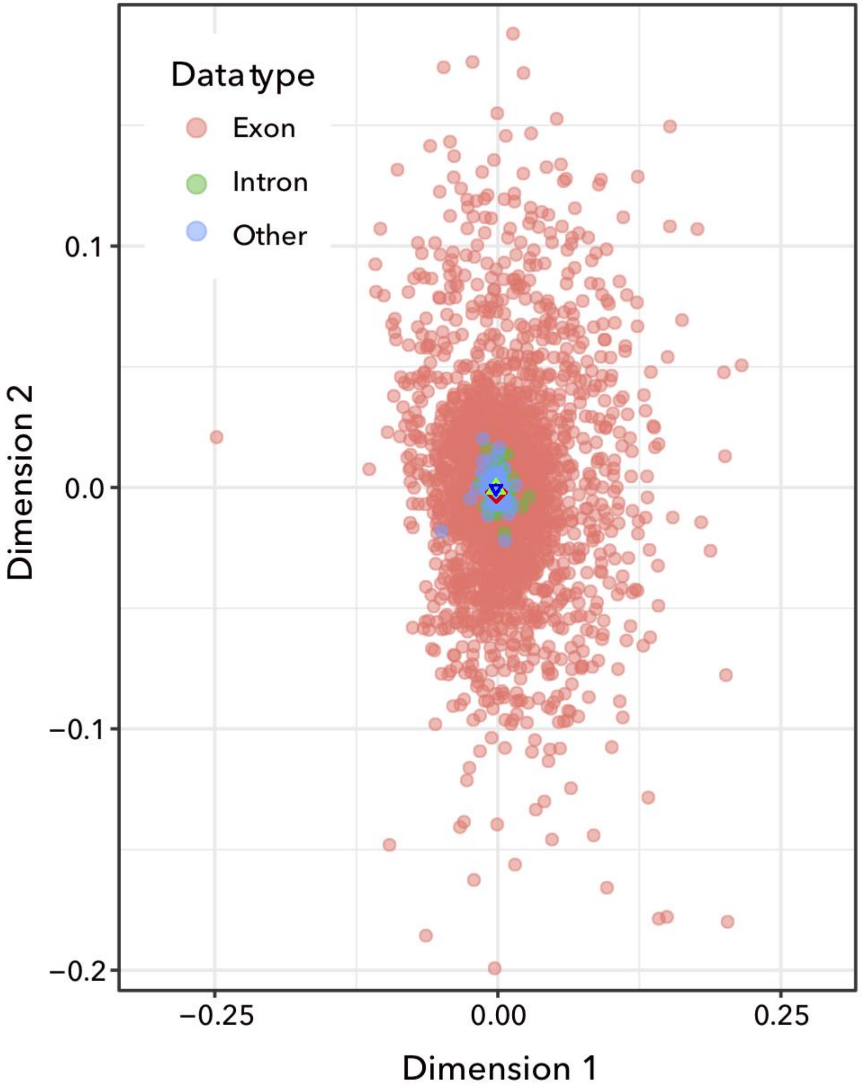
