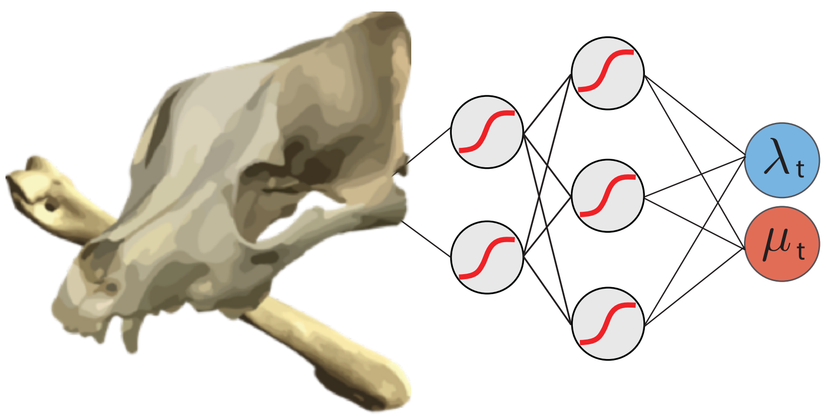
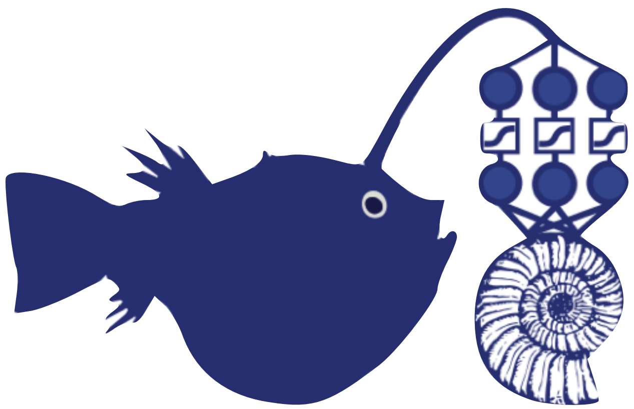
Bayesian and deep learning for macroevolutionary analyses

Gemini's take on our workshop

Learning objectives





PyRate and the BDNN model

Estimating trait and time dependent speciation and extinction rates form fossil occurrence data

DeepDive

Estimating diversity trajectories through time using deep learning and mechanistic simulations

(And more!)

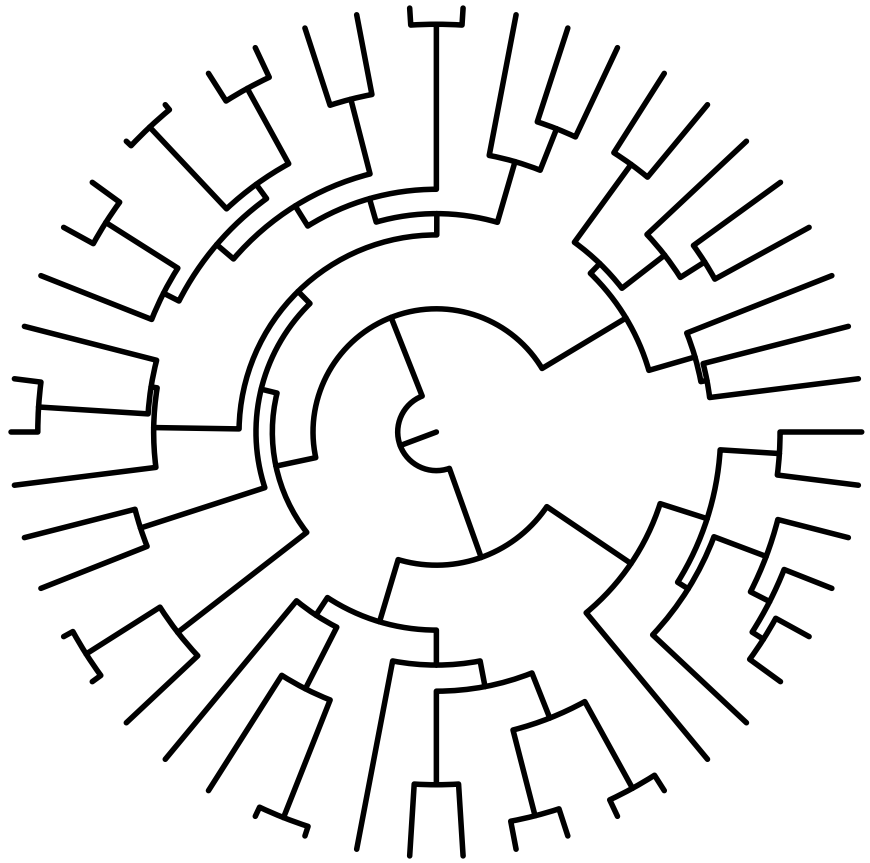
PyRate
Bayesian estimation of macroevolutionary rates from fossil occurrence data

Estimation of macroevolutionary rates from the fossil record

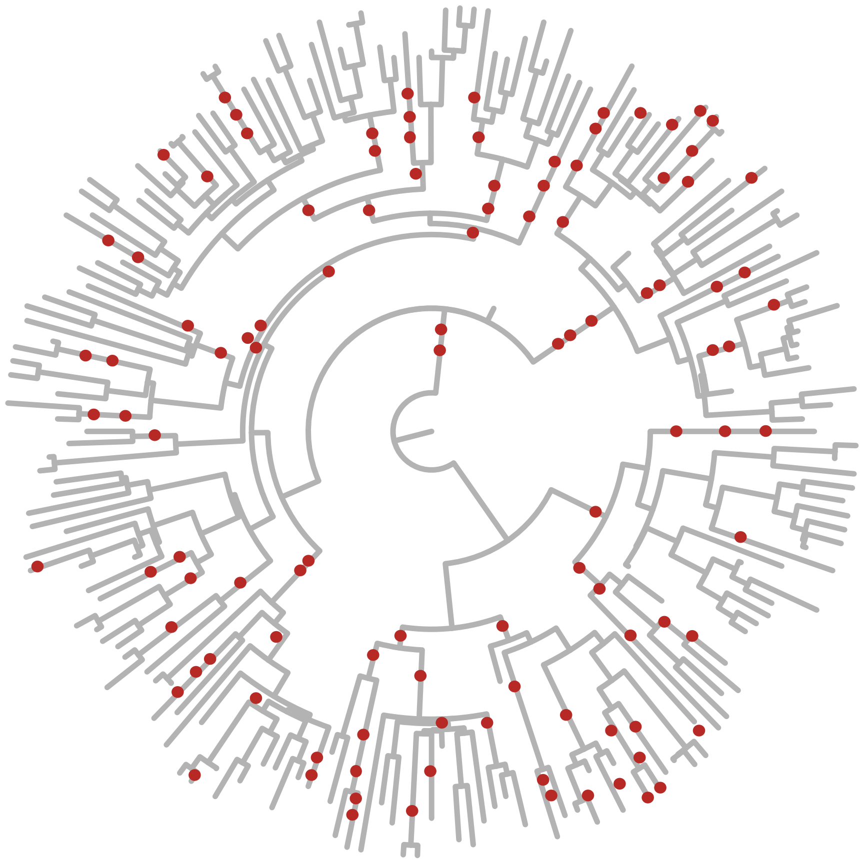




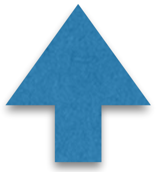

Complete evolutionary history of a clade



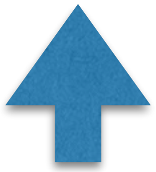
Reconstructed phylogeny of the extant taxa



Fossil record



Phylogenetic comparative methods

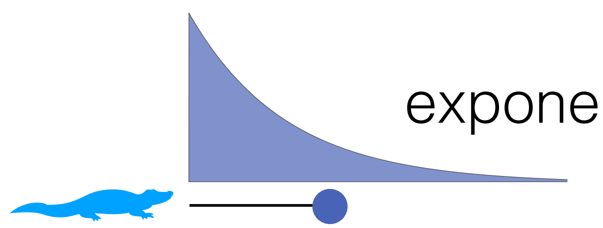


“Stratigraphic approaches”

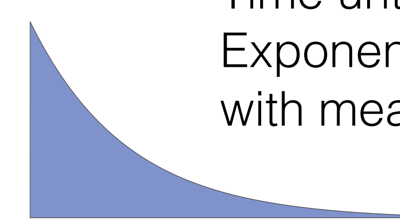
Species diversification as a birth-death stochastic process

λ: expected number of speciation events in 1 My per-lineage

μ: expected number of extinction events in 1 My per-lineage



exponential waiting time until speciation



Time until speciation:
Exponential distribution
with mean 1/λ

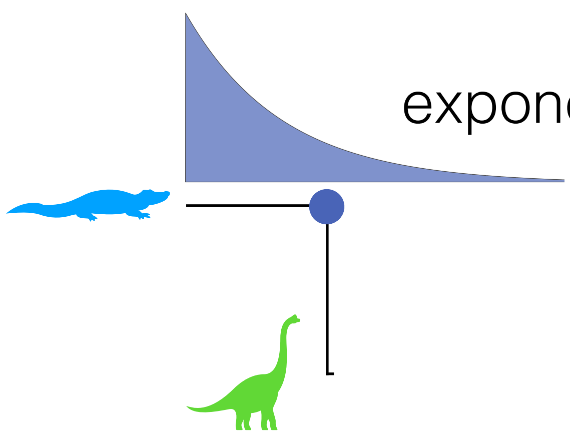
Time until extinction:
Exponential distribution
with mean 1/μ

Kendall 1948 Ann Math Statist; Gernhard 2008 J Theor Biol

Species diversification as a birth-death stochastic process

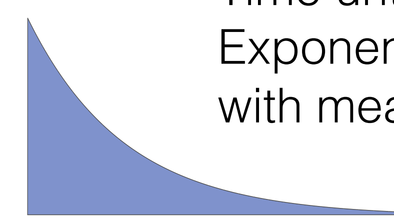
λ: expected number of speciation events in 1 My per-lineage

μ: expected number of extinction events in 1 My per-lineage



exponential waiting time until speciation

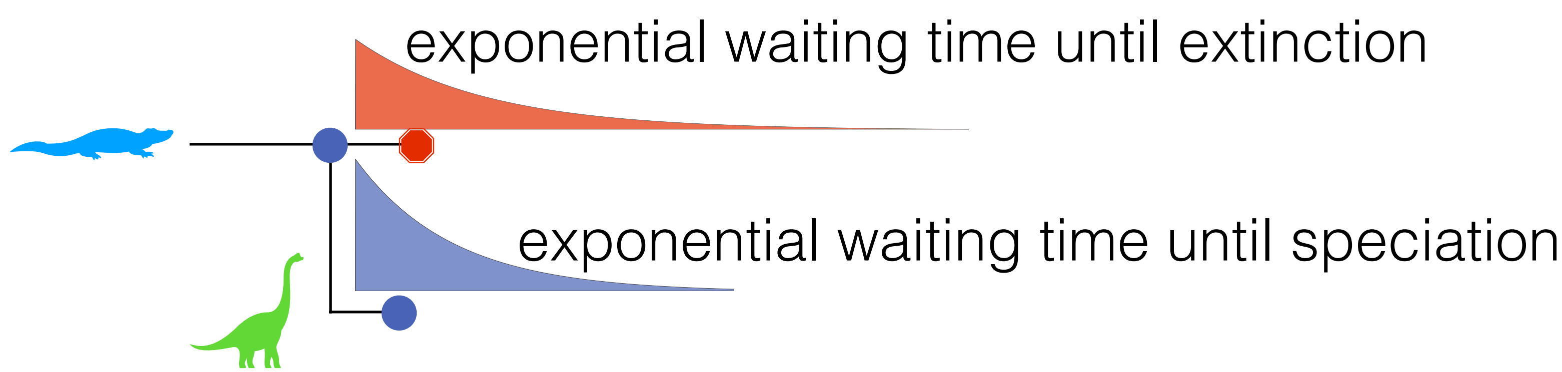
Time until speciation:
Exponential distribution
with mean 1/λ

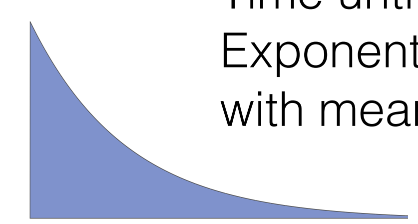


Time until extinction:
Exponential distribution
with mean 1/μ

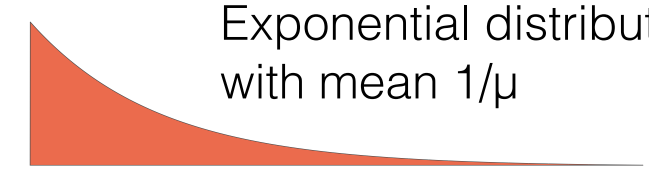
Species diversification as a birth-death stochastic process

λ: expected number of speciation events in 1 My per-lineage
μ: expected number of extinction events in 1 My per-lineage





Time until speciation:
Exponential distribution with mean $1/\lambda$

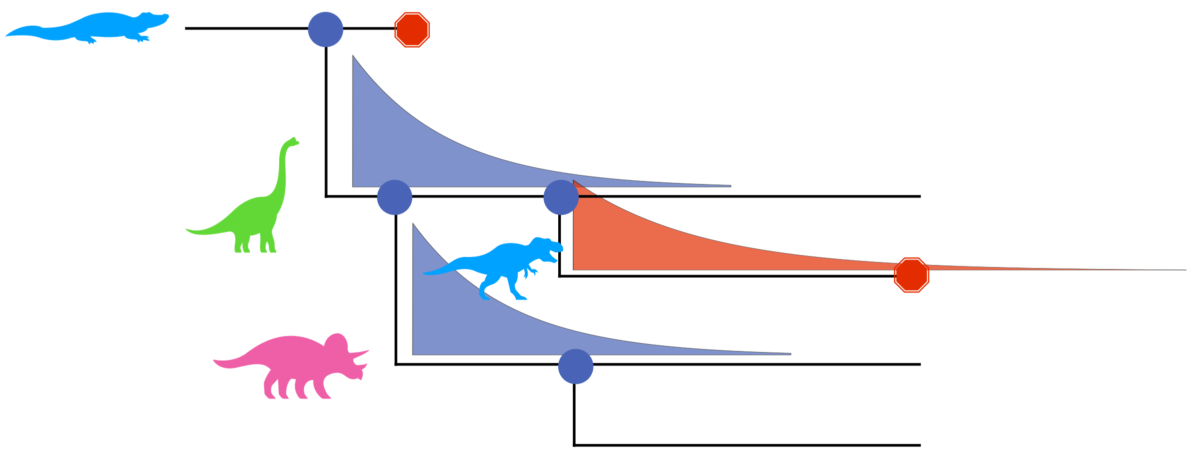


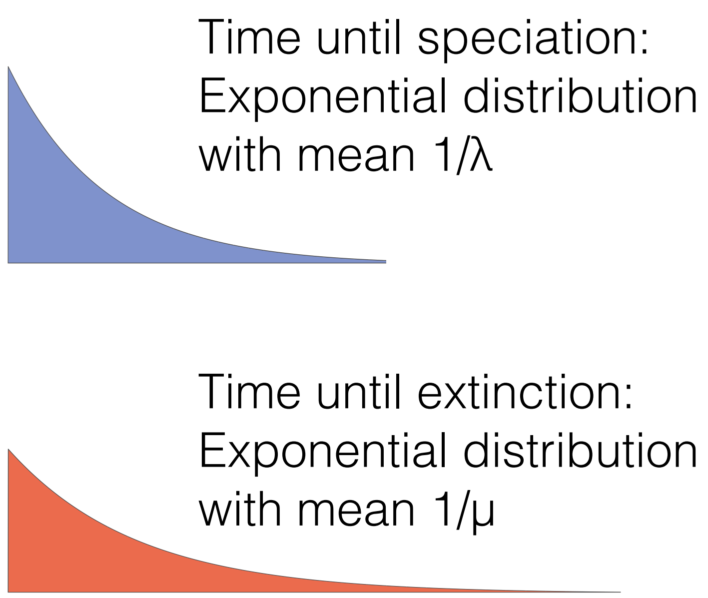
Time until extinction:
Exponential distribution with mean $1/\mu$

Species diversification as a birth-death stochastic process

λ: expected number of speciation events in 1 My per-lineage

μ: expected number of extinction events in 1 My per-lineage

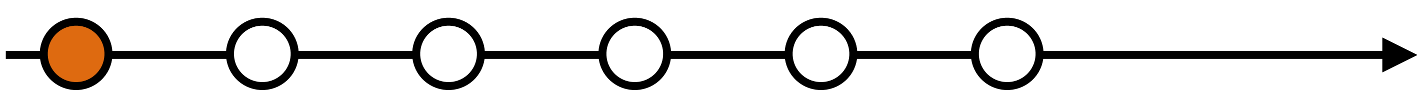




Species diversification as a birth-death stochastic process

Why an exponential waiting time?

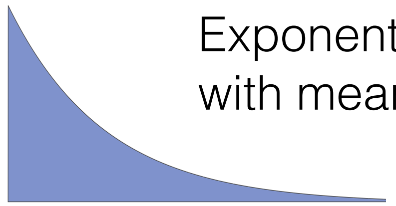
Expected pattern from a constant rate random process



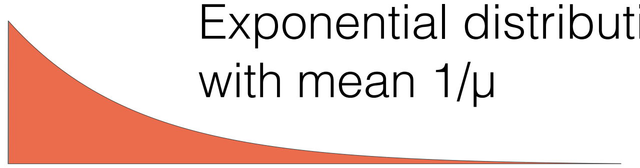
P_{event} (t = 0) = 0.5

Probability of the event

Time until speciation:
Exponential distribution
with mean 1/λ



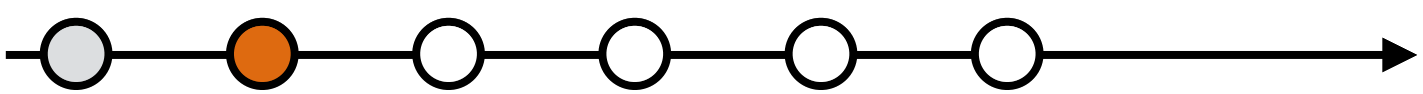
Time until extinction:
Exponential distribution
with mean 1/μ



Species diversification as a birth-death stochastic process

Why an exponential waiting time?

Expected pattern from a constant rate random process



$P_{event}\,(t = 2) = 0.5 \times (1 - 0.5)^{t-1} = 0.25$

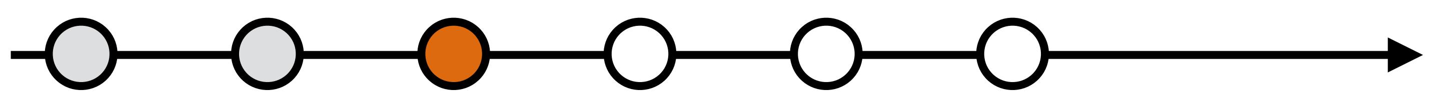
Probability of the event

Probability that it didn't occur before time *t*

Species diversification as a birth-death stochastic process

Why an exponential waiting time?

Expected pattern from a constant rate random process



$$P_{event}\,(t = 3) = 0.5 \times (1 - 0.5)^{t-1} = 0.125$$

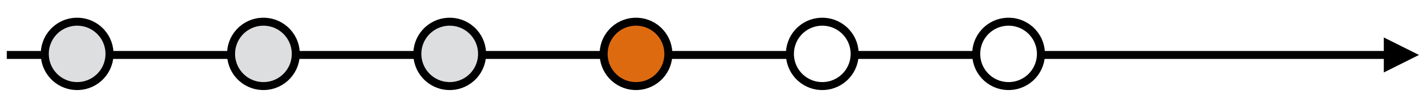
Probability of the event

Probability that it didn't occur before time *t*

Species diversification as a birth-death stochastic process

Why an exponential waiting time?

Expected pattern from a constant rate random process



$P_{event} (t = 4) = 0.5 \times (1 - 0.5)^{t - 1} = 0.063$

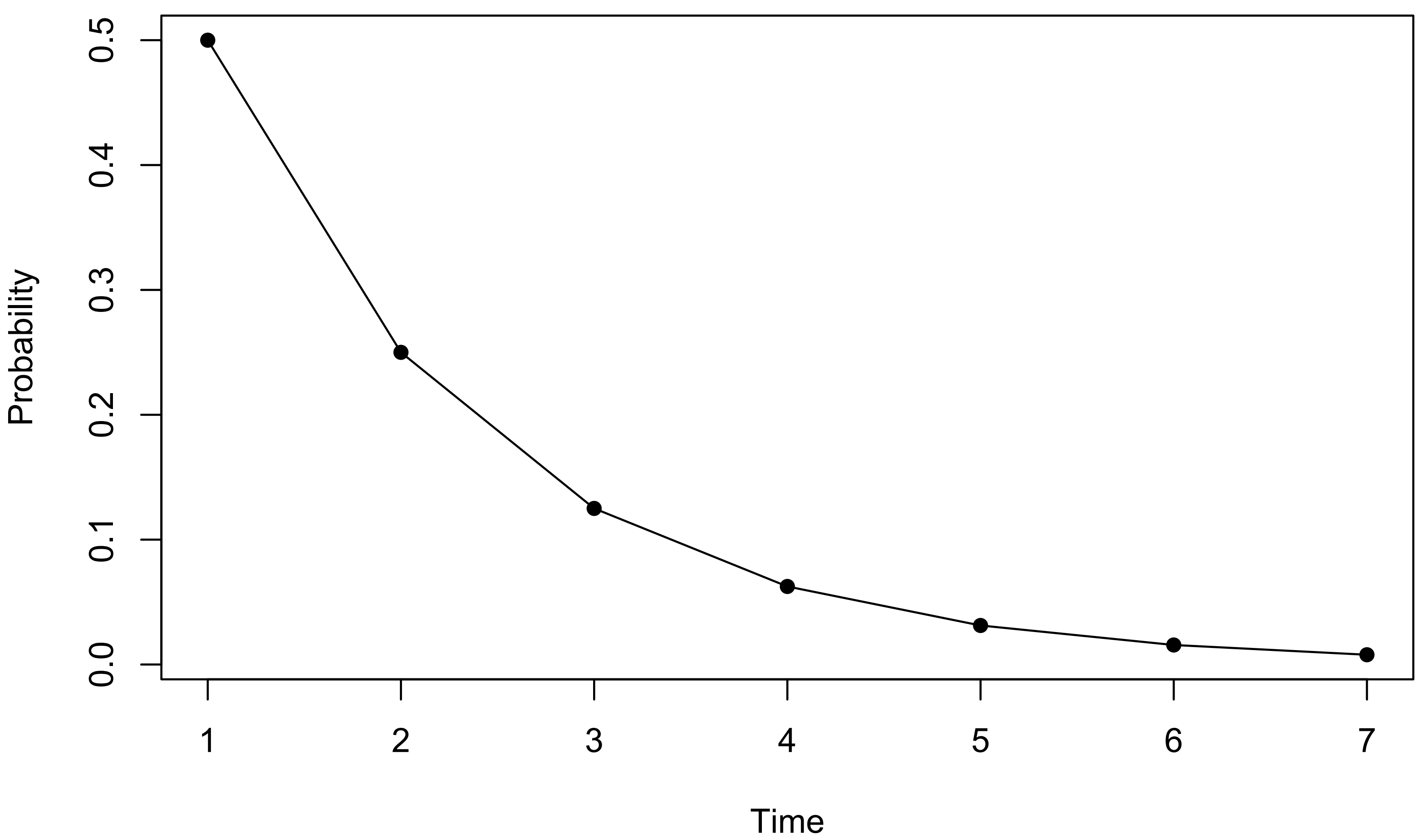
Probability of the event

Probability that it didn't occur before time *t*

Species diversification as a birth-death stochastic process

Why an exponential waiting time?

