CIS*6060 Assignment 1

Question 1: Transcription and Translation

Given the partial DNA coding segment from Homo sapiens CD4 is

5'ACCGGGGAGTCCCTTTTAGGCACTTGCTTCTGG3'

1) Produce corresponding RNA fragment:
   For producing the RNA fragment from given DNA segment we need to transcribe DNA sequence into RNA. During transcription DNA is read and RNA is synthesized is the same direction from left to right. The only difference is that the T (Thymine) in DNA is converted to U (Uracil) in RNA fragment.

   Given DNA segment:

   5'ACCGGGGAGTCCCTTTTAGGCACTTGCTTCTGG3'

   RNA fragment:

   5' ACCGGGGAGUCCCUUUUAGGCAUUGCUUCUGG 3'

2) Produce the partial primary protein sequence using reading frame 3:
   Reading frame 3 means we start reading the RNA fragment from 3rd nucleotide.
   The RNA fragment is divided into groups of 3 nucleotides each representing a specific amino.
   They are as follows: CGG GGA GUC CCU UUU AGG CAU UGC UUC UGG

RNA fragment	Amino
CGG	Arginine (R)
GGA	Glycine (G)
GUC	Valine (V)
CCU	Proline (P)
UUU	Phenylalanine (F)
AGG	Arginine (R)
CAU	Histidine (H)
UGC	Cysteine (C)
UUC	Phenylalanine (F)
UGG	Tryptophan (W)

Therefore, partial primary protein sequence using reading frame:

RGVPFRHCFW

Question 2: Summary

Combating the 2014 Ebola Outbreak using Bioinformatics:

The 2014 Ebola virus outbreak in West Africa was one of the deadliest epidemics in human history claiming approximately 11,000 lives. This showed the urgent need for advanced tools to understand and control the spread of this disease. Bioinformatics emerged as a critical field during this outbreak, enabling analysis of the Ebola virus genome and how to tackle this problem.

Overview: -

The Ebola virus outbreak began in Guinea in December 2013 and spread to other African nations like Liberia, Sierra Leone, etc. The virus caused severe hemorrhagic fever, had a high mortality rate, and put a strain on the healthcare systems in the affected regions. Traditional treatments could not contain the spread of the virus. This is where Bioinformatics came into the picture showing the use of advanced genomic and computation tools.

Bioinformatics Tools and Application in the 2014 Ebola Outbreak: -

Bioinformatics played an important role in decoding the Ebola virus. The whole-genome sequencing (WGS) tool helped scientists decode the genetic material of the virus from human samples. These sequences after decoding were fed into Bioinformatics software to identify mutations and trace the virus evolution. The most famous method used was the phylogenetic analysis to construct evolutionary trees in the Ebola virus. By comparing viral genomes from patients, scientists determined the cause of the Ebola virus which was a new variant of Zaire ebolavirus species that was present around 2004. This suggested that the virus was among animals before making its way into the human body.

Another important application was the use of molecular clock analysis, which estimates the rate of mutation over time. This helped scientists understand how fast the virus was mutating inside humans.

Significance and New Knowledge: -

The use of Bioinformatics during the Ebola virus outbreak highlighted several findings allowing scientists to keep track of its mutations and it also helped identify specific mutations that have resulted in its increased spread in humans. These applications helped the development of diagnostics, vaccines, therapeutics, etc. The most notable vaccine developed rVSV-ZEBOV vaccine, which has been used since then to control the Ebola virus.

This outbreak demonstrated the power of global collaboration in bioinformatics. Scientists from around the world shared genomic data and analysis using platforms like GenBank and GISAID. This collaboration was like what was seen during the 2011 E. coli outbreak and set a precedent for future epidemics.

Conclusion:

The 2014 Ebola virus outbreak showed the role of bioinformatics in addressing infectious diseases. By using bioinformatics techniques like genomic sequencing and analysis scientists learned about virus evolution, transmission, and biology. This helped in controlling the outbreak and showed how more effectively we can act for future epidemics. This Ebola case study serves as an example of how bioinformatics can bridge the gap between genomics and public health.

Question 3: Reading and summarizing fasta-file

asn1e3.py README3.txt

asn1e3.py > ...

```python
# CIS*6060 Bioinformatics
# Assignemnt 1 Question 3
# Under the guidance of : Prof. Yan Yan
# Atharva Vichare
# 1320832

from Bio import SeqIO

#This functions summarizes the contents of a multi-FASTA file by displaying the header, first 10 amino acids, and the total number of ami
def summarize_fasta(fasta):
    # Parse the FASTA file
    for record in SeqIO.parse(fasta, "fasta"):
        header = record.description
        sequence = str(record.seq)
        first_10 = sequence[:10]
        length = len(sequence)
        print(f">{header} {first_10} : {length}")

if __name__ == "__main__":
    fasta = input("Please input path of fasta file: ")
    summarize_fasta(fasta)
    print("Yayy you did it. Congrats :) :)")
```

Fig 1: Code snippet

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS Python

```
PS C:\Users\avich\OneDrive\Desktop\Bioinformatics> & C:/Users/avich/AppData/Local/Programs/Python/Python312/python.exe c:/Users/avich/OneDrive/Desktop/Bioinformatics/
asn1e3.py
Please input path of fasta file: C:\Users\avich\OneDrive\Desktop\Bioinformatics\multiprotein.fasta
>1433G_HUMAN (P61981) 14-3-3 protein gamma (Protein kinase C inhibitor protein 1) (KCIP-1) [Homo sapiens] VDREQLVQKA : 246
>ATP8_RAT (P11608) ATP synthase protein 8 (EC 3.6.3.14) (ATPase subunit 8) (A6L) (Chargerin II) [Rattus norvegicus] MPQLDTSTWF : 67
>ALR_LISIN (Q92DC9) Alanine racemase (EC 5.1.1.1) MVTGWHRPTW : 368
>CDCA4_HUMAN (Q9BXL8) Cell division cycle-associated protein 4 (Hematopoietic progenitor protein) [Homo sapiens] MFARGLKRKC : 241
Yayy you did it. Congrats :) :)
PS C:\Users\avich\OneDrive\Desktop\Bioinformatics>
```

Fig 2: Output