CS4054
Bioinformatics

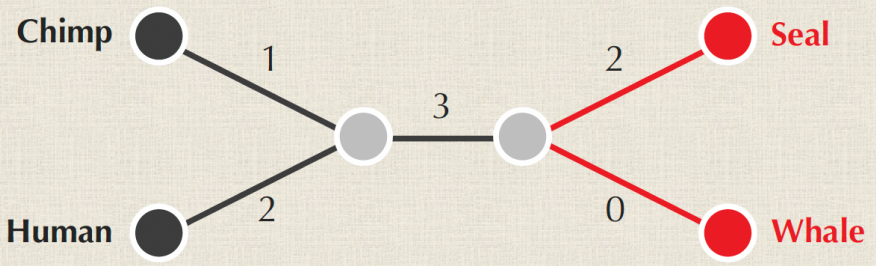
Spring 2025

Rushda Muneer

An Idea for Distance-Based Phylogeny

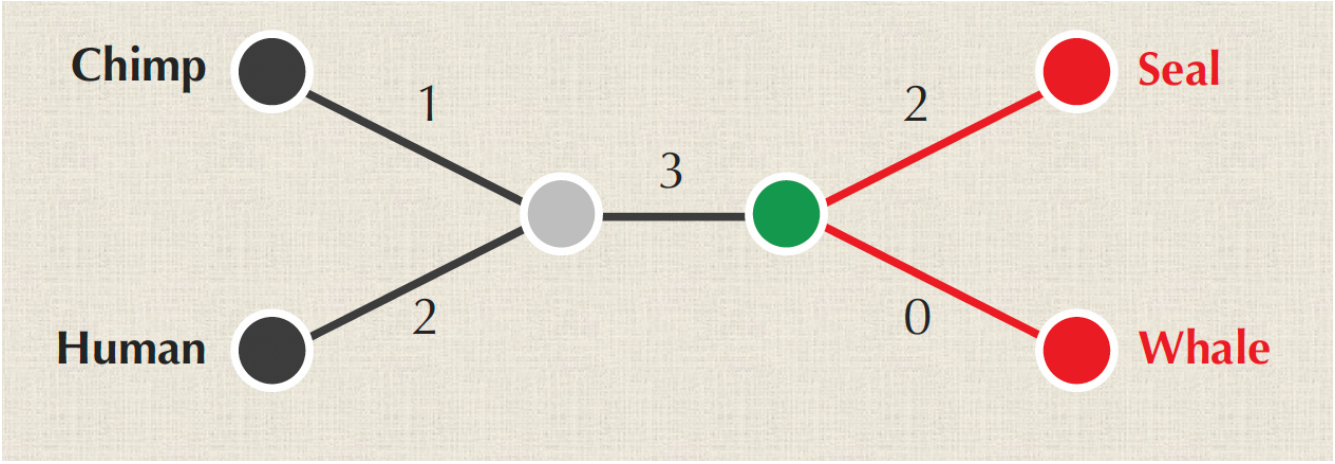
- Using an existing solved tree we can devise and algorithm for constructing a tree from the distance matrix
- Visually, the minimum distance between two species in the tree corresponds to selecting the leaves that are next to each other

	Chimp	Human	Seal	Whale
Chimp	0	3	6	4
Human	3	0	7	5
Seal	6	7	0	2
Whale	4	5	2	0



An Idea for Distance-Based Phylogeny

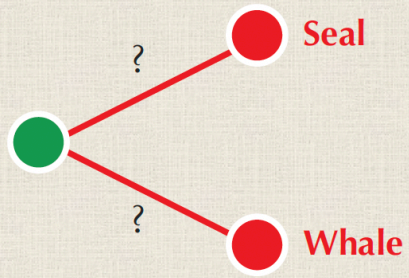
- **Seal and whale are neighbors (meaning they are leaves with the same parent).**



An Idea for Distance-Based Phylogeny

We don't know the distance between the leaves and their parent and can only use the fact the two leaves are neighbors to one another

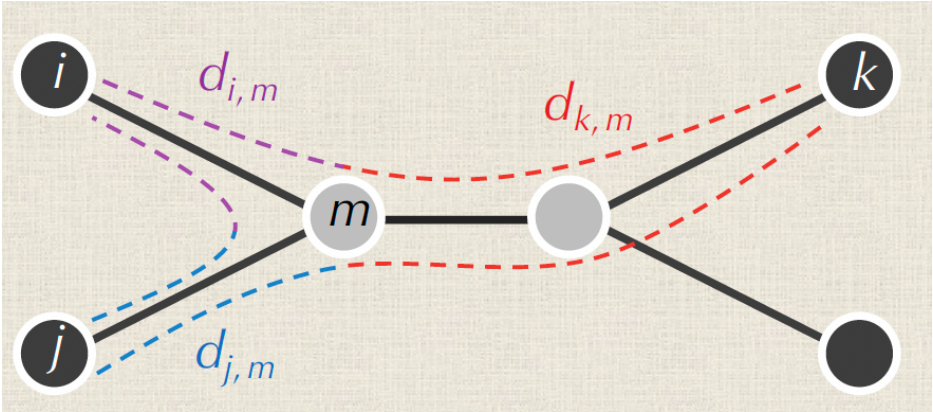
	Chimp	Human	Seal	Whale
Chimp	0	3	6	4
Human	3	0	7	5
Seal	6	7	0	2
Whale	4	5	2	0



