

## Aufgabe 2

(1) Seien  $u = \text{acgacgtag}$  und  $v = \text{ggacgtgcag}$  zwei Sequenzen über dem DNA-Alphabet  $A = \{a, c, g, t\}$ . Bestimmen Sie die  $q$ -Wort Distanz von  $u$  und  $v$  für  $q = 2$  und  $q = 3$ .

$q=2$	Auftreten in $u$	Auftreten in $v$	$ G_q(u)(w) - G_q(v)(w) $
aa	0	0	0
ac	2	1	1
ag	1	1	0
at	0	0	0
ca	0	1	1
cc	0	0	0
cg	2	1	1
ct	0	0	0
ga	1	1	0
gc	0	1	1
gg	0	1	1
gt	1	1	0
ta	1	0	1
tc	0	0	0
tg	0	1	1
tt	0	0	0

=> Um die  $q$ -Wort Distanz zu berechnen wird die Häufigkeit des Vorkommens von  $w$  in  $u$  bzw. von  $w$  in  $v$  aufsummiert (letzte Spalte).

=> Die  $q$ -Wort Distanz für  $q = 2$  beträgt:  $qdist(u,v) = 7$ .

=> Für  $q = 3$  wurde dieselbe Tabelle erstellt. Die  $q$ -Wort Distanz beträgt hier:  $qdist(u,v) = 9$

(2) Bestimmen Sie  $edist(u, v)$  und ein optimales Alignment von  $u$  und  $v$ .

$u = \text{acgacgtag}$  und  $v = \text{ggacgtgcag}$  über  $A = \{a, c, g, t\}$

```

a c g a c g t - - a g
  | | | | |   | |
- g g a c g t g c a g

```

Die Edit-Distanz beträgt:  $edist(u,v) = 4$

(3) Ersetzen Sie in u und v die Zeichen a und g jeweils durch r und die Zeichen c und t durch y. Welche Werte erhalten Sie nun für die q-Wort Distanz (für  $q = 2$  oder  $q = 3$ ) und für die Einheitskostendistanz?

$u = r y r r y r y r r$  und  $v = r r r y r y r y r r$  über  $A = \{r, y\}$

$q=2$	Auftreten in u	Auftreten in v	$ G_q(u)(w) - G_q(v)(w) $
rr	2	3	1
ry	3	3	0
yy	0	0	0
yr	3	3	0

=> Die q-Wort Distanz für  $q = 2$  beträgt:  $qdist(u,v) = 1$

$q=3$	Auftreten in u	Auftreten in v	$ G_q(u)(w) - G_q(v)(w) $
rrr	0	1	1
rry	1	1	0
ryr	3	3	0
ryy	0	0	0
yyy	0	0	0
yyr	0	0	0
yry	1	2	1
yrr	2	1	1

=> Die q-Wort Distanz für  $q = 3$  beträgt:  $qdist(u,v) = 3$

**Alignment:**

$r y r - r y r y r r$   
 | | | | | |  
 $r r r y r y r y r r$

$edist(u,v) = 2$

(4) Geben Sie Beispiele für Paare  $(u_1, v_1)$ ,  $(u_2, v_2)$ ,  $(u_3, v_3)$  von nicht leeren Sequenzen und Werte  $q_1, q_2, q_3$  an, für die gilt:  
Dabei können Sie  $q_i \geq 2$  für  $1 \leq i \leq 3$  frei wählen.

a)  $q\text{gdist } q_1(u_1, v_1) > \text{edist}_\delta(u_1, v_1)$

Beispiel:  $u = \text{aaggaa}$

$v = \text{aaaaaa}$

über Alphabet  $A = \{a, g\}$

$q=2$	Auftreten in u	Auftreten in v	$ G_q(u)(w) - G_q(v)(w) $
aa	2	5	3
ag	1	0	1
gg	1	0	1
ga	1	0	1

$\Rightarrow q\text{dist}(u, v) = 6$

$\Rightarrow \text{edist}(u, v) = 2$

b)  $q\text{gdist } q_2(u_2, v_2) = \text{edist}_\delta(u_2, v_2)$

Beispiel:  $u = \text{aaaa}$

$v = \text{aaaa}$

über Alphabet  $A = \{a\}$

$q=2$	Auftreten in u	Auftreten in v	$ G_q(u)(w) - G_q(v)(w) $
aa	3	3	0

$\Rightarrow q\text{dist}(u, v) = 0$

$\Rightarrow \text{edist}(u, v) = 0$

c)  $q\text{gdist } q_3(u_3, v_3) < \text{edist}_\delta(u_3, v_3)$

Beispiel:  $u = \text{aga}$

$v = \text{gag}$

über Alphabet  $A = \{a, g\}$

$q=2$	Auftreten in u	Auftreten in v	$ G_q(u)(w) - G_q(v)(w) $
aa	0	0	0
ag	1	1	0
gg	0	0	0
ga	1	1	0

$\Rightarrow q\text{dist}(u, v) = 0$

$\Rightarrow \text{edist}(u, v) = 2$