

Assembly and characterization analysis of the complete mitochondrial genome of *Lithocarpus litseifolius* (Hance) Chun

Xiaoyan Qiu

Hunan Agricultural University

Yuqiao Tian

Hunan Agricultural University

Ziqiang Li

Hunan Agricultural University

Xianjin Wu

Huaihua University

Zhuwen Xiang

Hunan Yaocha Engineering Technology Research Center

Yiqing Wang

Hunan Yaocha Engineering Technology Research Center

Jin Li

Shaoyang Academy of Agriculture science

Shengen Xiao

xiaosg@hunau.net

Hunan Agricultural University <https://orcid.org/0000-0003-3584-6359>

Research Article

Keywords: *Lithocarpus litseifolius* (Hance) Chun, Mitochondrial genome, Genome size variation, Repeats, Phylogenetic analysis

Posted Date: August 17th, 2023

DOI: <https://doi.org/10.21203/rs.3.rs-3230084/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Genetic Resources and Crop Evolution on April 20th, 2024. See the published version at <https://doi.org/10.1007/s10722-024-01989-2>.

**Assembly and characterization analysis of the complete mitochondrial
genome of *Lithocarpus litseifolius* (Hance) Chun**

Xiaoyan Qiu^{1, 2, 3}, Yuqiao Tian^{1, 3}, Ziqiang Li^{1, 3}, Xianjin Wu^{2, 3}, Zhuwen Xiang³, Yiqing Wang³,
Jin Li⁴ and Shengen Xiao^{1, 3*}

1 College of Horticulture, Hunan Agriculture University, Changsha 410000, China

2 College of Biology and Food Engineering, HuaHuai University, Huahuai 418000, China

3 Hunan Yaocha Engineering Technology Research Center, Xupu 419300, China

4 Institute of Tea Research, Shaoyang Academy of Agricultural Sciences, Shaoyang 422000,
China

***Corresponding Author: Shengen Xiao**

Tel:+0731-8418171, Fax:+0731-84618171, E-mail: Xiaosg@hunau.net

Abstract

Lithocarpus litseifolius (Hance) Chun (*L. litseifolius* 1837) is an evergreen tree of Fagaceae, commonly known as sweet tea. *L. litseifolius* is a natural sweetener with high levels of dihydrochalcone. In addition, *L. litseifolius* is a precious medicinal material, its phlorizin has a unique role in the treatment of diabetes. This investigation aimed to assemble and scrutinize the entire mitochondrial (mt) genome of *L. litseifolius*. The circular mt genome of *L. litseifolius* spans 573,177 bp and has a GC content of 45.61%. The mt genome of *L. litseifolius* comprises 61 genes, of which 21 are tRNA genes, 3 are rRNA genes, 36 are protein-coding genes (PCGs), and 1 is a pseudogene. Tetramer repeats made up 32.57% of all identified simple repeat sequences (SSRs), making them the most abundant type of SSR. 35 PCGs with a combined length of 32,208 bp were predicted to include a total of 461 RNA editing sites in the *L. litseifolius* mt genome. Besides, nine homologous genes between the chloroplast and mt genomes of *L. litseifolius* were identified. Furthermore, our findings demonstrated that while plant mt genome sizes vary considerably, the GC content of these genomes has remained largely constant. Seven genes were found to be associated with conservatism: *atp6*, *rps1*, *ccmC*, *rpl2*, *nad4*, *nad7*, and *trnY-GTA*. The phylogenetic analysis confirmed that *L. litseifolius* was genetically more clustered with *Quercus variabilis*. This study establishes the groundwork for investigations on the systematic evolution, genetic variability, and breeding of *L. litseifolius*.

Keywords *Lithocarpus litseifolius* (Hance) Chun · Mitochondrial genome · Genome size variation · Repeats · Phylogenetic analysis

Introduction

Lithocarpus litseifolius (Hance) Chun (*L. litseifolius* 1837) (Supplementary Figure 1), an evergreen plant of Fagaceae, commonly referred to as “sweet tea”. It has a long history of being consumed as a daily beverage for thousands (Cheng et al. 2016). *L. litseifolius* is a good source of natural sweetener compounds because it contains a large amount of dihydrochalcone (Cheng et al. 2018). Additionally, the medicinal values of *L. litseifolius* have also been identified in recent years, encompassing antioxidant (Shang et al. 2020), antidiabetic (Wang et al. 2016), anti-inflammatory (Gao et al. 2018), and hepatoprotective effects (Li et al. 2013). *L. litseifolius* is presently regarded as a special natural asset possessing noteworthy commercial, medicinal, and breeding worth.

The primary role of mitochondria within living cells by transforming biomass energy into chemical energy (Cheng et al. 2021). In addition, mitochondria participate in a multitude of life activities, including intracellular growth, division, differentiation, and apoptosis (Lu et al. 2022; Qiao et al. 2022). Mitochondria are believed to have originated from ancient endosymbiotic events and possess relatively autonomous genetic material (Qiao et al. 2022). In the majority of seed plants, mitochondrial DNA is inherited maternally, thereby eliminating the effects of patrilineal inheritance (Ye et al. 2022; Ma et al. 2022). Plant mt genomes exhibit numerous distinctive characteristics in comparison to the relatively conserved and compact mt genomes of animals, particularly a wide range of genome sizes and intricate genomic arrangements (Wu et al. 2022). Currently, there is a significant variation in the dimensions of

completely sequenced plant mt genomes. This variation covers a span of 100 times, with variation extending from 66 kb in *Viscum scurruloideum* (Skippington et al. 2015) to 11,000 kb in *Silene conica* (Sloan et al. 2012). Except for the typical circular structure, certain plant species have linear, and even multichromosomal architectures (Daniel et al. 2013). The factors underlying the size variability of plant mt genomes are different. The principal drivers of significant size differences are attributed to the amplification of repetitive elements, the acquisition of exogenous DNA, and the alteration of large intragenic segments (Wu et al. 2022). plant mitogenomes often use simple sequence repeats (SSRs) as molecular markers to identify species (Bi et al. 2020; Ma et al. 2017). Furthermore, indels and single nucleotide polymorphisms (SNPs) occurring in mitogenomes have been utilized as effective tools for rapid species differentiation and phylogenetic analyses (Seok et al. 2019; Mwamuye et al. 2020). Plant mt genomes are often large but have relatively small gene pools, including 24 core genes and 17 variant genes. The observed phenomenon can be explained by the disappearance or relocation of many genes to the nucleus throughout the evolution of angiosperms. However, the coding sequences of the remaining genes have remained highly stable (Li et al. 2023). Consequently, the mt genome holds significant importance in the fields of species identification, classification, and evolution.

The intricate characteristics of plant mt genomes, specifically, the abundance of repetitive sequences and frequent recombination events, pose challenges to achieving precise genome assembly (Wu et al. 2022), as a result, the assembly of plant mt genomes lags significantly behind that of animal mt genomes and plant plastid genomes (Lai et al. 2022; Hong et al. 2021). The read lengths of Illumina sequencing often fail to cover larger repetitive regions, resulting in the incomplete assembly of these areas (Wu et al. 2022). Third-generation sequencing technologies (TGS), like Oxford Nanopore and PacBio sequencing, produce individual reads ranging from 10 to 100 kb or even longer (Michael et al. 2020), which holds promise for enhancing coverage and enabling scaffolding across previously uncharted genomic areas (Li et al. 2023). Because nuclear and chloroplast DNA substantially contaminate the mt DNA readings, employing whole genomic DNA still remains problematic (Sloan and Wu. 2014). There is yet no convincing method that can be used to examine the whole structural information of plant mt genomes, assembling plant mt genomes is still a very challenging task. It was recently revealed that the pipeline SAG-BAC used a graph-based approach to construct plant mt genomes (Fischer et al. 2022). However, these algorithms still have a propensity to produce one or more circular sequences, making it impossible to confirm the mitogenome's whole structural information (He et al. 2023). Graph-based sequence assembly toolkit (GSAT) was recently created with the purpose of assembling and generating superior mt master graphs (MMGs) that accurately depict the diverse range of structural conformations observed in plant mt genomes. By utilizing high-throughput sequencing reads, both short and long, a well-supported MMG for two plant species served as models: rice (*Oryza sativa*) and thale cress (*Arabidopsis thaliana*) (He et al. 2023). The continuous progress in technology has led to a growing number of published plant mt genomes.

There are currently no reports on the *L. litseifolius* mt genome. Herein, the entire mt genome of *L. litseifolius* was sequenced and assembled, and the repeated sequences, codon preference, RNA editing, gene movement between chloroplast and mt genomes, and phylogenetic relationships were conducted

through analysis. It is anticipated that our results will serve as a theoretical basis for biological research and species identification, and will be crucial for ascertaining the provenance and evolutionary affiliations of various species.

Materials and methods

Plant materials, DNA extraction and genome sequencing

L. litseifolius plant was provided by Hunan Yaocha Engineering Research Center (Huaihua, China; 25°52'N, 111°08' E), the plant was placed in obscurity for 14 days in advance to get etiolated *L. litseifolius* seedlings. The leaf materials were collected to extract DNA by the hipure universal DNA kit (DP305-03, Tiangen Biotech, Beijing, China). Qubit 3.0 (Q33216, Invitrogen, Singapore) accurately quantified the concentration of DNA, in which DNA samples with a content of more than 1.5 micrograms were used for sequencing. The library was sequenced and constructed using the Oxford Nanopore PromethION sequencing device and Illumina Novaseq 6000 platform. Using Fastp (v0.20.0, <https://github.com/OpenGene/fastp>), the raw data were filtered to remove sequencing junctions and primer sequences, Sequence data from three separate sources were counted and filtered using the Perl script filling (v0.2.1, <https://link.zhihu.com/?target=https%3A//github.com/rrwick/Filtlong>).

Mitogenome assembly and annotation

The initial third-generation data were compared with the reference gene sequence (plant mt core gene) using the third-generation comparison software Minimap2 (v2.1) (Li 2018), the sequences with longer alignment lengths than 50 bp were selected as candidate sequences in the alignment. The sequence with the higher alignment quality (covering more complete core genes) and more aligned genes (one sequencing sequence encompasses multiple core genes) were chosen as the seed sequence. The third-generation assembly program canu (v1.4) was used to correct the third-generation data that were acquired (Koren et al. 2017), and bowtie2 (v2.3.5.1) was used to align the second-generation data to the corrected sequence. Using the default Unicycler (v0.4.8) parameter, the above-mentioned data of the second generation and the revised data of the third generation were then compared for concatenation. Finally, the *L. litseifolius* mt genome was obtained.

