

計畫編號：CCMP95-RD-025

行政院衛生署九十五年度科技研究發展計畫

台灣中草藥基因體基原鑑定及中英文基因體資料庫之建立
Traditional Chinese Medicine genomic identification and
construct the bioinformatic database in Taiwan

研究報告

計畫委託機關：中國醫藥大學

計畫主持人：謝長奇

研究人員：張永勳、郭昭麟、劉淑鈴、邱泰惠、陳方周、施維真

執行期間： 95 年 01 月 17 日至 95 年 12 月 31 日

計畫編號：CCMP95-RD-025

各機關研究計畫基本資料庫之計畫編號：

行政院衛生署九十五年度科技研究發展計畫

台灣中草藥基因體基原鑑定及中英文基因體資料庫之建立
Traditional Chinese Medicine genomic identification and
construct the bioinformatic database in Taiwan

研究報告

計畫委託機關：中國醫藥大學

計畫主持人：謝長奇

研究人員：張永勳、郭昭麟、劉淑鈴、邱泰惠、陳方周、施維真

執行期間： 95 年 01 月 17 日至 95 年 12 月 31 日

編號：CCMP95-RD-025

行政院衛生署中醫藥委員會九十五年度
委託研究計畫成果報告

台灣中草藥基因體基原鑑定及中英文基因體資
料庫之建立

Traditional Chinese Medicine genomic
identification and construct the bioinformatic
database in Taiwan

計畫委託機關：中國醫藥大學

計畫主持人：謝長奇

研究人員：張永勳、郭昭麟、劉淑鈴、邱泰惠、陳方周、
施維真

執行期間： 95 年 01 月 17 日至 95 年 12 月 31 日

目錄

	頁碼
中文摘要	1
英文摘要	2
前 言	3
材料與方法	6
結 果	9
討 論	34
結論與建議	35
誌 謝	35
參考文獻	36
圖	38

台灣中草藥基因體基原鑑定及中英文基因體資料庫之建立

謝長奇

中國醫藥大學

摘要

台灣市售中草藥材品種繁多而複雜，為確保藥材基原，中藥品管制度的鑑定標準是首要之工作，主要是將中藥材之規格包括基原、性狀及化學規格予以規範，隨著後基因體時代的來臨，利用基因體定序資料庫來鑑識與管理中草藥材，已成為鑑別標準化的趨勢。隨著分子生物學技術的發展，研究人員開始應用基因體(genome)生物多態性原則，以各種不同的基因體指紋鑑定方法來檢測屬(genus)、種(species)、變種(variety)、亞種(subspecies)甚至於特殊品系、分離株(strain, clone)之間的關係，近年來專利物種的最終鑑定方式乃是採用結合核糖體內轉錄間隔區(ribosomal RNA gene internal transcribed spacer, ITS)與內轉錄間隔區(intergenic spacers, IGS)作為專利植物鑑別的依據之一。我們整理了發表於Genbank中的資料發現，現行台灣傳統藥典中所載錄的藥材中，可搜尋到ITS的序列資料僅約32%，配合台灣傳統藥典的資料與其逐年更新的藥材庫，建立並確立基因體的資料庫是刻不容緩的工作。本計畫第一年主要工作為定序易混用、誤用之中草藥基因體資料，並定序市售使用量較大之中草藥，並建置『台灣中草藥基因體資訊網』。計畫執行一年的成果中，我們陸續將定序完整的資料發表在GenBank資料庫中，並於發表資料明列本計畫編號等，將相關資料整理完整，並以Perl程式語言進行序列分析，其分析結果表明其核苷酸之組成比例，雙核苷酸、三核苷酸比例等，並進行虛擬電泳膠之演化，不僅提供種屬之間比對之相異相同性之比較，更進一步將未來如需進行快速偵檢之限制酶切多型性分析立下基礎，這些成果與分析之資料經我們設計之網頁登載於中國醫藥大學中草藥生物資訊網中，將成果提供學研界參考，並希望學研界使用本系統後能提供建言未中草藥基因體生物資訊研究與基原鑑定能更趨完備。

關鍵詞：台灣中草藥、基因體鑑別、基因體資料庫、核糖體內轉錄間隔區

Traditional Chinese Medicine genomic identification and construct the bioinformatic database in Taiwan

Hsieh, Chang-Chi

China Medical University

ABSTRACT

The marketing and folk pharmaceutical medicines are very complicate in Taiwan. To assume the original of these pharmaceutical medicines, the quality control and identification standard will be the most important issue. They include the plant origin and chemical ingredients. But most important is to identify the genomic profiles of these pharmaceutical medicines in the post-genomic era. The researchers used genomic polymorphism to identify the plant in genus, species, variety even in strain and clone. In the recent years, the pattern plant was according to the ribosomal RNA gene internal transcribed spacer (ITS) to be one of the identification profiles. In our studies, there are only 32% of ITS profiles be published to Genbank in the "Taiwan traditional and folk medicine bioinformatics database". It presented as the first priority to complete the database of this genomic information for public healthy. In this study, we identify the most confused and mistake and the most popular pharmaceutical medicines in market, and construct the web page of the genomic database. The identical ITS sequences were published in GenBank in National Center for Biotechnology Information (NCBI). We constructed a Perl program to analysis the sequence composition in mono-nucleotide, bi-nucleotides and tri-nucleotides; restriction fragment length polymorphisms (RFLP) were also present as the virtual gel electrophoresis. These databases will give us the quick diagnosis in the identification of the confused or mistake used herbal medicine and further phylogenic research. We present our research result in the "Bioinformatic Center in China Medical University" to provide researcher a reference database and hope them can response us the new idea to improve this program.

Keywords : Traditional and folk medicine, genomic identification, genomic databases, internal transcribed spacer

壹、前言

台灣市售中藥材大部分仰賴進口，藥材來源分歧，品種繁多而複雜，品質不易控制，常因來源不同而有很大差異，為確保藥材基原，亟需針對常用中藥材進行相關檢驗之監測，以瞭解相關檢驗之差異性。行政院衛生署中醫藥委員會經數年努力完成國內第一部中華中藥典^[1]，於九十三年三月九日公告，並自九十三年五月一日起實施，並於九十四年八月三十一日公告將中華中藥典更名為台灣傳統藥典^[2]，提供有效的鑑定依據，確定藥材基原，為維護國人用藥安全與健康之重要課題。

隨著後基因體時代的來臨，利用基因體定序資料庫來鑑識與管理中草藥材，已成為鑑別標準化的趨勢。計畫執行期間將倚重中醫藥委員會、農委會等相關計畫主持人所建置之中草藥基原庫之經驗，邀請相關人員參與計畫之執行，並向主持人及研究中草藥基原之相關人員請教，提供未來台灣傳統藥典增編時較完整可行之參考規格。瞭解市售誤用混用與正品之相關藥材的差異性，以提供中草藥基原鑑定一個較完整可行之參考資料。隨著分子生物學技術的發展，研究人員開始應用基因體(genome)生物多態性原則，以各種不同的基因體指紋鑑定方法來檢測屬(genus)、種(species)、變種(variety)、亞種(subspecies)甚至於特殊品系、分離株(strain, clone)之間的關係，近年來專利物種的最終鑑定方式乃是採用結合核糖體內轉錄間隔區(ribosomal RNA gene internal transcribed spacer, ITS)與內轉錄間隔區(intergenic spacers, IGS)作為專利植物鑑別的依據之一，在中國大陸逐步進行中藥材的基因體鑑別工作，最近的研究成果顯示，針對麻黃屬 8 種藥用資源進行收集及鑑定^[3]。我們整理了發表於 Genbank 中的資料發現，現行台灣傳統藥典中所載錄的藥材中，可搜尋到的 ITS 與 IGS 僅約 32% (表)，配合台灣傳統藥典的資料與其逐年更新的藥材庫，建立並確立基因體的資料庫是刻不容緩的工作。

異種生物之間在核苷酸序列上的差異可由基因體中非轉錄片段鑑定出來，以基因體核酸序列為主的指紋分析中常應用於遺傳疾病的檢測，法醫學鑑定，流行病學監測(禽流感、口蹄疫、SARS 等)與植物的遺傳與育種鑑別，使用分子標記(Molecular marker)的優點有：

1. 數據穩定
2. 與生長發育無關，取材不受限制
3. 在真核生物中，因為基因組中存在著大量不編碼的 DNA 序列，它們的變異比較不受或較少受自然環境選擇的影響

而基因體鑑別的分標記包括下列種類：

第一類為以電泳、分子雜交為核心技術，代表的方法有：RFLP 等。

RFLP (restriction fragment length polymorphism)為限制酶切割片段長度多態性分析，是指物種在演化進程中，DNA 序列發生插入、缺失等突變，從而改變了限制酶(restriction endonuclease)的識別位點，最早被利用在鑑別噬菌體基因之多態性^[4]。因此，同種生物不同個體的 DNA 分子用同一種限制酶，會產生不同長

度的片段，在凝膠電泳後，經專一性標示探針偵測下，於 X 光底片上呈現不同的片段條帶多型性(length polymorphism)。只要限制酶選擇合適，對各植株均能顯示多態性和特異性，常用於“種”以下的分類。但這一類方法的缺點是必須搜尋不同的酶切點，不同物種間常無法共用同一種限制酶的試驗。本試驗最早於 1999 年應用於冬蟲夏草的鑑別^[5]，2003 年亦有人參的 RFLP 資料發表^[6]。

第二類結合 PCR、電泳核心技術，代表方法有 RAPD、SSR、AFLP 等。

RAPD (Random Amplified Polymorphism DNA)，隨機擴增核苷酸多態性分析為 1990 由 Dr. Williams 開發出來^[7]，利用單一股的短鏈引子(9-10 mers oligo primer)為 PCR 擴增的引子，經由非專一性的片段擴增在電泳呈相下，進行多態性分析，由於其 RAPD 引子短，因此黏著溫度低，一般為 35-37°C。實驗步驟少，省時、省力、速度快，不需先知道基因組的任何分子資訊；所使用材料不需為有性世代，亦可用任何部位的材料，在沒有有性世代的線粒體、葉綠體 DNA 研究中也可使用；引物沒有嚴格的種屬界限，同一套引物可用於任何一種植物的研究，具有廣泛性和通用性。最大的缺點為必須篩選多達上百個引子，且不同的多態性分析，其所用引子並不相同。

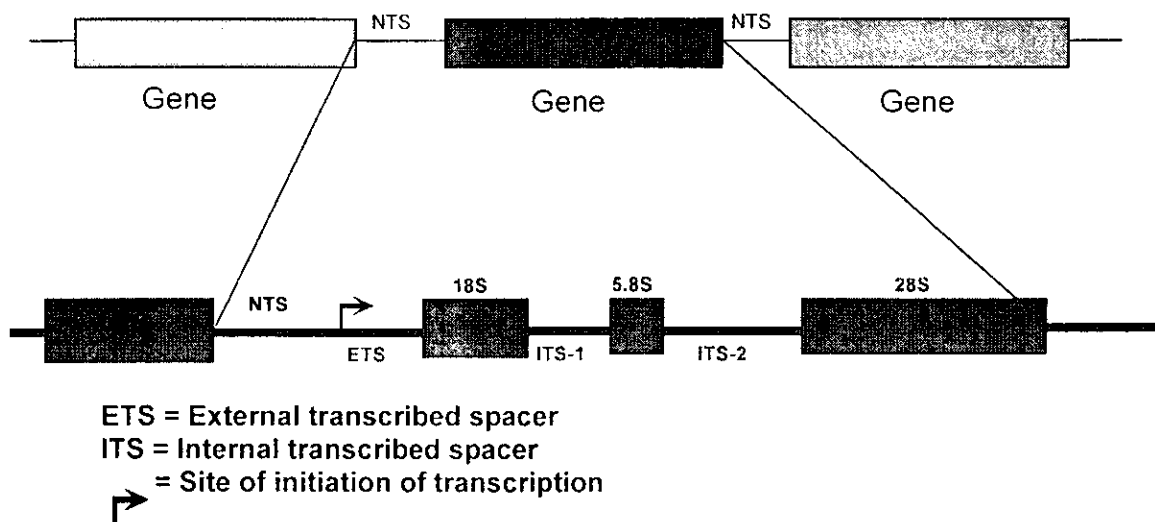
AFLP (Amplified Restriction fragment polymorphism) 限制酶片段擴增多態性分析為 1992 年 Dr. Zabeau 開發出來^[8]，結合了 RFLP 和 PCR 技術的特點，具有 RFLP 技術的可靠性和 PCR 技術高效性。基本原理為對基因組 DNA 進行酶切後，連上雙鏈人工接頭，根據接頭的核苷酸序列和酶切位點設計引物，進行選擇性擴增，它將 RAPD 隨機性和專一性擴增巧妙結合，再選用內切酶以達選擇的目的。它對於 DNA 品質要求較高，需避免其他 DNA 污染和抑制物質的存在，基因組的不完全酶切會影響實驗結果，並且需於於 4%-6%的變性聚丙烯醯胺凝膠(denatured PAGE)上分離微片段，技術層次較高；另外它是一種專利技術，試劑組價格昂貴。

SSR (Simple sequence Repeat)，單一重覆序列分析，也稱為微衛星分析(microsatellite)，或稱之為短鏈重覆序列分析(Short Tandem Repeat, STR)^[9]，所謂的衛星 DNA(Satellite DNA)是將真核生物基因組酶切、離心後，在主帶附近會出現(AT)含量很高的簡單高度重複序列。而微衛星 DNA 是另一類由幾個核苷酸(一般 1-5 個，序列很短)為重複單位串聯而成的 DNA 序列，在人或動物為重覆的(TG)_n，植物為(AT)_n，而微衛星 DNA 兩端多是高度保留的單拷貝序列(single copy sequence)，我們可以根據兩端序列的一致性，設計引子，進行 PCR，電泳分離，染色顯相以檢測微衛星序列多態性。

第三類結合 PCR 與 DNA 定序為核心技術，代表方法有 ITS、IGS 序列分析，可供鑑別的等級可精確到種株(strain, clone)的階段。

ITS (Internal transcribed spacer)，核糖體內轉錄間隔區序列分析，它是將介於 rRNA 基因序列中之間隔序列(圖一)，並不屬於基因表現序列^[10]，種屬間變異性大，並可鑑定到品系(strain)，分離株(clone)等，其優點是可用一組專屬引子，對所有物種進行分析，試驗方法異於標準化，所提供之 DNA 序列可作為演化數

分析之依據，並可推算演化距離。由於生物在演化過程中，突變是演化的必備要素，然而致死性突變會造成物種的滅絕，在兼顧演化的需求與物種的存活，功能性基因的高度保留性與轉錄間隔區之高變異性便是現今物種多樣性的一個寫照，生物學家在研究物種間的演化過程，傳統的方式為利用功能性器官型態的種原鑑定，在近代分子生物學的演化觀點與生物資訊學的發展之下，演化過程中的種原鑑別已經進展到使用基因體序列為鑑定的依據，並於 1993~1995 年間使應用於植物的鑑別工作^[11-13]。並於 1999 年首度應用於虫草的鑑定並比較四種虫草藥材之 ITS，並與傳統之 RFLP 相互比較^[5]。



