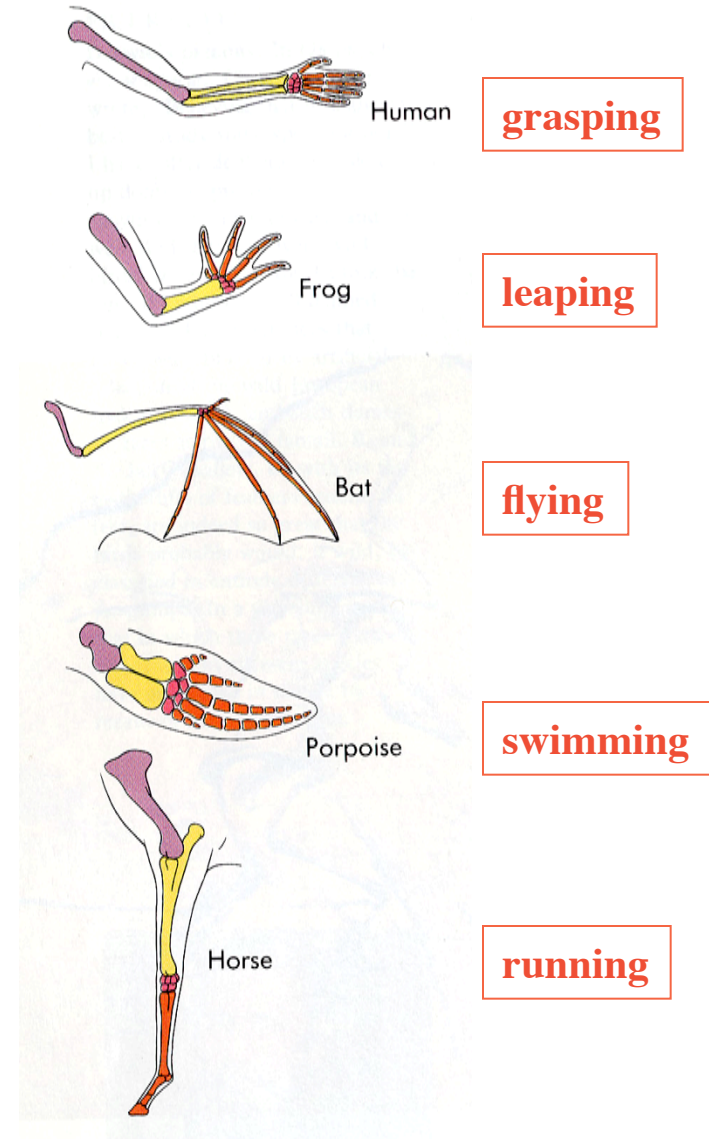
