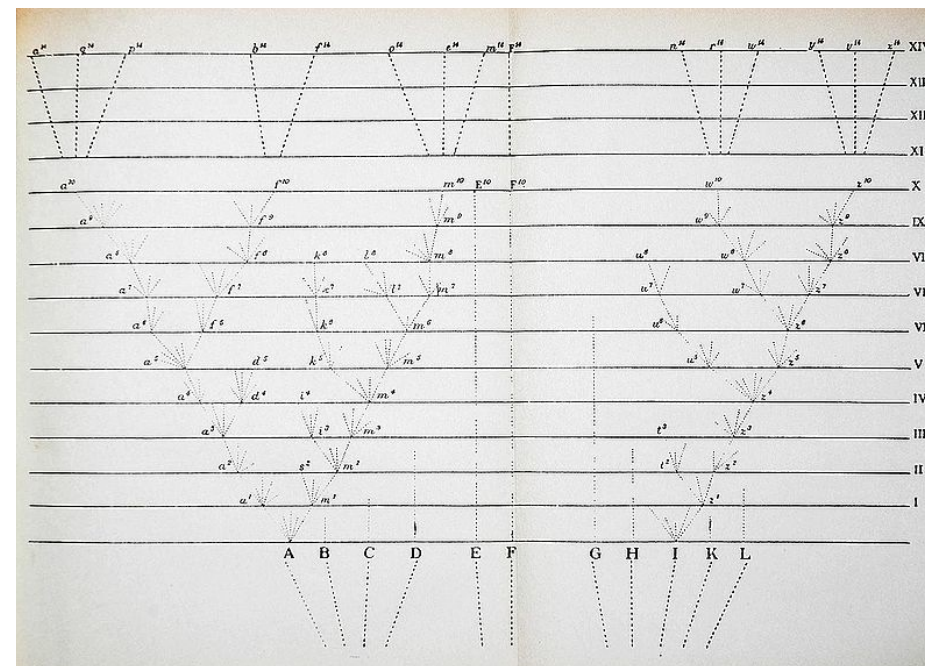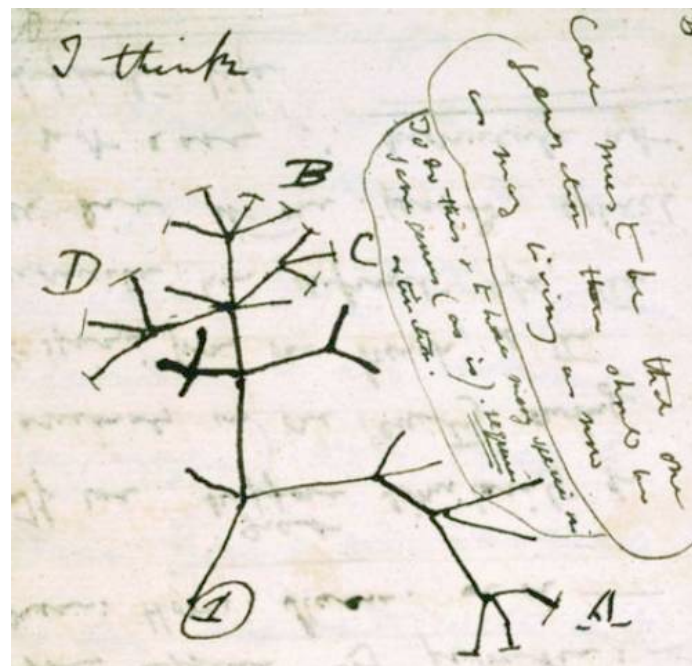# Phylodynamics in RevBayes
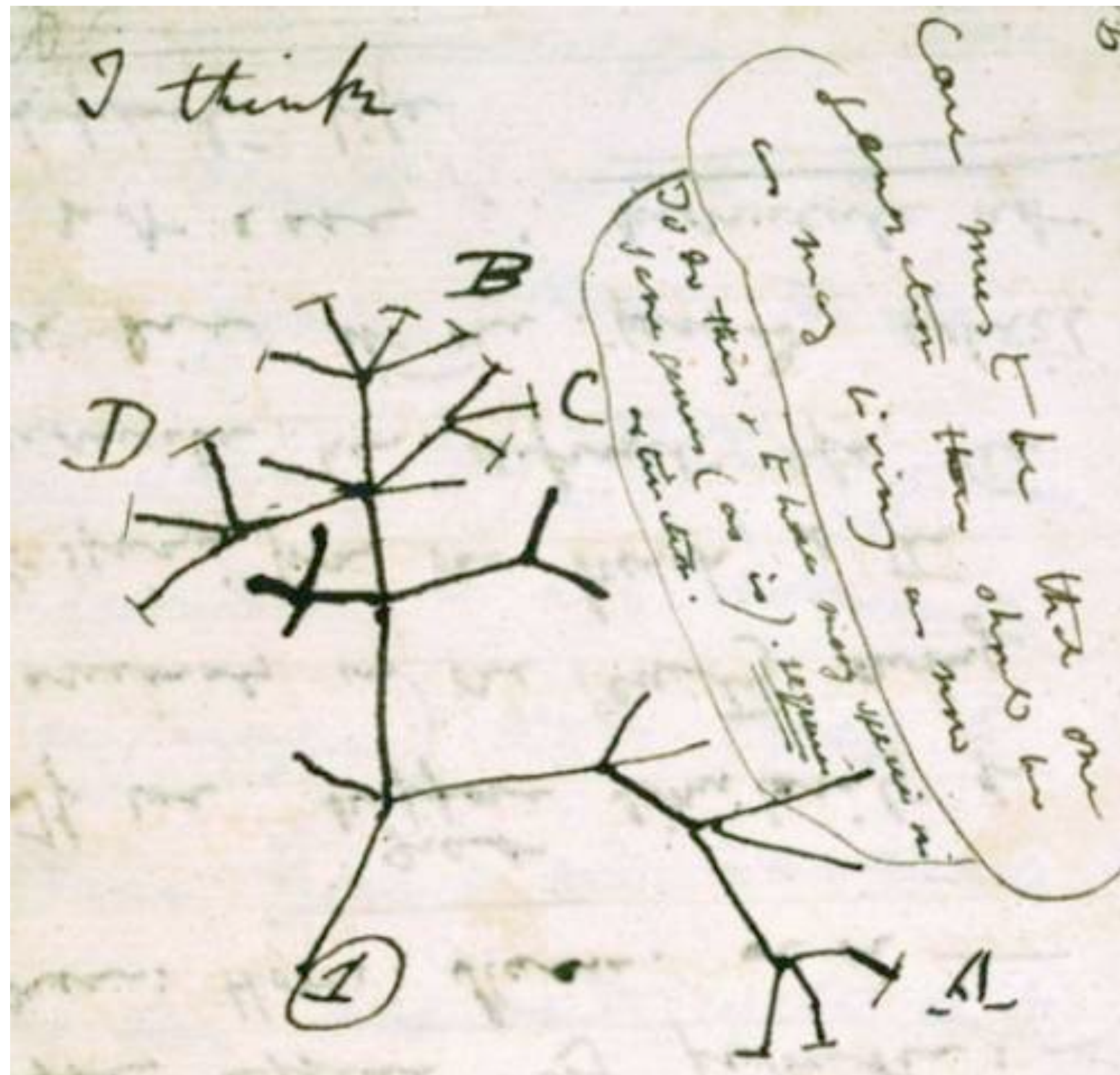
Full Bayesian inference
of time trees and population dynamic parameters

Tracy Heath & Tanja Stadler

Tanja Stadler
Computational Evolution group — Department of Biosystems Science and Engineering
tanja.stadler@bsse.ethz.ch

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich
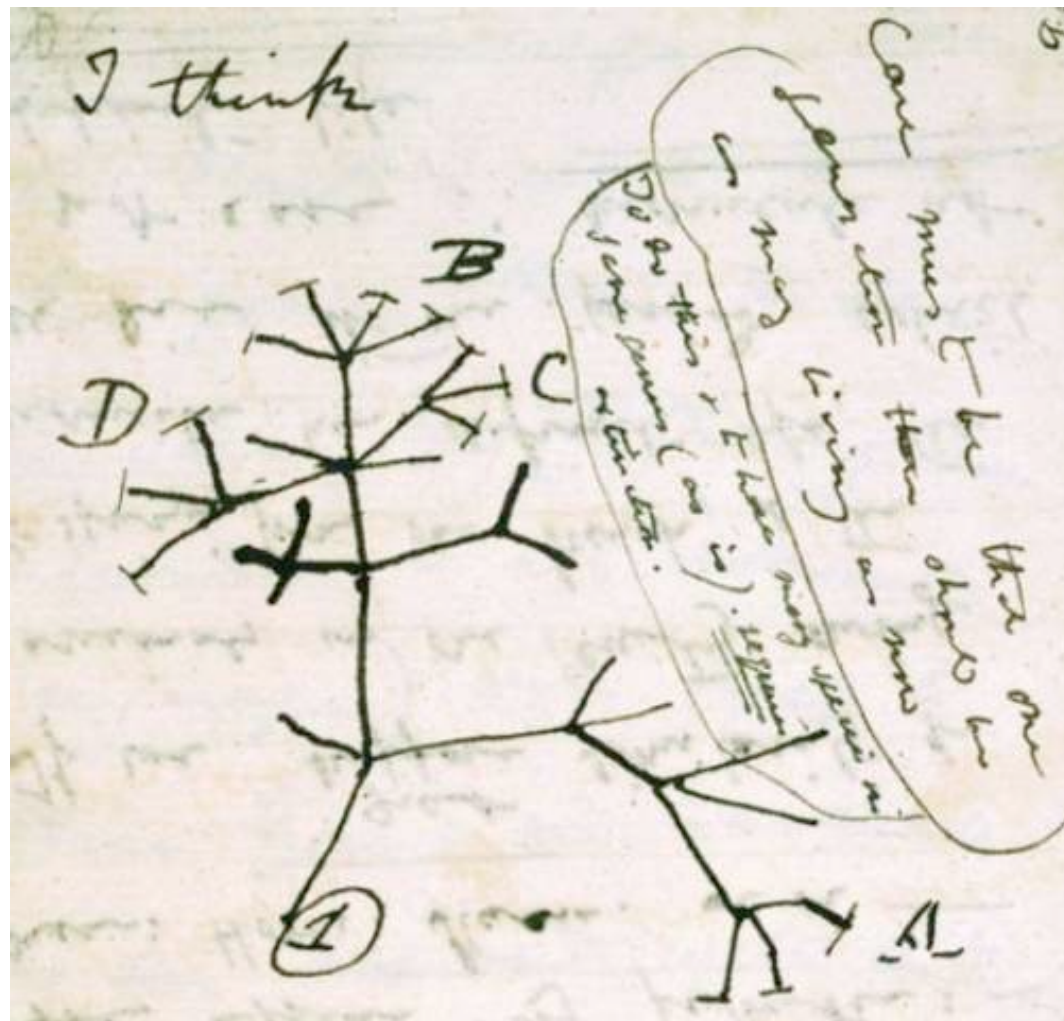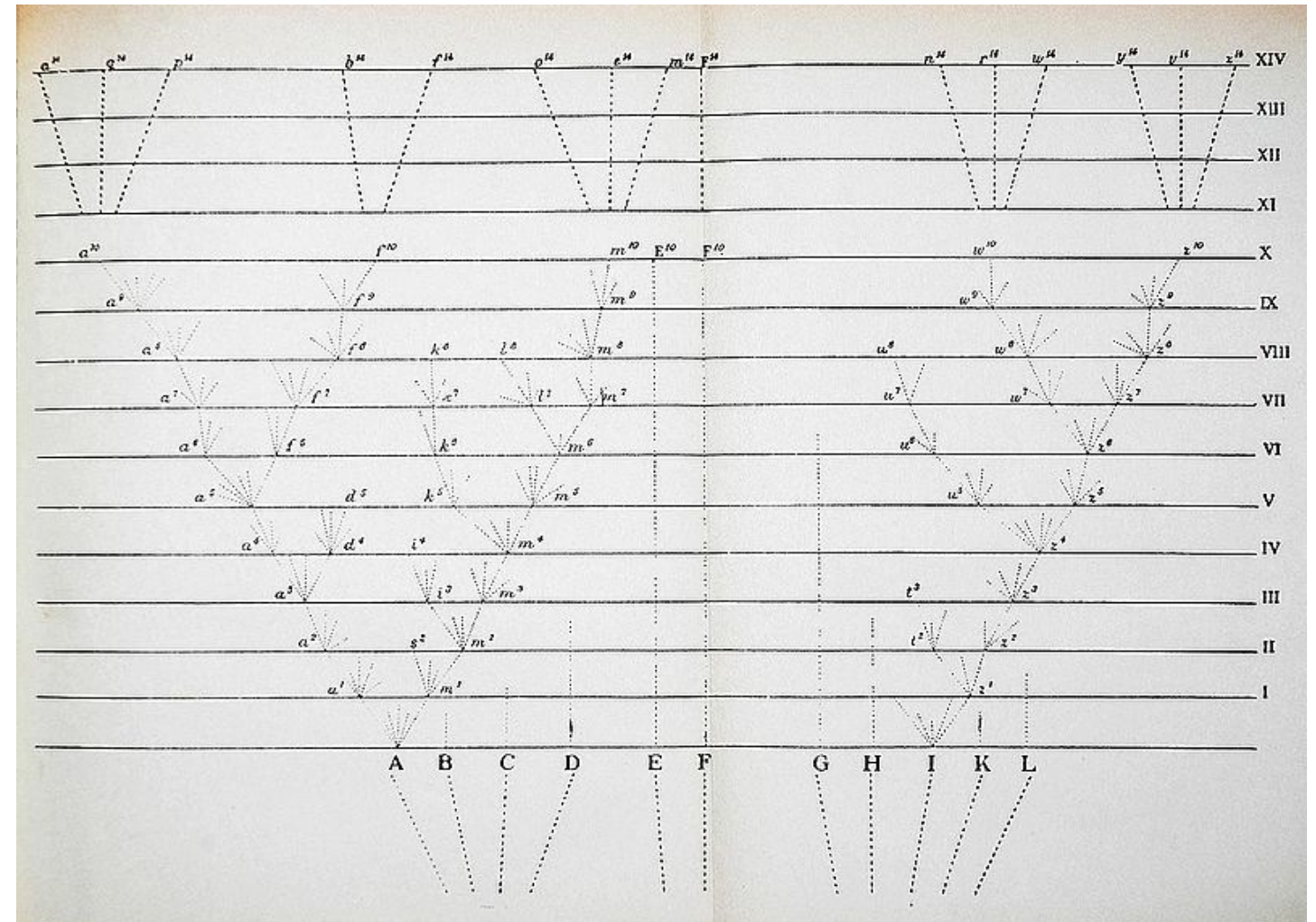
Darwin's first phylogenetic tree

