



A (very) brief introduction to RevBayes

October 21, 2017

Michael Landis & Rachel Warnock

michael.landis@yale.com

Modeling in theory

What parameters & processes explain the data?

Identify a problem

Define probabilities of outcomes

Implement model as software method

Apply model to data with method

Modeling in practice

Program new method

or

Modify existing method

- No method implements all models

- Existing software often difficult to extend

- Model specification complicated

Phylogenies are unusual



A simple model

Phylogenies are unusual



A simple model



A tree-based model

RevBayes

Flexible model specification

Graphical models

Easy and intuitive to use

Rev language interface

Fast computation and efficient inference

C++ backend

Bayesian terminology

D	data, observations
θ	model parameters
$P(D \mid \theta)$	model likelihood
$P(\theta)$	prior probability
$P(\theta \mid D)$	posterior probability

Bayes rule

$$P(\theta \mid D) = \frac{P(D \mid \theta)P(\theta)}{P(D)}$$

Model equations

How are the model variables interrelated?

$$D \sim \text{Exponential}(Y)$$

$$Y = (\theta_1 - \theta_2)^2$$

$$\theta_1 \sim \text{Exponential}(1)$$

$$\theta_2 \sim \text{Exponential}(1)$$

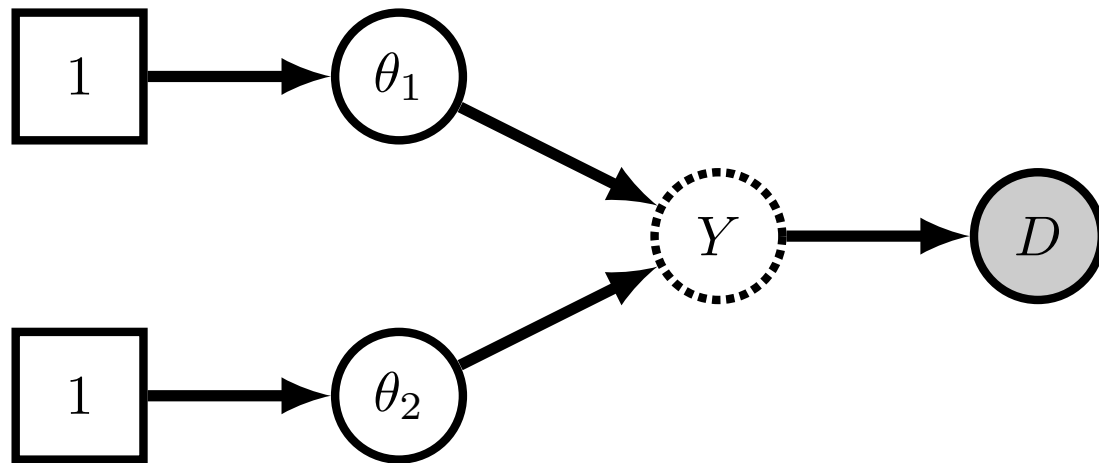
Model graph

$$D \sim \text{Exponential}(Y)$$

$$Y = (\theta_1 - \theta_2)^2$$

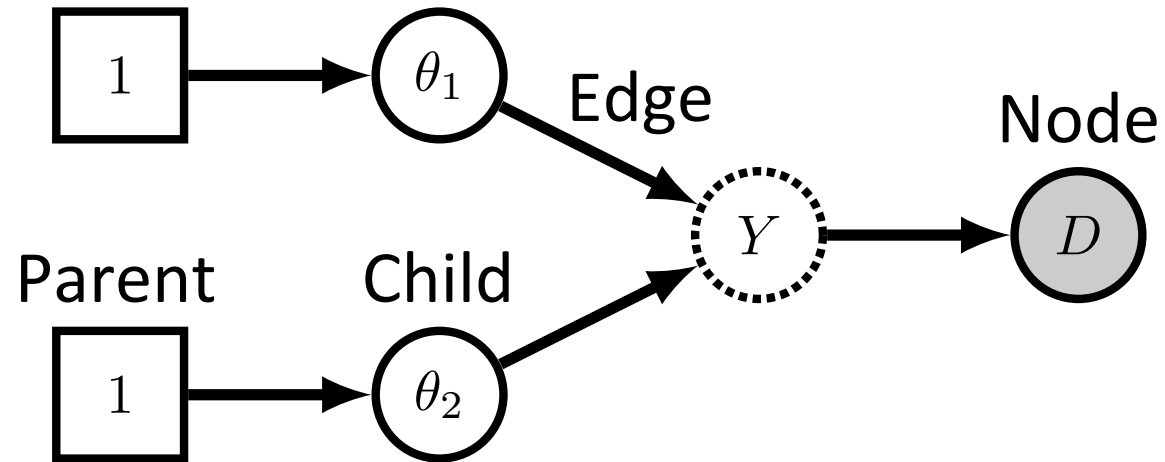
$$\theta_1 \sim \text{Exponential}(1)$$

$$\theta_2 \sim \text{Exponential}(1)$$



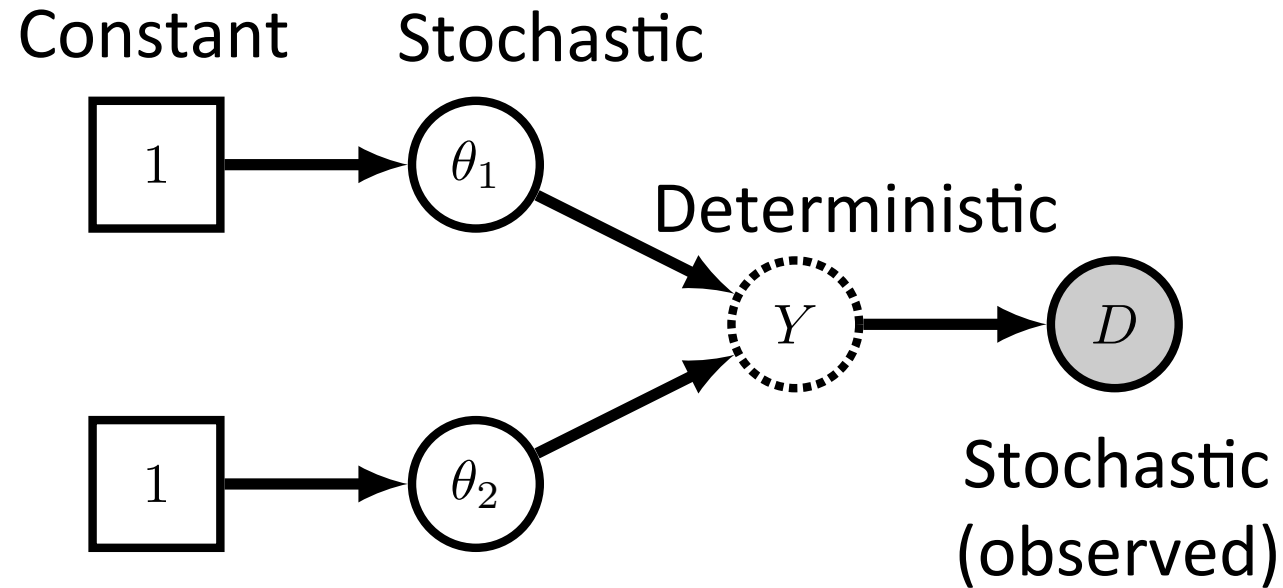
Model variable relationships

Edges indicate how child nodes
depend on parent nodes

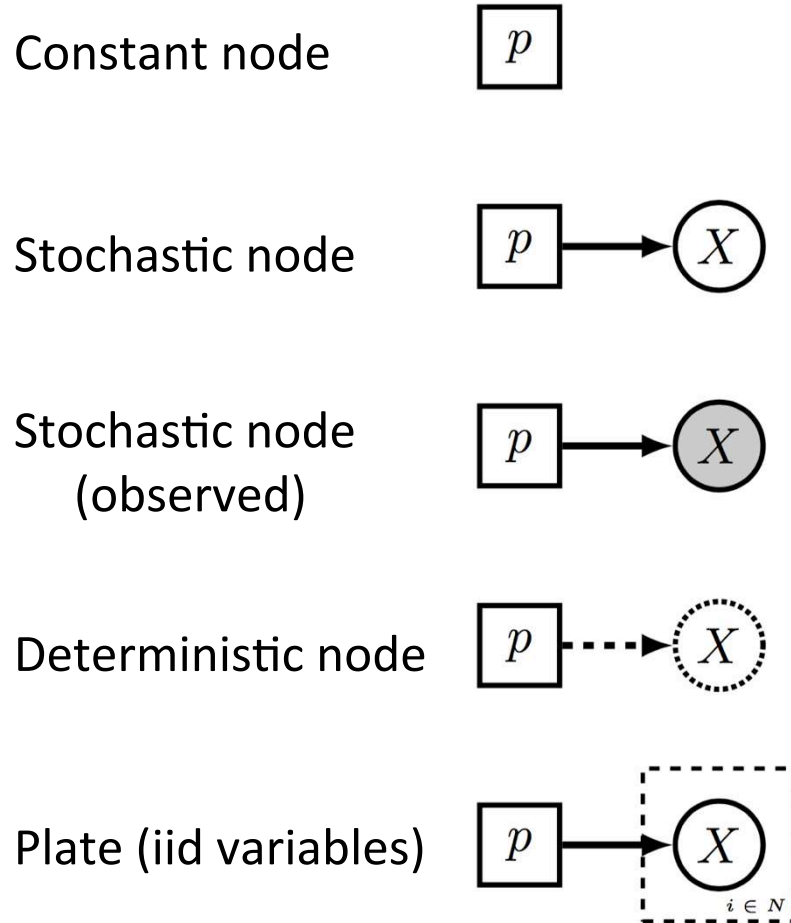


Model variable types

Nodes have different properties



The Rev Language



```
p <- 0.5
```

```
x ~ dnBernoulli(p)
```

```
x.clamp(1)
```

```
x := exp(p)
```

```
for (i in 1:N) {  
  x[i] ~ dnBernoulli(p)  
}
```

(short demo)

Bayesian phylogenetics

D

sequence alignment

θ

tree, rates

$P(D \mid \theta)$

prob. sequences evolved under
particular tree w/ rate

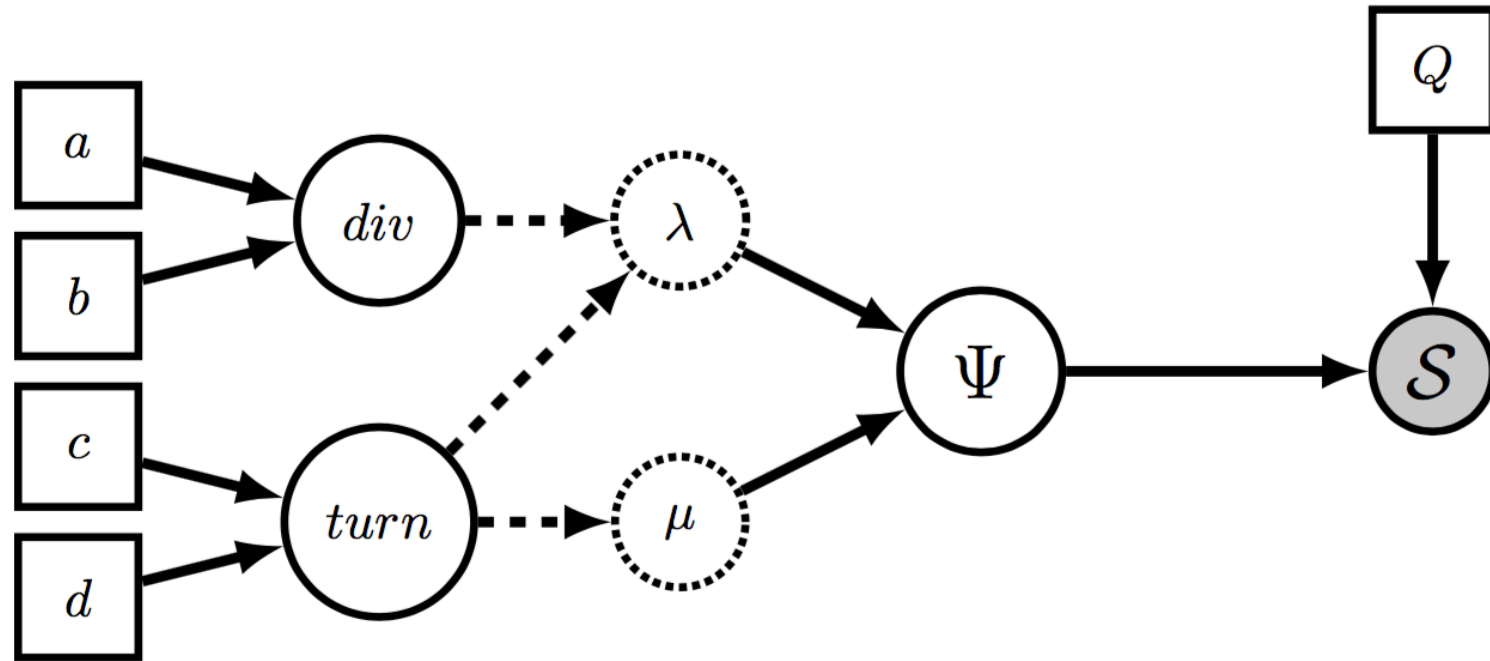
$P(\theta)$

birth-death process prior,
substitution rate prior, etc.

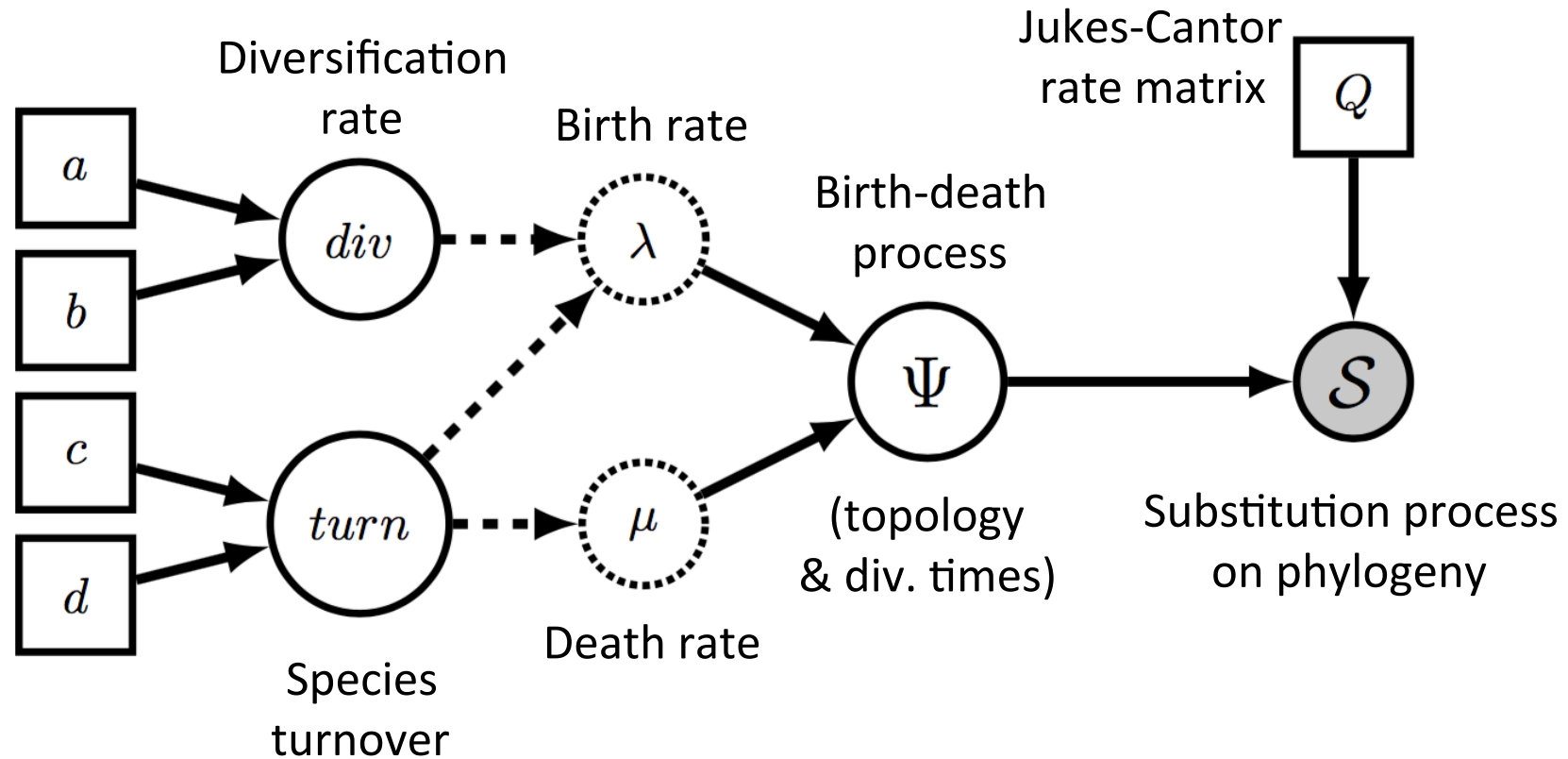
$P(\theta \mid D)$

prob. each tree w/ rate describe
particular evolved sequences

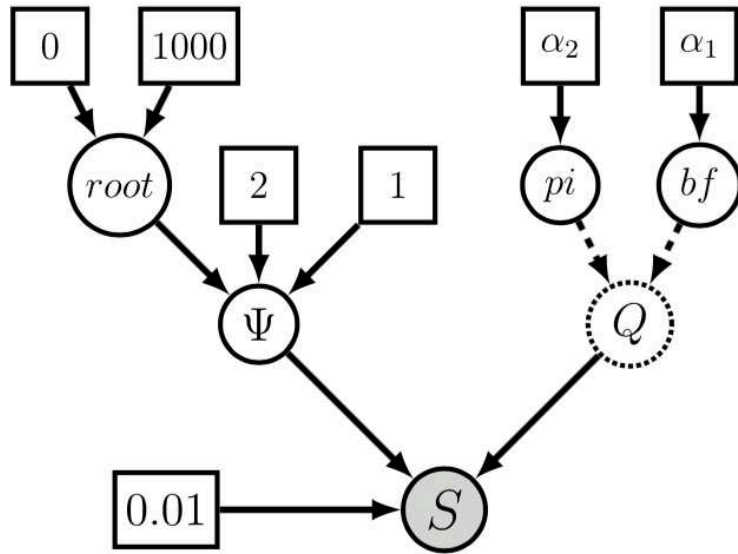
A simple phylogenetic model



A simple phylogenetic model



GTR, birth-death, strict clock



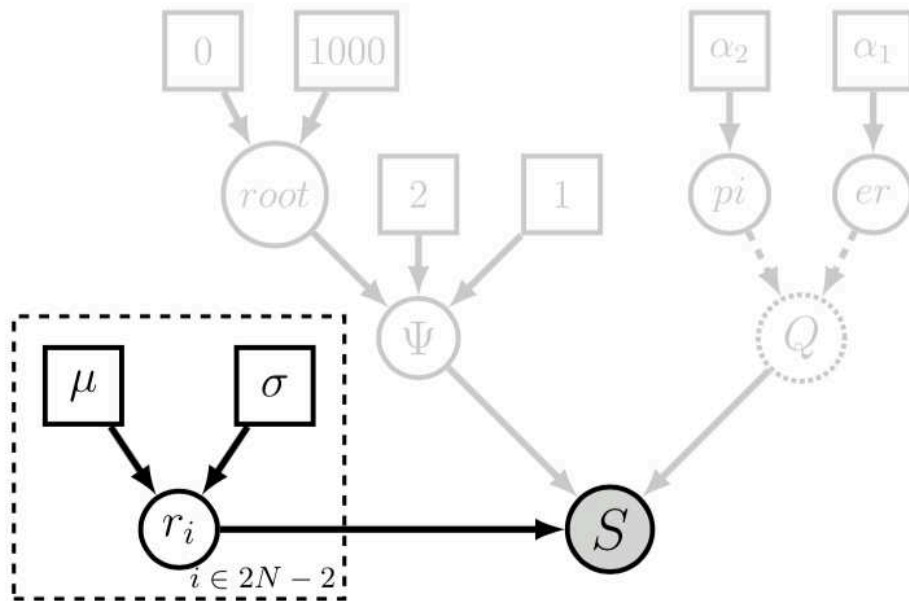
```
root ~ dnUniform(0,1000.0)
psi ~ dnBirthDeath(lambda=2.0, mu=1.0,
                    rootAge=root, taxon_names)
```

```
alpha1 <- v(1,1,1,1,1,1)
alpha2 <- v(1,1,1,1)
```

```
er ~ dnDirichlet( alpha1 )
pi ~ dnDirichlet( alpha2 )
Q_mol := fnGTR(er, pi)
```

```
seq ~ dnPhyloCTMC( tree=psi, Q=Q_mol, branchRates=0.01, type="DNA" )
seq.clamp( data )
```

