國立臺灣大學生命科學院基因體與系統生物學學位學程

博士論文

Genome and Systems Biology Degree Program

College of Life Science

National Taiwan University

Doctoral Dissertation

紅豆杉科與傘松科植物色質體基因組分析:洞悉其核質體 DNA 與重組基因群

Study on the Plastomic Organizations of Taxaceae and Sciadopityaceae: Insights into Their Nuclear Plastid DNAs and Chimeric Gene Clusters

許智堯

Chih-Yao Hsu

指導教授：趙淑妙 博士

Advisor: Shu-Miaw Chaw, Ph.D.

中華民國 105 年 7 月

July 2016

doi:10.6342/NTU201601257

國立臺灣大學博士學位論文
口試委員會審定書

紅豆杉科與傘松科植物色質體基因組分析:洞悉其核

質體 DNA 與重組基因群

Study on the Plastomic Organizations of Taxaceae and Sciadopityaceae: Insights into Their Nuclear Plastid DNAs and Chimeric Gene Clusters

本論文係許智堯君（D00b48013）在國立臺灣大學基因體與系統生物學學位學程完成之博士學位論文，於民國 105 年 1 月 26 日承下列考試委員審查通過及口試及格，特此證明

口試委員：＿＿＿＿＿＿＿＿（簽名）

（指導教授）

基因體與系統生物學學位學程主任 ＿＿＿＿＿＿＿＿（簽名）

doi:10.6342/NTU201601257

致謝

一轉眼間五年的時間過去了!在此特別感謝指導教授趙淑妙 博士在這段時間的指導與提攜；以及吳宗賢 博士於研究上的指導與經驗分享,使得我的研究可以順利進行;感謝 Edi Sudianto 不厭其煩地幫我修改論文英文;感謝科技部、中研院對於本研究的經費支持及台灣大學提供的研究生獎學金。感謝論文口試期間王亞男 教授、可文亞教授、林崇熙 教授、莊樹諄 教授、陳豐奇 教授及蔡怡陞 教授的批評與指正，使得本論文可以修改的更加完備。

此外亦要感謝王婷浈學妹當我口試的記錄,以及試驗室裡趙秀鳳小姐與許許多多的學弟妹,陪我度過了這幾年的實驗室生活,有他們的幫助與關心使得實驗室的生活充滿了樂趣與希望,在此致上最深的感謝與祝福。

當然也要感謝學程裡的同學們,有他們的陪伴讓我在學校時不會孤單一個人,真的很幸運可以與大家一起修課、討論、學習成長,我想這也是博士生涯中最值得懷念的部分了。祝福你們研究順利。

最後,本論文謹獻給我最親愛的家人,感謝你們的支持與鼓勵,讓我無後顧之憂的完成學業。

祝福各位身體健康!再次謝謝你們。

doi:10.6342/NTU201601257

TABLE OF CONTENTS

doi:10.6342/NTU201601257

doi:10.6342/NTU201601257

doi:10.6342/NTU201601257

LIST OF FIGURES

doi:10.6342/NTU201601257

doi:10.6342/NTU201601257

LIST OF TABLES

doi:10.6342/NTU201601257

摘 要

針葉樹植物（conifers)分為松科（Pinaceae）與柏門（Cupressophyta）兩大群，其中以柏門植物最具多樣性及高經濟價值。現生的柏門植物又分為五個科，共約有 400 多種。種子植物（seed plants）色質體基因組（plastomes)結構相當保守，但柏門植物的色質體基因組結構卻有高度的變異性。為了更深入的探討色質體基因組重組在演化上的意義與影響，本論文利用比較色質體基因體學的方法研究以下兩個題目。

第一，提出新的方法篩選及研究紅豆杉科植物核質體 DNA (nuclear plastid DNA or *nupt*)。核質體DNA是一群由色質體基因組轉移到核基因組中的DNA片段。核質體 DNA 的移轉為核基因組提供了豐富的遺傳資源，也提高了核基因組的遺傳多樣性。但目前為止，沒有針葉樹植物核質體 DNA 的相關研究，因此本研究定序了台灣穗花杉（*Amentotaxus formosana*）及台灣紅豆杉（*Taxus mairei*）的完整色質體基因組，並利用比較基因體學的方法分析穗花杉、紅豆杉及粗榧屬的色質體基因組排列方式，進而推測出紅豆杉科植物祖先色質體基因組的組成與排列方式。由此，我們設計專一性的引子來增幅祖先型及現生紅豆杉科植物色質體基因組的非共線型區域（non-syntenic region)。利用這方法我們一共篩選出 12.6 kb 的核質體 DNA，這些核質體 DNA 明顯地累積較多 GC 變成 AT 的突變。此外，藉由比較祖先型核質體 DNA 與現生核質體 DNA 的 *rps8* 基因，我們發現現生核質體 DNA 的 *rps8* 基因轉譯起始碼子包含了一個 C 變成 U 的 RNA 修飾。我們的研究進一步的指出紅豆杉科植物的色質體基因組大約在白堊紀就已經轉移到核基因組中，這些核質體 DNA 不僅保留了祖先型色質體基因組排列方式與核酸組成，也提供了線索，讓我們可以瞭解與研究針葉樹色質體基因組的演化過程。

第二，利用實驗的方法驗證與探討色質體基因組序列重組對於傘松科植物演化上的意義與影響。種子植物的色質體基因轉錄時大多是以操縱組（operon）的模式進行，也就是多個基因同時轉錄（polycistronic transcription)；種子植

doi:10.6342/NTU201601257

物的色質體基因組操縱組相當保守，即使像針葉樹的色質體基因組排列方式經過多次的重組，但它們的操縱組也很少遭到破壞。本研究定序了完整的傘松(*Sciadopitys verticillata*)色質體基因組。傘松科植物目前僅存一種，基因體比較分析後發現傘松的色質體基因組有三項特點: 色質體基因組排列經過多次的重組、包含了三組連續並退化的 tRNA 基因 (*trnV-GAC, trnQ-UUG* 及 *trnP-GGG*)、有一個特殊的反向重複序列（inverted repeat)，而這個反向重複序列可形成不同構型（isomeric）的色質體基因組。此外，傘松的色質體基因組因重組打斷了多個原本的操縱組，而這些斷裂的操縱組卻重新組合而形成四個重組基因群（chimeric gene clusters)。我們的數據顯示這些新的重組基因群保留有原本的啟動子(promoter），且可以順利地轉錄出其相對應的 RNA 序列。本研究結果使我們能更深入的了解為何針葉樹植物色質體基因組具有多樣性及高複雜度的特性。

關鍵詞: 色質體基因組、基因組重組、演化、核質體 DNA、基因群、紅豆杉科、傘松科、針葉樹

doi:10.6342/NTU201601257

ABSTRACT

Cupressophytes (Cupressophyta or conifers II) is the largest of the two conifer groups, the other being Pinaceae. Cupressophytes comprise about 400 species in five families. They are the most diversified and economically valuable group in conifers. Our previous studies showed that gene organizations of their plastid genomes (plastomes) are highly variable among the few elucidated plastomes of cupressophytes. To further decipher the evolution of plastomic re-organization, I carried out comparative plastome studies of two cupressophytes families; Taxaceae and Sciadopityaceae.

Here we reported two major findings. First, I proposed a new strategy for identification and evolutionary studies of nuclear plastid DNAs (*nupts*) in Taxaceae. Plastid-to-nucleus DNA transfer provides a rich genetic resource to the complexity of plant nuclear genome architecture. To date, the evolutionary fates of *nupt* remain unknown in conifers. We have sequenced the complete plastomes of two yews, *Amentotaxus formosana* and *Taxus mairei*. Comparative plastomic analyses revealed possible evolutionary scenarios for plastomic reorganization from ancestral to extant plastomes in the three sampled Taxaceae genera, *Amentotaxus, Cephalotaxus, and Taxus*. Specific primers were designed to amplify non-syntenic regions between ancestral and extant plastomes, and 12.6 kb of *nupts* were identified based on phylogenetic analyses. These *nupts* have significantly accumulated GC-to-AT mutations, suggesting a nuclear mutational environment shaped by spontaneous deamination of 5-methylcytosin. The ancestral initial codon of *rps8* is retained in the *Taxus nupts*, but its corresponding extant codon is mutated and requires C-to-U RNA-editing. These findings suggest that *nupts* can help

doi:10.6342/NTU201601257

recover scenarios of the nucleotide mutation process. We also demonstrated that the Taxaceae *nupts* we retrieved may have been retained since the Cretaceous period of Mesozoic Era and they carry the information of both ancestral genomic organization and nucleotide composition, which offer clues for understanding the plastome evolution in conifers.

Second, we used experimental data to show the evolutionary impact of plastomic rearrangements in Sciadopityaceae. Many genes in the plastid genomes of seed plants are organized in polycistronic transcription units known as operons. These plastid operons are highly conserved, even among conifers whose plastomes are highly rearranged. We sequenced the complete plastome sequence of *Sciadopitys verticillata* (Japanese umbrella pine), the sole member of Sciadopityaceae. The *Sciadopitys* plastome is characterized by extensive inversions, pseudogenization of tRNA genes after tandem duplications, and a unique pair of inverted repeats involved in the formation of isomeric plastomes. We showed that plastomic inversions in *Sciadopitys* have led to the shuffling of remote operons, resulting in the birth of four chimeric gene clusters. Our data also suggested that these chimeric gene clusters have adopted pre-existing promoters for the transcription of genes. This newly deciphered plastome of *Sciadopitys* advances our current understanding of how the conifer plastomes have evolved toward increased diversity and complexity.

Keywords: Plastome, Genomic reorganization, Evolution, *Nupt*, Gene cluster, Taxaceae, Sciadopityaceae, Conifer.

doi:10.6342/NTU201601257

CHAPTER 1

Background and Significance

1.1 General Characters of Plastids and Plastomes

Plastids are specialized organelles found in land plants and green algae. They usually contain diverse pigments and are responsible for photosynthesis. Chloroplast contains a type of plastids with green pigments, also known as chlorophylls. The chloroplast is enclosed by two envelope membranes that consist of an inner and an outer membrane (Wise et al., 2006). Chloroplasts use light energy and carbon dioxide as resources to produce starch and oxygen, which maintains the atmospheric oxygen level and provides essential energy for all of the lives on earth (Bryant et al., 2006).

Plastids were once free-living cyanobacteria engulfed by a eukaryotic precursor about two billion years ago (Martin et al., 2002; Archibald, 2009). They have their own genomes but the genome sizes are relatively small as compared to those of free-living prokaryotic ancestors (Figure 1). For example, the genomes of two reported cyanobacteria, *Nostoc* PCC7120 and *N. punctiforme,* are 6.4 Mb and 9 Mb in sizes with ~5,400 and ~7,200 proteins encoded, respectively (Timmis et al., 2004). In contrast, the plastid genomes (plastomes) of seed plants have an average size of about 145 kb with only 20 to 200 proteins encoded (Timmis et al., 2004; Jansen RK, 2012). This extreme reduction of plastomes was hypothesized as a result of bulk gene transfers from plastomes to the nucleus during early plastome evolution (Martin et al., 1998; Sheppard et al., 2008). Previous studies demonstrated that during the evolutionary history of plastomes, the vast majority of cyanobacterial genes were either lost or transferred to the nucleus and mitochondria of their host cells (Timmis and Scott, 1983; Mochizuki et

doi:10.6342/NTU201601257

al., 2008). For instance, a comparative study of the three genomes of *Arabidopsis* (nuclear, mitochondrial, and plastid genomes) and the cyanobacterial genomes indicated that *Arabidopsis* nuclear genome includes about 4,100 genes that are of cyanobacterial origin and approximately 1,300 of these were sent back to plastid (Kleine et al., 2009) (Figure 2). These data suggest that the DNA transfer from plastid to nucleus resulted in massive relocation of organelle genes and shaped the size of nuclear genome during organelle evolution.

Interestingly, the transfer of genes or DNA fragments between plastomes and nucleus is still an on-going process in flowering plants. Michalovova et al. (2013) analyzed the nuclear-encoded plastid DNA (*nupt*; Richly and Leister, 2004) in six completely sequenced plant species (*Arabidopsis thaliana, Vitis vinifera, Sorghum bicolor, Glycine max, Oryza sativa,* and *Zea mays*). Their results indicated that most *nupts* are close to centromeres and that longer *nupts* showed lower divergence from plastid DNA than shorter *nupts* in these six species. Taken together, these data suggested that *nupts* are newly transferred into nuclear genomes. Why do plastids tend to transfer genes to the nucleus? Population genetic study indicated that deleterious mutations can accumulate rapidly in asexual populations but can be minimized through sexual recombination (Muller, 1932, 1964; Moran, 1996; Allen, 2015; de Vries et al., 2016). Therefore, to avoid the deleterious mutations in the asexual plastids, gene transfer from the plastome to nucleus could increase the recombination rates, reduce the genetic load of plastomes (Martin et al., 1998), and contribute to a great diversity of new genetic materials for the generation of new genes (Timmis et al., 2004).

1.2 Plastomic Organization

Plastomes of most land plants are highly conserved in size and structure. In general,

doi:10.6342/NTU201601257

the plastomes of seed plants are circular with conserved genomic organization, including a large single copy (LSC), a small single copy (SSC), and a pair of inverted repeats (IRs) (Palmer 1983, 1991). The size of seed plant plastomes ranges from 120 to 160 kb, depending on different species (Palmer, 1985). The typical IRs of seed plant plastomes are about 20 to 25 kb.

However, a number of seed plant groups have lost (or highly reduced) one of their IRs. The IRs were likely independently lost at least five times in seed plants phylogeny (Jansen and Ruhlman, 2012). Within angiosperms, loss of an IR has been found in the plastomes of Fabaceae (Wojciechowski et al., 2004), Geraniaceae (Downie and Palmer, 1992) and two genera of Orobanchaceae (Palmer et al., 1991). Among gymnosperms, IR loss was only reported in conifers, Pinaceae, and cupressophytes (Raubeson and Jansen, 1992; Lin et al., 2010; Wu, Lin et al., 2011; Hsu et al., 2014). Recent comparative plastomic analyses suggested that Pinaceae lost IR_B and that cupressophytes lost $IR_{A,}$ and that plastomes of conifers not only have lost an IR but also exhibited frequent plastome rearrangements (Wu, Wang et al., 2011; Wu et al., 2014). However, whether these genome organizations are evolutionarily adaptive or associated with the loss of IR remain to be investigated.

1.3 Plastomic Rearrangements

Plastome rearrangements have been considered an evolutionary adaptation because they can create new gene clusters or isomeric forms (Cui et al., 2006). As mentioned earlier, most of the seed plant plastomes are structurally conserved. However, there are some exceptional cases in both angiosperms and gymnosperms.

In angiosperms, Campanulaceae, Fabaceae, and Geraniaceae have experienced bursts of gene order changes (Jansen and Ruhlman, 2012). Comparative analyses among

doi:10.6342/NTU201601257

18 Campanulaceae plastomes indicated that their genome orientation changed at least 42 times, including 18 large insertions (over 5 kb), five IR expansions/contractions, and several small inversions (Cosner, 1993; Cosner et al., 2004). Fabaceae are known to exhibit a high degree of gene order change in their plastomes, including the loss of IR, transfer of plastomic genes to the nucleus, intron losses, gene duplications, and inversions (Milligan et al., 1989; Gantt et al., 1991; Doyle et al., 1995; Doyle et al., 1996; Wojciechowski et al., 2004; Cai et al., 2008; Magee et al., 2010). The plastomes of Geraniaceae are highly rearranged because of the genome inversions and expansion/contraction of IR regions (Chumley et al., 2006; Guisinger et al., 2008; Blazier et al., 2011; Guisinger et al., 2011; Weng et al., 2013). For example, the plastome of *Pelargonium hortorum* has undergone at least 12 inversions and eight IR expansion/contraction changes (Chumley et al., 2006). However, reconstruction of evolutionary scenarios for these genome changes is not possible due to insufficient sampling. More genome sequencing data within each genus will be required to construct a reliable evolutionary model (Jansen and Ruhlman, 2012).

In gymnosperms, cupressophytes have been reported as the only group that exhibits a high level of gene order change. *Cryptomeria japonica*, the first completed plastome in cupressophytes, accumulates a lot of direct and inverted repeats, and at least 15 inversions in its plastome (Hirao et al., 2008). The gene order and genome structure of *Cryptomeria* plastomes are significantly different compared to previously reported land plant plastomes. In addition, similar to the plastome of *Cryptomeria*, those of *Agathis dammara, Nageia nagi, Calocedrus formosana,* and four *Juniperus* species also have extensive rearrangements (Guo et al., 2014; Wu and Chaw, 2014). However, to date, no experimental data has ever demonstrated that the rearranged plastomes of cupressophytes can create new co-transcriptional gene clusters.

doi:10.6342/NTU201601257

1.4 Transcription of Plastid Genes

Plastids are plant organelles that posses their own genomes and transcription systems (Allen, 2015). Because of their cyanobacterial origin, plastids retain a prokaryotic-type transcription system and a eubacterial RNA polymerase (Igloi and Kössel, 1992; Ishihama, 2000). Therefore, plastids and eubacteria have a similar gene expression system. Most of the plastid-encoded genes are arranged in operons. These genes in operons are co-transcribed into polycistronic RNA precursors, some of which may be post-processed and cut into monocistronic units for translation (Barkan, 1988; Stern et al., 2010; Meierhoff et al., 2003). Moreover, their operon organizations are usually conserved among higher plants (Jansen and Ruhlman, 2012). The plastids of higher plants include at least two types of plastomic transcription systems—the plastid-encoded RNA polymerase (PEP) and the nuclear-encoded plastid RNA polymerase (NEP) (Shiina et al., 2005).

PEP, a bacterial-type gene expression system, is a highly efficient transcription component in higher plants and contributes over 80% of all plastid transcripts (Zhelyazkova et al., 2012). Basically, it is composed of two major subunits, the core enzyme and sigma factor subunits. The core enzyme subunit is encoded by plastomes (*rpoA, rpoB, rpoC1,* and *rpoC2*) and catalyzes the activity of RNA synthesis (Ohyama et al., 1986; Hudson et al., 1988; Sexton et al., 1990). These four *rpo* genes are present in the plastomes of photosynthetic plants and algae and share high sequence similarity with their bacterial counterparts (Morden et al., 1991; Sakai et al., 1998). In contrast, the activity of the plant PEP system needs additional nuclear-encoded subunit, called sigma factor. These sigma factor subunits are encoded by nuclear genome (*SIG1, SIG2, SIG3, SIG4, SIG5,* and *SIG6*) and responsible for recognition of promoter regions and transcription initiations (Shirano et al., 2000; Kanamaru et al., 2001; Hanaoka et al.,

doi:10.6342/NTU201601257

2003; Privat et al., 2003; Tsunoyama et al., 2004; Ishizaki et al., 2005) (Table 1). Allison (2000) suggested that multiple sigma factors may lead the PEP to specific promoters during plastid and plant development.

NEP, a phage-type RNA polymerase, has high sequence similarity with both T3/T7 phage and mitochondrial polymerase enzymes (Allison et al., 1996; Kapoor et al., 1997; Gray and Lang, 1998; Hedtke et al., 2000;). NEP is specifically encoded by three nuclear genes (*RpoTp, RpoTm,* and *RpoTmp*) (Table 1) (Hedtke et al., 1997). The *rpoT* family is present in most plant species (Ortelt and Link, 2014). RpoTp proteins play an important role in a catalytic subunit of NEP and serve as the second transcription activity in higher plant plastids (Young et al., 1998; Kobayashi et al., 2001). RpoTmp proteins might be responsible for the early stage of seedling development and in the greening process of leaves. Experimental evidence from *Arabidopsis* showed that RpoTmp-deficient mutants had a significant delay of the greening process, altered leaf shape, and a defect in the light-induced accumulation of several plastid mRNAs (Baba et al., 2004). In addition, a biochemical study on spinach showed an unidentified nuclear-encoded RNA polymerase, NEP-2 (Bligny et al., 2000). In contrast to NEP, the NEP-2 containing fractions do not include a phage-like enzyme. NEP-2 can recognize the T7 and *rrn*Pc promoters, and is responsible for the transcription of rRNA operon (Shiina et al., 2005).

1.5 Regulation of Plastid Gene Transcription

Regulation of plastid gene transcription is still unclear. Three major mechanisms might account for their regulation. First, there is a partition of labor in which PEP transcribes genes associated with photosynthesis and NEP transcribes housekeeping genes (Hajdukiewicz et al., 1997; Pfannschmidt, Nilsson, Tullberg et al., 1999;

doi:10.6342/NTU201601257

Pfannschmidt, Nilsson, and Allen, 1999). Liere and Maliga (2001) proposed that NEP plays a major role at early stage of plastid development and transcribes plastome-localized genes involved in PEP core enzyme subunit. Subsequently, it was demonstrated that PEP does transcription during plastid development and is responsible for the transcription of photosynthesis-related genes (Liere and Maliga, 2001).

Second, transcription of plastid genes might be regulated by multiple promoters (Legen et al., 2002; Liere and Börner, 2007; Börner et al., 2015). An alternative promoter has evolved to perform plastomic transcription because many plastid genes are transcribed by both the NEP and PEP promoters. For example, Legen et al. (2002) analyzed the plastid transcription profiles of the entire plastome from a wild-type and PEP-deficient tobacco. Their results indicated that the functional integration of PEP and NEP is feasible and that the plastid genes are completely transcribed well in both wild-type and PEP-deficient tobacco.

Third, the entire plastome can be transcribed via read-through transcriptions which allow the expressions of downstream of 3' untranslated regions (Quesada-Vargas et al., 2005). Northern-blot analyses confirmed that read-through transcripts were present and took about 30% of total transcripts in the transgenic tobacco plastomes (Quesada-Vargas et al., 2005). Moreover, they constructed a transgenic line that disrupts the 16S *rrn* operon by insertion of a foreign operon. The transcription of disrupted 16S *rrn* operon should be affected by the terminator of the foreign operon (Quesada-Vargas et al., 2005). Then the downstream genes of the foreign operon would not be transcribed and the chloroplast protein synthesis would be changed. However, transcription of the whole 16S *rrn* operon was detected in this transgenic line and the phenotype of this transgenic lines and wild-type were similar (Quesada-Vargas et al., 2005). These results indicated that the read-through transcripts were sufficient for chloroplast development.

doi:10.6342/NTU201601257

1.6 Applications of Plastomic sequences for Addressing Plant Evolution

Plastomes provide rich information for resolution of phylogenies at different elevels of plants because of their lower nucleotide substitution rates. For example, Lin et al. (2010) reported a comparative plastomic study to resolve the previously controversial classifications in Pinaceae. They used 49 plastomic protein-coding genes common to 19 gymnosperms, including 15 species from eight Pinaceous genera, to reconstruct the phylogenetic trees of Pinaceous genera. Their phylogenetic tree suggested that *Cedrus* is clustered with *Abies-Keteleeria* and that *Cathaya* is closer to *Pinus* than to *Picea* or *Larix-Pseudotsuga*. Their molecular dating also suggested that Pinaceae first evolved during Early Jurassic and diversified during mid-Jurassic and Low Cretaceous. Furthermore, Seong and Offner (2013) used the *matK* gene to construct a phylogenetic tree of conifers. Their study linked the phylogeny and phenotype dates of conifers, like leaf type, seeds, and cones. They hypothesized that the environmental selection was the major force that drives the evolution of leaf phenotypes in conifers. Their study also supported the hypothesis that North American pines originated from Asian pines.

In addition, structural characters of plastomes, such as gene order inversions, genome rearrangements, expansion/contraction of IR, loss/gain of genes, and disruption of operons, can serve as good resources for phylogenetic inference (Raubeson and Jansen, 2005; Wu and Chaw, 2014; Hsu et al., 2016). A phylogenetic study of 81 plastid genes common to 64 seed plants showed a positive correlation among the numbers of gene order changes, gene losses, and lineage-specific rate accelerations (Jansen et al., 2007). The structural changes of plastomes generate a large amount of variability and some of these variations are accumulated during plastome evolution (Maréchal and Brisson, 2010).Wu and Chaw (2014) inferred phylogenetic trees of gymnosperms based on the plastome organizations. They used the matrices compiled from locally collinear

doi:10.6342/NTU201601257

blocks (LCBs) of plastomic architectures to construct a phylogenetic tree for extant gymnosperms. Their phylogenetic trees provided structural evidence for supporting the gnepines hypothesis and for the loss of different IR copies in Pinaceae and cupressophytes. They also suggested that at least two mechanisms, mutational burden (genome reduction) and rearrangement association (genome expansion), are involved in the variation of plastomic size in cupressophytes.

1.7 Plastid Inheritance

Studies have shown that plastid DNA is mainly inherited from the maternal parent in angiosperms (Corriveau and Coleman, 1988; Mogensen, 1996; Birky, 2003; Zhang et al., 2003). To date, only *Actinidia speciose* reportedly has the paternal inheritance of plastids in angiosperms (Testolin and Cipriani, 1997). About 80% of angiosperms species inherit plastids from their maternal parents, while the remaining species inherit them from both parents (biparental inheritance) (Jansen and Ruhlman, 2012). Hu et al. (2008) argued that maternal inheritance is the ancestral feature and that maternal inheritance has been converted to biparental inheritance and evolved independently among derived lineages. Phylogenetic studies also indicated that changes in the mode of inheritance are unidirectional. There have been no report of the occurence of maternal inheritance from biparental or paternal ancestors. However, Hansen et al. (2006) showed that intraspecific crosses of Passifloraceae resulted in primarily maternal-inherited plastids, whereas interspecific crosses had paternal-inherited plastids. This may be due to a functional incompatibility between interspecific genomes, which fails to exclude paternal DNA.

In contrast to angiosperm plastids, most of the gymnosperm plastids are paternally inherited (Stine et al., 1989). Mogensen (1996) further indicated that cycad, ginkgo, and

doi:10.6342/NTU201601257

gnetophyte plastids are of maternal inheritance, and most of the conifer plastids are of paternal inheritance except *Cryptomeria* (Ohba et al., 1971) and *Larix* (Szmidt et al., 1987). It is still a mystery why plastid inheritances are so variable among gymnosperms and angiosperms. To fully understand the variation of the modes of plastid inheritance, a broader examination of every representative genus from gymnosperms and angiosperms will be required.

1.8 What are Gymnosperms?

Gymnosperms are a group of seed plants, characterized by their “naked seeds” (Conway, 2013). Their ovules are naked prior to fertilization. Most of the gymnosperm seeds are borne on the surface of woody, scales, which form a cone. In contrast, angiosperm seeds are covered with mature ovaries or fruits. There are about 1,000 extant species of gymnosperms in five major groups: Pinaceae (conifers I), cupressophytes (conifers II), cycads, ginkgo, and gnetophytes (Gymnosperms on The Plant List, May 2016). Among living gymnosperms, conifers are the most species-rich group, followed by cycads, gnetophytes, and ginkgo.

Gymnosperms are extremely diversified and difficult to classify based on their morphological characteristics alone. In the past decades, molecular data have been widely used to re-examine the traditional classifications of gymnosperms. Chaw et al. (2000) suggested that the extant gymnosperm orders are monophyletic, and none of them alone is a sister of angiosperms. Later, phylogenetic studies further indicated that cycads are sister to ginkgo (Conway, 2013; Wu et al., 2013). However, the placement of gnetophytes has been controversial. After the studies of Chaw et al. (2000) and Bowe et al. (2000), most studies agreed that gnetophytes are a sister clade to the Pinaceae, widely known as the “gnepines” hypothesis.

doi:10.6342/NTU201601257

Conifers are evolutionarily more recent than the other groups of gymnosperms (Seong and Offner, 2013). They are the most abundant extant group of gymnosperms, containing six families and over 600 species (Conway, 2013). Almost all conifers are trees and evergreens. Although conifers contain much fewer species than the flowering plants, they are ecologically and economically important. They dominate forests in high latitudes of Northern and Southern Hemisphere and high altitudes of tropical and subtropical areas.

1.9 Why Cupressophyta?

Cupressophyta (common name: cupressophytes), the most diversified and valuable group of conifers, include ca. 400 species in five families: Araucariaceae, Cupressaceae, Podocarpaceae, Sciadopityaceae, and Taxaceae. They are of great economic value in terms of wood production, resins, pharmaceutical drugs, and horticulture. Notably, their plastome organization is diverse (Hirao et al., 2008; Wu, Lin et al., 2011), therefore, they are ideal materials for studying the evolution and mechanisms of plastome rearrangements.

Taxaceae (the yew family), a family of cupressophytes, comprises 28 species in six genera: *Amentotaxus, Austrotaxus, Cephalotaxus, Pseudotaxus, Taxus,* and *Torreya*. They are mainly distributed in the Northern Hemisphere. *Amentotaxus* includes five species restricted to subtropical Southeastern Asia, from West Taiwan across Southern China to Assam in the eastern Himalayas and south of Vietnam (Cheng et al., 2000). The genus *Taxus* include seven species, best known for their anti-cancer component taxol. They commonly occur in the understories of moist temperate or tropical mountain forests (de Laubenfels, 1988).

doi:10.6342/NTU201601257

Among the conifer families, Sciadopityaceae is the only one that includes the sole member *Sciadopitys verticillata* (abbreviated *Sciadopitys*), which is an evergreen tree that can reach 27 m tall. Its spectacular needle-like leaves are arranged in whorls, like an umbrella. Thus, *Sciadopitys verticillata* is commonly called the Japanese umbrella pine. Three genome-based (Chaw et al., 2000) and plastome-based (Rai et al., 2008) phylogenetic studies are congruent in placing the genus *Sciadopitys* as sister to Taxaceae and Cupressaceae. Recent molecular dating suggests that *Sciadopitys* diverged from other cupressophytes more than 200 million years ago (Crisp and Cook, 2011). Although *Sciadopitys* is considered a living fossil endemic to Japan, paleobiogeographic evidence indicates that its ancestors existed in China during the early and middle Jurassic (Jiang et al., 2012).

1.10 Research Purposes

As mentioned above, the plastomes of cupressophytes are highly rearranged. Therefore, they are ideal materials for studying the following two projects.

First, we aimed to propose a new strategy for the identification and evolutionary study of *nupts* in yews. Previously, most of the studies on *nupts* were based on a comparison between nuclear and plastid genomes. This approach, however, is impractical because the nuclear genome sizes of cupressophytes are huge (ranging from 6.3 to 20 Gb/1C; Murray BG., 1998). In this study, we propose an approach to study *nupts* based on comparative plastomes. A proof-of-concept study using yew plastomes is described (in Chapter 2 below).

Second, we aimed to understand the evolutionary effect of the plastomic rearrangements. Two questions were asked: (1) Do plastomic rearrangements alter the transcription, translation, or end products of plastomes? And (2) Can these

doi:10.6342/NTU201601257

rearrangements disrupt any functional operons and create new gene-clusters? To date, consequences of genomic rearrangements are still poorly understood in gymnosperms. Sciadopityaceae provides a unique opportunity to address these questions as its plastome is highly rearranged. Here we used experimental data to approach these two questions (Chapter 3).

doi:10.6342/NTU201601257

CHAPTER 2

Ancient Nuclear Plastid DNA in the Yew Family

2.1 Introduction

Previous comparative genomic studies indicated that on average about 14% of the nuclear-encoded proteins were acquired from the cyanobacterial ancestor of plastids (Deusch et al., 2008). Transgenic experiments also demonstrated a high frequency of plastid-to-nucleus transfers with one event per 11,000 pollen grains or per 273,000 ovules (Sheppard et al., 2008).

Nupts have been discovered in a large number of plant species (Smith et al., 2011). *Nupts* can contribute to nuclear exonic sequences (Noutsos et al., 2007) and play an important role in plant evolution. *Nupts* may be initially inserted close to centromeres and then fragmented and distributed by transposable elements (Michalovova et al., 2013). The amount of *nupts* in plants is associated with the nuclear genome size and the number of plastids per cell (Smith et al., 2011; Yoshida et al., 2014).

Studies on *nupts* remains limited to plant species that have complete sequences of both nuclear and plastid genomes. In nuclear genomes, *nupt* rearrangements may resemble that of plastomes or consisted of mosaic DNA derived from both plastids and mitochondria (Leister, 2005; Noutsos et al., 2005). Notably, a 131-kb *nupt* of rice was found to harbor a 12.4-kb inversion, which was likely the ancestral characters in the plastome before the transfer (Huang et al., 2005). Recently, Rousseau-Gueutin et al. (2011) proposed a PCR-based method to amplify *nupts* containing a specific ancestral sequence that was deleted from the plastomes of viable offspring. Hence, ancestral plastomic characteristics, such as unique indels and gene orders of specific fragments,

doi:10.6342/NTU201601257

may be retained in *nupts*. Construction of an ancestral plastomic organization should yield valuable clues to retrieve *nupts*. If a plastomic inversion can distinguish an ancestral plastome from its current counterpart, appropriate primers based on the ancestral plastomic organization should be able to amplify the corresponding *nupts* that were transferred to the nucleus before the inversion (Figure 3).

Although the first known *nupt* was identified more than 3 decades ago (Timmis and Scott, 1983), *nupts* of gymnosperms still remain poorly studied. Conifers, the most diverse gymnosperm group, possess huge nuclear genomes ranging from 8.3 to 64.3 pg (2C) (reviewed in Wang and Ran, 2014) and may have integrated many *nupts*. The plastomes of conifers are highly rearranged, possibly due to their common loss of a pair of large inverted repeats (Wicke et al., 2011). Numerous plastomic rearrangements have been identified and are useful in reconstructing phylogenetic relationships among taxa and inferring intermediate ancestral plastomes (Wu and Chaw, 2014). Therefore, the conifer plastomes are well suited for evaluating the feasibility of retrieving *nupts* and surveying their evolution.

In this study, we aim to demonstrate our approach (Figure 3) for mining *nupts* in yews, and to continue the understanding of the plastome evolution in conifers. To better reconstruct ancestral plastomes of yews, we sequenced two complete plastomes, one from each of the yew genera *Amentotaxus* and *Taxus*. The primers based on the recovered ancestral plastomic organization were used to amplify potential *nupts*. The origins of obtained *nupt* candidates were then examined by phylogenetic analyses and mutation preferences to ensure that they were indeed transferred plastomic DNA in the nucleus. Here, for the first time, we demonstrate that conifer *nupts* can be PCR-amplified using our approach and that ancestral plastomic characteristics retained in *nupts* can be compared with extant ones, providing valuable information for

doi:10.6342/NTU201601257

understanding plastome evolution in conifers.

2.2 Materials and Methods

2.2.1 DNA Extraction, Sequencing, and Genome Assembly

Young leaves of *Amentotaxus formosana* and *Taxus mairei* were harvested in the greenhouse of Academia Sinica and Taipei Botanical Garden, respectively. Total DNA was extracted with modified CTAB method with 2% polyvinylpyrrolidone (Stewart and Via, 1993). The DNA was qualified by a threshold of both 260/280 = 1.8–2.0 and 260/230 > 1.7 for next-generation DNA sequencing on an Illumina GAII instrument at Yourgene Bioscience (New Taipei City, Taiwan). For each species, approximately 4 GB of 73-bp paired-end reads were obtained. These short reads were trimmed with a threshold of error probability < 0.05 and then de novo assembled by use of CLC Genomic Workbench 4.9 (CLC Bio, Aarhus, Denmark). Contigs with sequence coverage of depth greater than 50X were blasted against the nr database of the National Center for Biotechnology Information (NCBI). Contigs with hits for plastome sequences with E-value < 10^{-10} were retained for subsequent analyses. Gaps between contigs were closed by PCR experiments with specific primers. PCR amplicons were sequenced on an ABI 3730xl DNA Sequencer (Life Technologies).

2.2.2 Genome Annotation and Sequence Alignment

Genome annotation involved the use of DOGMA with default option (Wyman et al., 2004). Transfer RNA genes were explored by using tRNA scan-SE 1.21 (Schattner et al., 2005). For each species, we aligned the annotated genes and their orthologous genes of other known conifer plastomes to confirm gene boundaries. Sequences were aligned using MUSCLE (Edgar, 2004) implemented in MEGA 5.0 (Tamura et al., 2011).

doi:10.6342/NTU201601257

2.2.3 Exploration of Single-Nucleotide Polymorphisms (SNPs), Indels, and Simple Sequence Repeat (SSR) Sequences

To estimate the distribution of both SNPs and indels between our newly sequenced plastome of *T. mairei* and the *T. mairei* voucher NN014 (NC_020321), the two genomes were aligned by using VISTA (Frazer et al., 2004). The alignment was then manually divided into non-overlapping bins of 200 bp according to the position of our newly sequenced *T. mairei* plastome. Both SNPs and indels in each bin were estimated by using DnaSP 5.10 (Librado and Rozas, 2009). SSRs of the *T. mairei* plastome were explored using SSRIT (Temnykh et al., 2001) with a threshold of repeat units > 3.

2.2.4 Construction of Ancestral Plastomic Organization

We performed whole-plastomic alignments between the two yews under study and other conifers, *Calocedrus formosana* (NC_023121), *Cephalotaxus wilsoniana* (NC_016063), *Cryptomeria japonica* (NC_010548), *Cunninghamia lanceolata* (NC_021437), and *Taiwania cryptomerioides* (NC_016065), to detect locally collinear blocks (LCBs) using Mauve 2.3.1 (Darling et al., 2010). The yielded matrix of LCBs was used to reconstruct the putative ancestral plastomic organizations on MGR 2.03 (Bourque and Pevzner, 2002), which seeks the minimal genomic rearrangements over all edges of a most parsimonious tree.

2.2.5 PCR Amplification, Cloning, and Sequencing

Ten pairs of specific primers used for amplification of *nupt* sequences in Taxaceae were manually designed and their sequences and corresponding locations are in Table 2 and Figure 4. PCR amplification involved the use of long-range PCR Tag (TaKaRa LA Taq, Takara Bio Inc.) under the thermo-cycling condition 98℃ for 3 min, followed by

doi:10.6342/NTU201601257

30 cycles of 98℃ for 15 s, 55℃ for 15 s, and 68℃ for 4 min, and a final extension at 72℃ for 10 min. Amplicons were checked by electrophoresis. Amplicons with expected lengths were collected and cloned into yT&A vectors (Yeastern Biotech Co., Taipei) that were then proliferated in *E. coli*. Sequencing the proliferated amplicons involved M13-F and M13-R primers on an ABI 3730xl DNA Sequencer (Life Technologies).

2.2.6 Phylogenetic Tree Analysis

Maximum likelihood trees were inferred from sequences of potential *nupts*, their plastomic counterparts, and their orthologs in other gymnosperms using MEGA 5.0 (Tamura et al. 2011) under a GTR + G (4 categories) model. Supports for nodes of trees were evaluated by 1,000 bootstrap replications.

2.2.7 Estimation of Mutations in Nuclear Plastid DNAs and Their Plastomic Counterparts

The sequence for each *nupt* was aligned to the homologous plastome sequences for *A. formosana, C. wilsoniana, T. maire* and *C. lanceolata* using MUSCLE (Edgar, 2004). To precisely calculate the mutational preference in *nupts*, all ambiguous sites and gaps were removed from our alignments. Nucleotide divergence between *nupts* and their plastomic counterparts were derived from mutations in either of these two sequences. A mutation in *nupt* or its plastomic counterpart was recognized when the corresponding site of the plastomic counterpart or *nupt* was identical to that of at least two other taxa. For example, a specific aligned site has “T”, “C”, “C”, “C”, and “C” in Cep-2 *nupt*, *C. wilsoniana, A. formosana, T. mairei* and *C. lanceolata*, respectively (also see the aligned position 32 in Figure 5). This site would be recognized as a nonsynonymous mutation

doi:10.6342/NTU201601257

from “C” to “T” in the Cep-2 *nupt* as the corresponding amino acid change from Alanine to Valine.

2.2.8 Plastome Mapping and Statistical Analyses

The plastome map of *T. mairei* was drawn using Circos (http://circos.ca), which is a flexible software for exploring relationships between objects or positions. It was written in Perl language. The physical map will be saved as PNG format in the input folder. In all statistical tests, including Pearson’s correlation test and Student’s t-test, Microsoft Excel 2010 was used.

2.3 Results

2.3.1 Reduction and Compaction of the Plastome of *T. mairei*

The plastomes of *A. formosana* (AP014574) and *T. mairei* (AP014575) are circular molecules with AT contents of 64.17% and 65.32%, respectively. The *T. mairei* (128,290 bp) plastome has lost five genes (*rps16, trnA-UGC, trnG-UCC, trnI-GAU, and trnS-GGA*) compared to that of *A. formosana* (136,430 bp), which leads to a relatively smaller plastome size. The coding regions occupy 61.27% of the plastome length in *A. formosana* and 64.18% in *T. mairei*. The gene density was estimated to be 0.88 and 0.90 (genes/kb) for the plastome of *A. formosana* and *T. mairei*, respectively. In addition, the other two published plastomes for Taxaceae species, *C. wilsoniana* (NC_016063) and *C. oliveri* (NC_021110), are 136,196 bp and 134,337 bp, respectively. Altogether, these data suggest that the plastome of *T. mairei* has evolved towards reduction and compaction.

Dot-plot analysis (Figure 6) reveals three genomic rearrangements between the plastomes of *A. formosana* and *T. mairei*, including a relocated fragment of ~18 kb from

doi:10.6342/NTU201601257

psbK to *trnC-GCA*, a relocated fragment of ~16 kb from *trnD-GUC* to *trnT-UGU*, and an inverse fragment of ~18 kb from *5'rps12* to *infA*. However, the two plastomes share a unique inverted repeat pair that contains *trnQ-UUG* in each copy, hereafter termed "trnQ-IR" (Figure 6).

2.3.2 Intra-species Variations in the Plastomes of *T. mairei*

To date, the plastomes of three *T. mairei* individuals (*T. mairei* voucher NN014: NC_020321, *T. mairei* voucher SNJ046: JN867590, and *T. mairei* voucher WC052: JN867591) have been published. Together with our newly sequenced plastome of *T. mairei*, these four plastomes vary slightly in size ranging from 127,665 to 128,290 bp. A neighbor-joining (NJ) tree inferred from the whole-plastomic alignment between these four individuals and *A. formosana* is shown in Figure 7. The tree topology indicates that although the plastome size of SNJ046 and WC052 is similar to that of NN014, and that the two plastomes SNJ046 and WC052 form a sister clade to that of our sampled *T. mairei*.

We also performed a pairwise genome comparison between our *T. mairei* and voucher NN014 because the latter was designated as the reference sequence (RefSeq) in NCBI GenBank. We detected 858 SNPs and 218 indels between the two plastomes. Figure 8 shows that the intergenic spacers and coding regions contained nearly equal numbers of SNPs. Most of the indels were found in the intergenic spacers and accounted the difference in plastome size between the two *T. mairei* individuals. We found 33 indels in the coding regions, but none caused frameshifts. Figure 9 illustrates the distribution of SNPs, indels, and SSRs in the plastome of our sampled *T. mairei*. Interestingly, the abundance of SSRs was positively correlated to those of SNPs (Pearson, $r = 0.52$, $p < 0.01$). However, no correlation was detected between SSRs and

doi:10.6342/NTU201601257

indels abundance (Pearson, r = 0.02, p = 0.89). In legumes, the region that contains *ycf4*, *psaI*, *accD*, and *rps16* was found to be hypermutable (Magee et al. 2010). In the plastome of *T. mairei*, three 200-bp bins that located in the sequence of *5'clpP* (position 55,001–55,200), *5'ycf1* (pos. 124,201–124,400), and the intergenic spacer between *rrn1*6 and *rrn23* (pos. 96,801–97,000) contained the highest sum of SNPs, indels, and SSRs (Figure 9). Therefore, these loci can be considered intra-species mutational hotspots in *T. mairei* and can be a potential high-resolution DNA barcodes for population genetics of *Taxus*.

2.3.3 Retrieval of Ancestral Plastome Sequences in Taxaceae

A matrix with 20 locally collinear blocks (LCBs) was generated on the basis of whole plastome alignments between the sampled three Taxaceae and four Cupressaceae species. This matrix of LCBs was then used in reconstructing ancestral plastomic organization. The most parsimonious tree with the corresponding ancestral plastomic organization is shown in Figure 4 and Figure 10, and that the three Taxaceae species form a monophyletic clade while *A. formosana* is closer to *C. wilsoniana* than to *T. mairei*. This topology is in good agreement with the recent molecular review of the conifer phylogeny by Leslie et al. (2012). Figure 4 shows the detailed evolutionary scenario of plastomic rearrangements with the intermediate ancestral plastomes in the three examined Taxaceae species. By comparing the ancestral and extant plastomes, we postulated that one, three, and two inversions might have occurred in *A. formosana*, *C. wilsoniana*, and *T. mairei*, respectively, after they had diverged from their common ancestor. Specific primer pairs were used for amplifying the corresponding ancestral fragments that differ from the extant plastomes in genomic organization (Figure 4). Five (Ame-2, Cep-2, Cep-5, Cep-6, and Tax-4) out of the ten primer pairs were able to

doi:10.6342/NTU201601257

produce amplicons totaling 16.6 kb (see Table 3 for accession numbers).

