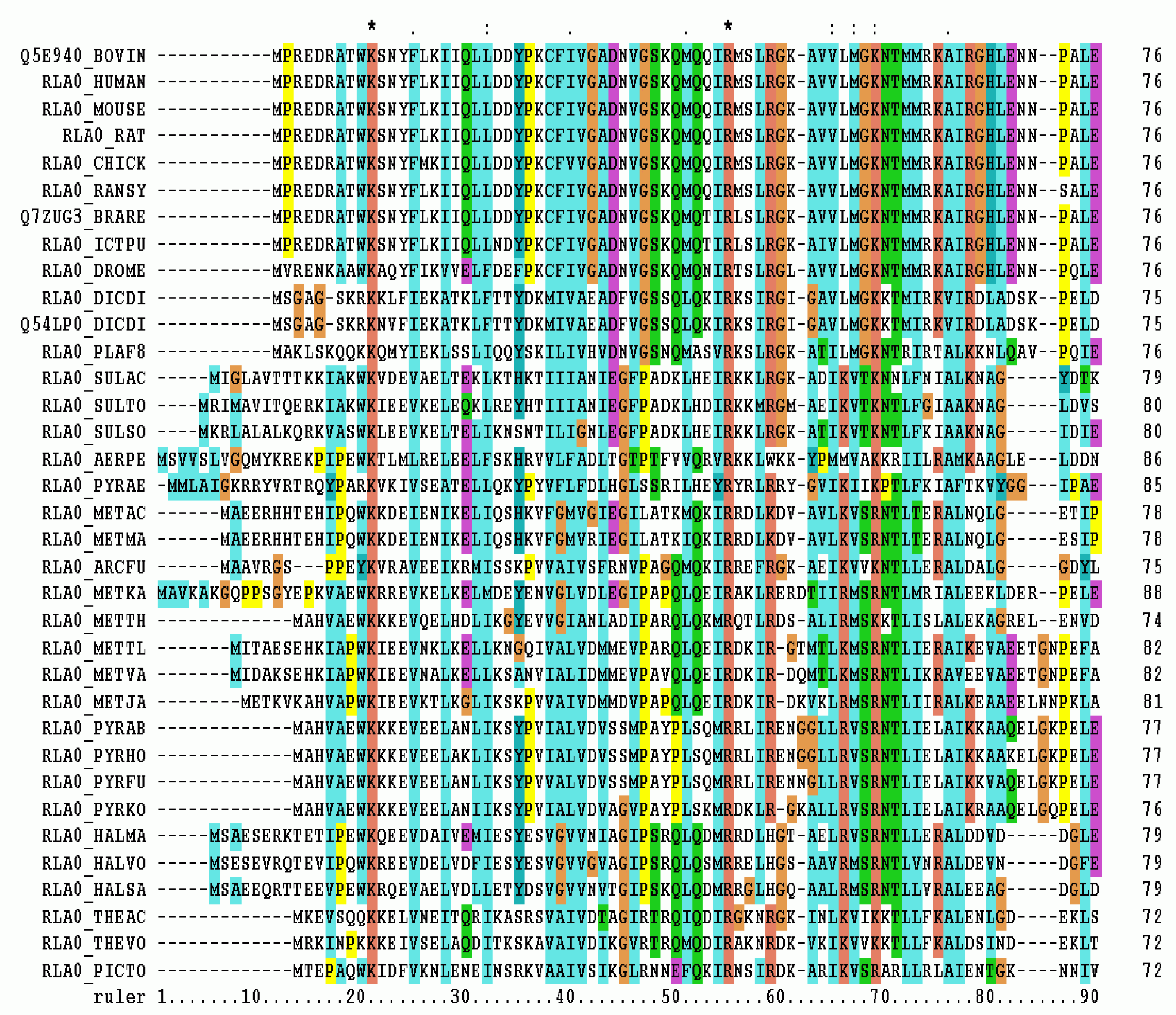


Multiple Sequence Alignments

BIOINFORMATICS

ARADHANA KUMARI

BIOINFORMATICS

Semester 2nd

TOPIC

Identification of Conserved Regions in Protein Sequences Through Multiple Sequence Alignments.

INDEX

Sr No.	CONTENT	Page No
1	**Introduction to Multiple Sequence Alignment (MSA)**	1-2
2	**Types and Tools Involved in MSA**	3-4
3	**Steps of MSA**	5-19
4	**Application of MSA**	20
5	**Conclusion**	21
6	**Reference**	22

Introduction

Proteins are fundamental macromolecules that perform a vast array of functions essential for life, including catalysis, structural support, signalling, and transport. The sequence of amino acids in a protein dictates its three-dimensional structure and, consequently, its function. **Across evolution, certain segments of protein sequences remain remarkably unchanged**, even among distantly related species. These conserved regions are critical to understanding protein function, stability, and evolutionary relationships.

Multiple Sequence Alignment (MSA) is a cornerstone technique in bioinformatics that enables researchers to compare and align three or more biological sequences simultaneously. By arranging sequences such that homologous residues are positioned in vertical columns, MSA reveals patterns of conservation and variation. These alignments are instrumental in identifying conserved regions, which often correspond to functional domains, active sites, or structural motifs essential for protein activity.

What A Multiple Sequence Alignment mean?
In a Multiple sequence alignment, **homologous** residues among a set of sequences are aligned together in columns. '**Homologous**' is meant in both the structural and evolutionary sense.

The significance of conserved regions extends beyond mere sequence similarity. They often represent evolutionary constraints—areas where mutations are deleterious and thus eliminated by natural selection**. As noted by Mount (2004), conserved regions are key to inferring functional importance and evolutionary relationships.** In structural biology, conserved residues frequently contribute to the protein's core or active site, maintaining structural integrity and catalytic efficiency (**Lesk, 2017**).

From an evolutionary perspective, conserved regions serve as molecular fossils, providing insights into ancestral sequences and speciation events. Phylogenetic analyses based on these regions help reconstruct evolutionary trees, elucidating relationships between species and protein families. **The work of Felsenstein (2004) underscores the power of sequence alignment in evolutionary biology, enabling hypotheses about common ancestry and functional divergence.**

This project employs MSA to identify and analyse conserved regions in a set of protein sequences, aiming to bridge sequence analysis with functional and evolutionary insights. By integrating bioinformatics tools, this study seeks to demonstrate how conserved regions inform our understanding of protein biology, from molecular function to evolutionary history.

Types

1. **Progressive Alignment Methods**

   Probably the most commonly used approach to multiple sequence alignment is progressive alignment. This works by constructing a succession of pairwise alignments. Initially, two sequences are chosen and aligned by standard pairwise alignment; this alignment is fixed. Then, a third sequence is chosen and aligned to the first alignment, and this process is iterated until all sequences have been aligned. Classic example: CLUSTAL.

2. **ITERATIVE Refinement Methods**

   One problem with progressive alignment algorithms is that the subalignments are 'frozen'. That is, once a group of sequences has been aligned, their alignment to each other cannot be changed at a later stage as more data arrive. Iterative refinement algorithms attempt to circumvent this problem.

Tools Involved In MSA

Clustal Omega

New MSA tool that uses seeded guide trees and HMM profile-profile techniques to generate alignments. Suitable for medium-large alignments.

Launch Clustal Omega

EMBOSS Cons

EMBOSS Cons creates a consensus sequence from a protein or nucleotide multiple alignment

Launch EMBOSS Cons

Kalign

Very fast MSA tool that concentrates on local regions. Suitable for large alignments.

Launch Kalign

MAFFT

MSA tool that uses Fast Fourier Transforms. Suitable for medium-large alignments.

Launch MAFFT

MUSCLE

Accurate MSA tool, especially good with proteins. Suitable for medium alignments.

Launch MUSCLE

MUSCLE 5

Muscle v5 is an extensive re-write of the MUSCLE code based on new algorithms. Suitable for medium-large alignments.

Launch MUSCLE 5

MView

Transform a Sequence Similarity Search result into a Multiple Sequence Alignment or reformat a Multiple Sequence Alignment using the MView program.

Launch MView

T-CoffeE

Consistency-based MSA tool that attempts to mitigate the pitfalls of progressive alignment methods. Suitable for small alignments.

Launch T-Coffee

WebPRANK

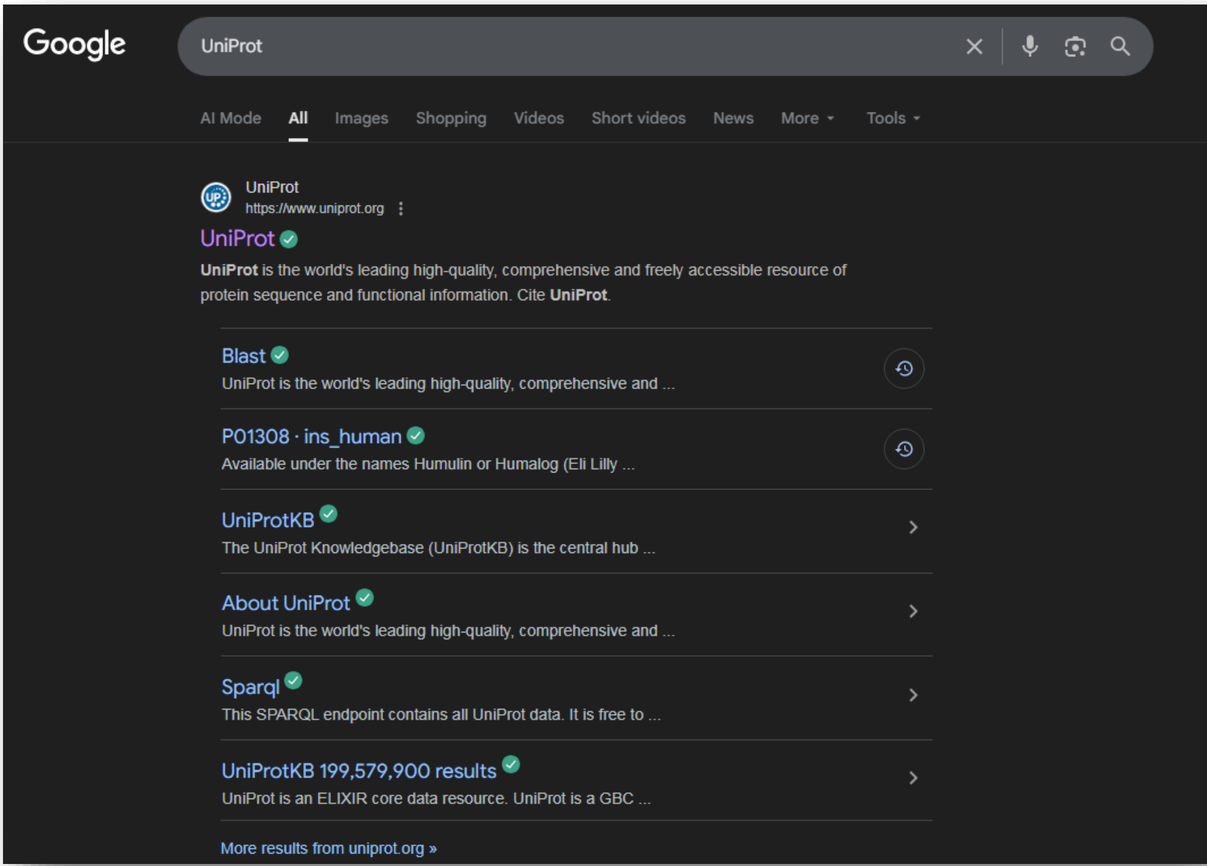
The EBI has a new phylogeny-aware multiple sequence alignment program which makes use of evolutionary information to help place insertions and deletions.