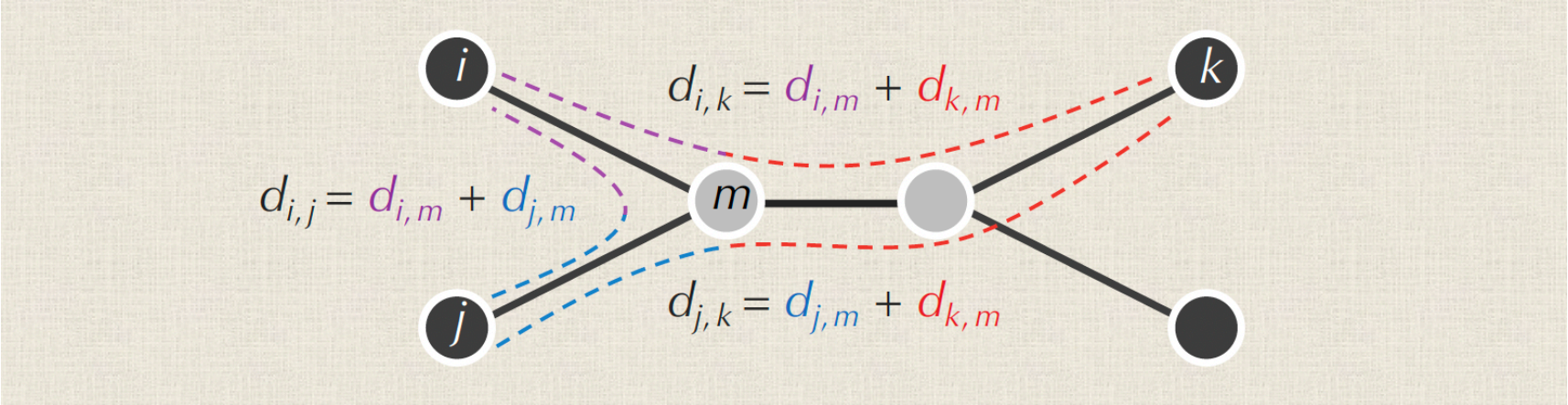
Key Point: How do we compute the unknown distances?

Towards a Recursive Algorithm

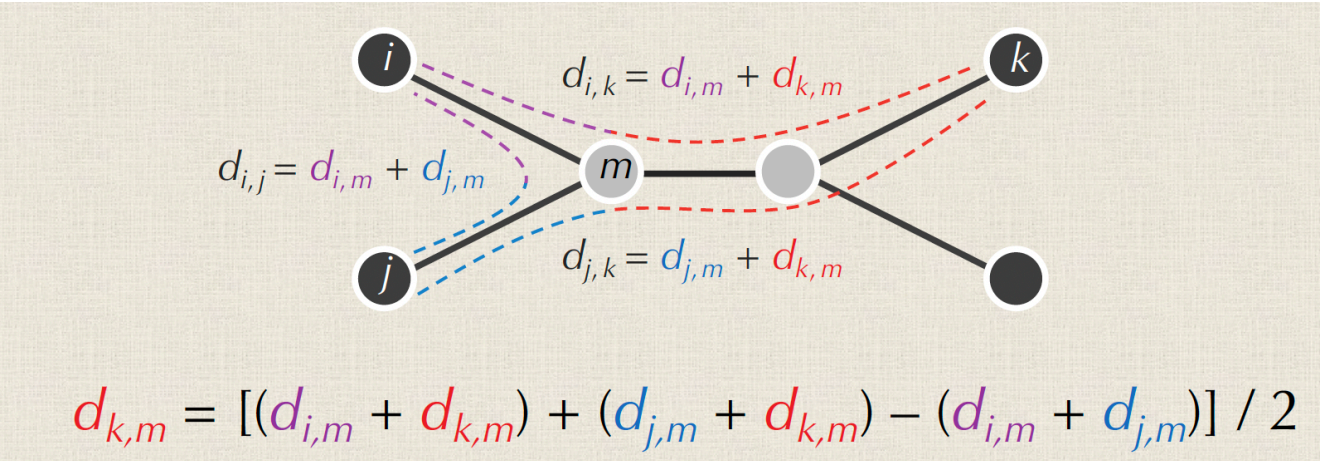
- Suppose we have two leaves i and j with a parent m
- We want to find out $d_{i,m}$ and $d_{j,m}$
- Using a third leaf k in the tree we want to use the distance $d_{k,m}$ to help us reconstruct the unknown values



Towards a Recursive Algorithm

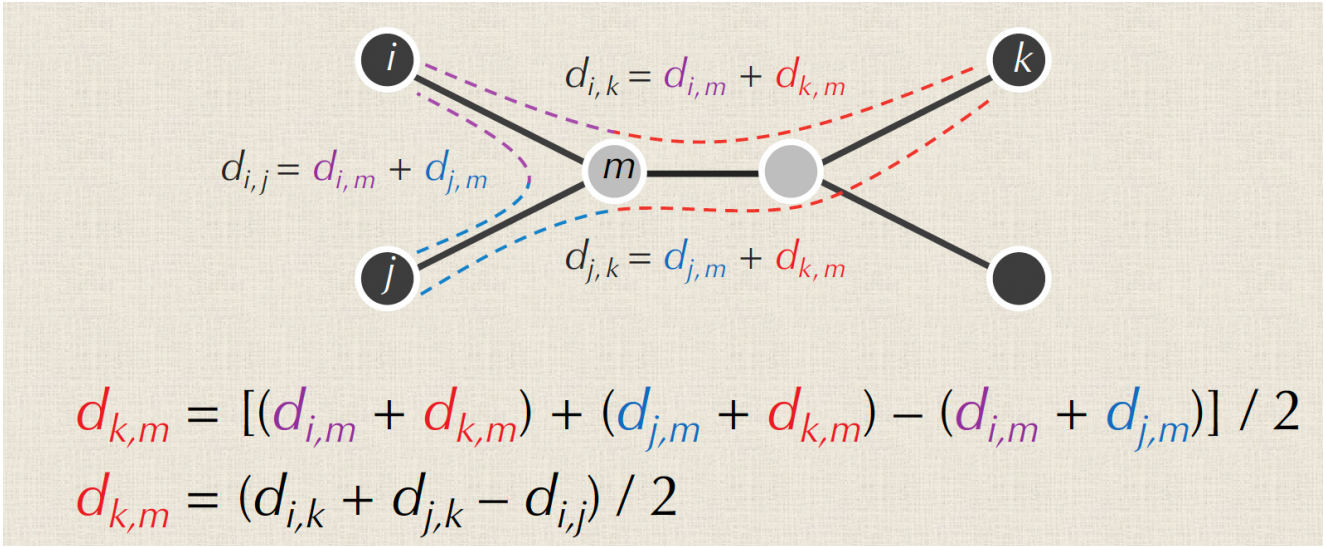


Towards a Recursive Algorithm



We don't know the distances between the internal nodes from the distance matrix
Hence we want to find a way to compute them via a priori information

