

**BIN 506 Proteins and DNA Sequence Analysis**  
**SPRING 2025 TAKE HOME Assignment**

The Take-Home Exam Presentations will be on **11<sup>th</sup> of June 9:40-12:30**

Reports due **20<sup>th</sup> of June 2025**.

- 1- It will be graded over 100 points. Answers to the questions have a potential of 90 points. A bonus of 10 points is reserved **for exceptional comments** that show your ability to use your overall knowledge to solve bioinformatics problems.
- 2- You are expected to submit your own work for each of the questions. Reviewing your classmates' applications and approaches is OK, but **your answers should stand out from each other.**
- 3- I will only answer questions submitted through the **ODTUCLASS forum so that** everyone will have access to the same explanations.
- 4- When possible, include screenshots or permanent links to your queries. Also, indicate the date of your query. Include a pdf or abstract of any literature you cite in your answers as supplementary data.
- 5- Some questions asking for your comment might not have a definitive answer. You are expected to build your discussions on the given information and available in the literature.
- 6- When submitting your results, **DO NOT EMAIL YOUR ANSWERS.**

**NO PRINTOUTS. UPLOAD** them to **ODTUCLASS** under **the TAKEHOME assignments**

1. Select a **HUMAN GENE** OF YOUR INTEREST WITH AN **ENTRY IN PDB** DATABASE (<https://www.rcsb.org/>) Introduce the general information available about “your gene” (5 points).
2. Find the genomic region of “your gene.” In which chromosomal band and clone or contig in the genomic assembly is your gene of interest located? Comment on the gene structure/model, possible mRNA and protein products, and their function? (15 points)
3. Comment about the supporting evidence for the annotation of the gene. **Is it a well-determined gene or not? Why?** Compare the experimental data and automated annotations. (15 points)
4. (20 points) Find out if any of the following species has proteins homologous to “your protein/gene of interest.” *M. musculus*, *G.gallus*, *O.lapites*, *A.mellifera*, *C.elegans*, *S.cerevisiae*.
  - a. Show the multiple sequence alignment and build a phylogenic tree using **only three of the homologous protein sequences**.
  - b. Do you see any clusters on the phylogenetic tree which are different from the tree of life?
  - c. Next, build the phylogenetic tree with all the sequences, describe any differences you have observed and discuss possible effects of adding more sequences to the phylogenetic analysis.
5. Find the protein sequence and which protein family or superfamily it belongs to. How many predicted or known proteins are there in the same family? Are there any conserved motifs, patterns, or domains in this family; mark these features on a multiple-sequence alignment commenting on functional annotations based on the sequence similarity. (15 points)
6. (20 points) Show the main secondary structures in the sequence.
  - a. Present and discuss the differences between the 3D structure of the protein from the PDB and AlphaFold prediction from the Uniprot database.
  - b. Label the regions with different colors with **helical or b-sheet** conformation.
  - c. **Also, identify the one great turn/loop/bend or break in the helix on the 3D structure. Label the amino acid residues mainly responsible for this structure and discuss their role**