

# 解碼牛樟基因體闡明被子植物的演化

趙淑妙<sup>1</sup>、吳宗賢<sup>1</sup>、蔡怡陞<sup>1</sup>

## 前言

牛樟(*Cinnamomum kanehirae*)是木蘭亞綱(Magnoliidae)、樟科(Lauraceae)、樟屬(*Cinnamomum*)林木。牛樟取其名「粗壯如牛」，是臺灣最雄偉的闊葉樹之一(圖1)。此一臺灣特有且珍貴的林木卻因其寄生真菌—牛樟芝—而受到無妄之災。其實，人類採取牛樟芝對牛樟樹而言是幫其移除病害，對人類及牛樟是雙贏的效果，但不肖人士利慾薰心，以殺雞取卵的方式盜伐牛樟木，製成椴木培養牛樟芝，反造成牛樟面臨滅絕的危機。撇開牛樟芝不談，牛樟樹本身即是非常難得的天然紀念物，是臺灣最珍貴的闊葉樹五木之一。解碼牛樟的基因體更為學術界的發展貢獻良多，使我們得以窺探被子植物的DNA分子及基因體演化之謎。



圖1 四人合抱的粗壯牛樟木。感謝特生中心陳志輝研究員的引路，得以目睹此天然紀念物的風采。

本文即以牛樟基因體解序為切入點，報導牛樟基因體在闡明被子植物的親緣關係、基因體複雜性、全基因體複製及萜類合成酶基因等方面的研究成果，亦從基因體的角度探討牛樟的保育方向。文中內容多截取於2019年發表於自然植物期刊(Nature Plants)的研究性論文(Chaw et al. 2019)。讀者若對牛樟的基因體深感興趣，想進一步獲得更多的資訊，請詳閱該期刊論文。

## 牛樟基因體的特性及複雜度

牛樟的染色體為二倍體，共12對。以流式細胞儀估算牛樟的基因體約為 $823.7 \pm 58.2$  Mb(百萬鹼基對)。利用最新的定序技術，我們組裝出它的12條染色體。牛樟基因體序列的雜合度(Heterozygosity)平均為0.54%，且在染色體不同區域上呈現高度變異性：部分區域的雜合度極低，可能是族群瓶頸效應造成選汰掃除(Selective sweep)的結果(圖2a)。進一步的分析亦指出在九百萬年前，牛樟的有效族群規模開始持續地萎縮(圖2b)，這可能與臺灣頻繁的造山運動及複雜的地形構造有關，而近期的不當伐木及盜伐則加速牛樟族群縮減的進程，因此維護或提高基因體的多樣性是牛樟保育工作的首要課題。

如同其他植物，牛樟富含多樣性的轉位子(Transposable element)及重複序列(Repetitive sequences)，佔基因體的五成左右。這些轉位子以長末端重複序列反轉錄轉位子(Long terminal

