

# PHYLOGENY & THE TREE OF LIFE

---





# PREFACE

- ❖ In this powerpoint we learn how biologists distinguish and categorize the millions of species on earth.
- ❖ Early we looked at the *process* of evolution here we look at the *patterns* of evolution.
- ❖ Specifically we will look at **phylogeny**, or the evolutionary history of a species or group of species.
- ❖ To construct these “histories” we will utilize **systematics**, a discipline focused on classifying organisms and determining evolutionary relationships.



# PHYLOGENY

## I. Main Idea:

This section explores **taxonomy**, the naming and classifying of organisms and later we explore how their evolutionary relatedness is derived.





# Binomial Nomenclature

- ❖ To avoid ambiguity when communicating about organisms scientists use Latin scientific names.

*Chrysaora fuscescens*



Jellyfish  
(cnidarian)

*Paranephrops planifrons*



Crayfish  
(crustacean)

*Lepisma saccharina*



Silverfish  
(insect)



# Binomial Nomenclature

- ❖ The two part format is commonly called binomial nomenclature (18th century, C. Linnaeus).
- ❖ The first part is the genus (genera)...capitalized and italicized
- ❖ The second part is the species...italicized

***Homo sapiens***

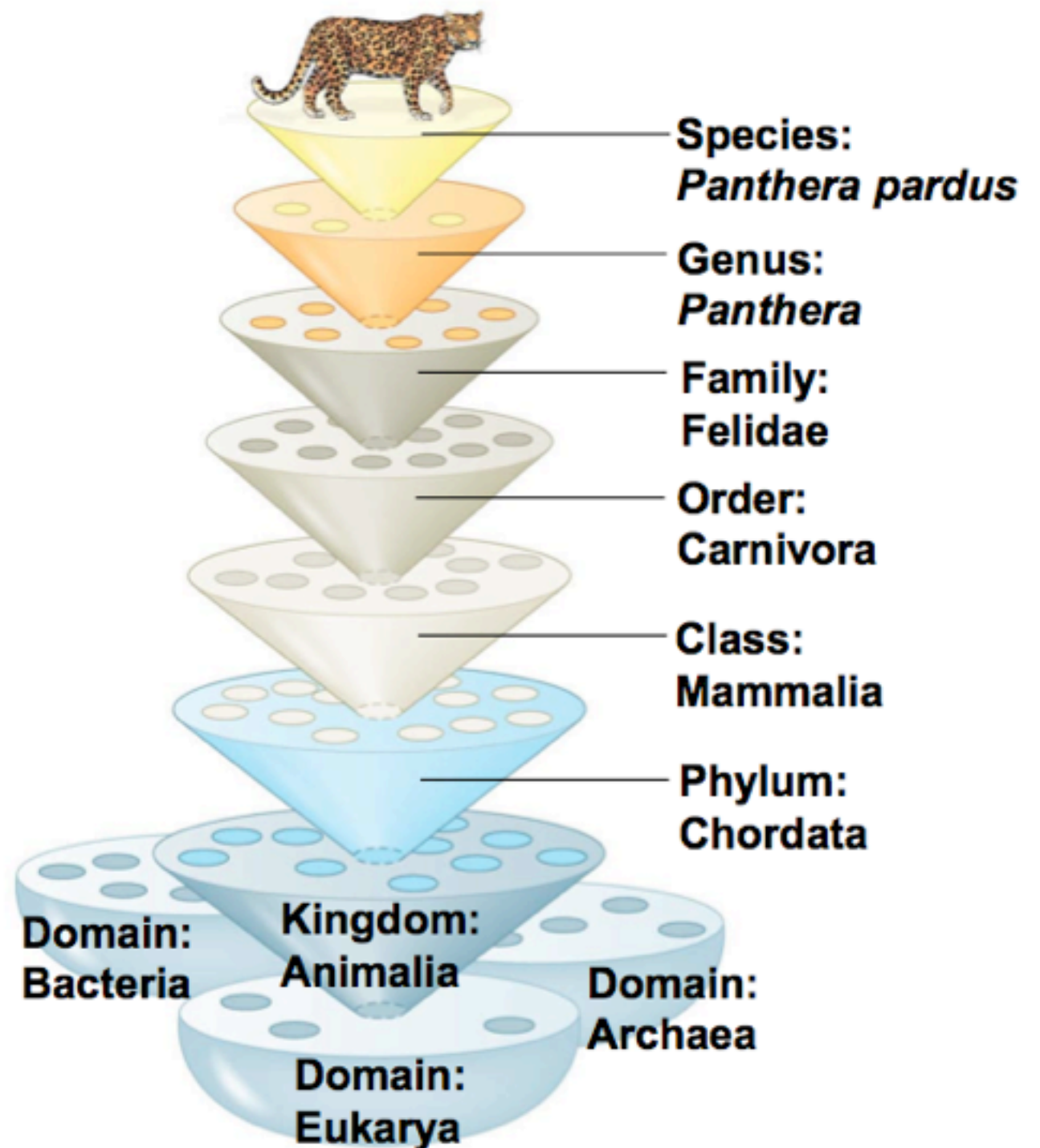
“man wise”



# Hierarchical Classification

- ❖ In addition to a naming system Carolus Linnaeus also developed a grouping system that relied on a hierarchy of groups of increasing inclusivity.

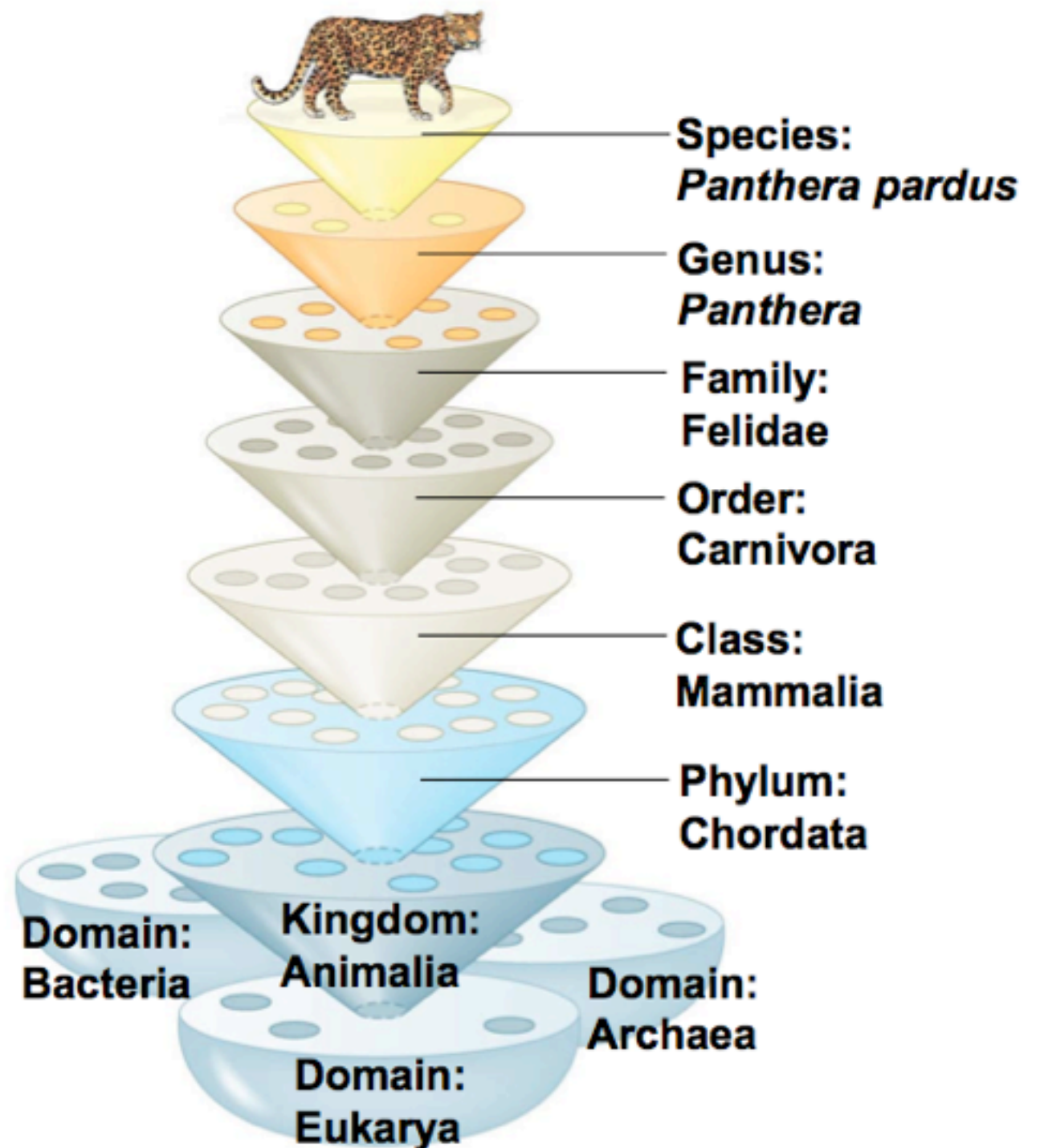
**Domains have  
recently been added**





# Hierarchical Classification

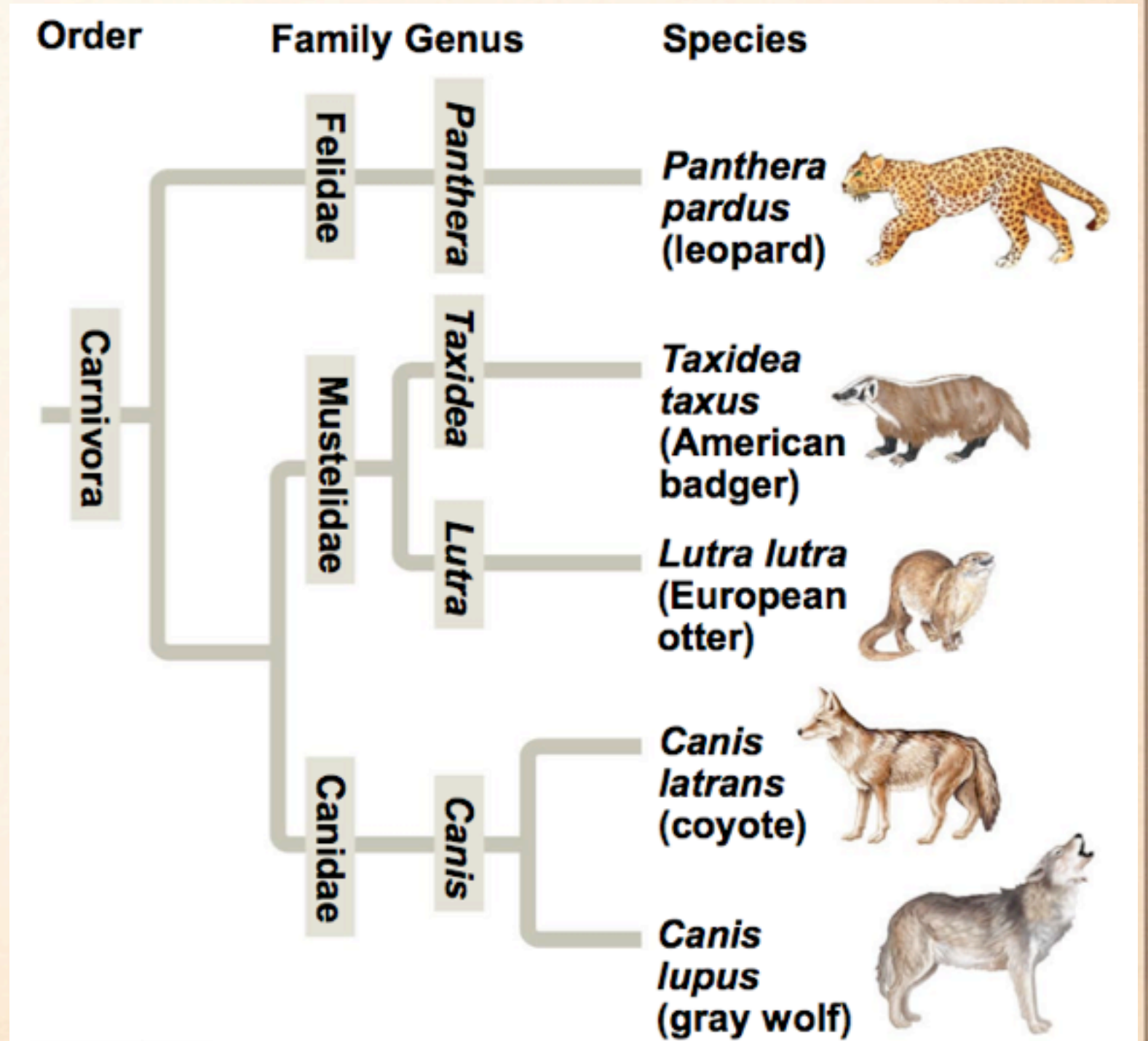
- ❖ Each level is called a **taxon**, taxons are not italicized, just capitalized.
- ❖ This system, structures life into a human view of the world.
- ❖ This system, does not necessarily reflect evolutionary relationships between groups.





