

基因轉殖作物外來基因比對資料庫之建立

The establishment of BLAST Database for foreign gene in transgenic plant

王昭能¹ 施仁國¹ 許承瑜¹ 王經篤¹ 葉錫東³ 范宗宸^{2*}

1. 亞洲大學生物資訊學系
2. 亞洲大學生物科技學系
3. 中興大學植物病理學系

*通訊：亞洲大學 生物科技學系 台中縣霧峰鄉柳峰路 500 號 mjfan@asia.edu.tw

摘要

全球基因轉殖作物的栽培正蓬勃發展，支持與反對基因轉殖的論述彼此爭論不休，建立可辨識基因轉殖或非基因轉殖生物的檢測方法，以讓大眾能有知道權利及選擇性是非常需要的。本研究的主要研究目的，為建立能基因轉殖作物外來基因(foreign gene)的 BLAST 資料庫，以提供快速的序列同源性比對，提供做為基因轉殖作物檢測系統所需的資訊。

本研究以基因轉殖作物外來基因(foreign gene)為研究對象，利用 AGBIOS 之 GM Database 取得已通過評估可釋放於環境之 GM 作物外來基因等資訊，且以 BioPerl 設計操作指令群對 NCBI Genbank 送出字串查詢核酸序列，以自動取得相關的核酸序列建立本機 BLAST 資料庫，使用 B2D Linux Server 平台及 WWWBLAST 介面，提供國內研究人員快速且有效率的比對出所研究的序列是否為 GMO 外來基因。本研究已立特定基因之 GMO BLAST 資料庫，並且可自行擴充資料庫，可於網際網路使用。

關鍵字：GMO、Foreign Gene、BLAST Database、WWWBLAST

The establishment of BLAST Database for foreign gene in transgenic plant

Chao-Neng Wang¹, shih-jen-kuo¹, Phillip C-y Sheu¹, Jing-Doo Wang¹, Ming-Jen Fan^{2*}

1. Department of Bioinformatics
2. Department of Biotechnology
3. Department of Plant Pathology

Addressed : ASIA UNIVERSITY Department of Biotechnology
500, Lioufeng Rd., Wufeng, Taichung 41354, Taiwan

mifan@asia.edu.tw

Abstract

The cultivation area in global scale of transgenic crop is increased rapidly. The open ion between support and opposition about the GMO is argued continuously. The mythos to recognize the GMO and NON-GMO are needed very necessary for consumer's right.

The goal of this research is to constructed a GMO BLAST Database system which can provide rapid sequence comparison. Established a GMO foreign gene database which can be as a part of GMO inspection system.

The GM database in AGBIOS, which content those authorization GM crop has been to through as appraisal and release to the environment, is the origin source of those GMO foreign gene come from. The BioPerl was used to design group operating instructions sent to NCBI GenBank to query and BLAST nucleotide sequence string interested. The free fee and stable B2D Linux Server platform was used to generated a local BLAST database which can automatically generate the associated nucleotide sequences for GMO Database. This platform also include a graphic WWWBLAST interface which can improve operational convenience and provide fast and efficient sequence alignment for GMO DNA sequencing.

Keyword : GMO 、 Foreign 、 BLAST Database 、 WWWBLAST

前言

全球基因轉殖作物的開發正蓬勃發展，目前全球已有百種以上的基因轉殖作物品系，涵蓋作物種類有數十種，基因轉殖作物外來基因(foreign gene)的數量也不斷的增加。基因轉殖作物為人類帶來許多利益與貢獻，如：減少農藥的使用量、降低地下水的污染、增加產量、容易管理、節約能源及人力成本、減少整地與土壤沖蝕、提供新的工業原料等

但根據世界衛生組織的官方說法(WHO, 2004)，「到目前為止，國際市場上已通過風險評估的基因轉殖食品，可能不會對人類健康產生危險」，亦即 WHO 並沒有對基改食品的長期性安全下斷言。基改產品所引發的問題有三，一是食品安全，二是環境安全，三是田間種植規範。現在支持與反對基因轉殖作物栽培及食用的論述彼此爭論不休，而在這仍屬於未定且需要長時間證明的情形下，可將基因轉殖與非基因轉殖生物區分開來並加以辨識的方法，使大眾可以有選擇的權利是非常需要的，基因轉殖作物外來基因種類不斷的增加，繼而可能產生如基因變化與基因遭人為修改以逃避檢驗的情形，增加了分子檢測的難度。因此需要建立一個外來基因資料庫，來輔助基因轉殖作物檢測的所需的資訊。

基因轉殖食品之鑑別檢驗方法目前主要以聚合酶鏈鎖反應(polymerase chain reaction, PCR)方法應用最廣，可供定性與定量分析。經 PCR 方法反應後之產物，可再行確認檢驗，常用方法如核酸定序、核酸內切酶圖譜、探針雜交等(Wolfram, 1997)。而核酸定序後的序列比對工作正是序列比對資料庫可應用之處。

本研究目的為建立 GMO BLAST Databases 用以比對及預測是否為以基因轉殖作物外來基因(foreign gene)，利用 AGBIOS 之 GMO Database 取得已通過評估，可釋放於環境之基因轉殖作物所舍之外來基因等資訊，並以 BioPerl 設計操作指令

對 NCBI Genbank 送出查詢字串，自動獲得所需的核酸序列，用以建立本機 BLAST 資料庫，並使用圖形化的 BLAST 介面提高操作便利性。可為 GMO 檢測後的核酸定序比對工作，提供一個免費快速且有效率的。BLAST Database 比對系統以提供研究人員所需的資訊。

本研究已建立 GMO BLAST 核酸資料庫與進行比對檢驗所需的序列來源，與運用的程式套件。並以 Papaya Ring Sport Virus Coat Protein 之 codon gene 為例，此 GMO BLAST Databases 系統平台可提供國內研究人員快速且有效率的比對出有興趣的序列是否為 GMO 外來基因之序列。

