

蝴蝶蘭基因資料庫建構 歷程與展望



撰文/蘇春霖·許仁弘·施明哲

緣起

台灣地處熱帶與亞熱帶之交，四面環海，中央有山脈超過三千公尺，生態環境複雜，使得生物群相多元化，生物資源豐沛，具有優良的生物多樣性研究條件。另一方面，台灣的農業自有獨特的發展歷程，曾經是稻米、蔗糖、鳳梨、香蕉等農產作物的大宗輸出國。近年來，芒果、木瓜、蝴蝶蘭、吳郭魚等農漁作物相繼興起，陸續形成有規模的經濟產業。任何產業要維持競爭優勢，永續發展，皆須有不斷的創新與研發，農業生物科技亦然，而推動現代化農業科技研發的基礎，在於基因資訊的開發。

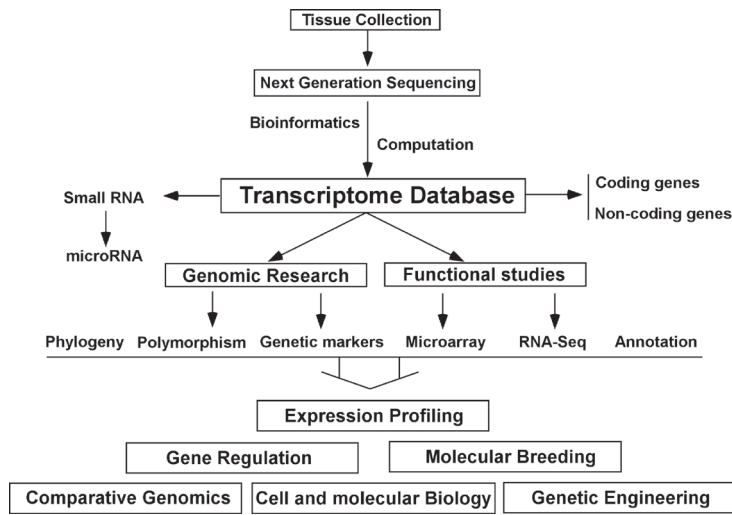
然而，無論是台灣特有的生物，或農產經濟作物，受到諸多主觀與客觀條件的限制，其基因體研究不甚發達。多數農業研究著重栽植生理與傳統育種、選種或引種；而生態研究則著眼調查與登錄。若要進一步深入瞭解基因與環境的互動、生物特性與基因功能的關係、基因的分子調控、物種演化的跡證，或者利用基因重組技術，改良農業作物的性狀，則往往受限於基因資料的不足。然而這些台灣特有生物或農產作物，並非如國際熱衷的模式物種，如水稻、玉米、油菜、大豆等，或者如人類、大小鼠、果蠅等模式物種，累積數量龐大的基因資料，配合積極開發的生物科技，對於其基因功能與調控，或品種改良，都有長足的突破與進步。台灣特有生物或作物往往不是模式物種，因此基因體

資料不足，無法支援生物科技的運作，阻礙研發速度，減緩研發能力的厚實。

基因體研究不發達的原因很多，最主要不外乎資源組合的困難，如成本耗費、資訊處理運算能力不足，與缺乏專業基因體研究人才。新一代的基因定序技術 (next generation sequencing)，不但能縮減成本，而且快速提供數量龐大的序列。但是相對也需要更大的運算能力，以及資訊處理技巧。處理資訊的能力，往往需要專業的生物資訊人才。建立基因體研發團隊與設備，整合多方面人才，加以訓練，需要資本與時間，非一般實驗室所容易建立。

中央研究院農業生物科技研究中心（以下簡稱農生中心）首先著手於台灣蝴蝶蘭的研究。蝴蝶蘭產業在近幾年出口額已超過一億美元，蘭花品種開發及溫室栽培技術已臻世界一流。雖然台灣蘭花育種技術享譽國際，但是在蝴蝶蘭的生長栽培上，仍有不少問題需要解決，例如生長速度緩慢、花期調控、病害防治、新品種需求等。若欲徹底了解基因對於生理表達性狀的調控機制，則應建立蘭花基因表達資料庫（圖一），以利學術與產業的共同發展。

全球蘭花品種繁多，據統計蘭科植物品種約有二萬五千種以上，在台灣文獻上已分類鑑定達三百多種以上，極具生物多樣性。由於地理分布廣，環境適應導致演化上的特異，蘭花具有許多特殊的生物性狀，是模式物種所無，例如蘭花與特殊蟲媒之間的交互演化；蝴蝶蘭採景天酸代謝 (crassulacean



