

Homework 3

Before starting, work through the Python primer. Let me know if you have trouble understanding how to install or use Python, or if you don't understand something in the Python primer.

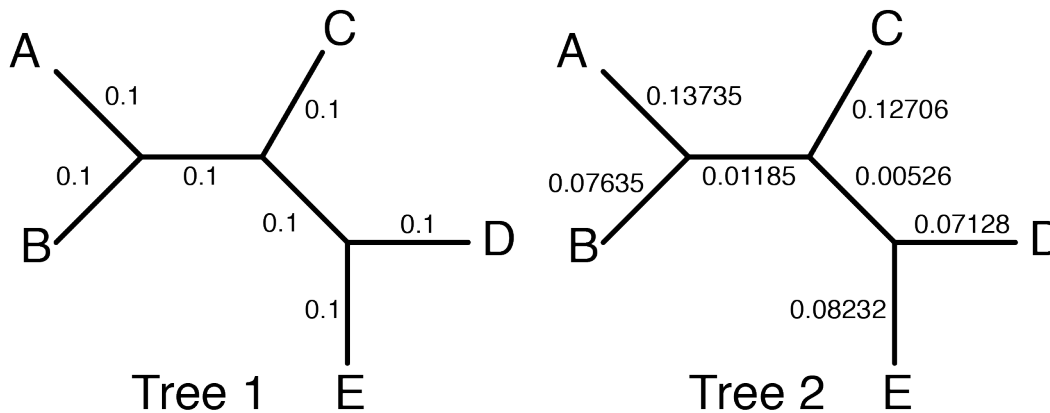
The numbers (x_{ij}) in the lower triangle of the table below represent the number of nucleotide differences between each possible pair of 5 DNA sequences. The total number of nucleotides is $n = 1296$. The raw dissimilarity (p-distance) between taxon i and taxon j would thus equal x_{ij}/n .

	A	B	C	D	E
A		0.213707			
B	241				
C	292	248			
D	262	187	231		
E	260	205	242	180	

Compute JC69 evolutionary distances (d_{ij}) using the x_{ij} numbers provided and **write the distances in the upper triangle** (use the script you wrote in the Python Primer lab to do the calculation). The distance between A and B has been done for you so that you can check to make sure you are doing it correctly. (Note: if you choose to do this without python, keep in mind that on many calculators you must use “ln” for natural logarithms rather than “log”.)

$$d_{ij} = -\frac{3}{4} \log \left(1 - \frac{4}{3} \frac{x_{ij}}{n} \right)$$

Using the JC69 distances (d_{ij}) and the lengths of corresponding paths through each tree (p_{ij}), compute the sum of squares (SS) for both trees shown below (tree 1 has all edges equal to 0.1, while tree 2 has edges estimated by PAUP*). Use power $k = 2$ when computing SS. (Again, use the script you wrote in the Python Primer lab.)



Fill in the unshaded empty squares in the table below for each of the trees. I've done “A vs. B” for you so that you can confirm that your python script is working correctly. Round everything to 5 decimal places. **Indicate which tree fits the data best according to least squares.**

	Tree 1			Tree 2		
	p_{ij}	d_{ij}	SS	p_{ij}	d_{ij}	SS
A vs. B	0.20000	0.21371	0.00411	0.21370	0.21371	0.00000
A vs. C						
A vs. D						
A vs. E						
B vs. C						
B vs. D						
B vs. E						
C vs. D						
C vs. E						
D vs. E						
total						