Bioinformatics Practical Manual

Protein & DNA Sequence Analysis with Phylogenetic Tree Construction

1. Aim
To analyze protein and DNA sequences using bioinformatics tools for functional annotation, structural prediction, and evolutionary relationship determination through phylogenetic tree construction.

2. Learning Outcomes
- Retrieve DNA and protein sequences from databases (NCBI, UniProt)
- Perform sequence similarity searches (BLASTn, BLASTp)
- Align sequences using Clustal Omega or MUSCLE
- Predict domains, motifs, and protein properties
- Model protein 3D structures using SWISS-MODEL or AlphaFold
- Construct and interpret phylogenetic trees (MEGA X / Phylogeny.fr)

3. Requirements
- Internet access and computer with browser
- Tools: NCBI BLAST, Clustal Omega, InterPro, ProtParam, SWISS-MODEL/AlphaFold, MEGA X or Phylogeny.fr

4. Procedure

Step 1: Sequence Retrieval
Retrieve DNA sequence () and protein sequence () from NCBI or UniProt. Save in FASTA format.

Step 2: Similarity Search
Run BLASTn for DNA and BLASTp for protein. Record top hits with % identity, coverage, E-value, and function.

Step 3: Multiple Sequence Alignment (MSA)
Align selected sequences using Clustal Omega or MUSCLE. Highlight conserved regions.

Step 4: Functional Analysis
Use InterPro/Pfam to identify protein motifs and domains. Use ProtParam for protein properties.

Step 5: Structural Modeling (Protein)

Check AlphaFold DB or model with SWISS-MODEL. Save and visualize the structure.

Step 6: Phylogenetic Tree Construction

Construct a phylogenetic tree using aligned sequences (MEGA X or Phylogeny.fr).

5. Observations & Results

A. BLAST Results

clu

Accession	Organism	% Identity	Coverage	E-value	Function
NP_001853.2	*Homo sapiens*	99.22%	100%	3e-89	cytochrome c oxidase subunit 5B, mitochondrial precursor
XP_054331893.1	*Pongo pygmaeus*	96.90%	100%	2e-86	cytochrome c oxidase subunit 5B, mitochondrial
KAF6429150.1	*Molossus molossus*	84.50%	100%	2e-78	hypothetical protein HJG59_003243
XP_055981796.1	*Sorex fumeus*	82.95%	100%	6e-78	cytochrome c oxidase subunit 5B, mitochondrial isoform X1

B. InterPro / Pfam Domains

Domain	Source DB	Residue Range	Function
Cytochrome c oxidase subunit Vb, copper-binding domain	Pfam	66-125	This domain is part of the final enzyme in the electron transport chain and is involved in binding a copper ion.

C. ProtParam Properties

Property	Value
Molecular Weight	13695.69
Theoretical pI	9.07
Instability Index	43.06 (This classifies the protein as unstable)
Aliphatic Index	78.06
GRAVY	-0.298

D. Phylogenetic Tree Notes

Closest Relatives	Evolutionary Insight
As you can see, the human (NP_001853.2) and the orangutan (XP_054331893.1) are paired together on the	The bat (KAF6429150.1) and the shrew (XP_055981796.1) are on more distant branches, showing they are less

	closest branches. This indicates they are the most closely related in this group, which makes sense as they are both primates.	related to the primates and to each other. This is exactly what we would expect to see.

E. Figures / Screenshots
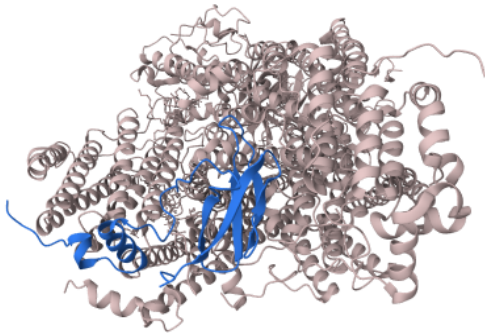
Insert screenshots of:

- BLAST results

	Cluster Composition Click the ⊞ to see the cluster contents	Cluster Ancestor	Cluster Representative Sequence	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	⊞ 13 member(s), 8 organism(s)	human	cytochrome c oxidase subunit 5B, mitochondrial precursor [...	265	265	100%	3e-89	99.22%	129	NP_001853.2
☑	⊞ 6 member(s), 6 organism(s)	Bornean orangutan	cytochrome c oxidase subunit 5B, mitochondrial [Pongo pyg...	259	259	100%	2e-86	96.90%	181	XP_054331893.1
☑	⊞ 1 member(s), 1 organism(s)	Pallas's mastiff bat	hypothetical protein HJG59_003243 [Molossus molossus]	239	239	100%	2e-78	84.50%	186	KAF6429150.1
☑	⊞ 3 member(s), 3 organism(s)	smoky shrew	cytochrome c oxidase subunit 5B, mitochondrial isoform X1 ...	236	236	100%	6e-78	82.95%	129	XP_055981796.1

