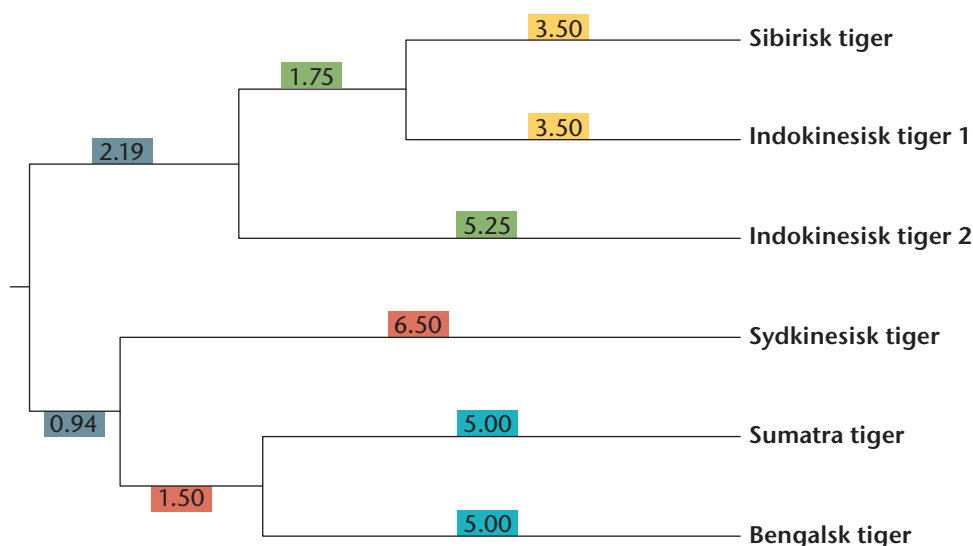


UPGMA-metoden

– eksempel truede tigre

UPGMA står for ‘Unweighted Pair Group Method with Arithmetic mean’ og er en metode til at udregne grenlængder i stamtræer ved en gennemsnitsberegning. Man kan læse teorien om metoden i fx Frank Grønlund Jørgensen (2012): Bioteknologi 6, side 31-34 fra Nucleus forlag. At gennemsnitsberegningen er ‘unweighted’ (uvægtet) betyder at alle værdier i afstandsmatrixen tæller lige meget. Når et sekvensanalyseprogram tegner stamtræer, kan UPGMA bruges som metode til at bestemme grenlængderne. Løsningen på opgaven om de truede tigre fra det vejledende eksamenssæt i biologi A fra 2006 udgivet af opgavekommissionen i biologi kan som vist i figur 1 tegnes som et fylogram med grenlængder målt i antal basesubstitutioner.



Figur 1. Stamtræ over tigre fra det vejledende opgavesæt 2006. Farverne på antallet af basesubstitutionerne vist over grenene er lavet, så det er nemmere at referere tilbage til grenlængderne i teksten.

Fremgangsmåden for at lave et UPGMA-stamtræ er:

- 1) Find det mindste antal forskelle i afstandsmatrixen – her vist i figur 2, men hvilken som helst afstandsmatrix kan anvendes.
- 2) Brug det tal til at tegne to-grene-stamtræet der forener de to taxa, som hver har den halve længde af tallet i afstandsmatrixen.
- 3) Herefter gøres de to taxa som er tegnet, til en gruppe.
- 4) Afstanden fra de andre taxa til denne gruppe udregnes som gennemsnittet af de andre arters afstand i forhold til de to arter der indgår i gruppen.
- 5) Herefter fremstilles med de beregnede tal en ny afstandsmatrix der på hvert led er en celle kortere, og man starter fra pkt. 1 igen.

Første afstandsmatrix der tegnes, er en optælling af forskelle i baser i de enkelte sekvenser. Det er svaret på ét af de spørgsmål, der er i opgaven. Se tabellen i figur 2.

	1	2	3	4	5
1 Sibirsk					
2 Sydkina	17				
3 Indokinesisk 1	7	18			
4 i Indokinesisk 2	9	16	12		
5 Sumatra	14	11	15	13	
6 Bengalsk	13	15	15	13	10

Figur 2. Afstandsmatrix som vist i opgaven. Numrene i øverste række referer til samme art/underart af tiger som er vist i venstre kolonne.

