Complete chloroplast genome of Castanopsis sclerophylla (Lindl.) Schott: Genome structure and comparative and phylogenetic analysis

Xuemin Ye, Dongnan Hu, Yangping Guo, Rongxi Sun

Jiangxi Provincial Key Laboratory of Silviculture, College of Forestry, Jiangxi Agricultural University, Nanchang, China

* rongxisun@163.com

Abstract

Castanopsis sclerophylla (Lindl.) Schott is an important species of evergreen broad-leaved tree in subtropical areas and has high ecological and economic value. However, there are few studies on its chloroplast genome. In this study, the complete chloroplast genome sequence of C. sclerophylla was determined using the Illumina HiSeq 2500 platform. The complete chloroplast genome of C. sclerophylla is 160,497 bp long, including a pair of inverted repeat (IR) regions (25,675 bp) separated by a large single-copy (LSC) region of 90,255 bp and a small single-copy (SSC) region of 18,892 bp. The overall GC content of the chloroplast genome is 36.82%. A total of 131 genes were found; of these, 111 genes are unique and annotated, including 79 protein-coding genes, 27 transfer RNA genes (tRNAs), and four ribosomal RNA genes (rRNAs). Twenty-one genes were found to be duplicated in the IR regions. Comparative analysis indicated that IR contraction might be the reason for the smaller chloroplast genome of C. sclerophylla compared to three congeneric species.

Sequence analysis indicated that the LSC and SSC regions are more divergent than IR regions within Castanopsis; furthermore, greater divergence was found in noncoding regions than in coding regions. The maximum likelihood phylogenetic analysis showed that four species of the genus Castanopsis form a monophyletic clade and that C. sclerophylla is closely related to Castanopsis hainanensis with strong bootstrap values. These results not only provide a basic understanding of Castanopsis chloroplast genomes, but also illuminate Castanopsis species evolution within the Fagaceae family. Furthermore, these findings will be valuable for future studies of genetic diversity and enhance our understanding of the phylogenetic evolution of Castanopsis.

Introduction

Castanopsis (Lindl.) Schott. is a monoecious, broad-leaved tree of the genus Castanopsis belonging to the Fagaceae family. The genus contains approximately 120 known species, of which 58 are native and 30 are endemic to China. However, C. sclerophylla is widely distributed in East and South Asia, and the tree has been introduced to North America[1, 2]. In
China, *C. sclerophylla* is a canopy tree widely distributed in subtropical evergreen forests[3]. Its fruit and wood are valuable, and it is regarded as a landscape and ornamental tree because of its glossy evergreen leaves and abundant white flowers[4]. In Jiangxi Province, the fruit of this tree has been used to make special foods such as sheet jelly, bean curd, and bean vermicelli[5]. Previous studies have mainly focused on the natural regeneration[6], biomass[7], morphology[8], chemotaxonomy[9], and genetic diversity[1, 10] of this species. However, because of the increasing economic value of the tree, natural trees are severely destroyed by humans, and their number is decreasing due to the slow growth rate of this species. As a result, the distribution of *C. sclerophylla* is severely fragmented and constantly threatened, requiring urgent conservation and restoration[11]. Phylogenetic and population genomics data are vital for developing effective conservation and management strategies. With the rapid development of next-generation sequencing technology such as Illumina sequencing, chloroplast genome assembly has become less expensive and easier than it was with the Sanger method. Additionally, comparative analysis of the complete chloroplast genome among closely related species has proven to be a valid and effective method for the studying evolutionary history, species conservation, and phylogenetic relationships[12–15].

Chloroplasts are essential organelles in plant cells that play very important roles in photosynthesis, carbon fixation, and synthesis of pigments, starch, fatty acids, and amino acids[16, 17]. The chloroplast genome of angiosperms typically consists of highly conserved circular DNA ranging from 120 to 180 kb in length with a typical quadripartite structure including a large single-copy (LSC) region, a small single-copy (SSC) region, and a pair of inverted repeats (IRs)[18]. The chloroplast genome encodes approximately 110 to 130 genes, including up to 80 unique protein-coding genes, four ribosomal RNAs (rRNAs), and approximately 30 transfer RNAs (tRNAs)[19]. In recent years, many complete chloroplast genome sequences of higher plants have been reported and used to study population structure and phylogenetic relationships.

In this study, we sequenced the *C. sclerophylla* chloroplast genome using Illumina technology. This is the first comprehensive analysis of the *C. sclerophylla* chloroplast genome in conjunction with the previously published whole-chloroplast genome sequences of three congeneric species. In addition, we used 22 complete chloroplast genome sequences from GenBank to analyze phylogenetic relationships and infer the phylogenetic position of *C. sclerophylla* but also enhance our understanding of *Castanopsis* species evolution within the family Fagaceae. Our data will contribute to our understanding of the genetic resources and evolution of *C. sclerophylla* based on the diversity in its chloroplast genome and also facilitate the exploration, utilization and application of conservation genetics of this species.