圖一、核糖體RNA內轉錄間隔區示意圖

在法醫學鑑定學的領域中有一個共識就是：我們的組成是甚麼那是不變的(“What we are”，是 constitution，不變的，是基因體的內容，就是藥材的本質、基原)，但我們會作甚麼是未知的(“Who we are”，基因體產物間的相互作用，與其衍生之生理生化反應，包括藥材之成份的變化)，傳統藥材基原鑑定為以外觀型態，配合顯微組織切片方式鑑別，近年來配合指標成份的指紋圖譜更能增加鑑識的可靠性，相信再配合上基因體的鑑別將使得鑑定工作更趨完美，由於早期基因體定序的複雜性，設備耗材的昂貴，常使得鑑定工作變得不切實際，但近年來由於分子生物技術的突飛猛進，定序鑑別的價格不斷下滑，定序精準度更高，更快速，以原始基因序列的指紋圖譜鑑識已儼然成為標準程序，由基因體資料庫(Genbank)中我們可以看見，專利物種的最終鑑定方式乃是採用結合核糖體內轉錄間隔區(ribosomal RNA gene internal transcribed spacer, ITS)與內轉錄間隔區(intergenic spacers, IGS)作為專利植物鑑別的依據之一。舉例來說：冬蟲夏草(*Cordyceps sinensis*)為珍貴的中藥材，但各地“品種”各有其特色，從 Genbank 的查詢中可以看出，以 ITS 來區分，可分出 30 種菌株 (strain)^[14]，藥材之分離株多已完成 ITS 的基因體定序工作，並作為專利保護之鑑別依據。

本研究計畫中，我們將進行首先進行台灣傳統藥典所載錄但未及登載 ITS/IGS 基因體資料之樣品，另協同主持人劉淑玲助理教授將進行台灣本土之草藥的 ITS/IGS 基因體資料建立，另協同主持人陳方周助理教授具備建立資料庫網

站之研究經驗，我們將借重網路資料庫的連結，達到將台灣傳統中草藥基因體基原資料庫之建置。

中草藥的鑑別，在於辨別藥材的真偽、摻雜和品質的優劣，用以保證藥材的療效。因為誤用混用的偽藥或劣藥，不但不能治好病，反而會誤病甚至害人。故本研究首先進行常用市售藥材基因體基原鑑別，建立其科學化的鑑別機制，期能對未來相似藥材之鑑定，能進行更快速及準確的比對，以提供中草藥基原鑑定一個較完整可行之參考資料，計畫中亦建立網路資料庫查詢機制，將結合 Genbank 資料進行完整的台灣傳統藥材核酸基因體資料庫。

貳、材料與方法

(一) 植物 DNA 抽取^[18]

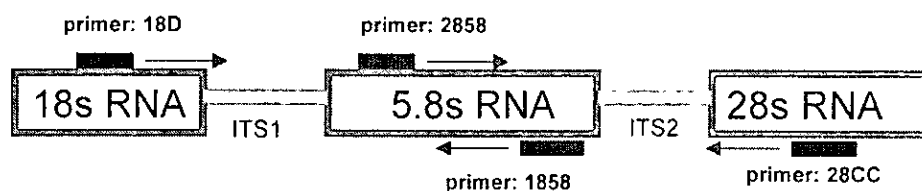
1. 磨碎植物 sample 成細粉狀，取量接近 0.1 mL 即可。
2. 加入 extraction buffer 800 μ L (EB, 100 mM Tris-HCl pH 8.0, 50 mM EDTA-2Na pH 8.0 500 mM NaCl, 10 mM beta-mercaptoethanol), 及 120 μ L 20% SDS, 再加入適量的 PVP, 強力搖晃 50 次後, 接著以 65°C 水浴(或乾浴) 10 min。
3. 加入 400 μ L of potassium acetate solution, 強力搖晃 50 次使之混合後冰浴 20 min。
4. Centrifuge 12 krpm at 4°C for 20 min, 抽上清液 500 μ L, 加入 1 ml isopropanol 使之沉澱, 然後搖晃使其均勻即可(不需要強力搖晃), 接著放入-20°C 靜置 5-10min
5. Centrifuge 12 krpm at 4°C for 15 min, 去除上清液, 倒立 10 min, 放在紙上使之乾燥。
6. 加入 500 μ L TE buffer (Buffer B, TE=5:1)會有懸浮物產生。加入 500 μ L 之 phenol。
7. 混合均勻後離心 (10 krpm 4°C for 5min)
8. 取上清液到新 eppendorf, 加入 0.25 mL 的 chloroform (chloroform : isoamylalcohol=24 : 1)
9. 再加入 0.25 ml 之 phenol, 混合均勻後離心(10 krpm 4°C for 5min)
10. 取上清液到新 eppendorf, 加入 0.5 mL 的 chloroform, 混合均勻後離心 (10 krpm 4°C for 5min)
11. 取上清液到新 eppendorf, 加入 0.9 ml 的絕對酒精和 0.15 ml of ammonium acetate (7.5 M)
12. 混合液放入-20°C for 30 min 使之增加沉澱。
13. 離心 12 krpm at 4°C for 15 min, 去上清液, 倒立乾燥 5~10 min(此步驟可直接加入 70% EtOH)
14. Rewash by used the 70% EtOH, 放置室溫下 15 min
15. 離心 10 krpm at 4°C for 5 min, 去上清液後再次使之乾燥

16. 加入 TE buffer (10 : 1) 0.2mL
17. 已分光光度計測定 OD260 之吸光值，以定量 DNA 濃度與純度
18. 將待測樣品放入 -20°C 中保存備用。

(二) 藥材基原鑑定-PCR, 一步驟法

藥材基源之鑑別乃是利用核糖體基因之 Internal transcribed spaces (ITS), ITS1 與 ITS2 作為基原鑑定之依據，所使用之引子係參考 Fu RZ et al.^[18] 所發表之論文並適當之修飾，5' 端引子：(18D)：5'-CACACCGCCCGTCGCTCCTACCGA-3'，3' 端引子：(28CC)：5'-ACTCGCCGTTACTAGGTG AA-3'，PCR 放大之反應條件為：25 μ l 反應中包含 15 ng 模板 DNA，10 mM Tris-HCl (pH 8.3)，50 mM KCl，1.0 μ M each primer，0.2 mM dNTP，2.0 mM MgCl₂，1.0 U Klen *Taq* DNA polymerase；反應循環為：one cycle of 94°C for 5 min, 50°C for 1 min and 72°C for 2 min; 40 cycles of 94°C for 1 min, 60°C for 1 min and 72°C for 1.5 min with a final extension of 72°C for 10 min。

經 PCR 擴增 ITS1/ITS2 序列後以 1%瓊脂電泳(agarose electrophoresis)進行長度判別，經分離約 800 bp 之片段以 spin column 進行純化動作，再交由合作廠商進行以 Sanger 法之定序工作。經重覆至少三次不同採集地點之同種藥材，經序列比對無誤後，並於自然科學博物館寄存標本樣品後，於 Genbank 登錄序列。



(四)、藥材基原鑑定-PCR, 兩步驟法

由於許多藥材經曬乾、烘乾、炮製等加工處理後，或染色體 DNA 抽取技術等因素，基因體序列常造成斷裂現象，所以當上述一步驟方法無法同時擴增 ITS1/ITS2 序列時，我們即利用兩步驟分段擴增法，第一組引子由 18D 與 1858 組成，用來擴增 ITS1 序列，第二組引子由 2858 與 28CC 組成，用來擴增 ITS2 序列，PCR 放大之反應條件為：25 μ l 反應中包含 15 ng 模板 DNA，10 mM Tris-HCl (pH 8.3)，50 mM KCl，1.0 μ M each primer，0.2 mM dNTP，2.0 mM MgCl₂，1.0 U Klen *Taq* DNA polymerase；反應循環為：one cycle of 94°C for 5 min, 50°C for 1 min and 72°C for 2 min; 40 cycles of 94°C for 1 min, 60°C for 1 min and 72°C for 1.5 min with a final extension of 72°C for 10 min。

經 PCR 擴增 ITS1/ITS2 序列後以 2%瓊脂電泳(agarose electrophoresis)進行長度判別，經分離約 500 bp 之片段以 spin column 進行純化動作，再交由合作廠商進行以 Sanger 法之定序工作。經重覆至少三次不同採集地點之同種藥材，經序列比對無誤後，並於自然科學博物館寄存標本樣品後，於 Genbank 登

錄序列。

(五)、電腦網路資料庫之建立

由製藥工業技術發展中心及工研院生物醫學工程中心指導的『藥材 e 藥網』^[19]將中藥材做過詳盡的整理，並以網頁呈現，是建構中草藥材電子資料庫的一個典範。透過藥材 e 藥網可以直接查詢到藥材的名稱、基原、產地、組織鑑定、化學成分等相關資料，甚至 HPLC 指紋圖譜也可由網站中獲得，但此網站有關 DNA 的鑑別資料有待補充。我們將台灣傳統藥典與中醫藥委員會委託試驗之台灣易混用、誤用之中草藥資料庫等計畫中，進行種屬名植物之 ITS/IGS 等資料之搜尋工作，並已建立相關檔案資料，將交由陳方周助理教授，將以現有資料進行登錄與網站之建置，待新登錄 Genbank 資料完成後，隨即增加於網站資料庫中以供查詢，即可以補充『藥材 e 藥網』的資料。

台灣中草藥基因體資料庫網站的網址，在測試過程中將附屬於中國醫藥大學的網域內，將主機建立在高速個人電腦中，以 Linux 為執行作業系統，安裝 Apache 網路伺服軟體以及 MySQL 資料庫，可搭配 PHP 以完成動態網頁的製作。網站分為管理階層以及檢視階層兩類。管理階層以密碼做為管制，在此階層中可以對資料庫進行編修，也可對網路流通進行監控。而檢視階層可以智慧式的查詢指令，查詢藥材的相關訊息，也可直接連接到藥材 e 藥網查詢相關資訊。

(六)、與傳統一般基源鑑定之比較

在我們之前的研究中顯示^[16]，以 ITS 基因體序列分析結果所繪製之演化圖譜與植物基源檢索表所列不盡相同，我們將在結果討論中加入 Flora of Taiwan^[20]與 eFlora.org^[21]所列資料之植物基源檢索表進行討論，比較其不同，我們在台灣產大戟科油柑屬藥材資源之基因體種源鑑別研究成果中已進行探討，並積極投稿中，本計畫將遵循相同模式進行分析比較，相信對於基源鑑定的工作可以提供一詳細的參考資料。

參、結果

(一)、定序分析：

期末成果已完成五十種藥材的定序工作，已確認的藥材已陸續發表至 NCBI 的 GenBank，已完成之藥材列表與發表到 Genbank 之範例如下：

編號	學名(中)	學名(英)	ITS1	ITS2
1	蒙古黃耆	<i>Astragalus mongholicus</i>	✓	✓
2	膜莢黃耆	<i>Astragalus membranaceus</i>	✓	✓
3	多序岩黃耆	<i>Hedysarum polybotrys</i>	✓	✓
4	川牛膝	<i>Cyathula officinalis</i>	✓	✓
5	懷牛膝	<i>Achyranthes bidentata</i>	✓	✓
6	菘藍 (北板藍)	<i>Isatis indigotica</i>	✓	✓
7	馬藍 (南板藍)	<i>Strobilanthes cusia</i>	✓	✓
8	麥藍菜	<i>Vaccaria segetalis</i>	✓	✓
9	野牡丹	<i>Melastoma candidum</i>	✓	✓
10	柳葉白前	<i>Cynanachum stauntonii</i>	✓	✓
11	蒲公英	<i>Taraxacum mongolicum</i>	✓	✓
12	兔耳菜 (小公英)	<i>Lactuca chinensis</i>	✓	✓
13	小飛揚草	<i>Euphorbia thymifolium</i>	✓	✓
14	黃花敗醬	<i>Patrinia scabiosaefolia</i>	✓	✓
15	白花敗醬	<i>Patrinia villosa</i>	✓	✓
16	何首烏	<i>Fallopia multiflora</i>	✓	✓
17	黃藥子	<i>Dioscorea bulbifera</i>	✓	✓
18	大丁黃	<i>Euonymus laxiflorus</i>	✓	✓
19	偽大丁黃 (黃金桂)	<i>Cudrania cochinchinensis var. gerontogea</i>	✓	✓
20	梔子	<i>Gardenia jasminoides</i>	✓	✓
21	白薇	<i>Cynanchum atratum</i>	✓	✓
22	細柱五加	<i>Acanthopanax gracilistylus</i>	✓	✓
23	白花蛇舌草	<i>Hedyotis diffusa</i>	✓	✓
24	關木通	<i>Aristolochia manshuriensis</i>	✓	✓
25	雞冠花(偽青箱)	<i>Euphorbia thymifolia</i>	✓	✓
26	歐菘藍	<i>Isstis tinctoria</i>	✓	✓
27	台灣馬藍	<i>Strobilanthes formosanus</i>	✓	✓
28	早蓮草	<i>Eclipta prostrata</i>	✓	✓
29	何首烏	<i>Fallopia multiflora</i>	✓	✓
30	黃藥子	<i>Dioscorea bulbifera</i>	✓	✓
31	沙苑子	<i>Astragalus complanatus</i>	✓	✓
32	升麻	<i>Cimicifuga foetida</i>	✓	✓
33	紫蘇	<i>Perilla frutescens</i>	✓	✓
34	三裂葉豚草	<i>Ambrosia trifida</i>	✓	✓
35	銀膠菊	<i>Parthenium hysterophorus</i>	✓	✓
36	魚腥草	<i>Houttuynia cordata</i>	✓	✓

37	豨薟	<i>Siegesbeckia orientalis</i>	✓	✓
38	昭和草	<i>Erechtites hierccifolia</i>	✓	✓
39	飛機草	<i>Erechtites valerianifolia</i>	✓	✓
40	大花咸豐草	<i>Bidens alba</i>	✓	✓
41	一支香	<i>Vernonia cinerea</i>	✓	✓
42	艾草	<i>Artemisia indica</i>	✓	✓
43	白頂飛蓬	<i>Erigeron annuus</i>	✓	✓
44	福山氏飛蓬	<i>Erigeron morrisonensis var. fukuyamae</i>	✓	✓
45	秋鼠麴草	<i>Gnaphalium hypoleucum</i>	✓	✓
46	鼠麴草	<i>Gnaphalium luteoalbum L.Subsp.affine</i>	✓	✓
47	龍葵	<i>Solanum nigrum</i>	✓	✓
48	光果龍葵	<i>Solanum americanum</i>	✓	✓
49	地膽草(毛蓮菜)	<i>Elephantopus mollis</i>	✓	✓
50	玉山抱莖籜簫	<i>Anaphalis morrisonicola</i>	✓	✓
51	尼泊爾籜簫	<i>Anaphalis nepalensis</i>	✓	✓

發表於 GenBank 的資料其查詢後檔案如下所示：

範例一：懷牛膝 (*Achyranthes bidentata*)

LOCUS DQ813300 778 bp DNA linear PLN 02-AUG-2006
DEFINITION *Achyranthes bidentata* CCMP95RD025-004 18S ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene, and internal transcribed spacer 2, complete sequence; and 26S ribosomal RNA gene, partial sequence.
ACCESSION DQ813300
VERSION DQ813300.1 GI:110826415
KEYWORDS .
SOURCE *Achyranthes bidentata*
ORGANISM *Achyranthes bidentata*
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; Caryophyllales; Amaranthaceae; *Achyranthes*.
REFERENCE 1 (bases 1 to 778)
AUTHORS Hsieh,C.-C., Chang,Y.-S., Kuo,C.-L., Liu,S.-L., Chiu,T.-H. and Chen,F.-J.
TITLE The genomic ITS sequence of nuclear ribosomal DNA identification and construction of the bioinformatics database of pharmaceutical botany in Taiwan
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 778)
AUTHORS Hsieh,C.-C., Chang,Y.-S., Kuo,C.-L., Liu,S.-L., Chiu,T.-H. and Chen,F.-J.
TITLE Direct Submission
JOURNAL Submitted (23-JUN-2006) Graduate Institute of Integration Chinese and Western Medicine, China Medical University, No. 91 Hsueh-Shih Rd., Taichung 404, Taiwan

```

FEATURES             Location/Qualifiers
    source            1..778
                     /organism="Achyranthes bidentata"
                     /mol_type="genomic DNA"
                     /specimen_voucher="CCMP95RD025-004"
                     /db_xref="taxon:384659"
                     /tissue_type="root"
    rRNA              <1..120
                     /product="18S ribosomal RNA"
    misc_RNA          121..351
                     /note="internal transcribed spacer 1"
    rRNA              352..508
                     /product="5.8S ribosomal RNA"
    misc_RNA          509..708
                     /note="internal transcribed spacer 2"
    rRNA              709..>778
                     /product="26S ribosomal RNA"

