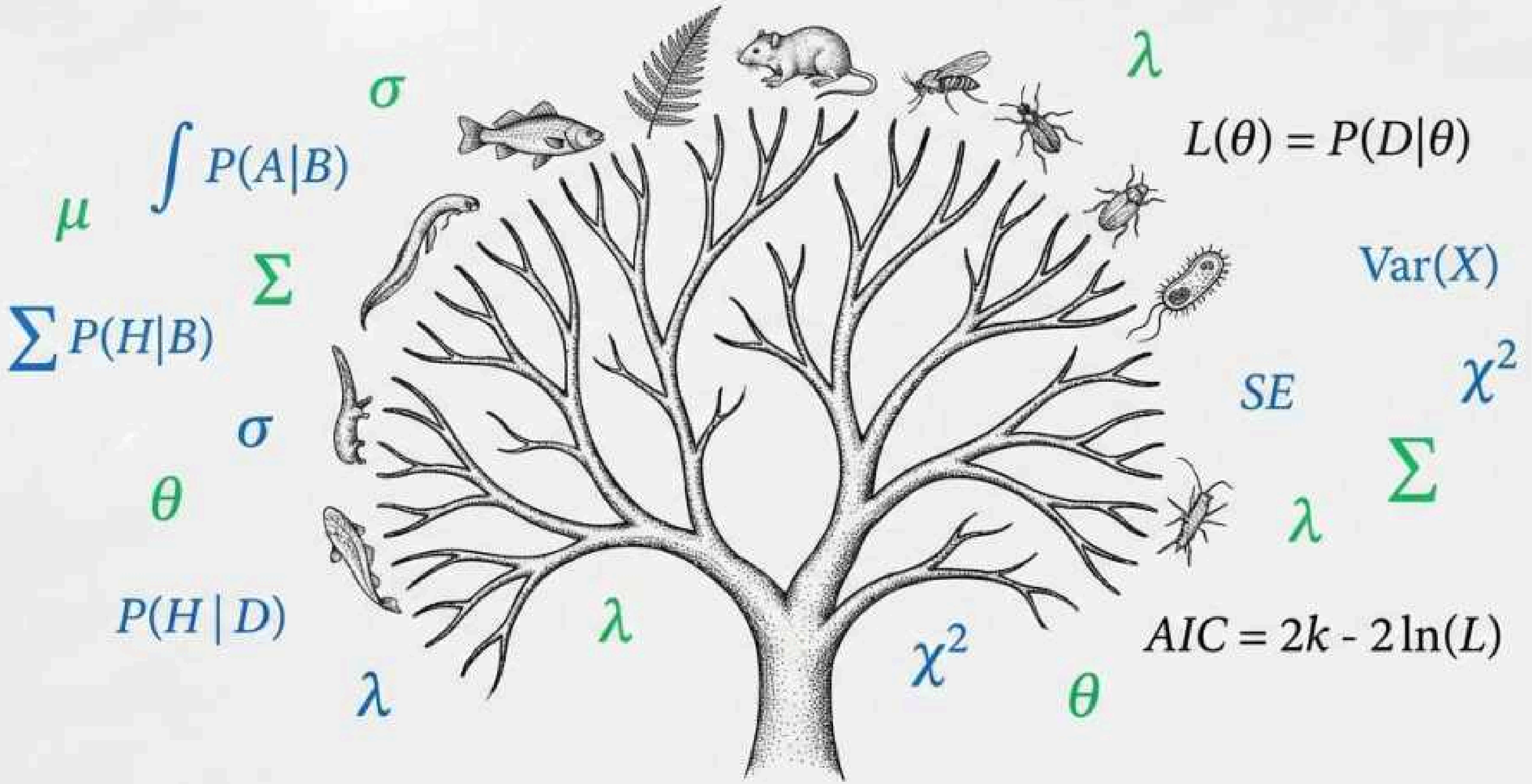
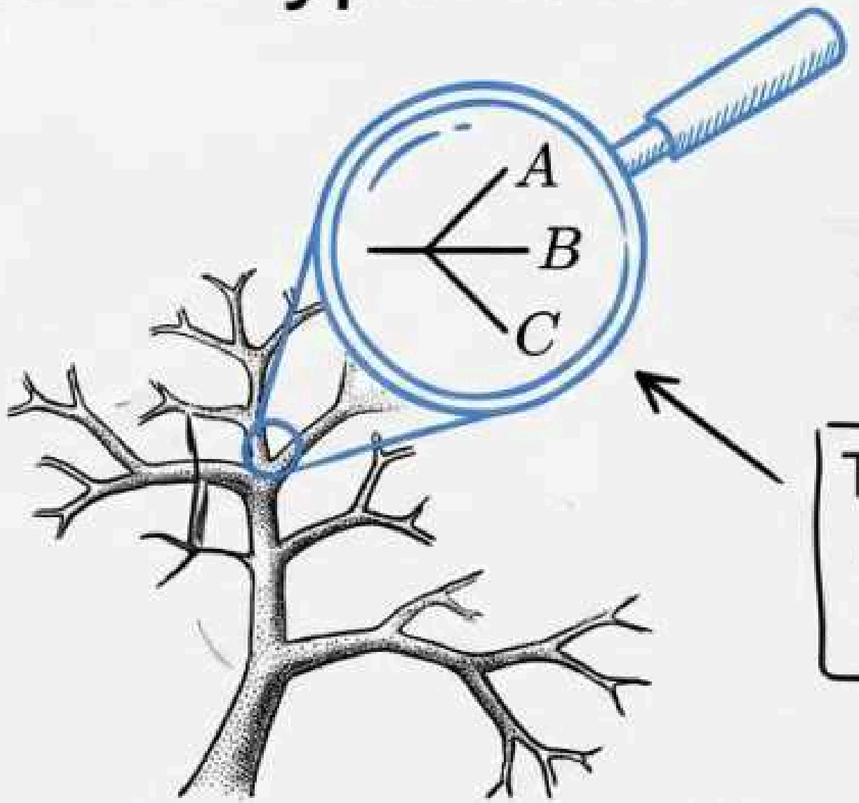


Building and Reading Trees

A Visual Guide to Evolutionary Phylogeny (BIO 340 B)

The Evolutionary Tree: A Hypothesis

¶ 2.1

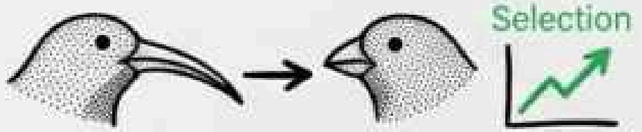


Trees are hypotheses of relationships, not absolute facts.