2.3.4 Characteristics of Potential *Nupt* Amplicons

The obtained PCR amplicons were sequenced and annotated (Table 3). With the exception of *chlB* of Cep-2, all putative protein-coding genes contain no premature stop codons. The coding sequence (CDS) of each amplicon was aligned with its plastomic counterparts and orthologs of other cupressophytes, *Ginkgo*, and *Cycas*. We used maximum likelihood (ML) trees inferred from concatenated CDSs to examine the origins of these PCR amplicons, with *Ginkgo* and *Cycas* as the outgroup (Figure 11). In each tree, the plastomic sequences were divided into three groups (i.e., the Cupressaceae clade, the Taxaceae clade, and the clade comprising Araucariaceae and Podocarpaceae). Notably, the placements of our PCR amplicons are incongruent among the four trees. For example, both Ame-2 and Cep-2 were clustered with their plastomic counterparts (Figure 11A). In contrast, Cep-5, Cep-6, and Tax-4 were placed remotely from their plastomic counterparts, indicating that they originated via horizontal transfer (Figure 11B, C, and D).

The ancestral plastomic organization that we used to design primers for amplification of Ame-2 and Cep-2 was rearranged by a 34-kb inversion flanked by trnQ-IRs. These trnQ-IRs were 564 and 549 bp in size for *A. formosana* and *C. wilsoniana*, respectively. IRs of similar sizes can mediate homologous recombination in the conifer plastomes (Tsumura et al., 2000; Wu et al., 2011; Yi et al., 2013; Guo et al., 2014). As a result, if the trnQ-IR-mediated isomeric plastome is present in our sampled taxa, our PCR approach shall be able to amplify isomeric plastomic fragments. Ame-2 has 100% sequence identity with its plastomic counterpart (Figure 11A) in the CDS, which strongly suggests its origin as an isomeric plastome. Cep-2 differs from its

doi:10.6342/NTU201601257

plastomic counterpart by several mutations, including two premature stop codons in *chlB*, of which one of the two cannot be replaced by neither U-to-C nor C-to-U RNA-editing (Figure 12). Therefore, the origin of Cep-2 is from a horizontal transfer rather than an isomeric plastome.

2.3.5 Evolution of *Nupt* Sequences in Taxaceae

The sequence identity between the four *nupts* and their plastomic counterparts ranges from 61.71% to 99.08% (Table 4). In fact, differences in aligned sites between *nupts* and their plastomic counterparts are derived from two types of mutations. One is the mutation in *nupts* and the other is that in plastomes. As shown in Table 4, with the exception of Tax-4, all *nupts* accumulated more mutations than their plastomic counterparts. The low sequence identity between Tax-4 and its plastome sequences (61.71% in Table 4) may be due to the unusually increased mutations in the latter. In all *nupts* except Cep-5, at least one potential protein-coding gene had the ratio of nonsynonymous (dn)/synonymous (ds) mutations > 1, which reflects the effect of relaxed functional constraints in *nupts*. Figure 13 illustrates nucleotide mutation classes in *nupts* and their corresponding plastome sequences. We excluded the plastomic counterpart of Cep-2 from the calculation because we observed only one mutation in the sequence. In all *nupts*, transitional mutations comprise over 50% of the total mutations. The mutation of G to A and its complement C to T (denoted GC-to-AT in Figure 13) had the highest frequency in both *nupts* and plastome sequences. To examine which of the mutation classes is statistically predominant, we compared the two most abundant classes of mutations. In *nupts*, the frequency was higher for GC-to-AT than AT-to-GC mutations (t-test, $p = 0.018$). However, GC-to-AT and AT-to-GC mutations did not differ in plastome sequences (t-test, $p = 0.379$), suggesting different mutational environments

doi:10.6342/NTU201601257

between *nupts* and their corresponding plastome sequences.

2.3.6 Ages of *Nupts* in Taxaceae

Molecular dating of sequences highly depends on mutation rates. Unfortunately, mutation rates in the nuclear genomes of Taxaceae species have not been directly measured. The *nupts* identified in this study were expected to evolve neutrally. The four-fold degenerated site is a useful indicator in measuring the rate of neutral evolution (Graur and Li, 2000). In nuclear genomes of conifers, the mutation rate at the four-fold degenerate sites was estimated to be 0.64×10^{-9} per site per year (Buschiazzo et al., 2012). In the *nupts* Cep-2, Cep-5, Cep-6, and Tax-4, we found 29, 117, 100, and 42 mutations among 2,961, 3,380, 2,207, and 1,466 sites, respectively (Table 4). Therefore, the ages of Cep-2, Cep-5, Cep-6, and Tax-4 were estimated to be approximately 15.3, 54.1, 70.8, and 44.8 million years (MY), respectively.

2.4 Discussion

2.4.1 Labile Plastomes of Yew Family and Their Impact on Phylogenetic Studies

The phylogenetic relationships among *Amentotaxus, Cephalotaxus,* and *Taxus* have not been resolved. Recent molecular studies placed *Amentotaxus* as sister to *Taxus* (e.g., Cheng et al. 2000; Mao et al. 2012) or to *Cephalotaxus* (e.g., Leslie et al. 2012). We found that a 34-kb inversion from *trnT* to *psbK* distinguished *A. formosana* and *C. wilsoniana* from *T. mairei* (Figure 4), which suggests that *A. formosana* is closer to *C. wilsoniana* than to *T. mairei*. However, the plastome of another *Taxus* species, *T. chinensis* (Zhang et al., 2014), cannot be distinguished from those of *A. formosana* and *C. wilsoniana* by this 34-kb inversion. Of note, this 34-kb inversion is flanked by a pair of trnQ-IR sequences. We found that the trnQ-IR sequence is commonly present in *A.*

doi:10.6342/NTU201601257

formosana (564 bp), *C. wilsoniana* (549 bp), *T. mairei* (248 bp), and *T. chinensis* (248 bp).

The presence of the trnQ-IR pair is able to generate isomeric plastomes in *C. oliveri* (Yi et al., 2013) and four *Juniperus* species (Guo et al., 2014). In Pinaceae, inverted repeats larger than 0.5 kb could trigger plastomic isomerization, and retention of an isomer was species- or population-specific (Tsumura et al., 2000; Wu et al., 2011). Indeed, Figure 11A revealed that Ame-2 was likely a PCR amplicon derived from the trnQ-IR-mediated isomeric plastome of *A. formosana*. Therefore, with the presence of an isomeric plastome, the synapomorphic character—the 34-kb inversion—in Figure 4 might be a false positive result caused by insufficient sampling. Nonetheless, our data also suggest that isomeric plastomes be cautiously treated when using genomic rearrangements in phylogenetic estimates.

Disruption of the plastomic operons is rare in seed plants (Jansen and Ruhlman, 2012). We found that the S10 operon of *T. mairei* was separated into two gene clusters (*rpl23-rps8* and *infA-rpoA*) by an 18-kb inversion (Figure 4). Because the transcriptional direction of the S10 operon is from *rpl23* to *rpoA* (Jansen and Ruhlman, 2012), the gene cluster *infA-rpoA* in *T. mairei* likely has to acquire a novel promoter sequence for transcription. Disruption of the S10 operon was previously reported in the plastome of Geraniaceae (Guisinger et al., 2011). However, the evolutionary consequence of plastomic operon disruption has never been studied. In the plastome of *T. mairei*, we detected prominently elevated mutations in the two separated gene clusters of the S10 operon as compared with their relative *nupts* (Table 4). Interestingly, two (i.e., *infA* and *rps11* in Table 4) of the three protein-coding genes on the plastomic gene-cluster *infA-rpoA* had dn/ds ratios larger than 1. Whether disruption of the S10 operon results in the positive selection of these two genes requires further investigation.

doi:10.6342/NTU201601257

2.4.2 PCR-Based Approach in Investigating *Nupts*: Pros and Cons

The immense growth of available sequenced nuclear genomes offers great opportunities for investigating nuclear organellar DNA (*norgs*). The number of *norgs* could vary depending on the use of different assembly software, versions of released genomes, and search strategies (Hazkani-Covo et al., 2010). A PCR-based approach, such as that of Rousseau-Gueutin et al. (2011) and ours, is free from this problem encountered in genome assembly. The *nupts* we amplified and reported here are a few examples of *nupts*. However, considering the huge nuclear genome of conifers which requires high cost and efforts for sequencing and assembly, our PCR-based approach provides a cost-effective way for studying the evolution of *nupts*.

Using a threshold of >70% sequence identity, Smith et al. (2011) extracted *nupts* of about 50 kb from the nuclear genome of *Arabidopsis*. The amount of *Arabidopsis nupts* decreased to approximately 17.6 kb when the threshold of sequence identity was increased to 90% (Yoshida et al., 2014). It seems that identification of possible *nupts* is largely influenced by the thresholds. Setting high thresholds might limit the exploration of *nupts* to only relatively recent transfers (Yoshida et al., 2014). Clearly, the problem of setting thresholds is absent from our PCR-based approach. In this study, sequence identity between *nupts* and their plastomic counterparts ranged from 61.71 to 99.08% (Table 4). Thus, one or three of the four presented *nupts* would not be obtained if we had considered the thresholds of Smith et al. (2011) or Yoshida et al. (2014), respectively.

Only five of our ten primer pairs worked well, and one amplified the DNA fragment of isomeric plastomes rather than *nupts*. This low success rate may be due to the unsuitable primers used in our PCR experiments. Multiple primer pairs for a specific locus may improve amplification of *nupts*, as noted by Rousseau-Gueutin et al. (2011).

doi:10.6342/NTU201601257

Plastid-to-mitochondrion DNA transfers are frequent in seed plants (Wang et al., 2007). Because the mitochondrial genome of Taxaceae is currently unavailable, the possibility that our PCR products were amplicons of mitochondrial plastid DNA could not be ruled out. The phylogenetic tree approach was previously used to examine horizontal DNA transfers (Bergthorsson et al., 2003; Rice et al., 2013), but our tree analyses in Figure 11 could not distinguish the transfer events between plastid-to-nucleus and plastid-to-mitochondrion origins. The mutation rate of nuclear genomes is higher than that of plastomes in plants (Wolf et al., 1987). All of our amplified *nupts*, except Tax-4, had more mutation sites than their plastomic counterparts (Table 4). Disruption of the S10 operon is likely associated with the elevated mutation in the plastomic counterpart of Tax-4, as mentioned above. Additionally, among our *nupts*, the AT-to-GC mutation was predominant (Figure 13). These data are similar to the findings for *nupts* in rice and *Nicotiana* (Huang et al., 2005; Rousseau-Gueutin et al., 2011), which reflects a nuclear-specific circumstance shaped by spontaneous deamination of 5-methylcytosin.

2.4.3 *Nupts* Are Molecular Footprints for Studying Plastomic Evolution

Although mutation rates are relatively low in plant organellar genomes, *norgs* can serve as “molecular fossils” for genomic rearrangements (Leister, 2005). Similarly, the Taxaceae *nupts* identified in this study retain the ancestral plastomic organization. In other words, *nupts* are footprints that are valuable in reconstructing the evolutionary history of plastomic organization and rearrangements.

Dating the age of *nupts* is critical for elucidating the evolution of *nupts*. For example, the estimated ages of Cep-2, Cep-5, and Cep-6 nupts are 15.3, 54.1, and 70.8 MY, respectively. Remarkably, these ages conflict with the scenario of plastomic rearrangements because the transfer of Cep-2 predated those of both Cep-5 and Cep-6

doi:10.6342/NTU201601257

(Figure 4). Two plastomic forms derived from trnQ-IR-mediated homologous recombination coexist in an individual of *C. oliveri* (Yi et al., 2013). This trnQ-IR is also present in the plastome of *C. wilsoniana* as previously mentioned. We suspect that in *C. wilsoniana*, the younger Cep-2 *nupt* might originate from a transferred fragment of the trnQ-IR-mediated isomeric plastome.

Most importantly, *nupts* can also help in probing RNA-editing sites and improving gene annotations. Figure 12 clearly reveals that the previously annotated *rps8* of *T. mairei* (vouchers NN014, WC052, and SNJ046) is truncated. Our newly predicted initial codon, “ACG”, locates 48 bp upstream of the previously predicted site. This “ACG” initial codon was predicted to be corrected to “AUG” via a C-to-U RNA-editing because the corresponding sequence of Tax-4 *nupt* and other conifers retain a normal initial codon of “ATG” (Figure 12). These data also imply that in *T. mairei*, the transfer of Tax-4 *nupt* predates the T-to-C mutation at the second codon position in the initial codon of *rps8*.

doi:10.6342/NTU201601257

CHAPTER 3

Birth of Four Chimeric Plastid Gene Clusters in *Sciadopitys verticillata*

3.1 Introduction

Due to the loss of many genes in early endosymbiosis, plastomes are much reduced compared to their cyanobacterial counterparts (Ku et al., 2015). To date, plastomes have invariably retained a small handful of prokaryotic features, including the organization of genes into polycistronic transcription units resembling bacterial operons (Sugiura, 1992; Wicke et al., 2011). A hallmark of seed plant plastomes is the presence of two 20- to 30-Kb IR (hereafter referred to as "typical IRs," including IR_A and IR_B), which typically contain four ribosomal RNAs. However, a few exceptions have been reported. For example, conifers—the largest gymnosperm group comprising cupressophytes and Pinaceae—have lost a typical IR copy from their plastomes (Raubeson and Jansen, 1992). Recent studies have further suggested that cupressophytes and Pinaceae might have lost different IR copies, with the former losing IR_A and the latter losing IR_B (Wu, Wang et al., 2011; Wu and Chaw, 2014).

Conifer plastomes are also characterized by extensive genomic rearrangements. The plastome of *Cryptomeria japonica*—the first completed plastome of cupressophytes (Hirao et al., 2008)—experienced at least 12 inversions after its split from the basal gymnosperm clade, cycads, whose plastomes have remained virtually unchanged for 280 million years (Wu and Chaw, 2015). The co-existence of four different plastome forms among Pinaceae genera is associated with intra-plastomic recombination mediated by three specific types of short IRs (Wu, Lin et al., 2011). Furthermore,

doi:10.6342/NTU201601257

Cephalotaxus oliveri (Cephalotaxaceae; Yi et al., 2013) and four *Juniperus* species (Cupressaceae; Guo et al., 2014) harbor isomeric plastomes that deviate from each other by an inversion possibly triggered by a trnQ-containing IR ("trnQ-IR"). Although conifer plastomes are highly rearranged (Wu and Chaw, 2014), disruptions in their operons are rare. Until recently, only one case was reported in the plastome of *Taxus mairei*, in which the S10 operon (trnI-rpoA region) was disrupted into two separate segments by a fragment of approximately 15 Kb (Hsu et al., 2014). However, the impact of such operon disruptions on plastid evolution remains poorly understood.

The 25 published cupressophyte plastomes available on GenBank (Dec 2015) represent four of the five cupressophyte families. However, no complete plastome is available for Sciadopityaceae. As part of our continuing efforts to decipher the diversity and evolution of conifer plastomes, we have completed and elucidated the plastome sequence of *Sciadopitys*. We found that the plastome of *Sciadopitys* is characterized by several unusual features. For the first time, this study reports the unusual shuffling of operons that results in the re-organization of plastid genes into new chimeric gene clusters.

3.2 Materials and Methods

3.2.1 DNA Extraction

Approximately 2 grams of fresh leaves were collected from an individual of *Sciadopitys verticillata* (voucher Chaw 1496) growing in the Floriculture Experiment Center, Taipei, Taiwan. The voucher specimen was deposited in the Herbarium of Biodiversity Research Center, Academia Sinica, Taipei (HAST). Total DNA of the leaves was extracted with 2X CTAB buffers (Stewart and Via, 1993). The extracted DNA was qualified with a threshold of DNA concentration >300 ng/μl, 260/280 = 1.8–

doi:10.6342/NTU201601257

2.0 and 260/230 > 1.7.

3.2.2 Sequencing, Plastome Assembly, and Genome Annotation

Sequencing was conducted on an Illumina MiSeq Sequencing System (Illumina, San Diego, CA) in Yourgene Bioscience (New Taipei City, Taiwan) to yield 300-bp paired-end reads of approximately 4 Gb. De novo assembly of the *Sciadopitys* plastome was performed using CLC Genomics Workbench 4.9 (CLC Bio, Arhus, Denmark). Plastid genes were predicted using DOGMA (Wyman et al., 2004) and tRNAscan-SE 1.21 (Schattner et al., 2005) with the default option that real tRNA genes should have ≥ 20 Cove scores. Boundaries of predicted genes were manually adjusted by aligning them with their orthologs of other gymnosperms. Sequences were aligned using MUSCLE (Edgar, 2004) implemented in MEGA 5.0 (Tamura et al., 2011).

3.2.3 Estimates of Dispersed Repeats and Plastomic Inversions

Repeat sequences were searched by comparing the plastome against itself using NCBI Blastn with the default settings, followed by manual deletion of overlapping or conjoined pairs. To assess the possible scenarios of plastomic inversions in *Sciadopitys*, the plastome of *Cycas taitungensis* (NC_009618) with its IRA removed was used for comparison. We identified the syntenic block of genes between *Sciadopitys* and *Cycas* using Mauve 2.3.1 (Darling et al., 2004). The resulting matrix of syntenic blocks was utilized to estimate the minimal inversion steps with MGR 2.0.1 (Bourque and Pevzner, 2002). The plastome map of *Sciadopitys* was drawn using Circos 0.67 (Krzywinski et al.,

doi:10.6342/NTU201601257

2009).

3.2.4 Detection of Isomeric Plastomes

Primer pairs listed in Table 5 were used to amplify DNA fragments specific to the two isomeric plastomes in *Sciadopitys* (i.e., *rpl33* + *rpoC2* and *rpoC1* + *rps18* for the presence of the A form; *rpl33* + *rps18* and *rpoC1* + *rpoC2* for the B form). PCR reactions were conducted with three different numbers of cycles. The conditions were 94℃ for 5 min, followed by 25, 30, or 35 cycles of 94℃ for 20 s, 55℃ for 20 s, and 72℃ for 2 min, and an extension of 72℃ for 10 min.

3.2.5 Detection of RNA Transcripts in Chimeric Gene Clusters

Total RNA was extracted from fresh leaves of *Sciadopitys* according to a modified RNA isolation protocol (Kolosova et al., 2004). We employed a RevertAid H Minus First Strand cDNA Synthesis Kit (Thermo Fisher Scientific, Waltham) to synthesize the first strand cDNA with four specific primers (SpsbNSrpoC1-3, SatpASrpl33-2, SrpoC2SpsbB-1, and Srps18 in Table 5). PCR reactions were conducted with the synthesized cDNA and four pairs of specific primers (atpF-1 + psbT for a 358-bp fragment; psbB-2 + atpA-2 for a 565-bp fragment; rpl33 + rpoC2 for a 687-bp fragment; rpoC1 + rps18 for a 939-bp fragment). The PCR conditions were 94℃ for 5 min, followed by 30 cycles at 94℃ for 20 s, 60℃ for 20 s, and 72℃ for 1 min, and an extension at 72℃ for 10 min.

3.3 Results and Discussion

3.3.1 Loss of IR_A from *S. verticillata* Plastome

The plastome of *Sciadopitys verticillata* (AP017299) is illustrated as a circular

doi:10.6342/NTU201601257

molecule with size of 138,309 bp (Figure 14). It is the largest among the known plastomes of Cupressales (including Sciadopityaceae, Taxaceae *s. l.*, and Cupressaceae *s. l.*). Flanking and adjacent genes of the typical IRs are informative markers for inferring the intact (or retained) IR copy in conifer plastomes. For example, the boundary of IRA or IRB is adjacent to the *psbA* or S10 operon (i.e., *trnI-rpoA* region), respectively (Wu, Wang et al., 2011). In the *Sciadopitys* plastome, the retained typical IR copy, which encompasses the region from *trnN-GUU* to *ycf2*, is adjacent to the S10 operon (Figure 14), indicating that it should be IRB. In other words, the lost IR copy is IRA. This observation reinforces the hypothesis that cupressophytes have lost IRA rather than IRB (Wu, Wang et al., 2011; Wu and Chaw, 2014).

The plastome of *Sciadopitys* contains a total of 121 genes, 83 of which are protein-coding genes and the rest are structural RNA genes (Table 6). Sixteen genes contain introns, but the intron of *rpoC1* has been lost. Remarkably, each of the three genes, *rrn5*, *trnI-CAU*, and *trnQ-UUG*, has two copies. The duplicated *rrn5* is located in the region between *psbN* and *psbT* (Figure 14). Duplicated *rrn5* has previously reported in the plastomes of *Agathis dammara* and *Wollemia nobilis* (Araucariaceae), but it was located in between *psbB* and *clpP* (Yap et al., 2015). Among the elucidated cupressophyte plastomes available in GenBank, only these three taxa contain two copies of plastid *rrn5*. However, since the locations of their extra *rrn5* differ among these two cupressophytes families, it is most parsimonious that duplications of *rrn5* occurred independently. Our data also show that *accD* was lost in the plastome of *Sciadopitys*. Li et al. (2016) had previously reported that the loss of *accD* occurred after the split of Sciadopityaceae from other cupressophytes and that the *accD* has been functionally transferred from plastid to nucleus.

doi:10.6342/NTU201601257

3.3.2 Pseudogenization of Four *tRNA* Genes after Tandem Duplications

Notably, four pseudo-tRNA genes (*ΨtrnV-GAC*, *ΨtrnQ-UUG*, and two copies of *ΨtrnP-GGG*) were detected in the *Sciadopitys* plastome (Figure 14). Both *ΨtrnV-GAC* and *ΨtrnQ-UUG* are close to their functional paralogs, implying that pseudogenization of these two genes might have occurred after tandem duplications. The plastid *trnP-GGG* of angiosperms likely has been lost for 150 MY (Chaw et al., 2004). In contrast, this tRNA gene is retained and commonly located in the region between *trnL* and *rpl32* in *Cycas*, *Ginkgo*, *Gnetum, Pinus* (Wu et al., 2007), and other cupressophyte families, such as Araucariaceae (Yap et al., 2015), Podocarpaceae (Vieira Ldo et al., 2014; Wu and Chaw, 2014), and Taxaceae *s. l.* (Yi et al., 2013; Hsu et al., 2014). In the *Sciadopitys* plastome, two *ΨtrnP-GGG* copies are separated by a distance of approximately 20-Kb; one is adjacent to *trnL-UAG* and the other is located near *rpl32* (Figure 14). Therefore, the two *ΨtrnP-GGG* copies might have resulted from tandem duplications, followed by subsequent 20-Kb plastomic inversion.

3.3.3 Evolution of Plastid *trnI-CAU* Genes in *S. verticillata*

Sciadopitys has two copies of plastid *trnI-CAU*: one located in between *trnC-GCA* and *psbA*, while the other is between *ycf2* and *rpl23* (Figure 14 and Figure 15). In *Cryptomeria*, one of the two plastid *trnI-CAU* copies was considered residual from the lost typical IR (Hirao et al., 2008). Indeed, the majority of cupressophyte plastomes have two *trnI-CAU* copies with sequence identity higher than 85% (Table 7), connoting their homologous origin.

In *Sciadopitys* plastome, both *trnI-CAU* are capable to fold into cloverleaf structures, but they differ in prediction scores. The copy that is located in between *ycf2* and *rpl23* has a score of 78.1 bits (Figure 15A), much higher than the score of the other

doi:10.6342/NTU201601257

copy (score = 48.5 bits; Figure 15B). Ten nucleotide substitutions (highlighted in gray in Figure 15) were detected among the two *trnI-CAU* copies, including three mismatches and four U•G abnormal pairings in the stems of the low-scoring *trnI-CAU* (Figure 15B). Although *trnI-CAU* is essential for plastid biology (Alkatib et al., 2012), the presence of two copies of *trnI-CAU* in *Sciadopitys* plastome remain to be investigated. Interestingly, the elucidated cupressophyte plastomes, such as *Cephalotaxus*, *Nageia*, and *Podocarpus*, contain only one copy of *trnI-CAU* (Table 7). Therefore, whether the low-scoring *trnI-CAU* of *Sciadopitys* is functionally redundant and subjected to relaxed structural constraint is worthy of further investigation.

3.3.4 Presence of Two Isomeric Plastomes in *S. verticillata*

Recent studies of conifer plastomes revealed that dispersed short IRs can trigger plastomic rearrangements to generate isomeric forms. In Pinaceae, a shift between different plastomic forms is often associated with homologous recombination (HR) mediated by the short IRs of approximately 949 bp (Tsumura et al., 2000; Wu, Lin et al., 2011). The short IRs that contain *trnQ-UUG* (*trnQ*-IR) can also promote the formation of isomeric plastomes in *Cephalotaxus* (Yi et al., 2013) and *Juniperus* (Guo et al., 2014).

Thirty-seven pairs of dispersed repeats were detected in the *Sciadopitys* plastome. Among them, the longest IR pair is 370 bp and contains the sequences of *3'rpoC1* and *5'rpoC2* (Figure 14). In the *Sciadopitys* plastome, only this 370-bp IR pair is longer than the 250-bp "*trnQ*-IR" of *Juniperus* (Guo et al., 2014). Hence, if the 370-bp IR were able to mediate HR in *Sciadopitys*, we would expect the presence of two plastomic forms, as depicted in Figure 16. We designate the plastomic form illustrated in Figure 14 as the A form, while the other is the B form. We have verified the presence of both

doi:10.6342/NTU201601257

the A and B forms by amplicons of four specific DNA fragments across the 370-bp IR in the PCR with 35 cycles (Figure 16). However, the amount of the PCR amplicons differs between the two forms. With 25 PCR cycles, only the two specific amplicons of the A form are evident, whereas those of the B form are undetectable (Figure 16). These results suggest that A form is the predominant form of *Sciadopitys* plastome populations, in agreement with our assembly results.

The plastomes of Cupressaceae and Taxaceae possess two copies of *trnQ*-IRs (Guo et al., 2014). Nonetheless, this IR is absent from the plastome of *Sciadopitys*. The plastomes of Araucariaceae (Yap et al., 2015) have an IR pair that is approximately 600-bp long and contains the gene *rrn5*. Such the length of IRs could potentially trigger HR. However, the presence of associated isomeric plastomes has not been experimentally demonstrated in Araucariaceae. Including the unique 370-bp IR of *Sciadopitys*, it is apparent that in cupressophytes, the presence of isomeric plastomes is overwhelming and associated with diverse short IRs.

3.3.5 Birth of Four Chimeric Gene Clusters

We identified a total of 16 syntenic blocks between *Cycas* and *Sciadopitys* plastomes (Figure 14; Figure 17). In addition to the loss of IRA from the *Sciadopitys* plastome mentioned above, eight plastomic inversions were detected to distinguish *Sciadopitys* from *Cycas* (Figure 17). Since *Cycas* was proposed to retain the ancestral gene order of seed plant plastomes (Jansen and Ruhlman, 2012), these eight inversions should have occurred after cupressophytes split from cycads.

In *Sciadopitys*, plastomic inversions have also disrupted four typical operons that are generally conserved among seed plants. These disrupted operons are *rps2–atpI–atpH–atpF–atpA* (hereafter, *rps2* operon), *psbB–psbT–psbH–petB–petD* (*psbB* operon),

doi:10.6342/NTU201601257

rpoB–proC1–rpoC2 (*rpoB* operon), and *petL–petG–psaJ–rpl33–rps18* (*petL* operon) (Figure 18A & B). Recombination between *rps2* and *psbB* operons is associated with inversion 8 (Figure 17), creating the *rps2–petD* and *psbB–atpA* gene clusters (Figure 18A). On the other hand, inversion 4 (Figure 17) recombined the *rpoB* and *petL* operons and then generated the *petL–rpoC2* and *rpoB–rps18* gene clusters (Figure 18B). Most genes in each of the four chimeric gene clusters have the same transcriptional direction (Figure 18A & B). Therefore, we postulated that genes in these chimeric gene clusters might be co-transcribed. We performed RT-PCR assays with specific primers designed from genes near the junction between different operon-derived segments to verify this proposition.

As shown in Figure 18C, our RT-PCR results indicate that (1) there was no DNA contamination in the assayed RNA because all of the negative controls failed to yield any signal; and (2) the expected size of amplicons was clearly detected in all experimental sets. These data suggest that shuffling between different operons could lead to the birth of new co-transcription units in plastids.

Disruptions of conserved plastid operons have been only reported in a few taxa, such as *Vigna* (Perry et al., 2002), *Trifolium* (Cai et al., 2008), *Trachelium* (Haberle et al., 2008), some genera of Geraniaceae (Guisinger et al., 2011), *Taxus* (Hsu et al., 2014), and *Sciadopitys* (this study). Of note, these taxa also have highly rearranged plastomes. Except for *Vigna* and *Sciadopitys*, none of the above taxa has experienced recombination between operons. Instead, their disrupted operons are separated rather than combined. In the *Vigna* plastome, recombination between two homologous operons; S10A and S10B, has led to the re-organization of genes in the operons (Perry et al., 2002). Nonetheless, novel chimeric gene clusters created by shuffling between heterologous operons (Figure 18) are documented for the first time in the present study.

doi:10.6342/NTU201601257

3.3.6 Evolutionary Effects of Novel Chimeric Gene Clusters

The chimeric gene clusters of *Sciadopitys* provide two novel insights into the evolution of plastomes. First, other than the gene cluster *rpoB–rpoC1–rps18*, the remaining three chimeric gene clusters do not alter their upstream regions, as the neighboring genes of their 5' regions are the same as those of *Cycas* (Figure 18A & B). This finding suggests that the promoter sequences of these gene clusters have not been altered after the associated inversions taken place. Figure 18D shows that the upstream sequence of *rpoB* harbors a YRTA motif of the nuclear-encoded RNA polymerase (NEP) promoters (Shiina et al. 2005). Furthermore, genes of different origins are able to be co-transcribed in the chimeric gene cluster (Fig. 18C). Therefore, we cannot rule out the possibility that the pre-existing promoters are adopted for transcription of the genes in these chimeric gene clusters.

Second, shuffling between *rpoB* and *petL* operons (Figure 18B) has relocated *rpoC2* to join the segment of the 5'*petL* operon whose transcription are associated with the plastid RNA polymerases (PEP) promoter (Finster et al., 2013). *RpoC2* codes for one of the core units of PEP (Hu and Bogorad, 1990). If the chimeric gene cluster *petL–petG–psaJ–rpl33–rpoC2* is exclusively transcribed by PEP, we would not expect any transcript of this gene cluster in *Sciadopitys*. Nonetheless, its associated transcript was observed in Figure 18C. Two possibilities might account for the presence of this transcript. First, the isomeric plastome of the B form (Figure 16) that contains an intact *rpoB* operon provides RPOC2 proteins. Second, an alternative promoter has evolved to perform transcription because many plastid genes are transcribed by both the NEP and PEP promoters (Börner et al., 2015). Third, this transcript may correspond to a read-through transcript which allows the expressions of downstream of 3' untranslated regions (Quesada-Vargas et al., 2005).

doi:10.6342/NTU201601257

Notably, functionally unrelated genes have been joined together in these chimeric gene clusters. For example, the four photosynthetic genes (*psbT*, *psbH*, *petB*, and *petD*) of the *psbB* operon have been relocated and joined with the segment of the 5'*rps2* operon whose promoter is of the NEP type (Kapoor and Sugiura, 1999). This suggests that the four photosynthetic genes are not transcribed by PEP, in disagreement with a partition of labor in which PEP transcribes genes associated with photosynthesis and NEP transcribes housekeeping genes (Hajdukiewicz et al., 1997). In contrast, this finding agrees with Liere and Börner (2007) that most of plastid genes can be transcribed by either NEP or PEP.

doi:10.6342/NTU201601257

CHAPTER 4

Conclusions

Plastomes of cupressophytes are highly variable in their size, genome organization, and gene content. These features provide a unique opportunity to study their evolution. In this study, we sequenced three complete plastomes, *A. formosana*, *T. mairei*, and *S. verticillata*. By comparing the three newly sequenced plastomes with published cupressophyte plastomes, we are able to obtain new insights into the plastomic evolution of cupressophytes.

We have shown that plastomic rearrangement events provide useful information for amplifying *nupts* in Chapter 2. Because it is difficult to avoid the amplification of isomeric plastomic or mitochondrial DNA, examining the origins of PCR amplicons was a prerequisite in this proposed PCR-based study. In angiosperms such as *Nicotiana*, *nupts* were experimentally demonstrated to be eliminated quickly from the nuclear genome (Sheppard and Timmis, 2009). However, we show that the oldest conifer *nupt* has been retained for at least 70.8 MY (i.e., since the late Cretaceous period). With an increase of available plastomes in conifers, comparative genomic analyses are expected to reveal more plastomic rearrangements. Using our approach, we are beginning to understand the evolution of *nupts* in diverse conifer species without the need to sequence and assemble their huge nuclear genomes.

In Chapter 3, we have shown that plastomic rearrangement events in *Sciadopitys* provide a unique opportunity to understand the evolutionary impact of plastomic rearrangements. The plastome of *Sciadopitys* is characterized by several unusual features, such as the loss of the typical IRA copy, the duplication and pseudogenization

doi:10.6342/NTU201601257

of four tRNAs, extensive genomic inversions, the presence of isomeric plastomes, and chimeric gene clusters derived from shuffling of remote operons. All these characteristics highlight the fact that the evolution of plastomes may be more complex than previously thought. The highly rearranged plastome of *Scidaopitys* advances our understanding of the dynamics, complexity, and evolution of plastomes in conifers.

doi:10.6342/NTU201601257

CHAPTER 5

Future Prospectives

In the first project, we proposed a PCR-based strategy to identify *nupts*. Although DNA transfer from plastomes to the nuclear genome is highly frequent, it is very rare to observe a functional organelle gene in the nuclear genome (Lloyd and Timmis, 2011). Most *nupts* are quickly deleted, decays, or alternatively scrapped during the plant evolution (Lloyd and Timmis, 2011). A *nupt* must acquire additional genetic elements if it is functional and can be retained in its new environment. The functional *nupts* should include at least three features. First, their DNA sequences should contain an intact open reading frame and could be transcribed correctly. Second, they should acquire a specific nuclear promoter and this promoter can regulate the *nupt* transcription during the plant development. Third, they should obtain a transit peptide of plastome-target to import plastome-specific proteins back to plastomes. It would be of interest to study if these newly identified *nupts* in gymnosperms have developed any novel functions.

Our study presented new insights into the plastome arrangements and intracellular gene transfer in non-model systems. After the completion of *Sciadopitys* plastome, at least one representative species for each gymnosperm families have been published. These plastomes provide an opportunity to systematically examine the plastid DNA evolution and to model plastomic orientation changes in gymnosperms. With the advent of new sequencing and bioinformatic technologies, such large scale systematic plastomic studies would be possible in near future, enabling a new era of the comparative genomics of organellar evolution.

doi:10.6342/NTU201601257

FIGURES

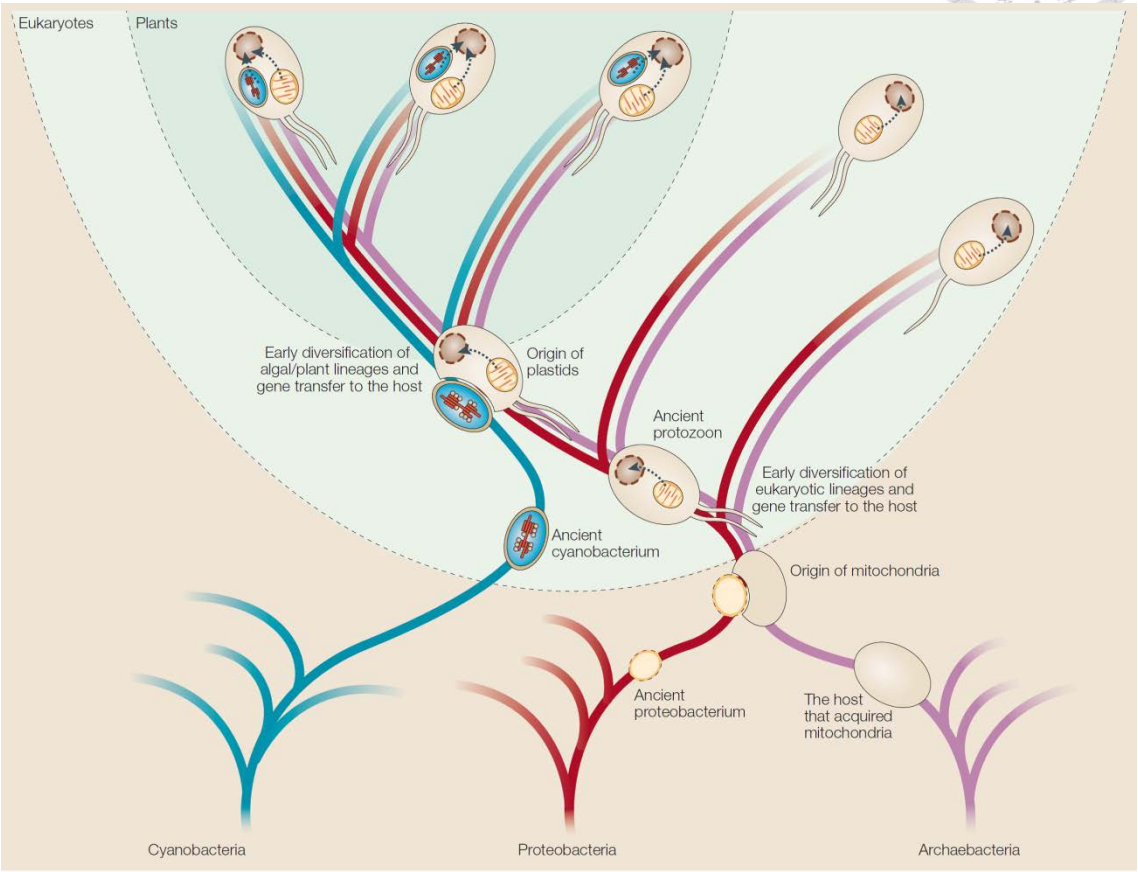

Figure 1

The phylogenetic tree of endosymbiotic evolution. (Image adapted from Timmis et al., 2004)

doi:10.6342/NTU201601257

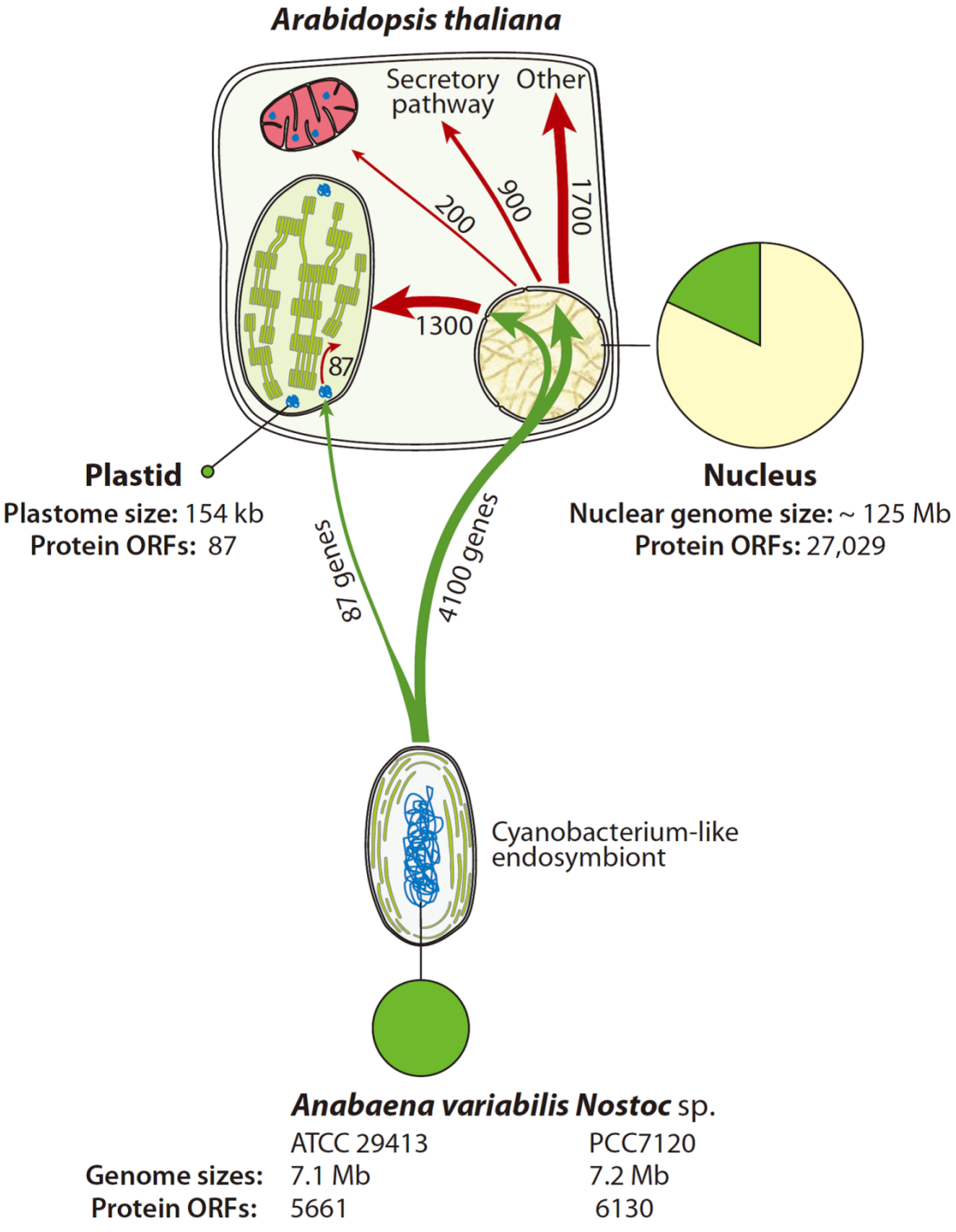

Figure 2

Fate of cyanobacterial genes and the intracellular targeting of their products in the flowering plant *Arabidopsis thaliana*. (Image adapted from Kleine et al., 2009)

doi:10.6342/NTU201601257

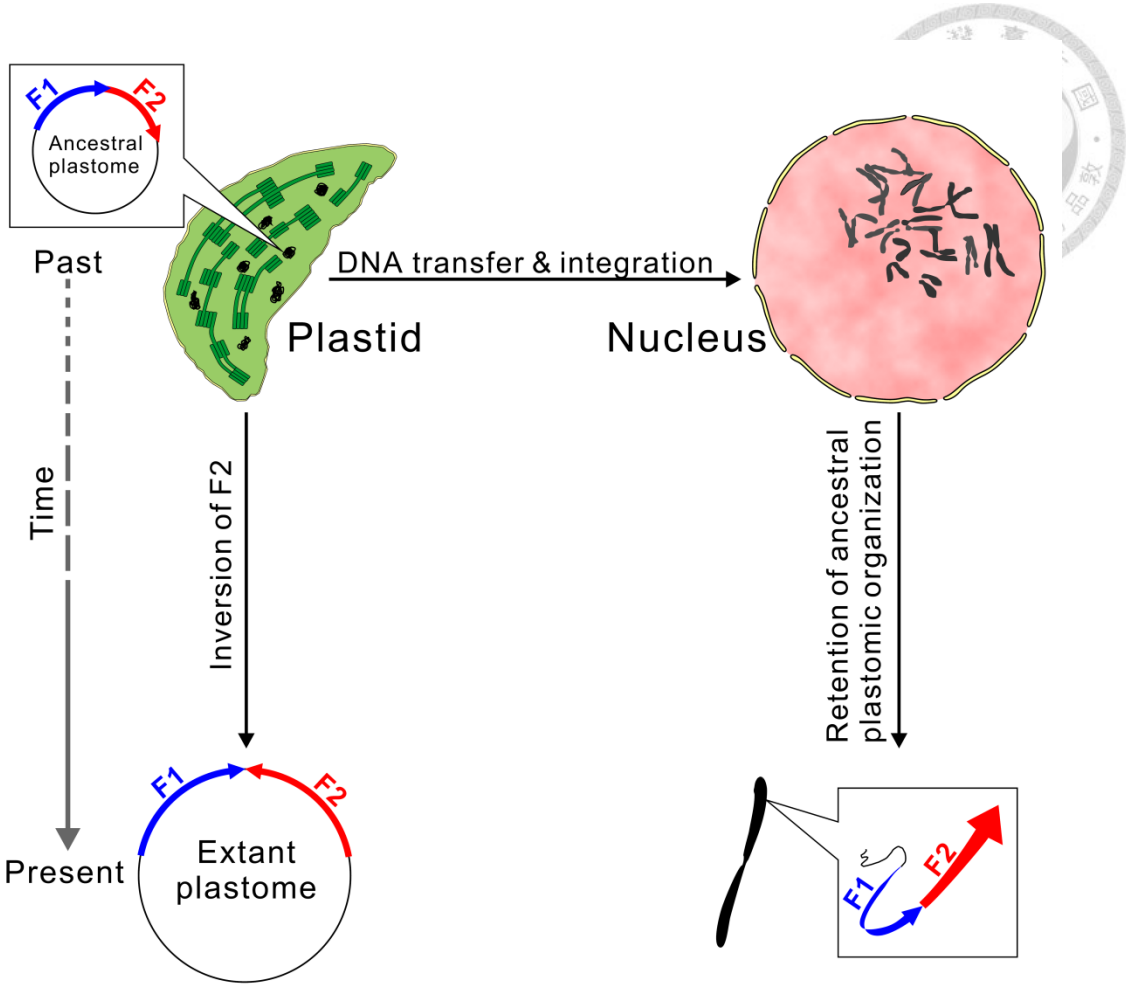

Figure 3

A schematic explanation for the amplification of ancestral plastomic DNAs transferred from plastids to the nucleus. Top left: an ancestral plastomic fragment that includes F1 and F2 sub-fragments with a head-to-tail arrangement was transferred to the nucleus (top right) in the past. After this transfer, an inversion of F2 occurred, which resulted in a head-to-head arrangement of F1 and F2 in the extant plastome. Primers based on distinctive arrangements between ancestral and extant plastomes can facilitate specific amplification of transferred ancestral plastomic fragments and avoid contaminants from amplification of the extant plastome.