Towards a Recursive Algorithm

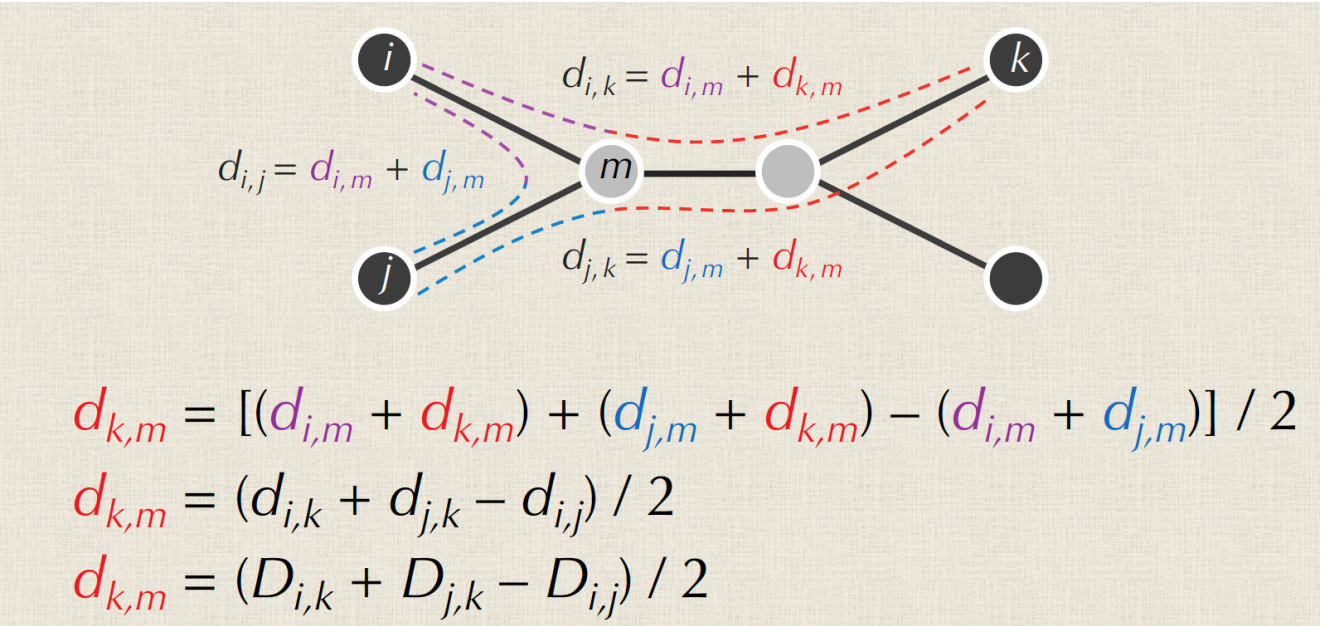


$$d_{k,m} = [(d_{i,m} + d_{k,m}) + (d_{j,m} + d_{k,m}) - (d_{i,m} + d_{j,m})] / 2$$

$$d_{k,m} = (d_{i,k} + d_{j,k} - d_{i,j}) / 2$$

The lower case d represents the intermediate distance between the internal nodes

Towards a Recursive Algorithm



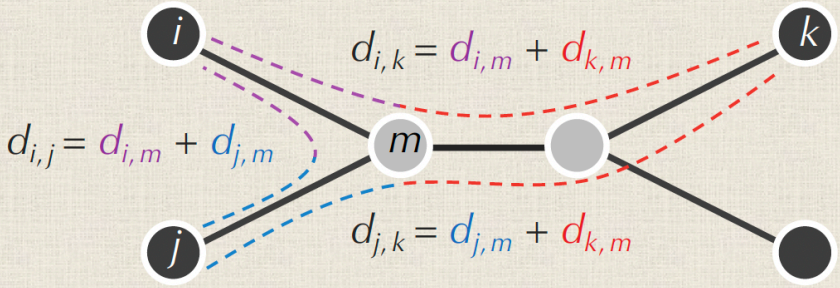
$$d_{k,m} = [(d_{i,m} + d_{k,m}) + (d_{j,m} + d_{k,m}) - (d_{i,m} + d_{j,m})] / 2$$

$$d_{k,m} = (d_{i,k} + d_{j,k} - d_{i,j}) / 2$$

$$d_{k,m} = (D_{i,k} + D_{j,k} - D_{i,j}) / 2$$

We can replace this distance with upper case D, which refers to the a priori distance we have available for the leaves in the distance matrix

Towards a Recursive Algorithm



$$d_{k,m} = [(d_{i,m} + d_{k,m}) + (d_{j,m} + d_{k,m}) - (d_{i,m} + d_{j,m})] / 2$$

$$d_{k,m} = (d_{i,k} + d_{j,k} - d_{i,j}) / 2$$

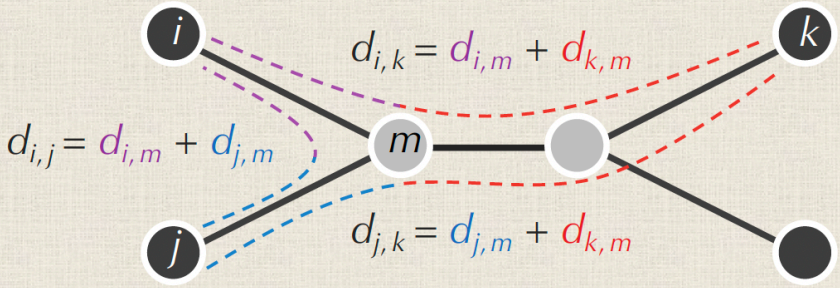
$$d_{k,m} = (D_{i,k} + D_{j,k} - D_{i,j}) / 2$$