<sup>1</sup> 中央研究院生物多樣性研究中心

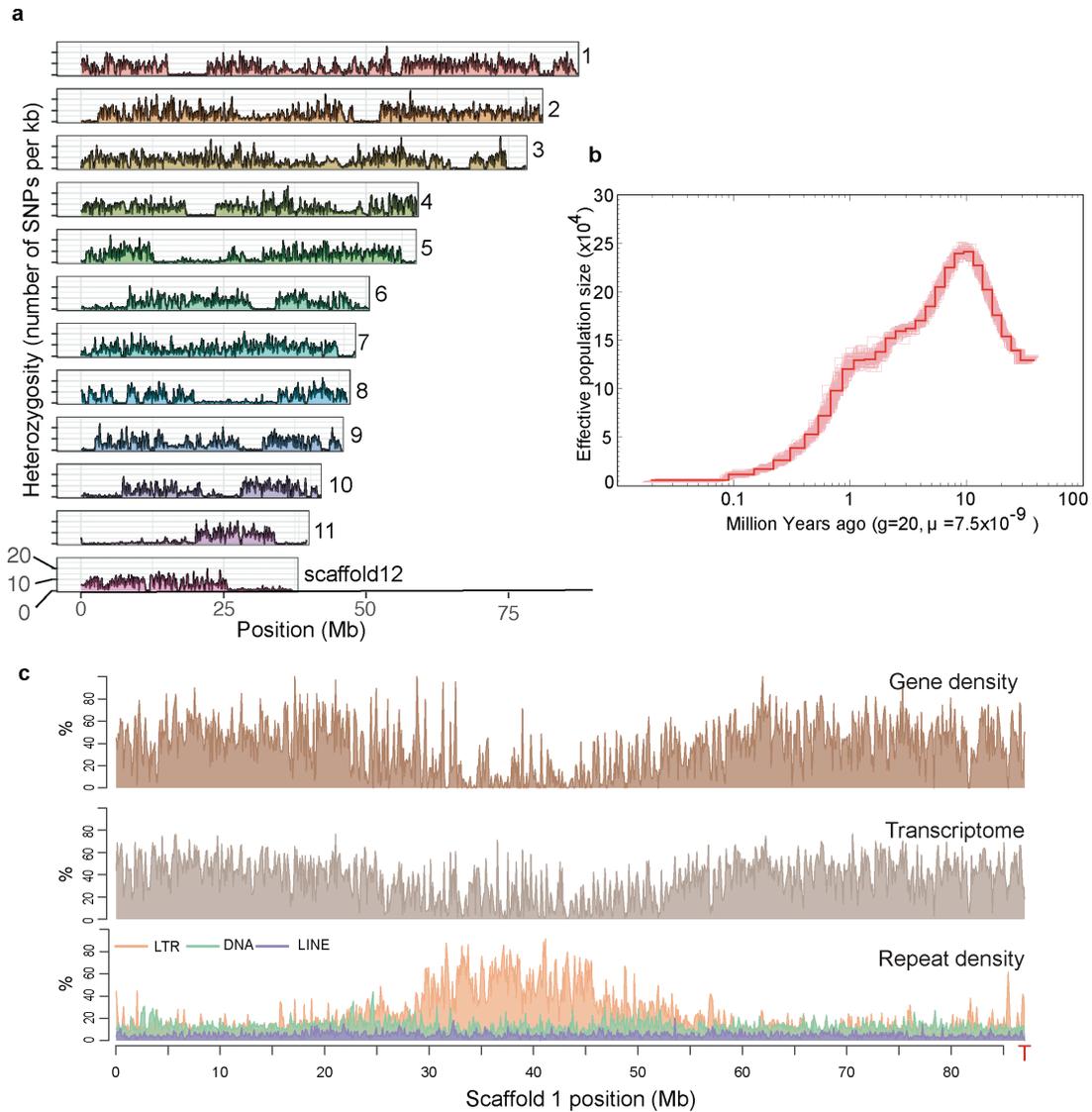


圖2 牛樟基因體的雜合度分析：a.雜合多型性位點在染色體的分布情形。b.牛樟有效族群大小的變動。c.牛樟基因密度(上方的直方圖)、對應基因的表現量(中間)及轉位子密度的分布情形(下方)。圖片截取自Chaw et al. (2019)。

repeat (LTR retrotransposon]為主，其次為DNA轉位子。我們分析並發現某些反轉錄轉位子具獨特性，表示牛樟近期發生過轉位子增殖事件，進而提高牛樟基因體的複雜度。此外，牛樟的基因與轉位子的分布呈現相反的趨勢，基因密

