Phylogenetics, Cladistics, Ontogeny, Ontology
“Ontogeny Recapitulates Phylogeny”

- Ernst Haeckel’s (1834-1919) theory that
  - The development of the individual organism
    - (its **ontogeny**)
      - onto- = existence, gen=development
  - ...is a speeded-up pseudo-copy
    - (**recapitulation** =“restatement”)
  - ...of the evolutionary history of its species
    - (**phylogeny**: phylo- =tribe, kind, phylum)
“Ontogeny Recapitulates Phylogeny” – In Perspective

- It’s cool
- It frequently works, sort of
- It’s explainable as follows:
  - Mutations can affect *earlier* or *later* developmental stages
  - Later ones are more likely to be viable
    - viable?
  - ... because later ones tend to have less dramatic (therefore risky) effects
Haeckel’s (in)famous drawing, 1866
+
modern pics (from National Geographic)
2nd-day chick embryo

Source:
http://www.talkorigins.org/faqs/wells/haeckel.html#chicken_embryo

from
“Wells and Haeckel's Embryos: A Review of Chapter 5 of *Icons of Evolution*”
by PZ Myers
Parrot chick embryo

Source:
AvianWeb.com
http://www.avianweb.com/images/birds/parrots/eggtoparrot.jpg

By permission of www.theparrotsocietyuk.org
"A single cell, less than half a millimeter across, develops into a baby elephant weighing more than 260 pounds. (National Geographic)"
Fetal Elephant, 16 weeks or 6 mo.

“At 16 weeks, the elephant fetus starts to look more like an elephant as the trunk develops. After about a year in the womb, the trunk grows longer than the legs. National Geographic”

“...image of an unborn elephant...approximately six months into the two-year gestation period...”
http://www.dailymail.co.uk/pages/galleries/index.html?in_gallery_id=9136&in_image_id=302394&in_page_id=1055, as of 10/7/07
Fetal Elephant, 12 mo.

“(National Geographic)”

About 18 inches long and 26 pounds
(per blog.wired.com/tableofmalcontents/2006/11/ as of 10/8/07)
Fetal Elephant, 12 months

“National Geographic Channel premiers a new documentary on Dec. 10 [2006] at 8 p.m. The documentary takes viewers inside the wombs of elephants, dolphins and dogs to trace their different paths from conception to birth. At 12 months, an elephant fetus is on average 18 inches long and weighs approximately 26 pounds. It can use its trunk, curling it right up into its mouth and over its head. (National Geographic)” - [http://abcnews.go.com/](http://abcnews.go.com/), 11/23/06
Fetal Dog

“Computer-generated images and intricately designed models showed off these animal fetuses at various stages of gestation. This photo shows a dog at very early stages of gestation. (National Geographic) ”
“At 39 days, the eyelids of a dog fetus fuse to protect the developing eyes. A golden retriever (pictured here) pregnancy lasts about nine weeks. All 400 breeds of dogs can produce multiple puppies in one litter. (National Geographic) ”
“At about 45 days the dog's nose is clearly visible. Approximately a week later the golden retriever fetus will have a full coat of light cream hair, and soon after the whiskers become visible as well. (National Geographic) ”
Why Infamous Instead of Famous?

Haeckel’s drawings were inaccurate...
The inaccuracies favored his theory, of course!

“The intentional ‘fraud’...in Haeckel's drawings is used by some creationists...against...evolution. However...the evidence is real but it exists in the embryos, not in Haeckel's drawings.”
http://en.wikipedia.org/wiki/Embryo_drawings, 10/15/06

“In 2000, Stephen Jay Gould...wrote: ‘We...have the right to be...astonished and ashamed by the...mindless...persistence of these drawings in...modern textbooks.’”
http://en.wikipedia.org/wiki/Embryo_drawings, 10/15/06
Haeckel’s distortions were used to back up (failed) anti-evolution legislation in Arkansas

http://www.talkorigins.org/faqs/wells/haeckel.html
What's Wrong with Haeckel's Idea?

- Mammalian “gill slits” have nothing to do with gills
- The period preceding the similar-looking stage differs across species
- His drawings were hand-drawn and distorted
- There *are* basic similarities in all vertebrate developmental programs
  - And that's why they look similar
- Haeckel was more Lamarkian than Darwinist
  - He “wanted” ontogeny to recapitulate phylogeny
What About Post-Natal Similarity?

