
Lecture 1

Introduction to molecular phylogenetics

Phylogenetic trees

What is a phylogenetic tree?

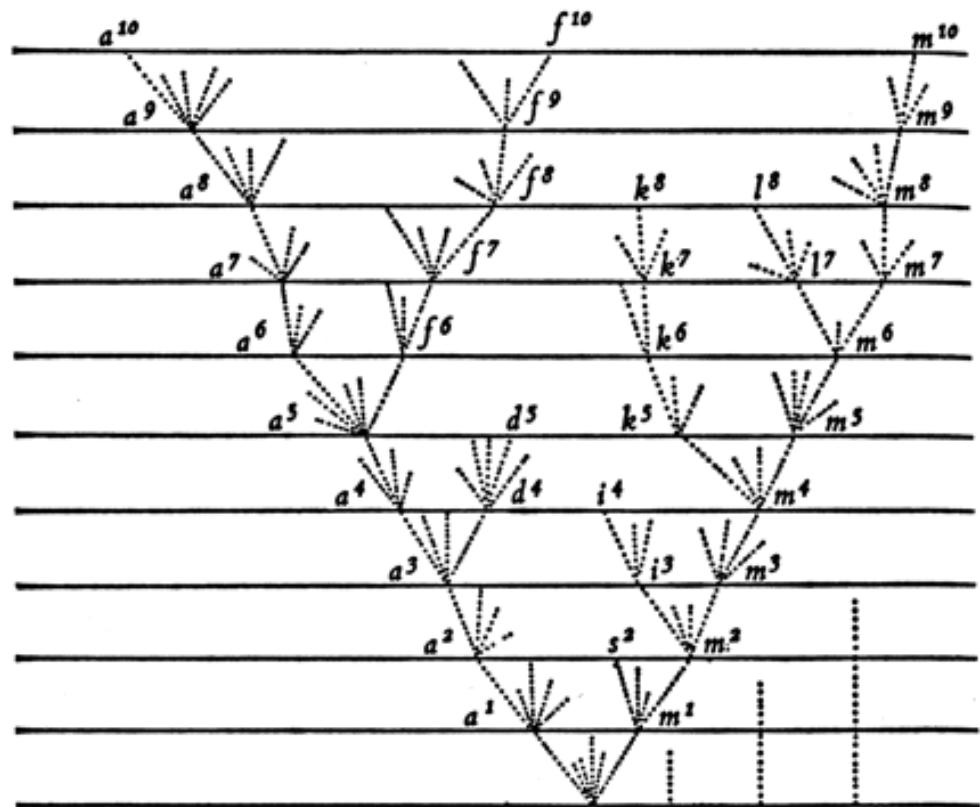
Phylogeny

Evolutionary relationships
among a group of organisms

I think



