

# Cycle 9: Evolutionary Relationships



## Introduction to Phylogeny

All organisms can be grouped into certain taxonomic groups to easily communicate biological information about various organisms. There are 8 taxonomic groupings which go from the most general to the most specific. These groupings are Domain, Kingdom, Phylum, Class, Order, Family, Genus, and Species (note: remember – Dear King Philip Came Over For Good Soup). For example, the classification of humans (scientific name: *Homo sapiens*) is:

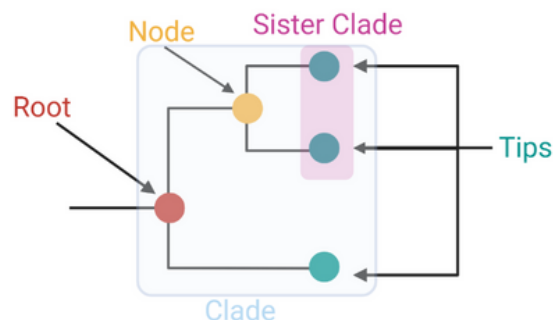
Domain	Kingdom	Phylum	Class	Order	Family	Genus	Species
Eukarya	Animalia	Chordata	Mammalia	Primates	Hominidae	Homo	sapiens

In order to describe lineages and evolutionary relationships, phylogenetic trees have been created. Phylogenetic trees are important tools to organize and communicate hypothesized evolutionary relationships among various groups of organisms. These trees are also very useful in answering questions about the molecular clock or the comparative method to test hypotheses about the evolution of traits. The molecular clock hypothesis states that differences in DNA sequence may index the time at which two species diverged. If there is a large difference it demonstrates that the divergence is in the distant past while a smaller difference means a more recent common ancestor. The comparative method compares anatomical and behavioural characteristics using phylogenetic trees.

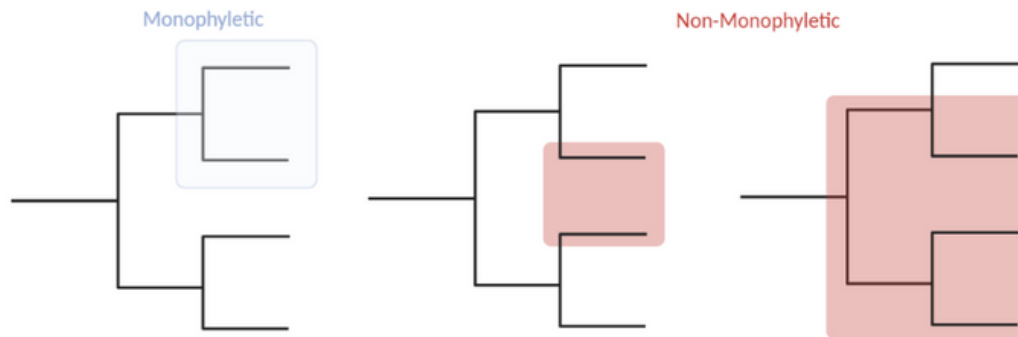
## Deciphering a Phylogenetic Tree

In the past, some historically recognized taxonomic groupings did not accurately correspond to evolutionary relationships. This was because they used to be classified strictly on how they looked and not by common ancestors. Thus, some groups looked alike and were historically grouped together but were never really related from an evolutionary standpoint. In the present, we create these trees using common ancestry and in some cases, molecular data (this will be touched upon later).

Some important vocabulary when deciphering and creating a phylogenetic tree are node, root, clade, and sister clade. A node is each branching point on a phylogenetic tree. A root is the oldest point on the tree, also known as the common ancestor of all species in that tree. A clade is a node with all the branches and tips. A sister clade refers to when two clades emerge from the same node.



