

Übung 3

3.1.1)

a) $p = D / L$

D: Anzahl d. Pos. an denen
sich die Sequenzen unterscheiden
L: Länge der Sequenzen

Seq 1 G A T A G
Seq 2 G A T A A
Seq 3 A T G G A
Seq 4 G A T A G

	Seq 1	Seq 2	Seq 3	Seq 4
Seq 1	0	0,2	1	0
Seq 2	0,2	0	0,8	0,2
Seq 3	1	0,8	0	1
Seq 4	0	0,2	1	0

	Seq 1	Seq 2	Seq 3	Seq 4
Seq 1	0	0,223	1	0
Seq 2	0,223	0	1,609	0,223
Seq 3	1	1,609	0	1
Seq 4	0	0,223	1	0

	Seq 1	Seq 2	Seq 3	Seq 4
Seq 1	0	0,233	/	0
Seq 2	0,233	0	/	0,233
Seq 3	/	/	0	/
Seq 4	0	0,233	/	0

b) $d_p = -\ln(1-p)$

c) $d_x = -\frac{3}{4} \ln(1 - \frac{4}{3}p)$

→ p-Distanz unterschätzt evolutionäre Distanz

→ p-Distanz mit Poisson-Korrektor berücksichtigt wiederholte Mutationen an der selben Position → führt bei größeren p zu größeren Distanzen

→ Jukes & Cantor nicht gültig für hohe p (ab 0,8)

3.1.2) $d_p = -\ln(I')$ bzw. $d_p = -\ln(1-p)$

I = Prozentuale Übereinstimmung

I' = Übereinstimmung ($I/100$) (nicht wie in Aufgabenstellung $I' = 100 * I$)

$$p = \frac{D}{L} \rightarrow p = \frac{L-A}{L} \rightarrow p = 1 - \frac{A}{L}$$

A: Anzahl d. Pos. an denen die Sequenzen gleich sind

$$L \rightarrow I' = \frac{A}{L} \quad (I = \frac{A}{L} \cdot 100)$$

$$d_p = -\ln(1-p) \rightarrow = -\ln(1 - (1 - \frac{A}{L}))$$

$$= -\ln(1 - 1 + \frac{A}{L})$$

$$= -\ln(\frac{A}{L}) = \underline{\underline{-\ln(I')}}$$

3.2)

3.2.1)

	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 ✓

Identität des Ununterscheidbaren: $d(\text{Sequenz 1, Sequenz 1}) = 0$

↳ Trifft zu ✓

Symmetrie: $d(\text{Sequenz 1, Sequenz 2}) = d(\text{Sequenz 2, Sequenz 1})$

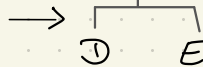
↳ Trifft zu ✓

⇒ valide Distanzmatrix

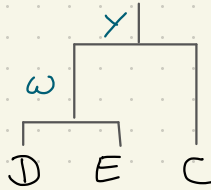
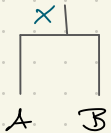
$$3.2.2) d_{\text{alt, neu}} = \frac{d_{\text{alt, Cluster 1 in neu}} + d_{\text{alt, Cluster 2 in neu}}}{\text{Elemente}_{\text{alt}} \cdot \text{Elemente}_{\text{neu}}}$$



	A	B	C	W
A	0	5	9	8,5
B		0	10	9,5
C			0	7,5



	C	W	X
C	0	7,5	9,5
W		0	9



←

	Y	X
Y	0	8,25

→