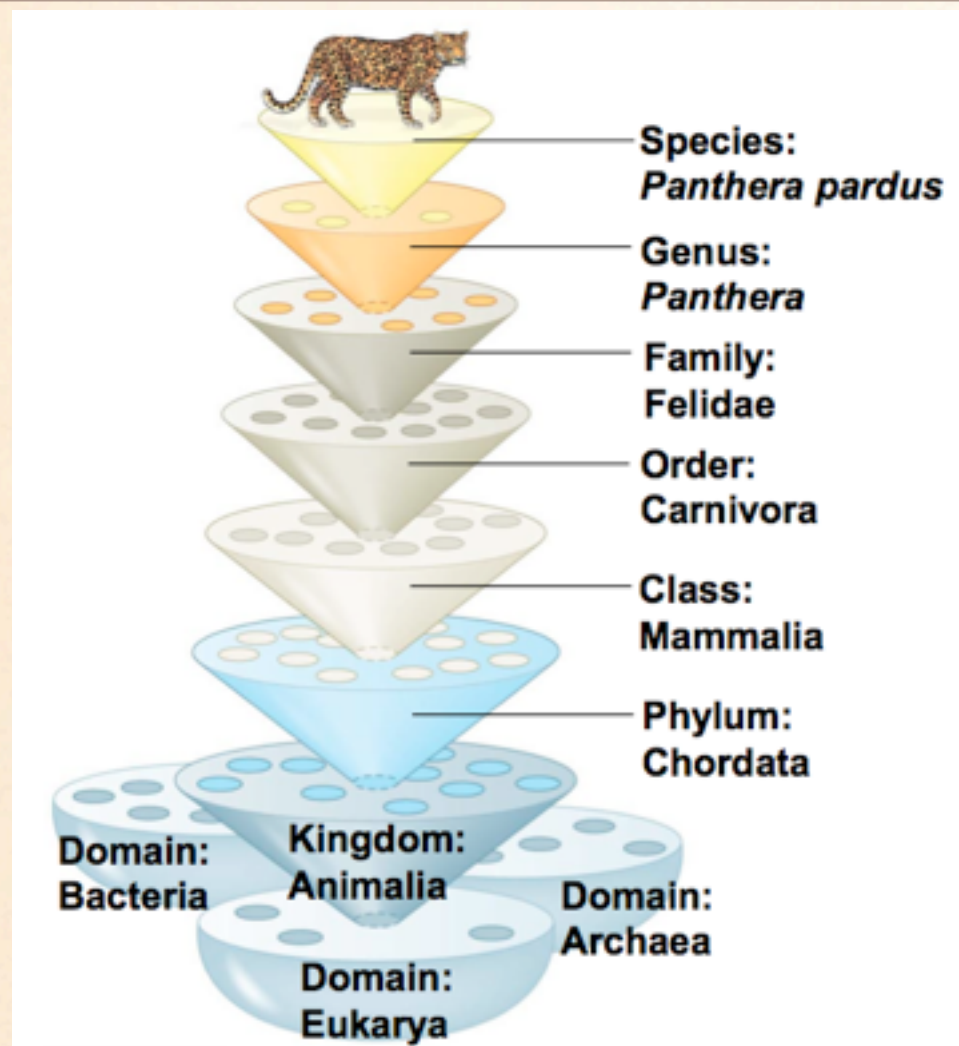
# Linking Classification & Phylogeny

- ❖ The evolutionary history of a group of organisms can be represented in a branching diagram called a **phylogenetic tree**

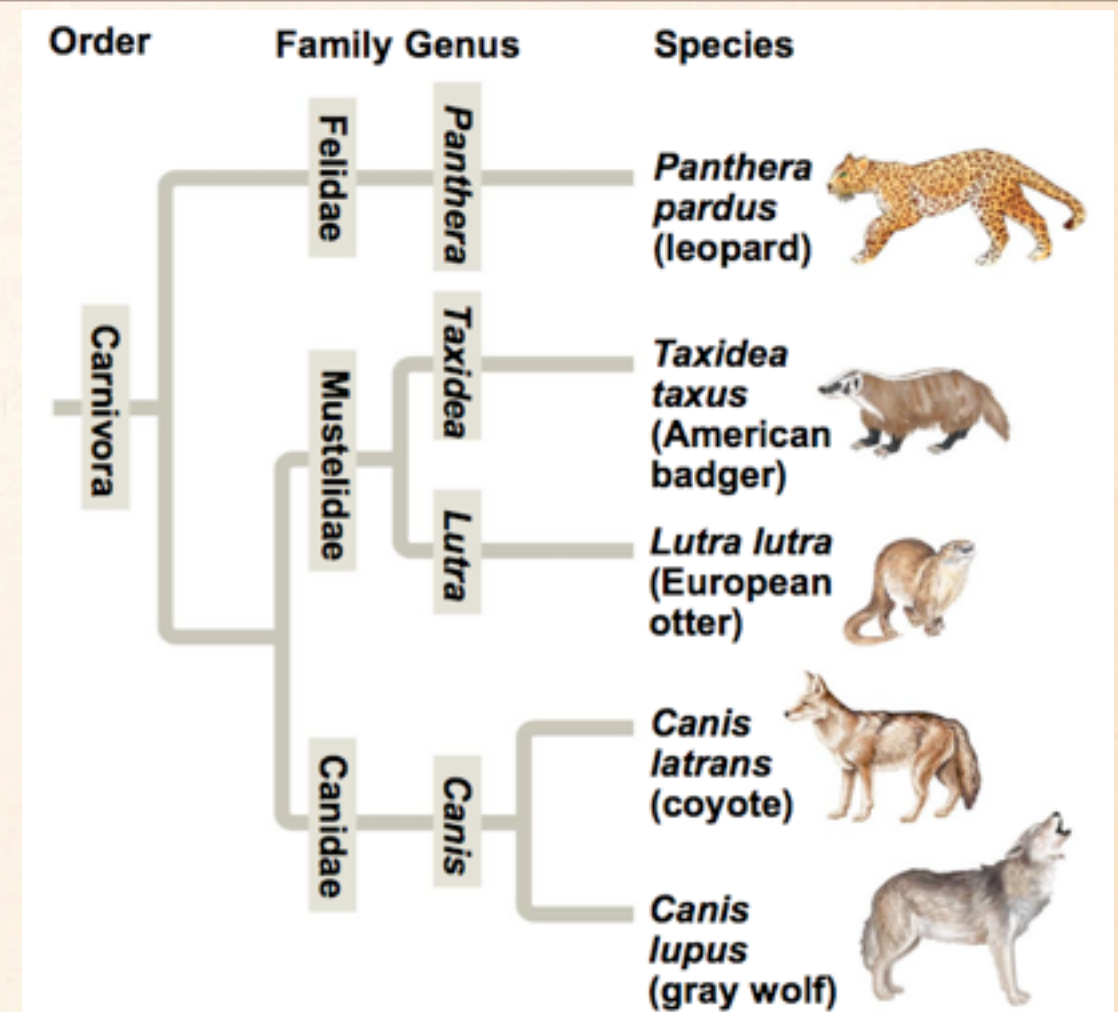




- ❖ Often the branching pattern match the nesting relationships of the taxons, sometimes however taxonomists classify a species into a group to which it is not closely related.



VS.



- ❖ One reason for this disagreement may lie in the fact a species lost a key feature shared by its close relatives, if DNA or other evidence supports this the species will be reclassified.



- ❖ In fact to avoid these disagreements some systematists have suggested a classification based entirely on evolutionary relationships a system called the...
- ❖ **PhyloCode** where groups would include a common ancestor and all its descendants.
- ❖ it would change how taxa are defined and recognized
- ❖ groups would no longer have ranks assigned to them
- ❖ their names would remain the same
- ❖ although still a bit controversial, many are adopting anyways



- ❖ **Regardless of which one is used, a phylogenetic tree represents a hypothesis about evolutionary relationships.**
- ❖ Each **branch point** represents the divergence of two species
- ❖ **Sister taxa** are groups that share an immediate common ancestor
- ❖ A **rooted** tree includes a branch to represent the last common ancestor of all taxa in the tree
- ❖ A **basal taxon** diverges early in the history of a group and originates near the common ancestor of the group
- ❖ A **polytomy** is a branch from which more than two groups emerge



**Branch point:  
where lineages diverge**

**ANCESTRAL  
LINEAGE**

**This branch point  
represents the  
common ancestor of  
taxa A–G.**

**This branch point forms a  
polytomy: an unresolved  
pattern of divergence.**

**Taxon A**

**Taxon B**

**Taxon C**

**Sister  
taxa**

**Taxon D**

**Taxon E**

**Taxon F**

**Taxon G**

**Basal  
taxon**



# What can we can learn, what we cannot learn from Phylogenetic trees

- ❖ **First, phylogenetic trees are intended to show patterns of decent NOT phenotypic similarity.**
- ❖ **Second, the sequence of branching shows only the pattern of descent** they do NOT necessarily indicate the actual or absolute ages of a particular species
- ❖ **Third, the common ancestors in a tree are extinct and they were distinct from todays extant species,** it should NOT be assumed that a taxon evolved from the taxon next to it.



# Applying Phylogenies

- ❖ Understanding phylogeny can have practical applications.
- ❖ Phylogeny provides important information about similar characteristics in closely related species
- ❖ Because phylogenies use DNA data to help create them we can use phylogenetic trees to identify closely related species
- ❖ *Consider corn, an important food crop, by identifying certain grasses (corn's closest relatives) scientists can protect or cultivate these grasses to save useful alleles for cross breeding.*



# Another Application

## EXPERIMENT

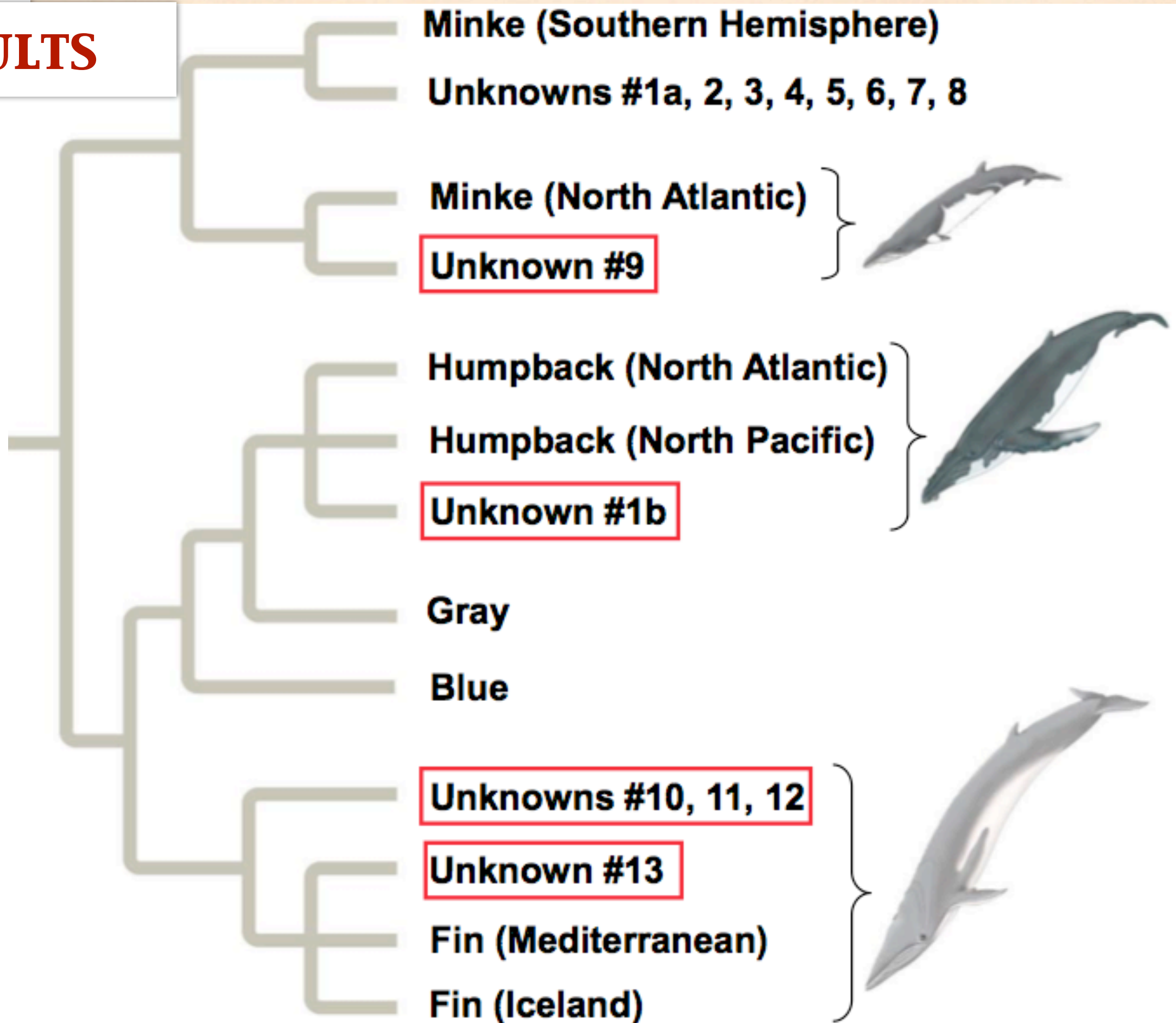
C.S. Baker (Univ. of Auckland) and S.R. Palumbi (Univ. of Hawaii) purchased 13 samples of “whale meat” from Japanese fish markets. They sequenced a specific part of the mitochondrial DNA from each sample and compared their results with comparable DNA sequence from known whale species. To infer identity of each sample, they constructed a gene tree, a phylogenetic tree that shows patterns of relatedness among DNA sequences rather than among taxa. **RESULTS**

## CONCLUSION

This analysis indicated that DNA sequences of six of the unknown samples (in red) were most closely related to DNA sequences of whales that are NOT LEGAL to harvest.



# RESULTS





# PHYLOGENY

## II. Main Idea:

The generate phylogenetic trees systematists gather information about morphology, biochemistry and of course genes.





# Morphological & Molecular Homologies

- ❖ Recall **Homologies**- phenotypic and genotypic similarities due to shared ancestry.
- ❖ In general organisms that share similar morphology or DNA are likely to be more closely related than others that do not share these similarities.
- ❖ Sometimes strikingly different looking organisms share much the same DNA sequences or vice versa and provides challenge to those trying to construct accurate phylogenetic trees.



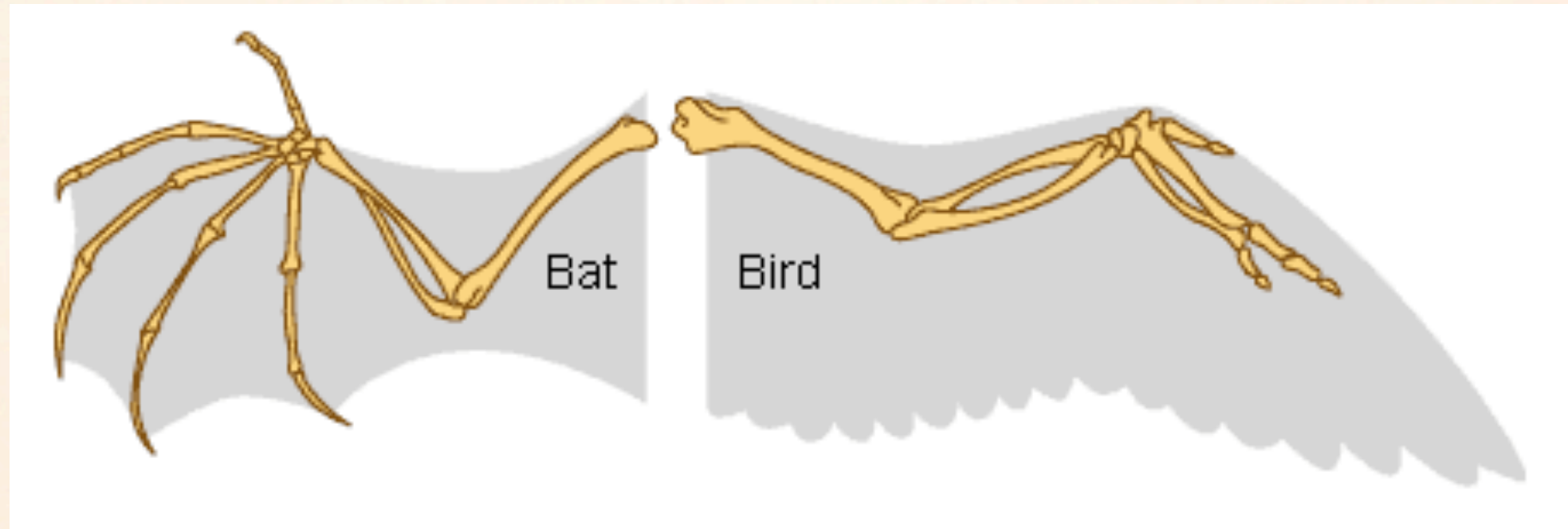
# Sorting Homology from Analogy

- ❖ One major obstacle in generating phylogenetic trees is distinguishing between homology and analogy.
- ❖ Recall Analogy results when organisms are similar due to convergent evolution not because of recent shared ancestry.
- ❖ Recall Convergent Evolution occurs when similar environmental pressures and natural selection produce similar adaptations in organisms from different lineages.