圖一 基因體學研究流程與應用領域

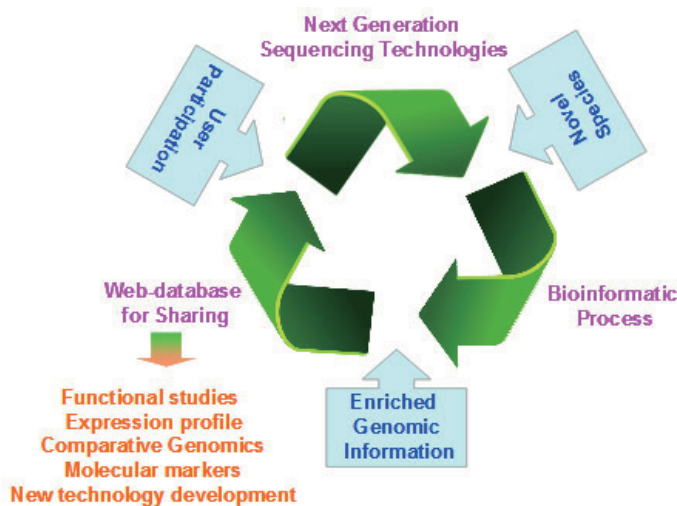
acid metabolism, CAM) 光合作用，以利水分的經濟使用，與一般 C3、C4 植物有異；其著生特性 (epiphytic)，附著於樹幹樹杈之間；極為特殊的種子發育過程，缺少胚乳或子葉，而經歷原球體 (protocorm) 的過渡階段等，都是獨特而值得生物學家探究的課題。但由於缺乏蘭花基因層次的研究工

具，例如育種判定基因標誌 (genetic markers) 與蘭花基因資料庫等，以致蘭花在生理研究方面無法快速發展。

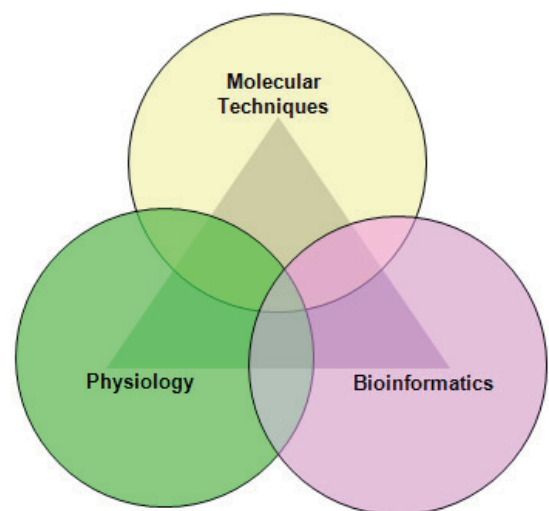
在蝴蝶蘭基因體研究上，由於原生種白花蝴蝶蘭「台灣阿嬤」(*Phalaenopsis aphrodite* Rchb. f.) 是商業雜交品系的重要親本，原生地分布於台灣東南部，具有大白花、多花、花型開放等優良性狀，因此中研院農生中心施明哲特聘研究員兼主任的研究團隊優先以「台灣阿嬤」(*Phalaenopsis aphrodite*) 為主題進行基因解碼。此外，近年來文心蘭的切花輸往日本，產量逐年增加，成為重要的蘭花產業，因此也針對其近親但迷你株型，且生命週期快速的扇型文心蘭 (*Erycina pusilla*) 進行基因體研究。

農業物種基因體學研究的推動

為了促進基因體學的研究，中研院農生中心推動基因體研發策略 (圖二 A)，結合最新的高速定序技術、生物資訊處理，以及資料庫分享三部曲，逐步收集不同蘭花的表達基因與基因組序列資料，加以註解、整理，並融入功能性研究成果，構築基



A. 基因體研究策略三構面
(Strategic trilogy of genomic research)



B. 結合跨領域能力團隊
(Cross-discipline talent team)

圖二 台灣蝴蝶蘭基因體研究分工圖

因資料庫。此一研發策略期望能充分發揮時間與資源的經濟效益，建立的流程強調連貫順暢，得到的基因序列組合兼具正確性與完整性。透過分享，使用者也能參與資訊的產生和校正。

基因解碼工作需要多種專業領域的專家參與，以詳細分工方式節省研究資源與時間。所幸獲得中研院經費輔助與農業生物技術產業化發展方案的產業化之創新生技研究計畫支持，在中研院農生中心施明哲主任、蘇春霖博士、台灣大學園藝系張耀乾副教授與元智大學資訊工程系趙雅婷助理教授的合作之下，結合分子生物技術、植物生理及生物資訊處理專長（圖二 B），以此分工模式進行蝴蝶蘭表達基因的研究工作。

