NAME:

LABORATORY 8; EXERCISE 4. CONSTRUCTION OF A DENDROGRAM

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.

Branching diagrams, formally known as dendrograms (most often called “Trees”), serve as the starting point for many Biological investigations. A basic understanding of how to construct and interpret Trees is an important skill for any student of Biology. Data from DNA sequences is especially useful for constructing Trees. Below you are asked to generate a simple dendrogram that illustrates relationships among the six DNA segments from Exercise 1.

SPECIFIC LABORATORY PROTOCOL –

1. As a biology student, you should have some idea how the six species you aligned in Exercise 1 being might be related to each other. If you do not know the identity of the Genus names of each species, a quick web search should provide you with that information. Once you have identified all your species, in the space below, draw your best professional guess as to the expected Relationships in the form of an Evolutionary Tree. (i.e. which species do you expect to be the closest in relationship?).

Construct a Nucleotide Difference Matrix Using Your Species from Exercise 1.

BioEdit provides multiple tools that can be used to generate a Nucleotide Differenece Matrix. You will use two different methods in the program to generate your specific matrix.

2. METHOD 1: VISUAL NUCLEOTIDE COUNTING – This is the method described in the Background portion of the laboratory.

2a. Using the aligned sequences in BioEdit from Exercise 1 (hopefully you didn't close the window!!), toggle the Dot format to best show the nucleotide differences between each species.

2b. Compare Sequence 1 with Sequence 2. Count the total number of bases that differ between these two sequences and record this in the matrix given below.

2c. Repeat the comparison with the Sequences for each species (1 – 3), (1 – 4), (1 – 5) and (1 – 6). Record each of the difference values in the proper blank in the matrix.

2d. Right CLICK on Sequence 2 (the "Pan" sequence) on the left hand side of the window. This will highlight this sequence and make it the "master" sequence. Repeat the comparisons above only this time, using Sequence 2 as the comparison sequence (e.g. (2 – 3), (2 – 4), etc.

2e. Repeat the above steps until you have completed filling in the matrix below with each specific nucleotide comparison difference. Be sure to include your "Unknown" Identity as part of the matrix

2f. You should now have a completed Nucleotide Difference Matrix to use to generate a Dendrogram of the relationship between your species.

VISUAL NUCLEOTIDE COUNTING MATRIX

	Sequence 1 (Human)	Sequence 2 (Pan)	Sequence 3 (Pongo)	Sequence 4 (Hippo)	Sequence 5 (Rattus)	Sequence 6
Sequence 1 (Human)	0					
Sequence 2 (Pan)		0				
Sequence 3 (Pongo)			0			
Sequence 4 (Hippo)				0		
Sequence 5 (Rattus)					0	
Sequence 6						0

3. METHOD 2: COMPUTER GENERATED MATRIX – This uses the power of the computer to generate a difference matrix.

3a. While holding down the SHIFT key, CLICK on each of the species name on the left side your BioEdit window.

3b. From the tool bar at the top of the page, Select the Alignment tab and on the dropdown menu, Select "Sequence difference count Matrix".

3c. Choose any file name you would like to call this file and enter that information in the box. **CHOOSE YOUR H: DRIVE AS THE LOCATION TO SAVE THE FILE TO!! Otherwise you will lose this file and information when the BioEdit program closes.**

4d. When you CLICK the SAVE button, you will have another window open titled "Difference Count Matrix". This is the computer generated matrix BioEdit produces.

4e. Enter the information from the computer generated matrix in the space below

COMPUTER GENERATED COUNTING MATRIX

	Sequence 1 (Human)	Sequence 2 (Pan)	Sequence 3 (Pongo)	Sequence 4 (Hippo)	Sequence 5 (Rattus)	Sequence 6
Sequence 1 (Human)	0					
Sequence 2 (Pan)		0				
Sequence 3 (Pongo)			0			
Sequence 4 (Hippo)				0		
Sequence 5 (Rattus)					0	
Sequence 6						0

Does the computer generated matrix agree with your visual nucleotide difference matrix? If not, BRIEFLY describe any differences and hypothesize on why you might be seeing those differences.

CONSTRUCT A NEAREST NEIGHBOR TREE

Follow the step-by-step guide in the Background for this laboratory and construct a Nearest Neighbot dendrogram from your sample data. You will be expected to create both LIST 1 and LIST 2 as described for all comparisons and finally draw out your best dendrogram based on the data you have accumulated. You should provide a title at the top of the box describing the protein you are comparing, you should provide a list of the species being compared and you should provide some kind of scale with appropriate numbers to resolve the relationships.

USING BIOEDIT TO CONSTRUCT A TREE

Because all tree-building algorithms require you to repeat the same steps over and over, computers are ideal machines to help construct trees. In fact, some of the earliest use of computers was to build topologies (networks). We will use BioEdit to produce a tree for comparison to yours.

4a. SELECT all six sequences by holding down the SHIFT key and CLICKING on the sequence names in the left side of the window.

4b. From the toolbar at the top of the page, Select Accesory Application and from the dropdown menu, Select "DNADist – Neighbor phylogenetic tree".

4c. A small window will appear. CLICK on the RUN APPLICATION button. Several windows will appear after you run the program. You can close all of the new windows except the one that contains the "Tree". It should be the window on top.

4d. Transfer this information to the box below.

Answer the following questions regarding both of the dendrograms you generated above.

1. Does your Visual Nucleotide Difference tree match the Computer generated tree? Why or why not? Discuss any reasons that might have caused these differences.

2. Do your generated trees match your Predicted Tree from the very beginning of the exercise? If not, identify the differences and discuss why your reasoning led you to the predicted tree versus what the generated tree tells you about the relationships between the species.

3. Describe in words what patters you see in dendrograms. For example, how many main groups are there? What species occur closest together and what does that tell you about their evolutionary relationship. What species did you find to be the most different? Why do you think they are the most different. Be brief but thorough in your explanations.