doi:10.6342/NTU201601257

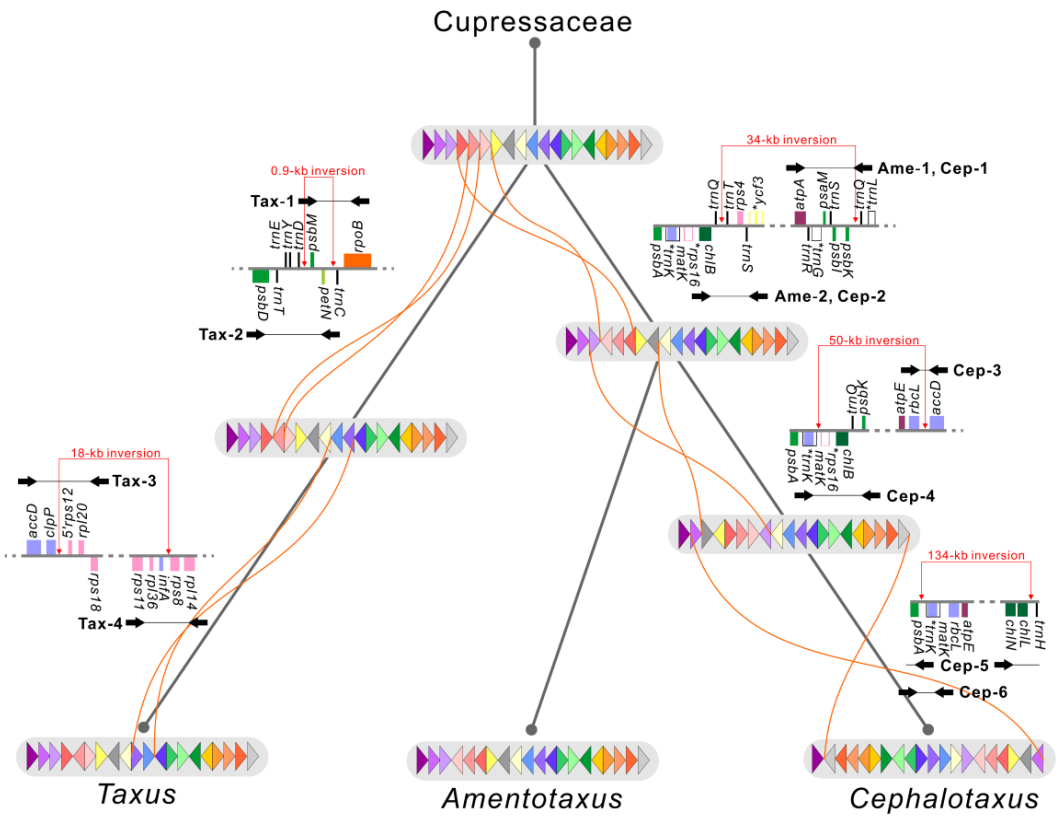

Figure 4

Hypothetical evolutionary scenarios for plastomic rearrangements in Taxaceae. Plastomes are circular but here are shown in grey horizontal bars (beginning at *psbA*) for pairwise comparisons. Color triangles within the grey horizontal bars denote locally collinear blocks with their relative orientations. Grey bars from top to bottom indicate the corresponding plastomes in the common ancestor of Taxaceae, intermediate ancestors, and extant representative species. Inversions between two plastomes are linked by orange curved lines. Ancestral gene orders before the occurrence of specific inversions are shown along tree branches. Primer pairs (black arrows) for amplification of the corresponding ancestral fragments are labeled: Tax-1 to 4 for *Taxus mairei*, Ame-1 to 2 for *Amentotaxus formosana*, and Cep-1 to 5 for *Cephalotaxus wilsoniana* (see Table S1 for primer sequences).

doi:10.6342/NTU201601257

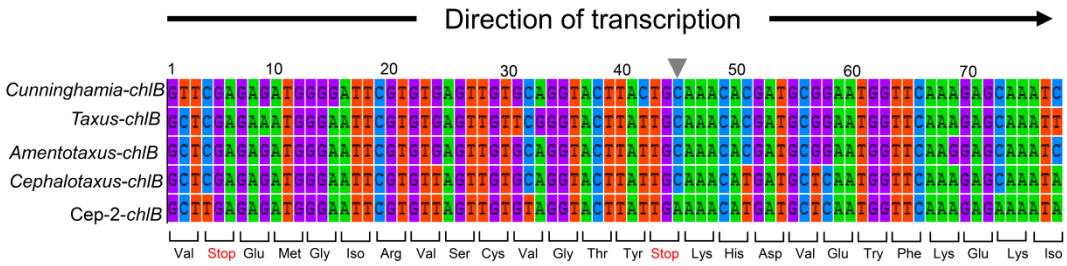

Figure 5

An alignment revealing two premature stop codons in the *chlB* sequence of Cep-2. The aligned position marked with a grey triangle indicates a point mutation in Cep-2, which results in a stop codon and cannot be replaced by RNA-editing. Positions of aligned sites are labeled with numbers. The translated amino acid for each codon of Cep-2 sequence is indicated.

doi:10.6342/NTU201601257

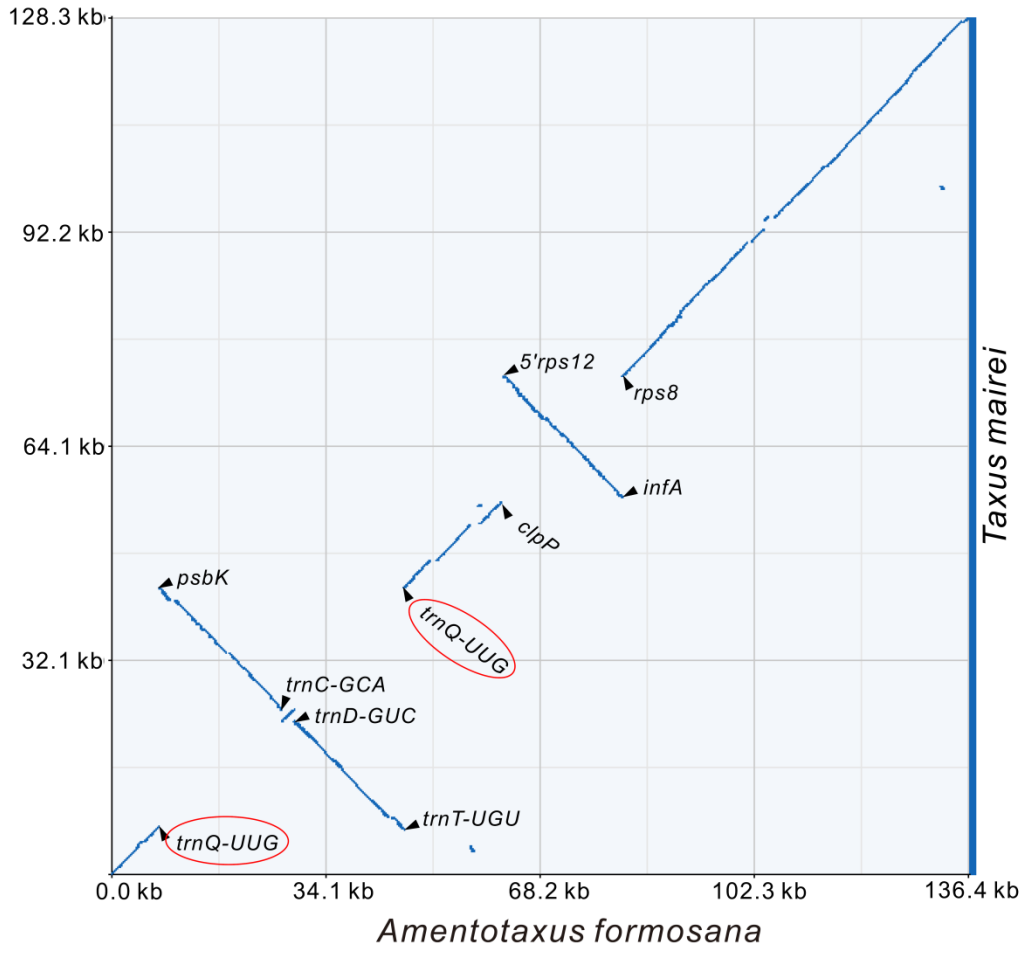

Figure 6

A dot-plot comparison of the plastomes of *Amentotaxus formosana* and *Taxus mairei*. Three relocations are revealed with their flanking genes. Transfer RNA genes, *trnQ-UUG*, inside the *trnQ*-IRs are highlighted in red circles.

doi:10.6342/NTU201601257

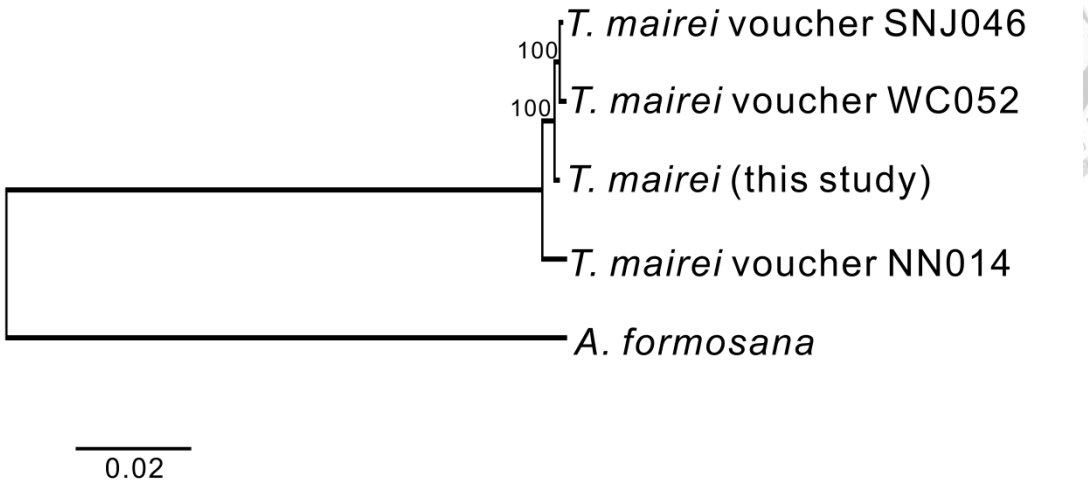

Figure 7

A neighbor-joining tree inferred from a whole-plastome alignment showing the relative relationship between *T. mairei* in this study and the other three published ones. *A. formosana* was used as the outgroup. Support values estimated from 1,000 bootstrapping analyses are indicated.

doi:10.6342/NTU201601257

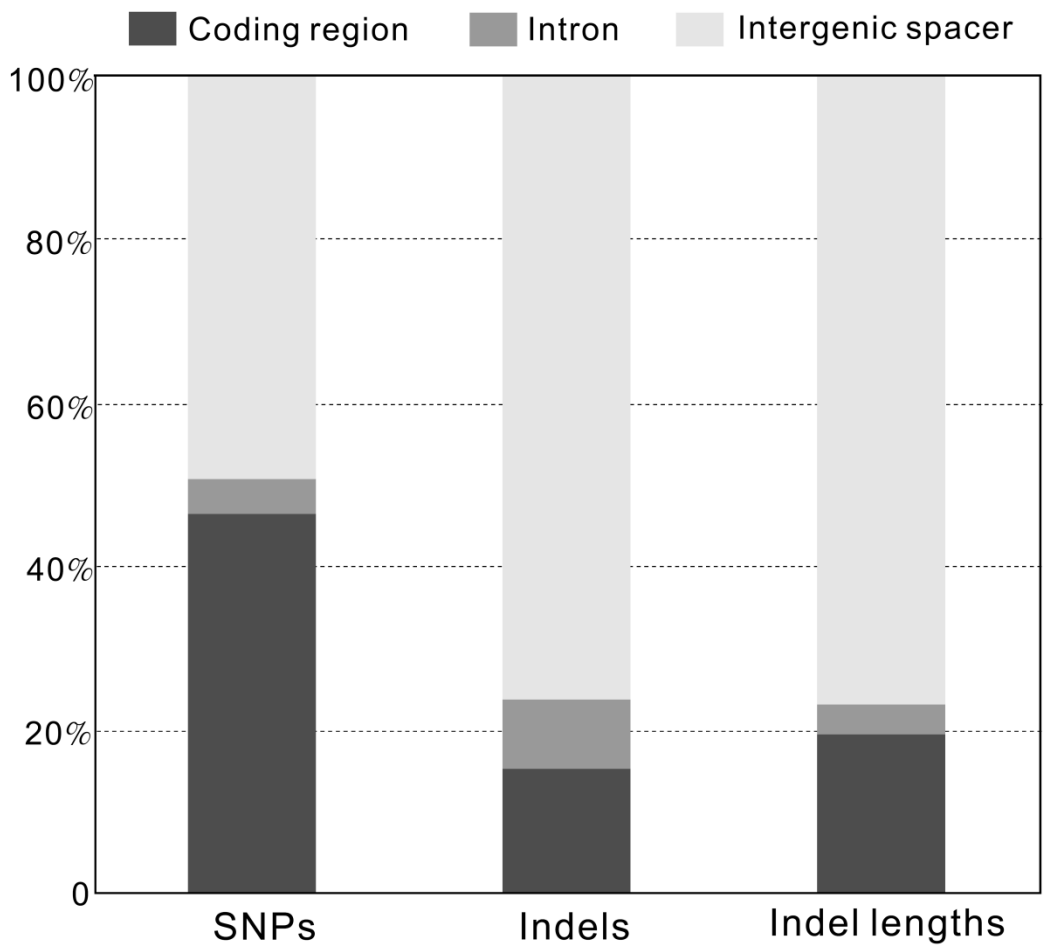

Figure 8

Stacked histogram for single-nucleotide polymorphisms (SNPs), indels, and indel lengths of the *T. mairei* plastome (AP014575) showing their relative proportions in coding, intronic, and intergenic regions.

doi:10.6342/NTU201601257

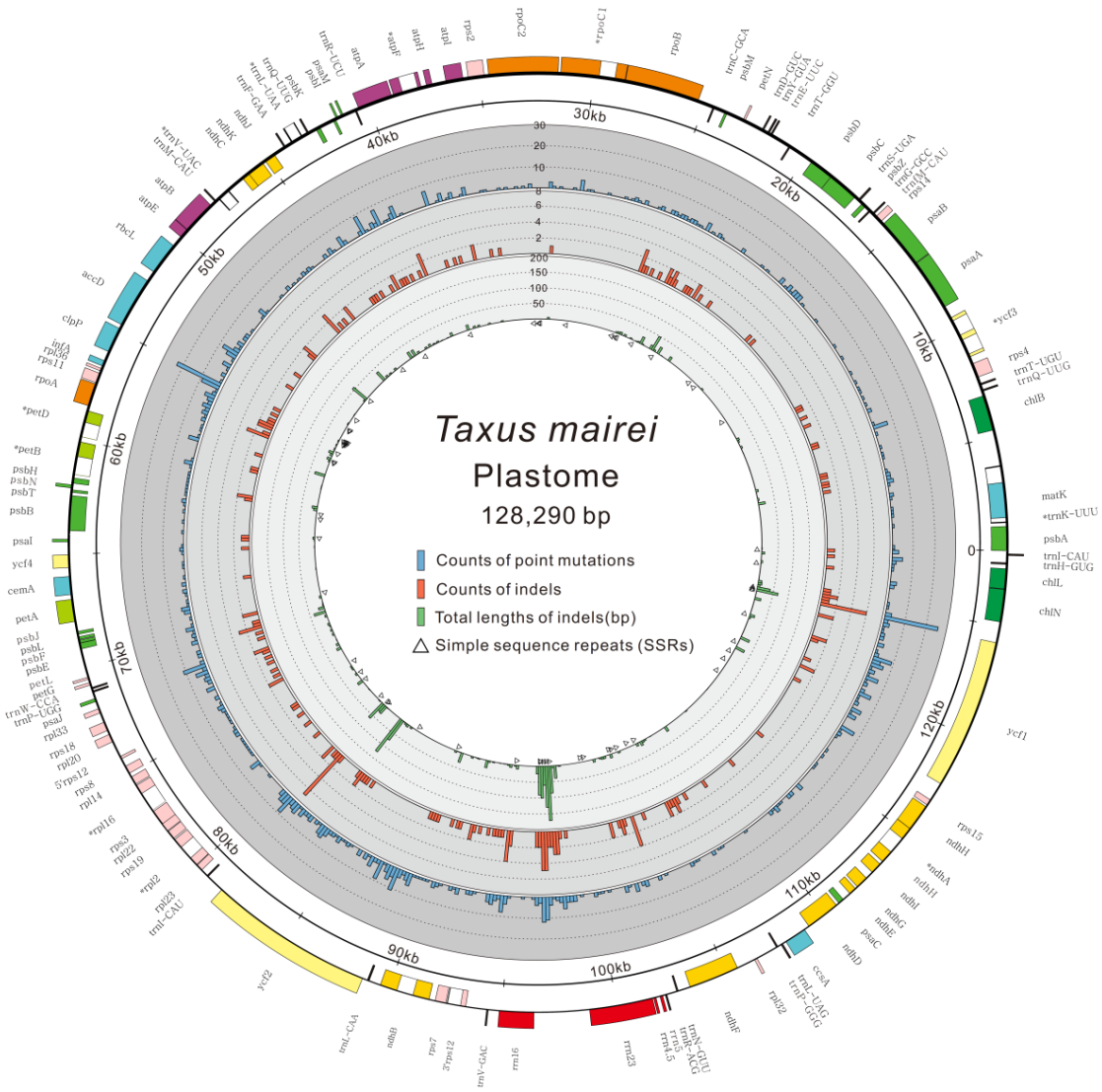

Figure 9

Distribution of single-nucleotide polymorphisms (SNPs), indels, and simple sequence repeats (SSRs) in the plastomes of *Taxus mairei*. The outermost circle is the plastome map of *T. mairei* (AP014575) with genes that are transcribed counter-clockwise (outer boxes) and clockwise (inner boxes), respectively. The immediately next circle denotes a scale of 5-kb units beginning at *psbA* gene (the 3 o'clock position). In the grey zone, three histograms from outer to inner are 1) counts of SNPs, 2) counts of indels, and 3) total indel lengths within non-overlapping 200-bp bins across the entire plastome. Triangles mark locations of SSRs.

doi:10.6342/NTU201601257

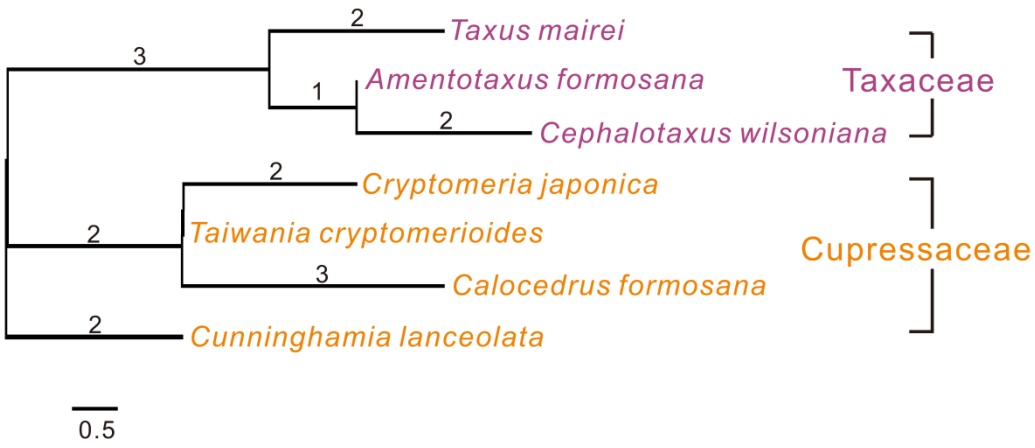

Figure 10

An unrooted tree inferred from the locally collinear block matrix generated from comparative plastomes among three Taxaceae and four Cupressaceae species. Values along branches denote numbers of rearrangements required for specific taxa or clades.

doi:10.6342/NTU201601257

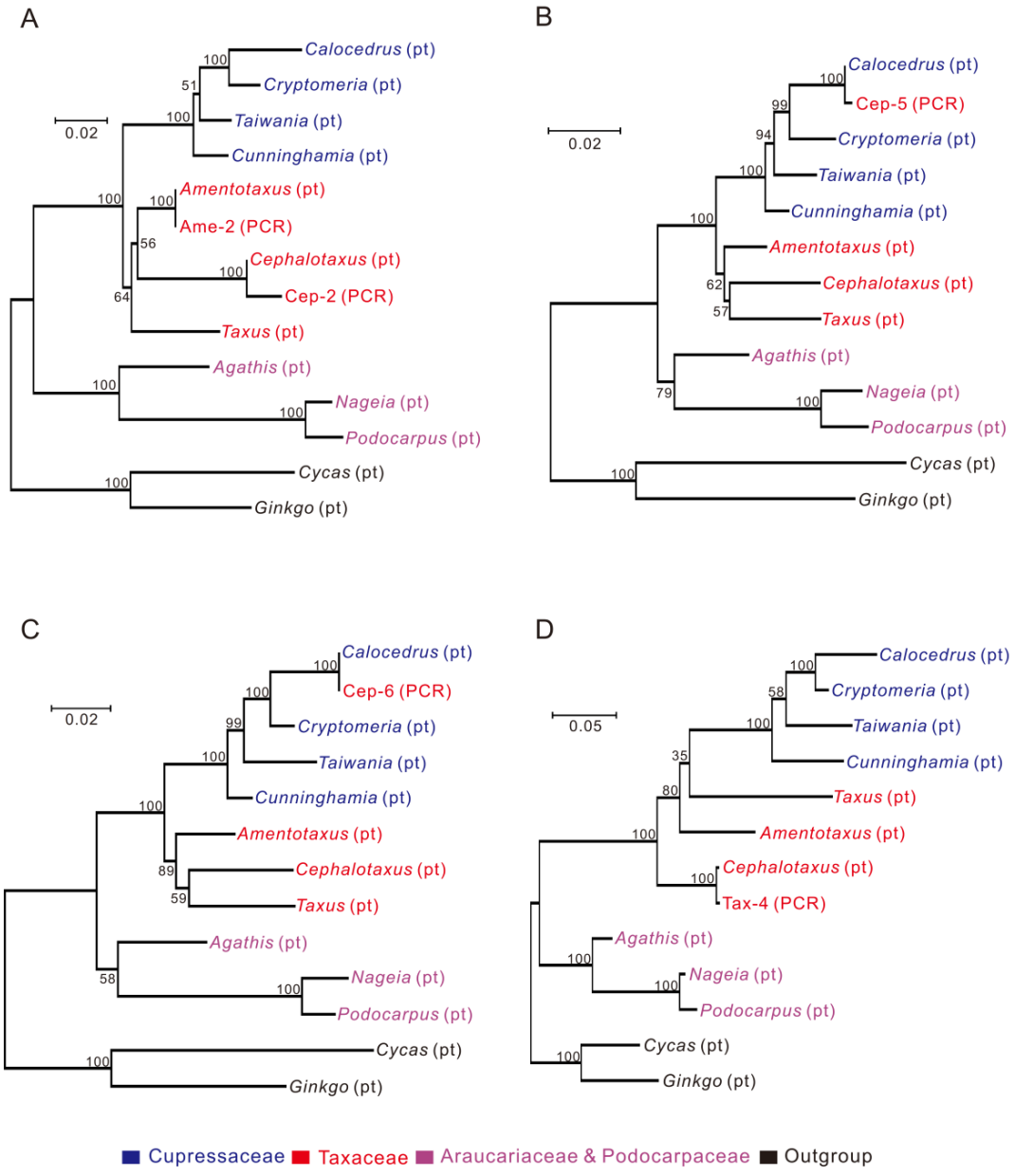

Figure 11

Origin of the obtained PCR amplicons examined by maximum-likelihood phylogenetic analyses. PCR amplicons are labeled “PCR”, and their plastomic counterparts and orthologs of other gymnosperms are labeled “pt”. Taxa of the same conifer family are in the same color. *Cycas* and *Ginkgo* together are the outgroup. Bootstrapping values assessed with 1,000 replicates are shown along branches.

doi:10.6342/NTU201601257

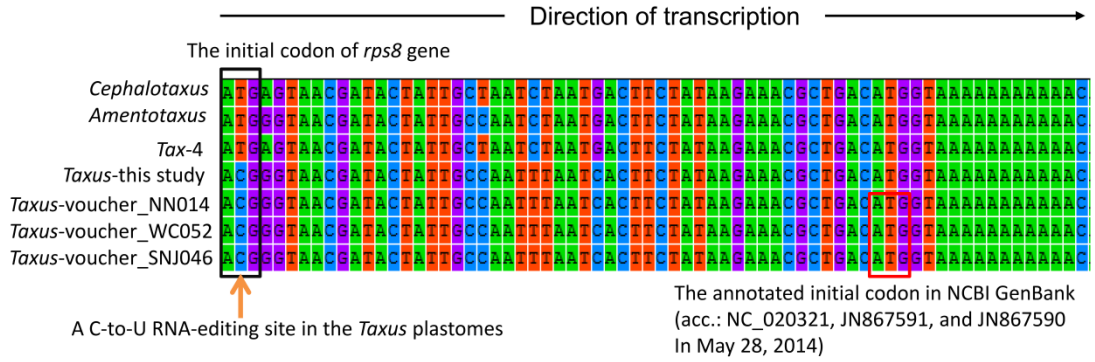

Figure 12

Alignment of seven *rps8* sequences. The left orange arrow highlights a specific C-to-U RNA-editing site at the second codon position of the initial codon in the four sampled *T. mairei* plastomes. A normal initial codon, ATG (black rectangle), was common among *Cephalotaxus*, *Amentotaxus*, and Tax-4. These data imply the creation of the "ACT" RNA-editing site after the transfer of Tax-4. The red rectangle denotes the initial codon annotated in the sequences from NCBI GenBank.

doi:10.6342/NTU201601257

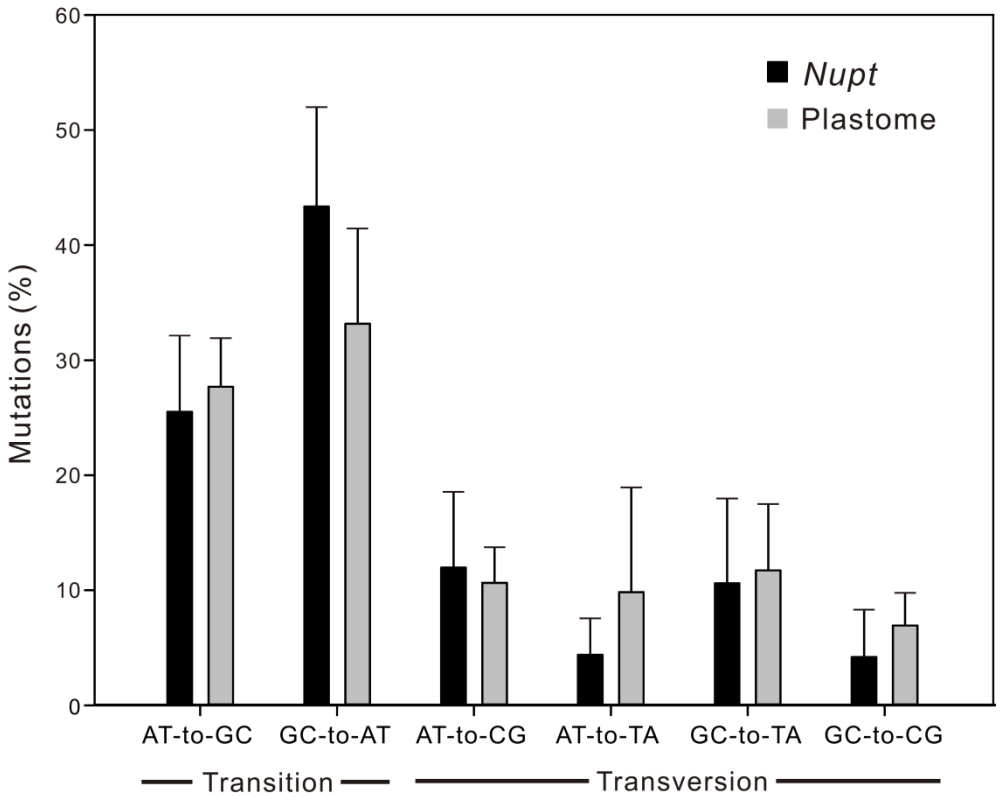

Figure 13

Percentage of nucleotide mutation classes in *nupts* and their plastomic counterparts. Types of mutations are divided into six classes. For example, the class AT-to-GC denotes the pooled percentage of the A-to-G mutations and its complement T-to-C. Data are mean ± SD.

doi:10.6342/NTU201601257

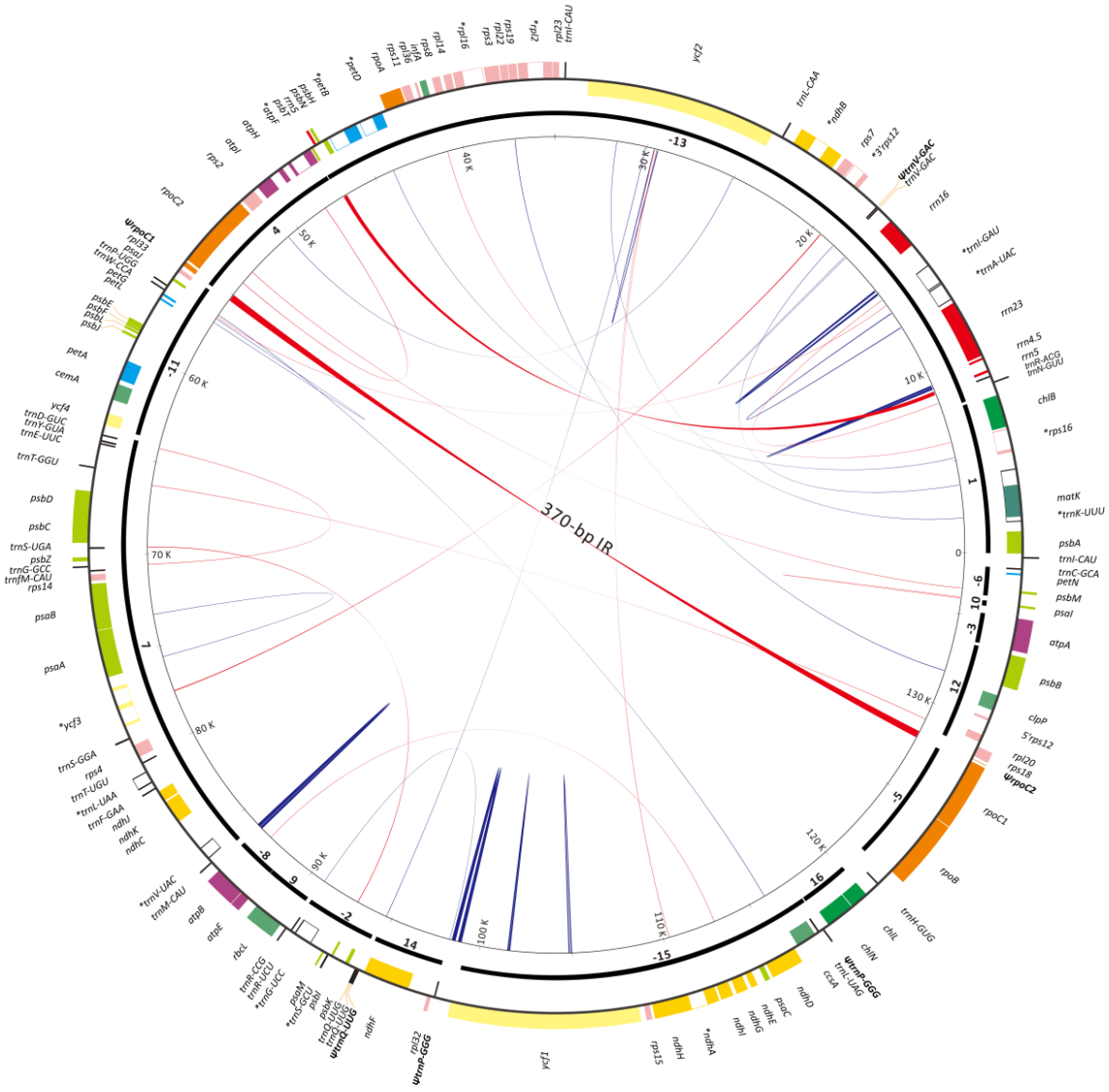

Figure 14

Plastome map of *Sciadopitys verticillata*. Colored boxes represent genes with counterclockwise (outer boxes) and clockwise (inner boxes) transcriptional directions. Syntenic blocks of genes between *Cycas* and *Sciadopitys* are depicted by thick black bars with Arabic numerals, where pluses or minuses indicate the corresponding syntenic blocks with the same or opposite directions between the two species, respectively. Pairs of dispersed repeats are connected by blue (direct repeats) or red (inverted repeats) lines, with their width proportional to the repeat size. Pseudogenes are bold and marked with a "Ψ." Intron-containing genes are indicated with an "*."

doi:10.6342/NTU201601257

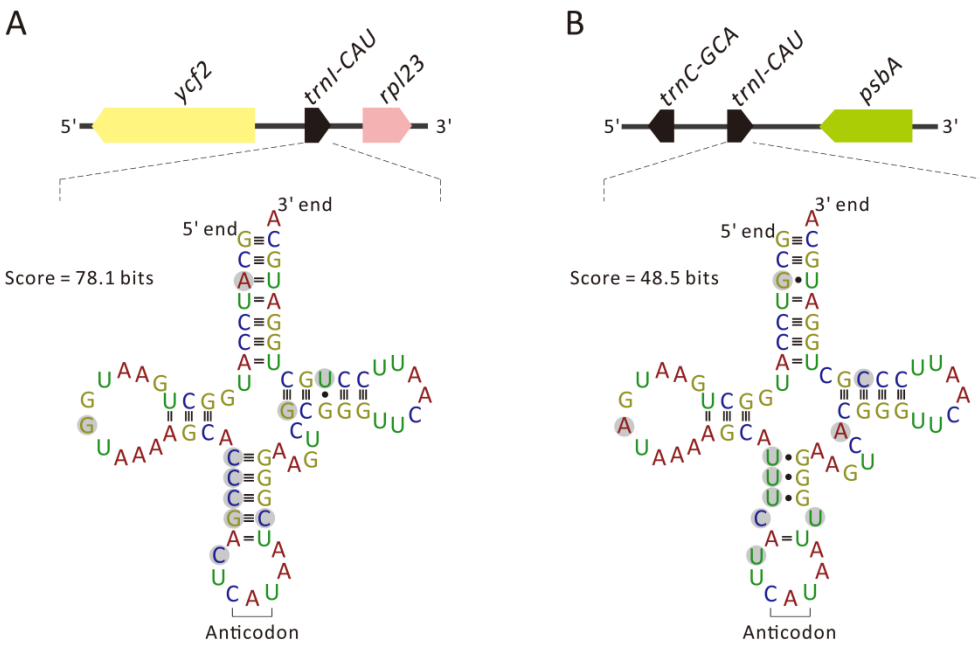

Figure 15

Comparison between the two copies of *trnI-CAU* genes in the *Sciadopitys* plastome. (A) Predicted cloverleaf structure of *trnI-CAU* located between *ycf2* and *rpl23* and (B) that of the other gene located between *trnC-GCA* and *psbA*. Pairwise substitutions of nucleotides between the two copies are highlighted in gray.

doi:10.6342/NTU201601257

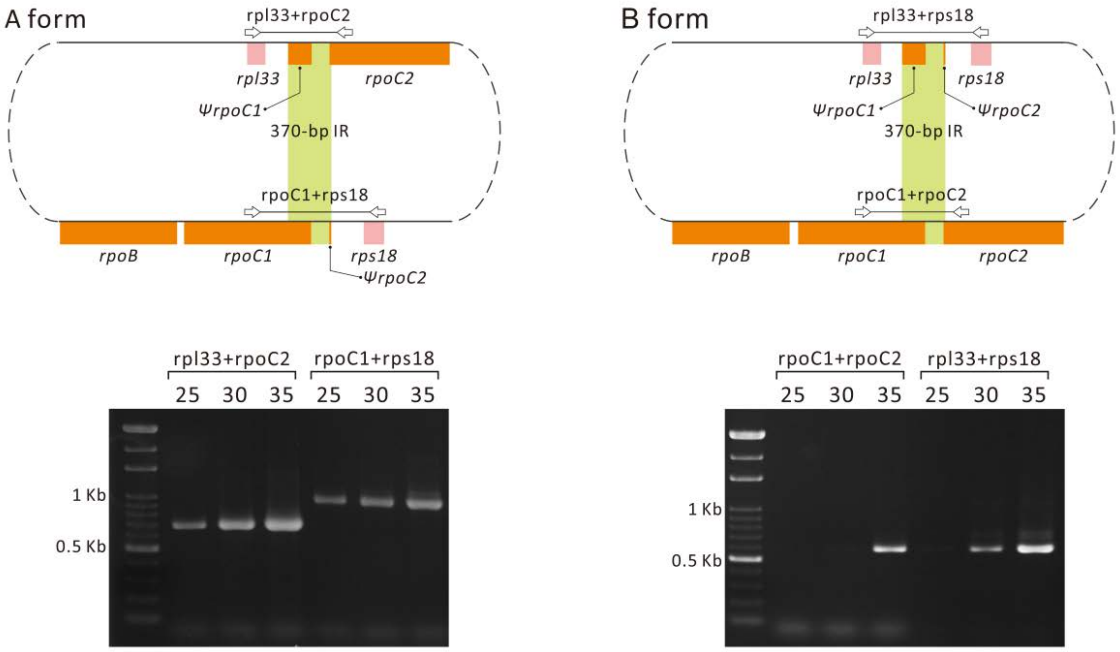

Figure 16

Co-existence of two isomeric plastomes in *Sciadopitys*. The A form is the plastome map obtained from our genome assembly and is shown in Figure 14. The B form differs from the A form by an inversion of the *rpoC2-rps18* (or *rpl33-rpoC1*) fragment. Light green areas are the 370-bp IRs involved in homologous recombination that allows for conversion between the two forms. Paired open arrows are primers specific for the PCR amplification of each form. The corresponding PCR amplicons are shown, and the numbers above each lane of gel photos denote the PCR cycles conducted.

doi:10.6342/NTU201601257

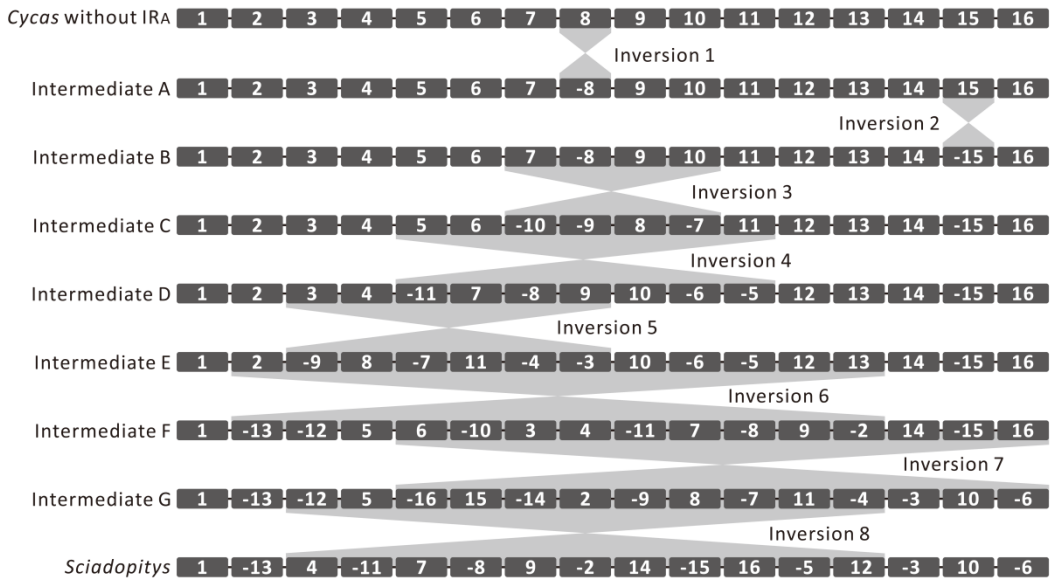

Figure 17

Postulated scenarios for the plastomic inversions in *Sciadopitys*. The plastome of *Cycas* with IRA moved (designated as *Cycas* without IRA) was used in comparison. Eight plastomic inversions that distinguish *Sciadopitys* from *Cycas* are revealed. Gray bars labeled with Arabic numerals represent syntenic blocks of genes between the two species. Syntenic blocks are not drawn to scale.

doi:10.6342/NTU201601257

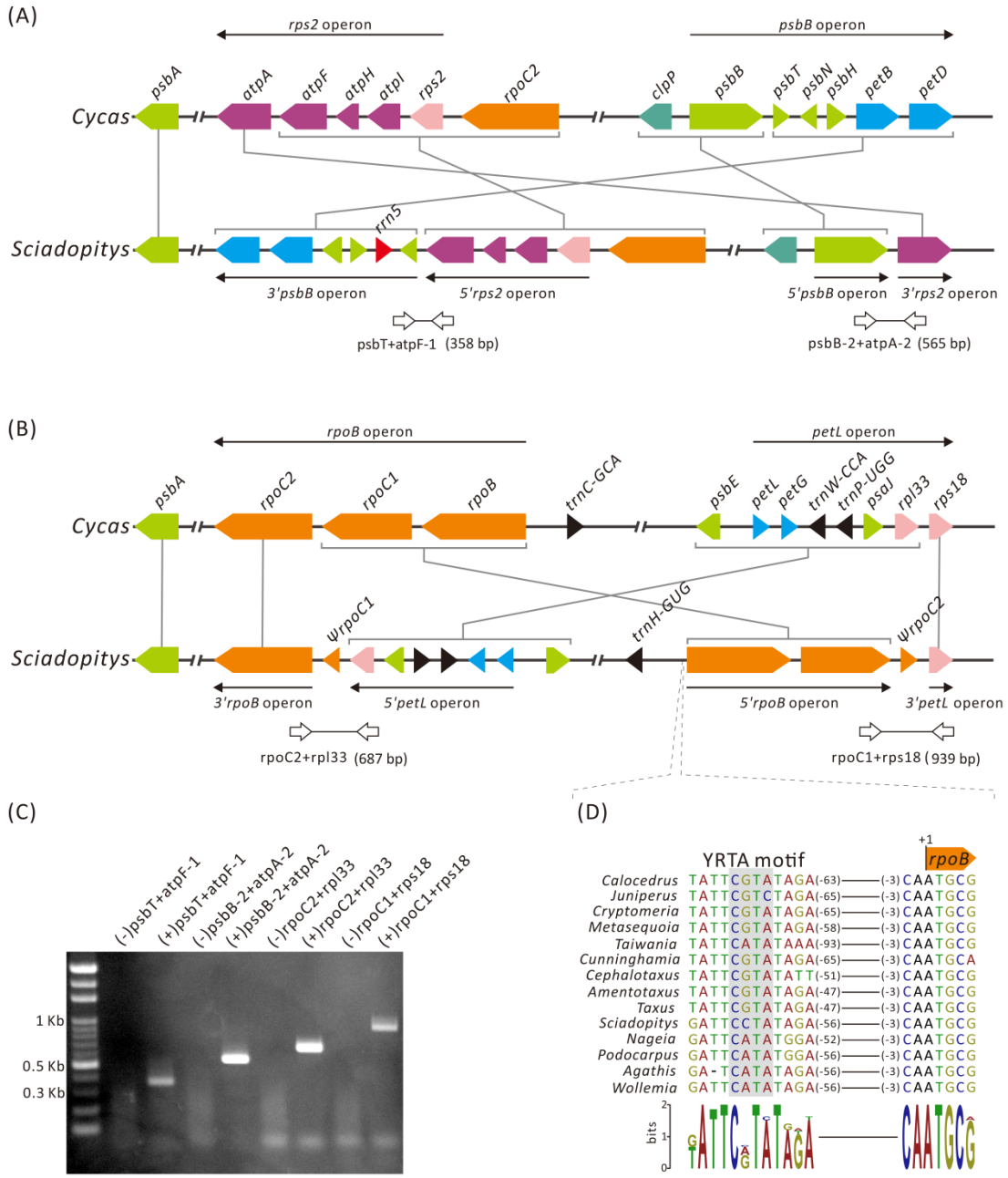

Figure 18

Birth of chimeric gene clustres in the *Sciadopitys* plastome. (A) Shuffling between *rps2* and *psbB* operons and (B) between *rpoB* and *petL* operons. Syntenic genes in the corresponding operons of *Cycas* are used as references. Operons and their transcriptional directions are indicated by solid arrows. Syntenic blocks of genes are connected with gray lines. Paired open arrows are primers for amplifying cDNA fragments across junctions between two recombined operons. The expected sizes of amplicons are shown in parentheses. (C) RT-PCR analysis for detecting the transcripts comprising the genes originated from different operons. Primer pairs used for RT-PCR

doi:10.6342/NTU201601257

assays are shown above the gel panel, with minus and plus signs (in parentheses) denoting the use of RNA (negative control) and cDNA (experimental set) as templates, respectively. (D) YRTA motif of the NEP promoter upstream of *rpoB*.

doi:10.6342/NTU201601257

TABLES

Table 1

Plastid and mitochondria RNA polymerases in higher plants[a]. (Table adapted from Shiina et al., 2005)

RNA polymerase	Subunit	Gene	Gene location
PEP			
Core	α subunit	*rpoA*	Plastid
	β subunit	*rpoB*	Plastid
	β′ subunit	*rpoC1*	Plastid
	β″ subunit	*rpoC2*	Plastid
σ factor	SIG1	*SIG1*	Nuclei
	SIG2	*SIG2*	Nuclei
	SIG3	*SIG3*	Nuclei
	SIG4	*SIG4*	Nuclei
	SIG5	*SIG5*	Nuclei
	SIG6	*SIG6*	Nuclei
NEP			
Plastid target NEP	RpoTp	*RpoTp*(*RpoT;3*)	Nuclei
Mitochondria target NEP	RpoTm	*RpoTm*(*RpoT;1*)	Nuclei
Dual-target NEP	RroTmp	*RroTmp*(*RpoT;2*)	Nuclei
NEP2	Unidentified	Unidentified	

[a] Original names for *RpoT* genes in Hedtke *et al.* (1997, 2000) are shown in parentheses

doi:10.6342/NTU201601257

Table 2

PCR primers used in the *nupt* study

	Primer name	Primer sequence (5'--->3')
Ame-1 & Cep-1	15F2	CAGTRGAAGAACAAATAGCTAYKATTTATRC
	4R1	GAATAGCTTCCGTTGAGTCTCTGC
Ame-2 &Cep-2	1F2	TRGCYGCRTACATTACTTCRAYAGTAAT
	2R2	TTGTCATTTYTYTGAGATCTAGGCAT
Cep-3	-16F3	GCTAAGGCTCATGGRGGBG
	-5R2	TGAAYAGCRTCGGKTAAACCTG
Cep-4	14F2	CGAACCAAAATYTCYGGATGART
	1R2	GGTTACGARGGTACKAATCAAATAGC
Cep-5	20F1	GCATGAGCCATTCCMGTRATRG
	13R1	ATGACYGCAATTYTAGAAAGACGC
Cep-6	13F1	TTACGYTCGTGCATMACTTCCA
	14R2	AYTCATCCRGARATTTTGGTTCG
Tax-1	3F2	ATGGAAGTAAATAHYCTYGCATTTMTTG
	15R2	TSTCCVACTCTTTYYCCATTAGGTA
Tax-2	2F1	TTAAAGAGCGTTTCCACGGG
	3R3	ATGGATATAGTYRDTATYGCTTGGGC
Tax-3	-5F3	GGTGGAGTSACTGCTAGTTTYGG
	-17R4	CYRTTAAACRAGCTCGTATTCTMTSTT
Tax-4	Rpl14F2	GGAACYCGRGTTTTTGGTTC
	Rps11R1	GAGGTCTACATCCRTTATGYGG

doi:10.6342/NTU201601257

Table 3

Characteristics of obtained PCR amplicons in the *nupt* study

ID[1]	Taxon	Accession	Length (bp)	Potential coding sequence
Ame-2	*A. formosana*	AB936749	3,965	partial *chlB*, *rps4*, *trnQ-UUG*, *trnG-UGU*, *trnS-GGA*
Cep-2	*C. wilsoniana*	AB936745	3,796	partial *chlB*[2] *rps4*,*trnQ-UUG*, *trnS-GGA*
Cep-5	*C. wilsoniana*	AB936746	3,788	partial *psbA*, *chlL*, partial *chlN*,*trnI-CAU*, *trnH-GUG*
Cep-6	*C. wilsoniana*	AB936747	2,717	partial *psbA*, partial *matK*, exon 2 of *trnK-UUU*
Tax-4	*T. mairei*	AB936748	2,298	partial *rps11*, *rps36*, *infA*, *rps8*, partial *rpl14*

[1] ID refers to the corresponding primer pairs in Figure 4.

