Chapter 25

Phylogeny and Systematics

Key Concepts

25.1 Phylogenies are based on common ancestries inferred from fossil, morphological, and molecular evidence

25.2 Phylogenetic systematics connects classification with evolutionary history

25.3 Phylogenetic systematics informs the construction of phylogenetic trees based on shared characters

25.4 Much of an organism's evolutionary history is documented in its genome

25.5 Molecular clocks help track evolutionary time

Framework

A goal of systematics is to reconstruct the phylogenetic history of species. Cladistic analysis is used to identify monophyletic taxa based on shared derived characters. The best phylogenetic hypotheses are parsimonious trees based on molecular homologies as well as on morphological and fossil evidence.

Chapter Review

Phylogeny is the evolutionary history of a species. The analytical study of the diversity of life and its phylogenetic history is called **systematics. Molecular systematics** uses DNA and RNA comparisons to infer relationships between genes and whole genomes.

25.1 Phylogenies are based on common ancestries inferred from fossil, morphological, and molecular evidence

The Fossil Record Fossils are preserved impressions or remnants of past organisms. The strata of sedimentary rocks form from the compression of layers of sand and silt deposits. The succession of organisms found in strata comprise the **fossil record.**

Paleontologists analyze a variety of fossils: preserved bones, teeth, or shells; mineralized fossils; sedimentary fossils that retain organic material that can be analyzed; molds or casts of organisms; trace fossils, such as footprints or burrows that give clues to an animal's behavior; and occasionally, an entire body of an organism preserved in ice, peat, or amber. The formation of a fossil is an unlikely occurrence. A large number of species that lived probably left no fossils; geologic processes destroy many fossils; and only a fraction of existing fossils have been found.

■ INTERACTIVE QUESTION 25.1

What types of organisms are most likely to appear in the fossil record?

Morphological and Molecular Homologies Homologies are similarities due to shared ancestry. Morphological and molecular homologies among living organisms can be used to infer phylogenetic history. The more similar morphologies or DNA sequences are, the more likely that organisms are closely related.

Sorting homology from analogy is essential in constructing phylogenies. **Analogy** is similarity due to convergent evolution, in which unrelated species develop similar features because they have similar ecological roles and natural selection has led to similar adaptations. Analogous structures are sometimes called **homoplasies.** In general, when two similar structures are very complex, they are more likely to be homologous and have a shared origin.

■ INTERACTIVE QUESTION 25.2

What two complications may make it difficult to determine phylogenetic relationships based on morphological similarities between species?

In molecular comparisons, the more nucleotide sequences shared between comparable genes, the more likely the genes are to be homologous. Molecular comparisons are complicated by insertion or deletion mutations that change the lengths of homologous regions of DNA. Computer programs are used to identify similar sequences and then insert gaps in order to align homologous DNA segments properly for nucleotide comparisons. The matching bases between organisms that are not closely related may be molecular homoplasies. Mathematical tools have been developed to identify "distant" homologies between extremely divergent sequencies.

■ INTERACTIVE QUESTION 25.3

Base deletions have changed the lengths and alignments of these two homologous regions of DNA. Determine the best possible fit between these two DNA sequences. How many deletions and base changes have occurred in these DNA segments?

A C G T G C A C G

A G T G A G G

25.2 Phylogenetic systematics connects classification with evolutionary history

Systematics dates back to Linnaeus, who named and classified organisms. **Taxonomy** divides organisms into categories based on a set of characteristics used to determine similarities and differences. Although not based on genealogy, parts of his system are useful in phylogenetic systematics.

Binomial Nomenclature Each species is assigned a two-part latinized name—a **binomial**—consisting of the **genus** and the **specific epithet,** or species.

Hierarchical Classification Linnaeus also organized species into broader categories. Related genera are grouped into **families,** which are then grouped into **orders, classes, phyla,** and **kingdoms,** and more recently, into **domains.**

A taxonomic unit at any level is called a **taxon.** Both words of the binomial are italicized, and the names for all taxa at the genus level and higher are capitalized. Taxa across different lineages are not comparable because grouping into higher classification levels is ultimately arbitrary.

