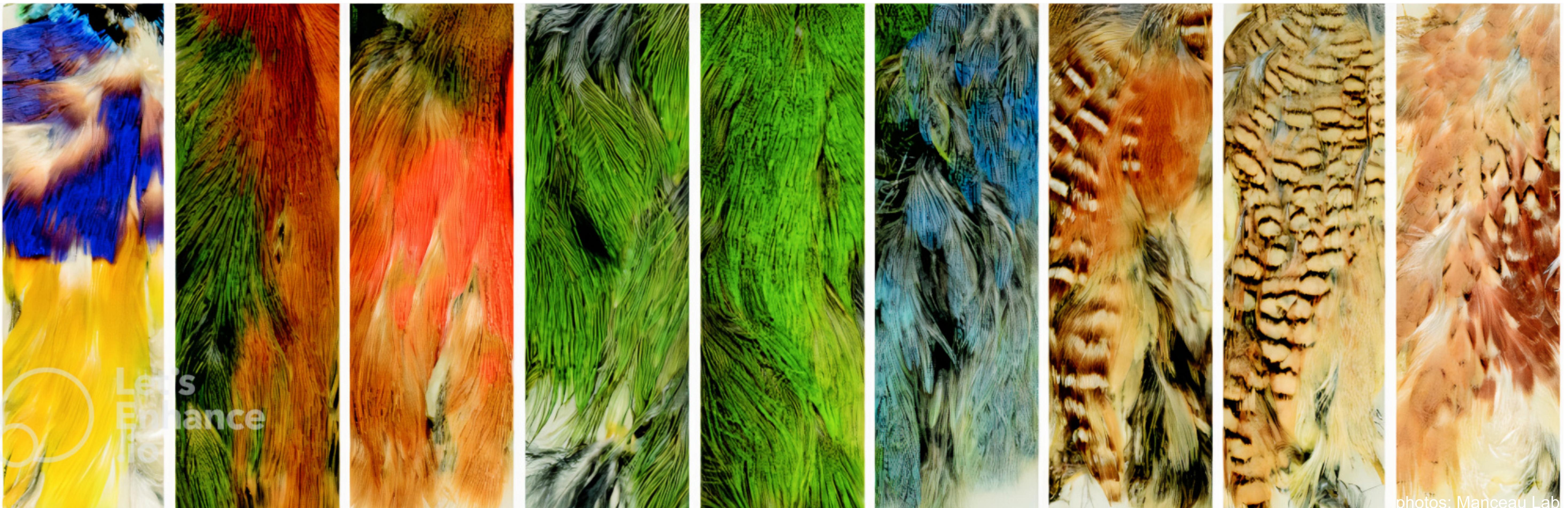
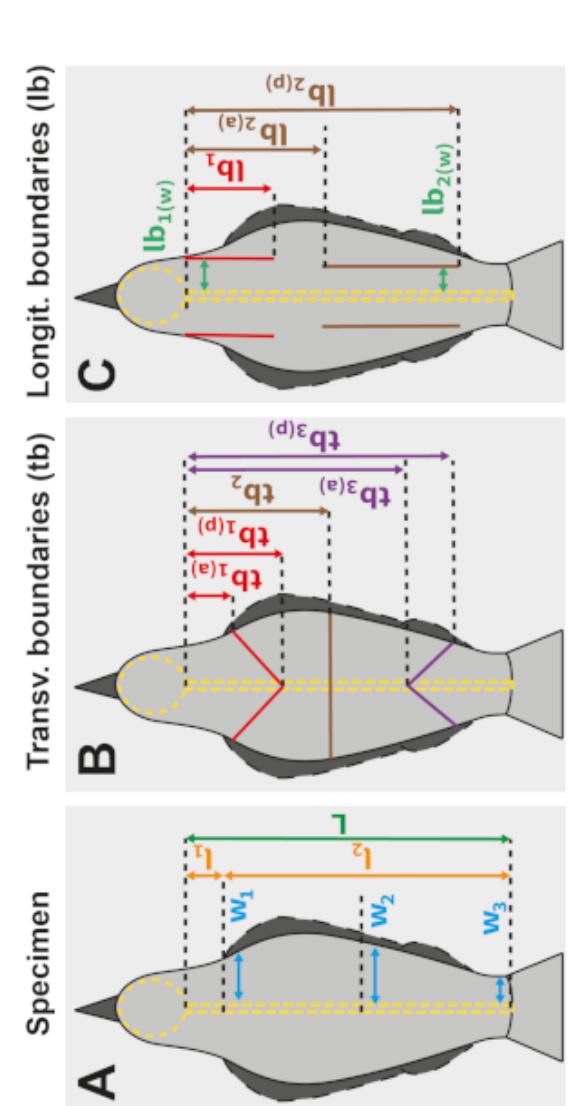
