Chapter 26: Phylogeny and the Tree of Life

1. Key Concepts Pertaining to Phylogeny
2. Determining Phylogenies
3. Evolutionary History Revealed in Genomes
1. Key Concepts Pertaining to Phylogeny
PHYLOGENY – the evolutionary history of species or group

TAXONOMY – system by which organisms are named and classified

• based on morphological and molecular (DNA, proteins) data

• not necessarily based on evolutionary history, though it can be

TAXONOMIC HIERARCHY – series of nested groupings used to categorize organisms

• developed by Carolus Linnaeus of Sweden in the mid-18th century and still used today
TAXON – a specific category or level in the taxonomic hierarchy (e.g., “class” or “family”)

BINOMIAL NOMENCLATURE – how organisms are named

- genus & specific epithet (e.g., *Homo sapiens*, *Escherichia coli*, *Tyrannosaurus rex*)
  - both terms are “latinized” and written in italics
  - the specific epithet is an adjective and is not capitalized
  - the genus is a noun and is capitalized
- also developed by Carolus Linnaeus
Phylogenetic Trees

A **PHYLOGENETIC TREE** is a branching diagram representing a hypothesis of evolutionary relationships

- each branch point indicates divergence from a common ancestor
- shows patterns of descent, NOT morphological similarity

**SISTER TAXA** – groups that share an immediate common ancestor (e.g., wolf and coyote)
**BASAL TAXON** – earliest lineage to diverge from the remainder of a group

**ROOTED** – if all taxa in a group have the same common ancestor they are said to be “rooted”

**POLYTOMY** – more than two lineages diverge from a single branch point (due to uncertain divergence)
Homology vs Analogy

**HOMOLOGY** – similarities due to common ancestry

- **Molecular Homology** – similarity in DNA, protein sequences
- **Morphological Homology** – skeletal & fossil similarities

**ANALOGY** – morphological similarity due to convergent evolution (e.g., bird vs bat wings)

- analogous structures that arise independently are referred to as homoplasies
2. Determining Phylogenies
All available information, molecular, morphological and physiological, are used in determining phylogenetic relationships.

When molecular and morphological data contradict, molecular data takes precedence.
Using DNA Homology to Construct Phylogenies

Sequences in non-conserved regions of DNA (i.e., sequences not subject to Natural Selection) are aligned since they are presumably free to mutate

- computers are used to find the best alignment accounting for deletions, insertions, etc

The degree of homology in similar regions can be used to construct a phylogeny

- greater homology = more recent common ancestor

NOTE: Random chance would result in 25% homology between sequences, so true homology must be >25%.
Using Cladistics to Construct Phylogenies

**Clade** – an ancestral species and all descendants

- all descendants have a single common ancestor
- all members of clade have *unique* shared derived characters
- clades are a monophyletic groups

**Shared Ancestral Characters (SAC)**

- characters that originated before the clade

**Shared Derived Characters (SDC)**

- characters found only within the clade

(a) Monophyletic group (clade)
PARAPHYLETIC group – consists of a single common ancestor and some (not all) of its descendants

POLYPHYLETIC group – consists of organisms with more than 1 common ancestor
Constructing phylogenies based on shared derived characters (SDCs):

- start with an **outgroup** (group that diverged before the lineage of interest) and construct the phylogeny based on multiple SDCs
- a character table can be used to identify successive branch points, each being due to a SDC found in successively fewer members of the phylogeny
Using the Principle of “Maximum Parsimony” to Construct Phylogenies

A phylogeny should be constructed based on maximum parsimony, i.e., the simplest solution that is consistent with the facts.

- this is the principle of “Occam’s Razor” in which the hypothesis with the fewest assumptions is favored

Applying the principle of maximum likelihood to phylogenies based on DNA sequences:

- the hypothesis that requires the least number of mutational events is favored
Technique

Species I  Species II  Species III

Three phylogenetic hypotheses:

1. Species I
2. Species II
3. Species III

Results

6 events
7 events
7 events
Phylogenetic bracketing is the process of inferring unknown characters in extinct species based on the principle of maximum parsimony:

- In this example the shared characters of crocodiles and birds are also attributed to 2 groups of dinosaurs.
- This example of phylogenetic bracketing exhibits maximum parsimony since the simplest explanation for these shared characters would be a single common ancestor from which all 4 groups evolved.
Evidence Supporting a Case of Phylogenetic Bracketing

This fossil supports a prediction based on phylogenetic bracketing – that dinosaurs, like crocodiles and birds, engaged in *brooding*, the bodily warming of eggs.

(a) Fossil remains of *Oviraptor* and eggs

(b) Artist’s reconstruction of the dinosaur’s posture based on the fossil findings
Phylogenetic Branch Lengths can be used to Indicate the Degree of Genetic Differences...
Phylogenetic Branches can also Indicate Time

Millions of years ago:

- PALEOZOIC: 542
- MESOZOIC: 251
- CENOZOIC: 65.5
- Present

Species:
- Drosophila
- Lancelet
- Zebrafish
- Frog
- Chicken
- Human
- Mouse
3. *Evolutionary History Revealed in Genomes*
Orthologous vs Paralogous Genes

ORTHOLOGOUS GENES – homologous genes in different species

PARALOGOUS GENES – homology between different genes in the same species due to duplication

Formation of orthologous genes: a product of speciation
Ancestral gene
Ancestral species
Speciation with divergence of gene
Orthologous genes
Species A
Species B

Formation of paralogous genes: within a species
Ancestral gene
Species C
Gene duplication and divergence
Paralogous genes
Species C after many generations

Gene families result from repeated duplication of an original gene followed by subsequent mutation.
Molecular Clocks

Certain genes or other regions of genomes tend to change (i.e., evolve) at a constant rate and thus serve as a **molecular clock**.

For example, the number sequence differences (i.e., nucleotide substitutions) in **orthologous** genes can reveal:

- the amount of time that has passed since 2 species diverged from a common ancestor
- the amount of time that has passed since the gene was duplicated

This is best determined with *neutral* mutations (mutations that are neither selected for or against).
Problems with Molecular Clocks

A molecular clock will *not* give accurate predictions if the rate of mutation is not constant which will be the case if:

- there is natural selection (i.e., a mutation is not neutral)
- changes in selective factors over time (i.e., under some conditions a mutation may be neutral and under other conditions it may not be)
- for these or other reasons the mutation rate may change over time

One can increase the number of genes and taxa analyzed as molecular clocks to dilute any inaccuracies associated with any particular gene, however the problems indicated above can never be entirely eliminated.
Current Model of the “Tree of Life”

• this model, as with previous models, will no doubt change over time as more information pertaining to current and extinct species becomes available

• as new information becomes available, we get closer and closer to the true evolutionary relationships among all species, both past and present