

## Chapter 26

# Phylogeny and the Tree of Life

### *Lecture Outline*

#### Overview: Investigating the Tree of Life

- Evolutionary biology is about both process and pattern.
  - The *processes* of evolution are natural selection and other mechanisms that change the genetic composition of populations.
  - These processes lead to evolutionary *patterns*, the products that evolution has produced over time.
- **Phylogeny** is the evolutionary history of a species or group of species.
- To reconstruct phylogeny, scientists use **systematics**, an analytical approach to classifying the diversity and determining the evolutionary relationships of living and extinct organisms.
  - Evidence used to reconstruct phylogenies can be obtained from the fossil record and from morphological, biochemical, and genetic similarities between organisms.
- Scientists are working to construct a universal tree of all life, which will be refined as new data are collected.

#### Concept 26.1 Phylogenies show evolutionary relationships.

- Organisms share homologous characteristics due to common descent.
  - Organisms share genes, metabolic pathways, and structural pathways with their close relatives.
- The scientific discipline of **taxonomy** determines how organisms are named and classified.

#### *Taxonomy employs a hierarchical system of classification.*

- The Linnaean system, first formally proposed by Carolus Linnaeus in the 18<sup>th</sup> century, has two main characteristics.
  1. Each species has a two-part name.
    - Under the binomial system, each species is assigned a two-part Latinized name, a **binomial**.
  2. Species are organized hierarchically into broader and broader groups of organisms.
- The first part of the name, the **genus**, is the closest group to which a species belongs.
  - The second part, the specific epithet, refers to one species within each genus.
  - The first letter of the genus name is capitalized, and both the genus and species names are italicized and Latinized.
  - For example, Linnaeus assigned to humans the optimistic scientific name *Homo sapiens*, which means “wise man.”
- A **hierarchical classification** groups species into increasingly inclusive taxonomic categories.

- Species that appear to be closely related are grouped into the same genus.
  - For example, the leopard, *Panthera pardus*, belongs to a genus that includes the African lion (*Panthera leo*) and the tiger (*Panthera tigris*).
- Genera are grouped into progressively broader categories: **family, order, class, phylum, kingdom, and domain**.
- Each taxonomic level is more comprehensive than the previous one.
  - As an example, all species of cats are mammals, but not all mammals are cats.
- The named taxonomic unit at any level is called a **taxon**.
  - For example, *Panthera* is a taxon at the genus level, and Mammalia is a taxon at the class level that includes all of the many orders of mammals.
- Higher classification levels are not defined by some measurable characteristic, such as the reproductive isolation that separates biological species.
- As a result, the larger categories are not comparable between lineages.
  - An order of snails does not necessarily exhibit the same degree of morphological or genetic diversity as an order of mammals.

***Classification and phylogeny are linked.***

- The evolutionary history of a group of organisms can be represented in a diagram called a **phylogenetic tree**.
- The branching of the tree may match the hierarchical classification of groups nested within more inclusive groups.
  - In other cases, similarities between organisms may have led taxonomists to place a species in other than the group to which it is most closely related.
  - Such species may be renamed so that their classification accurately reflects evolutionary history.
- The difficulty of aligning Linnaean classification with phylogeny has led systematists to propose that classification be based entirely on evolutionary relationships.
- **Phylocode** only names groups that include a common ancestor and all its descendents.
  - If this system were adopted, most taxonomic names would be unchanged. However, ranks such as family, order, and class would be abandoned.
  - Some commonly recognized groups would become part of other groups previously of the same rank.
  - Because birds descended from a taxon of reptiles, Aves (a Linnaean class) would be considered a subgroup of the Reptilia (another Linnaean class).

***A phylogenetic tree represents a hypothesis about evolutionary relationships.***

- Evolutionary relationships are often represented as a series of dichotomies, or two-way **branch points**.
  - Each branch point represents the divergence of two evolutionary lineages from a common ancestor.
- **Sister taxa** are groups of organisms that share an immediate common ancestor and are each other's closest relatives.
- A **rooted tree** includes the most recent common ancestor to all taxa in the tree.
- The term **basal taxon** refers to a lineage that diverges early in the history of a group, lying on a branch that originates near the common ancestor of the group.

