

Roll No.-----

B.Sc. - M.Sc. INTEGRATED (BIOCHEMISTRY)

SEMESTER IV, Examination 2019

PAPER CODE - BCH-GE-04 TH

PAPER TITLE: INTRODUCTION TO BIOLOGICAL DATA ANALYSIS

(Write your Roll No. at the top immediately on the receipt of this question paper)

Time: Three Hours

Maximum Marks: 75

Attempt ALL questions. All parts of a question must be answered in continuation. Start answer of a fresh question on a new page. Cite examples and draw labelled diagrams wherever necessary.

Q1. Attempt ALL of the following. State whether TRUE or FALSE.

15X1=15

- i. Unix is a programming language.
- ii. The output of "DNA"[1] will be N.
- iii. The output of len("GGAATTC CATG") will be 12.
- iv. \$ cat file1 file 2 >> new_file will append the contents of file 1 and file 2 to new_file.
- v. The move command can be used to rename a file in linux.
- vi. \$ wc -l file_name will print the number of characters in the file named file_name.
- vii. \$ sort -k filename.txt will sort the file on the basis of specific column.
- viii. The default directory is home directory when we start our terminal.
- ix. We can use set function to grab unique items from a list in python.
- x. If x = "GATC" then x[::-1] will print "CTAG".
- xi. If x = "DNA" then print (f'{x}== DNA') will give an error.
- xii. The command print("\n") can be used to print a new line in python.
- xiii. The expression GATC{2,4}TA will match "GATCCCTA" pattern.
- xiv. The expression GATC*GA will match "GATGA" pattern.
- xv. If n = ["G", "A", "T", "C"] then n.pop() will give "C" as result.

Q2. Attempt any ANY THREE of the following:

3X5=15

- A. Briefly describe the unix file system.
- B. Write short note on functions of operating system.

- C. You are provided with a file named dna.txt having several columns and rows. Write linux scripts to perform the following tasks:
- Print only those lines in column 2 which contains the word "DNA".
 - Print only column 4 of the file.
 - Write a single grep command to search both "DNA" and "dna" in a file.
 - Write a script to print first 50 lines of the file.
 - Write a script to substitute "ATG" with "GAT" in first two lines of the file.
- D. You have a file named dna.txt which contains multiple DNA sequences of varying length. Write a linux script to iterate through each line and find the word count of each line. Store the word count result in a new file named count.txt.
- E. Briefly describe piping and file permissions in unix with suitable examples.

Q3. Attempt ANY THREE of the following:

3X5=15

- Briefly describe conditional tests in python citing suitable examples.
- Describe various steps you will follow to read and write files in python.
- Study the gene_func dictionary and provide the output of following commands:
`gene_func = {'p53' : 'apoptosis', 'Stat' : {'Stat1' : 'Innate', 'Stat2' : 'survival' , 'Stat3' : 'Replication'}, 'Rab' : 'Signaling'}`
 - `print(gene_func['Stat'])`
 - `print(gene_func['Stat']['Stat2'])`
 - `print(len(gene_func))`
 - Write a script to retrieve value of "Rab" key from this dictionary.
 - Write a script to add a new "Atf3": "TLR signaling" pair to this dictionary.
- You are given a list named L1 which contains multiple DNA sequences. Write a python script to iterate through L1 and print True if "GGAATTC" pattern is present in the sequence or False if the pattern is not present.
- You have multiple DNA sequences of variable length in a file named dna. Using python regular expressions answer the following questions.
 - Search for EcoR1 restriction site (GAATTC) in the dna_file.
 - Write a scrip that will match either "GGGATGAAA" or "GGGAAA" sequences in dna_file.
 - Explain what pattern will this expression match `GCC{5}T`.

- iv. Search for sequences that start with "ATG" and ends with "TAG" with 100 nucleotides in between them.
- v. Search for sequences that have poly A tail of 5 to 10 nucleotides in length.

Q4. Attempt ANY THREE of the following:

3X5=15

- A. Briefly describe exception handling in python.
- B. Write a python script to generate a random sequence of DNA which is 100000 nucleotides long.
- C. You are given a list named list1 with some numbers. Write a script to calculate the sum of all numbers present inside the list.
- D. Write a script to read DNA sequence from a file (dna.txt) and calculate GC content.
- E. Define a function reverse_complement and use this function to find the complement of 'ACCATTG' sequence.

Q5. Write short note on ANY THREE of the following;

3X5=15

- A. Python Debugger (pdb).
 - B. Modules in python.
 - C. Text editors.
 - D. User interface in Python.
 - E. Virtual Environment (VE) and VEWwrapper.
-