

DNA-mutaties wiskundig berekend



Leerlingenbundel

Dit STEMOOV-project werd:

- Gemaakt door: Maartje De Weyer, Julie Scevenels, Sofie Vandemoortele en Kevin Watjou (studenten Educatieve Master Wetenschappen en Technologie - UHasselt)
- Onder begeleiding van: Ilse Engelen (UHasselt), prof. dr. Katleen Denolf (UHasselt)
- Gefinaliseerd door: Kristel Heerwegh (UHasselt)
- Met dank aan: Thomas Peeters (student Educatieve Master Wiskunde - UAntwerpen) voor hulp bij LaTeX en Python



INHOUDSTAFEL

INHOUDSTAFEL	1
1. Inleiding	2
Onderzoeksvraag	2
Doelstellingen	2
Volgorde van de onderdelen	3
Benodigdheden	3
2. Biologie	4
Introductie	4
Groepswerk	4
3. Wiskunde	5
Kansrekenen	5
Matrices	6
4. DNA-mutaties wiskundig berekend	7
DNA-mutaties wiskundig berekenen	7
Werking	7
DNA-mutaties in Python berekenen	7
Werking	7
Peerevaluatie	8
5. Fylogenetische Stamboom	9
Deel 1: Fylogenetische stamboom	9
Opstellen van een fylogenetische boom	10
Python	12
Toepassing primaten	12
Deel 2: Opdracht: Stel zelf een fylogenetische boom op	15
Peerevaluatie	17
Reflectie	18

1. Inleiding

Met het project 'DNA-mutaties wiskundig berekend' onderzoeken we hoe DNA verandert over de tijd. Tegenwoordig kan je redelijk eenvoudig en goedkoop een DNA testkit gebruiken om erachter te komen wat onze afkomst is. Maar wat betekent dit? Wat wordt er nu exact onderzocht in ons DNA dat informatie geeft over hoe mensen verschillen? Hoe kunnen we dit berekenen? En hoe veilig is het?

Niet enkel ons eigen DNA is een interessante bron van informatie. Ook veranderingen van een virus zoals het COVID 19 virus kunnen gerelateerd worden aan mutaties van DNA en processen die hierop inspelen. Wat betekent het wanneer er een 'nieuwe variant' van het virus gevonden is? En hoe kunnen we dit onderzoeken?

Onderzoeksvraag

Hoe verandert DNA overheen de tijd?

→ Deelvraag 1: Hoe wordt de basenverdeling van een DNA-sequentie na n generaties berekend op een efficiënte manier?

→ Deelvraag 2: Hoe wordt een fylogenetische stamboom opgesteld, die de verwantschap tussen verschillende soorten in kaart brengt?

Om deze onderzoeksvragen te beantwoorden hebben we eerst de juiste expertise nodig. Daarom gaan we verder kijken dan de individuele vakken op school en gaan de theoretische kennis van deze verschillende vakken combineren om echte uitdagingen en problemen op te lossen. Hiervoor heb je kennis nodig van zowel biologie, wiskunde als informatica.

Tijdens dit project zal je zowel individueel als in groep, zowel online als klassikaal werken.

Veel plezier in de zoektocht!

Doelstellingen

Deze doelstellingen heb je bereikt na het project:

- Kennis van DNA en mutaties en toepassing op verschillende cases
- Kritisch nadenken en reflecteren over een case
- Presentatie skills en leren samenwerken in groep
- Jezelf evalueren en feedback geven aan je medeleerlingen
- Kennis van kansrekenen en matrices en deze toepassen op DNA-berekeningen
- Leren werken met Python
- Onderzoekskompetenties

Volgorde van de onderdelen

Het project bestaat uit 5 grote onderdelen.

Door de opbouw van het project is het aangeraden om een bepaalde volgorde aan te houden bij het oplossen van de verschillende onderdelen:

1. Biologie: 4 lestijden
2. Wiskunde: 4 lestijden
3. DNA-mutaties wiskundig berekend: 1,5 lestijden
4. DNA-mutaties wiskundig berekend in Python: 2,5 lestijden
5. Fylogenetisch stamboom: 2 lestijden

Benodigdheden

Voor het onderdeel biologie en wiskunde heb je toegang nodig tot een computer met internetverbinding.

Voor het onderdeel informatica heb je toegang nodig tot een computer met internetverbinding en moet het programma Python/PyCharm geïnstalleerd zijn.

2. Biologie

Aantal lestijden: 4 uur (1,5 uur Bookwidget + 1,5 uur werken aan presentatie + 1 uur presenteren)

Benodigheden: computer met internetverbinding

Introductie

Het onderdeel biologie werkt met een Bookwidget¹.

Ga naar onderstaande website om te starten of volg de instructie van de leerkracht via smartschool:
www.bookwidgets.com/play/9BRWF92

De leerkracht zal op voorhand aangeven of je de bookwidget thuis maakt als voorbereiding of je tijd krijgt in de les om hieraan te werken.

