NAME: _____

LABORATORY 8; EXERCISE 3. ANALYSIS OF MULTIPLE DNA SEQUENCES

Purpose – This Laboratory will illustrate how information can be extracted from DNA segments using the techniques of BioInformatics. Practice in reading sequences, sequence alignment, data extraction and data summary are goals of this exercise. You should gain additional appreciation for the vast range of nucleotide variation within and among DNA segments.

This Laboratory will be divided into two data analysis exercises. In Exercise 1, you will use the BioEdit program to explore nucleotide sequences from mitochondrial DNA regions of six species of mammals. In Exercise 2, you will use nucleotide difference data as the basis for development of a Gene Tree.

SPECIFIC LABORATORY PROTOCOL –

1. Log in to your VM desktop through the COSAM Desktop Portal.

2. OPEN the Windows Start  Icon and OPEN the BioEdit program. **NOTE –** DO NOT close BioEdit through the entire lab, only minimize it if you need to. You will use this information for the whole series of experiments.

3. Click On the FILE tab and then OPEN. Highlight the "This PC"  icon.

4. Select the (J:) Drive labeled "Academic – Class Resources" by double-clicking it.

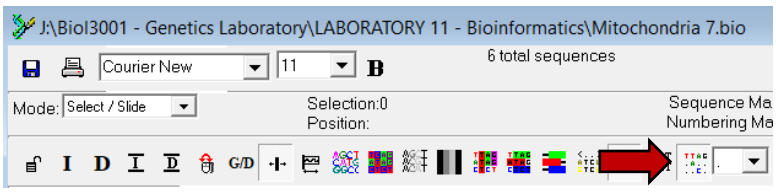
5. OPEN the "Biol3001 – Genetics Laboratory" subdirectory and be sure at the bottom of the "Open File …" box, that the "Files of Type:" box is showing "ALL FILES (*.*)". You can select this from the drop down menu.

6. OPEN the "LABORATORY 11 – Bioinformatics" subdirectory and then select one of the "*Mitochondria*" sequences (choose ANY single sequence 1-14!!) in the subdirectory.

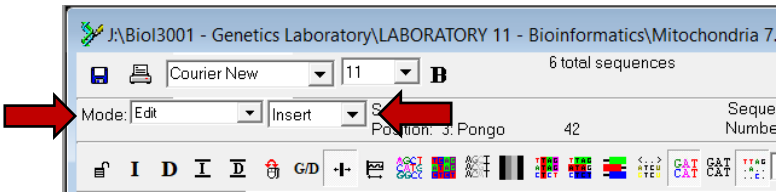
7. Click on and OPEN the data set and record the organisms you have listed in the spaces below.

Organism 1	
Organism 2	
Organism 3	
Organism 4	
Organism 5	
Organism 6	

8. SELECT the "Dot" Format option as you have before to more easily visualize differences in the sequences.



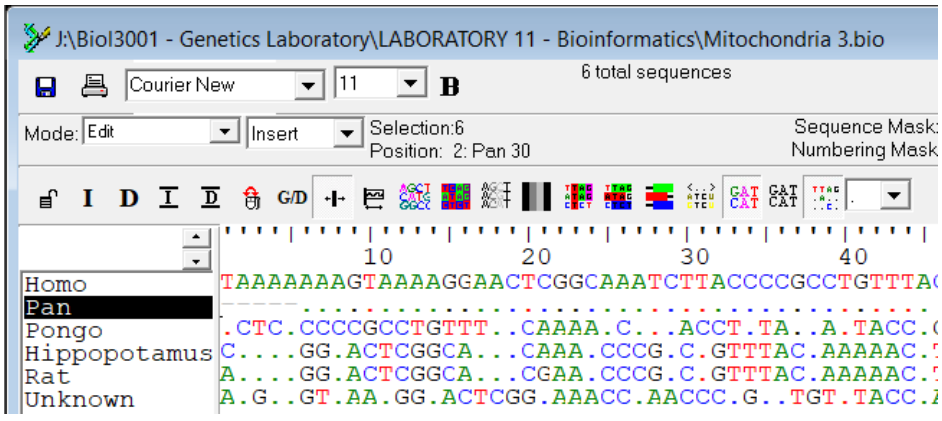
9. SELECT the Mode and change the dropdown menu to the “Edit and Insert” option.



