



# Taming the BEAST

Bayesian Evolutionary Analyses by Sampling Trees

Louis du Plessis

1. BEAST2 model building blocks
2. How to find a posterior
3. What are priors and why do they matter?
4. The anatomy of a BEAST2 analysis
5. Tutorial introduction and demo

## data

- Samples drawn from a realisation of some **stochastic process**
- Typically an alignment of DNA or RNA sequences

## model

- **Model** is a mathematical description of the **process** that generated the data
- Each model has a number of **model parameters**
- Parameters are **random variables**

## hypothesis

- **Hypothesis** is some statement about the model parameters
- Hypothesis assigns **values** to the model parameters
- Data used to decide if hypothesis is reasonable or not
- May only concern some model parameters. The rest are **nuisance parameters**

# Bayesian inference

---

## Prior → $P(\text{hypothesis} \mid \text{model})$

- Have some degree of belief in our hypothesis
- From external sources, previous analyses etc.
- All parameters have priors, whether you specify them or not!

## Likelihood → $P(\text{data} \mid \text{hypothesis, model})$

- Likelihood is proportional to the probability of observing the data given a hypothesis

## Posterior → $P(\text{hypothesis} \mid \text{data, model})$

- Updated probability for the model parameters in light of the data

# Bayesian inference

---

## Prior → $P(\text{hypothesis})$

- Have some degree of belief in our hypothesis
- From external sources, previous analyses etc.
- All parameters have priors, whether you specify them or not!

## Likelihood → $P(\text{data} \mid \text{hypothesis})$

- Likelihood is proportional to the probability of observing the data given a hypothesis

## Posterior → $P(\text{hypothesis} \mid \text{data})$

- Updated probability for the model parameters in light of the data

# Bayesian inference

---

$$P(\text{hypothesis} \mid \text{data}) = \frac{P(\text{data} \mid \text{hypothesis})P(\text{hypothesis})}{P(\text{data})}$$

The diagram illustrates the components of Bayesian inference:

- Likelihood:** Represented by a red curved arrow pointing from the "Posterior" term to the "Model evidence" term.
- Prior:** Represented by a red vertical arrow pointing upwards from the "Prior" term.
- Posterior:** Represented by a red curved arrow pointing from the "Prior" term to the "Posterior" term.
- Model evidence:** Represented by a red curved arrow pointing from the "Likelihood" term to the "Model evidence" term.

$$P(\text{hypothesis} \mid \text{data}) = \frac{P(\text{data} \mid \text{hypothesis})P(\text{hypothesis})}{P(\text{data})}$$

### Prior → $P(\text{hypothesis})$

- Have some degree of belief in our hypothesis
- From external sources, previous analyses etc.
- All parameters have priors, whether you specify them or not!

### Likelihood → $P(\text{data} \mid \text{hypothesis})$

- Likelihood is proportional to the probability of observing the data given a hypothesis

### Posterior → $P(\text{hypothesis} \mid \text{data})$

- Updated probability for the model parameters in light of the data

### Model evidence → $P(\text{data}) = P(\text{data} \mid \text{model})$

- Probability for data given model (any combination of parameters)
- Not generally used for parameter inference
- Used for Bayesian model selection

# We all have one thing in common...

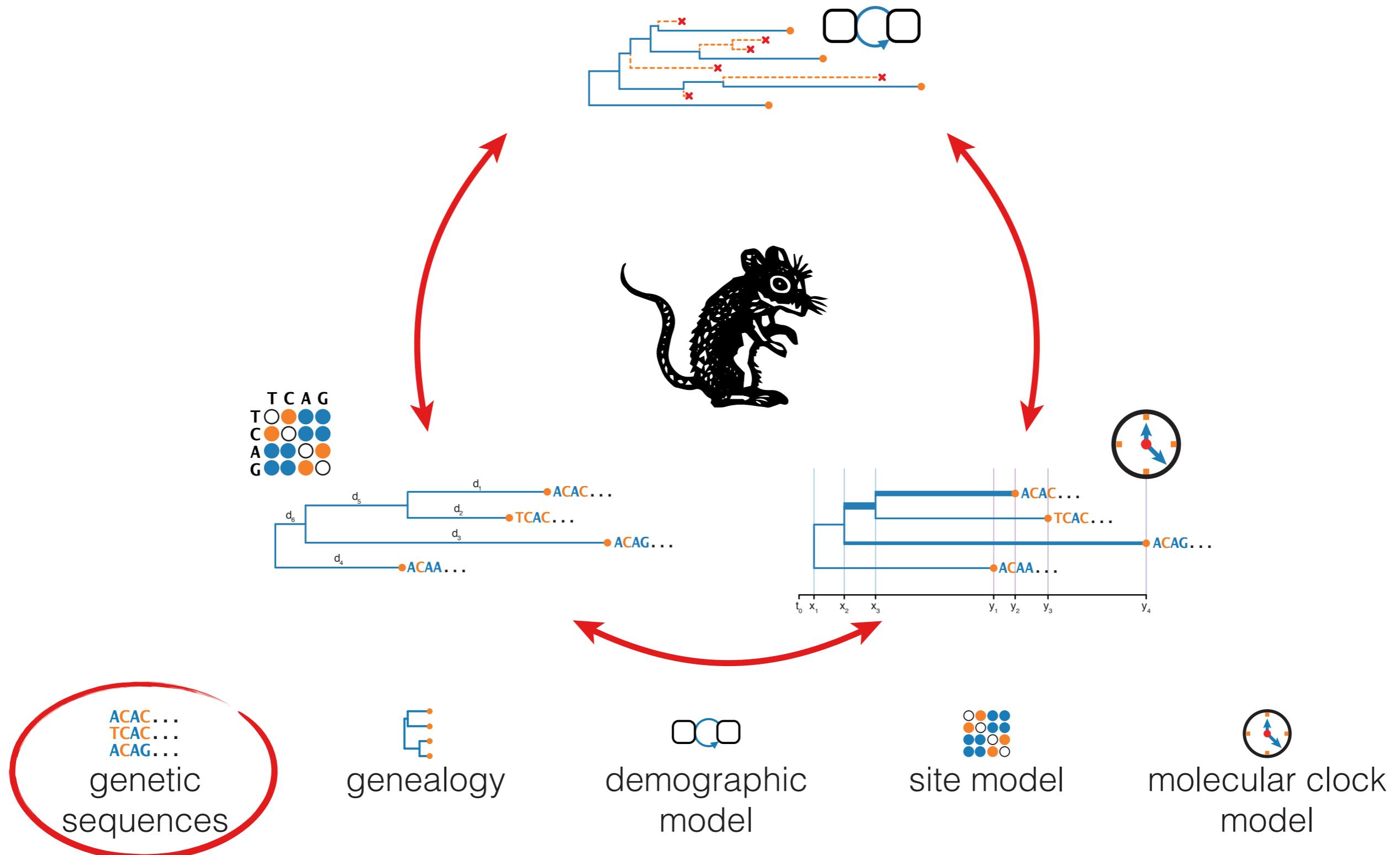
---



We all use (or want to use) **BEAST2** to answer questions about our data

**but *how?***

# What goes into a **BEAST** model?



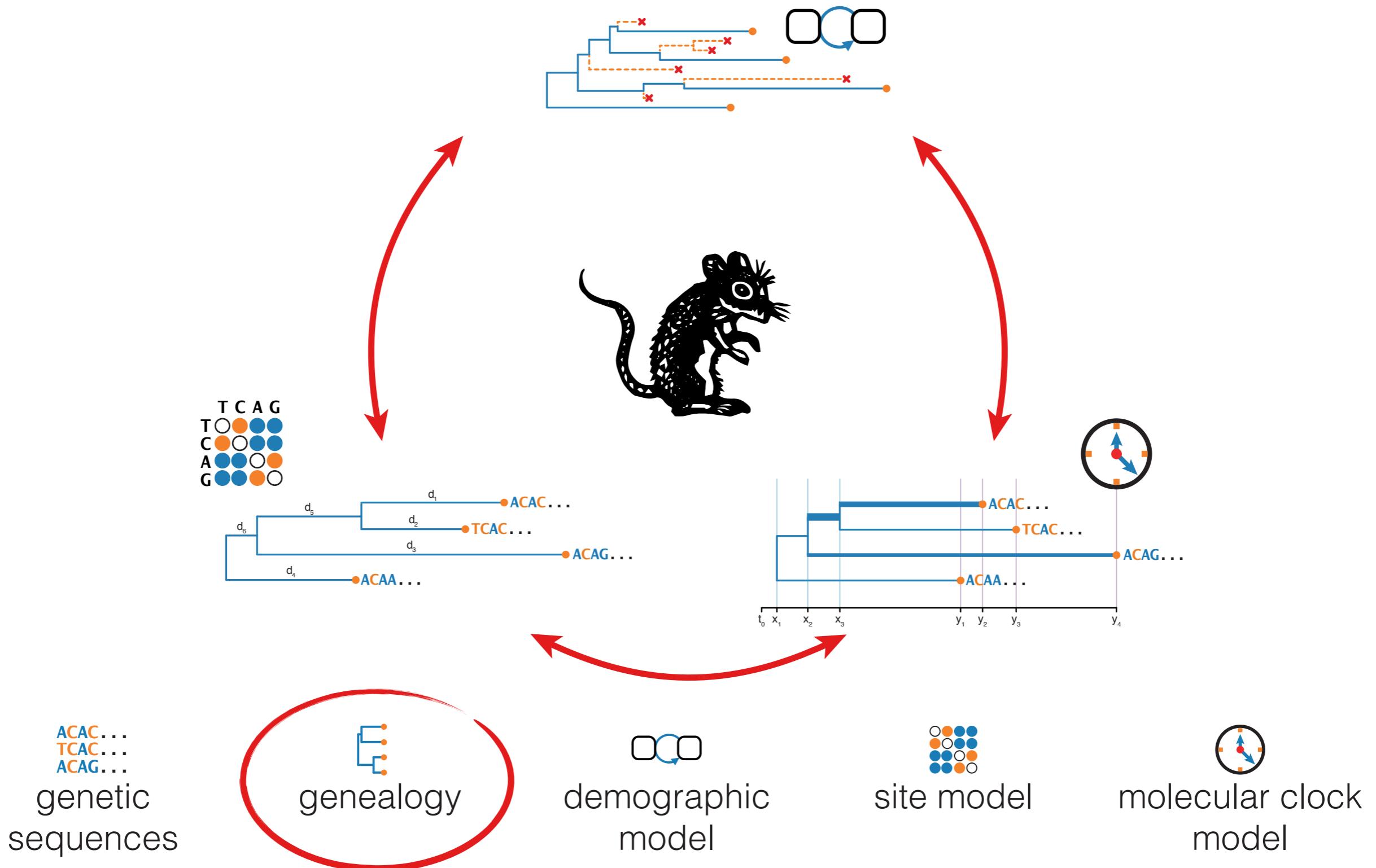
ACAC...  
TCAC...  
ACAG...

---

# The data

- Samples drawn from a realisation of some stochastic process
- **Assume that the data are correct**
- Typically one or more alignments of genetic sequencing data (DNA, RNA, amino acids, codons)
- Sampled at one or many time points
- May also contain sampling location, phenotypic trait data etc.

# What goes into a **BEAST** model?



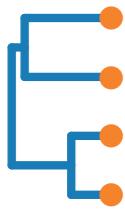
genetic  
sequences

genealogy

demographic  
model

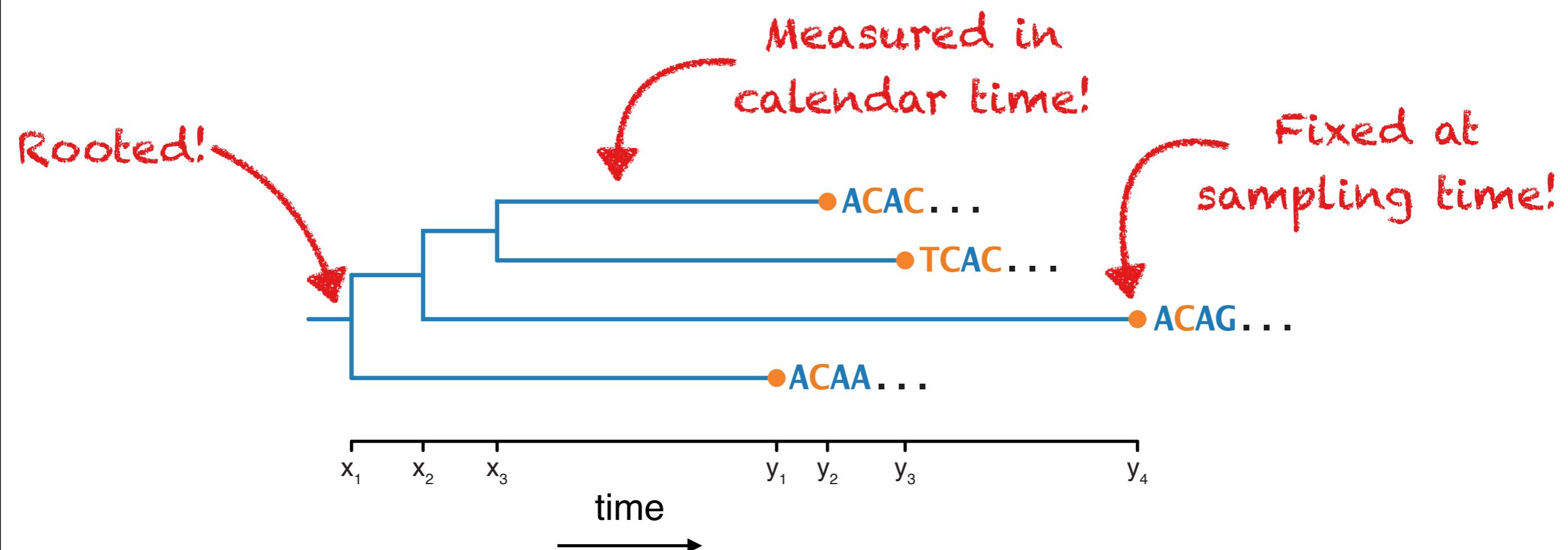
site model

molecular clock  
model

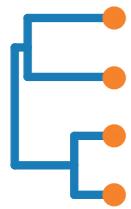


# The genealogy (tree)

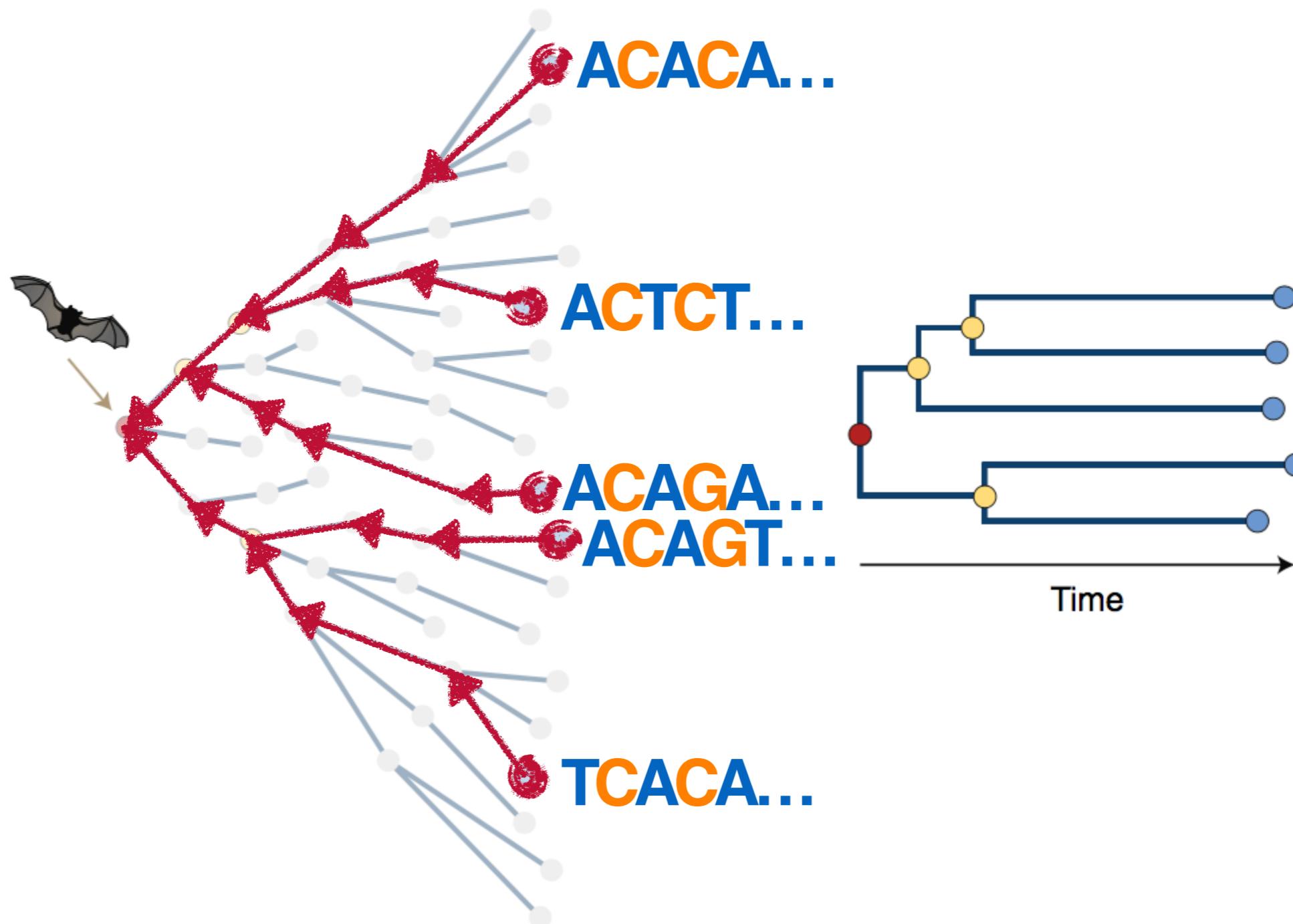
The fundamental genealogical structure  
in **BEAST2** is the **rooted time-tree**

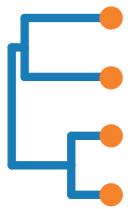


- This tree is a "sampled" or "reconstructed" tree
- Displays ancestral relationships between **sampled sequences** (individuals/taxa)



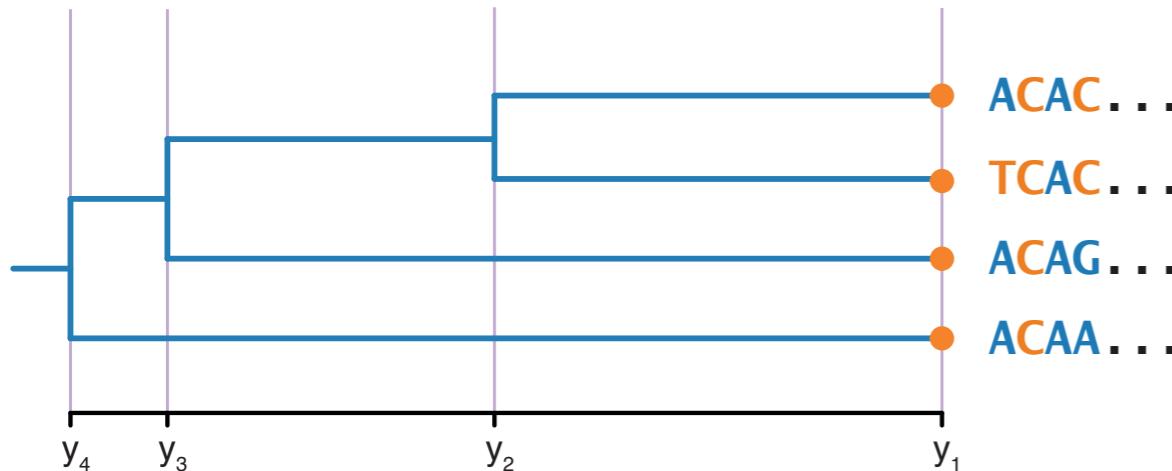
# The sampled/reconstructed tree





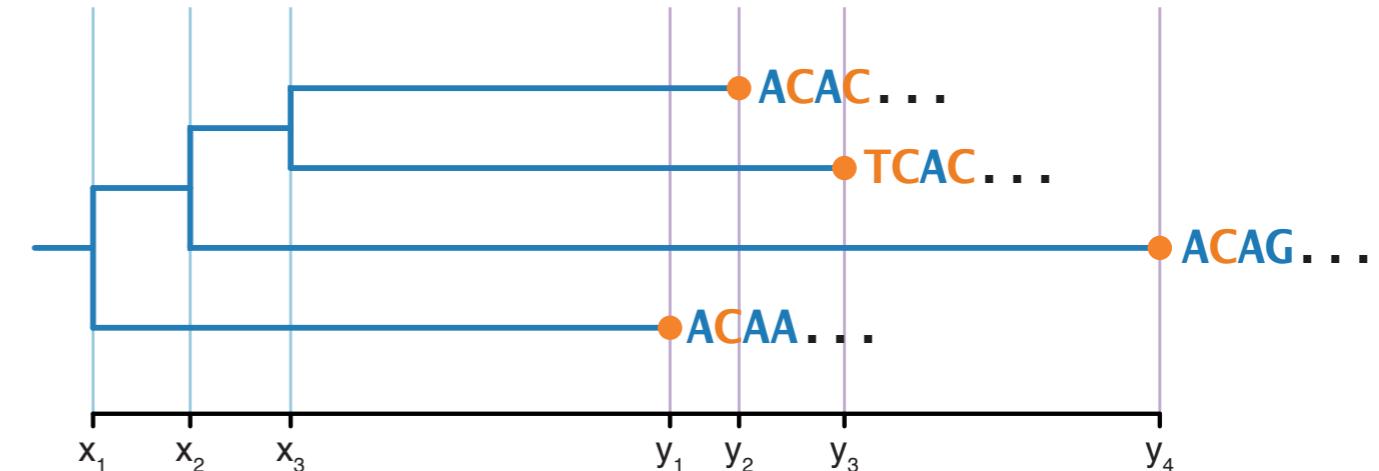
# The genealogy (tree)

Sequences sampled  
at one time point

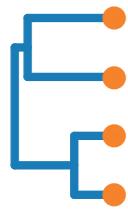


**homochronous tree**

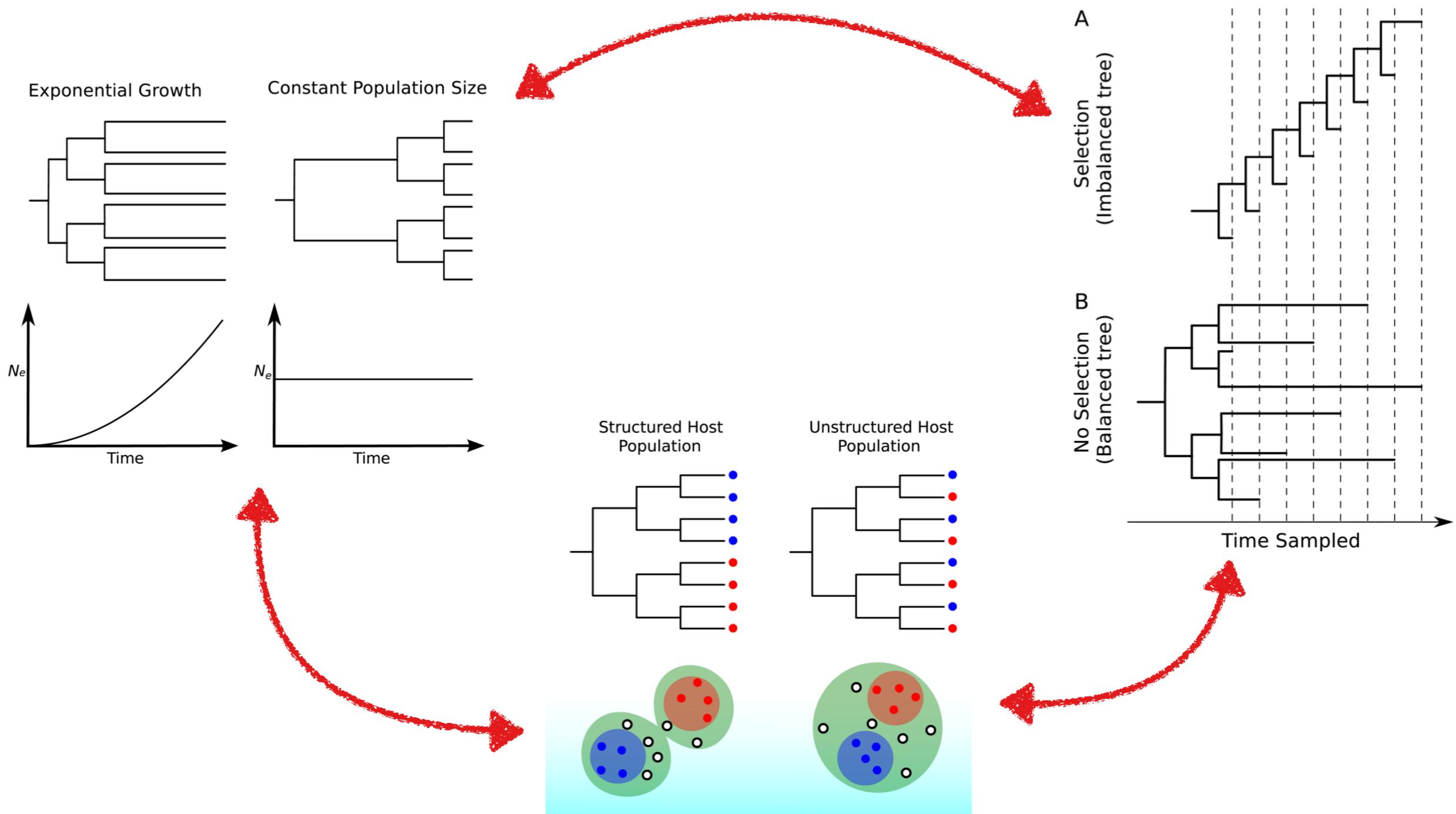
Sequences sampled  
at many time points



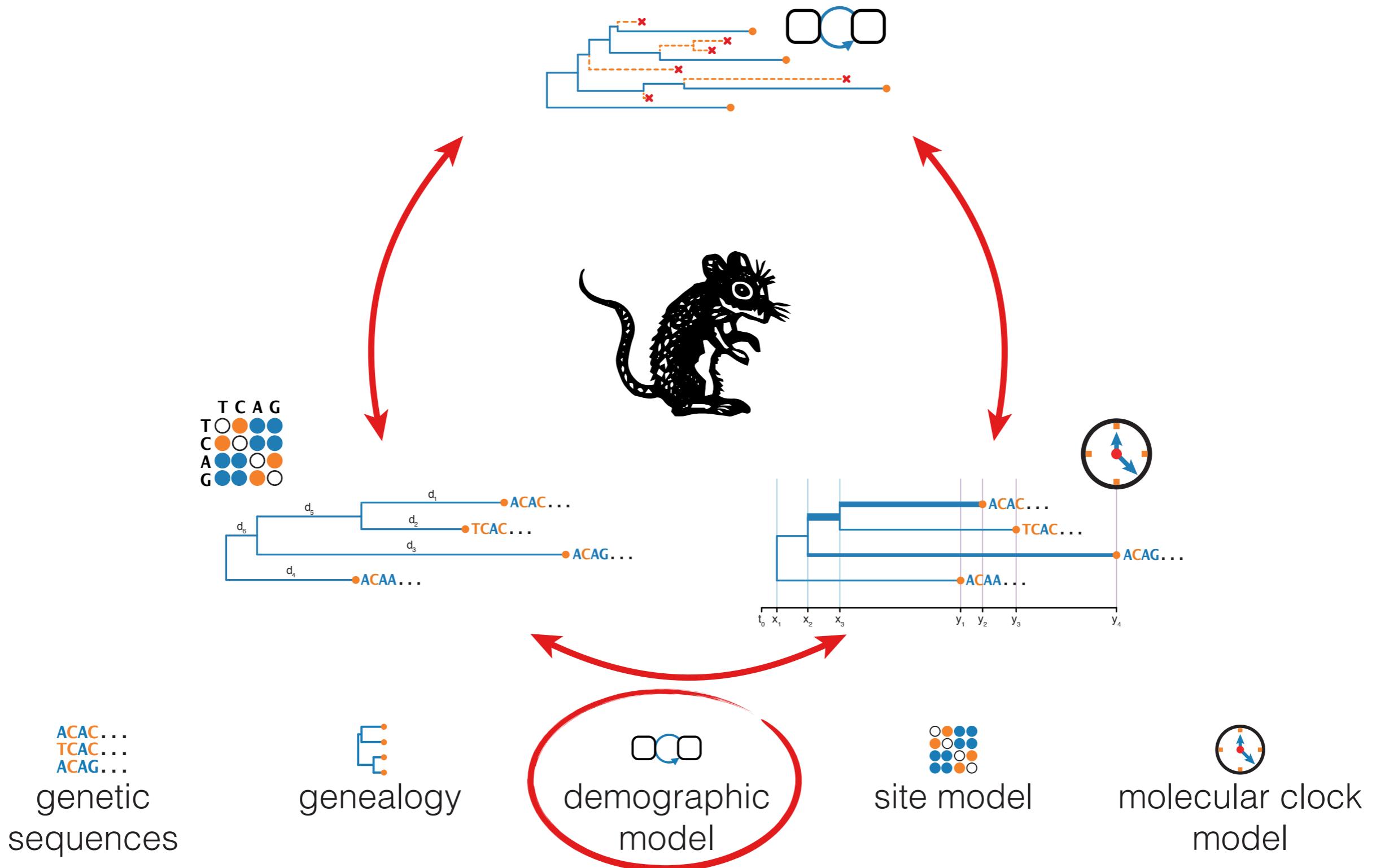
**heterochronous tree**



# Different population dynamics generate trees that look different

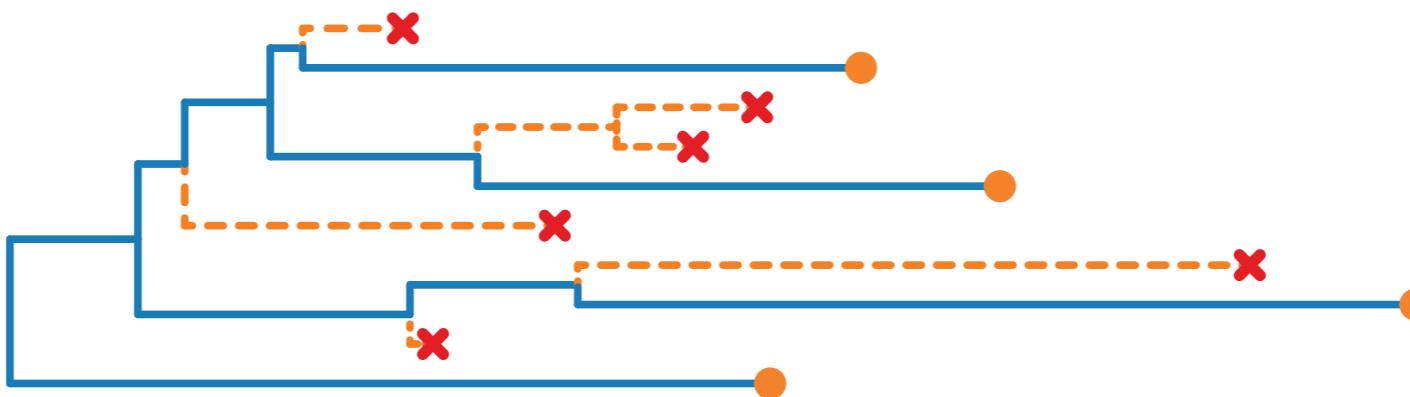


# What goes into a **BEAST** model?





# Demographic model



- Describes the population / speciation dynamics
- How does the population demographics / species diversity change over time?
- How likely is the genealogy given a demographic model?



# Demographic model / Tree prior



## Tree

Realisation of a stochastic process

$$P(F|oo)$$

## Demographic model

Describes the population dynamics (growth of the tree)

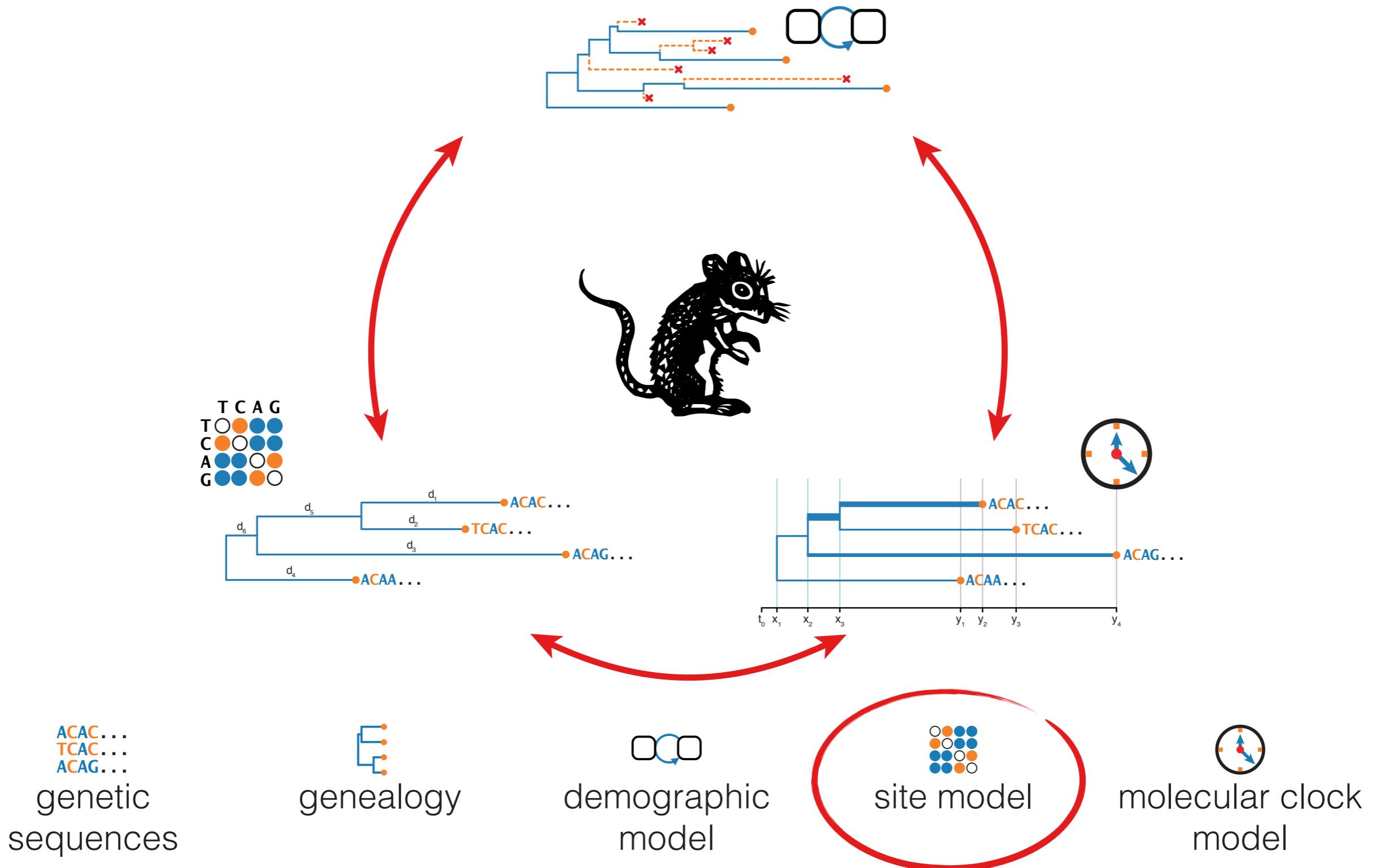
- **Coalescent:**

Given  $n$  sampling times and an estimate for  $N_e(t)$ , out of all the ways we can connect the samples, what is the probability of the current tree?

