**Running title**: chloroplast genome and historical biogeography of the three Magnolias

**chloroplast genome and historical biogeography of the three Magnolias**

**Min Zhang1, Sha Zhong 1, Shuai Guo1, Yanpeng Yin1, Luojing Zhou1, Bo Ren1, Li Wang2, Xiaodong Shi3, Feixia Hou1, Guoqing Zhuang2, Cheng Peng1, Xianmei Yin1\*, Jihai Gao1\***

1. *Pharmacy College, Chengdu University of Traditional Chinese Medicine, Key Laboratory of Distinctive Chinese Medicine Resources in Southwest China,* 611137, Chengdu, China

2. *Sichuan Academy of Forestry Sciences*, 610065, Chengdu, China

3. *Chengdu University*, 610106, Chengdu, China

\* For correspondence: gaojihai@cdutcm.edu.cn; [xianmeiyincm@163.com](mailto:xianmeiyincm@163.com)

\* Contributed equally to this work and are co-correspondent authors

**Novelty statement**

The related studies of *Magnolia officinalis*, *M. officinalis* subsp.*biloba* and *M. hypoleuca* mainly focus on the chemical constituents, clinical application, resource investigation and others. However, the theoretical guidance of molecular breeding and the divergence time of the three varieties have not been reported. This study compared the chloroplast genome of three Magnolias and developed the divergence time analysis. It will provide genetic resources for future research in the genus, and decipher the genetic relationship of three Magnolias to provide some reference value for molecular breeding of superior variety, and also provide useful information for identifing three Magnolias species, and providing insight into their evolutionary history.

**Abstracts**

*Magnolia officinalis* L, *M. officinalis* sub*s*p*.biloba* L.and *M. hypoleuca* L.are all typical medicinal plants, belonging to genus *Magnolia* and Family Magnoliaceae. Their molecular information, particularly genetic difference, were known less. In this study, the platform Illumina HiSeq was used to sequence and assemble a novel cp (chloroplast) genome of *M*. *hypoleuca* followed by cp purification. Combined with two published cp rawdata, gene cycles and function annotations were comparably performed for the three plant species. The results indicated that 19 791 019 clean readswas assembled for *M*. *hypoleuca* cp, Q30 being 91.33%, and genome 160 051 bp. Its GC content is 39.2%, including 37 tRNAs and 8 rRNAs. The *M. hypoleuca* had smaller cp genome and more introns (or exons) than *M. officinalis* and *M. officinalis* sub*s*p*.biloba*. And there were respectively 11 and 8 more functional genes in *M*. *hypoleuca* cp than that in the other two. Based on cp complete genomes sequences, we constructed the phylogenetic relationship and estimated the divergence time of the three species by ML method, with other 10 published Magnoliaceae species. The results showed that *M. officinalis* sub*s*p*.biloba* and *M. officinalis* might diverge from *M. hypoleuca* around 18.98 Ma, then they diverged from each other around 15.00 Ma. Additionally, the middle Miocene warming period might play an important role in the demographic and evolutionary histories of the three Magnolias*,* which provided a novel insight of the origin and dispersal routes of *M*. *hypoleuca*.

**Keywords** three Magnolias, *Magnolia hypoleuca*, chloroplast complete genome, phylogenetic tree, divergence time

**Introduction**

*M.* *officinalis*, *M. hypoleuca* and *M. officinalis* subsp.*biloba* are the deciduous tree belonging to the genus *Magnolia* (Iwasaki *et al.* 2012). Their bark have been used for thousands of years in Chinese and Japanese traditional medicines and are still widely employed as herbal preparations for their sedative, antioxidant, anti-inflammatory, antibiotic (Sarrica *et al.* 2018), dyspepsia (Shin *et al.* 1990) and antispastic effects (Luo *et al.* 2019). Besides, three Magnolias have many other uses besides medicine. For example, *M.hypoleuca*, famous in Korea, Japan and China (Youn *et al.* 2008), is widely utilized as a natural packaging material for traditional foods in Japan (Kawahara *et al.* 2014). Meanwhile, a comparision of three Magnoliasfound that *M. hypoleuca* owned unique biological characteristics, such as stronger cold resistance ability (Liu *et al.* 2006; Ravi *et al.* 2008), higher ß-cineol content (Yang *et al.* 2000) and faster growth and maturity rate.. Former researchers found that *M. hypoleuca* was the northernmost of all Magnolias plants and its cold tolerance was better than other Magnoliaceae plants，which is native to the south of the Kuril islands (Kwon *et al.* 2015). Till now, *M. hypoleuca* was widely distributed in Janpan, China and Korea. Instead, *M. officinalis* subsp.*biloba* and *M. officinalis* are native to China and were widely distributed in China. The origin of Janpanese islands was the eastern margin of the Eurasian Continent rifted approximately 700–750 Ma (Daniell *et al.* 2016). After the formation of Japan, those continental islands have experienced various paleogeographic and paleoclimatologic changes resulting in high endemism and biodiversity of terrestrial species (Liu *et al.* 2019; Richardson *et al.* 2001; Mao *et al.* 2012), such as *Liriodendron* (Liu *et al.* 1957), *Magnolia section Rytidospermum* (Maruyama *et al.* 1997), Some of them had been reported with the molecular phylogenetic techniques to study evolutionary patterns or estimate divergence times of disjuncts. It has been proposed that biogeographic studies of widely distributed plants can be performed to provide insight into the broader patterns of the evolutionary history and geographic diversification (Motokawa *et al.* 2017). However, there is no analysis of the evolutionary history and biogeography of the three Magnolias.