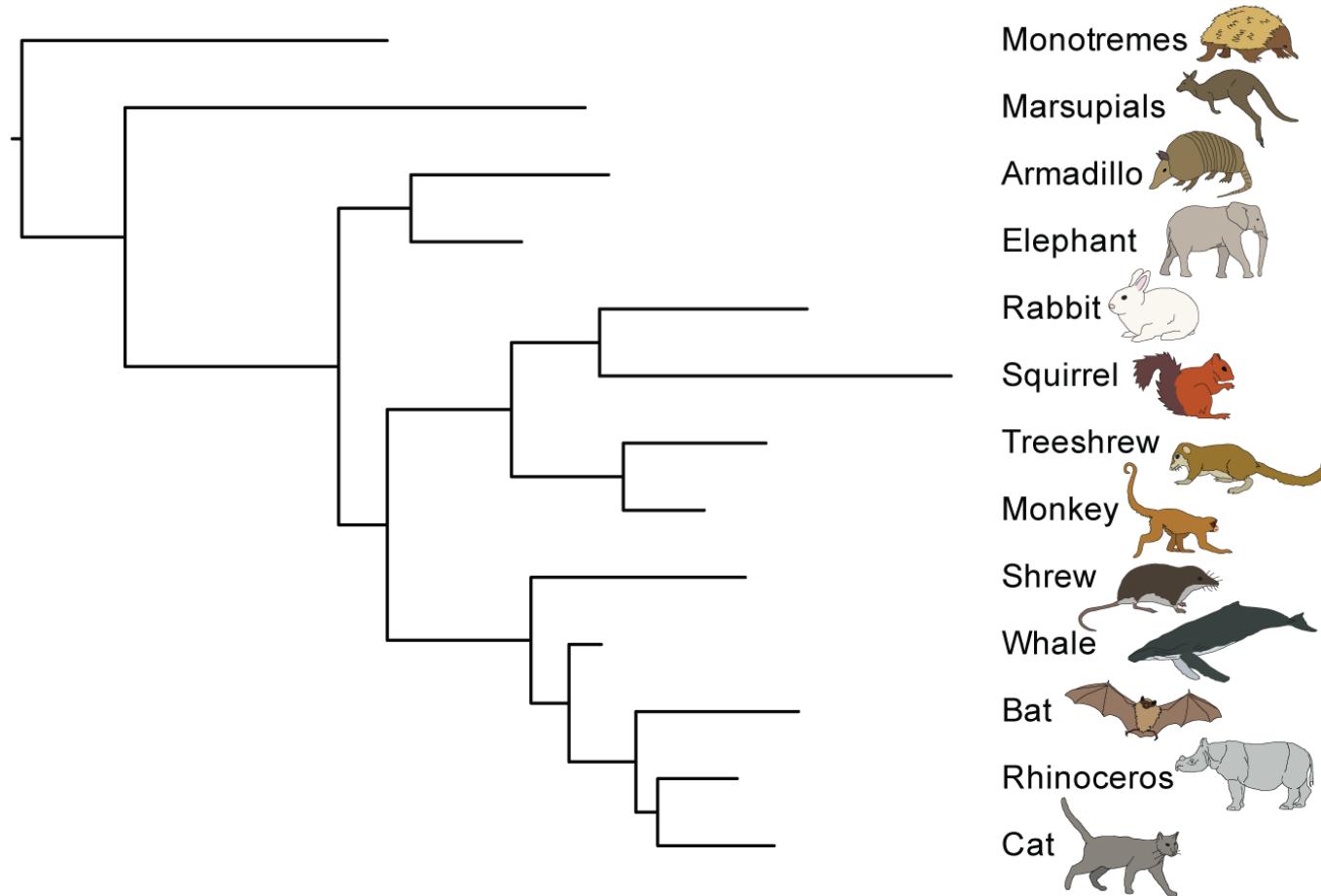
1837

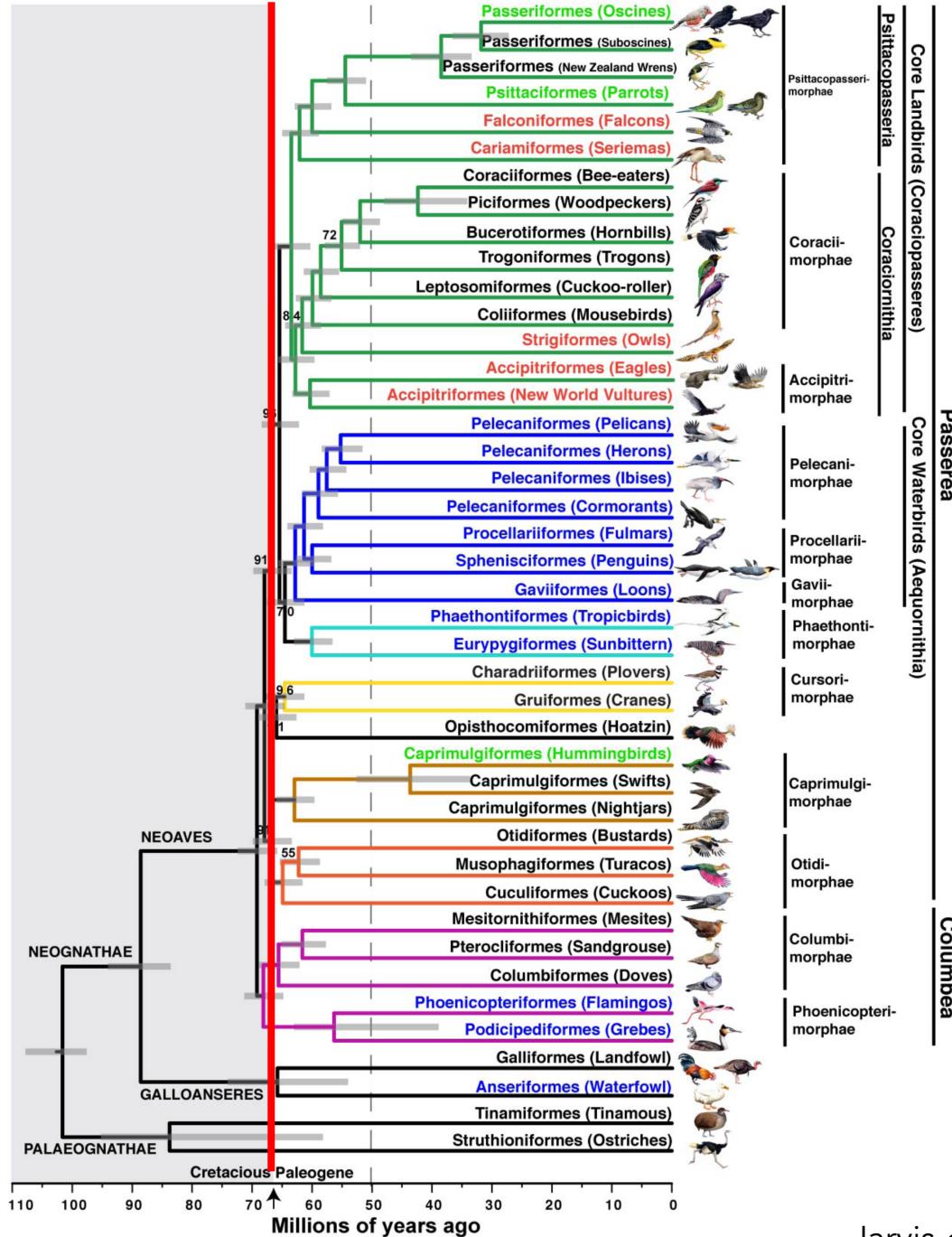


1859

Phylogenetic trees

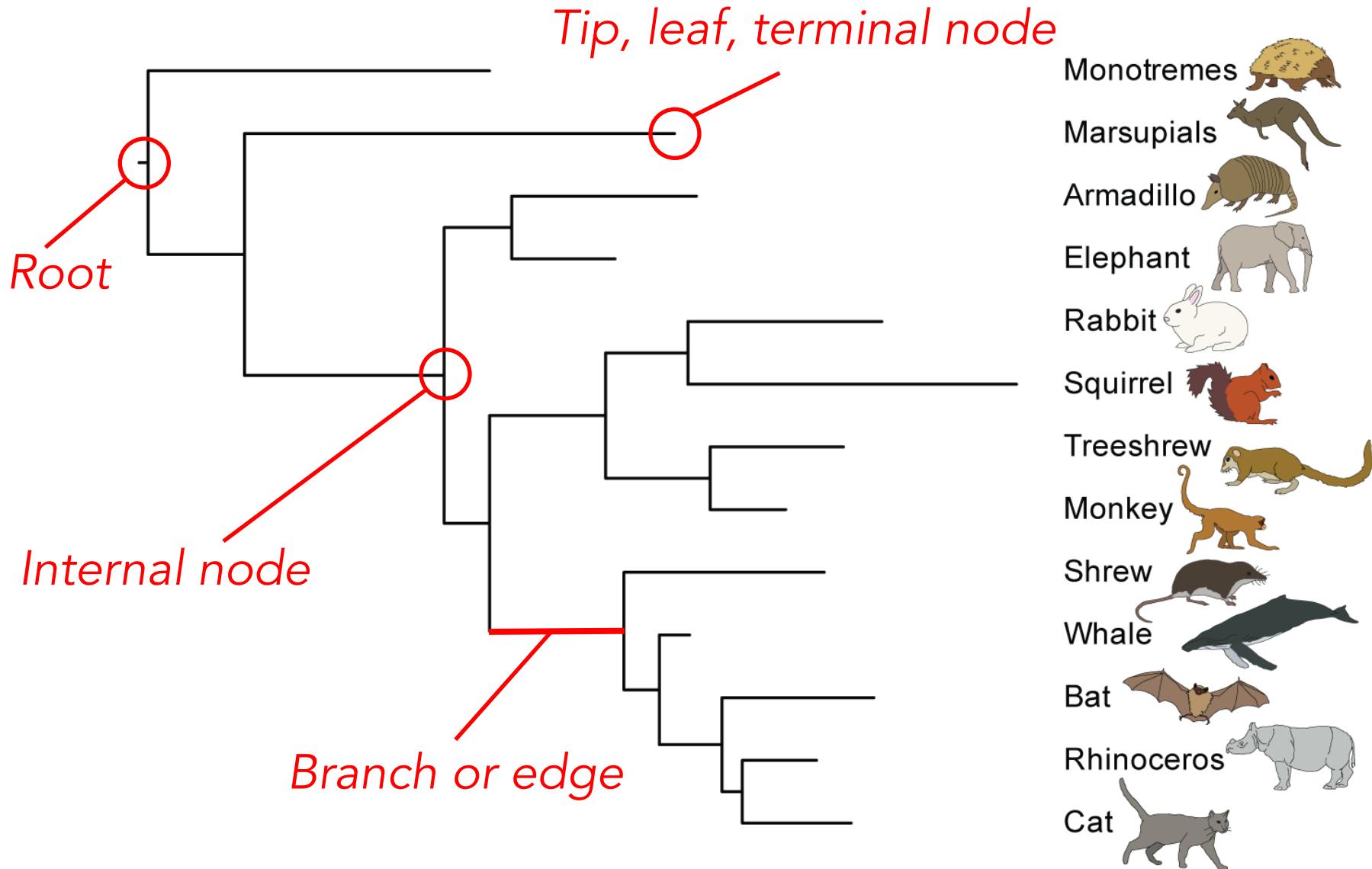
- Topology (relationships)
- Branch lengths (amount of evolutionary change or time)



