

Basic Bioinformatics Lab

This lab will help familiarize students with the basics of bioinformatics via the National Center for Biotechnology Information (NCBI) website. This very powerful public website states its goals as follows: “Established in 1988 as a national resource for molecular biology information, NCBI creates public databases, conducts research in computational biology, develops software tools for analyzing genome data, and disseminates biomedical information - all for the better understanding of molecular processes affecting human health and disease.”

This website has gene sequences for several organisms, including human, chimpanzees, mice, dogs, rats, chickens, mustard weeds, rice, fruit flies, nematodes, and *Plasmodium*. In this lab we will compare the amino acid sequences for certain proteins in these different organisms. Students will then form an evolutionary “tree” indicating the various relationships.

In a step-by-step process, the lab will help familiarize you with some aspects of the NCBI website. Feel free to explore on your own!

Part A. Hemoglobin Comparison

1. Open up your internet browser.
2. Go to the site: <http://www.ncbi.nlm.nih.gov>
3. In the upper left corner of the website, search “all databases” for hemoglobin. You will get a page showing a variety of articles and genetic information. Feel free to see where some of these sites take you by clicking on them. For example, “Pub Med” will list thousands of original papers related to hemoglobin, and “Nucleotide” will reveal the sequence of the mRNA for hemoglobin in a variety of organisms.
4. After you have explored for a bit, click on HomoloGene. This will give you dozens of choices where you can compare the protein sequences of the various organisms. Just below the tab for “Limits” change the “display” to FASTA. You will then see the amino acid sequence of each organism (in a one letter code – for interpretation of the code go to <http://wwwchem.csustan.edu/chem4400/code.htm>)
5. Choose one of the comparisons (with at least five organisms) and using the human as a base, count the number of amino acid differences for the other organisms.
6. Using this data, calculate the % of similarity of each organism to the human.

$$100 - (\text{Number of differences}/\text{total amino acids} \times 100) = \% \text{ similarity}$$

7. Interpret this result.

Part B. Gene of Interest

Repeat the process above with any gene or protein that is of interest to you.

Part C. How many genes?

For an up-to-date report of the number of genes in various organisms, go back to the original NCBI site in step 2, and search “HomoloGene” and leave the space after the word “for” blank. You will a recent count of identified genes for about 18 organisms.