Conflicting phylogenetic signals in plastomes of the tribe Laureae (Lauraceae)

Tian-Wen Xiao 1, 2, Yong Xu 1, 2, Lu Jin 1, 2, Tong-Jian Liu 1, Haifei Yan 1, Xuejun Ge Corresp. 1

1 Guangdong Provincial Key Laboratory of Applied Botany and Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, Guangdong, People's Republic of China
2 University of Chinese Academy of Sciences, Beijing, People's Republic of China

Corresponding Author: Xuejun Ge
Email address: xjge@scbg.ac.cn

Background. Gene tree discordance is common in phylogenetic analyses. Many phylogenetic studies have excluded non-coding regions of the plastome without evaluating their impact on tree topology. In general, plastid loci have often been treated as a single unit, and tree discordance among these loci has seldom been examined. Using samples of Laureae (Lauraceae) plastomes, we explored plastome variation among the tribe, examined the influence of non-coding regions on tree topology, and quantified intra-plastome conflict.

Results. We found that the plastomes of Laureae have low inter-specific variation and are highly similar in structure, size, and gene content. Laureae was divided into three groups, subclades I, II and III. The inclusion of non-coding regions changed the phylogenetic relationship among the three subclades. Topologies based on coding and non-coding regions were largely congruent except for the relationship among subclades I, II and III. By measuring the distribution of phylogenetic signal across loci that supported different topologies, we found that nine loci (two coding regions, two introns and five intergenic spacers) played a critical role at the contentious node.

Conclusions. Our results suggest that subclade III and subclade II are successively sister to subclade I. Conflicting phylogenetic signals exist between coding and non-coding regions of Laureae plastomes. Our study highlights the importance of evaluating the influence of non-coding regions on tree topology and emphasizes the necessity of examining discordance among different plastid loci in phylogenetic studies.
Conflicting phylogenetic signals in plastomes of the tribe Laureae (Lauraceae)

Tian-Wen Xiao\textsuperscript{1,2}, Yong Xu\textsuperscript{1,2}, Lu Jin\textsuperscript{1,2}, Tong-Jian Liu\textsuperscript{1}, Hai-Fei Yan\textsuperscript{1}, Xue-Jun Ge\textsuperscript{1}

\textsuperscript{1} Guangdong Provincial Key Laboratory of Applied Botany and Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, Guangdong, China.

\textsuperscript{2} University of Chinese Academy of Sciences, Beijing, China

Corresponding author:

Xue-Jun Ge

Email address: xjge@scbg.ac.cn

ORCID:

Xue-Jun Ge: orcid.org/0000-0002-5008-9475

Tian-Wen Xiao: orcid.org/0000-0002-7479-7089
Abstract

Background. Gene tree discordance is common in phylogenetic analyses. Many phylogenetic studies have excluded non-coding regions of the plastome without evaluating their impact on tree topology. In general, plastid loci have often been treated as a single unit, and tree discordance among these loci has seldom been examined. Using samples of Laureae (Lauraceae) plastomes, we explored plastome variation among the tribe, examined the influence of non-coding regions on tree topology, and quantified intra-plastome conflict.

Results. We found that the plastomes of Laureae have low inter-specific variation and are highly similar in structure, size, and gene content. Laureae was divided into three groups, subclades I, II and III. The inclusion of non-coding regions changed the phylogenetic relationship among the three subclades. Topologies based on coding and non-coding regions were largely congruent except for the relationship among subclades I, II and III. By measuring the distribution of phylogenetic signal across loci that supported different topologies, we found that nine loci (two coding regions, two introns and five intergenic spacers) played a critical role at the contentious node.

Conclusions. Our results suggest that subclade III and subclade II are successively sister to subclade I. Conflicting phylogenetic signals exist between coding and non-coding regions of Laureae plastomes. Our study highlights the importance of evaluating the influence of non-coding regions on tree topology and emphasizes the necessity of examining discordance among different plastid loci in phylogenetic studies.
Introduction

Gene tree discordance is relatively common in phylogenomic studies. The conflicts can be caused by biological factors like incomplete lineage sorting (ILS), hybridization, horizontal gene transfer, gene loss, and gene duplication (Maddison, 1997; Sun et al., 2015; Gonçalves et al., 2019; Sato et al., 2019). Most relevant studies have focused on incongruent tree topologies among different genomic compartments (Sun et al., 2015; Zhao et al., 2016; Walker et al., 2019) because these genes have evolved independently and their gene tree topologies have been influenced by biological processes. By contrast, relatively few studies have focused on tree conflicts among plastid genes (e.g., Foster, Henwood & Ho, 2018; Gonçalves et al., 2019; Walker et al., 2019; Zhang et al., 2020). Usually, plastomes are considered to be uniparentally inherited and to have evolved as a single unit, free from such biological sources of conflict (Birky, 1995; Wicke et al., 2011). However, the branched and linear structure of plastid DNA, which arose from recombination-dependent replication, is indicative of recombination (Oldenburg & Bendich, 2016; Ruhlman et al., 2017). In addition, biparental inheritance and heteroplasmy (e.g., the presence of different plastomes within an individual or a cell) have been reported in seed plants (Szmidt, Aldén & Hälgren, 1987; Johnson & Palmer, 1989; Reboud & Zeyl, 1994; Carbonell-Caballero et al., 2015). Heteroplasmy may, in rare cases, give rise to heteroplasmic recombination, which has been invoked to explain gene tree discordance (Marshall, Newton & Ritland, 2001; Sullivan et al., 2017; Sancho et al., 2018). In addition to recombination events, the transfer of genes among plastid, mitochondrial and nuclear genomes; positive selection; tree length (a proxy for evolutionary rate); and GC content may also generate
phylogenomic conflict (e.g., Stegemann et al., 2003; Smith, 2014; Wysocki et al., 2015; Piot et al., 2018; Saarela et al., 2018; Foster, Henwood & Ho, 2018). Aside from biological factors, non-biological factors (e.g., outlier genes, uninformative loci, and gaps) may cause conflict as well. For example, Duvall, Burke & Clark (2020) found that alternative topologies arose from alignment gaps. Given that most studies assume no conflict and treat the plastome as a single unit, taking biological and non-biological factors into consideration and quantifying the extent of conflict among different plastid loci is of great importance (Wolfe & Randle, 2004).

Owing to the rapid development of next-generation sequencing (NGS), more plastomes are becoming available at a reasonable cost, driving advances in phylogenomics and promoting a more comprehensive understanding of plant evolution (Li et al., 2019). Phylogenetic relationships among Lauraceae (Song et al., 2017), as well as many other groups (e.g., Eserman et al., 2014; Barrett et al., 2016), have been well resolved using plastome data. In phylogenomic studies of plastomes (Guo et al., 2017; Gonçalves et al., 2019; Xu et al., 2019; Li et al., 2019), plastome coding genes have generally been used, and non-coding regions have been excluded. Only a few studies have noted the potential impact of non-coding regions on tree topology. Parks, Cronn & Liston (2009) revealed that the phylogenetic position of Pinus albicaulis Engelm. based on complete plastomes differed from that based on exon sequences. A similar situation also occurred for phylogenetic relationships within Rubiaceae (Wikström, Bremer & Rydin, 2020), suggesting that there were conflicting phylogenetic signals between coding- and non-coding regions. Because tree topology is the foundation of comparative studies that infer biogeographic history, phylogenetic diversity and other evolutionary patterns (Walker et al.,
the influence of non-coding regions on phylogenetic inference should be evaluated.

Both ILS and hybridization are at play in tree species, which generally have high rates of outcrossing and large population sizes (Petit & Hampe, 2006; Crowl et al., 2019). Interspecific hybrids have been described in Persea (tribe Perseeae, sister to tribe Cinnamomeae and tribe Laureae), Cinnamomum and Aiouea (tribe Cinnamomeae) (van der Werff, 1984; Rohwer et al., 2019). These processes are perhaps also problematic in Laureae. When combined, such biological processes may make accurate inference of evolutionary relationships in Laureae difficult. Unfortunately, previous phylogenomic studies of Laureae have ignored potential conflicts among different plastid loci and the underlying processes that may have generated them (Zhao et al., 2018; Song et al., 2019; Tian, Ye & Song, 2019). These characteristics make Laureae an ideal group in which to explore intra-plastome conflict and its influence on phylogenetic inference.

