AMERICAN MUSEUM & NATURAL HISTORY

Seminars on Science

EVOLUTION WEEK 2 ASSIGNMENT: BUILD-A-TREE ACTIVITY

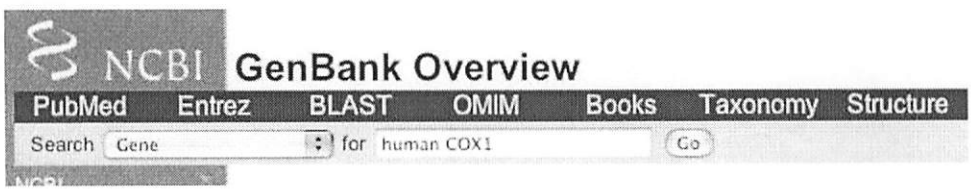
The following protocol will take you through the steps required to build a phylogenetic tree with molecular data using **SeaView**, a freely available computer program. Molecular data for many known organisms is available from **GenBank**, a database hosted by the National Center for Biotechnology Information (NCBI). NCBI provides a helpful resource for learning how to understand and interpret phylogenetic trees, which can be found at http://www.ncbi.nlm.nih.gov/About/primer/phylo.html . These procedures can be adapted for students to use in formulating and testing evolutionary hypotheses.

FIND A NUCLEOTIDE SEQUENCE

1. Go to **GenBank**:
   http://www.ncbi.nlm.nih.gov/Genbank/

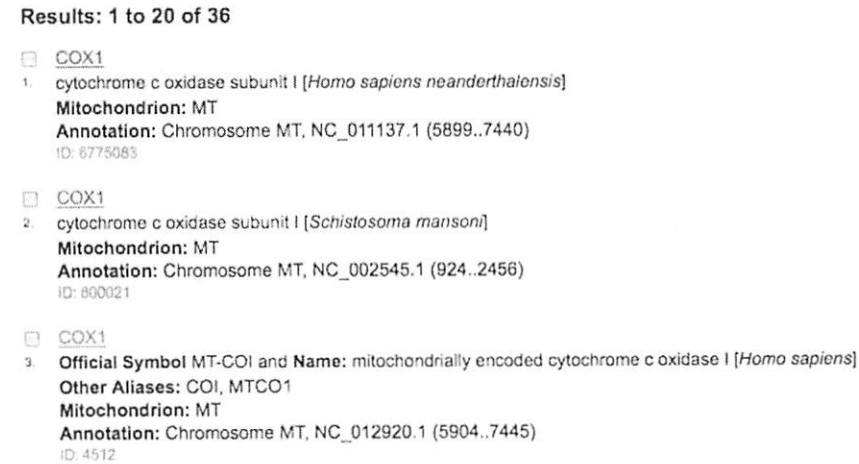
2. In the "Search" window, select "Gene" from the pull-down menu. Type in the organism you're looking for and the gene if you know it. In this case, start with COX1. Push **Go**.

   The cytochrome oxidase 1 (COX1) gene is a mitochondrial gene involved in respiration. It is a commonly sequenced gene and there are COX1 sequences available for a wide variety of organisms. You can use other genes, but you must choose the same gene for all the organisms you are comparing.

