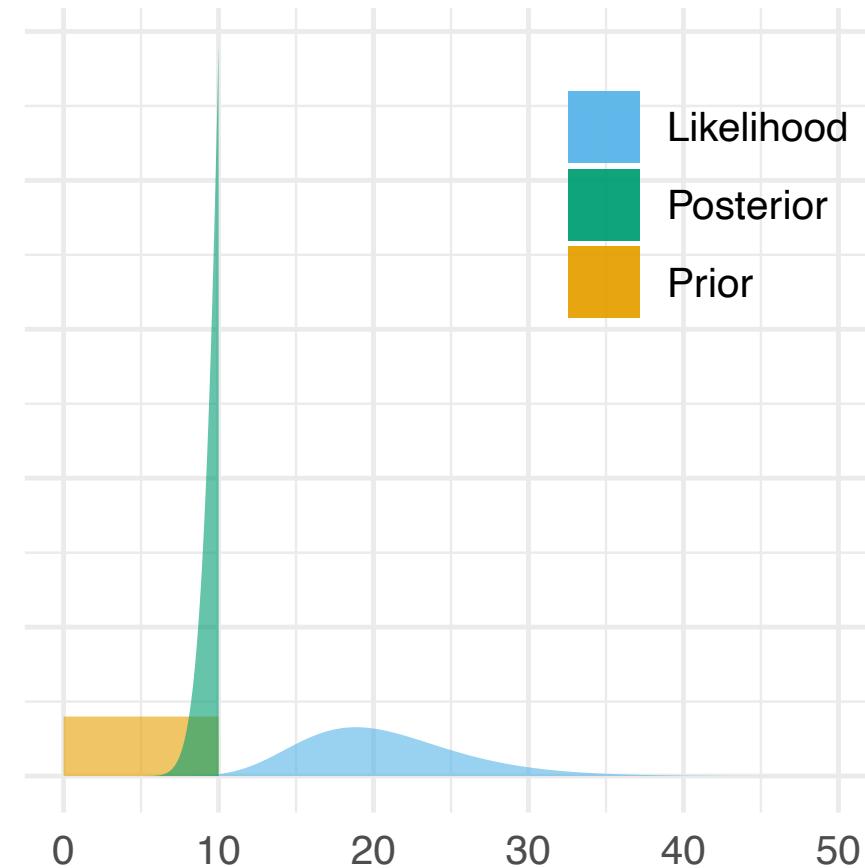
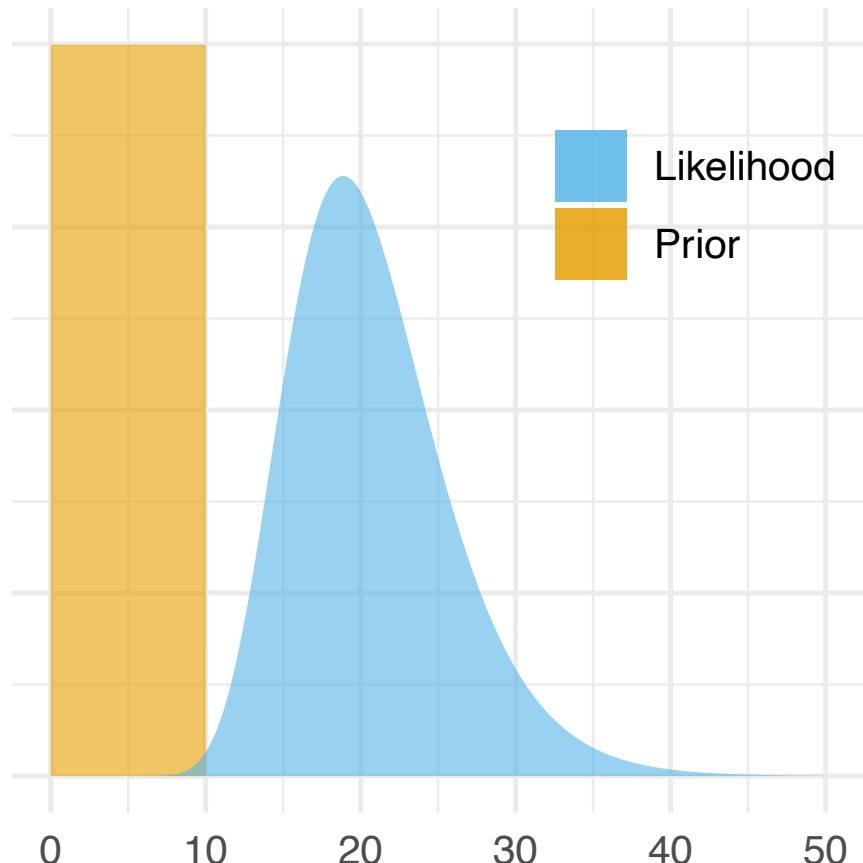
How do the Prior and Likelihood act together