Expected pattern from a constant rate random process



Species diversification as a birth-death stochastic process

Likelihood of a birth-death process

Times of speciation and extinction

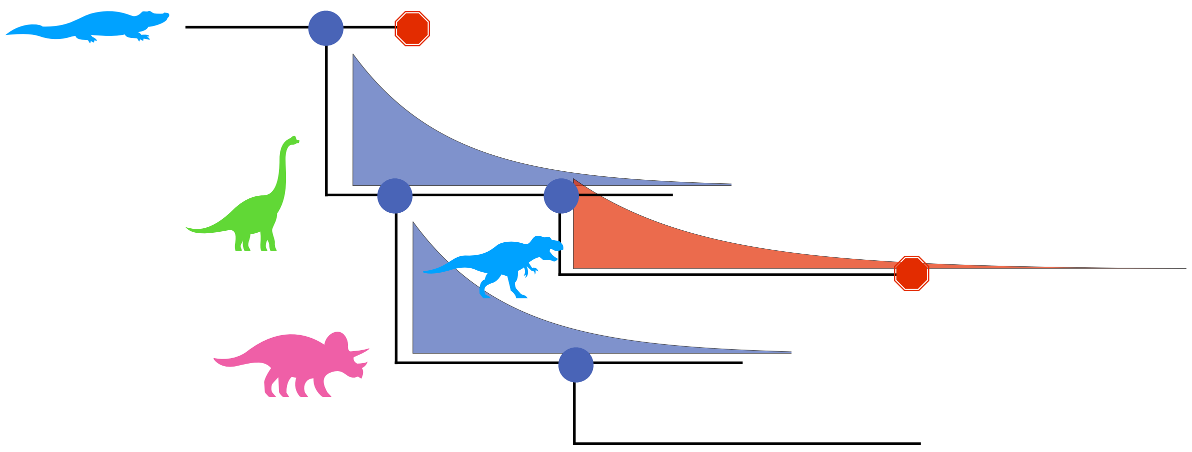
Rates of speciation and extinction

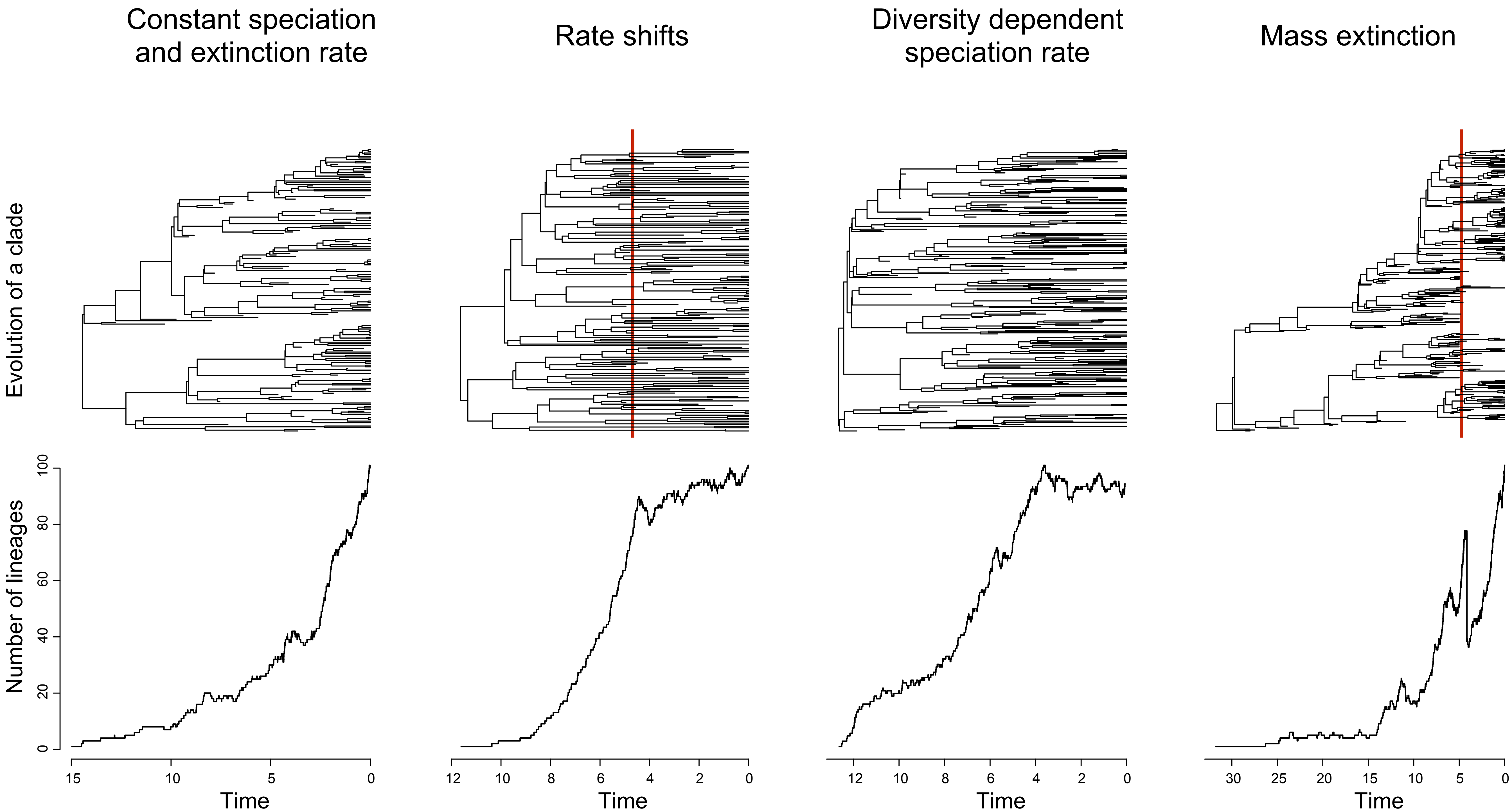
$$P\left(s, e, |\lambda, \mu\right) \propto \lambda_s \times \mu_e \times \exp\left(-\int_s^e \lambda_t + \mu_t \; dt\right)$$

Probability of the speciation event

Probability of the extinction event

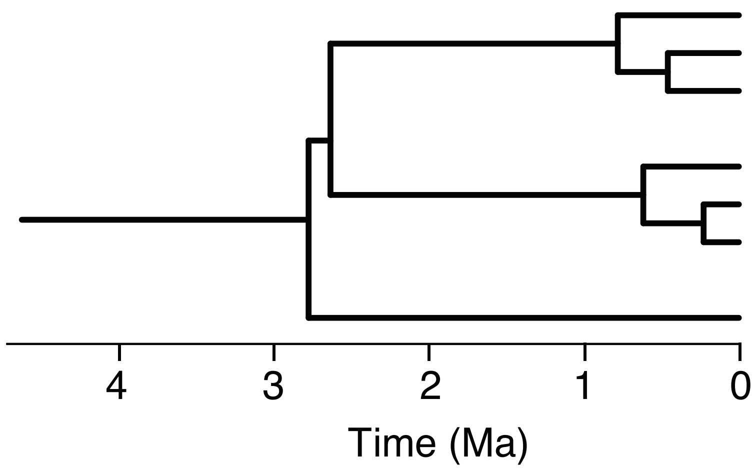
Probability of the waiting time between events



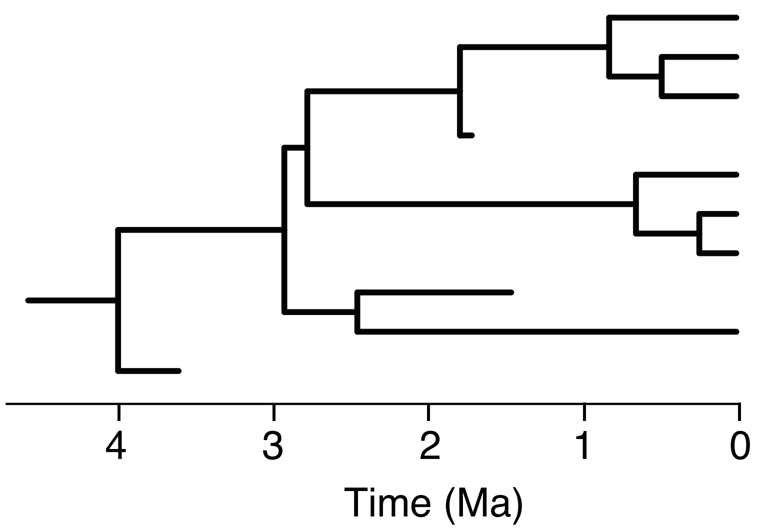

Constant speciation and extinction rate
Rate shifts
Diversity dependent speciation rate
Mass extinction
Evolution of a clade
Number of lineages
0
20
40
60
80
100
15
10
5
0
Time
12
10
8
6
4
2
0
Time
12
10
8
6
4
2
0
Time
30
25
20
15
10
5
0
Time

Methods to infer speciation and extinction rates

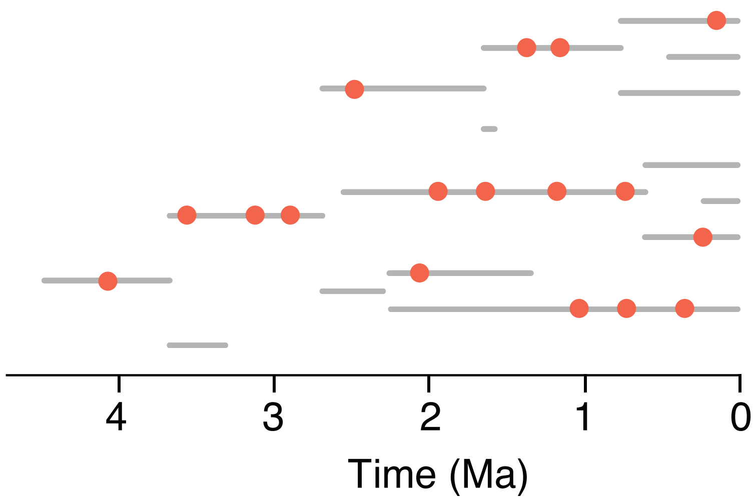
Phylogeny of extant species



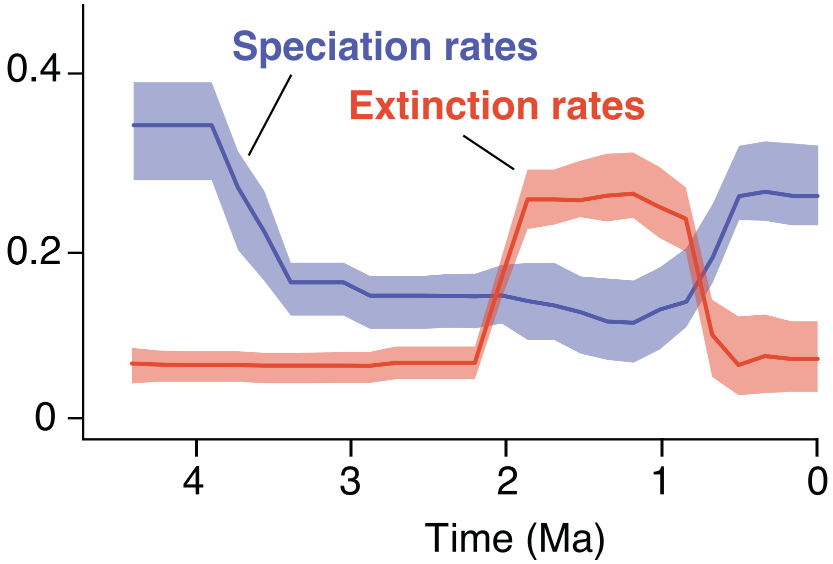
Complete evolutionary history



Fossil record

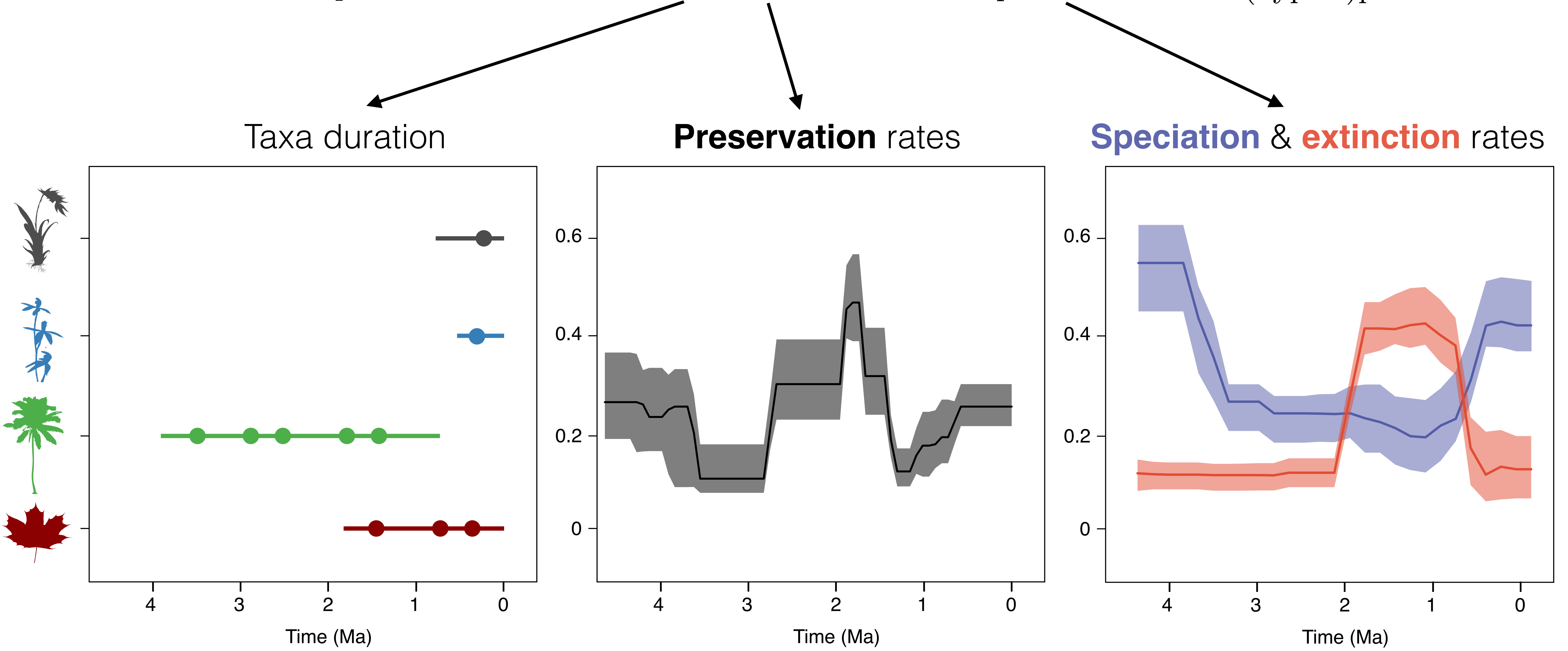


Diversification dynamics



Bayesian inference of macroevolutionary rates using fossil occurrences

$$\underbrace{P(q, \mathbf{s}, \mathbf{e}, \lambda, \mu | X)}_{\text{posterior}} \propto \underbrace{P(X | q, \mathbf{s}, \mathbf{e})}_{\text{likelihood}} \times \underbrace{P(\mathbf{s}, \mathbf{e} | \lambda, \mu)}_{\text{BD prior}} \times \underbrace{P(q) P(\lambda, \mu)}_{\text{other (hyper-)priors}}$$



PyRate

https://github.com/dsilvestro/PyRate

Silvestro et al. 2014 Syst Biol, MEE

A hierarchical probabilistic framework

$$P(s, e, q, \lambda, \mu | x) \propto P(x | s, e, q) \times P(s, e | \lambda, \mu) \times P(q) P(\lambda, \mu)$$

POSTERIOR LIKELIHOOD PRIOR HYPERPRIORS

Probability of:

1. times of speciation/extinction (*s,e*)
2. preservation rate (*q*)
3. speciation/extinction rates (λ,μ)

A hierarchical probabilistic framework

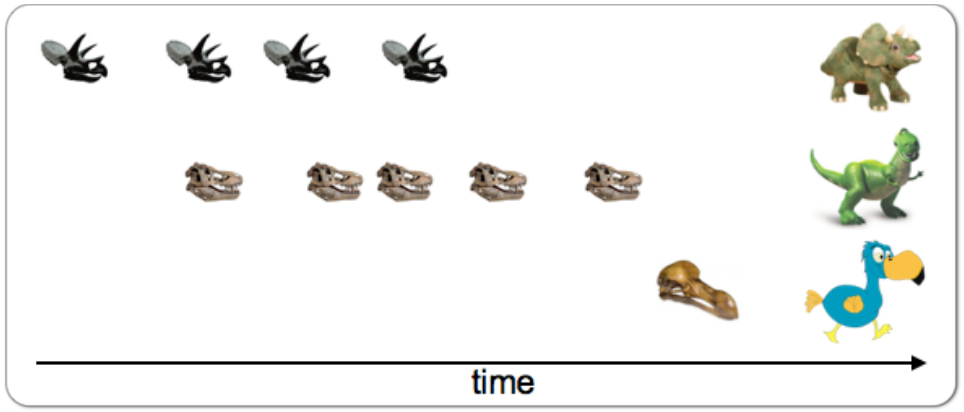
Probability of the data given speciation
and extinction times, preservation rate

$$P(s, e, q, \lambda, \mu | x) \propto P(x | s, e, q) \times P(s, e | \lambda, \mu) \times P(q) P(\lambda, \mu)$$

POSTERIOR LIKELIHOOD PRIOR HYPERPRIORS

Probability of:
1. times of speciation/extinction (*s,e*)
2. preservation rate (*q*)
3. speciation/extinction rates (λ,μ)

x =



A hierarchical probabilistic framework

Prior probability of the speciation and extinction times (Birth-Death process)

$$P(s, e, q, \lambda, \mu | x) \propto P(x | s, e, q) \times P(s, e | \lambda, \mu) \times P(q) P(\lambda, \mu)$$

POSTERIOR LIKELIHOOD PRIOR HYPERPRIORS

A hierarchical probabilistic framework

Prior distributions for
rate parameters

$$P(s, e, q, \lambda, \mu | x) \propto P(x | s, e, q) \times P(s, e | \lambda, \mu) \times P(q)P(\lambda, \mu)$$

POSTERIOR LIKELIHOOD PRIOR HYPERPRIORS

Bayesian inference of macroevolutionary rates using fossil occurrences

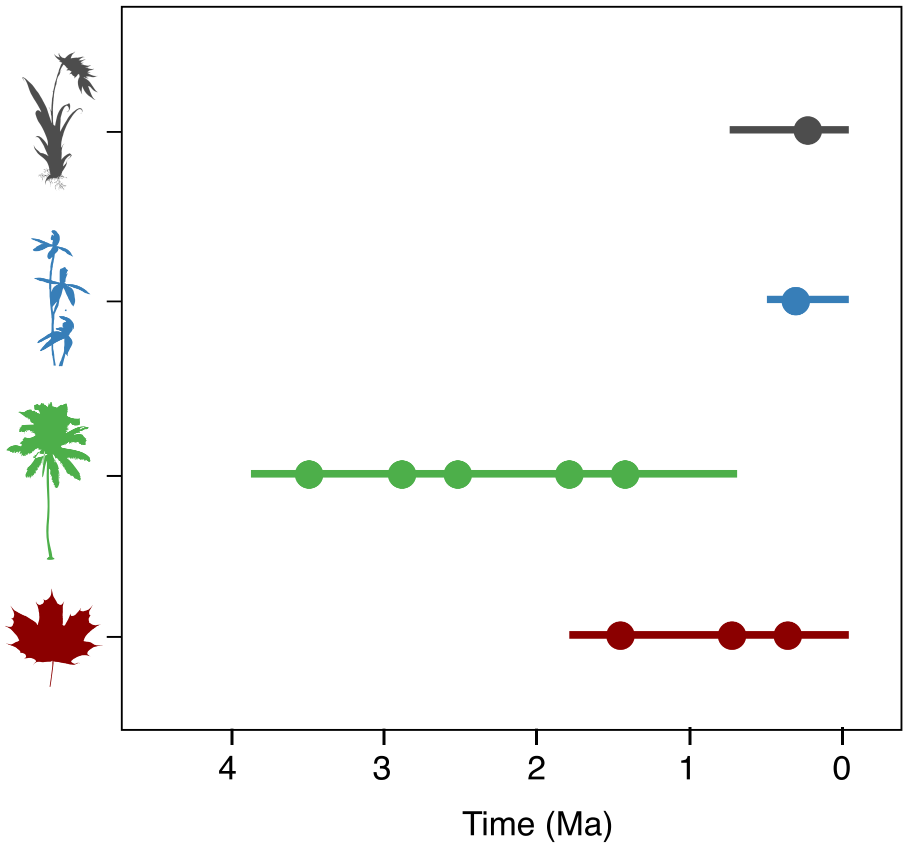
$$\underbrace{P(q,\mathbf{s},\mathbf{e},\lambda,\mu|X)}_{\text{posterior}} \propto \underbrace{P(X|q,\mathbf{s},\mathbf{e})}_{\text{likelihood}} \times \underbrace{P(\mathbf{s},\mathbf{e}|\lambda,\mu)}_{\text{birth-death prior}} \times \underbrace{P(q)P(\lambda,\mu)}_{\text{other (hyper)priors}}$$

Metropolis-Hastings MCMC

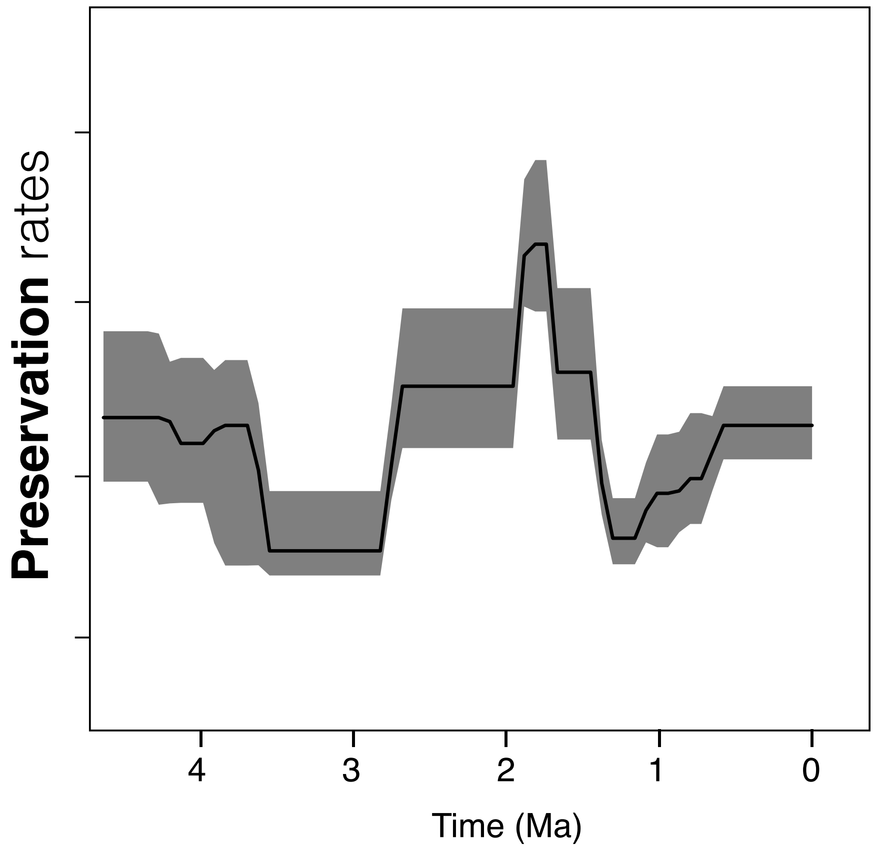
Reversible-jump MCMC

Sampled directly from conjugate posterior

Species duration



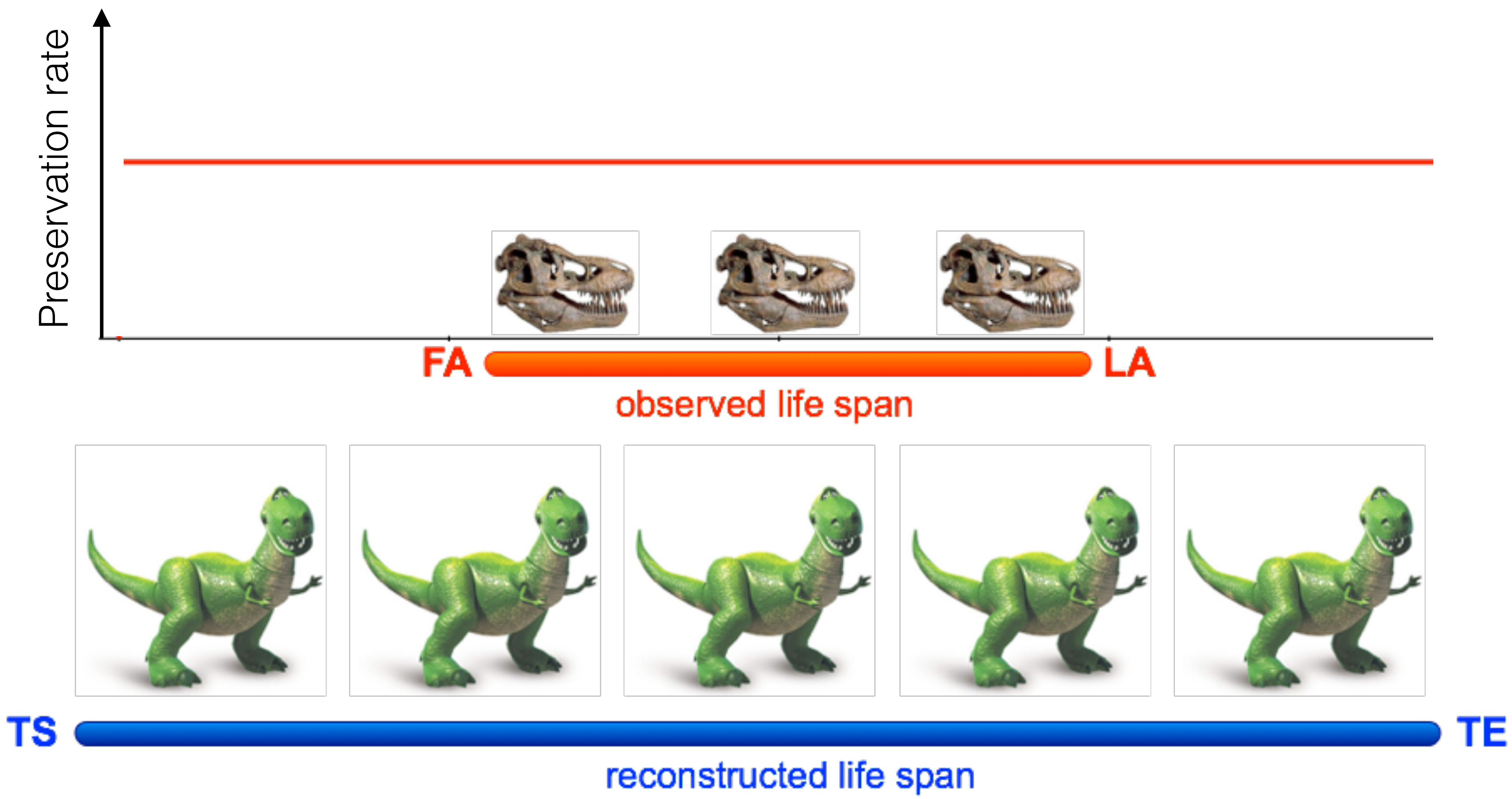
Poisson process



Estimated preservation rates and times of speciation and extinction

$$P(s, e, q, \lambda, \mu | x) \propto \underbrace{P(x | s, e, q)}_{\text{LIKELIHOOD}} \times P(s, e | \lambda, \mu) \times P(q) P(\lambda, \mu)$$

