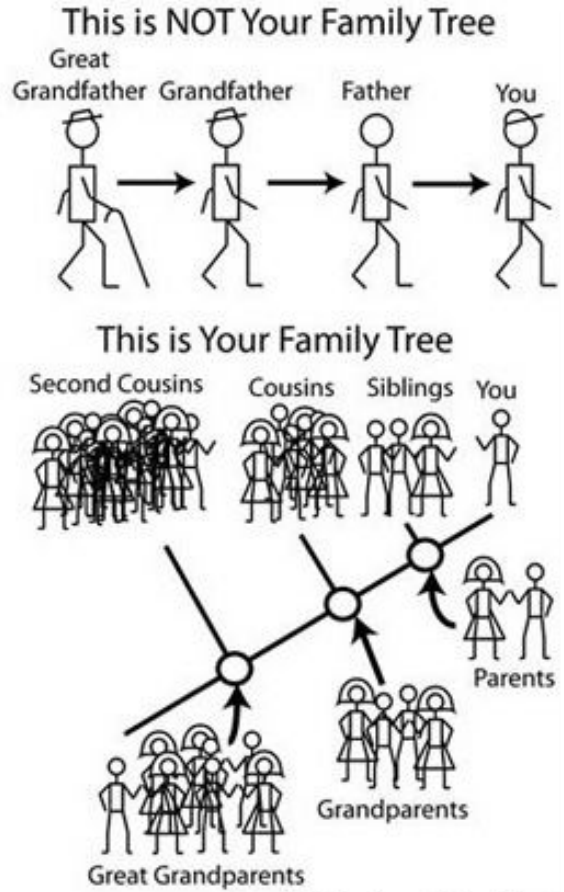
Phylogenetics

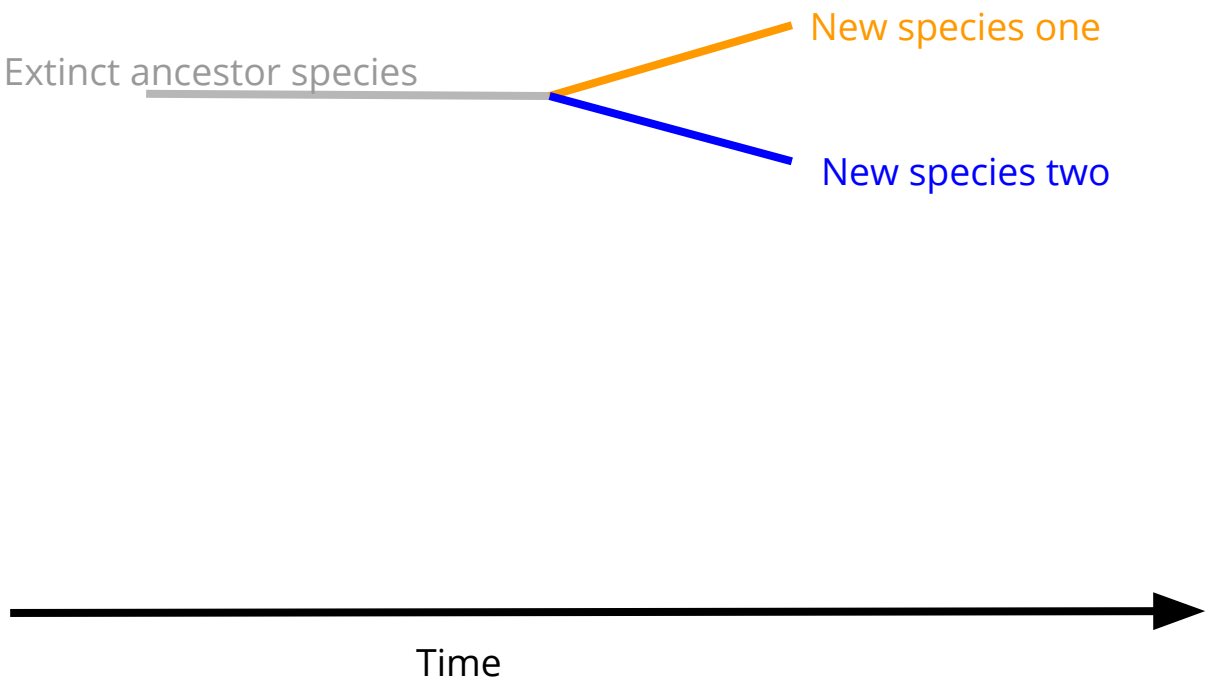
Introduction to Evolution and Scientific Inquiry

Dr. Stephanie J. Spielman; spielman@rowan.edu

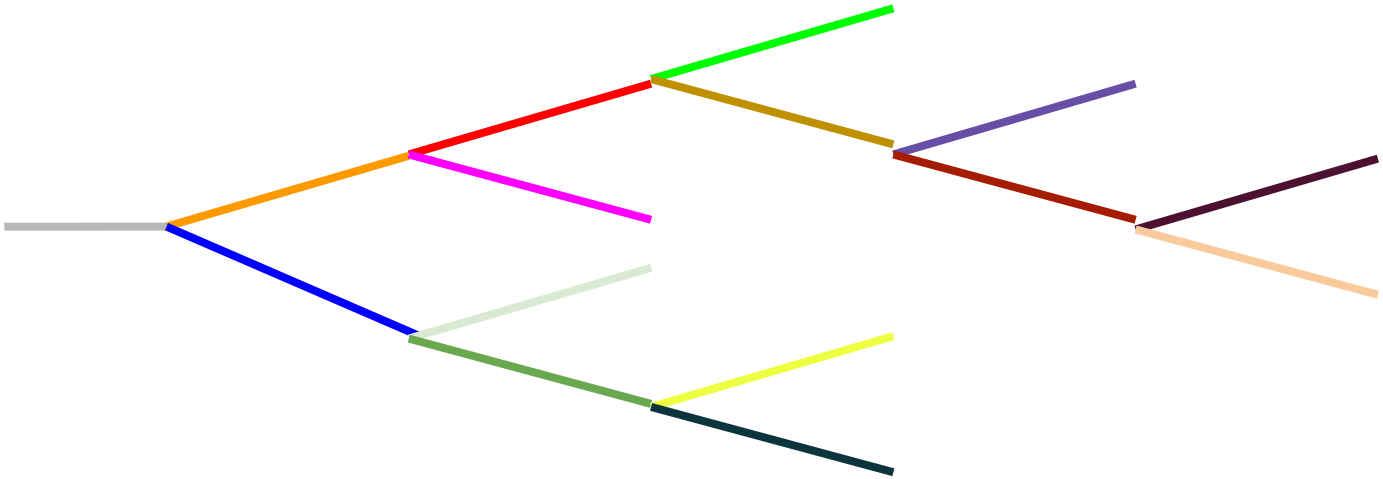


Cartoon by Matthew Bonnan of Macomb, IL, with kind permission of Florida Citizens for Science, Sept. 2010

Cladogenesis: birth of new "clades"



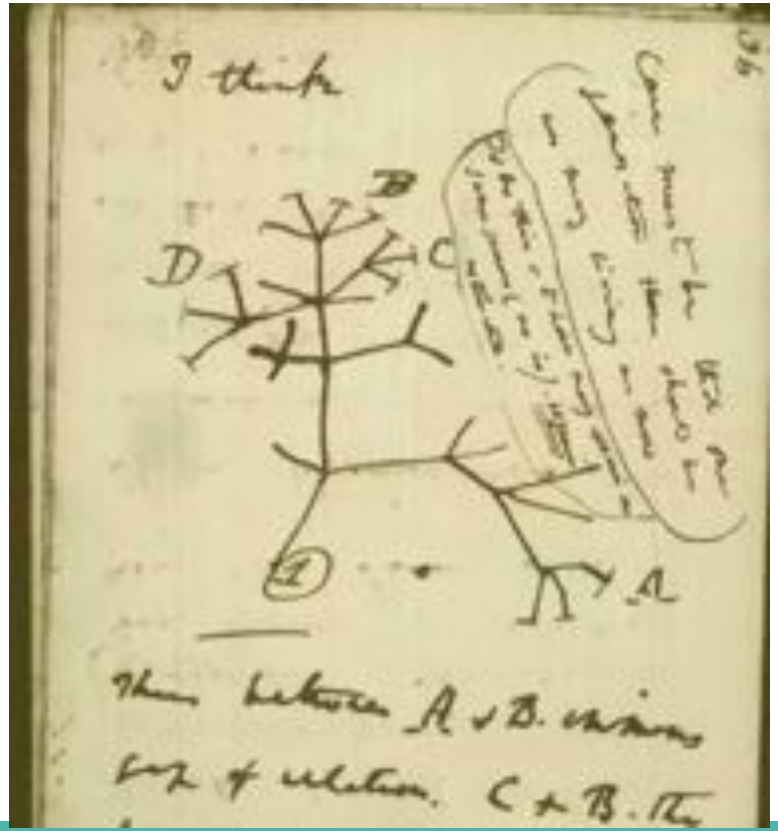
Cladogenesis again and again



This is an **evolutionary tree**, aka **phylogeny**.

