Technische Universität München
Institut für Informatik
Lehrstuhl für Effiziente Algorithmen
Volker Heun

SS 2000
Übungsblatt 7
4. Juli 2000

Algorithmen der Bioinformatik

Abgabetermin: 11. Juli

Aufgabe 1

Sei $s = s_1 \cdots s_n$ eine Zeichenreihe der Länge n. Ein k-mer t von s ist eine Teilwort der Länge k von s, d.h. es existiert ein $i \in [1 : n - k + 1]$ mit $s_i \cdots s_{i+k-1} = t$. Für eine Zeichenreihe s sei $\mathcal{M}_k(s) = \{s_i \cdots s_{i+k-1} \ : \ i \in [1 : n - k + 1\}$ die Multimenge der k-mere von s.

Zeigen Sie, dass es zwei Zeichenreihe s und s' gibt, so dass $\mathcal{M}_4(s) = \mathcal{M}_4(s')$ ist. Dies bedeutet, dass man aus einer Menge der k-mere die ursprüngliche Zeichenreihe nicht mehr eindeutig rekonstruieren kann.

Wie sieht es für beliebiges k aus?

Aufgabe 2

Konstruieren Sie mit Hilfe des in der Vorlesung vorgestellen Approximationsalgorithmus einen Superstring für

$$\begin{aligned} s_1 &= ACGACC, \\ s_2 &= ACCGAC, \\ s_3 &= CCGACG, \\ s_4 &= CGACGA, \\ s_5 &= CGACCG. \end{aligned}$$

Ist der konstruiere Superstring der kürzest mögliche?

Aufgabe 3

In der Übung erhalten Sie eine Kopie des Artikels *Combinatorial Algorithms for DNA Sequence Assembly* von *J.D. Kececioglu* und *E.W. Myers*, erschienen in *Algorithmica* Band 13 auf den Seiten 7–51 im Jahre 1995.

Lesen Sie den Artikel, so dass er in der Übung diskutiert werden kann.