Tribe Laureae, a species-rich group in the family Lauraceae, is phylogenetically sister to tribe Cinnamomeae (Song et al., 2019). It comprises approximately 500 species and 10 genera: Actinodaphne, Adenodaphne, Dodecadenia, Iteadaphne, Laurus, Lindera, Litsea, Neolitsea, Parasassafras and Sinosassafras (van der Werff & Richter, 1996; Chanderbali, van der Werff & Renner, 2001; Li et al., 2004; Li et al., 2008b). Species of this tribe are evergreen or deciduous and usually occur in the form of trees or shrubs (Li et al., 2008a). Their distribution ranges from the Mediterranean region, Asia, and Oceania to North America (Li et al., 2004). Some members of Laureae have great ecological and economic value. For example, Neolitsea sericea (Bl.) Koidz. is a dominant species found in various evergreen and deciduous broadleaf mixed forests.
and in evergreen broadleaf forests (Wang et al., 2009), and Laurus nobilis L. has been used in remedies for centuries (Nayak et al., 2006).

Although Laureae is monophyletic, generic delimitation within this tribe remains unclear (Kostermans, 1957; Hutchinson, 1964; Li et al., 2008b). Adenodaphne, endemic to New Caledonia, is closely related to Litsea (Chanderbali, van der Werff & Renner, 2001). However, morphological confusion still exists between this genus and Litsea, meaning that their distinctiveness and the monophyly of Adenodaphne require further study (Chanderbali, van der Werff & Renner, 2001). Actinodaphne is polyphyletic and closely related to the monophyletic genus Neolitsea (Li et al., 2007; Li et al., 2008b; Fijridiyanto & Murakami, 2009a, 2009b).

Although Fijridiyanto & Murakami (2009a, 2009b) argued that Actinodaphne was monophyletic, the species of Actinodaphne sampled in their analyses were totally different from those sampled in Li et al. (2007) and Li et al. (2008b). Furthermore, Lindera and Litsea have been shown to be polyphyletic, with Dodecadenia, Iteadaphne, Laurus, Parasassafras and Sinosassafras nested within them (Li et al., 2004; Li et al., 2008b). Liu et al. (2017) used three plastid barcode loci combined with the internal transcribed spacer (ITS) region for species identification and found that the Laureae tree was polytomic. Despite these efforts, phylogenetic relationships among and within these genera have been poorly resolved based on molecular markers like the ITS, the external transcribed spacer (ETS), matK, trnL-F and trnH-psbA.

Compared with these molecular markers, complete plastomes have better performance at the species level within Laureae, although generic delimitation still remains unclear due to limited taxon sampling (Zhao et al., 2018; Song et al., 2019; Tian, Ye & Song, 2019).
Thirty-five plastomes representing 28 species and six genera of Laureae have been published (Table S1). Compared with the vast diversity of Laureae, the published plastome data for this group are relatively limited. Hence, we now report 12 newly sequenced plastomes (Table 1) and combine them with existing plastomes to address three primary goals: (1) reinvestigation of phylogenetic relationships within Laureae; (2) examination of conflict between coding and non-coding regions; and (3) quantification of conflicts among different plastid loci.

Methods

Plant materials, DNA extraction and genome sequencing

Materials from 12 species in five genera (Actinodaphne obovata (Nees) Bl., Iteadaphne caudata (Nees) H. W. Li, Lindera erythrocarpa Makino, Litsea acutivena Hay., L. elongata (Wall. ex Nees) Benth. et Hook. f., L. glutinosa (Lour.) C. B. Rob., L. dilleniifolia P. Y. Pai et P. H. Huang, L. mollis Hemsl., L. monopetala (Roxb.) Pers., L. pungens Hemsl., L. szemaois (H. Liu) J. Li et H.W. Li, and Neolitsea pallens (D. Don) Momiy. et H. Hara) (tribe Laureae, Lauraceae) were collected and identified by the authors (Table 1). Voucher specimens were deposited in the herbarium of the South China Botanical Garden (IBSC) at the Chinese Academy of Sciences. No specific permissions were required for the relevant locations and activities. Including the plastomes downloaded from GenBank and the Lauraceae Chloroplast Genome Database (LCGDB, https://lcgdb.wordpress.com) (Table S1), this study included 47 Laureae plastomes, representing seven genera and all subclades identified by Song et al. (2019). Twelve plastomes from other tribes were also downloaded (Table S1).
Genomic DNA was extracted from silica-gel-dried leaf tissue using the cetyl trimethyl ammonium bromide (CTAB) method (Doyle & Doyle, 1987). The yields of genomic DNA extracts were quantified by fluorometric quantification on a Qubit instrument (Invitrogen, Carlsbad, California, USA) using the dsDNA HS kit, and the DNA size distribution was assessed visually on a 1% agarose gel. DNA libraries with an average insert size of 270 bp were prepared by the Beijing Genomics Institute (BGI, Shenzhen, China). Paired-end reads of 2 × 151 bp were generated on the Illumina X ten sequencing system (Illumina Inc.).

**Plastid genome assembly, annotation and comparison**

Low-quality reads and adaptors were removed using Trimmomatic v0.36 (Bolger, Lohse & Usadel, 2014), generating approximately 3 Gb of high-quality clean reads per sample. The clean reads were analyzed for quality control with FastQC (Andrews, 2010) and then used to assemble plastomes with NOVOPlasty v2.7.2 (Dierckxsens, Mardulyn & Smits, 2016). To guarantee assembly quality, clean reads were mapped to the assembled plastid genomes using the Burrows-Wheeler Aligner (BWA 0.7.17-r1188 (Li & Durbin, 2010)) and samtools 1.9 (Li et al., 2009), and were visually checked in Geneious Prime 2019.1.

Plastome annotation was performed using the program GeSeq - Annotation of Organellar Genomes (Tillich et al., 2017). Start and stop codons were inspected and manually adjusted in Geneious Prime when necessary. Plastomes were submitted to GenBank (MN274947, MN428456–MN428466). Maps of all 12 plastomes were drawn using the OrganellarGenomeDRAW tool (OGDRAW) (Lohse et al., 2013). A summary of the newly
sequenced plastomes is presented in Table 2.

To illustrate interspecific sequence variation within Laureae, plastomes of *A. obovata*, *I. caudata*, *Laurus nobilis* (KY085912), *Lindera erythrocarpa*, *Litsea acutivena*, *N. pallens* and *Parasassafras confertiflorum* (Meisn.) D. G. Long (MH729378) were aligned using MAFFT (Katoh & Standley, 2013) with default settings. Sequence identity was plotted with the mVISTA program using the LAGAN mode (Frazer et al., 2004), with *Lindera glauca* (Siebold et Zucc.) Bl. (MF188124) as a reference.

**Phylogenetic reconstruction and tests for selection**

To evaluate potential conflicts, phylogenetic trees were constructed using maximum likelihood (ML) methods based on six datasets: (1) complete plastome (CP), (2) coding regions (CDS), (3) non-coding regions (non-CDS), (4) large single copy region (LSC), (5) small single copy region (SSC), and (6) one inverted repeat region (IR).

Sequences were aligned using MAFFT with default settings and manually edited with BioEdit v7.2.5 (Hall, 1999) when necessary. The best-fitting DNA substitution models for the six unpartitioned datasets were selected using ModelTest-NG (Darriba et al., 2020) under the corrected Akaike Information Critierion (AICc). The aligned sequences and selected DNA substitution models were used for ML analyses, and ML trees were constructed using RAxML-NG (Kozlov et al., 2019). We also implemented a partitioning strategy on two datasets, the CP with one IR region removed (CP-reduced) and CDS (configuration details shown in Supplemental File 1). The optimal partitioning schemes for each dataset were inferred with
PartitionFinder 2 (Lanfear et al., 2016), and the optimal partitioning schemes, and nucleotide substitution models for each partition were used for ML analyses in RAxML-NG. Because gaps can affect tree topology (Duvall, Burke & Clark, 2020), we also performed the following analysis based on the CP dataset. ‘Mask Alignment’ in Geneious Prime was used to strip the gaps from the MAFFT alignment, with the threshold set to 0 (no gaps), 2%, 10%, 20%, 50% or 75%. The resulting alignments were used to infer ML trees in RAxML-NG.

