

# 生物序列比對分析之視覺與自動化系統—辨識 PAI 的雛形系統

## A Visualization and Automatic Analysis System for Bio-Sequence Alignment and Comparison — A Prototype System for Identifying PAI

郭嘉和	林耀鈴	蔡英德	許芳榮	張晃猷
Jia-Her Guo	Yaw-Ling Lin	Yin-Te Tsai	F.R. Hsu	Hwan-You Chang
靜宜大學資訊管 理研究所	靜宜大學資訊管 理研究所	靜宜大學資訊管 理研究所	靜宜大學會計研 究所	清華大學生命科 學研究所
ec86313@ms7.hi net.net	yllin@pu.edu.tw	ytsai@pu.edu.tw	frhsu@pu.edu.tw	hychang@life.nth u.edu.tw

### Abstract

Bacterial infections are still the major threat to human health. The recent progress on bacterial genome program provides many opportunities to identify novel targets for drug intervention. The pathogenic strain of a bacterium is different from its virulent relative in that it has acquired a large segment of DNA called *Pathogenicity Island (PAI)*. PAIs normally carry genes encoding several virulence factors. The identification of a PAI in a bacterial genome can greatly facilitate the understanding of bacterial pathogenesis and development of antimicrobial drugs. Based on several biological criteria, we develop a software system that allows one to identify PAIs rapidly from thousands of contig in the bacterial genome database.

Keywords : PAI, Sequence Visualization, G+C ratio, Sequence Comparison, Bio-characteristics database

### 中文摘要

利用 PAI (Pathogenicity island, 致病毒力基因島) 的特性來開發一個軟體系統, 以便幫助生物學家在大量基因體序列中分析、辨識 PAI。在這個系統中, 將建立一個基因序列資料庫, 來做 PAI 的序列比對。並且完成能縮短生物研究流程的實用軟體工具, 如 DNA G+C 的比例計算及蛋白質親水性比例計算等程式。希望藉由開發出來的軟體系統, 生物學家可以縮短生物研究流程、有更多的時間去更進一步深入地做 PAI 的研究, 以發現更多 PAI 的特性以及解決的方法, 用以製造更多有效的藥物來抑止細菌感染所造成的傳染病。此外, 此套軟體系統可進一步擴展功能用於研究動植物的基因體, 以促進生物學的發展。

關鍵詞：PAI、圖形化、G+C 比例、序列比對、生物特徵序列資料庫

### 一、導論

被科學界視為媲美曼哈坦原子彈計劃

(Manhattan Project) 和阿波羅登月計劃 (Apollo Project) 的人類基因體計劃 (Human Genome Project 簡稱 HGP) [2, 15] 於一九九〇年十月一日正式展開，最終目的在於解讀基因體核甘酸序列，並鑑別所有人類基因之功能。預訂以十五年的時間完成，耗資三十億美金。參與人類基因體計畫的有美、英、日、德、法、中等六國的十六個實驗室，已於 2000 年五月初宣告人類基因體初稿定序告一段落，並將工作重點轉入完稿階段。2000 年 6 月 26 日美國總統柯林頓和英國首相布萊爾聯合發表人類基因體計畫的成果。預計在 2003 年將完全破解人體 30 億組的遺傳密碼排列，繪製成人體基因地圖。在國科會以及台北榮總經費支持下，由台北榮總和陽明大學多位教授共同組成的榮陽團隊於 2000 年五月八日對外公佈：在人類第四號染色體已完成總數約一千萬鹼基的定序工作。

人類基因體計畫不僅僅是想知道這些基因的序列而已，更想了解或找出其生理功能及在一些生理程序上所扮演的角色。除了人類基因體外，目前也完成了一些其他生物基因體的序列，如 *Escherichia coli*, *Saccharomyces cerevisiae*, *Caenorhabditis elegans* and *Drosophila melanogaster* 等。隨著人類基因體計畫進入落幕階段，生物學正進入全新的數位時代—後基因體時代(post-genome era)。從前逐一解讀基因的方式將成為明日黃花，科學家所面對的是基因資料庫中天文數量之基因序列，而且總數還在不斷地快速累積當中，整個基因體功能的分析工作才剛剛開始。分析處理基因體圖譜和序列之龐大數據成為整個計劃不可缺失的重點工作，大致可分為兩方面：第一是生物資料庫；第二是發展更先進的序列分析工具。基因體新藥研發產業 (Genomic drug discovery industry) 將是在後基因體時代中所產生的一項新的生物技術研究領域。而如何利用龐大的基因庫，結

合生物資訊學 (Bioinformatics)，篩選出最具醫療價值的蛋白質分子，進而利用它開發新藥，已經被視為最重要的課題。基因科技的前景不只應用在治療疾病，還可用以了解生物演化、改良農產品等，幾乎掌握所有生命科學，也大幅影響人類生活，對人類生活的衝擊更不容小覷。