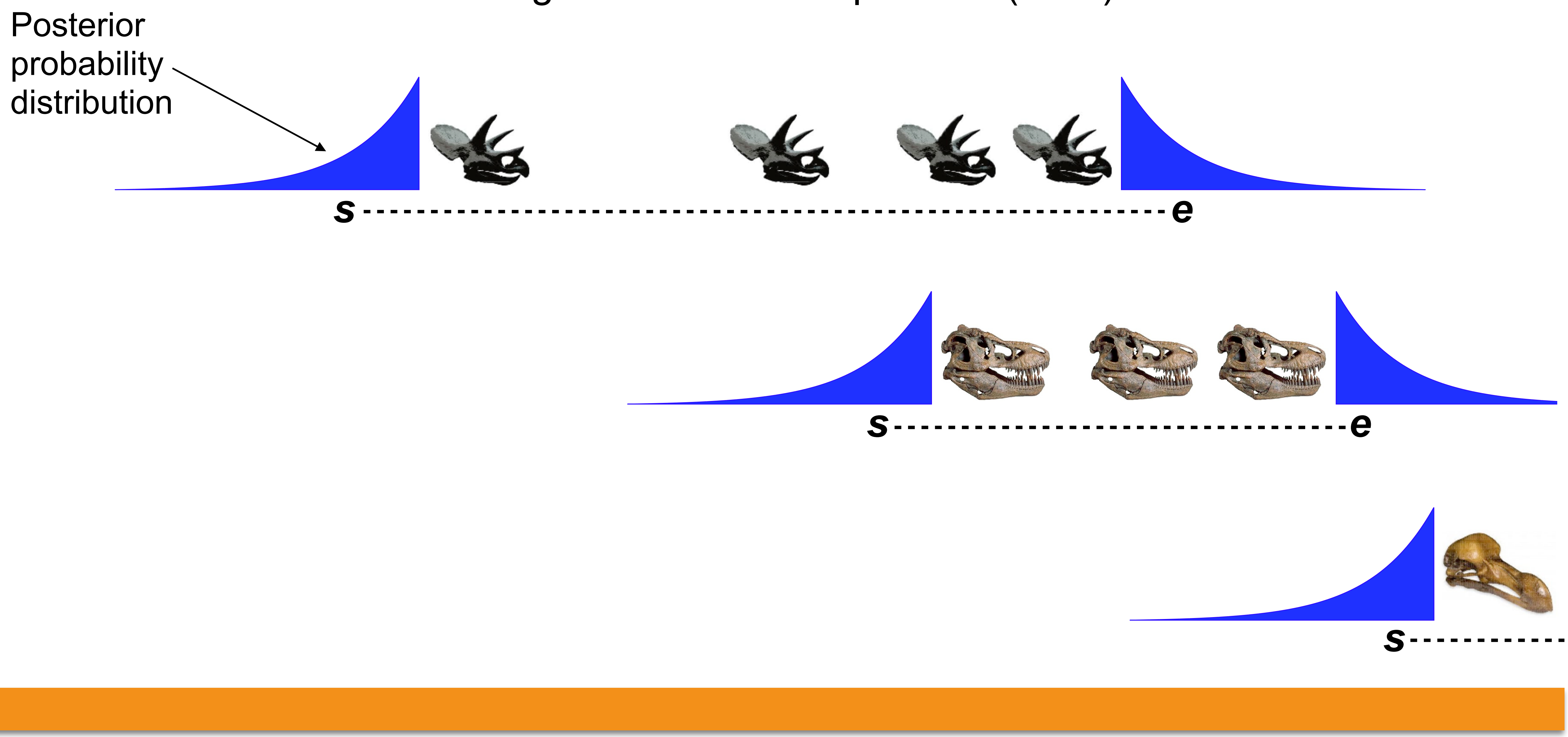
Homogeneous Poisson process (HPP)



Estimated preservation rates and times of speciation and extinction

$$P(s, e, q, \lambda, \mu | x) \propto \underbrace{P(x | s, e, q)}_{\text{LIKELIHOOD}} \times P(s, e | \lambda, \mu) \times P(q) P(\lambda, \mu)$$

Homogeneous Poisson process (HPP)

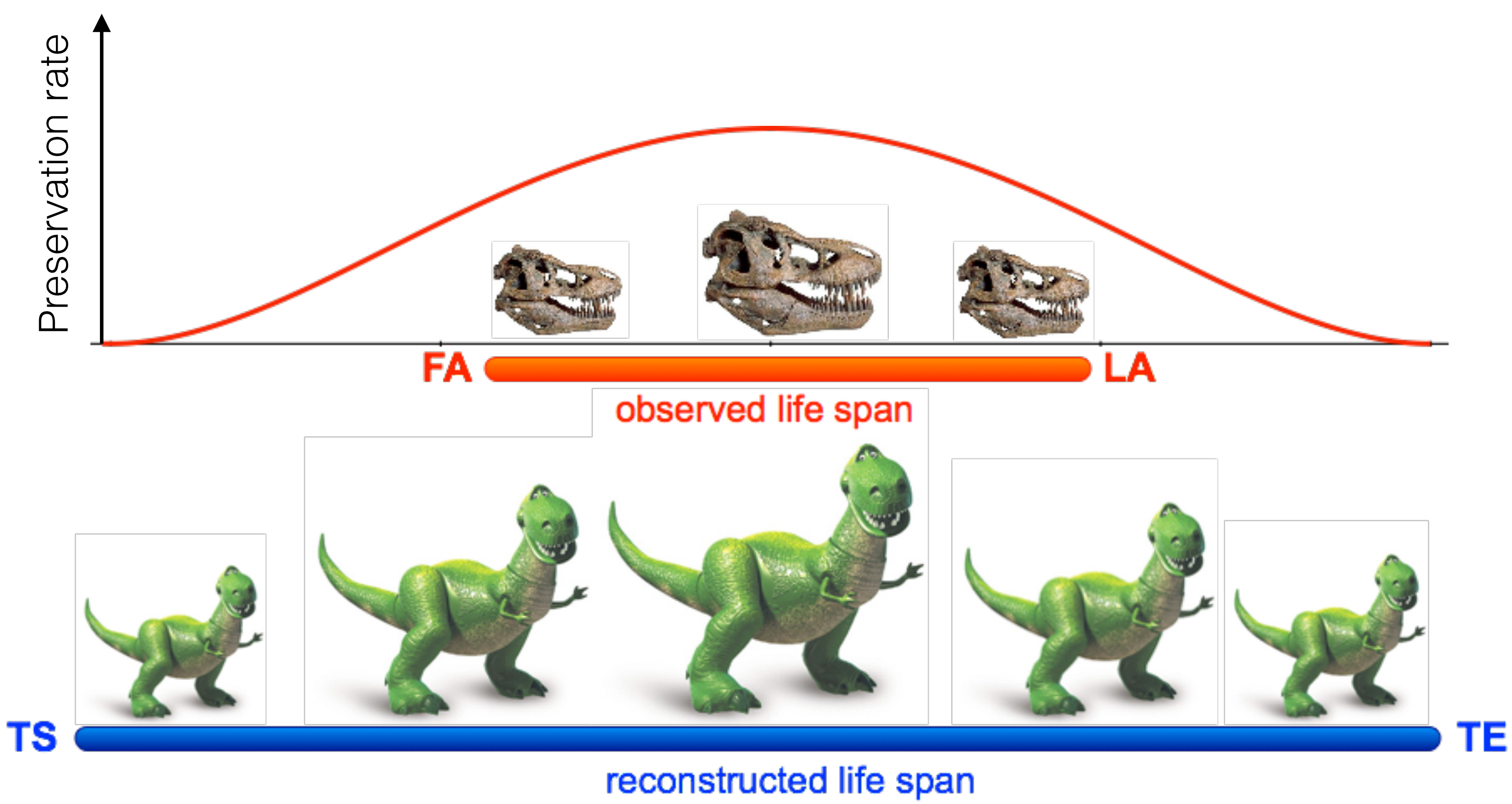


Constant preservation rate

Estimated preservation rates and times of speciation and extinction

$$P(s, e, q, \lambda, \mu | x) \propto \underbrace{P(x | s, e, q)}_{\text{LIKELIHOOD}} \times P(s, e | \lambda, \mu) \times P(q) P(\lambda, \mu)$$

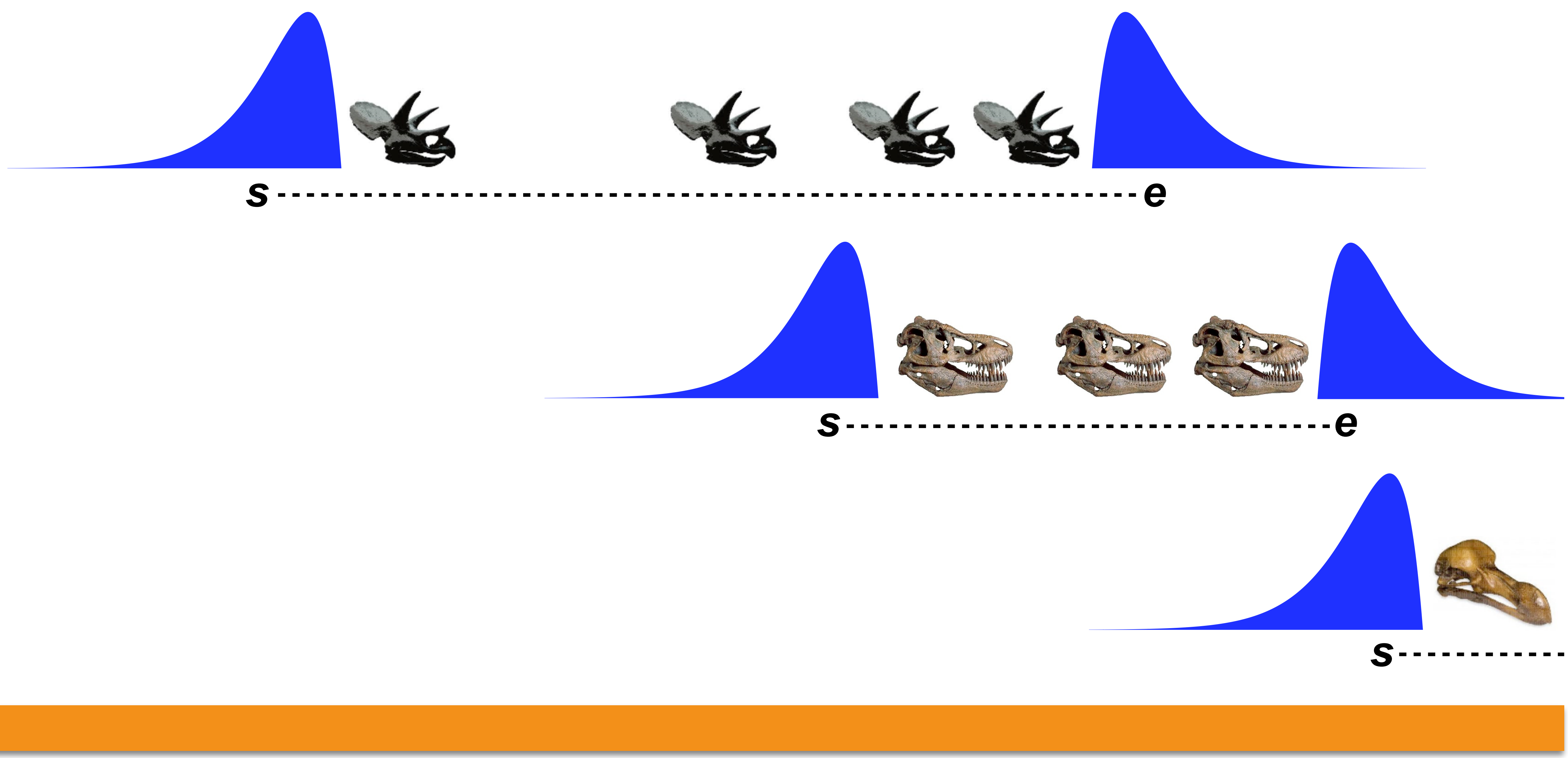
Non-homogeneous Poisson process (NHPP)



Estimated preservation rates and times of speciation and extinction

$$P(s, e, q, \lambda, \mu|x) \propto \underbrace{\boxed{P(x|s, e, q)}}_{\text{LIKELIHOOD}} \times P(s, e|\lambda, \mu) \times P(q)P(\lambda, \mu)$$

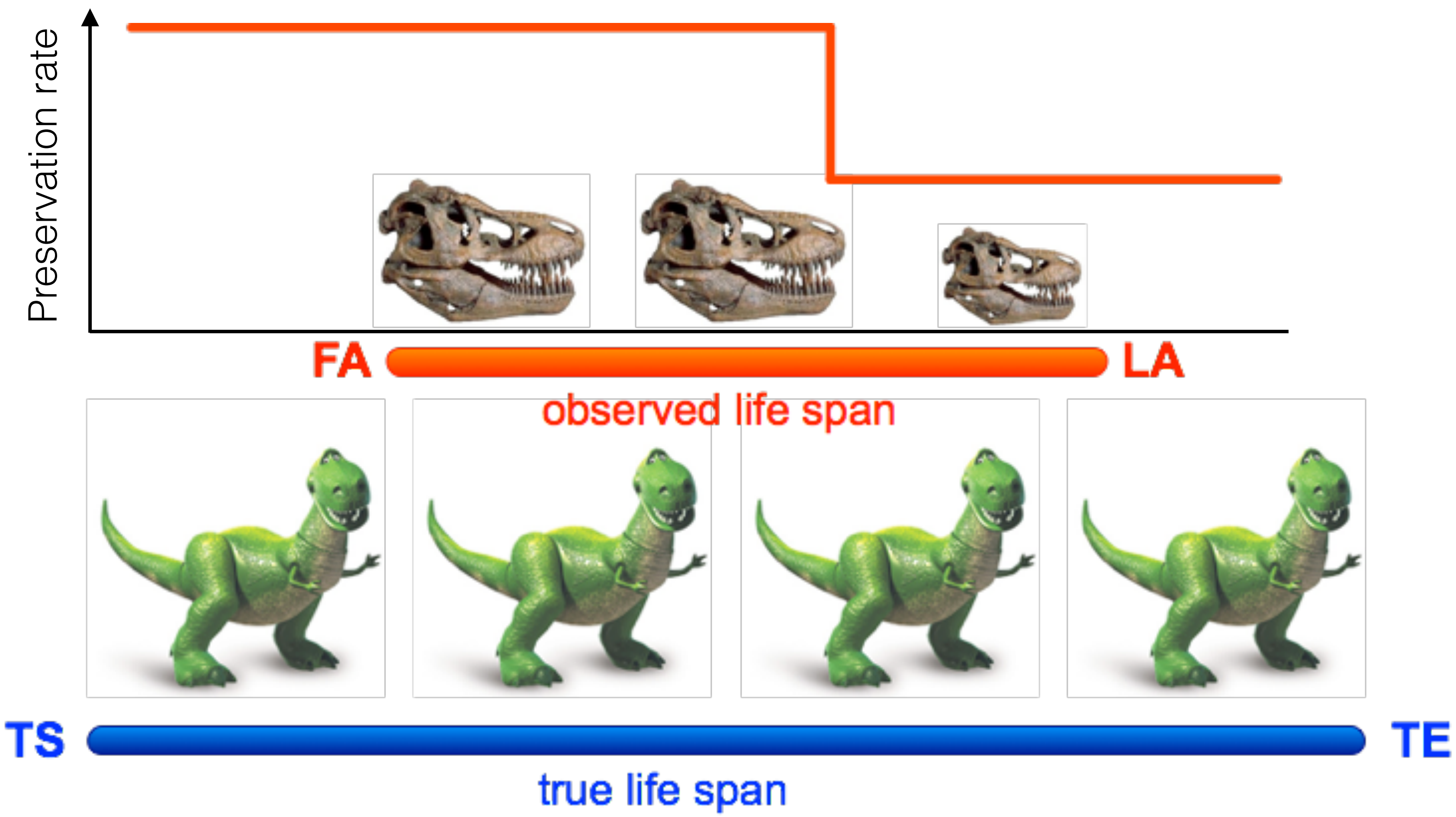
Non-homogeneous Poisson process (NHPP)



Constant mean preservation rate

$$P(s, e, q, \lambda, \mu | x) \propto \underbrace{P(x | s, e, q)}_{\text{LIKELIHOOD}} \times P(s, e | \lambda, \mu) \times P(q) P(\lambda, \mu)$$

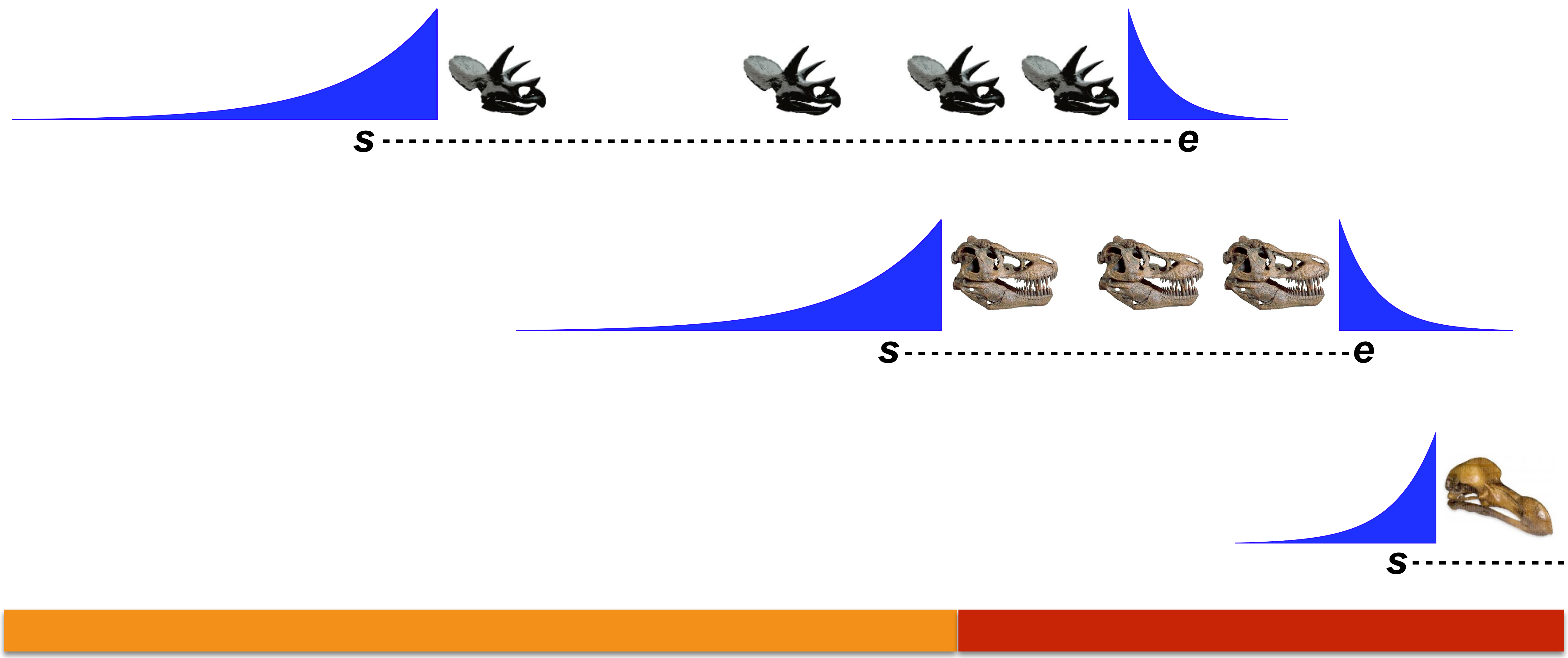
Time-variable Poisson process (TPP)



Estimated preservation rates and times of speciation and extinction

$$P(s,e,q,\lambda,\mu|x) \propto \underbrace{\boxed{P(x|s,e,q)}}_{\text{LIKELIHOOD}} \times P(s,e|\lambda,\mu) \times P(q)P(\lambda,\mu)$$

Time-variable Poisson process (TPP)



