

BMI/CS 576 – Day 14

- Today
 - Simulating evolution along a tree
 - Generating distance data
- Next week
 - Neighbor-joining and parsimony methods
 - Midterm (Wednesday)

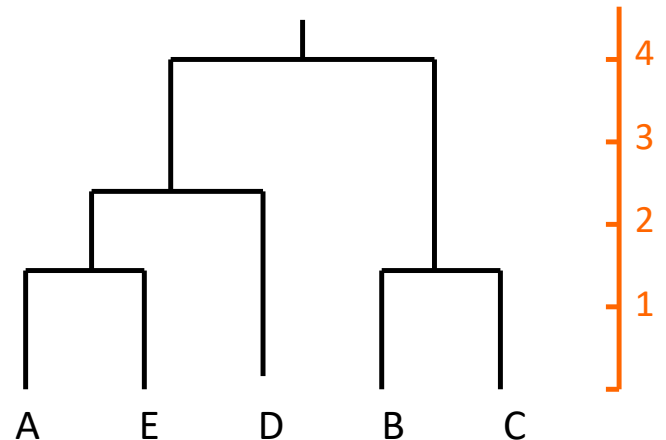
The molecular clock assumption & ultrametric data

- Ultrametric data: for all triplets of taxa, i, j, k , the pairwise distances between them are either all equal, or two are equal and the remaining one is smaller

alternative equivalent definition:

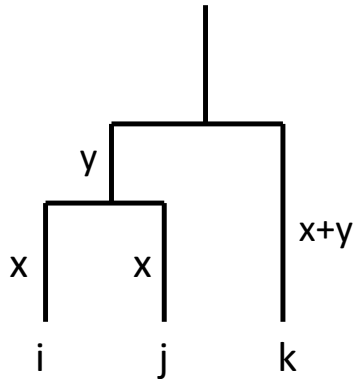
$$d_{ij} \leq \max(d_{ik}, d_{jk}) \quad \forall i, j, k$$

	A	B	C	D	E
A	0	8	8	5	3
B		0	3	8	8
C			0	8	8
D				0	5
E					0



Molecular clock => ultrametric

If molecular clock, then any tree relating three taxa looks like this:



The pairwise distances are thus:

$$d_{ij} = 2x$$

$$d_{ik} = 2(x + y)$$

$$d_{jk} = 2(x + y)$$

The ultrametric condition holds: $d_{ij} \leq d_{ik} = d_{jk}$

Questions

- When will the actual data be ultrametric?
 - Distance estimates from actual data are very unlikely to perfectly meet the ultrametric criteria
 - Even if the true, unobserved, rates of molecular evolution are ultrametric, randomness in mutation events often result in *estimated* rates not being ultrametric
 - Nevertheless, methods that assume ultrametric data can still be used on data that are not ultrametric
 - If the data are “close enough” to an ultrametric, these methods can succeed.

Questions

- Another example of UPGMA
 - We'll do this on Tuesday
 - Tuesday's notebook: UPGMA and NJ by hand