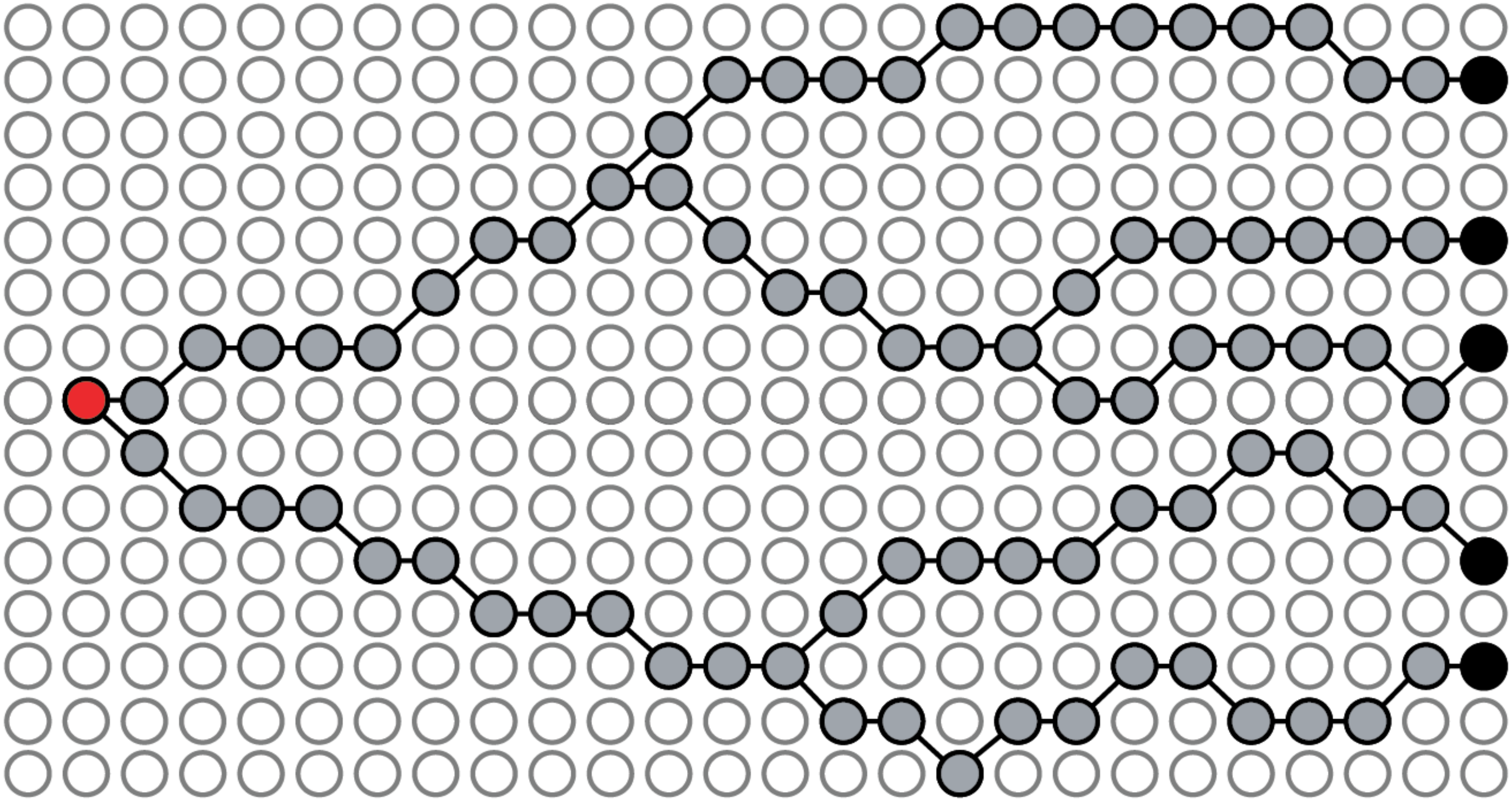
Gene trees

