

PHYLOGENETIC ANALYSIS WITH AMINO ACID DATA

Amino acids are the building blocks of proteins. These molecules are commonly studied to reconstruct evolutionary trees. This activity allows you to compare amino acid sequences of proteins from different species and use them to deduce evolutionary relationships.

Below is a comparison of amino acid sequences of the electron transport protein cytochrome c from three species of animals, designated by the capital letters A, B, and C. Each amino acid in the proteins is represented by a small letter. The differences in amino acid sequence are Bolded. (Note that for the purposes of this activity, the data are simplified.)

Species A: **a** v l i f **m** f a

Species B: v v l i f **p** w d

Species B: v v l i f **p** w d

Species C: **a** v l **g** f **m** f a

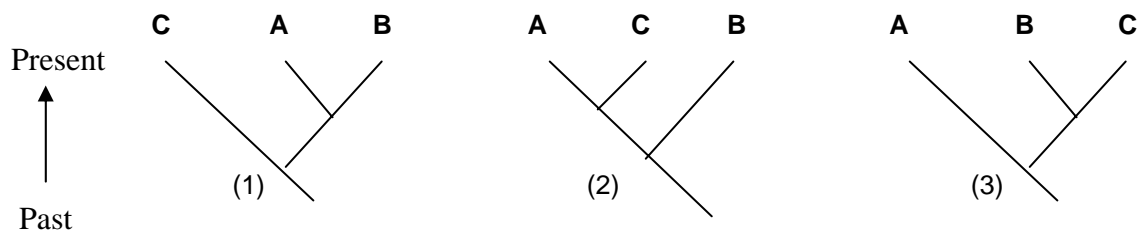
Species A: a v l i f m f a

Species C: a v l g f m f a

You can fill in this chart to summarize the numbers of differences in amino acid sequence among the three species. Count the differences and fill in a species character matrix below.

B		XXXXXXXXX
C		
	A	B

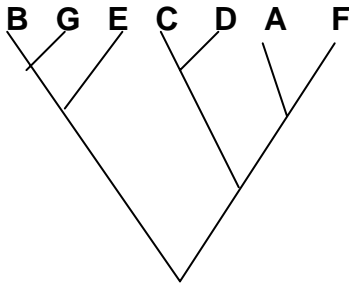
Based on the character matrix above, which of the three phylogenetic trees below, best describes the possible evolutionary relationships among species A, B, and C? The differences summed in the matrix can be used to infer that Species A and C are more closely related to each other than either is to B. In other words, species B is a distant relative of A and C, and species A and C diverged from each other more recently. By convention, time is the implicit vertical axis such that time proceeds from the past, low in the trees to the present at the top of the trees.



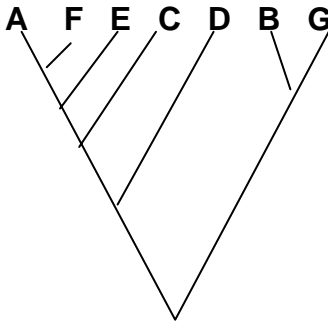
Incidentally, these data are based on comparisons of cytochrome c from a dog, a horse, and a rattlesnake. The dog and horse, which are mammals, are more closely related to each other than they are to the rattlesnake, which is a reptile.

Study the following expanded character matrix comparing the amino acid sequences of 7 species. From the trees below, determine the evolutionary tree that best corresponds with the amino acid sequence data. Remember, the more differences between a corresponding pair, the less related and thus the more distantly linked in a tree. Conversely, the fewer differences between a pair the more similarity and the closer the linkage in a tree.

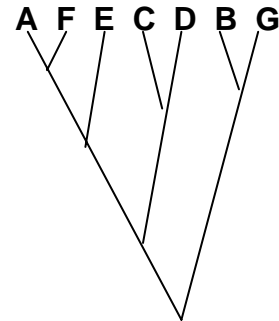
B	14					
C	5	15				
D	5	15	2			
E	3	14	4	4		
F	1	14	5	5	3	
G	14	7	15	15	14	14
	A	B	C	D	E	F



(1)



(2)



(3)

Phylogenetic tree number 3.) best corresponds to the amino acid sequence data. If species A is a horse, F is a donkey, E is a rabbit, C and D are birds, B is a moth and G is a fly, then clearly, the horses closest relative is the donkey or F. The horse, donkey, and rabbit are all mammals. The two birds are closely related to each other. The birds and the mammals are all vertebrates, and the tree shows their common heritage. The fruit fly and moth share a common evolutionary path, but it has been a long time since the insects diverged from the vertebrates.

On each tree, draw a circle around all taxa in the clades/groups called mammals, aves, vertebrates, invertebrates, and animals. The “best” tree according to phylogenetic criteria is the tree that contains only monophyletic clades/groups. Does the tree you selected contain only monophyletic groups? If not, you should re-analyze the character matrix.