Activity: Using Online Molecular Databases to Study Evolutionary Relationships

AP Biology

Introduction

How does an evolutionary biologist decide how closely related two different species are? The simplest way is to compare the physical features of the species (their "morphologies"). This method is very similar to comparing two people to determine how closely related they are. We generally expect that brothers and sisters will look more similar to each other than two cousins might. If you make a family tree, you find that brothers and sisters share a common parent, but you must look harder at the tree to find which ancestor the two cousins share. Cousins do not share the same parents; rather, they share some of the same grandparents. In other words, the common ancestor of two brothers is more recent (their parents) than the common ancestor of two cousins (their grandparents), and in an evolutionary sense, this is why we say that two brothers are more closely related than two cousins.

Similarly, evolutionary biologists might compare salamanders and frogs and salamanders and fish. More physical features are shared between frogs and salamanders than between frogs and fish, and an evolutionary biologist might use this information to infer that frogs and salamanders had a more recent common ancestor than did frogs and fish.

This methodology certainly has problems. Two very similar looking people are not necessarily related, and two species that have similar features also may *not* be closely related. Comparing morphology can also be difficult if it is hard to find sufficient morphological characteristics to compare. Imagine that you were responsible for determining which two of three salamander species were most closely related. What physical features would you compare? When you ran out of physical features, is there anything else you could compare? Many biologists turn next to comparing genes and proteins. Genes and proteins are not necessarily better than morphological features except in the sense that differences in morphology can be a result of environmental conditions rather than genetics, and differences in genes are definitely genetic. Also, there are sometimes more molecules to compare than physical features.

In the following exercise, you will use data in a public protein database of gene products (proteins) to evaluate evolutionary relationships. In the first part of the exercise, you will be examining the amino acid sequence of the hemoglobin beta protein, a protein common to most vertebrate organisms. You will also use the same database to generate phylogenetic trees for groups of organisms whose protein sequences you examine. For the second part of this exercise, you will determine what evolutionary relationships might exist among a group of organisms you choose based on a protein sequence of your choice. You will also draw a phylogenetic tree for these organisms.

You will obtain your data from a public online database that contains the amino-acid sequences of proteins coded for by many genes for many different organisms. Hemoglobin, the molecule that carries oxygen in our bloodstream, is composed of four subunits. In adult hemoglobin, two of these subunits are identical and coded for by the alpha hemoglobin gene. The other two are identical and coded for by the beta-hemoglobin genes. The hemoglobin genes are worthy of study themselves, but today we will just use the protein sequences as a set of traits to compare among species.

Materials for each student or pair of students:

- Background and Overview package
- Packet of tables for data collection
- Computer with internet access

Part One: Determining Relationships Among Groups of Vertebrates Using Morphologies

In this part of the lab, you will view several examples of vertebrate specimens and determine what characteristics they have in common with one another. You will also make a note of what characteristics distinguish them from one another. To do this, follow the directions given below:

1. Your teacher has uploaded a document containing images of the sample organisms. You need to view the photographs provided and fill in the following chart in your BILL. Be sure to give yourself enough space to fill in the required information!