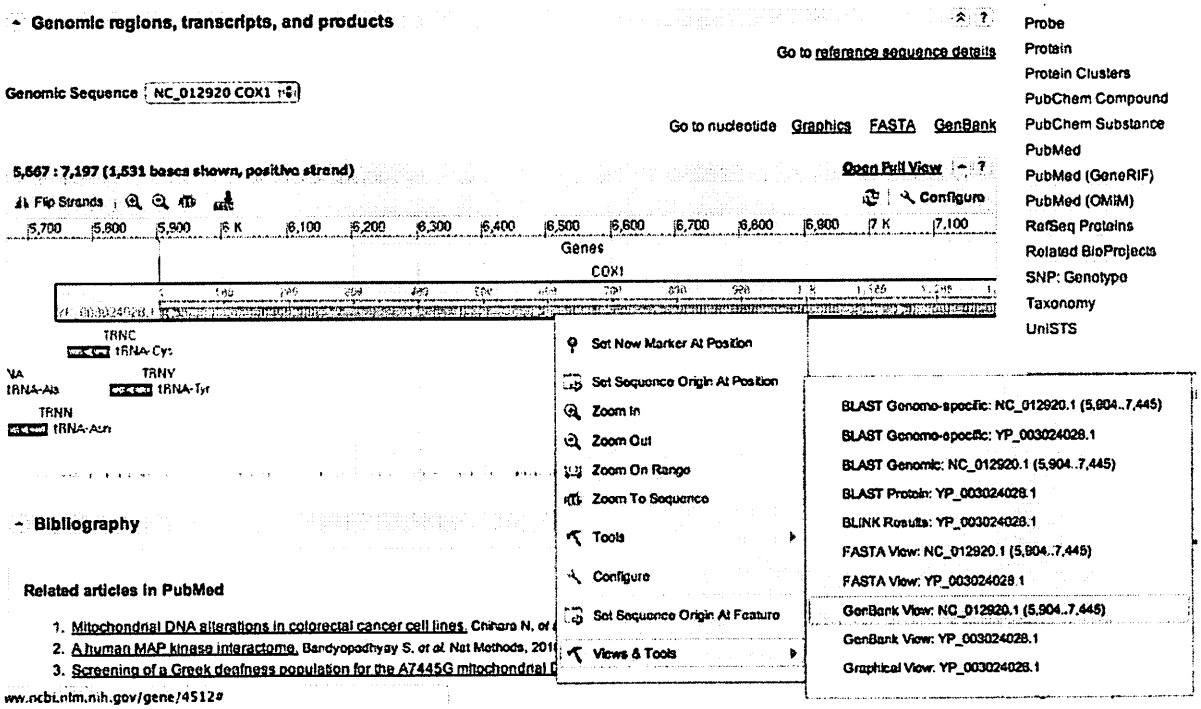


   Tip: For more unusual species, try a search using the genus name of the species you are looking for instead of the common name to enhance the specificity of your search.

3. In the search results, click on the "COX1" link for the species you want to investigate.



4. On the sequence summary page, scroll down to "Genomic, regions, transcripts, and products." Right click on the red banner. A dropdown menu box will appear. Scroll down and hover your cursor over "Views and Tools". Another menu box will appear. Select FASTA View: NC.



5. The sequence viewer will appear, showing many lines of DNA sequence (combinations of A, T, G, and C). Highlight all of this text, including the first line (this is the sequence header). Copy the text to your clipboard.

```
>gi|17981852:5905-7446 Homo sapiens mitochondrion, complete genome
ATGTTCGCCGACCGTTGACTATTCTCTACAAACCACAAAGACATTGGAACACTATACCTATTATTCGGCG
CATGAGCTGGAGTCCTAGGCACAGCTCTAAGCCTCCTTATTCGAGCCGAGCTGGGCCAGCCAGGCAACCT
TCTAGGTAACGACCACATCTACAACGTTATCGTCACAGCCCATGCATTTGTAATAATCTTCTTCATAGTA
ATACCCATCATAATCGGAGGCTTTGGCAACTGACTAGTTCCCCTAATAATCGGTGCCCCCGATATGGCGT
TTCCCCGCATAAACAACATAAGCTTCTGACTCTTACCTCCCTCTCTCCTACTCCTGCTCGCATCTGCTAT
AGTGGAGGCCGGAGCAGGAACAGGTTGAACAGTCTACCCTCCCTTAGCAGGGAACTACTCCCACCCTGGA
GCCTCCGTAGACCTAACCATCTTCTCCTTACACCTAGCAGGTGTCTCCTCTATCTTAGGGGCCATCAATT
TCATCACAACAATTATCAATATAAAACCCCCTGCCATAACCCAATACCAAACGCCCCTCTTCGTCTGATC
CGTCCTAATCACAGCAGTCCTACTTCTCCTATCTCTCCCAGTCCTAGCTGCTGGCATCACTATACTACTA
ACAGACCGCAACCTCAACACCACCTTCTTCGACCCCGCCGGAGGAGGAGACCCCATTCTATACCAACACC
TATTCTGATTTTTCGGTCACCCTGAAGTTTATATTCTTATCCTACCAGGCTTCGGAATAATCTCCCATAT
TGTAACTTACTACTCCGGAAAAAAAGAACCATTTGGATACATAGGTATGGTCTGAGCTATGATATCAATT
GGCTTCCTAGGGTTTATCGTGTGAGCACACCATATATTTACAGTAGGAATAGACGTAGACACACGAGCAT
ATTTCACCTCCGCTACCATAATCATCGCTATCCCCACCGGCGTCAAAGTATTTAGCTGACTCGCCACACT
CCACGGAAGCAATATGAAATGATCTGCTGCAGTGCTCTGAGCCCTAGGATTCATCTTTCTTTTCACCGTA
GGTGGCCTGACTGGCATTGTATTAGCAAACTCATCACTAGACATCGTACTACACGACACGTACTACGTTG
TAGCTCACTTCCACTATGTCCTATCAATAGGAGCTGTATTTGCCATCATAGGAGGCTTCATTCACTGATT
TCCCCTATTCTCAGGCTACACCCTAGACCAAACCTACGCCAAAATCCATTTCACTATCATATTCATCGGC
GTAAATCTAACTTTCTTCCCACAACACTTTCTCGGCCTATCCGGAATGCCCCGACGTTACTCGGACTACC
CCGATGCATACACCACATGAAACATCCTATCATCTGTAGGCTCATTCATTTCTCTAACAGCAGTAATATT
AATAATTTTCATGATTTGAGAAGCCTTCGCTTCGAAGCGAAAAGTCCTAATAGTAGAAGAACCCTCCATA
AACCTGGAGTGACTATATGGATGCCCCCCACCCTACCACACATTCGAAGAACCCGTATACATAAAATCTA
GA
```