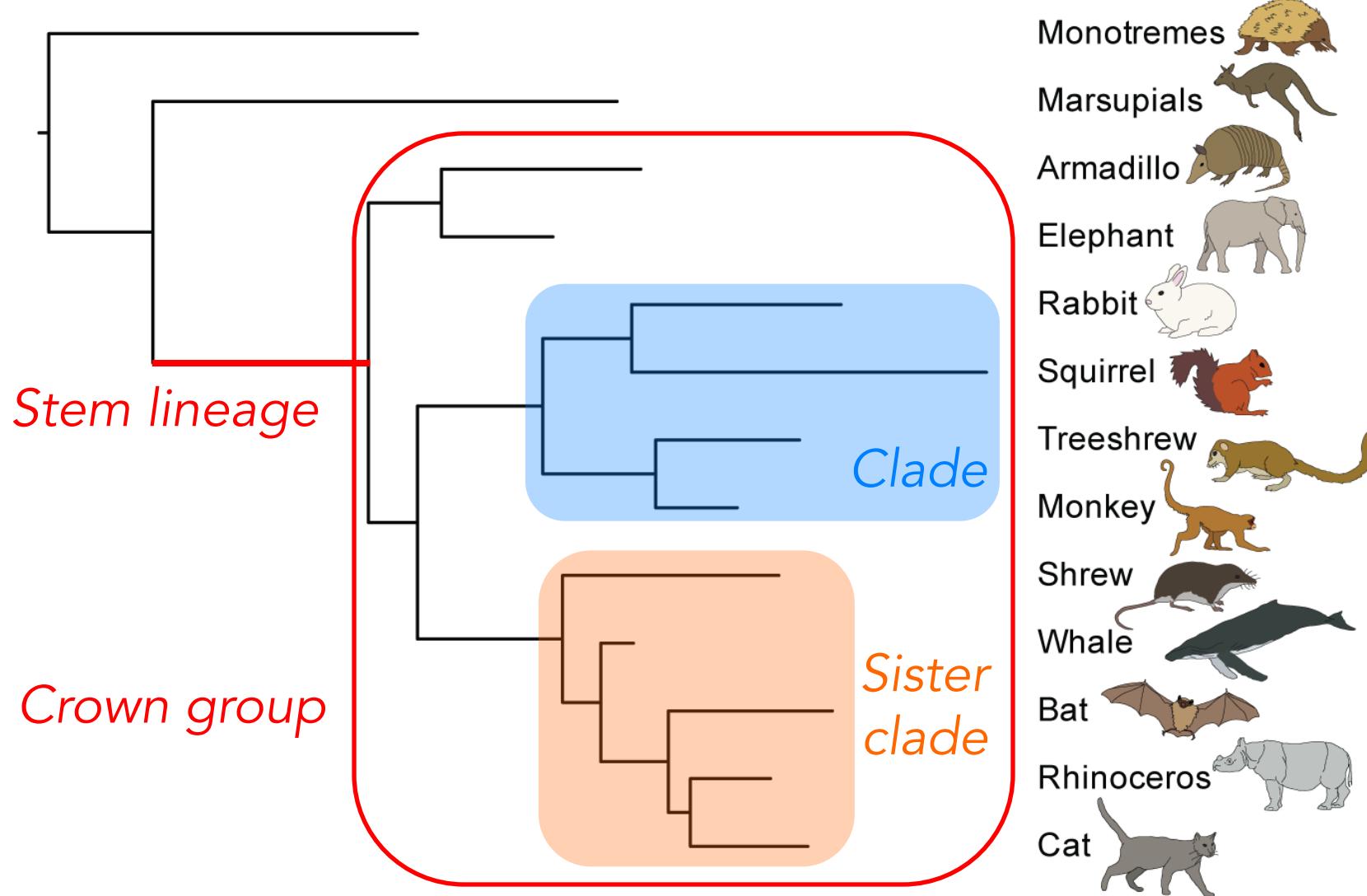


Tree thinking

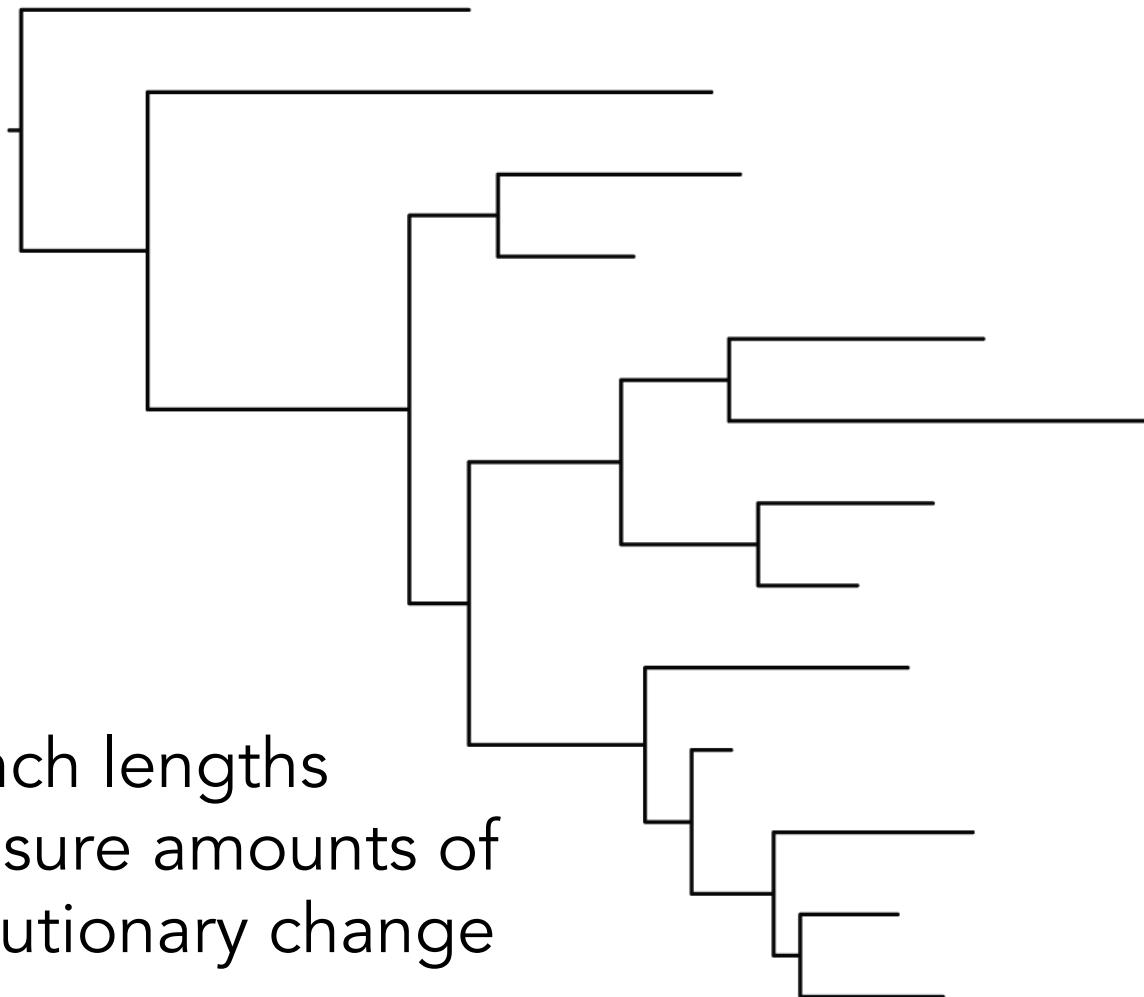
Phylogenetic trees



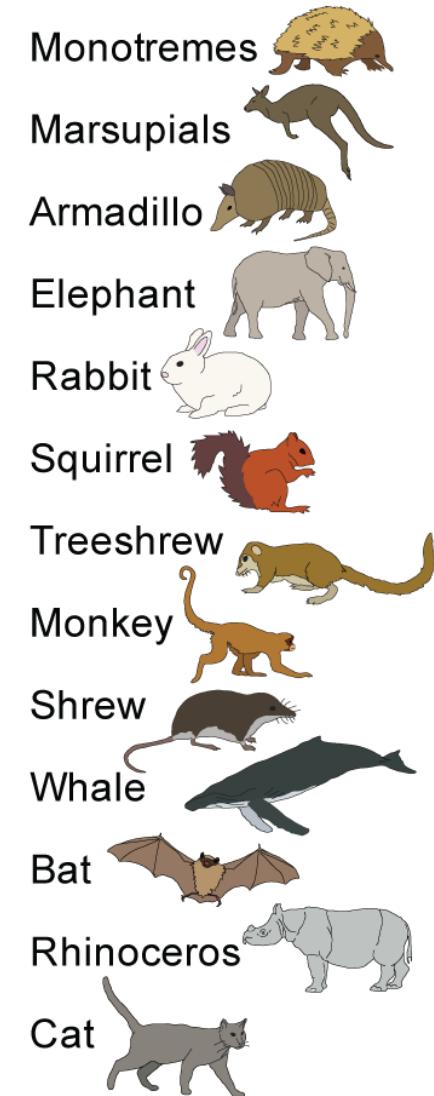