Linking Classification and Phylogeny A **phylogenetic tree** represents hypotheses about evolutionary relationships. Each dichotomous branch point represents the divergence of two species from a common ancestor. Phylogenetic systematics originated with Darwin, who recognized the evolutionary implications of the classification hierarchy.

25.3 Phylogenetic systematics informs the construction of phylogenetic trees based on shared characters

Cladistics A **cladogram** is a phylogenetic tree that shows the pattern of homologous shared characteristics among taxa. A **clade** consists of an ancestral species and all of its descendant species. Such a clade is **monophyletic.** A **polyphyletic** clade includes several clades but not their common ancestor, and a **paraphyletic** clade excludes some species that share a common ancestor with other species in the clade. **Cladistics** is the study of the relationships among clades.

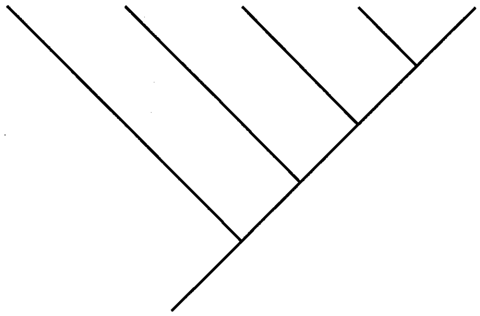
After separating similar characters that are homologous from those that are analogous, systematists identify **shared primitive characters,** which are common to more inclusive taxa and evolved at an earlier branch point, and **shared derived characters,** which are evolutionary novelties that are unique to a particular clade.

To determine the branching sequence of a group of related species, the group is compared to an **outgroup,** a species or group of species that is closely related to the group being studied. Any homologies that are common to both the outgroup and the **ingroup,** the taxa to be grouped, are shared primitive characters that were present in an ancestor common to both groups. A comparison of the numbers of characters that are present in each taxon of the ingroup indicates the sequence in which shared derived characters evolved and determines the branch points used to produce a cladogram.

■ INTERACTIVE QUESTION 25.4

Place the taxa (outgroup, A, B, C, and D) on the cladogram based on the presence or absence of the characters 1–4 as shown in this table. Indicate before each branch point the number for the shared derived character that evolved in the ancestor of the clade.

	Taxa				
Characters	Outgroup O	A	B	C	D
1	0	1	1	1	1
2	0	0	1	0	1
3	0	1	1	0	1
4	0	0	1	0	0



Phylogenetic Trees and Timing The branching pattern of a phylogenetic tree is relative rather than absolute; it indicates only the order in which members of each clade last shared a common ancestor.

In **phylograms,** which are based on homologous nucleotide sequences, the length of a branch correlates with the number of evolutionary changes (in a nucleotide sequence) that have occurred in the lineage. **Ultrametric trees** use the fossil record to locate branch points in the context of geologic time and also show all the branches from a common ancestor to the present as the same length.

Maximum Parsimony and Maximum Likelihood Systematists use morphological characters or molecular comparisons to choose among many possible phylogenetic trees using the principle of **maximum parsimony**—the smallest number of evolutionary changes is the simplest explanation and thus the best hypothesis to consider first. According to the principle of **maximum likelihood,** a tree can be found that represents the most likely sequence of events, given a certain set of assumptions. Computer programs search for the most parsimonious and most likely trees, using either "distance" methods that minimize the total of the percent differences among the sequences or "character-state" methods that minimize the total number of base changes or the most likely pattern of base changes among the sequences.

Phylogenetic Trees as Hypotheses The most parsimonious tree based on the fewest number of evolutionary changes represents the best hypothesis of the relationships among a set of species. A hypothesis of phylogenetic relationships becomes more reliable with the more data that can be compared (a large database of DNA sequence comparisons and several derived characters used to define each clade).

■ INTERACTIVE QUESTION 25.5

According to the principle of parsimony, the evolution of the four-chambered heart should place birds and mammals in the same clade. Why does the most accepted evolutionary tree show them as separate branches from the reptilian line?