Positive selection on plastid coding genes has the potential to bias phylogenies (e.g., Piot et al., 2018; Saarela et al., 2018), and we therefore performed positive selection tests using CODEML in PAML 4.9j (Yang, 2007). Coding genes were extracted and aligned in Geneious Prime using MAFFT, stop codons were removed manually, and the aligned sequences were converted to PAML format. Because site models allow dN/dS ratio to vary among different sites, we implemented M0, M1a, M2a, M3, M7 and M8 to identify positively selected sites. Likelihood ratio tests (LRTs) were performed using pchisq function in R 3.6.2 (R Core Team, 2018) to test if there was significant difference between models (M0 vs M3, M2a vs M1a, M8 vs M7). We manually deleted positively selected sites when LTRs was significant (M2a vs M1a and/or M8 vs M7 with p value less than 0.05). Coding gene alignments with positively selected sites removed were concatenated (CDS-reduced dataset), and used for ML tree inference to examine whether positive selection can bias phylogeny or not.

Node support investigation and tree topology tests

Because gene contents were not identical among Cryptocaryae, Cassytha, Caryodaphnopsis,
Neocinnamomum and other clades, the following analyses were performed using a dataset from which six plastomes had been removed (Beilschmiedia pauciflora H. W. Li, Caryodaphnopsis malipoensis Bing Liu et Y. Yang, Cassytha filiformis L., Cryptocarya chinensis (Hance) Hemsl. and Eusideroxylon zwageri Teijsm. et Binn.).

We extracted all loci (coding regions, introns, tRNA, rRNA and intergenic spacers) using a python script (Jin, 2019) and aligned them using MAFFT with default settings. These alignments were used to infer gene trees by rapid bootstrap analyses (option -f a) in RAxML (Stamatakis, 2014) with the GTRGAMMA model. The number of bootstrap replicates was set to 1000, as Simmons & Kessenich (2019) have suggested that fewer replicates may be insufficient to find the optimal gene tree topology. The best-scoring ML trees were used to estimate the species tree with local posterior probability (LPP) (Sayyari & Mirarab, 2016) in ASTRAL III (Zhang et al., 2018).

We performed constrained maximum likelihood analyses in IQ-TREE (Nguyen et al., 2014) to obtain the ML trees that supported different topologies. To understand which loci supported the alternative topologies, we calculated site-wise log-likelihood values for each topology in RAxML using option “-f G”. After obtaining site-wise lnL differences, we converted site-wise differences to locus-wise lnL differences (ΔlnL) in R 3.6.2. The lnL differences were plotted against each locus using ggplot2 (Wickham, 2016). It has been suggested that loci with an absolute ΔlnL > 2 are statistically significant (Edwards, 1984). Therefore, we conducted separate ML analyses on datasets from which these loci (absolute ΔlnL > 2) had been removed to test whether small subsets of sequence matrices determined tree topology (Shen, Hittinger & Rokas,
The Kishino–Hasegawa test (KH test) (*Kishino & Hasegawa, 1989*), Shimodaira-Hasegawa test (SH test) (*Shimodaira & Hasegawa, 1999*) and Approximately-Unbiased test (AU test) (*Shimodaira, 2002*) were used in IQ-TREE to assess the statistical significance of incongruence based on complete plastomes (including only one copy of the IR regions). We specified 10,000 RELL (resampling of estimated log-likelihoods) replicates for the topological tests.

**Results**

**Plastome features of Laureae**

The sizes of the 12 newly generated Laureae plastid genomes ranged from 152,132 bp (*Litsea szemaois*) to 152,916 bp (*Lindera erythrocarpa*) (Table 2), similar to previously published Laureae plastomes (152,211–153,011 bp, Table S1). All had a typical quadripartite structure and were assembled into a single, circular and double-stranded DNA sequence (Fig. 1). The length of the LSC, SSC and IR regions ranged from 93,119 bp (*Litsea szemaois*) to 93,921 bp (*Lindera erythrocarpa*), 18,796 bp (*N. pallens*) to 18,936 bp (*Litsea mollis*), and 20,057 bp (*A. obovata*) to 20,144 bp (*I. caudata*), respectively, with little variation in size (Table 2). The overall GC contents ranged from 39.1% to 39.2%. GC content was unequally distributed within the plastomes; it was highest in IR regions (44.4–44.5%), moderate in LSC regions (37.9–38.1%), and lowest in SSC regions (33.8–34.0%, Table 2).

The 12 newly sequenced plastomes contained 112 single-copy genes: 78 protein-coding genes, 30 tRNA genes, and 4 rRNA genes (Table 2 and Table S2). Sixteen genes had one intron,
and two genes had two introns. There were 13 duplicated genes in the IR regions (Table S2), and
$rps12$, $ycf1$, and $ycf2$ were partly duplicated in the IR regions (Fig. 1).

Phylogenetic reconstruction and positive selection tests

The GTR+I+G4 model was selected for the six unpartitioned datasets (CP, CDS, non-CDS, LSC, SSC and IR). Perseae was sister to Cinnamomeae and Laureae (Figs. 2 and S1-S5). All the ML trees indicated the monophyly of Laureae with high bootstrap (BS) support values (99–100%, Figs. 2 and S1–S4), except for the ML tree based on the IR region (71%, Fig. S5). This result was caused by the low variability of the IR region (Fig. S6). In the five ML trees (Figs. 2 and S1–S4), Laureae was divided into three groups. Subclade I included $Lindera$ communis Hemsl., $L$. glauca and $L$. nacusua (D. Don) Merr.; subclade II included $Laurus$ azorica (Seub.) Franco, $L$. nobilis, $Lindera$ megaphylla Hemsl., $Litsea$ acutivena, $L$. glutinosa, $L$. monopetala and $L$. pungens; and subclade III included the other Laureae species used in the study. In subclade I, $Lindera$ glauca was sister to $L$. communis and $L$. nacusua. In subclade II, $Laurus$ was sister to $Litsea$ acutivena, $L$. glutinosa and $Lindera$ megaphylla, and the position of $Litsea$ pungens was unstable (Figs. 2 and S1–S4). $Litsea$ monopetala (LAU00063) was embedded within three samples of $Litsea$ glutinosa in subclade II, highlighting the necessity of re-identification for $L$. monopetala (LAU00063). Topologies within subclade III based on different datasets were largely congruent (Figs. 2 and S1–S4). In subclade III, samples of $Litsea$, together with $Lindera$ obtusiloba Bl., were monophyletic. $Lindera$ erythrocarpa, $L$. latifolia Hook. f., $L$. metcalfiana Allen and $L$. robusta (Allen) Tsui were monophyletic as well. $Lindera$ aggregata, $L$. chunii
Lec., L. thomsonii Allen and L. thomsonii var. vernayana (Allen) H.P. Tsui formed a well-supported clade. Neolitsea was closer to Actinodaphne than to other Laureae species.

Subclade II was sister to subclade I based on four unpartitioned datasets (CP, non-CDS, LSC, SSC; Figs. 2 and S2–S4, respectively). However, subclade II was sister to subclade III rather than subclade I based on the unpartitioned CDS dataset (Fig. S1). Both topologies were strongly supported. The sister relationship of subclades I and II was supported in the ML tree based on partitioned plastomes (one IR removed, CP-reduced dataset; Fig. S7), and subclade II was sister to subclade III in the ML tree based on the partitioned CDS dataset (Fig. S8), indicating that our results were robust to different partitioning schemes.

The sister relationship of subclades I and II (BS values ranging from 80% to 92%) was consistently revealed even as the percentage of gaps increased (Table S3), indicating that gaps had no impact on our tree topology. Positively selected sites were detected in 27 coding genes (Table S4). The ML tree based on CDS-reduced dataset supported a sister relationship of subclades II and III (Fig. S9), consistent with ML trees based on CDS dataset (Figs. S1 and S8), suggesting that positive selection did not affect the relationship of the three subclades.

