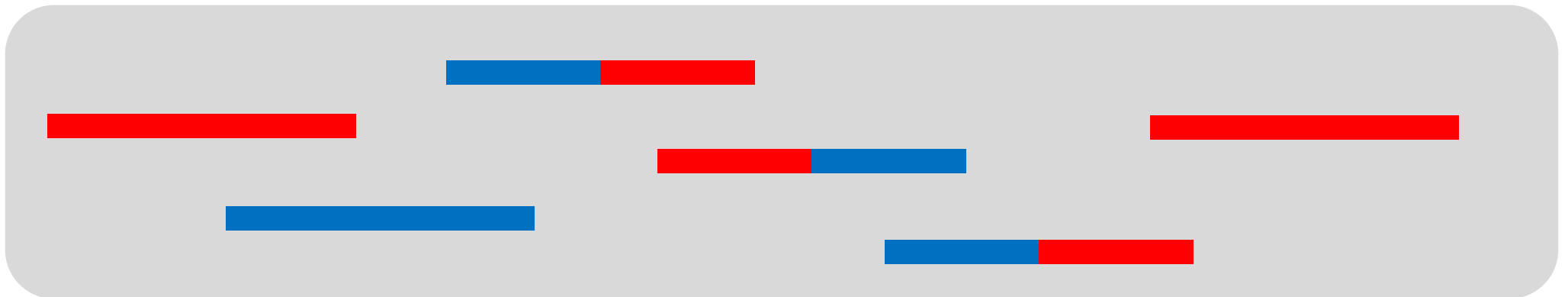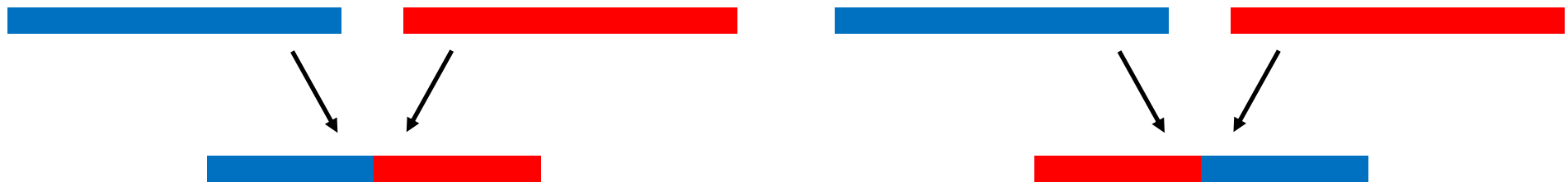