

Stamtræer med og uden rod

I denne tekst gennemgås hvorledes man kan sammenligne længden af grenene i forskellige typer stamtræer (UPGMA og Neighbour Joining) med de beregnede grenlængder angivet i afstandsmatrixen.

Metode

De 5 sekvenser der er anvendt for at fremstille de stamtræer, vi ønsker at analysere, er sekvenser af komplette mitochondriegenomer (mtDNA), og disse sekvenser er fundet på NCBI's database via søgefunktionen i programmet Geneious. Sekvenserne er navngivet med en kode som er unik. Kodernes navne kaldes 'Accession numbers' og fremgår af figur 1.

Art	Accession numbers
Chimpanse	NC_001643
Bonobo	NC_001644
Gorilla	NC_001645
Orangutang	NC_001646
Menneske	NC_012920

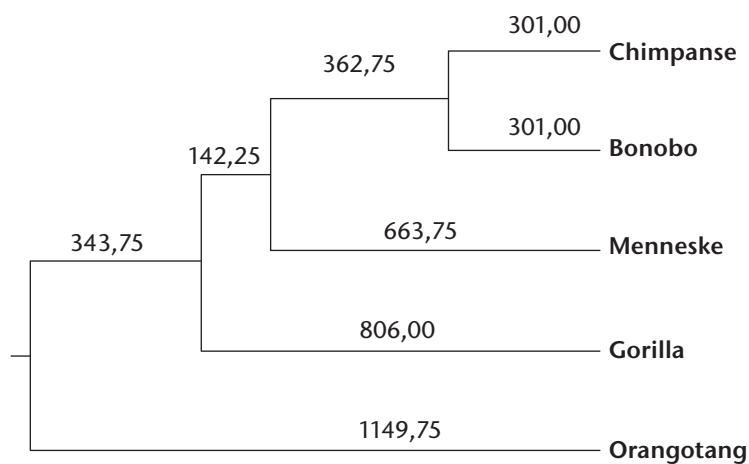
Figur 1. Accession numbers for de arter der indgår i de stamtræer, som bliver analyseret.

Sekvenserne alignes i programmet Genieous, og dette alignment er eksporteret til programmet MEGA som opstiller en afstandsmatrix, hvor antallet af forskelle i mtDNA mellem de forskellige sekvenser er vist. Det ses i figur 2 at antallet af forskelle mellem bonoboens og chimpansens sekvenser er 602 baser, og at forskellen mellem menneske og bonobo er 1321 basepar.

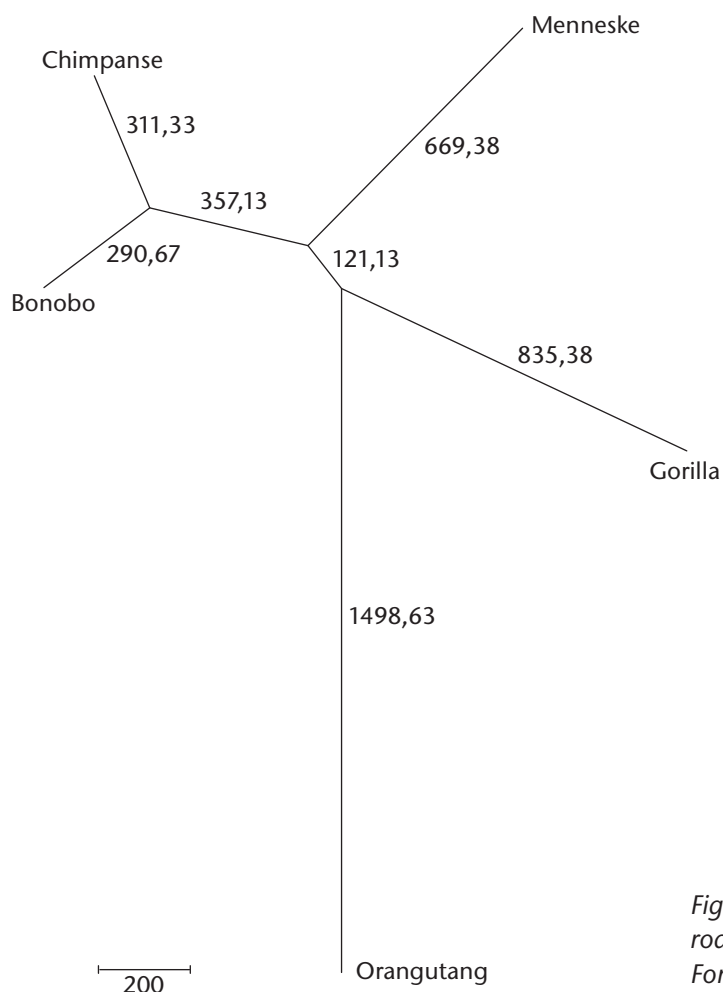
	1	2	3	4	5
1. NC 001645					
2. NC 001646	2334				
3. NC 001643	1608	2309			
4. NC 001644	1583	2285	602		
5. NC 012920	1645	2270	1334	1321	

Figur 2. Afstandsmatrix lavet i MEGA 7.0. I figuren er række 1 orangutang, række 2 er gorilla, række 3 er chimpanse, række 4 er bonobo og række 5 er menneske.

Der fremstilles et UPGMA-stamtræ med viste grenlængder i MEGA, se figur 3. Ligeledes fremstilles et Neighbour Joining-stamtræ med grenlængder som vist i figur 4.



Figur 3. UPGMA-stamtræ med grenlængder der symboliserer antal forskelle i basepar mellem de viste sekvenser. Forskellen er angivet langs de vandrette grene. Forskellen mellem chimpanse og bonobo er i dette træ $301 + 301 = 602$ basepar.