Let's check out animals on their hind legs
- Similar or not? How about this?
Ontogeny recapitulates phylogeny – sort of

- **Ontogeny**
  - Development of the individual

- **Phylogeny**
  - Development of the species

- **Ontology**
  - Named concepts and relationships in a field

- **Phylology**
  - No such thing (but see “Philo*logy”)

---

gen: development: gene, genus, genealogy, genesis, generation, cladogenesis...

-log-: field of study: biology, sociology, cardiology, ...
Phylogeny (study of species) II: Phylogenetics

- A basic principle:
  - **Similarity** between species correlates with **recency** of their common ancestor
  - ...so, **dissimilarity** correlates with **non-recent** (distant) common ancestry
Approaches in Phylogenetics

- Consider the problem of determining relatedness of species
- **Phenetics:**
  - Examine the similarity of phenotypes (the set of characteristics of organisms)
    - Cat paws, dog paws, human hands, chimp hands, fish fins
    - We could make a **cladogram** from this...
- **Cladistics:**
  - Focus on “new,” emergent characteristics
- **Evolutionary Systematics:**
  - A combination of phenetics and cladistics
We know that speciation occurs by **cladogenesis**: one species bifurcates (splits into two).

A “**clade** is made up of an ancestral species and all its descendants” - *New Scientist, 11 Sep 2004, p. 13*

- Do two species always form a clade?
- Do two species ever form a clade?
- (How many “clad-” words have we seen?)
How to do Cladistics

- Examine morphologies of organisms
  - What is morphology?
- Do appropriate biochemical assays in a lab
  - Why might this work?
- Do sequence similarity analyses
  - Bioinformaticists/-ians can do this
  - Similarity can be objectively measured
  - Haeckel’s hand drawings were less objective!
Phylogenetic Trees are Dendrograms

- Trees have a starting point
  - The “root”
- They have nodes
  - A node represents a species
- They have arcs (links, edges, “branches”, “lines”)
  - Arcs show evolution of one species into another
Phylogenetic Trees

http://www.acedb.org/Software/whelp/Dendrogram.html
A Phylogenetic Tree and Dendrogram

- http://www.mju.ac.th/biotech/dendrogram.jpg
Another Dendrogram
(not a phylogenetic tree)

http://149.170.199.144/multivar/images/dend5.gif
The following is a dendrogram showing the amount of protein sequence similarity among all known human chemokines.

Dendrograms V

itsa.ucsf.edu/~gram/home/gpcr/figure-2.gif

Where is the root?
Dendrograms VI

- free-living heterotrophic flagellates

http://www.bio.usyd.edu.au/Protsvil/dendrogram.jpg
Dendrograms – what if...?

- Dendrograms are a kind of tree.
- Is it possible for an ancestral species to split into 3 new species?
- What dendrogram would that lead to?
- “Cladograms” are binary.
- What does that say about cladograms as a model of evolutionary branching?
Dendrograms VIII

- Urban life in 2020

- Figure 2: Categorising the industrial cities.

More Dendrogram Concepts

- Dendrograms may be **binary** or not
  - Which of the foregoing are binary?
- Nodes without descendants are **leaves**
  - (also **terminal nodes** or **tips**)
- Cladistic theory: species **bifurcate**
  - So **binary** dendrograms are also called
    - **Cladograms**
- Let’s identify a few **clades** in the foregoing
Cladogram-Specific Concepts

- Arc (branch) length may mean something or not
  - Length can (if desired) be used to indicate “time”
    - or more accurately, evolutionary distance
      - or at least, sequence dissimilarity
        - (e.g., number of mutations)
  - If length is made to be significant...
  - the cladogram can be called a phylogram
- Length might also not mean anything
- Which foregoing dendrograms are...
  - Cladograms? Phylograms?
Data Structures Recap

- Graphs are networks of nodes and links
- Trees are graphs with no cycles or cross links
- Dendrograms are trees, with a clustering flavor
- Cladograms are binary dendrograms
- Phylograms are cladograms where link length represents time
- (Phylograms are named after phylogenetic trees)
Making a Cladogram

- Start with a similarity or distance table
  - Which is which below?

<table>
<thead>
<tr>
<th></th>
<th>Organism1</th>
<th>Organism2</th>
<th>Organism3</th>
<th>Organism4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organism1</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organism2</td>
<td>85</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organism3</td>
<td>55</td>
<td>80</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Organism4</td>
<td>70</td>
<td>65</td>
<td>40</td>
<td>100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Organism5</th>
<th>Organism6</th>
<th>Organism7</th>
<th>Organism8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organism5</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organism6</td>
<td>85</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organism7</td>
<td>55</td>
<td>80</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Organism8</td>
<td>70</td>
<td>65</td>
<td>40</td>
<td>0</td>
</tr>
</tbody>
</table>
Making the Cladograms

How might these tables be fully filled in?

