Quiz-1

MCQ:

1) You want to work with the reference sequences of a tree frog. Which database will you use:
   a) Unigene from NCBI
   b) GEO from NCBI
   c) **RefSeq from NCBI**
   d) DDBJ from Japan

2) Rosalind Franklin played a crucial role in the discovery of the DNA double helix structure, yet her contributions were often overshadowed. Which of the following statements about Rosalind Franklin is accurate?
   a) She was awarded the Nobel Prize in Physiology or Medicine for her work on DNA structure.
   b) **Franklin's X-ray crystallography images provided the first evidence of the DNA double helix.**
   c) She actively collaborated with James Watson and Francis Crick in the development of the DNA model.
   d) Franklin's research primarily focused on RNA rather than DNA structure.

3) If in a hypothetical biological system DNA directly produced functional proteins without the involvement of RNA, how does this observation challenge the central dogma of molecular biology?
   a) It supports the central dogma by demonstrating the linear flow of genetic information from DNA to RNA to protein.
   b) **It challenges the central dogma by suggesting a direct link between DNA and protein synthesis, bypassing the RNA intermediate.**
   c) It aligns with the central dogma, indicating that RNA is still essential for protein production.
   d) It confirms the traditional view that RNA acts solely as a messenger between DNA and proteins.

4) In the context of genetic databases, what is the relationship between GeneBank and NCBI, and how do they contribute to the storage and retrieval of genetic information?
   a) GeneBank and NCBI are unrelated; each serves as an independent repository for genetic data.
   b) **GeneBank is a part of NCBI, and together they facilitate the storage and retrieval of genetic sequences.**
   c) NCBI is a subsection of GeneBank, specializing in non-genetic data storage.
   d) GeneBank and NCBI operate separately, but share data periodically to ensure consistency.

5) If a DNA strand has the following sequence, choose the answer that has the correct mRNA sequence. 3' ATTAGCAC 5'
   a) **UAAUCGUG**
   b) TAATCGTG
   c) AUUAGCAC
   d) CACGATTA

6) Choose the correct flow of the events in central dogma
   a) **DNA > RNA > Protein**
   b) Translation > Transcription > Protein synthesis
   c) RNA > DNA > Protein
   d) Protein > RNA > DNA

7) What is Biopython?
   a) A programming language
   b) **A library for biological data analysis**
   c) A database of biological sequences
   d) A tool for gene editing

8) What function would you use to convert DNA to amino acid sequences in Biopython?
- a) **<u>transcribe()</u>**
- b) reverse_complement()
- c) **<u>translate()</u>**
- d) complement()

9) Which of the following about promoters in gene regulation is correct?
- a) Promoters are located downstream of the gene they regulate.
- b) **<u>Promoters are sequences on DNA that only influence transcription initiation</u>**
- c) Promoters are sequences on mRNA that only influence splicing
- d) Promoters are primarily responsible for the initiation of translation.

10) Considering the TATA box as a specific signal initiating transcription. How might a weakened TATA box influence the transcription?
- a) Transcription would proceed at an accelerated pace.
- b) **<u>Transcription initiation might be delayed or reduced.</u>**
- c) The TATA box has no impact on transcription.
- d) RNA polymerase would bypass the gene, representing a failure of the TATA box signal.

Close-end:

1) Imagine you have to identify the drugs through genomic study for a disease. What is the field of study from below you will choose and why?
- a) Genomics b) **<u>Pharmacogenomics</u>** c) Cheminformatics d) Pharmacogenetics

2) If you are having a gene called 'SOD1' in the species 'Homo Sapiens'. How will you retrieve its Nucleic acid sequence? Write a short snippet in Python/R for the same.

Answer –

```
from Bio import Entrez, SeqIO

Entrez.email = "your@email.com"  # Provide your email for NCBI request

gene_name = "SOD1"

search_term = f"{gene_name}[Gene] AND Homo sapiens[Organism]"

search_handle = Entrez.esearch(db="nucleotide", term=search_term)

gene_id = Entrez.read(search_handle)["IdList"][0]

sequence_handle = Entrez.efetch(db="nucleotide", id=gene_id, rettype="gb", retmode="text")

gene_sequence = SeqIO.read(sequence_handle, "genbank").seq

sequence_handle.close()

print(gene_sequence)
```

##If Entrez, SeqIO, efetch, with logical operators.

3) If the length of the DNA sequence is 3816, What would be the length of peptide?

Answer – 1272

4) A piece of DNA contains a higher amount of A and T bases than average. What can you comment about this thermal stability and why?

Answer – lower , because the number of hydrogen bonds are less.

5)   Write any two differences between **FASTA** and **GENBANK** file format

   Answer – a) FASTA format must begin with ">" followed by a unique SeqID (sequence identifier), immediately following the description line are the sequence representation.
   b) It contains minimal annotation or metadata.

   a)   **GENBANK is a comprehensive format used to store biological sequence data along with extensive metadata.**
   b)   **It contains sequence data, feature tables describing genes and other elements, organism information, references, and other annotations.**
   c)   **They are structured and can include detailed information about the sequence, such as gene locations, coding regions, promoters, and other functional elements.**

6)   Consider the DNA sequence "ATCGATCGATCGATCGA". If we perform slicing using seq[2:8], the extracted subsequence is **"CGATCG"**.

7)   True/False: Versioning and annotations are irrelevant in bioinformatics databases, as the stored sequences remain constant over time.
   Answer – False.

8)   If a DNA sequence has a high content of adenine (A), what nucleotide will be high in its complementary RNA sequence after transcription? Answer in one line.
   Answer - The RNA sequence will have a high uracil content.

9)   Differentiate between Origin of Replication (ORI) and a Transcription Start Site (TSS) in one line.

   Answer - The Origin of Replication (ORI) is a DNA region where DNA replication initiates, while a Transcription Start Site (TSS) is a site where RNA polymerase initiates transcription

10) True/False: An accession number is a unique identifier assigned to a biological sequence or a record in a biological database to facilitate easy and unambiguous reference.
   Answer - True