- **Birth-death:**

Given an estimate for the **origin** time, **birth**, **death** and **sampling** rates, if we simulate a tree forward-in-time from the origin to the time of the most recent sample, out of all the trees with  $n$  samples, what is the probability of the current tree?

# What goes into a **BEAST** model?



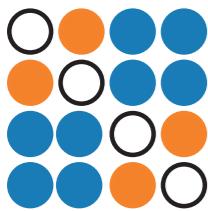
genetic  
sequences

genealogy

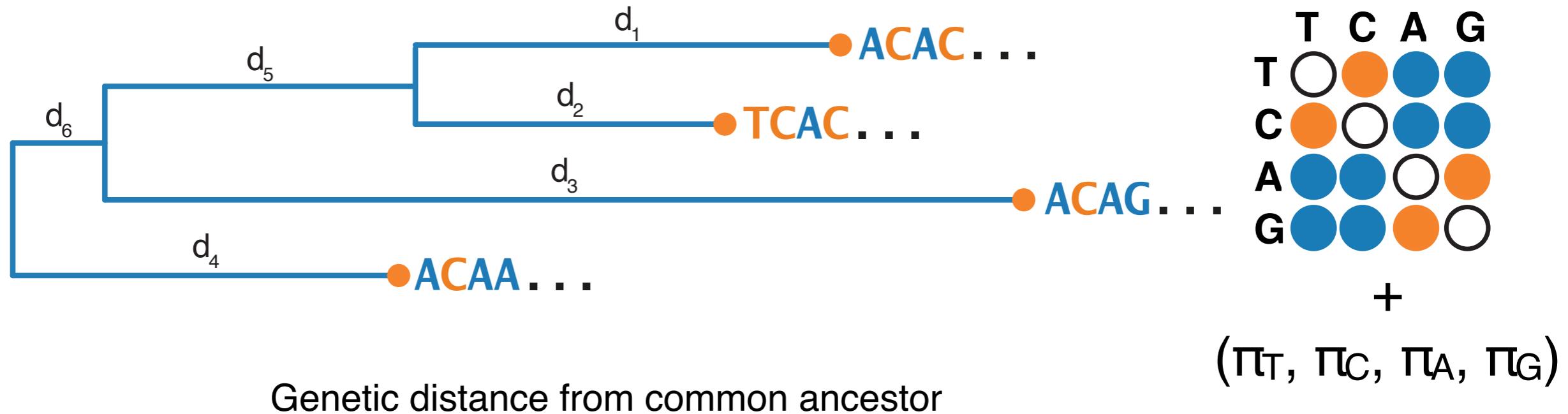
demographic  
model

site model

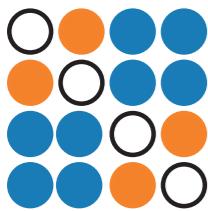
molecular clock  
model



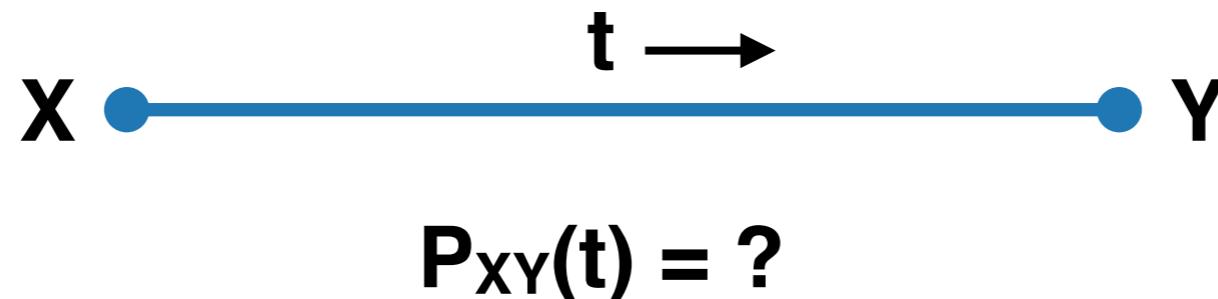
# Site model



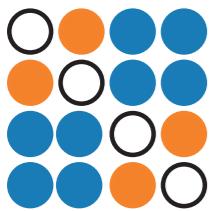
- We observe sequences at the tips, not at internal nodes
- **Substitution model** describes rates of substitution between available characters relative to genetic distance (expected substitutions/site), as well as equilibrium frequencies of characters
- **Site model** describes how the substitution model varies among sites
- **Site model links sequences to the genealogy**
  - using Felsenstein's pruning algorithm we can calculate the likelihood:  $P(\text{ACAC}\dots | \text{TCAC}\dots, \text{ACAG}\dots)$



# Substitution model



- What is the probability of observing **Y** at the end of the branch?
- Multiple substitutions at the same site means not all substitutions are observed
- Need to account for **all** possible trajectories from **X** to **Y**



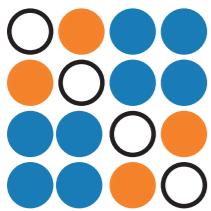
# Substitutions as a Markov process

## Assume:

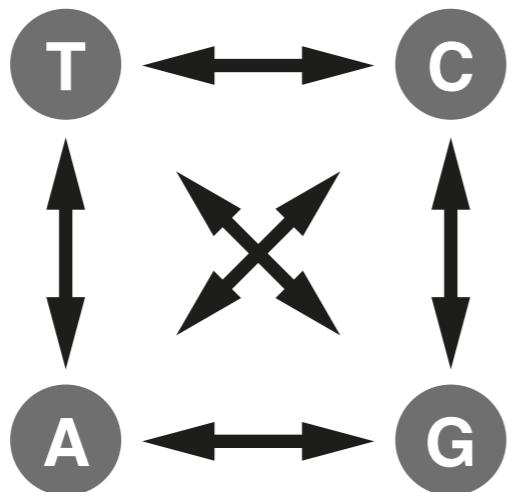
- Every site is evolving independently
- Substitutions at each site is governed by a (usually reversible) Markov process

$$Q = \begin{pmatrix} T & C & A & G \\ T & -(a+b+c) & a & b & c \\ C & d & -(d+e+f) & e & f \\ A & g & h & -(g+h+i) & i \\ G & j & k & l & -(j+k+l) \end{pmatrix}$$

**Process is governed by a rate matrix (Q) which gives the relative rates of substitutions between nucleotides**



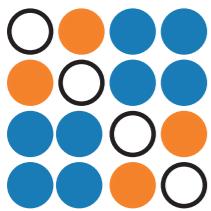
# Jukes-Cantor model (JC69)



$$\begin{matrix} & T & C & A & G \\ T & \cdot & \lambda & \lambda & \lambda \\ C & \lambda & \cdot & \lambda & \lambda \\ A & \lambda & \lambda & \cdot & \lambda \\ G & \lambda & \lambda & \lambda & \cdot \end{matrix}$$

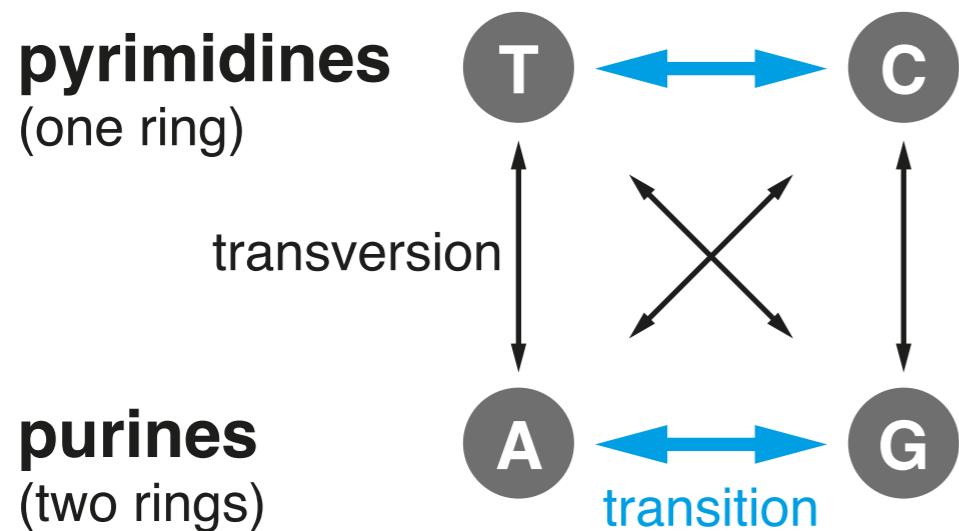
$$\pi_T = \pi_C = \pi_A = \pi_G$$

- Simplest model
- All rates and frequencies are equal!



# Kimura 2-parameter model (K80)

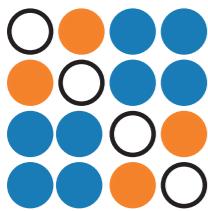
(courtesy of Carsten Magnus)



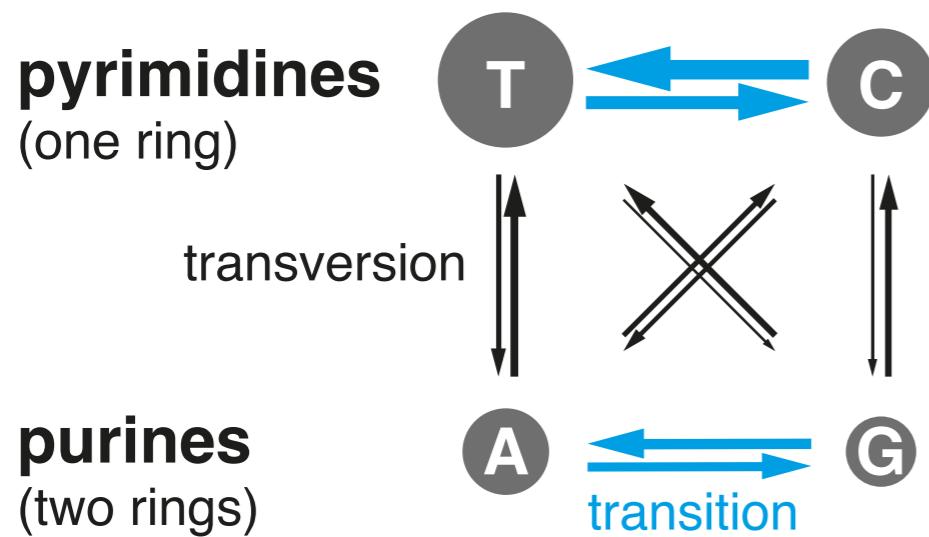
$$\begin{matrix} & T & C & A & G \\ T & \cdot & \alpha & \beta & \beta \\ C & \alpha & \cdot & \beta & \beta \\ A & \beta & \beta & \cdot & \alpha \\ G & \beta & \beta & \alpha & \cdot \end{matrix}$$

$$\pi_T = \pi_C = \pi_A = \pi_G$$

- Accounts for transition/transversion bias
- Still symmetric ( $r_{ij} = r_{ji}$ )
- Equilibrium frequencies still equal  
After a long period of evolution  $p(T) = p(C) = p(A) = p(G) = 0.25$



# HKY-model (HKY85)

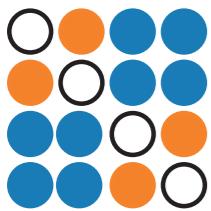


$$(\pi_T, \pi_C, \pi_A, \pi_G)$$

$$\begin{matrix}
 & T & C & A & G \\
 T & \cdot & \alpha\pi_C & \beta\pi_A & \beta\pi_G \\
 C & \alpha\pi_T & \cdot & \beta\pi_A & \beta\pi_G \\
 A & \beta\pi_T & \beta\pi_C & \cdot & \alpha\pi_G \\
 G & \beta\pi_T & \beta\pi_C & \alpha\pi_A & \cdot
 \end{matrix}$$

$$= \begin{pmatrix} \cdot & \alpha & \beta & \beta \\ \alpha & \cdot & \beta & \beta \\ \beta & \beta & \cdot & \alpha \\ \beta & \beta & \alpha & \cdot \end{pmatrix} \cdot \begin{pmatrix} \pi_T & 0 & 0 & 0 \\ 0 & \pi_C & 0 & 0 \\ 0 & 0 & \pi_A & 0 \\ 0 & 0 & 0 & \pi_G \end{pmatrix}$$

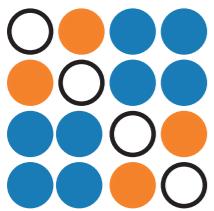
- Accounts for transition/transversion bias
- Accounts for unequal equilibrium frequencies
- Not symmetric anymore ( $r_{ij} \neq r_{ji}$ )
- Still time-reversible ( $\pi_i q_{ij} = \pi_j q_{ji}$ )



# General time-reversible model (GTR/REV) (courtesy of Carsten Magnus)

$$\begin{matrix} & T & C & A & G \\ T & \cdot & a\pi_C & b\pi_A & c\pi_G \\ C & a\pi_T & \cdot & d\pi_A & e\pi_G \\ A & b\pi_T & d\pi_C & \cdot & f\pi_G \\ G & c\pi_T & e\pi_C & f\pi_A & \cdot \end{matrix} = \begin{pmatrix} \cdot & a & b & c \\ a & \cdot & d & e \\ b & d & \cdot & f \\ c & e & f & \cdot \end{pmatrix} \cdot \begin{pmatrix} \pi_T & 0 & 0 & 0 \\ 0 & \pi_C & 0 & 0 \\ 0 & 0 & \pi_A & 0 \\ 0 & 0 & 0 & \pi_G \end{pmatrix}$$

- Most general time-reversible model
- More flexible models are possible, but mathematically inconvenient



# Transition probability matrix

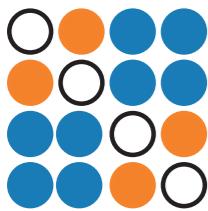
$$\mathbf{P}(t) = e^{\mathbf{Q}t} \quad \mathbf{P}(t) = \begin{pmatrix} T & C & A & G \\ T & p_{tt}(t) & p_{tc}(t) & p_{ta}(t) & p_{tg}(t) \\ C & p_{ct}(t) & p_{cc}(t) & p_{ca}(t) & p_{cg}(t) \\ A & p_{at}(t) & p_{ac}(t) & p_{aa}(t) & p_{ag}(t) \\ G & p_{gt}(t) & p_{gc}(t) & p_{ga}(t) & p_{gg}(t) \end{pmatrix}$$

- Transition probabilities take into account every possible evolutionary trajectory (Chapman-Kolmogorov theorem)

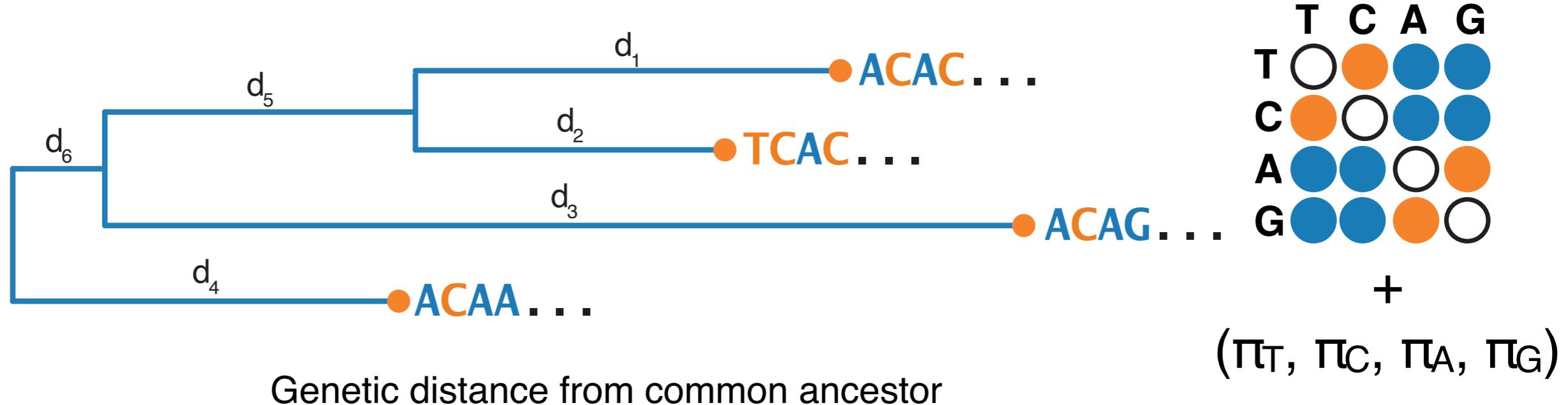


$$P_{XY}(t) = ?$$

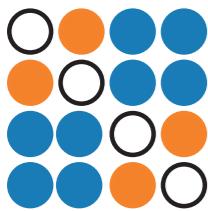
- $\mathbf{Q}$  only gives the **relative substitution rates**  
⇒ **distance is measured in expected number of substitutions per site**



# Site model



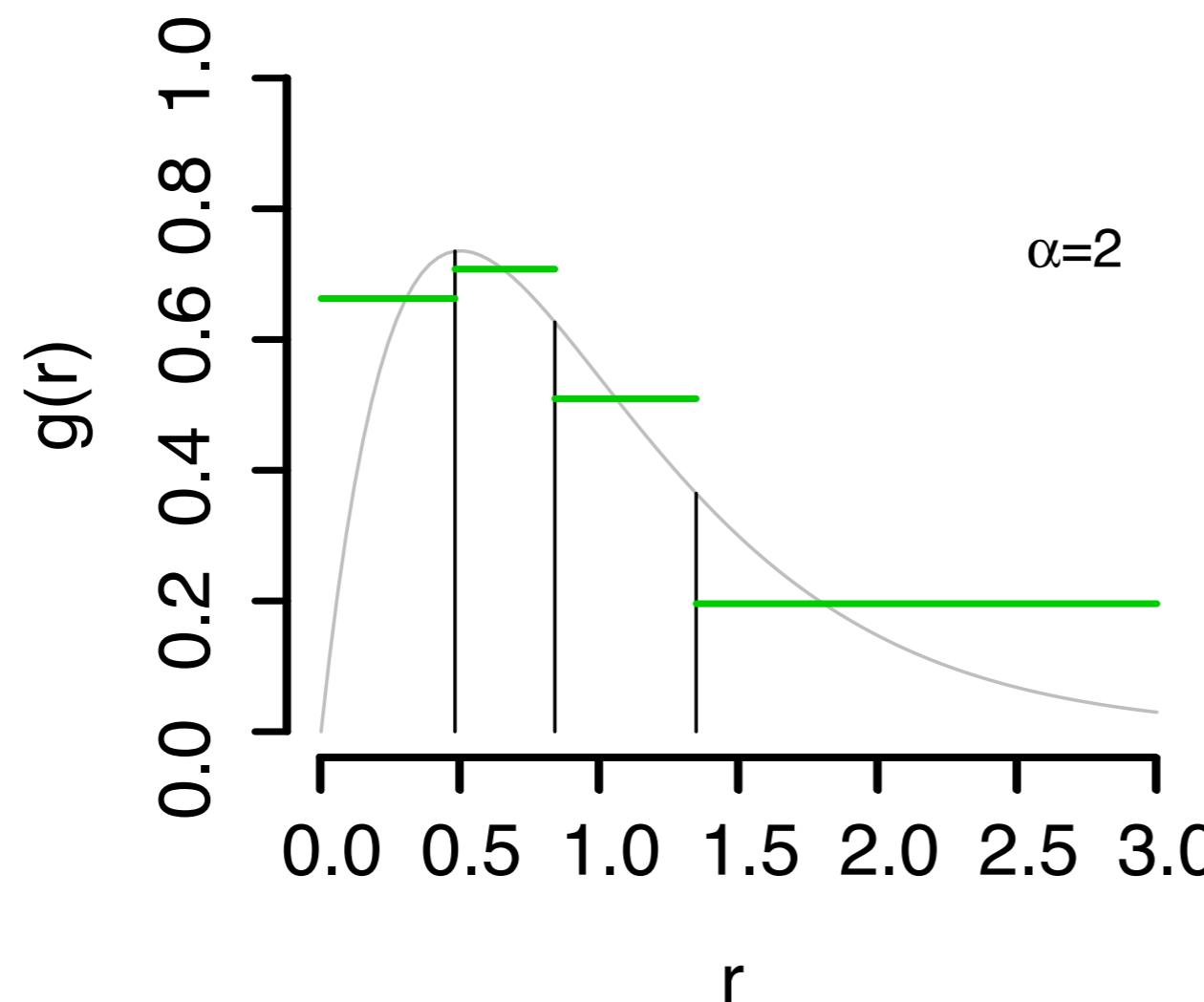
- Describes how the substitution model varies from site-to-site
- Assume every site is evolving independently
- Account for rate heterogeneity between sites:
  - Proportion of invariant sites
  - Gamma rate heterogeneity
  - Multi-locus models

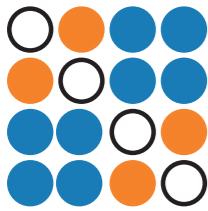


# Gamma rate heterogeneity

(courtesy of Carsten Magnus)

- Not all sites evolve at the same rate
- Assume rate heterogeneity among sites is  $\Gamma$ -distributed
- Discretise  $\Gamma$ -distribution to  $n$  discrete rate categories for computational reasons



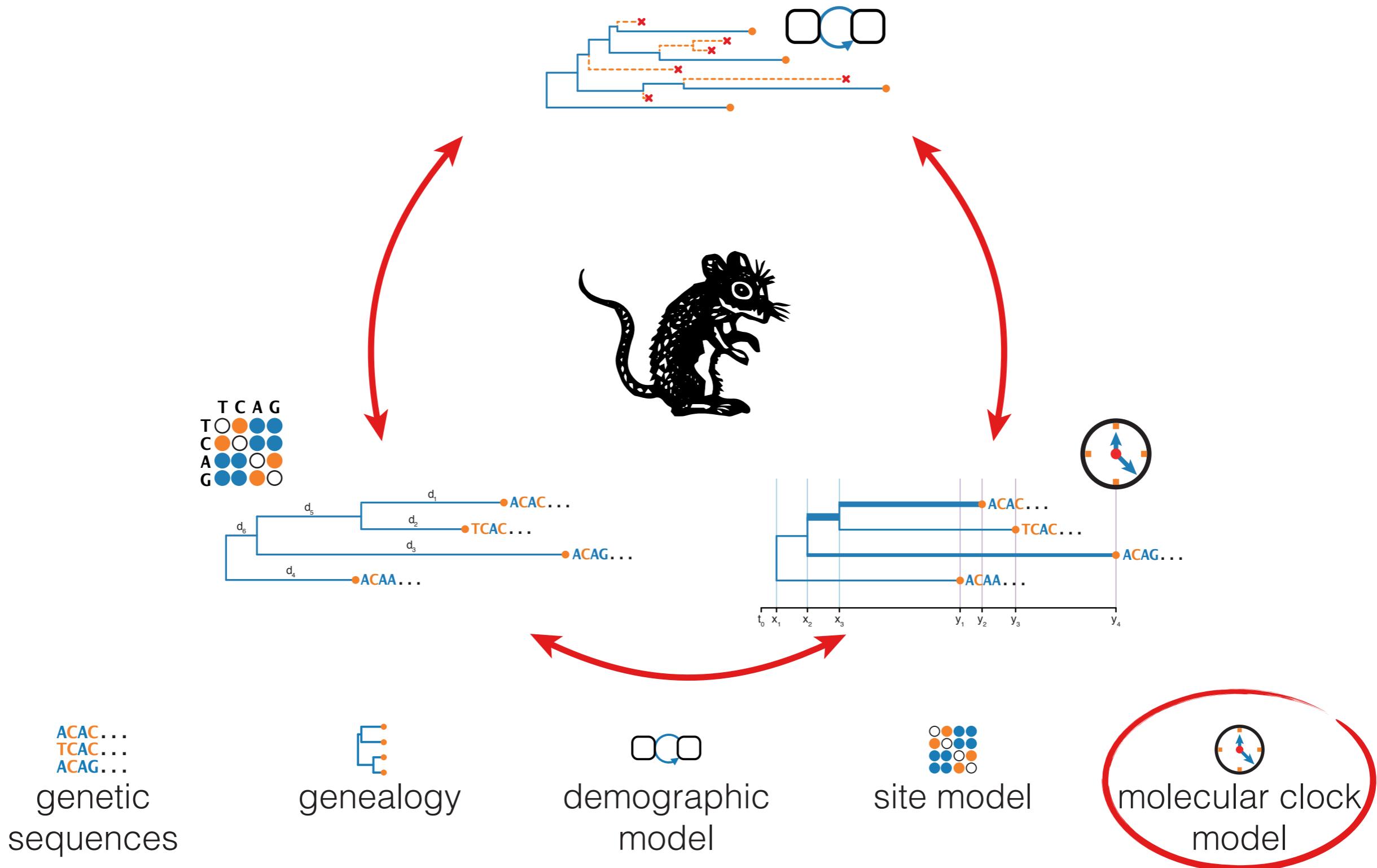


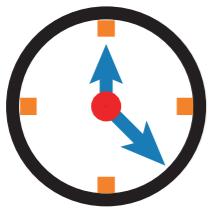
# Multi-locus models

---

- $\Gamma$ -distributed rate variation is not always flexible enough to model differences between different loci
- Use a separate substitution model for each locus
- Can also use separate models for different codon positions
- Loci could be different genes, different codon positions, etc.

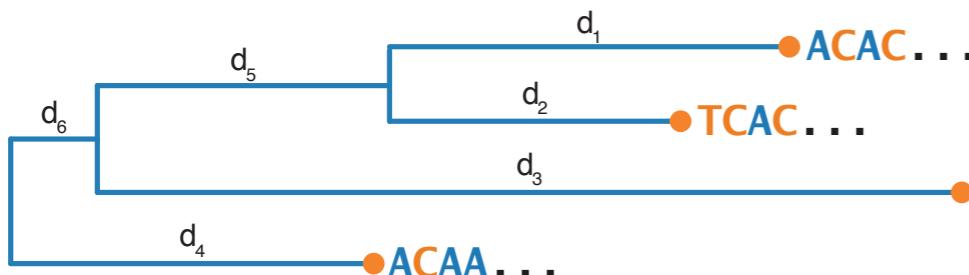
# What goes into a **BEAST** model?





# Molecular clock model

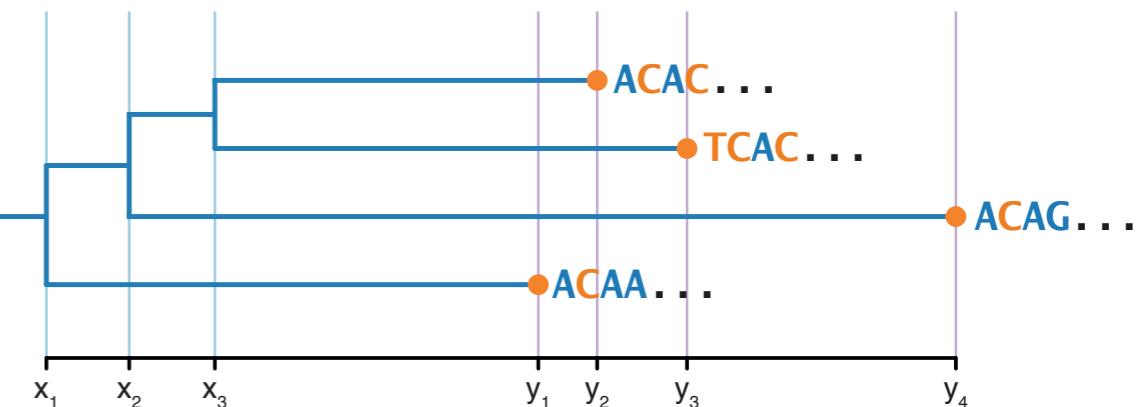
**genetic distance tree**  
(subst/site)



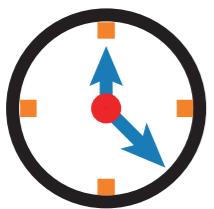
**clock rate**  
(subst/site/year)

$$= \mu \times$$

**time tree**  
(years)

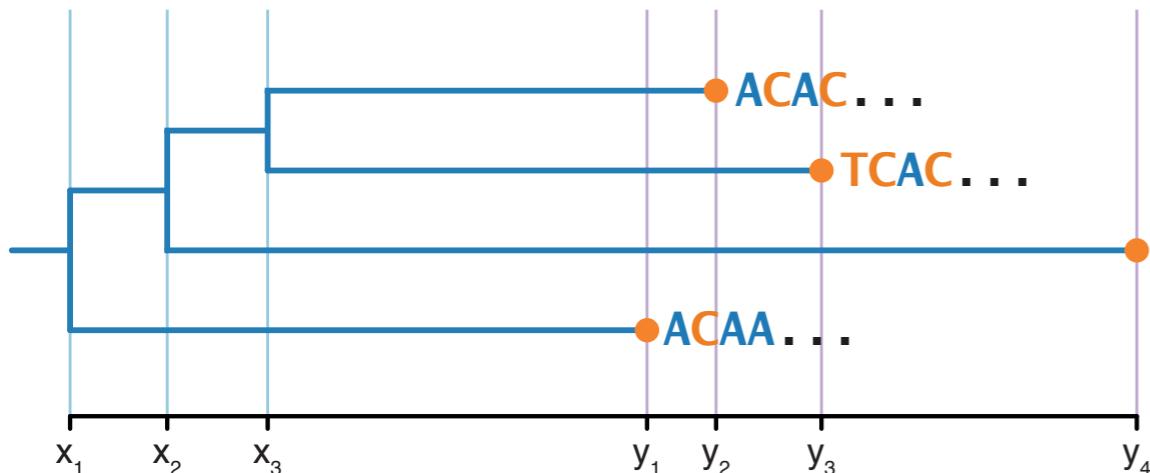


- Scales branch lengths to calendar time
- How long does it take for substitutions to appear?
- **Genetic distance = Rate x Time**

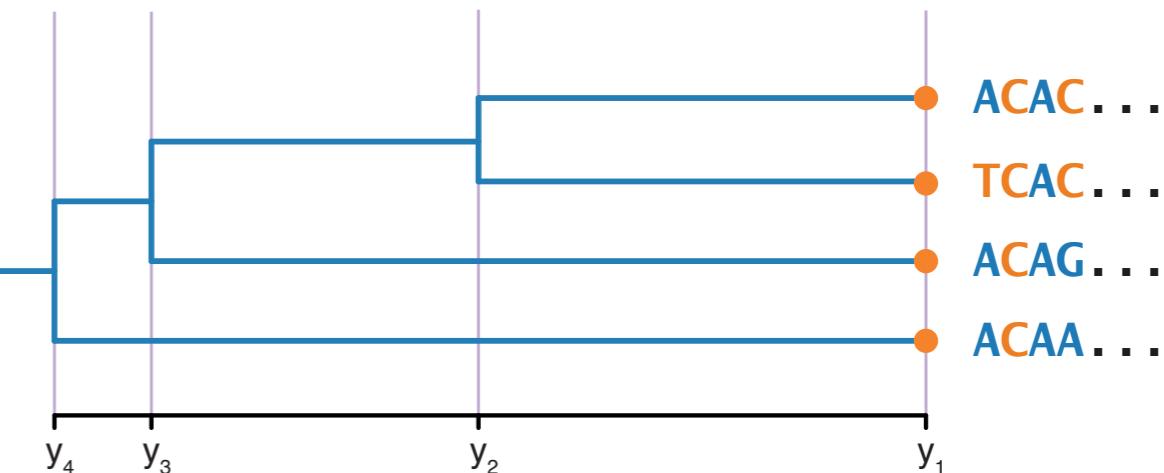


# Molecular clock model

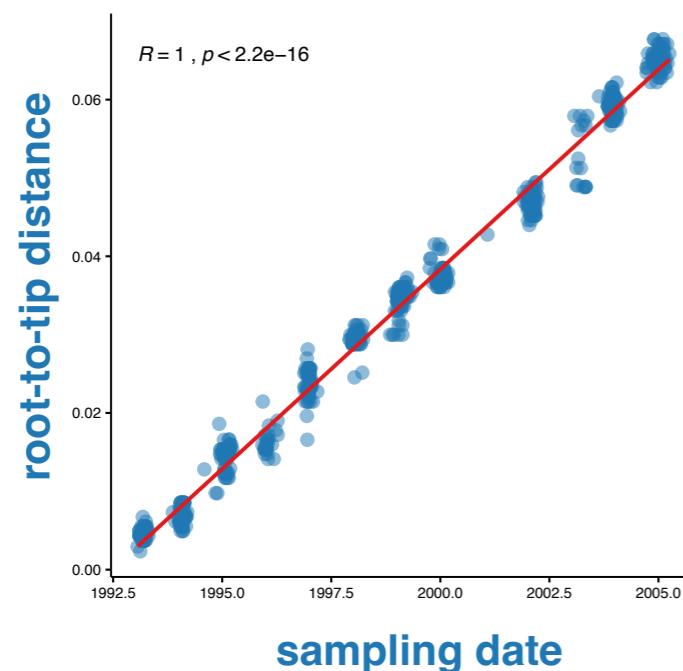
## heterochronous tree



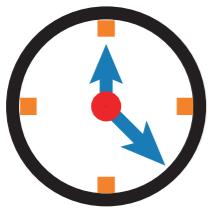
## homochronous tree



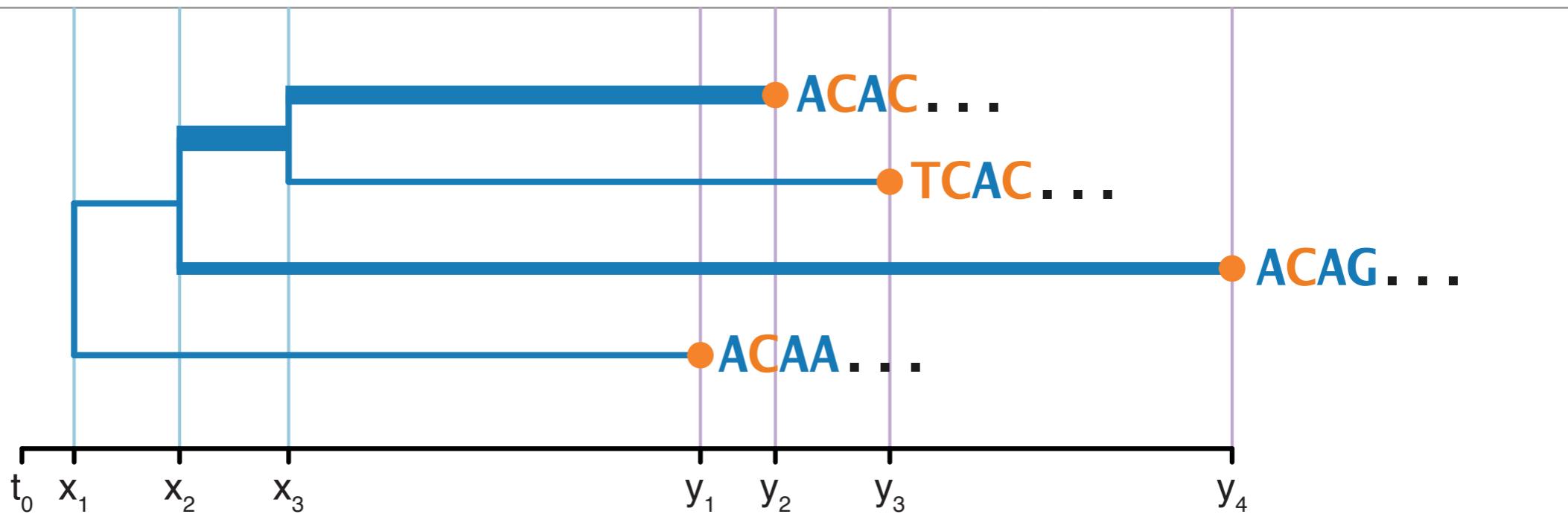
- Correlation between sampling time and genetic distance from the root (clock signal)



- Need external information to calibrate the clock
- Fix clock rate or calibrate some internal nodes with external information



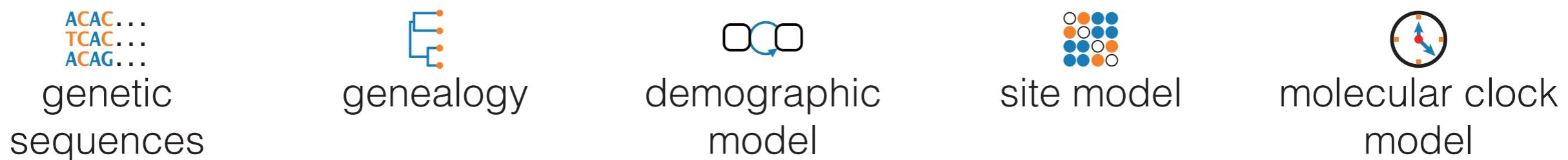
# Relaxed and local clocks



Some branches may have different clock rates!