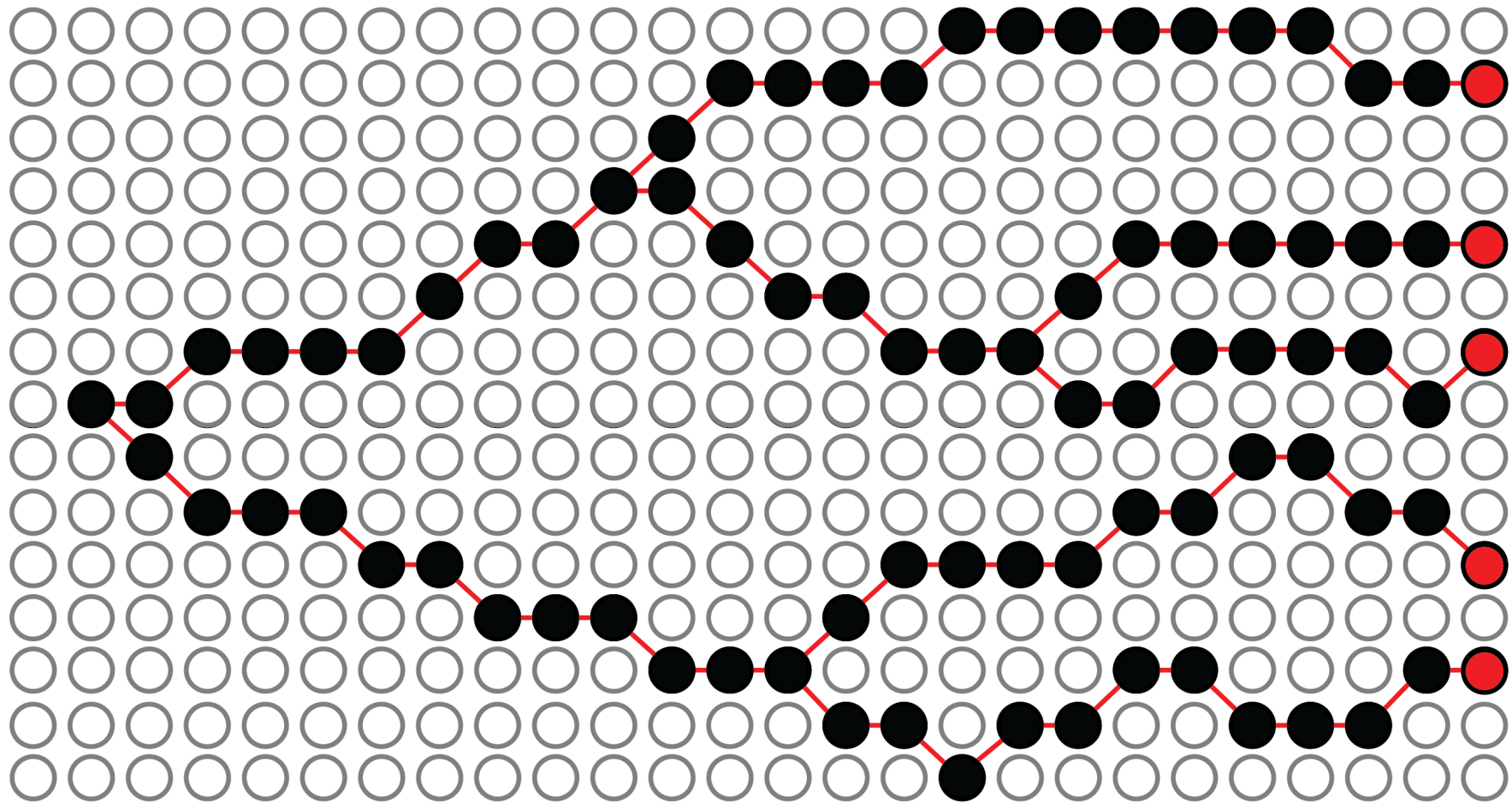# Phylogenomics

