Biological Trees

Gloria Rendon
SC11 Education
June, 2011
Biological trees are used for the purpose of classification, i.e. grouping and categorization of organisms by biological type such as genus or species.
Types of Biological trees

• Taxonomy trees, like the one hosted at NCBI, are hierarchies; thus classification is determined by position or rank within the hierarchy. It goes from kingdom to species.

• Phylogenetic trees represent evolutionary relationships, or genealogy, among species. Nowadays, these trees are usually constructed by comparing 16s/18s ribosomal RNA.

• Gene trees represent evolutionary relationships of a particular biological molecule (gene or protein product) among species. They may or may not match the species genealogy. Examples: hemoglobin tree, kinase tree, etc.
TAXONOMY TREES
Exercise 1: Exploring the Species Tree at NCBI

• There exist many taxonomies.
• In this exercise, we will examine the taxonomy at NCBI.
• NCBI has a taxonomy database where each category in the tree (from the root to the species level) has a unique identifier called taxid.
• The lineage of a species is the full path you take in that tree from the root to the point where that species is located.
• The (NCBI) taxonomy common tree is therefore the tree that results from adding together the full lineages of each species in a particular list of your choice.
Exercise 1: Exploring the Species Tree at NCBI

- Open a web browser on NCBI’s Taxonomy page
- Click on each one of the names here to look up the taxonomy id (taxid) of each one of the five categories of the taxonomy browser: Archaea, bacteria, Eukaryotes, Viroids and Viruses.
Exercise 1: Exploring a full Lineage

Still on the same page.
Click on the link “Homo sapiens” to explore the full lineage of this species.
Q: how long do you think the full lineage is, i.e. how many internal nodes there are between the root and this species?
Exercise 1: Exploring a full Lineage

- Q: how long do you think the full lineage is, i.e. how many internal nodes there are between the root and this species?
  - A: 29 categories

- Q: what is the taxid for this species?
  - A: 9606

- Q: Is this a terminal node (leaf)?
  - A: no, there are two more subspecies
Exercise 1: Species and subspecies

• Go back to NCBI’s Taxonomy page http://www.ncbi.nlm.nih.gov/Taxonomy/

• Use the search box to look up the tax id of these species:
  - Mouse
  - Chicken
  - Escherichia Coli
  - Lemna Minor

• After typing the name in the search box, click on the Go button to perform the search

The NCBI Taxonomy Homepage

These are direct links to some of the organisms commonly used in molecular research projects:
Exercise 1: Species and subspecies

When the name you typed has more than one match in the Taxonomy tree; it will display all results in a page.

Multiple matches are common when the name in the query matches the species and subspecies in the taxonomy as illustrated here.
Exercise 1: Exploring the Taxonomy Common Tree

• Go NCBI’s Taxonomy search page

• In the Search box type several tax ids separated by spaces:
  56636 10090 9031 562 5833 4472

• Then press the Search button
Exercise 1: Exploring the Taxonomy Common Tree

• Each one of these links takes you to the taxonomy page of the corresponding species.

• To see the Taxonomy common tree of these species, click on the Display Settings [a pull down menu should appear].
Exercise 1: Exploring the Taxonomy Common Tree

• From the Pull down menu select “Common Tree” as illustrated here

• Then click on the Apply button
Exercise 1: Exploring the Taxonomy Common Tree

This is the taxonomy common tree for these six species:

- Aeropyrum pernix
- Mouse
- Chicken
- Escherichia Coli
- Plasmodium falciparum
- Lemna Minor

How many categories there are in common between mouse (Mus musculus) and chicken (Gallus gallus)?