Phylogenetic trees



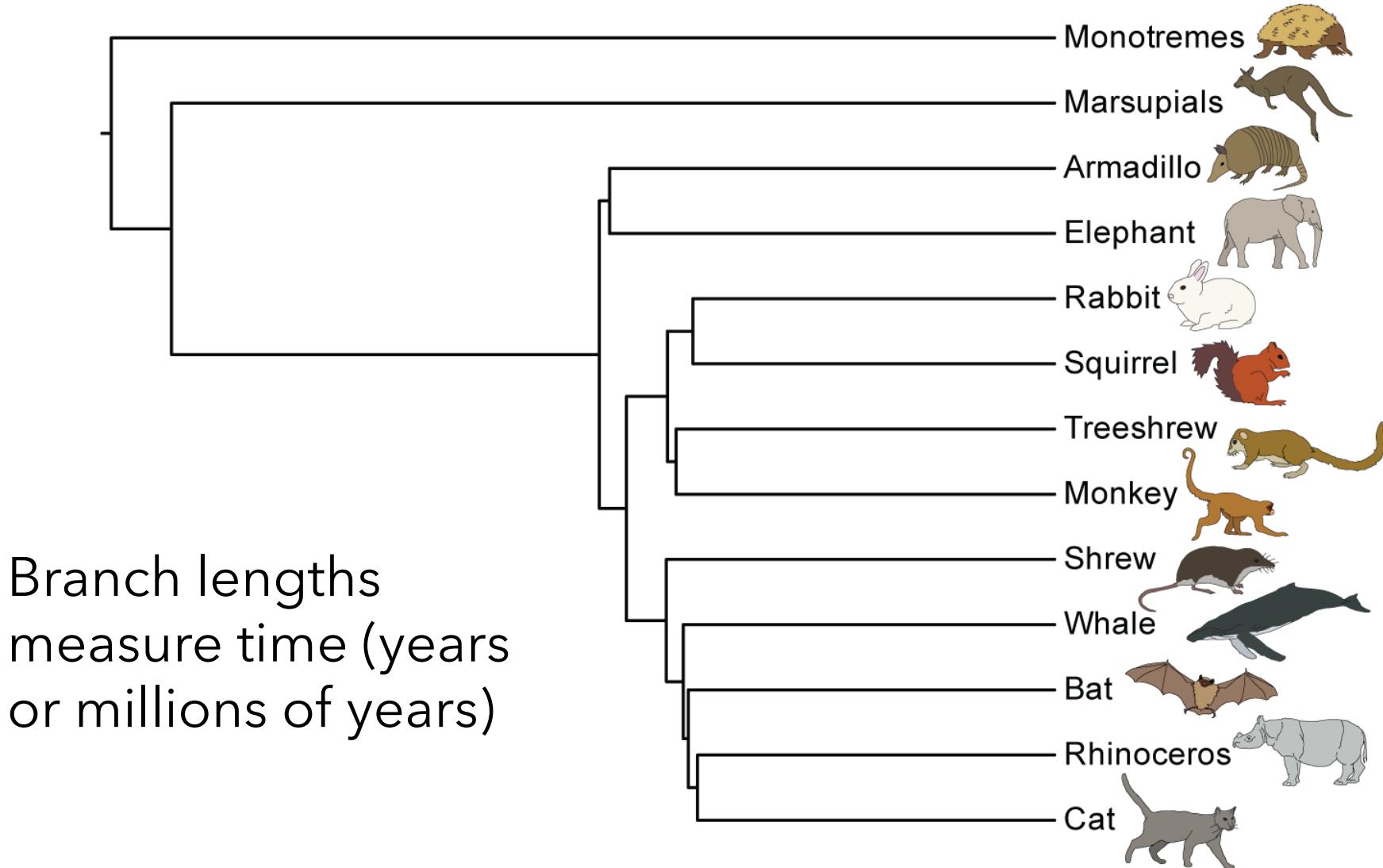
Trees: Phylogram



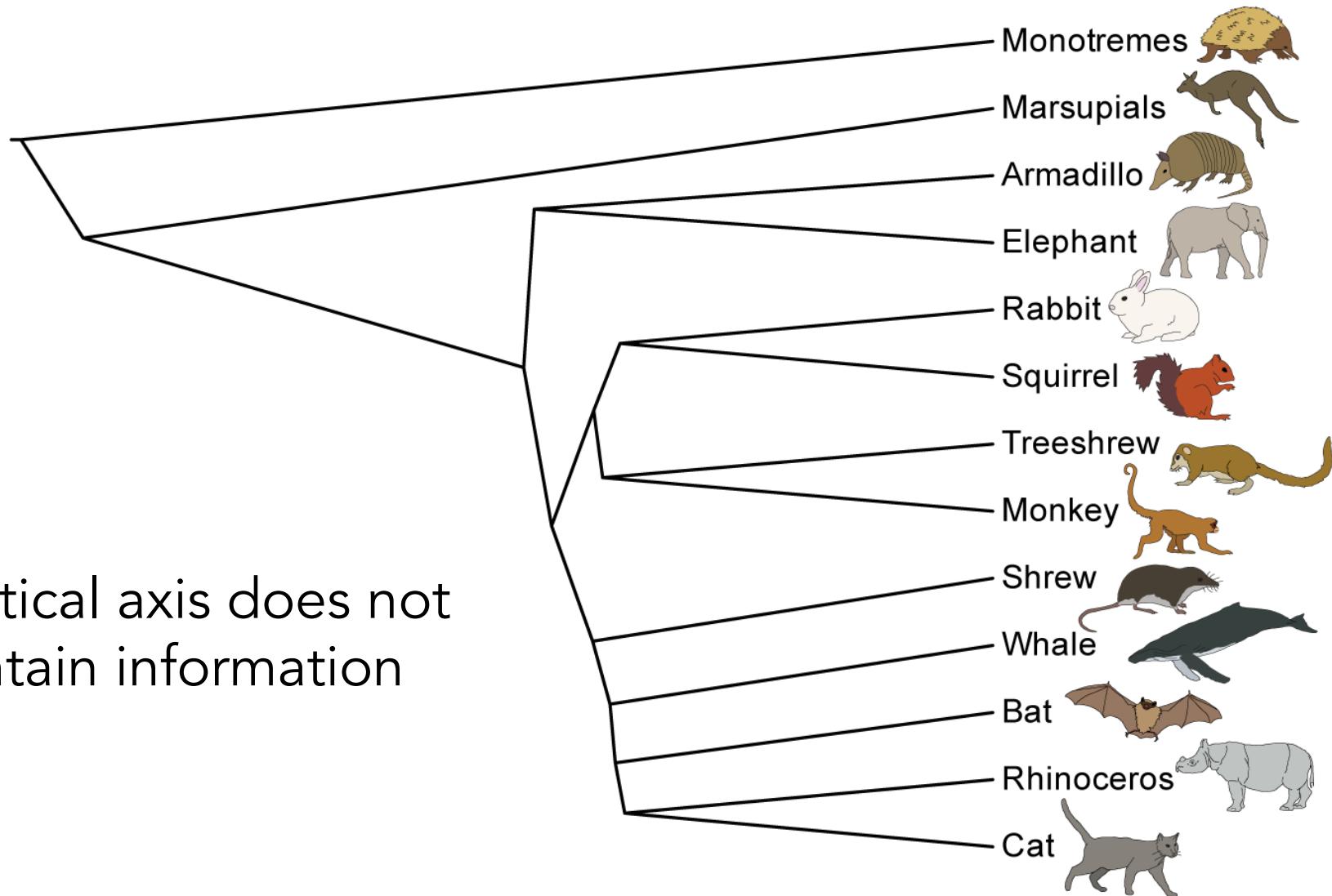
Branch lengths
measure amounts of
evolutionary change