Incongruence among gene trees

- Unlinked loci can have different gene trees
- Explore phylogenetic congruence across markers
- Topological distance metrics

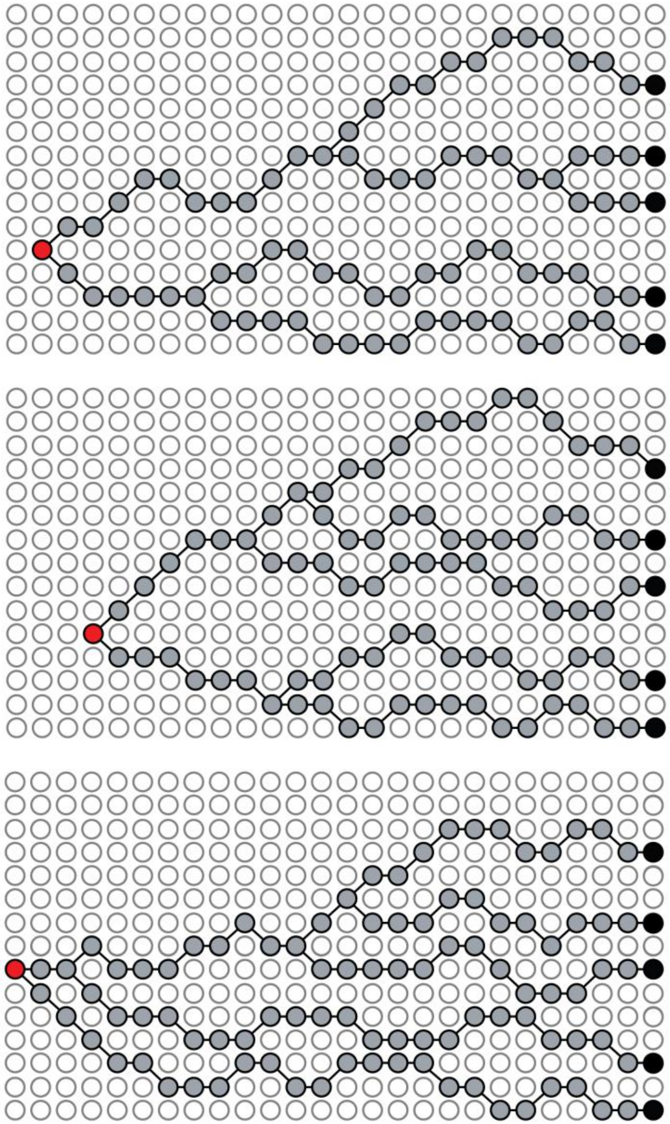


Coalescence theory



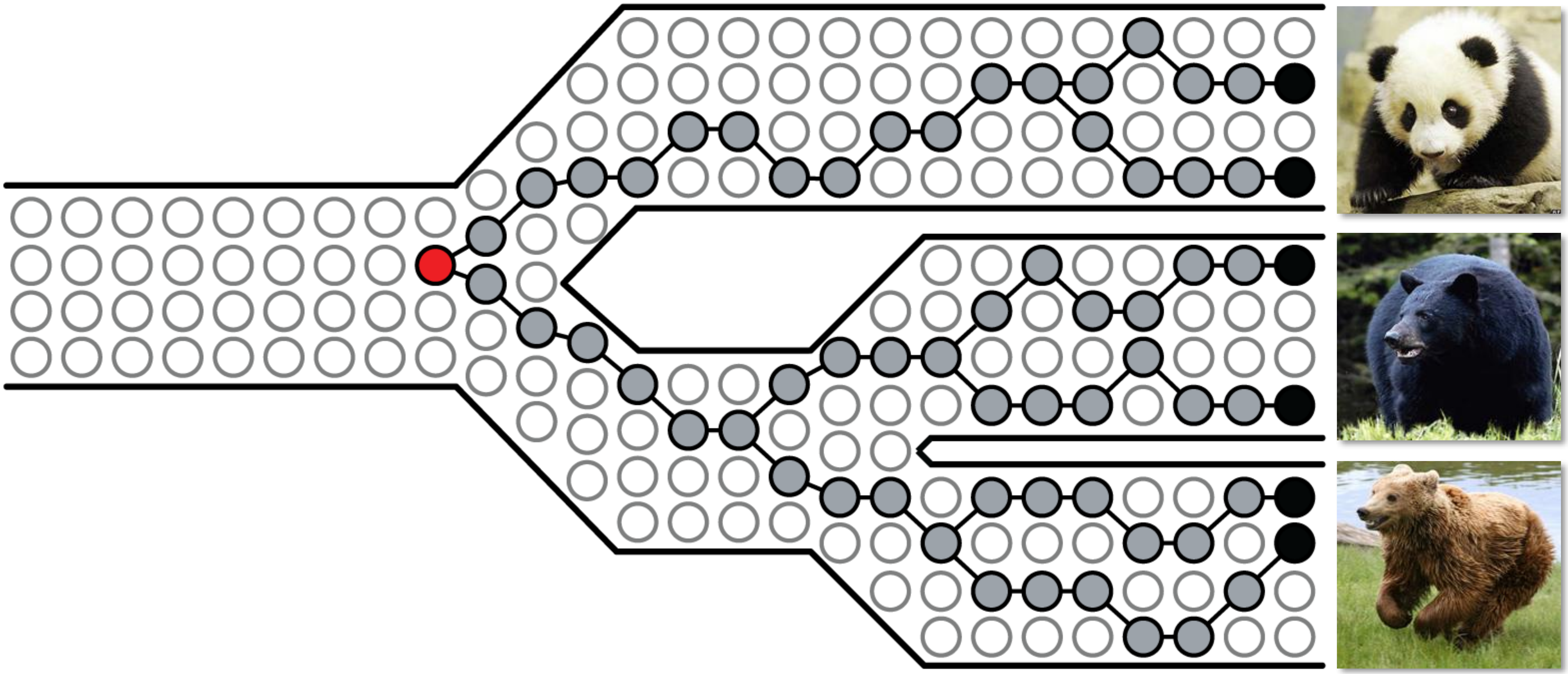
Gene trees within a species

- Gene genealogies vary stochastically between unlinked genes
- Independent genomic regions should not be grouped in analyses of a single species
    - They have different tree signals
    - They have different coalescent times