Cp-genome (chloroplast genome) possess a highly conserved tetrad structure, containing two inverted repeat (IR) regions (IRa and IRb), a small single-copy (SSC) region and a large single-copy (LSC) region (Jansen *et al.* 2005; Nie *et al.* 2008). In addition to photosynthesis, cp genome-encoded proteins are involved in other metabolic processes, such as responses to heat, drought, salt, and light (Kinoshita *et al.* 2019). Besides, cp-genome were widely used in species evolution (PARKS *et al.* 1990), phylogenetic (Kim *et al.* 2014; He *et al.* 2020; He *et al.* 2020) and biogeographic studies (Li *et al.* 2018; Roman *et al.* 2019; Gutiérrez *et al.* 2018). Meanwhile, organismic and environmental processes played a major role in organismal evolution (Yi *et al.* 2004; QIU *et al.* 1995). Currently, with the rapid development of high throughput sequencing technologies, made it possible for reseachers to obtain cp genomic sequences to study species evolution, phylogenetic and biogeographic (Wen *et al.* 1999). Therefore, whole cp genome of *M. hypoleuca* was urgently needed, which might significantly provide an insight in plant phylogenetic relationships and contributed to the domestication, and utilization of Magnoliaceae plants.

In this study, we sequenced cp genome of *M. hypoleuca* and performed the comparative genomes analysis to obtain comprehensive understanding the structure of cp genomes within three Magnolias. Meanwhile, we studied divergence time of three Magnolias*,* which indicated that it was consistent with Darwinian evolutionary theory. Our study will provide genetic resources for future research in the genus, and decipher the genetic relationship of three Magnolias to provide some reference value for molecular breeding of superior variety, and also provide useful information for identifing three Magnolias species, and providing insight into their evolutionary history.

**Materials and Methods**

**Sample material, DNA extraction, and sequencing**

The fresh and insect-free leaves samples of *M. hypoleuca* were harvested in the State Bank of Chinese Drug Germplasm Resources, Chengdu University of Traditional Chinese Medicine (Chengdu, China). We extracted cp complete genome from 20–30 mg of fresh leaf material (dried mass after lyophilization) by using modified CTAB (Matthes *et al.* 2020) method. Fresh material was lyophilized with liquid nitrogen. The specific extraction steps were as follows: The fresh leaves were smashed in liquid-nitrogen, and suspended in liquid A (50 mmol·L-1 Tris, 25 mmol·L-1 EDTA, 1.25 mol·L-1 NaCl, 0.25 mmol·L-1 Vc, 1.5% PVP, pH 3.6). Then filtering and retaining the supernatant; The Buffer B (50 mmol·L-1 Tris,25 mmol·L-1 EDTA,1.25 mol·L-1 NaCl,0.25 mmol·L-1 Vc,1 mmol·L-1 DTT,0.1% Bovine Serum Protein BSA,pH 8.0) was added in the supernatant and samples were placed at room temperature, extracted for 10 min at 2000 g. Moreover, the cpDNA of *M. hypoleuca* was stored at 4℃ to guarantee the quality. In addition, DNA of *M. hypoleuca* was qualified through Qubit 2.0 (BMKcloud http://www.biocloud.net/), then fragmented, purified, constructed sequencing library, and sequenced by a high-throughput sequencing platform (Illumina HiSeq PE14).

**Genome assembly and annotation**

Raw sequences were submitted to China National Center for Bioinformation (BioProject number PRJCA004348). Pair-end Illumina raw reads were cleaned from adaptors and barcodes, and then quality filtered by Trimmomatic (Bolger *et al.* 2014). After quality filtering, we retrieved the cpDNA sequence of *M. officinalis* (NC\_020316) and *M. officinalis* subsp.*biloba* (JN867581) as the references. The plastome was assembled using online platform of Galaxy (https://usegalaxy.org) (Yan *et al.* 2015). Then, the cp complete genomes of *M.hypoleuca,* *M. officinalis* and *M.officinalis* subsp.*biloba* were annotated using the online program CpGAVAS2 (Liu *et al.* 2012). And we drawed a circular representation of both sequences by using the online tool Genome VX (Conant *et al.* 2008).

**Evolutionary and Phylogenetic analysis**

To estimate phylogenetic relationships within three Magnolias, a total of 14 cp complete genomes were downloaded from the NCBI database. The Genbank accession numbers for each plant species were follows: *M. grandiflora* (JN867584), *M. sinostellata* (NC039941), *M. biondii* (KY085894), *M. sprengeri* (JX280401), *M. denudata* (JN227740), *M. yunnanensis* (KF753638), *M. zenii* (MH607378), *M. liliflora* (JX280397), *Liriodendron chinense* (NC030504), *Liriodendron Tulipifera* (NC008326) and *Ginkgo biloba* (NC016986). Among them, the last species, *Ginkgo biloba*, was used as outgroup.

The 14 cp complete genomes sequences were aligned using megaX. The program megaX was used to infer ClustalW, with Pairwise alignment of the sequence, according to the pairwise alignment to calculate the pairwise distance matrix. Moreover, model detection tool, "Model",was used for optional model detection. We used the optional model to construct the evolutionary tree using the ML method, with the bootstrap number set as 1000. Among them, only the bootstrap number of Phylogenetic tree was bigger or equal to 50%, it just valued to be reliable. The clades division in figures was mainly based on the clustering results of phylogenetic trees and the morphological classification of Magnoliaceae. Then the tree with “.nwk” format was downloaded for next analysis.

**Divergence time estimates**

We used the cp complete genomes data set for dating the divergence times. ML approaches based on a global clock model was usually used in the time estimates. A likelihood ratio test ruled out a global molecular clock (P < 0.05) for our data (Felsenstein *et al.* 1988). Time estimates were done based on a global molecular clock and fossil data. That was, we used the final aligned sequences with 14 species, which were converted to MEGA format by using MEGAX software (Kumar *et al.* 2018), and the phylogenetic tree of 14 species with .nwk format. An ML tree with lengths inferred from GARLI was used in the PL estimate with steps. All analyses were performed using the GTR model of nucleotide substitution with a gamma distribution with four rate categories. The tree calculation models was implemented in each analysis, with rate variation across branches assumed to be uncorrelated and lognormally distributed. Among them, three clades with an expanded outgroup species was used in the dating analyses, meaning the that individual DNA alignments were pruned to eliminate multiple accessions of the same species (Xu *et al.* 2017). The fossil times of *L. chinense* vs *L. Tulipifera* was 14.2 MYA, *M. zenii* vs *L. chinense* was 55 MYA and *M. sprengeri* vs *M. zenii* was 28.3 MYA*.*The outgroup fossil times *Ginkgo biloba* vs Magnoliaceae was 313 MYA. Therefore, the phylogeny was calibrated using four Magnoliaceae fossils that were applied to the stem nodes of *L. chinense*, *L. Tulipifera*, *M. zenii* and *M. sprengeri*. Additional fossil of Ginkgoaceae species was used to calibrate the stem node of Magnoliaceae and selecteed as an outgroup. All analyses were outputed in nwk format. The final estimates were obtained using the model that yielded the highest posterior probability. Posterior distributions of parameters were approximated using two independent analyses of 20 000 000 generations with 10% burn-in. Samples from the two chains which yielded similar results were combined.

