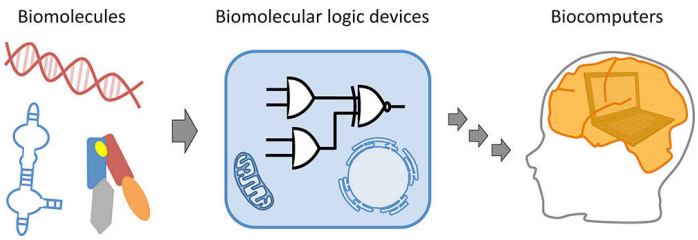


Name_____

Period_____

Have a BLAST with Biocomputation

The analysis of biological data using computational analysis is called **biocomputation** or **bioinformatics**. In this exercise, you will compare a specific DNA sequence among 5 different species and use this genetic information to construct a cladogram of their evolutionary "relatedness". You will also draw a cladogram based upon traditional phenotypic information, and then compare the two diagrams.

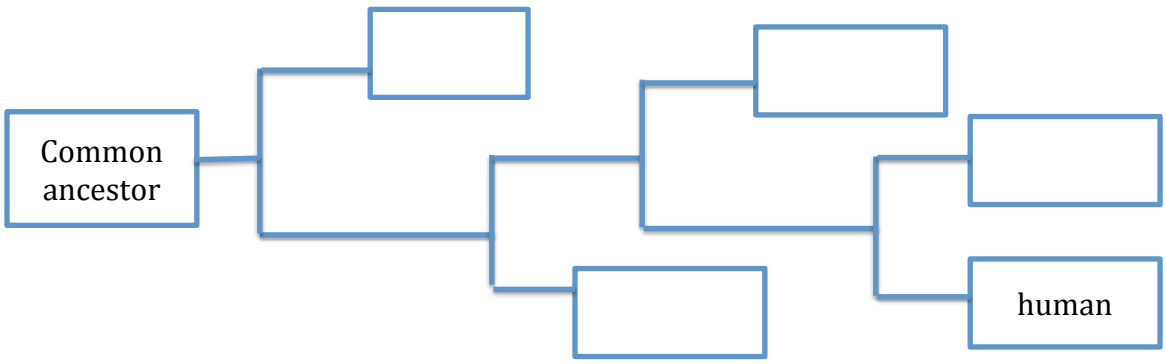
Pre-lab Activity: Answer questions 1and 2, using your background information and the Standard Genetic Code table.

1. Here is a list of various traits for humans and four unknown mammals. A '+' indicates that a species possesses that characteristic; a blank space means they lack it.

Species	Claws vs. Hooves		Placental vs Marsupial		Fur vs. hair		Quills
#1	+		+		+		
#2	+		+		+		+
#3	+			+	+		
#4		+	+			+	
#5 human			+			+	

Hint about fur and hair: Furred animals have: two types of hairs - soft under-hair called ground hair and coarse protective outer hair called guard hair). Haired animals just have the outer guard hair.

In the space below, use the data in the above table to fill in the cladogram that groups the animals using their phenotypic characteristics. Write the number of the organism in the appropriate box and then guess the name of each unknown animal and write it in the box. Finally, label the location (or junction point) of each adaptation (trait), showing where this adaptation occurred.



3. Use the Standard Genetic Code table provided to fill in the missing information. You will then need to figure out and write in the corresponding tRNA anticodon.

Amino Acid	3-letter Abbreviation	DNA codon	mRNA codon	tRNA anticodon
Methionine	Met		AUG	CAU
	Ser		UCA	
Arginine	Arg			CCU
	Val	GTA		UAC
		GGC	GGC	
	Pro		CCA	
Proline		CCA		UGG
		TTA		
Leucine				UAG
Arginine		CGC	CGC	

Lab Activity Part 1:
Use BLAST to compare gene sequences and draw a cladogram:
In this activity, you will now compare DNA segments from the gene BRCA1, and use that information to analyze similarities and differences between the five species. To do this, you will use a gene sequence comparison tool called BLAST (Basic Local Alignment Search Tool). BLAST is powerful software provided for free to all researchers (and students) by the National Center for Biotechnology Information (NCBI). This program is used millions of times a month by scientists all over the world to compare unknown DNA with the complete database of public DNA sequences.

1. To get started, refer to your BLAST tutorial sheet. Go to blast.ncbi.nlm.nih.gov and follow the instructions on the tutorial sheet to see how BLAST works. Have a BLAST!!

2. OK, now we are ready to figure out to which animal each of the four sequences belongs. Use BLAST to determine the genus and species for each of the unknown sequences and the animal's common name. To do this:

   Go to apbiostudio.stanford.edu/biocomputation/Unknown-Sequences.html
   Copy the first DNA sequence (just the sequence, not the label).

   Once you have copied the sequence, paste that sequence into the BLAST form at blast.ncbi.nlm.nih.gov

   Collect information about this unknown sequence and record in Data Table 1 below.

   Repeat for the other 3 sequences

Data Table 1: Key Data from BLAST

Unknown	Genus & Species of the best match	Expect Number	Percent Identity	What is the common name for this species
#1				
#2				
#3				

#4				

Hint for common name: Google it based upon the scientific name.

