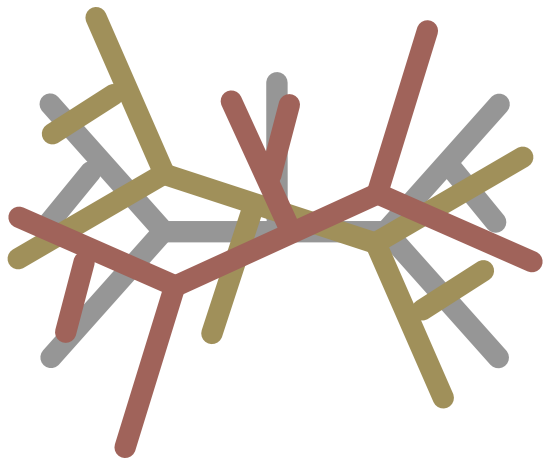


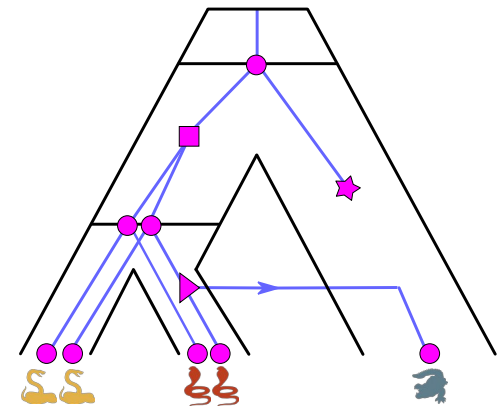
AleRax: phylogenetic inference from gene tree distributions

Benoit Morel

Heidelberg Institut for Theoretical Studies (HITS)
Karlsruhe Institute of Technology (KIT)



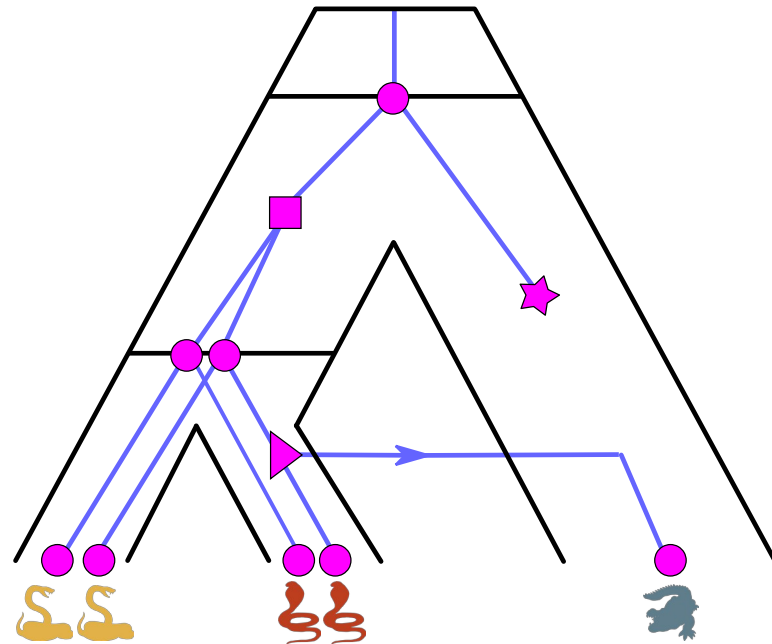
Oslo
23.03.2023



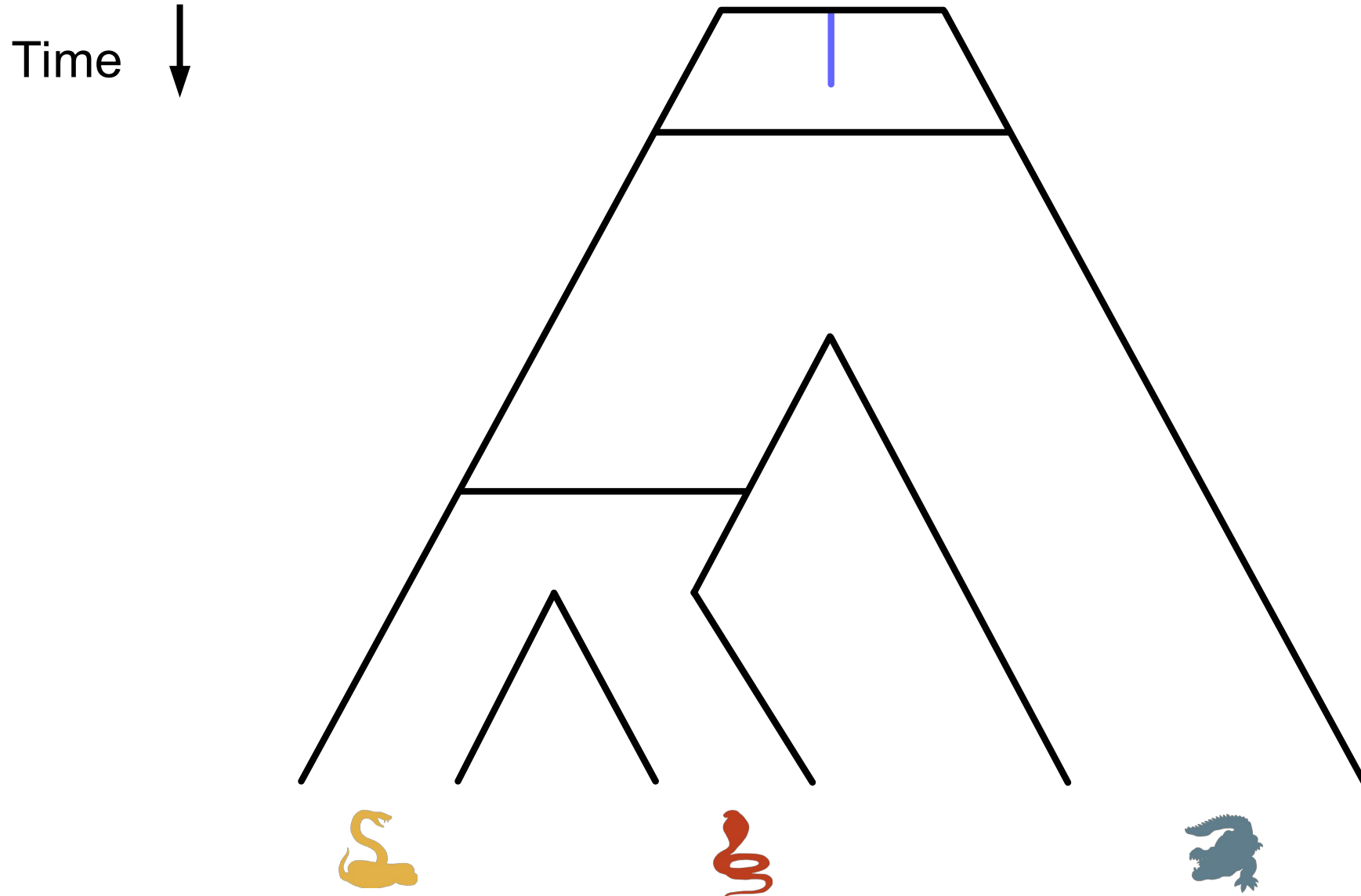
Motivation

Infer:

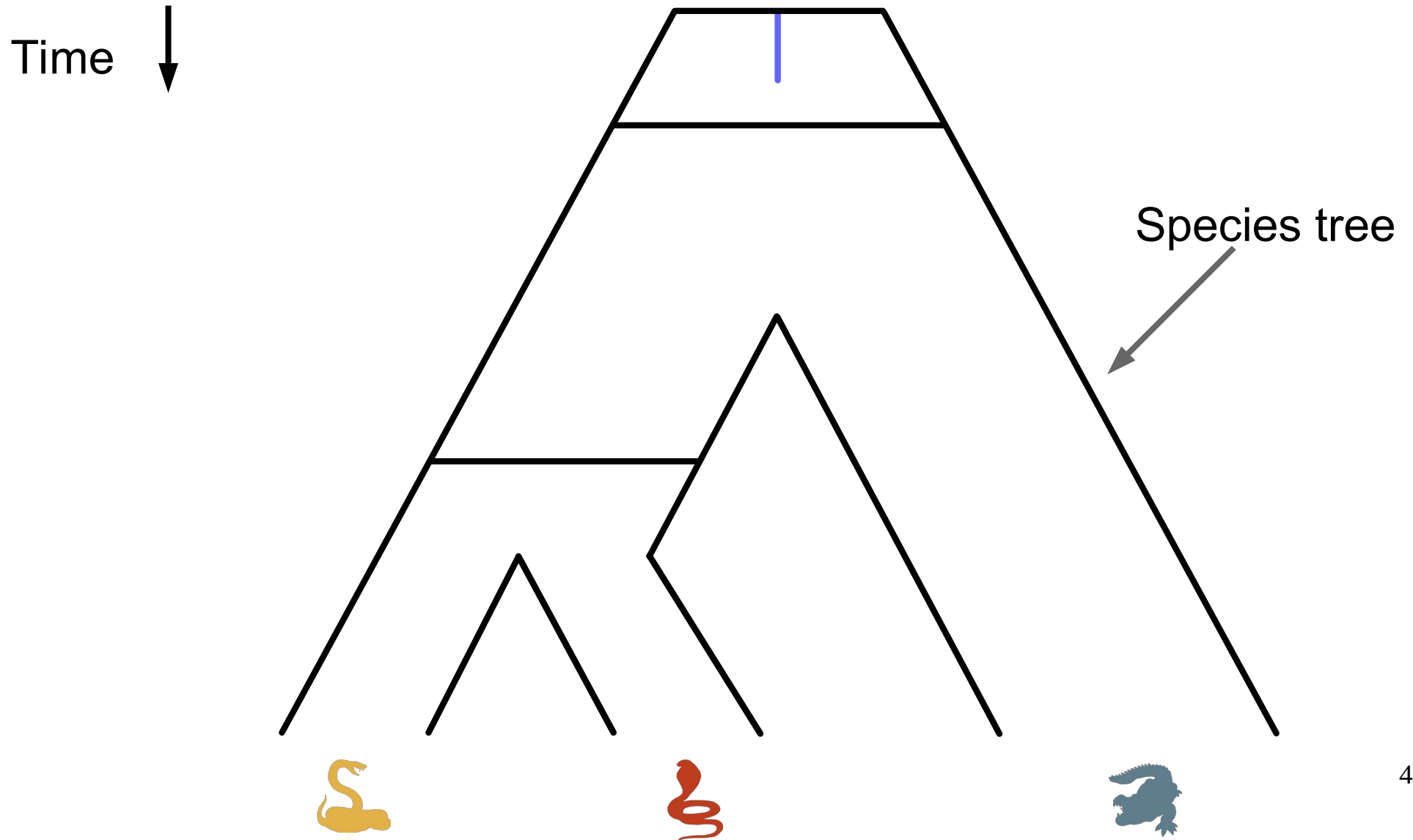
- Gene tree and species tree
- Gene duplication, transfer and loss events