Essential for dating evolutionary events

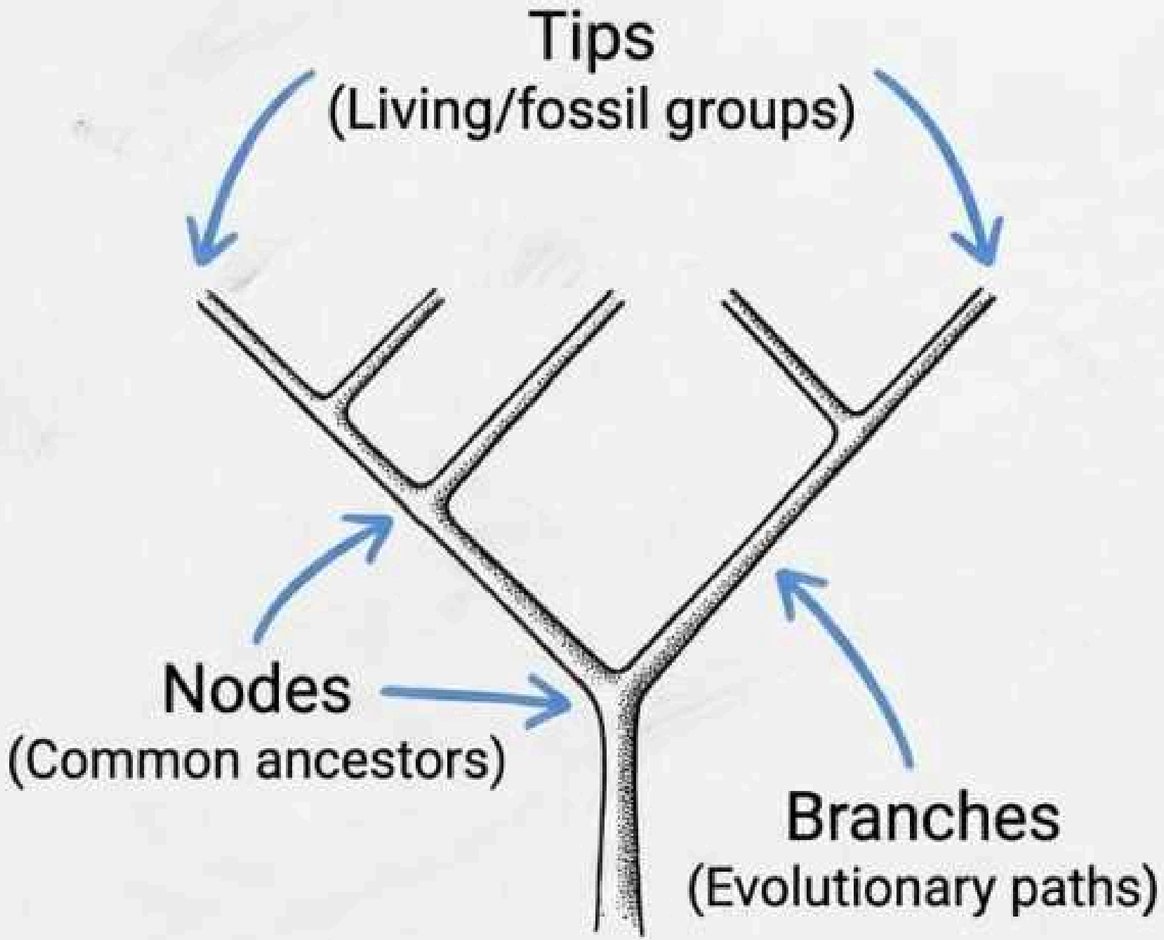


Crucial for detecting natural selection

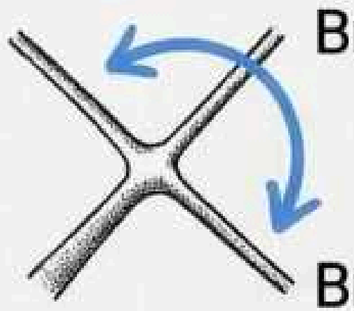


Anatomy of a Phylogeny

¶ 2.2



Caution

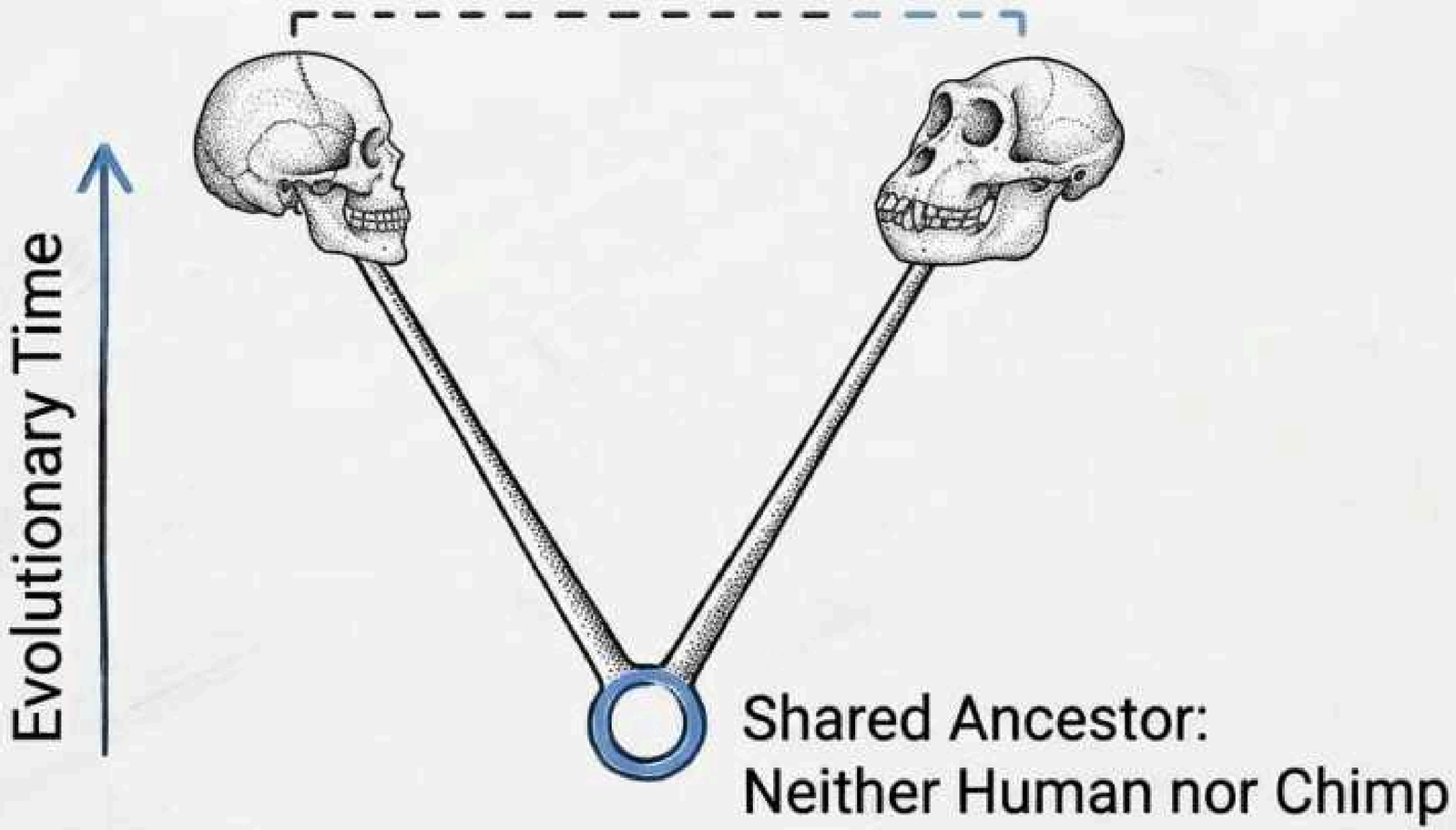


Order of tips carries no meaning.



