



Taming the BEAST

Bayesian Evolutionary Analyses by Sampling Trees

Louis du Plessis

1. Build-a-BEAST model
2. Crash course in MCMC
3. BEAST2 workflow

We all have one thing in common...



All of us use genomic sequencing data
to answer questions in **BEAST**

Bayesian phylogenetic and phylodynamic inference

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

Likelihood

Posterior

Prior

Marginal Likelihood of the data

Bayesian inference

(Data and model parameters are both described by probabilities)

Prior → $P(\text{model})$

- Have some degree of belief in our hypothesis
- All model parameters have priors, whether you specify them or not

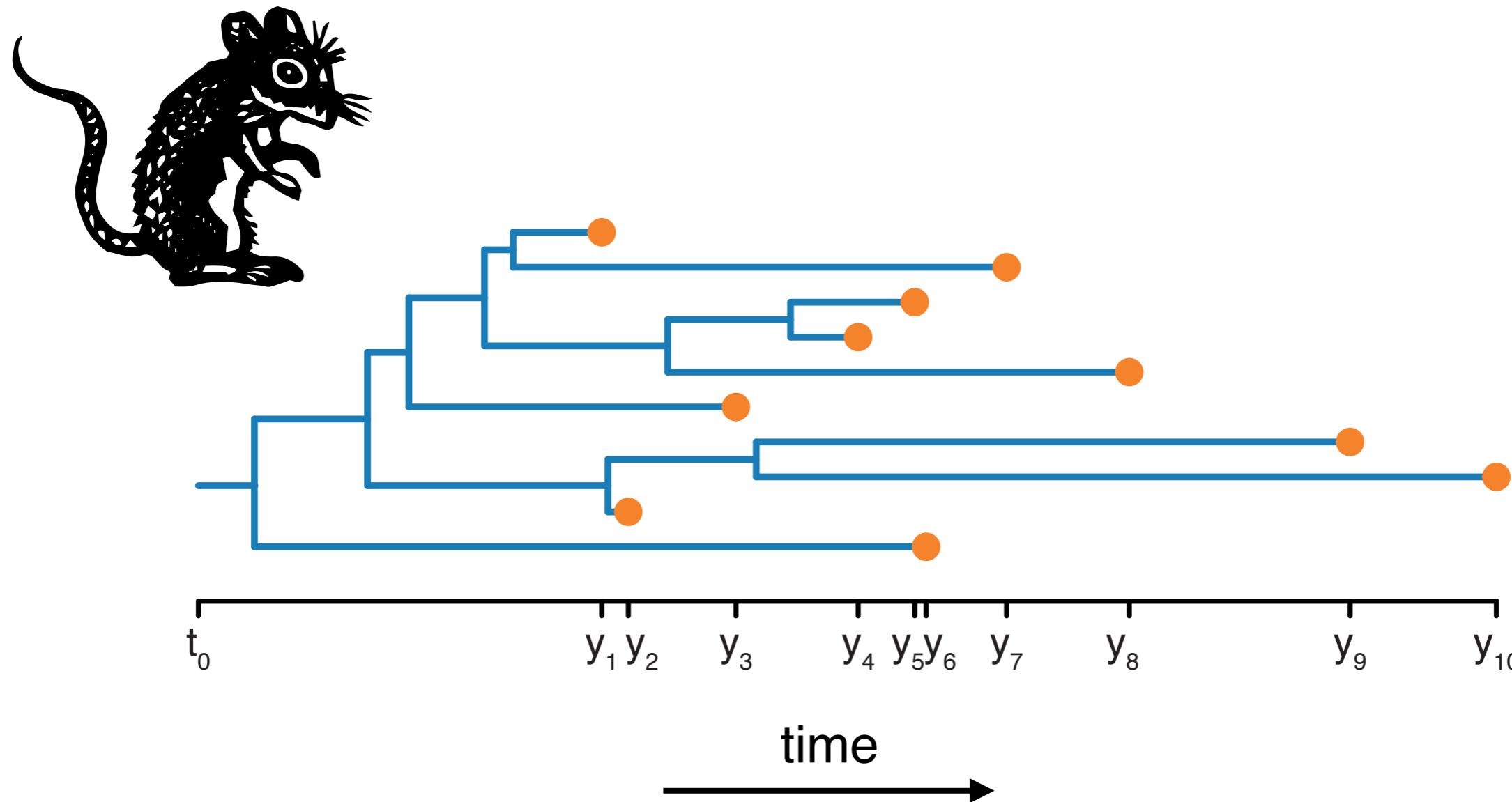
Likelihood → $P(\text{data} \mid \text{model})$

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

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

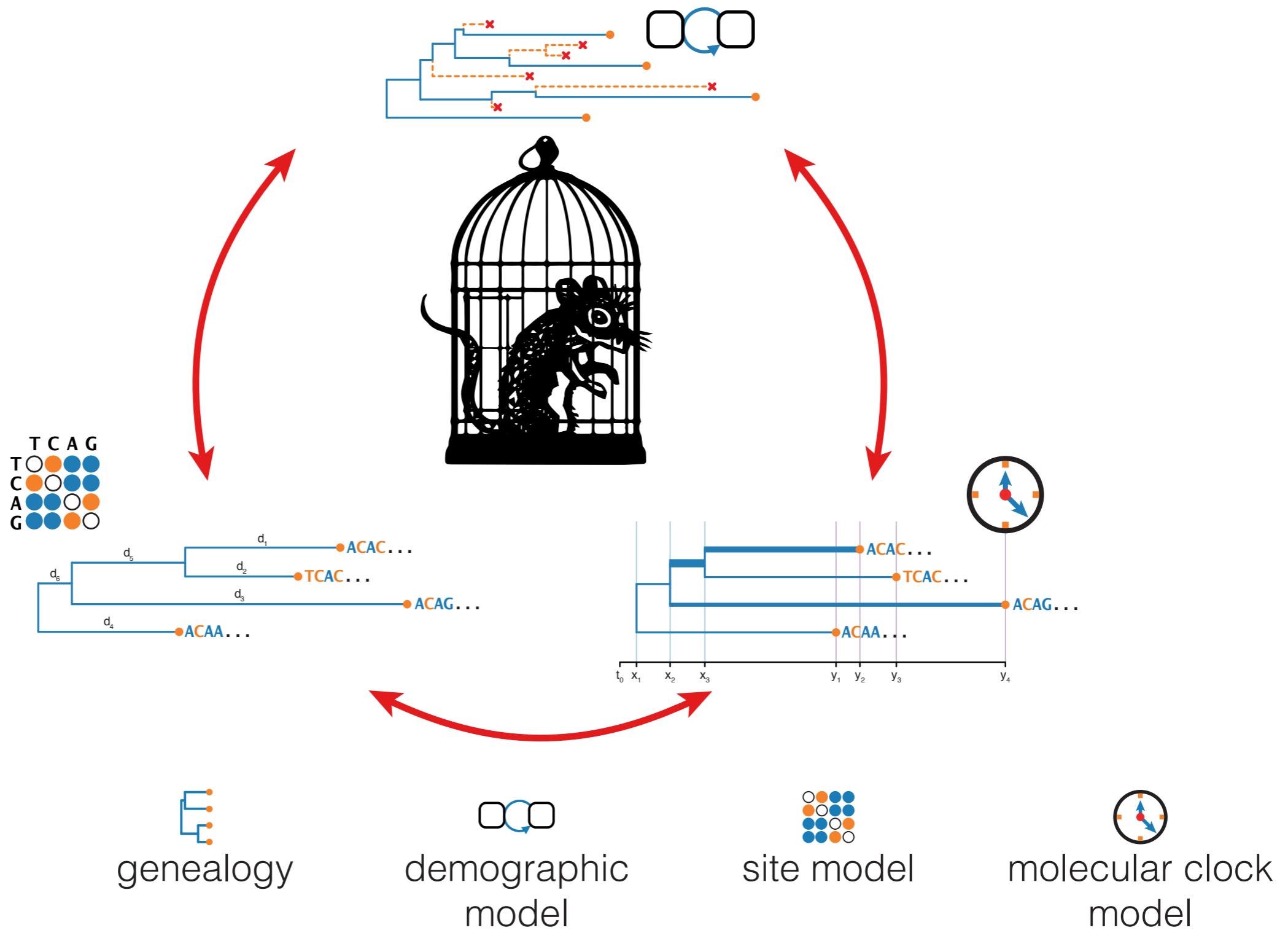
- Combines information from the data (**likelihood**) and previous knowledge (**prior**)

Rooted time-trees



Fundamental data structure in **BEAST**
is a **rooted time-tree**

What goes into a **BEAST** model?



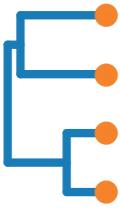
genetic
sequences

E
genealogy

OO
demographic
model

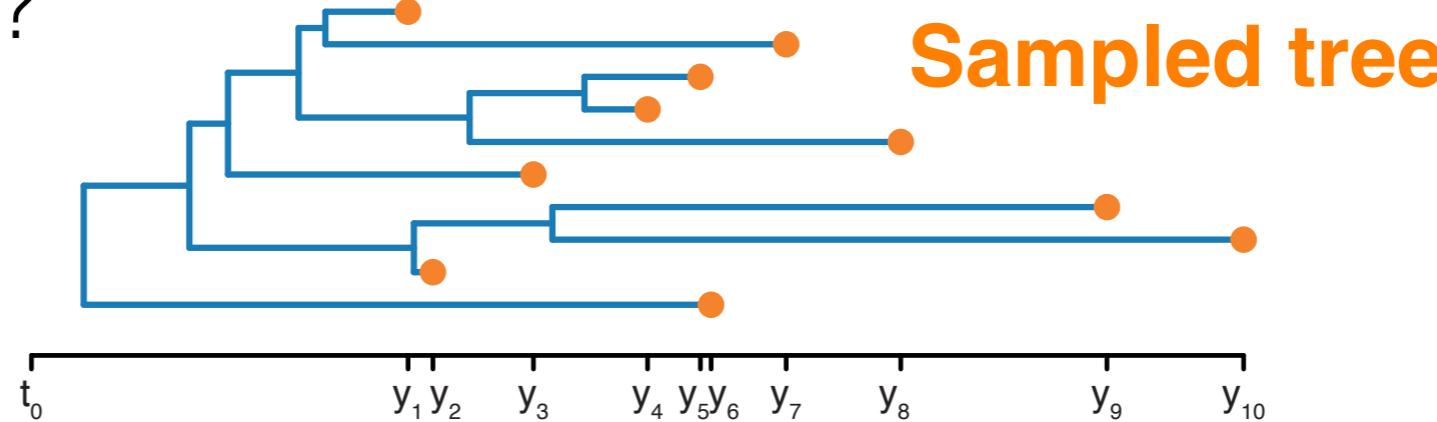
site model

clock
model

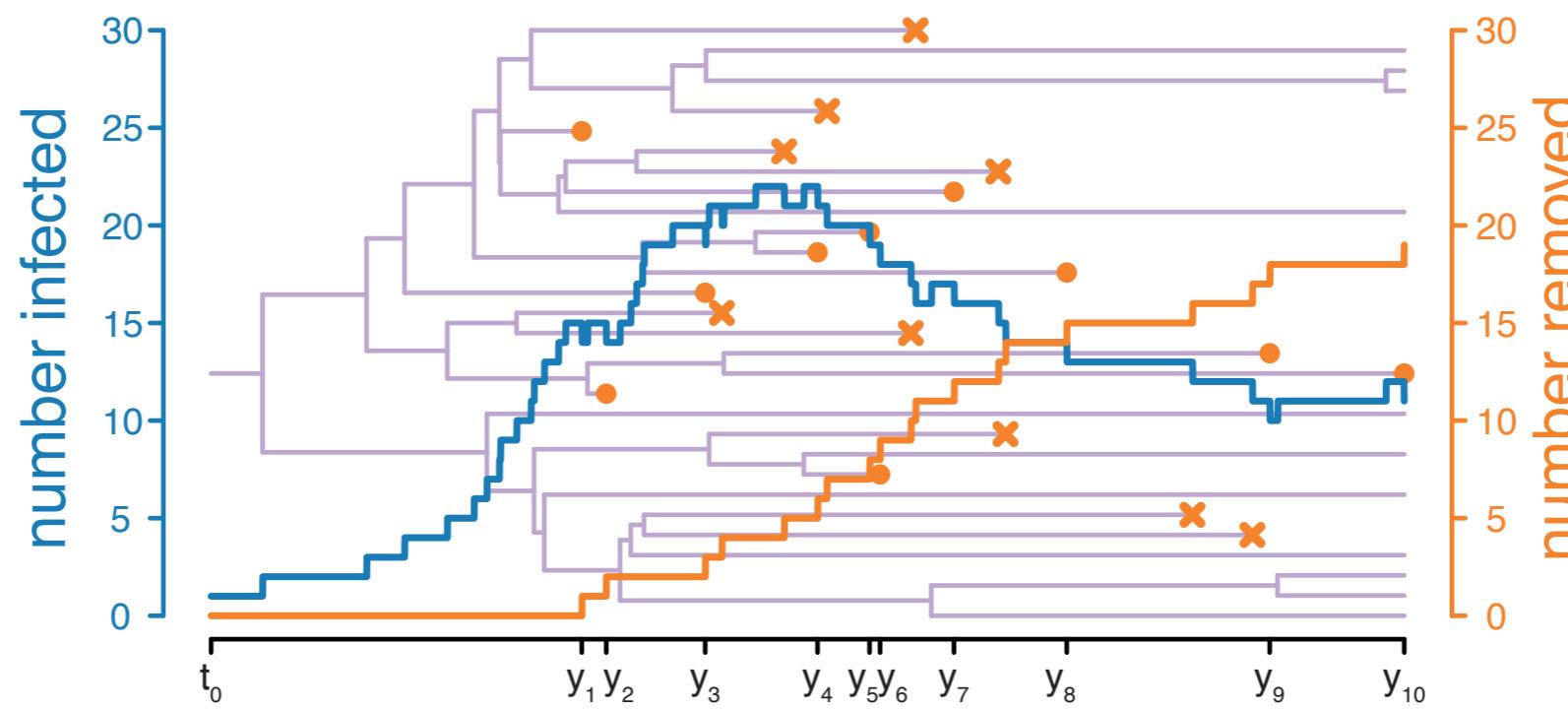


The genealogy (tree)

- What are the ancestral relationships between the sequences in our dataset?



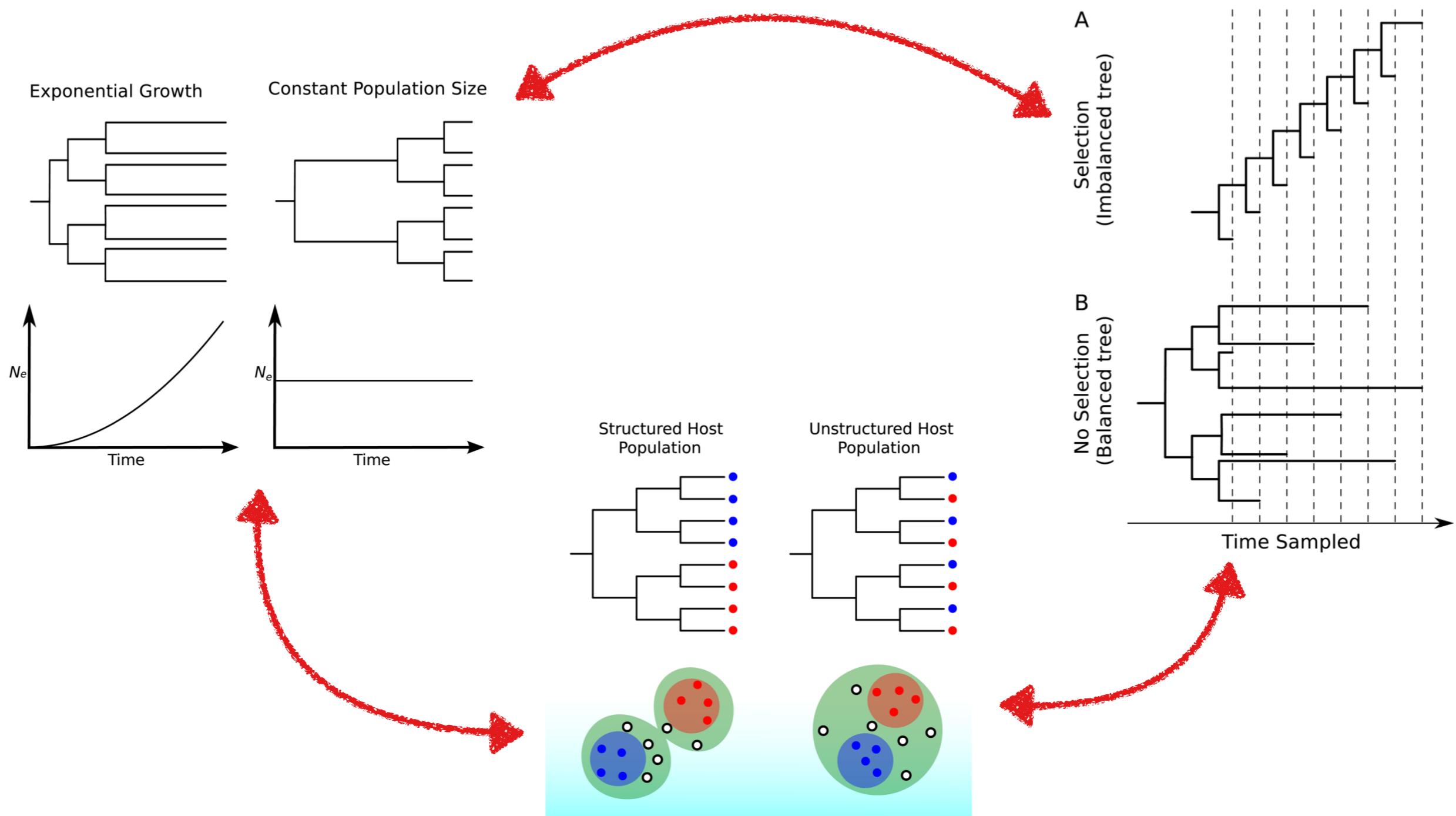
Full tree



- Only the relationships between the **sampled** sequences!

