**Final Exam**

**RUDRANIL MONDAL**

Email: rudranill.official@gmail.com

**Section A:**

1. True
2. False  
   NCBI stores not only DNA sequences but also RNA, protein sequences, literature (PubMed), structures, genomes, and more.
3. False  
   Sanger sequencing is known for high accuracy and long reads, but it is slower and lower throughput compared to modern NGS technologies.
4. True  
   Quality control of raw reads (e.g., using FastQC) is crucial to detect low-quality bases, adapters, or contamination.
5. False  
   Single-End sequencing provides only one read per DNA fragment. It’s Paired-End sequencing that gives two reads, which aids more accurate mapping.

**Section B:**

**1. What is bioinformatics? What is the purpose of bioinformatics in modern science?**

**Ans:** Bioinformatics is an interdisciplinary field that develops methods and tools for understanding biological data. It helps analyze large-scale data like DNA, RNA, and proteins to make scientific discoveries, improve healthcare, agriculture, and more.

**2. How does bioinformatics integrate multiple disciplines like computer science, biology, and statistics?**

**Ans:** Bioinformatics combines:

* **Biology** to understand living systems,
* **Computer science** to develop tools and manage data,
* **Statistics** to analyze and interpret data.  
  This integration allows solving complex biological questions efficiently.

**3. How do biological databases like NCBI, Ensembl, and PDB differ in their functionality?**

**Ans:**

* **NCBI**: Comprehensive data on sequences, genes, and publications.
* **Ensembl**: Focuses on genome annotation and comparative genomics.
* **PDB**: Stores 3D structures of proteins and nucleic acids.  
  Each serves a unique purpose in data storage and analysis.

**4. What are the practical applications of bioinformatics in fields such as agriculture?**

**Ans:** Bioinformatics is used for Identifying genes for drought or disease resistance, Developing genetically modified crops, Enhancing yield through genome-assisted breeding.

**5. How does bioinformatics support personalized medicine through genetic analysis?**

**Ans:** It helps analyze a patient’s genome to:

* Detect disease-causing mutations,
* Predict drug responses,
* Design personalized treatment plans for better outcomes.

**6. How has bioinformatics improved the study of hypothetical proteins?**

**Ans:** Bioinformatics tools can:

* Predict functions from sequence similarities,
* Model protein structures,
* Identify functional domains,  
  Thus giving insight into the role of hypothetical proteins.

1. **What are** **the primary differences between 1st, 2nd, and 3rd generation sequencing technologies?**

**Ans:** The primary differences between 1st, 2nd, and 3rd generation sequencing technologies are below:

* **1st Gen (Sanger)**: Accurate, low throughput, short reads.
* **2nd Gen (NGS)**: High throughput, short reads, lower cost.
* **3rd Gen**: Real-time, long reads, ideal for structural variation analysis.

**9. What** **are the practical applications of genome sequencing in fields like medical and agriculture?**

**Ans:** The practical applications of genome sequencing in fields like medical and agriculture are written below:

* **Medical**: Disease diagnosis, cancer genomics, drug development.
* **Agriculture**: Crop improvement, disease resistance, trait mapping.  
  Genome sequencing accelerates discovery in both fields.

**10. How do SRA, FASTQ, BAM, and VCF file formats contribute to genome data analysis workflows?**

* **SRA**: Stores raw sequencing reads.
* **FASTQ**: Contains sequences with quality scores.
* **BAM**: Stores aligned reads to reference genomes.
* **VCF**: Stores genetic variants.  
  Each plays a role in the NGS pipeline from raw data to variant analysis.

**11. What are the advantages and limitations of short-read sequencing versus long-read sequencing?**

* **Short-read**: High accuracy, low cost; but hard to resolve repeats.
* **Long-read**: Captures structural variants better, resolves complex regions; but may have higher error rates and cost.