Gene trees

- Most phylogenetic methods assume that there is a single tree that describes the evolution of the whole data set

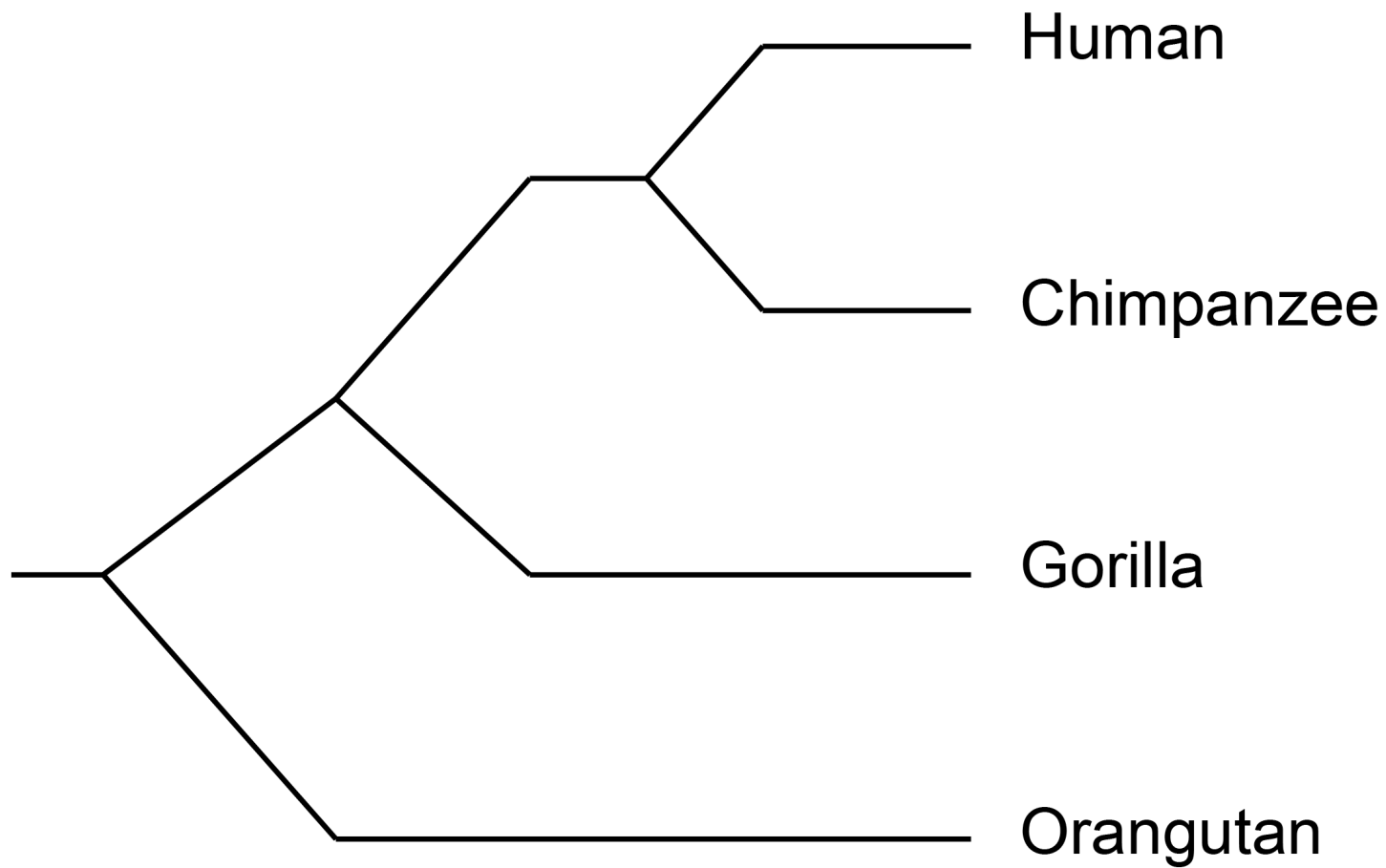- But recombination complicates this
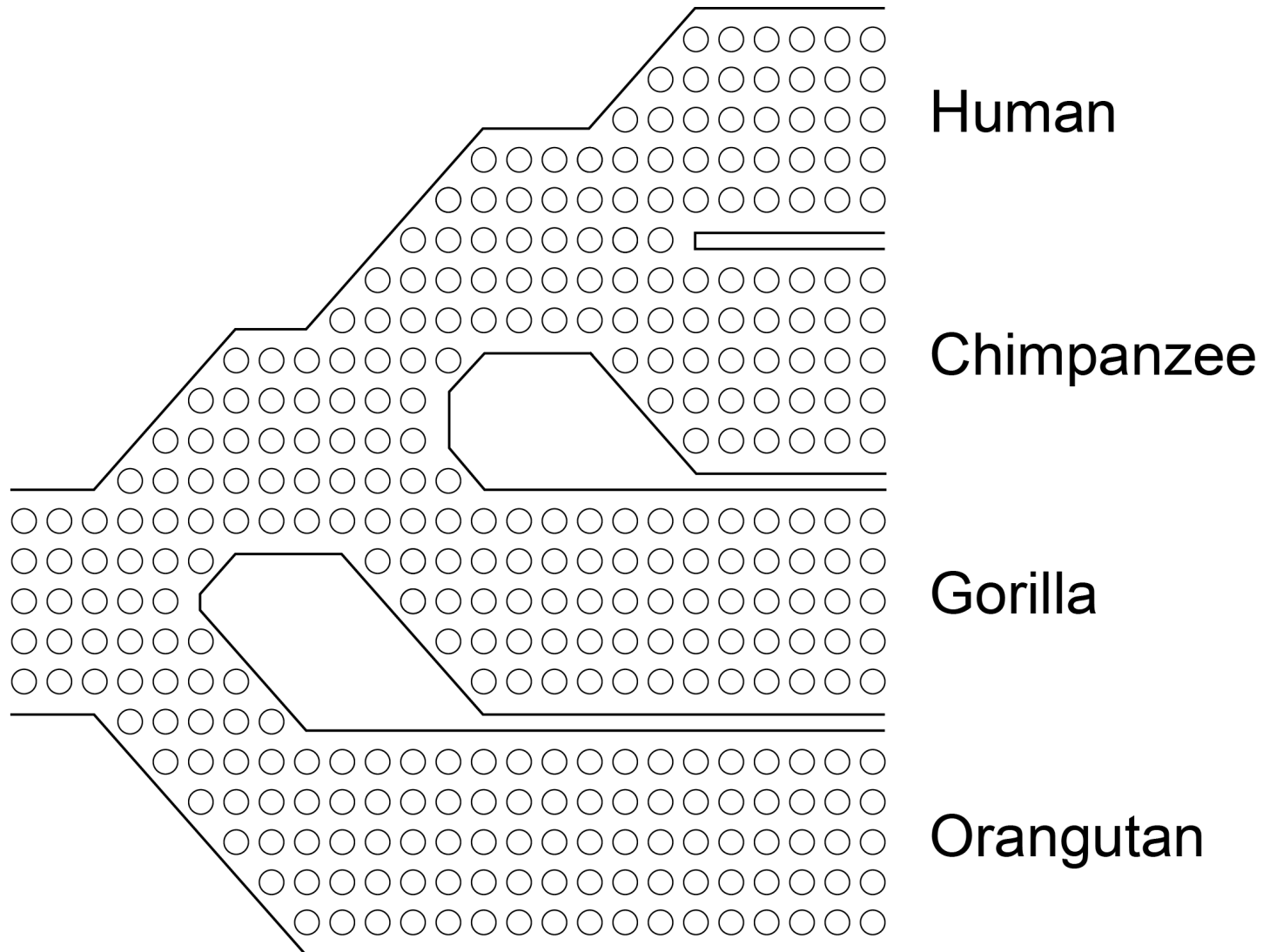
Coalescent theory

