

OWE1a Thematotoets Kans1

Naam student:

Klas:

Ga naar onderwijs online. Ga naar de course en kijk onder de praktijkopdrachten. Hier vind je alle documenten die je nodig hebt voor het theoriegedeelte en het praktijkgedeelte.

Theorie

Voor dit gedeelte heb je 1 klokuur de tijd (tussen 13.00 en 14.00 uur). Gebruik voor je antwoorden het digitale antwoordmodel:

BIN-OWE1a_Thematotoets1718_Kans1_Antwoordvel.docx

Bij iedere vraag staat het maximaal aantal punten dat je kan behalen vernoemd.

Bio-informatica (6 vragen, 20 pt)

Vraag 1 (3p)

Zoek van het organisme *Saccharomyces cerevisiae* het gen SIR1 op.

- A. Op welk chromosoom ligt dit gen? (2p)
- B. Welke genen omringen dit gen op het chromosoom? (1p)

Vraag 2 (3p)

Zoek van het menselijke gen Thyroid hormone receptor alpha (THRA) de lengte van de genoom sequentie en de lengte van de mRNA sequentie op (neem hiervoor isoform 3). Verklaar het verschil in lengte.

Vraag 3 (2p)

Wat is het gewicht van het eiwit Insulin van *Homo sapiens*?

Vraag 4 (4p)

Download van het eiwit Insulin (INS) van *Homo sapiens* en *Sus scrofa* en *Bos Taurus* de proteinesequenties in FASTA formaat en vergelijk deze in MAFFT van genome.jp.

- A. Geef een screenshot van de alignment en beschrijf wat je ziet. (3p)
- B. Geef de bijbehorende fylogenetische boom (UPGMA with branch length)(1p)

Vraag 5 (2p)

Leg uit wat het verschil is tussen een pairwise alignment en een multiple sequence alignment.

Vraag 6 (6p)

Ga naar PDB (RCSB.org) en open 4REA.

- A. Beschrijf de structuur; welke secundaire structuren heeft dit eiwit? (2p)
- B. Welke tertiaire structuur heeft dit eiwit: is het een monomeer of homo/hetero-x-meer? (2p)
- C. Geef een screenshot van de structuur. (2p)

Linux (8 vragen, 10 pt)

Voor het Linux gedeelte van deze toets geldt dat je op moet schrijven welk commando(s) je uitvoert om de gevraagde opdracht uit te voeren.

Vraag 1 (1p)

Start de terminal op (geen commando) en controleer in welke directory je je bevind.

Vraag 2 (1p)

Ga naar je home directory en maak daar een nieuwe map genaamd **Thematoets** aan.

Vraag 3 (1p)

Ga naar de aangemaakte map toe en maak hierin een nieuw bestand genaamd **vraag3.txt** aan.

Als inhoud kan je de volgende tekst kopiëren en plakken (geen commando):

*Lorem ipsum dolor sit amet, consectetur adipiscing elit.
Sed id nunc in risus pellentesque cursus vel et est.
Nunc urna enim, egestas nec nunc non, euismod dictum neque.
Cras lacinia lorem enim, ut pulvinar elit dapibus ut.
Phasellus a iaculis mi.
Curabitur tincidunt massa neque, id egestas magna rhoncus eu.
Sed a ante urna.
Vestibulum a turpis a justo aliquet ornare et eget nibh.
Sed tempor tortor rhoncus nulla auctor malesuada eu non justo.
Nunc aliquet finibus blandit.
In non nisl venenatis, efficitur ante viverra, egestas ante.*

Zorg ervoor dat je dit ook opgeslagen hebt (geen commando).

Vraag 4 (2p)

Toon de eerste 5 regels van dit document.

Vraag 5 (1p)

Wat zijn de rechten van dit document? Verander de rechten zodat iedereen dit document kan lezen, maar niemand mag schrijven of uitvoeren.

Vraag 6 (1p)

Maak een nieuw bestand genaamd vraag7.py. De inhoud is niet belangrijk. Geef vervolgens een overzicht van alle bestanden in de map die eindigen op .txt

Vraag 7 (1p)

Maak een kopie van vraag7.py, noem deze vraag7_kopie.py en verander de rechten van het bestand zodat alleen de eigenaar van bestand deze kan lezen en schrijven (dus geen uitvoerrechten).

Vraag 8 (2p)

Gebruik grep om alle regels uit het bestand vraag3.txt te tonen waar het woord "nunc" in voorkomt. De eerste "n" van het woord nunc kan een grote of kleine letter zijn.

Praktijk

Voor dit gedeelte heb je tot 17.00 uur de tijd.

Casus

Je bent gevraagd door het HAN BioCentre om hen te helpen met het identificeren van genen in het genoom van gist, *Saccharomyces cerevisiae*. Zoals wellicht bekend is *S. cerevisiae* een goed te onderzoeken model organisme doordat het een volledig bekend genoom heeft van iets meer dan 6000 genen.

