Pak. J. Bot., 50(6): 2189-2196, 2018.

COMPLETE CHLOROPLAST GENOME OF A VALUABLE ECONOMIC TREE, *LINDERA GLAUCA* (LAURACEAE) AND COMPARISON WITH ITS CONGENERS

BIAO XIONG[1], LIMEI ZHANG[2], LEI XIE[1], SHUBIN DONG[3] AND ZHIXIANG ZHANG[1*]

[1]*Laboratory of Systematic Evolution and Biogeography of Woody Plants, School of Nature Conservation, Beijing Forestry University, No.35 Tsinghua East Road, Beijing 100083, China*
[2]*College of landscape and architecture, Beijing Forestry University, No.35 Tsinghua East Road, Beijing 100083, China*
[3]*College of Biological Sciences and Technology, Beijing Forestry University, No.35 Tsinghua East Road, Beijing 100083, China*
[*]*Corresponding author's email: zxzhang@bjfu.edu.cn*

Abstract

Lindera glauca (Lauraceae), widely distributed in the subtropical region, is a tree of economic and ecological significance. Its fruits are rich in oil and **wild resources are abundant.** Here, we assembled the complete chloroplast genome sequence of *Lindera glauca* and compared it to chloroplast genomes of seven other species in Lauraceae. Cleaned reads from paired-end Illumina sequencing were mapped to a published plastome (*Litsea glutinosa*) and mapped onto the chloroplast genome of *L. glauca* using the BLAT and Geneious programs successively. The complete chloroplast genome size was 152,780 bp with 39.16% GC content, containing a pair of inverted repeats of 20,059 bp, which were separated by a large and a small single copy region of 93,809 bp and 18,853 bp, respectively. The genome encoded 125 genes, including 81 protein-coding, 36 transfer RNA, and 8 ribosomal RNA genes. In addition, 67 microsatellite sites were found, and 5 primer pairs were developed that were polymorphic in 90 *L. glauca* individuals. Phylogenetic analysis based on whole chloroplast sequences demonstrated that *L. glauca* was closely related to *Litsea glutinosa*, and *Cinnamomum micranthum* was sister to the rest of the species sampled. The complete chloroplast genome presented here is the first for *L. glauca* is and is the first in the genus *Lindera*. These results have the potential to be a valuable genome resource for further studies of the phylogenetics, genetic variation, and population genetic structure of the important economic plant.

Key words: *Lindera glauca*, Chloroplast genome, Chloroplast microsatellite, Illumina sequencing, Lauraceae.

Introduction

Lindera glauca (Sieb. et Zucc.) Blume (Lauraceae), a deciduous shrub or small tree, is widely distributed in low-altitude montane forests of central and southern of China, and is also found in Japan, Korea, Vietnam, and Taiwan (Wang *et al*., 1972; Chang, 1976; Ding *et al*., 1981; Tsui *et al*., 1982; Zheng, 1983). The fruits of *L. glauca* is rich in fatty acids and aromatic oils and contain terpenoids, flavonoids, and alkaloids, which are used to treat disease symptoms in folk, and are used as raw materials to produce medicines, lubricants, and biochemical products (Liu *et al*., 1982; Zheng, 1983; Wang *et al*., 1994; Sun *et al*., 2011; Kim *et al*., 2014; Suh *et al*., 2015). Some root extracts, such as N-methyllaurotetanine, exhibit apparent anticarcinogenic activities (Wang *et al*., 2011), and some volatile oils from leaves can be used in industrial spices (Sun *et al*., 2011). Owing to its increasing applications in diverse products and potentially great economic value, *L. glauca* has attracted increasing attention in recent years. However, phylogenetic relationships within the genus *Lindera* Thunb. (ca. 100 species) are still unclear because of insufficient molecular phylogenetic data. In addition, Dupont (2002) found that *L. glauca* could asexually reproducing by seed, and only female individuals are found in Japan. However, the male individuals of *L. glauca* are very rare in China mainland according to our successive four years of field survey, even though this species with sexual reproduction and male plants have been known to exist in continental East Asia for several decades (Wang, 1972; Tsui *et al*., 1982).

The chloroplast (cp) is a unique organelle and an essential role in the plant cell for conducting photosynthesis and carbon fixation (Neuhaus and Emes, 2000). Owing to its haploidy, conserved and stable genome structure, the rate of evolution, and gene content and gene order in most angiosperm plants (Daniell *et al*., 2016; Zhang *et al*., 2016), chloroplast genomes have been a focus of research in plant phylogeographic and molecular evolution. *L. glauca* belongs to a genus that **was revised based** on morphological taxonomy (Tsui *et al*., 1982; Tsui & Werff, 2008) and is not well circumscribed with the genera *Litsea* and *Neolitsea* having low (<50%) bootstrap support based on cpDNA sequences *trnL-trnF* and *psbA-trnH* (Chanderbali *et al*., 2001). Thus, sequencing the complete cp genome of *L. glauca* may not only elucidate the relationships within this genus, but also guide further exploration of the draft cp genome in closely related species. Further, the maternally inherited chloroplast genome of *L. glauca* may be helpful in exploring the reproductive mode (in particular apomixis) of this species.

In this study, complete chloroplast genome sequence data for *L. glauca* was obtained using Illumina sequencing technology. That is the first complete cp genome sequence in the genus *Lindera* Thunb. We compared this sequence with all seven previously published complete cp genomes in Lauraceae (up to July 1, 2017 in GenBank) to better understand the mutation patterns of the species cp genome. The resulting data may provide valuable information for phylogenetic and genome evolution among other species in genus *Lindera*. Furthermore, the resultant genomic resource may facilitate further studies of population structure, biogeography, and biochemistry for this species, especially the study of reproductive mode that includes seed propagation, intraspecific apomixis, and older root-sprout regeneration.