Inhoud van de leerstof in de Bookwidget: sommige onderdelen zijn herhaling, andere onderdelen zijn nieuw (afhankelijk van klas tot klas).

- Celkern
- Chromosomen
- DNA-structuur
- Replicatie
- Transcriptie en translatie
- Mutaties

Groepswerk

Nu je alle belangrijke leerstof rond DNA hebt herhaald, kunnen we deze toepassen op echte situaties. Hieronder vind je vier **vraagstukken** die allemaal gelinkt zijn aan een toepassing van DNA.

Zie bijlage voor de opdrachten.

Jullie moeten zich verdiepen in de onderwerpen en hierbij zeker de vragen oplossen die gegeven zijn bij de opgave. Gebruik hiervoor bronnen die jullie opzoeken via het internet. Denk na over de betrouwbaarheid van de bronnen! Jullie bevindingen **presenteren**



¹ Bij de Bookwidget werden onderstaande bronnen gebruikt:

- Lumen Learning. (2013). The Nucleus and Cytoplasm | Anatomy and Physiology. Lumenlearning.com. <https://courses.lumenlearning.com/nemcc-ap/chapter/the-cytoplasm-and-cellular-organelles/>
- Rye, C., Wise, R., Jurukovski, V., DeSaix, J., Choi, J., & Avissar, Y. (2016). 11.1 The proces of Meiosis [E-book]. In *Biology*. OpenStax. <https://openstax.org/books/biology/pages/11-1-the-process-of-meiosis>
- Berckmoes, R., Maesen, P., & Dirkx, M. (2014). Thema 1 Erfelijkheid. In *Naturalis 5 leerwerkboek* (pp. 80-151). Plantyn.
- DNA replication - 3D. (2015, 26 juni). [Video]. YouTube. https://www.youtube.com/watch?v=TNKWgcFPHqw&feature=emb_logo
- From DNA to protein - 3D. (2015, 7 januari). [Video]. YouTube. https://www.youtube.com/watch?v=qG7uCskUOrA&feature=emb_logo
- Talking Glossary of Genetics (Courtesy: National Human Genome Research Institute). (2012). Chromosome mutation. Commons.wikimedia.org. https://commons.wikimedia.org/wiki/File:Chromosome_mutation.png
- Sickel Cell Anemia. (2011, 10 november). [Video]. YouTube. https://www.youtube.com/watch?v=ZSpFVPBFeLk&feature=emb_logo

jullie aan elkaar op een manier die jullie zelf kiezen: powerpoint presentatie, poster, prezi, mindmap, keynote of andere opties. Voeg ook zeker een overzicht van de gebruikte bronnen toe. Jullie worden beoordeeld op de inhoud en op de presentatie zelf door de leerkracht en je medeleerlingen.

3. Wiskunde

Kansrekenen

Aantal lestijden: 1 uur

Benodigheden: computer met internetverbinding

Om DNA-mutaties wiskundig te berekenen en te begrijpen, hebben we ook een opfrissing nodig van kansrekenen. Verschillende aspecten en denkwijzen komen namelijk daarin terug.



Het onderdeel kansrekenen werkt met een Bookwidget². Hiervoor heb je een computer met internetverbinding nodig.

Ga naar onderstaande website om te starten of volg de instructie van de leerkracht via smartschool: www.bookwidgets.com/play/BBSUFBX

De leerkracht zal op voorhand aangeven of je de Bookwidget thuis maakt als voorbereiding of je tijd krijgt in de les om hieraan te werken.

In de Bookwidget onderstaande leerstof rond kansrekenen voorzien:

- Kansen inleiding
- Uitkomstenverzameling en gebeurtenis
- Formule van Laplace
- Complement
- Productregel voor onafhankelijke gebeurtenissen
- Somregel
- Voorwaardelijke kans
- Algemene productregel



² Bij de Bookwidget werden onderstaande bronnen gebruikt:

- Callaert, H. (2019). *Kansrekening voor de tweede graad*. https://www.uhasselt.be/documents/uhasselt@school/lesmateriaal/statistiek/Lesmateriaal/1_LEERLING%20kansrekening%20tweede%20graad_05.pdf.
- Faes, C. (2019). *Mutaties in ons DNA*. Universiteit Hasselt
- Khan Academy. (2017). *Probabilities in genetics*. Khan Academy. <https://www.khanacademy.org/science/high-school-biology/hs-classical-genetics/hs-introduction-to-heredity/a/probabilities-in-genetics>
- Delodderre, N., De Wilde, N., Op de Beeck, R., Somers, G., & Tytgat, P. (z.d.). *Delta Nova 5/6 Kansrekenen 6/8u* (pp 70-125). Plantyn.
- Wiskunde Academie. (2013). *Wiskunde Academie* [Video]. YouTube. https://www.youtube.com/channel/UCw_N3obzNBAbDXV0pSrr_xA

