Bayesian Phylogenetic Inference using RevBayes:

Introduction to RevBayes

Sebastian Höhna

GeoBio-Center
Ludwig-Maximilians Universität, München

Bastien Boussau

Tracy Heath

John Huelsenbeck

Fredrik Ronquist

RevBayes Team

Brian Moore

Michael Landis

Nicolas Lartillot

Sebastian Höhna

Will Freyman

April Wright

Mike May

RevBayes Team^2

Will Pett

Andy Magee

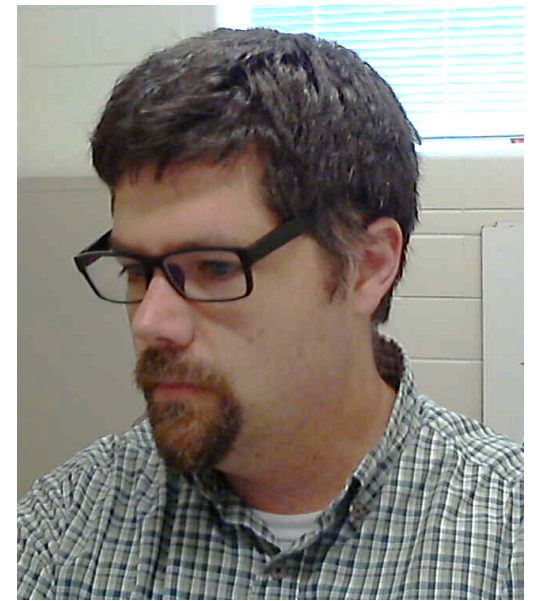Jeremy Brown

Rachel Warnock

Wade Dismukes

Dominik Schrempf

Joëlle Barido-Sottani

Ben Redelings

Jiansi Gao

RevBayes Team[3]

Which software to choose

?

Which software to choose

Does the software run the model/analysis?

Am I able to understand the software and to use it?

Is the software fast enough to give me an answer?

Why a single software

1. You only need to learn one software.

2. Lower chance of incompatibility.

3. Possibility to combine many models, methods & analysis.

4. Joint inferences are (theoretically) superior and easily feasible.

5. No redundant coding is necessary for recurring methods

   1. MCMC algorithm

   2. Likelihood computation

   3. Data structures

   4. etc.

Aims for RevBayes

1) General and flexible model specification

    a) Availability of (common) models

    b) Extendability

2) Easy to learn

    a) Well structured model specification

    b) Explicit models

    c) Documentation, examples and tutorials

3) Fast & Efficient

    a) Fast likelihood calculators
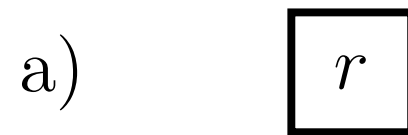
    b) Efficient (MCMC) algorithms, e.g., tree proposals

Graphical Model (DAG)

a) $\boxed{r}$
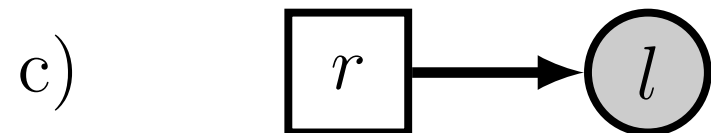
```
# constant node
r <- 10
```

Graphical Model (DAG)

a) $\boxed{r}$

b) $\boxed{r} \longrightarrow \bigcirc\!l$

```
# constant node
r <- 10



# stochastic node
l ~ dnExp(r)
```

Graphical Model (DAG)

a)   □ r

```
# constant node
r <- 10
```

b)   □ r → ◯ l

```
# stochastic node
l ~ dnExp(r)
```

c)   □ r → ◯ l

```
# stochastic node (observed)
l.clamp(0.1)
```

Graphical Model (DAG)

a) r # constant node
r <- 10

b) $r \rightarrow l$ # stochastic node
l ~ dnExp(r)

c) $r \rightarrow l$ # stochastic node (observed)
l.clamp(0.1)

d) $r \dashrightarrow l$ # deterministic node
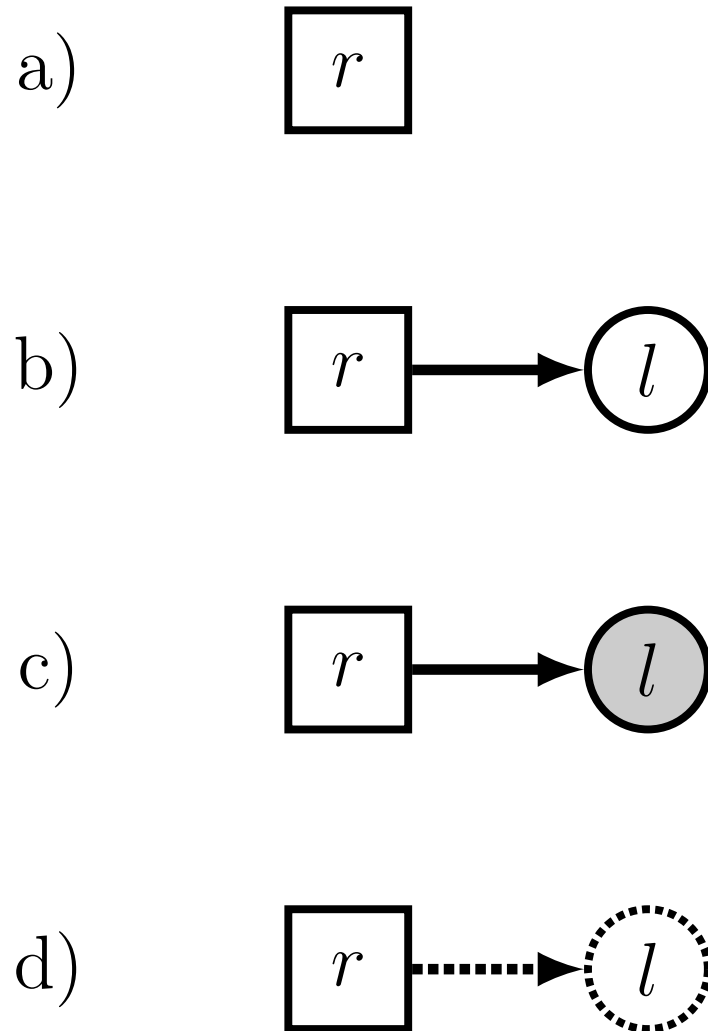l := exp(r)