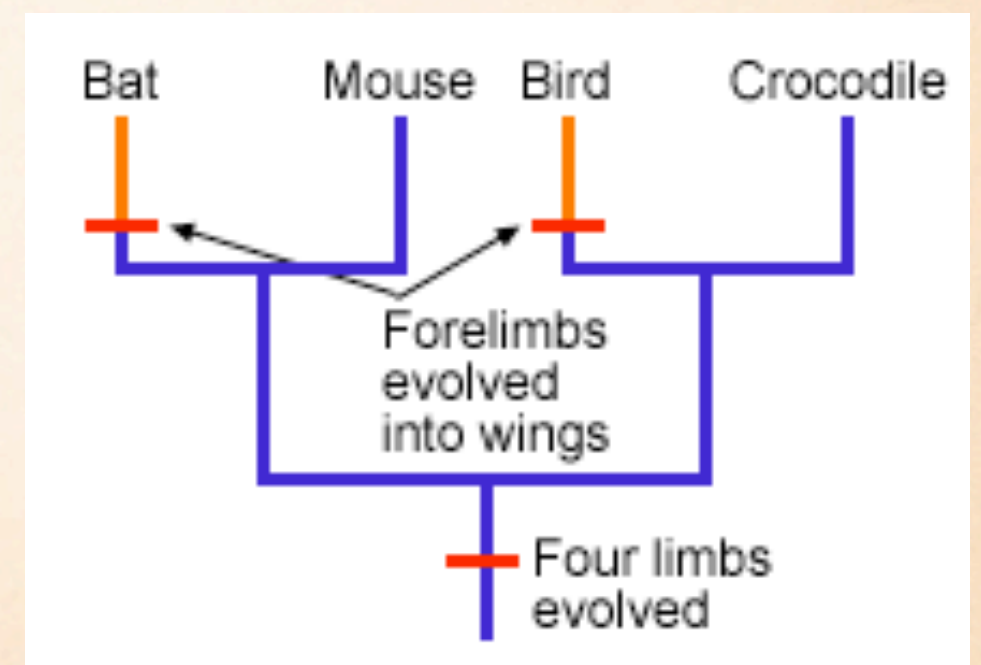




- ❖ Analogous structures that arose independently are also called **homoplasies**.



- ❖ The bat wing consists of stretched skin while the bird uses feathers for flight.
- ❖ Ironically birds are more related to crocodiles than bats.



# Sorting Homology from Analogy

- ❖ Another clue in distinguishing between homology and analogy lies in the complexity of the characters being compared.
- ❖ The more elements that are similar in a two complex structures the more likely share a recent common ancestor
- ❖ The same argument can be made at the level of the gene.



# Evaluating Molecular Homologies

- ❖ Comparing DNA sequences between organisms is likely the best way to establish relatedness but it also poses the greatest technological challenges as well.
- ❖ To start with DNA in each organism has to be sequenced (challenging by itself but science is better at sequencing everyday and to date has sequenced over 100 billion bases from 1000's of species).
- ❖ Next these sequences must be compared (perhaps even more challenging)
- ❖ In general, the more the sequences align the more similar they are

# Species 1

# Species 2

..**A**TACGGATACGG..

..**T**TACGGATACGG..

..**A**TACGGATACGG..

..**T**TACGGATACGG..

Sequence looks similar, difference is likely due to a single base substitution

**BUT... WHAT IF THE SINGLE ADENINE WAS LOST BY A DELETION MUTATION? COULD IT CHANGE OUR INTERPRETATION?**



# Species 1

# Species 2

..**A**TACGGATACGG..

..**T**TACGGATACGG..

..**T**ACGGATACGG**G**..

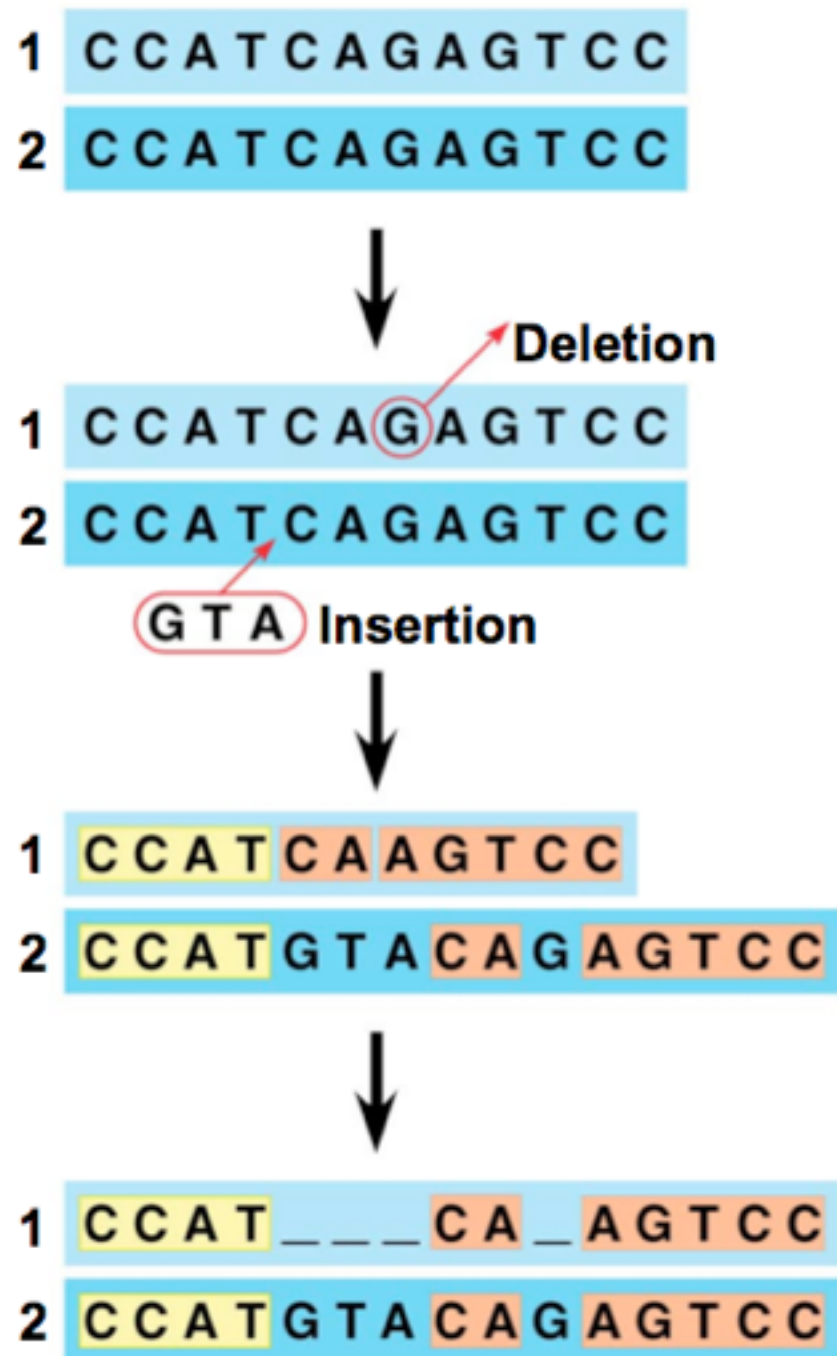
..**T**TACGGATACGG..

Now the sequence looks very different, when in fact only a single base deletion represents the difference

To address such problems researchers have developed computer programs and statistical tools that estimate the best way to align comparable DNA segments of differing lengths and differentiate among molecular homoplasies.

Species 1  
Species 2

Species 3  
Species 4



3  
4

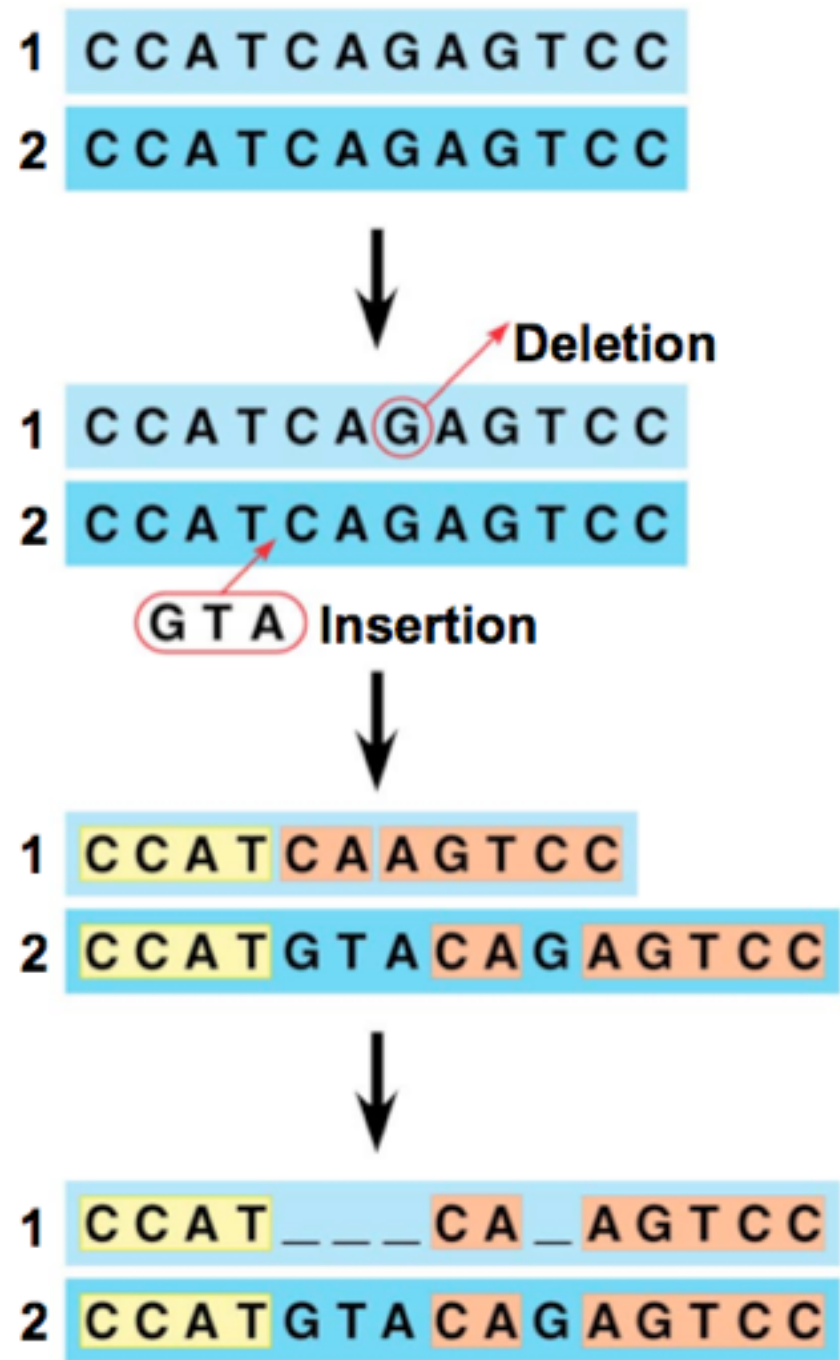
ACGGATAGTCCACTAGGCACTA  
TCACCGACAGGGTCTTTGACTAG

**Which set of species are more closely related 1&2 or 3&4?**

**Species 1&2 are more closely related than 3&4**



Turns out that the question on the last slide is very tricky and in reality systematists would rely on the help the of statistical tools and computers to draw the following...



Looking at the third image on the left one would likely conclude these species are very different due to many differences in their sequence, but the computer has determined that they actually share 11/13 bases and are therefore very similar



Turns out that the question on the last slide is very tricky and in reality systematists would rely on the help the of statistical tools and computers to draw the following...

## Molecular Homoplasy

**3**

A C G G A T A G T C C A C T A G G C A C T A

**4**

T C A C C G A C A G G T C T T T G A C T A G

Looking at the data above one would likely conclude these species are more similar than they are.

The similarities here are actually due to coincidence alone



# PHYLOGENY

## III. Main Idea:

The method used to infer phylogeny is called cladistics, an approach that uses common ancestry as its primary criterion for classifying organisms.





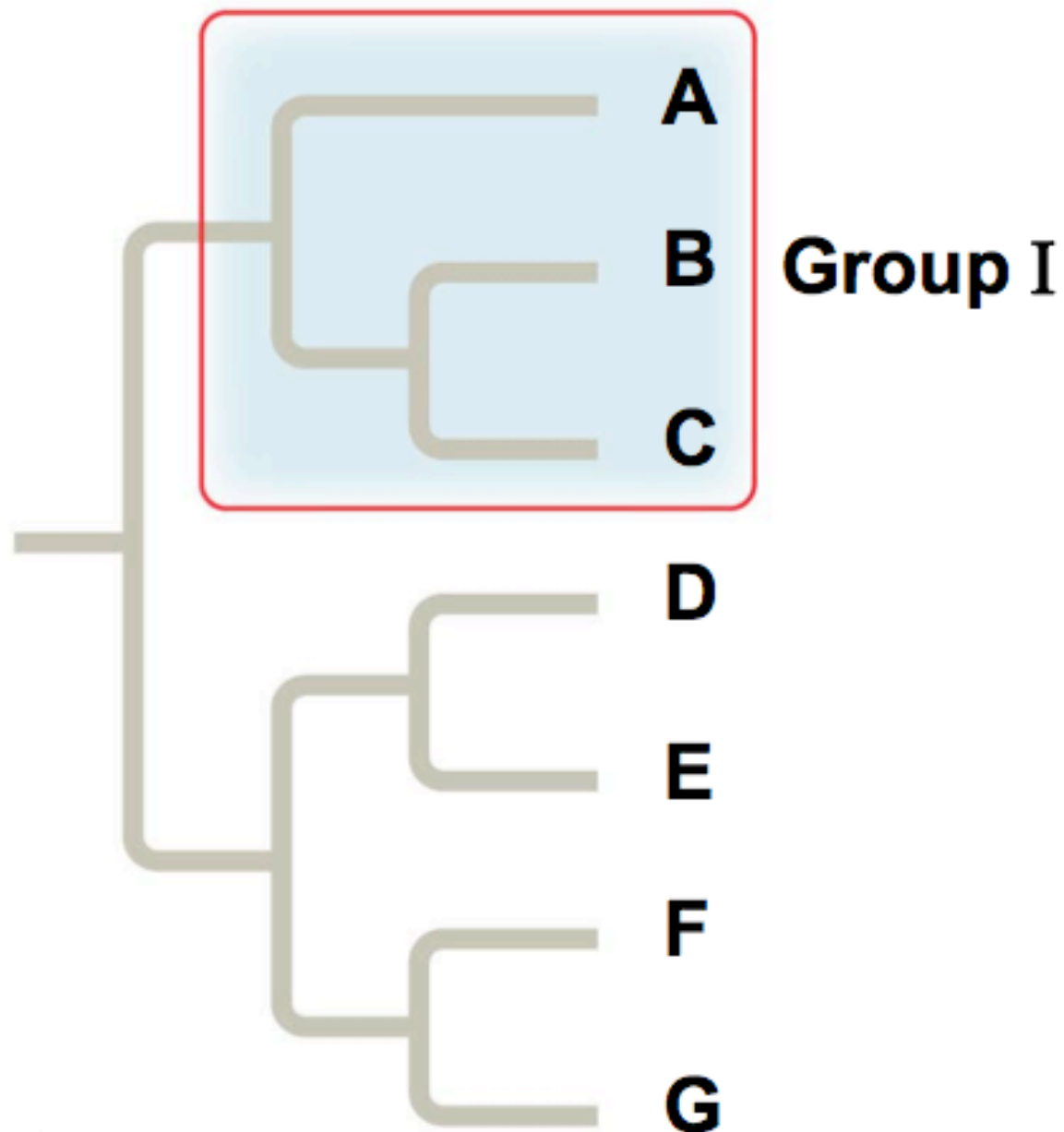
# Cladistics

- ❖ One approach to classify organisms relies on a system called **cladistics** where organisms are primarily classified together by common ancestry.
- ❖ In this approach systematists attempt to place species into groups called **clades**, each clade includes the common ancestor and all its descendants.
- ❖ Like taxonomic groups clades are nested within larger clades
- ❖ There are 3 types of clades.
  - ❖ **monophyletic, paraphyletic, polyphyletic**



