

Sara Kemmler 5760949
Robin Bonkaß 5769588

1	2	3	4	5	Σ

Übungsblatt Nr. 02

(Abgabetermin 12.05.22)

Aufgabe 1

Globales Alignment:

	0	G	A	G	A	T
0	0	-3	-6	-9	-12	-15
T	-3	-2	-5	-8	-11	-10
G	-6	-1	-4	-3	-6	-9
A	-9	-4	1	-2	-1	-4
T	-12	-7	-2	-1	-4	1
T	-15	-10	-5	-4	-3	-2
C	-18	-13	-8	-7	-6	-5
A	-21	-16	-11	-10	-5	-8
T	-24	-19	-14	-13	-8	-3

Mögliches globales Alignment von obigem Traceback:

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

Lokales Alignment:

	0	G	A	G	A	T
0	0	0	0	0	0	0
T	0	0	0	0	0	2
G	0	2	0	2	0	0
A	0	0	4	1	4	2
T	0	0	1	2	1	6
T	0	0	0	0	0	3
C	0	0	0	0	0	1
A	0	0	2	0	2	0
T	0	0	0	0	0	4

Mögliches lokales Alignment mit obigem Traceback:

G	A	T
G	A	T

Aufgabe 2

a)

$$E = mn \cdot 2^{-S'}$$

Formel nach S' umstellen:

$$\begin{aligned} \Leftrightarrow \frac{E}{mn} &= 2^{-S'} \\ \Leftrightarrow \log_2\left(\frac{E}{mn}\right) &= -S' \\ \Leftrightarrow -\log_2\left(\frac{E}{mn}\right) &= S' \end{aligned}$$

Gegebenen Werte in S' einsetzen:

$$\begin{aligned} S' &= \log_2 \frac{0.02}{175 \cdot 8 \cdot 10^8} \\ &= 42,6705 \end{aligned}$$

b)

i)

Der E-Wert berechnet sich wie folgt:

$$E = Knm \cdot e^{-\lambda S}$$

Sei nun die query Länge $n = 2n$ verdoppelt.

$$E = K(2 \cdot n)m \cdot e^{-\lambda S} = 2 \cdot (Knm \cdot e^{-\lambda S})$$

Daraus folgt, dass sich der E-Wert verdoppelt, wenn sich die Länge der query sequence verdoppelt.

ii)

Der E-Wert berechnet sich wie folgt:

$$E = Knm \cdot e^{-\lambda S}$$

Sei nun die query Länge $n = \frac{n}{2}$ halbiert.

$$E = K\left(\frac{n}{2}\right)m \cdot e^{-\lambda S} = \frac{1}{2} \cdot (Knm \cdot e^{-\lambda S})$$

Daraus folgt, dass sich der E-Wert halbiert, wenn sich die Länge der query sequence halbiert.

iii)

Der E-Wert berechnet sich wie folgt:

$$E = Knm \cdot e^{-\lambda S}$$

Hier ist S der score des minimum HSP. Dieser wird durch die scoring Matrix berechnet. Er hängt also direkt von der scoring Matrix ab und ändert sich auch, wenn die Matrix sich ändert. Daraus folgt, dass sich auch der E-Wert bei Änderung der scoring Matrix verändert.

Aufgabe 3

Da die gegebene Sequenz in der Datei `unknown_protein.fasta` eine Aminosäuresequenz handelt und somit ein Protein darstellt, nutzen wir Protein BLAST, kurz BLASTp.

Das Protein welches als einziges eine 100 prozentige Übereinstimmung mit der gegebenen Sequenz hat, ist das Lactase-Protein der Hausmaus (*mus musculus*). Es wird vorausgesagt, dass die Aktivitäten von Lactase und Transferase ermöglicht werden können. Das Lactase-Protein ist zudem in verschiedenen Prozessen beteiligt, unter anderem bei der Reaktion von Eisen-(II)-Ionen, Blei-Ionen und Nickel-Kationen.

Vermutet wird, dass das Lactase-Protein der Hausmaus in der sogenannten 'brush border'. Dabei handelt es sich um eine durch Mikrovilli geschützte Membran von Epithelien, die in vielen Bereichen des Körpers zu finden sind.

Das menschliche Ortholog zu diesem Protein ist beteiligt an angeborenem Laktase-Mangel. Dieses Ortholog wird bei uns Menschen als LCT (Laktase) bezeichnet.

Der E-Wert beträgt in unserem Fall 0.00. Dies bedeutet, dass es keine anderen möglichen Alignments gibt, die einen genau so guten bzw. höheren Score als das von BLAST berechnete Alignment, erreichen würden.

Die 'percentage identity' liegt bei 100%, womit man auf eine 100 prozentige Übereinstimmung der beiden Sequenzen schließen kann. Daraus resultiert, dass es keine gaps und keine mismatches gibt.

Eine 'query cover' von in unserem Fall 100% bedeutet, dass die beiden Sequenzen vollständig miteinander aligniert werden konnten. Somit erstreckt sich das lokale Alignment über die ganze Länge der query-Sequenz.

Fast alle der ersten 15 Treffer sind Laktase-Proteine von Mäusen oder Ratten, welche zur Unterfamilie der Murinae (Altweltmäuse) gehören. Daraus kann geschlossen werden, dass unsere obigen Resultate bestätigt werden können.

Aufgabe 4

Der Code, der sich im file: `Sara_Kemmler_Robin_Bonkass_A2.py` findet, kann mit folgendem Befehl ausgeführt werden:

```
python3 Sara_Kemmler_Robin_Bonkass_A2.py -f1 yersenia_1.fasta -f2 yersenia_2.fasta
```

Beim Ausführen der Datei wird `aligned_sequences.txt` erstellt.

Aufgabe 5

Der Code, der sich im file: `Sara_Kemmler_Robin_Bonkass_A2.py` findet, kann mit folgendem Befehl ausgeführt werden:

```
python3 Sara_Kemmler_Robin_Bonkass_A2.py -f1 yersenia_1.fasta -f2 yersenia_2.fasta
```

Die Ausgabe der auf dem Übungsblatt aufgeführten outputs finden sich unten in der generierten Datei: `aligned_sequences.txt`