

# SCIENTIFIC REPORTS

OPEN

## Comparative chloroplast genomics and phylogenetics of nine *Lindera* species (Lauraceae)

Mei-Li Zhao<sup>2,3</sup>, Yu Song<sup>1,4</sup>, Jun Ni<sup>2</sup>, Xin Yao<sup>1</sup>, Yun-Hong Tan<sup>1,4</sup> & Zeng-Fu Xu<sup>1,2</sup> 

*Lindera*, a core genus of the Lauraceae family, has important economic uses in eastern Asia and North America. However, its historical diversification has not been clarified. In this study, we report nine newly sequenced *Lindera* plastomes. The plastomes of these nine *Lindera* species range from 152,211 (*L. nacusua*) to 152,968 bp (*L. metcalifiana*) in length, similar to that of another Lauraceae species, *Litsea glutinosa* (152,618 bp). The length variation of these plastomes derived from the length variation in the loci *ycf1*, *ycf2*, *ψycf1*, and *ndhF-ψycf1*. Comparing our sequences with other available plastomes in the Lauraceae indicated that eight hypervariable loci, *ihbA-trnG*, *ndhA*, *ndhF-rpl32*, *petA-psbJ*, *psbK-psbI*, *rps16*, *trnS-trnG*, and *ycf1*, could serve as DNA barcodes for species delineation, and that the inverted repeats (IRs) showed contraction/expansion. Further phylogenetic analyses were performed using 32 complete plastomes of Lauraceae and seven barcodes from 14 additional species of *Lindera* and related species in the core Lauraceae. The results showed that these *Lindera* species grouped into two or four sub-clades, and that two *Litsea* species and *Laurus nobilis* were located in the same sub-clade as five *Lindera* species. These data support a close relationship between the genera *Laurus*, *Lindera*, and *Litsea*, and suggest that *Lindera* is polyphyletic.

In plants, the chloroplast is the main locus of photosynthesis and carbon fixation<sup>1</sup>. The chloroplast (cp) genome ranges from 120 to 180 kb in size and has a characteristic structure, in which two inverted repeat (IR) regions divide the cp genome into four parts: the IRs themselves, a large single copy region (LSC) and a small single copy region (SSC)<sup>2</sup>. In the LSC, three of the four core plant barcodes, *psbA-trnH*, *rbcL*, and *matK*, have been widely used for identification purposes and phylogenetic analyses in the past twenty years<sup>3</sup>. Currently, complete cp genomes as well as full-length cp gene sequences are available for an increasing number of taxa, and both have been the subjects of numerous phylogenies. Phylogenomics, a technique for estimating phylogenetic relationships based on high-throughput sequencing, can allow a comprehensive understanding of the evolutionary history of organisms. For instance, Ma, *et al.*<sup>4</sup> used cp phylogenomics to resolve the deep-level relationships of Arundinarieae. Yang, *et al.*<sup>5</sup> used complete cp genome sequences to infer phylogenetic relationships in the genus *Quercus*. Very recently, Zhang, *et al.*<sup>6</sup> provided important insights into deep phylogenetic relationships and the diversification history of the Rosaceae based on analyzing plastid phylogenomics.

Like Arundinarieae and Rosaceae, Lauraceae are a large monophyletic group, comprising approximately 3,500 known species from over 50 genera worldwide<sup>7,8</sup>, which are by far the largest family of the order Laurales<sup>9</sup>. The phylogenetic backbone is well resolved by now, but problems remain within most major clades<sup>10–13</sup>. Multiple classification schemes based on a variety of morphological and anatomical characteristics have been proposed, but none has been fully accepted. Therefore, other sources of data, such as genomic information, are needed for classification. As a foundation for further studies of Lauraceae phylogenomics, here we focus on *Lindera*<sup>14</sup>, a genus belonging to the core Lauraceae in the sense of Rohwer and Rudolph<sup>15</sup>, or the core Laureae in the sense of Chandrabali, *et al.*<sup>11</sup> *Lindera* species have not only popular ornamental and economic uses but also great medicinal and therapeutic value. *Lindera* is widely distributed in tropical, subtropical, and temperate zones in Asia and North America and includes approximately 100 species<sup>16</sup>, with *Lindera umbellata* Thunb. as the type species.

<sup>1</sup>Center for Integrative Conservation, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Menglun, Mengla, Yunnan, 666303, China. <sup>2</sup>Key Laboratory of Tropical Plant Resources and Sustainable Use, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Menglun, Mengla, Yunnan, 666303, China. <sup>3</sup>College of Life Sciences, University of Chinese Academy of Sciences, Beijing, 100049, China. <sup>4</sup>Southeast Asia Biodiversity Research Institute, Chinese Academy of Sciences, Yezin, Nay Pyi Taw, Myanmar. Correspondence and requests for materials should be addressed to Y.S. (email: [songyu@xtbg.ac.cn](mailto:songyu@xtbg.ac.cn)) or Z.-F.X. (email: [zfxu@xtbg.ac.cn](mailto:zfxu@xtbg.ac.cn))

	<i>L. benzoin</i>	<i>L. communis</i>	<i>L. glauca</i>	<i>L. latifolia</i>	<i>L. megaphylla</i>	<i>L. metcalfiana</i>	<i>L. nacusua</i>	<i>L. obtusiloba</i>	<i>L. robusta</i>
Total cpDNA size (bp)	152,478	152,778	152,706	152,779	152,711	152,968	152,211	152,773	152,852
Length of large single copy (LSC) region (bp)	93,573	93,748	93,650	93,792	93,651	93,888	93,735	93,714	93,860
Length of inverted repeat (IRs) region (bp)	20,048	20,066	20,054	20,070	20,066	20,051	20,070	20,075	20,061
Length of small single copy (SSC) region (bp)	18,809	18,898	18,948	18,847	18,928	18,978	18,336	18,909	18,870
Total GC content (%)	39.2	39.2	39.2	39.2	39.2	39.2	39.2	39.1	39.2
LSCGC content (%)	38	38	37.9	38	38	38	38	38	37.9
IRGC content (%)	44.5	44.4	44.5	44.4	44.4	44.5	44.4	44.4	44.5
SSCGC content (%)	34	34	33.8	33.9	33.9	34	34.1	33.8	34
Total number of genes	113	113	113	113	113	113	113	113	113
Total number of protein encoding genes	79	79	79	79	79	79	79	79	79
Total number of tRNA	30	30	30	30	30	30	30	30	30
Total number of rRNA	4	4	4	4	4	4	4	4	4

**Table 1.** Summary of nine complete plastomes of *Lindera*.

The fruits of most *Lindera* species, particularly *Lindera communis* Hemsley and *Lindera glauca* (Siebold & Zucc.) Blume, are rich in fatty oils and thus represent important wild woody oil plants<sup>17–19</sup>. *Lindera megaphylla* Hemsley is an economically important small deciduous tree, the wood of which can be used for buildings and furniture. Moreover, it is also a courtyard greening species, and its trunk and leaves are rich in alkaloids<sup>20,21</sup>. More importantly, *Lindera* plants are widely used in traditional medicine. *Lindera obtusiloba* Blume has been used as a traditional medicine for the treatment of fever, abdominal pain, extravasation, inflammation and poor blood circulation<sup>22–25</sup>. Thus, molecular methods for species delineation in the genus of *Lindera* are of considerable interest.