Launch WebPRANK

https://github.com/code-aradhana/bioinformatics-msa-project.git

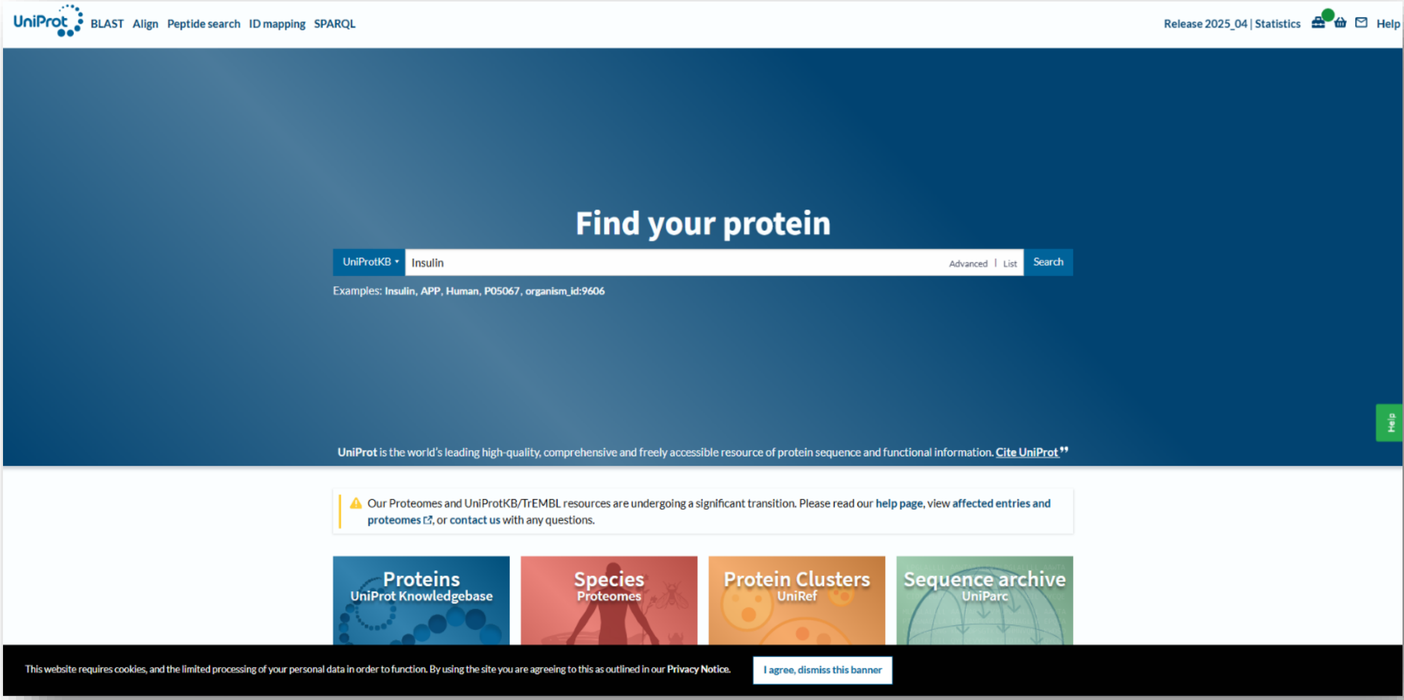
Steps In MSA

1. Search **UniPort** On Google and open it.

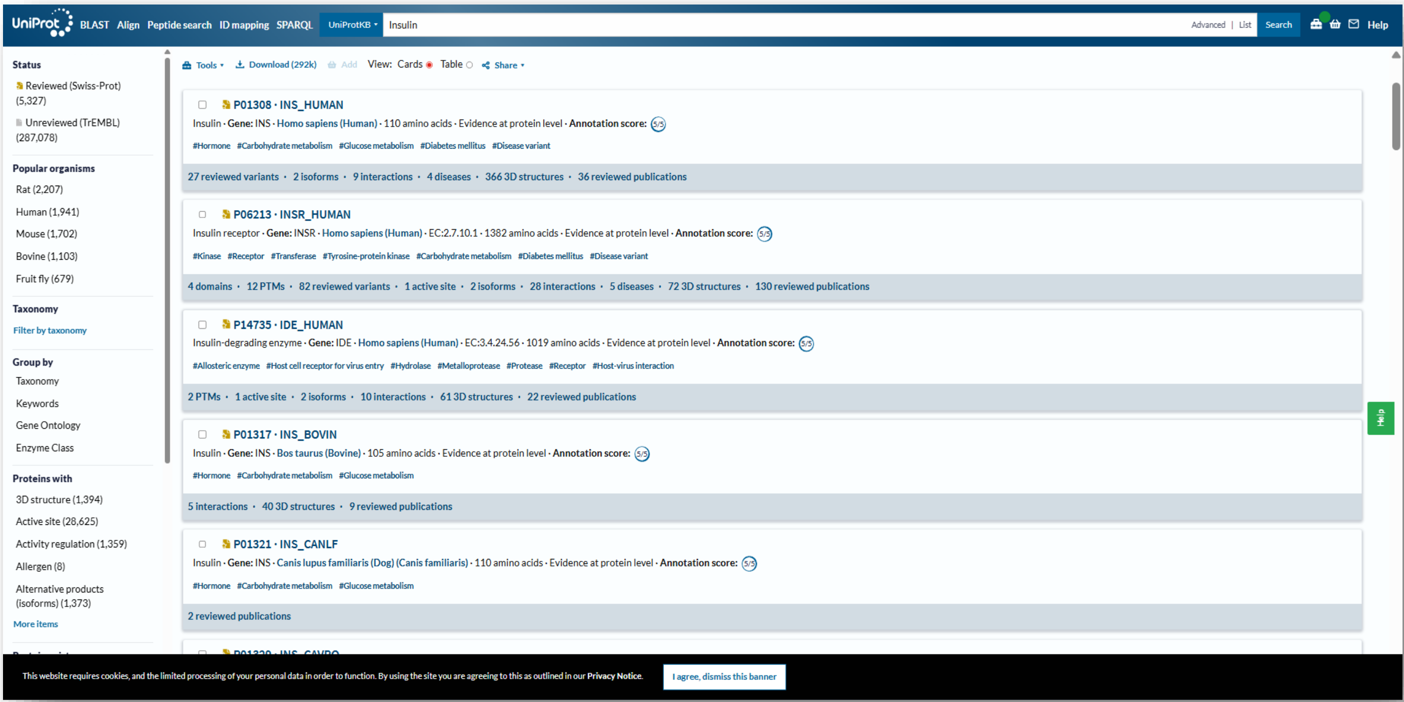


https://github.com/code-aradhana/bioinformatics-msa-project.git

2. Enter a **Query sequence** in search box.

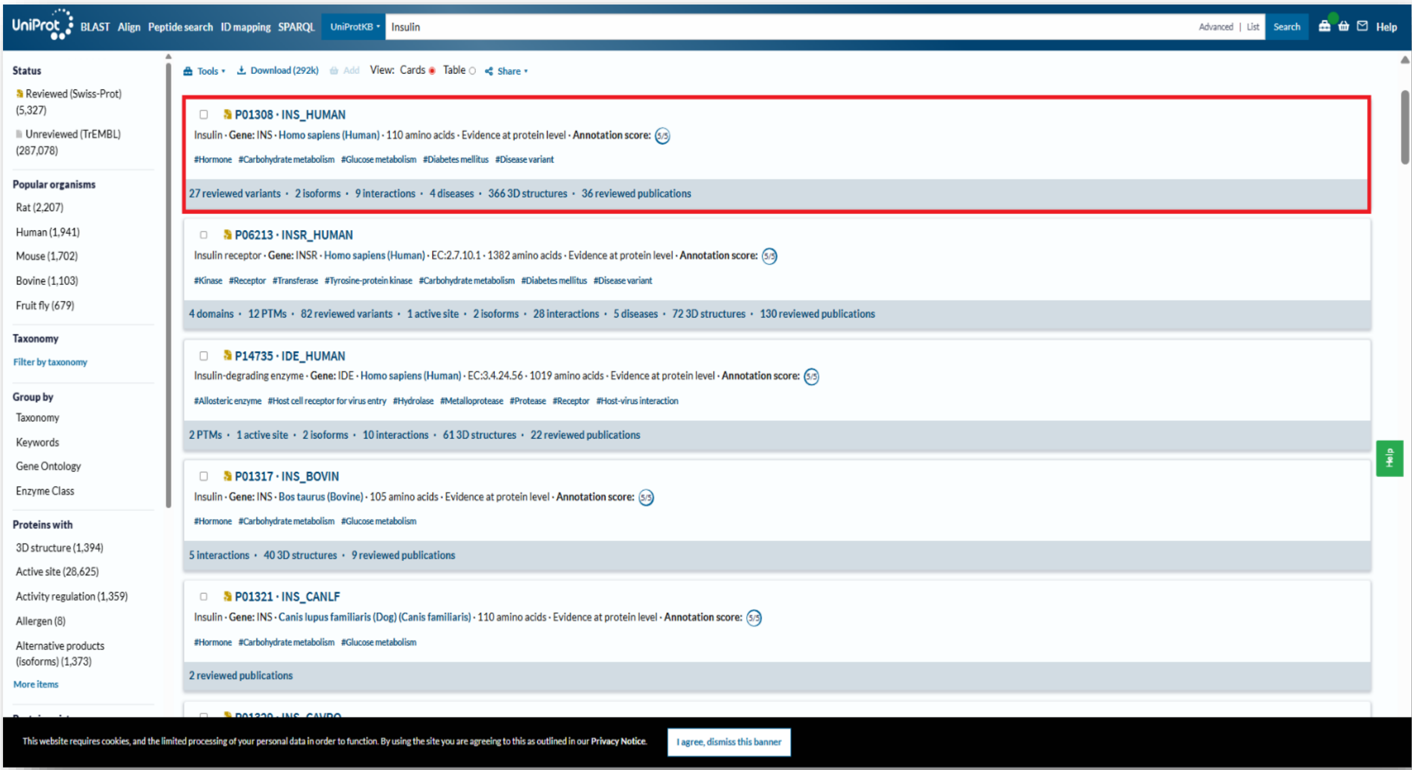


3. **Results** will appear in this format.

