NAME: _____

LABORATORY 9; EXERCISE 2. MUTATIONS

Purpose –These Exercises will use the BioEdit program. For this exercise, you will examine nucleotide variation in DNA sequences obtained from a set of similar species. Data from these sequences will be used to test predictions from the hypothesis that Transitions occur more frequently than do Transversions.

In this Exercise, you will examine nucleotide variation in DNA sequences obtained from a set of similar species. Data from these sequences will be used to test predictions from the hypothesis that Transitions occur more frequently than do Transversions. You will end this Exercise by creating a Consensus sequence. AConsensus sequence is produced by examining the nucleotides at each position across the sequences in the data set. At each position the most "common" nucleotide observed is placed in the Consensus. In positions where several different nucleotides are found, new letters are used to represent the observed combinations. Letters used in Consensus sequences have been standardized by the Nomenclature Committee of the International Union of Biochemists. The letters are:

Symbol[2]	Description	Bases represented					Complement
A	Adenine	A				1	T
C	Cytosine		C				G
G	Guanine			G			C
T	Thymine				T		A
U	Uracil				U		A
W	Weak	A			T	2	W
S	Strong		C	G			S
M	aMino	A	C				K
K	Keto			G	T		M
R	puRine	A		G			Y
Y	pYrimidine		C		T		R
B	not A (**B** comes after A)		C	G	T	3	V
D	not C (**D** comes after C)	A		G	T		H
H	not G (**H** comes after G)	A	C		T		D
V	not T (**V** comes after T and U)	A	C	G			B
N	any Nucleotide (not a gap)	A	C	G	T	4	N
Z	Zero					0	Z

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 10 – Mutations” subdirectory and then the “Mutation Sequences” subdirectory.

7. Click on and OPEN the labeled “PopSet1.gb”

8. In BioEdit you will see that there are FOUR DNA sequences that are present:

   - Species A
   - Species B
   - Species C
   - Species D

   For simplicity, you may assume that the sequences have already gone through Transcriptional Unit editing (all introns have been removed and exons are spliced together). The sequences have been aligned and you will also notice the first three bases of each sequence at “atg”. All four seqences can considered to be “in frame” at this point so that Translation can proceed.

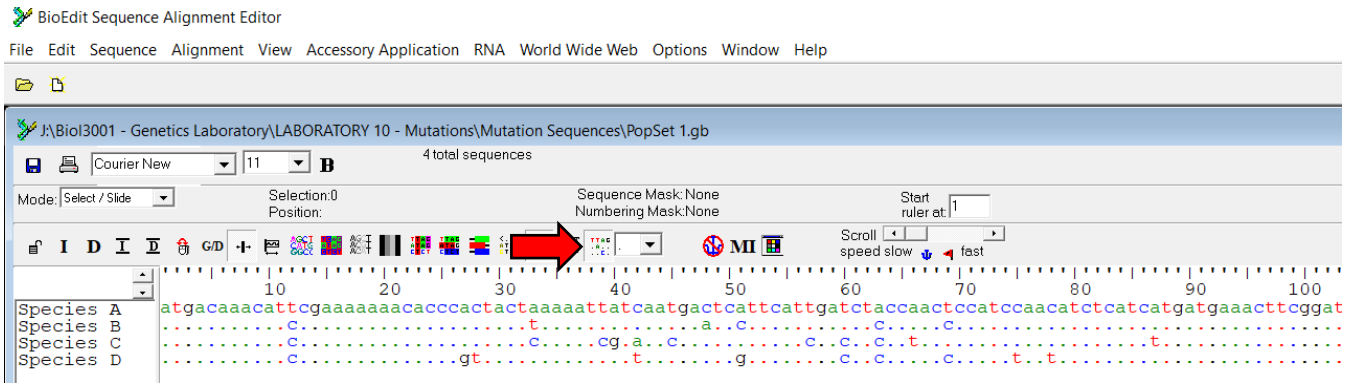
You will be testing TRANSVERSION / TRANSITION PREDICTION. Read the following hypothesis and predictions carefully to be sure you understand what questions you will be asking and answering with this Exercise.