- A **polytomy**, or branch point from which more than two descendent groups emerge, indicates our limited understanding of certain evolutionary relationships.
- Three key points about phylogenetic trees should be emphasized.
  1. Phylogenetic trees are intended to show patterns of descent, not phenotypic similarity. Closely related organisms may resemble one another due to common ancestry, but may not if their lineages have evolved at different rates or faced very different environmental conditions.
  2. The sequence of branching in a tree reflects patterns of descent and does not indicate the absolute ages of particular species.
  3. A taxon in a phylogenetic tree did not evolve from an adjacent taxon. Rather, both taxa evolved from a common ancestor.
- A species' phylogeny can provide useful information.
  - From a phylogeny of corn based on DNA data, researchers have identified two closely related species of wild grasses that may serve as “reservoirs” of beneficial alleles. These alleles may be transferred to cultivated corn by plant breeding or genetic engineering.
  - Phylogenetic trees played a role in demonstrating that “whale meat” sold in Japan was illegally harvested from protected species.

### **Concept 26.2 Phylogenies are inferred from morphological and molecular data.**

- Phenotypic and genetic similarities due to shared ancestry are called *homologies*.
- Organisms that share similar morphologies or DNA sequences are likely to be more closely related than organisms without such similarities.
- Morphological divergence between closely related species can be small or great, because morphological diversity may be controlled by relatively few genetic differences.
  - Consider the Hawaiian silversword plants, which vary from tall, twiggy trees to dense, ground-hugging shrubs. These phenotypic differences are based on small molecular divergences that arose over the last 5 million years, when the oldest of the Hawaiian Islands formed.
- Similarity due to convergent evolution is called **analogy**.
  - When two organisms from different evolutionary lineages experience similar environmental pressures, natural selection may result in convergent evolution.
  - For example, marsupial and eutherian moles are very similar in external appearance. However, they last shared a common ancestor 140 million years ago, when marsupial and eutherian mammals diverged.
  - This common ancestor was not mole-like. Analogous similarities evolved independently in these two mole lineages as they adapted to similar lifestyles.
- Distinguishing homology from analogy is critical in the reconstruction of phylogeny.
  - For example, both birds and bats have adaptations that allow them to fly.
  - A close examination of a bat's wing shows a greater similarity to a cat's forelimb than to a bird's wing.
  - Fossil evidence also documents that bat and bird wings arose independently from walking forelimbs of different ancestors.
  - Thus, a bat's wing is *homologous* to other mammalian forelimbs but is *analogous* in function to a bird's wing.
- Analogous structures that have evolved independently are also called **homoplasies**.

- In general, the more elements that are similar in two complex structures, the more likely it is that they evolved from a common ancestor.
  - For example, the skulls of an adult human and an adult chimpanzee are formed by the fusion of many bones.
  - The two skulls match almost perfectly, bone for bone.
  - It is highly unlikely that such complex structures have separate origins.
  - More likely, the genes involved in the development of both skulls were inherited from a common ancestor.
- The same argument applies to comparing genes, which are sequences of nucleotides.
  - If the genes in two organisms have very similar nucleotide sequences, it is highly likely that the genes are homologous.
- It may be difficult to carry out molecular comparisons of nucleic acids.
- The first step in molecular comparisons is to align nucleic acid sequences from the two species being studied.
  - In closely related species, sequences may differ at only one or a few sites.
  - Distantly related species may have many differences or sequences of different length.
- Over evolutionary time, insertions and deletions may accumulate, altering the lengths of the gene sequences.
- Deletions or insertions may shift the remaining sequences, making it difficult to recognize closely matching nucleotide sequences.
  - To deal with this, researchers use computer programs to analyze comparable DNA sequences of differing lengths and align them appropriately.
- The fact that molecules have diverged between species does not tell scientists how long ago their common ancestor lived.
  - Molecular divergences between lineages with reasonably complete fossil records can serve as a molecular yardstick to measure the appropriate time span of various degrees of divergence.
- As with morphological characters, it is necessary to distinguish homology from analogy to determine the usefulness of molecular similarities for reconstruction of phylogenies.
  - Very similar sequences are most likely homologies.
  - In distantly related organisms, identical bases in otherwise different sequences may simply be coincidental matches or molecular homoplasies.
- Scientists have developed mathematical tools that can distinguish “distant” homologies from coincidental matches in extremely divergent sequences.
- **Molecular systematics** uses DNA and other molecular data to determine evolutionary relationships.
- Scientists have sequenced more than 110 billion bases’ worth of nucleic acid data from thousands of species. The new data have supported hypotheses about many evolutionary relationships and have clarified others.

