

## Comparing DNA Sequences Cladogram Practice

See lecture questions 75, 122-123, 127, 137

Annotating Text	
<input type="checkbox"/>	<b>UNDERLINE</b> concepts you think might be useful for understanding or solving the problem
<input type="checkbox"/>	<b>Box</b> information you think might be helpful for designing your investigation
<input type="checkbox"/>	← Write <b>notes</b> in the left margin
<input type="checkbox"/>	→ Write <b>questions and answers</b> in the right margin
Each paragraph (including each step of the procedures) must have something underlined or boxed, <b>AND</b> have something written in the margins (a question and/or note).	

### BACKGROUND

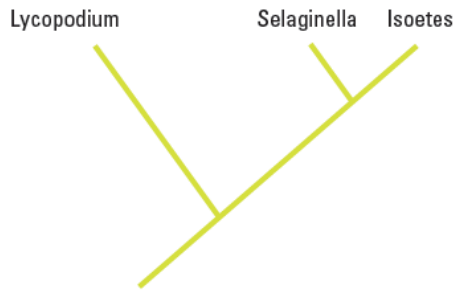
Between 1990–2003, scientists working on an international research project known as the Human Genome Project were able to identify and map the 20,000–25,000 genes that define a human being. The project also successfully mapped the genomes of other species, including the fruit fly, mouse, and *Escherichia 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? The answer to the former is most likely. The answer to the latter is maybe, but the disease profile in the fruit fly will likely provide important clues to understanding how it functions in humans. 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 is not 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.

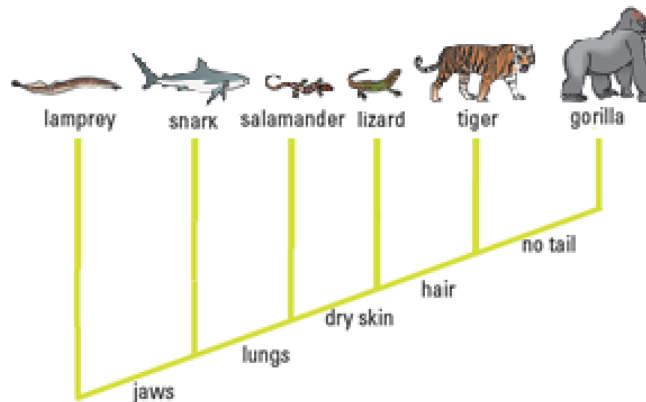
In this laboratory investigation, you will use BLAST to compare several genes, and then use the information to construct a cladogram. A cladogram (also called a phylogenetic tree) is a visualization of the evolutionary relatedness of species. Figure 1 is a simple cladogram.



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

Note that the cladogram is treelike, with the endpoints of each branch representing a specific species or larger taxonomic grouping. The higher two species share a branching point, the more recently they share a common ancestor. For example, Selaginella (spikemoss) and Isoetes (quillwort) share a more recent common ancestor than either does with Lycopodium. Importantly, Selaginella and Isoetes share the SAME common ancestor with Lycopodium and therefore are equally distantly related to Lycopodium.

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 above can be used to answer several questions:

- 1) Which organisms have lungs?
- 2) What three structures do all lizards possess?
- 3) According to the cladogram, which structure — dry skin or hair — evolved first?



- 7) Justify the cladogram above. A justification has 3 components: 1) scientific knowledge and/or theory; 2) specific data from your analysis related to the knowledge; and 3) an explanation of HOW the data supports the knowledge.

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.

**Table 2. Percentage Similarity Between the GAPDH Gene and Protein in Humans and Other Species**

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

- 8) Explain why the protein sequences are more similar than the DNA sequences. If DNA codes for the amino acid sequence of a protein, shouldn't the percent similarity of gene equal the percent similarity of protein? The answer is no, but why?

- 9) Draw a cladogram for the 4 species above and humans

10) Justify the cladogram above. A justification has 3 components: 1) scientific knowledge and/or theory; 2) specific data from your analysis related to the knowledge; and 3) an explanation of HOW the data supports the knowledge.