Resolving the Phylogenetic Position and Evolution of *Semiliquidambar* Based on Chloroplast Genomes (Postprint)

Authors: Zhou Yun, Wei Yanyan, Zhou Yun

Date: 2023-09-20T00:00:00+00:00

Abstract

To clarify the phylogenetic relationships of the nationally second-class protected plant Semiliquidambar cathayensis with its related taxa and analyze the adaptive evolution of chloroplast genes, this study utilized 24 chloroplast genome sequences from 22 species to construct maximum likelihood and Bayesian trees to analyze the phylogenetic relationships of S. cathayensis and its relatives, and examined the relationship between variation sites in chloroplast coding genes and selection pressure in S. cathayensis and related taxa through different models. The results showed that: (1) The chloroplast genome of S. cathayensis contains 133 genes, including 88 protein-coding genes (11 of which have introns), 37 tRNA genes, and 8 rRNA genes. (2) The chloroplast genomes of S. cathayensis and eight species from its related genera Altingia and Liquidambar were relatively conserved in terms of sequence length, gene number and composition, and GC content, with highly conserved boundaries between inverted repeat regions and small single-copy regions. The small single-copy and large single-copy regions showed higher degrees of variation, whereas the inverted repeat regions exhibited lower variation. (3) S. cathayensis clustered with species of Altingia and Liquidambar into the Altingiaceae clade, which could be divided into three subclades, with possible hybridization or incomplete lineage sorting among subclades or between species. (4) Adaptive evolution results indicated that under different models, species in the Altingiaceae clade experienced selective constraints (purifying selection) on chloroplast genes such as ndhA, and the site model also detected 28 sites across 10 genes with p-values greater than 0.99, suggesting that these coding gene variations may be associated with adaptive differentiation in Altingiaceae plants. This study supports the placement of S. cathayensis within Altingiaceae, and suggests that chloroplast genes in Altingiaceae species may have undergone adaptive evolution, providing reference materials for the conservation of medicinal resources with the same name but different species and for the innovative development of ethnic medicines.

Full Text

Preamble

DOI: 10.11931/guihaia.gxzw202209047
Title: Resolving Phylogenetic Position and Evolution of *Semiliquidambar cathayensis* Based on Chloroplast Genomes
Authors: ZHOU Yun*, WEI Yanyan (College of Pharmacy, Guangxi Medical University, Nanning 530021, China)

Abstract

To clarify the phylogenetic relationships between the nationally second-class protected plant *Semiliquidambar cathayensis* and its closely related taxa, and to analyze the adaptive evolution of chloroplast genes, this study constructed maximum likelihood and Bayesian trees using 24 chloroplast genome sequences from 22 species to analyze the phylogenetic relationships of *S. cathayensis* and its relatives. The correlation between variable sites in chloroplast coding genes and selective pressure among *S. cathayensis* and related taxa was examined using different models. The results showed: (1) The chloroplast genome of *S. cathayensis* contains 133 genes, including 88 protein-coding genes (11 with introns), 37 tRNA genes, and 8 rRNA genes. (2) Chloroplast genomes of eight species from *S. cathayensis* and its related genera *Altingia* and *Liquidambar* are relatively conserved in sequence length, gene number and composition, and GC content, with highly conserved boundaries between inverted repeat (IR) regions and small single-copy (SSC) regions. The large single-copy (LSC) and SSC regions exhibit relatively high variation, while IR regions show lower variation. (3) *S. cathayensis* and species from *Altingia* and *Liquidambar* clustered into the Altingiaceae clade, which could be divided into three subclades, with possible hybridization or incomplete lineage sorting among subclades or species. (4) Adaptive evolution analyses revealed that under different models, species in the Altingiaceae clade experienced selective constraints (purifying selection) on chloroplast genes such as *ndhA*. Site models also detected 28 sites across 10 genes with P-values greater than 0.99, suggesting these coding gene variations may be associated with adaptive differentiation in Altingiaceae. This study supports the placement of *S. cathayensis* within Altingiaceae and suggests that chloroplast genes among Altingiaceae species may have undergone adaptive evolution, providing reference materials for resource conservation of homonymic medicinal materials and innovative research on ethnomedicine.

Keywords: *Semiliquidambar cathayensis*, Altingiaceae, chloroplast genome, phylogeny, adaptive evolution

Introduction

The Yao medicine "Banfenghe" (also known as Banhefeng or Bianhebeng) is derived from the dried aerial parts of *Semiliquidambar cathayensis*, which contains

