



## COMPARING DNA SEQUENCES TO DETERMINE EVOLUTIONARY RELATIONSHIPS AMONG MOLLUSKS

### OVERVIEW

In the online activity “Biodiversity and Evolutionary Trees: An Activity on Biological Classification,” you generated a phylogenetic tree of mollusks using only shell morphology. In this exercise, you will revisit that classification and reconstruct the phylogenetic tree using ClustalX, software that aligns and compares DNA sequences. You will use a simple viewer program called NJplot to view the tree.



### SOFTWARE AND FILES

Install ClustalX, which is available at <http://www.clustal.org/clustal2/>. (For Windows, download clustalx-2.1-win.msi; for Mac OS, download clustalx-2.1-macosx.dmg.) Next, install NJplot, which is available at <http://pbil.univ-lyon1.fr/software/njplot.html>. Then download the zip file containing all the DNA sequence files used in this exercise, which is found under the “Biodiversity and Evolutionary Trees” section at <http://www.biointeractive.org/activities/index.html>.

### 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
GAACTAGGTCAGCCCGGTACTTTACTAGGTGACGATCAAATTTATAATGTCATYGTAAACCGCCCATGCTT...
```

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

```
demo.txt - Notepad
File Edit Format View Help
>sequence_1 other information about the sequence goes here (optional)
TCACTGTATACTTGCAGTCGA
>sequence_2 other information about the sequence goes here (optional)
TACTGTATACTTGGCGTACG
```

**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 species, which is used to identify the species it belongs to. For animals and many other eukaryotes, the mitochondrial cytochrome oxidase subunit I (*COI*) gene, which encodes part of an enzyme that is important 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, the *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.



### EXERCISE 1: AN OVERVIEW OF CLUSTALX

Let's use ClustalX to compare DNA sequences. For this exercise, use the test sequence file test.txt, which contains the three short DNA sequences (test1, test2, and test3) shown below:

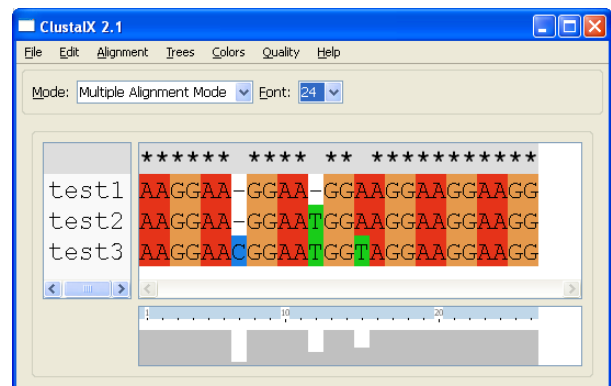
```
>test1
AAGGAAGGAAGGAAGGAAGGAAGG
>test2
AAGGAAGGAATGGAAGGAAGGAAGG
>test3
AAGGAACGGAATGGTAGGAAGGAAGG
```

**Load these sequences into ClustalX by choosing from the menu, File -> Load Sequences, and then selecting test.txt.** ClustalX displays these sequences as shown in the illustration on the right (PC version shown).



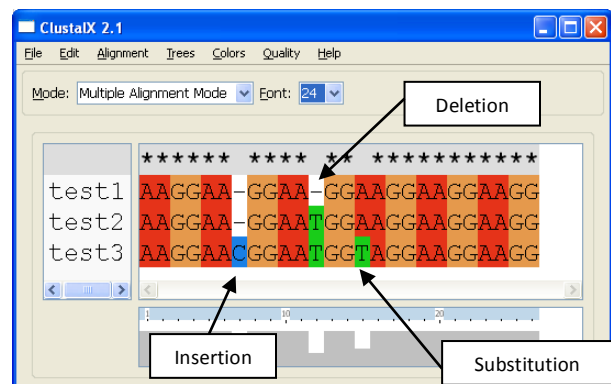
Before you can compare sequences, you have to "align" them, which means lining up the sequences and sliding them past one another until the best matching pattern is found. Alignment allows you to examine differences between related sequences; such differences reflect evolutionary relationships.