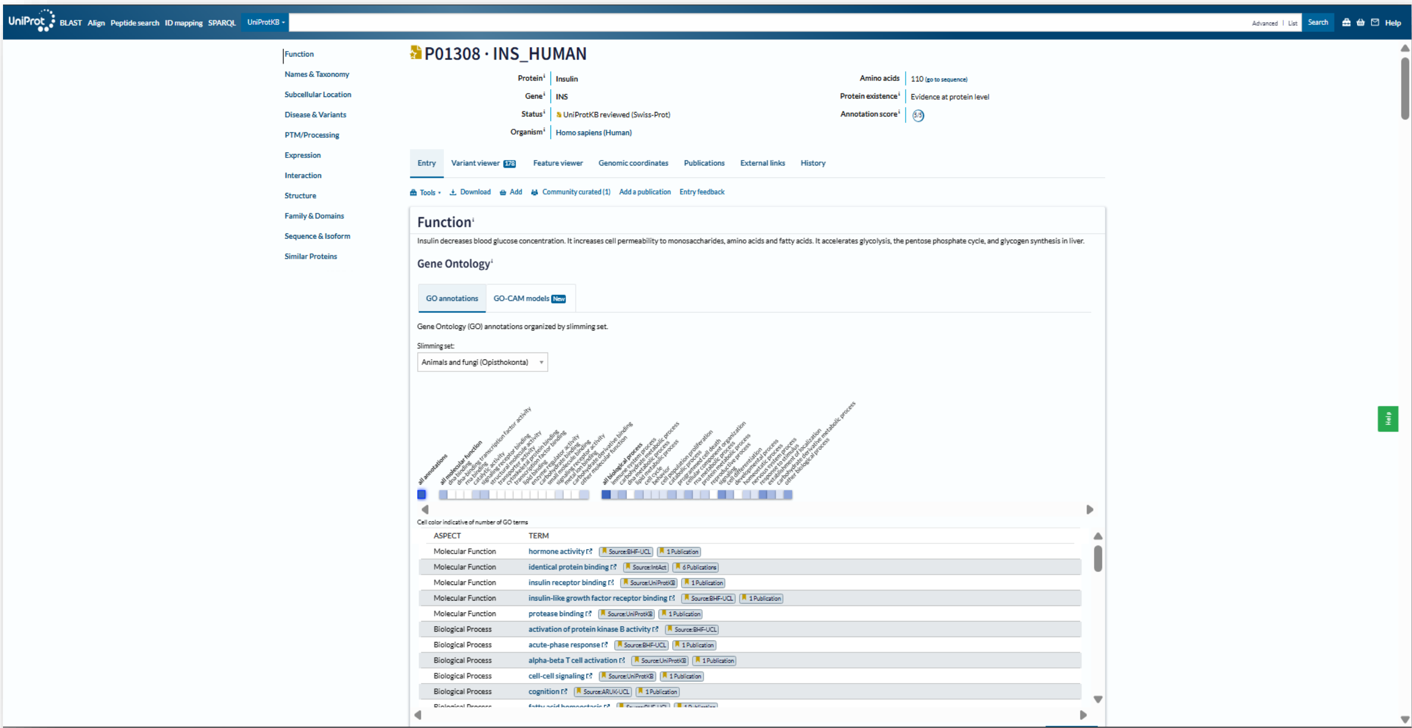


https://github.com/code-aradhana/bioinformatics-msa-project.git

4. Select the **required item** from the list.

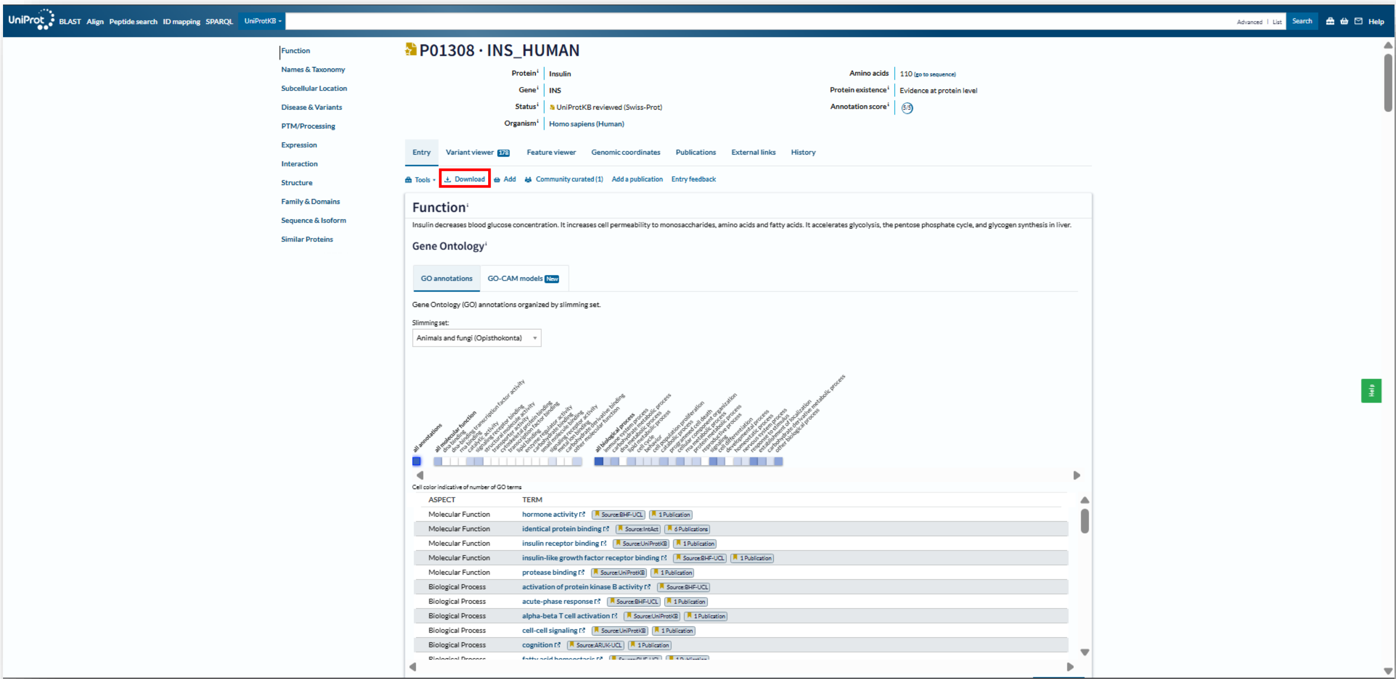


5. Page will appear like this.

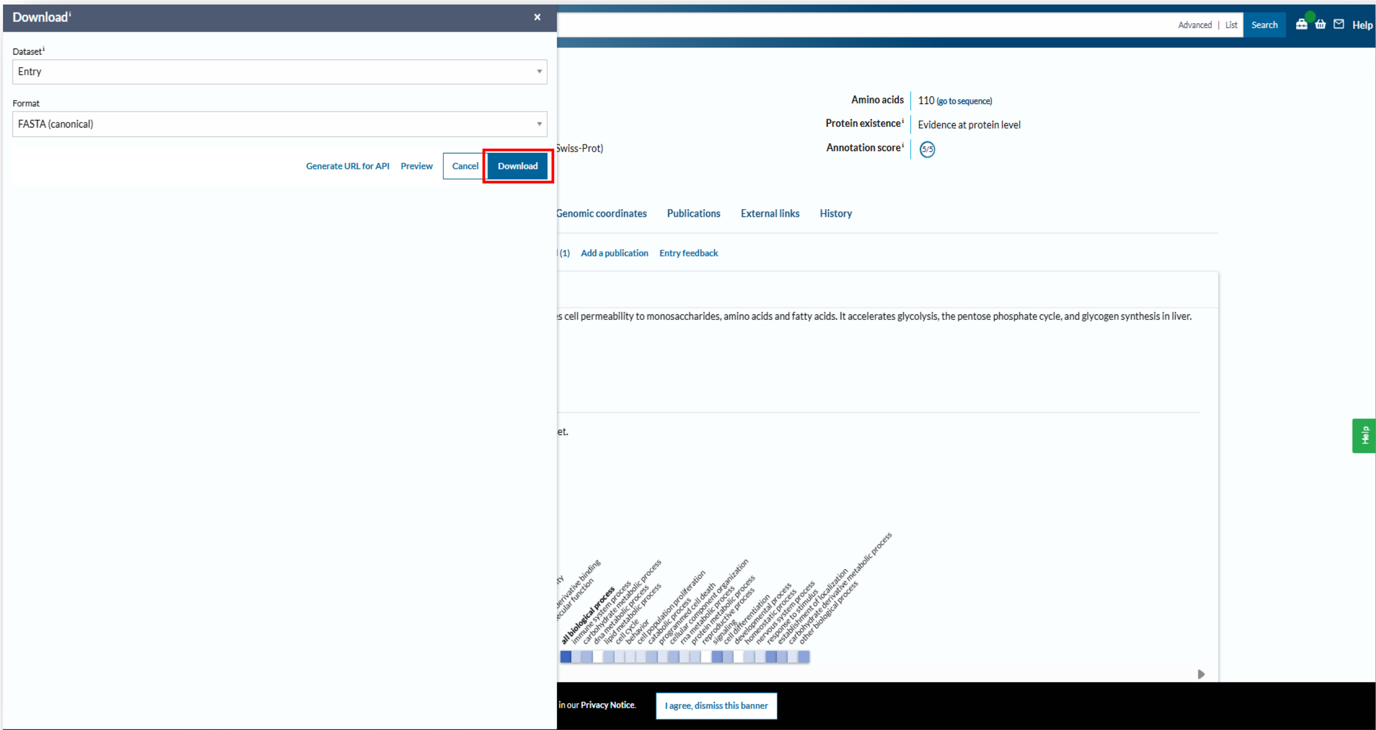


https://github.com/code-aradhana/bioinformatics-msa-project.git

6. Click on **Download.**

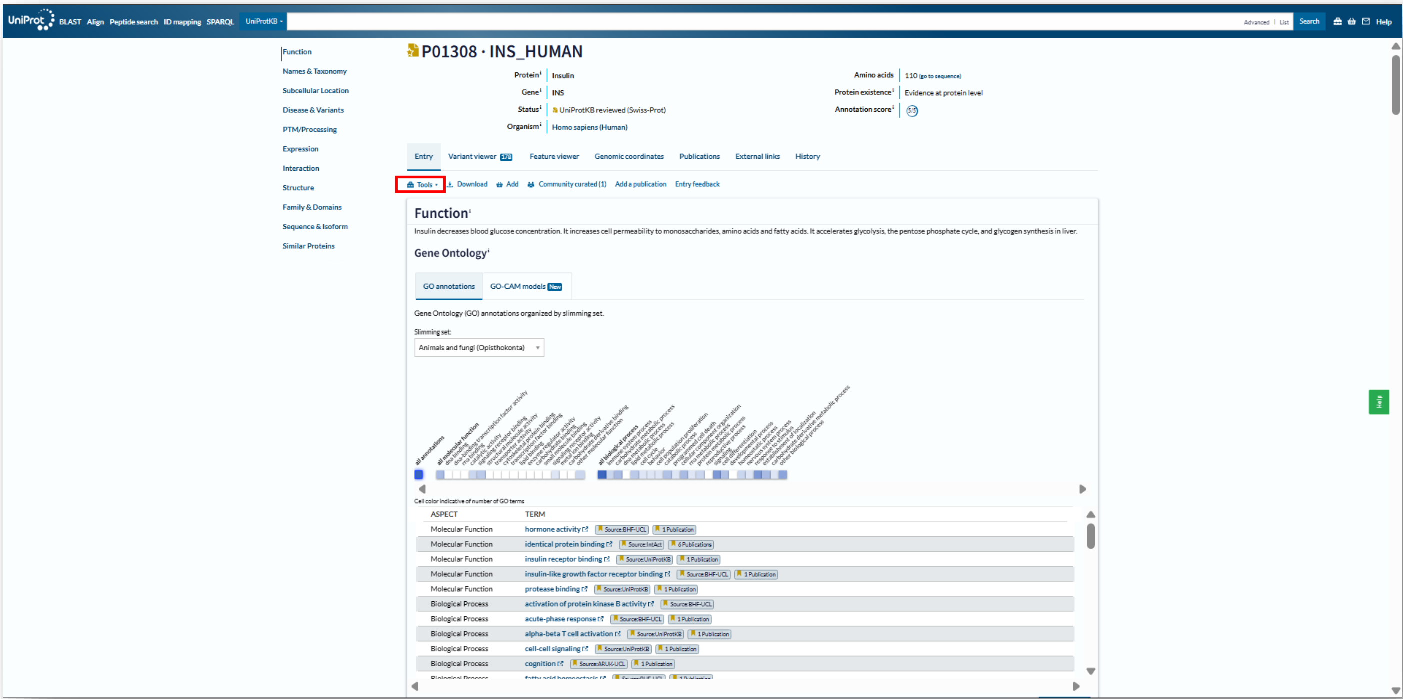


and select **FASTA** in Format drop-down and after that click on **Download** to retrieve it's FASTA sequence.

