

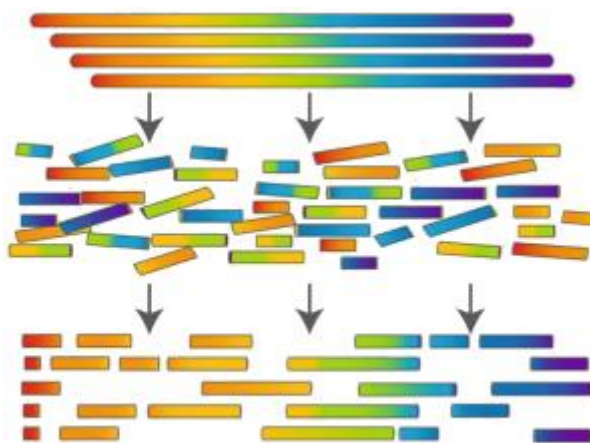
Ročník: 11

Kolo: 4.

Úloha: 6

Úloha o skladaní reťazcov

Biológovia v Korytnáčkove používajú špeciálny stroj, pomocou ktorého vedia určiť poradie báz (písmen) vo vlákne (reťazci) DNA. Tento proces sa nazýva *sekvenovanie DNA*. Má to však jeden háčik. DNA je veľmi dlhá a stroj vie určiť poradie len niekoľkých báz. Preto musí byť reťazec DNA pred sekvenovaním rozdelený na menšie reťazce, ktoré budú neskôr spojené do výsledného reťazca. Tento proces sa



ATGTTCCGATTAGGAAACCTTTCCTTAACTGTTTCATTTCAGTAAAAGGAGGAAATATAA

Zdroj: <http://rosalind.info/media/problems/long/fragment-assembly.thumb.png>

nazýva *DNA usporadúvanie sekvencií (sequence assembly)*. Delenie reťazca sa robí tak, že každý reťazec (okrem prvého a posledného) sa dá pripojiť len k jednému reťazcu na začiatok a len k jednému reťazcu na koniec.

Pre spájanie krátkych reťazcov do pôvodného DNA reťazca sú dôležité **prvé 4** a **posledné 4 písmená** krátkych reťazcov. Na základe nich je možné spojiť 2 reťazce dokopy. Teda, ak sú posledné 4 písmená jedného reťazca rovnaké ako prvé 4 písmená druhého reťazca, môžeme tieto reťazce spojiť.

Napríklad:

krátky reťazec 1: **AGCCGTATGCA**

krátky reťazec 2: **TGCAACGTTCA**

výsledný DNA reťazec: **AGCCGTATGCAACGTTCA**

Vedel by si korytnačím biológom pomôcť?



Imagine Logo

Vytvor procedúru **assembly** s vhodnými parametrami, ktorá vytvorí výsledný DNA reťazec pospájaním krátkych reťazcov.

Svoje riešenie ulož ako Imagine projekt **assembly.imp**.

Python

Vytvor funkciu **assembly** s vhodnými parametrami, ktorá vytvorí výsledný DNA reťazec pospájaním krátkych reťazcov.

Svoje riešenie ulož do súboru **assembly.py**.

Poznámka:

Na testovanie procedúry môžeš využiť vopred pripravené rozsekané reťazce, ktoré sú v súbore `retazce.txt`. Súbor `celeRetazce.txt` obsahuje výsledky jednotlivých pospájaní.