Trees: Chronogram

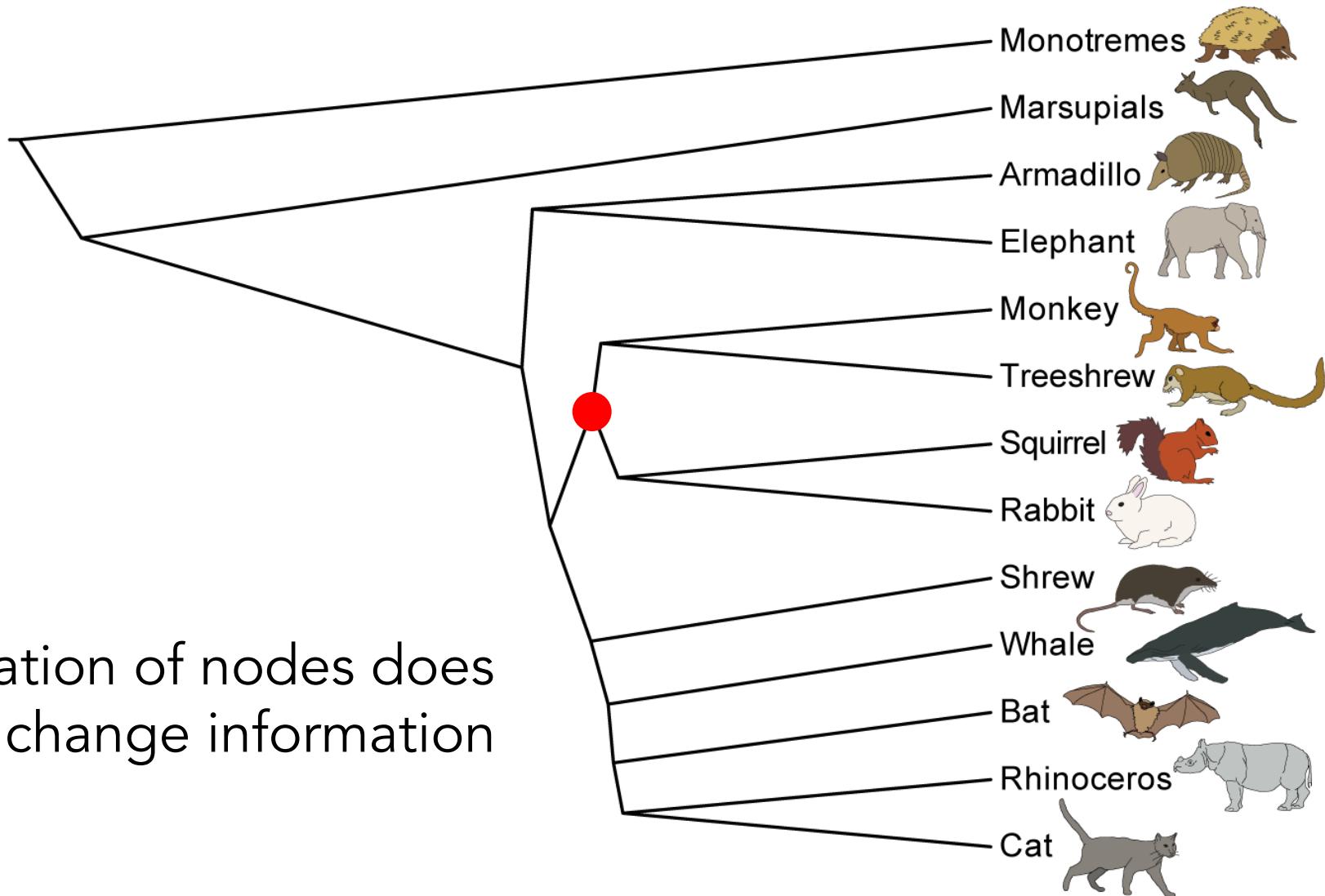


Phylogenetic trees



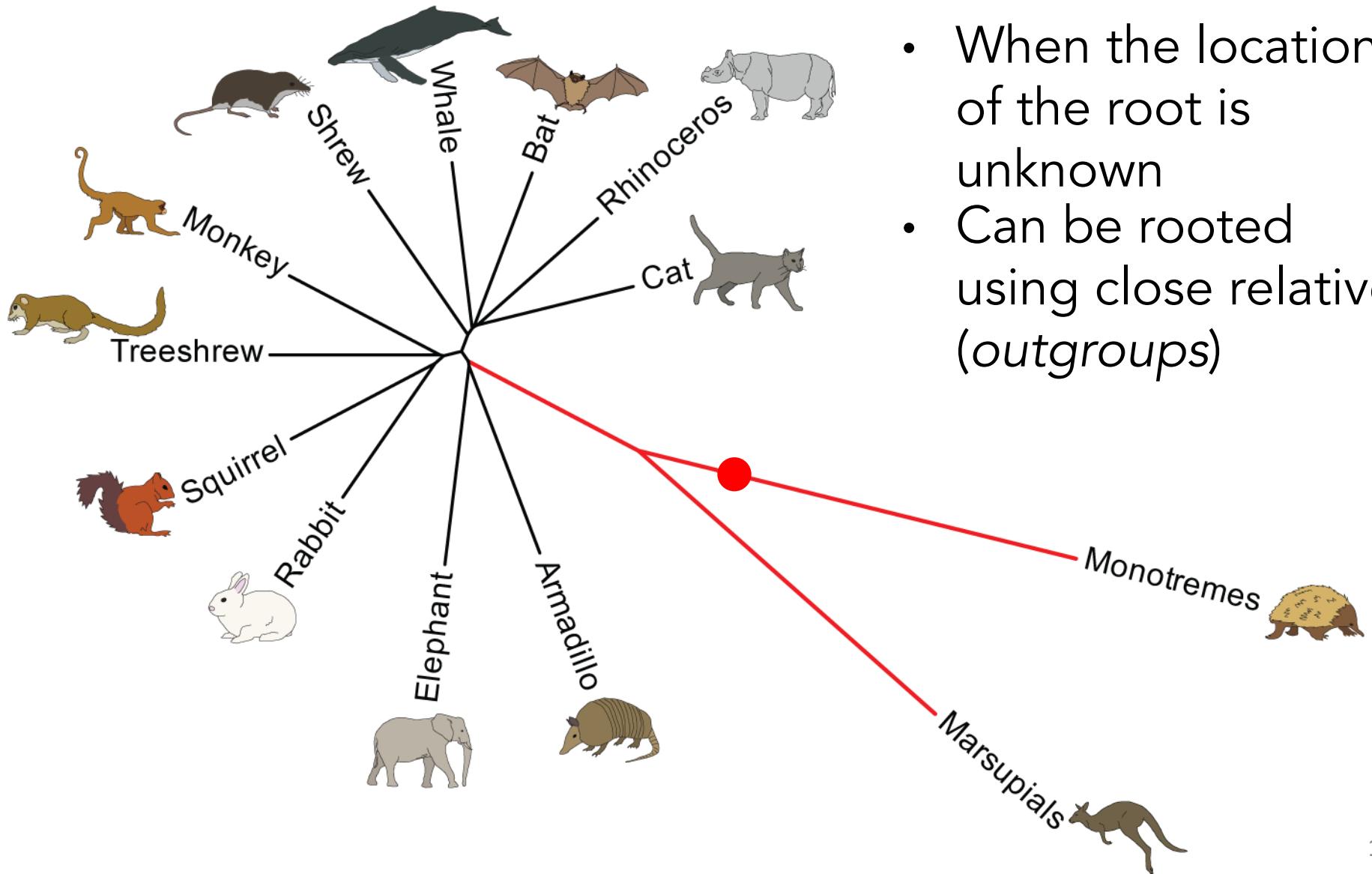
Vertical axis does not contain information

