Analyzing DNA Sequence Similarity on the Computer

In this exercise, you will obtain the sequence of a gene from an opposum and find all known genes that are related to that gene's DNA sequence. You will then compare the sequences of your gene to all of its relatives. You will learn how to use this information to determine the evolutionary relationships between different species.

Instructions:

Part A: Obtain the sequence of a specific gene from the opposum.

1. Open a web browser program and go the website NCBI:
   http://www.ncbi.nlm.nih.gov/entrez/

This brings you to a database called the NCBI Entrez database, which contains the sequence of every gene that has ever been sequenced.

2. In the search window at the top of the page, first select the word "Nucleotide" from the pull-down bar. Then, in the text box next to the pull-down bar, type in "*Monodelphis domestica* low density lipoprotein receptor." Click on the first result that appears: "**AY871266.1**". (Alternatively, you can just search for this file name in order to obtain the desired database entry.) This file contains the sequence of a gene from opposum called the LDL receptor. LDL stands for Low Density Lipoprotein, which is commonly known as "bad cholesterol." LDL receptors are used to bring "bad cholesterol" into cells.

3. Scroll down to the bottom of the page to the section entitled ORIGIN. Highlight the DNA sequence shown there, and then press "apple-C" to copy this sequence. Later, you will paste this sequence into another search box by clicking "apple-V."

Part B: Search a genome database for all known closely related genes to your gene of interest.

Now we will search a genome database for all known genes that have sequences that are similar to this LDL receptor protein from the opposum.

1. Open a web browser program and go the website BLAST:
   http://www.ncbi.nlm.nih.gov/blast
Under the heading "Basic BLAST," click on "nucleotide blast."

2. Paste the gene sequence (by clicking "apple-V) you obtained in Part A into the large search box on the BLAST website labeled: "**Enter accession number, gi, or FASTA sequence**." In the "Choose Search Set" section, select from the pull-down menu the option "Nucleotide Collection (nr/nt)." In the "Program Selection" section, pick the option "Somewhat similar sequences (blastn)." Click the blue button that says "BLAST."

The database of all known sequences is now being searched for those that resemble your sequence. The page will automatically update until the search is done. This may take a minute or two.

3. Scroll down past the box of red lines, and you will see information in a chart that looks like this. This chart is just an example – it is not the one you will actually see.

Sequences producing significant alignments:
(Click headers to sort columns)

Accession	Description	Max score	Total score
gi|110559104|DQ682505.1	Uncultured bacterium isolate SSCP band EJ10-2 16S ribosomal RNA gene, parti	464	464
gi|62866484|AY960266.1	Uncultured Enterobacteriaceae bacterium clone 900 16S ribosomal RNA gene, p	464	464
gi|154197335|EF704569.1	Uncultured proteobacterium clone MS123A1_D08 16S ribosomal RNA gene, part	459	459
gi|154197276|EF704510.1	Uncultured proteobacterium clone MS122A1_G07 16S ribosomal RNA gene, part	459	459
gi|154196798|EF704032.1	Uncultured proteobacterium clone MS111A1_C06 16S ribosomal RNA gene, part	459	459
gi|154195847|EF703081.1	Uncultured proteobacterium clone MS093A1_H06 16S ribosomal RNA gene, part	459	459
gi|154195764|EF702998.1	Uncultured proteobacterium clone MS092A1_F11 16S ribosomal RNA gene, part	459	459
gi|154192950|EF700184.1	Uncultured proteobacterium clone MS145A1_B08 16S ribosomal RNA gene, part	459	459
gi|154192947|EF700181.1	Uncultured proteobacterium clone MS145A1_B05 16S ribosomal RNA gene, part	459	459
gi|154192931|EF700165.1	Uncultured proteobacterium clone MS145A1_A01 16S ribosomal RNA gene, part	459	459
gi|154184881|EF707287.1	Uncultured proteobacterium clone MS171A1_E11 16S ribosomal RNA gene, part	459	459
gi|154184879|EF707285.1	Uncultured proteobacterium clone MS171A1_E09 16S ribosomal RNA gene, part	459	459
gi|154184876|EF707282.1	Uncultured proteobacterium clone MS171A1_E06 16S ribosomal RNA gene, part	459	459
gi|154184859|EF707265.1	Uncultured proteobacterium clone MS171A1_C12 16S ribosomal RNA gene, part	459	459
gi|154184839|EF707245.1	Uncultured proteobacterium clone MS171A1_B04 16S ribosomal RNA gene, part	459	459
gi|152218530|EU009194.1	Shigella sonnei strain FBD020 16S ribosomal RNA gene, complete sequence	459	459
gi|152218529|EU009193.1	Shigella sonnei strain FBD019 16S ribosomal RNA gene, complete sequence	459	459
gi|152218528|EU009192.1	Shigella sonnei strain FBD018 16S ribosomal RNA gene, complete sequence	459	459
gi|152218527|EU009191.1	Shigella sonnei strain FBD017 16S ribosomal RNA gene, complete sequence	459	459
gi|152218526|EU009190.1	Shigella sonnei strain FBD016 16S ribosomal RNA gene, complete sequence	459	459
gi|150408730|EF653419.1	Uncultured bacterium clone V5_2_90_B12/19-1507 16S ribosomal RNA gene, pa	459	459
gi|146286011|EF206275.1	Uncultured bacterium clone 57631 16S ribosomal RNA gene, partial sequence	459	459
gi|146285987|EF206251.1	Uncultured bacterium clone 57603 16S ribosomal RNA gene, partial sequence	459	459
gi|146285822|EF206086.1	Uncultured bacterium clone 54115 16S ribosomal RNA gene, partial sequence	459	459
gi|146285820|EF206084.1	Uncultured bacterium clone 54113 16S ribosomal RNA gene, partial sequence	459	459
gi|146285816|EF206080.1	Uncultured bacterium clone 54109 16S ribosomal RNA gene, partial sequence	459	459
gi|146285775|EF206039.1	Uncultured bacterium clone 52830 16S ribosomal RNA gene, partial sequence	459	459
gi|146285696|EF205960.1	Uncultured bacterium clone 47015 16S ribosomal RNA gene, partial sequence	459	459
gi|146285578|EF205842.1	Uncultured bacterium clone 7531 16S ribosomal RNA gene, partial sequence	459	459
gi|146285575|EF205839.1	Uncultured bacterium clone 7528 16S ribosomal RNA gene, partial sequence	459	459
gi|148913178|EF620926.1	Escherichia coli strain WD02 16S ribosomal RNA gene, partial sequence	459	459
gi|148913176|EF620924.1	Escherichia coli strain SYW003 16S ribosomal RNA gene, partial sequence	459	459

