

Algorithmen und Datenstrukturen in der Bioinformatik

Vierte Praktikumsaufgabe WS 12/13

Abgabe Montag, 21.01., 15:00 Uhr

MSA: C-optimale Schnittpositionen

Implementieren Sie den Branch-and-Bound Algorithmus zum Finden eines k -Tupels von C-optimalen Schnittpositionen für Sequenzen $s_1, \dots, s_k$. Schreiben Sie dazu ein Program, das:

- k Sequenzen einliest
- $c_2, \dots, c_k$ gegeben $c_1 = \lceil n_1/2 \rceil$ berechnet und ausgibt (**2 Punkte**)
- die *multiple additional cost* von $c_1, \dots, c_k$ berechnet und ausgibt (**1 Punkt**)

Für jedes Paar von Sequenzen sollen dabei die Matrizen D , D^{rev} und C berechnet werden (**3 Punkte**). Die Sequenzen werden Ihrem Programm als k Kommandozeilenparameter übergeben. Ausgegeben werden sollen der Wert des minimalen *multiple additional cost* und ein k -Tupel $c_1, \dots, c_k$ von Schnittpositionen, an dem dieser Wert angenommen wird.

Beispiel:

```
ktrappe@peking:~$ ./aufgabe6 CT AGT G
0
1,2,1
```

Zusatz: Zusätzlich kann die Iteration über die Schnittposition c_1 parallelisiert werden (**2 extra Punkte**).

Hinweise:

- Verwenden Sie das Levenshtein-Kostenschema.

- **Achtung:** In dieser Aufgabe werden Kosten minimiert (nicht Scores maximiert), der Needleman-Wunsch (aus P3) lässt sich aber mit geringen Änderungen anpassen.
- Eine Schnittposition ist die Länge der Sequenz links vom Schnitt.
- Setzen Sie alle paarweisen Gewichte $w_{i,j}$ auf 1.

Beachten Sie die Praktikumshinweise unter <https://www.mi.fu-berlin.de/w/ABI/A1DaBiWS12>.