25.4 Much of an organism's evolutionary history is documented in its genome

Molecular systematics uses nucleic acid or other molecular comparisons to uncover phylogenetic relationships, making it possible both to compare genetic divergence among individuals within a species and to reconstruct phylogeny among very distantly related species. The ability to span this length of time depends on genes that evolve both very slowly and very rapidly, such as rRNA sequences and mtDNA.

Gene Duplications and Gene Families Gene duplications are important evolutionary mutations, increasing the number of genes in the genome and the opportunity for change. **Orthologous genes** are homologous genes that diverge following speciation. **Paralogous genes** result from gene duplication and can diverge while in the same gene pool, providing new opportunities for evolutionary change.

Genome Evolution Widespread orthologous genes indicate that all organisms share many biochemical and developmental pathways. There is a relatively small difference in the numbers of genes in organisms of very different complexity.

■ INTERACTIVE QUESTION 25.6

a. Give an example of genes that would be compared to uncover phylogenetic relationships among the earliest branches on the tree of life.

b. Give an example of genes that evolve very rapidly and are used to discriminate among closely related species.

c. Give an example of paralogous genes.

25.5 Molecular clocks may track of evolutionary time

Molecular Clocks Some regions of DNA appear to evolve at constant rates, and comparisons of the number of nucleotide substitutions in orthologous genes can serve as **molecular clocks** to estimate the time since species branched from their common ancestor. The number of differences in paralogous genes is proportional to the time since the genes were duplicated.

Some genes appear to have a reliable average rate of evolution. Graphs that plot nucleotide or amino acid differences against times for known evolutionary branch points can be used to estimate phylogenetic branchings that are not evident from the fossil record.

According to the **neutral theory,** much of the evolutionary change in genes does not affect fitness. Harmful mutations are removed quickly from the gene pool, but neutral mutational changes should become fixed at a constant rate.

■ INTERACTIVE QUESTION 25.7

Using the neutral theory of molecular evolution, explain why different genes might have a different molecular clock rate.

Natural selection, which favors some DNA changes over others, should disrupt the smooth running of the molecular clock. Evolutionists disagree about the extent of neutral mutations, and thus about whether molecular clocks are reliable for timing evolution. Many are skeptical when molecular clocks are used to date evolutionary divergences that occurred billions of years ago.

By comparing sequences of HIV viruses from samples taken at various times during the epidemic, including one from 1959, researchers have observed a remarkably consistent rate of evolution. They estimated that the HIV-1 M strain first infected humans in the 1930s.

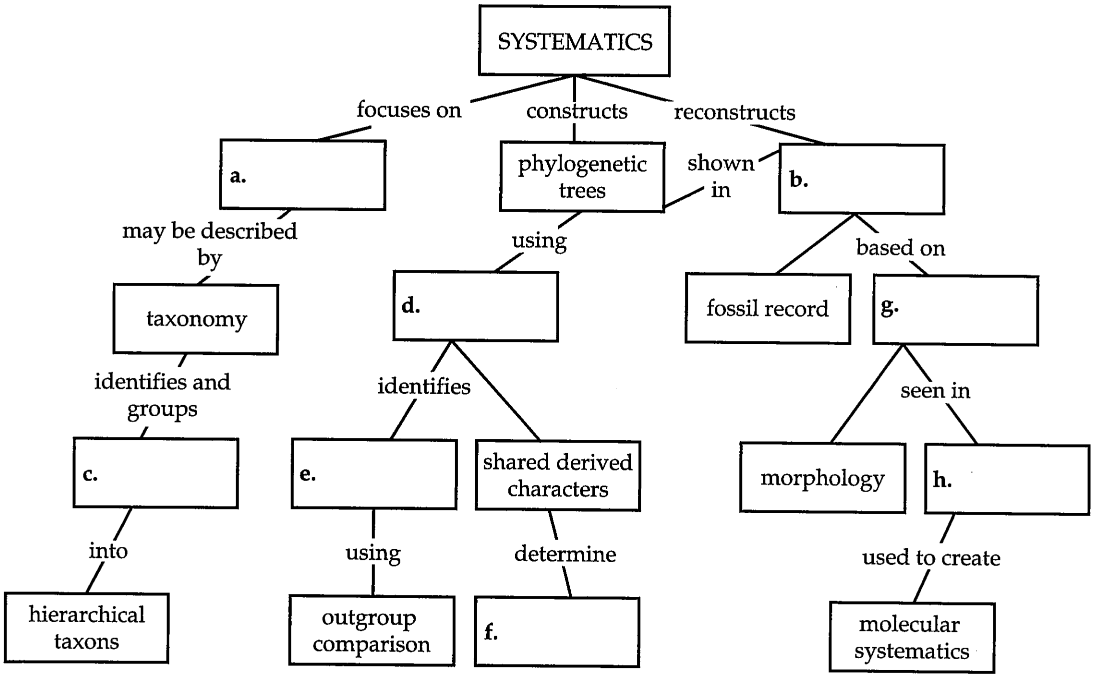
The Universal Tree of Life The rRNA genes have evolved so slowly that they have been used as the basis for constructing a universal tree of life. The tree consists of the domains Bacteria, Archaea, and Eukarya. Genome comparisons from the three domains indicate that horizontal gene transfer, perhaps through fusions of different organisms, occurred during the early history of life.

