WASP | WALLENBERG AI, AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM

digital futures

KTH VETENSKAP OCH KONST

Annotated Automatic Pruning in Miking CorePPL

Gizem Caylak
Collaborators: David Broman, Emma Granqvist, Fredrik Ronquist, Tim Virgoulay

2025-12-05

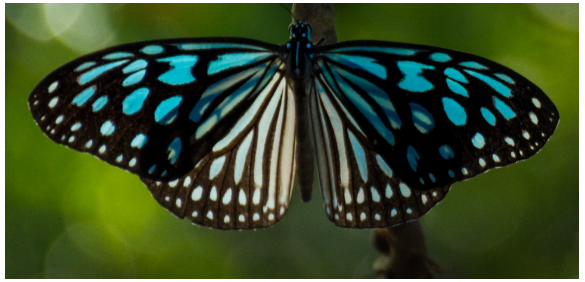
This work was partially supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation.

Introduction: Phylogenetics Tree Inference

- Understanding the relation between species

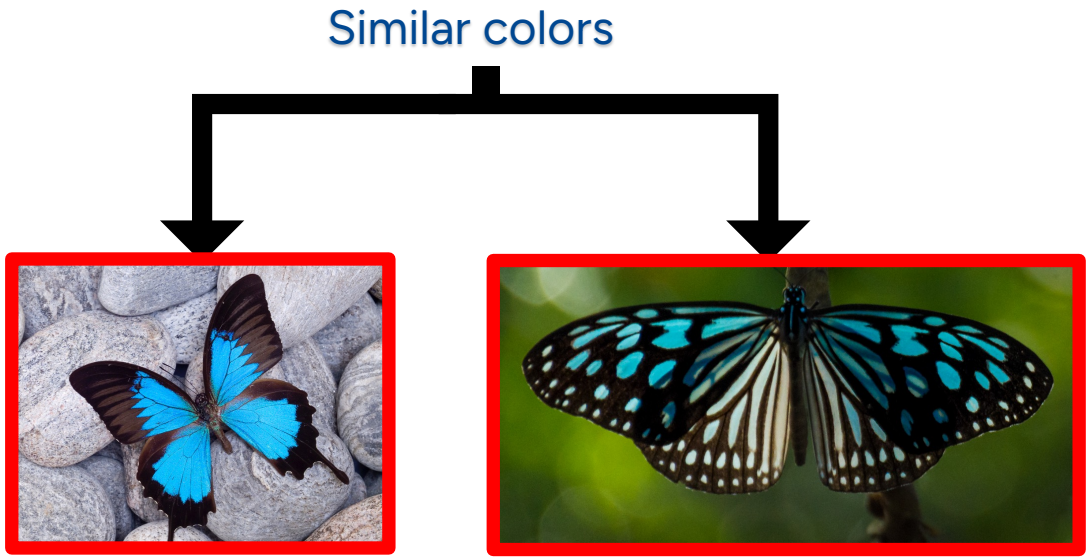






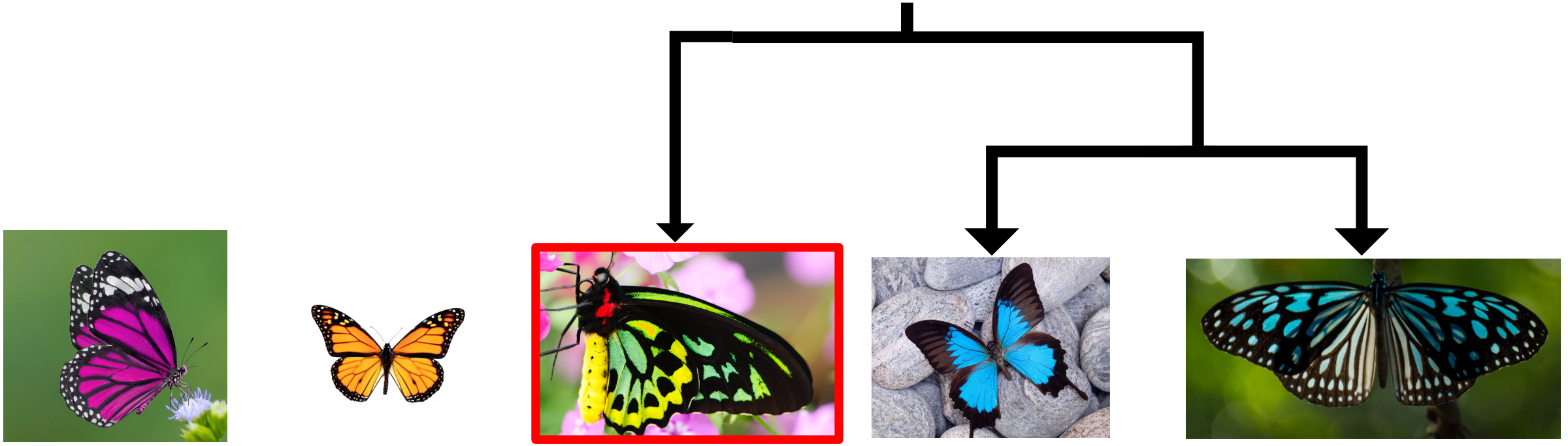
Introduction: Phylogenetics Tree Inference