alkaloids, flavonoids, terpenoids, and other compounds with therapeutic effects for dispelling wind-dampness and promoting blood circulation (Yang et al., 2019; Qiu et al., 2020). Its wild resources are sparsely distributed in mountainous regions of southern and southeastern China, and it is listed as a national second-class protected plant. Current molecular markers used to study genetic diversity in *S. cathayensis* populations include ISSR (Huang et al., 2021), SSR (Ye et al., 2020, 2021a), and SRAP (Ye et al., 2021b), which indicate that *S. cathayensis* populations have unstable structures and are endangered due to human disturbance and habitat destruction, making rational development and management extremely important and the search for new medicinal sources urgent (Fu, 1991; Ye et al., 2020b). Due to morphological characteristics such as heteromorphic leaves, spicate inflorescences, and monoecious plants, *S. cathayensis* has been classified in the subfamily Liquidambaroideae of Hamamelidaceae (Guangxi Institute of Botany, Chinese Academy of Sciences, 2005). Plants in this genus share morphological similarities with *Liquidambar* and *Altingia* in Altingiaceae, and *Altingia chinensis* from *Altingia* is also used as the medicinal material Banfenghe in some regions. There are 14 species from 7 genera across 5 families that share the same name and medicinal use as Banfenghe (Xie et al., 2018), necessitating clarification of the origins or alternative sources of these medicinal materials. Molecular phylogenetic studies have suggested that *S. cathayensis* should be placed in the family Altingiaceae (Angiosperm Phylogeny Group, 1998; Shi et al., 2001). Current molecular systematic studies involving *S. cathayensis* have focused on analyzing phylogenetic relationships between Hamamelidaceae and Altingiaceae or among genera within Altingiaceae (Shi et al., 1998, 2019; Ickert-Bond & Wen, 2006; Wu et al., 2010; Xiang et al., 2019; Tang et al., 2020; Ye et al., 2020a; Zhang et al., 2020a), while interspecific phylogenetic relationships among *S. cathayensis* and its close relatives require further analysis. Investigating the phylogenetic relationships between *S. cathayensis* and Altingiaceae or Hamamelidaceae plants is beneficial for molecular identification of *S. cathayensis* and its adulterants, thereby achieving the goal of clarifying origins. Furthermore, since the three genera of Altingiaceae are mainly distributed in southwestern China, whether there are divergences in gene structure or evolutionary rates among *S. cathayensis* and its close relatives that may affect pharmacological efficacy requires further analysis.

Comparative analysis of chloroplast genome sequences and construction of phylogenetic trees can be used to evaluate species' phylogenetic positions and evolutionary relationships (Sloan et al., 2014; Williams et al., 2019). Based on this, using differences in evolutionary rates of chloroplast coding genes to assess the relationship between gene variation and selective pressure in different plant groups can serve as a foundation for exploring new medicinal sources (Waldvogel et al., 2020; Zhao et al., 2020). Currently, chloroplast genomes of multiple species from *S. cathayensis* and its relatives in Altingiaceae and Hamamelidaceae have been reported. Therefore, this study utilizes publicly available chloroplast genomes of *S. cathayensis* and its relatives from public databases to address the following questions: (1) Reveal the phylogenetic position of *S. cathayensis*

within Altingiaceae or Hamamelidaceae through phylogenetic tree construction; (2) Analyze the relationship between evolutionary sites and selective pressure in chloroplast genes of *S. cathayensis* and its relatives, providing insights for identification and resource development of *S. cathayensis*.

1.1 Data Collection

Chloroplast genome sequence information for *S. cathayensis*, Altingiaceae, Hamamelidaceae, and other related taxa was retrieved from the NCBI database (The National Center for Biotechnology Information). A total of 24 chloroplast genome sequences from 22 species were obtained, including 2 individual sequences of *S. cathayensis* from the genus *Semiliquidambar*, 3 species from *Altingia*, and 4 species from *Liquidambar* (with 2 individual sequences of *Liquidambar formosana*). Additionally, 10 Hamamelidaceae species including *Hamamelis mollis* and *Sycopsis sinensis*, 1 Daphniphyllaceae species, 1 Cercidiphyllaceae species, and 2 Ranunculaceae species were selected. Details of the downloaded chloroplast genome sequences, species names, and accession numbers are provided in Table 1 .

1.2 Comparative Analysis of Chloroplast Genomes

Geneious R9 (Kearse et al., 2012) was used to analyze and compile information on the four boundaries (large single-copy region LSC, small single-copy region SSC, and inverted repeat region IR) and gene numbers/types for eight Altingiaceae species (Table 1). The IRscope program in R software (Amiryousefi et al., 2018) was used to visualize contraction and expansion of the four boundaries in these eight Altingiaceae species. mVISTA software (Frazer et al., 2004) was employed to conduct homology comparisons of chloroplast genome sequences from *S. cathayensis* and seven other Altingiaceae species using the Shuffle-LAGAN global alignment mode that detects gene rearrangements and inversions. Additionally, the Mauve multiple genome alignment method in Geneious R9 was used to perform collinearity comparisons of chloroplast genome sequences from these eight species.

1.3 Phylogenetic Analysis

This study used the online MAFFT v7 (Katoh et al., 2002) (https://mafft.cbrc.jp/alignment/software/) to align 24 chloroplast genome sequences from 22 species retrieved from NCBI (Table 1). DAMBE v6.4.29 was used to test for sequence substitution saturation to assess whether the alignment was suitable for phylogenetic analysis, evaluating whether the observed Iss value (index of substitution saturation) was significantly lower than the Iss.c value (critical index of substitution saturation) (Xia et al., 2003; Xia & Lemey, 2009). The Iss.c value was further divided into Iss.cSym (symmetrical substitution saturation index) and IssAsym (asymmetrical substitution saturation index). The CIPRES Web Portal 2.0 (http://www.phylo.org) was used to construct maximum likelihood (ML) trees

with RAxML. The ML analysis employed the GAMMA model with 1,000 rapid bootstrap replicates. jModelTest (Posada & Crandall, 1998; Darriba et al., 2012) was used to calculate AIC values and select the optimal model for tree construction for each dataset. MrBayes v3.2.6 (Ronquist & Huelsenbeck, 2003) was used to construct Bayesian (BI) trees, running for 3 million generations with sampling every 1,000 generations, discarding the first 25% of trees that did not reach stationarity, and using the remaining trees to calculate posterior probabilities.

