Learning unit 1: Phylogeny and Systematics

1.1 Introduction

To complete the learning unit, you will need to refer to pages 523–542 chapter 22 in Campbell et al. (2015)

You may always wonder as to why and how do biologists distinguish and categorise the millions of species on Earth. In this unit we will focus on how biologists trace phylogeny, the evolutionary history of a species or group of species. A phylogeny of snakes and lizards as an exemplar, shows that both eastern glass lizard and snakes evolved from lizards with legs –but they evolved from different lineages of legged lizards. This is demonstrated in figure 22.2 from your prescribed textbook. Thus it appears that their legs legless conditions evolved independently. We will look at how biologists reconstruct and interpret phylogenies using systematics. Systematics is a discipline focused on classifying organisms and determining their evolutionary relationships. We then focus attention on how systematists develop hypotheses about the evolutionary relationship of all the branches, twigs, and leaves on the tree of life.

1.2 Learning outcomes

By the end of this learning unit you should be able to

- explain the determination of phylogeny from common ancestries
- describe the binomial nomenclature system
- explain the hierarchical system of classification
- list the different hierarchical classification groupings
- discuss cladistic analysis on which systematics is based
- discuss the prince of parsimony and maximum likelihood
- construct the phylogeny tree

1.3 Phylogenies show evolutionary relationships

Recommended reading: pages 524–527 chapter 22 in Campbell et al. (2015)

Organisms share many characteristics because of common ancestry. As a result, you can learn a great deal about a species if you have explored its evolutionary history. For example, an organism is likely to share many of its genes, metabolic pathways, and structural proteins with its close relatives (Figure 1.1). We will consider practical applications of such information later in this unit, but first we will examine how organisms are named and classified, the scientific discipline taxonomy. We will also look at how we can interpret and use diagrams that represent evolutionary history.

Figure 1.1: Relationships between wolf-like canids (*Canis* species) capable of hybridizing

(https://commons.wikimedia.org/wiki/File:Wolf-like_canids_phylogeny_(ita).jpg)

1.3.1 Binomial nomenclature

The Swedish natural Carl von Linne (1707-1778), better known as the Latinised name, Carolus Linnaeus, was the first modern practitioner of taxonomy, the science that identifies, names, and classifies new species. Linnaeus invented the system of Binomial nomenclature, in which species are assigned a Latinised two part name, or binomial. The first part of a binomial is the name of genus (plural, genera) to which the species belongs. Note that genus is a group of species with similar characteristics. First letter of genus is capitalised, whilst the specific epithet is always in lower letters. The second part of binomial is called the specific epithet, is unique for each species within the genus. Both genus and specific epithet are always underlined or italicised. In order for biologists to avoid ambiguity and confusion when communicating about their research, Latin scientific names are used. There are many binomial names today, for an example, *Panthera pardus* and *Homo sapiens* are the scientific names of the leopard and “wise man,” respectively.

1.3.1.1 Activity 1.1

Do this activity and add it to your portfolio.

Remember, this could serve as part of your summary to use in preparing for the exam!

Refer to your textbook, and answer the following questions:

b. How does the system of binomial nomenclature minimise ambiguity in the naming and identification of species?

b. Are the following scientific names correct? Give reasons.

1. Acacia aerioloba
2. Panthera pardus
3. *Ophisaurus ventralis*
4. *Homo sapie*

1.3.1.2 Feedback on activity 1.1

In answering question a) the system of binomial nomenclature avoids ambiguity in the naming of species because it assigns a unique two-part name to each species.

b. Were you able to easily recognize the rules of binomial nomenclature before you attempt any of these scientific names? If yes, better keep up the good work but if not, please make sure that you revisit binomial nomenclature section and revise the simple rules on scientific names. In answering those questions rationally;

2. Yes, because the scientific name is written correctly and the first letter of genus has been written in capital letter, whilst both genus and specific epithet are underlined.
4. No, though the name has been well spelt, both genus and specific epithet should be either underlined or italicized.
6. Yes, all the rules have been followed.

4. No, the scientific name must always be spelled out correctly. In addition, the correct way is *Homo sapiens*.

1.3.2 Hierarchical classification

In addition to naming species, Linnaeus also grouped living organisms into a hierarchy of increasingly inclusive categories. Linnaeus' classification, called the taxonomic hierarchy, includes a nested series of formal categories which include domain, kingdom, phylum, class, order, family, genus, species and lastly, subspecies. To put this to our daily lives understanding, the resulting of biological classification of a particular organisms is more like a postal address identifying a person in a particular apartment, in a building with many apartments, on a street with many apartments buildings, in a city with many streets, and so on. Note that the organisms included within any category of the taxonomic hierarchy compose a taxon (plural, taxa). Leopard, for example, is a taxon (Felidae) at the family level, and *Panthera* is a taxon at a genus level (refer to figure 22.4 from your textbook). Species that are included in the same taxon at the bottom of the hierarchy (that is, in the same genus or family) generally share many characteristics. By contrast, species that are included in the same taxon only near the top of the hierarchy (that is, the same kingdom or phylum) generally share many fewer traits. Very important, in the Linnaean system, taxa broader than the genus are not italicised or underlined, though they are capitalised.

