Om inzicht te verkrijgen in de genetische variabiliteit van HIV-1 en om een verklaring te vinden voor zijn virulentie worden de subtypen van HIV-1 vergeleken met het HIV-2 en SIV op DNA- en proteïneniveau.

‘Wat is de fylogenetische oorsprong van HIV-1?’ & ‘Waarom is HIV-1 hoog virulent?’

Om hier antwoord te krijgen worden de volgende deelonderzoeksvragen geformuleerd:

1. Verschillen de GC percentages van de HIV/SIV subtypen?
2. Is er een verschil tussen het codon gebruik van het gen dat voor het HIV/SIV oppervlakteproteïne codeert en het codon gebruik van interne HIV/SIV genen?
3. Verschilt de aminozuur samenstelling van het oppervlakteproteïne met dat van de interne virale proteïne?
4. Wat is de mate van verschil tussen HIV en SIV subtypen als je vergelijkt op basis van een oppervlakteproteïne of op basis van een intern proteïne?

Stappenplan voor het uitvoeren van het HIV project:

1. Verzamelen sequenties
2. Onderzoek GC percentage
3. Onderzoek codon-gebruik
4. Karakterisering virale proteïne
5. Kenmerken oppervlakte proteïne
6. Fylogenetisch onderzoek

Weektaak1

Doelstellingen:

1. Uitleggen hoe een virus is opgebouwd en hoe het HIV verschilt van andere virussen
2. De levenscyclus van het HIV virus en zijn verspreiding in het menselijk lichaam uitleggen .
3. Relavante literatuur zoeken en bestuderen m.b.t. het centrale onderwerp.

Opdracht 1

1. Welke verschillende soorten virussen bestaan er?

Bacteriofagen, dierlijkevirussen, bacteriën die planten infecteren.

1. Hoe zijn deze virussen opgebouwd? Waaruit kan het ‘genoom’ bestaan?

Capside, capsomeren, ds DNA, ssDNA, dsRNA, ssRNA, membraan

1. Hoe infecteert een dierlijk virus een cel en hoe werkt, afhankelijk van zijn ‘genoom’, de levenscyclus van deze virussen?
2. Hoe is het HIV virus opgebouwd?

2 identieke RNA strands, 9 genen, 9 primaire eiwitten, 18 eiwitproducten, verschillende reading frames.

1. Hoe infecteert het HIV virus een cel en hoe vindt zijn levenscyclus plaats?
2. Wat is AIDS?

Weektaak 2

Doelstellingen:

1. Toelichten hoe het GC percentage in een genoom varieert.
2. Toelichten hoe het GC percentage tussen genomen varieert.
3. Het GC percentage van een DNA (of RNA) sequentie bepalen.

Je wil het GC percentage van de genomen onderzoeken en onderling vergelijken. Bij prokaryoten varieert het GC percentage behoorlijk. Bij eukaryoten varieert het GC percentage voornamelijk binnen het genoom.

Opdracht 1

Analyseer de GC percentages van verschillende genenomen.

1. Selecteer hiervoor:

1 humaan chromosoom

1. planten chromosoom
2. bacteriële genomen
3. Genereer voor elke DNA sequentie een ‘GC plot’. 🡪 Veranderingen van het GC percentage afhankelijk van de positie in het molecuu..
4. Bepaal (afhankelijk van de grootte van het chromosoom per 100, 10000 of 100000 bp het GC percentage.

Vergelijk voor de analyse van je resultaten het gemiddelde GC percentage en het verloop van de GC percentages met de locatie van genen op de chromosomen.

1. Waar in het menselijk/planten genoom is het GC percentage hoger: in coderende of in niet-coderende onderdelen van het genoom?

In de coderende onderdelen van het genoom zal het GC% hoger zijn dan in de niet-coderende delen. Dit komt doordat een hoger GC% betekent dat het DNA stabieler zal zijn, dit is een belangrijke factor voor coderende onderdelen.

1. Geldt hetzelfde voor bacteriën?
2. Waarom is het GC percentage niet altijd 50%

Omdat coderende sequenties beter geconserveerd moeten worden dan niet coderend DNA

1. Wat zijn de verschillen tussen eukaryoten en prokaryoten que GC percentages?

Het GC% bij prokaryoten kan variëren tussen de 25 en 75% bij eukaryoten varieert dit tussen de 40en 45%

1. Kun je de verschillen in GC percentages tussen eukaryoten en prokaryoten verklaren?

Prokaryoten hebben gemiddeld een kleiner genoom dan eukaryoten, daarom is de gendichtheid bij prokaryoten ook hoger en staat het aantal genen in verhouding met de grootte van het genoom.

1. En tussen de prokaryoten onderling?
2. Waarom wordt het GC percentage van genomen onderzocht?

Het GC% van genomen wordt onderzocht omdat het GC% een indicator is van waar genen zouden kunnen liggen.

Weektaak 3

Doelstellingen:

1. Toelichten hoe en vooral waarom het codongebruik (usage) bij verschillende soorten varieert.
2. Toelichten wat het verschil tussen codongebruik en codon bias is.
3. Toelichten wat wobble baseparing is tussen codons en anticodons.
4. Het codongebruik van genen bepalen
5. Een eindrapport schrijven waarin op een overzichtelijke manier de uitgewerkte resultaten van het onderzoek beschreven zijn.

Omdat meerdere codons voor hetzelfde aminozuur coderen, vindt je bij verschillende soorten organismen een ’voorkeur’ voor bepaalde codon. Zou het verschil in mutatiesnel tussen oppervlakteproteïne en interne virale genen een relatie hebben met het codongebruik? Je gaat onderzoeken of het codongebruik verschilt tussen interne genen en oppervlakte genen. Om het codongebruik te visualiseren maak je een grafiek waarin je per codon per aminozuur het percentage van voorkomen laat zien.