As you can see in the example above, the first two full names (genus and species) of organisms that come up are *Shigella sonnei* and *Escherichia coli*. Let's zoom in on the last two lines:

| gi\|148913178\|EF620926.1 | Escherichia coli strain WD02 16S ribosomal RNA gene, partial sequence | 459 | 459 |
| gi\|148913176\|EF620924.1 | Escherichia coli strain SYW003 16S ribosomal RNA gene, partial sequence | 459 | 459 |
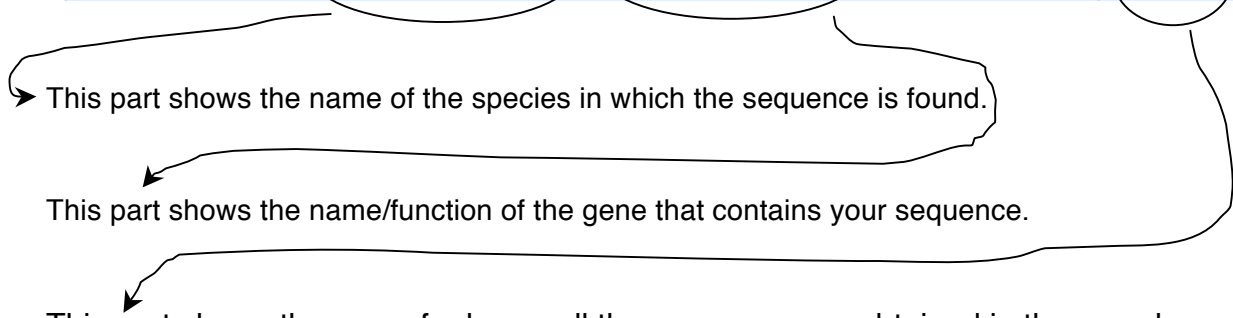
This part shows the name of the species in which the sequence is found.

This part shows the name/function of the gene that contains your sequence.

This part shows the score for how well the sequence you obtained in the search matched the sequence you entered into the database. The higher the score, the better a match it is. Scores in the range of 500 indicate very good matches. This score takes into account how identical the two sequences are (the "Query coverage" column) and over how long of a stretch of DNA the two sequences are that similar.

4. Record the information from your search in the table on the next page.

Record the **_seven different species_** that contain the closest known matches to your sequence. For each species, write its Latin name, its common name (i.e. what you would call it in a non-scientific conversation), and the score that indicates how good of a match your sequence was to that entry in the database.

** Do <u>NOT</u> list the same species more than once in your chart. List 7 **different** species.

	Species name in Latin *(Genus species)*	Common Name (obtained from Google Image searches)	Total Score
1			
2			
3			
4			
5			
6			
7			

Now answer the following questions, which relate to the phylogenetic trees that we made earlier.

1. Compare the total scores showing the similarity between the opposum, platypus, cow, and macaque genes for the LDL receptor. Are these results consistent with the tree you made earlier using DNA sequence? Why or why not?

2. If you had to add in the species *Homo sapiens* to the tree we made, where would you place it and why?

3. If the LDL receptor gene from elephant had been sequenced and was thus in the database, what total score would you predict that you would have acquired from this search? Explain your answer.