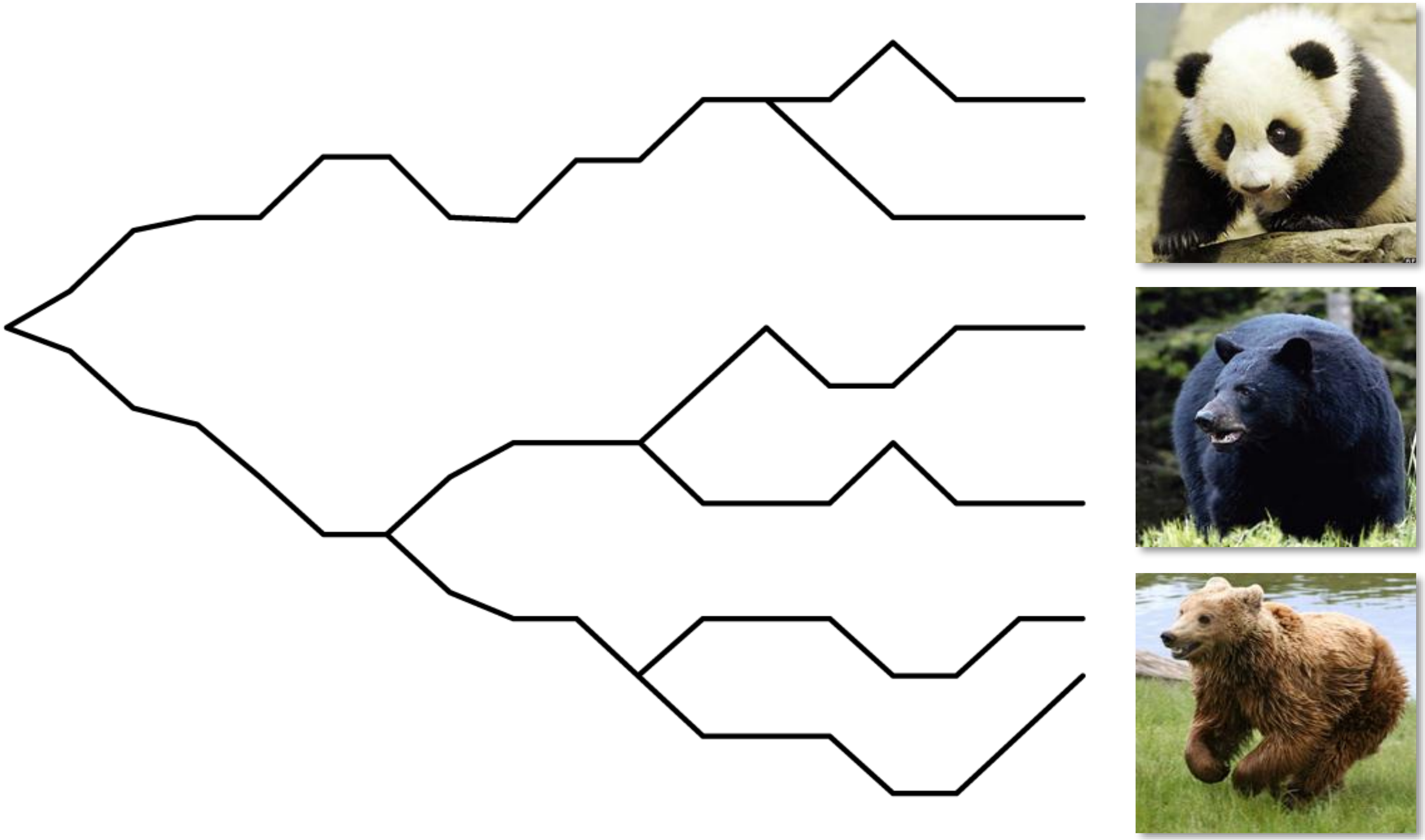
Gene trees across multiple species

- Gene trees within a tree of multiple species



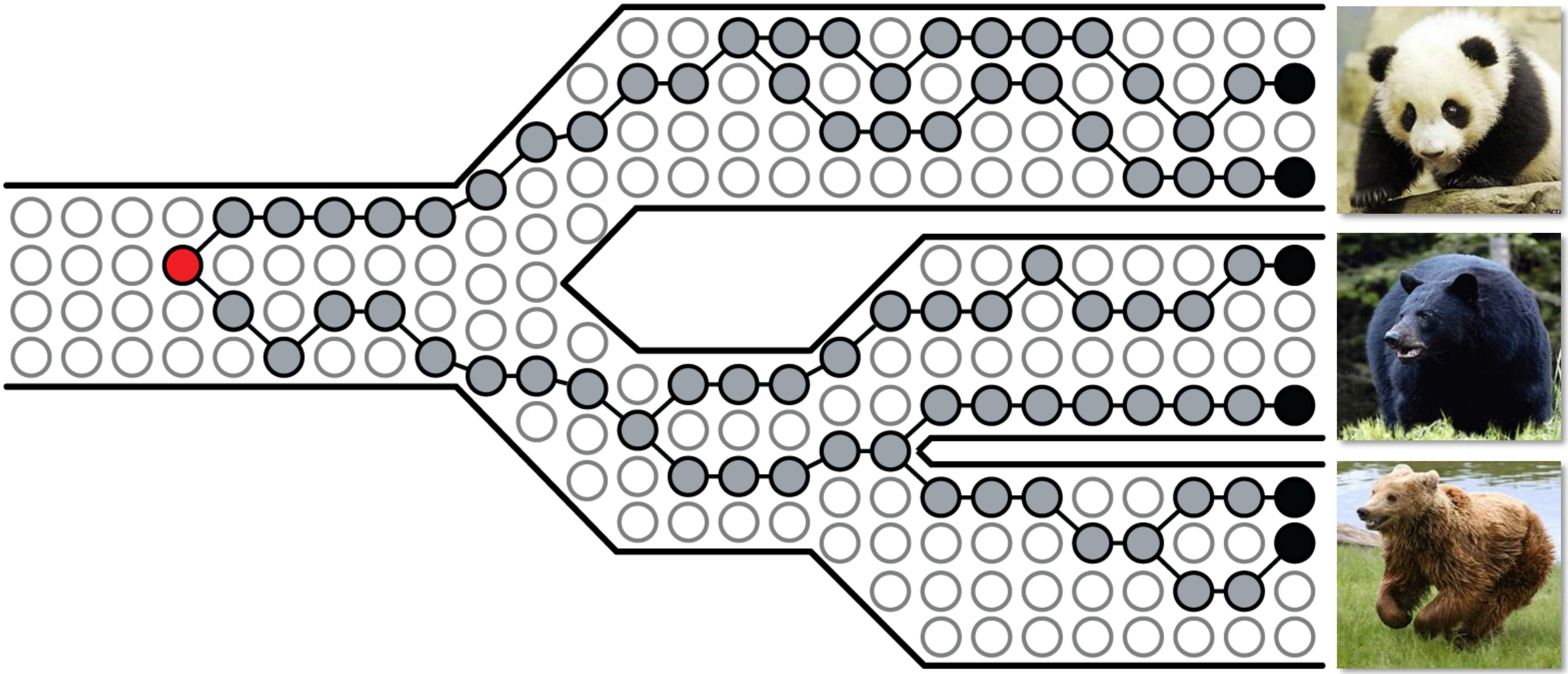
Gene trees across multiple species

- Gene trees within a tree of multiple species