### **Concept 26.3 Shared characters are used to construct phylogenetic trees.**

- In an approach to systematics called **cladistics**, common descent is the primary criterion used to classify organisms.
- Biologists place species into groups called **clades**, each of which includes an ancestral species and all of its descendents.

- Clades are nested within larger clades.
- A valid clade is **monophyletic**, consisting of an ancestral species and all its descendents.
  - When biologists lack information about some members of a clade, the result may be a **paraphyletic** grouping that consists of some, but not all, of the descendents.
  - The result may also be several **polyphyletic** groupings that include taxa with different ancestors.
  - Such situations call for further reconstruction to uncover species that tie these groupings together into monophyletic clades.

***Due to descent with modification, organisms share some, but not all, characteristics with their ancestors.***

- Systematists must sort through homologous features, or characters, to separate shared derived characters from shared ancestral characters.
  - A *character* is any feature that a particular taxon possesses.
  - A **shared derived character** is an evolutionary novelty unique to a particular clade.
  - A **shared ancestral character** originated in an ancestor of the clade.
- For example, the presence of hair is a good character to use to distinguish the clade of mammals from other vertebrates.
  - Hair is a shared derived character that uniquely identifies mammals.
- The presence of a backbone can qualify as a shared derived character, but at a deeper branch point that distinguishes all vertebrates from other mammals.
  - Among vertebrates, the backbone is a shared ancestral character because it evolved in the ancestor common to all vertebrates.
- Shared derived characters are useful in establishing a phylogeny, but shared ancestral characters are not.
  - The status of a shared derived character versus a shared ancestral character is relative, depending on the level at which the analysis is being performed.

***Phylogenies can be inferred from information about shared derived characters.***

- A key step in cladistic analysis is outgroup comparison, which is used to differentiate shared ancestral characters from shared derived characters.
- To do outgroup comparison, scientists need to identify an **outgroup**, a species or group of species from an evolutionary lineage that is known to have diverged before the lineage that includes the species we are studying (the **ingroup**).
  - A suitable outgroup can be determined based on evidence from morphology, paleontology, embryonic development, and gene sequences.
- To study the relationships among an ingroup of five vertebrates (a leopard, a turtle, a salamander, a tuna, and a lamprey) on a cladogram, an animal called the lancelet is a good choice.
  - The lancelet is a small member of the phylum Chordata that lacks a backbone.
  - The presence of a backbone, shared by *all* members of the ingroup but not the outgroup, is a useful character for the whole ingroup.
  - The presence of hinged jaws, absent in lampreys and present in the other ingroup taxa, helps to identify the earliest branch in the vertebrate cladogram.

***In some phylogenetic trees, branch lengths are proportional to the rates of phylogenetic change or the times at which given events occurred.***

- In many phylogenetic trees, the branch lengths of the tree do not have specific meaning.

- The chronology represented by the branching pattern of the tree is relative (earlier vs. later) instead of absolute (how many millions of years ago).
- In other trees, the branch length reflects the number of changes that have taken place in a particular DNA sequence in that lineage.
  - For example, if the *total length* of the horizontal lines from the base of such a tree to a mouse is shorter than the length of the line leading to the fruit fly *Drosophila*, more genetic changes have occurred in the *Drosophila* lineage than in the mouse lineage since the time the lineages leading to mouse and fly diverged from a common ancestor.
- All the different lineages that have descended from the common ancestor to all living things have survived for the same number of years.
  - Humans and bacteria share a common ancestor that lived more than 3 billion years ago.
  - Fossils and genetic evidence indicate that this ancestor was a single-celled prokaryote.
  - Even though bacteria have changed little in their morphology since that common ancestor, there has been 3 billion years of change in both the bacterial lineage and the eukaryotic lineage that includes humans.