Cousins, Not Ancestors

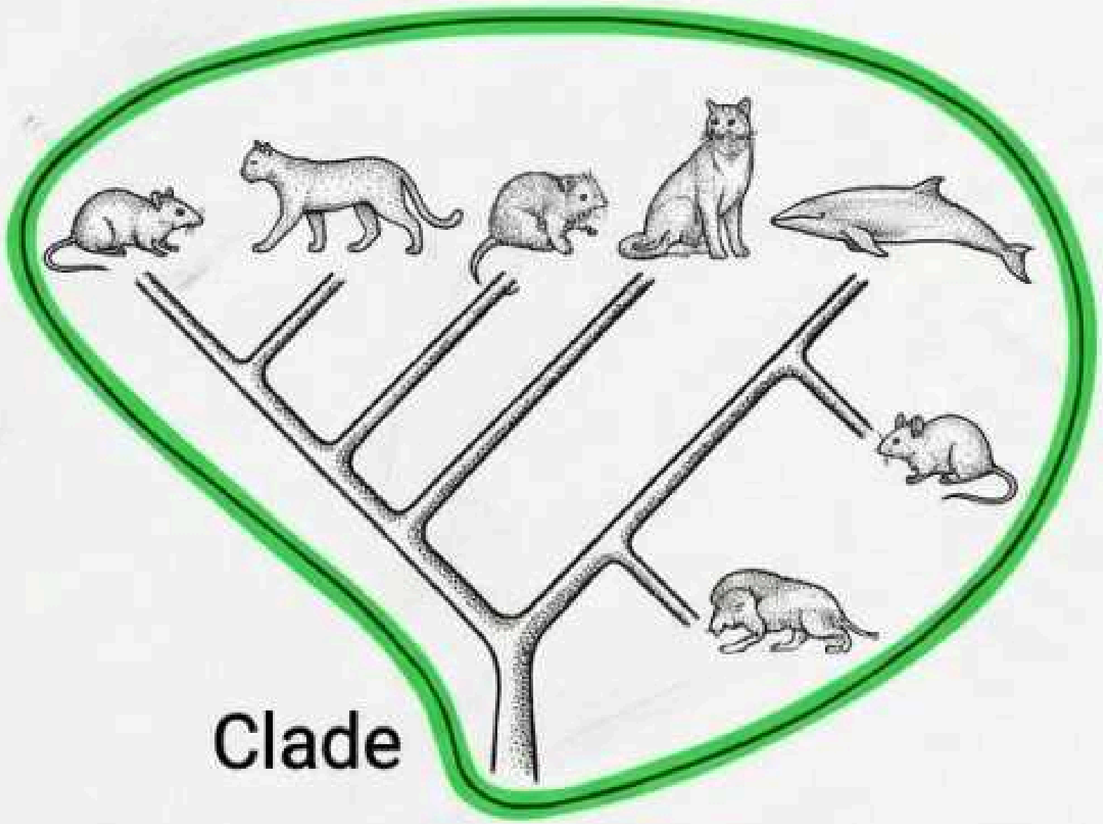
¶ 2.3



Humans did not descend from chimpanzees; they are sister groups. No living species is the ancestor of another.

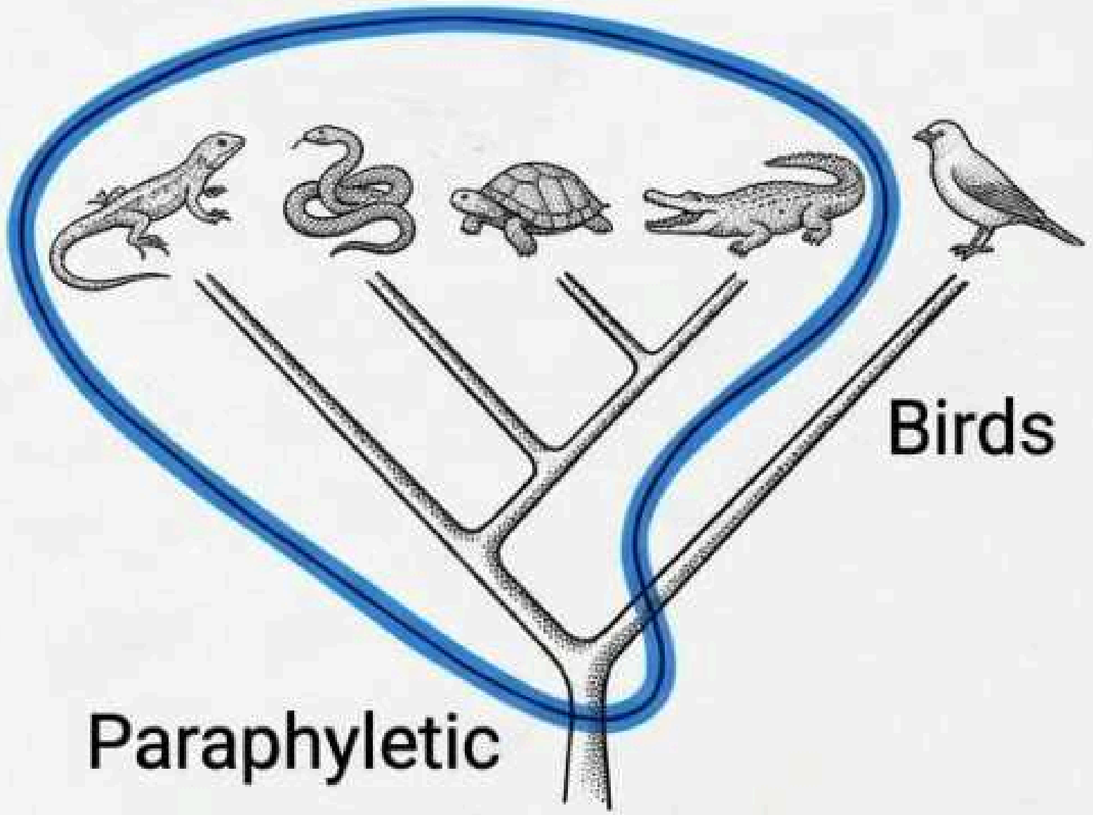
Naming the Tree: Clades vs. Paraphyly

¶ 2.4



Clade

An ancestor and all its descendants. (Modern classification standard).



Paraphyletic

Leaves out descendants (e.g., traditional reptiles exclude birds, despite a shared ancestor).