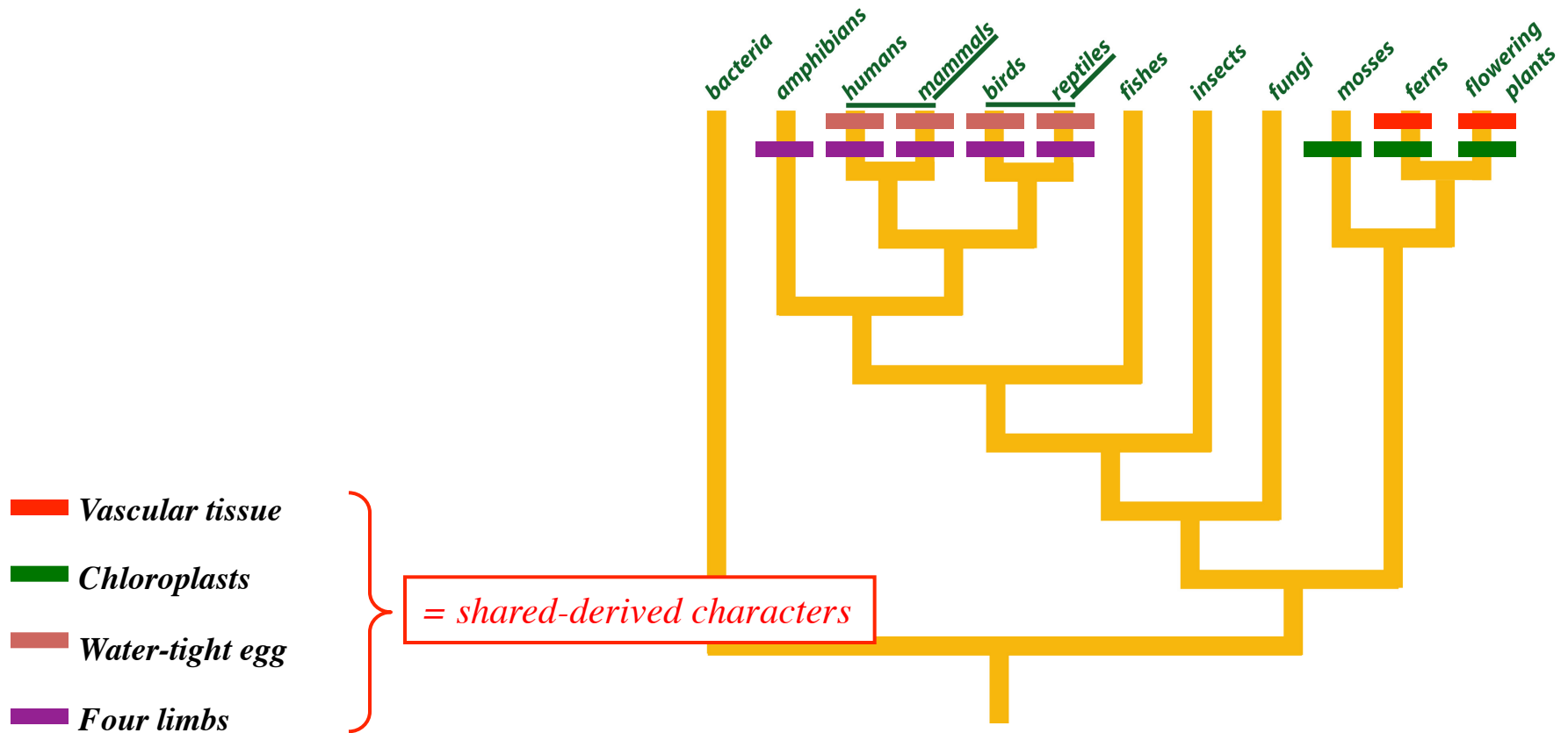