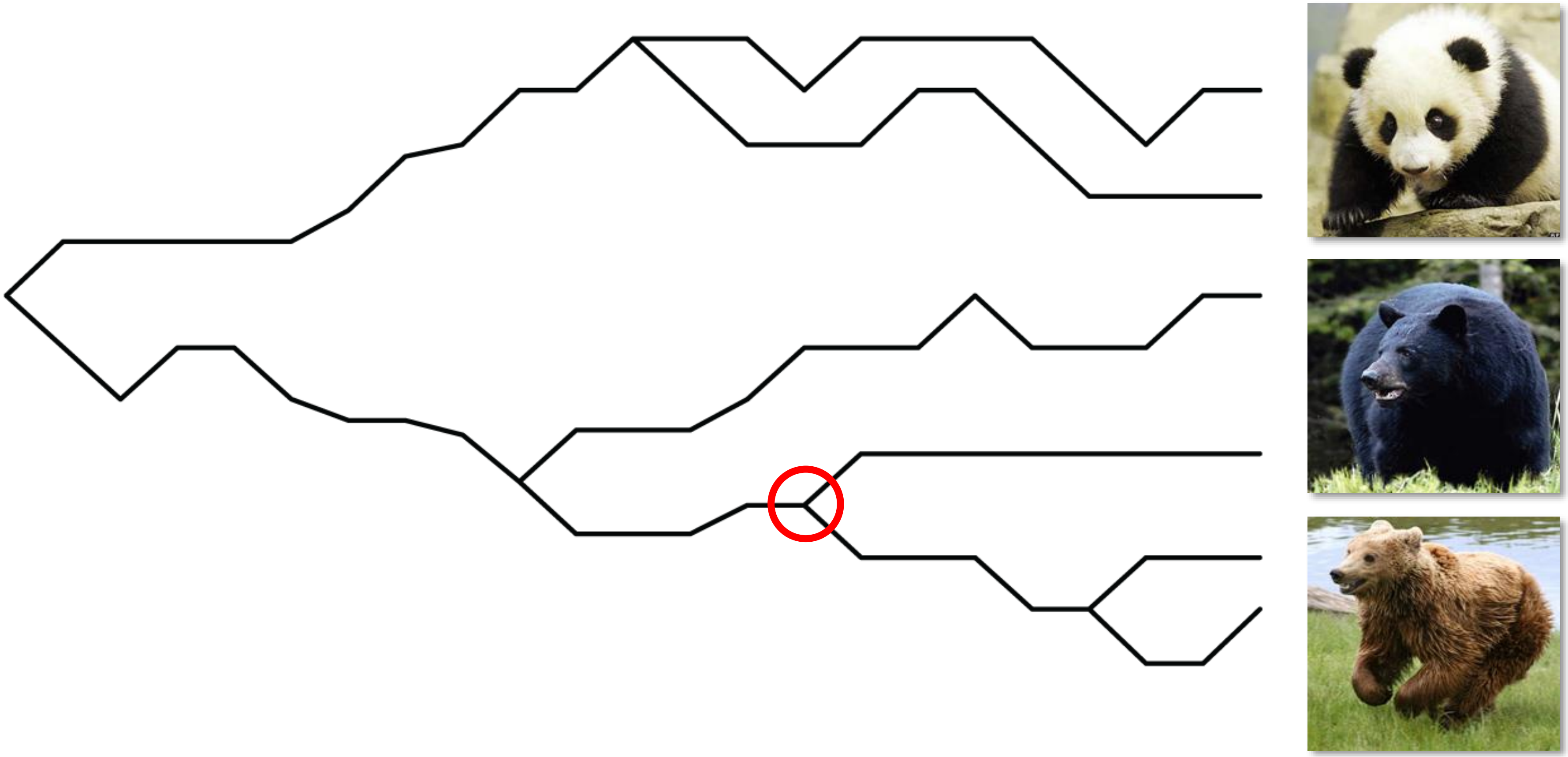
Gene trees across multiple species

- Incomplete lineage sorting



Gene trees across multiple species

- Incomplete lineage sorting

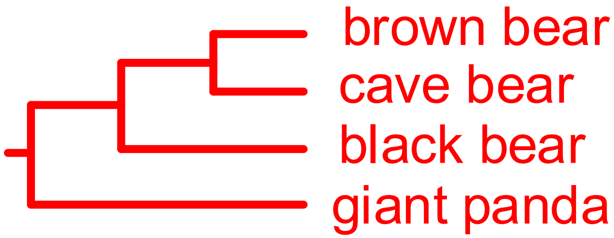


The species tree

- **Incomplete lineage sorting** can lead to gene trees that are different from the species tree
- We can still infer the species tree using multiple genes
- Three approaches
  1. Concatenation
  2. Consensus
  3. Coalescent methods