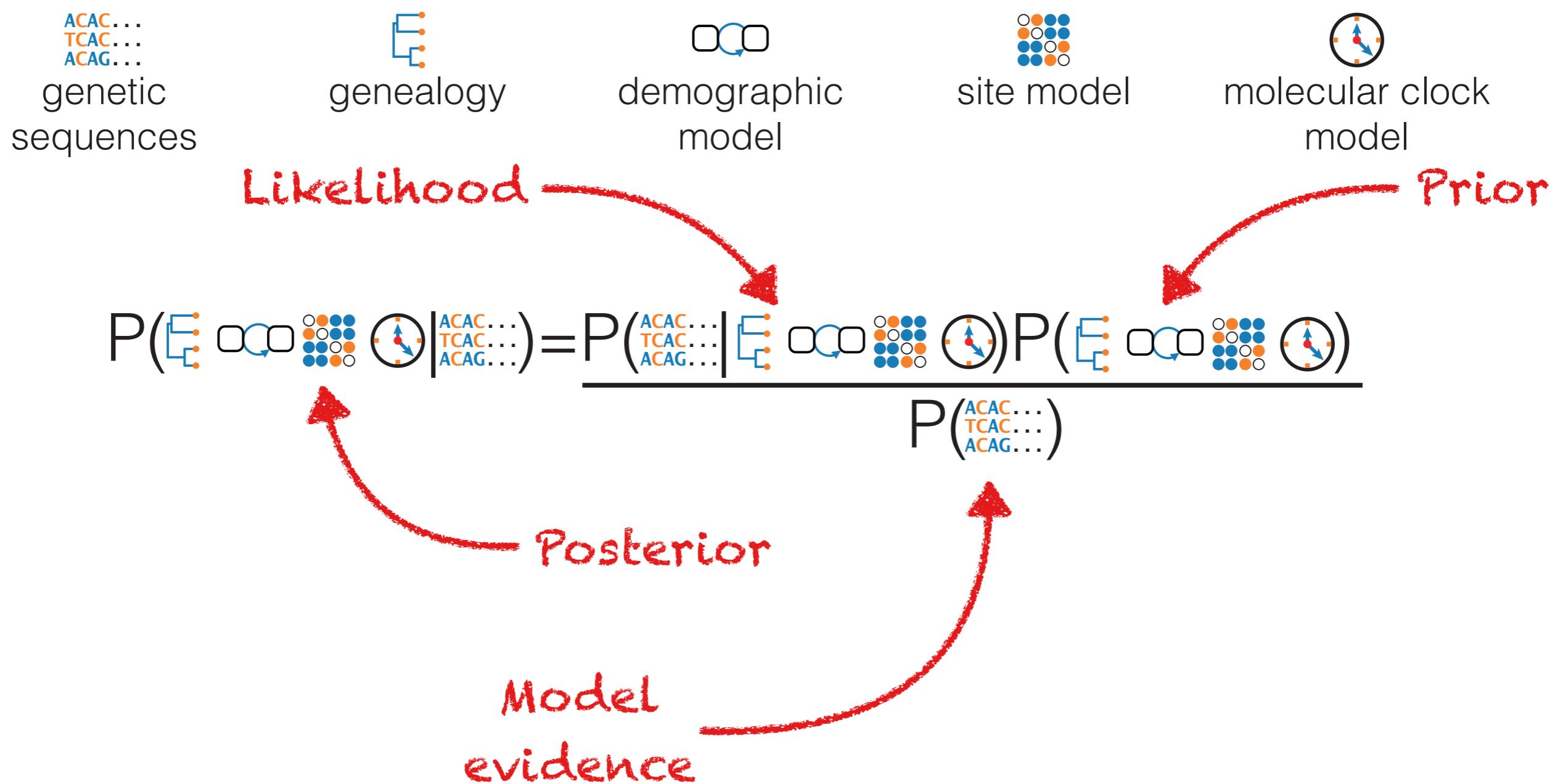
# Putting it all together

---

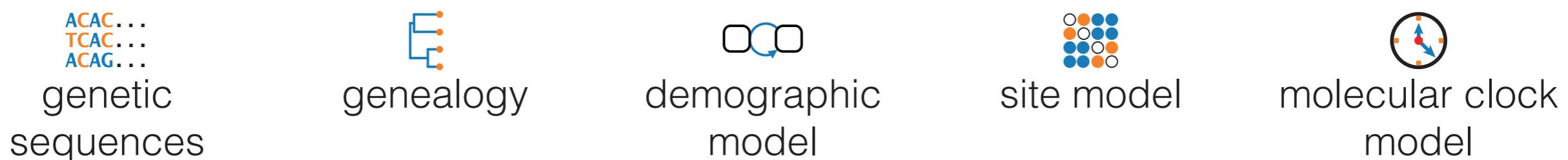


$$P(\text{hypothesis} \mid \text{data}) = \frac{P(\text{data} \mid \text{hypothesis})P(\text{hypothesis})}{P(\text{data})}$$

# Putting it all together



# Putting it all together

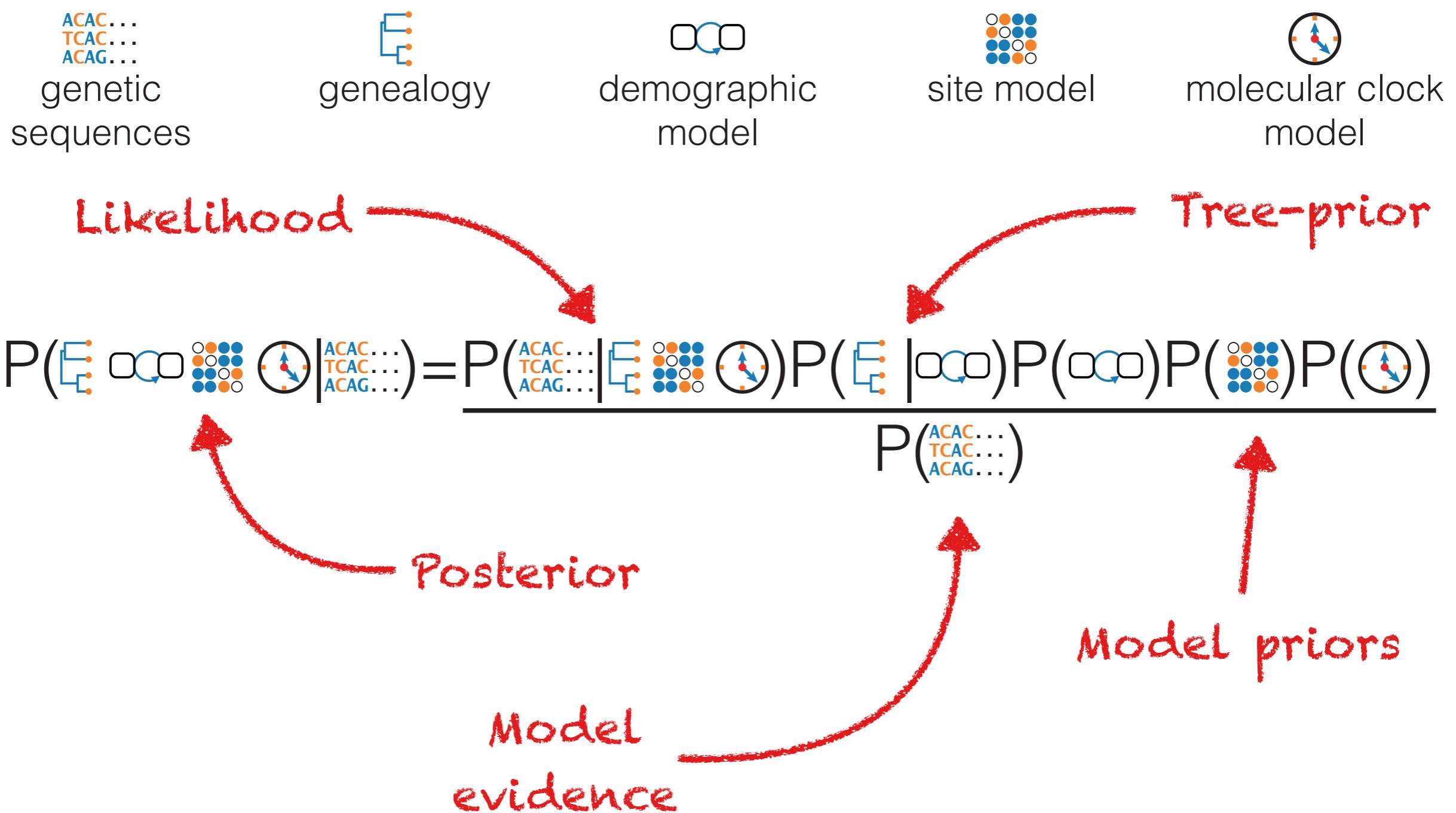


$$P(E \circ \circ \circ \bullet \bullet \bullet | ACAC \dots) = \frac{P(ACAC \dots | E \circ \circ \circ \bullet \bullet \bullet) P(E \circ \circ \circ \bullet \bullet \bullet)}{P(ACAC \dots)}$$

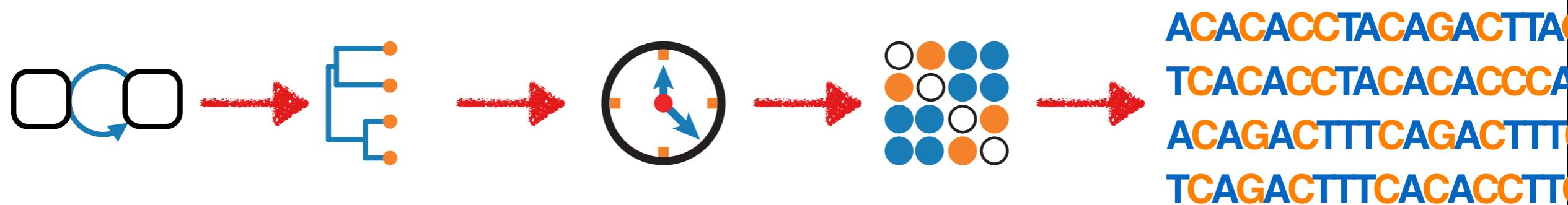
**Assume independence**

$$P(E \circ \circ \circ \bullet \bullet \bullet | ) = P(E | ) P(\circ \circ \circ) P(\bullet \bullet \bullet) P()$$

# Posterior distribution in BEAST2



# Neutrality assumption



**Assumed tree-generating process  
independent of sequence data**

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

# How to find a posterior



# How can we find the posterior?

- We want to calculate the posterior distribution

$$P(E \cap O \cap D \cap C | ACAC, TCAC, ACAG, \dots) =$$


- But we cannot easily calculate the marginal likelihood (model evidence)

$$P(A C A C \dots) \rightarrow ?$$

→ use **MCMC!** (Markov-chain Monte Carlo)

- MCMC is a stochastic algorithm that performs a directed random walk on the posterior, preferentially sampling high-density areas

# MCMC

(Markov-chain Monte Carlo)

---

- MCMC draws samples from the posterior
  - output is a list of values that can approximate the posterior
- Only need to compare which posterior density is higher
  - So we only need the ratio of posteriors  
(marginal likelihoods cancel out!)

$$\frac{P(\theta_1 | D)}{P(\theta_2 | D)} = \frac{\frac{P(D | \theta_1)P(\theta_1)}{P(\cancel{D})}}{\frac{P(D | \theta_2)P(\theta_2)}{P(\cancel{D})}}$$

# MCMC

(Markov-chain Monte Carlo)

---

- MCMC draws samples from the posterior
  - output is a list of values that can approximate the posterior
- Only need to compare which posterior density is higher
  - So we only need the ratio of posteriors  
(marginal likelihoods cancel out!)

$$\frac{P(\theta_1 | D)}{P(\theta_2 | D)} = \frac{P(D | \theta_1)P(\theta_1)}{P(D | \theta_2)P(\theta_2)}$$

# Operators

---

## Target distribution

- This is the **posterior** in BEAST2:  $P(\text{EvoSeq} | \text{ACAC...})$
- MCMC is a directed random walk through the state space and samples the target distribution
- How to pick the next state to sample?

## Proposal distribution

- Used to decide where to step to next
- The choice only affects the **efficiency** of the algorithm
- In BEAST1 and BEAST2 operators are used to propose the next step
- A parameter (or multiple parameters) are selected and perturbed to propose a step

# Operators

---

## Target

- This
- MC
- spa
- How
- Operators are a part of the MCMC **algorithm**, not the **model!**
- Tuning operators can help to improve efficiency, but should not change the results if chains are run long enough!

## Propo

- Used to decide where to step to next
- The choice only affects the **efficiency** of the algorithm
- In BEAST1 and BEAST2 operators are used to propose the next step
- A parameter (or multiple parameters) are selected and perturbed to propose a step

# MCMC in practice

---

## Before

- Decide on the length of the chain  
(total number of steps to take)
- Decide on the sampling frequency  
(how often to record samples so  
that they are uncorrelated)

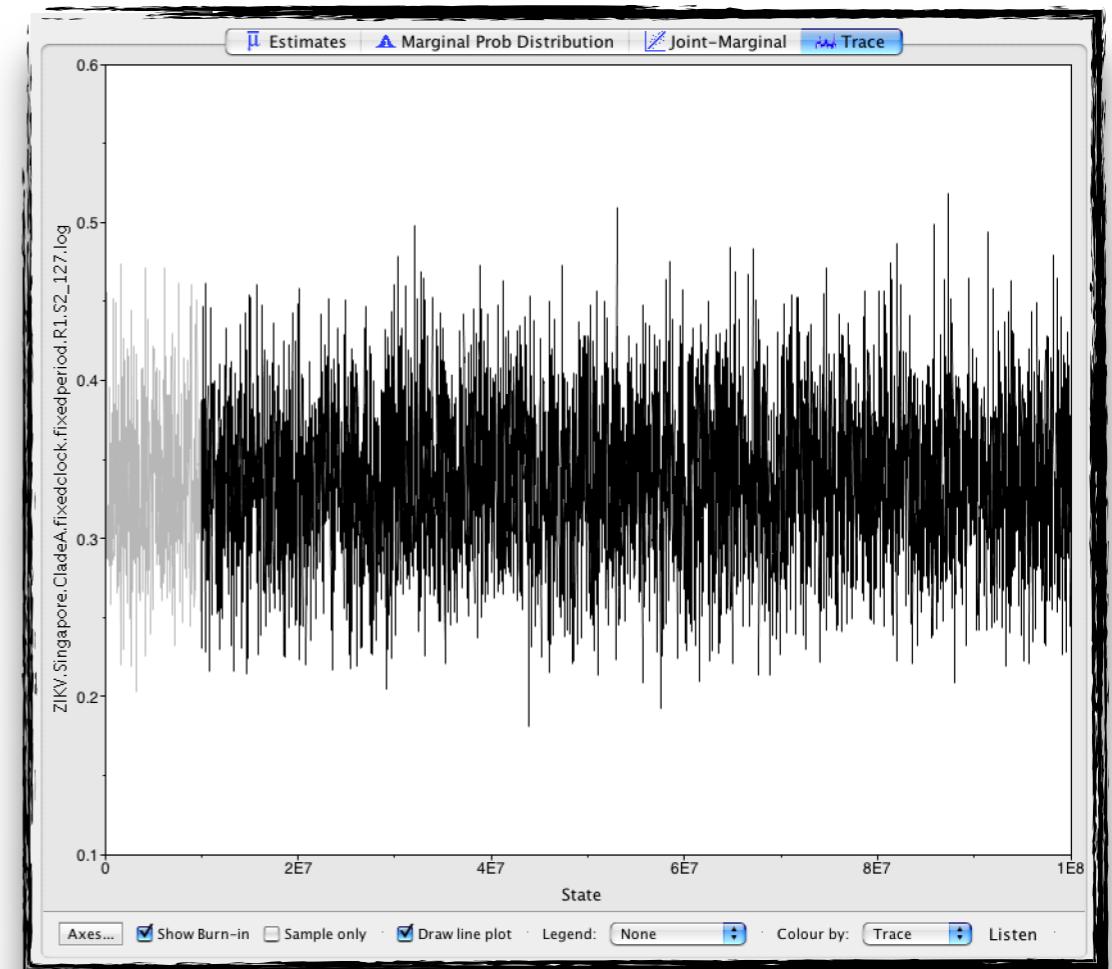
## After

- Discard burn-in  
(until stationary state is reached)
- Assess convergence and mixing
- Only then should we look at the estimates!

# What we hope will happen

---

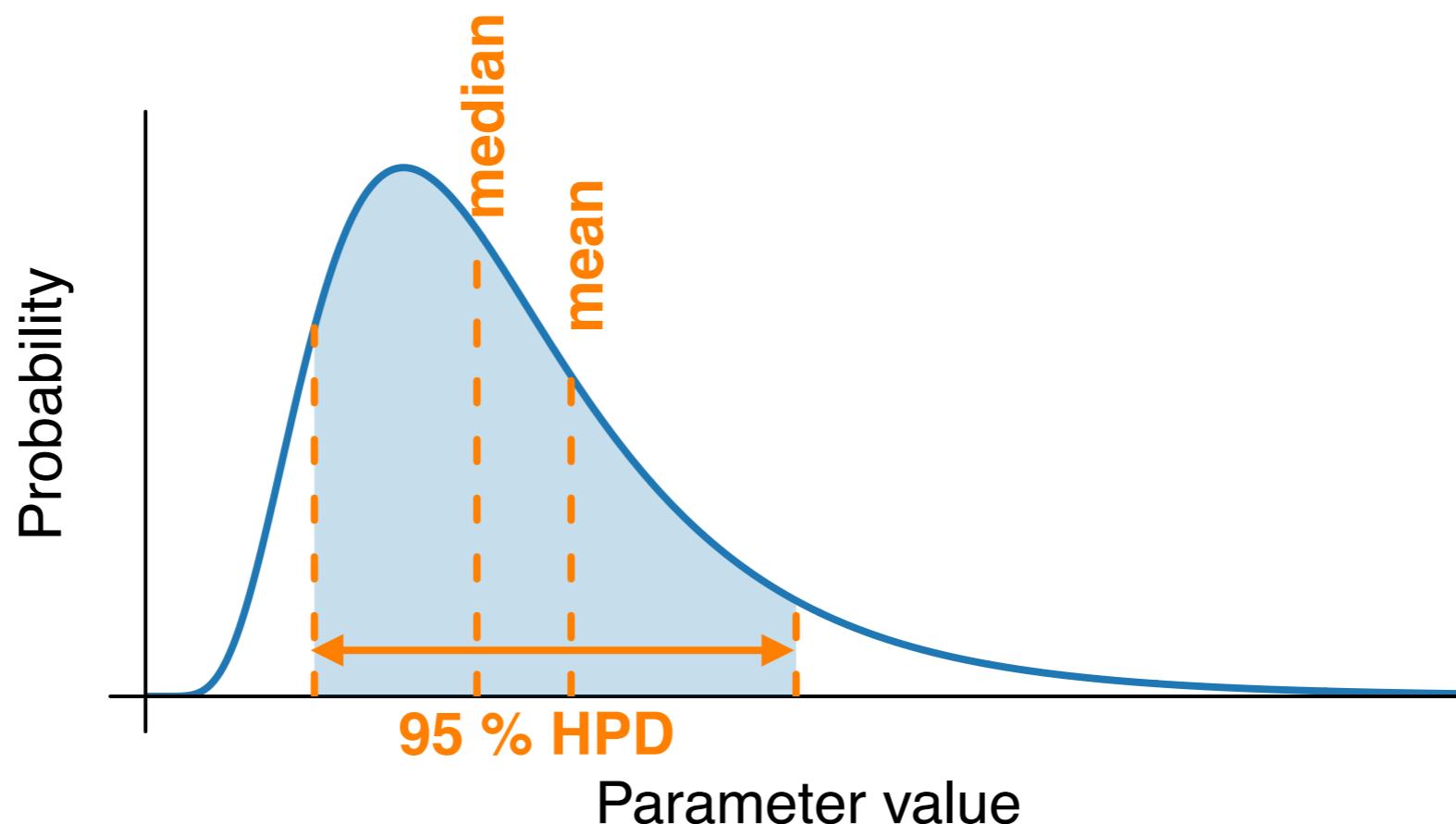
- The MCMC algorithm samples efficiently from high density areas of the posterior distribution
- We end up with a **good** approximation of the posterior distribution in **finite** time
- Samples are effectively **independent** (low autocorrelation)
- Appearance of white noise
- Everything is awesome!



Mixing well! 😊

# Summarising parameter posteriors

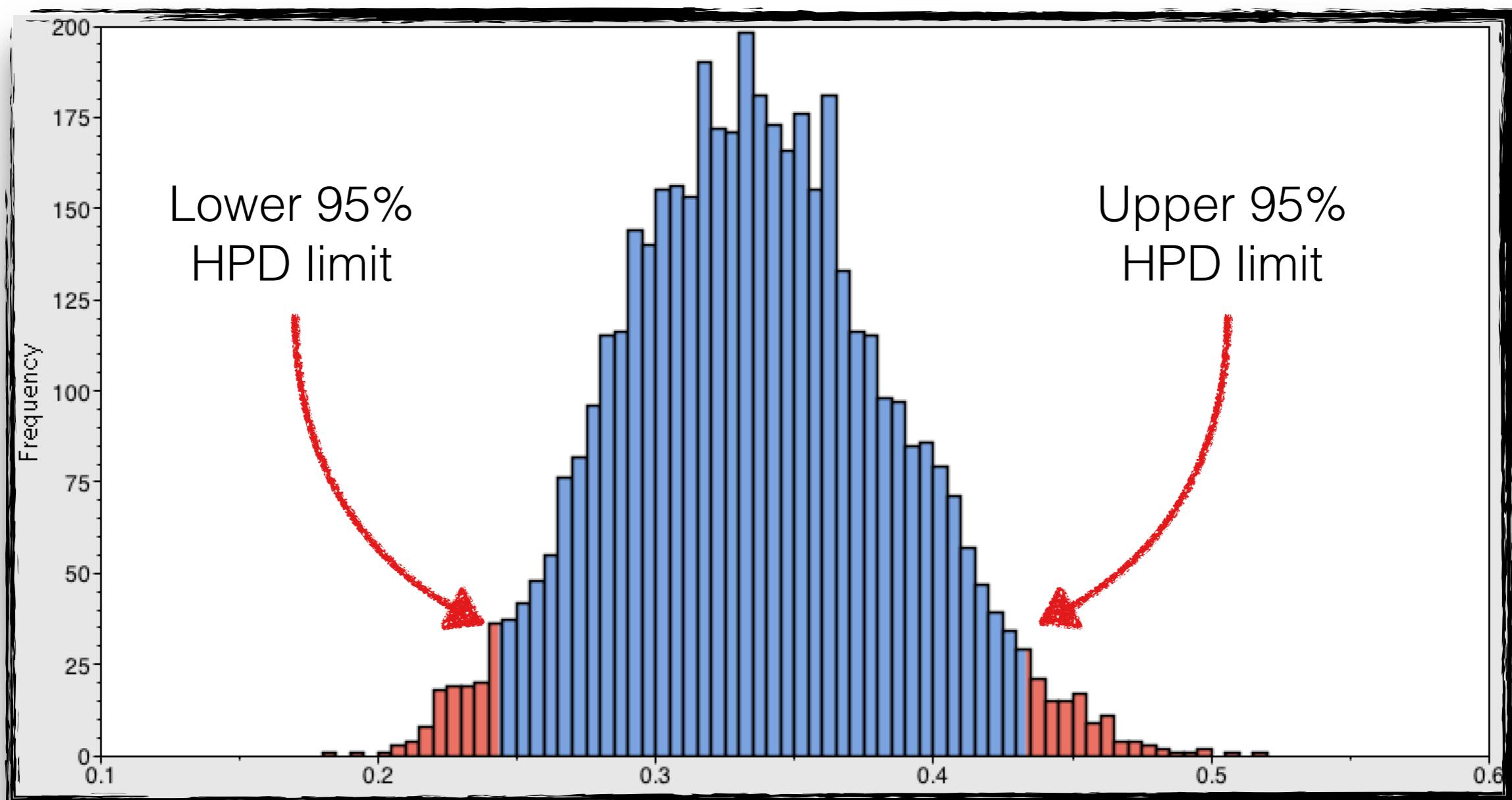
---



95% HPD (highest posterior density) interval is the smallest interval that contains 95% of the posterior probability

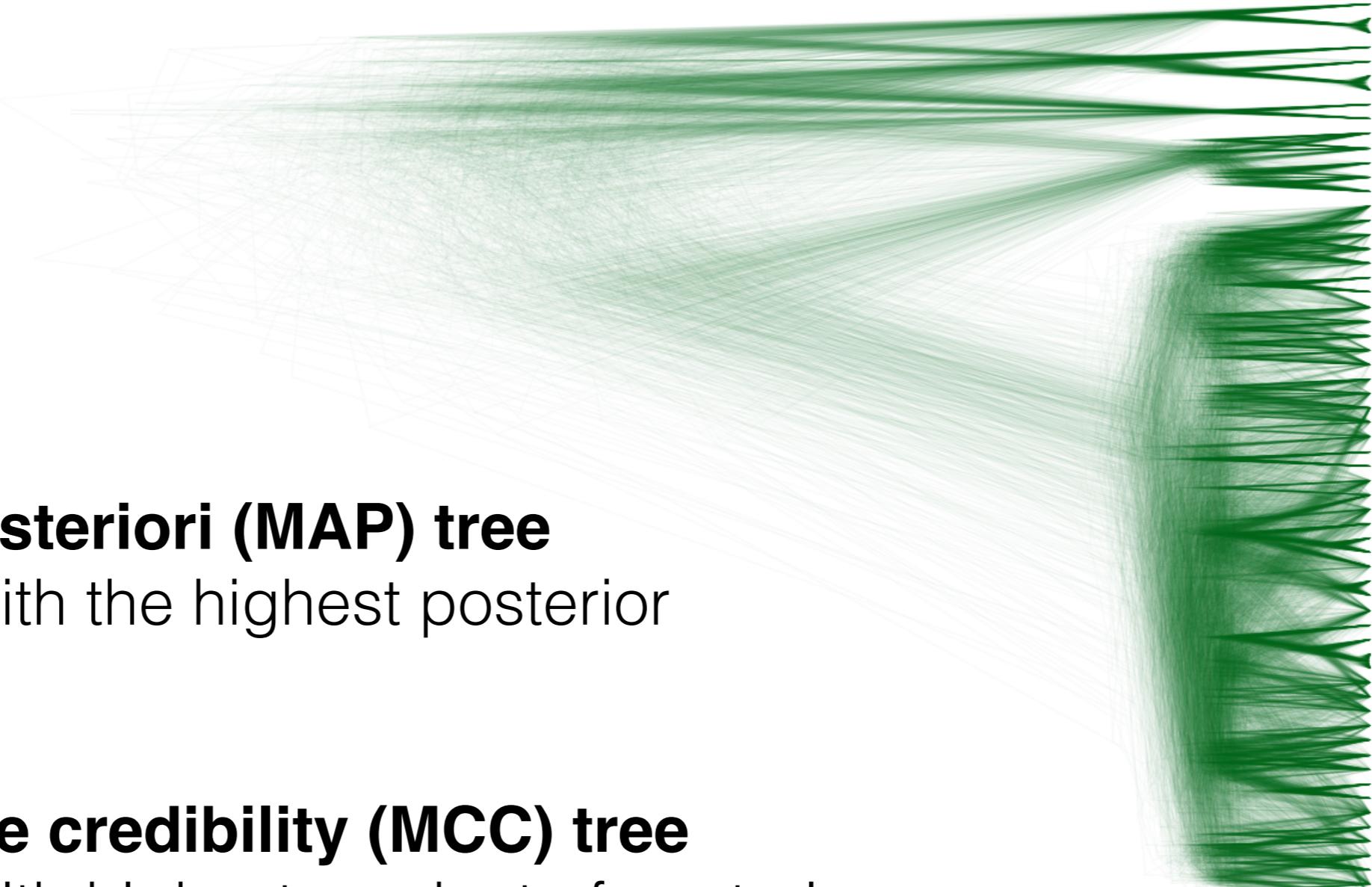
# In practice...

