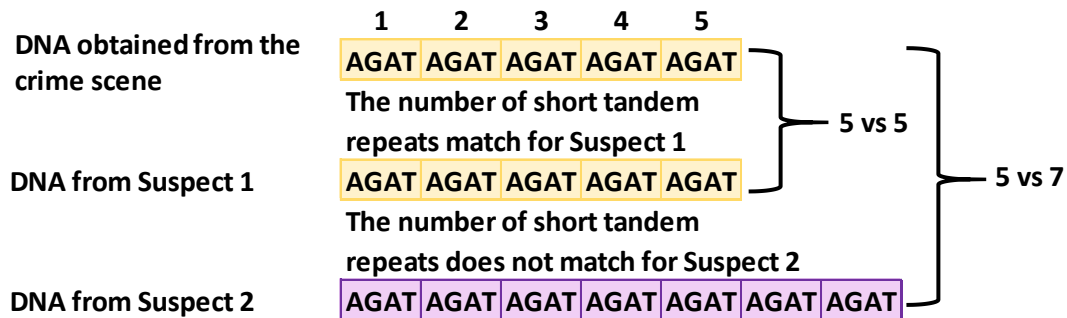


#### Q4. DNA Matching (30 Marks)

DNA is a molecule consisting genetic information which determines a person's characteristics. The genetic information is encoded by four chemical bases: A, T, C and G. These four chemical bases are important in performing DNA profiling used in a crime investigation. DNA profiling is based on the arrangement of these four chemical bases in a repetitive format termed Short Tandem Repeats (STRs). For example, considering the STR is AGAT, we have the following DNA fragments:



As shown above, the number of STRs (i.e. **AGAT**) in the DNA obtained from the crime scene is 5, which is similar to the number of STRs in the DNA obtained from suspect 1, indicating that suspect 1 probably appeared in the crime scene. By comparing the number of STRs in the DNA of suspect 2, there are two extra STRs which proposed that the DNA obtained from the crime scene does not belong to suspect 2.

Case scenario:

An investigation has been conducted in the laboratory towards **four suspects** via DNA profiling. Generally, a complete DNA sequence is provided and it is separated into fragments since the DNA sequence is normally very long. The matching process will be carried out between the fragments to determine the closest match. Later, the number of longest STRs in the matched DNA fragment sequences is extracted for any further investigation.

#### Write a program to

**Input, the following in sequence:**

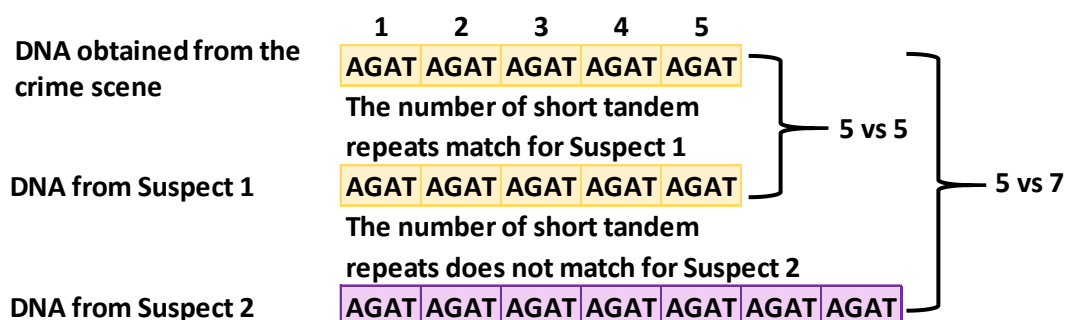
- **STR**
- The **DNA fragment** obtained from the crime scene
- The **DNA fragment** of Suspect 1
- The **DNA fragment** of Suspect 2
- The **DNA fragment** of Suspect 3
- The **DNA fragment** of Suspect 4

**Output, in sequence:**

- The suspect number (i.e., 1, 2, 3, or 4) whose DNA fragment matches with the DNA obtained from the crime scene. If none of the four suspect has a match, then output 0.
- The number of longest STRs in the matched **DNA fragment**. If there is no match, then output 0.

