

國立交通大學

生物資訊及系統生物研究所

碩士論文

**MicroOrganism Identifier: 微生物有效識別法
之網頁流程系統**

**Microorganism Identifier: A web service to
identify microorganism**

研究生：王煒驊

指導教授：林勇欣 博士

中華民國一百零一年八月

MicroOrganism Identifier:微生物有效識別法 之網頁流程系統

Microorganism Identifier: A web service to identify microorganism

研究生：王煒驊

Student : Wei-Hua Wang

指導教授：林勇欣 博士

Advisor : Yeong-Shin Lin



A Thesis Submitted to Institute of Bioinformatics and Systems Biology
National Chia Tung University in partial Fulfillment of the Requirements
For the Degree of Master in
Bioinformatics and Systems Biology
August 2012

Hsinchu, Taiwan, Republic of China

中華民國一百零一年八月

MicroOrganism Identifier:微生物有效識別法之網頁流程系統

學生：王煒驊

指導教授：林勇欣

國立交通大學生物資訊及系統生物碩士班

摘要

微生物在環境中扮演著重要的角色，不論是疾病、醫藥到工業都與微生物息息相關。因此要如何有效的鑑定微生物，成為一個很重要的課題。但是一般的鑑定方法需要花費大量的時間與昂貴的設備，為了快速且有效地鑑定微生物，在本研究中，我們開發了一個網頁流程系統，利用 structural RNA gene 間的距離作為一項 genomic feature，將這些 structural RNA 的 conserved region 設計成引子，憑藉著在不同基因體上引子的 binding site 間距不同來做鑑定的標準，最後利用演算法得到最佳解。

Microorganism Identifier: A web service to identify microorganism

Student: Wei-Hua Wang

Advisor: Yeong-Shin Lin

Institute of Bioinformatics and Systems Biology

National Chiao Tung University

Abstract

Microorganisms play an important role in the environment, and is closely related to disease ,drug and industry. In recent years, how to identify microorganisms effectively has become a very important issue. However the general identification methods often take a lot of time and resource. In order to identify microorganism quickly and cost-efficiently, we have developed a web service base on a genomic feature .We found that intergenic length between two neighboring structural RNAs are different between genome. Base on this feature ,we will design primer pairs by the conserved region of structural RNAs. And the primers will identify microorganism by PCR amplified fragment lengths. Finally, we will use the computational method to obtain the optimal solution.

謝 誌

兩年的研究所生涯，一路上受到許多老師與同學的幫助。首先，我要感謝林勇欣老師，老師在研究的過程中總是耐心的教導我，給予我許多精闢的建議，讓我在生資的領域從一開始什麼都不會到現在學了許多寶貴的知識。總是激發我的思考，並且指出問題所在，讓我在這兩年學到了許多。

在學習的過程中，感謝智欽學長、俊霖學長與羿喬學長總是當我在學習的困難中，向我伸出援手，並且總是極有耐心的與我一同討論問題，讓我總是非常抱歉花費他們這麼多珍貴的時間。還有同學之杭，不管在生活上或學業上，有人能一起努力，互相幫助，給予我許多的動力。

感謝楊進木老師、盧錦龍老師在口試的時候給予我寶貴的建議，並且提出更多面向的思考觀點，讓我能更進一步審視自己與系統，加以改進。

最後我要感謝一直支持我的家人與朋友，總是不吝給我關心與幫助，你們的支持讓我更能肯定自己的想法與目標。

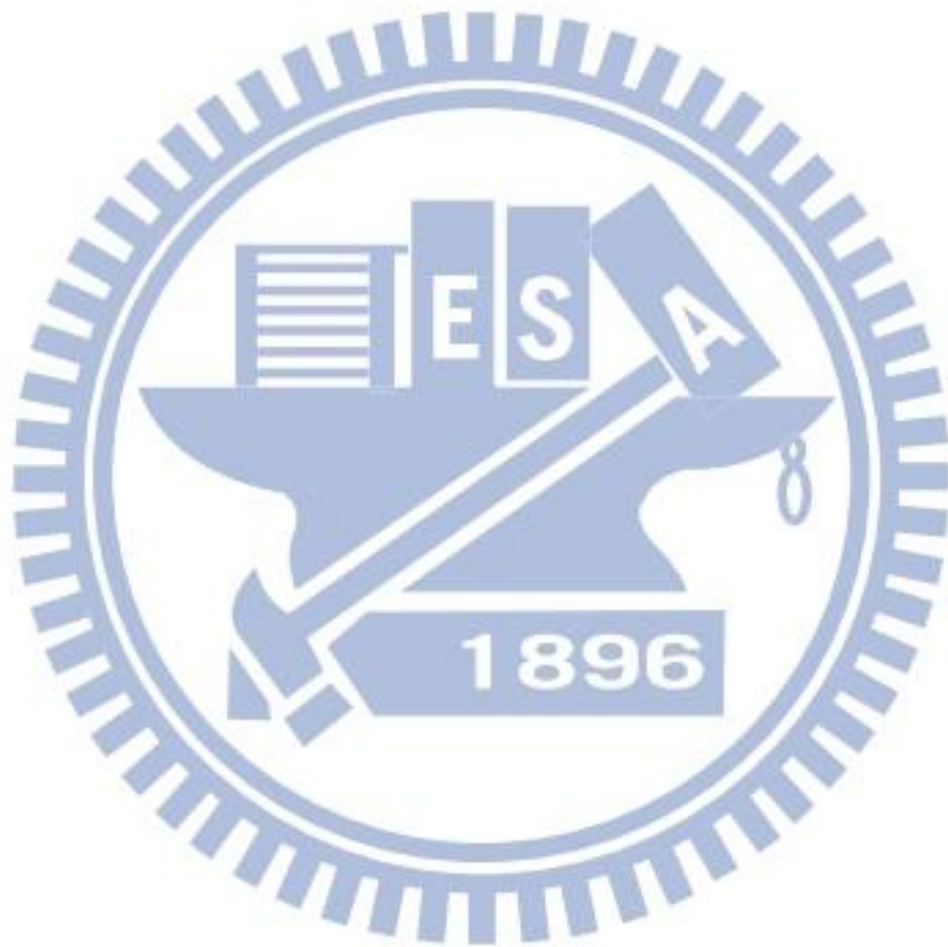
內文目錄

一、前言.....	1
1.1 研究動機.....	1
1.2 鑑定方法.....	2
1.3 概念簡述.....	3
1.4 研究成果.....	4
二、方法.....	5
2.1 資料庫.....	5
2.2 Structural RNA 的註解.....	5
2.3 挑選 conserved region.....	6
2.4 引子設計.....	7
2.5 組合分析.....	8
三、網頁實作.....	10
3.1 網頁工具概述.....	10
3.2 系統的操作方式.....	10
3.3 系統的輸出格式.....	10
3.4 實作環境.....	11
四、結果.....	12
5.1 限制.....	14
5.2 與上一版本的改進.....	14
5.3 未來工作.....	15



表目錄

表 117



圖目錄

圖 1	Genomic feature	18
圖 2	系統流程圖	19
圖 3	structural RNA 的註解	20
圖 4	篩選 structural RNA pair.....	21
圖 5	挑選具有鑑別力的 structural RNA pair.....	22
圖 6	引子設計流程(1)	23
圖 7	引子設計流程(2)	24
圖 8	判斷是否黏合	25
圖 9	搜尋路徑(1)	26
圖 10	搜尋路徑(2)	27
圖 11	系統首頁	28
圖 12	乳酸菌範例	29
圖 13	使用者上傳頁面	30
圖 14	結果圖	31
圖 15	引子對解(1)	32
圖 16	預測電泳圖(1)	32
圖 17	引子對解(2)	33
圖 18	預測電泳圖(2)	33
圖 19	引子對解(3)	34
圖 20	預測電泳圖(3)	34
圖 21	預測電泳圖(4)	35

一、前言

1.1 研究動機

在最近這半個世紀開始，由於生物科技的進步，人們漸漸發現微生物與我們的生活環境息息相關，除了一般所知的微生物在自然界有著的角色與行為外，也會直接或間接地影響動物與植物的疾病[1, 2]，另外微生物也是藥品[3]、工業品與食品的生產來源，因此，近代中有許多對微生物的研究與應用，例如，乳酸菌產產品(優格、優酪乳)[4]、酵母菌食品(麵包、酒類)[5]、工業酵素應用(清潔劑、動物飼料、紡織業)、能源與環保等等，都在探討微生物在自然界與人類環境中所扮演的角色。

由於微生物的多樣性與重要性，如何有效地鑑定成為了一個重要的課題，但是大部分的微生物鑑定都需要花費大量的資源與冗長的時間，因此，本研究的目的是在於建立一個低成本且快速的微生物鑑定工具。

1.2 鑑定方法

本研究的目的是在於建立一個低成本且快速的微生物鑑定工具。目前來說，微生物的鑑定方法主要分為表現型鑑定法與基因型鑑定法。表現型主要是利用生物表徵與生化反應等特徵[6, 7]，初期通常需要培養大量的菌株，在培養的過程易受環境變數的影響，因此培養不易且花費許多品質管理成本，另外還需要專業具有一定經驗的研究人員才能達到較準確的鑑定結果。

而相較於表現型鑑定法，基因型鑑定法結果較為簡易。基因型鑑定法分為兩種類型，第一種類型是根據大部分的微生物的序列都具有高度不變的特性(highly conserved)[8, 9]，例如 16S ribosomal RNA (rRNA)，常用來當作用來鑑定的研究目標[10-13]，而與其相關的基因型鑑定法有多位基因座序列分析法(Multilocus sequence typing, MLST)[14-17]。

另一種則為利用 genomic feature 來做鑑定，一般 DNA 中本身存在許多重複單位的序列，但會因為重複次數的不同，最後造成 DNA 長度的不同，利用這個特點，找出一些能切這些特殊重複片段的酵素，就可以依照切出來的片段長度不同作為鑑定的基準，或再去利用聚合酶連鎖反應與電泳進行分析。例如：限制切割片段長度多型性(Restriction Fragment Length Polymorphism, RFLP)[18-20]與增殖片段長度多型性(Amplified Fragment Length Polymorphism, AFLP)[21-23]、DNA 探針與核糖體基因分型分析(Analytical ribotyping)[24, 25]。

但這些基因型鑑定法大部份的鑑定研究都需要花費昂貴的資源與大量的時間，且為特定物種所設計[26, 27]，幾乎沒有辦法適用於其他微生物物種的鑑定，有鑑於此，我們希望開發一個快速又有效地微生物鑑定工具提供使用者使用。