1.3.2.1 Activity 1.2

Do this activity and add it to your portfolio.

Refer to your textbook and answer the following questions:

b. How does the taxonomic hierarchy help biologists organise information about different species?
d. List the major taxonomic categories from most to least inclusive.

1.3.2.2 Feedback on the activity 1.2

b. The taxonomic hierarchy helps biologists organise information about different species because it categorises them into increasingly inclusive groups. Species that are include in a lower taxonomic category share many characteristics, whereas those included only in the same higher category share fewer characteristics.
d. Kingdom, Phylum, Class, Order, Family, Genus, Species

1.4 Construction of phylogeny trees

Recommended reading: pages 529–535 chapter 22 in Campbell et al. (2015)

Patterns of shared characteristics can be depicted in a diagram called a cladogram. If shared characteristics are homologous and, thus, explained by common ancestry, then the cladogram forms the basis of a phylogenetic tree. A **clade** is defined as a group of species that includes an ancestral species and all its descendants. The study of resemblances among clades is called cladistics. Each branch, or clade, can be nested within larger clades. A valid clade is monophyletic, consisting of an ancestral species and all its descendants. When we lack information about some members of a clade, the result is a paraphyletic grouping that consists of some, but not all, of the descendants. The result may also be several

polyphyletic groupings that lack a common ancestor. Such situations call for further reconstruction to uncover species that tie these groupings together into monophyletic clades. Determining which similarities between species are relevant to grouping the species in a clade is a challenge. It is especially important to distinguish similarities that are based on shared ancestry or homology from those that are based on convergent evolution or analogy.

Systematists must also sort through homologous features, or characters, to separate shared derived characters from shared primitive characters. A “character” refers to any feature that a particular taxon possesses. A shared derived character is unique to a particular clade. A shared primitive character is found not only in the clade being analysed, but also in older clades. For example, the presence of hair is a good character to distinguish the clade of mammals from other tetrapods. It is a shared derived character that uniquely identifies mammals. However, 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 primitive character because it evolved in the ancestor common to all vertebrates. Shared derived characters are useful in establishing a phylogeny, but shared primitive characters are not. The status of a character shared derived versus shared primitive may depend on the level at which the analysis is being performed. A key step in cladistic analysis is outgroup comparison, which is used to differentiate shared primitive characters from shared derived ones. To do this, we need to identify an outgroup, a species or group of species that is closely related to the species that we are studying, but known to be less closely related than any members of the study group are to each other.

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 species making up the ingroup display a mixture of shared primitive and shared derived characters.

In an outgroup analysis, the assumption is that any homologies shared by the ingroup and outgroup are primitive characters that were present in the common ancestor of both groups. Homologies present in some or all of the ingroup taxa are assumed to have evolved after the divergence of the ingroup and outgroup taxa. In our example, a notochord, present in lancelets and in the embryos of the ingroup, is a shared primitive character and, thus, not useful for sorting out relationships between members of the ingroup. The presence of a vertebral column, shared by all members of the ingroup but not the outgroup, is a useful character for the whole ingroup. The presence of jaws, absent in lampreys and present in the other ingroup taxa, helps to identify the earliest branch in the vertebrate cladogram.

Analyzing the taxonomic distribution of homologies enables us to identify the sequence in which derived characters evolved during vertebrate phylogeny. A cladogram presents the chronological sequence of branching during the evolutionary history of a set of organisms. However, this chronology does not indicate the time of origin of the species that we are comparing, only the groups to which they belong. For example, a particular species in an old group may have evolved more recently than a second species that belongs to a newer group. A cladogram is not a phylogenetic tree.

To convert it to a phylogenetic tree, we need more information from sources such as the fossil record, which can indicate when and in which groups the characters first appeared. Any chronology represented by the branching pattern of a phylogenetic tree is relative (earlier versus later) rather than absolute (so many millions of years ago).