The steps for mt genome structure annotation are as follows: the Blast (v2.6, <https://blast.ncbi.nlm.nih.gov/Blast.cgi>) was used to annotate coding protein and rRNA, and the tRNA was annotated using the tRNA-scan-SE (<http://lowelab.ucsc.edu/tRNA-scan-SE/>) (Chan and Lowe 2019). The shortest length was fixed to 102 bp when ORF was annotated using Open Reading Frame Finder (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>), and redundant sequences and sequences that overlapped with known genes were ignored. Over 300 bp long sequence alignments are annotated in the NR collection. The aforementioned results were verified and manually revised to produce more precise

annotation results. Then, using OGDRAW (<https://chlorobox.mpimp-golm.mpg.de/OGDraw.html>), the mt genome was mapped, and the complete mt genome sequences of *L. litseifolius* were registered into NCBI with an accession number ON462106.

Analysis of repeated sequences

In the *L. litseifolius* mt genome, simple, tandem, and dispersed repeat sequences were discovered. Using MicroSATellite (MISA) identification tool Perl script (v1.0, parameter: 1-10 2-5 3-4 4-3 5-3 6-3), simple repeat patterns were detected (Thiel et al. 2003). Tandem repeat sequences (>6 bp repeat units) were discovered using the default settings of the Tandem Repeats Finder (v4.09, <http://tandem.bu.edu/trf/trf.submit.options.html>) program (Benson 1999). The blast (v2.10.1, parameters: -word size 7, value 1e-5, remove redundancy, remove tandem duplication) was used to detect dispersed repeat sequences. The circos (v0.69-5) was used to visualize these repeat sequences.

Analysis of homologous fragments of mitochondria and chloroplasts, and RNA editing analyses

From the NCBI Genome Resources Database, the chloroplast genome of *L. litseifolius* (OM048987) was obtained. The blast software (v2.6, <https://blast.ncbi.nlm.nih.gov/Blast.Cgi>) was employed with optimized parameters to achieve a 70% success rate of pairing, and the homologous genes of chloroplasts and mt genome were found. The Plant Predictive RNA Editor tool (<http://prep.unl.edu/>) was used to forecast the RNA editing events.

Analysis of codon composition

The effective number of codons (ENC) and the relative synonymous codon usage (RSCU) were computed utilizing the cloud platform (cloud.genepioneer.com.). The impact of genetic composition on codon usage preference was examined using ENC. Four mt genomes from the family of Fagaceae, including *L. litseifolius* (ON462106), *Quercus acuisissima* (MZ636519), *Quercus variabilis* (MN199236), and *Fagus sylvatica* (NC_050960), calculate their codon preference parameters, $ENC = 2 + GC3 + 29/(GC3^2 + (1-GC3)^2)$.

Ka/Ks analysis and phylogenetic tree construction

Homologous protein sequences between *L. litseifolius* (ON462106), *Q. acuisissima* (MZ636519), *Q. variabilis* (MN199236), *F. sylvatica* (NC_050960), and *Arabidopsis thaliana* (JF729201) were obtained using Blastn. The common protein-coding genes were then aligned using mafft (v7.427, <https://mafft.cbrc.jp/alignment/software/>). Using the Ka/Ks_Calculator (v 2.0, <https://sourceforge.net/projects/kakscalculator2/>), the non-synonymous (Ka) and synonymous (Ks) rates (Ka/Ks) were determined.

In the phylogenetic analysis, *L. litseifolius* and 19 other distinct species with complete mt genomes from a variety of families were used. The multiple sequence alignment between species was performed

using mafft (v7.427, <https://mafft.cbrc.jp/alignment/software/>), while Trimal (v1.4.rev15) was employed to concatenate the aligned sequences in a head-to-tail manner and subsequently prune them. The coding DNA sequences (CDS) were utilized to draw the evolutionary tree with maximum likelihood. RAxML (v8.2.10, <https://cme.h-its.org/exelixis/software.html>) estimated the GTRGAMM A model for the maximum likelihood phylogenetic tree using 1000 bootstrap replications. The jModelTest (v2.1.10, <https://github.com/ddarriba/jmodeltest2>) was used to calculate the optimal nucleotide substitution model, and the Bayesian inference phylogenetic tree was established by MrBayes (v3.2.7a, <http://nbiweden.github.io/MrBayes/>).

Genomic comparison of related species

The mt genome of *L. litseifolius* was compared to three other mt genomes of four Fagaceae plant species: *Q. acuisissima* (MZ636519), *Q. variabilis* (MN199236), and *F. sylvatica* (NC_050960). To find covariance regions, whole-genome covariance comparisons were made with the Lastz (v1.02.00). The following guidelines were established: the step is 20, and the seed pattern is 12 out of 19.

Results

Genomic features of the *L. litseifolius* mt genome

Table S1 displays the output of the second-generation and third-generation sequencing platforms. The former provided 24,930,314 clean reads, whereas the latter delivered 1,109,068 clean reads.

The mt genome of *L. litseifolius* exhibits a circular architecture spanning 573,177 bp (Fig. 1). The nucleotide composition of the genome is as follows: 27.33% A, 22.94% C, 27.06% T, and 22.67% G, resulting in a GC content of 45.61% (Table 1). The *L. litseifolius* mt genome contains 61 genes, including 36 PCGs, 21 tRNA genes, 3 rRNA genes, and 1 pseudogene. Among the 61 annotated genes, 10 genes contained introns, *ccmFC*, *cox2*, *trnI*-GAT, and *trnT*-TGT had a single intron, the gene *rpl2* was found to possess two introns, while the gene *nad4* exhibited three introns, additionally, the genes *nad1*, *nad2*, *nad5*, and *nad7* were observed to contain four introns each. The core genes could be classified into 10 distinct categories: nine NADH dehydrogenase genes (*nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *nad7*, and *nad9*), six ATP synthase genes (*atp1*(2), *atp4*, *atp6*, *atp8*, and *atp9*), five ribosomal proteins genes (*rps1*, *rps4*, *rps10*, *rps12*, and *rps19*), four cytochrome c biogenesis genes (*ccmB*, *ccmFC*, *ccmC*, and *ccmFN*), four ribosomal proteins genes (*rpl2*, *rpl5*, *rpl10*, and *rpl16*), three cytochrome c oxidase genes (*cox3*, *cox2*, and *cox1*), two succinate dehydrogenase genes (*sdh3* and *sdh4*), one ubiquinol cytochrome c reductase gene (*cob*), one maturase gene (*matR*), one transport membrane protein gene (*mttB*) (Table 2).

Repeat sequences analysis of *L. litseifolius* mt genome

Simple repeat sequences (SSRs), refer to a segment of DNA in the genome that is repeated numerous times and has a length ranging from 1-6 bp (Li et al. 2021). As shown in Table 3, 175 SSRs were

identified in the *L. litseifolius* mt genome, the detected SSRs included mononucleotide, dinucleotide, trinucleotide, tetranucleotide, and pentanucleotide repeats. Tetra and pentamer repeats were the most abundant and the least SSR type, constituting 32.57% and 5.14% of all identified SSRs, respectively. There were 51 monomer repeats (29.14%), 40 dimer repeats (22.86%) and 18 trimer repeats (10.29%) in all identified SSRs. The most frequently occurring motifs were A/T motifs (38), which made up 21.71% of all SSRs that were identified.

Tandem repeat sequences are characterized by the repetition of contiguous short sequences with a length of 1 to 200 bases (Paco et al. 2019). In the *L. litseifolius* mt genome, 8 tandem repeats were identified, with lengths varied between 13 and 21 bp and a matching degree of more than 95%, as presented in Table S2. As shown in Fig. 2, in the mt genome of *L. litseifolius*, 515 interspersed repeats, each with a length of 30 bp or more, were identified. Among these repeats, 271 (52.62%) were found to be in the forward orientation, while 244 (47.38%) exhibited palindromic characteristics. A majority of interspersed repeats (423 repeats, 82.14%) were 30-50 bp in length. The maximum length of a palindromic repeat was observed to be 118.1 bp, whereas the longest forward repeat measured 456.0 bp.

The prediction of RNA editing

In the mt genome of *L. litseifolius*, a sum of 461 RNA editing sites within 35 PCGs was predicted, involving 30 codon transitions (Fig. 3). Among them, the predicted values of editing sites of *atp1*, *atp8*, *rps19*, *sdh3*, and *sdh4* were the lowest (3), while those of *ccmFn* and *nad4* were the highest (38). The editing sites of 30.59% (141) and 66.38% (306) were located in the first and second positions of the triplet code, respectively. And in one particular editing instance, the first and second locations of the triplet codes were changed, causing the original proline (CCT) amino acid to be replaced with lysine (TTT). The hydrophobicity of 43.82% of amino acids was found to be unaltered after RNA editing, nevertheless, it was expected that 8.03% of the amino acids would turn hydrophilic while 47.72% of them would become hydrophobic (Table 4). In addition to changing the encoded amino acids, RNA editing can also cause early termination of the coding procedure. The above phenomenon occurred in coding genes *ccmFc* and *atp6* in the *L. litseifolius* mt genome. The predicted outcomes also demonstrate that after RNA editing, 45.99% (212 sites) of the amino acids underwent conversion to leucine, exhibiting the highest conversion tendency, additionally, 22.99% (106 sites) of the amino acids underwent conversion to phenylalanine, with the second-highest number.

Codon usage analysis of PCGs

The total length of *L. litseifolius*'s PCGs is 32,208 bp. ATG served as the start codon for all PCGs, while three sorts of stop codons were recognized, namely, TGA (36.11%), TAG (19.44%), and TAA (44.45%) (Table S3). The condon utilization examination uncovered that the most frequently occurring amino acids were Leucine (Leu) (10.58%-11.16%), Serine (Ser) (8.67%-9.32%), and Isoleucine (Ile) (7.63%-7.70%) (Fig. 4). In contrast, Tryptophan (Trp) (1.45%-1.63%) and Cysteine (Cys) (1.46%-1.54%) were identified

only frequently. Additionally, we analyzed 10,226 codons out of 36 PCGs of the *L. litseifolius* mt genome, and 6717 codons (65.69%) had RSCU values over 1.0, which shows that these codons are utilized more frequently than other synonymous codons (Table S4). Four mt genomes from Fagaceae plants contain 5768 and 6717 codons exhibiting an RSCU value exceeding 1.0. The condons GCU (Ala) and UAU (Tyr) were found to be the most prevalent among the four species, with an RSCU value greater than 1.5, whereas CUG (Met) and UUG (Met) were the least codons (RSCU<0.5). According to Table S5, there were 10,736 codons in all of the coding genes. Furthermore, the average GC content of the three-base codons in the *L. litseifolius* mt genome was lower than 50%, showing that the genome's codons were biased because both A and T bases were used. The ENC values of the mt genomes of four Fagaceae plants were greater than 35, the distribution range of GC3s is between 0.297 and 0.597, and ENC values are between 36.90 and 58.44. The results showed that the impact of mutation pressure on the codon usage preference of the mt genome in four species of Fagaceae species was relatively low. Most of the scattered points of the mitochondrial coding genes of the four Fagaceae plants deviated from the standard curve, indicating that the impact of selection pressure on codon preference is more important (Fig. 5).