[2] Reading frame with premature stop codons.

doi:10.6342/NTU201601257

Table 4

Mutations in *nupts* and their plastomic counterparts

Nupt	Identity[1] (%)	Length[2] (bp)	No. of mutations			
			Total	Potential protein-coding gene	*N*	*S*
Cep-2	99.08	2,961	29 (1)	*chlB*	19 (0)	7 (1)
				rps4	0 (0)	2 (0)
Cep-5	88.15	3,380	117 (75)	*psbA*	2 (4)	16 (10)
				chlL	3 (4)	24 (17)
				chlN	12 (15)	38 (42)
Cep-6	89.84	2,207	100 (67)	*psbA*	0 (2)	14 (12)
				matK	45 (37)	29 (10)
Tax-4	61.71	1,466	42 (135)	*rpl14*	2 (2)	0 (3)
				rps8	5 (9)	3 (12)
				infA	4 (26)	1 (22)
				rpl36	0 (1)	1 (1)
				rps11	13 (10)	14 (9)

[1] Refers to sequence identity between *nupts* and their plastomic counterparts. Gaps were included in calculating identity.

[2] Refers to lengths of unambiguous alignments where gaps and ambiguous sites were excluded. These alignments were used for calculating mutations.

Note: numbers in parentheses indicate mutations in corresponding plastomic sequences;

N, nonsynonymous; S, synonymous.

doi:10.6342/NTU201601257

Table 5.

Primers used in *Sciadopitys* project

Primer name	Primer sequence (5'--->3')
rpoC2	TCATTCCAAATTGATTTATAAATCTGGTA
rpl33	ACAAACATACCATTCACGGAGAAATA
rpoC1	AAAATAATATTGTTTCCTATAATAAACCTTCG
rps18	TTGGTCGCGGACGTTTACG
psbB-2	TAGTCCGAGGGGTTGGTTTACTT
atpA-2	ACTAATTCCCCTGCCATTACTTGAT
psbT	TTTTCTCCTCTATCCGGAACTTTG
atpF-1	AGAATTCAAAGAATGAGACCATTCACT
SpsbN	ATATCTCGTTTACTTGTAAGCTTTACTGGTT
SatpA	TTTTCTCGAAGTAGGAAAAGTTCGATAT
SrpoC2	CTATTTTTTTACTTGTCGTTTCAAATTTG

doi:10.6342/NTU201601257

Table 6.

Genes predicted in the plastome of *Sciadopitys*

Functional category	Gene[1]
Photosynthesis-related	
Photosystem I & II	*psaA, psaB, psaC, psaI, psaJ, psaM, psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ*
ATPase	*atpA, atpB, atpE, *atpF, atpH, atpI*
Cytochrome b6/f complex	*petA, *petB, *petD, petG, petL, petN,*
NADH dehydrogenase	**ndhA, *ndhB, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK*
RuBisCO synthesis	*rbcL*
Gene expression	
Ribosomal protein	**rpl2, rpl14, *rpl16, rpl20, rpl22, rpl23, rpl32, rpl33, rpl36, rps2, rps3, rps4, rps7, rps8, rps11, 5'rps12, *3'rps12, rps14, rps15, *rps16, rps18, rps19*
RNA polymerase	*rpoA, rpoB, rpoC1, rpoC2*
RNA structural gene	
Ribosomal RNA	*rrn4.5, rrn5*×2*, rrn16, rrn23*
Transfer RNA	**trnA-UAC, trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnfM-CAU, trnG-GCC, *trnG-UCC, trnH-GUG, trnI-CAU*×2*, *trnI-GAU, *trnK-UUU, trnL-CAA, *trnL-UAA, trnL-UAG, trnM-CAU, trnN-GUU, trnP-UGG, trnQ-UUG*×2*, trnR-ACG, trnR-CCG, trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-UGU, trnV-GAC, *trnV-UAC, trnW-CCA, trnY-GUA*
Other	*ccsA, cemA, chlB, chlL, chlN, clpP, infA, matK, ycf1, ycf2, *ycf3, ycf4*
Pseudogene	*ΨrpoC1, ΨtrnV-GAC, ΨtrnP-GGG*×2*, ΨtrnQ-UUG*

[1] "*": intron-containing genes; "×2": two copies

doi:10.6342/NTU201601257

Table 7.

Presence of *trnI-CAU* copies in the plastomes of cupressophytes

Family	Species (GenBank Accession)	Copy[1]	Seq. identity	Score
Cupressaceae	*Calocedrus formosana*	(+)*trnI-CAU*	98.63%	70.17
	(NC_023121)	(-)*trnI-CAU*		75.14
	Juniperus bermudiana	(+)*trnI-CAU*	98.63%	70.17
	(NC_024021)	(-)*trnI-CAU*		75.44
	Cryptomeria japonica	(+)*trnI-CAU*	98.63%	77.3
	(NC_010548)	(-)*trnI-CAU*		74.56
	Metasequoia glyptostroboides	(+)*trnI-CAU*	98.63%	69.43
	(NC_027423)	(-)*trnI-CAU*		74.71
	Cunninghamia lanceolate	(+)*trnI-CAU*	100%	75.44
	(NC_021437)	(-)*trnI-CAU*		75.44
	Taiwania cryptomerioides	(+)*trnI-CAU*	100%	75.44
	(NC_016065)	(-)*trnI-CAU*		75.44
Taxaceae	*Taxus mairei*	(+)*trnI-CAU*	98.63%	75.44
	(AP014575)	(-)*trnI-CAU*		75.39
	Amentotaxus formosana	(+)*trnI-CAU*	98.63%	75.44
	(NC_024945)	(-)*trnI-CAU*		75.67
Cephalotaxaceae	*Cephalotaxus wilsoniana*	(+)*trnI-CAU*	-	77.31
	(NC_016063)	-		
Sciadopityaceae	*Sciadopitys verticillata*	(+)*trnI-CAU*	86.30%	48.48
	(AP017299)	(+)*trnI-CAU*		78.15
Podocarpaceae	*Nageia nagi*	(+)*trnI-CAU*	-	75.44
	(NC_023120)	-		
	Podocarpus lambertii	(+)*trnI-CAU*	-	77.3
	(NC_023805)	-		
Araucariaceae	*Agathis dammara*	(+)*trnI-CAU*	97.26%	75.25
	(NC_023119)	(-)*trnI-CAU*		66.46
	Wollemia nobilis	(+)*trnI-CAU*	94.52%	75.25
	(NC_027235)	(-)*trnI-CAU*		55.15

[1] "+" and "-" in parentheses denote the transcriptional directions

doi:10.6342/NTU201601257

REFERENCS

Alkatib S, Fleischmann TT, Scharff LB, Bock R. 2012. Evolutionary constraints on the plastid tRNA set decoding methionine and isoleucine. Nucleic Acids Research 40:6713–6724.

Allen JF. 2003. Why chloroplasts and mitochondria contain genomes. Comparative and Functional Genomics 4:31–36.

Allison LA. 2000. The role of sigma factors in plastid transcription. Biochimie 82:537–548.

Allison LA, Simon LD, Maliga P. 1996. Deletion of *rpoB* reveals a second distinct transcription system in plastids of higher plants. The EMBO Journal 15:2802–2809.

Archibald, JM. 2009. The puzzle of plastid evolution. Current Biology: CB. 19:R81–88.

Baba K, Schmidt J, Espinosa-Ruiz A, Villarejo A, Shiina T, Gardeström P, Sane AP, Bhalerao RP. (2004). Organellar gene transcription and early seedling development are affected in the *rpoT*; 2 mutant of Arabidopsis. The Plant Journal 38:38–48.

Barkan A. 1988. Proteins encoded by a complex chloroplast transcription unit are each translated from both monocistronic and polycistronic mRNAs. The EMBO Journal 7:2637–2644.

Bergthorsson U, Adams KL, Thomason B, Palmer JD. 2003. Widespread horizontal transfer of mitochondrial genes in flowering plants. Nature 424:197–201.

Birky CW. 1995. Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. Proceedings of the National Academy of Sciences 92:11331–11338.

Blazier JC, Guisinger MM, Jansen RK. 2011. Recent loss of plastid-encoded *ndh* genes within Erodium (Geraniaceae). Plant Molecular Biology 76:263–272.

Bligny M, et al. 2000. Regulation of plastid rDNA transcription by interaction of CDF2 with two different RNA polymerases. The EMBO Journal 19:1851–1860.

doi:10.6342/NTU201601257

Bourque G, Pevzner PA. 2002. Genome-scale evolution: reconstructing gene orders in the ancestral species. Genome Research 12:26–36.

Börner T, Aleynikova AY, Zubo YO, Kusnetsov VV. 2015. Chloroplast RNA polymerases: role in chloroplast biogenesis. Biochim Biophys Acta. 1847:761–769.

Bryant, DA, Frigaard, NU. 2006. Prokaryotic photosynthesis and phototrophy illuminated. Trends in Microbiology 14:488–496.

Buschiazzo E, Ritland C, Bohlmann J, Ritland K. 2012. Slow but not low: genomic comparisons reveal slower evolutionary rate and higher dN/dS in conifers compared to angiosperms. BMC Evolionary Biology 12:8.

Cai Z, et al. 2008. Extensive reorganization of the plastid genome of *Trifolium subterraneum* (Fabaceae) is associated with numerous repeated sequences and novel DNA insertions. Journal of Molecular Evolution 67:696–704.

Chaw SM, Parkinson CL, Cheng Y, Vincent TM, Palmer JD. 2000. Seed plant phylogeny inferred from all three plant genomes: monophyly of extant gymnosperms and origin of Gnetales from conifers. Proceedings of the National Academy of Sciences 97:4086–4091.

Chaw SM, Chang CC, Chen HL, Li WH. 2004. Dating the monocot-dicot divergence and the origin of core eudicots using whole chloroplast genomes. Journal of Molecular Evolution 58:1–18.

Cheng Y, Nicolson RG, Tripp K, Chaw SM. 2000. Phylogeny of Taxaceae and Cephalotaxaceae genera inferred from chloroplast *matK* gene and nuclear rDNA ITS region. Molecular Phylogenetics and Evolution 14:353–365.

Conway S. 2013. Beyond pine cones: an introduction to gymnosperms. Arnoldia 70/4.

Corriveau JL, Coleman AW. 1988. Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperm species. American Journal of Botany 1443–1458.

Cosner ME. 1993. Phylogenetic and molecular evolutionary studies of chloroplast DNA

doi:10.6342/NTU201601257

variation in the Campanulaceae. PhD thesis, The Ohio State University, Columbus.

Cosner ME, Raubeson LA, Jansen RK. 2004. Chloroplast DNA rearrangements in Campanulaceae: phylogenetic utility of highly rearranged genomes. BMC Evolutionary Biology 4:27.

Crisp MD, Cook LG. 2011. Cenozoic extinctions account for the low diversity of extant gymnosperms compared with angiosperms. New Phytologist 192:997–1009.

Cui L, Leebens-Mack J, Wang LS, Tang J, Rymarquis L, Stern DB. 2006. Adaptive evolution of chloroplast genome structure inferred using a parametric bootstrap approach. BMC Evolutionary Biology 6:13.

Darling AC, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Research 14:1394–403.

Darling AE, Mau B, Perna NT. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One 5:e11147.

de Laubenfels, David J. 1988. Coniferales. Flora Malesiana, Series I, Dordrecht: Kluwer Academic 10:337–453.

de Vries J, Stanton A, Archibald JM, Gould SB. 2016. Streptophyte terrestrialization in light of plastid evolution. Trends in Plant Science 21:467–476.

Deusch O, Landan G, Roettger M, et al. 2008. Genes of cyanobacterial origin in plant nuclear genomes point to a heterocyst-forming plastid ancestor. Molecular Biology and Evolution 25:748–761.

Downie SR, Palmer JD. 1992. Use of chloroplast DNA rearrangements in reconstructing plant phylogeny. In Molecular systematics of plants. Springer US. 14–35.

Doyle JJ, Doyle JL, Palmer JD. 1995. Multiple independent losses of two genes and one intron from legume chloroplast genomes. Systematic Botany 272–294.

Doyle JJ, Doyle JL, Ballenger JA, Palmer JD. 1996. The distribution and phylogenetic significance of a 50-kb chloroplast DNA inversion in the flowering plant family Leguminosae. Molecular Phylogenetics and Evolution 5:429–438.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high

doi:10.6342/NTU201601257

throughput. Nucleic Acids Research 32:1792–1797.

Finster S, Eggert E, Zoschke R, Weihe A, Schmitz-Linneweber C. 2013. Light-dependent, plastome-wide association of the plastid-encoded RNA polymerase with chloroplast DNA. The Plant Journal 76:849–860.

Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. 2004. VISTA: computational tools for comparative genomics. Nucleic Acids Research 32:W273–279

Gantt JS, Baldauf SL, Calie PJ, Weeden NF, Palmer JD. 1991. Transfer of rpl22 to the nucleus greatly preceded its loss from the chloroplast and involved the gain of an intron. The EMBO Journal 10:3073–3078.

Graur D, Li WH. 2000. Fundamentals of Molecular Evolution. Sinauer Associates, Sunderland, MA.

Gray MW, Lang BF. 1998. Transcription in chloroplasts and mitochondria: a tale of two polymerases. Trends Microbiol 6:1–3

Guisinger MM, Kuehl JV, Boore JL, Jansen RK. 2008. Genome-wide analyses of Geraniaceae plastid DNA reveal unprecedented patterns of increased nucleotide substitutions. Proceedings of the National Academy of Sciences 105:18424–18429.

Guisinger MM, Kuehl JV, Boore JL, Jansen RK. 2011. Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: rearrangements, repeats, and codon usage. Molecular Biology and Evolution 28:583–600.

Guo W, et al. 2014. Predominant and substoichiometric isomers of the plastid genome coexist within *Juniperus* plants and have shifted multiple times during cupressophyte evolution. Genome Biology and Evolution 6:580–590.

Gymnosperms on The Plant List. http://www.theplantlist.org/browse/G/. Retrieved 2016/5/12.

Haberle RC, Fourcade HM, Boore JL, Jansen RK. 2008. Extensive rearrangements in the chloroplast genome of *Trachelium caeruleum* are associated with repeats and tRNA genes. Journal of Molecular Evolution 66:350–361

Hajdukiewicz PT, Allison LA, Maliga P. 1997. The two RNA polymerases encoded by

doi:10.6342/NTU201601257

the nuclear and the plastid compartments transcribe distinct groups of genes in tobacco plastids. The EMBO Journal 16:4041–4048.

Hanaoka M, Kanamaru K, Takahashi H, Tanaka K. 2003. Molecular genetic analysis of chloroplast gene promoters dependent on SIG2, a nucleus-encoded sigma factor for the plastid-encoded RNA polymerase, in Arabidopsis thaliana. Nucleic Acids Research 31:7090–7098.

Hansen AK, Escobar LK, Gilbert LE, Jansen RK. 2007. Paternal, maternal, and biparental inheritance of the chloroplast genome in Passiflora (Passifloraceae): implications for phylogenetic studies. American Journal of Botany 94:42–46.

Hazkani-Covo E, Zeller RM, Martin W. 2010. Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. PLoS Genetics 6:e1000834.

Hedtke B, Börner T, Weihe A. 1997. Mitochondrial and chloroplast phage-type RNA polymerases in *Arabidopsis*. Science 277:809–811.

Hedtke B, Börner T, Weihe A. 2000. One RNA polymerase serving two genomes. EMBO Reports 1:435–440.

Hirao T, Watanabe A, Kurita M, Kondo T, Takata K. 2008. Complete nucleotide sequence of the *Cryptomeria japonica* D. Don. chloroplast genome and comparative chloroplast genomics: diversified genomic structure of coniferous species. BMC Plant Biology 8:70.

Hsu CY, Wu CS, Chaw SM. 2014. Ancient nuclear plastid DNA in the yew family (Taxaceae). Genome Biology and Evolution 6:2111–2121.

Huang CY, Grünheit N, Ahmadinejad N, Timmis JN, Martin W. 2005. Mutational decay and age of chloroplast and mitochondrial genomes transferred recently to angiosperm nuclear chromosomes. Plant Physiology 138:1723–1733.

Hu Y, Zhang Q, Rao G. 2008. Occurrence of plastids in the sperm cells of Caprifoliaceae: biparental plastid inheritance in angiosperms is unilaterally derived from maternal inheritance. Plant and Cell Physiology 49:958–968.

Hu J, Bogorad L. 1990. Maize chloroplast RNA polymerase: the 180-, 120-, and

doi:10.6342/NTU201601257

38-kilodalton polypeptides are encoded in chloroplast genes. Proceedings of the National Academy of Sciences 87:1531–1535.

Hudson GS, Holton TA, Whitfeld PR, Bottomley W. 1988. Spinach chloroplast rpoBC genes encode three subunits of the chloroplast RNA polymerase. Journal of Molecular Biology 200:639–654.

Igloi GL, Kössel H. 1992. The transcriptional apparatus of chloroplasts. Critical Reviews in Plant Sciences10:525–558.

Ishihama A. 2000. Functional modulation of Escherichia coli RNA polymerase. Annual Reviews in Microbiology 54: 499–518.

Ishizaki Y, Tsunoyama Y, Hatano K, Ando K, Kato K, Shinmyo A, Kobori M, Takeba G, Nakahira Y, Shiina T. 2005. A nuclear-encoded sigma factor, Arabidopsis SIG6, recognizes sigma-70 type chloroplast promoters and regulates early chloroplast development in cotyledons. The Plant Journal 42:133–144.

Jansen RK, Cai Z, Raubeson LA, Daniell H, Leebens-Mack J, Müller KF, Lee SB. 2007. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. Proceedings of the National Academy of Sciences 104:19369–19374.

Jansen RK, Ruhlman TA. 2012. Plastid genomes of seed plants. In: Bock R, Knoop V, editors. Genomics of Chloroplasts and Mitochondria. Netherlands : Springer 103–126.

Jiang ZK, Wang YD, Zheng SL, Zhang W, Tian N. 2012. Occurrence of Sciadopitys-like fossil wood (Conifer) in the Jurassic of western Liaoning and its evolutionary implications. Chinese Science Bulletin 57:569–572

Kapoor S, Sugiura M. 1999. Identification of two essential sequence elements in the nonconsensus type II PatpB-290 plastid promoter by using plastid transcription

doi:10.6342/NTU201601257

extracts from cultured tobacco BY-2 cells. Plant Cell 11:1799–1810.

Kapoor S, Suzuki JY, Sugiura M. 1997. Identification and functional significance of a new class of non-consensus-type plastid promoters. The Plant Journal 11:327–337.

Kanamaru K, Nagashima A, Fujiwara M, Shimada H, Shirano Y, Nakabayashi K, Shibata S, Tanaka K, Takahashi H. 2001. An Arabidopsis sigma factor (SIG2)-dependent expression of plastid-encoded tRNAs in chloroplasts. Plant and Cell Physiology 42:1034–1043.

Kleine T, Maier UG, Leister D. 2009. DNA transfer from organelles to the nucleus: the idiosyncratic genetics of endosymbiosis. Annual Review of Plant Biology 60:115–138.

Kobayashi Y, Dokiya Y, Sugita M. 2001. Dual targeting of phage-type RNA polymerase to both mitochondria and plastids is due to alternative translation initiation in single transcripts. Biochemical and Biophysical Research Communications 289:1106–1113.

Kolosova N, et al. 2004. Isolation of high-quality RNA from gymnosperm and angiosperm trees. Biotechniques 36:821–824.

Krzywinski M, et al. 2009. Circos: an information aesthetic for comparative genomics. Genome Research 19:1639–1645.

Ku C, et al. 2015. Endosymbiotic origin and differential loss of eukaryotic genes. Nature 524:427–432.

Legen J, et al. 2002. Comparative analysis of plastid transcription profiles of entire plastid chromosomes from tobacco attributed to wild-type and PEP-deficient transcription machineries. The Plant Journal 31:171–188.

Leister D. 2005. Origin, evolution and genetic effects of nuclear insertions of organelle DNA. TRENDS in Genetics 21:655–663.

Leslie AB, Beaulieu JM, Rai HS, Crane PR, Donoghue MJ, Mathews S. 2012. Hemisphere-scale differences in conifer evolutionary dynamics. Proceedings of the

doi:10.6342/NTU201601257

National Academy of Sciences 109:16217–16221.

Librado P, Rozas J. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25: 1451–1452.

Li J, et al. 2016. Evolution of short inverted repeat in cupressophytes, transfer of *accD* to nucleus in Sciadopitys verticillata and phylogenetic position of Sciadopityaceae. Scientific Reports 6:20934.

Liere K, Börner T. 2007. Transcription and transcriptional regulation in plastids. In Cell and molecular biology of plastids. Springer Berlin Heidelberg 121–174.

Liere K, Maliga P. 2001. Plastid RNA polymerases in higher plants. In Regulation of photosynthesis. Springer Netherlands 29–49.

Liere K, Börner T. 2007. Transcription and transcriptional regulation in plastids. In Cell and molecular biology of plastids. Springer Berlin Heidelberg 121–174.

Lin CP, Huang JP, Wu CS, Hsu CY, Chaw SM. 2010. Comparative chloroplast genomics reveals the evolution of Pinaceae genera and subfamilies. Genome Biology and Evolution 2:504–517.

Lloyd AH, Timmis JN. 2011. The origin and characterization of new nuclear genes originating from a cytoplasmic organellar genome. Molecular Biology and Evolution 28:2019–2028.

Magee AM. et al. 2010. Localized hypermutation and associated gene losses in legume chloroplast genomes. Genome Research 20:1700–1710.

Mao K. 2012. Distribution of living Cupressaceae reflects the breakup of Pangea. Proceedings of the National Academy of Sciences 109:7793–7798.

Maréchal A, Brisson N. 2010. Recombination and the maintenance of plant organelle genome stability. New Phytologist 186:299–317.

Martin W, Stoebe B, Goremykin V, Hansmann S, Hasegawa M, Kowallik, K.V. 1998. Gene transfer to the nucleus and the evolution of chloroplasts. Nature 393:162–165.

Martin W, Rujan T, Richly E, Hansen A, Cornelsen S, Lins T, Leister D, Stoebe B, Hasegawa M, Penny D.. 2002. Evolutionary analysis of Arabidopsis, cyanobacterial,

doi:10.6342/NTU201601257

and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. Proceedings of the National Academy of Sciences 99:12246–12251.

Meierhoff K, Felder S, Nakamura T, Bechtold N, Schuster G. 2003. HCF152, an Arabidopsis RNA binding pentatricopeptide repeat protein involved in the processing of chloroplast psbB-psbT-psbH-petB-petD RNAs. The Plant Cell 15:1480–1495.

Michalovova M, Vyskot B, Kejnovsky E. 2013. Analysis of plastid and mitochondrial DNA insertions in the nucleus (NUPTs and NUMTs) of six plant species: size, relative age and chromosomal localization. Heredity 111:314–320.

Milligan BG, Hampton JN, Palmer JD. 1989. Dispersed repeats and structural reorganization in subclover chloroplast DNA. Molecular Biology and Evolution 6:355–368.

Mochizuki N, Tanaka R, Tanaka A, Masuda T, Nagatani A. 2008. The steady-state level of Mgprotoporphyrin IX is not a determinant of plastid-to-nucleus signaling in Arabidopsis. Proceedings of the National Academy of Sciences 105:15184–89.

Mogensen HL. 1996. The hows and whys of cytoplasmic inheritance in seed plants. American Journal of Botany 83:383–404.

Moran NA. 1996. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. Proceedings of the National Academy of Sciences 93:2873–2878.

Morden CW, Wolfe KH. 1991. Plastid translation and transcription genes in a non-photosynthetic plant: intact, missing and pseudo genes. The EMBO Journal 10:3281.

Muller HJ. 1932. Some genetic aspects of sex. The American Naturalist 66:118–138.

Muller HJ. 1964. The relation of recombination to mutational advance. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis 1:2–9.

Noutsos C, Kleine T, Armbruster U, DalCorso G, Leister D. 2007. Nuclear insertions of organellar DNA can create novel patches of functional exon sequences. Trends in Genetics 23:597–601.

Noutsos C, Richly E, Leister D. 2005. Generation and evolutionary fate of insertions of organelle DNA in the nuclear genomes of flowering plants. Genome Research 15:616–628.

Ohba K, Iwakawa M, Okada Y, Murai M. 1971. Paternal transmission of a plastid

doi:10.6342/NTU201601257

anomaly in some reciprocal crosses of Sugi, Cryptomeria japonica D. Don. Silvae Genet. 20:101–107.

Ohyama K, et al. 1986. Chloroplast gene organization deduced from complete sequence of liverwort Marchantia polymorpha chloroplast DNA. Nature 322:572–574.

Ortelt J, Link, G. 2014. Plastid gene transcription: promoters and RNA polymerases. Chloroplast Biotechnology: Methods and Protocols 47–72.

Palmer JD, Thompson WF. 1981. Rearrangements in the chloroplast genomes of mung bean and pea. Proceedings of the National Academy of Sciences 78:5533–5537.

Palmer JD. 1983. Chloroplast DNA exists in two orientations. Nature 301, 92–93.

Palmer JD. 1985. Comparative organization of chloroplast genomes. Annual Review of Genetics 19:325–354.

Palmer JD. 1991. Plastid chromosomes: structure and evolution. The Molecular Biology of Plastids 7:5–53.

Perry AS, Brennan S, Murphy DJ, Kavanagh TA, Wolfe KH. 2002. Evolutionary re-organisation of a large operon in adzuki bean chloroplast DNA caused by inverted repeat movement. DNA Research 9:157–162.

Quesada-Vargas T, Ruiz ON, Daniell H. 2005. Characterization of heterologous multigene operons in transgenic chloroplasts. Transcription, processing, and translation. Plant Physiology 138:1746–1762.

Pfannschmidt T, Nilsson A, Tullberg A, Link G, Allen JF. 1999. Direct transcriptional control of the chloroplast genes *psbA* and *psaAB* adjusts photosynthesis to light energy distribution in plants. IUBMB Life 48:271–276.

Pfannschmidt T, Nilsson A, Allen JF. 1999. Photosynthetic control of chloroplast gene expression. Nature 397:625–628.

Privat I, Hakimi MA, Buhot L, Favory JJ. Lerbs-Mache S. 2003. Characterization of Arabidopsisplastid sigma-like transcription factors SIG1, SIG2 and SIG3. Plant Molecular Biology 51:385–399.

doi:10.6342/NTU201601257

Rai HS, Reeves PA, Peakall R, Olmstead RG, Graham SW. 2008. Inference of higher-order conifer relationships from a multi-locus plastid data set. Botany 86: 658–669.

Raubeson LA, Jansen RK. 1992. A rare chloroplast DNA structure mutation is shared by all conifers. Biochemical Systematics and Ecology 20:17–24.

Richly E, Leister D. 2004. NUPTs in sequenced eukaryotes and their genomic organization in relation to NUMTs. Molecular Biology and Evolution 21:1972–1980.

Rice DW, Alverson AJ, Richardson AO, et al. 2013. Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm Amborella. Science 342:1468–1473.

Rousseau-Gueutin M, Ayliffe MA, Timmis JN. 2011. Conservation of plastid sequences in the plant nuclear genome for millions of years facilitates endosymbiotic evolution. Plant Physiology 157:2181–2193.

Sakai A, Saito C, Inada N, Kuroiwa T. 1998. Transcriptional activities of the chloroplast-nuclei and proplastid-nuclei isolated from tobacco exhibit different sensitivities to tagetitoxin: implication of the presence of distinct RNA polymerases. Plant and Cell Physiology 39:928–934.

Schattner P, Brooks AN, Lowe TM. 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Research 33(Web Server issue):W686–689.

Seong D, Offner S. 2013. A phylogenetic study of conifers describes their evolutionary relationships and reveals potential explanations for current distribution patterns. J. Emerg. Invest.. http://www. emerginginvestigators. org/2013/10.

Sexton TB, Christopher DA, Mullet JE. 1990. Light-induced switch in barley psbD-psbC promoter utilization: a novel mechanism regulating chloroplast gene expression. The EMBO Journal 9:4485.

Sheppard AE, Ayliffe MA, Blatch L, et al. 2008. Transfer of plastid DNA to the nucleus is elevated during male gametogenesis in tobacco. Plant Physiology 148:328–336.

doi:10.6342/NTU201601257

Sheppard AE, Timmis JN. 2009. Instability of plastid DNA in the nuclear genome. PLoS Genetics 5:e1000323.

Shiina T, Tsunoyama Y, Nakahira Y, Khan MS. 2005. Plastid RNA polymerases, promoters, and transcription regulators in higher plants. International Review of Cytology 244:1–68.

Shirano Y, Shimada H, Kanamaru K, Fujiwara M, Tanaka K, Takahashi H, Unno K, Sato S, Tabata S, Hayashi H, Miyake C, Yokota A, Shibata D. 2000. Chloroplast development in Arabidopsis thaliana requires the nuclear-encoded transcription factor sigma B. FEBS letters 485:178–182.

Stewart CN Jr, Via LE. 1993. A rapid CTAB DNA isolation technique useful for RAPD fingerprinting and other PCR applications. Biotechniques 14:748–750.

Stern DB, Goldschmidt-Clermont M, Hanson MR. 2010. Chloroplast RNA metabolism. Annual Review of Plant Biology 61:125–155.

Stine M, Sears BB, Keathley DE. 1989. Inheritance of plastids in interspecific hybrids of blue spruce and white spruce. Theoretical and Applied Genetics 78:768–774.

Sugiura M. 1992. The chloroplast genome. Plant Molecular Biology 19:149–168.

Szmidt AE, Aldén T, Hällgren JE. 1987. Paternal inheritance of chloroplast DNA in Larix. Plant Molecular Biology 9:59–64.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution 28:2731–2739.

Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch S. 2001. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Research 11:1441–1452.

Testolin R, Cipriani G. 1997. Paternal inheritance of chloroplast DNA and maternal inheritance of mitochondrial DNA in the genus Actinidia. Theoretical and Applied Genetics 94:897–903.

doi:10.6342/NTU201601257

Timmis JN, Ayliffe MA, Huang CY, Martin W. 2004. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat Rev Genet. 5:123–135.

Timmis JN, Scott NS. 1983. Sequence homology between spinach nuclear and chloroplast genomes. Nature 305:65–67.

Tsunoyama Y, Ishizaki Y, Morikawa K, Kobori M, Nakahira Y, Takeba G, Toyoshima Y, Shiina, T. 2004. Blue light-induced transcription of plastid-encoded psbD gene is mediated by a nuclear-encoded transcription initiation factor, AtSig5. Proceedings of the National Academy of Sciences 101:3304–3309.

Tsumura Y, Suyama Y, Yoshimura K. 2000. Chloroplast DNA inversion polymorphism in populations of *Abies* and *Tsuga*. Molecular Biology and Evolution 17:1302–1312.

Vieira Ldo N, et al. 2014. The complete chloroplast genome sequence of *Podocarpus lambertii*: genome structure, evolutionary aspects, gene content and SSR detection. PLoS One 9:e90618.

Wang D, Wu YW, Shih AC, Wu CS, Wang YN, Chaw SM. 2007. Transfer of chloroplast genomic DNA to mitochondrial genome occurred at least 300 MYA. Molecular Biology and Evolution 24:2040–2048.

Wang XQ, Ran JH. 2014. Evolution and biogeography of gymnosperms. Molecular Phylogenetics and Evolution 75C:24–40.

Weng ML, Blazier JC, Govindu M, Jansen RK. 2013. Reconstruction of the ancestral plastid genome in Geraniaceae reveals a correlation between genome rearrangements, repeats and nucleotide substitution rates. Molecular Biology and Evolution 31:645–659.

Wicke S, Schneeweiss GM, dePamphilis CW, Müller KF, Quandt D. 2011. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Molecular Biology 76:273–297.

Wise RR, Hoober JK. 2006. The structure and function of plastids. Springer Science Business Media.

Wolfe KH, Li WH, Sharp PM. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proceedings of the

doi:10.6342/NTU201601257

National Academy of Sciences 84:9054–9058.

Wojciechowski MF, Lavin M, Sanderson MJ. 2004. A phylogeny of legumes (Leguminosae) based on analysis of the plastid *matK* gene resolves many well-supported subclades within the family. American Journal of Botany 91:1846–1862.

Wu CS, Chaw SM, Huang YY. 2013. Chloroplast phylogenomics indicates that Ginkgo biloba is sister to cycads. Genome Biology and Evolution 5:243–254.

Wu CS, Chaw SM. 2014. Highly rearranged and size-variable chloroplast genomes in conifers II clade (cupressophytes): evolution towards shorter intergenic spacers. Plant Biotechnology Journal 12:344–353.

Wu CS, Chaw, SM. 2015. Evolutionary stasis in cycad plastomes and the first case of plastome GC-biased gene conversion. Genome Biology and Evolution 7:2000–2009.

Wu CS, Lin CP, Hsu CY, Wang RJ, Chaw SM. 2011. Comparative chloroplast genomes of Pinaceae: insights into the mechanism of diversified genomic organizations. Genome Biology and Evolution 3:309–319.

Wu CS, Wang YN, Hsu CY, Lin CP, Chaw SM. 2011. Loss of different inverted repeat copies from the chloroplast genomes of Pinaceae and cupressophytes and influence of heterotachy on the evaluation of gymnosperm phylogeny. Genome Biology and Evolution 3:1284–1295.

Wu CS, Wang YN, Liu SM, Chaw SM. 2007. Chloroplast genome (cpDNA) of *Cycas taitungensis* and 56 cp protein-coding genes of *Gnetum parvifolium*: insights into cpDNA evolution and phylogeny of extant seed plants. Molecular Biology and Evolution 24:1366–1379.

Wyman SK, Jansen RK, Boore JL. 2004. Automatic annotation of organellar genomes with DOGMA. Bioinformatics 20:3252–3255.

Yap JY, et al. 2015. Complete chloroplast genome of the wollemi pine (*Wollemia nobilis*): structure and evolution. PLoS One 10:e0128126.

doi:10.6342/NTU201601257

Yi X, Gao L, Wang B, Su YJ, Wang T. 2013. The complete chloroplast genome sequence of *Cephalotaxus oliveri* (Cephalotaxaceae): evolutionary comparison of *Cephalotaxus* chloroplast DNAs and insights into the loss of inverted repeat copies in gymnosperms. Genome Biology and Evolution 5:688–698.

Yoshida T, Furihata HY, Kawabe A. 2014. Patterns of genomic integration of nuclear chloroplast DNA fragments in plant species. DNA Research 21:127–140.

Young DA, Allen RL, Harvey AJ, Lonsdale DM. 1998. Characterization of a gene encoding a single-subunit bacteriophage-type RNA polymerase from maize which is alternatively spliced. Molecular and General Genetics MGG 260:30–37.

Zhang Q, Liu Y 2003. Examination of the cytoplasmic DNA in male reproductive cells to determine the potential for cytoplasmic inheritance in 295 angiosperm species. Plant and Cell Physiology 44:941–951.

Zhang Y, Ma J, Yang B, et al. 2014. The complete chloroplast genome sequence of *Taxus chinensis* var. *mairei* (Taxaceae): loss of an inverted repeat region and comparative analysis with related species. Gene 540:201–209.

Zhelyazkova P, et al. 2012. The primary transcriptome of barley chloroplasts: numerous noncoding RNAs and the dominating role of the plastid-encoded RNA polymerase. The Plant Cell 24:123-136.

doi:10.6342/NTU201601257

PUBLICATIONS

doi:10.6342/NTU201601257

Send Orders of Reprints at bspsaif@emirates.net.ae

Current Pharmaceutical Biotechnology, **2012**, *13*, 2721-2732

Tangy Scent in *Toona sinensis* (Meliaceae) Leaflets: Isolation, Functional Characterization, and Regulation of TsTPS1 and TsTPS2, Two Key Terpene Synthase Genes in the Biosynthesis of the Scent Compound

Chih-Yao Hsu[1,2,3], Pung-Ling Huang[2], Chih-Ming Chen[1], Chi-Tang Mao[1] and Shu-Miaw Chaw[1,*]

[1]*Biodiversity Research Center, Academia Sinica, 128 Academy Road, Section 2, Taipei 115, Taiwan;* [2]*Department of Horticulture and Landscape Architecture, National Taiwan University, Taipei 106, Taiwan;* [3]*Current address: Genome and Systems Biology, National Taiwan University, 1 Roosevelt Road, Section 4, Taipei 106, Taiwan*

Abstract: *Toona sinensis* (Chinese Mahogany; Meliaceae), a subtropical deciduous tree, has a tangy scent resembling a mix of shallots and garlic. *T. sinensis* has long been known for its medicinal efficacy for treating enteritis, dysentery, itch and some cancers. However, its volatile components and their biosynthesis remain unexamined. In this study, we identified the spectrum of volatile compounds, isolated and functionally characterized two terpene synthase genes, *Tstps1* and *Tstps2*, responsible for terpenoid synthesis in *T. sinensis* leaflets. TsTPS1 and TsTPS2 afford multiple products upon incubation with geranyl and farnesyl diphosphate respectively and mainly regulate the biosynthesis of (+) limonene and β-elemene *in vitro*, respectively. Headspace analyses show that 98% of leaflet volatiles were sesquiterpenoids and the developing leaflets released a greater diversity and quantity of volatiles than the mature leaflets did, and that β-elemene was the dominant component in both of them. These data suggested that tangy scent of *T. sinensis* consists of a combination of terpenoids and that *Tstps2* was the major gene involved in the terpenoid biosynthesis in *T. sinensis*. *In situ* hybridization revealed that glandular cells of the leaf rachises accumulated abundant *Tstps1* mRNA transcripts. Our GFP-based assay further unprecedentedly demonstrated that the transit-peptide of TsTPS1 targets specifically to the mitochondria.

Keywords: β-elemene, (+) limonene, terpene synthase, terpenoid, *Toona sinensis*, volatiles.

INTRODUCTION

Toona sinensis (Chinese Mahogany; Meliaceae; abbreviated as *Toona*) has been used widely in Asia for food and health. It is a deciduous tree native to eastern and southeastern Asia. Its leaves are compound, and when young, they are red to red-brown and turn green on maturation (after about 2 weeks) (Supplement **1**). *Toona* has a distinctively tangy scent that resembles a mixture of shallots and garlic. In Asia, the young leaflets and the rachises of the compound leaves (hereafter, termed "leaves" unless otherwise specified) are used as a vegetable and as seasoning, and the dried developing leaflets are used to make a health tea. Furthermore, the leaflets have long been used to treat enteritis, dysentery and pruritus in traditional Asian medicine [1]. *Toona* leaf extracts were found to induce apoptosis in human ovarian cancer cells and inhibit tumor growth in a rat model [2]. Recently, Chen and coworkers used *in vitro* experiments to show that extracts of the developing leaves inhibited the corona virus associated with severe acute respiratory syndrome (SARS) [3].