GTR, birth-death, UCLN clock



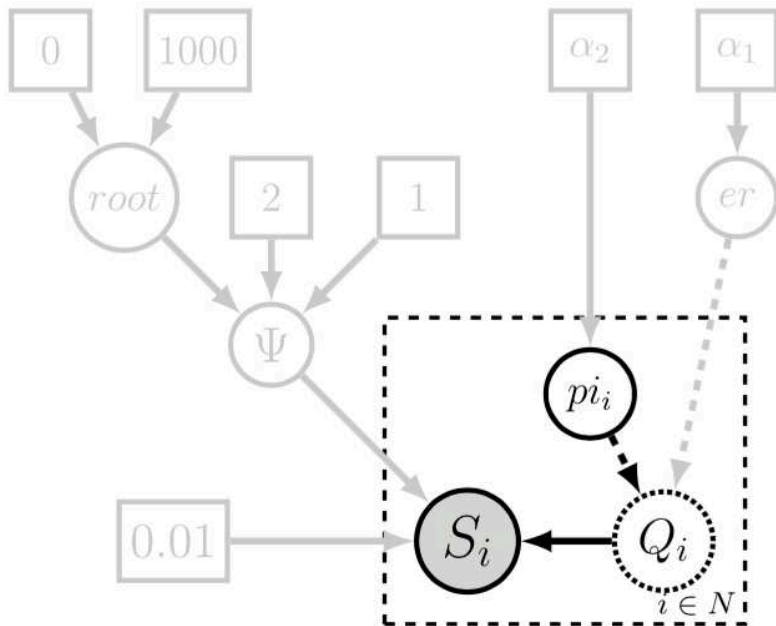
```
root ~ dnUniform(0,1000.0)
psi ~ dnBirthDeath(lambda=2.0, mu=1.0,
                    rootAge=root, taxon_names)
```

```
alpha1 <- v(1,1,1,1,1,1)
alpha2 <- v(1,1,1,1)
er ~ dnDirichlet( alpha1 )
pi ~ dnDirichlet( alpha2 )
Q_mol := fnGTR(er, pi)
```

```
mu <- 0
sigma <- 1
for (i in 1:(2*N-2)) {
  r[i] ~ dnLognormal( mu, sigma )
}
```

```
seq ~ dnPhyloCTMC( tree=psi, Q=Q_mol, branchRates=r, type="DNA" )
seq.clamp( data )
```

Partitions, GTR, birth-death, strict clock



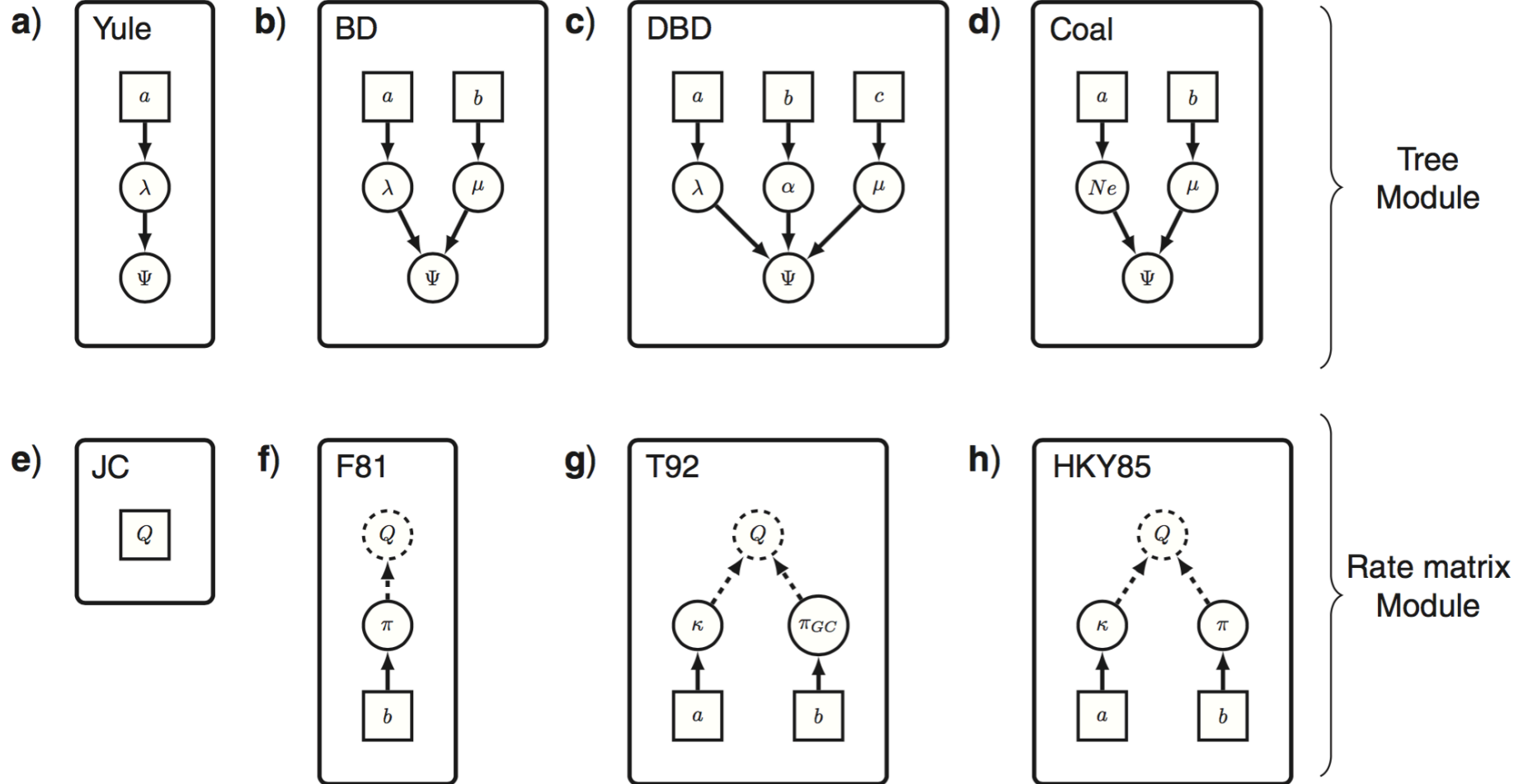
```
root ~ dnUniform(0,1000.0)
psi ~ dnBirthDeath(lambda=2.0, mu=1.0,
                    rootAge=root, taxon_names)

alpha1 <- v(1,1,1,1,1,1)
alpha2 <- v(1,1,1,1)

er ~ dnDirichlet( alpha1 )
for (i in 1:N) {
  pi[i] ~ dnDirichlet( alpha2 )
  Q_mol[i] := fnGTR(er, pi[i])

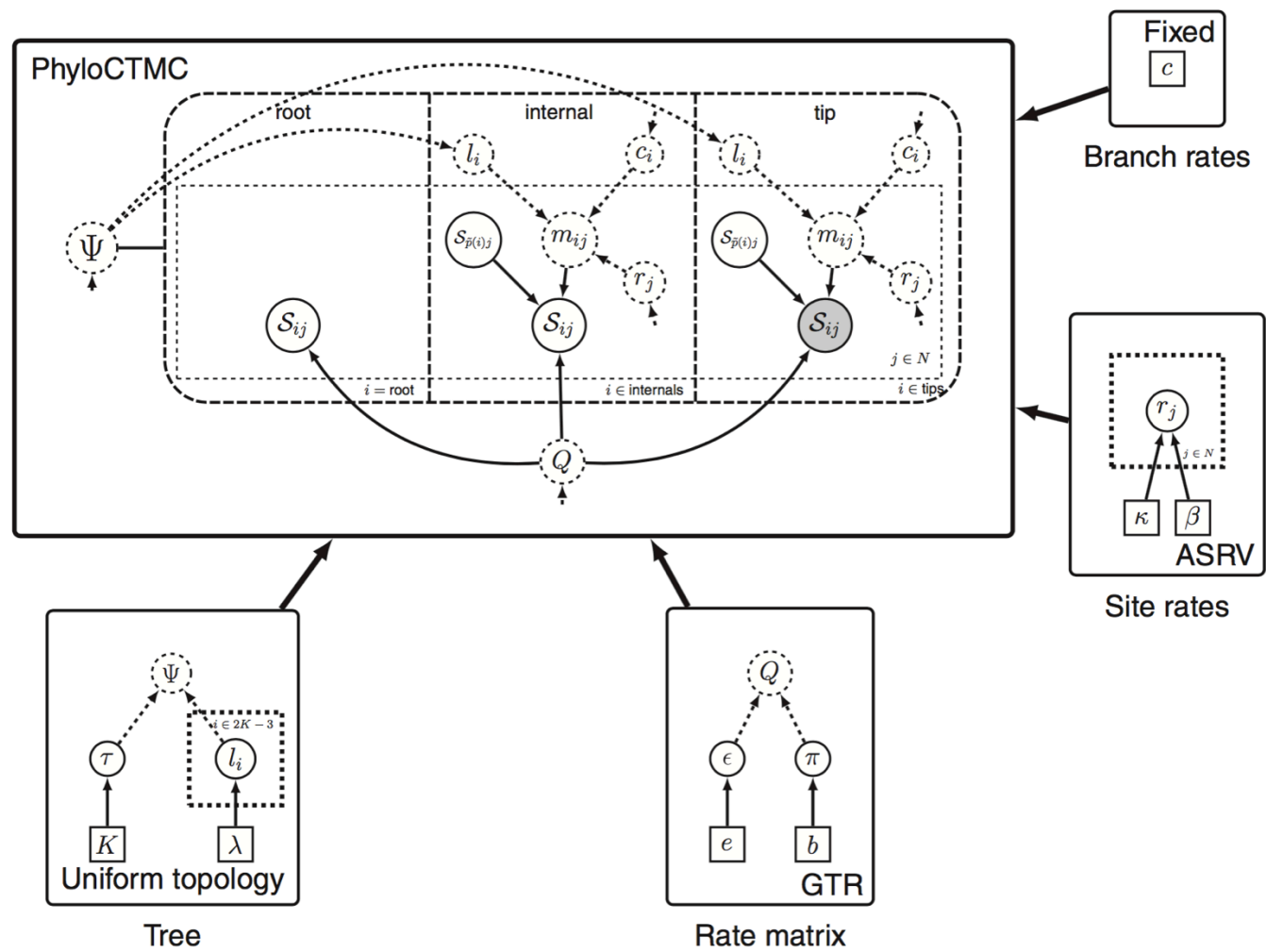
  seq[i] ~ dnPhyloCTMC( tree=psi, Q=Q_mol[i], branchRates=0.01, type="DNA" )
  seq[i].clamp( data[i] )
}
```

Modularity

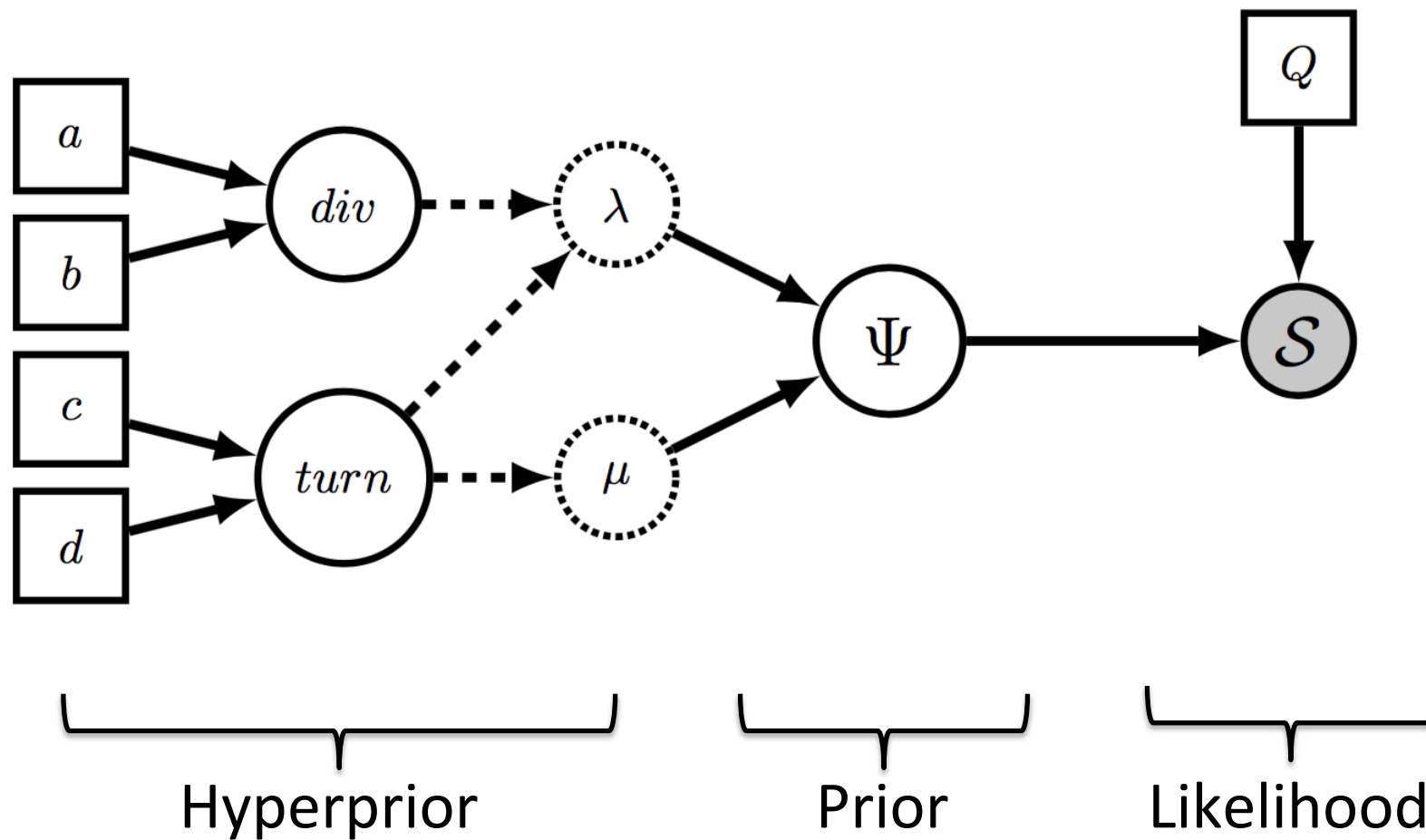


```
graph TD; Tree[Tree] --> PhyloCTMC[PhyloCTMC]; RateMatrix[Rate matrix] --> PhyloCTMC; SiteRates[Site rates] --> PhyloCTMC; BranchRates[Branch rates] --> PhyloCTMC;
```

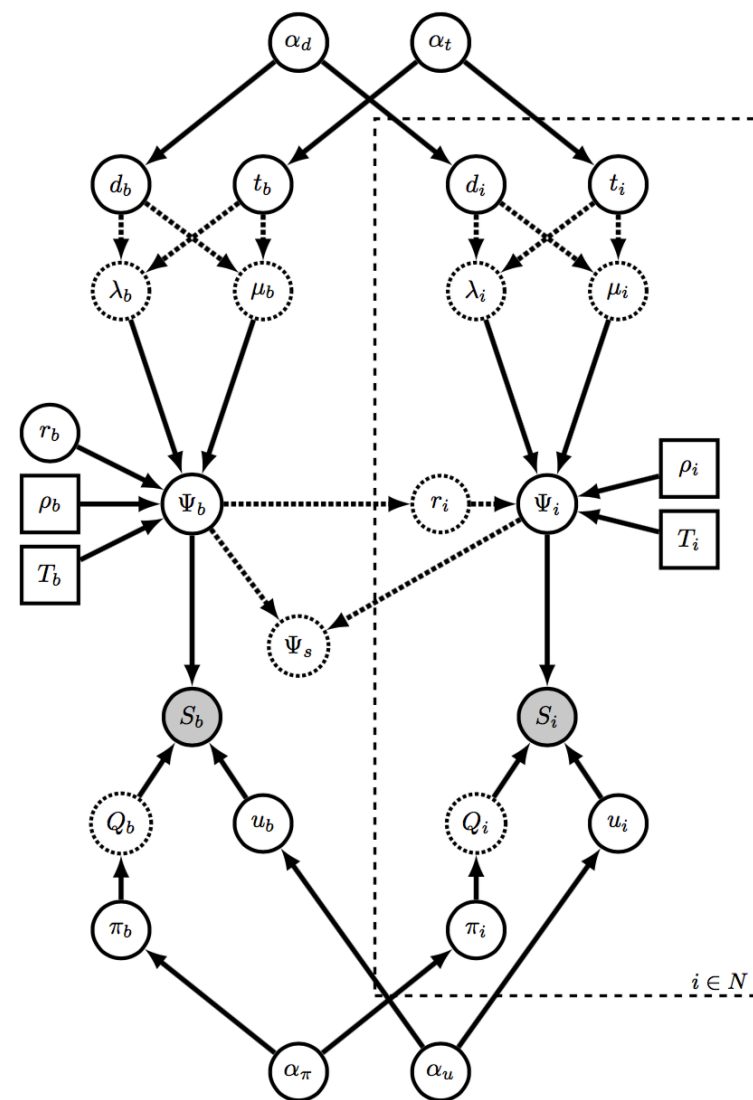
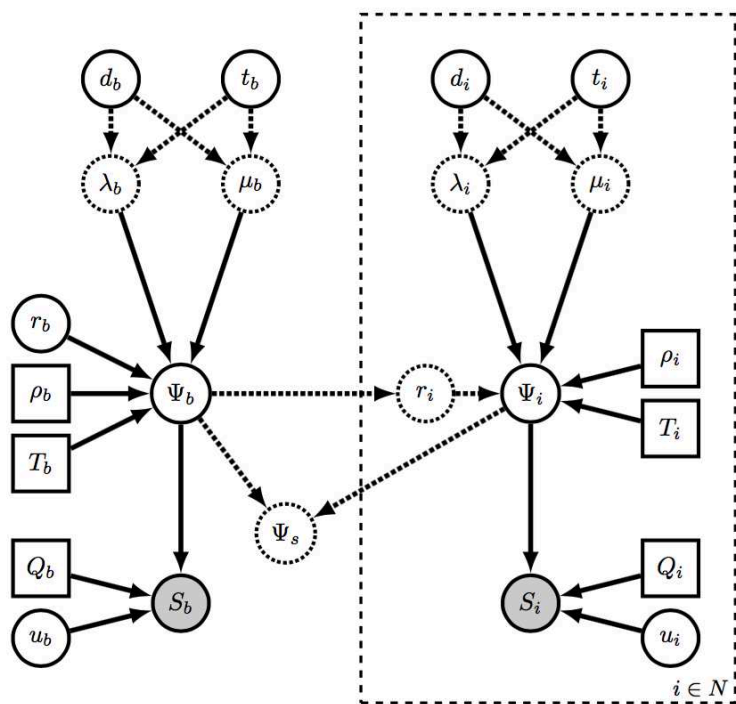
A diagram illustrating the inputs to the PhyloCTMC model. Four boxes at the top, labeled "Tree", "Rate matrix", "Site rates", and "Branch rates", each have an arrow pointing down to a central box labeled "PhyloCTMC".



