

Prof. Dr. Knut Reinert  
Rene Rahn  
Kathrin Trappe  
Kathleen Gallo  
Thomas Krannich

Institut für Informatik  
AG Algorithmische Bioinformatik

## Algorithmen und Datenstrukturen in der Bioinformatik

### Achtes Übungsblatt WS 12/13

Abgabe Freitag, 14.12.2012, 15:00 Uhr

Name:

Übungsgruppe:

A  B  C

Matrikelnummer:

---

Niveau I

---

#### Aufgabe 1: $q$ -gram-Lemma

Beweisen Sie das  $q$ -gram-Lemma: Zwei Strings der Länge  $w$ , die sich in höchstens  $k$  Stellen unterscheiden, haben mindestens  $w + 1 - (k + 1)q$  gemeinsame  $q$ -gramme.

---

#### Aufgabe 2: Quasar

Bauen Sie das Suffix-Array und die  $q$ -gram-Tabelle für  $k = 2$  und das Suchpattern  $P = \text{ACGACACA}$  auf. Erklären Sie, wie der  $q$ -gram-Index bei der Suche im Quasar-Algorithmus eingesetzt wird.

---

#### Aufgabe 3: Speicher

Berechnen Sie den Speicherverbrauch für ein  $q$ -Gramm Index über DNA in C++. Nehmen Sie  $q=4$  an und ein 32-bit Betriebssystem. Damit Sie eine Intuition für den Speicherverbrauch bekommen, geben Sie den verbrauchten Speicher in MB (Megabyte) an.

---

Niveau II

---

#### Aufgabe 3: Statistik

Gegeben ist eine Datenbank mit 10.000 zufälligen Nukleotid-Sequenzen der Länge 1.000. Die Nukleotide sind gleichverteilt und unabhängig voneinander. Sie haben eine queue von 100 Nukleotiden. Wie oft erwarten Sie im Schnitt die queue in der DB?