**Materials and methods**

**Plant material, DNA extraction and sequencing**

Fresh young leaves of *C. sclerophylla* were collected from the Jiangxi Agricultural University Arboretum in Nanchang, China (28˚45'N, 115˚49'E). Total genomic DNA was extracted using a Plant Genomic DNA Kit (TIANGEN, Beijing, China). Agarose gel electrophoresis and a microplate spectrophotometer (Molecular Device, Sunnyvale, CA, USA) were used to measure DNA quality and concentration, respectively. Shotgun libraries with an average insert size of 350 bp were constructed using pure DNA and sequenced from 150 bp paired-end read lengths with the Illumina HiSeq 2500 platform (Illumina, San Diego, California, USA). To obtain high-quality clean data, raw reads were filtered by removing the connector sequence and low-quality reads using NGS QC Toolkit_v.2.3.3[20].
Chloroplast genome assembly and annotation

The high-throughput raw reads were trimmed by FastQC. Next, the trimmed paired-end reads and references (C. hainanensis, C. echinocarpa, and C. concinna) were used to extract chloroplast-like reads, which were assembled by NOVOPlasty[21]. NOVOPlasty assembled partial reads and stretched them as far as possible until a circular genome is formed. A high-quality complete chloroplast genome was ultimately obtained. The assembled genome was annotated using CpGAVAS[22]. BLAST and Dual Organellar Genome Annotator (DOGMA) were applied to check the annotation results[23]. tRNAs were identified by tRNAscan-SE[24]. Circular gene maps of C. sclerophylla were drawn with the OGDRAW v1.2 program[25]. To analyze variation in synonymous codon usage, MEGA7 was used to compute relative synonymous codon usage (RSCU) values, codon usage, and GC content[26]. RSCU represent the ratio of the observed frequency of a codon to the expected frequency and is a good indicator of codon usage bias[27]. When the RSCU value is less than 1, synonymous codons are used less frequently than expected; otherwise, the value is greater than 1[28].

Comparative analysis and phylogenetic analysis

MUMmer[29] was employed for paired sequence alignment of the chloroplast genomes. Sequence divergence was computed pairwise distance between each two species adopting protein-coding sequences using MEGA 5.0 with Kimura 2-parameter model[30]. The mVISTA[31] program was used to compare the complete chloroplast genome of C. sclerophylla to three other published chloroplast genomes of the genus Castanopsis, i.e., Castanopsis concinna voucher Strijk_1489 (KT793041.1), C. echinocarpa (KJ001129.1), and C. hainanensis (MG383644.1), in Shuffle-LAGAN mode, adopting the annotation of C. concinna as a reference. In total, 20 chloroplast genomes belonging to Fagaceae were analyzed in this study, including the newly generated chloroplast genome C. sclerophylla and all of the published chloroplast genomes (data present in NCBI GenBank on 31.12.2018). The other 19 chloroplast (cp) genomes species are from the species Castanea henryi voucher CHEN20160703 (KX954615.1), Castanea mollissima (HQ336406.1), Castanopsis concinna voucher Strijk_1489 (KT793041.1), Castanopsis echinocarpa (KJ001129.1), Castanopsis hainanensis (MG383644.1), Fagus engleriana (KX852398.1), Lithocarpus balansae (KP299291.1), Quercus aliena (KP301144.1), Quercus aquifolioides (KP340971.1), Quercus baronii (KT963087.1), Quercus dolicholepis (KU240010.1), Quercus glauca (KX852399.1), Quercus rubra (JX970937.1), Quercus sichourensis (MF878253.1), Quercus spinosa (KM841421.1), Quercus taroensis (MF135621.1), Quercus tungmaiensis (MF593893.1), Quercus variabilis (KU240009.1), and Trigonobalanus doichangensis (KF990556.1). Phylogenies were constructed by maximum likelihood (ML) using the 20 cp genomes of the Fagaceae species in GenBank. Corylus fargesii (KX822767.2) and Eucalyptus umbra (KC180778.1) were used as outgroups. Sequences were initially aligned using MAFFFT[32], followed by visualization and manual adjustment of multiple sequence alignment in BioEdit[33]. The maximum likelihood (ML) analysis was conducted using RAxML web servers[34]. For ML analyses, general time reversible (GTR)+G model was used in as suggested by 1,000 bootstrap replicates with the default tree search algorithm of hill-climbing[30, 35, 36].