This study used coalescent simulation to detect the impact of incomplete lineage sorting on conflicts between nuclear gene trees and chloroplast trees. First, under the coalescent model, DendroPy v4.1.0 (Sukumaran & Holder, 2010) was used to simulate 10,000 chloroplast species trees from 24 chloroplast genome sequences. ASTRAL was used to construct a species tree as a reference, and these 10,000 simulated trees were summarized to obtain clade frequencies. Under incomplete lineage sorting, any phylogenetic relationships obtained in the empirical plastid tree should be reflected in the simulated plastid trees, thus each clade should have relatively high support. In cases of hybridization, some clades in the empirical plastid tree would have low or absent support in the simulated gene trees (Garía et al., 2017; Morales-Briones et al., 2018).

1.4 Adaptive Evolution Analysis

Geneious R9 was used to extract and align individual genes from the 24 chloroplast genomes, with stop codons removed. EasyCodeML (Gao et al., 2019) was used to batch-convert gene alignment sequences to .pml format. The maximum likelihood tree constructed from 24 chloroplast genomes was used as the tree file (.nwk format). EasyCodeML was then used to perform adaptive evolution analyses using clade model, branch model, and site model. The clade model can detect selective constraints on specific sites across an entire clade or focal lineage; the branch model can detect the intensity of selective constraints on the focal lineage; and the site model can detect positively selected sites without considering lineages. In both the clade model and branch model, the Altingiaceae clade was designated as the foreground branch.

2.1 Basic Characteristics of the *Semiliquidambar cathayensis* Chloroplast Genome Structure

The chloroplast genome of *S. cathayensis* ranges from 160,430 to 160,444 bp, with the LSC region spanning 88,969 to 88,991 bp, the SSC region 18,913 to 18,917 bp, and the IR region 26,261 to 26,281 bp. The GC content is uniformly 37.9% with no variation. Compared with seven other Altingiaceae species, the chloroplast genomes of all eight species contain 133 genes total, including 8 rRNA genes (4 located in IR regions), 37 tRNA genes (7 in IR regions), and 88 protein-coding genes (7 in IR regions). Eleven protein-coding genes contain introns: *rps16*, *atpF*, *rpoC1*, *ycf3*, *clpP*, *petB*, *petD*, *rpl16*, *rpl2*, *ndhB*, and

ndhA. Among these, *rpl2* and *ndhB* are distributed in IR regions, while *ycf3* and *clpP* have two introns. Variation in intron length leads to differences in gene length (Table 1). Among the eight species, the protein-coding sequences of *matK* and *ndhK* genes show length variation. In comparison, seven Hamamelidaceae species have 133 genes total, with 8 rRNA genes (4 in IR regions), 37 tRNA genes (7 in IR regions), and 88 protein-coding genes (7 in IR regions).

2.2 IR and SC Boundaries in *Semiliquidambar cathayensis* and Related Genera

Chloroplast genomes of the eight species from *S. cathayensis* and its relatives are relatively conserved in sequence length, gene order and number, and GC content. The transition regions between IR and SC boundaries show no differences in gene arrangement, with only variations in gene sequence length, demonstrating high conservation (Figure 1 [Figure 1: see original paper]). Except for *S. cathayensis*, where the IRb-LSC boundary lies in the spacer region between *rps19* and *rpl2*, the other seven species have the boundary within the *rps19* gene. The IR-SSC boundary is located within the *ycf1* gene in all species.

Figure 1 Comparison of borders of IR and SC regions in eight species of *Semiliquidambar cathayensis* and its related taxa. JLB: Junction of LSC and IRb; JSB: Junction of SSC and IRb; JSA: Junction of SSC and IRa; JLA: Junction of LSC and IRa.

2.3 Sequence Variation Analysis of Chloroplast Genomes in *Semiliquidambar cathayensis* and Related Genera

Using mVISTA software, homology of chloroplast genome sequences from *S. cathayensis* and seven other Altingiaceae species was analyzed to detect gene rearrangements and inversions under global alignment mode (Figure 2 [Figure 2: see original paper]). The results showed that the four regions of chloroplast genomes from the eight Altingiaceae species have highly conserved arrangement order, with obvious variation in non-coding regions but not in coding regions. Both SSC and LSC regions show significant variation, particularly in intergenic spacer regions of the LSC, while IR regions exhibit relatively low variation.

Figure 2 Alignment of eight chloroplast genomes from *Semiliquidambar cathayensis* (reference sequence) and its related taxa.

2.4 Collinearity Analysis of Chloroplast Genomes in *Semiliquidambar cathayensis* and Related Genera

The Mauve alignment method in Geneious R9 was used to detect rearrangements and collinearity in chloroplast genomes of *S. cathayensis* and seven Altingiaceae species (Figure 3 [Figure 3: see original paper]). Multiple genome alignment detected one locally collinear block among the chloroplast genomes of the eight

species, indicating high similarity among Altingiaceae genomes with no detected rearrangements or inversions.

Figure 3 Alignment of chloroplast genomes from eight species of *Semiliquidambar cathayensis* and its related taxa. Small blocks of various colors represent genes: black for transfer RNA (tRNA), red for ribosomal RNA, white for protein-coding genes, and green for intron-containing tRNA.

