

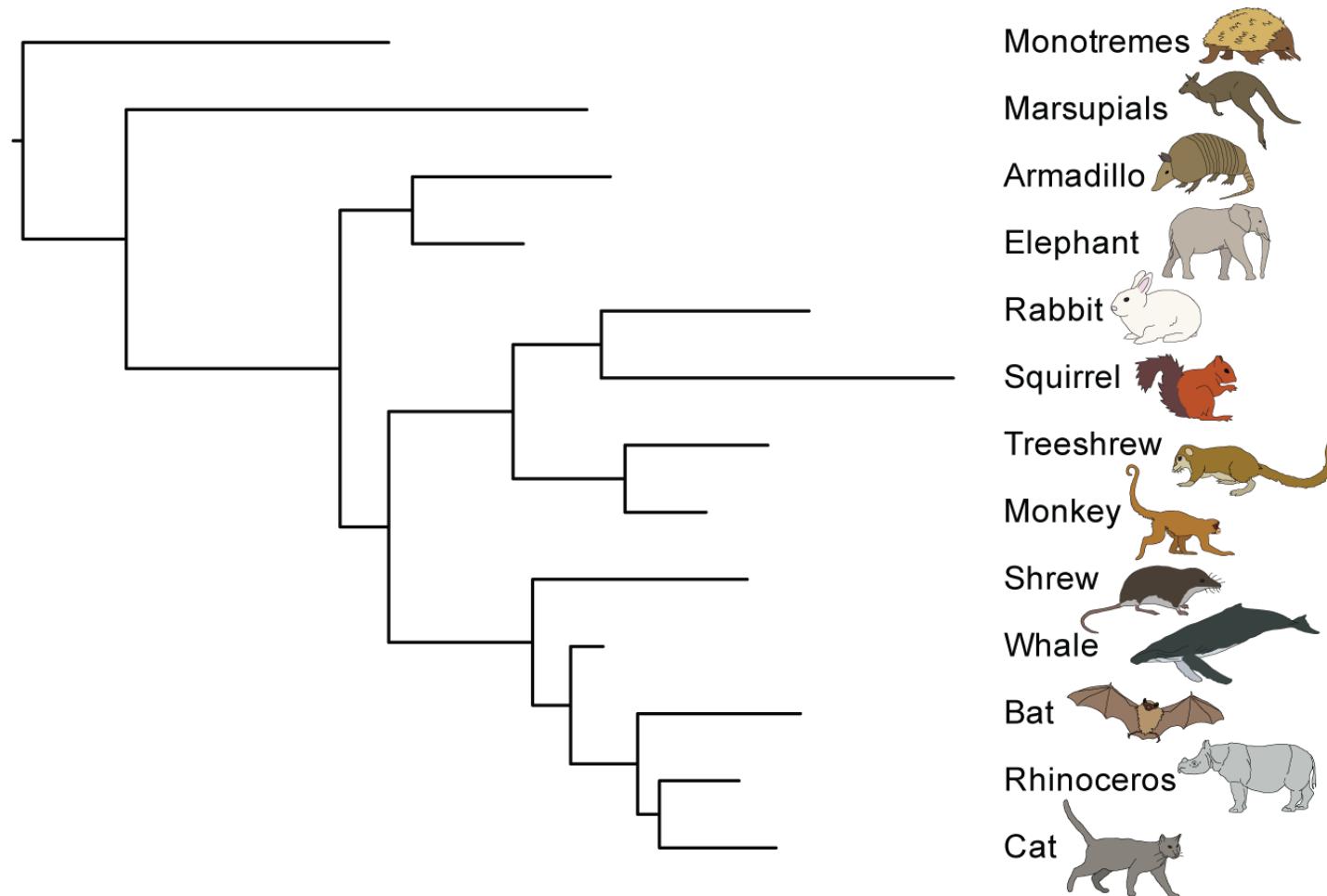
---

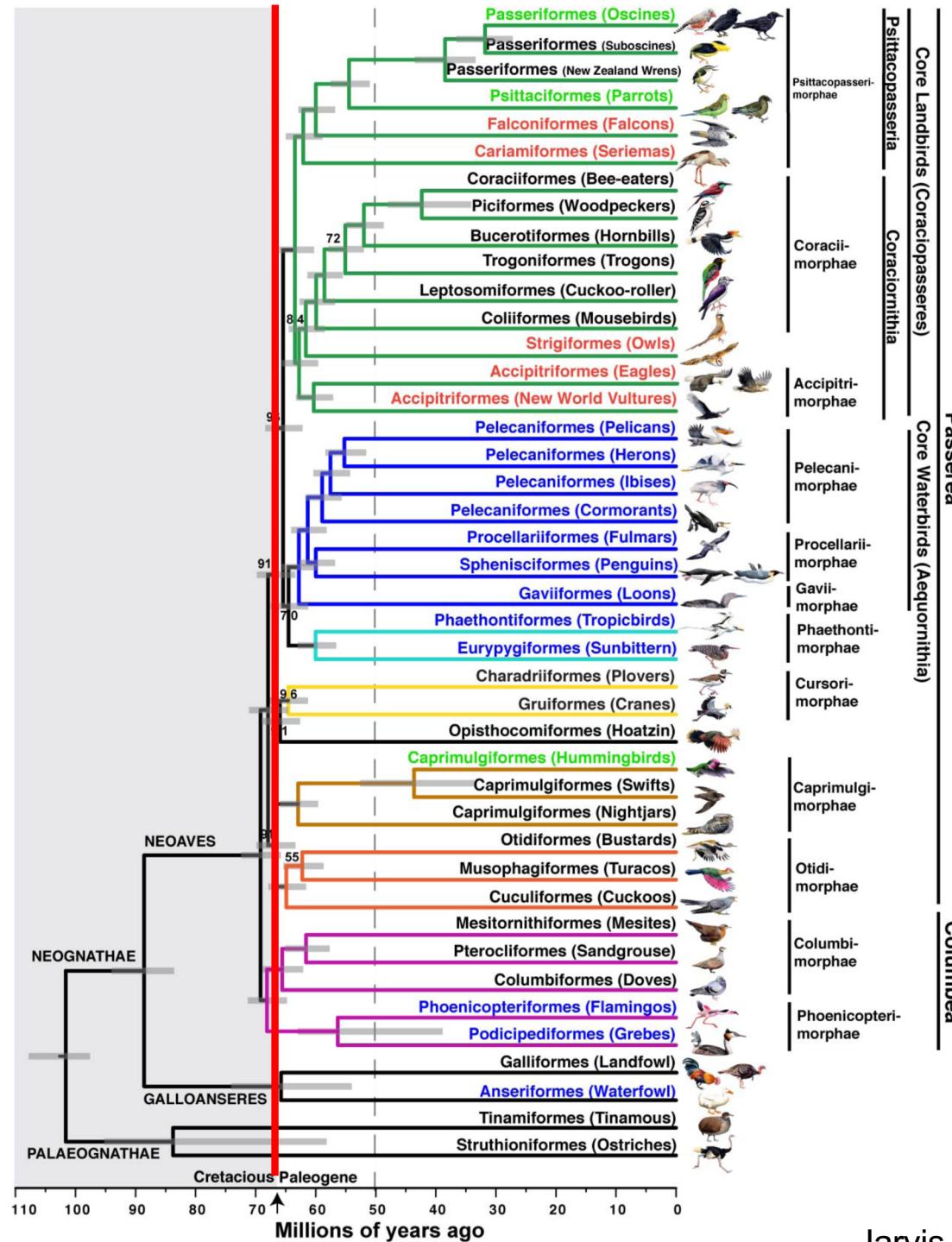
# Molecular Phylogenetics

---