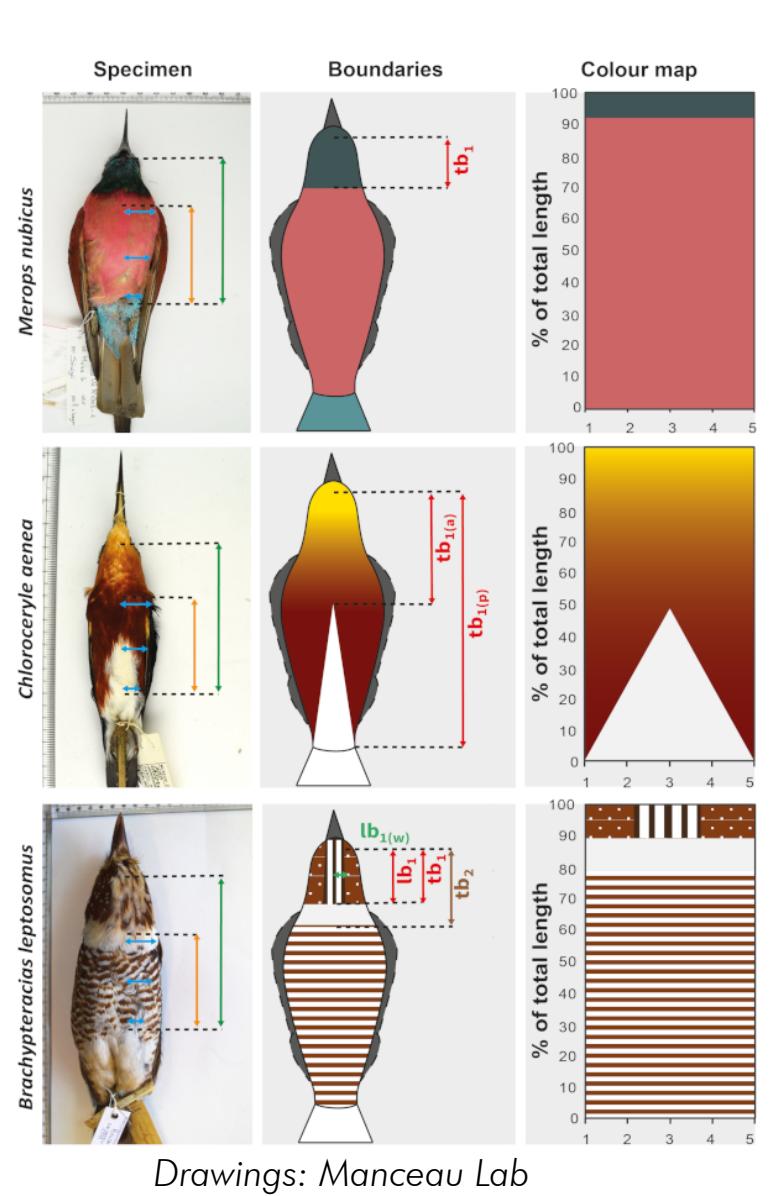