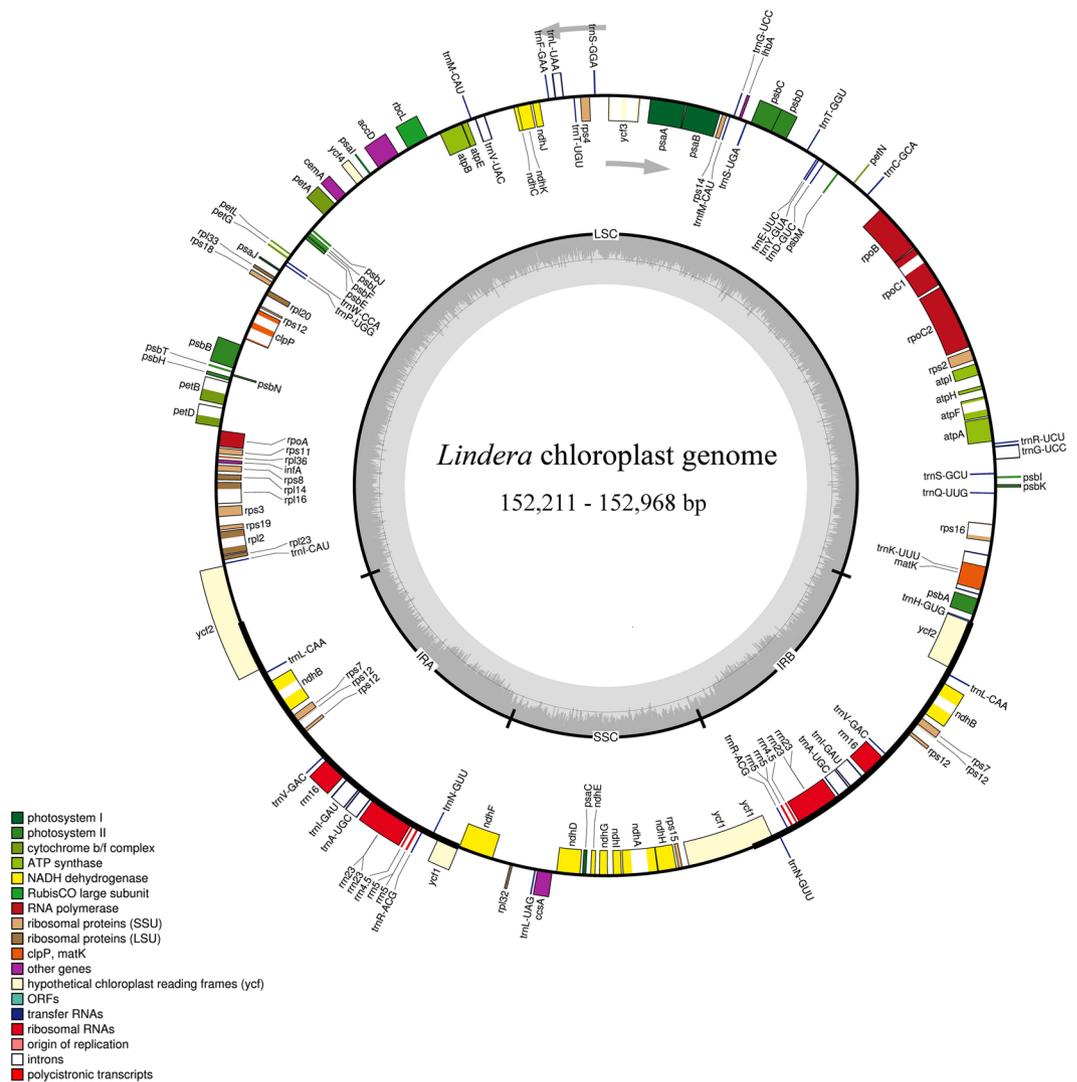
The first reported cp genomic markers in the Lauraceae were *rbcL* and *trnL-trnF*, which were used for phylogenetic analysis of the Laurales<sup>9,26</sup>. The *matK* gene was used to construct a phylogenetic tree to analyse the relationships among Lauraceae genera<sup>10</sup>, but in this analysis the Lauraceae (represented by a single species of *Actinodaphne*, *Laurus*, *Lindera*, *Litsea* and *Neolitsea*) remained unresolved. Then, Chandrabali, *et al.*<sup>11</sup> constructed larger phylogenetic trees for the Lauraceae using the chloroplast sequences *trnL-trnF*, *psbA-trnH*, *trnT-trnL*, and *rpl16* as well as the nuclear barcoding markers 26S rDNA and internal transcribed spacer (ITS) rDNA. The result of their *trnL-trnF* + *psbA-trnH* analysis showed *Lindera erythrocarpa* as sister to *Litsea glaucescens*, albeit without bootstrap support, whereas in their ITS analysis *Lindera erythrocarpa* appeared as sister to *Laurus nobilis*, likewise without bootstrap support. Li, *et al.*<sup>12</sup> and Nie, *et al.*<sup>27</sup>, who used ITS plus different chloroplast markers, found that the genus *Lindera* was not monophyletic. Fijridiyanto and Murakami<sup>28</sup> further analysed the phylogenetic relationships of *Litsea* and its related genera using the nuclear marker *rpb2*, also finding that *Lindera* was not a monophyletic group. Most recently, a report<sup>29</sup> showed a close relationship between the genera *Lindera* and *Litsea* while using *rbcL*, *matK*, *trnH-psbA*, and ITS to investigate the phylogenetic relationships in the Lauraceae.

In this study, we report the complete chloroplast genome sequences of nine *Lindera* species chosen for their economic importance. *Lindera communis*, *L. glauca*, *Lindera latifolia* Hook. f.<sup>14</sup>, and *Lindera nacusua* (D. Don) Merr.<sup>14</sup> are important wild woody oil plants, *L. megaphylla* and *Lindera robusta* (C. K. Allen) H. P. Tsui<sup>14</sup> are important timber plants, and *Lindera benzoin* (L.) Blume<sup>30</sup>, *Lindera metcalfiana* var. *dictyophylla* (C. K. Allen) H. P. Tsui<sup>14</sup>, and *L. obtusiloba* are widely used in traditional medicine. In addition, compared to other Lauraceae species, *L. benzoin* and *L. obtusiloba* are distributed in more northern areas, and their wide distribution shows the ability to adapt to cold environments. Based on cp sequence information, the characteristics and phylogenetic information of these species were further investigated.

## Results

**Characteristics of the cp genomes of *Lindera*.** The sizes of the cp genomes of the nine *Lindera* species range from 152,211 (*L. nacusua*) to 152,968 bp (*L. metcalfiana*) (Table 1). The sequences were assembled into a single, circular, double-stranded DNA sequence for each species. The cp genomes have a typical quadripartite structure, comprising the LSC, with a length of 93,573 (*L. benzoin*) to 93,888 bp (*L. metcalfiana*), the SSC, with a length of 18,336 (*L. nacusua*) to 18,978 bp (*L. metcalfiana*), and a pair of IR copies of 20,048 (*L. benzoin*) to 20,075 bp (*L. obtusiloba*) in length (Fig. 1 and Table 1). The cp sequences of *Lindera communis*, *L. glauca*, *L. latifolia*, *L. megaphylla*, *L. metcalfiana*, *L. obtusiloba*, and *L. robusta* are larger than that of *Litsea glutinosa* (152,618 bp, GenBank accession number KU382356)<sup>8</sup>, but shorter than those of *Phoebe omeiensis* and *P. shearerii* (152,855 bp, GenBank accession number KX427772; 152,876 bp, GenBank accession number KX427773)<sup>31,32</sup>. The cp sequences of *Lindera benzoin* and *L. nacusua* are all shorter than those of *Litsea glutinosa*, *Phoebe omeiensis* and *P. shearerii* (Table 1).

All nine *Lindera* cp genomes contain 113 single-copy genes, among which 79 encode proteins. Sixteen genes have one intron (*atpF*, *ndhA*, *ndhB*, *petB*, *petD*, *rpl2*, *rpl16*, *rpoC1*, *rps12*, *rps16*, *trnA-UGC*, *trnG-UCC*, *trnI-GAU*, *trnK-UUU*, *trnL-UAA* and *trnV-UAC*), and two genes have two introns (*clpP* and *ycf3*) (Table 2). The *ycf1* and *ycf2* genes are trans-spliced, and the nucleotide sequences of the *ycf1* and *ycf2* pseudogenes are 1,372, 1,373, 1,376, 1,377, 1,379, 1,383, and 1,389 bp (*Lindera communis*, *L. megaphylla*, *L. metcalfiana*, *L. latifolia* and *L. nacusua*, *L. robusta*, *L. glauca*, *L. benzoin*, and *L. obtusiloba*), and 3,162 bp (all nine *Lindera* species), respectively, being truncated at the IR boundaries (Fig. 1). Additionally, the *ycf15* gene, located in the LSC of the nine *Lindera* genomes, is



**Figure 1.** Circular gene map of *Lindera* species (*Lindera benzoin*, *L. communis*, *L. glauca*, *L. latifolia*, *L. megaphylla*, *L. metcalfiana*, *L. nacusua*, *L. obtusiloba*, and *L. robusta*) chloroplast genomes. The genes lying outside each circle are transcribed counter-clockwise, while those inside are transcribed clockwise. The coloured bars indicate different functional groups. The dashed darker grey area in the inner circle indicates genome GC content, while the lighter grey area shows AT content. IR = inverted repeat; SSC = small single copy; LSC = large single copy.

also a pseudogene (Table 2). The GC content of these *Lindera* cp genomes is 39.2%, except for *L. obtusiloba* (39.1%), which is similar to that of *Litsea glutinosa* (39.2%), but slightly higher than those of *Phoebe omeiensis* and *P. shearerii* (39.1%). The GC content of *L. obtusiloba* is the same as those of *Phoebe omeiensis* and *P. shearerii*<sup>32</sup> (Table 1).