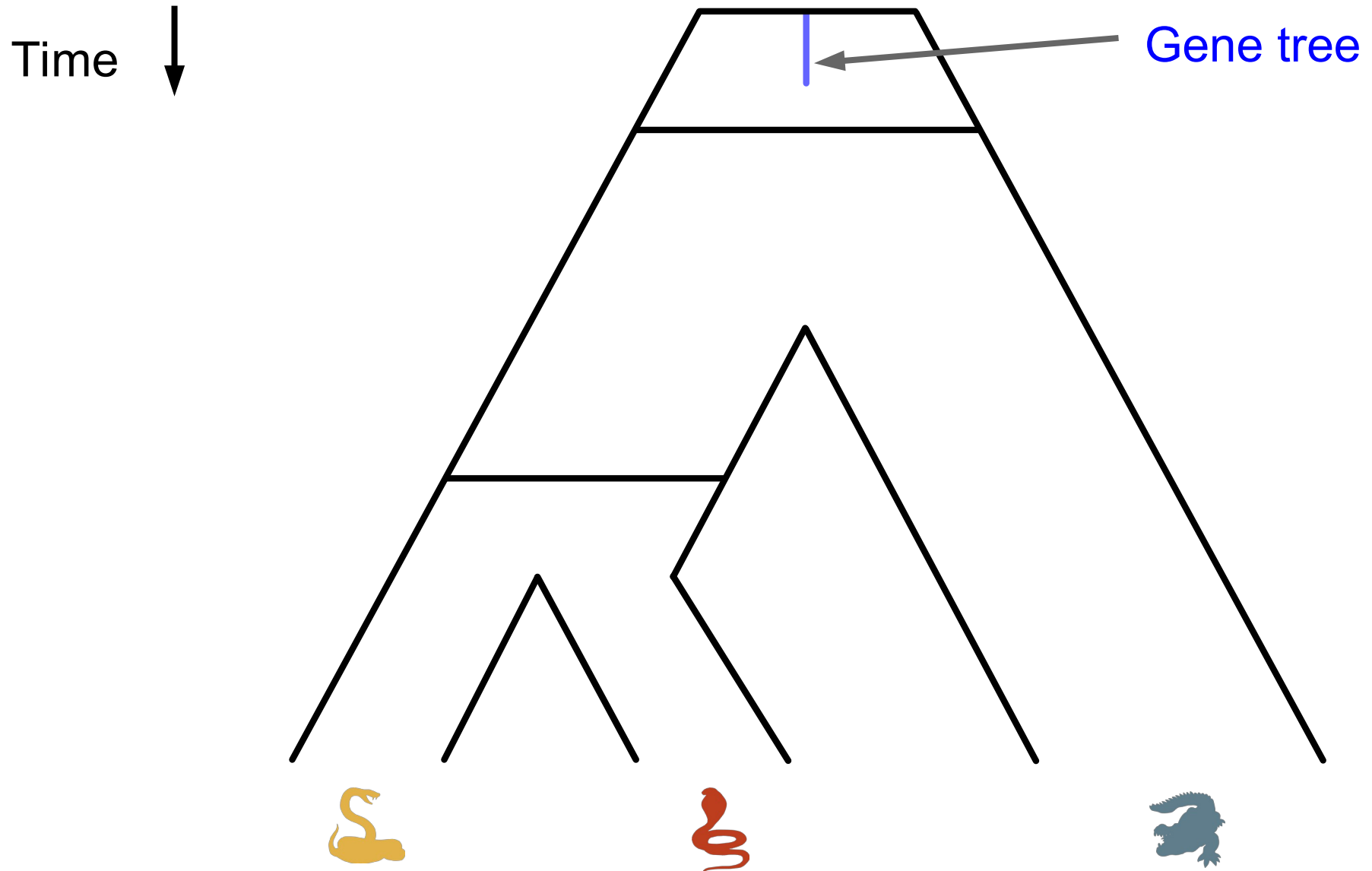
Gene evolution



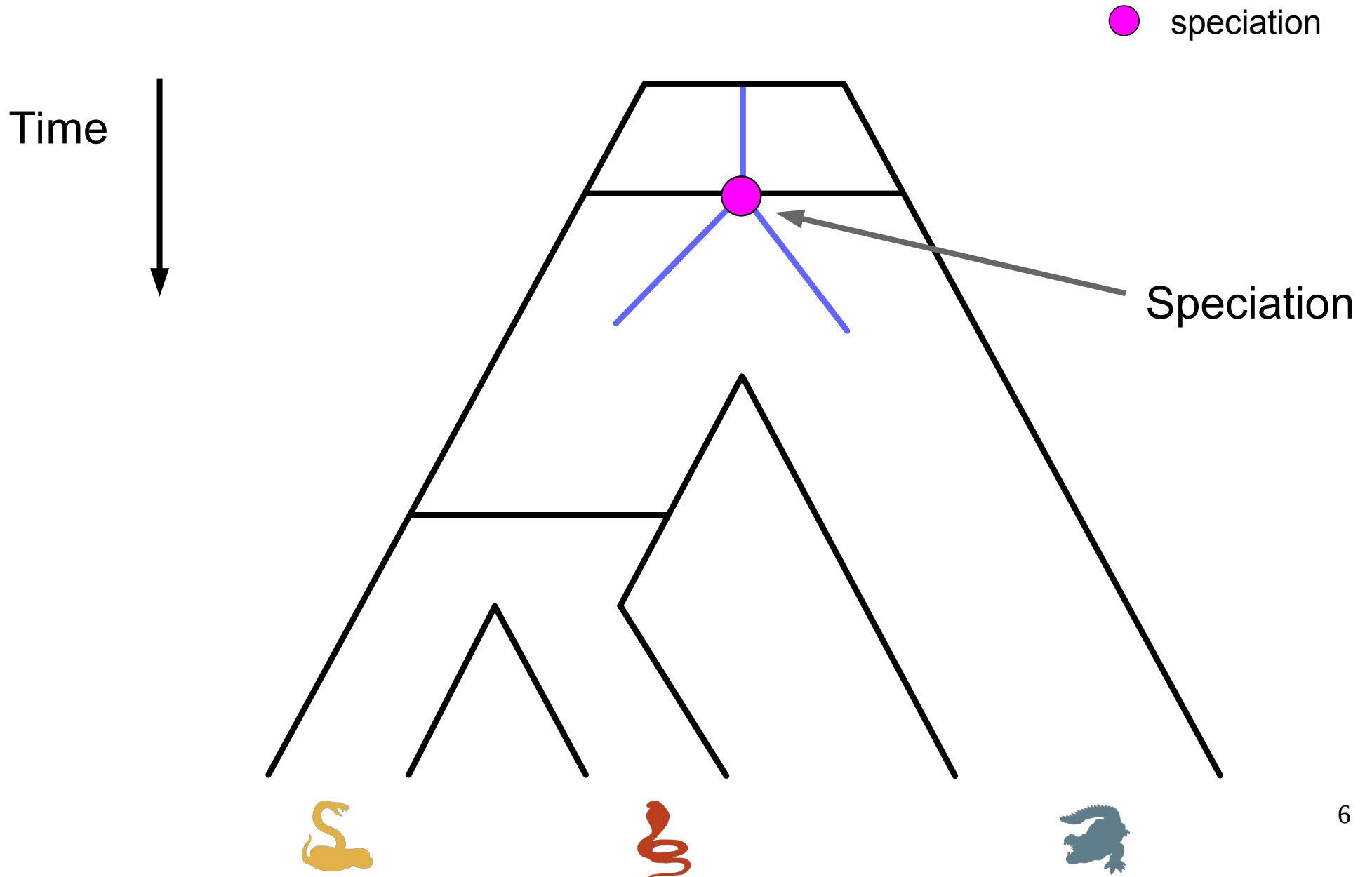
Gene evolution



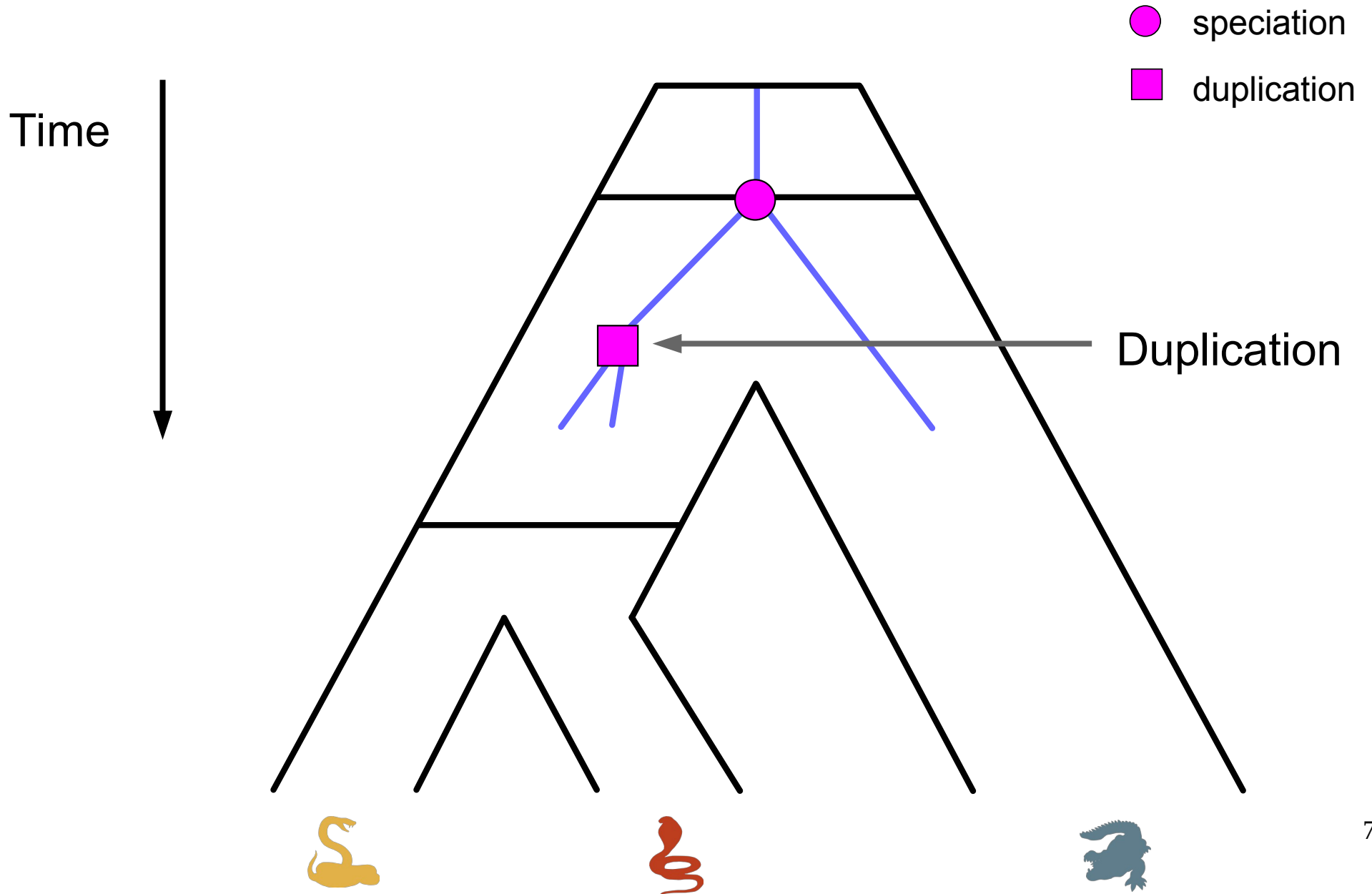
Gene evolution



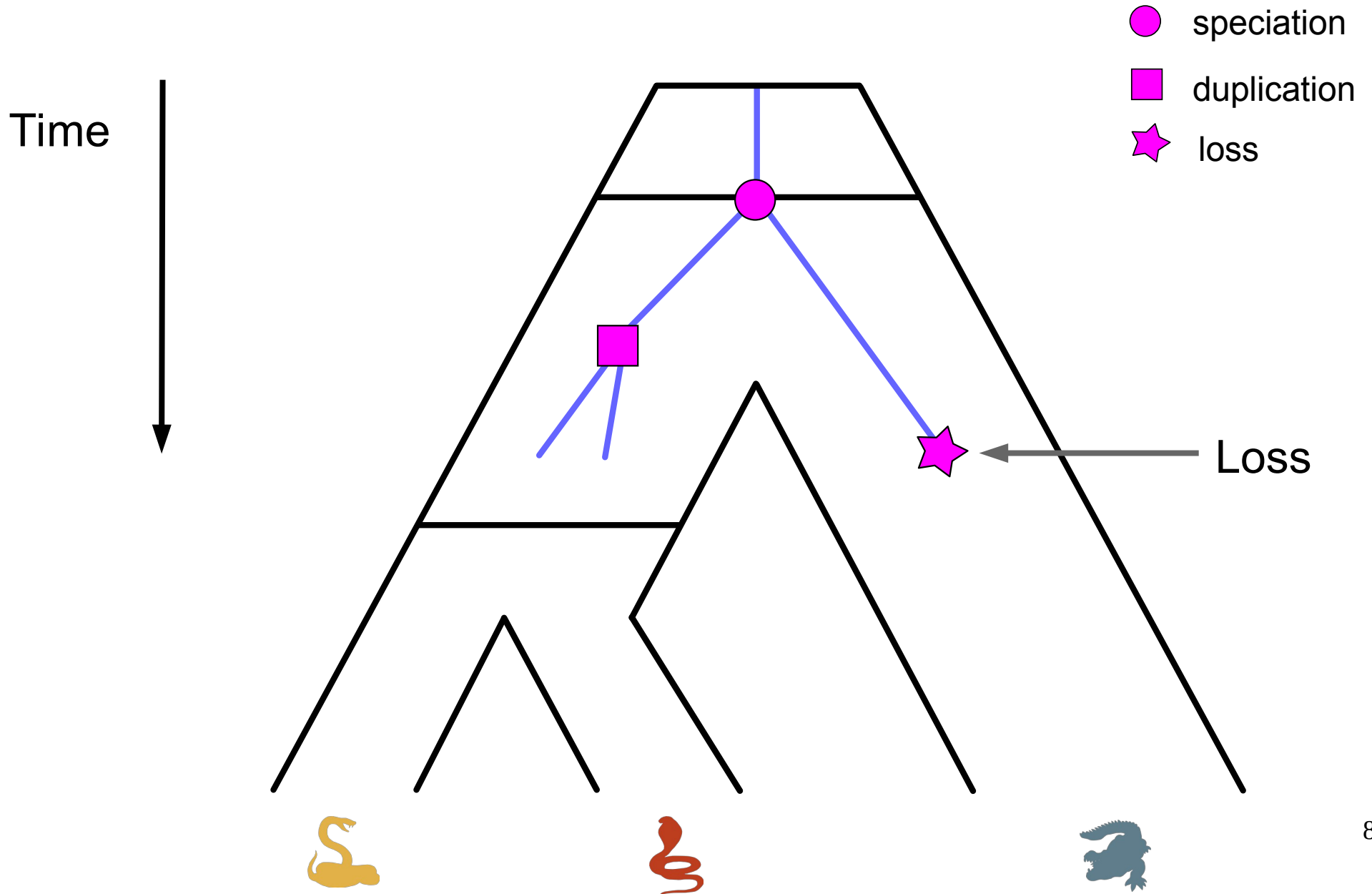
