

# Sequence Analysis SS 2013

Freie Universität Berlin, Institut für Informatik  
Knut Reinert, Sandro Andreotti  
Sommersemester 2013

7. Exercise sheet, June 9th, 2013  
Discussion: June 14th, 2013

## Exercise 1.

Prove the following  $q$ -gram lemma:

Let  $P$  and  $S$  be strings of length  $w$  with at most  $k$  mismatches. Then  $P$  and  $S$  share at least  $w + 1 - (k + 1)q$  common  $q$ -grams.

## Exercise 2.

SWIFT algorithm:

- Sketch the Function  $U(n, q, \epsilon)$  for increasing values of  $n$ . Let  $q = 7$ ,  $\epsilon = 0.1$  and  $n_0 = 30$ . Draw it for the intervall  $n = n_0 \dots n_0 + 35$ .
- The Lemma 2.2 in the SWIFT script contains a formula to compute  $w$  (for a  $w \times e$  parallelogram) Show that every local alignment with  $\tau$   $q$ -hits and  $e$  errors lies in a  $w \times e$  parallelogram.

## Exercise 3.

Suffix filters:

Determine **all** strong matches for the given weights and edit distances:

$$\begin{array}{l} i : 0 \ 1 \ 2 \ 3 \ 4 \ 5 \\ t_i : 2 \ 1 \ 1 \ 2 \ 1 \ 1 \\ \text{dist}(A_i, B_i) : 1 \ 1 \ 0 \ 2 \ 0 \ 1 \end{array}$$

## Exercise 4.

Factor filters:

Prove Theorem 3 (optimal factorization) from the script.