The leaves of *Toona* have been used extensively but the knowledge of their chemicals contributing to the *Toona* plant's distinctive taste and smell is limited. Mu and coworkers used microwave-assisted extraction and solid-phase microextraction methods to collect the volatile compound emitted from *Toona* leaves [4]. Their results revealed that the *Toona* leaves contained a great variety of terpenoids and the major compound of them was trans-caryophyllene, a sesquiterpenoid. Terpenoids are the largest family of natural plant products. They play crucial roles in the interactions of plants and their environments by emitting from the aerial or underground parts of plants. To date, more than 40,000 terpenoid molecules have been identified [5]. Terpenoids have economic value because both they and their derivatives are used extensively to flavor and add fragrance to food and cosmetics [6, 7]. Among terpenoids, monoterpenoids and sesquiterpenoids are the most common volatiles and serve as a direct or an indirect defense against potential herbivores and pathogens [8]. Terpenoids have also been shown to be effective against a wide range of tumors. For example, elemene, a sesquiterpenoid, was used to treat various cancers and elemene exists as an essential oil mixture of ß-, γ- and δ-forms [9]. Zhu and coworkers demonstrated that ß-elemene is an effective drug for lung cancer treatment by inhibiting the activity of the PIsK/Akt/mTOR/p70S6K1 signaling pathway [10]. Moreover, clinical studies have shown that ß-elemene can inhibit the growth of glioblastoma cells, the commonest and the most malignant type of brain tumor, through the activation and over expression of glia maturation factor ß [11].

The sub-cellular distribution of enzymes involved in plant terpenoid biosynthesis has been studied for decades.

*Address correspondence to this author at the Biodiversity Research Center, Academia Sinica, 128 Academy Road, Section 2, Taipei 115, Taiwan;
Tel: 886-2-27871155; Fax: 886-2-27898711; E-mail: smchaw@sinica.edu.tw

1873-4316/12 $58.00+.00

doi:10.6342/NTU201601257

The enzymes involved in terpenoid biosynthesis are terpene synthases (TPS). In general, monoterpene synthases (MTPSs) use geranyl pyrophosphate (GPP) as a substrate. GPP is derived from the 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway in plastids [12]. Whereas, sesquiterpene synthases (STPSs) catalyze farnesyl pyrophosphate (FPP) to generate sesquiterpenoids that are synthesized by the mevalonate pathway in the cytosol [13, 14]. In plants and mammals, isoprenoid biosynthesis is known to occur in the mitochondria [15, 16]. Furthermore, TPSs were reported to localize to mitochondria and chloroplast in strawberry [17] and southern magnolia [18]. However, there has been no evidence whether MTPSs are specifically targeted to mitochondria.

The objectives of this study were to identify the volatile components of *Toona* leaflets and isolate and characterize the key genes responsible for their biosynthesis. Therefore, we used degenerate primers for MTPS and STPS to screen the cDNA pool of *Toona* leaflets. Here we report for the first time the isolation and analysis of a MTPS gene, *Tstps1*, and a STPS gene, *Tstps2*, from *Toona* that are responsible for (+) limonene synthesis and β-elemene synthesis in *Toona*, respectively. We investigated the expression patterns of these two genes in *Toona*, and characterized their functions and protein targeting *in vitro*. Our data are also the first to demonstrate that the transit-peptide of a MTPS can target specifically to mitochondria. *In situ* hybridization and staining of cross sections of rachises revealed the glandular cells of rachis being rich in TPS. We also analyzed TPS sequences of known representative angiosperms to elucidate the evolution of *Tstps1* and *Tstps2*.

MATERIALS AND METHODS

Plant Material and RNA Extraction

The seedlings of *Toona* were collected on the Academia Sinica campus and replanted in a greenhouse at Academia Sinica. Total RNA was isolated from developing and mature leaflets, bark, rachises and roots, separately, according to a modified RNA isolation protocol [19]. A developing leaflet was defined as < 6 cm long with a red-brown color. A mature leaflet was defined as ≥ 6 cm long with a green color (Supplement **1**).

Collection of Headspace Volatile Compounds

To analyze the volatile compounds from the developing and mature leaflets of *Toona*, we collected a fresh leaflet from each and put it in a glass vial. The developing and mature leaflets were defined in the previous paragraph. The headspace volatile compounds were collected by solid phase microextraction (SPME) (100 μm, PDMS) for 15 min, and analyzed by use of gas chromatography-mass spectrometry (GC-MS) with an HP-5 capillary column [20]. The GC injection port temperature is 150°C [21]. The oven temperature was programmed to increase from 40°C to 150°C at 10°C min^{-1}, then from 150°C to 300°C at 20°C min^{-1}. The final temperature was maintained for 1 min.

Design of Degenerate Primers for PCR Amplification of TPS Genes

No TPS gene was characterized from Meliaceae. Therefore, we used some TPS genes from Rutaceae as template because Meliaceae and Rutaceae are close in the phylogeny. Comparison of the amino acid sequences of MTPSs and STPSs from Rutaceae revealed highly conserved regions, and these sequences were used to design degenerate primers [22, 23]. For PCR amplification of MTPS genes, the forward and reverse primers were 5'-GCRTTRGAGGCAAGGTGGTTCAT-3' and 5'-CCCGCATCAGTGAATAKCTCAAGTTC-3', respectively. For PCR amplification of STPS genes, the primers were sesqui-dF2, 5'-GAYGAYAYDTATGATGCWTATGGMACMW-3' and sesqui-dR2, 5'- TKWRCRAYRTCWSYCATKCHMACR-3'. Partial sequences of the two forward sequences located at the TPS specific motif "DDXXD".

cDNA Expression in *E. coli* and Enzyme Purification

To characterize the *Tstps1* cDNA, its putative transit-peptide sequence was truncated to generate *Tstps1ΔN*. *Tstps1ΔN* and *Tstps2* were cloned into expression vectors pET21b and pET28a-CBP-Factor Xa, respectively [24]. The 3' ends of the sequences were fused to a 6-His tag downstream of an inducible *lac* promoter. The resulting plasmids were transformed into BL21(DE3)-competent cells according to the manufacturer's protocol (Invitrogen, Carlsbad, CA). The recombinant proteins were purified following the method described in Tholl and coworkers [25].

Enzyme Assays and Product Identification

The recombinant proteins were purified using a nickel affinity column and assayed with 100 μM GPP, FPP and geranylgeranyl pyrophosphate (GGPP) used as substrates in a total volume of 500 μl TPS assay buffer (25 mM HEPES, 10 mM $MnCl_2$, 100 mMKCl, 10 mM $MgCl_2$, 5 mM DTT, 10% glycerol; pH 7.4). The reaction mixtures were incubated at 30°C for 2 h and the reaction products were collected by SPME. The collected compounds were analyzed by GC-MS with an HP-5 column. For TsTPS1 reaction, the oven temperature was programmed to increase from 40°C to 150°C at 10°C min^{-1}, and then from 150°C to 300°C at 20°C min^{-1}. The final temperature was maintained for 1 min. For TsTPS2 reaction, the inlet temperature of GC was 150°C [21] and the oven temperature was increased from 40°C to 180°C at 10°C min^{-1}, and the final temperature was kept for 5 min.

Phylogenetic Analysis

The deduced protein sequences of *Toona* TPS genes were aligned with elucidated TPS genes (Supplement **2**) [18, 26, 27]. Construction of phylogenetic trees involved the neighbor-joining method [28] in PAUP (v4.0b10) [29] and the Bayesian posterior probability methods in MrBayes (v3.1.2) [30]. The TPS sequences on our phylogenetic tree were labeled with the subfamilial names proposed by Bohlmann and coworkers [13], and Trapp and Croteau [26].

Anatomy of Rachis and Detection of Lipids and Proteins

Sections of rachises were fixed by submersion in freshly prepared FAA (50% ethanol, 10% formalin, 5% acetic acid) for 8 h. After a standard dehydration procedure, the tissues were embedded in paraffin, cut into 8 μm sections on a rotary microtome and placed sequentially onto slides. Finally, the sections were stained with Safranin/Fast green or Coomassie blue.

doi:10.6342/NTU201601257

In Situ Hybridization

Both antisense and sense RNA were labeled with digoxigenin (DIG) by use of the Roche DIG RNA Labeling kit (catalog no. 1 175 025). An approximate 400 bp fragment of *Tstps1* was amplified from the cDNA pool with gene-specific primers (TsT1F1, 5'-CGCACCCTATCTTGCTCGAGTTTG-3' and TsT1R4, 5'-GTCAACTCCTCTATCTTTACGGGCGTC-3'). For cell permeabilization and dehydration before hybridization and immunological detection, the protocol of Kramer was used [31].

Genomic DNA Amplification and Cloning

Genomic sequences of *Tstps1* and *Tstps2* were amplified by PCR with conserved primers for *Tstps1*, forward 5'-CCCAAGCTTCGAAGATCTGCCAA-3' and reverse 5'-TAAACTATGCGGCCGCGGTGGACAAGGAATGGGATT-3'; and *Tstps2*, forward 5'-ATGTCTGTCCCAGTTTCACAGATTCCG-3' and reverse 5'-CACTGGAATTGGATCAATTAGCAATGAAG-3'. The amplified fragments were cloned into the yT&A vector (Yeastern Biotech Co., Taipei).

Localization of *Tstps1* and *Tstps2* Transcripts

Total RNA was extracted from developing and mature leaflets, bark, rachises and roots of *Toona*. Total RNA (1 μg) was reverse transcribed by use of the Smart cDNA library Construction Kit (Clontech, Mountain View). PCR amplification involved the primers for *Tstps1*, forward, 5'-CGCACCCTATCTTGCTCGAGTTTG-3' and reverse, 5'-GTCAACTCCTCTATCTTTACGGGCGTC-3'; *Tstps2*, forward, 5'-TGCAAGAGACAGATTGGTTGAGTGCT-3', and reverse, 5'-GCATAGTGAACTCGGTATGACCATCCTT-3'; and actin, forward, 5'-TGYTGGAYTCTGGTGATGGTGT-3', and reverse, 5'-GCRACMACCTTRATCTTCATGCT-3'. Actin was used as an equal loading control in gel analysis.

TsTPS1 and TsTPS2 Transient Expression Assays

The ChloroP program (http://www.cbs.dtu.dk/services/ChloroP/) was used to predict a putative transit-peptide sequence. The putative transit sequence of TsTPS1 was amplified using the primer pair TsY1-GFPF (5'-TAAACTATTCTAGATGATGGCTTCTCACGTGCTAGC-3') and TsY1-GFPR (5'-TAAACTATGGATCCCGGAGACGAAGCCAT-3'). That of TsTPS2 was amplified with TsY5-GFPF (5'-TAAACTATTCTAGATGATGTCTGTCCCAGTTTCA-3') and TsY5-GFPR (5'-TAAACTATGGATCCCCACTGGAATTGGATC-3'). The amplified fragments were each cloned into the p326GFP vector at the *Xba*I and *Bam*HI restriction sites [32]. Constructs were transfected into *Arabidopsis thaliana* protoplasts by a PEG-mediated method. Transient expression of the GFP fusion proteins was observed after 16 h under a Zeiss confocal laser scanning microscope (LSM 510 META NLO DuoScan). MitroTracker orange (Invitrogen, Lot. 445400, Eugene) and the p326GFP vector alone were used as markers for identifying the mitochondrial and cytosolic compartments, respectively.

RESULTS

Analysis of Volatiles in Leaflets

Fig. (**1**) shows that the developing leaflets released a greater number of compounds, as well as higher concentrations of compounds, than the mature leaflets. The developing leaflets showed 7 volatile sesquiterpenoids: β-elemene, β-caryophyllene, γ-elemene, Guaiene, α-caryophyllene, β-selinene and α-selinene, with the β-elemene being the dominant one. These data suggest that sesquiterpenoids are be the

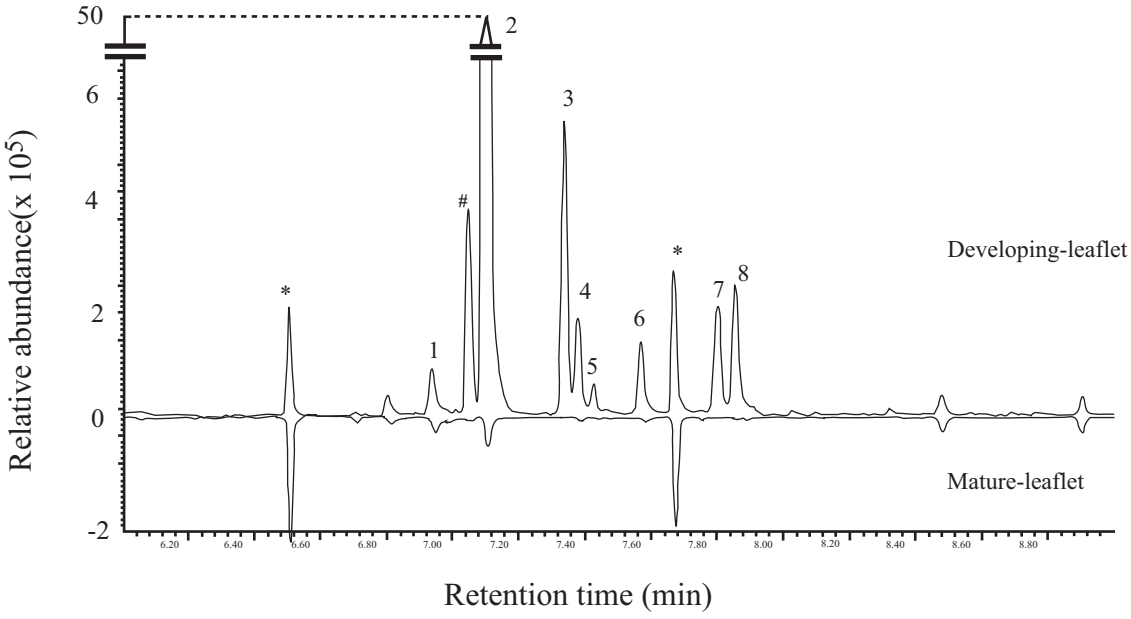

Fig. (1). Gas chromatography-mass spectrometry (GC-MS) analysis of headspace-collected volatile compounds from the leaflets of *Toona*. A developing and a mature leaflet were used. Peak 1, butanoic acid; 2, β-elemene; 3, β-caryophyllene; 4, γ-elemene; 5, Guaiene; 6, α-caryophyllene; 7, β-selinene; 8, α-selinene; #, a stereoisomer of β-elemene; *, column bleeds.

doi:10.6342/NTU201601257

major volatiles in the developing leaflets. In contrast, sesquiterpenoid levels were low or below detectable limits in mature leaflets. However, β-elemene was still the dominant volatiles in mature leaflets.

Characterization and Cloning of *Tstps1* and *Tstps2*

We investigated the biosynthesis of the main compounds of volatiles by cloning two TPS genes in *Toona*, *Tstps1* (Accession number AB303572) and *Tstps2* (Accession number AB509224). The full-length *Tstps1* clone was 1,788 bp and encoded a predicted protein of 595 amino acids with a calculated molecular mass of about 68.82 kDa (Fig. **2a**). The full-length *Tstps2* clone was 1,671 bp and encoded a predicted protein of 556 amino acids with a calculated molecular mass of 64.77 kDa. *Tstps1* was predicted by the Target P program (http://www.cbs.dtu.dk/services/TargetP/) to have a transit-peptide sequence in its N-terminal region. Thus, the 171 bp transit-peptide sequence was removed from upstream of the $RR(X)_8W$ motif to create *Tstps1ΔN*; and this 1,617 bp sequence, predicted to encode a protein of 538 amino acids, was cloned into the yT&A vector (Yeastern Biotech Co., Taipei).

Both TsTPS1 and TsTPS2 possessed two conserved motifs: $RR(X)_8W$ and DDXXD (Fig. **2a**). Alignment of the genomic DNA sequences of *Tstps1* (AB730584) and *Tstps2* (AB730585) and the corresponding cDNAs revealed that *Tstps1* contains six exons and five introns and that *Tstps2* contains seven exons and six introns (Fig. **2b**). On the basis of intron-exon organization, both *Tstps1* and *Tstps2* are assignable to the class III TPS according to the classification scheme of Trapp and Croteau [26]. However, *Tstps1* appears to be atypical, having lost the XI intron (Table **1**). The phases of intron insertion in both *Tstps1* and *Tstps2* were similar to those of TPS genes in other representative angiosperms (Table **1**), which suggests a common origin of the two genes to other angiosperms'. Our phylogenetic tree further indicates that TsTPS1 and TsTPS2 were nested in the TPS-b and TPS-a subfamilies, respectively (Fig. **3**) [13, 27, 33].

Functional Characterization of TsTPS1ΔN and TsTPS2

We used GPP, FPP, and GGPP as substrates to determine the functions of TsTPS1ΔN and TsTPS2. It turned out that both enzymes are capable of using only GPP and FPP, respectively. With GPP being used as a substrate, TsTPS1ΔN produced a major product, (+) limonene, and 3 side products: β-pinene, β-myrcene, terpinolene (Fig. **4a**). With FPP being as a substrate, TsTPS2 formed one major product, β-elemene, and 5 side products: β-caryophyllene, α-caryophyllene, germacrene D, α-selinene and δ-cadinene (Fig. **4b**).

Differential Expression of *Tstps1* and *Tstps2* mRNA in Plant Tissues

RT-PCR was used to detect *Tstps1* and *Tstps2* transcripts in the developing and mature leaflets, bark, rachises, and roots (Fig. **5**). The expression of both genes was higher in the roots than in the developing and mature leaflets. Notably, the mRNA level of *Tstps1* was higher in rachises than in other tissues, and *Tstps2* expression was barely detectable in rachises.

Terpenoid Storage in *Toona* Tissues

Microscopic observation of the cross sections of rachises revealed a high density of glandular cells (about 10 per mm^2) not seen in other tissues. The stained cross sections of rachises showed accumulation of both lipids and proteins in the glandular cells (Fig. **6a** and **6b**). *In situ* hybridization with a DIG-labeled antisense RNA probe also showed specific accumulation of *Tstps1* mRNA in the glandular cells (Fig. **6c**). In contrast, *in situ* hybridization with a sense RNA probe as a control resulted in little staining (Fig. **6d**). Thus, both *Tstps1* mRNA and other TPS gene products were accumulated in the rachises. Scanning electron microscopy further revealed the surfaces of rachises with numerous globular glandular cells (Fig. **6e**).

Localization of TsTPS1 and TsTPS2

We predicted that the N terminus of TsTPS1 had a 39 amino acid transit-peptide but that TsTPS2 did not contain any transit-peptide sequence. To determine the localization of the predicted TsTPS1 transit-peptide and TsTPS2, we fused the predicted TsTPS1 transit-peptide sequence (encoding 39 residues) and the full-length TsTPS2 to the N terminus of a GFP reporter gene. The control GFP reporter protein was localized to the cytoplasm (Fig. **7a**); and chloroplasts and mitochondria were visualized by chloroplast autofluorescence and MitoTracker staining, respectively. The merged images suggest that the transit-peptide of TsTPS1 was specifically localized to mitochondria rather than plastids (Fig. **7b**) and that TsTPS2 was localized to cytoplasm (Fig. **7c**). These results are consistent with previous studies that the TPS-a subfamily don't normally have transit peptides [18].

DISCUSSION

Volatile Compounds in Leaflets of *Toona*

Fig. (**1**) indicates that tangy scent from the developing leaflets of *Toona* was the result of a combination of sesquiterpenoids and β-elemene was the dominant volatiles. Our *in vitro* analysis further suggested that β-elemene was the major product of TsTPS*2* (Fig. **4b**). Apparently, TsTPS2 is responsible for the production of the major volatile emitted from the leaflets of *Toona*. Our headspace analysis revealed that the collected volatiles of developing leaflets are diverse and abundant than those of mature leaflets (Fig. **1**). The epidermis and cuticles are thicker in the mature leaflets than in the developing leaflets. They may act as barriers to the emission of water-insoluble terpenoids [34].

Mu and coworkers [4], and our data suggested consistently that sesquiterpenoids are the major volatiles in *Toona* leaves. However, they identified 45 volatile compounds and the highest content component was the sesquiterpene *trans*-caryophyllene (21.42%) [4]. In contrast, our headspace data showed that the β-elemene is dominant in both developing and mature leaflets. Nevertheless, the relative content of β-elemene was low (1.981%) in Mu and coworker's GC-MS analysis. Of note, before conducting GC-MS assays Mu and coworkers not only ground the leaves to fine power but also used "microwave-assisted extraction". Moreover, it is worthy of mentioning that the plant samples of Mu and coworkers and

doi:10.6342/NTU201601257

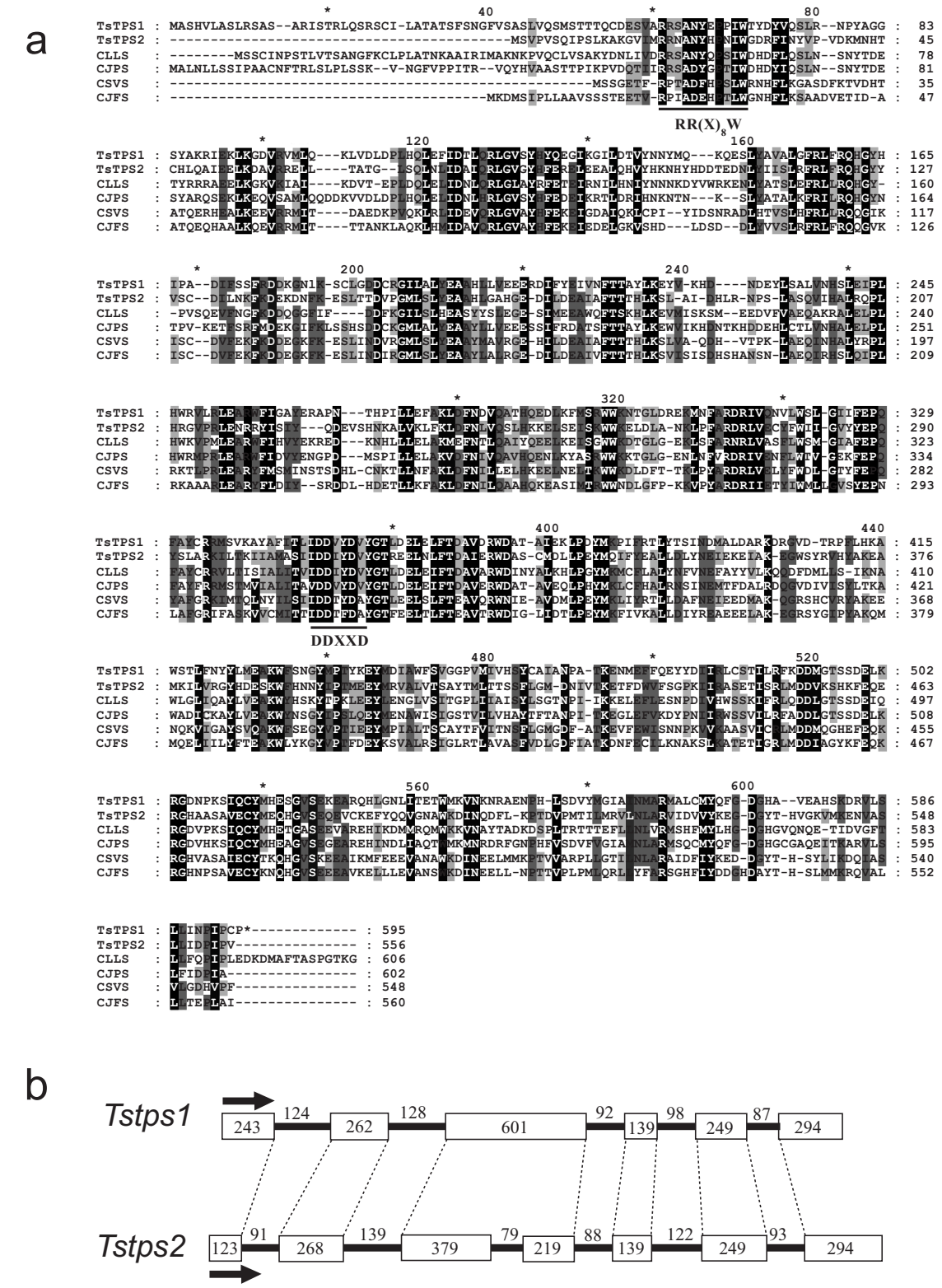

Fig. (2). Alignment of deduced amino acid sequences of *Tstps1, Tstps2,* and four terpene synthase genes from Rutaceae, and structures. **a)** CLLS, *Citrus limon* (+)-limonene synthase (AF5142891); CJPS, *Citrus jambhiri* β-pinene synthase (AF739331); CSVS, *Citrus sinensis* valencene synthase (AF4411241); CJFS, *Citrus junos* (E)-β-farnesene synthase (AAK542791). Dashes indicate gaps inserted for optimal alignment. Residues conserved among at least four of the six sequences are by a gray background, and those conserved in all six sequences are by a black background. The $RR(X)_8W$ and DDXXD motifs are underlined. **b)** The structures of the *Tstps1* and *Tstps2* genes. The open boxes and solid lines denote exons and introns, respectively. The arrows indicate transcription direction of the genes.

doi:10.6342/NTU201601257

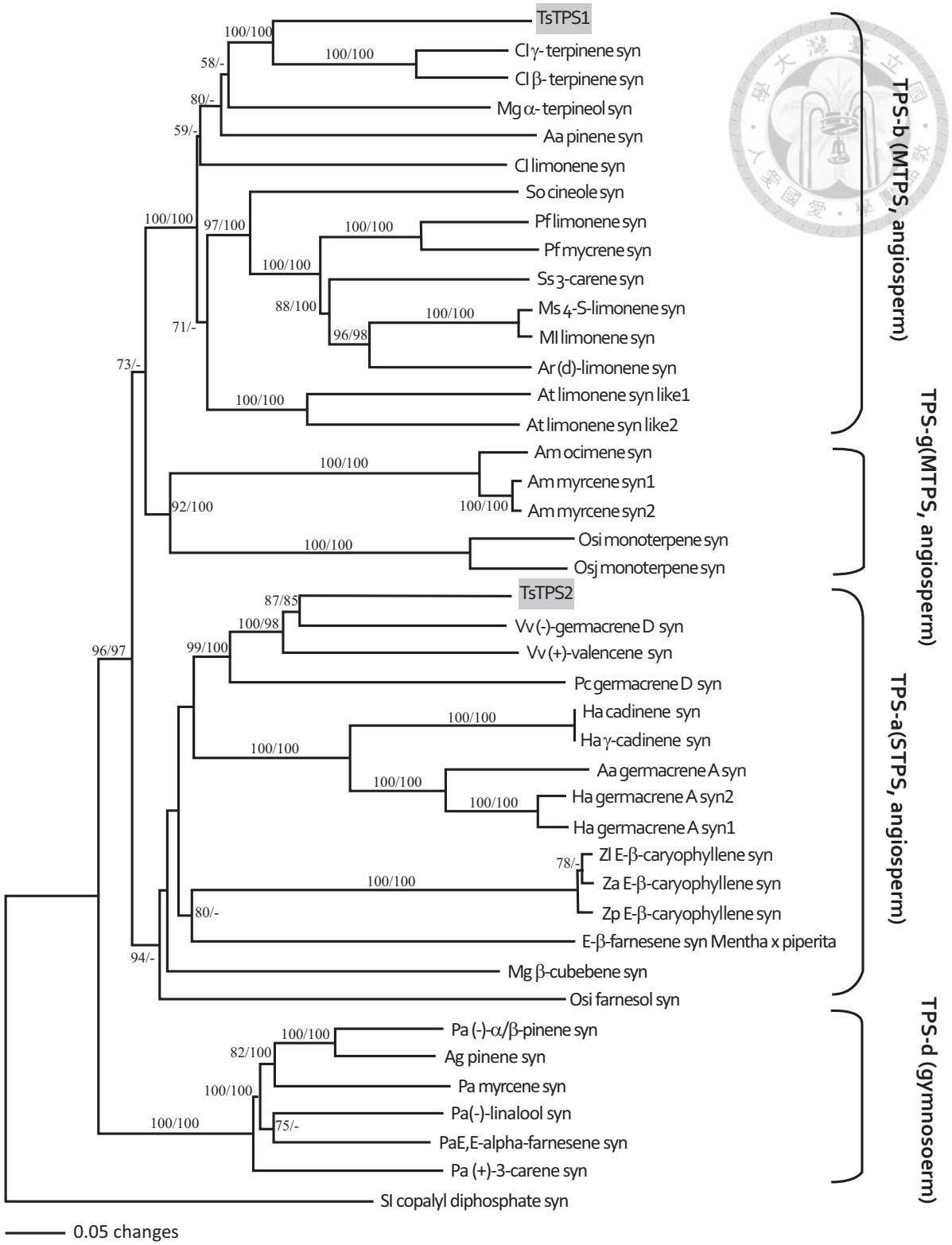

Fig. (3). A combined phylogenetic tree of plant TPS reconstructed by the neighbor-joining (NJ) and Bayesian methods. The numbers at the nodes denote NJ bootstrap values (before slash) and posterior probabilities of the Bayesian method (after slash). Groupings of genes were named according to Bohlmann and coworkers [12]. The gray boxes indicate *Toona* genes from this study. Aa, *Artemisia annua*; Ag, *Abies-grandis*; Am, *Antirrhinum majus*; Ar, *Agastache rugosa*; At, *Arabidopsis thaliana*; Cl, *Citrus limon*; Ha, *Helianthus annuus*; Mg, *Magnolia grandiflora*; Ml, *Mentha longifolia*; Ms, *Mentha spicata*; Os, *Oryza sativa*; Osj, *Oryza sativa* cv. *japonica*; Pa, *Picea abies*; Pc, *Pogostemon cablin*; Pf, *Perilla frutescens*; Sl, *Solanum lycopersicum*; So, *Salvia officinalis*; Ss, *Salvia stenophylla*; Vv, *Vitis vinifera*; Za, *Zea mays*; Zl, *Zea luxurians*; Zp, *Zea perennis*.

doi:10.6342/NTU201601257

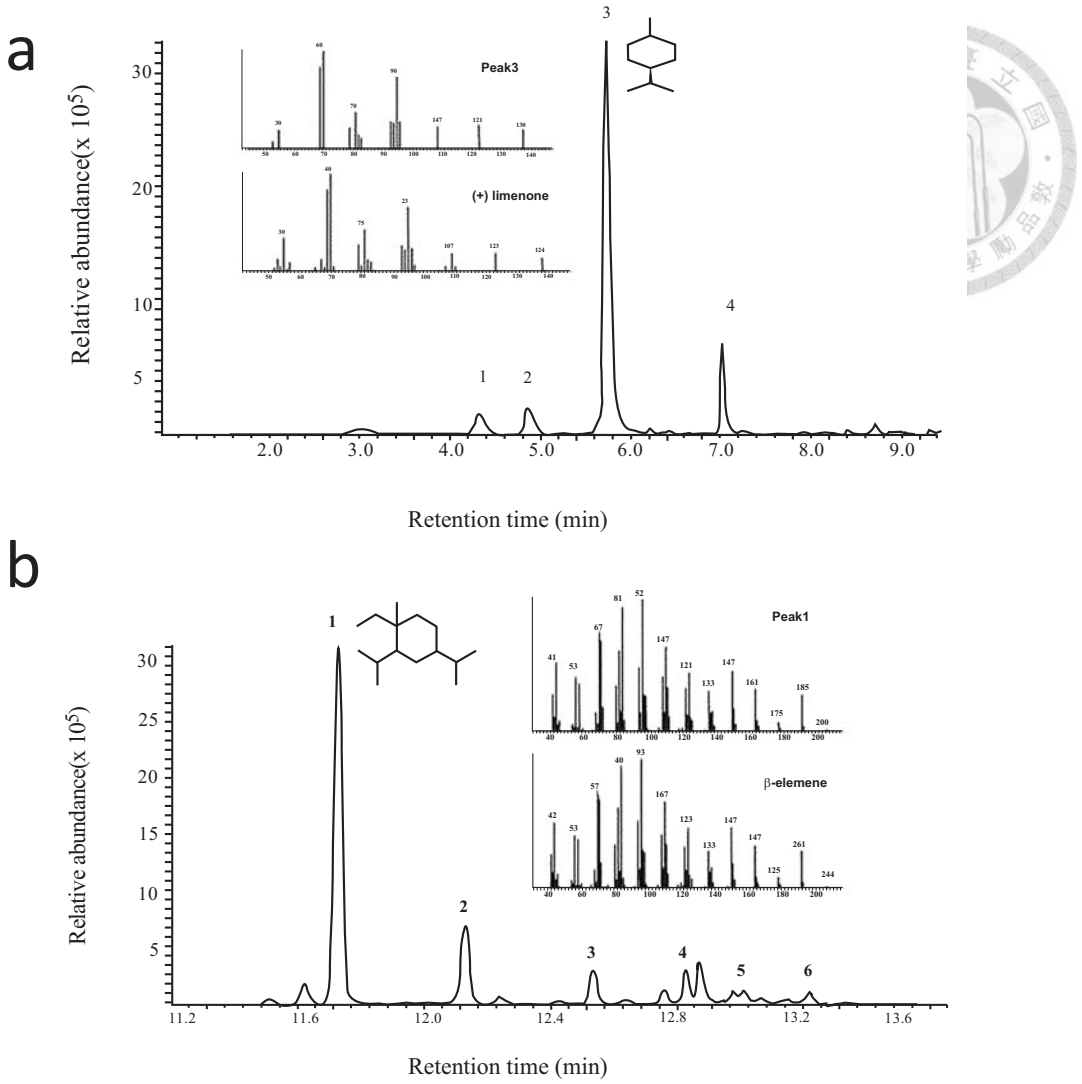

Fig. (4). GC-MS analysis of the terpene products produced by TsTPS1 and TsTPS2. **a)** GC-MS profile of products formed by TsTPS1ΔN with geranyl diphosphate used as a substrate. The main product was (+) limonene (peak 3). Other products included peak 1, β-pinene; 2, β-myrcene; 4, terpinolene. **b)** GC-MS profile of products formed by recombinant TsTPS2 with farnesyl diphosphate used as a substrate. The main product was β-elemene (peak 1). Other products included peak 2, β-caryophyllene; 3, α-caryophyllene; 4, germacrene D; 5, α-selinene; 6, δ-cadinene. The peaks labeled in the GC profiles were identified by comparison of their mass spectra with those in the WILEY275 library and authentic standards.

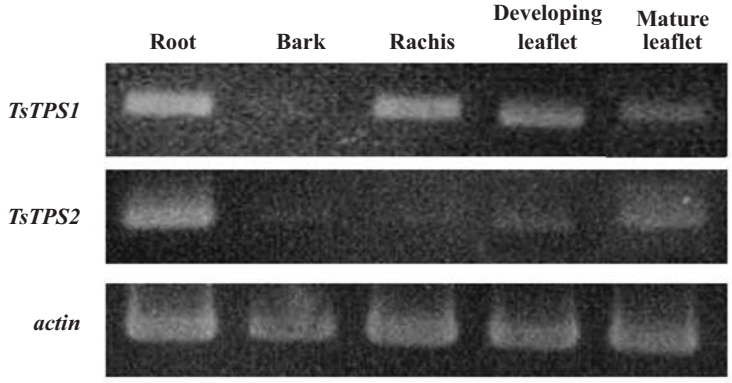

Fig. (5). Gene expression of *Tstps1* and *Tstps2* in *Toona* tissues as assayed by RT-PCR. Both gene expression patterns were examined in the root, bark, rachis, developing leaflets, and mature leaflets. Actin was used as a loading control.

doi:10.6342/NTU201601257

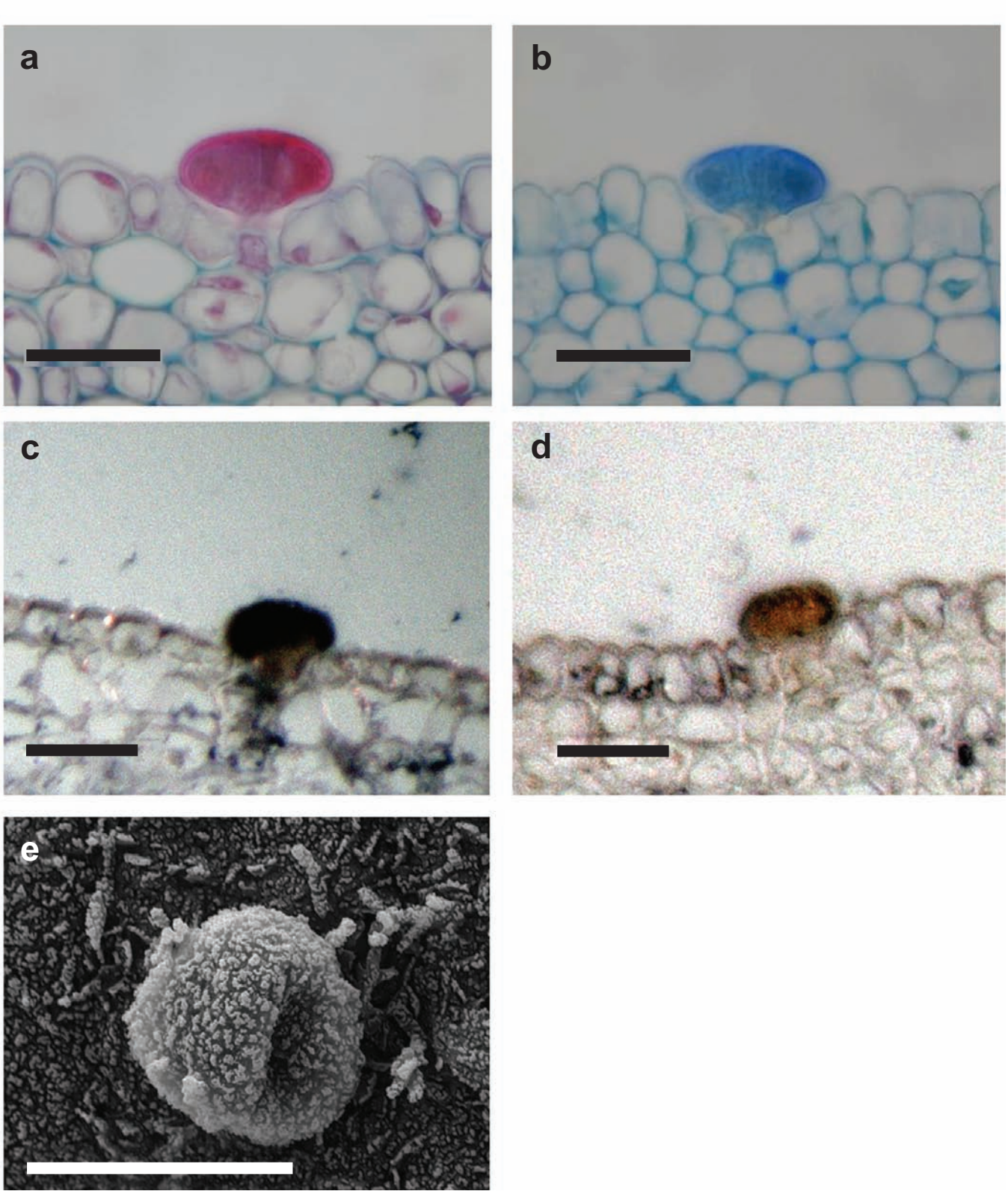

Fig. (6). Glandular cells on the rachises of *Toona*. **a)** A glandular cell positively stained with Safranin/Fast green indicating the presence of lipids. **b)** A glandular cell positively stained with Coomassie blue indicated the presence of proteins. **c)** *In situ* localization of *Tstps1* mRNA by hybridization with an antisense RNA probe. **d)** *In situ* hybridization with sense RNA probe as a negative control. **e)** Scanning electron microscopy image of a singular glandular cell. Scale bars = 25 μm.

doi:10.6342/NTU201601257

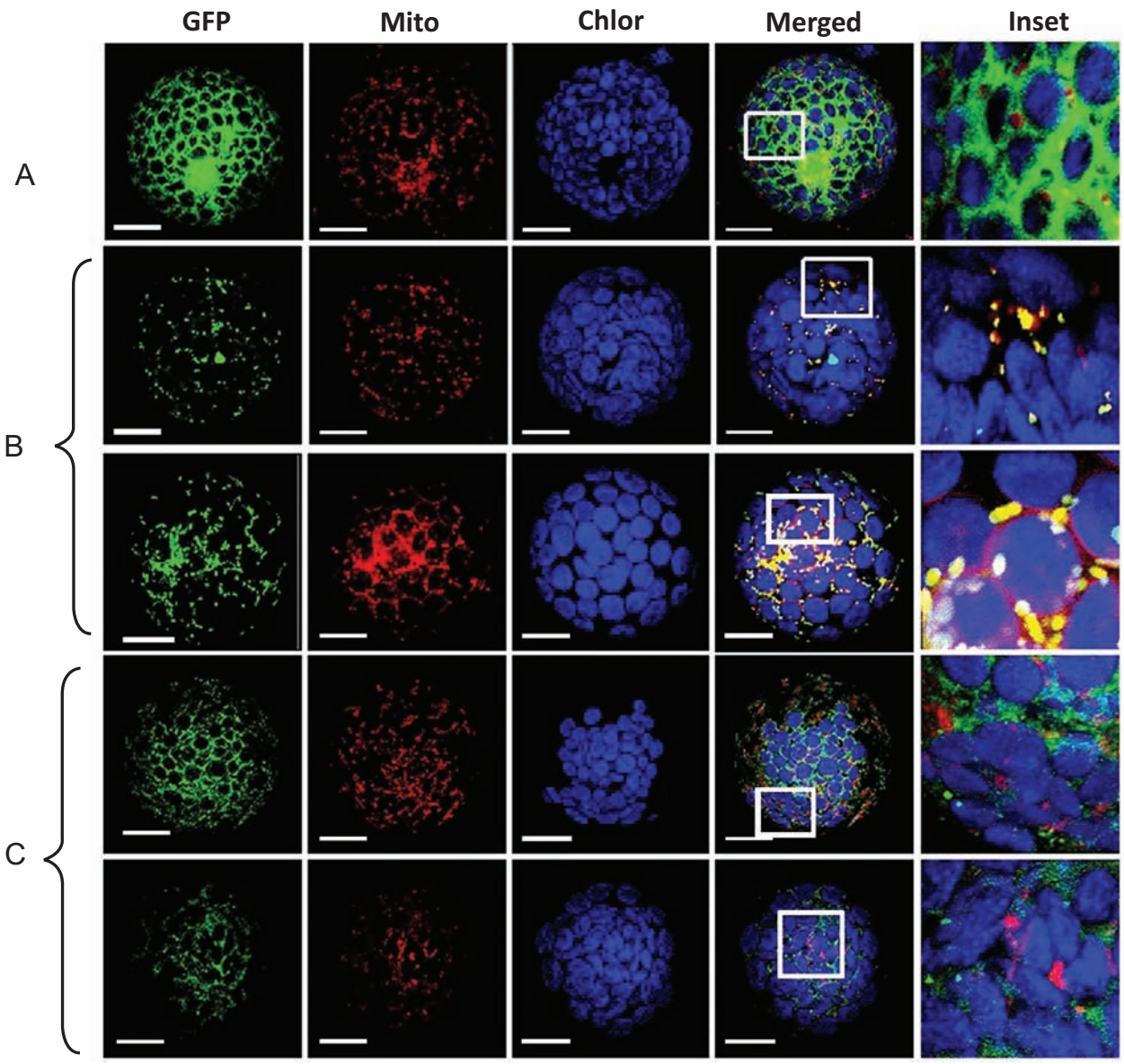

Fig. (7). Transient expression assays of TsTPS1 and TsTPS2 in *Arabidopsis* leaf protoplasts. **a)** GFP reporter protein alone as a control. **b)** The first 39 amino acids of TsTPS1 **c)**. The full-length TsTPS2 fused to the N terminus of a GFP reporter protein. The above constructs were transfected into *Arabidopsis* leaf protoplasts and analyzed by a confocal laser scanning microscope. GFP fluorescence is shown in the "GFP" column. Mitochondria were stained with red MitoTracker (Invitrogen) ("Mito" column), and chlorophyll auto fluorescence is the "Chlor" column. "Merged" column shows combined GFP, Mito and Chlor signals. White boxes in the "Merged" column are enlarged and shown in the "Inset" column. Scale bars = 15 μm.

ours were grown in two distinct environments. It has been suggested that environmental conditions such as temperature, day length, and light influence quantitative compositions of plant compound mixtures like volatiles and essential oils [35]. The location and average temperature of Shandong where the Mu and coworkers collected their samples is about 35 degree North latitude and 11^{o}C, respectively, whereas our samples were collected in Taipei, which is located at 25 degree North latitude with the average temperature being 23^{o}C. Taken together, these factors may contribute to the contrast differences in compound components and their relative contents in both studies. Further comparisons, in particular from the aspect of plant population genetics between the two sampled places are required before conclusions could be made.