研究材料及方法

一、材料

(一) GMO 序列材料

GMO 作物其外來基因眾多，本研究以 Agbios 網站上各國已核准釋放於環境的外來基因為依據。

Agbios 網站外來基因(foreign gene)選項欄共計 42 種基因。基因轉殖作物中一般使用最常使用的啟動子為花椰菜嵌紋病毒(Cauliflower mosaic virus, CaMV)35S Promoter，也列入本研究的第 43 種目標基因。

此外，鞘蛋白(coat protein, CP)基因視其抗病毒的標的植物而使用不同病毒的鞘蛋白基因；在 Agbios 網站上分為 5 種，分別為木瓜輪點病毒(Papaya ringspot potyvirus, PRSV)、黃瓜嵌紋病毒(Cucumber mosaic virus, CMV)、矮南瓜黃化嵌紋病毒(Zucchini yellow mosaic virus, ZYMV)、西瓜嵌紋病毒 2 號(Watermelon mosaic virus 2, WMV2)、馬鈴薯 Y 病毒 0 株系(Potato potyvirus Y strain 0, PVY)，也一併列入本研究的目標基因。綜合上述，本研究目標序列共源自 48 種基因(表 1)

(二) 測試序列

本研究測試序列來自中興大學植物病理系葉

錫東 教授與亞洲大學生物科技范宗宸 教授研究室，共 5 條，序列名稱與長度分別如下：

1. PRSV CP (Papaya ringspot potyvirus Coat protein), 1070 bp。
2. 35S Promoter, 835 bp。
3. GUS (β -glucuronidase), 1812 bp。
4. NPTII (neomycin phosphotransferase II), 795 bp。
5. NOS(nopaline synthase), 260 bp。

二、方法

(一) 自動擷取外來基因

本研究利用數個 Bioperl 模組撰寫程式，進行 GenBank 遠端資料庫自動搜尋與下載。所使用的 Bioperl 模組如下：

1. Bio::Perl、Bio::SeqIO：對序列進行操作或呼叫與轉換序列檔
2. Bio::DB::GenBank：獲取遠端資料庫(Genbank)資料
3. Bio::DB::Query::GenBank：對 Genbank 進行遠端資料庫查詢

先搜集全部目標基因名稱，利用 Qualifier(檢索限定子)中的基因名稱[GENE]與 [Title]兩個限定子進行搜尋，並將這些字串存於 gm_ID.txt。使用 Bioperl 中 Bio::DB::GenBank 模組對 NCBI Genbank 資料庫進行操作；並利用 Bio::DB::Query::GenBank 模組進行迴圈詢問 (query)，查詢 GMO 目標基因。得到最新 GMO 目標基因序列之 accession number(登錄碼)，再將所有得到的 accession number 清單儲存於 gm_AN.list。

接下來利用 Bioperl 中之 get_sequence 函數至 NCBI Genbank 下載 gm_AN.list 內全部 accession number 所對應之 FASTA 格式的 GMO 目標基因序列資料，並全部存放在 gmo.fasta 檔，為建立 BLAST 資料庫作準備。

find_AN.pl 程式至 Genbank 詢問使用 \$query_string 以及後面接著詢問字串。如「Aad\[Title\] AND Aad\[gene\]」。另外建立 \$query_obj 進行查詢操作時，使用參數

-db(protein or nucleotide database)與 -query(query string)，程式執行流程與說明如圖 1。

(二) 建立 GMO BLAST 資料庫

建立 BLAST 資料庫之前，需將序列格式化。由 NCBI FTP 下載 BLAST-2.2.16-ia32-win32.exe。解壓縮後，取得 formatdb.exe。formatdb.exe 是格式化 BLAST 資料庫的工具程式。利用 formatdb.exe 將 gmo.fasta 進行格式化，GMO BLAST 資料庫即建置完成。

(三) wwwBLAST

由 NCBI FTP 免費下載的 BLAST 伺服器版為 WWWBLAST 程式，版本為 2.2.16，亦可轉為 Linux 作業系統的單機版，需安裝於 Apache 伺服器上。本研究的使用平台為 B2D Server Linux (stable s1 版)，下載相對應的 WWWBLAST-2.2.16-ia32-linux.tar.gz。NCBI WWWBLAST 以仿照 NCBI 線上 BLAST 介面為主，是網頁版 (Web-base) 的 BLAST 系統。由 Regular BLAST、PSI/PHI BLAST、Mega BLAST、RPS BLAST、BLAST 2 sequences 五組程式所組成，每組並包含有無支援主從式架構(client-server)模式，共計 10 種查詢類別。解壓縮後進行權限設定，TmpGifs 資料夾權限為 777，logfiles (WWWBLAST.log and psiBLAST.log)權限為 666。本研究系統架構圖如圖 2。

結果

(一) 序列蒐集

本研究使用 Perl 程式對 NCBI 送出字串，為避免 Perl 對字串解譯錯誤，故在特殊符號前，如雙引號(" ")、中括號([])等，加上具轉譯作用的倒斜線"\\"，讓 Perl 略過倒斜線之後一個字元不去解譯它。全部基因檢索字串如表 2，共得到 2447 條 accession number，基因總數 48 個，41 個有序列資料的基因。其中 cry1F、cry34Ab1、cry35Ab1、cry3A、cry3Bb1、cry9c、mcry3A 皆為抗蟲基因，使用 [GENE] 限定檢索，查無資料，取消限定檢索改用 NCBI 廣義的全文搜尋，仍然無相關資料。全部

序列以 fasta 格式儲存在 gmo.fasta。利用 formatdb.exe 程式將 gmo.fasta 格式化，GMO Blast Database 即建置完成。

(二) 序列測試

一、PRSV CP(Papayaringt potyvirus Coat protein)

比對結果

利用 PRSV CP 比對在 BLAST Hits，會得到 83 個結果，每一條序列結果都有序列名稱及期望值 (E)，文字比對可以發現 X97251 與查詢序列間共有 975 個核酸相同，Hit Table 部份 X97251 與 X78557 經 NCBI 查詢結果，皆為 PRSV 台灣 YK 株系，X78557 核酸區段 1831-2751 為 coat protein 與本查詢序列一致 (1986-2960) 接近。