2.5 Phylogenetic Analysis

DAMBE was used to test for base substitution saturation in the chloroplast genome sequence matrix. The results showed that for randomly sampled groups of 4, 8, 16, and 32 taxa, the index of substitution saturation (Iss) was significantly lower than the critical index of substitution saturation (Iss.c) in both symmetrical and asymmetrical topologies, indicating the sequences were suitable for phylogenetic tree construction. The best-fit model selected was GTR+I+G. The maximum likelihood (ML) and Bayesian (BI) trees constructed from 24 chloroplast genome sequences showed consistent topologies (Figure 4 [Figure 4: see original paper]). Altingiaceae species formed one clade, while Hamamelidaceae species formed another. Within Altingiaceae, *S. cathayensis* clustered with *Altingia chinensis*, *Liquidambar formosana*, and *L. acalycina* to form Clade I (99/1.00), *A. excelsa* and *A. yunnanensis* formed Clade II (100/1.00), and *L. orientalis* and *L. styraciflua* formed Clade III (100/1.00). *Daphniphyllum oldhamii* and *Cercidiphyllum japonicum* each formed separate lineages as successive sister groups to Altingiaceae. Sampled Hamamelidaceae species could be divided into Fothergilleae, Hamamelideae, Eustigmateae, and Corylopsideae clades. Coalescent simulation to detect the impact of incomplete lineage sorting on chloroplast tree topology revealed low clade frequencies for multiple branches (Figure 4), suggesting that phylogenetic positions within Altingiaceae may be influenced by both chloroplast incomplete lineage sorting and hybridization.

Figure 4 Phylogenetic tree based on 24 chloroplast genomes, drawn based on the BI tree. The scale bar shows 0.02 substitutions/site. Numbers above branches are ML bootstrap values, BI posterior probabilities, and clade frequencies of simulated gene trees.

2.6 Adaptive Divergence

Using the branch model with Altingiaceae as the foreground clade, three genes were found to be under selective constraint ($0<\omega<1$): *ndhA* (ω=0.051, P<0.05), *ndhG* (ω=0.024, P<0.01), and *rps12* (ω=0.0001, P<0.01). Using the clade model with Altingiaceae as the foreground clade, 12 genes (*atpE*, *atpF*, *ndhA*, *ndhJ*, *psbM*, *rbcL*, *rpl14*, *rpoC2*, *rps2*, *rps3*, *rps4*, *rps12*, *rps14*) showed significant selective constraints (Table 2). Site model analysis to detect positively selected sites across chloroplast genome genes revealed that 10 genes including *accD*, *atpE*, *atpF*, *clpP*, *ndhA* were under selection pressure, with 45 sites showing positive selection at P>0.95 and 28 sites at P>0.99 (Table 3).

Table 2 Selection analysis of 75 genes in chloroplast genome based on the clade model. The table shows log-likelihood values and likelihood ratio test P-values for genes including *rpl14*, *rpoC2*, *rps12*, and *rps14*.

Table 3 Selection analysis of 75 genes in chloroplast genome based on the site model. The table shows positively selected sites with P>95% (*) and P>99% (**), listing specific amino acid positions and their corresponding P-values for genes such as *accD*, *clpP*, and *rbcL*.

3.1 Phylogenetic Relationships Among *Semiliquidambar* and *Altingia* and *Liquidambar* Genera

This study analyzed the phylogenetic relationships between *Semiliquidambar* and *Altingia* and *Liquidambar* based on chloroplast genomes. The results showed that species from the three genera could not each form monophyletic groups, supporting the placement of *Semiliquidambar* within Altingiaceae. This is consistent with molecular systematics based on chloroplast gene data and pollen morphology studies (Ickert-Bond & Wen, 2013). However, *Flora of China* classifies *Semiliquidambar*, *Altingia*, and *Liquidambar* within the subfamily Liquidambaroideae of Hamamelidaceae based on morphological traits such as heteromorphic leaves and unisexual flowers. Key morphological diagnostic characters for these genera—including leaf length and shape, petiole thickness and length, and calyx tooth length on infructescences—show continuous variation among congeneric or closely related species, lacking systematic quantitative standards. Additionally, *Semiliquidambar* exhibits transitional morphological traits with *Liquidambar* and *Altingia* within the same family, making it difficult to distinguish *Semiliquidambar* from *Altingia* and *Liquidambar* based on morphology alone.

Previous molecular systematics studies have indicated that *Semiliquidambar*, *Altingia*, and *Liquidambar* species form an East Asian clade, but relationships among species within the clade remain unclear and may require further development of high-resolution molecular markers (Ickert-Bond & Wen, 2013). Based on comparative chloroplast genome analysis, this study found that chloroplast genomes of eight Altingiaceae species show high conservation in gene structure, arrangement, and number, as well as IR-SC junction regions. Length variation among species' genomes primarily reflects differences in intron or coding sequence lengths. Future studies could develop high-resolution molecular markers from relatively variable regions such as intergenic spacers in chloroplast genomes for molecular identification and genetic differentiation studies of *S. cathayensis*.

