

PAST統計軟體

於DNA分子標誌多樣性分析之應用

▲ 農試所作物組 呂椿棠

▲ 農試所副所長 呂秀英

一、前言

DNA分子標誌以其多樣性豐富，利用遺傳標誌與重要外表型性狀間之連鎖關係，已廣泛用於植物分子遺傳圖譜建構、品種鑑別與品系純度測定、重要性狀基因定位與選殖、基因轉殖植物鑑定及分子標誌輔助選拔育種等方面。DNA分子標誌的擴增產物在電泳凝膠上分離後，必須讀帶與統計分析所獲得的數據資料。DNA電泳圖譜通常依各樣本在圖譜位置出現的條帶有和無兩種狀態(1或0)來呈現紀錄，然後再依其出現狀態的相似程度，計算出兩兩樣本間的相似性係數(similarity index)後，進行集群分析(cluster analysis)與主軸分析(PCoA, principal coordinate analysis, 或稱主座標分析)，利用這兩種統計方法所各自產生的樹狀圖(cluster dendrogram)及序

列分佈圖(ordination dendrogram)，以相輔相成來判斷各樣本的群別歸屬，該資料處理的完整流程可參閱本文作者之前作(呂等，2008)。在DNA分子標誌的分析流程中由於涉及相似性係數之計算，顯性分子標誌如RAPD和AFLP，建議使用Jaccard係數；共顯性標誌如SSR和RFLP，則採用Nei & Li係數或Dice係數(兩者公式相同)；若有針對同一試驗材料進行共顯性與顯性標誌所得之圖譜條帶紀錄，欲合併資料一起進行統計分析時，則應當使用Nei & Li係數或Dice係數(呂等，2008)。但一般統計套裝軟體並未提供這兩種相似性係數，因此往往必須藉助於特殊分析軟體，目前以需付費的NTSYS(Numerical Taxonomy System)生態學專用軟體較為使用，惟其操作較繁瑣不具親和性，且輸出圖形的編輯功能相當有限；或者可先利用其他軟體產生相似性係數後，再使用一般統計套裝軟體(如SAS、SPSS、SAS、STATISTICA)來分析。有鑑於此，本文擬介紹一個免費又容易操作的PAST統計軟體，可一氣呵成計算出DNA分子標誌資料之相似性係數並執行集群分析與主軸分析。

作者：呂椿棠副研究員
連絡電話：04-23317127

PAST is a free, easy-to-use data analysis package originally aimed at paleontology but now also popular in many other fields. It includes common statistical, plotting and modeling functions:

Download PAST

二、PAST之功能簡介

PAST為PAleontological Statistics（古生物統計學）的縮寫，為挪威Oslo大學自然歷史博物館之國家型專案計畫所開發之資料分析軟體，最初應用於古生物學研究，現今則廣泛應用於各領域，內含常用統計、繪圖及模式化功能，除此也涵蓋了一般統計套裝軟體所沒有的生態學、形態學、生物多樣性學、遺傳分類學、地理空間學、生物地層學等特殊領域之統計方法。官方網站（<http://folk.uio.no/ohammer/past/>）免費提供下載程式、完整操作手冊、實例研究、教學文件等。至2012年底PAST最新版本為2.16版，可於windows XP、Vista及Windows 7等作業系統下執行，該程式檔案容量甚小僅4.1MB，可免安裝，只要直接點選程式檔案，即可執行。

三、利用PAST進行分子標誌資料之樣本歸群的分析步驟說明

DNA分子標誌資料的處理流程，包含：電泳圖譜條帶出現紀錄以0和1來編碼登記；利用二元編碼資料，估算兩兩樣本間的相似性係數；由相似性係數進行集群分析與主軸分析，最後同時依據集群分析所得樹狀圖及主軸分析所得序列分佈圖，來判斷各樣本的群別歸屬。茲以一套6個樣本8個條帶位置之已編碼的分子標誌資料為例，說明使用PAST進行統計分析的步驟如下：