Hyperpriors



Scalable complexity



Performance

Estimate substitution process, tree fixed

Software	HKY	HYK+ Γ	GTR	GTR+ Γ
BEAST v1.8	65.3	188.4	75.8	213.4
BEAST v1.8 - BEAGLE	41.2	105.2	47.5	107.4
MrBayes 3.2	78.2	177.7	76.9	169.9
MrBayes 3.2 - BEAGLE	92.5	221.2	91.4	222.7
→ RevBayes	46.9	161.3	62.5	181.2

Estimate tree, substitution process fixed

Software	NNI	Node-Slide
BEAST v1.8	30.7	42.8
BEAST v1.8 - BEAGLE	21.0	28.3
MrBayes 3.2	37.2	38.1
MrBayes 3.2 - BEAGLE	42.6	31.9
→ RevBayes	17.8	23.5

Features

Complex models

Bayesian & maximum likelihood

Simulation

Model selection & adequacy

Ancestral state estimation

Custom variable monitoring

Parallel processing

Interactive & scripted languages

revbayes.com/tutorials

Introductions

- Getting Started with RevBayes

- Markov chain Monte Carlo

- Molecular phylogenetics

Substitution models

Data partitioning

Divergence time estimation

- node calibrations

- tip dating

- fossilized birth-death processes

Diversification rate estimation

- episodic models

- branch-heterogeneous rate variation

- state-dependent rate variation

Species tree estimation

- concatenated analyses

- multispecies coalescent

Biogeography

Chromosome evolution

Continuous trait evolution

Discrete morphological evolution

Methods

- MCMC Strategies

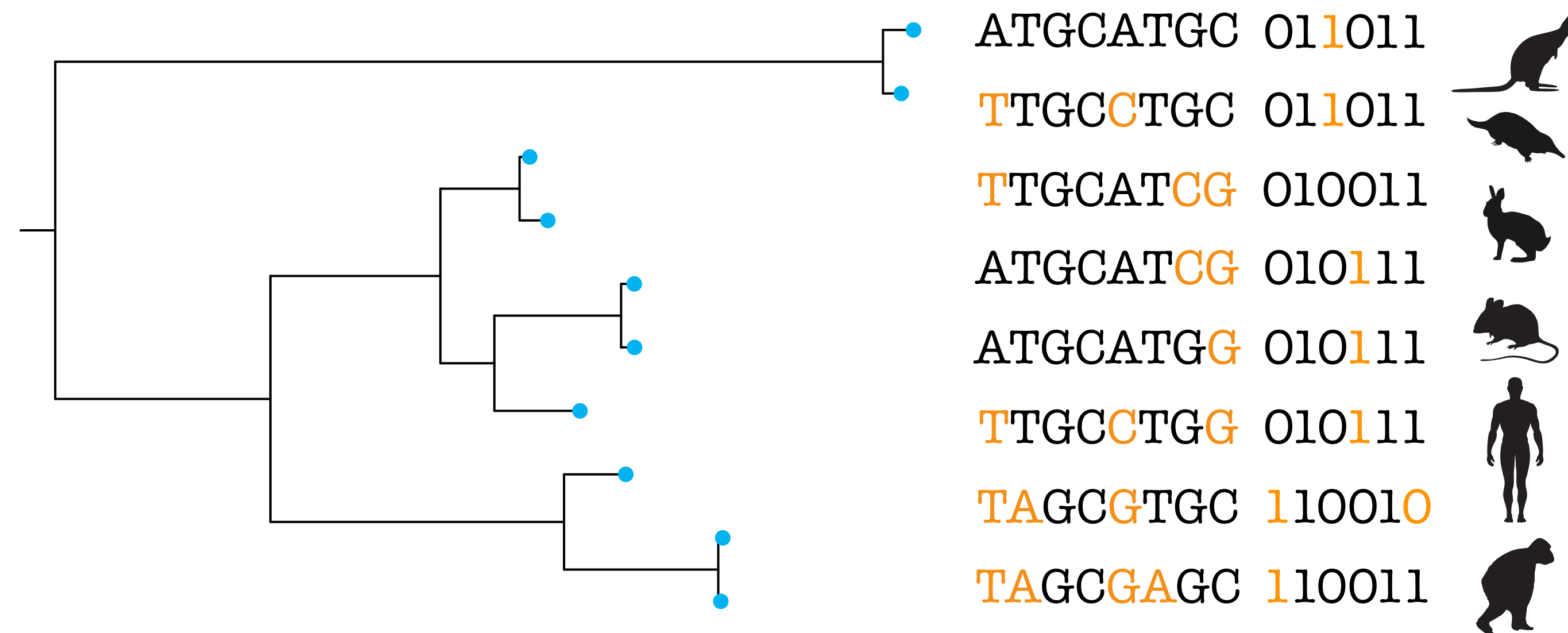
- Model Selection

- Model adequacy

Tip-dating and FBD tutorial with bears



Basic principles of divergence time estimation



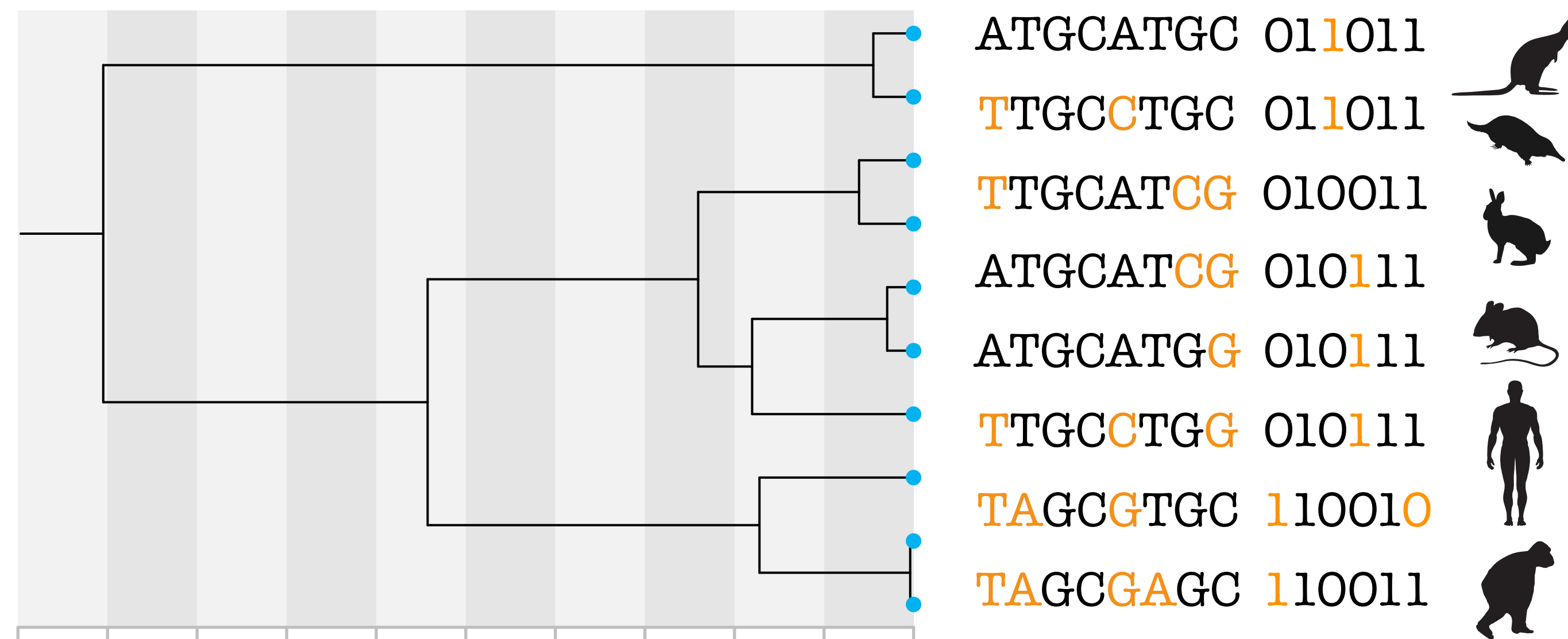
branch lengths = rate x time

$$v = rt$$



Zuckerkandl, Pauling, 1962, 1965

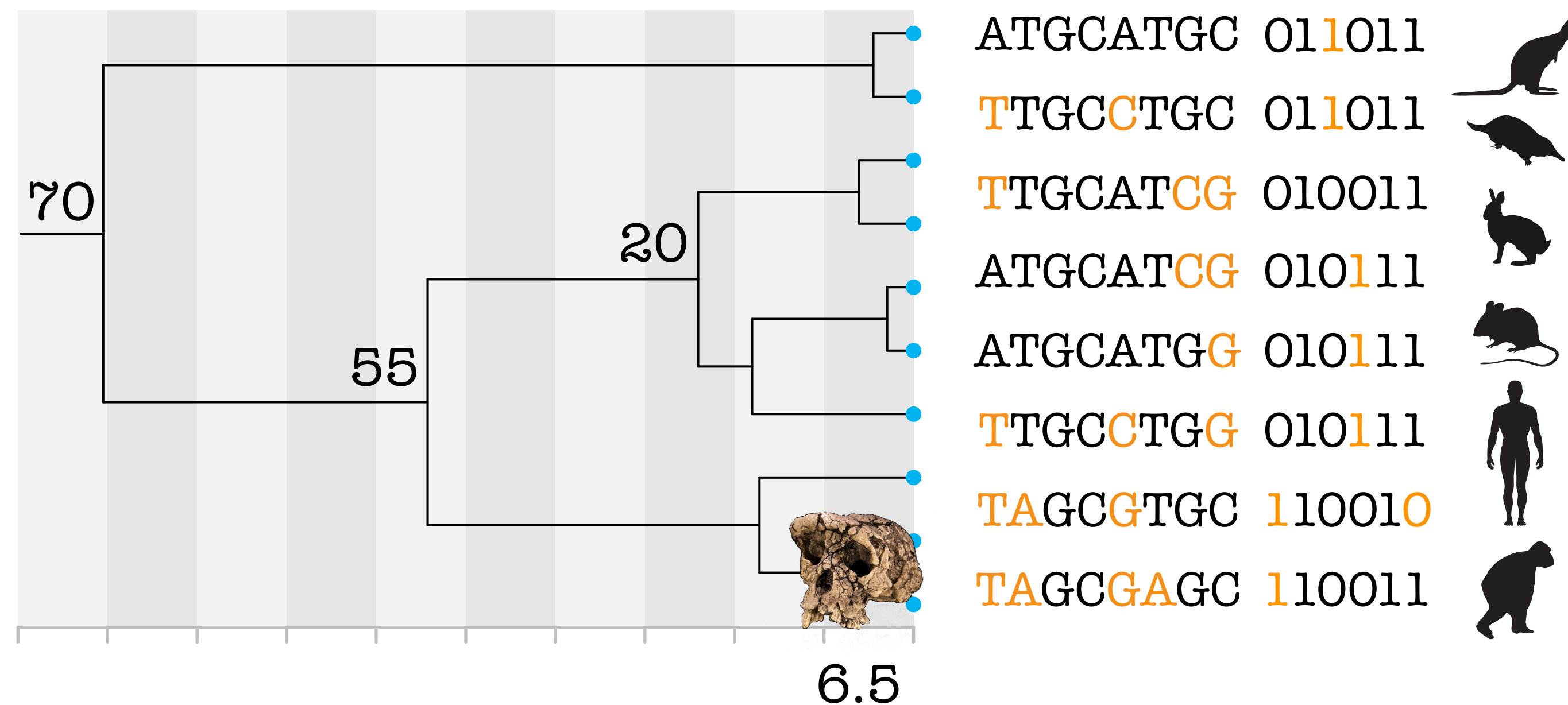
Basic principles of divergence time estimation



branch lengths = time



Basic principles of divergence time estimation

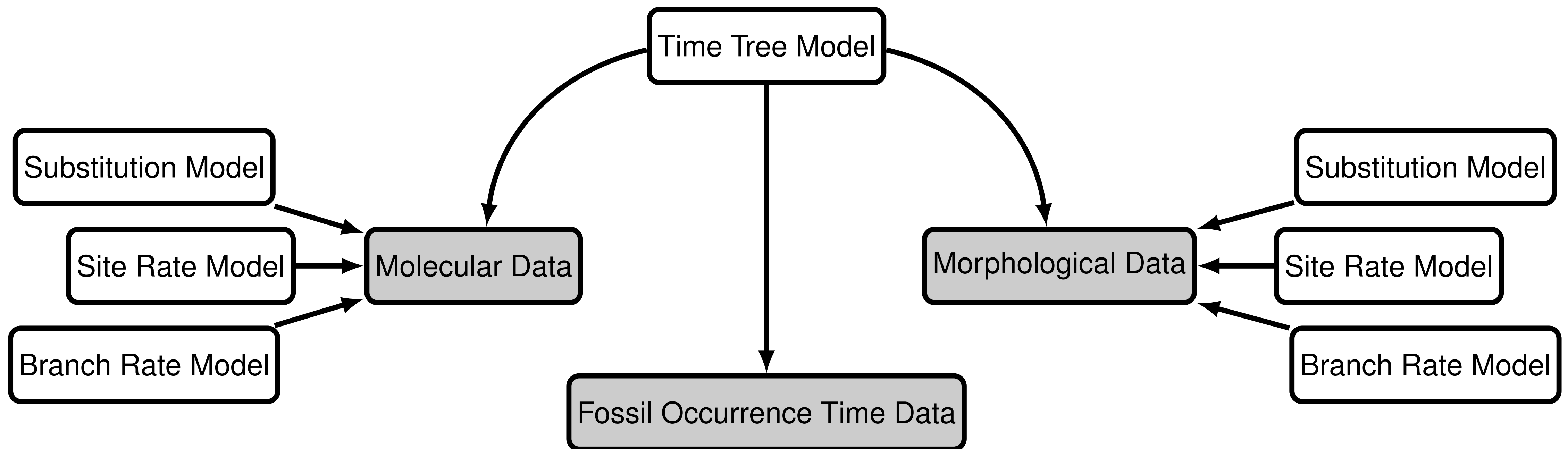


branch lengths = time



Zuckerkandl, Pauling, 1962, 1965

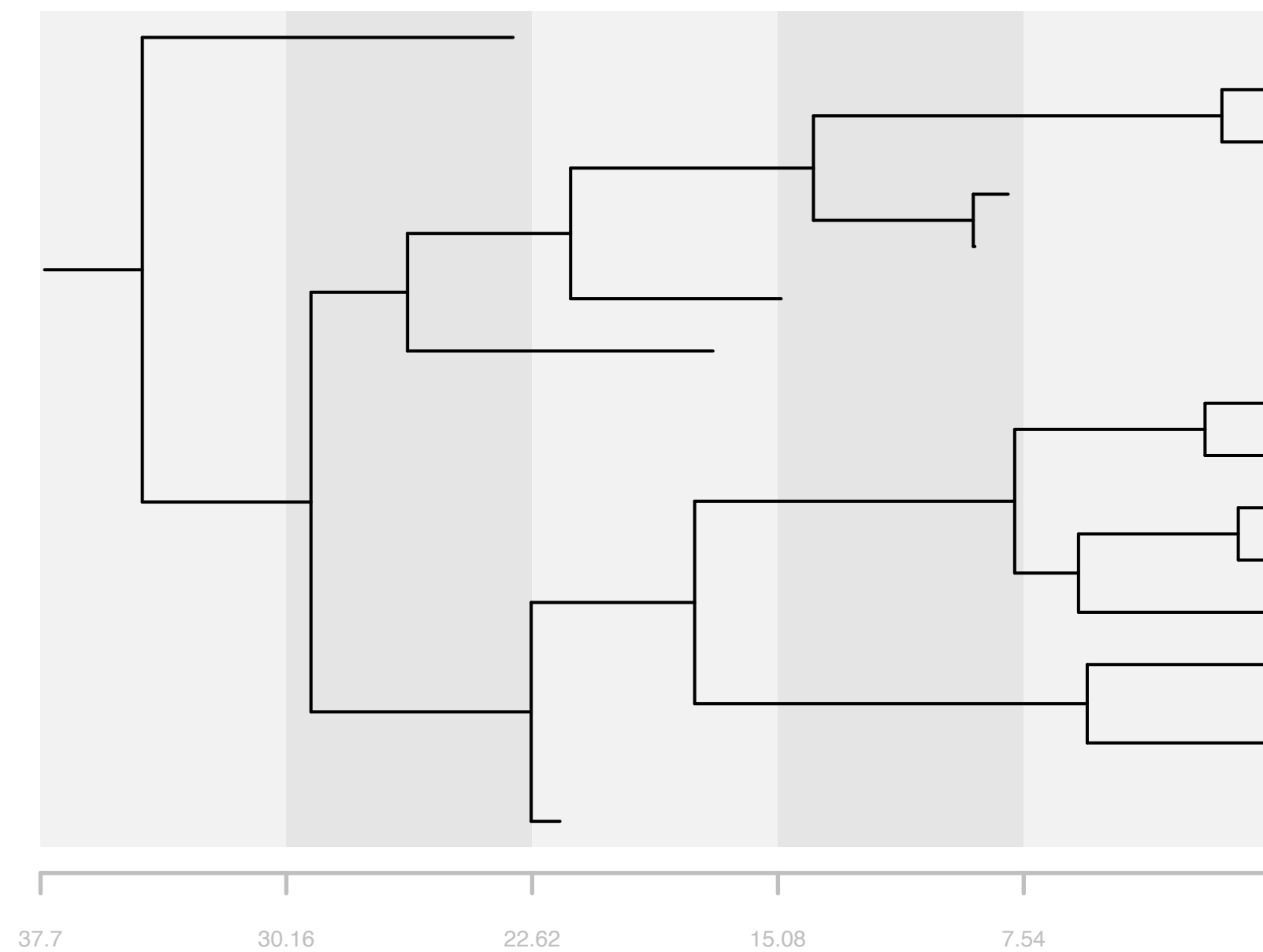
Divergence time estimation: a tricky business



Modular components of the graphical model used in the FBD tip-dating analysis described in the tutorial.