Matrices

Aantal lestijden: 3 uur

Het DNA bevat een enorme schat aan informatie. Dit maakt het soms moeilijk om het overzicht te bewaren. Om overzichtelijk te kunnen rekenen met zo een grote hoeveelheid gegevens worden in de praktijk meestal matrices gebruikt. In dit hoofdstuk wordt aan jullie een introductie gegeven over het onderwerp matrices

Om jullie bekend te maken met het onderwerp matrices maken jullie zelfstandig of per twee de **invulbundel matrices**.³ Hierbij leren jullie:

- het begrip matrix
- invullen en interpreteren van matrices
- bewerkingen uitvoeren met matrices (optellen, aftrekken, vermenigvuldigen, machten)

Er wordt hierbij een onderscheid gemaakt tussen **basisopdrachten** en **uitbreidingsopdrachten**. De basisopdrachten moeten zeker gemaakt worden om te kunnen werken aan het volgende hoofdstuk. De uitbreidingsopdrachten kunnen jullie ook nog maken om meer inzicht te verwerven in het onderwerp matrices. In onderstaande differentiatietabel kunnen jullie het overzicht terugvinden.

<u>Basisopdrachten</u>	<u>Uitbreidingsopdrachten</u>
Oefening 1, 2, 3, 5, 6, 7, 8, 10, 13, 14	Oefening 4, 9, 11, 12

³ Bij de invulbundel werden volgende bronnen gebruikt:

- Deloddere, N., De Wilde, N., Op de Beeck, R., Somers, G., & Tytgat, P. (z.d.). *Delta Nova 5/6 Matrices en Stelsels 6/8u* (1ste editie) (pp 8-21, 28-29, 34-37). Plantyn.
- Jones, P. W., & Smith, P. (2009). *Stochastic Processes: An Introduction, Second Edition (Chapman & Hall/CRC Texts in Statistical Science)* (pp 70-71, 88). Chapman and Hall/CRC.

4. DNA-mutaties wiskundig berekend

Aantal lestijden: 4 uur

DNA-mutaties wiskundig berekenen

In dit deel zullen we op een wiskundige manier het effect bestuderen van mutaties op de verdeling van de nucleotiden voor toekomstige generaties. Voor dit hoofdstuk werken we weer met een invulbundel. Deze bundel dienen jullie in te dienen na de vooropgestelde 4 lestijden en zal beoordeeld worden, zowel op basis van jullie geschreven oplossingen als jullie ingediende Python code. Meer details vinden jullie in de invulbundel terug!



Werking

Aantal lestijden: 1.5 uur

Je vindt de uitleg en opgave in de aparte **invulbundel DNA-mutaties wiskundig berekend**⁴.

DNA-mutaties in Python berekenen

Herinner je de onderzoeksvraag: we moesten niet enkel de verdeling na n generaties berekenen, maar ook op een efficiënte manier! Dit is de reden dat we dit in Python zullen doen. De opgave staat opnieuw in de **invulbundel DNA-mutaties wiskundig berekend**.

Werking

Aantal lestijden: 2.5 uur

Benodigdheden: computer met internetverbinding

Jullie werken in groepjes van 2 à 3 leerlingen. Elk groepje heeft zijn eigen opgave. Er zijn 10 verschillende opgaven, die elk drie bestanden bevatten:

- DNASequentieXa.txt
- DNASequentieXb.txt
- transitieMatrixX.txt

met X gelijk aan je groepsnummer.

⁴ Bij de invulbundel werd volgende bron gebruikt: Faes, C. (2019). *Mutaties in ons DNA*. Universiteit Hasselt

De gebruikte videotutorials⁵ in de invulbundel staan in de volgende tabel:

print()	len()	count()	matrices en vectoren in Python
			

Peerevaluatie

Geef de samenwerking met je groepsleden een score op 5. Geef jezelf ook een score.

Naam Groepslid:

1	2	3	4
Onvoldoende	Voldoende	Vaardig	Gevorderd

Naam Groepslid:

1	2	3	4
Onvoldoende	Voldoende	Vaardig	Gevorderd

Naam Groepslid:

1	2	3	4
Onvoldoende	Voldoende	Vaardig	Gevorderd



⁵ Onderstaande bronnen werden gebruikt:

- UHasselt Tutorials. (2018, 10 september). *De functie print() en het datatype "string"*. YouTube. https://www.youtube.com/watch?v=uoBdW-D9mg&list=PL2iW_rkiCt7UqxL47IGkBaTgApj_OHJ0K&index=4
- UHasselt Tutorials. (2020, 10 december). *De functie len()*. YouTube. https://www.youtube.com/watch?v=7krSoGswpdc&list=PL2iW_rkiCt7UqxL47IGkBaTgApj_OHJ0K&index=38
- Ryan Noonan. (2017, 5 april). *Python Count Occurrences of Letters, Words and Numbers in Strings and Lists*. YouTube. https://www.youtube.com/watch?v=szIFFw_Xl_M
- *Python Matrix and Introduction to NumPy*. (z.d.). [www.programiz.com](https://www.programiz.com/python-programming/matrix). Geraadpleegd op 28 december 2020, van <https://www.programiz.com/python-programming/matrix>