```

ORIGIN

```

1  cgcggcgacg tgggcggttc gccgccggcg acgtcgcgag aagttcactg aaccttatca
61  tttagaggaa ggagaagtgc taacaagggt tccgtagggtg aacctgcgga aggatcattg
121  tcgaaacctg cctagcagaa tgaccagcga acatgtttac atattgcatg gggaggggggt
181  actggcttgt ccagtccttc cctaagtgtg gggagttccc ccttgattgg tggtgctgcc
241  caacacaata acgaaccccg gcgtgatatg cgccaaggaa caaaaatgag agtgtgctta
301  tccttactcg gatttccggg tgaggatggt ggcacccaat ctaagtcatt aaatgactct
361  cggcaacgga tatctcggct ctgcacgca tgaagaacgt agcgaaatgc gatacttggg
421  gtgaattgca gaatcccgtg aaccatcgag tttttgaacg caagttgcgc ctgaagcctt
481  ttggccaagg cacgtctgcc tgggagtcac gcatagecgc tctccccacc tccaaagtgt
541  ggagggggaga ggaagatggc ctcccatgcc tcaccgggtg tggatggcct aaattaggaa
601  gcctcgggat acgagatgcc gcggcgattg gtggttgtat acatgggcct tccctcgtgt
661  cgtgcatcac gtagcccatg gggcctcgta ggacccttaa aaacctttgc gaccccaggt
721  caggcgggggt taccgcctga gtttaagcat atcaataagc ggaggaaaga aacttaca

```

//

範例二：歐菘藍 (*Isatis tinctoria*)

```

LOCUS      DQ813301                807 bp    DNA        linear    PLN 02-AUG-2006
DEFINITION Isatis tinctoria CCMP95RD025-010 18S ribosomal RNA gene, partial
sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene,
and internal transcribed spacer 2, complete sequence; and 26S
ribosomal RNA gene, partial sequence.
ACCESSION  DQ813301
VERSION    DQ813301.1  GI:110826416
KEYWORDS   .
SOURCE     Isatis tinctoria
  ORGANISM Isatis tinctoria
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;

```

rosids; eurosids II; Brassicales; Brassicaceae; Isatis.

REFERENCE 1 (bases 1 to 807)

AUTHORS Hsieh,C.-C., Chang,Y.-S., Kuo,C.-L., Liu,S.-L., Chiu,T.-H.,
Chen,F.-J. and Shih,W.-C.

TITLE The genomic ITS sequence of nuclear ribosomal DNA identification
and construction of the bioinformatics database of pharmaceutical
botany in Taiwan

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 807)

AUTHORS Hsieh,C.-C., Chang,Y.-S., Kuo,C.-L., Liu,S.-L., Chiu,T.-H.,
Chen,F.-J. and Shih,W.-C.

TITLE Direct Submission

JOURNAL Submitted (23-JUN-2006) Graduate Institute of Integration Chinese
and Western Medicine, China Medical University, No. 91 Hsueh-Shih
Rd., Taichung 404, Taiwan

FEATURES Location/Qualifiers

source 1..807
/organism="Isatis tinctoria"
/mol_type="genomic DNA"
/specimen_voucher="CCMP95RD025-010"
/db_xref="taxon:161756"
/tissue_type="root"

rRNA <1..120
/product="18S ribosomal RNA"

misc_RNA 121..390
/note="internal transcribed spacer 1"

rRNA 391..548
/product="5.8S ribosomal RNA"

misc_RNA 549..739
/note="internal transcribed spacer 2"

rRNA 740..>807
/product="26S ribosomal RNA"

ORIGIN

```

1 cgcggcgacg tgggtgggtc gccgtctgcg acgtcgcgag aagtccacta aaccttatca
61 tttagaggaa ggagaagtcg taacaagggt tccgtaggtg aacctgcgga aggatcattg
121 tcgatacctt atcgtaaaca gaacgaccog cgaacgattg atcatcactc tcggtgggct
181 ggtgtcttag ctgattccgt gcctgccgat tccgtgggta tgcgcgtggt ctcagccaag
241 attcatatct cggttgggtc atgcccctag ctcccgata tcaccaaac ccggcacgaa
301 aagtgtcaag gaacatgcaa ctaaacagcc tgcgttcgcc taccgggaga cgggtgttgc
361 gtggacgctg tgctgcaatc taaagtetaa aacgactetc ggcaacggat atcteggetc
421 tcgcatcgat gaagaacgta gcaaatgcg atacttgggt tgaattgcag aatcccgtga
481 accatcgagt ctttgaacgc aagtigcgcc ctaagccttc tggccgaggg cacgtctgcc
541 tgggtgtcac aaatcgctgt cccccatcc tctcgaggat aatggacgga agctgggtctc
601 cegtgtgtta ccgcacgcgg ttggccaaaa tccgagctaa ggacgcaagg agcgtctcga
661 catgcggtgg tgaattaaaa cctcgtcata ccgttggccg ctctgtctct gatgctctcg
721 atgacccaaa gtctcaacg cgaccccagg tcaggcggga tcaccgctg agtttaagca
781 tatcaataag cggaggaaag aactaca

```


//

正在發表中的資料如下所示：

範例一：台灣馬藍 (*Strobilanthes formosanus*)

>From gb-admin@ncbi.nlm.nih.gov

=====> bankit855397

Submission 1 of a total of 1 submission(s).

Comment: changchihsieh@hotmail.com

--To be released immediately after processing--

LOCUS bankit855397 823 bp DNA linear 09-NOV-2006
DEFINITION *Strobilanthes formosanus* 18S ribosomal RNA gene, partial sequence;
internal transcribed spacer 1, 5.8S ribosomal RNA gene, and
internal transcribed spacer 2, complete sequence; and 26S ribosomal
RNA gene, partial sequence.
ACCESSION 855397
VERSION
KEYWORDS
SOURCE *Strobilanthes formosanus*
ORGANISM *Strobilanthes formosanus*
Unclassified.
REFERENCE 1 (bases 1 to 823)
AUTHORS Hsieh,C.-C., Chang,Y.-S., Kuo,C.-L., Liu,S.-L., Chiu,T.-H.,
Chen,F.-J. and Shih,W.-C.
TITLE The genomic ITS sequence of nuclear ribosomal DNA identification
and construction of the bioinformatics database of pharmaceutical
botany in Taiwan
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 823)
AUTHORS Hsieh,C.-C., Chang,Y.-S., Kuo,C.-L., Liu,S.-L., Chiu,T.-H.,
Chen,F.-J. and Shih,W.-C.
TITLE Direct Submission
JOURNAL Submitted (09-NOV-2006) Graduate Institute of Integrated Medicine,
China Medical University, No. 91, Hsueh-Shih Rd., Taichung 404,
Taiwan
COMMENT Bankit Comment: changchihsieh@hotmail.com.
FEATURES Location/Qualifiers
source 1..823
/organism="Strobilanthes formosanus"
/mol_type="genomic DNA"
/specimen_voucher="CCMP95RD025-602"

rRNA 1..120
 /note="18S ribosomal RNA, partial sequence"
 misc_RNA 121..363
 /note="internal transcribed spacer 1"
 rRNA 364..520
 /product="5.8S ribosomal RNA"
 misc_RNA 521..751
 /note="internal transcribed spacer 2"
 rRNA 752..823
 /note="26S ribosomal RNA, partial sequence"
 BASE COUNT 161 a 262 c 261 g 139 t
 ORIGIN

```

1  tgagccgacg tgggcggttc gctgcccgcg acgtctcgag aagtccatta aaccttatca
61  tttagaggaa ggagaagtcg taacaagggt tccgtaggtg aacctgcgga aggatcattg
121 tgcaccctaa aacgagcaga ccgcgaacgc gttcaacaaa acgtggggggg ccgtcgtgcg
181 ggggtgggaa ttcccctccc gggcacgtcc cccgaccccc ggcggcgcgt gcgccgctcg
241 ggctaacgaa caccggcgcg gaaggcgcca aggaaaacat aaacgaagcg ttcgccccct
301 cgccccgcga ccgtgcgcgg tgcgcgctcg ggcggggggtg gccggacgcc tcttgaactt
361 tagaacgact ctcggaacg gatatctcgg ctctcgcatc gatgaagaac gtagcgaaat
421 gcgatacttg gtgtgaattg cagaatcccg tgaaccatcg gtttttgaac gcaagttagc
481 cccgaagcct tggggccgag ggcacgcctg cctggtcgcc acgaatcccg tcgccccctt
541 ccctcgtgct ccctaaccgg gtccgacggc cgagggcggt gcggagattg gcctcccgtg
601 cgccccctggc gcgcggccgg cttgaatgcg atccctcggc ggcgcccgtc gcgacaagtg
661 gtggatgact cctcagctgt ccctgtcgcg ctccgtggcg tcgtcctgac gggcatcacg
721 aacgaccoga gcgcacccgt gcgcaccgat tgcggcccca ggtcaggcgg ggattaccgg
781 ctgagtttaa gcatatcaat aagcggagga aaaagaaaca tac //
  
```

範例二：魚腥草 (*Houttuynia cordata*)

>From gb-admin@ncbi.nlm.nih.gov

=====> bankit855383

Submission 1 of a total of 1 submission(s).

Comment: changchihsieh@hotmail.com

--To be released immediately after processing--

LOCUS bankit855383 833 bp DNA linear PLN 09-NOV-2006
 DEFINITION *Houttuynia cordata* 18S ribosomal RNA gene, partial sequence;
 internal transcribed spacer 1, 5.8S ribosomal RNA gene, and
 internal transcribed spacer 2, complete sequence; and 26S ribosomal
 RNA gene, partial sequence.
 ACCESSION 855383
 VERSION

KEYWORDS .

SOURCE Houttuynia cordata

ORGANISM Houttuynia cordata
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; magnoliids; Piperales; Saururaceae;
Houttuynia.

REFERENCE 1 (bases 1 to 833)

AUTHORS Hsieh,C.-C., Chang,Y.-S., Kuo,C.-L., Liu,S.-L., Chiu,T.-H.,
Chen,F.-J. and Shih,W.-C.

TITLE The genomic ITS sequence of nuclear ribosomal DNA identification
and construction of the bioinformatics database of pharmaceutical
botany in Taiwan

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 833)

AUTHORS Hsieh,C.-C., Chang,Y.-S., Kuo,C.-L., Liu,S.-L., Chiu,T.-H.,
Chen,F.-J. and Shih,W.-C.

TITLE Direct Submission

JOURNAL Submitted (09-NOV-2006) Graduate Institute of Integrated Medicine,
China Medical University, No. 91, Hsueh-Shih Rd., Taichung 404,
Taiwan

COMMENT Bankit Comment: changchihsieh@hotmail.com.

FEATURES Location/Qualifiers

source 1..833
/organism="Houttuynia cordata"
/mol_type="genomic DNA"
/specimen_voucher="CCMP95RD025-605"
/db_xref="taxon:16752"

rRNA 1..120
/note="18S ribosomal RNA, partial sequence"

misc_RNA 121..351
/note="internal transcribed spacer 1"

rRNA 352..509
/product="5.8S ribosomal RNA"

misc_RNA 510..763
/note="internal transcribed spacer 2"

snRNA 764..833
/note="26S ribosomal RNA, partial sequence"

BASE COUNT 179 a 216 c 249 g 189 t

ORIGIN

```

1 cgcggcgacg tgggcggttc gctgccggcg acgtcgcgag aagtccactg aaccttatca
61 tttagaggaa ggagaagtcg taacaagggt tccgtaggtg aacctgcgga aggatcattg
121 ttgatgcgta ccaaaacaag atcgaccgaa gcgaacatgt gacccttgt tctttgcttg
181 cgtgcgagga ttgcegtctc ggtggctctc cgagcgcgtc gggcaccaac gaacaaaate
241 ccggcgcagt ccgcgccaaag gaattttctt aattgattgt ggcggtctcg tcgcccgttt
301 gcgccttgtg cgtcggcggg tgcgagttgg cgcacaaac tataatgtca atacgactct
361 cggcaacgga tatctcggct ctgcgcatcga igaagaacgt agcgaatgc gatacttggt
421 gtgaattgca gaatcccgtg aaccatcgag tctttgaacg caagttgcgc ccgaggcctt

```

```

481 ttaggttgag ggcacatctg cttgggcgtc aaacatcacg tcgctcccca caccactgcc
541 caacgggcag acaaggatgg gtttgcggag attggccgtc cgagtgcctc gagcacgcgg
601 ttggctgaaa agctctggcc tttggttgca cgcggtcaa caagtgggtg ttgtgggctc
661 ttttgggccg cgattgtcgg gatgttgtgt cgtgccttgc ccgtgttggc cagcgaggac
721 cctattcgat cgacctccgc actccggagg cacgaatcag attgcgaccc caagtcaggt
781 gggatcaccg getgaattta agcatatcaa taagcggagg aaaagaaact tac //

```

(二)、序列分析

定序完成之後的分析都利用我們在計畫中依照中草藥基因體分析的特點所設計的 Perl 程式進行分析，其分析的程式碼如下：

1. 以檔案陣列進行分析：ccmpfile.pl

```

#DNA sequences analysis program
#本程式承蒙行政院衛生署中醫藥委員會，計畫編號 CCMP95-RD-025 提供經費補助，使本程
式得以順利完成，特此致謝。
#以下是輸入檔案的地方
print "Please key in the filename in Genbank accession number array:\n";
$filearray=<STDIN>;
open FILE, "<$filearray";
while (<FILE>){
$array = $_;
}
#由 GenBank 下載序列資料以進行分析
@acn=split(/./,$array);
for ($an=0;$an<$#acn+1;$an++){
$acc1=$acn[$an];
use Bio::DB::GenBank;
use Bio::SeqIO::GenBank

$db = new Bio::DB::GenBank;
$seqobj = $db->get_Seq_by_acc("$acc1");
$ann_coll = $seqobj->annotation;
$string2=$acc1;
mkdir "/genbank";
mkdir "/genbank/$acc1";
open (OUTPUT,">/genbank/$acc1/$string2-annotation.txt");
for $ann ($ann_coll->get_Annotations) {
print OUTPUT $ann->as_text if ($ann->>tagname eq "comment");
}

$seqobj2 = $seqobj->seq();
open (OUTPUT,">/genbank/$acc1/$string2-sequence.txt");
print OUTPUT "$seqobj2";

$seqobj3 = $seqobj->desc();
open (OUTPUT,">/genbank/$acc1/$string2-description.txt");
print OUTPUT "$seqobj3";