A “simple” model of evolution and sampling

Time tree model



λ

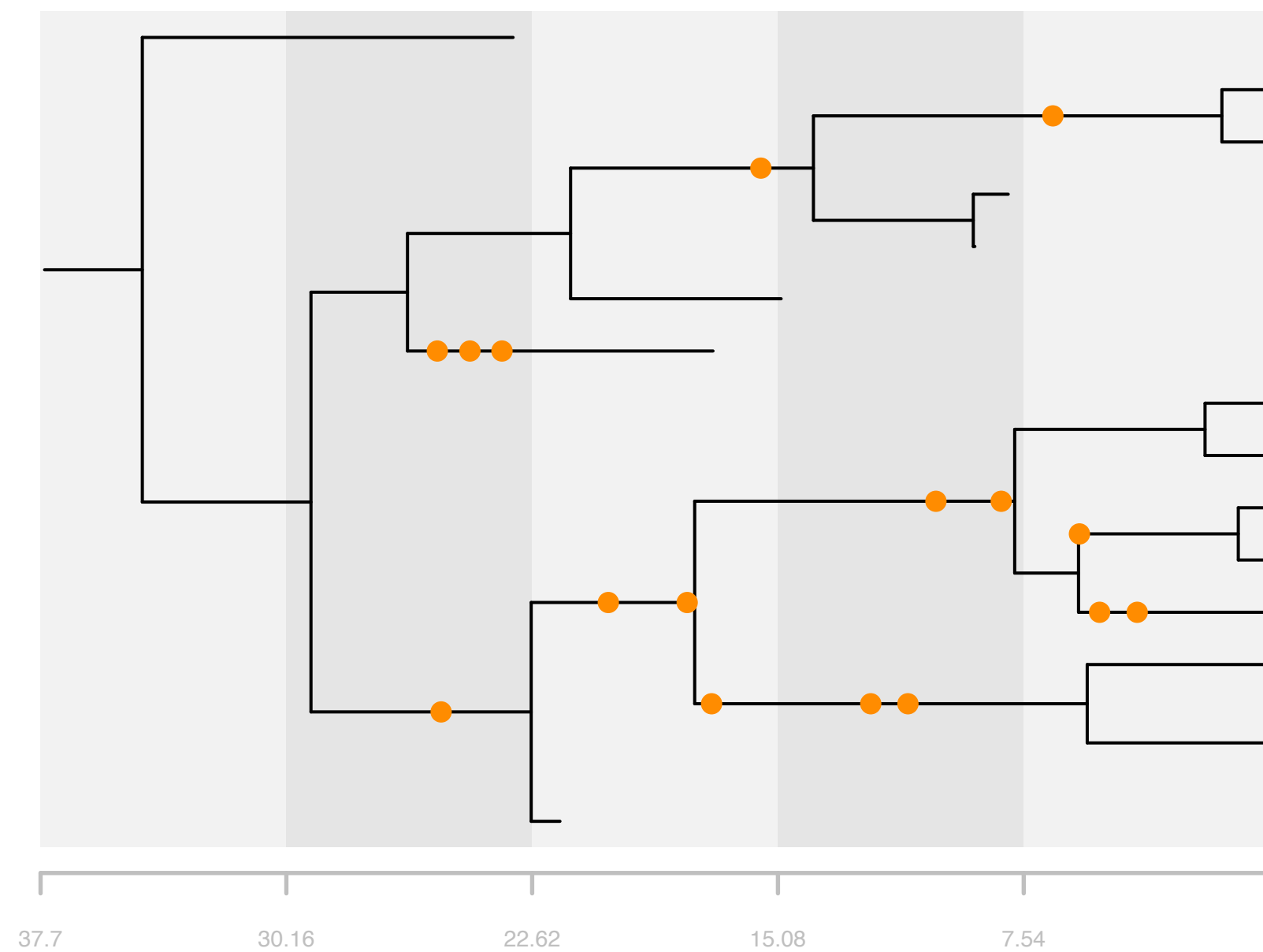
Branching (speciation) rate

μ

Extinction rate

A “simple” model of evolution and sampling

Time tree model



λ

Branching (speciation) rate

μ

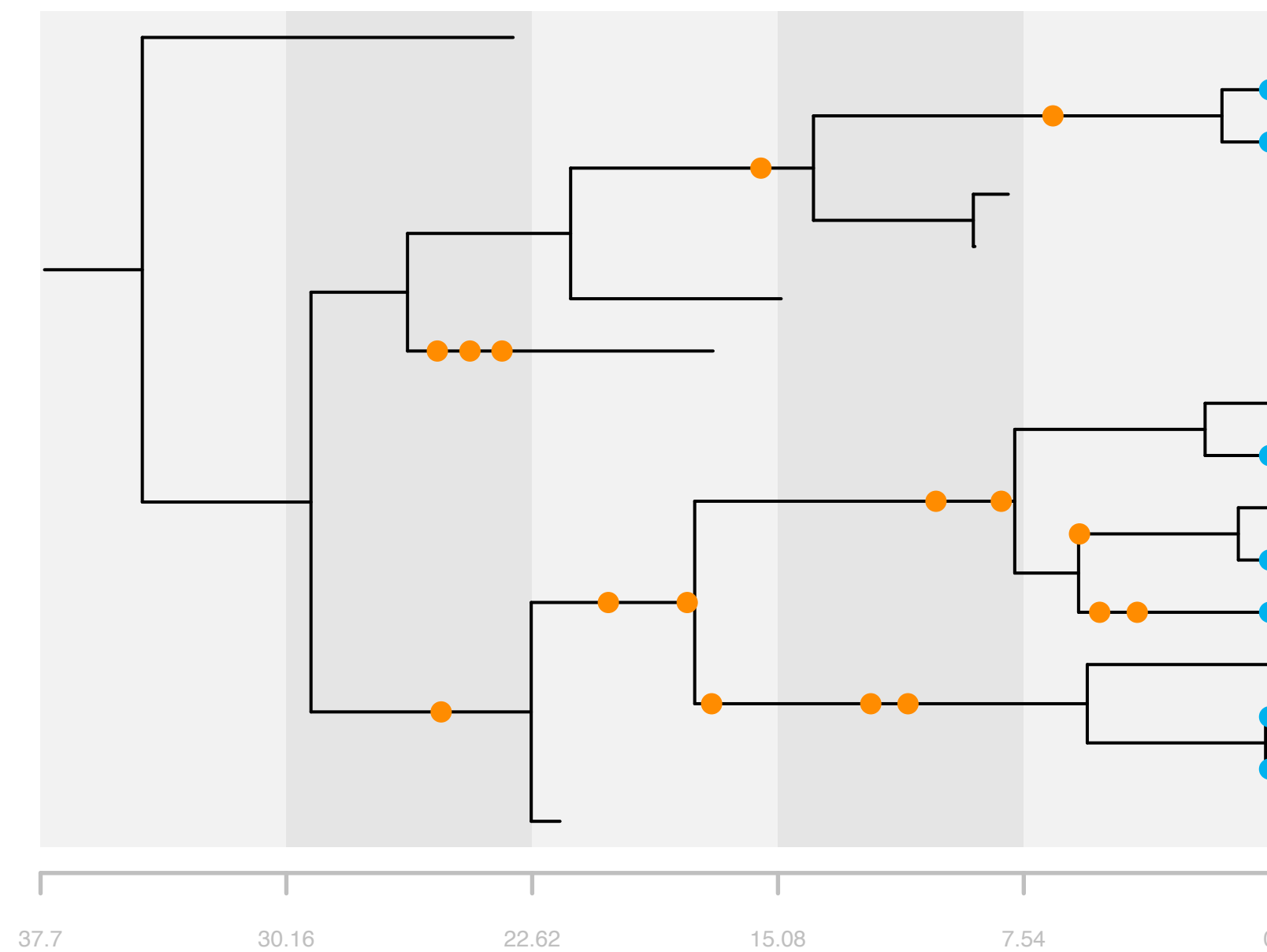
Extinction rate

ψ

Fossil sampling rate

A “simple” model of evolution and sampling

Time tree model



λ

Branching (speciation) rate

μ

Extinction rate

ψ

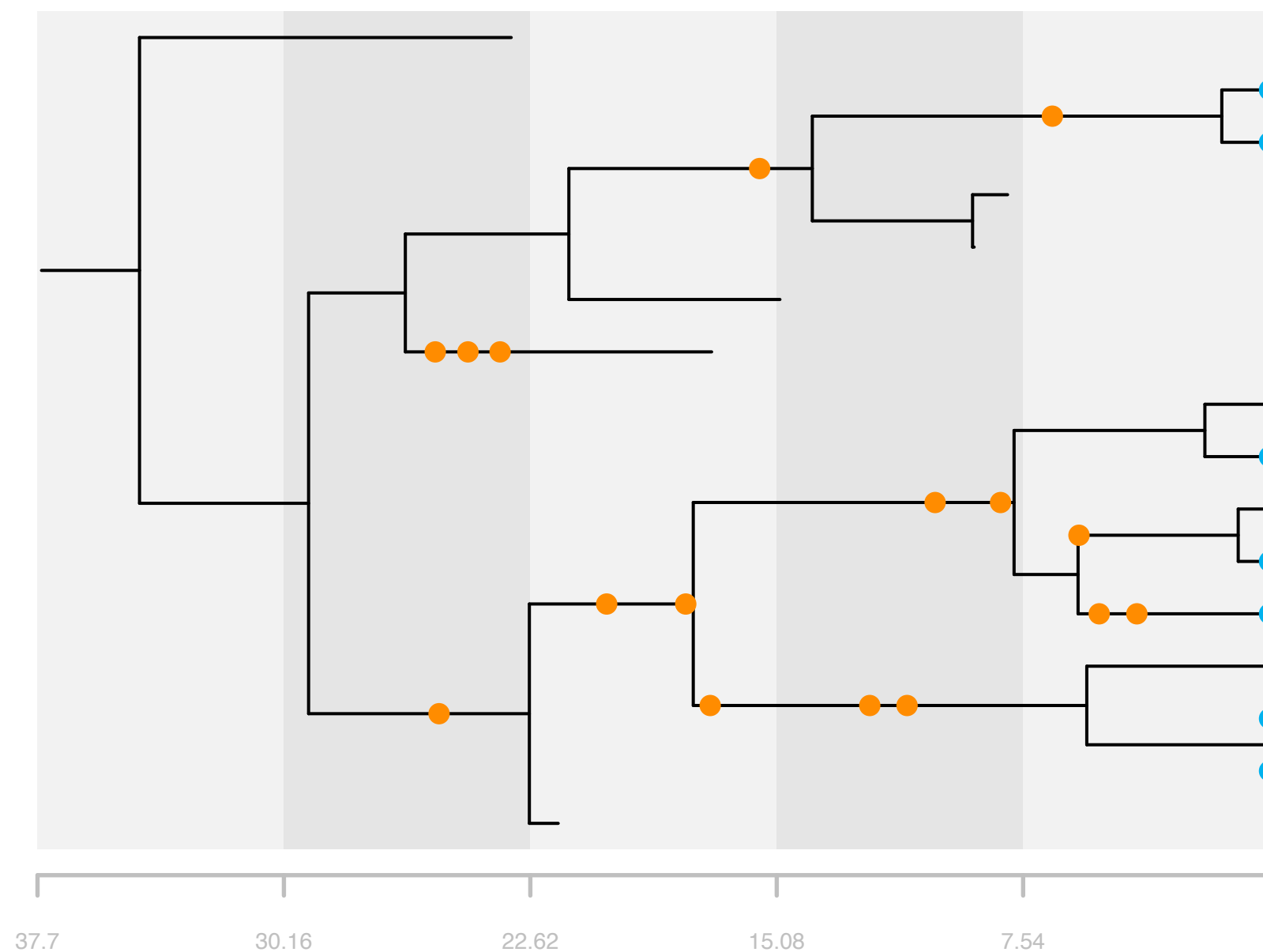
Fossil sampling rate

ρ

Extant species sampling probability

A “simple” model of evolution and sampling

Time tree model



λ

Branching (speciation) rate

μ

Extinction rate

ψ

Fossil sampling rate

ρ

Extant species sampling probability

t

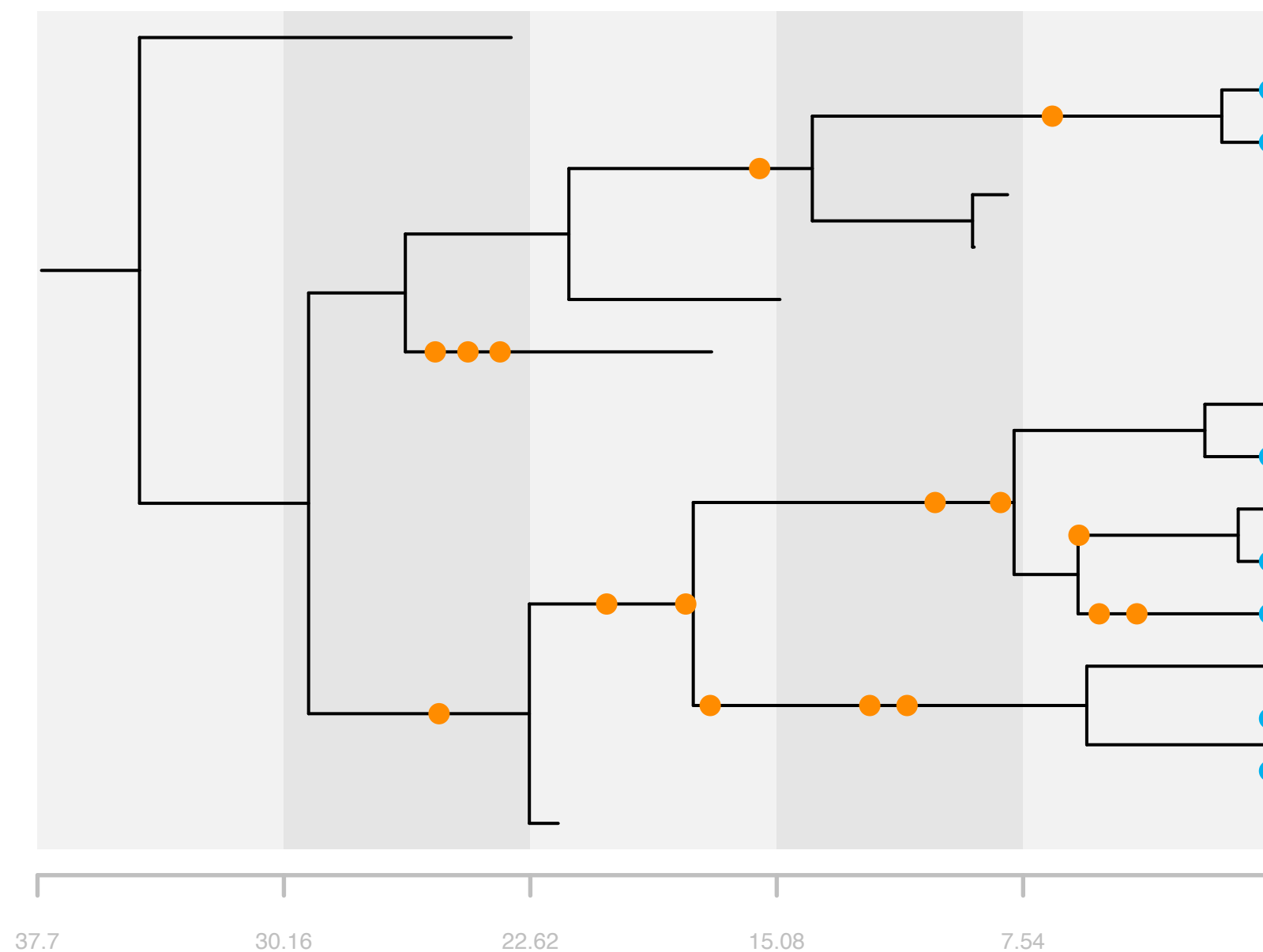
Divergence times (a stochastic outcome of this process)

F

Fossil occurrence times (a stochastic outcome of this process)

The fossilised birth-death process

Time tree model



λ

Branching (speciation) rate

μ

Extinction rate

ψ

Fossil sampling rate

ρ

Extant species sampling probability

t

Divergence times (a stochastic outcome of this process)