We did not detect the main product of TsTPS1, (+) limonene, in the headspace volatiles of *Toona* leaflets. Lee and Chappell observed that a MTPS is targeted to both mitochondria and chloroplasts [18]. However, they did not clarify whether the MTPS is capable of monoterpeniod biosynthesis in mitochondria. Previously there was no *in vivo* experiment demonstrating that MTPS can function inside mitochondria and whether GPP, a substrate of MTPS, can be stored in mitochondria. Moreover, limitation of the substrate in mitochondria may reduce the formation of volatile compounds

doi:10.6342/NTU201601257

Table 1. Comparison of Intron Phases among *Tstps1*, *Tstps2*, and Four Terpene Synthase Genes From Other Angiosperms.

Gene[a]	IN	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV
Tstps1	5	–	–	0	–	–	–	–	1	–	–	–	2	0	0
Tstps2	6	–	–	0	–	–	–	–	1	–	–	2	2	0	0
TS (Mg)	6	–	–	0	–	–	–	–	1	–	–	2	2	0	0
LS (Pf)	6	–	–	0	–	–	–	–	1	–	–	2	2	0	0
CS (At)	6	–	–	0	–	–	–	–	1	–	–	2	2	0	0
LS (At)	6	–	–	0	–	–	–	–	1	–	–	2	2	0	0
Consensus	6	–	–	0	–	–	–	–	1	–	–	2	2	0	0

[a], Roman numerals denote intron positions. Intron phase numbers 0, 1, and 2 indicate the intron insertion before the first, second, or third codon, respectively. The "IN" means the total number of introns. A dash indicates that an intron is absent in this region. TS (Mg): α-terpineol synthase from *Magnolia grandiflora*; LS (Pf): limonene synthase from *Perillafrutrscens*; CS (At): cadinene synthase from *Arabidopsis thaliana*; LS (At): limonene from *A. thaliana*. The elucidated angiosperm data were adopted from Trapp and Croteau [26] and Lee and Chappell [18].

[36]. The reduced amount of volatiles emitted from the leaflets of *Toona* is likely due to lacking the substrate GPP in mitochondria.

Tstps1 and *Tstps2* Yield Multiple Products and Highly Express in Roots

Our *in vitro* assays showed that both TsTPS1 and TsTPS2 produced multiple products, the major one being (+) limonene and β-elemene, respectively. A single TPS can synthesize multiple products since the enzyme can drive the multiple reactive mechanisms, a complex pattern of cyclizations, hydride shifts, and substitutions [37]. Examples include the TPSs from lemon [38], *Arabidopsis* [39] and sandalwood [40]. In addition, previous study showed that β-elemene is often a breakdown product of germacrene A because of the high temperature during GC process [21]. Therefore, we used the temperature below the breakdown point of germacrene A and compared the mass spectra of the main product of TsTPS2 with the authentic standard. In consistent with data from headspace analyses, β-elemene was found to be the dominant element.

Originally, we assumed the majority of volatile compounds, mostly terpenoids, are emitted from *Toona* leaflets. However, relatively higher levels of *Tstps1* and *Tstps2* transcripts were detected in roots than leaflets when plants were grown under normal conditions (Fig. **5**). An earlier observation reported massive amounts of secondary metabolites emitted from the roots of plants [41]. Terpenoids that defend against microbes (*i.e.*, bacteria, fungi and some other pathogens) were found released from the roots of *Arabidopsis* [42]. Nonetheless, whether the transcripts of *Tstps1* and *Tstps2* are constitutive and stimulated by soil or endosymbiotic microbes merit further investigation.

TsTPS1 is Stored in Glandular Cells and Specifically Targeted to Mitochondria

Many volatile monoterpenoids and sesquiterpenoids are produced and stored in the surface glands of plant vegetative tissues [43, 44]. In this study, only the rachises showed glandular cells filled with lipids and proteins with abundant *Tstps1* mRNA transcripts (Fig. **6**). No similar granular cells were found in the leaflets, barks, or roots of *Toona*. Previous studies showed that generally terpenoids are synthesized in the epidermal cells of leaflets or flowers, and volatiles can be released from these cells directly into the atmosphere [45]. Because of the high level of volatile emission from *Toona* leaflets, volatile terpenoids may be continually released from the stomata of leaflets rather than accumulated or stored in the leaflet tissues.

Previous studies have shown that MTPS transit-peptides target proteins to both chloroplasts and mitochondria. Examples include Mg17 in magnolias [18] and FaNES2 in strawberry [17]. Notably, Mg17 and FaNES2 are functionally different in that they are capable of generating monoterpenoids and sesquiterpenoids using GPP or FPP as substrates. However, TsTPS1 can only use GPP as the substrate and is specifically targeted to mitochondria, which is unprecedented among elucidated MTPSs. It has been shown that mitochondria have an isoprenoid biosynthetic pathway involved in the production of primary metabolites such as sterols and ubiquinones [16, 17]. If this is the case, MTPS may compete with other primary metabolism enzymes in mitochondria for the common substrate GPP or FPP. As a result, we hypothesize that the quantity of monoterpenoid products should be limited by the diverse localizations of enzymes in subcellular compartments, the abundance of substrates, and the affinity of enzymes.

Evolution of TsTPS1 and TsTPS2

Among the Class III TPS genes, *Tstps1* is atypical in that it lacks one intron [26]. Our comprehensive comparisons of the intron phases of previously elucidated MTPS & STPS genes conclude that TPS genes are highly conserved in terms of intron positions [18, 26], which suggests that TPS introns have existed at least since the common ancestor of seed plants and that some introns have been lost independently from diverse plant lineages during evolution. Furthermore, we reconstructed phylogenetic trees using NJ and Bayesian methods. TsTPS1 and TsTPS2 were aligned along with

doi:10.6342/NTU201601257

known TPSs of seed plants [18, 26, 27]. Fig. (**3**) suggests that TsTPS1 and TsTPS2 fall within the group TPS-b and TPS-a, respectively. These groupings suggested that despite differences in exon-intron organizations, the protein-coding sequences and functional domains of TsTPS1 and TsTPS2 are quite conserved. Moreover, our phylogenetic tree implies that the complex mono-, sesqui-, and di-terpene synthesis genes existed before the deep split of gymnosperms and angiosperms [13, 27]. Hence, TPSs must have played a crucial role in the development and growth of seed plants for 300 million years.

Further study to better understand the physiological role(s) of TsTPS1 in mitochondria is underway. We are also carrying experiments on the biosynthesis of the rest sesquiterpenoids that make up such a high proportion of the volatiles of *Toona* leaves.

SUPPLEMENTARY MATERIALS

Supplementary material is available on the publisher's web site along with the published article.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflicts of interest.

ACKNOWLEDGEMENTS

We thank Choun-Sea Lin and Fu-Hui Wu for assistance with the GFP assay and Shu-Chen Shen for assistance with confocal microscopy and data analysis. We thank Leek Chun-Yen Teng, Wen-Ke Huang, and Jen-Pan Huang for critical comments on an early version of the manuscript. We thank the three anonymous reviewers for their critical reading and helpful suggestions. This study was supported by a research grant from the Biodiversity Research Center, Academia Sinica and in part by the grant, NSC93WIA0100340, from National Science Council as well as the Investigator's Award of Academia Sinica to S.- M. C.

ABBREVIATIONS

FPP	=	Farnesyl pyrophosphate
GPP	=	Geranyl pyrophosphate
GGPP	=	Geranylgeranyl pyrophosphate
MC-MS	=	Gas chromatography-mass spectrometry
MEP	=	2-C-methyl-D-erythritol 4-phosphate
MTPS	=	Monoterpene synthase
SARS	=	Severe acute respiratory syndrome
SPME	=	Solid phase microextraction
STPS	=	Sesquiterpene synthase
Toona	=	*Toona sinensis*
TPS	=	Terpene synthase

REFERENCES

[1] Edmonds, J.M.; Staniforth, M. Toona sinensis (Meliaceae). *Curtis's Bot Mag.*, **1998**, *15*, 186-193.

[2] Chang, H.L.; Hsu, H.K.; Su, J.H.; Wang, P.H.; Chung, Y.F.; Chia, Y.C.; Tsai, L.Y.; Wu, Y.C.; Yuan, S.S. The fractionated *Toona sinensis* leaf extract induces apoptosis of human ovarian cancer cells and inhibits tumor growth in a murine xenograft model. *Gynecol. Oncol.*, **2006**, *102*, 309-314.

[3] Chen, C.J.; Michaelis, M.; Hsu, H.K.; Tsai, C.C.; Yang, K.D.; Wu, Y.C.;Cinatl, J.; Doerr, H.W. *Toona sinensis* Roem tender leaf extract inhibits SARS coronavirus replication. *J. Ethnopharmacol.*, **2008**, *120*,108-111.

[4] Mu, R.; Wang, X.; Liu, S.; Yuan, X.; Wang, S.; Fan, Z. Rapid determination of volatile compounds in *Toona sinensis* (A. Juss.) Roem. By MAE-HS-SPME followed by GC-MS. *Chromatographia*, **2007**, *65*, 463-467.

[5] Bohlmann, J.; Keeling, C.I. Terpenoid biomaterials. *Plant J.*, **2008**, *54*, 656-669.

[6] Pichersky, E.; Gershenzon, J. The formation and function of plant volatiles: perfumes for pollinator attraction and defense. Current opinion in plant biology **2002**, *5*, 237-243.

[7] Dudareva, N.; Pichersky, E.; Gershenzon, J. Biochemistry of plant volatiles. *Plant physiol.*, **2004**, *135*,1893-1902.

[8] Kessler, A.; Baldwin, I.T. Plant responses to insect herbivory: the emerging molecular analysis. *Annu. Rev. Plant Biol.*, **2002**, *53*, 299-328.

[9] Chen, S.L.; You, J. ; Wang, G.J. Supercritical fluid extraction of beta-elemene under lower pressure. *Se pu*, **2001**, *19*, 179-181.

[10] Zhu, T.; Xu, Y.; Dong, B.; Zhang, J.; Wei, Z.; Yao, Y. Beta-elemene inhibits proliferation of human glioblastoma cells through the activation of glia maturation factor beta and induces sensitization to cisplatin. *Oncol. Rep.*, **2011**, *26*, 405-413.

[11] Liu, J.; Hu, X.J.; Jin, B.; Qu, X.J.; Hou, K.Z.; Liu, Y.P. Beta-Elemene induces apoptosis as well as protective autophagy in human non-small-cell lung cancer A549 cells. *J. Pharm. Pharmacol.*, **2012**, *64*, 146-153.

[12] Lichtenthaler, H.K. The 1-Deoxy-D-Xylulose-5-phosphate pathway of isoprenoidbiosynthesis in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, **1999**, *50*, 47-65.

[13] Bohlmann, J.; Meyer-Gauen, G.; Croteau, R. Plant terpenoid synthases: molecular biology and phylogenetic analysis. *Proc. Natl. Acad. Sci. USA*, **1998**, *95*, 4126-4133.

[14] Pichersky, E.; Noel, J.P.; Dudareva, N. Biosynthesis of plant volatiles: nature's diversity and ingenuity. *Science*, **2006**, *311*,808-811.

[15] Cunillera, N.; Boronat, A.; Ferrer, A. The *Arabidopsis thaliana* FPS1 gene generates a novel mRNA that encodes a mitochondrial farnesyl-diphosphate synthase isoform. *J. Biol. Chem.*, **1997**, *272*, 15381-15388.

[16] Martin, D.; Piulachs, M.D.; Cunillera, N.; Ferrer, A.; Belles, X. Mitochondrial targeting of *farnesyl diphosphate* synthase is a widespread phenomenon in eukaryotes. *Biochim. Biophys. Acta*, **2007**, *1773*, 419-426.

[17] Aharoni, A.; Giri, A.P.; Verstappen, FW.; Bertea, C.M.; Sevenier, R.; Sun, Z.; Jongsma, M.A.; Schwab, W.; Bouwmeester, H.J. Gain and loss of fruit flavor compounds produced by wild and cultivated strawberry species. *Plant Cell*, **2004**, *16*, 3110-3131.

[18] Lee, S.; Chappell, J. Biochemical and genomic characterization of terpene synthases in *Magnolia grandiflora*. *Plant physiol.*, **2008**, *147*, 1017-1033.

[19] Kolosova, N.; Sherman, D.; Karlson, D.; Dudareva, N. Cellular and subcellular localization of S-adenosyl-L-methionine:benzoic acid carboxyl methyltransferase, the enzyme responsible for biosynthesis of the volatile ester methylbenzoate in snapdragon flowers. *Plant physiol.*, **2001**, *126*, 956-964.

[20] Deng, C.; Song, G.; Hu, Y. Application of HS-SPME and GC-MS to characterization of volatile compounds emitted from Osmanthus flowers. *Ann. Chim.*, **2004**, *94*, 921-927.

[21] de Kraker, J.W.; Franssen, M.C.; de Groot, A.; Konig, W.A.; Bouwmeester, H.J. (+)-Germacrene A biosynthesis. The committed step in the biosynthesis of bitter sesquiterpene lactones in chicory. *Plant physiol.*, **1998**, *117*, 1381-1392.

[22] Lücker, J.; El Tamer, M.K.; Schwab, W.; Verstappen, F.W.; van der Plas, L.H.; Bouwmeester, H.J.; Verhoeven, H.A. Monoterpene biosynthesis in lemon (*Citrus limon*). cDNA isolation and functional analysis of four monoterpene synthases. *Eur. J. Biochem.*, **2002**, *269*, 3160-3171.

[23] Sharon-Asa, L.; Shalit, M.; Frydman, A.; Bar, E.; Holland, D.; Or, E.; Lavi, U.; Lewinsohn, E.; Eyal, Y. Citrus fruit flavor and aroma biosynthesis: isolation, functional characterization, and develop-

mental regulation of Cstps1, a key gene in the production of the sesquiterpene aroma compound valencene. *Plant J.*, **2003**, *36*, 664-674.

[24] Shih, Y.P.; Kung, W.M.; Chen, J.C.; Yeh, C.H.; Wang, A.H.; Wang, T.F. High-throughput screening of soluble recombinant proteins. *Protein Sci.*, **2002**, *11*, 1714-1719.

[25] Tholl, D.; Kish, C.M.; Orlova, I.; Sherman, D.; Gershenzon, J.; Pichersky, E.; Dudareva, N. Formation of monoterpenes in Antirrhinum majus and Clarkia breweri flowers involves heterodimericgeranyldiphosphate synthases. *Plant cell*, **2004**, *16*, 977-992.

[26] Trapp, S.C.; Croteau, R.B. Genomic organization of plant terpene synthases and molecular evolutionary implications. *Genetics*, **2001**, *158*, 811-832.

[27] Martin, D.M.; Faldt, J.; Bohlmann, J. Functional characterization of nine Norway Spruce TPS genes and evolution of gymnosperm terpene synthases of the TPS-d subfamily. *Plant physiol.*, **2004**, *135*, 1908-1927.

[28] Saitou, N.; Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.*, **1987**, *4*, 406-425.

[29] Swofford, D.L. PAUP*: Phylogenetic Analysis Using Parsimony (and other methods) 4.0 Beta. Sunderland (MA): Sinauer Associates, Sunderland, Massachusetts.

[30] Huelsenbeck, J.P.; Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, **2001**, *17*, 754-755.

[31] Kramer, E.M. Methods for studying the evolution of plant reproductive structures: comparative gene expression techniques. *Methods Enzymol.*, **2005**, *395*, 617-636.

[32] Jin, J.B.; Kim, Y.A.; Kim, S.J.; Lee, S.H.; Kim, D.H.; Cheong, G.W.; Hwang, I. A new dynamin-like protein, ADL6, is involved in trafficking from the trans-Golgi network to the central vacuole in Arabidopsis. Plant cell, **2001**, *13*, 1511-1526.

[33] Dudareva, N.; Martin, D.; Kish, C.M.; Kolosova, N.; Gorenstein, N.; Faldt, J.; Miller, B.; Bohlmann, J. (E)-beta-ocimene and myrcene synthase genes of floral scent biosynthesis in snapdragon: function and expression of three terpene synthase genes of a new terpene synthase subfamily. *Plant cell*, **2003**, *15*,1227-1241.

[34] Goodwin, S.M.; Kolosova, N.; Kish, C.M.; Wood, K.V.; Dudareva, N.; Jenks, M.A. Cuticle characteristics and volatile emissions of petals in Antirrhinum majus. *Physiol. Plant*, **2003**, *117*, 435-443.

[35] Figueiredo, A.C.; Barroso, J.G.; Pedro, L.G.; Scheffer, J.J. Factors affecting secondary metabolite production inplants: volatile components and essential oils. *Flavour Fragr. J.*, **2008**, *23*, 213-226.

[36] Mahmoud, S.S.; Croteau, R.B. Menthofuran regulates essential oil biosynthesis in peppermint by controlling a downstream monoterpenereductase. *Proc. Natl. Acad. Sci. USA*, **2003,** *100*, 14481-14486.

[37] Kollner, T.G.; O'Maille, P.E.; Gatto, N.; Boland, W.; Gershenzon, J.; Degenhardt, J. Two pockets in the active site of maize sesquiterpene synthase TPS4 carry out sequential parts of the reaction scheme resulting in multiple products. *Arch. Biochem. Biophys.*, **2006**, *448*, 83-92.

[38] Lücker, J.; Schwab, W.; van Hautum, B.; Blaas, J.; van der Plas, L.H.; Bouwmeester, H.J; Verhoeven, H.A. Increased and altered fragrance of tobacco plants after metabolic engineering using three monoterpene synthases from lemon. *Plant physiology*, **2004**, *134*, 510-519.

[39] Chen, F.; Ro, D.K.; Petri, J.; Gershenzon, J.; Bohlmann, J.; Pichersky, E.; Toll, D. Characterization of a root-specific Arabidopsis terpene synthase responsible for the formation of the volatile monoterpene 1,8-cineole. *Plant physiol.*, **2004**, *135*, 1956-1966.

[40] Jones, C.G.; Keeling, C.I.; Ghisalberti, E.L.; Barbour, E.L.; Plummer, J.A.; Bohlmann, J. Isolation of cDNAs and functional characterisation of two multi-product terpene synthase enzymes from sandalwood, *Santalum album* L. *Arch. Biochem. Biophys.,* **2008**, *477*, 121-130.

[41] Walker, T.S.; Bais, H.P.; Grotewold, E.; Vivanco, J.M. Root exudation and rhizosphere biology. *Plant physiol.*, **2003**, *132*, 44-51.

[42] Steeghs, M.; Bais, H.P.; de Gouw, J.; Goldan, P.; Kuster, W.; Northway, M.; Fall, R.; Vivanco, J.M. Proton-transfer-reaction mass spectrometry as a new tool for real time analysis of root-secreted volatile organic compounds in Arabidopsis. *Plant physiol.*, **2004**, *135*, 47-58.

[43] Gershenzon, J.; McCaskill, D.; Rajaonarivony, J.I.; Mihaliak, C.; Karp, F.; Croteau, R. Isolation of secretory cells from plant glandular trichomes and their use in biosynthetic studies of monoterpenes and other gland products. *Anal. Biochem.*, **1992**, *200*, 130-138.

[44] McCaskill, D.; Croteau, R. Isoprenoid synthesis in peppermint (Mentha x piperita): development of a model system for measuring flux of intermediates through the mevalonic acid pathway in plants. *Biochem. Soc. Trans.*, **1995**, *23*, 290S.

[45] Dudareva, N.; Pichersky, E. Biochemical and molecular genetic aspects of floral scents. *Plant physiol.*, **2000,** *122*, 627-633.

Received: February 12, 2012 **Revised: July 06, 2012** **Accepted: July 29, 2012**

doi:10.6342/NTU201601257

GBE

Genome Biology and Evolution

SMBE

Ancient Nuclear Plastid DNA in the Yew Family (Taxaceae)

Chih-Yao Hsu[1,2,†], Chung-Shien Wu[1,†], and Shu-Miaw Chaw[1,*]

[1]Biodiversity Research Center, Academia Sinica, Taipei 11529, Taiwan

[2]Genome and Systems Biology Degree Program, National Taiwan University and Academia Sinica, Taipei 10617, Taiwan

*Corresponding author: E-mail: smchaw@sinca.edu.tw.

[†]These authors contributed equally to this work.

Accepted: July 25, 2014

Data deposition: This project has been deposited at DDBJ under the accessions AP014574, AP014575, AB936749, AB936745, AB936746, AB936747, and AB936748.

Abstract

Plastid-to-nucleus DNA transfer provides a rich genetic resource to the complexity of plant nuclear genome architecture. To date, the evolutionary route of nuclear plastid DNA (*nupt*) remain unknown in conifers. We have sequenced the complete plastomes of two yews, *Amentotaxus formosana* and *Taxus mairei* (Taxaceae of coniferales). Our comparative genomic analyses recovered an evolutionary scenario for plastomic reorganization from ancestral to extant plastomes in the three sampled Taxaceae genera, *Amentotaxus, Cephalotaxus, and Taxus*. Specific primers were designed to amplify nonsyntenic regions between ancestral and extant plastomes, and 12.6 kb of *nupts* were identified based on phylogenetic analyses. These *nupts* have significantly accumulated GC-to-AT mutations, reflecting a nuclear mutational environment shaped by spontaneous deamination of 5-methylcytosin. The ancestral initial codon of *rps8* is retained in the ~~T~~ *nupts*, but its corresponding extant codon is mutated and requires C-to-U RNA-editing. These findings suggest that *nupts* can help recover scenarios of the nucleotide mutation process. We show that the Taxaceae *nupts* we retrieved may have been retained ~~because~~ the Cretaceous and they carry information of both ancestral genomic organization and nucleotide composition, which offer clues for understanding the plastome evolution in conifers.

Key words: plastome, *Nupt*, ancestral genome, genomic reorganization, taxaceae, conifer.

Introduction

Plastids are cellular organelles descended from a free-living cyanobacterium (Martin et al. 2002). Their genomes (so-called plastomes) are extremely reduced with a large fraction of genes transferred to the nucleus. Transfer of plastid DNA to nuclear genomes is an ongoing process that increases the complexity of nuclear genomes (Timmis et al. 2004). Previous comparative genomic studies indicated that on average approximately 14% of the nuclear-encoded proteins were acquired from the cyanobacterial ancestor of plastids (Deusch et al. 2008). Transgenic experiments also demonstrated a high frequency of plastid-to-nucleus transfers with one event per 11,000 pollen grains or per 273,000 ovules (Sheppard et al. 2008).

Nuclear plastid DNA, termed *nupts* (Richly and Leister 2004), has been discovered in a large number of plant species (Smith et al. 2011). *Nupts* can contribute to nuclear exonic sequences (Noutsos et al. 2007) and plays an important role in plant evolution. *Nupts* may be initially inserted close to centromeres and then fragmented and distributed by transposable elements (Michalovova et al. 2013). The amount of *nupts* in plants is associated with the nuclear genome size and the number of plastids per cell (Smith et al. 2011; Yoshida et al. 2014).

Research of *nupts* remains limited to plant species with both nuclear and plastid genomes have been completely sequenced. In nuclear genomes, the arrangement of *nupts* resembled that of plastomes or consisted of mosaic DNA derived from both plastids and mitochondria (Leister 2005; Noutsos et al. 2005). Notably, a 131-kb *nupt* of rice was found to harbor a 12.4-kb inversion, which was considered to have taken place by homologous recombination in the plastome before the transfer (Huang et al. 2005). Recently, Rousseau-Gueutin et al. (2011) proposed a polymerase chain reaction (PCR)-based method to amplify *nupts* containing a specific ancestral sequence that was deleted from the plastomes of viable offspring. Hence, ancestral plastomic characteristics, such as unique indels and gene orders of specific fragments, may be retained in *nupts*. Construction of an

ancestral plastomic organization should therefore yield valuable clues to retrieve *nupts*. If a plastomic inversion can distinguish an ancestral plastome from its current counterpart, appropriate primers based on the ancestral plastomic organization should be able to amplify the corresponding *nupts* that were transferred to the nucleus before the inversion (fig. 1).

Although the first known *nupt* was identified more than three decades ago (Timmis and Scott 1983), *nupts* of gymnosperms still remain poorly studied. Conifers, the most diverse gymnosperm group, possess huge nuclear genomes ranging from 8.3 to 64.3 pg (2C) (reviewed in Wang and Ran 2014) and may have integrated many *nupts*. The plastomes of conifers are highly rearranged, possibly due to their common loss of a pair of large-inverted repeats (IR) (Wicke et al. 2011; Wu and Chaw 2014). Numerous plastomic rearrangements have been identified and are useful in reconstructing phylogenetic relationships between taxa and inferring intermediate ancestral plastomes (Wu and Chaw 2014). Therefore, the conifer plastomes are well suited for evaluating the feasibility of retrieving *nupts* and surveying their evolution (fig. 1).

Taxaceae (yews) is the smallest family of conifers, consisting of 28 species in six genera, *Amentotaxus*, *Austrotaxus*, *Cephalotaxus*, *Pseudotaxus*, *Taxus*, and *Torreya*. They are mainly distributed in the northern hemisphere. *Amentotaxus* includes five species restricted to subtropical southeastern Asia, from Taiwan west across southern China to Assam in the eastern Himalayas and south to Vietnam (Cheng et al. 2000). *Taxus* includes seven species, best known for containing anticancer agent taxol. They commonly occur in the understory of moist temperate or tropical mountain forests (de Laubenfels 1988).

In this study, we aim to demonstrate our approach (fig. 1) for mining *nupts* in yews and to continue the understanding of the plastome evolution in conifers. To better reconstruct ancestral plastomes of yews, we sequenced two complete plastomes, one from each of the yew genera *Amentotaxus* and *Taxus*. The primers based on the recovered ancestral plastomic organization were used to amplify potential *nupts*. The origins of obtained *nupt* candidates were then examined by phylogenetic analyses and mutation preferences to ensure that they were indeed transferred plastomic DNA in the nucleus. Here, for the first time, we demonstrate that conifer *nupts* can be PCR amplified using our approach and that ancestral plastomic characteristics retained in *nupts* can be compared with extant ones, providing valuable information for understanding plastome evolution in conifers.

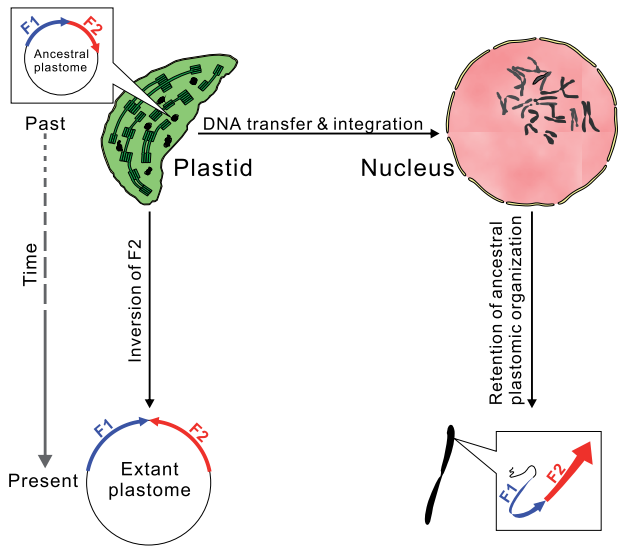

Fig. 1.—A schematic explanation for the amplification of ancestral plastomic DNAs transferred from plastids to the nucleus. Top left: An ancestral plastomic fragment that includes F1 and F2 subfragments with a head-to-tail arrangement was transferred to the nucleus (top right) in the past. After this transfer, an inversion of F2 occurred, which resulted in a head-to-head arrangement of F1 and F2 in the extant plastome. Primers based on distinctive arrangements between ancestral and extant plastomes can facilitate specific amplification of transferred ancestral plastomic fragments and avoid contaminants from amplification of the extant plastome.

Materials and Methods

DNA Extraction, Sequencing, and Genome Assembly

Young leaves of *Amentotaxus formosana* and *Taxus mairei* were harvested in the greenhouse of Academia Sinica and Taipei Botanical Garden, respectively, then ground with liquid nitrogen. Total DNA was extracted by a hexadecyltrimethylammonium bromide (CTAB) method with 2% polyvinylpyrrolidone (Stewart and Via 1993). The DNA was qualified by a threshold of both 260/280 = 1.8–2.0 and 260/230 > 1.7 for next-generation DNA sequencing on an Illumina GAII instrument at Yourgene Bioscience (New Taipei City, Taiwan). For each species, approximately 4 GB of 73-bp paired-end reads were obtained. These short reads were trimmed with a threshold of error probability <0.05 and then de novo assembled by use of CLC Genomic Workbench 4.9 (CLC Bio, Aarhus, Denmark). Contigs with sequence coverage of depth greater than 50× were blasted against the nr database of the National Center for Biotechnology Information (NCBI). Contigs with hits for plastome sequences with *E* value $<10^{-10}$ were retained for subsequent analyses. Gaps between contigs were closed by PCR experiments with specific primers. PCR amplicons were sequenced on an ABI 3730xl DNA Sequencer (Life Technologies).

Genome Annotation and Sequence Alignment

Genome annotation involved use of Dual Organellar Genome Annotator (DOGMA) with the default option (Wyman et al. 2004). Transfer RNA genes were explored by using tRNA scanSE 1.21 (Schattner et al. 2005). For each species, we aligned

the annotated genes and their orthologous genes of other known conifer plastomes to confirm gene boundaries. Sequences were aligned using MUSCLE (Edgar 2004) implemented in MEGA 5.0 (Tamura et al. 2011).

Exploration of SNPs, Indels, and SSRs

For estimating the distribution of both single-nucleotide polymorphisms (SNPs) and indels between our newly sequenced plastome of *T. mairei* and the *T. mairei* voucher NN014, the two genomes were aligned by using VISTA (Frazer et al. 2004). The alignment was then manually divided into non-overlapping bins of 200 bp according to the position of our newly sequenced *T. mairei* plastome. Both SNPs and indels in each bin were estimated by using DnaSP 5.10 (Librado and Rozas 2009). Simple sequence repeats (SSRs) of the *T. mairei* plastome were explored using SSRIT (Temnykh et al. 2001) with a threshold of repeat units >3.

Construction of Ancestral Plastomic Organization

We performed whole-plastomic alignments between the two yews under study and other conifers, *Calocedrus formosana* (NC_023121), *Cephalotaxus wilsoniana* (NC_016063), *Cryptomeria japonica* (NC_010548), *Cunninghamia lanceolata* (NC_021437), and *Taiwania cryptomerioides* (NC_016065), to detect locally collinear blocks (LCBs) using Mauve 2.3.1 (Darling et al. 2010). The yielded matrix of LCBs was used for constructing the putative ancestral plastomic organizations on MGR 2.03 (Bourque and Pevzner 2002), which seeks the minimal genomic rearrangements over all edges of a most parsimonious tree.

PCR Amplification, Cloning, and Sequencing

Ten pairs of specific primers used for amplification of *nupt* sequences were designed and their sequences and corresponding locations are in supplementary table S1, Supplementary Material online and figure 3. PCR amplification involved use of the long-range PCR Tag (TaKaRa LA Taq, Takara Bio Inc.) under the thermo-cycling condition 98 °C for 3 min, followed by 30 cycles of 98 °C for 15 s, 55 °C for 15 s, and 68 °C for 4 min, and a final extension at 72 °C for 10 min. Amplicons were checked by electrophoresis. Amplicons with expected lengths were collected and cloned into yT&A vectors (Yeastern Biotech Co., Taipei) that were then proliferated in *Escherichia coli*. Sequencing the proliferated amplicons involved M13-F and M13-R primers on an ABI 3730xl DNA Sequencer (Life Technologies).

Phylogenetic Tree Analysis

Maximum-likelihood (ML) trees were inferred from sequences of potential *nupts*, their plastomic counterparts, and their orthologs of other gymnosperms using MEGA 5.0 (Tamura et al. 2011) under a GTR + G (four categories) model. Supports for nodes of trees were evaluated by 1,000 bootstrap replications.

Calculation of Mutations in *Nupts* and Their Plastomic Counterparts

The sequence for each *nupt* was aligned with the homologous plastome sequences for *A. formosana*, *Ce. wilsoniana*, *T. maire*, and *Cu. lanceolata* using MUSCLE (Edgar 2004). To precisely calculate the mutational preference in *nupts*, all ambiguous sites and gaps were removed from our alignments. Nucleotide divergence between *nupts* and their plastomic counterparts were derived from mutations in *nupts* or their plastomic counterparts. A mutation of a *nupt* or the plastomic counterpart was recognized when the corresponding site of the plastomic counterpart or *nupt* was identical to that of at least two other taxa. For example, a specific aligned site has "T", "C", "C", "C", and "C" in Cep-2 *nupt*, *Ce. wilsoniana*, *A. formosana*, *T. maire*, and *Cu. lanceolata*, respectively (also see the aligned position 32 in supplementary fig. S5, Supplementary Material online). This site would be recognized as a nonsynonymous mutation from "C" to "T" in the Cep-2 *nupt* as the corresponding amino acid change from alanine to valine.

Genome Map and Statistical Analyses

The plastome map of *T. mairei* was drawn using Circos (http://circos.ca/, last accessed August 9, 2014). All statistical tests, including Pearson's correlation test and Student's *t*-test, involved use of Microsoft Excel 2010.

Results

Plastomic Evolution of *T. mairei* toward Reduction and Compaction

The plastomes of *A. formosana* (AP014574) and *T. mairei* (AP014575) are circular molecules with AT contents of 64.17% and 65.32%, respectively. The *T. mairei* (128,290 bp) plastome has lost five genes (*rps16*, *trnA-UGC*, *trnG-UCC*, *trnI-GAU*, and *trnS-GGA*) compared with that of *A. formosana* (136,430 bp), which leads to a relatively smaller plastome size. The coding regions occupy 61.27% of the plastome length in *A. formosana* and 64.18% in *T. mairei*. The gene density was estimated to be 0.88 and 0.90 (genes/kb) for the plastome of *A. formosana* and *T. mairei*, respectively. In addition, the other two published plastomes for Taxaceae species, *Ce. wilsoniana* (NC_016063) and *Ce. oliveri* (NC_021110), are 136,196 and 134,337 bp, respectively. Altogether, these data suggest that the plastome of *T. mairei* has evolved toward reduction and compaction.

Dot-plot analysis (supplementary fig. S1, Supplementary Material online) reveal three genomic rearrangements between the plastomes of *A. formosana* and *T. mairei*, including a relocated fragment of approximately 18 kb from *psbK* to

trnC-GCA, a relocated fragment of approximately 16 kb from *trnD-GUC* to *trnT-UGU*, and an inverse fragment of approximately 18 kb from *5′rps12* to *infA*. However, the two plastomes share a unique inverted repeat pair that contains *trnQ-UUG* in each copy, hereafter termed "*trnQ*-IR" (supplementary fig. S1, Supplementary Material online).

Intraspecies Variations in the Plastomes of *T. mairei*

To date, the plastomes of three *T. mairei* individuals (*T. mairei* voucher NN014: NC_020321, *T. mairei* voucher SNJ046: JN867590, and *T. mairei* voucher WC052: JN867591) have been published. Together with our newly sequenced plastome of *T. mairei*, these four plastomes vary slightly in size ranging from 127,665 to 128,290 bp. A neighbor-joining tree inferred from the whole-plastomic alignment between these four individuals and *A. formosana* is shown in supplementary figure S2, Supplementary Material online. The tree topology indicates that although the plastome size of SNJ046 and WC052 is similar to that of NN014, SNJ046, and WC052 form a sister clade to our locally sampled *T. mairei*.

We also performed a pairwise genome comparison between our *T. mairei* and voucher NN014 because the latter was designated as the reference sequence (RefSeq) in NCBI GenBank. We detected 858 SNPs and 218 indels. Supplementary figure S3, Supplementary Material online, shows that the intergenic spacers and coding regions contained nearly equal numbers of SNPs. Most of the indels were found in the intergenic spacers. We found 33 indels in the coding regions, but none caused frameshifts. Figure 2 illustrates the distribution of SNPs, indels, and SSRs in the plastome of our sampled *T. mairei.* Interestingly, the abundance of SSRs was positively correlated with that of SNPs (Pearson, $r = 0.52$, $P < 0.01$), with no correlation between SSRs and indels (Pearson, $r = 0.02$, $P = 0.89$). In legumes, the region that contains *ycf4*, *psaI*, *accD*, and *rps16* was found to be hypermutable (Magee et al. 2010). In the plastome of *T. mairei*, three 200-bp bins that locate in the sequence of 5′*clpP* (position 55,001–55,200), that of 5′*ycf1* (pos. 124,201–124,400), and the intergenic spacer between *rrn16* and *rrn23* (pos. 96,801–97,000) contained the highest sum of SNPs, indels, and SSRs (fig. 2). Therefore, these loci can be considered intraspecies mutational hotspots in *T. mairei* and can be potentially high-resolution DNA barcodes in the study of population genetics.

Retrieval of Ancestral Plastome Sequences in Taxaceae

A matrix with 20 LCBs was generated on the basis of whole-plastome alignments between the sampled three Taxaceae and four Cupressaceae species. This matrix of LCBs was then used in reconstructing ancestral plastomic organization. The most parsimonious tree with the corresponding ancestral plastomic organization is shown in supplementary figure S4, Supplementary Material online and figure 3, and that the three Taxaceae species form a monophyletic clade whereas *A. formosana* is closer to *Ce. wilsoniana* than to *T. mairei.* This topology is in good agreement with the recent molecular review of the conifer phylogeny by Leslie et al. (2012). Figure 3 shows the detailed evolutionary scenario of plastomic rearrangements with the intermediate ancestral plastomes in the three examined Taxaceae species. By comparing the ancestral and extant plastomes, one, three, and two inversions might have occurred in *A. formosana, Ce. wilsoniana* and *T. mairei,* respectively, after they had diverged from their common ancestor. Specific primer pairs were used for amplifying the corresponding ancestral fragments that differ from the extant plastomes in genomic organization (fig. 3). Five (Ame-2, Cep-2, Cep-5, Cep-6, and Tax-4) out of the ten primer pairs were able to produce amplicons totaling 16.6 kb (see supplementary table S2, Supplementary Material online, for accession numbers).

Characteristics of Potential *Nupt* Amplicons

The obtained PCR amplicons were sequenced and annotated (supplementary table S2, Supplementary Material online). With the exception of *chlB* of Cep-2, all putative protein-coding genes contain no premature stop codons. The coding sequence (CDS) of each amplicon was aligned with its plastomic counterparts and orthologs of other cupressophytes, *Ginkgo* and *Cycas*. We used ML trees inferred from concatenated CDSs to examine the origins of these PCR amplicons, with *Ginkgo* and *Cycas* as the outgroup (fig. 4). In each tree, the plastomic sequences were divided into three groups (i.e., the Cupressaceae clade, the Taxaceae clade, and the clade comprising Araucariaceae and Podocarpaceae). Notably, the placements of our PCR amplicons are incongruent among the four trees. For example, both Ame-2 and Cep-2 were clustered with their plastomic counterparts (fig. 4*A*). In contrast, Cep-5, Cep-6, and Tax-4 were placed remotely from their individual plastomic counterparts, indicating that they originated via horizontal transfer (fig. 4*B*–*D*).

The ancestral plastomic organization that we used to design primers for amplification of Ame-2 and Cep-2 was rearranged by a 34-kb inversion flanked by *trnQ*-IRs. This *trnQ*-IR is 564 and 549 bp for *A. formosana* and *Ce. wilsoniana*, respectively. IRs of similar sizes can mediate homologous recombination in the conifer plastomes (Tsumura et al. 2000; Wu et al. 2011; Yi et al. 2013; Guo et al. 2014). As a result, if the *trnQ*-IR-mediated isomeric plastome is present in our sampled taxa, our PCR approach may also be able to amplify isomeric plastomic fragments. Ame-2 has 100% sequence identity with its plastomic counterpart (fig. 4*A*) in the CDS, which strongly suggests its origin as an isomeric plastome. Cep-2 differs from its plastomic counterpart by several mutations, including two premature stop codons in *chlB*, of which one of the two cannot be replaced by neither U-to-C nor C-to-U RNA-editing (supplementary fig. S5, Supplementary

doi:10.6342/NTU201601257

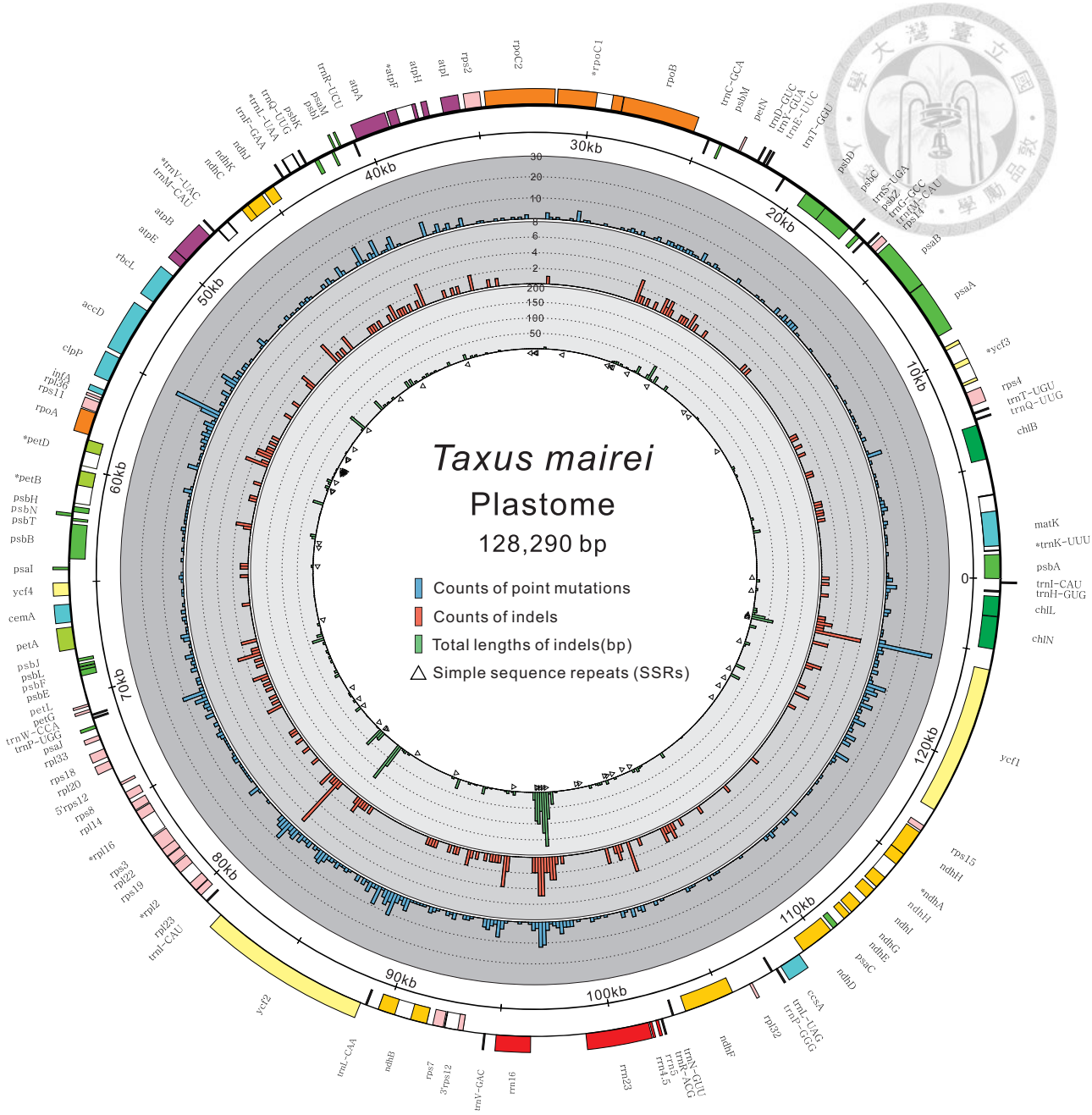

Fig. 2.—Distribution of SNPs, indels, and SSRs in the plastomes of *Taxus mairei*. The outermost circle is the plastome map of *T. mairei* (AP014575) with genes that are transcribed counter-clockwise (outer boxes) and clockwise (inner boxes), respectively. The immediately next circle denotes a scale of 5-kb units beginning at *psbA* gene (the 3 o'clock position). In the gray zone, three histograms from outer to inner are 1) counts of SNPs, 2) counts of indels, and 3) total indel lengths within nonoverlapping 200-bp bins across the entire plastome. Triangles mark locations of SSRs.

