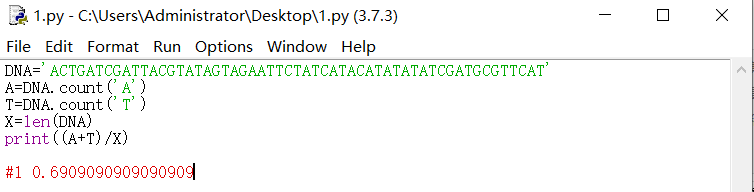
**实验课试题**

本试题共包含5道题，每道题编写程序后保存为一个以py结尾的文件。最后提交试题的文件是一个文件夹的压缩包，包括5个python程序文件，5个输入文件，2个输出文件，其中读取文件和写入文件的路径为相对路径。

题目需要将答案打印到屏幕上，答案必须在文件中以注释形式（#号）给出，并且代码中不含>>>，示例如下：



**第一题**

有一条短DNA序列seq1（在python中以seq1命名这个序列）

ATGCTTCCGGCTGTTCCCTGAGACCTCAAGTGTGAGGAATTCTGTACTATTGATGCTTCACACCT

1.将此条DNA序列互补成双链，计算双链氢键个数。

2.该序列含有两个外显子和一个内含子。外显子1从第一个脱氧核苷酸到第23个脱氧核苷酸，外显子2从第51个脱氧核苷酸到65个脱氧核苷酸。计算外显子总长度，并将序列以外显子大写，内含子小写的形式打印出来。

3.该DNA序列有EcoRI限制性内切酶的识别位点G＊AATTC（＊代表切割位点）。计算酶切后两个片段的大小，并将两个片段输出到屏幕上。

**第二题**

有多条DNA序列，读取文件input2.fa完成以下题目

1. 计算序列长度，并输出最长的序列到屏幕上。
2. 统计序列来自哪些物种，每个物种各有几条序列。

例：Homo sapiens 1条

**第三题**

以下是密码子与氨基酸的对应关系

'AUA':'I','AUC':'I','AUU':'I','AUG':'M',

'ACA':'T','ACC':'T','ACG':'T','ACU':'T',

'AAC':'N','AAU':'N','AAA':'K','AAG':'K',

'AGC':'S','AGU':'S','AGA':'R','AGG':'R',

'CUA':'L','CUC':'L','CUG':'L','CUU':'L',

'CCA':'P','CCC':'P','CCG':'P','CCU':'P',

'CAC':'H','CAU':'H','CAA':'Q','CAG':'Q',

'CGA':'R','CGC':'R','CGG':'R','CGU':'R',

'GUA':'V','GUC':'V','GUG':'V','GUU':'V',

'GCA':'A','GCC':'A','GCG':'A','GCU':'A',

'GAC':'D','GAU':'D','GAA':'E','GAG':'E',

'GGA':'G','GGC':'G','GGG':'G','GGU':'G',

'UCA':'S','UCC':'S','UCG':'S','UCU':'S',

'UUC':'F','UUU':'F','UUA':'L','UUG':'L',

'UAC':'Y','UAU':'Y','UAA':'','UAG':'',

'UGC':'C','UGU':'C','UGA':'','UGG':'W',

现有一条RNA序列seq2（在python中以seq2命名这个序列）

GCCAUUGUAAUGACGUGGGAAGGCCGCUGAAAGGGUGCCCGAUAG

1. 将此mRNA序列翻译成蛋白质序列（需要寻找起始密码子）。
2. 定义一个函数，有两个参数：蛋白质序列和氨基酸残基，返回蛋白质中该氨基酸的百分比含量。使用如下指令测试函数assert my\_function("MSRSLLLRFLLFLLLLPPLP","M") == 5，并计算1小题里蛋白质序列中氨基酸M的含量。

**第四题**

现有一KEGG聚类结果，读取文件完成以下题目。

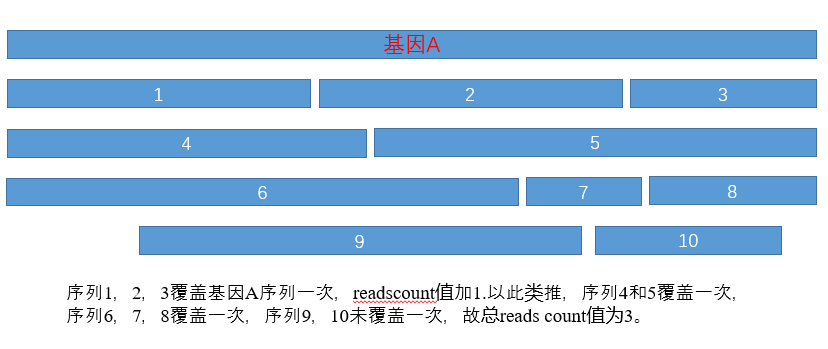
1. 找出Count数大于20的通路，并将通路名称打印到屏幕上。
2. 在结果中找出卵巢类固醇生成和雌激素信号通路，并将两个通路富集到的基因取交集，将交集里的基因打印到屏幕上。
3. 在表达量文件gene\_log2fc.txt中找到两个通路所包含的全部基因的表达差异倍数，将表达量上调的基因写入一个文件并命名为output4-up.txt，下调的写入一个文件并命名为output4-down.txt，文件每一行包含基因ID及其表达差异倍数，格式与输入文件保持一致。

**第五题**

在input5-1.fa和input5-2.fa文件中各有20条DNA短序列。基因A为：ACATCTCAAACTGGCAAAACTCAGTCTTAGCAGATTCAGTGTGGAAGCAGCTATCAAAAAGGCCATAAGGATTTTGTCCCCAAATTTCACATGAGCTACCTTGCTTCAAACTACTGAGATGAAGGGGGCAAGATTATTTGTCCTTCTTTCTAGTTTATGGAGTGGGGGCATTGGGCTTAACAACAGTAAGCATTCTTGGACTATACCTGA，基因B为：

ATTTCATTTCCACGCCAACCATCGTGCGCCGCGGTCTGAACGCTCCTGCCACAGAAAAAGAAAATAAAAGCAAGGAAAATTCTAATCGAATACCAAATATCGTGCTTGTGTGCTCTTTCCGCAATTGATTTTTTTTAAGTAGTGCATGACAATAACCGTTGAGTTGACTCCAACCGAAGTAACCATAAC，读取文件完成以下题目。

1. 计算基因A和基因B的readscount值。(共4个值，分别为input5-1.fa文件内基因A的readscount值，input5-1.fa文件内基因B的readscount值, input5-2.fa文件内基因A的readscount值, input5-2.fa文件内基因B的readscount值，按照顺序列出)



1. 计算基因A和基因B的log2fc值，并比较其大小，将二者较大的基因名及其log2fc值打印在屏幕上。（公式log2fc=log2（文件1基因A的reads count值/文件2基因A的reads count值））