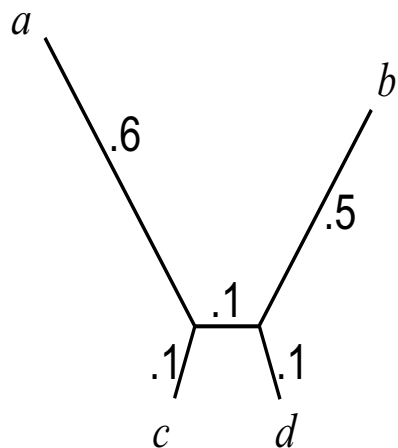


Slides for lecture on Neighbor Joining

Ex. Suppose 4 taxa are related by the following tree:



	a	b	c	d
a		1.2	.7	.8
b			.7	.6
c				.3

Which tree will UPGMA or FM construct?

Primate mitochondrial DNA sequences, HindIII

Hayasaka, K., T. Gojobori, and S. Horai. MBE (1988) 5:626-644.

	Gorilla	Orangutan	Human	Chimp	Gibbon
Gorilla	0	0.1890	0.1100	0.1130	0.2150
Orangutan	0.1890	0	0.1790	0.1920	0.2110
Human	0.1100	0.1790	0	0.0940	0.2050
Chimp	0.1130	0.1920	0.0940	0	0.2140
Gibbon	0.2150	0.2110	0.2050	0.2140	0

Primate mitochondrial DNA sequences, HindIII

Hayasaka, K., T. Gojobori, and S. Horai. MBE (1988) 5:626-644.

	Gorilla	Orangutan	Human	Chimp	Gibbon
Gorilla	0	0.1890	0.1100	0.1130	0.2150
Orangutan	0.1890	0	0.1790	0.1920	0.2110
Human	0.1100	0.1790	0	0.0940	0.2050
Chimp	0.1130	0.1920	0.0940	0	0.2140
Gibbon	0.2150	0.2110	0.2050	0.2140	0

$R_i = \sum_{j=1}^N d_{ij}$, total dissimilarity between taxon S_i and all other taxa

$$R_1 = R_{\text{Gorilla}} = .189 + .11 + .113 + .215 = .627$$

Neighbor Joining Algorithm

Step 1: Given dissimilarity data for N taxa, compute a new table of values of M using the equation below. Choose the smallest value in the table for M to determine which taxa to join.

$$M_{nm} = (N - 2)d_{nm} - R_n - R_m.$$

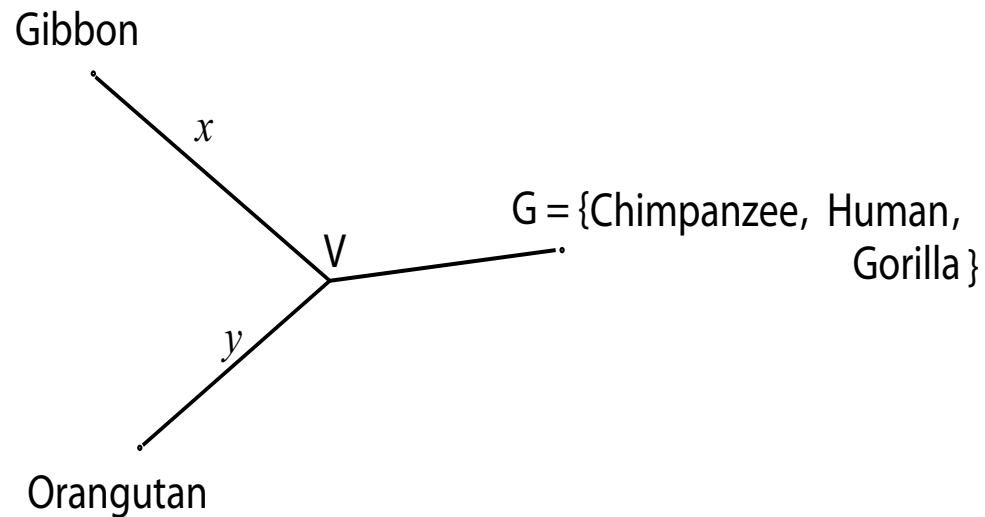
The matrix $M =$

	Gorilla	Orangutan	Human	Chimp	Gibbon
Gorilla		-0.8310	-0.8850	-0.9010	-0.8270
Orangutan			-0.8220	-0.8080	-0.9830
Human				-0.9190	-0.8180
Chimp					-0.8160