（一）電泳圖譜條帶出現紀錄的編碼輸入

PAST輸入資料的工作表類似於EXCEL工作表，輸入格式如圖一所示，行標題為樣本名稱（A、B…），欄標題為條帶位置（如1、2…），行和欄兩者不可顛倒。有關標題名稱之修改、資料

Row (Sample): 樣本

Column (Variable): 條帶位置

各樣本的條帶有無以0或1表示

	1	2	3	4	5	6	7	8
A	1	1	1	1	0	1	1	1
B	0	0	1	1	0	1	0	0
C	0	1	1	1	0	0	0	0
D	1	1	1	1	0	1	1	1
E	0	1	0	0	0	0	0	0
F	0	1	1	0	0	0	1	1

圖一、PAST的輸入資料視窗及格式。

PAST is a free, easy-to-use data analysis package originally aimed at paleontology but now also popular in many other fields. It includes common statistical, plotting and modeling functions.

Figure 3-1

A spreadsheet-like data entry form

工作表之存檔與開啟等視窗基本操作功能，可參考PAST使用手冊之詳細說明。

(二) 指定分析的資料範圍

PAST要求在執行任何分析之前，都必須先指定欲分析的資料範圍，如圖二所示。若全選，點選輸入畫面上功能選單的Edit之Select all；若部分選取，則在Edit mode勾選（內設）下按住鍵盤的shift鍵同時用滑鼠拉曳欲分析的部分資料範圍。

(三) 估算相似性係數並進行集群分析

點選功能選單上的Multivar之Cluster analysis，出現操作畫面如圖三-A所示，於其右側上方Algorithm（演算法）點選Paired group（即UPGMA法，是DNA分子標誌資料進行集群分析最常使用的方法），然後再於下方Similarity measure（相似度測量）點選Jaccard（顯性標誌適用）或Dice係數（共顯性標誌適用），此時集群樹狀圖立即顯現在畫面左側框內，於圖上點一下可進到圖形編輯視窗執行簡單的圖形編輯，如修改XY軸文字、選擇黑或藍色線條、加粗或不加粗線條、改變字體（圖三-B）；但只提供一種方向的樹狀圖（由上往下），此圖可另存成圖形檔。

(四) 估算相似性係數並進行主軸分析

重新進行上述（二）步驟後，點選功能選單上的Multivar之Principal coordinate，出現操作畫面如圖四所示，首先於其右側上方Similarity measure（相似度測量）點選Jaccard（顯性標誌適用）或Dice係數（共顯性標誌適用），再於下方Transformation exponent（轉換指數值）中嘗試改變c值（PAST提供c=1,2,4,6四種選項），檢視所估算出的特性根(Eigenvalue)的貢獻率(Percent)，選擇其中擬合效果最好的情況，亦即使第1軸和第2軸的累計貢獻率（解釋能力）

最大，如本例調整到c=2時前兩軸的貢獻率各為54.09%與34.38%，兩者累加值達88.47%，此為四種轉換指數值中擬合效果最好者，故採用此結果所繪製出的序列分佈圖；然後，點選View scatter功能鍵，可進入圖形瀏覽畫面內，勾選Row labels使圖中各點出現樣本代號，也可進而點選View number使顯示圖中各點數值（自由選項）；最後於圖上點一下可進行簡單的圖形編輯（畫面同圖三-B），該圖亦可存檔或直接列印。但必須注意，PAST只提供二維分佈圖。

(五) 結果解讀

綜合集群樹狀圖（圖三-A，解釋能力達88.47%）與序列分布圖（圖四）兩者結果，發現樣本A和D為100%相似，其餘樣本可與此兩樣本明顯區分開來；F和A或(D)的相似度僅57%，而B和C相似度僅50%，E和其他樣本之相似度不到20%。