Material online). Therefore, the origin of Cep-2 is from a horizontal transfer rather than an isomeric plastome.

Evolution of *Nupt* Sequences in Taxaceae

The sequence identity between the four *nupts* and their plastomic counterparts ranges from 61.71% to 99.08% (table 1). In fact, differences in aligned sites between *nupts* and their plastomic counterparts are derived from two types of mutations. One is the mutation in *nupts* and the other is that in plastomes. As shown in table 1, with the exception of Tax-4, all *nupts* accumulated more mutations than their plastomic counterparts. The low sequence identity between Tax-4 and its plastome sequences (61.71% in table 1) may be due to the unusually increased mutations in the latter. In all *nupts* except Cep-5, at least one potential protein-coding gene had the ratio of nonsynonymous (d*n*)/synonymous (d*s*)

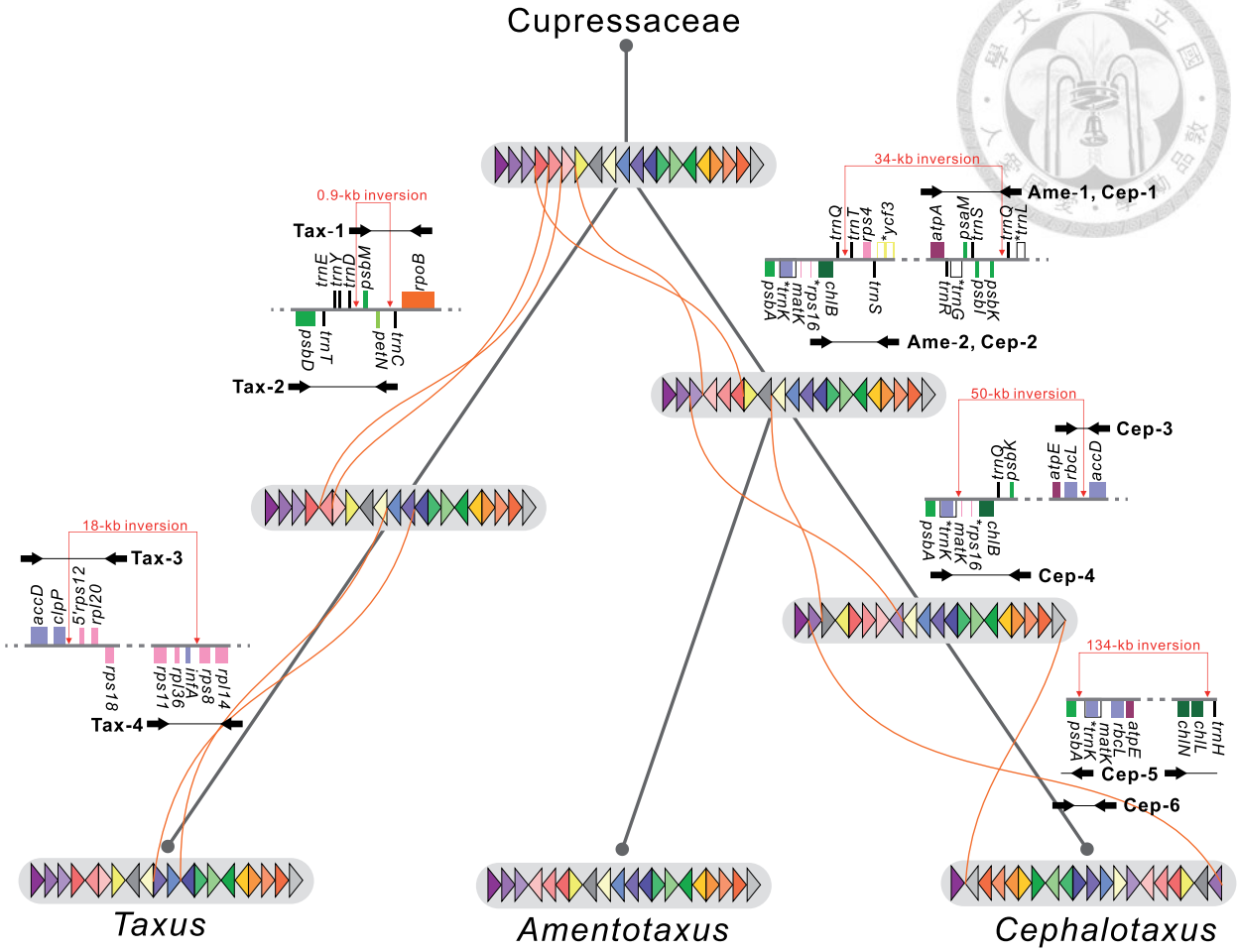

Fig. 3.—Hypothetical evolutionary scenarios for plastomic rearrangements in Taxaceae. Plastomes are circular but here are shown in gray horizontal bars (beginning at *psbA*) for pairwise comparisons. Color triangles within the gray horizontal bars denote LCBs with their relative orientations. Gray bars from top to bottom indicate the corresponding plastomes in the common ancestor of Taxaceae, intermediate ancestors, and extant representative species. Inversions between two plastomes are linked by orange curved lines. Ancestral gene orders before the occurrence of specific inversions are shown along tree branches. Primer pairs (black arrows) for amplification of the corresponding ancestral fragments are labeled: Tax-1 to 4 for *Taxus mairei*, Ame-1 to 2 for *Amentotaxus formosana*, and Cep-1 to 5 for *Cephalotaxus wilsoniana* (see supplementary table S1, Supplementary Material online, for primer sequences).

mutations >1, which reflects relaxed functional constraints in *nupts*. Figure 5 illustrates nucleotide mutation classes in *nupts* and their corresponding plastome sequences. We excluded the plastomic counterpart of Cep-2 from calculation because we observed only one mutation in the sequence. In all *nupts*, transitional mutations comprise over 50% of the total mutations. The mutation of G to A and its complement C to T (denoted GC-to-AT in fig. 5) had the highest frequency in both *nupts* and plastome sequences. To examine which of the mutation classes is statistically predominant, we compared the two most abundant classes of mutations. In *nupts*, the frequency was higher for GC-to-AT than AT-to-GC mutations (*t*-test, $P = 0.018$). However, GC-to-AT and AT-to-GC mutations did not differ in plastome sequences (*t*-test, $P = 0.379$), suggesting different mutational environments between *nupts* and their corresponding plastome sequences.

Ages of *Nupts* in Taxaceae

Molecular dating of sequences highly depends on mutation rates. Unfortunately, mutation rates in the nuclear genomes of Taxaceae species have not been directly measured. The *nupts* identified in this study were expected to evolve neutrally. The 4-fold degenerated site is a useful indicator in measuring the rate of neutral evolution (Graur and Li 2000). In nuclear genomes of conifers, the mutation rate at the 4-fold degenerate sites was estimated to be 0.64×10^{-9} per site per year (Buschiazzo et al. 2012). In the *nupts* Cep-2, Cep-5, Cep-6, and Tax-4, we found 29, 117, 100, and 42 mutations among 2,961, 3,380, 2,207, and 1,466 sites, respectively (table 1). Therefore, the ages of Cep-2, Cep-5, Cep-6, and Tax-4 were estimated to be approximately 15.3, 54.1, 70.8, and 44.8 Myr, respectively.

Genome Biology and Evolution

SMBE

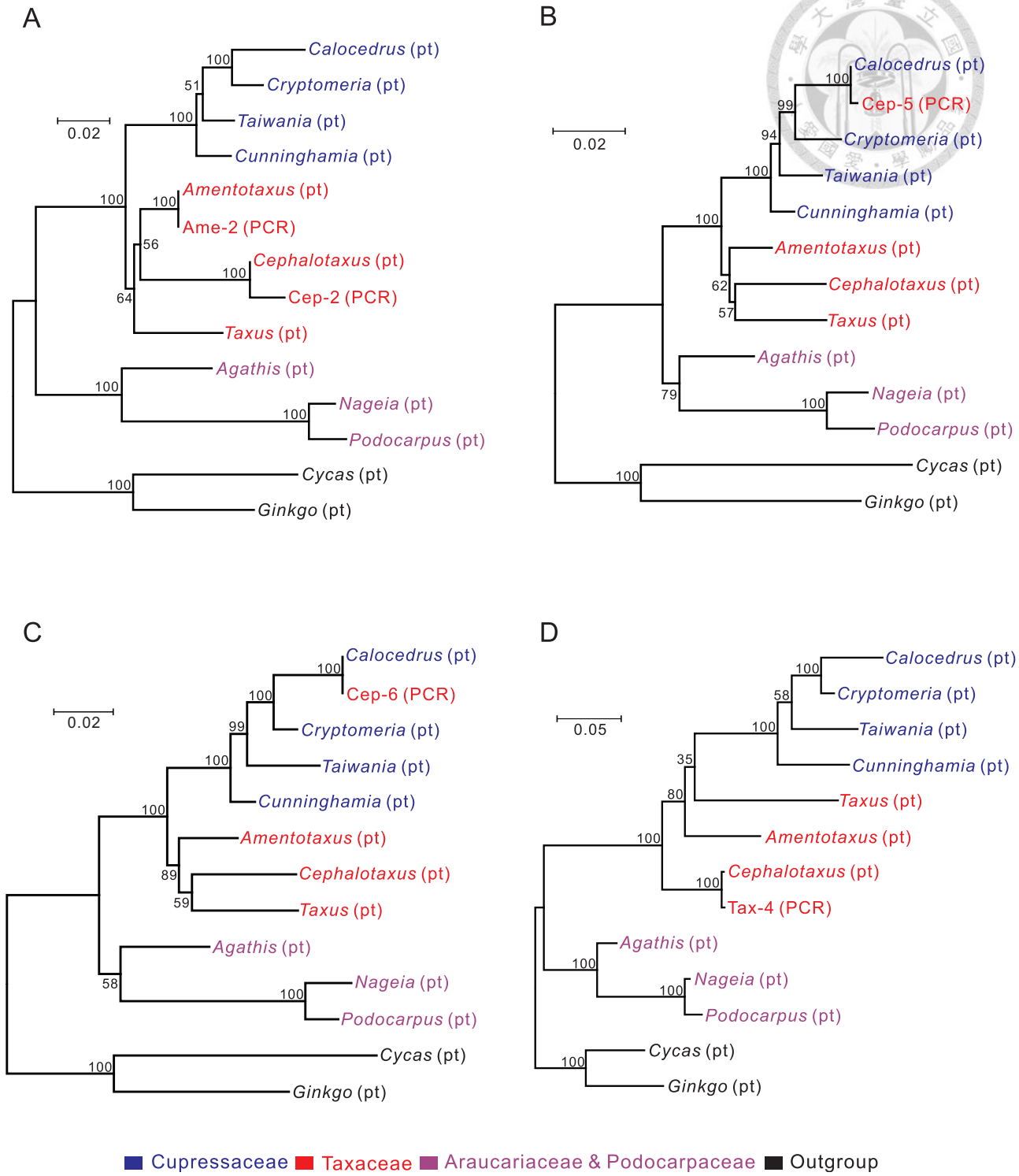

Fig. 4.—Origin of the obtained PCR amplicons examined by ML phylogenetic analyses. PCR amplicons are labeled "PCR," and their plastomic counterparts and orthologs of other gymnosperms are labeled "pt." Taxa of the same conifer family are in the same color. *Cycas* and *Ginkgo* together are the outgroup. Bootstrapping values assessed with 1,000 replicates are shown along branches.

Table 1
Mutations in *Nupts* and Their Plastomic Counterparts

Nupt	Identity[a]	Length[b]	Number of Mutations			
	(%)	(bp)	Total	Potential Protein-Coding Gene	*dn*	*ds*
Cep-2	99.08	2,961	29 (1)	*chlB*	19 (0)	7 (1)
				rps4	0 (0)	2 (0)
Cep-5	88.15	3,380	117 (75)	*psbA*	2 (4)	16 (10)
				chlL	3 (4)	24 (17)
				chlN	12 (15)	38 (42)
Cep-6	89.84	2,207	100 (67)	*psbA*	0 (2)	14 (12)
				matK	45 (37)	29 (10)
Tax-4	61.71	1,466	42 (135)	*rpl14*	2 (2)	0 (3)
				rps8	5 (9)	3 (12)
				infA	4 (26)	1 (22)
				rpl36	0 (1)	1 (1)
				rps11	13 (10)	14 (9)

Note.—Numbers in parentheses indicate mutations in corresponding plastomic sequences; *dn*, nonsynonymous; *ds*, synonymous.
[a]Refers to sequence identity between *nupts* and their plastomic counterparts. Gaps were included in calculating identity.
[b]Refers to lengths of unambiguous alignments where gaps and ambiguous sites were excluded. These alignments were used for calculating mutations.

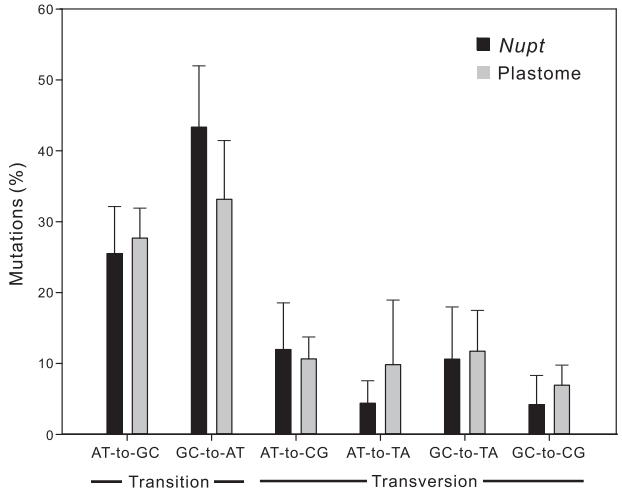

Fig. 5.—Percentage of nucleotide mutation classes in *nupts* and their plastomic counterparts. Types of mutations are divided into six classes. For example, the class AT-to-GC denotes the pooled percentage of the A-to-G mutations and its complement T-to-C. Data are mean ± SD.

Discussion

Labile Plastomes of Yew Family and Their Impact on Phylogenetic Studies

The phylogenetic relationships among *Amentotaxus*, *Cephalotaxus*, and *Taxus* have not been resolved. Recent molecular studies placed *Amentotaxus* as sister to *Taxus* (e.g., Cheng et al. 2000; Mao et al. 2012) or to *Cephalotaxus* (e.g., Leslie et al. 2012). We found that a 34-kb inversion from *trnT* to *psbK* distinguished *A. formosana* and *Ce. wilsoniana* from *T. mairei* (fig. 3), which suggests that *A. formosana* is closer to *Ce. wilsoniana* than to *T. mairei*. However, the plastome of another *Taxus* species, *T. chinensis* (Zhang et al. 2014), cannot be distinguished from those of *A. formosana* and *Ce. wilsoniana* by this 34-kb inversion. Of note, this 34-kb inversion is flanked by a pair of *trnQ*-IR sequences. We found that the *trnQ*-IR sequence is commonly present in *A. formosana* (564 bp), *Ce. wilsoniana* (549 bp), *T. mairei* (248 bp), and *T. chinensis* (248 bp).

The presence of the *trnQ*-IR pair was able to generate isomeric plastomes in *Ce. oliveri* (Yi et al. 2013) and four *Juniperus* species (Guo et al. 2014). In Pinaceae, inverted repeats larger than 0.5 kb could trigger plastomic isomerization, and retention of an isomer was species- or population-specific (Tsumura et al. 2000; Wu et al. 2011). Indeed, figure 4*A* revealed that Ame-2 was likely a PCR amplicon derived from the *trnQ*-IR-mediated isomeric plastome of *A. formosana*. Therefore, with the presence of an isomeric plastome, the synapomorphic character—the 34-kb inversion—in figure 3 might be a false positive result caused by insufficient sampling. Nonetheless, our data also suggest that isomeric plastomes be treated cautiously when using genomic rearrangements in phylogenetic estimates.

Disruption of the plastomic operons is rare in seed plants (Jansen and Ruhlman 2012). We found that the S10 operon of *T. mairei* was separated into two gene clusters (*rpl23-rps8* and *infA-rpoA*) by an 18-kb inversion (fig. 3). Because the transcriptional direction of the S10 operon is from *rpl23* to *rpoA* (Jansen and Ruhlman 2012), the gene cluster *infA-rpoA* in *T. mairei* likely has to acquire a novel promoter sequence for transcription. Disruption of the S10 operon was previously reported in the plastome of Geraniaceae (Guisinger et al. 2011). However, the evolutionary consequence of plastomic operon disruption has never been studied. In the plastome of

T. mairei, we detected prominently elevated mutations in the two separated gene clusters of the S10 operon as compared with their relative *nupts* (table 1). Interestingly, two (i.e., *infA* and *rps11* in table 1) out of the three protein-coding genes on the plastomic gene-cluster *infA-rpoA* had d*n*/d*s* ratios >1. Whether disruption of the S10 operon results in positive selection of these two genes requires further investigation.

Benefits and Cautions of PCR-Based Approach in Investigating *Nupts*

Explosive growth of available sequenced nuclear genomes offers great opportunities for investigating nuclear organellar DNA (*norgs*). The amount of *norgs* could vary depending on the use of different assembly versions of genomes and search strategies (Hazkani-Covo et al. 2010). A PCR-based approach, such as that of Rousseau-Gueutin et al. (2011) and ours, is free from this problem encountered in genome assembly. The *nupts* we amplified and report here are of course a few examples of conifer *nupts*. However, considering the huge nuclear genome of conifers whose sequencing and assembly require much cost and effort, our PCR-based approach provides a cost-effective way for studying the evolution of *nupts*.

Using a threshold of >70% sequence identity, Smith et al. (2011) extracted *nupts* of about 50 kb from the nuclear genome of *Arabidopsis*. The amount of *Arabidopsis nupts* decreased to approximately 17.6 kb when the threshold of sequence identity was increased to 90% (Yoshida et al. 2014). It seems that identification of possible *nupts* is largely influenced by the thresholds. Setting high thresholds might limit the exploration of *nupts* to only relatively recent transfers (Yoshida et al. 2014). Clearly, the problem of setting thresholds is absent from our PCR-based approach. In this study, sequence identity between *nupts* and their plastomic counterparts ranged from 61.71% to 99.08% (table 1). Thus, one or three of the four presented *nupts* would not be obtained if we had considered the thresholds of Smith et al. (2011) or Yoshida et al. (2014), respectively.

Only five of our ten primer pairs worked well, and one amplified the DNA fragment of isomeric plastomes rather than *nupts*. This low success rate may be due to the unsuitable primers used in our PCR experiments. Multiple primer pairs for a specific locus may improve amplification of *nupts*, as noted by Rousseau-Gueutin et al. (2011). Plastid-to-mitochondrion DNA transfers are frequent in seed plants (Wang et al. 2007). Because the mitochondrial genome of Taxaceae spp. is currently unavailable, the possibility that our PCR products were amplicons of mitochondrial plastid DNA could not be ruled out. The phylogenetic tree approach was previously used to examine horizontal DNA transfers (Bergthorsson et al. 2003; Rice et al. 2013), but our tree analyses in figure 4 could not distinguish the transfer events between plastid-to-nucleus and plastid-to-mitochondrion origins. The mutation rate of nuclear genomes is higher than that of plastomes in plants (Wolfe et al. 1987). All of our amplified *nupts*, except Tax-4, had more mutation sites than their plastomic counterparts (table 1). Disruption of the S10 operon is likely associated with the elevated mutation in the plastomic counterpart of Tax-4, as mentioned earlier. Additionally, among our *nupts*, the AT-to-GC mutation was predominant (fig. 5). These data are similar to the findings for *nupts* in rice and *Nicotiana* (Huang et al. 2005; Rousseau-Gueutin et al. 2011), which reflects a nuclear-specific circumstance shaped by spontaneous deamination of 5-methylcytosin.

Nupts Are Molecular Footprints for Studying Plastomic Evolution

Although mutation rates are relatively low in plant organellar genomes, *norgs* can serve as "molecular fossils" for genomic rearrangements (Leister 2005). Similarly, the Taxaceae *nupts* identified in this study do retain the ancestral plastomic organization. In other words, *nupts* are footprints that are valuable in reconstructing the evolutionary history of plastomic organization and rearrangements.

Dating the age of *nupts* is critical for elucidating the evolution of *nupts*. For example, the estimated ages of Cep-2, Cep-5, and Cep-6 *nupts* are 15.3, 54.1, and 70.8 Myr, respectively. Remarkably, these ages conflict with the scenario of plastomic rearrangements because the transfer of Cep-2 predated those of both Cep-5 and Cep-6 (fig. 3). Two plastomic forms derived from *trnQ*-IR-mediated homologous recombination coexist in an individual of *Ce. oliveri* (Yi et al. 2013). This *trnQ*-IR is also present in the plastome of *Ce. wilsoniana* as previously mentioned. We suspect that in *Ce. wilsoniana*, the younger Cep-2 *nupt* might originate from a transferred fragment of the *trnQ*-IR-mediated isomeric plastome.

Most importantly, *nupts* can also help in probing RNA-editing sites and improving gene annotations. Figure 6 clearly reveals that the previously annotated *rps8* of *T. mairei* (vouchers NN014, WC052, and SNJ046) is truncated. Our newly predicted initial codon, "ACG," locates 48 bp upstream of the previously predicted site. This ACG initial codon was predicted to be corrected to "AUG" via a C-to-U RNA-editing because the corresponding sequence of Tax-4 *nupt* and other conifers retain a normal initial codon of "ATG" (fig. 6). These data also imply that in *T. mairei*, the transfer of Tax-4 *nupt* predates the T-to-C mutation at the second codon position in the initial codon of *rps8*.

In conclusion, we have shown that plastomic rearrangement events provide useful information for amplifying *nupts*. Because avoiding the amplification of isomeric plastomic or mitochondrial DNA is difficult, examining the origins of PCR amplicons was a prerequisite in this proposed PCR-based study. In angiosperms such as *Nicotiana*, *nupts* were experimentally demonstrated to be eliminated quickly from the nuclear genome (Sheppard and Timmis 2009). However, we show that the oldest conifer *nupt* has been retained for

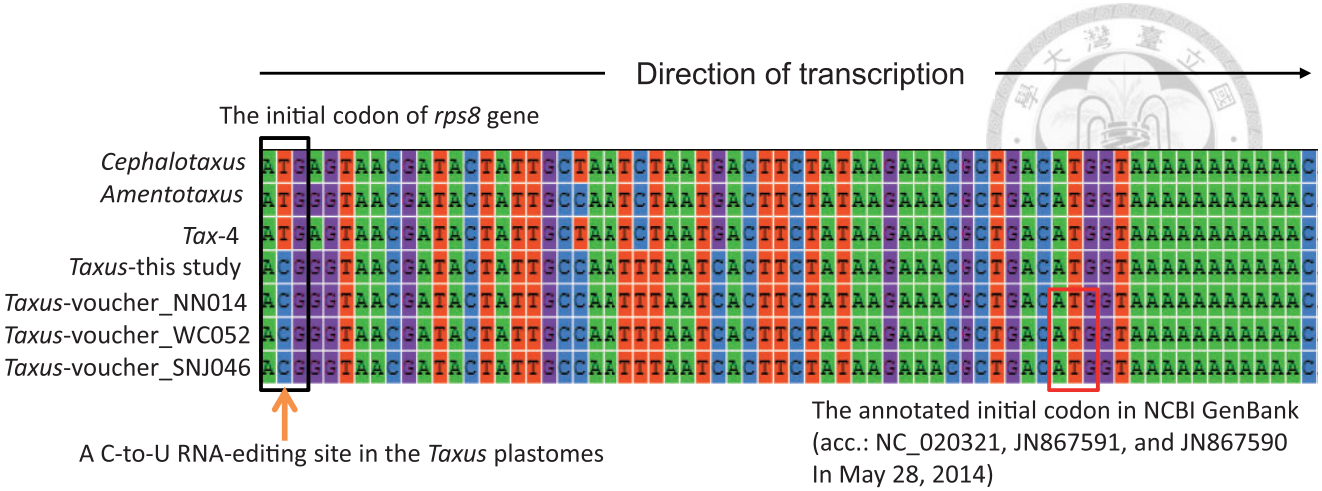

Fig. 6.—Alignment of seven *rps8* sequences. The left orange arrow highlights a specific C-to-U RNA-editing site at the second codon position of the initial codon in the four sampled *Taxus. mairei* plastomes. A normal initial codon, ATG (black rectangle), was common among *Cephalotaxus*, *Amentotaxus*, and Tax-4. These data imply the creation of the "ACT" RNA-editing site after the transfer of Tax-4. The red rectangle denotes the initial codon annotated in the sequences from NCBI GenBank.

70.8 Myr (i.e., since the Cretaceous period). With an increase of available plastomes in conifers, comparative genome analyses are expected to reveal more plastomic rearrangements. Using our approach, we are beginning to understand the evolution of *nupts* in diverse conifer species without the need to sequence and assemble their huge nuclear genomes.

Supplementary Material

Supplementary tables S1–S2 and figures S1–S5 are available at *Genome Biology and Evolution* online (http://gbe.oxfordjournals.org/).

Acknowledgments

This work was supported by research grants from the National Science Council, Taiwan (NSC 100-2621-B-001-003-MY3) and from the Investigator's Award of Academia Sinica to S.-M.C., and a doctoral student fellowship of jointed doctoral program by National Taiwan University and Academia Sinica to C.-Y.H., and a postdoctoral fellowship of Academia Sinica to C.-S.W. The authors thank two anonymous reviewers' helpful comments on the manuscript. The authors are indebted to Dr Isheng Tsai for his critical reading and editing of this revised version.

Literature Cited

Bergthorsson U, Adams KL, Thomason B, Palmer JD. 2003. Widespread horizontal transfer of mitochondrial genes in flowering plants. Nature 424:197–201.

Bourque G, Pevzner PA. 2002. Genome-scale evolution: reconstructing gene orders in the ancestral species. Genome Res. 12:26–36.

Buschiazzo E, Ritland C, Bohlmann J, Ritland K. 2012. Slow but not low: genomic comparisons reveal slower evolutionary rate and higher dN/dS in conifers compared to angiosperms. BMC Evol Biol. 12:8.

Cheng Y, Nicolson RG, Tripp K, Chaw SM. 2000. Phylogeny of Taxaceae and Cephalotaxaceae genera inferred from chloroplast matK gene and nuclear rDNA ITS region. Mol Phylogenet Evol. 14:353–365.

Darling AE, Mau B, Perna NT. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One 5: e11147.

de Laubenfels David J. 1988. Coniferales. In flora malesiana, Series I, Vol. 10. Dordrecht: Kluwer Academic. p. 337-453.

Deusch O, et al. 2008. Genes of cyanobacterial origin in plant nuclear genomes point to a heterocyst-forming plastid ancestor. Mol Biol Evol. 25:748–761.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797.

Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. 2004. VISTA: computational tools for comparative genomics. Nucleic Acids Res. 32:W273–W279.

Graur D, Li WH. 2000. Fundamentals of molecular evolution. Sunderland (MA): Sinauer Associates.

Guisinger MM, Kuehl JV, Boore JL, Jansen RK. 2011. Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: rearrangements, repeats, and codon usage. Mol Biol Evol. 28: 583–600.

Guo W, et al. 2014. Predominant and substoichiometric isomers of the plastid genome coexist within Juniperus plants and have shifted multiple times during cupressophyte evolution. Genome Biol Evol. 6: 580–590.

Hazkani-Covo E, Zeller RM, Martin W. 2010. Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. PLoS Genet. 6:e1000834.

Huang CY, Grünheit N, Ahmadinejad N, Timmis JN, Martin W. 2005. Mutational decay and age of chloroplast and mitochondrial genomes transferred recently to angiosperm nuclear chromosomes. Plant Physiol. 138:1723–1733.

Jansen RK, Ruhlman TA. 2012. Plastid genomes in seed plants. In: Bock R, Knoop V, editors. Genomics of chloroplasts and mitochondria. Netherlands. Springer. p. 103–126.

Leister D. 2005. Origin, evolution and genetic effects of nuclear insertions of organelle DNA. Trends Genet. 21:655–663.

Leslie AB, et al. 2012. Hemisphere-scale differences in conifer evolutionary dynamics. Proc Natl Acad Sci U S A. 109:16217–16221.

Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25: 1451–1452.
Magee AM, et al. 2010. Localized hypermutation and associated gene losses in legume chloroplast genomes. Genome Res. 20: 1700–1710.
Mao K, et al. 2012. Distribution of living Cupressaceae reflects the breakup of Pangea. Proc Natl Acad Sci U S A. 109:7793–7798.
Martin W, et al. 2002. Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. Proc Natl Acad Sci U S A. 99: 12246–12251.
Michalovova M, Vyskot B, Kejnovsky E. 2013. Analysis of plastid and mitochondrial DNA insertions in the nucleus (NUPTs and NUMTs) of six plant species: size, relative age and chromosomal localization. Heredity 111:314–320.
Noutsos C, Kleine T, Armbruster U, DalCorso G, Leister D. 2007. Nuclear insertions of organellar DNA can create novel patches of functional exon sequences. Trends Genet. 23:597–601.
Noutsos C, Richly E, Leister D. 2005. Generation and evolutionary fate of insertions of organelle DNA in the nuclear genomes of flowering plants. Genome Res. 15:616–628.
Rice DW, et al. 2013. Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm Amborella. Science 342:1468–1473.
Richly E, Leister D. 2004. NUPTs in sequenced eukaryotes and their genomic organization in relation to NUMTs. Mol Biol Evol. 21: 1972–1980.
Rousseau-Gueutin M, Ayliffe MA, Timmis JN. 2011. Conservation of plastid sequences in the plant nuclear genome for millions of years facilitates endosymbiotic evolution. Plant Physiol. 157:2181–2193.
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:W686–W689.
Sheppard AE, et al. 2008. Transfer of plastid DNA to the nucleus is elevated during male gametogenesis in tobacco. Plant Physiol. 148:328–336.
Sheppard AE, Timmis JN. 2009. Instability of plastid DNA in the nuclear genome. PLoS Genet. 5:e1000323.
Smith DR, Crosby K, Lee RW. 2011. Correlation between nuclear plastid DNA abundance and plastid number supports the limited transfer window hypothesis. Genome Biol Evol. 3:365–371.
Stewart CN Jr, Via LE. 1993. A rapid CTAB DNA isolation technique useful for RAPD fingerprinting and other PCR applications. Biotechniques 14: 748–750.
Tamura K, et al. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 28:2731–2739.
Temnykh S, et al. 2001. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Res. 11: 1441–1452.
Timmis JN, Ayliffe MA, Huang CY, Martin W. 2004. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat Rev Genet. 5:123–135.
Timmis JN, Scott NS. 1983. Sequence homology between spinach nuclear and chloroplast genomes. Nature 305:65–67.
Tsumura Y, Suyama Y, Yoshimura K. 2000. Chloroplast DNA inversion polymorphism in populations of Abies and Tsuga. Mol Biol Evol. 17: 1302–1312.
Wang D, et al. 2007. Transfer of chloroplast genomic DNA to mitochondrial genome occurred at least 300 MYA. Mol Biol Evol. 24: 2040–2048.
Wang XQ, Ran JH. 2014. Evolution and biogeography of gymnosperms. Mol Phylogenet Evol. 75C:24–40.
Wicke S, Schneeweiss GM, dePamphilis CW, Müller KF, Quandt D. 2011. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Mol Biol. 76:273–297.
Wolfe KH, Li WH, Sharp PM. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc Natl Acad Sci U S A. 84:9054–9058.
Wyman SK, Jansen RK, Boore JL. 2004. Automatic annotation of organellar genomes with DOGMA. Bioinformatics 20:3252–3255.
Wu CS, Chaw SM. 2014. Highly rearranged and size-variable chloroplast genomes in conifers II clade (cupressophytes): evolution towards shorter intergenic spacers. Plant Biotechnol J. 12:344–353.
Wu CS, Lin CP, Hsu CY, Wang RJ, Chaw SM. 2011. Comparative chloroplast genomes of Pinaceae: insights into the mechanism of diversified genomic organizations. Genome Biol Evol. 3:309–319.
Yi X, Gao L, Wang B, Su YJ, Wang T. 2013. The complete chloroplast genome sequence of *Cephalotaxus oliveri* (Cephalotaxaceae): evolutionary comparison of *Cephalotaxus* chloroplast DNAs and insights into the loss of inverted repeat copies in gymnosperms. Genome Biol Evol. 5:688–698.
Yoshida T, Furihata HY, Kawabe A. 2014. Patterns of genomic integration of nuclear chloroplast DNA fragments in plant species. DNA Res. 21: 127–140.
Zhang Y, et al. 2014. The complete chloroplast genome sequence of *Taxus chinensis* var. *mairei* (Taxaceae): loss of an inverted repeat region and comparative analysis with related species. Gene 540:201–209.

Associate editor: Bill Martin

Mitochondrial DNA

The Journal of DNA Mapping, Sequencing, and Analysis

ISSN: 1940-1736 (Print) 1940-1744 (Online) Journal homepage: http://www.tandfonline.com/loi/imdn20

The complete plastome sequence of Gnetum ula (Gnetales: Gnetaceae)

Chih-Yao Hsu, Chung-Shien Wu, Siddharthan Surveswaran & Shu-Miaw Chaw

To cite this article: Chih-Yao Hsu, Chung-Shien Wu, Siddharthan Surveswaran & Shu-Miaw Chaw (2015): The complete plastome sequence of Gnetum ula (Gnetales: Gnetaceae), Mitochondrial DNA

To link to this article: http://dx.doi.org/10.3109/19401736.2015.1079874

Published online: 15 Sep 2015.

Mitochondrial DNA

http://informahealthcare.com/mdn
ISSN: 1940-1736 (print), 1940-1744 (electronic)

Mitochondrial DNA, Early Online: 1–2
 DOI: 10.3109/19401736.2015.1079874

Taylor & Francis
Taylor & Francis Group

MITOGENOME ANNOUNCEMENT

The complete plastome sequence of *Gnetum ula* (Gnetales: Gnetaceae)

Chih-Yao Hsu[1,2], Chung-Shien Wu[1], Siddharthan Surveswaran[3], and Shu-Miaw Chaw[1]

[1]*Biodiversity Research Center, Academia Sinica, Taipei, Taiwan,* [2]*Genome and Systems Biology Degree program, National Taiwan University and Academia Sinica, Taipei, Taiwan, and* [3]*Centre for Ecological Sciences, Indian Institute of Science, Bangalore, India*

Abstract

This study reports the complete plastome sequence of *Gnetum ula*, a gymnosperm species of Gnetaceae (Gnetophyta). The plastome is 113 249 bp long. It has a quadripartite structure containing a pair of large inverted repeat regions of 19 772 bp each, a large single-copy region of 64 914 bp, and a small single-copy region of 8791 bp. One hundred sixteen genes were predicted in the plastome, including 68 protein-coding genes, eight ribosomal RNA genes, and 40 transfer RNA genes. The gene density is 1.024 (genes/kb). Similar to other known *Gnetum* plastomes, the *G.ula* plastome has lost 20 protein-coding genes commonly present in other seed plant plastomes. Our phylogenetic analyses indicate that the four sampled *Gnetum* species are monophyletic and that *G. ula* is close to the two other lianas rather than the only small tree species, *G. gnemon*. Our phylogenetic trees also indicate that gnetophytes have the fastest evolutionary rates among gymnosperms.

Keywords

Gnetum ula, gymnosperm, Illumina sequencing, plastome, reduced genome

History

Received 20 July 2015
Accepted 2 August 2015
Published online 11 September 2015

Gnetum ula, a large woody climber of the gymnosperm family, Gnetaceae (Gnetophyta; common name: gnetophytes), is endemic to the tropical rainforests of southern India. Gnetophytes are distinct from other gymnosperms by their angiosperm-like morphologies such as climbing or shrubby habits, broad leaves with net-like venation, and xylem with vessels (Doyle & Donoghue, 1986; Friedman, 1998). Gnetophytes comprise only three living genera, in which *Gnetum* is the only one distributed in tropical rainforests. In this study, we determined the complete plastome sequence of *G. ula*.

Young leaves of 2 g were collected to extract DNA using a modified CTAB protocol (Stewart & Via, 1993). The extracted DNA concentration was quantified as >300 ng/µl. Sequencing was conducted on an Illumina Hiseq 2500 Sequencing System (Illumina, San Diego, CA) in Yourgene Bioscience (New Taipei City, Taiwan) to yield 125 bp paired-end reads of approximately 2 Gb. Plastome assembly was performed using CLC Genome Workbench 4.9 (CLC Bio, Aarhus, Denmark). Gaps between plastome contigs were closed by PCR experiments with specific primers.

The plastome of *G. ula* (AP014923) is circular and 113 249 bp long. It is the smallest among the four *Gnetum* plastomes reported so far. It comprises a pair of large inverted repeat (IR) regions of 19 772 bp each, a large single-copy (LSC) region of 64 914 bp, and a small single-copy (SSC) region of 8791 bp. One hundred sixteen genes were predicted in this plastome, including 68 protein-coding genes, eight ribosomal RNA genes (four rRNA species), and 40 transfer RNA genes. Thirteen genes (*atpF, petB, petD, rpl2, rpl16, 3' rps12, rpoC1, trnA, trnG, trnI, trnK, trnL, and trnV*) contain one intron and only *ycf3* has two. The *G. ula* plastome has lost 20 protein-coding genes (i.e. *accD, rps16, rpl23, chlL, chlB, chlN, psaM, rpl32, rps15,* and *ndhA* to *ndhK*) that are commonly found in other seed plant plastomes. The genome-wide AT content is 61.5%. The gene density was estimated to be 1.024 (genes/kb). Previously, the plastomes of gnetophytes were demonstrated to have undergone reduction and compaction with a size of 109.5–118.9 kb and a gene density of 1.009–1.068 genes/Kb (Wu et al., 2009). Apparently, the plastome of *G. ula* falls into these ranges, indicating a plastomic wide reduction and compaction in gnetophytes. The reduction and compaction of *Gnetum* plastomes intrinsically mitigate the requirement of resources and might facilitate survival in competing with angiosperms in rainforest (Wu et al., 2009).

Fifty-six orthologous protein-coding genes from *G. ula* and other nine gymnosperm species were retrieved and aligned using MUSCLE (Edgar, 2004). Alignments of the 56 genes were concatenated to infer maximum likelihood, maximum parsimony, and neighbor-joining trees using MEGA 6.0 (Tamura et al., 2013). The resulting trees of all three methods congruently suggested that the four *Gnetum* species form a monophyletic group, in which *G. ula* is closer to the two liana species, *G. montanum* and *G. parvifolium*, than to the only small tree species of the genus, *G. gnemon* (Figure 1). The extremely long-branches leading to all three genera of gnetophytes indicate that they have evolved much faster than other gymnosperm lineages.

Correspondence: Shu-Miaw Chaw, Biodiversity Research Center, Academia Sinica, Taipei 11529, Taiwan. Tel: +886 2 27871155. Fax: +886 2 27899624. E-mail: smchaw@sinica.edu.tw

doi:10.6342/NTU201601257

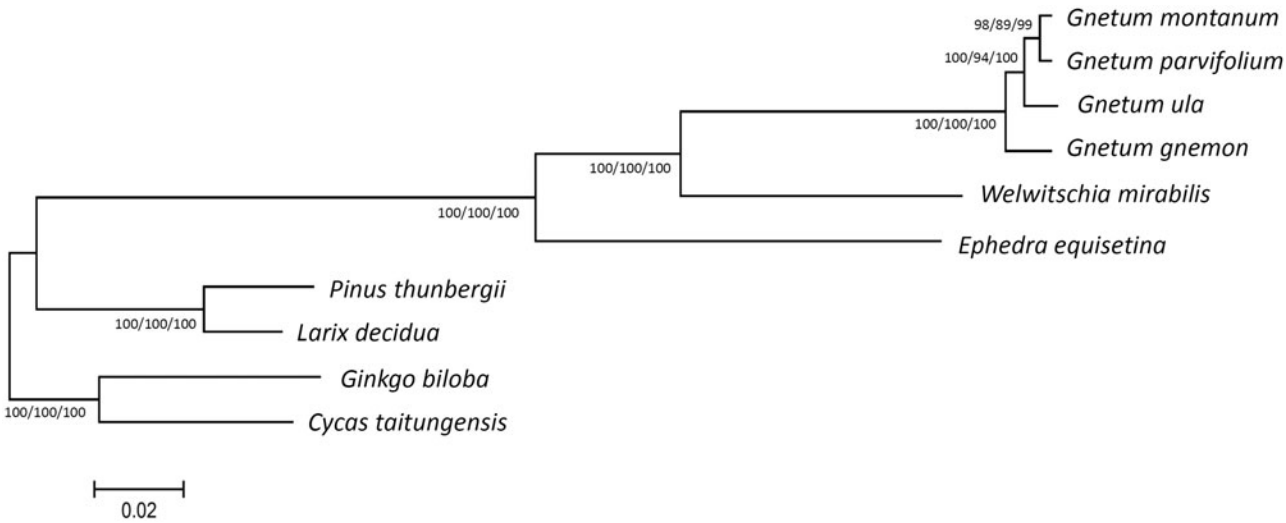

Figure 1. Phylogenetic analyses of 10 gymnosperm species. Trees were inferred from amino acid sequences of 56 concatenated chloroplast protein-coding genes using the ML method with a TJJ model, MP method, and NJ method with a Poisson model. Only the framework of the ML tree is presented. Values along branches denote bootstrap supports estimated from 1000 replicates with an arrangement of ML/MP/NJ methods. Accession numbers: *G. montanum* NC_021438, *G. parvifolium* NC_011942, *G. ula* AP014923, *G. gnemon* NC_026301, *W. mirabilis* NC_010654, *E. equisetina* NC_011954, *P. thunbergii* NC_001631, *L. decidua* NC_016058, *Ginkgo biloba* NC_016986, and *C. taitungensis* NC_009618.

Declaration of interest

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the article. This study was supported by research grants from the Ministry of Science and Technology, Taiwan (MOST 103-2621-B-001-007-MY3) and from the Investigator's Award of Academia Sinica to S.M.C.

References

Doyle JA, Donoghue MJ. (1986). Seed land phylogeny and the origin of the angiosperms: An experimental cladistic approach. Bot Rev 52: 321–431.

Edgar RC. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–7.

Friedman WE. (1998). The evolution of double fertilization and endosperm: An ''historical'' perspective. Sex Plant Reprod 11: 6–16.

Stewart Jr CN, Via LE. (1993). A rapid CTAB DNA isolation technique useful for RAPD fingerprinting and other PCR applications. Biotechniques 14:748–50.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30:2725–9.

Wu CS, Lai YT, Lin CP, Wang YN, Chaw SM. (2009). Evolution of reduced and compact chloroplast genomes (cpDNAs) in gnetophytes: Selection toward a lower-cost strategy. Mol Phylogenet Evol 52: 115–24.

doi:10.6342/NTU201601257

Birth of Four Chimeric Plastid Gene Clusters in Japanese Umbrella Pine

Chih-Yao Hsu[1,2,#], Chung-Shien Wu[1,#], and Shu-Miaw Chaw[1,*]

[1]Biodiversity Research Center, Academia Sinica, Nankang District, Taipei 11529, Taiwan

[2]Genome and Systems Biology Degree Program, National Taiwan University & Academia Sinica, Daan District, Taipei 10617, Taiwan

[#]These authors contributed equally to this work.

*Corresponding author: E-mail: smchaw@sinca.edu.tw.

Accepted: May 3, 2016

Data deposition: This project has been deposited at DDBJ under the accession AP017299.

Abstract

Many genes in the plastid genomes (plastomes) of plants are organized as gene clusters, in which genes are co-transcribed, resembling bacterial operons. These plastid operons are highly conserved, even among conifers, whose plastomes are highly rearranged relative to other seed plants. We have determined the complete plastome sequence of *Sciadopitys verticillata* (Japanese umbrella pine), the sole member of Sciadopityaceae. The *Sciadopitys* plastome is characterized by extensive inversions, pseudogenization of four tRNA genes after tandem duplications, and a unique pair of 370-bp inverted repeats involved in the formation of isomeric plastomes. We showed that plastomic inversions in *Sciadopitys* have led to shuffling of the remote conserved operons, resulting in the birth of four chimeric gene clusters. Our data also demonstrated that the relocated genes can be co-transcribed in these chimeric gene clusters. The plastome of *Sciadopitys* advances our current understanding of how the conifer plastomes have evolved toward increased diversity and complexity.