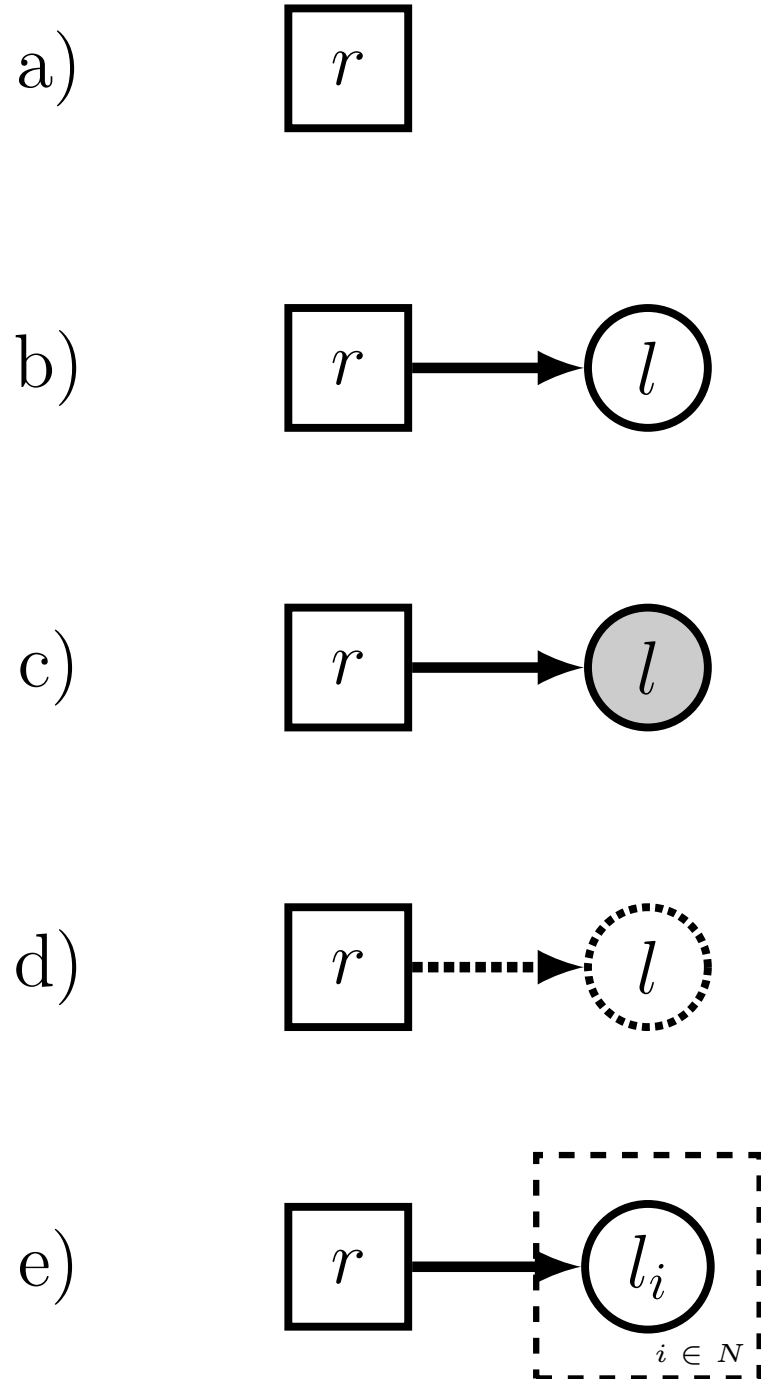# Graphical Model (DAG)

a)


```
# constant node
r <- 10
```
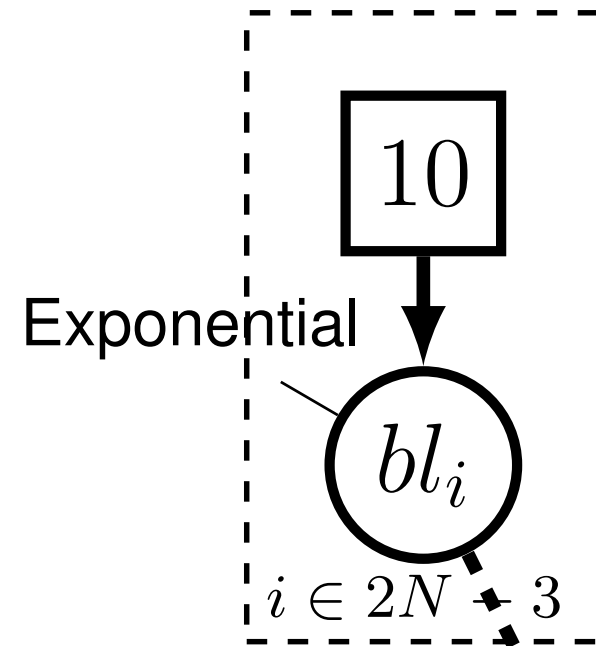
b)


```
# stochastic node
l ~ dnExp(r)
```

