Abstract

Basic Local Alignment Search Tool (BLAST) is one of the most widely used bioinformatics tools to determine similarities between genomic sequences. Ever since its inception several algorithmic improvements have been made to improve speed and runtime memory requirements without affecting the sensitivity and selectivity of the tool. Fast search algorithm (FSA) BLAST has been the most successful among such improvements with 20-30% faster processing rate. A DFA with hashed prefix word structures for the hit detection process in FSA BLAST has been proposed in the earlier work. Coding of the algorithms and testing on protein samples showed that the use of the new structure resulted in significant reduction in run time space but not the hit detection time. This paper proposes the use of a BLOSUM trie structure which eliminates the process of computing neighborhood words, resulted in a reduction of up to 75% in hit detection time. Tests were conducted with different BLOSUM matrices and threshold values and the proposed algorithm was found to be beneficial in terms of space, time complexity and accuracy without compromising on sensitivity and selectivity of the currently being used algorithm.
References

- WU-BLAST [http://blast. wustl. edu/]
- Delaney, S. Butler, G. Lam, C. and Thiel, L. Three Improvements to the BLASTP Search of Genome Databases, 0-7695-0686-0/00 $10. 00 _ 2000 IEEE


Index Terms

Computer Science
Bio-medical Sciences

Keywords

Deterministic finite automaton prefix word table neighborhood words query

pointers

hits

and BLOSUM Trie