Investigating incongruent nodes and differences in tree topology

The tree topology inferred from ASTRAL III (Fig. 3) was largely congruent with that of the ML trees (Figs. 2 and S1–S4), except that the former showed a sister relationship of subclade I and subclade III. We performed constrained maximum likelihood analyses in IQ-TREE and obtained
three suboptimal ML trees that supported the subclade II–subclade I (called T1 hereafter), subclade II–subclade III (T2) and subclade I–subclade III (T3) affinities. We extracted 243 loci and assessed how each locus supported one of the three topologies by examining the gene-wise log-likelihoods (Fig. 4). T1 was strongly supported by six loci (rpoC1 intron, trnG-trnfM, ndhA intron, psaJ-rpl33, rpl2-rpl23 and petN-psbM; absolute ΔlnL > 2); T2 was strongly supported by three loci (psaB, trnS-ycf3 and ycf2; absolute ΔlnL > 2); and T3 was moderately supported by one locus (clpP intron1; absolute ΔlnL > 1 and < 2) (Table S5). The sum of absolute ΔlnL of T1 was higher than that of T2 and T3 (Fig. 4), suggesting that our data support the topology of T1 rather than T2 or T3. After the removal of six loci (rpoC1 intron, trnG-trnfM, ndhA intron, psaJ-rpl33, rpl2-rpl23 and petN-psbM), a sister relationship of subclade II and subclade III was revealed (Fig. S10). After the removal of three loci (psaB, trnS-ycf3, and ycf2), subclade II was sister to subclade I (Fig. S11). These results underscore the decisive role played by small subsets of loci in phylogenetic inference.

The topological tests showed that T2 did not differ significantly from T1 (p > 0.05, Table S6). T3 was statistically rejected by the KH and AU tests (p < 0.05) but not by the Shimodaira-Hasegawa (SH) test (p = 0.0505). That T3 was rejected according to the KH and AU tests suggests that the sister relationship between subclades I and III may be misleading.

Discussion

Plastome features

It has been noted that most plastid genomes of land plants and algae range from 120 to 160
kilobase pairs (kb) in length (Palmer, 1985). In this study, the plastid genome sizes of 12 species from five Laureae genera ranged from 152,132 bp to 152,916 bp, indicating that plastid genome size was conserved within Laureae. GC content was highest in the IR region rather than in the single copy regions, owing to the presence of a ribosomal RNA gene cluster in the IR region, consistent with a previous study (Huotari & Korpelainen, 2012). GC contents of the IR, LSC and SSC regions of the newly sequenced plastomes were identical to those of nine Lindera species studied earlier (Zhao et al., 2018). In contrast to the gene losses recognized in several Lauraceae lineages (Song et al., 2017), our analysis revealed that gene content among Laureae was highly conserved. Song et al. (2017) suggested that plastome contraction in Lauraceae was largely driven by fragment loss events in the IR regions. In our study, we found no gene loss among Laureae plastomes.

Phylogenetic relationships within Laureae

Previous phylogenetic studies (Song et al., 2017; Zhao et al., 2018) based on complete plastomes suggested that Laureae was sister to Cinnamomeae and that together they were sister to Perseeae. The same phylogenetic relationships among these groups were recognized in our study (Figs. 2 and 3). In previous work, Actinodaphne and Neolitsea were resolved as monophyletic groups based on matK, ITS and rpb2 (Fijridiyanto & Murakami, 2009a, 2009b), but Actinodaphne was not a monophyletic group based on complete plastid genomes (Song et al., 2019). In this study, the non-monophyletic status of Actinodaphne was supported. The conclusion of Actinodaphne monophyly may have been caused by sampling bias in previous studies (Fijridiyanto &
The monophyly of *Neolitsea* can be neither rejected nor supported in the present study. *Actinodaphne cupularis* (Hemsl.) Gamble was grouped with *Neolitsea oblongifolia* Merr. et Chun, *N. pallens* and *N. chui* Merr. with low bootstrap support (54%; Fig. 2), and sampling of *Neolitsea* and related genera was limited. *Lindera* and *Litsea* were polyphyletic in our analysis, consistent with previous studies (Li et al., 2008b; Fijridiyanto & Murakami, 2009b). The phylogenetic position of *P. confertiflorum* was unresolved based on ETS and ITS (Li et al., 2008b), and the ambiguity of its position still remains, despite the integration of complete plastid genomes in our analysis and a previous study (Liao, Ye & Song, 2018).

Subclade III was sister to subclade I and II in our study, consistent with previous analyses (Zhao et al., 2018; Song et al., 2019; Tian, Ye & Song, 2019). The three *Lindera* species in subclade I share common morphological traits, such as alternate and pinninerved leaves, a persistent involucre, vegetative terminal buds in inflorescences and 3-merous flowers (Li et al., 2008a). However, these characters also occur in several members of the other two subclades (e.g., *Lindera benzoin* (L.) Bl. and *Laurus nobilis*), perhaps resulting from convergent and/or parallel evolution (Li et al., 2008b). These traits are not good indicators for delimiting the three subclades of Laureae. In subclade III, the trinerved or triplinerved species of *Lindera* (*Lindera aggregata*, *L. chunii*, *L. fragrans*, *L. limprichtii*, *L. pulcherrima*, *L. supracostata*, *L. thomsonii* and *L. thomsonii* var. *vernayana*) formed a well-supported clade in both our study and that of Tian, Ye & Song (2019). However, triplinerved leaves also exist in most species of *Neolitsea* (Li et al., 2008b; Li et al., 2008a). Therefore, traditional morphological traits are of limited use in taxon delimitation, even within subclades of Laureae. Given the limited samples and data in our
analyses, more sampling and DNA sequences are needed to further elucidate the relationships within Laureae.

Phylogenetic incongruence in the plastome

Although many studies have treated plastid protein-coding genes or the complete plastome as a single unit (e.g., Song et al., 2019; Tian, Ye & Song, 2019), potential conflicts among sequence types (i.e., coding vs. non-coding regions) have been reported in several studies. By comparing phylogenies based on complete plastomes and coding regions (Yu et al., 2017), it was inferred that non-coding regions did not significantly influence the tree topology of Theaceae. By contrast, non-coding regions had an impact on the phylogenetic relationships of several tribes in Rubiaceae (Wikström, Bremer & Rydin, 2020) and subtribes in Poaceae (Saarela et al., 2018). A conflicting signal between coding and non-coding regions was also reported in Leguminosae (Zhang et al., 2020). In this study, inclusion of non-coding regions altered tree topology in the tribe Laureae, suggesting the existence of a conflicting signal between coding and non-coding regions. Non-coding regions are often discarded for being uninformative, or for being misleading due to saturation at deep time scales (Foster, Henwood & Ho, 2018). In our study, tree topologies based on coding and non-coding regions were largely congruent, except for the relationships among the three subclades (Figs. S1–S2), indicating that non-coding regions are as informative as coding regions in Laureae. Thus, it is imperative to evaluate the influence of non-coding regions on tree topology rather than treating the whole plastome as a single unit or simply excluding non-coding regions from phylogenetic analysis.
To accommodate the conflicts among different plastid regions, a species tree was inferred through summary coalescent analysis. It has been suggested that the coalescent method is more robust than the concatenation method when the level of ILS is high (Liu, Xi & Davis, 2014; Mirarab, Bayzid & Warnow, 2014). High ILS tends to occur when the time interval between consecutive speciation events is short (Sun et al., 2015; Sato et al., 2019), and the core Lauraceae group (Perseeae, Cinnamomeae and Laureae) is thought to have undergone a rapid radiation (Chanderbali, van der Werff & Renner, 2001; Rohwer & Rudolph, 2005; Nie, Wen & Sun, 2007). We therefore chose to implement the coalescent method. Nonetheless, it should be noted that, with this method, short and uninformative loci may lead to problematic gene trees and therefore result in a less accurate species tree (Xi, Liu & Davis, 2015; Springer & Gatesy, 2016).

In our study, only nine of 243 loci (rpoC1 intron, trnG-trnfM, ndhA intron, psaJ-rpl33, rpl2-rpl23, petN-psbM, psaB, trnS-ycf3, and ycf2) had a strong phylogenetic signal at the contentious node. The other 234 loci with weak phylogenetic signals may have resulted in gene trees with uncertainties and led to inaccurate topology at this node.

