

**Adam Makuchowski**

SILESIAAN UNIVERSITY OF TECHNOLOGY

e-mail: adam.makuchowski@polsl.pl

**mgr Adam Makuchowski**

SILESIAAN UNIVERSITY OF TECHNOLOGY

**dr inż. Pokrzywa Rafał**

SILESIAAN UNIVERSITY OF TECHNOLOGY

**prof. dr hab. inż. Polaski Andrzej**

SILESIAAN UNIVERSITY OF TECHNOLOGY

## Discovering motifs in DNA sequences

One of the important aspects of molecular biology is to understand the complex mechanisms regulating a gene expression. One of the steps in the process of exploring regulatory mechanisms is discovering regulatory motifs that influence gene expression. Gene expression is transformed by the interaction of transcription factors with their corresponding binding sites. The purpose of presented algorithm is to detect the conservative motifs in DNA sequences, in order to identify regulatory sites.

New algorithm is presented in this paper that allows discovery of new motifs in a set of related regulatory DNA sequences and also in genome-wide search. This algorithm uses a heuristic approach based on the structure of suffix trie. For representation of motif sequences, we used a position specific scoring matrices (PSSMs), which are widely used for this purpose. In addition, two approaches have been examined: considering prior residue probability of background, and omitting real value probability. Taking into account the actual likelihood of the background during discovering of motifs, improves the quality of found motifs. Proposed algorithm was tested on reference genomes of human and mouse. The results obtained from the algorithm were compared with other known algorithms. The comparison of these algorithms are performed based on the following comparison measurements: nucleotide Performance Coefficient, Site Sensitivity, Site Positive Prediction, and Site Average Performance. From experiments on real biological data sets, we observed that the applications such as genome-wide search can be identified, in which the algorithm behaves better than other existing tools to search for motifs. But in the case of smaller data sets, average values of measurements were comparable to other existing motif finding tools.