

CHALLENGE 2

House Divided

Problem Description

At the *Princeton-Plainsboro Teaching Hospital*, the diagnostic team is tasked with solving a case of persistent, recurrent symptoms in multiple patients. After a deeper analysis, Dr. *Gregory House* suspects that a specific mutation in the DNA of these patients could be the common link.

Unfortunately, the lab equipment capable of identifying DNA sequences has been... “borrowed.” The team must now rebuild a computer program to analyze a DNA sequence for mutated patterns. Specifically, the task is to identify repeated DNA subsequences that are exactly one mutation (a single nucleotide difference) away from a given reference sequence. For instance:

ATCGATCGATCGAT
(original)

ATTGATCGATCGAT
(mutation)

Input Specification:

The first line is a string *S* of nucleotides representing the original DNA subsequence. The second line is a longer string *T* representing the entire DNA sequence with mutations present.

Output Specification:

A single integer that indicates the number of unique subsequences in *T* that:

- Are the same length as *S*.
- Differ by *S* by exactly one nucleotide.
- Appear more than once in *T*.

This is followed by the list of such sequences in **alphabetical order**, separated by a newline.

CHALLENGE 2

House Divided

Sample Input

Sample Output

AGA
AGTGAGGTAGT

1
AGT

The mutations of AGA that appear are AGT, TGA, and AGG. Of these, only AGT appears twice.

A G T G A G G T A G T

TGATG
TGCTGCTGTTGTTG

2
TGCTG
TTGTG

The mutations of TGATG that appear are TGCTG and TGTTG. Both of these appear twice (overlaps are permitted).

T G C T G C T G T T G T T G