**Results**

**Genome Sequencing, assembly and annotation**

A length of 19 816 708 raw readsof *M.hypoleuca* was obtained by Illumina HiSeq PE14 sequencing platform. Through removing the connector and low-quality reads, a length of 19 791 019 clean reads of *M.hypoleuca* were obtained and the Q30 was 91.33%. After sequence assembly and annotation, the results showed that the total length of the cp genome of *M.hypoleuca*, *M. officinalis* subsp.*biloba* and *M. officinalis* were 160 051 bp, 160 099 bp and 160 183 bp, respectively. (Fig 1 and Table 1). Among them, the genome of *M.hypoleuca* had a quadripartite structure with an SSC of 18 771 bp and an LSC of 88 146 bp, which were separated by two IR regions (IRa and IRb, each 26 562 bp). The GC content of the overall cp genome was 39.2%, comprising 8 rRNA and 37 tRNA, respectively (Table 1). Through the comparision of the cp genomes of three Magnolias, we concluded that the cp genomes structure had a typical quadripartite structure, with a circular molecule of 160 051 bp to 160 183 bp in length, and the content of GC, rRNA and tRNA were same in the cp genes of three Magnolias. On the contrary, the total number of genes and coding region were distincted in three Magnolias, with its total number of *M. hypoleuca* were 13 and 14 times higher than *M. officinalis* and *M. officinalis* subsp.*biloba*, respectively (Table 1).

**Comparative analysis of cp genomes**

Chloroplast is important organ for plant photosynthesis, which is very conservative in structure. Comparative analysis of cp genomes is an essential step in genomics (Chen *et al.* 2017; Liu *et al.* 1985). A comparison of the structural differences among cp genomes of three Magnolias indicated that the cp genome of *M. hypoleuca* was the smallest (Fig 1; *M. hypoleuca*, 160,051 bp; *M. officinalis* subsp.*biloba*, 160,099 bp; *Magnolia officinalis*, 160,183 bp). Instead, Comparison of the functional genes quantity showed that the cp genome of *M. hypoleuca* was the most which indicated *M. hypoleuca* had more functional effects. The result showed that *M. hypoleuca* had 11 more functional genes than *M. officinalis*, including *psbC*、*ycf14、ycf1、rps14、rps12、rps12、rpl22、rpl23、petD、petL、ropC2,* respectively, and had 8 more functional genes than *M. officinalis* subsp.*biloba*, including *psbC、ycf14*、*ycf1*、*rpl22*、*rps12、rpoC2、petD、petL,* respectively (Table 3). Studies on ribosomal protein rpl22 and rpl23 revealed they were essential of metabolism of organisms, whatever the plant was the stage of developmental or light phase. Moreover, the above all different genes were intensively located in Photosynthetic System Ⅱ Subunit, large ribosomal subunit and small ribosomal subunit, which mainly control photosynthesis and polypeptide formation in plants (Zhou *et al.* 2001; Xu *et al.* 2003).

Except for the length of total cp complete genome sequences and amount of functional genes, the frequently divergent regions between three Magnolias were mainly in the introns (or exons) content in genes (Table 2). Introns play important roles in regulating gene expression (Ping *et al.* 2020). The genes containing introns (or exons) of cp genomesof *M. hypoleuca* had 2 more than that in *M. officinalis* and *M. officinalis* subsp.*biloba*. They were respectively *petD* and *rpl16* gene (Table 2). It indicated that *petD* and *rpl16* genes of *M. hypoleuca* were more than that in *M. officinalis* and *M. officinalis* subsp.*biloba*. In this study, the length of exonⅠ of *ycf3* and *rps12* gene of *M. hypoleuca* were 102 bp longer and 114 bp smaller than the other two species. In all species, the two *rps12* gene were trans-spliced. There were two *rps12* gene in cp genomes of three Magnolias. Those were, the length of intronⅠ and exonⅠ of another *rps12* geneof *M. hypoleuca* were 535 bp longer and 227 bp smaller than the other two species. Moreover, *rps12* gene encoded the ribosome S12 protein (Liu *et al.* 2020), which is usually highly conserved, and its structural change was thought to be the result of evolution (Chen *et al.* 2019). It revealed differences among the three Magnolias.

**Divergence time estimation based on cp complete genomes**

We used 14 cp complete genomes of Magnoliaceae and Ginkgoaceae for phylogenetic reconstruction and estimation of the divergence times of three Magnolias. Partial phylogenetic tree of 14 species was constructed by ML method (Fig. 2). Results showed the ML phylogenetic tree based on the cpDNA of partial Magnoliaceae was divided into two main clades (Fig. 2). The first clade was Magnoliaceae and the second clade was outgroup, *Ginkgo biloba*. Among them, the first group could be further divided into two secondary groups, including 11 species of *Magnolia* genus as the first groups and *Linriodendron* genus as the second groups. The phylogenetic relationship of 14 species was mostly consistent with the study of *Lirianthe hodgsonii* (Chen *et al.* 2020; Parks *et al.* 1990). Moreover, through comparision of the phylogenic relationship of three Magnolias, *M. officinalis* was most closely related to *M. officinalis* subsp.*biloba*, followed by *M. hypoleuca*. The relationship between other Magnoliaceae plants and the three Magnolias was as follows: *M. sinostellata*, *M. yunnanensis*, *M. liliflora*, *M. biondii*, *M. grandiflora*, *M. denudata*, *M. zenii*, *L. tulipifera* and *L. chinense*. Besides, the result showed that the degree of genetic divergence between *M. officinalis* and *M. officinalis* subsp.*biloba* were later than *M. hypoleuca* in three Magnolias*.* The divergence time of *M. officinalis* was estimated to be 15.00 MYA and it was same as *M. officinalis* subsp.*biloba*, later than divergence time of *M. hypoleuca* which was estimated to be 18.98 MYA*.* Meanwhile, *Liriodendron* was estimated to have diverged in the mid Miocene based on allozyme and restriction fragment length polymorphism (RFLP) analyses of cpDNA and paleobotanical evidence (Li *et al.* 2012). It was consistent with the conclusion in this study (Fig 2). The divergence time of 13Magnoliaeace we selected was subdivided at the period of 14.20-21.78 MYA. It indicated that the evolutionary and biogeography of partial modern Magnoliaceae species mainly occurred in mid Miocene.