Analysis of multiple genes

1. **Concatenation.** Assume that all genes share the same evolutionary history

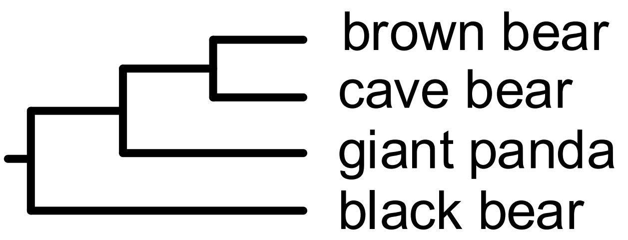


All genes

Ignores differences among gene trees

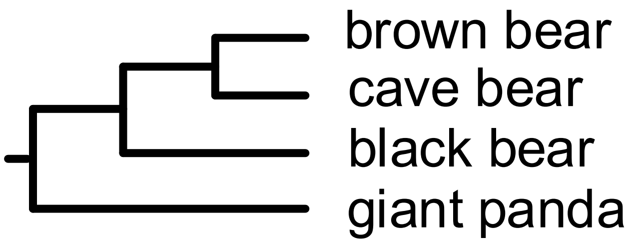
The species tree

2. **Consensus.** Estimate each gene genealogy and generate a consensus



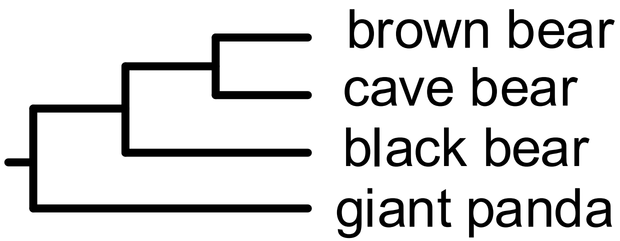
brown bear
cave bear
giant panda
black bear

Gene 1



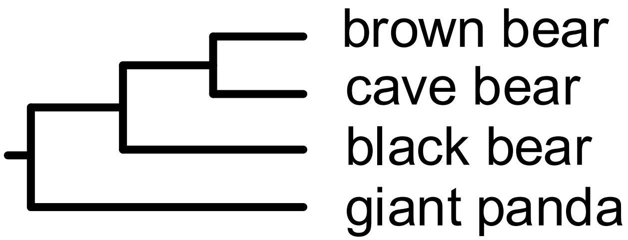
brown bear
cave bear
black bear
giant panda

Gene 2



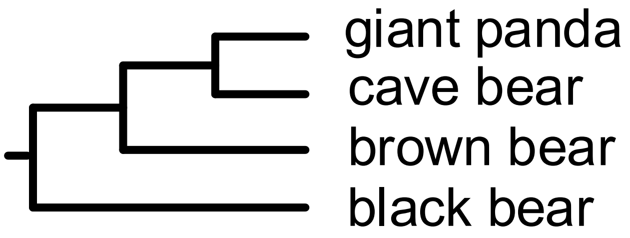
brown bear
cave bear
black bear
giant panda

Gene 3



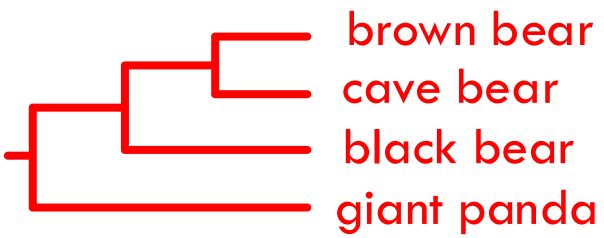
brown bear
cave bear
black bear
giant panda