Van je stagebegeleider heb je het volgende CSV bestand gekregen: **yeast_genes.csv**
Hierin staat per gen de accessiecode, de validatiestatus van het gen (verified, dubious, uncharacterized) en de lokalisatie. Het bestand ziet er als volgt uit:

```
gene,validation,localization
YAL008W,Verified,ion
YBR255W,Verified,biological_process
YGR164W,Dubious,biological_process
YGR131W,Verified,protein,ion
YNL003C,Verified,transport,ion
...
```

Taak

Je stagebegeleider wil dus dat je een programma schrijft dat nagaat welke van deze genen niet gevalideerd zijn.

- Hij wil graag weten hoeveel dit er zijn zijn, ofwel, hoeveel genen een andere status dan “Verified” hebben.
- Daarnaast wil hij ook weten welke accessiecodes horen bij deze niet gevalideerde genen.
- Van de *niet gevalideerde genen* wil hij ook graag weten hoeveel en welke er niet gevalideerde genenbetrokken zijn bij ionen, die dus ergens in de “localization” “ion” hebben staan.

Als output wordt het volgende ongeveer verwacht, maar voel je vrij om hier je eigen draai aan te geven:

```
*****
Number of unverified genes: 1051
Number involved in ion processes: 659
*****
Acc.code:  YAL008W
Ion?      False
*****
Acc.code:  YBR255W
Ion?      True
*****
. . .
```

N.B. De hier genoemde aantallen en accessiecodes zijn niet correct, deze zijn alleen maar om te illustreren hoe de output er uit kan zien.

Opdracht

Het op te leveren programma dient aan de volgende functionele eisen te voldoen:

1. Leest het bestand
2. Het script doorzoekt het bestand en berekend de gevraagde eigenschappen per gen en voor het totaal
3. Het programma geeft output waarbij de gebruiker een overzicht krijgt van de eigenschappen van de genen en het totaal aan genen

Het op te leveren programma dient aan de volgende technische (niet functionele) eisen te voldoen:

1. Er is een flowchart
 - Ontwerp een flowchart voor de functie `not_validated()` en schrijf de bijhorende code.
 - De flowchart mag digitaal of op papier.
 - De in te leveren flowchart beschrijft de functie nauwgezet.
2. Het script is opgedeelt in functies. De volgende functies zijn in ieder geval aanwezig.
 - De functie `read_file()` leest het bestand in en maakt een 2d lijst met hierin alle genen en hun eigenschappen. De functie retourneert de 2d lijst.
 - De functie `not_validated()` accepteert de 2d lijst en bekijkt welke van de genen er niet gevalideerd zijn. De functie retourneert een lijst met accessiecodes van deze genen.
 - De functie `ion_involved()` accepteert een variabele bepaald welke van de niet gevalideerde genen zijn bij ion processen. Wellicht heeft de functie hier de 2d lijst voor nodig, of gewoon een sublijst van de 2d lijst. Je mag dus zelf bepalen wat je aan deze functie geeft. De functie retourneert in ieder geval een lijst met accessiecodes voor alle genen die betrokken zijn bij ion processen.
 - De functie `main()` roept al deze functies aan en verwerkt eventuele output naar het scherm.
 - N.B. Voel je vrij om je eigen draai te geven aan deze functies als je voelt dat het efficiënter of handiger kan. De enige harde eis is dat je met 2d lijsten werkt en met functies die parameters accepteren en output retourneren.
3. De snelheid van het programma doet er niet toe.
4. Het programma is geschreven in Python.

Beoordeling

De beoordeling van het totale tentamen is als volgt.

Criterium			Max punten	Beoordeling
1	Linux		10	
2	Bio-informatica		20	
3	Git			
	a.	Alle bestanden op GitHub	5	
	b.	Commits	5	
4	Python Code			
	a.	Juiste datatypes en variabelen	2	
	b.	Juiste controle structuren	2	
	c.	Onderverdeling in functies	4	
	d.	Functies hebben parameters	4	
	e.	Functies retourneren waarden	4	
	f.	Documentatie van code	4	
5	Juiste werking *			
	a.	Werking <code>read_file()</code>	10	
	b.	Werking <code>not_validated()</code>	10	
	c.	Werking <code>ion_involved()</code>	10	
6	Flowchart			
	a.	Flowchart volgens standaard	5	
	b.	Flowchart matcht met code	5	
Totaal			100	

* Indien andere namen en/of verder opgedeelde functionaliteit, worden er 30 pt verdeeld over het inlezen van het bestand naar 2d lijst (10pt), het bepalen van de validatiestatus (10pt) en het bepalen van de ion betrokkenheid (10pt).