

ITBC BI1 Python 1 en Basiskennis voor Bioinformatici

Weektaak 2

OWE1 Python 1 en Basiskennis voor Bioinformatici

Taak 2 – Analyse van de sequenties

Doelstellingen

Na afloop van deze taak kun je:

- In python eenvoudige scripts schrijven om sequentiedata te analyseren
- Het script testen
- De scripts remote opslaan in GIT-classroom
- In Excel een overzicht maken van de data

Casus

Linux op je computer hebben en een heleboel data is een begin. Nu er nog wat mee doen. Je gaat deze week aan de slag met het analyseren van de sequenties met behulp van Python.

Als je resultaten gaat analyseren, is het belangrijk dat de gevonden resultaten ook kloppen. Het testen van het script is erg belangrijk, want je gaat conclusies verbinden aan de resultaten die je met behulp van het script hebt verkregen. Als je onverwachte resultaten krijgt, komt dat dan door een bug in het script of doordat het echt zo is? Doet het script wel wat het moet doen? Het testen van een script gebeurt vaak stapsgewijs. Elk onderdeel van het script kan je apart testen. Als je zover bent dat je wil testen of de resultaten die het script produceert ook wel echt kloppen, kan je gebruik maken van test-data. Dit is een kleine hoeveelheid data, waarvan je de analyseuitkomsten al kent. Wat deze test-data is, hangt af van wat je wilt gaan testen.

Tijdens het schrijven en testen van het programma zal je regelmatig dingen die je hebt geschreven moeten veranderen. De gamers onder ons weten dat het soms verrekke handig is als je een save-game hebt om op een eerder punt opnieuw te beginnen, als je iets hebt aangevallen wat je eigenlijk niet had moeten doen. Hetzelfde geldt voor programmeren. GIT kan je gebruiken voor versiebeheer van je scripts.

Opdracht 1: Samenwerkingscontract

Deze week gaan jullie starten met een projectgroep. Om goed samen te kunnen werken is het belangrijk om afspraken met elkaar te maken. Als de samenwerking lekker gaat, lijkt dit niet zo nodig, maar wat als het wat moeizamer gaat? Hieronder volgen een aantal punten die je binnen je projectgroep duidelijk moet afspreken. Misschien hebben jullie nog wel meer punten die je wilt bespreken. Zet alle afspraken op papier en onderteken met alle leden van de projectgroep dit samenwerkingscontract.

1. Tijd: Vergaderafspraken, werkdagen voor het project, hoeveel tijd inversteert iedereen? Wat te doen bij afwezigheid, te laat komen?
2. Informatiebeheer: Verslaglegging, notuleren.
3. Organisatie: Manier van vergaderen en informatieuitwisseling; voorzitten, notuleren, e-mailadres en telefoonnummers.
4. Samenwerking: Taakverdeling,
5. Geld: Kopieerkosten, reiskosten etc.
6. Kwaliteit: Voor welke kwaliteit van het product willen jullie gaan al projectgroep (Bijv. een 6 of een 10).

Mocht het niet lekker gaan met samenwerken in de projectgroep, dan kan de tutor hierin mediëren. Lever daarom het samenwerkingscontract in bij de tutor. Als er iets speelt, wacht dan niet te lang met het informeren van de tutor.

Opdracht 2: Python scripts voor analyse van sequentiedata

We willen graag het aantal nucleotiden van de RNA en genoom FASTA sequenties weten, om de verhouding in grootte tussen deze sequenties in kaart te brengen. Er is namelijk een verschil tussen genen van prokaryoten en eukaryoten. Daarnaast willen we graag weten wat het verschil is in GC% tussen het complete genoom van het organisme en de mRNA sequentie van het gekozen eiwit om te verifiëren of het verhaal klopt, dat in genen het GC% hoger is dan in niet coderende sequenties.

- Schrijf een python script dat het aantal nucleotiden van een sequentie kan tellen.
- Schrijf een python script dat het gc% van een sequentie kan berekenen.
- Test je script
- Sla je script remote op in GIT-classroom group assignment ([link](#)).
- Zoek in de genoom database (NCBI) het gc% van het organisme op

- Vergelijk het gc% van het organisme met het gc% van de mRNA sequentie.

Resultaat

Aan het eind van deze taak heb je

- een samenwerkingscontract, die je inlevert bij de tutor.
- een python script dat het aantal nucleotiden van een sequentie kan tellen.
- een overzicht van de verhouding in het aantal nucleotiden tussen de 5 uitgekozen genoom en mRNA sequenties.
- een python script dat het gc% van een sequentie kan berekenen.
- een vergelijking van het gc% van de 5 uitgekozen mRNA sequenties ten opzichte van het gc% van het genoom van het organisme.
- alle data en vergelijkingen overzichtelijk opgeslagen in Excel.
- de python scripts opgeslagen in GIT-classroom.

Presentatie

Presenteer je resultaten aan de tutorgroep. De tutor zal per projectgroep een student aanwijzen om het woord te doen. Van iedere toehoorder/student wordt verwacht dat hij kritische vragen stelt ten aanzien van de inhoud en/of de vorm van de presentatie.

Vergeet niet de bronvermelding in je presentatie (zie APA richtlijnen onder Algemene informatie op OnderwijsOnline).

Middelen

- NCBI genome: <https://www.ncbi.nlm.nih.gov/genome/>

Nabespreking

Besproken wordt onder andere:

- Op welke manieren kan je het script testen?
- Hoe kan je de gevonden resultaten verklaren?
- Wat verliep moeilijk dan wel gemakkelijk?
- Hoe verliep de samenwerking / communicatie?
 - Hoe hebben jullie de opdracht aangepakt?
 - Wie nam het initiatief?
 - Hoe ging de communicatie?
 - Hoe was de omgang met afspraken?
 - Hoe werd de deadline bewaakt?
 - Wat zouden jullie de volgende keer anders doen?
 - Als je 13 punten te verdelen had in gehele getallen (integers), wie zou hoeveel punten krijgen?

Copyright @ HAN, I. Paffen