## **BIO 101 Introductory Biology**

## Assignment 3

## Marks 30

## **Instructions:**

- Cite the resources you are using properly
- Attach screenshots/pictures in the document and convert it to a single PDF file
- Submit it on LMS
- 1) Download FASTA file for protein ID Po5205 from uniport.
  - a) What is the length of the coding region of the gene that codes for this protein? (2)
  - b) How many exons and introns are present in this protein? What are their sequences? (3)
  - c) How many splice variants are there for this protein? (2)
  - d) What is the function of this protein and what kind of domains are present? (3)
  - e) BLAST this protein in the following organisms. What kind of hits do you obtain? What does that tell you about this protein? (3+3+3) Attach screenshots while describing them as well.
    - i) Homo Sapiens
    - ii) Arabidopsis Thaliana
    - iii) Musca Domestica
  - f) Is the crystal structure of any region of this protein available? What techniques were used to visualize the structure? Attach any available pictures. (4)
- 2) Find the transition score for the following protein sequences using the BLOSUM62 matrix.
  - a) HENRYCAVILL into HARRYSTYLES (5)
  - b) CHRISPRATT into CHRISEVANS (5)

A	4																		
R	-1	5															400		
N	-2	0	6						BL	OSI	UM	62	SCC	orin	ig n	nati	rix		
D	-2	-2	1	6	2														
C	0	-3	-3	-3	9														
Q	-1	1	0	0	-3	5					(1	osit	tive	val	ues	are	sha	adec	1)
E	-1	0	0	2	-4	2	5				- 55								
G	0	-2	0	-1	-3	-2	-2	6											
н	-2	0	1	-1	-3	0	0	-2	8										
I	-1	-3	-3	-3	-1	-3	-3	-4	-3	4									
L	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4								
K	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5							
М	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5						
F	-2	-3	-3		-2	-3	-3	-3	-1	0	0	-3	0	6					
P	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7				
S	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4			
T	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5		
w	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11	
Y	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7
٧	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1
_	A	R	N	D	-	0	F	G	н	T		K	M	F	P	S	7	w	Y

The values for amino acid substitutions were obtained from Henikoff S & Henikoff JG (1992) Amino acid substitutions matrices from protein blocks. Proc. Natl. Acad. Sci. 89: 10915-10919.