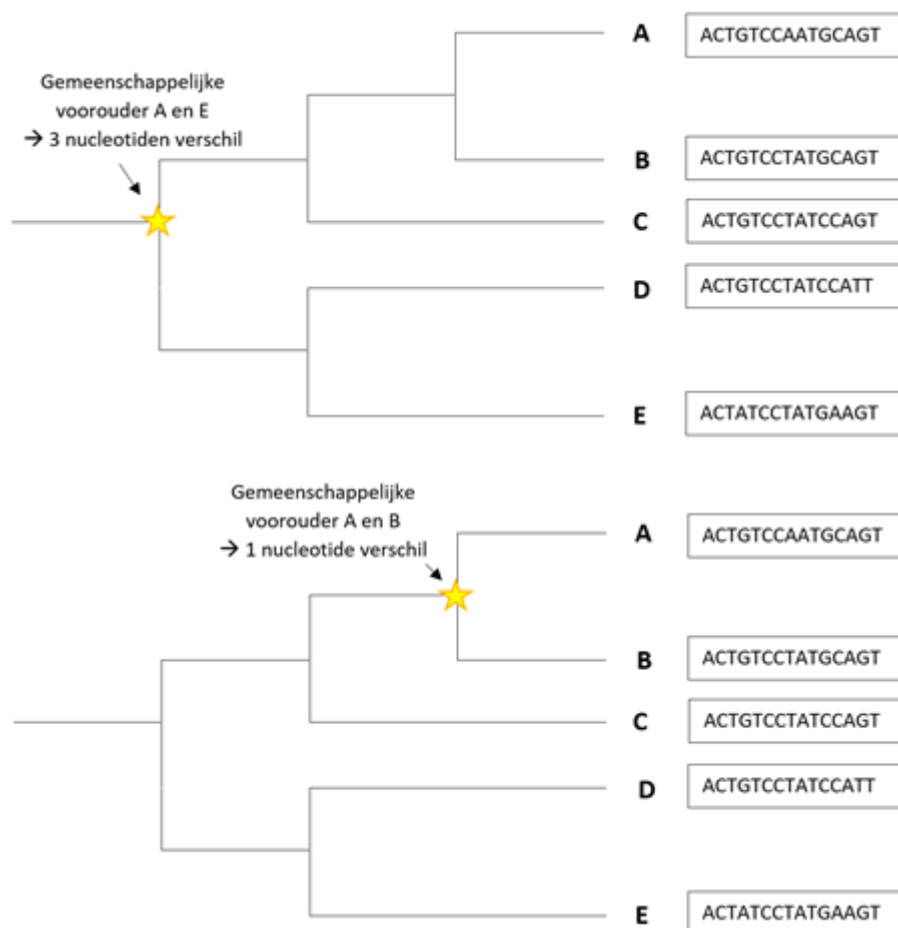
5. Fylogenetische Stamboom

Aantal lestijden: 2 uur

Benodigdheden: computer met internetverbinding en Python

Deel 1: Fylogenetische stamboom

Zoals je nu al ondervonden hebt uit de groepswerkjes voor biologie, bevat DNA een **enorme schat aan informatie** over ziektes, voorouders, verwantschappen, specifieke eigenschappen, enzoverder. De informatie in DNA kan ook gebruikt worden om de **verwantschappen** tussen verschillende soorten te bepalen. Hoe meer bepaalde soorten verschillen, oftewel hoe verder in de tijd de gemeenschappelijke voorouder ligt, hoe meer verschil in het DNA. Dit principe wordt ook wel **moleculaire of genetische klok** genoemd. Dit is geïllustreerd in de onderstaande afbeelding voor een stukje sequentie van 15 nucleotiden, maar in werkelijkheid is de code van het volledige genoom miljarden nucleotiden lang. Om verwantschappen te bepalen hebben we dus de hulp van informatica nodig⁶.



⁶ Rye, C., Wise, R., Jurukovski, V., DeSaix, J., Choi, J., & Avissar, Y. (2016). 11.1 Determining evolutionary relationships [E-book]. In *Biology*. OpenStax. <https://openstax.org/books/biology/pages/20-2-determining-evolutionary-relationships#0>

De afstand tussen genetische DNA-sequenties wordt ook wel de **fylogenetische afstand** genoemd. Hoe groter deze afstand tussen organismen, hoe minder deze bepaalde eigenschappen met elkaar gemeenschappelijk hebben. Indien deze schematisch worden voorgesteld zoals hierboven, spreken we van een **fylogenetische boom**. De positie van de splitsingen vertellen ons ook veel: hoe meer naar links, hoe langer het geleden is dat de soort afgesplitst is⁷.