Cladistics & The Derived Character

¶ 2.5

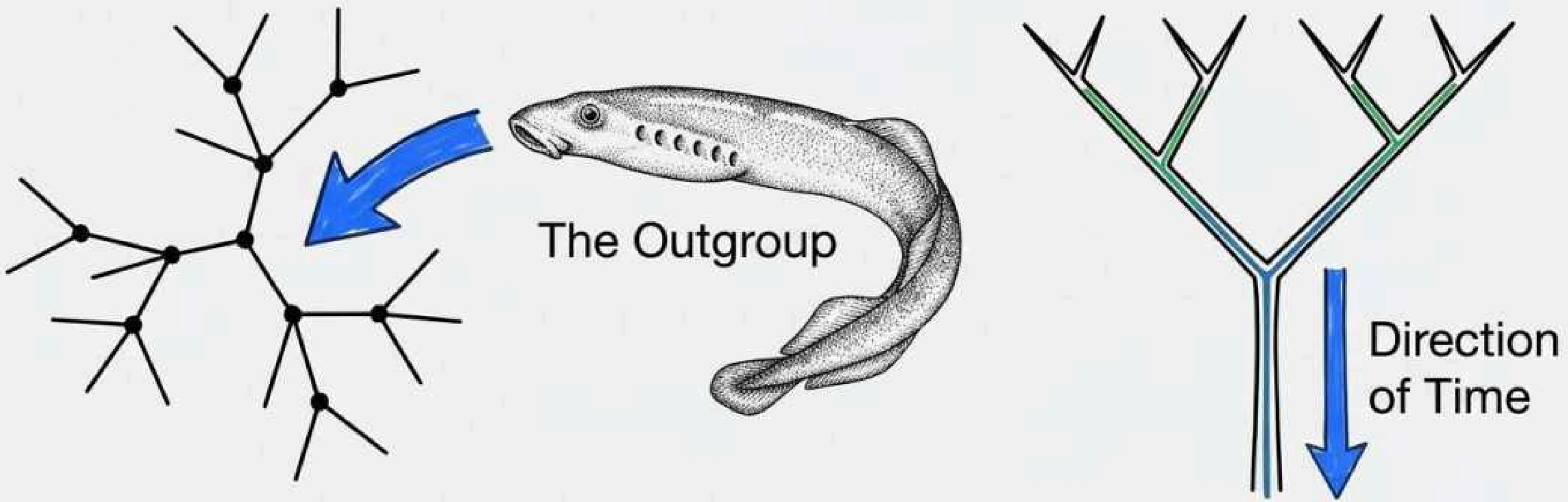


Only new features arising in a common ancestor reveal relationships.

Shared Derived Characters	Shared Ancestral Features
	
Novel traits that reveal recent relationships.	Inherited from far older ancestors. Tells us nothing about recent splits.

Rooting the Tree with an Outgroup

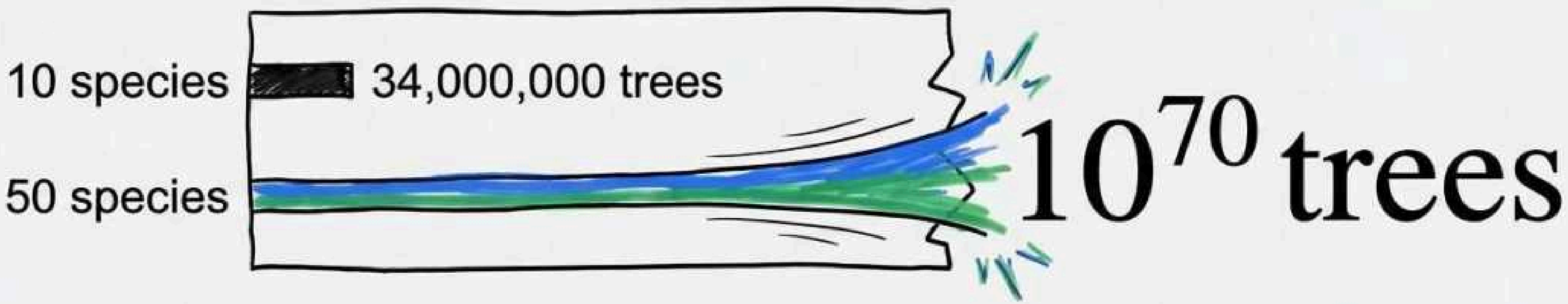
¶ 2.6



Function: A known relative outside the study group establishes which features are ancestral, providing both a root and a timeline.

