



P.O. Box 219 • Batavia, IL 60510
(800) 452-1261 • Fax (866) 452-1436
www.flinnsci.com • E-mail: flinn@flinnsci.com
©2013 Flinn Scientific, Inc. All Rights Reserved.

FLINN
SCIENTIFIC, INC
"Your Safer Source for Science Supplies"

Catalog No. FB2041

Publication No. 11143

Understanding Evolutionary Relationships

Big Idea 1, Investigation 3

An Advanced Inquiry Lab

Introduction

Bioinformatics is a powerful tool which can be used to determine evolutionary relationships and better understand genetic diseases. Explore the conservation of a popular enzyme, cytochrome C, and how it is present in different eukaryotic organisms.

Concepts

- Phylogeny
- Evolution
- Bioinformatics
- Cladograms
- Protein structure

Background

The Human Genome Project (HGP) was completed by scientists in 2003 and was coordinated by the U.S. Department of Energy and National Institutes of Health. The goals of the project were to:

- Identify all of the approximately 20,000–25,000 genes in human DNA.
- Store the genetic sequences in databases.
- Improve tools for data analysis.
- Transfer related technologies to the private sector.
- Address ethical, legal, and social issues arising from the identification of genetic data.

The project mapped not only the genome of humans but also of other species such as *Drosophila melanogaster* (fruit fly), mouse, and *Escherichia coli*. The locations and complete sequences of the genes in each of these species are available for anyone in the world to access on the Internet.

This information is important because the ability to identify the precise location and sequence of human genes will allow greater understanding of genetic diseases. Also, learning about the sequence of genes in other species helps us to understand evolutionary relationships among organisms. Many of our genes are similar if not identical to those found in other species.

For example, a gene in fruit flies is found to be responsible for a particular disease. Scientists might wonder if this gene is found in humans and does it cause similar disease. It would take years to read through the human genome to locate the same sequence of base pairs. Given time constraints, this is not practical—so a technological method was developed.

Bioinformatics is a study that combines statistics, mathematical modeling, and computer science to analyze biological data. Through bioinformatics, entire genomes may be quickly compared in order to detect and analyze their similarities and differences. BLAST (Basic Local Alignment Search Tool) is an extremely useful bioinformatics tool which allows users to input a gene sequence of interest and search entire genomic libraries for identical or similar sequences.

Classification of organisms based on evolutionary history is called *phylogenetic systematic*. Scientists study how different organisms are related to determine if they have common ancestry. Today most scientists practice *cladistics*. Cladistics is a taxonomic approach that classifies organisms according to the order in time at which branches arise along a phylogenetic tree without considering the degree of morphological divergence. A phylogenetic diagram based on cladistics is called a *cladogram*. It is a tree constructed from a series of two-way branch points. Each branch point represents the divergence of a common ancestor. The cladogram is treelike where the endpoint of each branch represents a specific species (see Figure 1 on the next page).

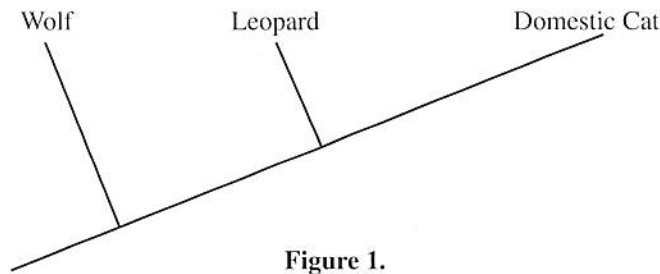


Figure 1.

The cladogram featured in Figure 2 includes additional details such as the evolution of particular physical structures called derived characters. Note that the placement of the derived characters corresponds to that character having evolved. Every species *above* the character label possess that structure. For example, lizards, tigers and gorillas all have dry skin. Whereas, salamanders, sharks and lamprey do not have dry skin.

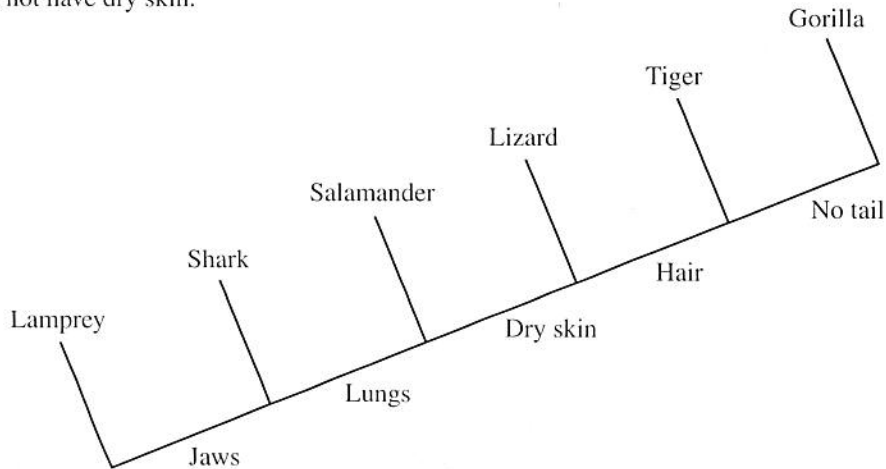


Figure 2. Cladogram of six animal species

Evolutionary changes stemming from random mutational events can alter a protein's primary structure. Some mutations do not allow the organism to survive. In order for the change to propagate, the mutation must either allow the organism to have the same evolutionary ability as it had previously or increase its probability to survive and reproduce. Sometimes a mutation can improve the fitness of a host in its natural environment. A classic Darwinian example is sickle cell anemia. This is a result of a single mutation whose adaptive consequences turned out to be beneficial to combat malaria. Normal hemoglobin cells have a high potassium concentration whereas hemoglobin sickle cells do not contain as much potassium. In order for a malaria parasite to survive it needs cells with a high potassium concentration. Thus they do not survive in sickle cells.