$$\therefore \quad d_{i,m} = D_{i,k} - (D_{i,k} + D_{j,k} - D_{i,j}) / 2$$

Towards a Recursive Algorithm

An analogous formula is available for j to m

Note that these values depend upon our choice of k



$$d_{k,m} = [(d_{i,m} + d_{k,m}) + (d_{j,m} + d_{k,m}) - (d_{i,m} + d_{j,m})] / 2$$

$$d_{k,m} = (d_{i,k} + d_{j,k} - d_{i,j}) / 2$$

$$d_{k,m} = (D_{i,k} + D_{j,k} - D_{i,j}) / 2$$

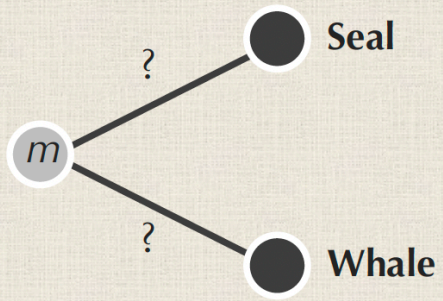
$$\therefore \; d_{i,m} = D_{i,k} - (D_{i,k} + D_{j,k} - D_{i,j}) / 2$$

$$d_{i,m} = (D_{i,k} + D_{i,j} - D_{j,k}) / 2$$

Now we can compute the distance from i to m

An Idea for Distance-Based Phylogeny

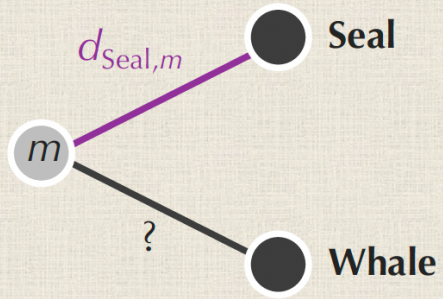
	Chimp	Human	Seal	Whale
Chimp	0	3	6	4
Human	3	0	7	5
Seal	6	7	0	2
Whale	4	5	2	0



$$d_{i,m} = (D_{i,k} + D_{i,j} - D_{j,k}) / 2$$

An Idea for Distance-Based Phylogeny

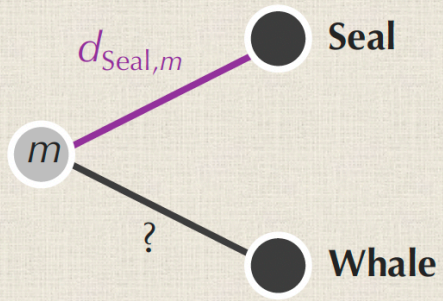
	Chimp	Human	Seal	Whale
Chimp	0	3	6	4
Human	3	0	7	5
Seal	6	7	0	2
Whale	4	5	2	0



$$d_{Seal,m} = (D_{Seal,k} + D_{Seal,j} - D_{j,k}) / 2$$

An Idea for Distance-Based Phylogeny

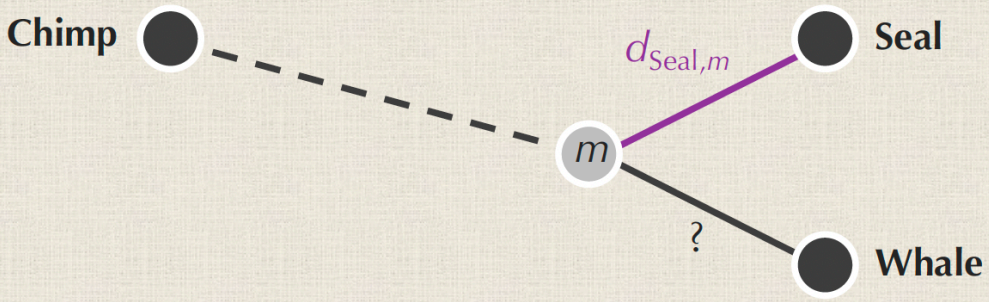
	Chimp	Human	Seal	Whale
Chimp	0	3	6	4
Human	3	0	7	5
Seal	6	7	0	2
Whale	4	5	2	0



$$d_{Seal,m} = (D_{Seal,k} + D_{Seal,Whale} - D_{Whale,k}) / 2$$

An Idea for Distance-Based Phylogeny

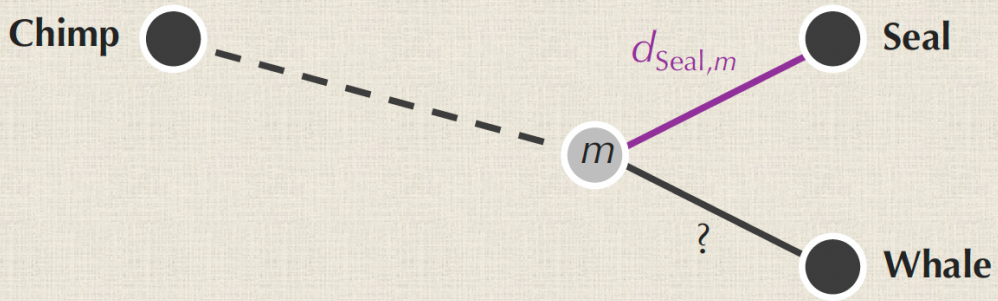
	Chimp	Human	Seal	Whale
Chimp	0	3	6	4
Human	3	0	7	5
Seal	6	7	0	2
Whale	4	5	2	0