# Types of Clades

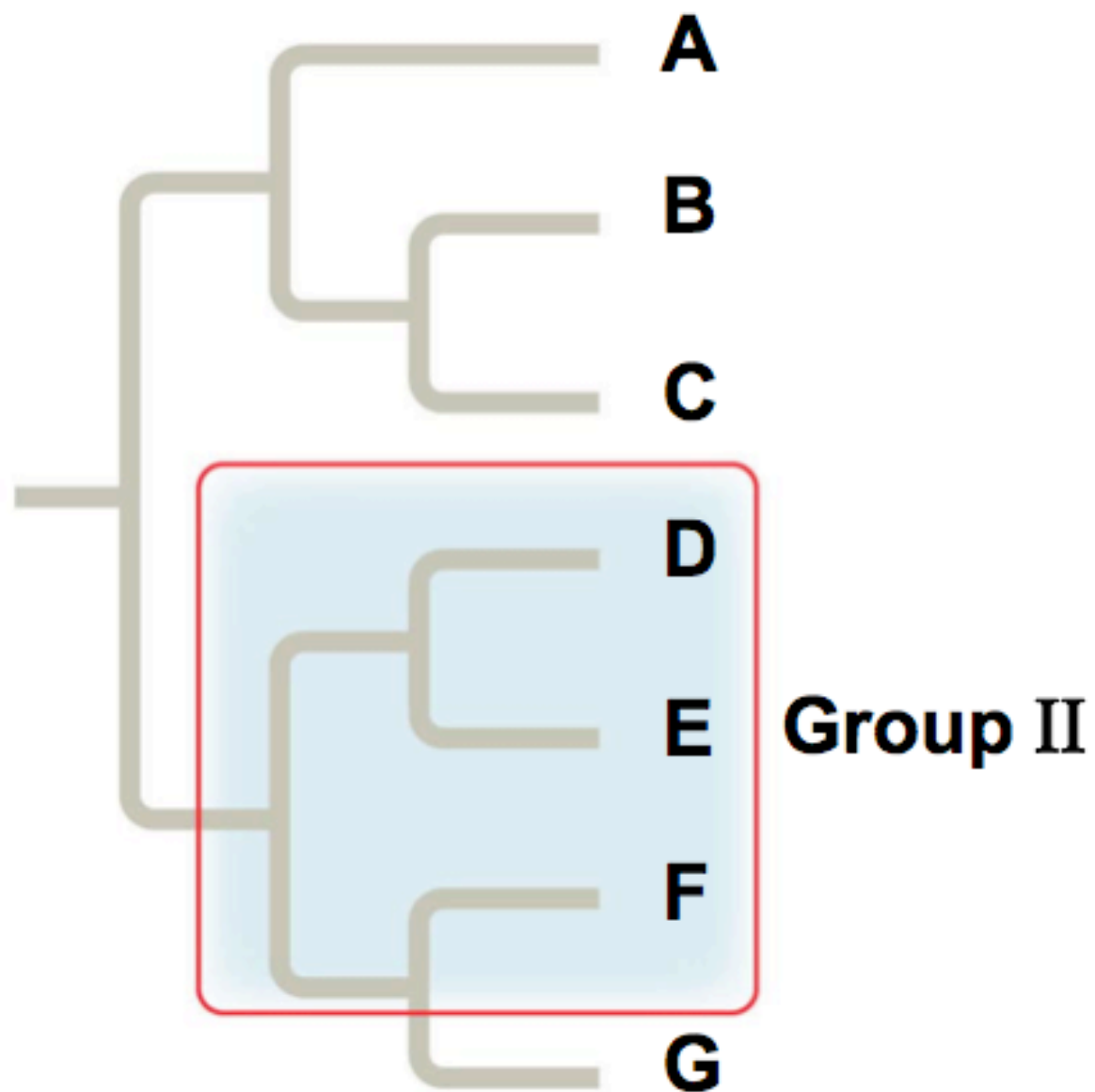
## (a) Monophyletic group (clade)



A valid clade is **monophyletic**, signifying that it consists of the ancestor species and all its descendants

# Types of Clades

## (b) Paraphyletic group

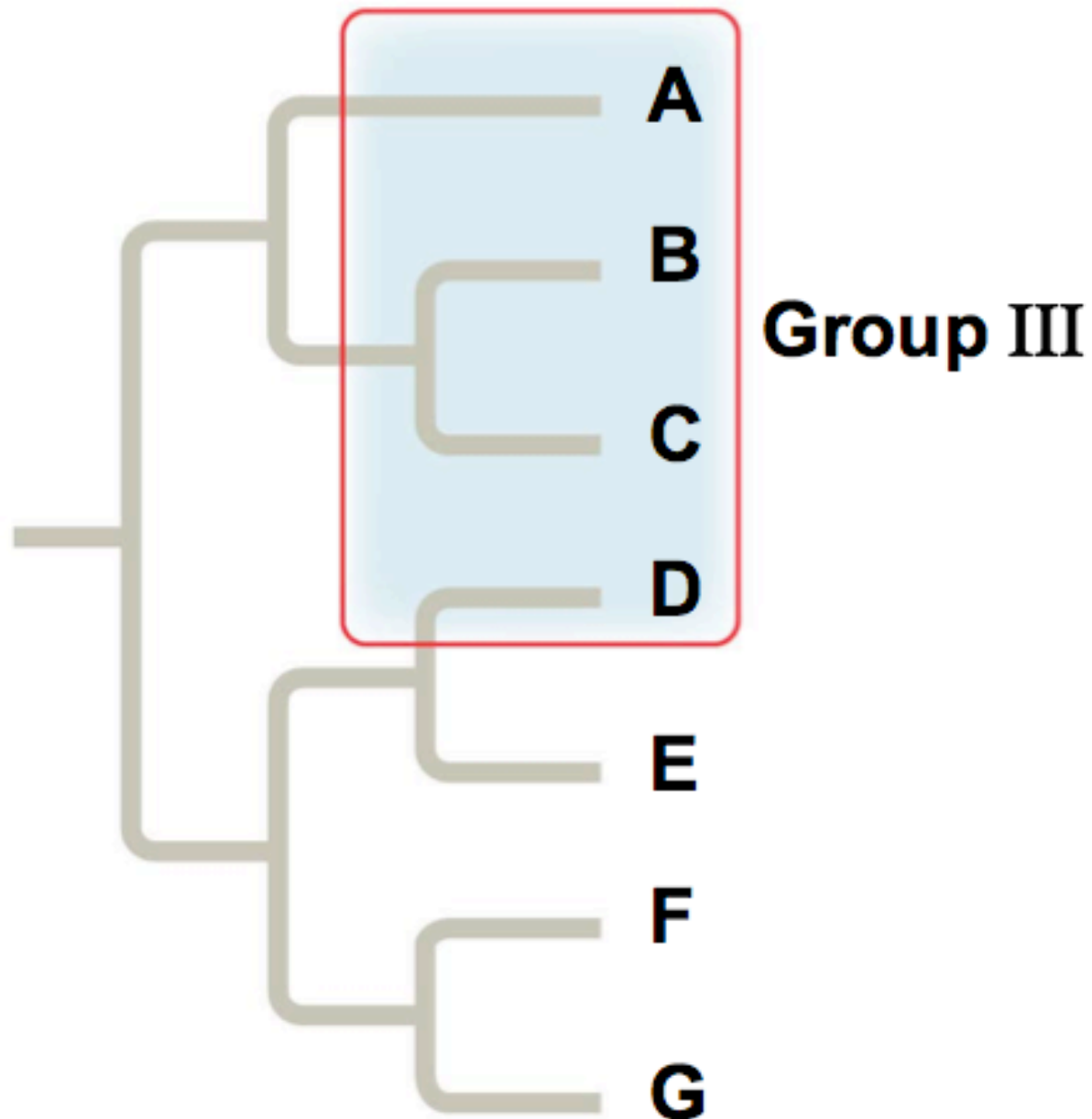


A **paraphyletic** grouping consists of an ancestral species and some, but not all, of the descendants



# Types of Clades

## (c) Polyphyletic group



A **polyphyletic** grouping consists of various species with different ancestors

# Ancestral & Derived Characters

- ❖ As a result of evolution, organisms both share traits and have their own unique traits compared to their ancestors.
- ❖ **Shared Ancestral Character**, one that originates in the ancestor of the group, taxon or clade
- ❖ Backbone in mammals, shared by all mammals including its common ancestor
- ❖ thus backbones predate the branching of mammals from other vertebrates



# Ancestral & Derived Characters

- ❖ As a result of evolution, organisms both share traits and have their own unique traits compared to their ancestors.
- ❖ **Shared Derived Character**, a novel trait is unique to a group, taxon or clade
- ❖ Hair in mammals, shared by all mammals but not found in the common ancestor of mammals
- ❖ *Ancestral & Derived traits are relative...remember the backbone example... if you go back to a deeper branch point you could find this trait separating vertebrates from other animals thus it now becomes derived*

# Inferring Phylogenies using Derived Characters

- ❖ Shared derived characters are unique to particular clades.
- ❖ All characters arose at some point in the history of life,
- ❖ It should be possible to determine the clade in which a derived character first appeared
- ❖ Then using information about shared derived characters one can begin to infer evolutionary relationships, or construct a phylogeny



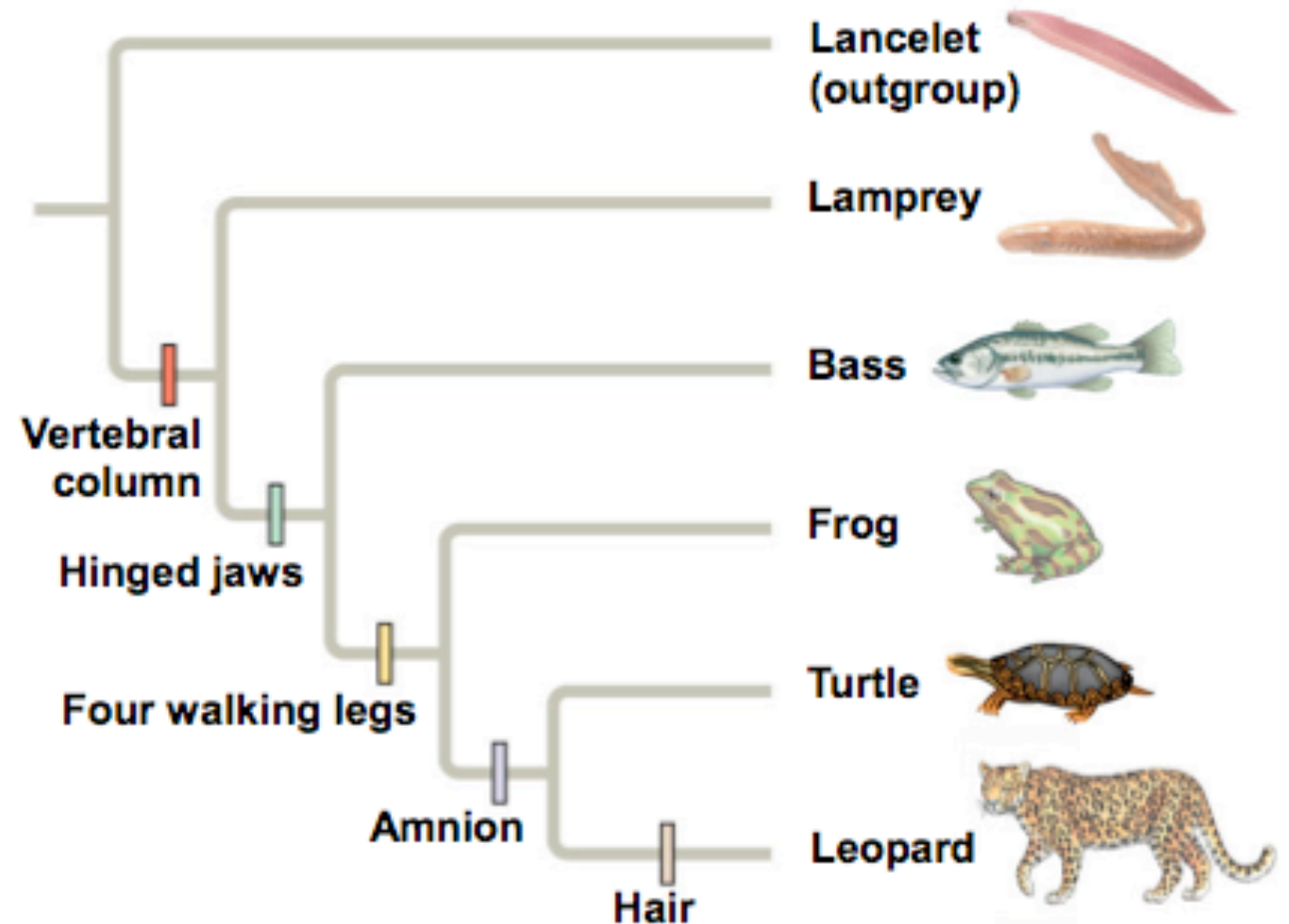
# Inferring Phylogenies using Derived Characters

- ❖ Start by selecting an outgroup as a basis of comparison.
- ❖ **Outgroup** is a species or group of species from lineage that is known to have diverged before the **Ingroup**, the species that are being studied.
- ❖ A suitable outgroup can be determined by morphology, paleontology, embryology or gene sequences.
- ❖ Comparing members of the ingroup with each other and the outgroup, we can deduce which characters were derived at various branch points

# Inferring Phylogenies using Derived Characters

CHARACTERS	TAXA					
	Lancelet (outgroup)	Lamprey	Bass	Frog	Turtle	Leopard
Vertebral column (backbone)	0	1	1	1	1	1
Hinged jaws	0	0	1	1	1	1
Four walking legs	0	0	0	1	1	1
Amnion	0	0	0	0	1	1
Hair	0	0	0	0	0	1