Darwin's two types of trees

Notebook, 1837

Origin of Species, 1859

General modeling framework
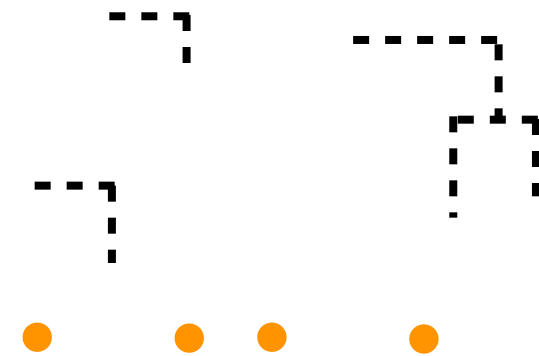
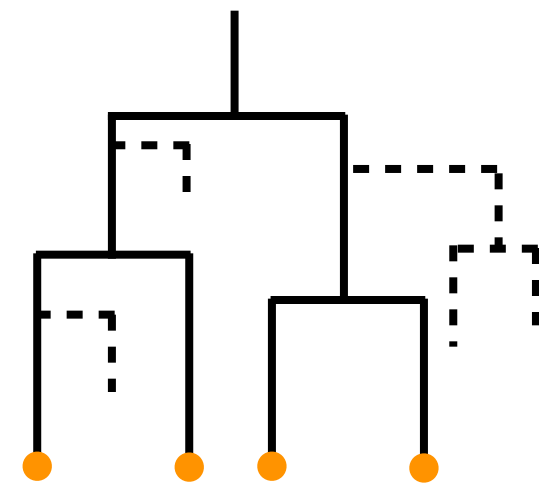| Population dynamic model	Growth of tree
	‣ parameters η

General modeling framework

Population dynamic model

Growth of tree
‣ parameters η

General modeling framework

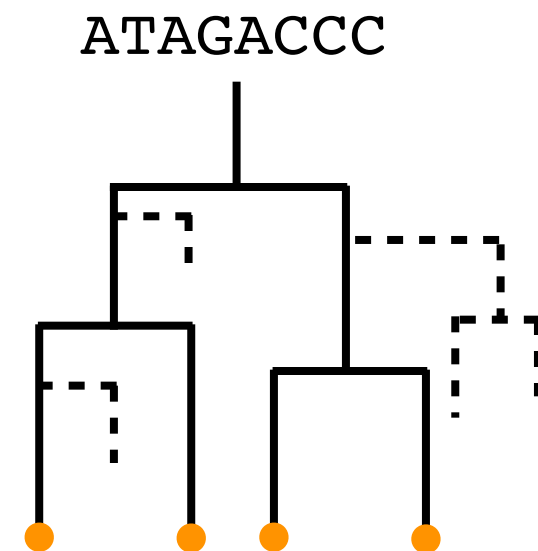|	
Population dynamic model	Growth of tree ‣ parameters η
Evolution-ary model	Evolution of sequences along tree ‣ parameters θ

ATAGACCC

General modeling framework
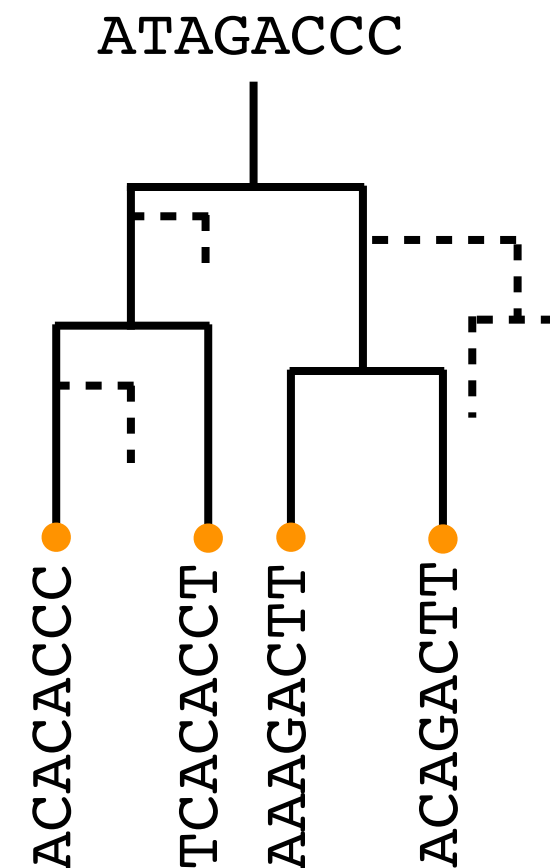
Population dynamic model

Growth of tree
- parameters η

Evolution-ary model

Evolution of sequences along tree
- parameters θ

Tree distri-bution

Defined through η and θ

ATAGACCC

ACACACCC   TCACACCT   AAAGACTT   ACAGACTT

General modeling framework

Bayesian approach for estimating trees and parameters

time tree   population dynamics
            evololutionary dynamics
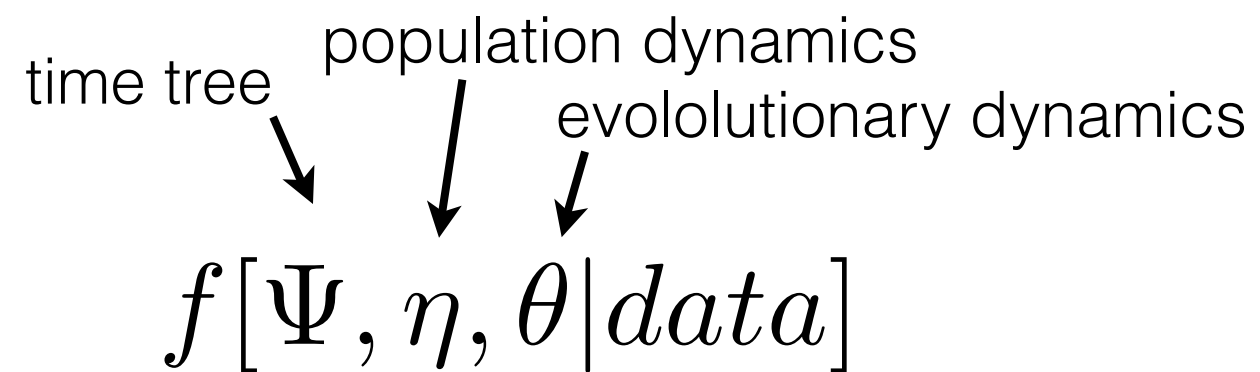
$$f\left[\Psi, \eta, \theta \,|\, data\right]$$

Bayesian approach for estimating trees and parameters

time tree population dynamics

evololutionary dynamics

$$f[\Psi, \eta, \theta | data] = \frac{f[data | \Psi, \theta] f[\Psi | \eta] f[\eta, \theta]}{f[data]}$$

Bayesian approach for estimating trees and parameters

time tree   population dynamics

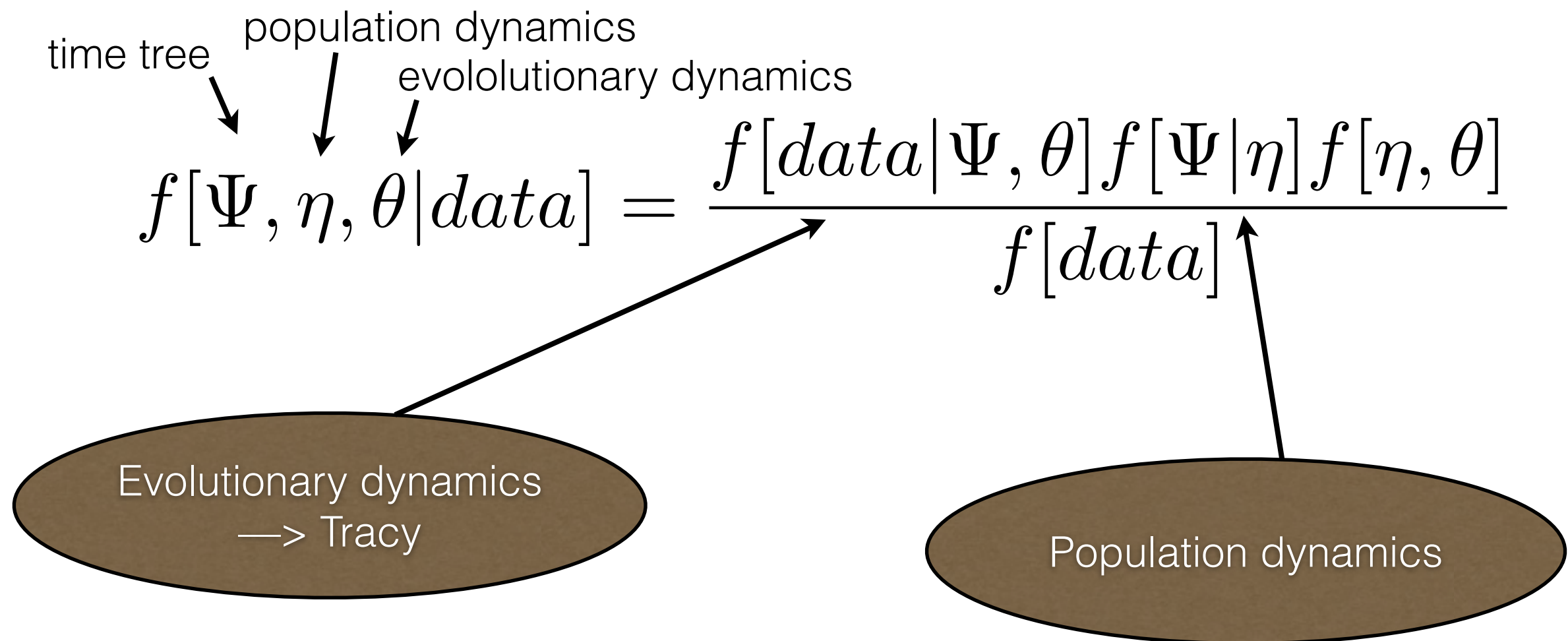evololutionary dynamics

$$f[\Psi, \eta, \theta | data] = \frac{f[data | \Psi, \theta] f[\Psi | \eta] f[\eta, \theta]}{f[data]}$$

Evolutionary dynamics
—> Tracy

Bayesian approach for estimating trees and parameters

time tree   population dynamics

evololutionary dynamics

$$f[\Psi, \eta, \theta | data] = \frac{f[data | \Psi, \theta] f[\Psi | \eta] f[\eta, \theta]}{f[data]}$$

