

Thema 1, script 7

Verplichte onderdelen:

- Zoek zowel de DNA- als de mRNA-referentiesequenties op behorende bij het gen dat taaislijmziekte veroorzaakt en download deze in GENBANK-formaat. Controleer dat het om versienummers NG_016465.4 en NM_000492.3 gaat.
- Schrijf een programma dat een door de gebruiker gespecificeerd GenBank-bestand inleest en de namen van de daarin voorkomende secties op het scherm toont. (Hint: de namen van de secties staan vooraan de regel; eventuele sub-secties op regels die beginnen met witruimte mag je negeren. Bv. *SOURCE* is wel een sectie, *ORGANISM* niet.)
- Pas je programma aan zodat het in plaats daarvan de inhoud van de genoemde secties opslaat in een variabele. (Hint: schrijf een functie die een dictionary retourneert.) Let op dat je programma correct omgaat met inhoud die over meerdere regels verspreid is.
- Gebruik de informatie uit de secties *VERSION* en *DEFINITION* om een geschikte header regel voor een bijbehorend FASTA-bestand te genereren. (Hint: kijk in een aantal FASTA-bestanden om te zien hoe deze headers typisch zijn opgebouwd.)
- Gebruik de sectie *ORIGIN* (tot aan de *"/*) om de nucleotide-sequentie te bepalen in de vorm van één lange string. (Hint: haal de spaties en cijfers en eventuele witruimte eruit.)
- Schrijf met behulp van de voorgaande gegevens een geldig FASTA-bestand weg naar de schijf. Pas de extensie van de bestandsnaam hiertoe aan. Bv., als het GenBank-bestand genaamd *mrna.gbk* wordt ingelezen, dan dient een FASTA-bestand genaamd *mrna.fa* te worden weggeschreven. Maak de regels met de sequentie niet langer dan 70 tekens.

Bonus-onderdelen:

- Pas je programma aan zodat het meerdere bestanden kan verwerken.
- Tel het aantal splice-sites in een GenBank-bestand en rapporteer dit op het scherm. Je kunt dit op meerdere manieren doen, steeds gebruik makend van de informatie in de sectie *FEATURES*. Bijvoorbeeld:
 - tel het aantal keer dat het sub-kopje *exon* voorkomt;
 - lees het sub-kopje *mRNA* en tel hoeveel komma's er in de *join(...)* staan.Komen de antwoorden van deze twee methoden overeen?
- Gebruik de DNA-sequentie (uit de sectie *ORIGIN*) en de gegevens over de splice-sites (uit de *join(...)* in het sub-kopje *mRNA* van de sectie *FEATURES*) om een gespliced mRNA te maken. (Hint: plak alle exons achter elkaar om één mRNA sequentie te verkrijgen.) Toon het resultaat op het scherm.

Lever je programma in via BlackBoard in de vorm van een python-bestand genaamd **script7.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.