

Chapter 18 Project

Molecular Evolution

Project Goal + Timeline

The goal of this project is to relate concepts of evolution in the context of organism lifestyle and appearance to evolution at the molecular level. You'll examine molecular evolution by comparing the amino acid sequences of a protein that is shared across species. This project should take between one and two hours of active time to complete. However, you may need to complete this project over two days to allow time for the software tools you'll be using to build alignments among amino acid sequences.

Directions

Part 1: Background and Predictions

The Ukrainian American geneticist Theodosius Dobzhansky once remarked, “Nothing in biology makes sense except in the light of evolution.” This argument is reflected in this activity, during which you will apply the lessons you have learned about evolution at the species level to the molecular level. Moreover, the insights we can get by applying evolutionary thinking to molecular structures will shed light onto how these molecules work. Evolution by natural selection is the framework by which many disparate parts of biology and the life sciences can be understood.

This project can be done in two phases—first, by completing the assignment using preselected sequences, and then secondly, using sequences that you find on your own. The sequences in question will be the order of amino acids that make up a specific protein. Just as you might analyze the structure of beaks across birds or the evolution of other physiological traits in an organism, we can examine the exact structure of a single protein; all of these features can be selected for and relate to the fitness of the organism.

Without diving into too much biochemistry, it is important to understand what the sequences represent. Proteins are long chains of connected amino acids; there are 20 different types of amino acids that can be found at any given position in a protein. A particular combination of amino acids will fold into a specific shape with a unique chemistry, allowing the protein to carry out a specific activity in the cell. This activity may, in turn, influence the survival, appearance, or behavior of the organism.

A protein sequence can be represented by a string of single-letter abbreviations for each amino acid. For example, the sequence for the human lysozyme protein reveals that the first amino acids in the protein are M (methionine), K (lysine), and A (alanine). The sequence is shown in FASTA format, which is a text-based format for representing nucleotide and amino acid sequences in bioinformatics.

```
>AAC63078.1 lysozyme precursor [Homo sapiens]
MKALIVLGLALLSVTVQGKVFERCELARTLKRLLGMDGYRGISLANWMCLAK-
WESGYNTRATNYNAGDRSTDYGFQINSRYWCNDGKTPGAVNACHLSCSALLQD-
NIADAAACAKRVVRDPQGVRAWAAWRNRCQDRDVRQYVQGCGV
```

You will use a bioinformatics approach, together with evolution, to gain insights into this protein.

1. Research what lysozyme is and how it works. What selective pressures do you think caused this enzyme to evolve in humans or a progenitor species?
2. What other organisms might you expect to have an enzyme like this? Why?
3. Which organisms do you think will have lysozyme sequences most like that of humans? Why?

Part 2: Sequence Search

Now you will search for an equivalent or similar sequence from different organisms—at least four additional sequences. You can do this in a few different ways:

1. Use the NCBI Protein Database (hawkes.biz/protein). In the search bar at the top of the page, search both the name of the enzyme and a species name (if needed, species names are provided in the Project Materials section). Select a relevant result, and on the page describing the result, select “FASTA” to access the complete amino acid sequence in FASTA format.
2. Use the NCBI protein BLAST tool (hawkes.biz/ProteinBLAST). Enter the amino acid sequence for the human lysozyme protein in the box for Step 2, then submit your job. When the results are ready, select sequences from different organisms to investigate.

Make sure to copy the FASTA sequence from your result into Question 3.

3. What are the sequences of lysozyme from other organisms—such as cow (*Bos taurus*) or chicken (*Gallus gallus*)? How do they compare to the sequence from humans?

Part 3: Align and Analyze

Now, you can examine the differences in a more systematic and informative way using a combination of the tools Clustal Omega and Mview. The first tool will identify the conserved sequences (similar sequences that suggest shared ancestry) in each protein, while the second will allow you to compare these sequences. **Note:** if you have collected the other sequences by running an NCBI BLAST+, you can simply ‘launch’ the Clustal Omega and Mview tools from the main results window by selecting that option in the dropdown box on the left.

Input your sequences into Clustal Omega (hawkes.biz/Clustal) and run the program. This program will show which sequences best align. **Note:** this program may take several hours to run. Return to your project once the results have been compiled.

Once the alignments have been completed, explore the output in Clustal Omega. Then, go to the “Results Viewer” tab, and select “view in Mview” near the bottom of the page. Submit your job in Mview. The Mview output highlights the shared amino acids at each position in the conserved sequences and calculates potential consensus sequences. The consensus sequence gives the most common amino acid at each position in the conserved sequences.

1. Are there any conserved sequences in the lysozyme protein across the species you selected? What sequences are similar, and how similar do these sequences appear to be?
2. Why do you think some sequences are conserved or the same across all species? How can you explain this result using evolution by natural selection?
3. What does the information you found tell you about the protein? What might you do with this information?

Part 4: Further Exploration

Now it’s your turn to be creative! Repeat all the above steps, but with a protein other than lysozyme. Answer the same questions, but with a protein that you or a partner has picked.

Project Materials

- Project worksheet and a pen
- Computer with internet access
- Access to bioinformatics websites:
 - NCBI Protein database (hawkes.biz/protein)
 - NCBI BLAST + (hawkes.biz/ProteinBLAST)
 - Clustal Omega (hawkes.biz/Clustal)
 - Mview (hawkes.biz/MView)
- Organism names and Taxonomic IDs. Possibilities include *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 Predictions questions
- ☐ Collect and record lysozyme sequences from at least four organisms
- ☐ Analyze sequences using Clustal Omega and Mview
- ☐ Complete Align and Analyze questions
- ☐ Repeat the procedure with a new protein of choice