Homework 1

Handed out: 9/4/08
Due: 9/11/08

For this homework you will need to do a bit of research online. You'll have to look up IUB/IUPAC codes for DNA ambiguity codes as well as conversion tables between amino-acids and the corresponding codons (genetic code). **Please list your bibliographic (online, library, etc.) source for this information.**

1. What is the reverse complement of the following DNA sequences. Note that the third sequence contains IUB/IUPAC ambiguity codes (representing sets of 2 or more nucleotides):

ACAGGATGTTCATAGGCATTCCTCAGACTACAGTC

ACTTGCTAAGAATCTGATTCAGATTCTTAGCAAGT

GGCATGTCWAGACCTAMCYGACTCVGTAGGCCATG

2. What is the amino-acid sequence encoded in the following DNA sequence (assume gene starts at the first start codon and ends at first stop codon).

TTCGAGGGGCATGTTTGTTGCTATGAATGATAATAAAACAATGCTTTTTATTCCGGGGGCAACCAATTAAGTAATTC

(Trivia: this is a piece of one of the plague's - *Yersinia pestis* - toxic factors)

3. Match the following amino-acid sequence to the corresponding location in the DNA string shown below.

KLFALTAVALMG

GTATGAAAAAACTAAAATTGTTTGCTCTTACAGCTGTAGCCCTAATGGGTGTTTCAGGTGTA

(Trivia: this is a piece of a bacterial rhodopsin gene - one of the genes involved in photosynthesis. Until 2000 it was believed that only plants were capable of photosynthesis. The discovery of bacterial rhodopsin was done computationally and is one of the advances made possible by genomic analysis).

4. Define the following biological terms (look them up on the Internet and write out a one sentence definition in **your own words**):

- Frameshift mutation

- Silent mutation

5. Write a simple parser for FASTA files. Specifically, write a program, in your favorite programming language, that reads in a FASTA file, then identifies all records that contain more than 500 nucleotides and outputs their identifiers to the screen, one per line. Note, for this assignment you are not allowed to use any of the publicly available bioinformatics libraries, the entire code must be written using standard constructs from the programming language of your choice.

Note: The code you write must compile and run on the glue machines.

Deliverables: (i) Your source code; (ii) The output obtained by running your code on the file
/afs/glue.umd.edu/class/fall2008/cmsc/423/0101/public/homework1.fa