F

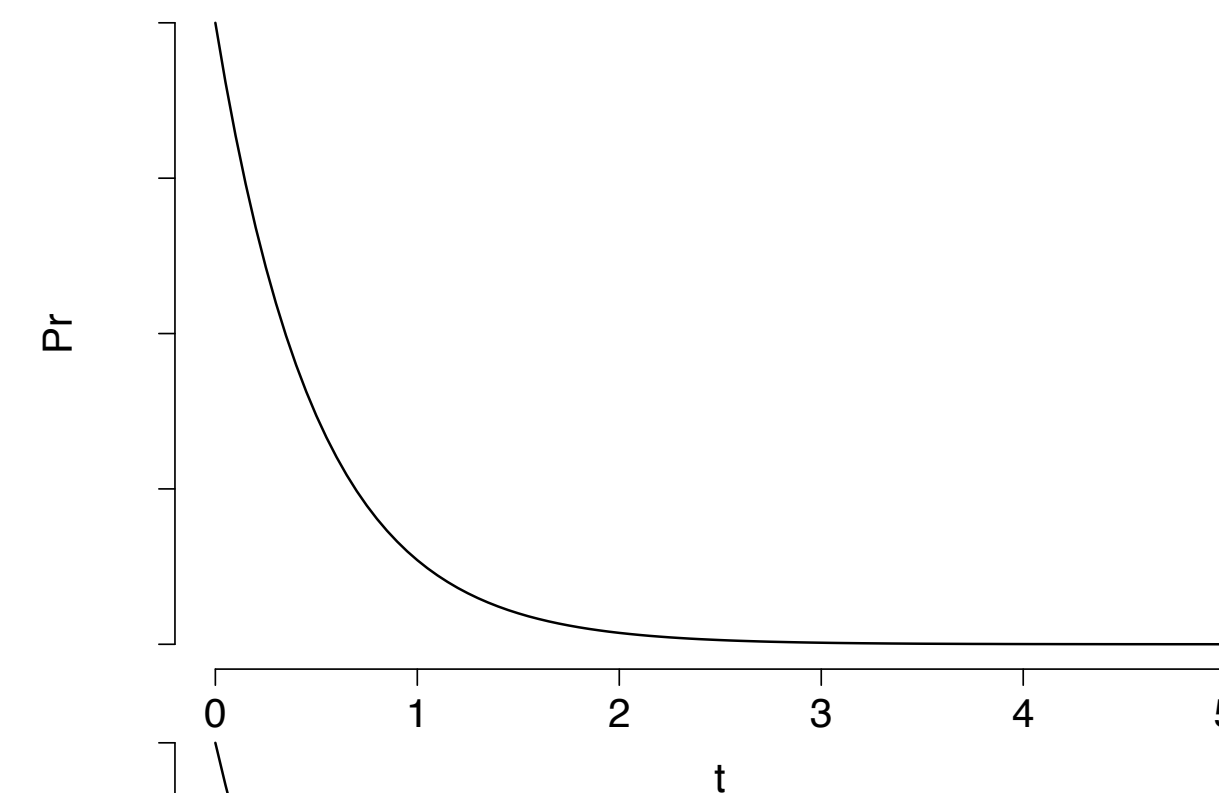
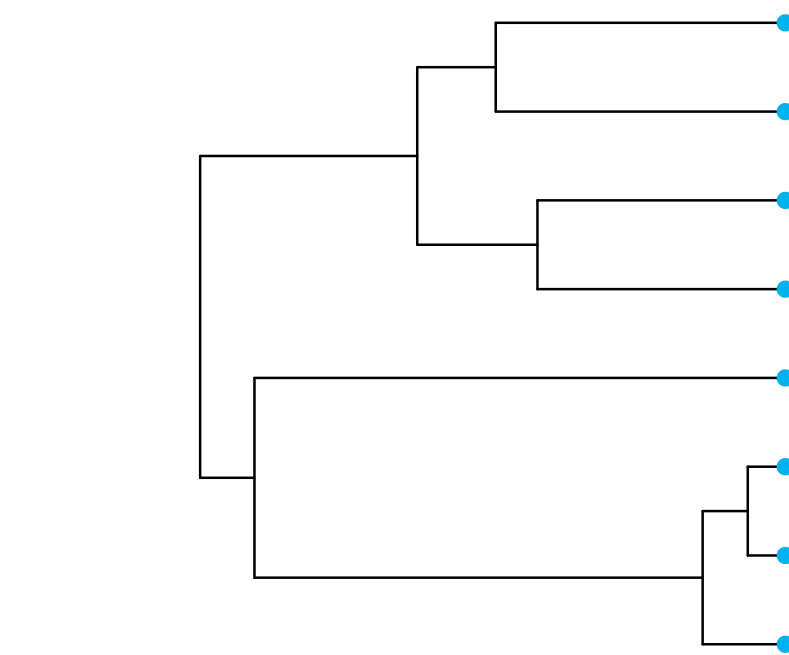
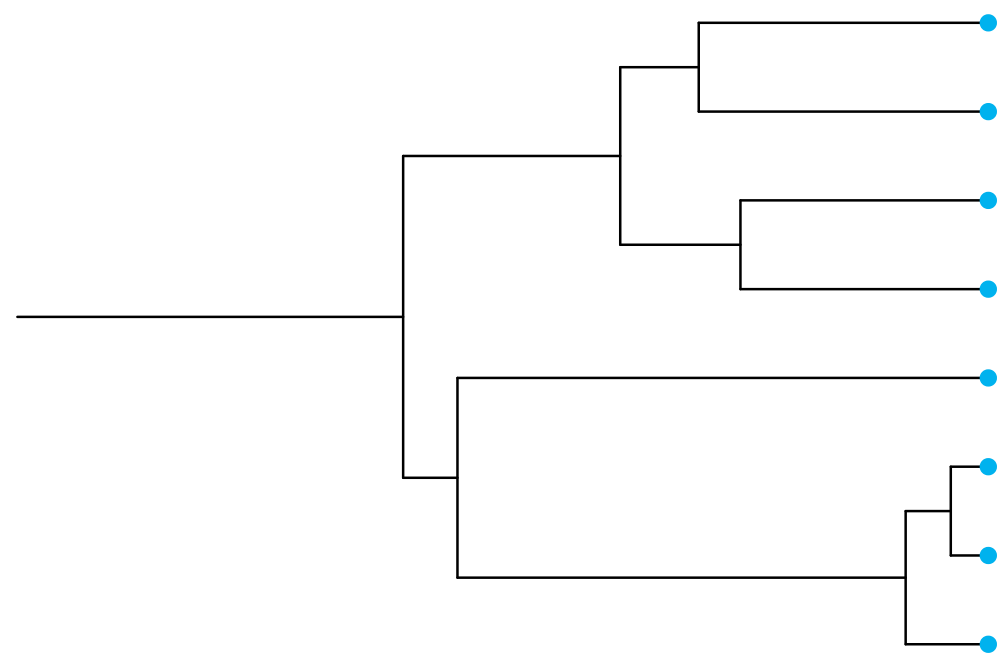
Fossil occurrence times (a stochastic outcome of this process)

Stadler, 2009, *JTB*, 26:58-66

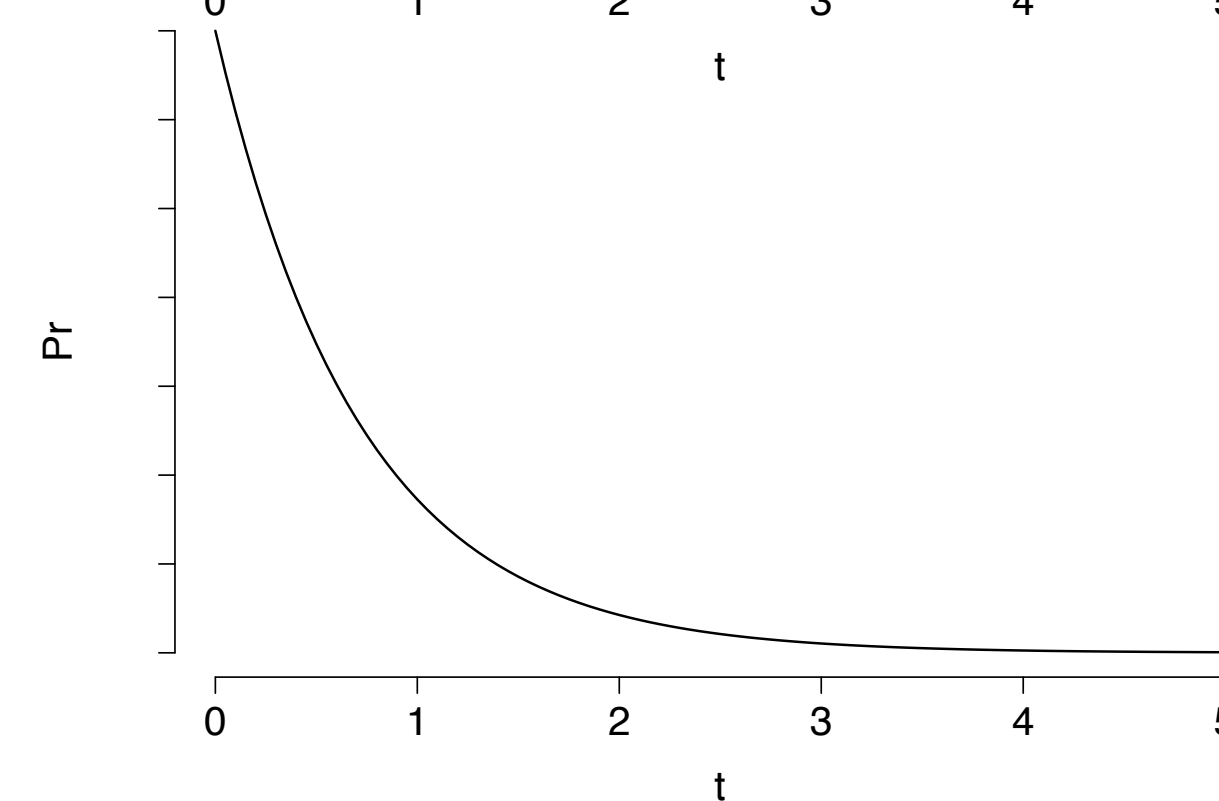
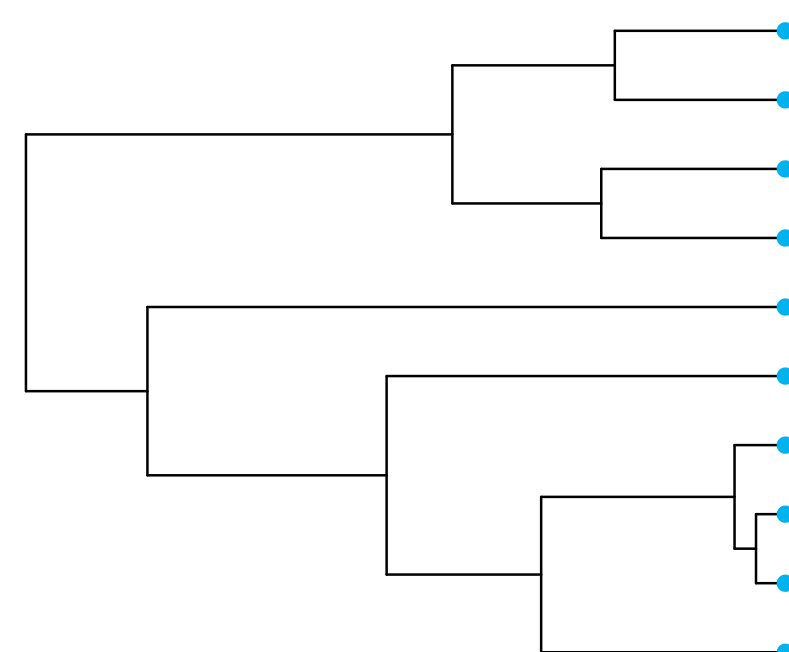
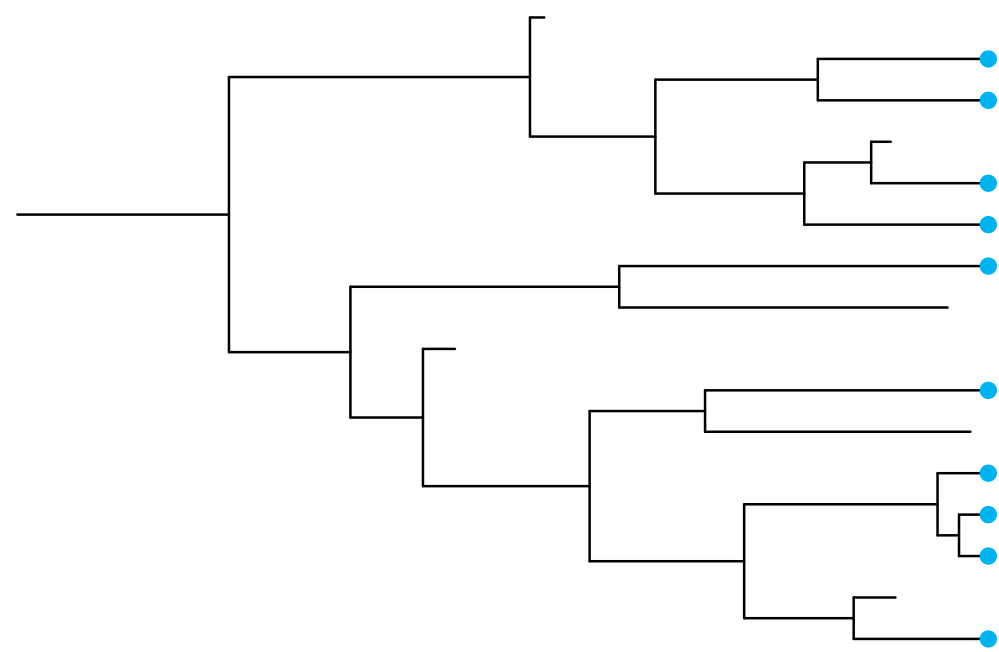
Previous birth-death tree priors (side note)

Time tree model

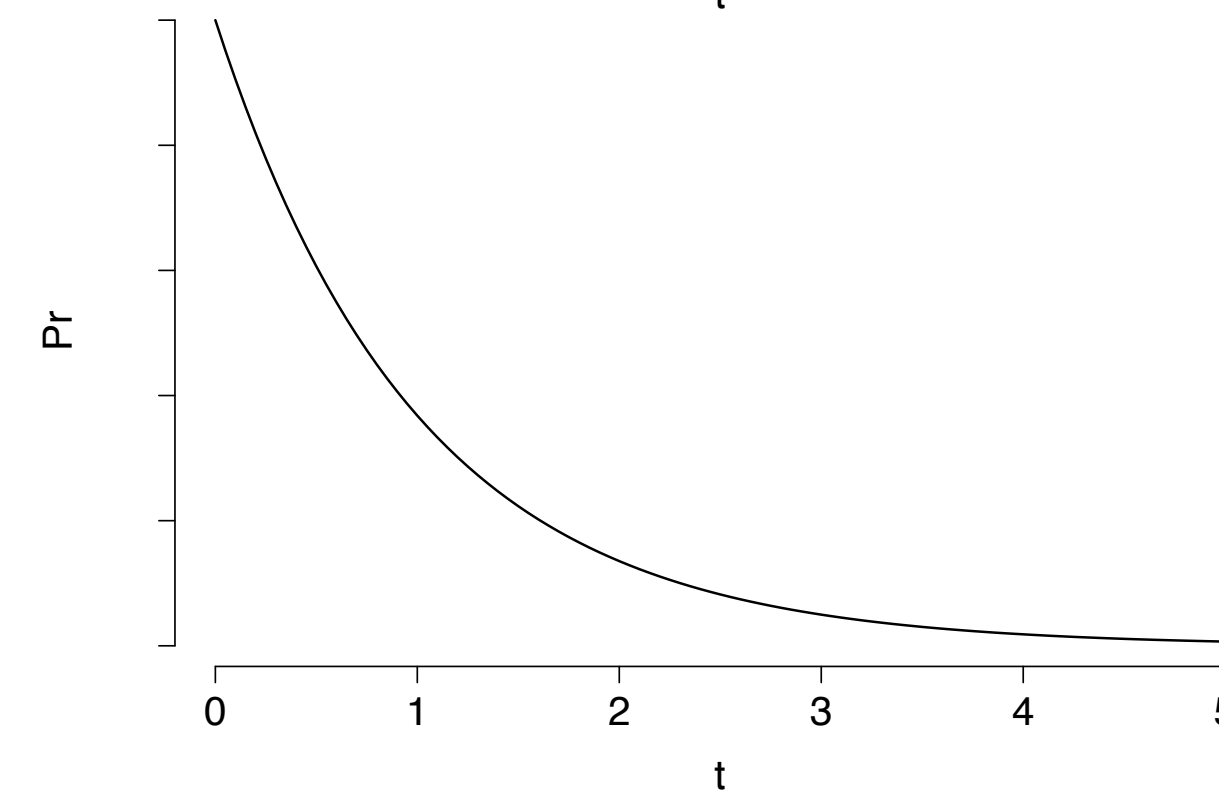
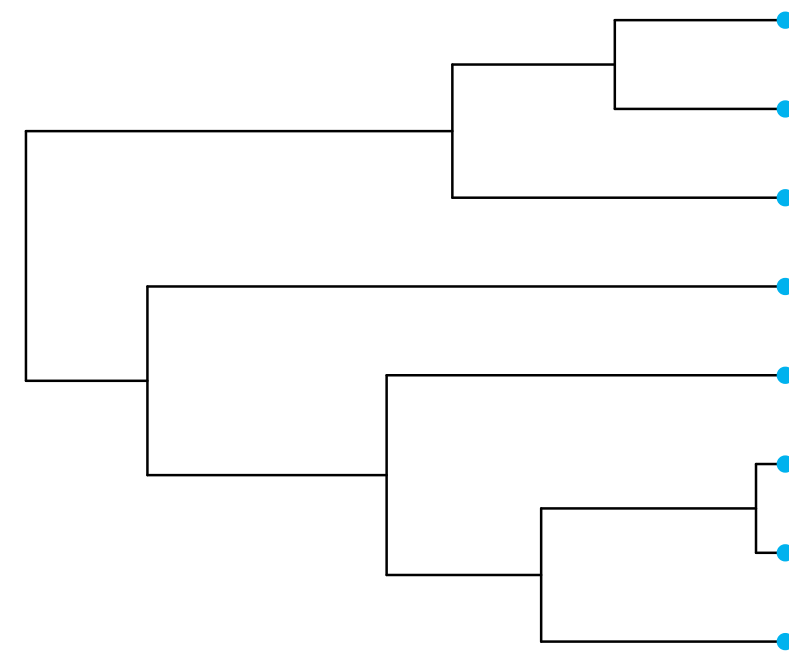
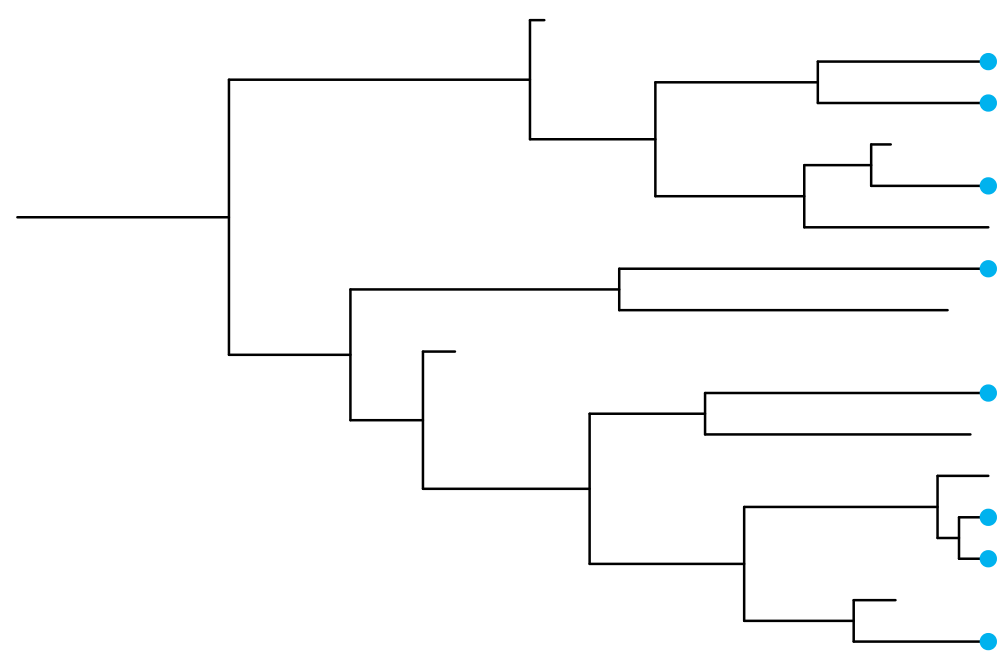
$\lambda = 1$