3.2 Phylogenetic Relationships of *Semiliquidambar cathayensis* and Its Close Relatives

This study analyzed phylogenetic relationships between *S. cathayensis* and its close relatives based on chloroplast genomes. The results showed that sampled individuals of *S. cathayensis* did not form a monophyletic group; instead,

S. cathayensis clustered with *Altingia chinensis*, *Liquidambar formosana*, and *L. acalycina* to form Clade I (99/1.00). Coalescent-based testing of chloroplast gene trees revealed low clade frequencies for multiple branches within Altingiaceae, suggesting that hybridization or chloroplast incomplete lineage sorting may exist among branches within the family, and thus the parentage of *S. cathayensis* cannot be definitively determined. This is consistent with previous molecular systematic results showing that *S. cathayensis* forms a monophyletic clade with *Liquidambar* species and *L. acalycina*, but with low support, leading to speculation that *S. cathayensis* may be a hybrid species (Shi et al., 2001; Ickert-Bond & Wen, 2013). Due to limited gene fragments and different samples used in existing studies, the degree of hybridization between *S. cathayensis* and *Liquidambar* and *Altingia* species remains unclear, necessitating population-level studies of the genetic background of *S. cathayensis*.

3.3 Adaptive Differentiation of Altingiaceae Plants

Environmental changes can drive adaptive evolution of plant genes. This study employed different models to detect selection pressure on multiple genes in Altingiaceae and its relatives, aiming to explore potential correlations between selection pressure at gene sites and related species or environments.

Based on the clade model results, when Altingiaceae was designated as the foreground clade, multiple *atp*, *ndh*, and *rps* genes showed significant selective constraints. These genes are mostly associated with photosynthesis, transcription, and translation functions in other plant groups. For example, the *ndh* gene family plays a crucial role in photosynthesis and is sensitive to environmental changes or plant stress (Martín & Sabater, 2010; Zhao et al., 2020). Ribosomal protein subunit genes *rpl* and *rps* are important for plant transcription and translation, and transcripts of genes such as *rps12* and *atpF* are also involved in splicing of chloroplast group II introns (Vogel et al., 1999). ATP synthase genes are essential for photosynthesis, while the *rpoC2* gene encodes the β subunit of chloroplast RNA polymerase, playing an important role in transcription. Since most Altingiaceae species are distributed in southwestern China, we hypothesize that their photosynthetic functions may differ from Hamamelidaceae species to adapt to relatively high-temperature and high-humidity environments. This is similar to previous transcriptome results for *S. cathayensis*, which found 92 unigenes mapped to photosynthesis regulatory pathways and 32 unigenes mapped to photosynthesis-antenna protein pathways (Tian et al., 2018). Thus, this study may provide a foundation for investigating light response in *S. cathayensis* at the gene level.

Site model analysis was used to examine whether amino acid sites in different genes experienced selection pressure. The results showed that 45 sites across 10 genes including *accD*, *clpP*, *rbcL*, *ycf1*, and *ycf2* were under positive selection with P>0.95, and 28 sites had P>0.99. Genes with more detected positively selected sites showed greater variation. These genes function in multiple processes including photosynthesis and plant metabolism. For instance, previous

studies have shown that the *accD* gene encodes the β subunit of acetyl-CoA carboxylase and functions in fatty acid biosynthesis; its frequent transfer or loss in angiosperm chloroplasts is considered an adaptation to the environment (Slabas & Fawcett, 1992), suggesting that changes in this gene may help Altingiaceae plants adapt to their habitats. The *clpP* gene encodes a Clp protease in the chloroplast genome that degrades polypeptides, helping control normal metabolic processes and playing an important role in plant stress resistance (Zheng et al., 2016), suggesting it may be crucial for Altingiaceae species adapting to the relatively hot and humid environment of southwestern China. The *rbcL* gene acts as a regulator of photosynthetic electron transport in plant photosynthesis, encoding the large subunit of Rubisco in chloroplasts, whose C-terminal region is significant in the photosynthetic system (Curmi et al., 1992). The *ycf1* and *ycf2* genes evolve rapidly and function in encoding chloroplast ATPase and regulating plant fruit development, showing similar patterns to other plants and indicating that *ycf* genes commonly undergo adaptive evolution (Zhou et al., 2019). In summary, these genes may have played important roles in the adaptation of Altingiaceae plants, including *S. cathayensis* medicinal sources, to the environmental conditions of southern China.

Currently, *Semiliquidambar* plants are mainly concentrated in southwestern China. Limited sampling ranges and molecular marker selection in existing studies have left intergeneric relationships within *Semiliquidambar* unclear. This study only analyzed chloroplast genome structural differences, phylogenetic relationships, and gene site selection pressure for *S. cathayensis* and its relatives based on publicly available database sequences, which was insufficient to infer relationships among *S. cathayensis*, *Liquidambar*, *L. acalycina*, and *Altingia* species. Therefore, future studies should expand sampling ranges for *S. cathayensis* and its relatives and select higher-resolution molecular markers to explore hybridization or incomplete lineage sorting between *S. cathayensis* and *Altingia* and *Liquidambar* species.

References

AMIRYOUSEFI A, HYVÖNEN J, POCZAI P, 2018. IRscope: an online program to visualize the junction sites of chloroplast genomes [J]. *Bioinformatics*, 34(17): 3030-3031.

ANGIOSPERM PHYLOGENY GROUP, 1998. An ordinal classification for the families of flowering plants[J]. *Ann Missouri Bot Gard*, 85(4): 531-553.

CHOI KS, HA YH, JEONG KS, et al., 2019. The complete chloroplast genome of *Corylopsis coreana* (Hamamelidaceae)[J]. *Conserv Genet Resour*, 11(3): 291-293.

