



Einführung in die Bioinformatik

Prof. Dr. Enno Ohlebusch, Dr. Karlheinz Holzmann,
Tobias Badura

WS 15/16

Übungsblatt 2

Abgabe bis 02.12.2014, 10:00. Lösungen bitte elektronisch (an tobias.badura@uni-ulm.de) mit Namen in Text und Quelltextdateien abgeben.

1. Aufgabe (7): Dotplot

Schreiben Sie eine Funktion die ein Dotplot generiert (in R). Eingabe sind zunächst zwei Nukleotid-Sequenzen. Erweitern Sie Ihre Funktion um die Filterung nach der k-Wortmethode. Testen Sie Ihre Funktion mit den Sequenzen S_1 und S_2 . Erzeugen Sie zusätzlich Dotplots für $k = 2$ und $k = 3$.

- > Sequenz S_1
AACTTTGGAGAAT
- > Sequenz S_2
AACTTGCGAGAAC

3. Aufgabe (10)

Ein globales Sequenzalignment der Sequenzen S_1 und S_2 , soll mit dem Needleman & Wunsch Algorithmus berechnet werden. Schreiben Sie dazu ein Programm in R das die alignierten Sequenzen ausgibt. Verwenden Sie die Einheitskostenmatrix aus der Vorlesung.

```
> Sequenz  $S_1$ 
TCAGTTGCCAAACCCGCT

> Sequenz  $S_2$ 
AGGGTTGACATCCGTTTT
```

4. Aufgabe (5): Homologie

Vergleichen Sie die Editierdistanzen des Cytoglobin von Mensch und Maus, sowie die Editierdistanz der Gene mit ihrem vorher implementierten Algorithmus. Erweitern Sie ggf. die Einheitskostenmatrix für Proteine.

- Wie kann es zu solchen Unterschieden kommen?
- Wie nennt man die Beziehung solcher Gene?
- Wie hat sich Myoglobin entwickelt?
- Proteine werden normalerweise nicht mit der Einheitskostenmatrix verglichen. Warum nicht und welche Matrizen kennen Sie um Proteine zu vergleichen?

Hinweis: Verwenden Sie die Datenbank von UniProt zur Suche.