1.3 概念簡述

在本研究中，我們提出一個新的基因型鑑定法，利用 **genomic feature** 的方式，與現階段已知被解碼的微生物序列，依靠它們之間的關係，去研究分析各個微生物物種之間的關聯。

首先，我們要事先得知要鑑定物種基因體序列，接著去找到它的 **structural RNAs**(ribosomal RNA 與 transfer RNA)在基因體上的位置，下一步利用這些 **structural RNAs** 在基因體上的相互之間的距離來判斷是否有鑑別力。換句話說，我們將 **structural RNA pair** 間的距離當作一個 **genomic feature**，利用同樣的 **structural RNA pair** 在不同基因體上有不同的距離，以此來鑑別物種。

我們利用這個概念開發了一個網頁服務工具來利用幫使用者做微生物的鑑定工作，讓使用者可以上傳自己感興趣目標基因體，利用這些基因體的 **structural RNA pair** 間距離的關係去設計聚合酶連鎖反應引子，下一步執行凝膠電泳去驗證這些引子在目標基因體執行聚合酶連鎖反應的結果，並模擬其實際結果產物圖，讓使用者參考。最後利用上一版本的“**Combination analyzer**”演算法，找出能有效鑑定目標基因體的最佳的引子對組合，並將結果利用電子郵件傳送給使用者。

1.4 研究成果

前一版的 MOI 系統是利用 *Lactobacillus* 屬為例[28]，設計出一個能鑑別乳酸菌的網頁工具，其中共有 17 株 *Lactobacillus* 菌株供使用者選擇。

本系統服務為前一版本 MOI 的衍生，除依舊保存 *Lactobacillus* 的範例供使用者參考外，另外提供了使用者上傳基因體的服務，還幫這些目標基因體做 structural RNA 註解的服務，改進上一版本需要準備目標基因體外，還需要其 structural RNA 註解的不便外，其中演算法與介面也經過改進，例如：利用 Electronic PCR(E-PCR)[29, 30]將原本運算時間冗長的方法置換，將其執行時間減省數倍；最後的電泳結果，利用圖表的呈現，讓使用者更能了解最後的模擬結果。

在本研究的最後，以 *Bacillus* 屬的菌株為例，驗證整個系統流程的執行程序與最後電泳的結果。

二、方法

2.1 資料庫

圖 2 所示為本網頁服務的系統流程，在網頁介面中，使用者可以使用基因體上傳的服務，依照使用者的個人需求，對所需要鑑別的微生物基因體，在網頁服務中上傳至本系統的資料庫。而系統會為了後續的系統需求及分析，將使用者上傳的目標基因體，經由後端程式的處理後，建立 MySQL 的關聯式資料庫的相關欄位及表格，以供後續的程式運算及分析使用。

2.2 Structural RNA 的註解

在本系統當中，引子設計是利用 structural RNA 來做引子分析的參考，然而，由於 structural RNA 在不同基因體上註解的不同，會影響到後續程式設計引子的結果。因此，當使用者上傳基因體後，我們會先將這些基因體序列做註解，標出 structural RNA 的相關資訊。這階段會根據使用者上傳的基因體序列分別做兩種註解：

(1)使用 NCBI Blastn 來為使用者上傳的基因體序列找出其 16S rRNA，Blastn 比對的資料庫為 NCBI 16SMicrobial 核酸序列資料，相似度門檻設為 e-Value 小於等於 10^{-5} ，

得到的比對結果依照 score 由高到低做排列，取分數最高但不重複的基因體序列，範圍為 16S RNA 註解範圍，如圖 3 所示。

(2)使用 tRNAscan-SE[31, 32]工具來為使用者上傳的基因體序列設計 tRNA 的註解，它的功能是能根據輸入的基因體序列進行處理，找出輸入基因體的 tRNA。

2.3 挑選 conserved region

根據上一步得到的註解資訊，分析使用者感興趣的基因體序列群組，找出這些基因體中 structural RNA 所在的位置，並且依據這些 structural RNA pair 在基因體上的間距(聚合酶連鎖反應的產物長度大約都在 100 bp 到 3000bp)，來做篩選的標準之一，如圖 4 所示。另外根據這些 structural RNA pair 在不同基因體上間距的差異，當作 genomic feature，若其差異未超過 100bp 以上或超過 3000bp 以上，則被視為不具鑑別力而排除，以圖 5 為例。

之後，在所有的目標基因體序列群組中，挑選出有間距差異相同的 structural RNA pair 的片段(具有鑑別力的 structural RNA pair)，並使用多重序列比對(multiple alignment)來找出所有 structural RNA pair 的 conserved region[33]，將這些結果做為設計引子的參考，流程如圖 6 所示。

2.4 引子設計

在得到 conserved region 後，由於需要 structural RNA pair 做為區別不同目標基因體，因此若是同類型的 structural RNA 得到的引子可以在其它不同的基因體上適用，我們則為它設計引子，稱之 common primer，若不能適用在大多數的基因體，則會再為其餘的 structural RNA 設計引子，此時稱為 specific primer，如圖 7 所示。

我們利用設計引子的工具為 Primer Hunter。Primer Hunter 除了能設計引子之外，還能預測是否黏合與預測黏合溫度等功能，但在本研究中僅利用 Primer Hunter 來設計引子。[34]

其相關設定如下：

- (1)保守序列的最小長度不得低於 16 bp。
- (2)保守序列中最多只能容許 3 個 mismatches。
- (3)一個 structural RNA 可能有 1 個至多個引子可以使用。
- (4) PrimerHunter 的參數設定：

Melting temperature (T_m): 60°-63°

Product length: 100-2000 bp

2.5 組合分析

在完成引子設計之後，我們得到許多針對目標基因體群具有鑑別能力的引子對，為了找出所有引子對目標基因體可能有非專一性的 binding site，在此我們使用 E-PCR 中的 Reverse-PCR 功能。此工具目的是利用雜湊(Hash)的方法，在目標基因體中，快速找出與引子相似的區域，將所有可能的 binding site 存入資料庫中。

E-PCR 的參數設定:

Mismatch (N): 5

Gap(G): 3

下一步，將所有的可能的 binding site 與引子進行 Tm 值的運算，其利用公式為 Melting Temperature $T_m = 81.5 + 0.41(\%GC) - 675/Length[35]$ ，得出的 Tm 值可用來判斷是否能與 binding site 黏合，如圖 8 所示。在 E-PCR 的過程中，我們略過目標引子對可能接合的片段中有過多 Insert 與 Delete 的搜尋，因為大量片段的 Insert 或 Delete 片段會使 Tm 值大量的降低，使得引子無法與其 binding site 黏合。

將引子在所有可能的 binding site 資訊存入資料庫，並建立索引表，已供後續搜尋與分析，重複以下兩個步驟:

- (1) 從搜尋索引表中找出只能由少數的引子對才能鑑別的組合，並記錄其搜尋路徑，如圖 9 所示。
- (2) 確認搜尋索引表的鑑別資訊，將(1)找到的可鑑別引子對從索引表中移除。並重複(1)，如圖 10 所示。

直至找到可以鑑別所有目標基因體的最佳引子對搜尋路徑(可能不是唯一，會有多組解)，並且也提供其他替補的最佳引子對搜尋路徑，供使用者參考選擇。



三、 網頁實作

3.1 網頁工具概述

本網頁系統分為兩部分，圖 11 為首頁。

- (1) 通過乳酸菌範例讓使用者從乳酸菌資料庫中選取菌株，並設定參數得到引子組合結果(圖 12)。
- (2) 讓使用者上傳微生物基因體序列，設計具有鑑別力的聚合酶連鎖反應引子。

3.2 系統的操作方式

本系統的操作畫面如圖 13 所示，使用者可以自行上傳想要鑑定的微生物基因體序列，在設定完相關參數後，填入使用者聯絡信箱與專案名稱，當系統分析完成後，系統會寄信通知使用者。其詳細參數預設值分別為 Tm Threshold: 5°及 distance threshold: 50 bp。

3.3 系統的輸出格式

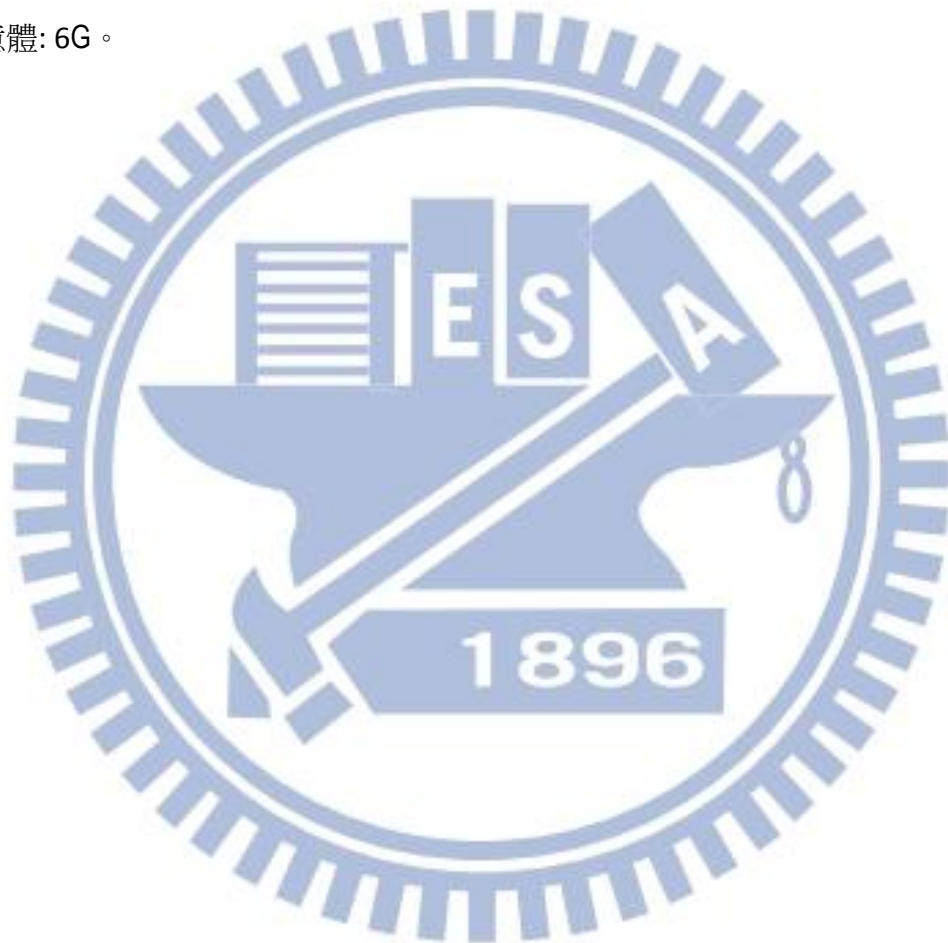
當系統完成鑑定工作之後，使用者會收到通知信函，可利用內嵌的網頁超連結，得到最佳引子組合(圖 14)與引子的相關資訊: 聚合酶連鎖反應產物的長度與