蝴蝶蘭基因解碼工程

研究材料來自於二倍體原生種蝴蝶蘭各器官組織（如根、莖、葉、花、果莢與種子）的表達基因（expressed sequences），利用中央研究院快速定序中心的高速定序儀（next generation sequencer, 主要有 Roche 454 Titanium，以及 Illumina Solexa 兩大技術平台）（圖三），解碼得到片段序列。接著進行序列組合（assembly），利用基因片段間的重疊序列銜接成為單一基因（unigene），再將這些基因與現有基因資料庫比對（例如阿拉伯芥、水稻基因庫等），將比

對的結果註解（annotation）記錄，彙整成基礎表達基因資料庫，並建立專屬網站，提供學界與業界的蘭花基因研究人員使用。

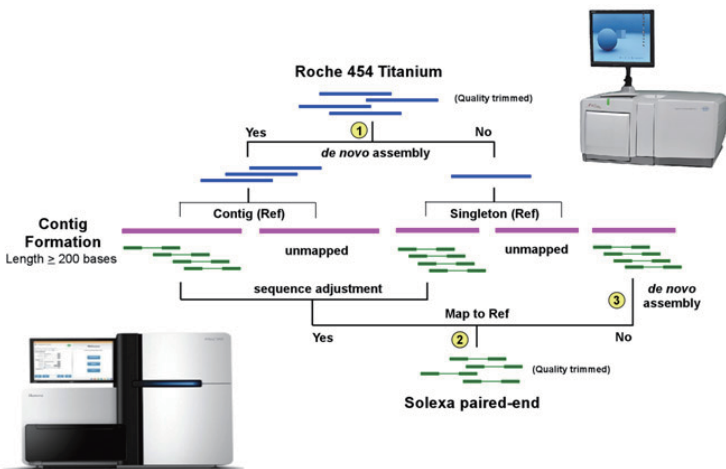
蝴蝶蘭表達基因對應於已知資料庫而能註解的基因，超過四萬多筆，屬於有蛋白質編碼的表達基因。此外，另有十九萬筆基因屬於非蛋白質編碼者，負擔基因調控功能，例如 non-coding RNA，或者 miRNA precursor。生物資訊處理團隊針對這類非蛋白質基因，也正進行註解與研究。

除了表達基因序列之外，研究團隊也致力於小片斷 RNA (small RNA) 的研究，由於 small RNA 之中，有一類 microRNA 具有調節基因表現量的功能，因此研究方法是把蘭花各部位組織如葉、根、花及發育中種子的 small RNA library，也送交高速定序儀 Solexa 進行序列分析，得到的序列比對 miRBase，若發現蘭花的 microRNA，亦將列入表達基因資料庫之中。

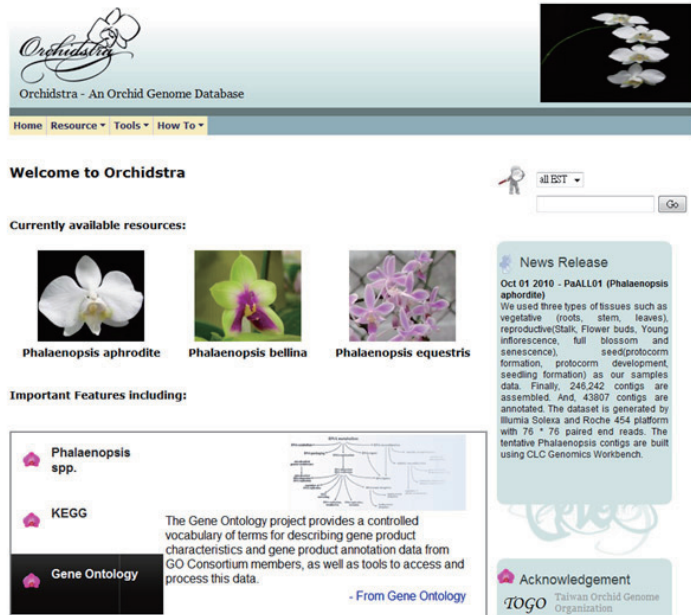
台灣蝴蝶蘭基因註解資料庫

目前已建立 Orchidstra 台灣蝴蝶蘭基因註解資料庫 (Phalaenopsis Genome Annotation Database) 網站 (<http://orchidstra.abrc.sinica.edu.tw/>)（圖四），Orchidstra 的命名是將 orchid（蘭花）與 orchestra（交響樂團）組合成為一個字，具有多重資訊交互輝映的意思。此資料庫內容包含台灣原生種蝴蝶蘭「台灣阿嬤」基因註解總計 233,823 筆重疊群（contigs），其中 42,590 筆重疊群已被註解為已知基因，191,233 筆重疊群為未知之轉錄序列（unknown transcripts）。