# Phylogenetic trees

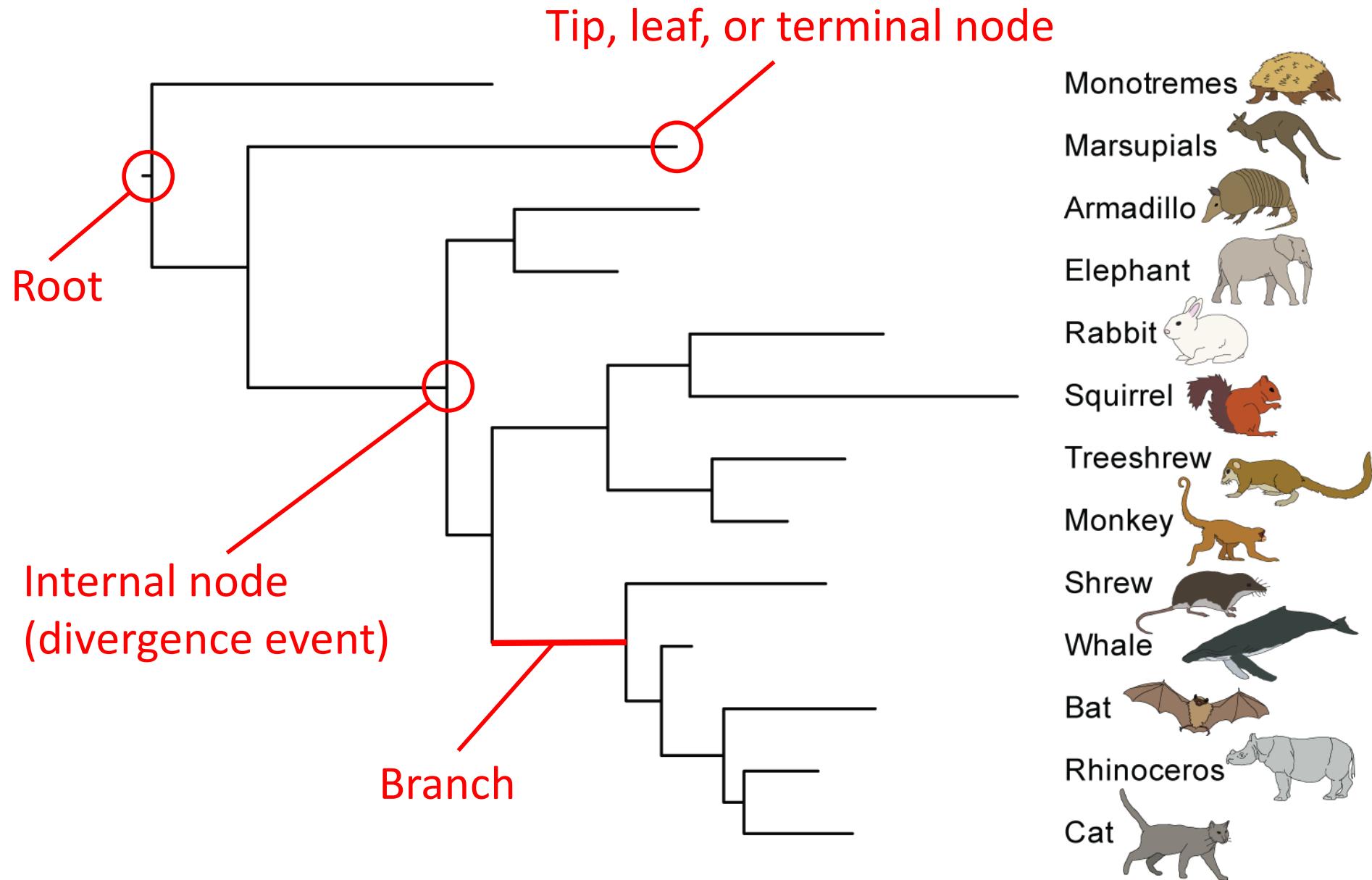
- Topology (relationships)
- Branch lengths (amount of evolutionary change or time)