目前細菌性的傳染病仍是導致人類死亡的主要原因之一。已往都採用抗生素治療，但由於抗藥性菌株的廣泛流行，已明顯的降低抗生素的效用。最近由於細菌基因體學的發展，揭露了許多前所未有的細菌基因及相關資訊，其中有許多新基因值得作為未來藥物研發的對象。許多生物學家透過一系列微生物學和免疫學研究，想要發現新的抗生素。一般而言，與細菌致病能力有關的基因是最值得作進一步的探討。生物學家發現許多致病的細菌都有一段特殊的 DNA 片斷，稱作致病毒力基因島 (PAI)[13]。透過 PAI 的研究，生物學家能夠發現這些致病細菌的許多重要特性，以便於能製造新的藥物。PAI 的常見特徵包括 1) 帶有數個致病相關基因[12, 14, 17]； 2) G+C 的比例和該細菌其他基因體的區段有所不同； 3) 兩端有特殊的插入序列，如 tRNA 基因序列或是同向的重複序列[5, 8, 20, 21]； 4) 具有協助轉移或嵌入至細菌染色體上的酵素基因。

目前 PAI 的研究仍要透過繁鎖的實驗分析，並無一方便迅速的方法幫助生物學家在大量的基因片段序列中偵測到 PAI。因此本文利用以上所述的 PAI 特徵，發展一軟體系統，用以迅速搜尋現有微生物基因庫中存在的 PAI。首先我們將建立與 PAI 特徵相關的資料庫，如插入序列資料庫。然後設計並發展快速搜尋 PAI 的序列比對演算法。此外，這個系統將提供能縮

短生物研究流程的實用軟體工具，如 DNA G+C 的比例計算及蛋白質親水性比例計算等工具程式。此一軟體將先以已知具有 PAI 的細菌基因庫如幽門曲狀桿菌作為測試對象，並進行修改。最後再針對包括肺炎克雷白氏菌等新完成的微生物基因庫進行比對。希望能及早發現這些微生物中基因體中所存有的 PAI，以供生物學家進一步之研究。相信此一軟體對於搜尋動物及植物病原細菌所帶有的 PAI 也將會有很大的幫助。

## 二、相關研究及現存系統

目前已有許多用來辨識分析生物序列的序列比對分析程式，如美國 NCBI 的 Blast [18]、日本 DDBJ 的 FastA [7]、Smith-Waterman、臺灣工研院的 FLAG [10] 等等。Blast [3, 4] 從 1990 年就一直發展至今，已經有很多系列版本被發展出來，可以運用在各種生物序列比對上。它是根據統計理論所設計的，雖然它也是一種 heuristic 運算法，其計分方式卻與像 dynamic programming 一樣，能模擬找出突變最少的狀況的過程。FastA [16, 19] 是在 1985 年就已開發出來，至今亦經過多次改良。它是利用序列片段的相似性來推測可能有親緣關係的區域，再將少量找到的區域做嚴謹的 Local Alignment 分析，可以節省許多計算的時間。Smith-Waterman [9] 利用 Dynamic programming 的方法，能有系統地尋找同源的序列，可是它的速度太慢，必須用硬體加速才實用。FLAG 是臺灣工研院生醫中心最近發展的 DNA 序列比對程式，特別是針對基因體與基因體之間的序列比對。

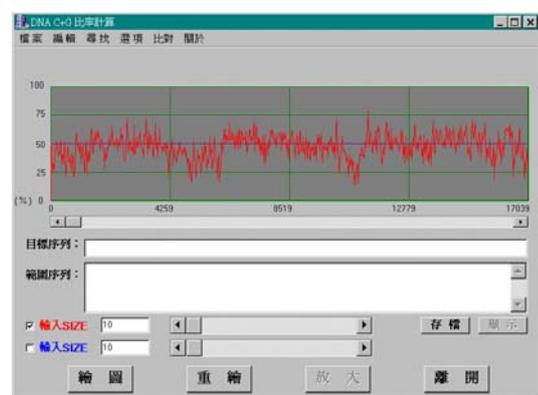
## 三、發展平台

本系統是使用 Borland C++ Builder 5 開發

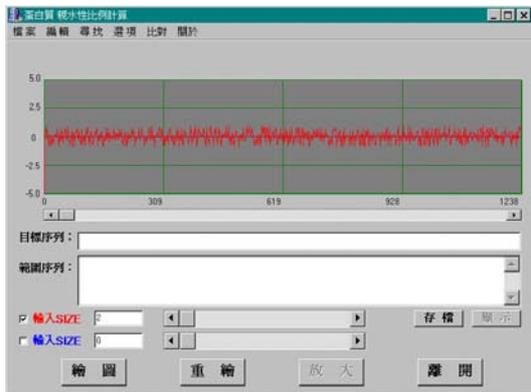
出來，相容於 Windows 系列（如 2000、NT、Me、98、95）作業系統。在本系統內的 Blast 序列比對演算法，是利用 NCBI 的 Blast 程式來執行，Local Alignment 程式則是作者依據演算法原理 [11, 9, 6] 撰寫出來的。本系統在一般個人電腦上即可安裝執行，增加使用者的方便及行動性，而且可以省卻很多麻煩，並可避免在網路上傳輸太大檔案的問題。