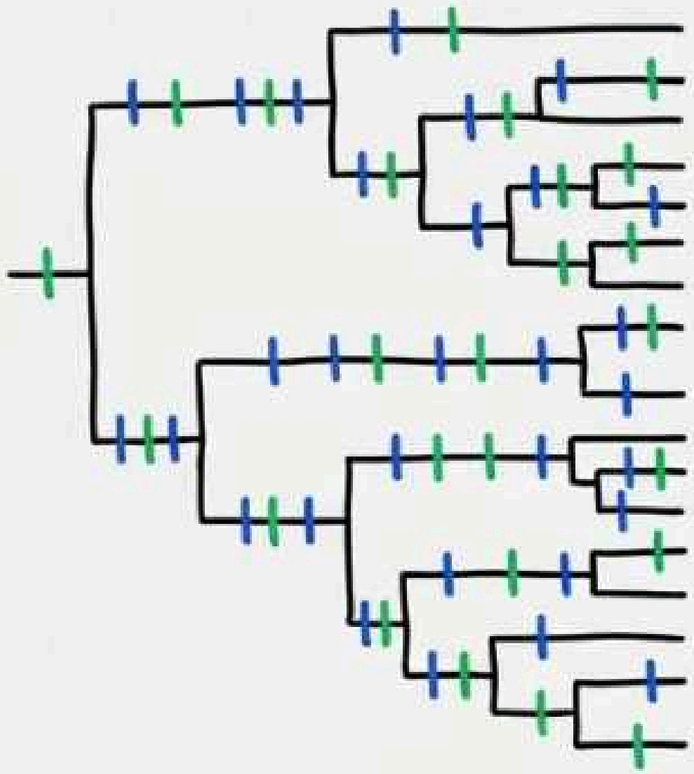
The Vastness of Tree Space

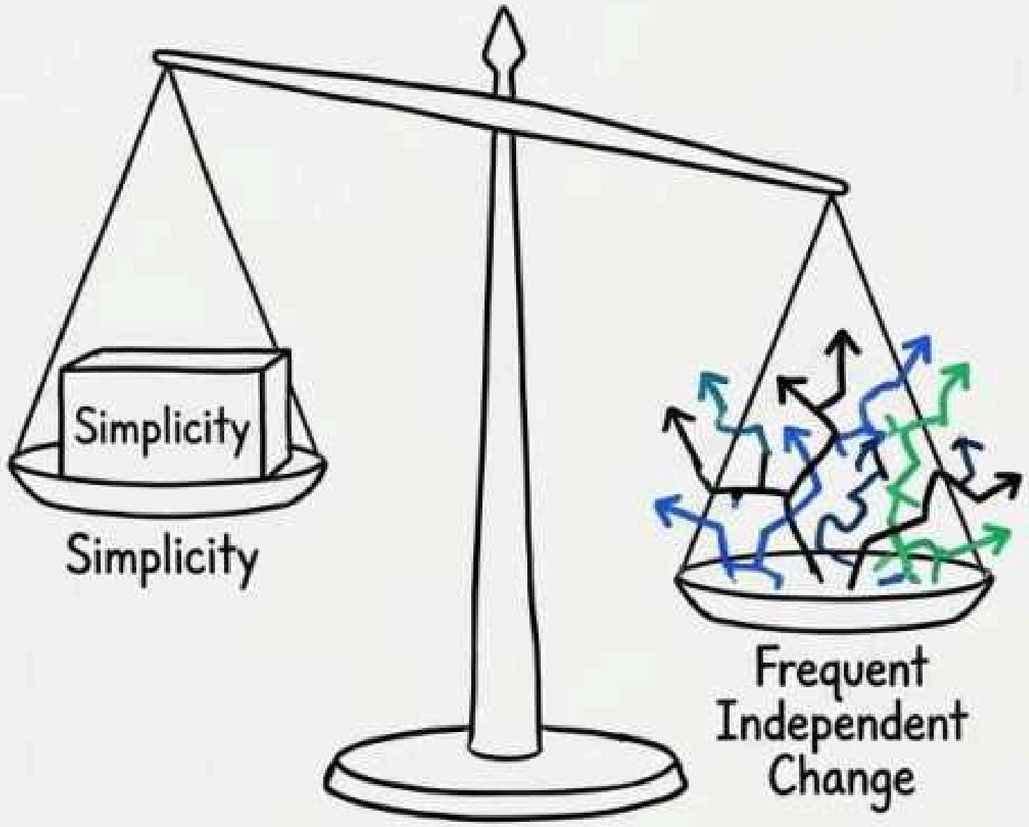
$$(2n\text{-}3)!!$$



Core Problem: The number of possible trees grows astonishingly fast. We cannot check every possibility; computers must use clever search algorithms rather than brute force.

Method 1: Maximum Parsimony





Mechanism: Choose the tree requiring the fewest evolutionary changes.

Vulnerability: Works well when change is rare. Misled when change is frequent (traits evolving independently). Lacks an explicit model of character change.

Method 2: Distance Methods

	A	B	C
A	0	0.15	0.35
B	0.15	0	0.2
C	0.35	0.2	0
D	0.15	0.15	0

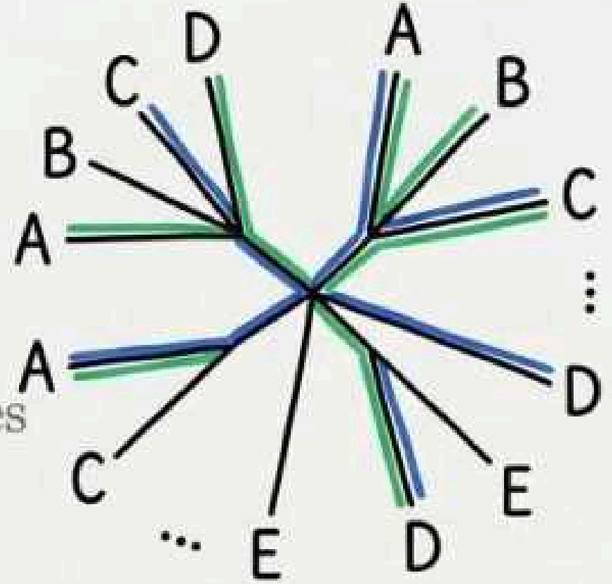
DNA Difference Matrix

$$D = -b \ln\left(1 - \frac{p}{b}\right)$$

where D = ~~corrected~~ distance

p = proportion of differences

b = a parameter



star-decomposition

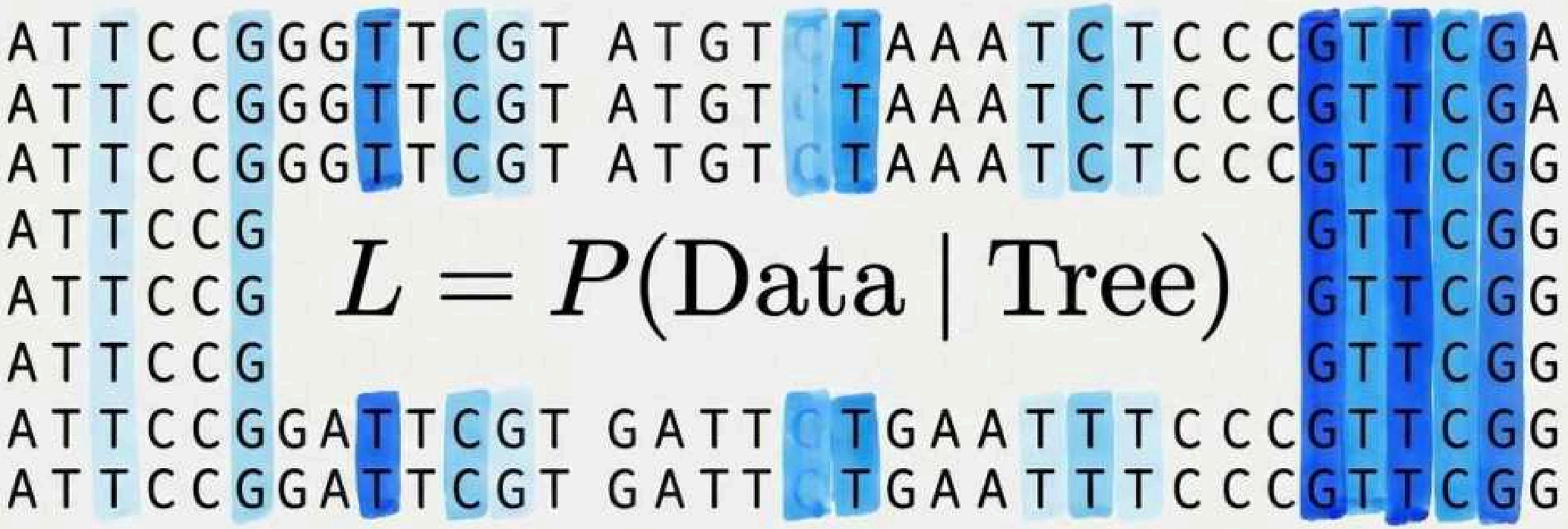
Mechanism: Calculate differences between pairs (e.g., DNA site differences) to build a tree fitting those distances, to build a tree fitting those distances.

Key Tool: Neighbor Joining (Saitou & Nei, 1987) — fast and widely used.

Requirement: Distances must be explicitly corrected for hidden multiple changes at a single site.

for hidden multiple changes at a single site.