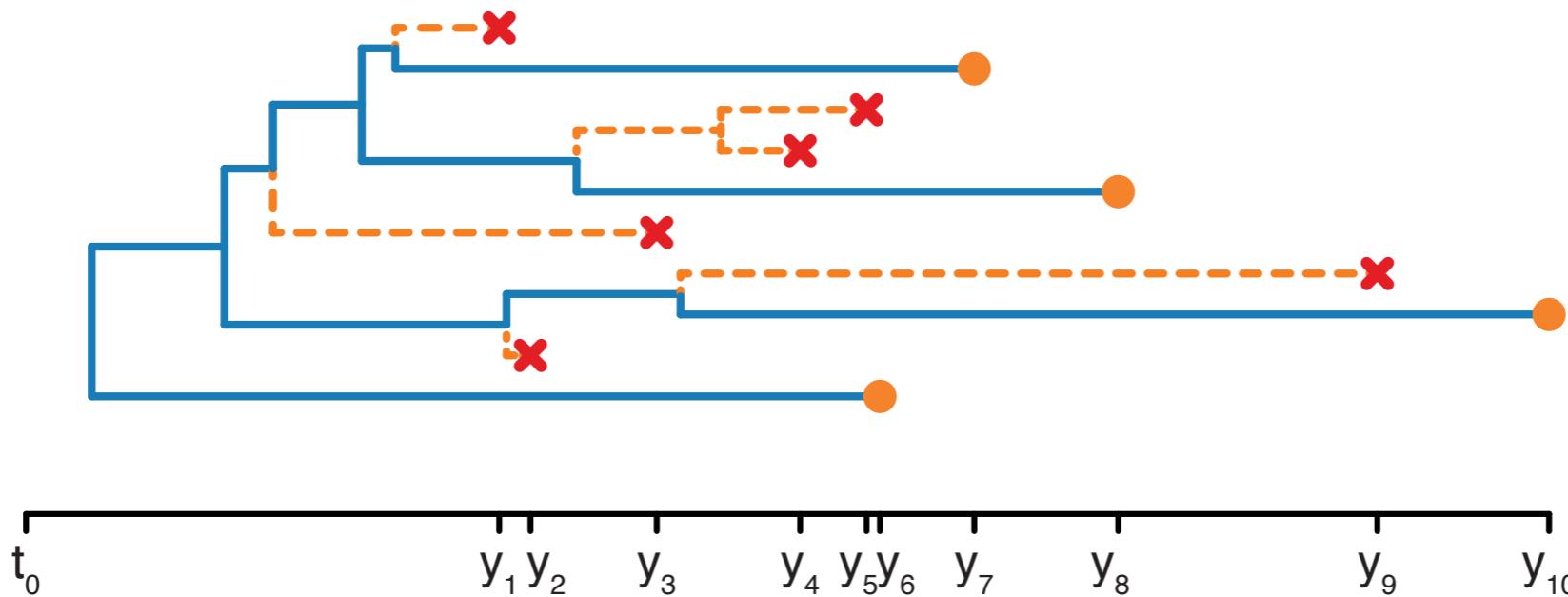


Different population dynamics generate different trees





Demographic model



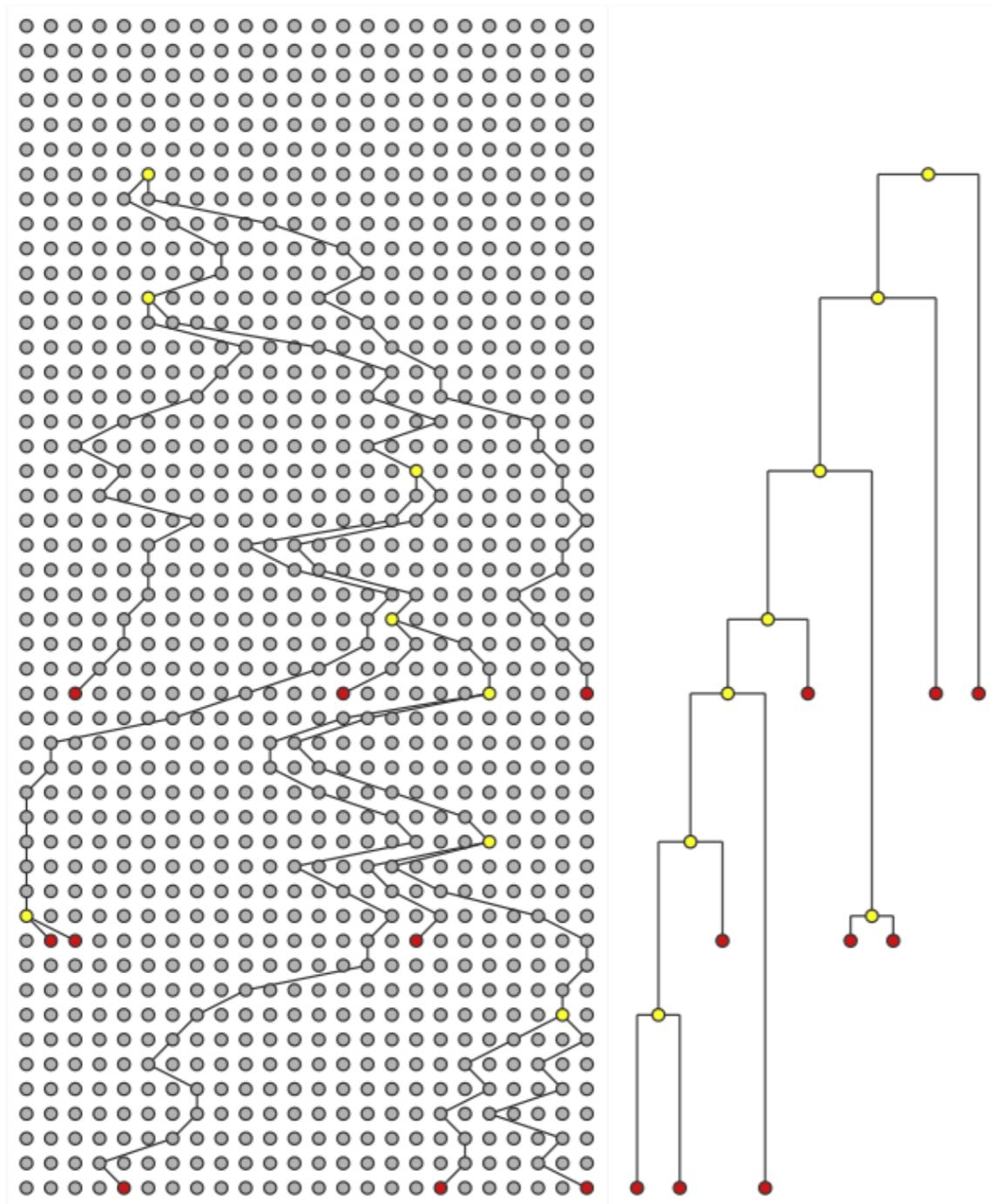
- Describes the population dynamics
- How does the population grow over time?

$$P(\text{E} | \text{OQO})$$

- How likely is the genealogy given a demographic model?
- Usually a birth-death or a coalescent model



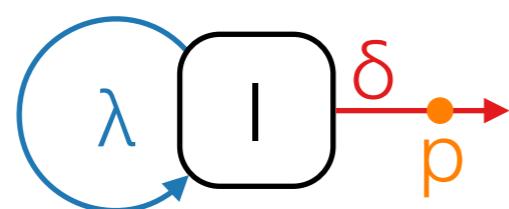
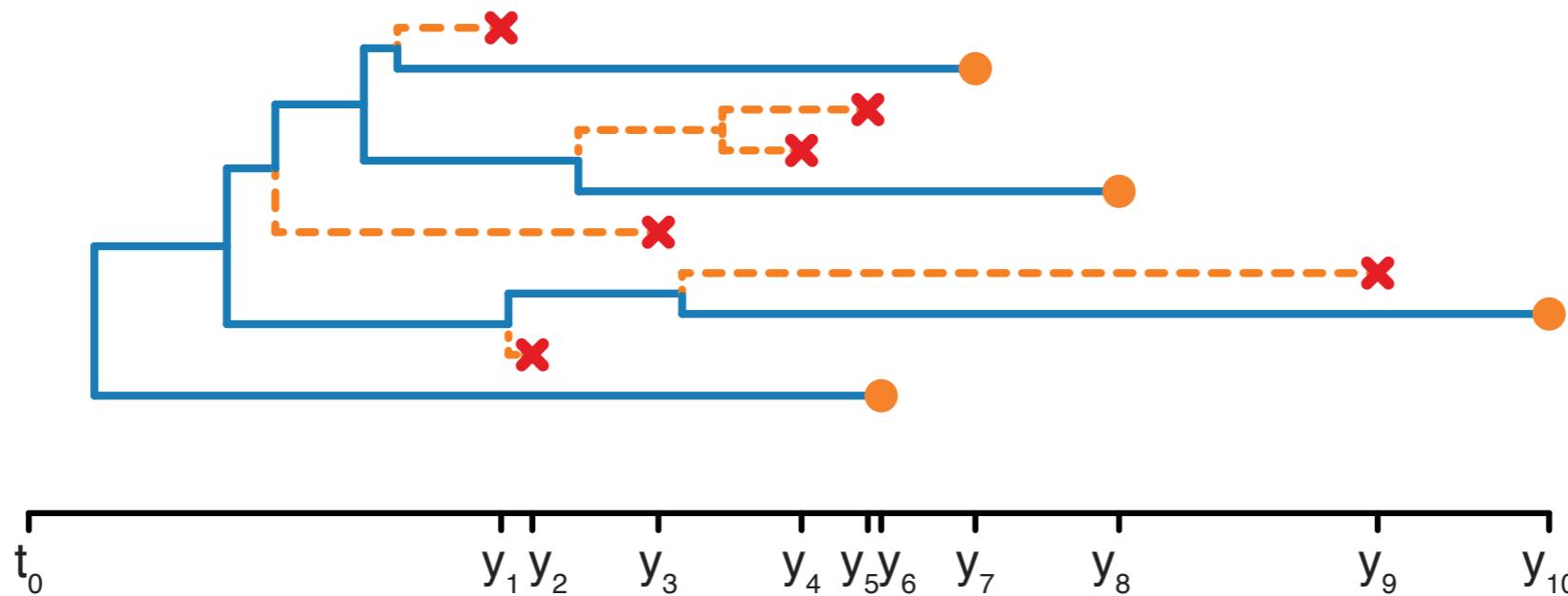
Coalescent model



- Assume Wright-Fisher like population dynamics
- Given effective population size (N_e)
- Calculate the probability for **2** nodes to coalesce in time **t**
- Calculate the probability of observing a given **tree** for a particular N_e

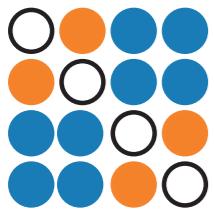


Birth-death model

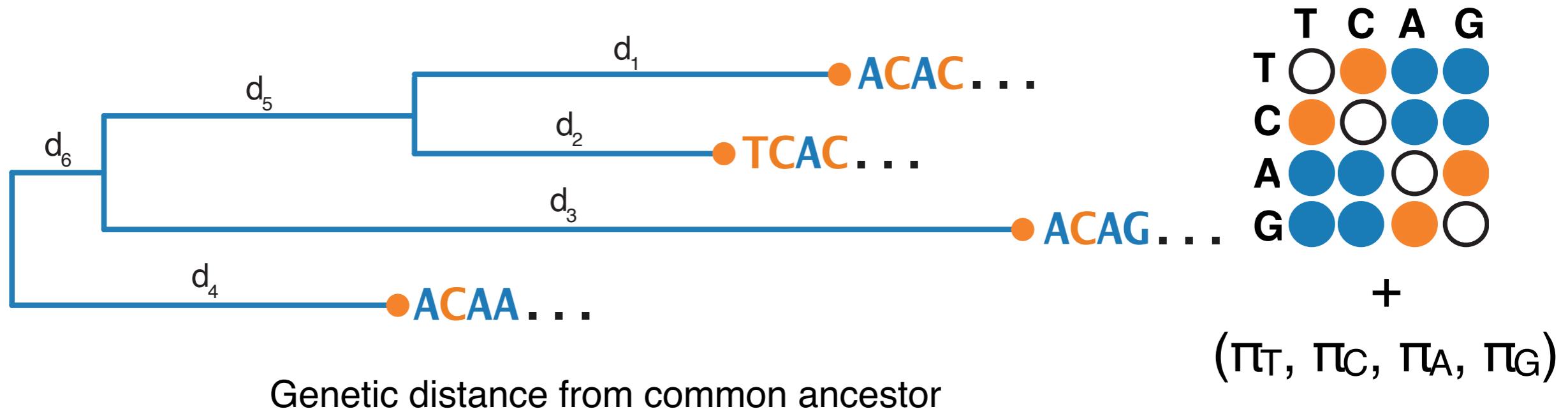


- λ — infection rate
- δ — becoming-noninfectious rate
- p — sampling probability

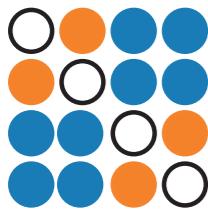
- Forward-in-time branching process
- Events happen at different rates
(speciation/infection, recovery/extinction etc.)
- Calculate the probability of a series of events happening at specific times to generate a tree



Site model



- Links the genome sequences to the genealogy
- We observe sequences at the tips, not their histories
- Multiple substitutions at the same site means not all substitutions are observed
- To infer the evolutionary history we need to take **all possible evolutionary trajectories** into account!

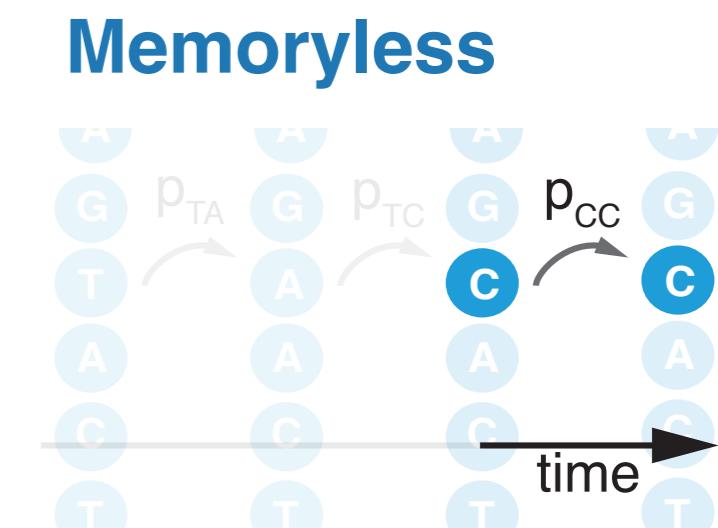
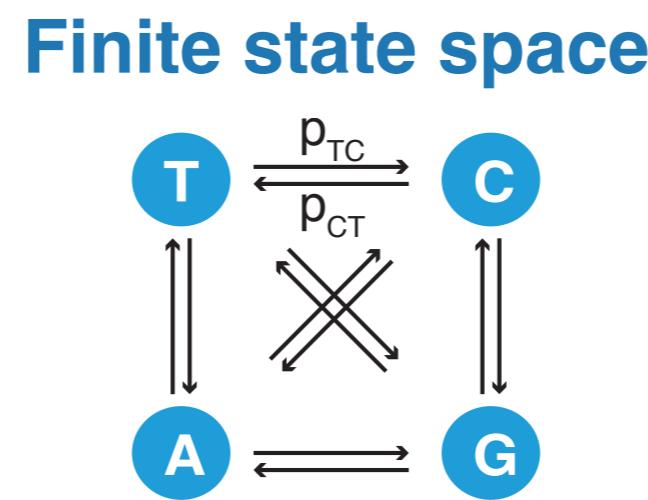
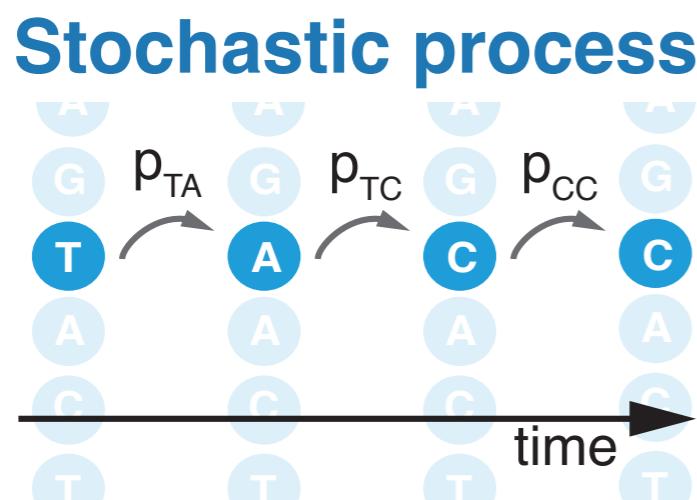


Substitutions as a Markov process

(courtesy of Carsten Magnus)

- Assume every site is evolving independently
- Assume nucleotide substitutions at each site is governed by a Markov process

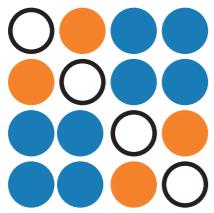
Markov process



Series of random experiments through time

Lives on a state space and jumps between different states

Probability of jumping to a state only depends on the current state



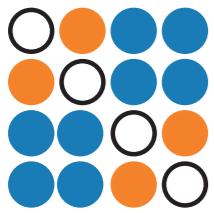
Probabilities and rates

(courtesy of Carsten Magnus)

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

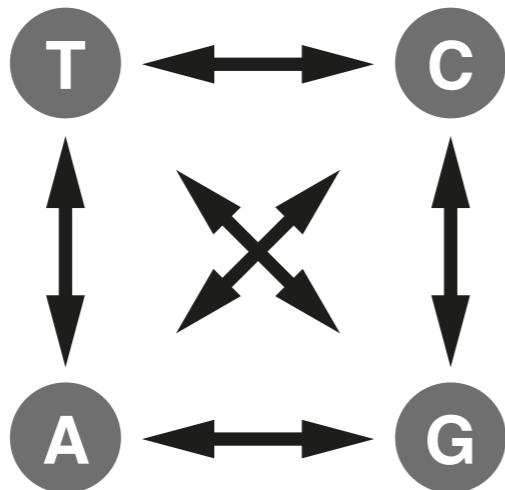
- Transition probabilities in $\mathbf{P}(t)$ take into account every possible evolutionary trajectory at each site (Chapman-Kolmogorov theorem)
- For convenience we work with the rate matrix \mathbf{Q} where q_{ij} is the relative rate of substitutions from state i to j

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



Jukes-Cantor model (JC69)

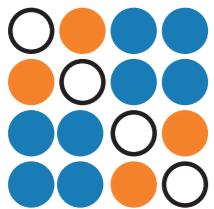
(courtesy of Carsten Magnus)



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

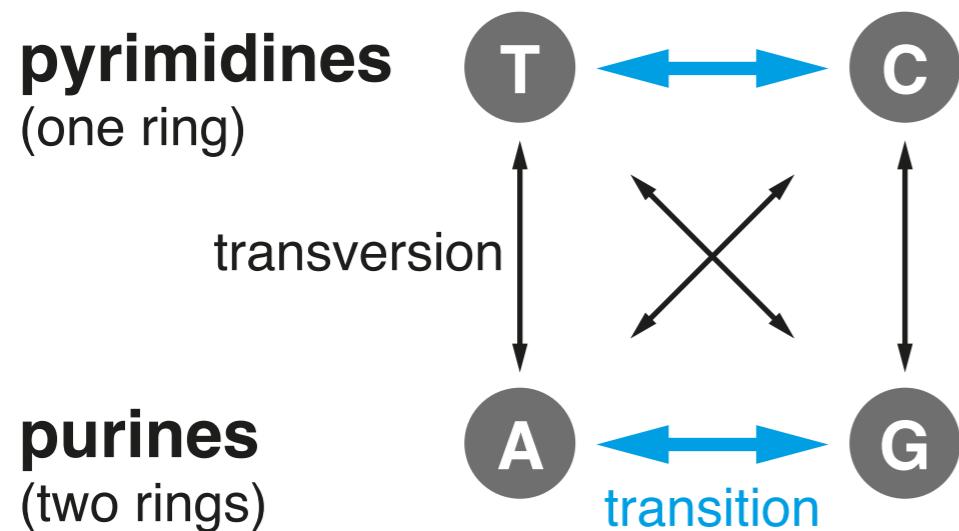
$$\pi_T = \pi_C = \pi_A = \pi_G$$

- Simplest model
- All rates and frequencies are equal!