Phylogenetic trees



Rotation of nodes does not change information

Phylogenetic trees: unrooted



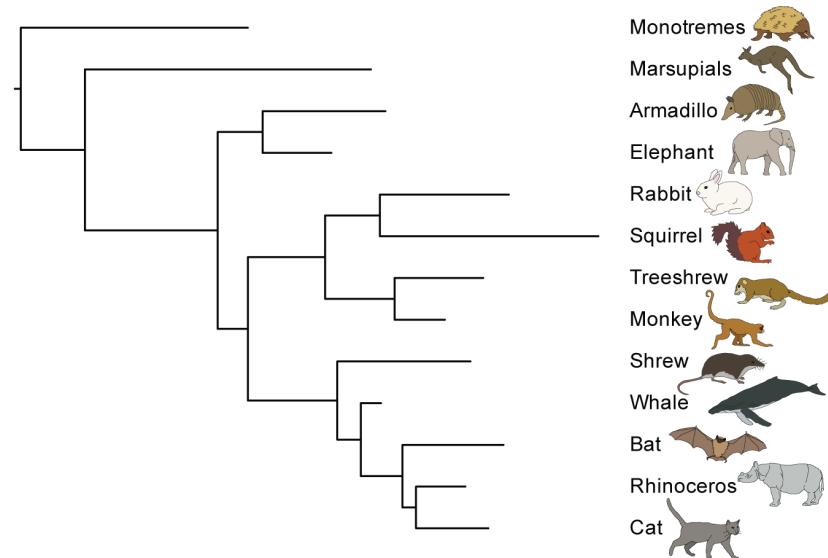
Phylogenetic trees: the Newick format

- Without branch lengths (cladogram):
 - `(Monotremes,(Marsupials,((Elephant,Armadillo),(((Squirrel,Rabbit),(Monkey,Treeshrew)),(Shrew,(Whale,(Bat,(Cat,Rhinoceros)))))));`
- With branch lengths (phylogram/chronogram):
 - `(Monotremes:12.0,(Marsupials:11.0,((Elephant:1.0,Armadillo:1.0):9.0,(((Squirrel:1.0,Rabbit:1.0):2.0,(Monkey:1.0,Treeshrew:1.0):2.0):5.0,(Shrew:4.0,(Whale:3.0,(Bat:2.0,(Cat:1.0,Rhinoceros:1.0):1.0):1.0):4.0):2.0):1.0):1.0);`

Molecular phylogenetics

Fundamental assumptions

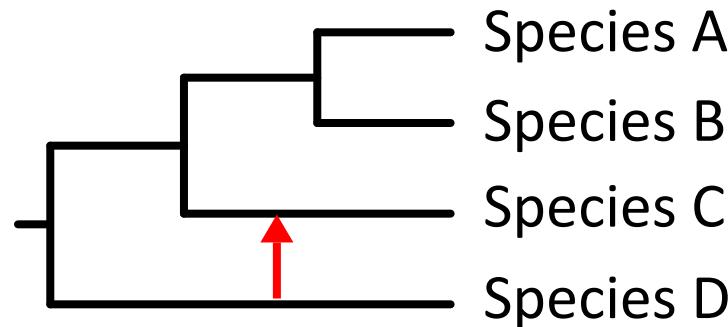
- Phylogenetic methods make several assumptions:
 - A tree represents the relationship among organisms
 - Homologous characters are compared
 - Independent characters are compared
 - Lineage evolve independently



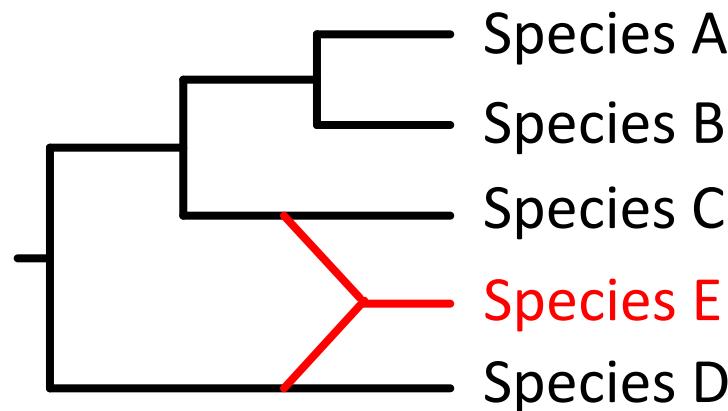
When is it incorrect
to assume a tree?

Non-treelike evolution

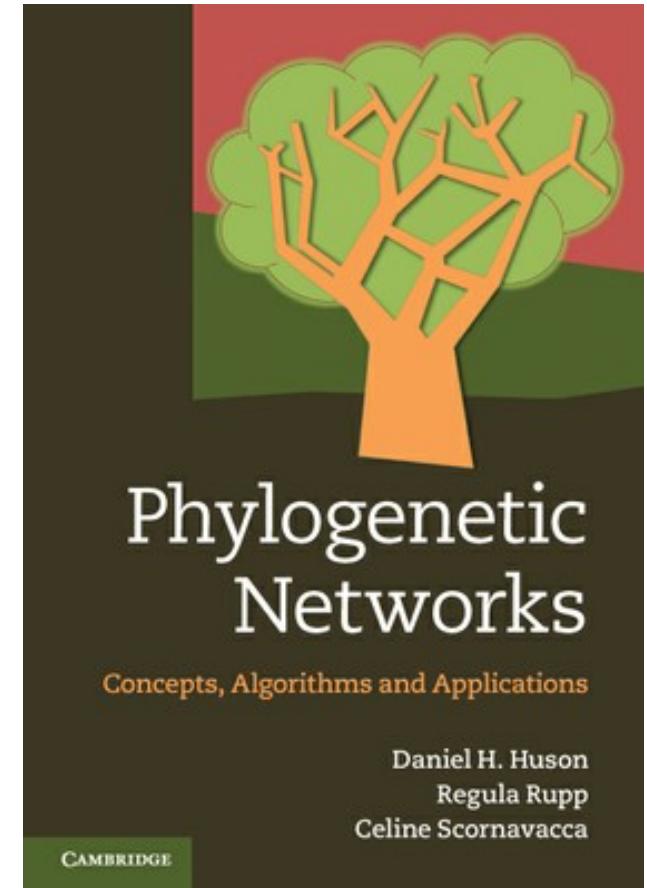
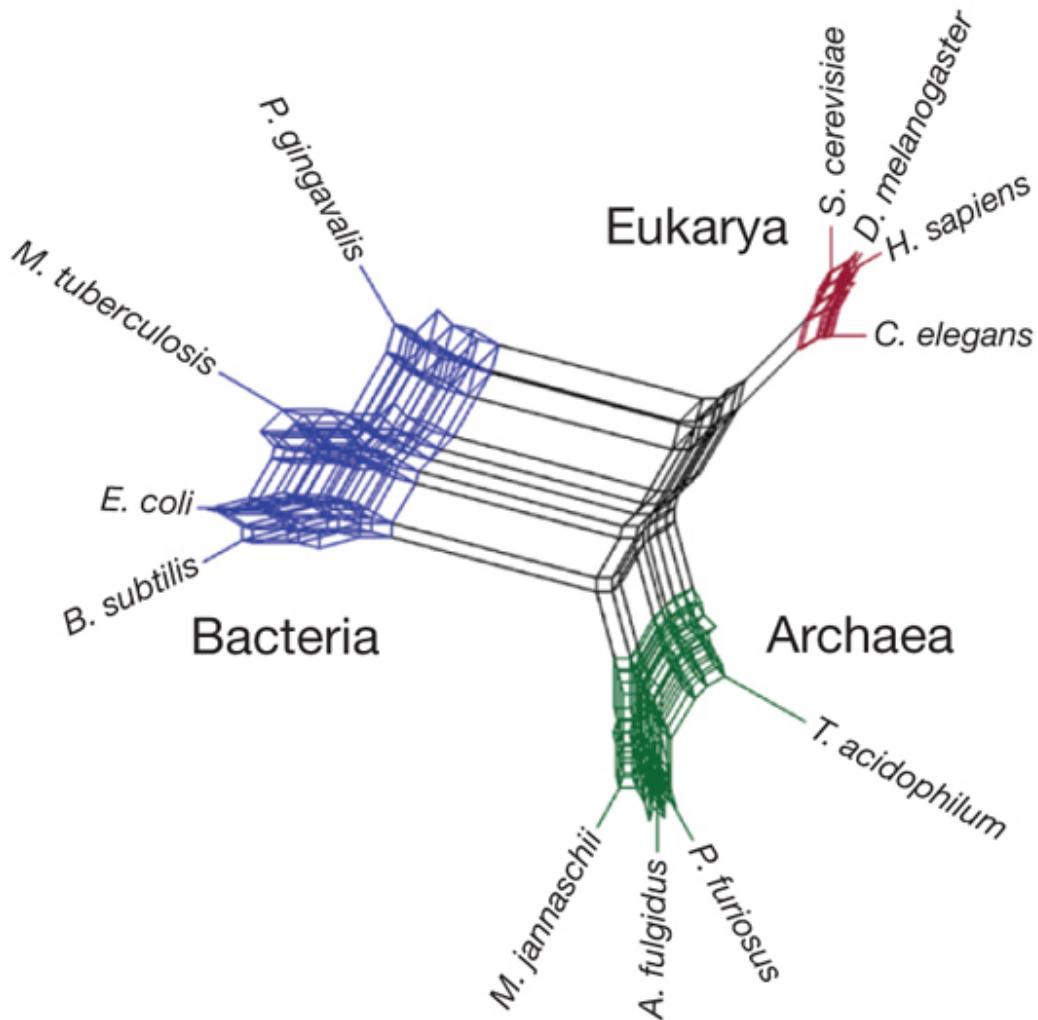
Horizontal transfer