---



# What about trees?

---



## **Maximum a posteriori (MAP) tree**

Sampled tree with the highest posterior probability

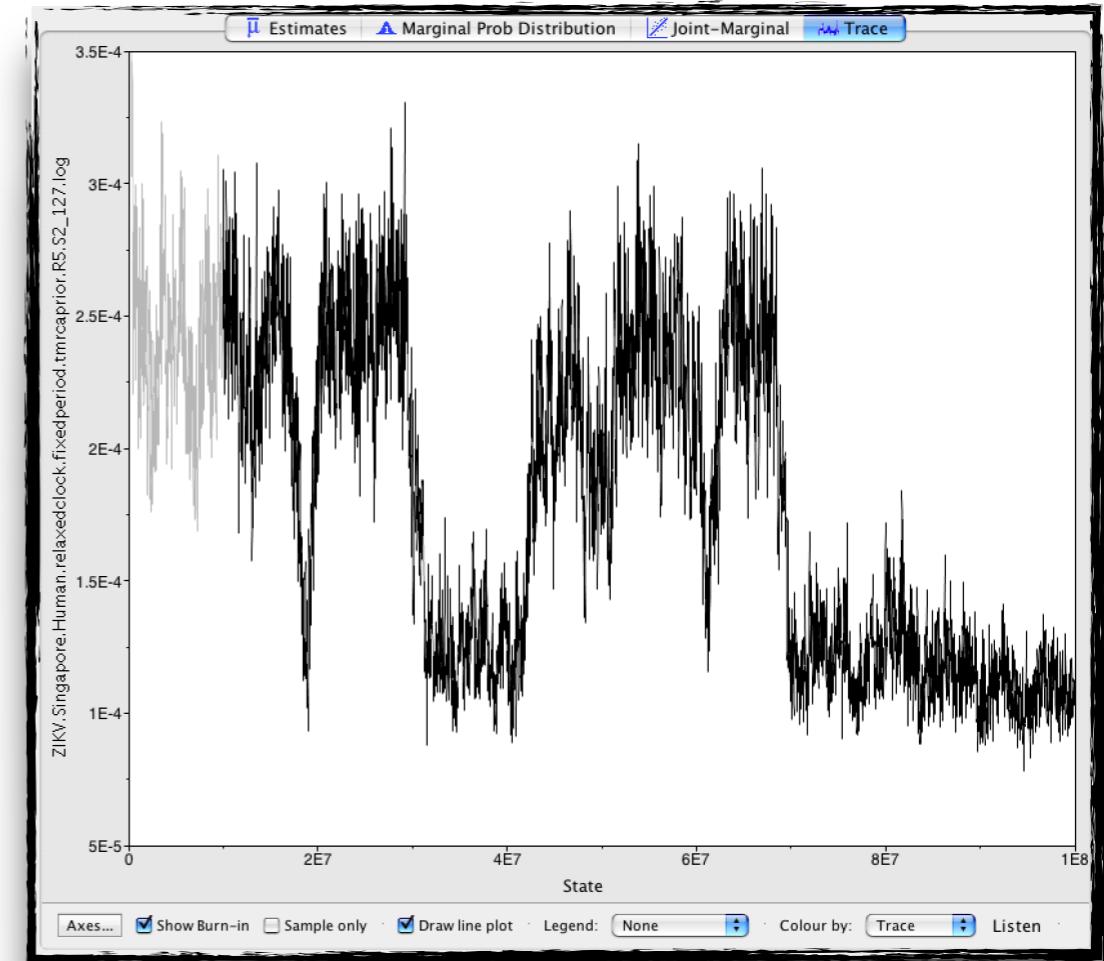
## **Maximum clade credibility (MCC) tree**

Sampled tree with highest product of posterior node probabilities

# Questions to ask...

---

- Is the chain **mixing** well?
- Are samples uniformly drawn from all over the stationary distribution?
- Do we have a “sticky chain?”



## Solutions

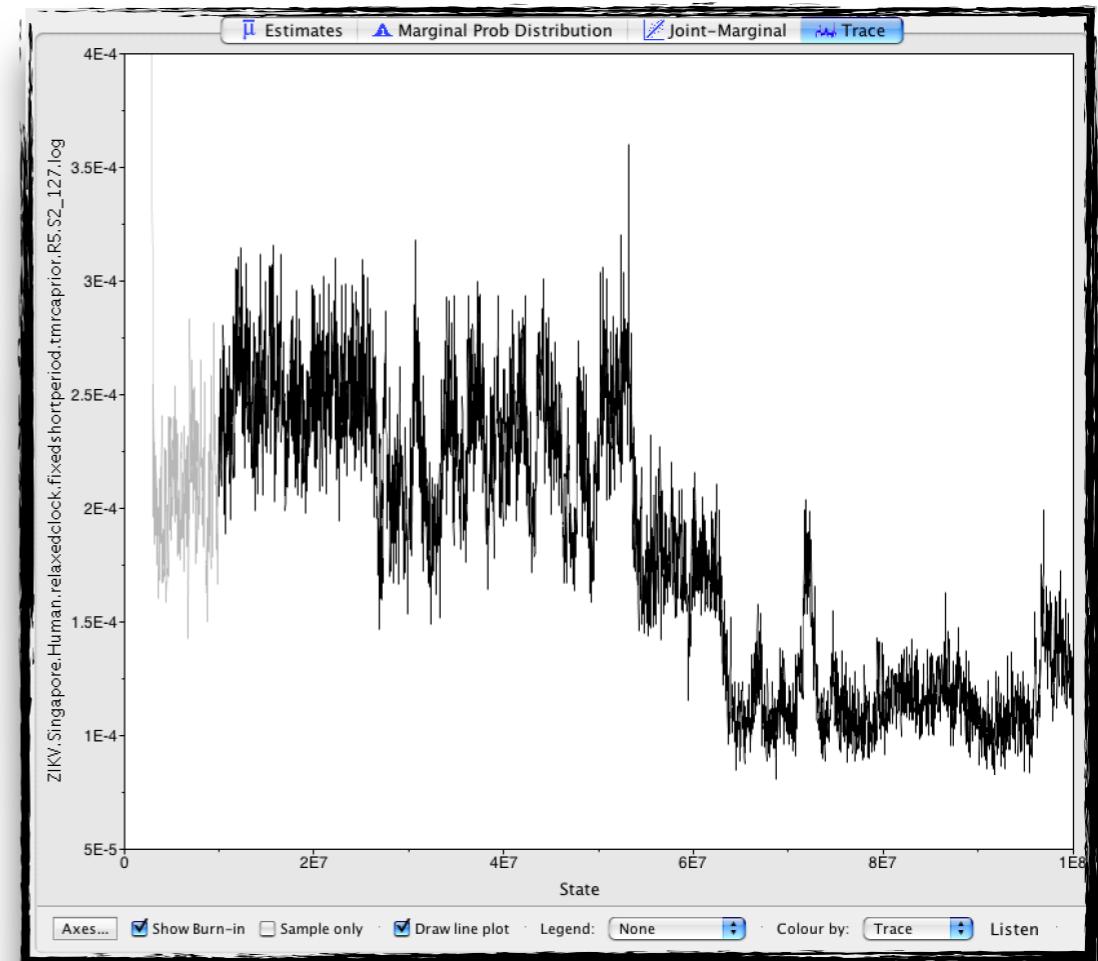
- MCMC gets stuck in some states for long times
- Tune operators to make better proposals

**Not mixing!** 😞

# Questions to ask...

---

- Has the chain **converged** to the stationary distribution?
- Did we pass the burn-in?



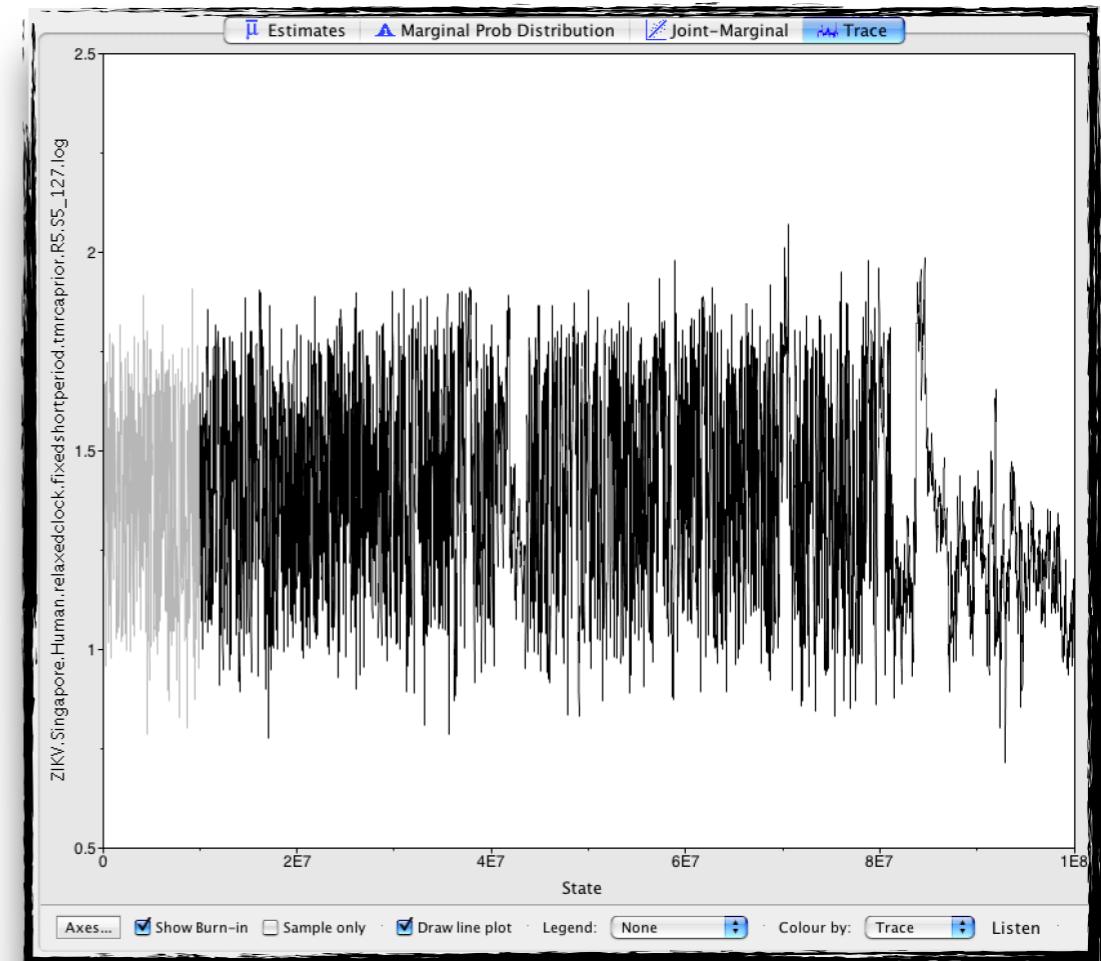
**Not converged!** 😓

**Solution:** Run for longer

# Questions to ask...

---

- Are we there yet?
- How do we know if the chain is long enough?



## Solution

- Run multiple chains
- Combine chains
- Check that all chains give the same result

**Still not converged!** 😞

The background of the slide features a wide-angle landscape of a mountain range. In the foreground, dark green forested slopes lead down to a bright blue lake. The middle ground shows more mountain ridges, some with patches of snow. The background consists of a range of mountains under a sky filled with white and grey clouds.

What are priors and  
why do they matter?

$$P(\text{hypothesis} \mid \text{data}) = \frac{P(\text{data} \mid \text{hypothesis})P(\text{hypothesis})}{P(\text{data})}$$

### Prior → $P(\text{hypothesis})$

- Have some degree of belief in our hypothesis
- From external sources, previous analyses etc.
- All parameters have priors, whether you specify them or not!

### Likelihood → $P(\text{data} \mid \text{hypothesis})$

- Likelihood is proportional to the probability of observing the data given a hypothesis

### Posterior → $P(\text{hypothesis} \mid \text{data})$

- Updated probability for the model parameters in light of the data

### Model evidence → $P(\text{data}) = P(\text{data} \mid \text{model})$

- Probability for data given model (any combination of parameters)
- Not generally used for parameter inference
- Used for Bayesian model selection

# What is a **prior** probability?

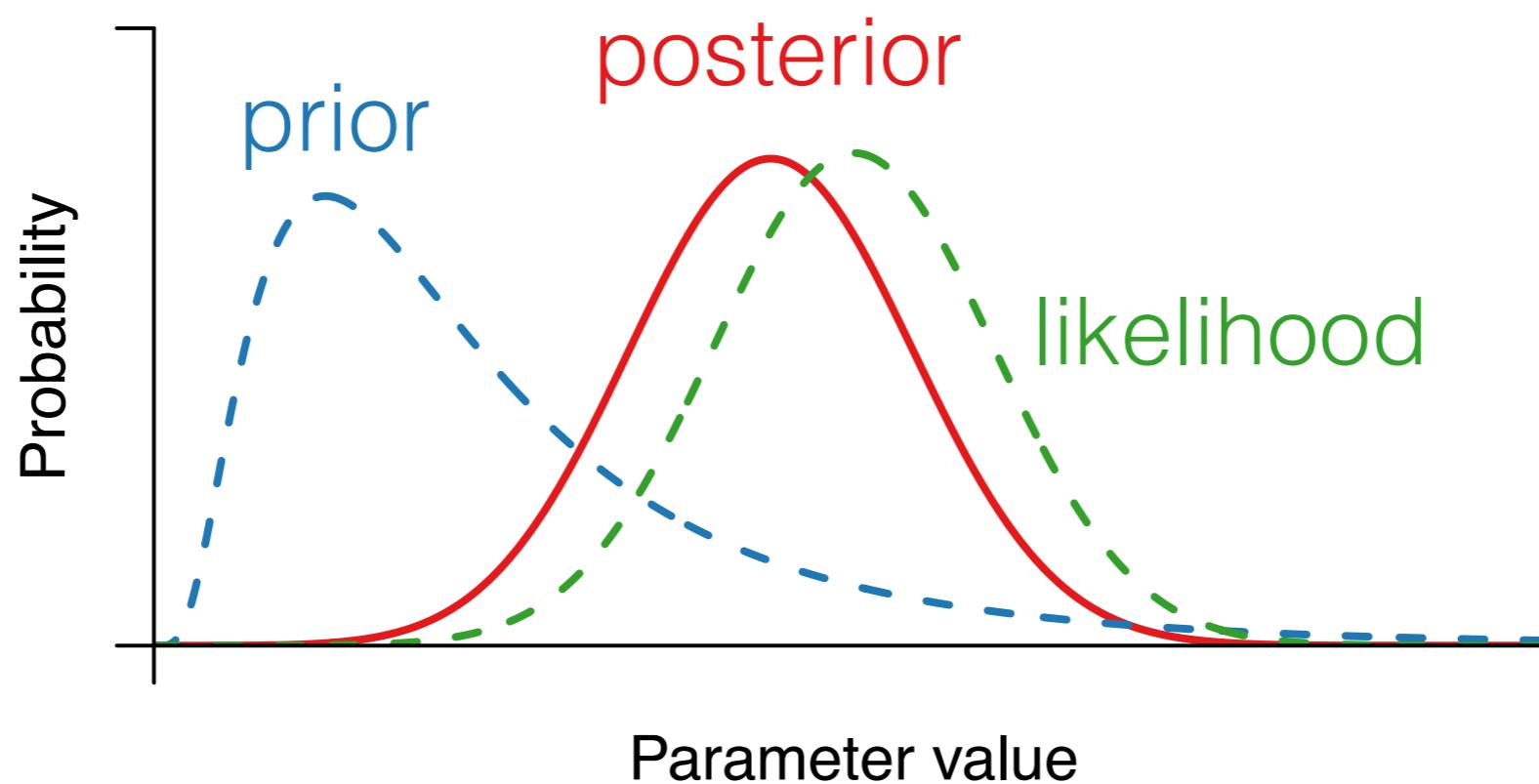
---

- The prior quantifies what you knew *prior* to taking the data into account, i.e. prior belief in your hypothesis
- This can include both:
  - Knowledge about specific model parameters (e.g. biologically realistic clock rates).
  - Aspects of the model (e.g. which clock model to use)
- When people talk about priors, they are usually referring to prior probability distributions over model parameters.
- Treatment of the model itself as part of the prior becomes important in the context of model selection.
- Priors can have priors (hyperpriors), which in turn can have priors, *ad infinitum*.
- All parameters have priors!  
(Even if you're not explicitly aware of them...)

# What is the effect of the prior? (i)

---

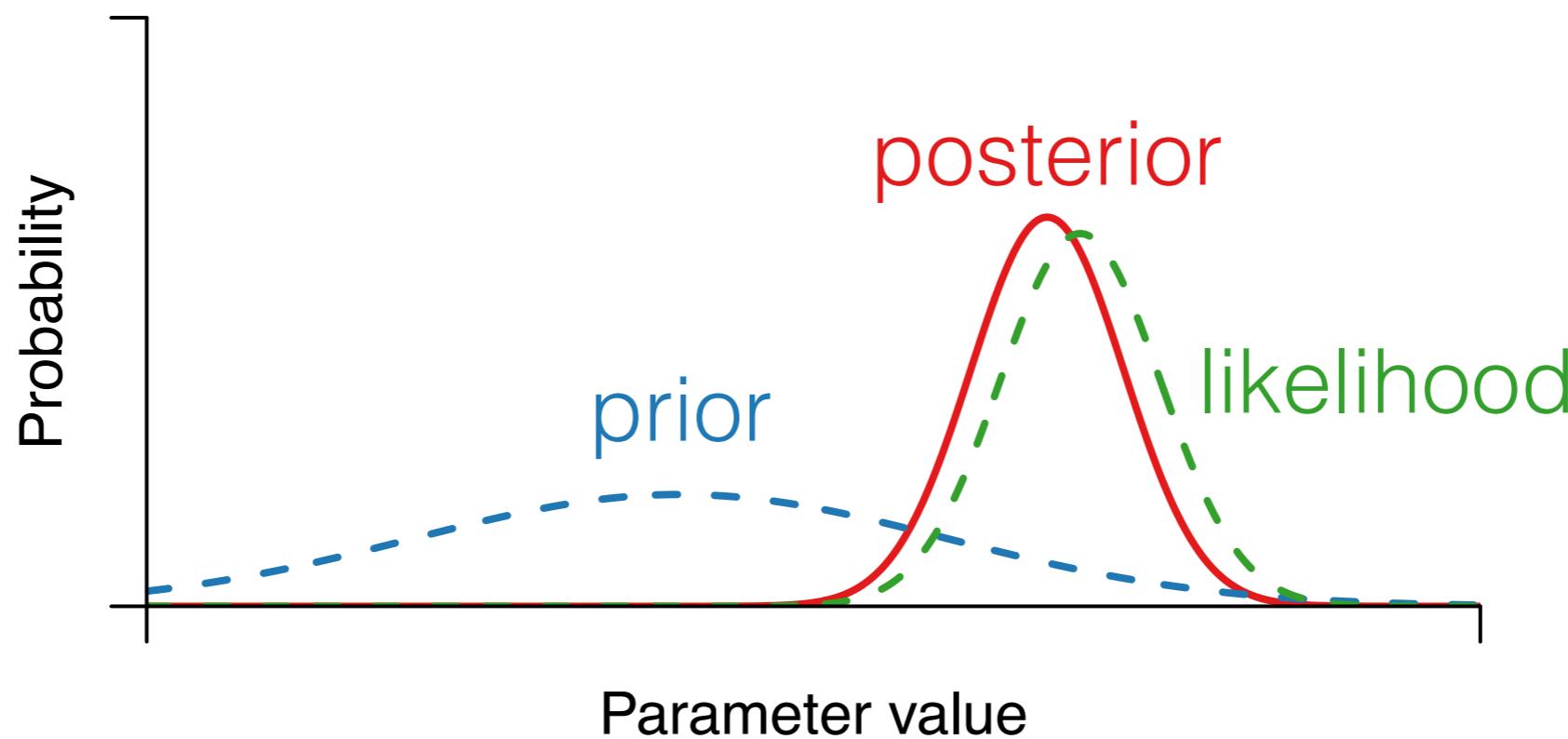
Prior distribution is combined (updated) with the likelihood to yield the **posterior distribution**



# What is the effect of the prior? (ii)

---

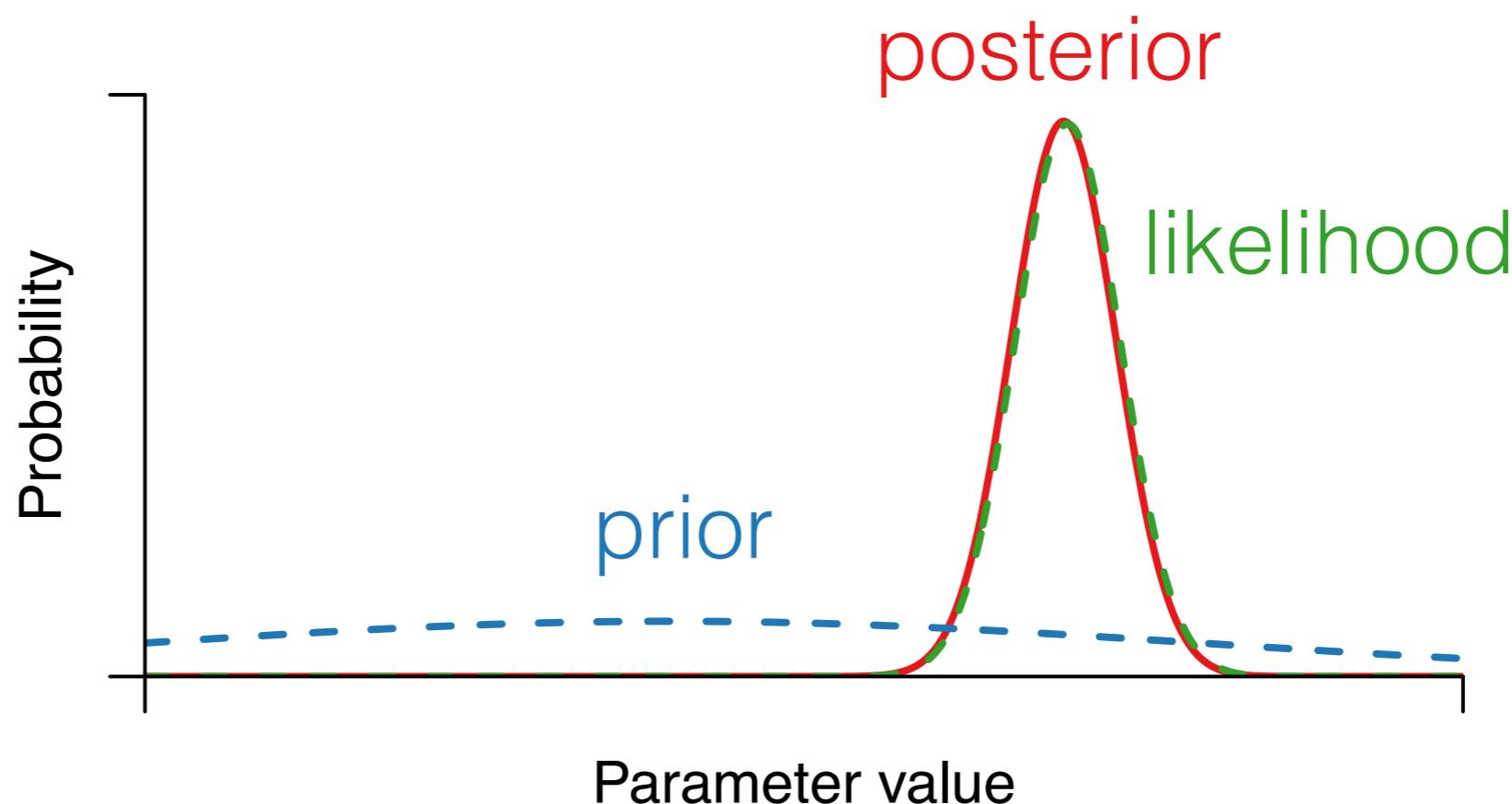
The more informative the data, the less effect the prior has



# What is the effect of the prior? (iii)

---

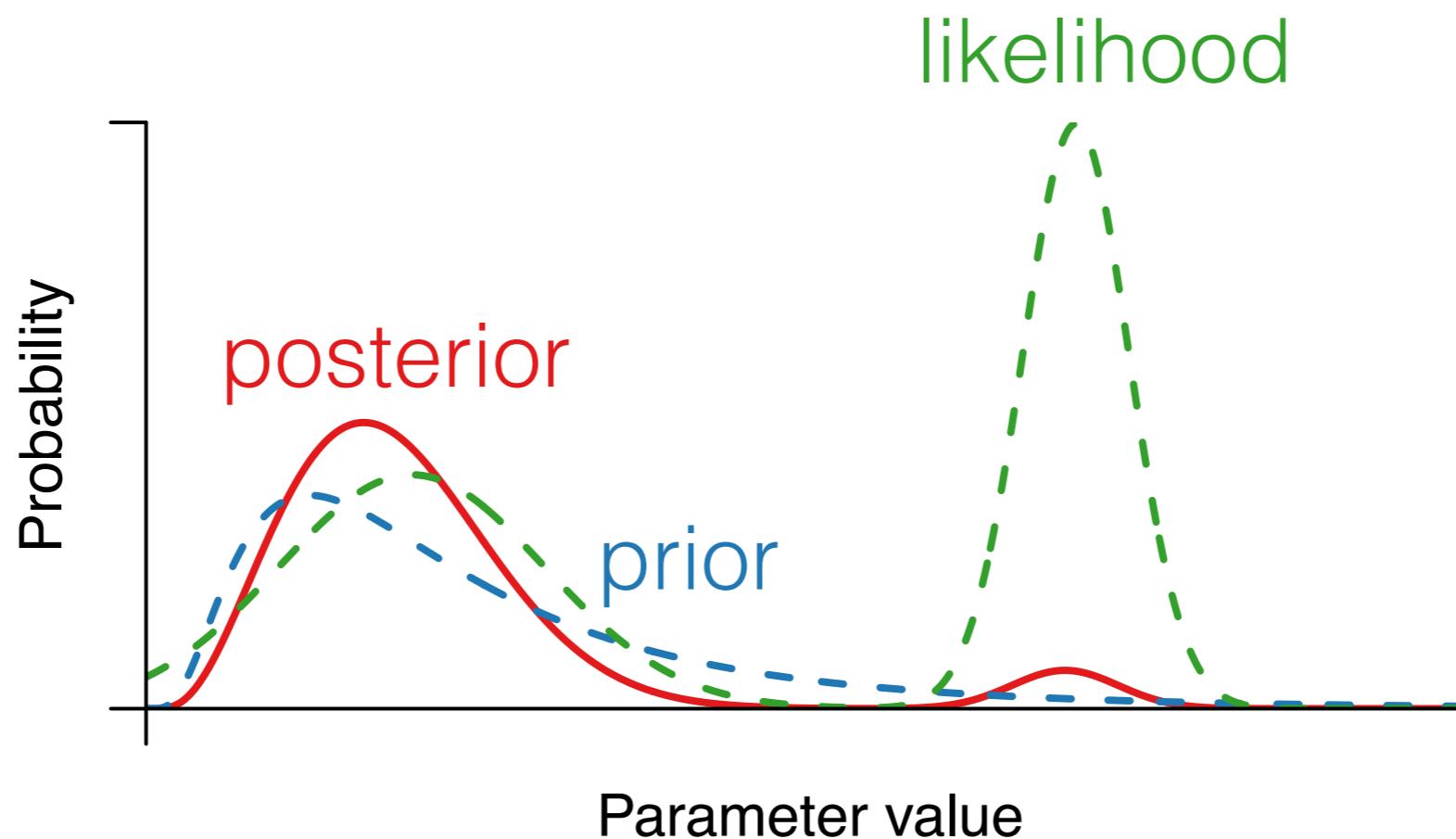
If the prior is very uninformative and the likelihood very informative the posterior will be almost exactly the likelihood



# What is the effect of the prior? (iv)

---

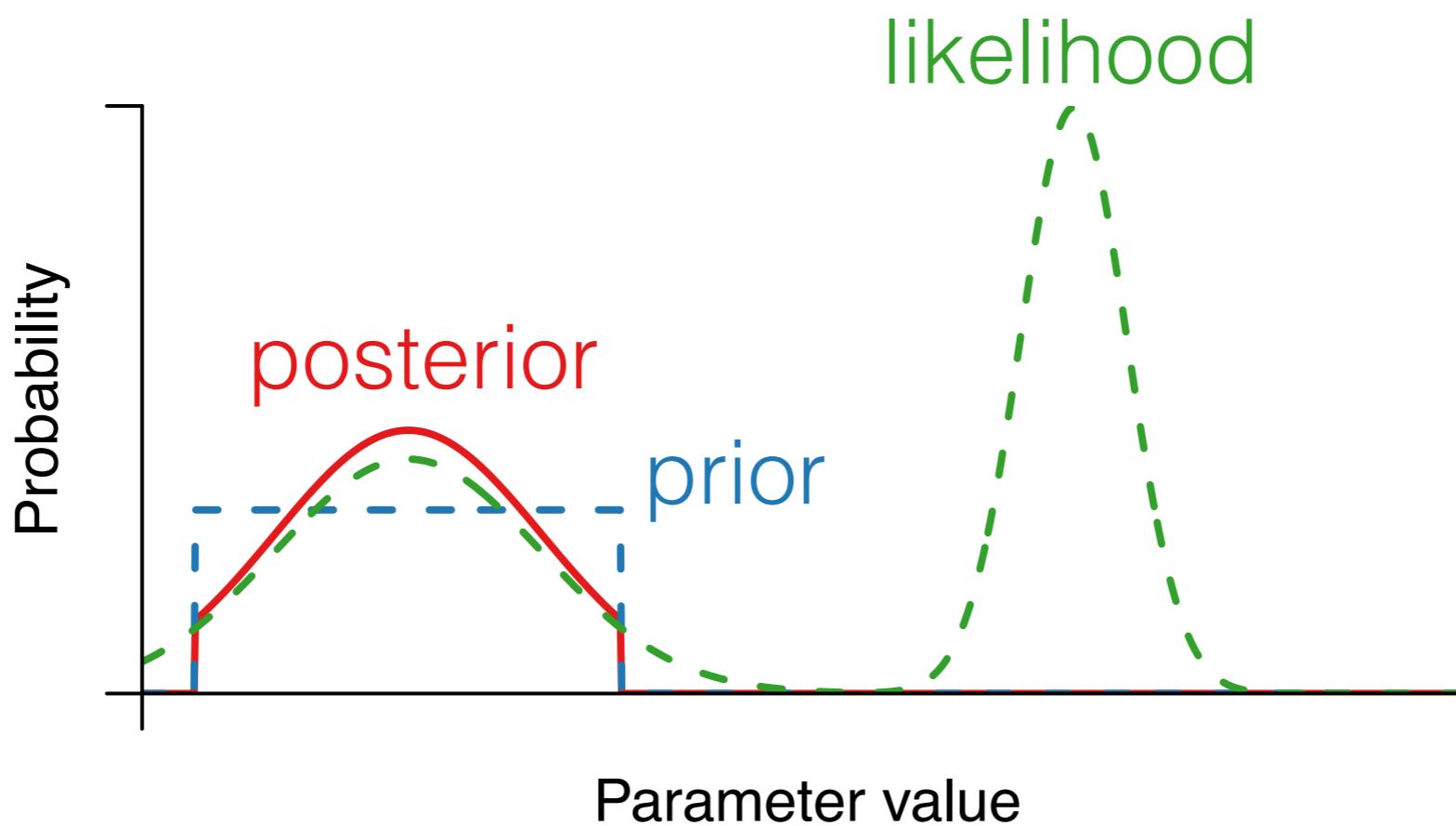
If the likelihood is multimodal the posterior is usually also multimodal even when the prior is informative



# What is the effect of the prior? (v)

---

Can use the prior to restrict the posterior to specific ranges,  
e.g. biologically realistic values



# How to pick a **prior**?

---

- Every prior distribution should be chosen with the particular analysis in mind (no priors are universal).
- This is because we must incorporate relevant prior knowledge into each distribution, including:
  - what we have learned from previous (independent) experiments or evidence (data),
  - constraints imposed by expert knowledge.
- Important to avoid making untenable assumptions (e.g. using a prior which is zero for parameter values outside of some range when there is no basis for this assumption.)
- Extremely important that the prior be chosen without recourse to the data to be analyzed.