Gene evolution



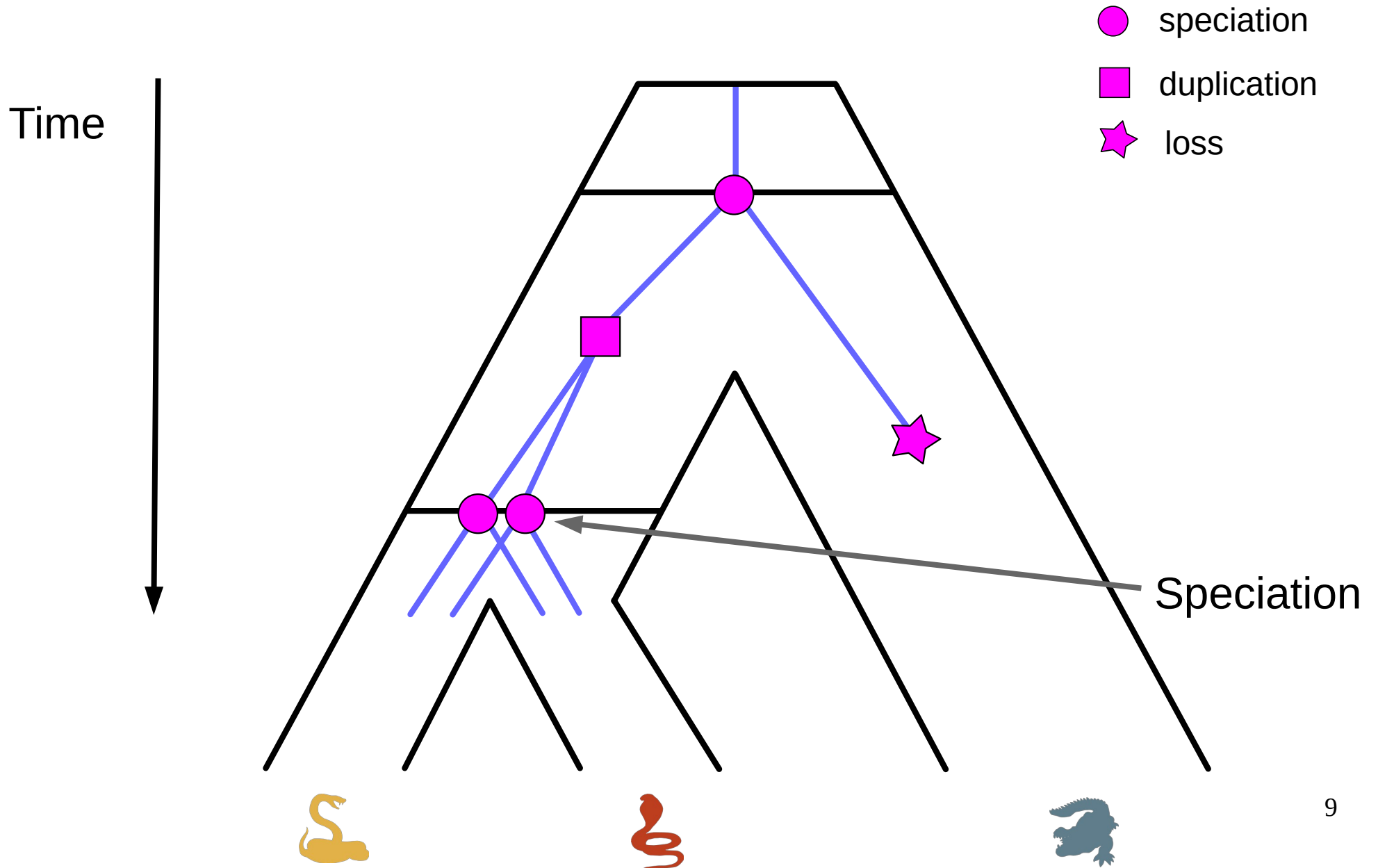
Gene evolution



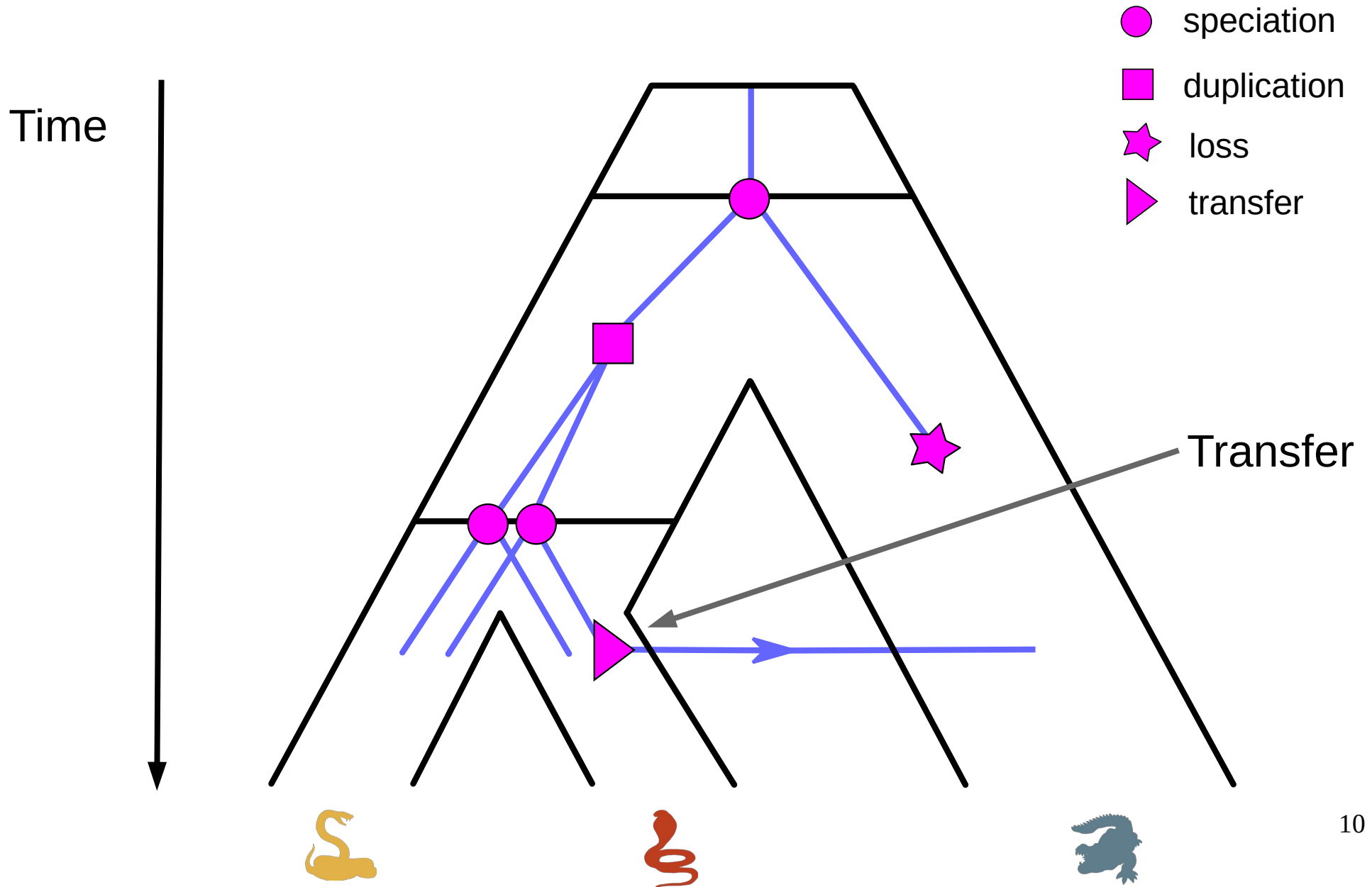
Gene evolution



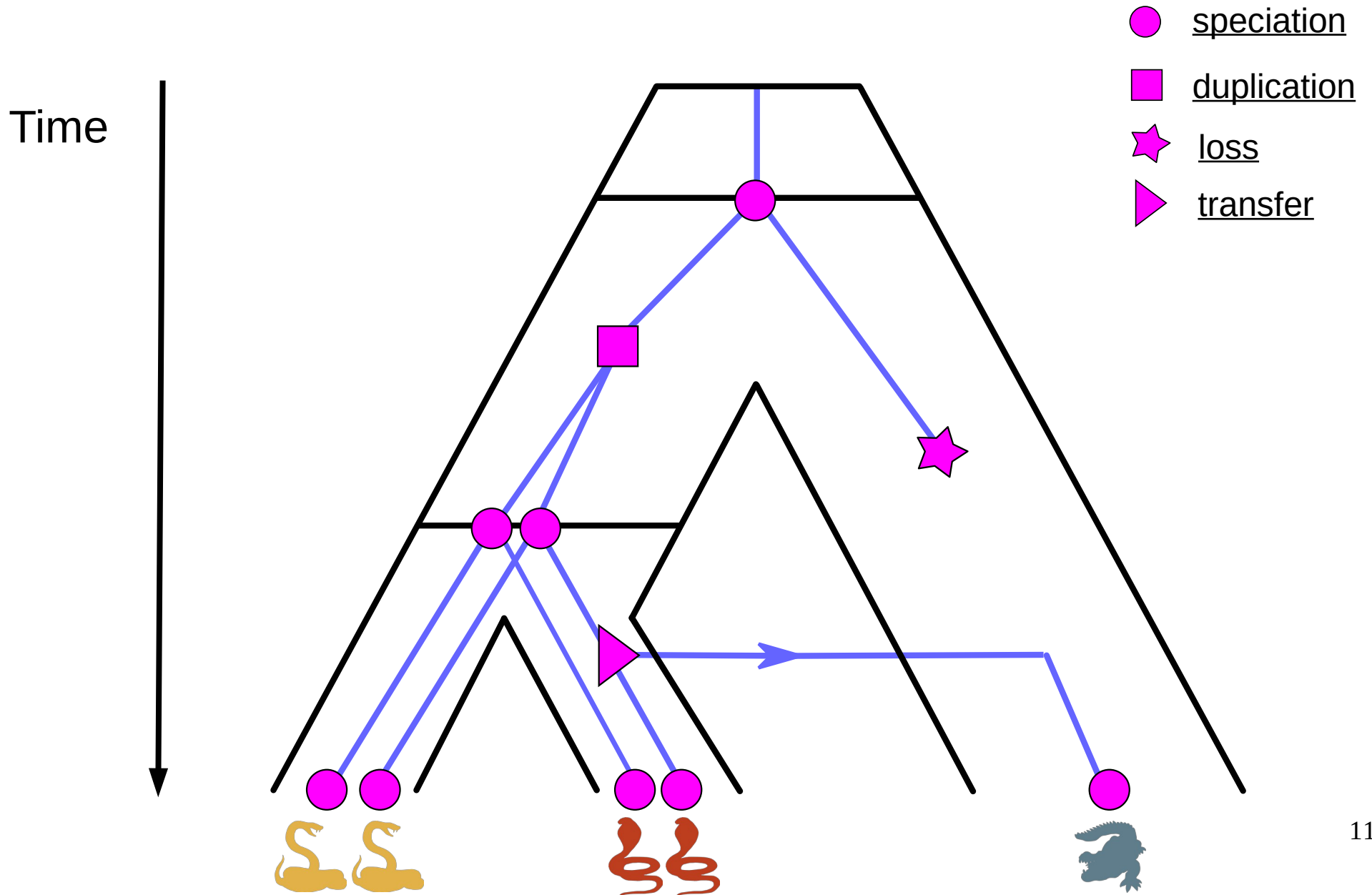
Gene evolution



Gene evolution

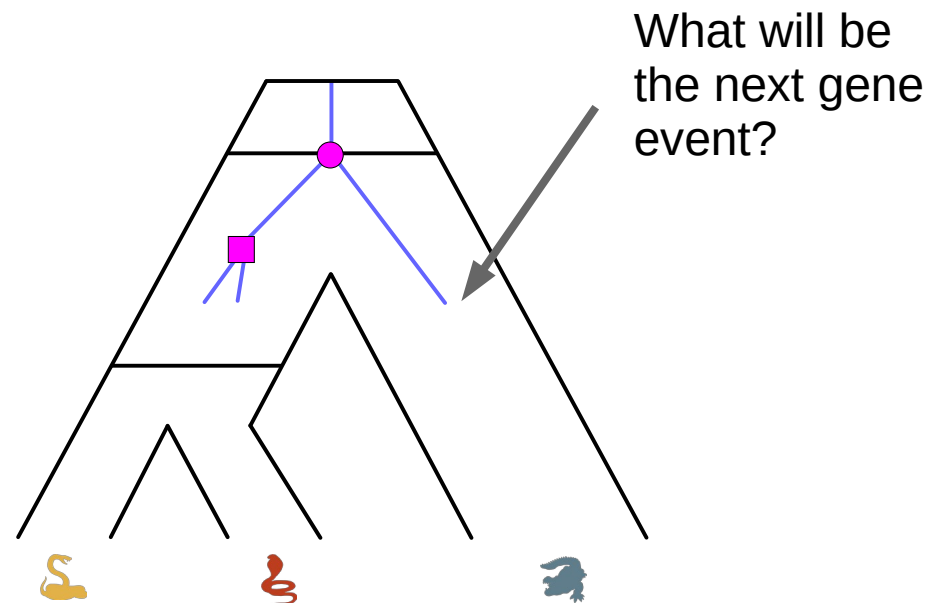