**From the menu, choose Alignment -> Do Complete Alignment. When prompted for output file names, use the default names given and click "OK."** The screen changes to look like the second illustration on the right.



Notice that it's a lot easier to see differences among DNA sequences after alignment. You can figure out what kinds of mutations have occurred in each sequence by how it compares to the others (as shown in the third illustration). The number of differences among sequences determines how closely or distantly related the corresponding organisms are.

Based on this information alone, which two sequences do you think are more closely related? To see if your answer was accurate, we can use ClustalX to generate a phylogenetic tree.

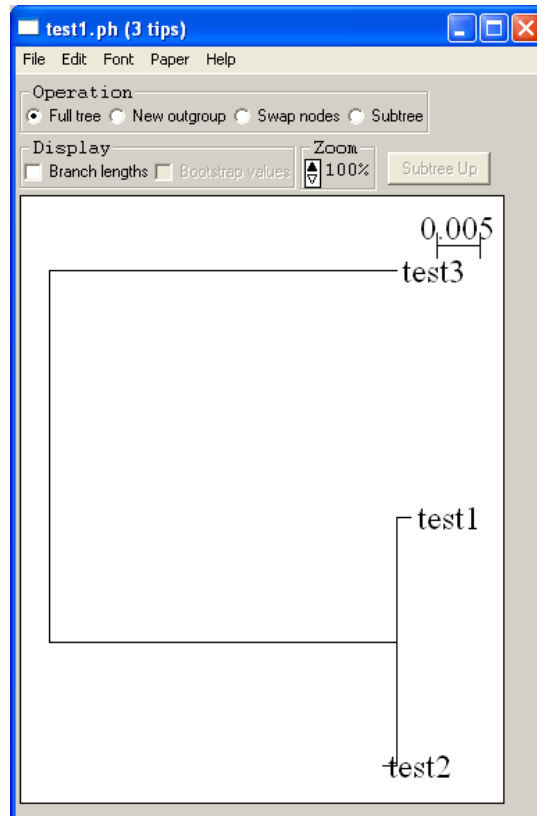


**From the menu, choose Trees -> Draw Tree.** This creates a phylogenetic tree file called test1.ph, which can be opened using NJplot.exe. **Launch NJplot, then from its menu, choose File -> Open, and select test1.ph.**

#### Comparing DNA Sequences Among Mollusks

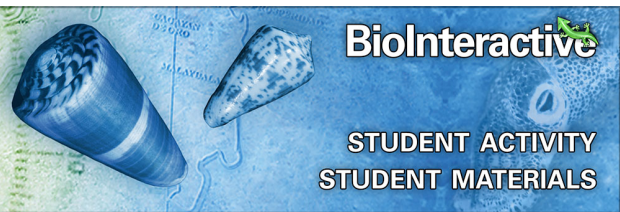


The result shows that test1 and test2 are on the same branch of the tree, indicating that they are more closely related to each other than to test3.



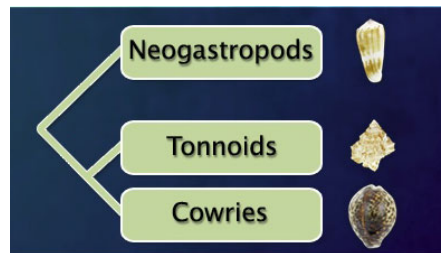


2009 HOLIDAY LECTURES  
**EXPLORING BIODIVERSITY**  
 THE SEARCH FOR NEW MEDICINES



**EXERCISE 2: 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.



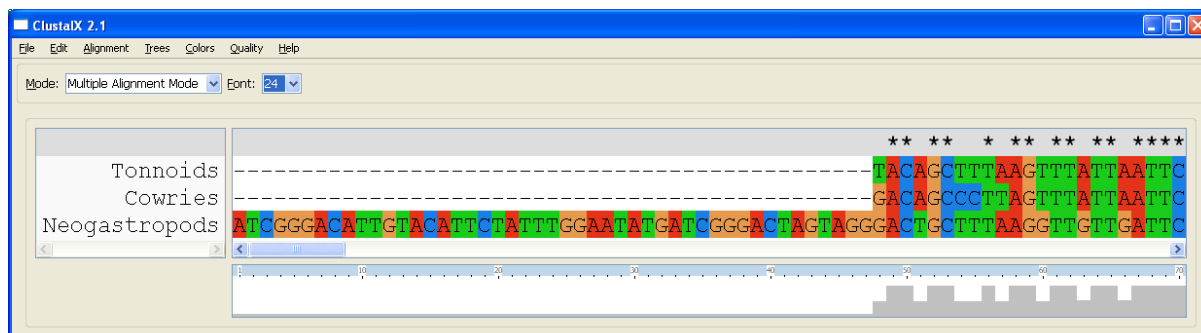
Let's confirm that conclusion using molecular data.

