

Taming the BEAST Online

FIRST STEPS IN BEAST

Slides adapted from a presentation by Louis du Plessis



By Ugnè Stoltz

We all have one thing in common...

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

BUT HOW?



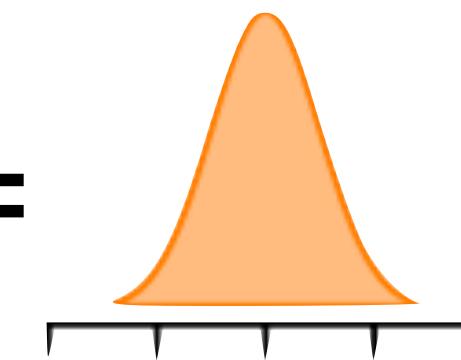
Bayesian
Evolutionary
Analysis by
Sampling
Trees
2

We all have one thing in common...

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

BUT HOW?

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



Estimate posterior probability
distribution of our model parameters

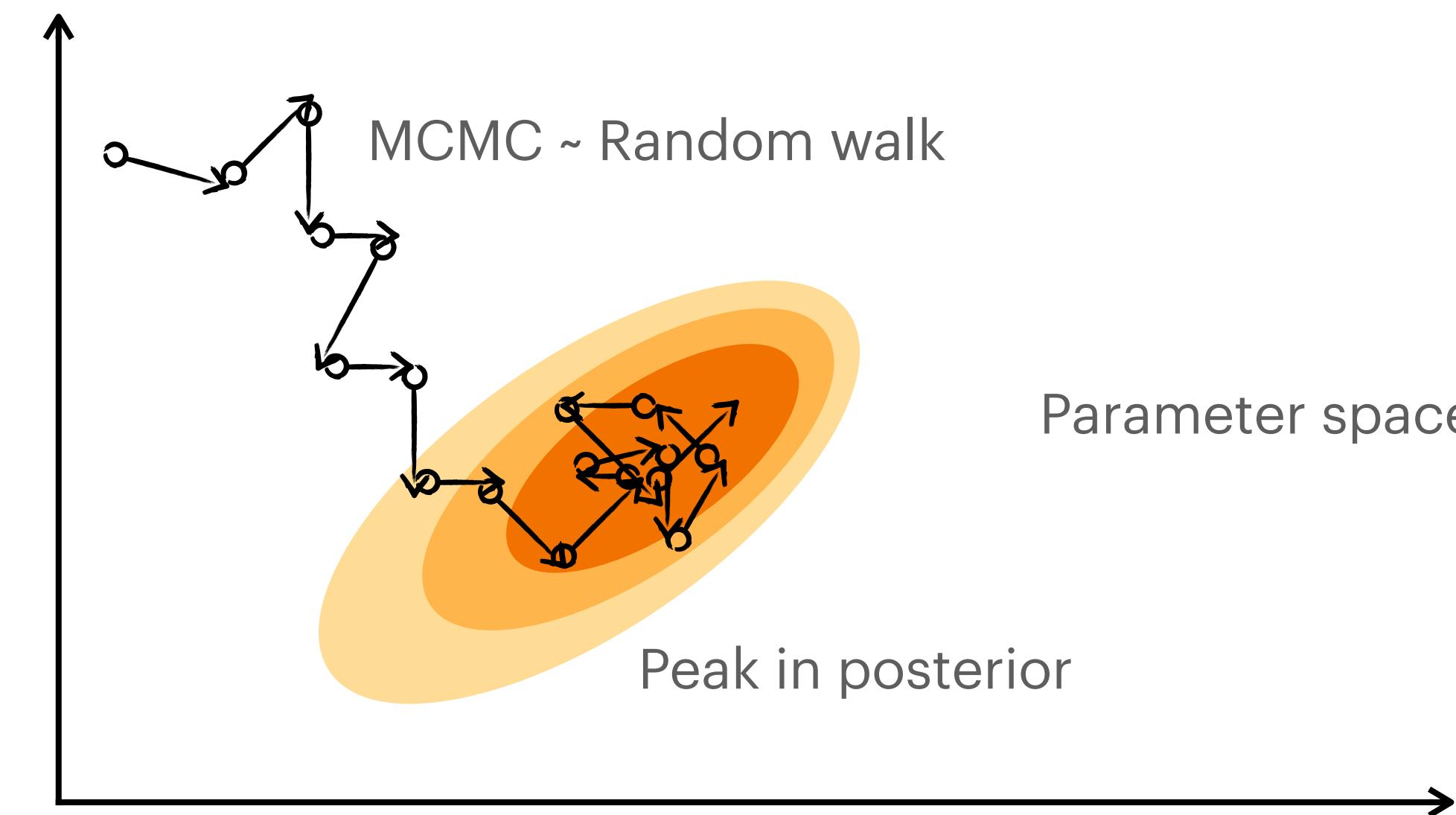
Bayesian
Evolutionary
Analysis by
Sampling
Trees
2

This morning

We all have one thing in common...

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

BUT HOW?



Bayesian
Evolutionary
Analysis by
Sampling
Trees
2

MCMC

We all have one thing in common...

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

BUT HOW?

CompEvol/**beast2**

Bayesian Evolutionary Analysis by Sampling Trees



28
Contributors

128
Issues

163
Stars

67
Forks

Bayesian
Evolutionary
Analysis by
Sampling
Trees

2

Modularity

BAYESIAN INFERENCE RECAP



Bayesian inference recap

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

PRIOR

- Original probability for the model parameters/components.
- All parameters have priors, whether you specify them or not!

Bayesian inference recap

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

LIKELIHOOD

- Probability of data given parameters (defined by the model)

Bayesian inference recap

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

POSTERIOR

- Updated probability for the model parameters in light of the data
- Result of our analysis

Bayesian inference recap

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

MODEL EVIDENCE

- Probability for data for any combination of parameters
- Used for Bayesian model selection.



**WHAT GOES
INTO A BEAST
ANALYSIS?**

What goes into a BEAST Analysis?

DATA

MODEL

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

DATA

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

Genetic Sequences, Trait Data, Roots of Words...

Gavryuskin et al.
Systematic Biology 2017

Bouckaert et al.
Science 2012

Often split into multiple "partitions" to allow specific sites to evolve at different rates, e.g.

- Coding/non-coding sites,
- Different codon positions (1+2 vs 3).

What goes into a BEAST Analysis?

