Documentation for the Codon Usage Frequency Table Tool

Tool Background

        Codons are sequences of three nucleotides (specifically, adenine (A), cytosine (C), guanine (G), and thymine (T)) found in messenger RNA (mRNA). They serve as the basic units of genetic instructions that determine the sequence of amino acids in a protein. Each codon corresponds to a specific amino acid or serves as a start or stop signal for protein synthesis. There are a total of 64 possible codons, and they are read during translation. These nucleotide triplets encode the information required for the synthesis of specific amino acids or serve as control signals. Triplets in DNA are transcribed into complementary codons in mRNA during the process of transcription. The mRNA codons are then translated into a sequence of amino acids during protein synthesis.

        Amino acids are the building blocks of proteins. There are 20 different amino acids commonly found in proteins. Each amino acid has a specific chemical structure and unique properties. The sequence of amino acids in a protein determines its structure and function. During translation, the codons in mRNA are matched to the corresponding amino acids, which are then joined together in a specific order to form a polypeptide chain, ultimately folding into a functional protein.

Tool Functionality

This Codon Usage Frequency Tool requires only one basic input: DNA sequence. The tool will convert the sequence into an amino acid sequence and create a table that will display triplets, amino acids, number of occurrences, and frequency of occurrence. The data in the table will be stored in MySQL.
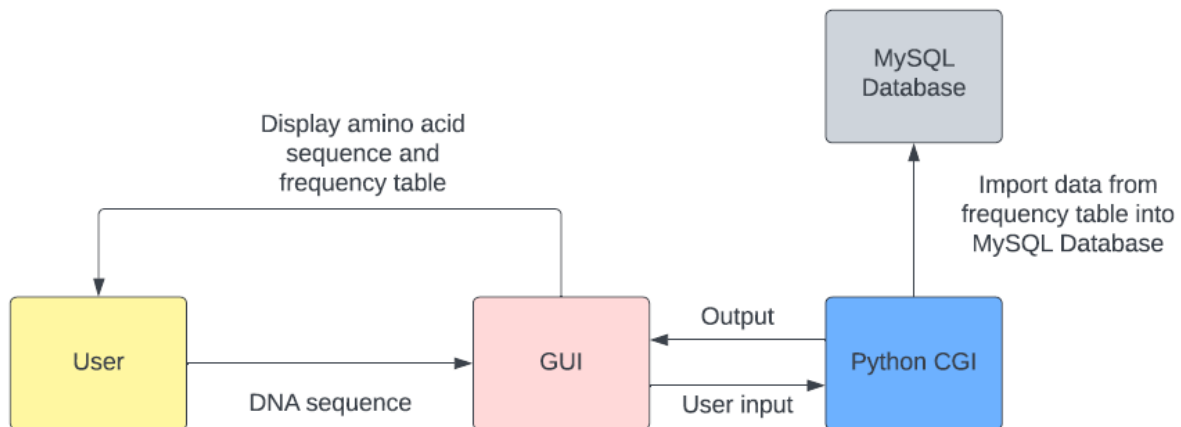
Example Codon Frequency Table

Triplet	Amino Acid	Number of Occurrences	Frequency of Occurrence
TTT	F	5	.1
TTC	F	10	.2
TTA	L	15	0.3
TTG	L	20	0.4
TAT	Y	5	.1
TAC	Y	10	.2
TAA	*	15	0.3
TAG	*	20	0.4

Tool Description

This Codon Usage Frequency Table Tool will utilize three main software technologies to carry out its functionality:

1) **SQL relational database/table** – to store all the generated information from the table in a database

2) **Python-based Computer Gateway Interface (CGI)** – to carry out the conversion of DNA to amino acid sequence, count the number of occurrences of each codon, and calculate the frequency of each codon

3) **CSS/HTML/JavaScript-based graphic user interface (GUI)** – provides a simple, user-friendly and interactive user interface for user's input and visualization of the output after backend data generation and shuttle data.

Tool Design and Development

1) A basic relational database (vngo6) will created
2) A CSS/HTML/JavaScript-based GUI will be created to allow users to input their DNA sequence.
3) A python code will be written to convert the DNA sequence to amino acid sequence. It will then analyze the sequence by counting the number of occurrences of triplets and frequency of each triplet. This information will be sent back to the GUI via JSON.

Example HTML

Please enter your DNA sequence here:

[text input box]

Submit

Results:

Amino Acid Sequence

[text input box]

Example Codon Frequency Table

Triplet	Amino Acid	Number of Occurrences	Frequency of Occurrence
TTT	F	5	.1
TTC	F	10	.2
TTA	L	15	0.3
TTG	L	20	0.4
TAT	Y	5	.1
TAC	Y	10	.2
TAA	*	15	0.3
TAG	*	20	0.4