

William Mleziva

5481 - Homework 1

Task 7)

The most striking difference is with codons TAA, TAG, TAA. These are the least frequent in the Coding sequences and are all stop codons. Since they are all stop codons they can only occur once at the end of a coding sequence. However, with a random frameshift these can occur commonly and are shown to. Additionally, there are many codons that are significantly more common in the coding sequences such as GCT and GGT. These are less common in the whole genome as a random frameshift makes the codon distribution fairly random.

Other tasks and how to run below

Task 2)

Code is in the file: 'count_codons.py'

Task 3)

Test fasta file is named 'demoseq.fna'

Test file output (expected & actual) is named 'demoOutput.csv'

The program is executed with:

```
python3 count_codons.py demoseq.fna demoOutput.csv
```

Task 4)

Whole Genome:

```
python3 count_codons.py SARS-CoV-2_whole_genome.fna WholeGenome.csv
```

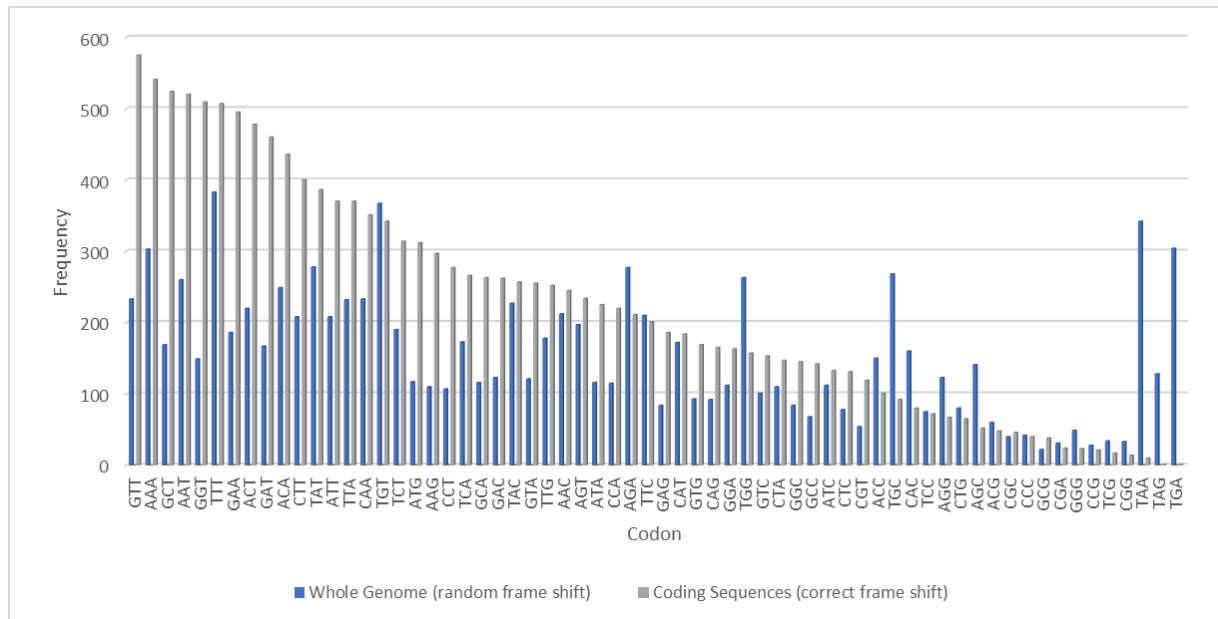
Split Genome:

```
python3 count_codons.py SARS-CoV-2_separate_genes.fna SplitGenome.csv
```

Output files are therefore named 'WholeGenome.csv' and 'SplitGenome.csv'

Task 5)

Barplots made in excel. 'CodonFrequency.jpg'



Task 6)

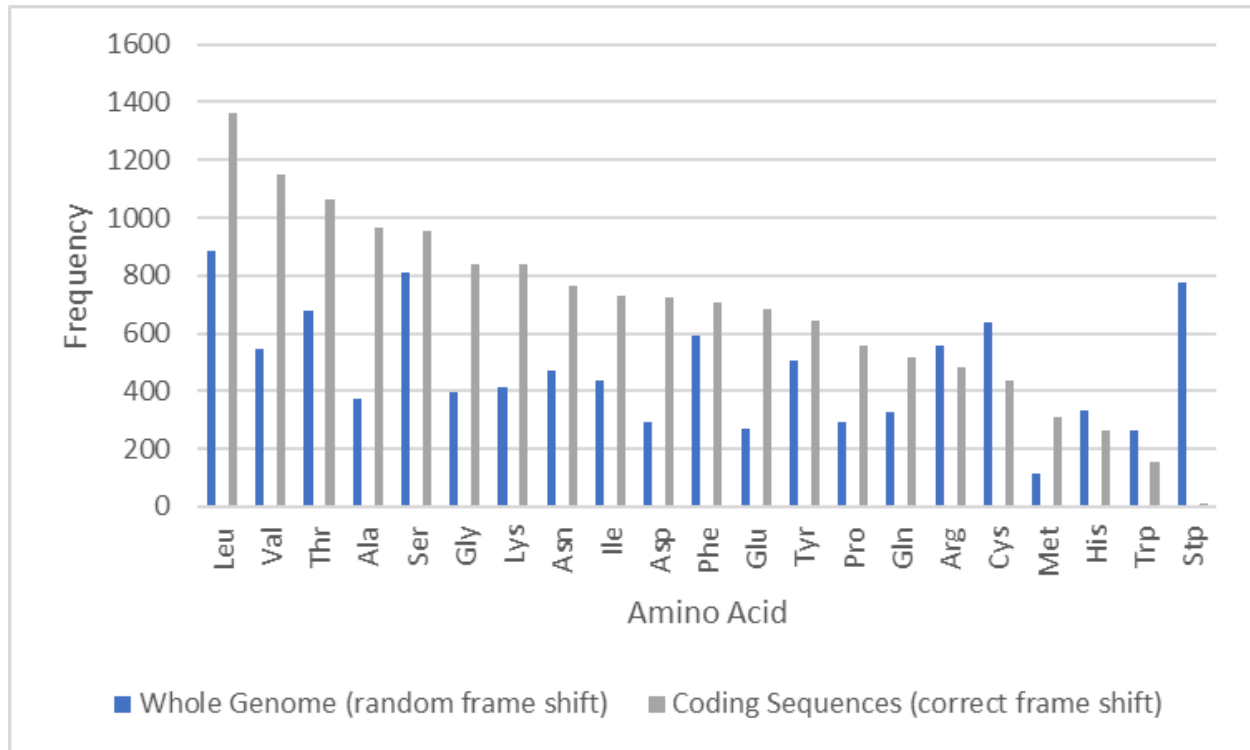
To convert to AA:

```
python3 CodonToAA.py WholeGenome.csv WholeGenome_AA.csv
```

```
python3 CodonToAA.py SplitGenome.csv SplitGenome_AA.csv
```

AA csv files are therefore named 'WholeGenome_AA.csv' and 'SplitGenome_AA.csv'

Barplot made in Excel. 'AminoAcidFrequency.jpg'



Task 7)

The most striking difference is with codons TAA, TAG, TGA. These are the least frequent in the Coding sequences and are all stop codons. Since they are all stop codons they can only occur once at the end of a coding sequence. However, with a random frameshift these can occur commonly and are shown to. Additionally, there are many codons that are significantly more common in the coding sequences such as GCT and GGT. These are less common in the whole genome as a random frameshift makes the codon distribution fairly random.