二、35S Promoter 比對結果

在 Blast Hits 比對到有 148 個結果，文字比對可以發現查詢出前 20 條序列一致性均為 100%，且全部為一般基因選殖所用的 Ti 質體 (Binary vector)。Hit Table 結果至 NCBI 查詢顯示都皆為我們測試序列 (35S Promoter)。

三、GUS (β -glucuronidase)比對結果

Blast hits 比對到 20 條一致性程度高，後 10 條差異性較大，文字比對方面發現前 20 條序列的期望值均 100%，主要都為載體，hit Table 結果顯示 1812、1802、1808、1800、1794，共五組，NCBI 查詢結果都均為本測試序列 (GUS) 基因片段。

四、NPTII (neomycin phosphotransferase II)

比對結果

Blast Hits 比對到 100 條序列結果，文字比對前二十條序列一致性 100%，且全部為一般基因選殖所用的 Ti 質體，Hit Table 序列相似區全一致，到 NCBI 查詢結果顯示該區段為基因為 nptII。

五、NOS(nopaline synthase)比對結果

Blast Hits 查詢到 230 序列，文字比對前 20 條序列一致性均 100%，且全部都為一般基因轉殖所用的 Ti 質體，Hit Table 序列均相似，經 NCBI 查詢結果顯示應區段基因為 NOS Promoter。

NCBI 整合性搜尋系統 Entrez 進行核酸資料庫檢索時，直接在檢索欄位 (Search Field) 鍵入欲搜尋基因名稱，大小寫並無差異，以 gus(β -glucuronidase) 基因為例，結果如圖 3。

仔細檢查搜尋結果，可發現有些非目標序列亦在搜尋結果之內，須再次加以過濾，以第二條序列 NM_202701 為例，如圖 4，該序列全長 1580bp，基因名為「WR11」並非我們所要的 gus，若使用此全文檢索方式搜尋，結果會夾雜許多非目標基因名稱的序列。為了進行有效率的搜尋，需利用 Entrez 的限定搜尋功能，常用檢索限定詞如表 3。將 gus 限定為基因名稱搜尋 (即 [GENE] 限定詞)，結果如圖 5，可獲得 30 條序列，序列數量減少且結果較未限定搜尋的結果精確。最後，將序列輸出為 FASTA 格式，以建立 GMO Blast Database。

NCBI Entrez 支援布林邏輯，如一般常用的交集 (AND)、聯集 (OR)、差集 (NOT)，使用這些邏輯運算子時需使用大寫以作辨識，使用布林邏輯可將多個限定詞組合，提高搜尋結果的品質。為了能獲得更精準且符合需求的結果，避免包含過多全基因體 (complete genome) 與全染色體序列 (complete chromosome sequence) 的資料。本研究使用檢索限定詞，以基因名稱的 [GENE] 與標題說明 [Title] 組合為檢索字串，進一步篩選以提高效率。某些基因的公開資料較少，故僅用 [GENE] 為限定詞；部分基因如 APH4 (hygromycin-B phosphotransferase) 使用 [GENE] 限定詞查詢無資料，改用一般較廣泛的名稱 HPT (hygromycin phosphotransferase) 作為搜尋字串。

(二) GMO Blast Database 的更新與效能探討

隨著全球 GMO 作物的蓬勃發展，GMO 外來基因的研究資料也逐漸增加。以本研究 GMO 外來基因的 *accd* 基因為例，透過 NCBI Entrez 的限定詞 [MDAT] 查詢歷年公開序列累積筆數，歷年累計成長至 2007 年 7 月如圖 6。序列資料的增加快速成長，資料的蒐集也需同步更新，以免發生資料過時，而降低資料庫的實用性。

本研究建立的 GMO Blast Database 系統在序列蒐集部分，於程式設計之初已將序列更新列入考

討論

(一) NCBI Entrez 之檢索技巧

量。更新步驟可以利用批次檔 (Batch file)，將 find_AN.pl、get_GM_Seq.pl 與 formatdb.exe 這三支程式集合而成單一檔案。批次檔就是將一群可以在 DOS 下執行的命令 (Command) 集合而成的檔案，副檔名為「.bat」。透過批次檔將所有更新動作包裹成單一檔案 autoget.bat，只要執行 autoget.bat 即可將所有序列重新下載完成更新，並產生新資料庫取代舊資料，對系統維護人員非常方便。若有新增 GMO 外來基因，則僅需將新的基因檢索條件鍵入 gm_ID.txt，再執行 autoget.bat，使資料庫更新即可。本資料庫首次建置完成於 2007 年 5 月 5 日，內有 2447 條序列，設定一個月更新一次，2007 年 7 月 5 日為 2489 條序列，新增 42 條序列。

以往蒐集序列是從 NCBI 的查詢結果中勾選，用人工的方式來來逐一判斷是否符合需求，需要花費大量的時間與人力才能完成。以上述 *accd* 基因為例，自 NCBI Entrez 鍵入 *accd* 搜尋，結果共計 1798 條序列，決定該條序列是否含有 *accd* 基因序列，必須逐筆檢視是否於注釋欄位中包含 *accd* 基因，若有再回上一頁勾選該條序列，最後再將全部勾選的序列匯出 Fasta 檔。整個作業共需檢視 1798 條序列資料，選取 681 條，若不考慮網路連線速度，以檢視一條序列是否會有目標基因需 30 秒計算，共花費 900 分鐘，本資料庫共計 48 個基因，需檢視超過 20,000 條序列，選出 2247 條序列，費時超過 166 個工作天，若考慮連線速度與之後的資料庫建置，所花費的人力物力非常驚人。透過本研究建立的自動擷取序列與建立資料庫系統，只需要 4~5 小時(視連線狀況)即可完成，若使用作業系統的排程設定，將更新建置的時間設定在凌晨，研究人員不需要花費任何工作天數即可完成，此項自動更新資料庫也是本系統的一項特色。

參考文獻

[1] Baxevanis A.D. and Ouellette B.F.F. (2005), A Practical Guide to the Analysis of Genes and Proteins, Bioinformatics.