**Load the file molluscs1.txt into ClustalX.**

(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*.)



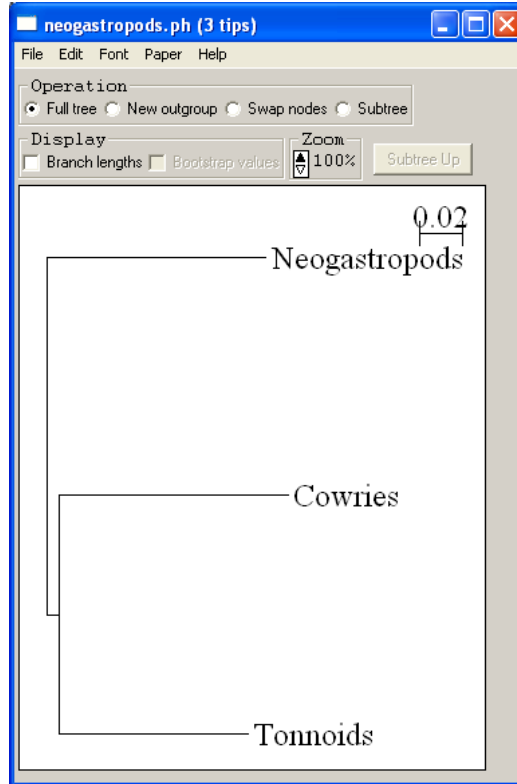
**Align the sequences by choosing from the menu, Alignment -> Do Complete Alignment.**



**Then choose from the menu, Trees -> Draw Tree.**



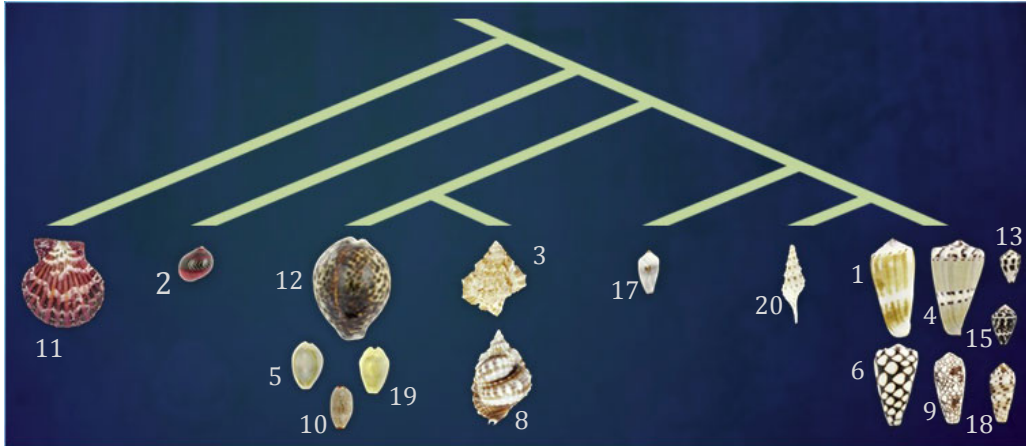
The result is shown below; it confirms that the tonnoids and the cowries are more closely related to each other than to the neogastropods.





