

# Präsenzübungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, WS 2013/2014

Prof. Dr. Jens Stoye · Linda Sundermann

<http://wiki.techfak.uni-bielefeld.de/gi/Teaching/2013winter/SequenzAnalyse>

## Präsenzübungsblatt 8, Woche 50/2013

### Aufgabe 1 (Das $q$ -gram Lemma)

Das  $q$ -gram Lemma lautet wie folgt:

*Gegeben ein lokales Alignment der Länge  $L$  mit maximal  $e$  Fehlern, dann enthält die alignierte Region zweier Strings mindestens*

$$T(l, q, e) := l + 1 - q \cdot (e + 1)$$

*gemeinsame  $q$ -grams.*

Beweise diese Aussage.

### Aufgabe 2 (Datenbanksuche)

1. Welche unterschiedlichen Annahmen werden bei

- a) der Suche nach einem optimalen Alignment von zwei Sequenzen und bei
- b) der Suche nach einer ähnlichen Sequenz in einer Datenbank

getroffen? Welche Anforderungen werden dadurch an die Software gestellt, die diese Probleme lösen soll?

2. Erkläre die Unterschiede zwischen *on-line* und *index-basierter* Datenbanksuche. Wie sind die Laufzeiten zur Vorverarbeitung und bei der Suche? Welche Laufzeit ergibt sich bei  $k$  Suchen?

In welchen Fällen lohnt sich eine index-basierte Suche, wann ist die on-line Variante vorzuziehen?