Classification and phylogeny
Chapter 2

Phylogeny defined
- Phylogeny is the history of descent of a group of taxa such as species from their common ancestors, including the order of branching and sometimes the absolute times of divergence
- Or the evolutionary history of a group

Other terms defined
- Systematics: Classification of organisms
- Taxonomy: Naming of organisms

Phylogeny is represented by a phylogenetic tree

How can we find phylogenetic history?
- Phylogenetic trees are based on comparison of traits - individuals with common traits are placed together
- Using characters
  - Phenotypic – external and internal morphology
  - Behavior
  - Cell structure
  - Biochemistry
  - DNA
Creating a phylogenetic tree

- Character states or traits are different possible forms of each character
  - Example: character - flower color, character state – blue or red
  - Example: character – nucleotide, character state - T

Character states or traits

- Character states inherited from a common ancestor are termed homologous
- Character states that differ from the ancestor are termed derived
- Character states that are the same as the ancestor are termed ancestral
- Phylogenetic inference based on synapomorphy = shared, derived character states
  - This phylogenetic tree is called a cladogram

Synapomorphies indicated by bars on cladogram

How can we tell which character state is derived and which is ancestral?

- Use an outgroup, a closely related group that is distantly related
  - Can use fossil evidence of divergence time
  - Can use a group with many different character states from your group of interest
The problem of homoplasy

- Homoplasy complicates the building of phylogenetic trees
- Homoplasy is the possession by two or more groups of a similar or identical character state that has not been derived by both species from their common ancestor; includes convergence, parallel evolution and evolutionary reversal

Convergent evolution

- Similar character states evolve independently in different lineages due to similar natural selection pressures

Parallel evolution

- Similar character states evolve independently in related lineages

Evolutionary reversal

- Reversals ("back mutations") can remove synapomorphies

Solution to homoplasy

- Use slowly evolving characters
- Use multiple lines of evidence
  - Morphology
  - DNA sequences

Homoplasy
Which tree is correct?

- Use principle of parsimony
  - The simplest explanation is the best explanation
  - Most widely used method, but not perfect
  - Best tree is the one that has the fewest evolutionary changes

Tree length and maximum parsimony

Methods other than parsimony

- Other methods are better
  - Neighbor-joining method
  - Maximum likelihood method
  - Bayesian method
### Phylogeny exercise

<table>
<thead>
<tr>
<th>Derived trait</th>
<th>Taxon</th>
<th>Jaws</th>
<th>Lungs</th>
<th>Claws or nails</th>
<th>Feathers</th>
<th>Fur</th>
<th>Mammary glands</th>
<th>Four-chambered heart</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hagfish</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Perch</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Salamander</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Lizard</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Crocodile</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Pigeon</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Mouse</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Chimpanzee</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

### Create a phylogenetic tree

- Species 1: GCTGATGAGT
- Species 2: ATCAATGAGT
- Species 3: GTTGCAACGT
- Species 4: GTCAATGACA

### Solution

- Species 1: ATGGACT
- Species 2: ATGGTCT
- Species 3: ATTCAGT
- Species 4: CGTCTCT
- Species 5: CGTCTCA

### Relative evolutionary time

**Ancient events**

**Recent events**

### Create a phylogenetic tree

![Phylogenetic tree diagram]
Whale evolution

• Highly evolved for aquatic life
• Features:
  – Torpedo shape, Fins
  – Migration of Nostrils to top of head
  – No external ear
• Clearly Mammalian
  – Warm Blooded
  – Mammary Glands
  – Hair in fetuses

Whale evolution

• Earliest Mammals are Terrestrial
  → Whales evolved from terrestrial ancestor
• Predictions:
  – Fossil Record should reveal transitional forms between terrestrial and aquatic
  – Evidence from ALL other biological disciplines should also reveal similarities between whales, other mammals, and their ancestors
• Evidence from different fields can be considered independent lines of evidence

Whale evolution

• Based on skeletal characters, whales are close relatives of ungulates
  – Two major groups
    • Perissodactyls (horses and rhinos)
    • Artiodactyls (cows, deers, hippos, pigs, peccaries and camels)

Of extant mammals, who is closest ancestor?

• Construct phylogeny based on morphological data

Astragalus
Of extant mammals, who is closest ancestor?

- Construct phylogeny based on DNA sequence and parsimony

Beta cassein gene

Digression: parasitic genes

- In animals, much DNA is non-coding, "fossil DNA"
  - Repeat sequences = 43.95% of total content of human chromosome 6
- After losing function, such genes often replicate and insert into new locations
- These are inherited along with coding genes
  - INterspersed Elements
    - Short length = SINE
    - Long = LINE

What are the advantages of using SINE's and LINE's in phylogenetic analysis?

- They are selectively neutral, so similarity should represent synapomorphy and not convergence, but some reverse mutation might occur
- Not likely that independent lineages gain insertions in same location

Presence (1) or absence (0) of specific LINES and SINES

<table>
<thead>
<tr>
<th>Locus</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cow</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Deer</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Whale</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Hippo</td>
<td>0</td>
<td>?</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Pig</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Camel</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>
Of extant mammals, who is closest ancestor?