c)


```
# stochastic node (observed)
l.clamp(0.1)
```
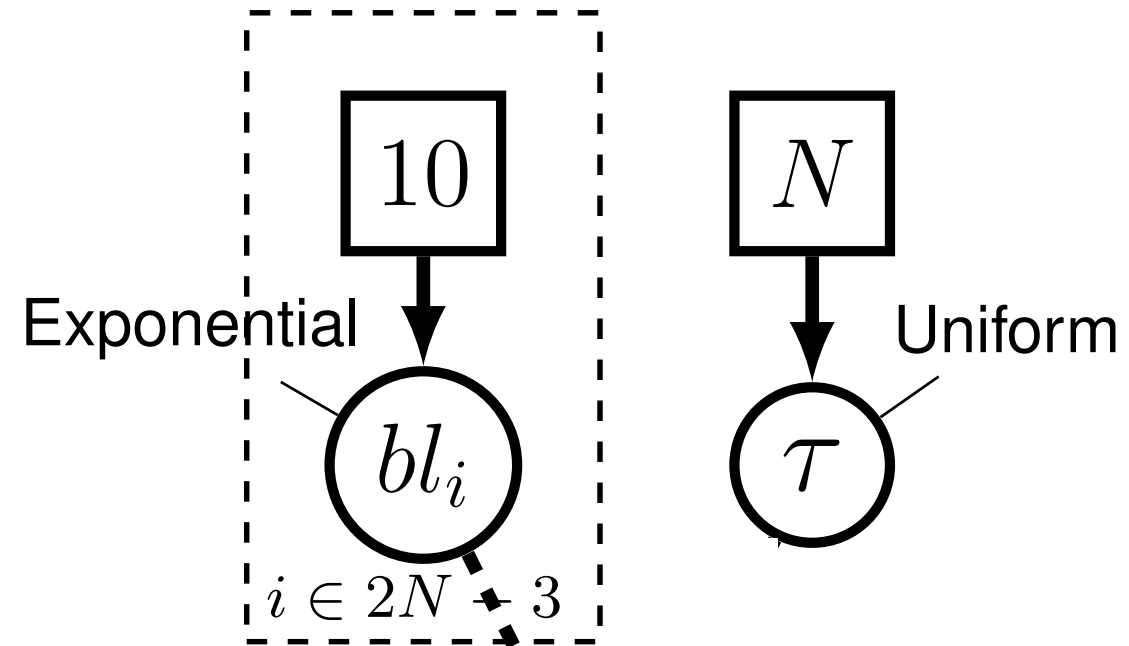
d)


```
# deterministic node
l := exp(r)
```