Det laveste tal i afstandsmatrixen figur 2 findes (7) som i dette eksempel er forskellen mellem indokinesisk tiger 1 og sibirsk tiger. Herved kan de 2 tigre tegnes i et stamtræ hvor grenene der forbinder dem, samlet er 7 baseforskelle lange. Derfor er grenene fra deres fælles stamform ud til hver tigerart $\frac{7}{2}$ baseforskelle lange, svarende til 3,5 baseforskelle; markeret med gult i stamtræet i figur 1. Sibirsk tiger og indokinesisk tiger gøres til en fællesgruppe (sibirsk-indo1) som de øvrige tigre i afstandsmatrixen skal sammenlignes med, og det gøres ud fra gennemsnit. Når gruppe 1 og 3 lægges sammen til én gruppe, udregnes den øvrige forskel ved at tage gennemsnittet i de farvede felter i tabel i figur 3. Værdien 7 udgår i den sammenhæng af afstandsmatrixen, fordi den er brugt.

	1	2	3	4	5
1 Sibirsk					
2 Sydkina	17				
3 Indokinesisk 1	7	18			
4 i Indokinesisk 2	9	16	12		
5 Sumatra	14	11	15	13	
6 Bengalsk	13	15	15	13	10

Figur 3. Afstandsmatrix, hvor de tal der skal udregnes et gennemsnit af, er farvet med samme farve.

Udregningen giver et resultat som ses i afstandsmatrix i figur 4.

	1	2	3	4
1 Sibirsk-indo1				
2 Sydkina	17,5			
3 Indokinesisk 2	10,5	16		
4 Sumatra	14,5	11	13	
5 Bengalsk	14,0	15	13	10

Figur 4. Afstandsmatrix udregnet på baggrund af figur 3, hvor sibirsk og indokinesisk tiger 1 er grupperede.

Det laveste tal i afstandsmatrixen i figur 4 findes (10) og det er det antal basesubstitutioner der forbinder den bengalske tiger med sumatratigeren. Den afstand er 10, og de to tigre forbindes med to grene der hver er 5 baseforskelle lange (de blå tal i stamtræet i figur 1). Dernæst skal sumatratigeren og den bengalske tiger grupperes, og gennemsnittet skal udregnes mellem de 2 grupper. De gule felter i figur 5 består af to værdier; nemlig 14,5 som er et gennemsnit af 14 og 15, og af 14 som er et gennemsnit af 13 og 15. Den samlede værdi der forener de to gule felter, bliver derfor gennemsnittet af tallene 14;15;13;15. Det er derfor at man taler om at gennemsnittet er uvægtet. Alle grenlængder der indgår, tæller lige meget i udregningen af det samlede gennemsnit.

	1	2	3	4
1 Sibirsk-indo1				
2 Sydkina	17,5			
3 Indokinesisk 2	10,5	16		
4 Sumatra	14,5	11	13	
5 Bengalsk	14,0	15	13	10

Figur 5. Afstandsmatrix der viser hvordan gennemsnittet skal udregnes, når man sætter suamatratigeren og den bengalske tiger i samme gruppe.

Efter at have grupperet sumatratigeren og den bengalske tiger får man en afstandsmatrix som vist i tabellen i figur 6. Værdien 10 udgår denne gang af afstandsmatrixen, fordi den er brugt.

	1	2	3
1 Sibirsk-indo1			
2 Sydkina	17,50		
3 Indokinesisk 2	10,50	16	
4 Sumatra-bengalsk	14,25	13	13

Figur 6. Sammenlagt afstandsmatrix udarbejdet fra data i figur 5. Mange af disse tal er nu gennemsnitstal.

For at bevare overblikket kan det være godt at se hvordan de enkelte tal i tabellen er udregnet. Alle tal i figur 6 på nær 16 er nemlig gennemsnit. Med den notation der anvendes i Excel, er der i tabellen i figur 7 angivet hvilke gennemsnit der er tale om i de enkelte celler.

	1	2	3
Sibirsk-indo1			
Sydkina	=MIDDEL(17;18)		
Indokinesisk 2	=MIDDEL(9;12)	16	
Sumatra-bengalsk	=MIDDEL(15;14;13;15)	=MIDDEL(11;15)	=MIDDEL(13;13)

Figur 7. Data fra figur 6 vist med de tal der indgår i de viste gennemsnit i figur 7. Notationen der anvendes, er den samme som anvendes i Excel til udregning af gennemsnit.