$$d_{Seal,m} = (D_{Seal,Chimp} + D_{Seal,Whale} - D_{Whale,Chimp}) / 2$$

An Idea for Distance-Based Phylogeny

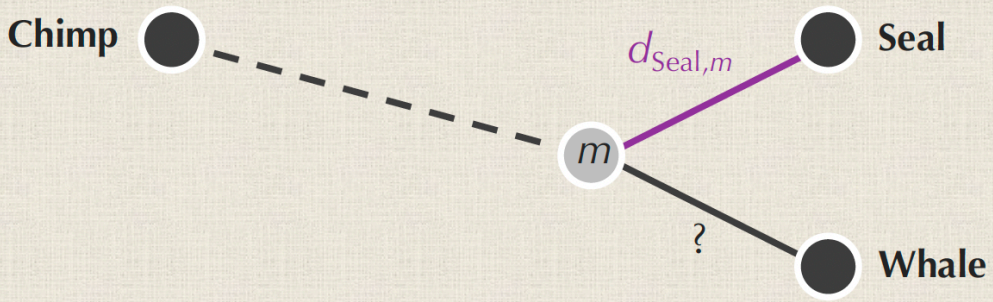
	Chimp	Human	Seal	Whale
Chimp	0	3	6	4
Human	3	0	7	5
Seal	6	7	0	2
Whale	4	5	2	0



$$d_{Seal,m} = (\quad 6 \quad + D_{Seal,Whale} - D_{Whale,Chimp}) / 2$$

An Idea for Distance-Based Phylogeny

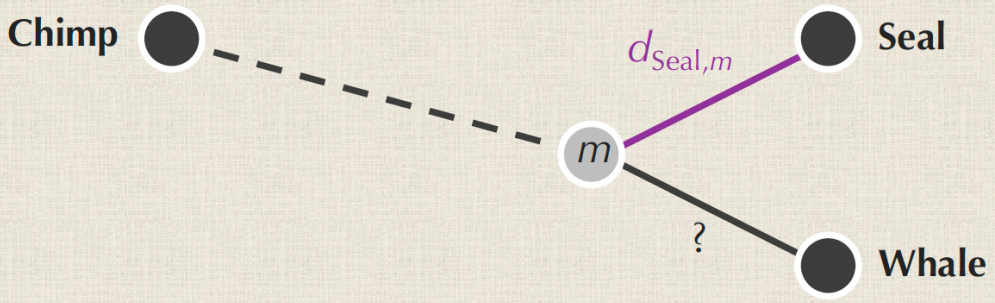
	Chimp	Human	Seal	Whale
Chimp	0	3	6	4
Human	3	0	7	5
Seal	6	7	0	2
Whale	4	5	2	0



$$d_{Seal,m} = (\ 6\ +\ 2\ - D_{Whale,Chimp}) / 2$$

An Idea for Distance-Based Phylogeny

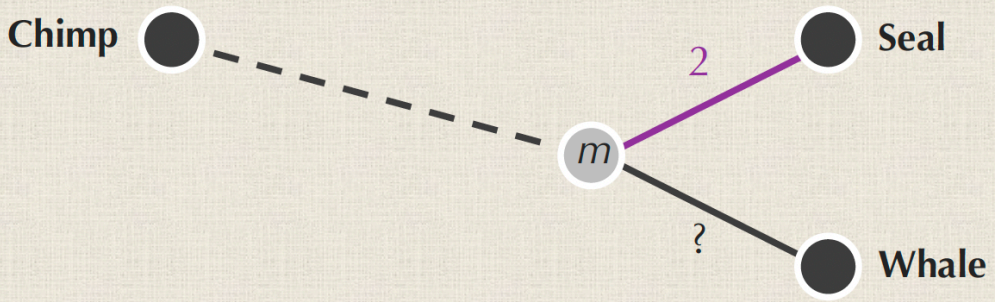
	Chimp	Human	Seal	Whale
Chimp	0	3	6	4
Human	3	0	7	5
Seal	6	7	0	2
Whale	4	5	2	0



$$d_{Seal,m} = (\quad 6 \quad + \quad 2 \quad - \quad 4 \quad) / 2$$

An Idea for Distance-Based Phylogeny

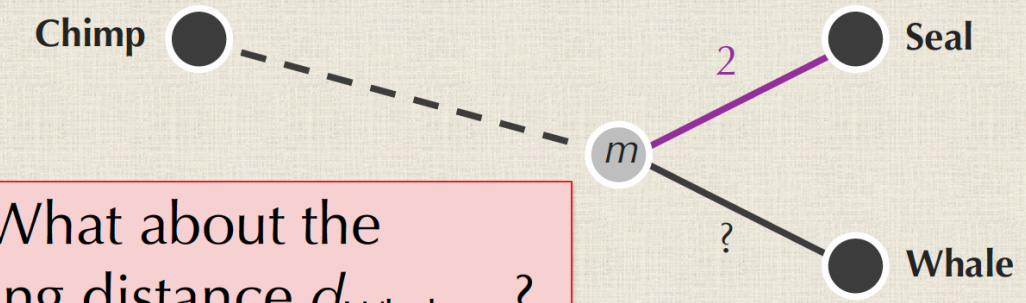
	Chimp	Human	Seal	Whale
Chimp	0	3	6	4
Human	3	0	7	5
Seal	6	7	0	2
Whale	4	5	2	0



$d_{Seal,m} = 2$

An Idea for Distance-Based Phylogeny

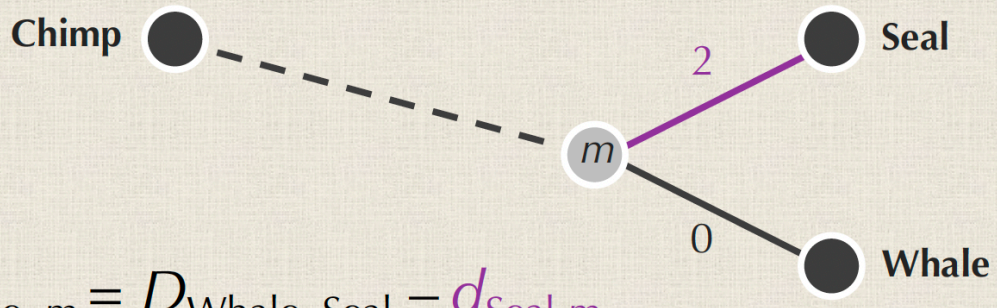
	Chimp	Human	Seal	Whale
Chimp	0	3	6	4
Human	3	0	7	5
Seal	6	7	0	2
Whale	4	5	2	0



STOP: What about the remaining distance $d_{\text{Whale}, m}$?

