Bhanuteja Pujari

Lab 05: Central Dogma I: Integrated pipeline for Building Web-based DNA Translation Tool

Your Task

To receive full credit for this lab, you must **answer** all following questions.

Task 1: Get protein sequence for DNA sequence (https://www.ncbi.nlm.nih.gov/nuccore/NM_207618.2), and compare it with sequence on NCBI website. Do you get same results?

```
GGTCAGAAAAGCCCTCTCCATGTCTACTCACGATACATCCCTGAAAACCACTGAGGAAGTGGC
TTTTCAGATCATCTTGCTTTGCCAGTTTGGGGTTGGGACTTTTGCCAATGTATTTCTCTTTGTC
TATAATTTCTCTCCAATCTCGACTGGTTCTAAACAGAGGCCCAGACAAGTGATTTTAAGACACA
TGGCTGTGGCCAATGCCTTAACTCTCTTCCTCACTATATTTCCAAACAACATGATGACTTTTGC
TCCAATTATTCCTCAAACTGACCTCAAATGTAAATTAGAATTCTTCACTCGCCTCGTGGCAAGA
AGCACAAACTTGTGTTCAACTTGTGTTCTGAGTATCCATCAGTTTGTCACACTTGTTCCTGTTA
ATTCAGGTAAAGGAATACTCAGAGCAAGTGTCACAAACATGGCAAGTTATTCTTGTTACAGTTG
TTGGTTCTTCAGTGTCTTAAATAACATCTACATTCCAATTAAGGTCACTGGTCCACAGTTAACA
GACAATAACAATAACTCTAAAAGCAAGTTGTTCTGTTCCACTTCTGATTTCAGTGTAGGCATTG
TCTTCTTGAGGTTTGCCCATGATGCCACATTCATGAGCATCATGGTCTGGACCAGTGTCTCCAT
GGTACTTCTCCTCCATAGACATTGTCAGAGAATGCAGTACATATTCACTCTCAATCAGGACCCC
AGGGGCCAAGCAGAGACCACAGCAACCCATACTATCCTGATGCTGGTAGTCACATTTGTTGGCT
TTTATCTTCTAAGTCTTATTTGTATCATCTTTTACACCTATTTTATATATTCTCATCATTCCCT
GAGGCATTGCAATGACATTTTGGTTTCGGGTTTCCCTACAATTTCTCCTTTACTGTTGACCTTC
AGAGACCCTAAGGGTCCTTGTTCTGTGTTCTTCAACTGTTGAAAGCCAGAGTCACTAAAAATGC
CAAACACAGAAGACAGCTTTGCTAATACCATTAAATACTTTATTCCATAAATATGTTTTTAAAA
GCTTGTATGAACAAGGTATGGTGCTCACTGCTATACTTATAAAGAGTAAGGTTATAATCACTT
GTTGATATGAAAGATTTCTGGTTGGAATCTGATTGAAACAGTGAGTTATTCACCACCCTCCAT
TCTCT
```

- Your Protein Sequence:
MSTHDTSLKTTEEVAFQIILLCQFGVGTFANVFLFVYNFSPISTGSKQRPRQVILRHMAVA
NALTLFLTIFPNNMMTFAPIIPQTDLKCKLEFFTRLVARSTNLCSTCVLSIHQFVTLVPVNS
GKGILRASVTNMASYSCYSCWFFSVLNNIYIPIKVTGPQLTDNNNNSKSKLFCSTSDFSV
GIVFLRFAHDATFMSIMVWTSVSMVLLLHRHCQRMQYIFTLNQDPRGQAETTATHTILM
LVVTFVGFYLLSLICIIFYTYFIYSHHSLRHCNDILVSGFPTISPLLLTFRDPKGPCSVFFNC

- NCBI's Protein Sequence:
```
MSTHDTSLKTTEEVAFQIILLCQFGVGTFANVFLFVYNFSPIST
GSKQRPRQVILRHMAVANALTLFLTIFPNNMMTFAPIIPQTDLKCKLEFFTRLVARST
NLCSTCVLSIHQFVTLVPVNSGKGILRASVTNMASYSCYSCWFFSVLNNIYIPIKVTG
PQLTDNNNNSKSKLFCSTSDFSVGIVFLRFAHDATFMSIMVWTSVSMVLLLHRHCQRM
QYIFTLNQDPRGQAETTATHTILMLVVTFVGFYLLSLICIIFYTYFIYSHHSLRHCND
ILVSGFPTISPLLLTFRDPKGPCSVFFNC
```