Results and discussion

Characteristics of C. sclerophylla cpDNA

A total of 65 million paired-end reads were obtained, and 10.44 Gb of high-quality clean data with a mean Q30 higher than 88.28% were obtained by removing low-quality reads and
connector sequence. The remaining high-quality reads were utilized in the further assembly. The complete chloroplast genome sequence of *C. sclerophylla* is 160,497 bp in length; it has been deposited in GenBank under accession number MK387847. The genome has a typical quadripartite structure including a pair of IR (IRa and IRb) regions of 25,675 bp that are separated by an LSC region of 90,255 bp and an SSC region of 18,892 bp (Fig 1, Table 1). The overall GC content of the chloroplast genome is 36.82%, which is similar to that of other Fagaceae species[37–39]. However, a few differences in GC content were found among the chloroplast genomes. The GC contents of the LSC, SSC, and IR regions are 34.65%, 30.94%, and 42.78%, respectively.

Fig 1. Chloroplast genome annotation map for *C. sclerophylla*. Genes inside the circle are transcribed in a clockwise direction; genes outside are transcribed in a counterclockwise direction. Different colors represent different functional genes. The darker gray and lighter gray in the inner circle show the GC and AT contents of the chloroplast genome, respectively.

https://doi.org/10.1371/journal.pone.0212325.g001
respective (Table 2). The GC content is highest in IR regions (42.78%), likely due to the presence of four duplicated ribosomal RNA genes in this region, a pattern also found in the chloroplast genome of *C. hainanensis* [38]. The overall GC content is an important species indicator [40].

A total of 131 genes were found in the *C. sclerophylla* chloroplast genome, including 86 protein-coding genes, 37 tRNA genes, and 8 rRNA genes (Fig 1, Table 1). Of these 131 genes, 110 genes are unique and annotated and divided into three categories: 79 protein-coding genes, 27 tRNA genes, and four rRNA genes (Table 3). In addition, 21 functional genes (seven protein-coding genes, four rRNA genes, and 10 tRNA genes) are duplicated in the IR regions (Fig 1). The LSC region comprises 62 protein-coding genes and 22 tRNA genes, whereas the SSC region comprises 11 protein-coding and one tRNA gene (S1 Table). There are 14 intron-containing genes, including eight protein-coding genes and six tRNA genes. Twelve genes contain one intron, and *clpP* and *ycf3* have two introns. *trnK-UUU* contains the longest intron (2,511 bp); and *trnL-UAA* the shortest (485 bp) (Table 4). A similar phenomenon is also present in *Quercus acutissima* [41]. *ycf3* gene expression results in stable accumulation of photosystem I complexes [42]. Therefore, we herein focus on the *ycf3* intron gain in *C. sclerophylla*, which may be helpful for further study of the photosynthesis mechanism.

**Codon usage analysis**

Relative synonymous codon usage frequency (RSCU) values were computed for the *C. sclerophylla* chloroplast genome using protein-coding sequences (S2 Table), as codon usage plays a vital role in shaping chloroplast genome evolution [43]. In total, 23,131 codons are present. Leucine (10.61%) is the most commonly encoded amino acid, with 2,454 codons, followed by isoleucine (8.85%) with 2,048 codons; cysteine (1.13%) is the least commonly encoded amino acid, with 262 codons (Fig 2). Similar ratios for amino acids were previously reported for chloroplast genomes [44, 45]. Moreover, methionine and tryptophan are encoded by only one codon, indicating no codon bias for these two amino acids (RSCU = 1). Nearly all of the codons ending with A and U had RSCU values of more than one (RSCU > 1), whereas the codons ending with

| Region | A (%) | T (%) | C (%) | G (%) | A+T (%) | G+C (%) |
|--------|------|------|------|------|--------|--------|
| LSC    | 31.94| 33.4 | 17.74| 16.91| 65.34  | 34.65  |
| SSC    | 34.4 | 34.66| 16.29| 14.65| 69.06  | 30.94  |
| IR     | 28.61| 28.61| 21.39| 21.39| 57.22  | 42.78  |
| Total  | 31.65| 32.23| 18.47| 17.65| 63.18  | 36.82  |

https://doi.org/10.1371/journal.pone.0212325.t002
C and G had RSCU values of less than one. The AU contents for the first, second, and third codon positions were calculated to be 54.07%, 56.29% and 70.20%, respectively. The results of high AU content at the third codon position were similar to reports for other plants[46].

