Day 7 (project day 6) Making a protein alignment using CLUSTAL Omega

1. Use CLUSTAL Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) to make a multiple sequence alignment of all the protein sequences you selected and including your original *Drosophila* sequence. Remember to just use one sequence per organism, typically the best BLAST “hit”.
2. Paste the sequences into the Clustal omega box in FASTA format.
3. The default Clustal omega settings, including the output format as ClustalW with character counts, are good. It shouldn’t take too long, so email notification probably is not necessary. Click submit. Note: Input sequence formatting issues are the most common reason for a failure at this step!
4. You can click on “show colors” if you want similar types of amino acids to have the same color.
5. An example alignment is shown below (it is truncated, so it just shows the N-terminal portion)



1. I recommend copying the whole alignment and pasting it in to a Word file in landscape format. Make sure the font is Courier or Courier New because those fonts have the same spacing for all letters. Making the margins more narrow may also help it all fit better.

**Notes about the alignments:** I like the Clustal format because you get little symbols (\*, :, .) that show you conserved residues. The \* will be conserved in all sequences, the two dots (:) show amino acids that have about the same size and polarity in that position in all the sequences, and one dot (.) shows where either the size or the polarity has been preserved. Your alignment probably won’t have nearly as many \* as the one I put in above because I only had three relatively similar sequences that I aligned. I also truncated the alignment, the second half wasn’t as well conserved as the part I showed you!

**Another view that is informative is MView.** Explore MView by clicking on the “Send to MView” button on the page where the clustal omega alignment results are given.

1. Analyze your alignments. Do the whole sequences align well? Typically the ends (N and C termini) don’t align as well because they aren’t as conserved. Which regions are more conserved in your case?
2. Can you find key amino acids or sequence regions from the paper I recommended in your alignment? For example, cytochrome b561 sequences have conserved histidine residues. For the memo you need to have an annotated alignment, which means you highlight or otherwise denote known important amino acid residues and discuss what is known about their contribution to the protein’s function.
3. If the alignment in the paper I recommended doesn’t include your actual *Drosophila* sequence it may help to run an alignment with a sequence or two from that published alignment and your *Drosophila* sequence so you can find key amino acid residues. Once you have that information then you can apply it to the alignment for all eight insect sequences.