



## INTRODUCTION

How do we use genetic data to determine the evolutionary relatedness of species? In this activity, you will explore this question using marine **mollusks**: invertebrate animals (such as snails) with soft, unsegmented bodies. In Part 1, you'll see how DNA can be used to study the evolutionary relationships among these mollusks. In Part 2, you'll examine differences among DNA sequences using online software. In Parts 3 and 4, you'll use DNA sequences to build diagrams of evolutionary relationships called phylogenetic trees.

By doing this activity, you'll explore how DNA sequences can be used to determine evolutionary relationships, build phylogenetic trees using online software, and compare hypotheses about species' relatedness. You will also practice skills — such as displaying and interpreting data in different formats, supporting claims with evidence, and making sense of detailed texts — that are essential to your success in science and life beyond school.

## MATERIALS

- access to the online software [MAFFT](#)
- DNA sequence files (test.txt, molluscs1.txt, molluscs2.txt), either from your instructor or from downloading [this ZIP file](#)

## PART 1: Comparing DNA Sequences to Determine Evolutionary Relationships

In the past, scientists determined evolutionary relationships among organisms based on morphological characteristics. After the development of molecular sequencing technology, it became more common to determine evolutionary relationships by comparing amino acid sequences, and later, DNA sequences.

1. What is a “morphological characteristic”? Give an example of a morphological characteristic in mollusks.
2. What might be some advantages of using DNA sequences, rather than morphological characteristics, to determine evolutionary relationships?

What types of DNA sequences are compared to determine evolutionary relationships? The sequences must be similar enough across different species so that you can compare them to each other. But they also need to vary enough so that you can use the patterns of differences to tell which species are more similar than others.

One such type of DNA sequence in animal species is the mitochondrial **COI gene sequence**. *COI*, which stands for cytochrome c oxidase subunit I, encodes part of an enzyme that is important for cellular respiration. *COI* has been sequenced for many animals, and its sequences are often unique to a particular species. This means that *COI* sequences usually vary among organisms from different species, but do not vary much among organisms from the same species.

3. Why would a *COI* sequence *not* be a good choice for studying variations within the *same* species?

4. *COI* sequences in plants have a much lower mutation rate than those in animals. Based on this information, why might *COI* sequences be less useful for determining evolutionary relationships among plant species than among animal species?

There are different formats for representing *COI* and other DNA sequences. Figure 1 shows a partial *COI* sequence in **FASTA format**. FASTA format starts with a “>,” followed by information about the sequence, a return, then the DNA sequence itself.

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

**Figure 1.** A partial sequence of a *COI* gene in FASTA format.

5. What type of animal did the *COI* sequence in Figure 1 come from? (*Hint:* Look at the information before the DNA sequence.)

## PART 2: Generating DNA Sequence Alignments and Phylogenetic Trees

In order to compare DNA sequences, it’s necessary to line up the sequences and slide them past one another until the best matching pattern is found. This technique, called **alignment**, allows you to see differences between sequences more easily. The differences between sequences can indicate how closely or distantly related the corresponding organisms are.

6. Would you expect organisms that are *more* closely related to have more or fewer differences between sequences? Explain your answer.

You can compare DNA sequences using a specialized software designed to rapidly align multiple sequences. In this activity, you will now use an online software called MAFFT to align three partial DNA sequences. You can either copy the sequences from Figure 2 below or use the “**test.txt**” file described in the “Materials” section above.

- Open [MAFFT](#) in your browser.
  - Under Step 1, choose “DNA” from the first menu, then enter your sequences. You can cut-and-paste them into the box or upload a text file. Make sure to use the FASTA format, or you may get errors.
  - Under Step 2, choose “Pearson/FASTA” as the output.
  - Submit the job and wait for the results. To make the differences among sequences easier to see, you may want to go to the “Results Viewer” tab, click on “View in MView,” and submit your results.
7. Based on your MAFFT results, label the differences in sequences in Figure 2 as deletions, insertions, or substitutions.

```
>test1
AAGGAAGGAAGGAAGGAAGG

>test2
AAGGAAGGAATGGAAGGAAGG

>test3
AAGGAACGGAATGGTAGGAAGG
```