DATA

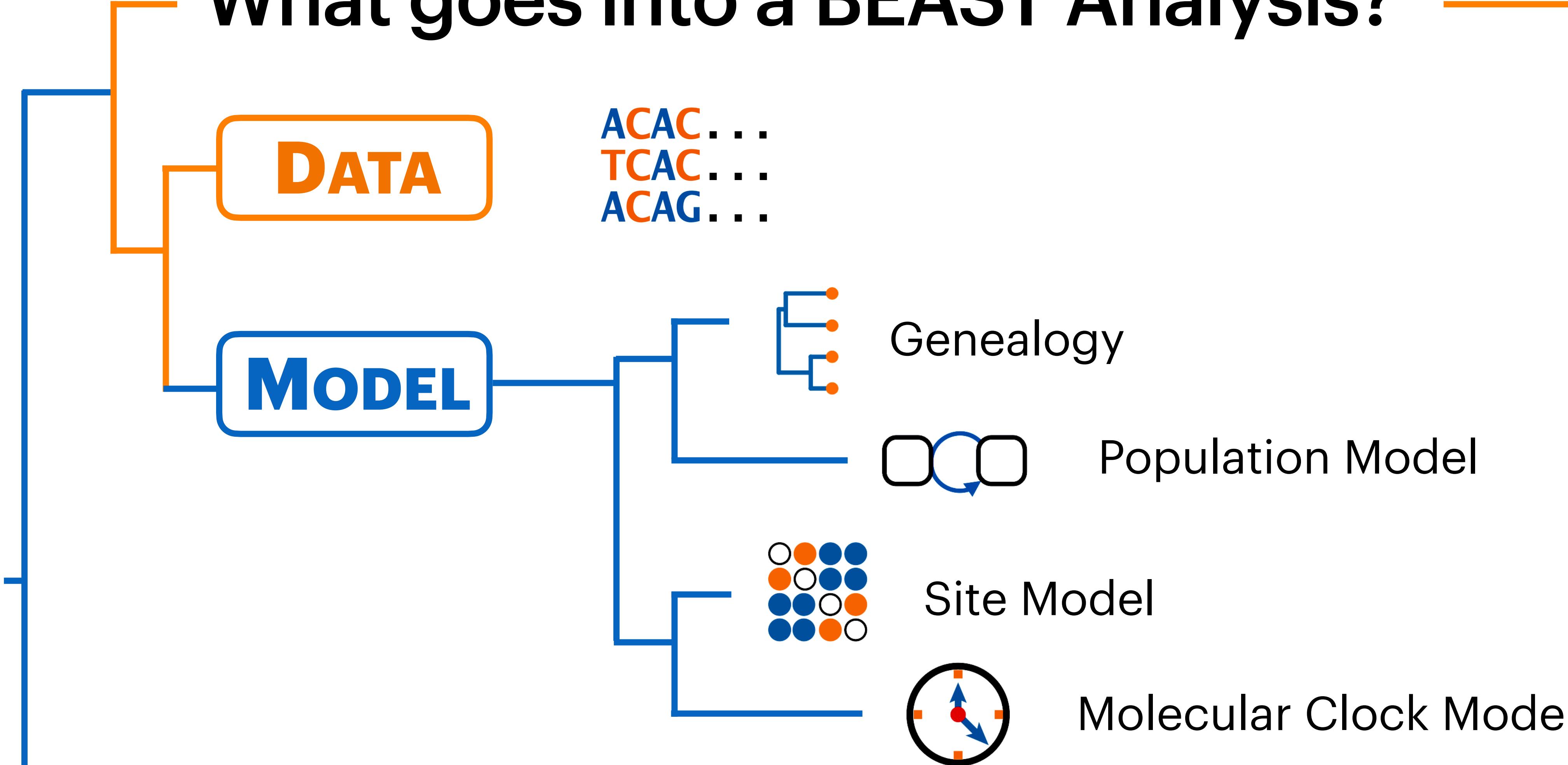
ACAC...
TCAC...
ACAG...

MODEL

$P(\text{model} | \text{ACAC...}, \text{TCAC...}, \text{ACAG...})$

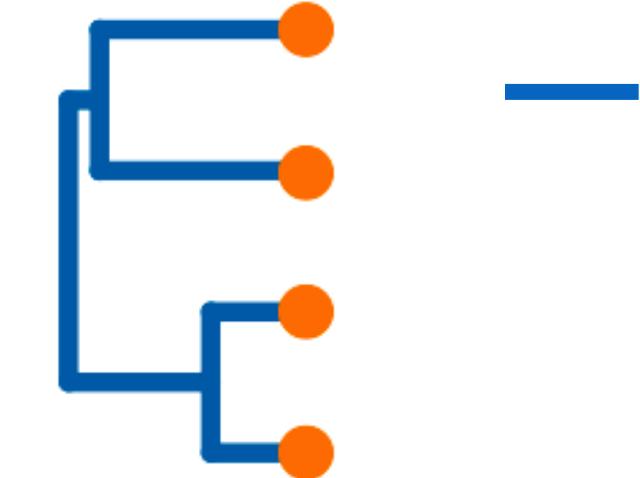
$$P(\text{model} | \text{ACAC...}, \text{TCAC...}, \text{ACAG...}) = \frac{P(\text{ACAC...}, \text{TCAC...}, \text{ACAG...} | \text{model}) P(\text{model})}{P(\text{ACAC...}, \text{TCAC...}, \text{ACAG...})}$$

What goes into a BEAST Analysis?