Gene evolution



The undatedDTL model

For a given gene copy, the next event is either:

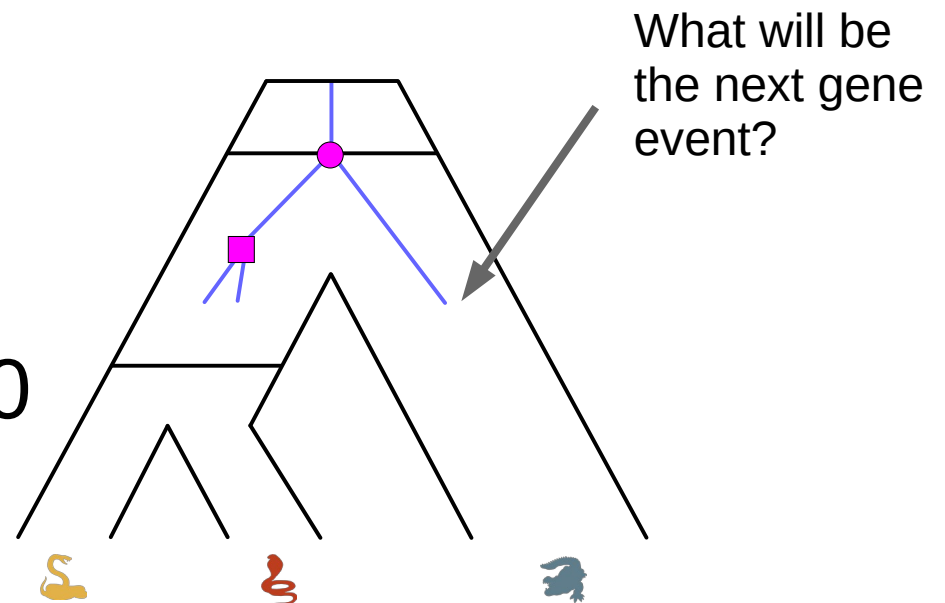


The undatedDTL model

For a given gene copy, the next event is either:

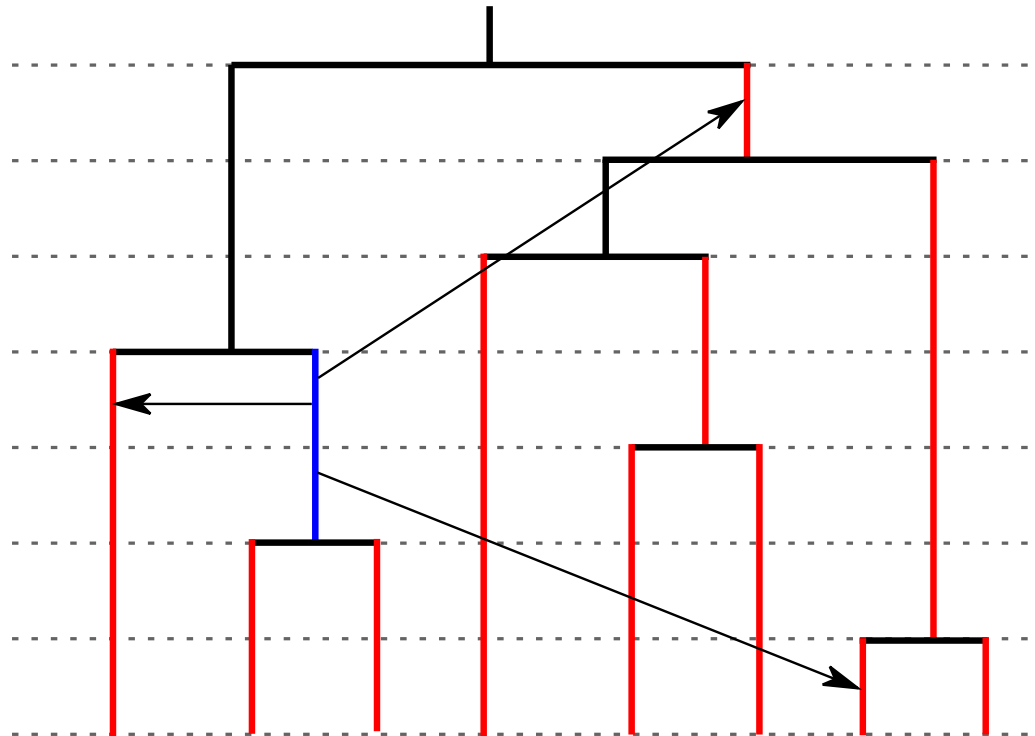
- a speciation (pS)
- a duplication (pD)
- a transfer (pT)
- a loss (pL)

$$pS + pD + pL + pT = 1.0$$



Gene transfer

The receiving species branch is chosen uniformly

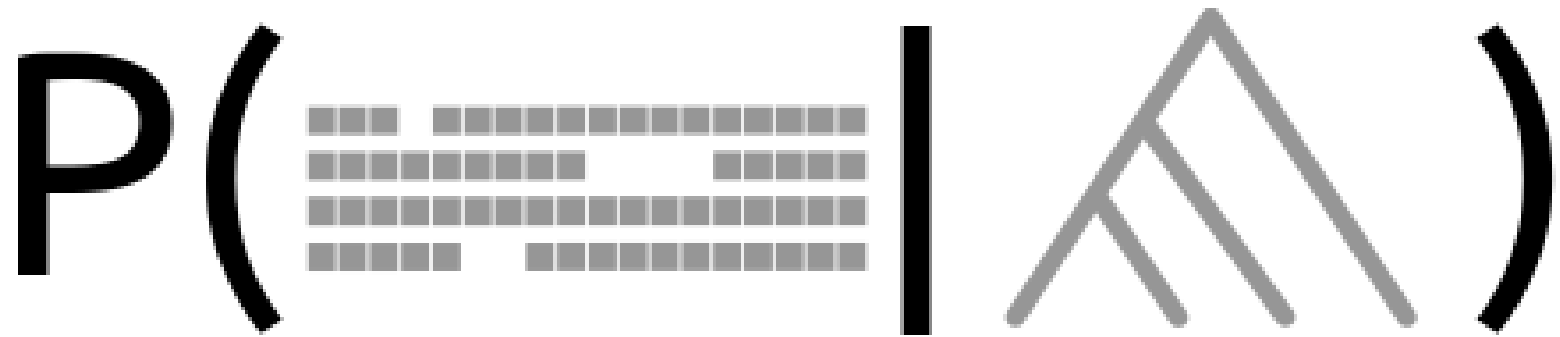


Reconciliation likelihood

$$P(\text{Gene Tree} \mid \text{Species Tree})$$

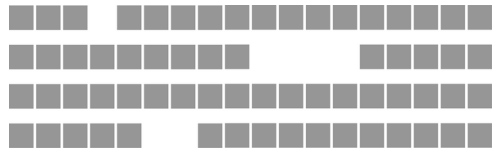
Probability of observing a gene tree given a species tree

Phylogenetic likelihood



Probability of observing the sequences given the gene tree

We only have the sequences



A block of four lines of gray rectangular bars representing DNA sequences. The bars are arranged in a staggered pattern, with some bars starting at different horizontal positions, suggesting a multi-sequence alignment or a set of related sequences.



A block of four lines of gray rectangular bars representing DNA sequences. The bars are arranged in a staggered pattern, with some bars starting at different horizontal positions, suggesting a multi-sequence alignment or a set of related sequences.



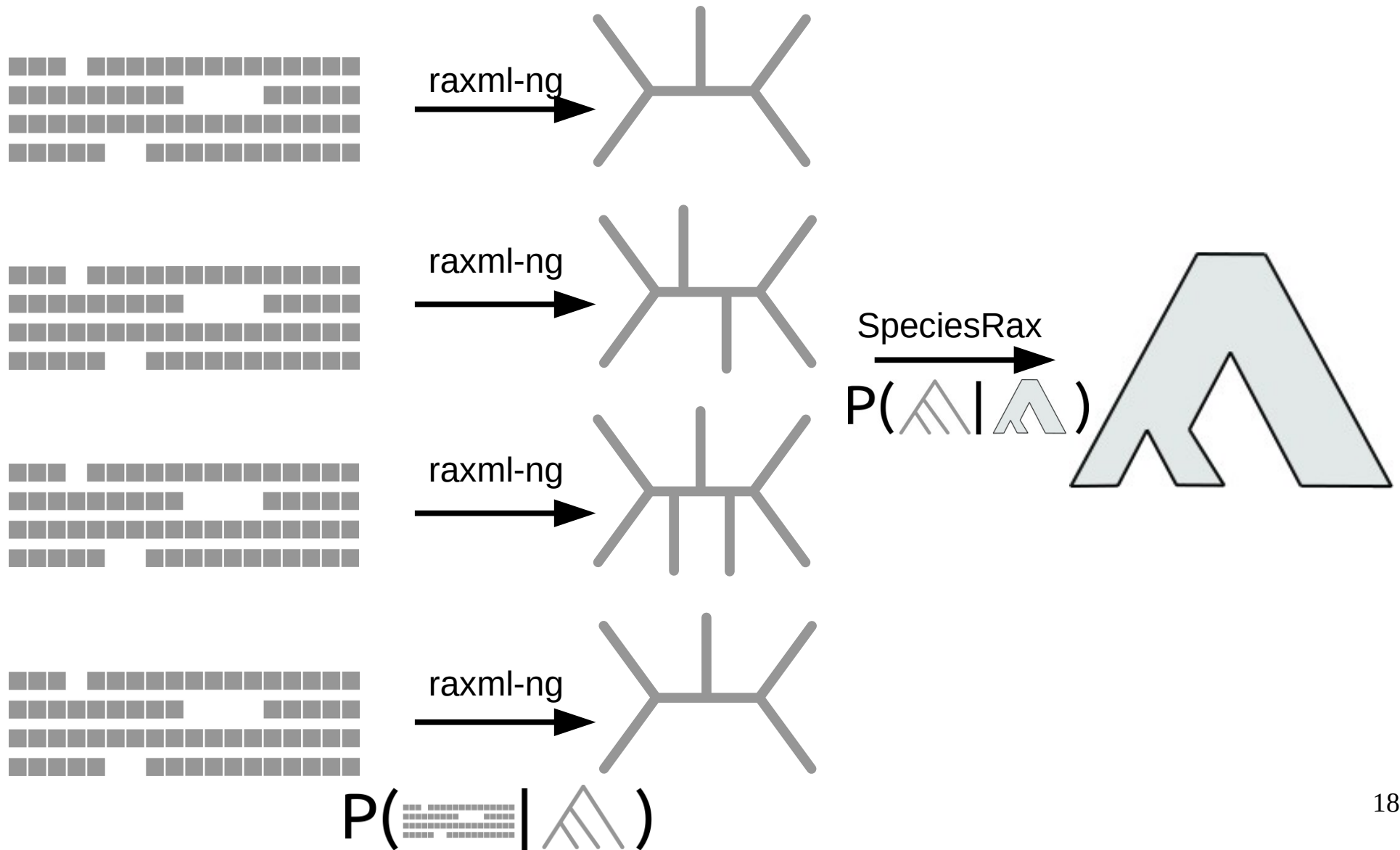
A block of four lines of gray rectangular bars representing DNA sequences. The bars are arranged in a staggered pattern, with some bars starting at different horizontal positions, suggesting a multi-sequence alignment or a set of related sequences.