度在染色體兩端較高而在中心區域較低。反之，轉位子則多聚集在染色體中間區域(圖2c)。

### 從牛樟基因體推論全基因體複製事件 全基因體複製(Whole genome duplication)

是驅動植物演化的重要機制(Jiao et al., 2011)。然而，在缺漏木蘭亞綱基因體的情況下，使學界無法建立被子植物全基因體複製的演化藍圖，而發表牛樟基因體則填補了此一重要的缺口。牛樟染色體間的互相比較，發現超過七成的區域呈現共線性(Synteny)，證實牛樟曾經歷全基因體複製事件(圖3a)。我們比對無油樟(*Amborella*)，目前被認為是最早演化出的被子植物與牛樟的基因體，結果顯示同源基因的比對數為1:1~4，也就是1個無油樟的基因可以對到1至4個牛樟的基因(圖3b)。由於無油樟的演化沒有經歷過全基因體複製事件(Zhong & Betancur-R, 2017)，這表明牛樟至少經歷過兩次的全基因體複製事件。而全基因體複製事件的發生時間可經由比較物種間同源基因的同義性(Synonymous substitution rate)取代速率來進行評估，我們推論木蘭亞綱的演化至少經歷兩次全基因體複製事件，一次發生在樟目(Laurales)與木蘭目(Magnoliales)分歧之前，另一次則發生在樟科與其他植物分歧之後。

## 牛樟萜類合成酶基因的演化

牛樟全株(如葉子和樹幹)含豐富且多樣的芳香精油。我們從牛樟基因體註解出101種萜類合成酶(Terpenoid synthase)基因，其數量遠高於其他被子植物。這些萜類合成酶基本上可分成六個主要的基因家系：TPS-a、TPS-b、TPS-c、TPS-e、TPS-f、及TPS-g(圖4)。TPS-c及TPS-e參與雙萜類合成，其他家系則與單萜類及倍半萜類合成相關。牛樟的萜類合成酶基因在染色體上呈現聚落狀地分布，亦即同一家系的萜類合成酶基因往往聚在一起，形成類似基因簇(Gene cluster)的現象，暗示串連重複(Tandem duplication)機制可能是驅動牛樟萜類合成酶巨量演化的重要因子。

牛樟芝的專一性寄生可能與牛樟豐富的萜類相關。牛樟芝的子實體可生產高達78種萜類，其中包含31種三萜類。大部分的三萜類是以單萜及倍半萜類為骨架，再通過環化作用產生。因此可合理推測，牛樟芝經由分解牛樟木

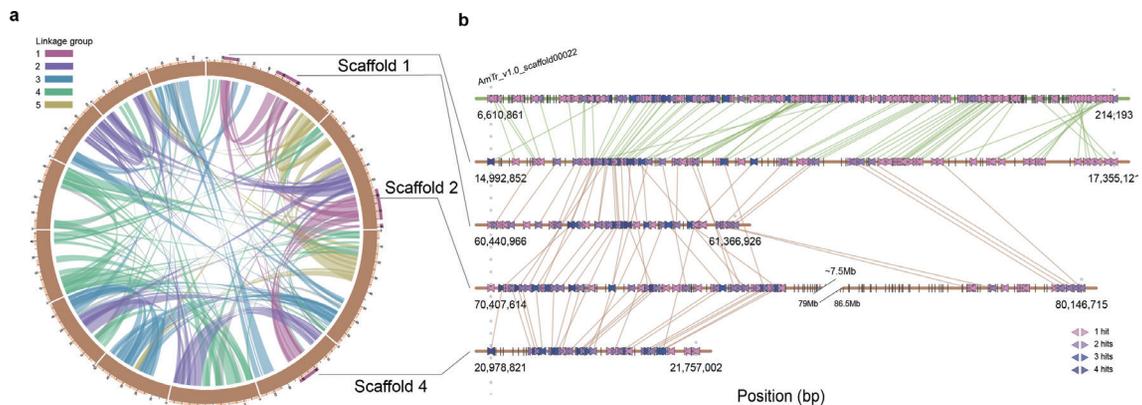


圖3 牛樟基因體比對分析：a.牛樟染色體相互比對顯示約72.7%的區域呈現共線性。b.比較無油樟與牛樟的基因體，發現1:1~4的關係。圖片截取自Chaw et al. (2019)。