Chloroplast to Mitochondrion DNA Transformation

As shown in Table 5 and Fig. 6, a total of 28 homologous fragments, spanning a combined length of 19,904 bp, were identified, making up 2.08% of the *L. litseifolius* mt genome. Four of them were longer than 1,000 bp, and fragments 1 and 2 were the longest, both of which were 6,080 bp. The following 9 mt genes were detected: *trnI*-GAT, *trnV*-GAC, *trnP*-TGG, *trnW*-CCA, *trnH*-GTG, *trnN*-GTT, *trnD*-GTC, *trnM*-CAT, and *rrn18*.

Analysis of synonymous and nonsynonymous replacement rates

The Ka/Ks ratio is very important in evolutionary analysis as it serves as a determinant of whether the PCG is subject to selection pressure. It is generally believed that a Ka/Ks ratio higher than one has a forward selection effect, a ratio of Ka/Ks equal to one has a neutral selection effect, and a ratio of Ka/Ks less than one has an effect of purification and selection. In this investigation, as depicted in Fig. 7, the Ka/Ks ratio of 35 PCGs in the mt genomes of *L. litseifolius*, *Q. acuisissima*, *Q. variabilis*, *F. sylvatica*, and *A. thaliana* was analyzed. The Ka/Ks ratio of 26 PCGs between *L. litseifolius* and *Q. variabilis* (MN199236) was zero. For the 25 shared PCGs between *L. litseifolius* and *Q. acuisissima*, the Ka/Ks ratio was also zero. The above results show that the PCGs shared between *L. litseifolius* and *Q. variabilis*, *Q. acuisissima* were close homologs. Additionally, the observation that the vast majority of Ka/Ks values were below 1.0 implies that most PCGs underwent stabilizing selection during their evolutionary history. In contrast, the Ka/Ks proportions of the three genes (*atp4*, *ccmB*, and *nad1*) were more prominent than 1.0, indicating that these genes had encountered positive choices all through advancement.

Comparison of the mt genome size and GC content between *L. litseifolius* and other species

The size and GC content of plant mt genome are their main features. In this investigation, we performed a comparative analysis of the mt genome size and GC content among 19 plant species, including 3 species each from Fagaceae, Asteraceae, Cruciferae, and Fabaceae families, 2 species from Theaceae family, 3 species from Poaceae family, and 1 species each from Grimmiaceae and Ginkgoaceae families, with that of the *L. litseifolius* mt genome (Table S6). These plant mt genomes range widely in size, from 107,186 bp (*Racomitrium emersum*) to 1,081,966 bp (*Camellia duntsa*), while the distinction in GC content of mt genomes was somewhat little (typically about 45%) (Fig. 8).

Mitochondrial genome comparison in Four Fagaceae plants

The mt genome of *L. litseifolius* was utilized as a reference to assess the results of a whole-genome correlation analysis conducted on four distinct Fagaceae plants, as depicted in Fig. 9. The mt genome's gene region of the related species showed a greater degree of resemblance to that of *L. litseifolius*, the conserved genes were *atp6*, *rps1*, *ccmC*, *rpl2*, *nad4*, *nad7*, and *trnY-GTA* (Table S7). The mt genomes of *L. litseifolius* and *Q. variabilis* were more similar to one another than those of the other two Fagaceae species.

Duplication and loss of mt genome in Fagaceae

With rapidly evolving sequencing technology, the complete plant mt genome has been assembled and reported, making it possible to compare mt genome characteristics among different plant species. Four mt genomes from the family of Fagaceae can now be accessed: *L. litseifolius*, *Q. acuissima*, *Q. variabilis*, and *F. sylvatica* (Table S8). There are a total of 36, 32, 36, and 36 PCGs in each of these four mt genomes, respectively. Different species have shown both gene duplication and gene loss. For example, *atp1* was replicated in the *L. litseifolius* mt genome, *atp9* was replicated in the *Q. acuissima* mt genome, and *atp1* was replicated in the *Q. variabilis*. The *rps14* and *rps3* were lost in the *L. litseifolius* mt genome, *rps14* and *rps4* from *Q. acuissima* being lost, *mttB*, *nad1*, *nad4L*, *rps16*, *rps1*, *rps14*, *rps4*, from *Q. variabilis* being lost, *rpl2* and *rps10* from *F. sylvatica* being lost.

Phylogenetic analysis

The *L. litseifolius* mt genome and the other 19 plant mt genomes were downloaded from GenBank to conduct phylogenetic trees by utilizing a fusion of the maximum likelihood and bayesian methodologies (Table S6). The result showed that Poaceae, Theaceae, Asteraceae, Fagaceae, Cruciferae, and Fabaceae were well-clustered (Fig. 10). Both analytical methodologies yield consistent outcomes regarding the phylogenetic tree, and the clustering of the phylogenetic tree conforms to the family and genus affiliations of these species, proving the validity of the genome-based clustering findings. *L. litseifolius*, which is closely linked to *Q. acutissima*, belonged to the family Fagaceae of the order Fagales.

Discussion

Generally, the mt genome of plants has many unique characteristics, including complex composition, a broad distribution in genome size, and so on (Zhou et al. 2022). The key characteristics of the *L. litseifolius* mt genome are described in this study. Many mt genomes have circular structures, but some are linear, and even multichromosomal architectures, for example, the mt genome of *Polytomella parva* is linear (Daniel et al. 2013). The published *L. litseifolius* mt genome here is circular with 573,177 bp and 45.61% GC content. The GC content was determined to be similar to that observed in previously sequenced plant mt genomes, for instance, *Camellia sinensis* var. *Assamica* cv. *Duntsa*, 45.62% (Li et al. 2023); *Acer truncatum*, 45.68% (Ma et al. 2022).

The mt genome contains numerous repeat sequences, including two main categories: interspersed and tandem repeats (Gualberto et al. 2014). Repeated sequences are frequently crucial to intermolecular recombination of mitochondria, and usually, the biggest repeats inside a species (frequently more than 1 kb in angiosperms) are discovered to undergo structural recombination, resulting in isomerization (Wynn et al. 2018; Guo et al. 2016). Numerous repetitive sequences have been found in the *L. litseifolius* mt genome, and the longest interspersed repeat sequence exceeded 1 kb (4,560 bp), which might be evidence of heterodimerization. In addition, the analysis shows that the number and sequence length of SSR in the *L. litseifolius* mt genome is out of proportion between the gene-coding region and the non-coding region, which indicates that the distribution of SSR loci is uneven.

The sizes of the mt genomes for the four Fagaceae plants ranged from 412,886 to 573,177 bp. The assessment of phylogenetic relationships and the rate of species change within Fagaceae plant species could potentially be achieved by analyzing the shape and dimensions of the mt genome. Notably, closely related species such as *R. pulchrum* and *R. simsii* exhibit quite similarities in the size and structure of their mt genomes (Xu et al. 2021). Therefore, the mt genome could serve as a valuable resource for comprehending plant evolution and establishing taxonomic classifications (Li et al. 2023).

L. litseifolius has thirty-six PCGs, twenty-one tRNAs, three rRNAs, and one pseudogene in its mt genome, with the longest mt genome among all Fagaceae plant species identified to date. Comparative analysis of mt genomes from four distinct Fagaceae plant species revealed the presence of two copies of the *atp1* gene in both *L. litseifolius* and *Q. variabilis*, and gene *atp9* has two copies in *Q. acutissima*. Additionally, the *L. litseifolius* mt genome had more annotated genes than the *Q. variabilis*. They were *mttB*, *nad1*, *nad4L*, *rpl16*, *rps1*, *rps19*, and *rps4*. These findings highlight the importance of plant mt genomes in future research, as they provide valuable insights into the evolution of genomes that have not been previously explored.

Codon usage bias is often used in phylogenetic and evolutionary analyses of mammals and insects, but less research has been done on plants. The range of ENC values was 20-61, values closer to 20 were indicative of a stronger preference for specific codons, while values closer to 61 indicated a weaker preference (Ikemura 1981). The codon of the *L. litseifolius* mt genome exhibited an ENC value of 53.64, indicating a low degree of codon preference in this organism. Four Fagaceae plants' mt genomes had ENC values greater than 35, suggesting that natural or artificial selection may have changed the codon usage bias of those plants' genes, but not mutation (Li et al. 2023). These findings suggested that mt genome ENC analysis may be useful for tracing the evolutionary history of Fagaceae plant species.

The phenomenon of genetic material exchange, specifically in the form of DNA, occurring between nuclear genomes and plant organelles, and between different species, is a common occurrence (Nguyen et al. 2020). Sequencing analysis has facilitated the identification of multiple occurrences of DNA transfer events among diverse genomes, encompassing mitochondrial (mt), nuclear, and chloroplast genomes, across a wide range of plant species (Qiao et al. 2022). In most plants, the chloroplast DNA content in the mt genomes ranges from 3% to 6%, and can sometimes reach up to 10% (Adams et al. 2002). In the case of the *L. litseifolius* mt genome, a total of 28 chloroplast insertions were identified, with lengths ranging from 29 to 6,080 bp, and a cumulative length of 11.90 kb, which accounts for 2.08% of the total length of the genome. The ratio of transferred fragments observed exhibits similarity to the earlier research results in *Acer truncatum* (2.36%) (Ma et al. 2022), however, it is inferior to that of *Vitis vinifera* (8.8%) (Yin et al. 2021). In this study, 9 mt genes were annotated, with 8 tRNA genes, and 1 rRNA gene. The occurrence of tRNA gene transfer from chloroplast to mt DNA is prevalent among angiosperms (Bi et al. 2016), and the tRNA gene plays an indispensable role in the mt genome. The chloroplast genome is very conservative in higher plants, and it is rarely reported that the fragments in the mitochondria go into the chloroplast. Furthermore, the interplay between plastomes and mitogenomes in plants has conventionally been perceived as a unidirectional mechanism, wherein plastomes are transferred to the mitogenome (Yue et al. 2022). However, several studies, such as those on *Daucus carota* (Iorizzo et al. 2012) and *Asclepias syriaca* (Straub et al. 2013), have documented the transfer of mt sequences to the plastomes. But in *L. litseifolius*, there is currently no empirical evidence indicating the occurrence of mt sequence transfer to the chloroplast genome.