An Idea for Distance-Based Phylogeny

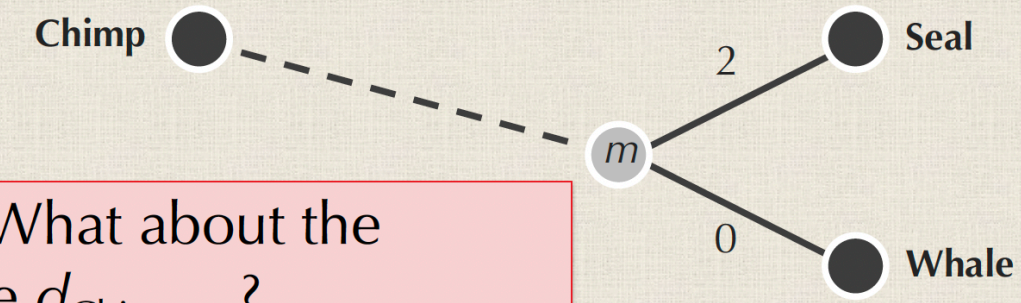
	Chimp	Human	Seal	Whale
Chimp	0	3	6	4
Human	3	0	7	5
Seal	6	7	0	2
Whale	4	5	2	0



$$d_{\text{Whale, } m} = D_{\text{Whale, Seal}} - d_{\text{Seal},m}$$
$$= 2 - 2 = 0$$

An Idea for Distance-Based Phylogeny

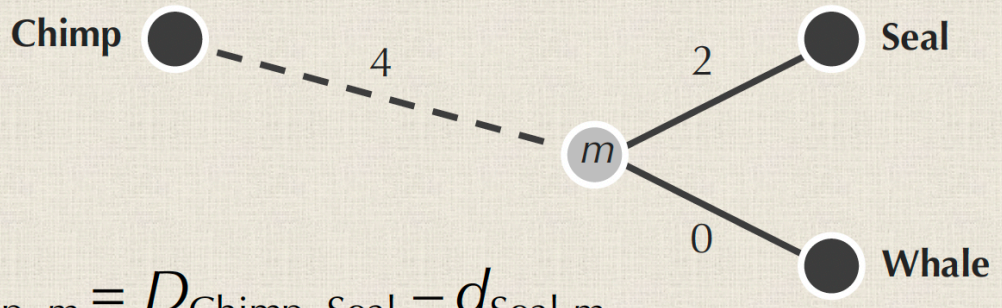
	Chimp	Human	Seal	Whale
Chimp	0	3	6	4
Human	3	0	7	5
Seal	6	7	0	2
Whale	4	5	2	0



STOP: What about the distance $d_{Chimp,\ m}$?

An Idea for Distance-Based Phylogeny

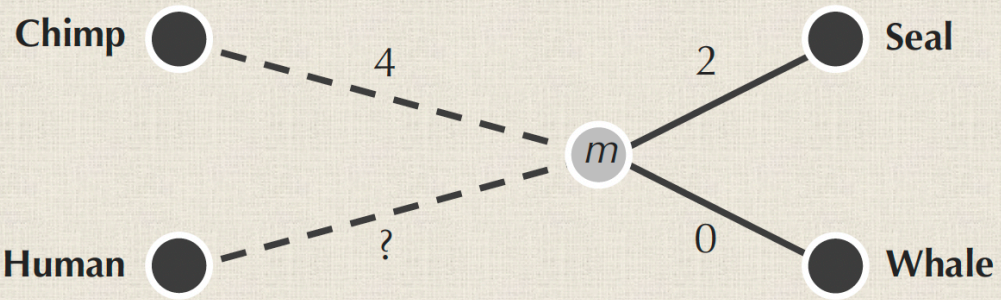
	Chimp	Human	Seal	Whale
Chimp	0	3	6	4
Human	3	0	7	5
Seal	6	7	0	2
Whale	4	5	2	0



$$d_{\text{Chimp, } m} = D_{\text{Chimp, Seal}} - d_{\text{Seal},m}$$
$$= 6 - 2 = 4$$

An Idea for Distance-Based Phylogeny

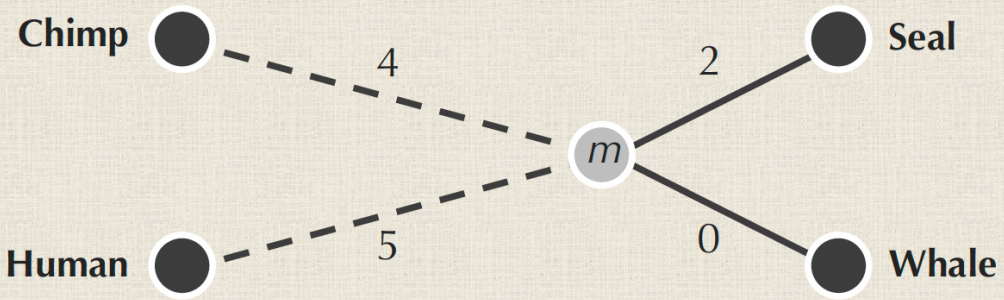
	Chimp	Human	Seal	Whale
Chimp	0	3	6	4
Human	3	0	7	5
Seal	6	7	0	2
Whale	4	5	2	0



STOP: And $d_{\text{Human, } m}$?