## 四、圖形化呈現方式

由於 DNA 序列與蛋白質序列全是一串一串的文字，資料一多起來就會變得很繁瑣，而且很難迅速地去了解其中代表的意思。所以，本系統採用圖形方式來呈現 DNA 序列與蛋白質序列及所有的處理結果。在 DNA 序列方面，利用 G+C 比例計算的結果來呈現。（如圖一）在蛋白質序列方面，利用親水性比例計算的結果來呈現。（如圖二）除了圖形化的結果外，本系統也會將一些指定的結果輸出成序列文字型式，在稍後的系統功能介紹中都會詳細描述。使用這種圖形呈現方式，相信使用者可以很快地發現他所要關心的部份。



圖一：SPI3 的 G+C 比例圖形



圖二：BVGS 的親水性比例圖形

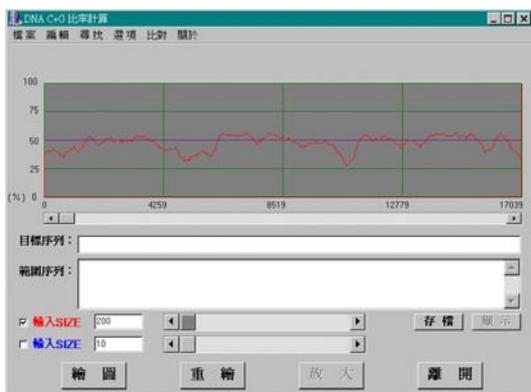
## 五、系統功能

本系統含有多種功能模組及一些便利的小功能，都是為幫助使用者辨識 PAI 而設計的，將於下面詳細描述之：

### 5.1 工具程式

- G+C 比例計算：

由於 PAI 的特徵：G+C 的比例和該細菌其他基因體的區段有所不同，以及圖形化的考量，所以本系統具備 G+C 比例計算的功能。只要在輸入 DNA 序列檔案或直接輸入 DNA 序列時，選擇 DNA 型式，即可輸出 G+C 比例圖形。(如圖一) 在圖形底下的「輸入 SIZE」物件可以控制 G+C 比例，使之放大或縮小。(如圖三)



圖三：SPI3 的 G+C 比例放大圖形

- 蛋白質親水性比例計算：

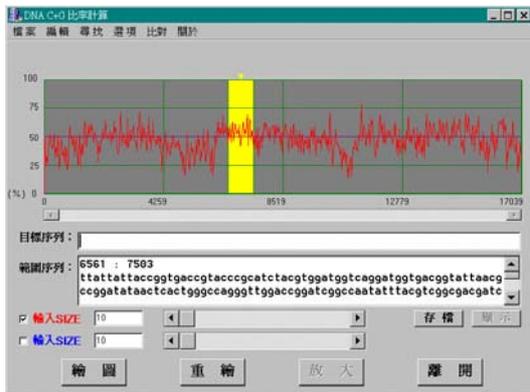
為了增加本系統的功能及多元性，加入蛋白質親水性比例計算的功能，讓使用者可以從更多的生物特性方面來觀察序列資料。只要在輸入蛋白質序列檔案或直接輸入蛋白質序列時，選擇蛋白質型式，即可輸出親水性比例圖形。(如圖二) 在圖形底下的「輸入 SIZE」物件可以控制親水性比例，使之放大或縮小。

### 5.2 序列尋找

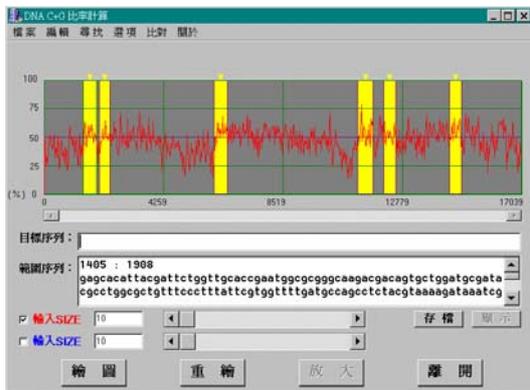
為了滿足生物學家的需要，以便他們可以做更多種類的分析、研究，本系統加入了序列尋找的功能。此功能提供多種條件選項讓使用者自己組合出符合需求的條件情況，條件選項可複選也可單選，會有不同的結果出現。(如圖四) 若只選擇「(% )以上」和「(% )以下」，會尋找出符合條件並最長的一個序列。(如圖五) 若只選擇「bp 以上」和「bp 以下」，則不會找到任何序列。若複選這 2 種條件選項，會尋找出符合條件的所有序列。(如圖六) 除了圖形結果輸出外，尋找到的序列文字資料也會同步輸出到「範圍序列」中。