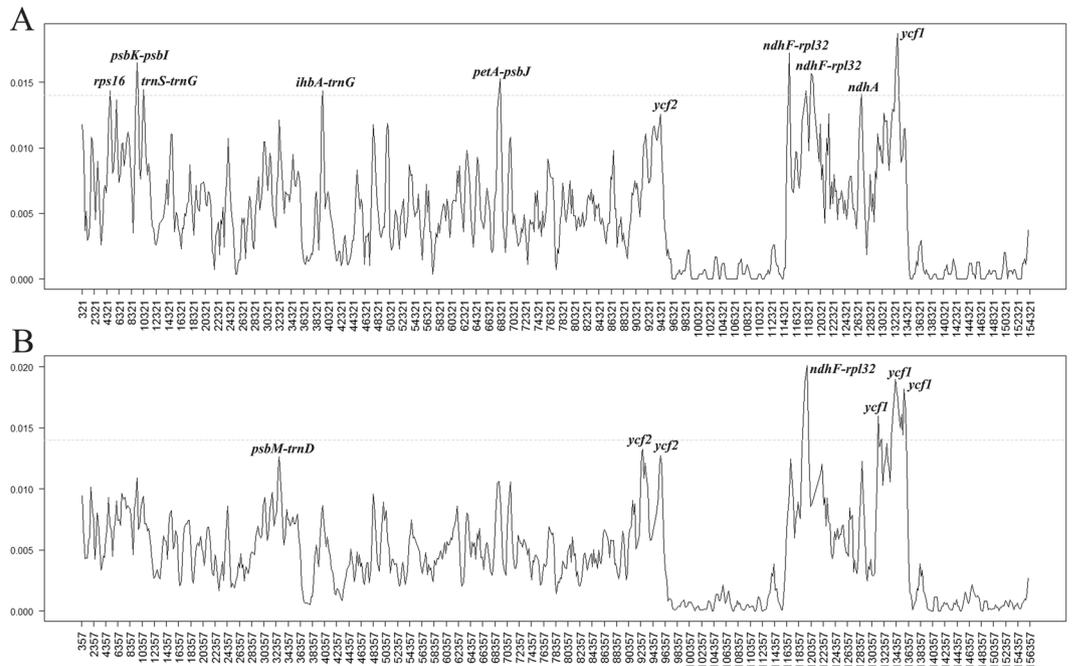
**Identification of the most variable regions.** To elucidate levels of sequence divergence, we calculated the nucleotide variability ( $P_i$ ) values. The  $P_i$  values within 600 bp across the nine genomes vary from 0 to 0.0187, and the mean value is 0.0048 (Fig. 2A), indicating that these sequences have high similarity. However, we identified nine hypervariable loci ( $P_i > 0.014$ ), which are *ihbA-trnG*, *ndhA*, *ndhF-rpl32*, *petA-psbJ*, *psbK-psbI*, *rps16*, *trnS-trnG*, and *ycf1*. The *ndhA* and *ndhF-rpl32* loci are in the SSC region; *ihbA-trnG*, *petA-psbJ*, *psbK-psbI*, *rps16*, and *trnS-trnG* are in the LSC region; and *ycf1* is in the IR region (Fig. 2A). To investigate the levels of sequence divergence among the genera, we calculated the genetic divergence of the sequenced cp genomes of core Lauraceae, including *Alseodaphne*, *Cinnamomum*, *Laurus*, *Lindera*, *Litsea*, *Machilus*, *Nectandra*, *Persea*, *Phoebe*, and *Sassafras*. The  $P_i$  values vary from 0 to 0.0201 in these 29 sequences (Fig. 2B), indicating that the variation at genus level is significantly larger than that at the species level. All these genomic features are shown in the sequence alignment of the nine *Lindera* species, *Laurus nobilis*, and *Litsea glutinosa* (Fig. S1). According to the alignment results (Fig. S1), all of these species share the same order and orientation of syntenic blocks, indicating that no rearrangement occurred in gene organization. These results accord with those of Male, *et al.*<sup>33</sup> and Asif, *et al.*<sup>34</sup>, which illustrate that cp genomes tend to be conserved and perfectly collinear, especially in the same plant family.

Category for genes	Group of genes	Name of genes
Photosynthesis related genes	Rubisco	<i>rbcL</i>
	Photosystem I	<i>psaA, psaB, psaC, psaI, psaJ</i>
	Assembly/stability of photosystem I	** <i>ycf3, ycf4</i>
	Photosystem II	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, ihbA</i>
	ATP synthase	<i>atpA, atpB, atpE, *atpF, atpH, atpI</i>
	cytochrome b/f complex	<i>petA, *petB, *petD, petG, petL, petN</i>
	cytochrome c synthesis	<i>ccsA</i>
	NADPH dehydrogenase	<i>*ndhA, *ndhB, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i>
Transcription and translation related genes	transcription	<i>rpoA, rpoB, *rpoC1, rpoC2</i>
	ribosomal proteins	<i>rps2, rps3, rps4, rps7, rps8, rps11, rps12, *rps12, rps14, rps15, *rps16, rps18, rps19, *rpl2, rpl14, *rpl16, rpl20, rpl23, rpl32, rpl33, rpl36</i>
	translation initiation factor	<i>infA</i>
RNA genes	ribosomal RNA	<i>rrn4.5, rrn5, rrn16, rrn23</i>
	transfer RNA	<i>*trnA-UGC, trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnG-UCC, *trnG-UCC, trnH-GUG, trnI-CAU, *trnI-GAU, *trnK-UUU, trnL-CAA, *trnL-UAA, trnL-UAG, trnM-CAU, trnM-CAU, trnN-GUU, trnP-UGG, trnQ-UUG, trnR-ACG, trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-UGU, trnV-GAC, *trnV-UAC, trnW-CCA, trnY-GUA</i>
Other genes	RNA processing	<i>matK</i>
	carbon metabolism	<i>cemA</i>
	fatty acid synthesis	<i>accD</i>
	proteolysis	** <i>clpP</i>
Genes of unknown function	conserved reading frames	<i>ycf1, ycf2</i>
Pseudogenes		<i>ycf15</i>

**Table 2.** Genes encoded by nine *Lindera* plastomes. Note: Asterisks (\*) before gene names indicate intron containing genes, and double asterisks (\*\*) indicate two introns in the gene.

**Comparative analysis of cp genomes.** Size variation in cp genomes is partly a result of contraction and expansion at the borders of the IR regions<sup>1</sup>. To trace the size differences among *Lindera* cp genomes, the IR-LSC and IR-SSC boundaries, with full annotations for the adjacent genes, were re-examined across nine *Lindera* cp genomes (Fig. 3). The entire *ycf1* gene crosses the SSC/IRB boundary, while another fragment of  $\psi$ *ycf1* is located at the IRA/SSC boundary. A  $\psi$ *ycf1* fragment with a length of 1,372–1,389 bp was found in the IRA region because the boundary between the SSC and IRB extended into the *ycf1* gene. In the nine *Lindera* species, the distances between  $\psi$ *ycf1* and *ndhF* vary from 6 (*L. obtusiloba*) to 37 bp (*Litsea glutinosa*) in length (Fig. 3).

**Phylogenomic analysis of sequenced Lauraceae plastomes.** To determine the phylogenetic relationships of the nine *Lindera* species with other Lauraceae, we reconstructed a phylogenetic tree based on our nine complete cp genomes of *Lindera* and 23 fully sequenced cp genomes of related Lauraceae (Fig. 4). The tree shows that the *Lindera* species can be divided into two sub-clades. Sub-clade I (*Lindera benzoin*, *L. latifolia*, *L. metcalfiana*, *L. obtusiloba*, and *L. robusta*) is sister to sub-clade II, which contains the remaining species.



**Figure 2.** Comparison of the nucleotide variability (Pi) values of the nine *Lindera* plastomes (A) and 29 plastomes of the core Lauraceae (B). X axis: position of the midpoint of a window, Y axis: nucleotide diversity of each window.