$\lambda = 1, \mu = 0.3$

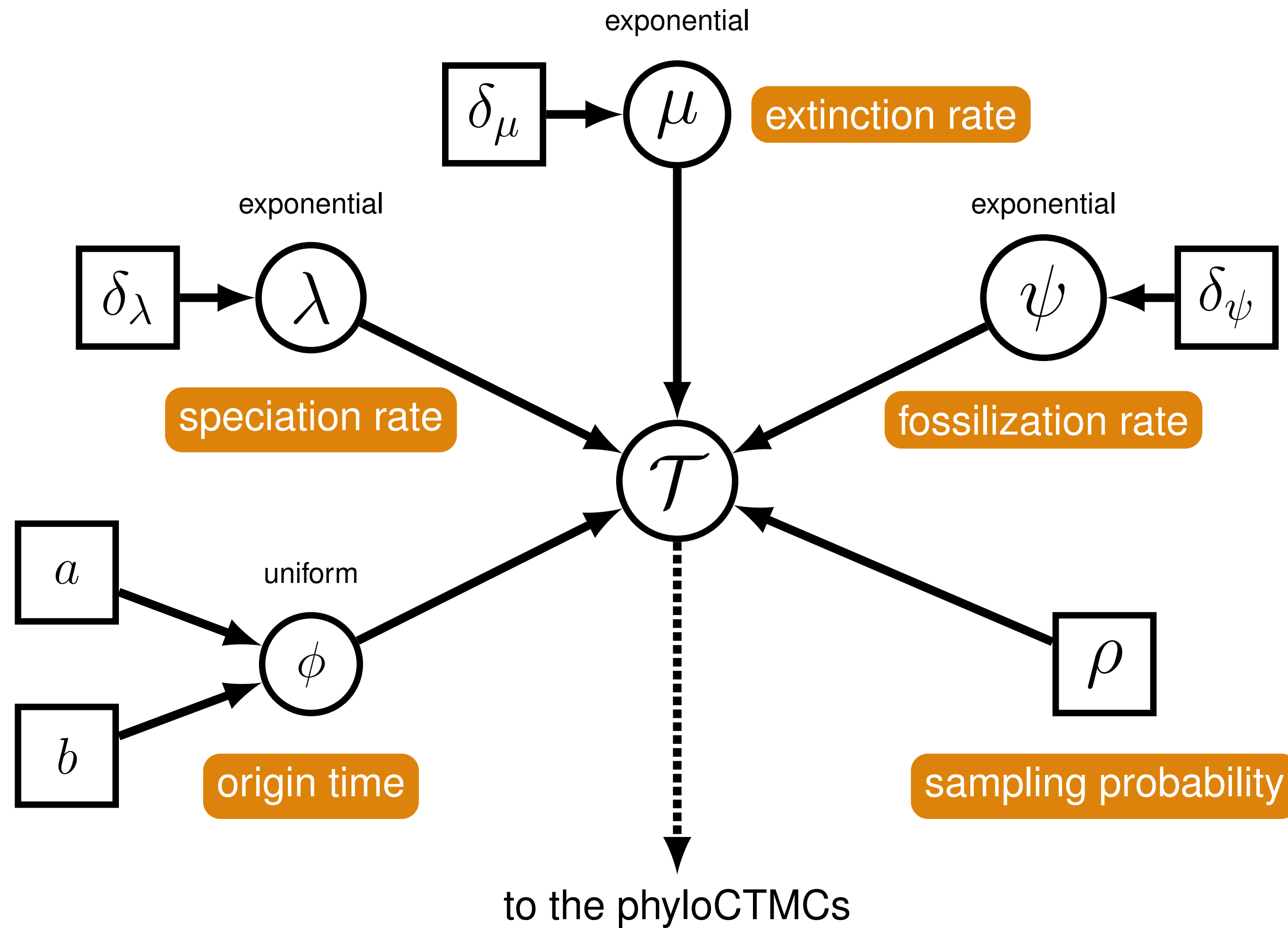


$\lambda = 1, \mu = 0.3, \rho = 0.5$



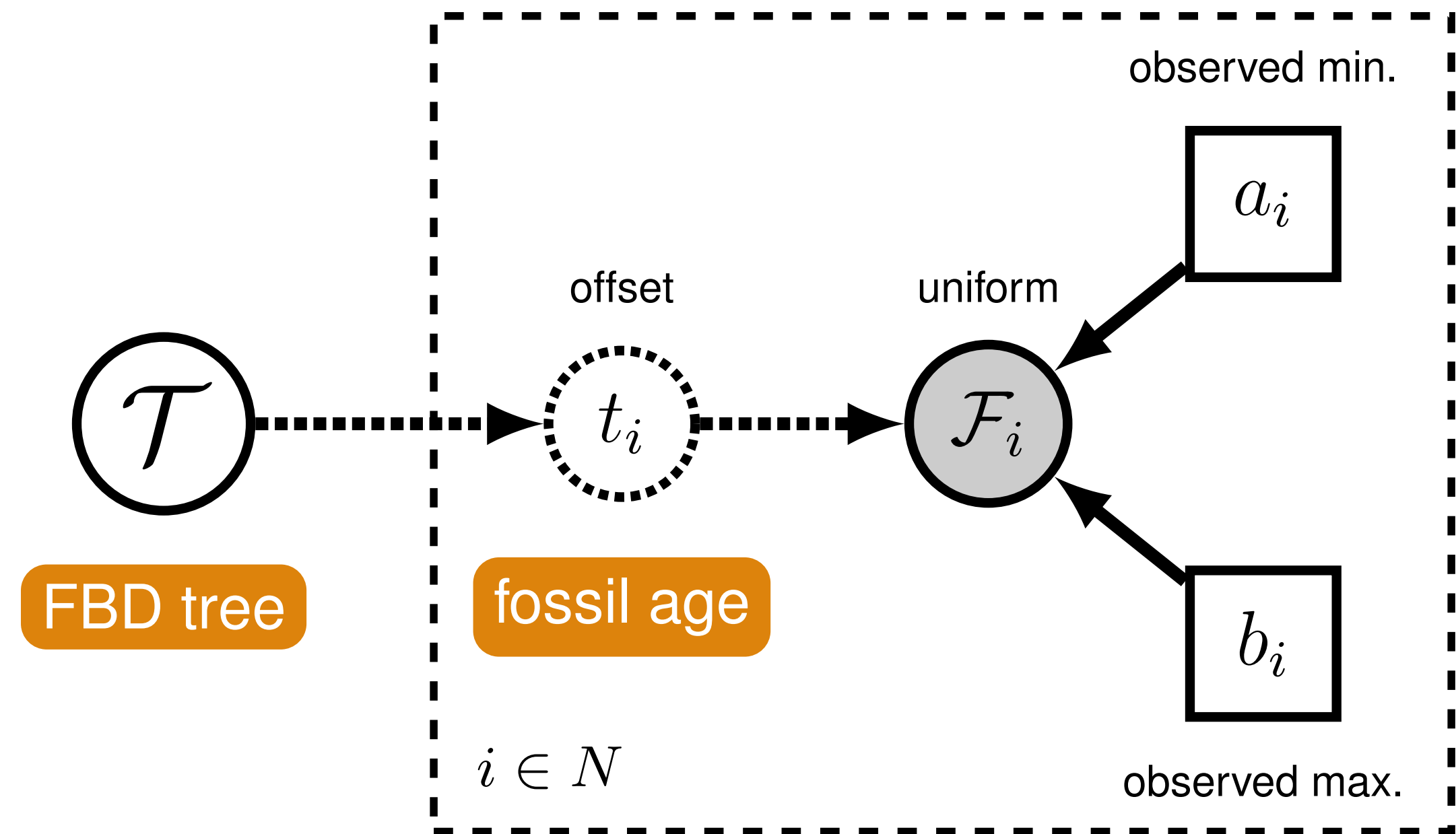
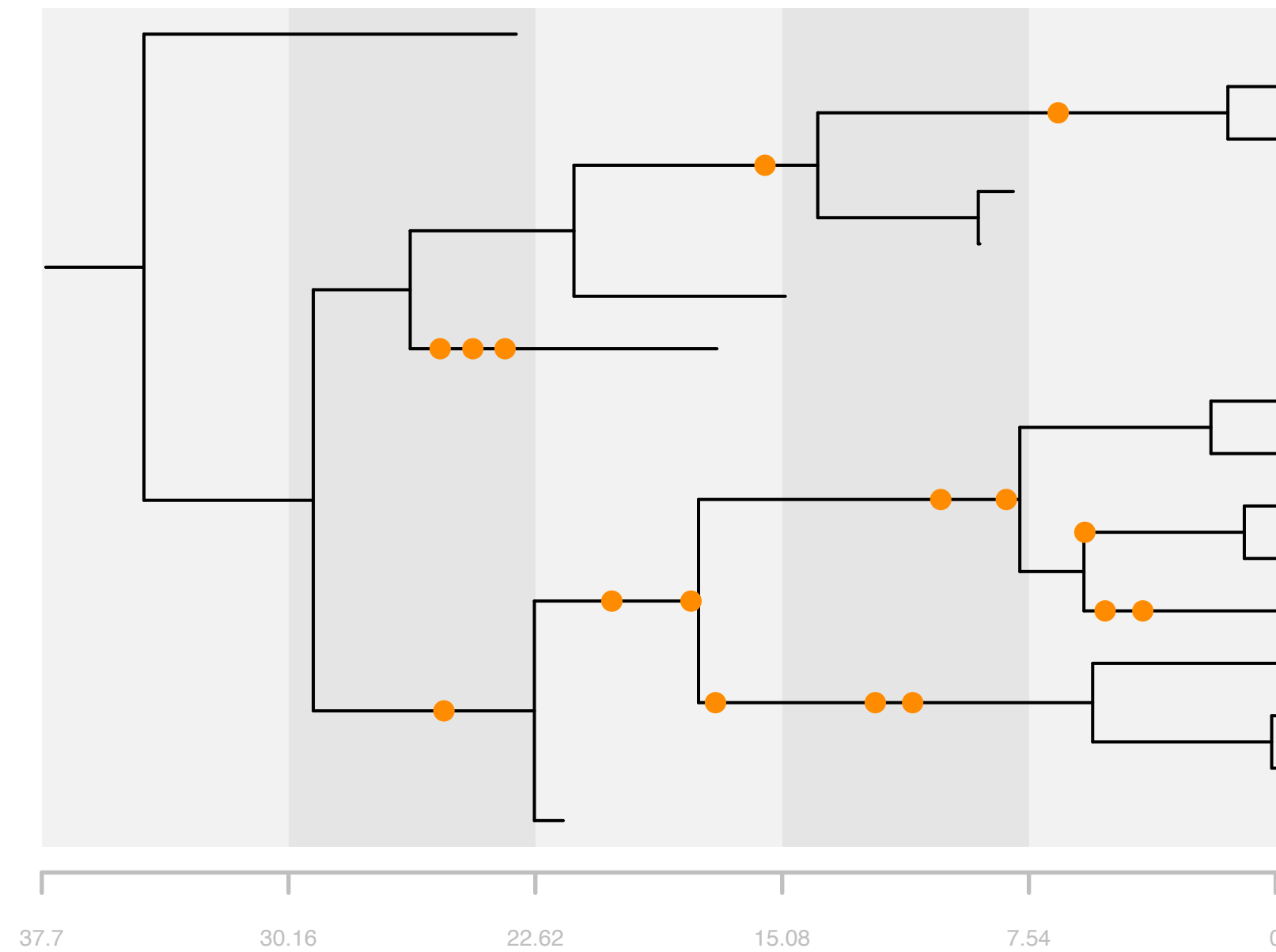
The fossilised birth-death process (= tree prior)

Time tree model



Stratigraphic age uncertainty

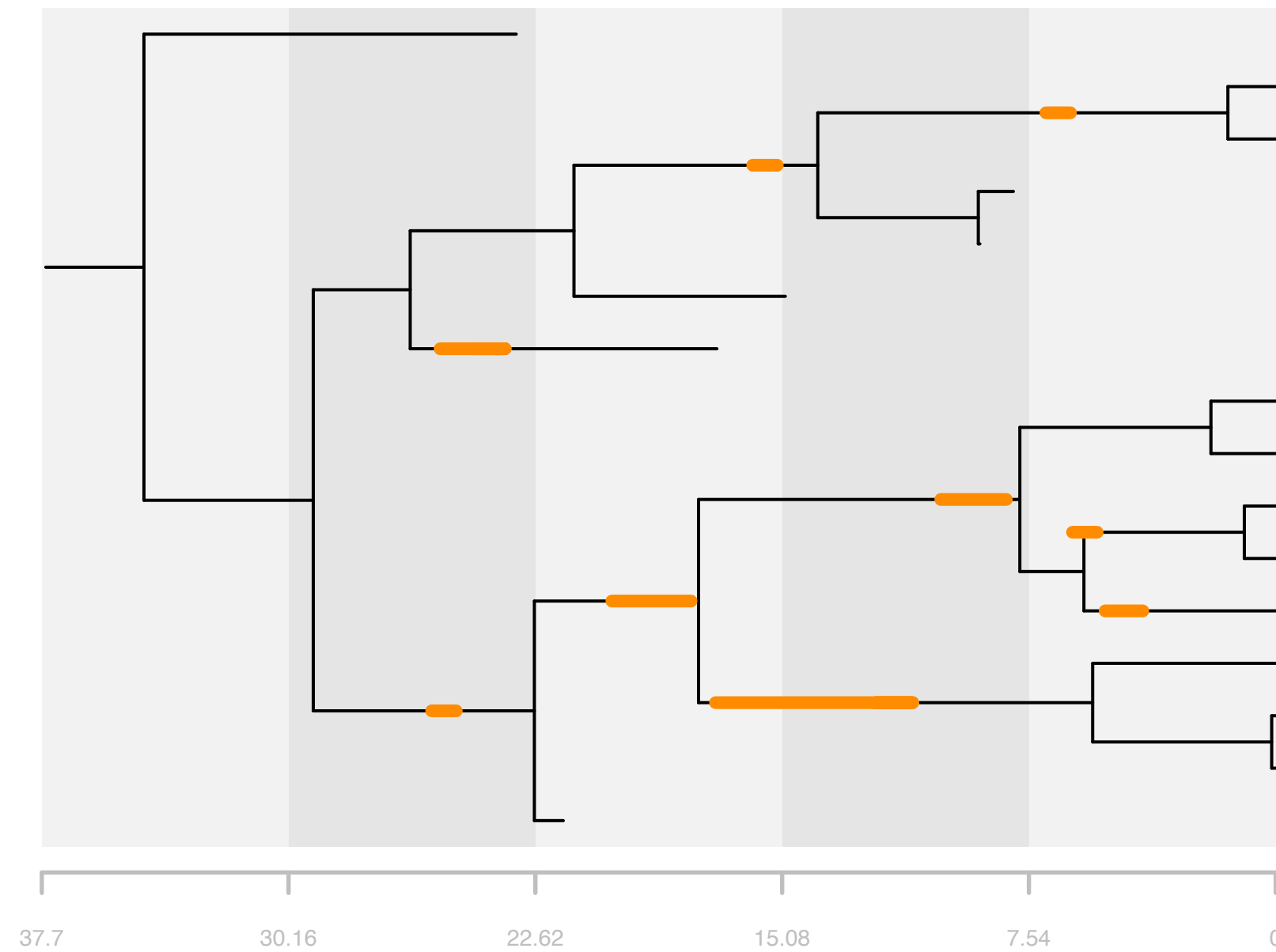
Fossil Occ.



Stratigraphic range data

Fossil Occ.

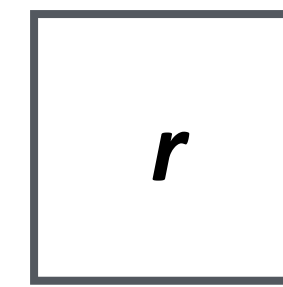
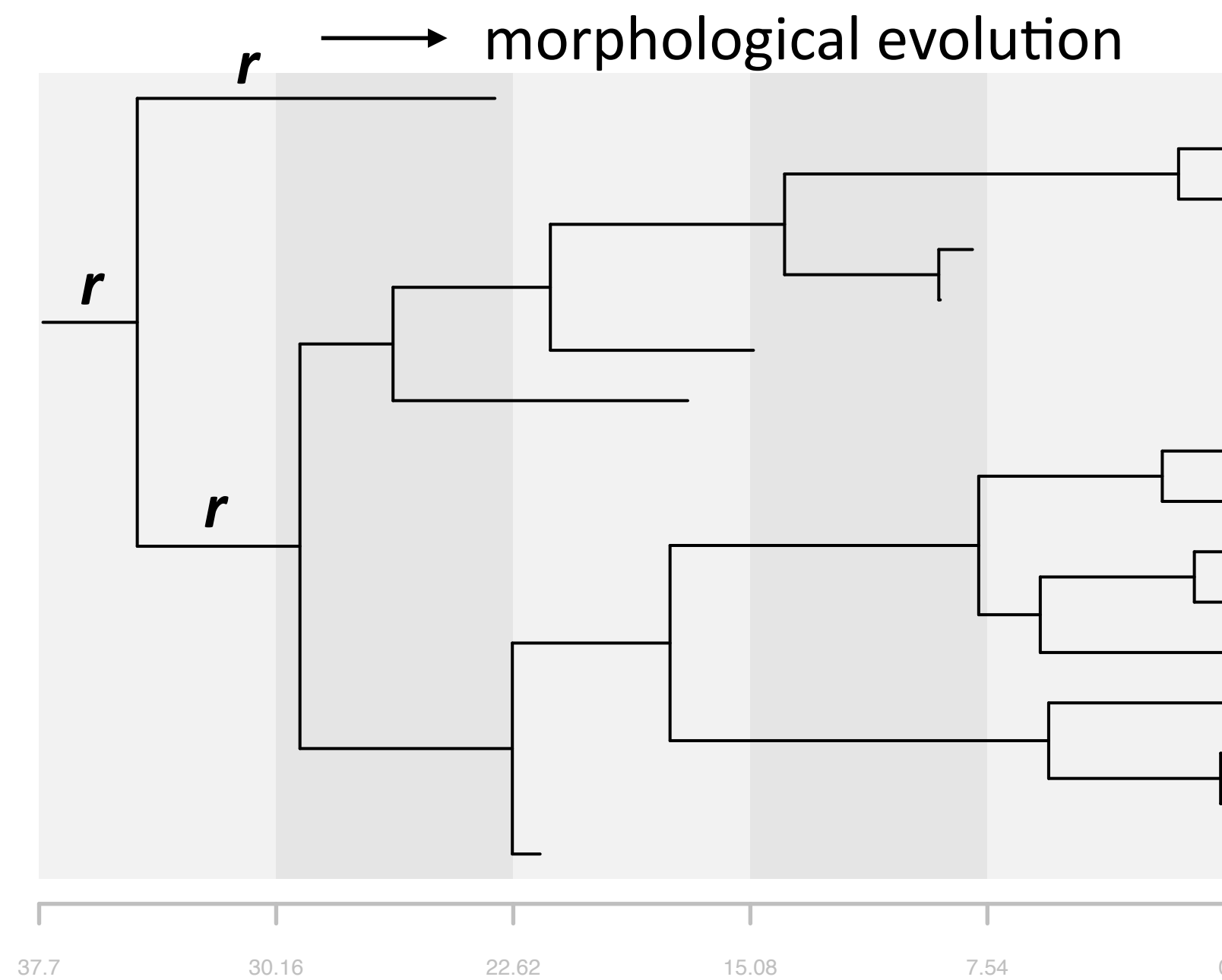
Time tree model



Watch this RevBayes space!!