e)


```
# stochastic nodes (iid)
for (i in 1:N) {
    l[i] ~ dnExp(r)
}
```
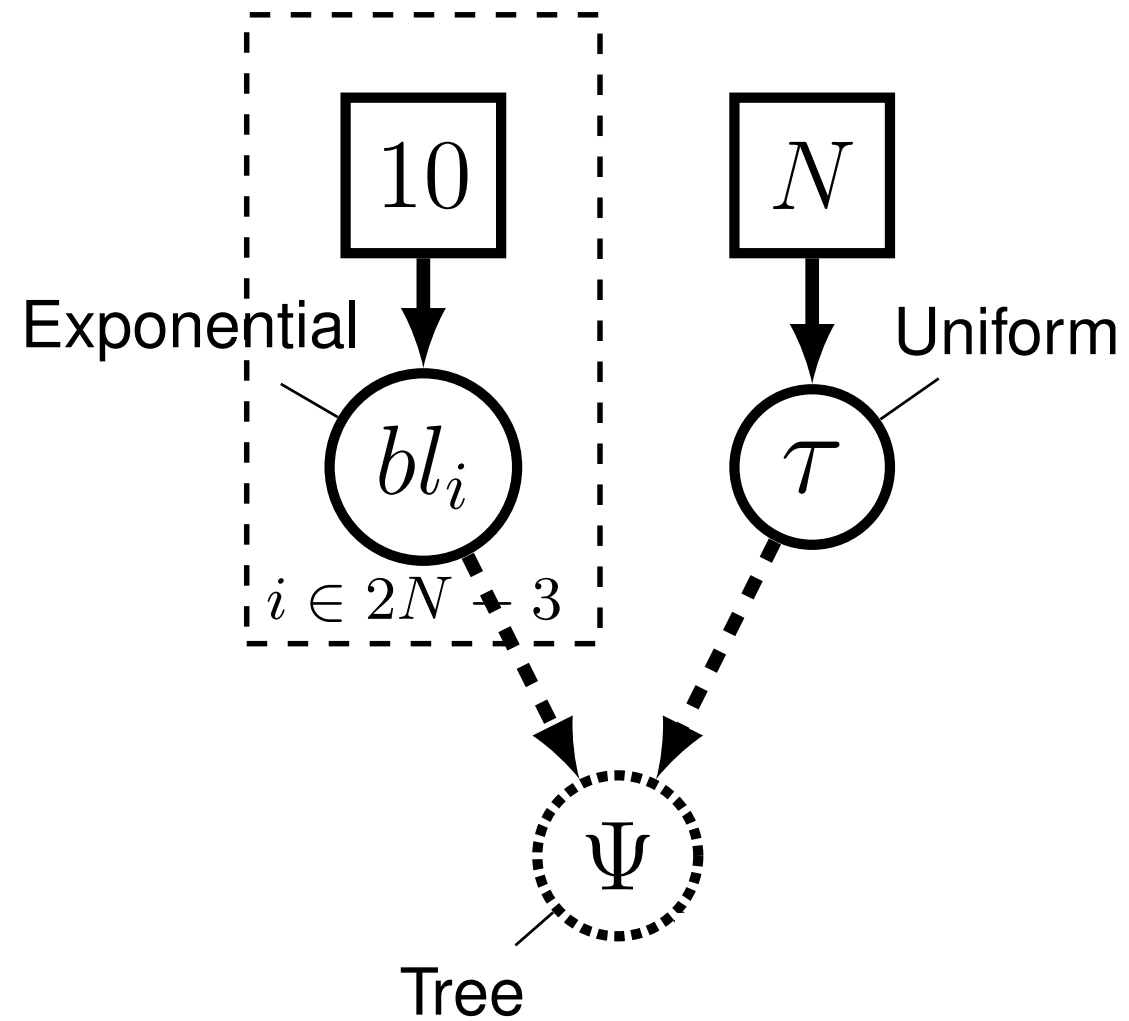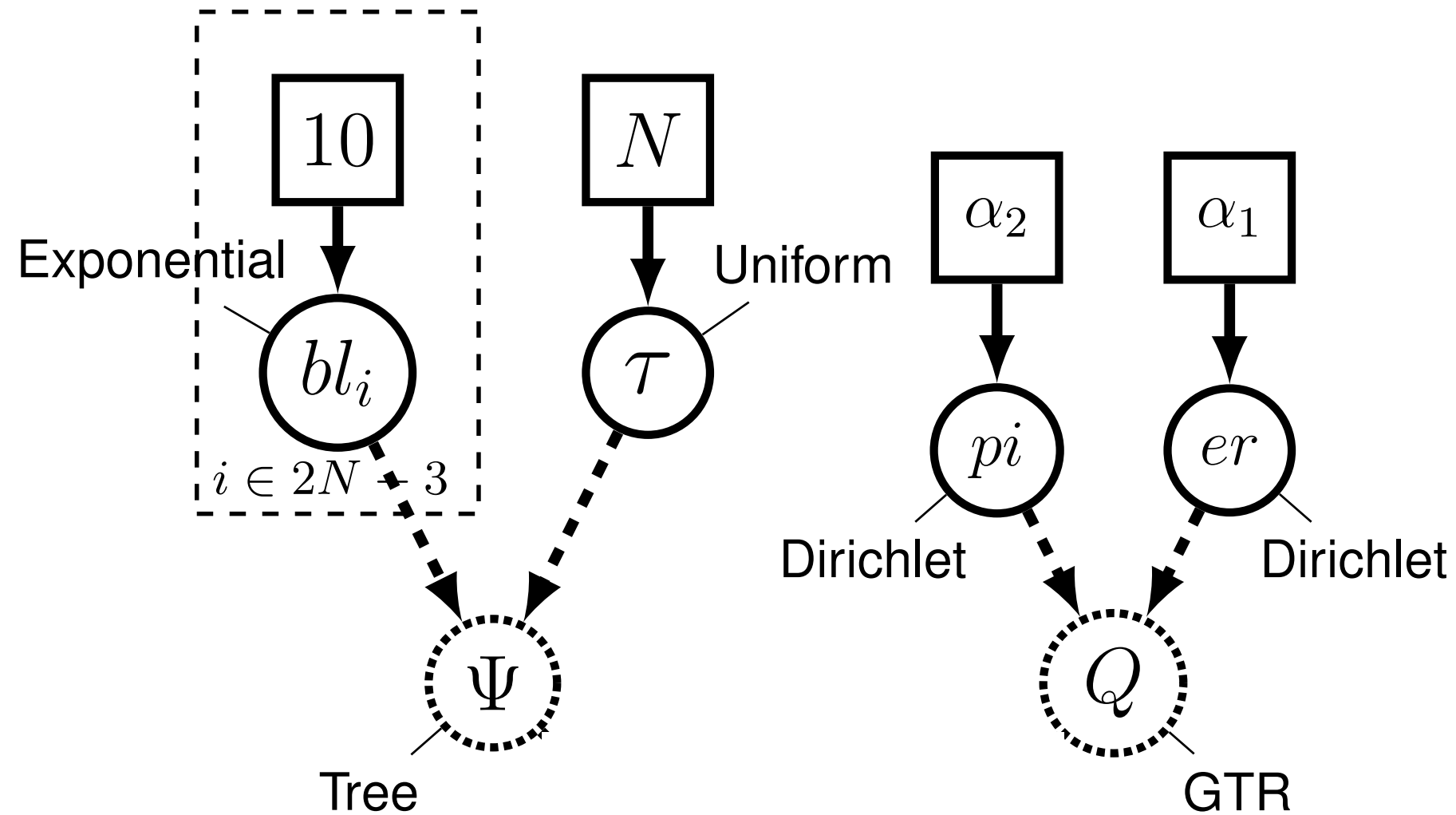
GTR+Gamma Model with Prior