**EXERCISE 3: PHYLOGENY OF MOLLUSKS**

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

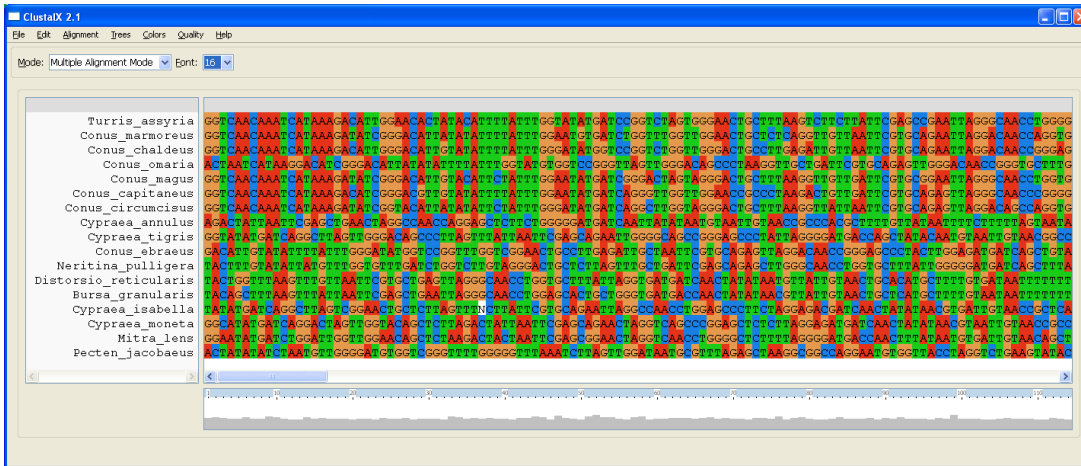


- |                             |                             |                                 |
|-----------------------------|-----------------------------|---------------------------------|
| 1. <i>Conus magus</i>       | 8. <i>Distorsio anus</i>    | 15. <i>Conus chaldeus</i>       |
| 2. <i>Neritina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

(Blank numbers are due to multiple specimens from the same species in the online seashell-sorting activity.)

We will compare the phylogenetic tree above, obtained using morphological information, with a tree obtained using DNA sequence data.

**The file molluscs2.txt contains COI DNA sequences from all of the shells in this tree\*. Load that file into ClustalX.**



Comparing DNA Sequences Among Mollusks

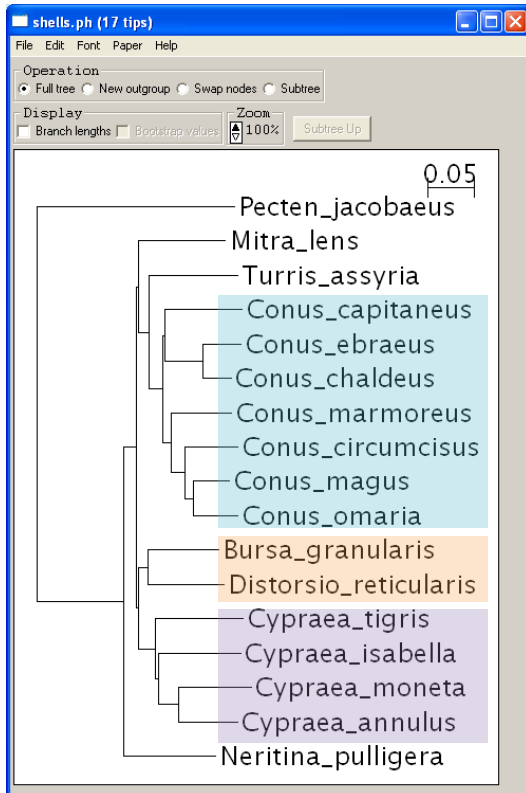
2009 HOLIDAY LECTURES  
**EXPLORING BIODIVERSITY**  
 THE SEARCH FOR NEW MEDICINES



**BioInteractive**

STUDENT ACTIVITY  
 STUDENT MATERIALS

**Align the sequences and make a phylogenetic tree. Examine the resulting tree in NJplot and compare it to the one above.** Confirm that the patterns of the various groupings are the same.



Even though the order of species in the tree on the left is not the same as in the tree above, the two trees are the same topologically. That means if you imagine the tree on the left is like a mobile hanging from a ceiling, you could rotate each branch point in the tree to make it the same as the tree above. However, in the ClustalX tree, individual cone snails and cowries have their own branches.

Cone Snails

Tonnoids

Cowries

\* 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 nobilis*;

*Mitra lens* for *Imbricaria conularis*;

*Pecten jacobaeus* for *Pecten pallium*.