The outcomes of the mt phylogenetic investigation of four Fagaceae botanical species and sixteen additional botanical species were in accordance with the taxonomic categorization of said species, indicating the feasibility of utilizing organelle genomes in plant phylogenetic investigations (Wu et al. 2012). Classification trees in this study demonstrate that the affinity between *L. litseifolius* and *Q. acutissima* is closer.

In this investigation, the *L. litseifolius* mt genome served as a reference. Through a match analysis of the whole-genome correlation, it was observed that three species of Fagaceae plants exhibit a higher degree of genomic similarity with *L. litseifolius* compared to interval sections. This finding suggests that the regions of the mt genome that encompass genes are more conserved among Fagaceae plant species, in contrast to the regions containing intervals (Li et al. 2023). Seven conserved genes were discovered in this study: *atp6*, *rps1*, *ccmC*, *rpl2*, *nad4*, *nad7*, and *trnY-GTA*. These genes will aid in the classification of Fagaceae plant species and further understanding of evolution.

In different species, there are variable numbers of RNA-editing sites. In the mt genome of *L. litseifolius*, a sum of 461 RNA editing sites within 35 PCGs was predicted, which is lower compared to those found in gymnosperms with larger mt genomes, such as *Ginkgo biloba* (1,306) (Kan et al. 2020), and more than those in *Suaeda glauca* (261) (Cheng et al. 2021). The frequency of RNA editing events located at the first codon position is approximately 30.59% (141), less than that at the second codon position (306; 66.38%). This is in accordance with the outcomes of prior studies (Cheng et al. 2021; Robles and Quesada 2021). The absence of an editing site located at the third position of triplet codons in the mt genome of *L. litseifolius* is in accordance with the infrequent occurrence of RNA editing sites in

the mt genomes of plants (Cheng et al. 2021; Verhage 2020). The mt genome of *L. litseifolius* contains numerous RNA editing sites, with a limited number of editing types. Specifically, there were only 30 types of codon, corresponding to 20 types of amino acid transfers, among the 461 RNA editing sites. These transfer types are comparatively smaller than those observed in monocotyledonous and dicotyledonous plants, which typically exhibit 50-60 codons and approximately 30 amino acids (Verhage 2020; Pinard et al. 2019; Alverson et al. 2010). However, they are similar to those observed in most gymnosperms, which typically exhibit 30-40 codons and around 20 amino acids. These findings are in agreement with prior research (Kan et al. 2020; Edera and Sanchez-Puerta 2021). In this study, following RNA editing, it was observed that the hydrophobicity of 43.82% of the amino acids remained unaltered, and it is expected that the hydrophobicity of other amino acids will change, which enables the protein to better fold and function. 30 codons are edited into new starting codons by RNA, which makes the encoded proteins more conservative and more homologous to the corresponding proteins encoded by other plants, thus helping genes to be expressed more completely.

The mt genomes of *L. litseifolius*, *Q. acuisissima*, *Q. variabilis*, *F. sylvatica*, and *A. thaliana* were subjected to Ka/Ks analysis. The results indicated that negative selection dominated the evolution of most genes, suggesting that the PCGs of the *L. litseifolius* mt genome have been conserved. However, PCGs such as *atp4*, *ccmB*, and *nad1* exhibited Ka/Ks values > 1, indicating that the evolutionary history of these genes has been influenced by positive selection. The identification of genes with high Ka/Ks ratios is a crucial aspect of gene selection and evolutionary research within the Fagaceae family. Additionally, we investigated the GC content present in the mt genomes of *L. litseifolius* and nineteen other plant species. Our results support the existing theory that higher plants' GC content maintains a high level of constancy throughout time (Cheng et al. 2021).

Conclusions

The entire mt genome of *L. litseifolius* was assembled and characterized in this investigation. The mt genome of *L. litseifolius* exhibits a circular structure spanning 573,177 bp. The nucleotide composition of the genome is as follows: 27.33% A, 22.94% C, 27.06% T, and 22.67% G, resulting in a GC content of 45.61%. Sixty-one genes in the mt genome of *L. litseifolius* have been annotated, including 36 PCGs, 21 tRNA genes, 3 rRNA genes, and 1 pseudogene. Then, the repeat sequences, RNA editing, codon usage, and chloroplast to mitochondrion DNA transformation were analyzed. The Ka/Ks analysis revealed that a significant proportion of genes experienced negative selection during the course of evolution, suggesting that the PCGs within the mt genome of *L. litseifolius* are conserved overall. In addition, our findings also demonstrate that despite significant variations in the magnitude of plant mt genomes, their GC content has remained relatively stable throughout the course of development. Seven genes were found to be associated with conservatism: *atp6*, *rps1*, *ccmC*, *rpl2*, *nad4*, *nad7*, and *trnY-GTA*. The phylogenetic analysis confirmed that *L. litseifolius* was genetically more clustered with *Q. variabilis*. Overall, This research offers extensive information about the *L. litseifolius* mt genome. Further study on breeding, genetic variation, and systematic evolution of *L. litseifolius* using the mt genome would benefit from the

findings.

Author Contributions X.Q. and S.X. conceived and designed the study. X.Q. and Y.T. performed the experiments and wrote the paper. X.W., Y.W., and Z.X. contributed to resource investigation and sample collection. Z.L. and J.L. analyzed and interpreted the data.

Funding This investigation received financial backing from the bioengineering Hunan provincial key laboratory of "double first-class" applied characteristic disciplines and ethnic medicinal plant resources in Hunan colleges and universities [2022SWG05], the key project of Huaihua "Study on selection of excellent germplasm resources and key techniques of deep processing of Xupu Yao cha" [2022], Furong Plan "the key technology research and development of *L. litseifolius* and industrialization innovation and entrepreneurship team" [2022], the provincial natural science foundation [2022JJ50044], the technical regulations for dwarfing cultivation of *L. litseifolius*" Xiang Caihang Index [2022] 91", Hunan province traditional Chinese medicine industry technology system project [2022] 67.

Data Availability Statement The *L. litseifolius* genomic sequences have been stored in the GenBank database (accession number ON462106, <https://www.ncbi.nlm.nih.gov/nuccore/ON462106>). SRA accession number is SRR17629386.

Declarations

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare no conflict of interest.

References

- Adams KL, Qiu YL, Stoutemyer M et al (2002) Punctuated evolution of mitochondrial gene content: high and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. *Proc Natl Acad Sci USA* 99: 9905–9912. <https://doi.org/10.1073/pnas.042694899>
- Alverson AJ, Wei X, Rice DW et al (2010) Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). *Mol Biol Evol* 27: 1436–1448. <https://doi.org/10.1093/molbev/msq029>
- Bi C, Lu N, Xu Y et al (2020) Characterization and analysis of the mitochondrial genome of common bean (*Phaseolus vulgaris*) by comparative genomic approaches. *Int J Mol Sci* 21: 3778. <https://doi.org/10.3390/ijms21113778>

- Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* 27: 573–580. <https://doi.org/10.1093/nar/27.2.573>
- Cheng J, Lyu LS, Shen YB et al (2016) Population structure and genetic diversity of *Lithocarpus litseifolius* (Fagaceae) assessed using microsatellite markers. *Nord J Bot* 000: 001–009. <https://doi.org/10.1111/njb.01111>
- Cheng YB, Liu FJ, Wang CH et al (2018) Bioactive triterpenoids from the leaves and twigs of *Lithocarpus litseifolius* and *L. corneus*. *Planta. Med* 84: 49–58. <https://doi.org/10.1055/s-0043-113826>
- Chan PP, Lowe TM (2019) tRNAscan-SE: Searching for tRNA Genes in Genomic Sequences. *Methods Mol Biol* 1962: 1–14. https://doi.org/10.1007/978-1-4939-9173-0_1
- Cheng Y, He XX, Priyadarshani SVGN et al (2021) Assembly and comparative analysis of the complete mitochondrial genome of *Suaeda glauca*. *BMC Genom* 22: 167. <https://doi.org/10.1186/s12864-021-07490-9>
- Daniel BS (2013) One ring to rule them all? Genome sequencing provides new insights into the ‘master circle’ model of plant mitochondrial DNA structure. *New Phytol* 200: 978–985. <https://doi.org/10.1111/nph.12395>
- Edera AA, Sanchez-Puerta MV (2021) Computational detection of plant RNA editing events. *Methods Mol Biol* 2181: 13–34. https://doi.org/10.1007/978-1-0716-0787-9_2
- Fischer A, Dotzek J, Walther D et al (2022) Graph-based models of the *Oenothera* mitochondrial genome capture the enormous complexity of higher plant mitochondrial DNA organization. *NAR Genom Bioinform* 4: lqac027. <https://doi.org/10.1093/nargab/lqac027>
- Gao JM, Liu S, Xu F et al (2018) Trilobatin protects against oxidative injury in neuronal PC12 Cells through regulating mitochondrial ROS homeostasis mediated by AMPK/Nrf2/Sirt3 signaling pathway. *Front Mol Neurosci* 11: 1–16. <https://doi.org/10.3389/fnmol.2018.00267>
- Guo W, Grewe F, Fan W (2016) Ginkgo and Welwitschia Mitogenomes reveal extreme contrasts in gymnosperm mitochondrial evolution. *Mol Biol Evol* 33: 1448–1460. <https://doi.org/10.1093/molbev/msw024>
- Hong Z, Liao XZ, Ye YJ et al (2021) A complete mitochondrial genome for fragrant Chinese rosewood (*Dalbergia odorifera*, Fabaceae) with comparative analyses of genome structure and intergenomic sequence transfers. *BMC Genom* 22: 672. <https://doi.org/10.1186/s12864-021-07967-7>
- He WH, Xiang KL, Chen CJ et al (2023) Master graph: an essential integrated assembly model for the plant mitogenome based on a graph-based framework. *Brief Bioinform* 24: 1–13. <https://doi.org/10.1093/bib/bbac522>
- Ikemura T (1981) Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. *J Mol Biol* 151: 389–409. [https://doi.org/10.1016/0022-2836\(81\)90003-6](https://doi.org/10.1016/0022-2836(81)90003-6)
- Iorizzo M, Senalik D, Szklarczyk M et al (2012) De novo assembly of the carrot mitochondrial genome using next generation sequencing of whole genomic DNA provides first evidence of DNA transfer

into an angiosperm plastid genome. BMC Plant Biol 12: 61. <https://doi.org/10.1186/1471-2229-12-61>

