

ITBC BI1 Python 1 en Basiskennis voor Bioinformatici

Weektaak 6

OWE1 Python 1 en Basiskennis voor Bioinformatici

Taak 6 – Vergelijken van de eiwitsequenties - deel 2

Doelstellingen

Na afloop van deze taak kun je:

- Met behulp van bio-informatica tools een multiple sequentie-alignment uitvoeren.
- Met behulp van bio-informatica tools een fylogenetische boom berekenen en vergelijken met de soorten boom.
- In python eenvoudige scripts schrijven om sequenties te analyseren, gebruik makend van functies.
- Voorbereid de workshop "samenvatting schrijven" binnenkomen.

Casus

De organismen die je aan het vergelijken bent, hebben een gemeenschappelijke voorouder, ergens heel lang geleden. In deze gemeenschappelijke voorouder was de sequentie die je aan het bekijken bent al aanwezig. Tijdens de evolutie vinden er mutaties plaats in de genoomsequentie. Door deze mutaties veranderen langzaam de sequenties van genen in het genoom. Sommige mutaties resulteren in een veranderde functie van het eiwit waarvoor dat bepaalde gen codeert. Als er voldoende mutaties zijn opgetreden dat twee organismen geen vruchtbaar nageslacht meer kunnen produceren, spreek je van een nieuwe soort. In deze nieuwe soorten vinden dan weer nieuwe mutaties plaats waardoor het verhaal zich kan herhalen. Uiteindelijk zijn vanuit de gemeenschappelijke voorouder de organismen ontstaan die je nu aan het vergelijken bent.

In de vorige weektaak heb je de 5 sequenties paarsgewijs met elkaar vergeleken. Een hele klus. Maar het zou fijn zijn om in een overzicht alle 5 de sequenties onder elkaar te hebben. Dit noemen we een multiple sequence alignment. Dit is mogelijk omdat de sequenties die je aan het bekijken bent, afkomstig zijn van een gemeenschappelijke voorouder. De functie van de 5 sequenties is vergelijkbaar, maar door het verloop van de evolutie zijn de sequenties niet meer precies hetzelfde. In de multiple sequence alignment kan je dan zien welke aminozuren belangrijk zijn voor het uitvoeren van de functie van dit eiwit, omdat deze aminozuren hetzelfde zijn gebleven tijdens de evolutie (i.e. geconserveerd). Als een belangrijk aminozuur namelijk

gemuteerd wordt, dan zal de nakomeling met deze mutatie niet levensvatbaar zijn. Die mutatie zie je daarom niet terug in de populatie.

Opdrachten

Opdracht 1: Alignen van eiwitsequenties

Align de eiwitsequenties van de 5 organismen met elkaar in een multiple alignment met de tool 'MAFFT' bij Genome.JP: <http://www.genome.jp/tools-bin/mafft>.

Een multiple alignment tool heeft de sequenties in een bestand onder elkaar in FASTA format als input nodig. De tool herkent waar de volgende sequentie begint aan de hand van de header die begint met >.

Mafft parameters: output order = aligned. Verder kan je de standaard instellingen gebruiken.

- Wat is het verschil in output tussen de pairwise en multiple alignment tools? Zijn de alignments vergelijkbaar? Met andere woorden, staan dezelfde aminozuren onder elkaar als je de pairwise alignments met de multiple alignment vergelijkt?
- Kan je het verschil in alignment tussen de pairwise alignments en de multiple alignment verklaren?

Opdracht 2: Phylogenetische boom

We kunnen de verschillen tussen de sequenties ook visualiseren in een fylogenetische boom. De uiteindes van de takken van de boom geven de verschillende sequenties weer. Een splitsing van twee takken geeft aan dat die twee sequenties een gemeenschappelijke voorouder hadden, ergens lang geleden in een organisme dat er nu niet meer is. De wortel (root) van de boom geeft de gemeenschappelijke voorouder van al deze organismen aan.

- Maak een fylogenetische boom van je multiple alignment.

Als je de alignment met MAFFT bij genome.jp hebt gemaakt, zie je bovenaan de resultatenpagina de mogelijkheid om een 'tree' te maken. Kies in dit menu voor "Rooted phylogenetic tree with branch length (UPGMA)". De lengte van de horizontale takken is nu een maat voor de afstand (het aantal mutaties) tussen de sequenties. De root is aangegeven met een klein streepje.