Kimura 2-parameter model (K80)

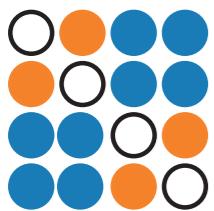
(courtesy of Carsten Magnus)



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

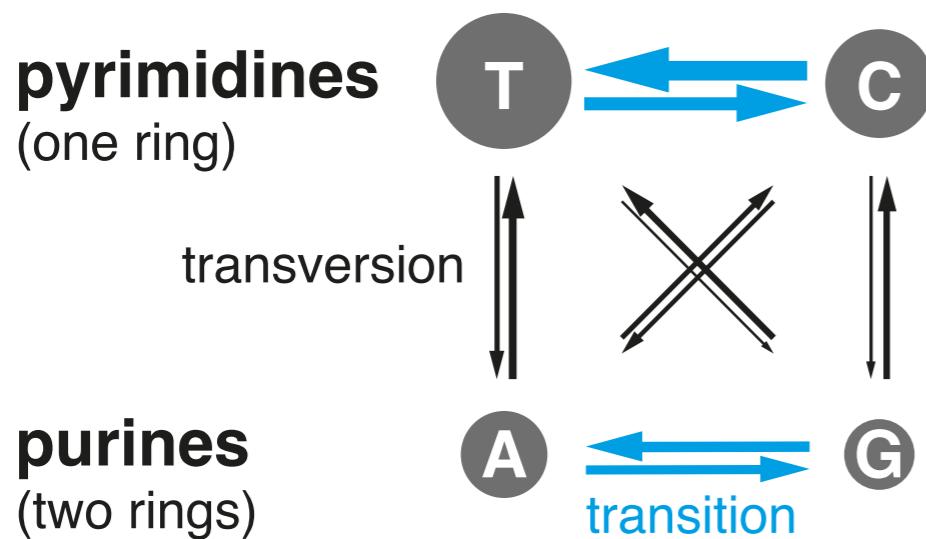
$$\pi_T = \pi_C = \pi_A = \pi_G$$

- 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)

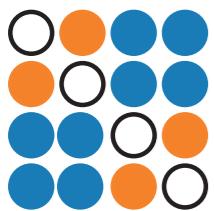
(courtesy of Carsten Magnus)



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

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

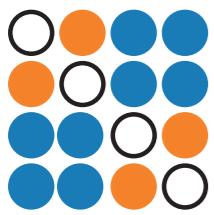
- Accounts for transition/transversion bias
- Not symmetric anymore ($r_{ij} \neq r_{ji}$)
- Still time-reversible ($\pi_i q_{ij} = \pi_j q_{ji}$)



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

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

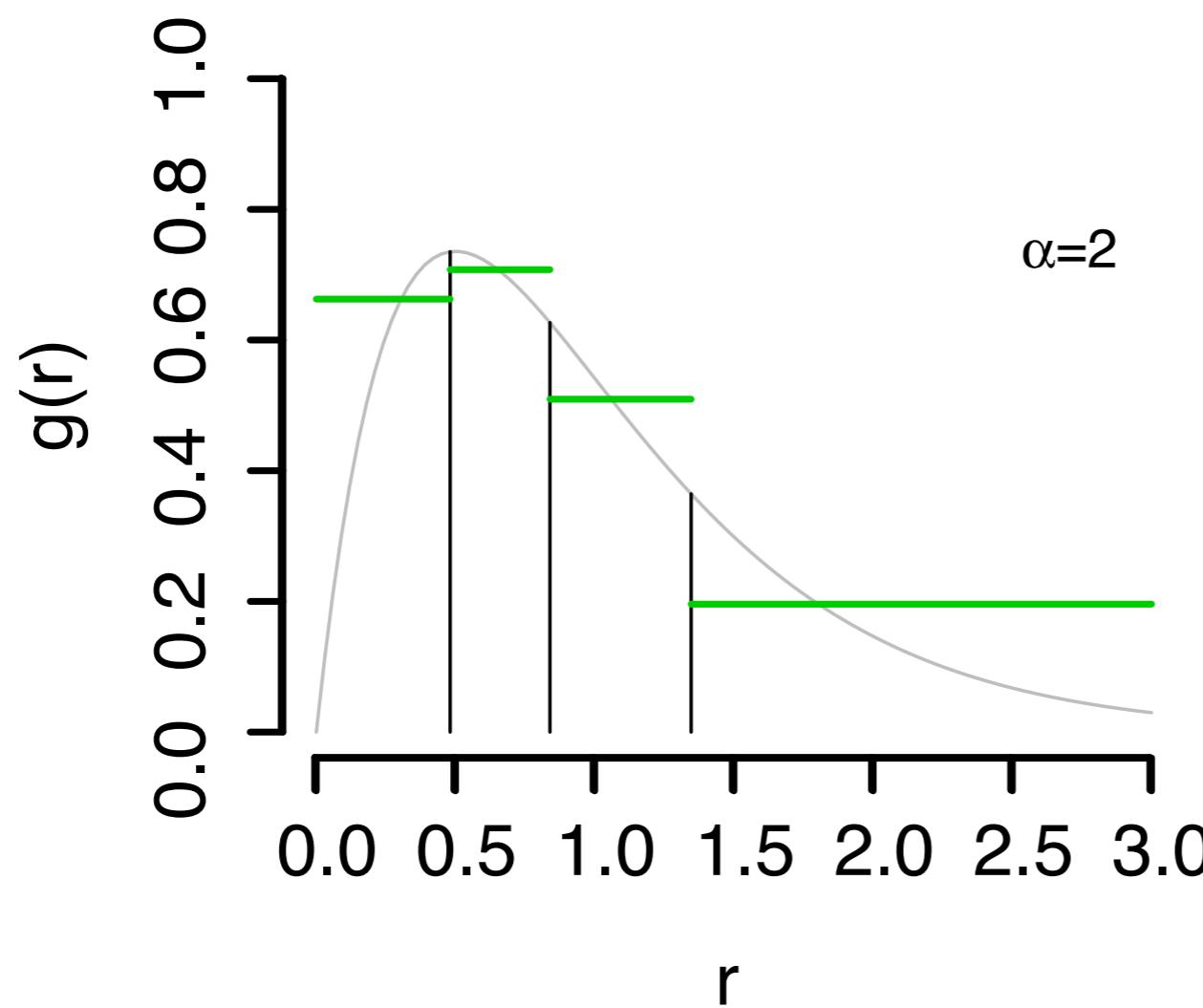
- Most general time-reversible model
- More flexible models are possible, but mathematically inconvenient

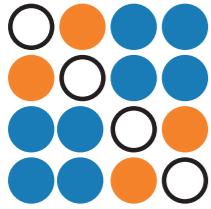


Gamma rate heterogeneity

(courtesy of Carsten Magnus)

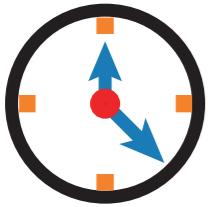
- Not all sites evolve at the same rate
- Assume rate heterogeneity is Γ -distributed
- Discretise Γ -distribution to n discrete rate categories for computational reasons



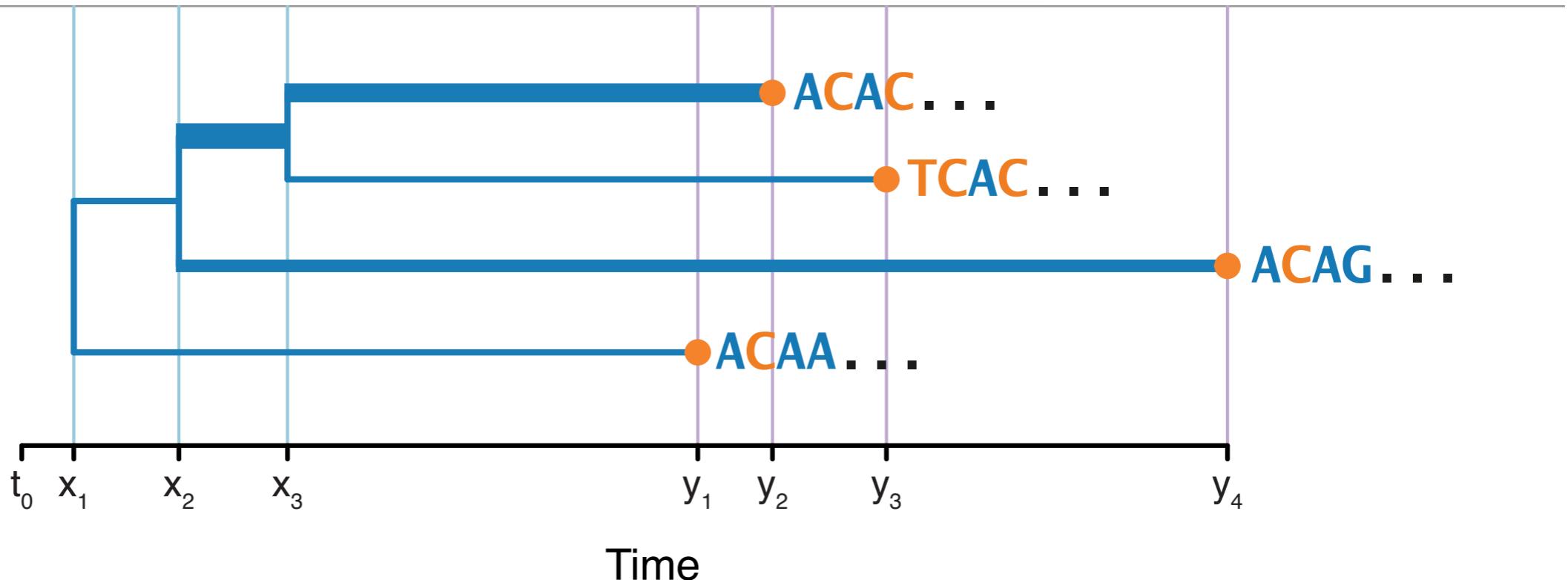


Multi-locus models

- Γ -distributed rate variation is not 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



Molecular clock model



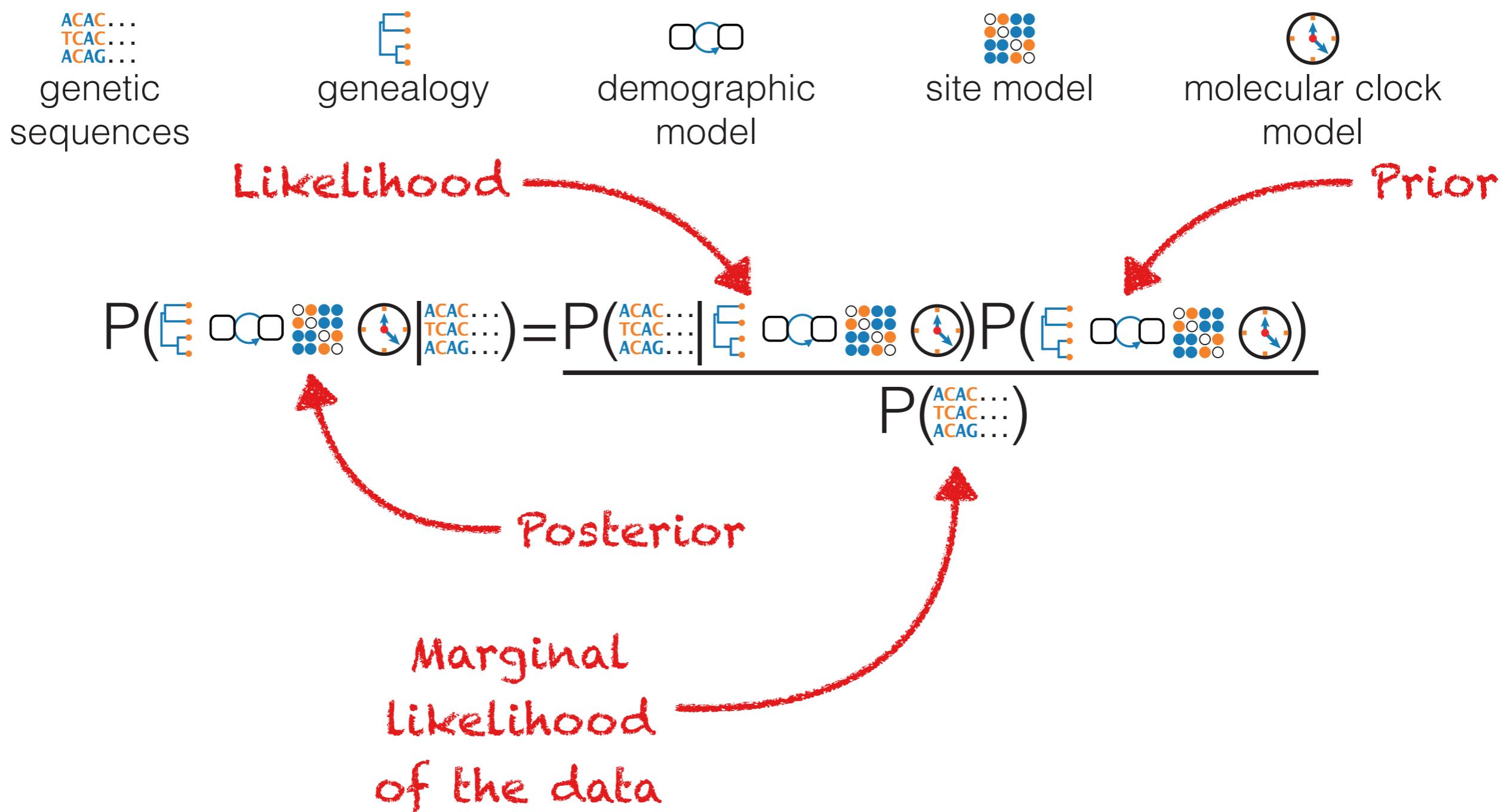
- Scales branch lengths to calendar time
- How long does it take for substitutions to appear?
- Different branches may have different clock rates
- Priors on internal nodes can help to calibrate the clock

Putting it all together



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

Putting it all together



Putting it all together

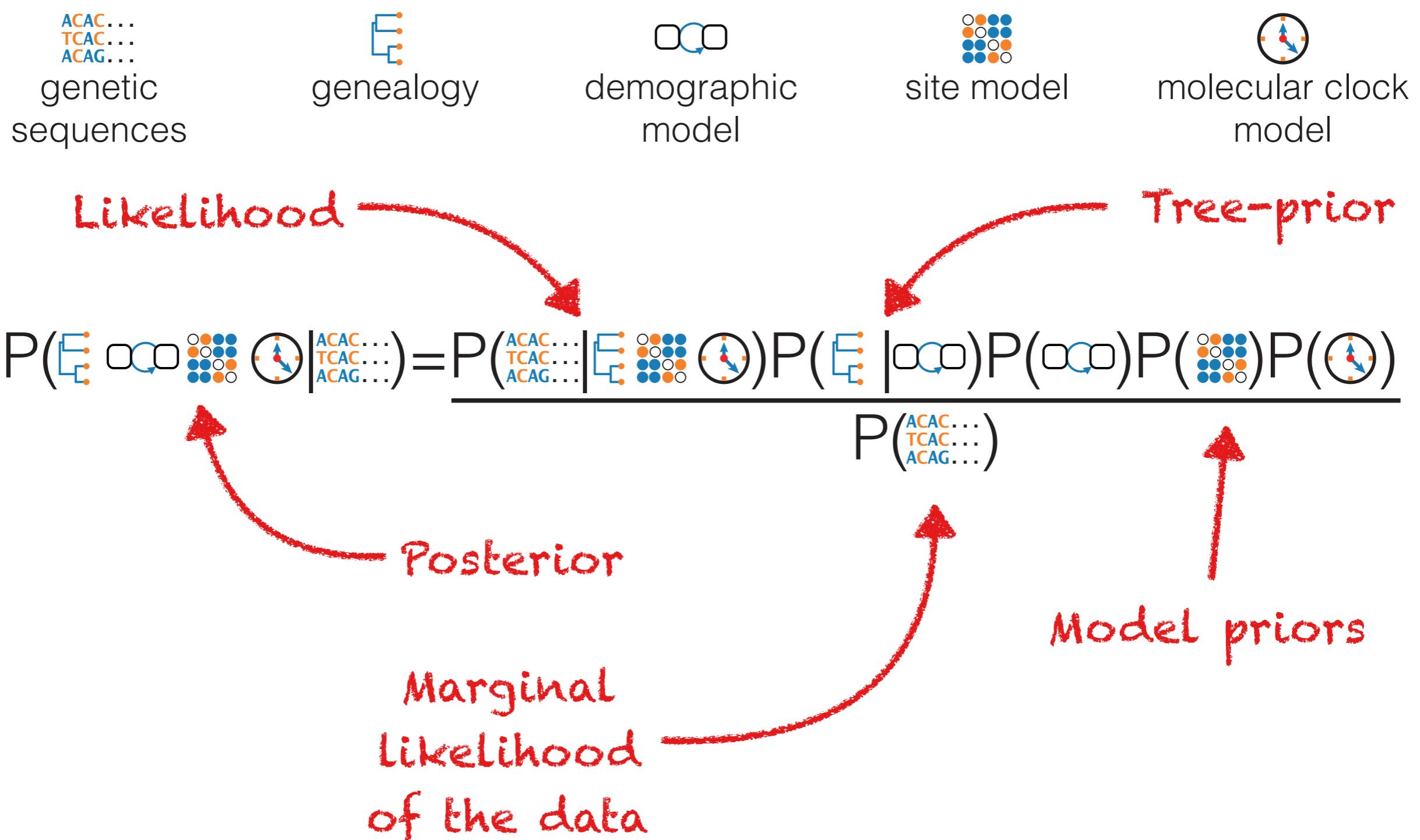


$$P(\text{Genealogy} \cap \text{Demographic Model} \cap \text{Site Model} \cap \text{Molecular Clock Model} | \text{Genetic Sequences}) = \frac{P(\text{Genetic Sequences} | \text{Genealogy} \cap \text{Demographic Model} \cap \text{Site Model} \cap \text{Molecular Clock Model}) P(\text{Genealogy} \cap \text{Demographic Model} \cap \text{Site Model} \cap \text{Molecular Clock Model})}{P(\text{Genetic Sequences})}$$