Task 2: Get protein sequence for DNA sequence (https://www.ncbi.nlm.nih.gov/nuccore/ NM_008220.5), and compare it with sequence on NCBI website. Do you get same results?

```
TTACGTTTGCTTCTGATTCTGTTGTGTTGACTTGCAACCTCAGAAACAGACATCATGGTGCACC
TGACTGATGCTGAGAAGGCTGCTGTCTCTGGCCTGTGGGGAAAGGTGAACGCCGATGAAGTTGG
TGGTGAGGCCCTGGGCAGGCTGCTGGTTGTCTACCCTTGGACCCAGCGGTACTTTGATAGCTTT
GGAGACCTATCCTCTGCCTCTGCTATCATGGGTAATGCCAAAGTGAAGGCCCATGGCAAGAAG
TGATAACTGCCTTTAACGATGGCCTGAATCACTTGGACAGCCTCAAGGGCACCTTTGCCAGCCT
CAGTGAGCTCCACTGTGACAAGCTGCATGTGGATCCTGAGAACTTCAGGCTCCTGGGCAATATG
ATCGTGATTGTGCTGGGCCACCACCTGGGCAAGGATTTCACCCCCGCTGCACAGGCTGCCTTCC
AGAAGGTGGTGGCTGGAGTGGCTGCTGCCCTGGCTCACAAGTACCACTAAGCCCCTTTTCTGCT
ATTGTCTATGCACAAAGGTTATATGTCCCCTAGAGAAAAACTGTCAATTGTGGGGAAATGATGA
AGACCTTTGGGCATCTAGCTTTTATCTAATAAATGATATTTACTGTCATCTCAAAAAAAAAAAA
AAAAAA
```

- Your Protein Sequence:

MVHLTDAEKAAVSGLWGKVNADEVGGEALGRLLVVYPWTQRYFDSFGDLSSASAIMG
NAKVKAHGKKVITAFNDGLNHLDSLKGTFASLSELHCDKLHVDPENFRLLGNMIVIVLG
HHLGKDFTPAAQAAFQKVVAGVAAALAHKYH

- NCBI's Protein Sequence:

MVHLTDAEKAAVSGLWGKVNADEVGGEALGRLLVVYPWTQRYFD
SFGDLSSASAIMGNAKVKAHGKKVITAFNDGLNHLDSLKGTFASLSELHCDKLHVDPE
NFRLLGNMIVIVLGHHLGKDFTPAAQAAFQKVVAGVAAALAHKYH

Task 3: Get protein sequence for DNA sequence (https://www.ncbi.nlm.nih.gov/nuccore/ NM_025580.2), and compare it with sequence on NCBI website. Do you get same results?

```
CGACGGGGAGCGCAGTGAAGGGGGGTGGGACCCGAACATGGCGGCGGTGGTAGCTGCTACGGCG
CTGAAGGGCCGGGGGGCGAGAAATGCCCGCGTCCTCCGGGGGATCCTCTCTGGAGCCACAGCTA
ACAAGGCTTCCCAGAACAGGACCAGAGCACTGCAGAGCCACAGCTCACCAGAATGCAAGGAGGA
GCCCGAGCCCCTCTCCCCTGAGCTAGAATACATTCCCAGAAAGAGGGGCAAGAACCCCATGAAA
GCCGTGGGGCTAGCCTGGGCCATCGGCTTCCCCTGTGGTATCCTCTTCTTCGTCCTCACCAAGC
AGGAAGTGGACAAGGACCGATTGAAGCAGATGAAAGCTCGTCAGAACATGCGGGTGTCCAACAC
GGGCGAGTATGAGAGCCAGAGGTTCAGGGCTTCACCCCAACAAGCCCAGTTCCCTGAGGTTGGG
TCTGGGGCCCAGACCTGAGTGAGGATCACTGCAACCTTGTGCTAGACTATTGACTGCCGTCTTG
TGTGGCCATGGCTCCTGAGGCCCACCAGAGGGCGCAGGATGCCCAAGCTGCTGCCAGCACCCAA
CTCTTCCACACCGTGGAAGGAGCCAAATGCAAATAAAGTTATTAAGTACTGACACGGAAAAAAA
AA
```

- Your Protein Sequence:

MAAVVAATALKGRGARNARVLRGILSGATANKASQNRTRALQSHSSPECKEEPEPLSPEL
EYIPRKRGKNPMKAVGLAWAIGFPCGILFFVLTKQEVDKDRLKQMKARQNMRVSNTGE
YESQRFRASPQQAQFPEVGSGAQT

- NCBI's Protein Sequence:

```
MAAVVAATALKGRGARNARVLRGILSGATANKASQNRTRALQSH
SSPECKEEPEPLSPELEYIPRKRGKNPMKAVGLAWAIGFPCGILFFVLTKQEVDKDRL
KQMKARQNMRVSNTGEYESQRFRASPQQAQFPEVGSGAQT
```

