Project 1 Documentation: Sequence Alignment using NCBI BLAST

Overview

This project demonstrates how to perform sequence alignment using the BLAST (Basic Local Alignment Search Tool) interface provided by NCBI, with the help of Biopython. It is divided into two parts:

- Part 1: Nucleotide sequence alignment using BLASTn

- Part 2: Protein sequence alignment using BLASTp

Requirements

- Python 3.x

- Biopython

- Internet access for querying NCBI BLAST services

- Python notebook environment (Kaggle, Jupyter, or Google Colab)

Part 1: Nucleotide BLAST (BLASTn)

Objective:

To compare the TP53 gene sequence against NCBI's nucleotide database (nt) using BLASTn.

Dataset:

- Source: https://www.ncbi.nlm.nih.gov/gene/7157

Workflow:

1. Load the TP53 gene sequence from the FASTA file.

2. Submit to NCBI BLASTn using qblast.

3. Parse results with SearchIO.

4. Display top hit (ID, description, E-value, alignment).

Part 2: Protein BLAST (BLASTp)

Objective:

To compare the TP53 protein sequence against NCBI's protein database (pdb) using BLASTp.

Dataset:

- Source: https://www.uniprot.org/uniprotkb/P04637/entry

- FASTA: https://rest.uniprot.org/uniprotkb/P04637.fasta

Workflow:

1. Load TP53 protein sequence.

2. Submit to NCBI BLASTp using qblast.

3. Parse with SearchIO.

4. Display top hit (chain ID, description, E-value, alignment).

Key Takeaways

- BLASTn and BLASTp are used to find sequence homology.

- Biopython simplifies accessing and parsing NCBI BLAST.

- Alignments help identify related genes or proteins across organisms.

References

- NCBI BLAST: https://blast.ncbi.nlm.nih.gov/Blast.cgi

- Biopython: https://biopython.org/wiki/Documentation

- TP53 Gene: https://www.ncbi.nlm.nih.gov/gene/7157

- TP53 Protein: https://www.uniprot.org/uniprotkb/P04637/entry