Assume independence

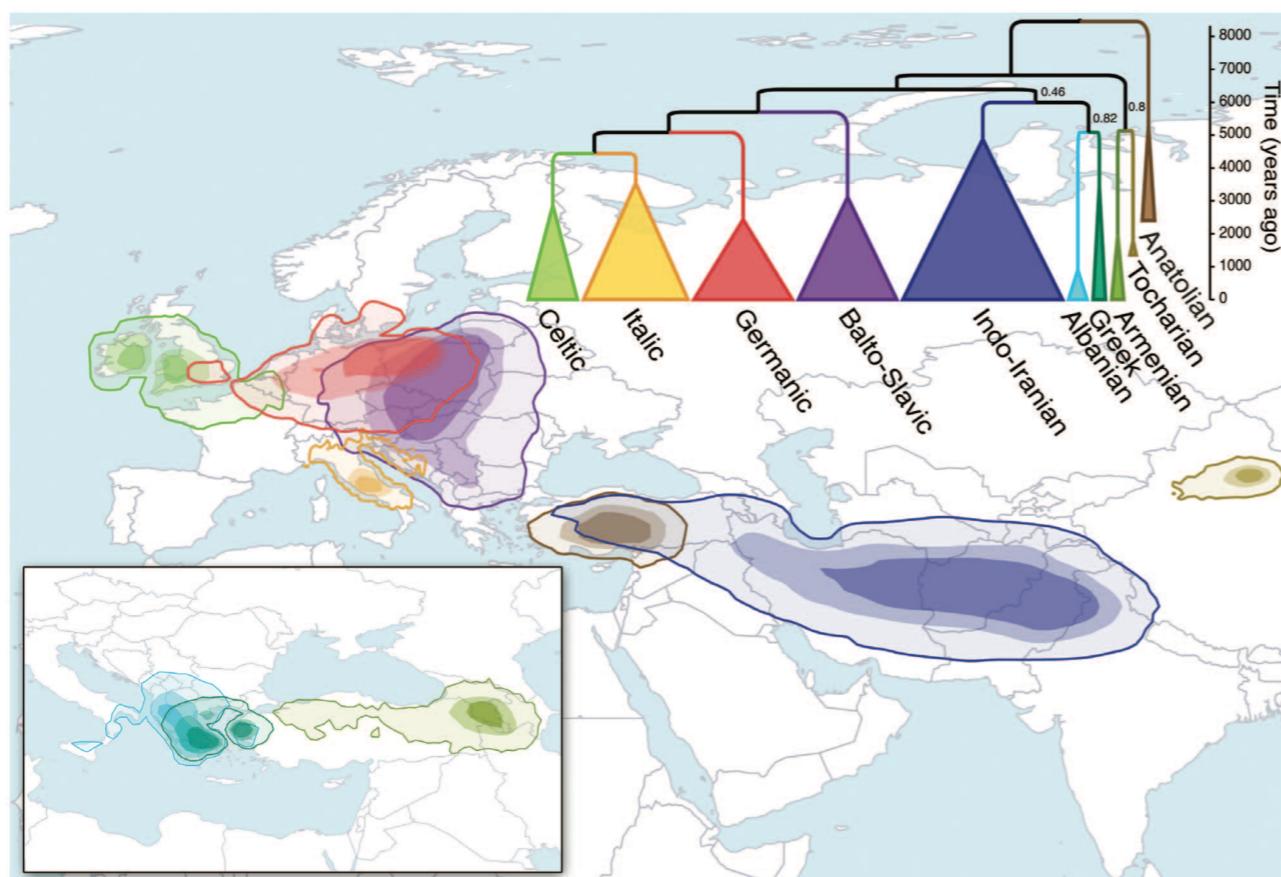
$$P(\text{Genealogy} \cap \text{Demographic Model} \cap \text{Site Model} \cap \text{Molecular Clock Model} | \text{Genetic Sequences}) = P(\text{Genealogy}) P(\text{Demographic Model}) P(\text{Site Model}) P(\text{Molecular Clock Model})$$

Posterior distribution in BEAST2



Exceptions I

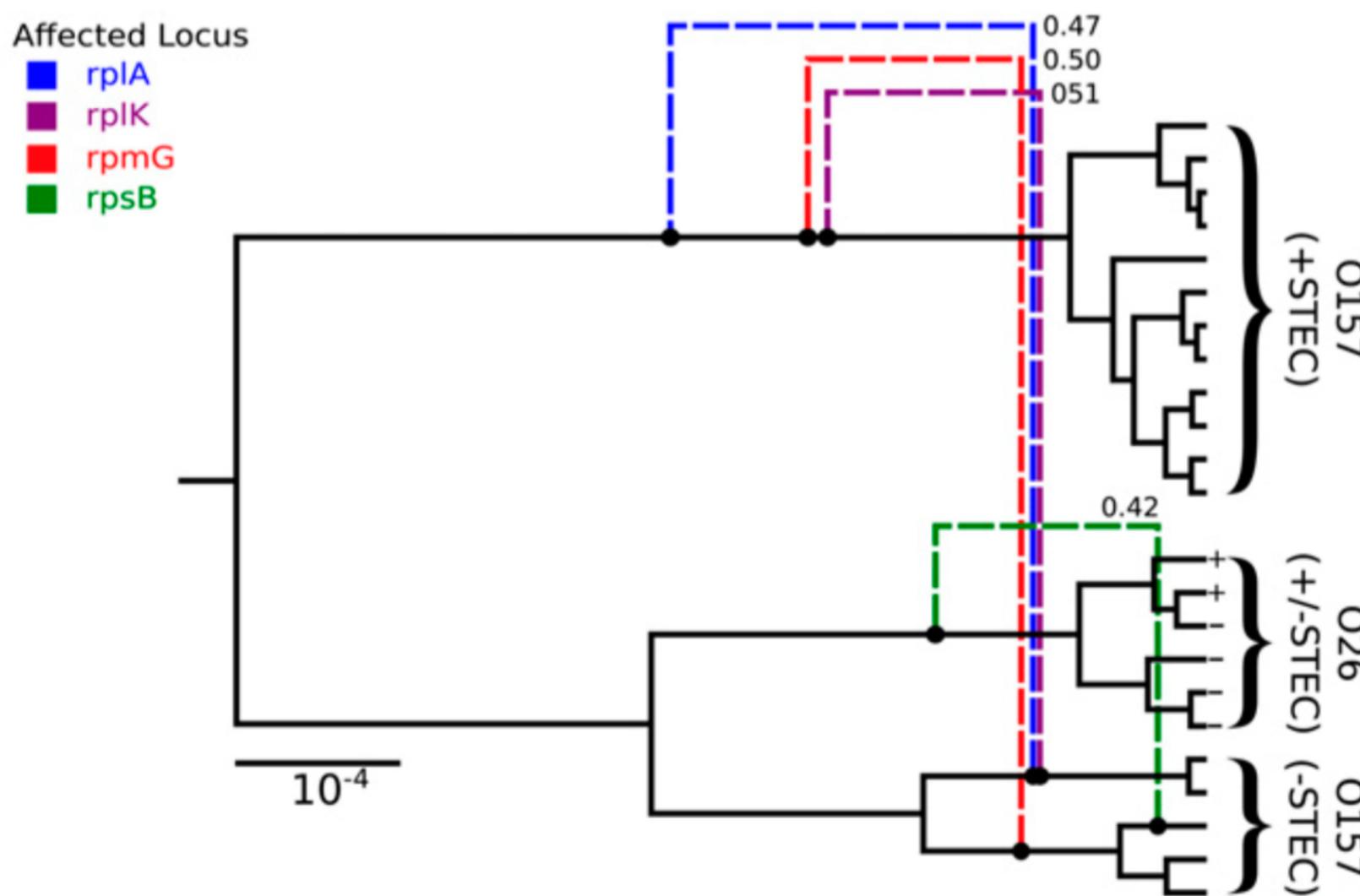
- Site models don't have to be on nucleotides
- Could be on amino acids, morphological traits, roots of words etc.



Bouckaert *et al.* **Science** 2012

Exceptions II

BEAST2 doesn't always use trees...



How can we find the posterior?

- We want to calculate the posterior distribution

$$P(\text{Emissions} \mid \text{Model}, \text{Sequence}) = \text{Posterior Distribution}$$


- But we cannot easily calculate the marginal likelihood
→ use **MCMC!** (Markov-chain Monte Carlo)

Markov-chain

- Stochastic process
- Jumps between different states
- Memoryless

Monte Carlo algorithm

- Randomized algorithm
- Deterministic runtime (it **will** finish)
- Output may **not** be correct (with some small probability)

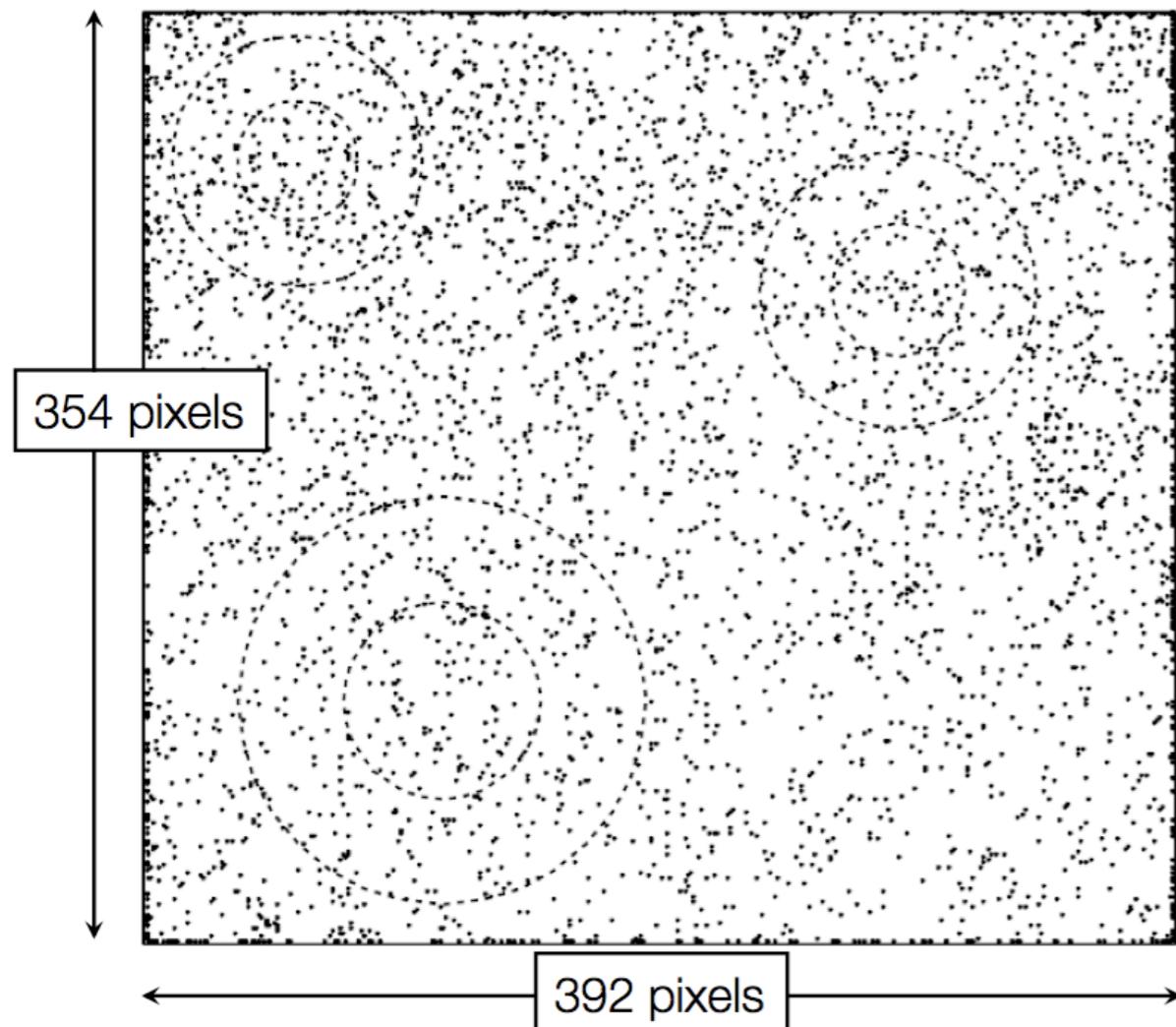
MCMC

(Markov-chain Monte Carlo)

- MCMC performs a random walk on the posterior, preferentially sampling high-density areas
- 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!

$$\frac{P(\text{model}_1 \mid \text{data})}{P(\text{model}_2 \mid \text{data})} = \frac{\frac{P(\text{data} \mid \text{model}_1)P(\text{model}_1)}{P(\cancel{\text{data}})}}{\frac{P(\text{data} \mid \text{model}_2)P(\text{model}_2)}{P(\cancel{\text{data}})}}$$

Pure random walk (courtesy of Paul Lewis)



Random walk

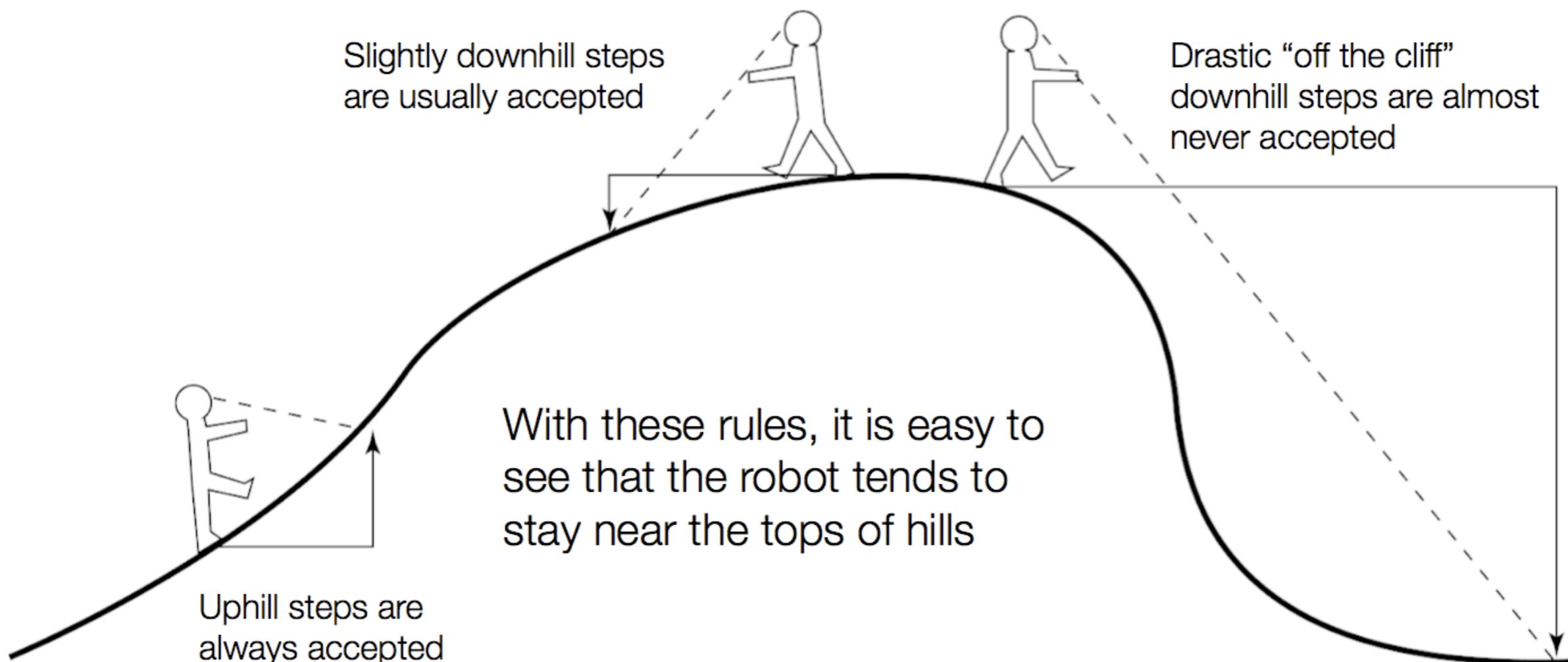
- Random direction
- Gamma distributed step size
- Reflection at edges

Target distribution

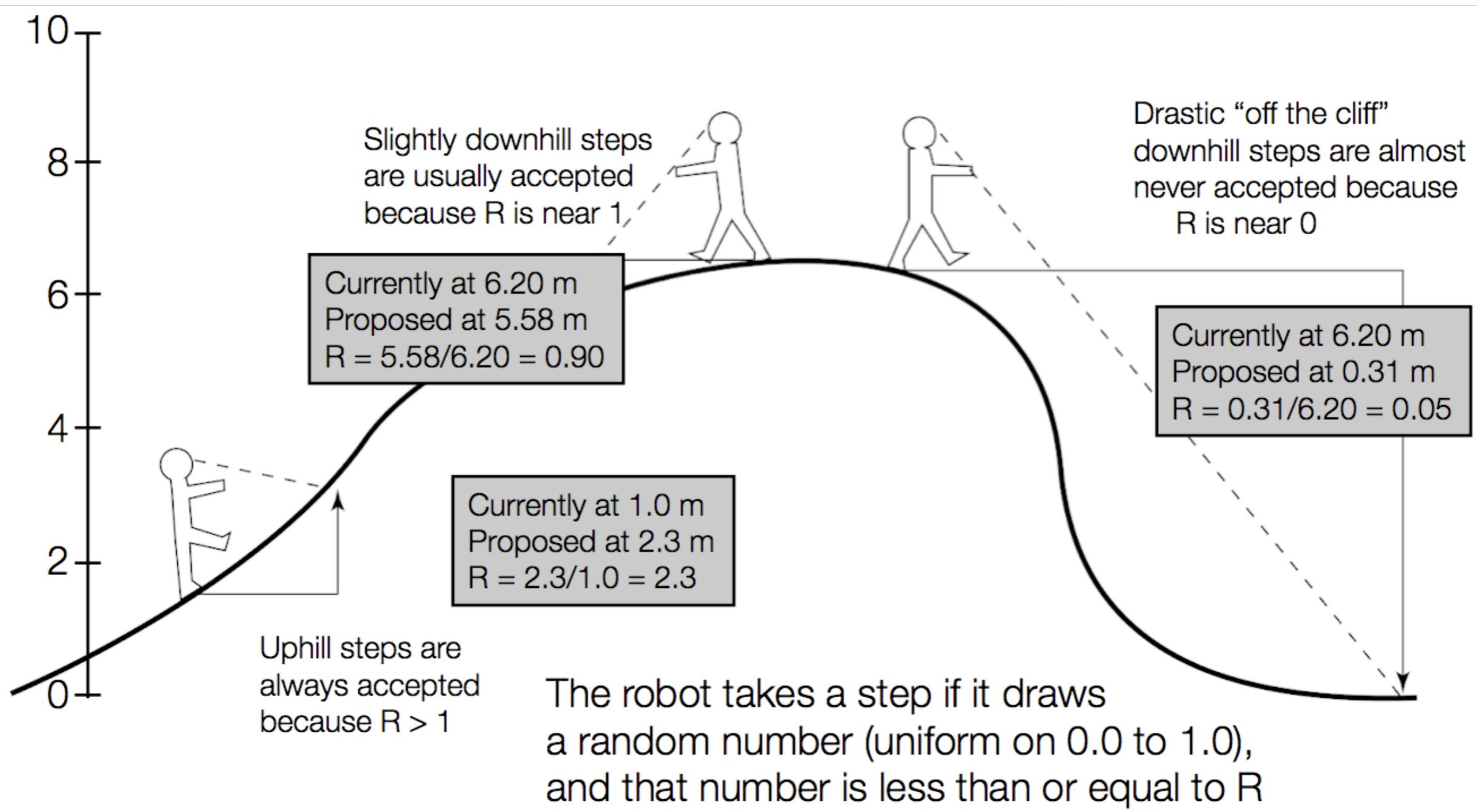
- Equal mixture of 3 bivariate normal hills
- Inner contours: 50%
- Outer contours: 95%

5000 steps by the random walk - not informative at all!

