JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST -3 EXAMINATION- 2024

M.Tech-I Semester (BT)

COURSE CODE (CREDITS): 13M11BT112 (3) MAX. MARKS: 35

COURSE NAME: Advanced Bioinformatics

COURSE INSTRUCTORS: Dr. Raj Kumar MAX. TIME: 2 Hours

Note: *(a) All questions are compulsory.*

(b) The candidate is allowed to make Suitable numeric assumptions wherever required for solving problems

Q.No	Question	Marks
Q1	You have a genome of 2×10^9 base pairs (bp). After sequencing, you generated 3×10^7 reads, with an average read length of 150 bp. What is the mean coverage of the genome?	5
Q2	For given nucleotides GNGC, the Sanger FASTQ values are 32, 2, 29, and 29, respectively. Analyze and interpret these results in your own words.	3
Q3	In Next-Generation Sequencing (NGS) data analysis, the Mapping Quality (MapQ) values typically range from 0 to 60 in the SAM/BAM file format. What is the impact of these values on the quality of the alignment of a read to a reference genome?	3
Q4	Genome assemblies offer a consensus representation of a genome, spanning all the chromosomes. Discuss important steps in a typical mapping assembly process.	3
Q5	A CIGAR string for an alignment is given as: 8M2I5M3D7M. What does this CIGAR string represent and what does each component mean? What does this suggest about the alignment's length and its relationship to the reference sequence?	5
Q6	Homology modeling involves several steps in the model-building process. Discuss each of these steps in detail, emphasizing their impact on the quality and accuracy of the final model.	7
Q7	You have run FASTQC on a dataset of raw sequencing reads. What can you infer from the FASTQC report regarding the quality of your sequencing data?	9