A model of character evolution across branches (clock model)

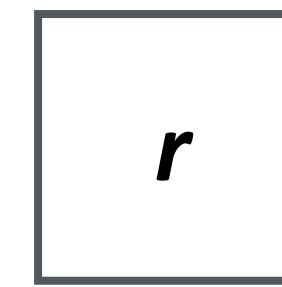
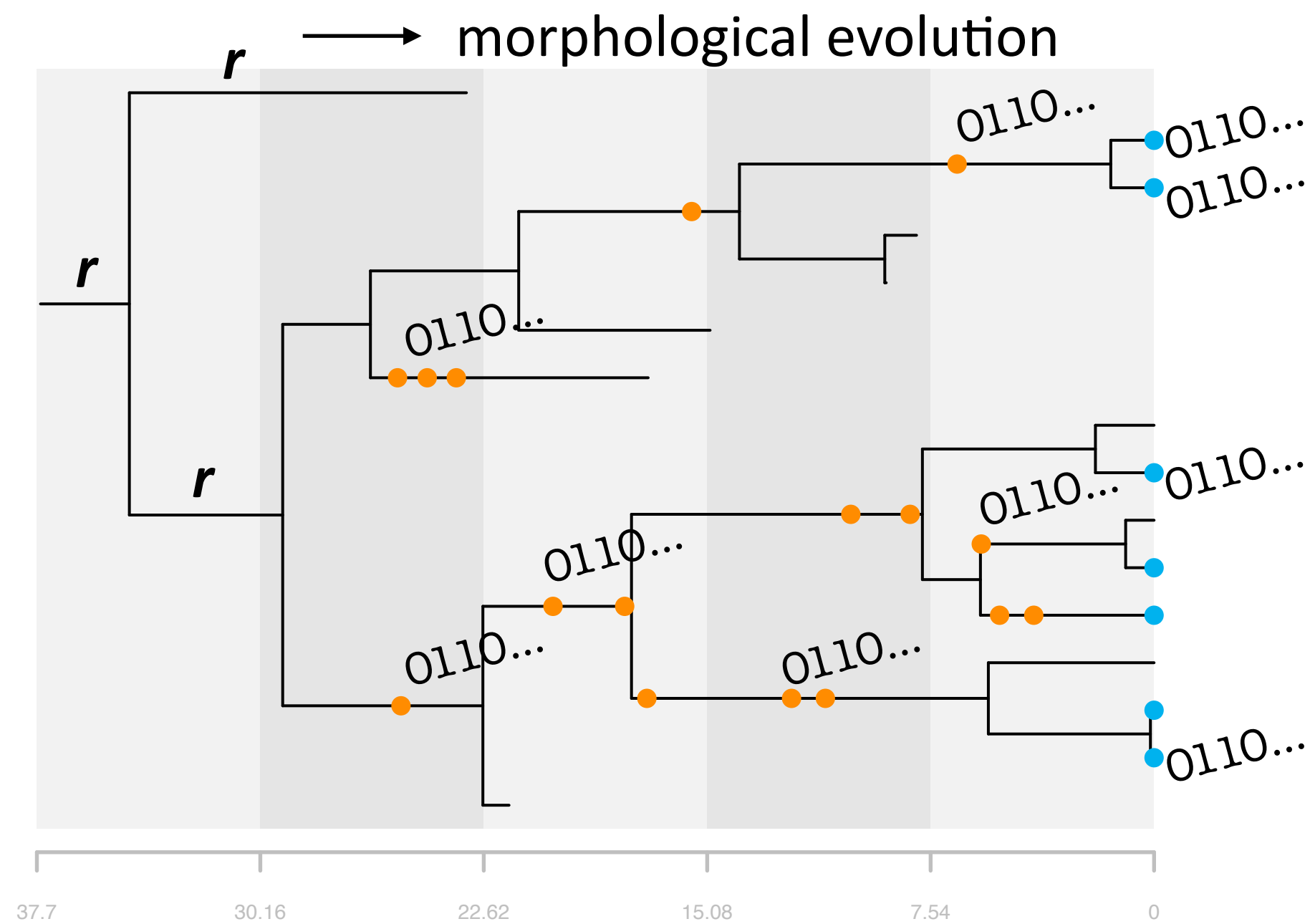
Branch rates



Substitution rate
substitution/character/myr

A model of character evolution across branches (clock model)

Branch rates

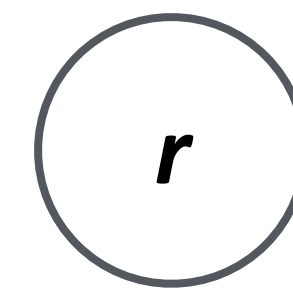
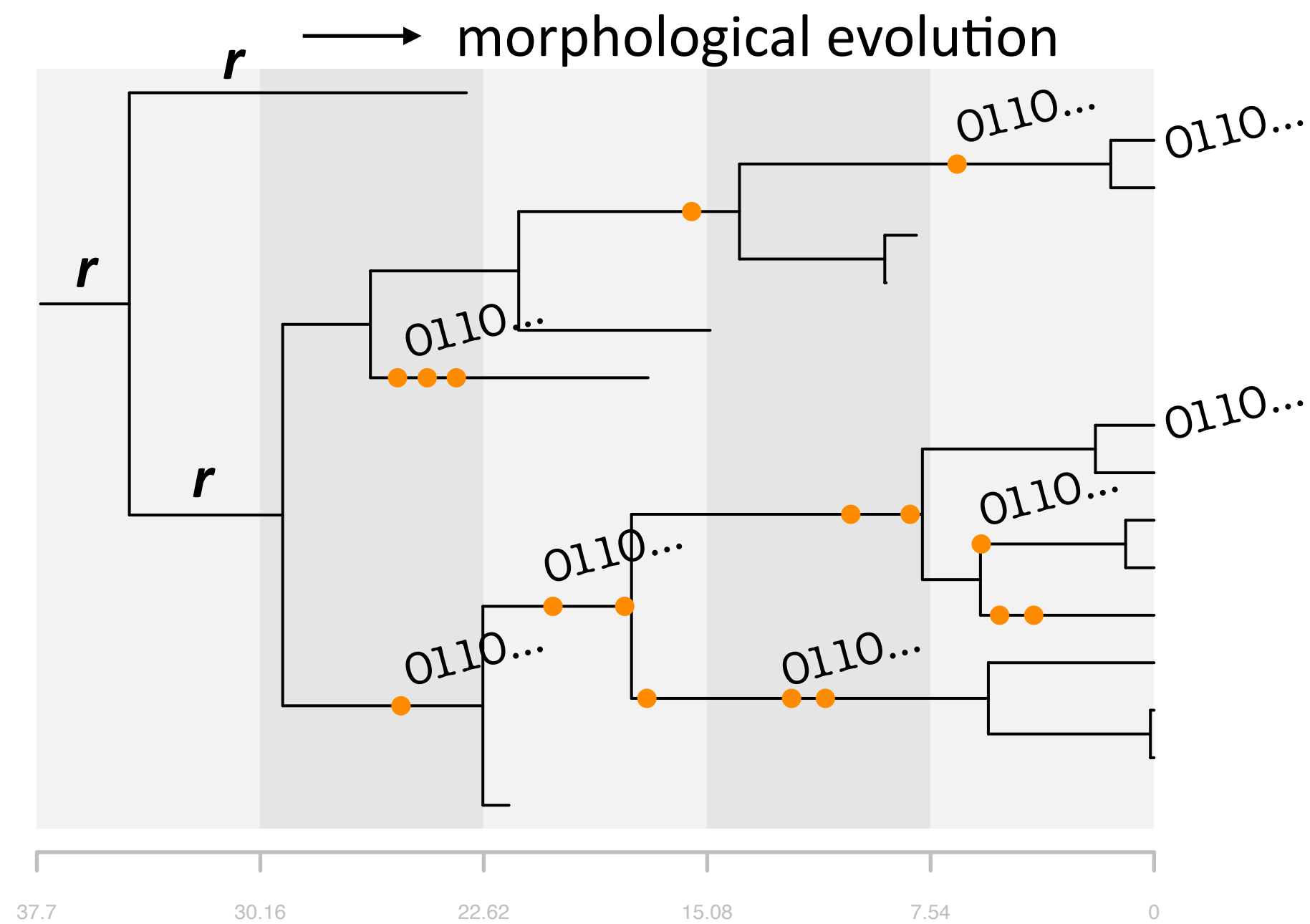


Substitution rate
substitution/character/myr

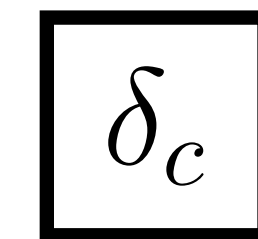
Not all fossils need to have data to be useful!
Fossils are valuable even without character data!

A model of character evolution across branches (clock model)

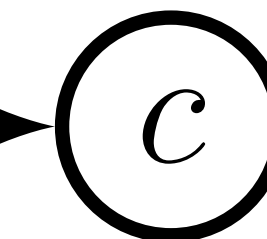
Branch rates



Substitution rate
substitution/character/myr



exponential



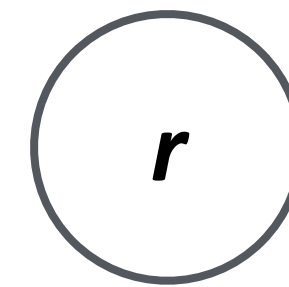
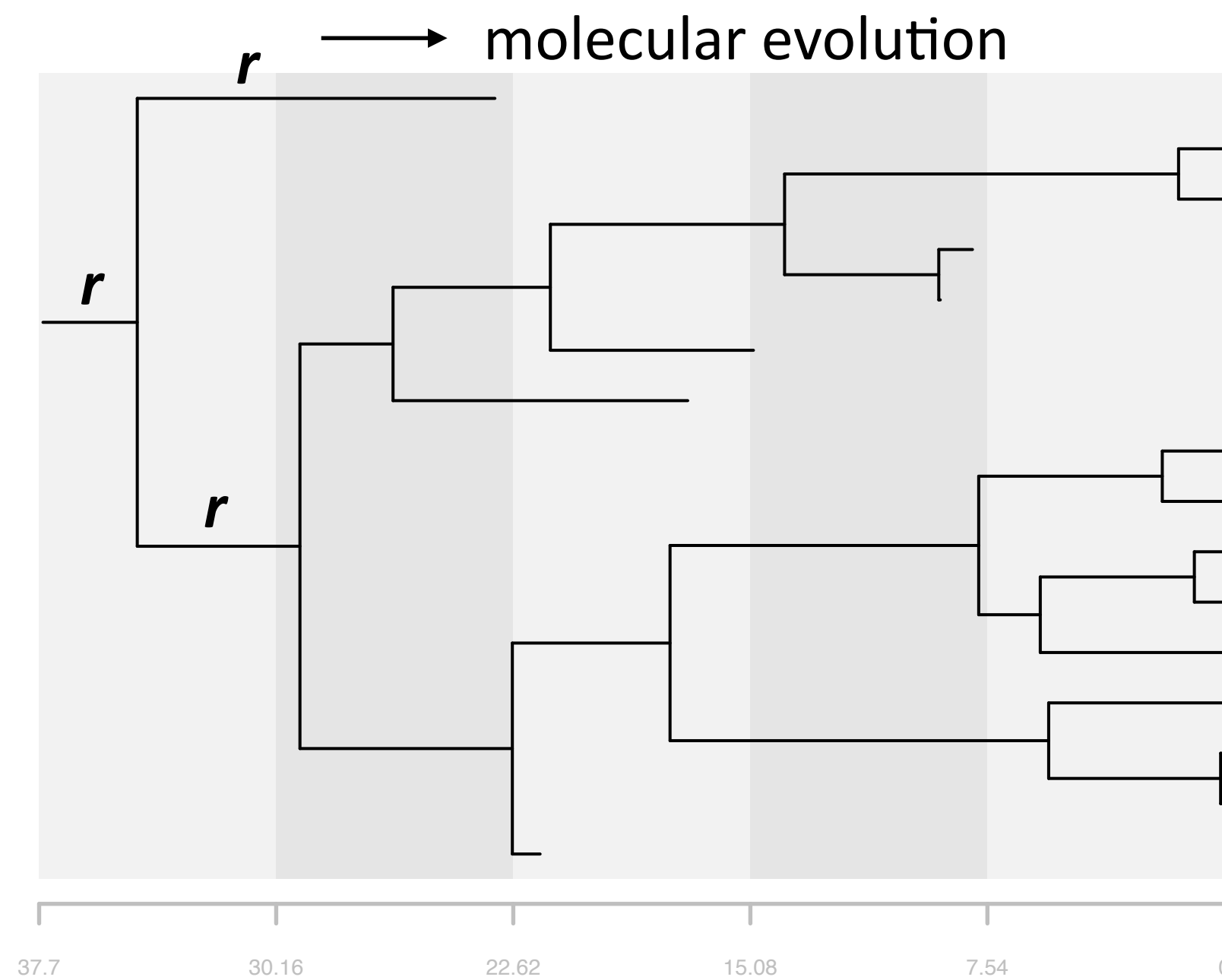
clock rate

to the phyloCTMC

Strict clock model

A model of character evolution across branches (clock model)

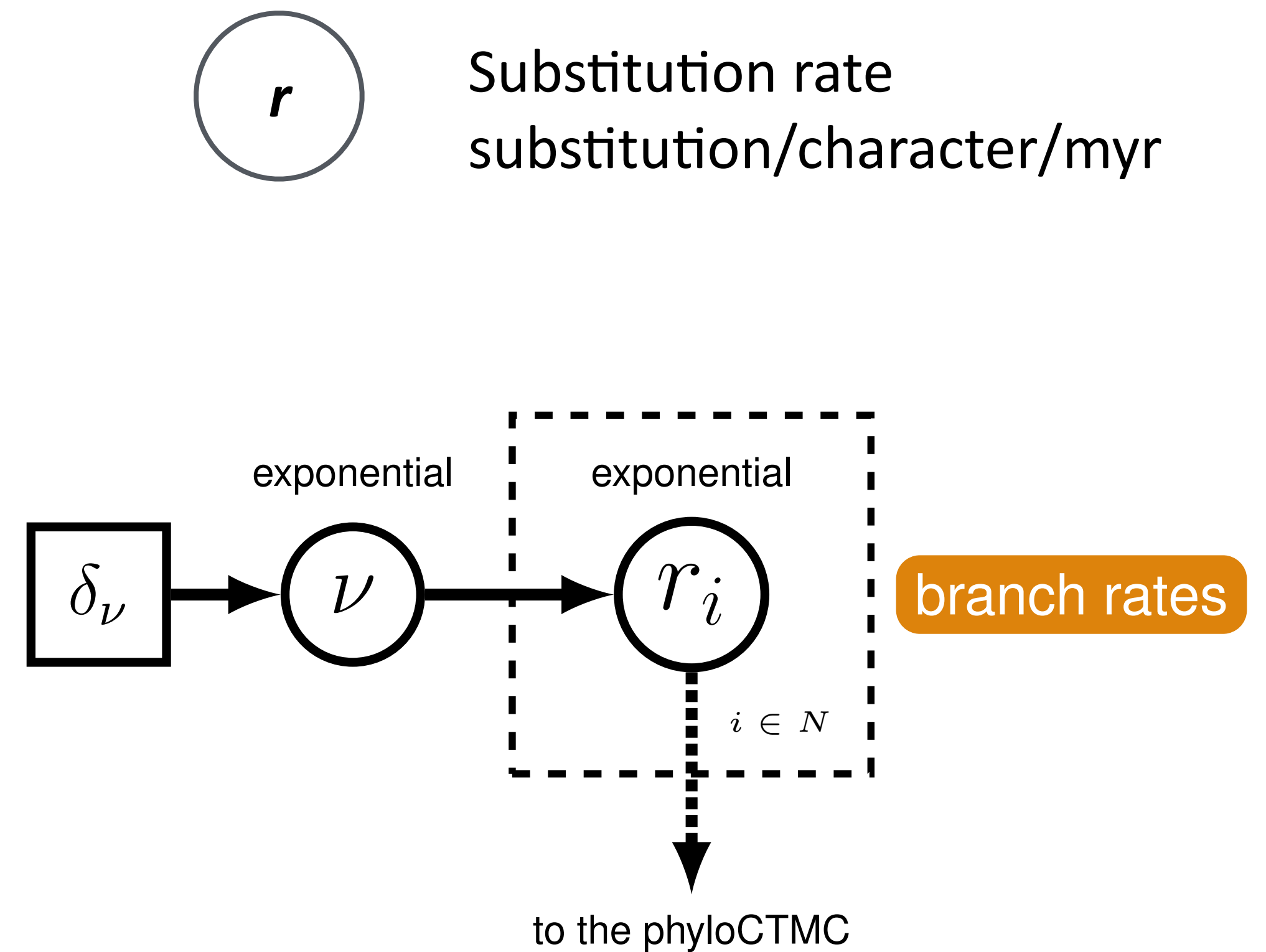
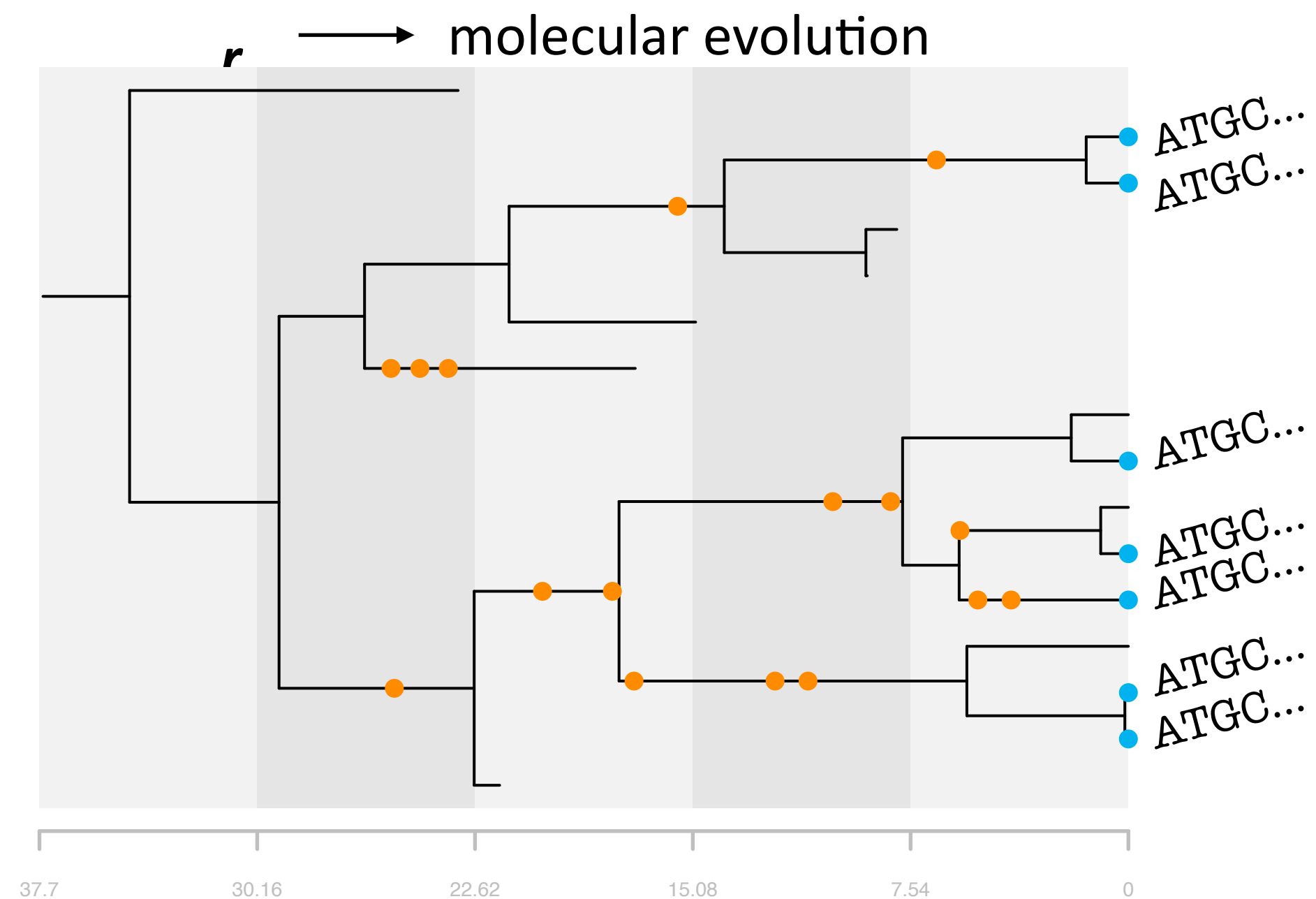
Branch rates