圖四：DNA G+C 比例的條件選項



圖五：符合 30% 以上，80% 以下



圖六：符合 30% 以上，400bp 以下、600bp 以下

### 5.3 序列比對

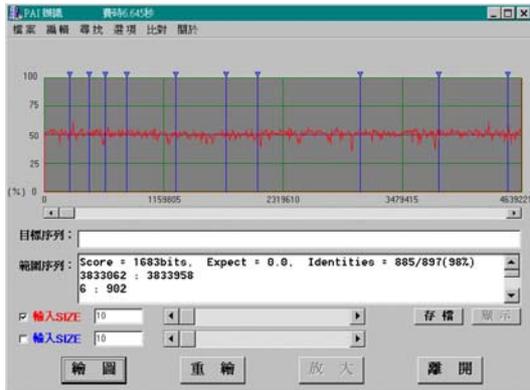
目前有很多的序列比對程式，而使用最廣的是 FastA 與 Blast 系列的程式。我們以這兩個程式加上最嚴謹的序列比對演算法(Local Alignment)，就它們的優缺點來考量：

- FastA 系列: 利用序列片段的相似性來推測可能有親緣關係的區域，再將少量找到的區域做嚴謹的 Local Alignment 分析，可以節省許多計算的時間。靈敏、但是速度較 Blast 慢。只能以核酸序列比對核酸資料庫，或以蛋白質序列比對蛋白質資料庫。TFastA 則以蛋白

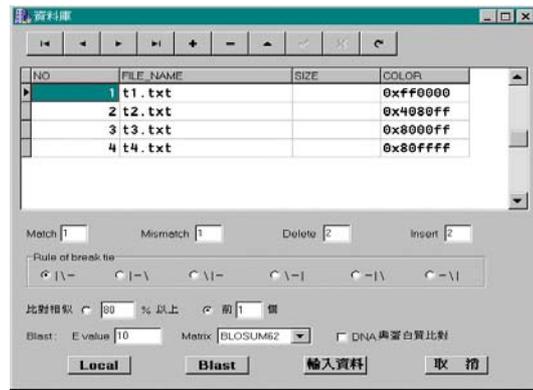
質序列比對核酸序列資料庫。

- Blast 系列: 比對速度非常快，但在序列相似性較低時會有失誤，有可能會漏掉一些相關的序列。它是一組程式，可以根據輸入的序列與所選擇的資料庫種類而執行適當的程式。例如：BlastN 以 DNA 序列去查詢 DNA 資料庫、BlastP 以蛋白質序列去查詢蛋白質資料庫、Blastx 以 DNA 序列去查詢蛋白質資料庫、TblastN 以蛋白質序列去查詢 DNA 資料庫。
- Smith-Waterman ( Local Alignment ) 系列: 是最嚴謹的序列比對，但處理速度太慢，而且會耗用很大的儲存空間，例如兩個 1Kb 的序列互相比對，至少要佔用  $10^6$  個位置存得分。在改善儲存空間方面，可以利用 Linear space 的方法來降低對記憶體之需求量。可是在改善速度方面，雖然可用硬體加速來增加執行速度，但資料量一多起來，還是很花時間。[1]

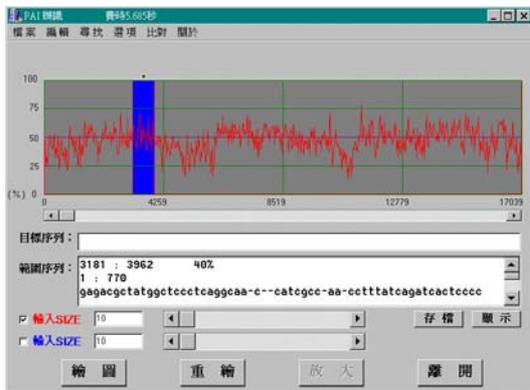
基於速度較快、功能較多的理由，本系統採用 Blast 程式來做序列比對。至於它的缺點：不夠嚴謹，本系統則另外增加 Local Alignment 演算法來輔助、補強。資料量很大時，建議使用 Blast 來做序列比對 (如圖七)，若要再做細微一點的比對時，便可使用 Local Alignment 演算法來做較精確的局部比對。資料量比較小又不使用特殊功能時，即可直接利用 Local Alignment 演算法來做序列比對。(如圖八) 除了圖形結果輸出外，比對出來的序列文字資料也會同步輸出到「範圍序列」中。