CREATE A FASTA FILE

6. Open your text-editing program (**Notepad** (PC) or **TextEdit** (Mac) – please do not use Microsoft Word). Open a new document. Paste the DNA sequence text from your clipboard.

7. Save the file as **sequences.txt**

   Tip: Mac users must convert the file to plain text before saving. To do this, click on the Format menu at the top of your screen and then choose the option "Make plain text". If the file extension is .rtf, the file will not be read.

8. Keep the file open. You'll need it in the next step.

ADD MORE SEQUENCES

9. Return to **GenBank**:
   http://www.ncbi.nlm.nih.gov/Genbank/

10. Search for additional organisms that you want to add to your tree. Follow steps 2-5 again.

11. Return to your text-editing program. Instead of creating a new file, copy and paste the new sequences into your FASTA file below the first one. Make sure you include the sequence header beginning with ">" at the start of each sequence. Save your file as you go.

12. Repeat Steps 2-11 until you have a minimum of 8-10 sequences in your FASTA file.

13. COX1 is a useful gene for comparing organisms within relatively closely related groups and is especially good for studying the relationships between animals. The mutation rate in this gene is too high for accurate comparison between distantly related organisms, such as a paramecium and a human. For best results, choose the majority of your organisms from **WITHIN** a group that you are interested in, such as Carnivora (carnivores) or Aves (birds). In addition, you must choose one organism that is **OUTSIDE** this group, which will become your ***outgroup.***

EDIT THE FASTA FILE

14. Let's look at the sequence header at the top of each DNA sequence. The species title that will show up on your tree will be the first line of each set of sequence data following the ">" symbol. This header can be edited for clarity, but you **MUST** preserve the ">" symbol. You can use the scientific name or the common name to identify your sequence.

    For example, the human sequence begins like this:
    ```
    >gi|251831106:5904-7445 Homo sapiens mitochondrion, complete genome
    ```

    This can be edited to look like this:
    ```
    >Human
    ```

    Or this:
    ```
    >Homo_sapiens
    ```

    Tip: Be sure to add an underscore "_" instead of a space between words. That way all words you include will show up as labels on your tree.

15. Make sure you have a return after your header and after the end of the sequence. Your list of sequences should appear as follows:

```
>Homo_sapiens
ATGTTCGCCGAC...

>Pan_paniscus
ATGTTCACCGAC...
```

Tip: Each sequence should be approximately 1500 bp, which should appear about the length of a "paragraph" in your file. If any seem to be much longer, see Step 19.

ALIGN YOUR SEQUENCES

16. Open **SeaView.**

17. In the File menu, choose **Open.**

18. Select your file. Your sequences should show up in the SeaView window. Check to see that they are all present, labeled correctly, and that the first few bases in the SeaView window correspond to the first few bases of each sequence in your text file.

Tip: If your file will not load into SeaView, or does not load correctly, check for the following common problems:

a. *Your file is in .doc or .rtf format. Look at the extension after the file name. It must end in .fasta or .txt. Open it in Notepad or Textedit and save as a plain text file.*
b. *You have accidentally deleted the ">" character at the beginning of one or more sequence headers. Simply add ">"back to the front of each sequence header.*
c. *You are missing one or more carriage returns at the end of each header and sequence. To fix this, place your cursor at the end of each sequence and header and consciously add a return even if one appears to be there already.*
d. *If a.-c. do not help, try loading the test file primatecox1.txt provided in your download folder to check the application. In addition, post your .txt file to the Assignment discussion board for assistance.*

19. Scan your sequences. They should all be roughly 1500 bp in length. If any of them are much longer than the others (eg. 16,000 bp), you may have accidentally downloaded the entire mitochondrial genome for that organism. Go back to Step 4 and make sure that you have extracted the COX1 portion only.