GTR+Gamma Model with Prior

GTR+Gamma Model with Prior

GTR+Gamma Model with Prior

GTR+Gamma Model with Prior

GTR+Gamma Model with Prior
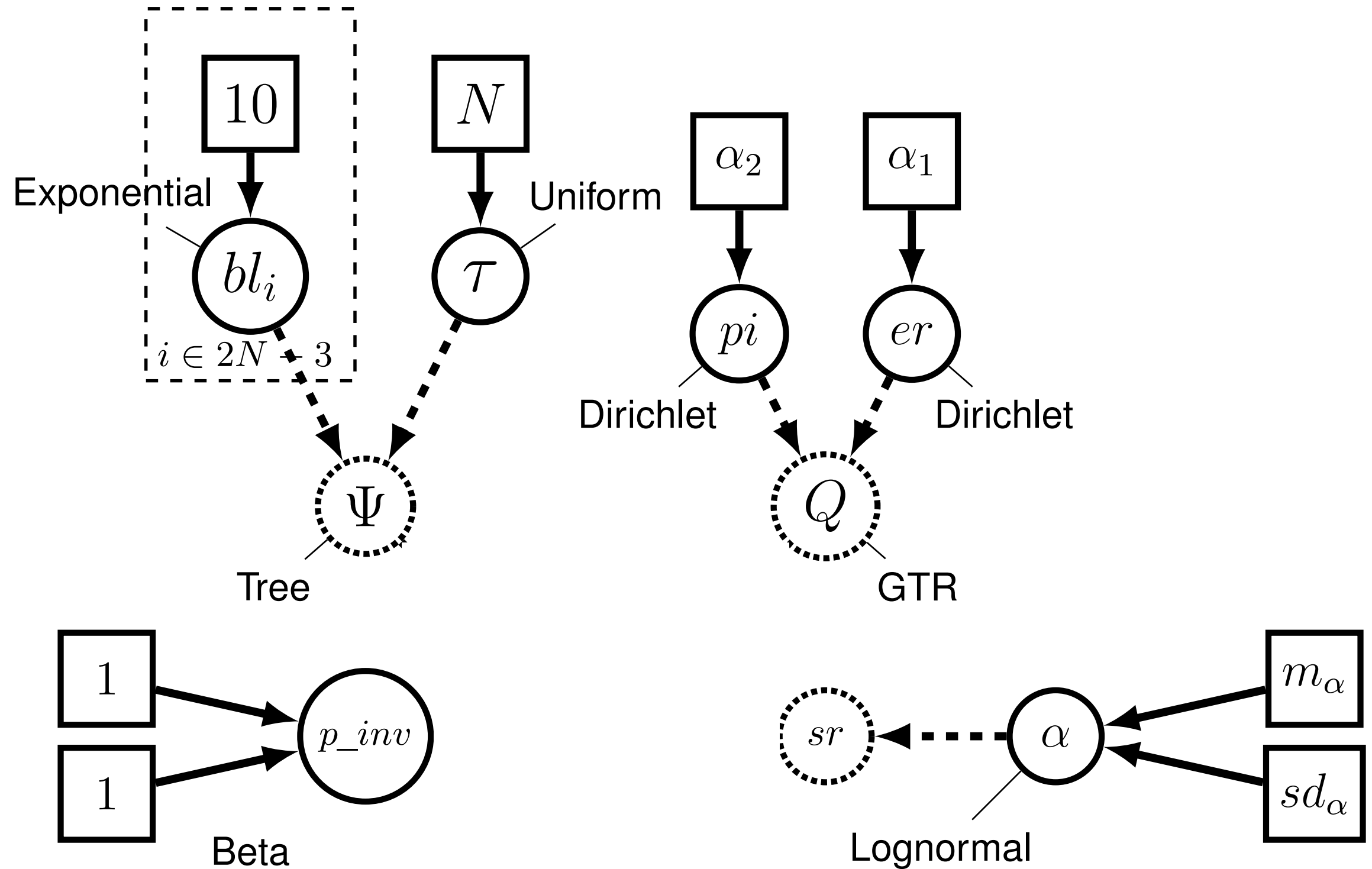
GTR+Gamma Model with Prior

Aims for RevBayes

1) General and flexible model specification

    a) Availability of (common) models

    b) Extendability   ✔

2) Easy to learn

    a) Well structured model specification

    b) Explicit models

    c) Documentation, examples and tutorials

3) Fast & Efficient

    a) Fast likelihood calculators
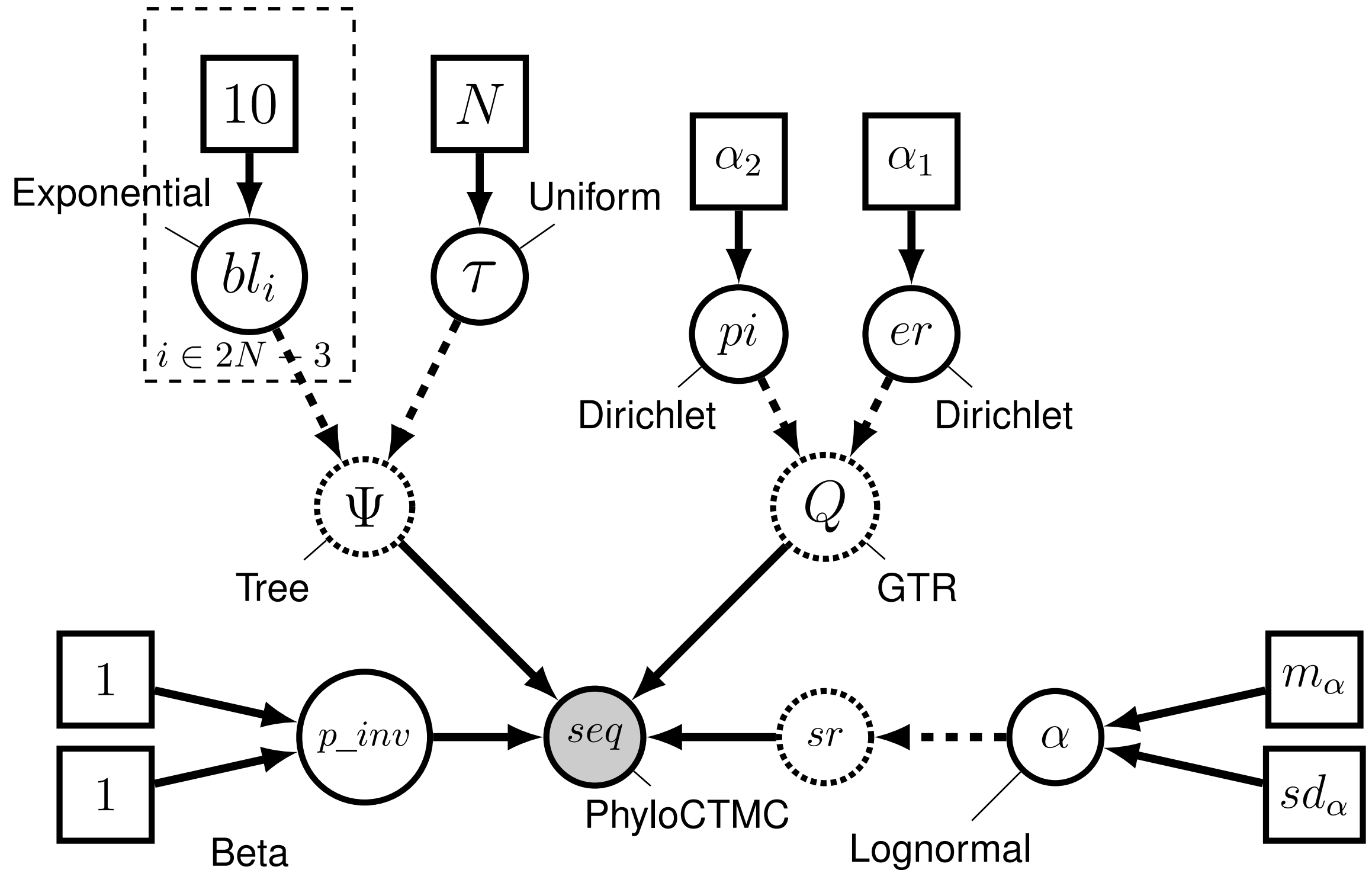
    b) Efficient (MCMC) algorithms, e.g., tree proposals

