Workshop
New approaches to phylogenetic inference

Session 3
Bayesian polymorphism-aware phylogenetic inference

Rui Borges & Carolin Kosiol

17th February 2021

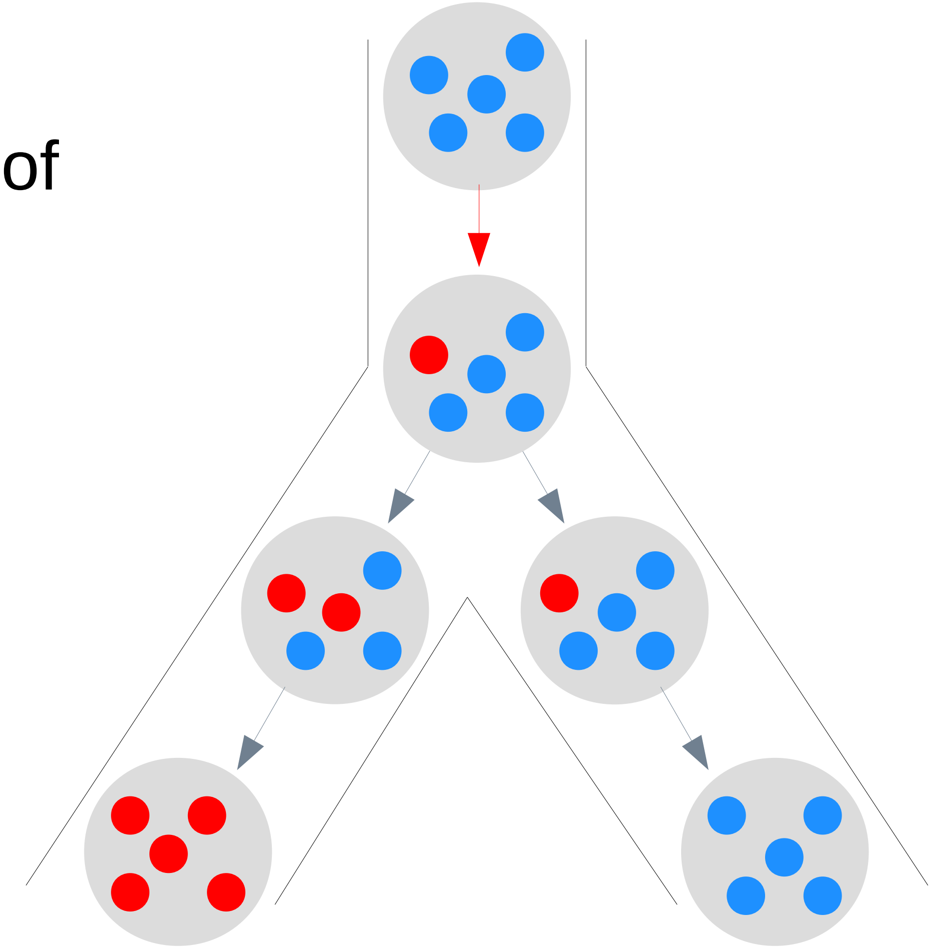
Plan for today:

- PoMos with selection
- Virtual PoMos
- Tutorial: Bayesian inference of phylogeny using polymorphism-aware models