Materials and Methods

Sampling, DNA extraction and sequencing: We collected fresh young leaves of wild *L. glauca* from the Jigong Mountain National Nature Reserve (JGS) in Xinyang City of the Henan Province (Table 1); samples were immediately frozen in dry ice and stored at −80°C until further use. Total genomic DNA was extracted from 1 g of the leaves using a modified cetyltrimethylammonium bromide (CTAB) method (Doyle & Doyle, 1987), and was then sequenced by Majorbio (Shanghai, China) using an Illumina Hiseq 4000 platform (San Diego, California, USA). A sequencing library (PE 150×2) was then established. In addition, in order to verify chloroplast microsatellite (cpSSR) loci polymorphism, genomic DNA of 90 individuals of plant materials from our previous work (Xiong *et al*., 2016) were extracted based on same modified CTAB method.

Chloroplast genome assembly: Adapters and barcodes were cleaned from Illumina paired-end (PE 150×2) raw reads and then quality filtered using Trimmomatic (Bolger *et al*., 2014). Reads were trimmed from both ends, and all individual bases with a Phred quality score <20 were removed, as well as those with more than three consecutive uncalled bases. After trimming, all reads of a median quality score lower than 25 bp in length were discarded (BioProject ID: PRJNA399837; BioSample accession: SAMN07551814; SRA accession no. SRR5980347). Resulting clean reads from this strategy were mapped to a plastome of *Litsea glutinosa* (KU382356) published on GenBank (Hinsinger and Strijk, 2017), using the BLAT v.32×1 program with a Python script (Kent, 2002; Weitemier *et al*., 2014) to exclude nuclear and mitochondrial reads. Using this method, we obtained the cp reads (gene fragments) of *L. glauca* from high-quality reads (after cleaning) using High Performance Computing (HPC) at the University of Florida. All selected cp reads were subsequently mapped to the complete cp genome of *Litsea glutinosa*. This species has the closest relationship with *L. glauca* based on morphological taxonomy among the seven members of Lauraceae with published complete cp genomes (Tsui & Werff, 2008). The selected cp reads were then clustered into the cp genome using Geneious v8.1.9 (Biomatters, Auckland, New Zealand). An initial annotation of the chloroplast genome was conducted using both Chloroplast Genome Annotation, Visualization, Analysis, and GenBank Submission (CPGAVAS; Liu *et al*., 2012) and Dual Organeller Genome Annotator (DOGMA; Wyman *et al*., 2004). The two software programs identify putative protein-coding genes by running BLASTX searches against a custom database of published chloroplast sequences. Since results of the annotations from both software programs are not very accurate, we compared them with each other and then performed manual adjustments. Furthermore, the tRNAscan-SE v1.21 program (Schattner *et al*., 2005) was used to check the identified tRNA-coding and rRNA-coding genes. The circle cp genome map of *L. glauca* was drawn using the Organellar Genome DRAW program (OGDRAW; Lohse *et al*., 2007).

Detection of chloroplast SSRs, primer analysis, and markers development: For chloroplast microsatellites (cpSSR), the distribution, type, and presence of microsatellites were studied in the complete cp genome of *L. glauca* using the Microsatellite Identification Tool (MISA; Thiel *et al*., 2003) with at least 10, 6, 5, 4, 3, 3, and 2 SSR motif repeats for mono-, di-, tri-, tetra-, penta-, hexa-, and higher-order nucleotides, respectively. In addition, cpSSRs of all seven previously published complete cp genomes in Lauraceae were detected using the same parameters: *Cinnamomum micranthum* (GenBank accession no. KR014245), *Machilus yunnanensis* (KT348516), *Machilus balansae* (KT348517), *Persea americana* (KX437771), *Phoebe omeiensis* (KX437772), *Phoebe sheareri* (KX437773), and *Litsea glutinosa* (KU382356).

Additionally, the program PRIMER 3 (PRIMER-E, Auckland, New Zealand) was used to design primer pairs with primers length varying from 18 to 28 bp, annealing temperatures between 55°C and 65°C, GC content from 40% to 60%, and sizes of amplification products from 100 to 400 bp. Owing to high false positives for mononucleotide repeats, we initially selected primer pairs from non-mononucleotide repeat motifs to amplify fragments from 18 individuals of *L. glauca* using polyacrylamide gel electrophoresis. Then, the primer pairs that successfully amplified fragments and were found to be polymorphic in the 18 individuals were labeled with fluorescently labeled nucleotides (forward primer with M13F) and were used to amplify in all 90 individuals, using capillary gel electrophoresis. The next steps followed methods and PCR assay described in previous work (Xiong *et al*., 2016). Fragment analysis was performed with an ABI 3730XL DNA Analyzer (Applied Biosystems) with a GeneScan 500 LIZ Size Standard, and the alleles were determined and manually scored using GeneMarker version 2.2.0 software (SoftGenetics, State College, Pennsylvania, USA). To evaluate the polymorphism of each cpSSR locus, the resulting genotypic data were analyzed to obtain standard descriptive statistics using GENEPOP v4.4 software (Rousset, 2008), including gene diversity (*H*), observed number of alleles (*Na*), and effective number of alleles (*He*).

Table 1. Sampling information for fresh leaves of *Lindera glauca* collected in this study.

Populations	Samples accession no.	Geographic coordinates		Elevation (m)	Province (China)	County
		Latitude	Longitude			
JGS	J14-09-28	31°52′15″N	114°05′13″E	293	Henan	Xinyang

Note: Samples accession numbers refer to voucher specimens deposited in the Museum of Beijing Forestry University (BJFC); Geographic coordinates and elevationare obtained with portable GPS receiver

Table 2. Sequencing data statistics of *Lindera glauca* chloroplast.

Sample	Raw data pair reads	Raw data total bases (bp)	Q20 (%)	Clean data pair reads	Clean data single reads	Clean data total bases (bp)
Lindera glauca	28,845,467×2	8,711,331,034	92.86	28,309,392×2	442,766	8,174,085,667

Table 3. Characteristics of *Lindera glauca* chloroplast genomes.

