Prof. Dr. Daniel Huson
Dr. Benjamin Albrecht
Zentrum für Bioinformatik
Fachbereich Informatik
Mathematisch-Naturwissenschaftliche Fakultät

EBERHARD KARLS

UNIVERSITÄT
TÜBINGEN

Advanced Java for Bioinformatics

WS 2016/17
Due: 27. October 2016

1   Getting started

- Download and install newest JDK (Java SE Development Kit 8u102).

- Download and install newest Intellij IDEA - Java IDE we will be using during the course (https://www.jetbrains.com/idea/).

If you have not used version control before read up on it and make sure you know basic git functionalities (one of the many online git tutorials: https://www.atlassian.com/git/tutorials/).

- Install git and create a GitHub account.

2   Refresh your Java (5 points)

Write a command line tool called FastaTool in Java. FastaTool should be able to read multi fastA, containing RNA sequences and print them formatted on the screen. For the example file (data01.fna) the formatting should look like this:

```
                    1                                                          60
AA03409.1/96-167a   UUUCCCAG-A-UCGCAUG---UUG---AUAUGU-C-CGCCG-GGUCAAUUAU-UAG----
AA09835.1/1-71      CAGGAUAG-U-ACGUAA----AU----CAAUGA-C-GUUGC-GCGGAAACCA-GGG----
AA13372.1/8-81      UCUCCCUG-A-UCGCAAGG--UUAG--AUAUCU-G-GGUAC-GACUUAGCAA-AAU----
AA13373.1/3754-3825 UCUUAAUG-A-UGGCAUG---UUG---AUAACA-C-CACCG-GUCCAAUUGU-UUU----


                    61                                                        119
AA03409.1/96-167    -------------------UC--CCGU-UUG-GCU-AA--GCCCA-U-GUUUGCC-A
AA09835.1/1-71      -------------------AA--GAUGAAAG-GAU-CA--CGGAC-G-GCGAAGU-A
AA13372.1/8-81      -------------------UG--GAUU-UUG-GCU-AC--GCCGC-G-CUUUCUC-U
AA13373.1/3754-3825 -------------------UG--CUCU-UUG-GCU-AA--GCCCU-G-CGGCCUC-G


Number of sequences: 4
Shortest length: 119 (excluding '-'s: ??)
Average length:  119 (excluding '-'s: ??)
Longest length:  119 (excluding '-'s: ??)
Counts: A: ??, C: ??, G: ??, U: ??, -: ??
```

FastaTool should take the file name from the parameter list. The tool should consist of at least four classes: Nucleotide, Sequence, FastaTool, CommandLine.

3   Submission of your work

Please send the name of your GitHub repository to your tutor/ They will check out the repository on the morning due and grade your code.