Results from repeated rounds of speciation.

The first ever phylogeny



Systematics, Cladistics, Phylogenetics

- Systematics: The study of diversity of life and identification of **taxa** (singular: taxon)

- Cladistics: The systematic classification of groups of organisms using *shared characteristics derived from a common ancestor*

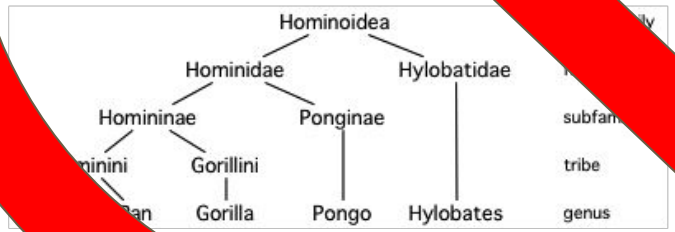
- Phylogenetics: The science of creating evolutionary trees that reveal how organisms are related to each other, based on common ancestry

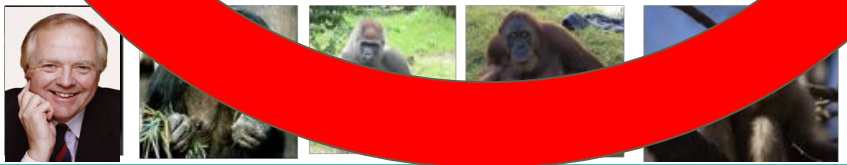
The Linnaean Hierarchy

Kingdom
Phylum
Class
Order
Family
Genus
Species

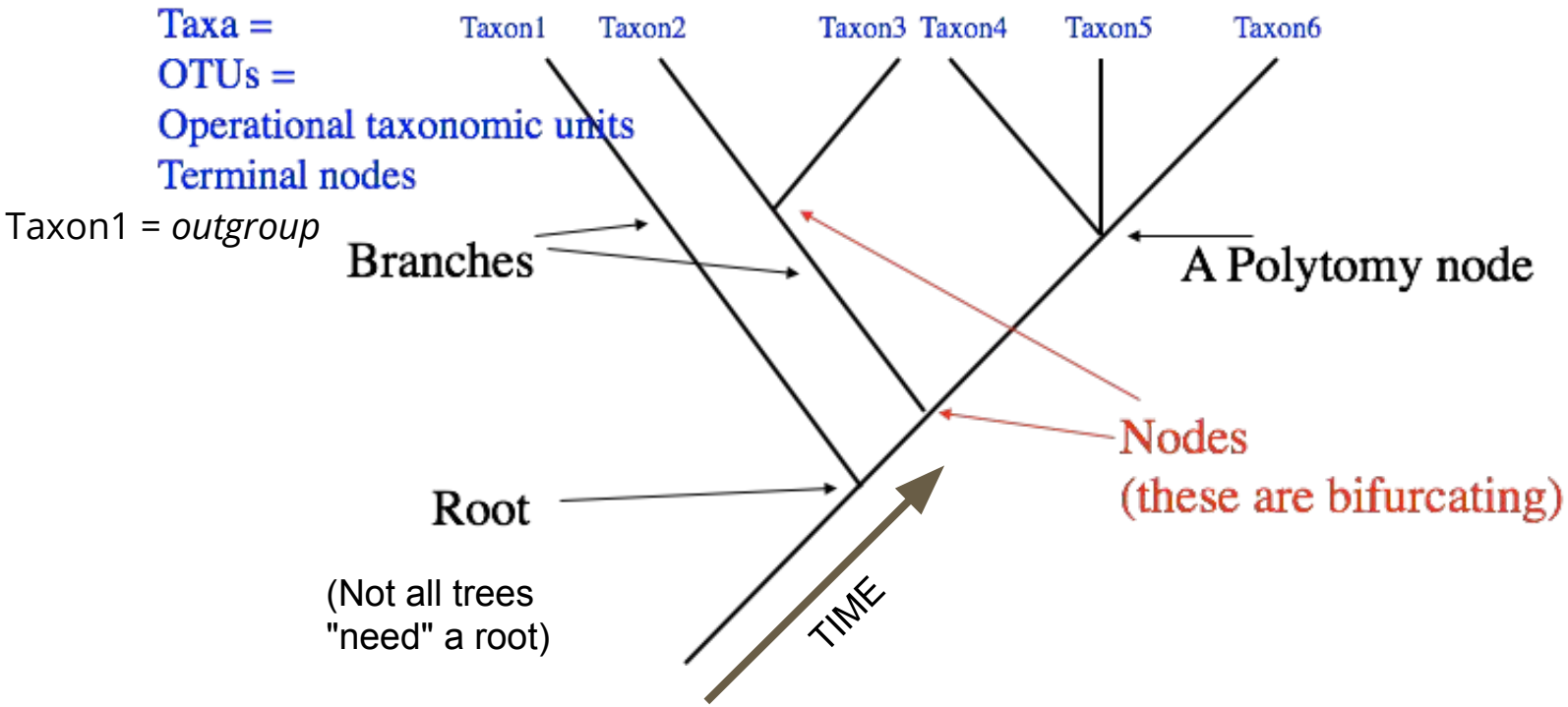
Is a "[illegible] a "tribe" a real evolutionary entity?

	Family	[illegible]	Tribe	Genus	Species
[illegible]oidea	*Hylobatidae*			*Hylobates*	*Hylobates sp.*
	Hominidae	*Ponginae*		*Pongo*	*Pongo pygmaeus*
		Homininae	*Panini*	[illegible]	*Pan paniscus*
					Pan troglodytes
					Gorilla gorilla
			Hominini		*Homo sapiens*

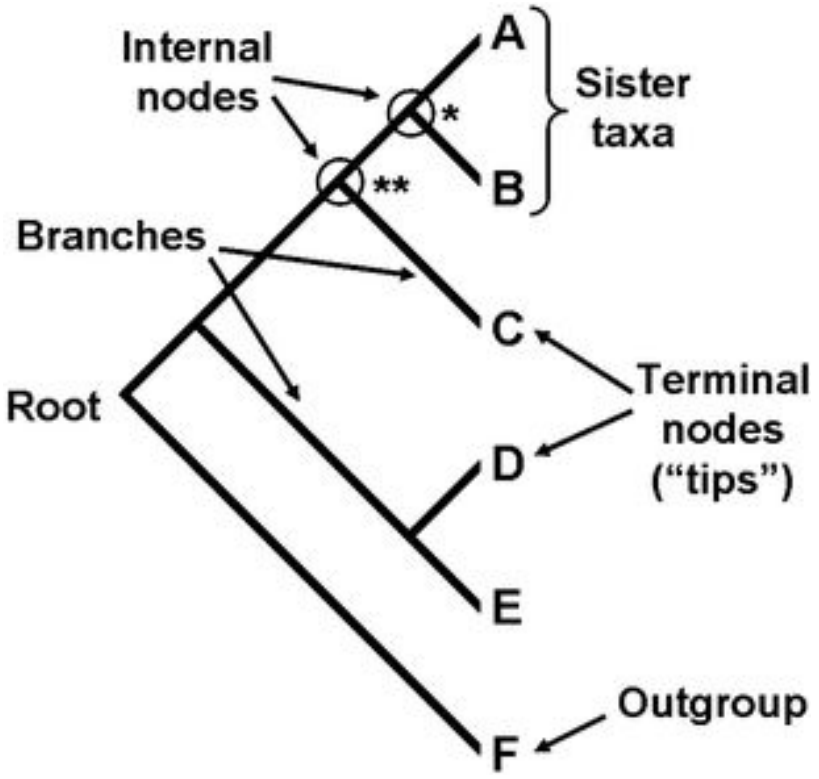




Anatomy of a phylogenetic tree



Put another way..

