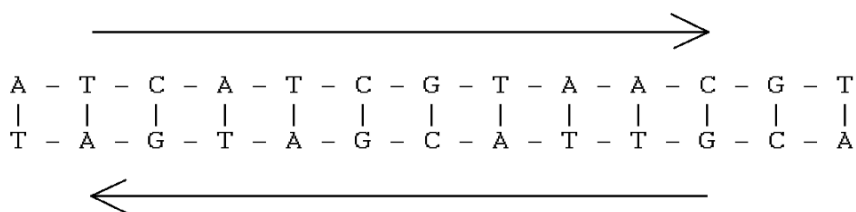


2019.1.2 - Identificare gene ADN

Maximum number of files: 5

Type of work: Individual work

În cadrul domeniului bioinformaticii, echipamentele de secvențiere de ADN preiau o probă biologică și produc un fișier text ce conține secvența de ADN detectată. Se cunoaște faptul că o secvență de ADN este reprezentată printr-un lung șir de caractere A, T, C și G, reprezentând nucleotidele Adenină, Timină, Guanină și Citozină. De asemenea, se cunoaște faptul că ADN-ul este format din două secvențe în oglindă, precum se vede în imaginea de mai jos, unde o nucleotidă A se leagă întotdeauna cu o nucleotidă T, iar o nucleotidă C cu una G, și invers. În plus, a doua secvență este citită întotdeauna de echipamentul de secvențiere în sens invers decât prima (de la sfârșit la început).



Cerință

Dându-se o secvență de ADN primită de la un echipament de secvențiere și o a doua secvență (mai scurtă) de interes, să se caute cea din urmă în cea dintâi. Nu se cunoaște care din cele două secvențe din perechea ADN-ului este dată de echipament, deci trebuie tratate ambele cazuri.

Date de intrare

Se vor citi de la tastatură (fluxul *stdin*) următoarele date:

- Pe prima linie se află un șir **S** de majuscule format exclusiv din caracterele A, T, C și G, reprezentând secvența dată de aparat;
- Pe a doua linie se află un alt șir **G**, similar cu primul dar mai scurt, reprezentând secvența de căutat.

Toate liniile conținând date de intrare sunt finalizate cu caracterul *newline* (tasta Enter).

Date de ieșire

Programul va afișa pe ecran (*stream*-ul standard de ieșire):

- pe prima linie: un număr întreg pozitiv **n** reprezentând numărul de poziții unde s-a găsit șirul căutat.
- pe următoarea linie: **n** valori întregi pozitive, separate prin spațiu, reprezentând pozițiile din secvență unde s-a găsit șirul căutat; acestea se afișează în ordine crescătoare.

ATENȚIE la respectarea cerinței problemei: afișarea rezultatelor trebuie făcută EXACT în modul în care a fost indicat! Cu alte cuvinte, pe stream-ul standard de ieșire nu se va afișa nimic în plus față de cerința problemei; ca urmare a evaluării automate, orice caracter suplimentar afișat, sau o afișare diferită de cea indicată, duc la un rezultat eronat și prin urmare la obținerea calificativului „Respins”.

Restricții și precizări

- 1024 \geq Lungime(**S**) \geq 2
- Lungime(**S**) \geq Lungime(**G**) \geq 1
- Numărătoarea indecșilor începe de la 0 (poziția primului caracter din secvență este 0).
- Dacă secvența căutată este periodică, pozițiile acesteia în secvența de intrare se pot suprapune și deci trebuie raportate toate pozițiile de potrivire (vezi exemplul 2).

5. **Atenție:** În funcție de limbajul de programare ales, fișierul ce conține codul trebuie să aibă una din extensiile .c, .cpp, .java, sau .m. Editorul web nu va adăuga automat aceste extensii și lipsa lor duce la imposibilitatea de compilare a programului!
6. **Atenție:** Pentru cei ce utilizează limbajul Matlab, este recomandat ca fișierul sursă să fie numit sub forma: <nume>.m.

Exemple

Intrare	Ieșire
ACTCGTTTCGT CGT	2 3 8
Explicație Secvența CGT apare de două ori la intrare, o dată pe poziția 3, apoi pe poziția 8.	
CGCTCTCTCCCG TCTC	2 3 5
Explicație Secvența căutată este periodică, ea apare pe pozițiile 3 și 5, și se raportează ambele, chiar dacă se suprapun.	
ACTCGTTACGT CGT	3 3 7 8
Explicație În situația de mai sus, secvența CGT este căutată și găsită în două poziții: 3 și 8. Dar conform cerinței, pentru că nu știm care din cele două secvențe a fost dată la intrare (prima sau cea în oglindă), vom căuta și secvența ACG, care reprezintă șirul căutat, oglindit și inversat. Acesta este găsit pe poziția 7.	