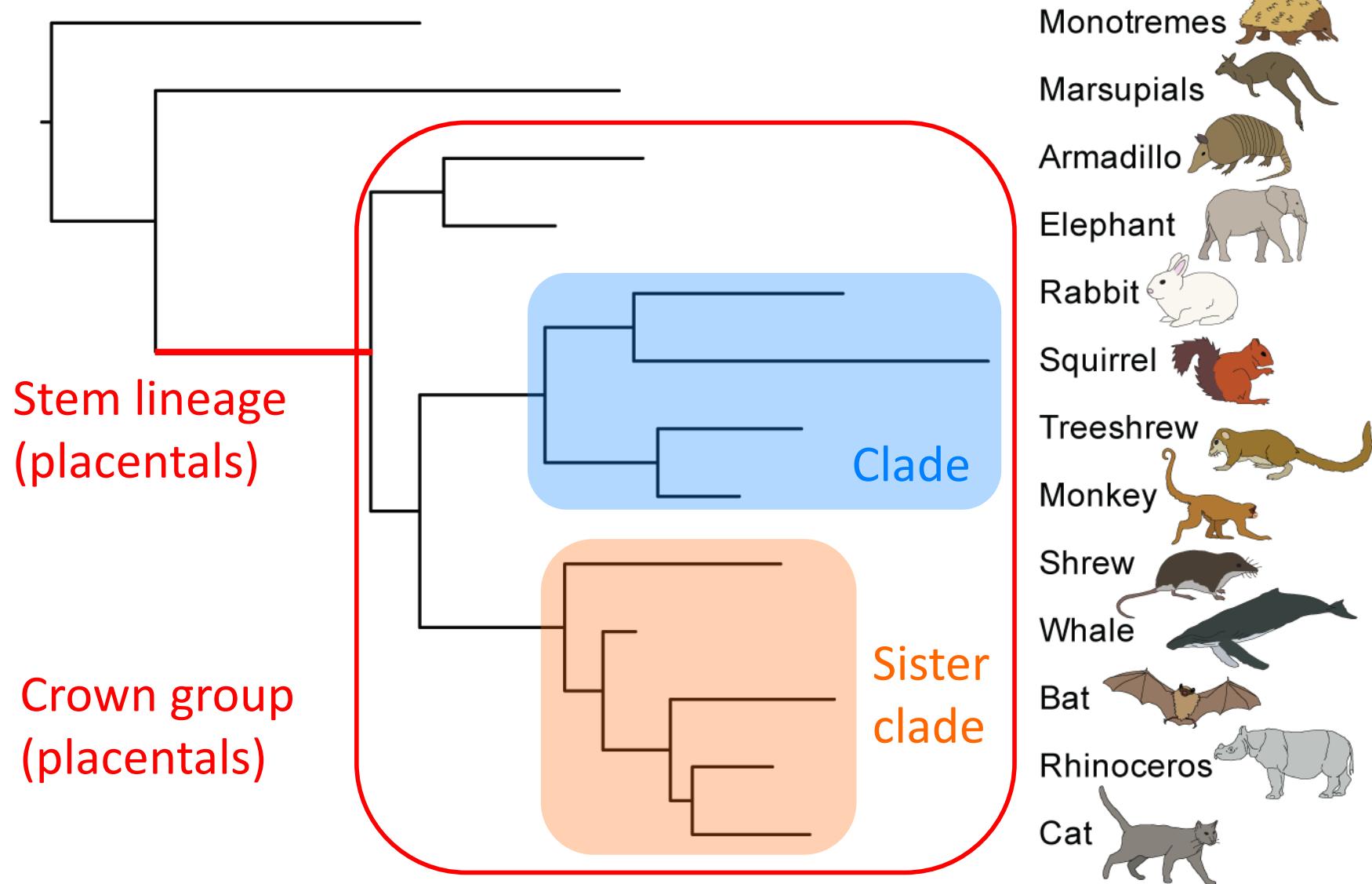


# Tree Thinking

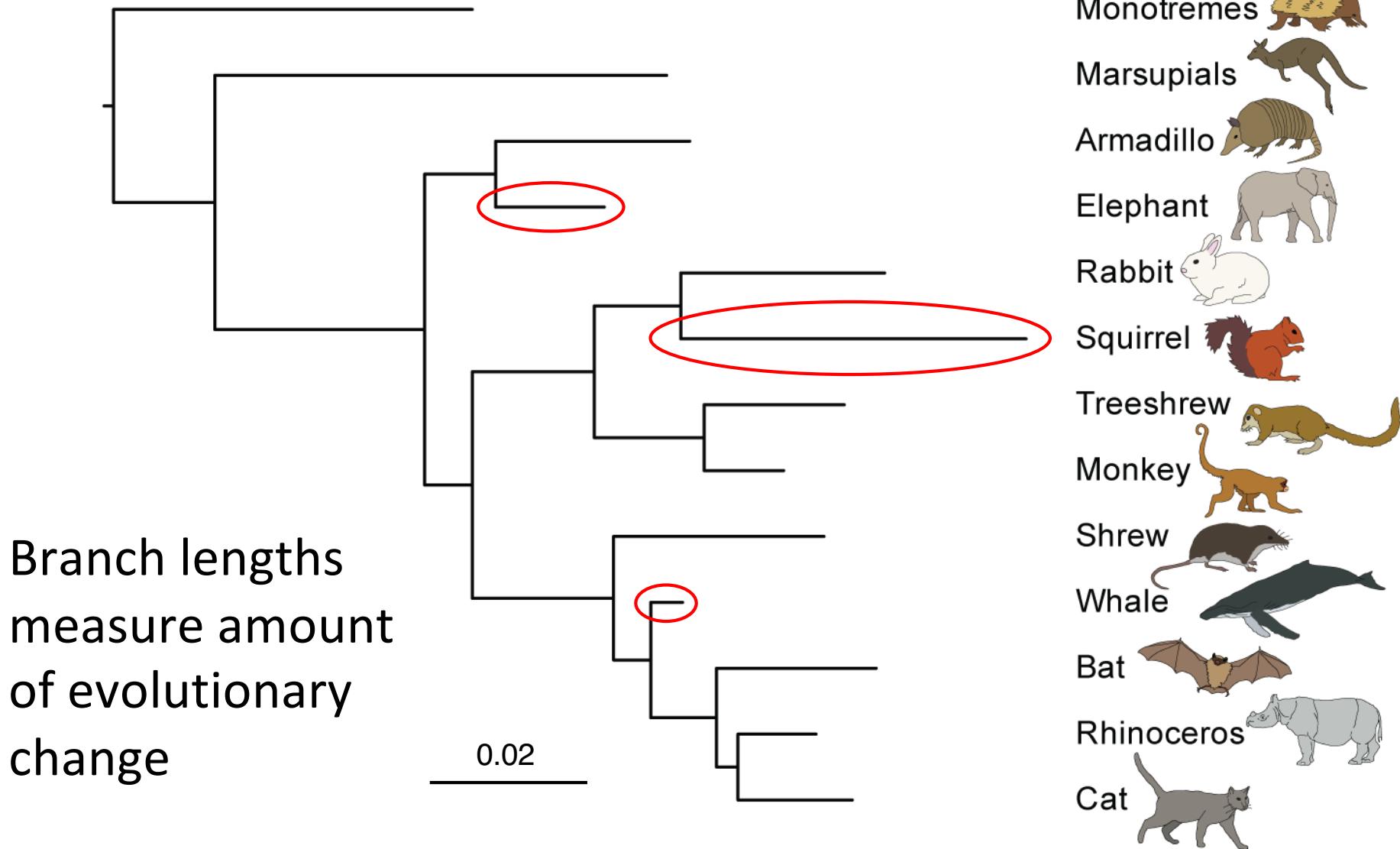
# Phylogenetic trees



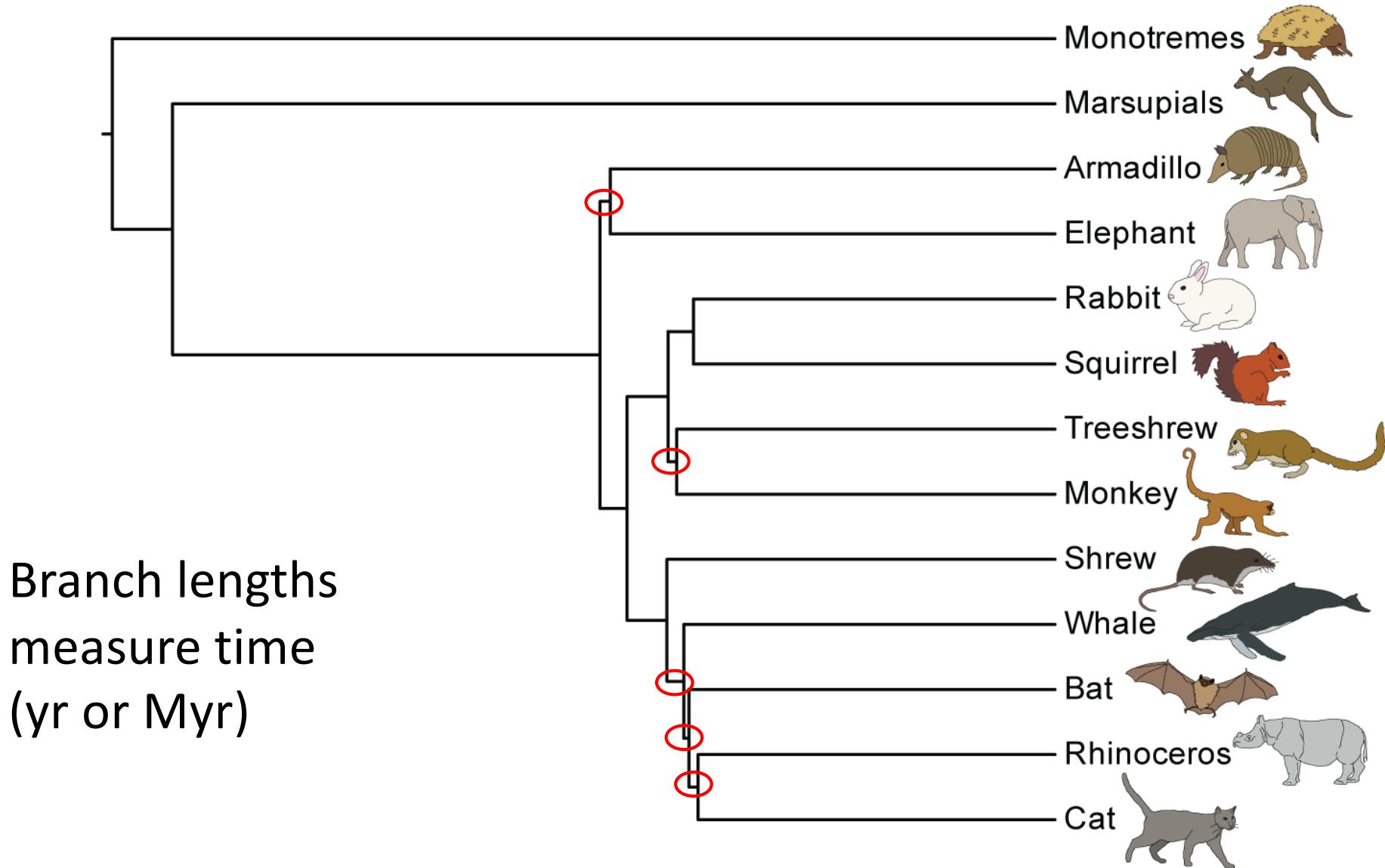
# Phylogenetic trees



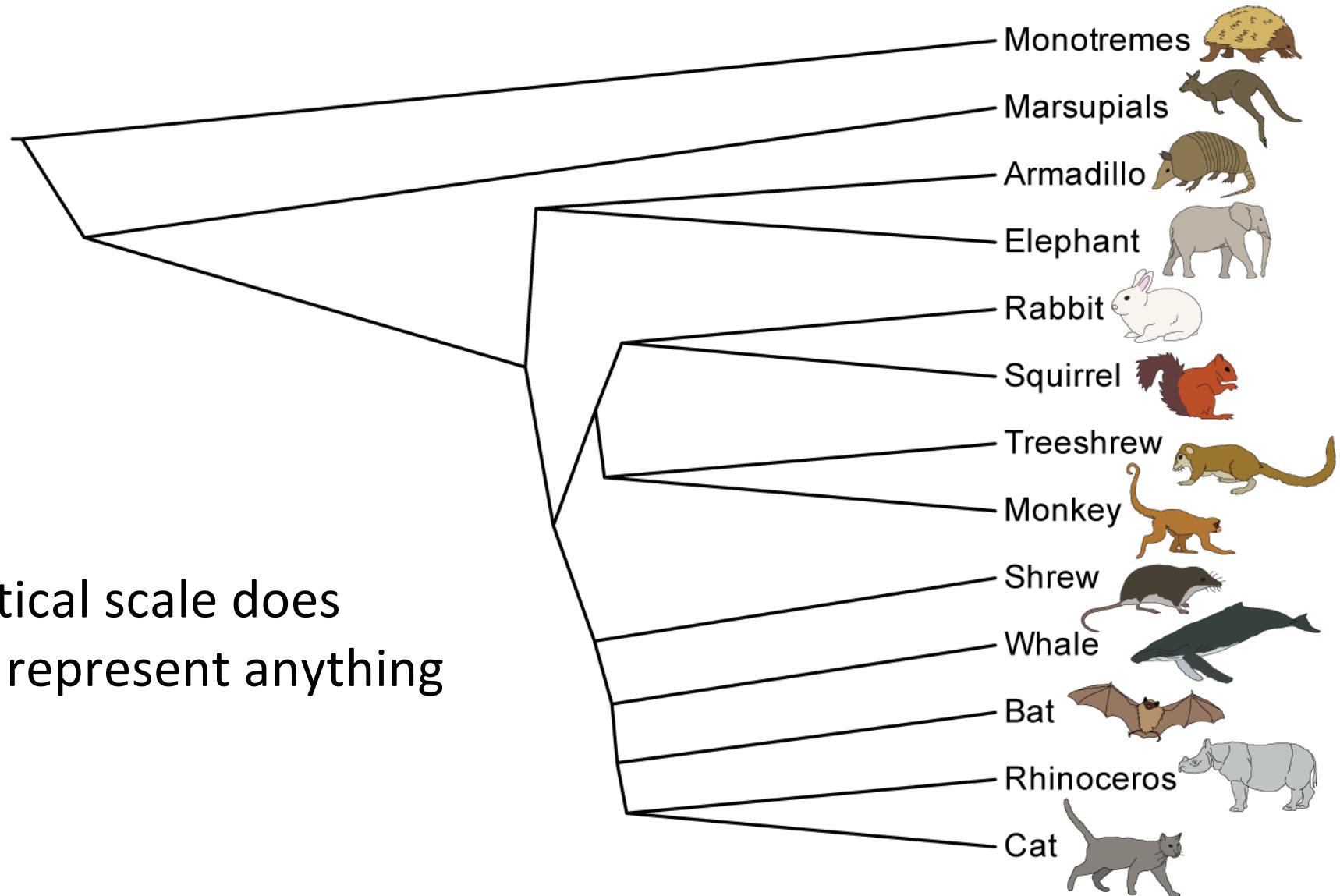
# Trees: Phylogram



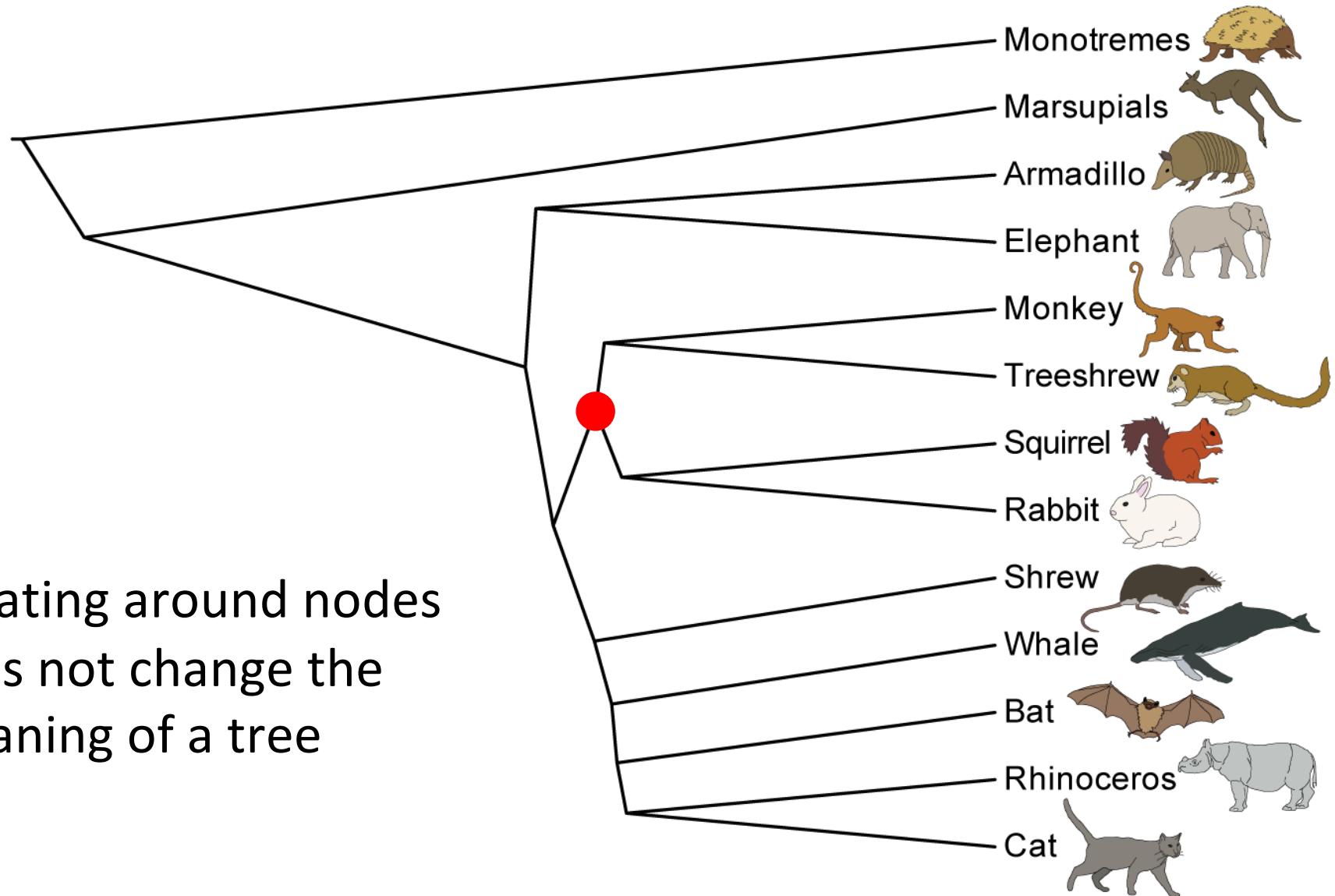
# Trees: Chronogram or time-tree



# Phylogenetic trees

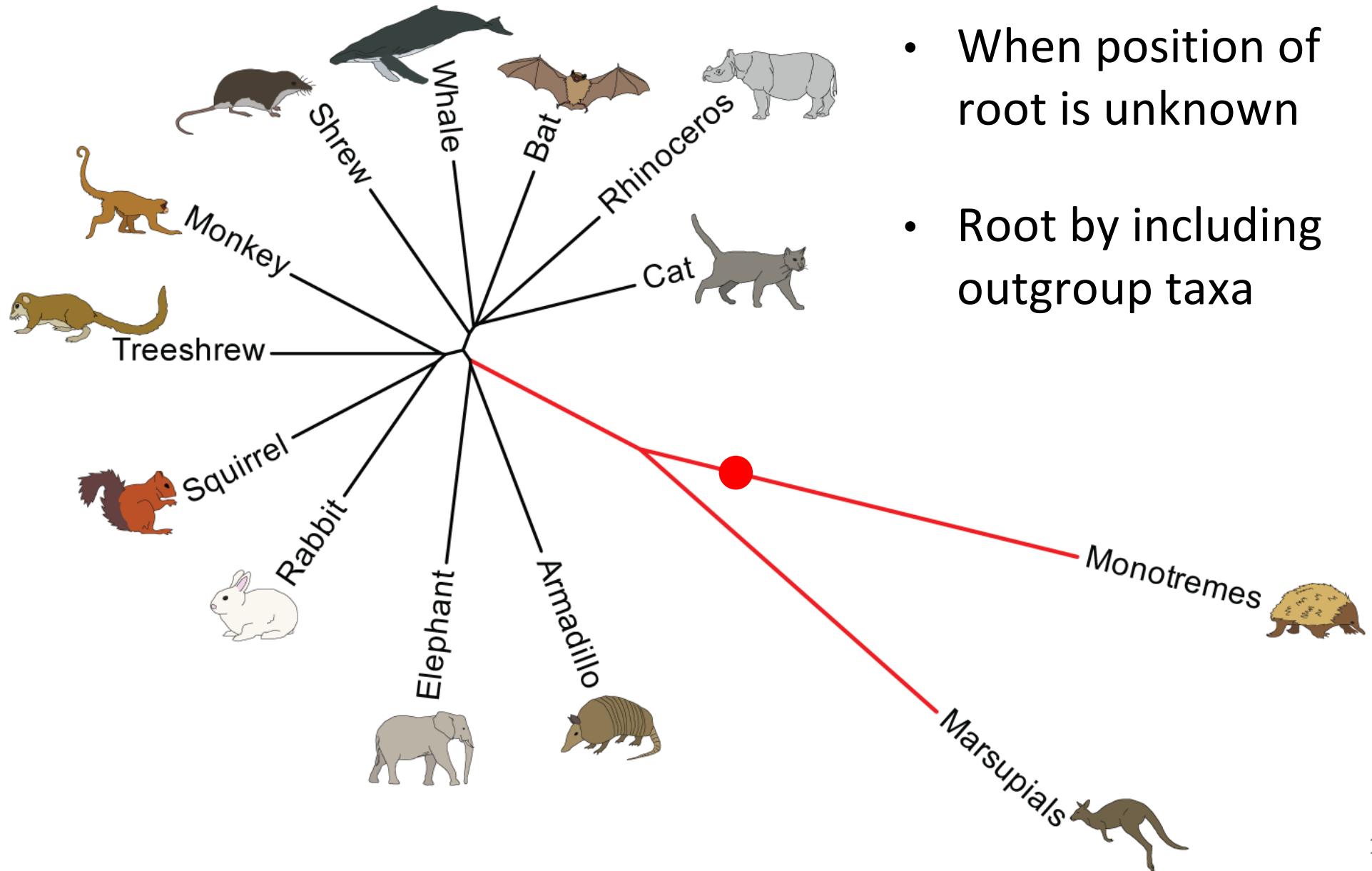


# Phylogenetic trees



Rotating around nodes  