```

```

#Sequence analysis and database output
$All=$seqobj2=~tr/ATCG//;
open (OUTPUT,">>/genbank/$acc1/$string2-analysis.txt");
print OUTPUT "nts\t$All\n";
@nucleotides=('A','T','G','C');
for ($k=0;$k<4;$k++){@k=$seqobj2=~m/$nucleotides[$k]/gi;
    $mono_number=scalar(@k);
    $mono_percentage=($mono_number/$All)*100;
    $monopercentage=sprintf("%.1f%", $mono_percentage);
    push @nucleotide1, "$monopercentage";
    push @DNAbase, "$nucleotides[$k]=$monopercentage";
    print OUTPUT "$nucleotides[$k]\t$mono_percentage\n";}

```

```

#ITS Mononucleotides analysis and pie figure output
use GD;
use GD::Graph::pie;
my $graph=GD::Graph::pie->new(550,550);
$graph->set(title=>"$string2 ITS Mononucleotides analysis",
    start_angle=>6,
    '3d'=>0,
    label =>"Mononucleotide composition",
    transparent=>0,
    x_all_ticks=>1,
    suppress_angle=>5);
$graph->set_title_font(gdGiantFont);
$graph->set_label_font(gdGiantFont);
$graph->set_value_font(gdGiantFont);
$image=$graph->plot([[@DNAbase],[@nucleotide1]]) or die $graph->error;
open (f1,">/genbank/$acc1/$string2-nucleotides.png");
binmode (f1);
print f1 $image->png;
print "@ING";
@DNAbase = ();
@nucleotide1 = ();

```

```

#ITS Dinucleotides analysis and bar figure output
open (OUTPUT,">>/genbank/$acc1/$string2-analysis.txt");
@dinucleotides=('AA','AG','AC','AT','TT','TA','TG','TC','CC','CA','CT','CG','GG','GA','GT','GC');
for ($l=0;$l<16;$l++){@l=$seqobj2=~m/$dinucleotides[$l]/gi;
    $di_number=scalar(@l);
    $di_All=$di_All+$di_number;}
for ($i=0;$i<16;$i++){@i=$seqobj2=~m/$dinucleotides[$i]/gi;
    $di_number1=scalar(@i);
    $di_percentage=($di_number1/$di_All)*100;
    $dipercentage=sprintf("%.1f%", $di_percentage);
    push @dinucleotide1, "$dipercentage";
    print OUTPUT "$dinucleotides[$i]\t$di_percentage\n"; }
@di_nucleotide=sort {$b<=>$a}@dinucleotide1;
my $ymax=$di_nucleotide[0]+1;
use GD;
use GD::Graph::bars;
my $graph=GD::Graph::bars->new(800,600);
$graph->set(show_values =>"@dinucleotide1",

```

```

transparent=>0,
title=>"$string2 ITS Dinucleotides analysis",
x_label=>'Dinucleotides composition',
x_label_position=>1/2,
y_label=>'Percentage(%)',
x_all_ticks=>1,
y_all_ticks=>0,
y_all_numbers=>0,
zero_axis=>0,
zero_axis_only=>0,
bar_spacing=>5,
y_max_value=>"$ymax",
y_long_ticks=>1,
y_tick_length=>"$xmax",
bgclr=>'white',
boxclr=>'lgray',
borderclrs=>[qw(black)],
dclrs=>[qw(lred
        lgreen
        lblue
        lyellow
        dbrown
        lpurple
        cyan
        orange
        )],
cycle_clrs=>1,
shadowclr=>'dgray',
shadow_depth=>2
);
$graph->set_title_font(gdGiantFont);
$graph->set_x_label_font(gdGiantFont);
$graph->set_y_label_font(gdGiantFont);
$graph->set_x_axis_font(gdLargeFont);
$graph->set_y_axis_font(gdLargeFont);
$imager=$graph->plot([[@dinucleotides],[@dinucleotide1]]) or die $graph->error;
open(f2,">/genbank/$acc1/$string2-dinucleotides.png");
binmode (f2);
print f2 $imager->png;
print "@ING";
$di_All = "";
@dinucleotides = ();
@dinucleotide1 = ();

```

#ITS Trinucleotides analysis and bar figure output

```

open (OUTPUT,">>/genbank/$acc1/$string2-analysis.txt");
@trinucleotides=('TTT','TTC','TTA','TTG','TCT','TCC','TCA','TCG','TAT','TAC','TAA','TAG','TGT','
TGC','TGA','TGG','CTT','CTC','CTA','CTG','CCT','CCC','CCA','CCG','CAT','CAC','CAA','CAG','CG
T','CGC','CGA','CGG','ATT','ATC','ATA','ATG','ACT','ACC','ACA','ACG','AAT','AAC','AAA','AAG'
.'AGT','AGC','AGA','AGG','GTT','GTC','GTA','GTG','GCT','GCC','GCA','GCG','GAT','GAC','GAA','
GAG','GGT','GGC','GGA','GGG');
for ($n=0;$n<64;$n++){@n=$seqobj2=~m/$trinucleotides[$n]/gi;
        $tri_number=scalar(@n);

```

```

        $tri_All=$tri_All+$tri_number;}
for ($j=0;$j<64;$j++){@j=$seqobj2=~m/$trinucleotides[$j]/gi;
        $tri_number1=scalar(@j);
        $tri_percentage=(($tri_number1/$tri_All)*100;
        $stripercentage=sprintf("%.1f%", $tri_percentage);
        print OUTPUT "$trinucleotides[$j]\t$tri_percentage\n";}

@trinucleotides1=('TTT','TTC','TTA','TTG','TCT','TCC','TCA','TCG','TAT','TAC','TAA','TAG','TGT',
'TGC','TGA','TGG');
for ($o=0;$o<16;$o++){@o=$seqobj2=~m/$trinucleotides1[$o]/gi;
        $tri_number2=scalar(@o);
        $tri_percentage1=(($tri_number2/$tri_All)*100;
        $stripercentage1=sprintf("%.1f%", $tri_percentage1);
        push @trinucleotide, "$stripercentage1";}

@tri_nucleotide=sort{$b<=>$a}@trinucleotide;
my $ymax1=$tri_nucleotide[0]+1;
use GD;
use GD::Graph::bars;
my $graph=GD::Graph::bars->new(800,600);
$graph->set(show_values =>"@trinucleotide",
        transparent=>0,
        title=>"$string2 ITS Trinucleotides(T) percentage data",
        x_label=>'Trinucleotides composition',
        x_label_position=>1/2,
        y_label=>'Percentage(%)',
        x_all_ticks=>1,
        y_all_ticks=>0,
        y_all_numbers=>0,
        zero_axis=>0,
        zero_axis_only=>0,
        bar_spacing=>5,
        y_max_value=>"$ymax1",
        y_long_ticks=>1,
        y_tick_length=>"$xmax",
        bgcolor=>'white',
        boxclr=>'lgray',
        borderclrs=>[qw(black)],
        dclrs=>[qw(lred
                lgreen
                lblue
                lyellow
                dbrown
                lpurple
                cyan
                orange
                )],
        cycle_clrs=>1,
        shadowclr=>'dgray',
        shadow_depth=>2
        );
$graph->set_title_font(gdGiantFont);
$graph->set_x_label_font(gdGiantFont);
$graph->set_y_label_font(gdGiantFont);

```

```

$graph->set_x_axis_font(gdLargeFont);
$graph->set_y_axis_font(gdLargeFont);
$image=$graph->plot([[@trinucleotides1],[@trinucleotide]]) or die $graph->error;
open (f3, ">/genbank/$acc1/$string2-trinucleotides(T).png");
binmode (f3);
print f3 $image->png;
print "@ING";
@trinucleotides1 = ();
@trinucleotide = ();

@trinucleotides2=('CTT','CTC','CTA','CTG','CCT','CCC','CCA','CCG','CAT','CAC','CAA','CAG','CGT','CGC','CGA','CGG');
for ($q=0;$q<16;$q++){@q=$seqobj2=~m/$trinucleotides2[$q]/gi;
    $stri_number3=scalar(@q);
    $stri_percentage2=($stri_number3/$stri_All)*100;
    $stripercentage2=sprintf("%.1f%", $stri_percentage2);
    push @trinucleotide1, "$stripercentage2";}
@tri_nucleotide1=sort{$b<=>$a}@trinucleotide1;
my $ymax2=$tri_nucleotide1[0]+1;
use GD;
use GD::Graph::bars;
my $graph=GD::Graph::bars->new(800,600);
$graph->set(show_values =>"@trinucleotide1",
    transparent=>0,
    title=>"$string2 ITS Trinucleotides(C) percentage data",
    x_label=>'Trinucleotides composition',
    x_label_position=>1/2,
    y_label=>'Percentage(%)',
    x_all_ticks=>1,
    y_all_ticks=>0,
    y_all_numbers=>0,
    zero_axis=>0,
    zero_axis_only=>0,
    bar_spacing=>5,
    y_max_value=>"$ymax2",
    y_long_ticks=>1,
    y_tick_length=>"$xmax",
    bgclr=>'white',
    boxclr=>'lgray',
    borderclrs=>[qw(black)],
    dclrs=>[qw(lred
        lgreen
        lblue
        lyellow
        dbrown
        lpurple
        cyan
        orange
    )],
    cycle_clrs=>1,
    shadowclr=>'dgray',
    shadow_depth=>2
);

```



```

$graph->set_title_font(gdGiantFont);
$graph->set_x_label_font(gdGiantFont);
$graph->set_y_label_font(gdGiantFont);
$graph->set_x_axis_font(gdLargeFont);
$graph->set_y_axis_font(gdLargeFont);
$imagedata=$graph->plot([[@trinucleotides2],[@trinucleotide1]]) or die $graph->error;
open (f4,">/genbank/$acc1/$string2-trinucleotides(C).png");
binmode (f4);
print f4 $imagedata->png;
print "@ING";
@trinucleotides2 = ();
@trinucleotide1 = ();

@trinucleotides3=('ATT','ATC','ATA','ATG','ACT','ACC','ACA','ACG','AAT','AAC','AAA','AAG','A
GT','AGC','AGA','AGG');
for ($s=0;$s<16;$s++){ @s=$seqobj2->get($s)/gi;
    $stri_number4=scalar(@s);
    $stri_percentage3=($stri_number4/$stri_All)*100;
    $stripercentage3=sprintf("%.1f%", $stri_percentage3);
    push @trinucleotide2, "$stripercentage3";}
@tri_nucleotide2=sort{$b<=>$a}@trinucleotide2;
my $ymax3=$tri_nucleotide2[0]+1;
use GD;
use GD::Graph::bars;
my $graph=GD::Graph::bars->new(800,600);
$graph->set(show_values =>"@trinucleotide2",
    transparent=>0,
    title=>"$string2 ITS Trinucleotides(A) percentage data",
    x_label=>'Trinucleotides composition',
    x_label_position=>1/2,
    y_label=>'Percentage(%)',
    x_all_ticks=>1,
    y_all_ticks=>0,
    y_all_numbers=>0,
    zero_axis=>0,
    zero_axis_only=>0,
    bar_spacing=>5,
    y_max_value=>"$ymax3",
    y_long_ticks=>1,
    y_tick_length=>"$xmax",
    bgcolor=>'white',
    boxclr=>'lgray',
    borderclrs=>[qw(black)],
    dclrs=>[qw(lred
        lgreen
        lblue
        lyellow
        dbrown
        lpurple
        cyan
        orange
    )],
    cycle_clrs=>1,

```

```

shadowclr=>'dgray',
shadow_depth=>2
);
$graph->set_title_font(gdGiantFont);
$graph->set_x_label_font(gdGiantFont);
$graph->set_y_label_font(gdGiantFont);
$graph->set_x_axis_font(gdLargeFont);
$graph->set_y_axis_font(gdLargeFont);
$image=$graph->plot([[@trinucleotides3],[@trinucleotide2]]) or die $graph->error;
open (f5, ">/genbank/$acc1/$string2-trinucleotides(A).png");
binmode (f5);
print f5 $image->png;
print "@ING";
@trinucleotides3 = ();
@trinucleotide2 = ();

@trinucleotides4=('GTT','GTC','GTA','GTG','GCT','GCC','GCA','GCG','GAT','GAC','GAA','GAG','G
GT','GGC','GGA','GGG');
for ($u=0;$u<16;$u++){@u=$seqobj2=~m/$Strinucleotides4[$u]/gi;
    $stri_number5=scalar(@u);
    $stri_percentage4=($stri_number5/$stri_All)*100;
    $stripercentage4=sprintf("%.1f%", $stri_percentage4);
    push @trinucleotide3, "$stripercentage4";}
@tri_nucleotide3=sort{$b<=>$a}@trinucleotide3;
my $ymax4=$tri_nucleotide3[0]+1;
use GD;
use GD::Graph::bars;
my $graph=GD::Graph::bars->new(800,600);
$graph->set(show_values =>"@trinucleotide3",
    transparent=>0,
    title=>"$string2 ITS Trinucleotides(G) percentage data",
    x_label=>'Trinucleotides composition',
    x_label_position=>1/2,
    y_label=>'Percentage(%)',
    x_all_ticks=>1,
    y_all_ticks=>0,
    y_all_numbers=>0,
    zero_axis=>0,
    zero_axis_only=>0,
    bar_spacing=>5,
    y_max_value=>"$ymax4",
    y_long_ticks=>1,
    y_tick_length=>"$xmax",
    bgcolor=>'white',
    boxclr=>'lgray',
    borderclrs=>[qw(black)],
    dclrs=>[qw(lred
        lgreen
        lblue
        lyellow
        dbrown
        lpurple
        cyan

```

```

        orange
    ]],
cycle_clr=>1,
shadow_clr=>'dgray',
shadow_depth=>2
);
$graph->set_title_font(gdGiantFont);
$graph->set_x_label_font(gdGiantFont);
$graph->set_y_label_font(gdGiantFont);
$graph->set_x_axis_font(gdLargeFont);
$graph->set_y_axis_font(gdLargeFont);
$imagedata=$graph->plot([[@trinucleotides4],[@trinucleotide3]]) or die $graph->error;
open (f6,">/genbank/$acc1/$string2-trinucleotides(G).png");
binmode (f6);
print f6 $imagedata->png;
print "@ING";
$tri_All = "";
@trinucleotides4 = ();
@trinucleotide3 = ();

#Restriction digest and subsequent gel run
use Bio::Seq;
use Bio::Tools::RestrictionEnzyme;
use Bio::Tools::Gel;
use GD;
use GD::Graph;
use GD::Graph::points;
@restriction_name=('Ladder','AvaII','BamHI','BsiI','BsmFI','EcoRI','EcoRV','HaeII','HaeIII','HhaI','HindIII');
@Laddermolecular=('2000','1000','900','800','700','600','500','400','300','200','100','50','25','10');
for($g=0;$g< $#Laddermolecular+1;$g++){push @Ladder,log10($g);}
sub log10{1.362529288*(10*log($Laddermolecular[$g])/log(10)+10)-13.60279288};
$new_band_NB=scalar(@Ladder);
push @number_data,"$new_band_NB";
for($d=1;$d< $#restriction_name+1;$d++){$seq1 = Bio::Seq->new(-id=>'groundhog
day',-seq=>$seqobj2);

        $enzyme =
Bio::Tools::RestrictionEnzyme->new(-NAME=>$restriction_name[$d]);
        @cuts= $enzyme->cut_seq($seq1);
        $gel = Bio::Tools::Gel->new(-seq=>\(@cuts,-dilate=>10);
        %d = $gel->bands;
        @new_bands=sort values(%d);
        $new_band_NB1=scalar(@new_bands);
        push @number_data,"$new_band_NB1";
        for($f=0;$f< $#new_bands+1;$f++){push
@NEWband,1.362529288*(50-"$new_bands[$f]")-13.60279288;}
        push @NewBands,[@NEWband];
        splice(@NEWband,0);
        }

@max=sort {$b<=>$a} @number_data;
$max1=$max[0];
@data=([@restriction_name]);
for($t=0;$t<$max1;$t++){push @t,"$Ladder[$t]";