[2] Chiter, A., Forbes, J. M., and Blair, G.

E.(2000), DNA stability in plant tissues: implications for the possible transfer of genes from genetically modified food, FEBS Letters, 481, 164-168.

[3] Conner AJ, Glare TR, Nap JP. (2003), The release of genetically modified crops into the environment, Overview of ecological risk assessment Plant J, 33, 19~46.

[4] Fagan, J., Schoel, B., Haegert, A., Moore, J., Beeby, J. (2001), Performance assessment under field conditions of a rapid immunological test for transgenic soybeans, Int. J. Food Sci. Technol, 36, 357-367.

[5] Huang, H. Y. and Pan, T. M.(2004), Detection of genetically modified maize MON810 and NK603 by multiplex and real-time polymerase chain reaction methods, J. Agric. Food Chem, 52, 3264-3268.

[6] Jakubowicz M (2002), Structure, catalytic activity and evolutionary relationships of 1-aminocyclopropane- 1-carboxylate synthase, the key enzyme of ethylene synthesis in higher plants, Acta Biochim Polon, 49, 757-774.

[7] James, C. (2007), Global status of commercialized biotech/GM crops, ISAAA Briefs 35.

[8] James, C.(2006),Global status of commercialized biotech/GM crops, ISAAA Briefs 34.

[9] Lee, M.K., et al. (2003), The mode of action of the Bacillus thuringiensis vegetative insecticidal protein Vip3A differs from that of Cry1Ab –endotoxin, Appl. Environ.Microbiol, 69(8),4648–4657.

[10] Mori,S.,Kobayashi,H.,Hoshi,Y.,Kondo,M.,and Nakano,M.(2004), Heterologous expression of the flavonoid 3',5'-hydroxylase gene of Vinca major alters flower color in transgenic Petunia hybrida, Plant Cell Rep, 22, 415-421.

[11] Owen, M.D.K. (2000), Current use of transgenic herbicide-resistant soybean and corn in the USA., *Crop Protection*, 19, 765-771.

[12] Oliveira Souza (2000), "Genetically Modified Plants: A Need for International Regulation", 6 Annual Survey

[13] Randal L. Schwartz and Tom Phoenix, "Learning Perl", O'REILLY, 2002, ISBN:957-8247-89-3

[14] Robbins-Manke JL, Zdraveski ZZ, Marinus M, Essigmann JM. (2005), Analysis of global gene expression and double-strand-break formation in DNA adenine methyltransferase- and mismatch repair-deficient *Escherichia coli*. *J Bacteriol*, 187, 7027-37.

[15] Stajich JE et. al. (2002), The Bioperl toolkit: Perl modules for the life sciences, *Genome Res*, 12(10), 1611-8.

[16] Spaar, A., Dammer, C., Gabdoulline, R.R., Wade, R.C., Helms, V. (2006), Diffusional encounter of barnase and barstar. *Biophys. J*, 90, 1913-1924.

[17] Ussuf K.K., Laxmi N. H. and Mitra R. (2001) Proteinase inhibitors: Plant-derived genes of insecticidal protein for developing insect-resistant transgenic plants, *Curr. Science*, 80, 847-853.

of International & Comparative Law 138-139 ◦

表 1 目標序列基因一覽表

編號	Gene code	所 coding 的蛋白質或基因名稱	類別
1	<i>aad</i>	3" (9)-0-aminoglycoside adenyltransferase	抗生素
2	<i>acc</i>	1-amino-cyclopropane-1-carboxylic acid synthase (<i>Dianthus caryophyllus</i> L.)	果實成熟
3	<i>accase</i>	acetyl-CoA-carboxylase (<i>Z. mays</i>)	耐除草劑
4	<i>accd</i>	1-amino-cyclopropane-1-carboxylic acid deaminase (<i>Pseudomonas chlororaphis</i>)	果實成熟
5	<i>als</i>	acetolactate synthase (<i>Brassica napus</i>)	耐除草劑
6	<i>aph4(hpt)</i>	hygromycin-B phosphotransferase (<i>Escherichia coli</i>)	抗生素
7	<i>barnase</i>	barnase ribonuclease (<i>Bacillus amyloliquefaciens</i>)	雄性不育
8	<i>barstar</i>	barnase ribonuclease inhibitor (<i>Bacillus amyloliquefaciens</i>)	雄性不育回復
9	<i>bla</i>	beta lactamase	抗生素
10	<i>dhdps</i>	dihydrodipicolinate synthase (<i>Corynebacterium glutamicum</i>)	改變離胺酸
11	PRSV CP	viral coat protein (<i>Papaya ring spot virus</i> (PRSV))	病毒
12	CMV CP	viral coat protein (<i>Cucumber mosaic virus</i>)	病毒
13	ZYMV CP	viral coat protein (<i>Zucchini yellow mosaic virus</i>)	病毒
14	WMV2 CP	viral coat protein (<i>Watermelon mosaic virus 2</i>)	病毒
15	PVY CP	viral coat protein (<i>Potato potyvirus Y</i> (PVY) strain 0 (common strain))	病毒
16	<i>cry1Ab</i>	Cry1Ab delta-endotoxin (<i>Btk</i> HD-1) (<i>Bacillus thuringiensis</i> subsp. <i>kurstaki</i> (<i>Btk</i>))	抗蟲
17	<i>cry1Ac</i>	Cry1Ac delta-endotoxin (<i>Bacillus thuringiensis</i> subsp. <i>kurstaki</i> (<i>Btk</i>))	抗蟲