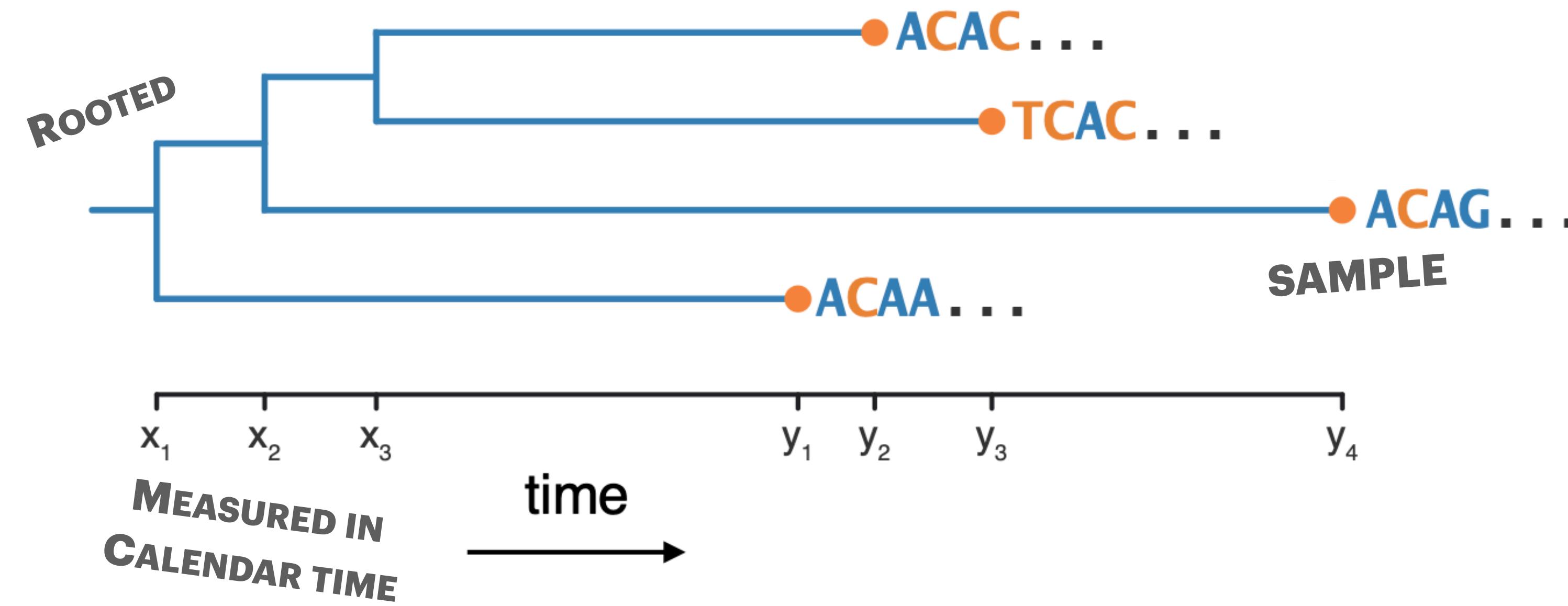


MODEL

Genealogy

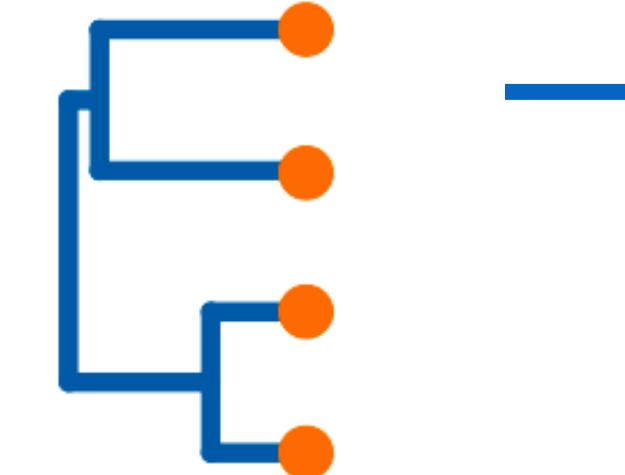


The fundamental genealogical structure
in **BEAST2** is the **rooted time-tree**



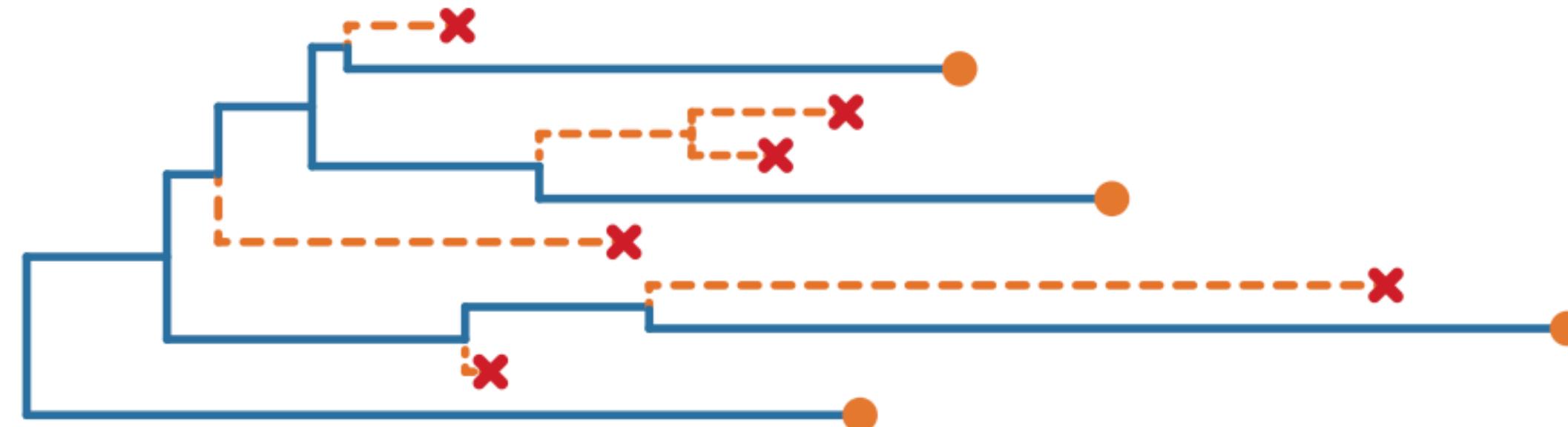
MODEL

Genealogy



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

FULL TREE



SAMPLED/RECONSTRUCTED TREE

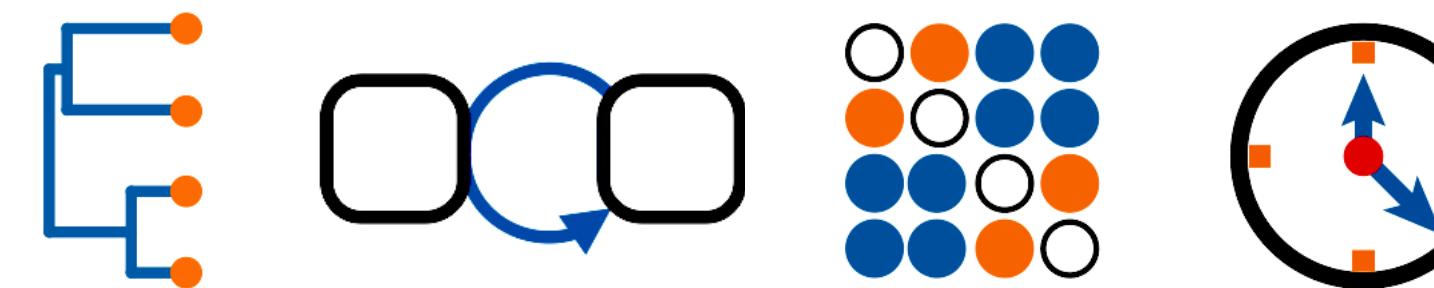


What goes into a BEAST Analysis?

DATA

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

MODEL



$$P(\text{tree, model} \mid \text{ACAC...}, \text{TCAC...}, \text{ACAG...}) = \frac{P(\text{ACAC...}, \text{TCAC...}, \text{ACAG...} \mid \text{tree, model}) P(\text{tree, model})}{P(\text{ACAC...}, \text{TCAC...}, \text{ACAG...})}$$

MODEL

Population Model



Describes the population/speciation/transmission dynamics

- How likely is the genealogy given a demographic model?

TREE
Realisation of a stochastic process

$$P(E | \text{POPULATION MODEL})$$

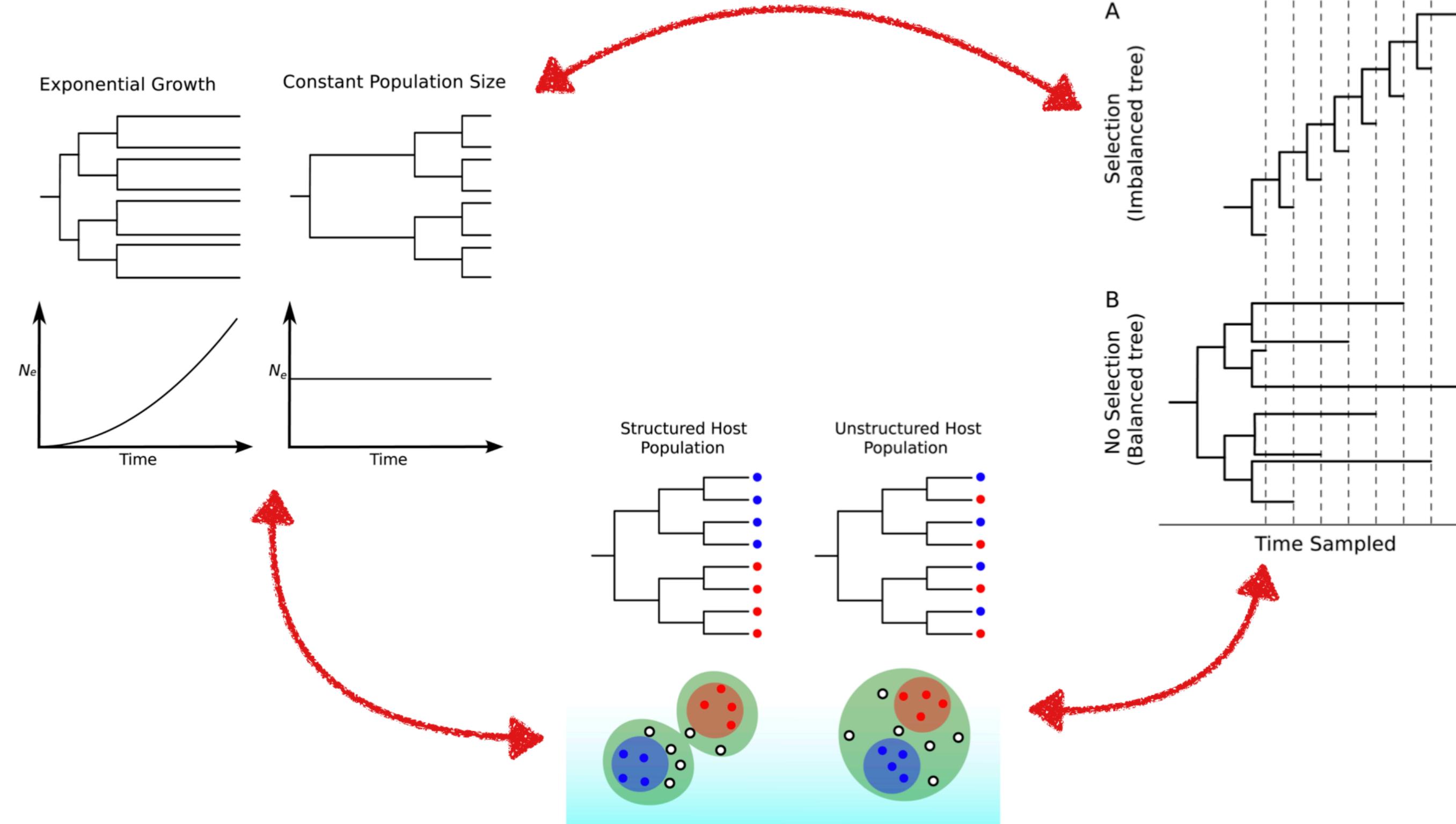
POPULATION MODEL
Describes the population dynamics (growth of the tree)