There are also different groupings within a phylogenetic tree. A monophyletic taxon describes a grouping which consists of one clade, an ancestral species (with node), and all the descendants. Another type of grouping is the non-monophyletic taxon which includes polyphyletic taxon and paraphyletic taxon. A non-monophyletic taxon refers to a grouping where some but NOT all descendants from a common ancestor are included. A polyphyletic taxon describes a grouping which includes species from different clades, the common ancestor is not included, while a paraphyletic taxon refers to a grouping which includes the ancestor and only some of its descendants.



There are also phylogenetic trees that are rooted and others that are unrooted. An unrooted tree is a phylogenetic tree with no element of time (no root). If given an unrooted tree, we must root it by designating an organism to be the outgroup (organism that branched off the earliest and will become a reference point) which will then become the root.

An evolutionary tree is quite useful in determining whether organisms are relatively close or distant relatives. This is determined by tracing the lineage of the organisms to their most recent common ancestor. The organisms that share a more recent common ancestor (node) are more closely related and those who share a common ancestor (node) that are further apart are more distant relatives. Note: Remember organisms may share a common trait but that may not mean that they are more closely related (similarities do not always correspond to relatedness).

## Anagenesis and Cladogenesis

Anagenesis and cladogenesis are both two types of evolutionary mechanisms which can lead to speciation. Anagenesis is a phyletic change that slowly accumulates in lineage as the environment shifts over time (in other words, the gradual transformation of one species into another as the characteristics shift over time). This does not increase biodiversity as the number of species does not increase. On a phylogenetic tree, anagenesis is represented by a single straight line.

Cladogenesis describes an ancestral species which produces two descendant species. This will increase biodiversity as the number of species increases. On a phylogenetic tree, this is illustrated by a branching pattern (where two descendants arise from a single branch). To put it simply anagenesis is a type of phylogenetic evolution which occurs within a single species, while cladogenesis is a type of branching evolution that leads to more than one species.

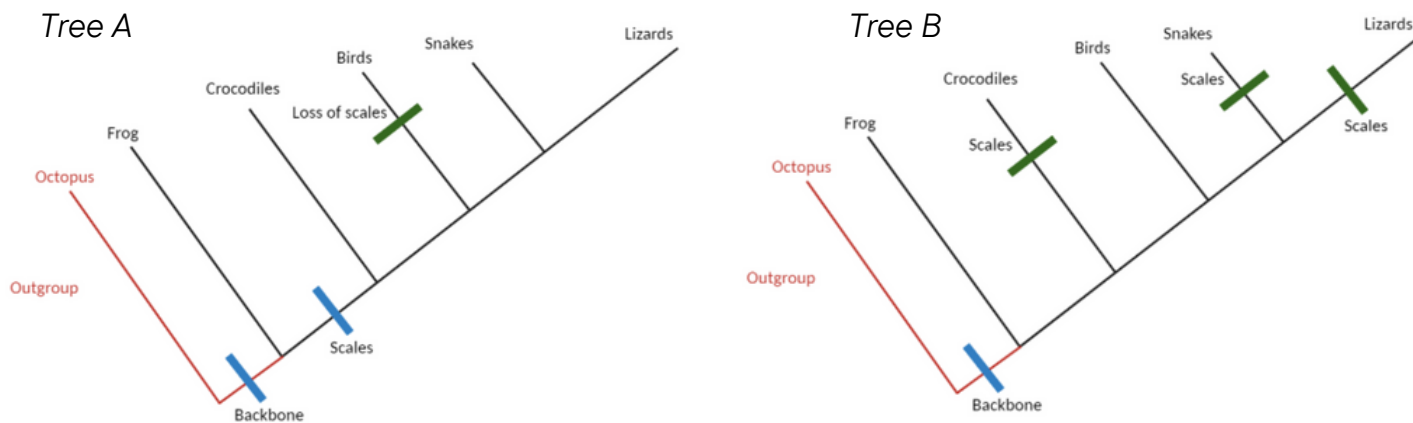
## Ancestral and Derived Character States

As organisms evolve, they develop new characteristics which are then passed onto future generations. Those organisms may then continue to evolve and become different species but still exhibit those same characteristics. In an organism, certain characteristics/traits can be ancestral character states or derived character states. Ancestral character states are traits that evolved before the most recent common ancestor of a group, while derived character states are traits that evolved after the most recent common ancestor of a group.