Some kinds of tree diagrams can be used to provide more specific information about timing. In a phylogram, the length of a branch reflects the number of genetic changes that have taken place in a particular DNA or RNA sequence in a lineage. Even though the branches in a phylogram may have different lengths, all the different lineages that descend from a common ancestor have survived for the same number of years. Humans and bacteria had a common ancestor that lived more than 3 billion years ago. This ancestor was a single-celled prokaryote and was more like a modern bacterium than like a human. Even though bacteria have apparently changed little in structure since that common ancestor, there have nonetheless been 3 billion years of evolution in both the bacterial and eukaryotic lineages. These equal amounts of chronological time are represented in an ultrameric tree.

In an ultrameric tree, the branching pattern is the same as in a phylogram, but all the branches that can be traced from the common ancestor to the present are of equal lengths. Ultrameric trees do not contain the information about different evolutionary rates that can be found in phylograms. However, they draw on data from the fossil record to place certain branch points in the context of geological time.

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

According to the principle of maximum parsimony, we 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 phylograms based on DNA sequences, the most parsimonious tree requires the fewest base changes in DNA. The principle of maximum likelihood states that, given certain rules about how DNA changes over time, a tree should reflect the most likely sequence of evolutionary events. Maximum likelihood methods are designed to use as much information as possible. Many computer programs have been developed to search for trees that are parsimonious and likely: “Distance” methods minimise the total of all the percentage differences among all the sequences. More complex “character-state” methods minimise the total number of base changes or search for the most likely pattern of base changes among all the sequences. Although we can never be certain precisely which tree truly reflects phylogeny, if they 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. A hypothesis may be modified when new evidence compels systematists to revise their trees. Many older phylogenetic hypotheses have been changed or rejected since the introduction of molecular methods for comparing species and tracing phylogeny. Often, in the absence of conflicting information, the most parsimonious tree is also the most likely. Sometimes there is compelling evidence that the best hypothesis is not the most parsimonious. Nature does not always take the simplest course.

In some cases, the particular morphological or molecular character we are using to sort taxa actually did evolve multiple times. For example, the most parsimonious assumption would be that the four-chambered heart evolved only once in an ancestor common to birds and mammals but not to lizards, snakes, turtles, and crocodiles. But abundant evidence indicated that birds and mammals evolved from different reptilian ancestors. The hearts of birds and mammals develop differently, supporting the hypothesis that they evolved independently. The most parsimonious tree is not consistent with the above facts, and must be rejected in favour of a less parsimonious tree. The four-chambered hearts of birds and mammals are analogous, not homologous. Occasionally misjudging an analogous similarity in morphology or gene sequence as a shared derived homology is less likely to distort a phylogenetic tree if several derived characters define each clade in the tree. The strongest phylogenetic hypotheses are those supported by multiple lines of molecular and morphological evidence as well as by fossil evidence.

1.5 An organism's evolutionary history is documented in its genome

Recommended reading: pages 535–536 chapter 22 in Campbell et al. (2015)

Molecular systematics is a valuable tool for tracing an organism's evolutionary history. The molecular approach helps us to understand phylogenetic relationships that cannot be measured by comparative anatomy and other non-molecular methods. For example, molecular systematics helps us uncover evolutionary relationships between groups that have no grounds for morphological comparison, such as mammals and bacteria.

Molecular systematics enables scientists to compare genetic divergence within a species. Molecular biology has helped to extend systematics to evolutionary relationships far above and below the species level. Its findings are sometimes inconclusive, as in cases where a number of taxa diverged at nearly the same time. 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 comparisons of DNA sequences in these genes can be used to sort out relationships between taxa that diverged hundreds of millions of years ago. 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.

Gene duplication has provided opportunities for evolutionary change

Gene duplication increases the number of genes in the genome, providing opportunities for further evolutionary change. Gene duplication has resulted in gene families, which are groups of related genes within an organism's genome. Like homologous genes in different species, these duplicated genes have a common genetic ancestor. There are two types of homologous genes: orthologous genes and paralogous genes. The term orthologous refers to homologous genes that are found in different gene pools because of speciation. The ß hemoglobin genes in humans and mice 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. Now that we have compared entire genomes of different organisms, two remarkable facts have emerged. Orthologous genes are widespread and can extend over enormous evolutionary distances. Approximately 99% of the genes of humans and mice are demonstrably 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 five 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 yeast and can carry out a wide variety of tasks in various body tissues.

1.6 Molecular clocks help track evolutionary time

Recommended reading: pages 536–540 chapter 22 in Campbell et al. (2015)

In the past, the timing of evolutionary events has rested primarily on the fossil record. One of the goals of evolutionary biology is to understand the relationships among all living organisms, including those for which there is no fossil record. Molecular clocks serve as yardsticks for measuring the absolute time of evolutionary change. They are based on the observation that some regions of the genome evolve at constant rates. For these regions, the number of nucleotide substitutions in orthologous genes is proportional to the time that has elapsed since the two species last shared a common ancestor. In the case of paralogous genes, the number of substitutions is proportional to the time since the genes became duplicated.