- Understanding the relation between species



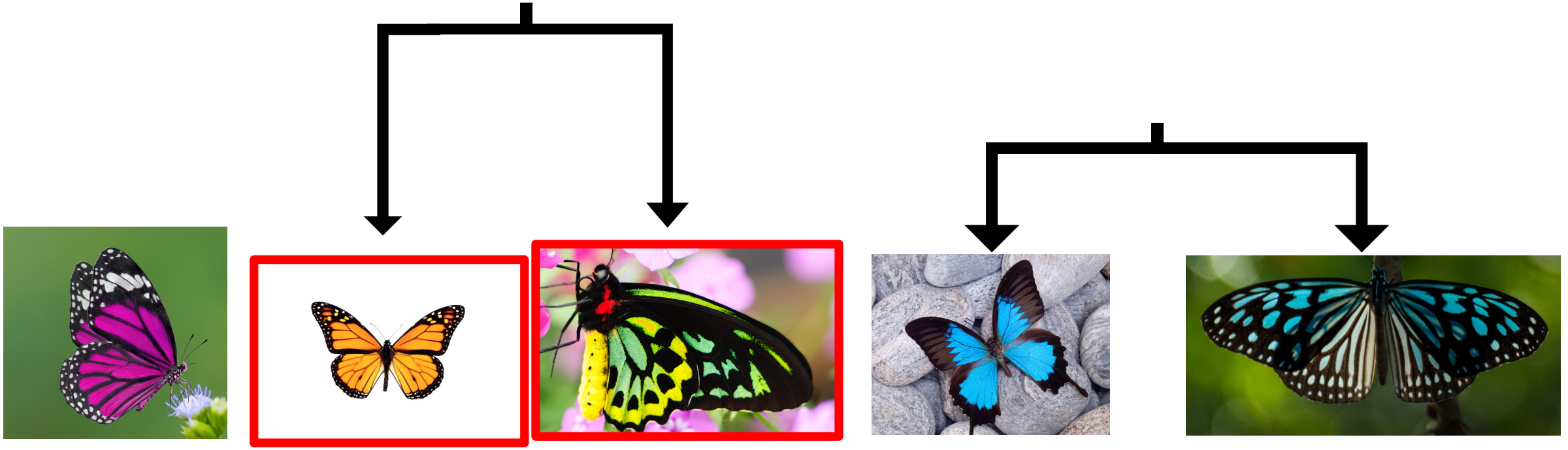
Introduction: Phylogenetics Tree Inference