To determine if a trait is ancestral or derived, we can also use outgroup analysis. An outgroup is an organism related to the clade but not included which can serve as a reference to the ingroup. The ingroup is simply the clade under study. The outgroup analysis will make the tree rooted which can then be used to determine the character states.

Note: Not all members will have the ancestral traits as they can be lost as organisms evolve. Character states observed in the outgroup are considered ancestral, and those observed only in the ingroup are considered derived.

*Example 1: Determine the most accurate tree between tree A and tree B.*



In this tree (tree A), the crocodile, snakes, and lizards are the only organisms with scales, however, all organisms in the ingroup have a backbone. The backbone and scales are derived character states as they evolved after the most recent common ancestor. This means that having no backbone and no scales is the ancestral character state as those are the traits that evolved before the most recent common ancestor. Now that we know what ancestral and derived character states and how to construct an evolutionary tree based on those states, how do we know which evolutionary tree is more accurate? This is solved using parsimony, which states that the simplest plausible explanation of any phenomenon is the best. In other words, the phylogenetic tree with the smallest number of evolutionary changes is likely the most accurate.

From the example above, you may have noticed that there is another method of constructing the tree where only the crocodiles, snakes and lizards have scales. Using parsimony, we can determine the most accurate tree (recall: the one with the smallest number of evolutionary changes). We see that tree A has three evolutionary changes, while tree B has four evolutionary changes. Thus, tree A is the most accurate tree.

## Homology, Convergence, and Homoplasy

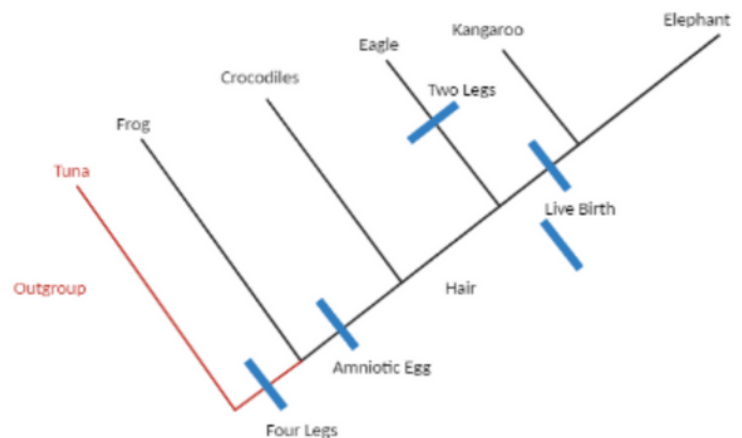
In constructing and deciphering evolutionary trees, it is also important to distinguish between homologous, homoplastic, or convergent traits. Homology refers to similar characteristics due to common ancestry, while homoplasy is characteristics shared across clades in a phylogeny that is not present in their common ancestor. Convergence is a subcategory of homoplasy and it refers to an independent evolution of similar features in different species. Given a phylogenetic tree, a trait is homologous if a trait between two sister clades is present in the common ancestor. However, if the trait is not present in a common ancestor, then it is homoplastic.

*Example 2: Build an evolutionary tree based on the following data*

	Number of Legs	Egg	Hair	Reproduction
Tuna - outgroup	0	Anamniotic Egg	N/A	Lay eggs
Elephant	4	Amniotic egg	Yes	Live Birth
Frog	4	Anamniotic Egg	N/A	Lay eggs
Crocodile	4	Amniotic egg	N/A	Lay eggs
Kangaroo	4	Amniotic egg	Yes	Live Birth
Eagle	2	Amniotic egg	Yes	Lay eggs

When building an evolutionary tree based on data, we must first identify the outgroup which is Tuna. Given that the tuna is the outgroup, we must find the ancestral and derived traits. For the number of legs, four and two are derived as the ancestral trait is zero legs. Amniotic eggs, hair, and live birth are also derived traits as they are not in the outgroup.