Tm 值、電泳的預測圖，如圖 15。

3.4 實作環境

MOI 是使用 PHP 所開發，伺服器作業系統:CentOS 5.4，CPU: AMD Phenom II X4 945，

記憶體: 6G。



四、結果

在這個研究中，我們到 NCBI 下載 *Bacillus* 與 *Lactobacillus* 的基因體來做系統的微生物物種鑑定測試，分別是: (1) *Bacillus subtilis* subsp. *subtilis* str. 168。 [36](2) *Bacillus subtilis* subsp. *spizizenii* TU-B-10。 [37](3) *Bacillus subtilis* subsp. *spizizenii* str. W23。 [38](4) *Bacillus anthracis* str. Ames[39]。 (5) *Bacillus atrophaeus* 1942。 (6) *Bacillus cellulosilyticus* DSM 2522。 (7) *Bacillus clausii* KSM-K16[40]。 (8) *Lactobacillus reuteri* JCM 1112。 [41](9) *Lactobacillus fermentum* IFO 3956[42]。將這些目標基因體上傳至系統做鑑定後，我們希望能通過系統設計出來的引子，利用最少的引子對組合，將這些基因體快速地鑑別出來。

此次運算時間大概處理 50 分鐘，最終得到的結果(圖 14)。這邊可以發現，本系統提供了許多引子對解提供使用者選擇，此範例用 3 組引子就可以鑑定出這九株(圖 15 至圖 20)，可從連結進入觀看詳細的結果，這邊可以看出由於 *Lactobacillus reuteri* JCM 1112 與 *Lactobacillus fermentum* IFO 3956 同為是 *Lactobacillus*，因此相較於其他 *Bacillus* 的基因體而言，它們之間能鑑定的聚合酶引子對較少(表 1)。換句話說，在兩個物種比較相近的基因體中，若要將其有效的鑑別，需要使用 *specific primer* 去鑑別，才能有效地將相近的物種作出區別。

在每一組引子對的聚合酶連鎖反應產物結果，我們分別會呈現出 $TM = 40$ 、45、50、55、60、65 六張不同 TM 的電泳預測圖，提供給使用者參考。這次實

驗的另一組引子對解中，其中有一組引子對隨著 TM 的變化，它能預測出的產物會隨著 TM 變化而有所不同，如圖 21 所示，在每一組引子對的聚合酶連鎖反應產物結果，黃色的線條代表預測會出現的產物，綠色則代表不是預期在這組 TM 會出現的產物，但有可能是雜訊或其他實驗原因所造產生，隨著溫度的上升，其產物越來越少，當 TM 超過 55 以後，就幾乎不會有產物的產生。這些數據圖像化的呈現，我們希望提供給使用者實際實驗的參考。



五、討論

5.1 限制

目前本系統提供了使用者上傳基因體序列的服務，但相較於基因體序列相比，要如何使用在這些序列上作正確 structural RNA 的註解並不是那麼容易。在本研究一開始我們將 23S rRNA 與 5S rRNA 列入影響設計引子的參數，但經過數次測試發現，因為在目前的 NCBI 資料庫中，其 structural RNA 在不同的基因體上存在著許多註解上的差異，尤其以 23S rRNA 最為嚴重，且因為 16S rRNA、5S rRNA、23S rRNA 在基因體上大部分都以 gene cluster 的方式聚集在相同的位置，不同 copy 間兩兩的距離差異往往很小，因此，為了之後設計引子的準確度與速度，在最後決定只對 16S rRNA 作註解的動作。另外，在 Tm 值的設定上，若使用者將聚合酶連鎖反應的 Tm 值區間設定過低，可能會造成許多候選引子無法列入考慮，最後造成能鑑定基因體的引子對組合過少，甚至最後可能會無法得出具有鑑別能力的引子對組合之情況發生，因此，一般建議 Tm 值差定在 10 左右。

5.2 與上一版本的改進

在上一版本中，未開放讓使用者上傳基因體的服務，在此版本中，除了改進讓使用者上傳基因體的客製服務外，也會對使用者所上傳的基因體序列作 tRNA

與 16S rRNA 作註解的動作。另外在候選引子對基因體有可能黏合片段上，原本是採取 *sliding window search* 的方法，由於此演算法的處理時間過於冗長，因此在本版本使用 E-PCR 工具中的 Reverse-PCR 功能，能利用雜湊的方法找出較快速地預測可能黏合片段，相較之前節省了數倍的時間。

另外，在最後的引子對在聚合酶連鎖反應黏合的片段資訊中，除了提供詳細的引子資訊與 T_m 值外，還另外呈現了聚合酶連鎖反應的結果預測圖，讓使用者在實際使用引子作聚合酶連鎖反應與電泳實驗時可以做參考。

5.3 未來工作

目前 MOI 在鑑定物種大都是依據 16s RNA 與 tRNA 兩種 structural RNA，未來可能會加入 *housekeeping gene*，以及提供使用者自行加入註解，希望能提供使用者更多引子對的選擇。在最後模擬電泳圖的結果呈現上，希望能準確的預測出當 T_M 變化時，模擬電泳圖也隨著 T_M 變化而造成產物形狀、顏色也會隨之不同。

。

六、結論

在這個研究中，我們利用了在不同物種之間，它們的 structural RNA 間距離不同作為一個 genomic feature，讓使用者可以利用的網頁介面，將想要鑑別的微生物群組上傳到系統網頁，系統會依照使用者的需求設計最佳引子對組，進而利用這些詳細的引子資訊去鑑別出使用者感興趣的微生物，而使用者可以得到詳細的引子資訊，與聚合酶連鎖反應產物詳細資料，且會呈現模擬電泳圖供使用者參考。



Organism	Sn	1	2	3	4	5	6	7	8	9
<i>Bacillus subtilis</i> subsp. <i>subtilis</i> str. 168	1	X	7	10	4	7	7	5	9	6
<i>Bacillus anthracis</i> str. Ames	2	X	X	7	5	7	8	8	6	9
<i>Bacillus clausii</i> KSM-K16	3	X	X	X	5	7	9	9	7	9
<i>Lactobacillus reuteri</i> JCM 1112	4	X	X	X	X	3	4	5	6	5
<i>Lactobacillus fermentum</i> IFO 3956	5	X	X	X	X	X	7	8	7	8
<i>Bacillus subtilis</i> subsp. <i>spizizenii</i> str. W23	6	X	X	X	X	X	X	3	9	3
<i>Bacillus atrophaeus</i> 1942	7	X	X	X	X	X	X	X	8	4
<i>Bacillus cellulosilyticus</i> DSM 2522	8	X	X	X	X	X	X	X	X	9
<i>Bacillus subtilis</i> subsp. <i>spizizenii</i> TU-B-10	9	X	X	X	X	X	X	X	X	X

表 1 引子對索引表



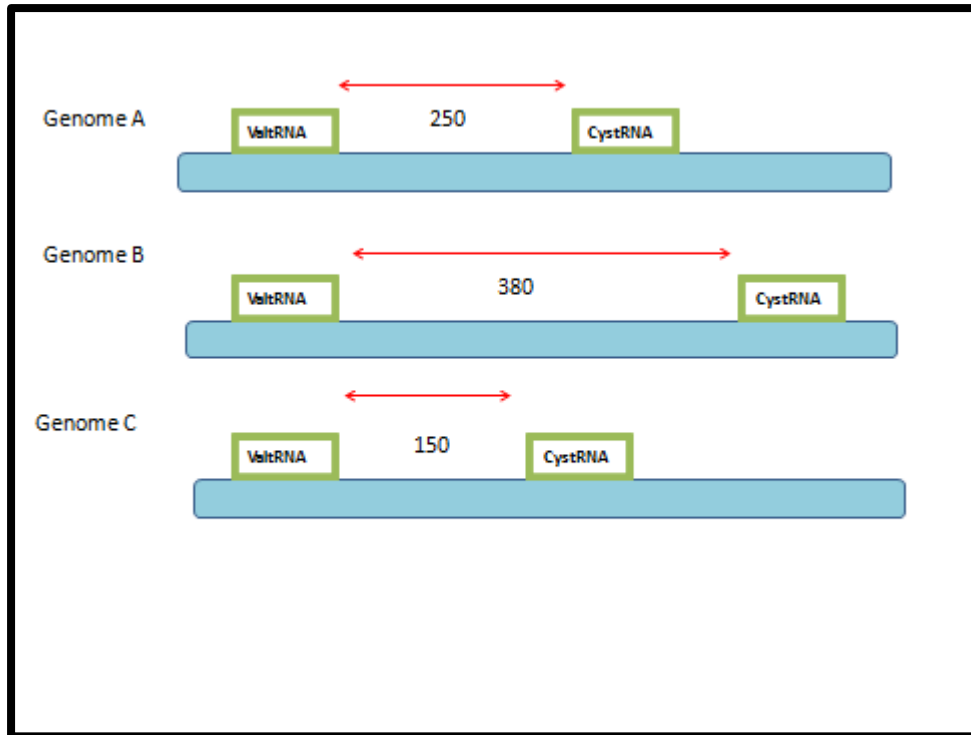


圖 1 Genomic feature

圖中三條 Genome 都具有 ValrRNA 與 CystRNA，故利用 ValrRNA 與 CystRNA 間的距離當作一項 Genomic feature，當這段距離在 A、B、C 三條不同 Genome 時，若彼此之間會有明顯的不同，就可以此來當作鑑別 Genome 的依據，這邊限制的差距必須在 50bp 以上，才會列入考慮。



