# SWE

An application that applies sets of modules to nucleotide and protein sequences using the biopython programming language. NCBI sequence retrieval, BLAST, MSA, Phylogenetic tree, and FASTA statistics are the five modules that make up our project. The objective is to streamline the procedure for retrieving sequences from NCBI and applying several technologies, including pairwise and multiple alignment, to them in a single application.

## Technologies Used

- VS code

## Setup

You have to allocate the application directory along with the documents he’s going to work on. Mucscle.exe and clustalw.exe must be downloaded and document its path in the code. Good internet connection is required for the usage of Entrez library. Once the directory and the exe all installed, the application is ready to work.

## Usage

1. NCBI: you should have a protein or nucleotide ID and seek to download its FASTA file on your pc. Once you entered either a single or multiple IDs and by clicking the database type, the file is automatically generated in your working directory along with details about the sequence that appears in the GUI.
2. BLAST: you have two sequences as text needs to be aligned globally, so you enter them in their textbox and then by clicking on the BLAST! button, the output appears in its textbox after being aligned in addition to the score listed.
3. MSA: here you could use the same FASTA file you generated from the NCBI module that consists of multiple of sequences, either browse the file or write it down in the textbox. The MSA is generated automatically, and you could see along with it the details about the conserved residues, gaps, strongly and weakly similar residues. There’s also an option to view it in HTML view by clicking on the button and choose the FASTA file to be aligned.
4. Phylogenetic tree: we applied two algorithms; UPGMA and NEWICK. Either of the two algorithms, you should enter your FASTA file that contains the multiple sequences and the choose your preferred algorithm. A figure then appears with the tree with a feature of searching and navigating through it.
5. FASTA statistics: In this module you only enter one sequence FASTA file by browsing through your files then the result appears with the statistics of the specific sequence right away.