**Figure 2.** Sample DNA sequences.

8. Which two sequences do you think are from the *most* closely related organisms, and why?

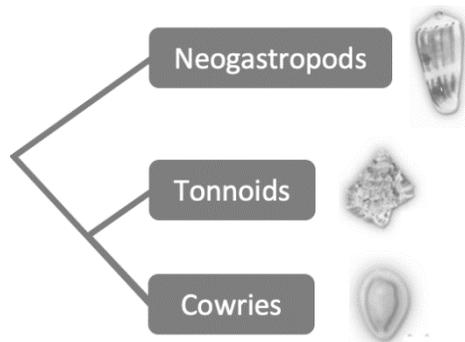
A **phylogenetic tree** is a diagram that can show which organisms are more closely related. We will use a program called Simple Phylogeny to generate a phylogenetic tree based on these DNA sequences.

- Return to the window with your MAFFT results, and open the “Results Viewer” tab.
  - Scroll down and click on “Send to Simple Phylogeny.”
  - The results from MAFFT should already be loaded in the box of Step 1. In Step 2, go to “Clustering Method” and choose “UPGMA.” (UPGMA is a method that uses DNA sequences to generate the type of phylogenetic tree we are interested in.)
  - Submit the job and wait for your phylogenetic tree to appear.
9. Sketch or provide a copy of the phylogenetic tree that was generated.

10. Based on this tree, would you revise your answer to Question 8? Why or why not?

### PART 3: Phylogenetic Trees of Neogastropods, Tonnoids, and Cowries

We can use DNA sequences to support or contradict hypotheses about evolutionary relationships based on morphological characteristics. Figure 3 shows a phylogenetic tree with three groups of snails: neogastropods (which includes cone snails), tonnoids, and cowries. This phylogenetic tree was based on morphological characteristics, such as shell shape.



**Figure 3.** Proposed evolutionary relationships of neogastropods, tonnoids, and cowries based on morphological characteristics.

11. According to the phylogenetic tree in Figure 3, which two groups of snails are most closely related?

12. Based on Figure 3, how might the *COI* gene sequences of these three groups vary? What would you expect to see in terms of similarities and differences?

Let's see if the tree in Figure 3 is supported by DNA evidence. For this, you will use the "molluscs1.txt" file described in the "Materials" section above, which contains *COI* sequences from representative species in each group of snails (neogastropods: *Conus magus*, tonnoids: *Bursa granularis*, cowries: *Cypraea tigris*).

Using the same procedures as in Part 2, run a sequence alignment and create a phylogenetic tree for the "molluscs1.txt" data. Be sure to use **UPGMA** as your clustering method when running Simple Phylogeny.

13. Sketch or provide a copy of the phylogenetic tree that was generated after your alignment.
  
14. Do your results support or contradict the evolutionary relationships proposed in Figure 3? Support your answer using evidence from the phylogenetic tree you generated.

**PART 4: Phylogenetic Trees of More Mollusks**

You will now use DNA sequences to build a phylogenetic tree for the 17 different mollusk species listed in Table 1. For this, you will use the "molluscs2.txt" file described in the "Materials" section above, which contains *COI* sequences from each species.

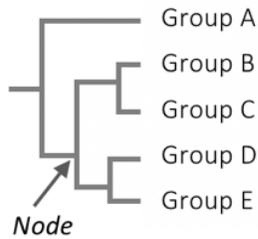
**Table 1.** Seventeen mollusk species and the groups to which they belong.

Mollusk Name	Group
<i>Bursa granularis</i>	Tonnoids
<i>Conus capitaneus</i>	Cone snails
<i>Conus chaldeus</i>	Cone snails
<i>Conus circumcissus</i>	Cone snails
<i>Conus ebraeus</i>	Cone snails
<i>Conus magus</i>	Cone snails
<i>Conus marmoreus</i>	Cone snails
<i>Conus omaria</i>	Cone snails
<i>Cypraea annulus</i>	Cowries
<i>Cypraea isabella</i>	Cowries
<i>Cypraea moneta</i>	Cowries
<i>Cypraea tigris</i>	Cowries
<i>Distorsio reticularis</i>	Tonnoids
<i>Mitra lens</i>	Miters
<i>Neritina pulligera</i>	Nerites
<i>Pecten jacobaeus</i>	Scallops
<i>Turris assyria</i>	Turrids

Using the same procedures as in Part 2, run a sequence alignment and create a phylogenetic tree for the "molluscs2.txt" data. Be sure to use **UPGMA** as your clustering method when running Simple Phylogeny.

