

# Comparing DNA Sequences to Understand Evolutionary Relationships With Blast

How can bioinformatics be used as a tool to determine evolutionary relationships and to better understand genetic diseases?

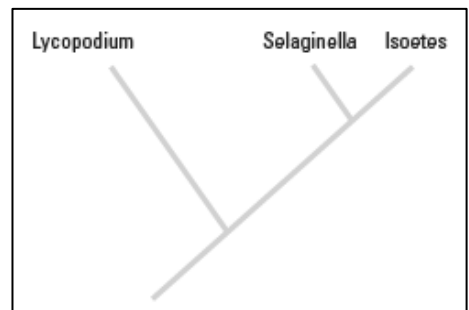
## ■ ■ BACKGROUND

Between 1990–2003, scientists working on an international research project known as the Human Genome Project were able to identify and map the 20,000–25,000 genes that define a human being. The project also successfully mapped the genomes of other species, including the fruit fly, mouse, and *Escherichia coli*. The location and complete sequence of the genes in each of these species are available for anyone in the world to access via the Internet.

Why is this information important? Being able to identify the precise location and sequence of human genes will allow us to better understand genetic diseases. In addition, learning about the sequence of genes in other species helps us understand evolutionary relationships among organisms. Many of our genes are identical or similar to those found in other species.

Suppose you identify a single gene that is responsible for a particular disease in fruit flies. Is that same gene found in humans? Does it cause a similar disease? It would take you nearly 10 years to read through the entire human genome to try to locate the same sequence of bases as that in fruit flies. This definitely isn't practical, so a sophisticated technological method is needed.

Bioinformatics is a field that combines statistics, mathematical modeling, and computer science to analyze biological data. Using bioinformatics methods, entire genomes can be quickly compared in order to detect genetic similarities and differences. An extremely powerful bioinformatics tool is BLAST, which stands for Basic Local Alignment Search Tool. Using BLAST, you can input a gene sequence of interest and search entire genomic libraries for identical or similar sequences in a matter of seconds.

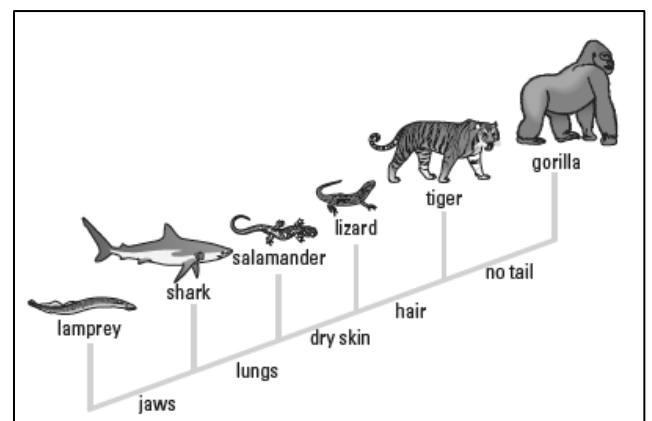


**Figure 1. Simple Cladogram Representing Different Plant Species**

In this laboratory investigation, you will use BLAST to compare several genes, and then use the information to construct a *cladogram*. A cladogram (also called a phylogenetic tree) is a visualization of the evolutionary relatedness of species. Figure 1 is a simple cladogram.

Note that the cladogram is treelike, with the endpoints of each branch representing a specific species. The closer two species are located to each other, the more recently they share a common ancestor. For example, *Selaginella* (spikemoss) and *Isoetes* (quillwort) share a more recent common ancestor than the common ancestor that is shared by all three organisms.

Figure 2 includes additional details, such as the evolution of particular physical structures called shared derived characters. Note that the placement of the derived characters corresponds to when (in a general, not a specific, sense) that character evolved; every species above the character label possesses that structure. For example, tigers and gorillas have hair, but lampreys, sharks, salamanders, and lizards do not have hair.



**Figure 2. Cladogram of Several Animal Species**

The cladogram in Figure 2 can be used to answer several questions. Which organisms have lungs? What three structures do all lizards possess? According to the cladogram, which structure- dry skin or hair- evolved first?

Historically, only physical structures were used to create cladograms; however, modern-day cladistics relies heavily on genetic evidence as well. Chimpanzees and humans share 95%+ of their DNA, which would place them closely together on a cladogram. Humans and fruit flies share approximately 60% of their DNA, which would place them farther apart on a cladogram. Can you draw a cladogram that depicts the evolutionary relationship among humans, chimpanzees, fruit flies, and mosses?

## Learning Objectives

- To create cladograms that depict evolutionary relationships
- To analyze biological data with a sophisticated bioinformatics online tool
- To use cladograms and bioinformatics tools to ask other questions of your own and to test your ability to apply concepts you know relating to genetics and evolution

## Pre-Lab Questions (Lab Notebook)

Consider the following questions to see how much you understand concepts related to cladograms before you conduct your investigation:

1. Use the following data to construct a cladogram of the major plant groups:

Organisms	Vascular Tissue	Flowers	Seeds
Mosses	0	0	0
Pine trees	1	0	1
Flowering plants	1	1	1
Ferns	1	0	0
Total	3	1	2

2. GAPDH (glyceraldehyde 3-phosphate dehydrogenase) is an enzyme that catalyzes the sixth step in glycolysis, an important reaction that produces molecules used in cellular respiration. The following data table shows the percentage similarity of this gene and the protein it expresses in humans versus other species. For example, according to the table, the GAPDH gene in chimpanzees is 99.6% identical to the gene found in humans, while the protein is identical.

Species	Gene Percentage Similarity	Protein Percentage Similarity
Chimpanzee ( <i>Pan troglodytes</i> )	99.6%	100%
Dog ( <i>Canis lupus familiaris</i> )	91.3%	95.2%
Fruit fly ( <i>Drosophila melanogaster</i> )	72.4%	76.7%
Roundworm ( <i>Caenorhabditis elegans</i> )	68.2%	74.3%

- a. Why is the percentage similarity in the gene always lower than the percentage similarity in the protein for each of the species? (Hint: Recall how a gene is expressed to produce a protein.)
- b. Draw a cladogram depicting the evolutionary relationships among all five species (including humans) according to their percentage similarity in the GAPDH gene.

## Optional Online Activities

“The Evolution of Flight in Birds”

<http://www.ucmp.berkeley.edu/education/explorations/reslab/flight/main.htm> This activity provides a real-world example of how cladograms are used to understand evolutionary relationships.

“What did T. rex taste like?”

<http://www.ucmp.berkeley.edu/education/explorations/tours/Trex/index.html>

“Journey into Phylogenetic Systematics” <http://www.ucmp.berkeley.edu/clad/clad4.html>