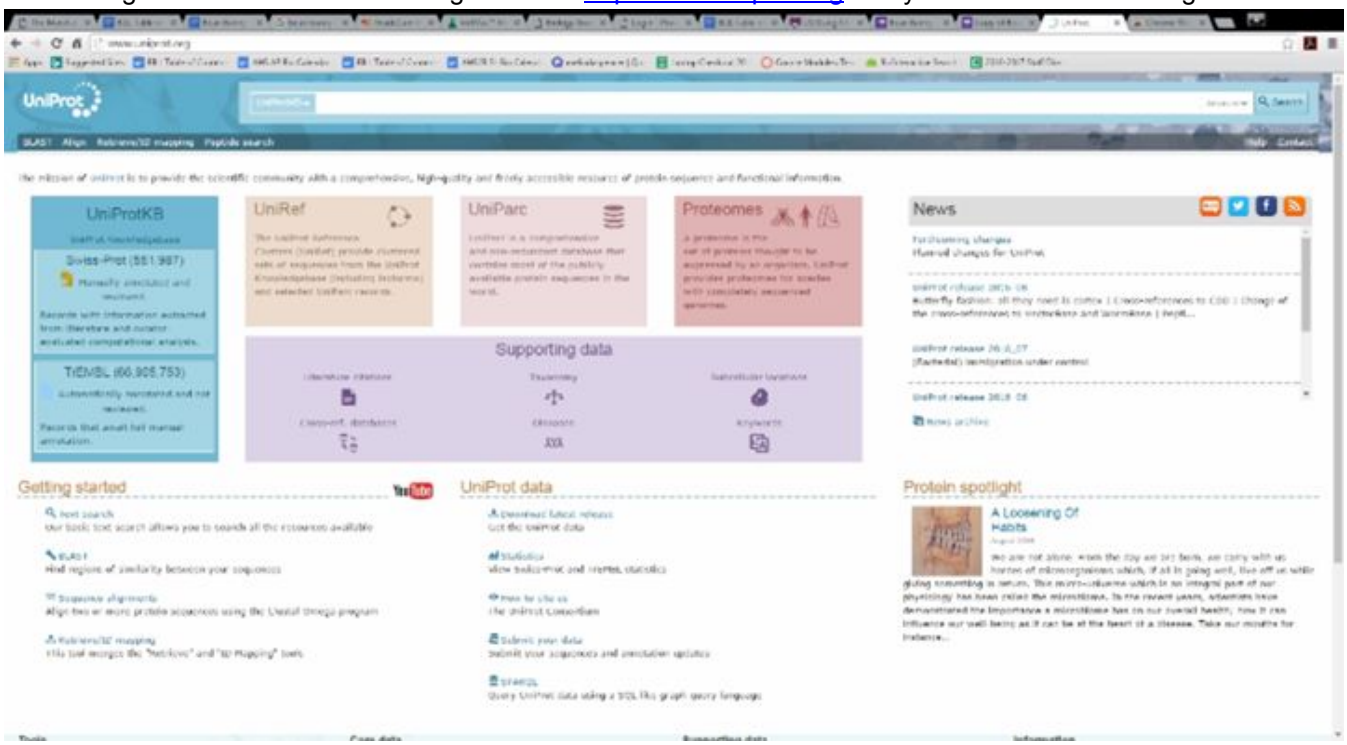
Species name	**Organism type: is it a fish, bird, etc?**	**List at least 5 distinguishing physical features of this organism. They must be viewable from the photograph.**
Homo sapiens (Human)		
Pan troglodytes (Chimpanzee)		
Bos Taurus (Cow/bull)		
Anser anser anser (Greylag goose)		
Ovis aries (Sheep)		
Canis familiaris domesticus (Dog)		
Gorilla gorilla (Gorilla)		
Caretta caretta (Loggerhead sea turtle)		
Trematomus bernacchii (Emerald rockcod)		

2. Based on the data you have gathered above, create a phylogenetic tree based solely on morphological differences. To construct your tree, you will do it in the same way we have done in class previously—by looking for shared derived characteristics the organisms may have based on your observations above. You are trying to answer the research question: **What evolutionary relationships exist among a group of vertebrates?**

3. **State a claim** about the evolutionary relationship these organisms share with one another based on your observations and initial phylogenetic tree construction. You will gather evidence in the second part of this activity that you will use to either provide support for your claim or refute your claim.

Part Two: Determining Relationships Among Groups of Vertebrates Using Molecular Data for a Known Protein Sequence

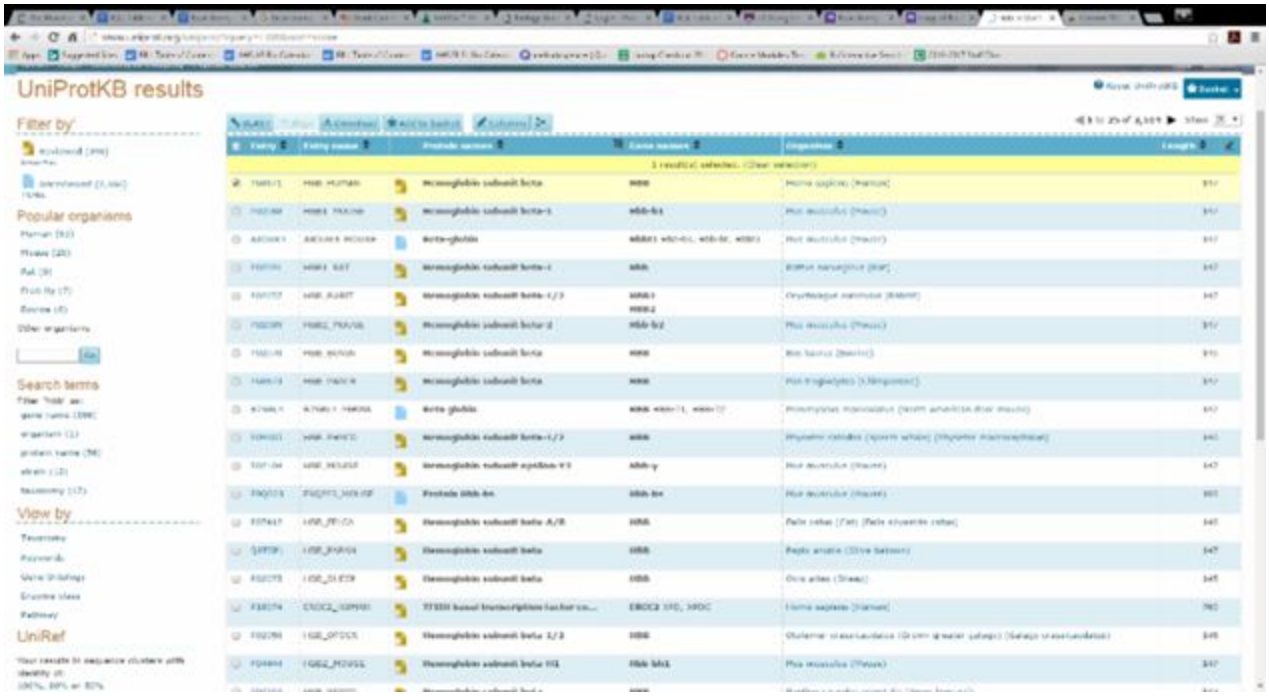
4. Now you will use a publicly available database known as Uniprot to find molecular information about each of these organisms. Go to the following website: http://www.uniprot.org and you will see the following screen:



For the first activity you will do, you will search in the database for information about the hemoglobin beta protein. Hemoglobin beta (HBB) is a subunit of the hemoglobin protein, an oxygen carrying pigment with four subunits found on red blood cells. In the box at the top of the screen, type in the search term: **HBB** and click "Search." This will search for all information in the database about the hemoglobin beta protein. It is important that you type the term in as all capital letters, as a search for hbb finds information about an entirely different protein.

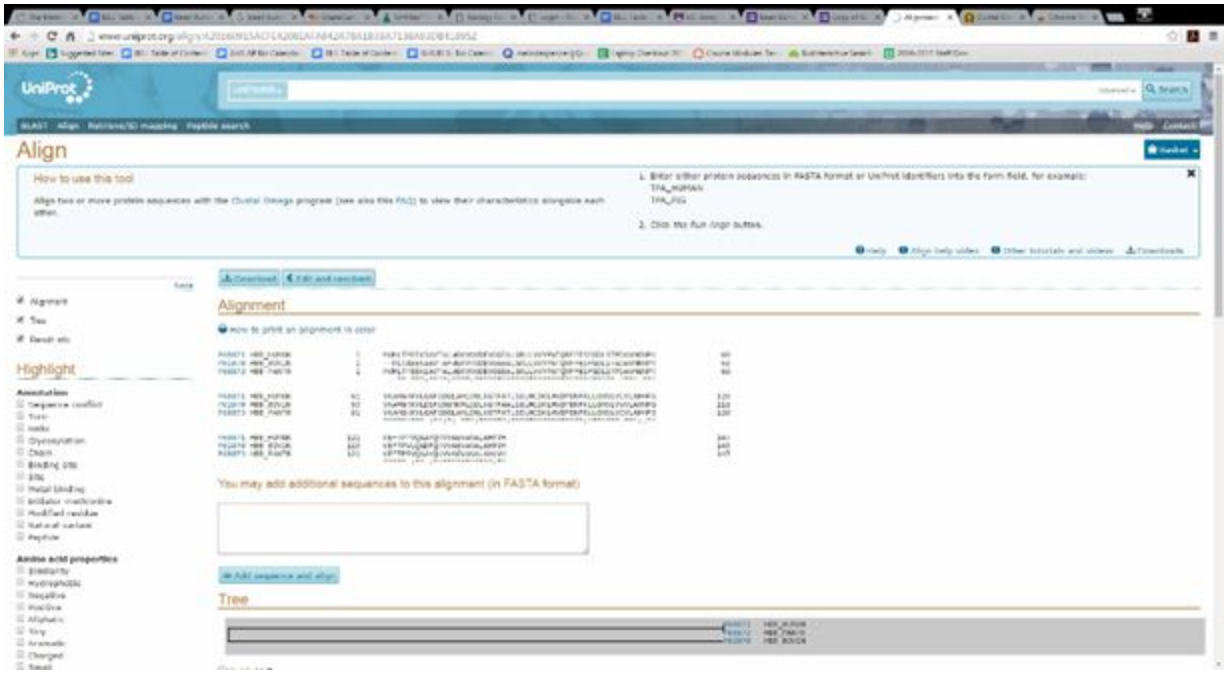
5. Because you are trying to locate specific information about the species listed in your table, you can also search for HBB *species name*. You may find this proves to be easier when searching. Once you have located the species you are searching for, check the box out to the left side of its name so that its amino acid sequence is chosen for alignment.