Word Roots

analog- = proportion (*analogy:* similarity due to convergence)
bi- = two; **nom-** = name (*binomial:* a two-part latinized name of a species)
clado- = branch (*cladogram:* a dichotomous phylogenetic tree that branches repeatedly)
homo- = like, resembling (*homology:* similarity in characteristics resulting from a shared ancestry)
mono- = one (*monophyletic:* pertaining to a taxon derived from a single ancestral species that gave rise to no species in any other taxa)
parsi- = few (*principle of parsimony:* the premise that a theory about nature should be the simplest explanation that is consistent with the facts)
phylo- = tribe; **-geny** = origin (*phylogeny:* the evolutionary history of a taxon)

Structure Your Knowledge

1. Complete this concept map to help you review some key ideas about the field of systematics.



Test Your Knowledge

MULTIPLE CHOICE: *Choose the one best answer.*

1. The richest source of fossils is found
   - **a.** in coal and peat moss.
   - **b.** along gorges.
   - **c.** within sedimentary rock strata.
   - **d.** encased in volcanic rocks.
   - **e.** in amber.

2. Which of the following is *least* likely to leave a fossil?
   - **a.** a soft-bodied land organism such as a slug
   - **b.** a marine organism with a shell such as a mussel
   - **c.** a vascular plant embedded in layers of mud
   - **d.** a freshwater snake
   - **e.** a human

3. Which of the following is the best example of structures that are homoplasies?
   - **a.** forelimbs of bat and mole
   - **b.** silversword plants of Hawaii
   - **c.** four-chambered hearts of birds and bats
   - **d.** skulls of apes and humans
   - **e.** hindlegs of Australian and North American moles

4. Which of these taxonomic units are comparable across all eukaryotes?
   - **a.** genera
   - **b.** species
   - **c.** phyla
   - **d.** kingdoms
   - **e.** All of these units are determined by set criteria, and all are comparable across different lineages of eukaryotes.

5. Which of the following is a shared derived character for marsupial and eutherian mammals?
   - **a.** parental care
   - **b.** internal fertilization
   - **c.** amniote egg
   - **d.** production of milk for young
   - **e.** complete embryonic development inside a uterus

6. Related families are grouped into the next-highest taxon called a
   - **a.** class.
   - **b.** order.
   - **c.** phylum.
   - **d.** genus.
   - **e.** kingdom.

7. Which of the following would provide the best data for determining the phylogeny of three very similar extant species?
   a. the fossil record
   b. a comparison of embryological development
   c. a quantitative analysis of their morphological similarities and differences
   d. comparisons of DNA sequences
   e. comparison of their taxonomic assignments

8. Convergent evolution may result
   a. when older structures are co-opted for new functions.
   b. when homologous structures are adapted for different functions.
   c. from adaptive radiation.
   d. when species are widely separated geographically.
   e. when species have similar ecological roles.

