

## Comp Bio Project Part 4

(Due Wed, Nov 30<sup>th</sup> before 9am)

**Goal:** Attempt to find distant relatives for your family, analyze your 3D protein structure, practice your phylogenetic skills, and jump back into the virtual mass spectrometry laboratory experience

### Assignment:

1. First remove any sequences in your alignment that are obviously not related to your family (missing large sections, barely have any conservation, etc). Don't feel pressured to keep every paralog from the last project, especially the ones with very little conservation.
2. Perform psi-blast searches to try to uncover more distant relatives:
  - a) Create a psi-blast checkpoint file using a section of your alignment.
  - b) Use this checkpoint file to search phylum-specific databases (under "UniProt Taxonomic Subsets") at <http://www.ebi.ac.uk/Tools/sss/psiblast/>.
  - c) Search *at least 3* of these phylum for homologs:
    - In terms of evolutionary distance from humans = rodents - arthropods – nematode – fungi – viridiplantae - bacteria – archae - viruses
3. Select the three top matches from each of your psi-blast searches & add to your alignment
  - Don't worry if your matches don't align very well – just trying them out in alignment to confirm/reject whether they fit → fine to say: "these are definitely not in my family" in your discussion.
4. Locate your important region/residue from project part 3 on your 3D-structure (pdb) in VMD, and create a nice illustration highlighting this region/residue for your report.
5. Provide the species identity of your unknown sample from the Virtual Mass Spectrometry lab (VMSL)

### Required Report Elements (3)

1. A Microsoft Word document (.doc/.docx) containing:
  - a) All the information on your family you've gathered so far, and any comments/questions addressed from project part 3.
  - b) An illustration of your protein's 3D-structure with your region/residue highlighted
  - c) A discussion of the conservation of your region/residue in your new alignment, and where this region/residue is located in your 3D structure. Does the location of your region/residue on your structure jibe with what you know about its function?
  - d) The *identity of your unknown from the VMSL lab* (also don't forget to take the Post-Quiz afterward for 20 homework points)

2. A copy of your latest alignment (**.msf**) file
  - With the names edited to show species & SeqID
3. A copy of your latest Excel (**.xls/.xlsx**) file
  - Update your first sheet with new additions to your alignment.
  - Add additional sheets with the results of each new database search.