***The principles of maximum parsimony and maximum likelihood help systematists reconstruct phylogeny.***

- As available data about DNA sequences increase, it becomes more difficult to draw the phylogenetic tree that best describes evolutionary history.
  - If you are analyzing data for 50 species, there are  $3 \times 10^{76}$  different ways to form a tree.
- According to the principle of **maximum parsimony**, scientists should look for the simplest explanation that is consistent with the facts.
  - In the case of a tree based on morphological characters, the most parsimonious tree is the one that requires the fewest evolutionary events to have occurred in the form of shared derived characters.
  - For phylogenies based on DNA sequences, the most parsimonious tree requires the fewest base changes.
- The principle of **maximum likelihood** states that, given certain probability rules about how DNA sequences change over time, a tree should reflect the most likely sequence of evolutionary events.
  - Maximum-likelihood methods are complex and designed to use as much information as possible.
  - Consider two possible, equally parsimonious trees. A tree that assumes that DNA changes have occurred at equal rates along all branches of the tree from the common ancestor is more likely than a tree that assumes widely different rates of evolution in different branches.
- Many computer programs have been developed to search for trees that are parsimonious and likely.
  - Although we can never be certain precisely which tree truly reflects phylogeny, if the trees are based on a large amount of accurate data, the various methods usually yield similar trees.

***Phylogenetic trees are hypotheses.***

- Any phylogenetic tree represents a hypothesis about how the organisms in the tree are related.
  - The best hypothesis is the one that best fits all the available data.
- Scientists can make and test predictions based on the assumption that a phylogeny is correct.
  - For example, in an approach known as *phylogenetic bracketing*, scientists can predict by parsimony that features shared by two groups of closely related ancestors are present in their common ancestor and all of its descendents.
  - Evidence suggests that birds descended from a group of bipedal Saurischian theropod dinosaurs.
  - The closest living relatives of birds are crocodiles.

- Birds and crocodiles share a number of features: Both have four-chambered hearts. Both “sing” to defend territories and attract mates. Both *brood* their eggs, birds by sitting on them and crocodiles by covering their eggs with their neck.
- Based on these observations, biologists predict that dinosaurs had four-chambered hearts, sang, and brooded eggs in nests.
- The fossil record does not provide evidence of dinosaur heart structure or singing behavior. However, fossilized dinosaur nests have been found with fossilized adults crouched over the eggs in a brooding posture.
- This evidence provides independent data supporting the hypothesis that birds descended from dinosaurs.

### **Concept 26.4 An organism’s evolutionary history is documented in its genome.**

- Molecular systematics is a valuable approach for tracing an organism’s evolutionary history.
- The molecular approach helps scientists understand phylogenetic relationships that cannot be measured by comparative anatomy and other nonmolecular methods.
  - For example, molecular systematics helps biologists uncover evolutionary relationships between groups that have little ground for morphological comparison, such as animals and fungi.
  - This approach allows biologists to construct phylogenies among groups of living organisms with no fossil record.
  - Molecular biology has helped to extend systematics to evolutionary relationships far above and below the species level.
- The ability of molecular trees to encompass both short and long periods of time is based on the fact that different genes evolve at different rates, even in the same evolutionary lineage.
- For example, the DNA that codes for ribosomal RNA (rRNA) changes relatively slowly, so DNA sequences in these genes can be compared to sort out relationships between taxa that diverged hundreds of millions of years ago.
  - Studies of rRNA sequences indicate that fungi are more closely related to animals than to plants.
- In contrast, mitochondrial DNA (mtDNA) evolved relatively recently and can be used to explore recent evolutionary events, such as relationships between groups within a species.
  - One research team has used mtDNA to trace the relationships between Native American groups.

### ***Gene duplication has provided opportunities for evolutionary change.***

- Gene duplication increases the number of genes in the genome, providing opportunities for further evolutionary change.
- Repeated duplications have resulted in *gene families*, groups of related genes within an organism’s genome.
- There are two types of homologous genes: orthologous genes and paralogous genes.
- **Orthologous genes** are homologous genes found in different species because of speciation.
  - The cytochrome *c* genes (which code for an electron transport chain protein) in humans and dogs are orthologous.
- **Paralogous genes** result from gene duplication and are found in more than one copy in the same genome.
  - Olfactory receptor genes have undergone many gene duplications in vertebrates.
  - Humans and mice each have huge families of more than 1,000 of these paralogous genes.