Exploration of the factors that underlie conflicts in phylogenetic signals is of great importance—but it is also challenging. Previous studies have examined whether biological and non-biological factors contribute to such conflicts (e.g., Duvall, Burke & Clark, 2020; Zhang et al., 2020). For example, gaps have been found to cause alternate, but conflicting topologies in Poaceae (Duvall, Burke & Clark, 2020). However, the inclusion of alignment gaps did not alter our tree topology (Table S3). Although previous studies indicated that partitioning improves phylogenetic inference (Xi et al., 2012), ML tree topologies based on partitioned and
unpartitioned datasets did not differ significantly in our study. It has been suggested that plastid
genes under positive selection may bias phylogenies (e.g., Piot et al., 2018; Saarela et al., 2018),
however, we found that the relationship among subclades I, II and III was not affected by
positively selected sites, suggesting that positive selection was not the cause of tree conflicts. In
this study, the low support values and short branch lengths of the estimated species tree (Fig. 3)
suggested that each locus had a significantly incongruent topology and may indicate the
existence of ILS. High levels of ILS are thought to yield similar numbers of loci supporting
alternative topologies (Huson et al., 2005). In our study, the numbers of loci supporting each
topology were different (six for T1, three for T2, and zero for T3 after exclusion of loci with
absolute $\Delta\ln L \leq 2$), suggesting that ILS may not be the primary cause of the discordance among
loci. Another plausible explanation for the conflict is heteroplasmic recombination, which can
occur in species with biparental plastome inheritance (Walker et al., 2019). Although
heteroplasmic recombination has been reported with clear evidence in Brachypodium and Picea
(Sullivan et al., 2017; Sancho et al., 2018), to our knowledge it has never been documented in
Lauraceae. Based on the data reported here, it is too early to draw a firm conclusion about the
causes of the conflict in phylogenetic signals. Although fully resolved phylogenies may still
remain elusive based on different genomic compartments (i.e., nuclear, mitochondrial and
plastid), phylogenomic studies that incorporate these compartments can provide new insights
into tree discordance and its underlying causes (Koenen et al., 2020). Therefore, more genetic
information (e.g., nuclear genes) will be required to solve this problem in future work.
Conclusion

In summary, this study revealed that Laureae plastomes are conserved in structure, size and gene content. A conflicting phylogenetic signal was detected between coding and non-coding regions, suggesting that the plastid genome should not be treated as a single unit. ML trees based on coding and non-coding regions were largely congruent except at the contentious node, indicating that coding regions are as informative as non-coding regions and that the influence of non-coding regions on tree inference should be evaluated. We also found that small subsets of plastome loci determined the topology at specific nodes, consistent with the results of a previous study (Shen, Hittinger & Rokas, 2017). Through quantification and analysis of intra-plastome conflicts, the sister relationship of subclade I (including Lindera communis, L. glauca and L. nacusua) and II (including Laurus azorica, L. nobilis, L. megaphylla, Litsea acutivena, L. glutinosa, L. monopetala and L. pungens) was supported by our study. Biological factors may contribute to the conflicts among plastid loci; however, more information is needed to determine the underlying mechanism(s).

Acknowledgements

The authors thank Yu-Ying Zhou and Chen-Xin Ma for assistance with the DNA experiments. They also thank Feng Song for her kind help in plastid genome assembly and annotation. The authors would like to thank TopEdit (www.topeditsci.com) for English language editing of this manuscript.

Reference
Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. Available at http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Barrett CF, Baker WJ, Comer JR, Conran JG, Lahmeyer SC, Leebens-Mack JH, Li J, Lim GS, Mayfield-Jones DR, Perez L, Medina J, Pires JC, Santos C, Wm. Stevenson D, Zomlefer WB, Davis JI. 2016. Plastid genomes reveal support for deep phylogenetic relationships and extensive rate variation among palms and other commelinid monocots. New Phytologist 209:855-870 DOI: 10.1111/nph.13617.

Birky CW. 1995. Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. Proceedings of the National Academy of Sciences of the United States of America 92:11331-11338 DOI: 10.1073/pnas.92.25.11331.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114-2120 DOI: 10.1093/bioinformatics/btu170.

Carbonell-Caballero J, Alonso R, Ibañez V, Terol J, Talon M, Dopazo J. 2015. A phylogenetic analysis of 34 chloroplast genomes elucidates the relationships between wild and domestic species within the genus Citrus. Molecular Biology and Evolution 32:2015-2035 DOI: 10.1093/molbev/msv082.

Chanderbali AS, van der Werff H, Renner SS. 2001. Phylogeny and historical biogeography of Lauraceae: Evidence from the chloroplast and nuclear genomes. Annals of the Missouri Botanical Garden 88:104-134 DOI: 10.2307/2666133.

Crowl AA, Manos PS, McVay JD, Lemmon AR, Lemmon EM, Hipp AL. 2019. Uncovering the genomic signature of ancient introgression between white oak lineages (Quercus). New Phytologist DOI: 10.1111/nph.15842.

Darriba D, Posada D, Kozlov AM, Stamatakis A, Morel B, Flouri T, Crandall K. 2020. ModelTest-NG: A new and scalable tool for the selection of DNA and protein evolutionary models. Molecular Biology and Evolution 37:291-294 DOI: 10.1093/molbev/msz189.

Dierckxsens N, Mardulyn P, Smits G. 2016. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Research e18-e18 DOI: 10.1093/nar/gkw955.

Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin 19:11-15.

Duvall MR, Burke SV, Clark DC. 2020. Plastome phylogenomics of Poaceae: alternate topologies depend on alignment gaps. Botanical Journal of the Linnean Society 192:9-20 DOI: 10.1093/botlinnean/boz060.

Edwards AWF. 1984. Likelihood. Cambridge: Cambridge University Press.

Eserman LA, Tiley GP, Jarret RL, Leebens-Mack JH, Miller RE. 2014. Phylogenetics and diversification of morning glories (tribe Ipomoeeae, Convolvulaceae) based on whole plastome sequences. American Journal of Botany 101:92-103 DOI: 10.3732/ajb.1300207.

Fijridiyanto IA, Murakami N. 2009a. Molecular systematic of Malesian Litsea Lam. and putative related genera (Lauraceae). Acta Phytotaxonomica et Geobotanica 60:1-18 DOI: 10.18942/apg.KJ00005576218.

Fijridiyanto IA, Murakami N. 2009b. Phylogeny of Litsea and related genera (Lauraeae-Lauraceae) based on analysis of rpb2 gene sequences. Journal of Plant Research 122:283-298 DOI: 10.1007/s10265-009-0218-8.

Foster CSP, Henwood MJ, Ho SYW. 2018. Plastome sequences and exploration of tree-space help to resolve the phylogeny of riceflowers (Thymelaeaceae: Pimelea). Molecular Phylogenetics and Evolution 127:156-167 DOI: 10.1016/j.ympev.2018.05.018.

Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. 2004. VISTA: computational tools for comparative genomics. Nucleic Acids Research 32:W273-W279 DOI: 10.1093/nar/gkh458.
Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Subgroup GPDP. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078-2079 DOI: 10.1093/bioinformatics/btp352.

Li J, Christophel DC, Conran JG, Li H-W. 2004. Phylogenetic relationships within the ‘core’ Laureae (*Litsea* complex, Lauraceae) inferred from sequences of the chloroplast gene matK and nuclear ribosomal DNA ITS regions. *Plant Systematics and Evolution* 246:19-34 DOI: 10.1007/s00606-003-0113-z.

Li J, Conran JG, Christophel DC, Li Z-M, Li L, Li H-W. 2008b. Phylogenetic relationships of the *Litsea* complex and core Laureae (Lauraceae) using ITS and ETS sequences and morphology. *Annals of the Missouri Botanical Garden* 95:580-600 DOI: 10.3417/2006125.9504.

Li J, Christophel DC, Conran JG, Li H-W. 2004. Phylogenetic relationships within the ‘core’ Laureae (*Litsea* complex, Lauraceae) inferred from sequences of the chloroplast gene matK and nuclear ribosomal DNA ITS regions. *Plant Systematics and Evolution* 246:19-34 DOI: 10.1007/s00606-003-0113-z.

Li J, Conran JG, Christophel DC, Li Z-M, Li L, Li H-W. 2008b. Phylogenetic relationships of the *Litsea* complex and core Laureae (Lauraceae) using ITS and ETS sequences and morphology. *Annals of the Missouri Botanical Garden* 95:580-600 DOI: 10.3417/2006125.9504.

