

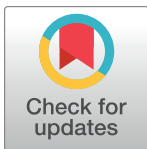
RESEARCH ARTICLE

Complete chloroplast genome sequences of Myristicaceae species with the comparative chloroplast genomics and phylogenetic relationships among them

Changli Mao , Fengliang Zhang, Xiaoqin Li, Tian Yang, Qi Zhao, Yu Wu*

Yunnan Institute of Tropical Crops, Xishuangbanna, China

* hhyw20030105@126.com



Abstract

Background

Myristicaceae was widely distributed from tropical Asia to Oceania, Africa, and tropical America. There are 3 genera and 10 species of Myristicaceae present in China, mainly distributed in the south of Yunnan Province. Most research on this family focuses on fatty acids, medicine, and morphology. Based on the morphology, fatty acid chemotaxonomy, and a few of molecular data, the phylogenetic position of *Horsfieldia pandurifolia* Hu was controversial.

Results

In this study, the chloroplast genomes of two *Knema* species, *Knema globularia* (Lam.) Warb. and *Knema cinerea* (Poir.) Warb., were characterized. Comparing the genome structure of these two species with those of other eight published species, including three *Horsfieldia* species, four *Knema* species, and one *Myristica* species, it was found that the chloroplast genomes of these species were relatively conserved, retaining the same gene order. Through sequence divergence analysis, there were 11 genes and 18 intergenic spacers were subject to positive selection, which can be used to analyze the population genetic structure of this family. Phylogenetic analysis showed that all *Knema* species were clustered in the same group and formed a sister clade with *Myristica* species support by both high maximum likelihood bootstrap values and Bayesian posterior probabilities; among *Horsfieldia* species, *Horsfieldia amygdalina* (Wall.) Warb., *Horsfieldia kingii* (Hook.f.) Warb., *Horsfieldia hainanensis* Merr. and *Horsfieldia tetratapa* C.Y.Wu. were grouped together, but *H. pandurifolia* formed a single group and formed a sister clade with genus *Myristica* and *Knema*. Through the phylogenetic analysis, we support de Wilde' view that the *H. pandurifolia* should be separated from *Horsfieldia* and placed in the genus *Endocomia*, namely *Endocomia macrocoma* subsp. *prainii* (King) W.J.de Wilde.

Conclusion

The findings of this study provide a novel genetic resources for future research in Myristicaceae and provide a molecular evidence for the taxonomic classification of Myristicaceae.

OPEN ACCESS

Citation: Mao C, Zhang F, Li X, Yang T, Zhao Q, Wu Y (2023) Complete chloroplast genome sequences of Myristicaceae species with the comparative chloroplast genomics and phylogenetic relationships among them. PLoS ONE 18(3): e0281042. <https://doi.org/10.1371/journal.pone.0281042>

Editor: Michael Scott Brewer, East Carolina University, UNITED STATES

Received: April 21, 2022

Accepted: January 14, 2023

Published: March 20, 2023

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0281042>

Copyright: © 2023 Mao et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All complete chloroplast genome sequences are available from the NCBI database (<https://www.ncbi.nlm.nih.gov/>)

under accession numbers MK285561-MK285565, MN683753-MN683756, and MH445411.

Funding: This work was supported by the National Natural Science Foundation of China (No.31960289). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

The family Myristicaceae includes about 20 genera and 500 species of evergreen trees distributed from tropical Asia and Pacific islands to Africa and tropical America; there are 3 genera and 10 species of Myristicaceae in China [1]. The seeds of some Myristicaceae plants contain about 57.39% solid oil and are mainly composed of Myristic acid and tetradecenoic acid, such that the relative average content of these two acids exceeds 90% [2, 3]. Since their seeds are rich in C14 fatty acid [4], which can be used as a condensing agent in medical, beauty, cosmetics, and other industrial products, they are considered a high-quality raw material [3–6], implying a significant economic opportunity in exploring the resource of Myristaceae species.

Because of the use of their oil and wood in industry, Myristicaceae plants have attracted increasing attention from researchers. The studies conducted on this family have mainly focused on their fatty acid content [2], chemical composition [7], and taxonomy [8–12]. We conducted a field survey for this family in Yunnan province, China and found that some species appear in small population with few individuals, such as only 1 *Horsfieldia pandurifolia* Hu (location: 99° 46.956'E, 23° 12.521'N, altitude of 1 000m) individual has been found in the North of Xiaohei River valley, Shuangjiang county, Yunnan. The original large population of *H. pandurifolia* has been destroyed and only small isolated population exist in the upper of the distribution area [13]. so, it is important to conserve the species using the effective measure. However, the systematic position of some species within the family remains controversial, such as *H. pandurifolia*. Therefore, the precise identification and delimitation of the species was the key.