# Quantitative Tools for Assessing the Correlated Evolution of Genes and Phenotypes: Application to Birds' Color Patterns (Q-T-Birds)

Viktor Senderov, Amaury Lambert, Marie Manceau, Carole Desmarquet, Caitlyn Jean-Baptiste, Ingrid Lafontaine, Hélène Morlon

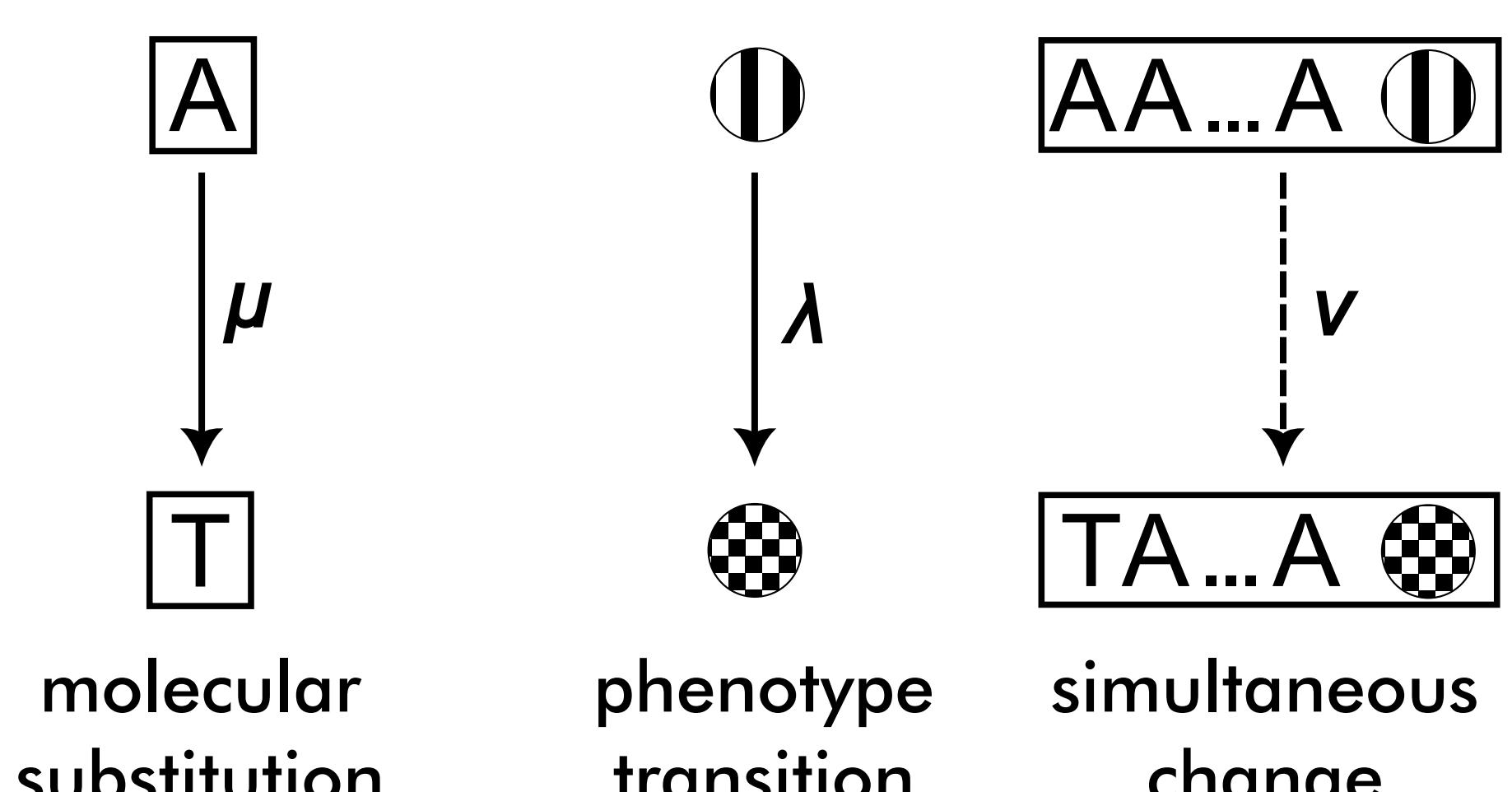


We use probabilistic programming languages (PPLs) to uncover the phylogenomic basis of bird color pattern complexity



Drawings: Manceau Lab

Each bird species follows a general schema for pattern formation



In the QT-Birds model we assume three rates ( $\lambda$  - phenotype change,  $\mu$  - molecular change,  $\nu$  - simultaneous change) to be unknown. We infer the rates using phenotypic, genomic, and phylogenetic data. We have a discrete phenotypic character and a genomic locus of length N.