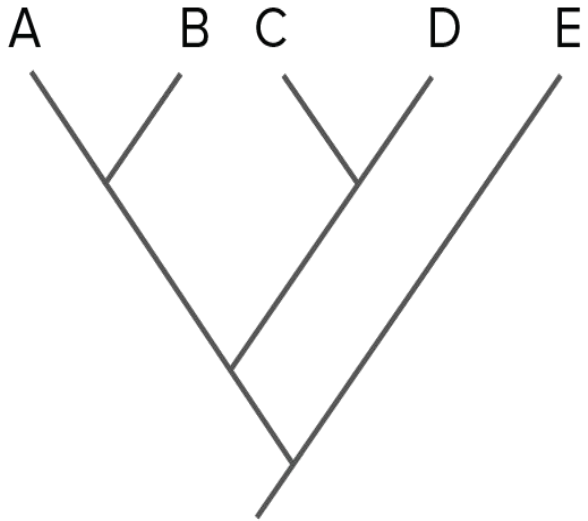


Sister taxa or clades share a common ancestor

A and B are sister taxa

(A,B) and (C,D) are sister clades

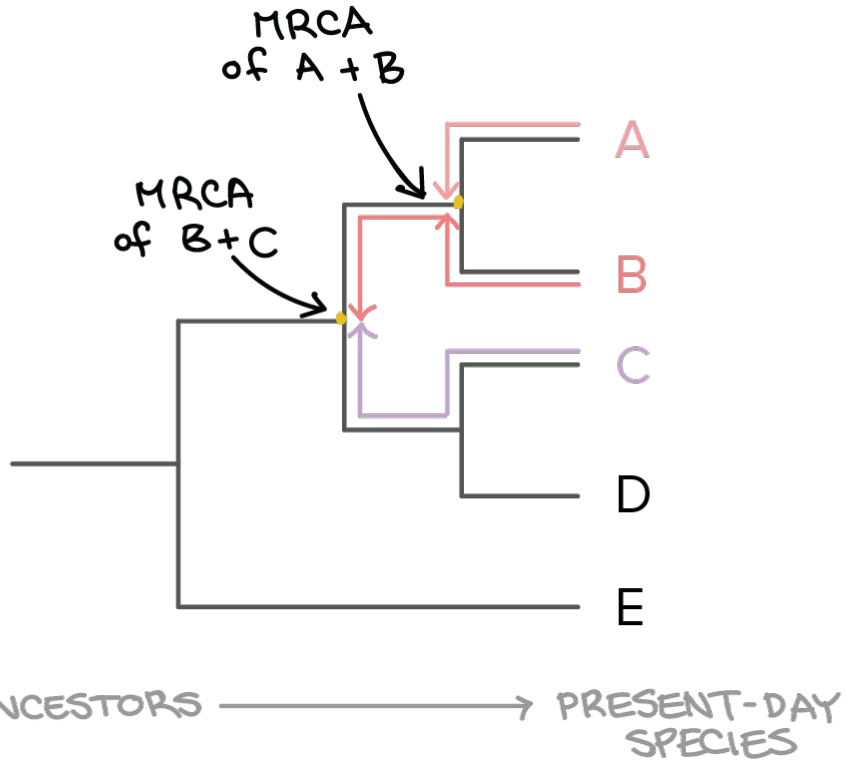
Who is E sister to?



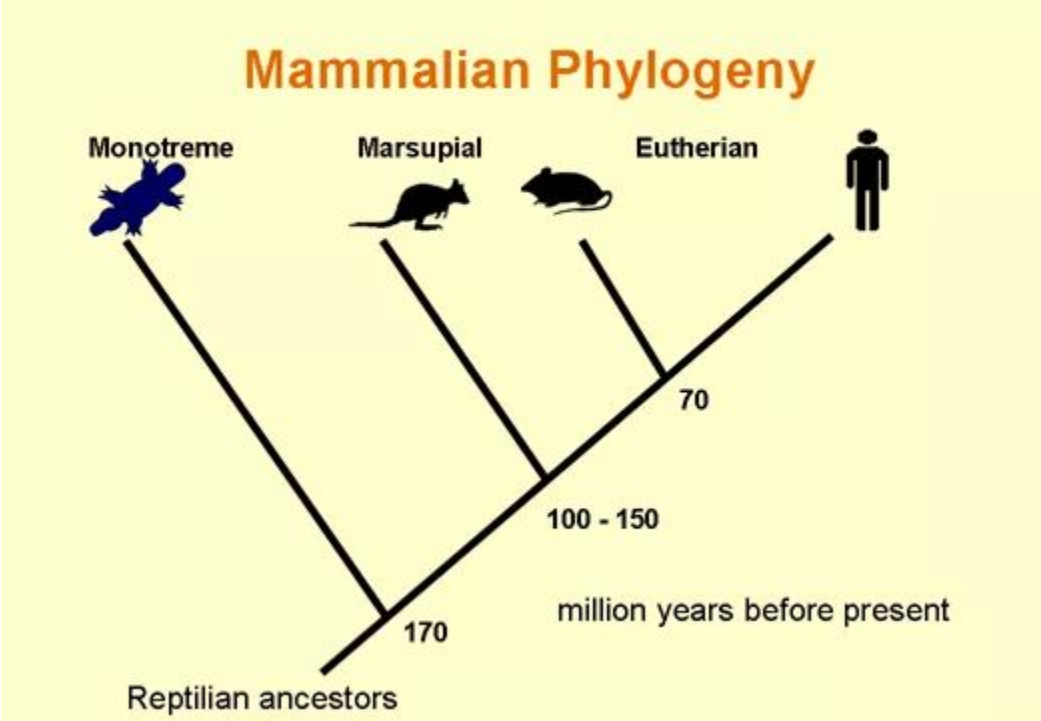
Read trees by tracing back through branches

MRCA = Most Recent Common Ancestor

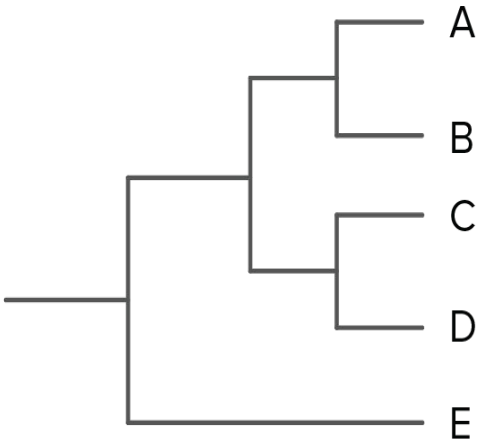
The more recently your MRCA, the more closely you are related.

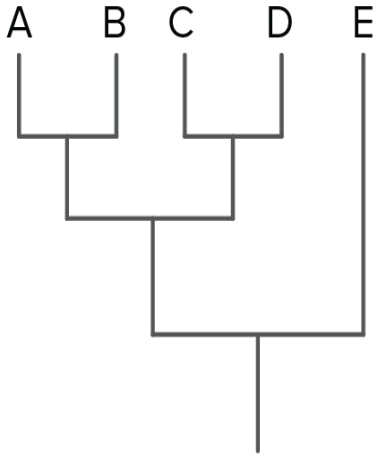


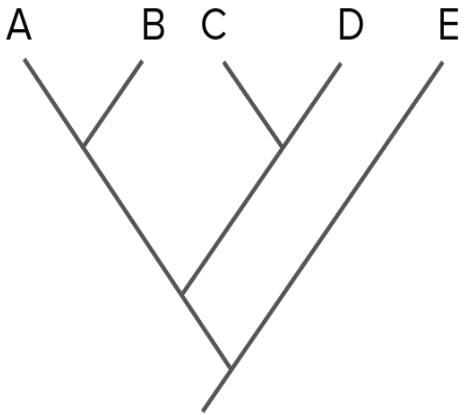
Reading phylogenies: Follow the path from the ancestor



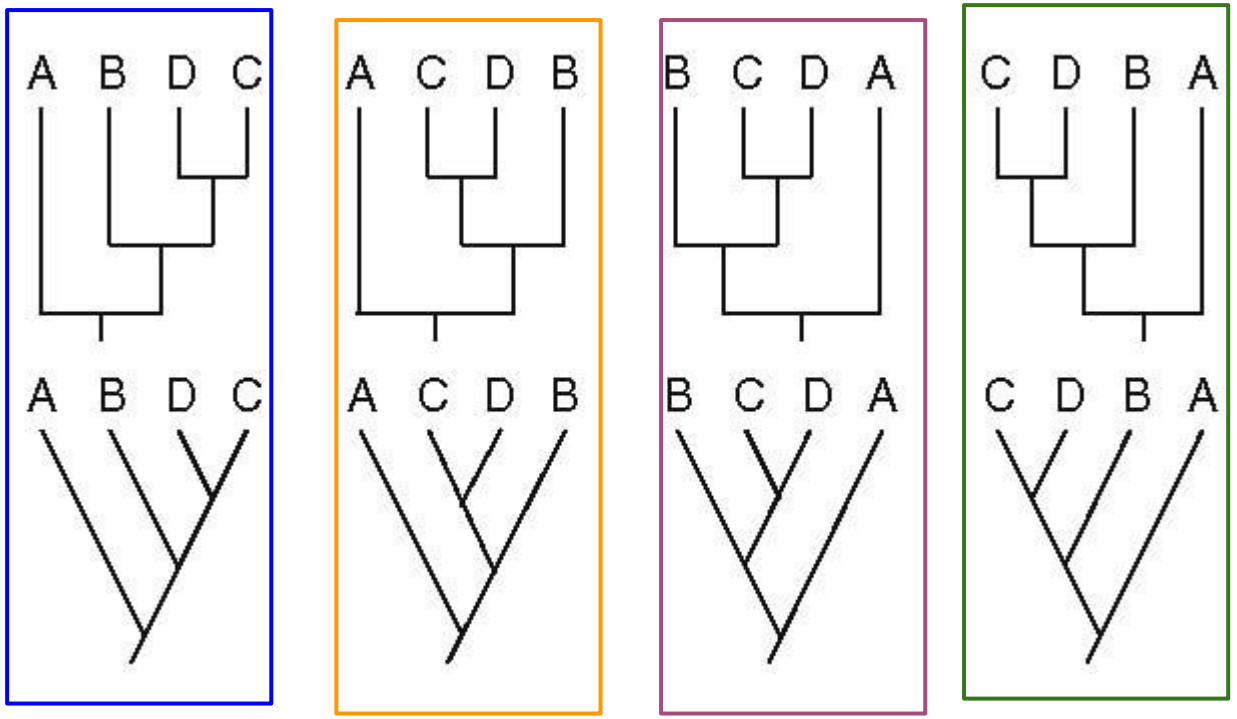
Angled or boxy shapes are still the same.