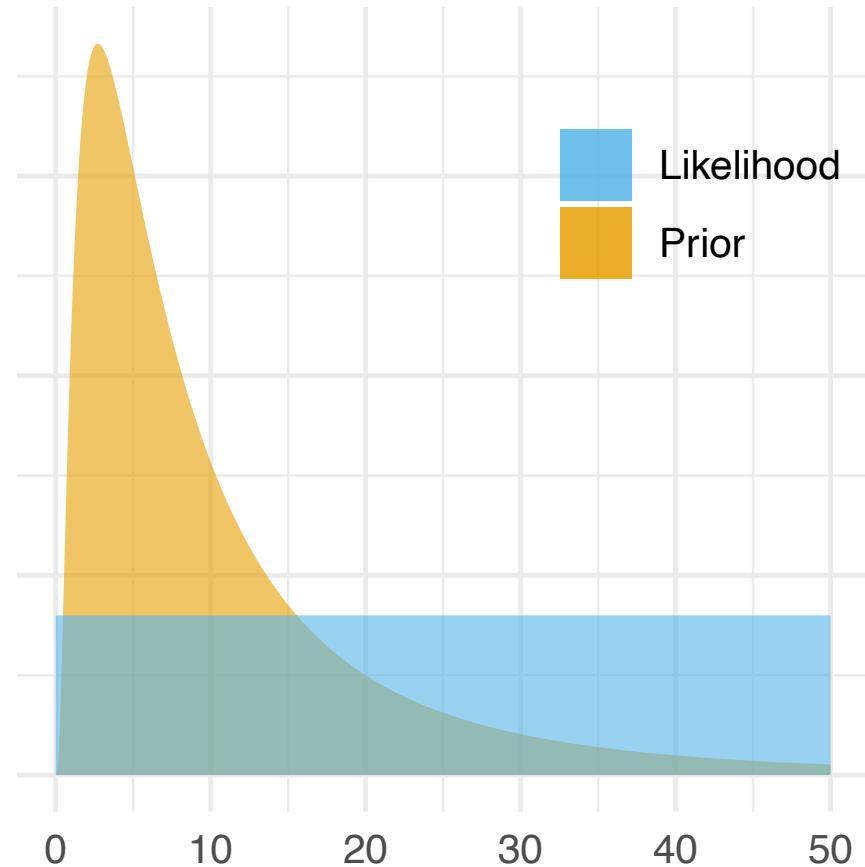
If the prior largely excludes the most likely values of the Likelihood



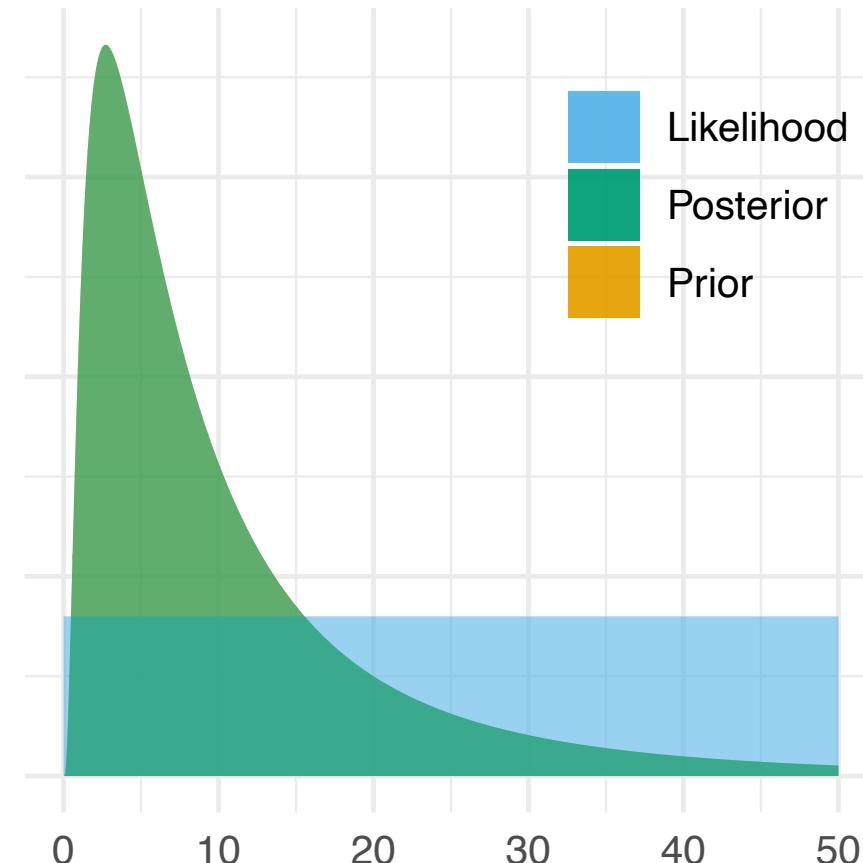
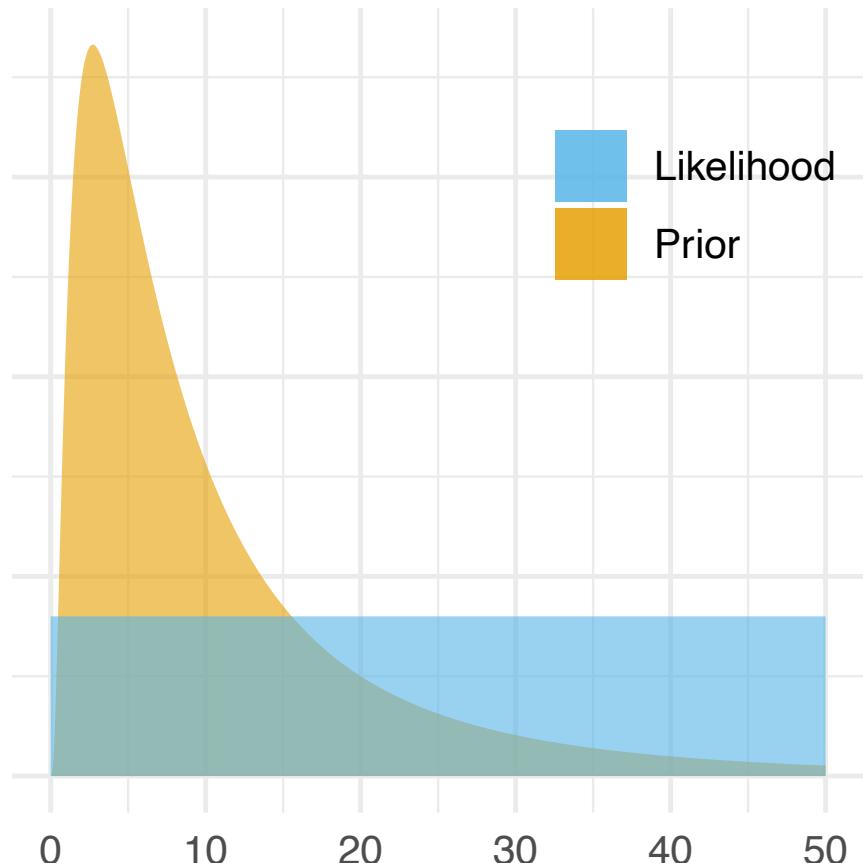
The so does the Posterior