Opstellen van een fylogenetische boom

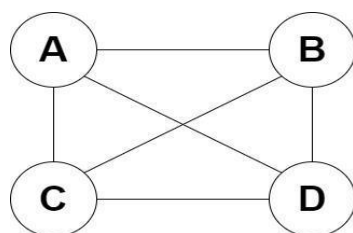
Om zo een fylogenetische boom op te stellen voor een aantal DNA-sequenties, moeten volgende stappen uitgevoerd worden:

1. **DNA-sequenties 'op elkaar leggen':** (= 'DNA-aligning' in het Engels) DNA-sequenties van verschillende soorten hebben vaak een verschillende lengte. Dit is omdat er soms 'gaten' komen in het DNA door mutaties die informatie verwijderen. Soms worden er ook extra nucleotiden geïntroduceerd en wordt de DNA-sequentie langer.

Kijk eens naar het verschil tussen het bestand 'primaten_unaligned.fasta' voordat de DNA sequenties op elkaar gelegd worden en het bestand 'primaten_aligned.fasta' nadat deze op elkaar gelegd zijn. Hoe worden deze gaten opgevuld?

Hoeveel soorten staan er in deze bestanden?

2. **Afstand tussen de DNA-sequenties berekenen:** Tussen elke soort wordt een afstand berekend. Deze afstanden worden dan in een afstandsmatrix (= 'distance matrix') gezet. *Stel dat er 4 soorten A, B, C en D zijn. Als je naar volgend diagram kijkt, hoeveel afstanden moeten er berekend worden? Schrijf deze afstandsparen op. (Bv. A-->B)*



Deze afstanden worden dan in een afstandsmatrix gezet:

	A	B	C	D
A	0	0	0	0
B	0.7	0	0	0
C	0.35	0.81	0	0
D	0.4	0.26	0.1	0

⁷ Rye, C., Wise, R., Jurukovski, V., DeSaix, J., Choi, J., & Avissar, Y. (2016). 11.1 Determining evolutionary relationships [E-book]. In *Biology*. OpenStax. <https://openstax.org/books/biology/pages/20-2-determining-evolutionary-relationships#0>

Welke soort matrix is dit?

Er staan erg veel nullen in deze matrix. Hoe komt dit denk je? (Beantwoord de deelvragen hieronder):

- *Wat zeggen de nullen op de diagonaalelementen?*

- *Waarom staan er nullen in de rechterbovenhoek?*

Indien je de fylogenetische afstand tussen 100 soorten wil berekenen, zou je al 4900 afstanden moeten berekenen!

3. **Fylogenetische boom opstellen:** hiervoor bestaan verschillende methodes: wij maken gebruik van het UPGMA-algoritme.

Een **algoritme** is een vast stappenplan dat altijd op dezelfde manier wordt gevolgd, zelfs als dit niet geprogrammeerd wordt op een computer! Een algoritme is immers een vaste reeks van instructies die altijd op dezelfde manier wordt uitgevoerd. De reden dat algoritmes toch meestal op computers worden geprogrammeerd, is omdat een computer zich uitstekend leent om deze algoritmes snel en efficiënt uit te voeren en dit duizenden tot miljoenen keren na elkaar.

Kan je zelf op zo een algoritme (=vast stappenplan) komen dat je al kent? (Tip: denk eens aan de wiskundeles)

Python

De code hiervoor staat in een **Python** programma dat al voor jou is geschreven. Als je ooit met software te maken krijgt, zal je vaker met code van anderen te maken krijgen. Je moet niet elk klein detail kennen van deze code, maar je weet wel wat deze ongeveer doet.



Probeer de drie stappen van het algoritme hierboven aan te duiden in de volgende code:

```

1  # Import modules
2  from Bio import Phylo
3  from Bio.Phylo.TreeConstruction import DistanceCalculator
4  from Bio.Phylo.TreeConstruction import DistanceTreeConstructor
5  from Bio import AlignIO
6
7  # Read the sequences and align
8  aln = AlignIO.read('primaten_aligned.fasta', 'fasta')
9
10 # Print the alignment
11 print(aln)
12
13 # Calculate the distance matrix
14 calculator = DistanceCalculator('identity')
15 dn = calculator.get_distance(aln)
16
17 # Print the distance Matrix
18 print('\nDistance Matrix\n=====')
19 print(dn)
20
21 # Construct the phylogenetic tree using UPGMA algorithm
22 constructor = DistanceTreeConstructor()
23 tree = constructor.upgma(dn)
24
25 # Draw the phylogenetic tree
26 Phylo.draw(tree)
27
28 # Print the phylogenetic tree in the terminal
29 print('\nPhylogenetic Tree\n=====')
30 Phylo.draw_ascii(tree)

```

Deze code staat op ook op SmartSchool. Open deze code in PyCharm CE. Voordat je deze code kan runnen, moet je eerst de pakketten 'Bio' en 'Matplotlib' installeren. Dit doe je op exact dezelfde manier als numpy.