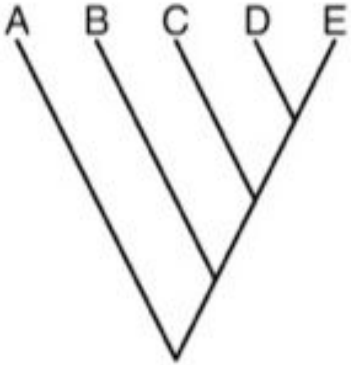


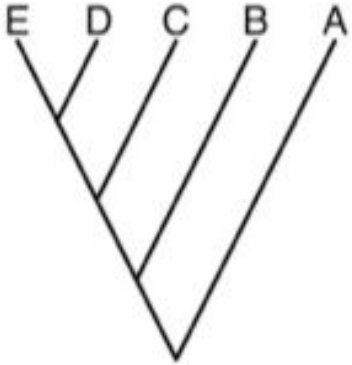


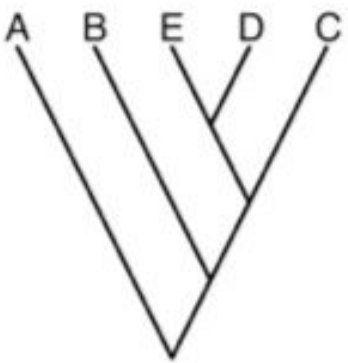
Each box contains the same tree drawn two different ways



How many different A/B/C/D/E relationships are here?



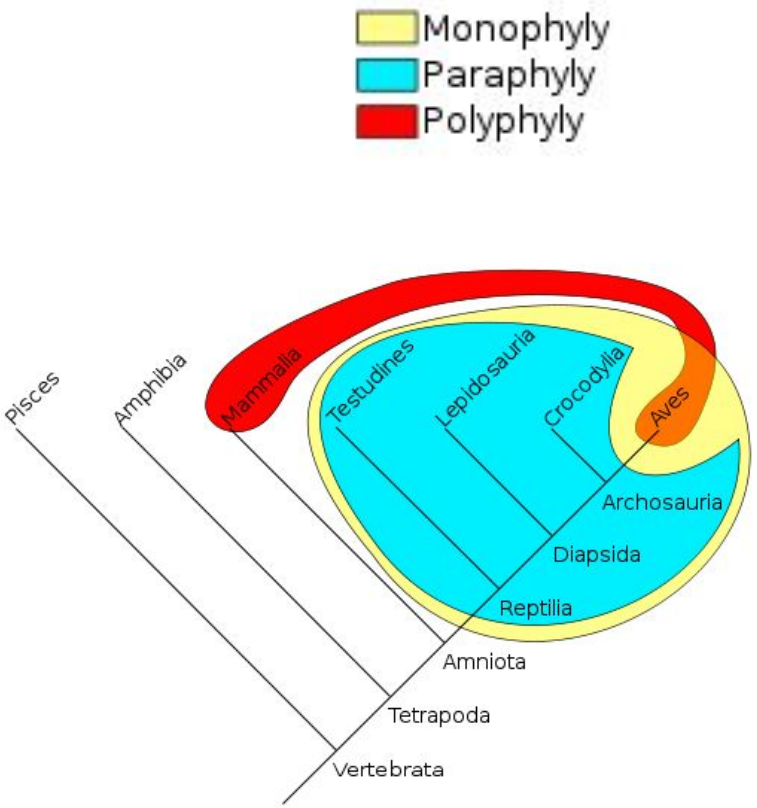




Tree-thinking about traits

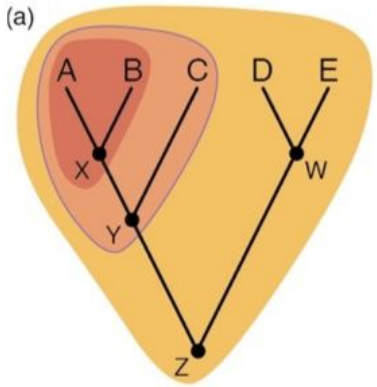
Which is a true evolutionary group?

- **Animals with a four-chambered heart**
  - **Birds and mammals**
- **Birds ("Aves")**
- **Reptiles**
  - **lizards, turtles, snakes, crocodiles**



Types of groupings on trees

- **Monophyletic (monophyly or clade)** is the only type of "true" evolutionary grouping in a tree.
  - Contains common ancestor + **all** descendents

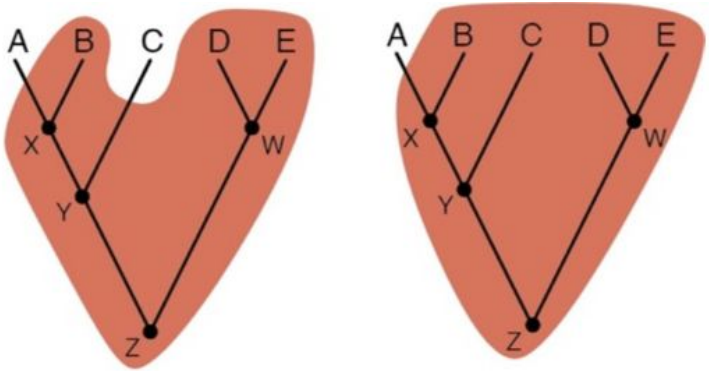


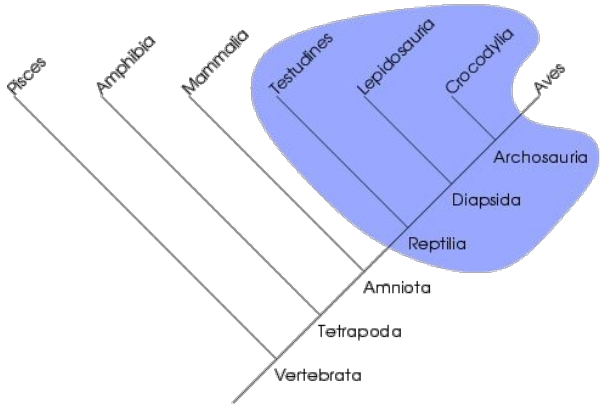
- Each color is a monophyletic group (clade)

- Phylogenies themselves are nested "phylogenetic" groups

Types of groupings on trees

- **Paraphyletic groups** contain only **some** of the descendents from a common ancestor
  - Usually defined based on trait values (has wings or not? has scales or not? etc.)
  - Question: what "might have happened" to the trait in the other taxa?

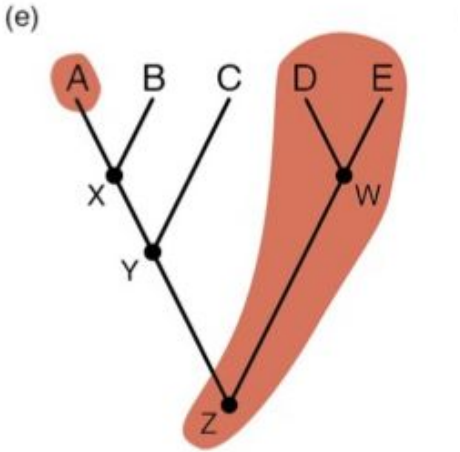




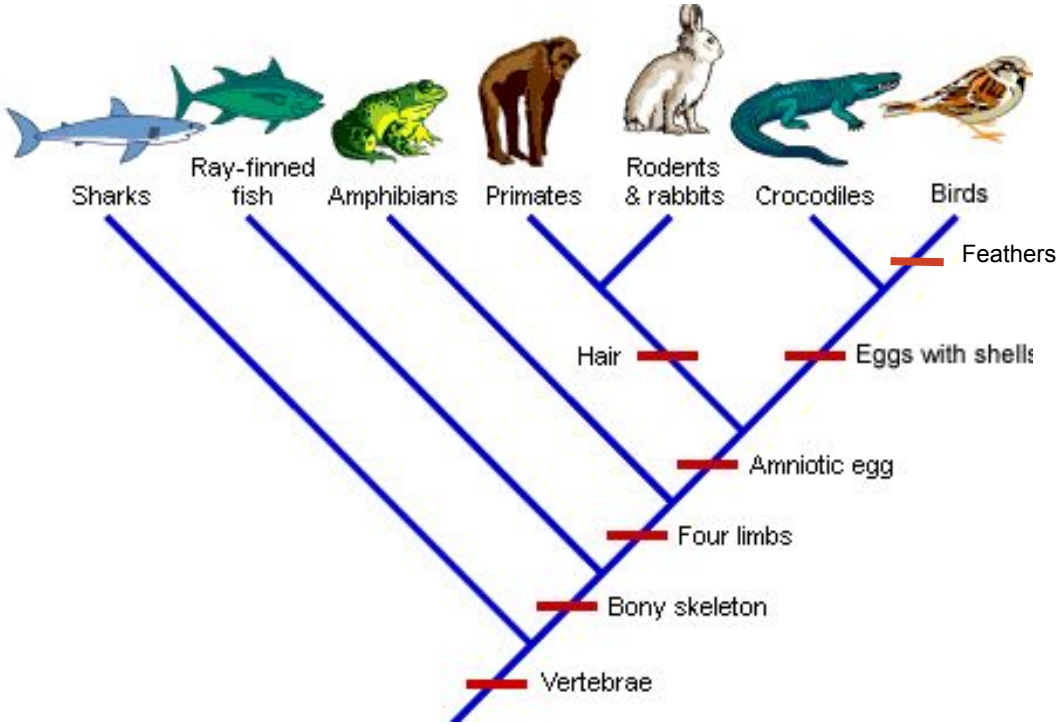
Types of groupings on trees

- **Polyphyletic groups** has at least one ancestral node excluded from the group, *so the internal nodes are not "connected"*
  - Usually defined based on trait values (has wings or not? has scales or not? etc.)
  - It usually looks pretty arbitrary

