

行政院國家科學委員會補助專題研究計畫 期末報告

利用同源基因資訊開發新演算法提供有效鑑種 方法及基因水平轉移研究

計畫類別： 個別型計畫 整合型計畫

計畫編號：NSC - 98-2621-B -009 - 001

執行期間：98年08月01日至101年07月31日

執行機構及系所：國立交通大學 生物科技學系(所)

計畫主持人：林勇欣博士

共同主持人：

計畫參與人員：王俊霖、方羿喬、殷潔、瓦蘭蒂、潘承宗、
李應達、張智欽、何宜佩、陳盈潔、蔡宗
翰、陳之杭、王煒驊、張瑋婷

本計畫除繳交成果報告外，另須繳交以下出國報告：

出席國際學術會議心得報告及發表之論文

處理方式：除列管計畫及下列情形者外，得立即公開查詢

涉及專利或其他智慧財產權， 一年 二年後可公開查詢

中 華 民 國 101 年 10 月 日

目錄

中英文摘要及關鍵詞.....	3
中文摘要.....	3
英文摘要.....	3
關鍵詞.....	4
報告內容.....	5
前言.....	5
研究目的.....	5
文獻探討.....	6
研究方法.....	7
Structural RNA 註解	7
挑選具鑑別力的 conserved region	8
引子設計.....	8
非專一性 binding site 檢測.....	8
組合分析- 挑選引子片段之近似最佳解演算法	8
<i>Stigmatella aurantiaca</i> 的水平基因轉移事件.....	8
結果與討論.....	9
參考文獻.....	11
附錄.....	13
附件四 國科會補助專題研究計畫項下出席國際學術會議心得報告.....	13
附件四 國科會補助專題研究計畫項下出席國際學術會議心得報告.....	15

中英文摘要及關鍵詞

中文摘要

第一部分

微生物對人類的日常生活甚為重要，因此如何能快速有效的鑑別微生物樣本的種類一直都是生物科學上的重要課題。傳統的方式以物種型態表現與生化血清型鑑定為主。在後基因體時代，由於許多微生物的基因體序列已被完全解碼且公開，使得科學家得以從基因體序列資訊中找尋新的遺傳特徵以作為物種快速鑑別使用。在本研究中，基於觀察到 Structural RNA 在原核生物基因體中的廣泛分佈，我們提出一個新的遺傳特徵- Structural RNA intergenic distance。並且在本研究中發展了有效率的演算法，一般使用者得以利用分子生物學實驗室中 PCR 技術，在相近微生物物種的 reference genome 可得的情況下，設計引子且利用少數的幾個 PCR 反應便能有效率的區別微生物樣本的物種。為了驗證了此新遺傳特徵的有效性，我們也以乳酸桿菌屬的基因體序列為例進行分析。本計畫最後也開發了網頁服務以供使用者處理其有興趣的微生物樣本，使本計畫成果能有更廣泛的應用。

第二部分

水平基因轉移 (Horizontal Gene Transfer, HGT)，又稱 (Lateral Gene Transfer, LGT)，是遺傳物質在物種間水平轉移的特定現象，在原核生物間較為常見。此轉移現象可讓接受基因轉移的物種得到新的特徵，增加其環境適應性 (比如對於抗生素的抗性)。本研究開發了一套可廣泛地檢測水平基因轉移事件的演算法，比較目標物種以及其他參考物種的基因體序列，挑選出在目標物種上可能的水平基因轉移候選基因。本研究以 *Stigmatella aurantiaca* 為例作為分析的目標物種。*S. aurantiaca* 是一種黏球菌 (兼性補食菌)，具有非常龐大的基因體。如此特殊的基因體大小可能起因於其補食其它細菌的特性，讓 *S. aurantiaca* 有更多機會接觸外界的遺傳物質而提高其水平基因轉移的機會。最後針對這些可能的候選基因再檢視其演化樹的型態來確認是否曾發生過水平基因轉移現象

英文摘要

Part I

Microorganisms are important for human daily life. The accurate identification of microorganisms is an important issue for biological science. Most traditional identification methods rely on the distinction of phenotypes and serotypes. In the post-genome era, with many reference prokaryotic genomic sequences available, we could use these data to develop a novel genetic fingerprint. In this study, we used intergenic distances between structural RNAs to distinguish microorganisms. A publicly-available web service was developed so that users can use the pipeline to design specific PCR primers that can distinguish microorganisms by PCR amplified fragment lengths.

Part II

Horizontal gene transfer (HGT), also named lateral gene transfer (LGT) event is usually discovered between prokaryotic species. This gene transfer event could

provide novel characters for the receiver species and thus increase its fitness. In this study, we developed a new algorithm that can extensively identify HGT events in a target species by comparing other reference genomic sequences. We used *Stigmatella aurantiaca* as an example. In the end, the identified HGT candidate genes would be verified by comparing their phylogenetic trees.