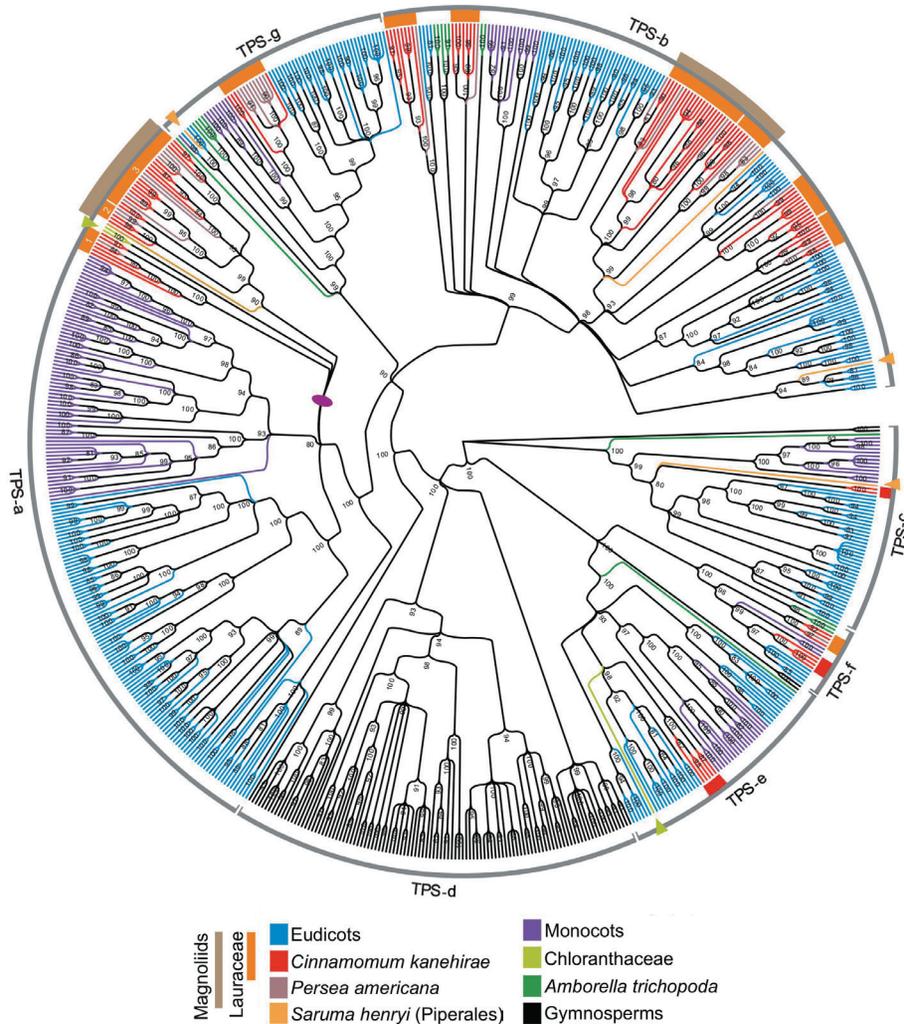


圖4 萜類合成酶基因樹。TPS-d為裸子植物特有，不存在於牛樟基因體。修改自Chaw et al. (2019)。

材得到中間產物，合成多樣且豐富的萜類。

### 牛樟基因體釐清木蘭亞綱的演化地位

木蘭亞綱包含約9,000種植物，是被子植物的第三大類群，僅次於真雙子葉植物 (Eudicots)及單子葉植物。木蘭亞綱的成員與人類文明發展息息相關，例如：胡椒及肉豆蔻

可製成香料及調味料、酪梨及釋迦為重要的經濟水果、樟樹與牛樟為常見的庭院樹種，也是製造家具和雕刻素材的良好材料。然而學界多年來對木蘭亞綱的演化地位爭議不斷，總結出五種主要的演化假說，分別是：木蘭亞綱(1)與金粟蘭科(Chloranthaceae)為姊妹群 (Moore et al., 2007, APG IV system, 圖5a-1)；