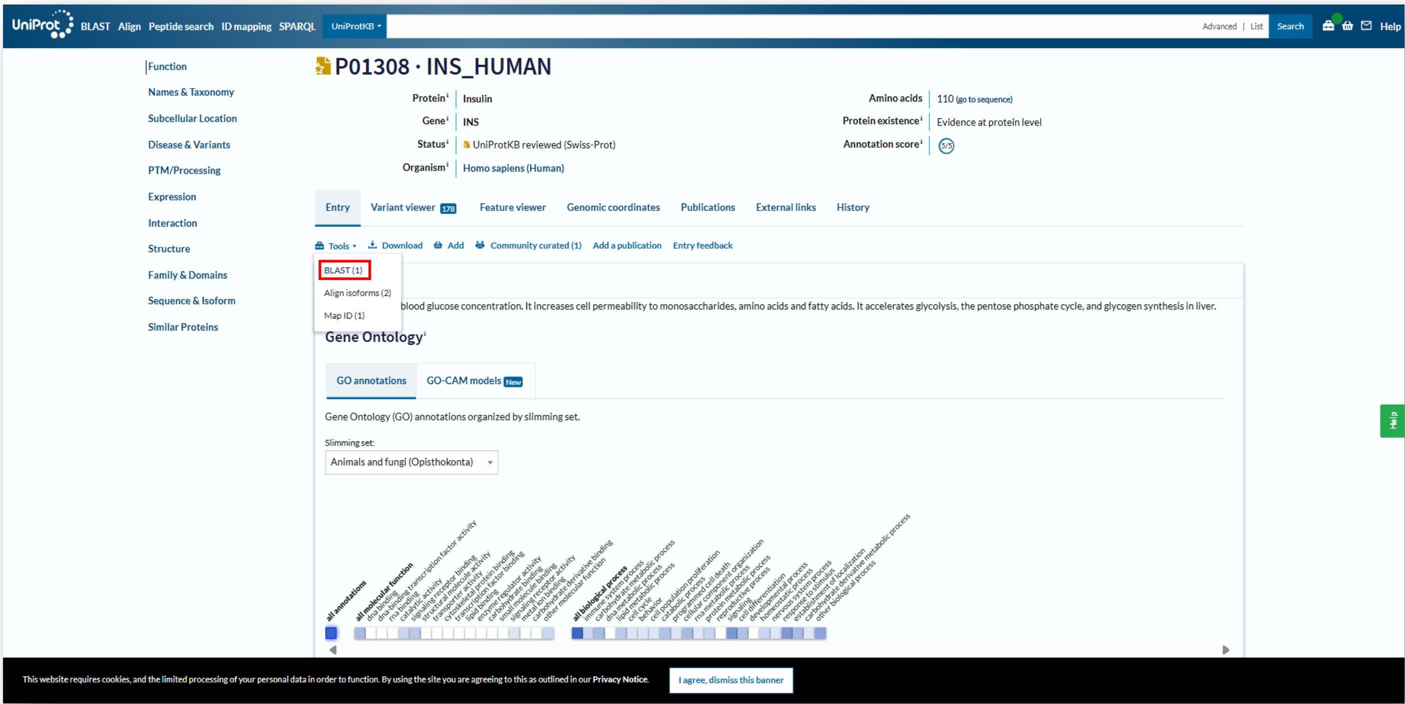


https://github.com/code-aradhana/bioinformatics-msa-project.git

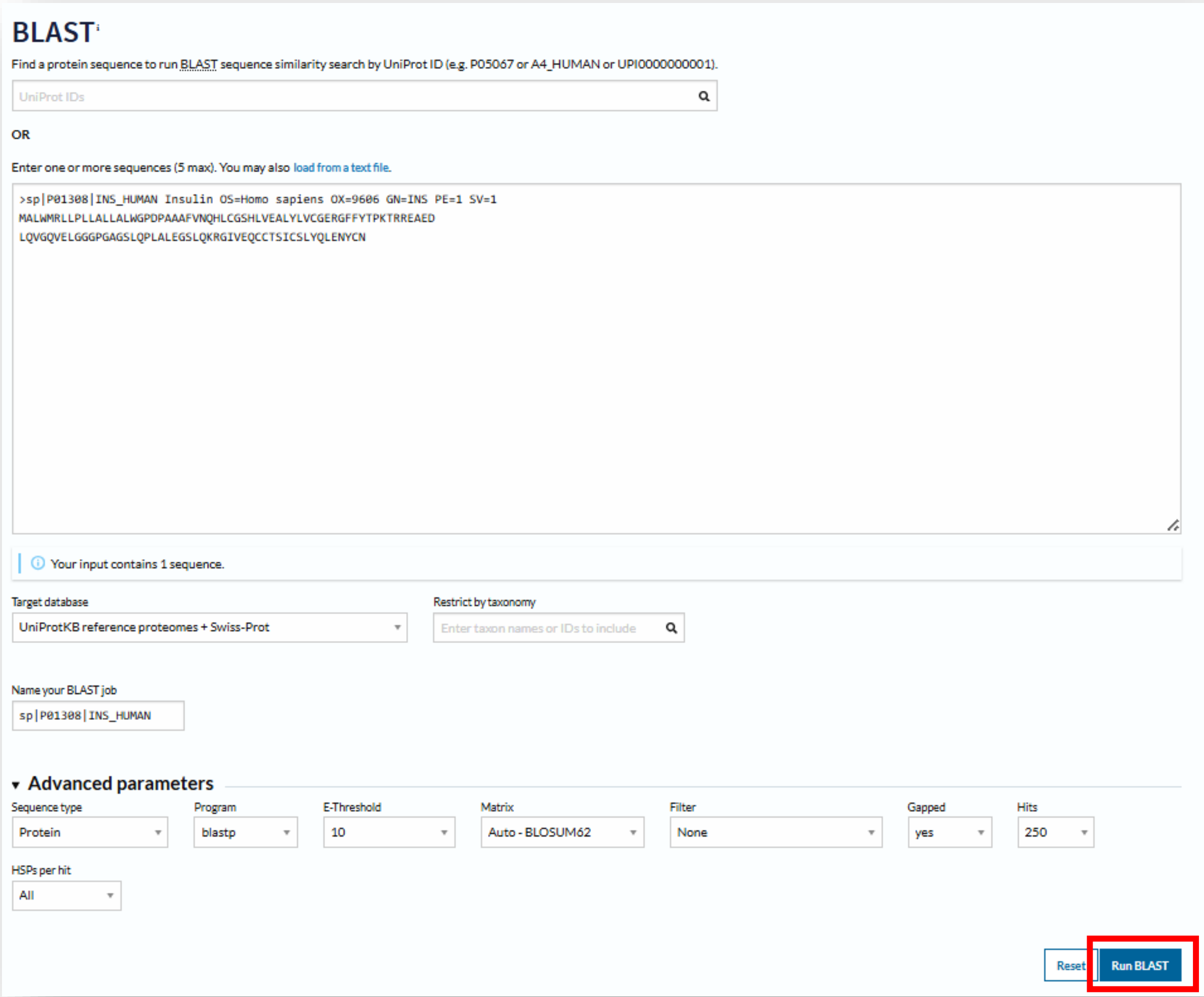
7. Click on **Tools.**



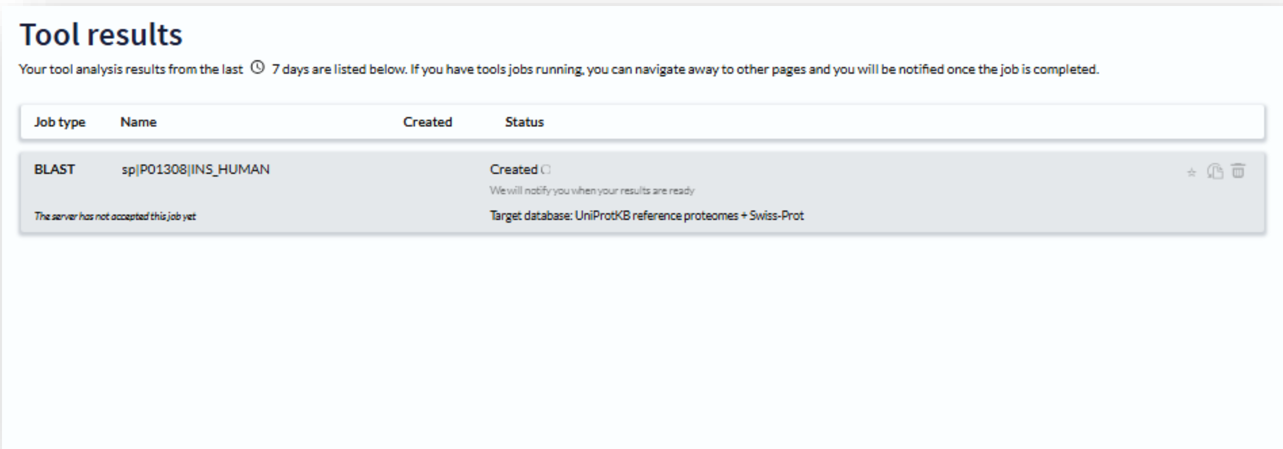