Koren S, Walenz BP, Berlin K et al (2017) Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. Genome Res 27: 722–736. <https://doi.org/10.1101/gr.215087.116>

Kan SL, Shen TT, Gong P et al (2020) The complete mitochondrial genome of *Taxus cuspidata* (Taxaceae): eight protein-coding genes have transferred to the nuclear genome. BMC Evol Biol 20: 10. <https://doi.org/10.1186/s12862-020-1582-1>

Li SH, Zeng JY, Tan J et al (2013) Antioxidant and hepatoprotective effects of *Lithocarpus polystachyus* against carbon tetrachloride-induced injuries in rat. Bangl J Pharmacol 8: 420–427. <https://doi.org/10.3329/BJP.V8I4.16458>

Lu C, Gao LZ, Zhang QJ et al (2022) A High-Quality Genome Assembly of the Mitochondrial Genome of the Oil-Tea Tree *Camellia gigantocarpa* (Theaceae). Diversity 14: 850. <https://doi.org/10.3390/d14100850>

Li J, Tang H, Luo H et al (2023) Complete mitochondrial genome assembly and comparison of *Camellia sinensis* var. *Assamica* cv. *Duntsa*. Front Plant Sci 14: 1117002. <https://doi.org/10.3389/fpls.2023.1117002>

Lai CJ, Wang J, Kan SL et al (2022) Comparative analysis of mitochondrial genomes of *Broussonetia* spp. (Moraceae) reveals heterogeneity in structure, synteny, intercellular gene transfer, and RNA editing. Front Plant Sci 13: 1052151. <https://doi.org/10.3389/fpls.2022.1052151>

Li H (2018) Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34: 3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>

Li QY, Su XJ, Ma HH et al (2021) Development of genic SSR marker resources from RNA-seq data in *Camellia japonica* and their application in the genus. Camellia Sci Rep 11: 9919. <https://doi.org/10.1038/s41598-021-89350-w>

Ma QY, Wang YX, Li SS et al (2022) Assembly and comparative analysis of the first complete mitochondrial genome of *Acer truncatum* Bunge: a woody oil-tree species producing nervonic acid. BMC Plant Biol 22: 29. <https://doi.org/10.1186/s12870-021-03416-5>

Ma Q, Li S, Bi C et al (2017) Complete chloroplast genome sequence of a major economic species, *Ziziphus jujuba* (Rhamnaceae). Curr Genet 63: 117–129. <https://doi.org/10.1007/s00294-016-0612-4>

Mwamuye M, Obara I, Khawla E et al (2020) Unique mitochondrial single nucleotide polymorphisms demonstrate resolution potential to discriminate *Theileria parva* vaccine and Buffaloderived strains. Life 10: 334. <https://doi.org/10.3390/life10120334>

Michael TP, VanBuren R (2020) Building near-complete plant genomes. Curr Opin Plant Biol 54: 26–33. <https://doi.org/10.1016/j.pbi.2019.12.009>

Nguyen VB, Giang VNL, Waminal NE (2020) Comprehensive comparative analysis of chloroplast genomes from seven panax species and development of an authentication system based on species-unique single nucleotide polymorphism markers. J Ginseng Res 44: 135–144. <https://doi.org/10.1016/j.jgr.2018.06.003>

- Pinard D, Myburg AA, Mizrahi E (2019) The plastid and mitochondrial genomes of *Eucalyptus grandis*. BMC Genom 20: 132. <https://doi.org/10.1186/s12864-019-5444-4>
- Paço A, Freitas R, Vieira-da-Silva A (2019) Conversion of DNA sequences: from a transposable element to a tandem repeat or to a gene. Genes 10: 1014. <https://doi.org/10.3390/genes10121014>
- Qiao YG, Zhang XR, Li Z et al (2022) Assembly and comparative analysis of the complete mitochondrial genome of *Bupleurum chinense* DC. BMC Genom 23: 664. <https://doi.org/10.1186/s12864-022-08892-z>
- Robles P, Quesada V (2021) Organelle genetics in plants. Int J Mol Sci 22: 2104. <https://doi.org/10.3390/ijms22042104>
- Shang A, Luo M, Gan RY et al (2020) Effects of microwave-assisted extraction conditions on antioxidant capacity of sweet tea (*Lithocarpus polystachyus* rehd.). Antioxidants 9: 1–17. <https://doi.org/10.3390/antiox9080678>
- Skippington E, Barkman TJ, Rice DW et al (2015) Miniaturized mitogenome of the parasitic plant *Viscum scurruloideum* is extremely divergent and dynamic and has lost all nad genes. Proc. Natl Acad Sci USA 112: 3515–3524. <https://doi.org/10.1073/pnas.1504491112>
- Sloan DB, Alverson AJ, Chuckalovcak JP et al (2012) Rapid Evolution of Enormous, Multichromosomal Genomes in Flowering Plant Mitochondria with Exceptionally High Mutation Rates. PLoS Biol 10: e1001241. <https://doi.org/10.1371/journal.pbio.1001241>
- Seok J, Kim A, Wang J et al (2019) Single-nucleotide polymorphism markers in mitochondrial genomes for identifying *Varroa destructor*-resistant and -susceptible strains of *Apis mellifera* (Hymenoptera: Apidae). Mitochondrial DNA Part A 30: 477–489. <https://doi.org/10.1080/24701394.2018.1551385>
- Sloan DB, Wu ZQ (2014) History of plastid DNA insertions reveals weak deletion and AT mutation biases in angiosperm mitochondrial genomes. Genome Biol Evol 6: 3210–3221. <https://doi.org/10.1093/gbe/evu253>
- Straub SCK, Cronn RC, Edwards C et al (2013) Horizontal transfer of DNA from the mitochondrial to the plastid genome and its subsequent evolution in milkweeds (apocynaceae). Genome Biol Evol 5: 1872–1885. <https://doi.org/10.1093/gbe/evt140>
- Thiel T, Michalek W, Varshney RK et al (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theor Appl Genet 106: 411–422. <https://doi.org/10.1007/s00122-002-1031-0>
- Verhage L (2020) Targeted editing of the Arabidopsis mitochondrial genome. Plant J 104: 1457–1458. <https://doi.org/10.1111/tpj.15097>
- Wang JF, Huang YM, Li KX et al (2016) Leaf extract from *Lithocarpus polystachyus* Rehd. Promote glycogen synthesis in T2DM mice. PloS One 11: e0166557. <https://doi.org/10.1371/journal.pone.0166557>
- Wu ZQ, Liao XZ, Zhang XN et al (2022) Genomic architectural variation of plant mitochondria-A review of multichromosomal structuring. J Syst Evol 60: 160–168. <https://doi.org/10.1111/jse.12655>
- Wynn EL, Christensen AC (2018) Repeats of Unusual Size in Plant Mitochondrial Genomes: Identification, Incidence and Evolution. G3-Genes Genom Genet 9: 549–559. <https://doi.org/10.1534/g3.118.200948>

- Wu ZQ, Ge S (2012) The phylogeny of the BEP clade in grasses revisited: Evidence from the whole-genome sequences of chloroplasts. *Mol Phylogenet Evol* 62: 573–578. <https://doi.org/10.1016/j.ympev.2011.10.019>
- Xu J, Luo H, Nie S et al (2021) The complete mitochondrial and plastid genomes of *rhododendron simsii*, an important parent of widely cultivated azaleas. *Mitochondrial DNA Part B Resour* 6: 1197–1199. <https://doi.org/10.1080/23802359.2021.1903352>
- Ye WT, Wang JC, Zhao XY et al (2022) Mitochondrial Genomes of two *Lycosa* spiders (Araneae, Lycosidae): Genome Description and Phylogenetic Implications. *Diversity* 14: 538. <https://doi.org/10.3390/d14070538>
- Yin X, Gao Y, Song S et al (2021) Identification, characterization and functional analysis of grape (*Vitis vinifera* L.) mitochondrial transcription termination factor (mTERF) genes in responding to biotic stress and exogenous phytohormone. *BMC Genom* 22: 136. <https://doi.org/10.1186/s12864-021-07446-z>
- Yue JW, Lu Q, Ni Y et al (2022) Comparative analysis of the plastid and mitochondrial genomes of *Artemisia giraldii* Pamp. *Sci Rep* 12: 13931. <https://doi.org/10.1038/s41598-022-18387-2>
- Zhou DG, Liu Y, Yao JZ et al (2022) Characterization and Phylogenetic Analyses of the Complete Mitochondrial Genome of Sugarcane (*Saccharum* spp. Hybrids) Line A1. *Diversity* 14: 333. <https://doi.org/10.1080/23802359.2020.1820395>

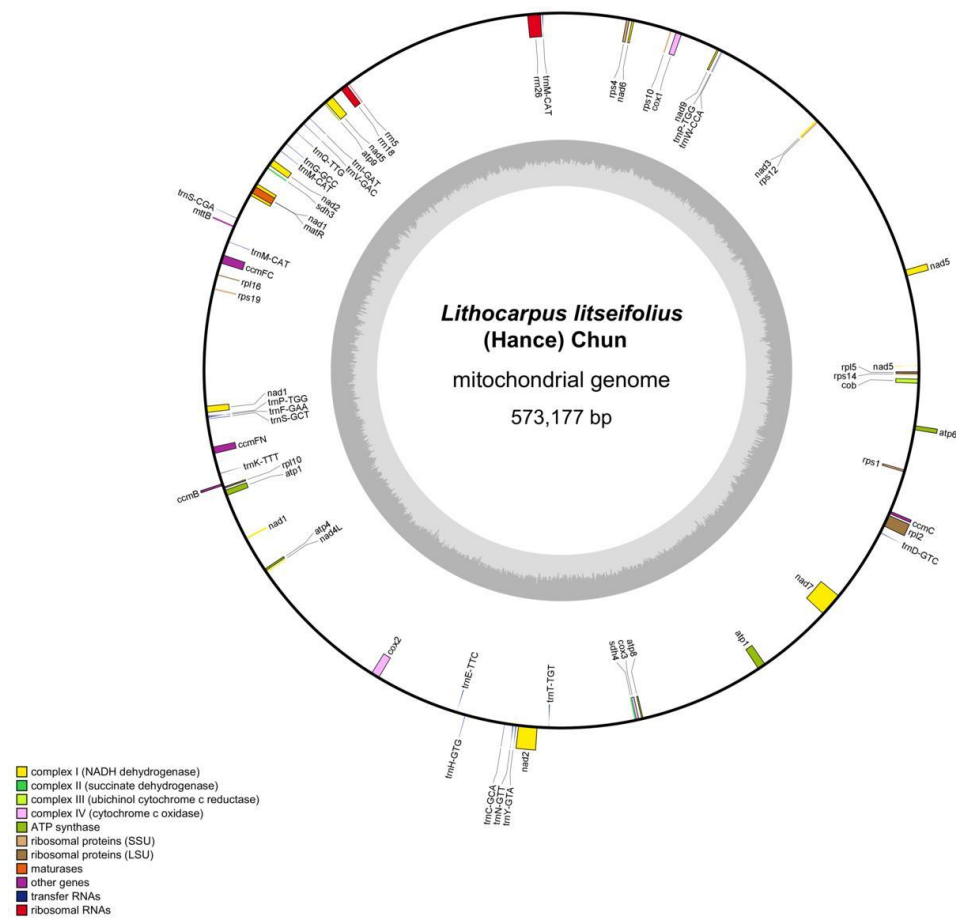


Fig. 1 The circular map of the mt genome of *L. litseifolius*. The gene map illustrates a total of 61 annotated genes belonging to various functional groups. Genes located on the outer and inner regions of the circle are transcribed in a clockwise and counterclockwise direction, respectively. The gray circle positioned inside the map represents the GC content

