

Sequence Bioinformatics

WS 2022/23

Tutor: Anupam Gautam

1	2	3	4	5	6	Σ

Marina Dittschar & Clarissa

Auckenthaler

Blatt 8

(Abgabe am 21.12.2022)

Task 1: Implement method `sk()` (1 point)

The code for the method `sk()` starts at line 81 in file `Minimap_Auckenthaler_Dittschar.java`.

Task 2: Implement DNA hash function `h()` (1 point)

The code for the method `h()` starts at line 111 in file `Minimap_Auckenthaler_Dittschar.java`.

Task 3: Implement sketch function `minimizerSketch()` (2 points)

The code for the method `minimizerSketch()` starts at line 134 in file `Minimap_Auckenthaler_Dittschar.java`.

Task 4: Implement indexing function `computeTargetIndex()` (2 points)

The code for the method `computeTargetIndex()` starts at line 176 in file `Minimap_Auckenthaler_Dittschar.java`.

Task 5: Implement mapping function `mapQuerySequence()` (3 points)

The code for the method `mapQuerySequence()` starts at line 216 in file `Minimap_Auckenthaler_Dittschar.java`.

Task 6: Run the program (1 point)

Because of an error in class FastA() from Assignment01 (not able to read in sequences from multiple lines) we needed to transform the input-files. To do this we used the following tool, to remove the break lines, it is available by the following link:

<https://www.gillmeister-software.com/online-tools/text/remove-line-breaks.aspx> Please only use the files which are provided by us.

To run our code direct to the src folder and enter the following command:

```
java assignment08/Minimap_Auckenthaler_Dittschar.java
```

```
"assignment08-08-three-no-breaks.fasta" "assignment08-0820.fasta"
```

Result:

Minimap_Auckenthaler_Dittschar w=10 k=15 targets=1 queries=3

Query: read1

Matches: 1

Target: 1, query: 3 - 177, target: 3 - 177, reverse: 0

```
AGAGTGTTTCAGACATTCTTTAAGCTTGTAATAAAATTTTTGGCTTTGTGTGCTGAC
TCTATCATTATTGGTGGAGCTAAACTTAAAGCCTTGAATTTAGGTGAAACATTTGT
CACGCACTCAAAGGGATTGTACAGAAAGTGTGTTAAATCCAGAGAAGAAACTGGC
CTACTCA
AGAGTGTTTCAGACATTCTTTAAGCTTGTAATAAAATTTTTGGCTTTGTGTGCTGAC
TCTATCATTATTGGTGGAGCTAAACTTAAAGCCTTGAATTTAGGTGAAACATTTGT
CACGCACTCAAAGGGATTGTACAGAAAGTGTGTTAAATCCAGAGAAGAAACTGGC
CTACTCA
```

Query: read2

Matches: 0

Query: read3

Matches: 1

Target: 1, query: 5 - 180, target: 365 - 540, reverse: 0

```
AAAGACACAGAAAAGTACTGTGCCCTTGACCTAATATGATGGTAACAAACAATACC
TTCACACTCAAAGGCGGTGCACCAACAAAGGTTACTTTTGGTGATGACACTGTGA
TAGAAGTGCAAGGTTACAAGAGTGTGAATATCACTTTTGAACCTGATGAAAGGAT
TGATAAAG
AAAGACACAGAAAAGTACTGTGCCCTTGACCTAATATGATGGTAACAAACAATACC
TTCACACTCAAAGGCGGTGCACCAACAAAGGTTACTTTTGGTGATGACACTGTGA
TAGAAGTGCAAGGTTACAAGAGTGTGAATATCACTTTTGAACCTGATGAAAGGAT
```

TGATAAAG

As we can see, there must be an error for finding the reverse complement, we were not able to fix it in time.