Liao Q, Ye T, Song Y. 2018. Complete chloroplast genome sequence of a subtropical tree, *Parasassafras confertiflorum* (Lauraceae). *Mitochondrial DNA Part B* 3:1216-1217 DOI: 10.1080/23802359.2018.1532331.

Liu L, Xi Z, Davis CC. 2014. Coalescent methods are robust to the simultaneous effects of long branches and incomplete lineage sorting. *Molecular Biology and Evolution* 32:791-805 DOI: 10.1093/molbev/msu331.

Liu Z-F, Ci X-Q, Li L, Li H-W, Conran JG, Li J. 2017. DNA barcoding evaluation and implications for phylogenetic relationships in Lauraceae from China. *PLOS ONE* 12:e0175788 DOI: 10.1371/journal.pone.0175788.

Lohse M, Drechsel O, Kahlaus S, Bock R. 2013. OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Research* 41:W575-W581.

Maddison WP. 1997. Gene Trees in Species Trees. *Systematic Biology* 46:523-536 DOI: 10.1093/sysbio/46.3.523.

Marshall HD, Newton C, Ritland K. 2001. Sequence-repeat polymorphisms exhibit the signature of recombination in Lodgepole pine chloroplast DNA. *Molecular Biology and Evolution* 18:2136-2138 DOI: 10.1093/oxfordjournals.molbev.a003757.

Mirarab S, Bayzid MS, Warnow T. 2014. Evaluating summary methods for multilocus species tree estimation in the presence of incomplete lineage sorting. *Systematic Biology* 65:366-380 DOI: 10.1093/sysbio/syu063.

Nayak S, Nalabothu P, Sandiford S, Bhogadi V, Adogwa A. 2006. Evaluation of wound healing activity of *Allamanda cathartica* L. and *Laurus nobilis* L. extracts on rats. *BMC Complementary and Alternative Medicine* 6:12 DOI: 10.1186/1472-6882-6-12.

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2014. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* 32:268-274 DOI: 10.1093/molbev/msu300.

Nie Z-L, Wen J, Sun H. 2007. Phylogeny and biogeography of *Sassafras* (Lauraceae) disjunct between eastern Asia and eastern North America. *Plant Systematics and Evolution* 267:191-203 DOI: 10.1007/s00606-007-0550-1.

Oldenburg DJ, Bendich AJ. 2016. The linear plastid chromosomes of maize: terminal sequences, structures, and implications for DNA replication. *Current Genetics* 62:431-442 DOI: 10.1007/s00294-015-0548-0.

Palmer JD. 1985. Comparative organization of chloroplast genomes. *Annual Review of Genetics* 19:325-354 DOI:
564 10.1146/annurev.ge.19.120185.001545.

565 Parks M, Cronn R, Liston A. 2009. Increasing phylogenetic resolution at low taxonomic levels using massively
566 parallel sequencing of chloroplast genomes. BMC Biology 7:84 DOI: 10.1186/1741-7007-7-84.

567 Petit RJ, Hampe A. 2006. Some evolutionary consequences of being a tree. Annual Review of Ecology, Evolution,
568 and Systematics 37:187-214 DOI: 10.1146/annurev.ecolsys.37.091305.110215.

569 Piot A, Hackel J, Christin P-A, Besnard G. 2018. One-third of the plastid genes evolved under positive selection in
570 PACMAD grasses. Planta 247:255-266 DOI: 10.1007/s00425-017-2781-x.

571 R Core Team. 2018. R: A language and environment for statistical computing. Available at http://www.R-
572 project.org/ (accessed 12 Dec 2019).

573 Reboud X, Zeyl C. 1994. Organelle inheritance in plants. Heredity 72:132-140 DOI: 10.1038/hdy.1994.19.

574 Rohwer JG, Rudolph B. 2005. Jumping genera: the phylogenetic positions of Cassytha, Hypodaphnis, and
575 Neocinnamomum (Lauraceae) based on different analyses of trnK intron sequences. Annals of the Missouri
576 Botanical Garden 1:153-178.

577 Rohwer JG, Trofimov D, Mayland-Quellhorst E, Albach D. 2019. Incongruence of morphological determinations
578 and DNA barcode sequences: a case study in Cinnamomum (Lauraceae). Willdenowia 49:383-400 DOI:
579 10.3372/wi.49.49309.

580 Ruhlman TA, Zhang J, Blazier JC, Sabir JSM, Jansen RK. 2017. Recombination-dependent replication and gene
581 conversion homogenize repeat sequences and diversify plastid genome structure. American Journal of
582 Botany 104:559-572 DOI: 10.3732/ajb.1600453.

583 Saarela JM, Burke SV, Wysocki WP, Barrett MD, Clark LG, Craine JM, Peterson PM, Soreng RJ, Vorontsova MS,
584 Duvall MR. 2018. A 250 plastome phylogeny of the grass family (Poaceae): topological support under
585 different data partitions. PeerJ 6:e4299 DOI: 10.7717/peerj.4299.

586 Sancho R, Cantalapiedra CP, López-Alvarez D, Gordon SP, Vogel JP, Catalán P, Contreras-Moreira B. 2018.
587 Comparative plastome genomics and phylogenomics of Brachypodium: flowering time signatures,
588 introgression and recombination in recently diverged ecotypes. New Phytologist 218:1631-1644 DOI:
589 10.1111/nph.14926.

590 Sato JJ, Bradford TM, Armstrong KN, Donnellan SC, Echenique-Diaz LM, Begué-Quiala G, Gámez-Diez J,
591 Yamaguchi N, Truong Nguyen S, Kita M, Ohdachi SD. 2019. Post K-Pg diversification of the mammalian
592 order Eulipotyphla as suggested by phylogenomic analyses of ultra-conserved elements. Molecular
593 Phylogenetics and Evolution 141:106605 DOI: 10.1016/j.ympev.2019.106605.

594 Sayyari E, Mirarab S. 2016. Fast coalescent-based computation of local branch support from quartet frequencies.
595 Molecular Biology and Evolution 33:1654-1668 DOI: 10.1093/molbev/msw079.

596 Shen X-X, Hittinger CT, Rokas A. 2017. Contentious relationships in phylogenomic studies can be driven by a
597 handful of genes. Nature Ecology & Evolution 1:0126 DOI: 10.1038/s41559-017-0126.

598 Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree selection. Systematic Biology 51:492-508
599 DOI: 10.1080/10635150290069913.

599 Shimodaira H, Hasegawa M. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic
600 inference. Molecular Biology and Evolution 16:1114-1114.

601 Simmons MP, Kessenich J. 2019. Divergence and support among slightly suboptimal likelihood gene trees.
602 Cladistics DOI: 10.1111/cla.12404.

603 Smith DR. 2014. Mitochondrion-to-plastid DNA transfer: it happens. New Phytologist 202:736-738 DOI:
Song Y, Yu W-B, Tan Y, Jin J, Wang B, Yang J-B, Liu B, Corlett RT. 2019. Plastid phylogenomics improve phylogenetic resolution in the Lauraceae. *Journal of Systematics and Evolution* DOI: 10.1111/jse.12536.

Song Y, Yu WB, Tan Y, Liu B, Yao X, Jin J, Padmanaba M, Yang JB, Corlett RT. 2017. Evolutionary comparisons of the chloroplast genome in Lauraceae and insights into loss events in the Magnoliids. *Genome Biology and Evolution* 9:2354-2364 DOI: 10.1093/gbe/evx180.

Springer MS, Gatesy J. 2016. The gene tree delusion. *Molecular Phylogenetics and Evolution* 94:1-33 DOI: 10.1016/j.ympev.2015.07.018.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312-1313 DOI: 10.1093/bioinformatics/btu033.

Stegemann S, Hartmann S, Ruf S, Bock R. 2003. High-frequency gene transfer from the chloroplast genome to the nucleus. *Proceedings of the National Academy of Sciences of the United States of America* 100:8828 DOI: 10.1073/pnas.1430924100.

Sullivan AR, Schiffthaler B, Thompson SL, Street NR, Wang X-R. 2017. Interspecific plastome recombination reflects ancient reticulate evolution in *Picea* (Pinaceae). *Molecular Biology and Evolution* 34:1689-1701 DOI: 10.1093/molbev/msx111.