PoMos re-re-visited

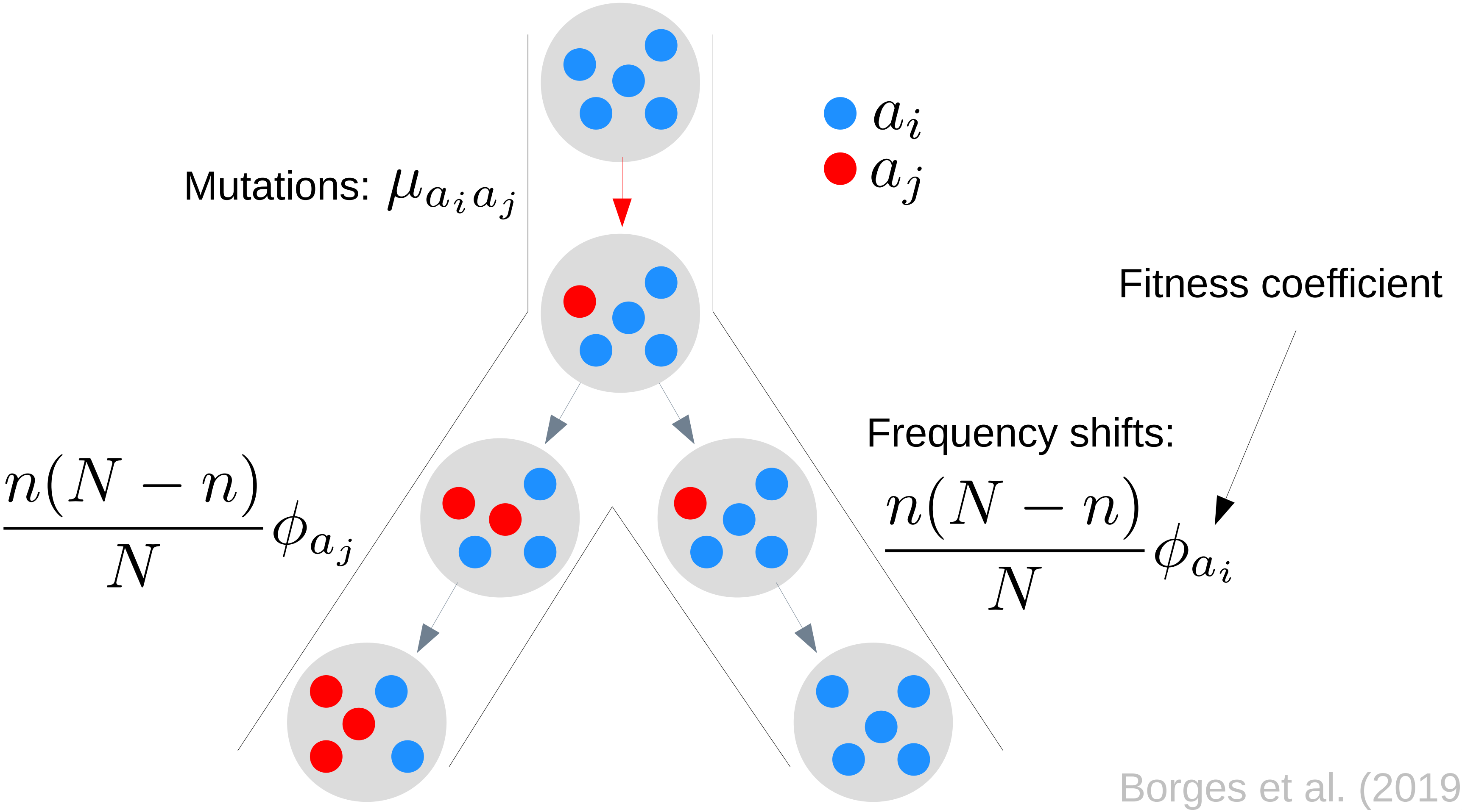
PoMos describe the evolution of several populations where changes in allele content and frequency are both possible.

- Mutational bias
- Genetic drift
- Selection
- ...



de Maio et al. (2013)

PoMos with selection



● a_i

● a_j

Mutations: $\mu_{a_i a_j}$

$$\frac{n(N-n)}{N}\phi_{a_j}$$

Fitness coefficient

Frequency shifts:

$$\frac{n(N-n)}{N}\phi_{a_i}$$

Borges et al. (2019)

Why including selection?

- Measuring allelic selection (e.g., nucleotides, or other evolutionary ‘units’)
- Accounting for nucleotide usage bias: e.g., GC-bias
- In RevBayes:

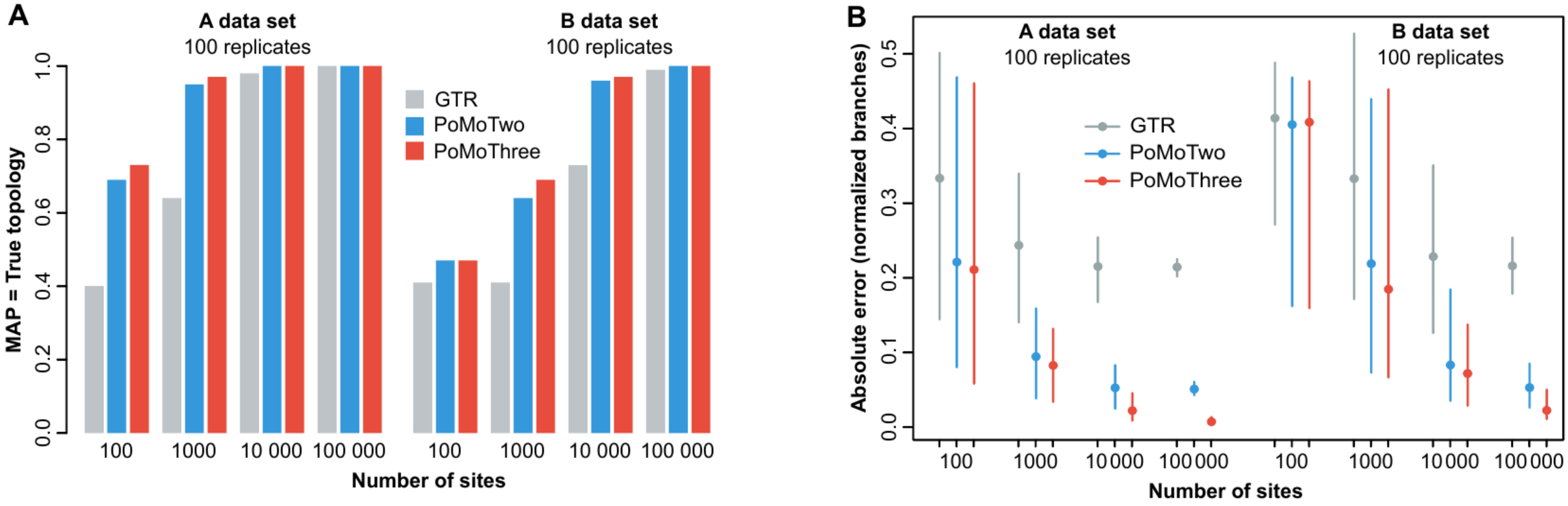
```
fnReversiblePoMo4N( N, pi, rho, phi )
fnPoMo4N( N, mu, phi )

fnReversiblePoMo2N( N, pi, rho, phi )
fnPoMo2N( N, mu, phi )

fnReversiblePoMoKN( N, pi, rho, phi )
fnPoMoKN( N, mu, phi )
```

Borges et al. (2019)

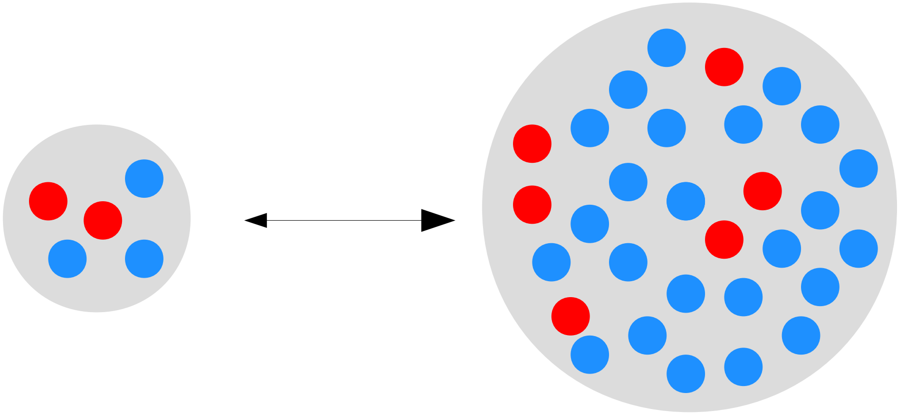
Why including selection?



Borges et al. (2020)

Virtual PoMos

- Using the effective population size is out of question : 10^4–10^6 in multicellular eukaryotes Lynch et al. (2016)
- Virtual population *versus* effective population