10. Now take time to examine your sequence alignments. What you should notice is that there is not much agreement between the sequences (i.e. there are very few dots that actually show up).This does not necessarily mean that the sequences do not align with each other. It could also mean that whoever generated the sequeces did not start reading them at exactly the same position in each sequence. In order to compare the sequences, you will need to actually align them. This can be done in three different ways: Visual Alignment, Pairwise Alignment or Group Alignment.

11. VISUAL ALIGNMENT – This method essentially is you as the researcher doing the alignment by hand and is the orginal way that sequence alignments were performed before scientist really had the power of computers behind them. This is the most time consuming method and thus is also the one that results in the most errors. But at one point in time, this was cutting edge science. To help you understand where the following methods come from, you will perform ONE alignment using this method.

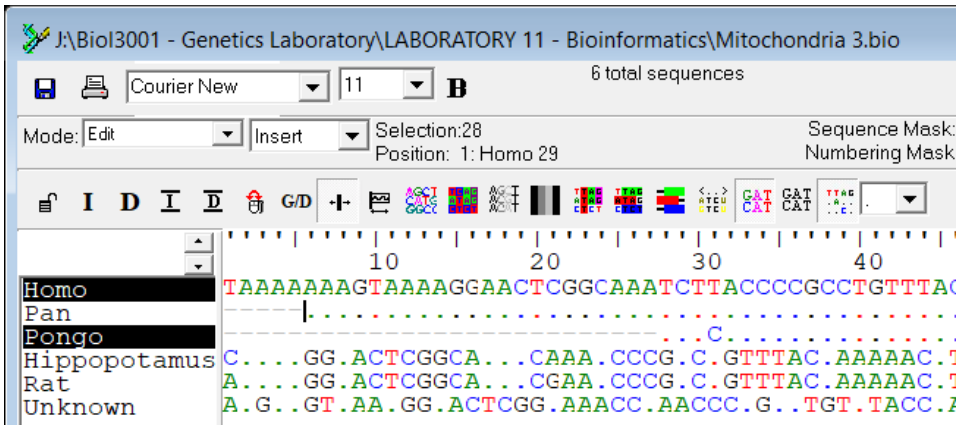
- CLICK on and Highlight the organism labeled PAN.
- Place your cursor on the first nucleotide of that sequence and CLICK on it. You should see a flashing black line appear before that nucleotide.
- Using the DASH key (“-“) on your keboard, slowly begin to type in one dash at a time to the sequence. After adding 4-5 dashes, you should see that most of the nucleotides suddently become Dots. This indicates that the two sequences are now aligned. Each Dot indicates that the two nucleotides in each sequence are identical. Any differences between the two sequences will appear as the alternate nucleotide.
- If ALL of the sequences you are working with are VERY similar in their overall sequences, this method can work very will still today. However, for sequences with many different differences, it is WAY TOO TEDIOUS to perform alignments in this method.



Visual Alignment of the Pan sequence to the Human sequence.

12. PAIRWISE ALIGNMENT – This is the first of two methods that use a computer algorithm to decide on the best alignment between sequences. Pairwise sequence alignment methods are used to find the best-matching piecewise (local or global) alignments of TWO query sequences. Pairwise alignments can only be used between two sequences at a time, but they are efficient to calculate and are often used for methods that do not require extreme precision (such as searching a database for sequences with high similarity to a query).