As the frog is the only other organism (other than the outgroup) to have an anamniotic egg they are most likely going to be next to each other in the evolutionary tree. The crocodile and frog are also the only other organisms to have no hair which means that it is likely next to the frog in the tree. Lastly, the elephant and kangaroo are the only ones with live birth and will thus be at the end of the evolutionary tree. This should be the final tree:



Note: Make sure to follow the principle of parsimony.

## Molecular Data and Constructing a Phylogeny

Molecular data, such as DNA sequences or amino acid sequences, can be used to build evolutionary trees by examining similarities and differences in the sequences. This method of constructing a tree presents its advantages and disadvantages.

Advantages:

- Provide abundant data which can be used to compare distantly related organisms that share no phenotypic characteristics and organisms with minor morphological differences
- The genome can be sampled, quantified, and analyzed directly and objectively

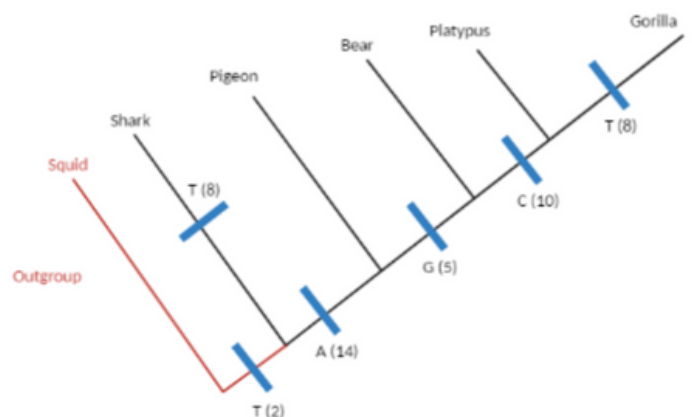
Disadvantages:

- There are a limited number of character states that exist in DNA and amino acids which can make it difficult to assess homology between species
- Researchers aren't aware of the functionality that the molecular differences present

*Example 3: Build an evolutionary tree based on the following molecular data.*

	Sequence
Squid - outgroup	GATTCATCGTCTTGC
Pigeon	GTTTCATCGTCTTAC
Platypus	GTTTGATCGCCTTAC
Bear	GTTTGATCGTCTTAC
Shark	GTTTCATTGTCTTGC
Gorilla	GTTTGATTGCCTTAC

The first thing we must do is align and notice the similarities and differences (outlined in red). As the shark is the only organism with the ancestral sequence of G(14) and not the derived sequence of A(14), then it will be next to the outgroup (squid). The organism following the shark will be the pigeon as it has the derived sequence of A(14), but it is the only other organism (other than the shark and squid) without the nucleotide G(5). Lastly, the bear does not have the nucleotide of C(10) which the gorilla and platypus possess and will thus be next to the pigeon on the tree. It's important to note that the nucleotide T(8) appeared in both shark and gorilla, yet they don't share the same most recent common ancestor. This is because T(10) is a result of homoplasy. The resulting tree should be:



*Disclaimer: We cannot guarantee that this resource will stand the test of time and therefore we are not responsible for any outdated information. This resource is student-made, and should be supplementary to resources provided by your instructors. It is not an alternative to your lectures and office hours. We are not responsible for the outcome of anyone's course evaluations based on this resource.*

*We will do our best to update this resource if there are any drastic changes. Please reach out to us at [team@webstraw.org](mailto:team@webstraw.org) if there are any issues with our current version and we will do our best to make changes promptly. We appreciate you using our resource!*