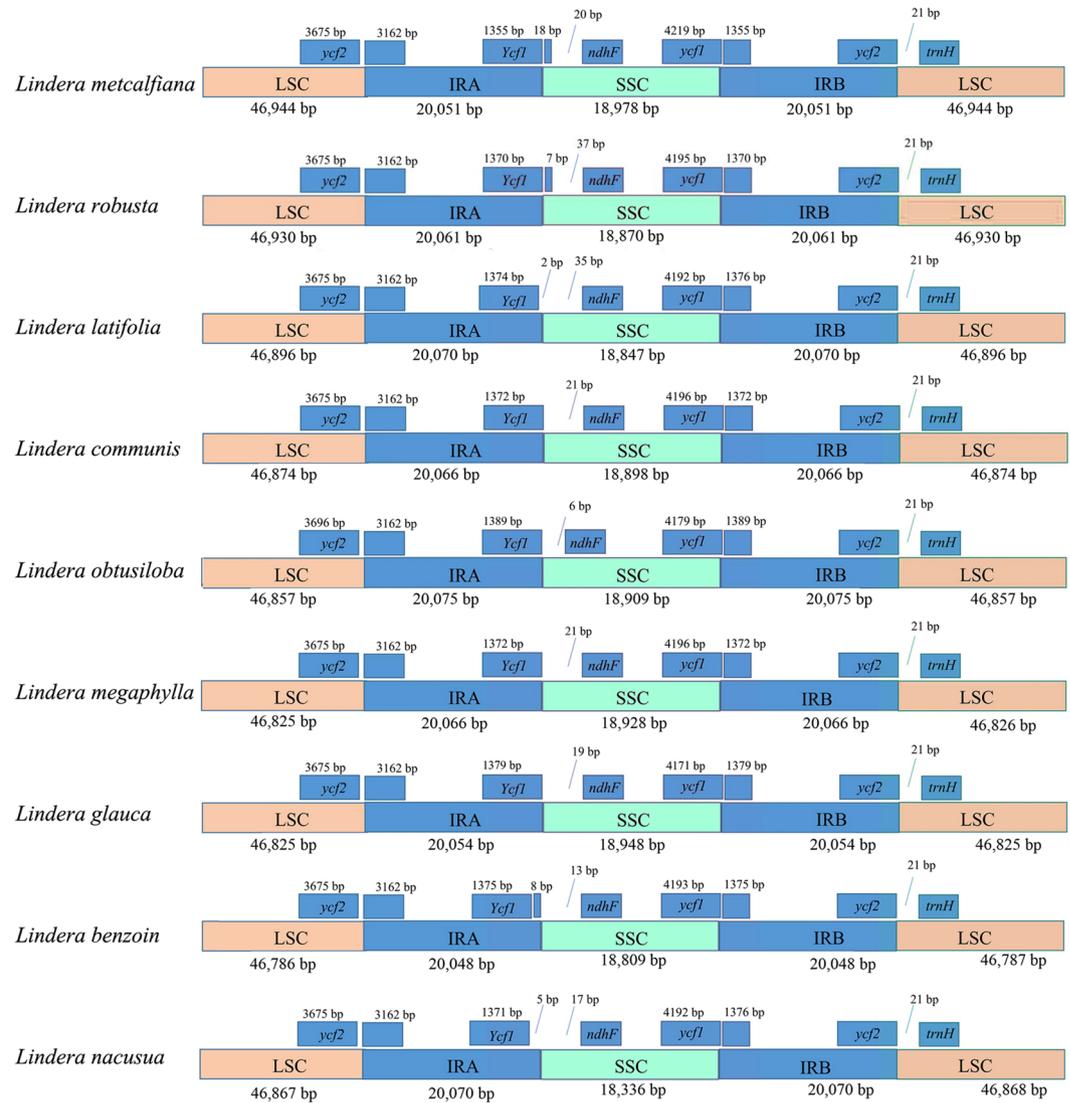
Sub-clade I, however, has only 78% bootstrap support, whereas sub-clade II is 100% supported. Within sub-clade I, part I (*Lindera communis*, *L. glauca*, and *L. nacusua*) is sister to part II (*Laurus nobilis*, *Lindera megaphylla*, and *Litsea glutinosa*) (Fig. 4). These data indicate that (among the species investigated) *Laurus nobilis* and *Litsea glutinosa* are most closely related to *Lindera megaphylla*, and the *Lindera* group, which is most closely related to the *Cinnamomum-Ocotea* clade, comprises nine *Lindera* species, *Laurus nobilis* and *Litsea glutinosa*.

**Phylogenetic analysis of *Lindera* species.** To better understand the phylogenetic relationships between our sequenced nine *Lindera* species and the other *Lindera* taxa with reported barcoding data, we downloaded available sequences from GenBank including *rbcL*, *matK*, *trnL-trnF*, *psbA-trnH*, *ndhF*, *ITS* and *rpb2* of all 33 core Lauraceae species (Table S1). The result of this phylogenetic analysis supports the grouping of all *Lindera* species with three *Litsea* species and *Laurus nobilis* (Fig. 5). This tree is further divided into four main sub-clades. Sub-clade I (Bayesian inference posterior probability, BI-PP = 1.00) includes *Lindera communis*, *L. fragrans*, *L. glauca*, *L. megaphylla*, *L. nacusua*, plus *Laurus nobilis*, *Lindera glutinosa*, and *Litsea tomentosa*. Sub-clade II includes only *Lindera obtusiloba*. Sub-clade III (BI-PP = 1.00) includes *Lindera erythrocarpa*, *L. latifolia*, *L. longipedunculata*, *L. lucida*, *L. metcalfiana*, *L. polyantha*, *L. robusta*, and *Litsea cubeba*. Sub-clade IV (BI-PP = 1.00) includes *Lindera aggregata*, *L. benzoin*, *L. chunii*, *L. fruticosa*, *L. kariensis*, *L. pulcherrima*, *L. reflexa*, *L. triloba*, *L. umbellata*, and *L. villipes*.

## Discussion

In this study, the complete cp genomes of nine *Lindera* species were sequenced using Illumina sequencing technology. These nine cp genomes possess the typical angiosperm quadripartite structure, which has a large single copy (LSC) region and a small single copy (SSC) region, separated by two short inverted repeat (IR) regions (Fig. 1). Similar to those of other sequenced Lauraceae species, the nine *Lindera* cp genomes have less length variation in the IR regions but more variation in the LSC and SSC regions. The length of the IR regions ranges from 20,048 bp in *L. benzoin* to 20,075 bp in *L. obtusiloba*. As Song, *et al.*<sup>35</sup> reported, both IR regions of *Machilus balansae* and *M. yunnanensis* are 20,074 bp in length. The LSC region of *Lindera metcalfiana* is 93,888 bp in length, which is 315 bp larger than that of *L. benzoin*. The SSC region of *L. metcalfiana* is 18,978 bp in length, which is 642 bp larger than that of *L. nacusua*. A comparative analysis was conducted to explain these size differences, and the results suggest the following: the length of the entire *ycf1* gene ranges from 5,550 bp in *L. glauca* to 5,574 bp in *L. megaphylla*; the length of the truncated *ycf1* gene ranges from 1,372 bp in *L. communis* and *L. megaphylla* to 1,389 bp in *L. obtusiloba*; the length of the entire *ycf2* gene ranges from 6,837 bp in *L. metcalfiana* to 6,858 bp in *L. obtusiloba* (Fig. 3). It has been reported that the *ycf1* and *ycf2* genes are located in the boundaries between the IR regions and the LSC and SSC regions, and these two genes experienced incomplete duplication<sup>35,36</sup>. Thus, length changes in *ycf1*, *ycf2*,  $\psi ycf1$ , and *ndhF-ψycf1* drive the contraction and expansion of the IR regions in the cp genomes of *Lindera*.

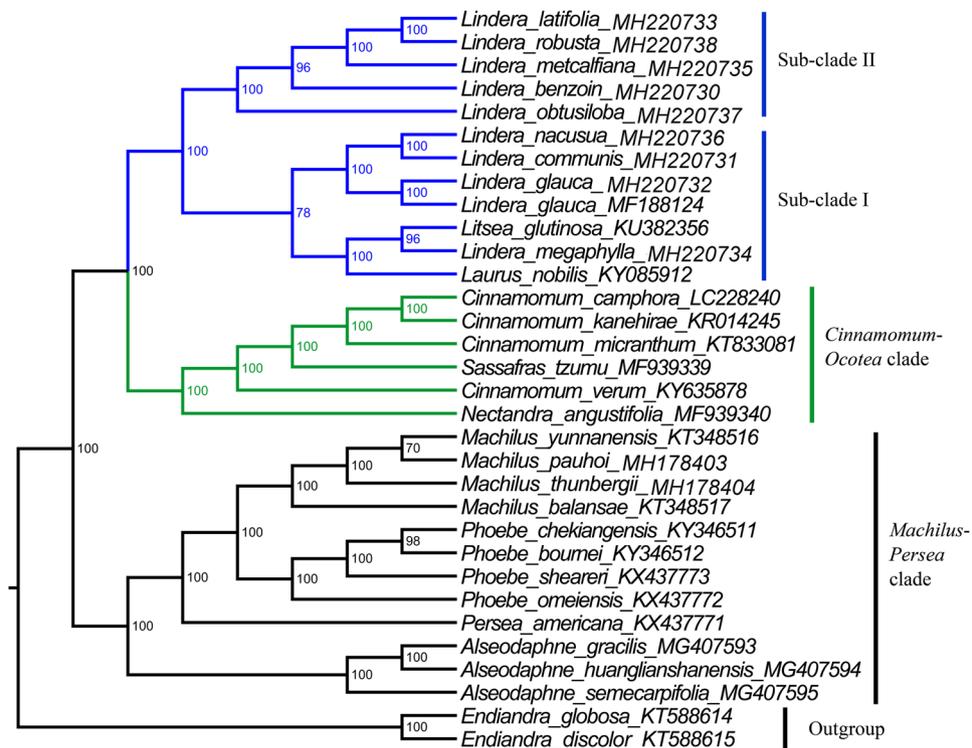
The *ycf1* was identified as a hypervariable locus at the species level within *Lindera*, as were *ihbA-trnG*, *ndhA*, *ndhF-rpl32*, *petA-psbJ*, *psbK-psbI*, *rps16*, and *trnS-trnG*. At the genus level, we identified *ndhF-rpl32* and *ycf1*