Task 4: Get protein sequence for DNA sequence (https://www.ncbi.nlm.nih.gov/nuccore/NM_009735.3), and compare it with sequence on NCBI website. Do you get same results?

```
TTTCAGTGGCTGCTACTCGGCGCTTCAGTCGCGGTCGCTTCAGTCGTCAGCATGGCTCGCTCGG
TGACCCTGGTCTTTCTGGTGCTTGTCTCACTGACCGGCCTGTATGCTATCCAGAAAACCCCTCA
AATTCAAGTATACTCACGCCACCCACCGGAGAATGGGAAGCCGAACATACTGAACTGCTACGTA
ACACAGTTCCACCCGCCTCACATTGAAATCCAAATGCTGAAGAACGGGAAAAAAATTCCTAAAG
TAGAGATGTCAGATATGTCCTTCAGCAAGGACTGGTCTTTCTATATCCTGGCTCACACTGAATT
CACCCCCACTGAGACTGATACATACGCCTGCAGAGTTAAGCATGCCAGTATGGCCGAGCCCAAG
ACCGTCTACTGGGATCGAGACATGTGATCAAGCATCATGATGCTCTGAAGATTCATTTGAACCT
GCTTAATTACAAATCCAGTTTCTAATATGCTATACAATTTATGCACGCAGAAAGAAATAGCAAT
GTACACATCACCTTCTTTATATCTTACTTTAAATATTTTATGCATGTTTTCAAAAAATTGGAAA
TATCCTAGATAGCTGAGCAATAAATCTTCAATAAGTATTTTGATCAGAATAATAAATATAATTT
TAAGAACAATAGTTGATCATATGCCAAACCCTCTGTACTTCTCATTACTTGGATGCAGTTACTC
ATCTTTGGTCTATCACAACATAAGTGACATACTTTCCTTTTGGTAAAGCAAAGAGGCCTAATTG
AAGTCTGTCACTGTGCCCAATGCTTAGCAATTCTCACCCCCAACCCTGTGGCTACTTCTGCTTT
TGTTACTTTTACTAAAAATAAAAAAC
```

- Your Protein Sequence:

MARSVTLVFLVLVSLTGLYAIQKTPQIQVYSRHPPENGKPNILNCYVTQFHPPHIEIQMLK
NGKKIPKVEMSDMSFSKDWSFYILAHTEFTPTETDTYACRVKHASMAEPKTVYWDRDM

- NCBI's Protein Sequence:

MARSVTLVFLVLVSLTGLYAIQKTPQIQVYSRHPPENGKPNILN
CYVTQFHPPHIEIQMLKNGKKIPKVEMSDMSFSKDWSFYILAHTEFTPTETDTYACRV
KHASMAEPKTVYWDRDM

Task 5: Get protein sequence for DNA sequence (https://www.ncbi.nlm.nih.gov/nuccore/NM_011662), and compare it with sequence on NCBI website. Do you get same results?

GTCACACCAGGTCCCACCAGCCCCTGGACTGTGGTGTCCAGTGCATATCTGGCCACCATGGGGG
CTCTGGAGCCCTCCTGGTGCCTTCTGTTCCTTCCTGTCCTCCTGACTGTGGGAGGATTAAGTCC
CGTACAGGCCCAGAGTGACACTTTCCCAAGATGCGACTGTTCTTCCGTGAGCCCTGGTGTACTG
GCTGGGATTGTTCTGGGTGACTTGGTGTTGACTCTGCTGATTGCCCTGGCTGTGTACTCTCTGG
GCCGCCTGGTCTCCCGAGGTCAAGGGACAGCGGAAGGGACCCGGAAACAACACATTGCTGAGAC
TGAGTCGCCTTATCAGGAGCTTCAGGGTCAGAGACCAGAAGTATACAGTGACCTCAACACACAG
AGGCAATATTACAGATGAGCCCACTCTATGCCCATCAGCGGCCTGATGCCCGGATCCGGTCATT
CCAGATGCCTACTCAACAAGCCCTCTCTGGGATCAGGACTCCCGTTGGAATACAGATCCACAGG
GTACCTCCCTGAGATATCTGACATTGTACCATTTCTGTCCCCAAATAAAAGACAGAGCAAAAAA
AAAAAAAAAAAAAAA

- Your Protein Sequence:
MGALEPSWCLLFLPVLLTVGGLSPVQAQSDTFPRCDCSSVSPGVLAGIVLGDLVLTLLIA
LAVYSLGRLVSRGQGTAEGTRKQHIAETESPYQELQGQRPEVYSDLNTQRQYYR

- NCBI's Protein Sequence:

MGALEPSWCLLFLPVLLTVGGLSPVQAQSDTFPRCDCSSVSPGV
LAGIVLGDLVLTLLIALAVYSLGRLVSRGQGTAEGTRKQHIAETESPYQELQGQRPEV
YSDLNTQRQYYR

Task 6: Give a brief overview of your work, discussing any thoughts, successes or challenges you faced.

On the code portion for this project, I have made the mistake of implementing a return function (incorrectly) in the middle of the code. Hence, when I inputed a large string of a DNA sequence with stop codons, the results obtained contained the three letter amino acids and not the single letter final sequence. Hence, this was the one difficulty I faced during the course of this project. Other than that, the written worksheet was fairly feasible because the HTML code was provided and the loops needed to define the function 'get_protein_from_DNA' were already implemented in the previous lab.