

## Správa projektu predmetu BIF

### 1. Vyhľadanie génov

Ako je zrejmé z obrázkov zadání, táto štruktúra obsahuje isté špecifické konzervované nukleotidy. Práve tieto boli použité na lokalizáciu génov. Algoritmus vyhľadáva podreťazec reprezentujúci nukleotidy GTTC, v každom mieste kde ich nájde začne vykonávať dôslednejšiu analýzu okolia. Od tejto časti smerom ku koncu existuje na konštantných pozíciách ešte niekoľko sekvencií, ktorých prítomnosť je overovaná. Tu je v programe možnosť meniť, počet požadovaných zhôd, ale zistil som že algoritmus funguje najlepšie ak očakávame že sa tam dané nukleotidy nachádzajú. Ďalej sa overovala prítomnosť sekvencií pred pozíciou „GTTC“. Toto bolo zložitejšie nakoľko sa tu nachádzajú variabilne dlhé sekvenície, ktoré tieto konzervované časti oddelujú. Tu sa pracuje postupom takým, že sa v cykle uvažuje so všetkými možnosťami. Ak sa nájde jedna vyhovujúca predpokladáme, že tu sekvencia existuje.

### 2. Výsledky

Algoritmus našiel, okrem existujúcich sekvencií 3 ďalšie. Tieto by bolo pravdepodobne možné vylúčiť ďalšou filtráciou.

Algoritmus však nenašiel 2 výskyty. Toto je spôsobené tým, že sa jedná o neštandardné sekvencie z hľadiska konzervovaných sekvencií. V prvom prípade sa jedná o sekvenciu trna s antikodonom Pseudo???, u ktorej nebolo možné nájsť v špecifické konzervované sekvencie. V druhom prípade sa jedná o tRNA pre aminokyselinu Selenocystein (SeC), dôvody nenájdenia sú podobné ako v prvom prípade.

### 3. Zložitosť

Algoritmus prehľadáva vstup a snaží sa nájsť podreťazec GTTC, čo je možné vykonať so zložitosťou  $n^4$ , na každom tomto indexe sa spustí algoritmus so zložitosťou  $2l^4 \cdot 3$ , čo vo výsledku dáva  $n^4 \cdot l^4 \cdot 3$ , čo je lineárne.