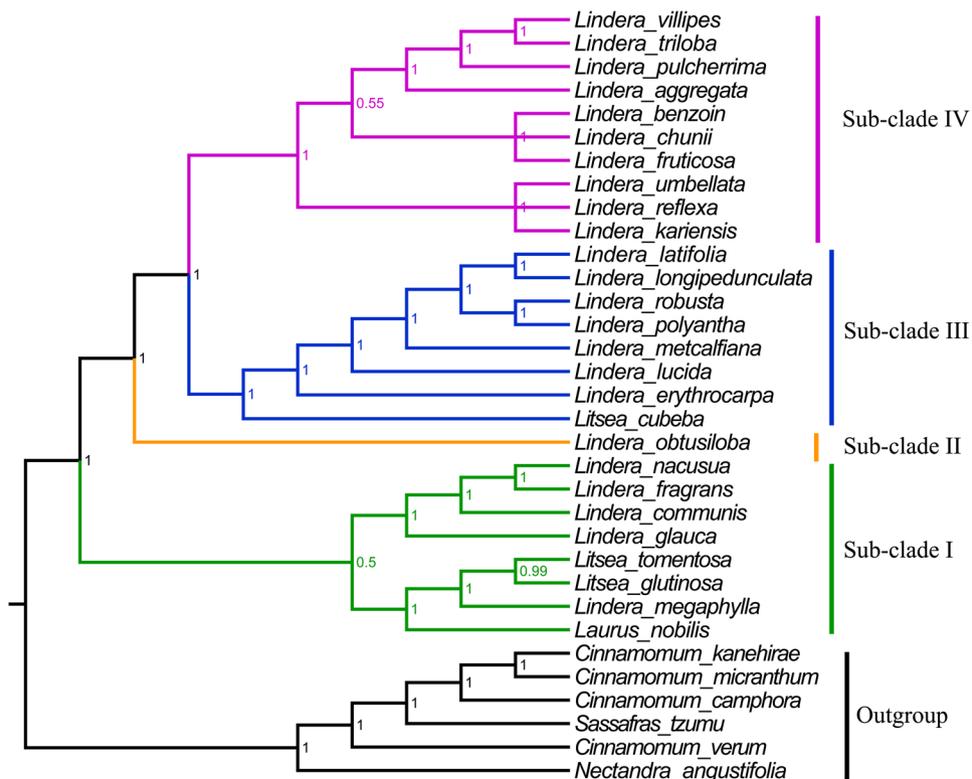
**Figure 3.** Comparison of LSC, IR, and SSC junction positions among nine *Lindera* chloroplast genomes.

as variable regions among 29 core Lauraceae species from *Alseodaphne*, *Cinnamomum*, *Laurus*, *Lindera*, *Litsea*, *Machilus*, *Nectandra*, *Persea*, *Phoebe*, and *Sassafras*. Among these regions, *ndhF-rpl32*, and *ycf1* have been shown to be variable enough across seed plants for use as barcodes in plant taxonomy and phylogeny<sup>31,35,37–40</sup>. The four fragments *ndhF-rpl32*, *petA-psbJ*, *trnS-trnG*, and *ycf1*, have previously been identified as hypervariable regions in *Litsea glutinosa*, *Machilus balansae*, *M. yunnanensis*, *Persea americana*, *Phoebe omeiensis*, and *P. shearerii*<sup>8,31,33,35</sup>. In addition to the previous results, we found that the fragments *ihbA-trnG*, *ndhA*, *petA-psbJ*, *psbK-psbI*, *rps16*, and *trnS-trnG* seemed to be especially variable loci in *Lindera* plastomes, and they showed promising levels of variation for application in DNA barcoding or intraspecific studies.

Previously, hypervariable loci have been used as barcoding markers for taxon identification and phylogenetic analysis<sup>3,38,41</sup>. The chloroplast gene fragments *rbcl*, *matK*, and *psbA-trnH* and the nuclear internal transcribed spacer (ITS) have been reported as available markers for plant barcoding<sup>3,29,38,41,42</sup>. Our phylogenetic analysis using 17 complete cp genome sequences, five chloroplast regions (*rbcl*, *matK*, *trnL-trnF*, *psbA-trnH*, and *ndhF*), and two nuclear sequences (*ITS* and *rpb2*) of another 14 *Lindera* and two *Litsea* species shows that a group containing 23 *Lindera* species, three *Litsea* species, and *Laurus nobilis* was most closely related to a *Cinnamomum-Ocotea* clade, with strong support<sup>11,43–45</sup>, as in previously published phylogenetic trees. The two species *Litsea glutinosa* and *L. tomentosa* as well as *Laurus nobilis* are located in a sub-clade with five *Lindera* species, *Lindera communis*, *L. fragrans*, *L. glauca*, *L. megaphylla*, and *L. nacusua*, while *Litsea cubeba* was located in another sub-clade with seven *Lindera* species, *Lindera erythrocarpa*, *L. latifolia*, *L. longipedunculata*, *L. lucida*, *L. metcalfiana*, *L. polyantha*, and *L. robusta*, which is in agreement with a previous phylogenetic result by Fijridiyanto and Murakami<sup>28</sup> who defined the relationships among seven *Lindera* species. And the clade containing *Laurus*, *Lindera*, and *Litsea* was the sister group of the *Cinnamomum-Ocotea* clade containing *Cinnamomum*, *Nectandra*, and *Sassafras* species, as found in previous studies<sup>11,43–45</sup>. In our study, we further determined the relationships of 16 additional *Lindera* species, *Lindera benzoin*, *L. chunii*, *L. communis*, *L. fragrans*, *L. fruticosa*, *L. kariensis*, *L.*



**Figure 4.** Molecular phylogenetic tree of 32 species of Lauraceae based on complete plastome sequences. Numbers at the nodes represent bootstrap percentages.



**Figure 5.** Phylogenetic relationships of 33 core Lauraceae species based on concatenated sequences of seven loci. Numbers at the nodes represent Bayesian inference posterior probabilities (BI-PP).

Name	Herbarium	Taxon	Voucher	Geographic origin	GenBank Accession number
<i>L. benzoin</i>	HITBC-BRG	<i>Lindera benzoin</i> (L.) Blume	SY34259	Beijing Botanical Garden Institute of Botany Beijing, China	MH220730
<i>L. communis</i>	HITBC-BRG	<i>Lindera communis</i> Hemsley	SY01432	Xishuangbanna Tropical Botanical Garden Yunnan, China	MH220731
<i>L. glauca</i>	HITBC-BRG	<i>Lindera glauca</i> (Siebold & Zuccarini) Blume	SY34253	Beijing Botanical Garden Institute of Botany Beijing, China	MH220732
<i>L. latifolia</i>	HITBC-BRG	<i>Lindera latifolia</i> Hook. f.	SY33219	Xishuangbanna Tropical Botanical Garden Yunnan, China	MH220733
<i>L. megaphylla</i>	HITBC-BRG	<i>Lindera megaphylla</i> Hemsley	SY33127	Kunming Institute of Botany Yunnan, China	MH220734
<i>L. metcalfiana</i>	HITBC-BRG	<i>Lindera metcalfiana</i> var. <i>dictyophylla</i> (C. K. Allen) H. P. Tsui	SY34595	Xishuangbanna Tropical Botanical Garden Yunnan, China	MH220735
<i>L. nacusua</i>	HITBC-BRG	<i>Lindera nacusua</i> (D. Don) Merr.	SY34537	South China Botanical Garden Guangzhou, China	MH220736
<i>L. obtusiloba</i>	HITBC-BRG	<i>Lindera obtusiloba</i> Blume	SY34165	Beijing Botanical Garden Institute of Botany Beijing, China	MH220737
<i>L. robusta</i>	HITBC-BRG	<i>Lindera robusta</i> (C. K. Allen) H. P. Tsui	SY34225	South China Botanical Garden Guangzhou, China	MH220738

**Table 3.** Sampled species and their voucher specimens used in this study.

*latifolia*, *L. longipedunculata*, *L. megaphylla*, *L. metcalfiana*, *L. nacusua*, *L. pulcherrima*, *L. reflexa*, *L. robusta*, *L. triloba*, and *L. villipes*. The results of our study are fully compatible with those of Fijridiyanto and Murakami<sup>28</sup>, as far as well-supported clades are concerned. A topological difference is found in the placement of *Lindera glauca*, in a clade (with *Actinodaphne* and *Neolitsea*, not examined here) that is sister to all taxa examined in both studies in Fijridiyanto and Murakami, vs. in sub-clade I here. This sub-clade, however, has practically no support.