圖 2 系統流程圖



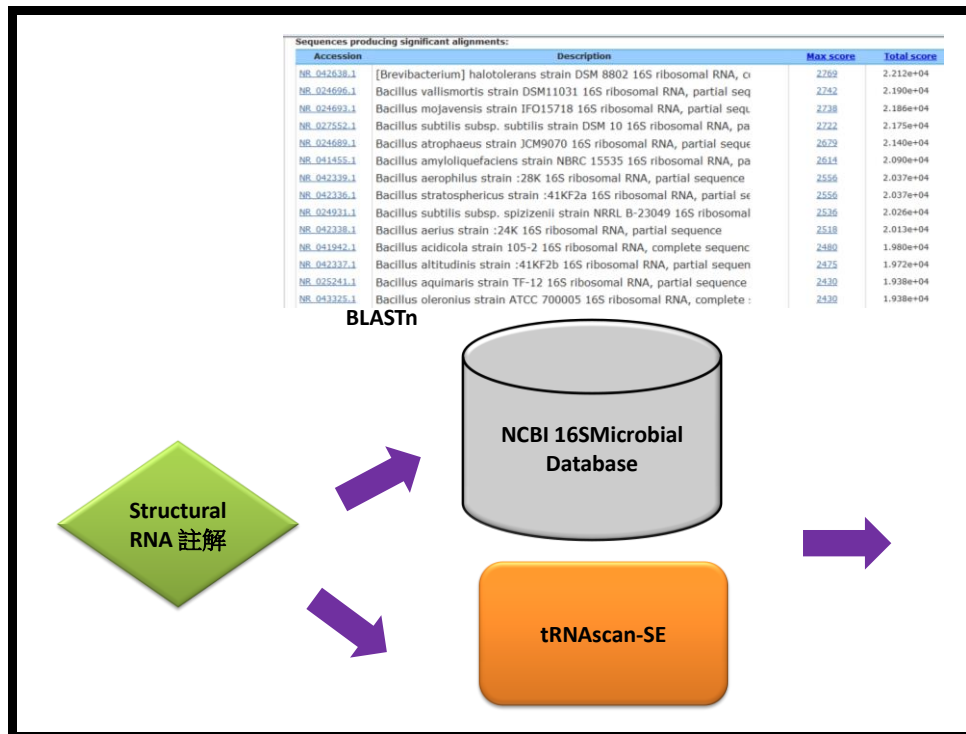


圖 3 structural RNA 的註解

Structural RNA 的註解分為兩部分(1)16S rRNA 部分利用 Blastn 比對 NCBI 的 16S Microbial Database (2)其餘 tRNA 利用 tRNAscan-SE 此工具作註解。

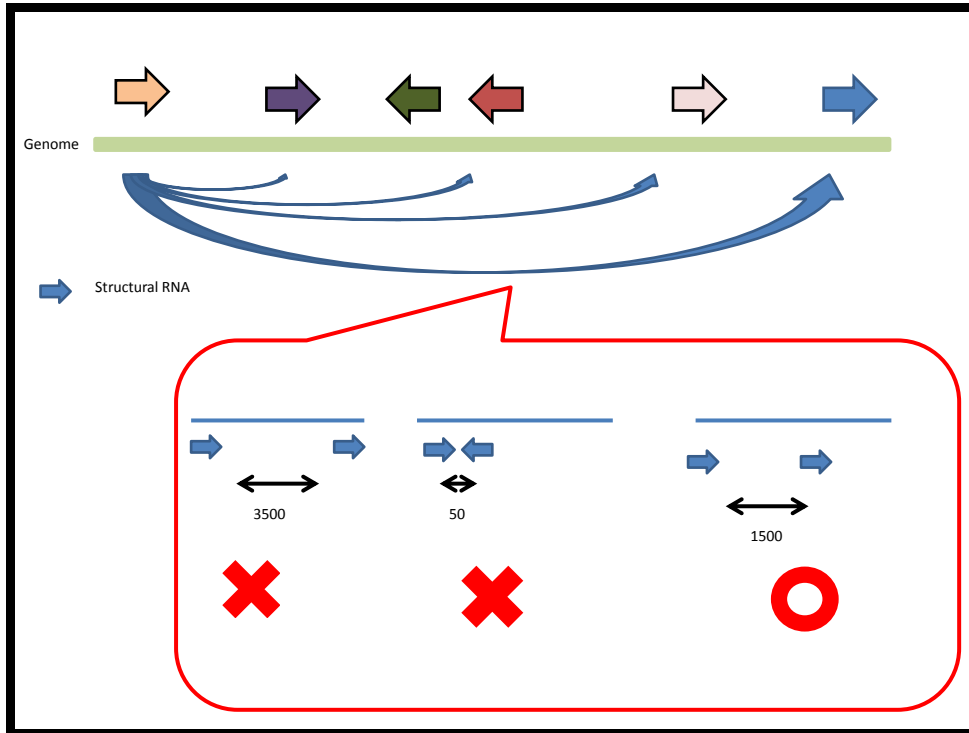


圖 4 篩選 structural RNA pair

聚合酶反應的產物通常都在 100bp 到 3000bp 之間，過長容易不穩定，過短則不易觀察，因此在這步驟裡會逐步篩選 structural RNA pair，若在範圍內的才會進入下一步篩選。

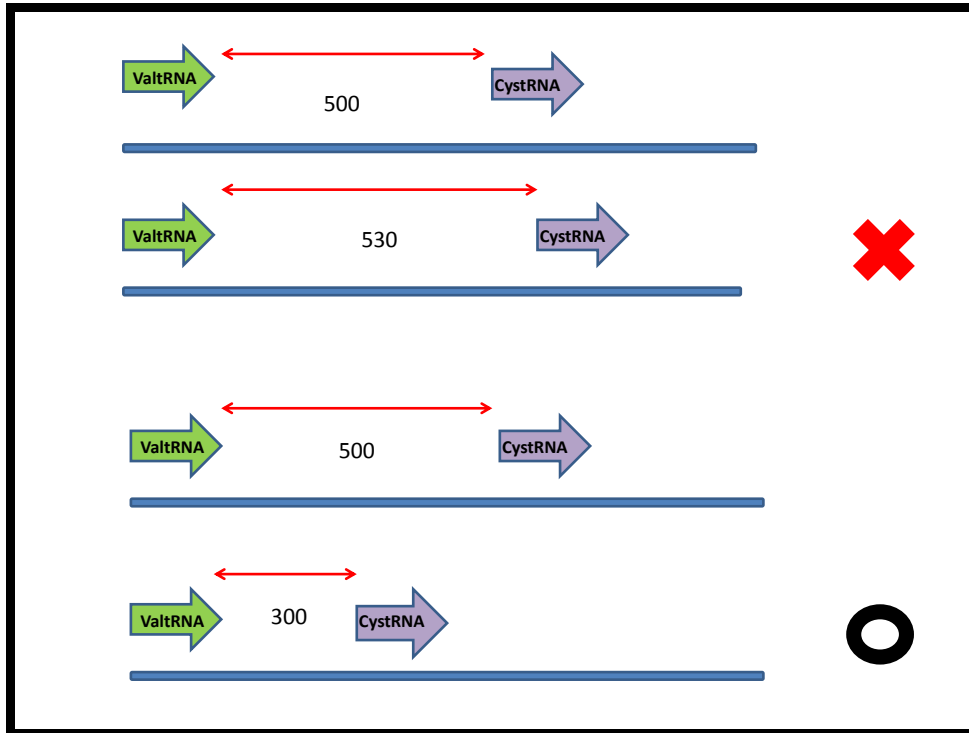


圖 5 挑選具有鑑別力的 structural RNA pair

我們設定兩個 structural RNA pair 間的距離至少需要超過 50bp 才有具有鑑別力。

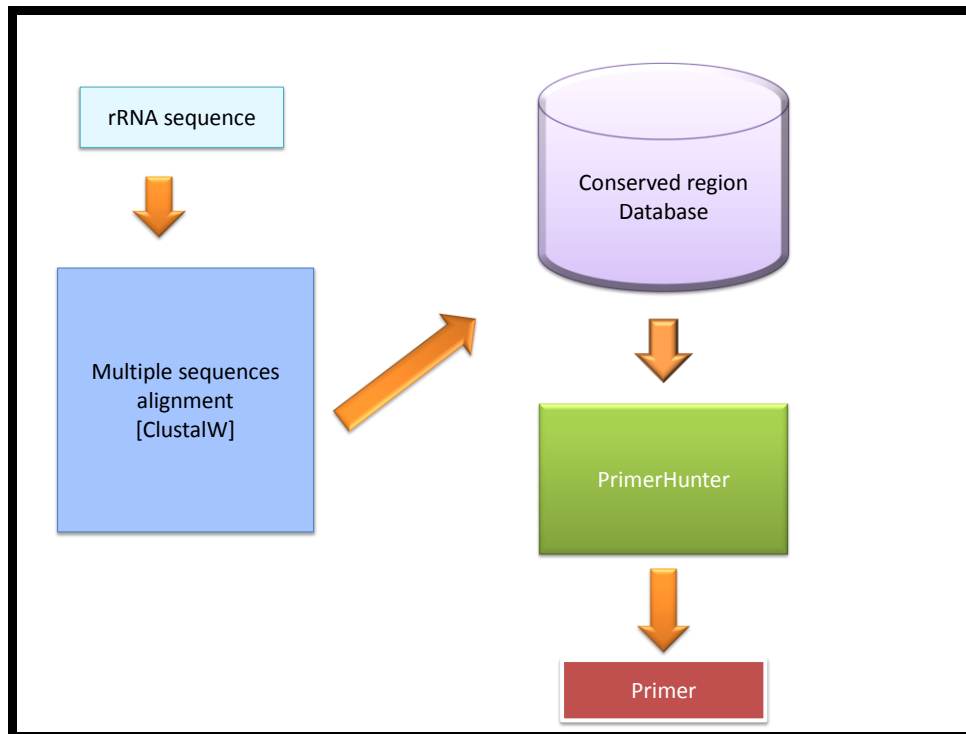


圖 6 引子設計流程(1)

當我們得到 structural RNA 的註解時，我們會將得到這些 rRNA sequence 去執行 ClustalW，找到一或多組的 conserved region，而這些 conserved region 會先存入資料庫，讓 PrimerHunter 做為設計引子的候選。