MCMC robot (courtesy of Paul Lewis)

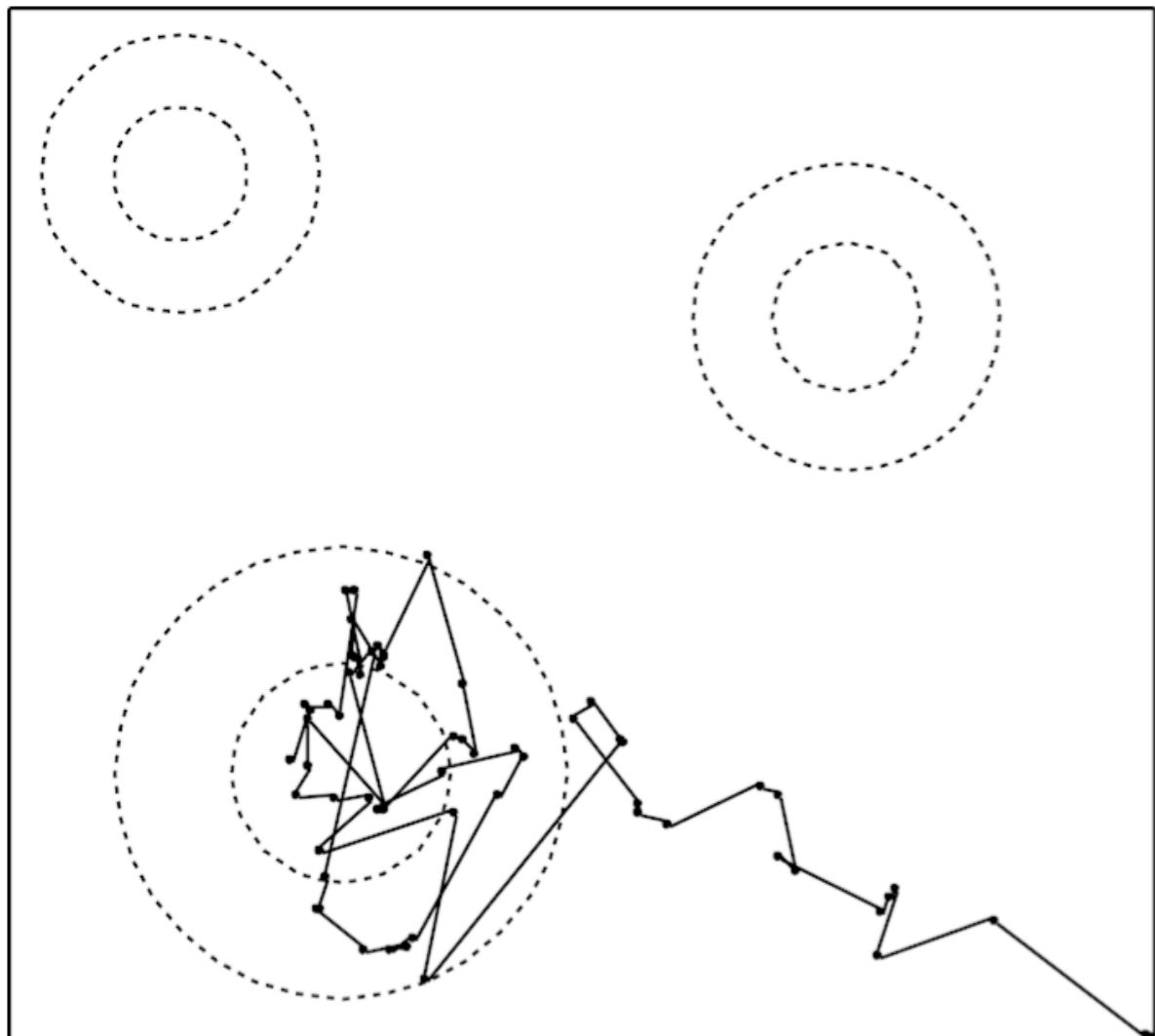


MCMC robot (courtesy of Paul Lewis)



(R is the ratio between the posterior densities)

Burn in (courtesy of Paul Lewis)

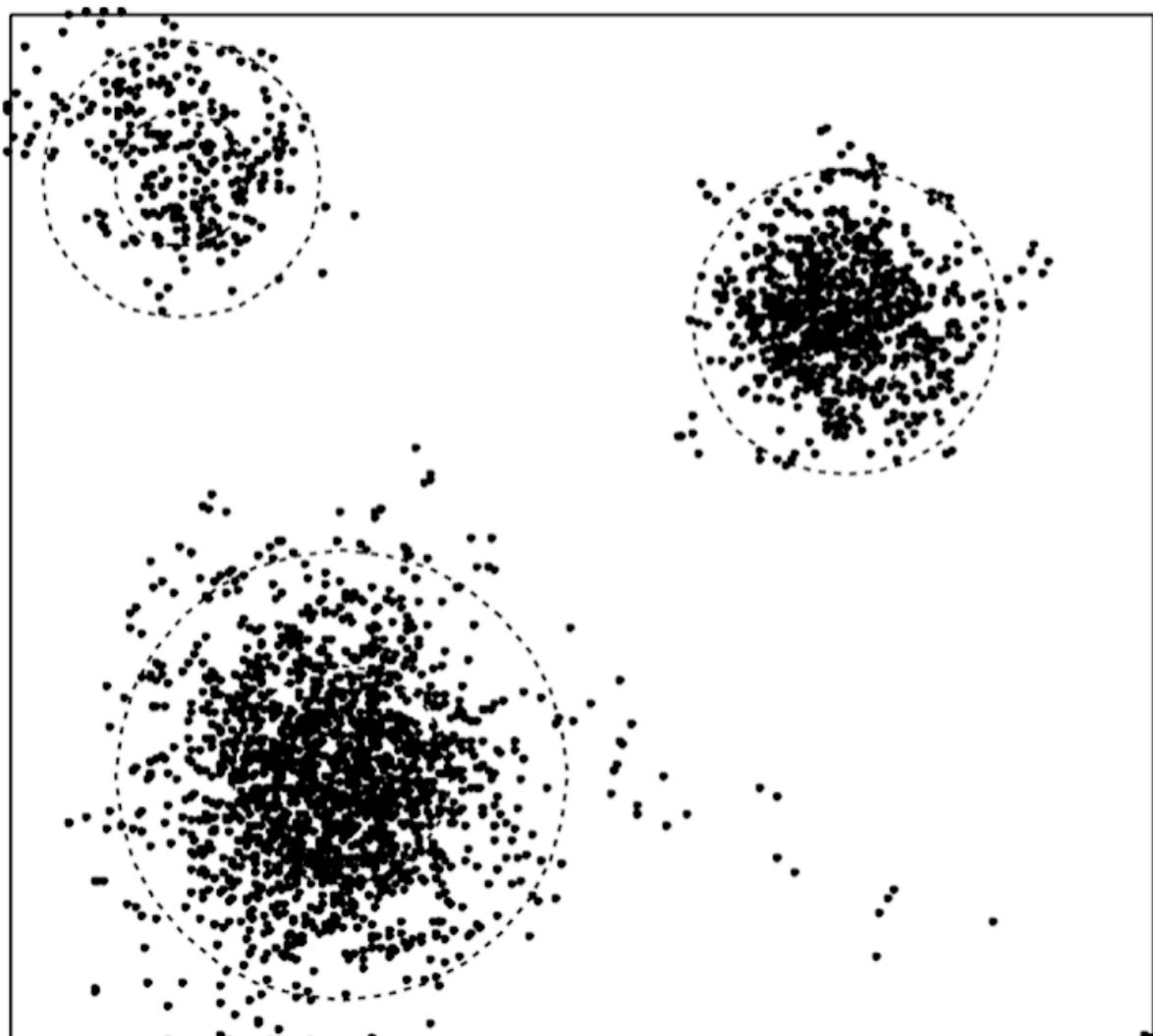


- Using MCMC rules the robot quickly finds one of the 3 hills
- First few steps are not representative of the distribution

First 100 steps by the robot

MCMC approximation

(courtesy of Paul Lewis)



How good is the approximation?

- 51.2% of points inside 50% contours
- 93.6% of points inside 95% contours

The more steps, the better the accuracy

5000 steps by the robot

Marginal distributions

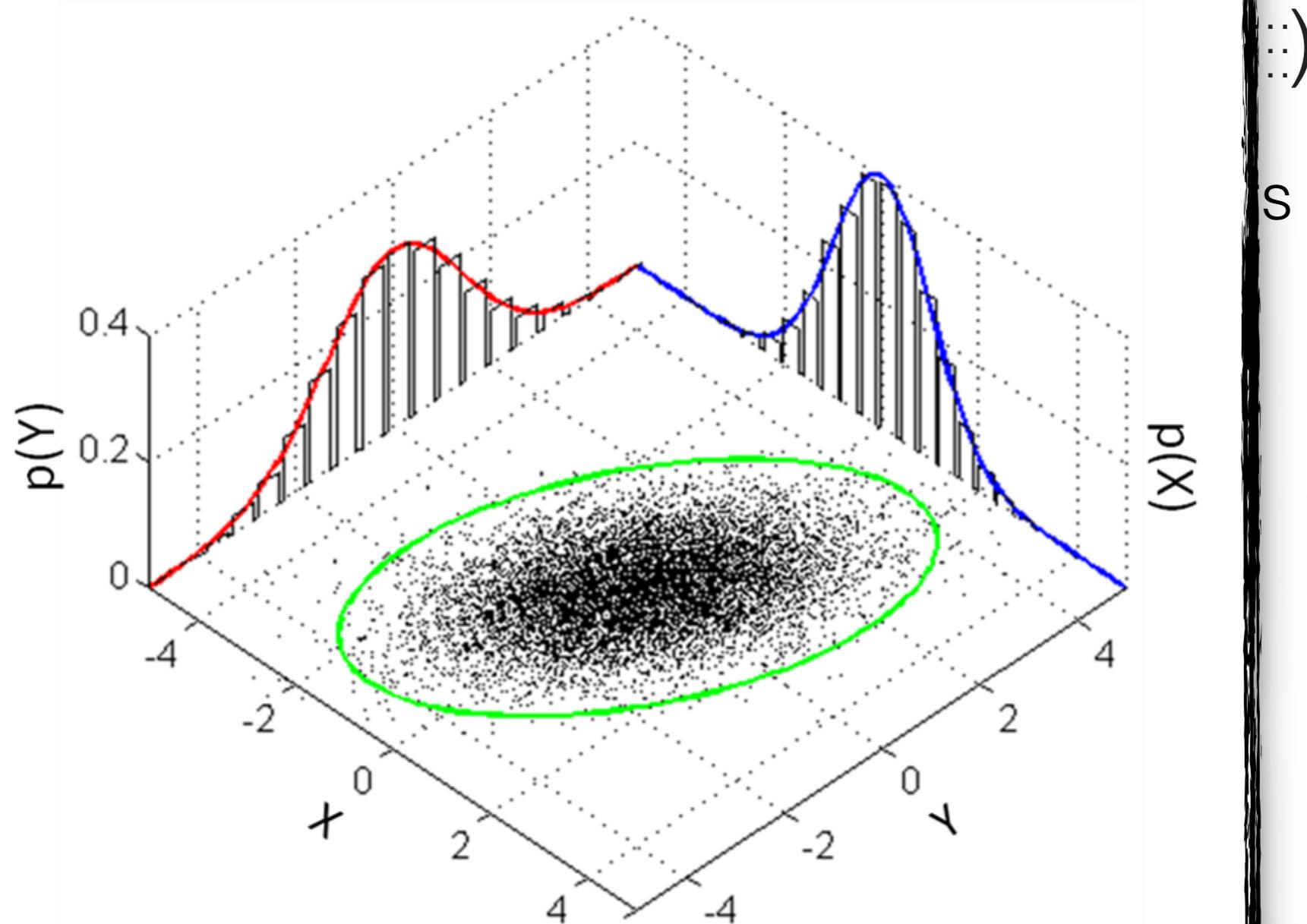
- We only have the joint posterior: $P(E \text{ } o \text{ } o \text{ } | \text{ } \theta)$
- But we want distributions for each of the parameters we are interested in → marginalize

$$P(\phi) = \int_{\Theta} P(\phi|\theta)P(\theta)d\theta$$

Margin

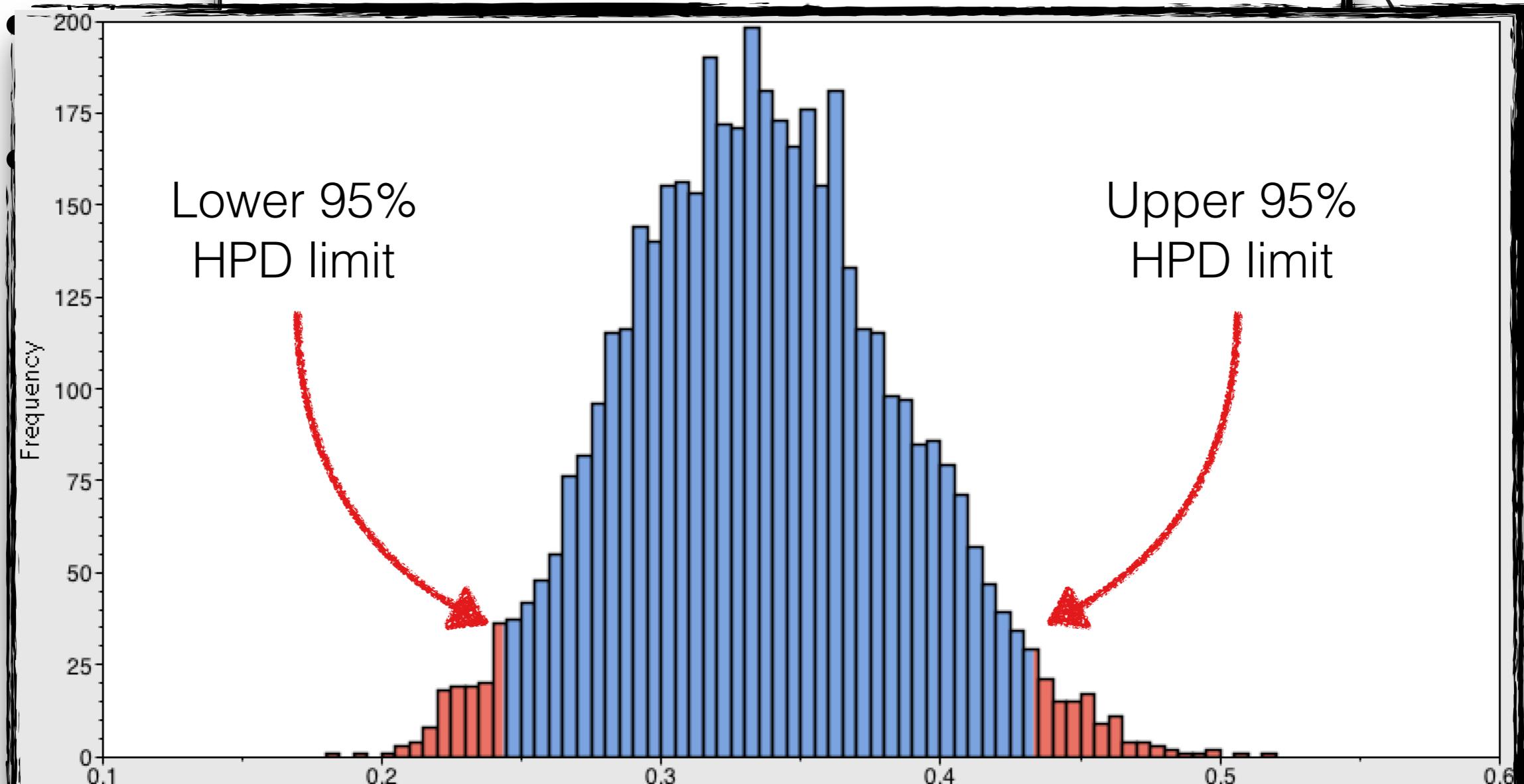
In practice

- We often
- But we
- we are



Margin

In practice



Operators

Target distribution

Proposal distribution

- Used to decide where to step to next
 - The choice only affects the **efficiency** of the algorithm
 - In BEAST and BEAST2 operators are used to propose the next step
 - Operators are a part of the MCMC **algorithm**, not the **model**
 - Tuning operators can help to improve efficiency, but should not change the results

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)

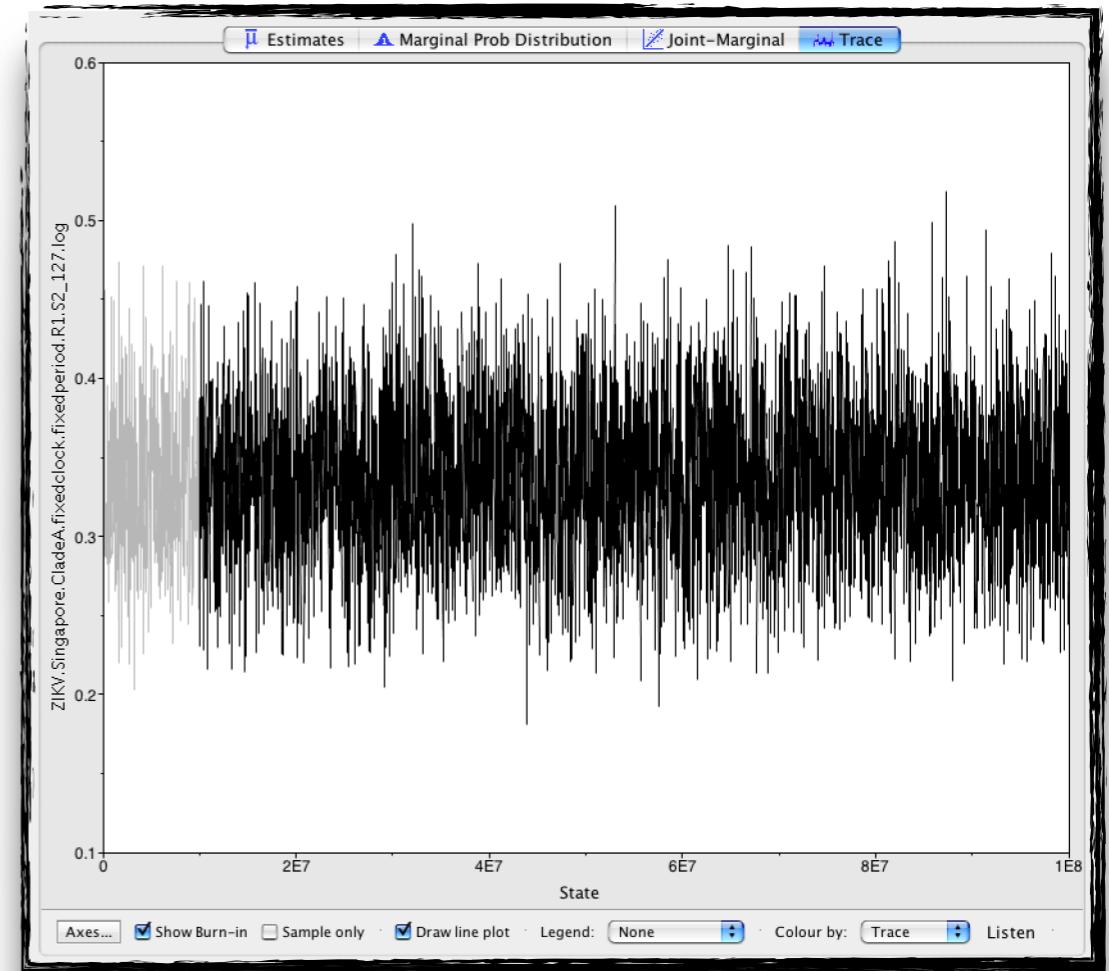
After

- Discard burn-in (until stationary state is reached)
- Assess convergence and mixing

