Inhoud

[Inleiding 3](#_Toc151974365)

[Materiaal en methoden 4](#_Toc151974366)

[Resultaten 5](#_Toc151974367)

[Discussie en Conclusie 6](#_Toc151974368)

[Referenties 7](#_Toc151974369)

[Bijlagen 8](#_Toc151974370)

## Inleiding

Zoals Charles Darwin ooit zei: “It is not the strongest of the species that survives, not the most intelligent that survives. It is the one that is the most adaptable to change.” Er is geen kader waar deze wijze woorden meer waar blijken dan bij het kader van fylogenie. In dit verslag wordt er een onderzoek naar de fylogenie van verschillende typen hemoglobine beschreven, en wordt de vraag “Hoe varieert de fylogenetische geschiedenis van hemoglobine-eiwitten?” beantwoord.

### Fylogenie

Fylogenie is een wetenschappelijk kader dat zich bezighoudt met het onderzoeken en vaststellen van relaties tussen verschillende soorten of groepen. Dit is een tak van de biologie die probeert te begrijpen hoe verschillende organismen op aarde zich ontwikkelen en evolueren. Fylogenetische analyse wordt gebruikt om stambomen te construeren, ook wel fylogenetische bomen genoemd, die een visuele weergave bieden van de evolutionaire afstammingslijnen en relaties van verschillende taxa.

### Fylogenetische bomen

Er zijn meerdere manieren om een fylogenetische boom in elkaar te zetten, en de verschillen tussen deze manieren wordt hieronder kort uitgelegd.

De meest gebruikelijke manier is “maximum parsimony”, waarbij de boom wordt geconstrueerd volgens het principe van minimale evolutionaire verandering. Met andere woorden: de boom die minder genetische veranderingen of mutaties nodig heeft om de waargenomen gegevens te verklaren, wordt als waarschijnlijker beschouwd. [1, 3]

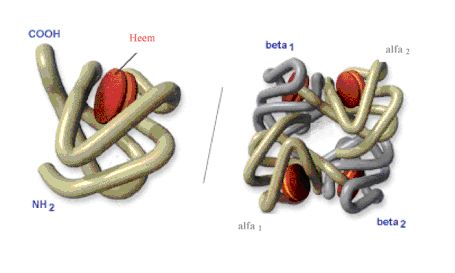
Een andere benadering is de methode “maximum likelihood”, waarbij de boom wordt gevonden die de hoogste waarschijnlijkheid heeft van de waargenomen gegevens gegeven het evolutionaire model. Deze benadering houdt rekening met probabilistische overwegingen en kan de complexiteit van evolutionaire processen beter modelleren. [2, 3]

De derde methode is UPGMA (Unweighted Pair Group Method with Arithmetic Mean), die is gebaseerd op een hiërarchische groeperingsmethode. Dit is de methode die bij dit onderzoek is gebruikt. UPGMA construeert een boom op basis van de gemiddelde afstand tussen taxa, ervan uitgaande dat de snelheid van evolutionaire verandering in alle takken hetzelfde is. Dit is een snelle en eenvoudige methode die vaak wordt gebruikt om een ​​eerste benadering van fylogenetische relaties te geven. [3] Om een boom in elkaar te zetten door middel van UPGMA worden er eerst pairwise alignments gemaakt van alle sequenties die in de boom moeten komen. Meer uitleg over deze methode en UPGMA is te vinden in de materiaal en methoden.

### Hemoglobine

De rode kleur van bloed is te danken aan de ijzerhoudende hemoglobine. Hemoglobine is een essentieel eiwit dat een centrale rol speelt in het transport van zuurstof in het bloed van veel gewervelde dieren, waaronder zoogdieren, vogels en zelfs sommige ongewervelde dieren zoals weekdieren. De voornaamste functie van hemoglobine is het binden en dragen van zuurstofmoleculen van de longen naar de weefsels en organen, en tegelijkertijd het vervoeren van koolstofdioxide, een bijproduct van de celstofwisseling, van de weefsels terug naar de longen voor uitademing. [4]

De unieke structuur van hemoglobine maakt het geschikt voor deze cruciale functie. Hemoglobine is een tetrameer-eiwit, wat betekent dat het bestaat uit vier subeenheden. Bij mensen en veel andere zoogdieren bestaat hemoglobine uit twee alfa-subeenheden en twee beta-subeenheden, die samen een stabiel kwartet vormen, zoals weergegeven in figuur 1. [4] Elk van deze subeenheden bevat een ijzeratoom, dat zich kan binden aan een zuurstofmolecuul. Hierdoor kan een enkel hemoglobinemolecuul maximaal vier zuurstofmoleculen binden.