# Homology

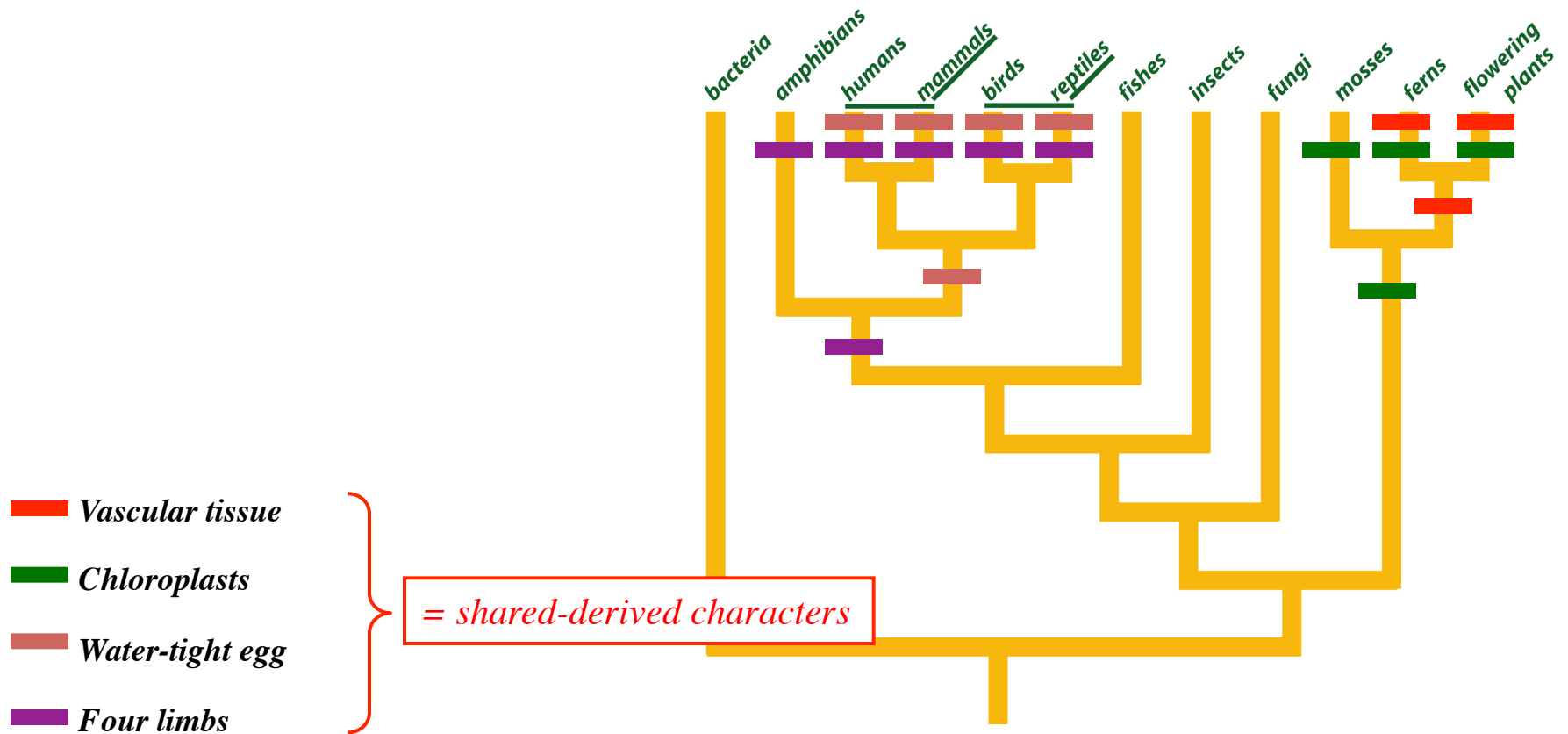
- Distantly related organisms share structural similarities
- Function varies
- Explicable by common ancestry