Category	*Lindera glauca*
Total cpDNA Size (bp)	152,780
Length of large single copy (LSC) region	93,809
Length of inverted repeat (IR) region	20,059
Length of small single copy (SSC) region	18,853
Number of total genes	125
Number of protein-coding genes[a]	81(3)
Number of tRNA genes[a]	36(6)
Number of rRNA genes[a]	8(4)
Overall GC content (%)	39.16%
GC content in LSC	37.98%
GC content in SSC	33.85%
GC content in IR	44.45%

Note: The numbers in brackets indicate the genes duplicated in one IR region

Table 4. Statistics of different SSR motif types in *Lindera glauca* chloroplast genomes.

SSR motif type	SSR abundances
Mono-(A)	25
Mono-(T)	33
Di-(AT)	4
Di-(TC)	2
Di-(TA)	2
Di-(GA)	1
Total	67

Phylogenetic analysis and genome comparison: To investigate the phylogenetic relationships of *L. glauca*, its complete cp genome sequences were compared to that of the seven species in Lauraceae, one species in Laurales (*Calycanthus floridus* var. *glaucus*; AJ428413), and one species in Magnoliales (*Liriodendron tulipifera*; DQ899947) as outgroup. The whole cp genome sequences were aligned using MAFFT-win (Katoh & Standley, 2013) with default settings. The eight sequences were aligned to explore a series of data partitions based on the complete data set using maximum-parsimony (MP) analysis with PAUP 4.0a152 (http://paup.csit.fsu.edu/). Bootstrap support values for individual clades were calculated by running 10,000 bootstrap replicates of the data. To investigate levels of sequence divergence across the chloroplast, we calculated the genetic divergence of the whole cp genome between *L. glauca* and six other species of the Lauraceae family, i.e., *M. yunnanensis*, *M. balansae*, *P. americana*, *P. omeiensis*, *P. sheareri*, and *Litsea glutinosa*, using DnaSP v5.0 software (Librado & Rozas, 2009). In addition, the mVISTA program was used to compare the cp genome of *L. glauca* with six species of Lauraceae (*M. yunnanensis*, *M. balansae*, *P. americana*, *P. omeiensis*, *P. sheareri*, and *Litsea glutinosa*) in a Shuffle-LAGAN mode (Frazer *et al*., 2004) using the *L. glauca* annotation as a reference.

Results

Chloroplast genome content of *L. glauca*: Overall, 56,618,784 paired-end reads (150 bp ×2; clean data) with an insert size of ~150 bp were constructed, and 442,766 single reads were generated using a total of 8,174,085,667 bp (Table 2) A total of 121,704 reads (0.215%) that exclude nuclear and mitochondrial reads were generated and the per-site coverage was from 53X to 174X across whole cp genome sequence of *L. glauca*, with a mean of 118X. This assembly was of high quality and without gaps. The complete cp genome sequence of *L. glauca* (GenBank accession no.MF188124) had a length of 152,780 bp and showed a characteristic circular structure of plastome (Fig. 1), consisting of a pair of inverted repeats (IRs) (20,059 bp each) separated by a large single copy (LSC) (93,809bp) and a small single copy (SSC) (18,853bp) region. The assembled cp genomes encoded 125 genes: 81 protein-coding genes, 36 transfer RNA (tRNA) genes, and 8 ribosomal RNA (rRNA) genes. The IRs region contained 13 genes: 3 protein-coding genes, 6 tRNA genes, and 4 rRNA genes per IR region. Furthermore, the overall GC content of the whole genome was 39.16% (Table 3).

Identification, development and validation of chloroplast SSR: A total of 67 cpSSRs of *L. glauca* were detected with the MISA program and compared to cpSSRs of seven other species in Lauraceae based on the same parameters (Fig. 2). *L. glauca* had the most cpSSRs among the 8 Lauraceae species. The most abundant motifs of *L. glauca* were mononucleotide repeats, accounting for about 86.57% of the total cpSSRs, followed by dinucleotides and hybrid-nucleotides. In contrast, tri-, tetra-, penta-, and higher-order nucleotides were not detected within the cp genome (Table 4). Of these abundant motifs, the LSC regions contained 54, while the IRs and SSC regions contained 2 and 11, respectively. Among all cpSSRs, 9 non-mononucleotide cpSSRs were detected and five of them were successfully amplified and found to be polymorphic in a total of 90 *L. glauca* individuals. The *Na* ranged from 2 to 3 with a mean of 2.400. The *He* and *H* varied from 1.069 to 1.254 and from 0.064 to 0.203, respectively (Table 5).

Phylogenetic and comparative analysis of chloroplast genome: Among the Lauraceae, the complete cp genome of *L. glauca*(152,780 bp) was larger than complete cp genome of five species, i.e., *C. micranthum* (152,700 bp), *M. yunnanensis* (152,721 bp), *M. balansae* (152,622 bp), *P. americana* (152,733 bp), and *Litsea glutinosa* (152,618 bp), and it was smaller than two species, i.e., *P. omeiensis* (152,855 bp) and *P. sheareri* (152,876 bp). Sixteen distinct genes contained one intron, i.e., *rps*16, *atp*F, *rpo*C1, *pet*B, *pet*D, *rp*l16, *ndh*A, *ndh*B, *trn*K-UUU, *trn*G-UCC, *trn*L-UAA, *trn*V-UAC, *trn*L-GAU, *trn*A-UGC, *trn*A-UGC, and *trn*L-GAU, while three genes contained two introns (*rps*12, *clp*P, and *ycf*3).