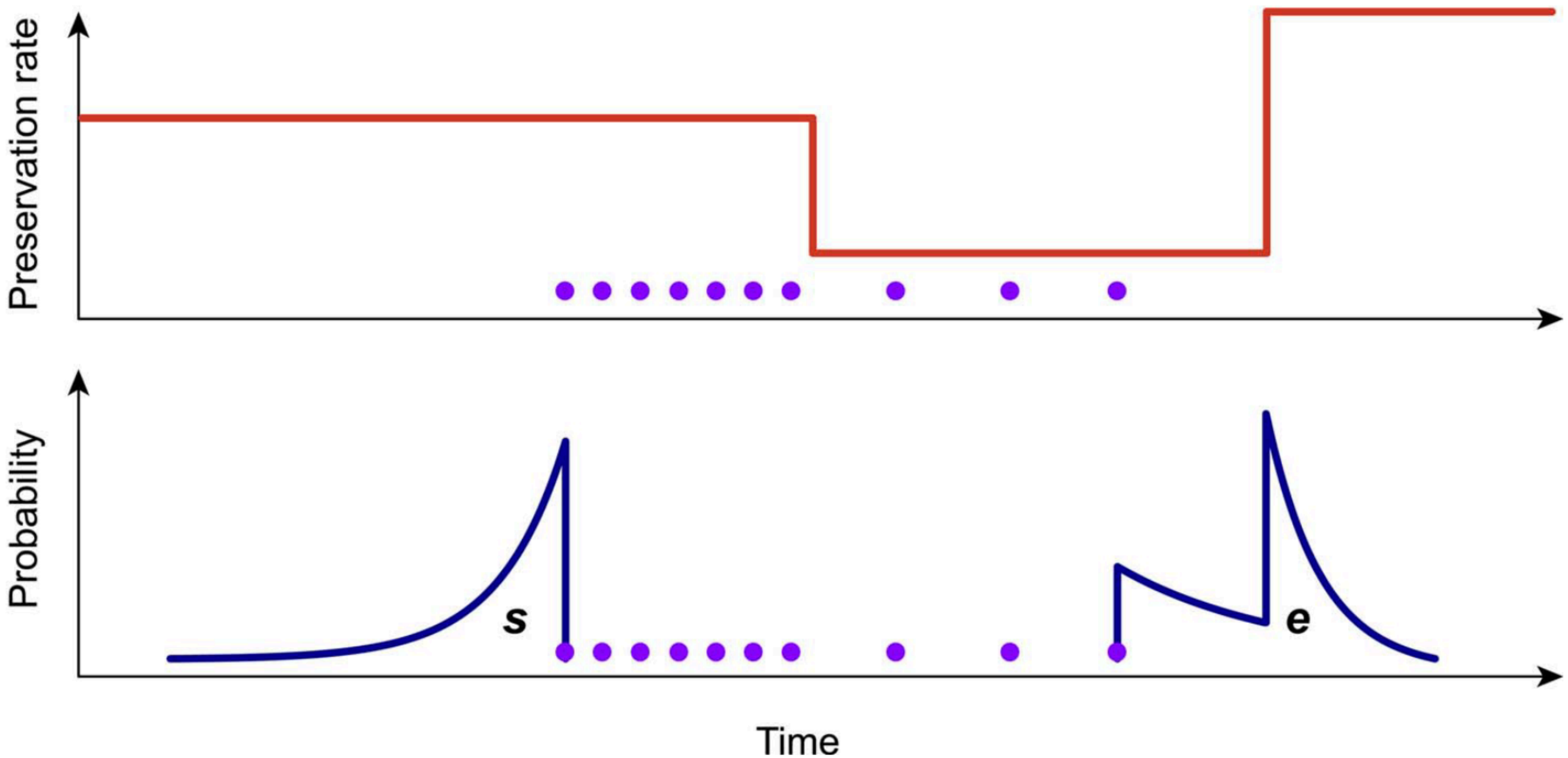
Low preservation rate

High preservation rate

Estimated preservation rates and times of speciation and extinction

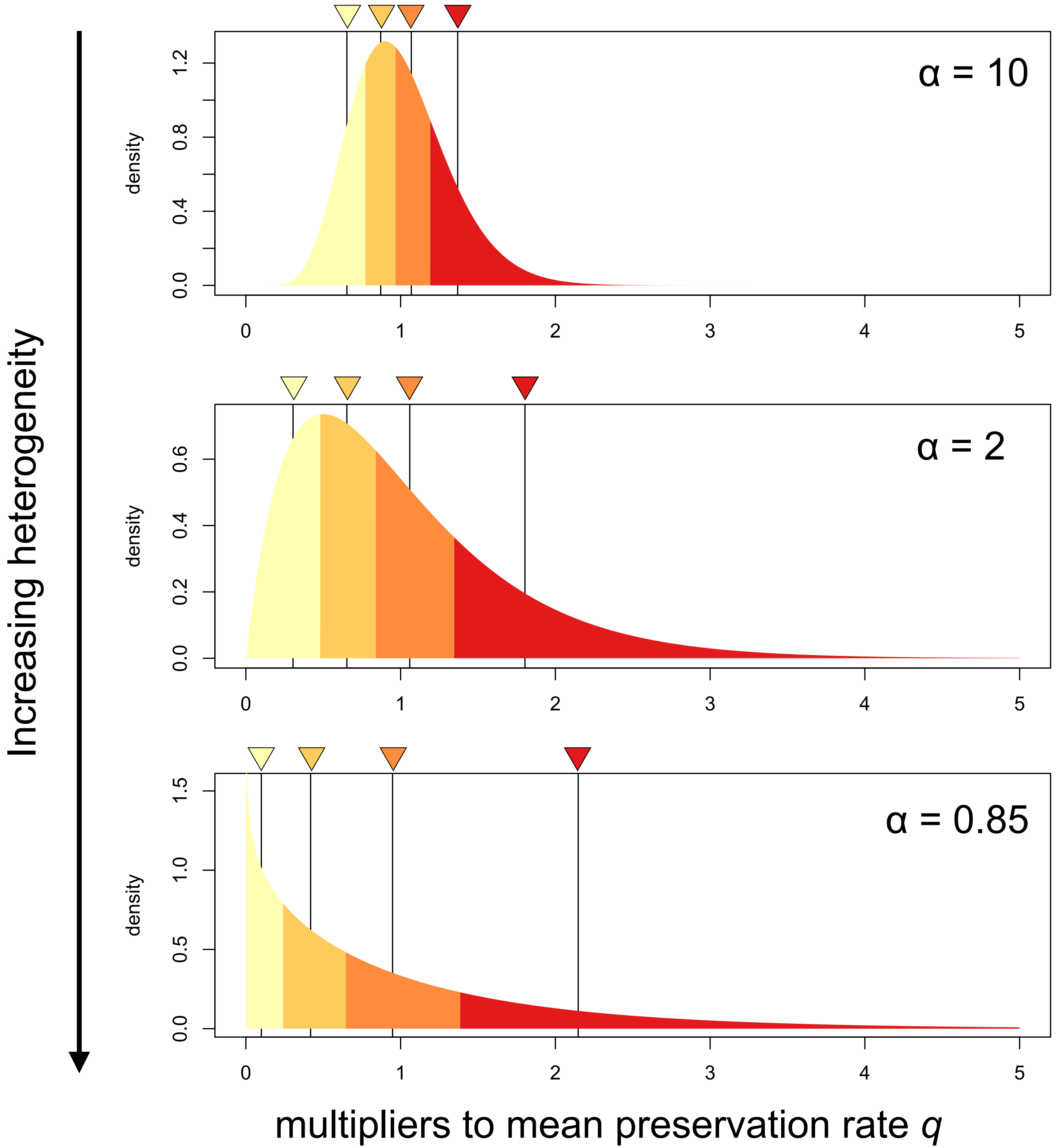
$$P(s, e, q, \lambda, \mu | x) \propto \underbrace{P(x | s, e, q)}_{\text{LIKELIHOOD}} \times P(s, e | \lambda, \mu) \times P(q)P(\lambda, \mu)$$

Time-variable Poisson process (TPP)

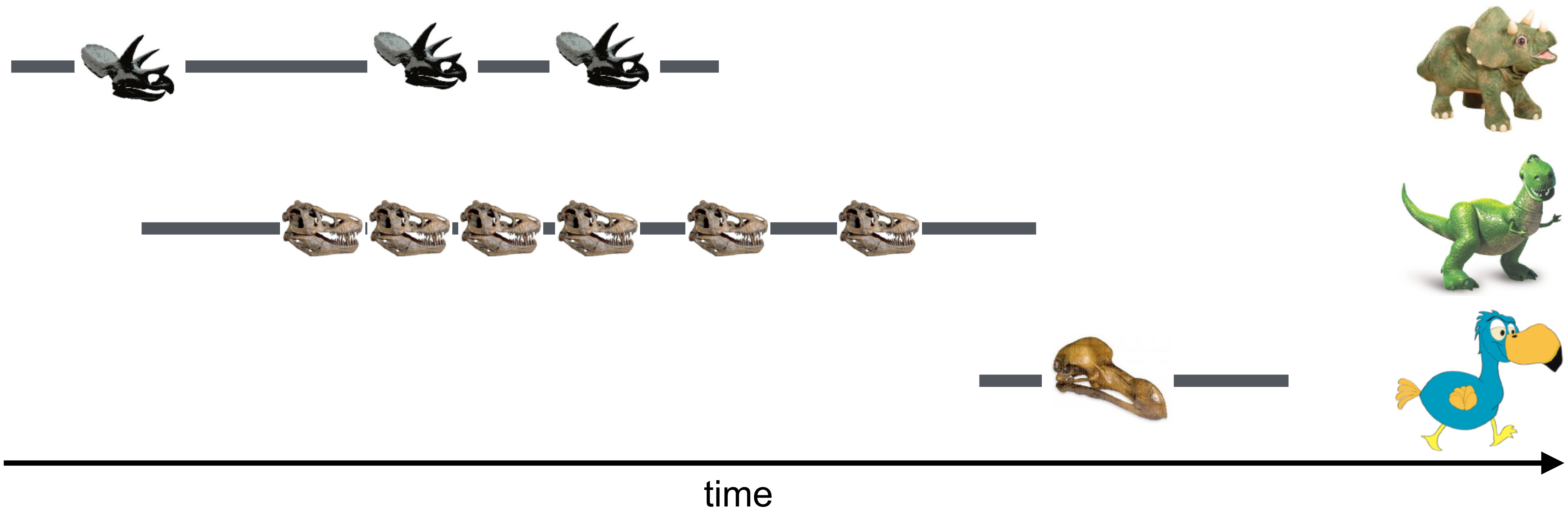


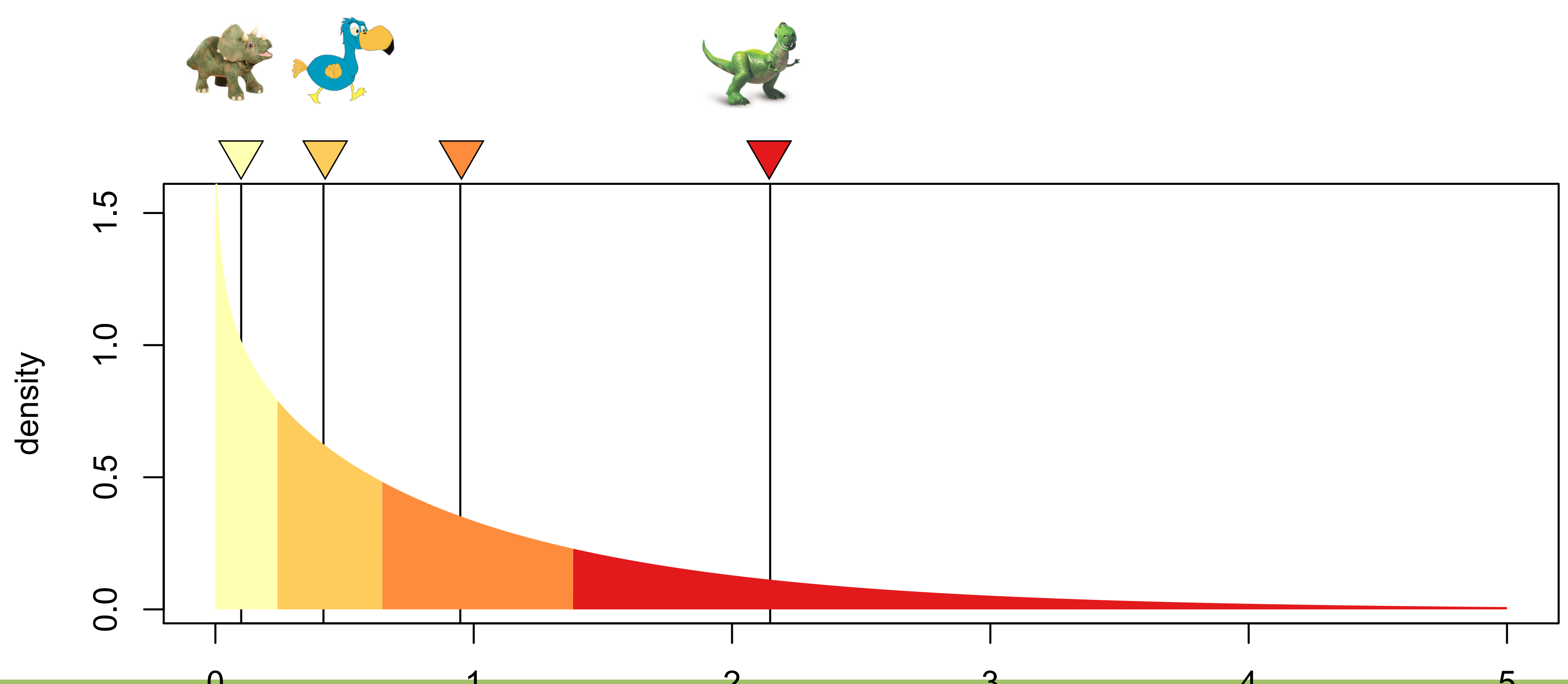
Silvestro et al. 2019 Paleobiology

Gamma model of rate heterogeneity

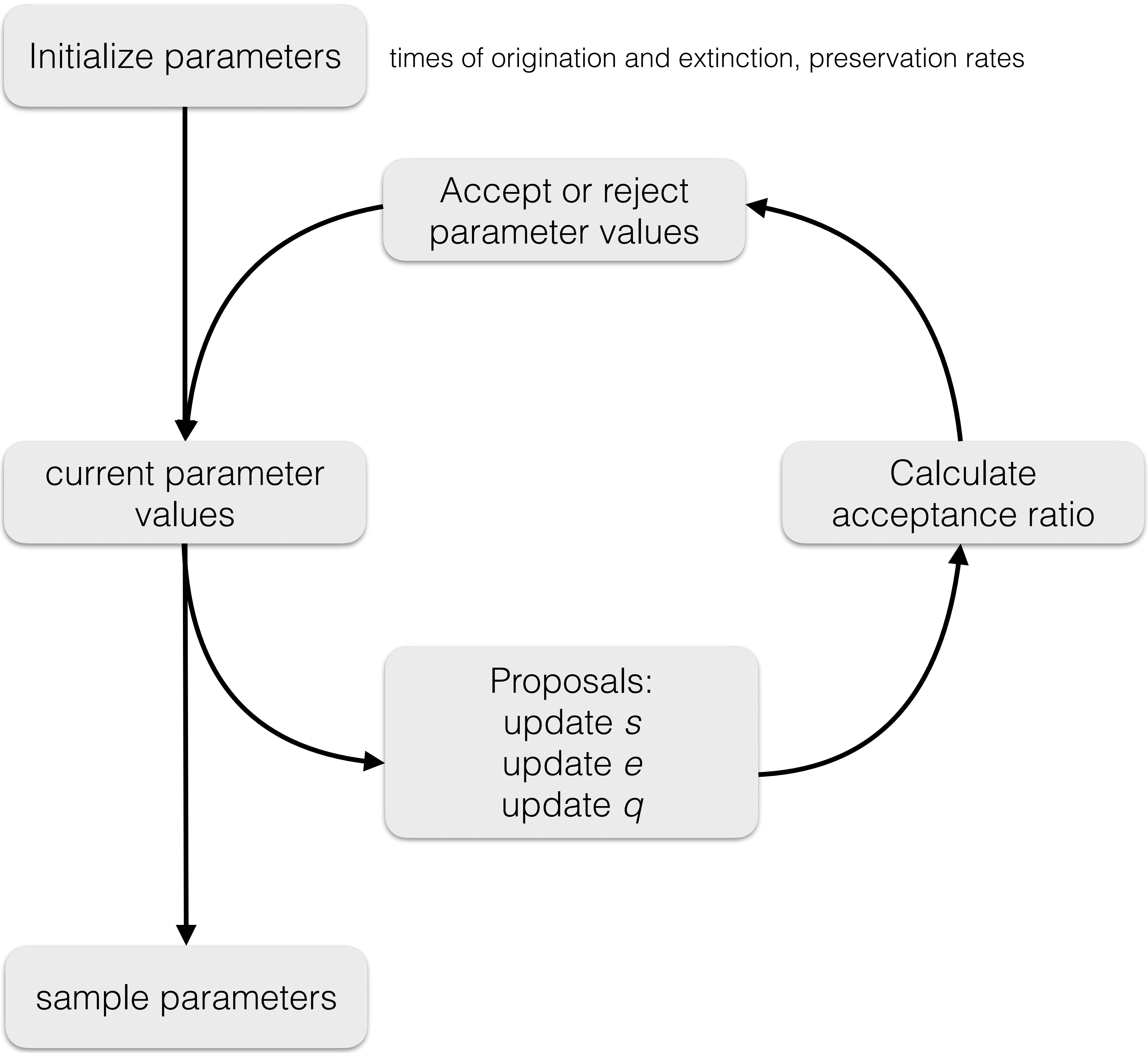


Poisson processes of preservation



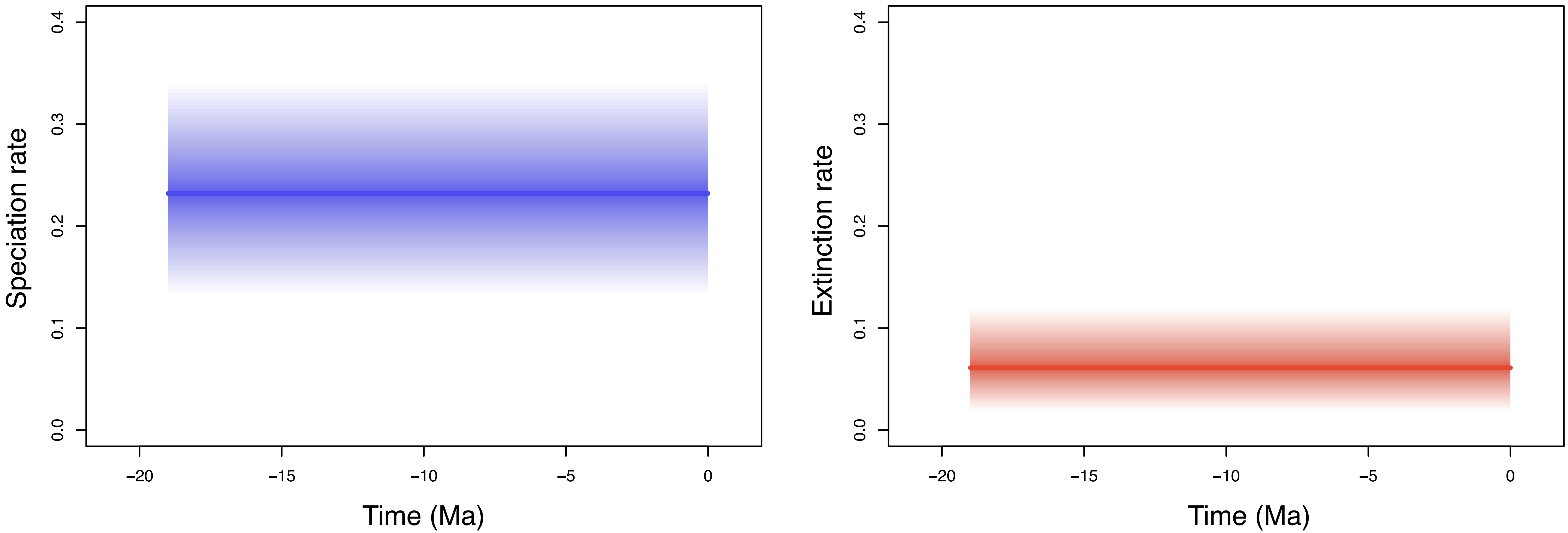


Estimating times of origination, extinction and preservation – MCMC



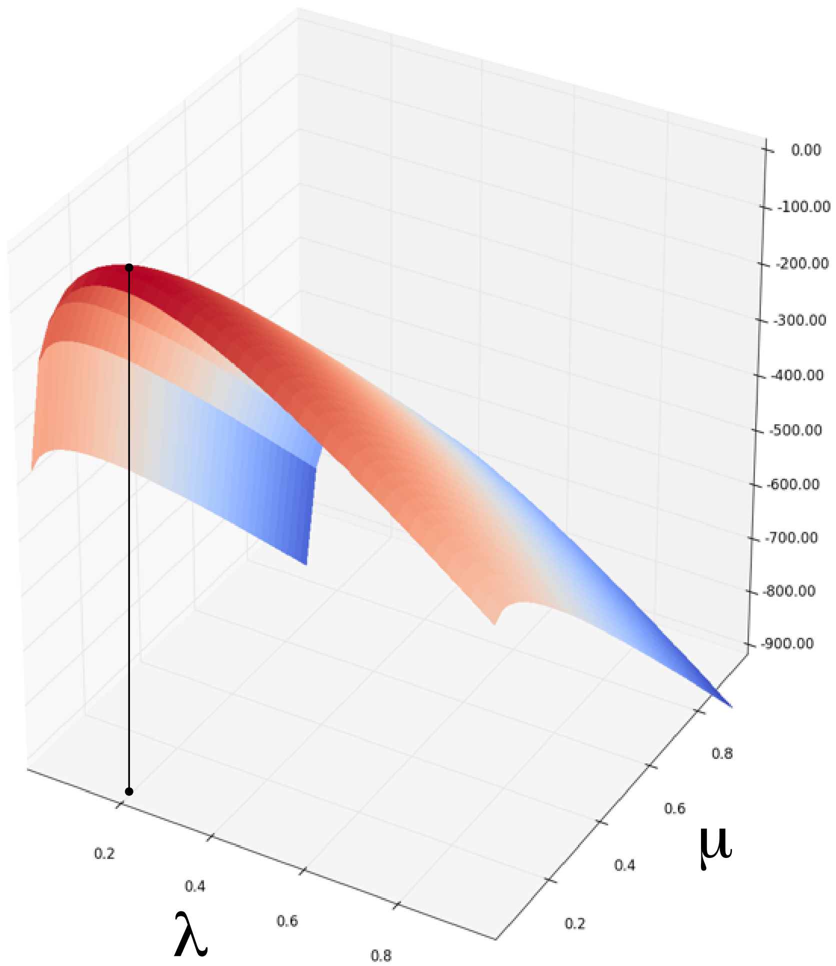
Birth-death models:
exploratory ("hypothesis-free") algorithms

$$\underbrace{P(q, \mathbf{s}, \mathbf{e}, \lambda, \mu | X)}_{\text{posterior}} \propto \underbrace{P(X | q, \mathbf{s}, \mathbf{e})}_{\text{likelihood}} \times \underbrace{P(\mathbf{s}, \mathbf{e} | \lambda, \mu)}_{\text{birth-death prior}} \times \underbrace{P(q) P(\lambda, \mu)}_{\text{other (hyper)priors}}$$