We can calibrate the molecular clock of a gene by graphing the number of nucleotide differences against the timing of a series of evolutionary branch points that are known from the fossil record. The slope of the best line through these points represents the evolution rate of that molecular clock. This rate 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. No genes mark time with a precise tick-tock accuracy in the rate of base changes.

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. 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 Darwinian selection. Researchers supporting this theory point out that many new mutations are harmful and are removed quickly. However, if most of the rest are neutral and have little or no effect on fitness, the rate of molecular change should be clocklike in their 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 favoured by natural selection. This 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 mark elapsed time approximately. 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.

The molecular clock approach has been used to date the jump of the HIV virus from related SIV 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-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. From their analysis, they project that the HIV-1 M strain invaded humans in the 1930s.

There is a universal tree of life

The genetic code is universal in all forms of life. From this, researchers infer that all living things have a common ancestor. Researchers are working to link all organisms into a universal tree of life. Two criteria identify regions of DNA that can be used to reconstruct the branching pattern of this tree. The regions must be able to be sequenced. They must have evolved slowly, so that even distantly related organisms show evidence of homologies in these regions. Ribosomal-RNA genes, coding for the RNA component of ribosomes, meet these criteria.

Two points have emerged from this effort

i) The tree of life consists of three great domains: Bacteria, Archaea, and Eukarya.

Most prokaryotes belong to Bacteria. Archaea includes a diverse group of prokaryotes that inhabit many different habitats. Eukarya includes all organisms with true nuclei, including many unicellular organisms as well as the multicellular kingdoms.

ii) The early history of these domains is not yet clear.

- Early in the history of life, there were many interchanges of genes between organisms in the different domains. One mechanism for these interchanges was horizontal gene transfer, in which genes are transferred from one genome to another by mechanisms such as transposable elements. Different organisms fused to produce new, hybrid organisms. It is likely that the first eukaryote arose through fusion between an ancestral bacterium and an ancestral Archaean.

1.7 Activity 1.3

Do this activity and add it to your portfolio.

Refer to your textbook and answer the following questions:

b. Distinguish between phylogeny and systematics.
d. In the following cladogram, which node occurred earliest in time?

c. In the cladogram for question (a): Which node represents the most recent common ancestor of terminal taxa B and C?
d. In the cladogram for question (a): Which terminal taxon is B more closely related to, A or C?

e. Explain how shared derived characteristics can be used to construct a phylogenetic diagram.
g. Describe the evidence that suggests that there is a universal tree of life.

1.8 Feedback on activity 1.3

a) In answering the first question, Phylogeny is the evolutionary history of a species or group of related species. Systematics is the study of biological diversity in an environmental context, encompassing taxonomy and involving the reconstruction of phylogenetic history.

b) Since the common ancestor is located at the base, your answer should be node 1.

c) Surely is node 2.

d) The correct answer is terminal taxon C

e) Biologists hypothesise that all of the chromosomes were inherited from the same ancestor. It's possible that in one of the descendants, one chromosome became two or two chromosomes became one, therefore, they can conclude that there is evolutionary history between the two species.

f) The tree of life is based on ribosomal RNA sequences. All life on earth can be placed in one of the three major categories, leading biologists to believe that all life started from a common ancestor.

1.9 Summary

Linnaeus's binomial classification system gives organisms two-part names: a genus and specific epithet. Species are grouped in increasingly broad taxa by Linnaean system. Thus, related genera are placed in the same family, families in orders, orders in classes, classes in phyla, phyla in kingdoms, and lastly, kingdom in domains. Systematics depict evolutionary relationships as branching phylogenetic trees. However, many systematics propose that classification be based entirely on evolutionary relationships.

A clade is a monophyletic grouping that includes an ancestral species and all of its descendants. Clade can be distinguished by their shared derived characters. Among phylogenies, the most parsimonious tree is the one that requires the fewest evolutionary changes.

Orthologous genes are homologous genes found in different species as a result of speciation. Whereas, paralogous genes are homologous genes within a species that result from gene duplication. Distantly related species often have many orthologous genes. Some regions of DNA change at a rate consistent enough to serve as a molecular clock, in which the amount of genetic changes is used to estimate the date of past evolutionary events. Molecular clock analyses suggest that the most common strain of HIV transmitted from primates to humans in the early 1900s.