

Ihre Namen: _____ Gruppe: _____

Evolutionsbiologie 2, WS2016/2017: Bioinformatik – Übung 2-1

Webseite zur Vorlesung/Übung:

<http://evol.bio.lmu.de/teaching/bachelor2/>

Teil 2-1-1: Alignments

1.1 Gegeben sind folgende Sequenzen:

Seq1 ATGCAT
Seq2 TTGCAC
Seq3 AGGCAT
Seq4 ATTGGA

(Q1) Berechnen Sie alle paarweisen Hamming-Distanzen.

Distanz

Seq1 – Seq2 _____

Seq1 – Seq3 _____

Seq1 – Seq4 _____

Seq2 – Seq3 _____

Seq2 – Seq4 _____

Seq3 – Seq4 _____

(Q2) Welches Sequenzpaar hat die größte Ähnlichkeit?

1.2 Gegeben sind folgende Sequenzen:

Seq1 AATCGA
Seq2 AAACCGA
Seq3 AATCGT
Seq4 ATCG

(Q3) Berechnen Sie alle paarweisen Levenshtein-Distanzen und geben Sie jeweils eine mögliche Sequenz an Änderungen an.

Distanz

Sequenz an Änderungen

Seq1 – Seq2 _____

Seq1 – Seq3 _____

Seq1 – Seq4 _____

Seq2 – Seq3 _____

Seq2 – Seq4 _____

Seq3 – Seq4 _____

(Q4) Welches Sequenzpaar hat die größte Ähnlichkeit?

1.3 Vervollständigen Sie den Dotplot für die Sequenz GAATTCCTTAAG gegen sich selbst.

(Q5) Was fällt Ihnen auf? Wie nennt man solche Sequenzabschnitte?

	G	A	A	T	T	C	C	T	T	A	A	G
G	X											X
A												
A												
T												
T												
C						X	X					
C												
T												
T												
A												
A												
G												

1.4 Betrachten Sie die beiden folgenden Alignments:

A A A T C G - G - A T C T - T A G
 A - - C G C G C A T - T A T A G

B A A T C G G A T C T T A G
 A C G C G C A T T A T A G

Verwenden Sie als Scoring-Schema: Match = 1, Mismatch = -1, Gap = 0

(Q6) Bestimmen Sie den Score für jedes Alignment. Welches Alignment hat den höheren Score?

Verwenden Sie nun für die Gap-Kosten -5.

(Q7) Bestimmen Sie den Score für jedes Alignment. Welches Alignment hat den höheren Score?

1.5 Gegeben ist folgende Substitutionsmatrix.

	A	T	C	G
A	0	5	5	1
T	5	0	1	5
C	5	1	0	5
G	1	5	5	0

(Q8) Handelt es sich bei dieser Substitutionsmatrix um eine Distanzmatrix oder um eine Ähnlichkeitsmatrix?

(Q9) Berechnen Sie mit Hilfe dieser Matrix den Score für folgendes Alignment:

```
A A T C G G C G C C
A G C A G A T A A C
```

(Q10) Wie viele Transitionen treten im obigen Alignment auf?

(Q11) Wie viele Transversionen treten im obigen Alignment auf?

(Q12) Die gegebene Substitutionsmatrix berücksichtigt die Häufigkeitsverteilung von Transitionen und Transversionen. Woran können Sie das sehen?

1.6 Verwenden Sie den Needleman-Wunsch-Algorithmus aus der Vorlesung zur Berechnung eines Alignments zwischen den Protein-Sequenzen MAETHE und AEHREN (Match = 0, Mismatch = -3, Gap = -5). Vergessen Sie nicht die Matrix zu initialisieren!

	ε	M	A	E	T	H	E
ε							
A							
E							
H							
R							
E							
N							

(Q13) Geben Sie ein optimales Alignment zu diesen beiden Sequenzen an. Gibt es mehr als eines? Warum bzw. warum nicht?

1.7 Gegeben ist folgendes multiple Sequenzalignment. **Hinweis: Bei der Lösung dieser Aufgabe sind die Vorlesungsfolien hilfreich.**

```
A A T C G A T G T G
A G T C A A T A C G
A A A C C C T G C G
A G C C T A T G T G
```

(Q14) Geben Sie die Konsensussequenz an.

(Q15) Vervollständigen Sie das Sequenzlogo zu diesem Alignment.

	A								

(Q16) Welche Position hat die geringste Konservierung?

Teil 2-1-2: Phylogenetische Analysen

In diesem Teil der heutigen Übung werden Sie zunächst einige Aufgaben auf dem Papier bearbeiten um die Inhalte zu trainieren, die Sie in der Vorlesung gelernt haben. Im Anschluss führen Sie eine kurze phylogenetische Analyse mit allen dafür nötigen Schritten mit internet-basierten Programmen durch.

Aufgaben zur Vorlesung

(Q17) 1.8 Zeichnen Sie einen möglichen Baum zur folgenden Newick-Darstellung:

((A, B), (C, D)), (E, F), G)

(Q18) Ist dies der einzig mögliche Baum für diese Darstellung? Warum bzw. warum nicht?

(Q19) Geben Sie eine alternative Newick-Darstellung an.

1.9 Gegeben sei die folgende Distanzmatrix für die Sequenzen A, B, C und D.

(Q20) Nutzen Sie die UPGMA-Methode und zeichnen Sie eine Baumtopologie. Berechnen Sie die Astlängen nach der Mittelpunkt-Methode (siehe Beispiel in der Vorlesung).

	A	B	C	D
A	-	0.30	0.32	0.40
B		-	0.08	0.20
C			-	0.22
D				-

(Q21) 1.10 Wie viele ungewurzelte und wie viele gewurzelte Bäume gibt es für 4 Sequenzen A, B, C und D?

1.4 Betrachten Sie folgendes Alignment:

A AATCGA
 B AACCGA
 C ATCCGA
 D CTGGAA

(Q22) Geben Sie je ein Beispiel für eine Symplesiomorphie, eine Autapomorphie und eine Synapomorphie.

(Q23) Geben Sie alle möglichen ungewurzelten (Bäume) Topologien für die Sequenzen aus (Q22) an und zeichnen Sie die Änderungen ein, die nötig sind, um jede Topologie zu realisieren. Welche Position ist parsimonie-informativ?

(Q24) Für welchen der in (Q23) erstellten Bäume würden Sie sich nach der Maximum-Parsimonie-Methode entscheiden? Ist die Möglichkeit eindeutig? Warum?

	Positions						
Baum	1	2	3	4	5	6	Summe