- Orthologous genes diverge after speciation and are found in separate gene pools.
  - For example, although the cytochrome *c* genes in humans and dogs serve a similar function, the gene's sequence in humans has diverged from that in dogs in the time since these species last shared a common ancestor.
- Paralogous genes can diverge within a species because they are present in more than one copy in the genome.
  - The paralogous genes of the human olfactory receptor gene family have diverged from each other during our long evolutionary history.
  - They now specify proteins that confer sensitivity to a wide variety of molecules, ranging from food odors to sex pheromones.

***Comparison of genomes of different organisms provides information about genome evolution.***

- Now that we have compared entire genomes of different organisms, two remarkable patterns have emerged.
- Lineages that diverged long ago can share orthologous genes.
  - Approximately 99% of the genes of humans and mice are detectably orthologous, and 50% of human genes are orthologous with those of yeast.
  - All living things share many biochemical and development pathways.
- The number of genes seems not to have increased at the same rate as phenotypic complexity.
  - Humans have only four times as many genes as yeast, a simple unicellular eukaryote, although we have a large, complex brain and a body that contains more than 200 different types of tissues.
  - Many human genes are more versatile than those of yeast and can carry out a wide variety of tasks in various body tissues.

**Concept 26.5 Molecular clocks help track evolutionary time.**

- One of the goals of evolutionary biology is to understand the relationships among all organisms, including those for which there is no fossil record.
- **Molecular clocks** serve as yardsticks for measuring the absolute time of evolutionary change.
  - Molecular clocks are based on the observation that some genes and other regions of the genome evolve at constant rates.
  - For these regions, the number of nucleotide substitutions in orthologous genes is proportional to the divergence time, the time that has elapsed since the two species shared a common ancestor.
  - In the case of paralogous genes, the number of substitutions is proportional to the time since the genes became duplicated.
- Scientists calibrate the molecular clock of a gene by graphing the number of genetic differences (nucleotide, codon, or amino acid differences) against the dates of evolutionary branch points that are known from the fossil record.
  - The average rate of genetic change inferred from such a graph can be used to estimate the absolute date of evolutionary events that have no fossil record.
- No molecular clock is completely accurate.
  - Genes that make good molecular clocks have fairly smooth *average* rates of change.
  - Over time, there may be chance deviations above and below the average rate.
- Rates of change of various genes vary greatly.
  - Some genes evolve a million times faster than others.



**Many of the changes in genes used as molecular clocks are assumed to be selectively neutral, fixed in a population by genetic drift.**

- The molecular clock approach assumes that much of the change in DNA sequences is due to genetic drift and is selectively neutral.
  - The **neutral theory** suggests that much evolutionary change in genes and proteins has no effect on fitness and, therefore, is not influenced by natural selection.
  - Supporters of this theory point out that many new mutations are harmful and are removed quickly.
  - However, if most other mutations are neutral and have little or no effect on fitness, then the rate of molecular change should be clocklike in its regularity.
- Differences in the rates of change of specific genes are a function of the importance of the gene.
  - If the exact sequence of amino acids specified by a gene is essential to survival, most mutations will be harmful and will be removed by natural selection.
  - If the sequence of genes is less critical, more mutations will be neutral, and mutations will accumulate more rapidly.
- Some DNA changes are favored by natural selection, which leads some scientists to question the accuracy and utility of molecular clocks for timing evolution.
  - Evidence suggests that almost 50% of the amino acid differences in proteins of two *Drosophila* species have resulted from directional natural selection.
- Over very long periods of time, fluctuations in the rate of accumulation of mutations due to natural selection may even out.
  - Even genes with irregular clocks can serve as approximate markers of elapsed time.
- Biologists are skeptical of conclusions derived from molecular clocks that have been extrapolated to time spans beyond the calibration in the fossil record
  - Few fossils are older than 550 million years old.
  - Estimates for evolutionary divergences prior to that time may assume that molecular clocks have been constant over billions of years.
  - Such estimates have a high degree of uncertainty.
- Molecular clocks may be calibrated with many genes, rather than one or a few.
  - When many genes are used, fluctuations in the evolutionary rate due to natural selection or other factors that vary over time may average out.
  - One group of researchers constructed molecular clocks of vertebrate evolution from published data for 658 nuclear genes.
  - The molecular and fossil-based estimates of divergence times agreed closely.
- The molecular clock approach has been used to date the jump of the HIV virus from related viruses that infect chimpanzees and other primates to humans.
- The virus has spread to humans more than once.
  - The multiple origins of HIV are reflected in the variety of strains of the virus.
- HIV's genetic material is RNA, which evolves quickly.
  - HIV-1 M is the most common HIV strain.
- Investigators have calibrated the molecular clock for the virus by comparing samples of the virus collected at various times, including a sample from 1959.
  - From their analysis, researchers project that the HIV-1 M strain invaded humans in the 1930s.