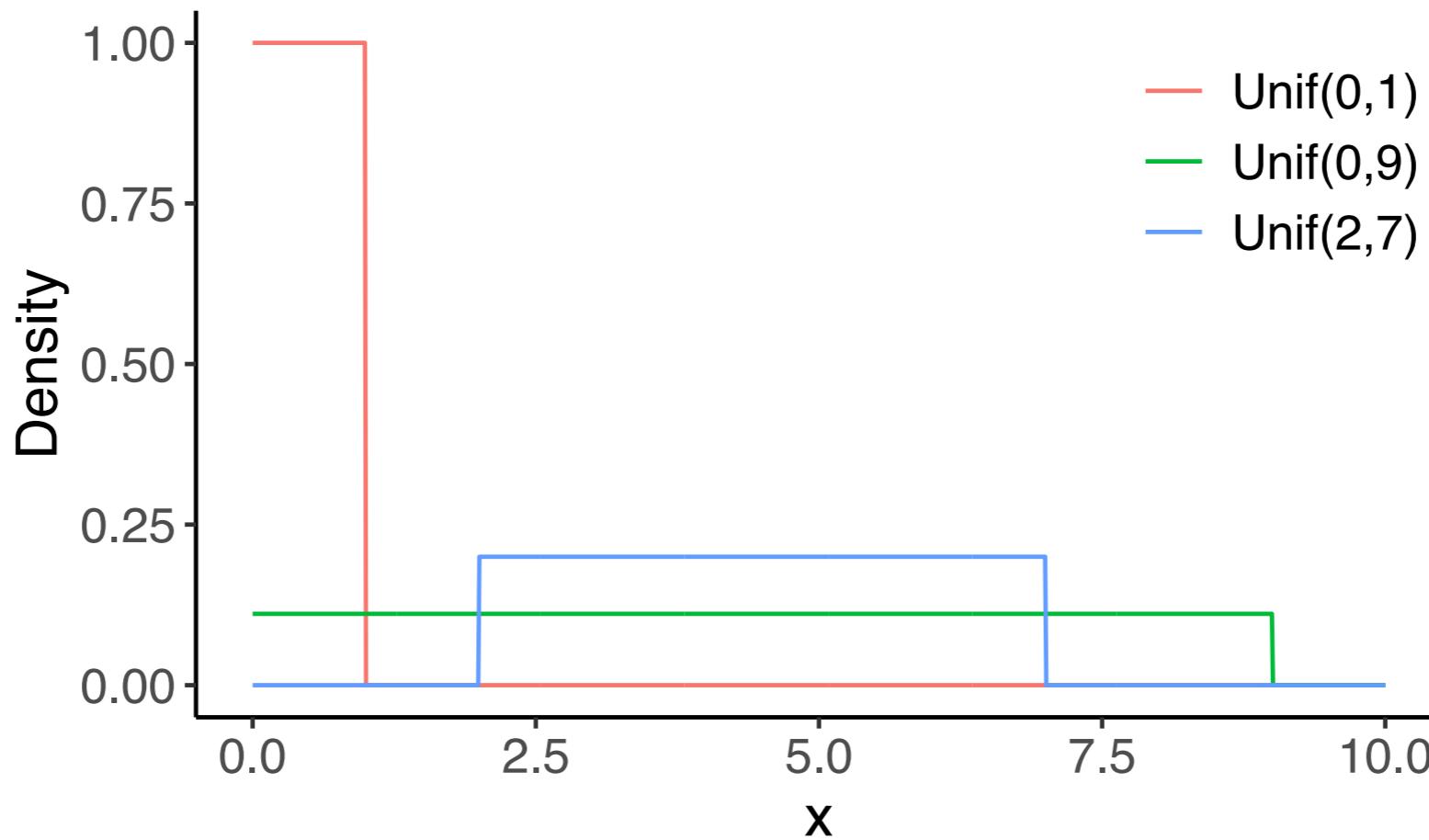
# Parametric priors in **Bayesian phylodynamics**

---

- Usually use parametric prior distributions on all continuous model parameters in a phylodynamic analysis:
  - clock rates
  - substitution model parameters  
(e.g. transition/transversion rate ratio)
  - phylodynamic model parameters  
(birth rates / effective population sizes)
  - ...
- Most of these priors are specified using members of a few well-known univariate probability densities
- Parameter bounds are part of the prior  
(e.g. normal distribution with lower bound 0)

# Uniform distribution

---



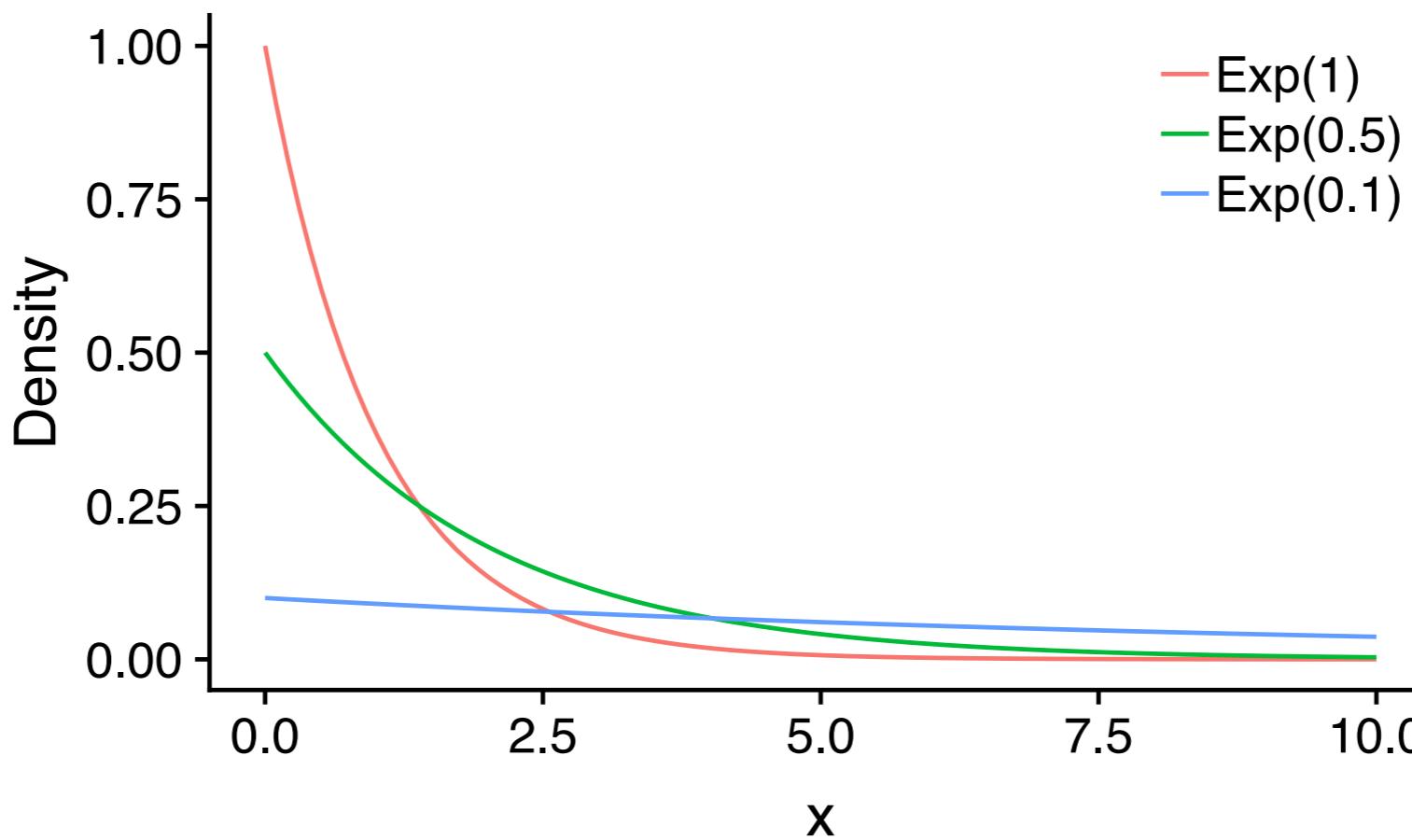
- Probability density function for **Uniform( $x_{\min}$ ,  $x_{\max}$ )**

$$f(x) = \begin{cases} \frac{1}{x_{\max} - x_{\min}} & \text{if } x_{\min} < x < x_{\max}, \\ 0 & \text{otherwise} \end{cases}$$

- Used to restrict parameter value to definite bounds

# Exponential distribution

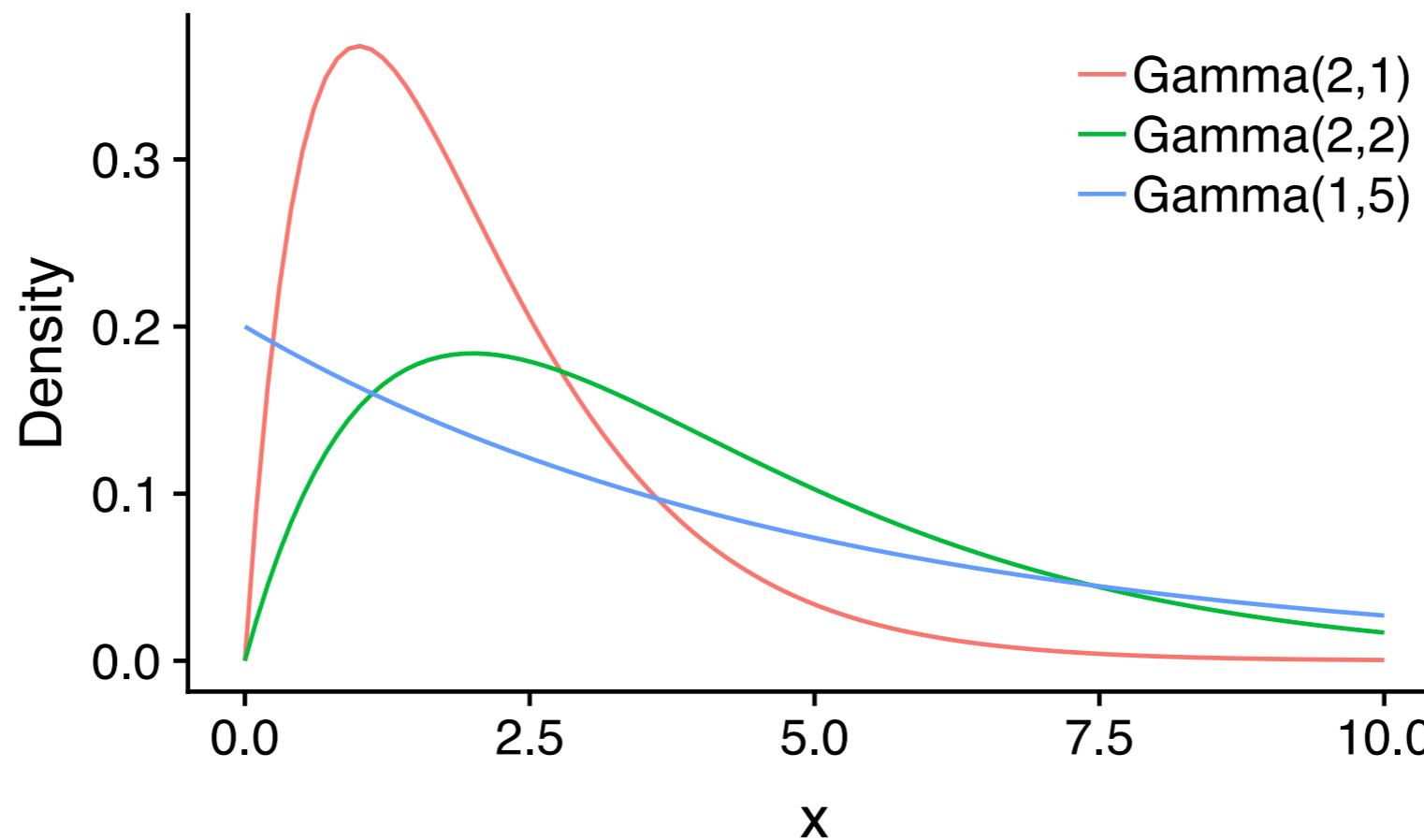
---



- Probability density function for **Exp(r)** with  $x \geq 0$ 
$$f(x) = e^{-rx}r$$
- Can be quite informative and favours parameter values closer to 0

# Gamma distribution

---



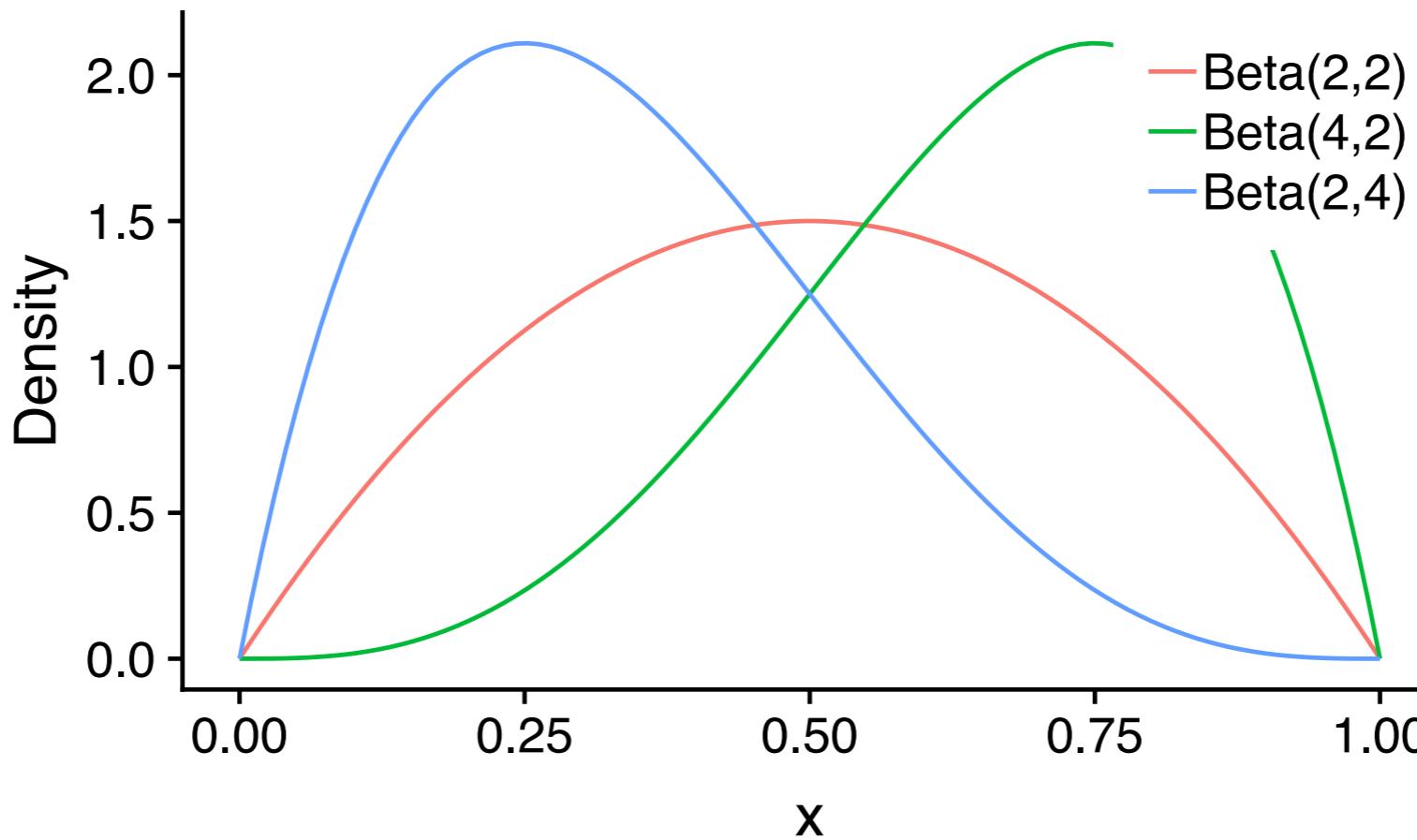
- Probability density function for **Gamma( $\alpha, \beta$ )** with  $x \geq 0$  where  $\alpha$  and  $\beta$  are the “shape” and “scale” parameters

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}$$

- A flexible generalisation of the exponential distribution

# Beta distribution

---



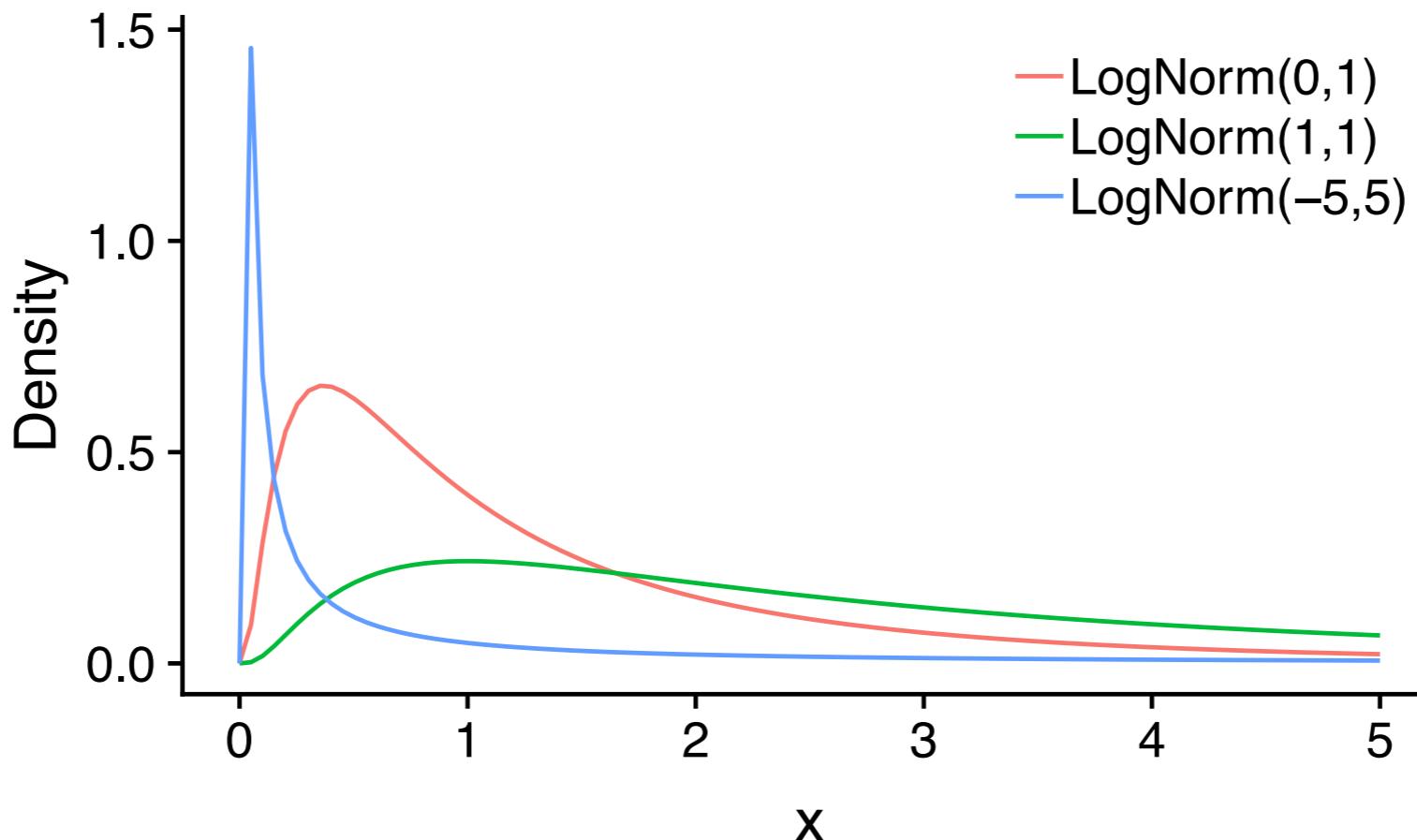
- Probability density function for **Beta(a,b)** with **0 ≤ x ≤ 1**

$$f(x) = \frac{\Gamma(a + b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1 - x)^{b-1}$$

- Flexible prior for quantities defined on the **[0,1]** interval  
(e.g. probabilities)

# Log-normal distribution

---



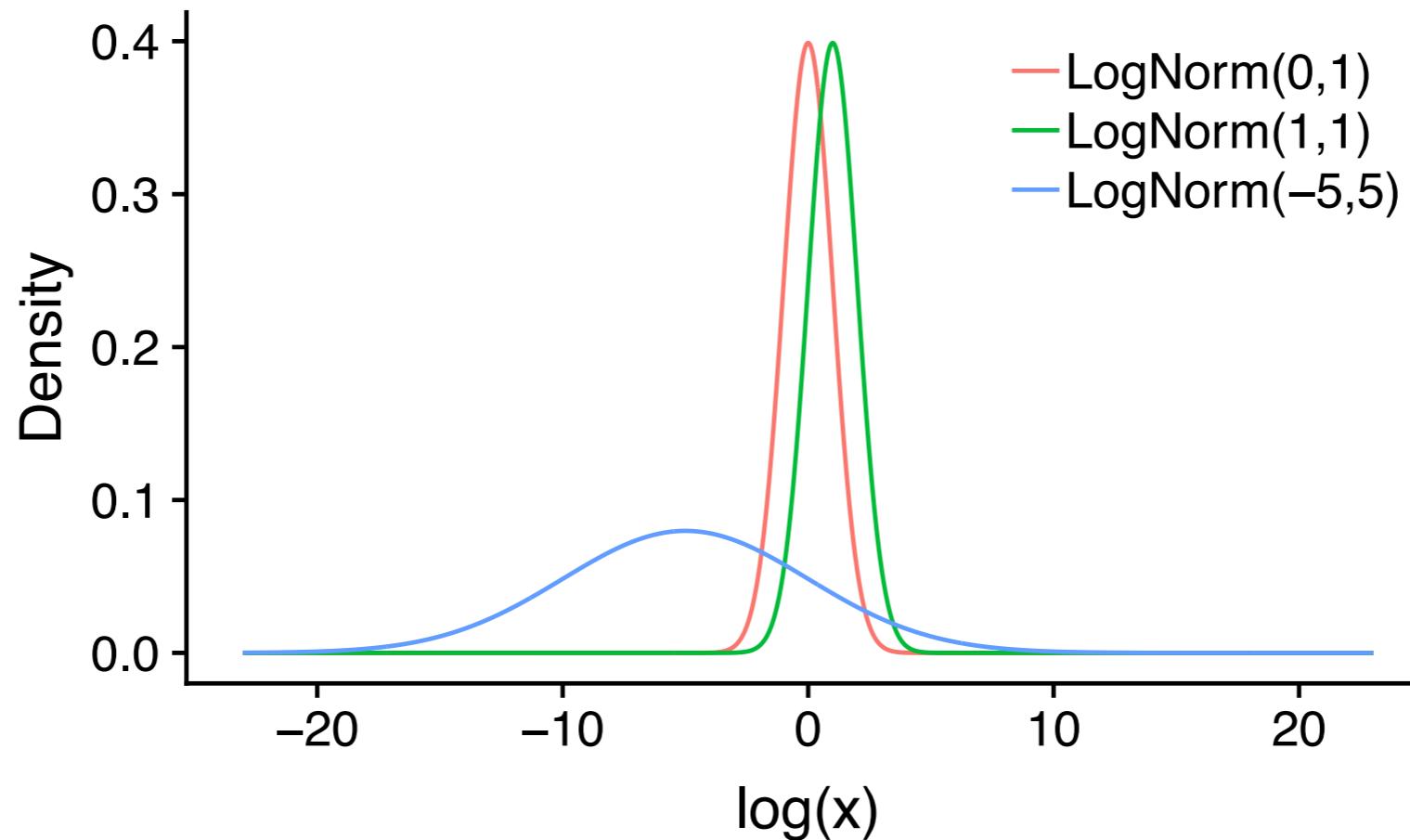
- Equivalent of a normal distribution on the logarithm of the parameter
- Probability density function for **LogNorm(M,S)** with  $x \geq 0$

$$f(x) = \frac{1}{\sqrt{2\pi}Sx} e^{-(\log(x)-M)^2/(2S^2)}$$

- **log(M)** is the median and **S** determines the standard deviation in log space

# Log-normal distribution

---



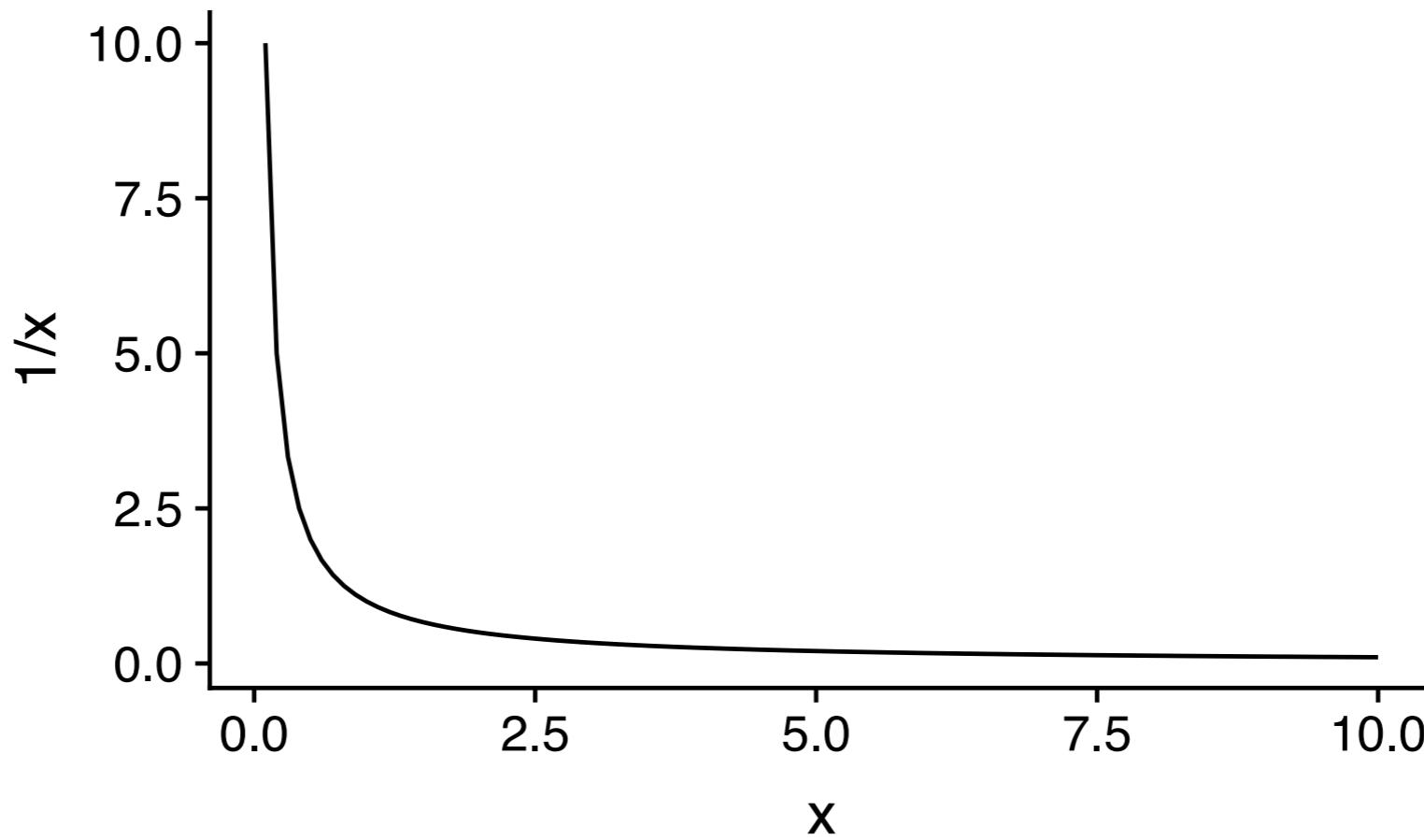
- Equivalent of a normal distribution on the logarithm of the parameter
- Probability density function for **LogNorm(M,S)** with  $x \geq 0$

$$f(x) = \frac{1}{\sqrt{2\pi}Sx} e^{-(\log(x)-M)^2/(2S^2)}$$

- **log(M)** is the median and **S** determines the standard deviation in log space

# 1/X distribution

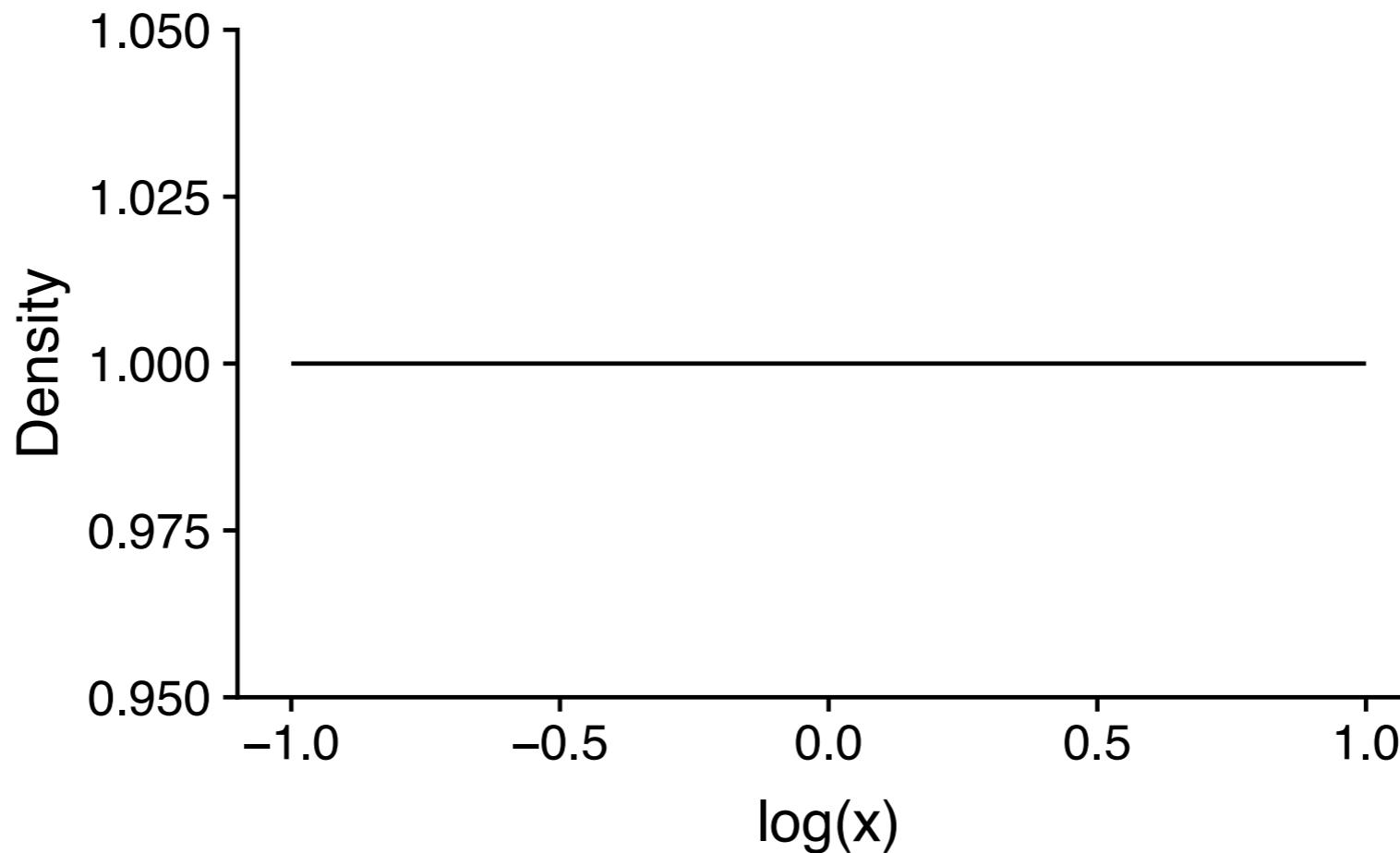
---



- Equivalent of a uniform distribution on the logarithm of the parameter
- Without bounds this distribution cannot be normalised:  
$$\frac{1}{k} \int_0^\infty \frac{dx}{x} = \infty$$
 for all possible normalising constants k

# 1/X distribution

---

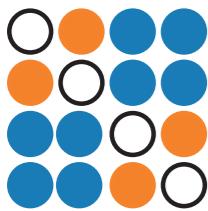


- Equivalent of a uniform distribution on the logarithm of the parameter
- Without bounds this distribution cannot be normalised:  
$$\frac{1}{k} \int_0^\infty \frac{dx}{x} = \infty$$
 for all possible normalising constants k
- May be a naturally uninformative prior for rate parameters:  
uncertainty equally distributed amongst orders of magnitude

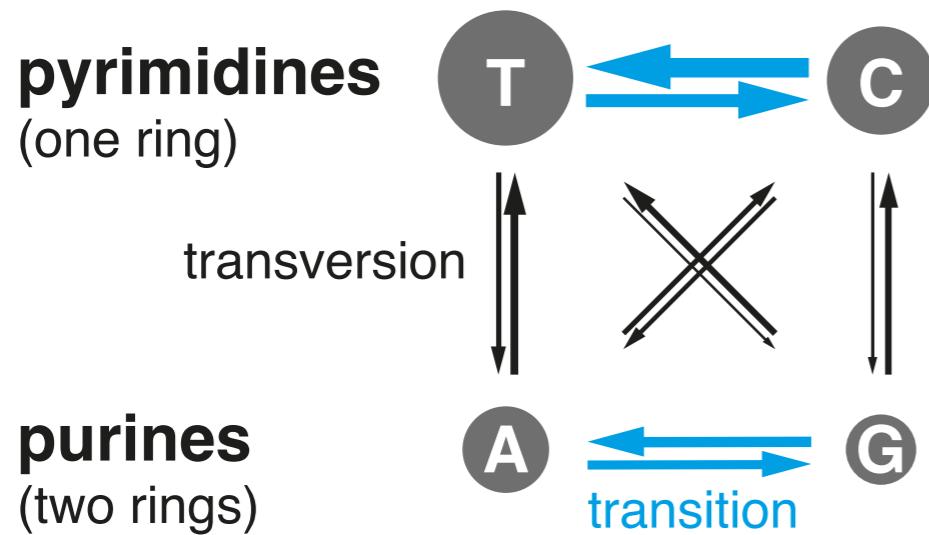
# Proper and improper priors

---

- Any true probability distribution can be used as a “proper” prior.  
(Integrates to one)
- Occasionally practitioners use priors that are not true normalizable probability distributions, e.g. Uniform( $0, \infty$ ) or a  $1/X$  prior with no upper bound. Such distributions are called “improper” priors.
- When using an improper prior, there is always a danger that the resulting posterior will also become improper.
- You should always avoid using improper priors in practice: provided you understand the model parameter you should always be able to place some loose bounds on the values it can reasonably take.