2009 HOLIDAY LECTURES  
**EXPLORING BIODIVERSITY**  
THE SEARCH FOR NEW MEDICINES



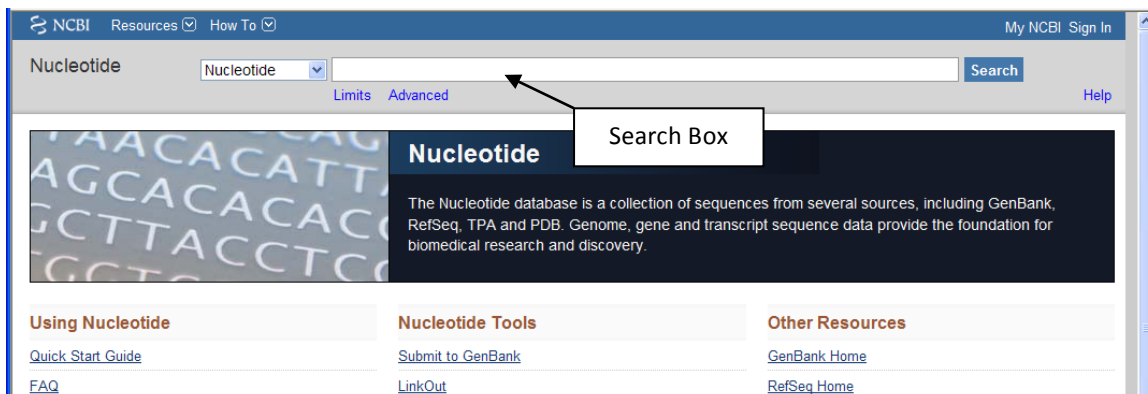
**BioInteractive**

STUDENT ACTIVITY  
STUDENT MATERIALS

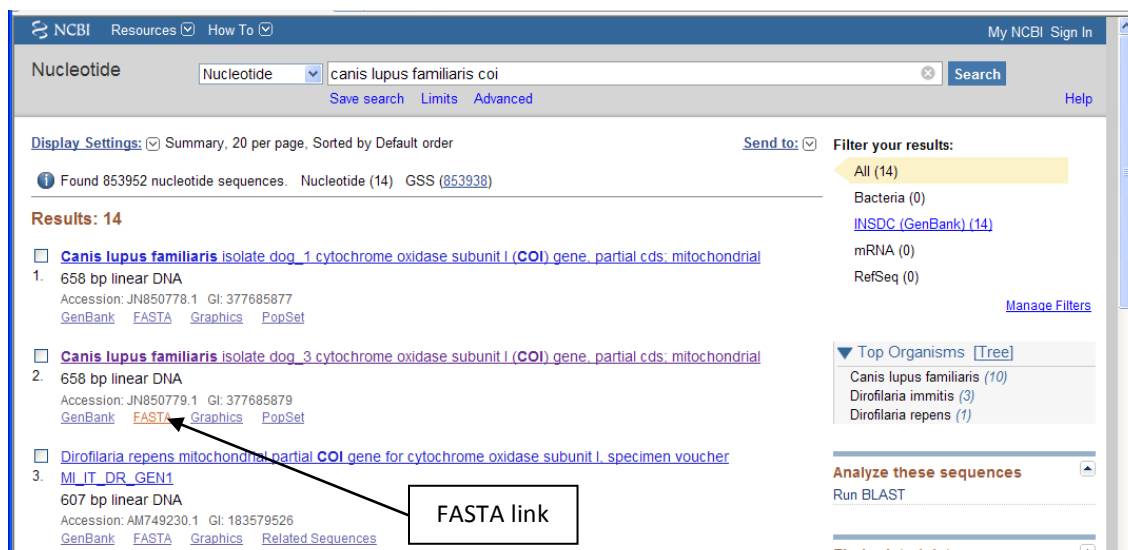
**EXERCISE 4: INQUIRY-BASED ACTIVITY—COMPARING SEQUENCES OF YOUR CHOICE**

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

To find DNA sequences, go to the National Center for Biotechnology Information's (NCBI's) nucleotide resource page at <http://www.ncbi.nlm.nih.gov/nucleotide>. NCBI website images are current as of March 2013.