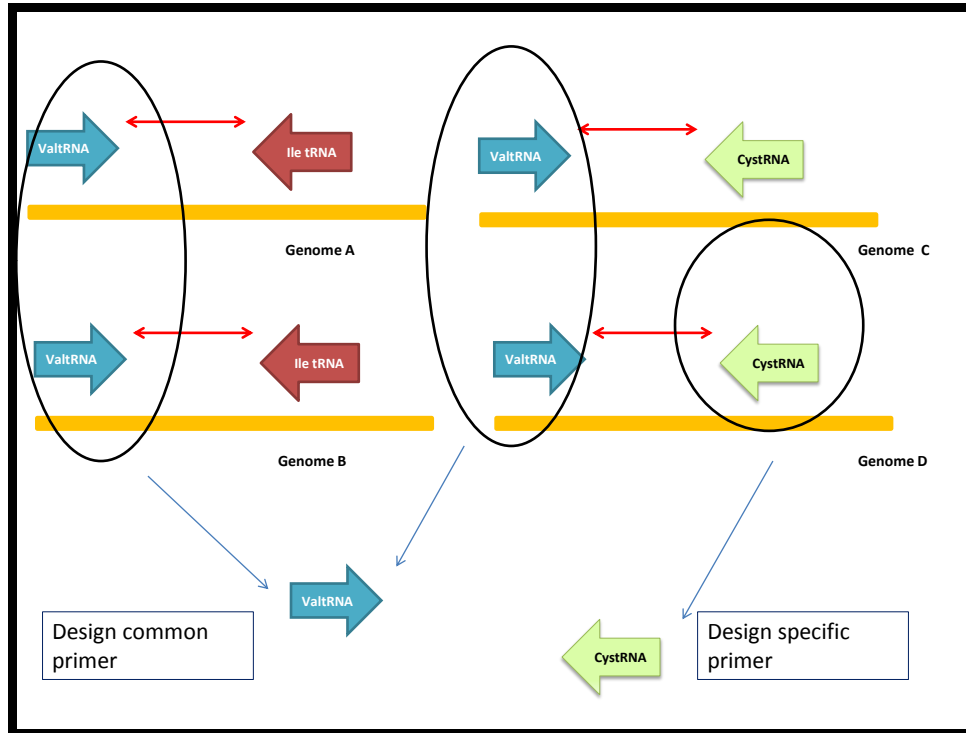


圖 7 引子設計流程(2)

若在不同的基因體上，某組 structural RNA 具 highly conserved 現象，在大多數的基因體上都有重複，我們會將它設計為 common primer，若否，則另外設計 specific primer，如此可以避免設計重複的引子，節省大量的時間。

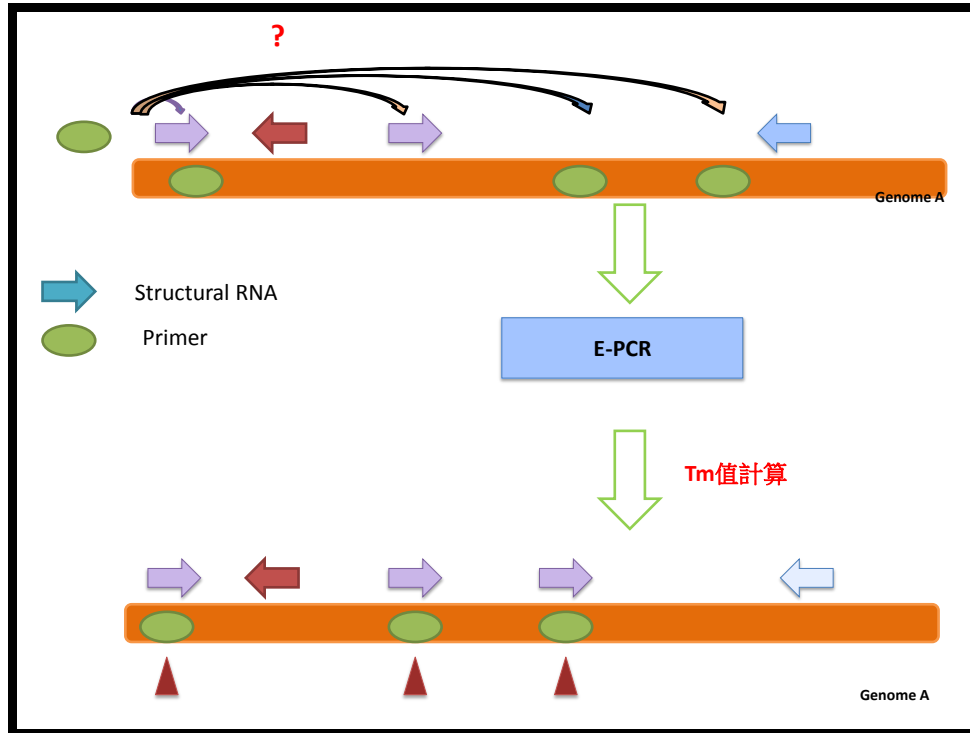


圖 8 判斷是否黏合

得到引子之後，雖然引子都是針對 structural RNA 量身定做的，但由於不確定設計出的來引子會不會黏到基因體上其他位置，所以我們利用 E-PCR 預測出可能接合的位置，再利用 T_m 值計算公式去判斷是否會接合。三角形代表實際可以黏合位置。

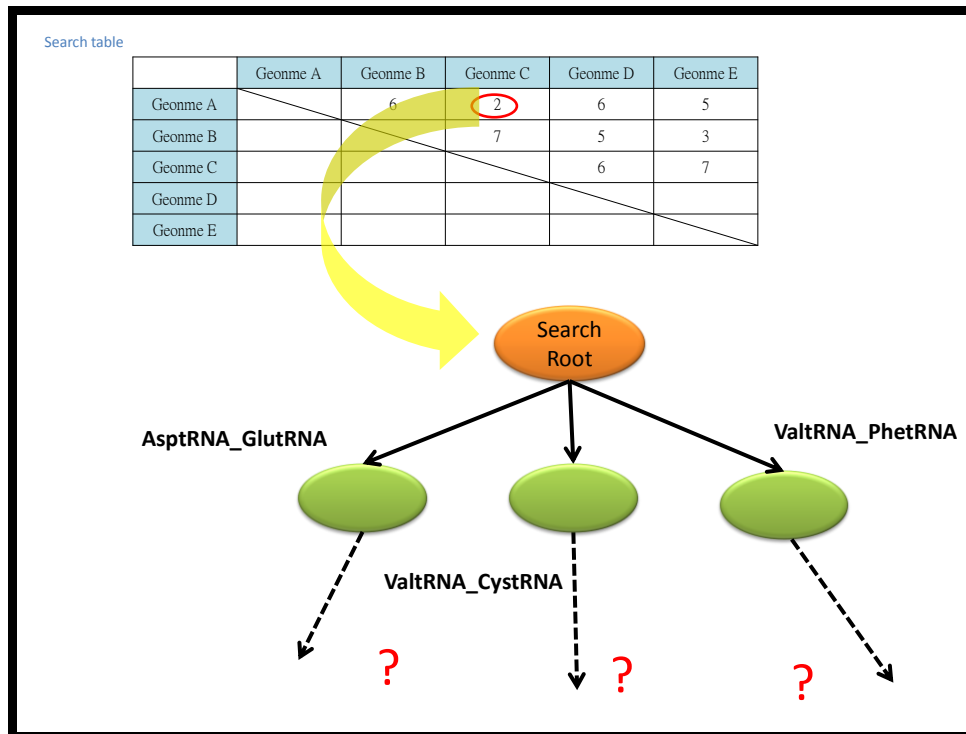


圖 9 搜尋路徑(1)

當建立 Search Table 後，我們可以根據表格上的數據，瞭解到兩兩基因體之間，有幾組引子對可以有效地鑑別。而搜尋起始點會從引子對數最少的開始，因為在最佳解的搜尋路徑規劃中，此組引子對通常是不可或缺的。

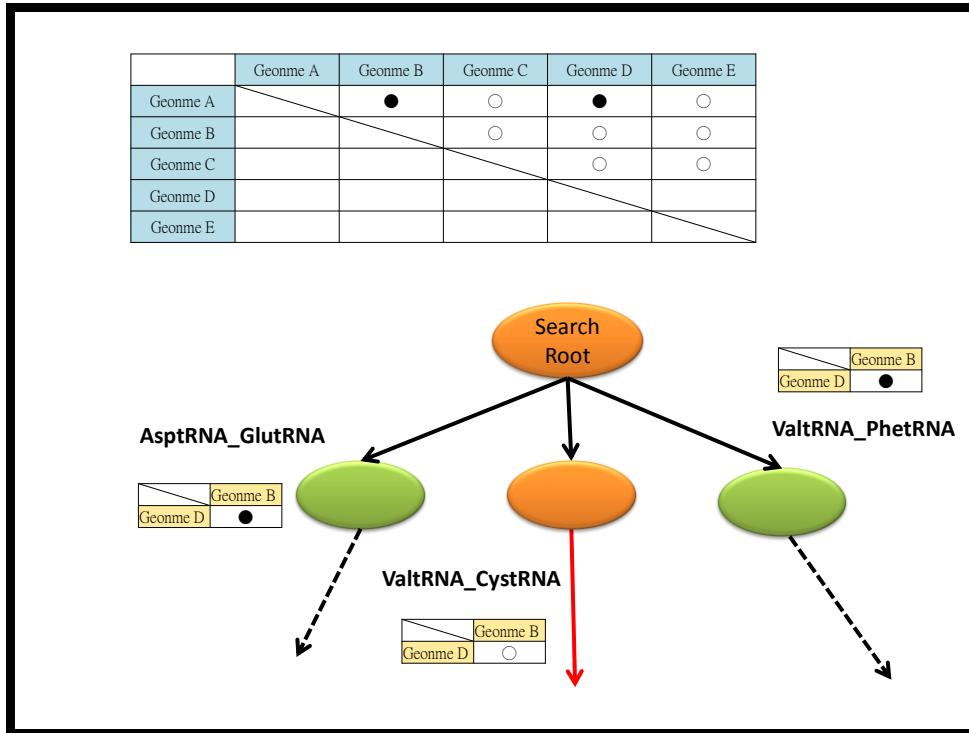


圖 10 搜尋路徑(2)

選定初始引子對後，我們同時也可得到此組引子對還可以區分哪些的基因體，剩下不能區分的基因體，在依照前一步的動作，去搜尋較少引子對可以區分的節點，重複直到完成最佳引子組合為止。

