中國醫藥大學

醫務管理研究所碩士論文

論文編號:IHAS-292

序列組共同具有短序列樣式之詳盡搜尋演算法

Finding Motif by exhaustive algorithm

指導教授：莊　振　村　博士

研 究 生：楊　鎮　嘉　　撰

中華民國九十三年六月

目 錄

表目錄

圖目錄

摘 要

Motif Finding 這個問題引起許多人的興趣，因此也造就了許多搜尋演算法的產生。然而許多的演算法在面對Pevnzer 和 Sze所提出的Challenge Problem無法得到一個非常好的結果。而本研究主要目的是利用自行構思的詳盡式搜尋來解決Motif finding的問題，嘗試突破PROJECTION在搜尋*(9,2)*、*(11,3)*、*(13,4)*、*(15,5)*、*(17,6)*等信號其結果不佳的問題，以及解決當序列長度(N)增加時所造成搜尋motif執行效率和準確度下降等問題。

一般人認為詳盡搜尋的方式雖然可以準確的找到 motif，但是當基因序列和搜尋的 motif 長度過長的時候，在運算時間容易呈指數的倍數成長，因此本研究透過一些輔助的技巧可以避免因長度的增加使得運算時間的增長，又能夠兼顧效率及精準的方式來找到最好的結果。可以預期的是此方式可以得到不錯的結果外，在方法上更可擴展到其他 motif 相關的研究領域上。

Abstract

Recently, motif finding became a very popular area in bioinformatics, thus more and more researches are interested in discover motif. However, many algorithms can not solve the Pevnzer and Szes ' challenge problem. It motivates my algorithm to purpose construct an exhaustive method to improve the motif finding performance in discover signals such as: *(9,2), (11,3), (13,4), (15,5),* and *(17,6)* and to solve the problem that accuracy will be descending while sequence length is increasing.

Although exhaustive search method could find motifs accuracy, it still needs to face the problem that computing time will be grown exponentially by length increasing of genomic sequence and motif. The research through provide assist skills not only to avoid the length of sequence increasing effect computing time, but also efficiency and accuracy to discover the optimal result. We could expect the research will bring well performance and apply in other bioinformatics domain related motif finding is expandable.

誌 謝

首先要感謝指導教授 莊振村博士引導我在踏入資訊的領域並且細心的指導與教悔，在這兩年的時間學習到做研究的精神、增進自我的邏輯與獨立思考的能力，此外，感謝吳禮字老師給予生物上的教學指導，如此才得以順利完成論文研究。另外，還要口試委員王旭正博士、周志賢博士於百忙之中審閱論文，並提供建議使得本論文更加完善。

在面臨問題的討論過程感謝梅芬在生物方面的知識的協助。最後，要感謝我的父母給我的支持與鼓勵，並且提供一個舒適安穩的求學環境，讓我能毫無牽掛的完成碩士學位。在此，謹將這份榮耀獻給我的家人。

研究生 楊鎮嘉 謹致

民國93年6月