
Algorithmische Bioinformatik I

Abgabetermin: Donnerstag den 14. Dezember, vor der Vorlesung

Aufgabe 1 (Programmieraufgabe 16P)

Hinweise MSA Carillo-Lipman

Aufgabenteil 1: Vorberechnungen (8P, bis 15.12.2016)

Implementieren Sie, für eine gegebene Menge von Sequenzen $s_1, \dots, s_k \in \Sigma^*$, Gap-Open/Gap-Extend Kosten $g \in \mathbb{N}$, Mismatch-Kosten $m \in \mathbb{N}$ und eine globale obere Schranke $C \in \mathbb{N}$, die Berechnung der Vorwärts- und Rückwärts-Matrix $P_{xy}, S_{xy} \in \mathbb{N}^{|x| \times |y|}$ sowie der oberen Schranken $C_{xy} \in \mathbb{Z}$ für alle paarweisen Alignments der Sequenzen $x = s_i$ und $y = s_j$, für $1 \leq i < j \leq k$. Verwenden Sie den Needleman-Wunsch-Algorithmus für globale Alignments mit **linearen** Gapkosten.

Als Ausgabe erwarten wir eine Datei *matrices.mats* im Ausgabeordner *output* in der für alle paarweisen Alignments die Indizes der Sequenzen i_j , sowie deren obere Schranke C_{ij} samt der Matrix $\Delta_{xy} = P_{xy} + S_{xy}$ ausgegeben wird. Die Matrix soll Komma-separiert ausgegeben werden. Ein Beispiel für $k = 3$, $g = 3$ und $m = 2$ ist hier angegeben:

0_1,-5	0_2,-2	1_2,-11
,,G,A,T,T,C	_,_,G,G,A,T,T,C	_,_,G,G,A,T,T,C
_,9,9,15,21,27,33	_,12,12,12,18,24,30,36	_,3,3,9,15,21,27,33
A,13,11,9,15,21,27	A,16,12,14,12,18,24,30	G,9,3,3,9,15,21,27
T,17,13,11,9,15,21	T,20,16,12,14,12,18,24	A,15,9,5,3,9,15,21
T,21,17,13,11,9,15	T,22,20,16,12,14,12,18	T,21,15,11,7,3,9,15
C,23,21,17,11,11,9	C,24,22,20,16,12,14,12	T,27,21,17,13,7,3,9
G,27,23,21,15,11,9	G,30,24,22,20,16,12,12	C,33,27,23,19,13,9,3
G,33,27,23,19,15,9	G,36,30,24,22,20,16,12	

Beachten Sie folgende vorgehensweise bei der **Implementierung**:

- Für Sequenzen $\{s_2, \dots, s_k\}$ berechnen Sie die paarweisen Alignments $(s_1, s_2), (s_1, s_3), \dots, (s_1, s_k), (s_2, s_3), \dots, (s_{k-1}, s_k)$. Geben Sie die Alignments in dieser **Reihenfolge** aus!

Beantworten Sie folgende **Fragen**:

1. Was ist die Interpretation eines Wertes in der Matrix Δ_{ij} ?
2. Welche Werte erwarten Sie auf dem Backtrack-Pfad eines optimalen Alignments?

Beachten Sie bitte folgendes **Interface** zum Aufruf des ersten Aufgabenteils:

```
java -jar eurenachnamen.jar 0 < $s_1, \dots, s_k$ > <g> <m> <C> <output-folder>
```

Die Abgabe erfolgt per **eMail** an algo2@bio.ifi.lmu.de als ausführbare *jar*-Datei in der auch die **Source**-Files enthalten sind! Überlegen Sie sich geeignete Testfälle um Ihren Code zu überprüfen!

Aufgabenteil 2: Carillo-Lipman (Bonus 8P, bis 12.01.2017)

Im zweiten Aufgabenteil implementieren Sie das eigentliche Alignment nach Carillo-Lipman.

Geben Sie hier bitte eine Datei *alignments.aln* aus, in der das Alignment für alle Sequenzen angegeben ist. Schreiben Sie in die erste Zeile den Alignment Score. Die nächste Zeile soll enthalten, wie viele Zellen besucht wurden. Darunter sollen die Alignments in Ihrer Eingabereihenfolge ausgegeben werden:

```
20
294
--ATTGGG
-GATTC--
GGATTC--
```

Beachten Sie bitte folgendes **Interface** zum Aufruf des zweiten Aufgabenteils:

```
java -jar eurenachnamen.jar 1 < $s_1, \dots, s_k$ > <g> <m> <C> <output-folder>
```

Die Abgabe erfolgt per **eMail** an algo2@bio.ifi.lmu.de als ausführbare *jar*-Datei in der auch die **Source**-Files enthalten sind! Überlegen Sie sich geeignete Testfälle um Ihren Code zu überprüfen!