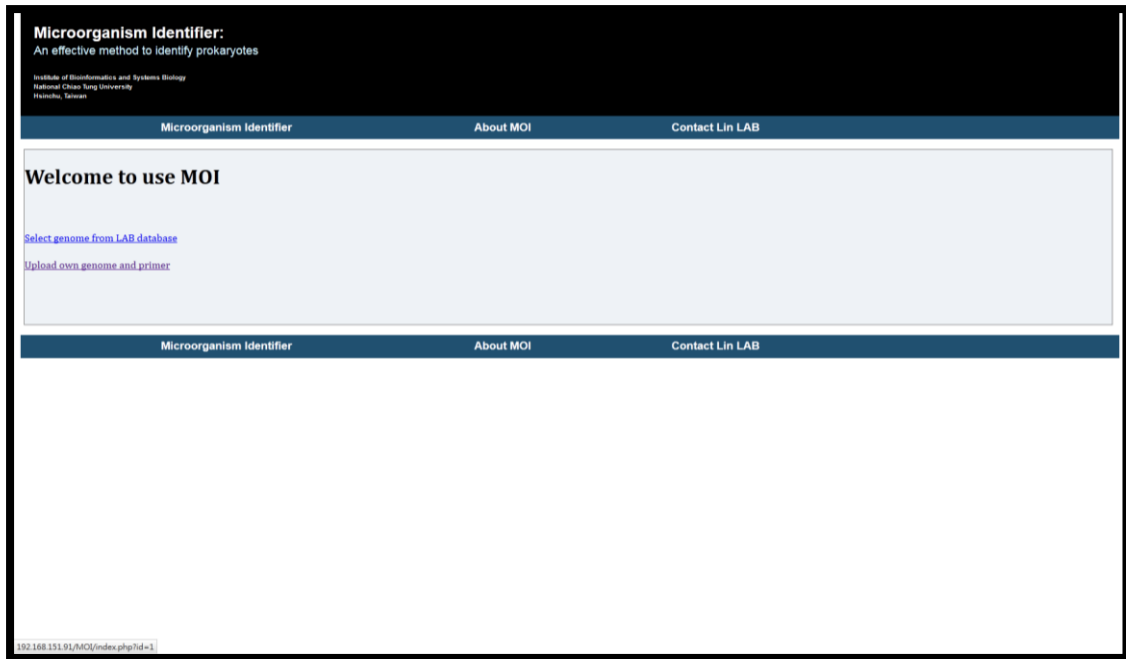


圖 11 系統首頁

在此選擇要使用乳酸菌範例或者自訂上傳序列。



Microorganism Identifier:
An effective method to identify prokaryotes

Institute of Bioinformatics and Systems Biology
National Tsing Hua University
Hsinchu, Taiwan

Microorganism Identifier About MOI Contact Lin LAB

Discriminatory Threshold(TM difference and PCR product length) [Example1](#) [Example2](#)
 * Default setting:TM: 40~ 63 Length:50 nt
 ◉ User-defined 3 ~ TM 50 ~ Length

Selected	Genome seq id	Organism
<input type="checkbox"/>	NC_006814	<i>Lactobacillus acidophilus</i> NCFM
<input type="checkbox"/>	NC_008497	<i>Lactobacillus brevis</i> ATCC 367
<input type="checkbox"/>	NC_008526	<i>Lactobacillus casei</i> ATCC 334
<input type="checkbox"/>	NC_010999	<i>Lactobacillus casei</i> BL23
<input type="checkbox"/>	NC_008054	<i>Lactobacillus delbrueckii</i> subsp. <i>bulgaricus</i> ATCC 11842
<input type="checkbox"/>	NC_008529	<i>Lactobacillus delbrueckii</i> subsp. <i>bulgaricus</i> ATCC BAA-365
<input type="checkbox"/>	NC_010610	<i>Lactobacillus fermentum</i> IFO 3956
<input type="checkbox"/>	NC_008530	<i>Lactobacillus gasseri</i> ATCC 33323
<input type="checkbox"/>	NC_010080	<i>Lactobacillus helveticus</i> DPC 4571
<input type="checkbox"/>	NC_005362	<i>Lactobacillus johnsonii</i> NCC 533
<input type="checkbox"/>	NC_012984	<i>Lactobacillus plantarum</i> JDM1
<input type="checkbox"/>	NC_004567	<i>Lactobacillus plantarum</i> WCFS1
<input type="checkbox"/>	NC_010609	<i>Lactobacillus reuteri</i> JCM 1112
<input type="checkbox"/>	NC_013198	<i>Lactobacillus rhamnosus</i> GG
<input type="checkbox"/>	NC_013199	<i>Lactobacillus rhamnosus</i> Le 705
<input type="checkbox"/>	NC_007576	<i>Lactobacillus sakei</i> subsp. <i>sakei</i> 23K
<input type="checkbox"/>	NC_007929	<i>Lactobacillus salivarius</i> UCC118

Microorganism Identifier About MOI Contact Lin LAB

圖 12 乳酸菌範例

選擇乳酸菌菌株與設定參數，可設定溫度差與產物長度差異的門檻限制。



Microorganism Identifier:
An effective method to identify prokaryotes

Institute of Bioinformatics and Systems Biology
National Chiao Tung University
Hsinchu, Taiwan

Microorganism Identifier
About MOI
Contact Lin LAB

File upload, genome fasta...
Please select genome files
 Caution: "the file must be in fasta format, The genome must be in a single FASTA file"
 For instance:

```
>gi.1183545685
  ATACTAGAGAGA...
```

SELECT FILES

Upload Files

Parameter setting
 Maximum Tm difference between primer pairs(TM: 50° - 73°):
 Maximum PCR product distance between primer pairs:
 Your email:
 Your Job Title:

Microorganism Identifier
About MOI
Contact Lin LAB

圖 13 使用者上傳頁面

使用者可自行上傳欲鑑定的基因體序列，可設定溫度差與產物長度差異的門檻限制，並需要填入信箱與專案名稱，當完成分析時會傳送信件通知使用者。



Primer pair combination 26		
Total primer number:6		
PCR reaction	Forward primer	Reverse primer
PCR reaction 1	Primer1	Primer2
PCR reaction 2	Primer3	Primer4
PCR reaction 3	Primer5	Primer6
Primer1(5'-3'):GGGAACTGGGAGACGC Primer2(5'-3'):CCAAGCTGAGCTACTCCCG Primer3(5'-3'):GGAGGATTAGCTCAGCTGGGAGAG Primer4(5'-3'):CGCTCGACTTGATGTATTAGGCA Primer5(5'-3'):GGCTCAGAGCGGCA Primer6(5'-3'):GGTTCAAATCCCGCAGC		
Primer pair combination 27		
Total primer number:6		
PCR reaction	Forward primer	Reverse primer
PCR reaction 1	Primer1	Primer2
PCR reaction 2	Primer3	Primer4
PCR reaction 3	Primer5	Primer6
Primer1(5'-3'):GGGAACTGGGAGACGC Primer2(5'-3'):CCAAGCTGAGCTACTCCCG Primer3(5'-3'):GGAGGATTAGCTCAGCTGGGAGAG Primer4(5'-3'):CGCTCGACTTGATGTATTAGGCA Primer5(5'-3'):GTGACAGCGAGGATTCTAACCA Primer6(5'-3'):GTGACTACCGTCCGGTTC		
Primer pair combination 28		
Total primer number:6		
PCR reaction	Forward primer	Reverse primer
PCR reaction 1	Primer1	Primer2
PCR reaction 2	Primer3	Primer4
PCR reaction 3	Primer5	Primer6
Primer1(5'-3'):GGGAACTGGGAGACGC Primer2(5'-3'):CCAAGCTGAGCTACTCCCG Primer3(5'-3'):GGAGGATTAGCTCAGCTGGGAGAG Primer4(5'-3'):CGCTCGACTTGATGTATTAGGCA Primer5(5'-3'):CTGTACACACCGCCGTC Primer6(5'-3'):CGACCTCAGCTTATCAGGC		

圖 14 結果圖

Primer pair combination 28 代表第 28 組引子對解，此組結果用 6 個引子即可鑑定。

PCR reaction 為網頁超連結，點擊觀看引子對的詳細資料。

Organism	Su	1	2	3	4	5	6	7	8	9
NC_000964	1	*	x	x	x	x	x	x	x	x
NC_003997	2	*	*	V	V	V	x	V	x	x
NC_006582	3	*	*	*	V	V	x	V	x	x
NC_010609	4	*	*	*	*	V	x	V	x	x
NC_010610	5	*	*	*	*	*	x	V	x	x
NC_014479	6	*	*	*	*	*	*	x	x	x
NC_014639	7	*	*	*	*	*	*	*	x	x
NC_014829	8	*	*	*	*	*	*	*	*	x
NC_016047	9	*	*	*	*	*	*	*	*	*

NC_000964	Primer sequence (5'3')	Tm	Genome position	Product length
NC_000964	Forward primer	42o	3155773	2874
	Reverse primer	44o	3158647	
NC_003997	Forward primer	56o	746845	167
	Reverse primer	42o	747012	
	Forward primer	50o	4651062	352
	Reverse primer	42o	4651414	
NC_006582	Forward primer	50o	3450495	1310
	Reverse primer	44o	3451805	
NC_010609	Forward primer	71o	1045882	137
	Reverse primer	54o	1046019	
	Forward primer	54o	637970	164
	Reverse primer	42o	638134	
	Forward primer	48o	810196	294
	Reverse primer	42o	810490	

圖 15 引子對解(1)

V代表可以鑑別，x則代表否。

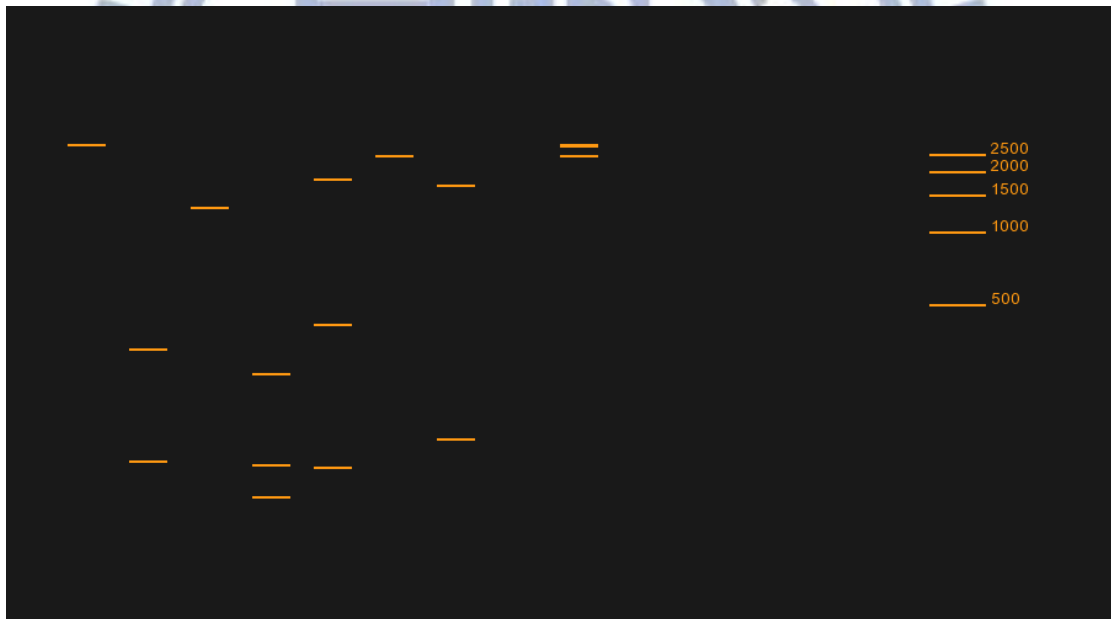


圖 16 預測電泳圖(1)