15. Sketch or provide a copy of the phylogenetic tree that was generated after your alignment. Label or color-code the species according to the groups shown in Table 1.

Remember that a phylogenetic tree represents a hypothesis about how species or groups of species evolved from a common ancestor. Each **node**, or branch point, represents a common ancestor (Figure 4). Two species, or groups, are *more* closely related if they have a *more* recent common ancestor. They are *less* closely related if they have a *less* recent common ancestor.

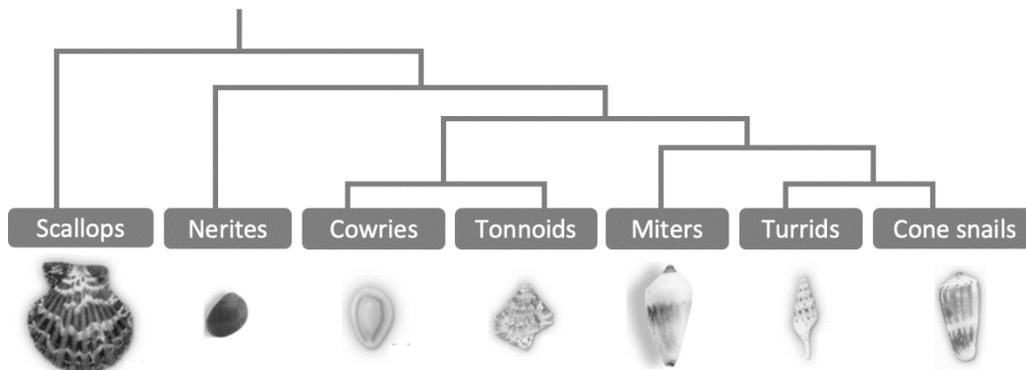


**Figure 4.** An example of a phylogenetic tree with one node labeled.

16. According to the phylogenetic tree you generated, which mollusks are *most* closely related to miters?

17. According to the phylogenetic tree you generated, are turrids more closely related to miters or to cone snails?

Figure 5 shows a phylogenetic tree for the same 17 mollusks that was based on morphological characteristics rather than DNA sequences.



**Figure 5.** Proposed evolutionary relationships of the 17 mollusks, based on morphological characteristics.

18. Do the DNA sequences you examined support all the evolutionary relationships proposed in Figure 5? Why or why not?

19. Imagine showing what you’ve been working on to a friend or family member who is not familiar with the science. Explain what a phylogenetic tree is, the information that is used to make it, and its purpose, in a way that your friend or family member could understand.

**EXTENSION: Build Your Own Phylogenetic Tree**

You will now choose your own species to build a new phylogenetic tree based on *COI* sequences. You can choose any species that you are interested in or that you'd like to learn more about. Follow the steps below to get the data you'll need:

- Open the [Nucleotide database](#) from the National Center for Biotechnology Information (NCBI).
- In the search box at the top of the page, enter the species you are interested in and “*COI*” (for the *COI* gene). Be careful in selecting which files to use, as there may be sequences from different species or different genes.
  - Your search may be more effective if you use the scientific name for your species. For example, instead of “dog,” use “*canis lupus familiaris*.”
  - Look for a sequence about 700 bases long, listed as “cytochrome oxidase subunit I (*COI*) gene” and “partial cds; mitochondrial.”
- Once you select a sequence, click on the FASTA link, then copy-and-paste the text into a **plain text (.txt) file**.
  - Remember that in FASTA format, the first line starts with “>” and contains information about the sequence, similar to the following example:

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

- The label at the beginning (in this case, “gi|377685879|gb|JN850779.1|”) will be used in the phylogenetic tree. You may want to edit it to make it easier to read. 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
```
- Collect *COI* sequences for **three or more** species you want to compare, and put all the sequences into the same .txt file. If you want to see what that looks like, you can look at “molluscs2.txt” as a model.

Once you have finished selecting your species and sequences, answer the following questions.

1. Sketch or describe the evolutionary relationships you expect for your species.
2. Using the same procedures as in Part 2, run a sequence alignment and create a phylogenetic tree for your species. Sketch or provide a copy of the phylogenetic tree that was generated after your alignment.
3. How does the tree you generated compare with the relationships you expected?