(a) Character table



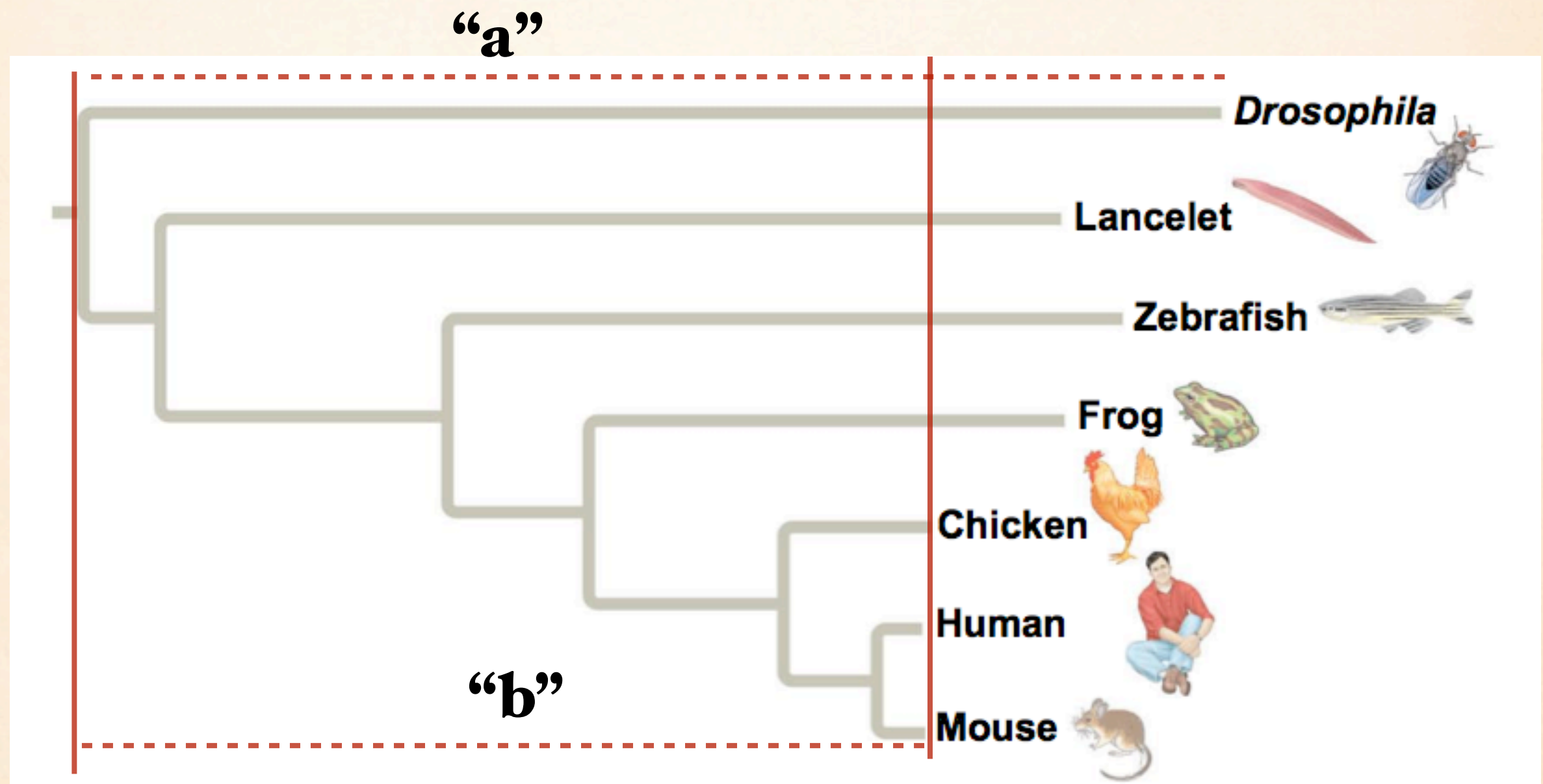
(b) Phylogenetic tree



# Phylogenetic Trees with Proportional Branch Lengths

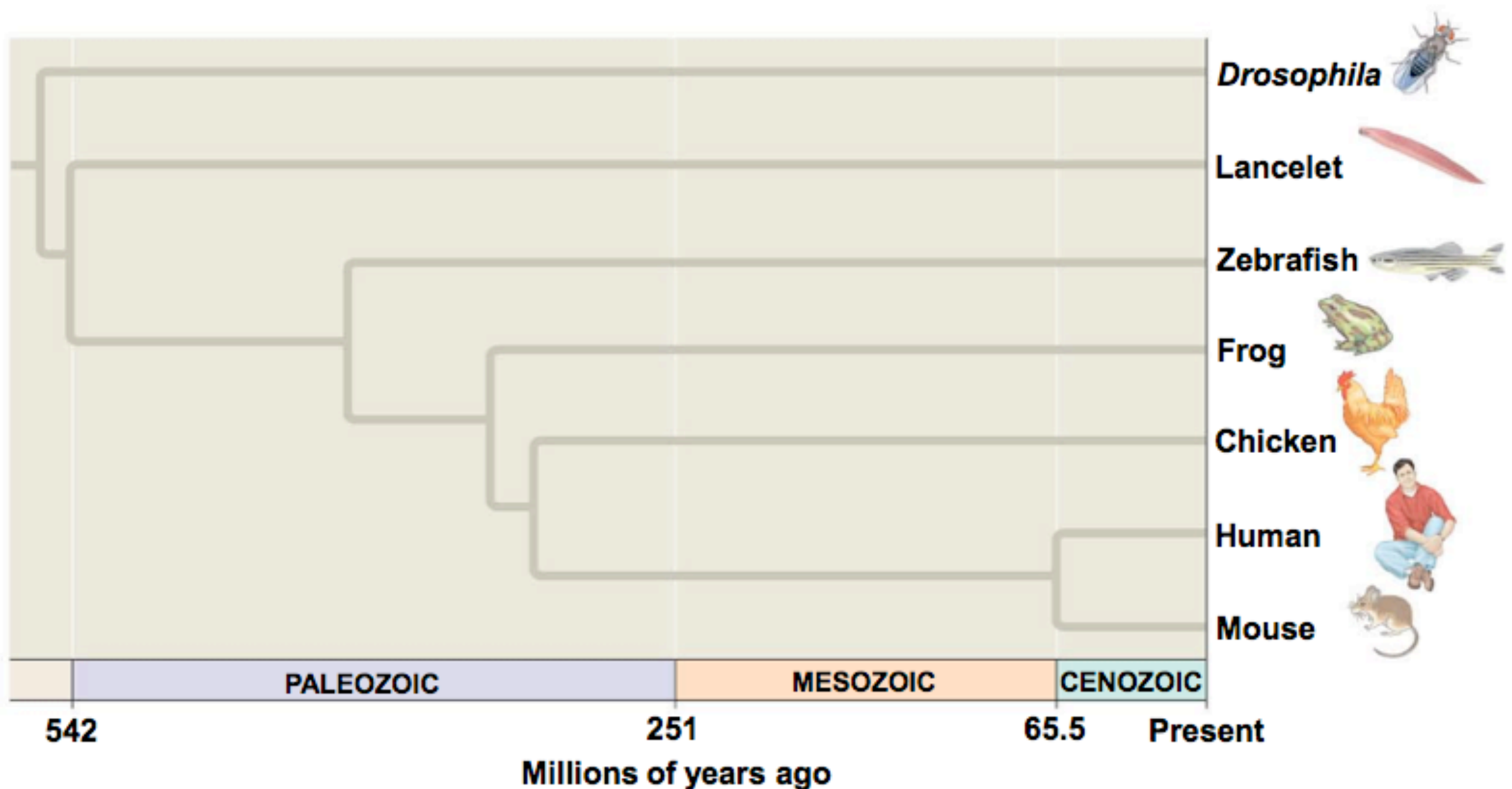
- ❖ The phylogenetic trees presented thus far and most that you will likely encounter are “relative”, the branches do NOT indicate degrees of change.
- ❖ They are also “relative” in time, the branching pattern does NOT reflect absolute time but rather relative time in the sense that one event happened earlier than another.
- ❖ However, some phylogenetic trees are built to show degrees of change, absolute time or both.

- ❖ Line “a” is longer than line “b” this implies that more genetic changes occurred in the fruit fly lineage than in the mouse lineage from the time they diverged from a common ancestor to present.





- ❖ This tree was developed using fossil data, its branch lengths are proportional to absolute time, all lineages have diverged from the common ancestor for equal amounts of time.



# Maximum Parsimony & Maximum Likelihood

- ❖ A growing database of gene sequences allows systematists to better develop phylogenies however the wealth of data are also makes it more difficult.
- ❖ In studying just 50 species there are  $3 \times 10^{76}$  different ways to arrange the trees, Which of these trees is the correct one?
- ❖ Science can never know for certain, but they can get close by narrowing the possibilities by applying the principles of maximum parsimony and maximum likelihood.



# Maximum Parsimony

- ❖ The principle of **maximum parsimony**, states that one should first investigate the simplest explanation that is consistent with the facts, also called “Occam’s Razor”
- ❖ William Occam was a 14th century English philosopher who advocated this minimalist problem solving technique
- ❖ If you build a tree use morphological data, the most parsimonious tree requires the fewest evolutionary events.
- ❖ If you build a tree use genetic data, the most parsimonious tree requires the fewest base changes.

# Maximum Likelihood

- ❖ The principle of **maximum likelihood**, states that given certain probability rules about how DNA sequences change over time, a tree can be found that reflects the most likely sequence of evolutionary events.
- ❖ Maximum likelihood methods can get complex, not suprisingly computers can help crunch data in the most likely and parsimonious ways.
- ❖ The next slides offer insight into a simple case where we might illustrate these methods.

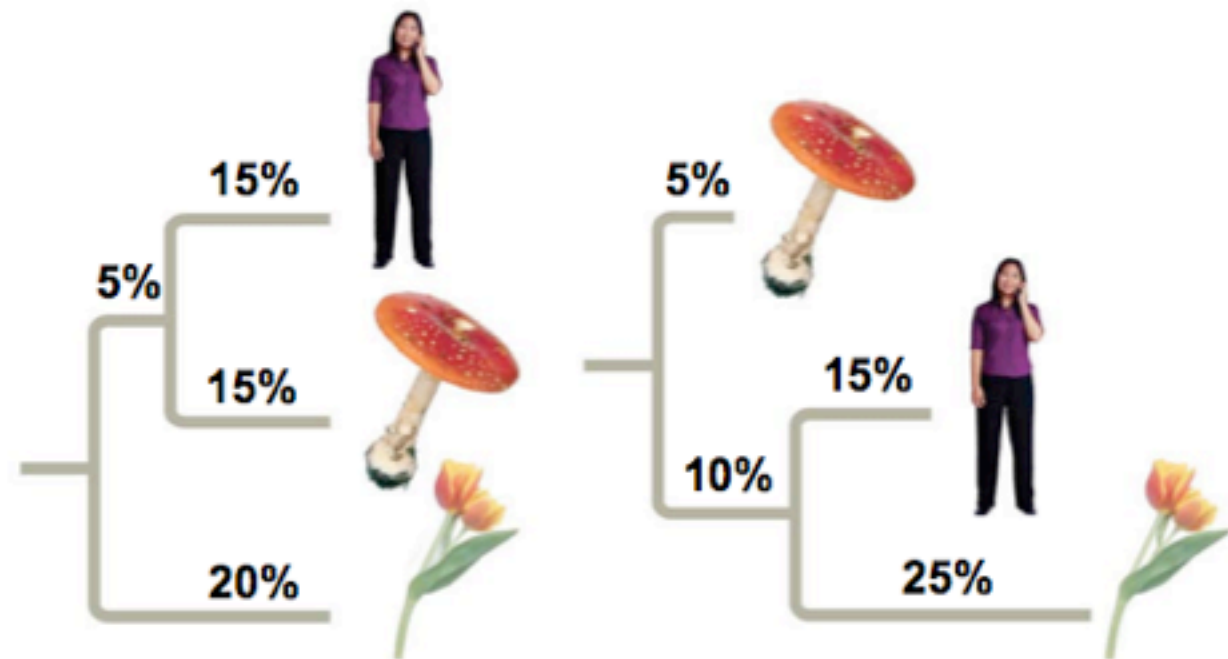


# Maximum Likelihood

Assuming that changes in DNA occur at the same rate tree 1 is more likely, for tree 2 to be right the rate of evolution in the mushroom lineage would have to slow while the rate in the tulip lineage would have to speed up

	Human	Mushroom	Tulip
Human	0	30%	40%
Mushroom		0	40%
Tulip			0

(a) Percentage differences between sequences



(b) Comparison of possible trees

## APPLICATION

In considering possible phylogenies for a group of species, systematists compare molecular data for the species. An efficient way to begin is by identifying the most parsimonious hypothesis-the one that requires the fewest evolutionary events (molecular changes) to have occurred

### TECHNIQUE



**Species I**



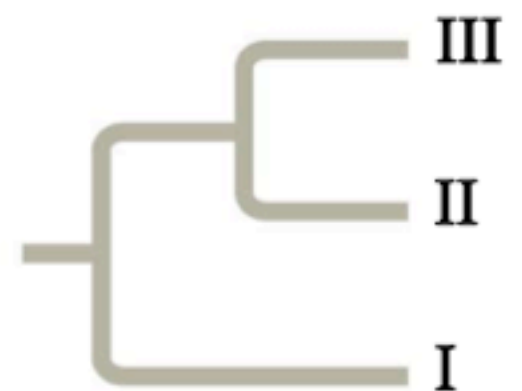
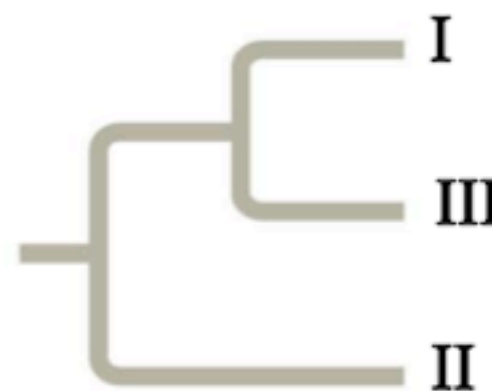
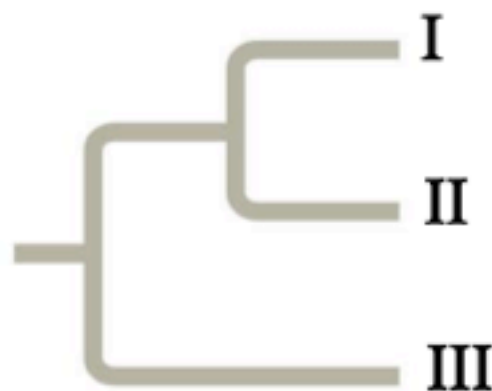
**Species II**



**Species III**

1

#### Three phylogenetic hypotheses:



Draw the three possible phylogenies for the species.

