

Algorithmen und Datenstrukturen in der Bioinformatik

Sechstes Übungsblatt WS 14/15

Abgabe Donnerstag 30.10., 12:00

Niveau I

Aufgabe 1: FastA

Erstellen Sie mit dem **FastA**-Algorithmus ein lokales Alignment für die Sequenzen $S_1 = \text{mississippi}$ und $S_2 = \text{mymissisahippi}$. Führen Sie dabei folgende Schritte aus:

- Erstellen Sie eine *Hash-Tabelle* für S_1 mit $ktup=2$.
- Finden Sie mit Hilfe dieser Hash-Tabelle die *Hot-spots* in S_2 .
- Tragen Sie die Hot-spots in einen *Dot-plot* ein und finden Sie *Diagonal runs*. Hierbei dürfen Lücken der Länge 1 (entspricht einem Mismatch) zugelassen werden.
- Rescoring*: Berechnen Sie die Scores Ihrer Diagonal runs mit folgenden Kosten: Match = 4, Mismatch = -3

Aufgabe 2: Chaining

- Erzeugen Sie durch *topologische Sortierung* einen DAG aus den gefundenen Hot-spots aus Aufgabe 1.
- Modifizieren Sie diesen Graphen und
- wenden Sie anschließend den *Dijkstra*-Algorithmus an.

Niveau II

Aufgabe 3: Heuristik

- An welchen Stellen des FastA-Algorithmus wird deutlich, dass es sich um eine *Heuristik* handelt? Unter welchen Umständen wird ein optimales Alignment nicht gefunden?
- Wie wirken sich große und kleine k-tup-Werte auf *Sensitivität* und *Spezifität* des Algorithmus aus?