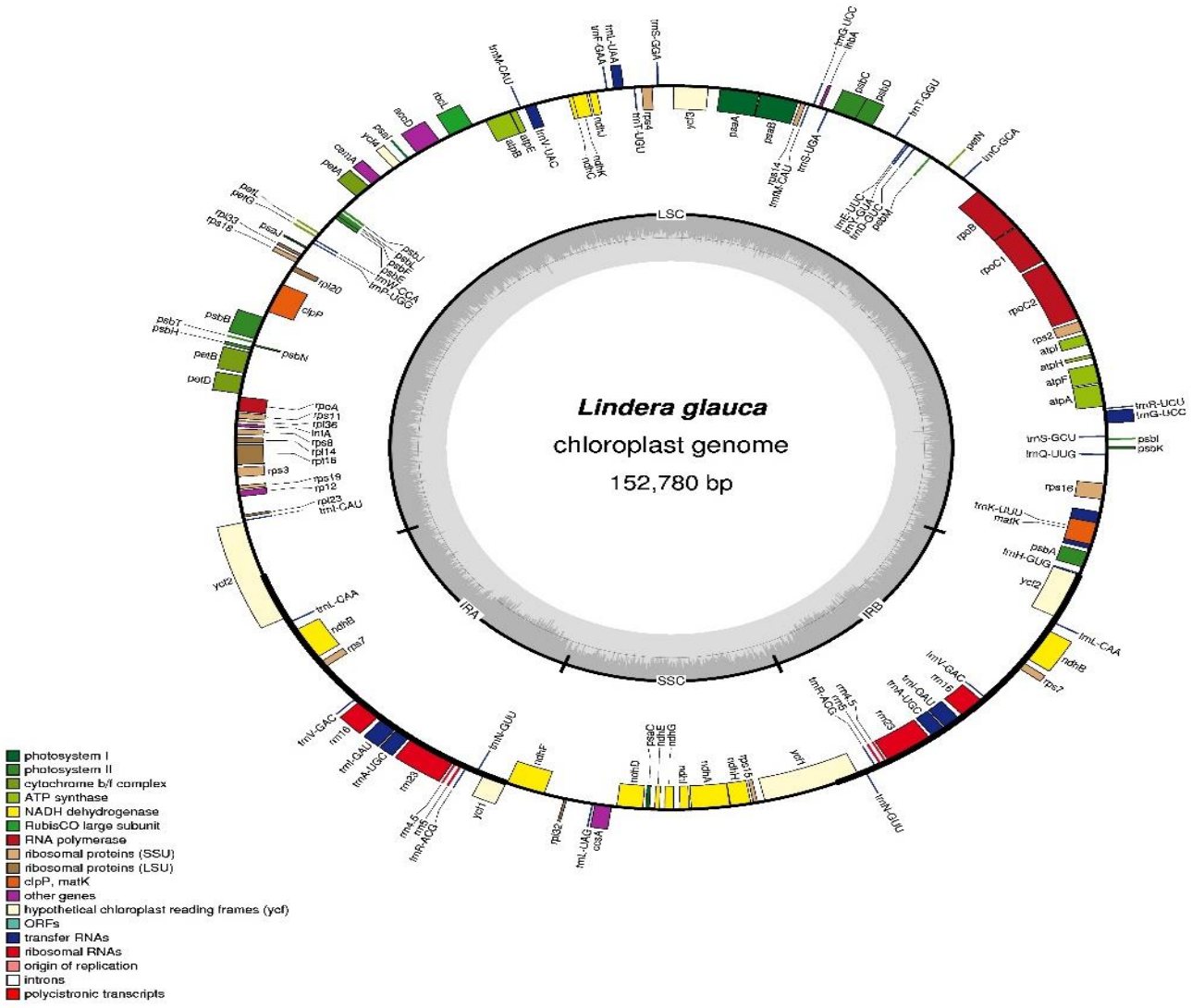

Fig. 1. Gene map of the *Lindera glauca* chloroplast genome. The annotation of the genome was performed using CPGAVAS and DOGMA. The genes that are drawn outside of the circle are transcribed clockwise, while those inside are counterclockwise. Small single copy (SSC), large single copy (LSC), and inverted repeats (IRa, IRb) are indicated.

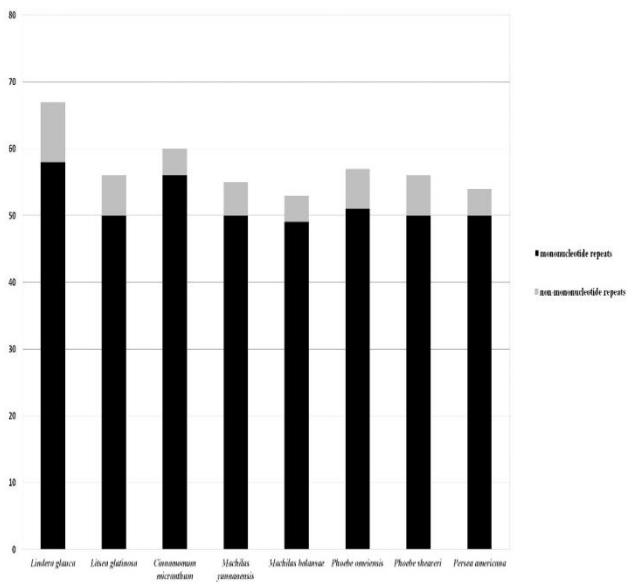

Fig. 2. Number of simple sequence repeats (SSRs) in eightcomplete chloroplast genomes of Lauraceae.

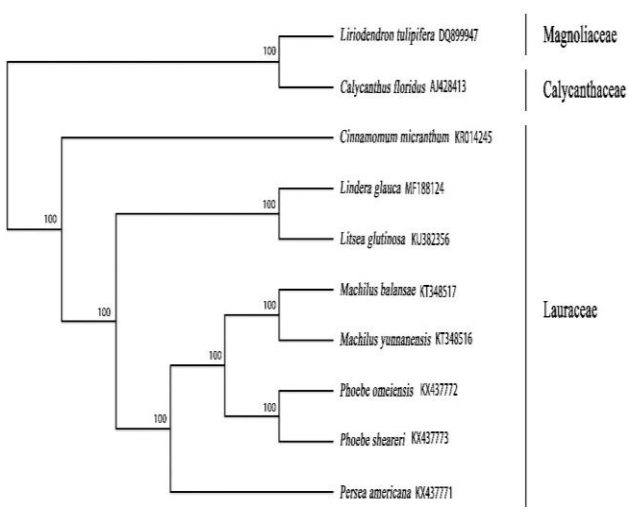

Fig. 3. Maximum Parsimony (MP) phylogenetic tree of eight species of Lauraceae based on complete plastome sequences. Bootstrapping analysis (10000 replicates) in the Paup program. The tree is rooted with the plastome sequence of *Calycanthus fertilis* and *Cinnamomum micranthum*.