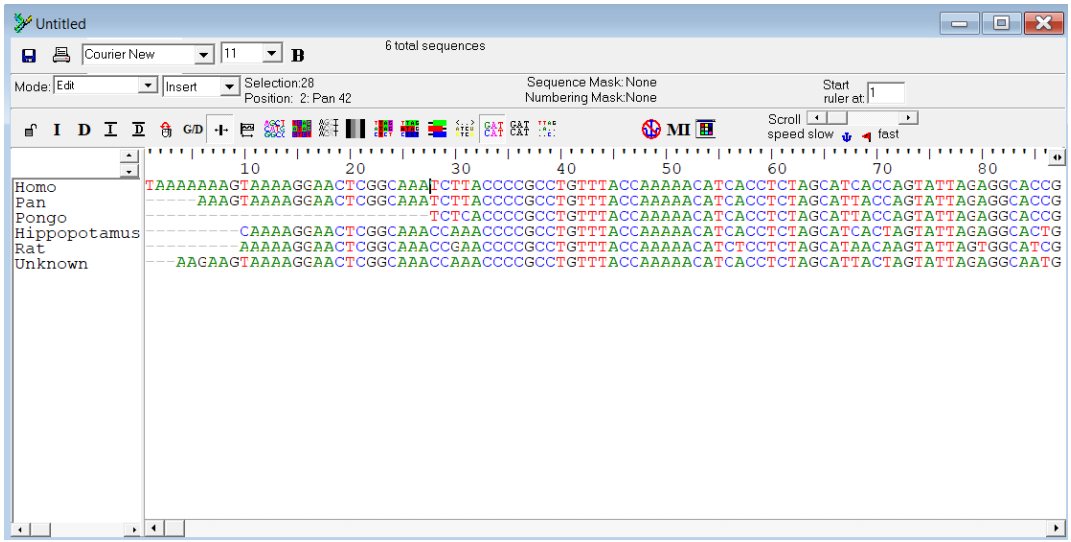
- CLICK on the sequence labeled HOMO (i.e. Homo Sapien) to select it as your first sequence.
- While Holding down the CTRL key on your keyboard, CLICK on the sequence labeled PONGO. Both sequence names should now be highlighted.
- In the toolbar at the top of the page, CLICK on the **Sequence** tab and select "Pairwise Alignment" from the dropdown menu. This should open another dropdown menu and select "Alighn two sequences (allow ends to slide)" from that last menu. Two new windows will open. The smallest one labeled "Pairwise Alignment Summary" can be closed as it just summarizes the parameters of the alignment performed and for your purposes, does not provide any pertinent information.
- Your ALIGNED sequences will appear in the other window. You can toggle how you view the sequences in this window by using the Dot option as above. You should notice that when the appropriate number of dashes are added to the beginning of the PONGO sequence, it aligns almost exactly with the HUMAN sequence.
- Using the computer generated pairwise alignment, MANUALLY add the appropriate number of DASHES ("-") to the beginning of your PONGO sequence in the original window listing all your organisms. You will need to select just the PONGO sequence by clicking on it and then clicking on the first nucleotide as you did above for the Visual Alignment.



Pairwise Alignment of the Pongo sequence to the Human sequence.

13. MULTIPLE ALIGNMENT – While Pairwise Alignment is excellent for two sequences, it quickly becomes difficult when you have many different sequences. To solve this problem, Multiple Alignment algorithms have been developed. Multiple Alignment methods try to align all of the sequences in a given query set. Multiple alignments are often used in identifying conserved sequence regions across a group of sequences hypothesized to be evolutionarily related. Such conserved sequence motifs can be used in conjunction with structural and mechanistic information to locate the catalytic active sites of enzymes. Alignments are also used to aid in establishing evolutionary relationships by constructing phylogenetic trees.

- To align all six of your sequences, CLICK on the sequence labeled “Homo” and while holding the SHIFT key down on your keyboard, CLICK on the last sequence on your list (“Unknown”). You should see all of your listed sequences Highlighted.
- From the toolbar at the top of the page, Select the **Accessory Application** tab and from the dropdown menu, select the “ClustalW Multiple Alignment” option.
- A new screen will appear with “ClustalW Options”. Do not change any of the default settings and then CLICK on the “Run ClustalW” button at the bottom of the window. You will get another window titled “BioEdit Sequence Alignment Editor. Again, this is simply information about where files are being generated and stored as part of the integration between ClustalW and BioEdit. Bypass this window by CLICKING the “OK” button.
- You will have a new window open with all six of your sequences aligned. This is the window you will use for the remainder of this lab