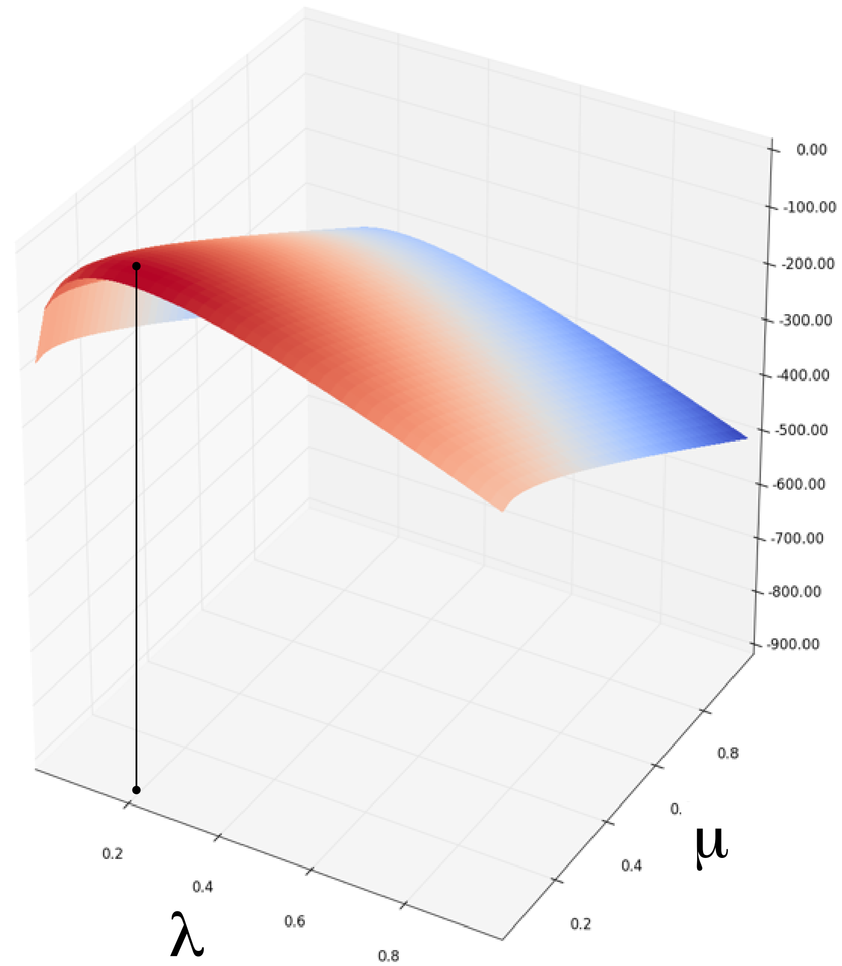


Robustness to incomplete taxon sampling

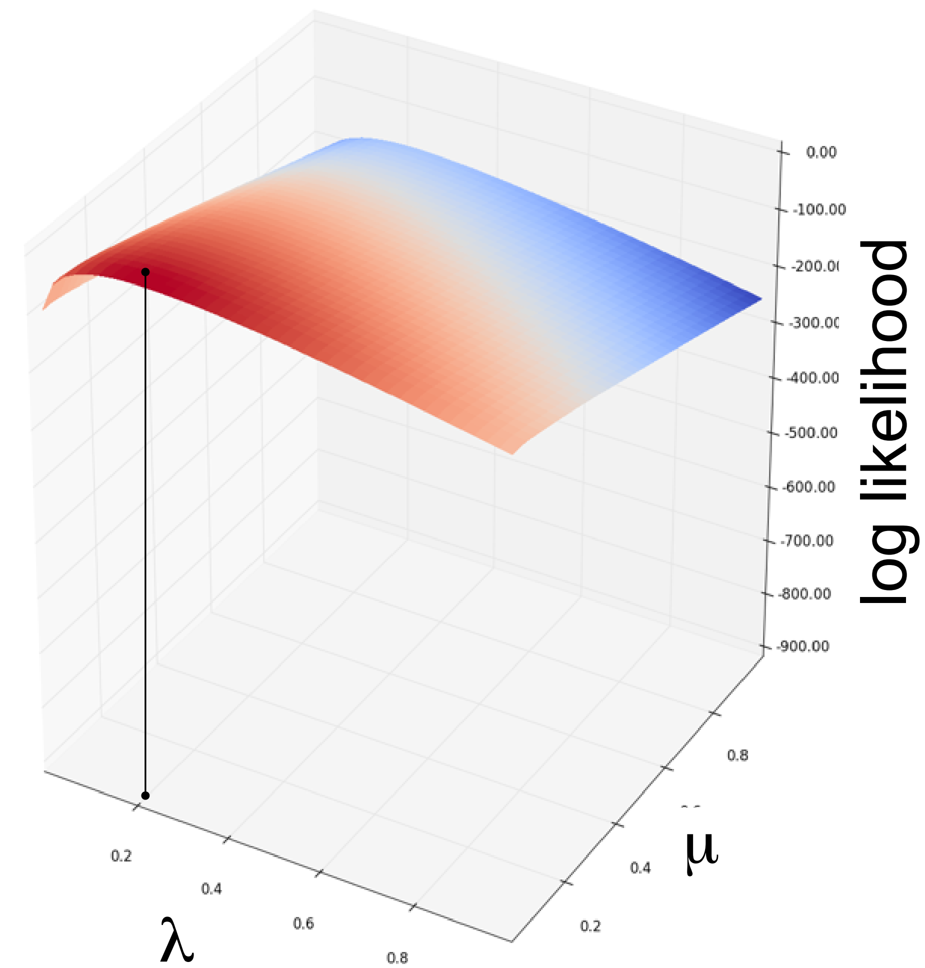
sampling: 100%



sampling: 50%



sampling: 25%

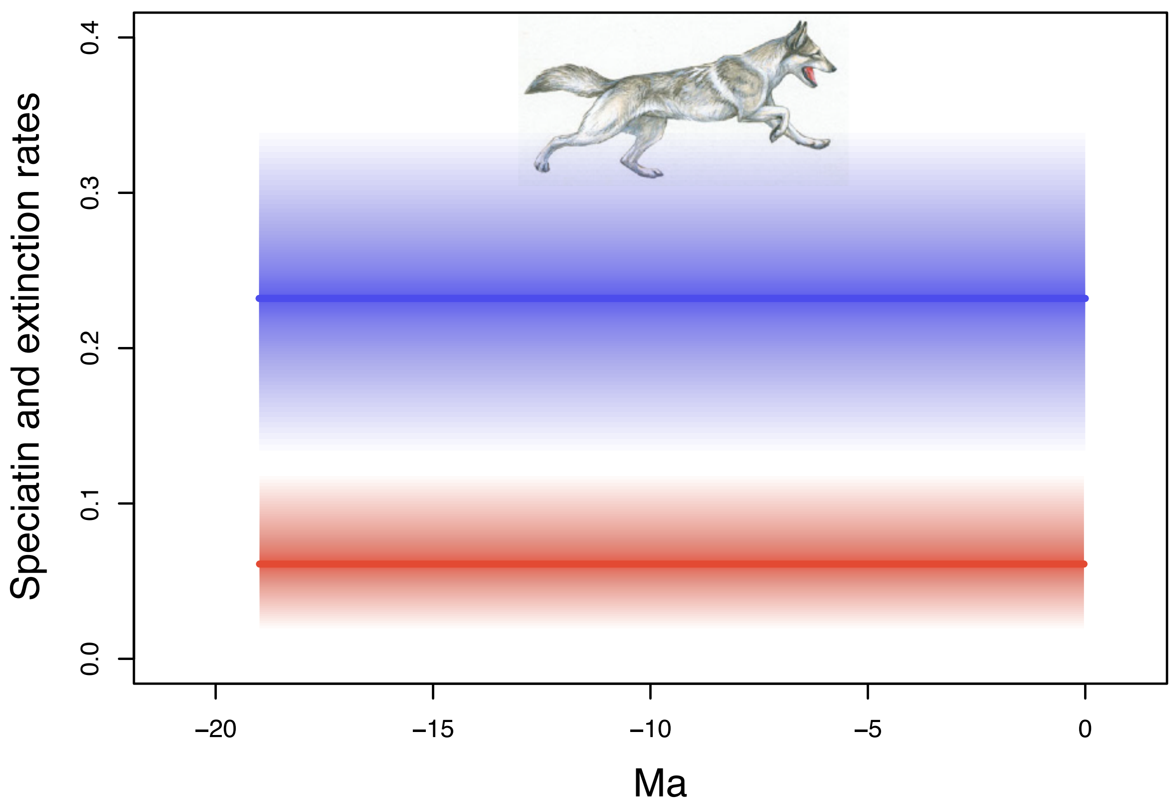


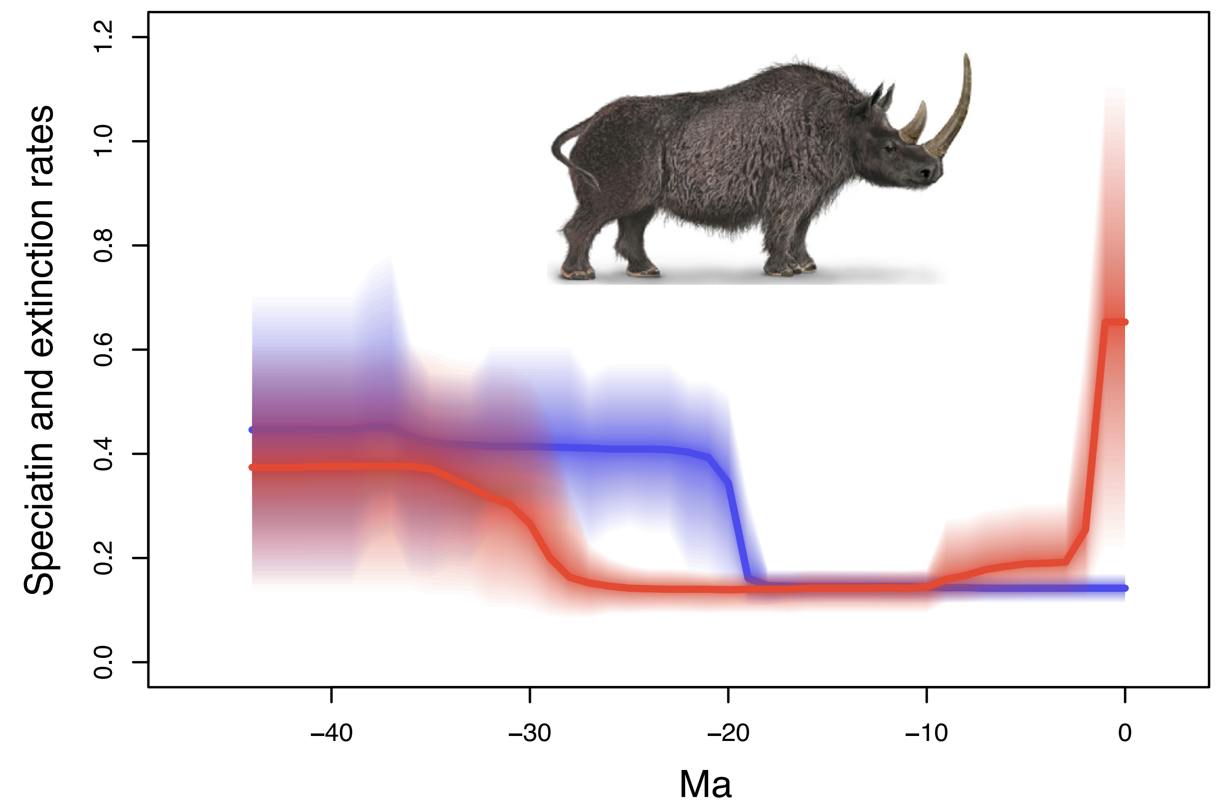
Random taxon sampling does not bias the estimation of speciation and extinction rates

Birth-death models: exploratory ("hypothesis-free") algorithms

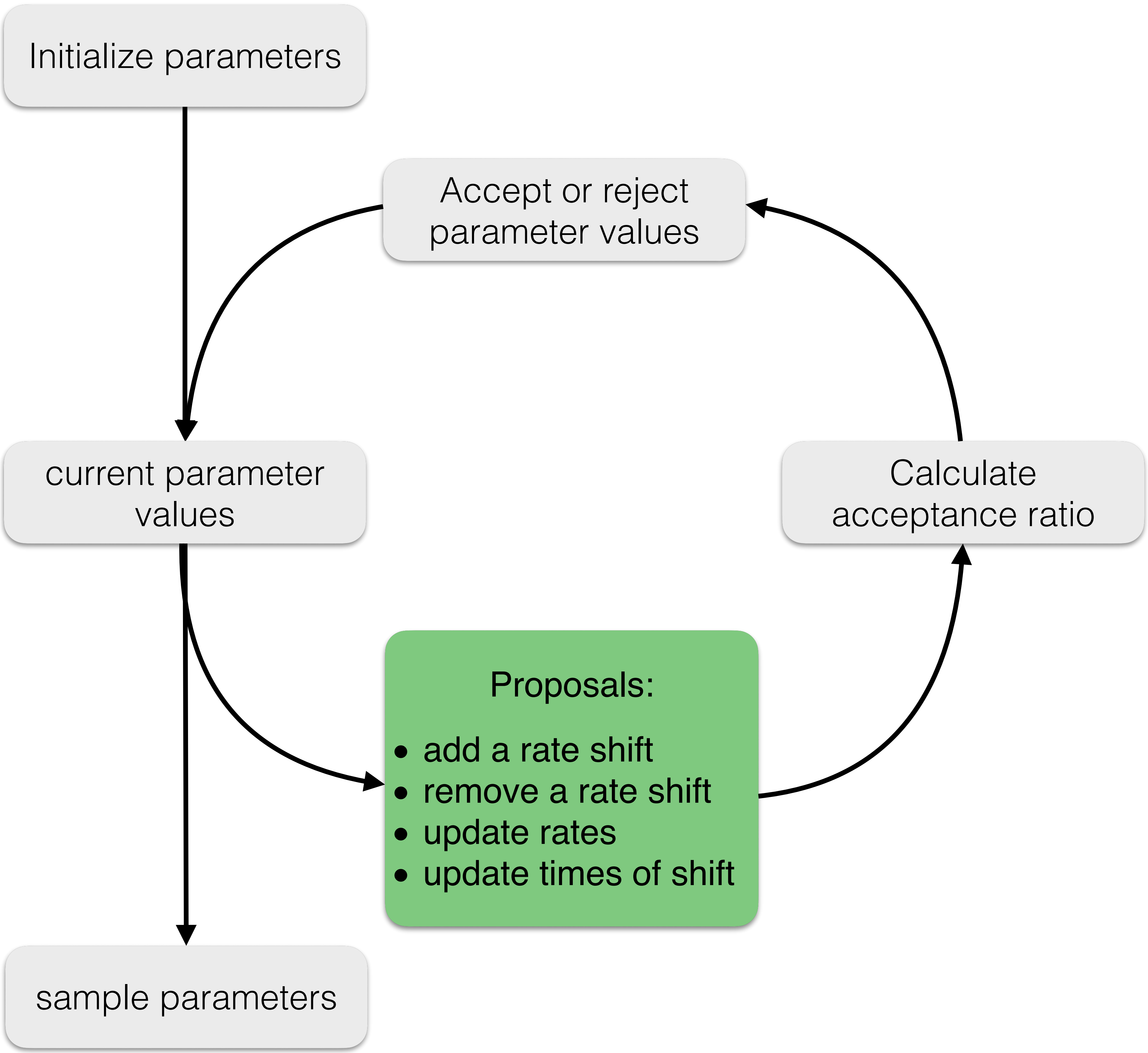
$$\underbrace{P(q,\mathbf{s},\mathbf{e},\lambda,\mu|X)}_{\text{posterior}} \propto \underbrace{P(X|q,\mathbf{s},\mathbf{e})}_{\text{likelihood}} \times \underbrace{P(\mathbf{s},\mathbf{e}|\lambda,\mu)}_{\text{birth-death prior}} \times \underbrace{P(q)P(\lambda,\mu)}_{\text{other (hyper)priors}}$$

Reversible-jump MCMC

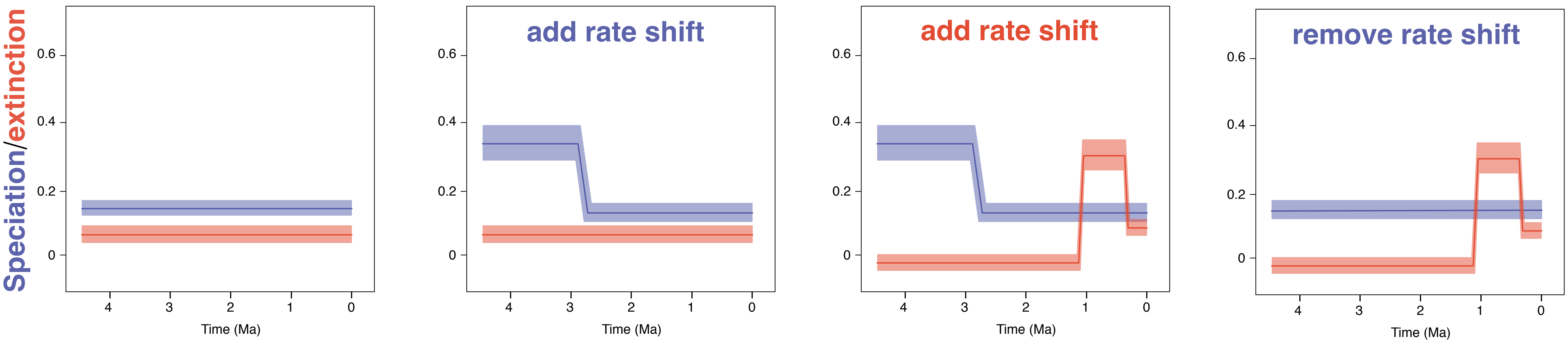




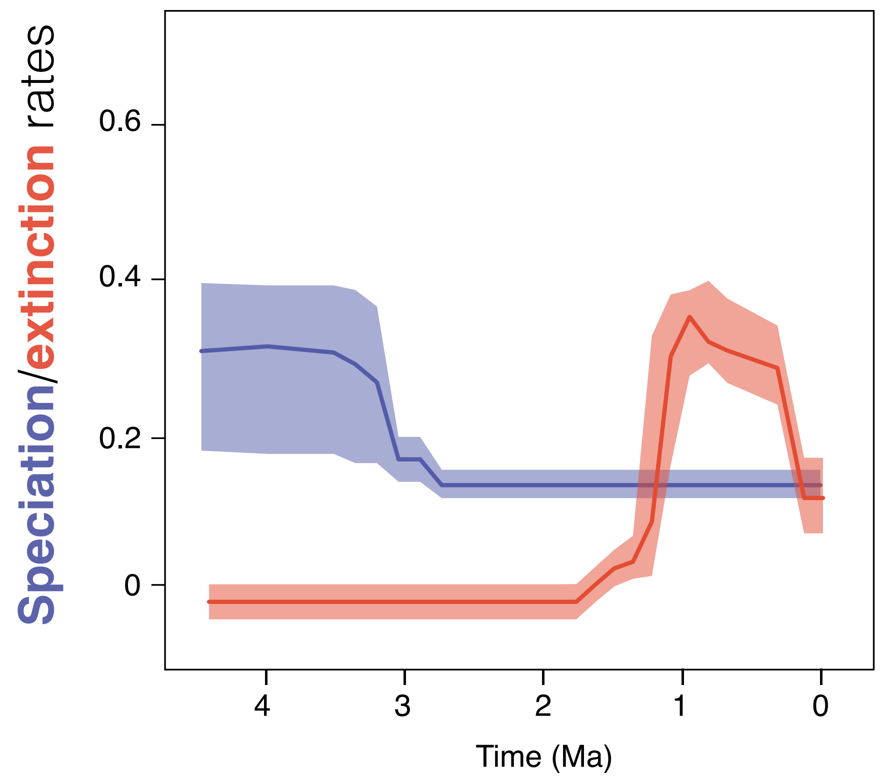
Estimating rate heterogeneity through time – RJMCMC



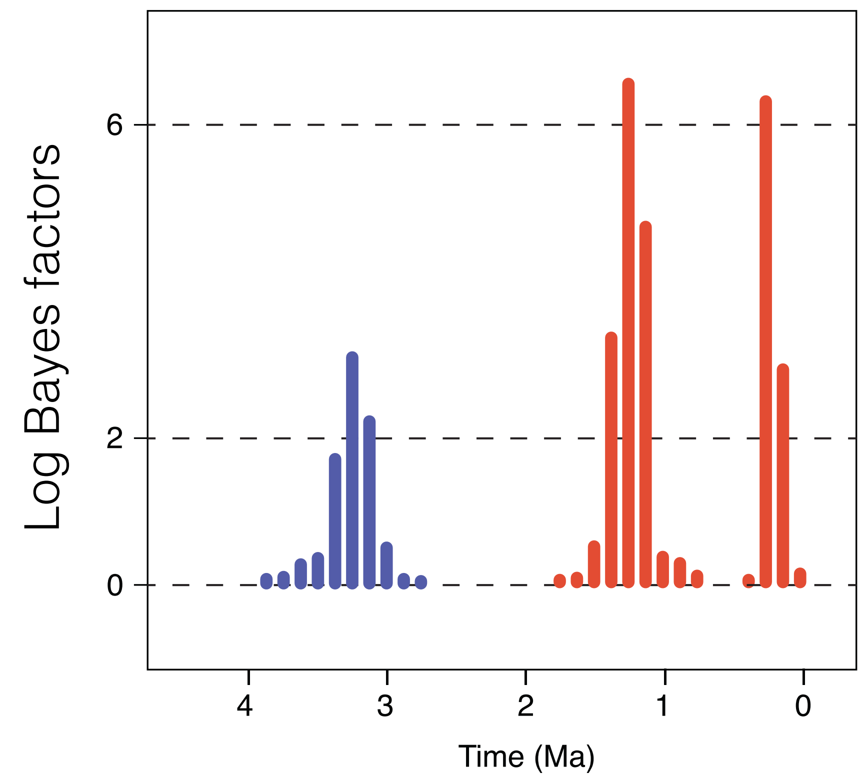
Estimating rate heterogeneity through time – RJMCMC