關鍵詞

菌種鑑定、16S rRNA、tRNA、基因體、PCR、水平基因轉移
identification of microorganism; 16S rRNA; tRNA; genome; PCR; Horizontal Gene Transfer; Lateral Gene Transfe

報告內容

前言

傳統微生物鑑定工作通常需要將微生物樣本在微生物實驗室內先加以培養後，藉由觀察微生物的表現型與血清型特徵來鑑定其種類或品系。這樣的流程通常需要花費幾天到幾周的時間，而且有許多微生物目前在實驗室內仍然是無法培養的(Sedlacek et al. 2010)。分子生物學 DNA 定序技術發展後，除了表現型與血清型外，使用 DNA 遺傳特徵也成為一種新的微生物鑑定技術發展的方向。比起傳統方法，DNA 遺傳特徵的方法較為快速且節省成本，較適合日常快速大量鑑定使用(Keswani and Whitman 2001; Rossello-Mora and Amann 2001)。

另外，水平基因轉移(Horizontal Gene Transfer, HGT)在原核生物中經常發生，偶爾在真核生物中亦可發現。此基因轉移現象能幫助生物快速得到新的特徵以適應環境，比如抗生素的抵抗或環境物質的代謝等。藉由轉形(Transformation)、結合 (Conjugation)、轉導(transduction)等方法，遺傳物質便有機會能跨越物種傳播。事實上，水平基因轉移的速率往往高於我們所能觀察到的。我們能藉由一些特徵來檢視水平基因轉移現象，比如演化分析上分類的 inconsistence—水平基因轉移候選基因在演化樹分類的位置會較接近其來源基因而與一般認知的物種樹不同，或由被懷疑是水平基因轉移基因的 GC 含量 (GC Content)、密碼子偏性 (Codon usage bias)等組成與周圍基因不同的特性來判斷。但現在並沒有一個簡單快速的方法能自動分析一個基因是否有水平基因轉移的傾向，因此我們嘗試利用兩組基因與目標基因間的同源關係，應用比例原則去找出距離特異的點，藉此快速篩選可能為水平基因轉移候選的基因，以進行接下來的演化樹分析。

研究目的

DNA 遺傳特徵的微生物鑑定技術，無論是基於 DNA 序列本身的方法亦或是基於長度片段多樣性的各種方法，普遍而言都有缺點是為特定物種而設計。當實驗室使用者研究工作上需要確認某認兩種微生物時，缺乏廣泛適用於任意微生物間的微生物鑑定方法。基於此動機，在觀察到後基因體時代，許多微生物基因體已經被完全解碼，若使用者能獲得相近物種的微生物基因體序列供參考，則如何能在避免利用局部遺傳特徵或會造成鑑定偏差的情況下，亦能快速幫助使用者鑑定為生物物種，是本研究想突破的課題之一。

為此，本研究發展了一個新的演算法，此演算法利用原核生物 structural RNAs 基因間的距離作為每一種原核生物不同的遺傳指紋，並且將此新穎的遺傳指紋應用在微生物樣本的快速鑑定上。

本研究第二部分則希望能開發一個能廣泛自動檢測水平基因轉移現象的方法，利用 BLAST 之蒐尋結果及 PhP 程式語言為基礎，將欲知物種之全基因組對細菌及真菌類做全面性的搜尋以快速確定任一基因之水平基因轉移關係。

文獻探討

利用分子生物學 DNA 遺傳特徵進行原核生物物種鑑定的方法大致可分為兩大類型。第一大類的方法是基於演化過程中，不同物種同源基因會逐漸演化而略有差異的觀察，進一步使用該同源基因 DNA 序列本身的差異。這類方法中最常見的例子是 16S ribosomal RNA (rRNA) 序列(Stackebrandt and Goebel 1994; Rossello-Mora and Amann 2001; Gevers et al. 2005)，因為這個方法被認為在確認是否為新物種上極為有效。在微生物命名的過程中，必須提供候選新物種的 16S rRNA 序列資料以證明其跟已知物種不同(Coenye et al. 2005; Gevers et al. 2005; Janda and Abbott 2007)。使用單一同源基因的方法，除了 16S rRNA 外，粒線體上的 cytochrome c oxidase subunit 1 (COI)(Hebert et al. 2003; Seifert et al. 2007)、ribosomal gene cluster 上的 internal transcribed spacer (ITS) region(Chase et al. 2005; Meyer et al. 2009)、DNA gyrase (Wang et al. 2009; Wang et al. 2010)等基因都被作為基因特徵的標的。

過去經常利用 16S rRNA 對原核生物進行物種鑑定，因為它們的演化速率比一般的基因慢且不易被水平基因轉移事件影響。Booyesen, Dicks and Meijering 運用 16S rRNA 分析來鑑定 *Leuconostoc* 和 *Weissella* 菌種(Booyesena et al. 2002)。但在分化時間較短的物種之間，16S rRNA 常因序列過於保守而無法提供足夠的鑑別力。Felis (Felis et al. 2001)、Torriani (Torriani et al. 1999; Torriani et al. 2001)和 Bringel (Bringel et al. 2001)以 *recA* 基因的序列分析來取代 16S rRNA 序列分析，也成功的完成了 *Lactobacillus casei* 及 *Lactobacillus plantarum* 菌種的鑑定。