**Discussion**

Through comparison of the cp genomes of the three Magnolias, and the main difference was found in the number of functional genes. Especially functional genes numbersof *M. hypoleuca* were more 11 than *M. officinalis*, and more 8 than *M. officinalis* subsp.*biloba*. Among them, *psbC* and *psbD* genes were key genes of photosynthetic system Ⅱ. Some researchers found the synthesis of *psbC* gene product might occur in the transcript containing *psbD* sequence. That indicated that *psbD* and *psbC* genes jointly coordinated the effect on light (Xu *et al.* 2015). Meanwhile, studies in many higher plants have shown that light-induced transcription of *psbD-psbC* gene increased and enhanced the synthesis and maintenance of Ⅱ subunit of chloroplast photosynthetic system, which is related to the need for relatively high level of *psbD* gene products during rapid cp synthesis (Gamble *et al.* 1989). It indicated the reason why *M. hypoleuca* has stronger adaptability to environment. Furthermore, some scholars (Liu *et al.* 2020) found that *psbD* gene in the *Phyllostachys japonicus* leaves played the protective role of photosynthetic system and reduce strong sunlight damage, when the plants were in a stage of growth and color differences. In this study, *psbC* gene number of *M. hypoleuca* was more than that of *M. officinalis* and *M. officinalis* subsp.*biloba*. The different number of *psbC* gene of three Magnolias might be one of the factors affecting photosynthesis, growth and development.

Besides, according to the *petD* gene of *Ginkgo biloba* (Yang *et al.* 2019), it found that the expression level of *petD* was highest in leaves, next to the stem, which was related to the formation of cytochrome b/f complex. The main function of cytochrome b/f complex was to participate in the electron transport process of photosynthesis and enhanced the photosynthetic ability of leaves. In this study, the *petD* gene of *M. officinalis* and *M. officinalis* subsp.*biloba* was probably pseudogenes that they did not gene expression. That also revealed that the photosynthetic capacity and growth ability of *M. hypoleuca* were stronger than the other two species.

It was well known that *M. hypoleuca* had more cold resistent ability than the other two species. In this study, based on the comparison of the contents and types of tRNA in the three Magnolias cp genomes, it showed that *M. hypoleuca* had 3 more tRNA in alanine transport than the other two species and it had unique types of tRNA in histidine transport. Some studies showed that the types and quantities of tRNA will change in different periods at the same organism and will be adjusted accordingly with environmental changes (Bhaskaran *et al.* 2012). For example, the natural structure formation and expression of tRNA was controled and affected by temperature and ionic strength (Mustroph *et al.* 2014). Alanine was one of the main products of anaerobic metabolism in plants, which could accumulate rapidly under stress conditions (Wei *et al.* 2020), and its free form could resist environmental stimulation. The content of free histidine in cp was positively correlated with the SOD content and it counld resist cold and oxidation. In short, it indicated the reason why *M. hypoleuca* had highest cold resistent ability of three Magnolias.

Comparing three Magnolias species cp genes introns (or exons), the results showed that *rps12* gene introns of *M. officinalis* and *M. officinalis* subsp.*biloba* was missing, and its exonsⅠ content was 0.50 times of *M. hypoleuca*. The plastid ribosomal protein S12 encoded by the *rps12* gene was a highly conserved protein, which was located in the functional center of the 30S subunit of the ribosome. Some scholars found that *rps12* was a trans-splicing gene in ferns, and the evolution rate of species was affected due to its change in intron deletion and exon position. It indicated that intron loss accelerated the evolution rate (Xu *et al.* 2010). In this study, the different length of *rps12* introns of three Magnolias revealed the evolutionary relationship between three Magnolias. Furthermore, some scholars found that the diversified expressionof *rps12* gene caused the different phenotypes in plants, such as sharp and narrow leaves, serrated or defective leaves. It was known that the leaf shape of Magnolias was divided into three types: the first type of leaf shape was sharply pointed or blunt, the second type had slightly absent or obtuse leaves at the apex and the apex of the leaves of the third type was concaved into 2 shallow lobes, but there was no obvious concave at the apex of the seedlings, which were more obtuse or slightly emarginate (Zachos *et al.* 2001). According to the theory of species evolution, the first type is relatively primitive, and the third type has a higher degree of evolution. Among them, the leaf shape of *M. hypoleuca, M. officinalis* and *M. officinalis* subsp.*biloba* belonged to the first type, second type and third type, respectively. It indicated that *M. hypoleuca* was relatively more primitive than *M. officinalis* and *M. officinalis* subsp.*biloba* by the diversity of leaf shape of three Magnolias. In short, *rps12* gene evolutionary study and leaf shape evolutionary researches were consistent that the evolution rate of *M. officinalis* and *M. officinalis* subsp.*biloba* was faster than that of *M. hypoleuca*.

