

An Efficient In-Exact Algorithm for Finding Approximate Alignment in DNA Sequences

Prepared By

Rufayda Mahmoud Eqailan

Supervised by

Dr. Akram Othman Al Mashaykhi

Abstract

In bioinformatics the sequence alignment which depends on string matching algorithms is the basic of data mining and analysis, it is the way of represent the arranging of sequences of DNA, RNA and protein. It is very important to find the homology and the similarity that may lead to know the consequence of functional, structural or evolutionary relationship between the sequences.

The algorithms that are presented in this area to solve sequence alignment matching problem have many different types, all of them try to find the optimal alignment solutions with focusing on achieving high accuracy in the results within the most minimum time for matching time. These are the challenges face these algorithms. This study presents new algorithm that accomplish this equation of challenges, area of research was on biology data, pair wise, indexed and local alignment algorithm. The expected results appeared that the new algorithm performed better than very popular algorithms in this field in most of the cases.

الخوارزمية الفعالة لإيجاد التطابق التقريبي باستخدام البحث التقريبي في سلاسل الحمض

النوي

إعداد

رفيدة عقيلان

إشراف

الدكتور أكرم المشايخي

الملخص

في المعلوماتية الحيوية مطابقيه السلاسل الجينية هي وسيلة لترتيب سلاسل الحمض النووي الريبي، او البروتين

لتحديد مناطق التشابه بينها و التي ربما تكون نتيجة لعلاقة وظيفية أو هيكلية او تطوريه بين هذه السلاسل.

في هذه الاطروحة نقدم خوارزمية مفهرسه محلية للمطابقة بين سلاسل الحمض النووي باستخدام المطابقة

المتماثلة و غير المتماثلة.

الخوارزميه المقدمة في هذه الاطروحة تركز على ايجاد حل للمشكلتين الاساسيتين اللتان تواجهان ايجاد هذه

المطابقة و هما ايجاد التطابقات الامثل في اقل وقت تنفيذ ممكن. اظهرت النتائج التجريبية ان اداء الخوارزمية

افضل من أداء خوارزميات تعتبر اساسية فس هذا المجال في معظم الحالات.