Lecture Outline: Trees

1. Nomenclature

Phylogeny A *phylogeny* is a tree that shows the evolutionary relationships among a group of organisms.

Taxon A *taxon* is a generic name for a taxonomic group. Examples are species, but also populations, genera, families, orders, phyla, and so on. The plural is *taxa*. Each leaf on a phylogenetic tree represents a taxon.

Tree A *tree* is a graph that represents evolutionary relationships among taxa. Here, *graph* is a technical mathematical term that stands for a set of nodes and a set of edges that connect pairs of nodes. A tree is a special graph is connected and has no cycles. A graph is *connected* when there is a path from any node to any other node. A graph has no *cycles* when there is only one unique path (without backtracking) between any two nodes.

Node A *node* in a phylogeny represents a taxon or a common ancestor for a set of taxa. Nodes are also refered to as *vertices*. The *degree* of a node is the number of edges that are connected to it.

Leaf A *leaf* is a node with degree one. In a phylogeny, a leaf usually represents a single present-day taxon. We will typically have a DNA sequence measured for each leaf.

Internal Node An *internal node* is node with degree greater than one. Internal nodes represent common ancestors. We typically do not have DNA data for internal nodes.

Edge An *edge* is the part of a graph that connects two nodes, and is represented by a line. An edge is also called a *branch*. An edge represents the evolutionary transition from an ancestral taxon to a descendant taxon.

Topology A tree *topology* represents all of the evolutionary relationships, but does not represent time or genetic distance. The topology is the graph and the leaf labels. The left/right orientation does not affect the topology.

Root The *root* of a tree is the node that represents the common ancestor of all taxa in the tree.

Rooted Tree A *rooted tree* is drawn so that time is represented by a single direction. This direction may be up, down, right, or sometimes even left, but the root is at the oposite end of the graph from the leaves.

Unrooted Tree An *unrooted tree* is drawn without reference to the direction of time. An unrooted tree represents all of the rooted trees consistent with it. The root would typically be on one of the edges of the tree.

Binary Tree A *binary tree* is a tree that represents an evolutionary history where all speciation events produce two ancestors from one. In a *rooted binary tree*, leaves have degree one, the root has degree two, and all other internal nodes have degree three. In an *unrooted binary tree*, all internal nodes have degree three. Trees that are not binary are said to contain *polytomies*, or nodes with more than two descendants. If there are n leaves, there are $n - 1$ internal nodes in a rooted binary tree and $n - 2$ internal nodes in an unrooted binary tree.

Clade In a rooted tree, a *clade* is a group of leaves that form a *monophyletic group* meaning they have a common ancestor that is not a common ancestor for any other leaf in the tree. In an unrooted tree, a clade would be any group of taxa that can be separated from the rest by removing a single edge.

Split A *split* is a partition of the taxa (leaves) into two nonempty sets. Each edge in a tree represents a split.

Subtree A *subtree* is a subset of a tree that is a tree. Typically, we think of subtrees as defined by a root and all of the descendant nodes and descendent edges, but we can also get subtrees by pruning away nodes and edges.

Edge Length An *edge length* is a number associated with an edge. The number may represent time or it may represent a measure of expected genetic distance.

Ultrametric Tree An *ultrametric tree* is a rooted tree with edge lengths where all leaves are equidistant from the root. Often, ultrametric trees represent *the molecular clock* which states that the rate of mutation is the same across all lineages of the tree.

2. Trees and common ancestry

(a) **Activity 1:** Show example tree and ask questions.

(b) **Activity 2:** Show six example trees: which trees have the same topology?

3. **Unrooted trees**

   (a) **Activity 3:** Sketch all rooted trees consistent with the given unrooted tree.

4. **Labeled histories**

   Labeled history A *labeled history* is a rooted ultrametric tree in which the speciation events occur in a specified order. A single tree topology may have one or many possible labeled histories.

   (a) **Activity 4:** Count labeled histories for two given trees.

5. **Counting**

   (a) There are $1 \times 3 \times \cdots \times (2n-5) \equiv (2n-5)!! \equiv u(n)$ unrooted binary trees with n leaves ($n > 2$).

   (b) There are $1 \times 3 \times \cdots \times (2n-3) \equiv (2n-3)!! \equiv r(n)$ rooted binary trees with n leaves ($n > 2$).

   (c) There are $\dfrac{n!(n-1)!}{2^{n-1}} \equiv \ell(n)$ labeled histories for n taxa.

   (d) Show where each formula comes from.

6. **Probability distributions on trees**

   (a) Bayesian prior probability distribution on trees are often uniform over a set of rooted or unrooted tree topologies.

   (b) A birth process or the coalescent put uniform distributions on labeled histories.

   (c) **Activity 5:** Compare the probabilities under these two distributions for two trees with six taxa.

7. **Probability distributions on clades**

   (a) **Activity 6:** For rooted tree with four leaves, what is the prior probability of a clade with 2 or 3 taxa under uniform rooted topologies?

   (b) **Activity 7:** For rooted tree with five leaves, what is the prior probability of a clade with 2 or 3 taxa under uniform rooted topologies?

   (c) **Activity 8:** For rooted tree with ten leaves, what is the prior probability of a clade with 6 taxa under uniform rooted topologies?

8. **Types of trees**

   Cladogram A *cladogram* is a tree that only represents a branching pattern. The edge lengths do not represent anything.

   Phylogram A *phylogram* is a phylogenetic tree where edge lengths represent time or genetic distance.

   Ultrametric tree An *ultrametric tree* or *chronogram* is a phylogenetic tree where edge lengths represent time (so current taxa are equidistant from the root).