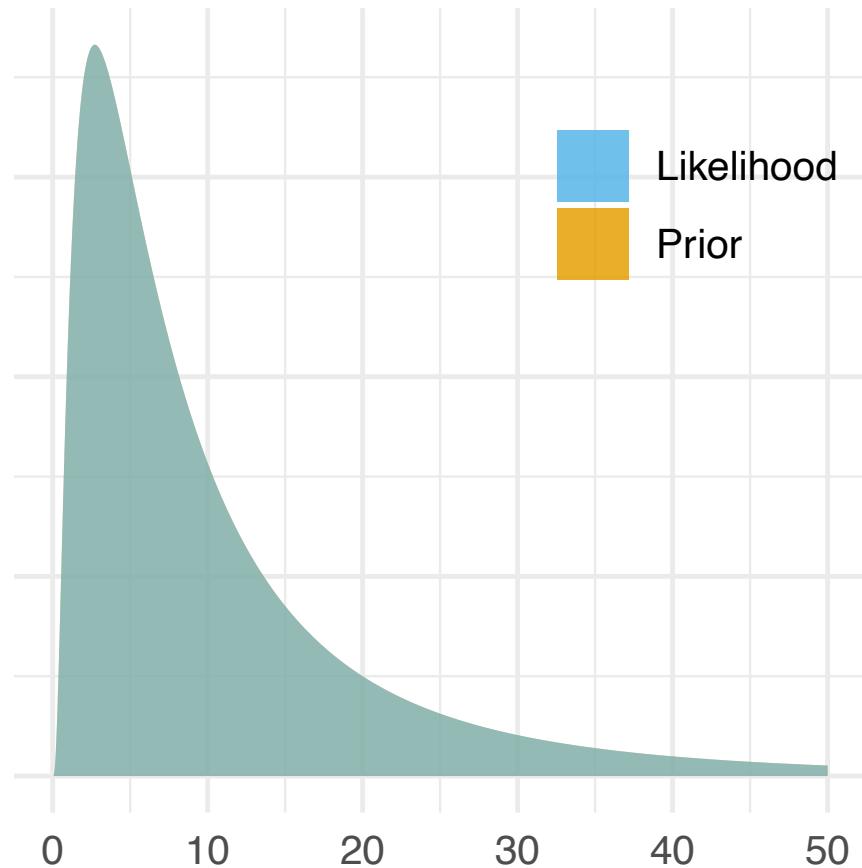
If the Likelihood provides no information



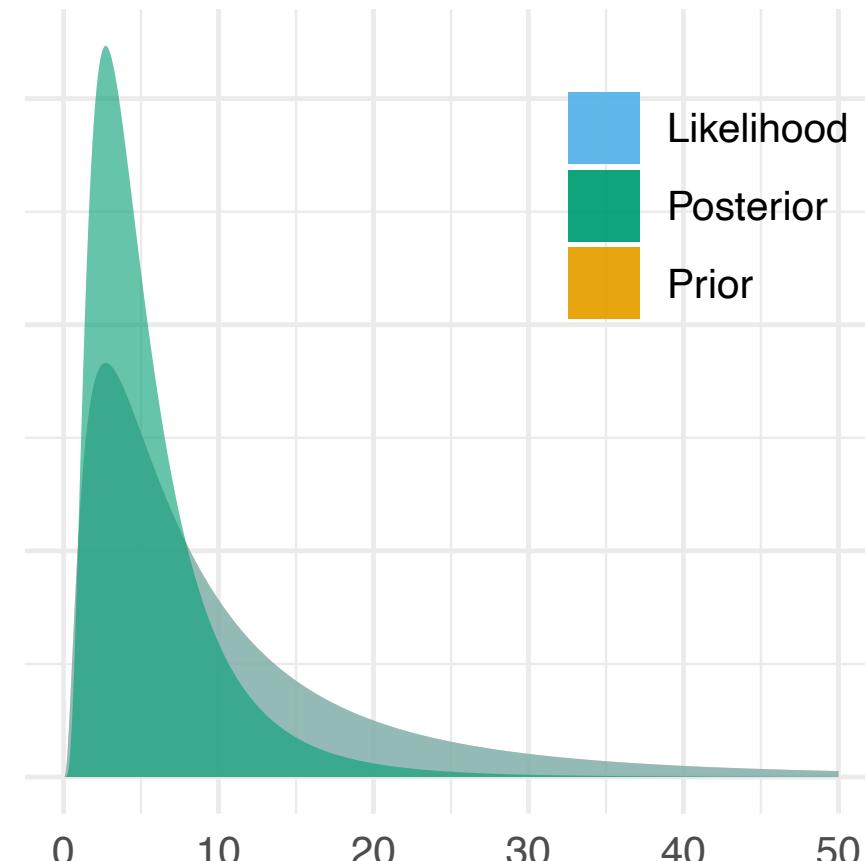
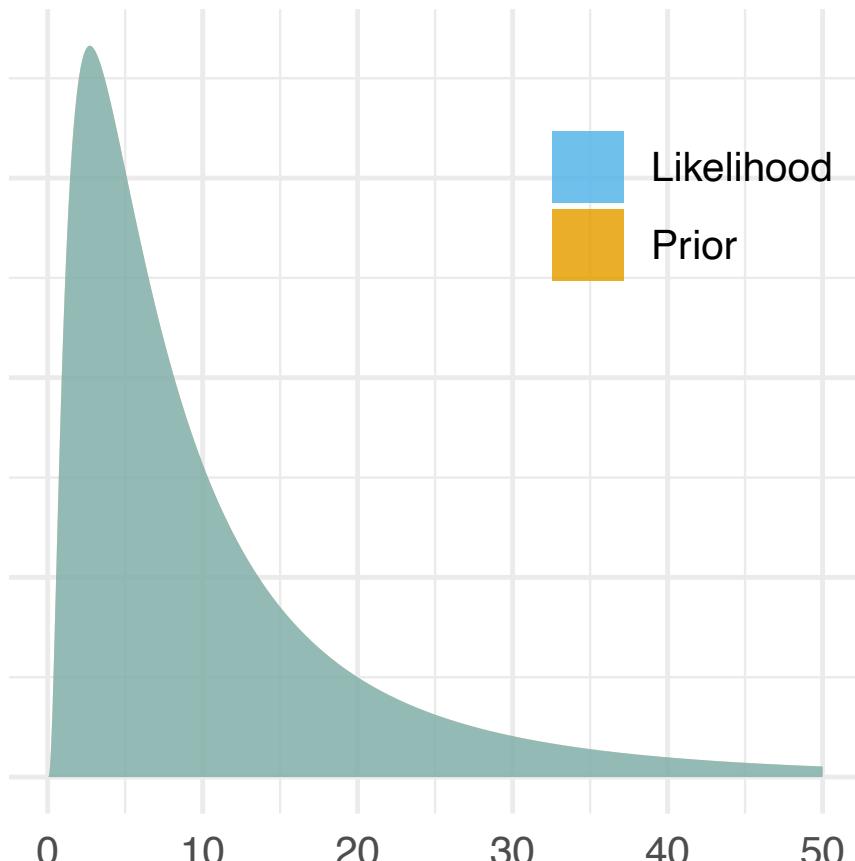
Then the Posterior will be equal to the prior