Hybridization

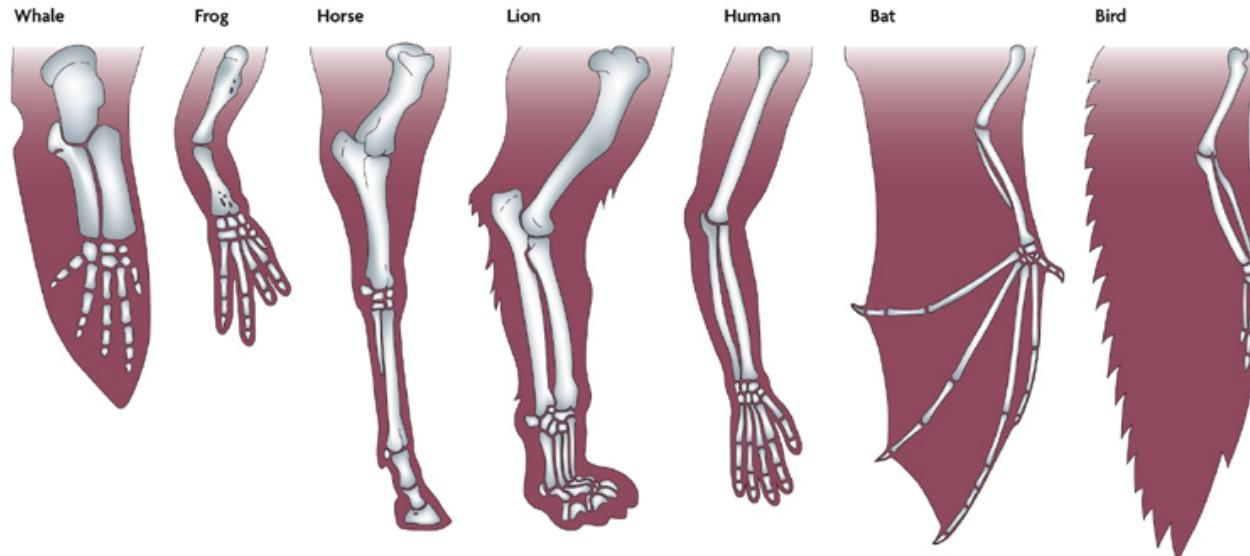


Phylogenetic networks



Fundamental assumptions

- Phylogenetic methods make several assumptions:
 - A tree represents the relationship among organisms
 - Homologous characters are compared
 - Independent characters are compared
 - Lineage evolve independently



Character homology

- Chains of nucleotides are compared
- Each nucleotide site contains a character
- But DNA sequences can vary in length