Furthermore, we did the cp-genome of three Magnolias phylogenetic analysis. Through a comparision of the divergence time of three Magnolias, the results indicated that divergence time of *M. hypoleuca* was earlier than the other two. This provided further evidence of the evolutionary relationship of the three Magnolias and contributed to the domestication, and utilization of Magnoliaceae plants. Meanwhile, the Miocene was a period with globally warmer climates than those in the preceding Oligocene, or the subsequent Pliocene (Wolfe *et al.* 1984). Among them, the middle Miocene warming period was from 13 to 18 MYA (Poortvliet *et al.* 2015). Some researchers found that the Miocene floras had many elements in common with the modern mesophytic floras of eastern Asia and eastern North America, supporting the proposal that the divergence of the modern north temperate elements occurred during that period. There are some else to support this proposal. such as, Manta, devil rays, *Pterocarya* (Poortvliet *et al.* 2015) and Magnoliaceae in the Northern Hemisphere and so on. This study also indicated that time-division of three Magnolias might occur in the period of middle Miocene warming period.

**Conclusions**

In this study, we reported and analyzed the cp complete genomes of *M. hypoleuca*, which is the components of deciduous broad-leaved forest in the cold temperate zone. And we compared cp complete genomes of three Magnolias, the results revealed that the genome size of *M. hypoleuca* was smaller than *M. officinalis* subsp.*biloba* and *M. officinalis*. Moreover, we found the distinct difference of three Magnolias from the content and length of introns and exons in genes, and the types and quantity of functional genes. We detected the GC content of three Magnolias cp genome was 39.2%, comprising 8 rRNA and 37 tRNA. In addition, the result showed that *M. hypoleuca* had 11 and 8 more functional genes than *M. officinalis* and *M. officinalis* subsp.*biloba*, respectively. And through constructing the relationship between phylogenetic and divergence time of the three Magnolias, it indicated that the divergence time of three Magnolias might take place during the middle Miocene (13-18 Ma).

**Ackowledgements**

The study was supported financially by Sichuan Provincial Administration of Traditional Chinese Medicine Program (2018QN001); Innovation Team of Sichuan Science and Technology Department (2017TD0001); Genomics Innovation Platform of Chinese Medicine Resources with Southwestern Characteristics (2020ZYD058).

**Author contributions**MZ and SZ made the write up and made illustrations, GG tatistically analyzed the data, FH, CP and JG planed the experiments, PY, LZ, BR, LW, XS, XY and GZ interpreted the results.

**References**

Bhaskaran H, Rodriguez-Hernandez A, Perona JJ (2012). Kinetics of tRNA folding monitored by aminoacylation. *RNA* 18(3):569-80

Bolger, A. M., Lohse, M., Usadel, B (2014). Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics* 30: 2114–2120

Chacón J, Luebert F, Selvi F, Cecchi L, Weigend M (2019). Phylogeny and historical biogeography of Lithospermeae (Boraginaceae): Disentangling the possible causes of Miocene diversifications. Mol *Phylogenet Evol* 141:106-626

Chen K (2019). Study on the structural variation of chloroplast genome in Magnoliaceae and screening of hypervariable genes. *Zhejiang Agri University* 7-15.

Chen S, Xu J, Liu C, Zhu Y, Nelson DR, Zhou S, Li C, Wang L, Guo X, Sun Y, Luo H, Li Y, Song J, Henrissat B, Levasseur A, Qian J, Li J, Luo X, Shi L, He L, Xiang L, Xu X, Niu Y, Li Q, Han MV, Yan H, Zhang J, Chen H, Lv A, Wang Z, Liu M, Schwartz DC, Sun C (2012). Genome sequence of the model medicinal mushroom *Ganoderma lucidum*. *Nat Commun* 3: 913

Chen SY, Wu T, Ma HF, Fu YP, Zhu YF, Hao JB, Jiang H, Sima YK (2020). The complete chloroplast genome sequence of *Lirianthe hodgsonii*, a tree species of Magnoliaceae as least concern. *Mitochondrial DNA B Resour* 5(3):3064-3066

Conant, G. C., Wolfe, K. H (2008). GenomeVx: simple web-based creation of editable circular chromosome maps. *Bioinformatics* 24: 861

Daniell H, Lin CS, Yu M, Chang WJ (2016). Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol* 17(1):134

Felsenstein J (1988). Phylogenies from molecular sequences: inference and reliability. *Annu Rev Genet* 22:521-65

Feng DJ, Liu X, Li XG, Zhu Z (2002). Relationship between tRNA abundance and gene expression. *Chin J of Biol* 06: 4-8.

Gamble PE, Mullet JE (1989). Blue light regulates the accumulation of two *psbD-psbC* transcripts in barley chloroplasts. *EMBO J* 7(5):1289-97

Gamble PE, Sexton TB, Mullet JE (1988). Light-dependent changes in *psbD* and *psbC* transcripts of barley chloroplasts: accumulation of two transcripts maintains *psbD* and *psbC* translation capability in mature chloroplasts. *EMBO J* 8(10):2785-94

Gutiérrez-Ortega JS, Salinas-Rodríguez MM, Martínez JF, Molina-Freaner F, Pérez-Farrera MA, Vovides AP, Matsuki Y, Suyama Y, Ohsawa TA, Watano Y, Kajita T (2018). The phylogeography of the cycad genus Dioon (Zamiaceae) clarifies its Cenozoic expansion and diversification in the Mexican transition zone. *Ann Bot* 121(3):535-548

He SL, Tian Y, Yang Y, Shi CY (2020). Chloroplast genome and phylogenetic analyses of *Morus alba* (Moraceae). *Mitochondrial DNA B Resour* 5(3):2203-2204

He SL, Tian Y, Yang Y, Shi CY (2020). Chloroplast genome and phylogenetic analyses of *Poncirus trifoliata* (Rutaceae). *Mitochondrial DNA B Resour* 5(3):2205-2206

Iwasaki T, Aoki K, Seo A, Murakami N (2012). Comparative phylogeography of four component species of deciduous broad-leaved forests in Japan based on chloroplast DNA variation. *J Plant Res* 125(2):207-21

Jansen RK, Raubeson LA, Boore JL, dePamphilis CW, Chumley TW, Haberle RC, Wyman SK, Alverson AJ, Peery R, Herman SJ, Fourcade HM, Kuehl JV, McNeal JR, Leebens-Mack J, Cui L (2005). Methods for obtaining and analyzing whole chloroplast genome sequences. *Methods Enzymol* 395:348-84