Table 5. Characteristics of five chloroplast micros atellite loci developed for 90 *Lindera glauca* individuals.

Locus	Primer sequence (5′-3′)	Repeat motif	Size (bp)	Ta	Na	Ne	H
CPLG01	F: CTCACCCTTCGTTGAACCAT R: GGCAGAATTTTGTTTCCAGG	(AT)6	221	56	3	1.094	0.086
CPLG02	F: TCCAACGGAATCCCACTTAC R: CCATAAATCCCGAGATGGAA	(TC)6(T)10	381	58	2	1.117	0.105
CPLG03	F: ACGATGACTTTGGTTTTCGC R: AGAAAGACCCGCCTGTCATA	(TC)7	207	56	2	1.193	0.162
CPLG04	F: GGGAAAAGACCCGTATCCAT R: GGATCGGATCGAATTGAAAA	(GA)6	244	56	2	1.069	0.064
CPLG05	F: TTGTACTGATTGGGGGCTTC R: GGGGTTCTTAAGCTTTTCGATT	(AT)10	190	56	3	1.254	0.203
Mean					**2.400**	**1.146**	**0.124**

Note: Ta = Annealing temperature, Na = Observed number of alleles, Ne= Effective number of alleles, H = Nei's (1973) gene diversity

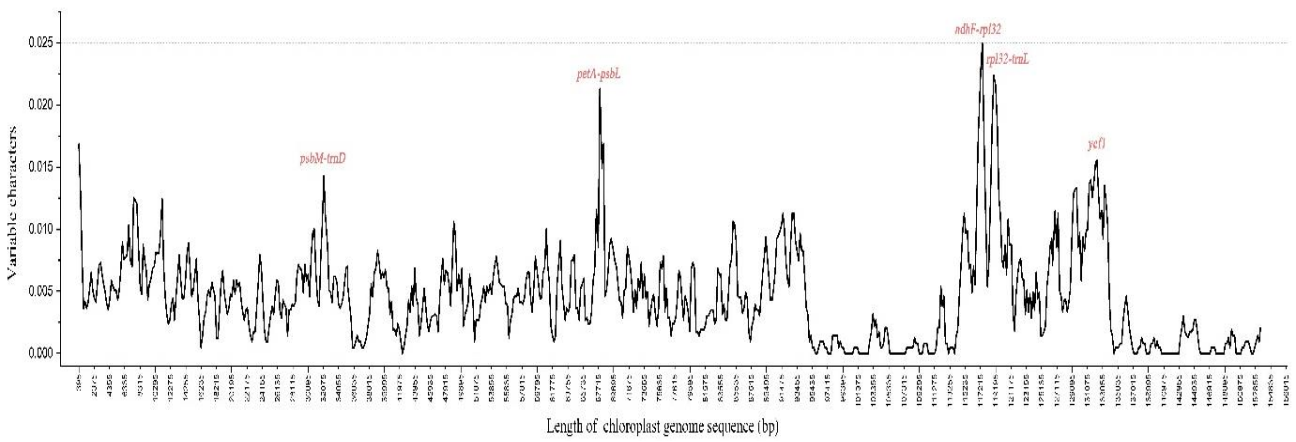

Fig. 4. Sliding window analysis of genetic divergence among the whole plastomes of *L. glauca*, *M. yunnanensis*, *M. balansae*, *P. Americana*, *P. omeiensis*, *P. sheareri*, and *L. glutinosa*.(window length: 600 bp, step size: 200 bp). X-axis: position of the midpoint of a window, Y-axis: nucleotide diversity of each window.

The maximum parsimony (MP) phylogenetic tree, constructed from whole cp sequences of these 8 species and 2 additional species (*Liriodendron tulipifera* and *Calycanthus floridus* var. *glaucus* as outgroup), supported clear lineage distinctions with 100% bootstrap values (Fig. 3). A total of 8,717 variable sites and 648 informative sites were found in ingroup taxa. In detail, *C. micranthum* was sister to the remaining seven species in the Lauraceae family, and *L. glauca* and *Litsea glutinosa* clustered into one group. The genera *Machilus* (*M. yunnanensis* and *M. balansae*) and *Phoebe* (*P. omeiensis* and *P. sheareri*) formed an independent clade, with *Persea americana* as their sister species.

Among the regions, the genetic divergence of the SSC region, the LSC region, and the IR region decreased in order, based on analysis with DnaSP software. Based on sequence divergence of the eight species in Lauraceae, the variable characters of the whole cp genome were from 0 to 0.025 (Fig. 4). Five loci were highly variable (*ndhF-rpl32*, *rpl32-trnL*, *ycf1*, *psbM-trnD*, and *petA-psbL*). Furthermore, the first three of these loci were located in the SSC region and the last two were in the LSC region, while none were located in the IRs regions.

The mVISTA percent identity plot of *L. glauca* and the 6 other Lauraceae chloroplast genomes is shown in Fig. 5. Other than several fragments that showed high levels of polymorphism in intergenic regions, no region was found to be highly variable in all comparisons. Among genes in the coding region, *rpoB*, *psbD*, *psbC*, *cemA*, *petA*, *ndhB*, and rRNA showed very little divergence among sequences. In contrast, substantial divergence was found for the intergenic regions, i.e., *ndhF-rpl32*, *rpl32-trnL*, *ycf1*, *psbM-trnD*, and *petA-psbL*. Furthermore, *trnG-rps16* and *trnG-trnM* showed variation between the genus *Machilus* and *Litsea*, while little difference was observed among the genera *Phoebe*, *Persea*, and *Lindera*.

Discussion

In mainland China, *Lindera glauca* has various practical applications in the medical and biochemical. Most previous studies on this plant focused on its metabolites (Sun *et al*., 2011; Kim *et al*., 2014; Suh *et al*., 2015), transcriptome sequencing (Niu *et al*., 2015), and analysis of fatty acid composition and content (Qi *et al*., 2016). However, the study of its plastid with matrilineal inheritance and phylogenetic relationships based on genome sequences are lacking, both of which could provide insights into apomixis in this species ((Dupont, 2002; Xiong *et al*., 2016; Zhu *et al*., 2016).

