1) Sometimes information flows from RNA to DNA such as when you convert RNA into cDNA molecules for microarray analysis or for what we commonly call as RNA-sequencing using an illumina sequencer. Write a program to reverse transcribe RNA to DNA for the two fasta RNA sequence files (RNA-sequence1.fna and RNA-sequence2.fna). Print only the forward strand from 5' to 3' for each of the two reverse transcribed molecules. Please also compute the di-nucleotide and tri-nucleotide frequency (both absolute and percentage) of the nucleotides present in the original RNA sequences in the two files. The output should be in the following three column format for each pair of RNA base combinations for each input file:

   1) RNA dinucleotides or RNA trinucleotides
   2) Absolute Frequency
   3) Percentage

2) Now use this program to compare the differences in the percentage abundance of a dinucleotide and trinucleotide composition between the two sequences (in the above fasta files) to identify those, which differ by three fold (3X times) in percentage between pairs of sequences that are provided.

Hint: There are a total of 16 combinations of dinucleotides and 64 combinations of trinucleotides in an RNA sequence. Measurements are computed by scanning for overlapping nucleotides as shown in the below example AGGAC would yield two occurrences of A, two occurrences of G and one occurrence of C at the mono-nucleotide level. Likewise, one occurrence each for AG, GG, GA and AC di-nucleotides when you count for adjacent overlapping nucleotides.