CURMI PM, CASCIO D, SWEET RM, et al., 1992. Crystal structure of the unactivated form of ribulose-1,5-bisphosphate carboxylase/oxygenase from tobacco refined at 2.0-Å resolution[J]. *J Biol Chem*, 267(24): 16980-16989.

DARRIBA D, TABOADA GL, DOALLO R, et al., 2012. jModelTest 2: more models, new heuristics and parallel computing[J]. *Nat Methods*, 9: 772.

DONG WP, XU C, CHENG T, et al., 2013. Sequencing angiosperm plastid genomes made easy: a complete set of universal primers and a case study on the phylogeny of Saxifragales[J]. *Genome Biol Evol*, 5(5): 989-997.

DONG WP, XU C, WU P, et al., 2018. Resolving the systematic positions of enigmatic taxa: manipulating the chloroplast genome data of Saxifragales[J]. *Mol Phylogenet Evol*, 126: 321-330.

FRAZER KA, PACHTER L, POLIAKOV A, et al., 2004. VISTA: computational tools for comparative genomics[J]. *Nucleic Acids Res*, 32 (Web Server issue): W273-W279.

FU LG, 1991. *Red book of Chinese plants: rare and endangered plants*[M]. Beijing: Science Press. [Fu LG, 1991. Red book of Chinese plants: rare and endangered plants[M]. Beijing: Science Press.]

GAO FL, CHEN CJ, ARAB DA, et al., 2019. EasyCodeML: A visual tool for analysis of selection using CodeML[J]. *Ecol Evol*, 9(7): 3891-3898.

GARÍA N, FOLK RA, MEEROW AW, et al., 2017. Deep reticulation and incomplete lineage sorting obscure the diploid phylogeny of rain-lilies and allies (Amaryllidaceae tribe Hippeastreae)[J]. *Mol Phylogenet Evol*, 111: 231-247.

GUANGXI INSTITUTE OF BOTANY, CHINESE ACADEMY OF SCIENCES, 2005. *Flora of Guangxi*: Vol. 2[M]. Nanning: Guangxi Science Press: 690-692. [Guangxi Institute of Botany, Chinese Academy of Sciences, 2005. Flora of Guangxi: Vol. 2[M]. Nanning: Guangxi Science Press: 690-692.]

HUANG LH, CHEN QT, XIAO YS, et al., 2021. Optimization and primers screening of ISSR-PCR reaction system for *Semiliquidambar cathayensis* Chang[J]. *Mol Plant Breed*, 19(20): 6782-6789. [Huang LH, Chen QT, Xiao YS, et al., 2021. Optimization and primers screening of ISSR-PCR reaction system for Semiliquidambar cathayensis Chang[J]. Molecular Plant Breeding, 19(20): 6782-6789.]

ICKERT-BOND SM, WEN J, 2006. Phylogeny and biogeography of Altingiaceae: evidence from combined analysis of five noncoding chloroplast regions[J]. *Mol Phylogenet Evol*, 39(2): 512-528.

ICKERT-BOND SM, WEN J, 2013. A taxonomic synopsis of Altingiaceae with nine new combinations[J]. *PhytoKeys*, 31: 21-61.

KATOH K, MISAWA K, KUMA K, et al., 2002. MAFFT: A novel method for rapid multiple sequence alignment based on a fast Fourier transformation[J]. *Nucleic Acids Res*, 30(14): 3059-3066.

KEARSE M, MOIR R, WILSON A, et al., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data[J]. *Bioinformatics*, 28(12): 1647-1649.

KELLY M, 2019. Adaptation to climate change through genetic accommodation and assimilation of plastic phenotypes[J]. *Philos Trans R Soc Lond B Biol Sci*, 374(1768): 20180176.

KIM SC, SHIN S, AHN JY, et al., 2019. Complete chloroplast genome of *Corylopsis spicata* and phylogenetic analysis[J]. *Mitochondrial DNA Part B*, 4(2): 2700-2701.

LAI JX, LIN FR, HUANG P, et al., 2020. Characterization of the complete chloroplast genome of *Liquidambar acalycina* Chang[J]. *Mitochondrial DNA Part B*, 5(2): 1697-1698.

LEE M, PARK JH, GIL J, et al., 2019. The complete chloroplast genome of *Paeonia lactiflora* Pall. (Paeoniaceae)[J]. *Mitochondrial DNA Part B*, 4(2): 2715-2716.

LI HL, CHENG XL, CHEN Y, et al., 2019. Complete plastome sequence of *Rhodoleia championii* Hook. f.(Hamamelidaceae)[J]. *Mitochondrial DNA Part B*, 4(2): 3458-3459.

MARTÍN M, SABATER B, 2010. Plastid *ndh* genes in plant evolution[J]. *Plant Physiol Biochem*, 48(8): 636–645.

MORALES-BRIONES D, LISTON A, TANK DC, 2018. Phylogenomic analyses reveal a deep history of hybridization and polyploidy in the Neotropical genus *Lachemilla* (Rosaceae)[J]. *New Phytol*, 218(4): 1668-1684.

PENG Y, YANG LM, WEI J, 2020. The complete chloroplast genome of *Sycopsis sinensis* Oliver[J]. *Mitochondrial DNA Part B*, 5(3): 2984-2985.