<table>
<thead>
<tr>
<th>Organism1</th>
<th>Organism2</th>
<th>Organism3</th>
<th>Organism4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organism1</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organism2</td>
<td>85</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Organism3</td>
<td>55</td>
<td>80</td>
<td>100</td>
</tr>
<tr>
<td>Organism4</td>
<td>70</td>
<td>65</td>
<td>40</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Organism5</th>
<th>Organism6</th>
<th>Organism7</th>
<th>Organism8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organism5</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organism6</td>
<td>85</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Organism7</td>
<td>55</td>
<td>80</td>
<td>0</td>
</tr>
<tr>
<td>Organism8</td>
<td>70</td>
<td>65</td>
<td>40</td>
</tr>
</tbody>
</table>
Agglomerative Heirarchical Clustering

- A way to construct phylograms from tables
- Given a set of organisms...
  - Start by partitioning the set into singleton clusters
  - What are these clusters for the tables below?

<table>
<thead>
<tr>
<th>Organism1</th>
<th>Organism2</th>
<th>Organism3</th>
<th>Organism4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organism1</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organism2</td>
<td>85</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Organism3</td>
<td>55</td>
<td>80</td>
<td>0</td>
</tr>
<tr>
<td>Organism4</td>
<td>70</td>
<td>65</td>
<td>40</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Organism1</th>
<th>Organism2</th>
<th>Organism3</th>
<th>Organism4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organism1</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organism2</td>
<td>85</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Organism3</td>
<td>55</td>
<td>80</td>
<td>100</td>
</tr>
<tr>
<td>Organism4</td>
<td>70</td>
<td>65</td>
<td>40</td>
</tr>
</tbody>
</table>
Clusters of Organisms

A singleton cluster consists of one organism

For example, circle one in the tables below!

<table>
<thead>
<tr>
<th></th>
<th>Organism1</th>
<th>Organism2</th>
<th>Organism3</th>
<th>Organism4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organism1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organism2</td>
<td>85</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organism3</td>
<td>55</td>
<td>80</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Organism4</td>
<td>70</td>
<td>65</td>
<td>40</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Organism1</th>
<th>Organism2</th>
<th>Organism3</th>
<th>Organism4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organism1</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organism2</td>
<td>85</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organism3</td>
<td>55</td>
<td>80</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Organism4</td>
<td>70</td>
<td>65</td>
<td>40</td>
<td>100</td>
</tr>
</tbody>
</table>
Agglomerative Heirarchichal Clustering II

- Find the most similar two clusters
  - What are they for each table below?
- Merge them into one and add to phylogram
- Repeat until everything is in one cluster

<table>
<thead>
<tr>
<th>Organism1</th>
<th>Organism2</th>
<th>Organism3</th>
<th>Organism4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organism1</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organism2</td>
<td>85</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Organism3</td>
<td>55</td>
<td>80</td>
<td>0</td>
</tr>
<tr>
<td>Organism4</td>
<td>70</td>
<td>65</td>
<td>40</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Organism1</th>
<th>Organism2</th>
<th>Organism3</th>
<th>Organism4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organism1</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organism2</td>
<td>85</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Organism3</td>
<td>55</td>
<td>80</td>
<td>100</td>
</tr>
<tr>
<td>Organism4</td>
<td>70</td>
<td>65</td>
<td>40</td>
</tr>
</tbody>
</table>
Agglomerative Heirarchical Clustering: *single linkage method*

- What is the similarity/diff. between clusters?
  - You can read it from the table only at first
  - “Single linkage method”: if a & b in a cluster,
    - $\text{distance}(ab,c) = \min(\text{distance}(a,c), \text{distance}(b,c))$

<table>
<thead>
<tr>
<th>Organization1</th>
<th>Organization2</th>
<th>Organization3</th>
<th>Organization4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organization1</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organization2</td>
<td>85</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Organization3</td>
<td>55</td>
<td>80</td>
<td>0</td>
</tr>
<tr>
<td>Organization4</td>
<td>70</td>
<td>65</td>
<td>40</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Organization1</th>
<th>Organization2</th>
<th>Organization3</th>
<th>Organization4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organization1</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organization2</td>
<td>85</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Organization3</td>
<td>55</td>
<td>80</td>
<td>100</td>
</tr>
<tr>
<td>Organization4</td>
<td>70</td>
<td>65</td>
<td>40</td>
</tr>
</tbody>
</table>

