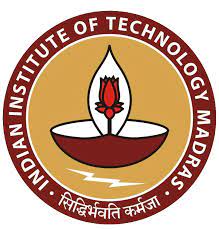
**BT 3040: Bioinformatics**

**Assignment 5**

..

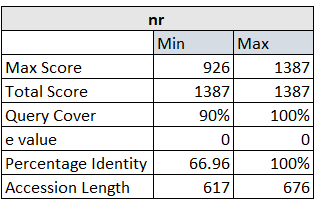


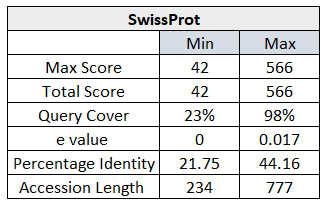
Atharva Mandar Phatak | BE21B009

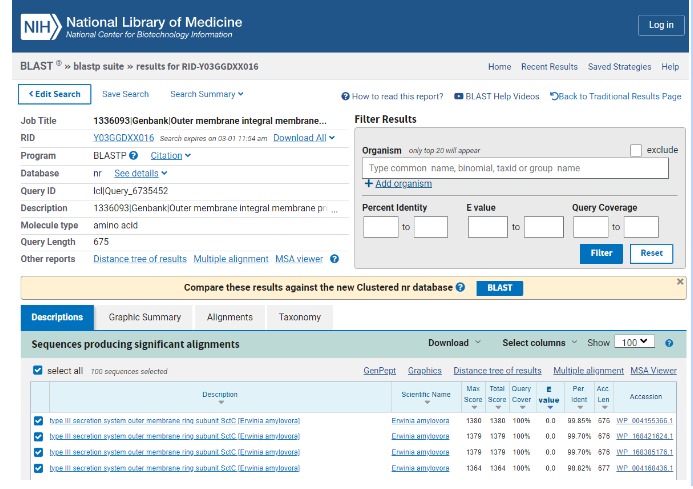
Indian Institute of Technology Madras

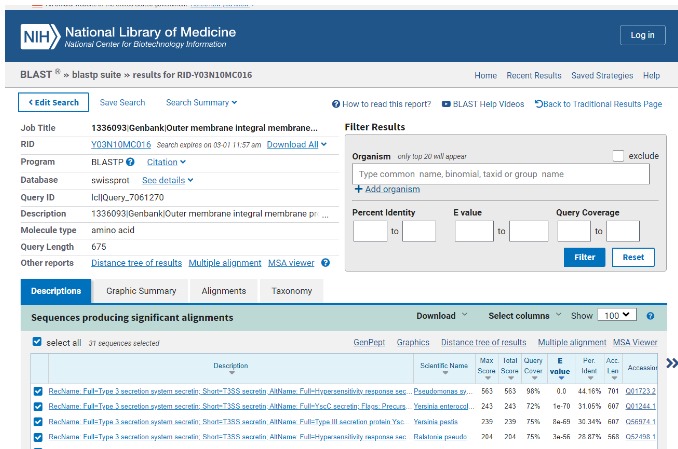
## **Q1) Analyze the occurrence of similar proteins in “nr” and SWISS-PROT database for the sequence given below:**

The following tables provide the analysis of occurrence of similar protiens.