H. pandurifolia was first named by Hu in 1963 [14], and recorded as *H. pandurifolia* Hu in Flora Yunnanica [15], but the *H. pandurifolia* was incorporated into the *Horsfieldia Prainii* (King) Warb., and classified into *Horsfieldia macrocoma* (Miq) Warb. as a subspecies, and established genus *Endocomia* as a model species by de Wilde in 1984, namely *Endocomia macrocoma* ssp. *prainii* [9]. In 2004, Ye argued that the taxonomic boundaries of the genera *Horsfieldia* and *Endocomia* was not obvious, annulled *Endocomia*, restoring *H. macrocoma* [12], and in Flora of China (2008), the genus *Endocomia* was also annulled and *H. pandurifolia* was named as *H. prainii* [1]. In recent years, based on results of molecular biology approaches, the systematic position of *H. pandurifolia* was also differently suggested by Wu [8] and Cai [16]. it is argued about *H. pandurifolia* position, so it is necessary to study the relationship of Myristicaceae, especially for *H. pandurifolia*.

The chloroplast organelle is the most noticeable feature in plants, and its plastome is conserved than mitochondrial and nuclear genomes. Due to the low rate of nucleotide substitution, the chloroplast genome is used frequently in phylogeny studies [17], at the same time, the chloroplast genome can be provide large sequence information, so it serves as good candidates for high resolution DNA barcoding [16]. With the advent of sequencing technology such as Illumina, Nanopore, there are increasing reports of complete chloroplast genome for plant phylogenetic analysis, such as in Acanthoideae [18], Magnoliaceae [19], and Zingiberaceae [20] etc. including the phylogenetic relationships study on the Chinese *Horsfieldia* based on the chloroplast genome analysis [16]. However, the analysis of phylogenetic relationship for the species of Myristicaceae that distributed in China using chloroplast genome has not been reported.

In this study, we sequenced, assembled, and characterized the chloroplast genome of two species of Myristicaceae, *Knema globularia* (Lam.) Warb. and *Knema cinerea* (Poir.) Warb., using the Roche/454 sequencing platform. To analyze the organization, gene contents, patterns of nucleotide substitution, simple sequence repeats (SSRs), and phylogenetic relationships, the previously reported chloroplast genomes of eight Myristicaceae species were considered for comparative analyses with the two newly assembled chloroplast genomes. Our study aims were as follows: (1) evaluate the variations within the 10 species and the structural diversity of

the chloroplast genome among Myristicaceae species; (2) upgrade our understanding of the application value of the chloroplast genome of Myristicaceae and provide novel genetic resources for future research in Myristicaceae; and (3) to provide a molecular evidence for the taxonomic classification of Myristicaceae.

Materials and methods

Plant material and DNA extraction

The species used in this study were identified according to the records in Flora of China. The samples information, including the two newly sequenced species and eight reported species, were listed in the Table 1, and the species names in this table were derived from the Flora Republicae Popularis Sinicae and Flora of China. The collection sites included Nangunhe National Nature Reserve, Xishuangbanna National Nature Reserve and Yunnan Institute of Tropical Crops (YITC), Yunnan, China. All the live specimen was planting on site, so, the herbarium were not been made. Fresh leaves of Myristicaceae plants were sampled and immediately dried with silica gel, and total genomic DNA per germplasm was extracted from 100 mg dried leaves using the DNeasy Plant Mini Kit (QIAGEN, Valencia, CA, USA).

DNA sequencing

DNA concentrations were quantified using Life Invitrogen Qubit® 3.0 (Life, Invitrogen, USA). The total genomic DNA per germplasm with over 30 ng μL^{-1} were used for subsequent steps. Purified DNA (5 mg) was sheared by nebulization with compressed nitrogen gas, which yielded DNA fragments of 400–800 bp in length, with single strands of DNA being recovered by denaturing treatment after the modification of terminal repair and specific adaptor sequence connections. Specific proportions of single-stranded DNA libraries were immobilized on DNA capture magnetic beads for Emulsion PCR and then sequenced in GS FLX reagents. After reaction, 4–6 billion base information sites were obtained.