Overall, our study reports nine chloroplast genomes of *Lindera* for the first time and compares their organizations with those of other Lauraceae species. Four divergent regions were found at the genus level, providing a valuable source of markers for future studies on species delineation and resolution of phylogenetic relationships among the Lauraceae. Our phylogenomic results also suggest that plastid phylogenomics can be regarded as a robust method for tackling difficult phylogenies and could be extended across the genera *Lindera* and *Litsea* with additional systematic sampling.

## Materials and Methods

**Plant materials.** Nine *Lindera* species were used for this research. Fresh young leaves of *Lindera benzoin*, *L. communis*, *L. glauca*, *L. latifolia*, *L. megaphylla*, *L. metcalfiana*, *L. nacusua*, *L. obtusiloba*, and *L. robusta* were harvested from mature plants growing in botanical gardens (Table 3) and then immediately frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$ . Specimens were deposited in the herbarium of the Biodiversity Research Group (BRG) of Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences (CAS).

**DNA preparation and chloroplast sequencing.** The cetyltrimethylammonium bromide (CTAB) method was used to extract total genomic DNA. Next-generation sequencing was performed according to Yang *et al.*<sup>46</sup>, and nine universal primer pairs from their study were also taken to perform long-range PCR. Then, the PCR products were purified and combined. Following the manufacturer's instructions (Illumina Nextera XT library), the mixture was fragmented and used to construct 500-bp short-insert libraries. All nine complete *Lindera* cp genomes were sequenced using a Genome Analyzer (Illumina HiSeq2000) at the Germplasm Bank of Wild Species, Kunming Institute of Botany, CAS.

**Cp genome assembly and annotation.** All nine *Lindera* raw sequencing reads were filtered using the NGS QC Tool Kit to obtain high-quality short reads. Then, the raw reads were trimmed according to their quality, removing bases from the 5' and 3' ends until no base with  $Q < 20$  was found. After that, the chloroplast genome was de novo assembled using the trial version of CLC v.8 (<http://www.qiagenbioinformatics.com>). The contigs were aligned using the publicly available cp genome of *Litsea glutinosa*<sup>8</sup> in Geneious 4.8 (<http://www.geneious.com/>) as a reference. Dual Organellar GenoMe Annotator (DOGMA, <http://dogma.cccb.utexas.edu/>) software was used to annotate the cp genomes and identify genes encoding proteins<sup>47</sup>, transfer RNAs (tRNAs), and ribosomal RNAs (rRNAs). The OrganellarGenomeDRAW tool (OGDRAW, <http://ogdraw.mpimp-golm.mpg.de/>) was used to draw the genome maps of *Lindera benzoin*, *L. communis*, *L. glauca*, *L. latifolia*, *L. megaphylla*, *L. metcalfiana*, *L. nacusua*, *L. obtusiloba*, and *L. robusta*.

**Sliding window analysis to identify hypervariable regions.** MAFFT (<http://mafft.cbrc.jp/alignment/server>) was used to align the nine cp genomes with one another. Afterward, we manually adjusted these sequences using BioEdit software (<http://www.mbio.ncsu.edu/bioedit/bioedit.html>). DnaSP version 5.0 with a sliding window analysis was used to calculate the nucleotide variability values ( $\pi$ ) within the chloroplast genomes. The window length was set to 600 bp and the step size to 200 bp. Then, the R program was used to plot values. In addition, we used this method to analyse the hypervariable regions among the nine *Lindera* cp genomes and those of *Alseodaphne gracilis* (GenBank accession number MG407593)<sup>48</sup>, *A. huanglianshanensis* (GenBank accession

number MG407594)<sup>48</sup>, *A. semecarpifolia* (GenBank accession number MG407595)<sup>48</sup>, *Cinnamomum camphora* (GenBank accession number LC228240)<sup>49</sup>, *C. kanehirae* (GenBank accession number KR014245)<sup>50</sup>, *C. micranthum* (GenBank accession number KR014245)<sup>50</sup>, *C. verum* (GenBank accession number KY635878)<sup>9</sup>, *Laurus nobilis* (GenBank accession number KY085912)<sup>13</sup>, *Lindera glauca* (GenBank accession number MF188124), *Litsea glutinosa* (GenBank accession number KU382356)<sup>8</sup>, *Machilus balansae* (GenBank accession number KT348517)<sup>35</sup>, *M. pauhoi* (GenBank accession number MH178403), *M. thunbergii* (GenBank accession number MH178404), *M. yunnanensis* (GenBank accession number KT348516)<sup>35</sup>, *Nectandra angustifolia* (GenBank accession number MF939340)<sup>13</sup>, *Persea americana* (GenBank accession number KX437771)<sup>31</sup>, *P. chekiangensis* (GenBank accession number KY346511), *Phoebe omeiensis* (GenBank accession number KX437772)<sup>31,32</sup>, *P. shearereri* (GenBank accession number KX437773)<sup>31,32</sup>, *P. zhenan* (GenBank accession number KY346512), and *Sassafras tzumu* (GenBank accession number MF939339)<sup>13</sup>.

**Phylogenetic analyses.** The plastome sequences of *Alseodaphne gracilis*, *A. huanglianshanensis*, *A. semecarpifolia*, *Cinnamomum camphora*, *C. kanehirae*, *C. micranthum*, *C. verum*, *Endiandra discolor* (GenBank accession number KT588615)<sup>51</sup>, *E. globosa* (GenBank accession number KT588614)<sup>51</sup>, *Laurus nobilis*, *Lindera glauca*, *Litsea glutinosa*, *Machilus balansae*, *M. pauhoi*, *M. thunbergii*, *M. yunnanensis*, *Nectandra angustifolia*, *Persea americana*, *Phoebe chekiangensis*, *P. omeiensis*, *P. shearereri*, *P. zhenan*, and *Sassafras tzumu* were downloaded from the NCBI GenBank. Thereafter, we used MAFFT to align these 23 sequences and our nine *Lindera* sequences and manually edited where necessary with BioEdit software. After these steps, the jModelTest 2.0 program was used to calculate nucleotide substitution, and the results indicated that the optimal model was “GTR + G”<sup>52</sup>. The “GTR + G” model was used for all maximum likelihood (ML) analyses implemented in RAXML version 8.0.20, as suggested in the manual<sup>53</sup>. Nonparametric bootstrapping was performed with the “fast bootstrap” algorithm of RAXML and 1000 replicates. The cp genomes of *Endiandra discolor* and *E. globosa* were used as the outgroup.

We assembled sequence alignments for 33 taxa of Lauraceae, 17 (*Cinnamomum camphora*, *C. kanehirae*, *C. micranthum*, *C. verum*, *Laurus nobilis*, *Lindera benzoin*, *L. communis*, *L. glauca*, *L. latifolia*, *L. megaphylla*, *L. metcalfiana*, *L. nacusua*, *L. obtusiloba*, *L. robusta*, *Litsea glutinosa*, *Nectandra angustifolia*, and *Sassafras tzumu*) with complete cp genome sequences, plus ITS and *rpb2* sequences, and 14 *Lindera* species (*Lindera aggregata*, *L. chunii*, *L. erythrocarpa*, *L. fragrans*, *L. fruticosa*, *L. kariensis*, *L. longipedunculata*, *L. lucida*, *L. polyantha*, *L. pulcherrima*, *L. reflexa*, *L. triloba*, *L. umbellata*, and *L. villipes*), and two *Litsea* species (*Litsea cubeba* and *L. tomentosa*) with the DNA sequences of *rbcL*, *matK*, *trnL-trnF*, *psbA-trnH*, *ndhF*, ITS<sup>54</sup> and *rpb2* from GenBank (Table S1). MAFFT software was used to analyse these sequences, and incongruous sequences of the same species were removed. The seven-sequence matrix was then manually adjusted, and Sequencher 4.10 (<http://www.genecodes.com>) was used to merge identical sequences. Then, a joint matrix was constructed using Sequence Matrix v.1.7.8, and jModelTest 2.0 was used to calculate the nucleotide substitution. The optimal model was chosen (“GTR + I + G”) (freqA = 0.3003, freqC = 0.1988, freqG = 0.1919, freqT = 0.3090, R(a) [AC] = 0.8998, R(b) [AG] = 2.0890, R(c) [AT] = 0.2648, R(d) [CG] = 0.4178, R(e) [CT] = 1.9183, R(f) [GT] = 1.0000, p-inv = 0.7020, gamma shape = 0.3000) to construct the phylogenetic tree<sup>52</sup>. Phylogenetic relationships were reconstructed using Bayesian inference (BI) and ML methods in MrBayes version 3.1.2<sup>55</sup>.