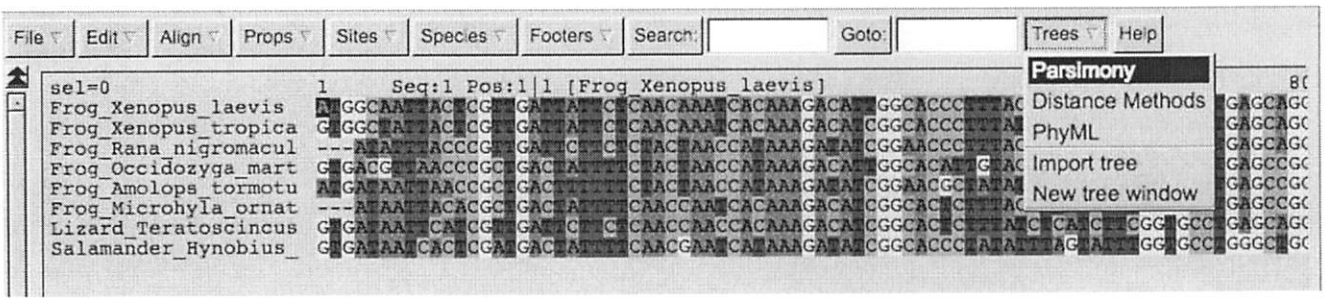
20. Under the Align menu choose **Align All.**

21. A new window will pop up that shows the alignment algorithm running in real time. When it is complete, click **OK.**

22. Scan the sequences again. Do you notice any changes? You may see that gaps have been inserted into some sequences. Think about why the program might be doing this. It's also possible you won't see any changes at all, particularly if your chosen organisms are very closely related.

BUILD THE TREE

23. Go to the Trees menu towards the right of the menu bar. Click "Trees" then "Parsimony". This will build a tree using the principle of Maximum Parsimony.



24. Click "OK" to agree to the default settings for the various options available.

25. A new window will appear with your tree. The relationships it shows may appear strange or counterintuitive to you at first. It is **VERY** important to set your outgroup (see next step) before analyzing your tree.

MANIPULATE THE TREE

26. Click **Re-root**. Black boxes will appear on your tree. Click the black box next to the organism that will serve as the outgroup. This organism should be the one you selected in Step 13 as the organism outside your main group of interest, making it the **most different** from all the other organisms on your tree.

27. Use the **Swap** command to swap the branches around the nodes by clicking the black box representing the node that you would like to swap. You can also zoom in on a subsection of your tree by clicking **Subtree** and then the black box representing the node you want to zoom in on. Click **Full** to eliminate buttons on the tree or to return to the main tree.

SHARE YOUR TREE

28. Take a screenshot of your tree. There are different methods to do this, depending on your computer:

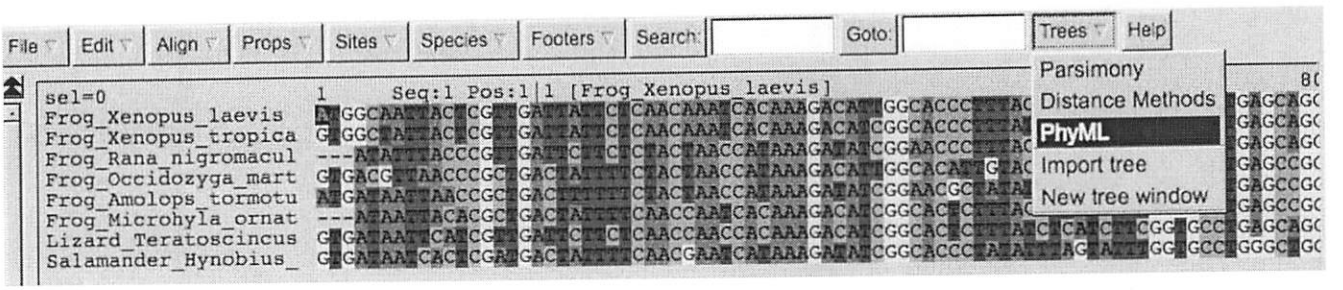
    Using Windows:
    a. Press the **Print Screen** button on your keyboard.
    b. Open **Microsoft Word** or **Microsoft Paint**.
    c. In the Edit menu, choose **Paste**. Your image will appear in the document.
    d. Save the document to share with others.

Using Mac:

a. Press **Command-Shift-3** using your keyboard (do not use the 3 on the number pad of a full size keyboard).
b. A file will appear on your desktop called **Picture 1**.
c. This image is ready to insert into **Microsoft Word**.

