

Hausaufgabe 3

Aufgabe 3.1 Evolutionärer Distanzen

Teilaufgabe 3.1.1 Berechnung evolutionärer Distanzen (20%)

Die berechneten Distanzen finden sich in der Datei [Distanzen.txt](#)

Zwei der Sequenzen (1 und 4) sind identisch, Sequenz 3 ist von diesen beiden vollständig verschieden, weswegen eine Verwendung der Poisson- und Jukes-Cantor-Distanzen keine sinnvollen Werte ergibt.

Sequenz 2 und 3 gleichen sich nur in einem Nukleotid, was in der Poisson-Distanz einen deutlich größeren Wert produziert.

Teilaufgabe 3.1.2 Prozentuale Übereinstimmung (10%)

Es sei $I' = \frac{M}{L} \cdot 100\%$, wobei $M \hat{=}$ Übereinstimmung zwischen zwei Sequenzen
 $L \hat{=}$ Länge der Sequenzen

Dann liegt eine Binomial-Verteilung vor, wobei gilt: $1 = I + p$ ($p = \frac{D}{L}$, $D \hat{=}$ Unterschiede und $I = \frac{I'}{100\%}$)

Ich nehme an, dass in der Aufgabe entweder $I' = \frac{I}{100}$ oder $d_f = -\ln(I)$ gemeint ist

Somit ist $d = -\ln(I) = -\ln(1-p) = d_p$ q.e.d.

Aufgabe 3.2 Konstruktion phylogenetischer Bäume (40%) 3.2.1.

Sie haben durch den Vergleich der DNA-Sequenzen die folgende Distanzmatrix erhalten

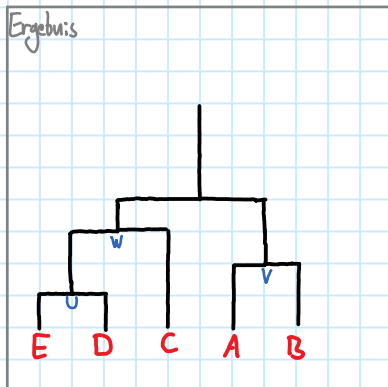
	A	B	C	D	E
A	0	5	9	9	8
B	5	0	10	10	9
C	9	10	0	8	7
D	9	10	8	0	3
E	8	9	7	3	0

(3)

Eine Distanzmatrix muss

- symmetrisch sein. Dies ist hier der Fall ✓
- identitär sein. Dies ist hier der Fall, denn die Werte der Hauptdiagonalen sind alle 0. ✓

3.2.2.



Schritt 1: E & D vereinigen

	A	B	C	U
A	0	5	9	$\frac{17}{2}$
B	5	0	10	$\frac{19}{2}$
C	9	10	0	$\frac{15}{2}$
U	$\frac{17}{2}$	$\frac{19}{2}$	$\frac{15}{2}$	0

Schritt 2: A & B vereinigen

	C	V	U
C	0	$\frac{19}{2}$	$\frac{15}{2}$
V	$\frac{19}{2}$	0	9
U	$\frac{15}{2}$	9	0

$$d_{WU} = \frac{1}{4}(9+8+10+9)$$

Schritt 3: C & U vereinigen

	W	U
W	0	$\frac{17}{2}$
U	$\frac{17}{2}$	0

$$d_{WU} = \frac{1}{6}(9+8+10+9+8+7) = \frac{17}{2}$$

→ W & U vereinigen

Aufgabe 3.3 Konstruktion phylogenetischer Bäume mit R (30%)

Nutzen Sie die R library PHANGORN um phylogenetische Bäume zu erstellen. Führen Sie dazu die folgenden Befehle aus um einen Teil der DNA Sequenz von vierzehn Primaten zu laden.

Code und Antworten in [Übung 3.R](#) zu finden.