The matrix M =

	Gorilla	Orangutan	Human	Chimp	Gibbon
Gorilla		-0.8310	-0.8850	-0.9010	-0.8270
Orangutan			-0.8220	-0.8080	-0.9830
Human				-0.9190	-0.8180
Chimp					-0.8160

Step 2: If S_i and S_j are to be joined at a new vertex V , temporarily collapse all other taxa into a single group G , and determine the lengths of the edges from S_i and S_j to V by using using the 3-point formulas for S_i , S_j , and G as in the algorithm of Fitch and Margoliash.



Original

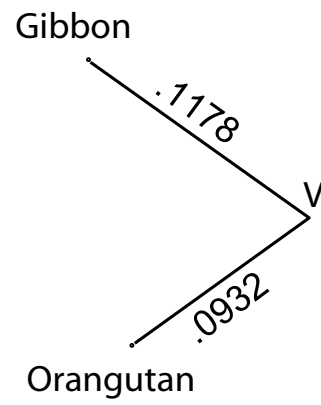
	Gorilla	Orangutan	Human	Chimp	Gibbon
Gorilla		0.1890	0.1100	0.1130	0.2150
Orangutan			0.1790	0.1920	0.2110
Human				0.0940	0.2050
Chimp					0.2140

Collapsed Table

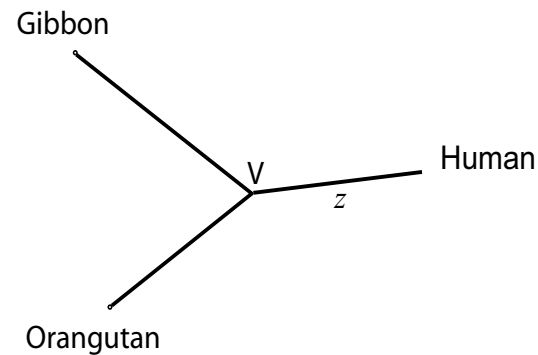
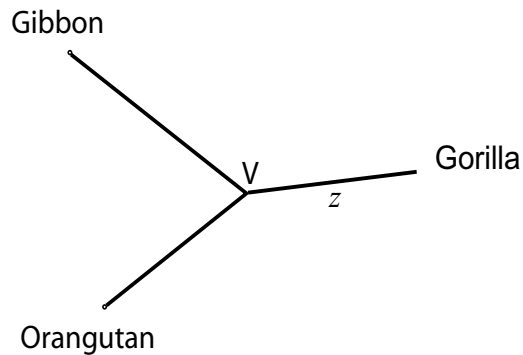
	Orangutan	Gibbon	Gorilla-Human-Chimp
Orangutan		.2110	.1867
Gibbon			.2113

$$x \approx .1178, \quad y \approx .0932$$

We get....