3. Use the percent identify in the table above and draw a new cladogram for these species. Place your answer in the space below.

4. Did you get the same answer with DNA data as with the phenotype data? Explain and describe reasons for the similarities and differences between the two data sources. Your answer should consider the source of each type of data, breadth of information available from each source (DNA vs. phenotypes). Write your answers in the space below.

Lab Activity, Part II. Analyze DNA sequences.
Shown below are the five sequences we've been working with so far. Note that this sequence is just a small portion of the full BRCA1 gene from the 3' end of the gene.

```
      5335  (coordinates of H. sapiens BRCA1 gene)                              5414
      |                                                                          |
human TGGTCAATGGAAGAAACCACCAAGGTCCAAAGCGAGCAAGAGAATCCCAGGACAGAAAGATCTTCAGGGGGCTAGAAATC
#1    --GTGAATGGAAGAAACCACCAGGGTCCAAAGAGAGCAAGAGAATCCCGGGACAGGAAGATCTTCAGGGGCCTAGAAGTC
#2    --GTTAACGGAAGAAACCACCAAGGTCCAAAGCGAGCAAGAGAATTCCACGACAGAAAGATCTTCGAGGGCCTAGAAGTC
#3    --GTCAATGGAAGAAACCATCAAGGCCCAAAACGAGCAAGAGAATCCCGGGACAAGAAGATCTTCAAGGGCCTGGAAATC
#4    ---TCAATGGGAGAAACCACCGAGGCCCAGAAAGAGCAAGAGAATCTCAGGGGATGAAGATCTTCAGGGGCCTGGAAATC
         * ** **.******** *..** ***.*..************  *. *. * .*********..*** **.***.**
```

Here is more information that will help you interpret the DNA sequences:

- **Coordinates:** There are two numbers that mark the ends of the DNA segment, 5335 and 5414 (5335 means it is the 5,355th nucleotide in BRCA1). These coordinates are important, and you will use them later in Part 3.
- **Sequences aren't perfectly aligned:** T is the 1st nucleotide in the human sequence at position 5335. Note that the other sequences are not all the same length (dashes are a space holder representing a gap or missing nucleotide).
- **Interpreting the * and _ symbols in the last row:** The last row of asterisks and dots shows which bases are conserved in the five sequences. For example, at human coordinate 5338, all five sequences have a T and a * appears in the last row. At position 5412, you find an A in three of the sequences and a G in the other two. These are nearly the same, and a _ appears in that position. If there is no conservation, then there is no mark as found at position 5339 and 5442.

The next step: Figure out which amino acids are conserved in the protein. These are a bit trickier because we are in the middle of this gene. We can't identify the usual DNA start codon (5' ATG 3' on the coding strand or 3' TAC 5' on the template strand) and begin there.

Here's a helpful trick for solving this problem: recall that multiple codons can code for the same amino acid and that the differences tend to be in the third nucleotide. For example, GCU, GCC, GCA, and GCG all code for the same amino acid, alanine (ala). Knowing this, take a look at the sequences above. Do you see a pattern that might help you figure out which is the correct reading frame? In other words, is there a pattern of two conserved and one not conserved repeated over and over?

1. Using this idea, find the first amino acid shared by all five DNA sequences and write your answer here_____.

2. Check your answer: The 1st amino acid shared by all five sequences has a DNA code of AAT, which corresponds to the amino acid asparagine (asn).

3. **Now you are ready to translate all five proteins.** To make things easier:
   - Work in groups of five students – each will translate one of the DNA sequences.
   - Translate the DNA code to mRNA code and then to an amino acid sequence. Write your answer in Data Table 2 below.
   - When you are finished, check each other's work before proceeding.

Data Table 2: DNA, mRNA and amino acid sequences

Organism	DNA (positions 5335-5414), mRNA code and amino acid sequence (25 amino acids)
Human DNA	TGGTCAATGGAAGAAACCACCAAGGTCCAAAGCGAGCAAGAGAATCCCAGGACAGAAAGATCTTCAGGGGGCTAGAAATC
mRNA	
amino acids	
1 DNA	--GTGAATGGAAGAAACCACCAGGGTCCAAAGAGAGCAAGAGAATCCCGGGACAGGAAGATCTTCAGGGGCCTAGAAGTC
mRNA	
amino acids	
2 DNA	--GTTAACGGAAGAAACCACCAAGGTCCAAAGCGAGCAAGAGAATTCCACGACAGAAAGATCTTCGAGGGCCTAGAAGTC
mRNA	
amino acids	
3 DNA	--GTCAATGGAAGAAACCATCAAGGCCCAAAACGAGCAAGAGAATCCCGGGACAAGAAGATCTTCAAGGGCCTGGAAATC
mRNA	
amino acids	
4 DNA	---TCAATGGGAGAAACCACCGAGGCCCAGAAAGAGCAAGAGAATCTCAGGGGATGAAGATCTTCAGGGGCCTGGAAATC
mRNA	
amino acids	

4. **Analyze your translated proteins in order to identify the conserved amino acids**. Conserved amino acids remain consistent over a wide range of organisms for a reason – they tend to be critical areas for the protein to properly function. **Go back to Data Table 2 and underline the conserved amino acids.**

Lab Activity, Part III:
Study the Impact of Different Mutations in BRCA1
The gene named Breast Cancer (BRCA1) is an important tumor suppressor gene that is responsible for repairing DNA. The BRCA1 gene isn't just found in humans, it is present in all mammals (and some other animals too!). As you might guess from its name, when this gene isn't working right, it can lead to breast cancer in humans. A large proportions of inherited human breast and ovarian cancers are caused by BRCA1 mutations.