# Example 1: HKY-model (HKY85)

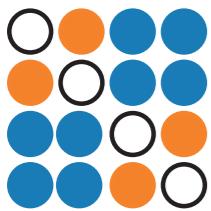


$$(\pi_T, \pi_C, \pi_A, \pi_G)$$

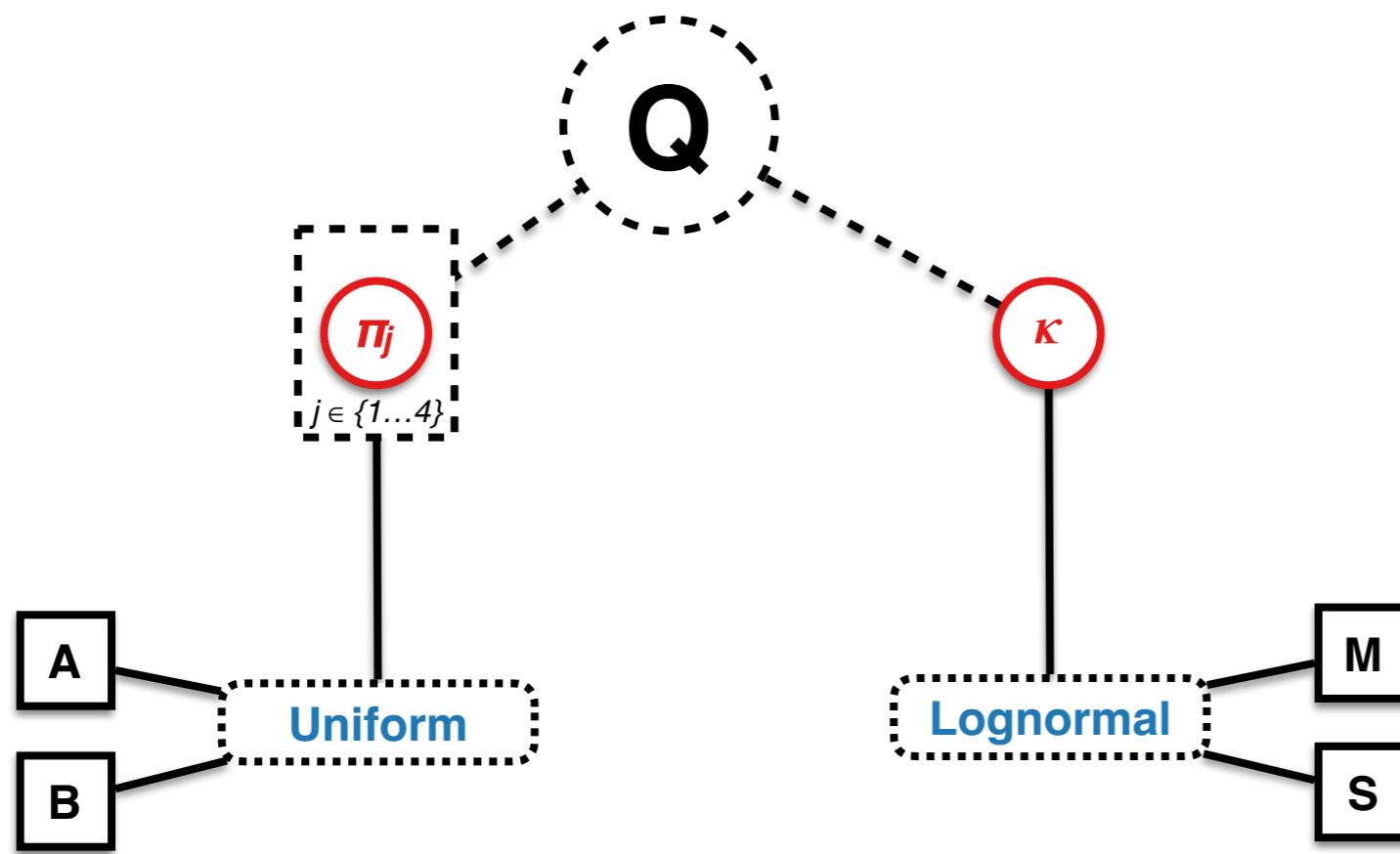
$$\begin{matrix}
 & \text{T} & \text{C} & \text{A} & \text{G} \\
 \text{T} & \cdot & \alpha\pi_C & \beta\pi_A & \beta\pi_G \\
 \text{C} & \alpha\pi_T & \cdot & \beta\pi_A & \beta\pi_G \\
 \text{A} & \beta\pi_T & \beta\pi_C & \cdot & \alpha\pi_G \\
 \text{G} & \beta\pi_T & \beta\pi_C & \alpha\pi_A & \cdot
 \end{matrix}$$

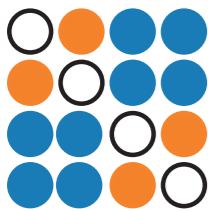
$$= \begin{pmatrix} \cdot & \alpha & \beta & \beta \\ \alpha & \cdot & \beta & \beta \\ \beta & \beta & \cdot & \alpha \\ \beta & \beta & \alpha & \cdot \end{pmatrix} \cdot \begin{pmatrix} \pi_T & 0 & 0 & 0 \\ 0 & \pi_C & 0 & 0 \\ 0 & 0 & \pi_A & 0 \\ 0 & 0 & 0 & \pi_G \end{pmatrix}$$

- **Q** matrix gives relative rates of substitution between nucleotides
- 5 parameters:
  - $\kappa = \alpha/\beta$
  - $\pi_T, \pi_C, \pi_A, \pi_G$  — assume these all have the same prior probability

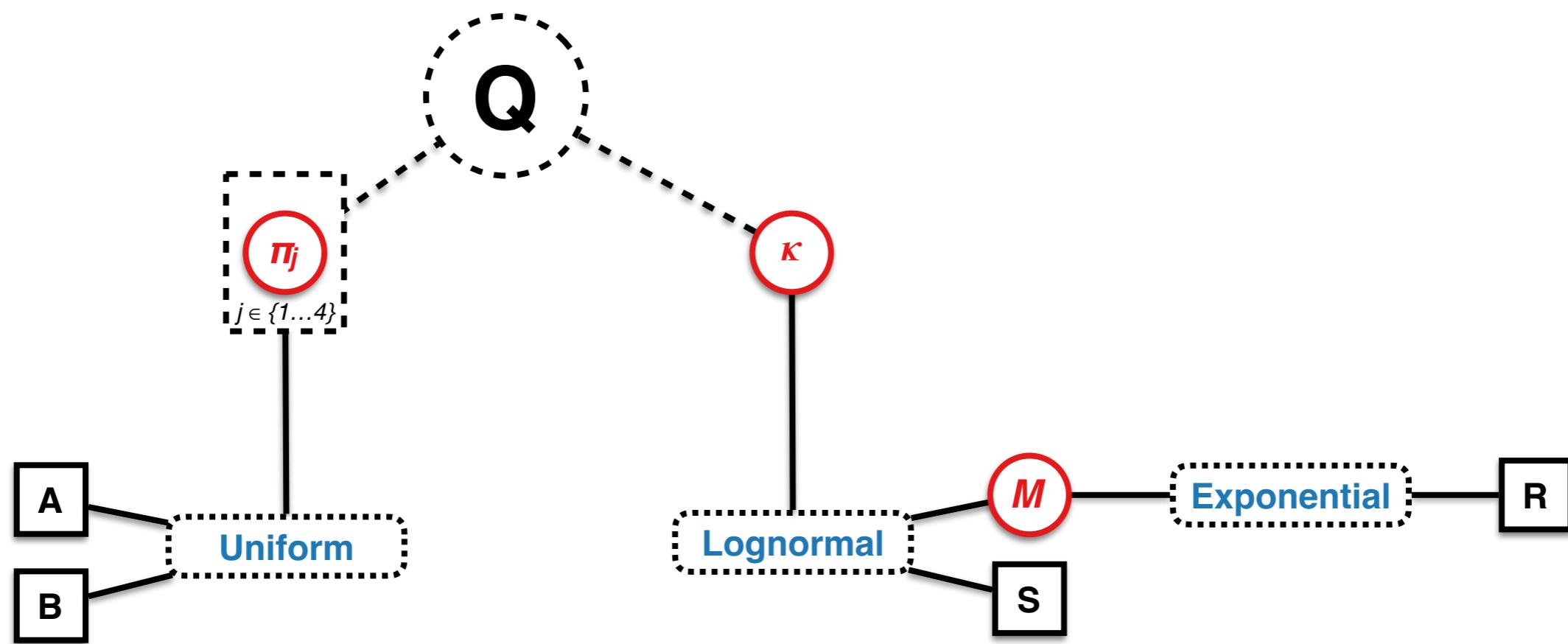


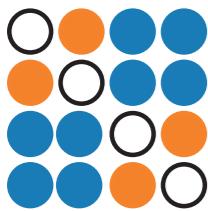
# Example 1: HKY-model (HKY85)



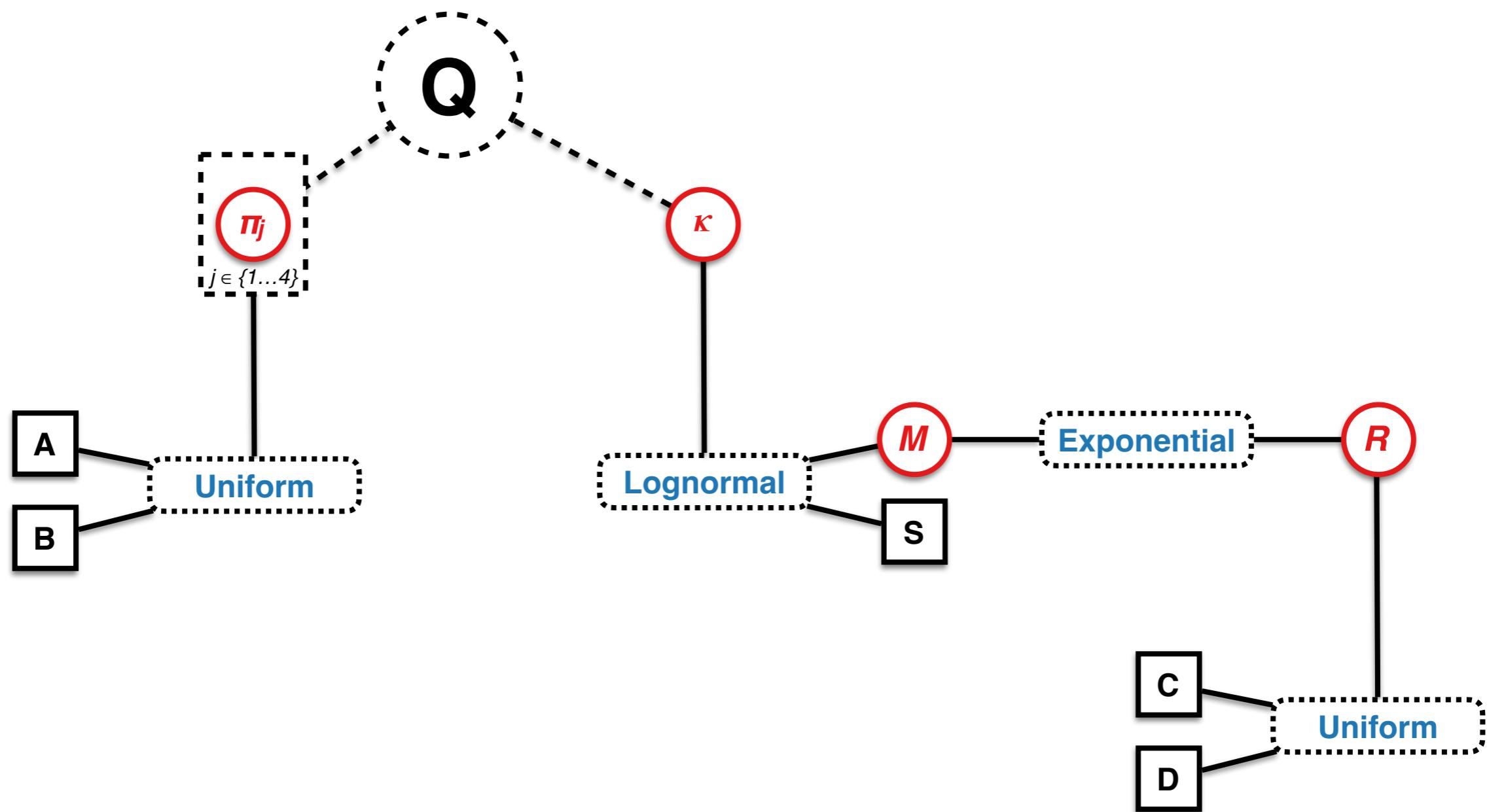


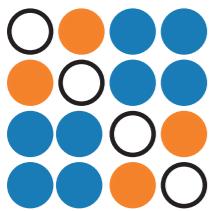
# Example 1: HKY-model (HKY85)



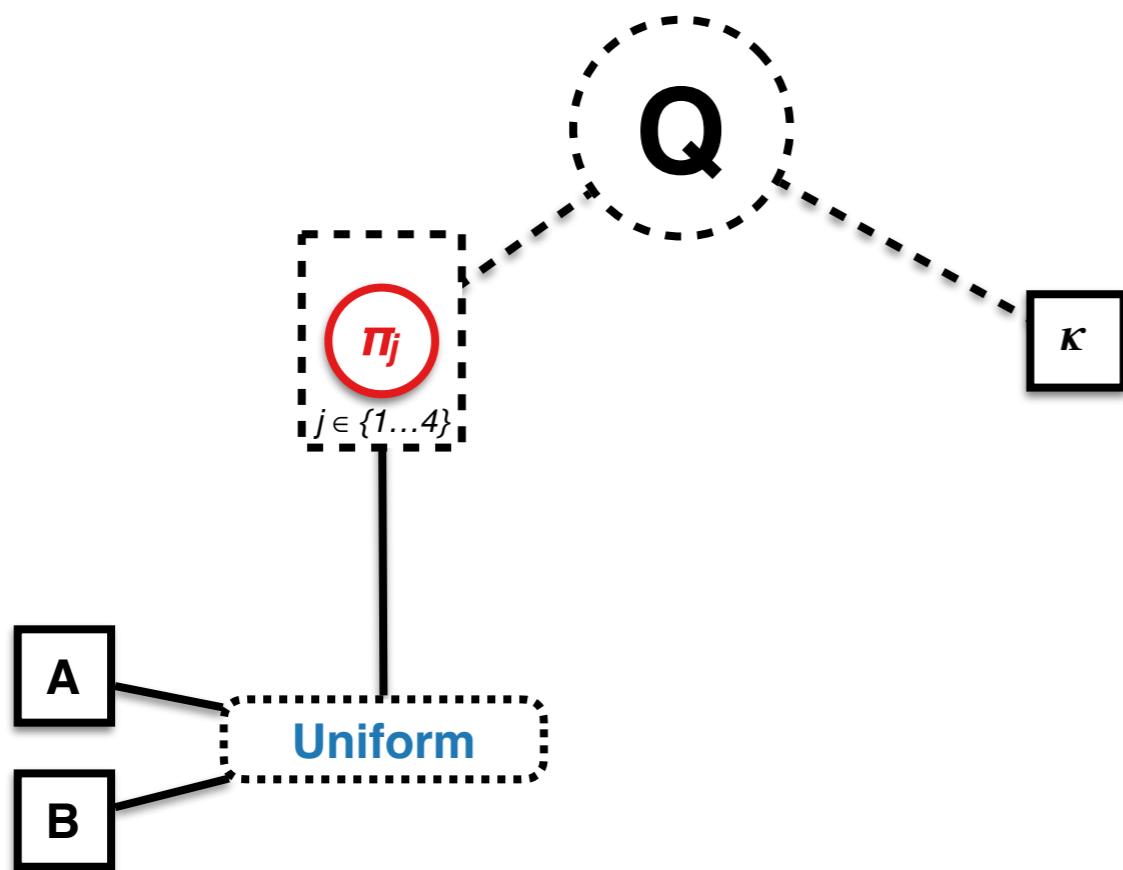


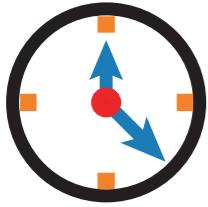
# Example 1: HKY-model (HKY85)





# Example 1: HKY-model (HKY85)





## Example 2: Molecular clock rate

---

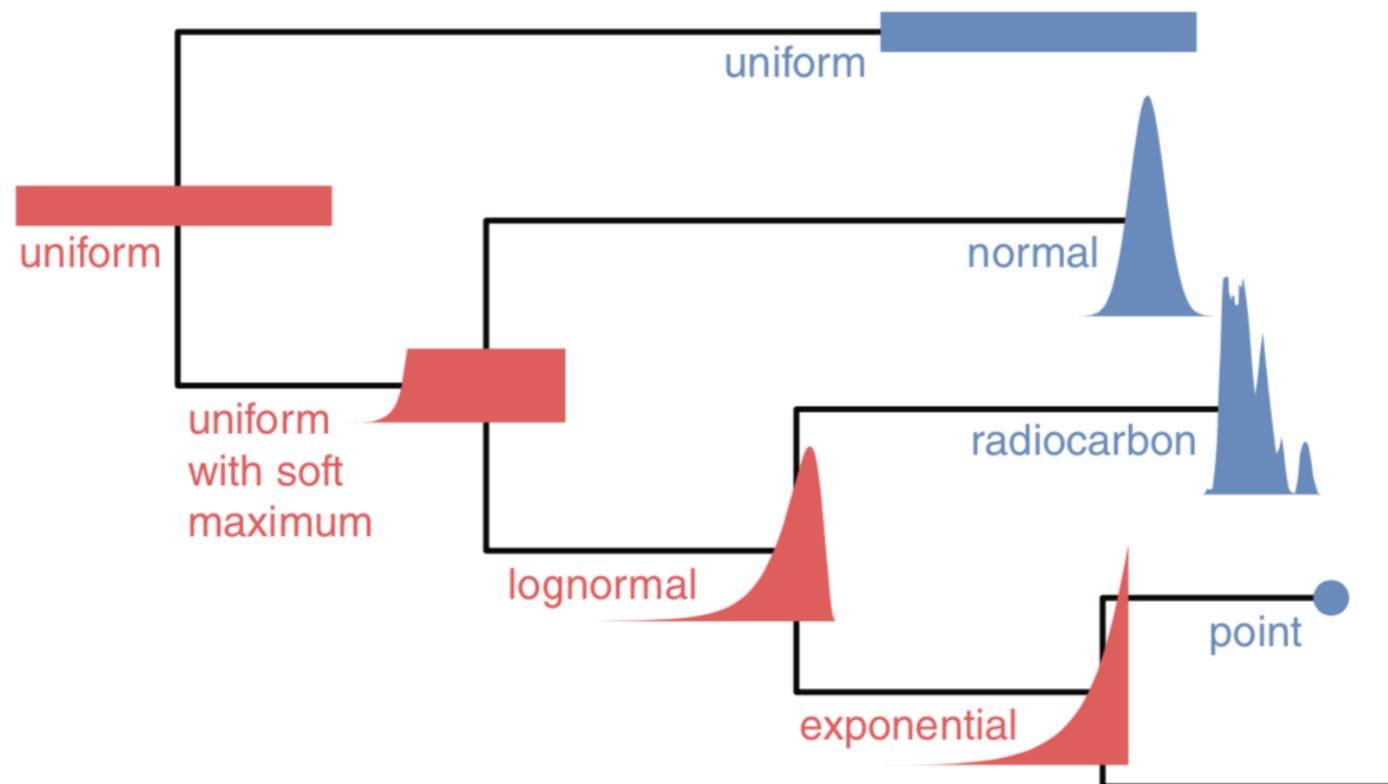
Suppose we want to pick a prior for the molecular clock rate,  $\mu$ , and we want to use an uninformative prior

- What about a uniform prior (constant)?  $f(\mu) = c$

### **Is this uninformative?**

- But this is defined from 0 to infinity!
- Places almost all probability on very large (and biologically unrealistic) values!
- There is no normalising constant such that it integrates to 1!
- $f(\mu) = 1/\mu$  is a better choice for rates (uniform in log space)
  - But it's also impossible to find a normalising constant!
- Probably better to impose an upper bound or use a proper prior like a lognormal distribution

# Calibration nodes



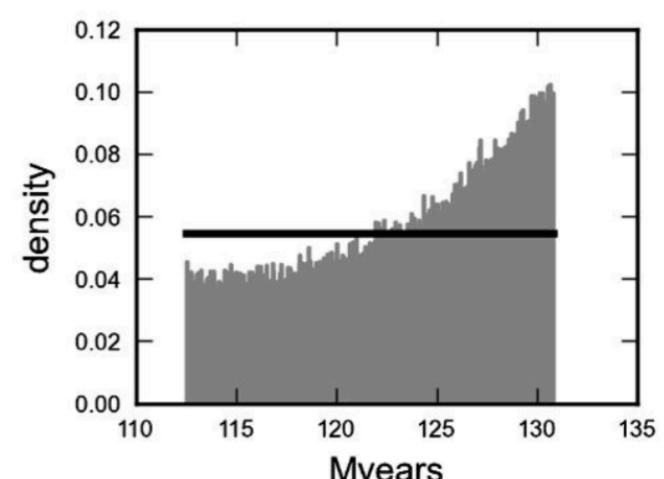
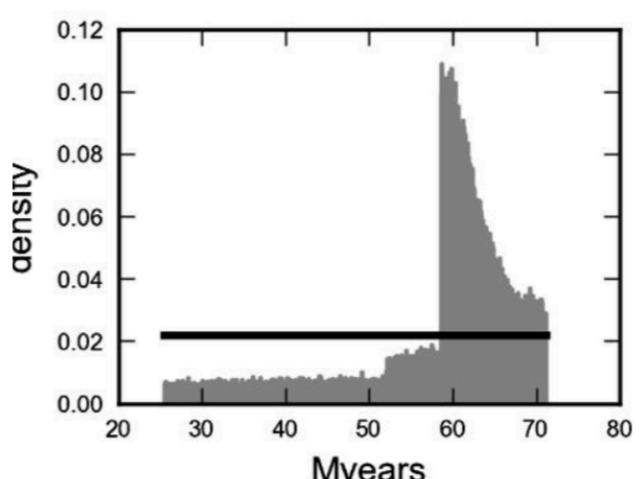
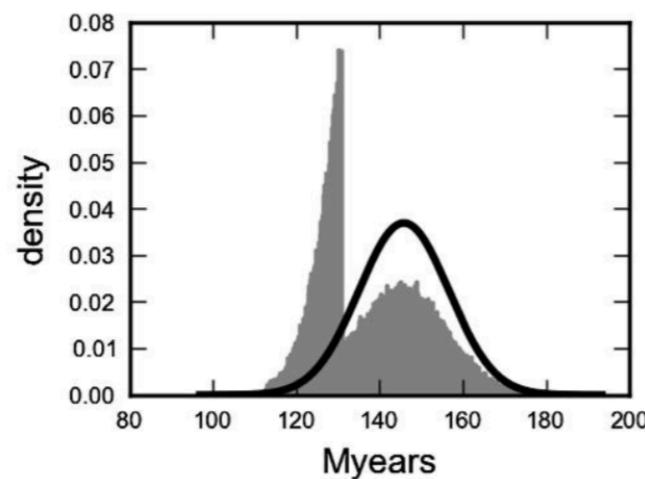
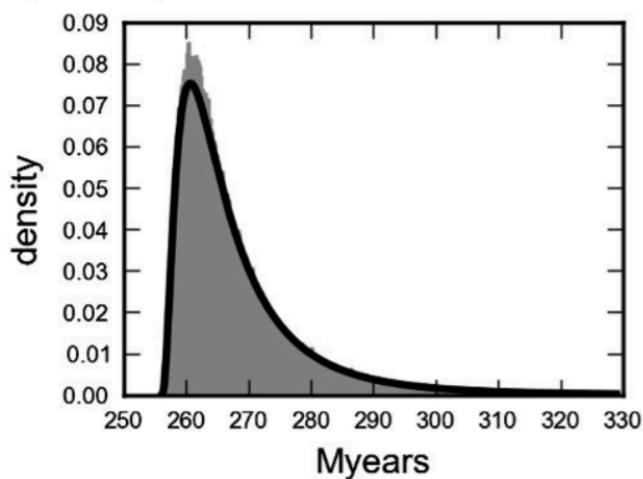
- May have additional information about divergence dates or occurrence times:
  - e.g. from fossils, radiocarbon dating or the timing of geographic events
- Homochronous trees **need** external information to be scaled to calendar time:
  - e.g. fix the clock rate or use a calibration node

# Calibration nodes

---

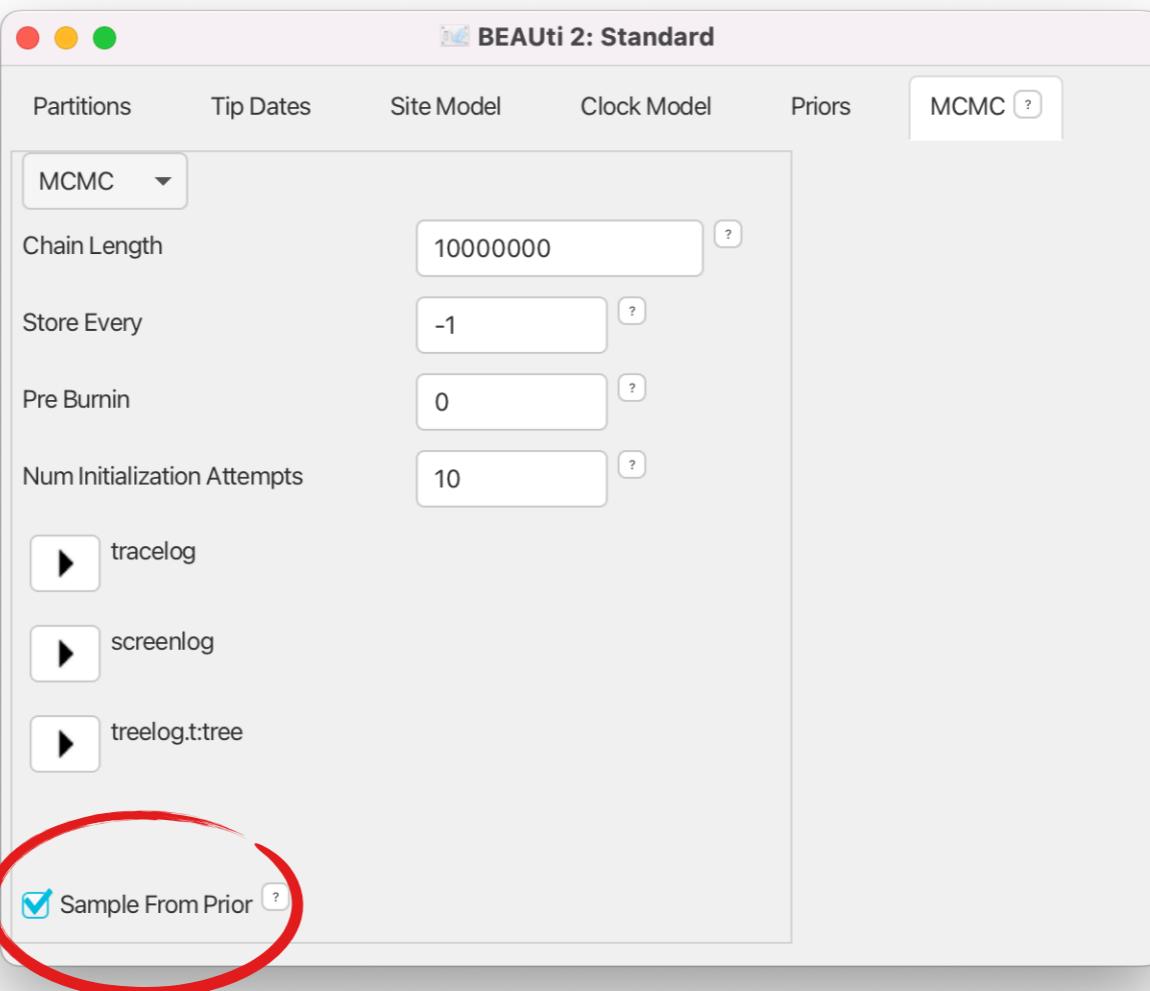
More calibration nodes mean more information to calibrate the clock

but is more always better?



Induced prior may be different to the prior you set!

# Setting priors best practice

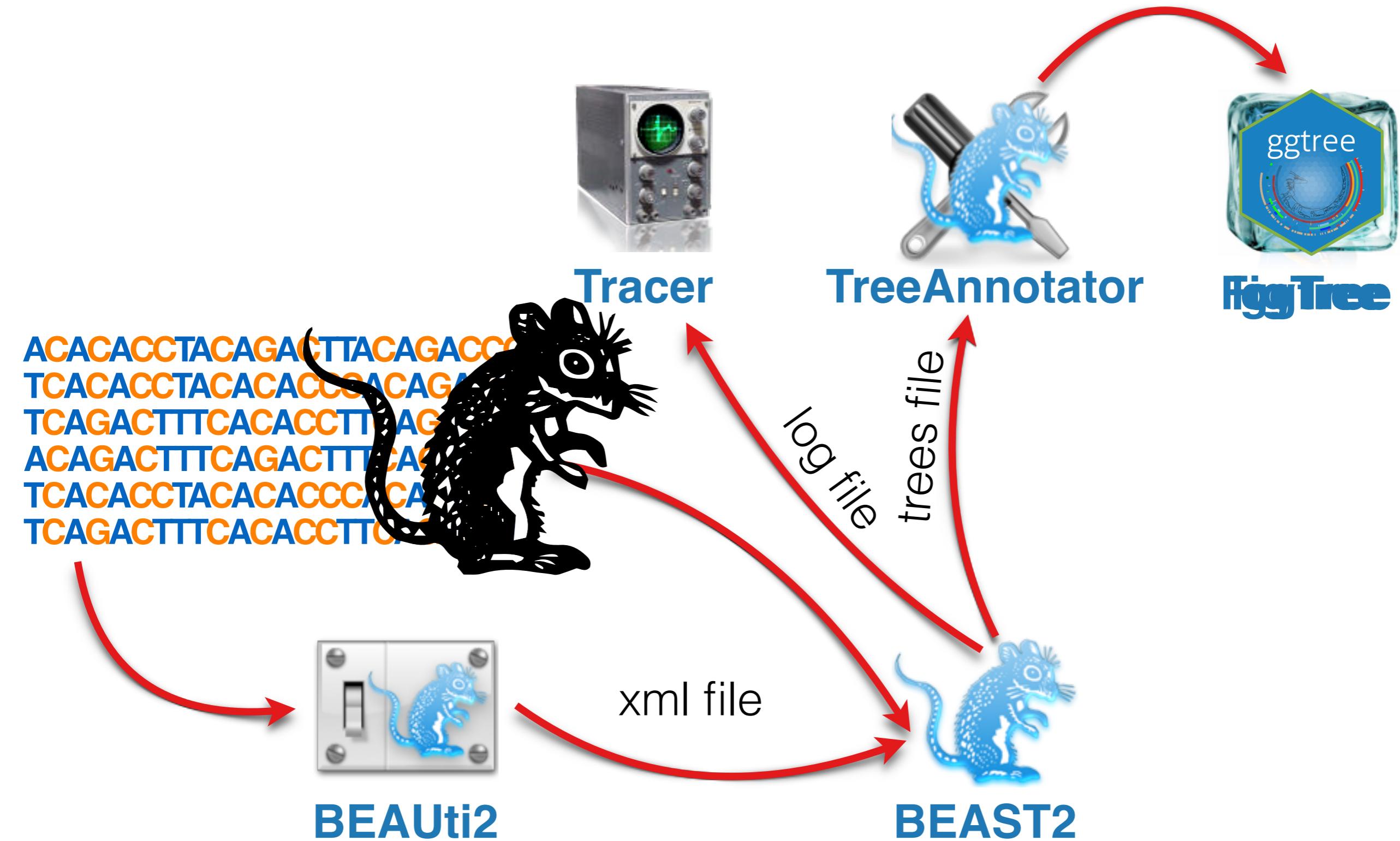


- Sample or simulate from prior
- For key parameters plot prior and posterior together (are the data informative?)
- Mess with priors and see how sensitive/robust posterior is to prior perturbations

# FANTASTIC BEASTS

AND WHERE  
TO FIND THEM

# BEAST2 workflow



# BEAUti2

(<http://beast2.org>)



Graphical tool for setting up a BEAST2 analysis

## Input:

- Genetic sequence data
- Optional:
  - Sampling times
  - Sampling locations
  - Traits
  - etc.