**Data archiving statement.** The complete cp genome sequence data of the nine *Lindera* taxa have been submitted to the GenBank of NCBI. The GenBank accession numbers are MH220730 (*Lindera benzoin*), MH220731 (*L. communis*), MH220732 (*L. glauca*), MH220733 (*L. latifolia*), MH220734 (*L. megaphylla*), MH220735 (*L. metcalfiana*), MH220736 (*L. nacusua*), MH220737 (*L. obtusiloba*), and MH220738 (*L. robusta*).

## References

- Zhao, Y. B. *et al.* The complete chloroplast genome provides insight into the evolution and polymorphism of *Panax ginseng*. *Frontiers in Plant Science* **5**, 696, <https://doi.org/10.3389/fpls.2014.00696> (2014).
- Palmer, J. D. Comparative organization of chloroplast genomes. *Annual Review of Genetics* **19**, 325–354, <https://doi.org/10.1146/annurev.ge.19.120185.001545> (1985).
- Hollingsworth, P. M. *et al.* A DNA barcode for land plants. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 12794–12797, <https://doi.org/10.1073/pnas.0905845106> (2009).
- Ma, P. F., Zhang, Y. X., Zeng, C. X., Guo, Z. H. & Z., L. D. Chloroplast phylogenomic analyses resolve deep-level relationships of an intractable bamboo Tribe *Arundinarieae* (Poaceae). *Systematic Biology* **63**(6), 933–950, <https://doi.org/10.1093/sysbio/syu054> (2014).
- Yang, Y. C. *et al.* Comparative analysis of the complete chloroplast genomes of five *Quercus* species. *Frontiers in Plant Science* **7**, 959, <https://doi.org/10.3389/fpls.2016.00959> (2016).
- Zhang, S. D. *et al.* Diversification of Rosaceae since the Late Cretaceous based on plastid phylogenomics. *The New Phytologist* **214**, 1355–1367, <https://doi.org/10.1111/nph.14461> (2017).
- Rohwer, J. G. Lauraceae. In: Kubitzki, K., Rohwer, J. G. & Bittrich, V. (eds) *The Families and Genera of Vascular Plants*, vol. 2. Springer-Verlag, Berlin, Heidelberg, New York, London, Paris, Tokyo, Hong Kong, Barcelona, Budapest, ISBN 3-540-55509-9 and 0-387-55509-9, [https://doi.org/10.1007/978-3-662-02899-5\\_46](https://doi.org/10.1007/978-3-662-02899-5_46) (1993).
- Hinsinger, D. D. & Strijk, J. S. Toward phylogenomics of Lauraceae: The complete chloroplast genome sequence of *Litsea glutinosa* (Lauraceae), an invasive tree species on Indian and Pacific Ocean islands. *Plant Gene* **9**, 71–79, <https://doi.org/10.1016/j.plgene.2016.08.002> (2017).
- Renner, S. S. Circumscription and phylogeny of the Laurales: evidence from molecular and morphological data. *American Journal of Botany* **86**(9), 1301–1315 (1999).
- Rohwer, J. G. Toward a phylogenetic classification of the Lauraceae: evidence from *matK* sequences. *Systematic Botany* **25**(1), 60–71, <https://doi.org/10.2307/2666673> (2000).
- Chanderbali, A. S., van der Werf, H. & Renner, S. S. Phylogeny and historical biogeography of Lauraceae: evidence from the chloroplast and nuclear genomes. *Annals of the Missouri Botanical Garden* **88**, 104–134, <https://doi.org/10.2307/2666133> (2001).
- Li, J., Christophel, D. C., Conran, J. G. & Li, H.-W. 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, <https://doi.org/10.1007/s00606-003-0113-z> (2004).

