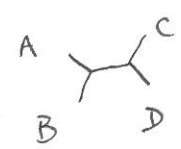


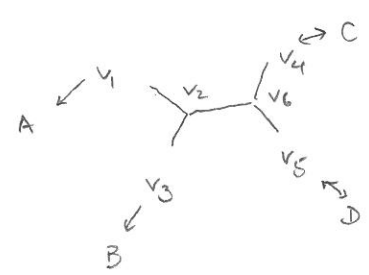
PHYLOGENETIC TREES

(aka add the missing ingredient of leaf-labelled tree)



Let X denote a finite set of labels or taxa names.

Then a PHYLOGENETIC TREE or formally PHYLOGENETIC X -tree is a tree $T = (V, E)$ together with a ^{bijection} labelling MAP $\varphi: \mathcal{L} \rightarrow X$ from the leaves to X .



$$\mathcal{L} = \{v_1, v_3, v_4, v_5\}$$

Since one goal of phylogenetic inference is to determine the 'best' tree relating sequence data, we next need to count how many trees there are.

Notation: Fix $n = \text{number of taxa}$ $|X| = n$

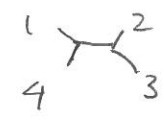
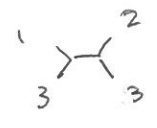
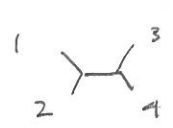
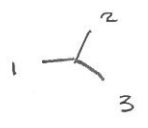
Define $b(n) = \text{number of distinct unrooted binary phylogenetic } X\text{-trees}$
see Figure.

Question: What is $b(n)$?

$n=2$

$n=3$

$n=4$



$b(2) = 1$

$b(3) = 1$

$b(4) = 3$

Isomorphic versions.

n=5: Think proof ...

$$(\# \text{ of } 3\text{-taxon trees}) (\text{number of edges to graft new edge onto}) = 15$$

$$(3) (5)$$

$b(5) = 15$

n=6: $b(5)$ (number of edges in 5-taxon unrooted tree)

Comments: The formula for $b(n)$ is recursive and we need to know the number of edges in any $(n-1)$ -taxon tree.

Theorem: If T_n is an unrooted, n -taxon, binary tree, then

T_n has _____ vertices and _____ edges

n	vertices	edges
2	2	1
3	4	3
4	6	5
5	8	7
6	10	9
n	$2n-2$	$2n-3$

Calc II

$2n$

Formal proof by induction.

Informal proof

Given an $(n-1)$ -taxon tree, adding an edge adds 2 vertices and 2 edges

Theorem 2: If $T=(V,E)$ is an unrooted phylogenetic X-tree with $|X|=n \geq 2$,

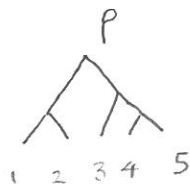
then $b(n) = (2n-5)!! = 1 \cdot 3 \cdot 5 \cdot \dots \cdot (2n-5)$

n	$b(n) = 1 \cdot 3 \cdot 5 \cdot \dots \cdot (2n-5)$
3	1
4	$1 \cdot 3 = 3$
5	$1 \cdot 3 \cdot 5 = 15$

6	$1 \cdot 3 \cdot 5 \cdot 7 = 105$	etc!
7	$105 \cdot 9 = 945$	

$b(n)$ grows super exponentially so for n -taxon datasets, there are many, many tree shapes to test for goodness of fit. Searching tree-space is hard!

To count the number of rooted phylogenetic trees, observe that each rooted tree T^p on $n = |X|$ taxa corresponds to



an unrooted tree on $n+1$ taxa.

Add a branch from p and view the tree as unrooted.

Theorem: The number of rooted phylogenetic X-trees on n taxa is

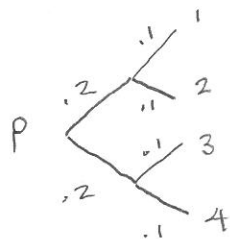
$$b(n+1) = (2n-3)!! = 1 \cdot 2 \cdot 3 \cdot \dots \cdot (2n-3)$$

To quantify the amount of evolutionary change relating taxa, we use

METRIC TREES: A metric tree (T, w) is a rooted or unrooted

tree together with a function $w: E \rightarrow \mathbb{R}^{\geq 0}$ which assigns nonnegative weights or lengths to the edges in T .

Graph:

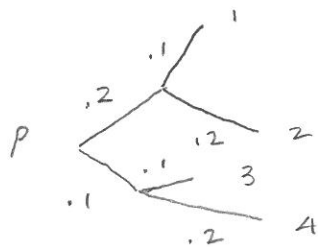


Given a rooted phylogenetic X-tree with weights or edge lengths, we can compute pairwise distances

 T_1

	1	2	3	4
1		.2	.6	.6
2			.6	.6
3				.2

Now consider the tree T_2 in comparison to T_1



T_2 suggests that "more" mutation occurred in taxon 2's descent from a common ancestor than taxon 1's.

≡ Differing amounts of mutation

This might be caused by a change (speeding up or slowing down) of the mutation rate for instance.

Defn: A rooted metric tree is called ULTRAMETRIC or we have a MOLECULAR CLOCK if all tips/leaves are equidistant from the root.

i.e. for all $l_1, l_2 \in \mathcal{L}$ leaves, $d(p, l_1) = d(p, l_2)$

One benefit of a molecular clock at work is that one might assume that the mutation rate is constant so that edge lengths are proportional to time.