Table 1. The information of ten samples used in this analysis.

Sample No.	Sample name	Collector number	Date of collection	GPS coordinates	Site of specimen preservation	Accession number in GenBank
1	<i>Horsfieldia amygdalina</i> (Wall.) Warb.	20140443	2019-02-13	100°47'E 22°00'N	YITC	MK285561
2	<i>Horsfieldia pandurifolia</i> Hu	20090402	2019-02-13	100°47'E 22°00'N	YITC	MH445411
3	<i>Horsfieldia kingii</i> (Hook.f.) Warb.	20140404	2019-02-13	100°47'E 22°00'N	YITC	MK285562
4	<i>Knema elegans</i> Warb.	20140606	2019-02-26	100°07.38'E 21°52.82'N	Xishuangbanna National Nature Reserve	MK285564
5	<i>Knema cinerea</i> (Poir.) Warb.	20140469	2019-02-27	101°13.18'E 21°54.00'N	Xishuangbanna National Nature Reserve	MN683756
6	<i>Knema furfuracea</i> (Hook.f. & Thomson) Warb.	20140428	2019-02-25	100°02.39'E 21°13.77'N	Xishuangbanna National Nature Reserve	MK285563
7	<i>Knema linifolia</i> (Roxb.) Warb.	20140475	2019-02-13	99°04.39'E 23°16.43'N	Nangunhe National Nature Reserve	MN683753
8	<i>Knema conferta</i> (King) Warb.	20161101	2019-02-14	99°01.013'E 23° 16.943'N	Nangunhe National Nature Reserve	MN683754
9	<i>Knema globularia</i> (Lam.) Warb.	20140487	2019-02-27	101°33.77'E 21° 36.75'N	Xishuangbanna National Nature Reserve	MN683755
10	<i>Myristica yunnanensis</i> Y.H.Li	20140492	2019-02-13	100°47'E 22°00'N	YITC	MK285565

<https://doi.org/10.1371/journal.pone.0281042.t001>

Genome assembly and annotation

The chloroplast genome was assembled using CLC Genomic Workbench v3.6 (<http://www.clcbio.com>), the contigs were aligned ($\geq 90\%$ similarity and query coverage) and ordered, according to the reference chloroplast genome of *K.elegans* which was reported by author. The genes in the chloroplast genome were predicted using the Dual Organellar GenoMe Annotator (DOGMA) program [21]. The protein-encoding genes, location of the ribosomal RNA (rRNA) and transfer RNA (tRNA) genes were annotated using BLASTX and BLASTN searches. To accurately determine the boundaries between introns and exons and the starting and stopping codon positions of protein-coding genes, the annotated results were manually examined, and the codon position was adjusted by comparing them with homologous genes in the reference genomes, namely *H. pandurifolia*, *H. kingii* and *K. elegans*, based on phylogenetic proximity. The chloroplast genome map was drawn using Genome Vx software [22], and the chloroplast genome sequences of *K.globularia* and *K.cinerea* have been deposited to the GenBank, the accession numbers were listed in Table 1.

Repeats analysis

Microsatellite Identification Tool (Misa-web-IPK Gatersleben (ipk-gatersleben.de)) [23] was used to identify the SSRs with the following parameters: 10 for mononucleotide, 5 for dinucleotides, 4 for trinucleotides, and 3 for more than 4-base SSR motifs. Long-repeat analyses of 10 species was done using the program REPuter (<https://bibiserv.cebitec.uni-bielefeld.de/reputer>) [24] involving the default parameters.

Genome comparison

Performing comparative analysis for the 10 Myristaceae species, the chloroplast genome of *Liriodendron tulipifera* L. from the same order Magnoliales was downloaded from GenBank (NC_008326.1, https://ncbi.nlm.nih.gov/search/all/?term=NC_008326.1) and used as a reference. The sequence identity of 11 species chloroplast genomes were plotted using the mVISTA program with the LAGAN mode [25]. To detect the rearrangement and inverse evolutionary events, multiple genome alignments were conducted using the progressive mauve algorithm [26], as implemented in the Geneious software package (Biomatters, Auckland, New Zealand), in the Mauve options, change the alignment algorithm to MCM (the Mauve Contig Mover) alignment. The borders of small single-copy (SSC), large single copy (LSC), and inverted repeat (IR) regions among Myristicaceae species were visually displayed and compared using Irscope [27] based on these species annotation GenBank (.gb) files. To detect the hotspots of the intergenetic divergence, 74 coding sequences, and 41 intergenic space sequences were extracted for each species using Phylosuite ver. 1.1.152 [28] and aligned in batches using MAFFT [29], using the—auto = strategy and the codon alignment mode. Following this step, the nucleotide diversity (Pi) value was calculated for each of the 115 loci using DnaSP ver.6.12.03 [30]. To determine whether coding genes in 10 species had selective pressure during evolution, DnaSP software [30] was used to calculated synonymous (dS) and nonsynonymous (dN) values of coding genes of 10 species as well.

