

Xiheng He

Lisanne Friedrich

Exercises for Algorithmic Bioinformatics II

Assignment 3

Xiheng He

November 2021

Exercise 3 (FASTA offsets, 10P):

Construct manually, as demonstrated in the lecture *FASTA*, the FASTA offset list and offset vector for the database sequence $s = \text{HARFYAAQIVLHARFYAAQIVL}$ and the query $t = \text{VDMAAQIA}$. Use $ktup = 2$ and hash the query.

Database sequence s:

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 |
|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|
| H | A | R | F | Y | A | A | Q | I | V | L | H | A | R | F | Y | A | A | Q | I | V | L |

Query sequence t:

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|
| V | D | M | A | A | Q | I | A |

hash table:

| ktup | HA | AR | RF | FY | YA | AA | AQ | QI | IV | VL | LH |
|---------------|------|------|------|------|------|------|------|------|------|-------|----|
| position in s | 1,12 | 2,13 | 3,14 | 4,15 | 5,16 | 6,17 | 7,18 | 8,19 | 9,20 | 10,21 | 11 |
| position in t | - | - | - | - | - | 4 | 5 | 6 | - | - | - |
| offset | - | - | - | - | - | 2,13 | 2,13 | 2,13 | - | - | - |

offset vector:

| -7 | -6 | -5 | -4 | -3 | -2 | -1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 |
|----|----|----|----|----|----|----|---|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|
| - | - | - | - | - | - | - | - | - | 3 | - | - | - | - | - | - | - | - | - | - | 3 | - | - | - | - | - | - | - | - |