An Idea for Distance-Based Phylogeny

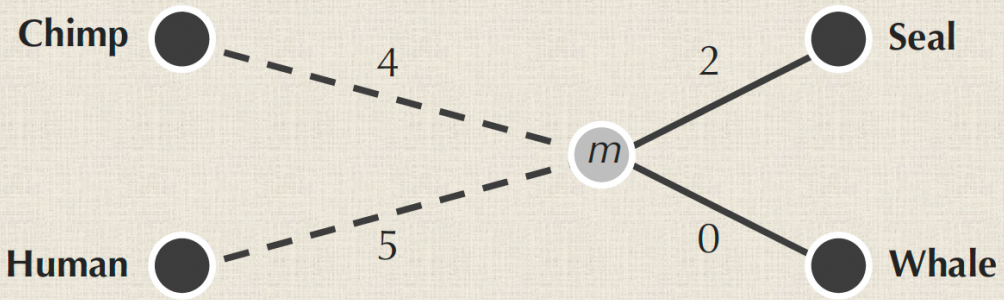
	Chimp	Human	Seal	Whale
Chimp	0	3	6	4
Human	3	0	7	5
Seal	6	7	0	2
Whale	4	5	2	0



An Idea for Distance-Based Phylogeny

Adding a new row, column for m in the distance matrix

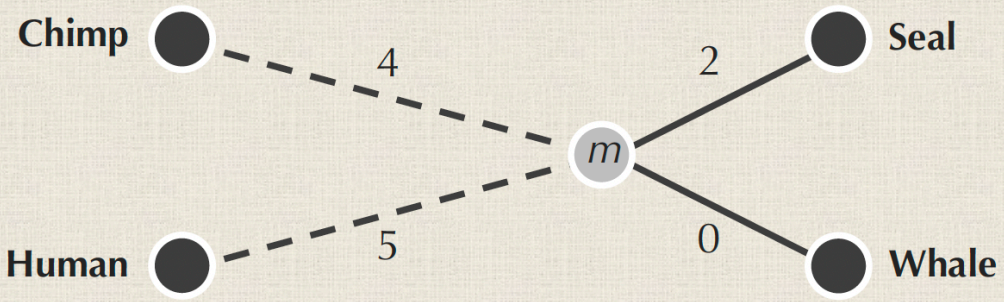
	Chimp	Human	Seal	Whale	*m*
Chimp	0	3	6	4	4
Human	3	0	7	5	5
Seal	6	7	0	2	2
Whale	4	5	2	0	0
m	4	5	2	0	0



An Idea for Distance-Based Phylogeny

Recurse to a 3x3 matrix

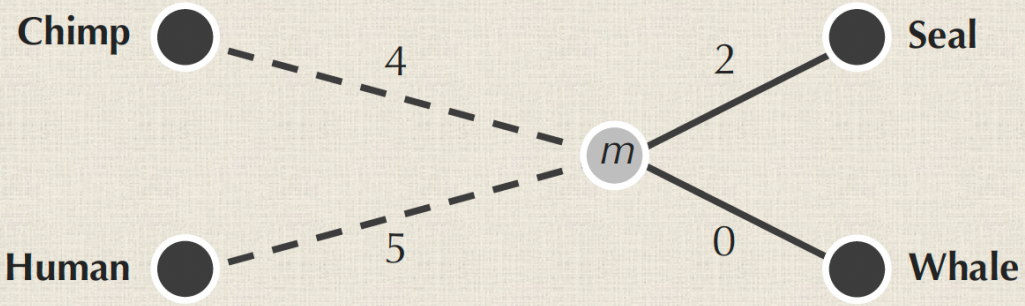
	Chimp	Human	Seal	Whale	*m*
Chimp	0	3	6	4	4
Human	3	0	7	5	5
Seal	6	7	0	2	2
Whale	4	5	2	0	0
m	4	5	2	0	0



An Idea for Distance-Based Phylogeny

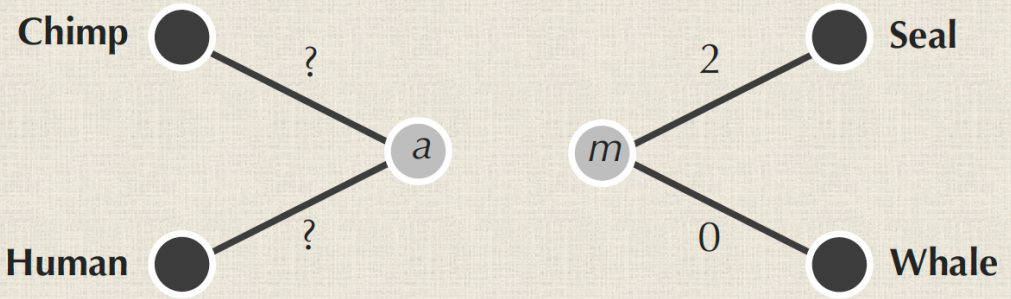
Recurse to a 3x3 matrix

	Chimp	Human	*m*
Chimp	0	3	4
Human	3	0	5
m	4	5	0



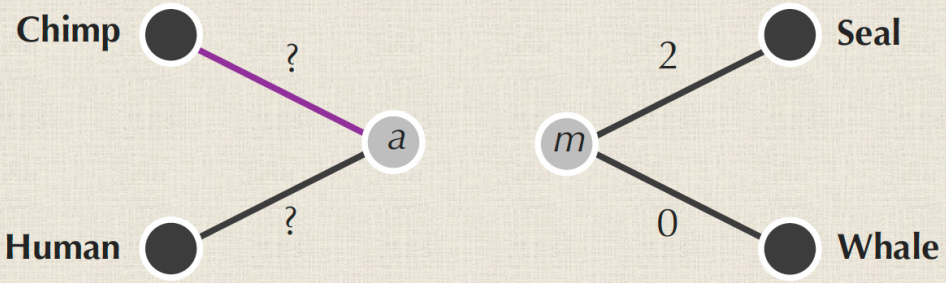
An Idea for Distance-Based Phylogeny

	Chimp	Human	*m*
Chimp	0	**3**	4
Human	**3**	0	5
m	4	5	0



An Idea for Distance-Based Phylogeny

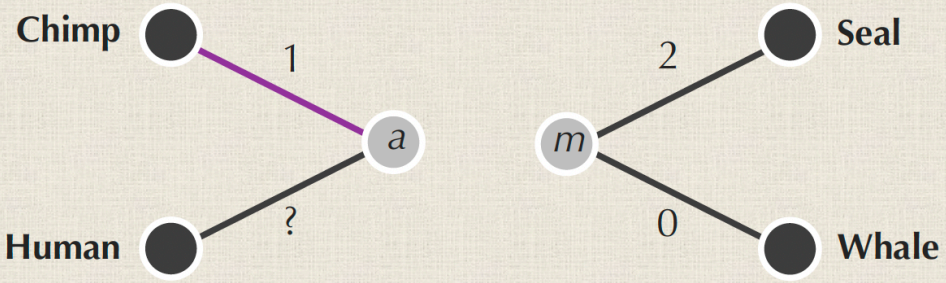
	Chimp	Human	*m*
Chimp	0	3	4
Human	3	0	5
m	4	5	0