## **Concept 26.6 New information continues to revise our understanding of the tree of life.**

- Early taxonomists classified all known species into two kingdoms: plants and animals.
- In the late 1960s, a five-kingdom system was widely adopted.
  - The kingdoms were Monera (prokaryotes), Protista (primarily unicellular eukaryotes), Plantae, Fungi, and Animalia.
  - Monera included all prokaryotic cells. Eukaryotic cells were divided among the other five kingdoms.
- Phylogenies based on genetic data revealed a fundamental problem with the five-kingdom system: Some prokaryotes differed as much from each other as they did from eukaryotes.
- Biologists have now adopted a three-domain system. The three domains—Bacteria, Archaea, and Eukarya—are at a taxonomic level higher than the kingdom level.
  - Analyses of nearly 100 completely sequenced genomes support the validity of these domains.
- The domain Bacteria contains most of the currently known prokaryotes, including the bacteria most closely related to chloroplasts and mitochondria.
- The domain Archaea consists of a diverse group of prokaryotes inhabiting a wide variety of environments.
  - Bacteria and archaea differ in many key structural, biochemical, and physiological characteristics.
  - Some archaea can use hydrogen as an energy source, and others were the chief source of the natural gas deposits that are found throughout Earth's crust.
- The domain Eukarya includes all organisms with cells containing true nuclei.
- Of the five kingdoms previously recognized by taxonomists, Plantae, Fungi, and Animalia are used.
  - The kingdom Monera, which included organisms in two different domains, is obsolete.
  - The kingdom Protista is polyphyletic, including organisms that are more closely related to plants, animals, or fungi than to other protists.
- Much of the history of life has been about single-celled organisms.
  - Most evolutionary branches on the tree of life are dominated by single-celled organisms.
- The first major split in the history of life occurred when the bacteria diverged from the other two domains.
  - Eukaryotes and archaea are more closely related to each other than either is to bacteria.
- The tree of all life is based largely on rRNA genes, which code for the RNA within ribosomes.
  - These genes have evolved so slowly that homologies between distantly related organisms can be detected.
- Other genes reveal a different set of relationships.
  - Many of the metabolic genes in yeast are more similar to bacterial genes than to archaea genes.
  - This finding suggests a more recent common ancestor between eukaryotes and bacteria.
- Early in the history of life, there were many interchanges of genes between organisms in different domains.
- These interchanges took place through **horizontal gene transfer**, in which genes are transferred from one genome to another by mechanisms such as transposable elements, plasmid exchange, and viral infection.
  - In some cases, different organisms fused to produce new, hybrid organisms.
- Recent research reinforces the view that horizontal gene transfer is important.

- A 2008 analysis indicated that, on average, 80% of the genes in 181 prokaryotic genomes had moved between species at some point during the course of evolution.
- The occurrence of horizontal gene transfers helps explain why universal genes built from different genes give inconsistent results.
- Some scientists suggest that horizontal gene transfers were so common that the early history of life should be represented as a tangled network instead of a dichotomously branched tree.
  - Other scientists have suggested that the history of life should be represented by a ring, not a tree.
- An analysis based on hundreds of genes supports the hypothesis that eukaryotes arose as a fusion between an early bacterium and an early archaean.
  - If eukaryotes arose as an endosymbiosis between an early bacterium and an early archaean, then eukaryotes are related equally closely to bacteria *and* archaea.
  - Such an evolutionary relationship cannot be depicted in a tree of life, but can be depicted in a *ring* of life.