**Algorithm sketch**  
 $Q$  - rate matrices;  
 $P$  - prob. matrices  
Branch of length  $t$   
Evolve each nucleotide  $i$ :  

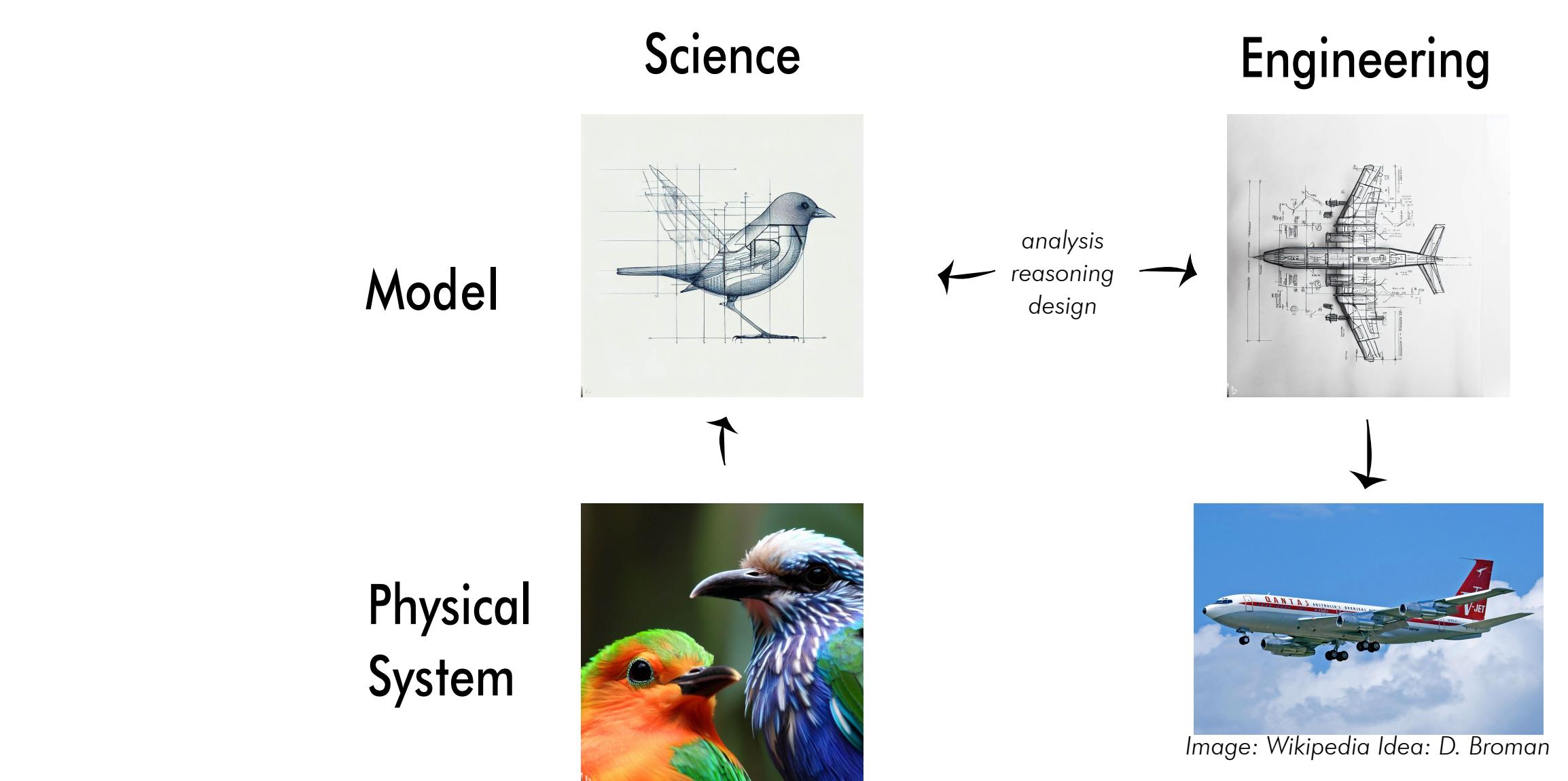
$$u \leftarrow \frac{1}{N}$$

$$n_i \sim \text{Poisson}(uvt)$$

$$P_\mu(t) = P_\mu^{n_i} \exp(\mu t Q_\mu)$$
  
Evolve phenotype:  