A menu will appear and then select **BLAST.**



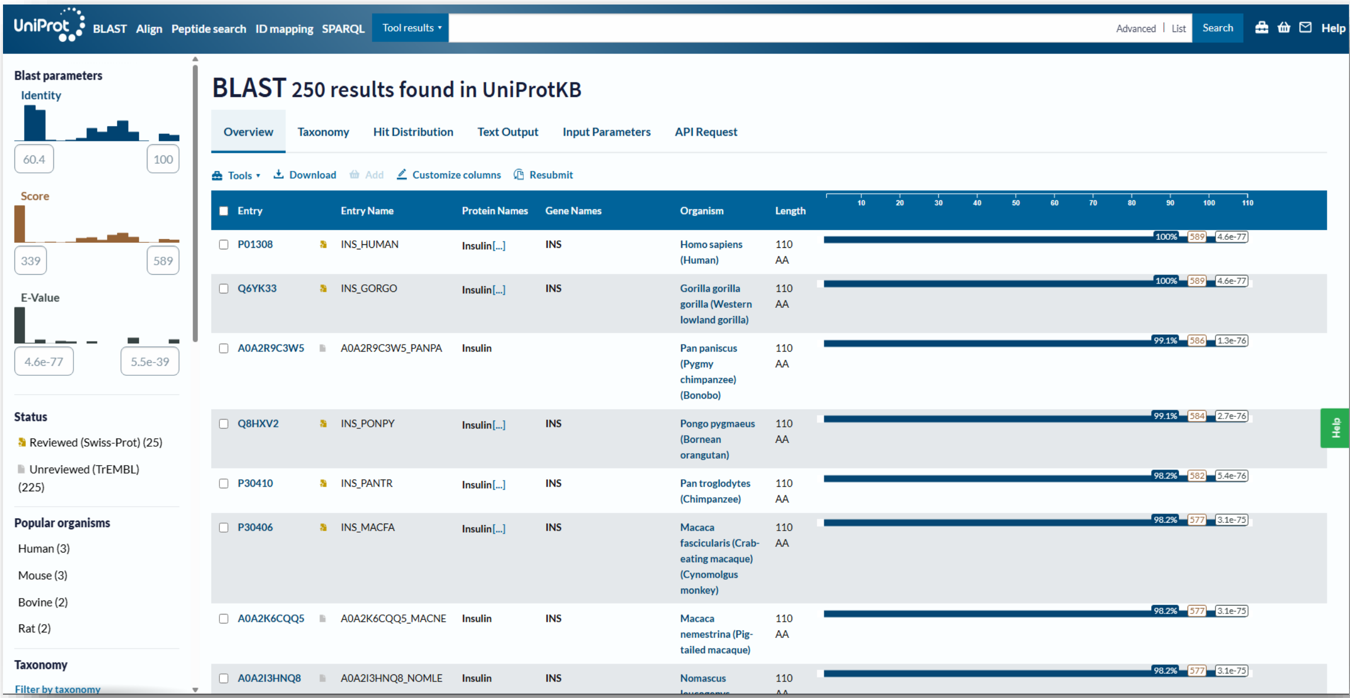
8. Scroll down and click **Run BLAST.**



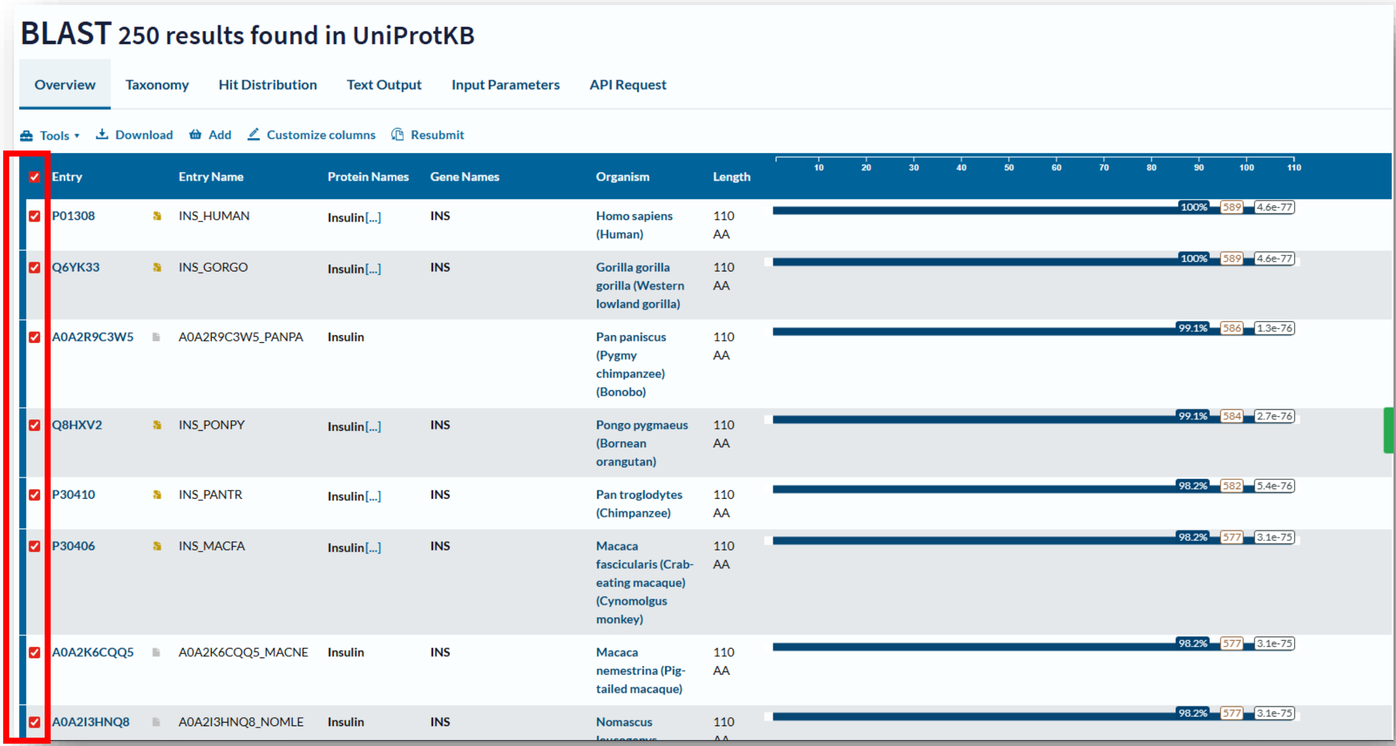
9. A page will come as shown below.