- Understanding the relation between species



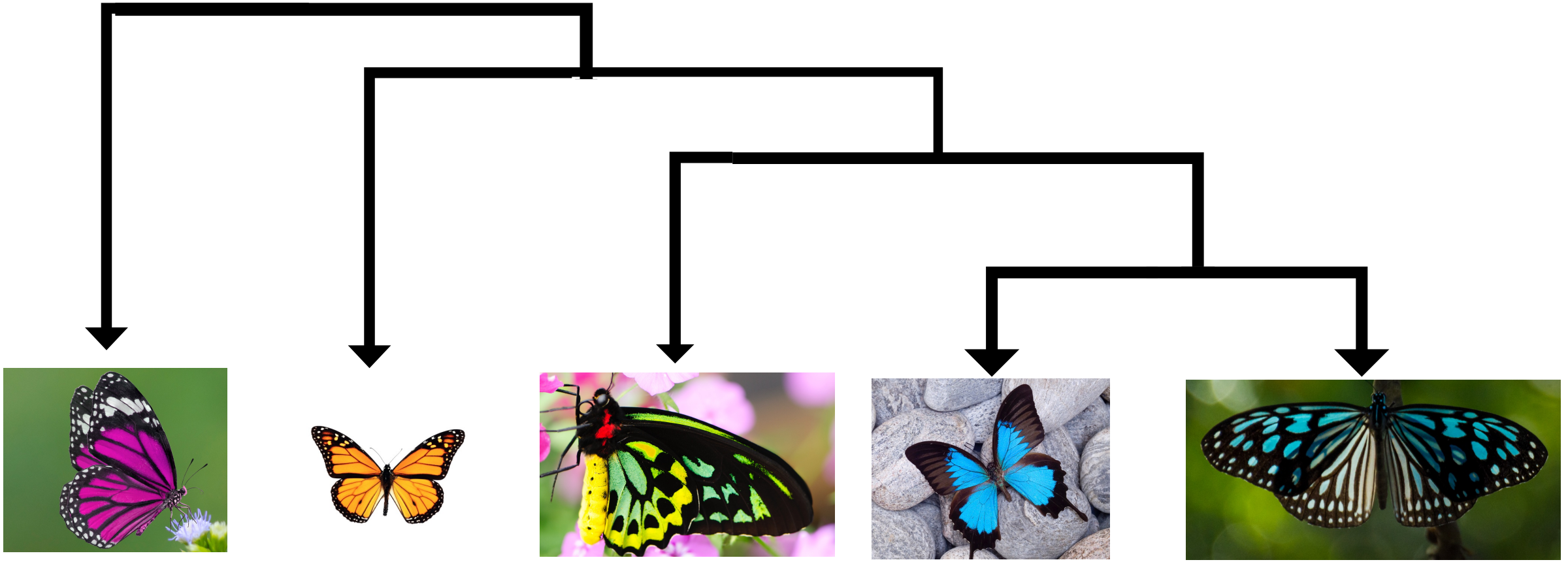
Introduction: Phylogenetics Tree Inference