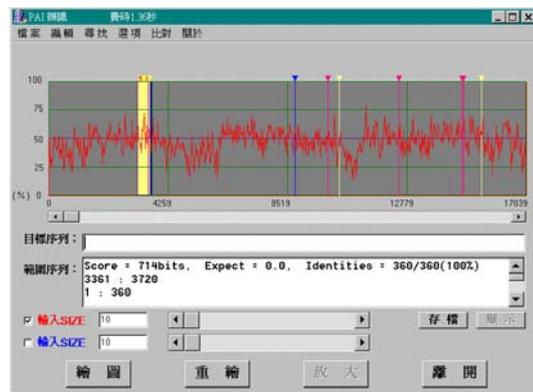
圖七：資料量很大時，使用 Blast 比對



圖九：資料庫



圖八：資料量較少時，使用 Local Alignment 比對



圖十：資料庫比對結果

#### 5.4 生物特徵序列資料庫

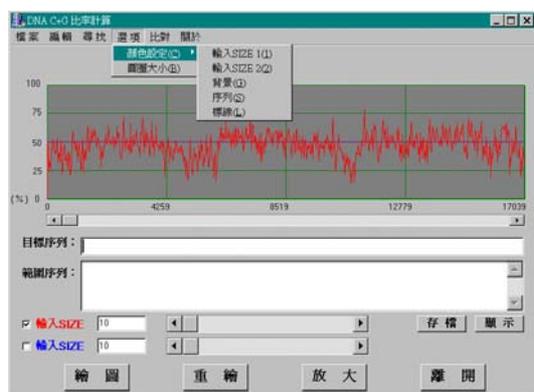
本系統提供生物特徵序列資料庫，讓 Blast、Local Alignment 做資料庫序列比對時使用。採用動態的方式來建構資料庫，使用者除了可以利用工具列編輯資料庫外，還可以用 Excel 依照格式編輯，並存成 CSV 檔案格式，需要使用時再輸入即可。因此，使用者可以依據需要，編輯好幾組資料庫表單，等於是使用者可以自己幫資料庫分門別類，增加比對時的多樣性、機動力。(如圖九)在對資料庫做比對時，會將結果依據資料庫表單上的顏色設定分別塗色，以幫助使用者容易分辨。(如圖十)

#### 5.5 使用者設定

由於本系統是採用圖形化呈現方式，所以顏色就變得相對重要。考慮到不同的使用者可能對不同的顏色比較敏感，在某些顏色組合下，能看得比較清楚、容易辨識。所以，本系統提供顏色設定的功能，讓使用者可以自己控制二組圖形、背景、序列、標線的顏色。(如圖十一)其中「序列」指的是 20 個胺基酸，每一個胺基酸有其代表的顏色，並提供「儲存」的功能，以便以後還可以使用同樣的顏色設定。(如圖十二)此選項的使用時機，請參照 5.6 的「對照序列」功能說明。

另外，為了讓使用者清楚分辨胺基酸或鹼

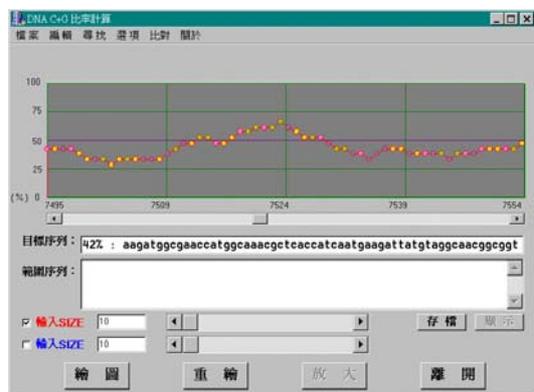
基在圖形中的位置，本系統設定只要出現在畫面上的序列長度少於 100 時，就會在每一個胺基酸或鹼基的位置上畫一個圓形並塗色以便辨識。(如圖十三) 使用者可以自己設定是否使用此功能及圓圈大小，在圖十上有個功能選項「圓圈大小」，就是用來設定此功能的。



圖十一：顏色設定



圖十二：序列的顏色設定



圖十三：畫面上的序列長度等於 60 時

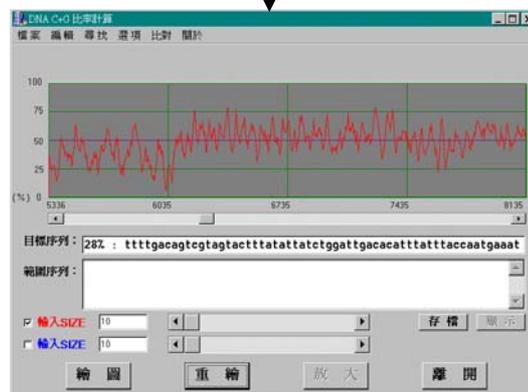
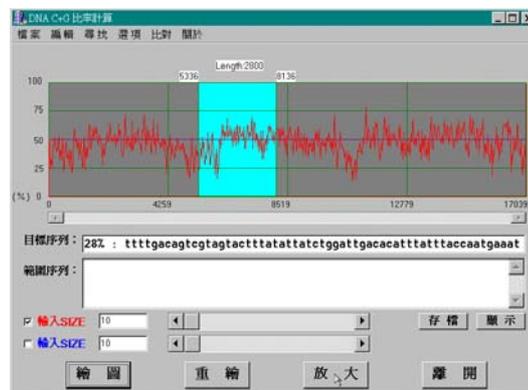
## 5.6 便利的小功能