$d_{Chimp,a} = (D_{Chimp,m} + D_{Chimp,Human} - D_{Human,m}) / 2$

An Idea for Distance-Based Phylogeny

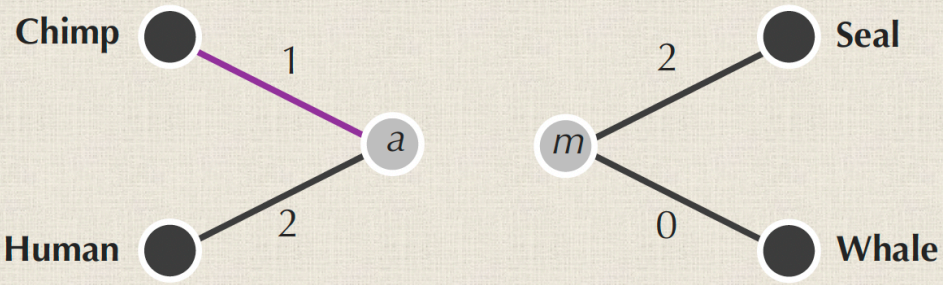
	Chimp	Human	*m*
Chimp	0	3	4
Human	3	0	5
m	4	5	0



$d_{Chimp,a} = 1$

An Idea for Distance-Based Phylogeny

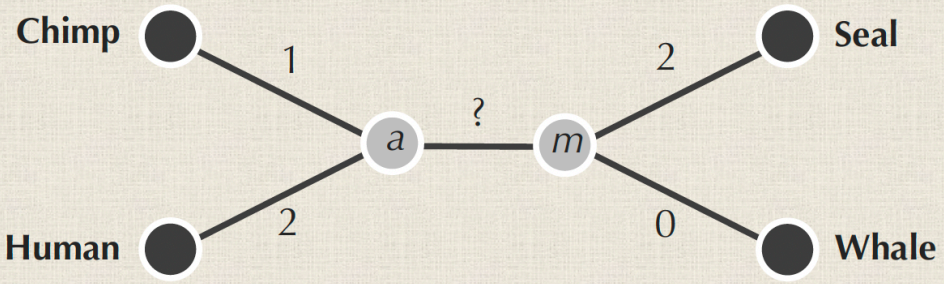
	Chimp	Human	*m*
Chimp	0	3	4
Human	3	0	5
m	4	5	0



$d_{Chimp,a} = 1$, so $d_{Human,a} = d_{Human,\ Chimp} - d_{Chimp,a} = 2$

An Idea for Distance-Based Phylogeny

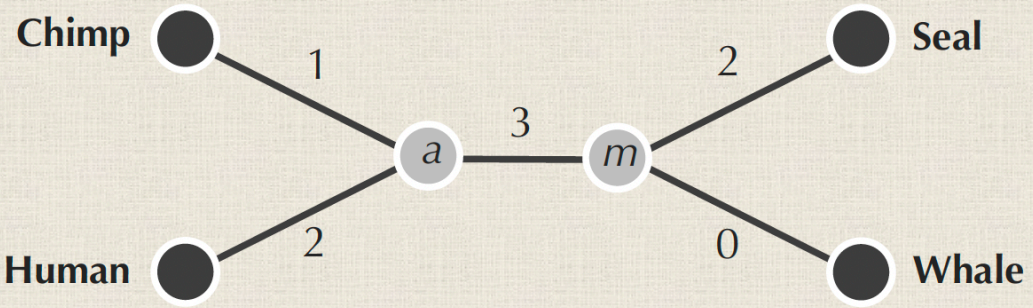
	Chimp	Human	*m*
Chimp	0	3	4
Human	3	0	5
m	4	5	0



STOP: And $d_{a,\,m}$?

An Idea for Distance-Based Phylogeny

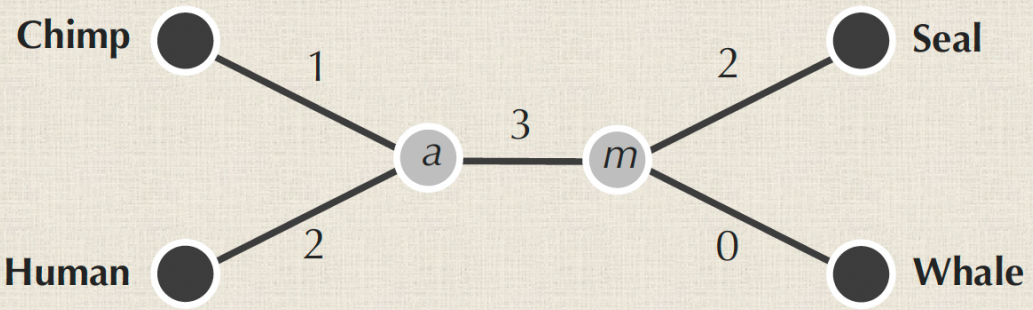
	Chimp	Human	*m*
Chimp	0	3	4
Human	3	0	5
m	4	5	0



An Idea for Distance-Based Phylogeny

The matrix fits for the constructed evolutionary tree

	Chimp	Human	Seal	Whale
Chimp	0	3	6	4
Human	3	0	7	5
Seal	6	7	0	2
Whale	4	5	2	0



Class Activity 1

	A	**B**	**C**	**D**
A	0	5	4	7
B	5	0	5	2
C	4	5	0	7
D	7	2	7	0

Construct an evolutionary tree for the given table if possible

Class Activity 2

	A	**B**	**C**	**D**
A	0	5	12	13
B	5	0	9	10
C	12	9	0	5
D	13	10	5	0

Construct an evolutionary tree for the given table if possible