

# 運用 APPA 擷取演算法探測 DNA 序列上多元重要基因片段

研究生：林宛嫻

指導教授：胡毓志博士

國立交通大學資訊科學所

## 摘要

由於基因定序技術發展的快速以及眾多基因計劃的進行，使得基因序列資訊量的增加以倍數方式在成長。面對如此急速增加的資料量，單純使用人工實驗由生物序列中分析出重要基因片段已不再具有可行性。然而如今大多探測多元重要基因片段的研究方法皆採用同時探測單元重要基因片段所組成的多元重要基因片段，致使其不僅僅在效能上受制於片段區間的長度，更無法擴展探測由三個或三個以上單元重要基因片段所組成的多元重要基因片段。因此，本篇論文提出由 Apriori 擷取演算法概念所延伸的 APPA 演算法，其以 MERMAID (Hu, 2003) 為前置處理程序找出單元重要基因片段，進而架構推論出所有可能的多元重要基因片段。我們利用含有已知多元重要基因片段的生物 DNA 序列和人工 DNA 序列，驗證 APPA 在探測多元重要基因片段上的準確度。而實驗結果亦證明，我們的方法不僅能找出較為複雜多元的多元重要基因片段，且其結果比 Dyad Analysis (van Helden, 2000) 來得精確。