- Paraphyletic vs polyphyletic
  - Para excludes descendents. *All* nodes can be traced to ancestor.
  - Poly excludes ancestors. *Some* nodes can be traced to ancestor.



Example of tracing the evolution of traits along a tree



The tree-thinking challenge

Complete all questions with your group.

What are all possible trees for three taxa A, B, C?

Number of Species	Number of Possible Trees
3	3
4	15
5	105
6	954
7	10,395
8	135,135
9	2,027,025
10	34,459,425
11	654,729,075
12	13,749,310,575 ← **about 14 billion**
13	316,234,143,225 ← **over 300 billion**

There are two broad ways to make trees

- Using **distance** among all sequences

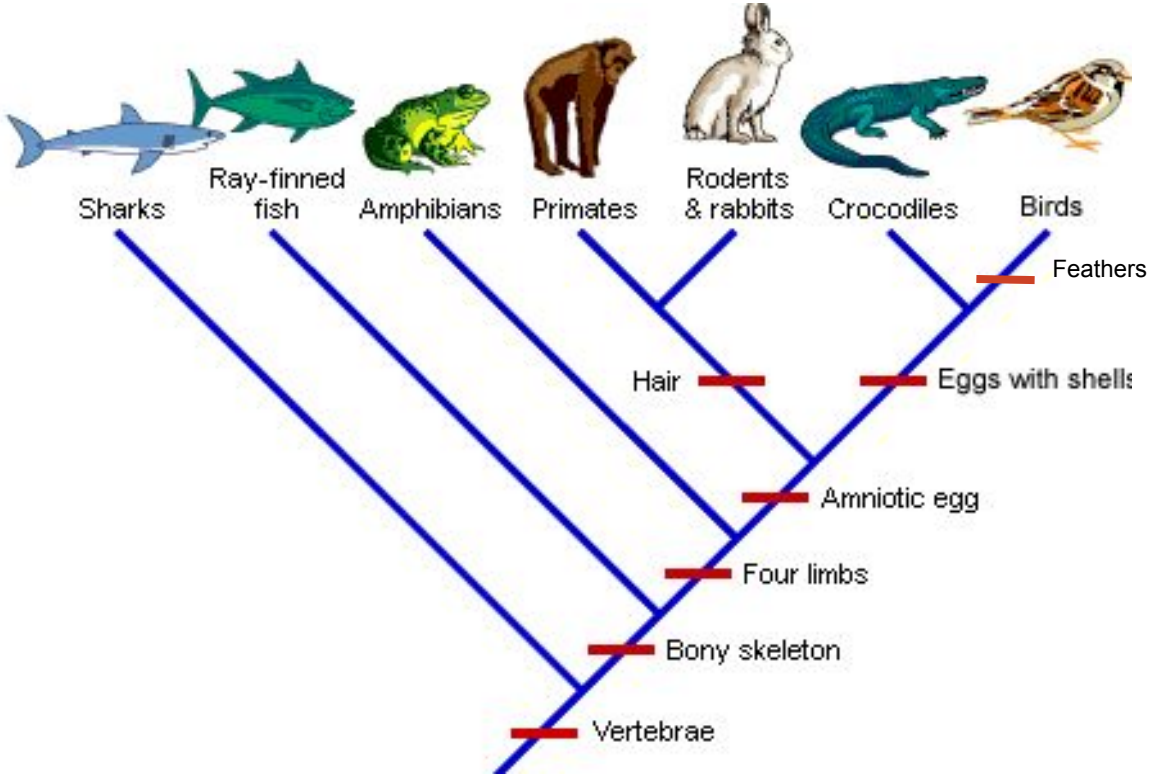
- Using an **optimality criterion**

- Either way, we use **homologous characters (traits)**

We use homologous characters to create phylogenies

- **Character** is any trait (or DNA sequence!) you use to make your tree
  - For the past ~25 years, almost ALL TREES are made from DNA sequences
  - Previously morphology was used
  - Morphology is still used when studying fossils

- The character value in the ancestor is **ancestral**. The character value in a descendent is **derived**.



Imagine we had no DNA and had to make the tree from these traits.

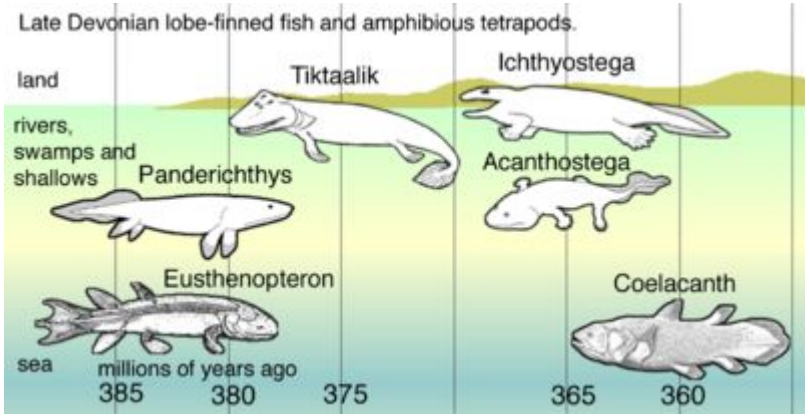
Which characters are **informative**? Which are not informative?

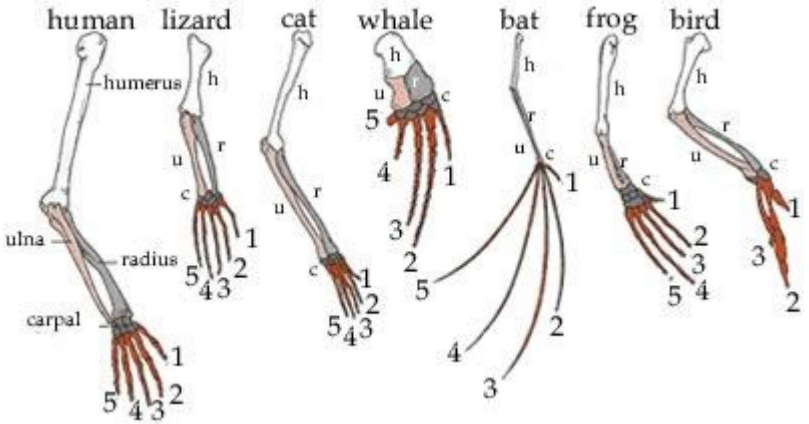
We create trees by tracing change in homologous characters

- **Homology**: traits shared due to **common ancestry**
  - "homologous traits"
  - The trait evolved in the common ancestor, and evolution has "tweaked" the trait as it **diverged** in descendents

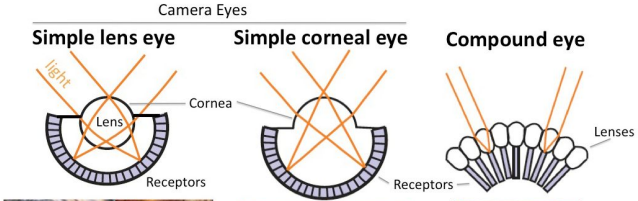

- **Homoplasy**: traits shared due to "acquired" similarity, i.e. **convergent evolution**
  - "analogous traits"
  - The same trait evolves several times independently


Homology of tetrapod limbs





Examples of convergent evolution



Vertebrates, cephalopods, gastropods, jellies, annelids, at least one copepod crustacean

Arachnids, some vertebrates, some larval insects

Insects, crustaceans, some mollusks and annelids

Niche	Placental Mammals	Australian Marsupials
Burrower	Mole	Marsupial mole
Anteater	Lesser anteater	Numbat (anteater)
Mouse	Mouse	Marsupial mouse
Climber	Lemur	Spotted cuscus
Glider	Flying squirrel	Flying phalanger
Cat	Ocelot	Tasmanian "tiger cat"
Wolf	Wolf	Tasmanian wolf