Substitution rate
substitution/character/myr

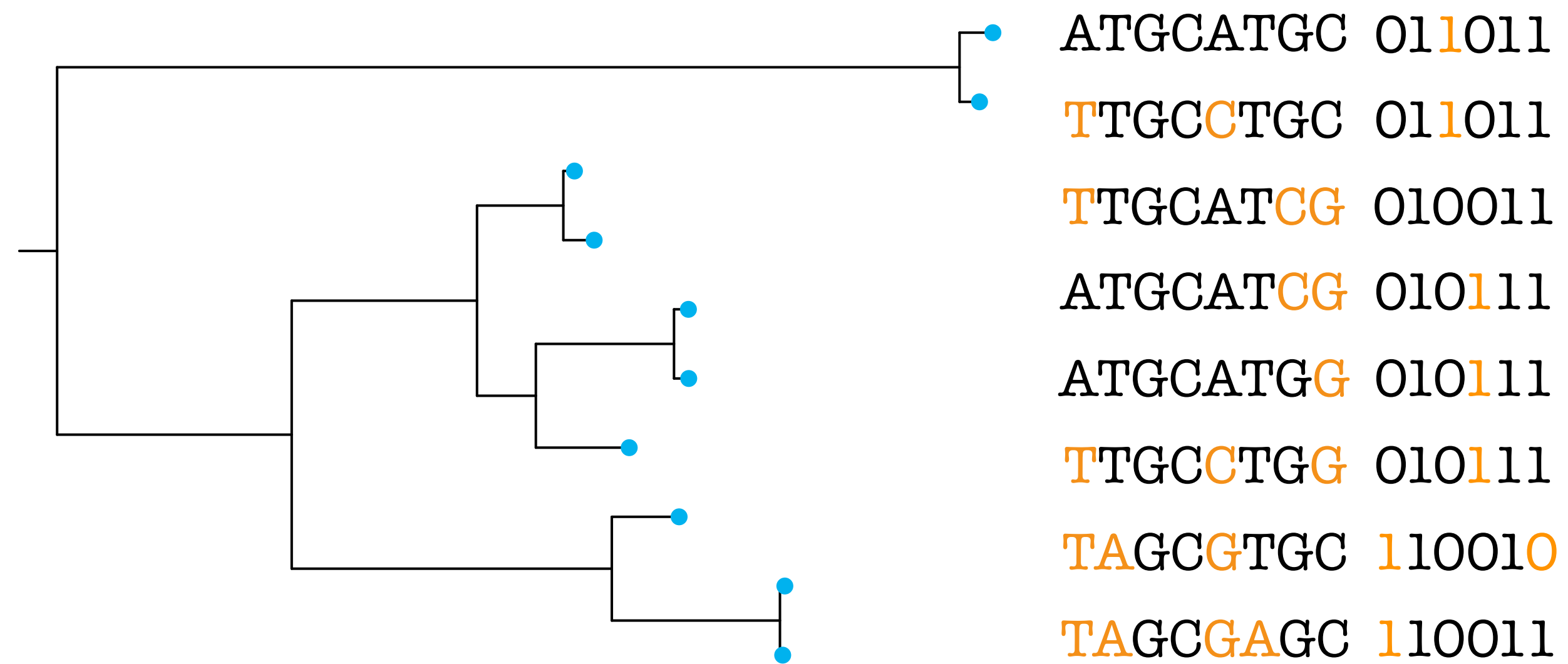
A model of character evolution across branches (clock model)

Branch rates



A relaxed clock model

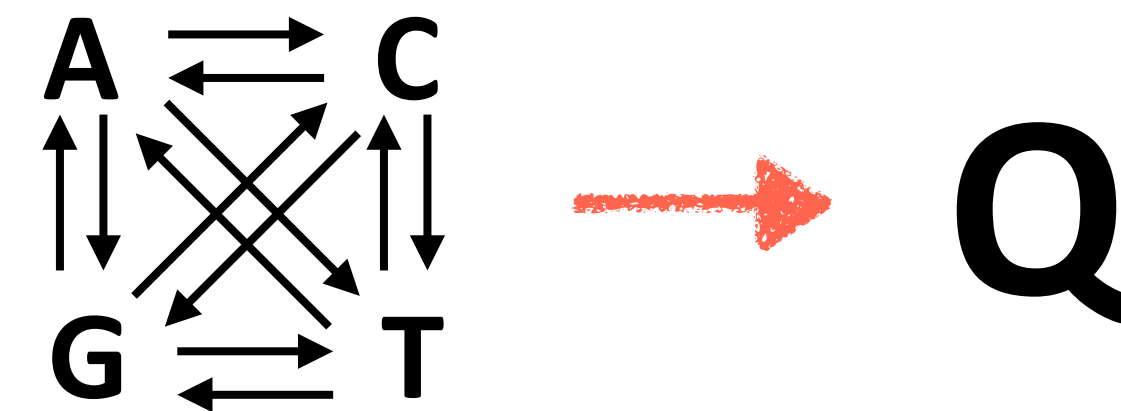
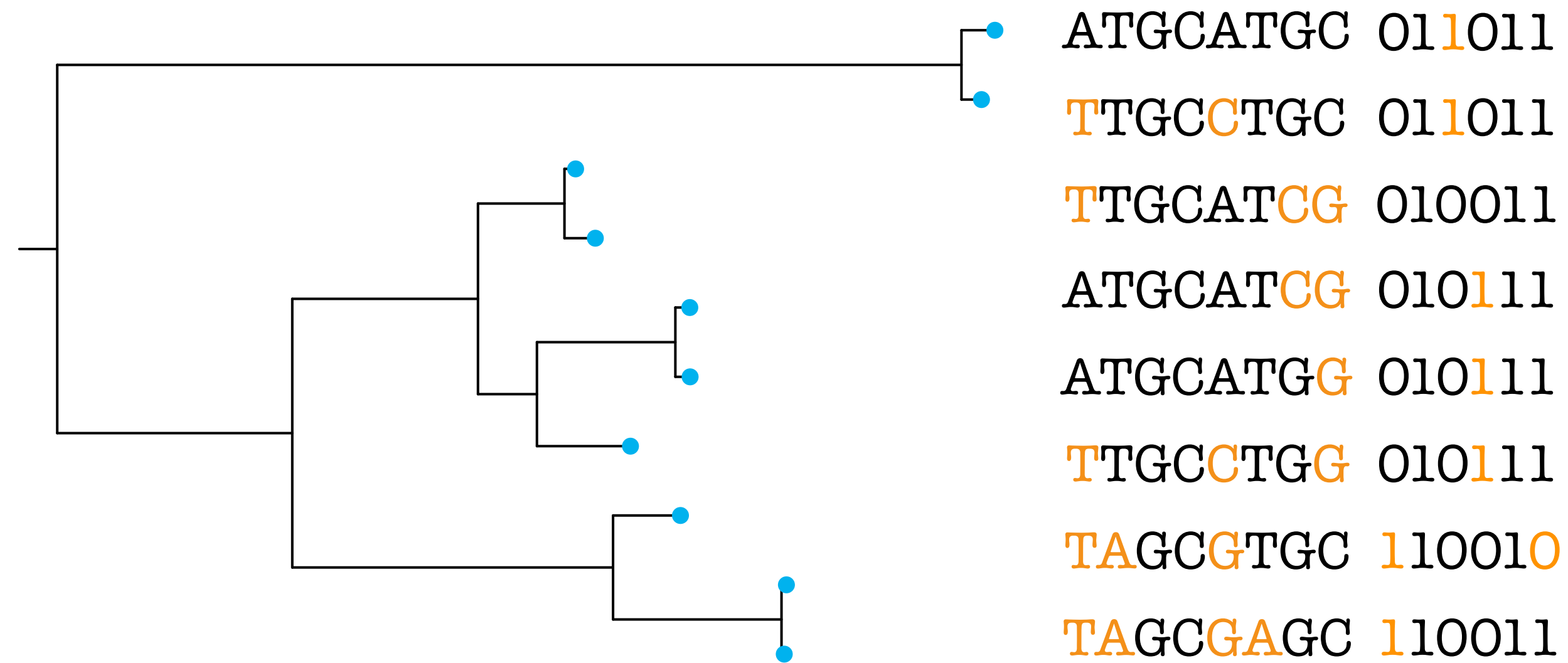
A simple model of character evolution



branch lengths = rate x time

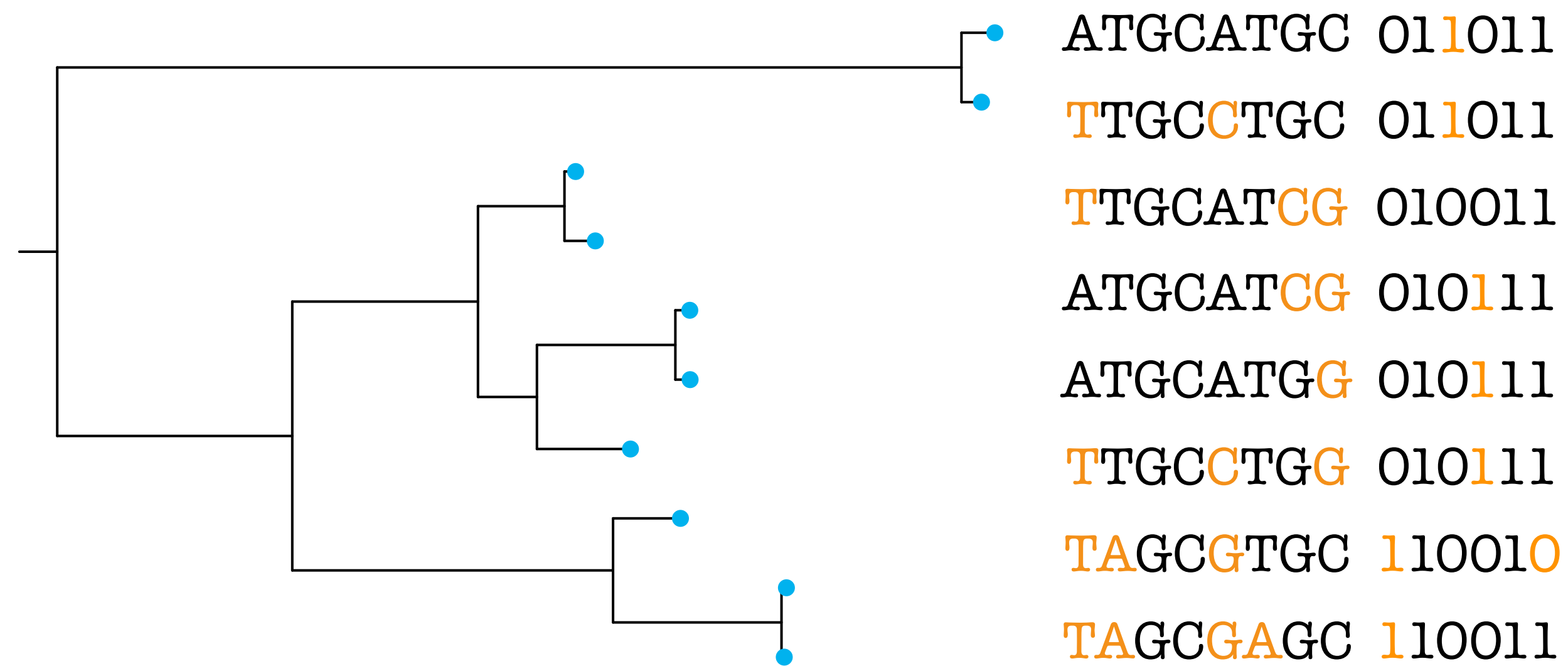
A simple model of character evolution

Sub. model



A simple model of character evolution

Sub. model



$0 \rightleftharpoons 1 \rightarrow Q$

$0 \rightleftharpoons 1$
 $0 \rightleftharpoons 2 \rightleftharpoons 1 \rightarrow Q$

A simple model of character evolution

```
> Q <- fnJC(2)
```

← Jukes cantor model with 2 characters

```
> Q  
[ [ -1.0000, 1.0000 ] ,  
  [ 1.0000, -1.0000 ] ]
```

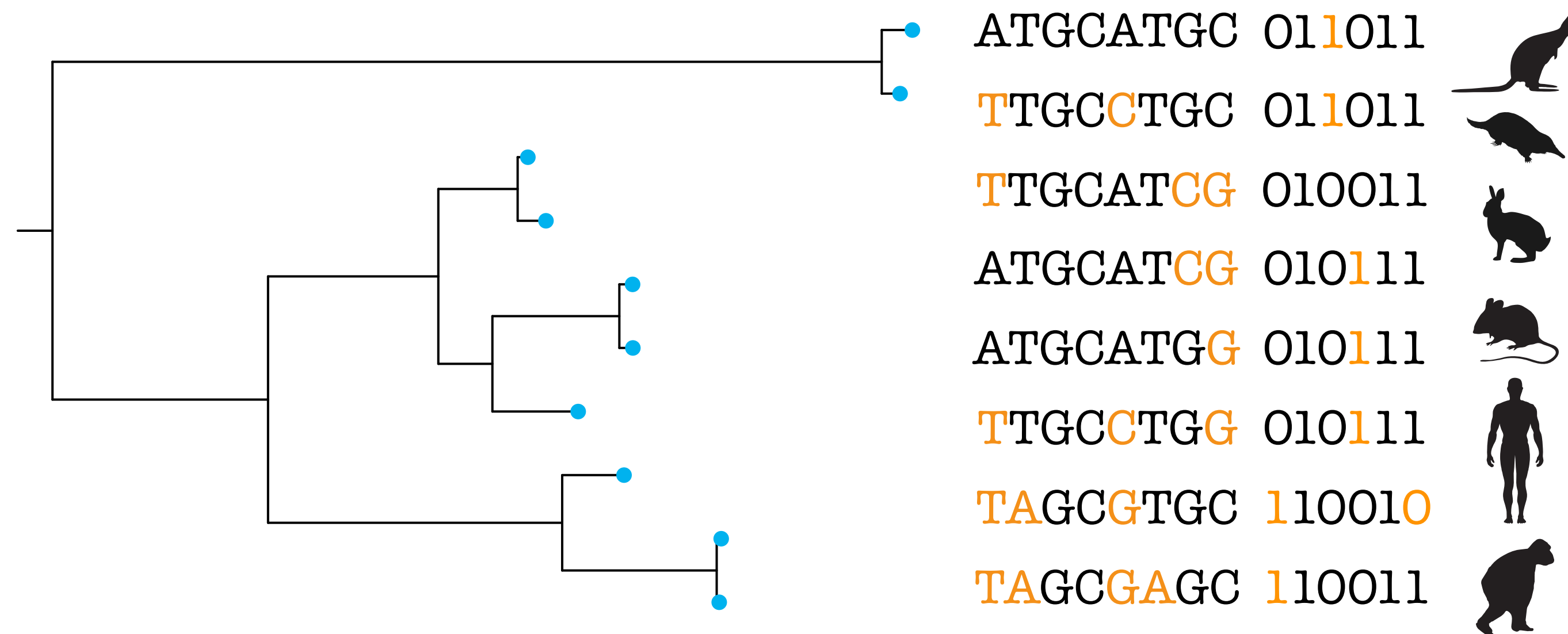
```
> Q.getTransitionProbabilities(0.1)  
[ [ 0.909, 0.091],  
  [ 0.091, 0.909] ]
```

← 0.1 myr

```
> Q.getTransitionProbabilities(100)  
[ [ 0.500, 0.500],  
  [ 0.500, 0.500] ]
```

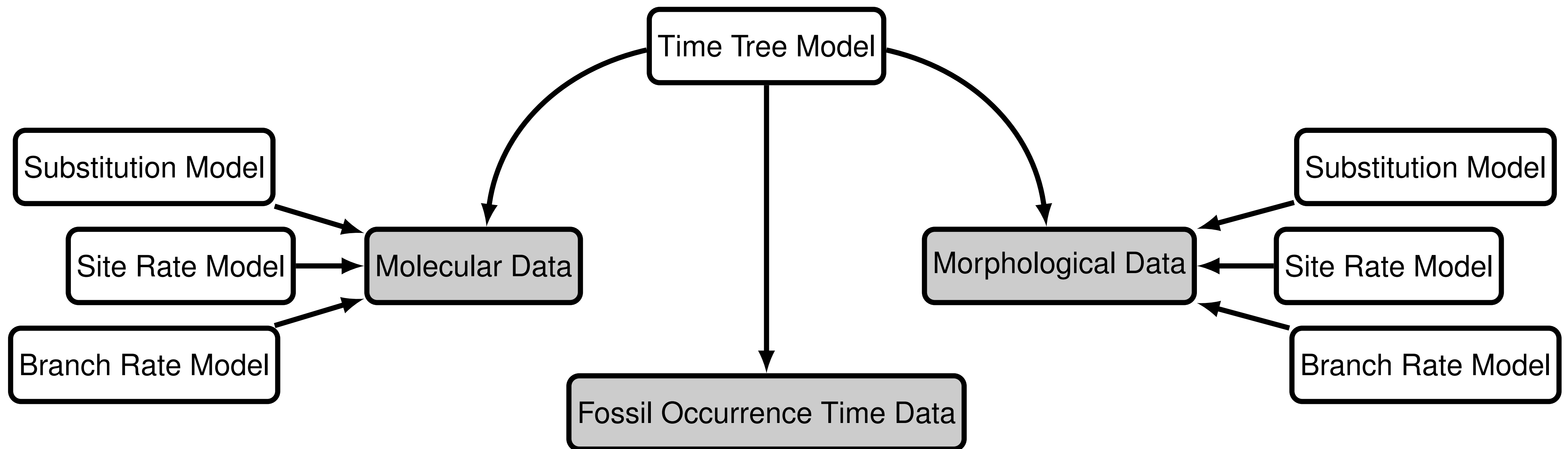
← 100 myr

Rate and time are only semi-identifiable



branch lengths = rate x time

Divergence time estimation: a tricky business



Modular components of the graphical model used in the FBD tip-dating analysis described in the tutorial.