Evolutionary dynamics
—> Tracy

Population dynamics

Bayesian approach for estimating trees and parameters
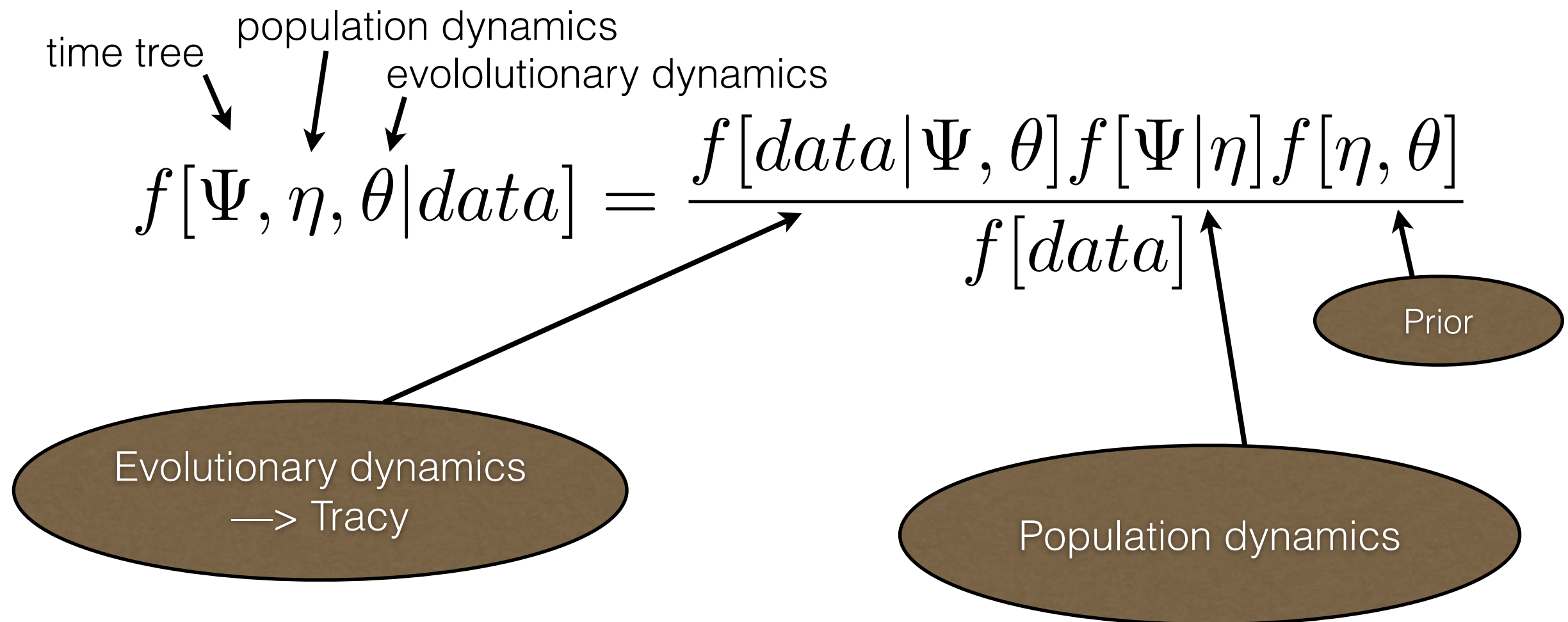
time tree population dynamics

evololutionary dynamics

$$f[\Psi, \eta, \theta | data] = \frac{f[data | \Psi, \theta]\, f[\Psi | \eta]\, f[\eta, \theta]}{f[data]}$$

Prior

Evolutionary dynamics
—> Tracy

Population dynamics

Bayesian approach for estimating trees and parameters

time tree population dynamics

evololutionary dynamics

$$f[\Psi, \eta, \theta | data] = \frac{f[data | \Psi, \theta] \, f[\Psi | \eta] \, f[\eta, \theta]}{f[data]}$$
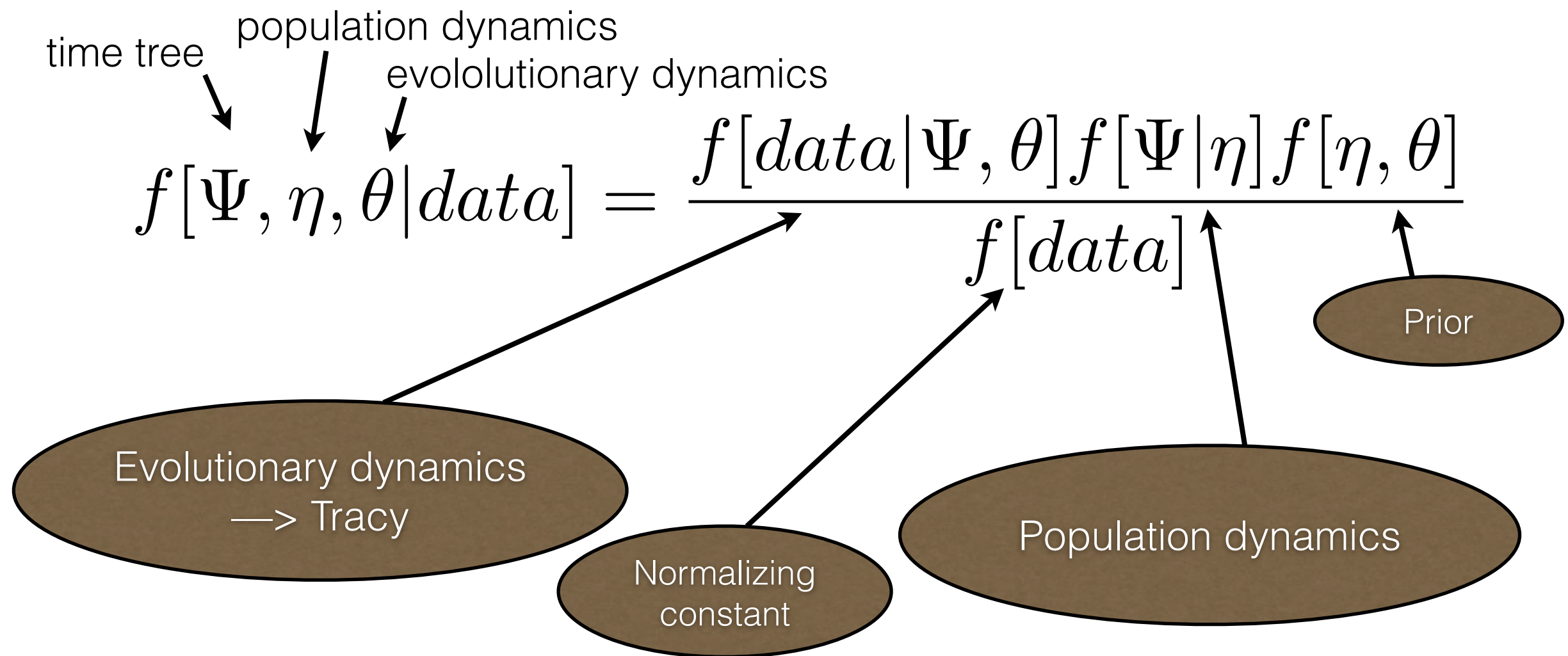
Prior

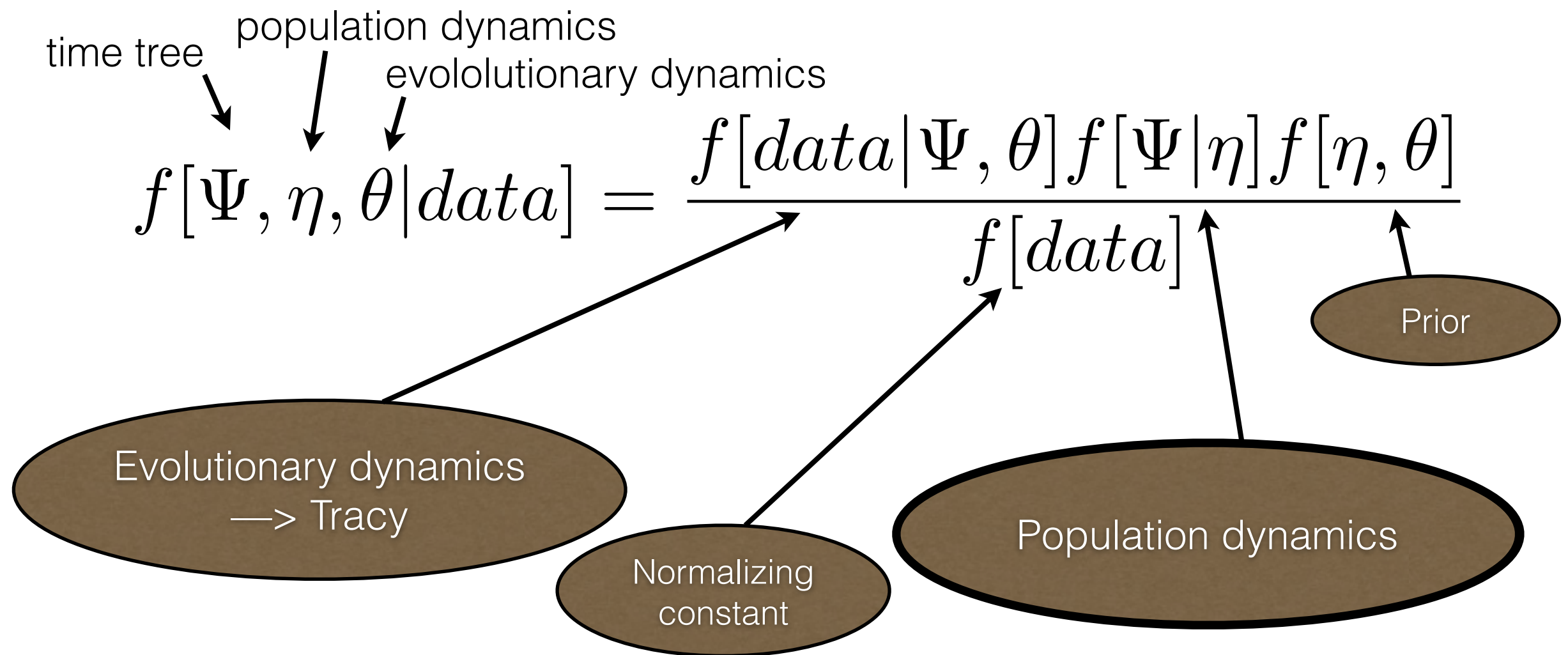Evolutionary dynamics
—> Tracy

Normalizing constant

Population dynamics

Bayesian approach for estimating trees and parameters

Bayesian approach for estimating trees and parameters

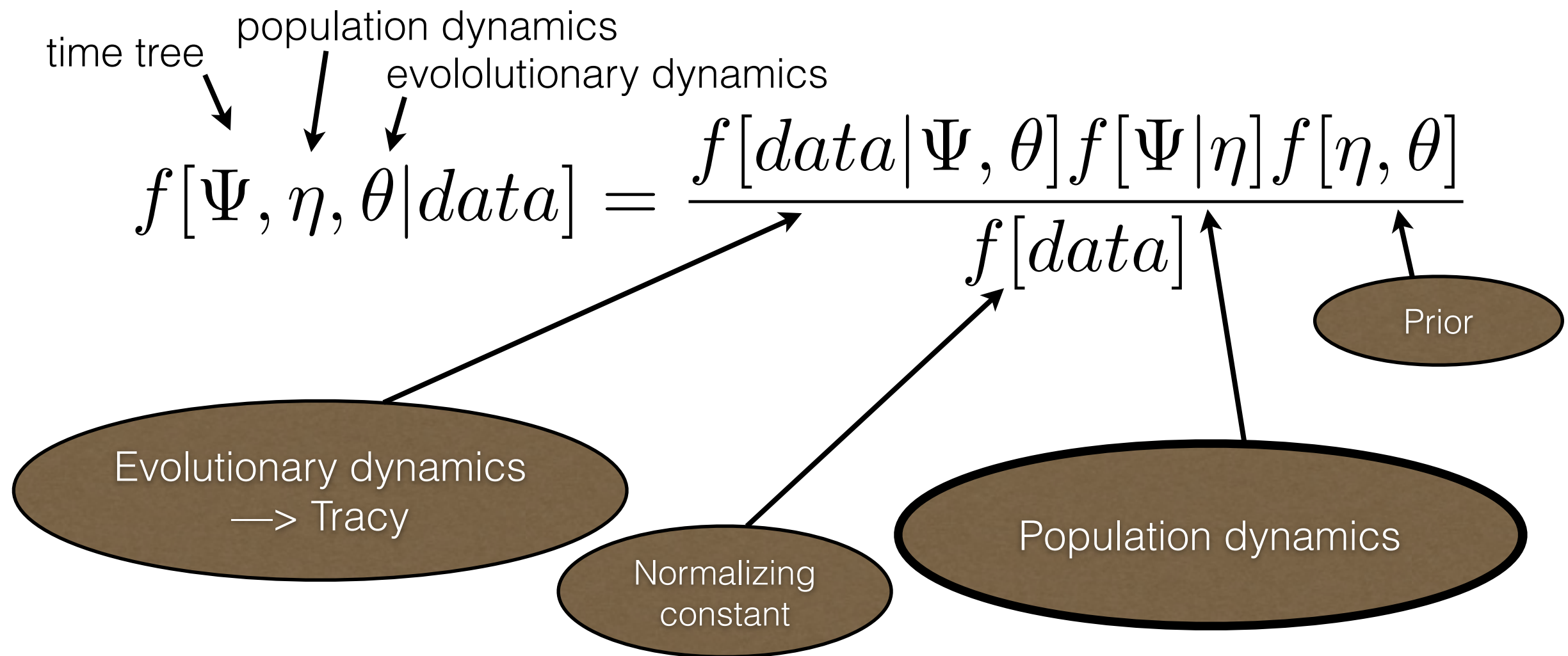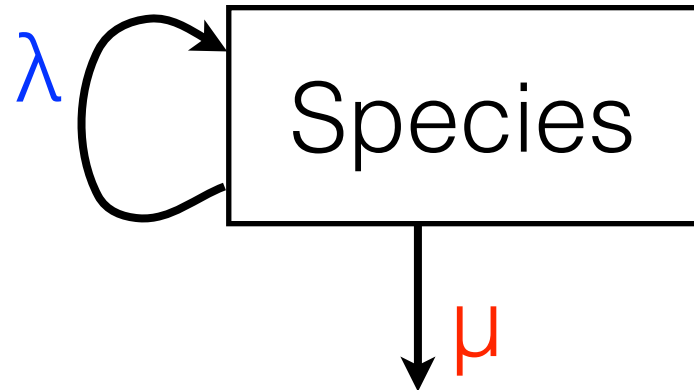RevBayes samples the posterior distribution of trees and parameters

