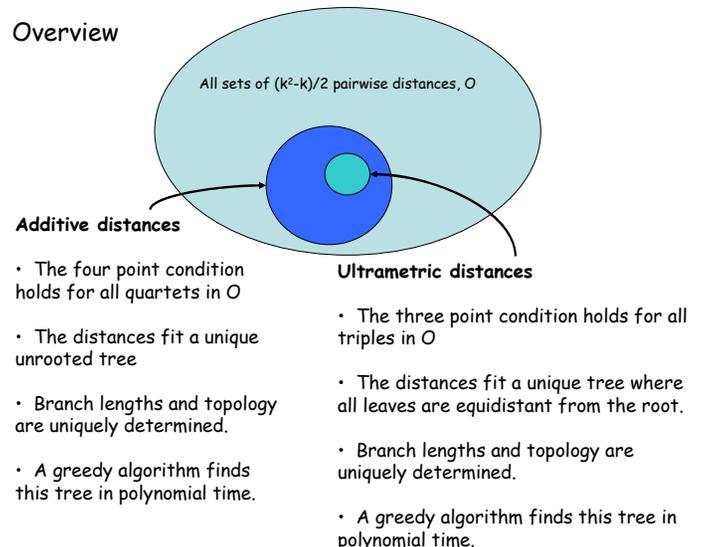


Overview



Summary

- A matrix is *additive* if it satisfies the four point condition.
- A tree defines a *tree metric*, $T[i,j]$; i.e., the pairwise distances between all pairs of leaves.
- All tree metrics are additive.
- If a matrix, $O[i,j]$, is additive
 - there exists a unique tree topology with branch lengths such that $T[i,j] = O[i,j]$.
 - This tree can be obtained in polynomial time.
- In real life, observed distance matrix, $O[i,j]$ is never additive.

Summary, cont'd

- A matrix is *ultrametric* if it satisfies the three point condition.
- All ultrametric matrices fit rooted trees.
- Not all rooted tree metrics are ultrametric.
- An ultrametric tree
 - satisfies the molecular clock hypothesis.
 - All distances from the root to a leaf are the same.
 - Its branch lengths are proportional to time.
- For $k > 3$,
 - All ultrametric matrices are additive
 - But, an additive matrix is *not necessarily* ultrametric.

Greedy distance-based methods

Given k taxa with pairwise distances O .

- Initialization: F = forest of k trees of size 1. $D=O$
- Repeat
 - Find subtrees u and v in F such that u and v are neighbors in tree.
 - Merge u and v to obtain a new subtree w .
 - Compute branch lengths in w
 - Remove u and v from F . Insert w into F .
 - Recompute D , replacing u and v with w until $(|F|=1)$

The algorithms differ in

- How neighbors to be merged are identified.
- How the branch lengths are computed.
- How the distance matrix is updated.

Unweighted Paired Group Method with Average Means (UPGMA)

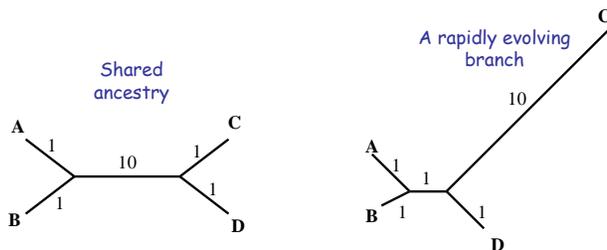
- The [UPGMA algorithm](#) assumes that the rate of change is the same in all lineages. The consequences of that assumption are:
 - At each step, the two closest taxa are selected as neighbors.
 - $\text{height}(\text{parent}(i,j)) = D[i,j]/2$
 - Root is midpoint between two most distant taxa.
- This is true for ultrametric matrices, but not generally true, even for additive matrices.
- If UPGMA is applied to a matrix that deviates greatly from the assumptions of ultrametricity, a tree with the wrong topology and/or branch lengths may result.

Neighbor Joining

The [NJ algorithm](#) adjusts the distance matrix for variations in the rate of change. The “adjusted” distance between a pair of nodes is calculated by subtracting the average of the distances to all other leaves.

- **Thm:**
 - If D is additive, the pair of taxa that minimize this “corrected” distance matrix are neighbors in the true tree.
- **Proof:**
 - Durbin *et al.*, 7.8
- If D is additive, then NJ will reconstruct the correct *unrooted* tree in quadratic time.

NJ intuition



Does a long branch indicate shared ancestry or a change in the substitution rate?