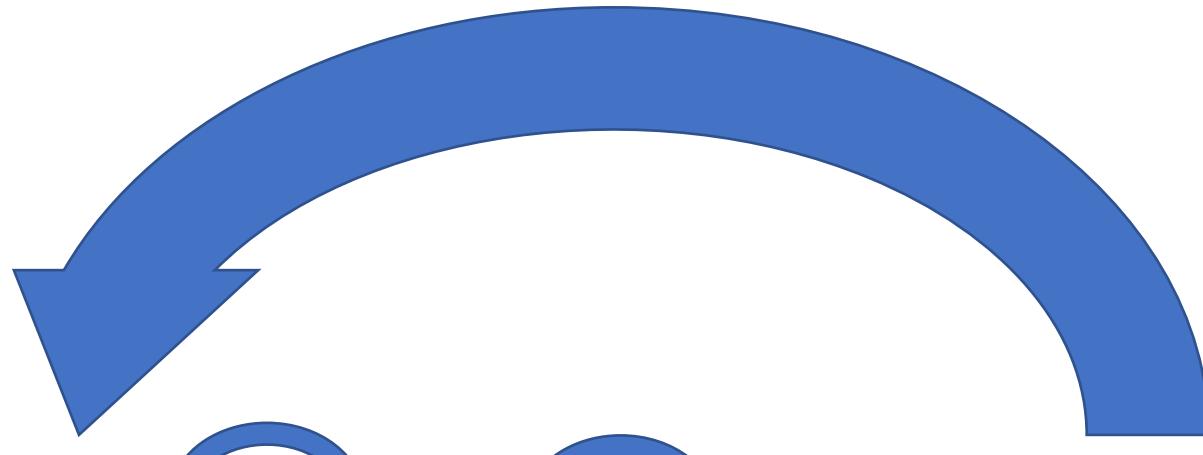
If the prior is equal to the likelihood



Then the posterior will be more
"peaked/narrower" than either Prior or
Likelihood



Everything affects everything.


$$P(E \text{ } \boxed{\text{ }} \text{ } \boxed{\text{ }} \text{ } | \text{ } \begin{matrix} ACAC... \\ TCAC... \\ ACAG... \end{matrix}) = \frac{P(\begin{matrix} ACAC... \\ TCAC... \\ ACAG... \end{matrix} | E \text{ } \boxed{\text{ }} \text{ } \boxed{\text{ }} \text{ } | \text{ } \boxed{\text{ }}) P(E | \boxed{\text{ }} \text{ } \boxed{\text{ }}) P(\boxed{\text{ }} \text{ } \boxed{\text{ }}) P(\boxed{\text{ }} \text{ } \boxed{\text{ }} \text{ } | \text{ } \boxed{\text{ }}) P(\boxed{\text{ }} | \text{ } \boxed{\text{ }})}$$

The equation illustrates the joint probability of a sequence of events (E, boxed symbols, and a sequence of three symbols) given a sequence of hidden states (ACAC, TCAC, ACAG). The probability is calculated as the product of the conditional probabilities of each event given the previous state and the hidden state, divided by the probability of the hidden state sequence.

True prior distributions are revealed by sampling under the prior.

$$P(\text{E} \text{ } \text{ } \text{ } \text{ } | \text{X}) = \frac{P(\text{X} | \text{E} \text{ } \text{ } \text{ } \text{ }) \text{ } P(\text{E} | \text{o}) \text{ } P(\text{o}) \text{ } P(\text{E} | \text{o}) \text{ } P(\text{o})}{P(\text{X})}$$

MCMC Convergence and troubleshooting

Failed initialization

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.