6. You will notice that when you check the box next to the species names, the screen will look like this. When you select a species, it will be highlighted in yellow and you will have the option to “add to basket.”

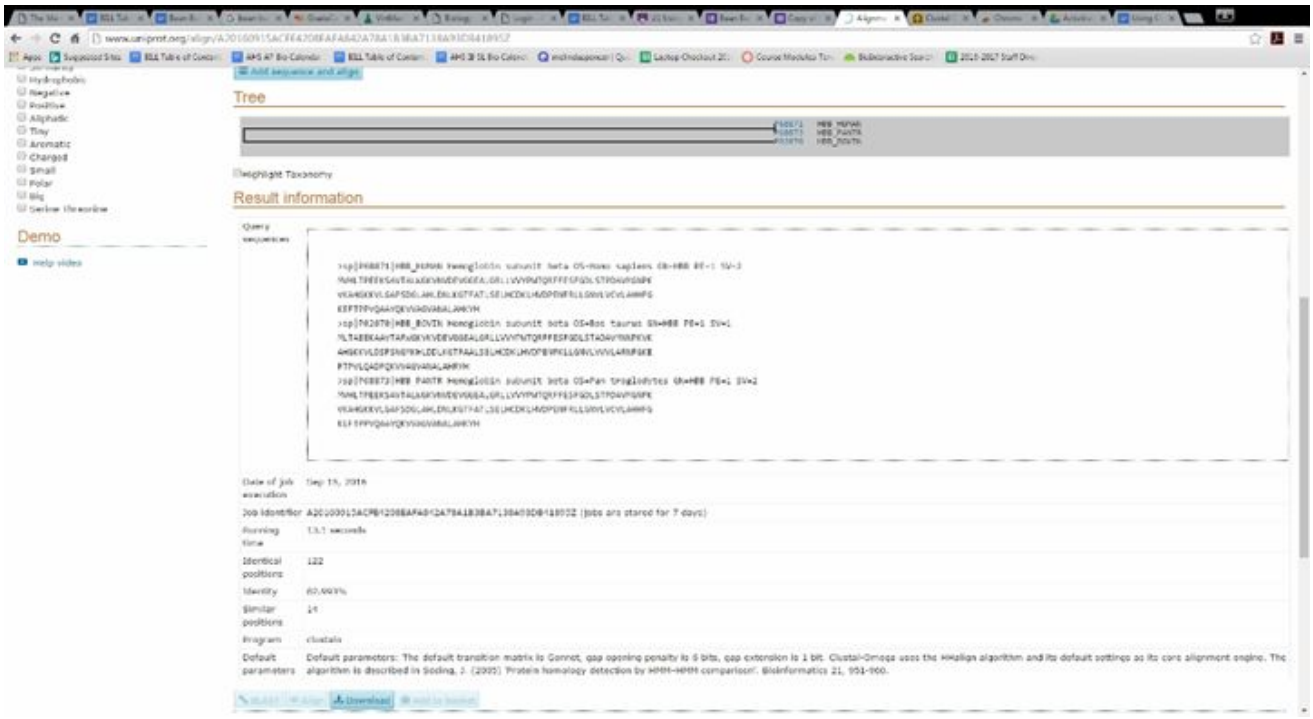


This bar will contain the records you have selected as well as a few buttons that you will use during this activity: **add to basket**, which will allow you to then align sequences for further data analysis, **align**, which will allow you to align multiple amino acid sequences and **BLAST**, which is used to align nucleic acid sequences; and download, which will allow you to download the data that are generated.

7. Once you have selected all 9 sequences, click **align**. You will get a screen that looks like this:



8. Under the section that says “alignment,” you will see multiple protein sequences with asterisks, dashes or colons underneath. If you continue scrolling down, you will find a screen that looks like this:



You will see a phylogenetic tree at the top, which provides a graphical representation of the relationships the organisms have.

You will also see the amino acid sequences that were aligned, and how they lined up with each other. You will also see how many amino acids were present in the protein sequences that you aligned. You will also see a number expressed as a percentage called “identity.” This value is essentially the percent of amino acids in the selected sequences that are similar. If all the amino acids were the same, the percent would be 100%. Identical amino acids are marked with two dots between them (:). If there is one dot, the change in amino acid is conservative (both amino acids have similar properties and charge), and if there are no dots, then the two amino acids have different biochemical properties. The amino acids with an asterisk underneath are those which are identical.