為了讓使用者能夠更方便、迅速辨識 PAI，本系統含有很多便利的小功能來幫助使用者。以下便是其中的一些：

- 局部放大

若資料量太大，使用者想看清楚局部區域的圖形時，便可利用此功能，把滑鼠拖曳選擇的部份，依與整個畫面的比例大小放大在畫面上。(如圖十四)

另外，正如圖五、圖六、圖七、圖八、圖十所示：在序列尋找、比對完之後，結果所顯示的區域上方都會一個指示的倒三角形。以滑鼠左鍵點選這些標記，會將此標記代表的序列放大在整個畫面上。

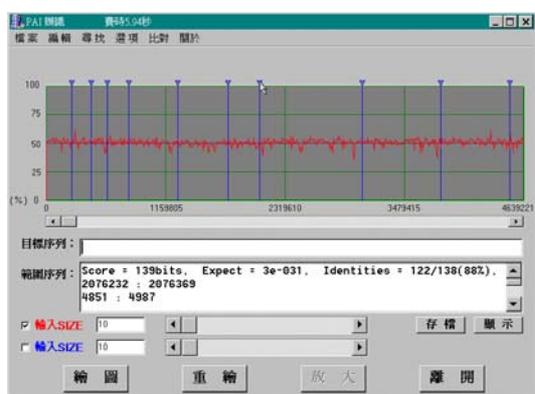


圖十四：局部放大

- 局部顯示

若資料量太大，使用者想看清楚局部區域的序列資料時，便可利用此功能，把滑鼠拖曳選擇的部份，將其代表的序列資料顯示在「範圍序列」裡。

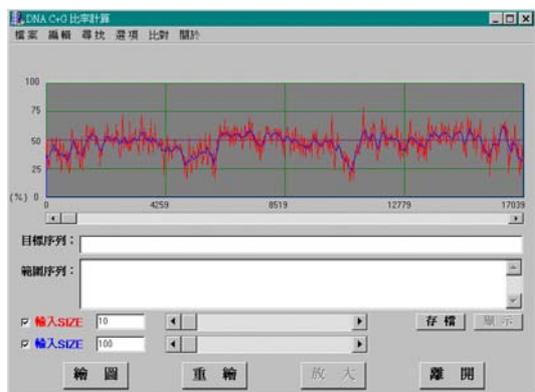
另外，正如圖五、圖六、圖七、圖八、圖十所示：在序列尋找、比對完之後，結果所顯示的區域上方都會一個指示的倒三角形。以滑鼠右鍵點選這些標記，會將此標記代表的序列及相關資訊顯示在「範圍序列」中。(如圖十五)



圖十五：比對結果序列顯示

- 對照比較

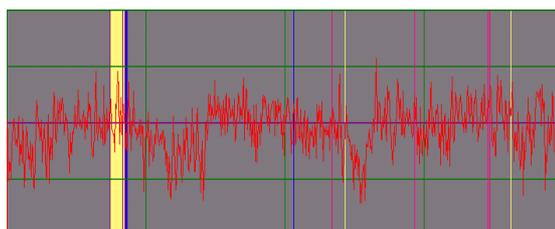
本系統可同時提供二組比例大小不一樣的圖形，以便使用者可以對照比較。只要在「輸入 SIZE」設定即可，字型的顏色是代表圖形的顏色。(如圖十六)



圖十六：二組圖形對照比較

- 儲存圖片

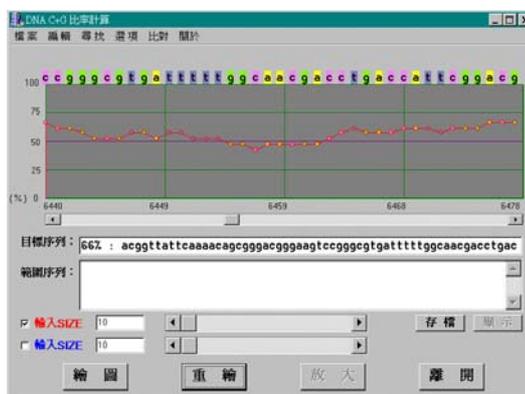
為了方便使用者將圖形用在別的地方，本系統提供儲存圖片的功能，可以將畫面中的圖形存成 BMP 點陣圖。(圖十七)