- Understanding the relation between species



Introduction: Phylogenetics Tree Inference

- Understanding the relation between species

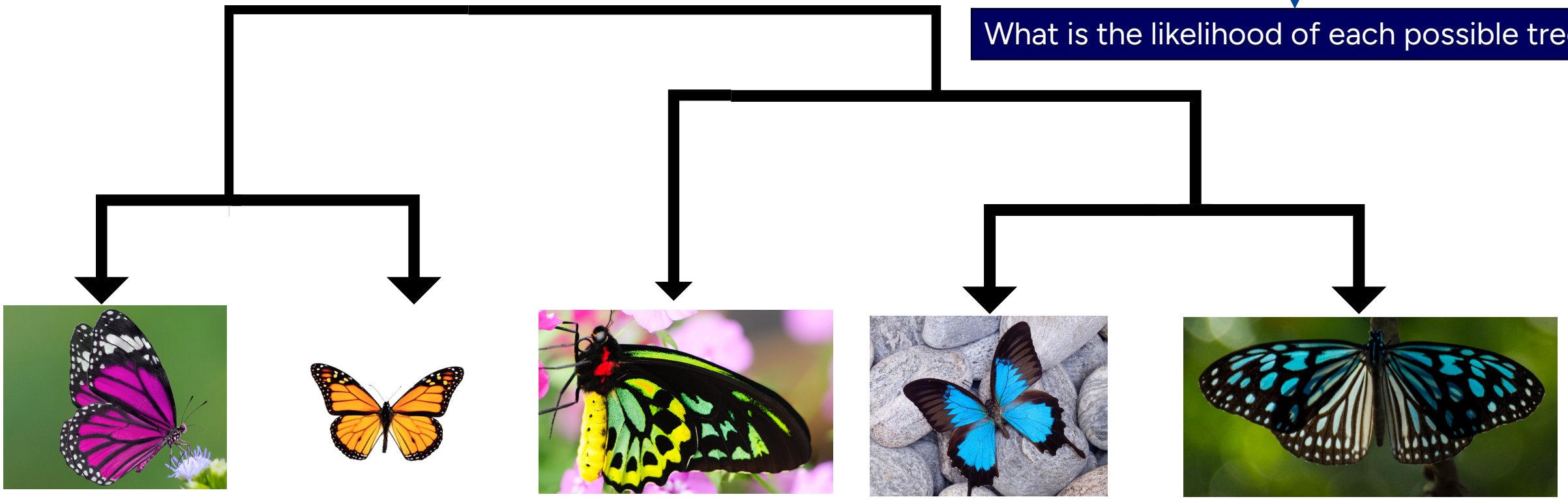


Introduction: Phylogenetics Tree Inference

- Understanding the relation between species
- Tree inference is a fundamental problem

Which tree is more likely?

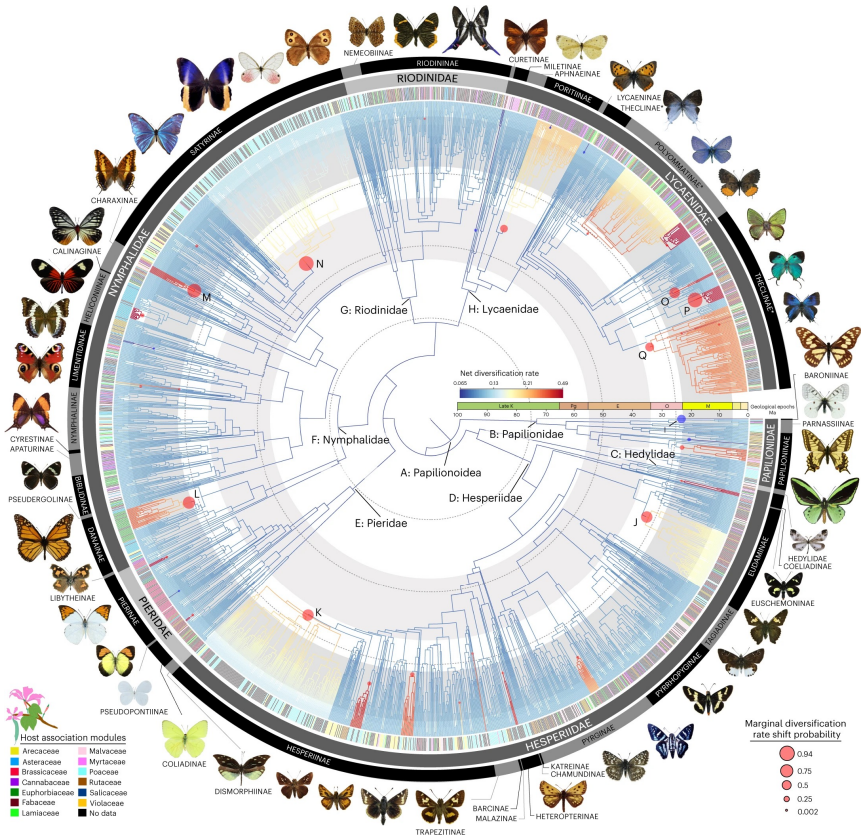
What is the likelihood of each possible tree?