Phylogenetic analysis

To obtain a more comprehensive results, 16 previously reported chloroplast genomes of Myristicaceae by Cai et al. [16] were also downloaded from GenBank (MN486685, MN486686, and MN495958–MN495971) and performed the phylogenetic analyses together with these 10 sequences in this study (the accession numbers are listed in Table 1), and using *L.tulipifera* as

outgroup. After the alignment of all sequences with MAFFT software, the phylogenetic analyses were performed using maximum likelihood (ML) and Bayesian inference (BI) with MEGA 7.0 [31] and MrBayes version 3.2.6 [32] respectively. Conditions were set as bellow: using the model of GTR+ Γ for ML, and GTR+ Γ +I for BI analysis, the node support values were given in the form of posterior probability (PP) and bootstrap value (BV), when perform the BI analysis, the Markov chain Monte Carlo (MCMC) was run for 2,000,000 generations with two parallel searches using four chains.

Results

Characteristics of Myristicaceae chloroplast genome

The 10 complete chloroplast DNA sequences ranged from 154,527 bp (*K. furfuracea*) to 155,923 bp (*M. yunnanensis*), and they all displayed a quadripartite structure typical of angiosperms, which consisted of an SSC (15,072–30,998 bp), an LSC (86,188–92,561 bp), and a pair of IRs (37,754–48,154 bp) (Fig 1 and Table 2). The base composition of the 10 chloroplast genome sequences were analyzed and counted, and the GC content ranged from 39.19% (*M.*

