**Final Project Proposal**

Nikhil Panu, Tom Pavarini, Will Zvagelsky

We will create a tool that will align DNA strands to a database of known proteins and we will compare/contrast our approach to those of similar tools such as BLAST and PAUDA. Once our program is given a DNA strand, it will immediately generate the complement strand. Now, we will have both strands necessary for evaluation. Then, we will cover three cases for each strand, whether the read starts from the first, second, or third codon position. For each case, the program will determine which codon matches which protein by using the codon table (Figure 1). Once the protein sequence is determined, we will align it to an existing database and compare our result to the alignment determined by BLAST and PAUDA. BLAST is very flexible, it allows the user a number of possible inputs, our program will start by only accepting the genome.3 PAUDA runs approximately ten thousand times faster than BLASTX, hopefully our program can run as fast as BLASTX.2

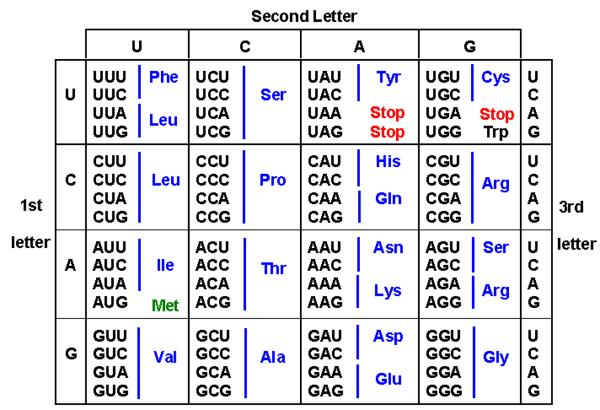


Figure 1

In order for our program to run, we will need a genome sequence that we can first transcribe into the corresponding mRNA sequence, and then translate into the corresponding protein sequence.

The main milestones that we need to accomplish are: (1) DNA transcription to mRNA, (2) mRNA translation to the corresponding protein, (3) translation from all six possible reading frames, (4) protein alignment with database. Once all of these goals have been met, then we can begin to expand the program by possibly being able to provide the tool a FASTA file which will first assemble the genome and then perform the typical tasks of transcription and translation. Another stretch goal would be to match the time complexity and speed of the BLASTX program.

References

1. <http://bitesizebio.com/21223/how-does-blast-work/>
2. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3866550/>
3. <http://www.ncbi.nlm.nih.gov/pubmed/2231712>