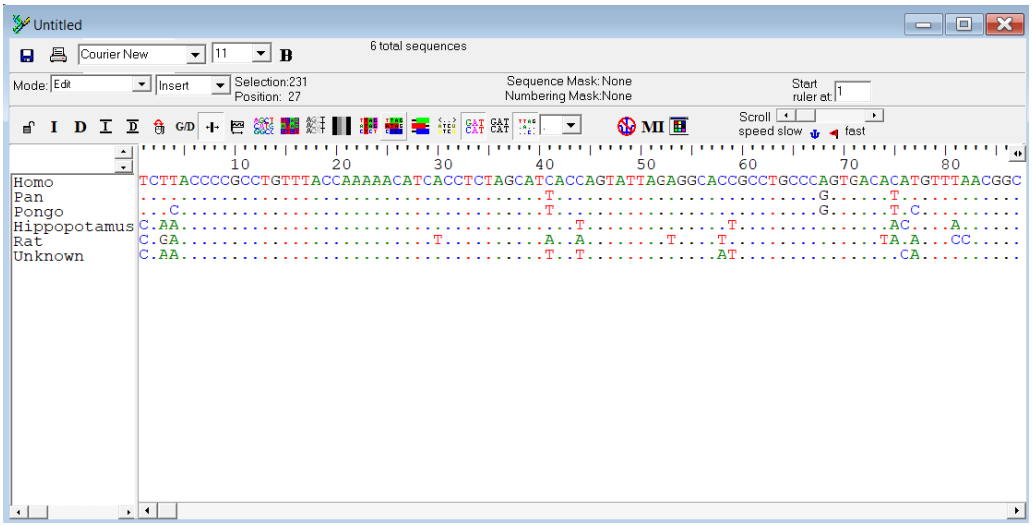


Multiple Alignment of ALL sequences in your database.

14. Since it should now be obvious that each of these sequences was started at a different point when they were entered in to the database, you will probably find it easier to work with them if they all look a little more alike than they do now. To accomplish that, you will need to EDIT each sequence a little bit using the same methods you have used previously.

- Change the Mode to EDIT and chose the INSERT option from the dropdown box.
- Choose any nucleotide you wish to get rid of in each sequence and using the BACKSPACE key on your keyboard to remove it (you may need to CLICK on the nucleotide immediately after the one you wish to get rid of.
- Remove all the nucleotides that you deem unnecessary from each sequence. HOWEVER, REMEMBER you want to maintain your alignment!!! So make sure all the sequences wind up being the same length and that the overall alignment pattern you have just worked so hard to produce maintains itself.

15. You should now be able to toggle the Dot button to see how much sequence homology you have between all of your organisms.



Use BLAST to Identify your DNA Region and your Unknown Organism.

15. Determine the identity of your assigned mitochondrial DNA region and determine the Species name and information of the Unknown by searching GENBANK as you have done previously.

- CLICK on the sequence label of the Unknown Mitochondrial sequence after you have completed your reorganization above.
- Select **Edit** from the toolbar at the top of the BioEdit Window and then “Copy Sequences to Clipboard (Fasta Format)”.
- OPEN your browser of choice (remember to NOT CLOSE BioEdit or you will have to start from scratch again!!) and navigate to https://blast.ncbi.nlm.nih.gov/Blast.cgi.
- BLAST your sequence information and answer the questions on the following pages.

BIOL 3001 GENETICS LABORATORY

A. Organism (Genus and Species)

Accession Number

Identification

B. The DNA found in mitochondria is generally one of three types: 1. Control Region (controls replication in the mitochondria); 2. Gene Regions (one of the 13 different genes that are present in the mitochondrial genome that code for mostly the Electron Transport Chain); or tRNA Regions (provides the information to make tRNA molecules used in translation of the mitochondrial genome.

In what region of the mitochondrail genome is your fragment located and what is the function of this region? (may require you to do a web search or PubMed search; both can be done at the NCBI site)

C. Open back up your BioEdit window (remember the one you did not close out so you don't have to redo everything!), and look at the aligned sequences. Do you see any patterns in the alignments? For example, are there nucleotides that are different throughout all the sequences? Are there differences that seem to cluster together? Do you see differences in alignment between different sets of organisms?

Those are just some of the questions you can think about as your look at your sequences. In the space below, discuss BRIEFLY what those observations mean to you. Do they indicate anything to you about how your Unknown compares to the other sequences (functionally or maybe evolutionarily)?

D. List the sequence your Unknown was most similar to and the sequence your Unknown was the most different from. Discuss BRIEFLY what this might tell you regarding relationships of the known species to your Unknown.