Det ses i figur 6 at det laveste tal er 10,5, og at det forbinder den indokinesiske tiger 2 med gruppen af sibirsk tiger og indokinesisk 1 tiger. Det vil sige at hver af de to linjestykker skal være 5,25 lange (med grønt i stamtræet i figur 1). Sibirsk tiger og indokinesisk tiger 1 allerede forbundet af to linjestykker der hver er 3,5 lange. Derfor skal der fra gruppen af indokinesisk tiger 1 og sibirsk tiger tegnes et linjestykke, der er $5,25 - 3,5 = 1,75$ langt, og det skal forbindes med et linjestykke der er 5,25 langt hen til indokinesisk tiger 2. Så får man en gruppe der består af indokinesisk tiger 1 og 2 og af sibirsk tiger. Gennemsnittet af disse sammenlægninger udregnes ved at lave en gennemsnitsberegning af den indokinesiske tiger 2 og henholdsvis den sydkinesiske tiger og gruppen Sumatra-bengalsk. Se de grønne og gule felter i tabellen i figur 8.

	1	2	3
1 Sibirsk-indo1			
2 Sydkina	17,50		
3 Indokinesisk 2	10,50	16	
4 Sumatra-bengalsk	14,25	13	13

Figur 8. Afstandsmatrix der viser hvordan gennemsnittet skal udregnes, når man sætter indokinesisk tiger 2 i samme gruppe som sibirsk og indokinesisk tiger 1.

Ved at se på tabellen i figur 8 kan man se at sammenlægningen af de gule tal giver en værdi skrevet i Excel-notation på =MIDDEL(17;18;16), og de grønne tal bliver =MIDDEL(15;14;13;15;13;13) hvilket svarer til henholdsvis 17 og 13,833. I figur 9 ses det opsat i tabelform i det laveste tal (10,5) udgår til resultatet der kan ses i tabel 8.

	1	2	3
1 Sibirsk-indo1-2			
2 Sydkina	17		
3 Sumatra-bengalsk	13,8333	13	

Figur 9. Afstandsmatrix bestående af gruppen sibirsk og indokinesisk tiger 1 og 2. Den sydkinesiske tiger er for sig selv og der er en gruppe af Sumatratiger og bengalsk tiger.

Af tabellen i figur 9 kan man se at den laveste værdi er 13, og at det giver 2 linjestykker på 6,5 der kan forbinde den sydkinesiske tiger med gruppen af sumatra- og bengalsk tiger (tal med rødt på stamtræet i figur 1). Sumatra- og bengalsk tiger er allerede forbundet med linjestykker der er 5 lange. Derfor skal der tegnes et linjestykke der er $6,5 - 5 = 1,5$ mod venstre i stamtræet, og dette knudepunkt skal forbindes med et linjestykke på 6,5 som forbinder den sydkinesiske tiger med sumatratigeren og den bengalske tiger. Herved udgår 13 af afstandsmatrixen og man har to grupper af tigre. Den ene består af sibirsk tiger og de to indokinesiske tigere, og den anden gruppe består af den sydkinesiske tiger og sumatra- og bengalsk tiger. Det vil sige at for at forbinde de 2 grupper skal der tegnes en forskel som er et samlet gennemsnit af {15;14;13;15;13;13;17;18;16}. Det giver i alt 14,889 og dermed 2 linjestykker med en længde på 7,44 (med grå markering i stamtræet i figur 1). For gruppen sydkinesisk, bengalsk og sumatratiger er der allerede tegnet linjestykker der er 6,5 lange. De skal derfor yderligere forlænges med $7,44 - 6,5 = 0,94$. For gruppen sibirsk og de 2 indokinesiske tigre er der allerede tegnet linjestykker der er 5,25 lange. De skal derfor yderligere forlænges med $7,44 - 5,25 = 2,19$. Herefter får begge grupper et knudepunkt der forener hele stamtræet.

Kilde

Opgavekommissionen i biologi: 'Opgave 2. Truede tigre' fra *Biologi A (stx)*
 – *Vejledende opgavesæt 2 – maj 2006*, Undervisningsministeriet, 2006.