- Construct phylogeny based on DNA sequence and distance methods

Genetic distance matrix

<table>
<thead>
<tr>
<th></th>
<th>Cow</th>
<th>Deer</th>
<th>Whale</th>
<th>Hippo</th>
<th>Pig</th>
<th>Peccary</th>
<th>Camel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deer</td>
<td>0.073</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Whale</td>
<td>0.150</td>
<td>0.197</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hippo</td>
<td>0.148</td>
<td>0.197</td>
<td>0.053</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pig</td>
<td>0.264</td>
<td>0.270</td>
<td>0.197</td>
<td>0.217</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peccary</td>
<td>0.340</td>
<td>0.412</td>
<td>0.268</td>
<td>0.287</td>
<td>0.129</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Camel</td>
<td>0.284</td>
<td>0.347</td>
<td>0.216</td>
<td>0.236</td>
<td>0.291</td>
<td>0.340</td>
<td></td>
</tr>
<tr>
<td>Outgroup</td>
<td>0.306</td>
<td>0.340</td>
<td>0.241</td>
<td>0.261</td>
<td>0.311</td>
<td>0.306</td>
<td>0.210</td>
</tr>
</tbody>
</table>

- “Distance” is calculated as 1 - similarity (r)

What does fossil record say?

- Numerous fossil whales now found
- Many intermediates
  - Traces loss of legs
  - Evolution of ear & echolocation
  - Migration of nasal passage

Ambulocetus, 48 mya
**Rodhocetus 46 mya**
- Well-developed hip bones
- Legs somewhat functional
- Muscle attachment points to vertebrae suggest powerful tail muscles for swimming
- Nostrils higher on skull

**Dorodon 36 mya**

**Hot off the presses**
- New ungulate fossil described in December 2007 as missing link between whales and land-dwelling mammals
- Similar skull and ears
  - *Indohyus* (48 mya)

**Significance of alternative character sets**
- Means to independently test hypotheses of species’ derivation
- Classes of data
  - Fossil record (combined with geologic dating)
  - Morphology
  - Behavior
  - Protein (a.a. sequence) similarity
  - Nuclear DNA sequence similarity
    - Genes and non-coding
    - Non-nuclear DNA sequence similarity

**A summary of whale evolution (from 2001)**

The Origin of Whales and the Power of Independent Evidence

http://www.talkorigins.org/features/whales/
Molecular clocks

- DNA sequences may evolve at a constant rate
- This "molecular clock" may allow us to estimate the absolute time of divergence
  - Clock will vary from gene to gene, lineage to lineage and base to base

Evidence for a molecular clock

- Sequence differences evolve most slowly at second base positions within codons.
Using molecular clocks

- How long ago since humans and chimpanzees have diverged?

- Average rate of base pair substitutions in a lineage can be estimated if there is an absolute time of divergence
  - 310 base pair changes on the lineage from Old World Monkeys (Cercopithecidae) and Homo
  - Oldest fossils of Cercopithecidae are dated at 25 mya

- Average rate for Rhesus monkey lineage = 457/25 my = 1.83 x 10^{-3} per my
- Average rate for Homo line is 310/25 my = 1.24 x 10^{-3} per my
- Average rate is 1.54 x 10^{-3} per my

- D = 2rt
- D = proportion of base pairs that differ between sequences
- r = rate of divergence per base pair per my
- T = time (what we want to know)

- D = 2rt
- t = D/2r
- t = 0.0256/2 * 0.00154
- t = 8.3 my

Thoughts for consideration:

- Descent with modification applies to “species”
- Descent with modification applies to “genes”
- Out of thousands of genes, different species can share some identical genes from a common ancestor
- Gene and species lineages have a degree of independence
  - Not every gene has to change as new species are created
Difficulties in phylogenetic analysis

- Scoring characters is difficult
  - Independence of characters is not known, i.e. characters may change together
- Homoplasy is common

Difficulties

- Evolution can erase past evolution
  - Divergence time very distant
  - Evolution very quick

Rapidly evolving sequences are successful for taxa that have diverged recently

Slowly evolving sequences are successful for taxa that diverged in the distant past

Difficulties

- Some events (such as adaptive radiation) happen too quickly to develop distinct synapomorphies

Hybridization (reticulate evolution) may occur
Horizontal gene transfer may occur

Biological evolution is defined as
1. change in the properties of an organism over the course of development.
2. directed change toward some fixed goal.
3. the belief that each organism has an immutable essence.
4. change in the properties of groups of organisms over the course of generations.
5. a completely random biological process.

Darwin’s biological views were influenced by
1. the writings of Gregor Mendel.
2. his work as a naturalist aboard the H.M.S. Beagle.
3. decades of military service.
4. the theory of plate tectonics.
5. correspondence with Ernst Mayr.

Which of the following principles of evolution is true?
1. Mutation alone is sufficient to cause large allele frequency changes.
2. Genotype and phenotype are synonymous.
3. Hereditary variation can be maintained because of the particulate nature of genes.
4. Acquired characteristics are inherited.
5. Natural selection needs to be strong for there to be substantial evolutionary change.

Which of the following statements is true?
1. Natural selection acts on populations, not on individuals.
2. Evolutionary change is a population process.
3. Evolutionary change is an organismal process.
4. Developmental change is a population process.
5. Species are simply those organisms that share similar phenotypes.

Which of the following does not cause complications in inferring a phylogeny?
1. Homoplasy
2. The presence of synapomorphies (shared derived characteristics)
3. Variation in rates of evolution for different taxa
4. Convergent evolution
5. Absence of DNA sequence information for many extinct taxa
Which of the following is an accepted rule of species nomenclature?

1. Names need not have meaning.
2. The genus and species epithet need not agree in gender.
3. The name of a species is always in boldface type.
4. Taxonomic names cannot be revised.
5. No two species of animals or plants can bear the same name.

A situation in which lineages diverge too rapidly for a phylogeny to be constructed is called

1. synaptomorphy.
2. homoplasy.
3. haplotype divergence.
4. DNA sequence polymorphism.
5. adaptive radiation.