3

Species tree

Multispecies coalescent
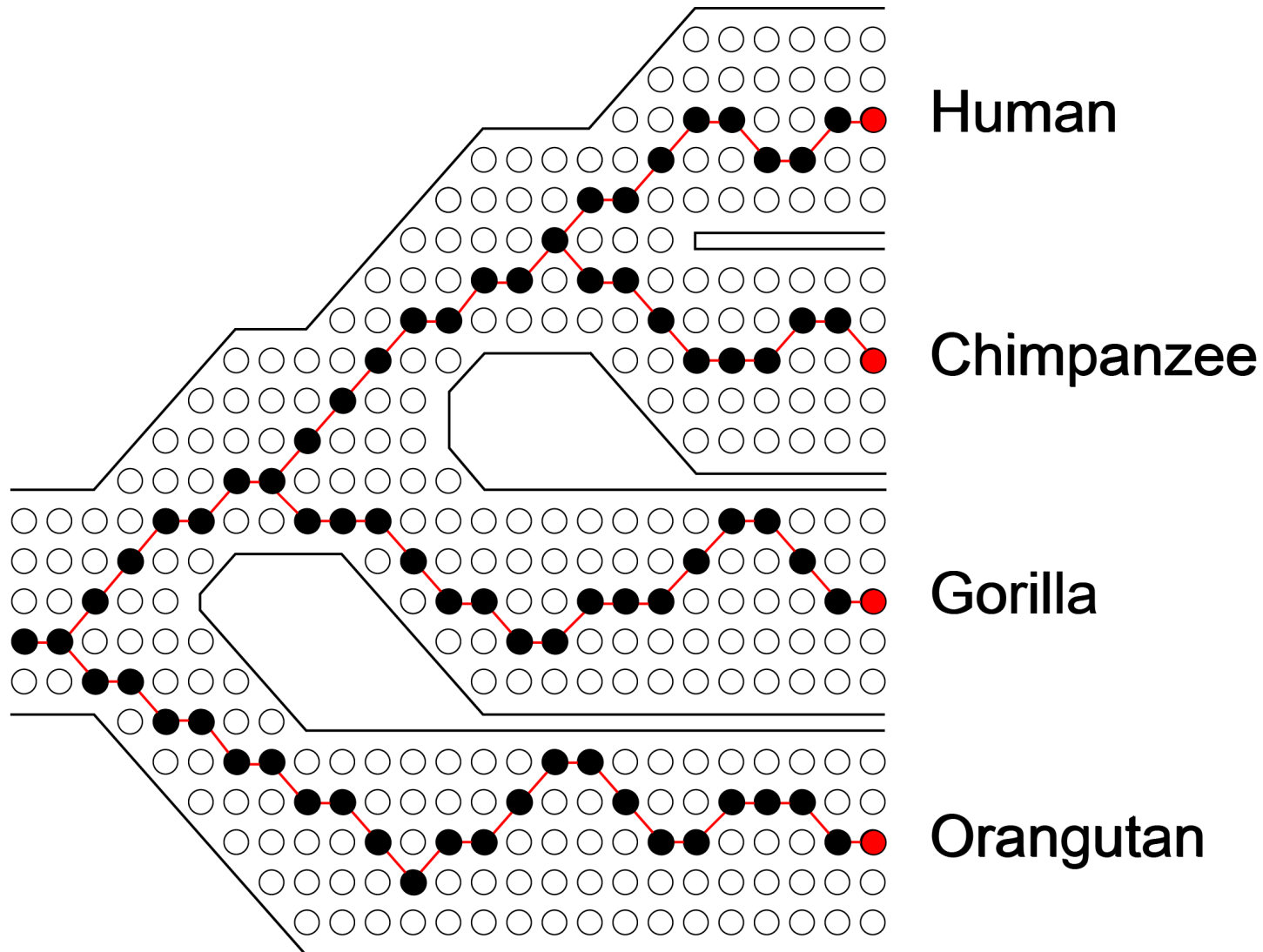

Human

Chimpanzee

Gorilla

Orangutan

Gene tree (concordant)

Human

Chimpanzee

Gorilla

Orangutan

Gene tree (discordant)