```

```

        for($e=0;$e< $#restriction_name;$e++){push @t,"$NewBands[$e][$t]";}
        push @data,[@t];
        splice(@t,0);
    }

$long=80*12+150;
$graph = GD::Graph::points->new($long,800);

sub y_format{$value = shift;
    $ret;
    if ($value>=0){$ret = sprintf("%.f",exp(($value-0.0225)/13.62529288*log(10)));}
}

$graph->set(zero_axis=>0,
zero_axis_only=> 0,
y_tick_number => 10,
y_number_format=>\&y_format,
transparent=>0,
x_label_position=>1/2,
boxclr=>'black',
dclrs=>[qw(white)],
x_label => 'Restriction endonuclease',
y_label => 'DNA fragment length, bp',
title => $string2.' - Restriction map, gel electrophoresis-1',
);
$graph->set_title_font(gdGiantFont);
$graph->set_x_label_font(gdGiantFont);
$graph->set_y_label_font(gdGiantFont);
$graph->set_x_axis_font(gdLargeFont);
$graph->set_y_axis_font(gdLargeFont);
$graph->set( markers => [9] );
$graph->set( marker_size => 30 );
$graph->plot(\@data);
open(OUT, ">/genbank/$acc1/$string2-electrophoresis-1.png") or die $!;
binmode OUT;
print OUT $graph->gd->png;
@NewBands=();
@data=();
close OUT;

@restriction_name1=('Ladder','HinI','HinPI','KpnI','MspI','SmaI','SthI32I','TaqI','TspRI','XbaI','XhoI');
@Laddermolecular=('2000','1000','900','800','700','600','500','400','300','200','100','50','25','10');
for($g=0;$g< $#Laddermolecular+1;$g++){push @Ladder1,log10($g);}
sub log10{1.362529288*(10*log($Laddermolecular[$g])/log(10)+10)-13.60279288};
$new_band_NB1=scalar(@Ladder1);
push @number_data1,"$new_band_NB1";
for($d1=1;$d1< $#restriction_name1+1;$d1++){$seq2 = Bio::Seq->new(-id=>'groundhog day',-seq=>$seqobj2);

    $enzyme =
Bio::Tools::RestrictionEnzyme->new(-NAME=>$restriction_name1[$d1]);
    @cuts1 = $enzyme->cut_seq($seq2);
    $gel =

Bio::Tools::Gel->new(-seq=>\@cuts1.-dilate=>10):

```

```

                                %d1 = $gel->bands;
                                @new_bands1=sort values(%d1);
                                $new_band_NB2=scalar(@new_bands1);
                                push @number_data1,"$new_band_NB2";
                                for($f1=0;$f1<$#new_bands1+1;$f1++){push
@NEWband1,1.362529288*(50-"$new_bands1[$f1]")+13.60279288;}
                                push @NewBands1,[@NEWband1];
                                splice(@NEWband1,0);
                                }

@max1=sort{$b<=>$a}@number_data1;
$max2=$max1[0];
@data1=({@restriction_name1});
for($t1=0;$t1<$max2;$t1++){push @t1,"$Ladder1[$t1]";
                                for($e1=0;$e1<$#restriction_name1;$e1++){push
@t1,"$NewBands1[$e1][$t1]";}
                                push @data1,[@t1];
                                splice(@t1,0);
                                }

$long=80*12+150;
$graph = GD::Graph::points->new($long,800);

sub y_format{$value = shift;
    $ret;
    if ($value>=0){$ret = sprintf("%.f",exp(($value-0.0225)/13.62529288*log(10)))};
    }}

$graph->set(zero_axis=>0,
zero_axis_only=> 0,
y_tick_number => 10,
y_number_format=>\&y_format,
transparent=>0,
x_label_position=>1/2,
boxclr=>'black',
dclrs=>[qw(white)],
x_label => 'Restriction endonuclease',
y_label => 'DNA fragment length, bp',
title => $string2.' - Restriction map, gel electrophoresis-2',
);
$graph->set_title_font(gdGiantFont);
$graph->set_x_label_font(gdGiantFont);
$graph->set_y_label_font(gdGiantFont);
$graph->set_x_axis_font(gdLargeFont);
$graph->set_y_axis_font(gdLargeFont);
$graph->set( markers => [9] );
$graph->set( marker_size => 30 );
$graph->plot(\@data1);
open(OUT, ">/genbank/$acc1/$string2-electrophoresis-2.png") or die $!;
binmode OUT;
print OUT $graph->gd->png;
@NewBands1=();
@data1=();
close OUT;
print $string2." analysis OK!\n";

```

```
}
```

```
print "Thx use this program\n":
```

2. 以 accession number 輸入進行查詢分析：ccmpacc.pl

```
#DNA sequences analysis program
#本程式承蒙行政院衛生署中醫藥委員會，計畫編號 CCMP95-RD-025 提供經費補助，使本程
式得以順利完成，特此致謝。
#以下是輸入 Accession 的地方

#DNA sequences analysis program
print "Please key in the Accession number:\n";
$acc1=<STDIN>;
chomp($acc1);
use Bio::DB::GenBank;
use Bio::SeqIO::GenBank

$db = new Bio::DB::GenBank;
$seqobj = $db->get_Seq_by_acc("$acc1");
$ann_coll = $seqobj->annotation;
$string2=$acc1;
mkdir "/genbank/$acc1";
open (OUTPUT,">/genbank/$acc1/$string2-ann.txt");
for $ann ($ann_coll->get_Annotations) {
print OUTPUT $ann->as_text if ($ann->>tagname eq "comment");
}

$seqobj2 = $seqobj->seq();
open (OUTPUT,">/genbank/$acc1/$string2-seq.txt");
print OUTPUT "$seqobj2";
$All=$seqobj2=~tr/ATCG//;
open (OUTPUT,">>/genbank/$acc1/$string2-analysis.txt");
print OUTPUT "nts\t$All\n";
@nucleotides=('A','T','G','C');
for ($k=0;$k<4;$k++){ @k=$seqobj2=~m/$nucleotides[$k]/gi;
    $mono_number=scalar(@k);
    $mono_percentage=($mono_number/$All)*100;
    $monopercentage=sprintf("%.1f%", $mono_percentage);
    push @nucleotide1, "$monopercentage";
    push @DNAbase, "$nucleotides[$k]=$monopercentage";
    print OUTPUT "$nucleotides[$k]\t$mono_percentage\n";}

use GD;
use GD::Graph::pie;
my $graph=GD::Graph::pie->new(550,550);
$graph->set(title=>"$string2 ITS Mononucleotides analysis",
    start_angle=>6,
    '3d'=>0,
    label =>"Mononucleotide composition",
    transparent=>0,
    x_all_ticks=>1,
    suppress_angle=>5);
```

```

$graph->set_title_font(gdGiantFont);
$graph->set_label_font(gdGiantFont);
$graph->set_value_font(gdGiantFont);
$image=$graph->plot([[@DNAbase],[@nucleotide1]]) or die $graph->error;
open (f1,">/genbank/$acc1/$string2-nucleotides.png");
binmode (f1);
print f1 $image->png;
print "@ING";

open (OUTPUT,">>/genbank/$acc1/$string2-analysis.txt");
@dinucleotides=('AA','AG','AC','AT','TT','TA','TG','TC','CC','CA','CT','CG','GG','GA','GT','GC');
for ($l=0;$l<16;$l++){@l=$seqobj2=~m/$dinucleotides[$l]/gi;
    $di_number=scalar(@l);
    $di_All=$di_All+$di_number;}
for ($i=0;$i<16;$i++){@i=$seqobj2=~m/$dinucleotides[$i]/gi;
    $di_number1=scalar(@i);
    $di_percentage=($di_number1/$di_All)*100;
    $dipercentage=sprintf("%.1f%",$di_percentage);
    push @dinucleotide1,"$dipercentage";
    print OUTPUT "$dinucleotides[$i]\t$di_percentage\n"; }

@di_nucleotide=sort{$b<=>$a}@dinucleotide1;
my $ymax=$di_nucleotide[0]+1;
use GD;
use GD::Graph::bars;
my $graph=GD::Graph::bars->new(800,600);
$graph->set(show_values=>"@dinucleotide1",
transparent=>0,
title=>"$string2 ITS Dinucleotides analysis",
x_label=>'Dinucleotides composition',
x_label_position=>1/2,
y_label=>'Percentage(%)',
x_all_ticks=>1,
y_all_ticks=>0,
y_all_numbers=>0,
zero_axis=>0,
zero_axis_only=>0,
bar_spacing=>5,
y_max_value=>"$ymax",
y_long_ticks=>1,
y_tick_length=>"$xmax",
bgclr=>'white',
boxclr=>'lgray',
borderclrs=>[qw(black)],
dclrs=>[qw(lred
    lgreen
    lblue
    lyellow
    dbrown
    lpurple
    cyan
    orange
)],
cycle_clrs=>1,

```

```

shadowclr=>'dgray',
shadow_depth=>2
);
$graph->set_title_font(gdGiantFont);
$graph->set_x_label_font(gdGiantFont);
$graph->set_y_label_font(gdGiantFont);
$graph->set_x_axis_font(gdLargeFont);
$graph->set_y_axis_font(gdLargeFont);
$imager=$graph->plot([[@dinucleotides],[@dinucleotide1]]) or die $graph->error;
open(f2,">/genbank/$acc1/$string2-dinucleotides.png");
binmode (f2);
print f2 $imager->png;
print "@ING";

open (OUTPUT,">>/genbank/$acc1/$string2-analysis.txt");
@trinucleotides=('TTT','TTC','TTA','TTG','TCT','TCC','TCA','TCG','TAT','TAC','TAA','TAG','TGT','
TGC','TGA','TGG','CTT','CTC','CTA','CTG','CCT','CCC','CCA','CCG','CAT','CAC','CAA','CAG','CG
T','CGC','CGA','CGG','ATT','ATC','ATA','ATG','ACT','ACC','ACA','ACG','AAT','AAC','AAA','AAG'
,'AGT','AGC','AGA','AGG','GTT','GTC','GTA','GTG','GCT','GCC','GCA','GCG','GAT','GAC','GAA','
GAG','GGT','GGC','GGA','GGG');
for ($n=0;$n<64;$n++){@n=$seqobj2=~m/$trinucleotides[$n]/gi;
    $stri_number=scalar(@n);
    $stri_All=$stri_All+$stri_number;}
for ($j=0;$j<64;$j++){@j=$seqobj2=~m/$trinucleotides[$j]/gi;
    $stri_number1=scalar(@j);
    $stri_percentage=(($stri_number1/$stri_All)*100;
    $stripercentage=sprintf("%.1f%", $stri_percentage);
    print OUTPUT "$trinucleotides[$j]\t$stri_percentage\n";}

@trinucleotides1=('TTT','TTC','TTA','TTG','TCT','TCC','TCA','TCG','TAT','TAC','TAA','TAG','TGT'
'TGC','TGA','TGG');
for ($o=0;$o<16;$o++){@o=$seqobj2=~m/$trinucleotides1[$o]/gi;
    $stri_number2=scalar(@o);
    $stri_percentage1=(($stri_number2/$stri_All)*100;
    $stripercentage1=sprintf("%.1f%", $stri_percentage1);
    push @trinucleotide, "$stripercentage1";}

@tri_nucleotide=sort{$b<=>$a}@trinucleotide;
my $ymax1=$stri_nucleotide{0}+1;
use GD;
use GD::Graph::bars;
my $graph=GD::Graph::bars->new(800,600);
$graph->set(show_values =>"@trinucleotide",
    transparent=>0,
    title=>"$string2 ITS Trinucleotides(T) percentage data",
    x_label=>'Trinucleotides composition',
    x_label_position=>1/2,
    y_label=>'Percentage(%)',
    x_all_ticks=>1,
    y_all_ticks=>0,
    y_all_numbers=>0,
    zero_axis=>0,
    zero_axis_only=>0,
    bar_spacing=>5,

```



```

        y_max_value=>"$ymax1",
        y_long_ticks=>1,
        y_tick_length=>"$xmax",
        bgcolor=>'white',
boxclr=>'lgray',
borderclrs=>[qw(black)],
dclrs=>[qw(lred
        lgreen
        lblue
        lyellow
        dbrown
        lpurple
        cyan
        orange
)],
cycle_clrs=>1,
shadowclr=>'dgray',
shadow_depth=>2
);
$graph->set_title_font(gdGiantFont);
$graph->set_x_label_font(gdGiantFont);
$graph->set_y_label_font(gdGiantFont);
$graph->set_x_axis_font(gdLargeFont);
$graph->set_y_axis_font(gdLargeFont);
$image=$graph->plot([[@trinucleotides1],[@trinucleotide]]) or die $graph->error;
open (f3,">/genbank/$acc1/$string2-trinucleotides(T).png");
binmode (f3);
print f3 $image->png;
print "@ING";

@trinucleotides2=('CTT','CTC','CTA','CTG','CCT','CCC','CCA','CCG','CAT','CAC','CAA','CAG','CG
T','CGC','CGA','CGG');
for ($q=0;$q<16;$q++){@q=$seqobj2=~m/$trinucleotides2[$q]/gi;
        $tri_number3=scalar(@q);
        $tri_percentage2=($tri_number3/$tri_All)*100;
        $stripercentage2=sprintf("%.1f%", $tri_percentage2);
        push @trinucleotide1, "$stripercentage2";}

@tri_nucleotide1=sort{$b<=>$a}@trinucleotide1;
my $ymax2=$tri_nucleotide1[0]+1;
use GD;
use GD::Graph::bars;
my $graph=GD::Graph::bars->new(800,600);
$graph->set(show_values =>"@trinucleotide1",
        transparent=>0,
        title=>"$string2 ITS Trinucleotides(C) percentage data",
        x_label=>'Trinucleotides composition',
        x_label_position=>1/2,
        y_label=>'Percentage(%)',
        x_all_ticks=>1,
        y_all_ticks=>0,
        y_all_numbers=>0,
        zero_axis=>0,
        zero_axis_only=>0,

```

```

        bar_spacing=>5,
        y_max_value=>"$ymax2",
        y_long_ticks=>1,
        y_tick_length=>"$xmax",
        bgcolor=>'white',
boxclr=>'lgray',
borderclrs=>[qw(black)],
dclrs=>[qw(lred
        lgreen
        lblue
        lyellow
        dbrown
        lpurple
        cyan
        orange
)],
cycle_clrs=>1,
shadowclr=>'dgray',
shadow_depth=>2
);
$graph->set_title_font(gdGiantFont);
$graph->set_x_label_font(gdGiantFont);
$graph->set_y_label_font(gdGiantFont);
$graph->set_x_axis_font(gdLargeFont);
$graph->set_y_axis_font(gdLargeFont);
$image=$graph->plot([[@trinucleotides2],[@trinucleotide1]]) or die $graph->error;
open (f4,">/genbank/$acc1/$string2-trinucleotides(C).png");
binmode (f4);
print f4 $image->png;
print "@ING";

@trinucleotides3=('ATT','ATC','ATA','ATG','ACT','ACC','ACA','ACG','AAT','AAC','AAA','AAG','A
GT','AGC','AGA','AGG');
for ($s=0;$s<16;$s++){ @s=$seqobj2=~m/$trinucleotides3[$s]/gi;
        $tri_number4=scalar(@s);
        $tri_percentage3=($tri_number4/$tri_All)*100;
        $stripercentage3=sprintf("%.1f%", $tri_percentage3);
        push @trinucleotide2, "$stripercentage3";}

@tri_nucleotide2=sort{$b<=>$a}@trinucleotide2;
my $ymax3=$tri_nucleotide2[0]+1;
use GD;
use GD::Graph::bars;
my $graph=GD::Graph::bars->new(800,600);
$graph->set(show_values =>"@trinucleotide2",
        transparent=>0,
        title=>"$string2 ITS Trinucleotides(A) percentage data",
        x_label=>'Trinucleotides composition',
        x_label_position=>1/2,
        y_label=>'Percentage(%)',
        x_all_ticks=>1,
        y_all_ticks=>0,
        y_all_numbers=>0,
        zero_axis=>0,