### Comparative analysis of genomic structure

Three complete chloroplast genomes within the *Castanopsis* genus (*C. hainanensis*, *C. echinocarpa*, and *C. concinna*) were selected for comparison with that of *C. sclerophylla*. *C. sclerophylla* has the smallest chloroplast genome (160,497 bp); *C. echinocarpa* has the largest

### Table 3. List of genes annotated in the sequenced *C. sclerophylla* chloroplast genome.

| Category        | Function          | Genes                      |
|-----------------|-------------------|----------------------------|
| Photosynthesis  | Photosystem I     | *psaA*, *psaB*, *psaC*, *psaL*, *psaF* |
|                 | Photosystem II    | *psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbH*, *psbf*, *psbK*, *psbL*, *psbM*, *psbN*, *psbT* |
|                 | Cytochrome b/f complex | *petA*, *petB*, *petD*, *petG*, *petL*, *petN* |
|                 | ATP synthase      | *atpA*, *atpB*, *atpE*, *atpF*, *atpH*, *atpI* |
|                 | NADH dehydrogenase | *ndhA*°, *ndhB*°(X2), *ndhC*, *ndhD*, *ndhE*, *ndhF*, *ndhG*, *ndhH*, *ndhI*, *ndhJ*, *ndhK* |
|                 | Rubisco large subunit | *rbcL* |
| Self-replication| RNA polymerase    | *rpoA*, *rpoB*, *rpoC1*°, *rpoC2* |
|                 | Ribosomal proteins (LSU) | *rpl14*, *rpl16*, *rpl2*°(X2), *rpl20*, *rpl22*, *rpl23*(X2), *rpl32*, *rpl33*, *rpl36* |
|                 | Ribosomal proteins (SSU) | *rps11*, *rps12*°, *rps14*, *rps15*, *rps16*°, *rps18*, *rps19*, *rps2*, *rps3*, *rps4*, *rps7*(X2), *rps8* |
|                 | Transfer RNAs     | *trnA-UGC*°(X2), *trnC-ACA*°, *trnC-GCA*°, *trnD-GUC*°, *trnE-UCC*°(X3), *trnF-GAA*°, *trnG-GCC*°, *trnH-GUG*°, *trnK-UUU*°, *trnL-CAAX*(X2), *trnL-CAAX*, *trnL-UAA*°, *trnL-UAG*°, *trnM-CAUX*(X4), *trnN-GUX*(X2), *trnP-UUG*°, *trnQ-UUG*°, *trnR-ACGX*(X2), *trnR-UCUX*, *trns-GCU*°, *trns-GGA*°, *trns-UAG*°, *trns-UGU*°, *trnV-GAC*(X2), *trnV-GAC*, *trnW-CA*°, *trnY-GUA*° |
|                 | Ribosomal RNAs    | *rrn16*, *rrn23*, *rrn4.5*, *rrn5* |
| Others          | Hypothetical chloroplast reading frames | *ycf1*(X2), *ycf2*(X2), *ycf3*°°, *ycf4* |
|                 | Other genes       | *ccsA*, *cemA*, *clpP*°°, *infA*, *lhbA*, *matK* |

(*) number of gene copies in the IR.

° Genes containing one intron

°° genes containing two introns.

https://doi.org/10.1371/journal.pone.0212325.t003

### Table 4. Lengths of exons and introns for genes with introns in the *C. sclerophylla* chloroplast genome.

| Gene    | Location | Exon I (bp) | Intron I (bp) | Exon II (bp) | Intron II (bp) | Exon III (bp) |
|---------|----------|-------------|---------------|--------------|---------------|---------------|
| clpP    | LSC      | 70          | 851           | 291          | 654           | 227           |
| trnK-UUU| LSC      | 36          | 2511          | 34           |               |               |
| rpoC1   | LSC      | 429         | 838           | 1618         |               |               |
| trnC-ACA| LSC      | 37          | 610           | 55           |               |               |
| ndhB    | IRA      | 776         | 681           | 755          |               |               |
| ndhA    | SSC      | 550         | 1049          | 540          |               |               |
| rpl2    | IRB      | 390         | 685           | 433          |               |               |
| trnA-UGC| IRB      | 36          | 801           | 35           |               |               |
| trnL-UAA| LSC      | 34          | 485           | 49           |               |               |
| trnE-UUC| LSC      | 31          | 956           | 39           |               |               |
| trnT-CGU| LSC      | 34          | 720           | 42           |               |               |
| rps16   | LSC      | 41          | 903           | 227          |               |               |
| ycf3    | LSC      | 125         | 727           | 225          | 768           | 154           |
| atpF    | LSC      | 144         | 789           | 409          |               |               |