Wait for few minutes and then results will appear.

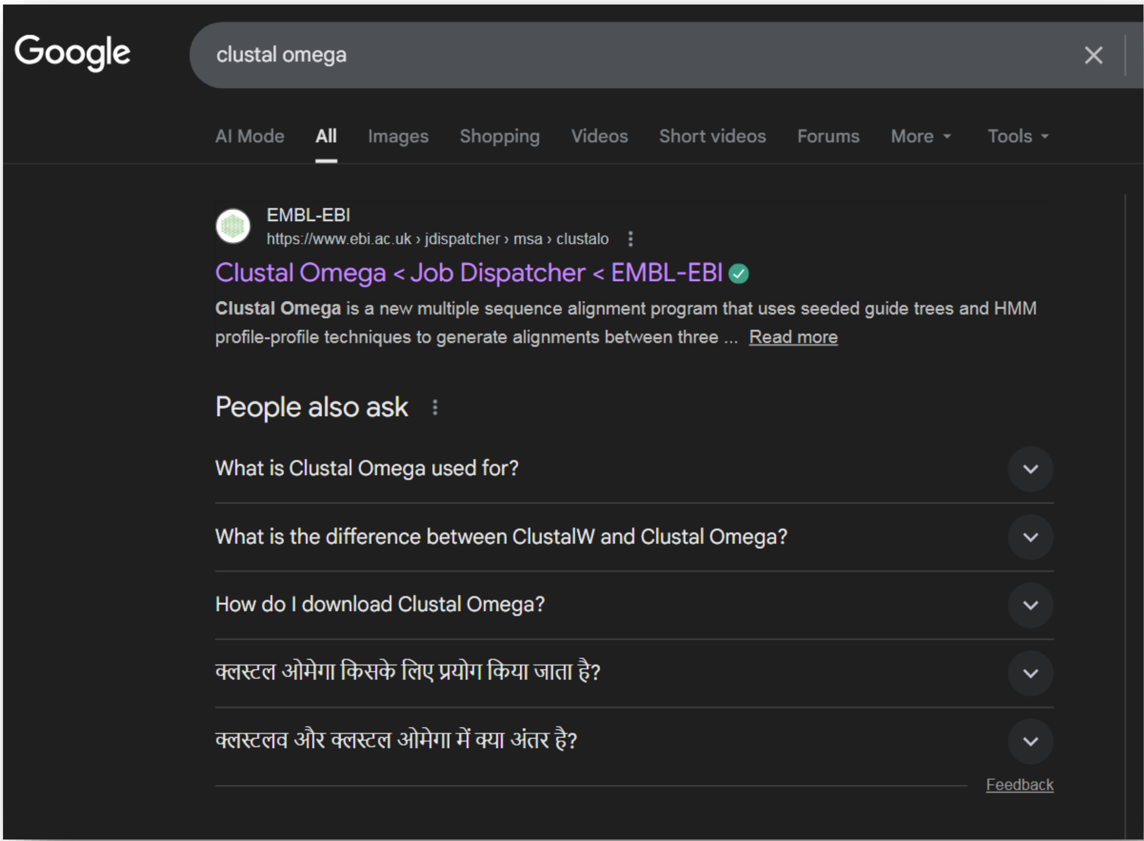


10. We are selecting **20 sequences and download it**. Download the Multiple Sequence Alignment.

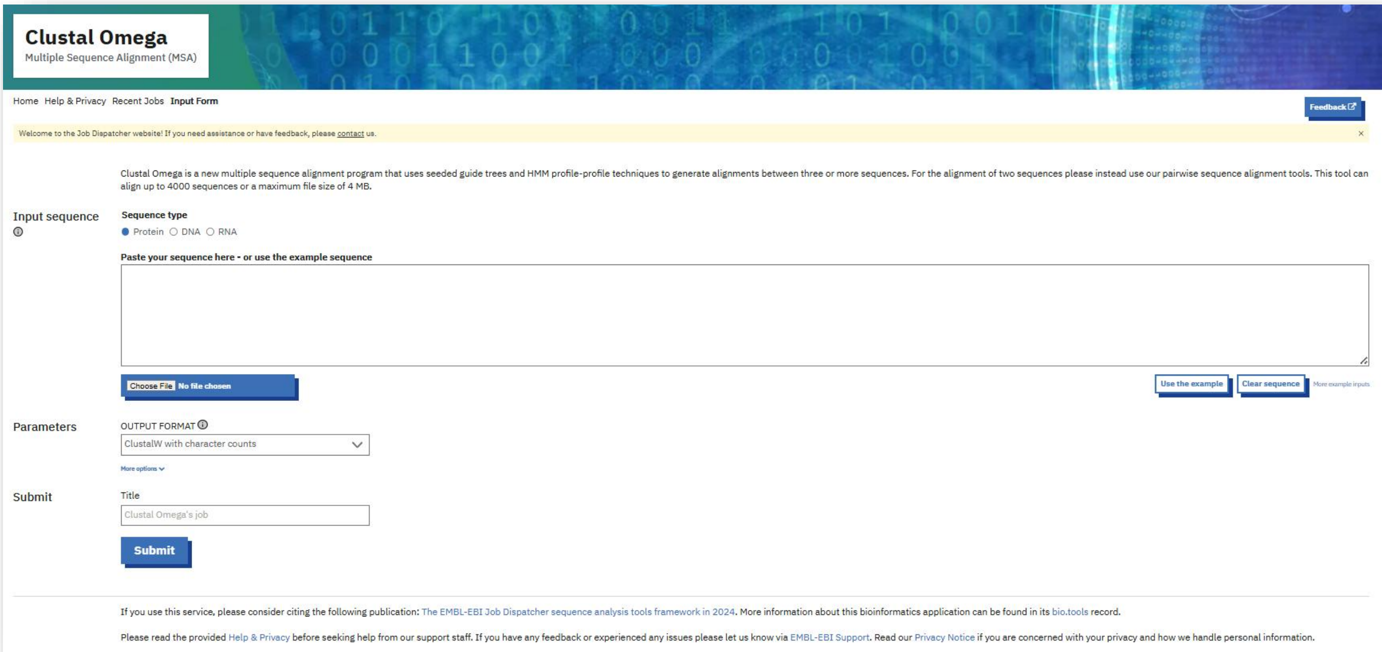


https://github.com/code-aradhana/bioinformatics-msa-project.git

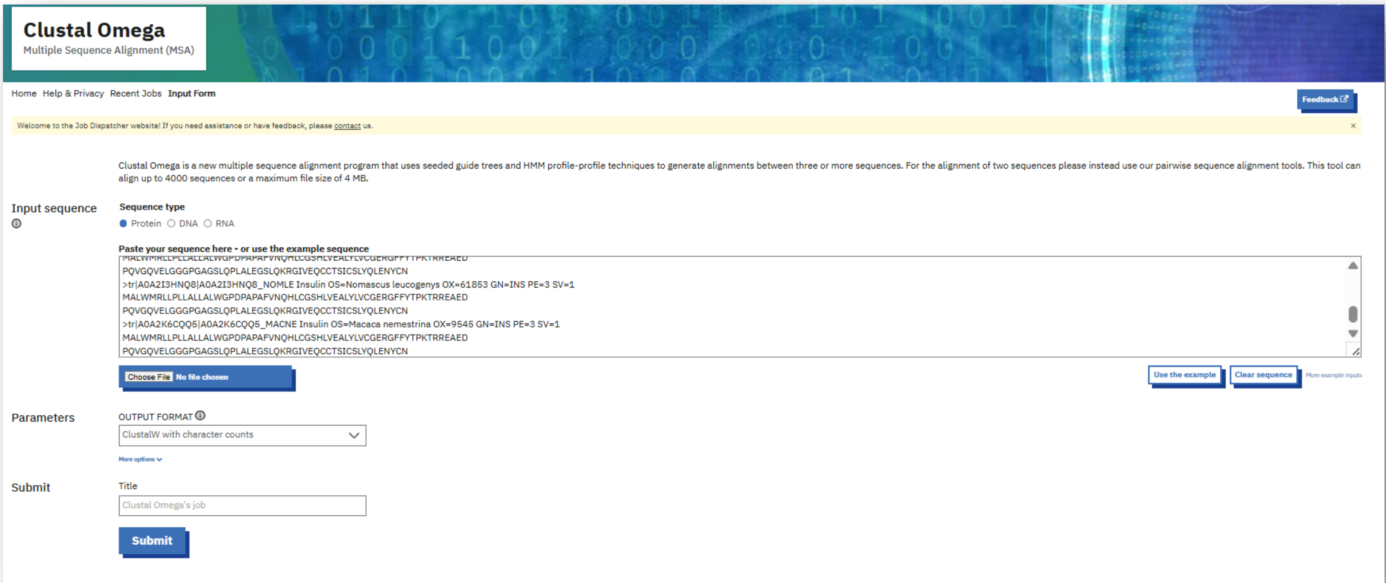
11. Search **Clustal Omega** on google.



