

Softwarewerkzeuge der Bioinformatik

Prof. Dr. Volkhard Helms
PD Dr. Michael Hutter, Markus Hollander,
Marie Detzler
Wintersemester 2020/2021

Universität des Saarlandes
Zentrum für Bioinformatik

Übungsblatt 2

Sequenzanalyse: Paarweise Alignments

Lernziel: Sie sollen lernen, wann Sie welche BLAST-Suche (Protein-BLAST, Nucleotide-BLAST, MEGA-BLAST, PSI-BLAST) verwenden sollten, und welche Parameter (E-value, Matrix, query database etc.), abhängig von der Suche, sinnvoll sind. Zusätzlich sollen Sie die theoretischen Fragen beantworten und ein paarweises Alignment mit Needleman-Wunsch berechnen.

Aufgabe 2.1: Dynamisches Alignment

Führen Sie ein globales Alignment nach Needleman-Wunsch der Sequenzen **ACDEF AFGHI** und **KDELA FG** durch.

		A	C	D	E	F	A	F	G	H	I
K											
D											
E											
L											
A											
F											
G											

Globales Alignment:

Aufgabe 2.2: ProteinBLAST

Die Vorlesungsfolien könnten für die Beantwortung der folgenden Fragen hilfreich sein.

- a) Wie ist die erwartete Schranke (*expected threshold*, *E-Value*) definiert? Warum ist eine Schranke von 10 für den E-Value nicht unbedingt sinnvoll? Was sind sinnvolle E-Values?
- b) Welchen Einfluss hat die Wortlänge (*word size*) auf die Laufzeit und die Genauigkeit?
- c) Welche Besonderheit hat der erste Treffer eine BLAST Suche gegen eine übliche Datenbank?
- d) Führen Sie eine **Protein-Protein BLAST** Suche (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) mit dem Protein **P00042** durch:
 - i. gegen die **UniProtKB/Swiss-Prot** Datenbank unter Verwendung von *default* Werten. Finden Sie die 10 Proteine mit größter Homologie zu P00042 und lassen Sie sich deren Sequenzen anzeigen. Um was für Proteine handelt es sich?
 - ii. gegen die **non-redundant** Datenbank mit einer E-Value-Schranke von 0.001. Was für Unterschiede gibt es? Für welche Organismengruppen wurden Treffer gefunden?

Aufgabe 2.3: MegaBLAST

Wählen Sie **Human** als Genom auf der BLAST-Startseite. Suchen Sie nach der mRNA **NM_175054** des menschlichen Gens *HIST4H4* mit **megaBLAST** in der Datenbank **Genome (GRCh38.p13)**.

- a) Auf welchem Chromosom liegt *HIST4H4*?
- b) Gibt es ein Paralog?
- c) Finde zwei bis drei direkt benachbarten Gene zu *HIST4H4*.

Aufgabe 2.4: PSI-BLAST

- a) Suchen Sie mit **ProteinBLAST** viele weit entfernte Homologe des Proteins **Q57997** in der **non-redundant** Datenbank mit einem E-Value von 0.02 als oberste Schranke. Welche Substitutionsmatrix ist geeignet?
- b) Führen Sie dieselbe Suche mit **PSI-BLAST** durch, und verwenden Sie eine Schranke von 0.001 für den maximalen E-Value der Sequenzen, die zur Erstellung der PSSM verwendet werden.
- c) Wo sind Unterschiede zwischen dem Ergebnis aus a) und dem der 1. Iteration aus b)?
- d) Wie verändert sich das Ergebnis von Teil b), wenn weitere Iterationen durchgeführt werden?

Viel Spaß!