POSADA D, CRANDALL KA, 1998. jModeltest: testing the model of DNA substitution[J]. *Bioinformatics*, 14: 817–818.

QIU Q, YANG DJ, XU LH, et al., 2020. The complete chloroplast genome sequence of *Altingia yunnanensis*[J]. *Mitochondrial DNA Part B*, 5(1): 1050-1051.

QIU S, CHEN YY, YAN XJ, et al., 2020. Chemical constituents from the leaves of *Semiliquidambar cathayensis*[J]. *J Chin Med Mater*, 43(5): 1136–1139. [Qiu S, Chen YY, Yan XJ, et al., 2020. Study on chemical constituents from the leaves of Semiliquidambar cathayensis[J]. Journal of Chinese Medicinal Materials, 43(5): 1136-1139.]

REN XL, DU XM, XIN GL, et al., 2018. The complete chloroplast genome of *Sinowilsonia henryi* (Saxifragales: Hamamelidaceae), an endangered relict species[J]. *Conserv Genet Resour*, 10(4): 643-645.

RONQUIST F, HUELSENBECK JP, 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models[J]. *Bioinformatics*, 19(12): 1572-1574.

SHI S, CHANG HT, CHEN YQ, et al., 1998. Phylogeny of the Hamamelidaceae based on the ITS sequences of nuclear ribosomal DNA[J]. *Biochem Syst Ecol*,

26(1): 55-69.

SHI S, HUANG Y, ZHONG Y, et al., 2001. Phylogeny of the Altingiaceae based on cpDNA *matK*, *PY-IGS* and nrDNA ITS sequences[J]. *Plant Syst Evol*, 230: 13-24.

SHI YC, DUAN N, LIU BB, 2019. Complete chloroplast genome sequence of *Semiliquidambar cathayensis* (Hamamelidaceae), a rare and endangered species endemic to China[J]. *Mitochondrial DNA B*, 4(2): 3252-3253.

SLABAS AR, FAWCETT T, 1992. The biochemistry and molecular biology of plant lipid biosynthesis[J]. *10 Years Plant Mol Biol*, 19(1): 169-191.

SLOAN DB, TRIANT DA, FORRESTER NJ, et al., 2014. A recurring syndrome of accelerated plastid genome evolution in the angiosperm tribe Sileneae (Caryophyllaceae)[J]. *Mol Phylogenet Evol*, 72: 82-89.

SUKUMARAN J, HOLDER MT, 2010. DendroPy: A Python library for phylogenetic computing[J]. *Bioinformatics*, 26(12): 1569-1571.

TANG XH, FAN HH, ZHANG J, et al., 2020. The complete chloroplast genome of *Semiliquidambar cathayensis* HT Chang 'T5' (Hamamelidaceae)[J]. *Mitochondrial DNA Part B*, 5(2): 1267-1268.

TIAN XM, ZENG LZ, YAN LH, et al., 2018. Study on transcriptome characteristic of *Semiliquidambar cathayensis* Chang[J]. *Hunan Forestry Sci Tech*, 45(5): 40-50. [Tian XM, Zeng LZ, Yan LH, et al., 2018. Study on transcriptome characteristics of Semiliquidambar cathayensis Chang[J]. Hunan Forestry Science and Technology, 45(5): 40-50.]

VOGEL JC, RUMSEY FJ, RUSSELL SJ, et al., 1999. Genetic structure, reproductive biology and ecology of isolated populations of *Asplenium csikii* (Aspleniaceae, Pteridophyta)[J]. *Heredity*, 83(5): 604-612.

WALDVOGEL AM, FELDMEYER B, ROLSHAUSEN G, et al., 2020. Evolutionary genomics can improve prediction of species' responses to climate change[J]. *Evol Lett*, 4(1): 4-18.

WANG Y, LI YQ, YUAN XL, et al., 2019. The complete chloroplast genome sequence of *Mytilaria laosensis*[J]. *Mitochondrial DNA Part B*, 4(2): 3916-3917.

WILLIAMS AM, FRISO G, VANWIJK KJ, et al., 2019. Extreme variation in rates of evolution in the plastid Clp protease complex[J]. *Plant J*, 98(2): 243-259.

WU W, ZHOU RC, HUANG YL, et al., 2010. Molecular evidence for natural intergeneric hybridization between *Liquidambar* and *Altingia*[J]. *J Plant Res*, 123(2): 231-239.

XIA X, LEMEY P, 2009. Assessing substitution saturation with DAMBE. In: Philippe L. (ed.), *The Phylogenetic Handbook: A Practical Approach to DNA and Protein Phylogeny*[M]. London: Cambridge University Press, 615-630.

XIA X, XIE Z, SALEMI M, et al., 2003. An index of substitution saturation and its application[J]. *Mol Phylogenet Evol*, 26(1): 1-7.

XIANG XG, XIANG KL, ORTIZ RDC, et al., 2019. Integrating palaeontological and molecular data uncovers multiple ancient and recent dispersals in the pantropical Hamamelidaceae[J]. *J Biogeogr*, 46(11): 2622-2631.

XIE SY, YAO KL, WU XJ, 2018. Overview of pharmacological research on *Semiliquidambar cathayensis* H. T. Chang[J]. *J Fujian Forestry Sci Tech*, 45(4): 122-127. [Xie SY, Yao KL, Wu XJ, et al., 2018. Overview of pharmacological research on Semiliquidambar cathayensis H. T. Chang[J]. Journal of Fujian Forestry Science and Technology, 45(4): 122-127.]