Marginal rates

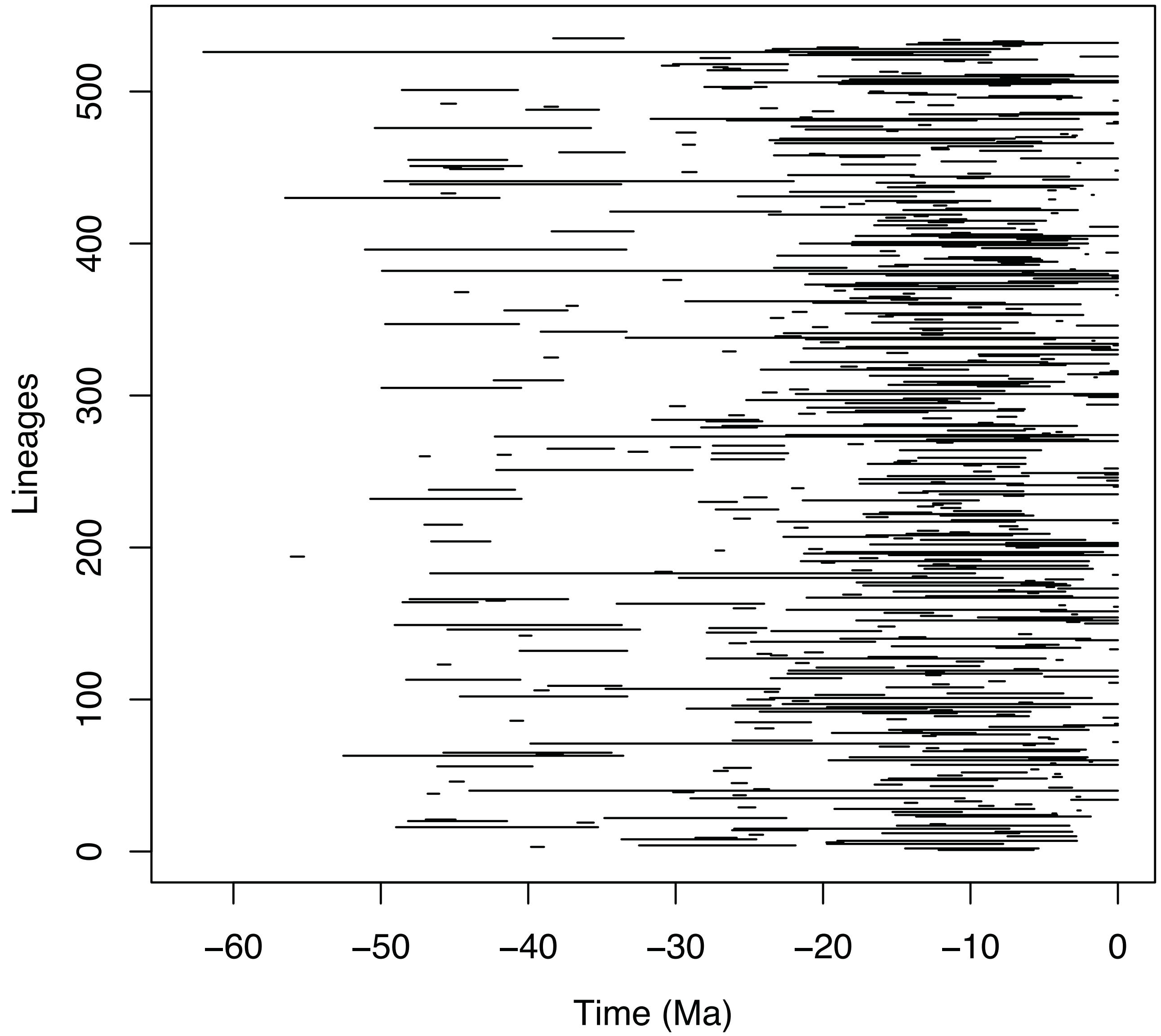


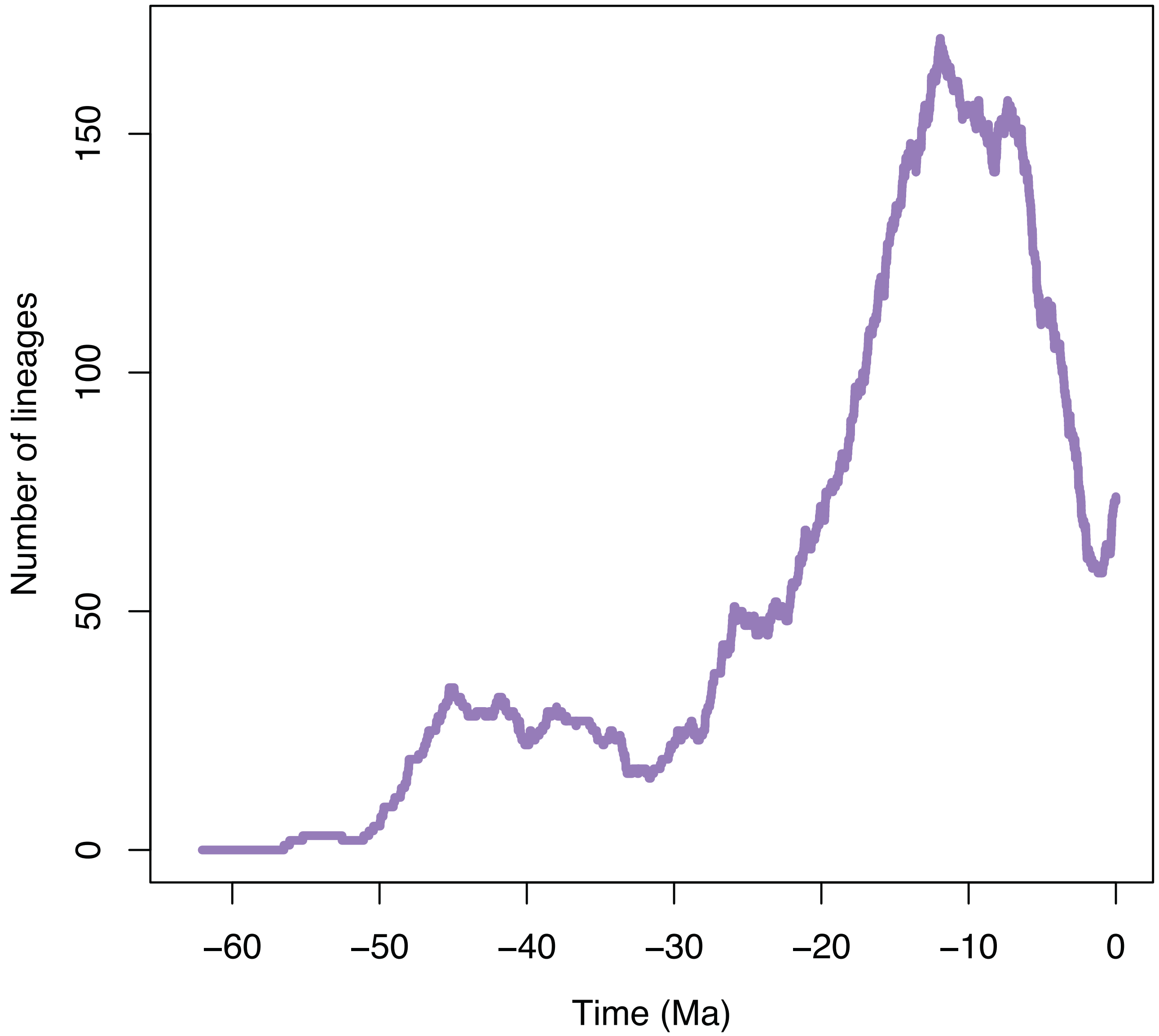
Times of rate shifts



Silvestro et al. 2019 Paleobiology

Diversification and extinction of marine mammals

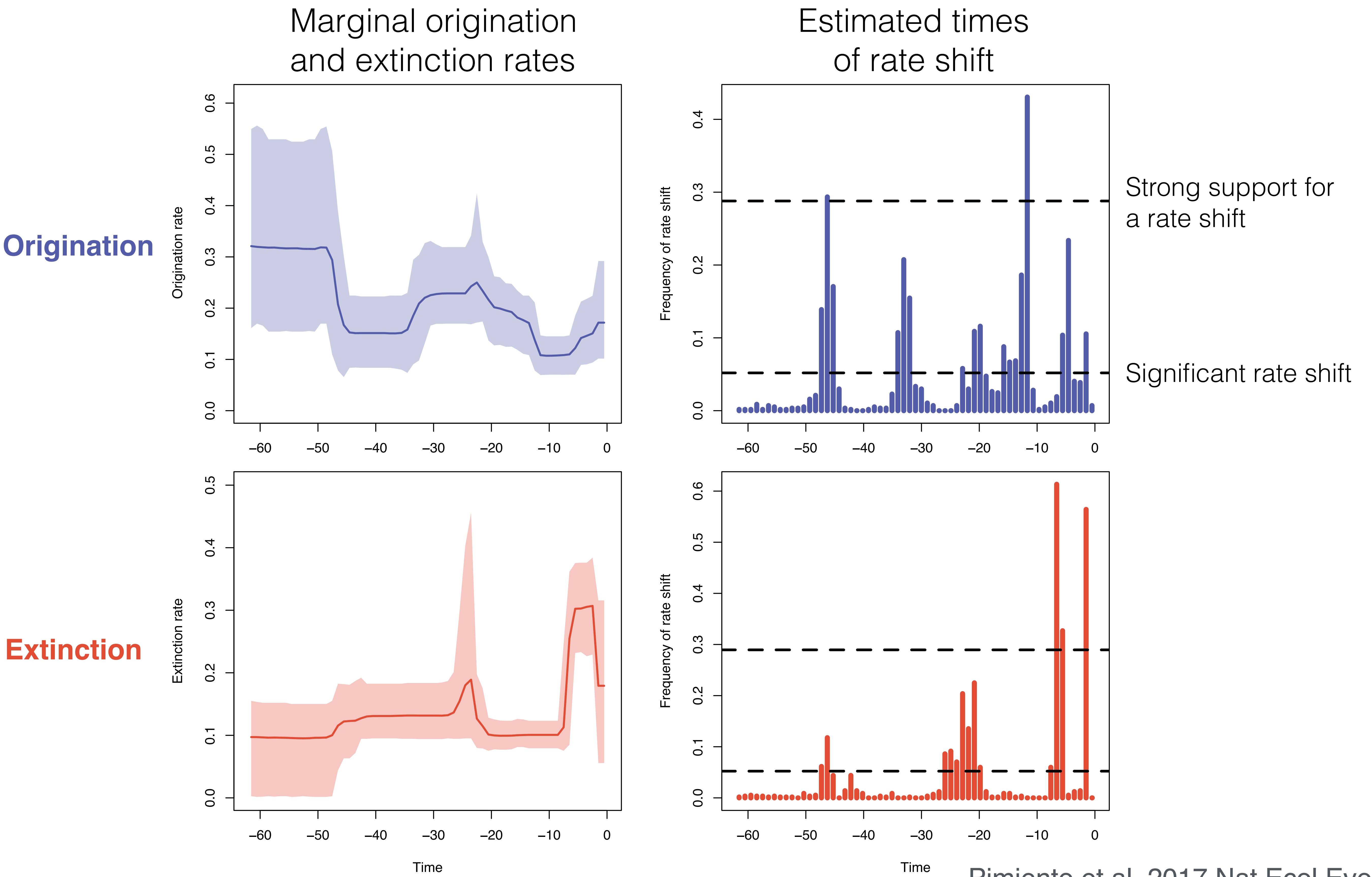




C Pimiento

Pimiento et al. 2017 Nat Ecol Evol

Diversification and extinction of marine mammals



Pimiento et al. 2017 Nat Ecol Evol

Hyper-prior distributions on the rate parameters

$$P(s, e, q, \lambda, \mu | x) \propto P(x | s, e, q) \times P(s, e | \lambda, \mu) \times \boxed{P(q)P(\lambda, \mu)}$$

POSTERIOR LIKELIHOOD PRIOR HYPERPRIORS

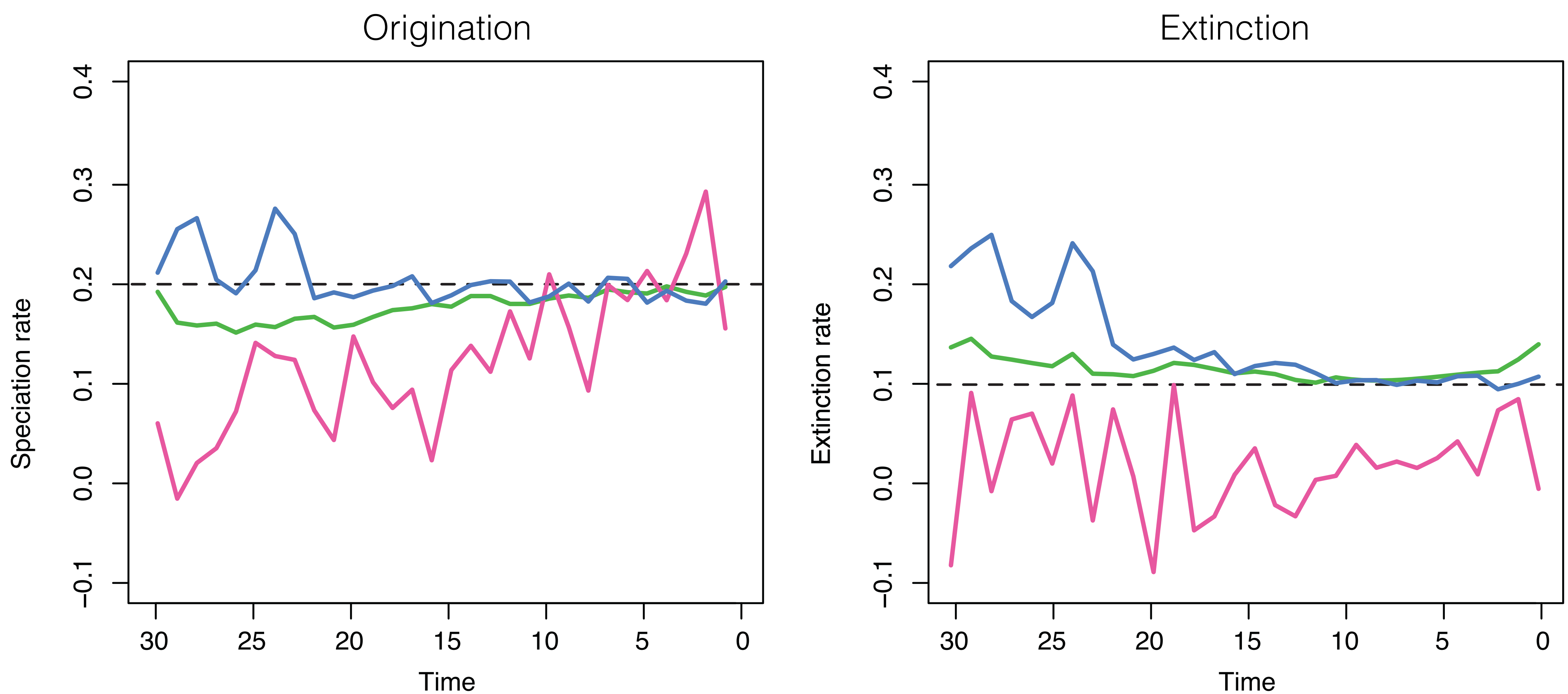
Gamma prior on preservation rate

Uniform prior on alpha (rate heterogeneity)

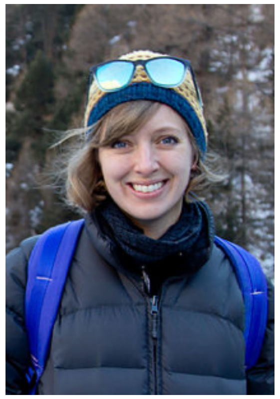
Gamma or half-Cauchy priors on BD rates

Effects of PyRate's regularization through RJMCMC

Constant birth-death simulation with constant, low preservation



Mean per capita (Foote 2000)
Three-timer (Alroy 2008)
CMR (Liow & Finarelli 2014)

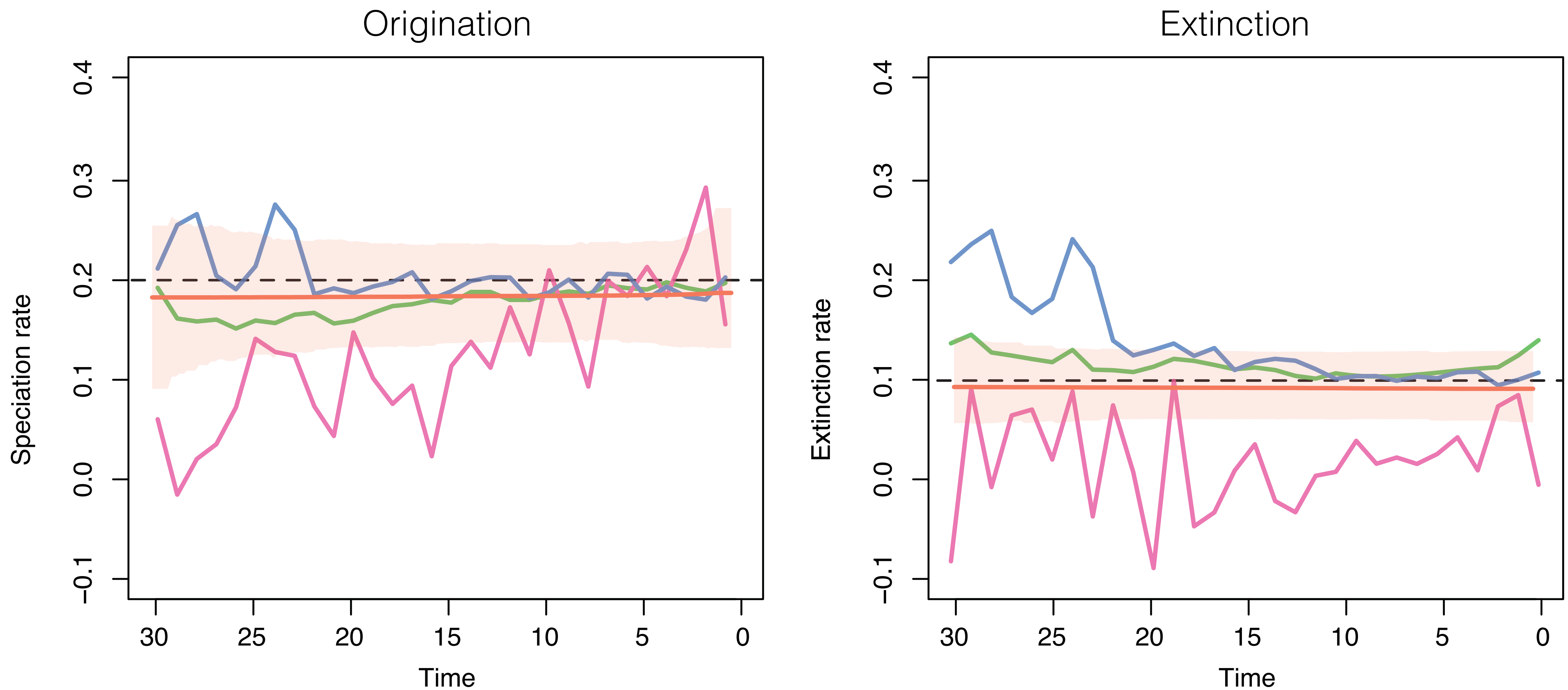


T M Smiley

Smiley 2018 Paleobiology

Effects of PyRate's regularization through RJMCMC

Constant birth-death simulation with constant, low preservation



Mean per capita (Foote 2000)

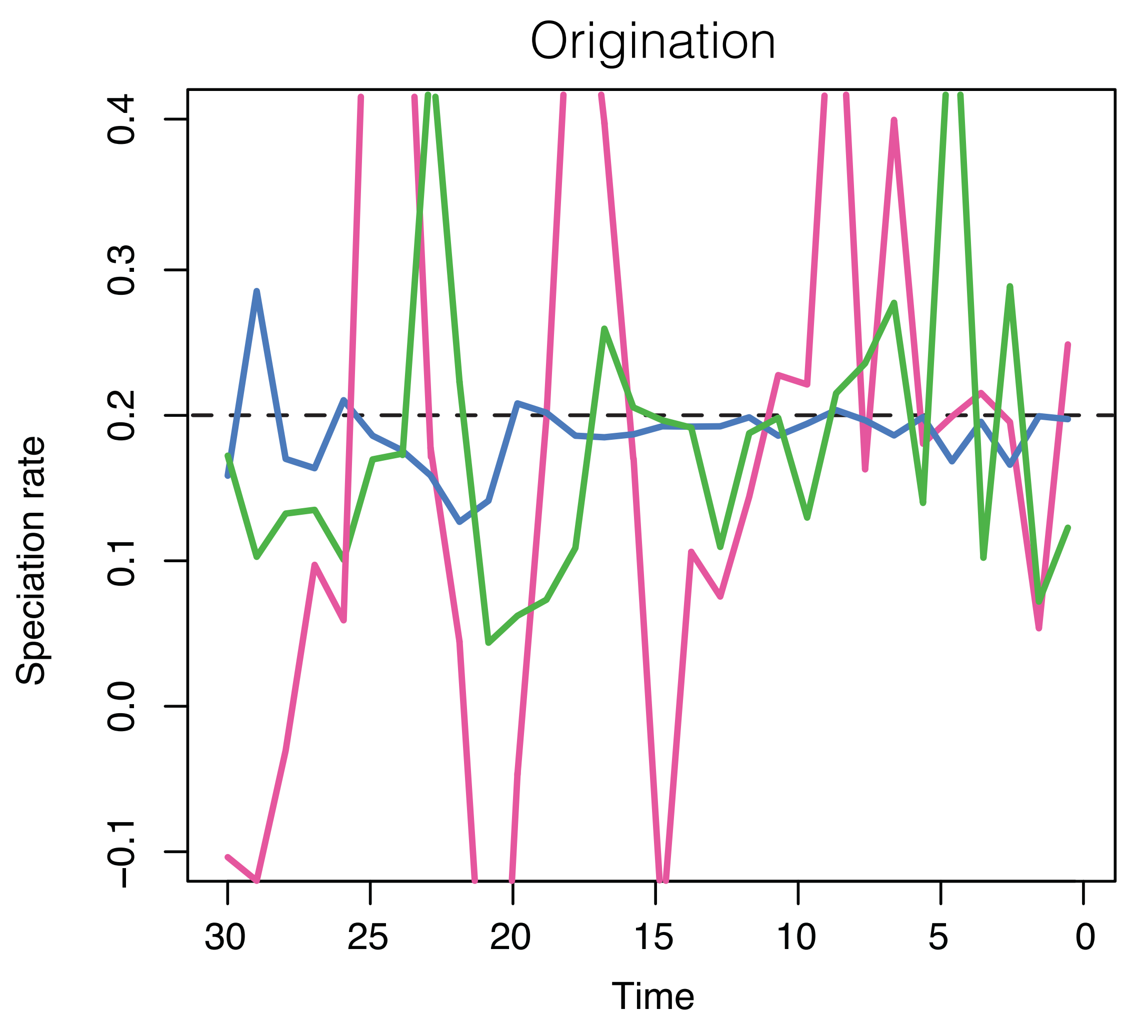
Three-timer (Alroy 2008)

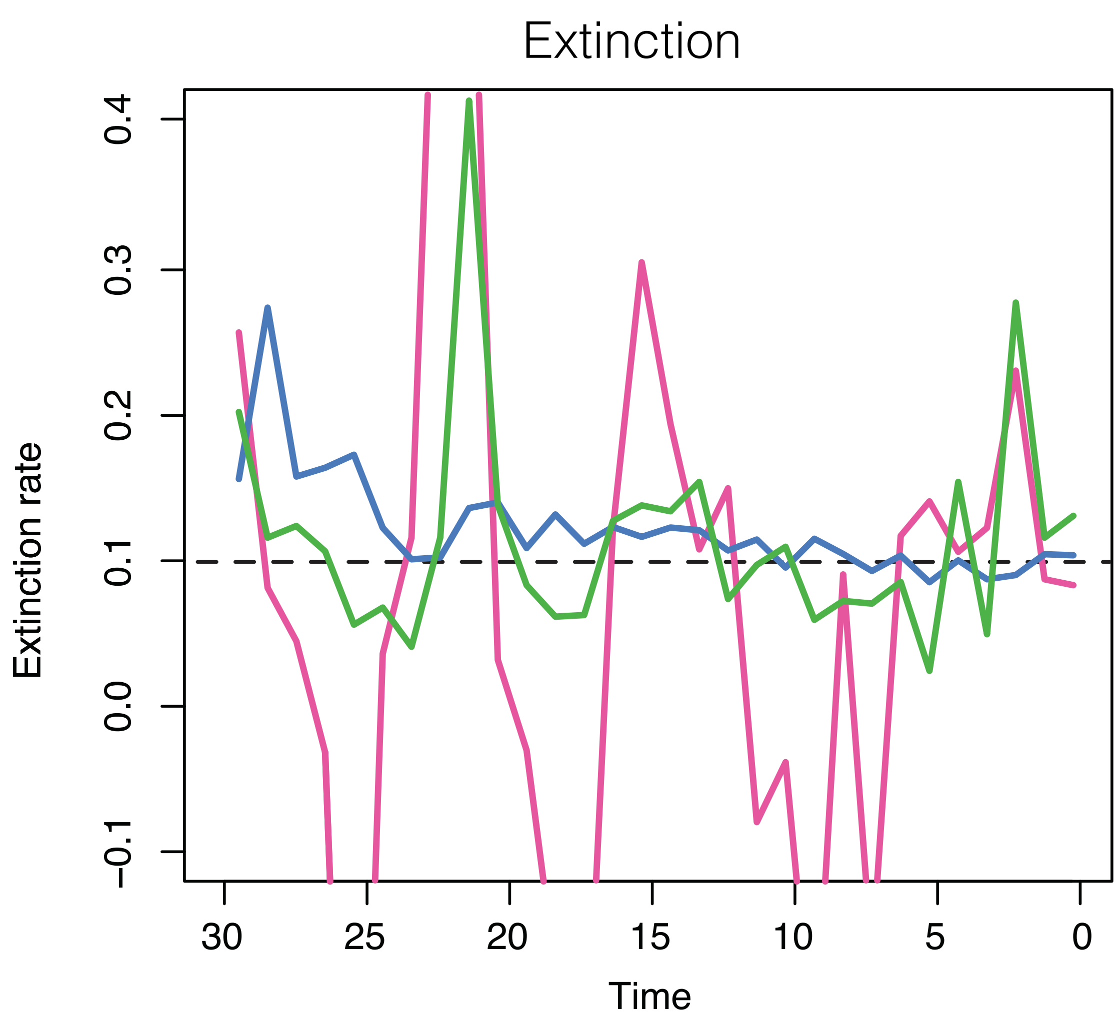
CMR (Liow & Finarelli 2014)

PyRate

Effects of PyRate's regularization through RJMCMC

Constant birth-death simulation with time-variable preservation





Mean per capita (Foote 2000)

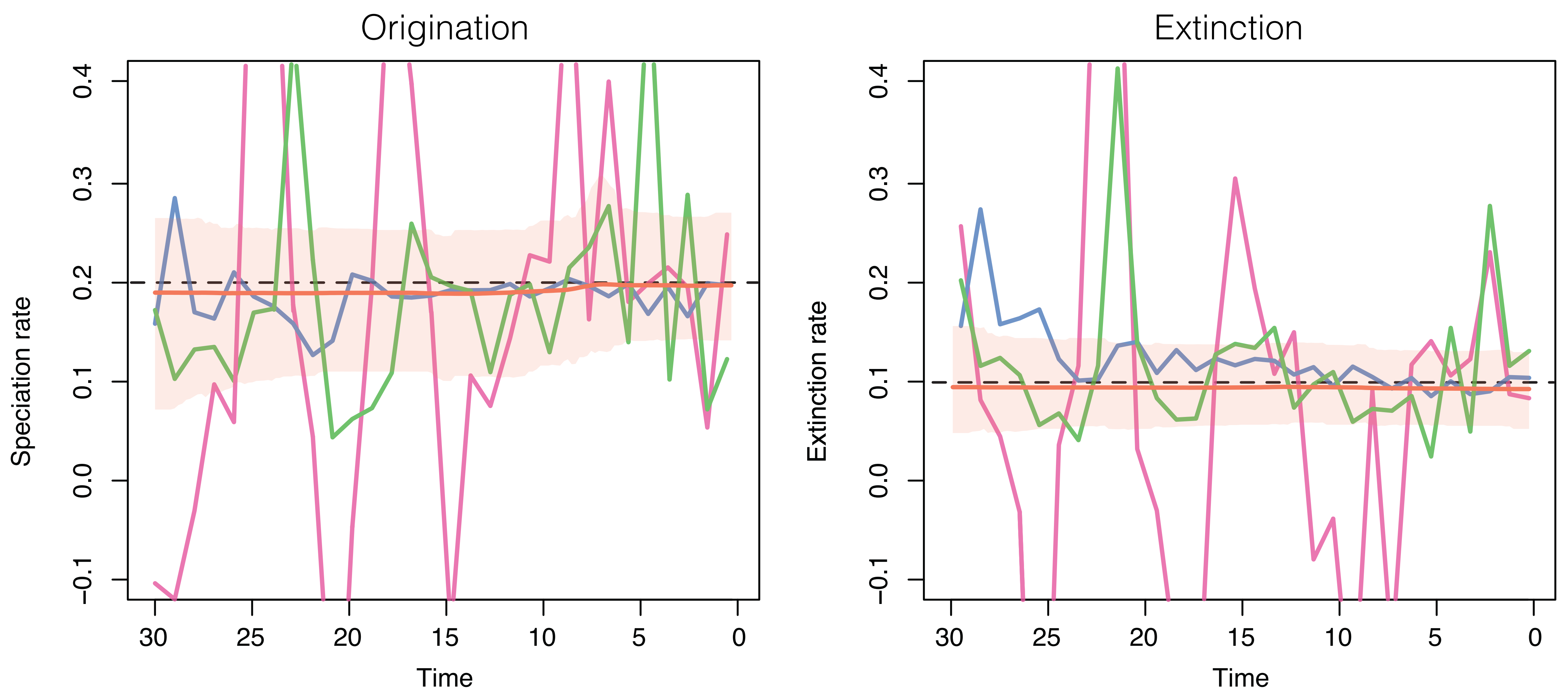
Three-timer (Alroy 2008)