does not change the  
meaning of a tree

# Phylogenetic trees: Unrooted



- When position of root is unknown
- Root by including outgroup taxa

# Phylogenetic trees: 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):1.0):4.0):2.0):1.0):1.0);

# Phylogenetic Analysis

# Phylogenetic analysis

---

- Sometimes we know the phylogeny
  - Viral transmission histories
  - Pedigrees (humans, domesticated animals, lab organisms, etc.)
- Usually we do not know the phylogeny but we can estimate it
  - Morphological data
  - Molecular data

# Data types

---

- **Sequence data**
  - Nucleotides
  - Amino acids
- **Binary data** (presence/absence of genomic features)
- **Microsatellites** (repeat numbers)
- **Single-nucleotide polymorphisms** (SNPs)
- **Reduced-representation sequences**

# Sequence data

---

- **Coding sequences**
  - Ribosomal RNA
  - Protein-coding genes
- **Non-coding sequences**
  - Intergenic sites
  - Introns
- **Amino acid sequences**



# Sequence data

protein-coding gene

		M	R	E	P	Y	S	R
brown bear	CGTTAG--CATGAGGGAACCTACTCTAGG							
		M	R	E	P	Y	S	R
cave bear	CGATAG-TCATGAGGGAACCTACTCTAGG							
		M	R	E	S	Y	P	R
black bear	CGTTAG-TTATGAGGGAATCCTACCCCTAGG							
		M	R	H	S	-	S	R
giant panda	CA--GGTTIATGAGGCATTCC---TCTAGG							

# Transcriptomes and exon capture

---

- Large panels of protein-coding loci
- Sequences are easier to align
- Good for inferring deep relationships
- Issues to consider:
  - **Variability**  
Might not be much variation at the population level
  - **Selection**  
Differences in selection will lead to rate differences across exons

# Whole genome sequencing

- Typically NOT (yet) the entire genome
- Many challenges: Jarvis et al. (2014)  
*Science*
  - >400 years of computing using a single processor
- **Issues to consider**
  - Single-copy genes
  - Selectively neutral
  - Unlinked loci



# Popular phylogenetic methods

Maximum parsimony

Distance-based methods

Maximum likelihood

Bayesian inference



# Australian marsupials

- One of the 3 major groups of mammals
- A distinctive component of the Australian fauna
- About 334 species in 22 families

18 of 22 families

45 species

1550 exons

