

CMSC423: Bioinformatic Algorithms, Databases and Tools

Phylogenetic inference

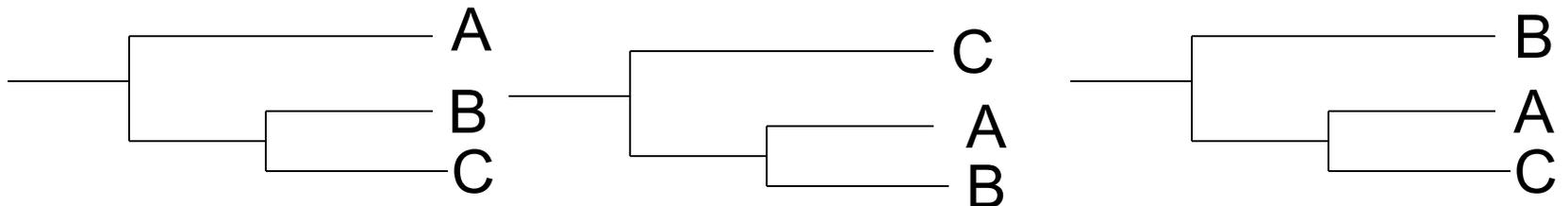
Phylogeny questions

Given several organisms & a set of features (usually sequence, but also morphological: wing shape/color...)

- A. Given a phylogenetic tree – figure out what the ancestors looked like (what are the features of internal nodes)



- B. Find the phylogenetic tree that best describes the common evolutionary heritage of the organisms



Phylogeny questions

- A. Easy-ish – can be done with dynamic programming
- B. Hard – Many possible trees

$$\frac{(2n-3)!}{2^{n-2} (n-2)!}$$

rooted trees with n leaves

Phylogeny estimation methods

- Distance-based methods
- Maximum parsimony
- Maximum Likelihood

Distance-Based Phylogeny Problem

Reconstruct an evolutionary tree fitting a distance matrix

Input: A distance matrix

Output: A tree fitting this distance matrix

Note: A weighted unrooted tree T fits a distance matrix D if $d_{i,j}(T) = D_{i,j}$ for every pair of leaves i and j

Distance matrix

Distance matrix

- Symmetric (for all i, j $D_{i,j} = D_{j,i}$)
- Non-negative
- satisfy triangle's inequality (for all i, j , and k , $D_{i,j} + D_{j,k} \geq D_{i,k}$)

	Carp	Zebrafish	Salmon	Trout
Carp	0	3	7	9
Zebrafish		0	6	8
Salmon			0	6
Trout				0

Construct a phylogenetic tree

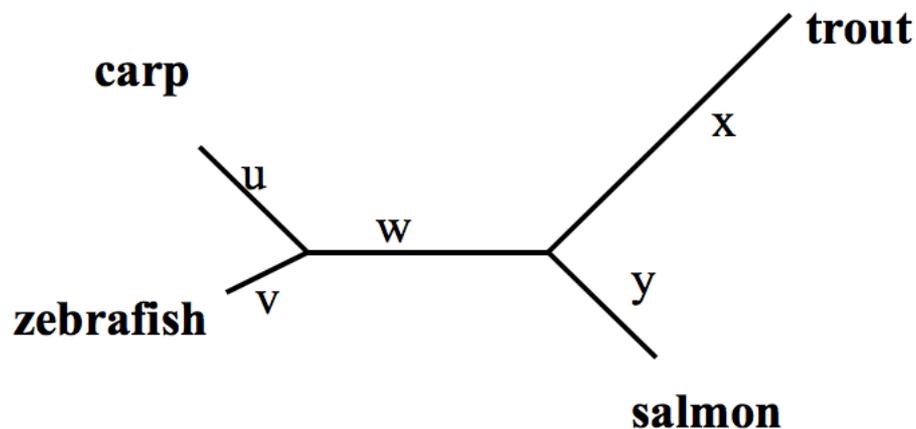
	Carp	Zebrafish	Salmon	Trout	
Carp	0	3	7	9	
Zebrafish		0	6	8	Observed distances
Salmon			0	6	
Trout				0	

- Tree topology? Think hierarchical clustering
- Branch lengths?