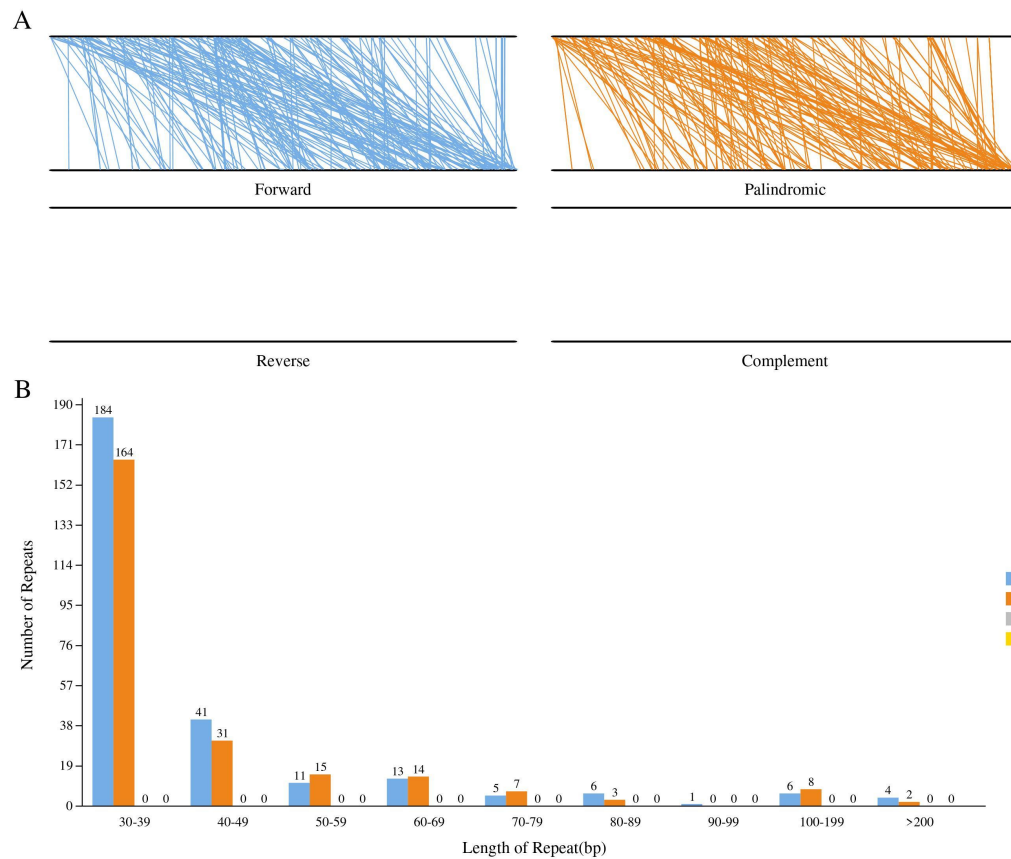


Fig. 2 Statistical graphs of the distribution and length of interspersed repeat sequences in *L. litseifolius* mt genome. A: the genomic distribution of four types of interspersed repeat sequences. The genome is represented by two black lines, and the repeat sequences of the same type are connected by line segments. B: Abscissa is the type of interspersed repeat sequences, and the ordinate is the number of interspersed repeat sequences. C: Complementary repetition, R: Reverse repetition, P: Palindrome repetition, F: Forward repetition

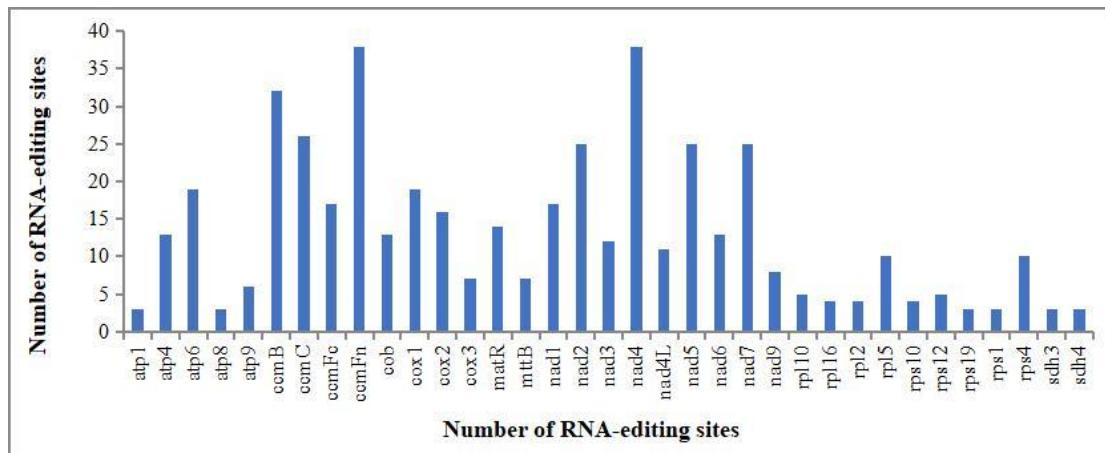


Fig. 3 Forecasting of RNA editing sites in the *L. litseifolius* mt PCGs. The blue bars represent the number of RNA editing sites for each gene

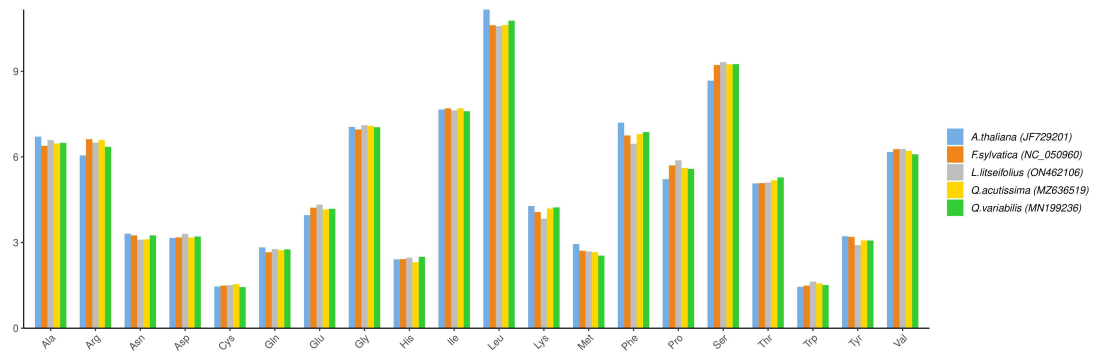


Fig. 4 The mt codon usage pattern of *L. litseifolius* compared to *Q. acutissima* (MZ636519), *Q. variabilis* (MN199236), *F. sylvatica* (NC_050960), and *A. thaliana* (JF729201)