Figur 4. Neighbour Joining-stamtræ uden rod. Forskellen i basepar er vist på grenene. Forskellen mellem bonobo og chimpanse er her $290,67 + 311,33 = 602$ basepar.

Beregninger

Man kan se hvor mange forskelle der er mellem de forskellige arter i afstandsmatrixen i figur 2, og man kan herefter prøve at se hvilket stamtræ der bedst viser de forskelle angivet i matrixen. Af afstandsmatrixen fremgår det at afstanden mellem menneske og chimpansé er 1334 basepar, og at afstanden mellem gorilla og orangutang er 2334 basepar. Disse tal er de observerede forskelle. De skal nu sammenlignes med grenlængderne i stamtræerne i figur 3 og 4.

Herefter beregnes de kvadrerede forskelle ved brug af følgende formel:

$$\text{Kvadreret forskel} = (\text{Afstand basepar}_{\text{matrix}} - \text{Afstand basepar}_{\text{stamtræ}})^2$$

Det vil sige at jo større forskel der er på det talte i matrixen og det afbillede i stamtræet, desto større bliver den kvadrerede forskel. Grunden til at man kvadrerer forskellen er at man derved kun kommer til at arbejde med positive tal, og når man summerer for et helt stamtræ tæller alle forskelle positivt.

UPGMA-stamtræ, figur 3

Afstanden mellem menneske og chimpansé udgøres af grenlængderne:
 $301,00 + 362,75 + 663,75 = 1327,50$ basepar.

Afstanden mellem gorilla og orangutang udgøres af grenlængderne:
 $806,00 + 343,75 + 1149,75 = 2299,50$ basepar.

Matrixværdierne for de to afstande er henholdsvis 1334 basepar og 2334 basepar. Udregning for menneske-chimpansé bliver derfor:

$$(1334 - 1327,50)^2 = 42,25$$

For gorilla-orangutang:

$$(2334 - 2299,50)^2 = 1190,25$$

Neighbour Joining-stamtræ

Afstanden mellem menneske og chimpansé udgøres af grenlængderne:
 $311,33 + 357,13 + 669,38 = 1337,38$

Afstanden mellem gorilla og orangutang udgøres af grenlængderne:
 $1498,63 + 835,38 = 2334,01$

Matrixværdierne for de to afstande er henholdsvis 1334 basepar og 2334 basepar. Udregning for menneske-chimpansé bliver derfor:

$$(1334 - 1337,38)^2 = 11,42$$

For gorilla-orangutang:

$$(2334 - 2334,01)^2 = 0,00$$

Det ses af de to stamtræer figur 3 og figur 4 at hvis der kun er én knude, der forbinder 2 arter, kan man sagtens lave grenlængder der passer. I stamtræet i figur 3 (UPGMA) er der én knude

mellem chimpanse og bonobo. Der er en afstand i afstandsmatrixen på 602 basepar, og den afstand afbilledes også med to grenlængder på hver 301 basepar. Alle andre arter i stamtræet i figur 3 har grenlængder de deler med andre arter, og derfor må man vha. gennemsnit finde de mindst mulige forskelle mellem data i matrixen og grenlængderne i stamtræet.

I stamtræet uden rod (figur 4) er der to slægtskaber hvor to arter kun deler én knude, nemlig mellem chimpanse og bonobo og mellem gorilla og orangutang. I stamtræet uden rod er det derfor muligt at tegne grenlængder der passer med de observerede i afstandsmatrixen for netop disse to slægtskaber. I stamtræet med rod er det alene chimpansen og bonobo der er forbundet med én knude, og derfor er det kun mellem disse to at der ikke er forskel på det observerede i afstandsmatrixen og det tegnede i stamtræet. Man kan sige at jo flere knuder der er mellem taxa i stamtræet, desto større er forskellen mellem det observerede i afstandsmatrixen og det tegnede i stamtræerne. Man mister simpelthen frihedsgrader for hver knude der er mellem to taxa i et stamtræ.

Udregner man den samlede kvadrerede forskel mellem matrixen og de to stamtræer, fås at den kvadrerede forskel for UPGMA med rod er 4175,50 og for Neighbour Joining, uden rod, er 2084,45. Udregningerne fremgår af figur 5. Læg mærke til at ingen af de to stamtræer i hhv. figur 3 og 4 viser det observerede antal af forskelle som de fremgår af afstandsmatrixen figur 2.

Grenlængder			Matrix	Kvadreret forskel	
	Med rod	Uden rod		Med rod	Uden rod
Bonobo-Chimpanse	602,00	602,00	602	0,00	0,00
Chimpanse-Menneske	1327,50	1337,84	1334	42,25	14,75
Bonobo-Menneske	1327,50	1317,18	1321	42,25	14,59
Chimpanse-Gorilla	1612,00	1624,97	1608	16,00	287,98
Bonobo-Gorilla	1612,00	1604,31	1587	625,00	299,64
Menneske-Gorilla	1612,00	1625,89	1645	1089,00	365,19
Chimpanse-Orangutang	2299,50	2288,22	2309	90,25	431,81
Bonobo-Orangutang	2299,50	2267,56	2285	210,25	304,15
Menneske-Orangutang	2299,50	2289,14	2270	870,25	366,34
Gorilla-Orangutang	2299,50	2334,01	2334	1190,25	0,00
			SUM	4175,50	2084,45

Figur 5. Tabel der viser antallet af forskelle mellem de observerede sekvenser i to forskellige stamtræer og i en afstandsmatrix. Desuden udregnes den kvadrerede forskel mellem matrixen og stamtræerne.