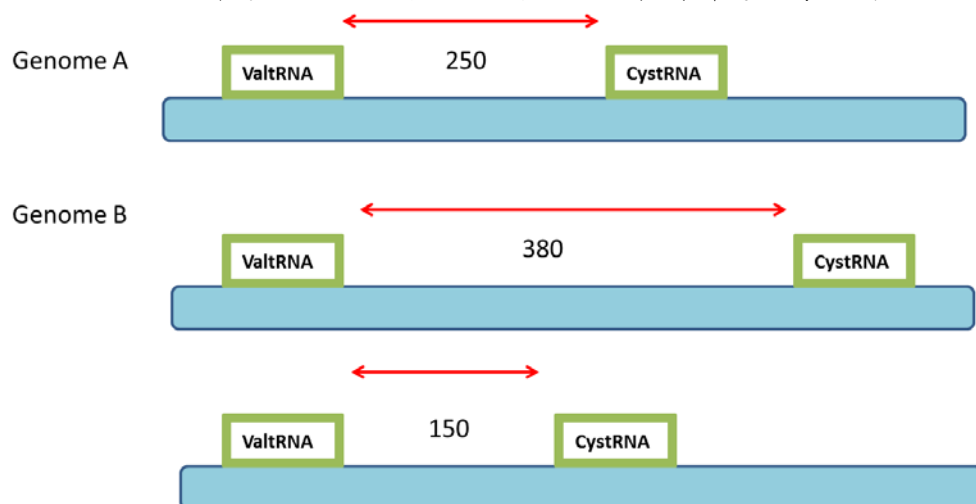
第二大類的方法不直接使用 DNA 序列本身，而是利用 DNA 片段長度的多樣性，例如擴增片段長度多態性(Amplified Fragment Length Polymorphism, AFLP)(Pongam et al. 1998)，restriction fragment length polymorphisms (RFLP)(Zelinski 1991; McNicholas 1994)，ribotyping(Hiramatsu 1992; Bingen et al. 1994; Papasian et al. 1996; Pavlic and Griffiths 2009)等。擴增片段長度多態性、16S rDNA 限制性酶切片段長度多態性分析(Amplified rDNA restriction analysis)(Deak 1999)和即時定量聚合酶連鎖反應(Real-time PCR)等都是架構在聚合酶連鎖反應的應用。

至於傳統研究水平基因轉移現象的方法多半是利用 BLAST 過濾掉相同或相近的同源基因，找出序列相似度高但物種關係很遠的基因來當做水平基因轉移現象的候選者，並進行後續的演化與基因組成分析。最近的研究顯示廣泛發生水平基因轉移現象會使物種間的演化關係變成網路的形式，即 Phylogenetic network，此網路的建立，利用最小 SPR distance 的方法，可加入 inner node 融合不同的演化樹，將水平基因轉移的概念直接表達在演化樹上(Nakhleh 2010)。如果能用數值法自動篩選出最可能有水平基因轉移現象的基因的話，也許我們就能更有效率地建立 Phylogenetic network，可以更廣泛地研究物種早期演化的歷史。

研究方法

針對物種鑑定，本研究中提出一個新的基因型鑑定法，即利用 structural RNAs 基因間的距離作為不同物種的遺傳指紋，與現階段已知被解碼的微生物序列比對，作為鑑定的核心。

方法中，需先收集目標鑑定物種基因體序列，接著去辨識出其 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。



為了驗證及方便實作這新遺傳特徵及演算法，本研究利用此演算法開發了一個網頁服務工具(稱 MOI)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"組合分析-挑選引子片段之近似最佳解演算法"，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。

整個網頁工具 MOI 的流程簡示於下圖：



Structural RNA 註解

目前大多數已定序完成的基因體其基因種類及位址都是經由生物資訊的方法預測註解得到的，不同研究者所提供的註解基準可能並不相同。然而標準一致的 structural RNA 註解對本演算法而言相當重要，若直接將上述不一致的註解結果

引用於本研究上，會有許多誤差產生。因此在 reference genome 上傳後，我們的網頁系統中，會先以 tRNAscan(el-Mabrouk and Lisacek 1996; Lowe and Eddy 1997; Schattner et al. 2005) 及 NCBI 16S microbial Database 中的序列，將 structural RNA 於基因體中的位置正規化。

挑選具鑑別力的 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 在基因體上的間距來做篩選的標準。以 structural RNA pair 的距離差距在 50 base pair 且差距 5% 為具鑑別力的標準。以此標準在所有的目標基因體序列群組中，挑選出有間距鑑別力的 structural RNA pair 的片段，並使用多重序列比對(multiple alignment)來找出所有 structural RNA pair 的 conserved region。

引子設計

在得到 conserved region 後，此研究中選擇利用設計引子的工具為 Primer Hunter(Duitama et al. 2009)。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 (Tm): 60°~63°
Product length: 100 ~ 2000 bp

非專一性 binding site 檢測

在完成引子設計之後，會得到許多針對目標基因體群具有鑑別能力的引子對，但這些引子對是否有 non-specific binding site 需要另外檢測。在研究中使用另一工具 e-PCR(Schuler 1997) 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

組合分析- 挑選引子片段之近似最佳解演算法

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。
- (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)。