Gene 4



giant panda
cave bear
brown bear
black bear

Gene 5



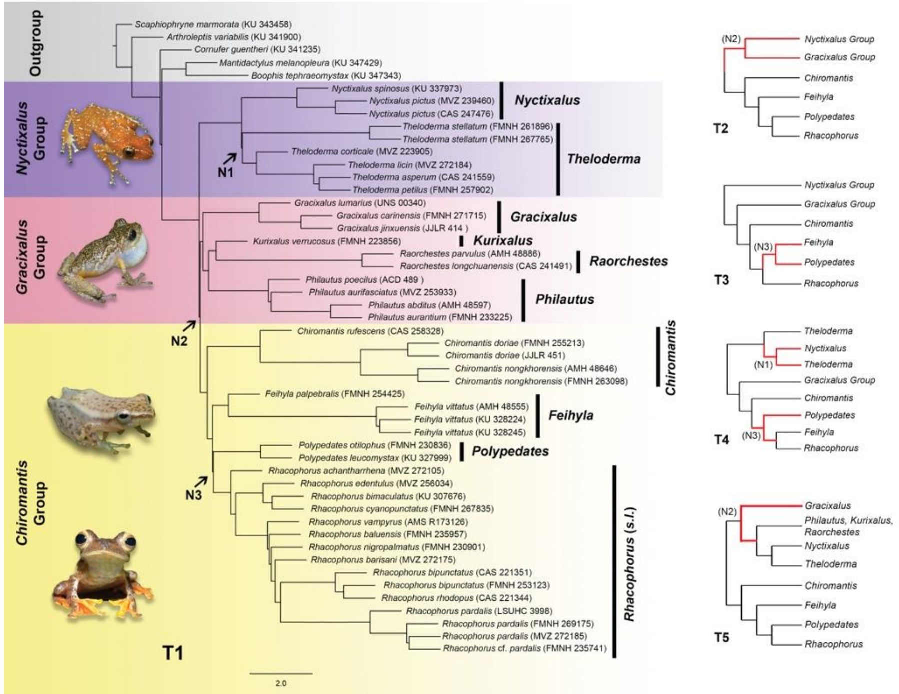
brown bear
cave bear
black bear
giant panda

Consensus

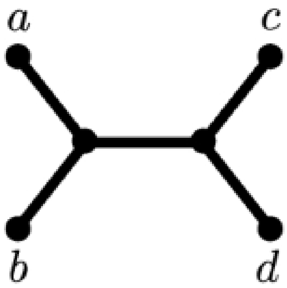
But the most common gene tree is not always similar to the species tree ("anomaly zone")

The anomaly zone

When it is more probable that gene trees are <u>not</u> consistent with the species tree



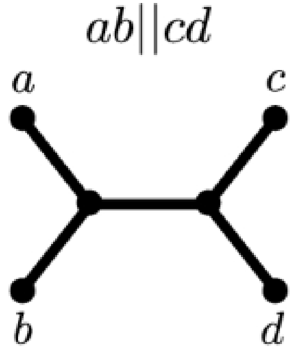
Quartets



$ab||cd$

Quartets

Quartets do not suffer from the anomaly zone



$$\sum_{q \in Q(T)} w(q, \mathcal{T})$$

BIOINFORMATICS

Vol. 30 ECCB 2014, pages i541–i548
doi:10.1093/bioinformatics/btu462

ASTRAL: genome-scale coalescent-based species tree estimation

S. Mirarab[1], R. Reaz[1], Md. S. Bayzid[1], T. Zimmermann[1,2], M. S. Swenson[3] and T. Warnow[1,*]

The species tree

3. **Other coalescent methods**
   Estimate the species tree from gene trees

- Assume that gene trees are independent samples from a stochastic process within the species tree (the multi-species coalescent)
- Various methods
  - **BEAST* co-estimates gene- and species-trees

Useful references

- **Gene tree discordance, phylogenetic inference and the multispecies coalescent**
  Degnan & Rosenberg (2004) *Trends Ecol Evol*, 24: 332–340.
- **The Concatenation Question.** David Bryant, Matthew W. Hahn. Phylogenetics in the Genomic Era, pp.3.4:1--3.4:23, 2020. ⟨hal-02535651⟩