https://doi.org/10.1371/journal.pone.0212325.t004
chloroplast genome (160,647 bp) with the smallest IR region (25,629 bp). Additionally, the
lengths of LSC regions varied among these four species, from 90,255 bp in C. sclerophylla to
90,394 bp in C. echinocarpa (Table 1). The different lengths of the LSC region are the main rea-
son for the difference in sequence length among the four species, consistent with the results
for the genus Oryza[44]. To investigate levels of genome divergence, the program mVISTA
was used to plot sequence identity for the chloroplast genomes of the four species using C. con-
cinna as a reference (Fig 3). The results of sequence analysis revealed the LSC and SSC regions
to be more divergent than the IR regions among the four Castanopsis genomes; furthermore,
greater divergence was found in noncoding regions than in coding regions. Coding regions
with significant variation in the four chloroplast genomes included ndhF, ndhG, and ycf1,
which are all located in SSC regions. Nonetheless, the most divergent regions were observed in intergenic regions, including *trnK-rps16*, *trnS-trnT*, *atpA-atpF*, *trnC-petN*, *trnT-psbD*, *IhbA-trnG*, *ycf3-trnS*, *rps4-trnT*, *trnT-trnL*, *atpB-rbcL*, *petA-psbJ*, *psbE-petL*, *rpl16-rpl3*, and *ndhF-rpl32*.

The expansion and contraction of IR regions at the borders are the major reason for chloroplast genome size variation and play vital roles in evolution[47–49]. A detailed comparison of four junctions (JLA, JSB, JSA, and JLA) between the two single-copy regions (LSC and SSC) and the two IRs (IRa and IRb) was performed for *C. sclerophylla*, *C. hainanensis*, *C. echinocarpa* and *C. concinna* by analyzing exact IR border positions and adjacent genes (Fig 4). Overall IR regions are relatively conserved in the genus *Castanopsis*, and this result agrees with reports for the genus *Quercus*[41]. The *rps19* gene is located between the junction of the LSC and IRb regions in *C. concinna*. However, in the *C. sclerophylla*, *C. hainanensis*, and *C. echinocarpa* chloroplast genomes, the *rps19* gene is located in the LSC region and is 11 bp, 11 bp, and
10 bp from the border of the LSC region, respectively. Some studies have indicated that ycf1 is required for plant viability and encodes Tic214, which is a vital component of the TIC complex in Arabidopsis[50, 51]. The ycf1 gene crosses the SSC/IRb and SSC/IRa regions. The SSC/IRb junction is located in the ycf1 region in the chloroplast genome of all four Castanopsis species and extends into the SSC region by different lengths depending on the genome (C. sclerophylla, 21 bp; C. hainanensis, 24 bp; C. echinocarpa, 59 bp; and C. concinna, 59 bp); the IRb region includes 1,131, 1,157, 1,107, and 1,092 bp of the ycf1 gene. The SSC/IRa junction also extends into the SSC region by different lengths depending on the genome (C. sclerophylla, 4,581 bp; C. hainanensis, 4,568 bp; C. echinocarpa, 4,608 bp; and C. concinna, 4,581 bp); the IRa region includes 1,092, 1,114, 1,107, and 1,092 bp of the ycf1 gene.

**Phylogenetic analysis**

Phylogenetic analysis was performed by ML based on the 22 aligned sequences of chloroplast genomes (Fig 5). C. fargesii and E. umbra were used as outgroups. The ML-based phylogenetic analysis showed that these four species of the genus Castanopsis form a monophyletic clade and that C. sclerophylla is closely related to C. hainanensis with strong bootstrap values. The ML tree indicated that Castanopsis is closely related to Castanea. Surprisingly, Quercus species do not form a clade, and Quercus is not divided into two clusters containing either evergreen or deciduous tree species. The phylogenetic status of these genera is consistent with a previous report[41, 52, 53]. The relatively high variation in Quercus may be related to the widely distributed range which need to local adaptation to different environments. Notably, F. engleriana is the first to diverge in Fagaceae, which indicates the relatively high genetic divergence between F. engleriana and others, followed by T. doichangensis, which indicates that they are early diverging taxa in Fagaceae[54]. Moreover, the same topology results of genus Fagus was confirmed by the research based on nuclear marker[55].

Little is known to date about the chloroplast genome of Castanopsis, and only three chloroplast genome sequences of Castanopsis species can be found in GenBank, which has greatly
hampered the study of the phylogenetic relationships of this genus. Therefore, more research on the complete chloroplast genomes of Castanopsis species needs to be conducted in the future.