blue whale

CGTTAGTACACT

humpback whale

CGATAGTTCACT

gray whale

CGTTAGTTTACC

fin whale

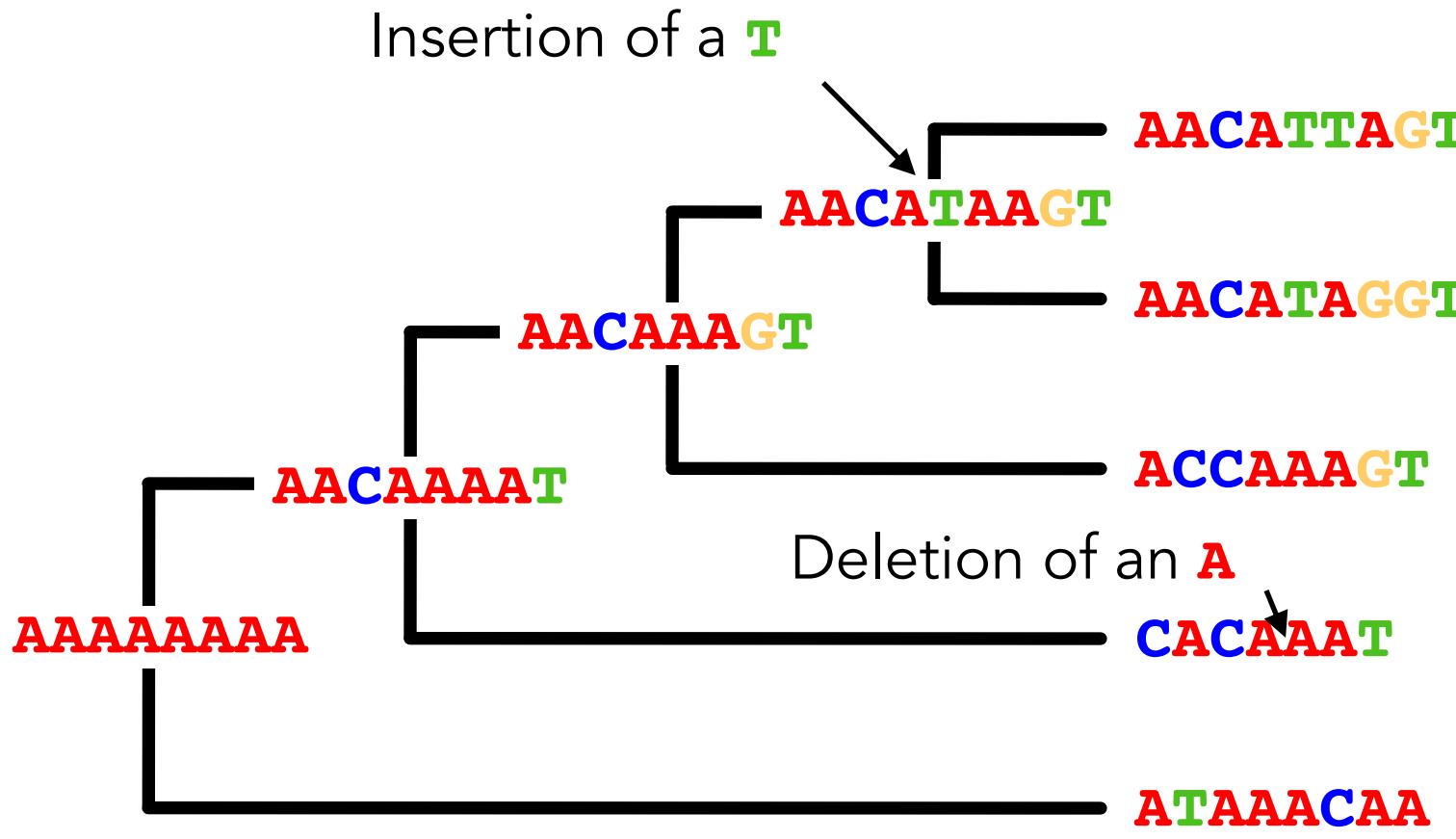
CATTGGATTACT

right whale

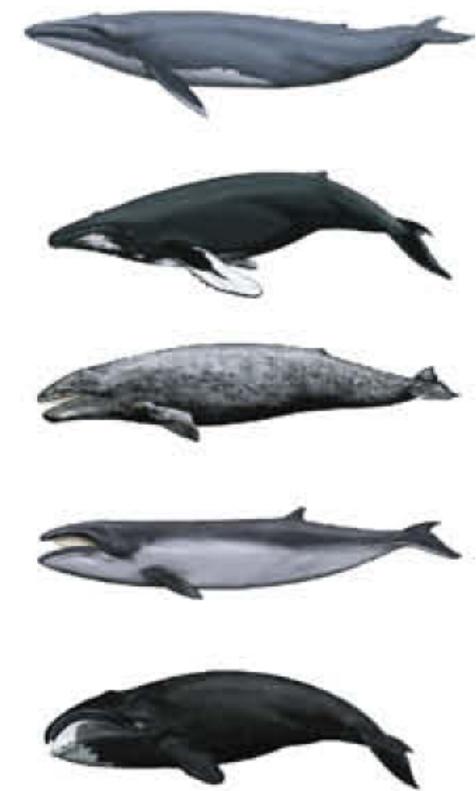
CATTGGTTTACT



Example: Whales



DNA sequence alignment



AACATTAGT

AACATAGGT

ACCAAAAGT

CACAAAT

ATAAACAA



AACATTAGT

AACATAGGT

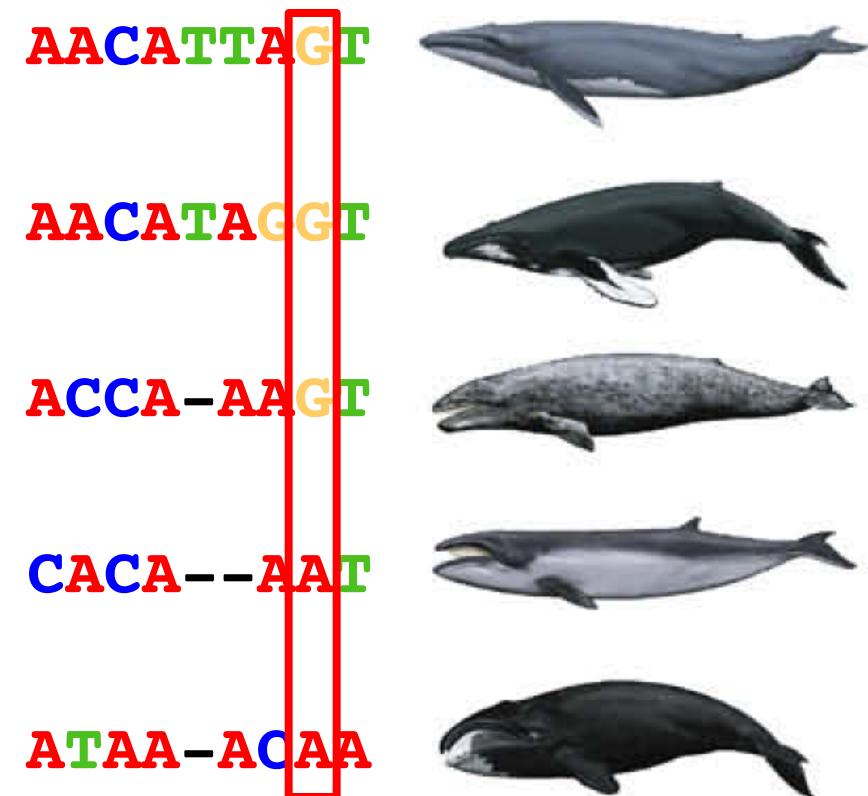
ACCA-AAGT

CACA--AAT

ATAA-ACAA

DNA sequence alignment

- Homologous sites
- Inherited from the common ancestor of all sequences in an alignment
- The aim of alignment is to maximise the number of sites for which homology can be assumed



DNA sequence alignment

- The first three sequences group together
- The last two sequences group together
- One site provides information for all phylogenetic methods

AACATTAGT
AACATAGGT
ACCA-AAGT
CACAA--AAT
ATAA-ACAA



DNA sequence alignment

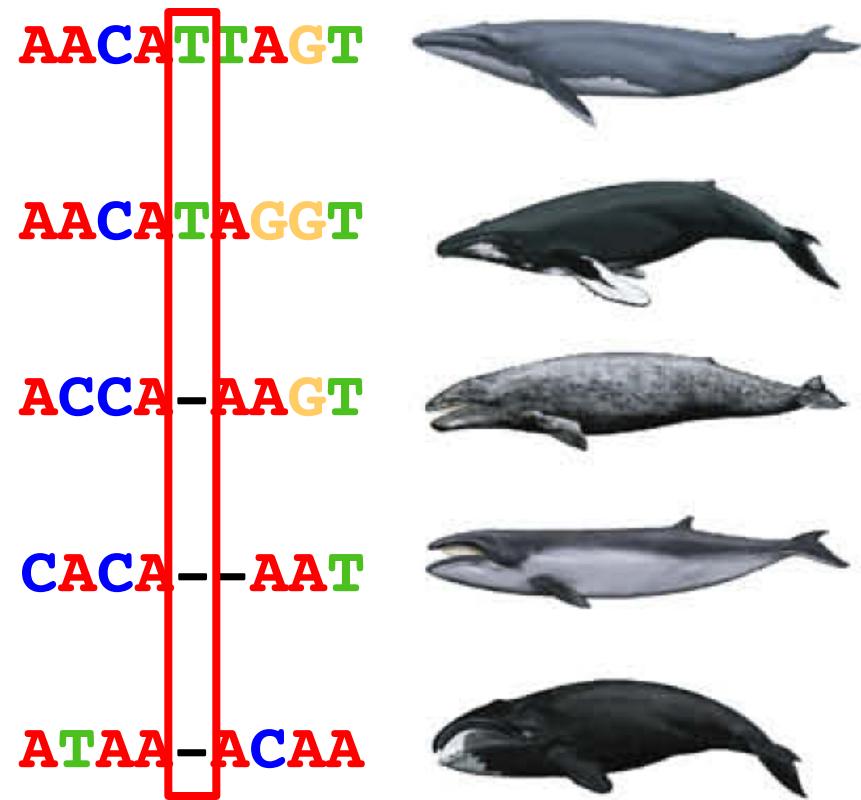
- Sequences are not grouped
 - Does not inform parsimony
- But contains information about the amount of evolutionary change
 - Useful for other phylogenetic methods

AACATTAGT
AACATAAGT
ACCA-AAGT
CACA--AAT
ATAA-ACAA



DNA sequence alignment

- Indel – insertion or deletion
- Potentially informative
- Most phylogenetic methods do not use this information



A practical approach

Sequences are aligned using automated methods

CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice

Julie D.Thompson, Desmond G.Higgins⁺ and Toby J.Gibson*

Software

Open Access

MUSCLE: a multiple sequence alignment method with reduced time and space complexity
Robert C Edgar*

MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform

Kazutaka Katoh, Kazuharu Misawa¹, Kei-ichi Kuma and Takashi Miyata*

A practical approach

Sequences are aligned using automated methods



Alignments are adjusted "by eye"

CTATGTGGCACCCAGCCCATGCA--AGC

ATATGTGGCA-----CCCAGGCA--AG-

ATATGTGGCACCCAGCCCATGCATT---

A practical approach

Sequences are aligned using automated methods



Alignments are adjusted "by eye"



Sites with uncertain homology are excluded

CTATGTGGCACCCAGCCCATGCA -- AGC

ATATGTGGCA ----- CCCAGGCA -- AG - ?

ATATGTGGCACCCAGCCCATGCA TTT --

Useful reading