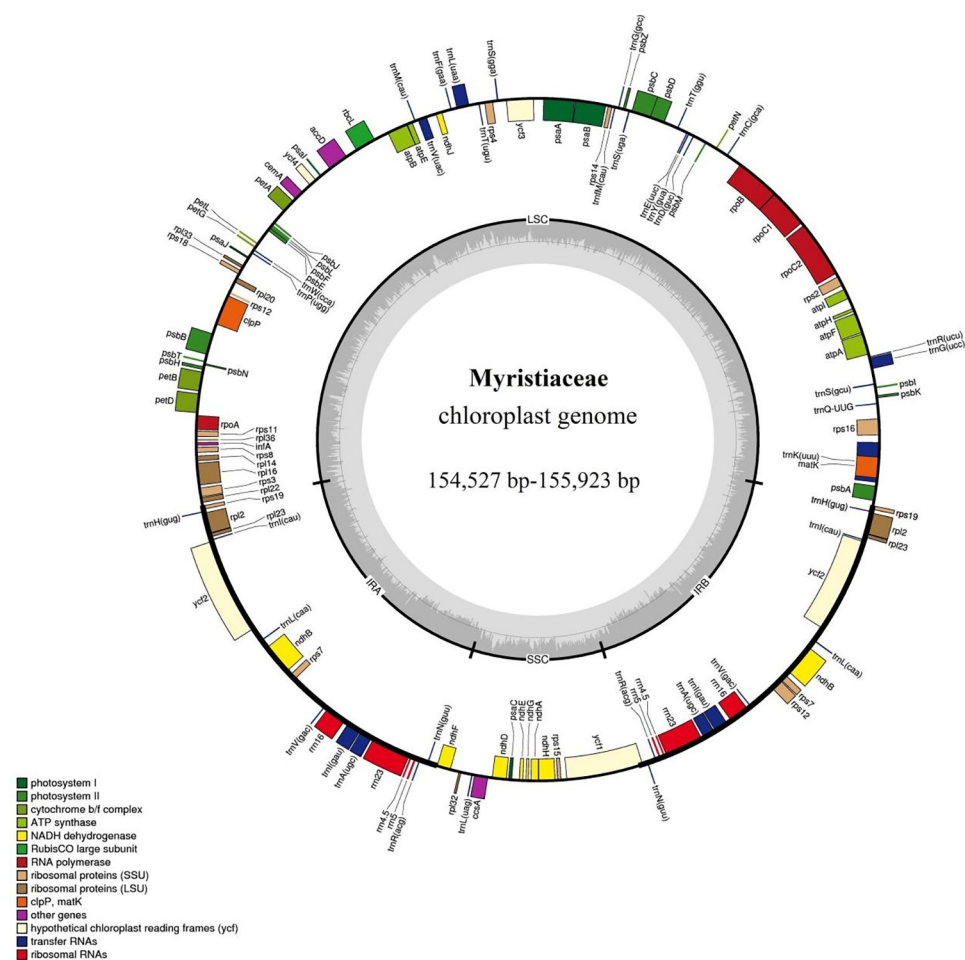


Fig 1. Gene map of the Myristicaceae chloroplast genome. Genes on the inside of the circle are transcribed in the counterclockwise direction, while genes on the outside of circle are transcribed clockwise. The different gene functional groups were labeled with different colors. In inner circle, the gray indicates GC content, while light gray represents AT content. LSC, Large Single Copy; SSC, Small Single Copy; IR, Inverted Repeat.

<https://doi.org/10.1371/journal.pone.0281042.g001>

Table 2. Summary of features of ten species of Myristicaceae.

genome features	<i>H. amygdalina</i>	<i>H. pandurifolia</i>	<i>H. kingii</i>	<i>M. yunnanensis</i>	<i>K. furfuracea</i>	<i>K. elegans</i>	<i>K. linifolia</i>	<i>K. conferta</i>	<i>K. globularia</i>	<i>K. cinerea</i>
cp length	155,683	155,695	155,655	155,923	154,527	155,691	155,754	155,744	155,726	155,690
LSC length	86,931	92,561	86,913	87,088	86,188	86,883	86,991	86,926	86,898	86,881
SSC length	30,998	15,072	20,691	20,731	20,229	20,686	20,683	20,770	20,674	20,761
IR length	37,754	48,062	48,052	48,104	48,110	48,122	48,080	48,052	48,154	48,048
Genes	124	121	123	124	124	123	128	128	131	131
CDS	86	84	85	85	87	85	89	89	92	92
tRNA	27	26	27	27	27	27	31	31	31	31
rRNA	8	8	8	8	8	8	8	8	8	8
GC%	39.24	39.21	39.23	39.19	39.23	39.20	39.19	39.20	39.22	39.21

<https://doi.org/10.1371/journal.pone.0281042.t002>

yunnanensis) to 39.24% (*H. amygdalina*), while the AT content ranged from 60.76% (*H. amygdalina*) to 60.81% (*M. yunnanensis*). The number of genes ranged from 121 to 131, with varied numbers of CDS, rRNA, and tRNA in different species (Table 3); among these genes, 11 genes (*trnQ-UUG*, *rps19*, *psbB*, *trnS-GGA*, *rpoB*, *atpH*, *rps7*, *trnV-GAC*, *ndhH*, *rpl23*, and *trnL-CAA*) had one intron, while *rps7* contained two introns in *H. amygdalina*, *K. elegans*, and *M. yunnanensis*.

Repeat analysis

Tandem repeat sequences (TRSs) have an important influence in terms of gene structure, function and evolution, and so on [33]. SSRs are tandemly repeated motifs with a length of 1–6 bp, which have been widely used as molecular markers in evolutionary biology and population genetics [34, 35]. To explore the genetic changes evident in the Myristicaceae species analyzed, we performed tandem repeat and SSRs analysis. In this study, we identified an average of 62 SSR loci in the complete chloroplast genome of the studied species, including 45 mononucleotide SSR loci, 5 dinucleotide SSR loci, 2 trinucleotide SSR loci, 8 tetranucleotide SSR loci, and 2 pentanucleotide SSR loci. In all species, only one hexanucleotide was found in *H. pandurifolia*, (AATAAA)₃, located in the *matK~rps16* region (Fig 2A). SSR information, from trinucleotides to hexanucleotides, are displayed in Table 4, because of the large number of mononucleotide and dinucleotides SSR sites, they are listed in S1 Table.