The phylogenetic dendrogram based on eight whole cp genomes of Lauraceae demonstrated that *L. glauca* was closely related to *Litsea glutinosa* with *Cinnamomum micranthum* as the sister taxon. The result differs from that of morphological taxonomy which placed the genus *Cinnamomum* (Trib. Cinnamomineae) as more closely related to the genera *Machilus* and Phoebe than to the genera *Lindera* and *Litsea* (Trib. Litseeae) (Tsui *et al*., 1982; Tsui and Werff, 2008). In contrast to the study of Chanderbali *et al*., (2001) which showed that the phylogenetic relationships among the core Laureae (i.e., genera *Lindera*, *Litsea*, *Neolitsea*, *Adenodaphne*, and *Actinodaphne*) based on several fragments (*trnL-trnF*, *psbA-trnH*, and ITS) were not well defined because of low bootstrap values, the phylogenetic dendrogram presented here, based on whole cp sequences, was well supported with 100% bootstrap values. A previous report suggested that *Litsea glutinosa* and *C. micranthum* had a close relationship with 90% bootstrap values (Hinsinger and Strijk, 2017). In our study, we found that *L. glauca* and *Litsea glutinosa* had a closer relationship with a 100% bootstrap value (Fig. 3). The phylogenic relationship of *Persea americana* and the genera *Machilus* and Phoebe is consistent with Song *et al*., (2016). Additionally, the largest divergence in cp SSC region (from 114K to 133K) was observed between *C. micranthum* and the other seven Lauraceae species, consistent with Hinsinger and Strijk (2017). This divergence may result from the presence of a special structure in the chloroplasts of different species such as inversion of some genes (Joseph *et al*., 2014) or another unidentified cause.

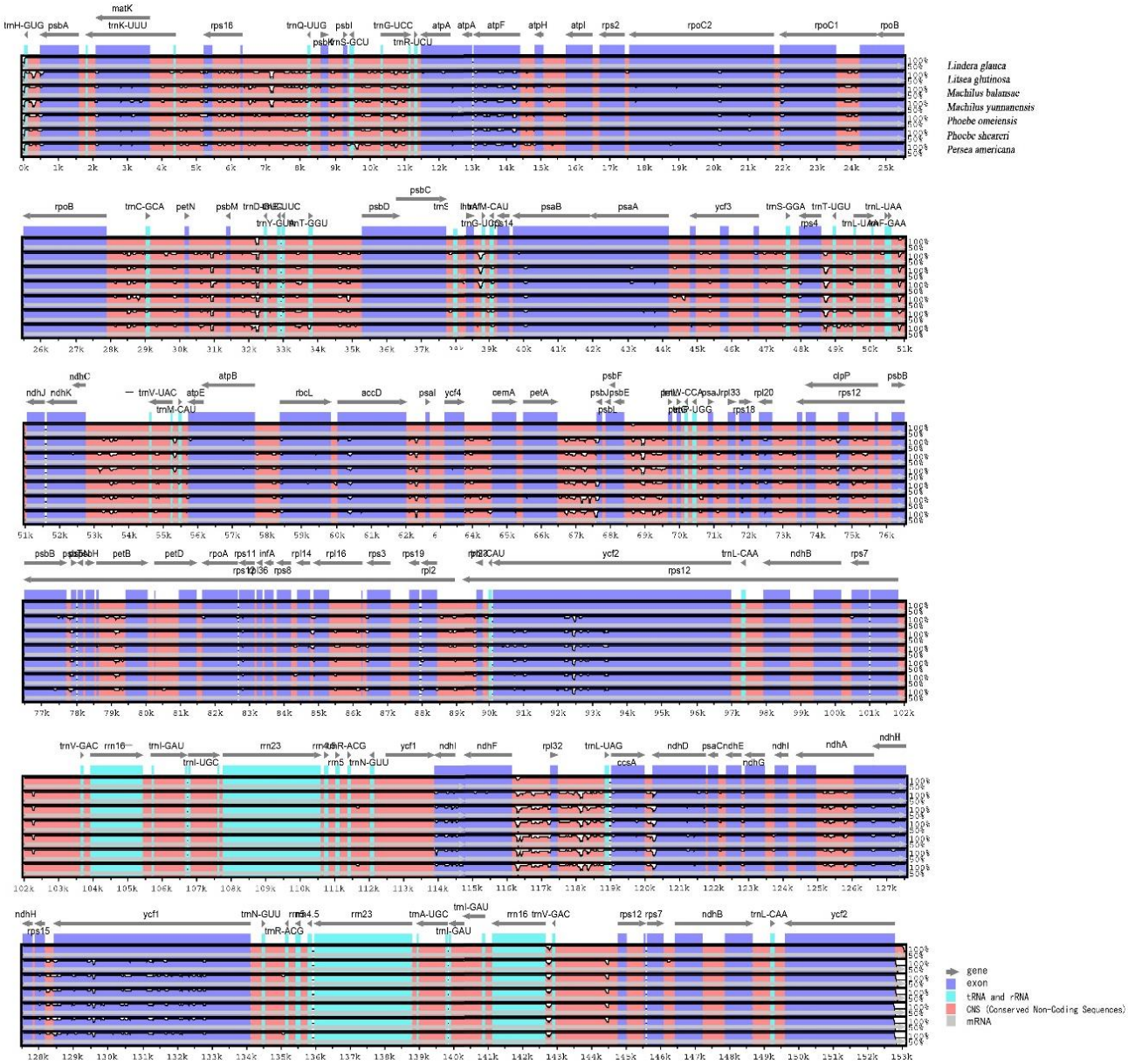

Fig. 5. mVISTA percent identity plot of seven Lauraceae chloroplast genomes with the annotation of *Lindera glauca* as a reference.The top line shows genes in order (transcriptional direction indicated by arrow). The sequence similarity of the aligned regions among the seven species is shown as horizontal bars. The x-axis represents the coordinate in the chloroplast genome, and the Y-scale represents the percent identity between 50 and 100%. Genome regions are color coded as protein-coding (exon, blue), tRNA or rRNA (cyan), and conserved noncoding sequences (CNS, red).

