LAB 1: Manipulate fasta/fastq files in Python
ASSIGNMENTS

Assignment 1: Statistics extraction

Write a python program for extracting statistics from fasta/fastq files. The program must take as a first argument from the command line the name of the input fasta file to be analyzed and write to an output text file (whose name is passed as a second argument from the command line) a summary of the computed statistics.

The following are the expected output statistics:

- Statistics of single bases across all the reads: Number of A,T,C,G

- Number of reads having at least one low complexity sequence: AAAAAA, TTTTTT, CCCCCC or GGGGGG.

- Number of reads having the number of GC couples (so called **GC content**) higher than a threshold GC_THRESHOLD passed as third argument from the command line

- For each read having a GC content higher than GC_THRESHOLD, report the read_id and the number of GC couples

Assignment 2: Fasta comparison

Write a python program to compare two fasta files. The two fasta files are passed as first and second argument from the command line.

The two fasta files have the following characteristics:

-The fasta format of the two files is correct (no need to check the format)

-Each read can take up one or multiple lines

-Each input file does not contain duplicated reads (i.e. identical reads)

The program must write as output a third fasta file containing only the reads that are in common between the input files. The read ids in the output file should be composed by the read id of the first file concatenated with the read id of the second file.