18	<i>cry1F</i>	Cry1F delta-endotoxin (<i>Bacillus thuringiensis</i> var. aizawai)	抗蟲
19	<i>cry2Ab</i>	Cry2Ab delta-endotoxin(<i>Bacillus thuringiensis</i>)	抗蟲
20	<i>cry34Ab1</i>	Cry34Ab1 delta-endotoxin (<i>Bacillus thuringiensis</i> strain PS149B1)	抗蟲
21	<i>cry35Ab1</i>	Cry35Ab1 delta-endotoxin (<i>Bacillus thuringiensis</i> strain PS149B1)	抗蟲
22	<i>cry3A</i>	cry3A delta-endotoxin (<i>Bacillus thuringiensis</i> subsp. Tenebrionis)	抗蟲
23	<i>cry3Bb1</i>	cry3Bb1 delta-endotoxin (<i>Bacillus thuringiensis</i> subsp. kumamotoensis)	抗蟲
24	<i>cry9c</i>	cry9c delta-endotoxin (<i>Bacillus thuringiensis</i> subsp. Tolworthi)	抗蟲
25	<i>dam</i>	DNA adenine methylase (<i>Escherichia coli</i>)	雄性不育
26	<i>dfr</i>	dihydroflavonol reductase (<i>Petunia hybrida</i>)	花色
27	<i>epsps</i>	5-enolpyruvylshikimate-3-phosphate synthase (<i>Agrobacterium tumefaciens</i> CP4)	耐除草劑
28	<i>fad2</i>	fatty acid desaturase	改變脂肪酸
29	<i>hfl</i>	flavonoid 3p, 5p hydroxylase (<i>Petunia hybrida</i>)	花色
30	<i>gmfad2-1</i>	delta(12)-fatty acid dehydrogenase (<i>Glycine max</i>)	改變脂肪酸
31	<i>gox</i>	glyphosate oxidoreductase (<i>Ochrobactrum anthropi</i>)	耐除草劑
32	<i>gus</i>	beta-D-glucuronidase (<i>Escherichia coli</i>)	染色標記 報導基因
33	<i>hel</i>	helicase (<i>Potato leafroll luteovirus</i> (PLRV) orf 2)	病毒
34	<i>mcry3A</i>	Cry3A delta-endotoxin (<i>Bacillus thuringiensis</i> subsp. tenebrionis)	改良型抗蟲

		nitrilase	
35	<i>nitrilase</i>	(<i>Klebsiella pneumoniae</i> subspecies ozanae)	耐除草劑
		nopaline synthase	調控基因
		(<i>Agrobacterium tumefaciens</i>)	(Promoter)
36	<i>nos</i>	<i>A. tumefaciens nopaline synthase (nos)</i> <i>3'-polyadenylation signal</i>	調控基因 (Terminator)
37	<i>neo</i> (<i>nptII</i>)	neomycin phosphotransferase II (<i>Escherichia coli</i>)	抗生素
38	<i>pat</i>	phosphinothricin N-acetyltransferase (<i>S. viridochromogenes</i>)	耐除草劑
39	<i>pg</i>	polygalacturonase (Tomato)	軟化
40	<i>pinII</i>	protease inhibitor (<i>Solanum</i> <i>tuberosum</i>)	抗蟲
41	<i>pmi</i>	Phosphomannose isomerase	代謝，可作為 篩選標記
		nicotinate-nucleotide	
42	<i>qtpase</i>	pyrophosphorylase(carboxylating) (<i>Nicotiana tabaccum</i>)	改變尼古丁
		replicase (RNA dependent RNA polymerase)	
43	<i>rep</i>	(<i>Potato leafroll luteovirus</i> (PLRV) orf 1)	抗病毒
44	<i>samase</i>	S-adenosylmethionine hydrolase (<i>E. coli bacteriophage T3</i>)	延緩成熟
45	<i>spc</i>	Spectinomycin/streptomycin resistance	抗生素
46	<i>bay te</i>	thioesterase (<i>Umbellularia</i> <i>californica</i>)	降低月桂酸與 豆蔻酸
47	<i>vip3A</i>	VIP3A vegetative insecticidal protein (<i>Bacillus thuringiensis</i> strain AB88)	抗蟲
			調控基因
48	35S	35S RNA of <i>cauliflower mosaic virus</i>	(Promoter) 調控基因 (Terminator)

表 2 基因檢索字串總表與取回筆數

編號	Gene code	檢索字串	搜尋筆數
1	<i>aad</i>	Aad\[Title\] AND Aad\[gene\]	9
2	<i>acc</i>	ACC\[Title\] AND ACC \[gene\]	33
3	<i>accase</i>	ACCcase\[Title\] AND ACCCase\[gene\]	61
4	<i>accd</i>	ACCd\[Title\] AND ACCd \[gene\]	681
5	<i>als</i>	Als\[Title\] AND Als \[gene\]	87
6	<i>aph4(hpt)</i>	Hpt\[Title\] AND Hpt\[gene\]	22
7	<i>barnase</i>	Barnase\[gene\]	2
8	<i>barstar</i>	Barstar\[gene\]	1
9	<i>bla</i>	Bla\[Title\] AND Bla\[gene\]	114
10	<i>dhdps</i>	DHDPS\[Title\] AND DHDPS\[gene\]	4
11	<i>PRSV CP</i>	PRSV AND CP\[gene\]	149
12	<i>CMV CP</i>	CMV AND CP\[gene\]	244
13	<i>ZYMV CP</i>	ZYMV AND CP\[gene\]	11
14	<i>WMV2 CP</i>	WMV 2 AND CP\[gene\]	2
15	<i>PVY CP</i>	PVY AND CP\[gene\]	142
16	<i>cry1Ab</i>	cry1Ab\[gene\]	7
17	<i>cry1Ac</i>	cry1Ac\[gene\]	11
18	<i>cry1F</i>	cry1F\[gene\]	0
19	<i>cry2Ab</i>	cry2Ab\[gene\]	5
20	<i>cry34Ab1</i>	cry34Ab1\[gene\]	0
21	<i>cry35Ab1</i>	cry35Ab1\[gene\]	0
22	<i>cry3A</i>	cry3A\[gene\]	0
23	<i>cry3Bb1</i>	cry3Bb1\[gene\]	0
24	<i>cry9c</i>	cry9c\[gene\]	0
25	<i>dam</i>	Dam\[Title\] AND Dam\[gene\]	16
26	<i>dfr</i>	DFR\[Title\] AND DFR\[gene\]	188
27	<i>epsps</i>	EPSPS\[gene\]	21
28	<i>fad2</i>	fad2\[Title\] AND fad2\[gene\]	80
29	<i>hfl</i>	F3'5'H\[gene\]	2
30	<i>gmfad2-1</i>	\ "Glycine max" AND Fad2-1\[gene\]	2
31	<i>gox</i>	Gox\[Title\] AND Gox\[gene\]	8
32	<i>gus</i>	Gus\[gene\]	30
33	<i>hel</i>	helicase [title]AND Hel[gene]	78
34	<i>mcry3A</i>	mcry3A[gene]	0