Rev (the RevBayes language)
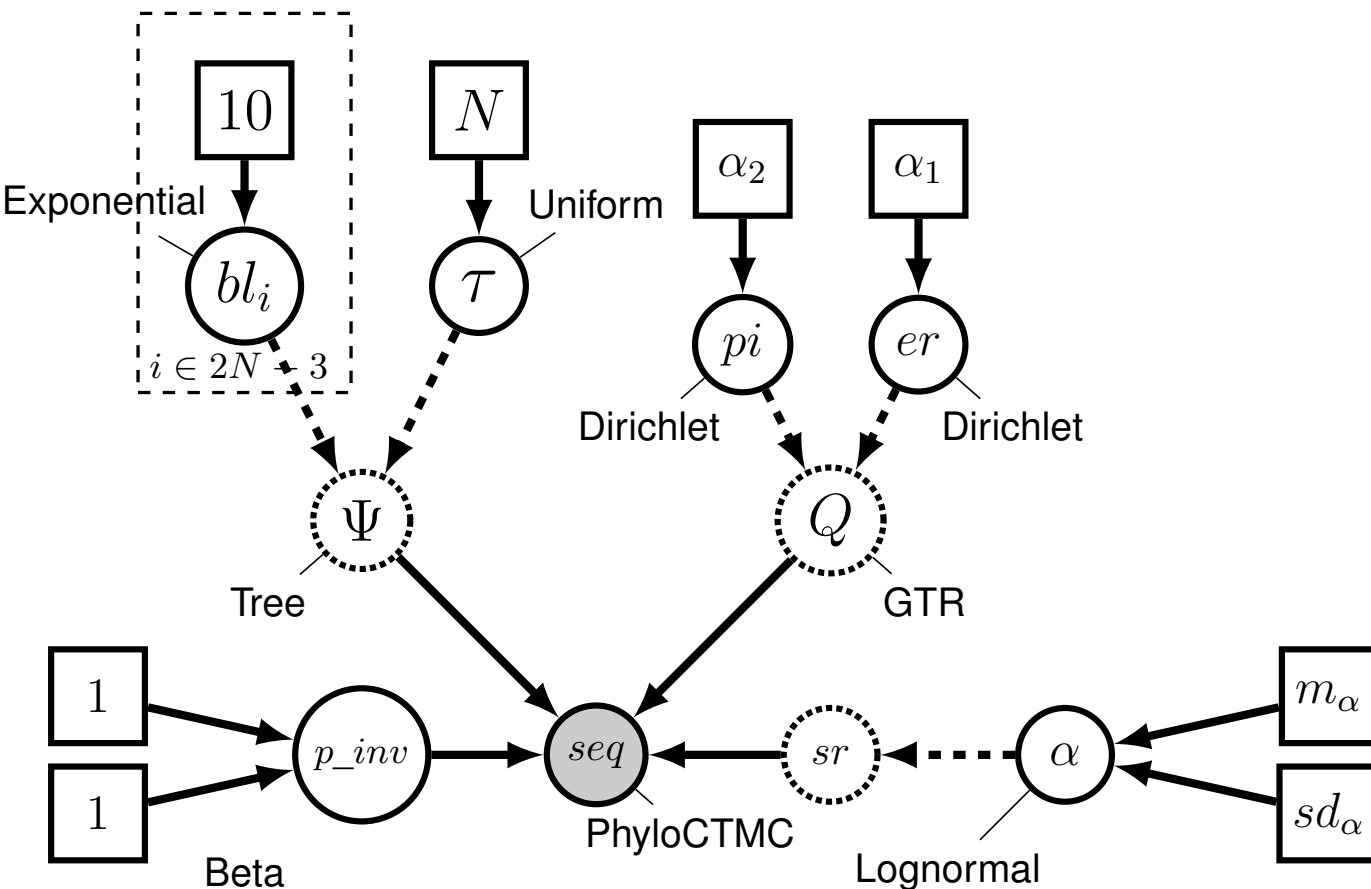

```
for (i in 1:n_branches) {
        bl[i] ~ dnExponential(10.0)
}
topology ~ dnUniformTopology(taxa)
psi := treeAssembly(topology, bl)

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

m_alpha <- ln(5.0)
sd_alpha <- 0.587405
alpha ~ dnLognormal( m_alpha, sd_alpha )
sr := fnDiscretizeGamma( alpha, alpha, 4, false )

p_inv ~ dnBeta(1,1)

seq ~ dnPhyloCTMC( tree=psi, Q=Q, pInv=p_invar,
                   siteRates=sr, type="DNA" )

seq.clamp( data )
```

Manual (Compilation of tutorials)

Currently covered topics:

1. Introduction

2. Models of molecular evolution.

3. Inference methods.

4. Divergence time estimation

5. Diversification rate estimation

6. Gene tree - species tree estimation

7. Historical biogeography

8. Comparative phylogenetic method

Over 400 pages !!!