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。

***Stigmatella aurantiaca* 的水平基因轉移事件**

本研究第二部分的目的是以 *S. aurantiaca* 為例，找出此目標物種全基因各自與哪些資料庫中物種的基因體特定部分有水平基因轉移現象，因此我們由 NCBI 上抓取 *S. aurantiaca* 之全基因做為目標物種和真細菌、古細菌、真菌之完整基因體共三千餘件做為資料庫。首先利用 tBLASTn，以 *S. aurantiaca* 之基因為

query 對資料庫進行 BLAST。根據 NCBI Handbook 所述，經過標準化後的 bit-score 可以用來比較兩序列間分化的程度，因此對於每筆 *S. aurantiaca* 之基因與其在資料庫內的 hit，我們便以 bit-score 做為判斷該組序列同源性的數值。得到大量數據後我們將資料庫中的所有物種進行 all-against-all 的兩兩比對，並同 *S. aurantiaca* 之基因在此兩物種上有源性之基因的 BLAST hit 做一張以兩物種為 x-y 軸，bit-score 為軸上數值的散布圖，如此一來此圖包含了三條序列的意義，即 *S. aurantiaca* 之基因，x 軸代表之物種與 y 軸之物種，每一個與此兩物種有同源關係的 *S. aurantiaca* 基因在此圖上便會是一個點。在一般的情況下，雖然不同的基因會有不同的演化速率，但是 *S. aurantiaca* 與物種 x 的分化時間與 *S. aurantiaca* 與物種 y 的分化時間之間的比值應該是固定的，因此在 bit-score 的散佈圖上點的分布應該會是呈一線性回歸的模式。但如果某一個 *S. aurantiaca* 基因與兩物種其一或其近親曾發生水平基因轉移現象，另一個物種與 *S. aurantiaca* 的分化時間則在此水平基因轉移現象之前，該基因所對應的點便可能會獨立於線性回歸群集之外形成 outlier。本研究利用 php 程式語言自動將所有物種進行兩兩比較後，使用 R 語言計算每個點各自的 Mahalanobis distance，此 distance 在該比較組中較大的基因點(outlier) 便是可能發生水平基因轉移現象的候選。挑出這些候選基因並排序後，程式會自動抓取該基因與其關聯兩物種上基因體的序列片段，利用 CLUSTAL W 和 MEGA 將演化樹建出，加上該比較組內其他正常基因之演化樹(隨機挑選 10 個非水平基因轉移之基因，同樣取該基因與其關聯兩物種上基因體的序列片段所見之樹)，便能簡單的比較兩演化樹呈現的分類差異以確定是否真是水平基因轉移現象。

結果與討論

本研究發展了一個新的演算法，此演算法利用原核生物 structural RNAs 基因間距離作為每一種原核生物不同的遺傳指紋，並且將此新穎的遺傳指紋應用在微生物樣本快速鑑定上。此外，在本研究中，亦完成了一使用本演算法及遺傳指紋的網頁服務，可供網頁使用者自行在上傳原核生物基因體序列後，提供對應的 PCR 引子，供做快速鑑定參考。

在本研究中，實際使用了乳酸桿菌的基因體序列為資料驗證了本演算法及遺傳指紋的高有效性，挑選乳酸桿菌屬的原因是本屬微生物多屬於益生菌，廣泛研究完成多株全基因體定序，且實際在應用上常有鑑別菌種的需要；從演算法得到的解挑選了二組引子(MGlu_M23s 和 MGly_M23s，如下表)及 7 株 *Lactobacillus* 菌株(*L. brevis* ATCC 367, *L. gasseri* ATCC 33323, *L. casei* BL23, *L. plantarum* WCFS1, *L. rhamnosus* GG, *L. delbrueckii subsp. bulgaricus* ATCC 11842, *L. acidophilus* NCFM)透過聚合酶連鎖反應及凝膠電泳分析來驗證。

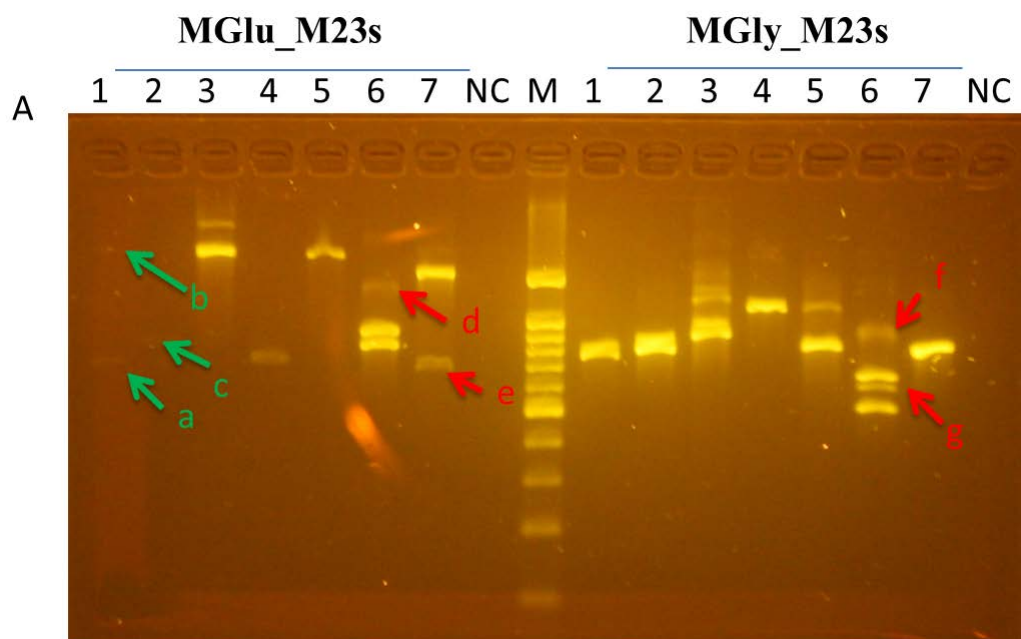
Primer pair	Forward primer(5' -> 3')	Reverse primer(5' -> 3')
MGlu_M23s	CCGTACGGGATTTGA ACCCA	CCTTAGTACGAGAGG ACCGGGA
MGly_M23s	ACCATGGCAAGGTGA TGTTCTAC	CCTTAGTACGAGAGG ACCG

