

Mini Project 2 - Genomic Data Analysis

Learning Objectives

1. Understanding DNA & RNA sequences
2. Basic sequence operations: transcription (DNA \rightarrow RNA), reverse complement, and GC content calculation
3. What a Single Nucleotide Polymorphism is and how to find one by comparing reference and mutated sequences
4. Visualize GC content to compare sequences

Project Breakdown

1. Understanding DNA & RNA sequences
 - a. Show DNA and RNA sequences side by side
 - b. Highlight main differences (T \rightarrow U)
2. With chosen example CFTR (gene that leads to Cystic Fibrosis):
 - a. Transcription (DNA \rightarrow RNA)
 - b. Reverse complement
 - c. GC content calculation

```
def gc_content(seq):  
    gc_count = seq.count("G") + seq.count("C")  
    return (gc_count / len(seq)) * 100  
  
print("GC Content:", gc_content(dna_seq), "%")
```

- i.
3. Explain what an SNP (Single Nucleotide Polymorphism): a single base change in DNA compared to a reference sequence
4. Compare CFTR gene to mutated gene with a missense mutation
 - a. Code for checking each nucleotide
 - b. Code for translating each sequence to proteins \rightarrow check if there is a difference
 - c. Visualize the GC content via heatmap
5. Reflection
 - I. Small 3 question survey
 - II. Difficulty of the module
 - III. Accessibility/understandability of module
 - IV. Any points of improvement

Deliverables

1. Repeat steps 2-4 with given gene: Hemoglobin Gene (HBA1) which has a silent mutation
 - a. Transcription
 - b. Reverse Complement
 - c. GC content
 - d. Silent mutation - main check
 - e. GC heatmap

https://colab.research.google.com/drive/1vEAJvfD3r14_xXwYrPmSpVSL-3HvN9R7?authuser=1#scrollTo=dJOgT1BD71PH

Mini Project 3 - Exploring Protein Structures

Learning Objectives

- Understand how to obtain protein sequences from NCBI database
- Learn how to use alpha fold to predict protein structure
- Compare protein structures to experimentally generated ones
- Identify structural features and relate them to protein function

Project Breakdown

1. Introduction & Background
 - a. Explain what protein structure is
 - b. Importance of protein structure in function
 - c. Overview of the databases that store protein structure
2. Task (w/ tutorial + example)
 - a. Choose a gene - will provide a list of genes and their importance
 - b. Get the sequence - write up/demonstrate how to retrieve the sequence from NCBI
3. Analysis (w/ tutorial + example)
 - a. Take that gene and search NCBI for the structures of that protein
 - b. Take the sequence from NCBI for that protein and output it into Alphafold
4. Compare structures (w/ general guidance + example)
 - a. Write a paragraph explaining if there are any differences between the structures include
 - i. Visual differences
 - ii. Can the function change due to that
 - iii. Check for quantitative differences as well
5. Reflection
 - a. Small 3 question survey
 - i. Difficulty of the module
 - ii. Accessibility/understandability of module
 - iii. Any points of improvement

Deliverables

- Short 1 page report that includes
 - The protein chosen
 - The sequence of the protein

- Screenshots of both the NCBI structure and the alpha fold structure
- Observations and insights
- Filled out survey

Task List for Development

- Write up the introduction and background
- Create a document of the genes and a tutorial on how to access the sequences from NCBI
- Create a tutorial + document of how to search NCBI and how to use alpha fold
- Create a small analysis example for students to reference
- Tutorial
 - Pick a gene different from the genes given to students
 - Run through the task with the example gene
 - Show how to run a small analysis on the gene

Mini Project 4 - Unix commands for Bioinformatics

Learning Objectives

- Learn basic unix commands
- Learn how BLAST is used in the command line
- Basic example of how genes from different species are compared to get students motivated and involved in the learning process
- Showcase how performing BLAST locally on your computer can be faster than using the NCBI online BLAST tool inspiring them to gain more computational knowledge through workshops available on campus

Project Breakdown

1. Introduction & Background

- Explain how BLAST can be used locally
- Explain how Unix allows us to use BLAST locally
- Explain a few basic UNIX commands to get them started
 - New Command - How to use it - DIY
- Game for learning terminal commands: Terminus
- Direct the students to workshops where they can learn more about Unix

2. Task w/ Tutorial

Walk through :

- Pick 2 Protein
 - Compare Spike protein between Wuhan & Omicron
- Run Blast Locally
 - `!blastn -query query.fasta -subject subject.fasta -outfmt 6 -out result.txt`
- Visualize/Summarize the similarities and difference
- Then separately the student will run through I to III on their own with the genes of their choice (*Mitochondrial DNA* across species)

3 . Analysis

Questions:

How similar are these genes?

Where are the differences?

4. Reflection

- V. Small 3 question survey
- VI. Difficulty of the module 1 - 5 Stars
- VII. Accessibility/understandability of module
- VIII. Any points of improvement

https://colab.research.google.com/drive/13EM_sDJXh1xyET9XSdMx4e80jQWENHJR

1. Introduction + BLAST + Mini Project 1
2. Genomic Data Analysis + Mini-Project 2 (Sovie)
 - a. NCBI - sequence
 - b. Visualize
 - c. Compare two heat maps
3. Protein Structure & Function + Mini-Project 3 (Gunica)

Databases

How proteins are made

NCBI

- a. Galaxy
 - b. Alpha fold
4. Unix commands for Bioinformatics + Mini-Project 4 (Brenda)
 - a. Commands
 - b. Apply to bioinformatics

<https://colab.research.google.com/drive/1pDXbUWtY80UeC6q7-u8yA2mFP99eSPue?usp=sharing>

- Annotate code
- Game for learning terminal commands for mac: [Terminus](#)