Introduction: Phylogenetics Tree Inference

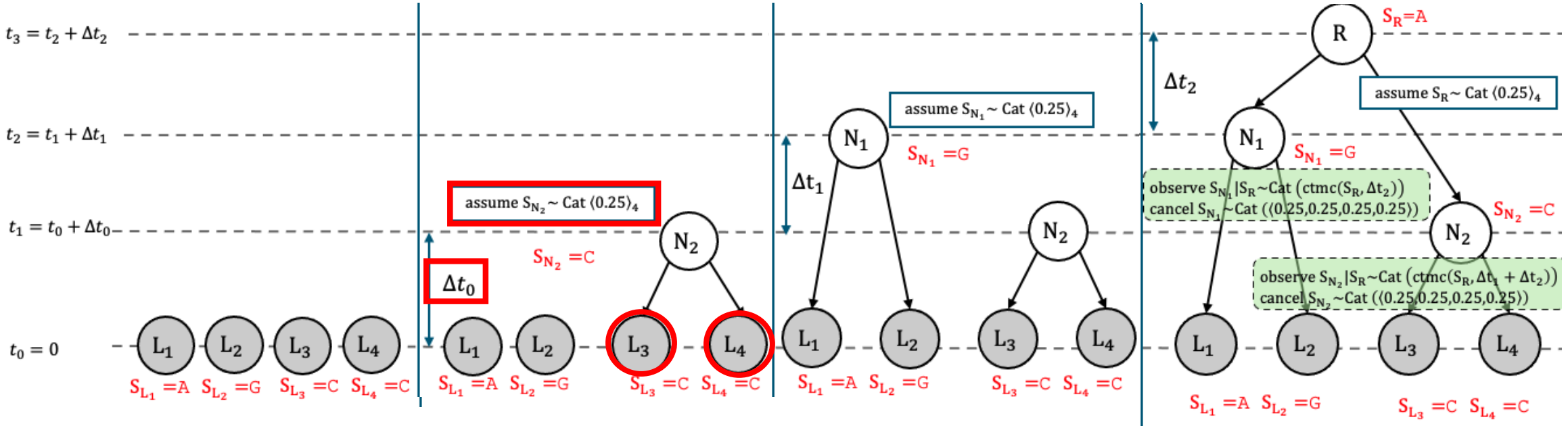
- Understanding the relation between species
- Tree inference is a fundamental problem
- Using genetic data makes the problem harder

DNA may consist of billions of nucleotides (features)



The figure is retrieved from: Kawahara, A.Y., Storer, C., Carvalho, A.P.S. *et al.* A global phylogeny of butterflies reveals their evolutionary history, ancestral hosts and biogeographic origins. *Nat Ecol Evol* **7**, 903–913 (2023). https://doi.org/10.1038/s41559-023-02041-9

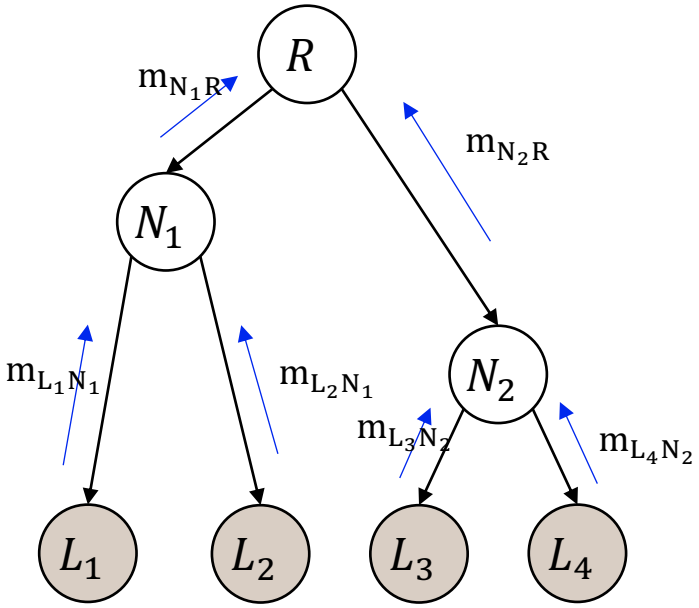
Naive Backward Tree Inference Problem



The likelihood of observing C at L_3 and L_4 given Δt_0 and the parent's genetic sequence

- Internal nodes are **latent** (unobserved)
- We **sample** them during inference