凝膠電泳照片圖，及預測之 PCR 產物長度如下圖。其中得到的產物大部分與預測相符，但圖中綠色箭頭指的 a、b、c 三個點，由於兩端 primer 溫度的差異，

只得到微弱訊號。紅色箭頭指的 d、e、f、g 四個點，推測是非專一性的 binding 造成的不再預期內的產物。

儘管 PCR 驗證實驗，產物與預測並非 100% 相符，但是 PCR 產物長度的 pattern 仍能讓使用者清楚不模糊的分辨所要辨別的微生物樣本。

由演算法得到的結果，僅由 3 個 primer 完成 2 個 PCR 反應即可將 7 個不同乳酸桿菌屬的樣本加以區分。



B

Lane No.	Organism	PCR 1 (primers MGluc and M23S)		PCR 2 (primers MGlyc and M23S)			
1	<i>Lactobacillus brevis</i>	676	1891	738			
2	<i>Lactobacillus gasseri</i>	787	2804	793			
3	<i>Lactobacillus casei</i>	1887	2903	810	915	1218	
4	<i>Lactobacillus plantarum</i>	696	2346	1123			
5	<i>Lactobacillus rhamnosus</i>	1883		806	824	1127	
6	<i>Lactobacillus delbrueckii</i>	780	882	~1400	493	633	~600 ~900
7	<i>Lactobacillus acidophilus</i>	1451	1573	~700	804		
8	no template control						

在水平基因轉移的研究結果方面，掃描完 *Stigmatella aurantiaca* 全基因之後，有六組 outlier 顯示「STAUR_2131」為水平基因轉移候選基因。在詳細比對其演化樹後可以看出，「STAUR_2131」與一群 *Aspergillus* 屬的真菌分在同一個分支，相異於其真正的物種演化關係，確實為水平基因轉移事件的結果。然而此方法仍在開發階段，其分析時間較長，且尚未確認是否真的可以找出全部的候選基因，由於多樣性的影響，同種的物種間仍有少量序列差異，這會造成相當明顯的 outlier 而干擾我們的分析，因此必須要設立一合適的閾值來進行過濾，這部分也需要更進一步的研究。另外，關於 *S. aurantiaca* 基因體大小是否

受到水平基因轉移現象影響的問題，也需要更進一步的分析其他參考物種來進行比較才能確定。無論如何，此研究初步開發出了能自動且有系統地找出水平基因轉移事件的演算法，而且是以全物種基因體分析為基底、輔以每個候選基因之挑選都有一個比較的序列在，能夠避免資料缺失或演化時間差異不足而造成水平基因轉移分析困難或有遺漏的問題。這樣的研究成果如搭配上 phylogenetic network 的建立方法，便能夠有效的建立各物種間一目了然的水平基因轉移關係，協助研究者針對基因功能或來源進行更進一步的研究。