- Sometimes called a **tree prior**
- Usually a **coalescent** or **birth-death** model

MODEL

Population Model

Different population dynamics generate trees that look different



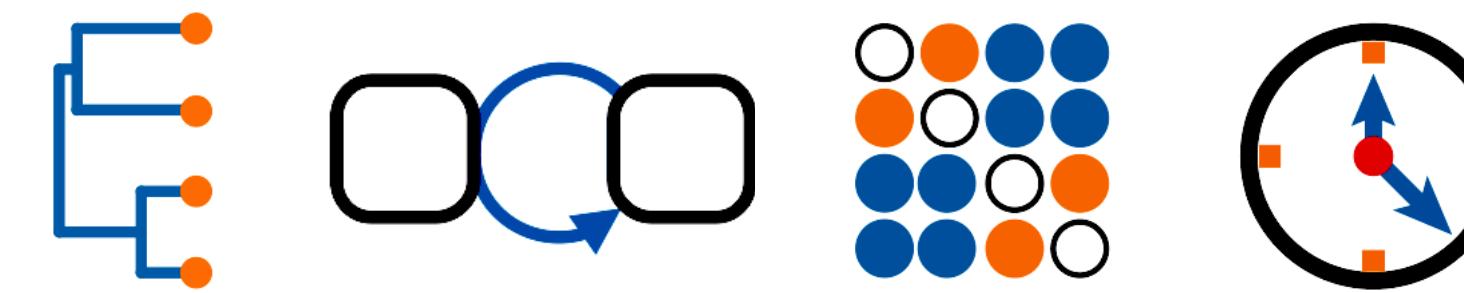
Volz et al. PLoS Comp Biol 2013
Grenfell et al. Science 2004

What goes into a BEAST Analysis?

DATA

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

MODEL

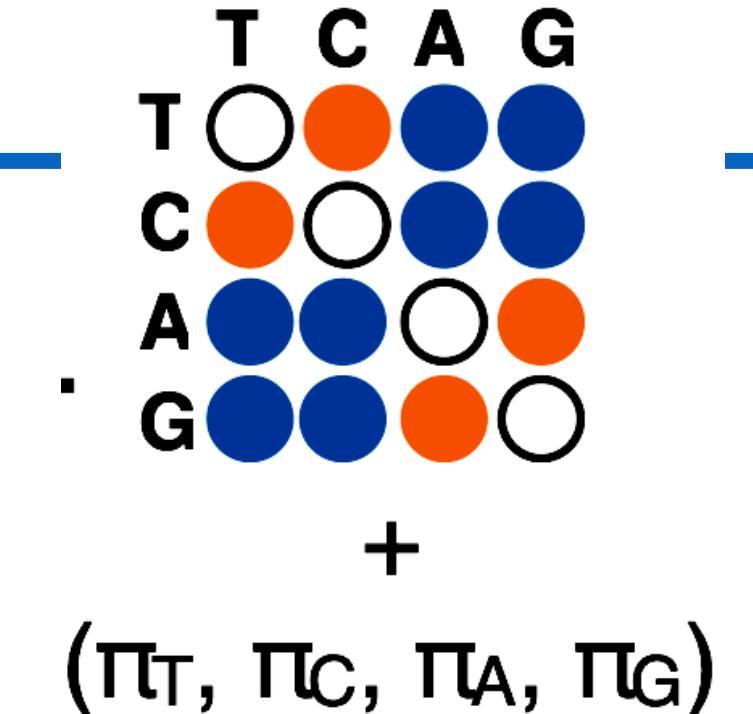


$$P(\text{E} \circ \square, m | \text{ACAC...}, \text{TCAC...}, \text{ACAG...}) = \frac{P(\text{ACAC...} | \text{E} \circ \square, m) P(\text{E} \circ \square, m)}{P(\text{ACAC...})}$$

MODEL

Site Model

Links data sequences to the genealogy



- Describes **rates of substitution** between available characters
- The **tree likelihood** sums over all possible substitution histories leading to the observed sequences:

$$P(\text{ACAC...} | \text{TCAC...} \text{ ACAG...}) \quad \text{TREE LIKELIHOOD}$$

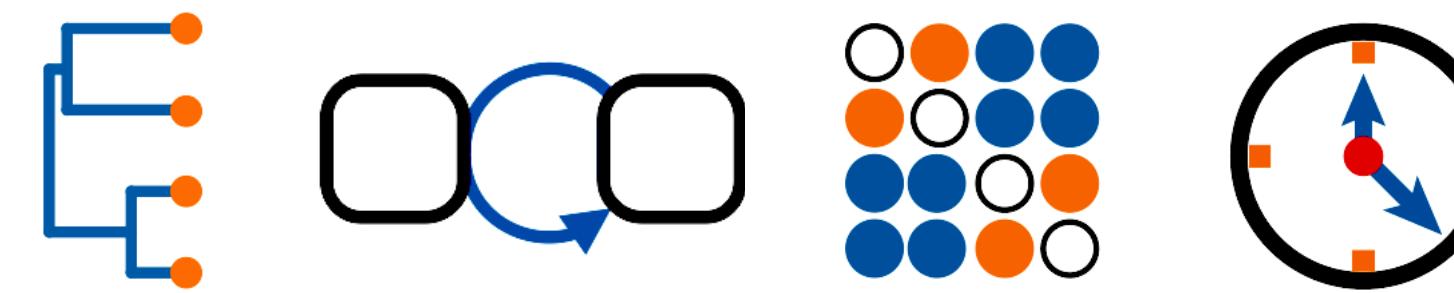
- Popular models are GTR, HKY, K80, JK69...