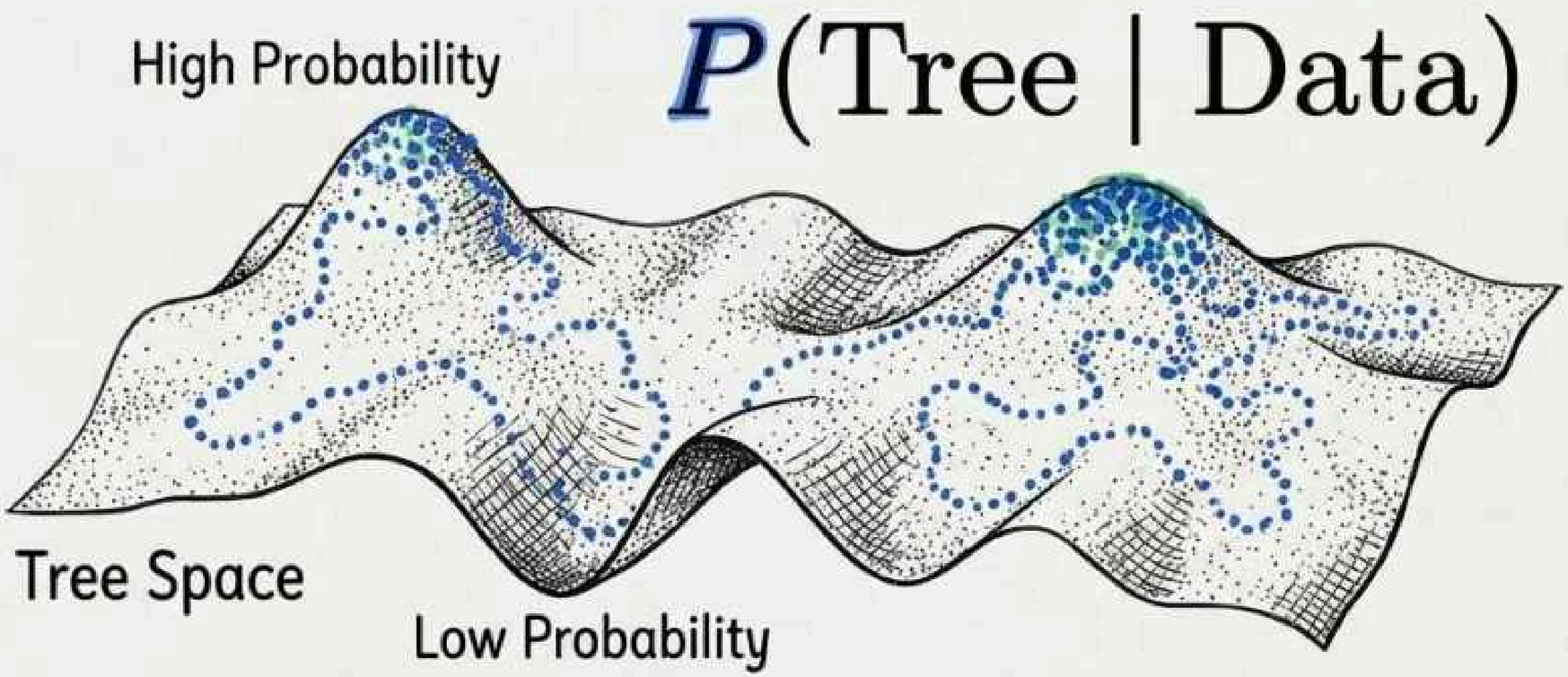
Method 3: Maximum Likelihood



Mechanism: Asks which tree makes the observed data most probable under an explicit model of sequence change (Felsenstein, 1981).

Advantage: Explicitly models how often bases change and how rates vary across sites. Robust against parsimony errors.

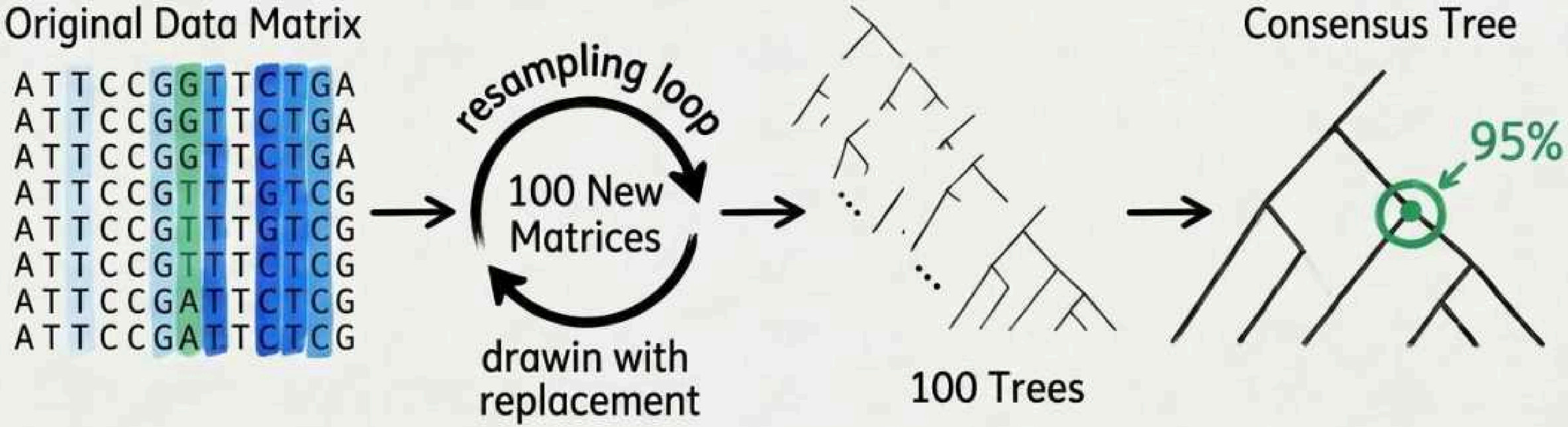
Method 4: Bayesian Inference



Mechanism: Combines likelihood with prior beliefs. MCMC wanders among trees in proportion to their probability.

Result: Yields a large sample of trees, directly showing uncertainty, rather than just one single tree.