In this case with species there are only 3 possibilities



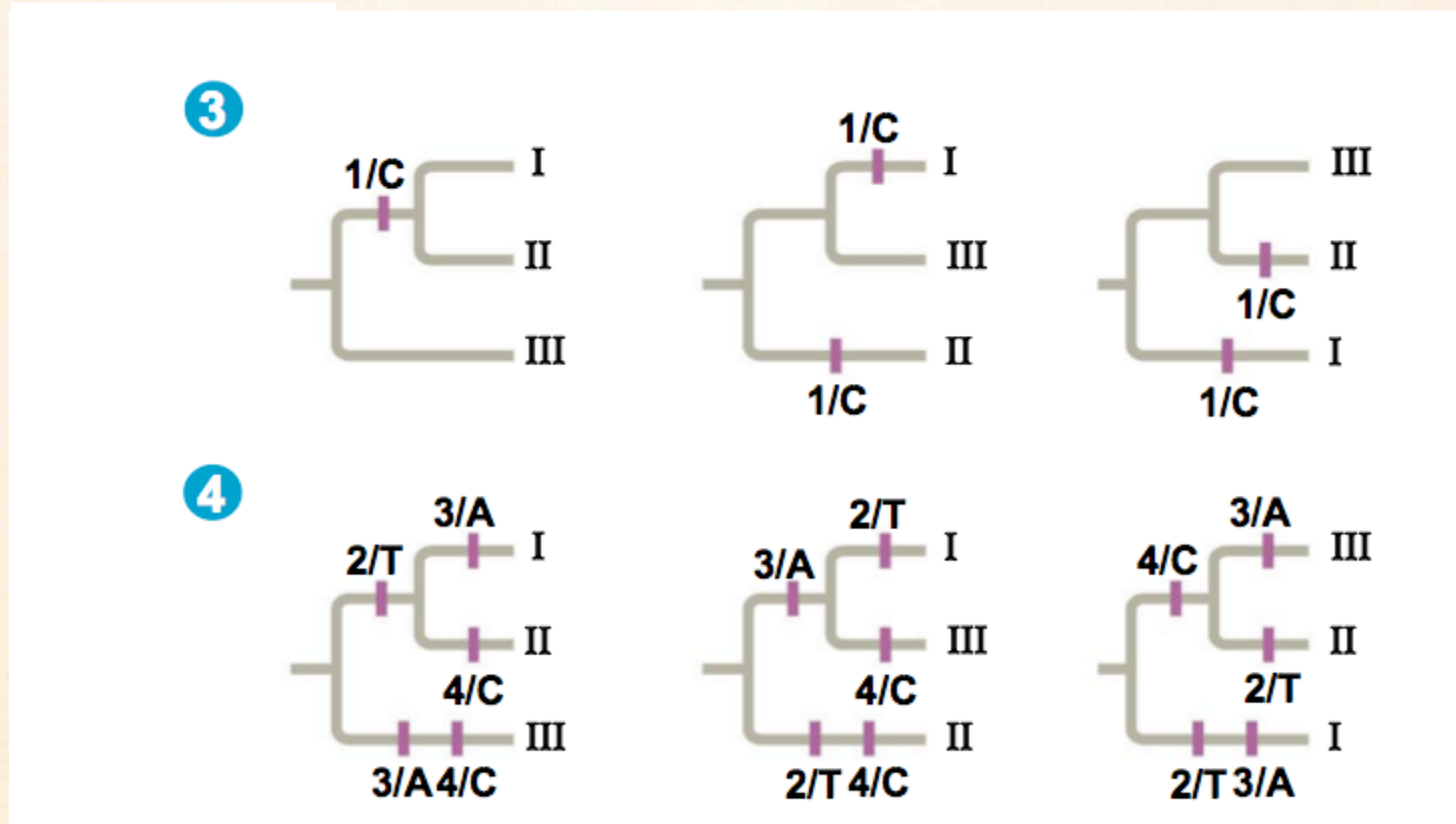
Tabulate the molecular data for the species. In this simplified example, the data represent a DNA sequence consisting of just four nucleotide bases. Data from several outgroup species (not shown) were used to infer the ancestral DNA sequence.

## TECHNIQUE

2

	Site			
	1	2	3	4
Species I	C	T	A	T
Species II	C	T	T	C
Species III	A	G	A	C
Ancestral sequence	A	G	T	T

Now focus on site I in the DNA sequence. In the tree on the left, a single base change event, represented by the purple hatchmark on the branch leading to species I and II (and labeled 1/c, indicating a change at site 1 to nucleotide c), is sufficient to account for the site 1 data. In the other two trees, two base-change events are needed.

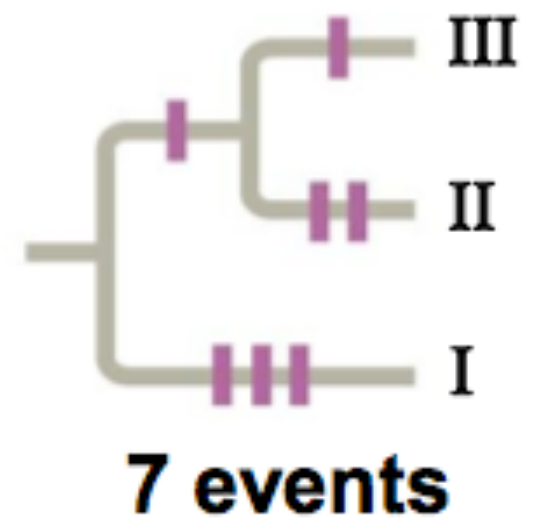
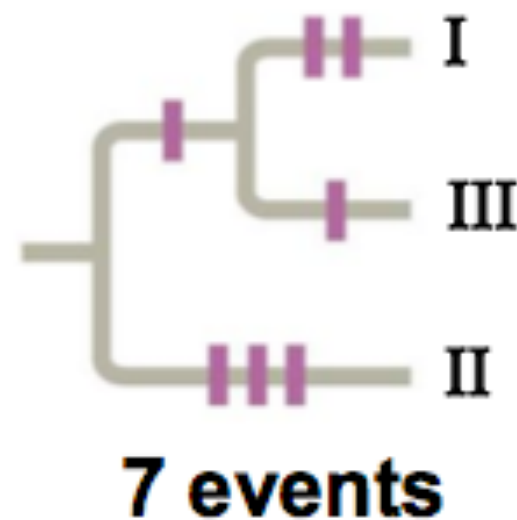
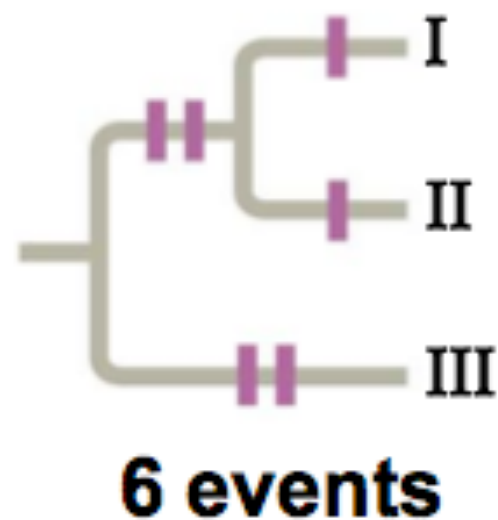


Continuing the comparisons of bases at sites 2, 3 and 4 reveals that each of the three trees requires a total of five additional base-change events (purple hatchmarks)



# RESULTS

To identify the most parsimonious tree, we total all the base-change events noted in steps 3 and 4. We conclude the first tree is the most parsimonious tree of the three possible phylogenies. (In a real example, many more sites would be analyzed. Hence, the trees would often differ by more than one base-change event.)

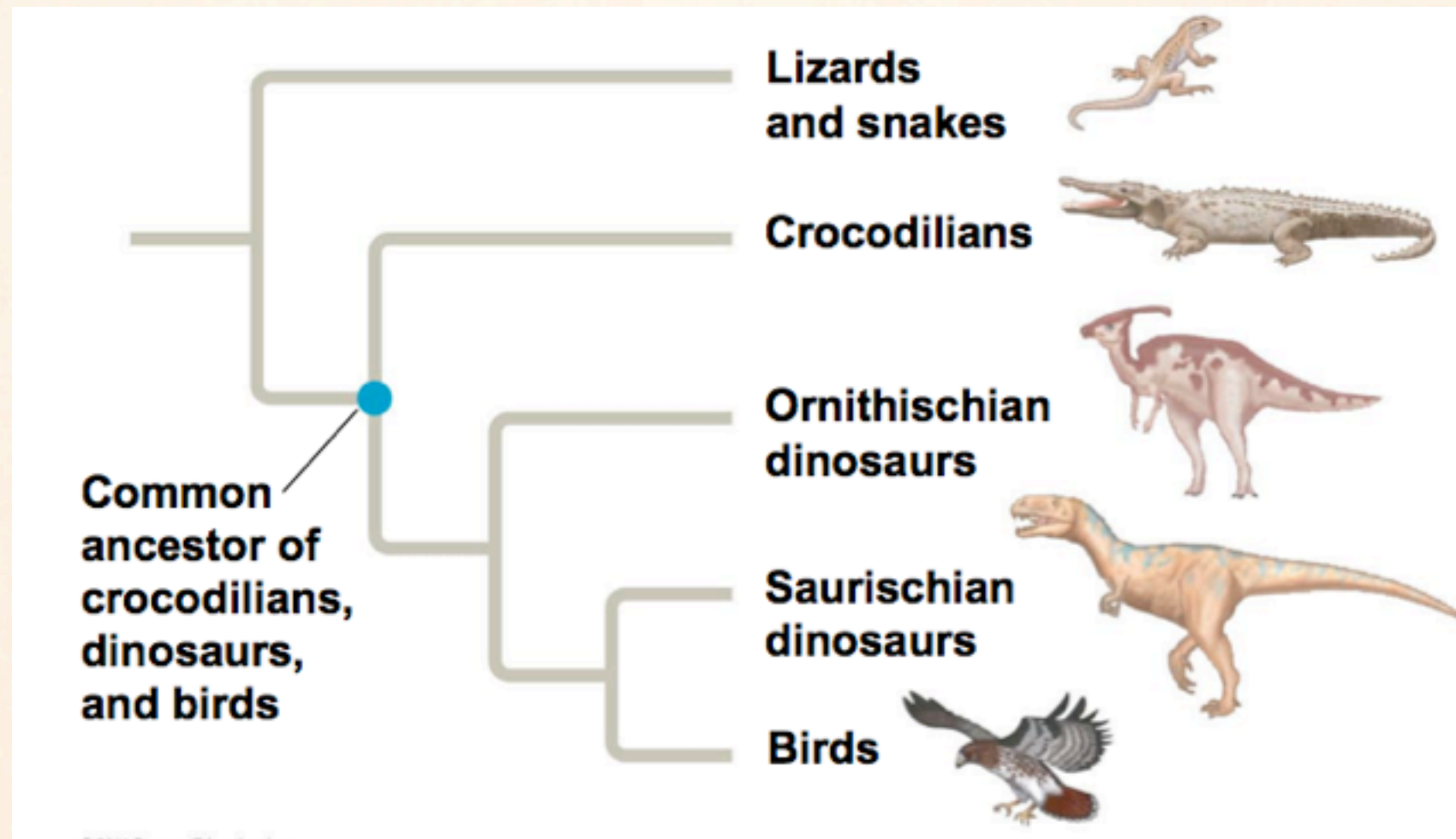


# Phylogenetic Trees as Hypotheses

- ❖ Remember phylogenies are possible explanations, they may change or be modified with new and compelling evidence.
- ❖ Scientists can make and test predictions by using and assuming that a phylogeny is correct
- ❖ For example we can compare extant crocodiles and birds, any similarities in structure or behavior is assumed present in their common ancestor, in this same way we might make certain predictions about the structure and behavior of dinosaurs



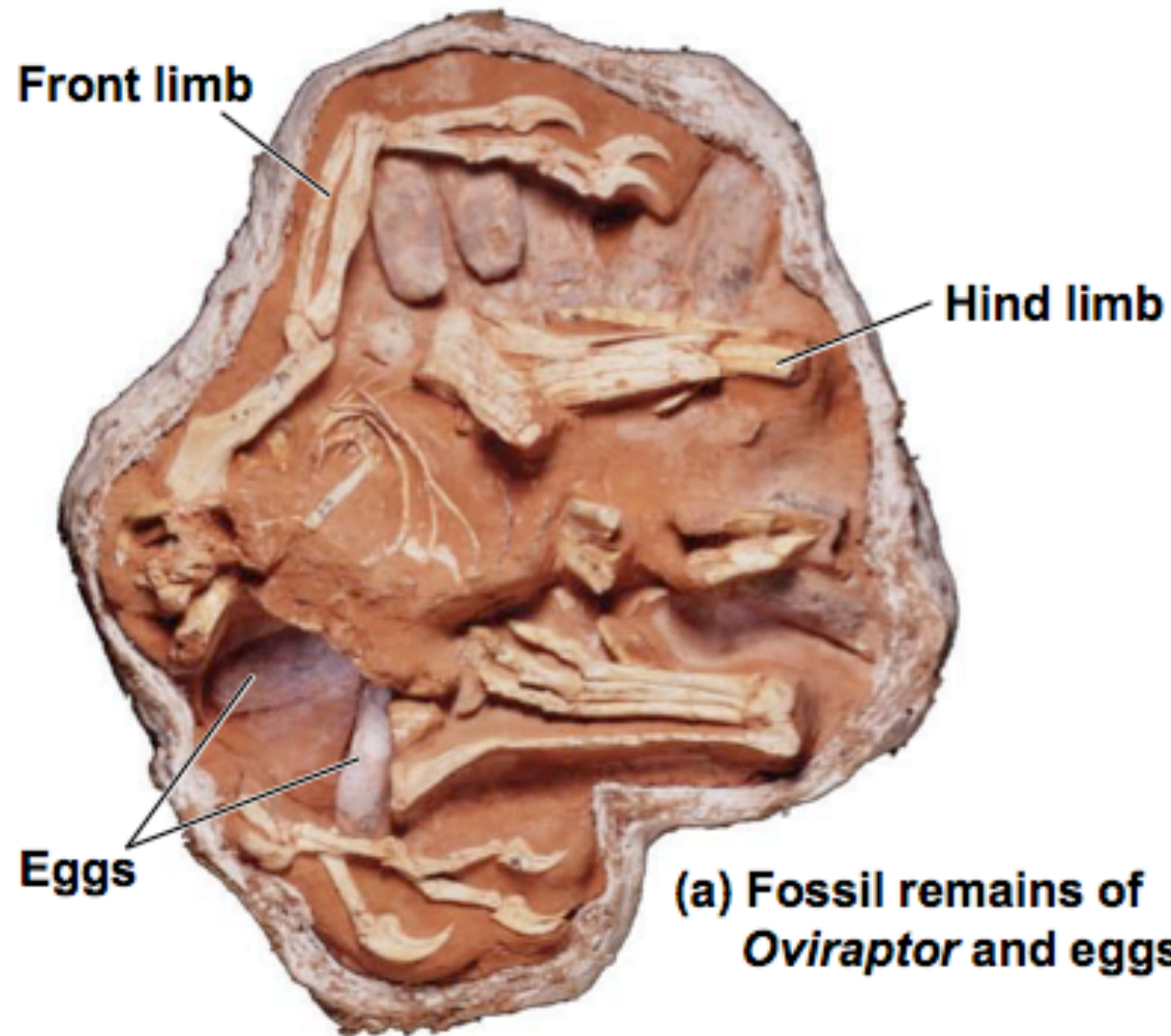
- Birds and crocodiles share several features: four-chambered hearts, song, nest building, and brooding



- These characteristics likely evolved in a common ancestor and were shared by all of its descendants, including dinosaurs



- The fossil record supports nest building and brooding in dinosaurs



(a) Fossil remains of *Oviraptor* and eggs



(b) Artist's reconstruction of the dinosaur's posture based on the fossil findings



# PHYLOGENY

## IV. Main Idea:

Genomes hold clues to understanding evolutionary history and the relatedness of different species.