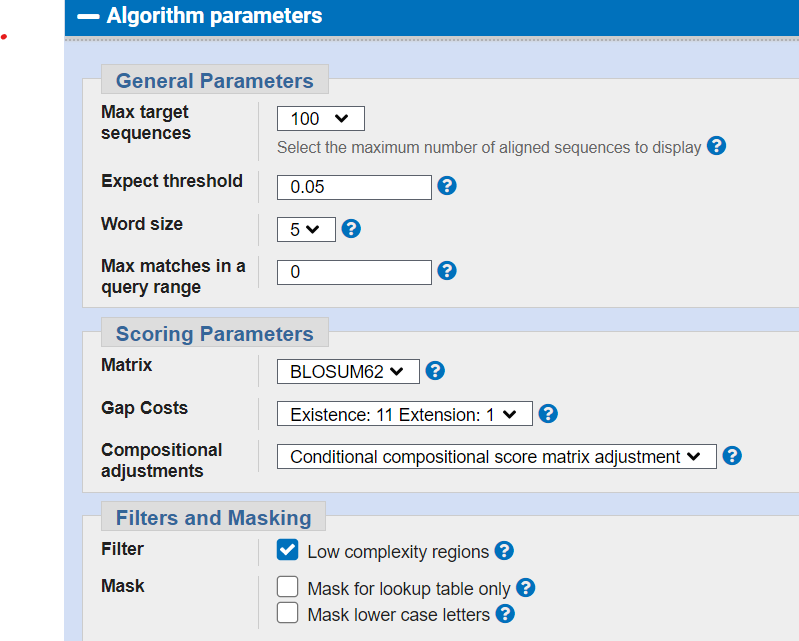








## **Q2) List the algorithm parameters used for the search (Q1)**



### General parameters displayed:

#### 1. Max target sequences

#### 2. Expected threshold

#### 3. Word size

#### 4. Maximum matches in a query range

### Scoring Parameters

#### 1. Matrix

#### 2. Gap costs

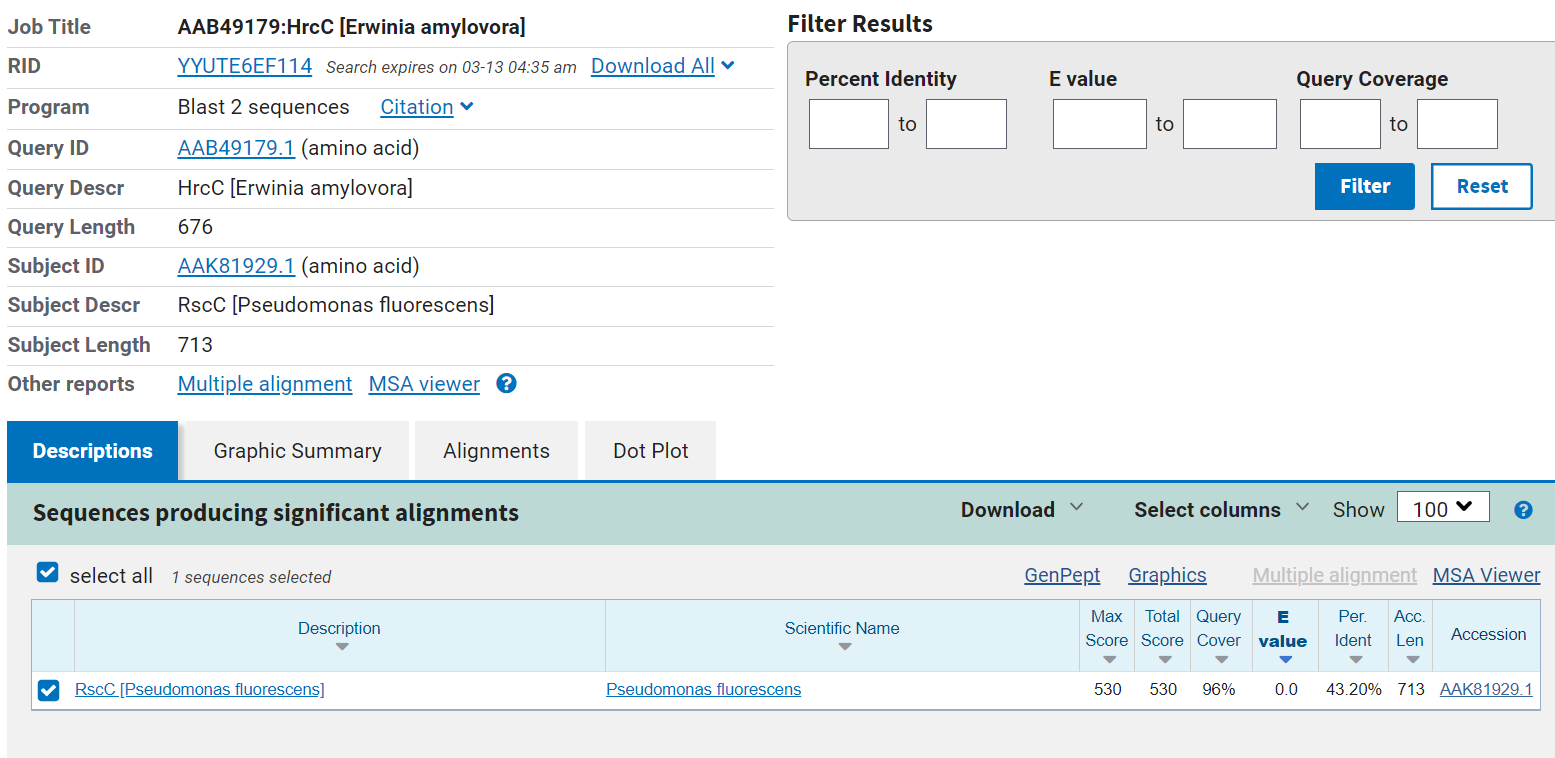
#### 3. Compositional adjustments

### Filter and Masking

#### 1. Filter

#### 2. Mask

## **Q3)** **What is the sequence identity of the query sequence (given in Q1) with AAK81929.1**



#### Max Score = 530

#### Total Score = 530

#### Query Cover = 96%

#### E Value = 0

#### **Per. Ident = 43.20%**

#### Acc. Len = 713

## **Q4) How far are hemoglobin (beta) sequences in humans and chicken similar?**

#### When searched for Haemoglobin sequences in UniProt, we get Human and Chicken protein sequences which are analysed in blast and the Percentage Identity is 69.39%. (Where the query cover is 100%). The percentage Identity is **69.39%**

## **Q5) Write a program to list all the matching pentapeptides (which occur in both the sequences) and their frequency of occurrence in given sequences.**

Code: <https://colab.research.google.com/drive/1w2RDIZpFqqW5EFqo6AZ7fBpcZ-ZNdsNA?usp=sharing>

Output:

The frequency of WTQRF in Human Hb Sequence is: 1 and in Chicken Hb Sequence is: 1

The frequency of HVDPE in Human Hb Sequence is: 1 and in Chicken Hb Sequence is: 1