圖中黃線代表電泳的結果，右側則為 marker。

Organism	Su	1	2	3	4	5	6	7	8	9
NC_000964	1	*	V	V	x	x	V	x	V	x
NC_003997	2	*	*	V	x	x	V	x	V	V
NC_006582	3	*	*	*	x	x	V	x	V	V
NC_010609	4	*	*	*	*	x	x	x	x	x
NC_010610	5	*	*	*	*	*	x	x	x	x
NC_014479	6	*	*	*	*	*	*	x	V	V
NC_014639	7	*	*	*	*	*	*	*	x	x
NC_014829	8	*	*	*	*	*	*	*	*	V
NC_016047	9	*	*	*	*	*	*	*	*	*

NC_000964	Primer sequence (5'-3')	Tm	Genome position	Product length
Forward primer	GGAGGATTAGCTCAGCTGGGAGAG	75.0	166253	292
Reverse primer	CGCTCGACTTGCATGTATTAGGCA	54.0	166545	
Forward primer	GGAGGATTAGCTCAGCTGGGAGAG	77.0	95375	1062
Reverse primer	CGCTCGACTTGCATGTATTAGGCA	42.0	96437	
NC_003997	Primer sequence (5'-3')	Tm	Genome position	Product length
Forward primer	GGAGGATTAGCTCAGCTGGGAGAG	77.0	245265	987
Reverse primer	CGCTCGACTTGCATGTATTAGGCA	42.0	246252	
NC_006582	Primer sequence (5'-3')	Tm	Genome position	Product length
Forward primer	GGAGGATTAGCTCAGCTGGGAGAG	77.0	141833	1150
Reverse primer	CGCTCGACTTGCATGTATTAGGCA	54.0	142983	
NC_014479	Primer sequence (5'-3')	Tm	Genome position	Product length
Forward primer	GGAGGATTAGCTCAGCTGGGAGAG	75.0	160606	291
Reverse primer	CGCTCGACTTGCATGTATTAGGCA	42.0	160897	
NC_014829	Primer sequence (5'-3')	Tm	Genome position	Product length
Forward primer	GGAGGATTAGCTCAGCTGGGAGAG	77.0	106626	1087
Reverse primer	CGCTCGACTTGCATGTATTAGGCA	54.0	107713	
Forward primer	GGAGGATTAGCTCAGCTGGGAGAG	77.0	373702	1312

圖 17 引子對解(2)

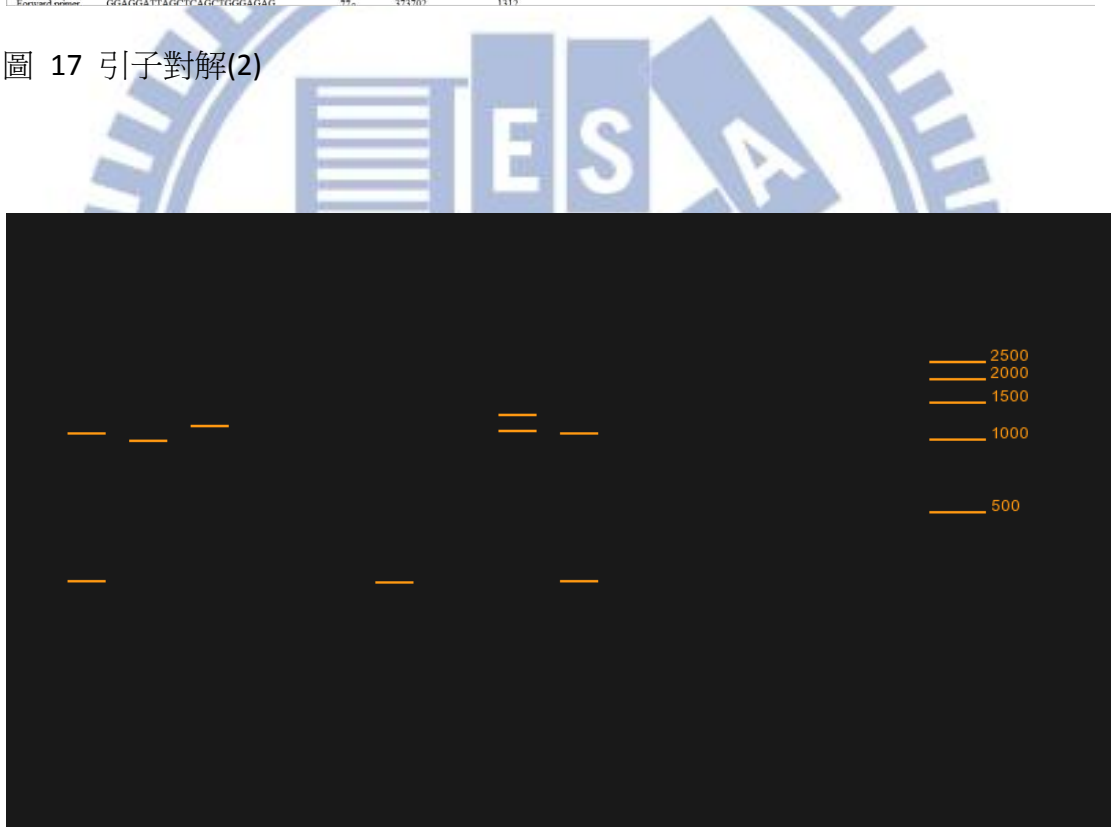


圖 18 預測電泳圖(2)

Organism	Su	1	2	3	4	5	6	7	8	9
NC_000964	1	*	x	V	x	x	x	x	V	V
NC_003997	2	*	*	V	x	x	x	x	x	V
NC_006582	3	*	*	*	V	V	V	V	V	V
NC_010609	4	*	*	*	*	x	x	x	V	V
NC_010610	5	*	*	*	*	*	x	x	V	V
NC_014479	6	*	*	*	*	*	*	x	V	V
NC_014639	7	*	*	*	*	*	*	*	V	V
NC_014829	8	*	*	*	*	*	*	*	*	V
NC_016047	9	*	*	*	*	*	*	*	*	*

Organism	Primer sequence (5'-3')	Tm	Genome position	Product length
NC_000964	CTTGTACACACCGCCGTC	62o	11207	287
Forward primer	CGACCTCAGCTTATCAGGC	42o	11494	
Reverse primer	CTTGTACACACCGCCGTC	62o	31675	287
Forward primer	CGACCTCAGCTTATCAGGC	42o	31962	
Reverse primer				
NC_003997	CTTGTACACACCGCCGTC	62o	10705	317
Forward primer	CGACCTCAGCTTATCAGGC	44o	11022	
Reverse primer	CTTGTACACACCGCCGTC	62o	30499	317
Forward primer	CGACCTCAGCTTATCAGGC	42o	30816	
Reverse primer				
NC_006582	CTTGTACACACCGCCGTC	62o	935033	501
Forward primer	CGACCTCAGCTTATCAGGC	42o	935534	
Reverse primer				
NC_010609	CTTGTACACACCGCCGTC	62o	178761	279
Forward primer	CGACCTCAGCTTATCAGGC	54o	179040	
Reverse primer				
NC_010610				

圖 19 引子對解(3)

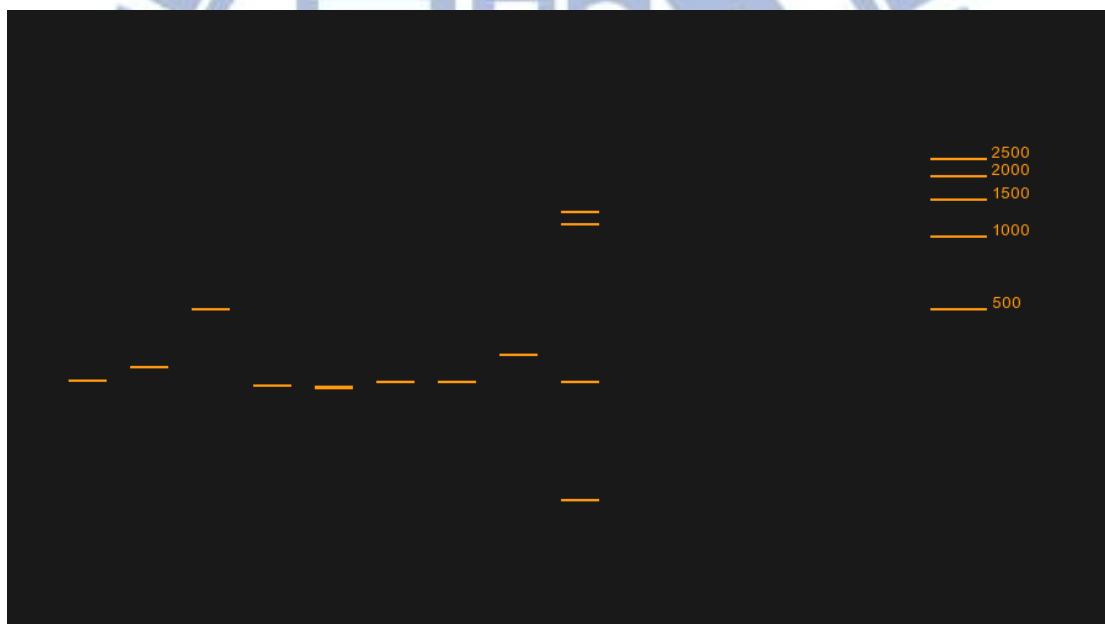


圖 20 預測電泳圖(3)