What goes into a BEAST Analysis?

DATA

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

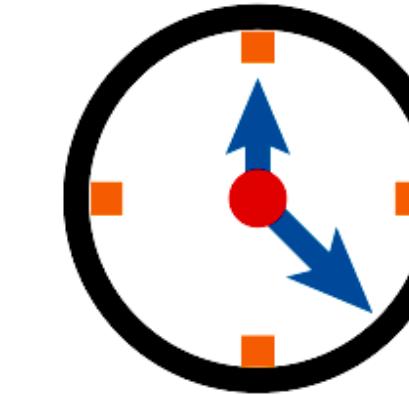
MODEL



$$P(\text{Tree} \text{, Clock } | \text{Data, Model}) = \frac{P(\text{ACAC...} | \text{Tree, Clock, Model})}{P(\text{ACAC...})}$$

MODEL

Molecular Clock Model



Determines how quickly sequences evolve along the tree

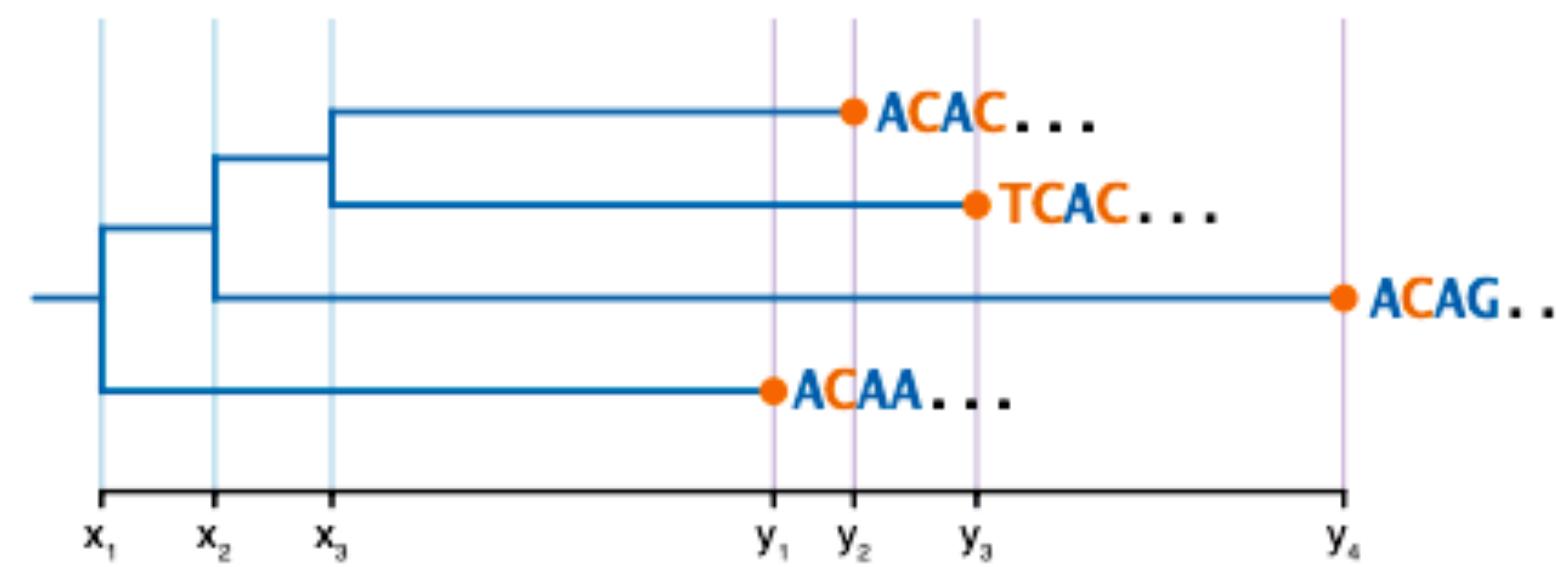
genetic distance tree
(subst/site)



clock rate
(subst/site/year)

$$= \mu \times$$

time tree
(years)



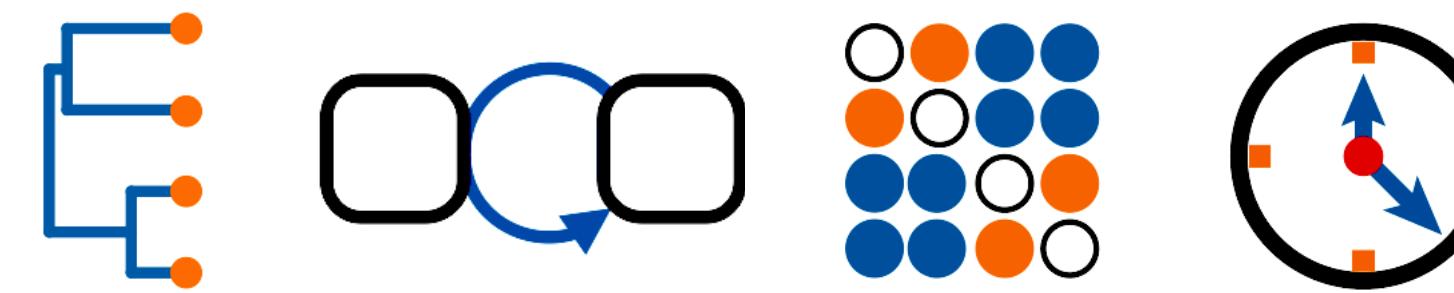
- **Relaxed vs strict clocks:** Some branches may have different clock rates

What goes into a BEAST Analysis?

DATA

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

MODEL



$$P(\text{Phylogenetic Tree} \mid \text{Clock}, \text{Grid}, \text{Data}) = \frac{P(\text{Data} \mid \text{Phylogenetic Tree}, \text{Clock}, \text{Grid}) P(\text{Clock} \mid \text{Phylogenetic Tree}) P(\text{Grid} \mid \text{Phylogenetic Tree})}{P(\text{Data} \mid \text{Grid})}$$

What goes into a BEAST Analysis?

Putting it all together...

$$P(\text{Tree} | \text{Data}) = \frac{P(\text{Data} | \text{Tree}) P(\text{Tree})}{P(\text{Data})}$$

The equation shows the joint probability of the tree and data. The numerator is the product of the tree likelihood (P(Data | Tree)) and the tree prior (P(Tree)). The denominator is the marginal probability of the data (P(Data)).

