

# Introduction to Bioinformatics HW 1

Submission deadline: 10 April 2022

PS: HWs submitted after the deadline will not be accepted

Please upload your HW to YTU-LMS:

Any question should be addressed to [naydin@yildiz.edu.tr](mailto:naydin@yildiz.edu.tr)

You must name your Perl file by using your matriculation number (for example, 07011068.pl)

Write a Perl program that takes a DNA sequence in FASTA format and implements DNA sequence translation to amino acid sequence process.

Please note that the input data file must be in **FASTA** format.

(An example data file that can be used as input is provided as **DNA-ex.txt**)

*FASTA format: A sequence file in FASTA format can contain several sequences.*

*One sequence in FASTA format begins with a single-line description, followed by lines of sequence data. The description line must begin with a greater-than (">") symbol in the first column.*

An example sequence in FASTA format is:

```
>U03518 Aspergillus awamori internal transcribed spacer 1 (ITS1)
AACCTGCGGAAGGATCATTACCGAGTGCGGGTCCCTTTGGGCCCAACCTCCCATCCGTGTC
TATTGTACCCTGTTGCTTCGGCGGGCCCCGCCGTTGTTCGGCCGCCGGGGGGGCGCCTCTG
CCCCCGGGCCCCGTGCCCGCCGGAGACCCCAACACGAACACTGTCTGAAAGCGTGCAGTC
TGAGTTGATTGAATGCAATCAGTTAAAACCTTCAACAATGGATCTCTTGGTTCCGGCATT
```

|              |   | Second letter                            |                                      |  |   |                  |  |
|--------------|---|--|--------------------------------------|--|---|------------------|--|
|              |   | U  | C                                    | A  | G   |                  |  |
| First letter | U | UUU } Phe<br>UUC }<br>UUA } Leu<br>UUG } | UCU }<br>UCC } Ser<br>UCA }<br>UCG } | UAU } Tyr<br>UAC }<br>UAA Stop<br>UAG Stop | UGU } Cys<br>UGC }<br>UGA Stop<br>UGG Trp | U<br>C<br>A<br>G |  |
|              | C | CUU }<br>CUC } Leu<br>CUA }<br>CUG }     | CCU }<br>CCC } Pro<br>CCA }<br>CCG } | CAU } His<br>CAC }<br>CAA } Gln<br>CAG }   | CGU }<br>CGC } Arg<br>CGA }<br>CGG }      | U<br>C<br>A<br>G |  |
|              | A | AUU }<br>AUC } Ile<br>AUA }<br>AUG Met   | ACU }<br>ACC } Thr<br>ACA }<br>ACG } | AAU } Asn<br>AAC }<br>AAA } Lys<br>AAG }   | AGU } Ser<br>AGC }<br>AGA } Arg<br>AGG }  | U<br>C<br>A<br>G |  |
|              | G | GUU }<br>GUC } Val<br>GUA }<br>GUG }     | GCU }<br>GCC } Ala<br>GCA }<br>GCG } | GAU } Asp<br>GAC }<br>GAA } Glu<br>GAG }   | GGU }<br>GGC } Gly<br>GGA }<br>GGG }      | U<br>C<br>A<br>G |  |

A=Ala=Alanine  
C=Cys=Cysteine  
D=Asp=Aspartic acid  
E=Glu=Glutamic acid  
F=Phe=Phenylalanine  
G=Gly=Glycine  
H=His=Histidine  
I=Ile=Isoleucine  
K=Lys=Lysine  
L=Leu=Leucine

M=Met=Methionine  
N=Asn=Asparagine  
P=Pro=Proline  
Q=Gln=Glutamine  
R=Arg=Arginine  
S=Ser=Serine  
T=Thr=Threonine  
V=Val=Valine  
W=Trp=Tryptophan  
Y=Tyr=Tyrosine