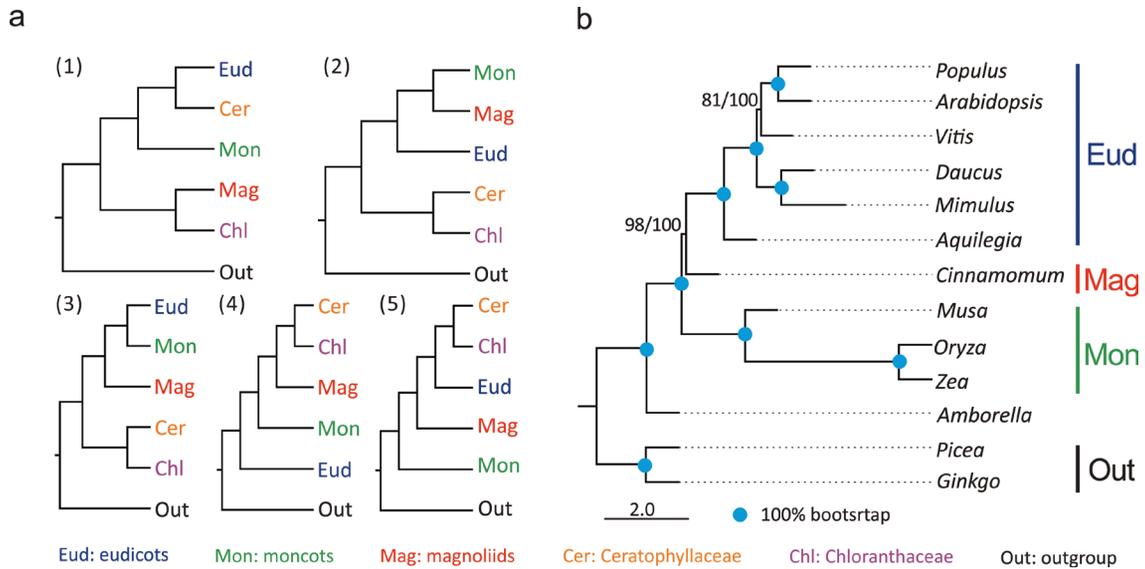


圖5 木蘭亞綱的演化爭議：a.五種主要的爭議性假說。b.211個單拷貝基因所建構的演化樹，演化樹結構以合併法(Coalescence)為基礎，支持度小於100%的支系則標出其阿拉伯數字的百分比，其排列為：合併法/串聯法(Concatenation)。演化樹修改自Chaw et al. (2019)。

(2)或與單子葉植物為姊妹群(Endress & Doyle, 2009, 圖5a-2)；(3)或與真雙子葉及單子葉植物為姊妹群(Qiu et al., 2010, 圖5a-3)；(4)或與金魚藻科(Ceratophyllaceae)及金粟蘭科為姊妹群(Zhang et al., 2012, 圖5a-4)；(5)或與金魚藻科、金粟蘭科及真雙子葉植物所構成的分類群為姊妹群(Zeng et al., 2014, 圖5a-5)。

牛樟代表第一個基因體被完整解序的木蘭亞綱植物，使我們得以從基因體的角度，重新檢驗木蘭亞綱的演化地位。我們篩選出211個單拷貝(Single-copy)基因，以串聯法(Concatenation)及近期發展出來的合併法(Coalescence)，建構親緣演化樹。串聯法是將所有基因串聯成超級矩陣(Super-matrix)再建構演化樹；合併法則是將單獨基因所建構的演化樹進行彙整，此法較能詮釋個別基因在演化上的差異。兩種不同方法獲得一致的結果：皆支

持木蘭亞綱與真雙子葉植物為姊妹群，且較疏遠於單子葉植物(圖5b)。前人的研究多利用胞器(Organelle)基因，或轉錄體建構演化樹。為避免被多拷貝(Multiple-copy)基因所產生的親緣演化雜訊(Phylogenetic noise)干擾，我們僅以單拷貝基因建構演化樹，而結果顛覆以往學界對木蘭亞綱演化地位的看法，亦暗示木蘭亞綱的演化史並非如傳統分類學者所想像的古老。

