**BLAST Lab Part II**

[**http://blast.ncbi.nlm.nih.gov/Blast.cgi**](http://blast.ncbi.nlm.nih.gov/Blast.cgi)

## Overview

In the online activity “Biodiversity and Evolutionary Trees: An Activity on Biological Classification,” you generated a phylogenetic tree of molluscs using only shell morphology information. In this exercise, you will revisit that classification and reconstruct the phylogenetic tree using BLAST, software that compares DNA sequences.

## Format of DNA Sequence Information

There are different formats for representing DNA sequences. Shown below is a partial sequence from the dog’s *cytochrome oxidase subunit I* (*COI*) gene in FASTA format. FASTA format starts with a “>,” followed by information about the file to the end of the first line, followed by the DNA sequence.

>gi|377685879|gb|JN850779.1| Canis lupus familiaris isolate dog\_3 cytochrome oxidase subunit I (COI) gene, partial cds; mitochondrial

TACTTTATACTTACTATTTGGAGCATGAGCCGGTATAGTAGGCACTGCCTTGAGCCTCCTCATCCGAGCC

GAACTAGGTCAGCCCGGTACTTTACTAGGTGACGATCAAATTTATAATGTCATYGTAACCGCCCATGCTT…

A file containing FASTA format sequence information may contain multiple sequences one after another. For example:


## What Sequences Do We Choose to Compare?

In modern taxonomic practice, scientists routinely analyze the DNA from specimens they collect to obtain a “DNA barcode,” a short DNA sequence unique to a particular organism, which is used to identify the species it belongs to. For animals and many other eukaryotes, the mitochondrial cytochrome c oxidase subunit I (*COI*) gene, that encodes an important enzyme for cellular respiration, has been used to generate such barcodes. As a result, *COI* sequences are available from a wide range of species, making it possible to use this gene sequence to explore phylogenetic relationships.

*COI* is a good choice for DNA barcoding, because, in general, there is little variation in *COI* sequences of organisms within the same species, while there is significant variation in *COI* sequences of organisms from different species. Therefore, a *COI* sequence provides a unique sequence signature for a particular species. For the same reasons, *COI* gene is suitable for comparing phylogenetic relationships between species.

Because the *COI* sequences are so similar within the same species, the *COI* gene is not a good choice for studying variations within the same species, or even among species that have recently speciated. *COI* sequences also have a low mutation rate among many species of plants, and cannot be used for DNA barcoding or phylogenetic comparisons of those species.

## Phylogeny of Neogastropods, Tonnoids, and Cowries

In the online seashell sorting activity, “Biodiversity and Evolutionary Trees: An Activity on Biological Classification,” one of the questions you had to answer was which two groups of gastropods among neogastropods, tonnoids, and cowries are more closely related. Based on the information in that activity, the correct answer was tonnoids and cowries.

Let’s confirm that conclusion using molecular data.

1. **Open the file molluscs1.txt from the webpage documents section.**
2. **Open the BLAST program – from the home page select the following:**
	1. **Select “Nucleotide BLAST”**
	2. **Under the large box, check the box that says “Align two or more sequences”**
	3. **Copy and paste the first “neogastropod” sequence into the FASTA sequence into the box.**
	4. **Copy and paste the second TWO sequences into the SECOND box**
	5. **Click the radio button that says “More dissimilar sequences (discontinguous megablast)”**
	6. **Hit “BLAST”!**

(For this exercise, the DNA sequences used were obtained from representative species from each group as follows: neogastropods = *Conus magus*; tonnoids = *Bursa granularis*; and cowries = *Cypraea tigris*)

1. **Once your search is complete – click on the link that says “[Distance tree of results]” – it should be near the top.**
2. **Record your results in on a separate sheet of paper. *Explain your results.***

## Phylogeny of Molluscs

The final phylogeny obtained from the online seashell sorting activity is shown below.

1. *Conus magus*
2. *Neritina communis*
3. *Bursa ignobilis*
4. *Conus capitaneus*
5. *Cypraea annulus*
6. *Conus marmoreus*
7. *-*
8. *Distorsio anus*
9. *Conus omaria*
10. *Cypraea isabella*
11. *Pecten pallium*
12. *Cypraea tigris*
13. *Conus ebraeus*
14. *-*
15. *Conus chaldeus*
16. *-*
17. *Imbricaria conularis*
18. *Conus circumcisus*
19. *Cypraea moneta*
20. *Turris babylonia*

We will compare the phylogenetic tree above, obtained using morphological information, with a tree obtained using DNA sequence data. (Note for the instructions – BLAST may have saved your information – but double-check all settings before you BLAST!)

1. **Open the file molluscs2.txt from the webpage documents section.**
2. **Open the BLAST program – from the home page select the following:**
	1. **Select “Nucleotide BLAST”**
	2. **Under the large box, check the box that says “Align two or more sequences”**
	3. **Copy and paste the first sequence ONLY into the FASTA sequence into the box.**
	4. **Copy and paste ALL of the remaining sequences into the SECOND box**
	5. **Click the radio button that says “More dissimilar sequences (discontinguous megablast)”**
	6. **Hit “BLAST”!**

NOTE: \* Because *COI* gene sequences for five of the species used in the online seashell activity were not available, we used substitute sequences from closely related species. Substitutions were as follows:

*Neritina pulligera* was substituted for *Neritina communis;
Distorsio reticularis* for *Distorsio anus;
Bursa granularis* for *Bursa ignobilis;
Mitra lens* for *Imbricaria conularis;
Pecten jacobaeus* for *Pecten pallium.*

1. **Once your search is complete – click on the link that says “[Distance tree of results]” – it should be near the top.**
2. **Record your results. *Explain your results.***

## Inquiry-Based Activity–Comparing Sequences of Your Choice

In this exercise, you will find the DNA sequences you want to compare.

To find DNA sequences, go to the NCBI’s nucleotide resource page at <http://www.ncbi.nlm.nih.gov/nuccore>. NCBI website images are current as of March 2012.

In the search box at the top of the page, enter the species you are interested in, and “coi” (for cytochrome oxidase I gene). Your search will be more effective if you use the scientific species name. For example, instead of “dog,” use “canis lupus familiaris.” You may need to Google the common name of an organism to get the scientific name!

**Note – do NOT use a dog in this example – because there are so many breeds and its well-studied, you won’t have any other species that appear.**

**Try to pick out a bird species that you know or a fish or marine mammal. These are suggestions – if you cannot make a “tree” that shows different species and their relationships, you will have to try again!**

Once you know which sequence you want to use, click on the FASTA file link, then select all the text, copy it, and paste it into a new document as a plain text file. You can approach this two ways: (either one at a time or all at once)

1. Pick ONE organism – then **record organisms that you think might be related to the one you selected. Predict at least 3 and record in your lab notebook. Use only ONE FASTA sequence for your original organism and BLAST that one only. Uncheck the “Align two or more sequences” box!**

**OR**

1. Pick at least 3 organisms that you believe are related to one another – **record these in your lab notebook and find the FASTA sequences for all of them. Then follow the instructions for testing multiple relationships.**

Once you have the distance tree this time, you may have to “hover” on the tree “branch” and select “Show Subtree” – if it shows you the EXACT same species (keep in mind that there are MULTIPLE uploads of these genes to this website!) then select “Reset” at the top and try opening another subtree.

Record your results – drawing the tree with the common name of the species shown in relationship is a ***very good idea*!** You may need to Google the scientific name (remember it’s only two words) to determine the common name of your organism. Find me if you need assistance when moving around in the tree site!