參考文獻

- Bingen EH, Denamur E, Elion J. 1994. Use of ribotyping in epidemiological surveillance of nosocomial outbreaks. *Clin Microbiol Rev* **7**(3): 311-327.
- Booyensa C, Dicks LM, Meijering I, Ackermann A. 2002. Isolation, identification and changes in the composition of lactic acid bacteria during the malting of two different barley cultivars. *Int J Food Microbiol* **76**(1-2): 63-73.
- Bringel F, Quenee P, Tailliez P. 2001. Polyphasic investigation of the diversity within *Lactobacillus plantarum* related strains revealed two *L. plantarum* subgroups. *Syst Appl Microbiol* **24**(4): 561-571.
- Chase MW, Salamin N, Wilkinson M, Dunwell JM, Kesanakurthi RP, Haidar N, Savolainen V. 2005. Land plants and DNA barcodes: short-term and long-term goals. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* **360**(1462): 1889-1895.
- Coenye T, Gevers D, Van de Peer Y, Vandamme P, Swings J. 2005. Towards a prokaryotic genomic taxonomy. *FEMS Microbiology Reviews* **29**(2): 147-167.
- Deak T. 1999. Molecular taxonomy of yeasts. *Acta Microbiol Immunol Hung* **46**(2-3): 181-186.
- Duitama J, Kumar DM, Hemphill E, Khan M, Mandoiu, II, Nelson CE. 2009. PrimerHunter: a primer design tool for PCR-based virus subtype identification. *Nucleic Acids Res* **37**(8): 2483-2492.
- el-Mabrouk N, Lisacek F. 1996. Very fast identification of RNA motifs in genomic DNA. Application to tRNA search in the yeast genome. *J Mol Biol* **264**(1): 46-55.
- Felis GE, Dellaglio F, Mizzi L, Torriani S. 2001. Comparative sequence analysis of a *recA* gene fragment brings new evidence for a change in the taxonomy of the *Lactobacillus casei* group. *Int J Syst Evol Microbiol* **51**(Pt 6): 2113-2117.
- Gevers D, Cohan FM, Lawrence JG, Spratt BG, Coenye T, Feil EJ, Stackebrandt E, Van de Peer Y, Vandamme P, Thompson FL et al. 2005. Re-evaluating prokaryotic species. *Nature Reviews Microbiology* **3**(9): 733-739.
- Hebert PD, Ratnasingham S, deWaard JR. 2003. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings Biological sciences / The Royal Society* **270** Suppl 1: S96-99.
- Hiramatsu K. 1992. [Application of ribotyping to the identification and molecular epidemiology of staphylococci]. *Nihon Rinsho* **50** Suppl: 333-338.
- Huntley S, Hamann N, Wegener-Feldbrugge S, Treuner-Lange A, Kube M, Reinhardt R, Klages S, Muller R, Ronning CM, Nierman WC et al. 2011. Comparative genomic analysis of fruiting body formation in Myxococcales. *Mol Biol Evol* **28**(2): 1083-1097.

- Janda JM, Abbott SL. 2007. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *Journal of clinical microbiology* **45**(9): 2761-2764.
- Keswani J, Whitman WB. 2001. Relationship of 16S rRNA sequence similarity to DNA hybridization in prokaryotes. *International Journal of Systematic and Evolutionary Microbiology* **51**(Pt 2): 667-678.
- Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* **25**(5): 955-964.
- McNicholas PA. 1994. DNA typing: analysis of restriction fragment length polymorphisms. *Australasian biotechnology* **4**(2): 97-98.
- Meyer W, Aanensen DM, Boekhout T, Cogliati M, Diaz MR, Esposto MC, Fisher M, Gilgado F, Hagen F, Kaocharoen S et al. 2009. Consensus multi-locus sequence typing scheme for *Cryptococcus neoformans* and *Cryptococcus gattii*. *Med Mycol* **47**(6): 561-570.
- Nakhleh L. 2010. Evolutionary Phylogenetic Networks: Models and Issues. In *"The Problem Solving Handbook for Computational Biology and Bioinformatics"* (Heath L, and Ramakrishnan N, Eds), pp. 125-158. Springer.
- Papasian CJ, Kinney J, Coffman S, Hollis RJ, Pfaller MA. 1996. Transmission of *Citrobacter koseri* from mother to infant documented by ribotyping and pulsed-field gel electrophoresis. *Diagn Microbiol Infect Dis* **26**(2): 63-67.
- Pavlic M, Griffiths MW. 2009. Principles, applications, and limitations of automated ribotyping as a rapid method in food safety. *Foodborne Pathog Dis* **6**(9): 1047-1055.
- Pongam P, Osborn TC, Williams PH. 1998. Genetic Analysis and Identification of Amplified Fragment Length Polymorphism Markers Linked to the *alm1* Avirulence Gene of *Leptospaeria maculans*. *Phytopathology* **88**(10): 1068-1072.
- Rossello-Mora R, Amann R. 2001. The species concept for prokaryotes. *FEMS Microbiology Reviews* **25**(1): 39-67.
- Schattner P, Brooks AN, Lowe TM. 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res* **33**(Web Server issue): W686-689.
- Schuler GD. 1997. Sequence mapping by electronic PCR. *Genome Res* **7**(5): 541-550.
- Sedlacek I, Novakova D, Svec P. 2010. Ribotyping and biotyping of *Lactobacillus helveticus* from the koumiss. *European Food Research and Technology* **230**(5): 753-758.
- Seifert KA, Samson RA, Dewaard JR, Houbraken J, Levesque CA, Moncalvo JM, Louis-Seize G, Hebert PD. 2007. Prospects for fungus identification using CO1 DNA barcodes, with *Penicillium* as a test case. *Proc Natl Acad Sci U S A* **104**(10): 3901-3906.
- Stackebrandt E, Goebel BM. 1994. A Place for DNA-DNA Reassociation and 16s Ribosomal-Rna Sequence-Analysis in the Present Species Definition in Bacteriology. *International Journal of Systematic Bacteriology* **44**(4): 846-849.
- Torriani S, Clementi F, Vancanneyt M, Hoste B, Dellaglio F, Kersters K. 2001. Differentiation of *Lactobacillus plantarum*, *L. pentosus* and *L. paraplantarum* species by RAPD-PCR and AFLP. *Syst Appl Microbiol* **24**(4): 554-560.
- Torriani S, Zapparoli G, Dellaglio F. 1999. Use of PCR-based methods for rapid differentiation of *Lactobacillus delbrueckii* subsp. *bulgaricus* and *L. delbrueckii* subsp. *lactis*. *Appl Environ Microbiol* **65**(10): 4351-4356.