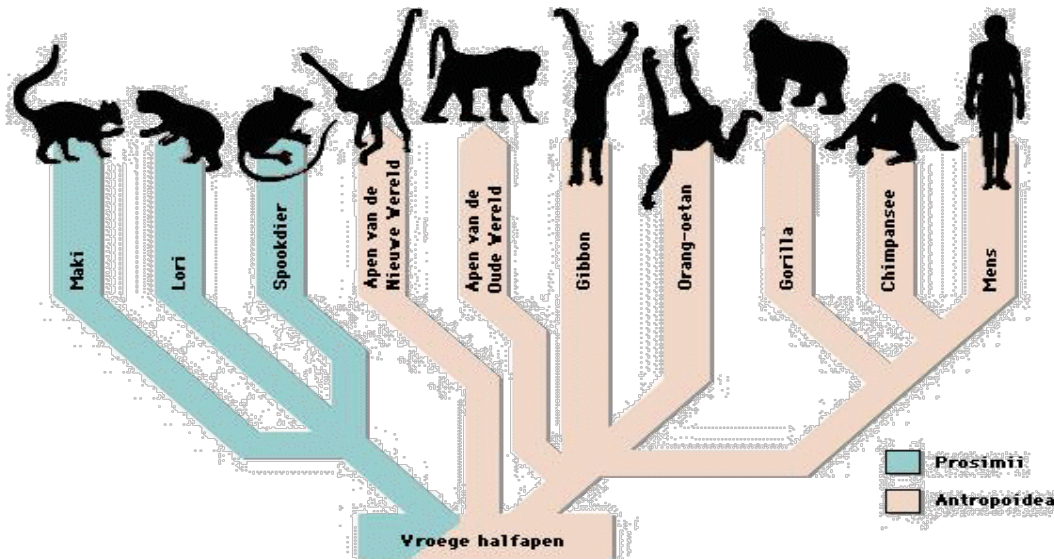
Toepassing primaten

We willen nu graag weten hoe de volgende primaten verwant zijn: mens (*Homo sapiens*), chimpansee (*Pan troglodytes*), gorilla (*Gorilla gorilla*) en oerang-oetang (*Pongo pygmaeus*). Via een databank aangeboden door The National Center for Biotechnology Information (<https://www.ncbi.nlm.nih.gov/>) hebben we het mitochondriale genoom van deze soorten kunnen verkrijgen. Dit zijn de soorten die in het bestand 'primaten_aligned.fasta' staan. Ook is er nog een andere soort toegevoegd die tot een heel andere familie hoort dan de rest van de vier primaten, nl.

de noordelijke giraf (*giraffa camelopardalis*). Dit is een testsoort: als de code correct is, dan moet de giraf het minste verwant zijn.

Probeer de code⁸ eens te runnen, dit is het bestand 'FylogenetischeBoom.py'. Je ziet dan de fylogenetische boom van deze soorten. De eerste keer dat je dit doet, zal dit even duren. Zorg ook dat je console breed genoeg uitgetrokken is, dat je de boom zeker goed kan zien.

9



Beantwoordt onderstaande vragen en leg uit:

- Juist of fout: De giraf en de gorilla zijn dicht bij elkaar verwant dan de Chimpansee en de orang-oetan.
- Wat is de afstand tussen de orang-oetan en de mens? (rond af op 2 getallen na de komma).

Wat betekent deze afstand concreet?

Welke twee andere soorten hebben dezelfde afstand?

- Welk soortenpaar is het meest verwant?
- Bekijk de fylogenetische boom. Beargumenteer, in de context van de evolutie van de mens, waarom de stelling "De mens heeft een gemeenschappelijke voorouder met hedendaagse apen." meer wetenschappelijk correct is dan "De mens stamt af van apen".

⁸Mallawaarachchi, V. (2017). *TreeConstructionDistanceBased.py*. Gist.github. <https://gist.github.com/Vini2/0723caa046c6b8e1d745a623f2695abb>

⁹ <https://www.tongerlo.org/2016/06/03/schepping-of-evolutie/>

Deel 2: Opdracht: Stel zelf een fylogenetische boom op

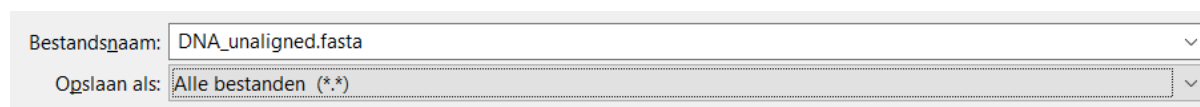
Kies nu zelf 4 soorten van eenzelfde familie (bijv. katachtigen, vissen, kikkers, ...).

Schrijf deze op, inclusief de Latijnse naam:

-
-
-
-

We kiezen ook weer een testsoort van een andere familie, we nemen hiervoor de mens. De latijnse benaming hiervoor is de homo sapiens.

Doe het kladblok open in Windows. Je moet hier nog niets inschrijven, maar sla dit alvast op onder de naam 'DNA_unaligned.fasta'. Zorg dat je dit onder 'Alle bestanden' opslaat. Zorg ook dat dit in dezelfde map staat als je Python project.