#### Q4. DNA 匹配 (30 分)

DNA 是一种由遗传信息组成的分子，它决定了一个人的特征。遗传信息由四种化学碱基编码：A、T、C 和 G。这四种化学碱基对于犯罪调查中使用的 DNA 分析非常重要。DNA 分析基于这四种化学碱基以重复格式排列，称为短串联重复序列 (STRs, Short Tandem Repeats)。例如，假设 STR 是 AGAT，且我们得到了以下的 DNA 片段：



如上图，从犯罪现场获得的 DNA 中连续 STR（即 AGAT）的数量为 5，与从嫌疑人 1 获得的 DNA 中的 STR 数量相等，这说明了嫌疑人 1 很可能曾出现在犯罪现场。另一方面，比较嫌疑人 2 的 DNA 中 STR 的数量，嫌疑人 2 有两个额外的 STR，这表明了从犯罪现场获得的 DNA 不属于嫌疑人 2。

考虑以下案例：

实验室通过 DNA 分析对四名嫌疑人进行了调查。一般来说，由于 DNA 序列通常很长，在调查前完整的 DNA 序列通常会先被分成片段。之后，再将获得的片段之间进行配对以确定最接近的匹配。之后，提取匹配的 DNA 片段序列中最长 STR 的数量以供进一步调查。

#### 试写一程式以

##### 依序输入

- STR
- 从犯罪现场获得的 DNA 片段
- 嫌疑人 1 的 DNA 片段
- 嫌疑人 2 的 DNA 片段
- 嫌疑人 3 的 DNA 片段
- 嫌疑人 4 的 DNA 片段

##### 输出

- 与犯罪现场的 DNA 片段匹配的犯罪嫌疑人的编号（即 1，2，3 或 4）。若四位嫌疑人的 DNA 都无法和犯罪现场的 DNA 匹配，则输出 0。
- 匹配 DNA 片段中最长 STR 的数量。若无法匹配，则输出 0。

### Example (例子)

Input (输入)	Output (输出)
AGAT GATCAGATAGATAGATAGATAGATATGCATAATTCCTGAC GATCATGCATGCAGATAGATATGCATGCATGCAGTCAGCA GATCAGATAGATAGATAGATAGATATGCATAATTCCTGAC AATCATAGATAGATAGATAGATACTTATCGACTAGGATCC TATCATAATATGCATGCATGCATGCATGCATGCATGCTCC	2 5
ACG ACGAGCACGGTCATAGATAGATACTTATCGACTAGGATCC ACGAGCACGGTCATAGATAGATACGGATCGACTAGGATCC ACGAGCACGGTCATAGATAGATACTTATCGAATAGGATCC ACGAGCACGGTCATAGATAGATACTTATCGACTACGATCC ACGAGCACGGTCATAGATAGATACTTATAGACTAGGATCC	0 0
CTA CTACTACTACTAATAACTGCTACTATAGATAGATAGATACT CTACTACTACTAACTGATAGAGATAGATAGATAGATACT CTAGAGATAGATAGATAGATACTAGACGATAACTGCTACT CTACTACTACTAATAACTGCTACTATAGATAGATAGATACT CTAGAGATAGATAGATAGATAGATAACGATAACTGCTACT	3 4
GATA CTAGAATAGCTAGACGATAACTGCTAGATAGATAGATAGA CTACTAGACGATAACTGATAGATAGATATAGATAGATACT CTAGAGATAGATAGATAGATAGATAACGATAACTGCTACT CTAGAATAGCTAGACGATAACTGCTAGATAGATAGATAGA CTAGAATAGATAGATAGATAGCTAGACGATACCTGCTACT	3 3
CTA ATAACTGCTACTATAGATCTATAGATACTCTACTACTACTA ATAACTGCTACTAACTGATAGAGATAGATAGATAGATACT ATAACTGCTACTAGATAGATACTAGACGATAACTGCTACT ATAACTGCTACTAGATAGATAGATAACGATAACTGCTACT ATAACTGCTACTATAGATCTATAGATACTCTACTACTACTA	4 4