### 牛樟、樟樹、及冇樟的親緣關係

牛樟與冇樟是形態上極為相似的樹種，這兩樹種的混淆造成苗木培育及造林上的困擾。牛樟與冇樟在形態上最明顯的區別為果實與胚軸：牛樟果實為僧帽狀，胚軸與果柄成一直線；冇樟果實則呈長橢圓形，胚軸常歪生(林, 1993)。但上述的特徵僅見於結果期的成熟林木，幼苗或非結果期的植株仍難以判斷種

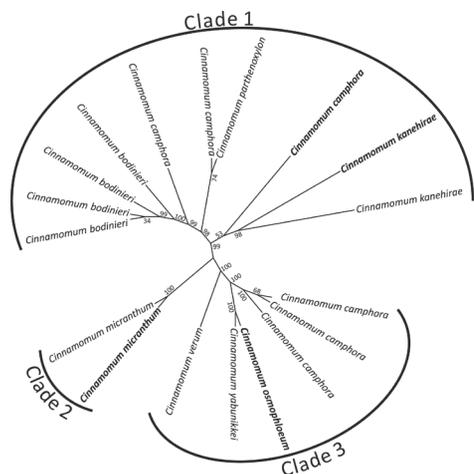


圖6 質體基因體親緣關係樹。粗體者為筆者團隊定序的個體。阿拉伯數字為節點的支持度(%)。

別。此外，牛樟與冇樟相似的形態，是否意味它們的親緣關係非常緊密？而它們與樟樹的關係又為何？這些演化議題值得深入探討。

受惠於次世定序所帶來的便利性與成本降低，基因資料庫的質體基因體(Plastid genome = plastome)，近幾年在數量上呈現飛躍式的成長。迄今有7種樟屬植物15個個體的質體基因體序列被發表，加上我們自己定序的三種樟屬植物，分析這些質體基因體有助於評估牛樟、樟樹及冇樟的親緣關係。我們以全質體基因體序列及最大似然法(Maximum likelihood)建構親緣演化樹(圖6)。18個樟屬植物被區分成三個分類群(Clade)，牛樟被歸類在第一分類群，冇樟為第二分類群，但樟樹被歸類在第一及第三分類群，呈現并系群(Paraphyly)的現象。

我們採自臺灣天然族群的樟樹個體與牛樟被歸類至同一分類群，且是牛樟的最近親種，雖然支持度僅53%；而冇樟被歸類在另一分類群，與牛樟的關係較疏遠。這結果揭

示：(1)牛樟與冇樟的取樣個體分別形成獨立且疏遠的單系群(Monophyly)，因此質體基因體是區別牛樟與冇樟的利器，適合開發成基因條碼系統(Genetic barcoding system)。(2)造成樟樹并系群現象的原因不明，追溯定序的團隊分別來自臺灣及中國，可能因族群地域性而造成并系群。總之，我們的初步分析顯示：臺灣產的樟樹與牛樟親緣關係最近，但此結論仍需更多樣化的資料，如核基因體及粒線體基因體，加以佐證。

## 結論與展望

在水稻基因體(2005)圖譜問世的15年後，第一個木蘭亞綱植物—牛樟的基因體被完整解序且發表，此重要物種的基因體資料，填補了我們被子植物基因體演化研究上的缺口，亦改寫以往學界對木蘭亞綱演化地位的認知，同時也宣告後牛樟基因體時代的來臨。我們發表的基因體資料達到染色體層級，將扮演完善的參考指標，加速牛樟族群基因體學的研究進展，亦可應用在開發功能性基因位點及分子標誌等方面，促進牛樟的保育與林木產業應用。

## 致謝

本研究通訊作者為中央研究生物多樣性中心趙淑妙特聘研究員和蔡怡陞副研究員。美國喬治亞大學植物生物系教授James H. Leebens-Mack提供轉錄體數據和闡釋分析結果。研究經費則由中央研究院深耕計劃和主題計畫(趙淑妙)及前瞻計畫(蔡怡陞)支持。質體基因體定序的部分感謝行政院科技部經費 (MOST 103-2621-B-001-007-MY3) 和中央中央研究院生物多樣性研究中心的研究經費(趙淑妙)。