

Chapter 20 Project

Planting Phylogenetic Trees

Project Goal + Timeline

In this project, you will explore how phylogenetic trees are made using protein sequence data. In fact, you will compare this method against your intuitive understanding of evolutionary relationships and trees based on physical appearance and physiology of the organisms. Creating these trees and establishing these relationships can be important for better understanding how different parts of an organism work. This project should take between one and two hours to complete.

Directions

Part 1: Background and Predictions

As organisms evolve, they change in many ways—for example, in appearance, physiology, and behavior. Underlying all of these are changes at the genetic level to the sequence in DNA and the corresponding sequence of amino acids in proteins encoded for by the DNA. One way in which scientists can establish evolutionary relationships and build phylogenetic trees is by examining these sequence differences among organisms.

For this project, you will need to collect protein sequences from a variety of organisms. Well-known organisms are great choices, especially those that are used frequently in experiments, are of importance to society, or are otherwise well studied. These include our own species, *Homo sapiens*, as well as the chimpanzee, cow, or chicken (see the Project Materials section for an expanded list). You can find the FASTA format protein sequences by searching for them in the NCBI Protein Database (hawkes.biz/protein), among other sites.

1. Decide what set of at least five species you will compare. Based on existing phylogenetic trees or your own intuitive understanding, predict what the evolutionary relationships are among these organisms. What is the rationale for your prediction?
2. Draw a cladogram to represent your hypothesized relationships among the organisms.

Next, you will collect sequences for the same protein from each organism you've selected. You can pick a variety of different proteins—from ones that might be specific to an organism's metabolism (such as insulin), to defense proteins in multicellular organisms (such as lysozyme), to something that might be universal to all cells (such as EF-Tu, a translation factor found in all cells).

3. Decide which protein you'd like to analyze. You may find it helpful to do a search for some proteins that are common across a particular group of animals. Give the name of your protein and describe its function.

Part 2: Sequence Search

You'll collect the amino acid sequences in the FASTA format. In FASTA format, the sequence of amino acids in a protein is represented using a one-letter abbreviation for each of the twenty possible amino acids at each position in the protein. The specific combination observed gives the protein a unique structure and chemistry, endowing it with a particular ability in the cell.

Collect the sequences by searching in the NCBI Protein database (hawkes.biz/protein). In the search bar at the top of the page, search both the name of the protein and a species name. Select a relevant result, and on the page describing the result, select "FASTA" to access the complete amino acid sequence in FASTA format.

1. What are the sequences of this protein from at least five different organisms?

Part 3: Tree Construction

Once you have the sequences collected, you can use it to create a phylogenetic tree using the Phylogeny.fr tool (hawkes.biz/phylogeny). Select the "one click" version of the tool and input all your sequences in FASTA format.

1. What are the evolutionary relationships evident from this analysis? Provide the phylogram, cladogram, or tree generated by the software.
2. Explore different types of representations. Which is most informative for this particular analysis?
3. How does this tree compare to the one that you created before the analysis? What are some possible reasons for any discrepancies?

Part 4: Further Exploration

Now you should expand on your analysis. You can do this by repeating the bioinformatics analysis but using a different target protein. You can also increase the number species you are examining.

1. When selecting another protein to carry out the comparison, does this change your answer to the questions in Part 3? Provide the new phylogenetic tree and any new conclusions.
2. What type of gene or protein do you think is most useful for making comparisons over a long evolutionary timescale?

Project Materials

- Project worksheet and a pen, or a computer with a word processor
- Access to bioinformatics websites:
 - NCBI Protein Database (hawkes.biz/protein)
 - Phylogeny.fr (hawkes.biz/phylogeny)
- Organism names and Taxonomic IDs for *Homo sapiens* (human, 9606), *Bos taurus* (cow, 9913); *Gallus gallus* (chicken, 9031); *Pan troglodytes* (chimpanzee, 9598); *Danio rerio* (zebra fish, 7955); *Mus musculus* (mouse, 10090); *Rattus norvegicus* (rat, 10116); *Xenopus laevis* (African clawed frog, 8355); *Arabidopsis thaliana* (thale cress, 3702); *Caenorhabditis elegans* (roundworm, 6239); *Drosophila melanogaster* (fruit fly, 7227). Others can be found through NCBI's taxonomy browser (hawkes.biz/taxonomy).

Student Checklist

- ☐ Complete Background and Prediction questions
- ☐ Collect and record sequences
- ☐ Create a phylogenetic tree
- ☐ Complete Tree Construction questions
- ☐ Repeat the procedure with a new protein of choice from the same set of species
- ☐ Complete Further Exploration questions