Kawahara T, Tomono T, Hamauzu Y, Tanaka K, Yasui H (2014). Inhibitory Effect of a Hot-Water Extract of Leaves of Japanese Big-Leaf Magnolia (*Magnolia obovata*) on Rotavirus-Induced Diarrhea in Mouse Pups. *Evid Based Complement Alternat Med* 2014:365831

Kim HT, Chung MG, Kim KJ (2014). Chloroplast genome evolution in early diverged leptosporangiate ferns. *Mol Cells* 37(5):372-82

Kinoshita G, Nunome M, Kryukov AP, Kartavtseva IV, Han SH, Yamada F, Suzuki H (2019). Contrasting phylogeographic histories between the continent and islands of East Asia: Massive mitochondrial introgression and long-term isolation of hares (Lagomorpha: Lepus). *Mol Phylogenet Evol* 136:65-75

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol* 35(6):1547-1549

Kwon Oh Jung, Oh Choong Hyeon (2015). Naturalization of Landscaping Woody Plant, *Magnolia obovata* Potentially Invasive Species. *J of mt Sci* 12(01):30-38

Li DM, Lv FB, Zhu GF, Cao JX, Sun YB, Liu HL, Wang Z (2012). Correlation Analysis of Codon Usage in Chloroplast Genome of *Oncidium*. *Guangdong Agr sci* 39(10): 61-65.

Li X, Li Y, Zang M, Li M, Fang Y (2018). Complete Chloroplast Genome Sequence and Phylogenetic Analysis of *Quercus acutissima*. *Int J Mol Sci* 19(8):2443

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

Liu S, Wang Z, Wang H, Su Y, Wang T (2020). Patterns and Rates of Plastid *rps12* Gene Evolution Inferred in a Phylogenetic Context using Plastomic Data of Ferns. *Sci Rep* 10(1):9394

Liu XF, Zhu GF, Li DM, Wang XJ (2019). Complete chloroplast genome sequence and phylogenetic analysis of *Spathiphyllum* 'Parrish'. *PLoS One* 14(10):e0224038

Liu XY, Zhou GZ, Wang XY (2006). Research on Physiology of *Magnolia hypoleuca* Cold Resistance. *J Shenyang Agr University* 37(6): 845-848.

Liu ZC, Su DY (1985). Effects of high temperature on chloroplast ribosome and chloroplast protein biosynthesis in wheat. *J of Integrat Plant Biol* 1985(01): 63-67.

Luo H, Wu H, Yu X, Zhang X, Lu Y, Fan J, Tang L, Wang Z (2019). A review of the phytochemistry and pharmacological activities of *Magnoliae officinalis* cortex. *J Ethnopharmacol* 236:412-442.

Mao K, Milne RI, Zhang L, Peng Y, Liu J, Thomas P, Mill RR, Renner SS (2012). Distribution of living Cupressaceae reflects the breakup of Pangea. *Proc Natl Acad Sci U S A* 109(20):7793-8

Maruyama, S., Isozaki, Y ., Kimura, G., Terabayashi, M (1997). Paleogeographic maps of the Japanese Islands: Plate tectonic synthesis from 750 Ma to the present. *Island Arc* 6: 121–142

Matthes, N., Westphal, K., Haldemann, C., Egert, M., Jokisch, C., Speck, B (2020). Validation of a modified CTAB method for DNA extraction from protein-rich maize feedstuffs. *J of Consumer Protection and Food Safety* 4(24): 23-27

Motokawa, Masaharu (2017). "Land Emergence" and "Elevation Shift" Affect Diversification: A New Perspective Toward Understanding the High Species Diversity of Terrestrial Animals in Japan. Species Diversity of Animals in Japan. *Springer Japan* 1:3-23

Mustroph A, Barding GA Jr, Kaiser KA, Larive CK, Bailey-Serres J (2014). Characterization of distinct root and shoot responses to low-oxygen stress in Arabidopsis with a focus on primary C- and N-metabolism. *Plant Cell Environ* 37(10):2366-80

Nie ZL, Wen J, Azuma H, Qiu YL, Sun H, Meng Y, Sun WB, Zimmer EA (2008). Phylogenetic and biogeographic complexity of Magnoliaceae in the Northern Hemisphere inferred from three nuclear data sets. *Mol Phylogenet Evol* 48(3):1027-40

PARKS, C . R . , AND J. F. WENDEL (1990). Molecular divergence between Asian and North American species of *Liriodendron* (Magnoliaceae) with implications for interpretation of fossil floras. *American J of Botany* 77(10): 1736-1748

Ping JY, Zhu M, Su Yingjuan, Wang ting (2020). Molecular evolution of the *rps12* gene in the chloroplast of ferns. *Acta Plant Sci* 38(1): 1-9.

Poortvliet M, Olsen JL, Croll DA, Bernardi G, Newton K, Kollias S, O'Sullivan J, Fernando D, Stevens G, Galván Magaña F, Seret B, Wintner S, Hoarau G (2015). A dated molecular phylogeny of manta and devil rays (Mobulidae) based on mitogenome and nuclear sequences. *Mol Phylogenet Evol* 83:72-85

QIU, Y .-L., M. W . CHASE, AND C. R. PARKS (1995). A chloroplast DNA phylogenetic study of the eastern Asia: eastern North America disjunct section Rytidospermum of *Magnolia* (Magnoliaceae). *American J of Botany* 82: 1582–1588

Ravi V, Khurana JP, Tyagi AK, Khurana P (2008). An update on chloroplast genomes. *Plant Syst Evol* 271:101–122

Richardson JE, Pennington RT, Pennington TD, Hollingsworth PM (2001). Rapid diversification of a species-rich genus of neotropical rain forest trees. *Sci* 293: 2242–2245

Roman MG, Gangitano D, Houston R (2019). Characterization of new chloroplast markers to determine biogeographical origin and crop type of *Cannabis sativa*. *Int J Legal Med* 133(6):1721-1732

Sarrica A, Kirika N, Romeo M, Salmona M, Diomede L (2018). Safety and Toxicology of Magnolol and Honokiol. *Planta Med* 84(16):1151-1164

Wei D, Yan YH, Liu YP, Wu QH, Chen J, Pei J (2020). Comparative transcriptome and proteome profiling of chlorophyll metabolism pathway in four types of *Magnolia officinalis*. *Zhongguo Zhong Yao Za Zhi* 45(16):3826-3836.