```

```

        zero_axis_only=>0,
        bar_spacing=>5,
        y_max_value=>"$ymax3",
        y_long_ticks=>1,
        y_tick_length=>"$xmax",
        bgcolor=>'white',
boxclr=>'lgray',
borderclrs=>[qw(black)],
dclrs=>[qw(lred
        lgreen
        lblue
        lyellow
        dbrown
        lpurple
        cyan
        orange
        )],
cycle_clrs=>1,
shadowclr=>'dgray',
shadow_depth=>2
    );
$graph->set_title_font(gdGiantFont);
$graph->set_x_label_font(gdGiantFont);
$graph->set_y_label_font(gdGiantFont);
$graph->set_x_axis_font(gdLargeFont);
$graph->set_y_axis_font(gdLargeFont);
$image=$graph->plot([[@trinucleotides3],[@trinucleotide2]]) or die $graph->error;
open (f5,">/genbank/$acc1/$string2-trinucleotides(A).png");
binmode (f5);
print f5 $image->png;
print "@ING";

@trinucleotides4=('GTT','GTC','GTA','GTG','GCT','GCC','GCA','GCG','GAT','GAC','GAA','GAG','G
GT','GGC','GGA','GGG');
for ($u=0;$u<16;$u++){@u=$seqobj2=~m/$trinucleotides4[$u]/gi;
        $stri_number5=scalar(@u);
        $stri_percentage4=($stri_number5/$stri_All)*100;
        $stripercentage4=sprintf("%.1f%", $stri_percentage4);
        push @trinucleotide3, "$stripercentage4";}
@tri_nucleotide3=sort{$b<=>$a}@trinucleotide3;
my $ymax4=$tri_nucleotide3[0]+1;
use GD;
use GD::Graph::bars;
my $graph=GD::Graph::bars->new(800,600);
$graph->set(show_values =>"@trinucleotide3",
        transparent=>0,
        title=>"$string2 ITS Trinucleotides(G) percentage data",
        x_label=>'Trinucleotides composition',
        x_label_position=>1/2,
        y_label=>'Percentage(%)',
        x_all_ticks=>1,
        y_all_ticks=>0,
        y_all_numbers=>0,

```

```

        zero_axis=>0,
        zero_axis_only=>0,
        bar_spacing=>5,
        y_max_value=>"$ymax4",
        y_long_ticks=>1,
        y_tick_length=>"$xmax",
        bgcolor=>'white',
boxclr=>'lgray',
borderclrs=>[qw(black)],
dclrs=>[qw(lred
        lgreen
        lblue
        lyellow
        dbrown
        lpurple
        cyan
        orange
)],
cycle_clrs=>1,
shadowclr=>'dgray',
shadow_depth=>2
);
$graph->set_title_font(gdGiantFont);
$graph->set_x_label_font(gdGiantFont);
$graph->set_y_label_font(gdGiantFont);
$graph->set_x_axis_font(gdLargeFont);
$graph->set_y_axis_font(gdLargeFont);
$imager=$graph->plot([[@trinucleotides4],[@trinucleotide3]]) or die $graph->error;
open (f6,">/genbank/$acc1/$string2-trinucleotides(G).png");
binmode (f6);
print f6 $imager->png;
print "@ING";

#Restriction digest and subsequent gel run
use Bio::Seq;
use Bio::Tools::RestrictionEnzyme;
use Bio::Tools::Gel;
use GD;
use GD::Graph;
use GD::Graph::points;
@restriction_name=('Ladder','AvaI','BamHI','BsiI','BsmFI','EcoRI','EcoRV','HaeII','HaeIII','HhaI','HindIII','HinfI','HinPII','HpaII','KpnI','MspI','SmaI','SthI32I','TaqI','TspRI','XbaI','XhoI');
@Laddermolecular=(2000,'1000','900','800','700','600','500','400','300','200','100','50','25','10');
for($g=0;$g< $#Laddermolecular+1;$g++){push @Ladder,log10($g);}
sub log10{1.362529288*(10*log($Laddermolecular[$g])/log(10)+10)-13.60279288};
$new_band_NB=scalar(@Ladder);
push @number_data,"$new_band_NB";
for($d=1;$d< $#restriction_name+1;$d++){$seq1 = Bio::Seq->new(-id=>'groundhog
day',-seq=>$seqobj2);
        $enzyme =
Bio::Tools::RestrictionEnzyme->new(-NAME=>$restriction_name[$d]);
        @cuts= $enzyme->cut_seq($seq1);
        $gel = Bio::Tools::Gel->new(-seq=>\@cuts,-dilate=>10);

```

```

                                %d = $gel->bands;
                                @new_bands=sort values(%d);
                                $new_band_NB1=scalar(@new_bands);
                                push @number_data,"$new_band_NB1";
                                for($f=0;$f<$#new_bands+1;$f++){push
@NEWband,1.362529288*(50-"$new_bands[$f]"-13.60279288);}
                                push @NewBands,[@NEWband];
                                splice(@NEWband,0);
                                }

    @max=sort{$b<=>$a}@number_data;
    $max1=$max[0];
    @data=([@restriction_name]);
    for($t=0;$t<$max1;$t++){push @t,"$Ladder[$t]";
                                for($e=0;$e<$#restriction_name;$e++){push @t,"$NewBands[$e][[$t]"];}
                                push @data,[@t];
                                splice(@t,0);
                                }

    $long=80*22+150;
    $graph = GD::Graph::points->new($long,800);

sub y_format{$value = shift;
    $ret;
    if ($value>=0){$ret = sprintf("%.f",exp(($value-0.0225)/13.62529288*log(10)));
    }}

    $graph->set(zero_axis=>0,
zero_axis_only=> 0,
y_tick_number => 10,
y_number_format=>\&y_format,
transparent=>0,
x_label_position=>1/2,
boxclr=>'black',
dclrs=>[qw(white)],
x_label => 'Restriction endonuclease',
y_label => 'DNA fragment length, bp',
title => $string2.' - Restriction map, gel electrophoresis',
    );
    $graph->set_title_font(gdGiantFont);
    $graph->set_x_label_font(gdGiantFont);
    $graph->set_y_label_font(gdGiantFont);
    $graph->set_x_axis_font(gdLargeFont);
    $graph->set_y_axis_font(gdLargeFont);
    $graph->set( markers => [9] );
    $graph->set( marker_size => 30 );
    $graph->plot(\@data);
    open(OUT, ">/genbank/$acc1/$string2-electrophoresis.png") or die $!;
    binmode OUT;
    print OUT $graph->gd->png;
    close OUT;
    print "Thx use this program\n";
    close INPUT;
    close OUTPUT;

```

(三)、中草藥基因體資訊網的建立與序列分析結果：

本計畫進行之中草藥基因體資料庫的建立與分析數據都已向學校電算中心申請到本計畫所使用的虛擬網址 tcmgdb.cmu.edu.tw(資料庫存放位址，圖一)與 bioinfo.cmu.edu.tw(序列分析與網頁存放位址，圖二)。在伺服器中已配備 MySQL 資料庫系統，並且使用 PHP 程式與資料庫系統互相支援，專門用來設計動態式網頁。本計畫的目標是為中藥材 ITS 基因體序列建立專屬的序列資料庫，須要大規模的資料存儲空間。然而在測試階段，資料庫的內容宜簡化，以利於在程式的執行階段進行除蟲作業(debug)。目前在 MySQL 系統中已建立 tcmgdb 資料庫，專為本計畫安排一個獨立的資料存儲空間。而在此資料庫僅加入三個資料表，其中 drugtb 資料表使用 drid、name、sname、web、ncbidb、gene 六個欄位，而 webtb 及 datatb 為兩個關連資料表，分別列出 NCBI 指向的網址及 NCBI 中所使用的資料庫名稱，我們進一步利用相關程式進行近緣相似物種之比對，並公告於網站中。

網頁的設計也是以簡單實用為現階段的主要考量，經點選或輸入興趣的藥材名稱，按下傳送按鈕即可將該藥材的名稱、學名、基因組態呈現，並使用超連結指向 NCBI 的資料庫。結果(圖三)將序列摘要與單核苷酸(圖四)、雙核苷酸(圖五)、三核苷酸(圖六~圖九)與 RFLP 虛擬電泳膠分析(圖十、圖十一)以懷牛膝為例標示出其分析結果。

肆、討論

本計畫以中藥飲片為鑑定對象，對於 genomic DNA 的萃取可能會因為中藥炮製與中藥材儲存的緣故，致使所抽取之 DNA 含有不完整之片段序列與微生物序列，大多是霉菌的污染，在試驗中有一半左右的定序資料無法使用，霉菌的污染問題一直影響著中草藥的品質，中醫藥委員會亦提供研究經費探討微生物的污染對於中藥品質的影響評估，由於基因體的分析可以分離到附著於藥材上的微生物，是否可作為進一步快速分析微生物污染與藥材品質的一項指標，所附著的微生物是否具有毒性，亦可由其分離微生物株的基因體分析中進一步確認，此項研究工作值得進一步探討。

80 年代初期，ribosomal DNA (rDNA)的結構漸漸明朗，由於它位於基因體中含有許多重複單位(拷貝，copy)，每個重複單位的長度約 7~13 kb，其結構包括 rDNA 編碼區域(coding region)，非編碼區(ITS, ETS)和非轉錄片段(NTS)組成，這三部份的變異幅度依次增加，而其重要性相對降低，因為編碼區域對物種的生存有重要的意義，當物種於演化的進程中存活下來後，其編碼區段的序列必然達成一個平衡並僅有極少的改變，由這些區段之重要度與物種之間的歧異度來選擇適當的分子標記來做為物種鑑別之基礎時，ITS 的序列分析便顯得非常重要，由於各物種之間的重複數目並不相同，rDNA 基因組在植物體中的 copy number 為 500~40,000 個，差異相當大，在進行藥材的基原鑑定時如果遇到 copy number 較低之物種時，再加上藥材的儲存條件不良，炮制的過程對細胞組織的破壞，都會影響 rDNA 的萃取與 PCR 擴增的結果，有了本次研究的經驗後我們發現，如果

可以建立各種藥材基因體的參考染色體 DNA，或許對於今後研究人員要進行藥材鑑識時的一個參照。

在考慮到藥材的微生物污染與炮製後之藥材基因體分析無法得到明確的答案時，建立完整的資料庫又刻不容緩，我們在期中報告之後希望以鮮品進行序列的確定，我們陸續收集藥材，並到各大藥園、山區或大陸地區尋找樣品以供染色體萃取，並進行 PCR 擴增與序列定序分析，並發表到 GenBank 中，以完整其資料庫。我們分析這些相進的混用、勿用或代用的藥材植物時發現，藥材的選用或許有地域性上的限制，如代用品的選擇有或許是因為該地區可以大量取得的物種，其藥理作用或許相近，如果可以經由藥材基因體的鑑別分析，瞭解其物種之相關演化，找到相近的物種，或許對於開發新穎的藥用植物，或許是一個快速的篩選方法，當然，在各物種間藥材的有效成份或許會因為功能性基因的相互作用而有所差異，但基因體分析在 PCR 技術與定序技術的支援下，未來或許對於功能性基因的分析會有更長足的進展，主持人就分析了常見藥材中的黃酮類成份常用來做為抗氧化、抗發炎與抗癌的相關研究，而推演黃酮類化合物的生合成可知道，許多的黃酮生成酵素參與黃酮類化合物的形成，如果對於這些一系列的酵素基因分子進行表現陣列分析(cDNA microarray)，或蛋白質體的分析，再高通量，高速度的篩選下，開發新穎藥材或天然物應該是一可行的方向。

伍、結論與建議

- 一、本計畫已逐漸完成相關基因體資料庫的基礎分析，由於藥材資源非常廣大，本土中草藥資源亦相當豐富，為避免混用與誤用中草藥資源，並確保國人用藥安全，建立此基本資料庫實為一重要工作，建議中醫藥委員會可以持續補助相關計畫，進一步將本土特有中草藥之基因體資料庫建置完全。
- 二、未來之相關研究計畫應可建立這些藥材的參考染色體備份，以提供未來研究人員如果須要比對資料時之參考，並與生醫所(工研院)討論，以避免重複之工作。
- 三、在各物種間藥材的有效成份或許會因為功能性基因的相互作用而有所差異，未來或許對於功能性基因的分析進行表現陣列分析(cDNA microarray)，或蛋白質體(Proteomic analysis)的分析，經過高通量，高速度的篩選下，開發新穎藥材或天然物應該是一可行的方向。

誌謝

本計畫承蒙行政院衛生署中醫藥委員會，計畫編號 CCMP95-RD-025 提供經費贊助，使本計畫得以順利完成，特此致謝。

陸、參考文獻

- [1]. 行政院衛生署中華藥典中藥集編修小組，中華中藥典，行政院衛生署，2004，台北
- [2]. 署授藥字第 0940004263 號，公告『自即日起，「中華中藥典」更名為「台灣傳統藥典」』行政院衛生署，2005/08/31
- [3]. Long C, Kakiuchi N, Takahashi A, Komatsu K, Cai S, Mikage M. Phylogenetic analysis of the DNA sequence of the non-coding region of nuclear ribosomal DNA and chloroplast of Ephedra plants in China. *Planta Med.* 2004 Nov;70(11):1080-4.
- [4]. Kiko H, Niggemann E, Ruger W. Physical mapping of the restriction fragments obtained from bacteriophage T4 dC-DNA with the restriction endonucleases SmaI, KpnI and BglIII. *Mol Gen Genet.* 1979;172(3):303-12.
- [5]. Fu RZ, Wang J, Zhang YB, Wang ZT, But PP, Li N, Shaw PC. Differentiation of medicinal *Codonopsis* species from adulterants by polymerase chain reaction-restriction fragment length polymorphism. *Planta Med.* 1999 Oct;65(7):648-50.
- [6]. Hon CC, Chow YC, Zeng FY, Leung FC. Genetic authentication of ginseng and other traditional Chinese medicine. *Acta Pharmacol Sin.* 2003 Sep;24(9):841-6.
- [7]. Williams JG, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.* 1990 Nov 25;18(22):6531-5.
- [8]. Thomas CM, Vos P, Zabeau M, Jones DA, Norcott KA, Chadwick BP, Jones JD. Identification of amplified restriction fragment polymorphism (AFLP) markers tightly linked to the tomato Cf-9 gene for resistance to *Cladosporium fulvum*. *Plant J.* 1995 Nov;8(5):785-94.
- [9]. Hearne CM, McAleer MA, Love JM, Aitman TJ, Cornall RJ, Ghosh S, Knight AM, Prins JB, Todd JA. Additional microsatellite markers for mouse genome mapping. *Mamm Genome.* 1991;1(4):273-82.
- [10]. Pellegrini M, Manning J, Davidson N. Sequence arrangement of the rDNA of *Drosophila melanogaster*. *Cell.* 1977 Feb;10(2):213-4.
- [11]. Gardes M, Bruns TD. ITS primers with enhanced specificity for basidiomycetes--application to the identification of mycorrhizae and rusts.