# Exploring Genomes

- ❖ Genomic data can help us compare organisms that have little morphology in common.
- ❖ Genomic data can help us reconstruct phylogenies of organisms with a poor fossil record
- ❖ Genomic data can help us discovered the “big branches” on the tree of life that occurred long ago.
- ❖ Genomic data can also help us understand the “smallest twigs” on the tree of life that have recently



# Exploring Genomes

- ❖ Different genes evolve at different rates.
- ❖ The DNA that codes for rRNA mutates very slowly as a result studying these sequences are useful in deducing evolutionary relationships of millions of years ago.
- ❖ Mitochondrial DNA on the other hand has a much higher rate of mutation and is thus useful in deducing recent evolutionary events on the order of thousands of years ago.

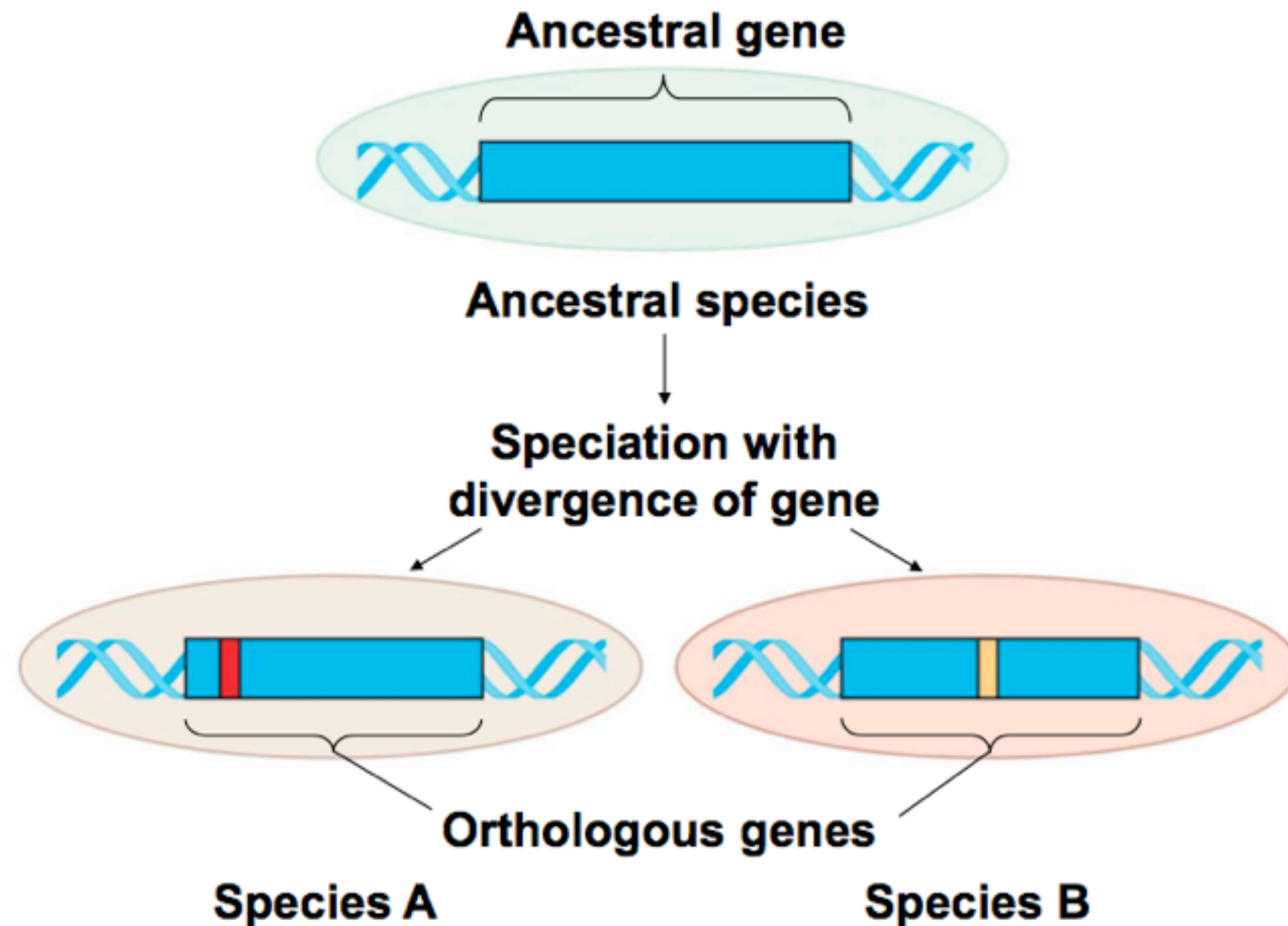
# Gene Duplication & Gene Families

- ❖ Gene duplication increases the number of genes in the genome, providing more opportunities for evolutionary changes
- ❖ Repeated gene duplications resulting gene families
  - ❖ Gene families- groups of related genes within a genome
- ❖ Like homologous genes, duplicated genes can be traced to a common ancestor
- ❖ There are two types of duplicated genes, **orthologous** and **paralogous**



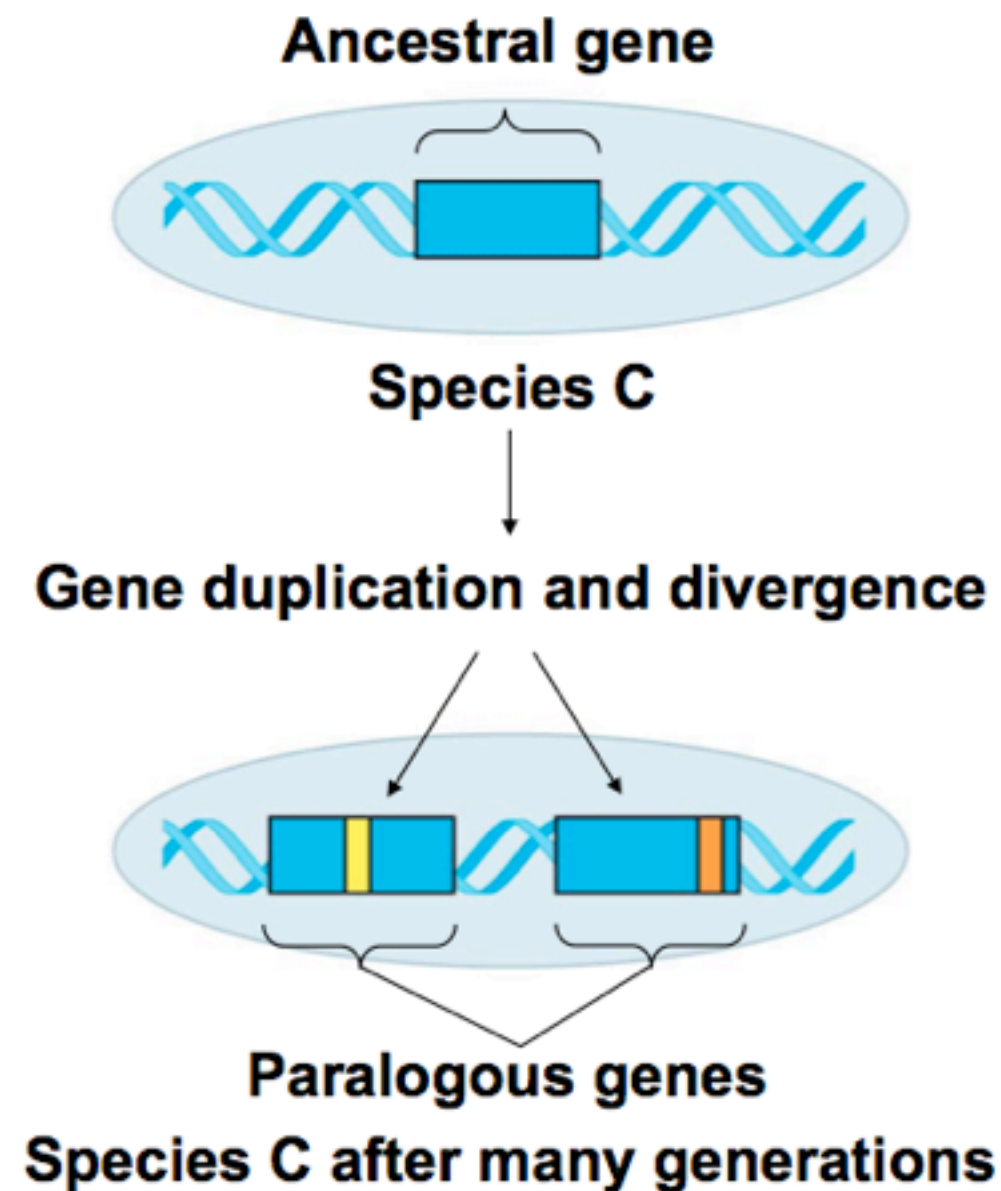
- **Orthologous genes** are found in a single copy in the genome and are homologous between species
- They can diverge only after speciation occurs

**Formation of orthologous genes:  
a product of speciation**



- **Paralogous genes** result from gene duplication, so are found in more than one copy in the genome
- They can diverge within the clade that carries them and often evolve new functions

**Formation of paralogous genes:  
within a species**





# Genome Evolution

- ❖ After comparing entire genomes of many different organisms two patterns have emerged.
- ❖ First, lineages that diverged long ago can share orthologous genes.
- ❖ Humans and mice diverged 65 mya and 99% of their genes are orthologous, and humans and yeast diverged 1 bya and 50% of their genes are orthologous.
- ❖ Such commonalities explain why such different organisms share so many biochemical and developmental pathways.

# Genome Evolution

- ❖ Second, the number of genes a species has does not seem to increase through duplication at the same rate as perceived phenotypic complexity.
- ❖ For example, humans only have 4x more genes than yeast despite being enormously more complex
- ❖ We are learning that human genes are more versatile and can encode multiple proteins that perform different tasks in different tissues



# PHYLOGENY

## V. Main Idea:

Molecular clocks will help evolutionary biologists understand relationships among all organisms, even those with no fossil record.





# Molecular Clocks

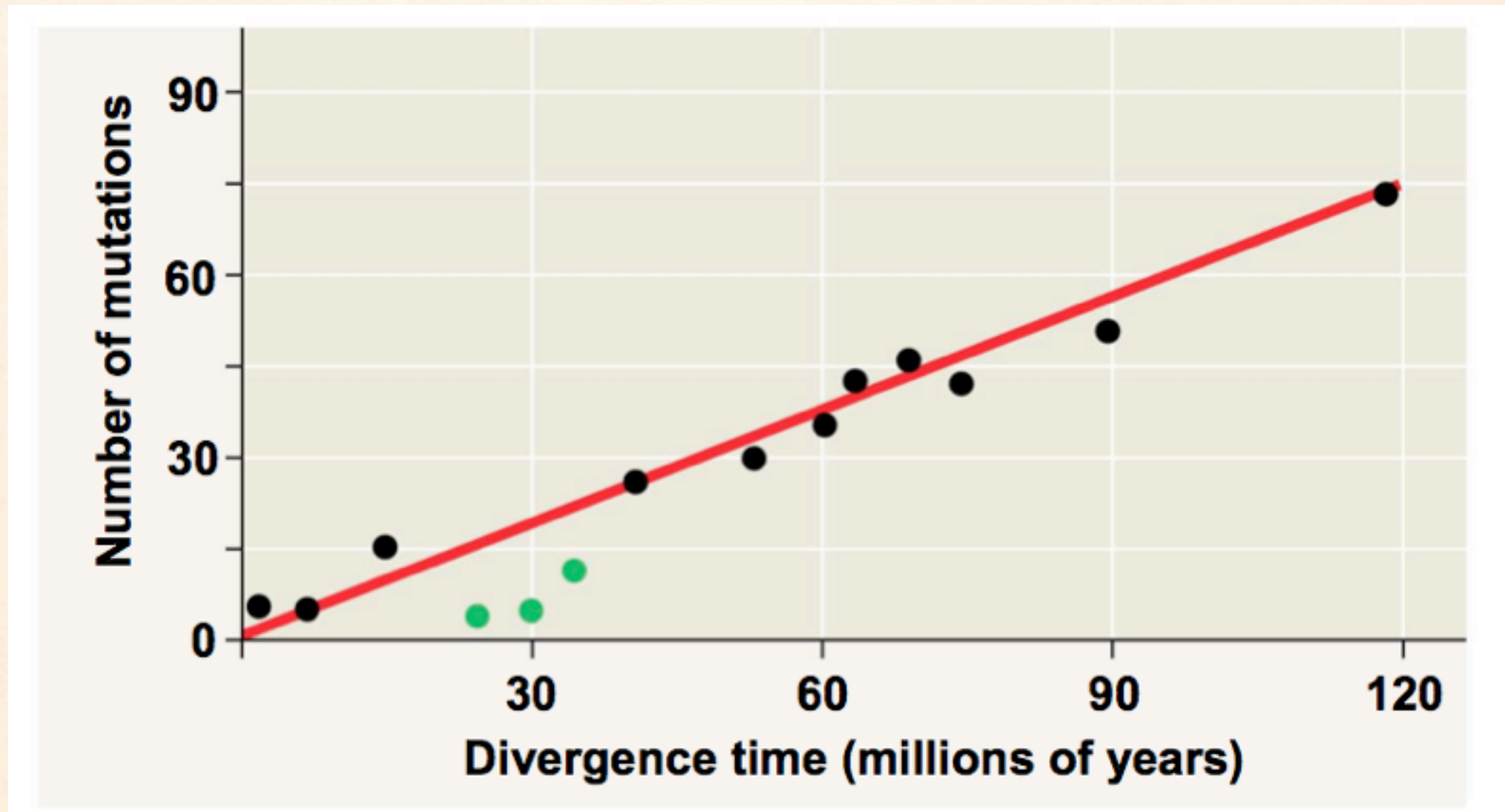
- ❖ **Molecular clocks** are “yardsticks” for measuring absolute time of evolutionary change based on the observation that some genes and other regions of genomes appear to evolve at constant rates.
- ❖ *The assumption is that the number of nucleotide substitutions in orthologous genes is proportional to the time that has elapsed since the species branched from their common ancestor.*
- ❖ *In paralogous genes nucleotide substitutions are proportional to the time since the genes became duplicated*



# Molecular Clocks

- ❖ We can calibrate the molecular clock of a gene that has a reliable rate of evolution by graphing the number of genetic differences (nucleotide, codon or amino acids) against dates of evolutionary branch points that are known from the fossil record.
- ❖ The average rates of change can be inferred from such graphs and can be used to estimate the dates of events that cannot be discerned from the fossil record, such as the origin of Silverswords on Hawaii.