Optimality criterion = a measurement of being optimal (the best)

- Find the tree with the **best value** (optimality) of some measurement (criterion) that tells us if the tree is a **good fit to the data**

- Good fit to the data = the tree and data match really well

- (Note, there are other ways also, but this is the modern-day standard)

Types of optimality criterion

- Parsimony
  - The tree with the fewest steps/evolutionary changes is the best tree
  - **We will learn this one**
  - For all possible trees, the one with the fewest number of changes is the "best"

- Some kind of complicated statistic
  - "Maximum Likelihood"
  - "Bayesian Posterior Probability"
  - **By FAR the most commonly-used approaches in modern-day phylogenetic research**
  - For all possible trees, the one with the highest PROBABILITY is the "best"

Character Matrices show different trait values in species

Rows are species/groups of organisms

Columns are character trait values

Binary (yes/no)character matrix

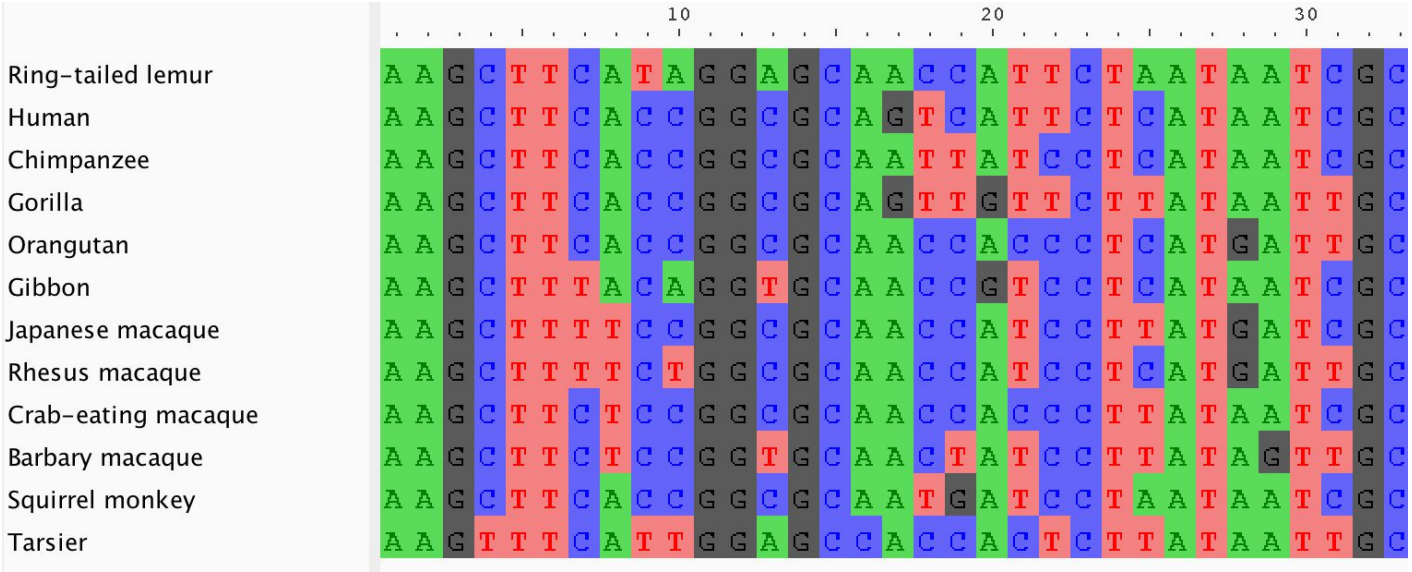
	Vertebrae?	Bony skeleton?	Four limbs?	Amniotic egg?*	Hair?	Two post-orbital fenestrae?**
Sharks and relatives	YES	no	no	no	no	no
Ray-finned fishes	YES	YES	no	no	no	no
Amphibians	YES	YES	YES	no	no	no
Primates	YES	YES	YES	YES	YES	no
Rodents and rabbits	YES	YES	YES	YES	YES	no
Crocodiles and relatives	YES	YES	YES	YES	no	YES
Dinosaurs and birds	YES	YES	YES	YES	no	YES

*amniotic egg:
an egg in which the embryo is surrounded by the moisture-retaining amnion membrane

**post-orbital fenestrae:
holes in the skull behind the eye

Making trees from DNA sequences

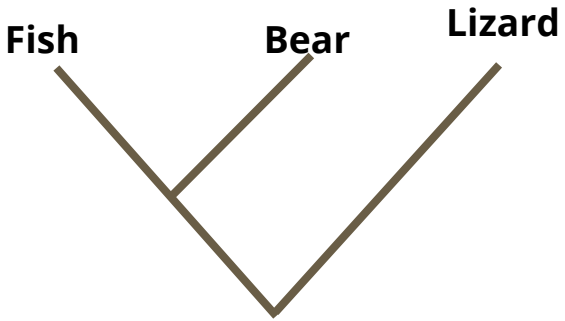
If the character matrix is DNA sequences, it is called a **sequence alignment**

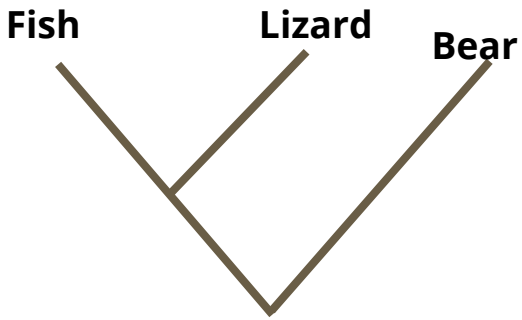


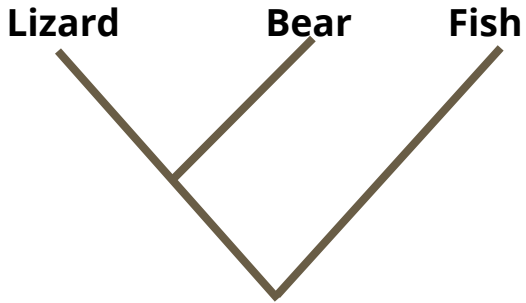
Let's find the best tree under parsimony

```
Fish:     CCAGT
Bear:     ACTGC
Lizard:   GCATC
```

Position 2 is constant,
1, 3, 4, 5 are variable







The tree with the **lowest tree length** is the best tree.

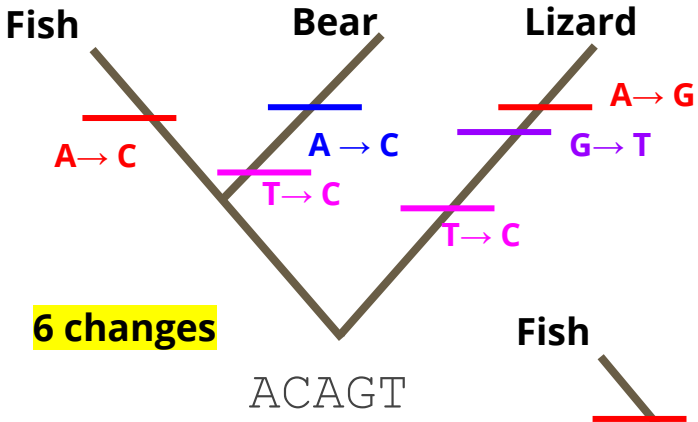
Tree Length = total number of changes along the tree.
This is our **optimality criterion**