CMR (Liow & Finarelli 2014)

Smiley 2018 Paleobiology

Effects of PyRate's regularization through RJMCMC

Constant birth-death simulation with time-variable preservation

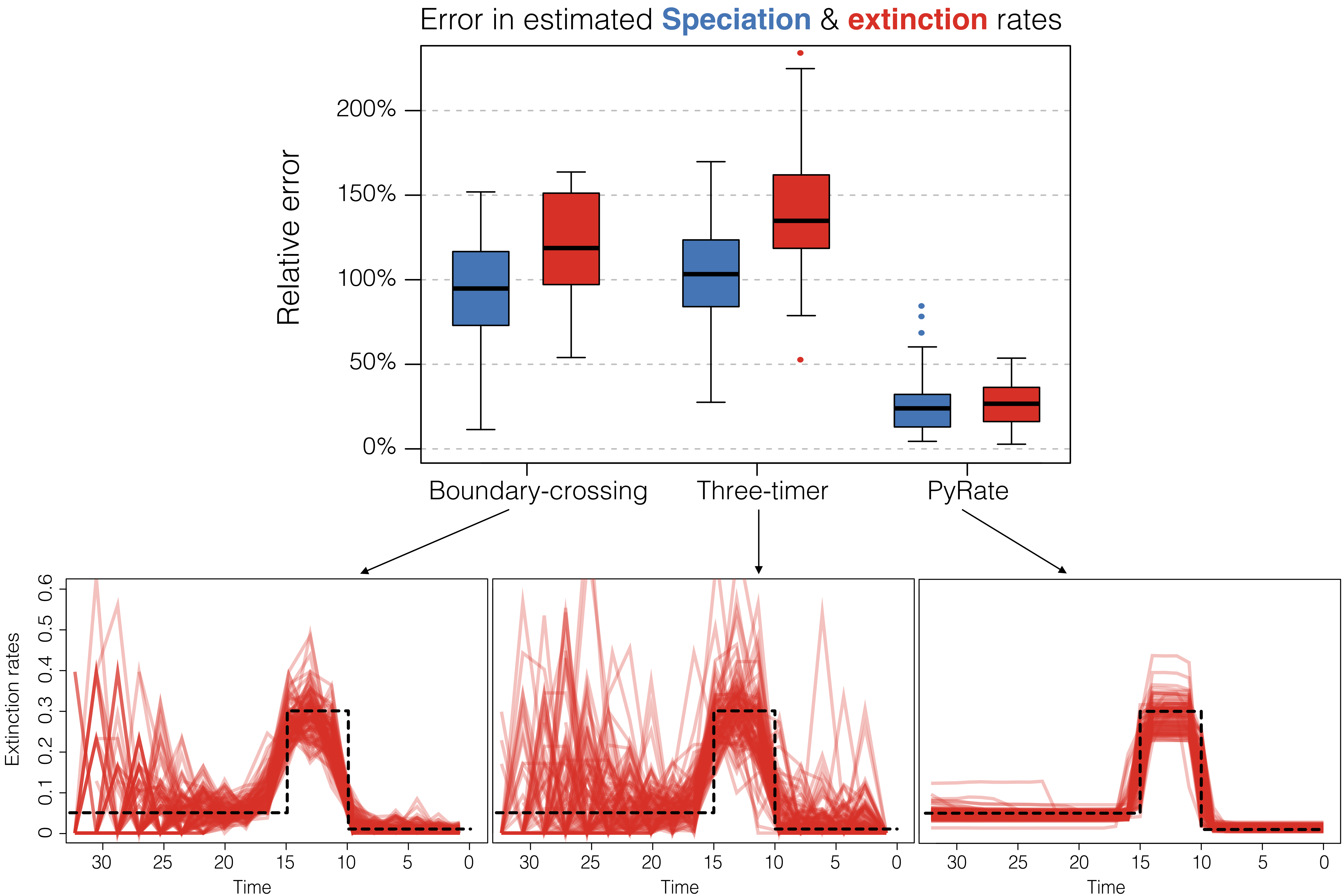


Mean per capita (Foote 2000)

Three-timer (Alroy 2008)

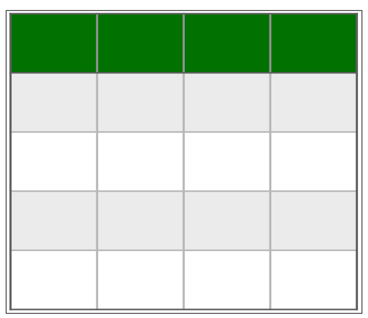
CMR (Liow & Finarelli 2014)

PyRate



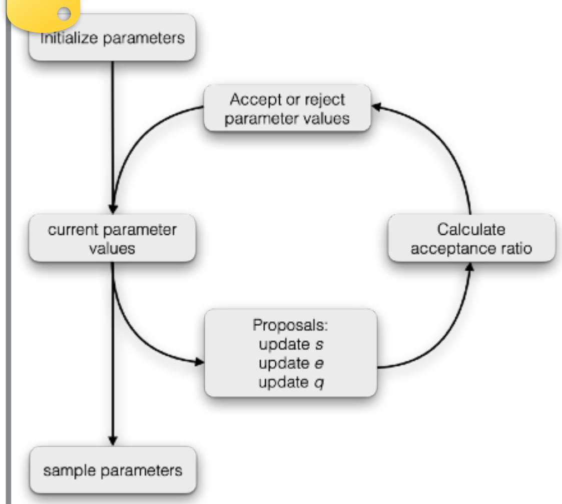
Silvestro et al. 2019 Paleobiology

How PyRate works



Fossil table and input preparation

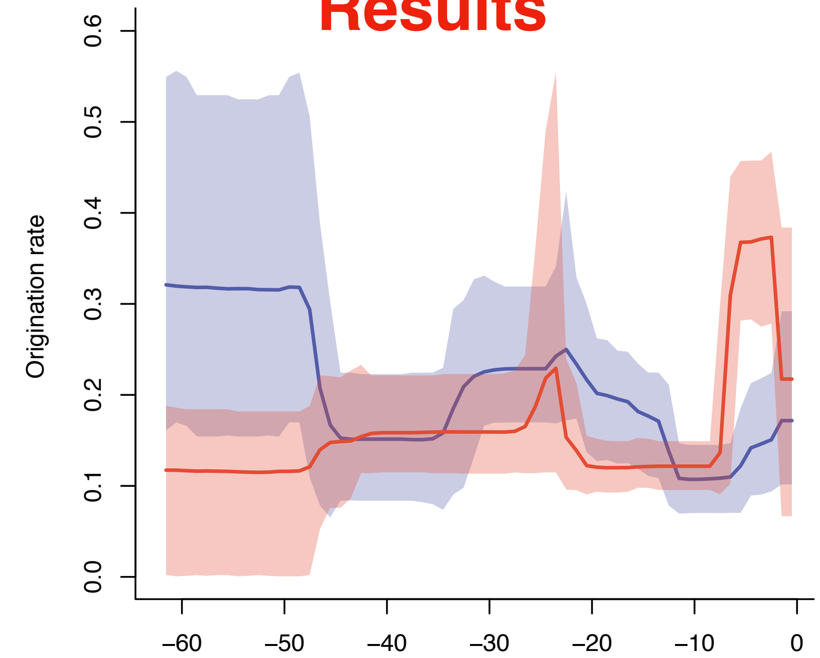
pyrate_utilities.R



MCMC

PyRate.py

Results



- Speciation, extinction, preservation rates through time
- Significance of rate shifts
- Times of origination and extinction of sampled species

How PyRate works

Taxon	Status	MinAge	MaxAge
Ailurus	Extant	0.6	1.3
Alopecocyon	Extinct	9.5	10.35
Alopecocyon	Extinct	9.7	11.11
...	...	...	...

Fossil occurrence data with age range and info on whether it's extant or extinct

Parse fossil table and create PyRate input file

```
# load PyRate utilities

source(".../PyRate-master/pyrate_utilities.r")
extract.ages(file = "path_to_your_fossil_table/Ursidae.txt",
             replicates = 10, # specifcy number of age randomizations
             cutoff = 10 # drop occurrences with wide age range (e.g. 10 myr)
             )


```

How PyRate works

Run PyRate in a Terminal window

PyRate — -zsh — 110×40

```
mbp:~ $ cd Software/PyRate
mbp:PyRate $ python PyRate.py path_to_input/Rhinocerotidae_PyRate.py
```

This will run a default analysis with RJMCMC

How PyRate works

Add options to the analysis: time variable preservation rates

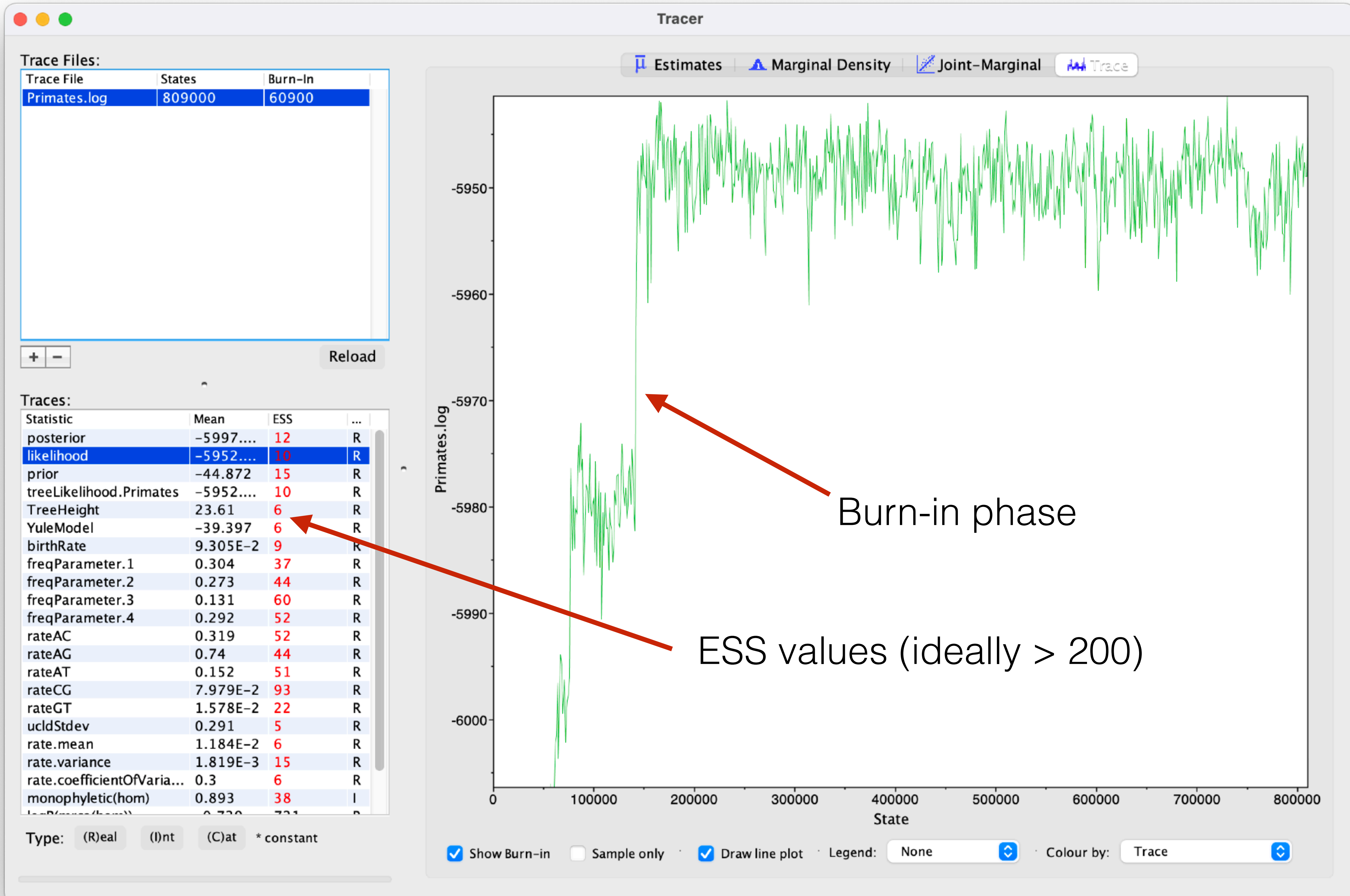
PyRate — -zsh — 110×40

```
mbp:~ $ cd Software/PyRate
mbp:PyRate $ python PyRate.py path_to_input/Rhinocerotidae_PyRate.py -qShift path_to_file/epochs.txt
```

Simple text file defining the time bins for piece-wise constant preservation rates

2.58
5.333
23.03
33.9
56.0
66.0

Check MCMC convergence

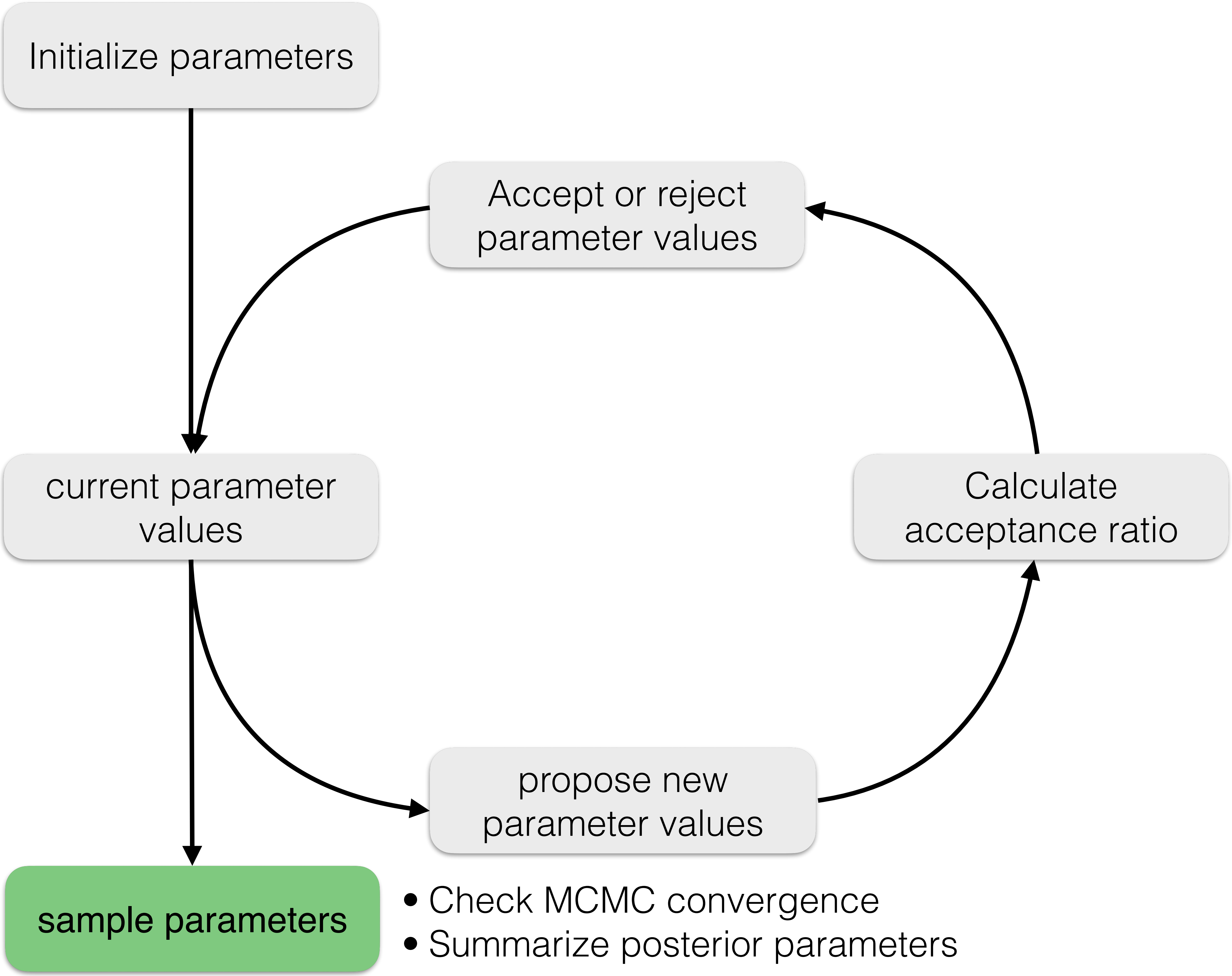


Burn-in phase

ESS values (ideally > 200)