Experiment Overview

In this experiment, the conservation of genes among species will be explored using both bioinformatics and basic comparative techniques.

Pre-Lab Questions

		TAXA			
		Pine Trees	Mosses	Flowering Plants	Ferns
Characteristics	Vascular Tissue	1	0	1	1
	Flowers	0	0	1	0
	Seeds	1	0	1	0

Table 1. Character Table. A zero (0) indicates that a character is absent; a one (1) indicates that the character is present.

1. Given that flowering plants have a “1” for vascular tissue, flowers, and seeds, where would it be located on the cladogram?
2. Given that mosses have a “0” in all categories where would they be located on the cladogram?
3. Draw a cladogram from the information provided in the Character Table.

Materials

Computer with Internet connection

Word processing program

Safety Precautions

This activity is considered nonhazardous. Please follow all classroom safety guidelines.

Baseline Activity

Part A. Gathering cytochrome C sequences from NCBI

1. Go to the National Center for Biotechnology Information Website: <http://ncbi.nlm.nih.gov>
2. In the search box type in “cytochrome C” *Note:* Make sure that the drop-down menu to the left of the search box is set to “All databases.” Click the search button and you will be directed to the ENTREZ page.
3. Select the “Protein: sequence database”
4. You will be directed to the protein database. In the upper left region of the page locate the drop-down menu and make sure it is selected on “protein.”
5. In the search box type “cytochrome C horse.” Click the search button.
6. Many choices will appear, select the record with the accession number for “P00004.2”
7. There is a plethora of information on this page. Locate the menu in the upper left portion of the page that says “display.” Click the drop-down menu and select “FASTA.” This will list the amino acid sequence in a simple format.
8. Open an MS Word document or another word processing program.
9. Go back to the NCBI page and then copy and paste the entry into the Word document.
10. Save the document as “cytochrome C sequences”.
11. Go back to the NCBI Website and arrow back until you arrive at the protein search page. Backspace to remove the organism horse and then type the next organism. *Note:* Cytochrome C should still precede the organism in the search box.
12. Continue copying and pasting the cytochrome C sequences until you have gathered all of the organisms listed after the Part A procedure.
13. Save your Word document. Record the species name and accession numbers for each organism.

Organisms: horse, chicken, tuna, human, *Ateles* (monkey), chimpanzee, rabbit, *Arabidopsis thaliana* (plant), rattlesnake, frog, dog, bee, cow.

Part B. Comparative Genomics and Bioinformatics

Before beginning you must set up an account at Biology Workbench (it’s free). This merely serves to allocate hard disk space for the session that you will run. Go to <http://workbench.sdsc.edu> and select “set up a free account” and follow the Website instructions.

1. There is a series of buttons near the middle of the page. Select “protein tools.”
2. Select the “Add new protein sequence” option and click “run.”
3. Two boxes will appear—type the name of the sequence in the smaller of the two and paste the sequence into the larger box. To paste the sequence, open the cytochrome C Word file saved in Part A. Copy the first sequence but only the amino acid, not the identifying information. Return to the Biology Workbench site and paste into the sequence box.
4. Scroll down to the bottom of the page and click “save.”

5. The Web site will default back to the “add sequence” page. *Note:* There is now a check box for your sequence on the page.
6. Click “add sequence” again and then “run.” Add the next organism sequence.
7. Repeat step six until all organisms have been added.
8. Once all organisms are added go back to the menu box and pull down until you see “Clustal W-Multiple Sequence Alignment.” Click on “Clustal W” and press “run.”
9. Scroll down until you see the “guided tree display.” Pull down to “rooted and unrooted display” then click the submit button.
 - a. Manually compare any of the four amino acid sequences you gathered from the database and count the number of differences between the human cytochrome C sequence and the four others. Does this data correlate with the phylogenetic tree?

Opportunities for Inquiry

1. Consider the following questions while reflecting upon your knowledge of evolutionary relationships.
 - a. Research what other genes are conserved among organisms tested in this activity. Do their phylogenetic trees appear to be the same or different than the cytochrome C tree?
 - b. What other organisms other than those researched in the baseline activity have similar evolutionary patterns?
2. Plan, discuss, execute, evaluate and justify an investigation to test a question regarding evolutionary relationships using the BLAST Web site.
 - a. Determine a gene and organism to explore.
 - b. Research the sequence, function, and structure of the protein created from the chosen gene.
 - c. How will the data be analyzed to test the developed hypothesis?
 - c. Review your questions and proposed analysis with your instructor prior to beginning the experiment.
 - e. Present and defend your findings to the class.
 - f. Make suggestions for a new or revised experiment to modify or retest your hypothesis.