Aims for RevBayes

1) General and flexible model specification

    a) Availability of (common) models

    b) Extendability   ✔

2) Easy to learn

    a) Well structured model specification

    b) Explicit models   ✔

    c) Documentation, examples and tutorials

3) Fast & Efficient

    a) Fast likelihood calculators

    b) Efficient (MCMC) algorithms, e.g., tree proposals

Performance Study — Likelihood Computation

Primates:
- 12 taxa
- 898 sites
- 412 patterns

MCMC:
- burnin of 10^5
- chain length of 10^6
- only substitution model parameters are updated

	HKY	HKY+G	GTR	GTR+G
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

* times are given in seconds

Performance Study — MCMC shortcuts

Primates:
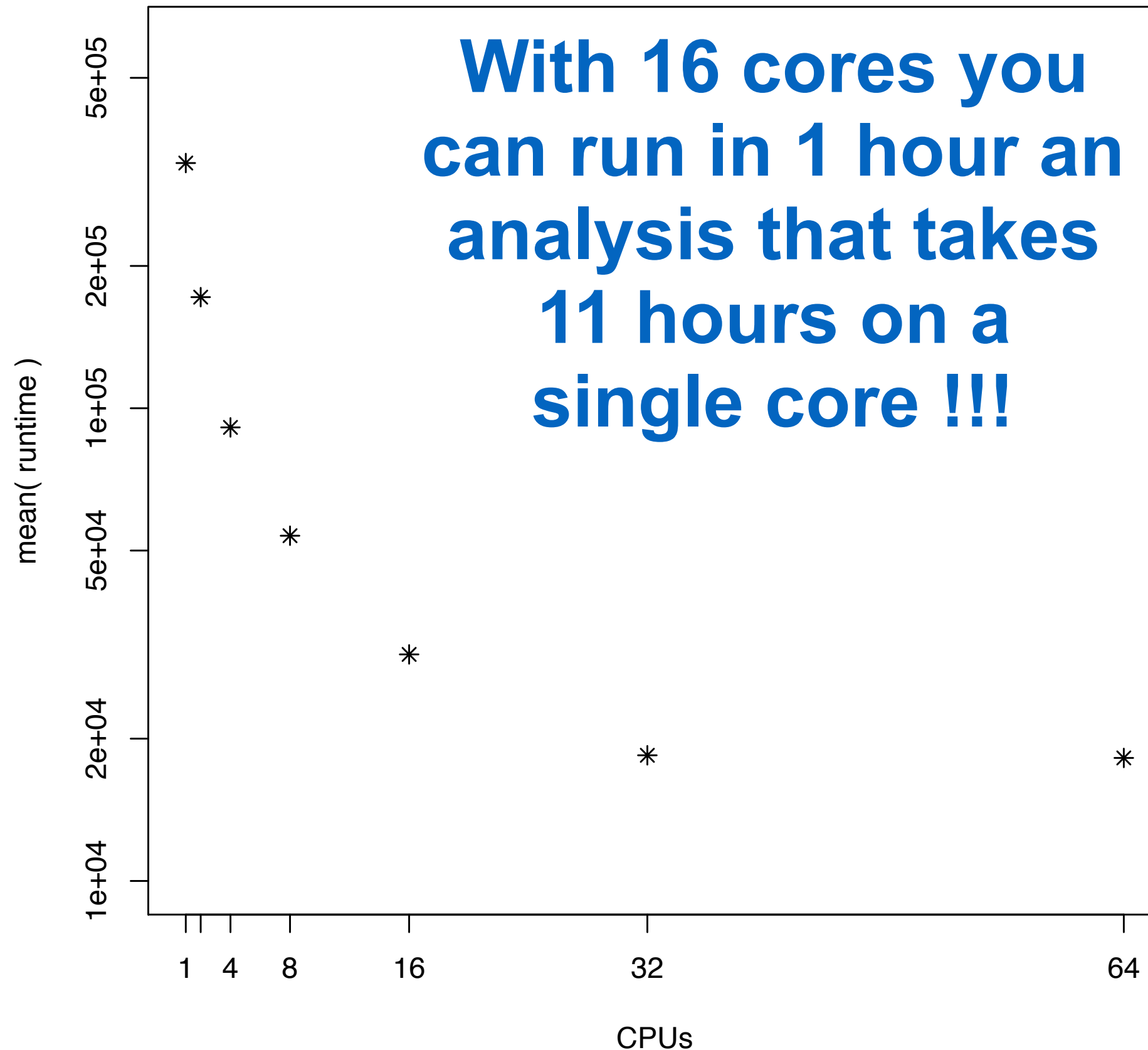- 12 taxa
- 898 sites
- 412 patterns

MCMC:
- burnin of 10^5
- chain length of 10^6
- only topology or node ages are updated

	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

* times are given in seconds

RevBayes on Computer Clusters

Aims for RevBayes

1) General and flexible model specification

   a) Availability of (common) models

   b) Extendability   ✔

2) Easy to learn

   a) Well structured model specification

   b) Explicit models   ✔

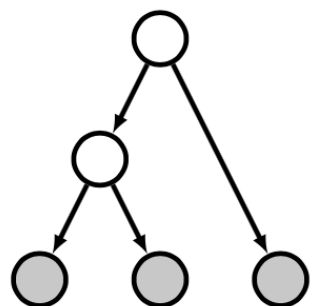   c) Documentation, examples and tutorials

3) Fast & Efficient

   a) Fast likelihood calculators

   b) Efficient (MCMC) algorithms, e.g., tree proposals   ✔

Resources

- General website: www.RevBayes.com

- Code repository: github.com/revbayes/revbayes

- Tutorials: http://revbayes.github.io/tutorials.html

- Help: https://revbayes.github.io/documentation/

- Discussion-group/mailing-list:

    https://groups.google.com/forum/#!forum/revbayes-users

RevBayes.com

RevBayes

Bayesian phylogenetic inference using probabilistic graphical models and an interpreted language

About

RevBayes provides an interactive environment for statistical computation in phylogenetics. It is primarily intended for modeling, simulation, and Bayesian inference in evolutionary biology, particularly phylogenetics. However, the environment is quite general and can be useful for many complex modeling tasks.