Failed initialization

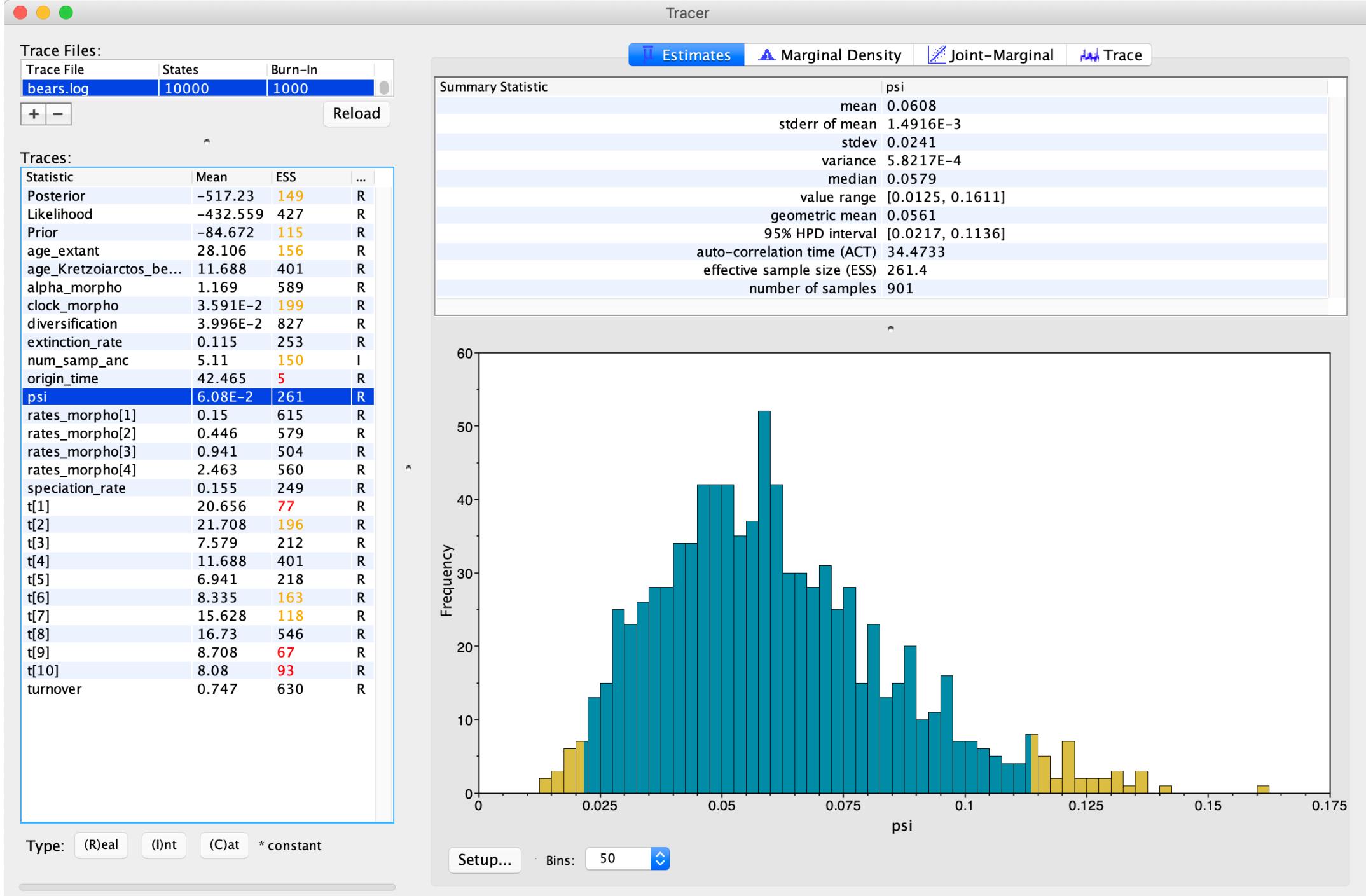
- Check initial values
- Check for incompatible priors
- Increase number of initialization attempts
- Talk to the developers / BEAST2 support group