Figuur 1 - Visuele weergave van hemoglobine [4]

In samenspel met andere componenten van het bloed, zoals rode bloedcellen, zorgt hemoglobine ervoor dat zuurstof efficiënt door het lichaam wordt getransporteerd om de energiebehoefte van de cellen te ondersteunen. Het begrip van de functie van hemoglobine is van groot belang bij het begrijpen van de ademhaling, het metabolisme en de algemene fysiologie van organismen.

## Materiaal en methoden

### Materiaal

Dit onderzoek begon met 5 aminozuur sequenties van hemoglobine eiwitten, waarvan de exacte oorsprong nog onbekend was. Dit was dus het eerste wat onderzocht werd. Dit was gedaan door middel van NCBI blastp. [5] Hieronder in tabel 1 staan deze sequenties weergegeven, en in tabel 2 staat achtergrondinformatie over deze sequenties weergegeven.

Tabel - De 5 onderzochte aminozuur sequenties

|  |  |
| --- | --- |
| Nummer | Aminozuur sequentie |
| 1 | mvhwtaeekqlitglwgkvnvaecgaealarllivypwtqrffasfgnlssptailgnpmvrahgkkvltsfgdavknldnikntfsqlselhcdklhvdpenfrllgdiliivlaahfskdftpecqaawqklvrvvahalarkyh |
| 2 | mvlsaadknnvkgiftkiaghaeeygaetlermfttypptktyfphfdlshgsaqikghgkkvvaalieaanhiddiagtlsklsdlhahklrvdpvnfkllgqcflvvvaihhpaaltpevhasldkflcavgtvltakyr |
| 3 | mvhltaeekslvsglwgkvnvdevggealgrllivypwtqrffdsfgdlstpdavmsnakvkahgkkvlnsfsdglknldnlkgtfaklselhcdklhvdpenfkllgnvlvcvlahhfgkeftpqvqaayqkvvagvanalahkyh |
| 4 | mvlspadktnikstwdkigghagdyggealdrtfqsfpttktyfphfdlspgsaqvkahgkkvadalttavahlddlpgalsalsdlhayklrvdpvnfkllshcllvtlachhpteftpavhasldkffaavstvltskyr |
| 5 | mpivdtgsvaplsaaektkirsawapvystyetsgvdilvkfftstpaaqeffpkfkglttadqlkksadvrwhaeriinavndavasmddtekmsmklrdlsgkhaksfqvdpqyfkvlaaviadtvaagdagfeklmsmicillrsay |

Tabel - Achtergrondinformatie over de 5 onderzochte aminozuur sequenties

|  |  |  |  |
| --- | --- | --- | --- |
| Nummer | Accession | Naam eiwit | Naam organisme |
| 1 | NP\_990820.1 | hemoglobin subunit beta | Gallus gallus (kip) |
| 2 | NP\_001004376.1 | hemoglobin subunit alpha-A | Gallus gallus (kip) |
| 3 | NP\_001257812.1 | hemoglobin subunit beta-like | Canis lupus familiaris (hond) |
| 4 | NP\_001257814.1 | hemoglobin subunit alpha-like | Canis lupus familiaris (hond) |
| 5 | XP\_032826327.1 | globin-5 | Petromyzon marinus (zeeprik) |

De volgende stap van dit onderzoek was om pairwise alignments uit te voeren met elke combinatie van aminozuur sequenties. Dit werd gedaan door middel van het schrijven van een eigen python [6] programma. Dit programma krijgt als input een fasta bestand van sequenties, maakt hier alle nodige pairwise alignments mee, maakt een distancematrix, en zet deze distancematrix tot slot om in Newick tree format [7] (uitleg hierover volgt nog). De stappen die het programma doorliep worden hieronder behandeld.