圖十七：將圖十的圖形存成圖片

- 對照序列

另外，為了讓使用者清楚分辨胺基酸或鹼基，本系統設定只要出現在畫面上序列長度少於 50 時，就會在每一個胺基酸或鹼基的正上方位置秀出代號並塗色以便辨識。(如圖十八)



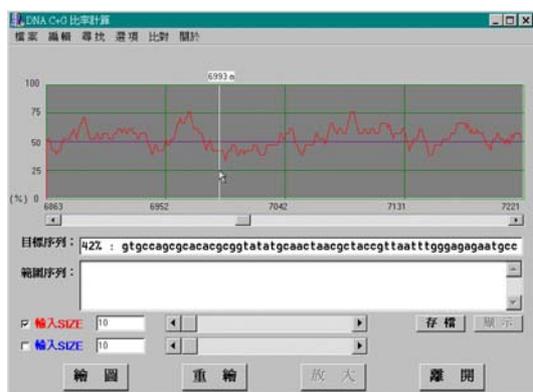
圖十八：畫面上的序列長度等於 39 時

- 水平捲軸

若已利用局部放大的功能，將局部區域放大到整個畫面上，便可使用水平捲軸的功能，藉移動捲軸來觀看其它區域的圖形。另外，也可利用滑鼠右鍵按住圖形不放，然後再水平移動，一樣會有跟水平捲軸相同的功能。

- 標線

按滑鼠左鍵點在圖形上，則會標示此位置的一條直線，並且在上方會顯示此位置的資訊及胺基酸或鹼基代號。除了出現上述的標線外，還會將此位置的數值及本身加左右各 30 個的胺基酸或鹼基顯示在「目標序列」中。(如圖十九)



圖十九：標線

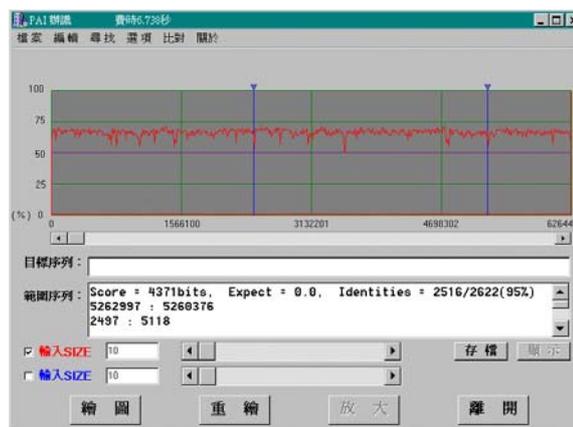
- 儲存序列資料

為了方便使用者記錄序列資料，本系統提供儲存序列資料的功能，可以將顯示在範圍序列中的序列資料存成 TXT 純文字檔。

## 六、系統應用測試

我們藉由以下的設備來對幾組 DNA 序列作 Blast 比對: Windows 98 SE 作業系統、Celeron 1000 CPU、256RAM 記憶體。輸入 *Pseudomonas aeruginosa* PA01 ( 6.2MB ) 的繪圖時間為 2.0 秒。比對的對象分別為 PAGI-1 ( 65KB, 如圖二十)、PAI\_O157 ( 56KB )、SPI3 ( 21KB )、PAI6 ( 7KB, partial ), E-value 設定為 1e-299, 處理時間分別為 6.7 秒、5.2 秒、4.8 秒、4.6 秒。輸入 *Escherichia coli* O157:H7 ( 5.6MB ) 的繪圖時間為 1.6 秒。比對的對象分別為 PAGI-1、

PAI\_O157、SPI3、PAI6, E-value 設定為 3e-11, 處理時間分別為 4.7 秒、5.9 秒、4.2 秒、4.2 秒。輸入 *Pyrococcus abyssi* ( 1.7MB ) 的繪圖時間為 0.7 秒。比對的對象分別為 PAGI-1、PAI\_O157、SPI3、PAI6, E-value 設定為 3e-11, 處理時間分別為 1.9 秒、2.0 秒、1.8 秒、1.7 秒。若比對出來的結果愈多，則處理的時間會相對地增加。



圖二十：Pseudomonas aeruginosa PA01 比對 PAGI-1

## 七、結論與未來方向

這個系統將幫助生物學家快速地辨識、分析 PAI, 並且預期發現新的 PAI。利用獲知的資訊製造新的藥物, 以抑止細菌感染所造的傳染病。更進一步, 希望此套系統可運用於動植物的 DNA 分析上, 以促進生物學的發展。由於本系統擁有眾多功能, 除了可以辨識、分析 PAI 外, 更可以運用在其它的生物資料分析上。今後, 也仍會繼續增加其它能幫助生物學家做分析工作的功能, 以期本系統能真正達到協助生物學家研究的目的。