35	<i>nitrilase</i>	Nitrilase\[Title\] AND Nitrilase\[gene\]	1
36	<i>nos</i>	Nos\[Title\] AND Nos\[gene\]	81
37	<i>neo(npt II)</i>	neomycin AND NPTII\[gene\]	128
38	<i>pat</i>	PAT\[Title\] AND PAT\[gene\]	71
39	<i>pg</i>	PG[<i>gene</i>]	32
40	<i>pin II</i>	pin-II2x[<i>gene</i>]	2
41	<i>pmi</i>	pmi\[Title\] AND pmi\[gene\]	11
42	<i>qtpase</i>	QPT[<i>gene</i>]	6
43	<i>rep</i>	replicase [title]AND Rep[<i>gene</i>]	44
44	<i>SAMase</i>	SAM[<i>gene</i>]	15
45	<i>spc</i>	spectinomycin AND Spc\[gene\]	18
46	<i>bay te</i>	thioesterase AND te\[gene\]	1
47	<i>vip3A</i>	vip3A\[gene\]	23
48	<i>35S</i>	CaMv AND 35s\[gene\]	4
	總計		2447

表 3 Entrez 常用檢索限定詞

索引標籤	說明	檢索限定詞
All Fields	包含所有搜尋欄位	[ALL] or [ALL FIELDS]
Accession	序列登錄碼	[ACCN] or [ACCESSION]
Gene Name	基因名稱	[GENE]
Keyword	關鍵字	[KYWD] or [KEYWORD]
Properties	屬性	[PROP]
Organism	物種名稱	[ORGN] or [ORGANISM]
Title	每筆資料的標題說明行	[Title]