What does the BRCA1 protein do? The protein encoded by the BRCA1 gene combines with other tumor suppressors, DNA damage sensors, and signal transducers to form a large protein complex known as the BRCA1-associated genome surveillance complex (BASC). The BRCA1 protein works with RNA polymerase II and also interacts with histone deacetylase (which de-acetylates histones and hinders transcription). That is why defective BRCA1 genes cause cancer – mutations aren't repaired and certain genes aren't transcribed.

Recall the human BRCA1 DNA sequence you used in Parts 1 and 2 of this lab. This DNA sequence is what scientists call the **reference sequence**, or the most common bases at each position. The reference sequence doesn't show all the individual variants within a population. As you know, we all have millions of DNA differences sprinkled across our genomes compared to the reference sequence. While most of these have little effect, a sizeable minority do. It is these that make you look and act a bit different than your friends. The BRCA1 reference sequence is shown below.

```
5335        5345        5355        5365        5375        5385        5395        5405
|           |           |           |           |           |           |           |
TGGTCAATGG AAGAAACCAC CAAGGTCCAA AGCGAGCAAG AGAATCCCAG GACAGAAAGA TCTTCAGGGG GCTAGAAATC
```

Reference sequences can also help scientists figure out if a change has an effect. For example, if all redheads have a certain difference in their MC1R gene, then this might explain their red hair (it does by the way). The argument for this mutation being the cause of red hair is made even stronger if the mutation changes a conserved amino acid. Looking for changes in conserved amino acids can also help scientists figure out which DNA changes (or mutations) led to a cell becoming pre-cancerous or cancerous.

1. You will now study 4 different variants of this sequence and determine their impact. **DNA variants** are noted with shorthand symbols using the nucleotide coordinate. A change in the nucleotide is written with the reference nucleotide, an arrow ('>) and the changed nucleotide. For example, 300G>C says that nucleotide 300 is changed from a 'G' to a 'C'.

   Practice: Write out the symbols for changing nucleotide 5385 to "A"_____

2. Insertion ('ins') or deletion ('del') of one nucleotide are represented with the nucleotide coordinate, "ins" or "del" and the changed nucleotide. For example, 300insA means that an 'A' was added after nucleotide number 300. An example deletion is 300delG, deleting nucleotide 300 which happens to be a G.

   Practice: Symbolically write out an insertion of "G" just after nucleotide 5365_____.

   Write a deletion of nucleotide 5405_____.

3. **You will now analyze 4 different BRCA1 variants**. For each, describe the number of changed amino acids and what the change is (e.g. from ala to val), whether the change affects a conserved amino acid, and what you think the impact to the protein will be. Record your answers in Data Table 3 below:

Data Table 3: BRCA1 mutations and their impact

Mutation	Were conserved amino acids affected?	Describe the changed amino acids.	Give an opinion as to whether this might cause cancer. Justify your opinion. (Hint big changes = dead protein = cancer)
5346 A>C			
5376 G>T			
5389 G>T			
5382 ins C			

4. **Now, record the location of each mutation on the reference DNA sequence** at the top of this page. Under the reference sequence, write in the changed amino acids and label the mutation #.

5. **Check your answers!** Particular variations, mutations, to the BRCA1 gene are associated with an increased risk of developing breast or ovarian cancer. With some variants the odds increase from a 12% chance of developing breast cancer in your life to 40-60% change.

   One of the variations you analyzed is included in DNA analysis conducted by the DNA testing company 23andMe – the5382insC variant. Select one or more of these resources and read about this mutation and the impact it has. Then, compare your assessment of 5382insC (answer #4 above) to the actual data.

Description	URL
BRCA Factsheet	National Cancer Institute, NIH
BRCA1 Information	Wikipedia
Genetics Home Reference for BRCA1	National Library of Medicine
23andme BRCA1, includes 5382insC	23andme.com
In depth description research and history of BRCA1 and disease	Online Mendelian Inheritance of Man
Angelina Jolie tests positive for a BRCA1 variant	PBS.org
Myriad Genetics controls testing for BRCA1	Forbes.com
ACLU on Myriad Genetics case at the Supreme Court	ACLU.org

6. **Compare your answer to the actual impact for the 5382insC mutation.**

   A. Were you correct or incorrect?

   B. Summarize the problems caused by the 5382insC mutation, and reasons why it has such a negative effect.

   C. Write your typed response on a separate sheet of paper, and make sure you give a thoughtful, thorough answer with supporting evidence from the suggested readings listed above.