**Conclusions**

*C. sclerophylla* is an important evergreen broad-leaved species in the Castanopsis genus of the Fagaceae family. In this study, the complete chloroplast genome sequence of *C. sclerophylla*
was determined using the Illumina HiSeq 2500 platform. The *C. sclerophylla* chloroplast genome exhibits a typical quadripartite and circular structure similar to that of the chloroplast genome of three congeneric species. Compared to the chloroplast genomes of the three other *Castanopsis* species, that of *C. sclerophylla* is the smallest (160,497 bp). In the ML phylogenetic tree, the phylogenetic relationships among 22 angiosperms strongly support the known classification of *C. sclerophylla*, and ML analysis showed that these four *Castanopsis* species form a monophyletic clade and that *C. sclerophylla* is closely related to *C. hainanensis* with strong bootstrap values. In addition, *Castanopsis* is closely related to *Castanea*. The genus *Castanopsis* contains approximately 120 known species, nearly half of which are native to China. Indeed, China has a large amount of *Castanopsis* germplasm resources, and the availability of chloroplast genomes provides a powerful genetic resource for phylogenetic analysis and biological study. Therefore, further research of the complete chloroplast genome of the genus *Castanopsis* is necessary in the future. The data will contribute to the development of genetic resources and the identification of evolutionary relationships and also facilitate the exploration, utilization and application of conservation genetics for the genus.

**Supporting information**

S1 Table. The number of genes in the *C. sclerophylla* chloroplast genome. (DOCX)

S2 Table. Codon–anticodon recognition pattern and codon usage for the *C. sclerophylla* chloroplast genome. (DOCX)

**Author Contributions**

Conceptualization: Xuemin Ye, Rongxi Sun.

Funding acquisition: Rongxi Sun.

Methodology: Xuemin Ye.

Software: Xuemin Ye, Dongnan Hu, Yangping Guo.

Supervision: Rongxi Sun.

Writing – original draft: Xuemin Ye.

Writing – review & editing: Dongnan Hu, Yangping Guo, Rongxi Sun.

**References**

1. Shi YS, Zhang J, Jiang K, Cui MY, Li YY. Development and characterization of polymorphic microsatellite markers in *Castanopsis sclerophylla* (Fagaceae). American Journal of Botany. 2011; 98(2): 19–21.

2. Conden PJ. Propagation of *Castanopsis sclerophylla* and *Lindera umbellata* by Stem Cuttings and Nitrogen Nutrition of Containerized *Ternstroemia gymnanthera*. Thesis, North Carolina State University, 2002.

3. Wang R, Compton SG, Shi YS, Chen XY. Fragmentation reduces regional-scale spatial genetic structure in a wind-pollinated tree because genetic barriers are removed. Ecology and evolution. 2012; 2(9): 2250–2261. [https://doi.org/10.1002/ece3.344 PMID: 23139883].

4. Tu ZC, Liu GX, Dou YX, Liu CM, Lin DR. Study on extraction and purification total flavones from the fruits of *Castanopsis sclerophylla*. Food Res Dev. 2008; 1: 33–36.

5. Xiao QH, Zong CT, Xue CZ, Hui W, Hui X, Qiu TZ, et al. Isolation and Physicochemical Properties of Hardleaf Oatchestnut (*Castanopsis sclerophylla*) Starch. Applied Mechanics & Materials. 2012; 140(2): 360–368.
6. Xin Z, Xu G, Shen D, Gu Y, Hui G, Luo X, et al. Maintenance and natural regeneration of *Castanopsis sclerophylla* populations on islands of Qiandao Lake Region, China. Acta Ecologica Sinica. 2007; 27 (2): 424–430.

7. Tang X, Fehrmann L, Guan F, Forrester DI, Guisasola R, Pérez-Cruzado C, et al. A generalized algebraic difference approach allows an improved estimation of aboveground biomass dynamics of *Cunninghamia lanceolata* and *Castanopsis sclerophylla* forests. Annals of Forest Science. 2017; 74(1): 12.

8. Yue HF, Shao WH, Jing ZH, Lu P, Huang L, Jiang JM. Geographic variation of seed characters of *Castanopsis sclerophylla*. Forest Research. 2010; 23(3):453–456.

9. Huang YL, Tanaka T, Matsuo Y, Kouno I, Li DP, Nonaka GI. Two new phenolic glucosides and an ellagitannin from the leaves of *Castanopsis sclerophylla*. Phytochemistry. 2012; 74(1): 12.

10. Ye LJ, Wang J, Sun P, Dong SP, Zhang ZY. The Transfer ability of Nuclear Microsatellite Markers in *Castanopsis* Species to *Castanopsis tibetana* (Fagaceae). Plant Diversity & Resources. 2014; 36 (04): 443–448.

11. Moore MJ, Soltis PS, Bell CD, J Gordon B, Soltis DE. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. Proceedings of the National Academy of Sciences of the United States of America. 2010; 107(10): 4623–4628. https://doi.org/10.1073/pnas.0907801107 PMID: 20176954

12. Cecile A, J Gordon B, McMahon MM, Sanderson MJ. Covarion structure in plastid genome evolution: a new statistical test. Molecular Biology & Evolution. 2005; 22(4): 914–924.