Let’s try it!
Agglomerative Heirarchical Clustering: *complete linkage method*

“Complete linkage method”:
- if a & b in a cluster,
  - distance(ab,c) = \textbf{max}(distance(a,c), distance(b,c))

<table>
<thead>
<tr>
<th>Organism1</th>
<th>Organism2</th>
<th>Organism3</th>
<th>Organism4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organism1</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organism2</td>
<td>85</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Organism3</td>
<td>55</td>
<td>80</td>
<td>0</td>
</tr>
<tr>
<td>Organism4</td>
<td>70</td>
<td>65</td>
<td>40</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Organism1</th>
<th>Organism2</th>
<th>Organism3</th>
<th>Organism4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organism1</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organism2</td>
<td>85</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Organism3</td>
<td>55</td>
<td>80</td>
<td>100</td>
</tr>
<tr>
<td>Organism4</td>
<td>70</td>
<td>65</td>
<td>40</td>
</tr>
</tbody>
</table>

*Let’s try it!*
Agglomerative Hierarchical Clustering: *average linkage method*

“Average linkage method”:
- if a & b in a cluster,
  - \( \text{distance}(ab,c) = \text{average}(\text{distance}(a,c), \text{distance}(b,c)) \)
- But what does “average” mean?

<table>
<thead>
<tr>
<th>Organism1</th>
<th>Organism2</th>
<th>Organism3</th>
<th>Organism4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organism1</td>
<td>0</td>
<td>85</td>
<td>55</td>
</tr>
<tr>
<td>Organism2</td>
<td>85</td>
<td>0</td>
<td>40</td>
</tr>
<tr>
<td>Organism3</td>
<td>55</td>
<td>80</td>
<td>100</td>
</tr>
<tr>
<td>Organism4</td>
<td>70</td>
<td>65</td>
<td>40</td>
</tr>
</tbody>
</table>

Let’s try it!
### Average linkage methods

**UPGMA**

- = **unw**eighted pair group method using arithmetic mean
- if ab & cde are clusters,
  - distance(ab, cde)
    
    \[
    \text{distance(ab, cde)} = \frac{d(a,c) + d(a,d) + d(a,e) + d(b,c) + d(b,d) + d(b,e)}{6}
    \]

---

<table>
<thead>
<tr>
<th></th>
<th>Organism1</th>
<th>Organism2</th>
<th>Organism3</th>
<th>Organism4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organism1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organism2</td>
<td>85</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organism3</td>
<td>55</td>
<td>85</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Organism4</td>
<td>70</td>
<td>65</td>
<td>40</td>
<td>100</td>
</tr>
</tbody>
</table>

**Idea:** let’s not try it!
Various pair group methods

- UPGMA
  - Unweighted pair group method - arithmetic
- WPGMA
  - Weighted pair group method - arithmetic
- UPGMC
  - Unweighted pair group method - centroid
- WPGMC
  - Weighted pair group method - centroid
Other Methods

- **NJ – Neighbor Joining**
  - Used by ClustalW/X

- **Maximum Parsimony**
  - “distance”
    - = minimum # of mutations needed to convert one sequence to another
  - Basically, the edit distance
    - computationally expensive

- **Maximum Likelihood**
  - Like maximum parsimony but accounts for the fact that some mutations are more likely than others
Major Common Assumptions

- All branches evolve at the same rate
  - But they might not!
    - Some branches evolve slowly (why?)
    - Some evolve fast (why?)
      - Example of a branch that changes fast?

- All subsequences evolve at the same rate
  - Again, not true in general
    - Why not?

- Sequences are aligned optimally
  - To do so is too expensive, so we usually make do with results that are “good” but not guaranteed
Constructing phylogenetic trees:
- possible for very different organisms by sequence evolution methods
- can’t do that using morphology
- E.g. comparing a plant and an animal
Picking sequences to test

- Primates
  - (or other closely related groups)
  - mtDNA is good to use
    - Mitochondrial DNA
    - It evolves very quickly
      - Especially a non-coding region in it

- Very divergent organisms
  - E.g. comparing plants, animals, & bacteria
    - Use rRNA (ribosomal RNA)
      - It changes very slowly