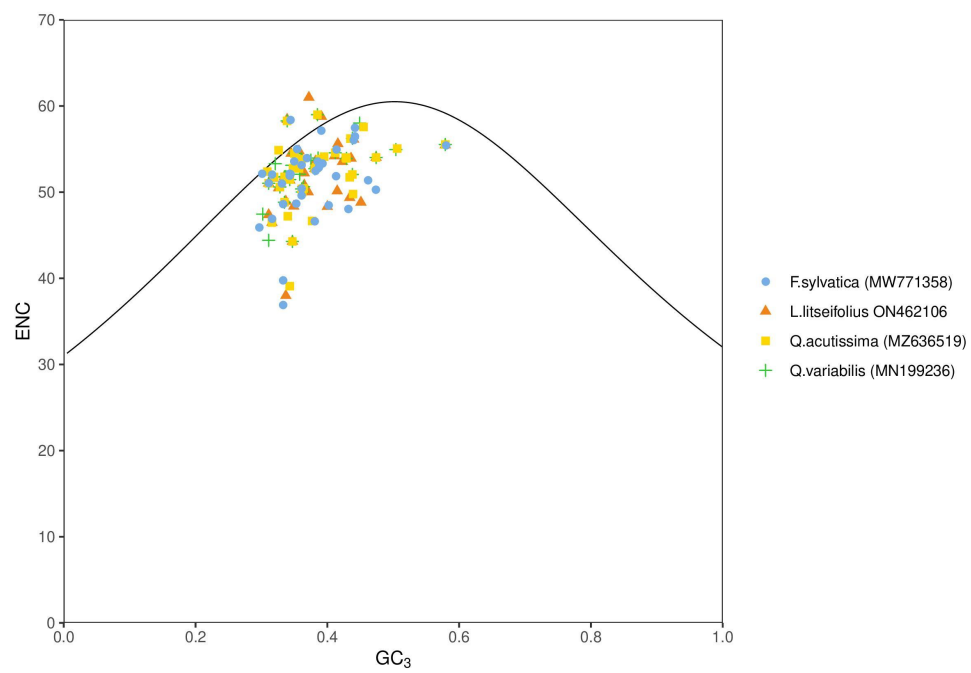


Fig. 5 Four Fagaceae plants mt genome's ENC plot

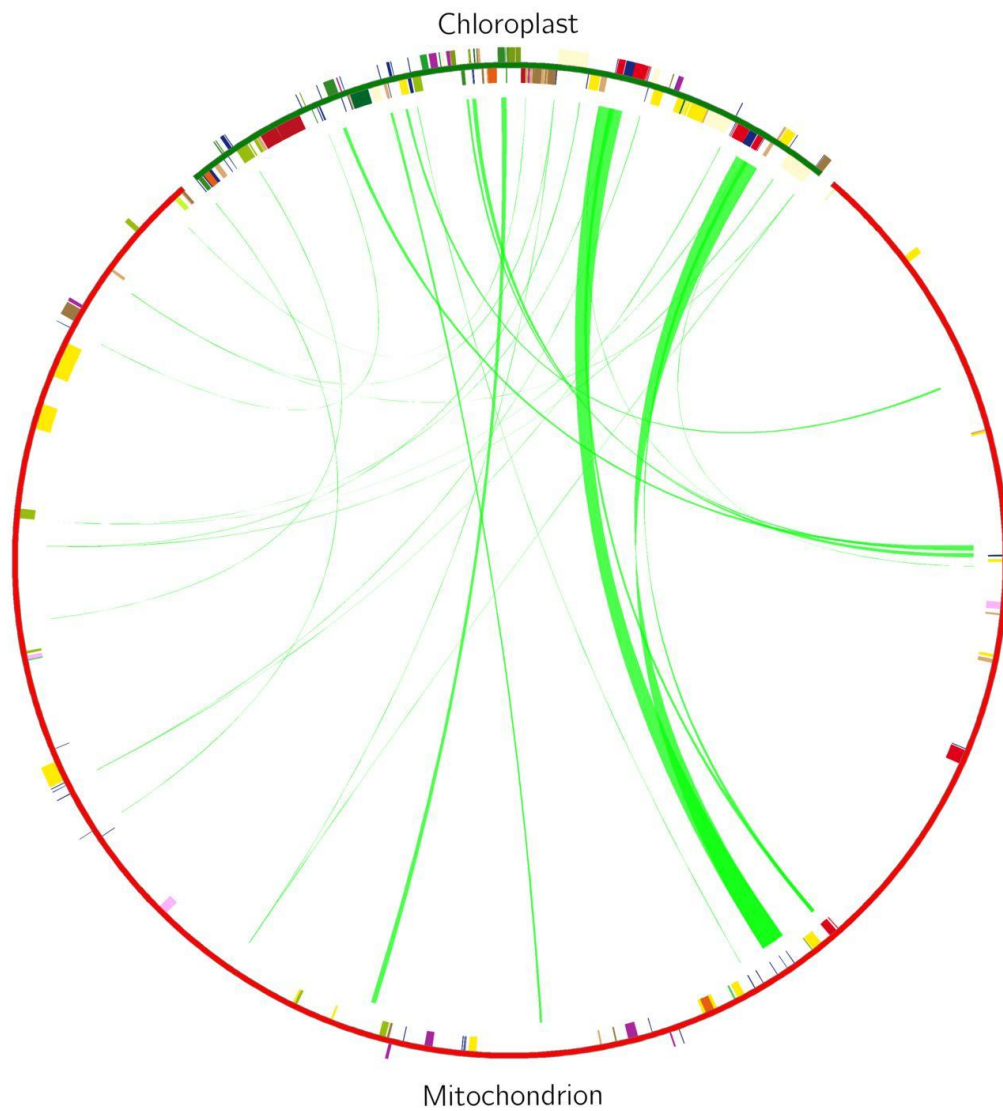


Fig. 6 A comparative analysis was conducted between a homologous fragment present in the chloroplast genome of *L. litseifolius* and its counterpart in the mt genome. The track visually represents the entire genomes of chloroplast and mt DNA, with the former displayed in green and the latter in red. The transplanted gene fragments' beginning and ending positions are joined by the green line segment in the map. The transferred fragment's size is shown by the breadth of the green line segment

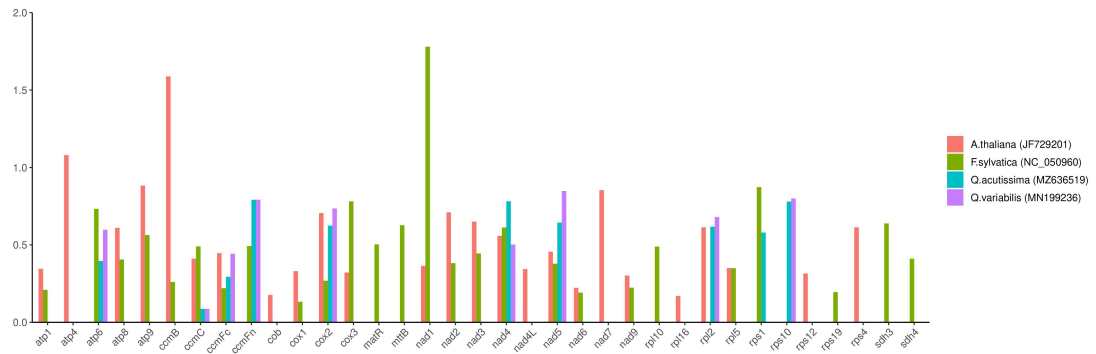


Fig. 7 Ka/Ks ratios of 35 PCGs in *L. litseifolius*, *Q. acutissima* (MZ636519), *Q. variabilis* (MN199236), *F. sylvatica* (NC_050960), and *A. thaliana* (JF729201)

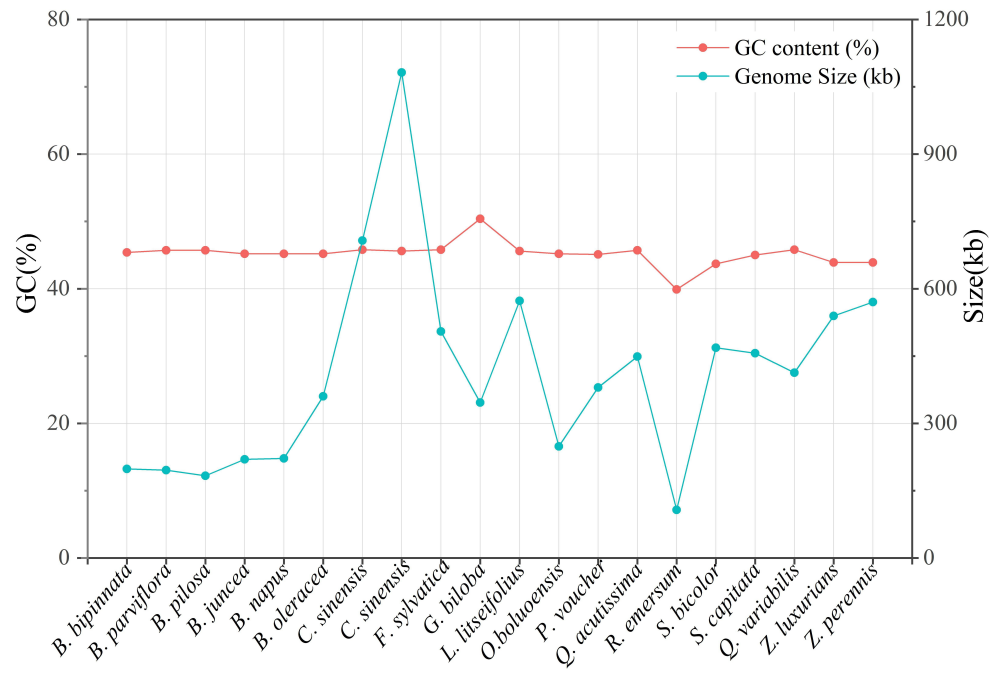


Fig. 8 Gene content (GC) and genome size (kb) of 20 plant mt genomes

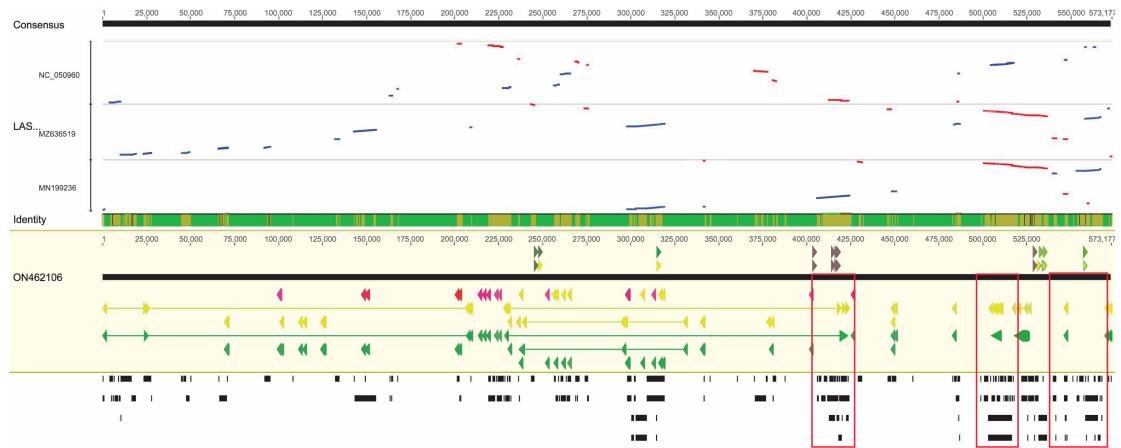


Fig. 9 Four Fagaceae mt genomes were aligned in a collinear manner. Co-directional pairings with high scores are represented by blue bars, while reversed pairings are represented by red bars. Coding sequences are indicated by yellow arrows, ribosomal RNA genes by red bars, transfer RNA genes by purple bars, protein-coding genes by green bars, and exonic areas by gray bars. The size of the black squares corresponds to the extent of shared genetic material between the two species. The conservative regions, which include *atp6*, *rps1*, *ccmC*, *rpl2*, *nad4*, *nad7*, and *trnY-GTA*, are highlighted in red

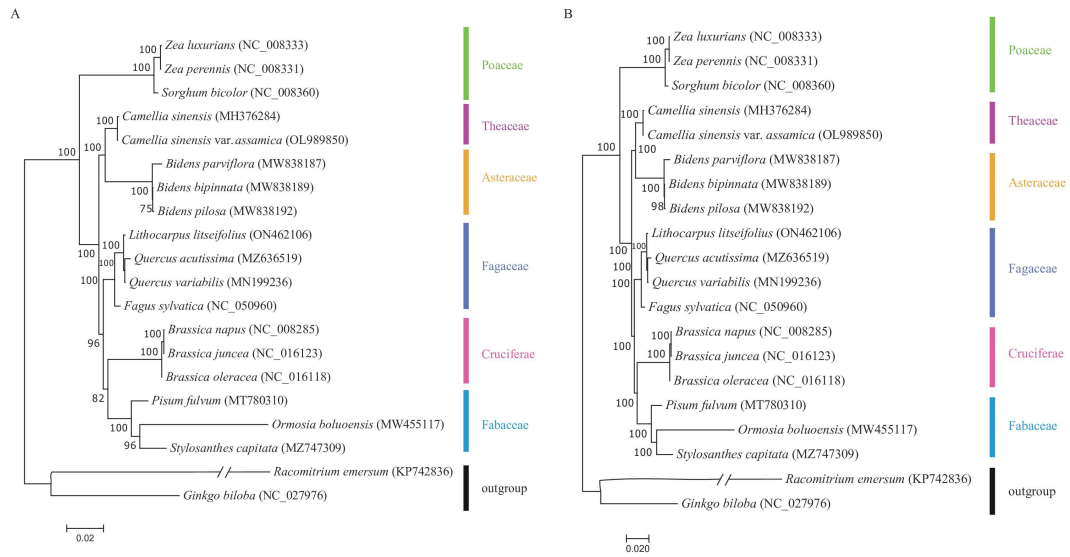


Fig. 10 The phylogenetic relationships of *L. litseifolius* with other 19 plant species. Outgroup: *Racomitrium Emersum* and *Ginkgo biloba*. (A) Maximum-likelihood phylogenies. (B) Bayesian phylogeny.