13. Raman G, Choi KS, Park S. Phylogenetic Relationships of the Fern *Cyrtomium falcatum* (Dryopteridaceae) from Dokdo Island Based on Chloroplast Genome Sequencing. Genes. 2016; 7(12): 15.

14. Park I, Kim WJ, Yeo SM, Choi G, Kang YM, Piao R, et al. The Complete Chloroplast Genome Sequences of *Fritillaria ussuriensis* Maxim. and *Fritillaria cirrhosa* D. Don, and Comparative Analysis with Other *Fritillaria* Species. Molecules. 2017; 22(6): 982.

15. Liu C, Shi L, Zhu Y, Chen H, Zhang J, Lin X, et al. CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences. Bmc Genomics. 2012; 13(1): 715.

16. Wyman SK, Jansen RK, Boore JL. Automatic annotation of organellar genomes with DOGMA. Bioinformatics. 2004; 20(17): 3252–3255. https://doi.org/10.1093/bioinformatics/bth352 PMID: 15180927

17. Lohse M, Drechsel O, Bock R. OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. Current Genetics. 2007; 52 (5–6): 267–274. https://doi.org/10.1007/s00294-007-0161-y PMID: 17957369

18. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Molecular Biology & Evolution. 2016; 33(7): 1870.

19. Sharp PM, Li WH. An evolutionary perspective on synonymous codon usage in unicellular organisms. Journal of molecular evolution. 1986; 24(1–2): 28–38. Epub 1986/01/01. PMID: 3104616.
28. Gupta SK, Bhattacharyya TK, Ghosh TC. Synonymous codon usage in Lactococcus lactis: mutational bias versus translational selection. Journal of biomolecular structure & dynamics. 2004; 21(4): 527–536. Epub 2003/12/25. https://doi.org/10.1080/07391102.2004.10506946 PMID: 14692797.

29. Kurtz S, Philippe A, Delichar AL, Smoot M, Shumway M, Antonescu C, et al. Versatile and open software for comparing large genomes. Genome Biology. 2004; 5(2): R12. https://doi.org/10.1186/gb-2004-5-2-r12 PMID: 14759262.

30. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. Journal of molecular evolution. 1980; 16(2): 111–120. Epub 1980/12/01. PMID: 7463489.

31. Mayor C, Brudno M, Schwartz JR, Poliakov A, Rubin EM, Frazer KA, et al. VISTA: visualizing global DNA sequence alignments of arbitrary length. Bioinformatics. 2000; 16(11): 1046–1047. https://doi.org/10.1093/bioinformatics/16.11.1046 PMID: 11159318.

32. Kazukata K, Kei-Ichi K, Hiroyuki T, Takashi M. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Research. 2005; 33(2): 511–518. https://doi.org/10.1093/nar/gki198 PMID: 15661851.

33. Hall TA. A User-Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/NT. Nucleic Acids Symposium Series. 1999; (41): 95–98.

34. Alexandros S, Paul H, Jacques R. A rapid bootstrap algorithm for the RAxML Web servers. Systematic Biology. 2017; 57(5): 758.

35. Yang Y, Zhu J, Feng L, Zhou T, Bai G, Yang J, et al. Plastid Genome Comparative and Phylogenetic Analyses of the Key Genera in Fagaceae: Highlighting the Effect of Codon Composition Bias in Phylogenetic Inference. Frontiers in plant science. 2018; 9: 82. Epub 2018/02/17. https://doi.org/10.3389/fpls.2018.00089 PMID: 29449857.

36. Dong W, Xu C, Li W, Xie X, Lu Y, Liu Y, et al. Phylogenetic Resolution in Juglandes Based on Complete Chloroplast Genomes and Nuclear DNA Sequences. Frontiers in plant science. 2017; 8: 1148. Epub 2017/07/18. https://doi.org/10.3389/fpls.2017.01148 PMID: 28713409.

37. Ye XM, Guo YP, Lei XG, Sun RX. The complete chloroplast genome of Castanopsis Fargesii Franch. (Fagaceae). Mitochondrial DNA Part B. 2019; 4(1): 1656–1657. https://doi.org/10.1089/dna.1992.11.269.

38. Chen X, Yang J, Yang Y, Zhang X, Du X, Zhao G. Characterization of the complete plastid genome of Castanopsis hainanensis. Conservation Genetics Resources. 2018; 10(4):825–828.