The birth-death model as a model for speciation and extinction
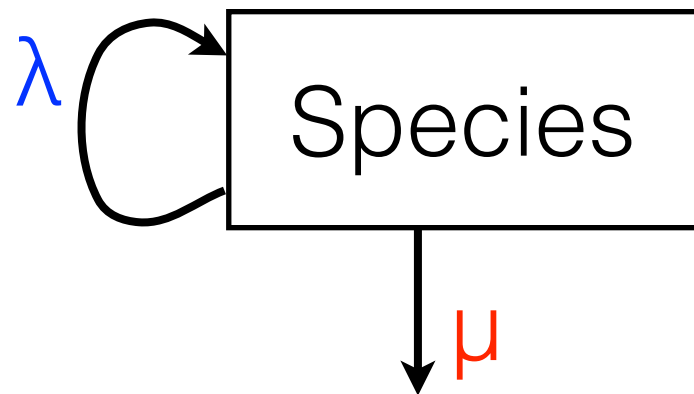
Birth-death model

- stem age of a clade T
- speciation rate λ
- extinction rate μ
- sampling probability ρ

Rates may depend on:
    1) constant
    2) time (environmental-dependence)
    3) # of species (diversity-dependence)
    4) type of species

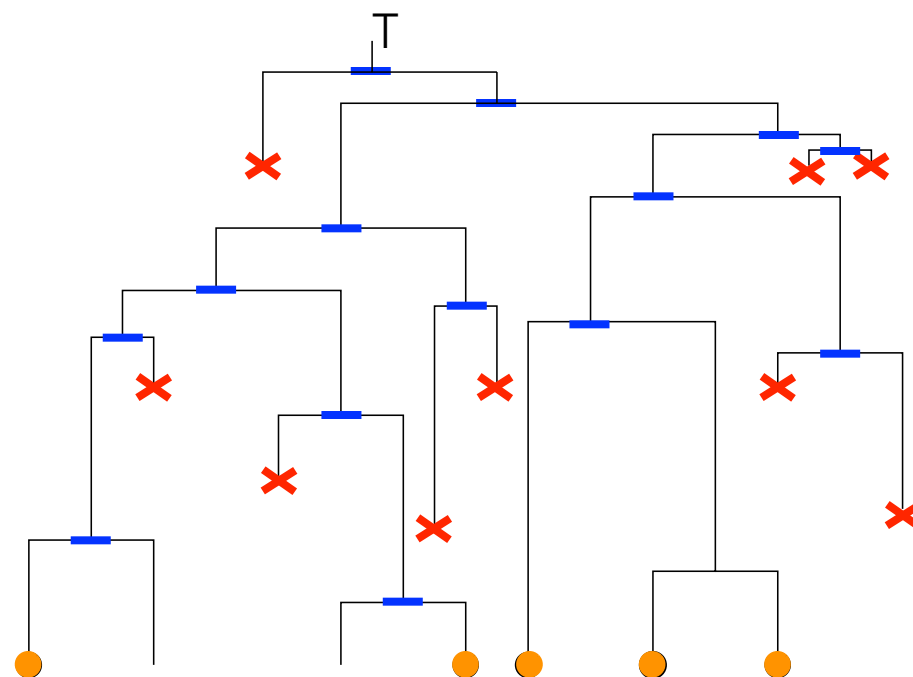# The birth-death model as a model for speciation and extinction

Birth-death model

- stem age of a clade T
- speciation rate λ
- extinction rate μ
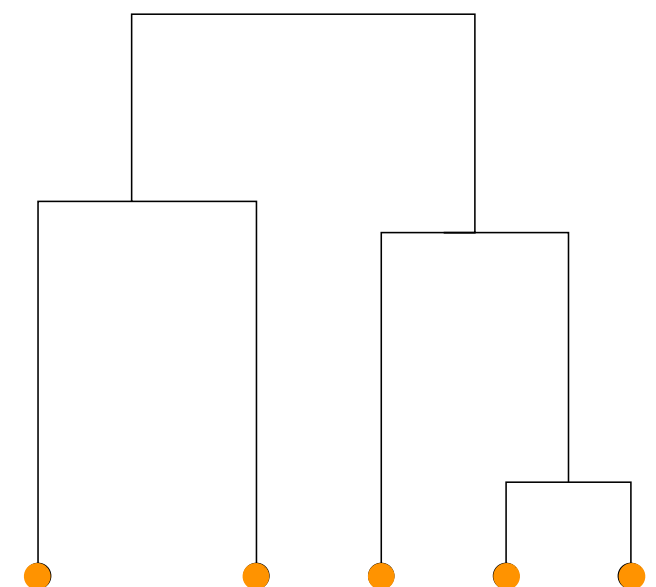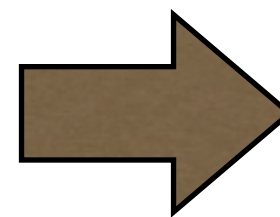- sampling probability ρ

Rates may depend on:
1) constant
2) time (environmental-dependence)
3) # of species (diversity-dependence)
4) type of species

Phylo-genetic trees

complete phylogeny

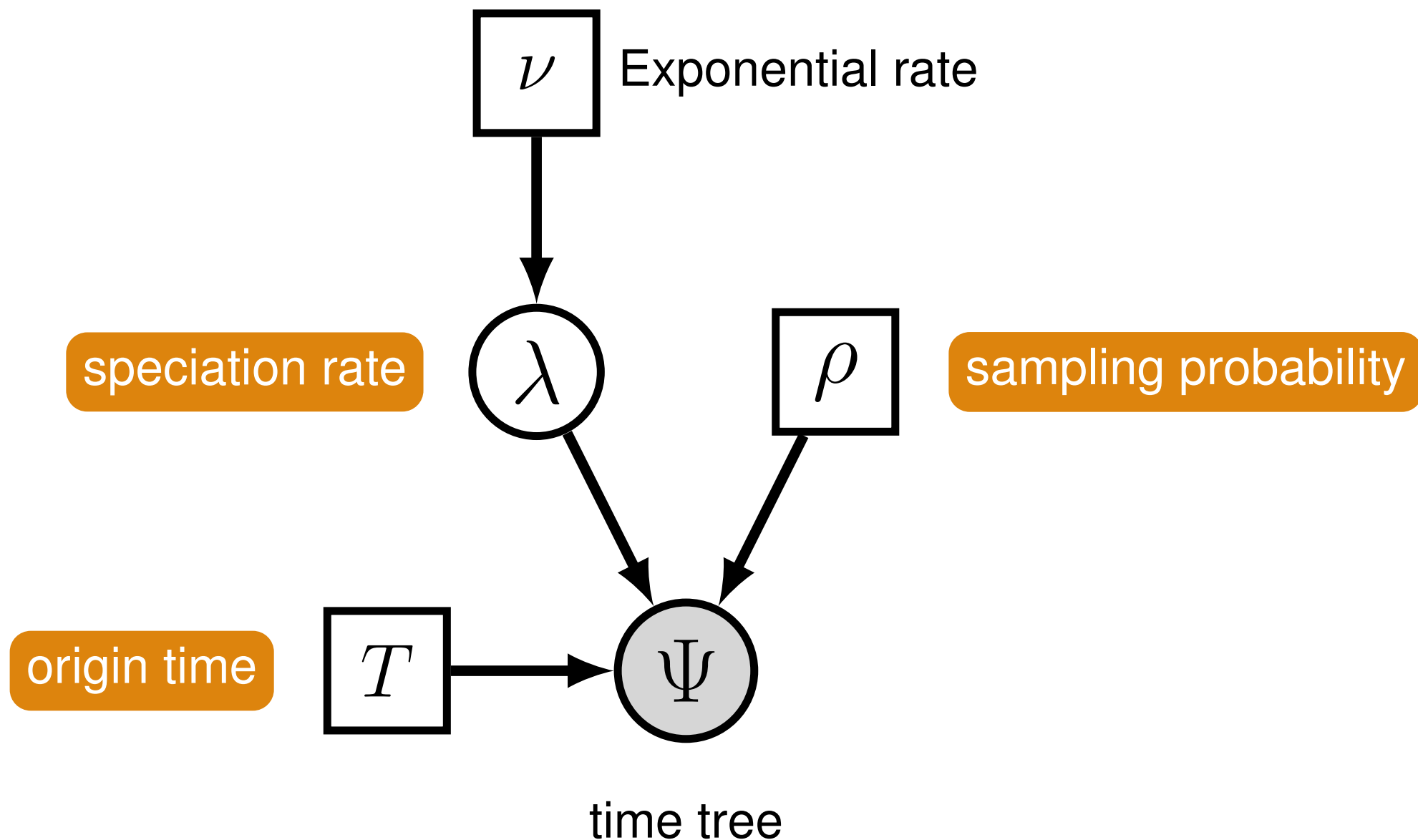time tree

Yule model in RevBayes

Yule model (simplest tree prior):
tree age T, constant birth rate λ, no extinction (μ=0), sampling probability ρ

time tree

Calculating the probability of a Yule tree with complete sampling $f[\Psi | \lambda, \rho = 1, T]$

Model

Age of a clade: T
Speciation rate: λ
Sampling probability: ρ

Constant rates means exponentially distributed waiting time

Constant rates means exponentially distributed waiting time

constant speciation rate λ means:

Constant rates means exponentially distributed waiting time

constant speciation rate λ means:
  Prob(speciation event during Δt) = λΔt

Constant rates means exponentially distributed waiting time

constant speciation rate λ means:
$$Prob(\text{speciation event during } \Delta t) = \lambda \Delta t$$

Constant rates means exponentially distributed waiting time

constant speciation rate λ means:
    Prob(speciation event during Δt) = λΔt

What is probability that a lineage did not speciate for time u?

Constant rates means exponentially distributed waiting time

constant speciation rate λ means:
    Prob(speciation event during Δt) = λΔt

What is probability that a lineage did not speciate for time u?
    Let u = kΔt.

Constant rates means exponentially distributed waiting time

constant speciation rate λ means:
    Prob(speciation event during Δt) = λΔt

What is probability that a lineage did not speciate for time u?
    Let u = kΔt.
    Prob(no speciation event during u/k) = 1-λu/k

Constant rates means exponentially distributed waiting time

constant speciation rate λ means:
    Prob(speciation event during Δt) = λΔt

What is probability that a lineage did not speciate for time u?
    Let u = kΔt.
    Prob(no speciation event during u/k) = 1-λu/k
    Prob(no speciation event during u) = $(1-λu/k)^k$

Constant rates means exponentially distributed waiting time

constant speciation rate λ means:
Prob(speciation event during Δt) = λΔt

What is probability that a lineage did not speciate for time u?
Let u = kΔt.
Prob(no speciation event during u/k) = 1-λu/k
Prob(no speciation event during u) = $(1-λu/k)^k$

Constant rates means exponentially distributed waiting time

constant speciation rate λ means:

  Prob(speciation event during Δt) = λΔt

What is probability that a lineage did not speciate for time u?

  Let u = kΔt.

  Prob(no speciation event during u/k) = 1-λu/k

  Prob(no speciation event during u) = $(1-\lambda u/k)^k$

  $(1-\lambda u/k)^k = e^{-\lambda u}$ for k→∞

Constant rates means exponentially distributed waiting time

constant speciation rate λ means:
Prob(speciation event during Δt) = λΔt

What is probability that a lineage did not speciate for time u?
Let u = kΔt.
Prob(no speciation event during u/k) = 1-λu/k
Prob(no speciation event during u) = $(1-\lambda u/k)^k$

$(1-\lambda u/k)^k = e^{-\lambda u}$ for k➙∞

Constant rates means exponentially distributed waiting time

constant speciation rate λ means:
    Prob(speciation event during Δt) = $\lambda\Delta$t

What is probability that a lineage did not speciate for time u?
    Let u = kΔt.
    Prob(no speciation event during u/k) = $1-\lambda u/k$
    Prob(no speciation event during u) = $(1-\lambda u/k)^k$

        $(1-\lambda u/k)^k = e^{-\lambda u}$ for k$\rightarrow\infty$

…which is the exponential distribution, Exp(λ).

Constant rates means exponentially distributed waiting time

constant speciation rate λ means:
  Prob(speciation event during Δt) = $\lambda\Delta t$

What is probability that a lineage did not speciate for time u?
  Let u = $k\Delta t$.
  Prob(no speciation event during u/k) = $1-\lambda u/k$
  Prob(no speciation event during u) = $(1-\lambda u/k)^k$

   $(1-\lambda u/k)^k = e^{-\lambda u}$ for k→∞

…which is the exponential distribution, Exp(λ).

Constant rates means exponentially distributed waiting time

What is probability that a **n** lineage did not speciate for time u?

Let $u = k\Delta t$.

Prob(no speciation event during u/k) $= (1-\lambda u/k)^{\mathbf{n}}$

Prob(no speciation event during u) $= (1-\lambda u/k)^{\mathbf{n}k}$

$$(1-\lambda u/k)^{\mathbf{n}k} = e^{-\mathbf{n}\lambda u} \text{ for } k \to \infty$$

…which is the exponential distribution, Exp(**n**λ).

