

A VERSATILE EXACT PATTERN MATCHING ALGORITHM FOR BIOLOGICAL SEQUENCES

Latha Parthiban

Assistant Professor of Computer Science
Pondicherry University Community College
Puducherry, India

ABSTRACT

Pattern matching is essential for locating nucleotide or amino acid sequence patterns in the biological databases. In this paper, Zhu-Takaoka Fast Search algorithm (ZTFS) algorithms are modified to reduce the computational time complexity and also the number of shifts done while comparing the pattern with the given string. The experimental results on biological sequence of Enterobacteria phage lambda and Measles virus proves that the modified algorithm works faster than other existing algorithms, especially in case of longer patterns of nucleotide sequences. The algorithm has also been extended to classify the input biological sequences into different possible combinations of tricodons and amino acids present in the sequence. Computational complexity calculations and performance analysis for character comparisons are presented to indicate the potential reduction in time complexity of the proposed algorithm.

KEY WORDS: Pattern matching, Amino acids, Character comparisons, Nucleotide sequences, Measles virus

تحديد الخوارزمية الدقيقة متعددة الإستخدامات لسلسلة البيولوجية
الملخص

يعتبر مطابقة النمط اساسي لتحديد موقع النوكليوتيدات nucleotide أو تسلسل أنماط الحمض الأميني amino acid في قواعد بيانات البيولوجية. في هذه الورقة تم تعديل خوارزمية (ZTFS) لتخفيض الوقت الحسابي المعقد و أيضا عدد التحولات المعمول بها خلال مقارنة نمط تلك السلسلة. و قد أثبتت نتائج التجريبية في المتتالية البيولوجية في فيروس Enterobacteria phage Lambda و فيروس الحصبة Measles virus أن عمل الخوارزمية المعدلة أسرع من الخوارزمية الموجودة و خصوصا في حالة سلسلة الطويلة لعينة النوكليوتيد Nucleotide. و قد تم أيضا توسيع خوارزمية لتصنيف مدخلات متواليات البيولوجية في مختلف التوليفات الممكنة في سلسلة الأحماض الأمينية و tricodons. و قد تم تمثيل التعقيدات الحسابية و تحليل الفعالية لمقارنة الموصفات لدلالة على انخفاض المحتمل في تعقيدات وقت الخوارزمية المقترحة.