**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,” enter “*canis lupus familiaris*.”**



Look for something about 700 bases long listed as “cytochrome oxidase subunit I (COI) gene,” and “partial cds; mitochondrial.” In the screenshot above, both search result 1 and 2 are DNA sequences from dog and would be good choices for this activity. But you have to be careful in selecting which file to use, as sequences from different species or sequences of different genes may be present. For example, the third result is for *Dirofilaria*, a common parasite of dogs.

**Comparing DNA Sequences Among Mollusks**



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.

NCBI Resources How To My NCBI Sign In

Nucleotide Nucleotide Search

Limits Advanced Help

Display Settings: FASTA Send:

**Canis lupus familiaris isolate dog\_3 cytochrome oxidase subunit I (COI) gene, partial cds; mitochondrial**

GenBank: JN850779.1

Bank Graphics PopSet

```
>gi|377685879|gb|JN850779.1| Canis lupus familiaris isolate dog_3 cytochrome oxidase subunit I (COI) gene, partial cds; mitochondrial
TACTTTATACTTACTATTGGAGCATGAGCCGGTATAGTAGGCCTGCTTGAGCCCTCCTCATCCGAGCC
GAACCTAGGTGAGCCGGTACTTTACTAGGTGACGATCAAATTTATAATGTCATYGTAAACCGCCCATGCTT
TCGTAATAATCTCTTCATAGTCATGCCCATCATAATTTGGGGCTTTGGAACTGACTAGTGCCTTAAAT
AATTGGTGCTCCGGACATGGCATTCCCCCGAATAAATAACATGAGCTTCTGACTCCTTCTCCATCCTTT
CTTCTACTATTAGCATCTTCTATGGTARAAGCAGGTGACGAAACGGGATGAACCATATACCCCCACTGG
CTGGCAATCTGGCCCATGCAGGAGCATCCGTTGACCTTACAATTTCTCCTTACACTTAGCCGGAGTCTC
TTCTATTTAGGGCAATTAATTTCACTACTATTACAACATAAAACCCGCTGCAATATCCAGTAT
CAAACCTCCCTGTTTGTATGATCAGTACTAATTACAGCAGTCTACTCTTACTATCCCTGCCTGTACTGG
CTGCTGGAATTACAATACTTTTAAACAGACCGGAATCTTAATACAACTTTTTTGTATCCCGCTGGAGGAGG
AGACCCATCTATATCAACACCTATT
```

Change region shown

Customize view

Analyze this sequence

- Run BLAST
- Pick Primers
- Highlight Sequence Features
- Find in this Sequence

Related information

- PopSet
- Protein
- PubMed

Remember that in FASTA format, the first line contains the information about the sequence and starts with ">" and ends at the end of the line.

```
>gi|377685879|gb|JN850779.1| Canis lupus familiaris isolate dog_3 cytochrome oxidase subunit I (COI) gene, partial cds; mitochondrial
```

ClustalX uses the letters after ">" until the first space as the label for the file. In this case, the label is "gi|377685879|gb|JN850779.1|" The label will be used onscreen and in the tree diagram, so it would help if you edited it to something simpler for this exercise. For example, you could insert "Dog" after the ">" so that it reads:

```
>Dog gi|377685879|gb|JN850779.1| Canis lupus familiaris isolate dog_3 cytochrome oxidase subunit I (COI) gene, partial cds; mitochondrial
```

Once you have the DNA sequence information from all the organisms you want to compare, assemble them all into a single text file. Be sure to save your file as a plain text (.txt) file. If you want to see what that looks like, you can look at molluscs2.txt as a model. Once you have all the sequences loaded, the procedure for aligning them and drawing a tree and looking at it is the same as before.

Be creative and explore obscure species that you only read about in books or websites! You may also want to repeat this exercise using different genes. For phylogenetic comparisons to work, the gene should be conserved among different species you are interested in, similar enough to be compared, yet different enough to give enough variability. You could try genes for actin, myosin, or ubiquitin. If you restrict the range of species, even the gene for hemoglobin may work.