# Nested hierarchical structure "groups within groups"



# Nested hierarchical structure "groups within groups"

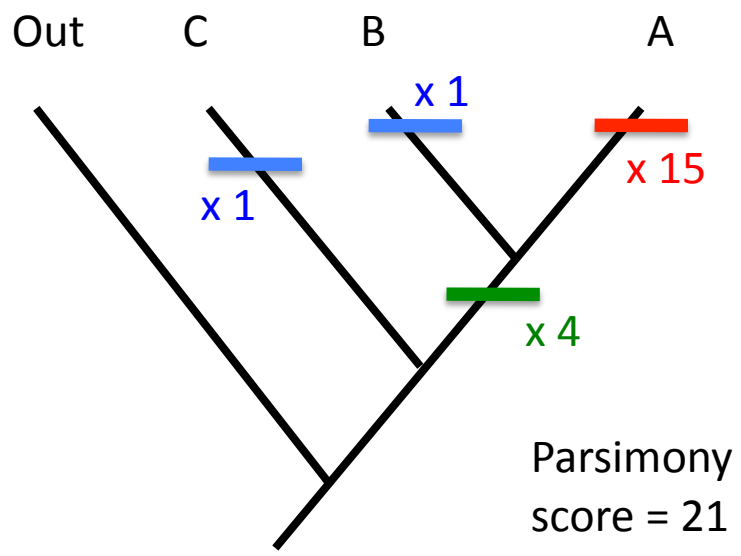
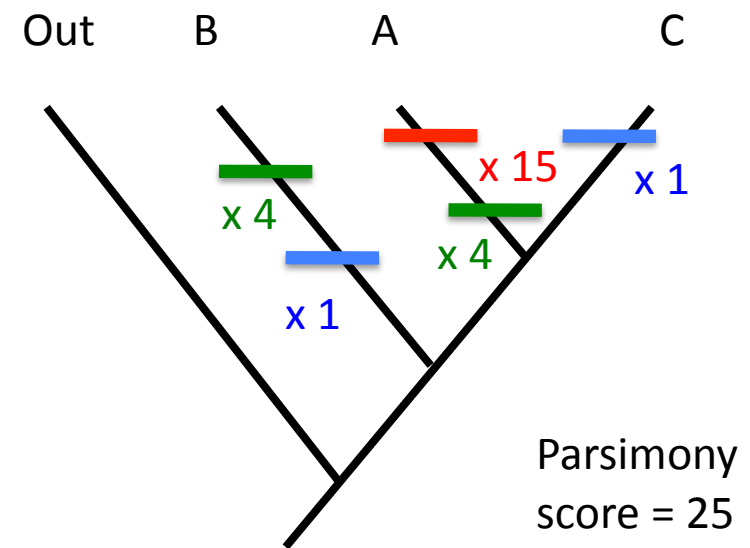
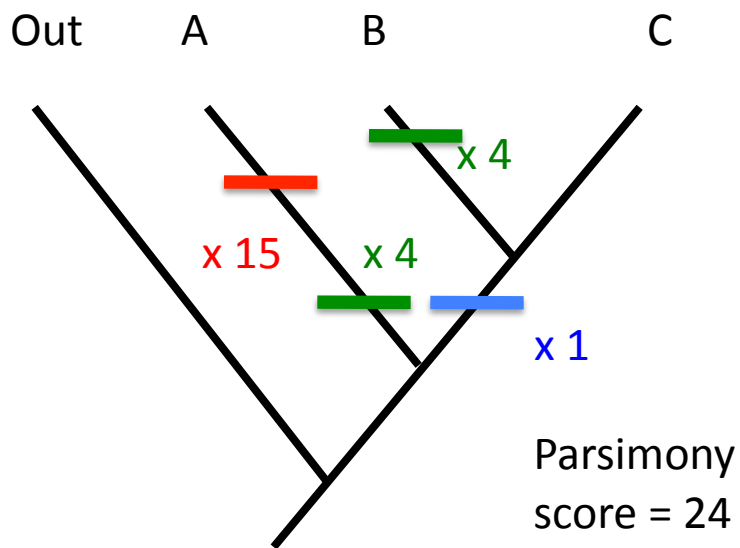


# Parsimony – simple example

	A	B	C
Characters 1-15	1	0	0
Characters 16-19	1	1	0
Character 20	0	1	1
Outgroup	0	0	0

Taxa B + C are more similar in terms of shared character states --- (they agree on 16/20 characters)

**But which tree is the most parsimonious?**



	A	B	C
Characters 1-15	1	0	0
Characters 16-19	1	1	0
Character 20	0	1	1
Outgroup	0	0	0

