STELLA MARIS COLLEGE (AUTONOMOUS) CHENNAI 600 086
(For candidates admitted from the academic year 2019 – 2020 & thereafter)

M. Sc. DEGREE EXAMINATION, NOVEMBER 2023
BIOINFORMATICS
THIRD SEMESTER

COURSE : CORE
PAPER : PYTHON AND R PROGRAMMING
SUBJECT CODE : 19BI/PC/PR34
TIME : 3 HOURS **MAX. MARKS: 100**

SECTION – A

ANSWER ALL THE QUESTIONS: (20x1=20)

1. How do you define a function in Python?
2. What does DNA parsing mean in the context of bioinformatics?
3. What is Jupyter Notebook, and what are its primary uses?
4. Differentiate between mutable and immutable data types in Python?
5. What is Biopython and what are its primary applications?
6. What is meant by 'translating' a DNA sequence in bioinformatics?
7. How is the Bio.Seq module utilized in Biopython?
8. What is the Entrez gene database, and how is it important for genetic research?
9. How can you use R as a calculator to compute the square root of 16?
10. In R, how do you assign the value of 5 to an object named **x**?
11. What is R, and what are its main applications?
12. In what format does R save graphical output by default?
13. What is the primary purpose of feature selection in gene expression data analysis?
14. Name two commonly used methods for data preprocessing in gene expression studies.
15. Define a heatmap and its role in visualizing gene expression data.
16. What role does R play in analyzing gene expression data?
17. What is Bioconductor and its primary purpose?
18. Name two popular packages available in Bioconductor.
19. Name a common R package used in sequence analysis.
20. Describe a basic function of the 'DESeq2' package in Bioconductor.

SECTION – B

ANSWER ANY FOUR OF THE FOLLOWING (4x10=40)

21. Explain the concept of variables in Python, including their naming conventions and common data types.
21. Describe in detail the various functionalities offered by the Bio.Seq module, giving examples of its practical applications.
23. Discuss the significance of creating objects in R and assigning values to them. Provide examples showcasing this process.
24. Explain the difference between simple plotting and advanced plotting in R. Include examples of the capabilities of each.
25. Discuss the significance of data preprocessing and normalization in gene expression data analysis, including the challenges faced without these steps.
26. Describe the process of creating and interpreting heatmaps in the context of gene expression studies, illustrating their importance.

..2

27. Discuss the architecture of Bioconductor and its advantages in handling and analyzing high-throughput genomics data.

SECTION – C

ANSWER ANY TWO OF THE FOLLOWING (2x20 =40)

28. Given a FASTA file containing DNA sequences, explain the steps and Python code snippets required to parse the information, locate specific genes, and extract their positions. Consider error handling and other essential practices.

29. Provide a comprehensive guide to getting started with Biopython, starting from its installation, the significance of its major modules, and a basic walkthrough of a simple bioinformatics task.

30. Provide a comprehensive guide to plotting in R, beginning with simple plotting techniques, progressing to advanced plotting, and culminating in customizing plots with colors, labels, subscripts, and superscripts. Support your guide with illustrative code snippets.

31. Delve deep into the applications of R and Bioconductor in phylogenetics and sequence analysis. Describe the tools, packages, and methodologies available, their functionalities, and demonstrate a complete analysis process using a hypothetical dataset.

****