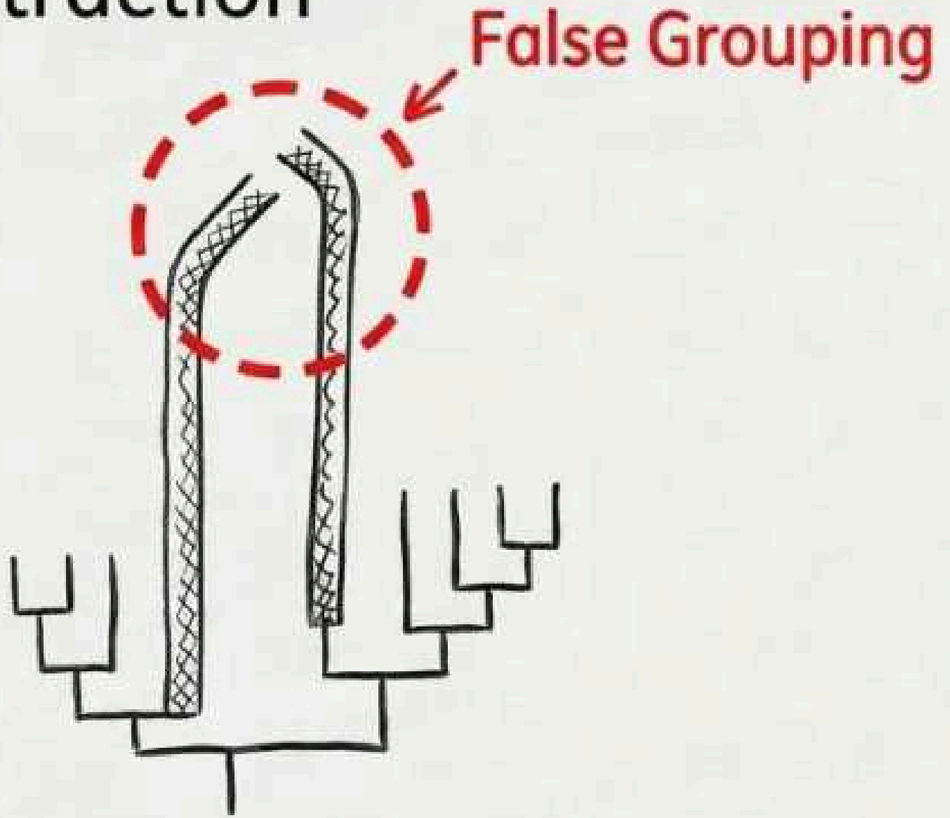
Measuring Confidence: The Bootstrap



Process: Introduced by Felsenstein (1985). Resample data, build trees, count how often a group appears.

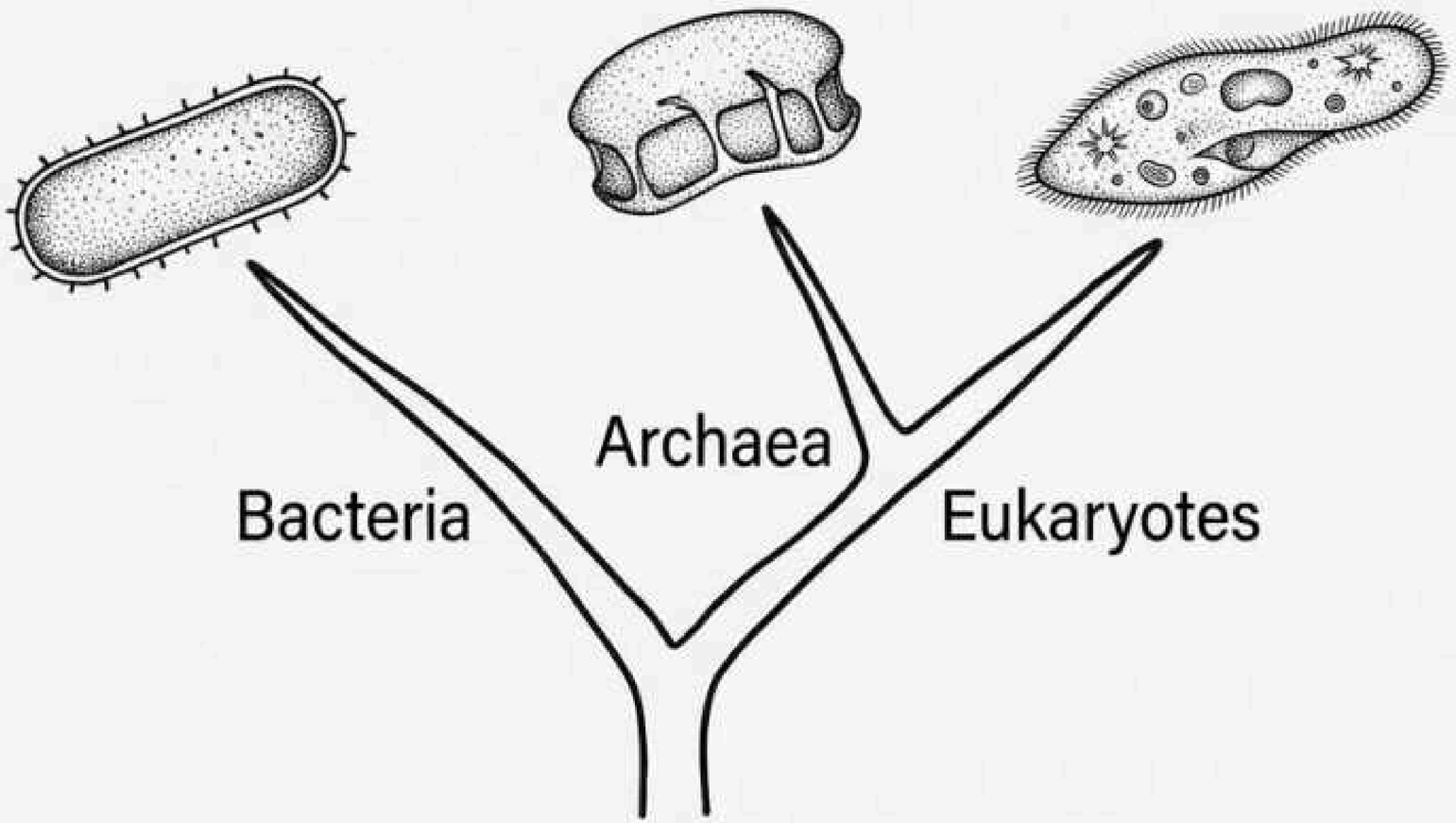
Scale: 95% = Strong support. 50% = Weak support. (Bayesian posterior probabilities play a similar, often higher, role).

Pitfall: Long-Branch Attraction



- The Trap: Unrelated, rapidly changing lineages share changes by chance. Parsimony often wrongly groups them together.
- Remedies: Add species to break up long branches, use better likelihood models, or remove the fastest-changing sites.