### 備註

本系統由國科會 NCS 89-2218-E-126-006 補助支持, 可從 <http://alzo.cs.pu.edu.tw/~bioinfor>

mation 下載本系統。

## 感謝

徐婕婷同學在 Blast 演算法相關資訊的介紹及 Blast 程式的應用指導、賴怡琪同學在生物領域相關資料的提供及程式的測試。

## 參考文獻：

- [1] 序列資料庫搜尋程式  
[http://binfo.ym.edu.tw/post/topics/srch\\_prg.htm](http://binfo.ym.edu.tw/post/topics/srch_prg.htm)
- [2] 淺談人類基因體計畫  
<http://life.nthu.edu.tw/~stass/animalfarm/5th/One-5.htm>
- [3] S.F. Altschul, W. Gish, W. Miller, E.W. Myers and D.J. Lipman, Basic local alignment search tools. *J. Mol. Biol.*, **215**: 403-410, 1990.
- [4] S.F. Altschul, T.L. Madden, Alejandro A.Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and D.J. Lipman, *Gapped BLAST and PSI-BLAST*: a new generation of protein database search programs, *Nucleic Acids Res.* **25**:3389-3402, 1997.
- [5] C. Buchrieser, R. Brosch, S. Bach, A. Guiyoule and E. Carniel, The high pathogenicity island of *Yersinia tuberculosis* can be inserted into any of the three chromosomal *asn* tRNA genes. *Mol. Microbiol.*, **30**: 965-978, 1998.
- [6] T.H. Cormen, C.E. Leiserson, R.L. Rivest, Introduction to Algorithms : Dynamic Programming, p301-328.
- [7] DNA Data Bank of Japan  
<http://www.ddbj.nig.ac.jp/>
- [8] U. Dobrindt, P.S. Cohen, M. Utley, I. Mühldorfer and J. Hacker, The *leuX*-encoded tRNA<sub>5<sup>Leu</sup></sub> but not the pathogenicity islands I and II influence the survival of the uropathogenic *Escherichia coli* strain 536 CD-1 mouse bladder mucus in the stationary phase. *FEMS Microbiol. Lett.*, **162**: 135-141, 1998.
- [9] Richard Durbin, S.R. Eddy, Anders Krogh, Graeme Mitchison, Biological sequence analysis : Pairwise alignment, p12-45.
- [10] Fast Local Alignment for Gigabases  
<http://flag.itri.org.tw/>
- [11] Gusfield , *Algorithms on Strings, Trees, and Sequences*, Cambridge University Press, 1997.
- [12] J. Hacker, L. Bender, M. Ott, J. Wingender, B. Lund, T. Marre and W. Goebel, Deletions of chromosomal regions coding for fimbriae and hemolysins occur *in vivo* and *in vitro* in various extraintestinal *Escherichia coli* isolates. *Micro. Pathog.*, **8**: 213-225, 1990.
- [13] J. Hacker, G. Blum-Oehler, I. Mühldorfer and H. Tschäpe, Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. *Mol. Microbiol.*, **23**: 1089-1097, 1997.
- [14] J. Hacker, S. Knapp and W. Goebel, Spontaneous deletions and flanking

- regions of the chromosomal inherited hemolysin determinant of *Escherichia coli* O6 strain. *J. Bacteriol.*, **154**: 1145-1152, 1993.
- [15] Human Genome Project  
<http://www.ornl.gov/hgmis/>
- [16] D.J. Lipman, W.R. Pearson: *Rapid and sensitive protein similarity searches*, *Science*, 227, 1435-1441, 1985.
- [17] D. Low, V. David, D. Lark, G. Schoolnik and S. Falkow, Gene clusters governing the production of hemolysin and mannose-resistant hemagglutination are closely linked in *Escherichia coli* serotype O4 and O6 isolates from urinary track infections. *Infect. Immun.*, **43**: 353-358, 1984.
- [18] National Center for Biotechnology Information  
<http://www.ncbi.nlm.nih.gov/>
- [19] W.R. Pearson, D.J. Lipman: *Imported tools for biological sequence comparison*, *Proc. Natl. Acad. Sci. USA*, 85, 2444-2448, 1988.
- [20] A. Ritter, D. Gally, P.B. Olsen, U. Dobrindt., A. Friedrich, P. Klemm and J. Hacker, The Pai-associated leuX specific tRNA<sup>Leu</sup> affects type 1 fimbriation in pathogenic *Escherichia coli* by control of FimB recombinase expression. *Mol. Microbiol.*, **25**:871-882, 1997.
- [21] H. Schmidt, J. Scheet, C. Janetzki-Mittermann, M. Datz and H. Karch, An *ileX* tRNA gene is located close to the Shiga toxin II operon in enterohemorrhagic *Escherichia coli* O157 and non-O157 strains. *FEMS Microbiol. Lett.*, **149**: 39-44, 1997.