HYPOTHESIS - Transition Mutations occur MORE FREQUENTLY than do Transversion Mutations.

NULL PREDICTION - Transition Mutations occur at the same frequency as to Transversion Mutations in the coding regions of a gene and the hypothesis is incorrect.

ALTERNATE PREDICTION - TransitionMutations occur do in fact occur more frequently than Transversion Mutations and the hypothesis is correct.

9. Select the DOT DISPLAY button to show where the differences occur among the sequences of the four species.



10. Generate an Overall Nucleotide Difference Matrix for your four species Sequences.

- Review the instructions from the previous Lab Exercise if needed.
- Highlight all four Species by holding down the SHIFT key and clicking on each Species.
- Select the ALIGNMENT tab on the tool bar at the top of the window.
- On the drodown menu list, Select SEQUENCE DIFFERENCE COUNT MATRIX.
- In the SAVE box that opens, you will need to identify a location to save your data. Open the dropdown box in the SAVE IN: box, and select YOUR PERSONAL H: DRIVE!! (*DO NOT save to the Mutations subdirectory in the BioEdit program. You will lose your information if you do so.*) You may place this matrix file in any subdirectory you choose on your H: Drive and name it whatever you like.
- Record the data in the matrix below:

NUCLEOTIDE DIFFERENCES

	Species A	Species B	Species C	Species D
Species A				
Species B				
Species C				
Species D				

11. Count the number of TRANSITION/TRANSVERSION changes and create a Difference Matrix for the four species.

- Follow the example in the Background and Introduction to this Exercise.
- Use the Nucleotide Counting Method from the Bioinformatics Laboratory.
- Using the aligned sequences in BioEdit, make a comparison between Species A and Species B.
    - Count the NUMBER OF TRANSITION differences and the NUMBER OF TRANSVERSION differences between the two sequences and record below.

Species A/Species B Transitions	
Species A/Species B Transversions	

    - Next, compare Species A and Species C. Count the NUMBER OF TRANSITION differences and the NUMBER OF TRANSVERSION differences between the two sequences and record below.

Species A/Species C Transitions	
Species A/Species C Transversions	

    - Last, compare Species A and Species D. Count the NUMBER OF TRANSITION differences and the NUMBER OF TRANSVERSION differences between the two sequences and record below.

Species A/Species D Transitions	
Species A/Species D Transversions	

- REPEAT the above steps and compare each sequence to the other ones. Record your observations below. NOTE – Making a comparison between Species A and Species B is the same as comparing Species B to Species A. Therefore, there is no need to count those numbers twice.

Species B/Species C Transitions	
Species B/Species C Transversions	
Species B/Species D Transitions	
Species B/Species D Transversions	
Species C/Species D Transitions	
Species C/Species D Transversions	

12. Generate an Difference Matrix for your four species Sequences for both the number of TRANSITIONS and the number of TRANSVERSIONS Record that information in the boxes below.

TRANSITIONS

	Species A	Species B	Species C	Species D
Species A				
Species B				
Species C				
Species D				

TRANSVERSIONS

	Species A	Species B	Species C	Species D
Species A				
Species B				
Species C				
Species D				

13. Create a CONSENSUS sequence between all four of your Species sequences. A consensus sequence is a nucleotide sequence of DNA, RNA, or an amino acid sequence of proteins that is generally used for inter- or intramolecular interactions and analysis. The consensus sequence will allow you to see what the "average" DNA sequence is between your species.

- Hold down the Shift key on your keyboard and Click on each of your four sequences to highlight their names.
- In the Toolbar at the top of the Window, Open the "Alignment" tab.
- Select and Click on the "Create Consensus Sequence" from the dropdown menu.
- You should see a new sequence show up in BioEdit labeled CONSENSUS consisting of upper case letters.

