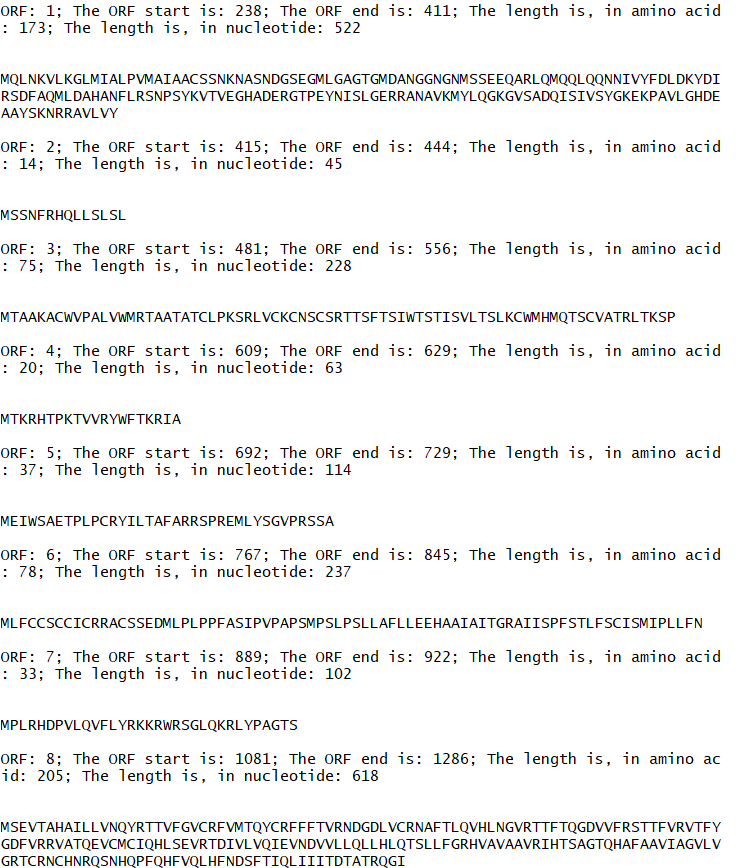
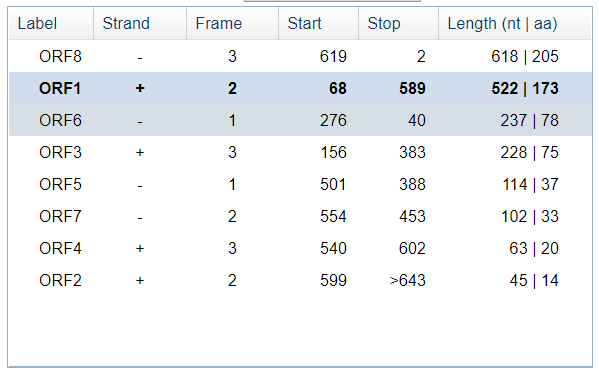
**Bioinformatics Assignment (Part 3) – Jennifer Nolan C16517636**

The following report will analyse the output from my program and the output from the NCBI ORF Finder.

My program, created in part two of this assignment, returned the below Open Reading Frames, seen in Figure 1, from the DNA sequence provided in the .fasta file. Mostly my programs output matches the output from the NCBI, shown in Figure 2.



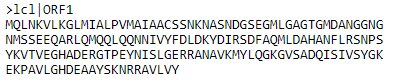
*Figure 1: Results from my program*



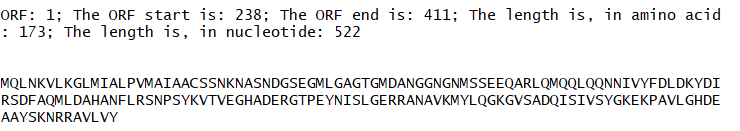
*Figure 2: NCBI results*

Firstly, the lengths of the Amino Acid Open Reading Frames from my program match those outputted from the NCBI application using the same DNA sequence. My program outputs the ORF’s in order of the ORF number, beginning with ORF1 and finishing with ORF8. Although the order is different to that of the NCBI, the corresponding ORF’s from my program and the NCBI output match. As well as the Amino Acid ORF lengths matching, the nucleotide ORF lengths outputted from my program also matched the nucleotide ORF lengths outputted from the NCBI application.

Secondly, the Amino Acid ORF sequences outputted from my program also match those outputted from the NCBI application. An example of the match can be seen below in Figures 3 and 4. My programs output, Figure 4, matches the amino acid output from the NCBI, Figure 3.

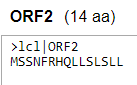


*Figure 3: ORF1 Amino Acid NCBI*



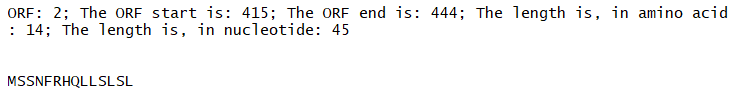
*Figure 4: ORF1 Amino Acid my program*

However, there was an issue with one of the Amino Acids. The NCBI application specifies in the table, shown in Figure 2, that ORF2 is 14 amino acids in length. However, when displaying the amino acid for ORF2, the NCBI application returns an amino acid that is 15 in length, as shown in Figure 5.



*Figure 5: ORF2 Amino Acid NCBI*

Whereas, my own program returns an amino acid that is 14 in length, as shown in Figure 6.



*Figure 6: ORF2 Amino Acid my program*

The only difference between the output from my program and the output from the NCBI application that I have noticed is the stop and start positions. As shown in Figure 1, my programs start and stop positions differ from those outputted from the NCBI, as shown in Figure 2. This is because my program outputs the amino acid start and stop positions and not the nucleotide start and stop positions, like the NCBI. I was able to get the amino acid start and stop positions easily but could not get the DNA nucleotide start and stop positions, especially with regards to the compliment. I struggled mostly with the compliment because I reversed the DNA sequence for the compliment so that it could be read from left to right and be used in the same functions as the original DNA sequence.