Assignment #4 due Thursday, February 7

We are going to write a UPGMA algorithm in Python and analyze some human and Neanderthal data.

Download "mitoData.py" from Content:

1. neandList (and humansList) are species lists. They are in the form of a list of trees represented by tuples. This form will be useful.
2. neandMatrix (and humansMatrix) are dictionaries specifying pairwise distances between species. They are in the form where the key is a tuple which is a pair of trees, e.g., (("species A",(),()), ("species B",(),())) and the value is the pairwise distance. This form will also be useful.

Human data from modern humans; distances based on mitochondrial sequence

San	San individual from southern Africa
Yoruba	Yoruba individual from western Africa
Finnish	Finnish individual from northern Europe
Kikuyu	Kikuyu individual from eastern Africa
Papuan	Papuan individual from New Guinea
Han	Han individual from China

Neanderthal data from fossils, modern humans and chimpanzee; distances based on mitochondrial sequence

Chimpanzee	common Chimpanzee
San	San individual from southern Africa
Yoruba	Yoruba individual from western Africa
Finnish	Finnish individual from northern Europe
Neanderthal	~38,000-year-old fossil of Neanderthal individual from Croatia
Kostenki	~30,000-year-old fossil of individual of Russia with modern human appearance

At the end are some hints to write the UPGMA algorithm. Use the upgma function to generate two trees:

```
>>> neandTree = upgma(neandList, neandMatrix)
>>> humansTree = upgma(humansList, humansMatrix)
```

Turn in your code, the output of the code for both the Neanderthal and human data sets, and two hand-drawn trees where you have assumed a human-chimpanzee divergence of 6 million years to calibrate the trees. In addition, answer the following questions:

1. Are Neanderthals more closely related to modern Europeans than to other modern humans? Does it appear that modern Europeans descended from Neanderthals?
2. Does this data support an African origin for modern humans?
3. Compare the divergence time of Neanderthals from modern humans with that among modern human groups. What is the ratio between these two?

HINTS:

Write the following functions:

1. **findClosestPair(speciesList, Distances)**
   Inputs speciesList (list of trees represented by tuples) and Distances dictionary and returns the pair of trees in speciesList that are closest. If the closest trees are treei and treej, the function should return the tuple (treei, treej).
2. **updateDist(speciesList, Distances, newTree)**
   Does not return anything. Updates the Distances dictionary by adding the distances between newTree and all the other trees in speciesList. This function uses the "countleaves" function from lecture.
3. **upgma(speciesList,Distances)**
   This function runs the UPGMA algorithm described in lecture. It repeats the following steps until there is only one tree left in speciesList, at which point this tree is returned.
   a. Find the closest pair of species in speciesList (use "findClosestPair")
   b. Merge these two species into a new tree. If the two species are treei and treej and their pairwise distance is d, the new tree will be the tuple (d/2, treei, treej). Using the distance in this way will be useful.
   c. Update the speciesList by removing these two species. Note: the function speciesList.remove(blah) will remove the item named blah from the list named speciesList.
   d. Update the Distances dictionary (use "updateDist")
   e. Add the new merged tree into speciesList