

Thema 1, script 8

Verplichte onderdelen:

- Kopieer de inhoud van je taaislijmziekte-mRNA GenBank-bestand naar de online GenBank Feature Extractor op http://www.bioinformatics.org/sms2/genbank_feat.html (*). Upload deze informatie met de Submit-knop en kopieer de inhoud van de pagina die dan verschijnt naar een tekst-bestand. (Hint: kopieer alleen de inhoud in monospaced font; de titel "GenBank Feature Extractor results" hoeft je niet mee te kopiëren.)
- Bestudeer de inhoud van de uitvoer nauwkeurig en vergelijk deze met de invoer die je had ingegeven. Analyseer hoe de uitvoer is verkregen uit de invoer. (Hint: de secties DEFINITION, FEATURES, en ORIGIN zijn met name relevant.)
- Schrijf een programma dat een door de gebruiker gespecificeerd GenBank-bestand inleest en dat een tekstbestand met alle features uitvoert op exact dezelfde wijze zoals de GenBank Feature Extractor dat doet. (Hint: maak zelf eerst een stappenplan hoe je dit aanpakt, met eigen "deelopdrachten".) Als de invoer bestaat uit een bestand genaamd *mrna.gb*, dan dient het programma de uitvoer weg te schrijven in een bestand genaamd *mrna_features.txt*.
- Zet met de online GenBank Feature Extractor ook de inhoud van je taaislijmziekte-DNA GenBank-bestand om in een tekstbestand met features. Controleer dat je eigen programma dezelfde uitvoer geeft; pas je code zonedig aan.

Bonus-onderdelen:

- Ga na hoe de uitvoer verandert als je in de online GenBank Feature Extractor de optie met het uitvoerformaat verandert van "separated" in "uppercased". (Hint: kijk niet alleen naar hoofdlettergebruik, maar ook naar de grootte van het geproduceerde bestand.) Voorzie je eigen programma ook van een optie waarmee de gebruiker kan kiezen voor het "separated" of "uppercase" formaat, en zorg dat de uitvoer daarmee overeenkomt.
- Zoek de eiwit-referentiesequentie op behorende bij het gen dat taaislijmziekte veroorzaakt en download deze in GenPept-formaat. Controleer dat het om versienummer NP_000483.3 gaat. Probeer ook deze om te zetten met de online GenBank Feature Extractor, en ga na dat dit niet helemaal goed gaat. Zorg ervoor dat je eigen programma deze omzetting wél kan maken.

Lever je programma in via BlackBoard in de vorm van een python-bestand genaamd **script8.py**.

Let hierbij op dat je code:

- ✓ op tijd wordt ingeleverd;
- ✓ zonder foutmeldingen uitvoerbaar is;
- ✓ de (verplichte) onderdelen van de opdracht correct en volledig verricht;
- ✓ liefst zo efficiënt mogelijk geïmplementeerd is;
- ✓ en, een netjes leesbare programmeerstijl gebruikt.

* Soortgelijke online GenBank Feature Extractors zijn te vinden op http://www.geneinfinity.org/sms/sms_gbfeatures.html, http://www.cellbiol.com/sequence_manipulation_suite/genbank_feat.php, http://groups.molbiosci.northwestern.edu/matouschek/links/sms2/genbank_feat.html, en elders.