# The Data - Genetic Sequences

## Sequence alignment issues

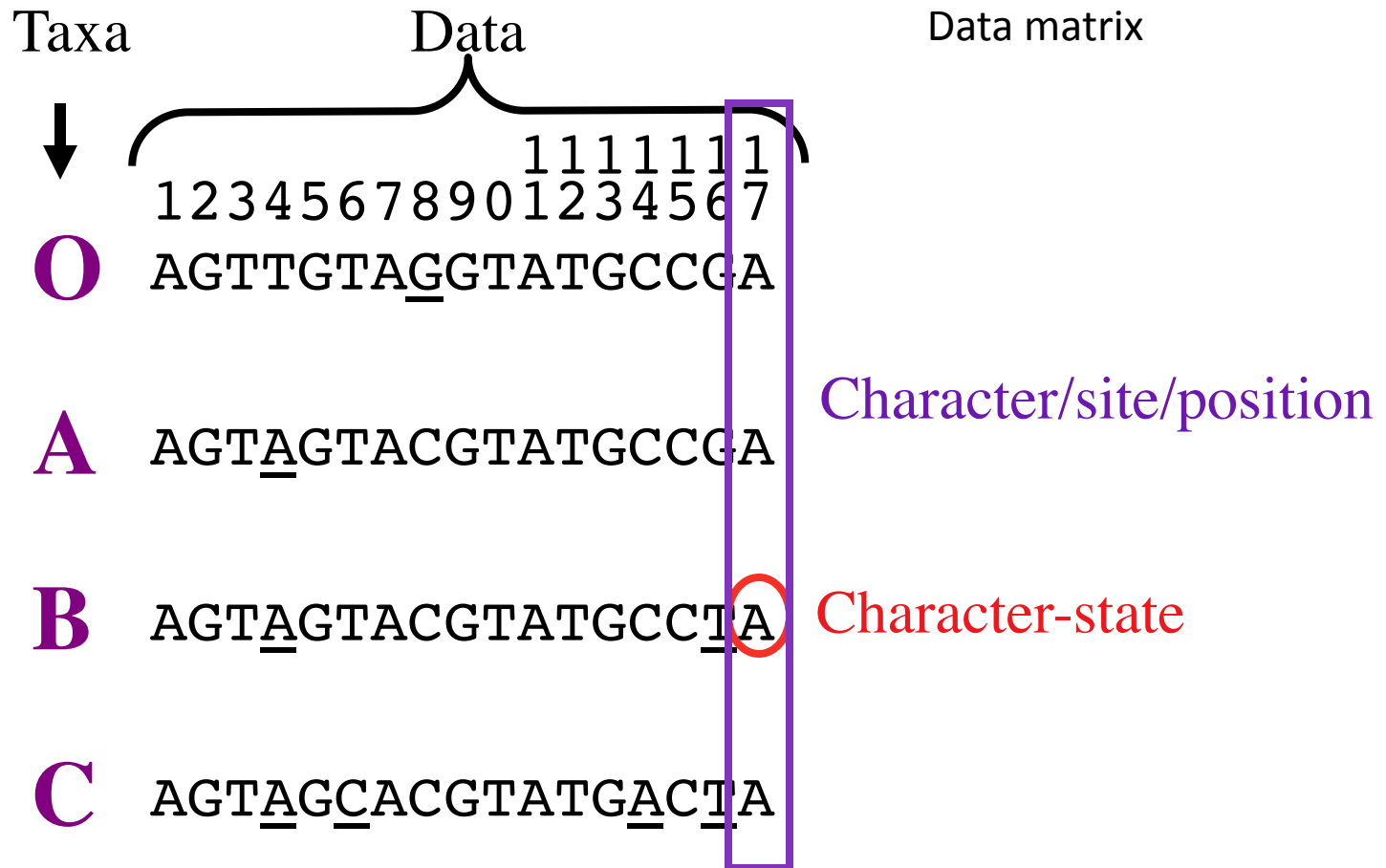
The screenshot displays the BioEdit Sequence Alignment Editor interface. The window title is "BioEdit: Sequence Alignment Editor". The menu bar includes "File", "Edit", "Sequence", "Alignment", "View", "Accessory Application", "RNA", "World Wide Web", "Options", "Window", and "Help". The main window shows a file named "A:\cox2,3.bio" with 9 total sequences. The font is set to Courier New, size 11, and bold. The mode is "Select / Slide". The sequence mask and numbering mask are both set to "None". The start ruler is at position 1. The scroll speed is set to "slow".

The alignment shows the following sequences:

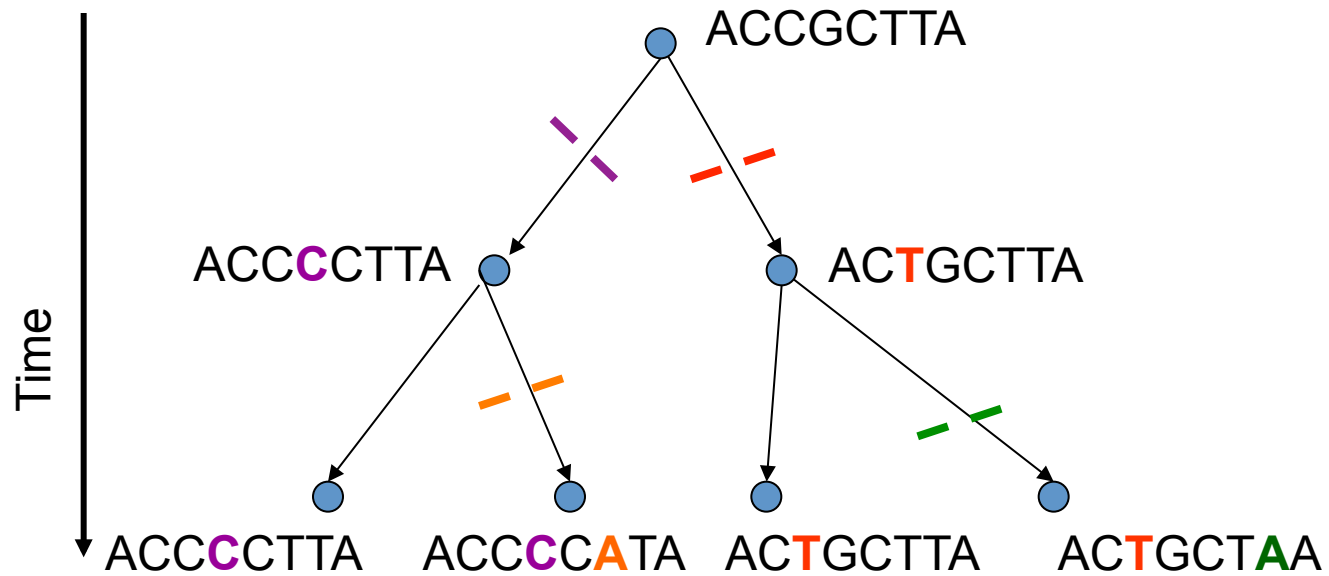
- KK01
- KK02-cox2 se
- KK03-cox2 se
- KK04-cox2 se
- KK05-cox2 se
- KK06-cox2 se
- KK07-cox2 se
- KK08-cox2 se
- KK09-cox2 se

A blue box highlights a region of misalignment between sequences KK01 and KK09. In KK01, the sequence is "...CCNACTNCGNTGGCAGGTTAAAGAGAGGGTATATACTATGGGACAGTGCAGCGAAATTTGTGGCATAAACCATGGTTTTATGCCATAGTTGTGGAAGCTGTTTCGTTACCA". In KK09, the sequence is "...NN~GGGGCGTCCNCCTGCGGAGCCANGTTAAAGAGAGGGNATNTNCTNTGGACAGTGCAGCGAAATTTGTGGCATAAACCATGGTTTTATGCCATAGTTGTGGAAGCTGTTTCGTTACCA". The blue box highlights the region from approximately position 40 to 60, where the sequences are not aligned.