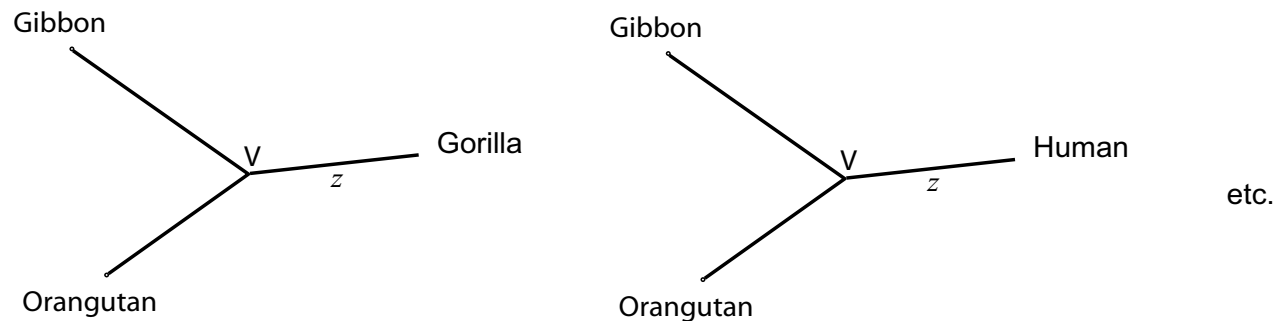
Step 3: Determine distances/dissimilarities from each of the taxa S_k in G to V by applying the 3-point formulas to the distance data for the three taxa S_i , S_j and S_k . Now include V in the table of dissimilarity data, and drop S_i and S_j .



etc.

	V	Gorilla	Human	Chimp
V		?	?	?
Gorilla			0.1100	0.1130
Human				0.0940

Step 3: Determine distances/dissimilarities from each of the taxa S_k in G to V by applying the 3-point formulas to the distance data for the three taxa S_i , S_j and S_k . Now include V in the table of dissimilarity data, and drop S_i and S_j .



	V	Gorilla	Human	Chimp
V		0.0965	0.0865	0.0975
Gorilla			0.1100	0.1130
Human				0.0940

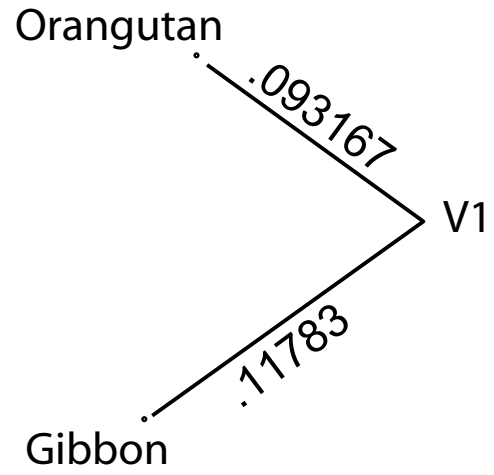
Step 4: The distance table now includes $N - 1$ taxa. If there are only 3 taxa, use the 3-point formulas to finish. Otherwise go back to step 1.

	V	Gorilla	Human	Chimp
V		0.0965	0.0865	0.0975
Gorilla			0.1100	0.1130
Human				0.0940

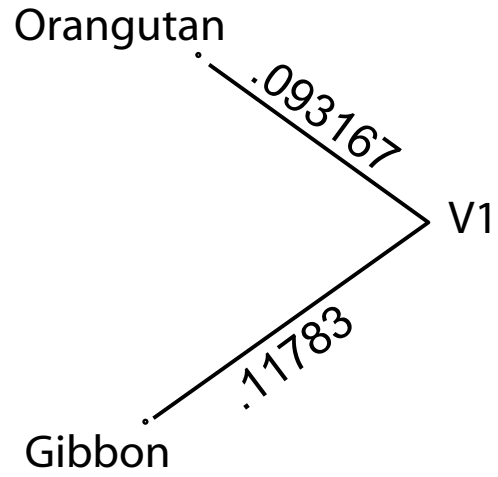
Go back to Step 1

Go back to Step 1 [with MATLAB](#).

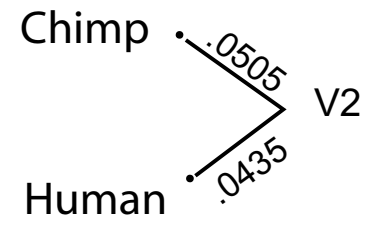
First iteration:



First iteration:



Second iteration:



Last iteration:

