

# Sequence X Sequence

Project presentation

# The Idea

Bio Informatics :

is an interdisciplinary field that develops methods and software tools for understanding biological data. As an interdisciplinary field of science, bioinformatics combines Computer Science, Statistics, mathematics and engineering to study and process biological data.

Computers became essential in molecular biology when Protein sequences became available after Fredrick Sanger determined the sequence of Insulin in the early 1950s.

Comparing multiple sequences manually turned out to be Impractical.

A comparison of genes within a species or between different species can show similarities between protein functions, or relations between species

- Some times large parts of the DNA do not serve any obvious
- purpose.
- This so-called junk DNA may, however, contain unrecognized
- functional elements.
- Comparing sequences can help determine similarities and
- differences, between two different species or a two creatures
- of the same species.
- That can offer great help in evolution theory and other related theories in BIO

# Sequence X Sequence

It is a software application that compines BioInformatics techniques with Python coding and libraries to mainly perform acomparison between two DNA sequences to find out the percentage of similarity between them to help in various BIO related topics.

The idea is to help scientists discover the percentage of similarity between two DNA sequences in a much easier way through a desktop application that can save them a lot of time and efforts and make their life easier .

# How it works

You can enter sequences in 3 different ways

- 1- Direct access
- 2- Uploading your sequence file
- 3- Entering the sequence ID number

DNA sequence consists of 4 letters [A, G, T, C]

The four letters are divided into two groups: [A,G]  
[T,C]

the app will compare every two letters of the two  
sequence in its turn .

If it finds the same letter it will give a result of (10)

if it finds different letter :

- from the same group, it will give a (0)
- from different groups, it will give a (-5)

- When you hit calculate a box will appear containing results
- the two sequences under each other, with a counter
- under them counting the elements, the score of the two
- elements and the percentage of similarity according to the
- number of sequences



Sequence X Sequence

✕

SEQUENCE  
X  
SEQUENCE

☒ FASTA

☐ GenBank

First Sequence

Choose File

File Path

ID Number

Second Sequence

Choose File

File Path

ID Number

Exit

Calculate



# Challenges

BioPython:

BioPython is a library dedicated to handle biological information through NCBI, it can help manipulate biological information in the form we need, we used it to make the application read sequences only from the files we upload, and to save the sequences it reads from the data base in a seperated file in the application

# Future Work

It's not hard but we can expand the application to read all kinds of sequences, Proteins, DNA or whatever .

Help scientists to manipulate sequences more by removing or adding elements to the sequences .

We could try to add database of sequences to remove online access from the app and make it easier for users