四、PAST在分子標誌資料分析之另類運用

PAST的優點是免費、操作簡易、功能完整，且執行程式檔案容量甚小免安裝，但為使檔案容量變小，也相對犧牲了一些非屬必要的功能，例如沒有太多額外的輸入選項、輸出結果簡單明瞭、以及圖形編輯功能仍顯陽春（但仍較NYSTS為佳）等。因此，從上述可知，PAST應用於分子標誌資料之多樣性分析，仍有些不足之處，如集群分析結果僅提供單一方向的樹狀圖，但我們可將該圖另存新檔後，再透過其他修圖軟體工具（如Powerpoint等）來調整圖的方向或再美化加工；另外，使用PAST執行主軸分析結果無法提供三維序列分佈圖。基此，若希望輸出結果更詳細且圖形編輯效果更佳，則可以只利用PAST求得的相似性係數作為輸入資

PAST is a free, easy-to-use data analysis package originally aimed at paleontology but now also popular in many other fields. It includes common statistical, plotting and modelling functions:

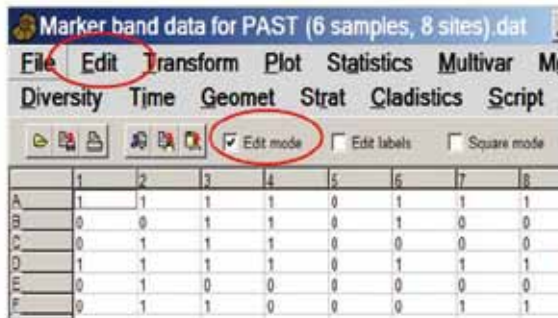
Download PAST

• A spreadsheet-type data entry form

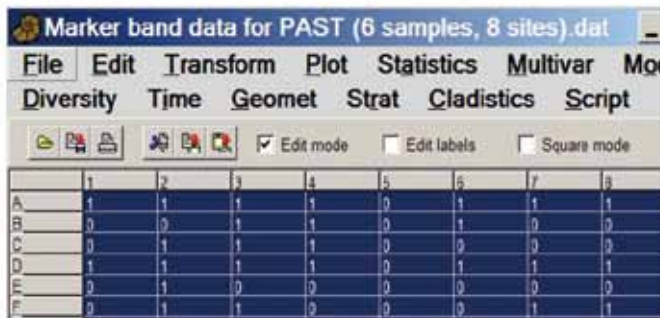
料，再進到其他一般統計套裝軟體進行後續的集群分析和主軸分析（但前提是該統計套裝軟體，須具有可直接讀取相似性係數的功能）。

單純使用PAST求算相似性係數的步驟很簡單，只要在前節所述之前兩個步驟的輸入編碼資料並指定資料範圍後，再點選功能選單上的Statistics之Similarity

生
物
統
計



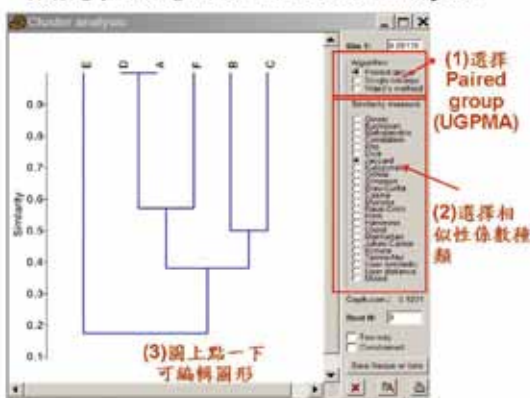
執行任何分析前：
Edit/ Select all
或在Edit mode勾選下
按住shift鍵同時用滑鼠
拉曳選擇範圍



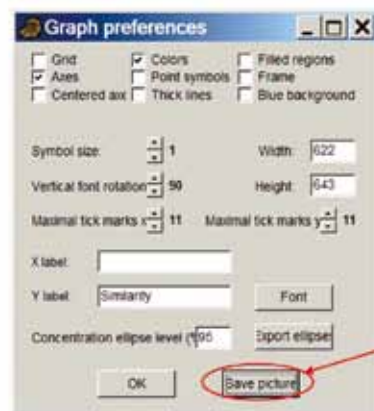
圖二、PAST在執行任何分析之前須先指定欲分析的資料範圍。全選：Select all/Edit；部分選取：勾選Edit mode下按住shift鍵，同時用滑鼠拉曳擬分析資料範圍。