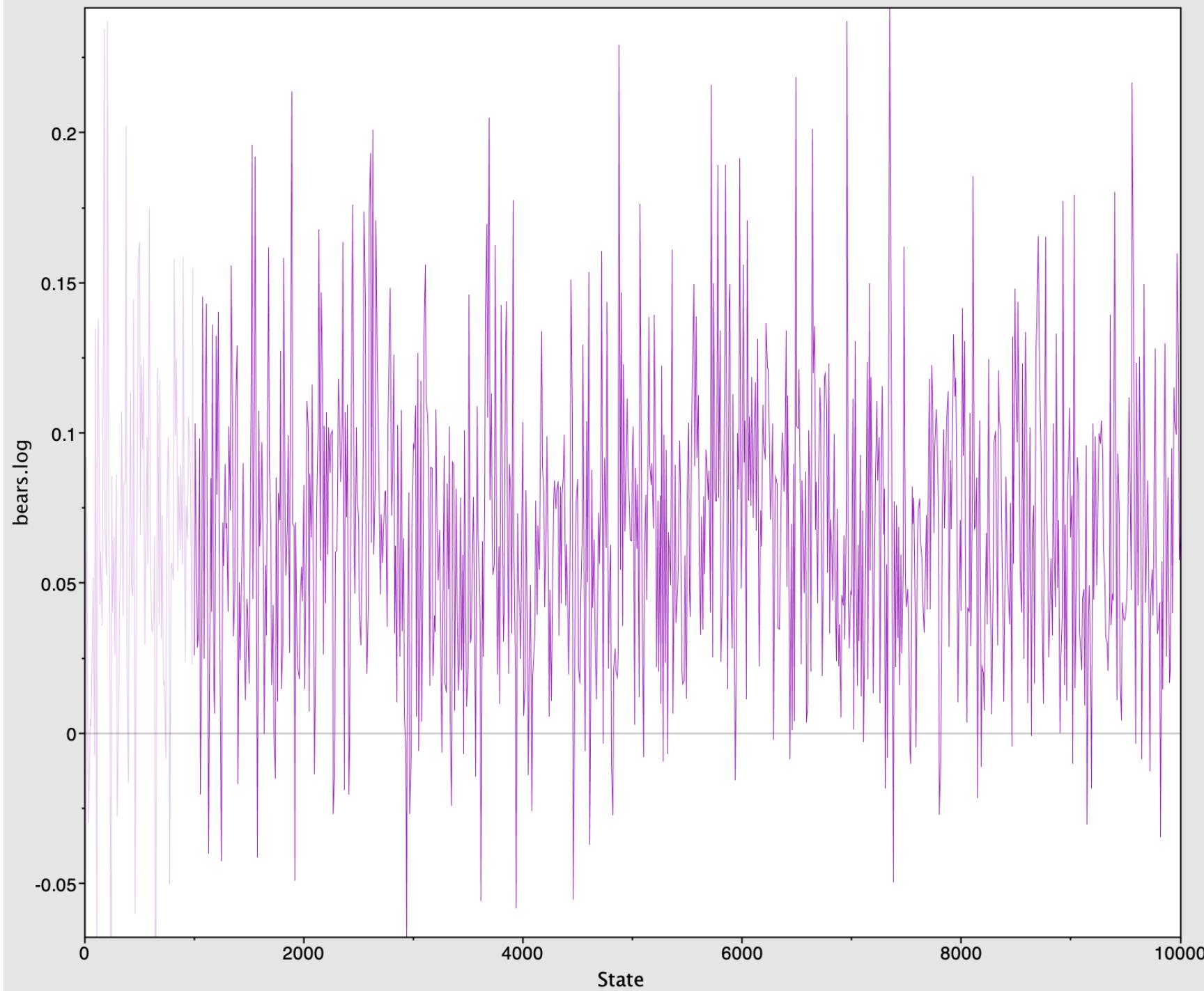
Convergence

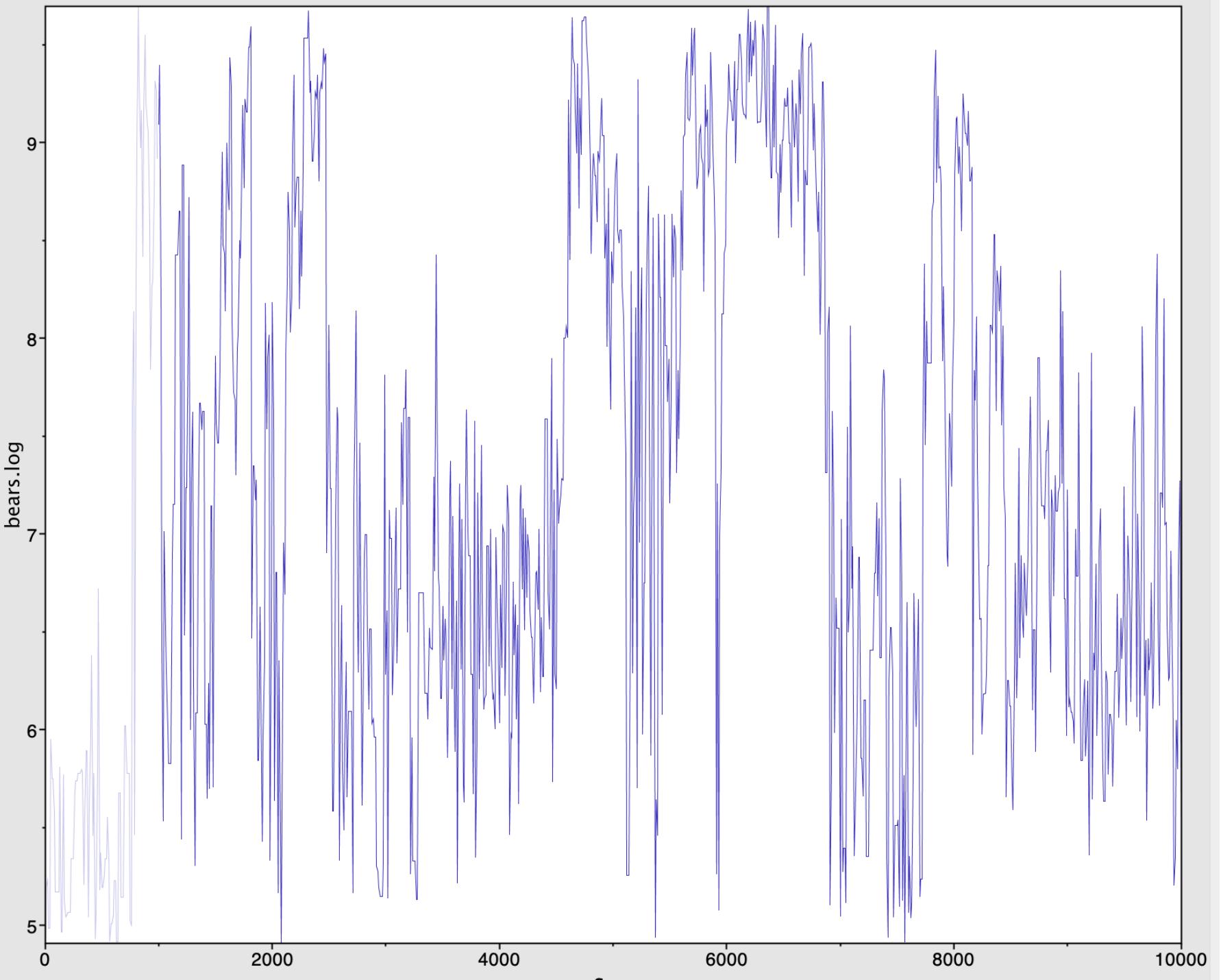
MCMC theory guarantees convergence to the posterior distribution