Wen, J (1999). Evolution of Eastern Asian and Eastern North American disjunct distributions in flowering plants. *Annu. Rev. Ecol.* *Syst* 30: 421–455

Wolfe, J.A (1985). Distribution of major vegetational types during the Tertiary: the carbon cycle and atmospheric CO2natural variations archean to present; Proceedings of the Chapman Conference on Natural Variations in Carbon Dioxide and the Carbon Cycle, Tarpon Springs, FL, January 9–13, 1984 (A86-39426 18–46). American Geophysical Union, Washington, DC,; 357–375.

Xu BQ, An MM, Jiang K, Xu LL, Yang HY, Zhou HB (2015). Cloning and functional analysis of the chloroplast *psbD* gene of *Phyllostachys japonicus*. *J Zhejiang Ag Forestry University* 32(4): 557-565.

Xu F, Zhang WW, Tang Y, Wang Y, Cheng SY (2010). Cloning and expression of *petD* gene of *Ginkgo* chloroplast. *Plant Physiol Commun* 46(01): 37-41

Xu J, Chu Y, Liao B, Xiao S, Yin Q, Bai R, Su H, Dong L, Li X, Qian J, Zhang J, Zhang Y, Zhang X, Wu M, Zhang J, Li G, Zhang L, Chang Z, Zhang Y, Jia Z, Liu Z, Afreh D, Nahurira R, Zhang L, Cheng R, Zhu Y, Zhu G, Rao W, Zhou C, Qiao L, Huang Z, Cheng YC, Chen S (2017). Panax ginseng genome examination for ginsenoside biosynthesis. *Gigascience* 6(11):1-15

Xu J, Feng D, Song G, Wei X, Chen L, Wu X, Li X, Zhu Z (2003). The first intron of rice EPSP synthase enhances expression of foreign gene. *Sci China C Life Sci* 46: 561–569

Yan L, Lai X, Li X, Wei C, Tan X, Zhang Y (2015). Analyses of the complete genome and gene expression of chloroplast of sweet potato [*Ipomoea batata*]. *PLoS One* 10(4):e0124083

Yang F (2019). Sequencing and structural analysis of chloroplast genome of *Rosa qilixiang*. *Genomics Appl Biol* 38(8): 3586-3594.

Yang L (2000). The contents of magnolol, honokiol and ß-eucalyptol in different varieties and producing areas of *Magnolia officinalis* and their correlation with the morphological characteristics of magnolia bark. *Foreign Med* 54(2): 61- 69.

Yi T, Miller AJ, Wen J (2004). Phylogenetic and biogeographic diversification of Rhus (Anacardiaceae) in the Northern Hemisphere. *Mol Phylogenet Evol* 33(3):861-79

Youn U, Chen QC, Lee IS, Kim H, Yoo JK, Lee J, Na M, Min BS, Bae K (2008). Two new lignans from the stem bark of *Magnolia obovata* and their cytotoxic activity. *Chem Pharm Bull (Tokyo)* 56(1):115-7

Zachos, J., Pagani, M., Sloan, L., Thomas, E., Billups, K. (2001). Trends, rhythms, and aberrations in global climate 65 Ma to present. *Sci* 292: 686–693

Zhou YH, Hou BK, Shi DQ, Xiao YH, Shi R, Hu ZM, Chen ZH (2001). Cloning and sequence analysis of *3 'rps12*, *rps7* and *ndhB* gene fragments in poplar. *Chin J of Biochem Mol Biol* 05: 606-616.



**Fig. 1**: Three Magnolias cp gene cycles a M. hypoleuca cp genome. b M. officinalis subsp.biloba cp genome. c M. officinalis cp genome. Genes shown outside the circle are transcribed clockwise and those inside counterclockwise. Genes belonging to different functional groups are color-coded



**Fig.2**: Fossil-calibrated phylogenetic tree of three Magnolias plants Species tree constructed using four cpDNA of 14 species in Magnoliaceae and one outgroup, *Ginkgo biloba*. The tree is a clade credibility tree; numbers along the branches of the tree are the average estimated divergence times.

**Table 1**: Basic information of three Magnolias cp genes

| Name | Length(bp) | GC(%) | Total number of genes | Coding region | rRNA | tRNA | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| *Magnolia* *officinalis* | 160 183 | 39.2 | 126 | 82 | 8 | 37 |
| *Magnolia officinalis* subsp.*biloba* | 160 099 | 39.2 | 128 | 84 | 8 | 37 |
| *Magnolia hypoleuca* | 160 051 | 39.2 | 137 | 97 | 8 | 37 |