Table 1 Genomic characteristics of the *L. litseifolius* mt genome

Characteristic	A (%)	T (%)	G (%)	C (%)	GC (%)	Size (bp)	Proportion in Genome (%)
Whole genome	27.33	27.06	22.67	22.94	45.61	573,177	100.00
Protein-coding genes	25.94	30.35	22.19	21.52	43.71	32,208	5.62
Cis-spliced introns	25.40	22.80	27.20	24.60	51.80	27,979	4.88
rRNA genes	25.91	22.08	29.08	22.93	52.01	5,472	0.95
tRNA genes	22.09	26.49	28.62	22.80	51.42	1,548	0.27
Non-coding regions	27.53	27.08	22.45	22.94	45.39	513,006	89.50

Table 2 Gene profile and organization of the *L. litseifolius* mt genome

Gene type	Name
ATP synthase	<i>atp1</i> (2) <i>atp8</i> <i>atp4</i> <i>atp9</i> <i>atp6</i>
Cytochrome c biogenesis	<i>ccmB</i> <i>ccmFN</i> <i>ccmFC</i> * <i>ccmC</i>
Ubichinol cytochrome c reductase	<i>cob</i>
Cytochrome c oxidase	<i>cox1</i> <i>cox3</i> <i>cox2</i> *
Maturases	<i>matR</i>
Transport membrane protein	<i>mttB</i>
NADH dehydrogenase	<i>nad1</i> **** <i>nad3</i> <i>nad2</i> **** <i>nad4L</i> <i>nad4</i> *** <i>nad6</i> <i>nad5</i> **** <i>nad9</i> <i>nad7</i> ****
Ribosomal proteins (LSU)	<i>rpl10</i> <i>rpl16</i> <i>rpl5</i> <i>rpl2</i> **
Ribosomal proteins (SSU)	# <i>rps14</i> <i>rps1</i> <i>rps12</i> <i>rps10</i> <i>rps19</i> <i>rps4</i>
Succinate dehydrogenase	<i>sdh4</i> <i>sdh3</i>
Ribosomal RNAs	<i>rrn18</i> <i>rrn26</i> <i>rrn5</i>
	<i>trnC</i> -GCA <i>trnE</i> -TTC <i>trnD</i> -GTC <i>trnH</i> -GTG <i>trnF</i> -GAA <i>trnI</i> -GAT*
	<i>trnG</i> -GCC <i>trnM</i> -CAT(3) <i>trnK</i> -TTT <i>trnN</i> -GTT <i>trnQ</i> -TTG
Transfer RNAs	<i>trnP</i> -TGG(2) <i>trnS</i> -GCT <i>trnS</i> -CGA <i>trnY</i> -GTA <i>trnT</i> -TGT* <i>trnW</i> -CCA <i>trnV</i> -GAC

* : number of introns; Gene(2): Number of multi-copy gene copies; # Gene: Pseudo gene.

Table 3 SSR distribution of in the mt genome of *L. litseifolius*

SSR form	Repetitive fragment	SSR numbers	Total
monomer	C/G	13	51
	A/T	38	
dimer	AT/AT	16	40
	AG/CT	19	
	CG/CG	1	
	AC/GT	4	
trimer	AAC/GTT	4	18
	AAT/ATT	3	
	ATC/ATG	1	
	ACT/AGT	1	
	AAG/CTT	8	
	ACC/GGT	1	
tetramer	AAAC/GTTT	2	57
	AACC/GGTT	2	
	AAGC/CTTG	2	
	AAGG/CCTT	5	
	AAAG/CTTT	18	
	AAAT/ATTT	3	
	AATC/ATTG	1	
	AACT/AGTT	1	
	ACCG/CGGT	2	
	ATCC/ATGG	1	
	ACAT/ATGT	1	
	CCCG/CGGG	3	
	AAGT/ACTT	2	
	AGCG/CGCT	1	
	AATG/ATTC	9	
	AGAT/ATCT	1	
	ATCG/ATCG	1	
	ACTC/AGTG	2	
pentamer	AAAAG/CTTTT	2	9
	AAACT/AGTTT	1	
	AATAT/ATATT	1	
	AATGC/ATTGC	1	
	AATGG/ATTCC	1	
	AATGT/ACATT	1	
	ACTAG/AGTCT	1	
	ACTAT/AGTAT	1	

Table 4 Forecasting of RNA editing sites

Type	RNA-editing	Number	Percentage
hydrophobic	CCG (P) => CTG (L)	30	32.32%
	CCT (P) => CTT (L)	24	
	CCA (P) => CTA (L)	46	
	CCC (P) => CTC (L)	7	
	CTC (L) => TTC (F)	8	
	CCT (P) => TTT (F)	10	
	CCC (P) => TTC (F)	4	
	CTT (L) => TTT (F)	12	
	GCT (A) => GTT (V)	2	
	GCA (A) => GTA (V)	1	
	GCG (A) => GTG (V)	5	
hydrophilic	CAC (H) => TAC (Y)	8	11.50%
	CGC (R) => TGC (C)	9	
	CAT (H) => TAT (Y)	13	
	CGT (R) => TGT (C)	23	
hydrophobic-hydrophilic	CCG (P) => TCG (S)	3	8.03%
	CCT (P) => TCT (S)	15	
	CCC (P) => TCC (S)	12	
	CCA (P) => TCA (S)	7	
hydrophilic-hydrophobic	TCT (S) => TTT (F)	38	47.72%
	TCC (S) => TTC (F)	31	
	TCA (S) => TTA (L)	65	
	TCG (S) => TTG (L)	40	
	ACC (T) => ATC (I)	1	
	ACA (T) => ATA (I)	5	
	ACT (T) => ATT (I)	5	
	ACG (T) => ATG (M)	6	
	CGG (R) => TGG (W)	29	
	CGA (R) => TGA (X)	1	
hydrophilic-stop	CAA (Q) => TAA (X)	1	0.43%

Table 5 Fragments transferred from chloroplasts to mitochondria in *L. litseifolius*

	Identity/%	Length (bp)	Mismatches	Gap opens	cp start	cp end	mt start	mt end	cp gene	mt gene
1	99.671	6,080	0	2	219,243	213,184	104,137	110,216	<i>rrn16/rps12/trnI-GAU/trnA-UGC/trnV-GAC</i>	<i>trnI-GAT/trnV-GAC</i>
2	99.671	6,080	0	2	213,184	219,243	141,658	147,737	<i>rps12/rrn16/trnA-UGC/trnV-GAC/trnI-GAU</i>	<i>trnV-GAC/trnI-GAT</i>
3	93.755	1,249	60	7	323,495	322,253	79,184	80,420	<i>psbB/psbT</i>	/
4	99.492	788	4	0	98,810	98,023	37,858	38,645	<i>psbC/psbD/psbD/psbC</i>	/
5	77.657	1,016	152	46	100,069	101,021	71,686	72,689	<i>petL/petG/trnW-CCA/trnP-UGG</i>	<i>trnW-CCA/trnP-TGG</i>
6	95.882	340	8	4	279,666	279,333	50,729	51,068	<i>rps4</i>	/
7	82.424	495	76	6	98,828	99,313	70,294	70,786	<i>psbE</i>	/
8	80.62	516	70	14	57,302	56,812	54,497	55,007	<i>ndhJ</i>	/
9	73.903	889	177	42	203,761	202,903	106,926	107,789	<i>rrn16</i>	<i>rrn18</i>
10	73.903	889	177	42	202,903	203,761	144,085	144,948	<i>rrn16</i>	<i>rrn18</i>
11	97.241	145	4	0	476,605	476,461	99,097	99,241	<i>ycf2</i>	/
12	97.241	145	4	0	476,461	476,605	152,633	152,777	<i>ycf2</i>	/
13	86.301	146	10	1	279,820	279,675	50,456	50,591	<i>trnS-GGA</i>	/
14	96.703	91	2	1	405,076	404,986	14	103	<i>trnH-GUG</i>	<i>trnH-GTG</i>
15	98.795	83	1	0	417,577	417,495	114,933	115,015	<i>trnN-GUU</i>	<i>trnN-GTT</i>
16	98.795	83	1	0	417,495	417,577	136,859	136,941	<i>trnN-GUU</i>	<i>trnN-GTT</i>
17	96.341	82	3	0	530,285	530,204	33,807	33,888	<i>trnD-GUC</i>	<i>trnD-GTC</i>
18	94.667	75	4	0	226,104	226,030	57,590	57,664	<i>trnT-GGU</i>	<i>trnM-CAT</i>
19	88.889	99	2	2	457,893	457,991	13,807	13,896	<i>atpF</i>	/
20	89.61	77	6	2	545,352	545,276	159,211	159,285	<i>trnI-CAU</i>	/
21	89.61	77	6	2	545,276	545,352	92,589	92,663	<i>trnI-CAU</i>	/
22	89.333	75	6	2	358,186	358,112	92,591	92,663	<i>trnI-CAU</i>	/
23	89.333	75	6	2	358,112	358,186	159,211	159,283	<i>trnI-CAU</i>	/
24	88.571	70	8	0	483,029	482,960	108,124	108,193	<i>rrn16</i>	/
25	88.571	70	8	0	482,960	483,029	143,681	143,750	<i>rrn16</i>	/
26	87.143	70	0	3	103,453	103,384	104,286	104,346	<i>rps12</i>	/
27	87.143	70	0	3	103,384	103,453	147,528	147,588	<i>rps12</i>	/
28	100	29	0	0	567,560	567,532	85,343	85,371	<i>rps11</i>	/

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplymentlymaterials.pdf](#)