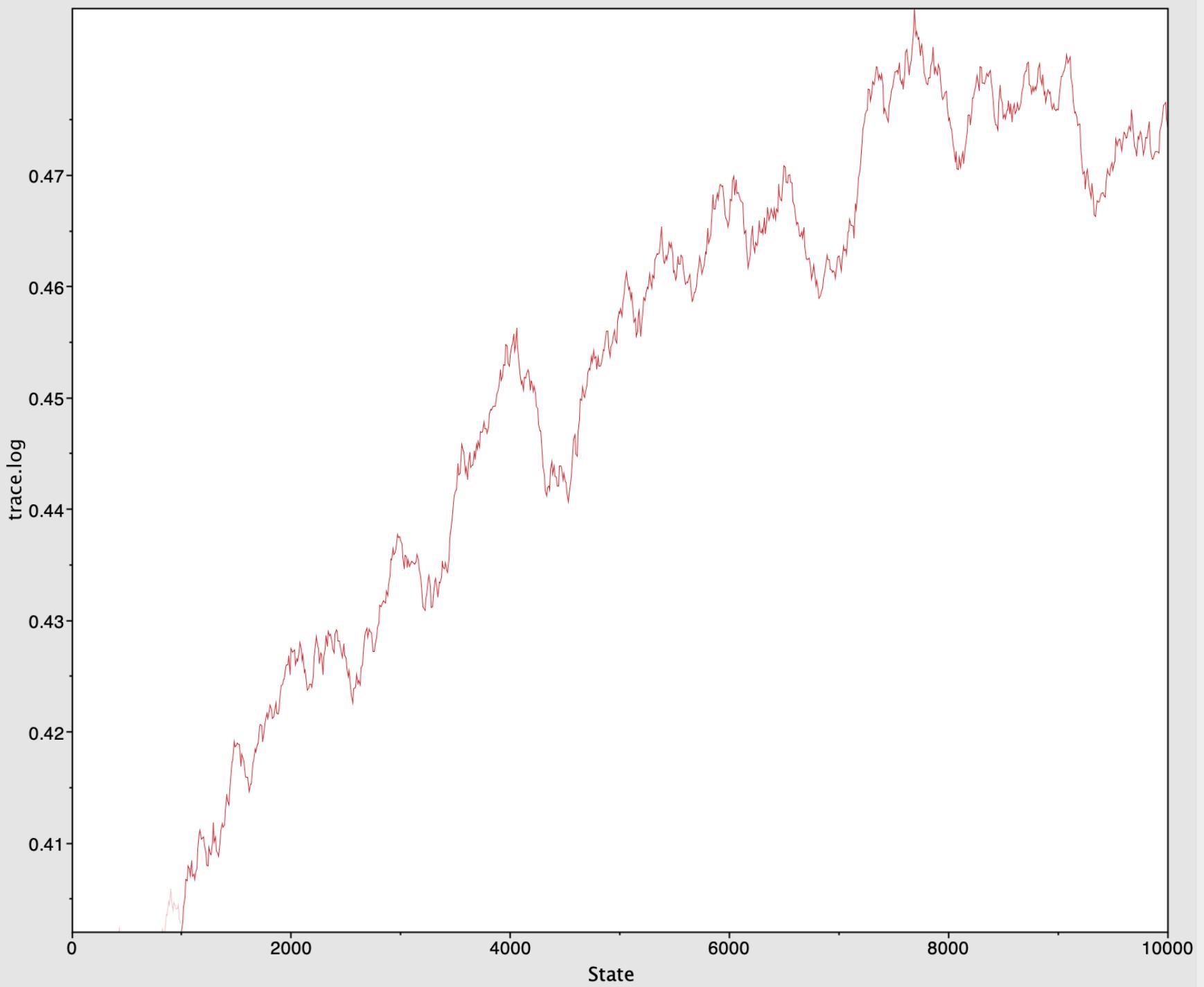
=> but not when !