14. Analyze this sequence visually and see what you notice about it. You should notice that there are "NEW" letters aside from the normal A, T, C and G nucleotide identifiers. Use the table at the beginning of this Exercise to identify the meanings of those new letters. With this greater understand of Consensus Sequence Notation, you should be able to analyze your consensus sequence with a little better knowledge base. As an example, when you see the letter "Y" in a consensus sequence, it indicates that position is held by a PYRIMIDINE (either a C or a T). This is a way Bioinformatic Scientists can generate the "best fit" sequence when comparing multiple unknowns.

15. Taking into account your new found knowledge on how to look at a Consensus Sequence, review the Consensus sequence you generated in BioEdit. Do you see any patterns emerge? Does the Consensus sequence match one of the Species sequences "better" than the others? Are the differences you see between the Consesnsus Sequence and the Species sequences spread evenly across the sequence or are they grouped into certain locations? Give specific examples of what you notice in this comparison and then offer some ideas about WHY you may be seeing what you are seeing. Think about what you know about how genes work to help you answer these questions.

16. Let's return to our Hypothesis and Null Prediction.

HYPOTHESIS - Transition Mutations occur MORE FREQUENTLY than do Transversion Mutations.

NULL PREDICTION - Transition Mutations occur at the same frequency as to Transversion Mutations in the coding regions of a gene and the hypothesis is incorrect.

ALTERNATE PREDICTION - TransitionMutations occur do in fact occur more frequently than Transversion Mutations and the hypothesis is correct.

You are predicting that the number of Transition mutation will be the same as the number of Transversion mutation. You should be able to look at the matrices you generated above and have some idea of whether your prediction is correct. However, being a true scientist, you need more that qualitative evidence. You MUST quantitate your prediction. To test if your Observed data matces your Expected data, you will need to use the Chi-Square Analysis much the way your did with your Fruit Fly experiments earlier in the semester. This way you can QUANTITATIVELY support your Null Precition or your Alternate Prediction. Because your prediction is that Transitions = Transversions, your Expected value is 50% of your total number of mutations.

Calculate the Chi-Square to analyze your data.

Chi-Squared Test (enter values from F2 Generation Page, combine sexes to one phenotype)

Phenotype	Observed	Expected	O - E	$(O – E)^2$	$(O – E)^2 / E$
Transitions					
Transversions					
TOTAL					

Observed Chi – Squared Value	=	
Degrees of Freedom (*df*)	=	
Table Value (0.05)	=	
Overall Conclusion	=	--

Provide an explanation for the results you obtained from your Chi-Square. Briefly explain how the data supports or rejects your prediction.

17. Now that you have made some infereneces about mutations in this particular sequence across different Species, let's determine what gene you are actually looking at. Highlight your Consensus Sequence and copy it to the clipboard by opening the Edit tab and "Copy Sequences to clipboard (Fasta Format)"

18. Use your browser of choice and direct it to https://blast.ncbi.nlm.nih.gov/Blast.cgi. You will BLAST your Consensus sequence against the nucleotide database to determine the identity of your gene.

19. Open a Nucleotide BLAST query and Paste your Consensus Sequence in the "Query Sequence" box.

20. Since you do not know much information about the Species you obtained your sequences from, you will need to "widen" the parameters of your BLAST analysis.

- In the Choose Search Set query box, Click the "Others (nr, etc.)" option. As you do not know if your sequence is from Human or Mouse genomic information, you do not need to limit your search options at this point. You can always focus your search parameters later. It is always best to start with a wider net.
- In the Program Selection query box, Click on the "Somewhat similar sequences (blastn)" option. Again, you are submitting a Consensus Sequence. You will probably not get a hit using the default "Highly similar sequences" option especially if you have a number of differences within your Species.

21. Perform the BLAST analysis and determine the best match for your Consensus Sequence. Enter that information below. You may have to perform a quick Web based search to answer some of the questions.

Most Likely Genus of your species –

What is this Genus (common name of the organism)? -

Name of Gene/Protein –

Function of Protein Product –