

Algorithms in Bioinformatics II, SS2004

Assignment sheet # 13

Daniel Huson

July 11, 2004

The aim of this assignment sheet is to write a program `albi2.assembly.Compomers` for sequencing by compomers, based on: <http://www-ab.informatik.uni-tuebingen.de/teaching/ss04/java/assign13.zip>.

Input are four files, each containing an MS spectrum obtained from a partial cleavage dataset for one of the four “cut” bases A, C, G or T. You can generate these from the file `sample` using the program `albi2.assembly.CompomerSimualator.java`.

We make the following assumptions.

- There exist two special “bases”, 0 and 1 that are situated at the first and last position of the sample sequence, respectively.
- The weights of the “bases” 0, 1, A, C, G and T, are very different, as specified in the Java code.
- The spectra are error free, that is, all reported weights are exact, and there are no false positive or false negative entries.
- The fragments corresponding to a spectra contain either 0 or 1 occurrences of the corresponding cut base.
- The sample sequence is known to have length < 100 .

1 Reading the Spectra (1 point)

Implement a method `albi2.assembly.Compomers.readSpectrum` that reads a spectrum from a file and puts it into a bit set. Apply it to all four input spectra.

2 Compomers from Spectra (4 points)

Implement a method `albi2.assembly.Compomers.computeCompomers` that determines the list of compomers from a given spectra. Apply it to all four input spectra.

3 The undirected seqeuncing graph (5 points)

Implement a method `albi2.assembly.Compomers.computeGraph` that computes the undirected sequencing graph from a list of compomers and apply in all four cases. Display the graph using the provided `showGraph` method.

4 Sequence reconstruction (5 additional points)

Implement a method `albi2.assembly.Compomers.reconstructSequence` that takes as input the four undirected sequencing graphs and produces as output a reconstruction of the original sequence.

Assignments due: **Monday, July 19, 10am**