13. Song, Y. *et al.* Evolutionary comparisons of the chloroplast genome in Lauraceae and insights into loss events in the Magnoliids. *Genome Biology and Evolution* **9**, 2354–2364, <https://doi.org/10.1093/gbe/evx180> (2017).
14. Cui, H. B. & Weff, H. S. *LINDERA* Thunberg, Nov. Gen. Pl. 64. 1783, nom. cons., not Adanson (1763). *Flora of China* **7**, 142–159 (2008).
15. Rohwer, J. G. & Rudolph, B. Jumping genera: the phylogenetic positions of *Cassytha*, *Hypodaphnis*, and *Neocinnamomum* (Lauraceae) based on different analyses of *trnK* intron sequences. *Annals of the Missouri Botanical Garden* **92**(2), 153–178 (2005).
16. Cao, Y. *et al.* The genus *Lindera*: a source of structurally diverse molecules having pharmacological significance. *Phytochemistry Reviews* **15**, 869–906, <https://doi.org/10.1007/s11101-015-9432-2> (2015).
17. Zhou, Z. Y., Wang, W. G., Duan, J. F., Zhu, L. W. & Jiang, J. X. Chinese spicehush (*Lindera communis*) seed oil catalyzed into biodiesel by enzymatic method. *Applied Mechanics and Materials* **448–453**, 1587–1591, <https://doi.org/10.4028/www.scientific.net/AMM.448-453.1587> (2013).
18. Wang, W.-G., Ma, L., Jiang, J.-X., Jiang, L. & Zhang, Z.-X. Transesterified Chinese spicehush (*Lindera communis*) seed oil as a biodiesel fuel. *Forestry Studies in China* **9**, 132–136, <https://doi.org/10.1007/s11632-007-0020-4> (2007).
19. Niu, J. *et al.* Transcriptome analysis of distinct *Lindera glauca* tissues revealed the differences in the unigenes related to terpenoid biosynthesis. *Gene* **559**, 22–30, <https://doi.org/10.1016/j.gene.2015.01.002> (2015).
20. Chou, C. J., Lin., L. C., Chen., K. T. & Chen, C. F. Northalifoline, a new isoquinolone alkaloid from the pedicels of *Lindera megaphylla*. *Journal of Natural Products* **57**(6), 689–694 (1994).
21. Chen, C. C., Lin, C. F. & Huang, Y. L. Bioactive constituents from the flower buds and peduncles of *Lindera megaphylla*. *Journal of Natural Products* **59**(9), 1426–1425 (1995).
22. Choi, H. G. *et al.* A new neolignan and lignans from the stems of *Lindera obtusiloba* Blume and their anti-allergic inflammatory effects. *Archives of Pharmacal Research* **37**, 467–472, <https://doi.org/10.1007/s12272-013-0239-7> (2014).
23. Hong, C. O., Lee, H. A., Rhee, C. H., Choung, S. Y. & Lee, K. W. Separation of the antioxidant compound quercitrin from *Lindera obtusiloba* Blume and its antimelanogenic effect on B16F10 melanoma cells. *Bioscience, Biotechnology, and Biochemistry* **77**, 58–64, <https://doi.org/10.1271/bbb.120562> (2013).
24. Freise, C., Kim, K. Y. & Querfeld, U. A *Lindera obtusiloba* extract blocks calcium-/phosphate-induced transdifferentiation and calcification of vascular smooth muscle cells and interferes with matrix metalloproteinase-2 and metalloproteinase-9 and NF- $\kappa$ B. *Evidence-based Complementary and Alternative Medicine: eCAM* **2015**, 679238, <https://doi.org/10.1155/2015/679238> (2015).
25. Ruehl, M. *et al.* Extracts of *Lindera obtusiloba* induce antifibrotic effects in hepatic stellate cells via suppression of a TGF- $\beta$ -mediated profibrotic gene expression pattern. *The Journal of Nutritional Biochemistry* **20**, 597–606, <https://doi.org/10.1016/j.jnutbio.2008.06.003> (2009).
26. Renner, S. S. Phylogenetic affinities of Monimiaceae based on cpDNA gene and spacer sequences. *Perspectives in Plant Ecology, Evolution and Systematics* **1**, 61–77 (1998).
27. Nie, Z. L., Wen, J. & Sun, H. Phylogeny and biogeography of *Sassafras* (Lauraceae) disjunct between eastern Asia and eastern North America. *Plant Systematics and Evolution* **267**, 191–203, <https://doi.org/10.1007/s00606-007-0550-1> (2007).
28. Fijridiyanto, I. A. & Murakami, N. Phylogeny of *Litsea* and related genera (Laureae-Lauraceae) based on analysis of *otrpb2* gene sequences. *Journal of Plant Research* **122**, 283–298, <https://doi.org/10.1007/s10265-009-0218-8> (2009).
29. Liu, Z. F. *et al.* DNA barcoding evaluation and implications for phylogenetic relationships in Lauraceae from China. *Plos One* **12**, e0175788, <https://doi.org/10.1371/journal.pone.0175788> (2017).
30. Tucker, O., Maciarello, M. J., Burbage, P. W. & George, S. Spicebush [*Lindera benzoin* (L.) Blume var. *benzoin*, Lauraceae]: a tea, spice, and medicine. *Economic Botany* **48**(3), 333–336 (1994).
31. Song, Y., Yao, X., Tan, Y., Gan, Y. & Corlett, R. T. Complete chloroplast genome sequence of the avocado: gene organization, comparative analysis, and phylogenetic relationships with other Lauraceae. *Canadian Journal of Forest Research* **46**, 1293–1301, <https://doi.org/10.1139/cjfr-2016-0199> (2016).
32. Song, Y. *et al.* Comparative analysis of complete chloroplast genome sequences of two subtropical trees, *Phoebe sheareri* and *Phoebe omeiensis* (Lauraceae). *Tree Genetics & Genomes* **13**, <https://doi.org/10.1007/s11295-017-1196-y> (2017).
33. Male, P. J. *et al.* Genome skimming by shotgun sequencing helps resolve the phylogeny of a pantropical tree family. *Molecular Ecology Resources* **14**, 966–975, <https://doi.org/10.1111/1755-0998.12246> (2014).
34. Asif, M. H. *et al.* Complete sequence and organisation of the *Jatropha curcas* (Euphorbiaceae) chloroplast genome. *Tree Genetics & Genomes* **6**, 941–952, <https://doi.org/10.1007/s11295-010-0303-0> (2010).
35. Song, Y. *et al.* Comparative analysis of complete chloroplast genome sequences of two tropical trees *Machilus yunnanensis* and *Machilus balansae* in the family Lauraceae. *Frontiers in Plant Science* **6**, 662, <https://doi.org/10.3389/fpls.2015.00662> (2015).
36. Li, P. *et al.* Comparative genomics and phylogenomics of East Asian tulips (*Amana*, Liliaceae). *Frontiers in Plant Science* **8**, 451, <https://doi.org/10.3389/fpls.2017.00451> (2017).
37. Korotkova, N., Nauheimer, L., Ter-Voskanyan, H., Allgaier, M. & Borsch, T. Variability among the most rapidly evolving plastid genomic regions is lineage-specific: implications of pairwise genome comparisons in *Pyrus* (Rosaceae) and other Angiosperms for marker choice. *Plos One* **9**, e112998, <https://doi.org/10.1371/journal.pone.0112998> (2014).
38. Dong, W. P. *et al.* *ycf1*, the most promising plastid DNA barcode of land plants. *Scientific Reports* **5**, 8348, <https://doi.org/10.1038/srep08348> (2015).
39. Dong, W. P., Liu, J., Yu, J., Wang, L. & Z. S. Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. *Plos One* **7**(4), e35071, <https://doi.org/10.1371/journal.pone.0035071.t001> (2012).
40. Sarkinen, T. & George, M. Predicting plastid marker variation: can complete plastid genomes from closely related species help? *Plos One* **8**, e82266, <https://doi.org/10.1371/journal.pone.0082266> (2013).
41. Dong, W. P. *et al.* Discriminating plants using the DNA barcode *rbclb*: an appraisal based on a large data set. *Molecular Ecology Resources* **14**, 336–343, <https://doi.org/10.1111/1755-0998.12185> (2014).
42. Rohwer, J. G. *et al.* Is *Persea* (Lauraceae) monophyletic? Evidence from nuclear ribosomal ITS sequences. *Taxon* **58**(4), 1153–1167 (2009).
43. Huang, J. F. *et al.* Origins and evolution of cinnamon and camphor: A phylogenetic and historical biogeographical analysis of the *Cinnamomum* group (Lauraceae). *Molecular phylogenetics and evolution* **96**, 33–44, <https://doi.org/10.1016/j.ympev.2015.12.007> (2016).
44. Trofimov, D., Rudolph, B. & Rohwer, J. G. Phylogenetic study of the genus *Nectandra* (Lauraceae), and reinstatement of *Damburneya*. *Taxon* **65**(5), 980–996, <https://doi.org/10.12705/655.3> (2016).
45. Rohde, R. *et al.* Neither *Phoebe* nor *Cinnamomum* – the tetrasporangiate species of *Aiouea* (Lauraceae). *Taxon* **66**(5), 1085–1111, <https://doi.org/10.12705/665.6> (2017).
46. Yang, J. B., Li, D. Z. & Li, H. T. Highly effective sequencing whole chloroplast genomes of angiosperms by nine novel universal primer pairs. *Molecular Ecology Resources* **14**, 1024–1031, <https://doi.org/10.1111/1755-0998.12251> (2014).
47. Wyman, S. K., Jansen, R. K. & Boore, J. L. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* **20**(17), 3252–3255, <https://doi.org/10.1093/bioinformatics/bath352>. PMID:15180927 (2004).
48. Song, Y., Yao, X., Liu, B., Tan, Y. & Corlett, R. T. Complete plastid genome sequences of three tropical *Alseodaphne* trees in the family Lauraceae. *Holzforchung* **72**, 337–345, <https://doi.org/10.1515/hf-2017-0065> (2018).
49. Liu, Z., Chen, S. L., Song, J. Y., Zhang, S. J. & Chen, K. L. Application of deoxyribonucleic acid barcoding in Lauraceae plants. *Pharmacognosy magazine* **8**, 4–11, <https://doi.org/10.4103/0973-1296.93301> (2012).

50. Wu, C. C., Chu, F. H., Ho, C. K., Sung, C. H. & Chang, S. H. Comparative analysis of the complete chloroplast genomic sequence and chemical components of *Cinnamomum micranthum* and *Cinnamomum kanehirae*. *Holzforchung* **71**(3), 189–197, <https://doi.org/10.1515/hf-2016-0133> (2017).
51. Rossetto, M., Kooyman, R., Yap, J. Y. & Laffan, S. W. From ratites to rats: the size of fleshy fruits shapes species' distributions and continental rainforest assembly. *Proceedings. Biological Sciences* **282**, 20151998, <https://doi.org/10.1098/rspb.2015.1998> (2015).
52. Darrriba, D., Taboada, G. L., Doallo, R. & Posada, D. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* **9**, 772, <https://doi.org/10.1038/nmeth.2109> (2012).
53. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313, <https://doi.org/10.1093/bioinformatics/btu033> (2014).
54. Li, L., Li, J., Conran, J. G., Li, X. W. & Li, H. W. Phylogeny of *Neolitsea* (Lauraceae) inferred from Bayesian analysis of nrDNA ITS and ETS sequences. *Plant Systematics and Evolution* **269**, 203–221, <https://doi.org/10.1007/s00606-007-0580-8> (2007).
55. Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574, <https://doi.org/10.1093/bioinformatics/btg180> (2003).

## Acknowledgements

The authors would like to thank Dr. Bing Liu from the State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, for providing plant tissues of *Lindera benzoin* and *L. glauca*, and acknowledge Jun-Bo Yang, Jing Yang, Juan-Hong Zhang, Chun-Yan Lin, Zheng-Shan He, and Ji-Xiong Yang at the Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences, for sequencing technology. This work was supported by the CAS “Light of West China” Programme (No. Y7XB061B01), the project of the Southeast Asia Biodiversity Research Institute (No. Y4ZK111B01), and the National Natural Science Foundation of China (No. 31600531). We sincerely thank two anonymous referees for their critical and invaluable comments that greatly improved our manuscript.

## Author Contributions

Mei-Li Zhao and Yu Song designed the experiments; Mei-Li Zhao, Yun-Hong Tan and Yu Song collected the samples; Mei-Li Zhao conducted the experiments; Mei-Li Zhao, Jun Ni and Xin Yao analysed the results; and Mei-Li Zhao, Yu Song and Zeng-Fu Xu wrote the manuscript. All authors read and approved the final manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-27090-0>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018