Opdracht 1

1. Waarom hebben organismen vaak een codon bias?
2. Niet alle genen hebben een codon usage die past bij de codon bias. Waarom is dat zo?
3. Wat wordt tijdens de evolutie beïnvloedt door wat?
   1. Het GC-percentage beïnvloedt de codon bias.
   2. De codon bias beïnvloedt het GC-percentage

Onderzoek het codongebruik van een gen dat voor een sterk geconserveerd enzym uit bijvoorbeeld de glycolyse of citroenzuurcyclus codeert. Selecteer vervolgens van 4 organismen (van vorige week) de coderende nucleotidesequenties van dit gen. Bepaal per gen het codongebruik en onderzoek eventuele verschillen.

Wat is de kans dat een willekeurige sequentie voor tryptofaan codeert?

Ga uit van A:C:G:T is 0.25:0.25:0.25:0.25.

Wat is de kans dat een willekeurige gesequentie voor alanine codeert?

Ga uit van A:C:G:T is 0.25:0.25:0.25:0.25.

Stel dat het coderende aandeel van een bepaald genoom bestaat uit: 3000A, 4000T, 5000C& 4500G.

Wat is de kans dat een coderende sequentie voor het codon 'CGT' codeert?

Wat is de kans dat een coderende sequentie zowel voor het codon 'TAT' als ook voor het codon 'CGT' codeert?

Wat is de kans dat een coderende sequentie voor de codons 'TAT' of 'CGT' codeert?

Weektaak 4

Doelstellingen

1. Van elk aminozuur de 1 leter en 3 letter codes opnoemen.
2. De aminozuren indelen in polaire en niet-polaire animozuren
3. De specifieke eigenschappen van aminozuren benoemen
4. De functionele eigenschappen van aminozuren verklaren
5. Eiwitsequenties opzoeken in een biologische database.
6. De opbouw in domeinen van eiwitsequenties opzoeken in een biologische database

Net als de genen kunnen de virale proteïnen verdeeld worden in oppervlakteproteïne en interne proteïne. Je wil onderzoeken of de eigenschappen van deze twee groepen proteïnen verschillen. Kan je door het vergelijken van de groep oppervlakteproteïnen of de groep interne proteïnen van HIV1 met de proteïnen van de andere virussen een verklaring vinden voor de pathogeniciteit van HIV1.

Opdracht 1

1. Wat zijn de 1 en 3 lettercodes van alle aminozuren?
2. Welke aminozuren zijn polair? Welke aminozuren zijn apolair?
3. Welke aminozuren zijn hydrofoob? Welke zijn hydrofiel?
4. Welke bijzondere rol hebben de aminozuren cysteïne en tryptofaan in een eiwitstructuur?
5. Zijn er nog meer aminozuren met belangrijke eigenschappen?

**Stap1 genoomsequenties**

De genoomsequenties, mRNA en proteïnesequenties van HIV-1, HIV-2, SIV en SIVmnd-2 virussen zijn via de NCBI website beschikbaar. De genoomsequenteis, de CDS sequenties en eiwitsequenties van de verschillende virussen worden gedownload in   
FASTA format voor verdere analyses.

**Stap 2 GC percentage**

Door middel van een Python script wordt het GC percentage van het genoom van elk virussubtype berekend. Op basis van de drie GC percentages worden het gemiddelde GC percentage, de variantie en de mediaan van het GC percentage berekent. Daarnaast wordt van het virusgenoom een GC plot gegenereerd om het verloop van het GC percentage over het genoom te visualiseren. Het GC percentage wordt vergeleken met het verloop van het GC percentage. Hierbij wordt rekening gehouden met de positionering van de genen op het genoom.

**Stap 3 codon usage**

Het codongebruik (‘codon usage’) van de coderende sequentie van de genen van elk subtype (HIV1, HIV2, SIV en SIVmnd2) wordt geanalyseerd met behulp van een python script. Voor elk subtype worden de genen in twee groepen opgedeeld. Groep 1 bevat het gen, dat voor het oppervlakteproteïne codeert. Groep 2 bevat de interne virale genen.

Om vast te kunnen stellen of er een codon bias is, wordt per aminozuur het percentage gebruikte codons bekeken. Door het gebruik van percentages kunnen de groepen onderling en de virussen onderling met elkaar vergeleken worden.

Eisen pyhton scripts:

Script 1: om het codon gebruik te analyseren leest het python script de CDS sequentie in fasta format en berkent het codongebruikt (de frequentie dat elk codon voorkomt). Dit wordt weergegeven in een plot.

**Stap 4 proteïnekarakterisering**

De virale proteïne van de subtypen worden in twee groepen verdeeld. Groep 1 bevat de oppvervlakteproteïnen en groep 2 bevat de interne virale proteïnen. Door middel van een python script worden de proteïnen van elke groep geanalyseerd.

Voor elke groep worden de volgende kenmerken berekend:

1. Het percentage cysteïne
2. Het percentage tryptofaan
3. Het percentage hydrofobe aminozuren
4. Het percentage hydrofiele aminozuren
5. De drie meest voorkomende aminozuren met bijbehorende percentages
6. De drie minst voorkomende aminozuren met bijbehorende percentages

Eisen python script:

Het python script leest een proteïnesequentie in fasta format in en bepaalt de frequenties voor elk aminozuur als ook voor groepen van aminozuren.