Prob(no speciation event during u) $= e^{-n\lambda u}$

Prob(speciation before u) $= 1-e^{-n\lambda u}$

Prob(speciation at u) $= \frac{d}{du}(1-e^{-n\lambda u}) = n\lambda e^{-n\lambda u}$
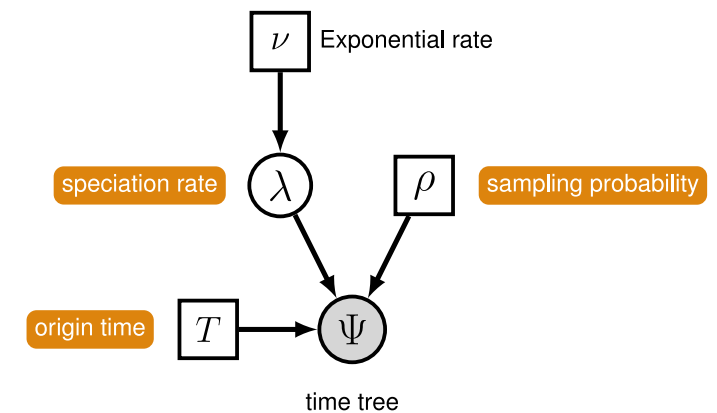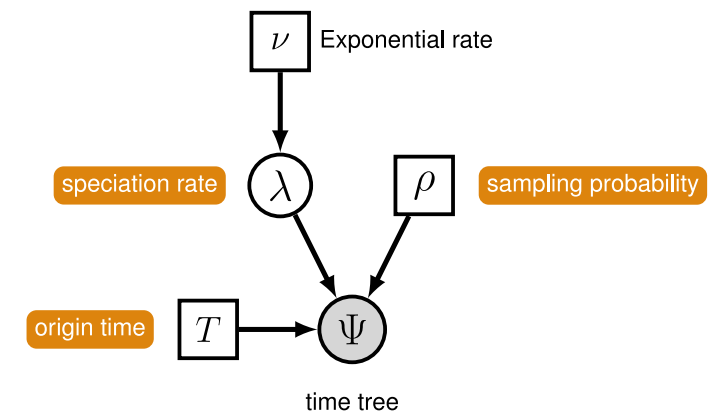
Calculating the probability of a Yule tree with complete sampling $f[\Psi|\lambda, \rho = 1, T]$

Model

Age of a clade: T
Speciation rate: λ
Sampling probability: ρ
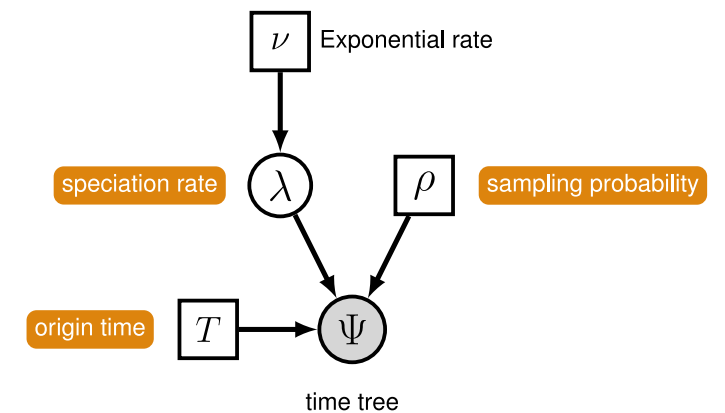
Calculating the probability of a Yule tree with complete sampling $f[\Psi|\lambda, \rho = 1, T]$

Model

Age of a clade: T
Speciation rate: λ
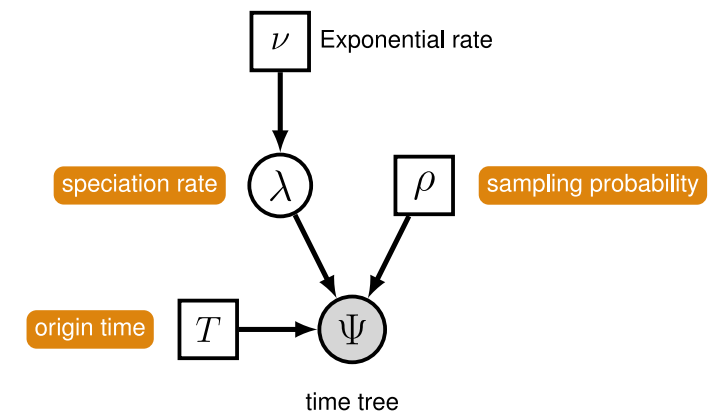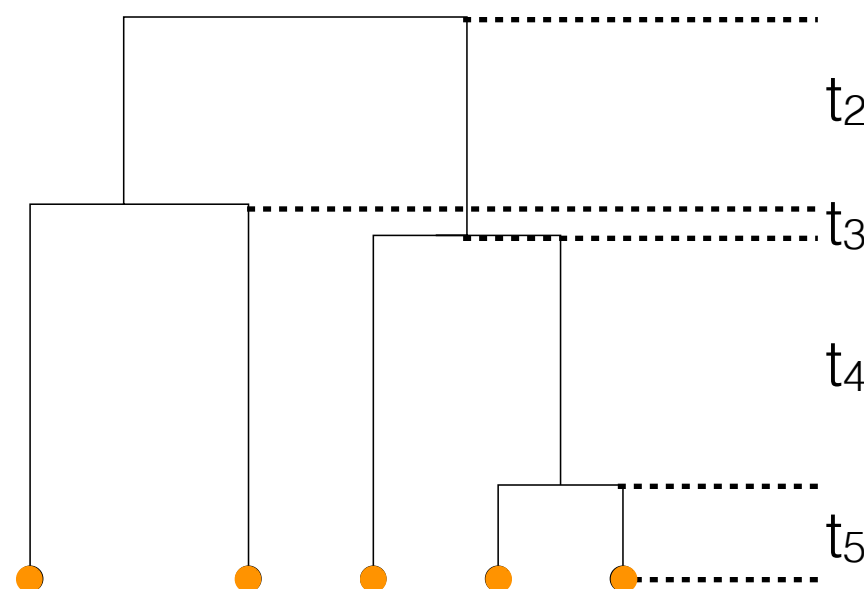Sampling probability: ρ

Probability of tree

Waiting time until a single lineage splits:

Waiting time until n lineages split:

Calculating the probability of a Yule tree with complete sampling $f[\Psi|\lambda, \rho = 1, T]$

Model	Age of a clade: T Speciation rate: λ Sampling probability: ρ

Probability of tree	Waiting time until a single lineage splits:	Exp(λ)
	Waiting time until n lineages split:	Exp(nλ)

Calculating the probability of a Yule tree with complete sampling $f[\Psi|\lambda, \rho = 1, T]$

Model
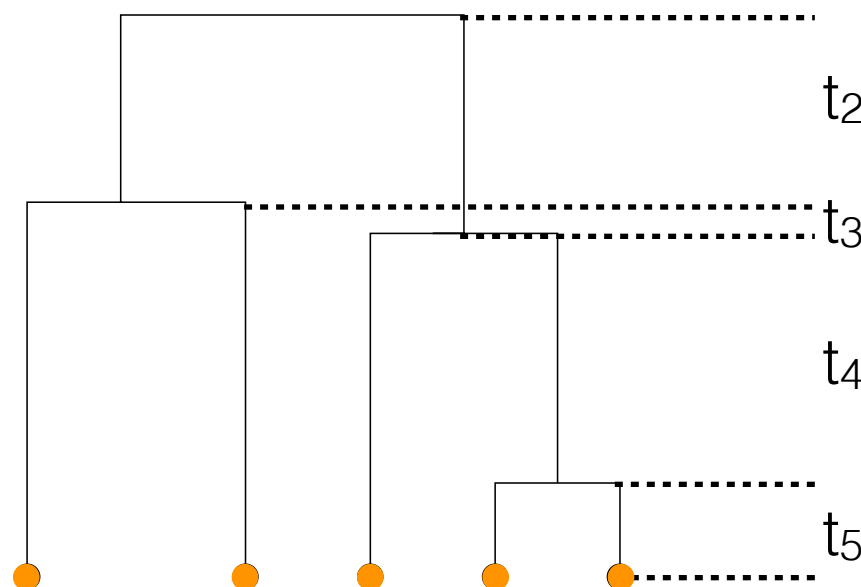
Age of a clade: T
Speciation rate: λ
Sampling probability: ρ

Probability of tree

Waiting time until a single lineage splits: Exp(λ)

Waiting time until n lineages split: Exp(nλ)

Example

$f(\Psi|\lambda, \rho = 1, T = t_2 + t_3 + t_4 + t_5) =$

Calculating the probability of a Yule tree with complete sampling $f[\Psi|\lambda, \rho = 1, T]$

Model

Age of a clade: T
Speciation rate: λ
Sampling probability: ρ

Probability of tree

Waiting time until a single lineage splits: Exp(λ)

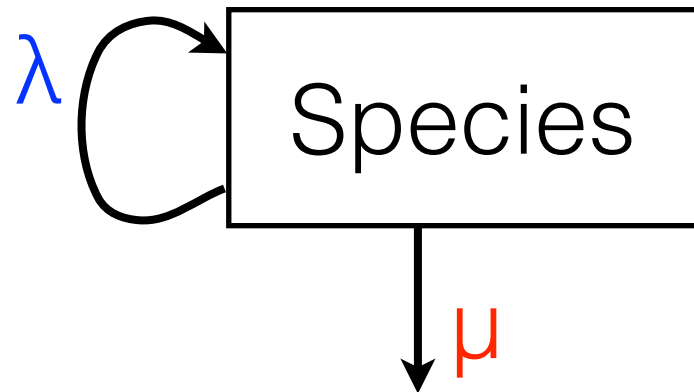Waiting time until n lineages split: Exp(nλ)

Example

t_2

t_3

t_4

t_5
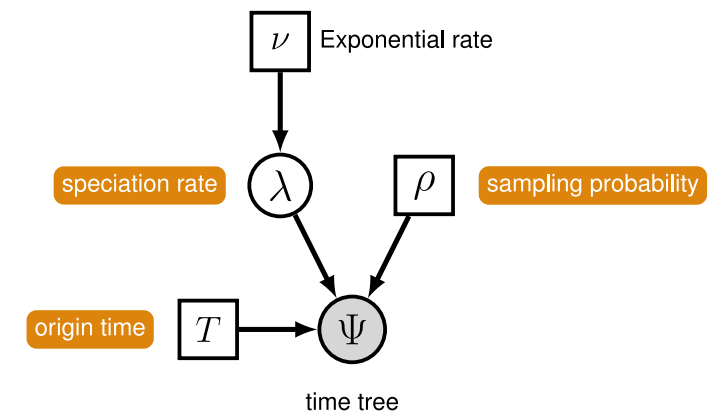
$f(\Psi| \lambda, \rho=1, T=t_2+t_3+t_4+t_5) =$
$2\lambda \exp(2\lambda t_2)/2 \times 3\lambda \exp(3\lambda t_3)/3 \times$
$4\lambda \exp(4\lambda t_4)/4 \times \exp(5\lambda t_5)$

The general birth-death model

- stem age of a clade T
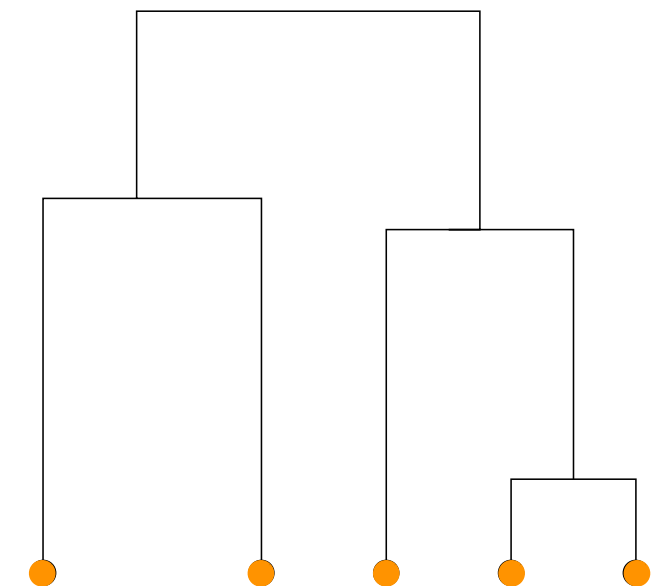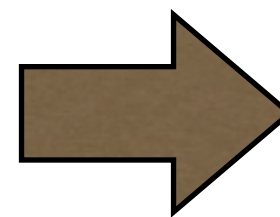- speciation rate λ
- extinction rate μ
- sampling probability ρ