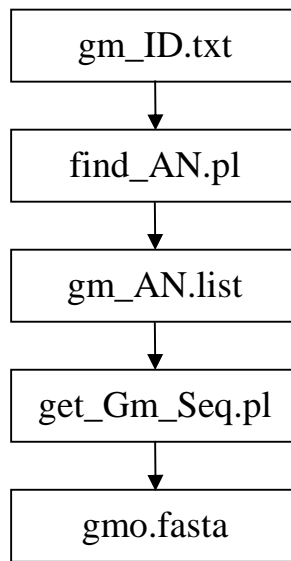


圖 1 序列擷取程式執行流程

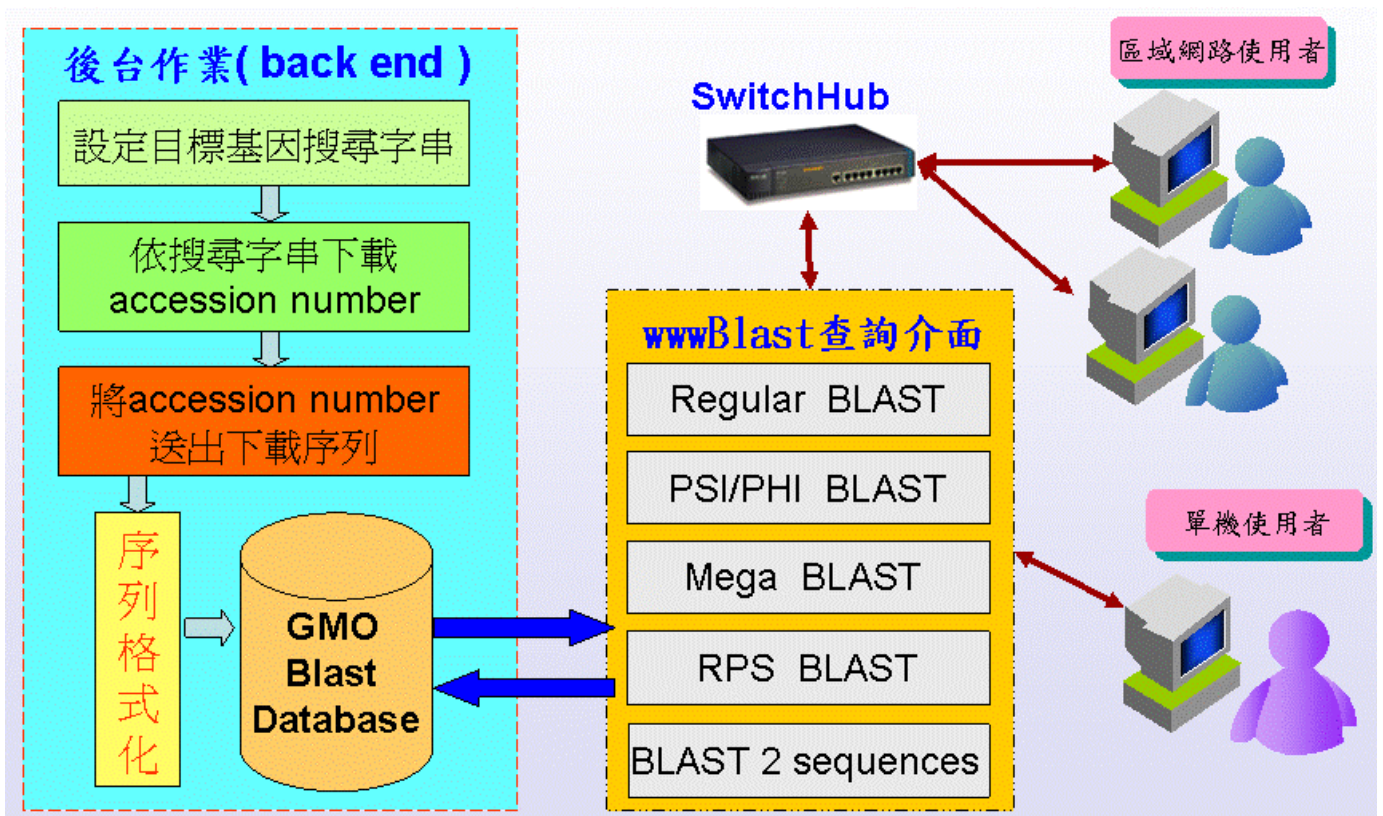


圖 2 GMO BLAST Database 系統架構圖

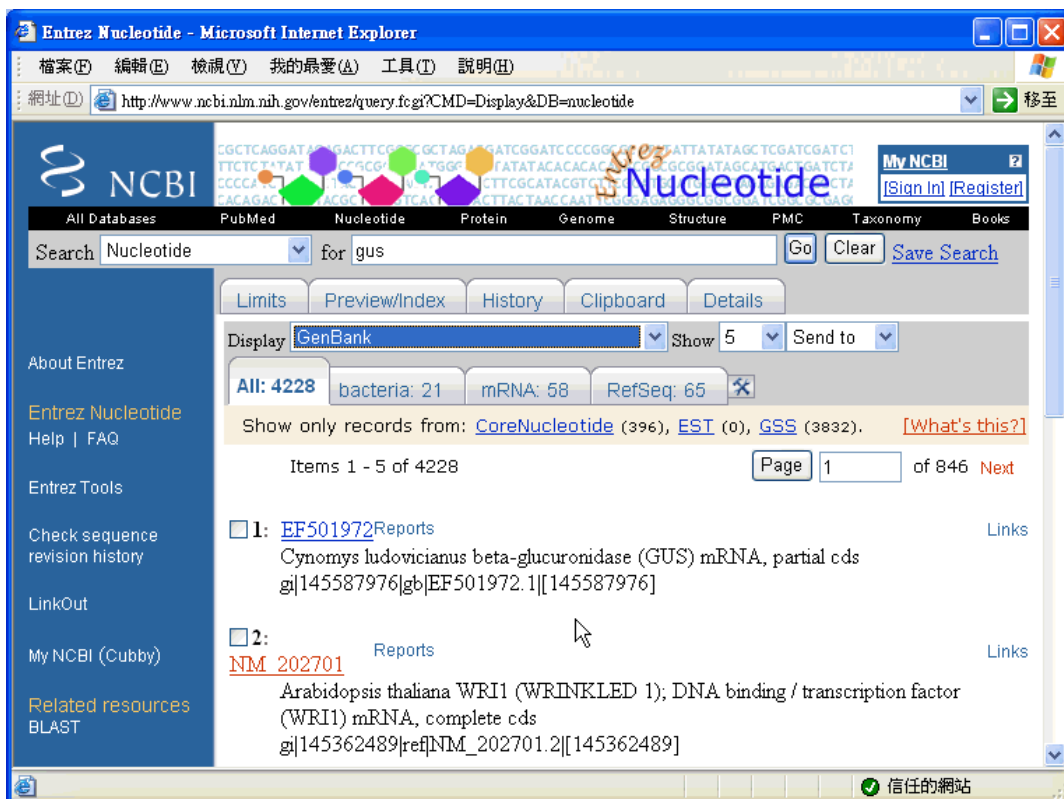


圖 3 NCBI 全文檢索 gus(β -glucuronidase)基因

說明:使用一般全文檢索，查詢 GUS 可得到 4228 條序列