Hint: click on “Expand All and then count categories from Eukaryotes downwards.”
Exercise 2: Another, more complete, common tree

• Go back to NCBI’s Taxonomy page

• Construct a Taxonomy Common Tree for the 13 species used to build the three kingdoms of C. Woese:

2214  Methanosarcina
498970  Bacillus_firmus strain QJGY18
187420  Methanothermobacter_thermautotrophicus
1097  Chloribium
83816  Methanobrevibacter_ruminantium
562  Escherichia_coli
39152  Methanococcus_maripaludis
10090  Mus_musculus
48776  Aphanocapsa sp. HBC6
4472  LemnaMinor and LemnaMimorChloroplast
1717  Corynebacterium_diphtheriae
4932  Saccharomyces_cerevisiae
PHYLOGENETIC TREES
A phylogeny, or species/evolutionary tree, represents the evolutionary relationships among a set of organisms or groups of organisms, called taxa (singular: taxon) that are believed to have a common ancestor.
• The tips of the phylogenetic tree represent groups of descendant taxa (often species)
• The internal nodes of the tree represent the common ancestors of those descendents.
• The tips are the present and the internal nodes are the past.
• The edge lengths in some trees correspond to time estimates – evolutionary time.
Sister Groups and a common ancestor

- Two descendents that split from the same node are called sister groups.
- In the trees above, species A & B are sister groups — they are each other's closest relatives; which means that:
  - i) they have a lot of evolutionary history in common and very little evolutionary history that is unique to either one of the two sister species and
  - ii) that they have a common ancestor that is unique to them.
Equivalent trees

- For any speciation event on a phylogeny, the choice of which lineage goes to the right and which one goes to the left is arbitrary.
- These three phylogenies are therefore equivalent.
• Many phylogenies also include an outgroup — a taxon outside the group of interest.
• All the members of the group of interest are more closely related to each other than they are to the outgroup. Hence, the outgroup stems from the base of the tree.
• An outgroup can give you a sense of where on the bigger tree of life the main group of organisms falls. It is also useful when constructing evolutionary trees.
Branches and clades

- Evolutionary trees depict clades.
- A clade is a group of an ancestor and all descendants.
- You can think of a clade as a branch of the tree of life.
- Some examples of clades and non-clades in a phylogenetic tree are shown here.
More on clades. Nested clades

- Clades can be nested within one another
- A clade may include many thousands of species or just a few.
- Some examples of clades at different levels are marked on the phylogenies above.
Types of trees: unrooted vs. rooted

- **A rooted** phylogenetic tree is a tree with a unique root node corresponding to the (usually imputed) most recent common ancestor of all the entities at the leaves (aka tips) of the tree. A rooted tree is a binary tree.

- **Unrooted** trees illustrate the relatedness of the leaf nodes without making assumptions about common ancestry. An unrooted tree has a node with three edges; the rest of the nodes have up to two edges.
Dendrogram, cladogram, phylogram

- Dendrogram is the ‘generic’ term applied to any type of diagrammatic representation of phylogenetic trees. All four trees depicted here are dendrograms.

- Cladogram (to some biologists) is a tree in which branch lengths DO NOT represent evolutionary time; clades just represent a hypothesis about actual evolutionary history. TREE1 and TREE2 are cladograms and TREE1 = TREE2.

- Phylogram (to some biologists) is a tree in which branch lengths DO represent evolutionary time; clades represent true evolutionary history (amount of character change). TREE3 and TREE4 are phylograms and TREE3 ≠ TREE4.
Phylogenetic Trees and classification

• Phylogenetic trees classify organisms into clades.

• By contrast, the Linnaean system of classification assigns every organism a kingdom, phylum, class, order, family, genus, and species.

• The phylogenetic tree depicted here identifies four clades
How evolution shapes the phylootree

--A ladder-like tree (left) would be consistent with a group of HIGHLY divergent species.

--A (nearly) complete binary tree (right) would be consistent with a group of CLOSELY related species.
How evolution shapes the phyloptree

A clade that shows SHORT edge lengths contains species that diverged recently from a common ancestor.

A clade that shows LONG edge lengths contains species that diverged a long time ago from a common ancestor.
How evolution shapes the phylotree

Species are not evenly distributed in the phylotree.

Clades corresponding to large families/genera will be represented by large subtrees.

Clades corresponding to small families/genera will be represented by sparsely populated subtrees.

How evolution shapes the phylotree

Clades corresponding to small families/genera will be represented by sparsely populated subtrees.
How tools/models shapes the phylogenetic tree

Tree 1: Delta and Epsilon branched off early from the rest of the family

Tree 2: Gamma and Beta branched off early from the rest of the family
Additional Readings


• Understanding Evolution http://evolution.berkeley.edu/