39. Cheng L, Huang W, Lan Y, Cao Q, Su S, Zhou Z, et al. The complete plastid genome sequence of the wild Chinese chestnut (Castanea mollissima). Conservation Genetics Resources. 2017; (3): 1–4.

40. Shen X, Wu M, Liao B, Liu Z, Bai R, Xiao S, et al. Complete chloroplast genome sequence and phylogenetic analyses of the Medicinal plant Artemisia annua. Molecules. 2017; 22(8): 1330.

41. Xuan L, Li Y, Zhang Y, Li M, Fang Y. Complete chloroplast genome sequence and phylogenetic analysis of Quercus acutissima. International Journal of Molecular Sciences. 2018; 19(8): 2443.

42. Boudreau E, Takahashi Y, Lemieux C, Turmel M, Rochaix JD. The chloroplast ycf3 and ycf4 open reading frames of Chlamydomonas reinhardtii are required for the accumulation of the photosystem I complex. Embo Journal. 2014; 16(20): 6095–6104.

43. Li B, Lin F, Huang P, Guo W, Zheng Y. Complete chloroplast genome sequence of Decaisnea insignis: genome organization, genomic resources and comparative analysis. Scientific reports. 2017; 7(1): 10073. https://doi.org/10.1038/s41598-017-10409-y PMID: 28855603.

44. Asaf S, Waqas M, Khan AL, Khan MA, Imran QM, et al. The complete chloroplast genome of wild rice (Oryza minuta) and its comparison to related species. Frontiers in plant science. 2017; 8: 304. https://doi.org/10.3389/fpls.2017.00304 PMID: 28326093.

45. Chen J, Hao Z, Xu H, Yang L, Liu G, Sheng Y, et al. The complete chloroplast genome sequence of the relict woody plant Metasequoia glyptostroboides Hu et Cheng. Frontiers in plant science. 2015; 6: 447. https://doi.org/10.3389/fpls.2015.00447 PMID: 26136762.

46. Ren T, Yang Y, Zhou T, Liu ZL. Comparative plastid genomes of Primula species: sequence divergence and phylogenetic relationships. 2018; 19(4). https://doi.org/10.3399/ijms19041050 PMID: 29614787.

47. Shetty SM, Md Shah MU, Makale K, Mohd-Yusuf Y, Khalid N, Othman RY. Complete chloroplast genome sequence of corroborates structural heterogeneity of inverted repeats in wild progenitors of cultivated bananas and plantains. Plant Genome. 2016; 9(2). https://doi.org/10.3835/plantgenome2015.09.0089.

48. Yao X, Tang P, Li Z, Li D, Liu Y, Huang H. The first complete chloroplast genome sequences in Actinidiaceae: genome structure and comparative analysis. PloS one. 2015; 10(6): e0129347. https://doi.org/10.1371/journal.pone.0129347 PMID: 26046631.
49. Yang Y, Dang Y, Li Q, Lu J, Li X, Wang Y. Complete chloroplast genome sequence of poisonous and medicinal plant Datura stramonium: organizations and implications for genetic engineering. PloS one. 2014; 9(11): e110656. https://doi.org/10.1371/journal.pone.0110656 PMID: 25365514.

50. Dong W, Xu C, Li C, Sun J, Zuo Y, Shi S, et al. ycf1, the most promising plastid DNA barcode of land plants. Scientific reports. 2015; 5: 8348. https://doi.org/10.1038/srep08348 PMID: 25672218

51. Shingo K, Jocelyn B, Minako H, Yoshino H, Maya O, Midori I, et al. Uncovering the protein translocon at the chloroplast inner envelope membrane. Science. 2013; 339(6119): 571–574. https://doi.org/10.1126/science.1229262 PMID: 23372012

52. Simeone MC, Grimm GW, Papini A, Vessella F, Cardoni S, Tordoni E, et al. Plastome data reveal multiple geographic origins of Quercus Group Ilex. PeerJ. 2016; 4: e1897. Epub 2016/04/29. https://doi.org/10.7717/peerj.1897 PMID: 27123376.

53. Vitelli M, Vessella F, Cardoni S, Pollegioni P, Denk T, Grimm GW, et al. Phylogeographic structuring of plastome diversity in Mediterranean oaks (Quercus Group Ilex, Fagaceae). Tree Genetics & Genomes. 2017; 13(1): 3.

54. Manos PS, Zhou ZC. Systematics of Fagaceae: Phylogenetic tests of reproductive trait evolution. International Journal of Plant Sciences. 2001; 162(6): 1361–1379.

55. Oh SH, Manos PS. Molecular Phylogenetics and Cupule Evolution in Fagaceae as Inferred from Nuclear CRABS CLAW Sequences. Taxon. 2008; 57(2): 434–451.