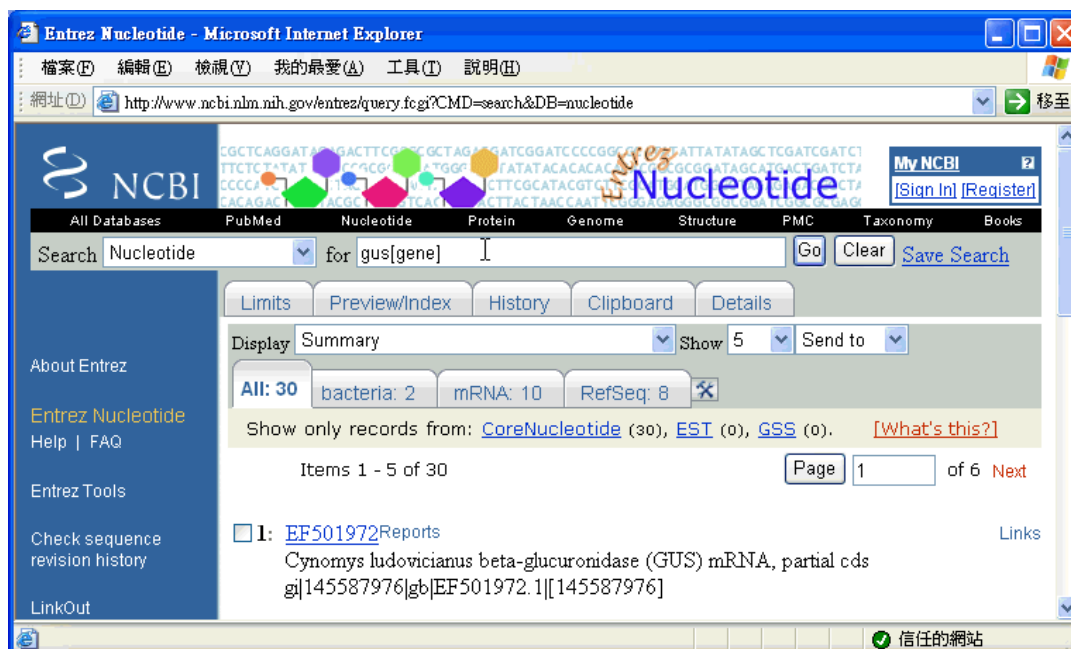


圖 4 使用限定詞檢索 gus 基因

說明:使用 GUS[GENE]為檢索字串，搜尋結果僅 30 條序列

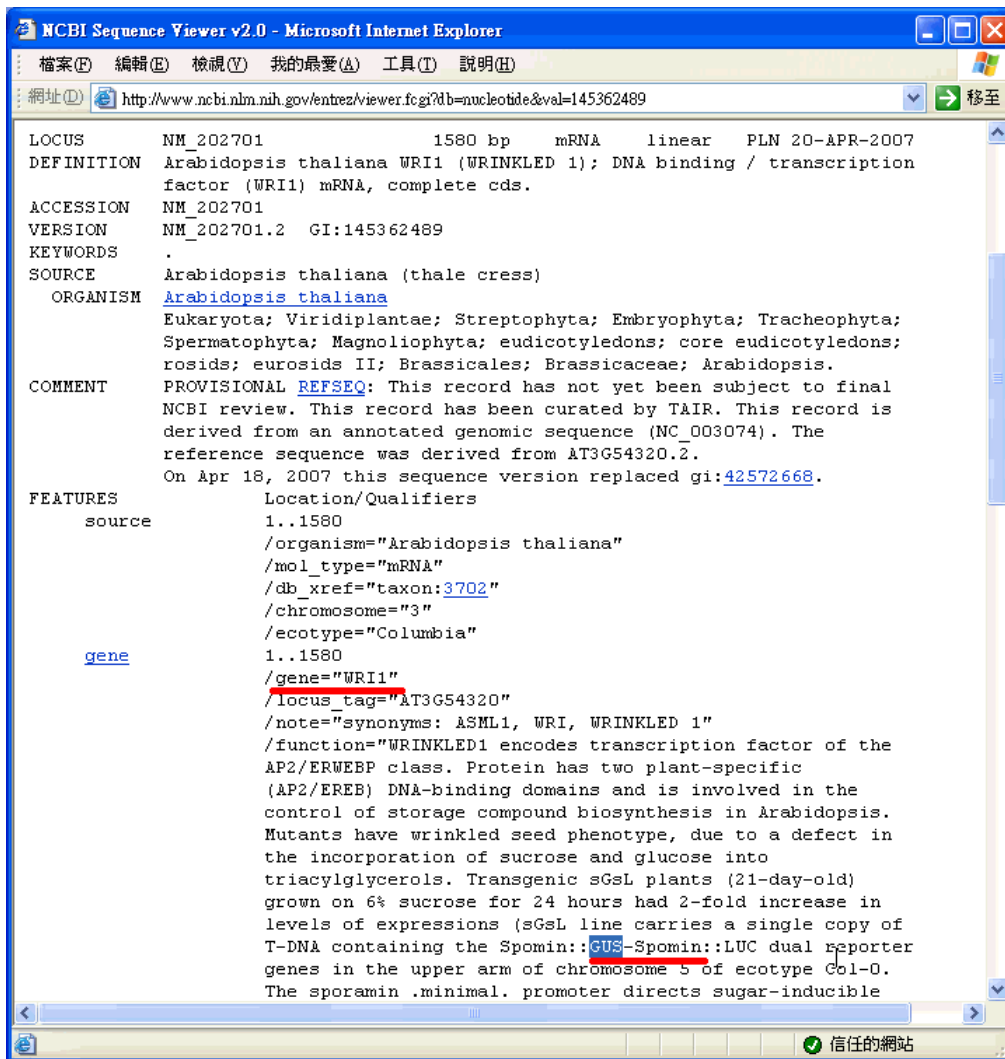


圖 5 非目標序列 NM_202701

說明：GUS 檢索結果中，NM_202701 的序列為 WRI1 基因，非檢索目標找尋的 GUS 基因。

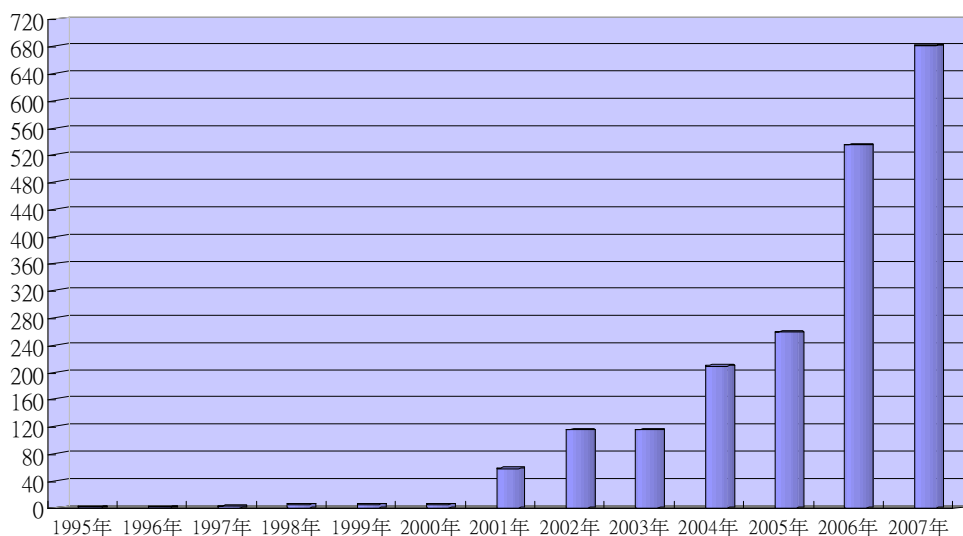


圖 6 NCBI 的 accd 序列資料歷年累計成長筆數