Belief Propagation on Tree Inference Problem

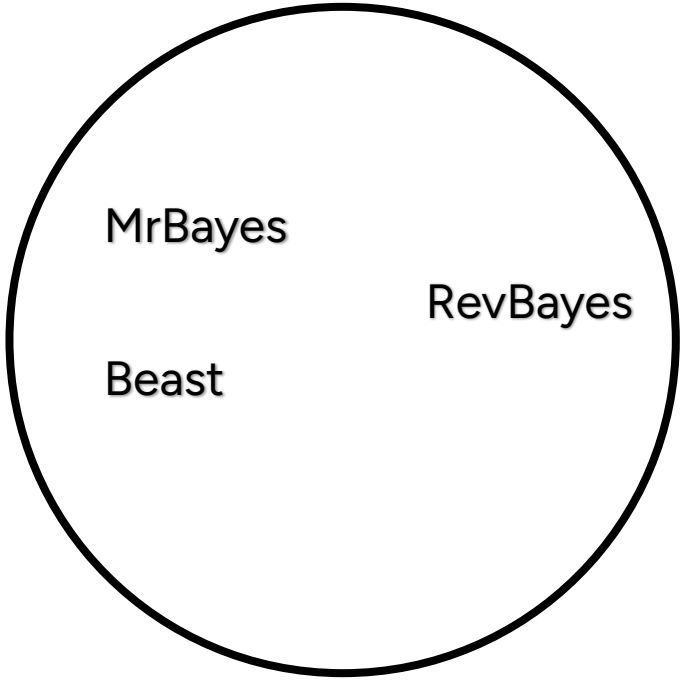


Propagate the messages to the root to calculate the likelihood of observations given the tree topology

Marginalize out the internal nodes:

Intuitively, we sum over all possible states that the internal node can take, weighted by their probabilities.

Phylogenetic tools

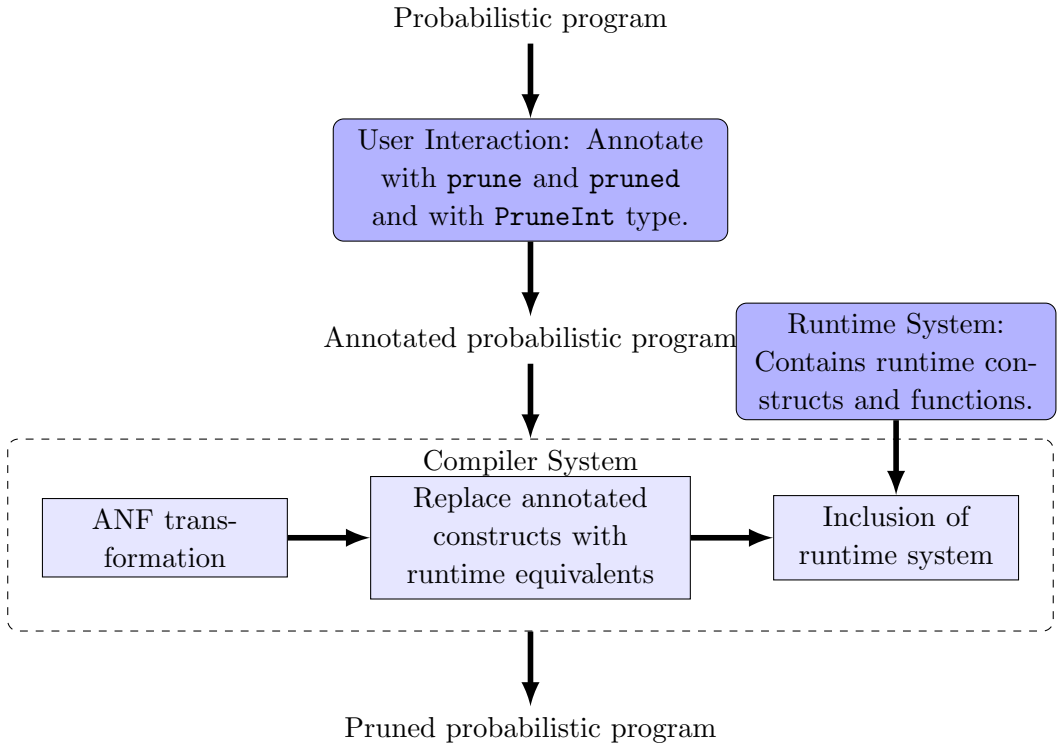


Efficient; however:

- Hardcoded into the model
- Restricted language

Probabilistic Programming Languages

- PPLs provide **flexibility**.
- However, flexibility makes it **hard to implement** domain-specific **optimizations** like pruning.