Furthermore, high-throughput sequencing technology, or next-generation sequencing (NGS), has changed the way and method that we view and conduct scientific approaches in basic, biological, medical and clinical study since its inception (Michael, 2010). Compared to cp genomes obtained by Sanger sequencing, plastid genomes produced by NGS technology are easier to obtain and are more effective in resolving difficult phylogenetic relationships, such as the phylogenetic position of *Dipteronia* species (Yang *et al*., 2014), *Machilus* species (Song *et al*., 2015), and *Epimedium* species (Zhang *et al*., 2016). All of these studies, however, used dozens or hundreds of pairs of normal and specific primers to bridge gaps in the plastomes, after NGS. In the present study, more than 8 GB of clean data per primer pair and 28,309,392×2 clean data pair reads were obtained using the Illumina Miseq platform. Using HPC and reference genomes, the fold coverage (mean 118X) of *L. glauca* was so high that no gaps were found. Thus, by using one species as the reference genome from the same genus or family, large sequencing data (or high fold coverage) is effective way to obtain whole cp genome, owing to reducing a series of experimental steps such as long-range PCR.

Lauraceae is an economically and ecologically important family that can be found worldwide. However, genomic resources for the genus *Lindera* are scarce. Some common features in Lauraceae, like conserved patterns of repeats and SSRs, clearly differ among genera. The dendrogram presented here shows the utility of selecting core generic representatives throughout Lauraceae, obtaining their complete plastome sequences, and using these sequences in future analyses to create a comprehensive phylogenetic framework. Phylogenetic analyses based on complete cp genomes, in contrast to a single or several genes, can solve the paraphyletic and polyphyletic complexities hampering our understanding of the family (Chanderbali *et al*., 2001; Li & Li, 2004). This study provides new information which will facilitate future evolutionary and demographic studies in Lauraceae species. In addition, the data could enhance further phylogenetic, genetic variation, and population studies for this important economic plant. Moreover, because the chloroplast genome (matrilineal inheritance) of *L. glauca* is haplotype but the nuclear genome (biparental inheritance) is not, five cpSSRs developed in this study can be combined with 25 EST-SSRs from previous work (Xiong *et al*., 2016) to facilitate studies of genetic diversity, population genetic structure, and genome evolution among populations of this species.

Conclusion

In this study, we firstly constructed and presented the complete cp genome of *L. glauca* that is the first one in the genus *Lindera*. Five divergence hotspots (*padM-trnD*, *petA-psbL*, *ndhF-rpl32*, *rpl32-trnL*, and *ycf1*) and total 67 SSRs (nine non- mononucleotide repeats) were identified across the whole genome sequence. Five polymorphic SSR markers developed here (cpDNA-SSR) were found to be polymorphic in total 90 *L. glauca* individuals, and proved useful in the evaluation of the genetic diversity in this species.

Acknowledgement

We are grateful to Jigongshan National Nature Reserve for partial materials collection. This research was supported by the Chinese Key Technology Research and Development Program of Twelfth Five-Year Plan (No.2013BAD01B06).

References

Bolger, A.M., M. Lohse and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 15: 2114-2120. doi: 10.1093/bioinformatics/btu170

Chanderball, A.S., H. Werrf and S.S. Renner. 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

Chang, C.E. 1976. Lauraceae. In: *Flora of Taiwan* Vol. 2. (Ed.): Liu, T.S. Epoch press, Taipei, pp. 406-468.

Daniell, H., C.S. Lin, M. Yu and W.J. Chang. 2016. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol*., 17: 1-29. doi: 10.1186/s13059-016-1004-2

Ding, B.Z. 1981. Flora of Henan. Henan People's Publishing House Press, Zhengzhou.

Doyle, J.J. and J.L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull*., 19: 11-15.

Dupont, Y.L. 2002. Evolution of apomixis as a strategy of colonization in the dioecious species *Lindera glauca* (Lauraceae). *Popul. Ecol*., 44: 293-297. doi: 10.1007/s101440200033

Frazer, K.A., L. Pachter, A. Poliakov, E.M. Rubin and I. Dubchak. 2004. VISTA: computational tools for comparative genomics. *Nucleic Acids Res*., 32: W273-W279. doi: 10.1093/nar/gkh458

Hinsinger, D.D. and J.S. Strijk. 2017. 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. doi: 10.1016/j.plgene.2016.08.002

Josephfw, J., Z. Michael and C.E. Nancy. 2014. Comparative analysis of complete chloroplast genome sequence and inversion variation in *Lasthenia Burkei* (Masieae, Asteraceae). *Amer. J. Bot*., 101: 722-729. doi:10.3732/ajb.1400049

Katoh, K. and D.M. Standley. 2013. MAFFT multiple sequence alignment Software Version 7: improvements in performance and usability. *Mol. Biol. Evol*., 30: 772-780. doi: 10.1093/molbev/mst010

Kent, W.J. 2002. BLAT—the BLAST-Like Alignment Tool. *Genome Res*., 12: 656-664. doi: 10.1101/gr.229202

Kim, K.H., E. Moon, K. Ha, W.S. Suh, H.H. Kim, S.Y. Kim, S.U. Choi and K.R. Lee. 2014. Bioactive Lignan Constituents from the Twigs of *Lindera glauca*. *Chem. Pharm. Bull*., 62: 1136-1140. doi: 10.1248/cpb.c14-00381

Li, J. and X.W. Li. 2004. Advances in Lauraceae systematic research on the world scale. *Acta Botanica Yunnanica*, 26: 1-11. ([In Chinese])

Librado, P. and J. Rozas. 2009. DNA SP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, 25: 1451-1452. doi: 10.1093/ bioinformatics/btp187

Liu, C., L.C. Shi, Y.J. Zhu, H.M. Chen, J.H. Zhang, X.H. Lin and X.J. Guan. 2012. CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences. *BMC Genom.*, 13: 715. doi: 10.1186/1471-2164-13-715.

