

DEVELOPMENT OF MATHEMATICAL MODELS AND METHODS FOR HIGHLY ACCURATE PREDICTION OF CODING SEGMENTS OF EUKARYOTIC DNA

A.N. Gaiduk, V.A. Galinskij, V.A. Voloshko

*Research Institute for Applied Problems of Mathematics and Informatics
Belarus State University, Minsk, Belarus*

The identification of coding segments has been one of the most difficult problems in bioinformatics. We propose probability models and methods, allowing to increase accuracy of prediction of protein-coding segments of eukaryotic DNA sequences.

At the present time the large amount of methods proposed and the number of software systems developed for prediction of protein-coding segments of eukaryotic DNA sequences. The basic lacks of the methods proposed and the programs developed are missing and concatenation of coding segments, as well as errors in determining the boundaries of the coding segments.

We propose probability models of protein-coding segments of eukaryotic DNA based on:

1. n -dimensional ($n=3, 4, 5, 6$) probability distribution of nucleotides.
2. parsimonious high order Markov chains.

We also present mathematical methods for prediction of protein-coding segments in eukaryotic DNA sequences based on our models