Human

Chimpanzee

Gorilla

Orangutan
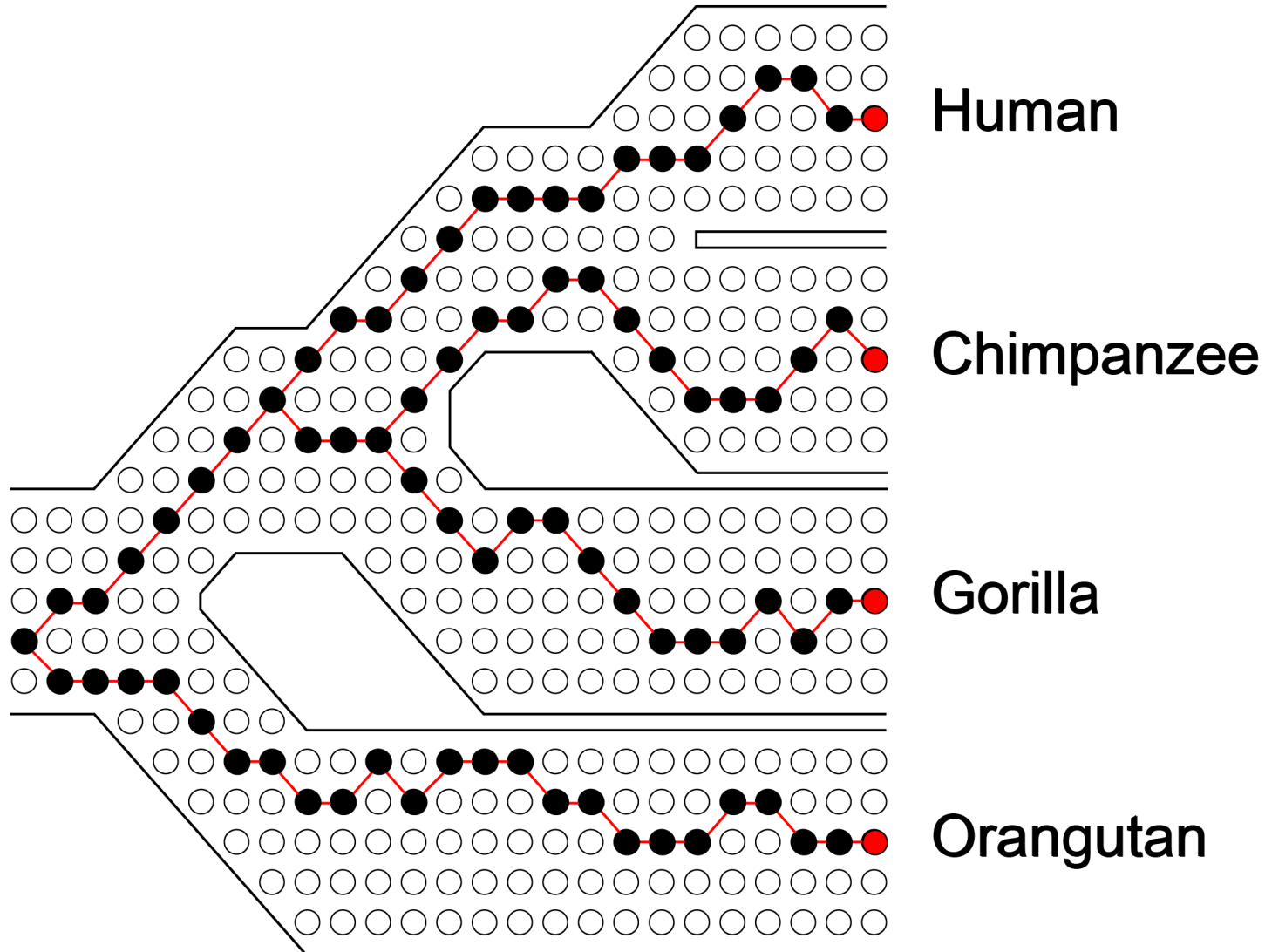
Incongruence among gene trees

- Phylogenetic analyses of genome-scale data sets must deal with incongruence among gene trees

  - **Incomplete lineage sorting**

    - Worse when speciation has been rapid

    - Worse when ancestral population sizes are large

  - Different direction and strength of selection

  - Stochastic variation in the mutational process
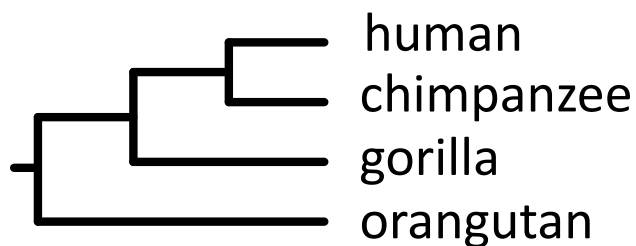
  - Biases in nucleotide composition

Species tree

- **Incomplete lineage sorting** can lead to gene trees that do not match the species tree

- We can infer the species tree from multiple gene trees even when they are incongruent

- Three approaches
  1. Consensus
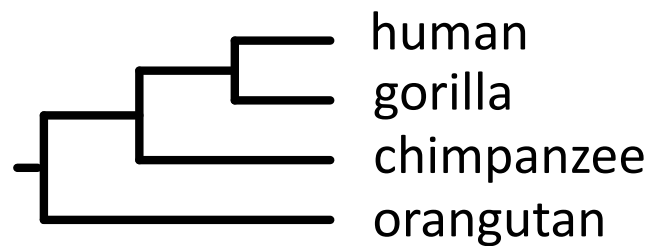 2. Concatenation
  3. Species-tree methods
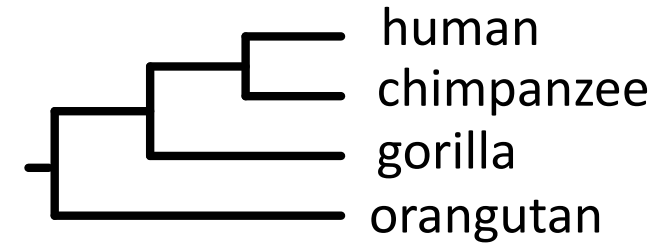
Species tree

1. **Consensus**

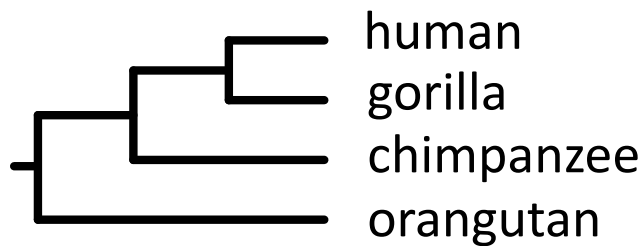   Estimate genealogy from each gene and find the consensus

Gene 1

human
chimpanzee
gorilla
orangutan

Gene 2

human
gorilla
chimpanzee
orangutan

Gene 5

human
chimpanzee
gorilla
orangutan

Gene 4

human
gorilla
chimpanzee
orangutan

Gene 5
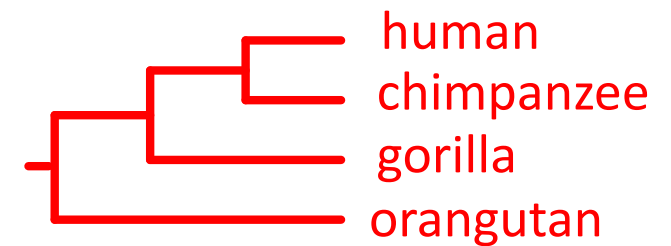
human
chimpanzee
gorilla
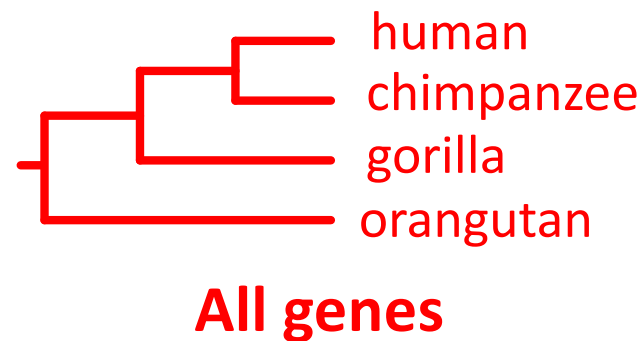orangutan

Consensus

human
chimpanzee
gorilla
orangutan
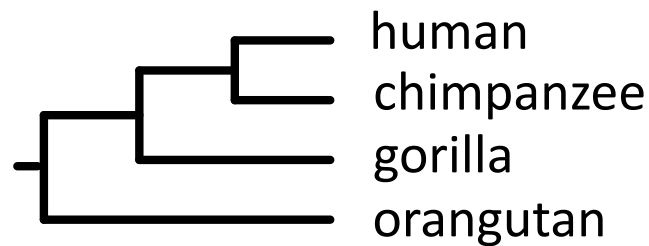
But the most frequent gene tree does not always match the true species tree ("anomaly zone")

Analysing multiple loci

2. **Concatenation**

    Assume that all genes share the same evolutionary history

All genes

But this ignores the occurrence of different gene trees

Species tree

3. Species-tree methods
   Estimate the species tree based on gene trees

- Gene trees are independent realisations of a stochastic process (the coalescent) on the same species tree

- Various methods

  - *BEAST* co-estimates gene trees and the species tree

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,*]

ASTRAL

- **A**ccurate **S**pecies **TR**ee **AL**gorithm

- Can analyse a genome-scale data set in minutes

- Finds the species tree with the highest agreement with 'quartets' among the gene trees

Useful references

- **Lineage sorting in apes**
  Mailund *et al*. (2014) *Annu Rev Genet*, 48: 519–535.

- **Estimating phylogenetic trees from genome-scale data**
  Liu *et al*. (2015) *Annals New York Acad Sci*, 1360: 36–53.