

Übungen zur Vorlesung Algorithmische Bioinformatik

Freie Universität Berlin, WS 2016/17

Martin Vingron · Annalisa Marsico · Alena van Bömmel · Edgar Steiger · Thimo Wellner

Weihnachtliches Spezialblatt 10 · Ausgabe am 15.12.2016

Abgabe am 3.1.2017 per Slack oder im Tutorium

Name:

Matrikelnummer:

Übungsgruppe:

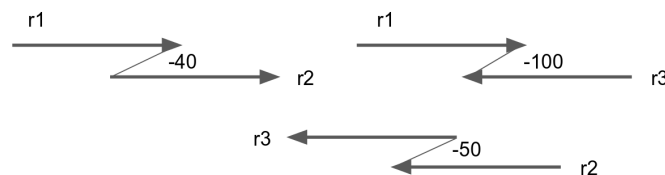
Sie können in diesem Blatt 25 reguläre Punkte und bis zu 45 Bonuspunkte erzielen.

Aufgabe 1 (30 Punkte; Programmieren). Gegeben ist ein fasta-File¹ mit mehreren Reads derselben Länge. Es sollen nun die Überlappungen zwischen den Reads gefunden werden (vgl. Abbildung auf Blatt 8, Aufgabe 2).

- (20 Punkte) Implementieren Sie dazu ein ausführbares Skript, dass
 - ein fasta-File mit Reads einliest,
 - für jede paarweise Kombination von Reads den größtmöglichen Overlap zwischen den beiden Reads bestimmt (achten Sie dabei auch auf die möglichen Richtungen der Reads).
 - die Ergebnisse als Tabelle in ein txt-File mit der folgenden Gestalt schreibt:

```
Seq1 direction1 Seq2 direction2 overlap
r1 + r2 + -40
r1 + r3 - -100
r3 - r2 - -50
```

Dies würde folgenden Überlappungen entsprechen:



- aus der Kommandozeile mit folgenden Parametern aufgerufen werden kann:
`inputfile.fasta outputfile.txt 10 350`, wobei 10 dem minimalen berücksichtigten Overlap zwischen zwei Reads und 350 dem maximalen Overlap entspricht.
- (10 Punkte) Wie kann aus der Tabelle in 1c der Minimal Spanning Tree (und damit auch das Layout und die globalen Koordinaten der Reads) konstruiert werden? Beschreiben Sie entweder das Vorgehen verbal oder geben Sie Pseudocode an oder implementieren Sie sogar eine Routine dafür.

¹Material 3: <https://www.molgen.mpg.de/3719136/reads.txt>

Aufgabe 2 (30 Punkte; Programmieren). Human miR-155 is an important microRNA playing a role in regulating genes of the immune response. The sequence of miR-155 (in fasta format) is given on the lecture website². The task is to calculate a combined score based on

- an alignment score of a local alignment between the gene and the microRNA and
- the minimum free energy of the miRNA:mRNA duplex,

for given 3' UTRs of putative microRNA target genes.

- (a) Implement the local alignment of miRanda-algorithm as explained in the recursive formula on the lecture slides, with the following scoring scheme:

$$S(a, b) = S(b, a) = \begin{cases} 5 & \text{if } a = G \text{ \& } b = C \\ 5 & \text{if } a = A \text{ \& } b = U \\ 2 & \text{if } a = G \text{ \& } b = U \\ -3 & \text{for any other base pairing} \end{cases}$$

Use a linear gap penalty: $g = -2$. In addition, take into account the 5' to 3' asymmetry such that the score of the alignment of the first 10 bases at the 5' of the microRNA (e.g. first 10 bases at the 3' of the target RNA) will be multiplied by 2. For simplicity, you can use the python implementation of the Smith-Waterman algorithm³ and then modify the scoring function.

- (b) Calculate the alignment score for all three putative target genes⁴.
- (c) Compute the energy of the miRNA:mRNA duplex for each putative target RNA separately (see lecture slide 28) using `RNAfold` implementation from the Vienna-RNA-Package⁵.
- (d) Compute a combined score from the alignment score and from free energy. Hint: you need to convert your scores to Z -scores; the two scores are on different scales and cannot be compared directly.
- (e) Compare your results with microRNA target predictions for miR-155 on www.microRNA.org. Do you get the same ranking? If not, what could be the reason?



6

²Material 1: <https://www.molgen.mpg.de/3716658/mir-155.txt>

³Check this code: <https://gist.github.com/radaniba/11019717#file-smithwaterman-py> or the `swalign` package: <https://github.com/mbreese/swalign>

⁴Material 2: <https://www.molgen.mpg.de/3716667/targetrnas.txt>

⁵<http://www.tbi.univie.ac.at/RNA/index.html>

⁶Christmas icons created by Hand-drawn goods