TRY ANOTHER TREE-BUILDING METHOD

29. Go to the Trees menu towards the right of the menu bar. Click "Trees" then "PhyML". This will build a tree using the principle of **Maximum Likelihood**.

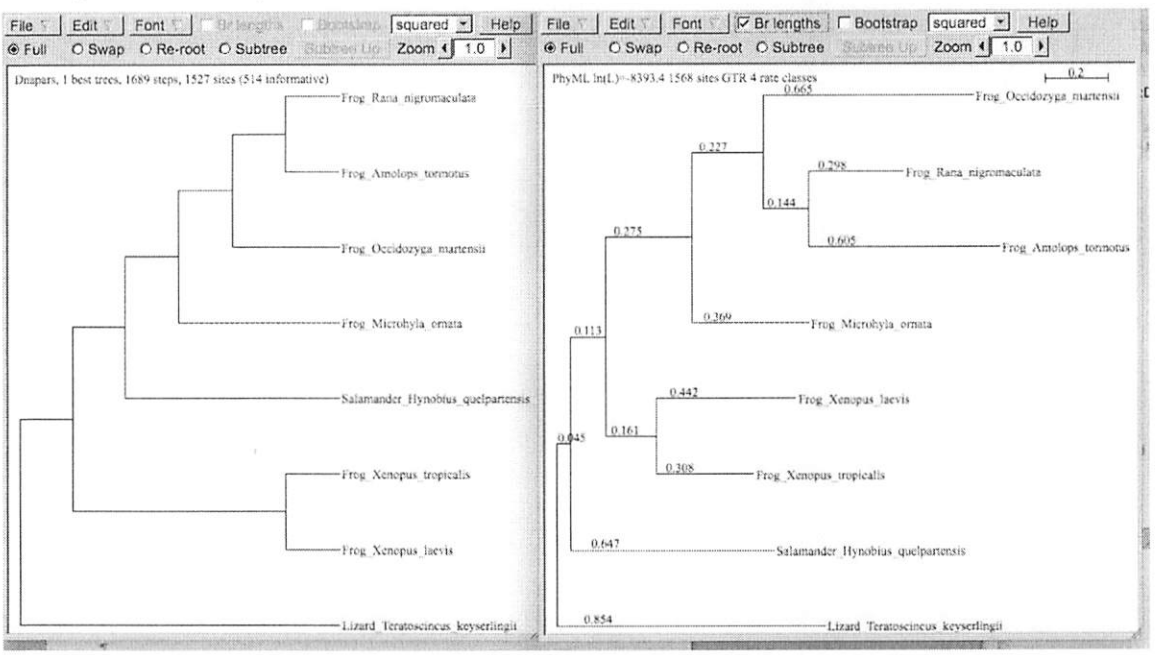


30. Click "Run" to agree to the default settings for the various options available.

31. A new window will pop up that shows the tree-building algorithm running in real time. When it is complete, click **OK**:

32. Your new Maximum Likelihood tree will appear. Repeat step 26 to set the outgroup for this tree before you analyze and compare your trees. You can also use the **Swap** command again to re-arrange the branches.

33. Click **Full** to return to the main tree view, then click **Br Lengths** near the top of your tree window. Numbers will appear along your branches. Put your two trees side by side so that you can compare and contrast their appearance and the relationships they show.



34. To complete this assignment, please submit a file to the Dropbox that contains both your tree images and your answers to the Discussion questions below.

DISCUSSION

- Why do you need to align the sequences (Steps 16-22) before building the tree?
- Why did you choose this group of organisms to build your trees?
- What organism did you pick as your outgroup? Why? What function does the outgroup serve?
- Describe some of the relationships depicted by your tree.
- When you swap nodes (Step 27) you are only rearranging the way the tree is displayed, but the relationships it depicts are unchanged. What are some of the advantages of being able to manipulate a tree in this manner?
- We asked you to create your tree using both the Maximum Parsimony and Maximum Likelihood methods. How do the two trees compare? What extra information does the Maximum Likelihood tree provide?

HELPFUL LINKS

Understanding Evolution – Phylogenetic systematics, a.k.a. evolutionary trees
http://evolution.berkeley.edu/evolibrary/article/phylogenetics_01

National Center for Biotechnology Information – Phylogenetics Fact Sheet
http://www.ncbi.nlm.nih.gov/About/primer/phylo.html

Tree Thinking Quizzes – Two quizzes to check your understanding of evolutionary trees
http://www.sciencemag.org/content/310/5750/979/suppl/DC1

SeaView –by Pôle Bioinformatique Lyonnais
http://pbil.univ-lyon1.fr/software/seaview.html