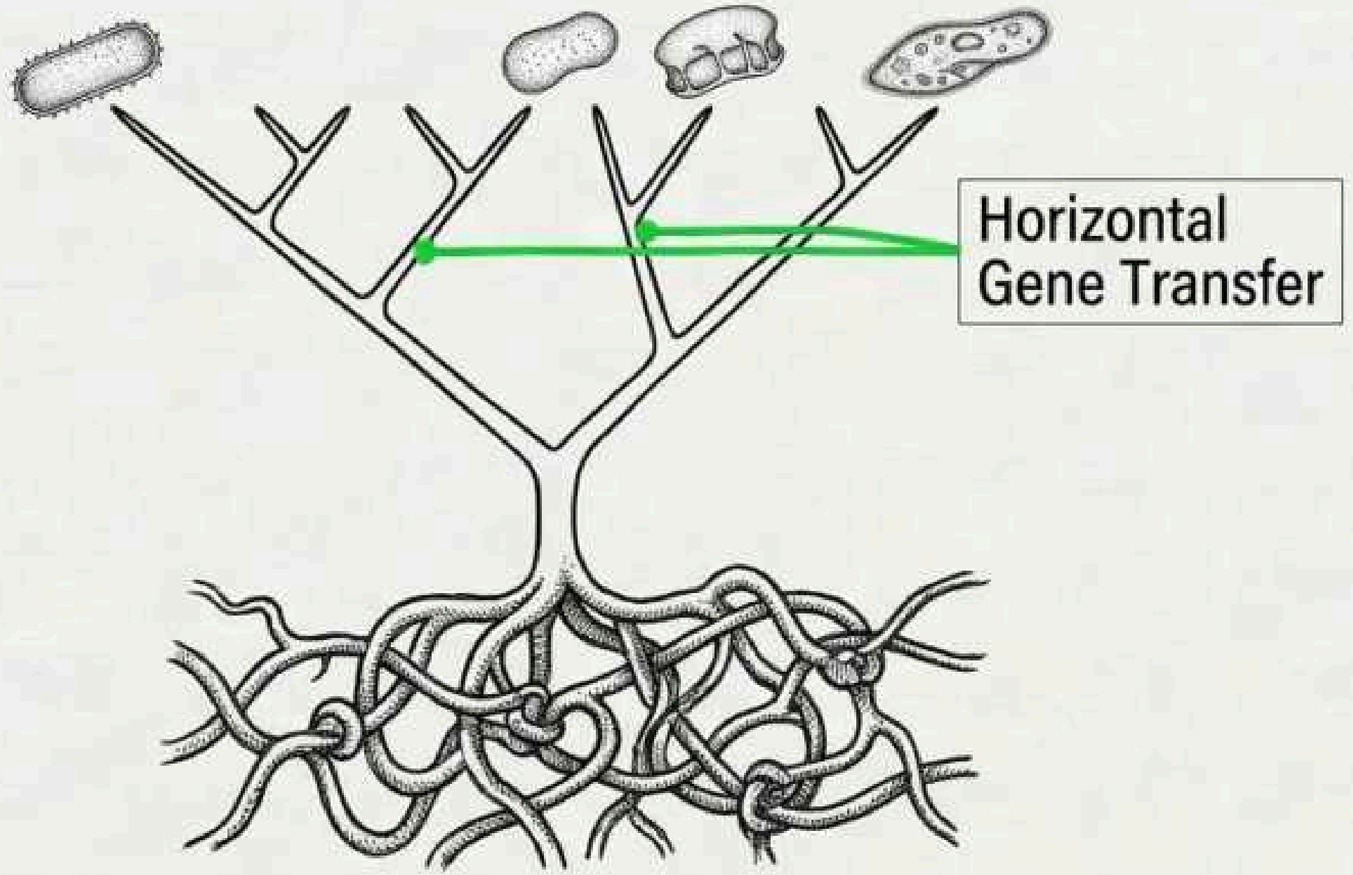
Molecules Transform the Tree



Discovery: Ribosomal RNA comparisons by Woese & Fox (1977) revealed Archaea.

The 3 Domains: Bacteria, Archaea, and Eukaryotes (plants, animals, fungi).

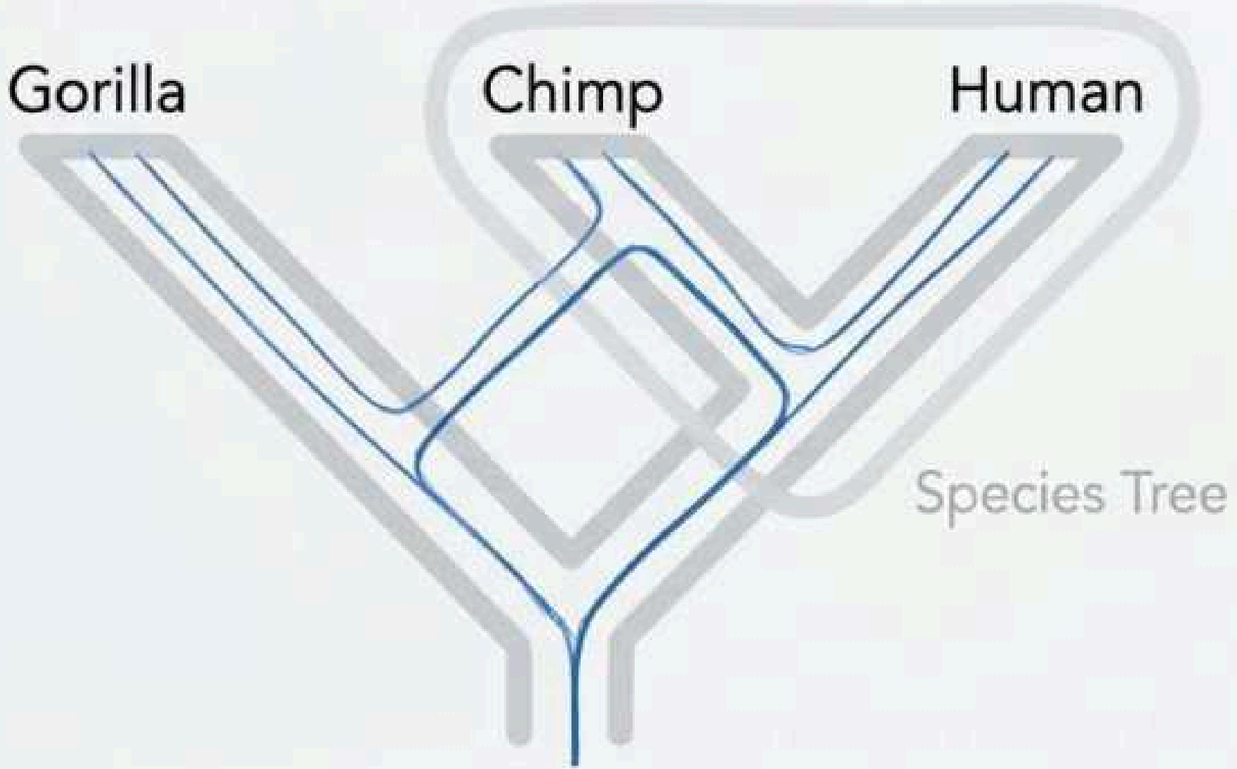
The Tangled Web of Microbes



Mechanism: Horizontal Gene Transfer (HGT). Bacteria and archaea exchange genes directly (e.g., antibiotic resistance).

Result: Different microbial genes have different histories. The root of life is a web, though animal/plant trees remain standard.

Incomplete Lineage Sorting



The Phenomenon: Ancestral populations carry multiple gene versions that sort differently into descendant species.

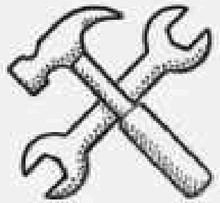
Impact: Roughly one-third of the human/chimp/gorilla genome contradicts the main species tree. Modern methods must analyze many genes together to find the true species tree.

Lecture Summary: Building and Reading Trees



Reading Trees

Tips are cousins; clades rely on shared derived characters.



Building Trees

Parsimony, Distances, Likelihood, Bayesian.



Pitfalls

Long-branch attraction, HGT, Incomplete lineage sorting.



Next Up

Putting dates on the tree.