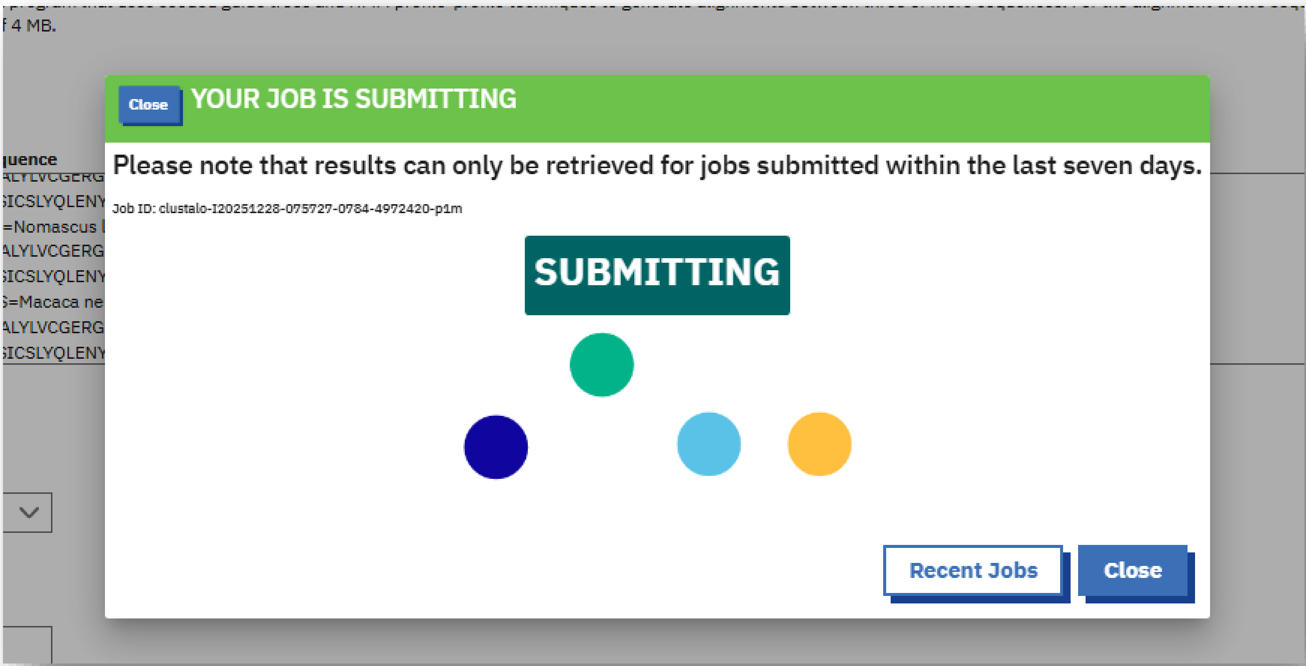
12. Open the **Clustal Omega.**



13. Paste all the 20 sequences in the query box.



14. Again, **a waiting pop-up** appear. Wait until the results appear on the screen. Clustal omega sometime take longer time to show results.



https://github.com/code-aradhana/bioinformatics-msa-project.git

15. Under **Tool Output** results will shown up.

Tool output

Download

```
CLUSTAL O(1.2.4) multiple sequence alignment


tr|A0A8I5TQT5|A0A8I5TQT5_PONAB      MALWMRLLPLLALLALWGPDPA-AFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAED    59
sp|P30410|INS_PANTR                 MALWMRLLPLLVLLALWGPDPASAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAED    60
sp|P01308|INS_HUMAN                 MALWMRLLPLLALLALWGPDPAAAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAED    60
sp|Q6YK33|INS_GORGO                 MALWMRLLPLLALLALWGPDPAAAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAED    60
tr|A0A2R9C3W5|A0A2R9C3W5_PANPA      MALWMRLLPLLALLALWGPDPASAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAED    60
sp|Q8HXV2|INS_PONPY                 MALWMRLLPLLALLALWGPDPAQAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAED    60
tr|A0A2K5P2L3|A0A2K5P2L3_CERAT      MALWMRLLPLLALLALWGPDPVPAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAED    60
tr|A0A8C9LMF1|A0A8C9LMF1_9PRIM      MALWMRLLPLLALLALWGPDPVPAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAED    60
tr|A0A8D2GB84|A0A8D2GB84_THEGE      MALWMRLLPLLALLALWGPDSVPAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAED    60
tr|A0A2K6R041|A0A2K6R041_RHIRO      MALWMRLLPLLALLALCGPDPVPAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAED    60
sp|P30406|INS_MACFA                 MALWMRLLPLLALLALWGPDPAPAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAED    60
tr|F7AUL3|F7AUL3_MACMU              MALWMRLLPLLALLALWGPDPAPAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAED    60
tr|A0A2I3HNQ8|A0A2I3HNQ8_NOMLE      MALWMRLLPLLALLALWGPDPAPAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAED    60
tr|A0A2K6CQQ5|A0A2K6CQQ5_MACNE      MALWMRLLPLLALLALWGPDPAPAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAED    60
tr|A0A096MTW9|A0A096MTW9_PAPAN      MALWMRLLPLLALLALWGPDPVPAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAED    60
tr|A0A0D9RBQ0|A0A0D9RBQ0_CHLSB      MALWMRLLPLLALLALWGPDPVPAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAED    60
tr|A0A2K5YKV7|A0A2K5YKV7_MANLE      MALWMRLLPLLALLALWGPDPVPAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAED    60
tr|A0AAJ7MUG7|A0AAJ7MUG7_RHIBE      MALWMRLLPLLALLALWGPDPVPAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAED    60
tr|A0A2K5JZH7|A0A2K5JZH7_COLAP      MALWMRLLPLLALLALWGPDPVPAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAED    60
sp|P30407|INS_CHLAE                 MALWMRLLPLLALLALWGPDPVPAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAED    60
                                    ***********.**** ***  .*************************************

tr|A0A8I5TQT5|A0A8I5TQT5_PONAB      LQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN  109
sp|P30410|INS_PANTR                 LQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN  110
sp|P01308|INS_HUMAN                 LQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN  110
sp|Q6YK33|INS_GORGO                 LQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN  110
tr|A0A2R9C3W5|A0A2R9C3W5_PANPA      LQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN  110
sp|Q8HXV2|INS_PONPY                 LQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN  110
tr|A0A2K5P2L3|A0A2K5P2L3_CERAT      PQVGQVELGGGPGAGSLQPLSLEGSLQKRGIVEQCCTSICSLYQLENYCN  110
tr|A0A8C9LMF1|A0A8C9LMF1_9PRIM      PQVGQVELGGGPGTGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN  110
tr|A0A8D2GB84|A0A8D2GB84_THEGE      PQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN  110
tr|A0A2K6R041|A0A2K6R041_RHIRO      PQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN  110
sp|P30406|INS_MACFA                 PQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN  110
tr|F7AUL3|F7AUL3_MACMU              PQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN  110
tr|A0A2I3HNQ8|A0A2I3HNQ8_NOMLE      PQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN  110
tr|A0A2K6CQQ5|A0A2K6CQQ5_MACNE      PQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN  110
tr|A0A096MTW9|A0A096MTW9_PAPAN      PQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN  110
tr|A0A0D9RBQ0|A0A0D9RBQ0_CHLSB      PQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN  110
tr|A0A2K5YKV7|A0A2K5YKV7_MANLE      PQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN  110
tr|A0AAJ7MUG7|A0AAJ7MUG7_RHIBE      PQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN  110
tr|A0A2K5JZH7|A0A2K5JZH7_COLAP      PQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN  110
sp|P30407|INS_CHLAE                 PQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN  110
                                     ************:*****:******************************
```

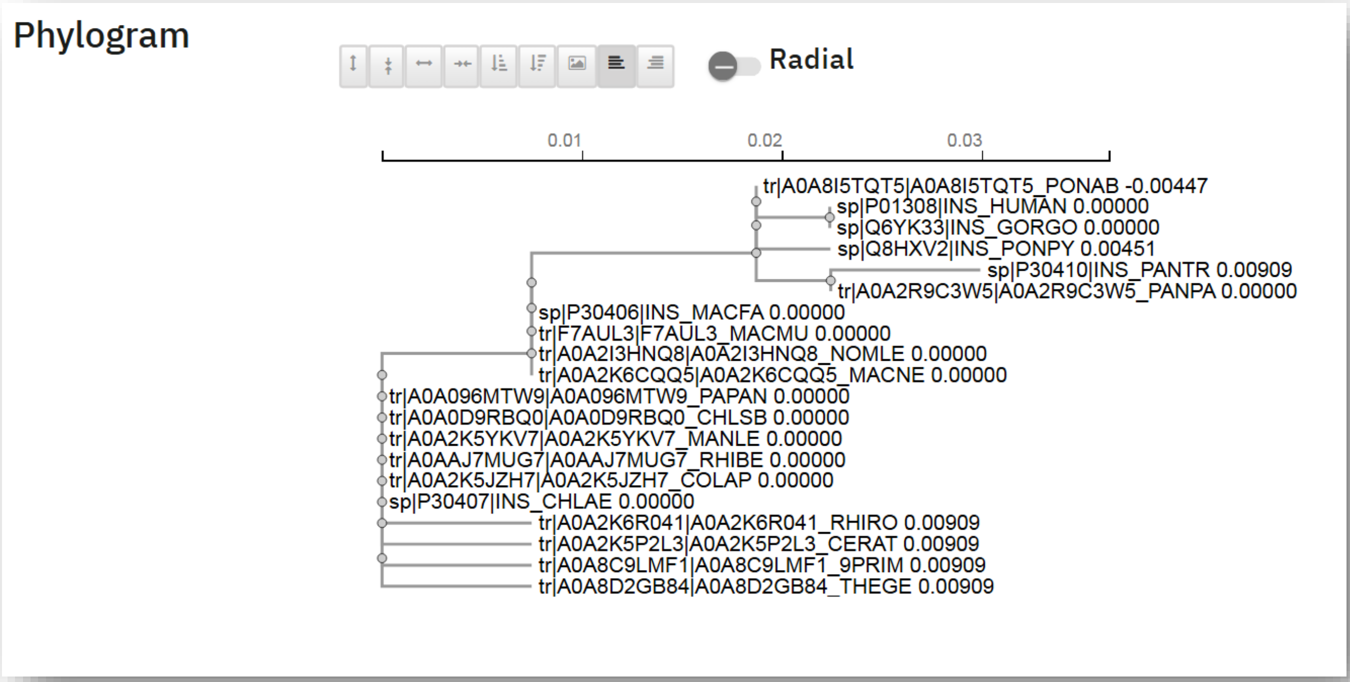
16. Scroll down to have **coloured sequences**.

