**Instructions for DNA sequence analysis**

1. On the desktop open the applications folder and click on the “DNAStar” folder to open it
2. Find and Click to Open the Lasergene 8 folder
3. Find and click to Open the MegAlign Program
4. Go to “File” in the MegAlign toolbar at the top of the computer and click “Enter Sequences”
5. Look on the Desktop of the computer for the “Worm Sequences Folder “ and open it. Find the “dpy-17 from WB unspliced.seq” and the “WT dpy-17.seq” files in this folder.
6. Drag “dpy-17 from WB unspliced.seq” and the “WT dpy-17.seq” files into the large empty alignment box of the MegAlign screen OR, if you can’t drag and drop, highlight each of those sequences from the Worm Sequences Folder on the Desktop of the computer and click “Add files” from the MegaALign menu. If you see the sequences traveling across the screen, click “Done”.
7. Both sets of bases should be visible across the top of the MegaAlign alignment window
8. Put the mouse cursor on the horizontal bar that divides the MegAlign window and use the mouse to drag that bar all the way to the far right (end of the screen).
9. Click on “Align” in the MegAlign Toolbar and select “Clustal W Method”
10. The program will think for a minute and then, in the alignment window, you will see your two sequences aligned.
11. Blue color means either mismatches or no sequence
12. Red color indicates that the bases of your mutant’s sequence and the wild type sequence match
13. Use the blue scroll bubble at the bottom of the alignment window to scroll across to the right and look for blue boxes at the top of the window indicating where the sequences do not match
14. Keep in mind that the beginning and the end of a sequence usually has some “stutter” or mismatch. Sequencing runs are only reliable for ~800-900 bases at best. Do not be concerned about the mismatches at the beginning of the sequence or at the end when there are blocks of blue when the sequencing reliability has ended.
15. N stands for a base that the computer could not call. Note that you can sometimes be “smarter” than the computer. Always examine the chromatogram to see if you can determine what an N (uncalled) base should be in an area other than the beginning or the end of a sequence.
16. To examine the chromatogram for a sequenced file: Go back to the DNAStar folder in Applications of the computer and Lasergene8 and then Open SeqMan
17. Click on “Sequences” and “Add” from the SeqMan Toolbar at the top of your computer
18. Choose the .ab1 version of the file you are interested in (for this exercise choose “WT dpy-17.abi”
19. Click Add File and then Done
20. In the Sequence menu on the SeqMan Tool bar, click “Show Original Trace/Flowgram Diagram”
21. You will see a window appear with lots of different color “peaks” --each color represents a different base.
	1. Red = T
	2. Black = G
	3. Blue = C
	4. Green = A
22. Make the screen bigger by dragging on the bottom right corner so you can see the peaks better
23. You can use the + and – magnifying glasses on the left hand side of the screen to make the peaks easier to see.
24. Using the scroll bar bubble at the bottom of the window, scroll to the right until you get to base 30 where you find an area with many uncalled bases, “N”, that MegAlign marked with a blue box as a “misalignment” and see if you can determine what some of these base really are. You may not be able to do if there overlapping peaks or no clear peaks such as those found at base#30. Look at bases #32 and #33. There is a pretty clear blue peak C at 33 and a red peak T at 32. The computer couldn’t decide, but you could tentatively call these and see if that’s what the compared sequence has called them. If so you have tentatively called a match here rather than a misalignment and you can continue to search for real mismatches..
25. You should be able to determine that there are no real mismatches between the good quality part of the WT dpy-17 sequence and the Wormbase sequence (there shouldn’t be any).
26. Repeat this analysis beginning with step 6, but this time drag the “dpy-17 mutant.seq” file into the MegAlign window with your “dpy-17 from WB unspliced.seq” file and the WT dpy-17.seq file.
27. Once you have determined the significant change in the DNA in the mutant (compared to the WormBase unspliced sequence), you must also figure out why this change translates into a defective gene product (usually a protein). Remember the central dogma: DNA encodes RNA which, in turn, encodes protein. Function relates to protein and structure of the protein is controlled, ultimately, by DNA.
28. To study your protein product, Open a NEW window in MegAlign by clicking on File NEW and using your mouse to drag the central horizontal bar to the far right.
29. Click on “File” and “Enter Sequences”
30. Find the Worm Sequences Folder on the Desktop menu and highlight it. Click Add File (NOT ADD FOLDER) and you will see the contents of the Desktop folder appear in the Add to Project upper Window. Click once on the “Dpy-17 protein WT.pro” to highlight it and then Click on Add FILE. Do the same for the “dpy-17 mutant protein2.pro” file. When they have both appeared in the lower Add to Project Window, click on any empty space in the lower window and then click “Done”
31. The two proteins should be aligned in the MegAlign main window(the bar is red at the top) and you should be able to scroll across using the bubble at the bottom of the window to find where the change in the DNA caused a change in the protein. Notice that the numbers don’t match the numbers in the DNA sequence. Why not? How do you figure out how the number of a DNA base correlates with the number of the amino acid that is listed here? How many bases are in a codon?
32. What kind of change is this? Notice that the amino acids are listed by their one letter symbol rather than the older three letter code. To find the key to this code go to <http://www.ebi.ac.uk/2can/tutorials/aa.html>. A period indicates a stop codon UAA, UAG, or UGA. How is the change you have detected in the gene likely to have affected the protein in the mutant?
33. Now it’s time to find out the potential larger significance of the protein in species other than C. elegans. We need to do a BLAST to find out if this protein or this gene sequence (in its normal form) has known homologues and, if so, what is the function of those genes or gene products in other species.