- Mol Ecol. 1993 Apr;2(2):113-8.
- [12].Liu JS, Schardl CL. A conserved sequence in internal transcribed spacer 1 of plant nuclear rRNA genes. Plant Mol Biol. 1994 Oct;26(2):775-8.
- [13].Sallares R, Allaby RG, Brown TA. PCR-based identification of wheat genomes. Mol Ecol. 1995 Aug;4(4):509-14.
- [14].http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?dbFrom=nucleotide&db=nucleotide&cmd=Search&cmd_current=Limits&orig_db=nucleotide&term=Cordyceps+sinensis, 2005/11.
- [15].蔡貴花、簡一治、謝長奇、張賢哲，保育類龜板應用 DNA 定序之鑑定，第 17 屆天然藥物研討會暨中草藥生物科技研討會，中國醫藥大學，2002/7/13，台中。
- [16].李昭瑩、謝長奇、朱俊男、邱泰惠，Identification of Phyllanthus species based on ITS sequence of nuclear ribosomal DNA in Taiwan，中國藥學會，臺北醫學大學藥學院，2004/12/18，台北。
- [17].http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?dbFrom=nucleotide&db=nucleotide&cmd=Search&cmd_current=Limits&orig_db=nucleotide&term=Lee+CY+AND+Hsieh+CC, 2005/11.
- [18].Fu RZ, Wang J, Sun YR, Shaw PC. Extraction of genomic DNA suitable for PCR analysis from dried plant rhizomes/roots. Biotechniques. 1998 Nov;25(5):796-8, 800-1.
- [19].<http://www.herbemed.org.tw/>, 2006/1/3
- [20].<http://tai2.ntu.edu.tw/fotdv/fotmain.htm>, 2006/1/3
- [21].<http://www.efloras.org/index.aspx>, 2006/1/3
- [22].<http://www.dcb.org.tw/adminz/index.asp>, 2006/1/3

柒、圖

圖一、基因體資料庫搜尋網頁，<http://tcmgdb.cmu.edu.tw/>，相關資料連結，建立搜尋與分析之功能的網站。

中醫藥材基因資料庫 - Microsoft Internet Explorer

檔案(F) 編輯(E) 檢視(V) 我的最愛(A) 工具(T) 說明(H)

網址(O) <http://tcmgdb.cmu.edu.tw/> 移至 上一頁 搜尋 連結

Google (G) 開始 9 已擱載 拼字檢查 傳送到 設定

中醫藥材基因資料庫 搜尋結果

阿膠

中藥名稱：阿膠

學名： [Abutilon theophrasti](#)

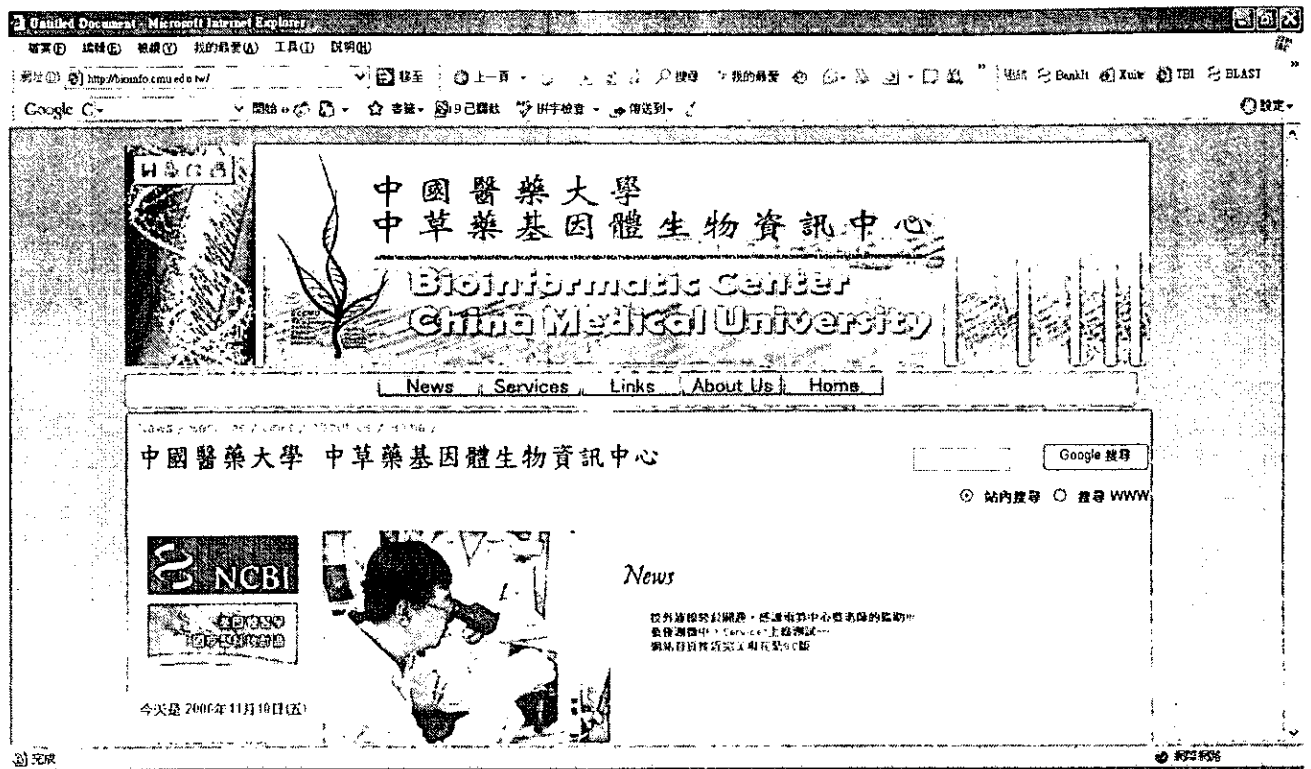
相關連結： [DQ006017](#)

基因：
aagtcgtaac aagggttccg taggtgaacc tgcggaagga 1cattgtcga aacctgccta
gcagaacgac ccglgaatgc gttatcatac aaaacaccga gaggggtcgg atgcaatatt
gccccaaccc ctctcgatgc ctctctcgt ttggtcttgc ctcatccac ctgtgaggg
lgagttgcaa ttcccalcca ctccaaggca aaaccaacaa cccccggcg gaa1tgcgc
aaggaaatta aalgaaaaga gtgcaacgtt actgtctccg acccgtcgc ggtgtttg
agggagtgc gctgttact ttgtctgaa aatacaaac gactctggc aacggatc
tcggctctc calcatgaa gaacgtagc aaatgcgata ctgtgttga attgcagaal
cccglgaacc atcgagctt lgaacgcaag ttgcgcccc agccaltagg ccgaggcac
gtctgcctgg gtgtcagca tcttgcctc calcaaacct ctagcttatt cggctcagg
caaattgtg gcggaalig gcttccctg tctcaccgt gcgcgctgg cctaaaaatg
agtttcggc gatgaagtc cgaacaatc ggtgggaatg ctctcagct cctcttca
agttgtgt gctctgat tgaacccta tgacctttt ggcatcaca ttttggct
cgatcgcga cccaggica ggcgggatt

公告暨資料搜尋

網際網路

圖二、中草藥基因體資訊網頁，<http://bioinfo.cmu.edu.tw>，中草藥基因體資料分析與結果展示網頁。



圖三、序列分析

Internal transcribed spacer (ITS) profile analysis system - Microsoft Internet Explorer

檔案(F) 編輯(E) 格式(O) 我的最愛(S) 工具(T) 說明(H)

地址: http://www.zcmu.edu.cn/~cchsh/ncgenbank/DQ813300/DQ813300.html

Google

BankIt BLAST

中國醫藥大學
中草藥基因體生物資訊中心
Bioinformatics Center
China Medical University

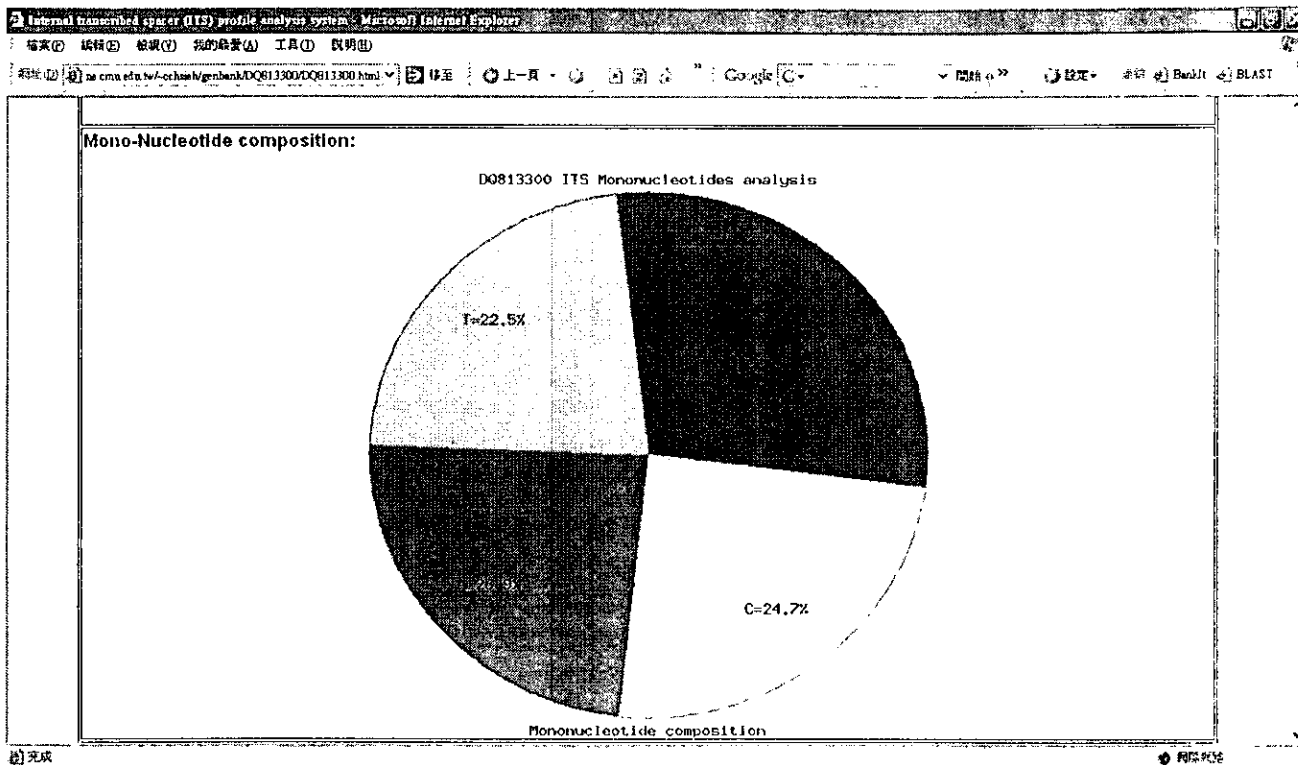
Achyranthes bidentata analysis profile

Species Name :
Achyranthes bidentata

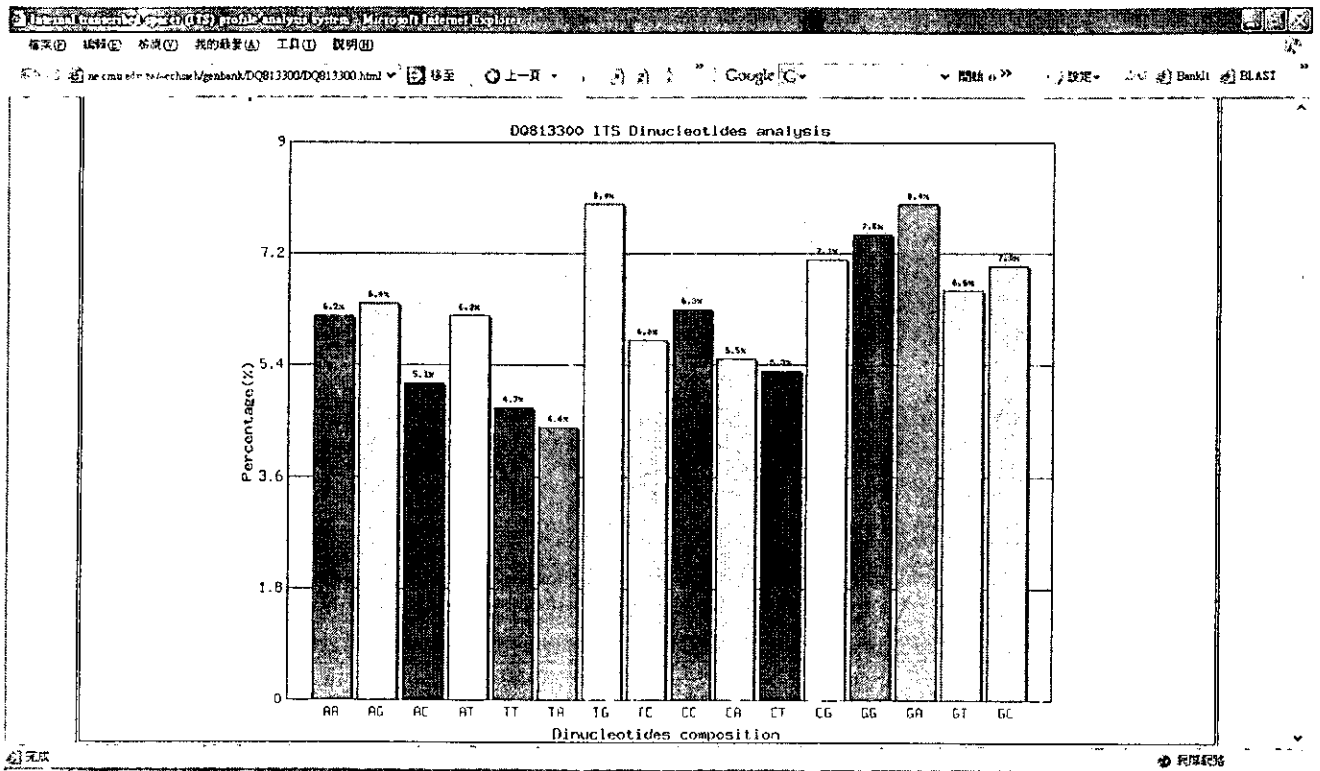
Genome list :

#	Herb	Accession number	Discription	size (bps)
1	懷牛膝	DQ813300	Achyranthes bidentata CCMP95RD025-004 18S ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene, and internal transcribed spacer 2, complete sequence; and 26S ribosomal RNA gene, partial sequence.	778