Nu moeten je deze DNA-sequenties opzoeken in een databank. Hieronder staat beschreven hoe je dit doet voor de Homo Sapiens:

1. Open de site <https://www.ncbi.nlm.nih.gov/>
2. Voor je in de zoekbalk typt, zorg dan dat je binnen de nucleotide databank zoekt.



In de databank schrijf je de latijnse naam en erachter schrijf je 'complete'. Dit is omdat er soms maar gedeeltelijke DNA-sequenties in de databank worden gezet. Voor ons onderzoek is dit echter niet voldoende en willen we een volledige sequentie van het mitochondriaal DNA. Duidt daarom ook zeker in de linkerkolom 'Mitochondrion' aan onder 'Genetic compartments'

**Genetic
compartments**
Chloroplast (2)
Mitochondrion (81,352)
Plasmid (30,332)
Plastid (16)

Je krijgt dan volgende zoekresultaten:

- ☐ [Homo sapiens isolate Igorot10 mitochondrion, complete genome](#)
 1. 16,569 bp circular DNA
Accession: KU752565.1 GI: 1017029704
[Protein](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)
- ☐ [Homo sapiens isolate Igorot23 mitochondrion, complete genome](#)
 2. 16,557 bp circular DNA
Accession: KU752564.1 GI: 1017029690
[Protein](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)
- ☐ [Homo sapiens isolate Igorot6 mitochondrion, complete genome](#)
 3. 16,559 bp circular DNA
Accession: KU752563.1 GI: 1017029676
[Protein](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)
- ☐ [Homo sapiens isolate Igorot17 mitochondrion, complete genome](#)
 4. 16,567 bp circular DNA
Accession: KU752562.1 GI: 1017029662
[Protein](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)

Al deze resultaten zijn complete genomen van de homo sapiens. Daarom zijn al deze sequenties goed. Merk ook op dat deze allemaal rond de 16000 nucleotiden zitten.

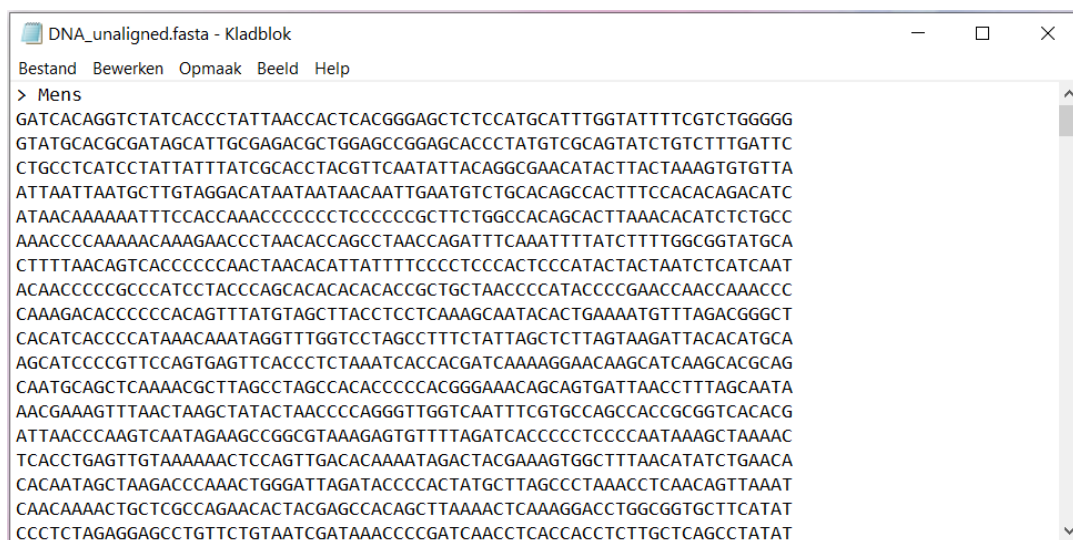
Klik bijvoorbeeld eens op de bovenste. Klik dan op FASTA, deze link staat onder de titel.

Homo sapiens isolate Igorot10 mitochondrion, complete genome

GenBank: KU752565.1

[FASTA](#) [Graphics](#)

