



Contribution ID: 94

Type: Oral Presentation

Finding motifs in DNA and protein sequences using set match

Monday, 11 July 2016 12:00 (20 minutes)

Abstract content
 (Max 300 words)
 http://events.saip.org.za/getFile.py/?target=_blank
 Formatting & Special chars

Background: Gene selection affects many biological traits that influence diseases. Accurately finding genes that influence traits of interest is very important as isolating such genes is useful in diagnosis and drug development. An important process in gene identification is isolation of motifs or short sequences of amino acids. These motifs lead to finding relationships among genes.

Results: This paper proposes using both exact match and expectation maximization string searches to find motifs in genes. String matching techniques are useful in finding motifs. A new string matching approach that combines Boyer-Moore and Expectation-Maximization is recommended after implementing and analysis of some of the classical methods. Motifs components were placed in sets rather than exact strings. The method improved motif identification of MBD found in human hemoglobin that are preserved but have some differences.

Conclusions: The problem of string matching with application to Bioinformatics and Genomics was studied. Implementation of the simple match, Knuth-Morris-Pratt, and Boyer-Moore methods shows that the Boyer-Moore method is the most efficient. To enable this efficient method be more applicable to genomic motif finding, a preprocessing phase that uses expectation-maximization and creates sets of equally likely residues is recommended.

Primary author: Mr CHIRWA, Robert (University of Louisville)

Presenter: Mr CHIRWA, Robert (University of Louisville)

Session Classification: Parallel Track B

Track Classification: Computational Biology