**Table 2**: Comparison of introns and exons of three Magnolias cp genome

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | *Magnolia* *officinalis* | | | | | *Magnolia officinalis* subsp.*biloba* | | | | | *Magnolia hypoleuca* | | | | |
| gene | EpI | InI | EpII | InII | EpIII | EPI | InI | EpII | InII | EpIII | EpI | InI | EpII | InII | EpIII |
| *trnK-UUU* | 37 | 2498 | 35 |  |  | 37 | 2492 | 35 |  |  | 36 | 2491 | 34 |  |  |
| *rps16* | 42 | 824 | 246 |  |  | 42 | 824 | 246 |  |  | 39 | 823 | 220 |  |  |
| *atpF* | 145 | 709 | 410 |  |  | 145 | 707 | 410 |  |  | 144 | 708 | 410 |  |  |
| *rpoC1* | 432 | 740 | 1614 |  |  | 432 | 734 | 1614 |  |  | 431 | 733 | 1614 |  |  |
| *ycf3* | 124 | 733 | 232 | 729 | 141 | 124 | 734 | 232 | 727 | 141 | 226 | 732 | 232 | 727 | 143 |
| *trnL-UAA* | 35 | 491 | 50 |  |  | 35 | 491 | 50 |  |  | 34 | 490 | 49 |  |  |
| *trnV-UAC* | 39 | 584 | 37 |  |  | 39 | 585 | 37 |  |  | 38 | 583 | 36 |  |  |
| *rps12* | 227 |  | 227 |  | 29 | 227 |  | 227 |  | 29 | 113 | 535 | 231 |  | 25 |
| *clpP* | 71 | 786 | 291 | 629 | 244 | 71 | 781 | 291 | 628 | 244 | 70 | 776 | 290 | 627 | 245 |
| *petB* | 6 | 784 | 642 |  |  | 7 | 784 | 641 |  |  | 5 | 783 | 641 |  |  |
| *petD* |  |  |  |  |  |  |  |  |  |  | 7 | 700 | 472 |  |  |
| *rpl16* |  |  |  |  |  |  |  |  |  |  | 8 | 968 | 398 |  |  |
| *ndhB* | 775 | 700 | 758 |  |  | 775 | 700 | 758 |  |  | 776 | 700 | 755 |  |  |
| *trnI-GAU* | 42 | 937 | 35 |  |  | 42 | 936 | 35 |  |  | 41 | 936 | 34 |  |  |
| *trnA-UGC* | 38 | 800 | 35 |  |  | 38 | 800 | 35 |  |  | 37 | 798 | 34 |  |  |
| *ndhA* | 553 | 1082 | 539 |  |  | 553 | 1102 | 539 |  |  | 552 | 1103 | 540 |  |  |
| *trnA-UGC* | 38 | 800 | 35 |  |  | 38 | 800 | 35 |  |  | 37 | 798 | 34 |  |  |
| *trnI-GAU* | 42 | 937 | 35 |  |  | 42 | 936 | 35 |  |  | 41 | 935 | 34 |  |  |
| *rps12* | 227 |  | 29 |  | 29 | 227 |  | 29 |  |  |  | 532 | 231 |  | 25 |
| *ndhB* | 775 | 700 | 758 |  |  | 775 | 700 | 758 |  |  | 776 | 699 | 755 |  |  |
| *rpl2* | 385 | 661 | 431 |  |  | 385 | 649 | 443 |  |  | 396 | 660 | 430 |  |  |
| *trnG-UCC/ trnG-GCC* | 24 | 770 | 48 |  |  | 24 | 767 | 48 |  |  | 23 | 768 | 47 |  |  |

Note: Ribosomal subunit (rps); RNA polymerase (rpo); Open reading frame (ycf); Protease gene (clp); Cytochrome (pet); Dehydrogenase complex (ndh); Ribosome large subunit (rpl); Expressed region (ep); Intron (in); Photosynthesis (psa)

**Table 3**: Comparison of cp functional genes of three Magnolias plants

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Gene function | Gene type | *Magnolia hypoleuca* | *Magnolia* *officinalis* | *Magnolia officinalis* subsp.*biloba* |
| Photosynthesis | ATP synthase subunit | atpA, atpA, atpB, atpE, atpF, atpF, atpH, atpH, atpI, atpI | atpA, atpB, atpE, atpF, atpH, atpI | atpA, atpB, atpE, atpF, atpH, atpI |
| Photosynthetic System I Subunit | psaA, psaB, psaC, psaI, psaJ | psaA, psaB, psaC, psaI, psaJ | psaA, psaB, psaC, psaI, psaJ |
| Photosynthetic System Ⅱ Subunit | psbA, psbA, psbB, psbC, psbC, psbD, psbD, psbE, psbF, psbI, psbJ, psbK, psbL, psbM, psbM, psbN, psbT, psbZ, ycf3 | psbA, psbB, psbC, psbD, psbE, psbF, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ, ycf3 | psbA, psbB, psbC, psbD, psbE, psbF, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ, ycf3 | |
| Self copy | NADH dehydrogenase subunit | ndhA, ndhB, ndhB, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK | ndhA, ndhB, ndhB, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK | ndhA, ndhB, ndhB, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK | |
| Cytochrome b/f complex | petA, petB, petD, petG, petL, petN, petN | petA, petB, petD, petG, petL, petN | petA, petB, petD,petG, petL, petN | |
| Ribulose diphosphate carboxylase subunit | rbcL | rbcL | rbcL | |
| Large ribosomal subunit | rpl14, rpl16, rpl2, rpl20, rpl23, rpl32, rpl33, rpl36 | rpl14, rpl16, rpl2, rpl2, rpl20, rpl23, rpl23, rpl32, rpl33, rpl36 | rpl14, rpl16, rpl2, rpl2, rpl20, rpl23, rpl23, rpl32, rpl33, rpl36 | |
| Small ribosomal subunit | rps11, rps12, rps12, rps14, rps16, rps18, rps19, rps2, rps2, rps4, rps7, rps7, rps8 | rps11, rps12, rps12,rps14, rps16, rps18, rps19, rps2, rps4, rps7, rps7, rps8 | rps11, rps12, rps12, rps14, rps14 rps16, rps18, rps19, rps2, rps3, rps4, rps7, rps7, rps8 | |
| Other gene | DNA-dependent RNA polymerase subunit | rpoA, rpoB, rpoB, rpoC1, rpoC1, rpoC2, rpoC2 | rpoA, rpoB, rpoC1, rpoC2 | rpoA, rpoB, rpoC1, rpoC2 | |
| Acetyl-CoA carboxylase subunit | accD | accD | accD | |
| C-type cytochrome synthase | ccsA | ccsA | ccsA | |
| Membrane Protein | cemA | cemA | cemA | |
| Protease | clpP | clpP | clpP | |
| Translation initiation factor | infA | infA | infA | |
| Mature enzyme | matK | matK | matK | |
| Unknown functional gene | Conservative open reading frame | ycf1, ycf4 | ycf1, ycf2, ycf2, ycf4 | ycf1, ycf2, ycf2, ycf4 | |

Note: Photosynthesis (psa); Acetyl-coa carboxylase (acc); Ribulose bisphosphate carboxylase (rbc); TypeC cytochrome synthesis gene (ccs); Membrane protein gene (cem); Mature enzyme gene (mat) ; Photosystem (psb)