

Date: 07-10-2019

Assignment # 01

Subject: Introduction to Bioinformatics

Total Marks: 100

Deadline for submission is **17:00 PKT, Monday 21st October, 2019**. Upload your code on Slate.

Question: The super healing of Markhors

According to the **Central Dogma of Molecular Biology**, a **gene** from a **genome** is first **transcribed** into a strand of **RNA**. Then, the **RNA** transcript is translated into an **amino acid** sequence of a **protein**. **DNA** (deoxyribonucleic acid) is a hereditary material in humans and almost all other organisms. Nucleotides in **DNA** contain four different nitrogenous bases: adenine (**A**), guanine (**G**), cytosine (**C**), and thymine (**T**). A **gene** is a segment of **DNA**. A **Genome** is an organism's complete set of **DNA**. The Markhor /'mɑ:rkɔ:r/ (Capra falconeri), is a large Capra species native to Central Asia, Karakoram and the Himalayas. It is listed on the IUCN Red List as Near Threatened since 2015. The Markhor is the national animal of Pakistan. It is also known as the screw horn goat, Pashto: مرغومی Margūmi and Persian/Urdu as مارخور .

Historically, in the ancient scriptures and folktales, Markhors are depicted to be miraculous holy animals – their blood could be used to clean a wound and flesh consumed by injured persons to help cure the disease instantaneously. Scientists have identified the exact protein that is responsible for this behavior. However, they are not sure of the exact **DNA** sequence which was eventually translated into this specific protein. They know the DNA sequence which is approximately same across all the Markhor species. The life of a bioinformatics scientist would be easy if these sequences were completely conserved across all the Markhors. But the reality is more complex, as these DNA subsequences may mutate, i.e., vary at some positions.

You are required to solve a much simpler version of this problem: you are given the DNA sequence corresponding to this protein, a chromosomal DNA sequence from Capra falconeri genome, and are asked to find the **best matching regions** between the genomes of a group of Markhors, **with exactly one mismatch**. **You are required to** discover these similar regions without any prior knowledge of how the regions look like.

So, the problem you are going to solve would be as follows: Given a random sample of *t* **DNA** sequences, find the **pattern** that is implanted in each of the individual **DNA** sequences, with exactly one **mutation** (variation).

As an example, consider the following randomly generated **DNA strings** with implanted pattern **TTTTCCC** of length **6**.

GAGGATTCGTTTGCCGAGG
AAATTTACCTAGAATGTCA
AGTTCCCCCTGGACACAAT
GAATTATCCCCAAGTCCTAA
AGACGAAGGTTCCCGTTGA

Notice that each of the implanted pattern has one mutation, as underlined.

Input

The input consists of multiple test cases. The first line of input is the number of test cases ($1 \leq N \leq 100$).

Each of the following N lines contains an integer t followed by an integer k , followed by a collection of t **DNA** strings. Note that both the t and k can have any value between 2 and 9, inclusive.

Output

Output *returns* all the **substrings**, each of which is of length k , and has exactly **one mutation**.

For each test configuration in the input first, output “Case n : ” where n is the test configuration number, followed by all the **substrings** of length k , separated by spaces.

Sample Input	Sample Output
4 4 2 TAT GGG CTC AGA 3 3 CGTTT GGTGG CCCCT 4 3 TAGACCG ACGGAAT GCCATAG CTTTAC 4 4 AAATACTG TACCAGTT GTCAAGGG AAATACGT	Case 1: GT TG Case 2: CGT GCT Case 3: GAC TAA TAT Case 4: AAGT CAAT

Disclaimer: The question is taken from a recent quiz competition held in the university.

xxx----- Good Luck! -----xxx