9. Which of the following is the best description of our current hypothesis of the universal tree of life?
   a. The tree of life consists of three great domains: Bacteria, Archaea, and Eukarya.
   b. The base of the tree of life is still uncertain because the molecular clock is not accurate for evolutionary events that occurred that long ago.
   c. The Domain Archaea is known to be the first branch; domains Bacteria and Eukarya are more closely related to each other.
   d. There was substantial horizontal gene transfer and perhaps even fusion of different organisms during the early history of life.
   e. Both a and d represent our current hypothesis.

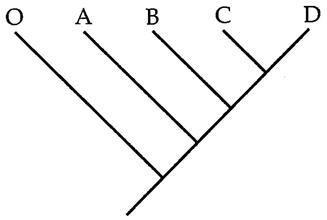
10. How could a molecular clock be useful for studying orthologous genes?
    a. It could help determine when the original gene duplication occurred.
    b. It could help date the divergence of different lineages that all contain that homologous gene.
    c. Because orthologous genes do not have neutral mutations, the molecular clock could not be used to study them.
    d. It could date when the gene first arose.
    e. Because the evolution of orthologous genes does not occur at a consistent rate, a molecular clock would not be useful.

11. Shared derived characters are
    a. those that characterize all the species on a branch of a dichotomous phylogenetic tree.
    b. determined through a computer comparison of orthologous genes in a group of species.
    c. homologous structures that develop during adaptive radiation.
    d. characters found in the outgroup but not in the species to be classified.
    e. used to identify species but not higher taxa.

12. A comparative study of which of the following would provide the best data on the early evolution of fungi and plants?
    a. mitochondrial DNA
    b. DNA from a chloroplast
    c. the amino acid sequence of chlorophyll
    d. DNA for ribosomal RNA
    e. the morphology of present-day specimens

13. A comparative study of which of the following would provide the best data on the ancestry of people from Germany, Italy, and Spain?
    a. mitochondrial DNA
    b. orthologous genes
    c. ribosomal RNA
    d. paralogous genes
    e. the fossil record

14. A taxon such as the class Reptilia, which does not include its relatives, the birds, is
    a. really an order.
    b. a clade.
    c. monophyletic.
    d. polyphyletic.
    e. paraphyletic.

15. 

    According to this phylogenetic tree created using cladistic analysis, C and D are most closely related because they
    a. do not share a common ancestor with O, A, or B.
    b. are monophyletic.
    c. evolved from a common ancestor a long time ago.
    d. have the most shared derived characters in common.
    e. have the youngest species.

16. A biologist is studying the evolution of four similar species of birds. Which approach would allow her to choose the best phylogenetic tree from all possible phylogenies?

- **a.** Draw the simplest cladogram and choose that one.
- **b.** From a comparison of nucleotide sequences, determine the number of evolutionary events required for each tree and choose the most parsimonious tree.
- **c.** Compare the entire genomes of each species; the two most similar genomes are the two species that are most closely related.
- **d.** Determine which species can interbreed; those that can evolved from a common ancestor most recently.
- **e.** Choose the tree that has the most evolutionary changes required as the most probable explanation for why these similar birds have evolved into four distinct species.

TRUE OR FALSE: *Indicate T or F, and then correct the false statements.*

_______ 1. A monophyletic taxon includes only species that share a common ancestor.

_______ 2. The more the sequences of amino acids in homologous proteins vary, the more recently the two species have diverged.

_______ 3. Phylogenetic trees determined on the basis of similar structures may be inaccurate when adaptive radiations have created large differences or when convergent evolution has created misleading analogies.

_______ 4. Phylogenetic trees designed with the concept of parsimony should show the most likely evolutionary relationships.

_______ 5. The strongest support for a phylogenetic hypothesis comes from the concurrence of molecular, morphological, and fossil evidence.

_______ 6. Phlyograms are phylogenetic trees that indicate chronological branch points as found in the fossil record.