DNA SEQUENCING AND ASSEMBLING — APPROACHES, GRAPH MODELS, AND ALGORITHMS

Summary

Reading genetic information of an organism, i.e. reading a sequence of nucleotides of a DNA fragment, can be done in two or three stages. In the first stage, the sequencing, one can obtain sequences up to a few hundreds of nucleotides. There are several approaches to carry out this stage. The historically oldest approach is gel electrophoresis, also called by the name of the author – the Sanger method. Another approach is sequencing by hybridization, which is technologically more sophisticated and it involves also algorithmic methods to process the experimental data (as opposed to the previous approach). The novel, fully automated approaches (owned by Roche, Illumina, Applied Biosystems) generate millions of short DNA sequences in short time. Next stage in reading a DNA sequence is the assembling: the output of the sequencing stage is assembled together into longer contigs of length up to even a few million nucleotides. The last stage, called the mapping or the finishing, consists in scheduling assembled sequences in the right order.

The methods presented in the paper are only a part of immensely rich literature, which is available for the DNA sequencing and assembling. They were chosen both from the point of view of their importance for the development of this research branch (historically most important approaches and algorithms) and for their attractiveness (interesting graph models). The meaning of the sequencing and the assembling as the first steps on the way of understanding genetic information of organisms, guarantees further development of associated biochemical and computational approaches..