## Output:

- Compact XML description of data, model and prior distributions that can be run in BEAST2

B

(ht)

**BEAUti 2: Standard**

Partitions Tip Dates Site Model Clock Model Priors MCMC

Tree.t:tree Calibrated Yule Model

birthRateY.t:tree Gamma[0.001,1000.0] initial = [1.0] [0.0, $\infty$ ] Calibrated Yule speciation process birth rate for t:noncoding

clockRate.c:clock Uniform[0.0,Infinity] initial = [1.0] [0.0, $\infty$ ]

gammaShape.s:1stpos Exponential[1.0] initial = [1.0] [0.1, $\infty$ ] Prior on gamma shape for partition s:1stpos

gammaShape.s:2ndpos Exponential[1.0] initial = [1.0] [0.1, $\infty$ ] Prior on gamma shape for partition s:2ndpos

gammaShape.s:3rdpos Exponential[1.0] initial = [1.0] [0.1, $\infty$ ] Prior on gamma shape for partition s:3rdpos

gammaShape.s:noncoding Exponential[1.0] initial = [1.0] [0.1, $\infty$ ] Prior on gamma shape for partition s:noncoding

kappa.s:1stpos Log Normal[1.0,1.25] initial = [2.0] [0.0, $\infty$ ]

kappa.s:2ndpos Log Normal[1.0,1.25] initial = [2.0] [0.0, $\infty$ ]

kappa.s:3rdpos Log Normal[1.0,1.25] initial = [2.0] [0.0, $\infty$ ]

kappa.s:noncoding Log Normal[1.0,1.25] initial = [2.0] [0.0, $\infty$ ] HKY transition-transversion parameter of partition s:noncoding

human-chi... Normal  monophyletic

+ Add Prior

Primates\_long.xml UNREGISTERED

Primates\_long.xml

```
39
40 <run id="mcmc" spec="MCMC" chainLength="2500000">
41   <state id="state" storeEvery="5000">
42     <tree id="Tree.t:tree" name="stateNode">
43       <taxonset id="TaxonSet.noncoding" spec="TaxonSet">
44         <alignment id="noncoding" spec="FilteredAlignment" filter="1,458-659,897-898">
45           <data idref="primate-mtDNA"/>
46         </alignment>
47       </taxonset>
48     </tree>
49     <parameter id="mutationRate.s:noncoding" name="stateNode">1.0</parameter>
50     <parameter id="gammaShape.s:noncoding" name="stateNode">1.0</parameter>
51     <parameter id="kappa.s:noncoding" lower="0.0" name="stateNode">2.0</parameter>
52     <parameter id="kappa.s:1stpos" lower="0.0" name="stateNode">2.0</parameter>
53     <parameter id="gammaShape.s:1stpos" name="stateNode">1.0</parameter>
54     <parameter id="mutationRate.s:1stpos" name="stateNode">1.0</parameter>
55     <parameter id="kappa.s:2ndpos" lower="0.0" name="stateNode">2.0</parameter>
56     <parameter id="gammaShape.s:2ndpos" name="stateNode">1.0</parameter>
57     <parameter id="mutationRate.s:2ndpos" name="stateNode">1.0</parameter>
58     <parameter id="kappa.s:3rdpos" lower="0.0" name="stateNode">2.0</parameter>
59     <parameter id="gammaShape.s:3rdpos" name="stateNode">1.0</parameter>
60     <parameter id="mutationRate.s:3rdpos" name="stateNode">1.0</parameter>
61     <parameter id="birthRateY.t:tree" name="stateNode">1.0</parameter>
62     <parameter id="clockRate.c:clock" name="stateNode">1.0</parameter>
63   </state>
64
65   <init id="RandomTree.t:tree" spec="beast.evolution.tree.RandomTree" estimate="false" initial="@Tree.t:tree" taxa=
@noncoding>
66     <populationModel id="ConstantPopulation0.t:tree" spec="ConstantPopulation">
67       <parameter id="randomPopSize.t:tree" name="popSize">1.0</parameter>
68     </populationModel>
69   </init>
70
71   <distribution id="posterior" spec="util.CompoundDistribution">
72     <distribution id="prior" spec="util.CompoundDistribution">
73       <distribution id="CalibratedYuleModel.t:tree" spec="beast.evolution.speciation.CalibratedYuleModel"
birthRate="@birthRateY.t:tree" tree="@Tree.t:tree"/>
74       <prior id="CalibratedYuleBirthRatePrior.t:tree" name="distribution" x="@birthRateY.t:tree">
75         <Gamma id="Gamma.0" name="distr">
76           <parameter id="RealParameter.0" estimate="false" name="alpha">0.001</parameter>
77           <parameter id="RealParameter.01" estimate="false" name="beta">1000.0</parameter>
78         </Gamma>
79       </prior>
80       <prior id="ClockPrior.c:clock" name="distribution" x="@clockRate.c:clock">
81         <Uniform id="Uniform.0" name="distr" upper="Infinity"/>
82       </prior>
83     </distribution>
84   </distribution>
85 
```

Line 1, Column 3

0 misspelled words

Spaces: 4

XML

# BEAST2

(<http://beast2.org>)



- Bayesian **e**volutionary **a**nalysis by **s**ampling **t**rees
- Performs MCMC analyses of sequences under selected sequence evolution and tree model
- Similar to BEAST 1.10 but completely separate and generally incompatible
- BEAST2 and BEAST1 have a common origin, have much of the same functionality but have diverged over time
- BEAST2 has a modular design that makes it easy to extend
- GUI interface but can also be run from command line (e.g. on a cluster)

## Input:

- xml model description file

## Outputs:

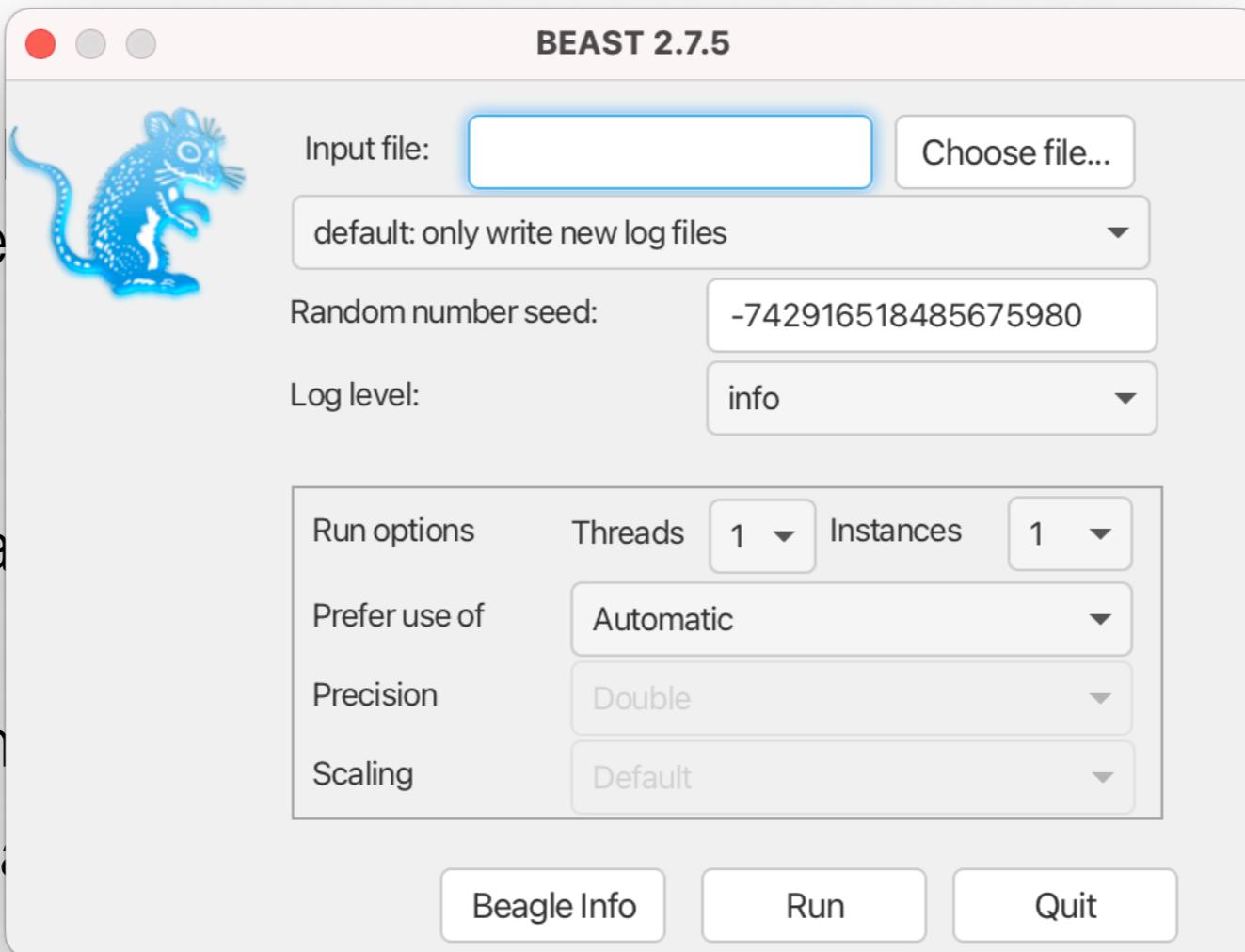
- log file
- trees file
- state file

# BEAST2

(<http://beast2.org>)



- Bayesian
- Performs sequence analysis
- Similar to incompatible
- BEAST2 also uses the same XML language
- BEAST2 has a GUI interface
- GUI interface (e.g. on a Mac OS X, Linux, or Windows)



lected  
d generally  
e much of  
e  
to extend  
line

## Input:

- xml model description file

## Outputs:

- log file
- trees file
- state file

# BEAST2 packages



- BEAST2 is organised into a central "core" together with a large number of separate packages
- Packages can be developed by anybody — **including you!**
- Can be directly integrated into BEAST2 and updated frequently without waiting for a full BEAST2 release
- Packages add new models or completely new functionality
  - phylogeography
  - bacterial ARG inference
  - morphological models
  - multispecies coalescent
  - model selection and averaging
  - stochastic simulations
  - linguistic analyses
  - ...
- Install new packages through the package manager in BEAUti

BEAST 2 Package Manager

List of available packages for BEAST v2.7.\*

Name	Installed	Latest	Dependencies	Link	
BEAST.base	2.7.5	2.7.5			BEAST base
BEAST.app	2.7.5	2.7.5	BEAST.base		BEAST base applications
Babel		0.4.2	BEAST.app, BEAST.base, BEAST...		BABEL = BEAST analysis backing effective linguistics
bacter		3.0.1	BEAST.app, feast, BEAST.base		Bacterial ARG inference.
BADTRIP		2.0.0	BEAST.base		Infer transmission time for non-haplotype data and epi data
BASTA		4.0.0	BEAST.base, BEAST.app		Bayesian structured coalescent approximation
bdmm		2.0.0	BEAST.base, MASTER, BEAST.ap...		Multitype birth-death model (aka birth-death-migration model)
<b>BDSKY</b>	<b>1.5.0</b>	<b>1.5.0</b>	<b>BEAST.base, BEAST.app</b>		<b>birth death skyline - handles serially sampled tips, piecewise cons</b>
bdtree		0.0.1	BEAST.base, BEAST.app		Birth-death sequential sampling
BEAST_CLASSIC		1.6.2	BEAST.base, BEAST.app		BEAST classes ported from BEAST 1 in wrappers
BEASTLabs	2.0.1	2.0.1	BEAST.base, BEAST.app		BEAST utilities, such as Script, multi monophyletic constraints
BEASTvntr		0.2.0	BEAST.app, BEAST.base		Variable Number of Tandem Repeat data, such as microsatellites
BICEPS		1.1.1	BEAST.base, BEAST.app		Bayesian Integrated Coalescent Epoch PlotS + Yule Skyline
bModelTest		1.3.3	BEAST.base, BEAST.app		Bayesian model test for nucleotide subst models, gamma rate het
BREAK_AWAY		1.2.0	BEAST.base, GEO_SPHERE, BEA...		break-away model of phylogeography
CA		2.1.0	BEAST.base, BEAST.app		Bayesian estimation of clade ages based on probabilities of fossil
ClaDS		2.0.3	BEAST.app, BEAST.base		Implementation of the ClaDS birth-death tree prior
CoalRe		1.0.3	BEAST.base, BEAST.app		Inference of Recombination networks
CodonSubstModels		2.0.0	BEAST.app, BEAST.base		Codon substitution models for DNA
contacTrees		1.2.0	BEASTLabs, BEAST.base, BEAST...		Phylogenetic model with horizontal transfer for linguistics
contraband		1.0.1	BEAST.base, BEAST.app		Scalable brownian models for continuous trait evolution
CoupledMCMC		1.2.1	BEAST.app, BEAST.base		Adaptive coupled MCMC (adaptive parallel tempering or MC3)
DENIM		1.1.1	BEAST.base, BEAST.app		Divergence Estimation Notwithstanding ILS and Migration

Latest

Install/Upgrade

Uninstall

Package repositories

?

Close

- Install new packages through the package manager in BEAUTi

# Tracer

(<http://beast.community>)



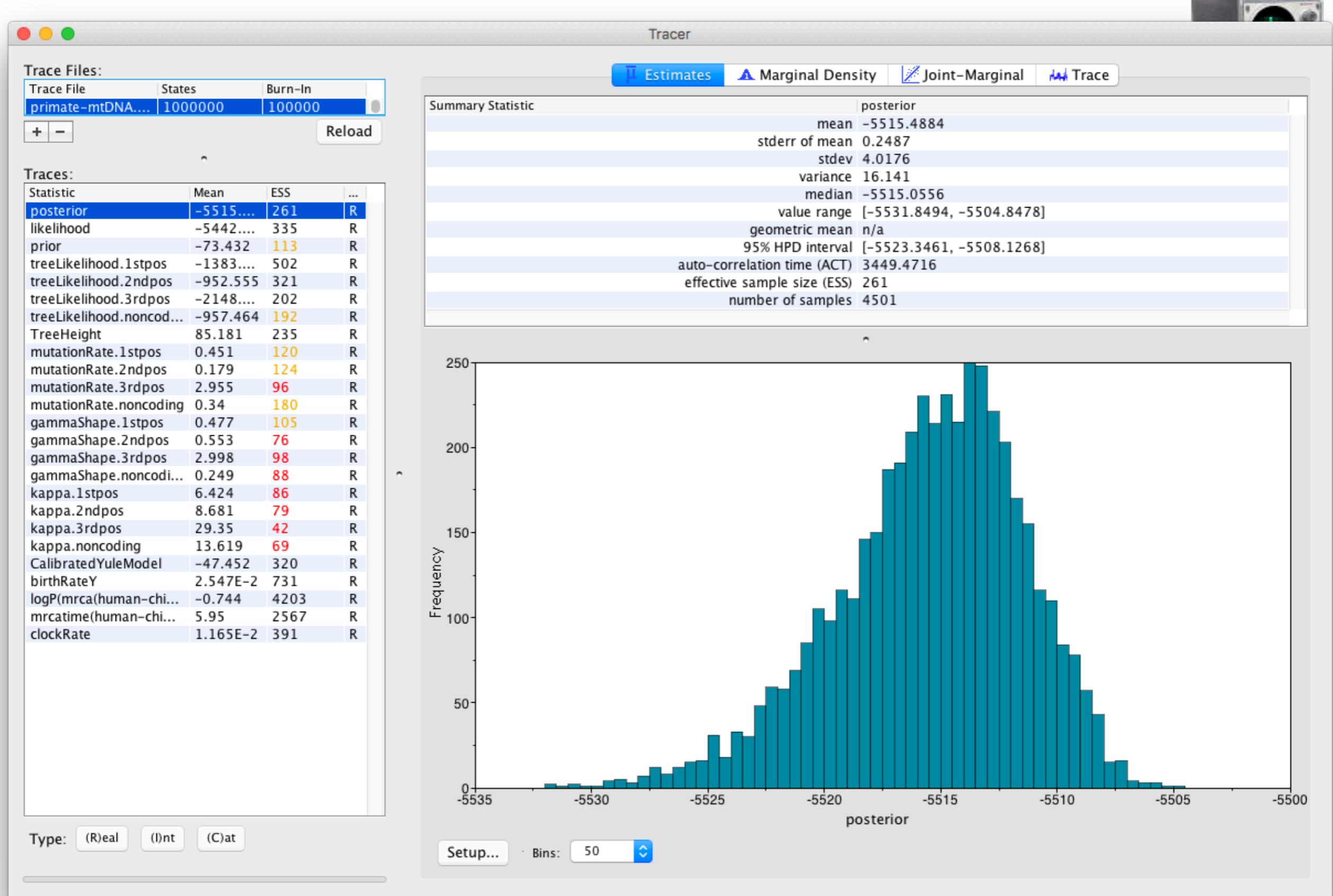
- Analyse (parameter) log files from BEAST2 runs
- Check mixing, ESS, ACT, parameter correlations
- Provides overview of posterior parameter estimates
- Easily compare several analyses
- Demographic reconstruction for some models (e.g. Bayesian Skyline Plot)
- Tracer is **primarily** a diagnostic tool — usually perform final analyses in a statistical package like R!

## Input:

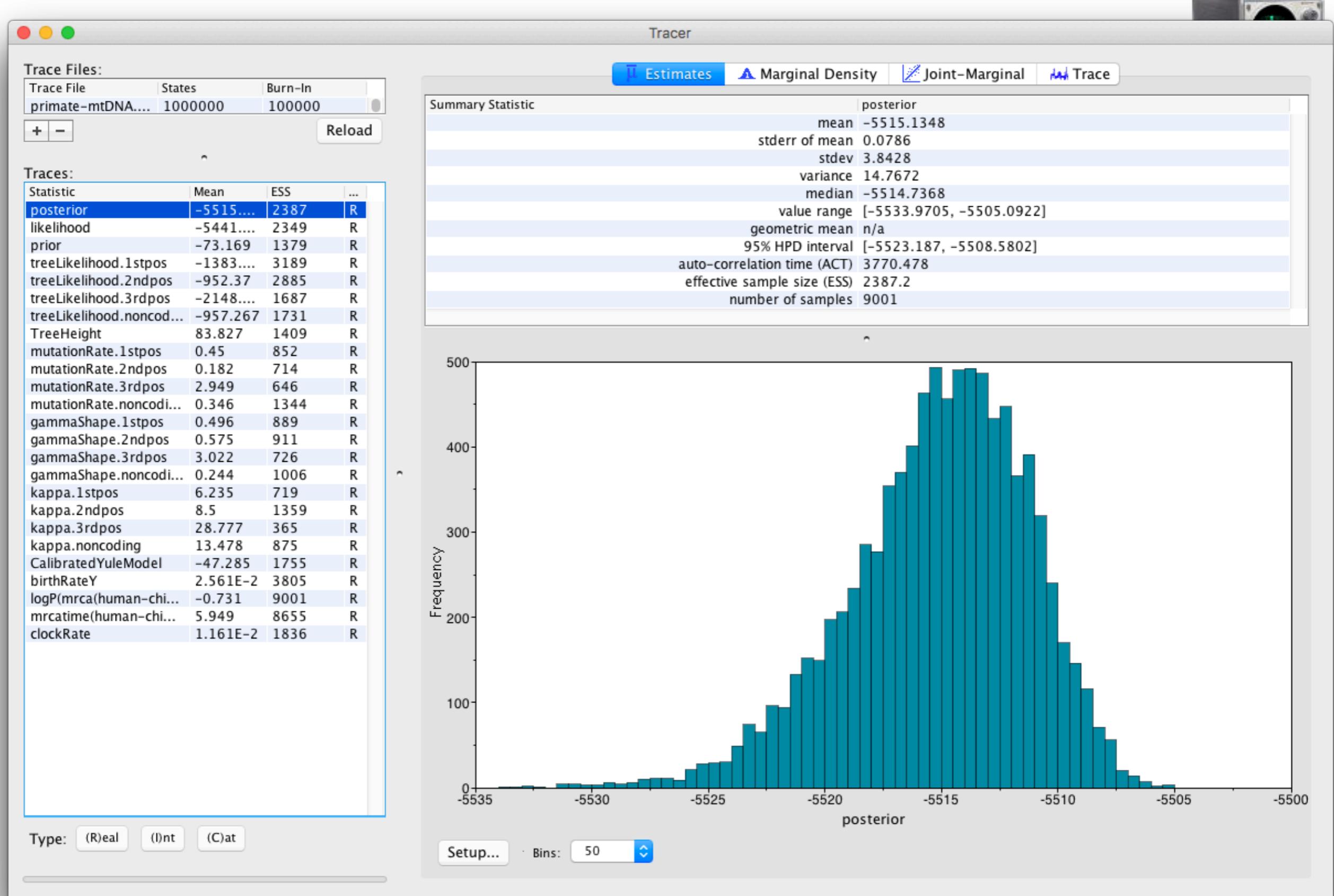
- log file

## Output:

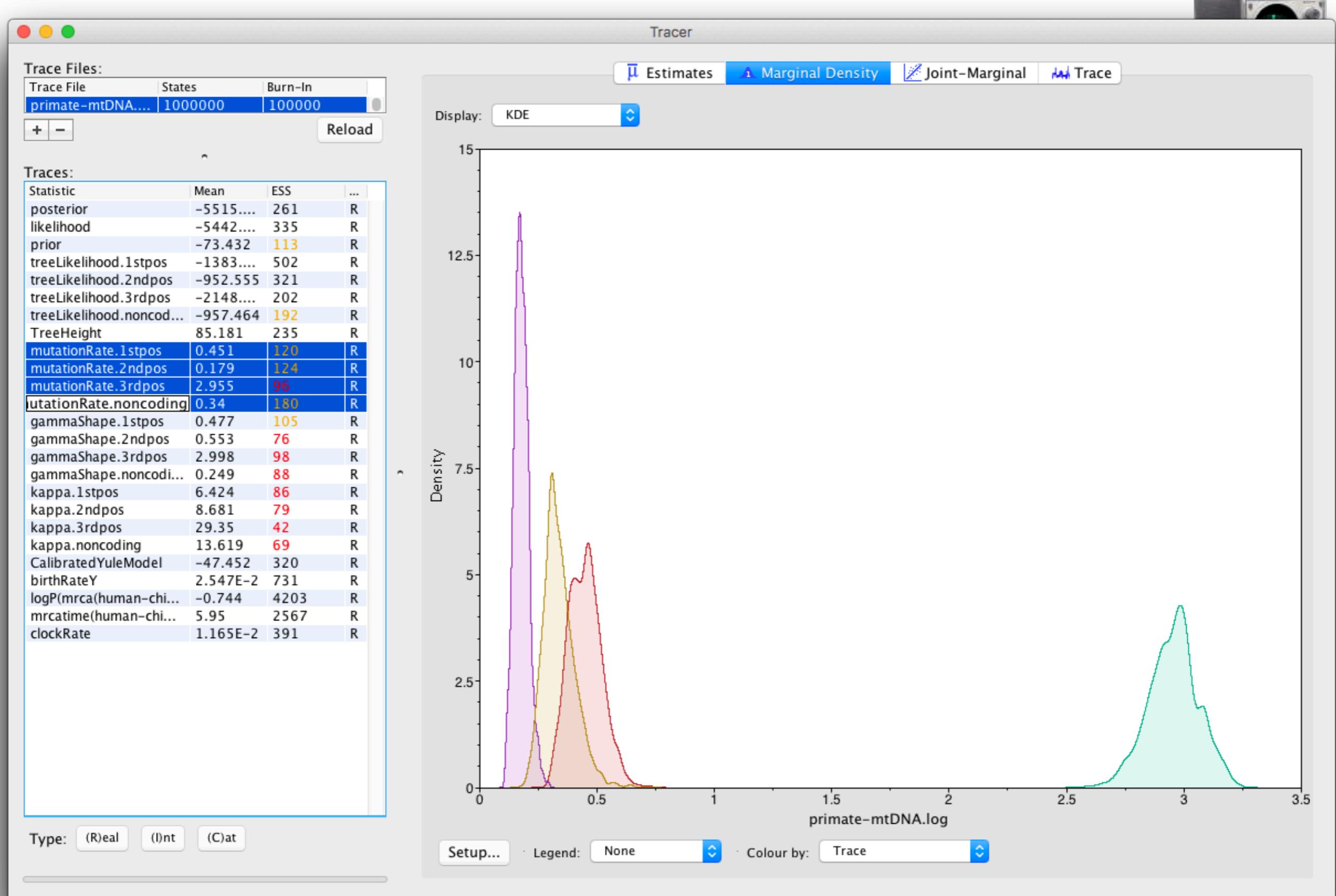
- Gain insight
- Demographic reconstructions