Accumulated mutations in seven proteins has increased over time in a consistent manner for most mammals. The green dots represent primates, whose proteins appear to have evolved more slowly than those of other mammals. The divergence time for data point was based upon fossil evidence.



*Of course, no gene marks time with absolute precision. Molecular clocks are only accurate in a statistical sense of showing a fairly smooth average rate of change*



# Neutral Theory

- ❖ **Neutral theory** states that much evolutionary change in genes and proteins has no effect on fitness and is not influenced by natural selection
- ❖ The harmful mutations are removed quickly and if the rest have no effect on fitness then the rate of molecular change in these genes and proteins should be regular like a clock.
- ❖ Differences in the clock rate for different genes are a function of how important the gene is.

# Neutral Theory

- ❖ If an exact amino acid sequence that a gene codes for is necessary for survival then most mutations will be harmful and few will be neutral.
- ❖ ...thus such genes would evolve slowly
- ❖ But if an exact amino acid sequence that a gene codes for is less critical for survival then most mutations will be neutral and few will be harmful.
- ❖ ...thus such genes would evolve more quickly



# Problems with Molecular Clocks

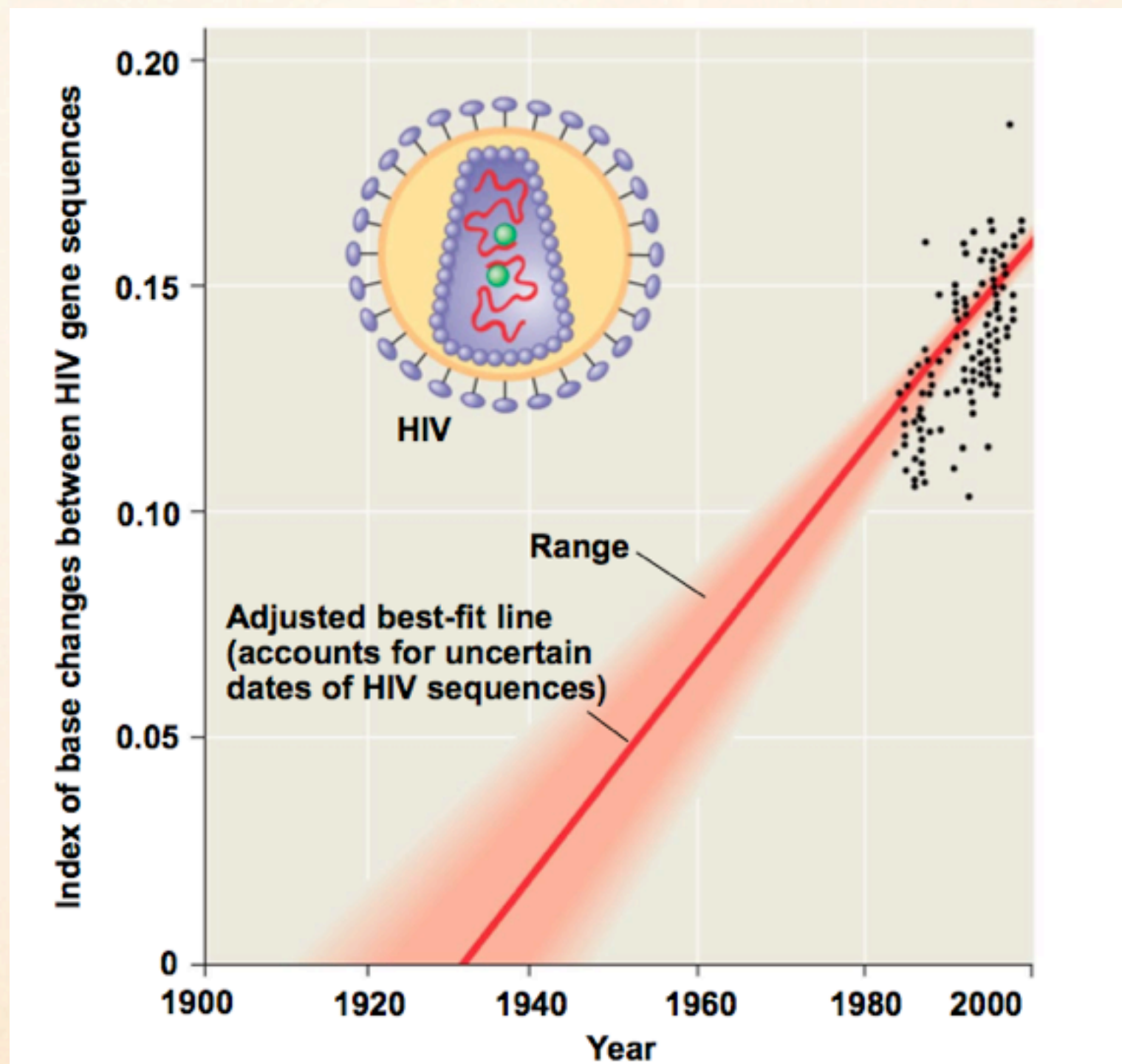
- ❖ The molecular clock does not run as smoothly as neutral theory predicts
- ❖ Irregularities result from natural selection in which some DNA changes are favored over others
- ❖ Estimates of evolutionary divergences older than the fossil record have a high degree of uncertainty
- ❖ The use of multiple genes may improve estimates

# Applying a Molecular Clock: HIV

- ❖ Phylogenetic analysis shows that HIVs descended from viruses that infect chimpanzees and other primates
- ❖ HIV spread to humans more than once
- ❖ Comparison of HIV samples shows that the virus evolved in a very clocklike way



- ❖ Application of a molecular clock to one strain of HIV suggests that that strain spread to humans during the 1930s



# PHYLOGENY

## VI. Main Idea:

New information continues to revise our understanding of the tree of life.





# From Two Kingdoms to Three Domains

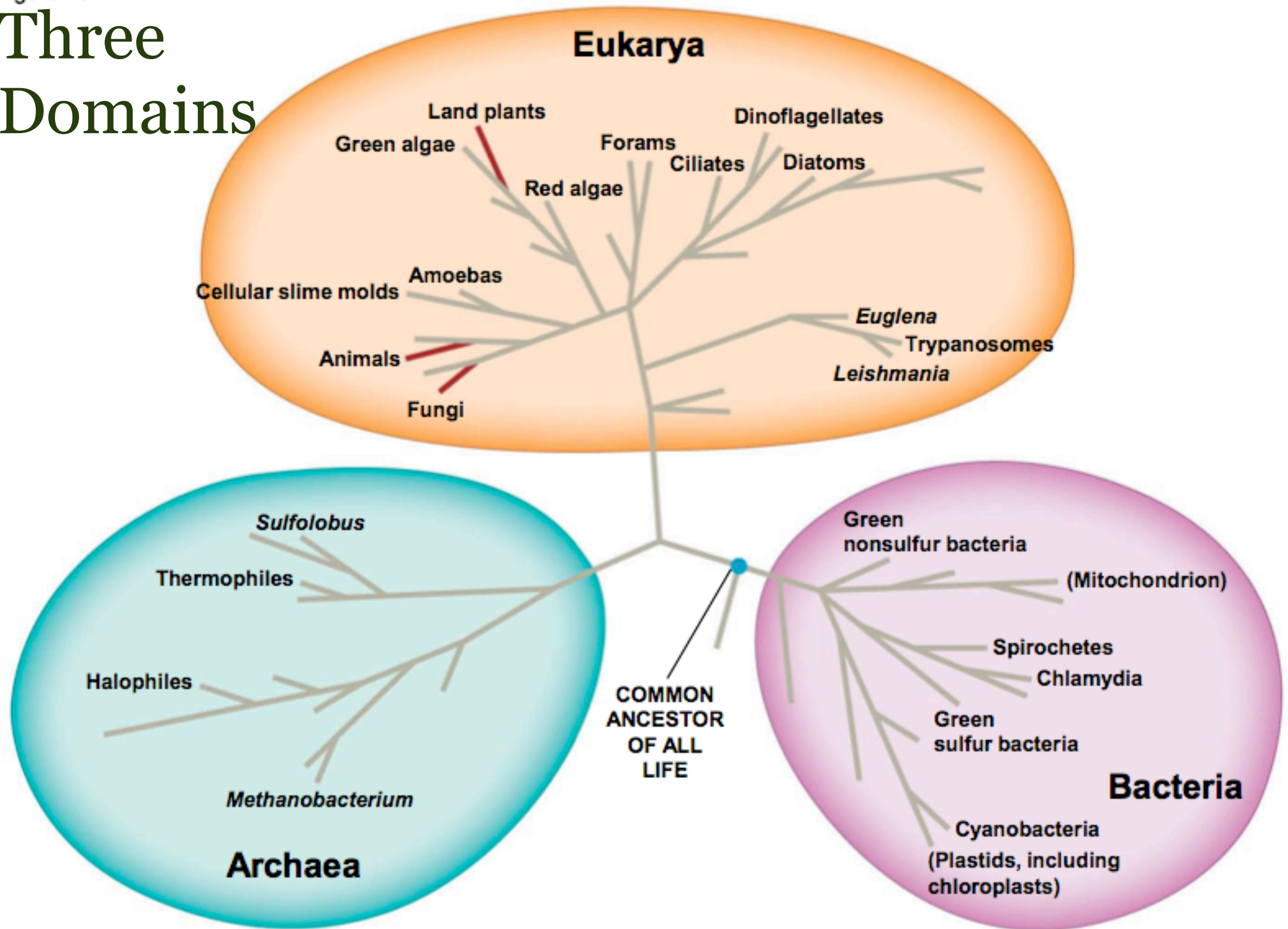
- ❖ Early taxonomists placed all life into one of two kingdoms: plants and animals!
- ❖ In the 1960's, with growing and gained acceptance, the five kingdom taxonomic scheme took hold.
- ❖ Monera, Protista, Fungi, Plants and Animals
- ❖ The scheme was highlighted by two fundamentally different cell types prokaryotic and eukaryotic cells

# From Two Kingdoms to Three Domains

- ❖ However, phylogenies based on genetic data soon revealed problem, some prokaryotes differed from each other as much as they differed from eukaryotic cells.
- ❖ Biologists responded by adopting three domains-  
**Bacteria, Archaea and Eukarya**
- ❖ ...a taxonomic level above kingdoms
- ❖ 4 kingdoms are still recognized, but Monera is now obsolete because it has members in two different domains (“protista is also now crumbling apart”)



# Three Domains



# A Simple Tree of Life

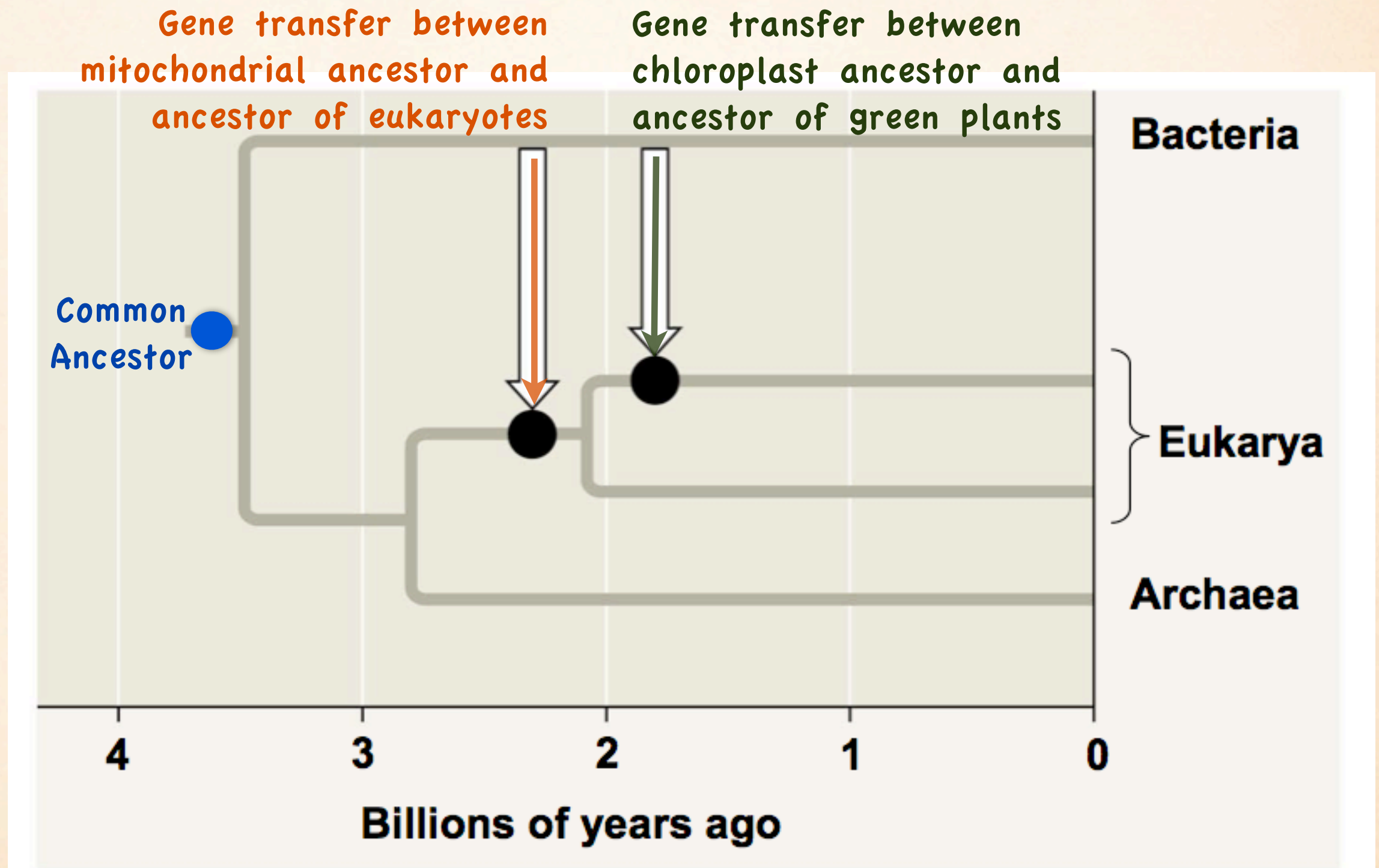
- ❖ The tree of life suggests that eukaryotes and archaea are more closely related to each other than to bacteria
- ❖ The tree of life is based largely on rRNA genes, as these have evolved slowly
- ❖ New evidence suggests there have been substantial interchanges of genes between organisms in different domains



# A Simple Tree of Life

- ❖ **Horizontal gene transfer** is the movement of genes from one genome to another
- ❖ Horizontal gene transfer occurs by exchange of transposable elements and plasmids, viral infection, and fusion of organisms
- ❖ Horizontal gene transfer complicates efforts to build a tree of life

# The role of horizontal gene transfer in the history of life.





# A Simple “Ring” of Life?

- ❖ Some researchers suggest that eukaryotes arose as an fusion between a bacterium and archaean
- ❖ If so, early evolutionary relationships might be better depicted by a ring of life instead of a tree of life