RevBayes uses its own language, Rev, which is a probabilistic programming language like JAGS, STAN, Edward, PyMC3, and related software. However, phylogenetic models require inference machinery and distributions that are unavailable in these other tools.

The Rev language is similar to the language used in R. Like the R language, Rev is designed to support interactive analysis. It supports both functional and procedural programming models, and makes a clear distinction between the two. Rev is also more strongly typed than R.

Software

Download and Install RevBayes

Mac OS X

[Download Executable (10.6+)]

Windows

[Download Executable (7+)]

Source code

[GitHub Repository]

Compiling from source

Linux

```
$ git clone https://github.com/revbayes/revbayes.git revbayes
$ cd revbayes/projects/cmake
$ ./build.sh
```

For the MPI version:

RevBayes Tutorials

This list shows all of the RevBayes tutorials for learning various aspects of RevBayes and Bayesian phylogenetic analysis. Each one explicitly walks you through model specification and analysis set-up for different phylogenetic methods. These tutorials have been written for new users to learn RevBayes at home, at workshops, and in courses taught at the undergraduate and graduate levels. You may find that the styles are somewhat different between tutorials and that some have overlapping content.

Please see the Tutorial Format guide for details about how to read the tutorials.

Please see Recommended Software for links to various software programs you may need to download in order to follow the tutorials.

Contribute!

Introduction to RevBayes and MCMC

Tutorials

Introduction to RevBayes and MCMC

Getting Started with RevBayes

A very basic overview on how to use RevBayes

Rev Language Syntax

A very short introduction to the Rev language

Introduction to Graphical Models

A gentle introduction to graphical models, probabilistic programming, and MCMC using a simple linear regression example.

Introduction to MCMC using RevBayes

A simple Archery example for building a hierarchical model and sampling under Markov Chain Monte Carlo

Introduction to MCMC using RevBayes

Introduction to MCMC Simulation using a simple Binomial Model

Understanding Continuous-Time Markov Models

Simulating DNA sequence evolution with a die

Basic introduction to Rev & MCMC

General Rev language features and simple Poisson regression

Diagnosing MCMC performance

How to assess the performance of MCMC simulations

Tutorials

Model Selection and Testing

General Introduction to Model selection

Comparing relative model fit with Bayes factors

Model selection of common substitution models for one locus

Comparing relative model fit with Bayes factors

Model averaging of substitution models

Reversible-jump MCMC over substitution models

Model selection of partition models

Comparing relative model fit with Bayes factors

Assessing Phylogenetic Reliability Using RevBayes and P^3

Model adequacy testing using posterior prediction (Data Version).

Assessing Phylogenetic Reliability Using RevBayes and P^3

Model adequacy testing using posterior prediction (Inference Version).

Tutorials

Standard tree inference

Substitution Models

Phylogenetic inference of nucleotide data using RevBayes

Inferring Selection through Codon Models

Estimating dN/dS

Partitioned data analysis

Current Protocols in Bioninformatics - Phylogenetic Inference using RevBayes (Protocol #2)

Relaxed Clocks & Time Trees

Comparing relaxed clock models & estimating rooted time trees

Discrete morphology - Tree Inference

Phylogenetic inference with discrete morphological data

Alignment concatenation

Species tree estimation from multiple concatenated gene sequence alignments

Phylogenetic Inference using RevBayes

Continuous trait evolution with Ornstein-Uhlenbeck models

Gene tree - species tree reconstruction

Incomplete lineage sorting and the multispecies coalescent

Discrete morphology - Ancestral State Estimation

Ancestral State Estimation and Testing for Irreversibility

Tutorials

Complex hierarchical models for phylogenetic inference

Combined-Evidence Analysis and the Fossilized Birth-Death Process for Analysis of Fossil and Extant Specimens

Joint inference of divergence times and phylogenetic relationships of fossil and extant taxa

Combined-Evidence Analysis and the Fossilized Birth-Death Process for Stratigraphic Range Data

Joint inference of divergence times and phylogenetic relationships of fossil and extant taxa

Tutorials

Diversification Rate Estimation

Introduction to Diversification Rate Estimation

Overview of Analyses, Models and Theory

Simple Diversification Rate Estimation

Comparing different constant-rate models of lineage diversification

Episodic Diversification Rate Estimation

Estimating Rate-Variation Through Time in Diversification Rates

Environmental-dependent Speciation & Extinction Rates

Estimating Correlation between Diversification Rates and Environmental Characters

Branch-Specific Diversification Rate Estimation

How to estimate branch-specific shifts in diversification rates

Background on state-dependent diversification rate estimation

An introduction to inference using state-dependent speciation and extinction (SSE) branching processes

State-dependent diversification with BiSSE and MuSSE

Inference using the

Diversification Rate Estimation with Missing Taxa

How to estimate

State-dependent diversification with HiSSE and ClaSSE

Hidden state and

Tutorials

Comparative methods

Continuous trait evolution and Brownian motion

Phylogenetic analysis of continous characters and the comparative method

Chromosome Evolution

Modeling chromosome evolution with ChromEvol, BiChroM, and ChromoSSE

Tutorials

Biogeography

Introduction to Historical Biogeography

Phylogeographic analysis under the dispersal-extinction-cladogenesis process

Historical Biogeographic Analysis using a Simple DEC Model

Ancestral range estimation of the silverswords under the dispersal-extinction-cladogenesis process

Historical Biogeographic Analysis using an epoch DEC Model

Ancestral range estimation of the silverswords under the dispersal-extinction-cladogenesis process

Biogeographic dating analysis

Biogeographic dating analysis under the dispersal-extinction-cladogenesis process

Forum/Mailing group