Idea behind the virtual PoMos

Effective dynamic

Virtual dynamic

Expected diversity

Virtual population size

M

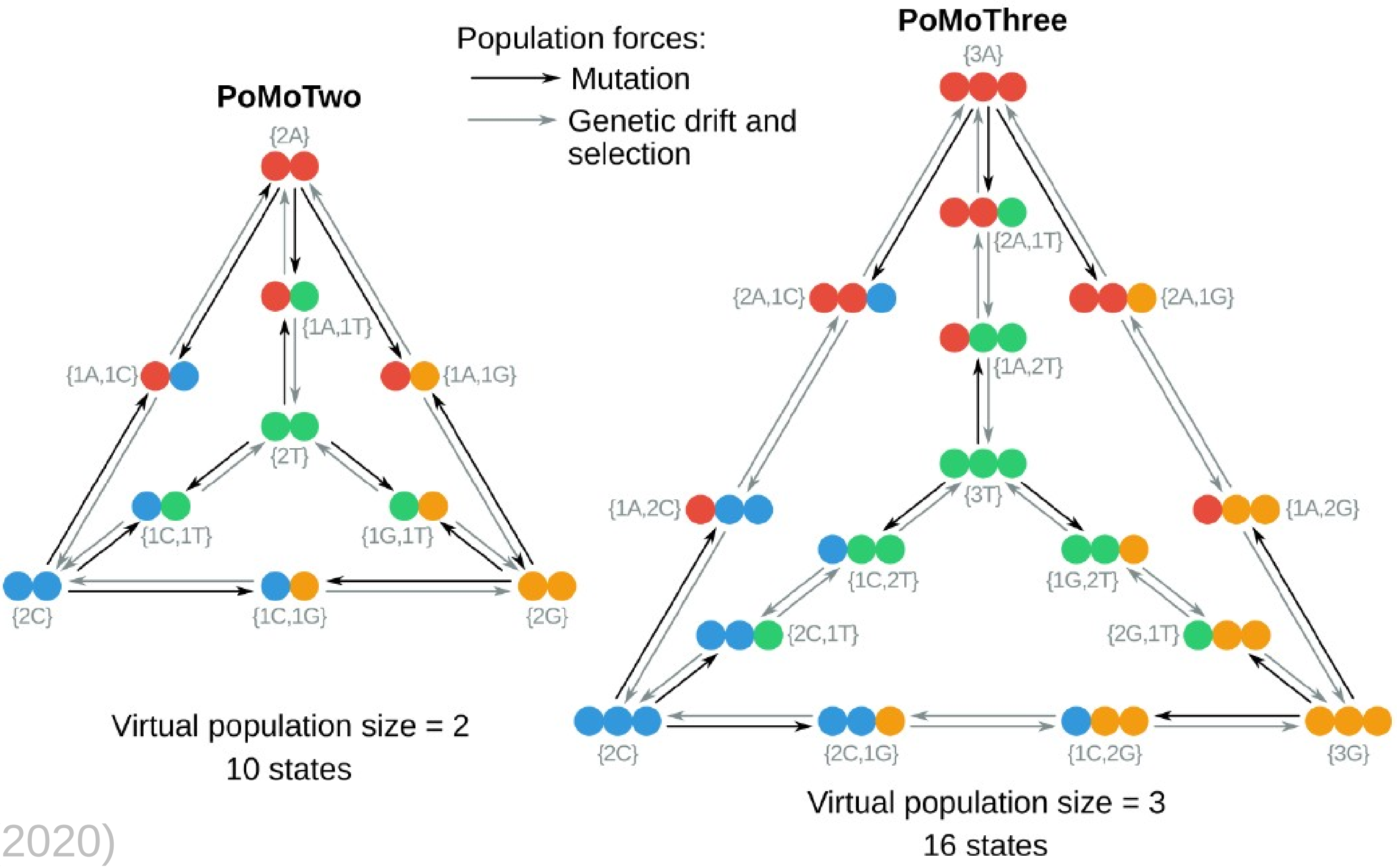
Effective population size

N

Neutral case: $\mu_{a_i a_j} = \frac{H_{M-1}}{H_{N-1}} \mu'_{a_i a_j}$

Borges et al. (2019)

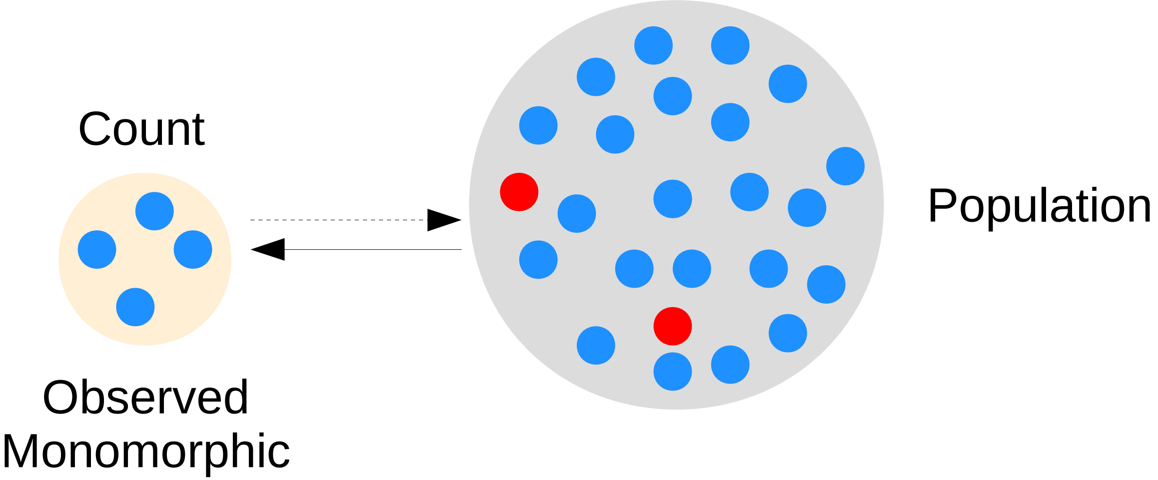
Virtual PoMos



Borges et al. (2020)