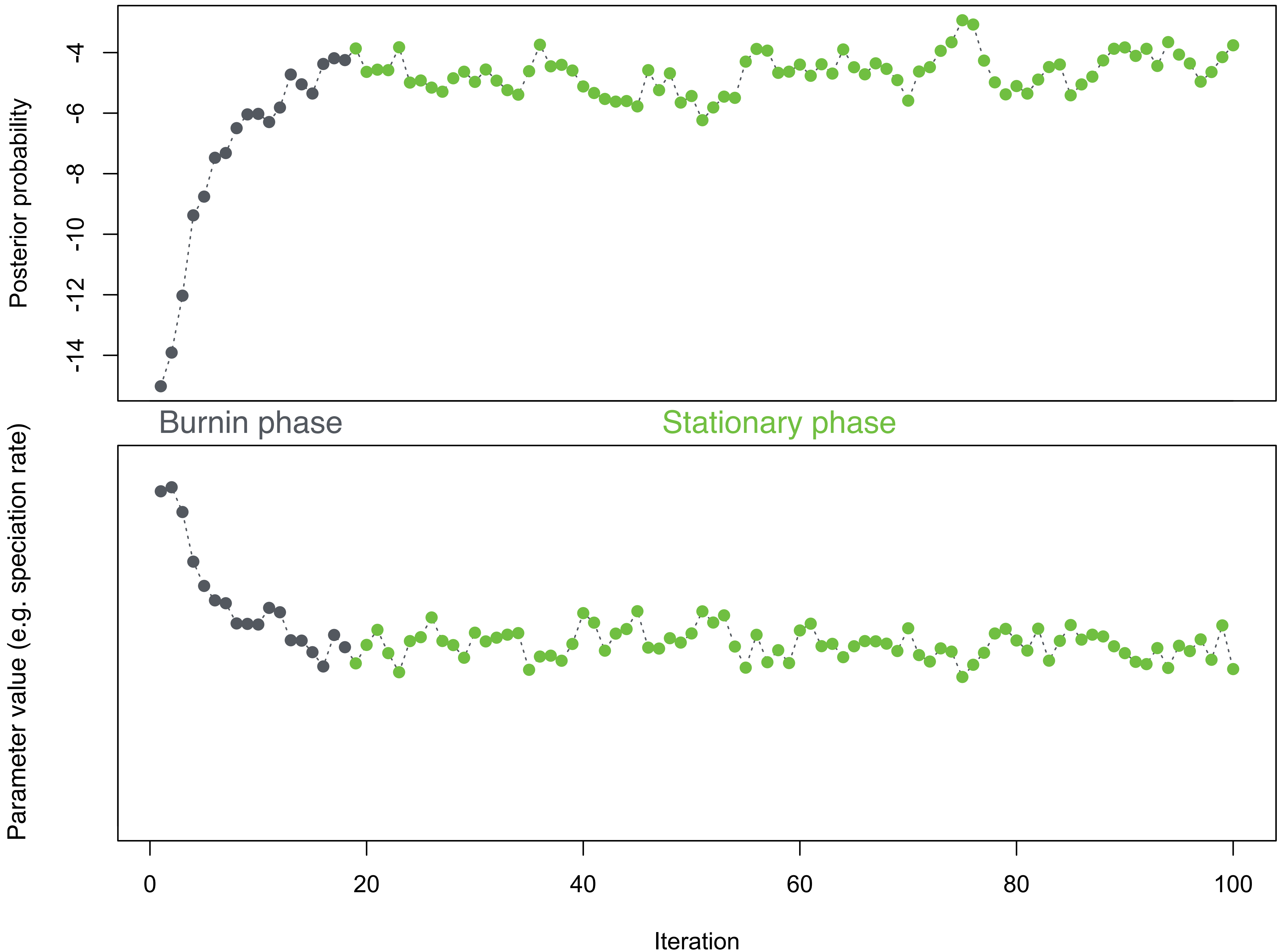
Tracer

Metropolis-Hastings MCMC

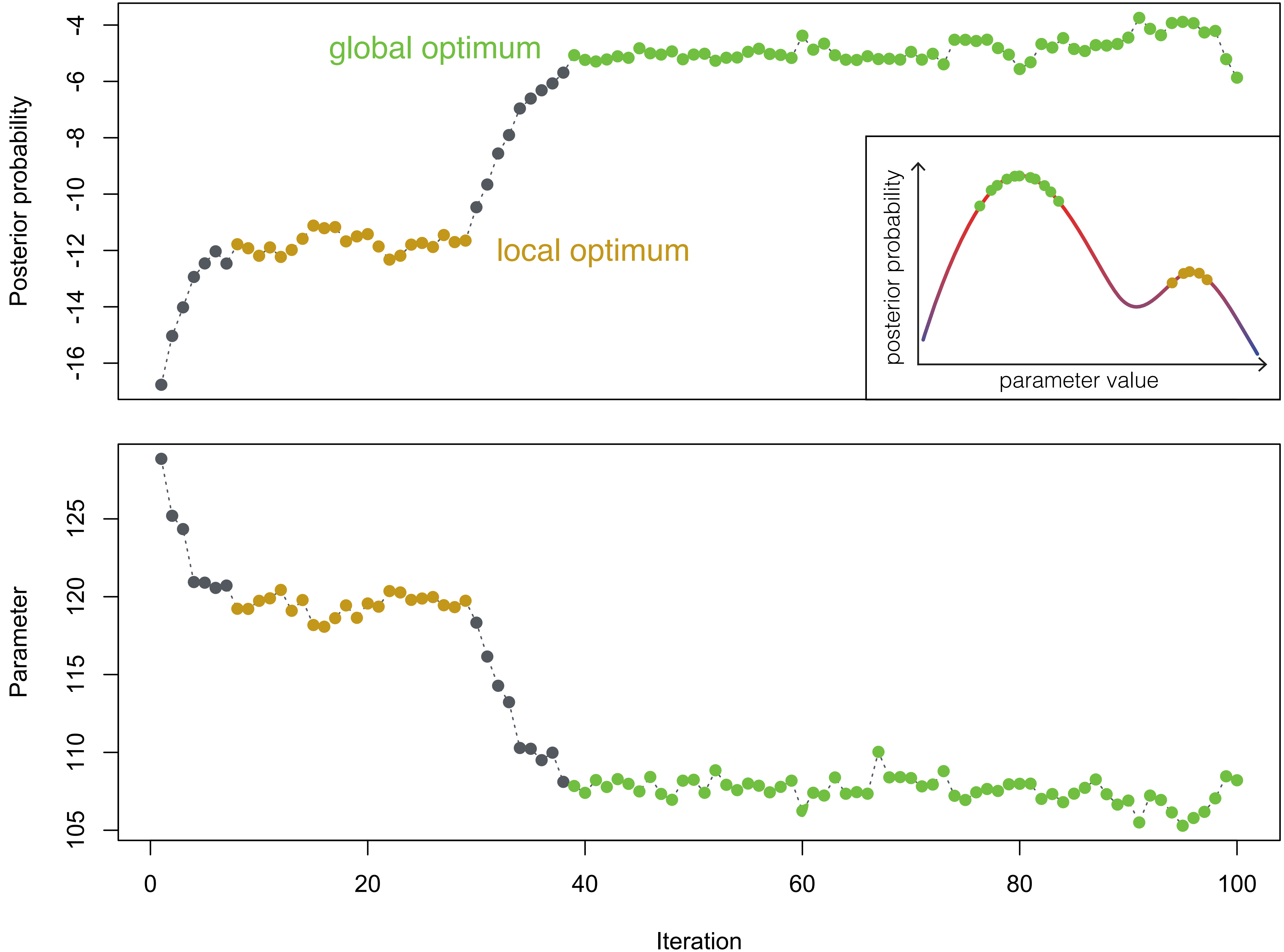


- Check MCMC convergence
- Summarize posterior parameters

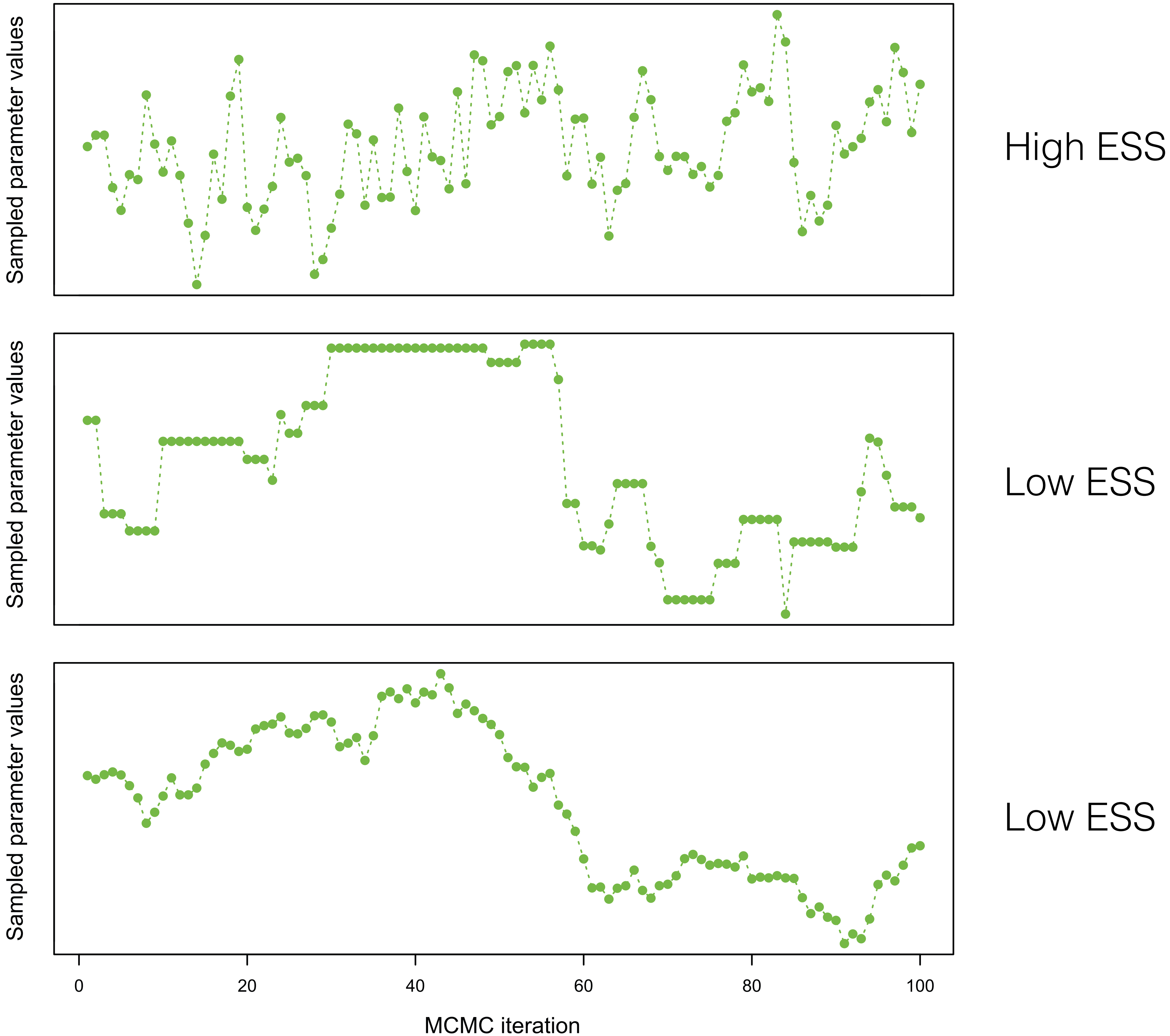
Burn-in and MCMC convergence



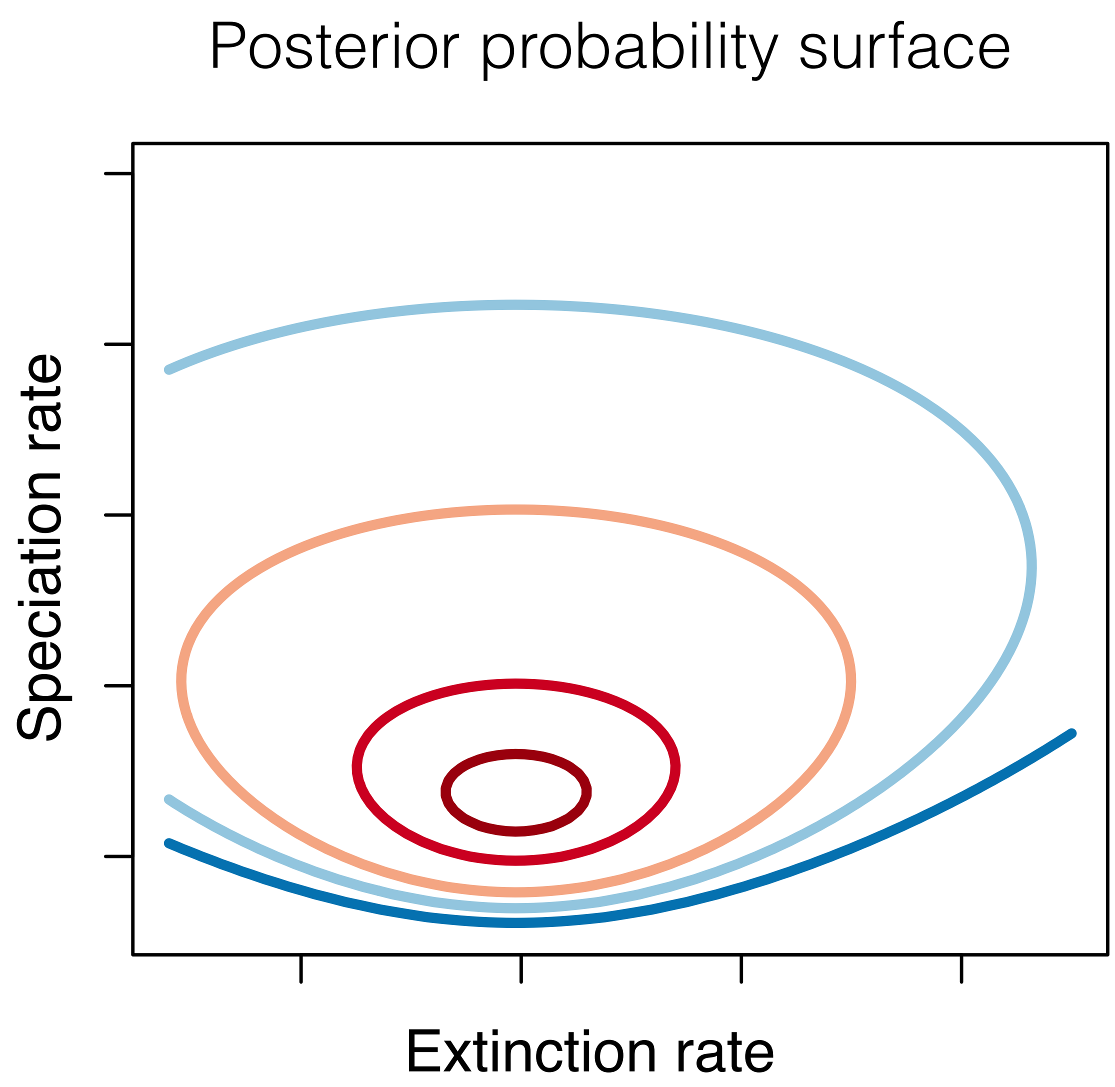
Burn-in and MCMC convergence



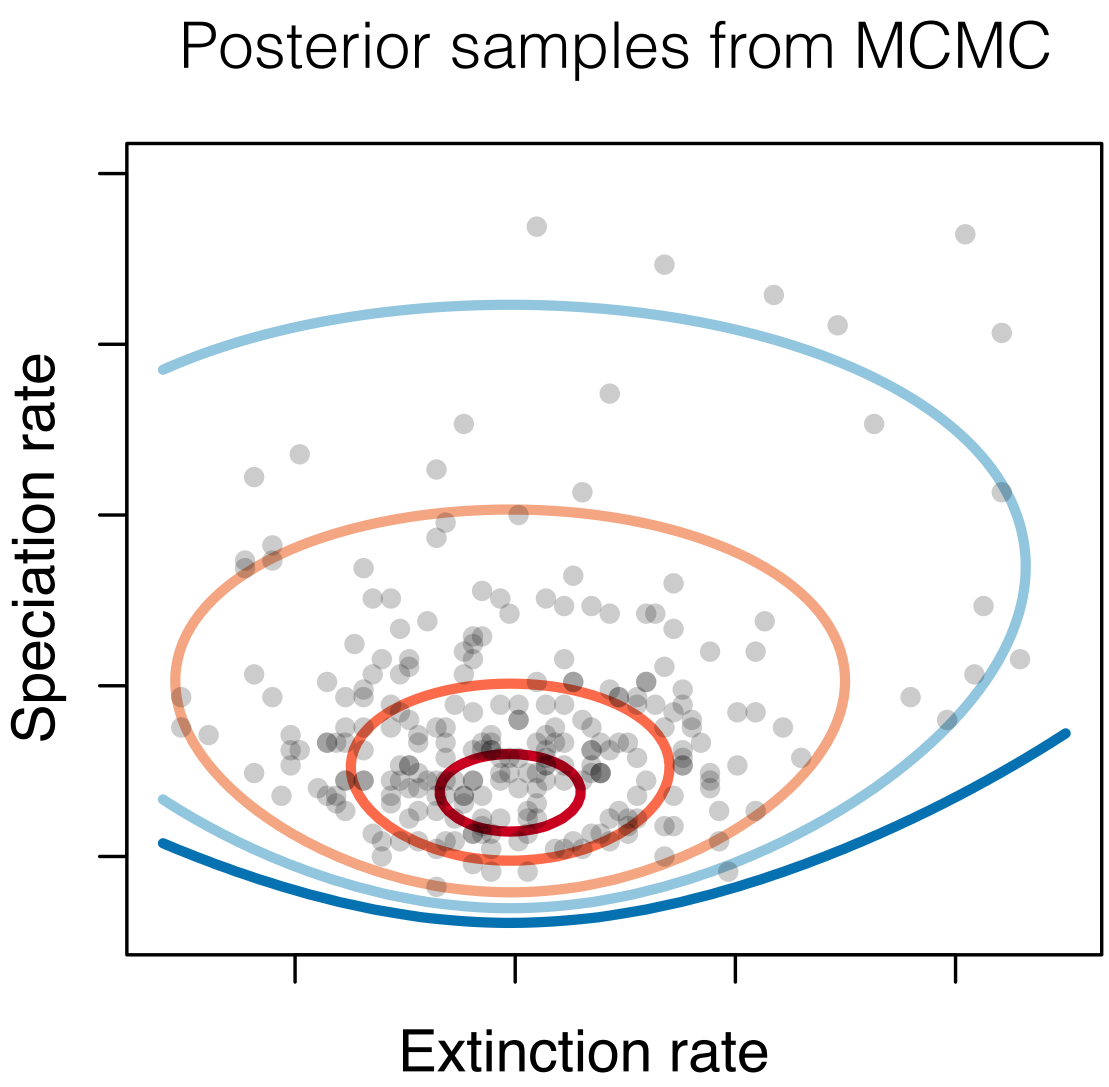
Posterior samples: Effective Sample Size (ESS)



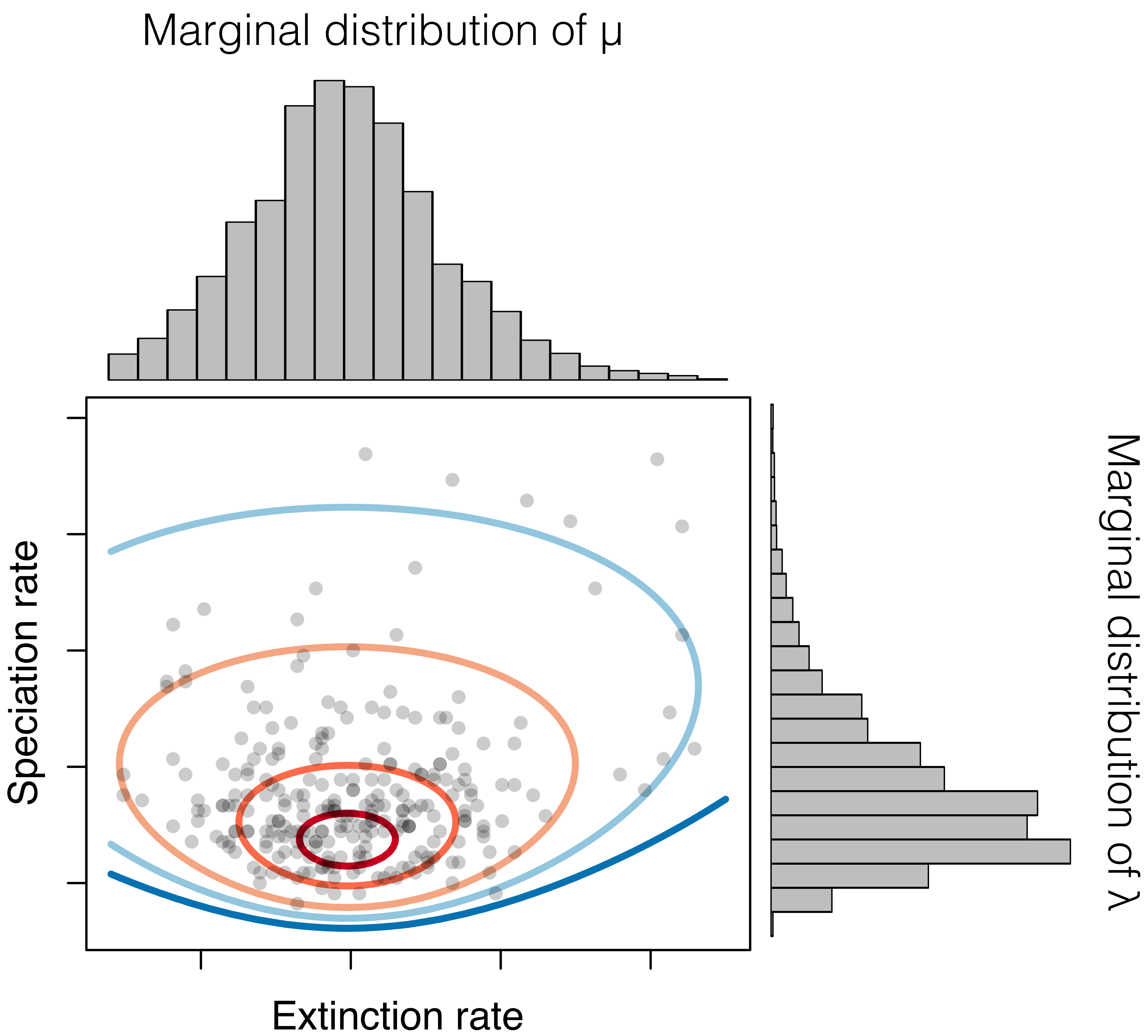
Summarizing MCMC samples



Summarizing MCMC samples



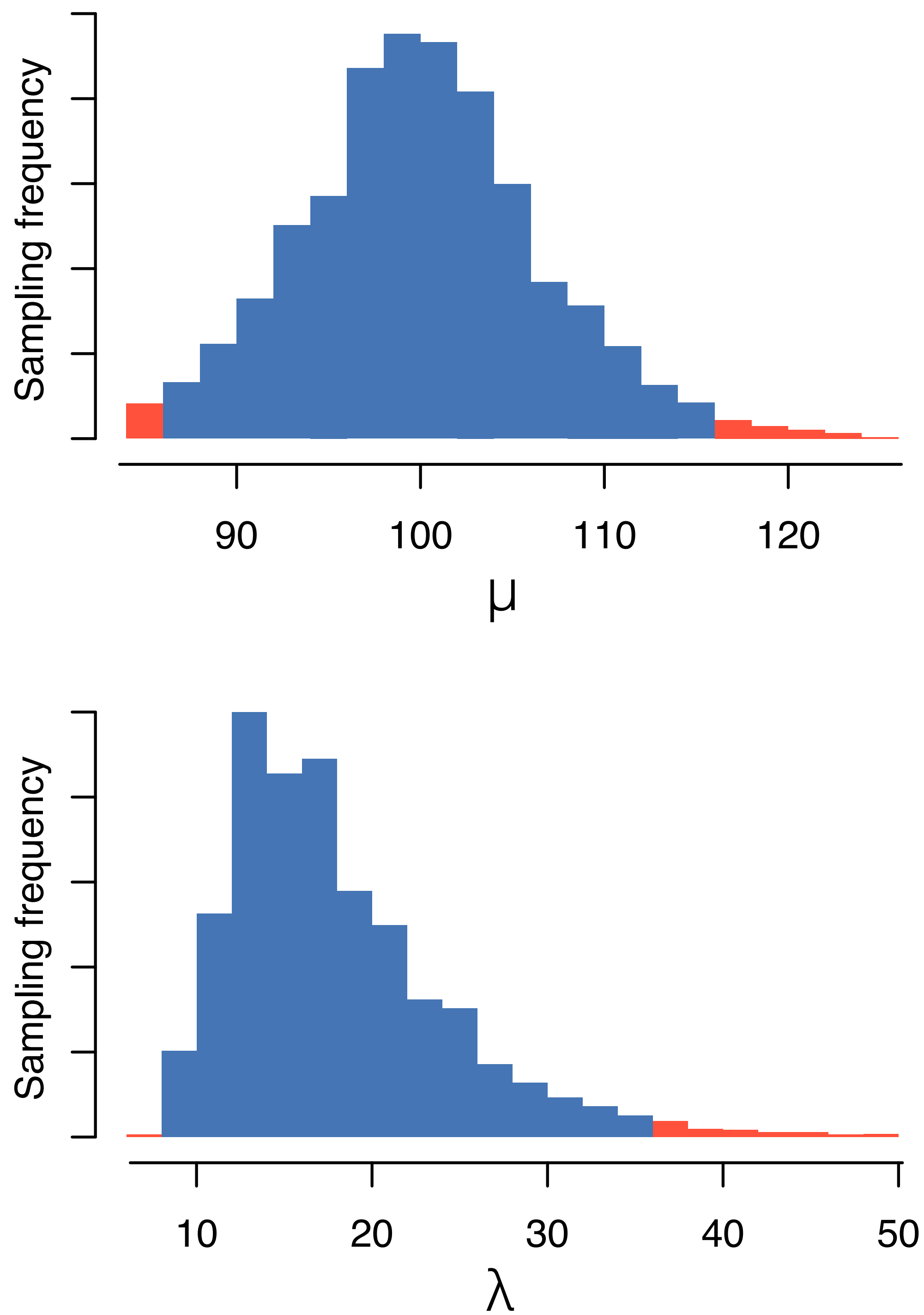
Summarizing MCMC samples



Summarizing MCMC samples

Posterior estimate: mean of MCMC samples

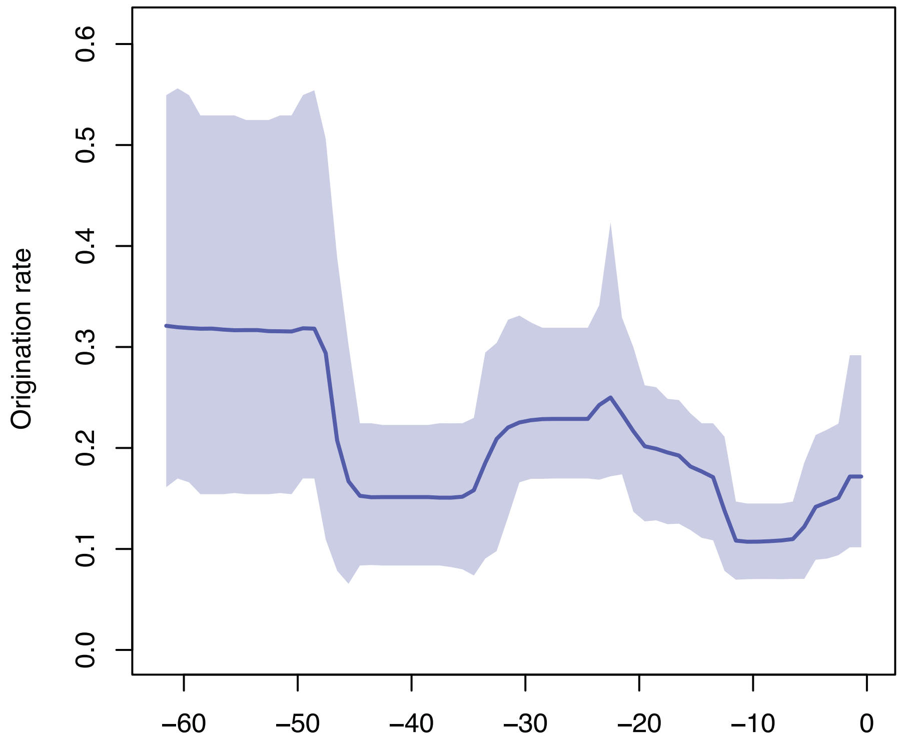
95% Credible intervals

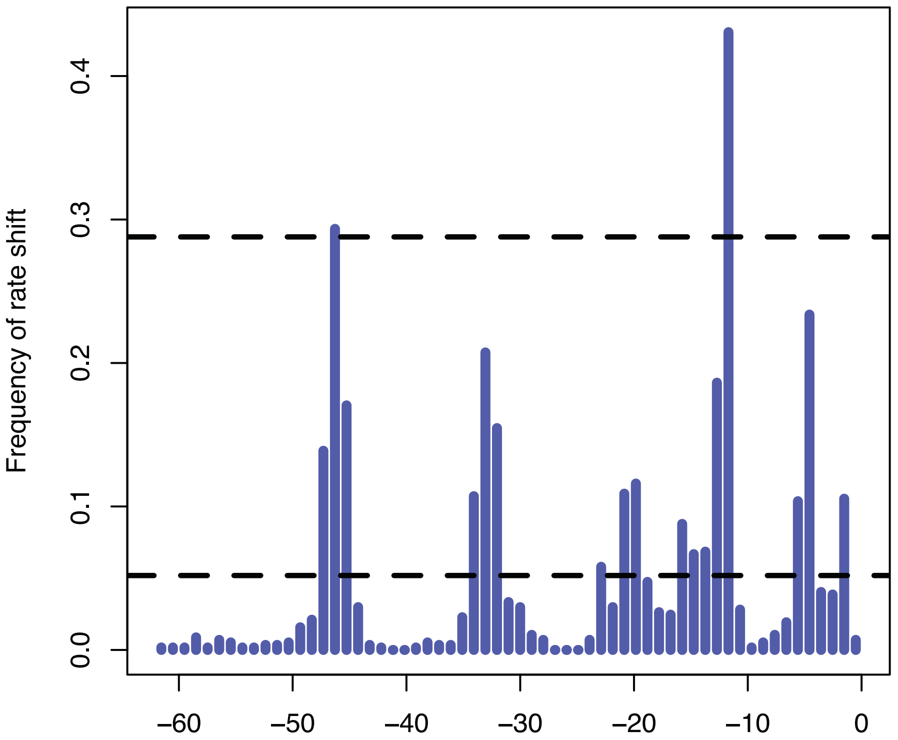


Plot results

PyRate — -zsh — 110×40

```
mbp:~ $ cd Software/PyRate
mbp:PyRate $ python PyRate.py -plotRJ your_path/pyrate_mcmc_logs -b 0.2
```





Strong support for a rate shift (logBF > 6)

Significant rate shift (logBF > 2)