對於蝴蝶蘭基因表達有興趣的使用者，在 Orchidstra 網站上不僅可以利用 Blast 尋找 homologous 基因，另外也提供其他的註解連結包括 KEGG PATHWAY（圖五）、GO (Gene Ontology)（圖六）、Pfam domain 等，並具備瀏覽、搜尋、比對、連結、下載等功能。此資料庫已經開放供國內外蘭花研究學者使用，無須註冊申請，資料陸續更新之中。



圖三 中央研究院快速定序中心的高速定序儀技術平台



資料來源：<http://orchidstra.abrc.sinica.edu.tw/>

圖四 Orchidstra台灣蝴蝶蘭基因註解資料庫 (Phalaenopsis Genome Annotation Database)

應用於設計蘭花生物晶片領域

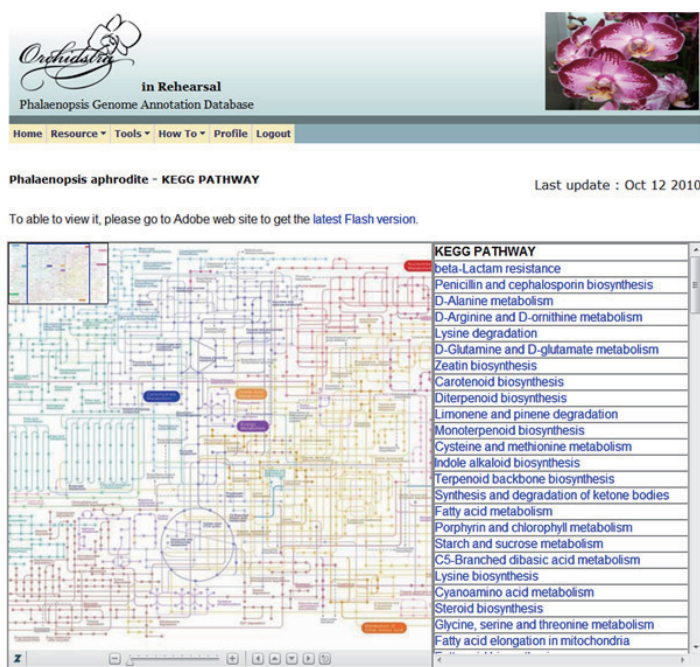
根據現有基因資料(共 43,662 target genes, 67,038 probes) 可應用於蘭花微陣列 (Orchid microarray) 生物晶片設計, 目前第一代蘭花生物晶片已完成測試, 此一晶片涵蓋大於 90% 台灣蝴蝶蘭註解基因, 品質穩定度與再現性極高。測試於商業雜交品種, 目前流行的大白花 V3 品系, 晶片訊號與原生種之間的關聯性亦高, 顯示可直接用於來自商業雜交種的試驗樣本。未來將可利用蘭花微陣列生物晶片進行基因表達圖譜分析 (gene expression profiling) 研究, 探討蘭花生理基因表達功能之機制。經過驗證與改良, 亦已完成第二代晶片更新版本 (42,957 target genes, 42,973 probes), 有效降低成本。蘭花專用晶片結合資料庫的註解, 兩者的價值更相得益彰。

結論與展望

由中研院農生中心施明哲主任團隊建構的 Orchidstra 蝴蝶蘭基因註解資料庫, 以台灣原生種蝴蝶蘭「台灣阿嬤」與迷你文心蘭為材料, 利用高速定序技術解碼與大量基因資訊的處理平台, 成功建立基因表達資料庫。根據這些蝴蝶蘭表達基因序列, 又進行專屬微陣列 (microarray) 晶片設計, 提升蘭花功能基因體 (Functional Genomics) 的研究能力。

在 small RNA 基因表現分析方面, 已定序完成根、花、葉、種子的 small RNA, 目前正在進行 microRNA 與 pre-miRNA 的分析。在蘭花生理研究方面, 仍持續針對種子發育過程、CAM 光合作用, 以及低溫催花現象, 利用基因表達圖譜研究 (gene expression profiling) 方式進行功能基因體研究, 期能找到蘭花獨特生理反應的關鍵基因, 改進蝴蝶蘭栽培的生理穩定性。