# Specific procedure



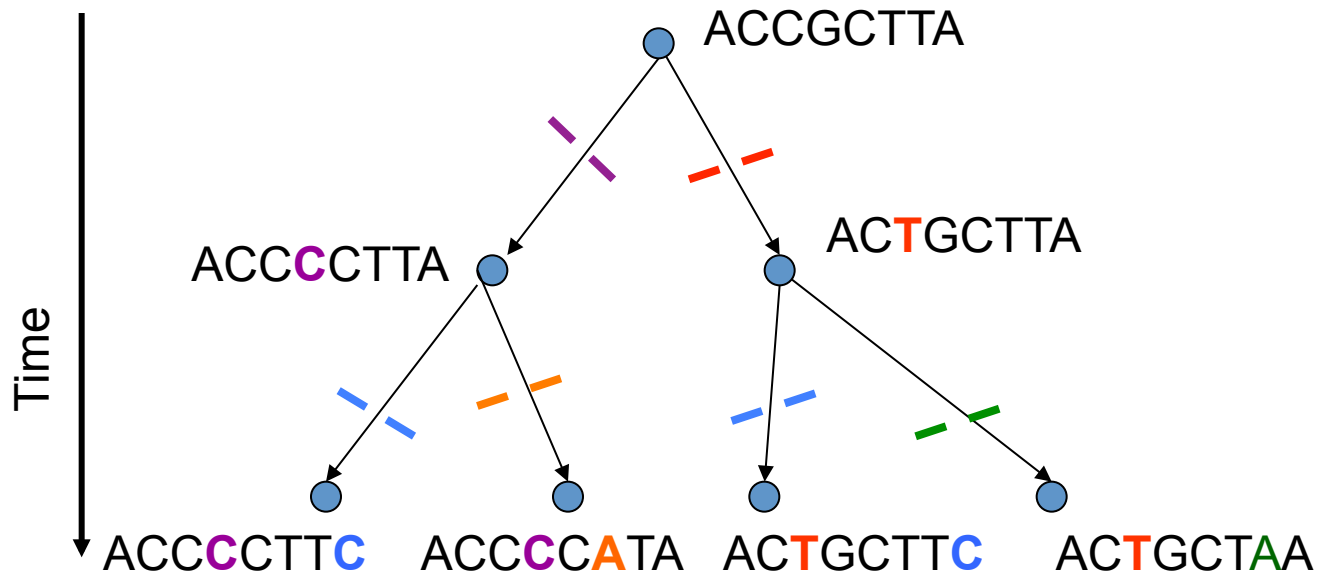
# Parsimony - data agrees



ACCCCTTA  
ACCCATA  
ACTGCTTA  
ACTGCTAA



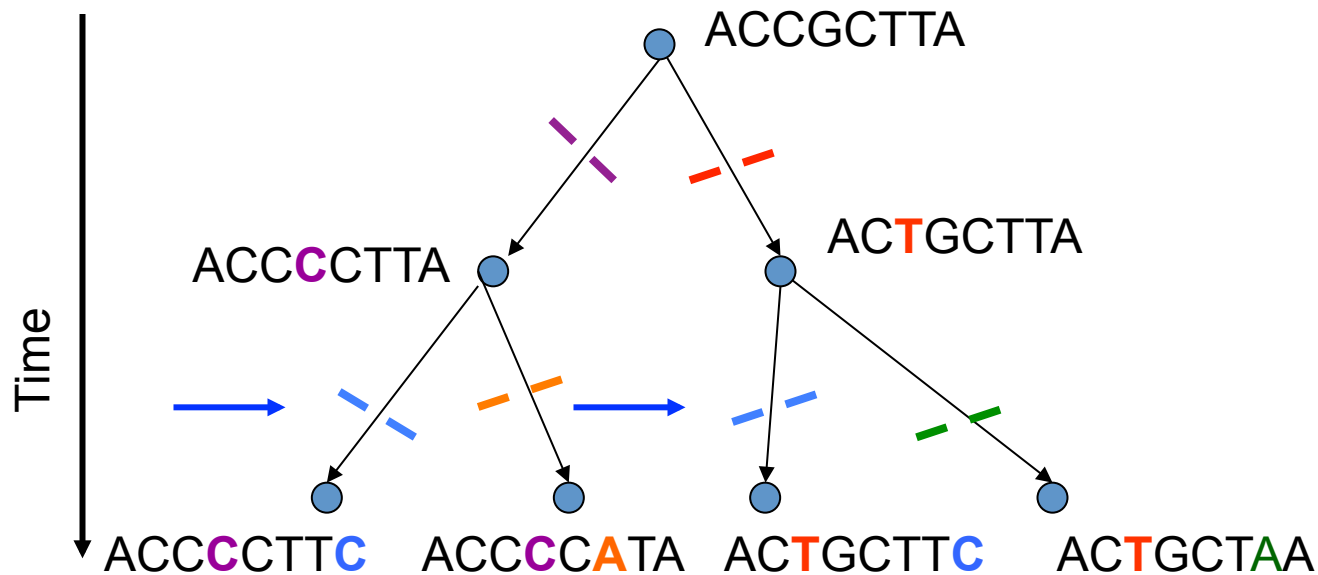
# Parsimony - data disagrees



ACCCCTTC
ACCCATA
ACTGCTTC
ACTGCTAA

# Homoplasy

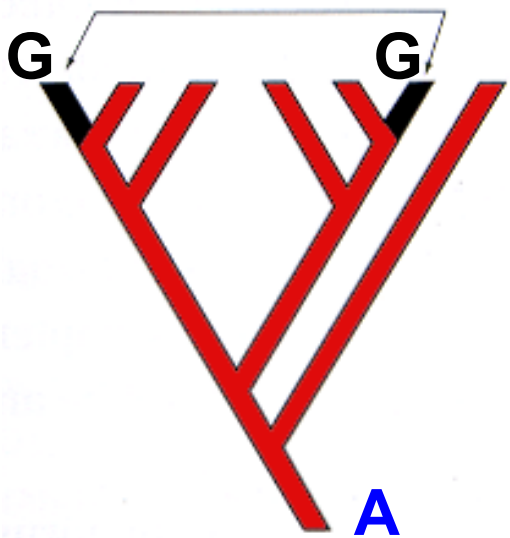
When two or more characters can't possibly fit on the same tree without requiring one character to undergo a parallel change or reversal



# Parsimony - homoplasy

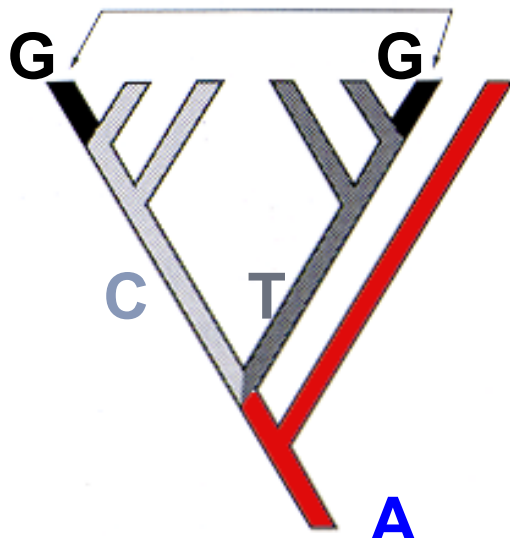
## *Parallelism*

Independent evolution of same feature from same ancestral condition



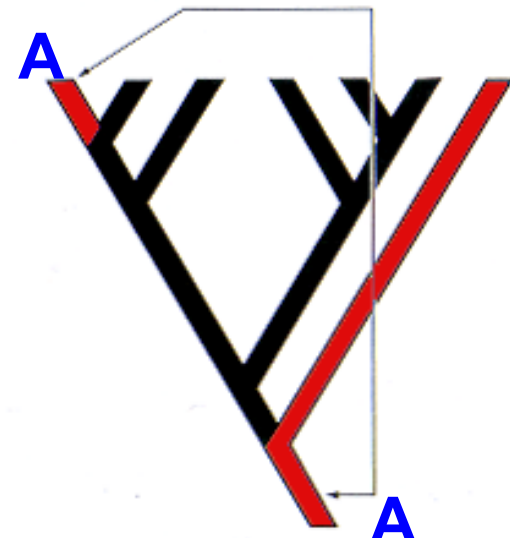
## *Convergence*

Independent evolution of same feature from different ancestral condition



## *Reversal*

Reversion to ancestral condition



Three kinds of homoplasy

# Parsimony can still work (According to some people)

- If characters are independent (a key assumption), **homoplasy will be randomly distributed**
- Homoplasies will tend to cancel each other out
- Non-homoplastic changes will tend to agree
- Therefore, **with enough characters** the shortest tree is a good estimate of the true tree

# Real Example

- Any real instance of phylogenetic inference will follow the logic of what we were just talking about, but with sophisticated statistical modeling, it is not always obvious.
- For pretty clear cases, see the most recent issue of *Cladistics* such as:
  - <http://dx.doi.org/10.1111/j.1096-0031.2009.00303.x/>