

Thema 1, script 6

Verplichte onderdelen:

- Schrijf een programma dat op basis van informatie over een mutatie en een FASTA-bestand met een DNA- of mRNA-sequentie, deze mutatie aanbrengt in de sequentie en naar het scherm schrijft. Bv., de invoer "dna.fasta 31 33 TAG" zou de drie nucleotiden op posities 31 t/m 33 uit het bestand genaamd *dna.fasta* moeten vervangen door de nucleotide-reeks TAG.
- Zorg dat je programma de volgende mutaties correct kan uitvoeren. Denk zelf na hoe een gebruiker deleties kan invoeren.

Casus	Beginpositie	Eindpositie	Mutatie
CFTR	4474	4474	T
CFTR	72	74	GTG
CFTR	554	585	-
CFTR	507	509	-

- Pas je programma aan zodat de gemuteerde sequentie naar een nieuw FASTA-bestand wordt weggeschreven. Pas hierbij de header-regel aan zodat het duidelijk is dat dit bestand een mutatie bevat. De naam van het uitvoer-bestand wordt door de gebruiker aangegeven.
- Breid je code uit zodat deze ook een multi-FASTA-bestand kan inlezen. (Hint: lees https://en.wikipedia.org/wiki/FASTA_format voor informatie omtrent dit formaat.) De gebruiker dient aan te geven in welke sequentie de mutatie dient te worden uitgevoerd door de identifier die in de header voorkomt te vermelden. Bv., de invoer "multi-in.fasta multi-out.fasta NG_016465.4 31 33 TAG" muteert alleen de sequentie met identifier NG_016465.4 uit het bestand *multi-in.fasta*, en schrijft alle sequenties vervolgens weg naar het bestand *multi-out.fasta*.

Bonus-onderdelen:

- Herschrijf je programma zodanig dat het niet een gehele sequentie in het geheugen hoeft te onthouden, maar de correct gemuteerde sequentie(s) reeds regel voor regel wegschrijft *terwijl* het de oorspronkelijke sequentie(s) aan het inlezen is.
- Veralgemeeniseer je code zodat je programma ook mutaties in FASTA-bestanden met aminozuur-sequenties aankan (de gebruiker geeft de begin-/eindposities en de mutatie reeds in termen van aminozuren, dus je hoeft geen codon-translatie uit te voeren).
- Pas de code zodanig aan dat je programma zelf een geschikte bestandsnaam voor het uitvoerbestand afleidt op basis van de naam van het invoerbestand indien de gebruiker er geen aangeeft.

Lever je programma in via BlackBoard in de vorm van een python-bestand genaamd **script6.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.