Given a tree topology, estimating branch lengths is easy...

	Carp	Zebrafish	Salmon	Trout
Carp	0	3	7	9
Zebrafish		0	6	8
Salmon			0	6
Trout				0

Observed
distances



Match distance matrix to branch lengths

Match distance matrix to branch lengths

	Carp	Zebrafish	Salmon	Trout
Carp	0	3	7	9
Zebrafish		0	6	8
Salmon			0	6
Trout				0

Observed
distances

$$u + v = 3$$

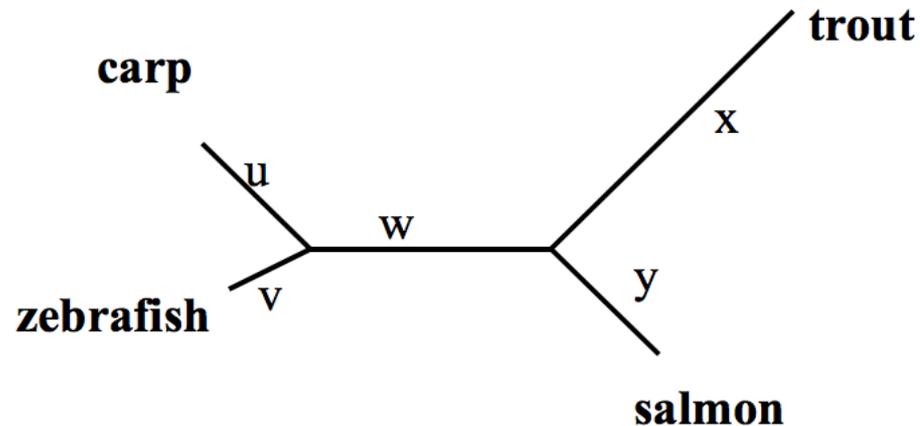
$$u + w + y = 7$$

$$u + w + x = 9$$

$$v + w + y = 6$$

$$v + w + x = 8$$

$$x + y = 6$$



Match distance matrix to branch lengths

	Carp	Zebrafish	Salmon	Trout
Carp	0	3	7	9
Zebrafish		0	6	8
Salmon			0	6
Trout				0

Observed
distances

$$u + v = 3$$

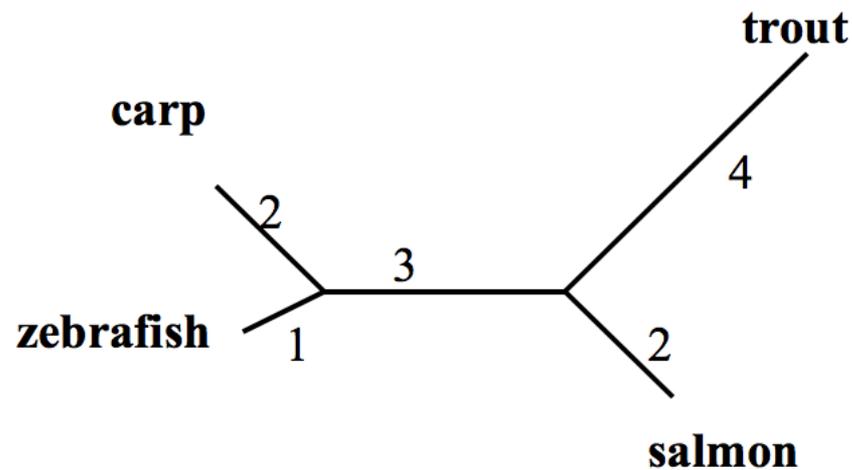
$$u + w + y = 7$$

$$u + w + x = 9$$

$$v + w + y = 6$$

$$v + w + x = 8$$

$$x + y = 6$$



Next: maximum parsimony

Maximum likelihood is a variant of maximum parsimony (at least algorithmically)