- Komt deze fylogenetische boom overeen met de evolutionaire hiërarchie van de soorten?

Via de taxonomy database kan je een soorten boom maken: <https://www.ncbi.nlm.nih.gov/Taxonomy/CommonTree/wwwcmt.cgi>.

- Zal de fylogenetische boom gebaseerd op een eiwit altijd overeenkomen met de evolutionaire boom van de soorten?

Opdracht 3: Verbeteren Python scripts

Je hebt de afgelopen weken heel wat Python scripts geschreven voor het analyseren van sequenties. We begonnen eenvoudig, en langzamerhand werd het steeds complexer. Waarschijnlijk kan je met de kennis die je nu hebt de scripts van de afgelopen weken verbeteren. Door gebruik te maken van functies wordt een script overzichtelijker en soms ook sneller.

Voordat we de afgelopen weken begonnen met het schrijven van een script, is eerst mbv pseudocode bedacht hoe het script zou moeten werken. Als je gebruik maakt van functies kan je in plaats van pseudocode ook gebruik maken van een flowchart. Daarin wordt de flow van het script door de functies gevisualiseerd.

- Verbeter de scripts van de weektaken door gebruik te maken van functies.
- Test de werking van het script.
- Laat per script in een flowchart de werking van het script zien.

Opdracht 4: Voorbereiding op het verslag

Wanneer onderzoekers een onderzoek afronden, presenteren zij dit in een wetenschappelijk artikel.

Aan het eind van je stage en je afstudeerproject schrijf je individueel een verslag. Dit verslag heeft de opbouw van een wetenschappelijk artikel: Samenvatting, Inleiding, Methode, Resultaten, Discussie, Literatuur. Hoe je zo'n verslag schrijft ga je dit jaar leren. We beginnen in periode 1 met de Samenvatting.

In de weektaak 7 gaan jullie zelf de samenvatting schrijven van het onderzoek dat je hebt gedaan dit project. Daarnaast krijgen jullie in week 6 een workshop over het schrijven van de samenvatting.

Lees als voorbereiding op weektaak 7 en de workshop, samenvattingen 1 en 2 door, die op OnderwijsOnline te vinden zijn.

Beantwoord de volgende vragen:

- Wat valt je op aan de samenvattingen?
- Waar is de samenvatting in een wetenschappelijk artikel te vinden, denk je? Waarom juist daar?
- Wat zal het doel zijn van een samenvatting in een wetenschappelijk artikel?
- Zijn er verschillende secties in een samenvatting te ontdekken? Zo ja, welke?
- Voor wie is de samenvatting geschreven?
- In welke zin(nen) worden de resultaten van het onderzoek beschreven?
- Wat is de onderzoeksvraag in de artikelen?

Resultaat

Aan het eind van deze taak heb je

- een multiple alignment en fylogenetische boom van je 5 sequenties
- vergelijking van de fylogenetische boom met de soorten hiërarchie.
- dmv functies verbeterde Python scripts van de afgelopen weektaken.
- van elk script een flowchart dat de werking van het script visualiseert.
- de python scripts opgeslagen in GIT-classroom.
- de antwoorden op de vragen van opdracht 4.

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

- MAFFT Genome.jp: <http://www.genome.jp/tools-bin/mafft>
- NCBI taxonomy common tree: <https://www.ncbi.nlm.nih.gov/Taxonomy/CommonTree/wwwcmt.cgi> <https://www.ncbi.nlm.nih.gov/protein/>

Andere tools om alignments te maken:

- EBI pairwise alignment tools: <http://www.ebi.ac.uk/Tools/psa/>
- EBI multiple alignment tools: <http://www.ebi.ac.uk/Tools/msa/>
- NCBI multiple alignment tool COBALT: <https://www.ncbi.nlm.nih.gov/tools/cobalt/cobalt.cgi>
- Genome.jp alignment tools: <http://www.genome.jp/tools-bin/mafft>

Nabespreking

In de nabespreking en tijdens de presentaties komen onder andere de volgende punten aan de orde:

- Bespreken resultaten.
- Vergelijking van scripts, oude versie tov nieuwe versie.
- Hoe ziet een goede flowchart eruit?
- Wat was eenvoudig en wat was lastig?
- Hoe verloopt de samenwerking in de projectgroep?