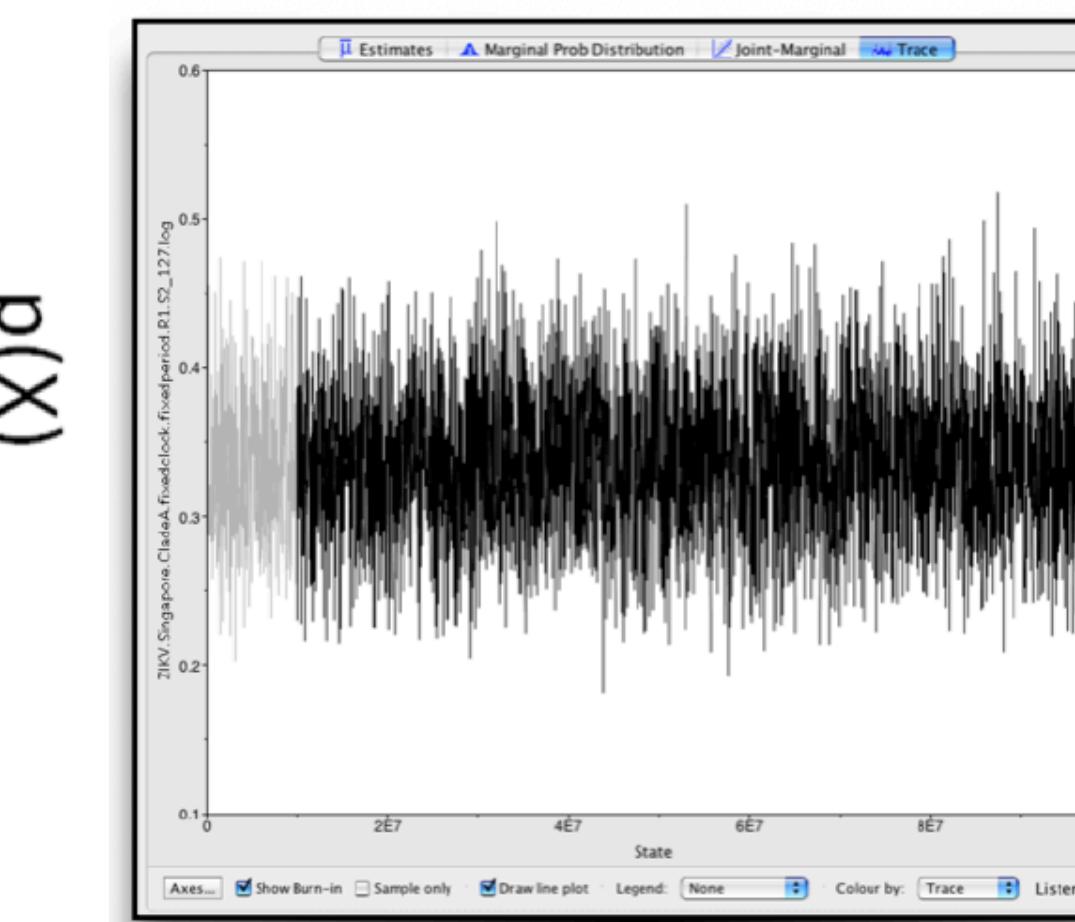
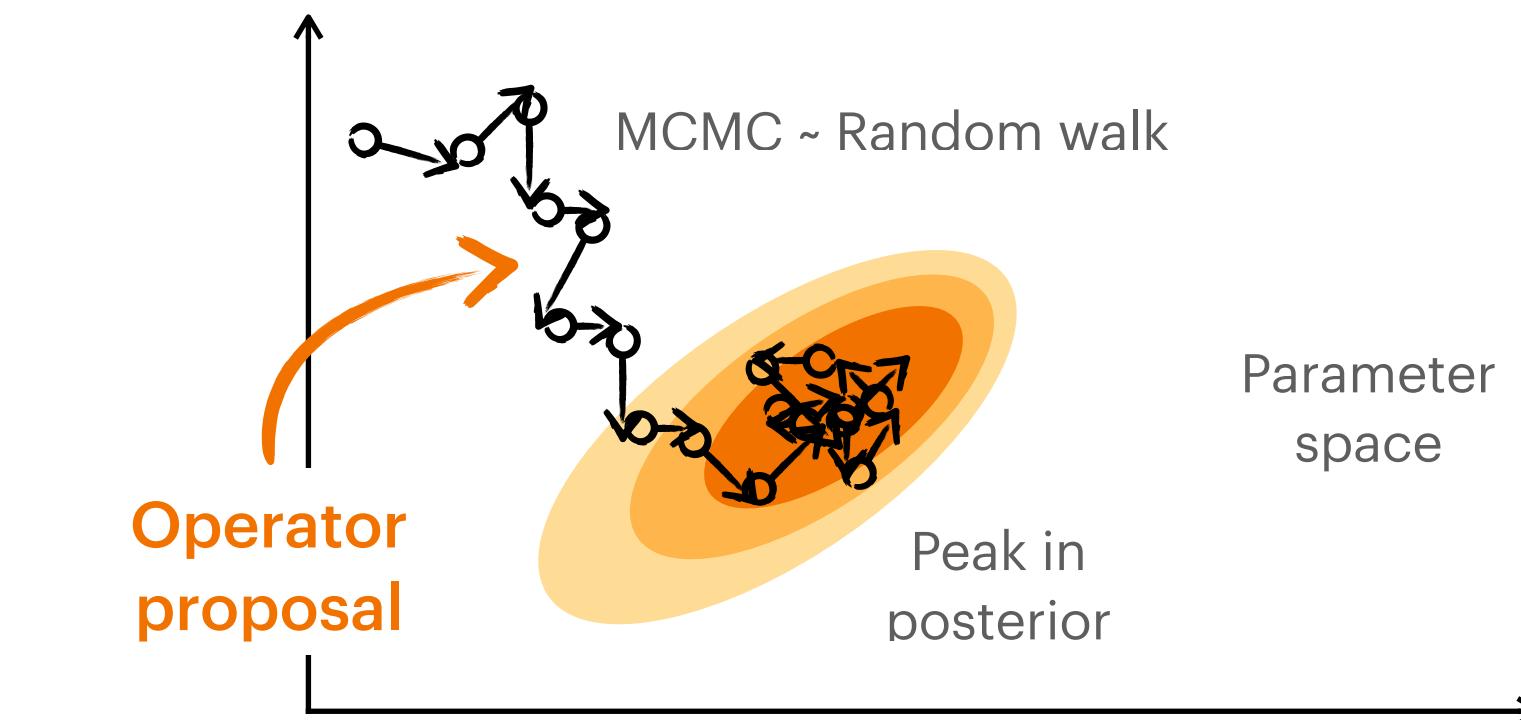
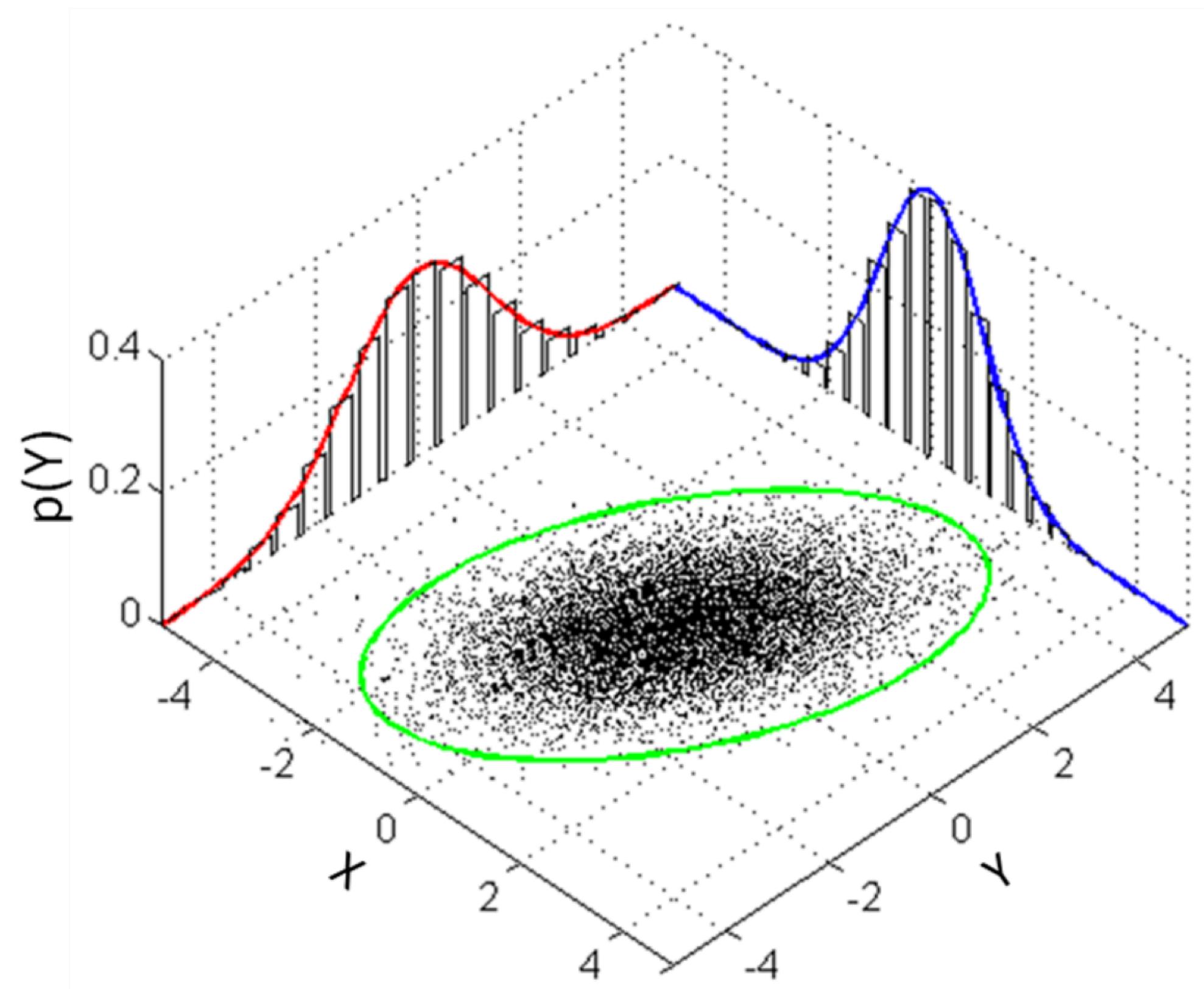
Assume independence:

$$P(\text{Data} | \text{Tree}) = P(\text{Tree Likelihood}) P(\text{Tree Prior}) P(\text{Parameter Priors})$$

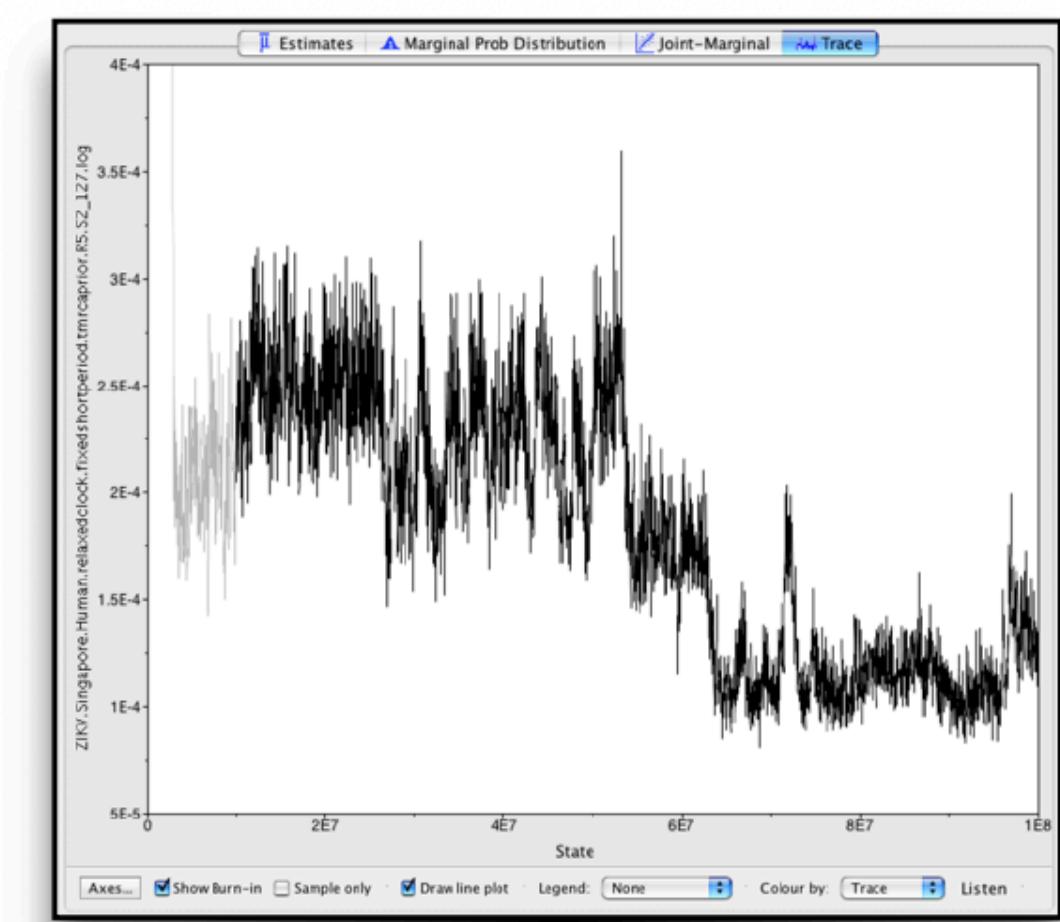
The equation shows the joint probability of the data given the tree, factored into three independent components: Tree Likelihood, Tree Prior, and Parameter Priors.

MCMC (Markov Chain Monte Carlo)

We sample from the joint posterior



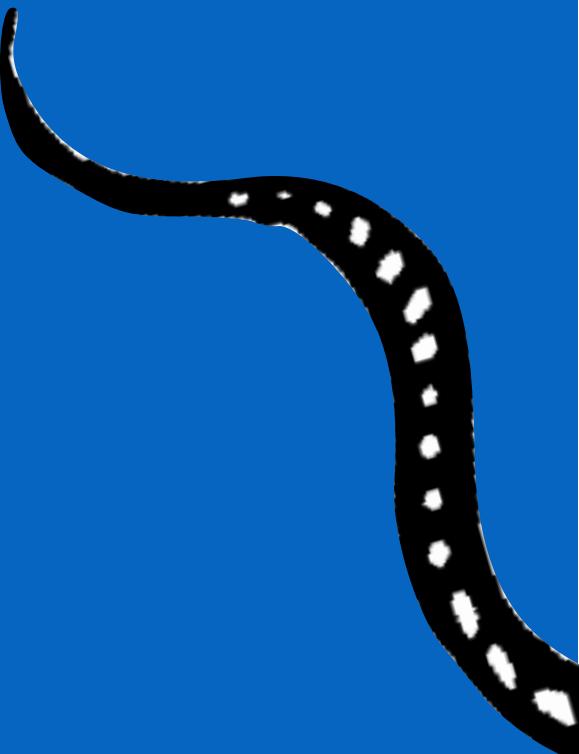
Mixing well! 😊



Not mixing! 😢

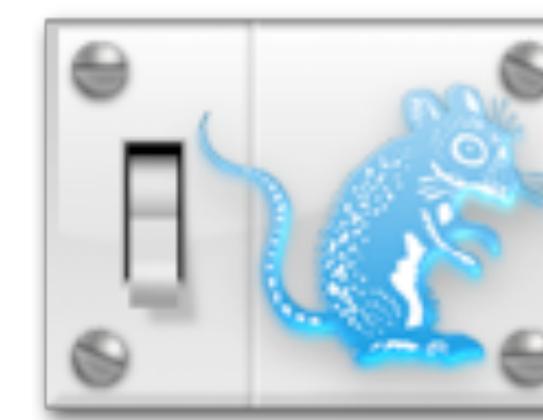
BEAST 2

WORKFLOW



BEAST2 Workflow

ACACACCTACAGAC
TCACACACCTACACAC
TCAGACTTTCACAC
ACAGACTTTCAGAC
TCACACACCTACACAC



BEAUti2

XML



BEAST2

Tracer



Trace file
.log

Trees file
.trees

TreeAnnotator



ggtree (R)