Liu, L.D., J.D. Cheng and S.G. Lan. 1982. Chemical composition test and research applications of *Lindera galuca* fruits. *J. Integ. Plant Biol*., 24: 252-258. ([In Chinese])

Lohse, M., O. Dechsel and R. Bock. 2007. Organellar Genome DRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr. Genet*., 52: 267-274. doi: 10.1007/s00294-007-0161-y

Michael, L.M. 2010. Sequencing technologies — the next generation. *Nature Reviews Genetics*, 11: 31-46. doi: 10.1038/nrg2626

Neuhaus, H. and M. Emes. 2000. Nonphotosynthetic metabolism in plastids. *Ann. Rev. Plant Biol*., 51: 111-140. doi: 10.1146/annurev.arplant.51.1.111

Niu, J., X.Y. Hou, C.L. Fang, J.Y. An D.L., Ha, L. Qiu, Y.X. Ju, H.Y. Zhao, W.Z. Du, J. Qi, Z.X. Zhang, G.N. Liu and S.Z. Lin. 2015. Transcriptome analysis of distinct *Lindera glauca* tissues revealed the differences in the UniGenes related to terpenoid biosynthesis. *Gene*, 559: 22-30. doi: 10.1016/j.gene.2015.01.002

Qi, J., B. Xiong, Z.X. Zhang, A.G. Zhen and F.G. Pu. 2016. Diversity analysis on fatty acid composition of *Lindera glauca* fruit from different provenances. *Journal of the Chinese Cereals and Oils Association*, 31: 72-78. ([In Chinese])

Rousset, F. 2008. GENEPOP'007: A complete re-implementation of the GENEPOP software for Windows and Linux. *Mol. Ecol. Resource*, 8: 103-106. doi: 10.1111/j.1471-8286.2007.01931.x

Schattner, P., A.N. Brooks and T.M. Lowe. 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res*., 33: W686-W689. doi: 10.1093/nar/gki366.

Song, Y., W. Dong, B. Liu, C. Xu, X. Yao, J. Gao and R.T. Corlett. 2015. Comparative analysis of complete chloroplast genome sequences of two tropical trees *Machilus yunnanensis* and *Machilus balansae* in the family Lauraceae. *Front. Plant Sci*., 6: 662. doi: 10.3389/fpls.2015.00662

Song, Y., X. Yao, Y.H. Tan, Y. Gan and R.T. Corlett. 2016. Complete chloroplast genome sequence of the avocado: gene organization, comparative analysis, and phylogenetic relationships with other Lauraceae. *Can. J. For. Res*., 46: 1293-1301. doi: 10.1139/cjfr-2016-0199

Suh, W.S., K.H. Kim, H.K. Kim, S.U. Choi and K.R. Lee. 2015. Three New Lignan Derivatives from *Lindera glauca* (Siebold et Zucc.) Blume. *Helvetica Chimica Acta*, 98: 1087-1094. doi: 10.1002/hlca.201500002

Sun, H.L., J.X. Wang, X.Z. Gu and W.Y. Kang. 2011. Analysis of volatile compounds from leaves and fruits of *Lindera glauca*. *Chin. J. Exp. Tradit. Med. Form*., 17: 94-97. ([In Chinese])

Thiel, T., W. Mivhalek, R.K. Varshney and A. Graner. 2003. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor. & App. Genetics*, 10: 411-422. doi: 10.1007/s00122-002-1031-0

Tsui, H.P. and H. Werff. 2008. Lauraceae and Hernandiaceae. In: *Flora of China* Vol. 7, (Ed.): Li, H.W/ Science Press & Missouri Botanical Garden Press, Beijing &St. Louis, pp. 393-394.

Tsui, H.P., Z.D. Xia and J.L. Li. 1982. Lauraceae and Hernandiaceae. In: *Flora of China*. (Ed): H.W. Li. Vol. 31. Science Press, Beijing, pp 393-394.

Wang, R., S. Tang, H. Zhai and H. Duan. 2011. Studies on anti-tumor metastatic constituents from *Lindera glauca*. *China J. Chinese. Mater. Med*., 36: 1032-1036. ([In Chinese])

Wang, W.C. 1972. Higher plants iconography of China. Science Press, Beijing.

Wang, Y., X. Gao, X. Yu, S. Cheng and L. Kong. 1994. Study on the resource and its utilizations of *Lindera glauca* in China. *Henan Science*, 12: 331-334. ([In Chinese])

Weitemier, K., S.C.K. Straub, R.C. Cronn, M. Fishbein, R. Schmickl, A. Mcdonnell and A. Liston. 2014. Hyb-Seq: Combinging target enrichment and genome skimming for plant phylogenomics. *App. Plant Sci*., 2: 1400042. doi: 10.3732/apps.1400042

Wyman, S.K., R.K. Jansen and J.L. Boore. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics*, 20: 3252-3255. doi: 10.1093/bioinformatics/bth352

Xiong, B., S.B. Dong, J. Qi, L.M. Zhang, D.L. Ha, Y.X. Ju and Z.X. Zhang. 2016. Development and characterization of microsatellite loci for *Lindera glauca* (Lauraceae). *Appl. Plant Sci*., 4: 1600088. doi: 10.3732/apps.1600088

Zhang, Y., L. Du, A. Liu, J. Chen, L. Wu, W. Hu, W. Zhang, K. Kim, S.C. Lee, T.J. Yang and Y. Wang. 2016. The complete chloroplast genome sequences of five *Epimedium* species: lights into phylogenetic and taxonomic analyses. *Front Plant Sci*., 7: 306. doi: 10.3389/fpls.2016.00306

Zheng, W.Q. 1983. Dendrology of China. Chinese Forestry Publishing House Press, Beijing.

Zhu, S.S., Y.Q. Ding, Z.Y. Yap and Y.X. Qiu. 2016. De novo assembly and characterization of the floral transcriptome of an economically important tree species, *Lindera glauca* (Lauraceae), including the development of EST-SSR markers for population genetics. *Mol. Biol. Rep*., 43: 1243-1250. doi: 10.1007/s11033-016-4056-1

(Received for publication 28 November 2017)