Sun M, Soltis DE, Soltis PS, Zhu X, Burleigh JG, Chen Z. 2015. Deep phylogenetic incongruence in the angiosperm clade Rosidae. *Molecular Phylogenetics and Evolution* 83:156-166 DOI: 10.1016/j.ympev.2014.11.003.

Szmidt AE, Aldén T, Hällgren J-E. 1987. Paternal inheritance of chloroplast DNA in *Larix*. *Plant Molecular Biology* 9:59-64 DOI: 10.1007/BF00017987.

Tian X, Ye J, Song Y. 2019. Plastome sequences help to improve the systematic position of trinerved *Lindera* species in the family Lauraceae. *PeerJ* 7:e7662 DOI: 10.7717/peerj.7662.

Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. 2017. GeSeq - versatile and accurate annotation of organelle genomes. *Nucleic Acids Research* 45:W6-W11 DOI: 10.1093/nar/gkx391.

van der Werff H. 1984. Notes on neotropical Lauraceae. *Annals of the Missouri Botanical Garden* 71:1180-1183 DOI: 10.2307/2399252.

van der Werff H, Richter HG. 1996. Toward an improved classification of Lauraceae. *Annals of the Missouri Botanical Garden* 83:409-418 DOI: 10.2307/2399870.

Walker JF, Walker-Hale N, Vargas OM, Larson DA, Stull GW. 2019. Characterizing gene tree conflict in plastome-inferred phylogenies. *PeerJ* 7:e7747 DOI: 10.7717/peerj.7747.

Wang G, Zhao C, Chen Y, Cheng C, Zhao Y. 2009. National key protected plant flora and distribution characteristic in Zhouran Islands. *Journal of Zhejiang Forestry Science and Technology* 29:43-47.

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 DOI: 10.1007/s11103-011-9762-4.

Wickham H. 2016. *ggplot2: elegant graphics for data analysis*. New York: Springer-Verlag.

Wikström N, Bremer B, Rydin C. 2020. Conflicting phylogenetic signals in genomic data of the coffee family (Rubiaceae). *Journal of Systematics and Evolution* DOI: 10.1111/jse.12566.

Wolfe AD, Randle CP. 2004. Recombination, heteroplasmy, haplotype polymorphism, and paralogy in plastid genes: implications for plant molecular systematics. *Systematic Botany* 29:1011-1020.

Wysocki WP, Clark LG, Attigala L, Ruiz-Sanchez E, Duvall MR. 2015. Evolution of the bamboos (Bambusoideae;
Poaceae): a full plastome phylogenomic analysis. *BMC Evolutionary Biology* 15:50 DOI: 10.1186/s12862-015-0321-5.

Xi Z, Liu L, Davis CC. 2015. Genes with minimal phylogenetic information are problematic for coalescent analyses when gene tree estimation is biased. *Molecular Phylogenetics and Evolution* 92:63-71 DOI: 10.1016/j.ympev.2015.06.009.

Xi Z, Ruhfel BR, Schaefer H, Amorim AM, Sugumaran M, Wurdack KJ, Endress PK, Matthews ML, Stevens PF, Mathews S, Davis CC. 2012. Phylogenomics and a posteriori data partitioning resolve the Cretaceous angiosperm radiation Malpighiales. *Proceedings of the National Academy of Sciences of the United States of America* 109:17519-17524 DOI: 10.1073/pnas.1205818109.

Xu W-Q, Losh J, Chen C, Li P, Wang R-H, Zhao Y-P, Qiu Y-X, Fu C-X. 2019. Comparative genomics of figworts (Scrophularia, Scrophulariaceae), with implications for the evolution of Scrophularia and Lamiales. *Journal of Systematics and Evolution* 57:55-65 DOI: 10.1111/jse.12421.

Yang Z. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology and Evolution* 24:1586-1591 DOI: 10.1093/molbev/msm088.

Yu XQ, Gao LM, Soltis DE, Soltis PS, Yang JB, Fang L, Yang SX, Li DZ. 2017. Insights into the historical assembly of East Asian subtropical evergreen broadleaved forests revealed by the temporal history of the tea family. *New Phytologist* 215:1235-1248 DOI: doi:10.1111/nph.14683.

Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* 19:153 DOI: 10.1186/s12859-018-2129-y.

Zhang R, Wang Y-H, Jin J-J, Stull GW, Bruneau A, Cardoso D, de Queiroz LP, Moore MJ, Zhang S-D, Chen S-Y, Wang J, Li D-Z, Yi T-S. 2020. Exploration of plastid phylogenomic conflict yields new insights into the deep relationships of Leguminosae. *Systematic Biology* DOI: 10.1093/sysbio/syaa013.

Zhao L, Li X, Zhang N, Zhang S-D, Yi T-S, Ma H, Guo Z-H, Li D-Z. 2016. Phylogenomic analyses of large-scale nuclear genes provide new insights into the evolutionary relationships within the rosids. *Molecular Phylogenetics and Evolution* 105:166-176 DOI: 10.1016/j.ympev.2016.06.007.

Zhao M-L, Song Y, Ni J, Yao X, Tan Y-H, Xu Z-F. 2018. Comparative chloroplast genomics and phylogenetics of nine *Lindera* species (Lauraceae). *Scientific Reports* 8:8844 DOI: 10.1038/s41598-018-27090-0.
**Table 1** (on next page)

Sampled species and voucher specimens of Laureae in this study.
| Taxon                              | Herbarium | Voucher  | Geographic origin                          | GenBank Accession number |
|-----------------------------------|-----------|----------|--------------------------------------------|--------------------------|
| *Actinodaphne obovata* (Nees) Bl. | IBSC      | XTBGLQM0236 | Xishuangbanna, Yunnan, China               | MN274947                 |
| *Iteadaphne caudata* (Nees) H. W. Li | IBSC      | XTBGLQM0582 | Xishuangbanna, Yunnan, China               | MN428456                 |
| *Lindera erythrocarpa* Makino     | IBSC      | 180923   | Baishanzu Mountain, Zhejiang, China        | MN428457                 |
| *Litsea acutivena* Hay.           | \         | \        | Chebaling, Guangdong, China                | MN428458                 |
| *Litsea dillenifolia* P. Y. Pai et P. H. Huang | IBSC      | XTBGLQM0095 | Xishuangbanna, Yunnan, China               | MN428459                 |
| *Litsea elongata* (Wall. ex Nees) Benth. et Hook. f. | IBSC      | WBGQXJ001 | Badagong Mountain, Hunan, China            | MN428460                 |
| *Litsea glutinosa* (Lour.) C. B. Rob. | IBSC      | XTBGLQM0653 | Xishuangbanna, Yunnan, China               | MN428461                 |
| *Litsea mollis* Hemsl.            | IBSC      | CFL2678   | Libo county, Guizhou, China                | MN428462                 |
| *Litsea monopetala* (Roxb.) Pers. | IBSC      | XTBGLQM0687 | Xishuangbanna, Yunnan, China               | MN428463                 |
| *Litsea pungens* Hemsl.           | IBSC      | WBGQXJ124 | Badagong Mountain, Hunan, China            | MN428464                 |
| *Litsea szemaois* (H. Liu) J. Li et H.W. Li | IBSC      | XTBGLQM0692 | Xishuangbanna, Yunnan, China               | MN428465                 |
| *Neolitsea pallens* (D. Don) Momiy. et H. Hara | IBSC      | 18371    | Dinghu Mountain, Guangdong, China          | MN428466                 |
Table 2 (on next page)

Summary of 12 complete plastomes of Laureae.
### Actinodaphne obovata
- **Total cpDNA size (bp)**: 152579
- **Length of LSC region (bp)**: 93655
- **Length of IR region (bp)**: 20057
- **Length of SSC region (bp)**: 18810
- **Total GC content (%)**: 39.1
- **LSC GC content (%)**: 37.9
- **IR GC content (%)**: 44.4
- **SSC GC content (%)**: 33.9
- **Total number of genes (unique)**: 127 (112)
- **Protein-coding genes (unique)**: 84 (78)
- **Total number of tRNA**: 36 (30)
- **Total number of rRNA**: 8 (4)