More than 10,000 samples is a waste of space
(but need to sample at the right frequency)

What we hope will happen

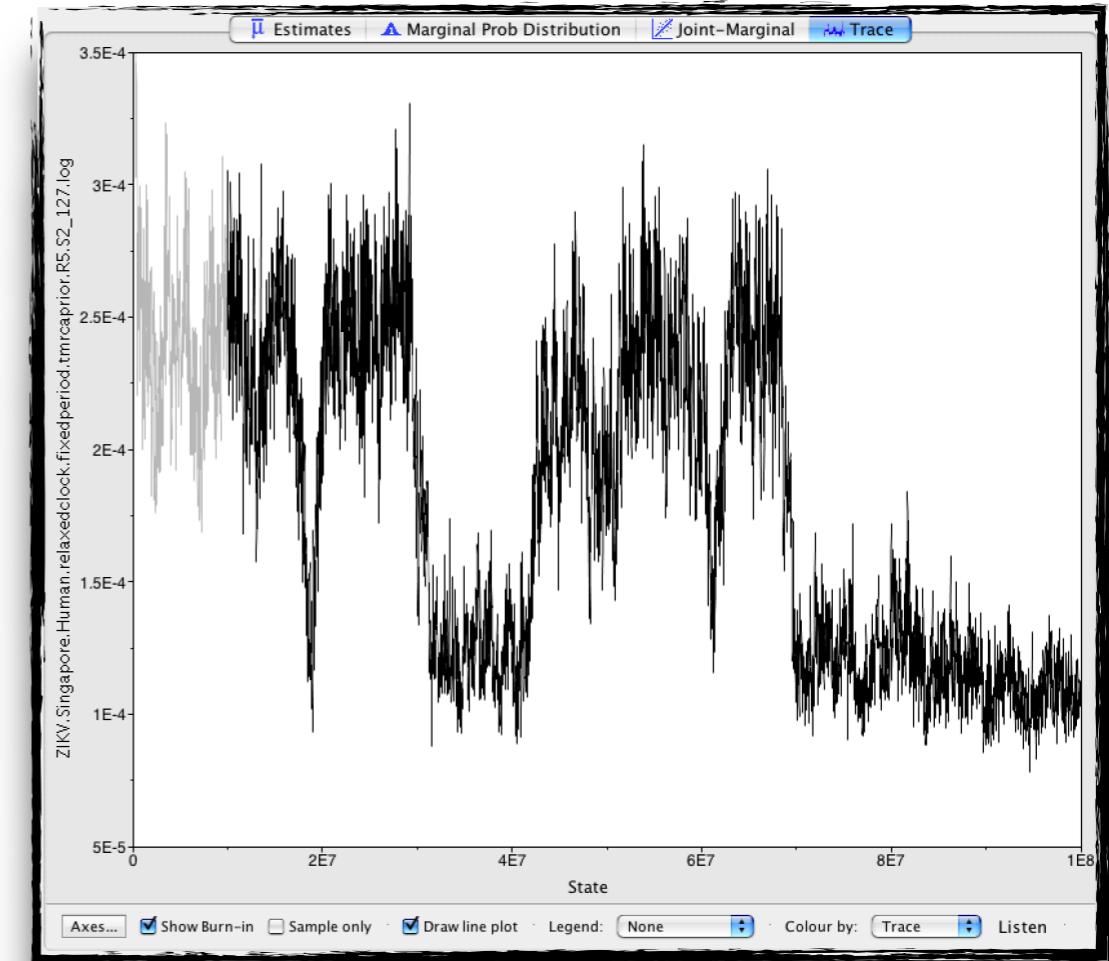
- The MCMC algorithm samples efficiently from high density areas of the posterior distribution
- We end up with a **good** approximation of the posterior distribution in **finite** time
- Appearance of white noise
- Everything is awesome!



Mixing well! 😊

Questions to ask...

- Is the chain **mixing** well?
- Are samples uniformly drawn from all over the stationary distribution?
- “Sticky chain”



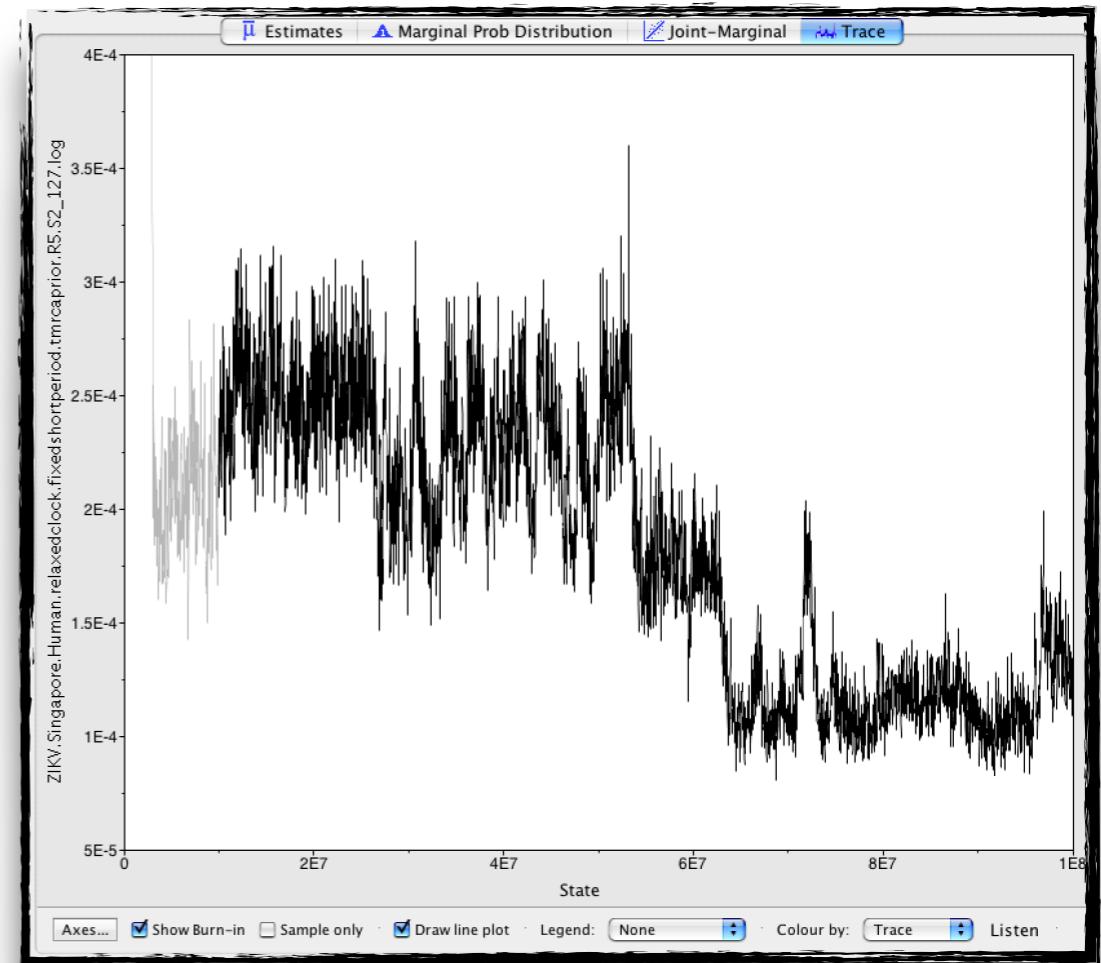
Solutions

- 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?

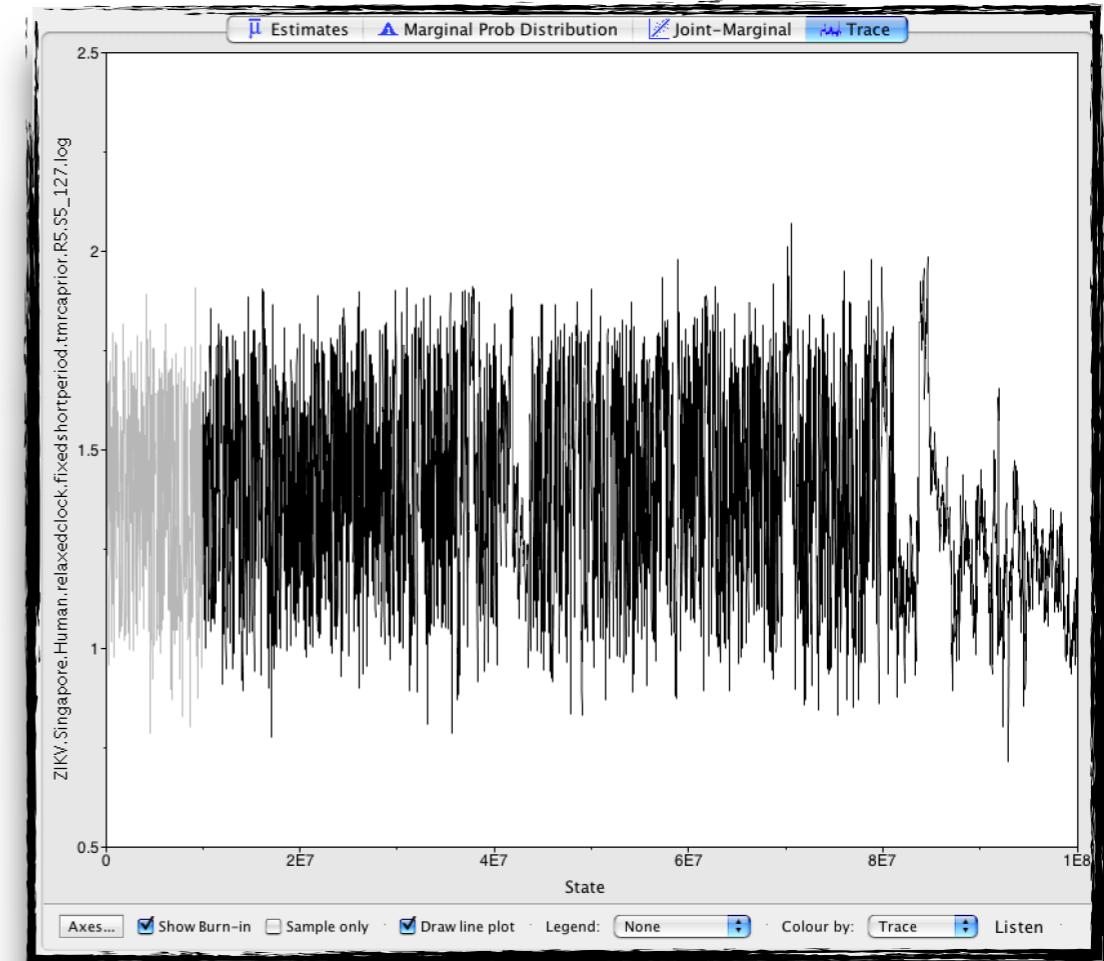


Not converged! 😓

Solution: Run for longer

Questions to ask...

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



Solution

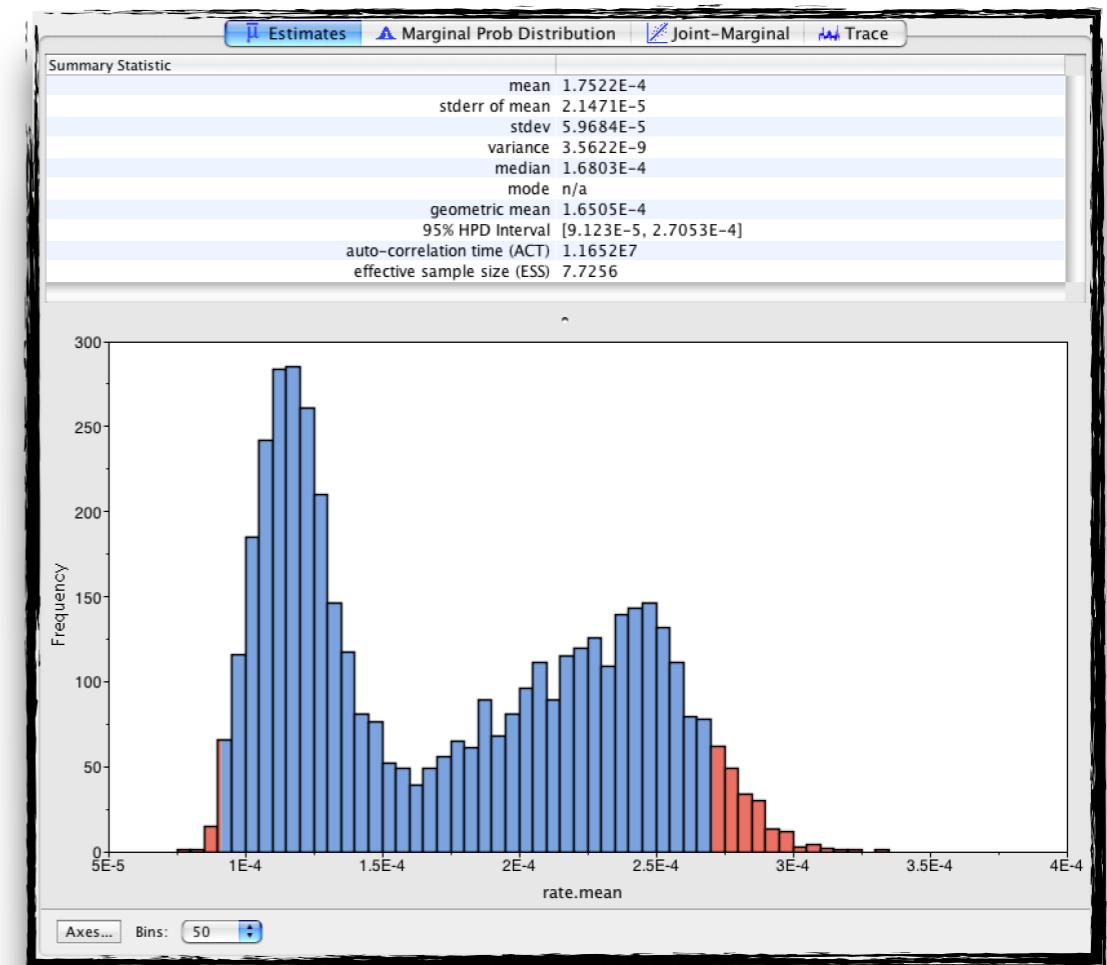
- Run multiple chains
- Combine chains
- Check that all chains give the same result

Still not converged! 😞

What if the answer is not what we wanted?

What is happening here?

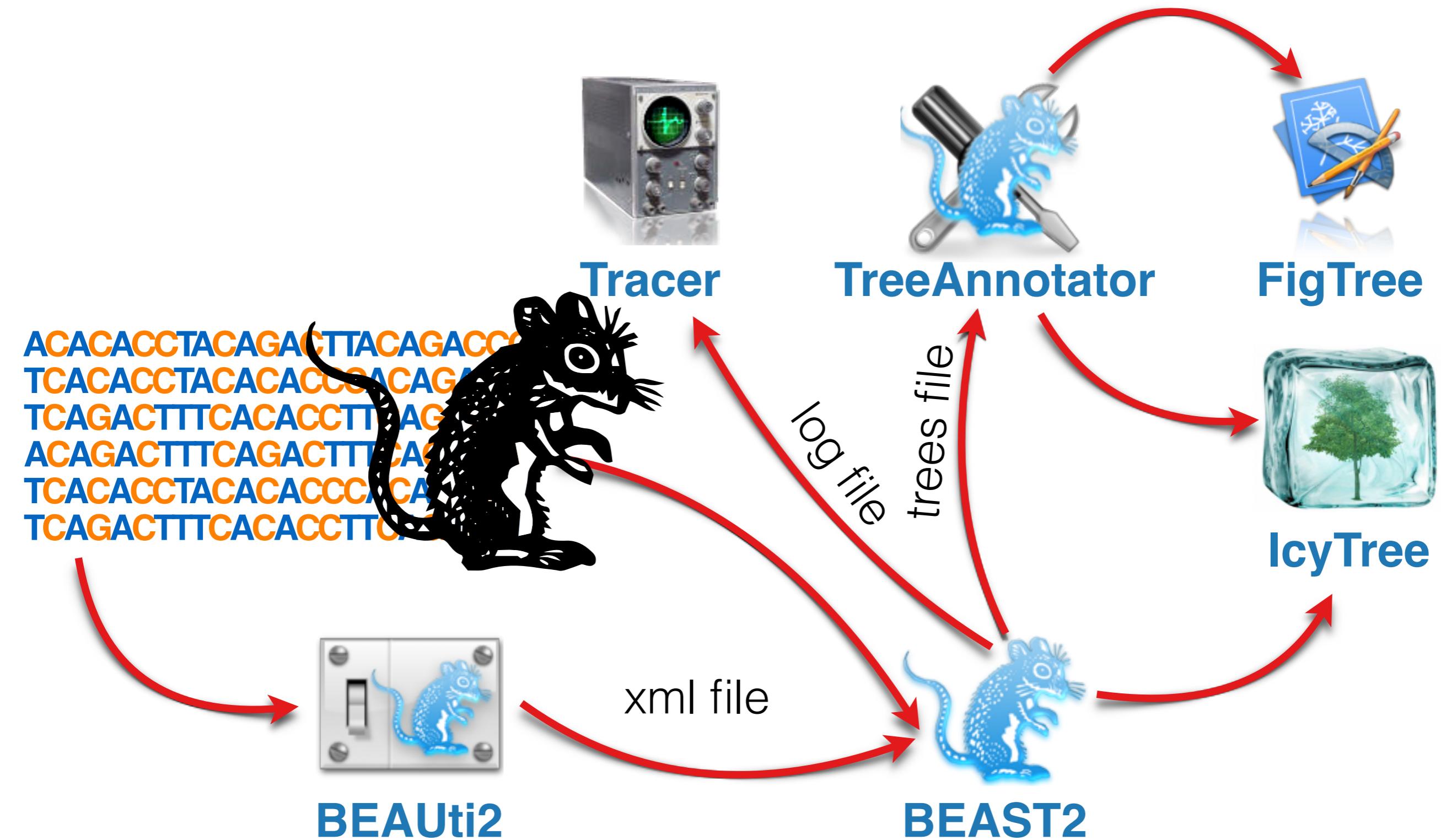
- If the chain converged and mixed well then this is due to the data and model choice
- The model supports a bimodal posterior distribution
- May not be the answer we wanted but it may be the truth
- Should we change the model or parameterisation?