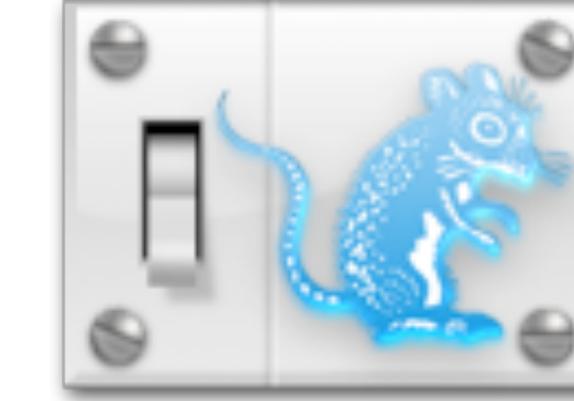


Icytree



FigTree

BEAUti



Graphical tool for setting up a BEAST2 analysis

INPUT

- 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
- **.xml** file



Bayesian Evolutionary Analysis by Sampling Trees

- Performs MCMC analyses of sequences under selected sequence evolution and tree (epidemiological/speciation) model.
- Similar to BEAST 1.8.4/1.10 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. **BEAST2 Packages**.

INPUT

- XML model description file

OUTPUT

- (Trace) log file
- Tree (log) file

Tracer



- Analyse (parameter) log files from BEAST2 runs
- Assess mixing, ESS, ACT, parameter correlations
- Provides overview of posterior parameter estimates
- Comparisons of several analyses.
- Tracer is primarily a diagnostic tool — usually want to perform final analyses in a statistical package like R.

INPUT

- One or more trace log files

OUTPUT

- Insight

TreeAnnotator



- Analyse trees file from BEAST2 runs.
- Produces single summary tree 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 edge lengths when topological uncertainty is large!

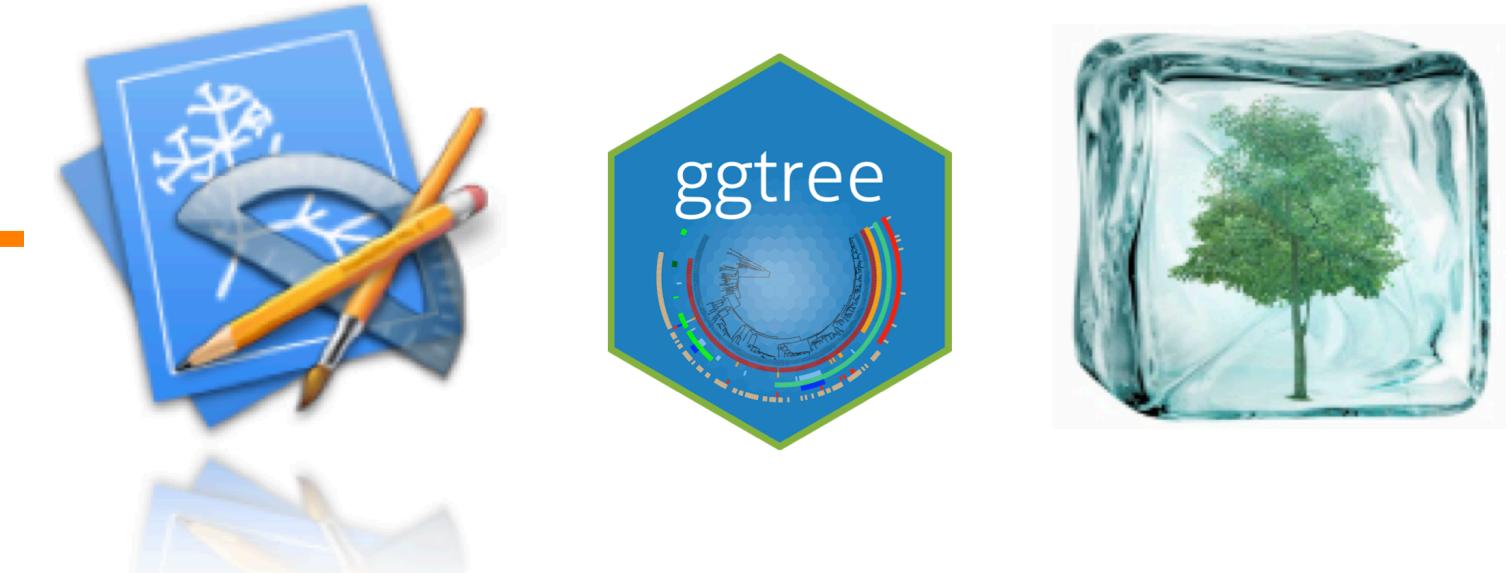
INPUT

- Trees log file

OUTPUT

- Summary tree

Tree Visualization



- Visualise trees from BEAST2 runs.
- Annotate branches and nodes with probabilities and labels.
- **FigTree:** Many different tree visualisation styles: circular, unrooted, etc.
- **Icytree:** Quick visualisation, web app, no installation required.
- **ggtree:** R package to visualize trees
- Allows highlighting of particular clades, colouring of edges and more

INPUT

- Tree file (e.g. TreeAnnotator output)

OUTPUT

- Tree visualisation

TUTORIAL YOUR TURN!



Taming the BEAST



news workshops

tutorials

contribute



Introduction to BEAST2

This is a simple introductory tutorial to help you get started with using BEAST2 and its accomplices.

Jūlia Pečerska , Veronika Bošková and Louis du Plessis
Updated 8 Aug 2019 by laduplessis

Skyline plots

Inference of past population dynamics using Bayesian Coalescent Skyline and Birth-Death Skyline plots.

Nicola F. Müller and Louis du Plessis
Updated 27 Aug 2019 by laduplessis

Prior selection

Prior selection and clock calibration using Influenza A data.

Veronika Bošková , Venelin Mitov and Louis du Plessis
Updated 8 Aug 2019 by laduplessis

Substitution model averaging

A tutorial using bModelTest

David A. Rasmussen , Carsten Magnus and Remco Bouckaert
Updated 13 Aug 2019 by rbouckaert

Troubleshooting

Post-processing and improving performance

David A. Rasmussen
Updated 18 Jun 2018 by Unknown

Divergence Time Estimation using BEAST v2.x

Dating Species Divergences with the Fossilized Birth-Death Process

Tracy A. Heath
Updated 16 Mar 2020 by tgvaughan

Let's start!

11:30 → 12:35 tutorial (breakout rooms)

12:35 → 12:45 wrap up (main room)

Follow the instructions at:

<https://taming-the-beast.org/tutorials/Introduction-to-BEAST2>

Do not hesitate to ask for help in the Slack channel!

room 8 

t-first-steps

Help in Breakout room 3!

Hey! I need help in breakout room 5