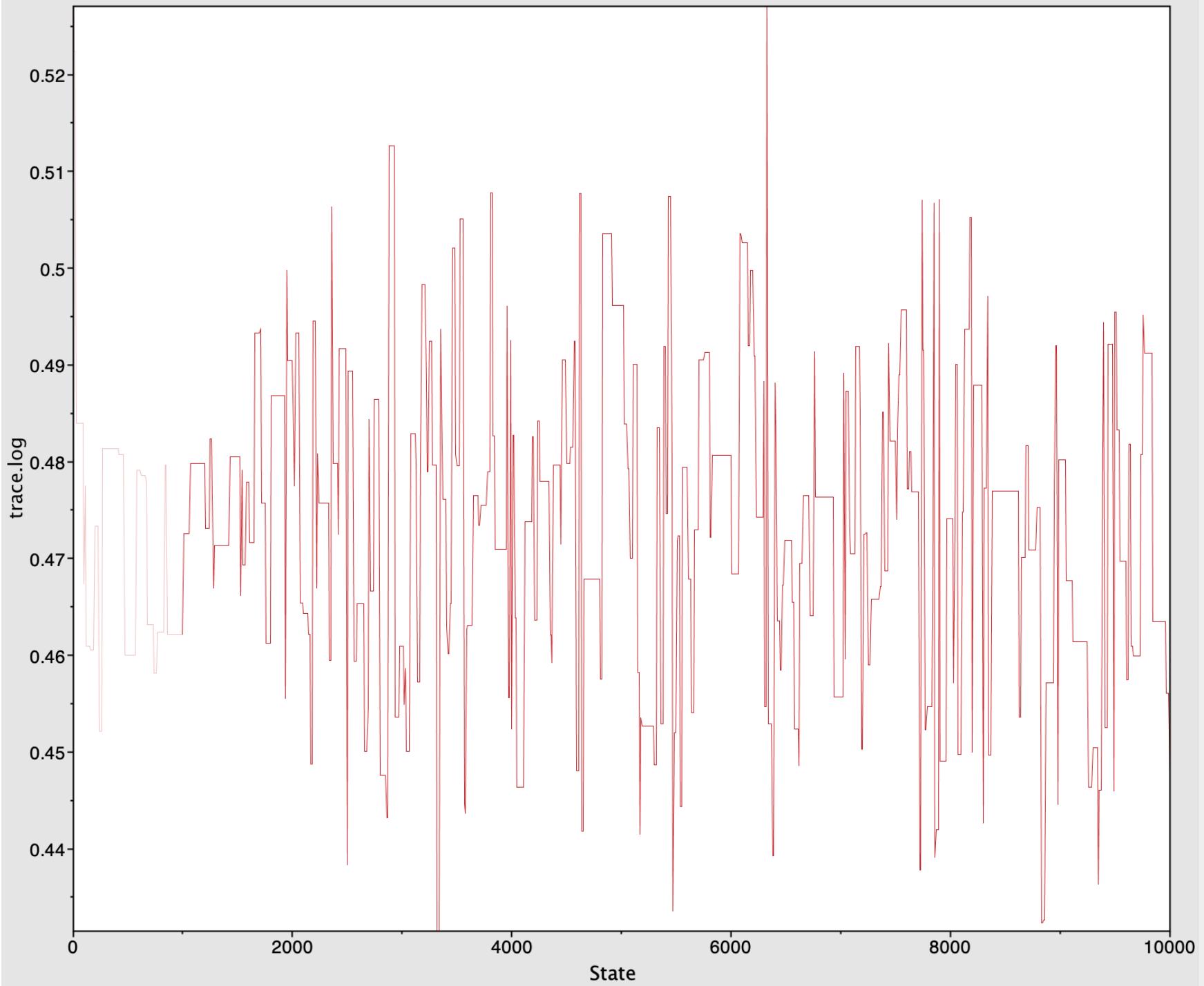
=> how do we know when our inference has converged ?











What can you do ?

- Increase sampling frequency
- Adjust operator weights
- Increase chain length
- Check for model misspecification
- Check for prior incompatibility
- Simplify the analysis setup
- Run heated chains (MCMCMC)
- Talk to the developers / BEAST2 support group

Adjusting operator weights

Statistic	Mean	ESS	...
Posterior	-517.23	149	R
Likelihood	-432.559	427	R
Prior	-84.672	115	R
age_extant	28.106	156	R
age_Kretzoiarctos_be...	11.688	401	R
alpha_morpho	1.169	589	R
clock_morpho	3.591E-2	199	R
diversification	3.996E-2	827	R
extinction_rate	0.115	253	R
num_samp_anc	5.11	150	I
origin_time	42.465	5	R
psi	6.08E-2	261	R
rates_morpho[1]	0.15	615	R
rates_morpho[2]	0.446	579	R
rates_morpho[3]	0.941	504	R
rates_morpho[4]	2.463	560	R
speciation_rate	0.155	249	R
t[1]	20.656	77	R
t[2]	21.708	196	R
t[3]	7.579	212	R
t[4]	11.688	401	R
t[5]	6.941	218	R
t[6]	8.335	163	R
t[7]	15.628	118	R
t[8]	16.73	546	R
t[9]	8.708	67	R
t[10]	8.08	93	R
turnover	0.747	630	R

▶ Scale: birthRate.t:primate	Scale birth rate of Yule prior of tree t:noncoding	3.0
▶ Scale: Tree.t:primate		3.0
▶ Scale: Tree.t:primate		3.0
▶ Uniform: Tree.t:primate		30.0
▶ Subtree Slide: Tree.t:primate		15.0
▶ Exchange: Tree.t:primate		15.0
▶ Exchange: Tree.t:primate		3.0
▶ Wilson Balding: Tree.t:primate		3.0
▶ Scale: rateAC.s:primate	Scale GTR A-C substitution parameter of partition s:noncoding	0.1
▶ Scale: rateAG.s:primate	Scale GTR A-G substitution parameter of partition s:noncoding	0.1
▶ Scale: rateAT.s:primate	Scale GTR A-T substitution parameter of partition s:noncoding	0.1
▶ Scale: rateCG.s:primate	Scale GTR C-G substitution parameter of partition s:noncoding	0.1
▶ Scale: rateGT.s:primate	Scale GTR G-T substitution parameter of partition s:noncoding	0.1
▶ Scale: Tree.t:primate	Scales all internal nodes for tree t:noncoding	3.0
▶ Scale: Tree.t:primate	Scales root node for tree t:noncoding	3.0
▶ Uniform: Tree.t:primate	Draws new internal node heights uniformly for tree t:noncoding	30.0
▶ Subtree Slide: Tree.t:primate	Performs subtree slide rearrangement of tree t:noncoding	15.0
▶ Exchange: Tree.t:primate	Narrow exchange performs local rearrangement of tree t:noncoding	15.0
▶ Exchange: Tree.t:primate	Wide exchange performs global rearrangement of tree t:noncoding	3.0
▶ Wilson Balding: Tree.t:primate	Performs Wilson-Balding global rearrangement of tree t:noncoding	3.0
▶ Delta Exchange: freqParameter.s:primate	Exchange values of frequencies of partition s:noncoding	0.1

Default Operator Schedule

Transform Auto Optimize Detailed Rejection

Auto Optimize Delay

10000