- Wang LT, Kuo HP, Wu YC, Tai CJ, Lee FL. 2009. *Lactobacillus taiwanensis* sp. nov., isolated from silage. *Int J Syst Evol Microbiol* **59**(Pt 8): 2064-2068.
- Wang LT, Tai CJ, Wu YC, Chen YB, Lee FL, Wang SL. 2010. *Pseudomonas taiwanensis* sp. nov., isolated from soil. *Int J Syst Evol Microbiol* **60**(Pt 9): 2094-2098.
- Zelinski T. 1991. The use of DNA restriction fragment length polymorphisms in conjunction with blood group serology. *Transfusion* **31**(8): 762-770.
- 張智欽 2011. MicroOrganism Identifier: An effective method to identify lactic acid bacteria. Master Thesis. National Chiao Tung University.
- 何宜佩 2011. The horizontal gene transfer events and alterations of genomic GC content in the genus *Aspergillus*. Master Thesis. National Chiao Tung University.
- 陳之杭 2012. The method of detecting horizontal gene transfer events by applying reference sequences: A case study of *Stigmatella aurantiaca*. Master Thesis. National Chiao Tung University.
- 王煒驊 2012. MicroOrganism Identifier: A web service to identify microorganism. Master Thesis. National Chiao Tung University.

附錄

附件四

附件四 國科會補助專題研究計畫項下出席國際學術會議心得報告

日期：2010 年 5 月 17 日

計畫編號	NSC - 98 - 2621 - B - 009 - 001		
計畫名稱	利用同源基因資訊開發新演算法提供有效鑑種方法及基因水平轉移研究		
出國人員姓名	李應達、潘承宗、何宜佩	服務機構及職稱	交通大學系統生物及生物資訊研究所 學生
會議時間	2010 年 5 月 17 日至 2010 年 5 月 21 日	會議地點	中國大陸，江蘇省蘇州市
會議名稱	(中文) (英文)CSH Asia Conference: Epigenetic, Chromatin & Transcription		
發表論文題目	(中文) (英文)		

一、參加會議經過

Cold Spring Harbor Asia Conferences (CSH Asia)是由 Cold Spring Harbor Laboratory 和中國蘇州工業園區(Suzhou Industrial Park, SIP)共同於獨墅湖畔

的大型會議中心所舉辦。自 2010 年四月起，CSH Asia 舉辦了多樣領域的會議，2010 年 5 月舉辦了 CSH Asia Conference: Epigenetic, Chromatin & Transcription。

這場會議邀請了 Robert Roeder (The Rockefeller University), Danny Reinberg (Howard Hughes Medical Institute, NYU School of Medicine), Richard Young (Whitehead Institute and Massachusetts Institute of Technology) 及 Michael G. Rosenfeld (Howard Hughes Medical Institute, University of California-San Diego) 做為 keynote speaker。

五月 18 及 19 日的下午各安排了一場 poster session，於會場有超過上百位學者展示他們的研究成果。在這段時間裡，交談聲此起彼落，眾多學者互相分享探討交流，十分熱烈分享成果同時提出問題討論。而我最有興趣的是與 transposable element 相關的主題。

二、與會心得

Epigenetic、Chromatin 及 Transcription 的研究在近幾年來已經成為生物學的中心議題，在參與會議的過程中我能夠體會到在這項領域的發展是多麼快速，必需要加緊腳步才能跟上。

非常感謝國科會的經費支持，使得我們這一次出國參加研討會得以成行。於求學其間能夠有一次參與國外大型研討會活動是一次非常難得的經驗，有助於協助國內學者於相關領域研究有了更深入的了解。

附件四

附件四 國科會補助專題研究計畫項下出席國際學術會議心得報告

日期：2012 年 06 月 30 日

計畫編號	NSC - 98-2621-B - 009 - 001		
計畫名稱	利用同源基因資訊開發新演算法提供有效鑑種方法及基因水平轉移研究		
出國人員姓名	方羿喬、王俊霖、陳之杭	服務機構及職稱	交通大學系統生物及生物資訊研究所 學生
會議時間	2012 年 6 月 23 日至 2012 年 6 月 26 日	會議地點	愛爾蘭共和國，都柏林
會議名稱	(中文) (英文) Society for Molecular Biology & Evolution 2012		
發表論文題目	(英文) The Role, Function, and Conservation of Upstream Open Reading Frames in Human (英文) <i>Aspergillus</i> genes found in <i>Zea mays</i> shotgun reads (英文) The horizontal gene transfer events in the genus <i>Aspergillus</i>		

一、參加會議經過

Society for Molecular Biology & Evolution 2012 (SMBE 2012)是國際演化學界年度最大的學術研討會。今年輪在愛爾蘭首都都柏林舉行，會議共 4 天，有來自各國約 1300 名學者參加。會議期間共發表了超過 240 場專題演講及超過 700 篇壁報論文，議題涵蓋從傳統田野演化研究到基因體分子演化的各層面主題。會議中參與投稿三篇壁報論文，並於大會的壁報論文時間與來自各國的參加者有許多討論。