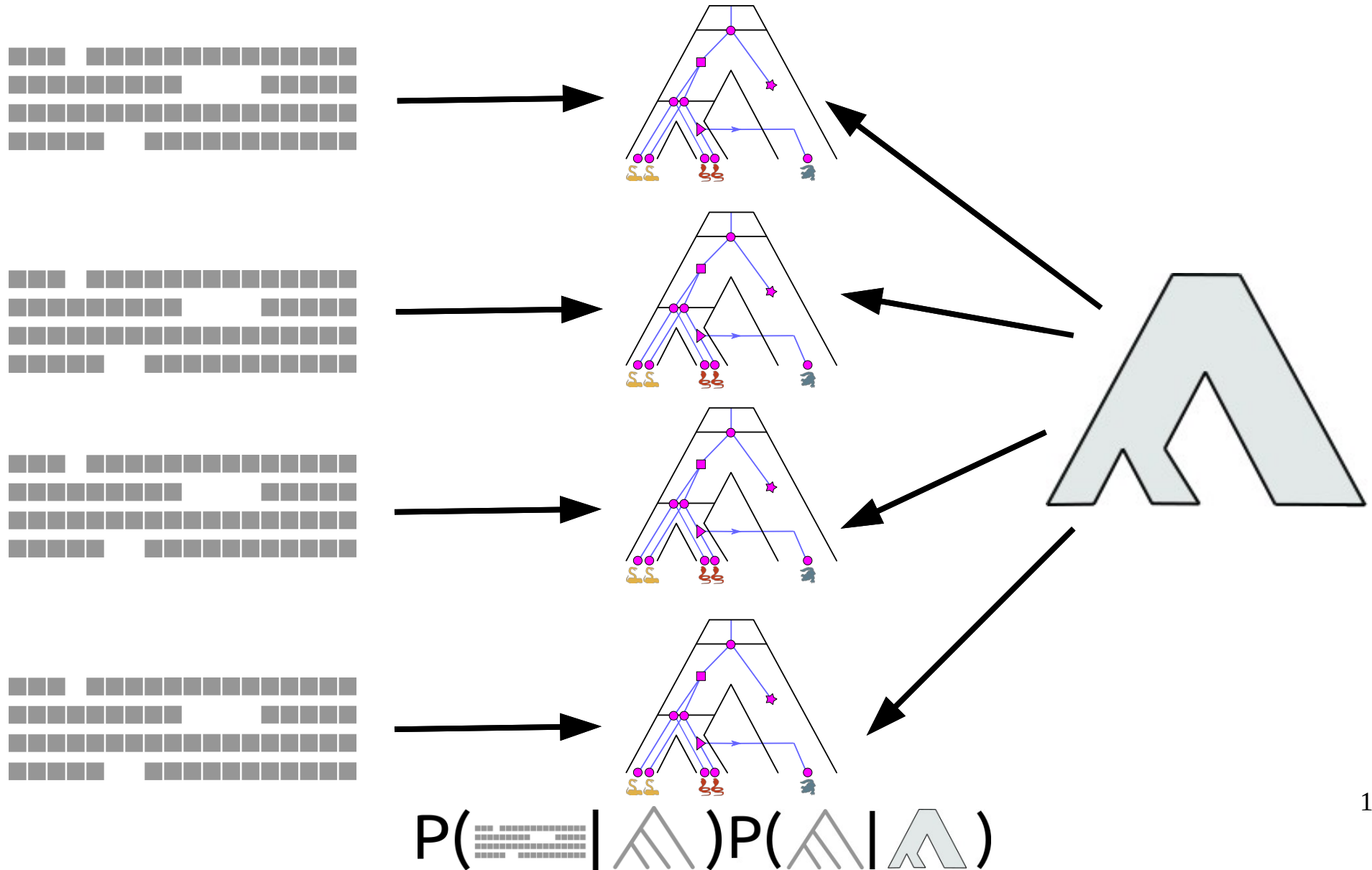
A block of four lines of gray rectangular bars representing DNA sequences. The bars are arranged in a staggered pattern, with some bars starting at different horizontal positions, suggesting a multi-sequence alignment or a set of related sequences.

How do we get from the
sequences to the reconciled
gene trees?