XU Y, XIAO TW, ZHAO N, et al., 2019. Characterization of the complete plastid genome of an endangered species *Fortunearia sinensis* (Hamamelidaceae)[J]. *Mitochondrial DNA Part B*, 4(1): 1432-1434.

YANG DJ, QIU Q, XU LH, et al., 2020. The complete chloroplast genome sequence of *Altingia excelsa*[J]. *Mitochondrial DNA Part B*, 5(1): 534-535.

YANG L, LIU RH, HE JW, 2019. Rapid analysis of chemical compositions from *Semiliquidambar cathayensis* roots by ultra high-performance liquid chromatography and quadrupole time-of-flight tandem mass spectrometry[J]. *Molecules*, 24(22): 4098.

YE XZ, WEN GW, ZHANG MZ, et al., 2021a. Genetic diversity and genetic structure of a rare and endangered plant *Semiliquidambar cathayensis* Hung T. Chang[J]. *Plant Sci J*, 39(4): 415-423. [Ye XZ, Wen GW, Zhang MZ, et al., 2021a. Genetic diversity and genetic structure of the rare and endangered plant Semiliquidambar cathayensis Hung T. Chang[J]. Plant Science Journal, 39(4): 415-423.]

YE XZ, YANG XJ, WANG MQ, et al., 2020. Analysis of SSR loci in transcriptome of rare and endangered plants of *Semiliquidambar cathayensis*[J]. *Mol Plant Breed*, 18(5): 1585-1592. [Ye XZ, Yang XJ, Wang MQ, et al., 2020. Analysis of SSR loci in the transcriptome of the rare and endangered plant Semiliquidambar cathayensis[J]. Molecular Plant Breeding, 18(5): 1585-1592.]

YE XZ, ZHANG MZ, JIANG YT, et al., 2020a. The complete chloroplast genome of *Altingia chinensis* (Hamamelidaceae)[J]. *Mitochondrial DNA B*, 5(1): 695-696.

YE XZ, ZHANG MZ, LIU YP, et al., 2021b. Analysis on genetic diversity of natural populations of *Semiliquidambar cathayensis* based on SRAP marker[J]. *J Plant Res Environ*, 30(4): 60-68. [Ye XZ, Zhang MZ, Liu YP, et al., 2021b. Analysis of genetic diversity of natural populations of Semiliquidambar cathayensis based on SRAP markers[J]. Journal of Plant Resources and Environment, 30(4): 60-68.]

YE XZ, ZHAO GH, ZHANG MZ, et al., 2020b. Distribution pattern of endangered plant *Semiliquidambar cathayensis* (Hamamelidaceae) in response to

climate change after the last interglacial period[J]. *Forests*, 11(4): 434.

YU JJ, HU GX, ZHAO F, et al., 2019. The complete chloroplast genome sequence of *Disanthus cercidifolius* Subsp. Longipes (Hamamelidaceae)[J]. *Mitochondrial DNA Part B*, 4(1): 1424-1425.

ZHANG MY, WANG XF, GAO J, et al., 2020b. Complete chloroplast genome of *Paeonia mairei* H. Lév.: characterization and phylogeny[J]. *Acta Pharm Sin*, 55(1): 168-176. [Zhang MY, Wang XF, Gao J, et al., 2020b. Complete chloroplast genome of Paeonia mairei H. Lév.: characterization and phylogeny[J]. Acta Pharmaceutica Sinica, 55(1): 168-176.]

ZHANG MZ, JIANG YT, YE XZ, et al., 2020a. The complete chloroplast genome of *Semiliquidambar cathayensis* (Hamamelidaceae)[J]. *Mitochondrial DNA B*, 5(1): 695-696.

ZHANG YY, CAI HX, DONG JX, et al., 2019. The complete chloroplast genome of *Loropetalum subcordatum*, a national key protected species in China[J]. *Conserv Genet Resour*, 11(4): 391-393.

ZHAO DN, REN Y, ZHANG JQ, 2020. Conservation and innovation: plastome evolution during rapid radiation of *Rhodiola* on the Qinghai-Tibetan Plateau[J]. *Mol Phylogenet Evol*, 144: 106671.

ZHEN CH, KONG XY, CHEN GX, et al., 2016. Screening, clustering and response to salinity stress of Clp family genes in Peanut (*Arachis hypogaea* L.)[J]. *Shandong Agric Sci*, 45(12): 1-5. [Zhen CH, Kong XY, Chen GX, et al., 2016. Screening, clustering and response to salinity stress of Clp family genes in peanut (Arachis hypogaea L.)[J]. Shandong Agricultural Sciences, 48(12): 1-5.]

ZHOU T, RUHSAM M, WANG J, et al., 2019. The complete chloroplast genome of *Euphrasia regelii*, pseudogenization of *ndh* genes and the phylogenetic relationships within Orobanchaceae[J]. *Front Genet*, 10: 444.

ZHU SS, YIN PP, YAP ZY, et al., 2019. Chloroplast genomes of two extant species of Tertiary relict *Cercidiphyllum* (Cercidiphyllaceae): comparative phylogenomic analyses[J]. *Mitochondrial DNA Part B*, 4(1): 1551-1552.

Note: Figure translations are in progress. See original paper for figures.