二、與會心得

因為定序技術突破帶來大量基因體序列資料，包括傳統 Sanger 定序及 NGS 定序的資料著實影響了演化的研究方向。尤其是在微生物的演化研究上，除單一 species/strain 定序單一 genome 的方式外，現在有更多人是在看 clone 內的不同個體間的差異，單一 strain 一次就可定序 1000 個 genome。

三、發表論文全文或摘要

The Role, Function, and Conservation of Upstream Open Reading Frames in Human

Yi-Chiao Fang, Arthur Chun-Chieh Shih, Yeong-Shin Lin

1National Chiao Tung University, Hsinchu City, Taiwan, 2Academia Sinica, Taipei, Taiwan

Upstream open reading frame (uORF) is an open reading frame whose start codon (Upstream AUG, uAUG) locate in the 5'UTR of an mRNA. Upstream ORF may affect the translation of main coding sequence by interfering the start of its translation. Although about half of known human transcripts contain at least one uORF, we don't know what roles uORFs play and how uORFs function in human. By analyzing the significant terms of Gene Ontology, we found that uORFs play important roles in metal ion binding and regulation of cellular processes, especially in the regulation of transcription. A uORF that interfere the translation of main coding sequence must be recognized by the ribosome on their uAUG. The features of uAUGs decide which uAUG may be recognized frequently. Beside the uAUG, there are several known features of uORFs that affects the starting of translation of uORF itself and the re-initiation of translation of following ORFs. According to these features, we can estimate how the main coding sequence of an mRNA is affect by the uORFs starting from 5'UTR. We found that dozens of uORFs that frequently affect the translation of main coding sequences are starting from a conserved uAUG. Our results suggest that uORFs may be conserved for their regulation functions and also influence the protein expression of many genes, play an important role in human.

***Aspergillus* genes found in *Zea Mays* shotgun reads**

Chun-Lin Wang, Yi-Pei Ho, Chien-Chi Chen, Shih-Hau Chiu, Li-Min Sung, Tsu-pei Chiu, Yeong-Shin Lin

Institute of Bioinformatics and Systems Biology, National Chiao-Tung University, 75 Po-Ai Street, Hsinchu, Taiwan,

Bioresource Collection and Research Center, Food Industry Research and Development Institute, 331 Shih-Pin Road, Hsinchu, Taiwan

Aspergillus species play an influential role in medical, agricultural and commercial areas. We retrieved the predicted *Aspergillus* open reading frame (AORF) sequences of three species (*A. fumigatus*, *A. nidulans* and *A. niger*) from *Aspergillus Genome Database* and performed TBLASTN analysis against NCBI Whole Genome Database. Surprisingly, we found that some of the AORF sequences were highly conserved in the *Zea mays* genome and EST shotgun reads. There are at least two possibilities to account for this phenomenon: (1) Horizontal gene transfer events might happen between *Aspergillus* and *Zea mays*. (2) *Aspergillus*-contaminated samples of *Zea mays* were used for conducting high throughput sequencing. Our further analyses indicated that although these AORFs hit *Zea mays* EST reads, we could not map them to *Zea mays* genome draft. At present, we postulate that the *Aspergillus* genetic material found might be unexpected contamination in *Zea mays* since *Aspergillus* sp. is usually found to live with crops. Furthermore, we performed *Gene Ontology* analysis and found that many AORFs were related to primary metabolic processes (e.g., carbohydrate, lipid, protein, or amino acid) while some others may have cation transmembrane activity. This unintended finding is important since it provides direct evidence for the expression of some *Aspergillus* genes in crops. Further studies could be useful for *Aspergillus* related researches.

The horizontal gene transfer events in the genus *Aspergillus*

Yi-Pei Ho, Chih-Hang Chen, Yeong-Shin Lin
National Chiao Tung University, Hsinchu, Taiwan

Horizontal Gene Transfer (HGT), also called Lateral Gene Transfer (LGT), plays an important role in evolution of prokaryotes and part of eukaryotes. HGT is the nonsexual process of genetic material transfer across species. In the phylogenetic view, the tree will show a network-like diagram distinct from the vertical transfer. It is well known that HGT events occur frequently among prokaryotes that can often increase fitness to colonize new environment. The common transfer events that people often mentioned is developing drug resistance. However, as genome sequencing era coming, HGTs found in eukaryotes increasingly. HGT events may explain some new traits of organisms and even new disease to human, animals and plants. There are several approaches, which based on genome-wide features, sequence similarity and phylogeny incongruence, can screen out potential HGT cases. Since, there are different possibilities that can contribute to the gene anomalies, HGTs should be confirmed by different approaches, especially phylogeny analysis. In this study, We find one HGT event, transferred from fungi to *Stigmatella aurantiaca*. From the analysis of phylogeny and genome characteristics, the orthologous gene of *Aspergillus clavatus* is the most closed gene sequence to the HGT gene, STAUR_2131. To find out where the sequence come from and any other gene take part in HGT event, we perform a large-scale scan for further study.

四、建議

感謝國科會對本次參與國際研討會的經費支持，使得此次的研討會得以成行，此行除了成功發表計畫相關的壁報論文也獲取了先端的演化研究進程，此行所獲得資訊將有助於往後本國演化領域研究。