### Pairwise alignments

Als eerst moet er voor een pairwise alignment een scoringsmatrix en het type alignment gekozen worden. Een scoringsmatrix is een tabel waarin staat hoeveel score één paar aminozuren geeft. Als je bijvoorbeeld Arganine met Methionine vergelijkt, kan je een scoringsmatrix opzoeken hoeveel score dit paar waard is. De scoringsmatrix die voor dit onderzoek gekozen werd was BLOSUM80 [8] (BLOck SUbstitution Matrix), wat een scoringsmatrix is dat gemaakt is door naar sequenties te kijken met niet meer dan 80% paarsgewijze identiteit. Voor het type alignment valt er te kiezen uit een global alignment, en een local alignment. Er was gekozen voor een local alignment, en dus was het Smith-Waterman algoritme [9] gebruikt. Vervolgens kon de alignment beginnen, en zijn de 2 sequenties tegen elkaar opgezet in een soort tabel (alignmatrix), en is elk vakje (een vergelijking van 2 aminozuren) ingevuld op basis van het Smith-Waterman algoritme en het BLOSUM80 scoringsmatrix. Dit werkt als volgt: voor elk vakje wordt gekeken naar het vakje erboven, links ervan, en linksboven ervan (diagonaal). Voor elk van deze vakjes wordt de mogelijke nieuwe score van het lege vakje berekend. Voor het verticale en horizontale vakje zou betekenen dat het lege vakje een gap is, en wordt dus de gap-score (staat in de scoringsmatrix) genomen, en wordt dit bij het horizontale of verticale vakje zijn score toegevoegd (de gap score is altijd negatief). Dit wordt dan 1 van de mogelijke 3 scores van het lege vakje. Bij het diagonale vakje, zou dit betekenen dat er geen gap is, maar een match of mismatch. Om de potentiële score van het lege vakje hier te berekenen wordt de score van het diagonale vakje genomen, en wordt hieraan toegevoegd de score van de 2 aminozuren die vergeleken worden (door middel van de scoringsmatrix). Maar omdat er voor een local alignment is gekozen met het Smith-Waterman algoritme, wordt er aan de 3 mogelijke waarden van het lege vakje, ook nog de waarde 0 toegevoegd. Er kunnen dus geen negatieve getallen in de alignmatrix staan. Tot slot wordt van deze 4 getallen het hoogste getal gekozen, wordt dit getal in het lege vakje ingevuld, en als de oorsprong van dit getal 1 van de 3 andere vakjes was (dus niet de 0 van het Smith-Waterman algoritme) wordt er een pijl getekend van het oorsprong-vakje naar het zojuist ingevulde vakje. Op deze manier wordt de alignmatrix dus van linksboven naar rechtsonder gevuld. Dit wordt gedaan voor elke combinatie van sequenties. De optimale alignment wordt gevonden door het vakje met de hoogste score te vinden, en dit als startpunt te gebruiken. Het enige wat er nu nog gedaan moet worden is de pijlen omdraaien, en deze pijlen terug volgen tot een 0 wordt gevonden. Dit geeft voor elke alignmatrix dus een alignment van de 2 sequenties. Tussen de twee sequenties komt nog een regel er tussenin om aan de geven of de twee gepaarde aminozuren een match, mismatch of gap vormen. Met deze ‘midline’ wordt een score gegeven aan elke alignment, voor matches en mismatches wordt aan de totale score de score van de 2 gepaarde aminozuren toegevoegd, maar voor gap openings en gap extensions wordt respectievelijk -10 en -1 toegevoegd aan de totaalscore. Tot slot worden de scores van de alignments weergegeven in een tabel (5x5 tabel, alignment score tussen elke sequentie combinatie).

### Distance matrix

Met de 5x5 tabel van alignment scores, werd vervolgens een 5x5 tabel opgezet, genaamd de distance matrix. Deze distance matrix laat ons de evolutionaire afstand tussen 2 sequenties zien. De score met de formule in figuur 2 berekend. 

Figuur - Berekening voor de afstand (d) tussen sequentie i en j, door middel van alignment scores (S)

In deze formule staat di,j voor de distance tussen twee sequenties (i en j), en staat Si,j voor de alignment score tussen sequentie i en sequentie j.

### Boom maken door middel van UPGMA

## Resultaten

[results]

## Discussie en Conclusie

[d&c]

## Referenties

1. Kannan L, Wheeler WC. Maximum Parsimony on Phylogenetic networks. *Algorithms Mol Biol*. 2012;7(1):9. Published 2012 May 2. <https://doi.org/10.1186/1748-7188-7-9>
2. Brooks-Bartlett, Jonny. "Probability Concepts Explained: Maximum Likelihood Estimation." Towards Data Science, Jan 4, 2018. <https://towardsdatascience.com/probability-concepts-explained-maximum-likelihood-estimation-c7b4342fdbb1>
3. Rainer Borriss, Christian Rueckert, Jochen Blom, Oliver Bezuidt, Oleg Reva, Hans-Peter Klenk, 18 - Whole Genome Sequence Comparisons in Taxonomy, Methods in Microbiology Volume 38, Academic Press, 2011, <https://doi.org/10.1016/B978-0-12-387730-7.00018-8>.
4. Faculteit Wetenschappen KU Leuven, Hemoglobine en de rode kleur van bloed, Nov 9, 2009, <https://wet.kuleuven.be/wetenschapinbreedbeeld/lesmateriaal_biochemie/waaromisbloedrood/bloed_rood>
5. National Center for Biotechnology Information (US). BLASTP [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [cited 30/11/2023]. Available from: <https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome>
6. Python Software Foundation. (2023). Python Language Reference, version 3.10.6 Available at <https://docs.python.org/3/>
7. Newick
8. Blosum80
9. Smith-Waterman

## Bijlagen