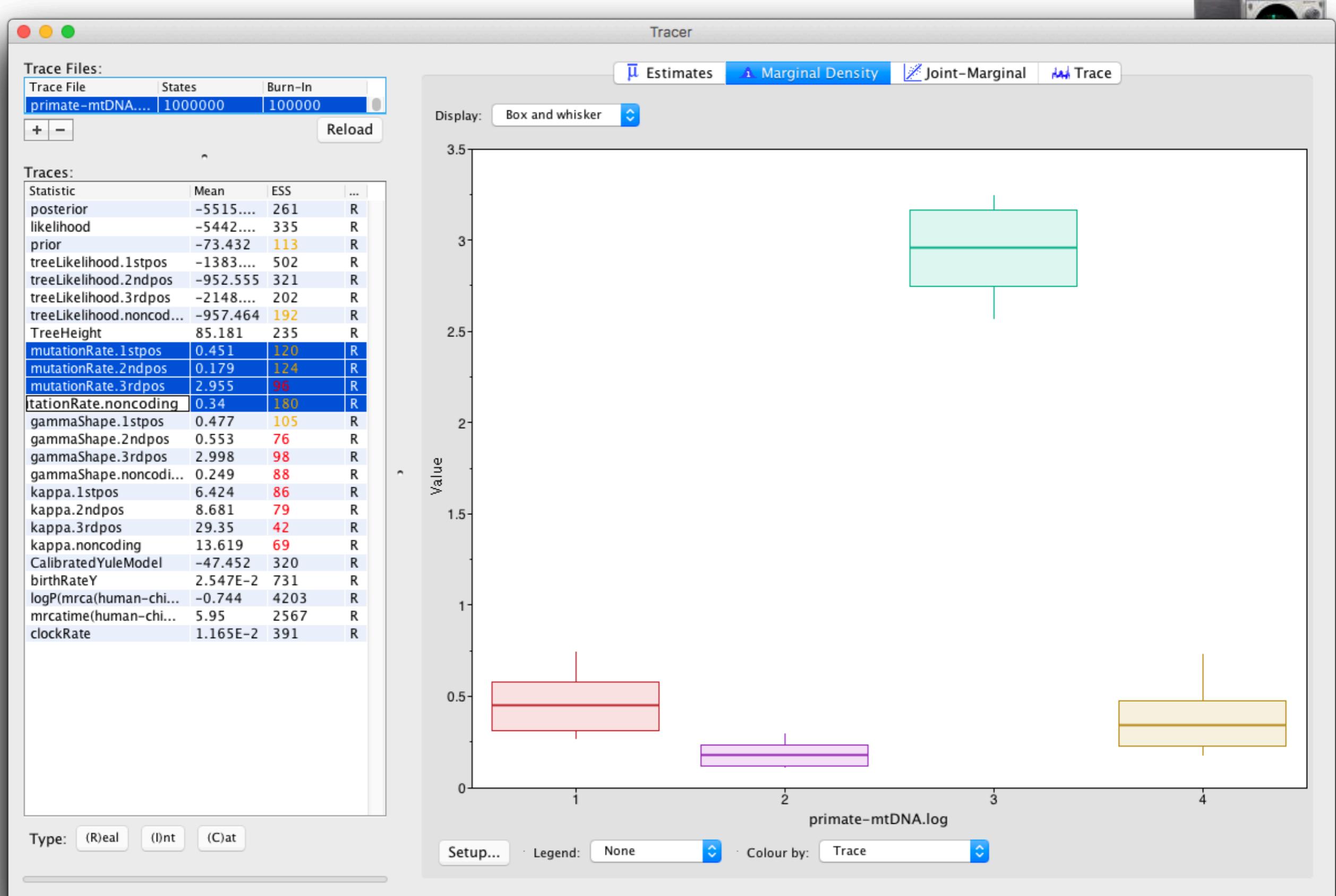
- Demographic reconstructions



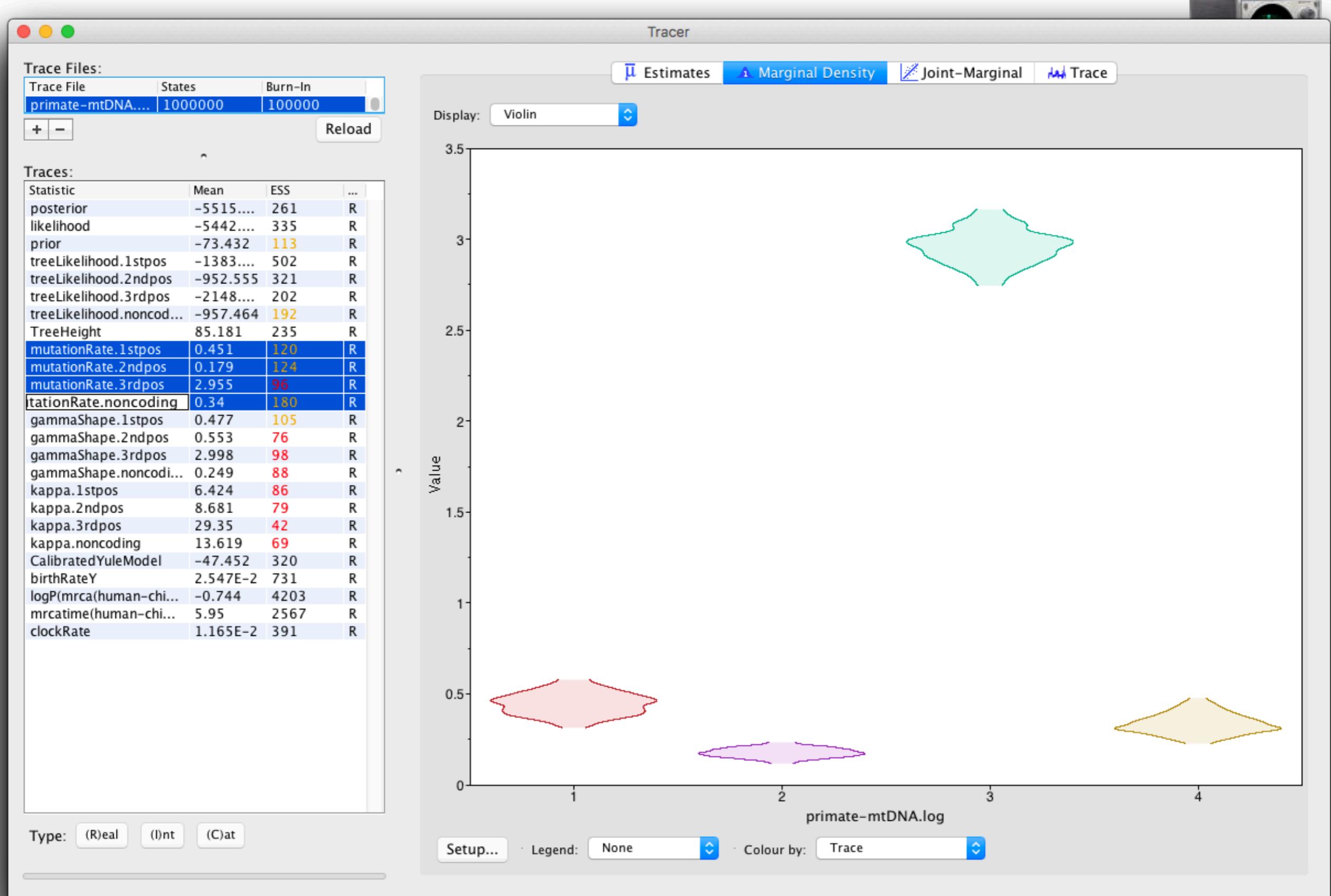
- Demographic reconstructions



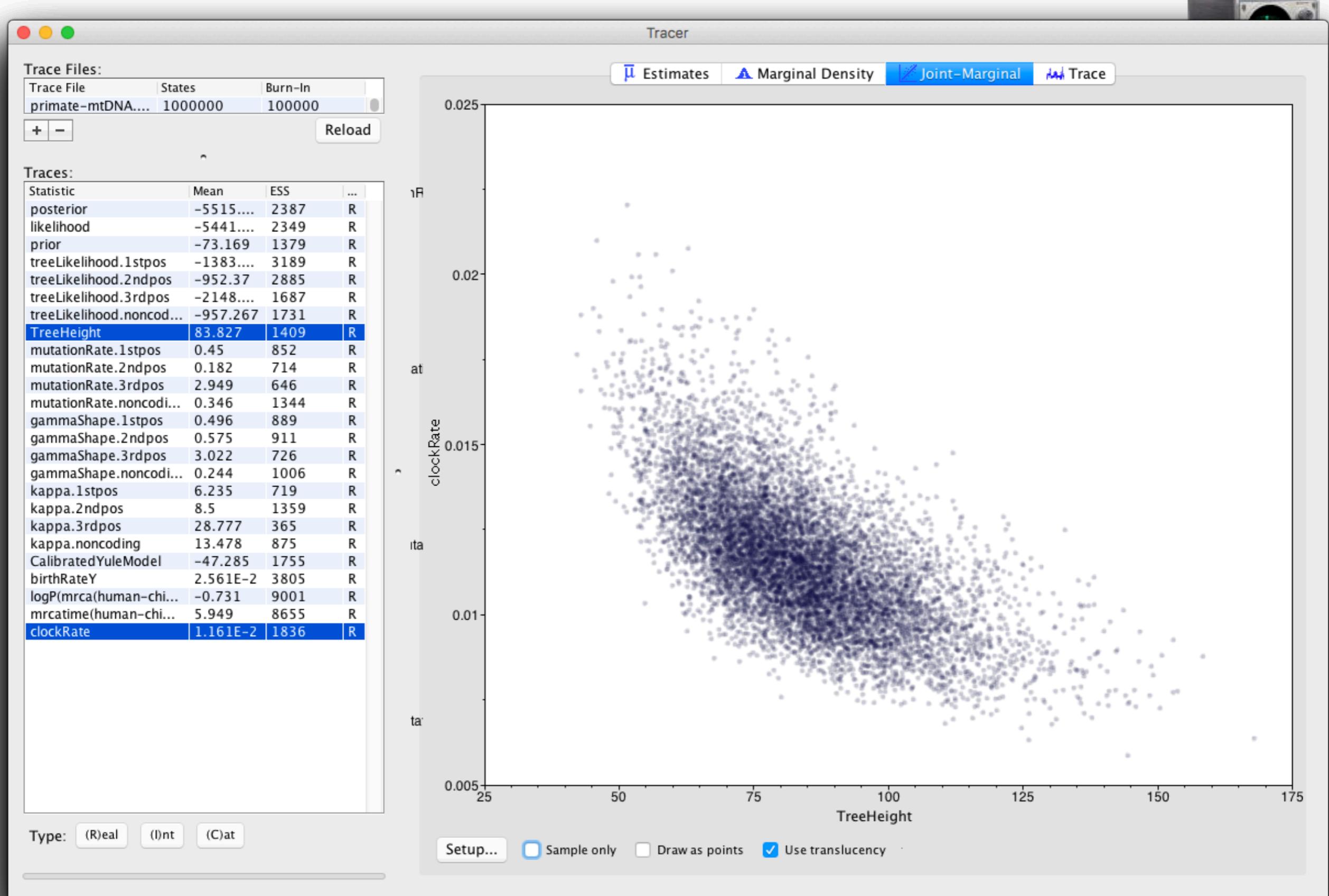
- Demographic reconstructions



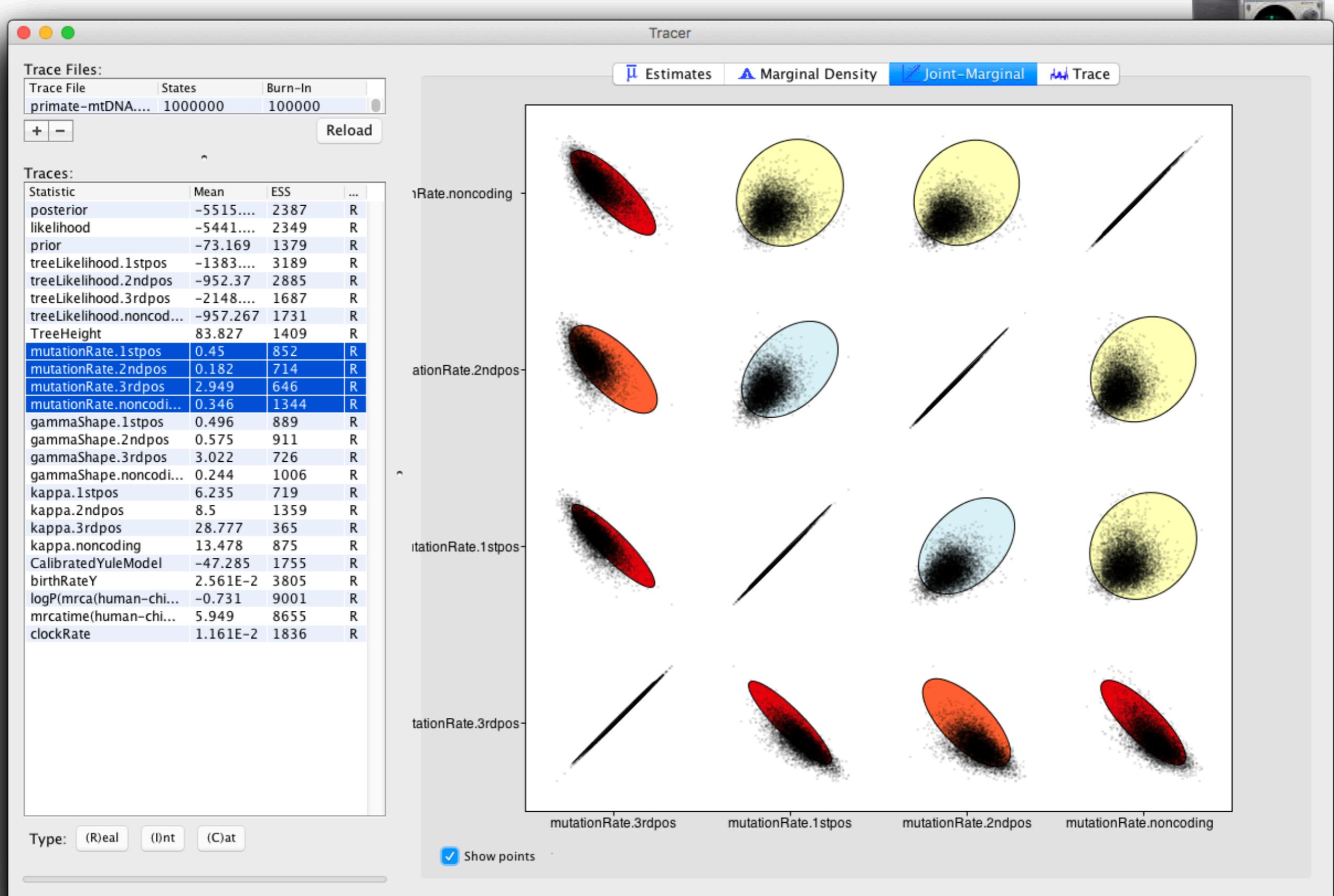
- Demographic reconstructions



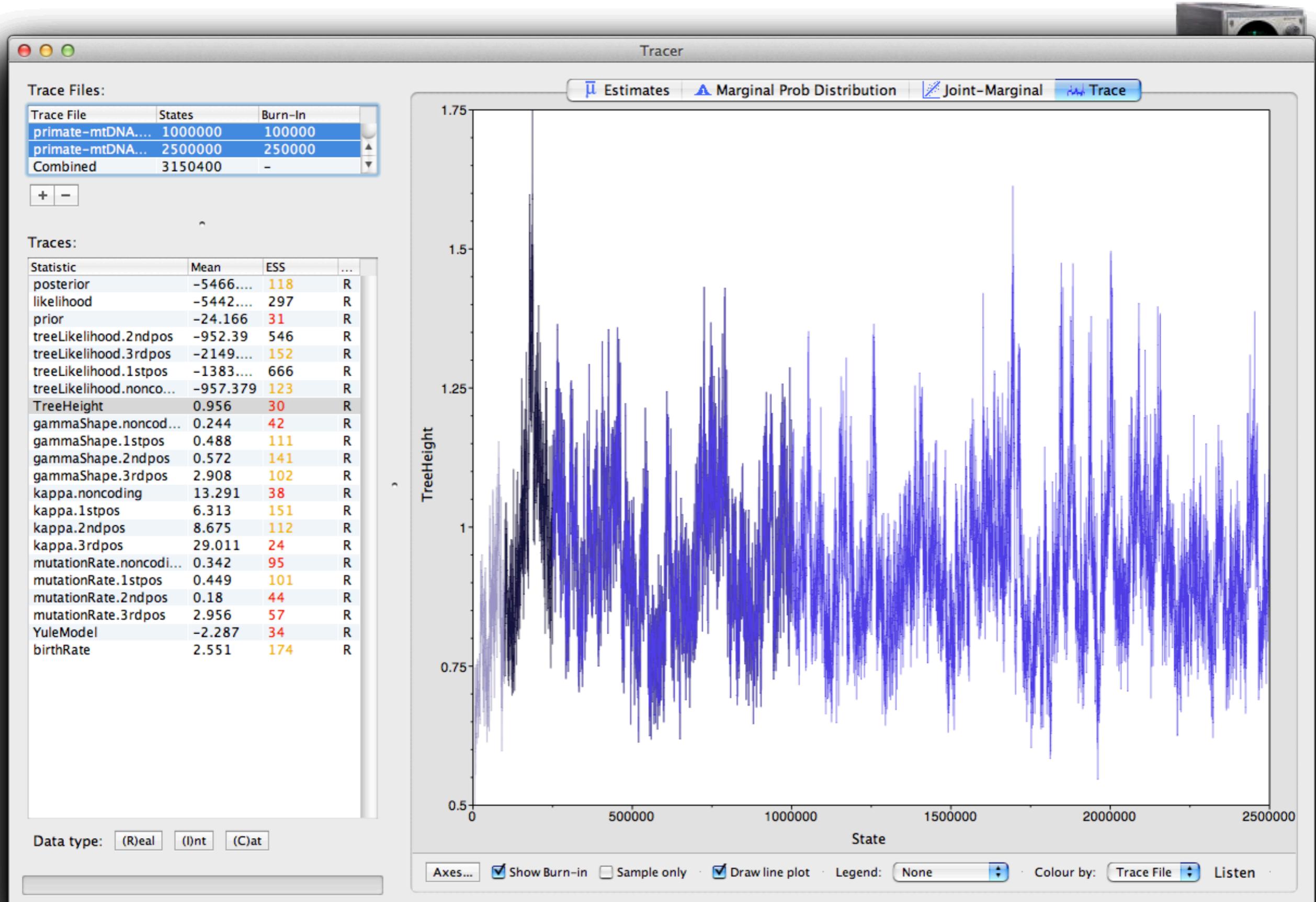
- Demographic reconstructions



- Demographic reconstructions

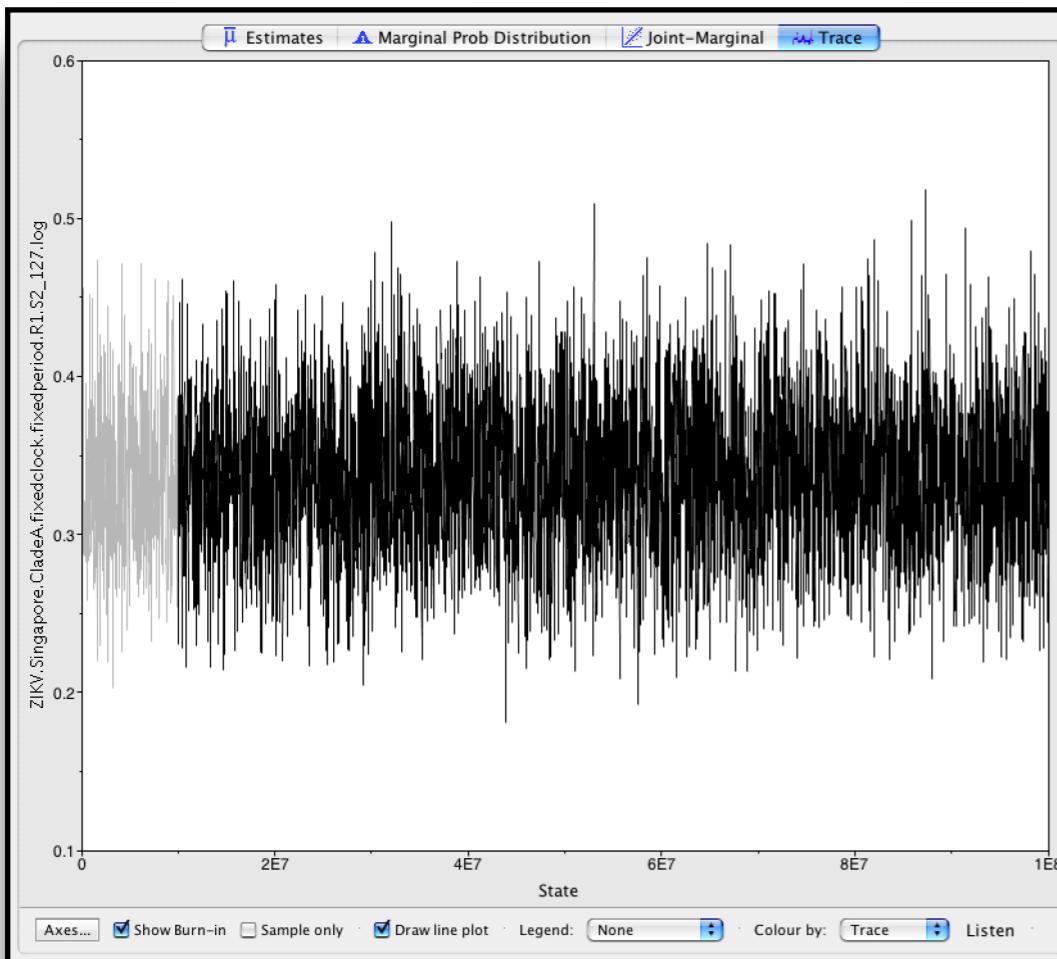


- Demographic reconstructions

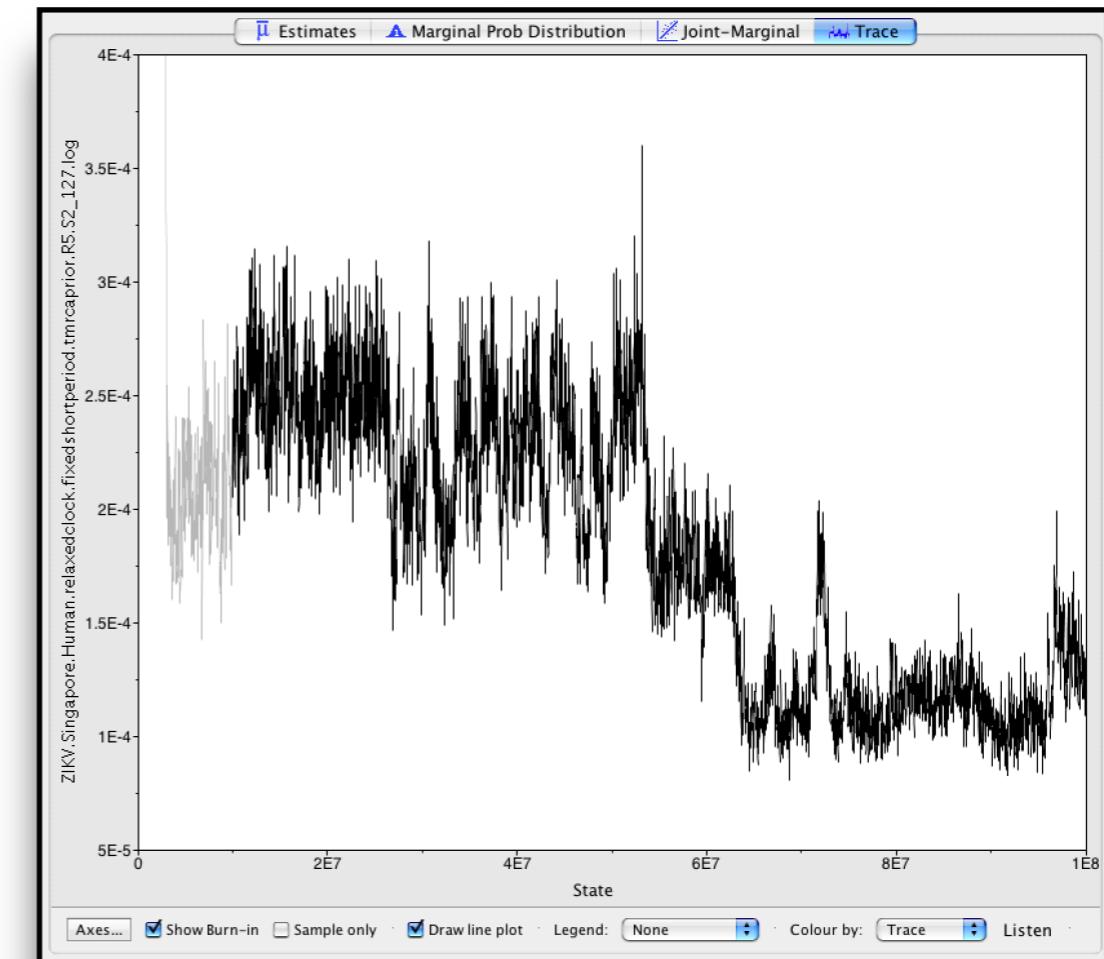


- Demographic reconstructions

# Look at the chains first!



Mixing well! 😊



Not mixing! 😢

reconstructions



# TreeAnnotator

(Included with BEAST2)

---

- Analyse trees file from BEAST2 runs
- Produces single summary tree (MCC) with node annotations (including clade posterior probabilities)
- Positions internal nodes according to average taxon set MRCA times in trees file
- Note that the MCC tree is just a heuristic summary: may produce negative branch lengths when topological uncertainty is large!

## Input:

- Tree log file  
(many trees)

## Output:

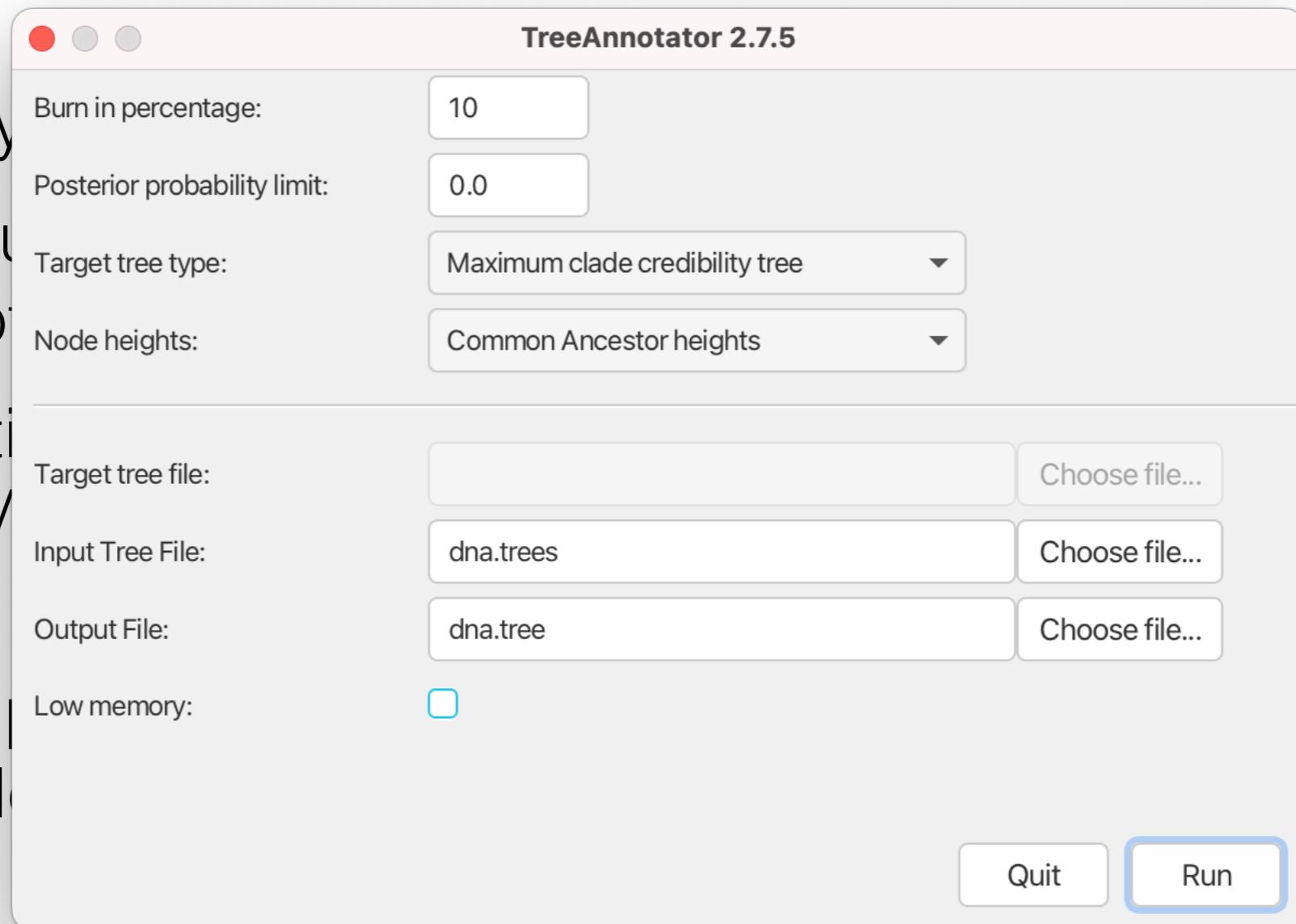
- MCC tree  
(one annotated summary tree)

# TreeAnnotator

(Included with BEAST2)



- Analyze tree log files
- Produce annotated trees
- Position nodes on a set MCC tree
- Note: may need to recompute topology



## Input:

- Tree log file  
(many trees)

## Output:

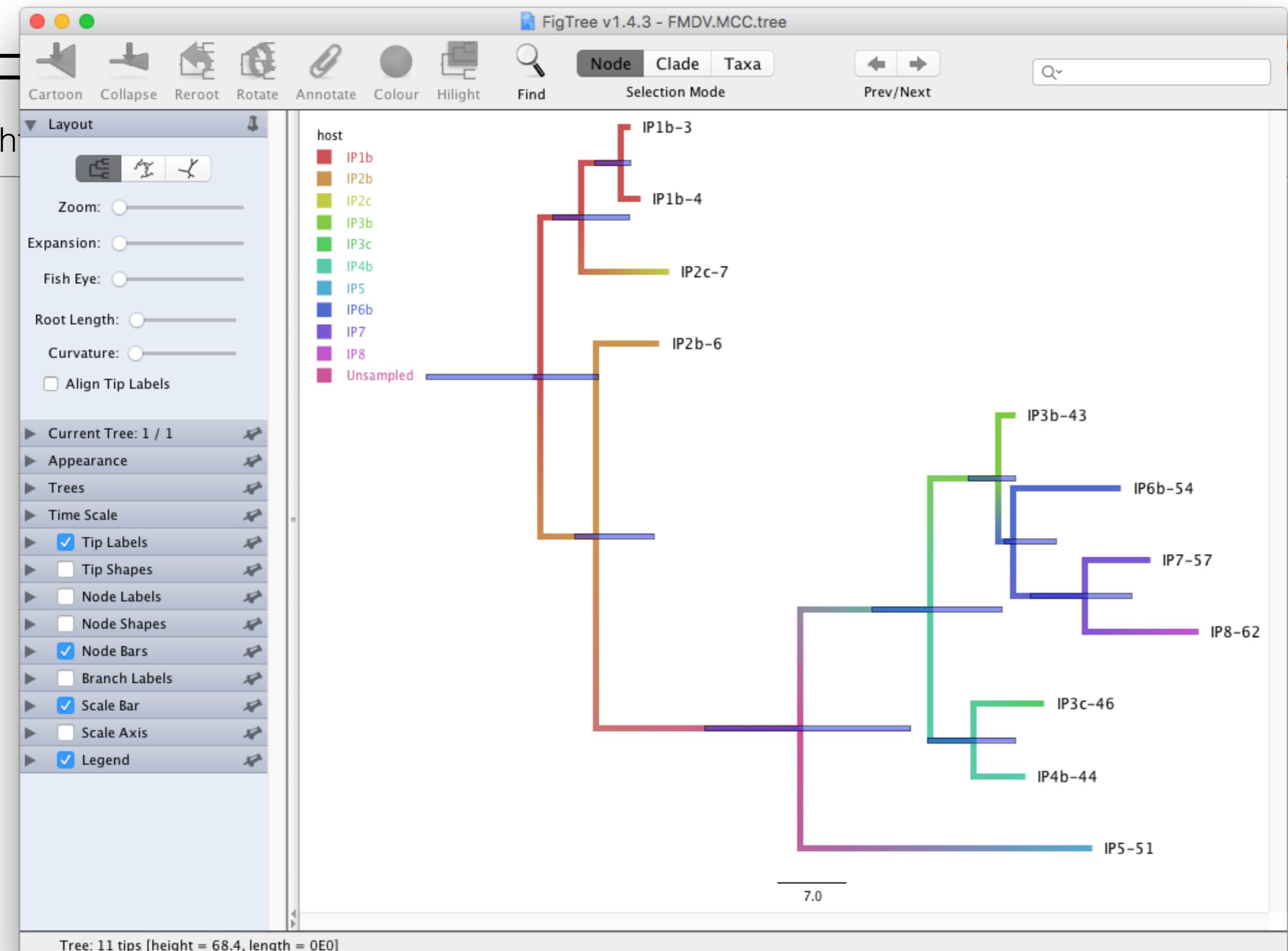
- MCC tree  
(one annotated summary tree)

# FigTree

(<http://tree.bio.ed.ac.uk/software/figtree/>)



- Visualise trees from BEAST2 runs
- Annotate branches and nodes with probabilities and labels



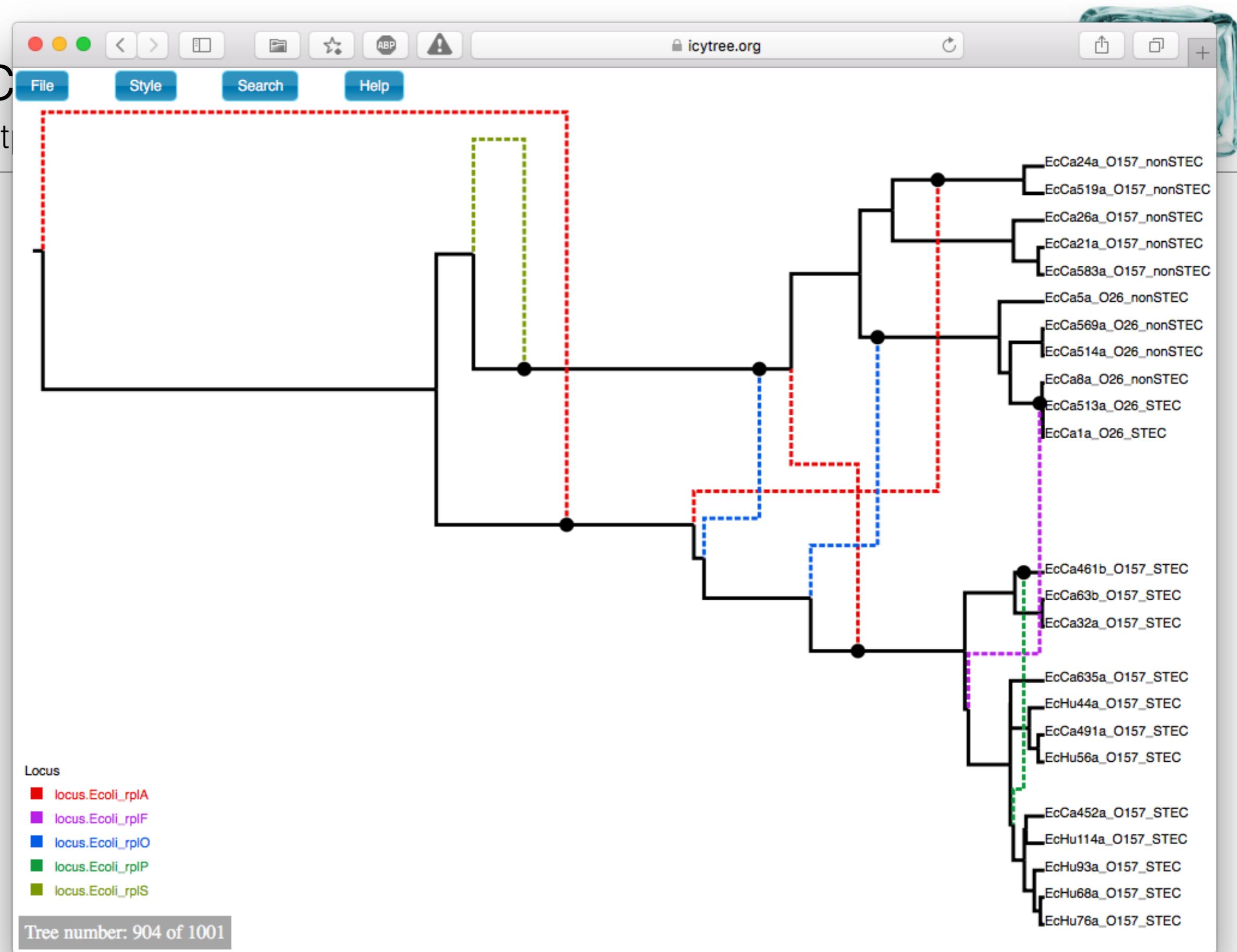
# IcyTree

(<https://icytree.org>)



- Similar to FigTree, but places an emphasis on quick visualisation rather than publication quality output
- Annotate branches and nodes with probabilities and labels
- Better suited for structured models and ancestral recombination graphs (ARGs)
- Faster than FigTree for analysing many trees
- Web app (no installation required)

|C  
(htt

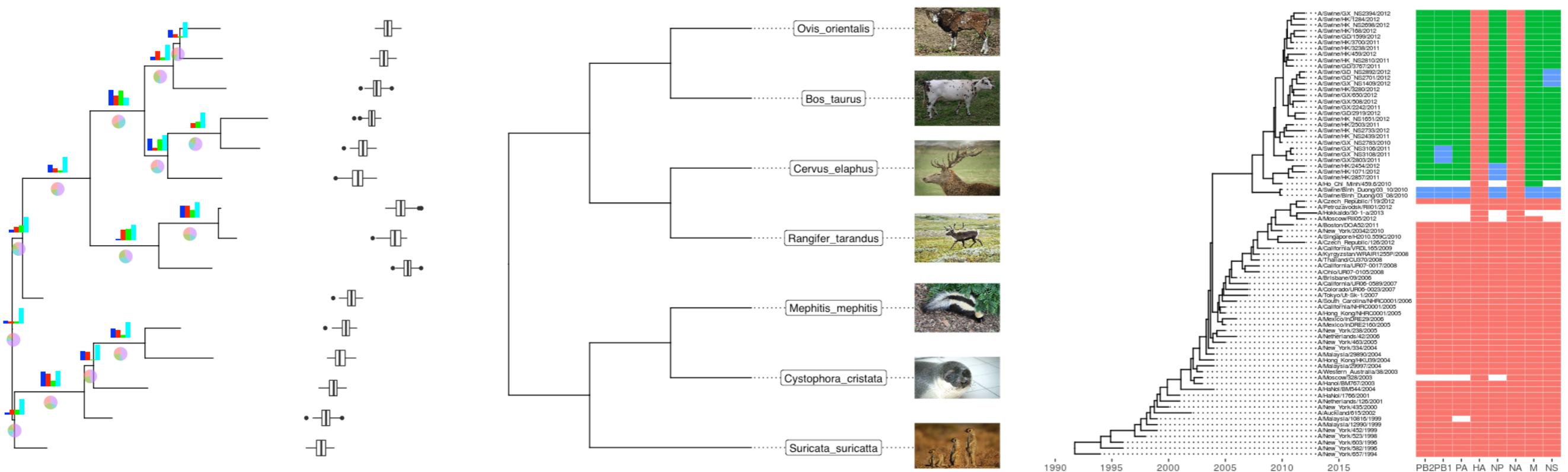


# ggtree

(<https://guangchuangyu.github.io/software/ggtree/>)

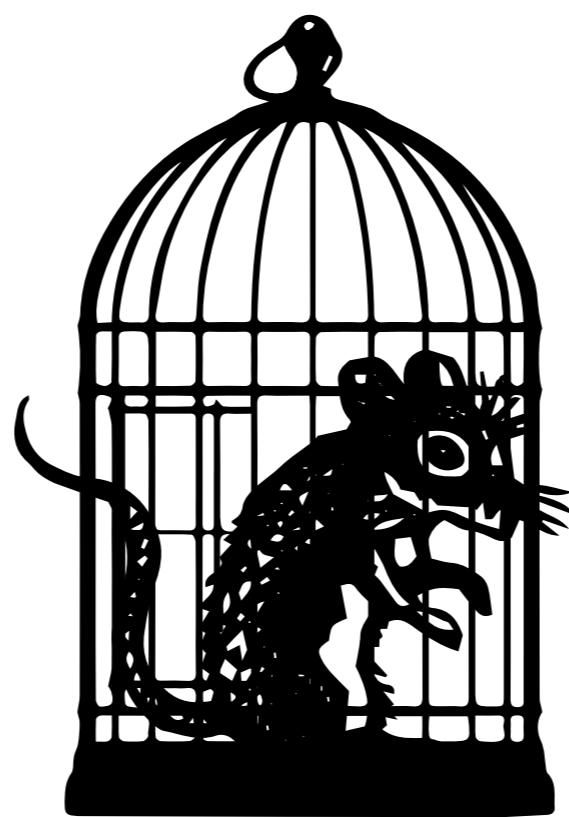


- R-package to visualise trees using ggplot grammar
- Works with BEAST2 tree files (and many other packages)
- Can easily annotate trees with other analyses in R



# Website demo...

---



# Tutorial acknowledgments

---

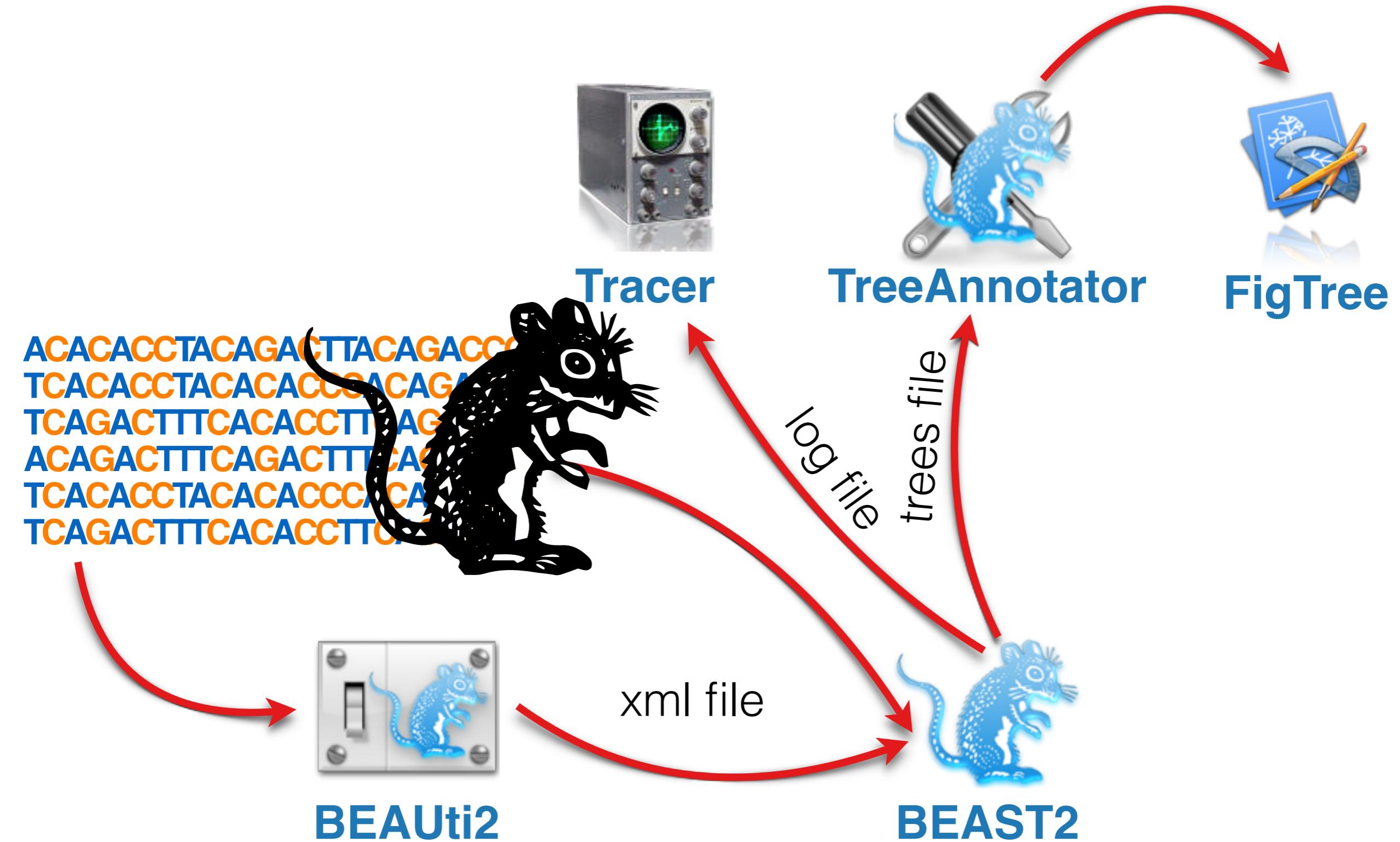
## **Tutorial authors**

Joëlle Barido-Sottani  
Veronika Bošková  
Denise Kühnert  
Carsten Magnus  
Venelin Mitov  
Nicola F. Müller  
Jūlija Pečerska  
David Rasmussen  
Remco Bouckaert  
Tracy A. Heath  
Huw Ogilvie  
Rachel Warnock

## **Updates for BEAST 2.7**

Bethany J. Allen  
Joëlle Barido-Sottani  
Tim Vaughan  
Etthel Windels  
Huw Ogilvie

# BEAST2 workflow



# Tools of the trade

---

## **BEAST2**

Software implementing MCMC for model parameter and tree inference

## **BEAUTi2**

Part of BEAST2 package for setting up the input file (.xml)

## **Tracer**

Analysis of BEAST output files (.log)

## **TreeAnnotator**

Analysis of BEAST output files (.trees)

## **FigTree, IcyTree, ggtree**

Visualisation of trees (.trees)

# BEAST best practice

(This is just a guideline and each analysis is unique)

---

## Before you begin

- 1) Know your data  
(check alignment, remove outliers, check clock signal etc.)
- 2) Plan your analysis carefully

## Before you run the analysis

- 3) Ask someone else to look at your XML file
- 4) Sample from the prior (run without data)

## Actually running the analysis

- 5) Run analysis with multiple chains

## After the analysis

- 6) Combine chains
- 7) Assess convergence and mixing
- 8) Ask someone else to look at your log files

# Thank you for listening!

Slides are my own, but some slides were inspired by (or copied from) slides by **Alexei Drummond**, **David Rasmussen**, **Carsten Magnus** and **Oliver Pybus**

