Trees. An undirected tree is an undirected, connected acyclic graph. A directed graph is called a directed tree if its underlying graph (i.e. the graph that is obtained from it by ignoring edge orientation) is an undirected tree. A rooted tree is a tree in which a vertex is distinguished as the root. Normally in a rooted tree, the edges are considered as directed, either away or toward the root.

Theorem 1. *The following statements are equivalent for an undirected graph G:*

1. *G is a tree.*
2. *Every two vertices of G are joined by a unique path.*
3. *G is connected and $|V(G)| = |E(G)| + 1$.*
4. *G is connected, and if we add a new edge between two nonadjacent vertices of G, then the resulting graph contains exactly one cycle.*

We call a vertex a leaf if its degree is one (or outdegree zero, in the case of directed trees). The other vertices are called interior/internal. An unrooted tree is binary if the degree of each interior vertex is three. A rooted tree is binary if the root vertex has degree two, and every other interior has degree three. In other words, if we consider the edges of a rooted tree as being directed away from the root, then it is binary if each interior vertex has exactly outgoing edges.

Theorem 2. *Every undirected tree with at least two vertices has at least two leaves.*

Proof. Let T be an undirected tree. We know that $\sum_{v \in V(T)} \deg(v) = 2|E(T)|$, and that $|E(T)| = |V(T)| - 1$. Thus, $\sum_{v \in V(T)} \deg(v) = 2|V(T)| - 2$. The claim now easily follows. $\square$

X-trees. Let X be a nonempty set of labels. An unrooted X-tree is an ordered pair (T, ϕ), where T is an unrooted tree, and ϕ is a map from X to $V(T)$ in which every vertex of degree at most two is labeled. It is noteworthy to observe that:

1. As ϕ is a map from X to $V(T)$, a vertex of T can have more than one label.
2. By definition, each leaf or a 2-degree vertex must have (at least) a label.
3. The definition does not forbid a vertex with degree larger than two not to have a label.

When no misunderstanding can arise, we refer to an X-tree (T, ϕ) as T for brevity.

A phylogenetic tree on X is an X-tree with the requirement that ϕ is a bijective map from X to the set of leaves of T. In other words, each leaf of a phylogenetic tree has exactly one label, and none of the interior vertices have a label. Note that there cannot be any vertex of 2-degree in a phylogenetic tree.

Theorem 3. *Let T be an undirected binary phylogenetic tree on X. Then T has $2|X| - 3$ edges.*

Rooted X-trees. The definition of a rooted X-tree (rooted phylogenetic tree on X) is entirely similar to that of an unrooted X-tree (unrooted phylogenetic tree on X).

Given a rooted phylogenetic tree tree T, we can obtain an unrooted tree associated with T as follows. Let ρ be the root vertex of T. If $\deg(\rho) \geq 3$, then we disregard ρ as the root, and consider as an ordinary vertex, hence obtaining an unrooted tree. If $\deg(v) = 2$, we then contract ρ—that is, we replace the two edges incident with ρ by a new edge joining the two neighbor vertices of ρ—to obtain an unrooted phylogenetic tree.

Conversely, let T be an unrooted phylogenetic tree on X. Fix a particular element $x \in X$. We remove x along with the edge incident to it, and designate the vertex originally adjacent to x as the root of the resulting tree. More formally, let $P(X)$ and $R(X)$ be the sets of unrooted and rooted, respectively, phylogenetic trees on X. Choose an element $x \in X$, and define $\psi_x\colon P(X) \to R(X \setminus \{x\})$ be a function that maps an unrooted tree $T \in P(X)$ to a rooted tree in $R(X \setminus \{x\})$ as described above. It can be verified that ψ_x is bijective.

Theorem 4. *The number of unrooted binary phylogenetic trees on X is* $(2|X| - 5)!!$.

Corollary 5. *The number of rooted binary phylogenetic trees on X is* $(2|X| - 3)!!$.

We introduce some additional notation. A cherry is a pair of leaves that have a common neighbor. A caterpillar (rooted or unrooted) is a binary phylogenetc tree for which the subgraph induced on the set of interior vertices is a path. Thus a rooted caterpillar has only one cherry, and an unrooted caterpillar has exactly two cherries (for the case the tree has more than three leaves). A phylogenetic tree with no internal edge is called a star tree.

Let X be a nonempty taxon set. A split of X (or an X-split) is a partition of X into two nonempty subsets, which we call the components of the split. A split with components A and B is denoted as $A|B$. If either $|A| = 1$ or $|B| = 1$, we say that the split $A|B$ is trivial. Let T be an unrooted X-tree. The removal of an edge of T results in two (unrooted) subtrees. Clearly the leaf-label sets of the two subtrees form an X-split, and we call it an induced X-split.

Let T be a rooted X-tree, and consider the edges of T as being directed away from the root of T. For a vertex v of T, denote by $T(v)$ the induced subgraph of T on the set of descendant vertices of v (including v itself). The leaf-label set of $T(v)$ is called a cluster induced by v, and is denoted by $C_T(v)$. Let A be a subset of X. The least common ancestor of A in T, denoted by $\mathrm{lca}_T(A)$, is the vertex v with the property that: (1) $A \subset C_T(v)$; (2) for any vertex w that is a proper descendant of v, $A \not\subset C_T(w)$.

Finally, for specifying a rooted phylogenetic tree in the text format, we can use the Newick format. In this format, parentheses are used to group siblings, which are separated by commas. For example, (a, b, c) represents a rooted tree with three leaves a, b, and c attached directly to the root; $((a, b), c), d)$ represents a rooted caterpillar tree in which a and b form a cherry (in fact, it is the only cherry in the tree). More details about the Newick format can be found at: `http://evolution.genetics.washington.edu/phylip/newicktree.html`.

Split compatibility. We say that two X-splits $A|B$ and $C|D$ are compatible if at least one of the four intersections $A \cap C$, $A \cap D$, $B \cap C$, and $B \cap D$ is empty. It can be verified that if $A|B$ and $C|D$ are distinct, then they are compatible if and only if exactly one of the four intersections above is empty.

For an X-tree T, let $\Sigma(T)$ be the set of X-splits induced by the edges of T.

Theorem 6. *Let Σ be a collection of X-splits. Then there exists an X-tree T such that $\Sigma = \Sigma(T)$ if and only if the splits in Σ are pairwise compatible.*