$$n_s \leftarrow \sum_{i=1}^N n_i$$

$$P_\lambda(t) = P_\lambda^{n_s} \exp(Q_\lambda \lambda t)$$

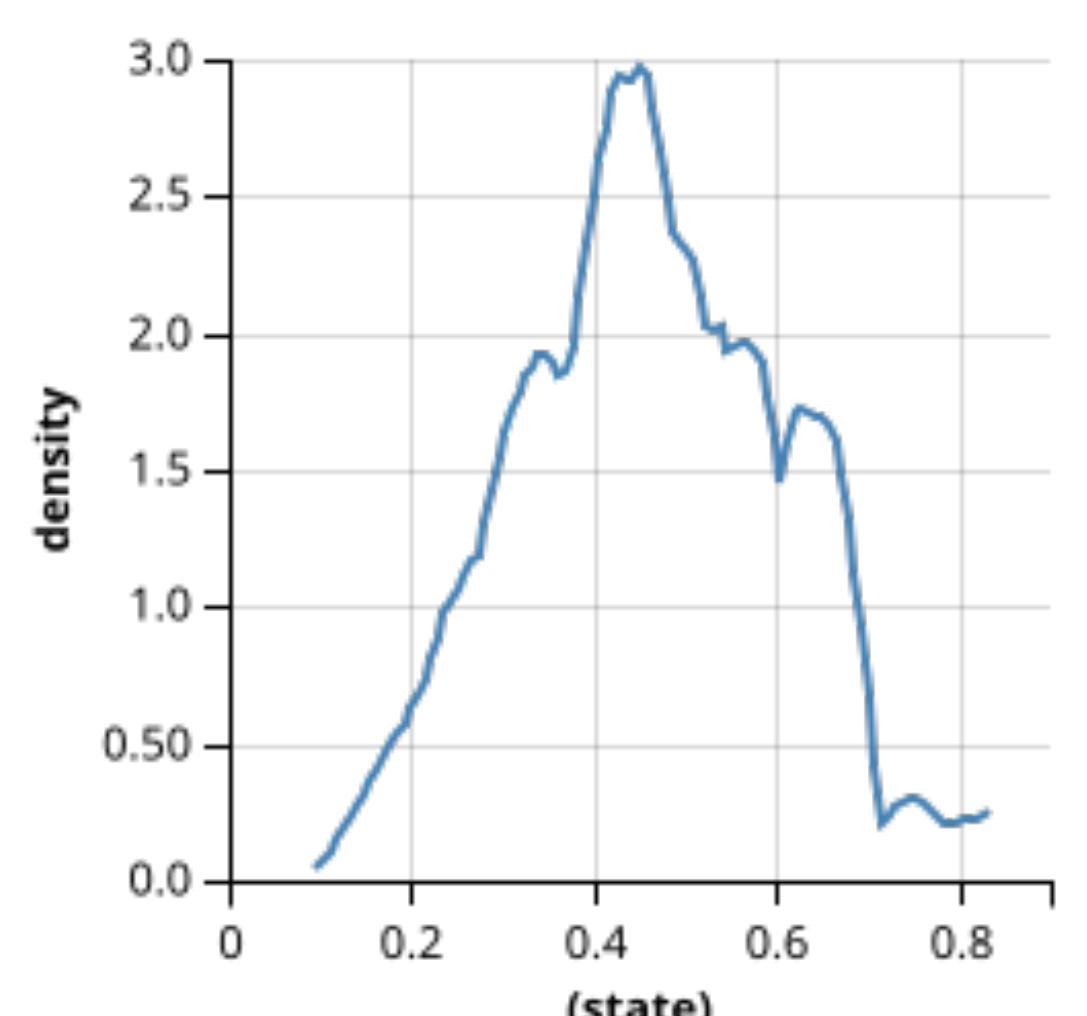


• Model comparison • Inference of