Key words: plastome, gene cluster, rearrangement, evolution, conifer.

Introduction

Due to the loss of many genes in early endosymbiosis, plastomes are reduced compared with their cyanobacterial counterparts (Ku et al. 2015). To date, plastomes have invariably retained a small handful of prokaryotic features, including organization of genes into polycistronic transcription units resembling bacterial operons (Sugiura 1992; Wicke et al. 2011). A hallmark of seed plant plastomes is the presence of two 20- to 30-Kb inverted repeats (IRs) (hereafter referred to as "typical IRs," including IR_A and IR_B), which usually contain four ribosomal RNAs. However, a few exceptions have been reported. For example, conifers—the largest gymnosperm group comprising Cupressophyta (cupressophytes) and Pinaceae—have lost a typical IR copy from their plastomes (Raubeson and Jansen 1992). Recent studies have further suggested that cupressophytes and Pinaceae might have lost different IR copies, with the former losing IR_A and the latter losing IR_B (Wu, Wang, et al. 2011; Wu and Chaw 2014).

Conifer plastomes are also characterized by extensive genomic rearrangements. The plastome of *Cryptomeria japonica*—the first completed plastome of cupressophytes (Hirao et al. 2008)—experienced at least 12 inversions after its split from the basal gymnosperm clade, cycads, whose plastomes have remained virtually unchanged for 280 million years (Wu and Chaw 2015). The co-existence of four different plastome forms among Pinaceae genera is associated with intra-plastomic recombination mediated by three specific types of short IRs (Wu, Lin, et al. 2011). Furthermore, *Cephalotaxus oliveri* (Cephalotaxaceae; Yi et al. 2013) and four *Juniperus* species (Cupressaceae; Guo et al. 2014) harbor isomeric plastomes that deviate from each other by an inversion possibly triggered by a *trnQ*-containing IR ("*trnQ*-IR"). Although conifer plastomes are highly rearranged (Wu and Chaw 2014), disruptions in their operons are rare. Until recently, only one case was reported in the plastome of *Taxus mairei*, in which the S10 operon (*trnI-rpoA* region) was disrupted into two separate segments by a fragment of approximately 15 Kb (Hsu et al. 2014). However, the impact of such operon disruptions on plastid evolution remains poorly understood.

Cupressophytes comprise five conifer families: Araucariaceae, Cupressaceae, Podocarpaceae, Sciadopityaceae, and Taxaceae. Among them, Sciadopityaceae is the only family with the single species *Sciadopitys verticillata* (hereafter referred to as *Sciadopitys*). *Sciadopitys* is an evergreen tree that can reach 27 m tall. Its spectacular needle-like leaves are arranged in whorls, like an umbrella. Thus, *Sciadopitys* is commonly called the Japanese umbrella pine. Three genome-based (Chaw et al. 2000) and plastome-based (Rai et al. 2008) phylogenetic studies are congruent in placing *Sciadopitys* as a sister to Taxaceae and Cupressaceae. Recent molecular dating suggests that *Sciadopitys* diverged from other cupressophytes more than 200 million years ago (Crisp and Cook 2011). Although *Sciadopitys* is considered a living fossil endemic to Japan, paleo-biogeographic evidence indicates that its ancestors existed in China during the early and middle Jurassic (Jiang et al. 2012).

The 25 published cupressophyte plastomes available on GenBank (December, 2015) represent four of the five cupressophyte families but no complete plastome is available for Sciadopityaceae. As a part of our continuing efforts to decipher the diversity and evolution of conifer plastomes, we have completed and elucidated the plastome sequence of *Sciadopitys*. We found that the plastome of *Sciadopitys* is characterized by several unusual features, including shuffling of the conserved plastid operons and re-organization of plastid genes into new chimeric gene clusters.

Materials and Methods

DNA Extraction

Approximately 2 g of fresh leaves were collected from an individual of *Sciadopitys verticillata* (voucher chaw 1496) growing in the Floriculture Experiment Center, Taipei, Taiwan. The voucher specimen was deposited in the herbarium of Biodiversity Research Center, Academia Sinica, Taipei (HAST). Total DNA of the leaves was extracted with 2X CTAB buffers (Stewart and Via 1993). The extracted DNA was qualified with a threshold of DNA concentration >300 ng/μl, 260/280 = 1.8–2.0 and 260/230 $>$ 1.7.

Sequencing, Plastome Assembly, and Genome Annotation

Sequencing was conducted on an Illumina MiSeq Sequencing System (Illumina, San Diego, CA) in Yourgene Bioscience (New Taipei City, Taiwan) to yield 300-bp paired-end reads of approximately 4 Gb. De novo assembly of the *Sciadopitys* plastome was performed using CLC Genomics Workbench 4.9 (CLC Bio, Arhus, Denmark). Plastid genes were predicted using DOGMA (Wyman et al. 2004) and tRNAscan-SE 1.21 (Schattner et al. 2005) with the default option that real tRNA genes should have $>=20$ Cove scores. Boundaries of predicted genes were manually adjusted by aligning them with their orthologs of other gymnosperms.

Estimate of Dispersed Repeats and Plastomic Inversions

Repeat sequences were searched by comparing the plastome against itself using NCBI BLASTn with the default settings, followed by manually deleted overlapping or conjoined pairs. To assess the possible scenarios of plastomic inversions in *Sciadopitys*, the plastome of *Cycas taitungensis* (NC_009618) with its IR_A removed was used for comparison. We identified the syntenic block of genes between *Sciadopitys* and *Cycas* using Mauve 2.3.1 (Darling et al. 2004). The resulting matrix of syntenic blocks was utilized to estimate the minimal inversion steps with MGR 2.0.1 (Bourque and Pevzner 2002). The plastome map of *Sciadopitys* was drawn using Circos 0.67 (Krzywinski et al. 2009).

Detection of Isomeric Plastomes

Primer pairs listed in supplementary table S1 (Supplementary Material online) were used to amplify DNA fragments specific to the two isomeric plastomes in *Sciadopitys* (i.e., rpl33 + rpoC2 and rpoC1 + rps18 for the presence of A form; rpl33 + rps18 and rpoC1 + rpoC2 for the presence of B form). PCR reactions were conducted with three different numbers of cycles. The conditions were 94 °C for 5 min, followed by 25, 30, or 35 cycles of 94 °C for 20 s, 55 °C for 20 s, and 72 °C for 2 min, and an extension of 72 °C for 10 min.

Detection of RNA Transcripts in Chimeric Gene Clusters

Total RNA was extracted from fresh leaves of *Sciadopitys* according to a modified RNA isolation protocol (Kolosova et al. 2004). We employed a RevertAid H Minus First Strand cDNA Synthesis Kit (Thermo Fisher Scientific, Waltham) to synthesize the first-strand cDNA with four specific primers (SpsbN, SatpA, SrpoC2, and Srps18 in supplementary table S1, Supplementary Material online). PCR reactions were conducted with the synthesized cDNA and four pairs of specific primers (atpF-1 + psbT for a 358-bp fragment; psbB-2 + atpA-2 for a 565-bp fragment; rpl33 + rpoC2 for a 687-bp fragment; rpoC1 + rps18 for a 939-bp fragment). The PCR conditions were 94 °C for 5 min, followed by 30 cycles at 94 °C for 20 s, 60 °C for 20 s, and 72 °C for 1 min, and an extension at 72 °C for 10 min.

Results and Discussion

Loss of IR_A from *Sciadopitys* Plastome

The plastome of *Sciadopitys verticillata* (AP017299) is illustrated as a circular molecule that consists of 138,309 bp (fig. 1). It is the largest among the known plastomes of Cupressales (including Sciadopityaceae, Taxaceae *s. l.*, and Cupressaceae *s. l.*). Flanking and adjacent genes of the typical IRs are informative markers for inferring the intact (or retained) IR copy in conifer plastomes. For example, the boundary of IR_A or IR_B is adjacent to the *psbA* or S10 operon (i.e., *trnI-rpoA* region), respectively (Wu, Wang,

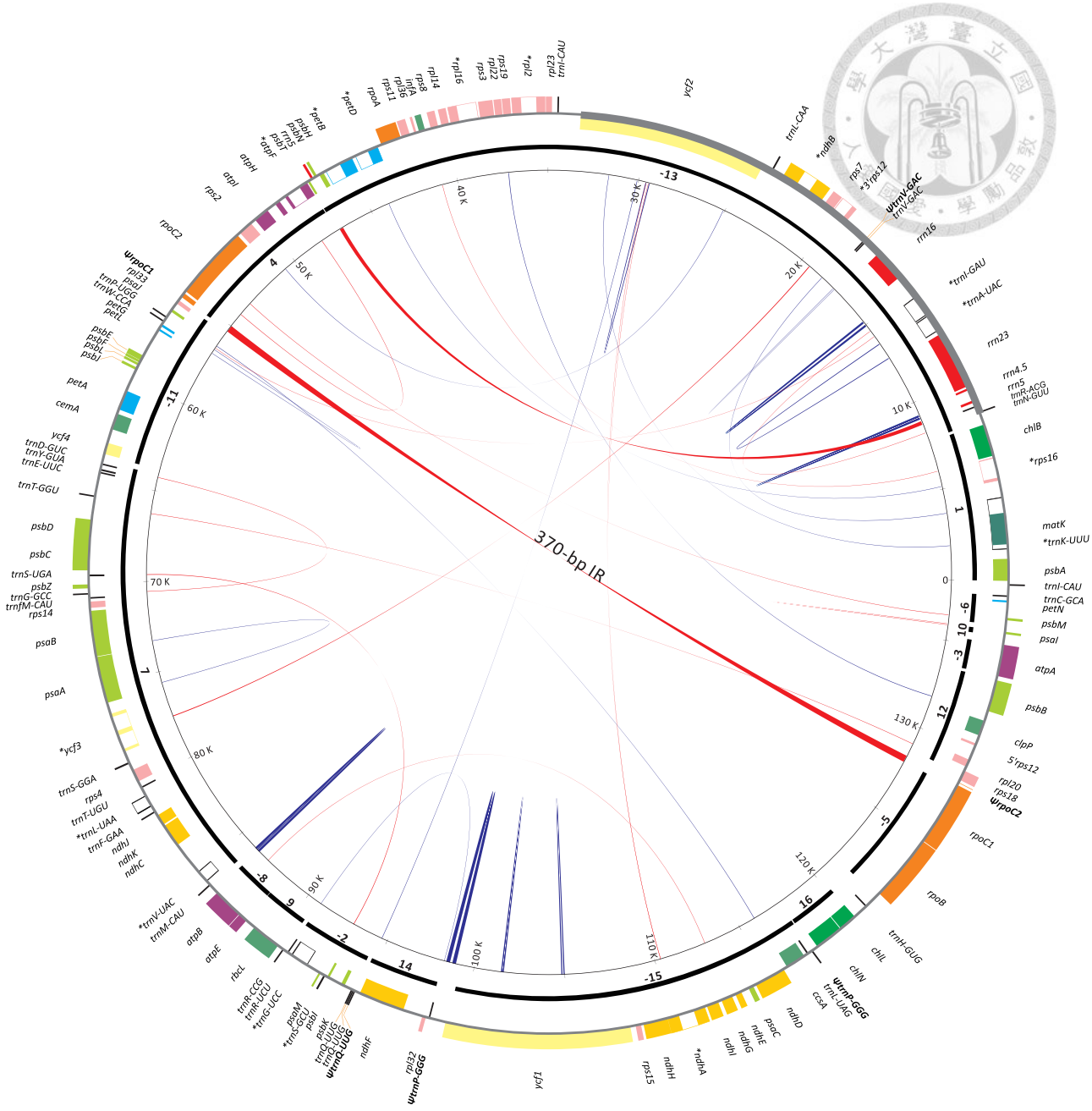

Fig. 1.—Plastome map of *Sciadopitys verticillata*. Colored boxes represent genes with counterclockwise (outer boxes) or clockwise (inner boxes) transcriptional directions. Thick and thin grey lines are IRB and single-copy region as compared to those of *Cycas*, separately. Syntenic blocks of genes between *Cycas* and *Sciadopitys* are depicted by thick black bars with Arabic numerals, where pluses or minuses indicate the corresponding syntenic blocks with the same or opposite directions between the two species, respectively. Pairs of dispersed repeats are connected by blue (direct repeats) or red (inverted repeats) lines, with their width proportional to the repeat size. Pseudogenes are bold and marked with a "Ψ." Intron-containing genes are indicated with an "*".

et al. 2011). In the *Sciadopitys* plastome, the retained typical IR copy, which encompasses the region from *trnN-GUU* to *ycf2*, is adjacent to the S10 operon (fig. 1), indicating that it should be IRB. In other words, the lost IR copy is IRA. This observation reinforces the hypothesis that cupressophytes have lost IRA rather than IRB (Wu, Wang, et al. 2011; Wu and Chaw 2014).

The plastome of *Sciadopitys* contains a total of 121 genes, 83 of which are protein-coding genes and the rest are structural RNA genes (supplementary table S2, Supplementary Material online). Sixteen genes contain introns, but the intron of *rpoC1* has been lost. Three genes, *rrn5*, *trnI-CAU*, and *trnQ-UUG*, have two copies. The duplicated *rrn5* is located in the region between *psbN* and *psbT* (fig. 1).

The plastomes of both *Agathis dammara* and *Wollemia nobilis* (Araucariaceae) also have a duplicated *rrn5*, but it is in the region between *psbB* and *clpP* (Yap et al. 2015). Among the elucidated cupressophyte plastomes available on GenBank, only these three taxa contain two copies of plastid *rrn5*. Because *Sciadopitys* (Sciadopityaceae) and Araucariaceae differ in the locations of their extra *rrn5*, it is most parsimonious that duplications of *rrn5* took place independently in the two families. Moreover, *Sciadopitys* has lost its plastid *accD*, which possibly has been transferred to the nucleus (Li et al. 2016).

Pseudogenization of Four tRNA Genes after Tandem Duplications

Four pseudo tRNA genes (*Ψtrn*V-GAC, *Ψtrn*Q-UUG, and two copies of *Ψtrn*P-GGG) were detected in the *Sciadopitys* plastome (fig. 1). Although the Cove scores for both *Ψtrn*V-GAC (score = 28.41) and *Ψtrn*Q-UUG (score = 26.8) were marginal, we predicted them as pseudogenes because their sequences could not form proper cloverleaf structures. Both *Ψtrn*V-GAC and *Ψtrn*Q-UUG are near their functional paralogs, implying that pseudogenization of these two genes might have occurred after tandem duplications. In angiosperms, the plastid *trn*P-GGG likely has been lost for 150 MY (Chaw et al. 2004). In contrast, this tRNA gene is retained and commonly located in the region between *trn*L and *rpl32* in *Cycas*, *Ginkgo*, *Gnetum*, and *Pinus* (Wu et al. 2007), and other cupressophyte families, such as Araucariaceae (Yap et al. 2015), Podocarpaceae (Vieira Ldo et al. 2014; Wu and Chaw 2014), and Taxaceae *s. l.* (Yi et al. 2013; Hsu et al. 2014). In the *Sciadopitys* plastome, two *Ψtrn*P-GGG copies are separated by a distance of approximately 20-Kb: one is adjacent to *trn*L-UAG and the other is located near *rpl32* (fig. 1). Therefore, the two *Ψtrn*P-GGG copies might have resulted from tandem duplications, and subsequently, a plastomic inversion of an approximately 20-Kb fragment separated them.

Evolution of the Plastid *trn*I-CAU in *Sciadopitys*

Sciadopitys has two plastid *trn*I-CAU copies, one between *trn*C-GCA and *psb*A, and the other between *ycf2* and *rpl23* (figs. 1 and 2). In *Cryptomeria*, one of the two plastid *trn*I-CAU copies was considered residual from the lost typical IR (Hirao et al. 2008). Indeed, the majority of cupressophyte plastomes have two *trn*I-CAU copies with the sequence identity higher than 85% (table 1), connoting their homologous origin.

In *Sciadopitys* plastome, both copies of *trn*I-CAU are able to fold into cloverleaf structures, but they differ in prediction scores. The gene located between *ycf2* and *rpl23* has a score of 78.1 bits (fig. 2*A*), much higher than the score of the other gene (score = 48.5 bits; fig. 2*B*). Ten nucleotide substitutions (highlighted in gray in fig. 2) were detected between the two *trn*I-CAU copies, including three mismatches and four U•G abnormal pairings in the stems of the low-scoring *trn*I-CAU (fig. 2*B*). Although *trn*I-CAU is essential for plastid biology (Alkatib et al. 2012), why two copies of *trn*I-CAU are required remains to be investigated. The elucidated cupressophyte plastomes, such as *Cephalotaxus*, *Nageia*, and *Podocarpus*, contain only one copy of *trn*I-CAU (table 1). Therefore, whether the low-scoring *trn*I-CAU of *Sciadopitys* is functionally redundant and subjected to relaxed structural constraint is worthy of further investigation.

Presence of Two Isomeric Plastomes in *Sciadopitys*

Recent studies of conifer plastomes revealed that dispersed short IRs can trigger rearrangements to generate isomeric forms. In Pinaceae, a shift between different plastomic forms is often associated with homologous recombination (HR) mediated by the short IRs of approximately 949 bp (Tsumura et al. 2000; Wu, Lin, et al. 2011). The short IRs that contain *trn*Q-UUG (*trn*Q-IR) can also promote the formation of isomeric plastomes in *Cephalotaxus* (Yi et al. 2013) and *Juniperus* (Guo et al. 2014).

Thirty-seven pairs of dispersed repeats were detected in the *Sciadopitys* plastome. Among them, the longest IR pair is 370 bp and contains the sequences of 3′*rpoC1* and 5′*rpoC2* (fig. 1). In the *Sciadopitys* plastome, only this 370-bp IR pair is longer than the 250-bp "*trn*Q-IR" of *Juniperus* (Guo et al. 2014). Hence, if the 370-bp IR is able to mediate HR in *Sciadopitys*, we would expect the presence of two plastomic forms, as depicted in figure 3. The plastome form illustrated in figure 1 is designated as the A form, and the other is the B form. We have verified the presence of both the A and B forms by amplicons of four specific DNA fragments across the 370-bp IR in the PCR with 35 cycles (fig. 3). However, the amount of the PCR amplicons differs between the two forms. With 25 PCR cycles, the two specific amplicons of the A form are evident, but those of the B form are undetectable (fig. 3). These results suggest that A form is predominant in *Sciadopitys* plastome populations, in agreement with our assembly results.

The plastomes of Cupressaceae and Taxaceae possess a *trn*Q-IR (Guo et al. 2014). Nonetheless, such a *trn*Q-IR is absent from the plastome of *Sciadopitys*. The plastomes of Araucariaceae (Yap et al. 2015) have an IR pair that is approximately 600-bp long and contains the gene *rrn5*. Such the length of IRs could potentially trigger HR; however, the presence of associated isomeric plastomes has not been experimentally demonstrated in Araucariaceae. Including the unique 370-bp IR of *Sciadopitys*, it is apparent that in cupressophytes, the presence of isomeric plastomes is overwhelming and associated with diverse short IRs.

Birth of Four Chimeric Gene Clusters

We identified a total of 16 syntenic blocks between *Cycas* and *Sciadopitys* plastomes (fig. 1 and supplementary fig. S1, Supplementary Material online). In addition to the loss of IR_A from the *Sciadopitys* plastome mentioned above, eight

Table 1

Presence of *trnI-CAU* copies in the plastomes of cupressophytes

Family	Species (GenBank accession)	Copy[a]	Seq. identity (%)[b]	Score
Cupressaceae	*Calocedrus formosana* (NC_023121)	(+)*trnI-CAU*	98.63	70.17
		(-)*trnI-CAU*		75.14
	Juniperus bermudiana (NC_024021)	(+)*trnI-CAU*	98.63	70.17
		(-)*trnI-CAU*		75.44
	Cryptomeria japonica (NC_010548)	(+)*trnI-CAU*	98.63	77.3
		(-)*trnI-CAU*		74.56
	Metasequoia glyptostroboides (NC_027423)	(+)*trnI-CAU*	98.63	69.43
		(-)*trnI-CAU*		74.71
	Cunninghamia lanceolate (NC_021437)	(+)*trnI-CAU*	100	75.44
		(-)*trnI-CAU*		75.44
	Taiwania cryptomerioides (NC_016065)	(+)*trnI-CAU*	100	75.44
		(-)*trnI-CAU*		75.44
Taxaceae	*Taxus mairei* (AP014575)	(+)*trnI-CAU*	98.63	75.44
		(-)*trnI-CAU*		75.39
	Amentotaxus formosana (NC_024945)	(+)*trnI-CAU*	98.63	75.44
		(-)*trnI-CAU*		75.67
Cephalotaxaceae	*Cephalotaxus wilsoniana* (NC_016063)	(+)*trnI-CAU*	—	77.31
		—		
Sciadopityaceae	*Sciadopitys verticillata* (AP017299)	(+)*trnI-CAU*	86.30	48.48
		(+)*trnI-CAU*		78.15
Podocarpaceae	*Nageia nagi* (NC_023120)	(+)*trnI-CAU*	—	75.44
		—		
	Podocarpus lambertii (NC_023805)	(+)*trnI-CAU*	—	77.3
		—		
Araucariaceae	*Agathis dammara* (NC_023119)	(+)*trnI-CAU*	97.26	75.25
		(-)*trnI-CAU*		66.46
	Wollemia nobilis (NC_027235)	(+)*trnI-CAU*	94.52	75.25
		(-)*trnI-CAU*		55.15

[a]"+" and "-" in parentheses denote the transcriptional directions.

[b]Estimates based on comparison of the two *trnI-CAU* copies within each species.

plastomic inversions were detected to distinguish *Sciadopitys* from *Cycas* (supplementary fig. S1, Supplementary Material online). Since *Cycas* was proposed to retain the ancestral gene order of seed plant plastomes (Jansen and Ruhlman 2012), these eight inversions should have occurred after cupressophytes split from cycads.

In *Sciadopitys*, plastomic inversions have also disrupted four operons that are generally conserved among seed plants. These disrupted operons are *rps2–atpI–atpH–atpF–atpA* (hereafter, *rps2* operon), *psbB–psbT–psbH–petB–petD* (*psbB* operon), *rpoB–proC1–rpoC2* (*rpoB* operon), and *petL–petG–psaJ–rpl33–rps18* (*petL* operon) (fig. 4*A* and *B*). On one hand, recombination between *rps2* and *psbB* operons is associated with inversion 8 (supplementary fig. S1, Supplementary Material online), creating the *rps2–petD* and *psbB–atpA* gene clusters (fig. 4*A*). On the other hand, inversion 4 recombined the *rpoB* and *petL* operons and then generated the *petL–rpoC2* and *rpoB–rps18* gene clusters (fig. 4*B*). Most genes in each of the four chemic gene clusters have the same transcriptional direction (fig. 4*A* and *B*). Therefore, we postulated that genes in these chemic gene clusters might be co-transcribed. To verify this, we performed RT-PCR assays with specific primers designed from genes near the junction between different operon-derived segments.

As shown in figure 4*C*, our RT-PCR results indicate that (1) there was no DNA contamination in the assayed RNA because all the negative controls failed to yield any product; and (2) the expected size of amplicons was clearly detected in all experimental sets. These data suggest that shuffling between different operons could lead to the birth of new co-transcription units in plastids.

Disruptions of conserved plastid operons have been only reported in a few taxa, such as *Vigna* (Perry et al. 2002), *Trifolium* (Cai et al. 2008), *Trachelium* (Haberle et al. 2008), some genera of Geraniaceae (Guisinger et al. 2011), *Taxus* (Hsu et al. 2014), and *Sciadopitys* (this study). These taxa also have highly rearranged plastomes. Except for *Vigna* and *Sciadopitys*, none of the above taxa has experienced recombination between operons. In the *Vigna* plastome, recombination between two homologous operons, S10A and S10B, has led to the re-organization of genes in the operons (Perry et al. 2002). Nonetheless, novel chimeric gene clusters created

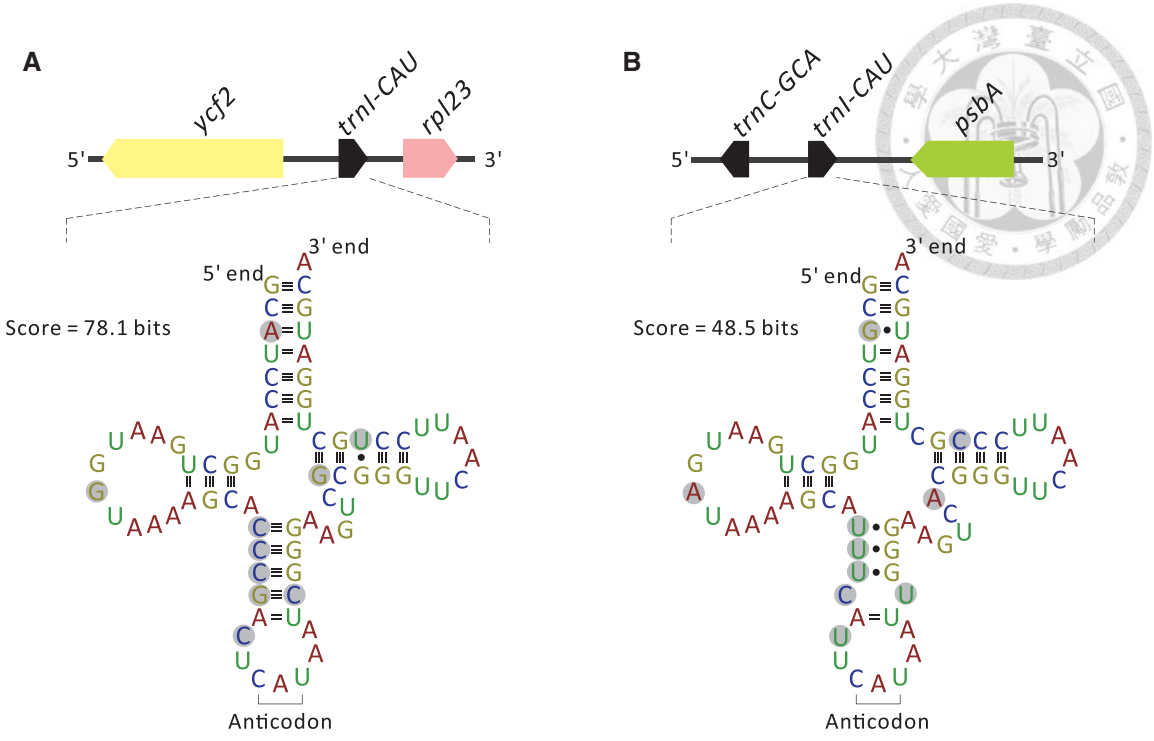

Fig. 2.—Comparison between the two copies of *trnI-CAU* in the *Sciadopitys* plastome. (A) Predicted cloverleaf structure of *trnI-CAU* located between *ycf2* and *rpl23* and (B) that of the other located between *trnC-GCA* and *psbA*. Pairwise substitutions of nucleotides between the two copies are highlighted in gray.

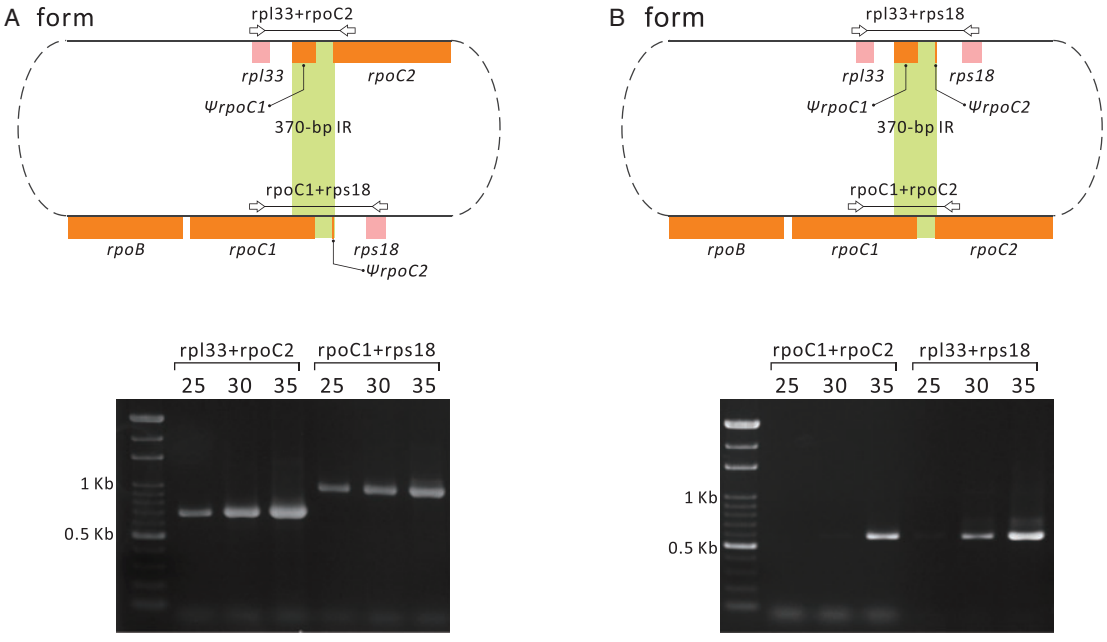

Fig. 3.—Co-existence of two isomeric plastomes in *Sciadopitys*. The A form is the plastome map obtained from our genome assembly and is shown in fig. 1. The B form differs from the A form by an inversion of the *rpoC2–rps18* (or *rpl33–rpoC1*) fragment. Light green areas are the 370-bp IRs involved in homologous recombination that allows for conversion between the two forms. Paired open arrows are primers specific for the PCR amplification of each form. The corresponding PCR amplicons are shown, and the numbers above each lane of gel photos denote the PCR cycles conducted.

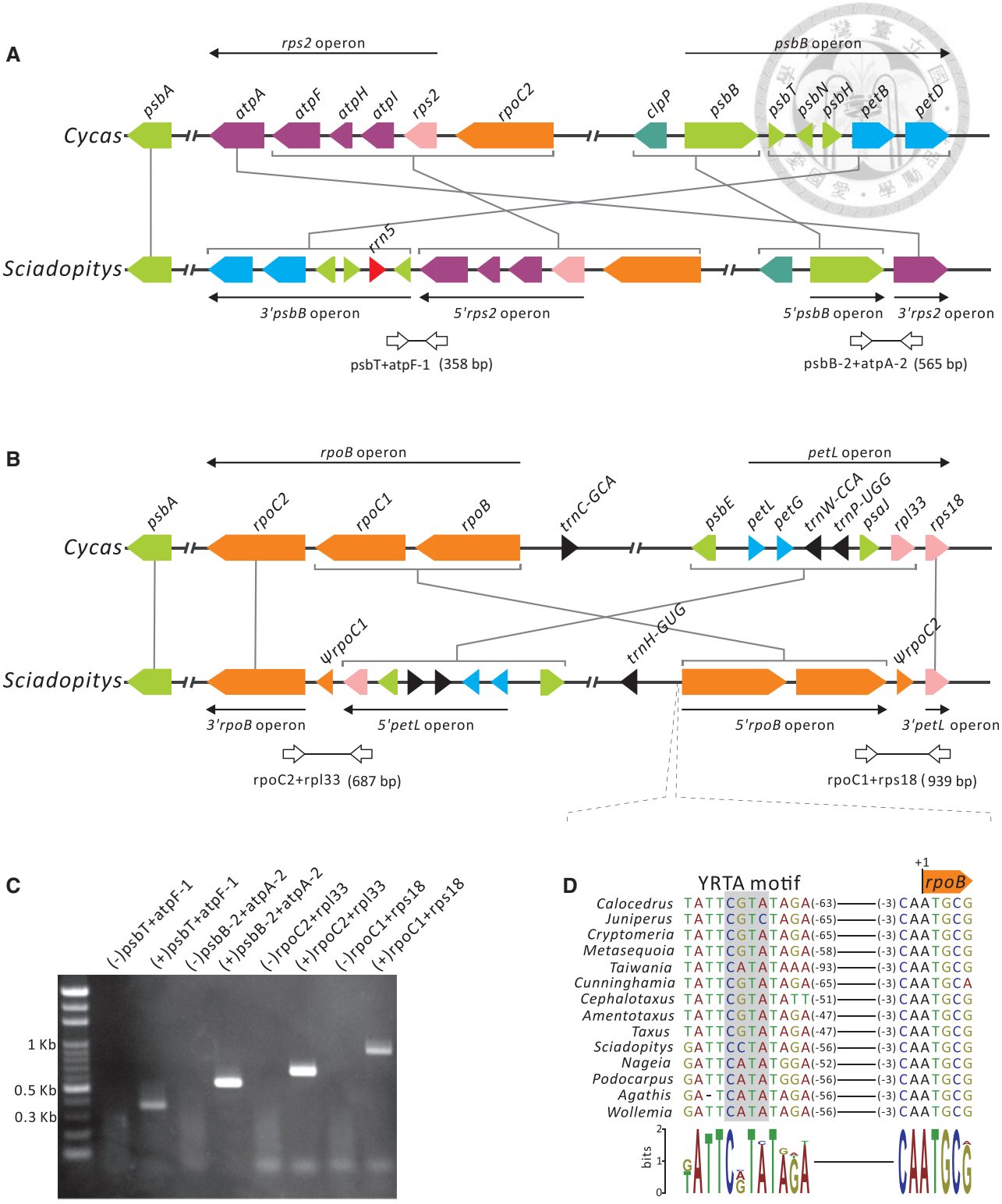

Fig. 4.—Birth of chimeric gene clusters in the *Sciadopitys* plastome. (A) Shuffling between *rps2* and *psbB* operons and (B) between *rpoB* and *petL* operons. Syntenic genes in the corresponding operons of *Cycas* are used as references. Operons and their transcriptional directions are indicated by solid arrows. Syntenic blocks of genes are connected with gray lines. Paired open arrows are primers for amplifying cDNA fragments across junctions between two recombined operons. The expected sizes of amplicons are shown in parentheses. (C) RT-PCR analysis for detecting the transcripts comprising the genes originated from different operons. Primer pairs used for RT-PCR assays are shown above the gel panel, with minus and plus signs (in parentheses) denoting the use of RNA (negative control) and cDNA (experimental set) as templates, respectively. (D) YRTA motif of the NEP promoter upstream of *rpoB*.

by shuffling between heterologous operons (fig. 4) are documented for the first time in the present study.

Evolutionary Impacts of Novel Chimeric Gene Clusters

The chimeric gene clusters of *Sciadopitys* provide two novel insights into the evolution of plastomes. First, other than the gene cluster *rpoB–rpoC1–rps18*, the remaining three chimeric gene clusters do not alter their upstream regions, as the neighboring genes of their 5′ regions are the same as those of *Cycas* (fig. 4*A* and *B*). This finding suggests that the promoter sequences of these gene clusters have not been altered after the associated inversions taken place. Figure 4*D* shows that the upstream sequence of *rpoB* harbors an YRTA motif of the nuclear-encoded RNA polymerase (NEP) promoter (Shiina et al. 2005). Furthermore, genes of different origins are able to be co-transcribed in the chimeric gene cluster (fig. 4*C*). Therefore, we cannot rule out the possibility that the pre-existing promoters are adopted for transcription of the genes in these chimeric gene clusters.

Second, shuffling between *rpoB* and *petL* operons (fig. 4*B*) has relocated *rpoC2* to join the segment of the 5′*petL* operon whose transcription is associated with the plastid RNA polymerase (PEP) promoter (Finster et al. 2013). *RpoC2* codes for one of the core units of PEP (Hu and Bogorad 1990). If the chimeric gene cluster *petL–petG–psaJ–rpl33–rpoC2* is exclusively transcribed by PEP, we would not expect any transcript of this gene cluster in *Sciadopitys*. Nonetheless, its associated transcript was observed in figure 4*C*. Two possibilities might account for the presence of this transcript: (1) the isomeric plastome of the B form (fig. 3) that contains an intact *rpoB* operon provides RPOC2 proteins; (2) an alternative promoter has evolved to perform transcription because many plastid genes are transcribed by both the NEP and PEP promoters (Börner et al. 2015).

Conclusion

The plastome of *Sciadopitys* is characterized by several unusual features, such as the loss of the typical IR_A copy, the duplication and pseudogenization of four tRNAs, extensive genomic inversions, the presence of isomeric plastomes, and chimeric gene clusters derived from shuffling of remote operons. All these characteristics highlight the fact that the evolution of plastomes may be more complicated than previously thought. The highly rearranged plastome of *Scidaopitys* advances our understanding of the dynamics, complexity, and evolution of plastomes in conifers.

Supplementary Material

Supplementary figure S1 and table S1 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

The authors thank Shu-Mei Liu for filling sequence gaps. This work was supported by research grants from the Ministry of Science and Technology, Taiwan (MOST 103-2621-B-001-007-MY3), and from the Investigator's Award of Academia Sinica (2011–2015) to S.-M.C. Special thanks are due to one anonymous reviewer for critical reading and helpful comments. The authors also like to thank Dr. Robert Jansen for valuable suggestions that improved the manuscript.

Literature Cited

Alkatib S, Fleischmann TT, Scharff LB, Bock R. 2012. Evolutionary constraints on the plastid tRNA set decoding methionine and isoleucine. Nucleic Acids Res. 40:6713–6724.

Bourque G, Pevzner PA. 2002. Genome-scale evolution: reconstructing gene orders in the ancestral species. Genome Res. 12:26–36.

Börner T, Aleynikova AY, Zubo YO, Kusnetsov VV. 2015. Chloroplast RNA polymerases: role in chloroplast biogenesis. Biochim Biophys Acta. 1847:761–769.

Cai Z, et al. 2008. Extensive reorganization of the plastid genome of *Trifolium subterraneum* (Fabaceae) is associated with numerous repeated sequences and novel DNA insertions. J Mol Evol. 67:696–704.

Chaw SM, Chang CC, Chen HL, Li WH. 2004. Dating the monocot–dicot divergence and the origin of core eudicots using whole chloroplast genomes. J Mol Evol. 58:18–11.

Chaw SM, Parkinson CL, Cheng Y, Vincent TM, Palmer JD. 2000. Seed plant phylogeny inferred from all three plant genomes: monophyly of extant gymnosperms and origin of Gnetales from conifers. Proc Natl Acad Sci U S A. 97:4086–4091.

Crisp MD, Cook LG. 2011. Cenozoic extinctions account for the low diversity of extant gymnosperms compared with angiosperms. New Phytol. 192:997–1009.

Darling AC, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 14:1394–1403.

Finster S, Eggert E, Zoschke R, Weihe A, Schmitz-Linneweber C. 2013. Light-dependent, plastome-wide association of the plastid-encoded RNA polymerase with chloroplast DNA. Plant J. 76:849–860.

Guisinger MM, Kuehl JV, Boore JL, Jansen RK. 2011. Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: rearrangements, repeats, and codon usage. Mol Biol Evol. 28:583–600.

Guo W, et al. 2014. Predominant and substoichiometric isomers of the plastid genome coexist within *Juniperus* plants and have shifted multiple times during cupressophyte evolution. Genome Biol Evol. 6:580–590.

Haberle RC, Fourcade HM, Boore JL, Jansen RK. 2008. Extensive rearrangements in the chloroplast genome of *Trachelium caeruleum* are associated with repeats and tRNA genes. J Mol Evol. 66:350–361.

Hirao T, Watanabe A, Kurita M, Kondo T, Takata K. 2008. Complete nucleotide sequence of the *Cryptomeria japonica* D. Don. chloroplast genome and comparative chloroplast genomics: diversified genomic structure of coniferous species. BMC Plant Biol. 8:70.

Hsu CY, Wu CS, Chaw SM. 2014. Ancient nuclear plastid DNA in the yew family (taxaceae). Genome Biol Evol. 6:2111–2121.

Hu J, Bogorad L. 1990. Maize chloroplast RNA polymerase: the 180-, 120-, and 38-kilodalton polypeptides are encoded in chloroplast genes. Proc Natl Acad Sci U S A. 87:1531–1535.

Jansen RK, Ruhlman TA. 2012. Plastid genomes of seed plants. In: Bock R, Knoop V, editors. Genomics of chloroplasts and mitochondria. Netherlands: Springer. p. 103–126.
Jiang ZK, Wang YD, Zheng SL, Zhang W, Tian N. 2012. Occurrence of Sciadopitys-like fossil wood (Conifer) in the Jurassic of western Liaoning and its evolutionary implications. Chin Sci Bull. 57:569–572.
Kolosova N, et al. 2004. Isolation of high-quality RNA from gymnosperm and angiosperm trees. Biotechniques 36:821–824.
Krzywinski M, et al. 2009. Circos: an information aesthetic for comparative genomics. Genome Res. 19:1639–1645.
Ku C, et al. 2015. Endosymbiotic origin and differential loss of eukaryotic genes. Nature 524:427–432.
Li J, et al. 2016. Evolution of short inverted repeat in cupressophytes, transfer of *accD* to nucleus in *Sciadopitys verticillata* and phylogenetic position of Sciadopityaceae. Sci Rep. 6:20934.
Perry AS, Brennan S, Murphy DJ, Kavanagh TA, Wolfe KH. 2002. Evolutionary re-organisation of a large operon in adzuki bean chloroplast DNA caused by inverted repeat movement. DNA Res. 9:157–162.
Rai HS, Reeves PA, Peakall R, Olmstead RG, Graham SW. 2008. Inference of higher-order conifer relationships from a multi-locus plastid data set. Botany 86:658–669.
Raubeson LA, Jansen RK. 1992. A rare chloroplast DNA structure mutation is shared by all conifers. Biochem Syst Ecol. 20:17–24.
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:W686–W689. (Web Server issue)
Shiina T, Tsunoyama Y, Nakahira Y, Khan MS. 2005. Plastid RNA polymerases, promoters, and transcription regulators in higher plants. Int Rev Cytol. 244:1–68.
Stewart CN Jr, Via LE. 1993. A rapid CTAB DNA isolation technique useful for RAPD fingerprinting and other PCR applications. Biotechniques 14:748–750.
Sugiura M. 1992. The chloroplast genome. Plant Mol Biol. 19:149–168.
Tsumura Y, Suyama Y, Yoshimura K. 2000. Chloroplast DNA inversion polymorphism in populations of *Abies* and *Tsuga*. Mol Biol Evol. 17:1302–1312.
Vieira Ldo N, et al. 2014. The complete chloroplast genome sequence of *Podocarpus lambertii*: genome structure, evolutionary aspects, gene content and SSR detection. PLoS One 9:e90618.
Wicke S, Schneeweiss GM, dePamphilis CW, Müller KF, Quandt D. 2011. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Mol Biol. 76:273–297.
Wu CS, Chaw SM. 2014. Highly rearranged and size-variable chloroplast genomes in conifers II clade (cupressophytes): evolution towards shorter intergenic spacers. Plant Biotechnol J. 12:344–353.
Wu CS, Chaw SM. 2015. Evolutionary stasis in cycad plastomes and the first case of plastome GC-biased gene conversion. Genome Biol Evol. 7:2000–2009.
Wu CS, Lin CP, Hsu CY, Wang RJ, Chaw SM. 2011. Comparative chloroplast genomes of Pinaceae: insights into the mechanism of diversified genomic organizations. Genome Biol Evol. 3:309–319.
Wu CS, Wang YN, Hsu CY, Lin CP, Chaw SM. 2011. Loss of different inverted repeat copies from the chloroplast genomes of Pinaceae and cupressophytes and influence of heterotachy on the evaluation of gymnosperm phylogeny. Genome Biol Evol. 3:1284–1295.
Wu CS, Wang YN, Liu SM, Chaw SM. 2007. Chloroplast genome (cpDNA) of *Cycas taitungensis* and 56 cp protein-coding genes of *Gnetum parvifolium*: insights into cpDNA evolution and phylogeny of extant seed plants. Mol Biol Evol. 24:1366–1379.
Wyman SK, Jansen RK, Boore JL. 2004. Automatic annotation of organellar genomes with DOGMA. Bioinformatics 20:3252–3255.
Yi X, Gao L, Wang B, Su YJ, Wang T. 2013. The complete chloroplast genome sequence of *Cephalotaxus oliveri* (Cephalotaxaceae): evolutionary comparison of *Cephalotaxus* chloroplast DNAs and insights into the loss of inverted repeat copies in gymnosperms. Genome Biol Evol. 5:688–698.
Yap JY, et al. 2015. Complete chloroplast genome of the wollemi pine (*Wollemia nobilis*): structure and evolution. PLoS One 10:e0128126.

Associate editor: Bill Martin