Practice problems and concepts to know for for BNFO 236 exam one

1. Do all homeworks from scratch on your own – cannot emphasize this enough

2. Write a Python program to read in two aligned DNA sequences in FASTA format and print the number of matches and mismatches. For example if the input in the file dna.fasta is

   ```
   >human
   ACG--TA
   >mouse
   ACTCCTC
   ```

   then your program should output matches=3 and mismatches=2.

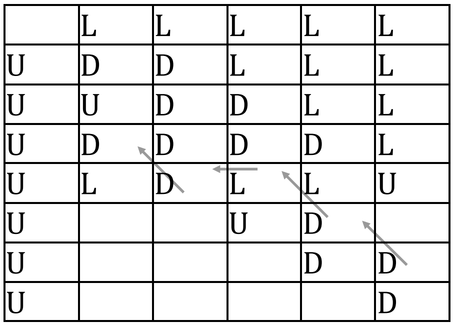
3. Write a program to translate a DNA sequence in a FASTA file into a protein sequence. You may assume the sequence length is a multiple of 3. Hint: if you use a dictionary your program will be shorter than otherwise.

4. Passing variables in Python by reference and by value. What is the difference?

5. Write a function that starts traceback from *i_start*, *j_start* and stops when i <= *i_end* or j <= *j_end* (5 pts). The function takes as parameters two DNA sequences, traceback matrix T, the start and end positions. Your function should return the two aligned sequences. The function definition is

```
def traceback(seq1, seq2, T, i_start, j_start, i_end, j_end):
```

As an example, suppose the traceback matrix is given by and i_start = 6, j_start = 5, i_end = 2, j_end = 1. Then the traceback path is given by the sequence of arrows below.



	L	L	L	L	L
U	D	D	L	L	L
U	U	D	D	L	L
U	D	D	D	D	L
U	L	D	L	L	U
U			U	D	
U				D	D
U					D

It starts at row 6 and column 5 and stops at row 3 and column 1. Note that the traceback stops because the column number is equal to to j_end even though the row number is greater than i_end.

For the above example the function would be called from the main program as

```
i_start = 6
i_end = 2
j_start = 5
j_end = 1
[aligned_s1, aligned_s2] = traceback(seq1, seq2, T, i_start, j_start,
i_end, j_end)
```

6. Suppose you are given a DNA sequence in a file called dna.fasta and random fragments of this sequence each of length 5 in a file called fragments.fasta. For example dna.fasta may contain

```
>Genome
ACACAGTGATGATTGAGGGGGGAGAGGACACACAGGGATTGAGATGGA
```

and fragments.fasta may be

```
>P0
GAGGG
>P1
ACAGT
>P2
TTGAG
>P3
AGGAC
```

Write a Python script that prints the fragment names in the correct order from left to right as they appear in the sequence in dna.fasta. For the above example the output would be

```
P1, P2, P0, P3
```

7. The FASTA non-interleaves format defines how to store DNA sequences in a file. For example

```
>human
ACCAGGGTAAACGTGGACAATCCGAAAATA
>mouse
AAAAGCGTTAACGTAGACAACCCGAAATTA
```

Consider the interleaved format where there are just 10 characters of a sequence per line:

```
>human
ACCAGGGTAA
ACGTGGACAA
TCCGAAAATA
>mouse
AAAAGCGTTA
ACGTAGACAA
CCCGAAATTA
```

Write a Python program that converts interleaved format to non-interleaved. Your script should print the interleaved format to the screen.