SpeciesRax: species tree inference



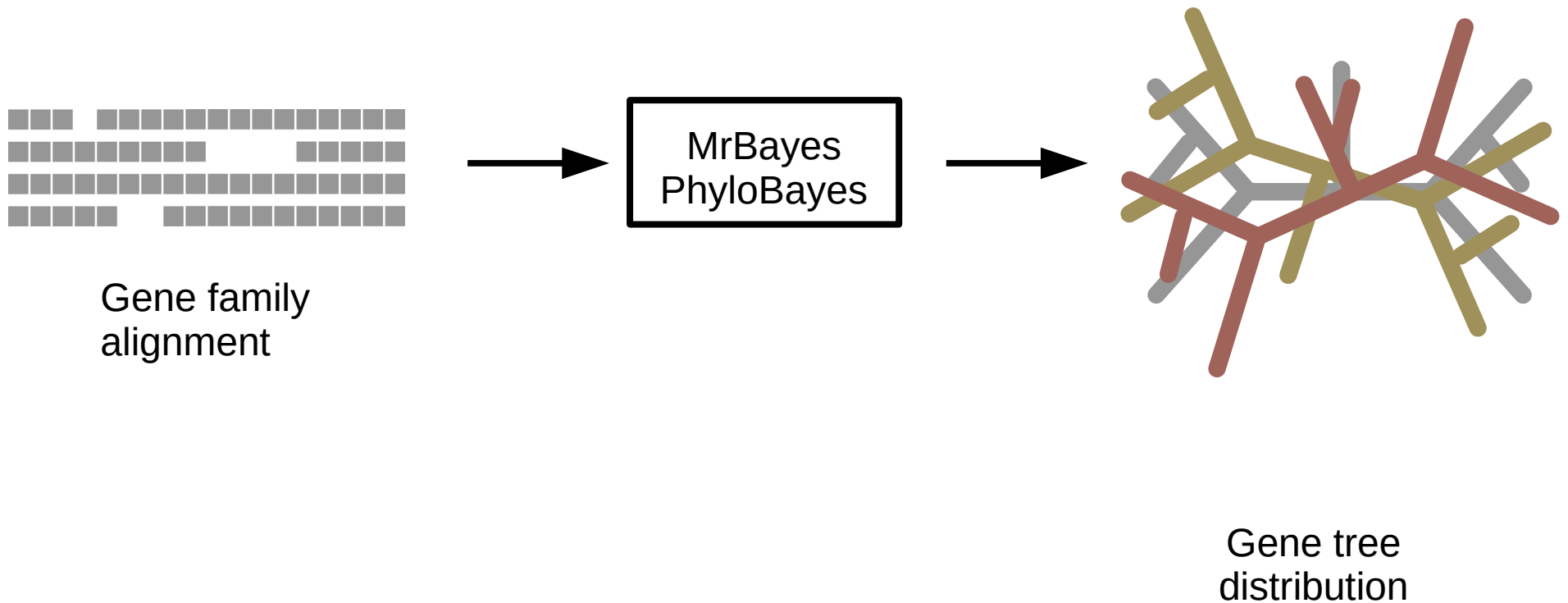
GeneRax: correction and reconciliation



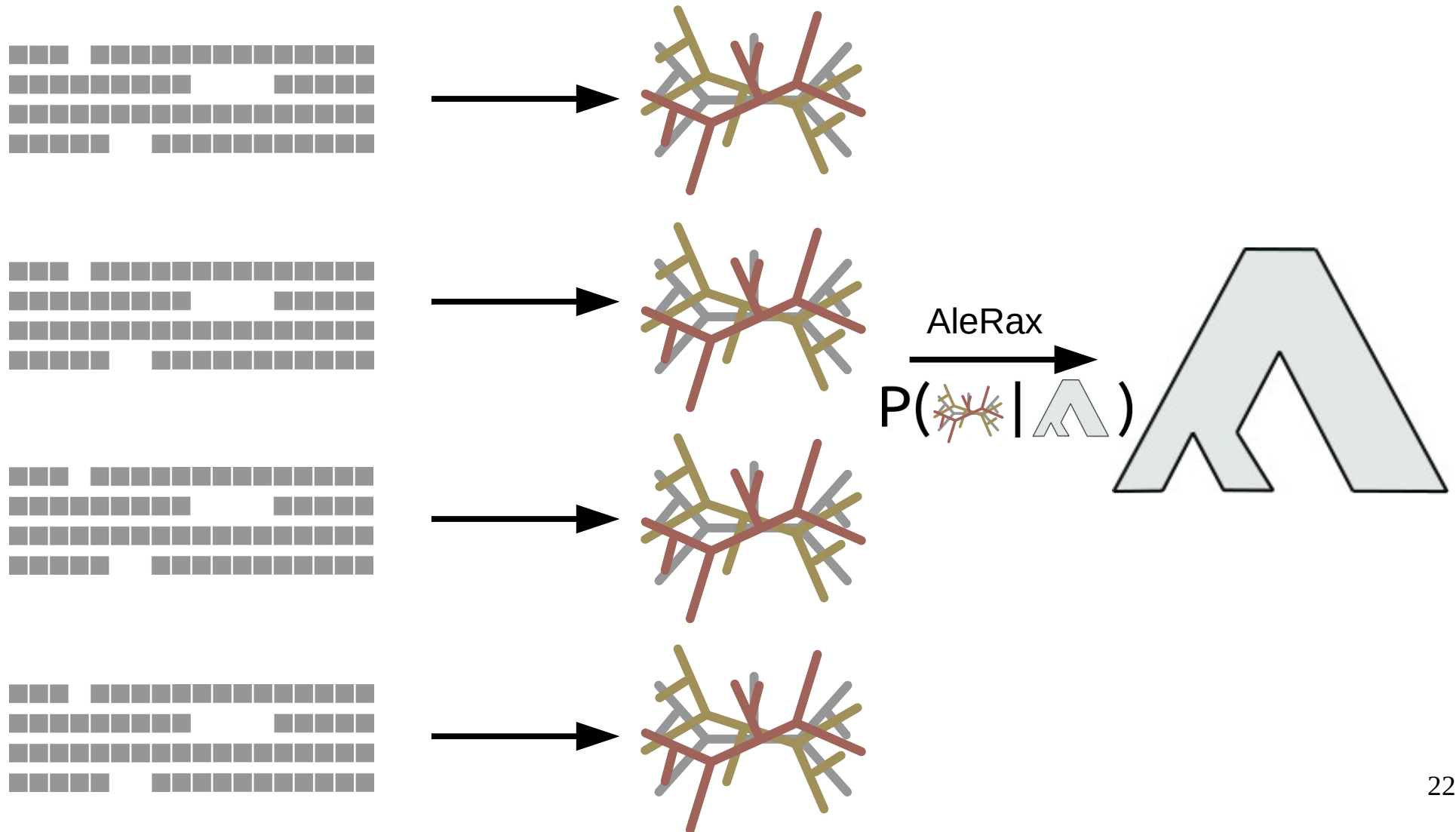
Limitations

- GeneRax: no support values
- SpeciesRax infers the species tree from **estimated** gene trees.

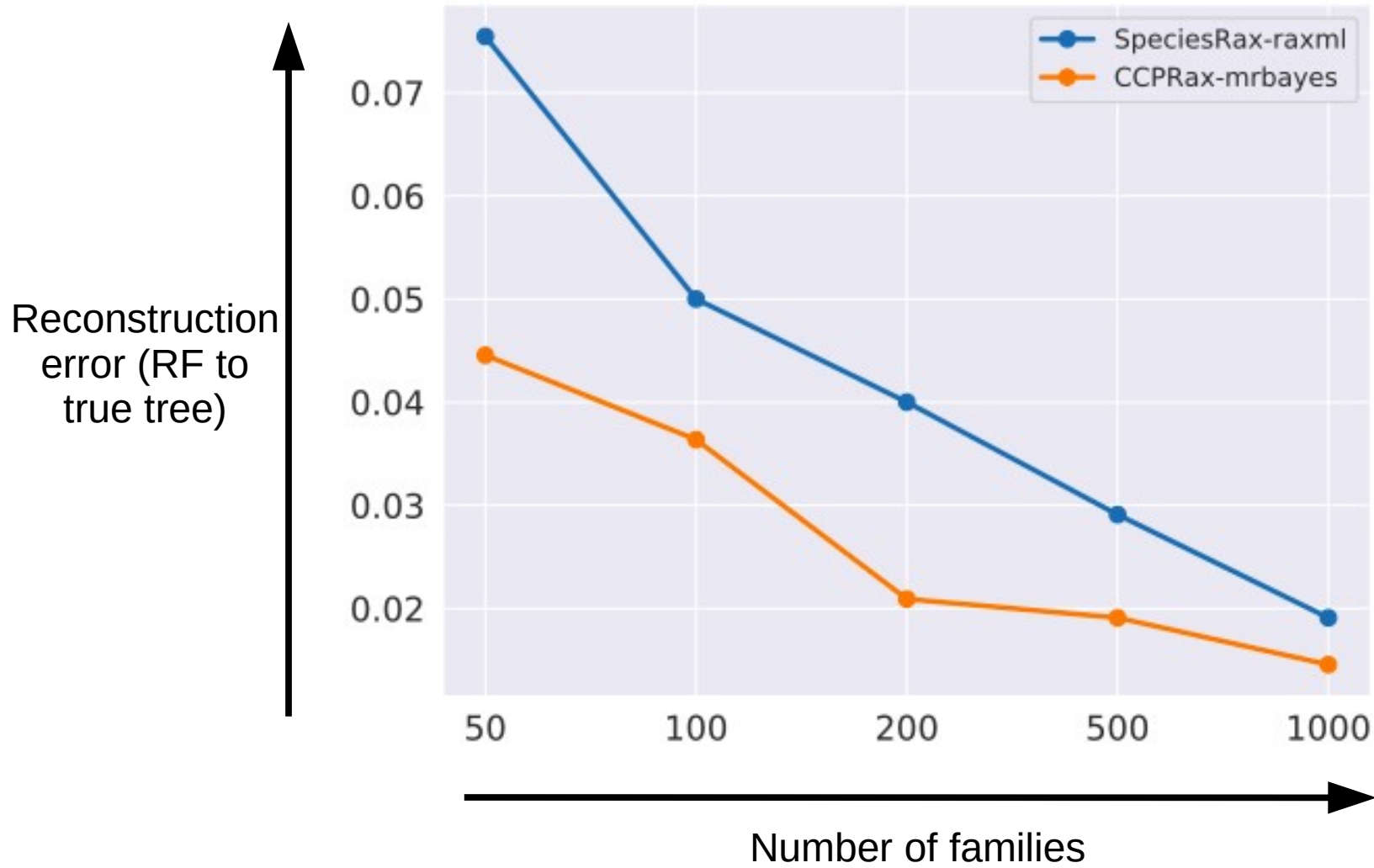
AleRax: integrate over the posterior distributions of gene trees



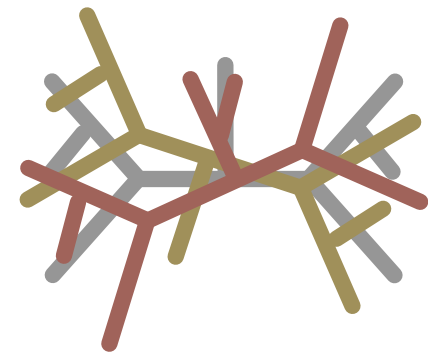
AleRax: integrate over the posterior distributions of gene trees



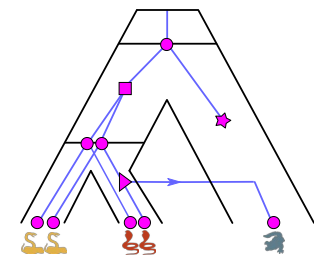
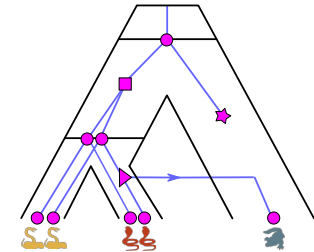
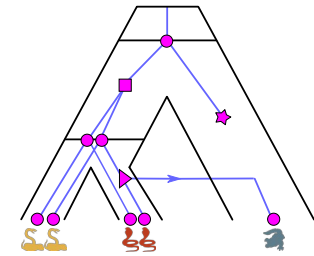
Benchmark



AleRax: sample gene trees (reimplementation of ALE)



Sample under
the UndatedDTL



Beyond tree inference

- Test whole-genome duplication events
- Identify “highways of transfers”
- Extend to species network
- ... and other exciting projects!

Questions?