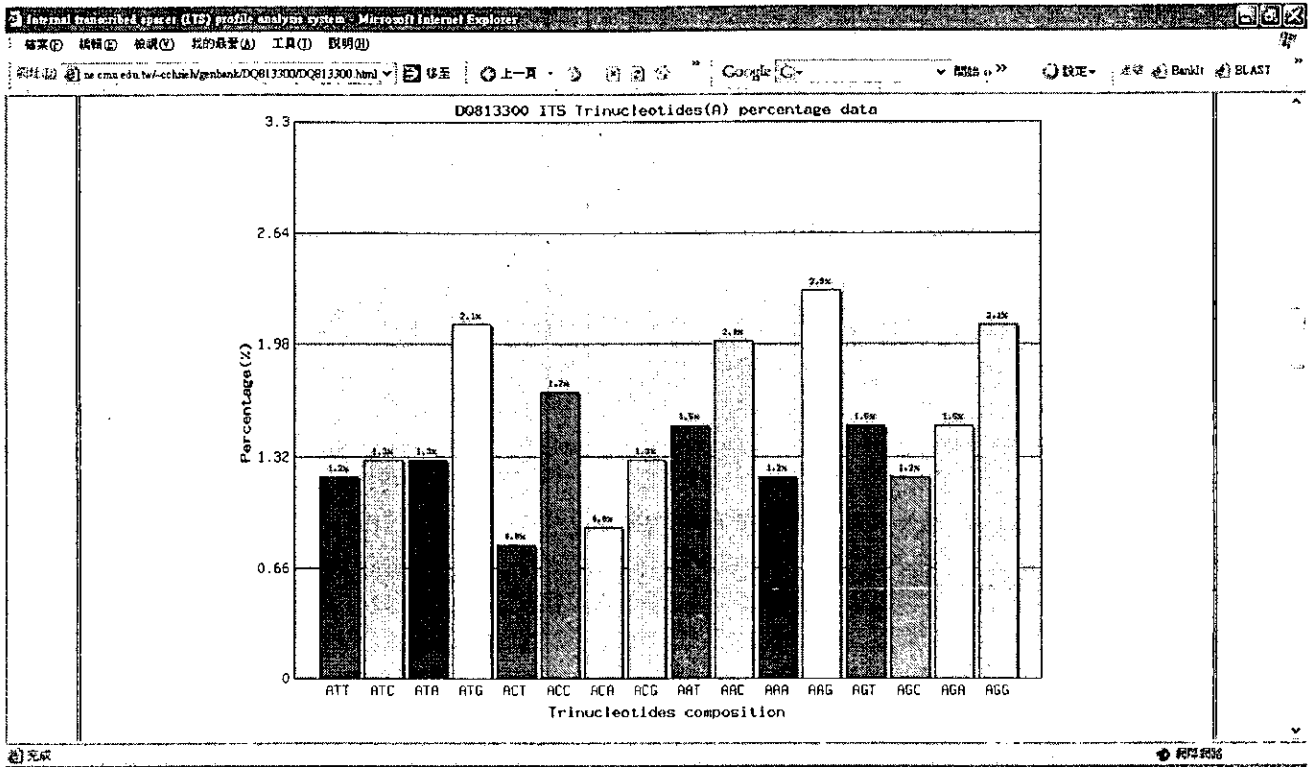
圖四、單核苷酸分析



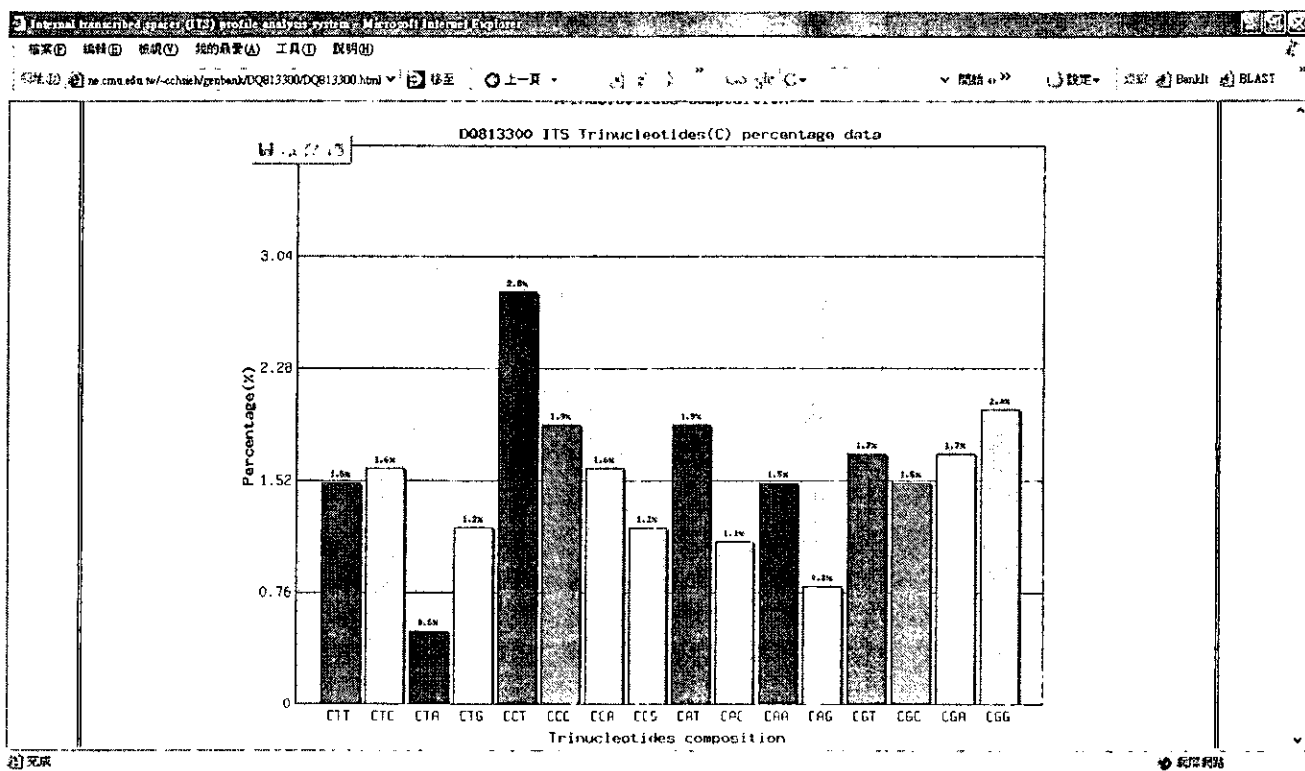
圖五、雙核苷酸分析



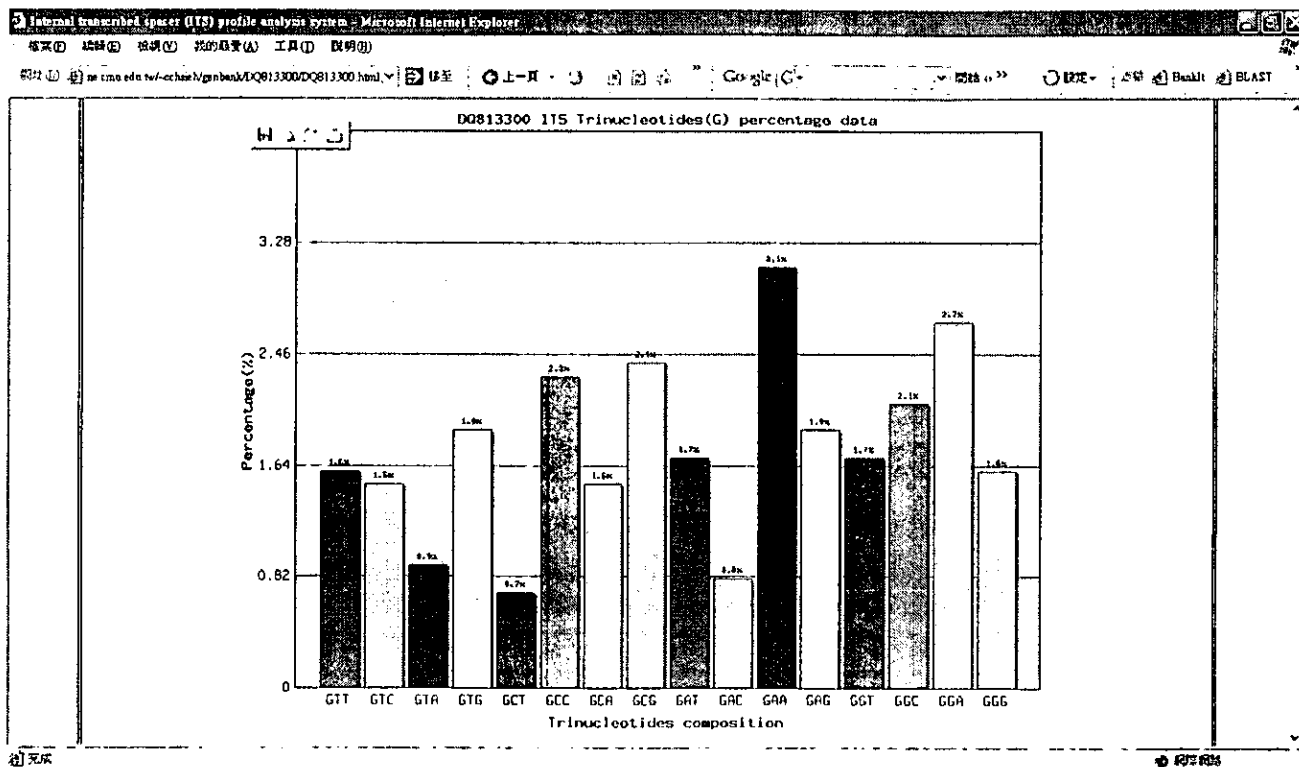
圖六、三核苷酸分析(A)



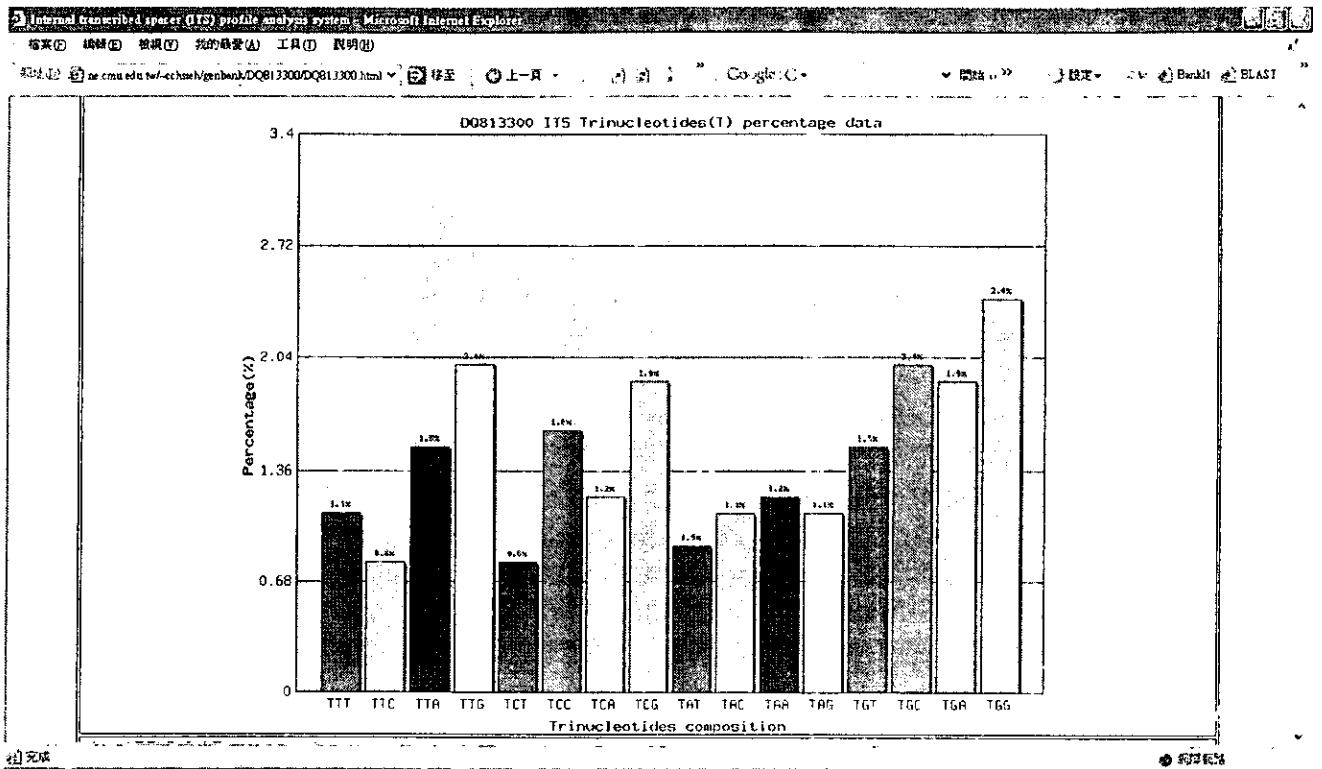
圖七、三核苷酸分析(C)



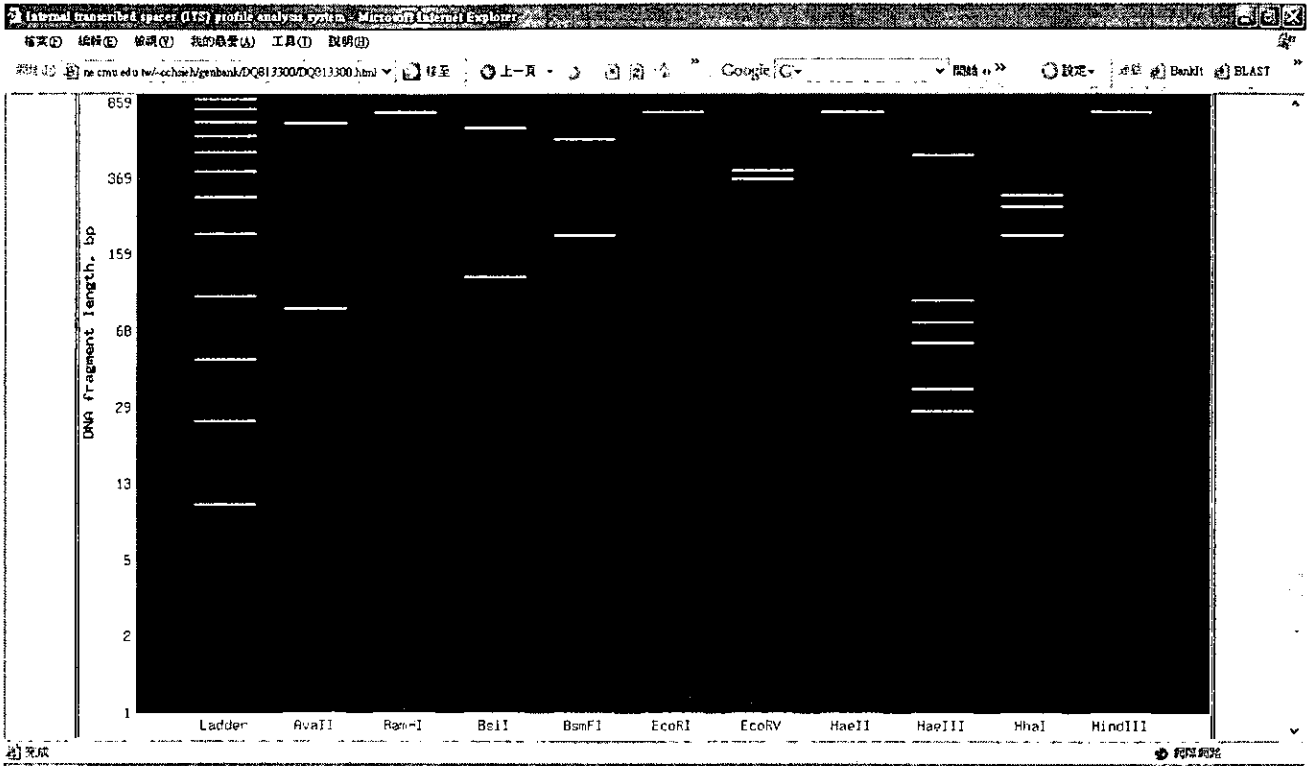
圖八、三核苷酸分析(G)



圖九、三核苷酸分析(T)



圖十、ITS-RFLP 虛擬電泳膠分析(一)：



圖十一、ITS-RFLP 虛擬電泳膠分析(二)：