- Multiple Sequence Alignment

```
NP_001853.2       ----------------------------------------------------------MA   2
XP_054331893.1    ------MRSFRPRHSLGSGRARGLGGSGKLTERRKKPAPRSPACLCSLFPEVLLVAEAMA   54
KAF6429150.1      MRILFRLRAFTGRVRRGTG-GSGAGGSRQLTPRRKQRSPEVPPISGACHGSAAVVAGVMA   59
XP_055981796.1    ----------------------------------------------------------MA   2
                                                                            **

NP_001853.2       SRLLRGAGTLAAQALRARGPSGAAAMRSMASGGGVPTDEEQATGLEREIMLAAKKGLDPY   62
XP_054331893.1    SRLLRGAGALAAQALRARGPSGAAAVRSMASGGGVPTDEEQATGLEREIMLAAKKGLDPY   114
KAF6429150.1      SRLLRGAGALVAQALRARGPTGAAAVRSMASGGGVPTDDDQATGLEREVMLAARKGLDPY   119
XP_055981796.1    SRLLRGAGALFAQALRTRGPNGVAAVRSMASAGGIPTDDEQATGLEREIMMAARKGLDPY   62
                  ********:* *****:***.*.**:*****.**:***::********:*:**:******

NP_001853.2       NVLAPKGASGTREDPNLVPSISNKRIVGCICEEDNTSVVWFWLHKGEAQRCPRCGAHYKL   122
XP_054331893.1    NVLAPRGASGTREDPNLVPSISNKRIVGCICEEDNTSVVWFWLHKGEAQRCPRCGAHYKL   174
KAF6429150.1      NMLAPKAASGTKEDPNLVPSITNKRIVGCICEEDNSAVIWFWLHKGETQRCPSCGTHYKL   179
XP_055981796.1    NLLAPKAASGTKEDPNLVPSISNKRIVGCICEEDNSTVIWFWLHKGEAQRCPSCGTHYKL   122
                  *:***:.****:*********:**********::*:*********:**** **:****

NP_001853.2       VPQQLAH       129
XP_054331893.1    VPQQLAH       181
KAF6429150.1      VPHQLAH       186
XP_055981796.1    VPHQLSH       129
                  **:**:*
```
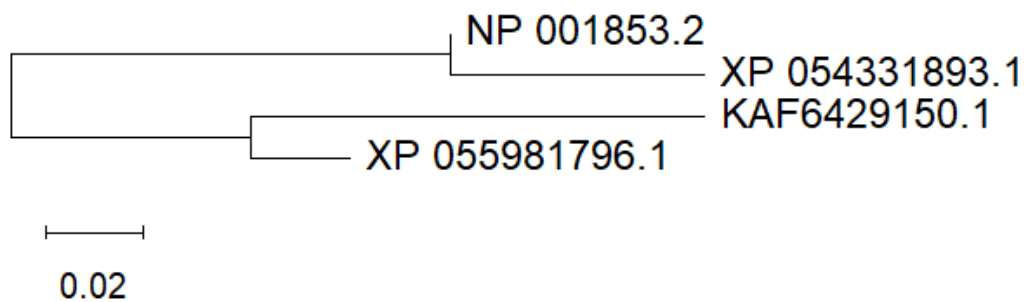
- Protein structure model

- Phylogenetic tree

6. Conclusion

This analysis investigated the protein sequence **AAA52061.1**, the cytochrome c oxidase precursor from *Homo sapiens*. A **BLASTp** search confirmed its identity, revealing high similarity to cytochrome c oxidase subunit 5B proteins in other mammals, including the orangutan (*Pongo pygmaeus*). Functional analysis using **InterPro** identified a key "Cytochrome c oxidase subunit Vb, copper-binding domain," confirming its role in cellular respiration. Physicochemical analysis via **ProtParam** determined the protein has a molecular weight of approximately 13.7 kDa and is classified as unstable. A 3D structural model was obtained from the AlphaFold database. Finally, a **multiple sequence alignment** showed high conservation across related species, and the **phylogenetic tree** constructed using MEGA X demonstrated evolutionary relationships consistent with known biology, grouping *Homo sapiens* most closely with *Pongo pygmaeus*. In summary, this series of bioinformatic analyses successfully characterized the protein's function, physicochemical properties, structure, and evolutionary context.