unobserved parameters • Sampling

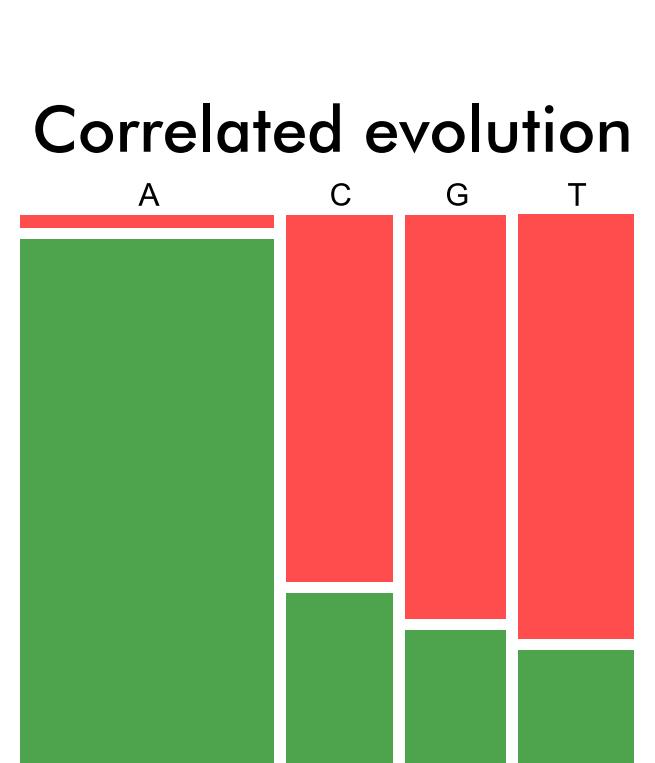
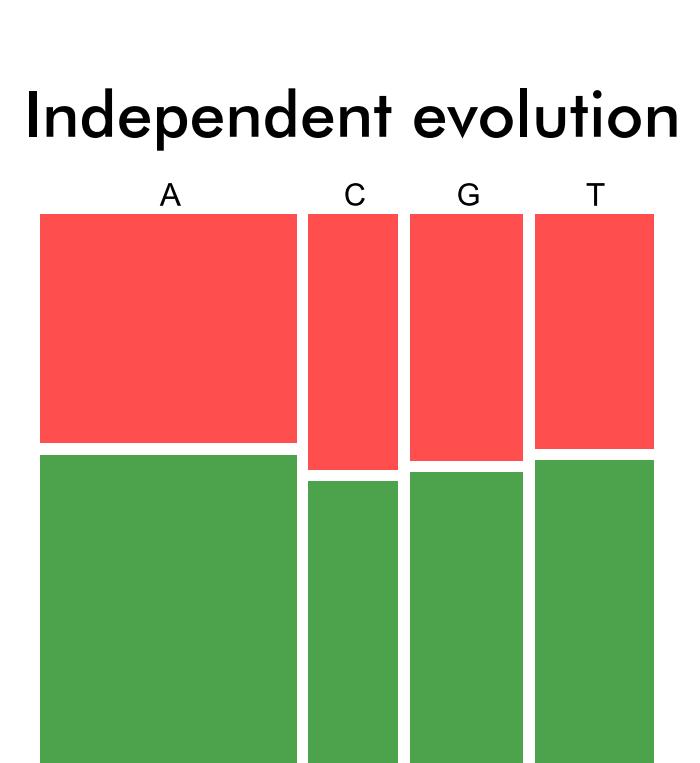
Universal PPLs offer a holistic approach to statistical modeling