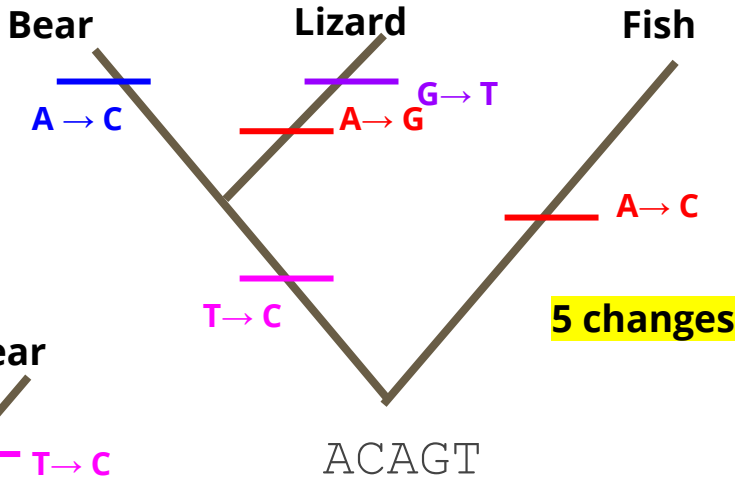
Fish: CCAGT

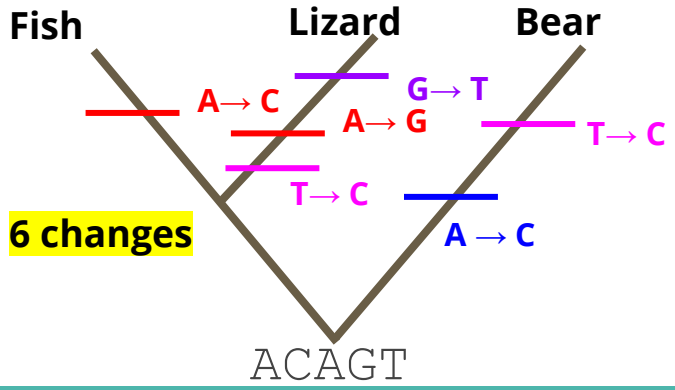
Bear: ACTGC

Lizard: GCATC

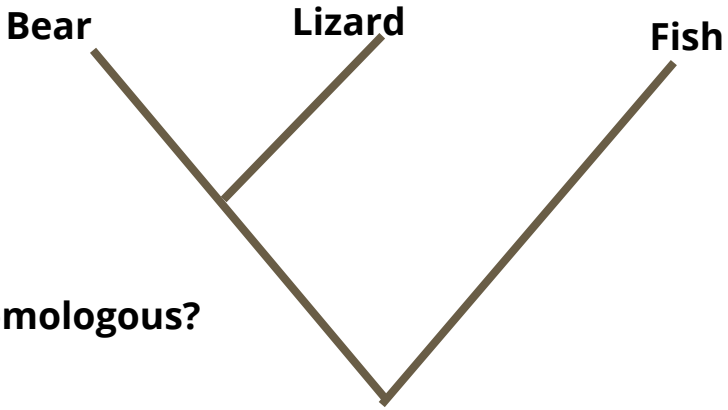
Ancestor: ACAGT [assume; I'll tell you when you need]







Once we have a tree, we can study evolution of traits

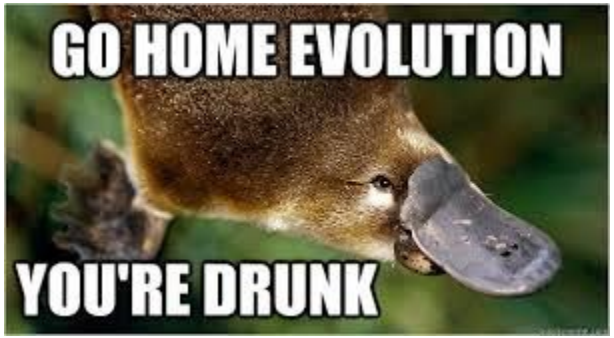


Which traits are likely homologous? Homoplasious?

	Four limbs	Lives on land	Eats insects
Bear	Yes	Yes	No
Lizard	Yes	Yes	Yes
Fish	No	No	Yes

Major question: Is evolution parsimonious??

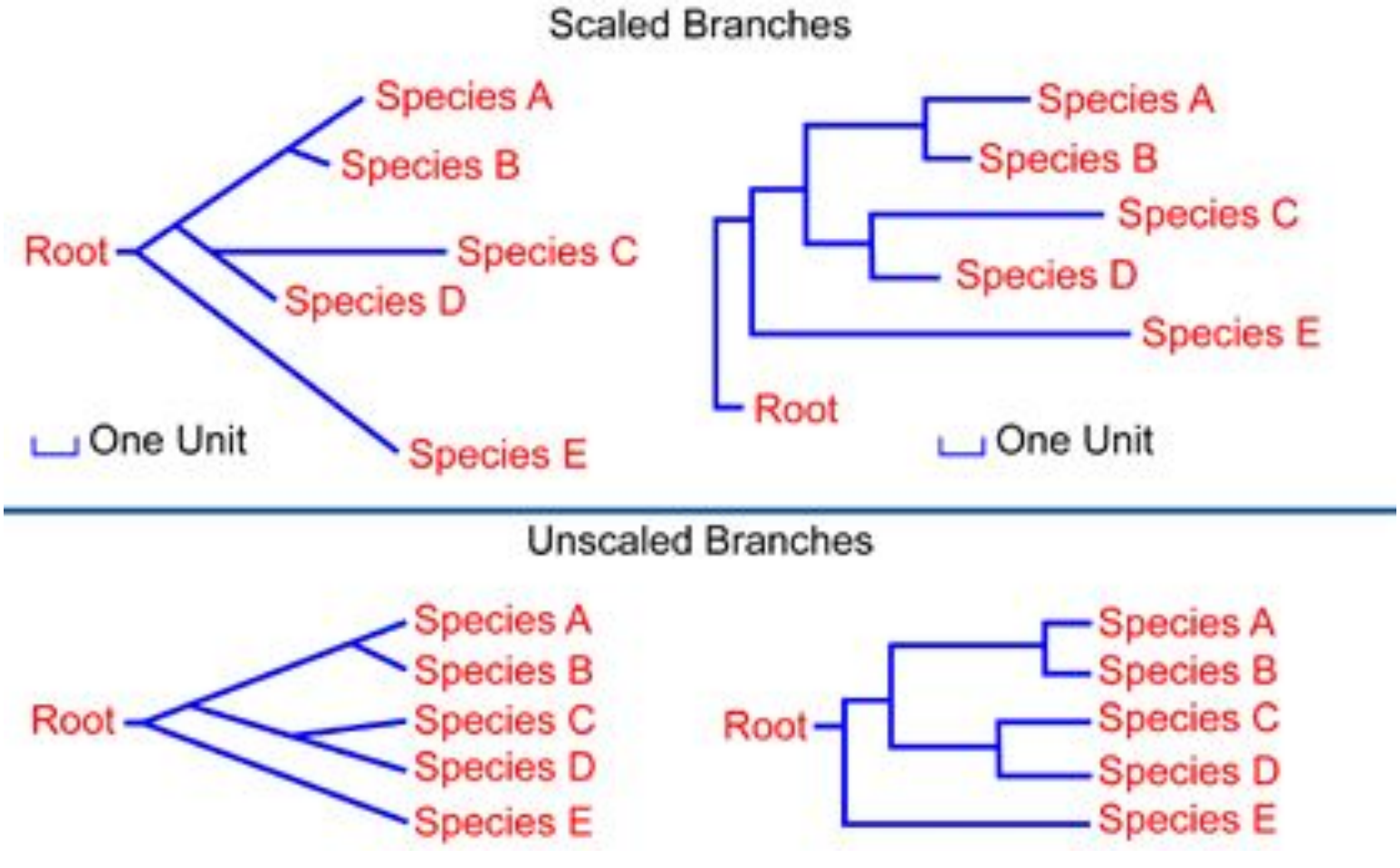
NO EVOLUTION IS OUT OF CONTROL AND DOES CRAZY THINGS



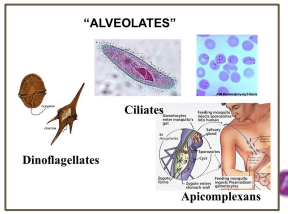
Cladograms vs phylogenies

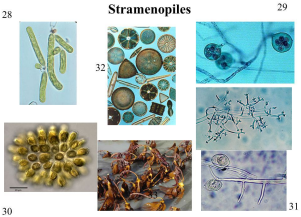
- Cladograms simply show relationships. We are technically making cladograms here

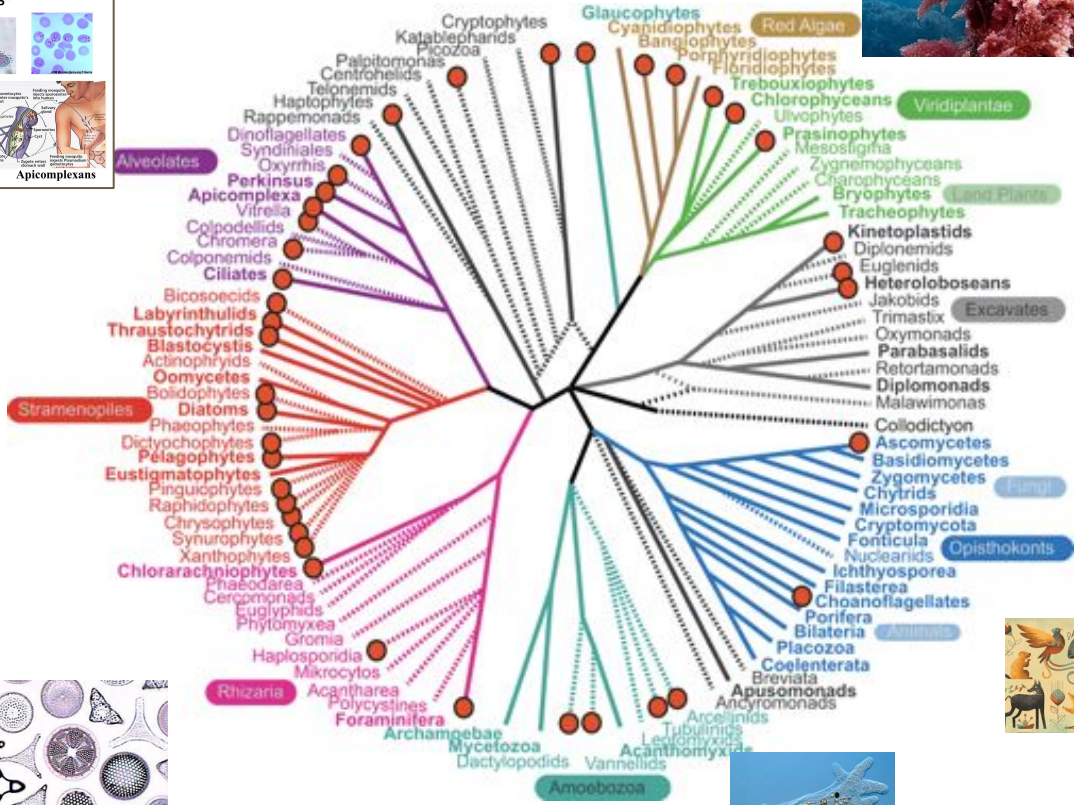
- Phylogenies are cladograms where the **branch lengths** represent the amount of evolutionary change
  - short branches = small amounts of change
  - long branches = large amounts of change
  - **We use genetic data (DNA or protein sequences) to make these**