The frequency of PWTQR in Human Hb Sequence is: 1 and in Chicken Hb Sequence is: 1

The frequency of FRLLG in Human Hb Sequence is: 1 and in Chicken Hb Sequence is: 1

The frequency of SELHC in Human Hb Sequence is: 1 and in Chicken Hb Sequence is: 1

The frequency of TQRFF in Human Hb Sequence is: 1 and in Chicken Hb Sequence is: 1

The frequency of LSELH in Human Hb Sequence is: 1 and in Chicken Hb Sequence is: 1

The frequency of GKKVL in Human Hb Sequence is: 1 and in Chicken Hb Sequence is: 1

The frequency of PENFR in Human Hb Sequence is: 1 and in Chicken Hb Sequence is: 1

The frequency of LHVDP in Human Hb Sequence is: 1 and in Chicken Hb Sequence is: 1

The frequency of LWGKV in Human Hb Sequence is: 1 and in Chicken Hb Sequence is: 1

The frequency of AHGKK in Human Hb Sequence is: 1 and in Chicken Hb Sequence is: 1

The frequency of YPWTQ in Human Hb Sequence is: 1 and in Chicken Hb Sequence is: 1

The frequency of WGKVN in Human Hb Sequence is: 1 and in Chicken Hb Sequence is: 1

The frequency of VYPWT in Human Hb Sequence is: 1 and in Chicken Hb Sequence is: 1

The frequency of KLHVD in Human Hb Sequence is: 1 and in Chicken Hb Sequence is: 1

The frequency of NFRLL in Human Hb Sequence is: 1 and in Chicken Hb Sequence is: 1

The frequency of VDPEN in Human Hb Sequence is: 1 and in Chicken Hb Sequence is: 1

The frequency of CDKLH in Human Hb Sequence is: 1 and in Chicken Hb Sequence is: 1

The frequency of HCDKL in Human Hb Sequence is: 1 and in Chicken Hb Sequence is: 1