Is this a problem? 🤔

Solution: Be more open-minded

BEAST2 workflow



BEAUti2

(<http://beast2.org>)



GUI for setting up BEAST2 input file in xml format

Input:

- Sequence alignment

Output:

- BEAST2 configuration file (xml file)

BEAUti 2: Standard /Users/louis/Documents/Taming_the_BEAST/Tutorials-Git/Introduction-to-BEAST2/xml/Primates.xml

Partitions Tip Dates Site Model Clock Model Priors MCMC

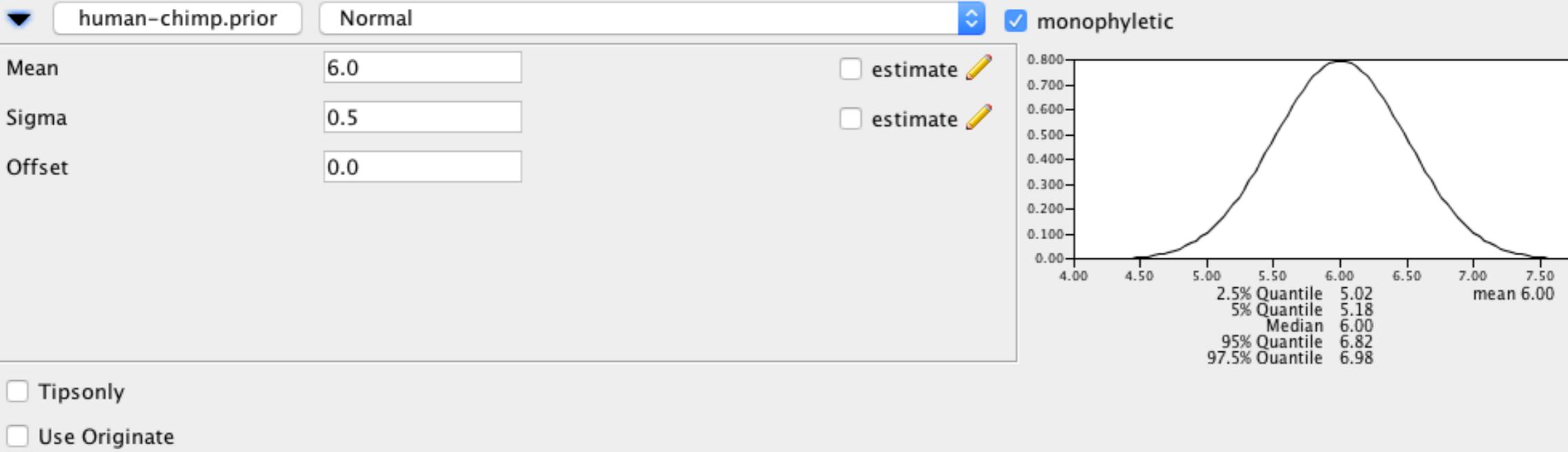
Link Site Models Unlink Site Models Link Clock Models Unlink Clock Models Link Trees Unlink Trees

Name	File	Taxa	Sites	Data Type	Site Model	Clock Model	Tree	...
noncoding	primate-mtDNA	12	205	nucleotide	noncoding	clock	tree	
1stpos	primate-mtDNA	12	231	nucleotide	1stpos	clock	tree	
2ndpos	primate-mtDNA	12	231	nucleotide	2ndpos	clock	tree	
3rdpos	primate-mtDNA	12	231	nucleotide	3rdpos	clock	tree	

+ - r Split

Partitions Tip Dates Site Model Clock Model Priors MCMC

▶ Tree.t:tree	Calibrated Yule Model		
▶ birthRateY.t:tree	Gamma	initial = [1.0] $[-\infty, \infty]$	Calibrated Yule speciation process birth rate for t:3rdpos
▶ clockRate.c:clock	Uniform	initial = [1.0] $[-\infty, \infty]$	substitution rate of partition c:3rdpos
▶ gammaShape.s:1stpos	Exponential	initial = [1.0] $[-\infty, \infty]$	Prior on gamma shape for partition s:1stpos
▶ gammaShape.s:2ndpos	Exponential	initial = [1.0] $[-\infty, \infty]$	Prior on gamma shape for partition s:2ndpos
▶ gammaShape.s:3rdpos	Exponential	initial = [1.0] $[-\infty, \infty]$	Prior on gamma shape for partition s:3rdpos
▶ gammaShape.s:noncoding	Exponential	initial = [1.0] $[-\infty, \infty]$	Prior on gamma shape for partition s:noncoding
▶ kappa.s:1stpos	Log Normal	initial = [2.0] $[0.0, \infty]$	HKY transition-transversion parameter of partition s:1stpos
▶ kappa.s:2ndpos	Log Normal	initial = [2.0] $[0.0, \infty]$	HKY transition-transversion parameter of partition s:2ndpos
▶ kappa.s:3rdpos	Log Normal	initial = [2.0] $[0.0, \infty]$	HKY transition-transversion parameter of partition s:3rdpos
▶ kappa.s:noncoding	Log Normal	initial = [2.0] $[0.0, \infty]$	HKY transition-transversion parameter of partition s:noncoding



Primates_long.xml UNREGISTERED

```
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 </run>
```

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 (epidemiological/speciation) model
- Similar to BEAST 1.8 but completely separate
- BEAST2 has most of the functionality of BEAST 1.8 and a lot more
- BEAST2 has a modular design that makes it easy to extend

Input:

- xml file

Outputs:

- log file
- trees file
- state file

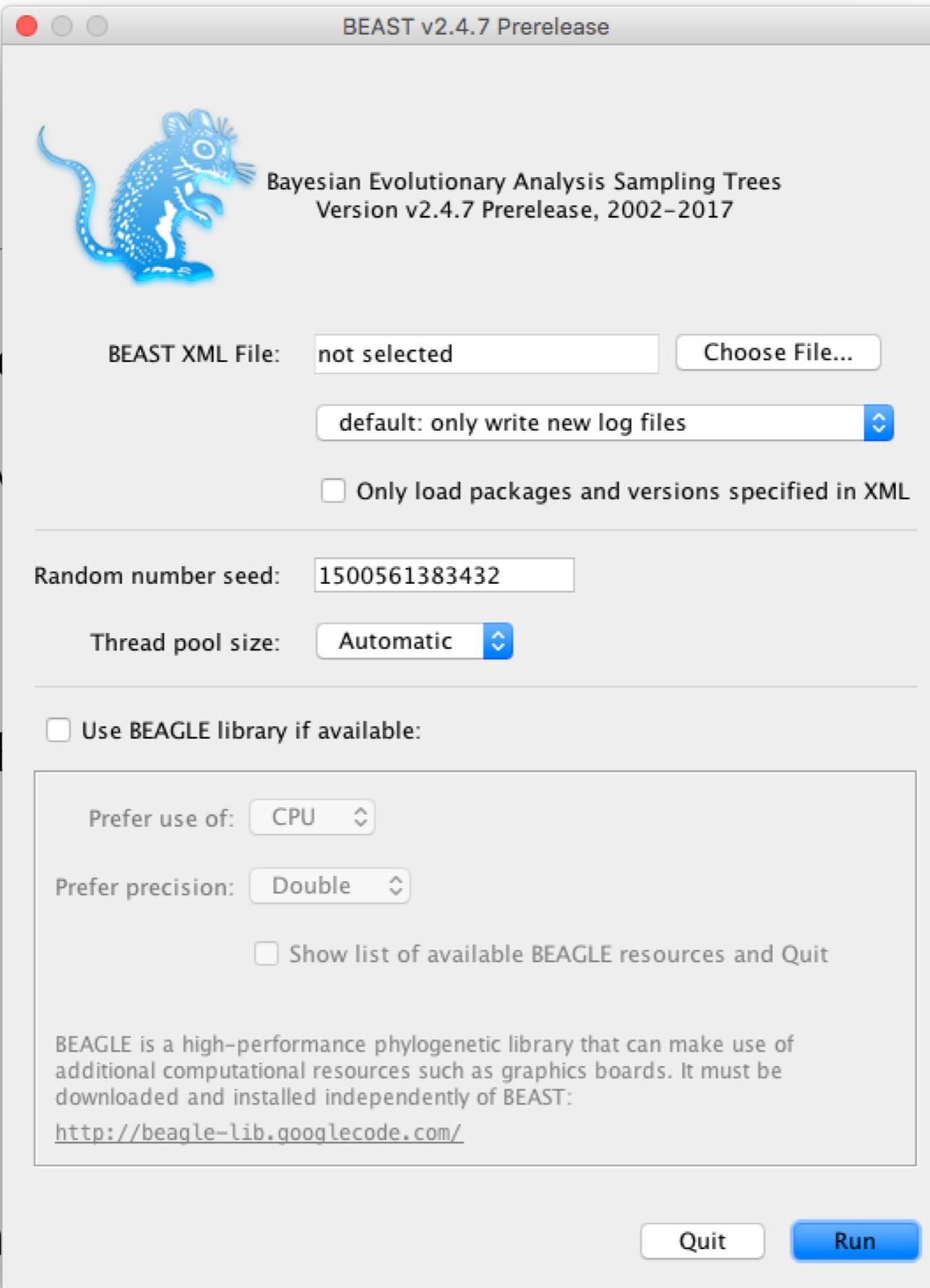
BEAST2

(<http://beast2.org>)

- Bayesian coalescent framework
- Performs MCMC sampling of phylogenetic trees under a sequence evolution model
- Similar to JModelTest
- BEAST2 has more features
- BEAST2 has more features

Input

- XML files



- trees file
- state file



selected
speciation)

8 and a lot

to extend

BEAST2 packages



- Independent researchers can easily develop their own BEAST2 packages
- Packages can be frequently updated without waiting for the next BEAST2 release
- Packages add new models or completely new functionality
- Phylogeography, bacterial ARG inference, morphological models, model selection and averaging, stochastic simulations etc.
- Install new packages through BEAUti

BEAST2 packages



- Independent BEAST2 packages for the next release
- Packages for the next release
- Packages for the next release
- Phylogenetic models for simulations etc.
- Install new packages through BEAUTi

BEAST 2 Package Manager

List of available packages for BEAST v2.4.*

Name	Installed	Latest	Dependencies	Link	Detail
BEAST	2.4.7	2.4.7			BEAST core
bacter	1.2.1	1.2.1			Bacterial ARG inference.
BASTA		2.3.1			Bayesian structured coalescent approximation
bdmm	0.2.0	0.2.0	MultiTypeTree		pre-release of multitype birth-death model (aka birth-death skyline - handles serially sampled tips, piec...
BDSKY	1.3.3	1.3.3			birth death skyline - handles serially sampled tips, piec...
BEAST_CLASSIC	1.3.0	1.3.0	BEASTLabs		BEAST classes ported from BEAST 1 in wrappers
BEASTLabs	1.7.0	1.7.1			BEAST utilities, such as Script, multi monophyletic c...
BEASTShell		1.3.0			BEAST Shell - BeanShell scripting for BEAST
BEASTvntr		0.1.1			Variable Number of Tandem Repeat data, such as micro...
bModelTest	1.0.4	1.0.4	BEASTLabs		Bayesian model test for nucleotide subst models, gamm...
CA		1.2.1			CladeAge aPackage for fossil calibrations
DENIM		0.3.0			Divergence Estimation Notwithstanding ILS and Migration
Epilnf		5.0.1	SA		Inference of epidemic trajectories
GEO_SPHERE		1.1.2	BEASTLabs		Whole world phylogeography
Mascot		0.0.2			Marginal approximation of the structured coalescent
MASTER		5.1.1			Stochastic population dynamics simulation
MGSM		0.2.1			Multi-gamma and relaxed gamma site models
MM		1.0.5			Enables models of morphological character evolution
MODEL_SELECTION		1.3.4	BEASTLabs		Select models through path sampling/stepping stone an...

Latest [Install/Upgrade](#) [Uninstall](#) [Package repositories](#) [Close](#) ?

Tracer

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



- Analyse log files from BEAST2 runs
- Check mixing, ESS, ACT, parameter correlations
- Overview of posterior parameter estimates
- Comparisons of several analyses

Input:

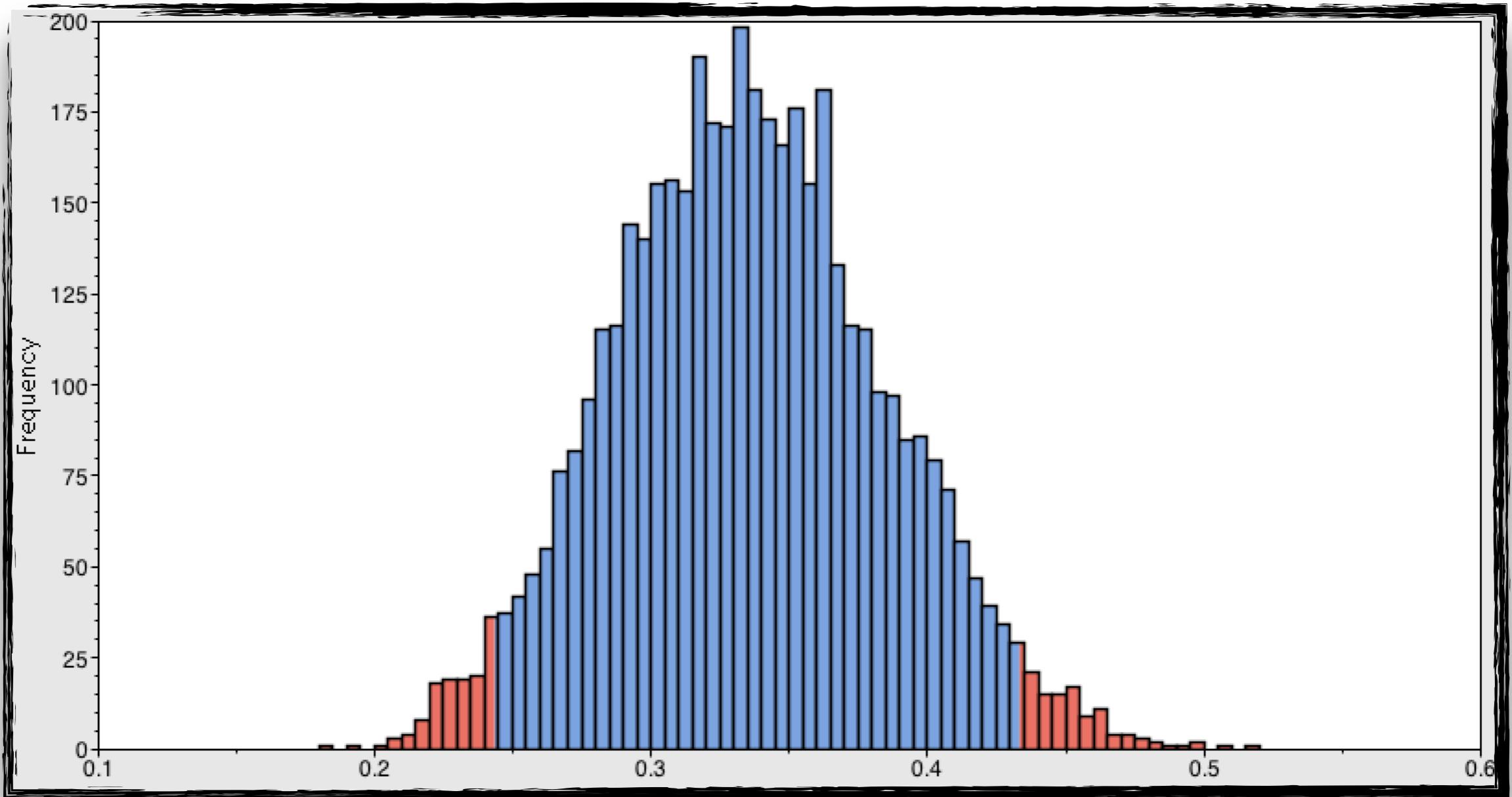
- log file

Output:

- Gain insight

Tracer

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

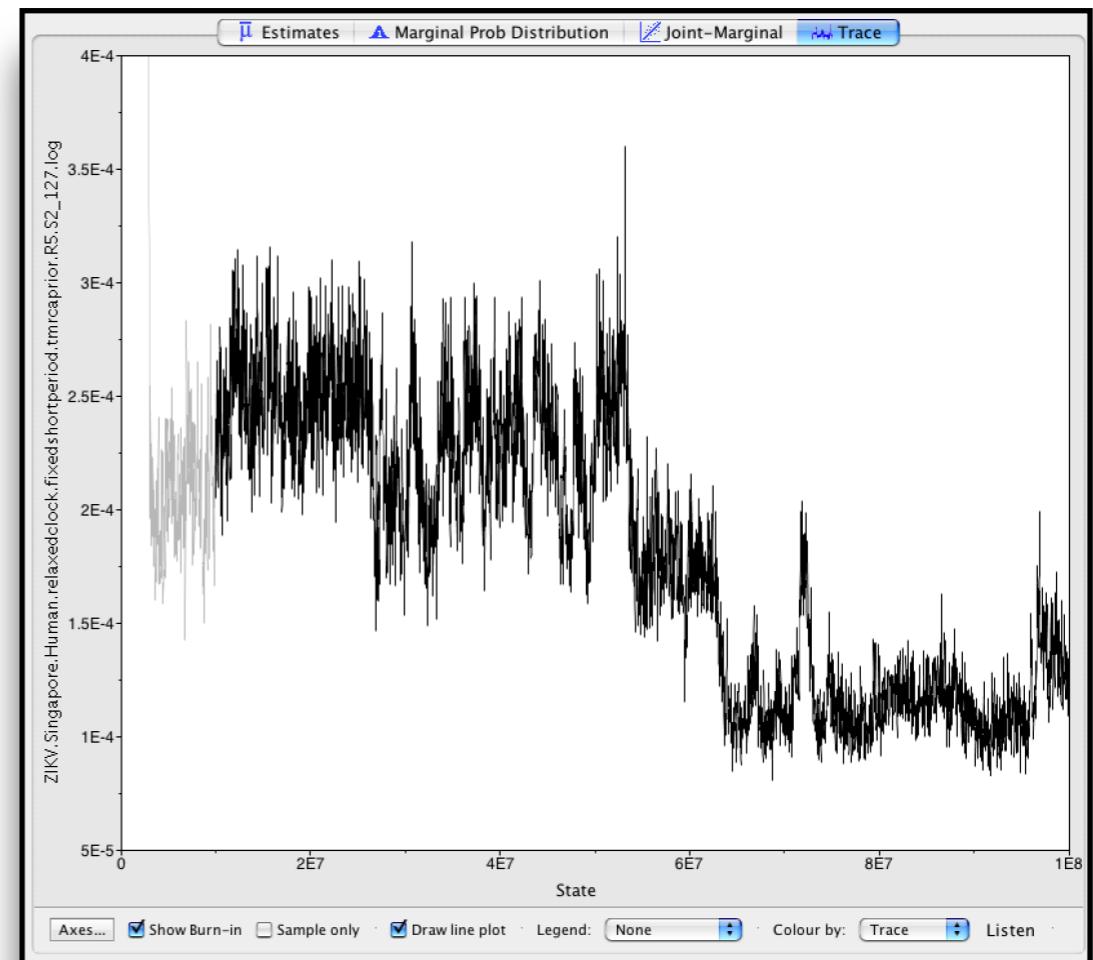
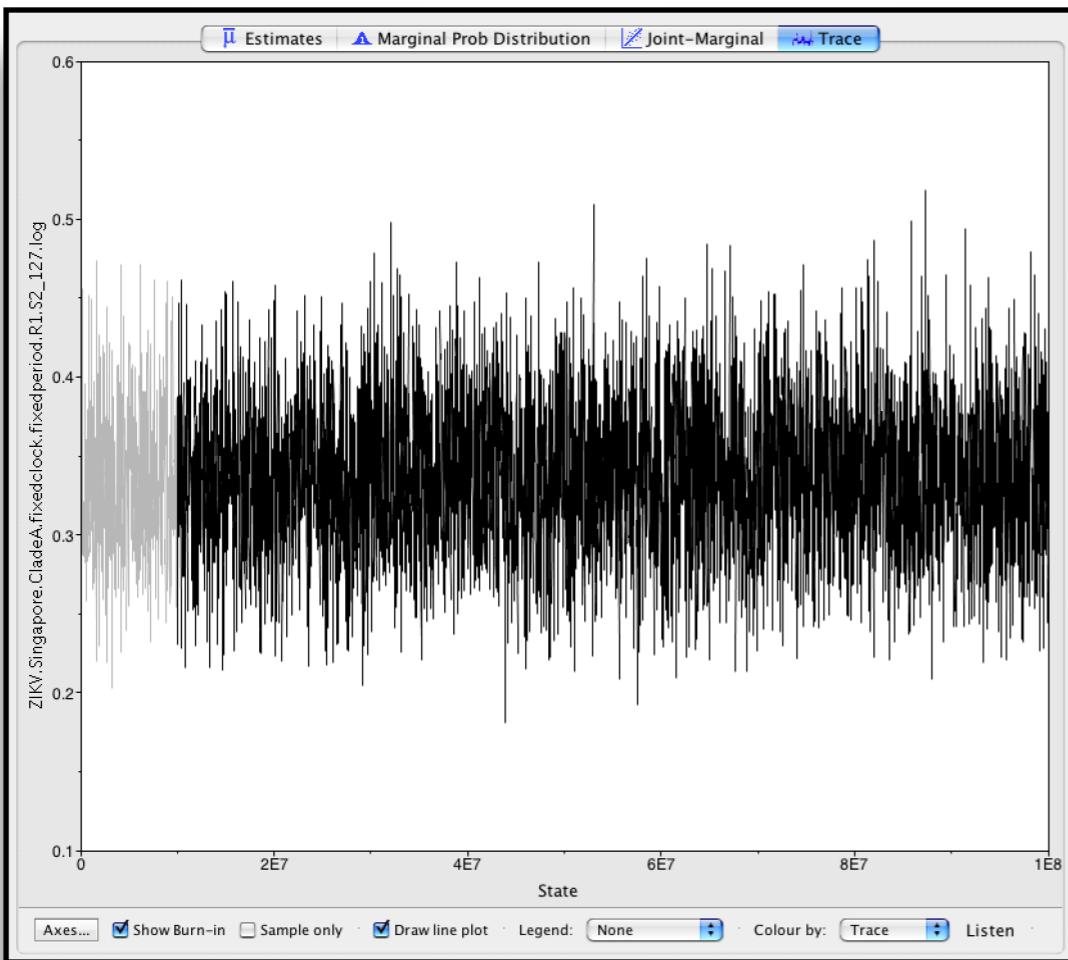


Tracer

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



Look at the chains first!