Rates may depend on:
1) constant
2) time (environmental-dependence)
3) # of species (diversity-dependence)
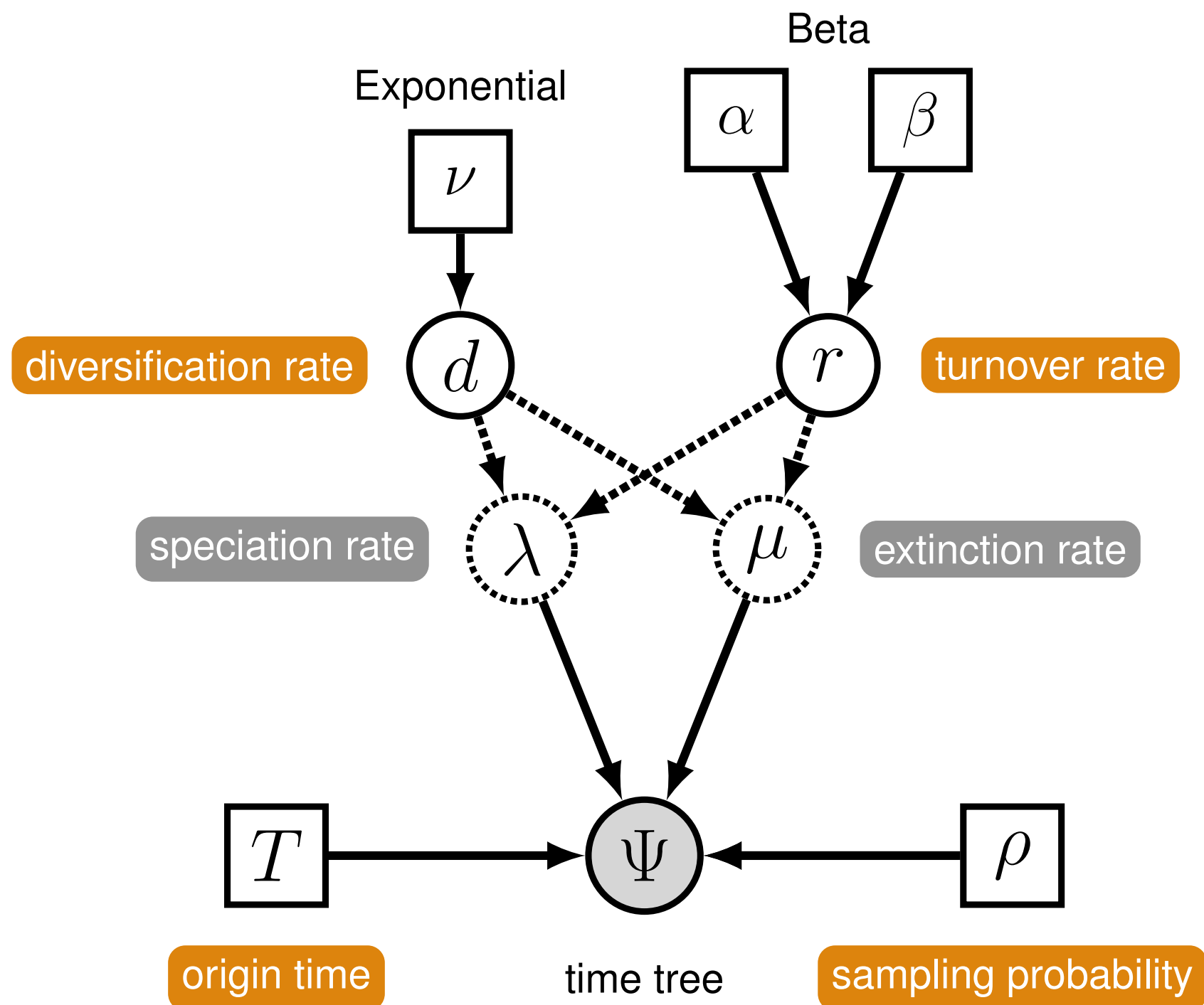4) type of species
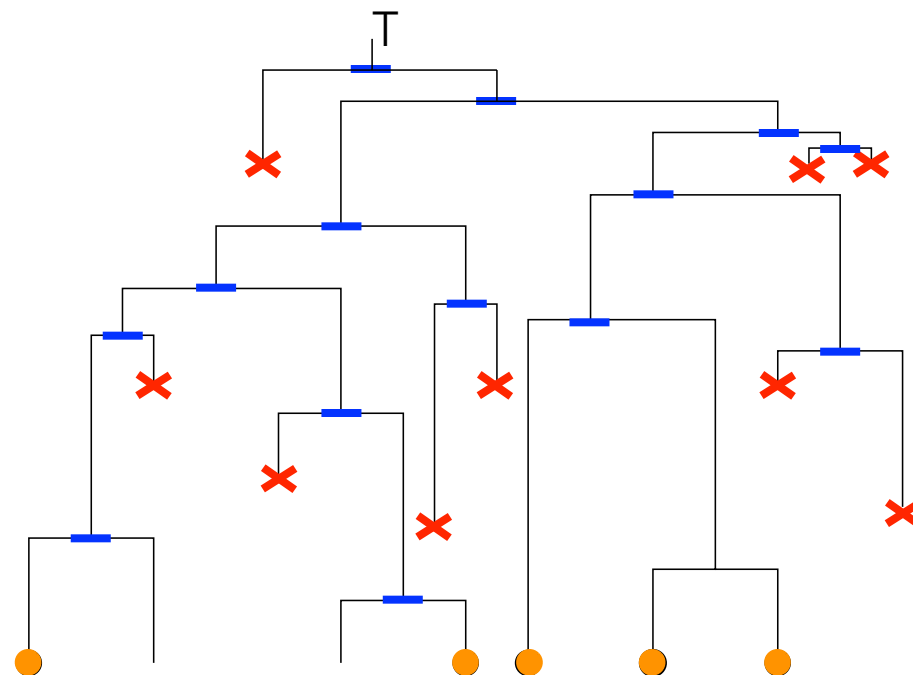
complete phylogeny

reconstructed phylogeny
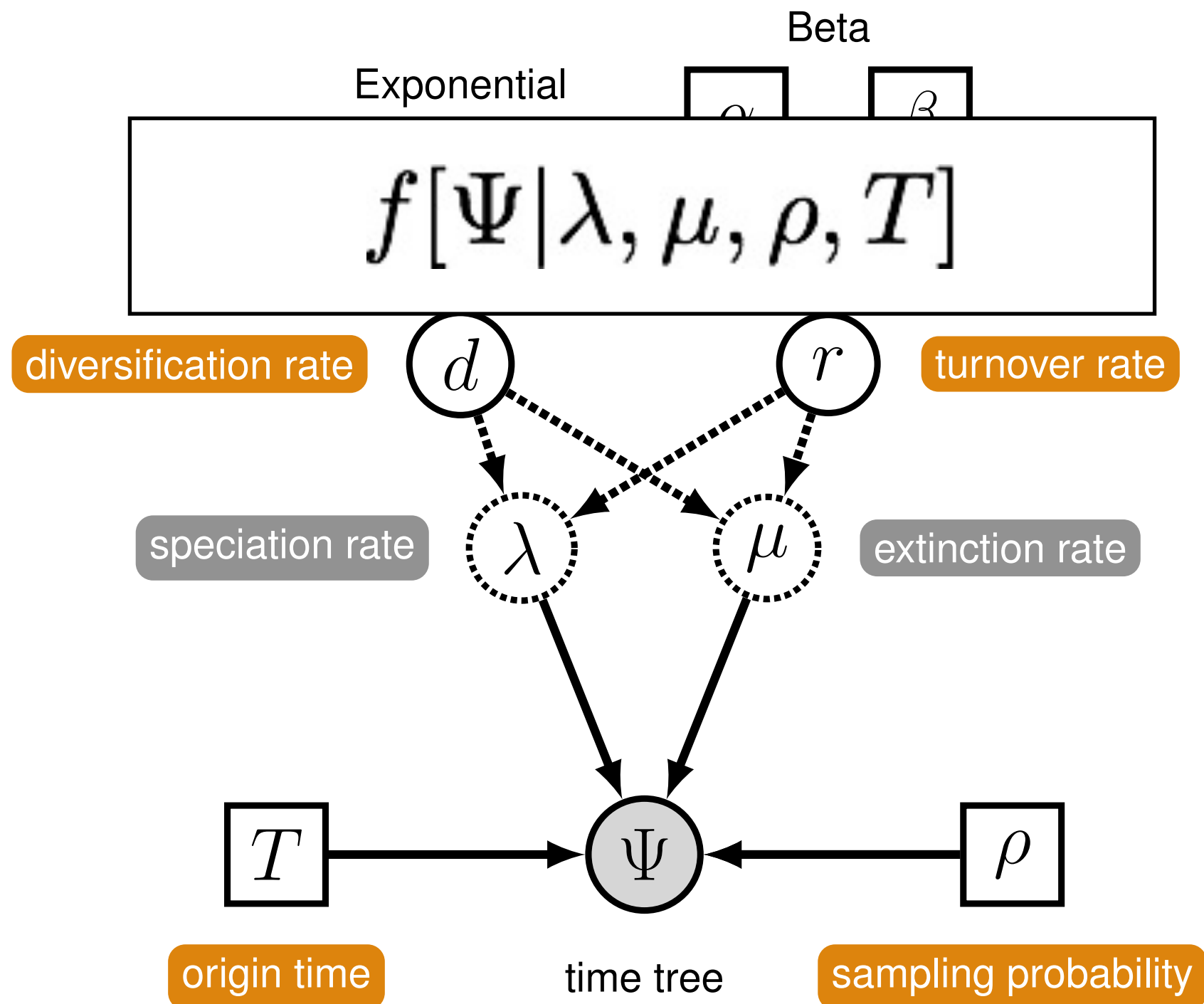
Stadler (JTB, 2009)

12

Constant rate birth-death model in RevBayes

Constant rate birth-death model in RevBayes

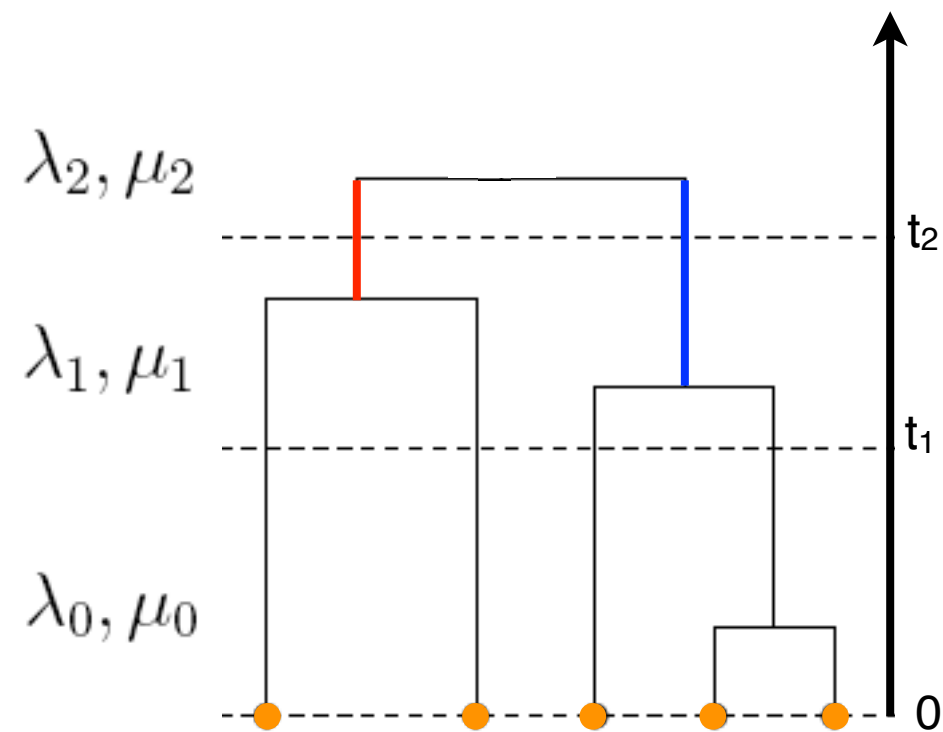# Piecewise constant birth-death model
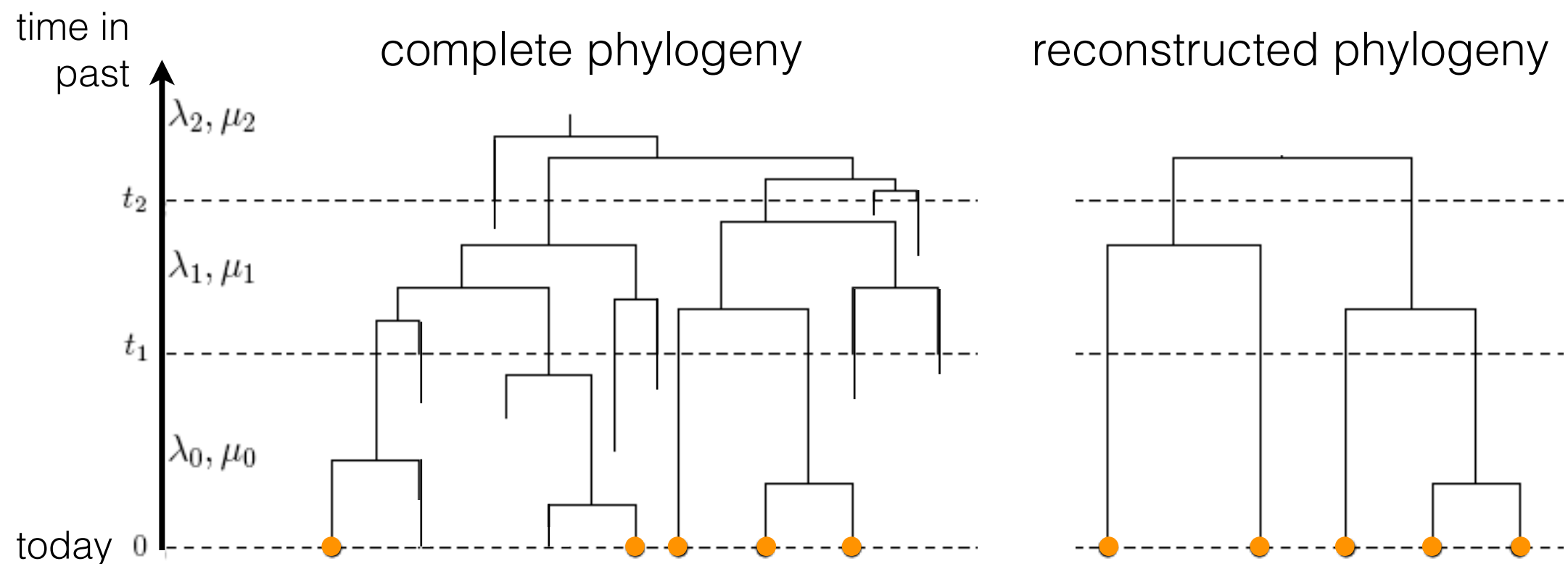
Piecewise constant birth-death model
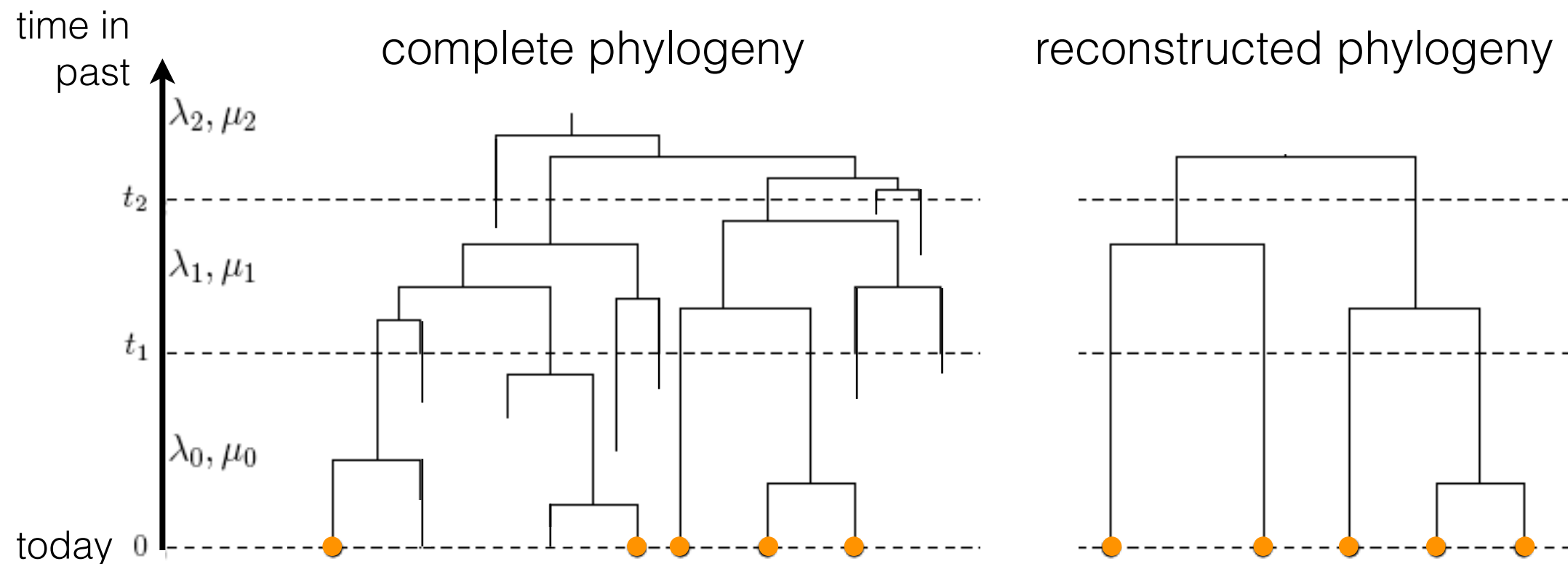
Calculate **likelihood** of reconstructed phylogeny by "integrating" over all complete phylogenies:

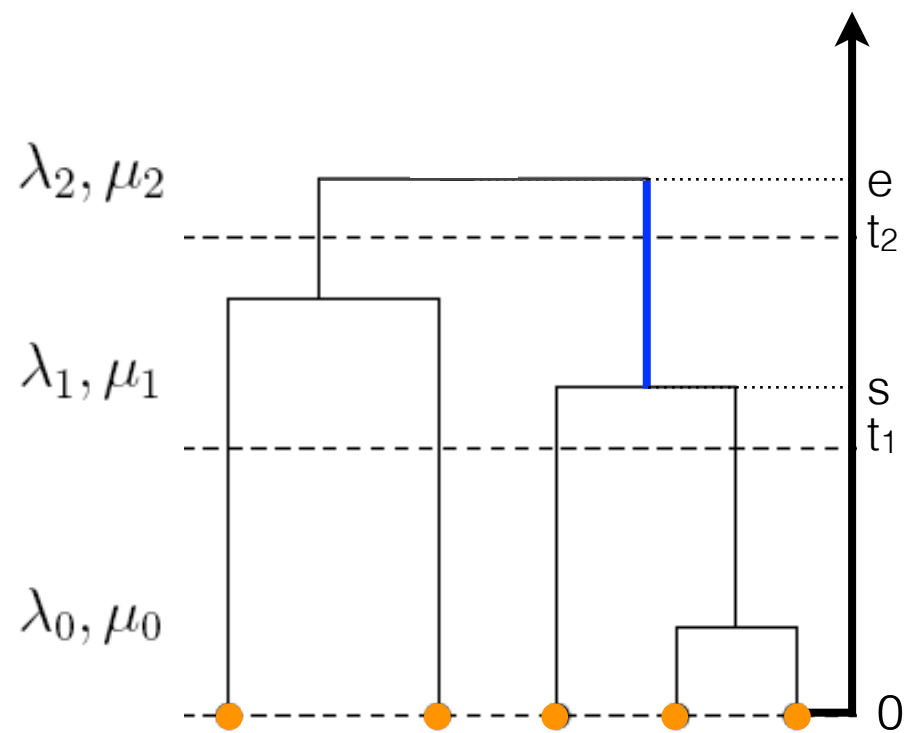**Efficiently done using differential equations and recursions**
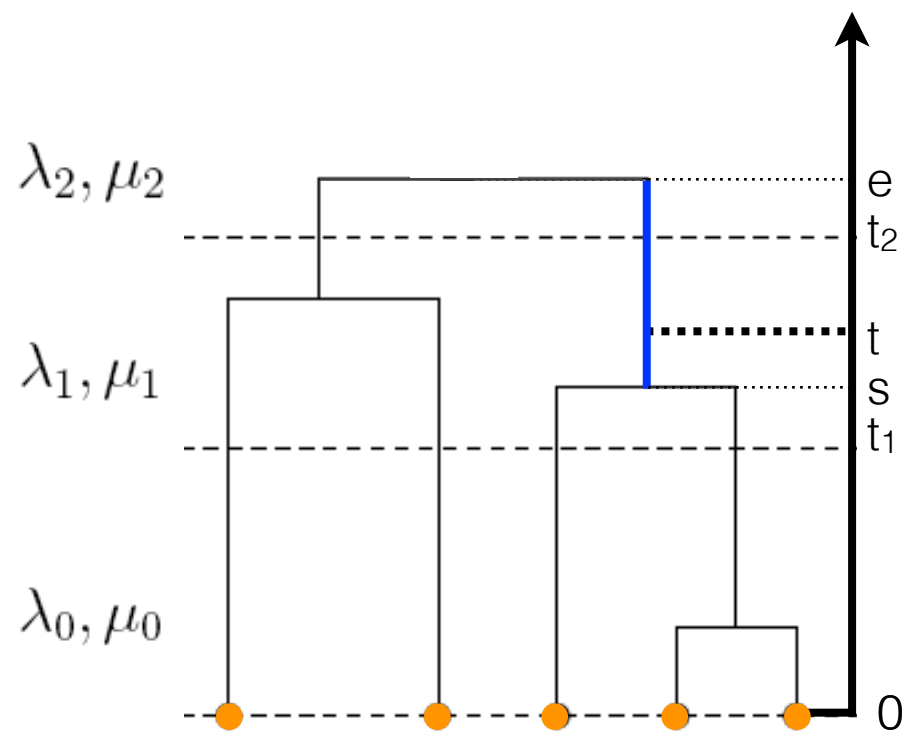
Calculating the likelihood: recursion

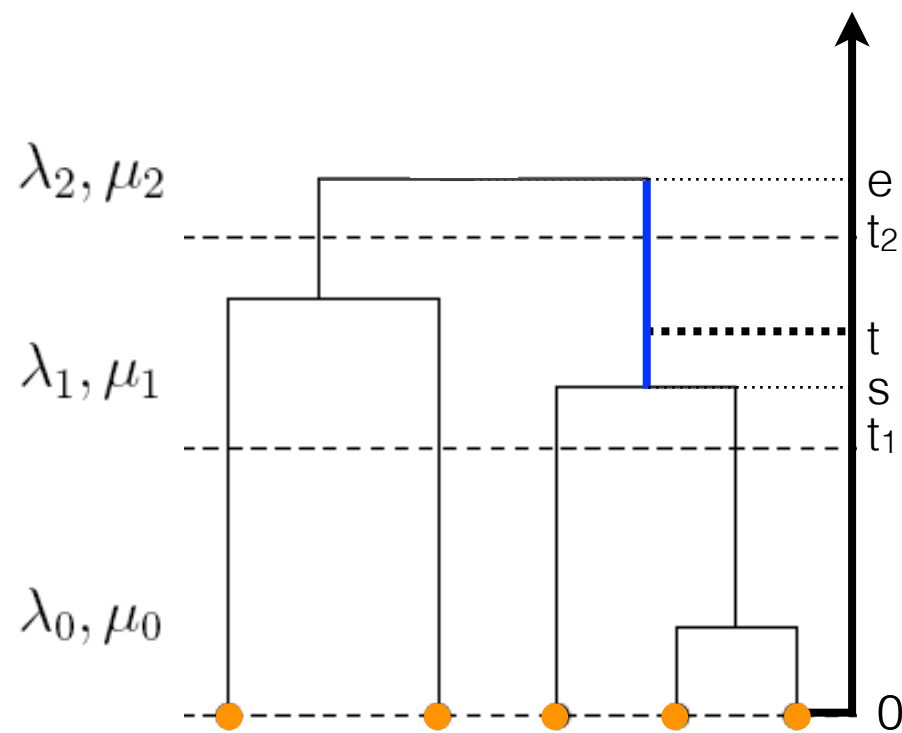# Calculating the likelihood: differential equation

Calculating the likelihood: differential equation

q(t): probability density that a given individual at time t produces the descending edge.

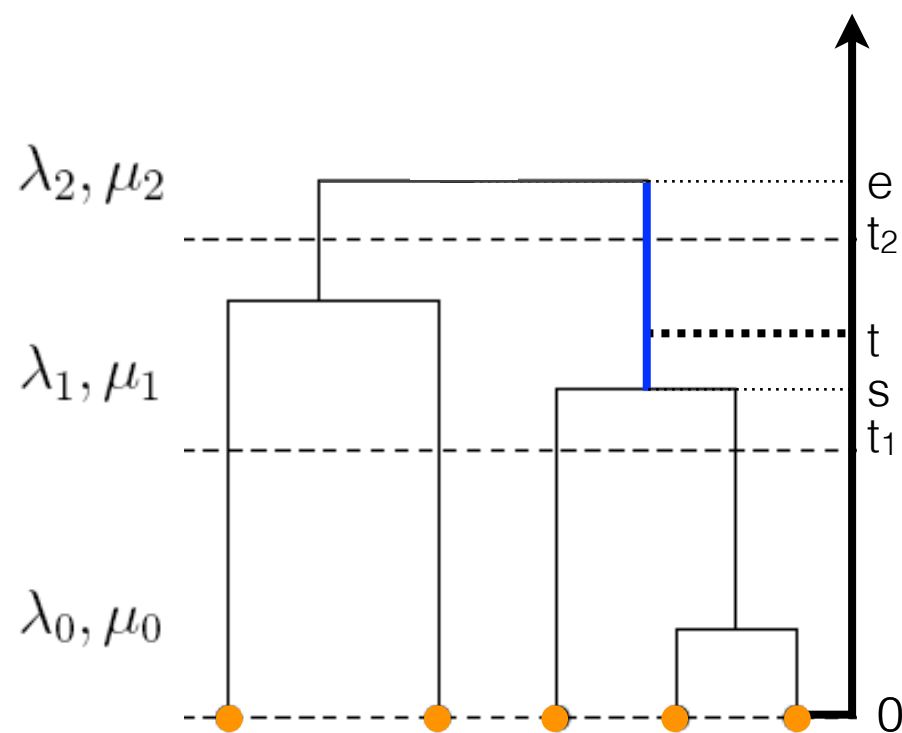# Calculating the likelihood: differential equation

q(t): probability density that a given individual at time t produces the descending edge.

q(e) = Prob (|)

q(s) = 1

Calculating the likelihood: differential equation

λ_2, μ_2

λ_1, μ_1

λ_0, μ_0

e
t_2
t
s
t_1
0

q(t): probability density that a given individual at time t produces the descending edge.

q(e) = Prob (|)

q(s) = 1

p(t): probability that a given individual at time t has no extant and sampled species offspring.