The frequency of ELHCD in Human Hb Sequence is: 1 and in Chicken Hb Sequence is: 1

The frequency of DPENF in Human Hb Sequence is: 1 and in Chicken Hb Sequence is: 1

The frequency of HGKKV in Human Hb Sequence is: 1 and in Chicken Hb Sequence is: 1

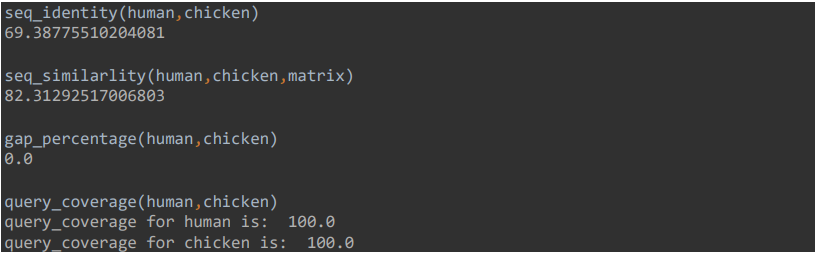
The frequency of ENFRL in Human Hb Sequence is: 1 and in Chicken Hb Sequence is: 1

The frequency of GKVNV in Human Hb Sequence is: 1 and in Chicken Hb Sequence is: 1

The frequency of LHCDK in Human Hb Sequence is: 1 and in Chicken Hb Sequence is: 1

The frequency of DKLHV in Human Hb Sequence is: 1 and in Chicken Hb Sequence is: 1

## **Q6) Write a program to compute sequence identity, similarity, query coverage and gap percentage from the alignment of human and chicken hemoglobin sequences (refer Q4).**