目前 Orchidstra 蝴蝶蘭基因註解資料庫的資料仍不斷地在更新中, 台灣阿嬤蝴蝶蘭的表達基因已完成解碼與註解工作, 並開放給學界與業界研發人員使用, 希望藉由基因表達層次的解析, 打開台

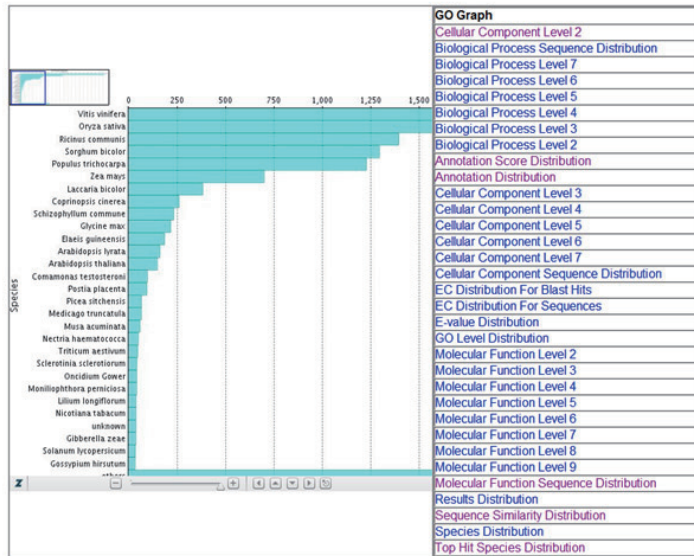


圖五 台灣蝴蝶蘭基因註解資料庫提供KEGG PATHWAY的註解連結功能

Phalaenopsis aphrodite - GO's Analysis chart

Last update : Oct 12 2010

To able to view it, please go to Adobe web site to get the latest Flash version.



圖六 台灣蝴蝶蘭基因註解資料庫提供GO (Gene Ontology) 的註解連結功能

灣蝴蝶蘭基因的神祕面紗，解開生理作用調控之奧秘，協助解決蘭花產業育種與生長栽培上的問題。借助這個計畫的演練，發展的高通量基因組合與註解流程已證明極有效率，不僅縮短研發時間，並產出更多、更精準的基因序列。未來希望能擴而大之，將其廣泛應用於更多台灣農業作物或特有生物之物種的基因研究，提升台灣生物科技與農產業長遠的發展。

AgBIO

蘇春霖 中央研究院 農業生物科技研究中心
資深博士後研究員
許仁弘 農業生物技術產業化推動辦公室 資深博士後研究員
施明哲 中央研究院 農業生物科技研究中心
特聘研究員兼主任
農業生物技術產業化推動辦公室 策略規劃組召集人

參考文獻

1. Metzker, M. L. (2010) *Sequencing technologies - the next generation*. Nat. Rev. Genet. 11:31-46.
2. Meyer, E., Aglyamova, G. V., Wang, S., Buchanan-Carter, J., Abrego, D., Colbourne, J. K., Willis, B. L. and Matz, M. V. (2009) *Sequencing and de novo analysis of a coral larval transcriptome using 454 GSFLX*. BMC Genomics 10:219.
3. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. and Wold, B. (2008) *Mapping and quantifying mammalian transcriptomes by RNA-Seq*. Nat. Methods 5:621-628.
4. Pridgeon, A. M., Cribb, P. J., Chase, M. W. and Rasmussen, F. N. (2005) *Genera Orchidacearum*, Vol. 4, Epidendroideae (part 1). Oxford University Press, Oxford, UK. pp. 435-437.
5. Shin, H., Hirst, M., Bainbridge, M. N., Magrini, V., Mardis, E., Moerman, D. G., Marra, M. A., Baillie, D. L., Jones, S. J. (2008) *Transcriptome analysis for Caenorhabditis elegans based on novel expressed sequence tags*. BMC Biol. 6:30.
6. Wall, P. K., Leebens-Mack, J., Chanderbali, A. S., Barakat, A., Wolcott, E., Liang, H., Landherr, L., Tomsho, L. P., Hu, Y., Carlson, J. E., Ma, H., Schuster, S. C., Soltis, D. E., Soltis, P. S., Altman, N. and dePamphilis, C. W. (2009) *Comparison of next generation sequencing technologies for transcriptome characterization*. BMC Genomics 10:347.
7. Orchidstra台灣蝴蝶蘭基因註解資料庫，From <http://orchidstra.abrc.sinica.edu.tw/>。
8. 農業生物技術產業化發展方案，From <http://dpiab.sinica.edu.tw/>。