Probabilistic Programming Languages

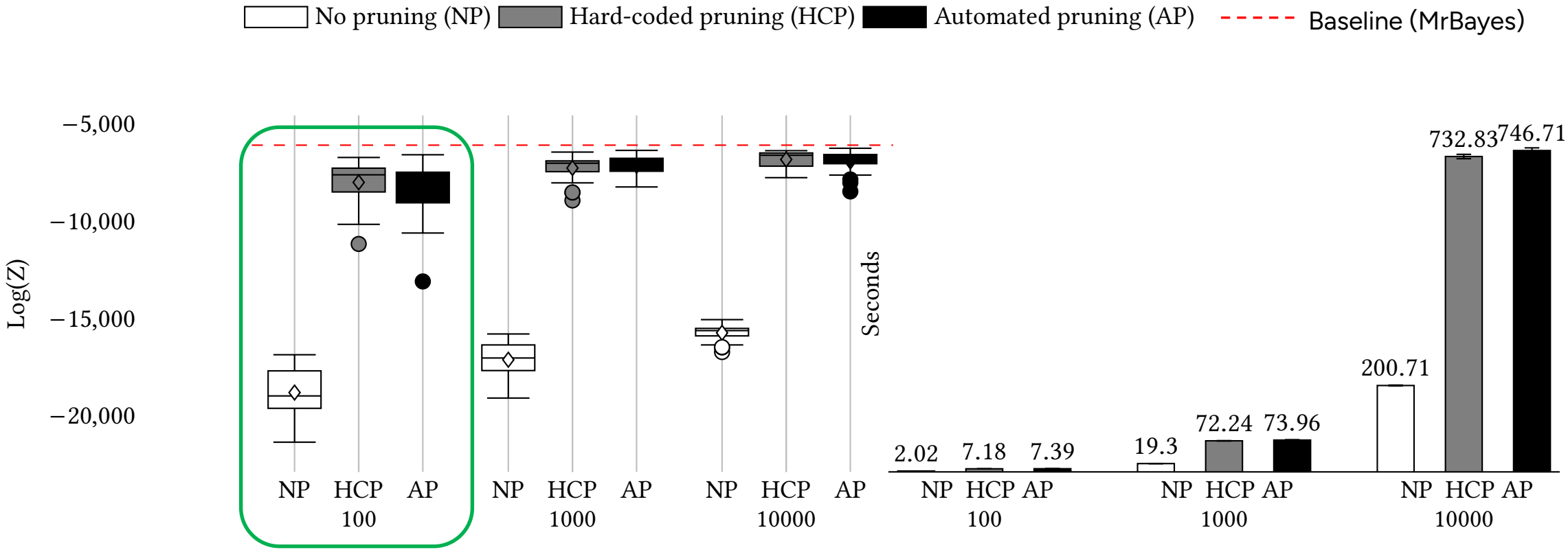
- PPLs provide **flexibility** - users write models without worrying about inference details.
- However, flexibility makes it hard to implement domain-specific optimizations like pruning.

```
(...)
con Node:{age: Float, seq: [PruneInt], left: Tree, right: Tree} -> Tree
con Leaf:{age: Float, seq: [Int]} -> Tree (...)
recursive let cluster = lam trees. lam maxAge. lam seqLen.(...)
  let t = assume (Exponential 10.0) in
  let age = addf t maxAge in
  let parentSeq = iid (lam p. prune (Categorical p)) [0.25,0.25,0.25,0.25] seqLen in
  iteri (lam i:Int. lam site:PruneInt. iter (lam child.
    let deltaT = (subf age (getAge child)) in
    let p = ctmc (pruned site) deltaT in
    match child with Node n then
      let s = get n.seq i in
      observe (pruned s) (Categorical p);
      cancel (observe (pruned s) (Categorical [0.25,0.25,0.25,0.25]))
    else match child with Leaf l in
      let s = get l.seq i in
      observe s (Categorical p);
      cancel (observe s (Categorical [0.25,0.25,0.25,0.25]))
  ) [leftChild, rightChild]) parentSeq;
  let parent = Node {age=age, seq=parentSeq,left=leftChild, right=rightChild} in(...)
```

Where we marginalize out internal nodes

Results

Generalized time reversible model– Primates data



(a) Log normalizing constants.

(b) Execution times

RESEARCH-ARTICLE | OPEN ACCESS |

Annotated Automatic Pruning of Universal Probabilistic Programming Languages

Authors: Gizem Çaylak, Emma Granqvist, Thimothée Virgoulay, Fredrik Ronquist, David Broman | Authors Info & Claims

ACM Transactions on Probabilistic Machine Learning, Volume 1, Issue 3 • Article No.: 15, Pages 1 - 33
https://doi.org/10.1145/3731457

Published: 26 August 2025 Publication History

Thank you!