Mixing well! 😊

Not mixing! 😢



TreeAnnotator

(Included with BEAST2)

- Analyse trees file from BEAST2 runs
- Produces MCC tree with node annotations (posterior probability)
- Note that the MCC tree is just a summary and may never actually appear in the trees file!

Input:

- trees file
(many trees)

Output:

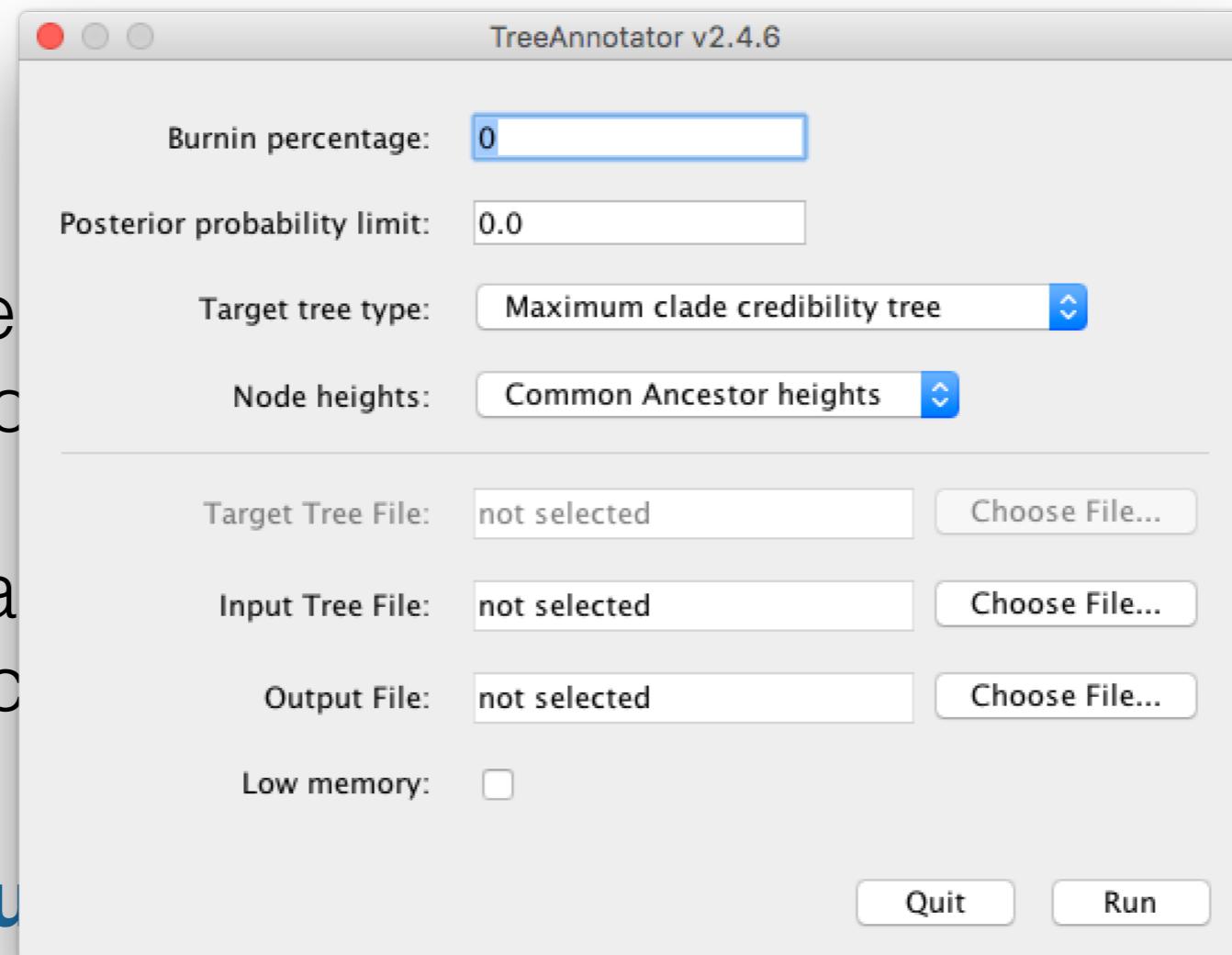
- MCC tree
(one tree)



TreeAnnotator

(Included with BEAST2)

- Analyse
- Produce
(posterior)
- Note that
never ac



Input

- trees file
(many trees)
- MCC tree
(one tree)

FigTree

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



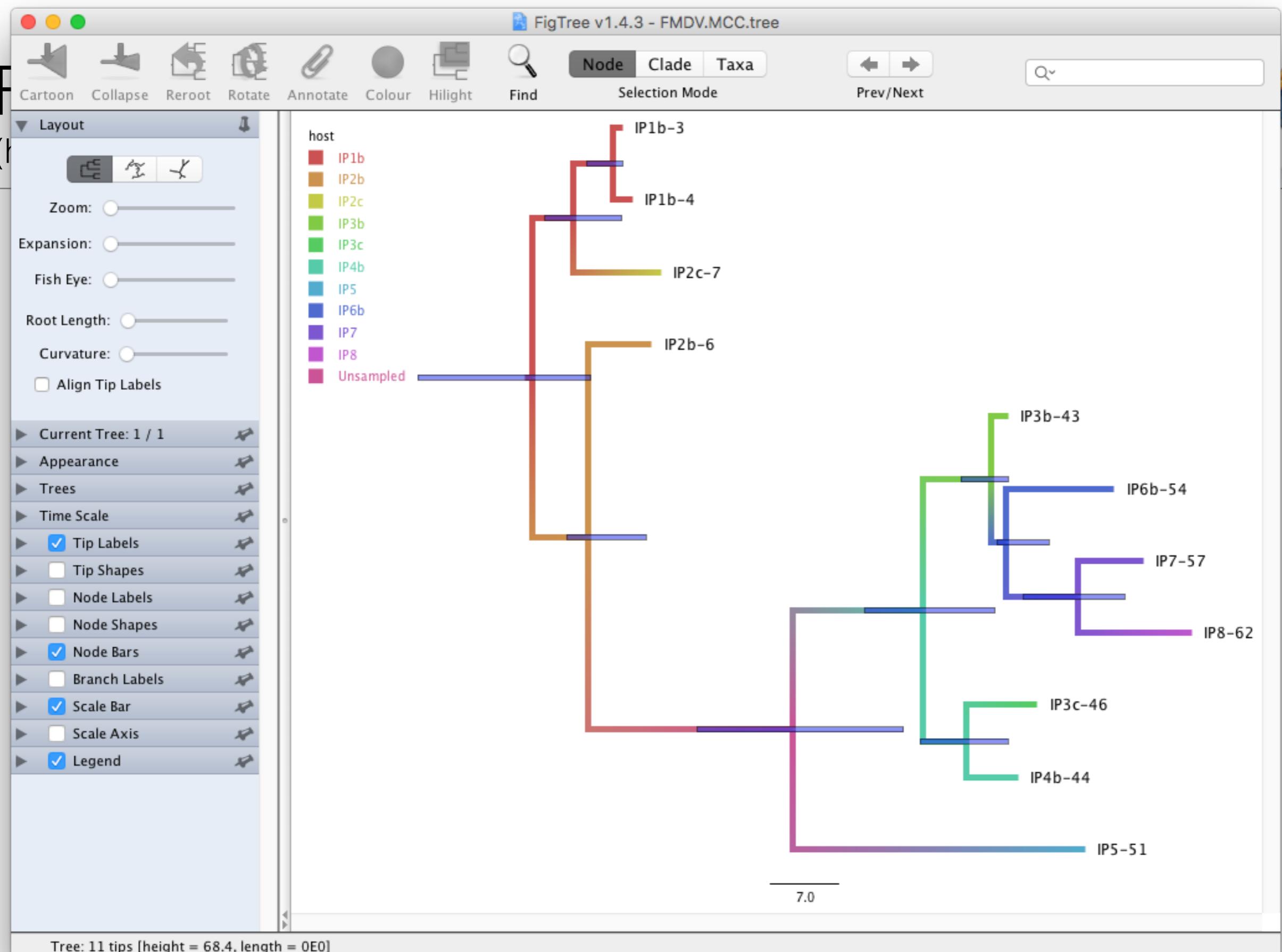
- Visualise trees from BEAST2 runs
- Annotate branches and nodes with probabilities and labels

Input:

- trees file

Output:

- Insight
- Figures



IcyTree

(<https://icytree.org>)



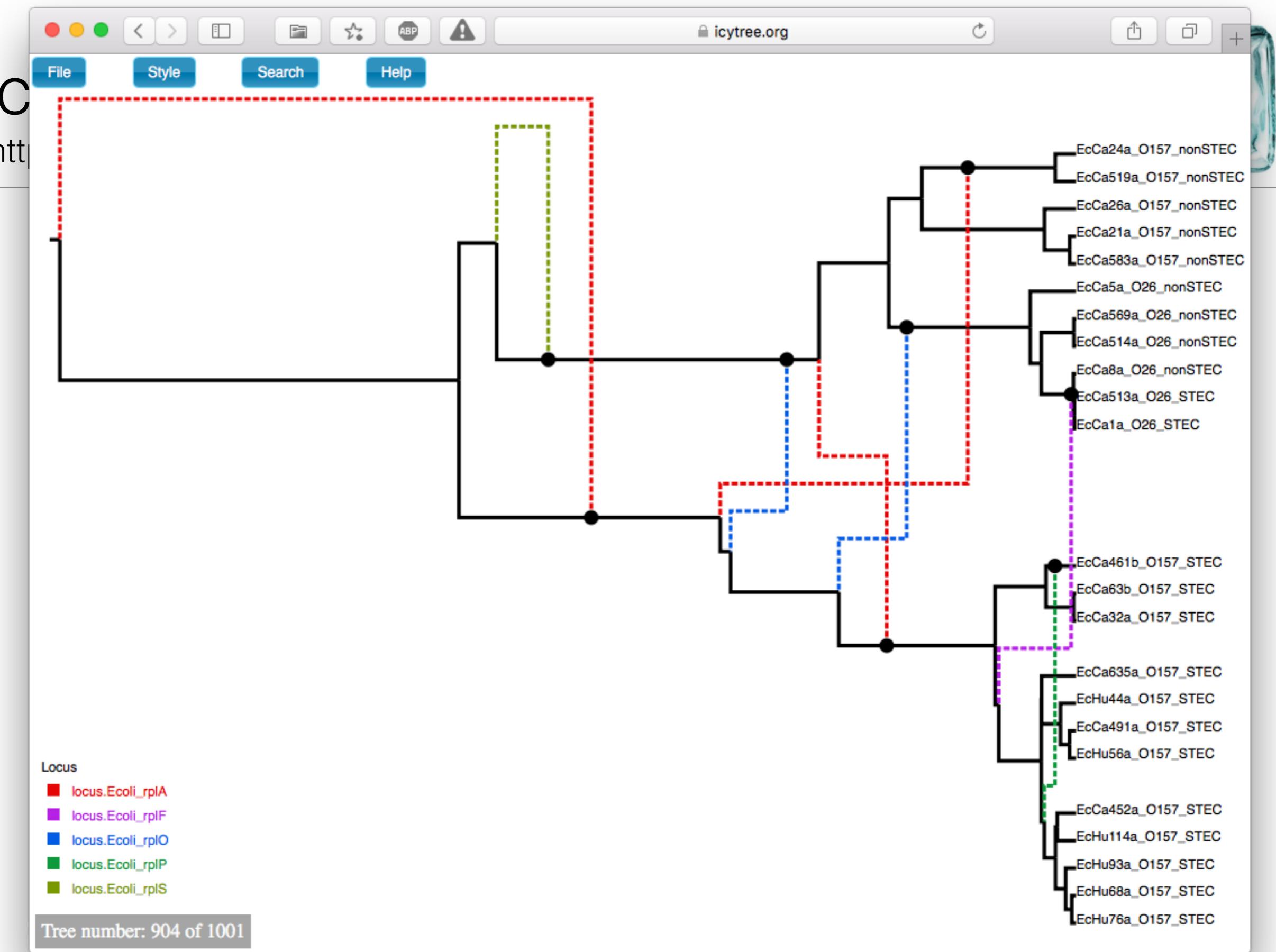
- Visualise trees from BEAST2 runs
- Annotate branches and nodes with probabilities and labels
- Better suited for structured models and ancestral recombination graphs (ARGs)
- Faster than FigTree for analysing many trees

Input:

- trees file

Output:

- Insight
- Figures



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 and BEAST2 output files (.log)

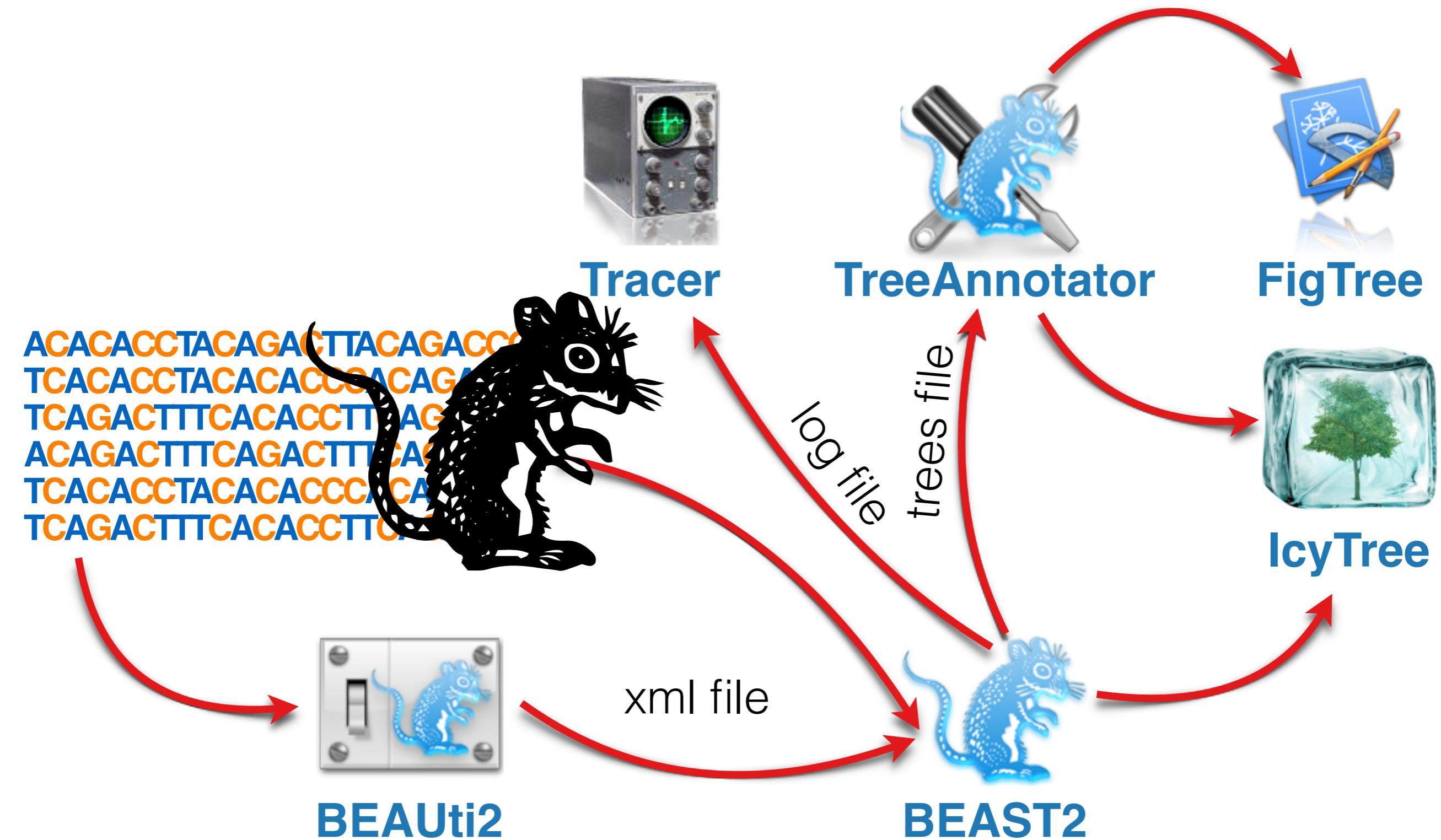
TreeAnnotator

Analysis of BEAST2 output files (.trees)

FigTree and IcyTree

Visualisation of trees (.trees)

BEAST2 workflow



BEAST best practice

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

Before you begin

- 1) Know your data
- 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



Thanks for listening!

Some slides inspired by (or shamelessly copied from) slides by Paul Lewis and Carsten Magnus

Questions?