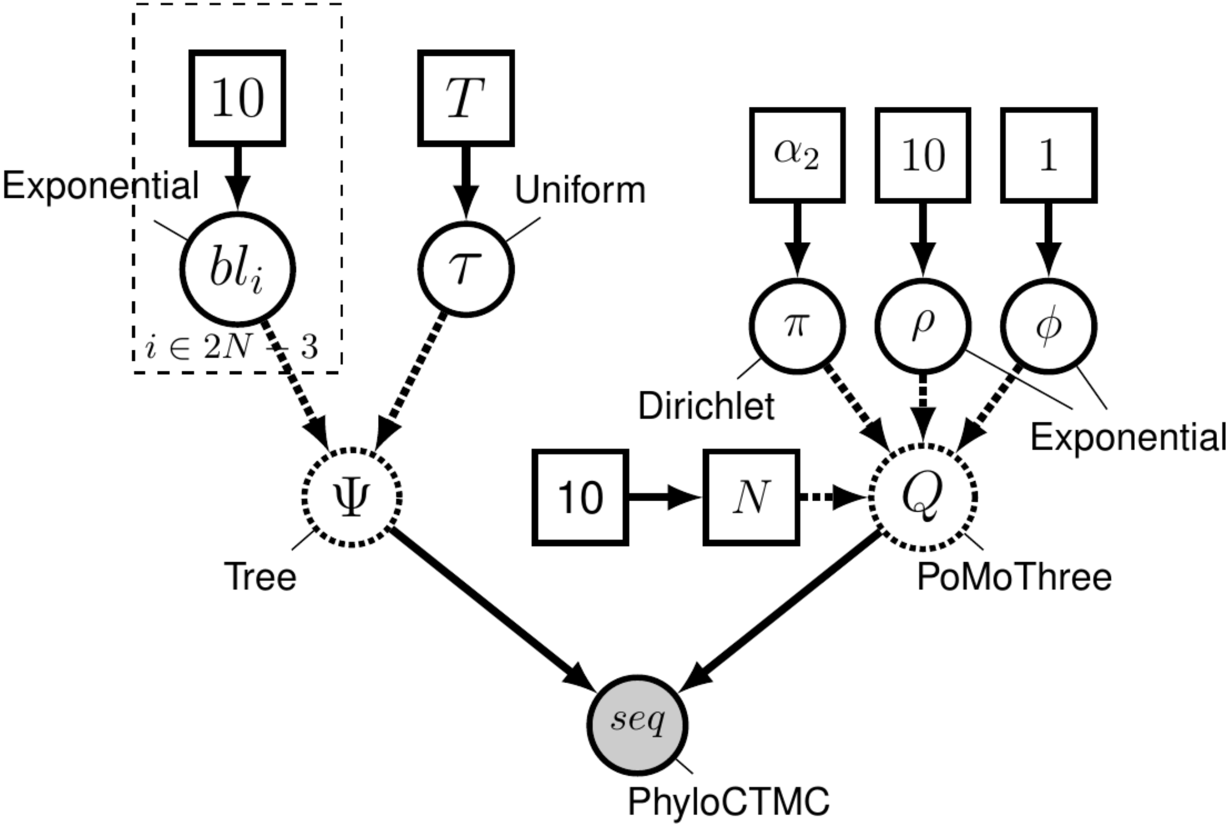
Weighted-Sampled method

- **Weighted method**: weight the PoMo states at each terminal node according to likelihood of the observed counts (e.g., using a binomial distribution) Schrempf et al. (2016)
- **Weighted-Sampled method**: weight the PoMo states and sample from them.



Let’s start!

DAG


10
T
Exponential
Uniform
bl_i
τ
$i \in 2N - 3$
α_2
10
1
π
ρ
ϕ
Dirichlet
Exponential
Ψ
10
N
Q
Tree
PoMoThree
seq
PhyloCTMC

Why Bayesian? Why RevBayes?

- Several highly maintained/used routines for parameter estimation and hypothesis testing
- Integrate different sources of information: e.g., mutation rates, population size, GC-bias rate, other biases
- Has sophisticated phylogenetic models: e.g., molecular clock, character evolution, …
- Powerful in combining different parametric hierarchies