****

#### Sequence identity: 69.38775%

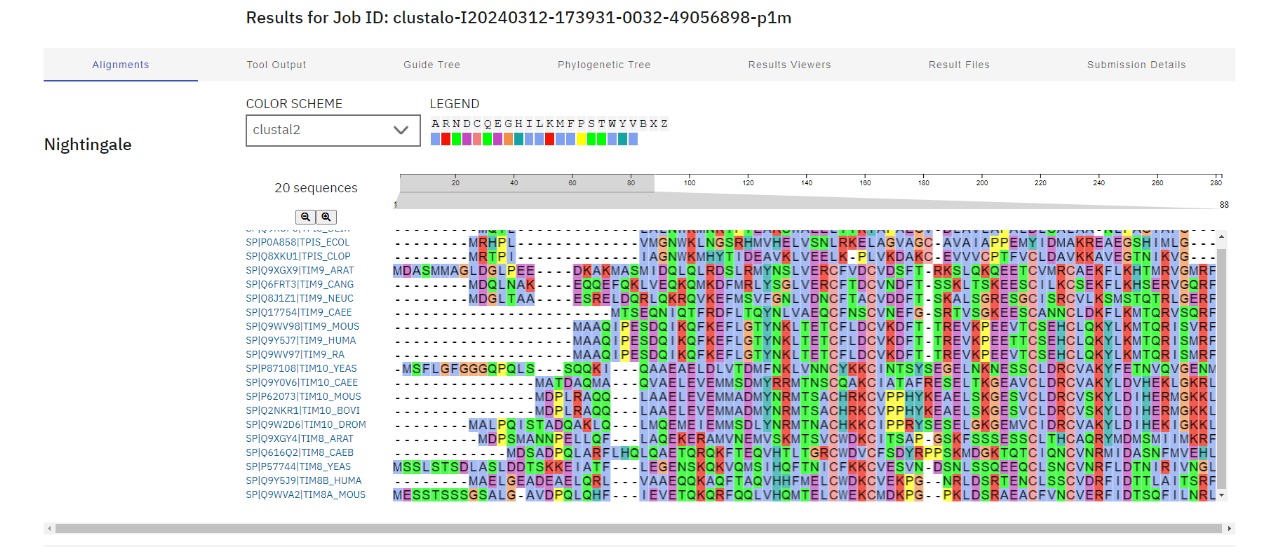
#### Sequence similarity: 82.3129%

#### Gap percentage: 0%

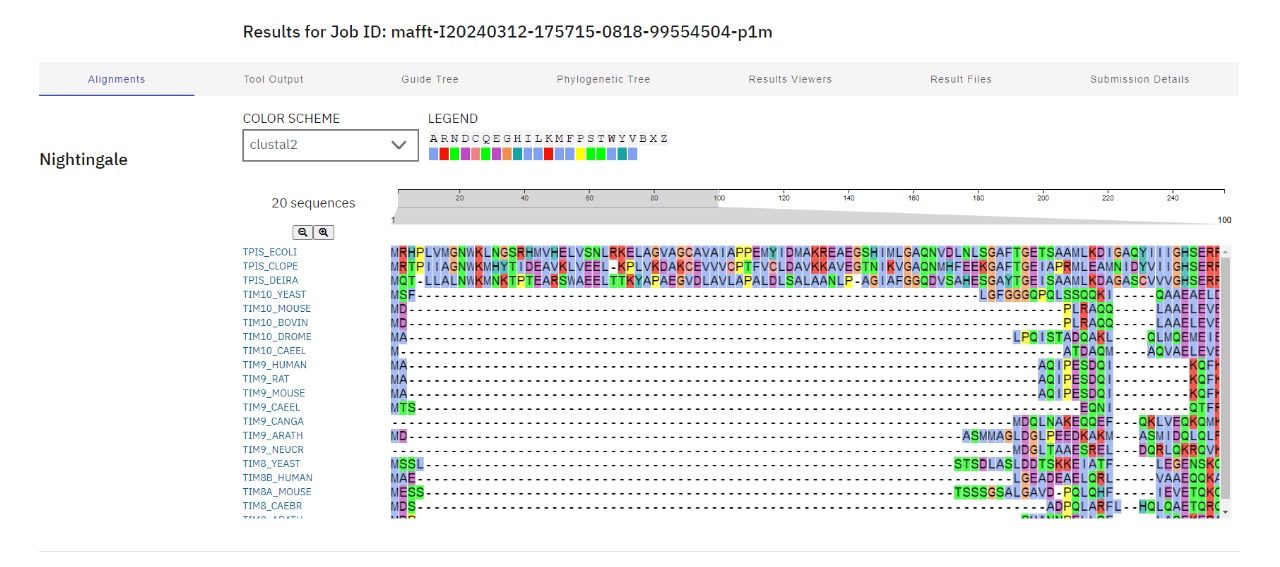
#### Query Coverage: 100%

## **Q7) Obtain the multiple sequence alignment for TIM barrel proteins from different organisms (select 20 proteins, for example). Compare the results obtained with Clustal Omega, MAFFT, and MUSCLE. List 5 residue positions which are aligned differently in these three methods.**

