Übungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, WS 2024
Prof. Dr. Jens Stoye · Leonard Bohnenkämper

`https://gi.cebitec.uni-bielefeld.de/teaching/2024winter/sa`

Übungsblatt 5 vom 14.11.2024
Abgabe am 21.11.2024 bis 10:00 Uhr (morgens)

Aufgabe 1 (Affine Gapkosten) (6 Punkte)

1. Warum sollten bei affinen Gapkosten die Gap-open-Kosten d nicht niedriger gewählt als die Gap-extension-Kosten e, also $d \geq e$?
2. Zeige, dass affine Gapkosten subadditiv sind, also dass gilt: $g(\ell_1 + \ell_2) \leq g(\ell_1) + g(\ell_2)$
3. Berechne ein optimales globales Alignment mit affinen Gapkosten der Sequenzen $x =$ `CAGT` und $y =$ `CAAGCTGA` effizient mit Hilfe des Gotoh-Algorithmus (berechne die Matrizen S, V und H) und gib dessen Gesamtscore an. Verwende dabei: Score für Match = 3, Score für Mismatch = 0, Kosten für Gap-open $d = 3$, sowie Kosten für Gap-extension $e = 1$.

Aufgabe 2 (Suboptimale Alignments) (6 Punkte)

1. Beschreibe in eigenen Worten, was man unter *überlappenden* lokalen Alignments versteht.
2. Warum ist man bei der Bestimmung von suboptimalen Alignments in erster Linie an *nichtüberlappenden* Alignments interessiert? Welche Probleme möchte man vermeiden?
3. Gegeben seien die Sequenzen $x =$ `ACAGCTTA` und $y =$ `ATTCATGTG` sowie die folgende Score-Funktion: Match = 3, Mismatch = −2, Indel = −2.
   (a) Berechne die Alignment-Matrix für lokales Alignments mit dem Smith-Waterman-Algorithmus und gib ein optimales Alignment an.
   (b) Berechne die Matrix für nichtüberlappende, suboptimale lokale Alignments nach *Waterman-Eggert* und gib das erste suboptimale Alignment an.
   (c) Aktualisiere die Matrix und gib das nächste suboptimale Alignment an.

Aufgabe 3 (Forward-Backward Technik) (8 Punkte)

Gegeben seien die Sequenzen $x =$ `ATGCAATC` und $y =$ `CTCAGAT`. Verwende Einheits**kosten** für globale Alignments. Für die Matrizen kannst du die Vorlage auf der nächsten Seite nutzen.

1. Berechne die Edit-Matrix D und die Edit-Matrix D^{rev} der reversen Strings.
2. Berechne die Gesamtkostenmatrix T und die Zusatzkostenmatrix C.
3. Gib ein optimales globales Alignment von s und t an und markiere es in D, D^{rev}, T und C.
4. Erkläre die Hirschberg-Methode an diesem Beispiel.
   (a) Welche Teile von D, D^{rev} und T werden jeweils berechnet? Markiere es in der jeweiligen Matrix.
   (b) An welchem Punkt in T "schneidet" der Algorithmus? Markiere auch diesen Punkt.
   (c) Welche Alignments werden danach rekursiv berechnet?

D	ϵ	C	T	C	A	G	A	T
ϵ								
A								
T								
G								
C								
A								
A								
T								
C								

D^{r}	C	T	C	A	G	A	T	ϵ
A								
T								
G								
C								
A								
A								
T								
C								
ϵ								

T	ϵ	C	T	C	A	G	A	T
ϵ								
A								
T								
G								
C								
A								
A								
T								
C								

C	ϵ	C	T	C	A	G	A	T
ϵ								
A								
T								
G								
C								
A								
A								
T								
C								