Virus Mutations   P41641_en

Recall that the genetic code (see problem P36671) is a set of rules that translates nucleotide sequences into proteins. When working with DNA instead of messenger RNA, nucleotide sequences change the base `U` to the base `T`.

Often, as a resistance mechanism against drugs, viruses mutate their DNA while maintaining their functionality, i.e., synthesizing the same protein. For example, if a virus had the sequence

$$(1)\ \texttt{GCCAATGACTAAGGCCTAAAGA}$$

it would then synthesize the protein *Thr–Lys–Ala*. If the virus mutated to sequence (2)

$$(1)\ \texttt{GCCAATGAC}\textit{T}\texttt{AAGGCCTAAAGA}$$
$$(2)\ \texttt{GCCAATGAC}\textbf{C}\texttt{AACGCCTAAAGA}$$

(only the tenth base has changed), it would continue to synthesize *Thr–Lys–Ala*. In this case, we say that one (or more) equivalent mutations have occurred.

On the other hand, if the virus mutated to sequence (3)

$$(1)\ \texttt{GCCAATG}\textit{A}\texttt{CTAAG}\textit{G}\texttt{CCTAAAGA}$$
$$(3)\ \texttt{GCCAATG}\textbf{T}\texttt{CTAAG}\textbf{A}\texttt{CCTAAAGA}$$

(only the eighth and fourteenth bases), it would then synthesize *Ser–Lys–Thr*, which is noted by indicating `Thr-1-Ser,Ala-3-Thr` (the number indicates which codon the mutation occurred in).

Write a program that indicates whether pairs of DNA sequences represent equivalent mutations and, if not, indicates which mutations have occurred. The program should also state how many bases are different between the `ATG` codon and the *Stop* codon.

Input

The input begins with an integer *k*. Then, *k* pairs of DNA sequences of the same length follow. Each DNA sequence consists of a string of characters 'A', 'C', 'T', and 'G' on a single line without spaces.

All DNA sequences have an `ATG` codon before a *Stop* codon. The `ATG` codon and the *Stop* codon never mutate in any pair of sequences, and no mutation can introduce a new *Stop* codon. Pairs of sequences always have the same length.

Output

For each pair of DNA sequences in the input, indicate whether they are equivalent or not. If they are not equivalent, indicate all the changes produced by the mutations according to the standard notation described above. The number of different bases should be written at the end of the line, in parentheses.

Follow the format of the example.

Sample input 1

```
4

GCCAATGACTAAGGCCTAAAGA
GCCAATGACCAAGGCCTAAAGA

GCCAATGACTAAGGCCTAAAGA
GCCAATGTCTAAGACCTAAAGA

GCCAATGACTAAGGCCTAAAGA
GCCAATGTCTAAAACCTAAAGA

GCCAATGACTAAGGCCTAAAGA
GCCAATGACTAAGGCCTAAAGA
```

Sample output 1

```
Equivalents (1)
Thr-1-Ser,Ala-3-Thr (2)
Thr-1-Ser,Ala-3-Thr (3)
Equivalents (0)
```

Problem information

Author: Lluís Alemany Puig
Translator: Jordi Petit

Generation: 2026-01-25T11:06:10.911Z