Kopieer nu deze DNA-sequentie. Open je kladblok. Voor je deze sequentie plakt, schrijf even de naam van de soort op, in dit geval de mens. Doe dit door eerst een '>' teken te schrijven en dan de naam van de soort (zonder spaties). Plak dan de DNA-sequentie op de volgende lijn. Dit zou er als volgt moeten uitzien:



```

DNA_unaligned.fasta - Kladblok
Bestand Bewerken Opmaak Beeld Help
> Mens
GATCACAGGTCTATCACCCCTATTAACCACTCACGGGAGCTCTCCATGCATTTGGTATTTTCGTCTGGGGG
GTATGCACGCGATAGCATTGCGAGACGCTGGAGCCGGAGCACCCCTATGTCGCAGTATCTGTCTTTGATTTC
CTGCCTCATCTATTATTTATCGCACCTACGTTCAATATTACAGGCGAACATACTTACTAAAGTGTGTTA
ATTAATTAATGCTTGTAGGACATAATAACAATTGAATGCTGACAGCCACTTTCCACACAGACATC
ATAACAAAAATTTCCACCAAAACCCCTCCCCCGCTTCTGGCCACAGCACTTAAACACATCTCTGCC
AAACCCCAAAACAAAGAACCTAACACAGCCTAACAGATTTCAAATTTTATCTTTTGGCGGTATGCA
CTTTTAACAGTCACCCCCCACTAACACATTATTTCCCTCCCACTCCCATACTACTAATCTCATCAAT
ACAACCCCGCCCATCTACCCAGCACACACACCGCTGCTAACCCCATACCCGAACCAACCAACCC
CAAAGACACCCCCACAGTTTATGTAGCTTACCTCCTCAAAGCAATACACTGAAAAATGTTTAGACGGGCT
CACATCACCCCATAAACAAATAGGTTTGGTCTTAGCCTTTCTATTAGCTCTTAGTAAGATTACACATGCA
AGCATCCCCGTTCCAGTGAGTTACCCCTCTAAATCACCAGATCAAAGGAACAAGCATCAAGCACGCAG
CAATGCAGCTCAAACGCTTAGCCTAGCCACACCCCCACGGGAACAGCAGTGATTAACCTTTAGCAATA
AACGAAAGTTTAAGCTAAGCTATACTAACCCAGGGTTGGTCAATTTCTGTCAGCCACCGCGGTACACG
ATTAACCAAGTCAATAGAAGCCGGGTAAGAGTGTTTATAGTACCCCTCCCAATAAAGCTAAAAC
TCACCTGAGTTGTAAAAAAGTCCAGTTGACACAAAATAGACTACGAAAGTGCTTTAACATATCTGAACA
CACAATAGCTAAGACCCAACTGGGATTAGATACCCCACTATGCTTAGCCCTAAACCTCAACAGTTAAAT
CAACAAAGTCTCTGCCAGAACACTACGAGCCACAGCTTAAAGTCAAAGGACCTGGCGGTGCTTCATAT
CCCTCTAGAGGAGCCTGTTCTGTAATCGATAAACCCCGATCAACCTCACCACCTCTTGCTCAGCCTATAT
  
```

Open nu het bestand 'Omzetting.py' en zet dit bestand samen met 'muscle_muscle3.8.31_i86win32.exe' in dezelfde map. Run dan het bestand 'Omzetting.py', dit kan even duren. Indien hier problemen mee zijn, is de naam van je bestand waarschijnlijk niet correct ofwel is er iets misgegaan bij het knippen en plakken. Kijk dit dus nog eens na!

We hebben nu nog maar 1 soort in het bestand staan, nl. de testsoort. De vier andere soorten van dezelfde familie die je gekozen hebt, moeten nu nog toegevoegd worden op dezelfde manier. Probeer zeker het bestand 'Omzetting.py' elke keer eens te runnen als je één soort hebt toegevoegd. Zo weet je meteen wanneer je een fout hebt gemaakt en moet je achteraf niet je hele bestand gaan afspeuren!

Indien je dit gedaan hebt, moet je de file 'FylogenetischeBoom.py' runnen. Je moet wel nog aangeven dat je nu met je eigen bestand 'DNA_aligned.fasta' gaat werken. Schakel hiervoor lijn 8 uit en schakel lijn 9 in. Je krijgt dat opnieuw een fylogenetische boom.

Bespreek deze figuur:

- *Is je testsoort het minst verwant met de andere soorten?*

- *Doe drie uitspraken over deze fylogenetische boom. Probeer telkens een ander aspect van de boom te bespreken. (Tip: Indien je niets weet, haal dan je inspiratie hiervoor bij de vragen bij de vorige opdracht)*
 -

 -

 -

Peerevaluatie

Geef de samenwerking met je groepslid een score op 5. Geef jezelf ook een score.

Naam Groepslid:

1	2	3	4
Onvoldoende	Voldoende	Vaardig	Gevorderd

Naam Groepslid:

1	2	3	4
Onvoldoende	Voldoende	Vaardig	Gevorderd



Reflectie

Wat vind je van het eindresultaat?

Welke onderdeel van het project vond je het moeilijkste?

Van welk onderdeel heb je het meeste geleerd?

Welke kennis kan je meenemen naar andere vakken?

Welke andere toepassingen van Python kan je nog bedenken?

-
-
-

Wat zijn je bevindingen over dit STEM-project?

Heeft dit project je geholpen om toepassingen van de verschillende STEM-vakken beter te begrijpen?

Wat zijn minpunten of verbeterpunten?

Wat vond je leuk, interessant?

Zou je dit project aanraden aan je medeleerlingen?