Calculating the likelihood: differential equation
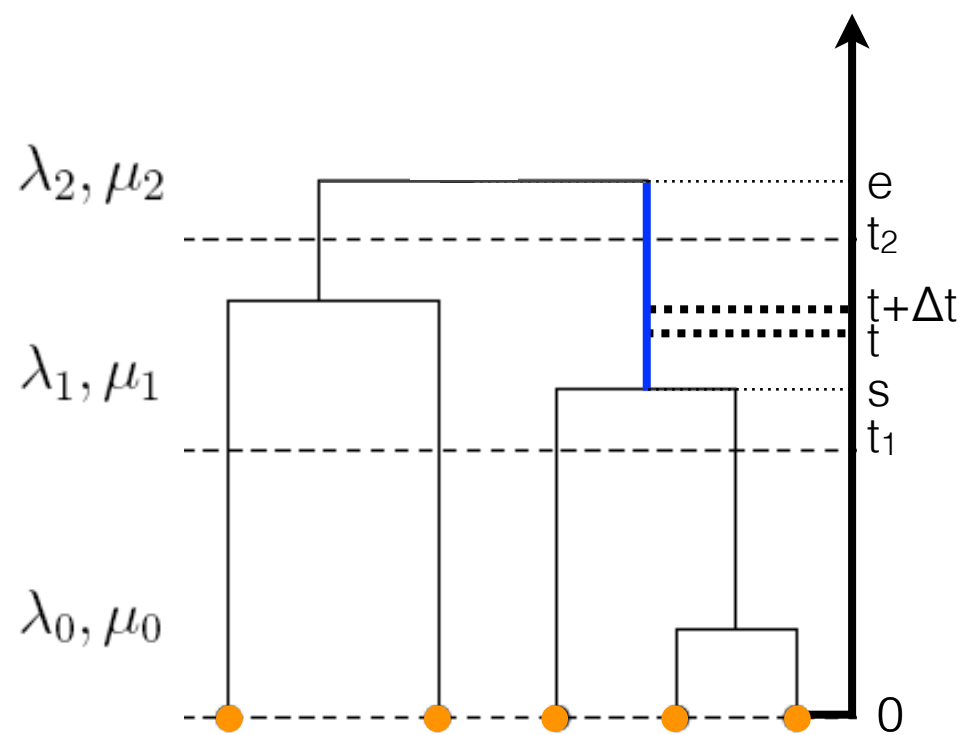

q(t): probability density that a given individual at time t produces the descending edge.

q(e) = Prob (|)

q(s) = 1

p(t): probability that a given individual at time t has no extant and sampled species offspring.

$$q(t + \Delta t) = (1 - (\lambda_i + \mu_i)\Delta t - O(\Delta t^2))q(t) + \lambda_i \Delta t 2 q(t)p(t) + O(\Delta t^2)$$

Calculating the likelihood: differential equation

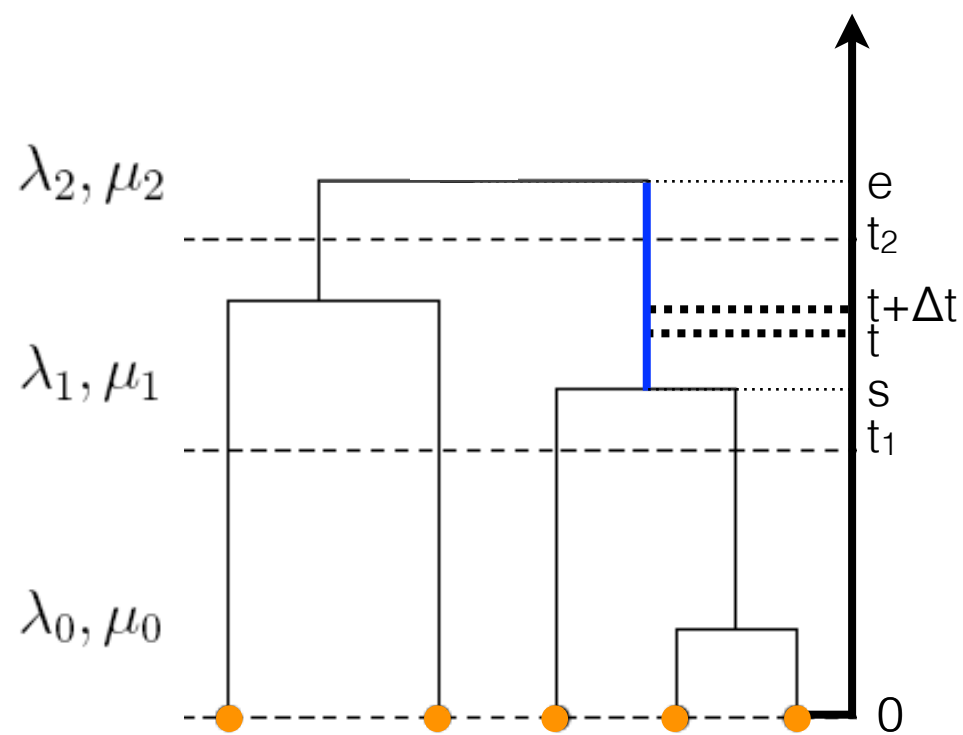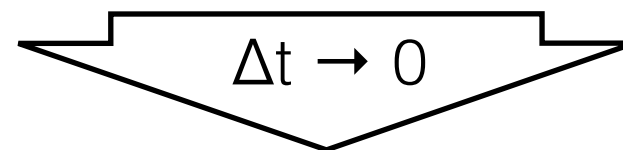q(t): probability density that a given individual at time t produces the descending edge.

q(e) = Prob (|)

q(s) = 1

p(t): probability that a given individual at time t has no extant and sampled species offspring.

$$q(t + \Delta t) = (1 - (\lambda_i + \mu_i)\Delta t - O(\Delta t^2))q(t) + \lambda_i \Delta t 2 q(t)p(t) + O(\Delta t^2)$$
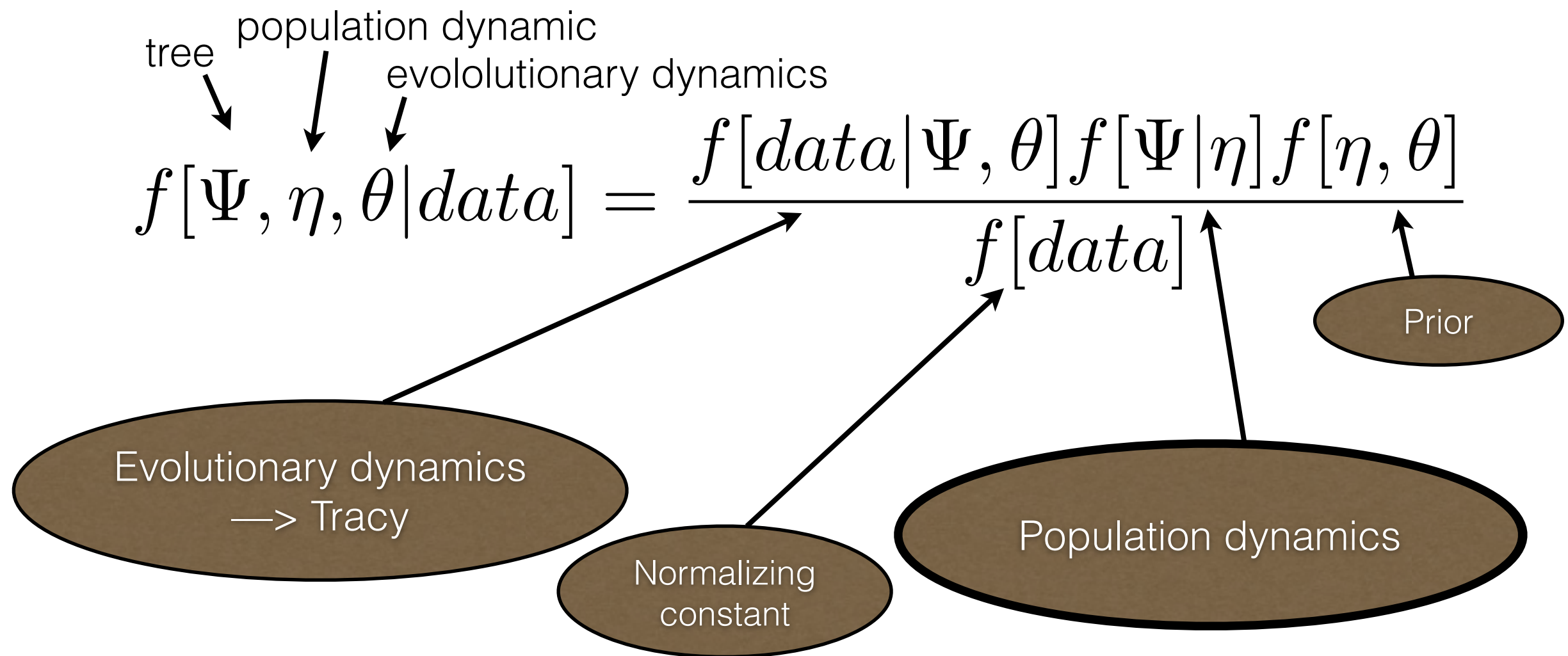
Δt → 0

$$\frac{d}{dt}q(t) = -(\lambda_i + \mu_i)q(t) + 2\lambda_i q(t)p(t)$$

Tree Probability density

$$f \left(\vcenter{\hbox{}} \right) = \rho^n (\lambda_0 - \mu_0)^{2n} \frac{q_{l(x_1)}(x_1)^2}{(1 - p_{l(x_1)}(x_1))^2} \prod_{i=2}^{n-1} \lambda_{l(x_i)} q_{l(x_i)}(x_i) \prod_{i=1}^{m} (\lambda_i - \mu_i)^{2n_i} q_{i-1}(t_i)^{n_i}$$

...with the x_i being the branching times and $p(x_i), q(x_i), l(x_i)$ being known functions (solutions for the differential equations) depending on the parameters λ, μ, ρ, t

Bayesian approach for estimating trees and parameters

tree

population dynamic

evololutionary dynamics

$$f[\Psi, \eta, \theta | data] = \frac{f[data|\Psi, \theta] \, f[\Psi|\eta] \, f[\eta, \theta]}{f[data]}$$

Prior

Evolutionary dynamics
—> Tracy

Normalizing constant

Population dynamics

Using a birth-death model allows us to estimate phylogenies together with speciation and extinction rates

Reminder: birth-death model in RevBayes
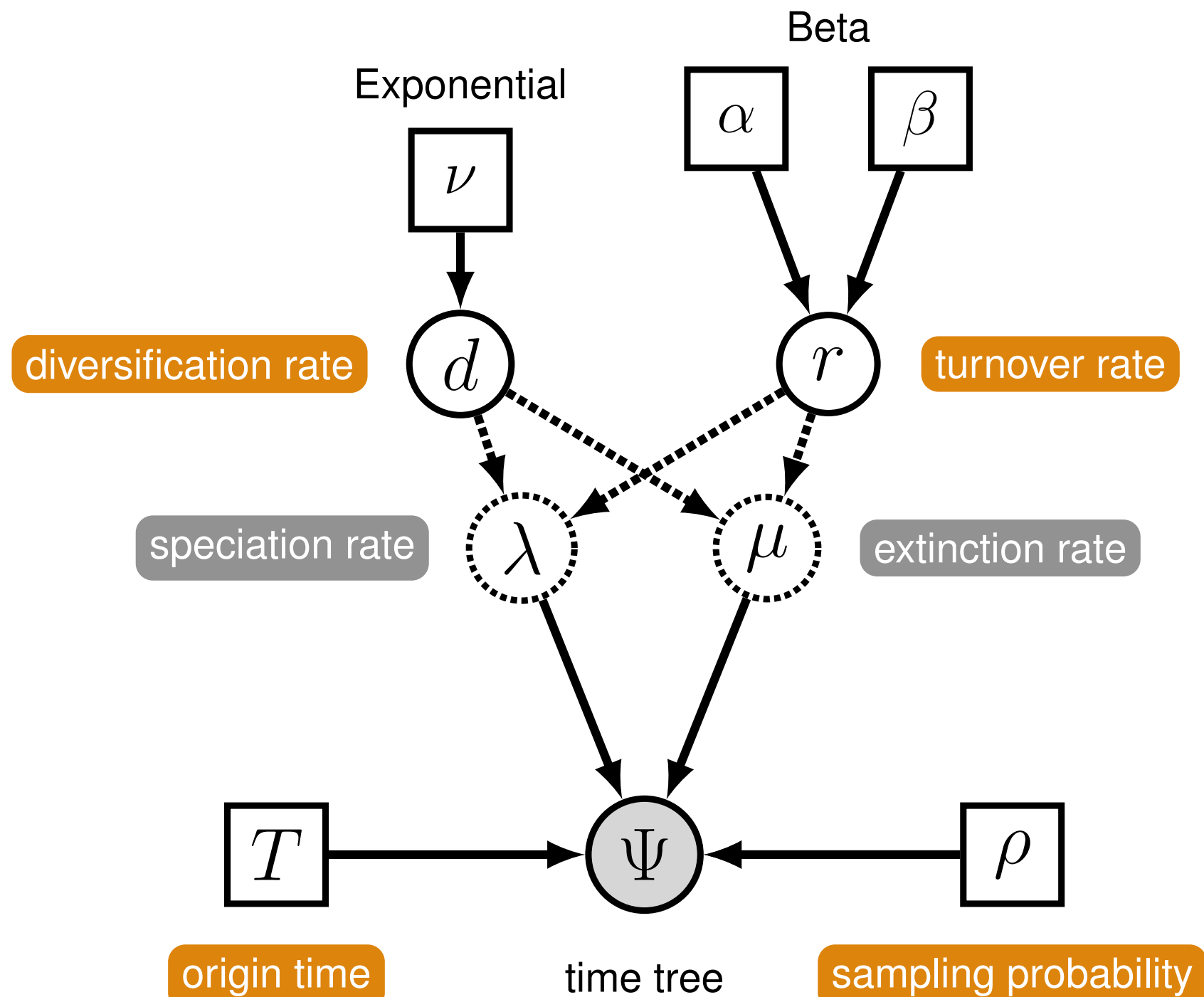
Exercises

Data

Fixed phylogeny of bears

MCMC

Pick best model (with or without extinction?)

Quantify speciation and extinction rate