9. Examine the section that has the phylogenetic tree:

Print a copy of the tree that was generated. You will need to replace the accession code (the numbers at the end of the branches) with the actual species name of the organism. Place this tree under the data charts for this lab in your notebook.

Part Two Data Table: Completing a Distance Matrix by Calculating Percentage Identity

Determine the percentage identity for the hemoglobin beta protein for each of the species in the table below by aligning each pair of organisms, then copying the "identity" data generated by Uniprot.

Species name	*H. sapiens:* Human	*P. troglodytes*: Chimpanzee	*Bos taurus:* Cow/bull	*Anser anser anser:* Greylag goose	*Ovis aries:* Sheep	*Canis familiaris domesticus:* Dog	*Gorilla gorilla* (Gorilla)	*Caretta caretta:* Logger-head sea turtle	*T. bernacchii:* Emerald rockcod
Homo sapiens (Human)									
Pan troglodytes (Chimpanzee)									
Bos Taurus (Cow/bull)									
Anser anser anser (Greylag goose)									
Ovis aries (Sheep)									
Canis familiaris domesticus (Dog)									
Gorilla gorilla (Gorilla)									
Caretta caretta (Loggerhead sea turtle)									
Trematomus bernacchii (Emerald rockcod)									

Remember that the research question we are investigating is: **What evolutionary relationships exist among a group of vertebrates?**

Using your data, state a claim about the evolutionary relationship between the following pairs of organisms:

- *Homo sapiens* and *Pan troglodytes*
- *Anser anser anser* and *Caretta caretta*
- *Bos taurus* and *Ovis aries*

Use the evidence you have gathered above to defend your claim, and then provide reasoning as to why your evidence does or does not support your claim about the relationships these organisms share. In your reasoning be sure to discuss any differences that may exist between the predicted phylogenetic tree you drew based on morphological data and the tree generated from the molecular data.

Part Three: You Choose!

Now you will choose a protein sequence to align for at least 6 different organisms using the UniProt database. You will conduct a similar analysis of this protein for the organisms you choose. When choosing organisms, you should choose at least one bacterial species, one plant species and if available, one fungal species in addition to the animal species you choose. Be sure to indicate in your lab notebook which of the proteins listed below you are choosing, as well as the scientific names of the organisms you are selecting. **Formulate a hypothesis about which organisms you think will be most closely related, and which organisms you think will be most distantly related to one another.**

You will need to create an appropriate data chart to record the data you collect. Think carefully about the data chart you will produce in order to organize your data in a way that makes sense to you.

Your protein choices are:

Protein Name	**Gene names**	**Function of this protein**
Cytochrome C	CYTC, Cytc1	A cytochrome protein involved in cellular respiration
Superoxide dismutase	SOD1	An antioxidant enzyme important in preventing cellular damage by reactive oxygen species
Catalase	CAT, kata	An enzyme that catalyzes the breakdown of hydrogen peroxide to water and oxygen gas
Aquaporin	AQP-1, AQP1, AYP1	A protein channel that regulates water flow into and out of cells
Cyclin T1	CCNT1	A protein that regulates cell division
Glyceraldehyde 3-phosphate dehydrogenase	GAPDH	An enzyme used in cellular respiration.
Prostaglandin G/H synthase	COX-2, PTGS2	A membrane-bound enzyme used in production of prostaglandins, proteins that mediate the pain response

You will record percentage similarity and copy the phylogenetic tree that is generated from the data you collect. Once you have collected your data and presented it in a table, **state a claim** about the relatedness of the organisms you selected. Then, use your data to determine if it supports or does not support your claim. Provide reasoning that explains why you are able/unable to derive the relationships between the organisms you chose based on your data. In your reasoning, discuss what you infer about the biological importance of the molecule you chose based on the data available.

REPORTING ON RESULTS--THE DELIVERABLE

For Part 1, you will turn in:

- The data chart comparing morphological characteristics of the group of animals you examined.
- Your predicted phylogenetic tree based on morphologies.
- Your claim, evidence and reasoning behind why you identified the relationships between the organisms the way you did.

For Part 2, you will turn in:

- The data chart comparing percentage identity among the organisms listed.
- The phylogenetic tree generated by Uniprot--you will need to screenshot this.
- Your claim, evidence and reasoning behind why you identified the relationships between the organisms the way you did.

For Part 3, you will turn in:

- The data chart comparing percentage identity for the protein you selected for the organisms you chose.
- The phylogenetic tree generated by Uniprot--you will need to screenshot this.
- Your claim, evidence and reasoning behind why you identified the relationships between the organisms the way you did.

You will turn this in as a Google Doc in Google Classroom. Remember to share the document with your teacher so they can **VIEW**.