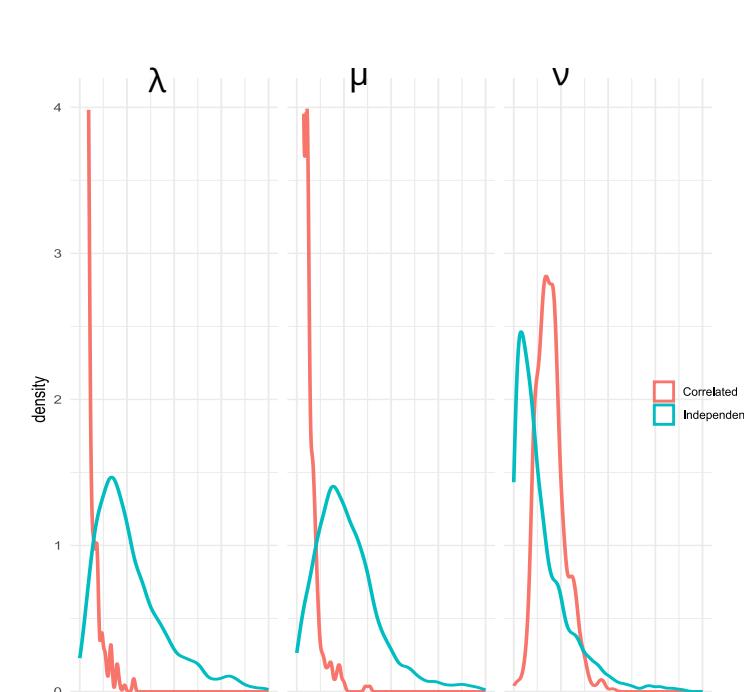
```
var data =
[false, false, true, false, false, true, true, false, true]
var model = function() {
  var p = sample(Beta({a: 1, b: 1}))
  map(function(d) {
    observe(Bernoulli({p: p}), d)
  }, data)
  return p
}
var dist = Infer({method: 'SMC', model: model, particles: 1000})
dist
```



Is this a fair coin?



Data simulated under the QT-Birds model under a regime of independent evolution and correlated evolution forms different patterns.

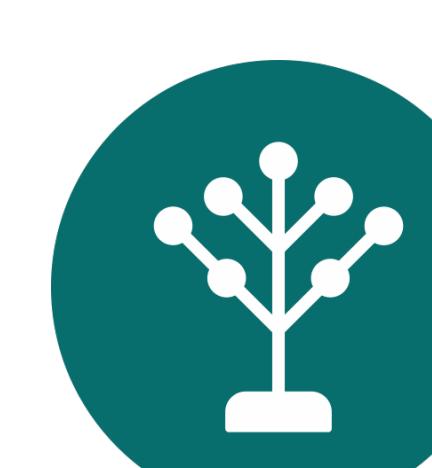


Inferring back the evolutionary rates correctly recovers high  $\nu$ , low  $\mu$ , low  $\lambda$  for correlated evolution, and low  $\nu$ , high  $\mu$ , high  $\lambda$  for independent evolution.

Experiments on simulated data



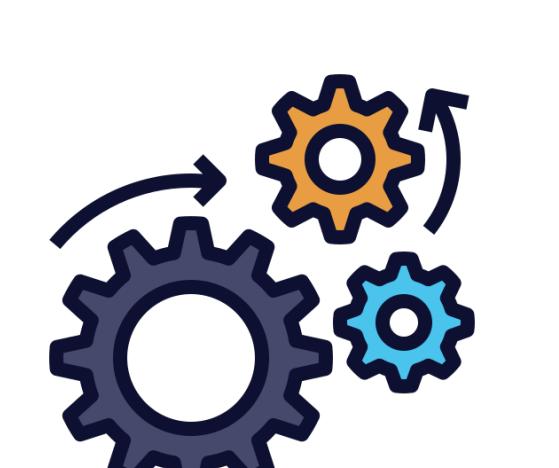
Simplicity  
Designed to meet the needs of computational biologists



Phylogenetic Data  
Supports natively the PhyJSON format for evolutionary trees



Rich Model Library  
Offers state-of-the-art diversification models as templates



Powerful Statistical Inference  
Sequential Monte-Carlo (SMC) and Markov-chain Monte-Carlo (MCMC) inference

**TreePPL** - a Probabilistic Programming Language for Statistical Phylogenetics