Scaled Branches
Species A
Species B
Root
Species C
Species D
Species E
One Unit
Species A
Species B
Species C
Species D
Species E
Root
One Unit
Unscaled Branches
Species A
Species B
Root
Species C
Species D
Species E
Species A
Species B
Root
Species C
Species D
Species E

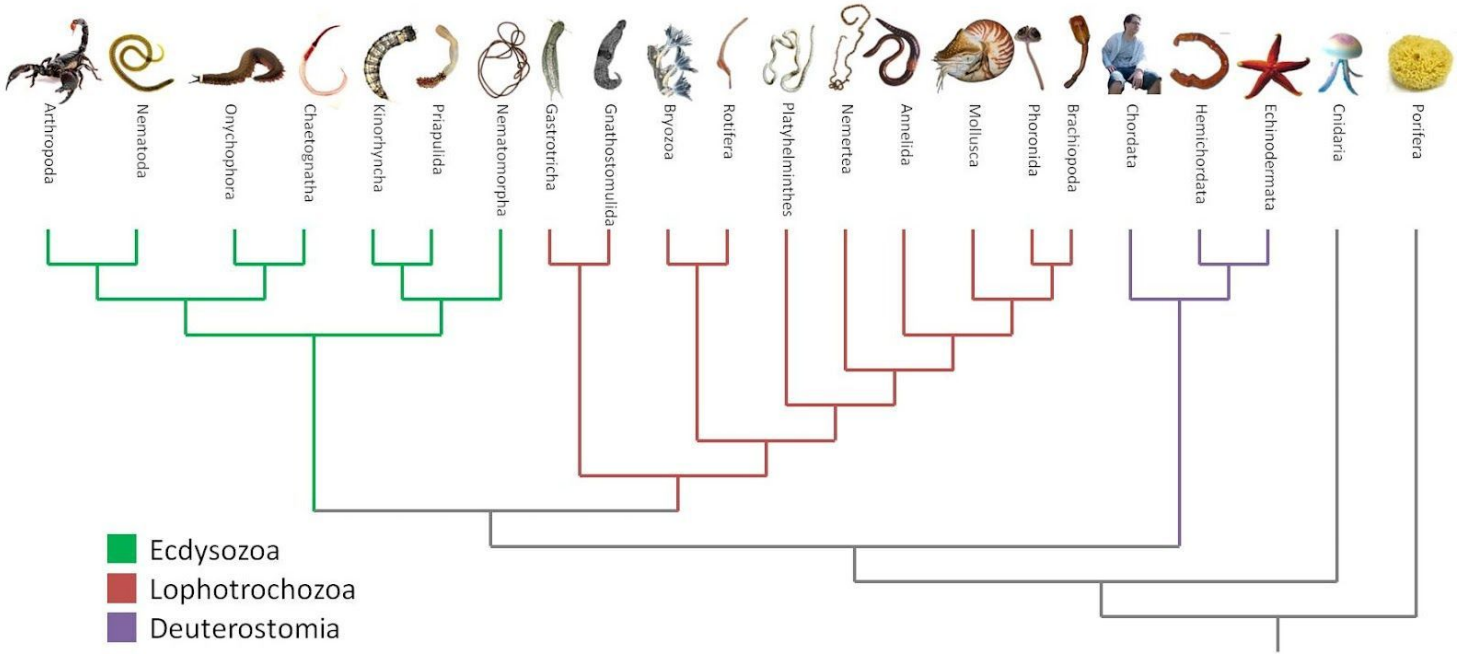
The eukaroytic tree of life





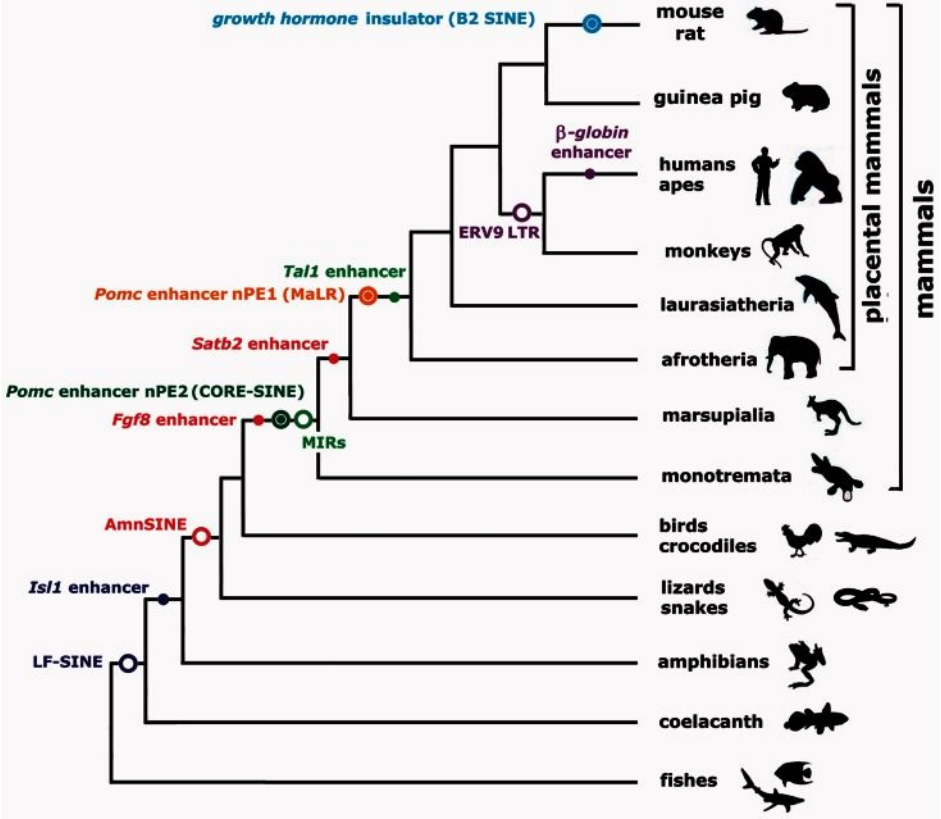


The Metazoan tree of life (animals!)

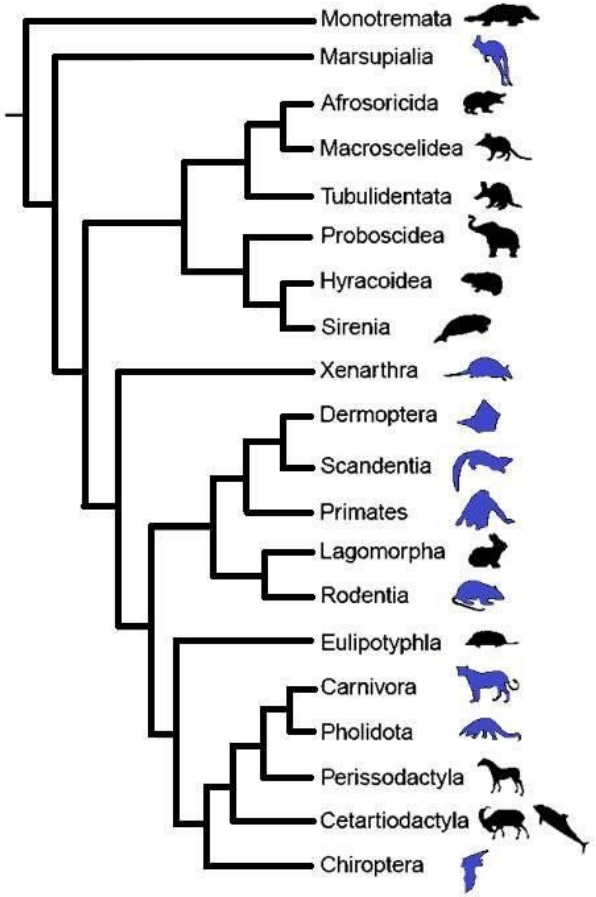


Vertebrate tree of life

(ignore marker dots at nodes)



Mammalian tree of life



The primate tree of life

(ignore branch colors)