(A)

功能選單上點選 Multivar / Cluster analysis



(B)

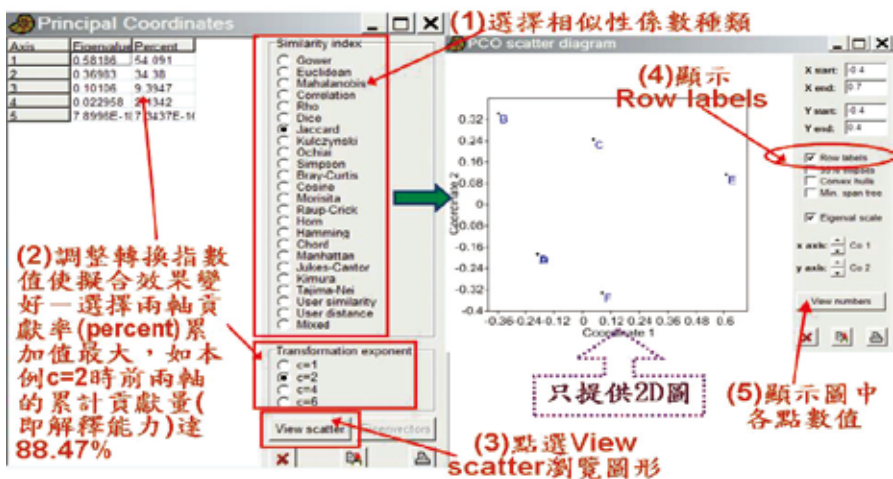


圖三、利用PAST估算相似性係數並進行集群分析的(A)操作畫面與執行步驟，及(B)圖形編輯與存檔。

PAST is a free, easy-to-use data analysis package originally aimed at paleontology but now also popular in many other fields. It includes common statistical, plotting and modeling functions:

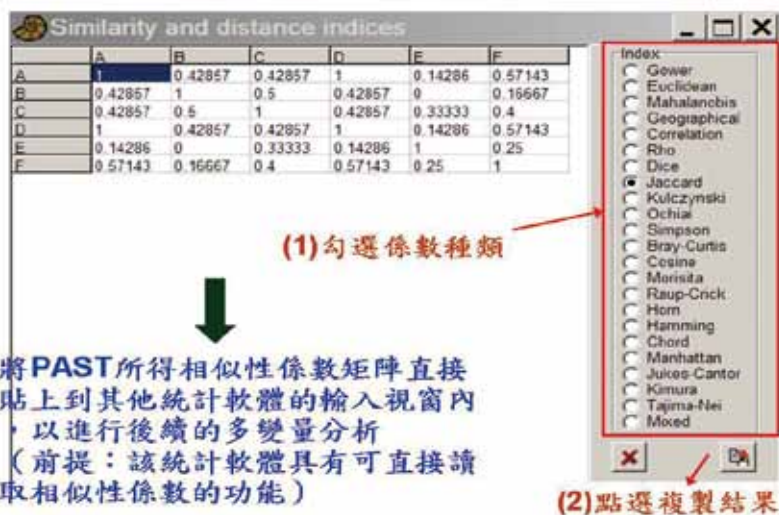
- A spreadsheet-type data entry form

功能選單上點選 Multivar / Principal coordinates



圖四、利用PAST估算相似性係數並進行主軸分析的操作畫面與執行步驟。

功能選單上點選 Statistics / Similarity and distance indices



圖五、利用PAST只計算相似性係數的操作畫面與執行步驟。

and distance indices，出現操作畫面如圖五所示，於其右側上方index（指標）點選係數種類，即在左視窗框內產生一個係數矩陣。該係數矩陣可直接利用複製與貼上到其他統計軟體的輸入視窗內，以進行後續的統計分析。

五、參考文獻

呂秀英、魏夢麗、呂椿棠。2008。DNA分子標誌資料之遺傳變異的歸群分析。技術服務73：25-27。