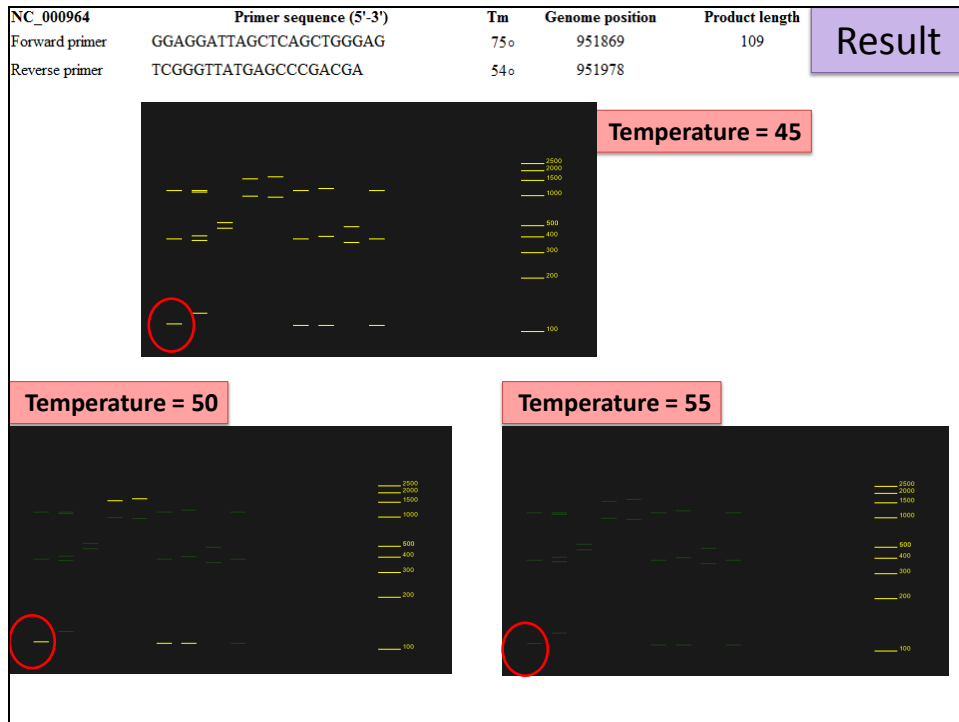


圖 21 預測電泳圖(4)

黃色的線條代表預測會出現的產物，綠色則代表不是預期在這組 T_m 會出現的產物。聚合酶連鎖反應的產物會隨著 T_m 的上升而有所改變。

參考文獻

1. Ege MJ, et al., *Exposure to environmental microorganisms and childhood asthma. N Engl J Med*, 2011. **364**(8): p. 701-709.
2. Dobrindt, U., et al., *Genomic islands in pathogenic and environmental microorganisms. Nature Reviews Microbiology*, 2004. **2**: p. 414-424.
3. PJ, v.d.B., *Antimicrobial drugs, microorganisms, and phagocytes. Rev Infect Dis*, 1989. **11**(2): p. 213-245.
4. Azcarate-Peril MA, Tallon R, and K. TR., *Temporal gene expression and probiotic attributes of Lactobacillus acidophilus during growth in milk. J Dairy Sci*, 2009. **92**(3): p. 870-876.
5. Rodríguez-Lerma GK, et al., *Microbial ecology studies of spontaneous fermentation: starter culture selection for prickly pear wine production. J Food Sci*, 2011. **76**(6): p. M346-352.
6. Mohania D, et al., *Molecular approaches for identification and characterization of lactic acid bacteria. J Dig Dis*, 2008. **9**(4): p. 190-198.
7. Kesmen Z, et al., *Combination of culture-dependent and culture-independent molecular methods for the determination of lactic microbiota in sucuk. Int J Food Microbiol*, 2012. **153**(3): p. 428-435.
8. Vincent Daubin , et al., *logenetics and the Cohesion of Bacterial Genomes. Science*, 2003. **301**(5634): p. 829-832.
9. Guohuq jin , et al., *Inferring Phylogenetic Networks by the Maximum Parsimony Criterion: A Case Study. Molecular Biology and Evolution*, 2006. **24**(1): p. 324-337.
10. Pillidge, C., et al., *Intragenomic 16S rRNA gene heterogeneity in Lactococcus lactis subsp. cremoris International Dairy Journal* 2009. **19**(4): p. 222-227.
11. Cilia V, Lafay B, and C. R, *Sequence heterogeneities among 16S ribosomal RNA sequences, and their effect on phylogenetic analyses at the species level. Mol Biol Evol*, 1996. **13**(3): p. 451-461.
12. Smits SA and O. CC, *Phylometrics: a pipeline for inferring phylogenetic trees from a sequence relationship network perspective. BMC Bioinformatics*, 2010. **11 suppl 6**: p. S18.
13. Qasem JA, Al-Zenki S, and A.-M. A., *Identification and characterization of Streptococcus agalactiae isolates using 16S rRNA sequencing and cellular fatty acid composition analysis. Pak J Biol Sci*, 2010. **13**(1): p. 9-15.
14. BG, S., *Multilocus sequence typing: molecular typing of bacterial pathogens in*

- an era of rapid DNA sequencing and the internet. *Curr Opin Microbiol*, 1999. **2**(3): p. 312-316.
15. Maiden MC, et al., Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci U S A*, 1998. **95**(6): p. 3140-3145.
 16. KA, J., Multi-locus sequence typing. *Methods Mol Med*, 2001. **67**: p. 173-186.
 17. Mark C. Enright, et al., Multilocus Sequence Typing for Characterization of Methicillin-Resistant and Methicillin-Susceptible Clones of *Staphylococcus aureus*. *J Clin Microbiol*, 2000. **38**(3): p. 1008-1015.
 18. WEN-TSO LIU, et al., Characterization of Microbial Diversity by Determining Terminal Restriction Fragment Length Polymorphisms of Genes Encoding 16S rRNA. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*, 1997: p. p.4516-4522.
 19. A.Mark Osborn, E. .B.Moore, and k. n.Tlmm, An evaluation of terminal-restriction fragment length polymorphism(T_RFLP) analysis for the study of microbial community structure and dynamics. *Environmental*, 2000. **2**(1): p. 39-50.
 20. N.J. Gawel , R.L. Jarret , and A.P.W. z, Restriction fragment length polymorphism (RFLP)-based phylogenetic analysis of *Musa*. *Theor*, 1992. **84**: p. 286-290.
 21. P. J. Maughan, et al., Amplified fragment length polymorphism (AFLP) in soybean: species diversity, inheritance, and near-isogenic line analysis *Biomedical and Life Sciences*, 1996. **93**: p. 392-401.
 22. X. Vekemans, et al., Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. *Molecular Ecology*, 2002(11): p. 139-151.
 23. G.M.MUKUVI, et al., Amplified fragment length polymorphism(AFLP) analysis of genetic variation in *Moringa oleifera* LAM. *Molecular Ecology*, 1999. **1999**(8): p. 463-470.
 24. Suzuki K, Ozaki K, and Y. H, Genetic marker for differentiating beer-spoilage ability of *Lactobacillus paracollinoides* strains. *J Appl Microbiol*, 2004. **97**(4): p. 712-718.
 25. MICHAEL WAGNER, et al., Probing Activated Sludge with Oligonucleotides Specific for *Proteobacteria*: Inadequacy of Culture-Dependent Methods for Describing Microbial Community Structure. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*, 1993. **59**(5): p. 1520-1525.

26. Cynthia L. Meaysa, et al., *Source tracking fecal bacteria in water: a critical review of current methods*. ELSEVIER, 2004. **73(1)**: p. 71-79.
27. Hans-Jürgen Busse, Ewald B.M. Denner, and W. Lubitz, *Classification and identification of bacteria: current approaches to an old problem. Overview of methods used in bacterial systematics*. ELSEVIER, 1996. **47(1)**: p. 3-38.
28. Chang, C.-C., *MicroOrganism Identifier: An effective method to identify lactic acid bacteria*. 2011: p. 41.
29. GD, S., *uence mapping by electronic PCR*. *Genome Res*, 1997. **7(5)**: p. 541-550.
30. Rotmistrovsky K, Jang W, and S. GD, *A web server for performing electronic PCR*. *Nucleic Acids Res*, 2004. **32(Web Server issue)**: p. W108-112.
31. Lowe TM and E. SR., *tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence*. *Nucleic Acids Res*, 1997. **25(5)**: p. 955-964.
32. Schattner P, Brooks AN, and L. TM., *The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs*. *Nucleic Acids Res*, 2005. **33(Web Server issue)**: p. W686-689.
33. Thompson JD, Higgins DG, and G. TJ., *CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice*. *Nucleic Acids Res*, 1994. **22(22)**: p. 4673-4680.
34. Duitama J, et al., *PrimerHunter: a primer design tool for PCR-based virus subtype identification*. *Nucleic Acids Res*, 2009. **37(8)**: p. 2483-2492.
35. W Rychlik, W J Spencer, and R.E. Rhoads, *Optimization of the annealing temperature for DNA amplification in vitro*. *Nucleic Acids Res*, 1990. **18(21)**: p. 6409-6412.
36. Barbe V, et al., *From a consortium sequence to a unified sequence: the Bacillus subtilis 168 reference genome a decade later*. *Microbiology*, 2009. **155(Pt6)**: p. 1758-1775.
37. Earl AM, et al., *Whole-genome sequences of Bacillus subtilis and close relatives*. *J Bacteriol*, 2012. **194(9)**: p. 2378-2379.
38. DR, Z., *The genome sequence of Bacillus subtilis subsp. spizizenii W23: insights into speciation within the B. subtilis complex and into the history of B. subtilis genetics*. *Microbiology*, 2011. **157(Pt7)**: p. 2033-20471.
39. Rasko DA, et al., *Genomics of the Bacillus cereus group of organisms*. *FEMS Microbiol Rev*, 2005. **29(2)**: p. 303-329.
40. Kageyama Y, et al., *Intragenomic diversity of the V1 regions of 16S rRNA genes in high-alkaline protease-producing Bacillus clausii spp*. *Extremophiles*,

2007. **11**(4): p. 597-603.

41. *ese SA, et al., The evolution of host specialization in the vertebrate gut symbiont Lactobacillus reuteri. PLoS Genet, 2011. 7*(2): p. e1001314.
42. *Morita H, et al., Comparative genome analysis of Lactobacillus reuteri and Lactobacillus fermentum reveal a genomic island for reuterin and cobalamin production. DNA Res, 2008. 15*(3): p. 151-161.