17. Click the **Phylogenetic Tree** in horizontal tabular menu bar.

Phylogenetic Tree

```
(
(
(
(
(
(
(
(
(
(
(
(
(
(
(
(
tr|A0A8I5TQT5|A0A8I5TQT5_PONAB:-0.00447,
(
sp|P01308|INS_HUMAN:0.00000,
sp|Q6YK33|INS_GORGO:0.00000)
:0.00447)
:0.00000,
sp|Q8HXV2|INS_PONPY:0.00451)
:0.00000,
(
sp|P30410|INS_PANTR:0.00909,
tr|A0A2R9C3W5|A0A2R9C3W5_PANPA:0.00000)
:0.00453)
:0.01365,
sp|P30406|INS_MACFA:0.00000)
:0.00000,
tr|F7AUL3|F7AUL3_MACMU:0.00000)
:0.00000,
tr|A0A2I3HNQ8|A0A2I3HNQ8_NOMLE:0.00000)
:0.00000,
tr|A0A2K6CQQ5|A0A2K6CQQ5_MACNE:0.00000)
:0.00909,
tr|A0A096MTW9|A0A096MTW9_PAPAN:0.00000)
:0.00000,
tr|A0A0D9RBQ0|A0A0D9RBQ0_CHLSB:0.00000)
:0.00000,
tr|A0A2K5YKV7|A0A2K5YKV7_MANLE:0.00000)
:0.00000,
tr|A0AAJ7MUG7|A0AAJ7MUG7_RHIBE:0.00000)
:0.00000,
tr|A0A2K5JZH7|A0A2K5JZH7_COLAP:0.00000)
:0.00000,
sp|P30407|INS_CHLAE:0.00000)
:0.00000,
tr|A0A2K6R041|A0A2K6R041_RHIRO:0.00909)
:0.00000,
tr|A0A2K5P2L3|A0A2K5P2L3_CERAT:0.00909)
:0.00000,
tr|A0A8C9LMF1|A0A8C9LMF1_9PRIM:0.00909,
tr|A0A8D2GB84|A0A8D2GB84_THEGE:0.00909);
```

18. Scroll down to have **Phylogram.**



Results of Multiple Sequences Alignment of 20 sequences -

>sp|P01308|INS_HUMAN Insulin OS=Homo sapiens OX=9606 GN=INS PE=1
SV=1MALWMRLLPLLALLALWGPDPAAAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLEN
YCN

>sp|Q6YK33|INS_GORGO Insulin OS=Gorilla gorilla gorilla OX=9595 GN=INS PE=3
SV=1MALWMRLLPLLALLALWGPDPAAAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLEN
YCN

>tr|A0A2R9C3W5|A0A2R9C3W5_PANPA Insulin OS=Pan paniscus OX=9597 PE=3
SV=1MALWMRLLPLLALLALWGPDPASAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLEN
YCN

>sp|Q8HXV2|INS_PONPY Insulin OS=Pongo pygmaeus OX=9600 GN=INS PE=3
SV=1MALWMRLLPLLALLALWGPDPAQAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLEN
YCN

>sp|P30410|INS_PANTR Insulin OS=Pan troglodytes OX=9598 GN=INS PE=1
SV=1MALWMRLLPLLVLLALWGPDPASAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLEN
YCN

>sp|P30406|INS_MACFA Insulin OS=Macaca fascicularis OX=9541 GN=INS PE=3
SV=1MALWMRLLPLLALLALWGPDPAPAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDPQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLEN
YCN

>tr|A0A2K6R041|A0A2K6R041_RHIRO Insulin OS=Rhinopithecus roxellana OX=61622 GN=INS PE=3
SV=1MALWMRLLPLLALLALCGPDPVPAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDPQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENY
CN

>tr|A0A8D2GB84|A0A8D2GB84_THEGE Insulin OS=Theropithecus gelada OX=9565 GN=INS PE=3
SV=1MALWMRLLPLLALLALWGPDSVPAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDPQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLEN
YCN

>tr|A0A8C9LMF1|A0A8C9LMF1_9PRIM Insulin OS=Piliocolobus tephrosceles OX=591936 GN=INS PE=3
SV=1MALWMRLLPLLALLALWGPDPVPAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDPQVGQVELGGGPGTGSLQPLALEGSLQKRGIVEQCCTSICSLYQLEN
YCN

>tr|A0A2K5P2L3|A0A2K5P2L3_CERAT Insulin OS=Cercocebus atys OX=9531 GN=INS PE=3
SV=1MALWMRLLPLLALLALWGPDPVPAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDPQVGQVELGGGPGAGSLQPLSLEGSLQKRGIVEQCCTSICSLYQLEN
YCN

>tr|A0A096MTW9|A0A096MTW9_PAPAN Insulin OS=Papio anubis OX=9555 GN=INS PE=3
SV=2MALWMRLLPLLALLALWGPDPVPAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDPQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLEN
YCN

>tr|A0A0D9RBQ0|A0A0D9RBQ0_CHLSB Insulin OS=Chlorocebus sabaeus OX=60711 GN=INS PE=3
SV=1MALWMRLLPLLALLALWGPDPVPAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDPQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLEN
YCN

>tr|A0A2K5YKV7|A0A2K5YKV7_MANLE Insulin OS=Mandrillus leucophaeus OX=9568 GN=INS PE=3
SV=1MALWMRLLPLLALLALWGPDPVPAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDPQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLEN
YCN

>tr|A0AAJ7MUG7|A0AAJ7MUG7_RHIBE Insulin OS=Rhinopithecus bieti OX=61621 GN=INS PE=3
SV=1MALWMRLLPLLALLALWGPDPVPAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDPQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLEN
YCN

>tr|A0A2K5JZH7|A0A2K5JZH7_COLAP Insulin OS=Colobus angolensis palliatus OX=336983 PE=3
SV=1MALWMRLLPLLALLALWGPDPVPAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDPQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLEN
YCN

>sp|P30407|INS_CHLAE Insulin OS=Chlorocebus aethiops OX=9534 GN=INS PE=1
SV=1MALWMRLLPLLALLALWGPDPVPAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDPQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLEN
YCN

>tr|A0A8I5TQT5|A0A8I5TQT5_PONAB Insulin OS=Pongo abelii OX=9601 GN=INS PE=3
SV=1MALWMRLLPLLALLALWGPDPAAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENY
CN

>tr|F7AUL3|F7AUL3_MACMU Insulin OS=Macaca mulatta OX=9544 GN=INS PE=3
SV=3MALWMRLLPLLALLALWGPDPAPAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDPQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLEN
YCN

>tr|A0A2I3HNQ8|A0A2I3HNQ8_NOMLE Insulin OS=Nomascus leucogenys OX=61853 GN=INS PE=3 SV=1MALWMRLLPLLALLALWGPDPAPAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDPQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLEN YCN

>tr|A0A2K6CQQ5|A0A2K6CQQ5_MACNE Insulin OS=Macaca nemestrina OX=9545 GN=INS PE=3 SV=1MALWMRLLPLLALLALWGPDPAPAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDPQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLEN YCN

Application of Multiple Sequence Alignment

- Recombinant protein synthesis
- Drugs production
- Antibiotic production
- Functional genomics
- Determination of protein folding patterns in bioinformatics
- It plays vital role in proteomics.
- Used for the prediction of final structure, function and location of protein.
- To find out location of gene coding for that protein.
- Genetic diseases.
- Identification of sequence differences and variations such as point mutations.
- Revealing the evolution and genetic diversity of sequence and organisms.

Conclusion

This project successfully demonstrated the fundamental bioinformatics workflow for identifying evolutionarily conserved regions in protein sequences using Multiple Sequence Alignment (MSA). The primary objective—to align homologous insulin protein sequences from various primate species and analyze patterns of conservation—was achieved through a systematic process.

The practical steps involved retrieving target sequences (Human Insulin, P01308) from the UniProt database, using BLAST to find closely related homologous sequences, and performing the alignment with the Clustal Omega tool. The resulting MSA of 20 primate insulin sequences provided a clear visual and data-driven output.

In summary, this project confirms that Multiple Sequence Alignment is an indispensable tool in computational biology. It provides a powerful lens to visualize evolution's fingerprint on protein sequences, enabling researchers to infer function, trace ancestry, and form hypotheses for experimental validation. The skills developed—from sequence retrieval and database searching to alignment execution and interpretation—form an essential foundation for any subsequent bioinformatics analysis.

Reference

- Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids
  Authors: Richard Durbin, Sean R. Eddy, Anders Krogh, Graeme Mitchison.
- https://www.scribd.com/document/479739675/Multiple-Sequence-Alignment
- https://www.ebi.ac.uk/jdispatcher/msa
- Mount, D. W. (2004). Bioinformatics: Sequence and Genome Analysis (2nd ed.). Cold Spring Harbor Laboratory Press.
- **Lesk, A. M. (2017).** *Introduction to Protein Science: Architecture, Function, and Genomics* (3rd ed.). Oxford University Press.
- Felsenstein, J. (2004). Inferring Phylogenies. Sinauer Associates.
  A comprehensive guide to phylogenetic methods, emphasizing the role of conserved sequences in evolutionary studies.
- **UniProt Consortium (2023).** UniProt: the Universal Protein Knowledgebase. *Nucleic Acids Research, 51*(D1), D523–D531.