

Genomics

Freie Universität Berlin, Institut für Informatik
Knut Reinert, Peter Robinson, Sebastian Bauer
Wintersemester 2012/2013

6. Übungsblatt vom 30. November 2012
Diskussion am 6. Dezember 2012

1. Outline a possible approach to estimate the relative abundance of species contained in the sample assuming that the genomes of the species are all given.
2. Define the *repeat resolution problem*. For which specific applications is it relevant? What is a *defined nucleotide position* (DNP) in this context?
3. What characteristics can be used to help distinguishing reads from a repeated part of the genome with a small amount of mutations from reads with small amount of sequencing errors?
4. Suppose that we should resolve possible repeats in a set of 5 reads. In previous steps, we already identified 3 DNPs in the reads as follows:

R1 ---T--A--G----

R2 ---T--A--G----

R3 ---T--G--C----

R4 ---G--G--C----

R5 ---G--G--T----

- (a) Specify the weighted K_5 for the k star problem.
- (b) Formulate the corresponding integer linear programming problem for $k = 2$.
- (c) Find the optimal solution to that problem.