### Iteadaphne caudata
- **Total cpDNA size (bp)**: 152863
- **Length of LSC region (bp)**: 93761
- **Length of IR region (bp)**: 20144
- **Length of SSC region (bp)**: 18814
- **Total GC content (%)**: 39.1
- **LSC GC content (%)**: 38.0
- **IR GC content (%)**: 44.4
- **SSC GC content (%)**: 33.8
- **Total number of genes (unique)**: 127 (112)
- **Protein-coding genes (unique)**: 84 (78)
- **Total number of tRNA**: 36 (30)
- **Total number of rRNA**: 8 (4)

### Lindera erythrocarpa
- **Total cpDNA size (bp)**: 152916
- **Length of LSC region (bp)**: 93921
- **Length of IR region (bp)**: 20071
- **Length of SSC region (bp)**: 18853
- **Total GC content (%)**: 39.2
- **LSC GC content (%)**: 38.0
- **IR GC content (%)**: 44.4
- **SSC GC content (%)**: 33.9
- **Total number of genes (unique)**: 127 (112)
- **Protein-coding genes (unique)**: 84 (78)
- **Total number of tRNA**: 36 (30)
- **Total number of rRNA**: 8 (4)

### Litsea acutivena
- **Total cpDNA size (bp)**: 152718
- **Length of LSC region (bp)**: 93677
- **Length of IR region (bp)**: 20066
- **Length of SSC region (bp)**: 18909
- **Total GC content (%)**: 39.2
- **LSC GC content (%)**: 38.0
- **IR GC content (%)**: 44.4
- **SSC GC content (%)**: 33.9
- **Total number of genes (unique)**: 127 (112)
- **Protein-coding genes (unique)**: 84 (78)
- **Total number of tRNA**: 36 (30)
- **Total number of rRNA**: 8 (4)

### Litsea elongata
- **Total cpDNA size (bp)**: 152793
- **Length of LSC region (bp)**: 93827
- **Length of IR region (bp)**: 20066
- **Length of SSC region (bp)**: 18844
- **Total GC content (%)**: 39.2
- **LSC GC content (%)**: 38.0
- **IR GC content (%)**: 44.4
- **SSC GC content (%)**: 33.9
- **Total number of genes (unique)**: 127 (112)
- **Protein-coding genes (unique)**: 84 (78)
- **Total number of tRNA**: 36 (30)
- **Total number of rRNA**: 8 (4)

### Litsea glutinosa
- **Total cpDNA size (bp)**: 152748
- **Length of LSC region (bp)**: 93698
- **Length of IR region (bp)**: 20062
- **Length of SSC region (bp)**: 18926
- **Total GC content (%)**: 39.2
- **LSC GC content (%)**: 38.0
- **IR GC content (%)**: 44.5
- **SSC GC content (%)**: 33.8
- **Total number of genes (unique)**: 127 (112)
- **Protein-coding genes (unique)**: 84 (78)
- **Total number of tRNA**: 36 (30)
- **Total number of rRNA**: 8 (4)

### Litsea dilleniifolia
- **Total cpDNA size (bp)**: 152298
- **Length of LSC region (bp)**: 93218
- **Length of IR region (bp)**: 20094
- **Length of SSC region (bp)**: 18892
- **Total GC content (%)**: 39.2
- **LSC GC content (%)**: 38.0
- **IR GC content (%)**: 44.4
- **SSC GC content (%)**: 34.0
- **Total number of genes (unique)**: 127 (112)
- **Protein-coding genes (unique)**: 84 (78)
- **Total number of tRNA**: 36 (30)
- **Total number of rRNA**: 8 (4)

### Litsea mollis
- **Total cpDNA size (bp)**: 152736
- **Length of LSC region (bp)**: 93655
- **Length of IR region (bp)**: 20063
- **Length of SSC region (bp)**: 18936
- **Total GC content (%)**: 39.2
- **LSC GC content (%)**: 38.0
- **IR GC content (%)**: 44.4
- **SSC GC content (%)**: 33.9
- **Total number of genes (unique)**: 127 (112)
- **Protein-coding genes (unique)**: 84 (78)
- **Total number of tRNA**: 36 (30)
- **Total number of rRNA**: 8 (4)

### Litsea monopetala
- **Total cpDNA size (bp)**: 152705
- **Length of LSC region (bp)**: 93758
- **Length of IR region (bp)**: 20074
- **Length of SSC region (bp)**: 18799
- **Total GC content (%)**: 39.2
- **LSC GC content (%)**: 38.0
- **IR GC content (%)**: 44.4
- **SSC GC content (%)**: 33.9
- **Total number of genes (unique)**: 127 (112)
- **Protein-coding genes (unique)**: 84 (78)
- **Total number of tRNA**: 36 (30)
- **Total number of rRNA**: 8 (4)

### Litsea pungens
- **Total cpDNA size (bp)**: 152655
- **Length of LSC region (bp)**: 93520
- **Length of IR region (bp)**: 20131
- **Length of SSC region (bp)**: 18873
- **Total GC content (%)**: 39.2
- **LSC GC content (%)**: 37.9
- **IR GC content (%)**: 44.4
- **SSC GC content (%)**: 34.0
- **Total number of genes (unique)**: 127 (112)
- **Protein-coding genes (unique)**: 84 (78)
- **Total number of tRNA**: 36 (30)
- **Total number of rRNA**: 8 (4)

### Litsea szemaois
- **Total cpDNA size (bp)**: 152132
- **Length of LSC region (bp)**: 93119
- **Length of IR region (bp)**: 20090
- **Length of SSC region (bp)**: 18843
- **Total GC content (%)**: 39.2
- **LSC GC content (%)**: 38.1
- **IR GC content (%)**: 44.4
- **SSC GC content (%)**: 34.0
- **Total number of genes (unique)**: 127 (112)
- **Protein-coding genes (unique)**: 84 (78)
- **Total number of tRNA**: 36 (30)
- **Total number of rRNA**: 8 (4)

### Neolitsea pallens
- **Total cpDNA size (bp)**: 152699
- **Length of LSC region (bp)**: 93761
- **Length of IR region (bp)**: 20071
- **Length of SSC region (bp)**: 18796
- **Total GC content (%)**: 39.1
- **LSC GC content (%)**: 37.9
- **IR GC content (%)**: 44.4
- **SSC GC content (%)**: 33.9
- **Total number of genes (unique)**: 127 (112)
- **Protein-coding genes (unique)**: 84 (78)
- **Total number of tRNA**: 36 (30)
- **Total number of rRNA**: 8 (4)
Figure 1

Complete plastid genome map of Laureae.

Different functional genes are color coded. Genes outside the circle are transcribed counterclockwise, genes inside the circle are transcribed clockwise. GC content is indicated by darker gray in the inner circle.
plastid genome of tribe Laureae
152132 - 152916 bp
Figure 2

Maximum likelihood phylogenetic tree of Laureae inferred with RAxML-NG based on complete plastomes.

Bootstrap values are indicated on branches. Subclades I, II and III are colored in red, green and blue, respectively.
Figure 3

Tree of Laureae inferred with ASTRAL III using a multispecies coalescent approach.

Local posterior probabilities (LPP) are indicated on branches. Subclades I, II and III are colored in red, green and blue, respectively.
Figure 4

Difference in the log-likelihood (InL) of each plastid locus between two alternative topologies.

The x axis indicates each locus, and the y axis indicates lnL difference. (A) Positive and negative values support the topology showing subclades I-II (T1) and subclades II-III (T2), respectively. (B) Positive and negative values support the topology showing subclades I-II (T1) and subclades I-III (T3), respectively. (C) Positive and negative values support the topology showing subclades II-III (T2) and subclades I-III (T3), respectively. Values starting with + or – indicate the sum of positive and negative values, respectively, and the number of supporting loci is shown in the parenthesis. Note that the order of loci on x axis are different among A, B and C.
A

T1
Subclade III
Subclade II
Subclade I

35 (124 loci)

T2
Subclade III
Subclade II
Subclade I

-19 (119 loci)

B

T1
Subclade III
Subclade II
Subclade I

35 (121 loci)

T3
Subclade III
Subclade I
Subclade II

-13 (122 loci)

C

T2
Subclade III
Subclade II
Subclade I

10 (130 loci)

T3
Subclade III
Subclade I
Subclade II

-4 (113 loci)