#### Clustal



#### MAFFT



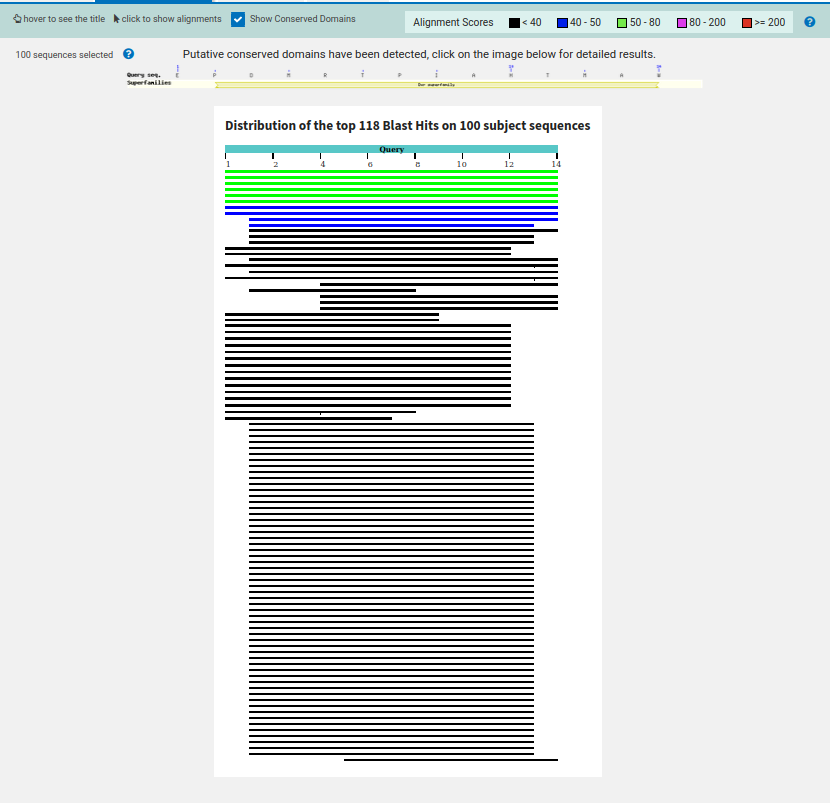
#### c) MUSCLE

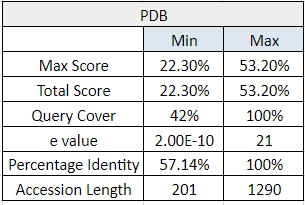
#### 

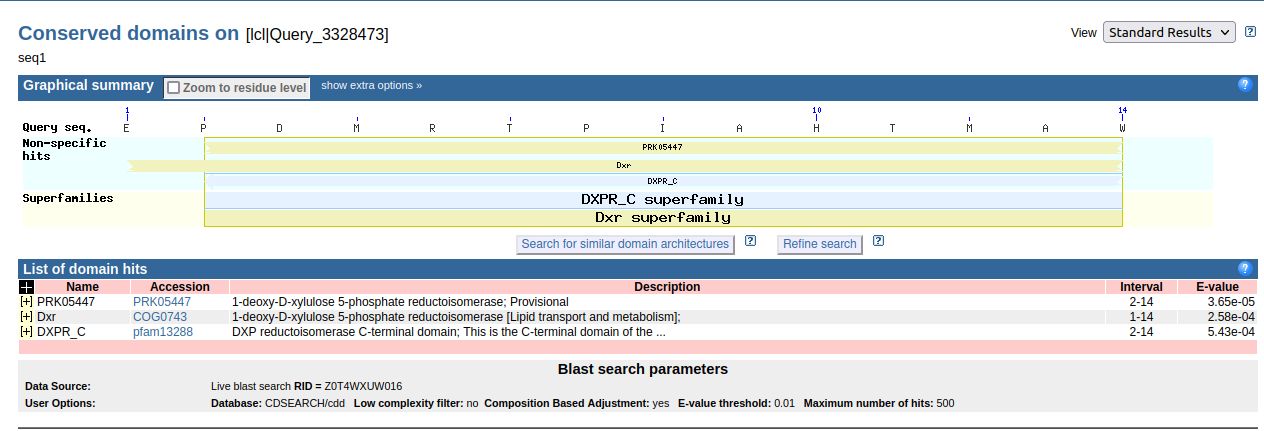
#### Analysis:

#### The lengths of the aligned sequences are 120, 117, and 116 for Clustal Omega, MAFFT, and MUSCLE, respectively. It’s evident that the alignment of residues at positions 1, 2, 3, and 4 varies across these three methods. As we move further along the residue positions, the alignments become increasingly diverse.

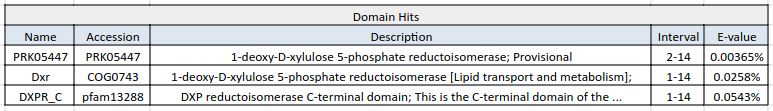
**Q8) Blast the below sequence 'EPDMRTPIAHTMAW' against the PDB database. Analyze the results and discuss the significance of the results.**

Graphical representation of the BLAST sequence





Thus there are three domain hits for the given sequence.



#### Residue 2-14 is a domain named PRK05447 which is conserved across many organisms.

#### BLAST results show this sequence is commonly found in - Escherichia coli, Klebsiella

#### pneumoniae.

#### 100 sequences, only 10% of the sequences have very minimal E value.