At the same time, a total of 38 types of TRSs were detected in all species, with the repeat type ranging from 16 to 24 per species (Fig 2B); these repeats are mainly distributed in the *ycf2*, *trnV-GAC~rps7*, and *trnN-GUU~trnR-ACG* in the IR regions, the *rps11*, *ndhB*, *petN~psbM*, *atpH~atpI*, *rpoB~trnC-GCA*, *atpB~rbcL*, *trnP-UGG~psaJ*, *psbZ~trnG-GCC*, *rpl20~rps12*, *rpl32~trnL-UAG*, *trnC-GCA~petN*, *trnD-GUC~trnY-GUA*, and *trnT-UGU~trnF-GAA* in the LSC, and the *ycf1*, *ycf1~trnN-GUU*, *ndhD~ccsA*, *ccsA~ndhF*, and *rpl32~trnL-UAG* in the SSC. Among them, the shortest TRSs with a base sequence of TTTATATAA were detected in the species *K. cinerea*, *K. elegans*, *K. furfuracea*, and *K. linifolia*, while the longest TRS, with a base sequence of AGAAAAATGGAGACTATTTCTTTTATTTAT was detected only in *K. linifolia* (Fig 3).

Sequence divergence

To detect the selection pressures in protein coding genes (PCGs) in Myristicaceae chloroplast genomes, nonsynonymous (dN) substitutions, synonymous (dS) substitutions, and their ratios (dN/dS) were calculated, using 74 PCGs of 10 Myristicaceae species (Fig 4, S2 Table). Most PCGs had dN/dS values that were less than 1, and only 11 (*accD*, *ccsA*, *matK*, *ndhF*, *ndhG*,

Table 3. Gene present in the chloroplast genome of Myristicaceae.

Category	Group of genes	Name of genes
Protein genes	Photosystem I	<i>psaA, psaB, psaC, psaI, psaJ, ycf3</i>
	Photosystem II	<i>psbA</i> ³⁽²⁾ , <i>psbB</i> ^{1,3(2)} , <i>psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbT, psbZ</i>
	Subunit of cytochrome	<i>petA, petB, petD, petG, petL, petN</i>
	Subunit of synthase	<i>atpA, atpB, atpE, atpF, atpH</i> ¹ , <i>atpI</i>
	Large subunit of rubisco	<i>rbcL</i>
	NADH dehydrogenase	<i>ndhA, ndhB</i> ³⁽²⁾ , <i>ndhC, ndhD, ndhE, ndhF, ndhG, ndhH</i> ¹ , <i>ndhI, ndhJ, ndhK</i>
	ATP dependent protease subunit P	<i>clpP</i>
	Chloroplast envelope membrane protein	<i>cemA</i>
RNA genes	Ribosomal RNA genes (rRNA)	<i>rrn5</i> ³⁽²⁾ , <i>rrn4.5</i> ³⁽²⁾ , <i>rrn16</i> ³⁽²⁾ , <i>rrn23</i> ³⁽²⁾
	Transfer RNA genes (tRNA)	<i>trnH-GUG, trnQ-UUG</i> ¹ , <i>trnS-GCU, trnR-UCU, trnC-GCA, trnD-GUC, trnY-GUA, trnE-UUC, trnT-GGU, trnS-UGA, trnG-GCC, trnS-GGA</i> ¹ , <i>trnT-UGU, trnF-GAA, trnM-CAU</i> ³⁽³⁾ , <i>trnW-CCA, trnP-UGG, trnI-GAU</i> ^{3,4} , <i>trnL-CAA</i> ^{1,3} , <i>trnV-GAC</i> ^{1,3(2)} , <i>trnA-UGC</i> ^{3(2),4} , <i>trnR-ACG, trnN-GUU, trnL-UAG</i>
Ribosomal proteins	Small subunit of ribosome	<i>rps2, rps3, rps4, rps7</i> ^{1,2,3(2)} , <i>rps8, rps11, rps12</i> ³⁽²⁾ , <i>rps14, rps15, rps16, rps18, rps19</i>
Transcription	Large subunit of ribosome	<i>rpl2</i> ³⁽²⁾ , <i>rpl14, rpl16, rpl20, rpl22, rpl23</i> ^{1,3(2)} , <i>rpl32, rpl33, rpl36</i>
	DNA dependent RNA polymerase	<i>rpoA, rpoB</i> ¹ , <i>rpoC1, rpoC2</i>
Other genes	Maturase	<i>matK</i>
	Subunit acetyl-coA carboxylase	<i>accD</i>
	C-type cytochrome synthesis	<i>ccsA</i>
	Hypothetical proteins	<i>ycf2</i> ³⁽²⁾ , <i>ycf4, ycf1</i>
	Translation initiation factor IF-1	<i>infA</i>

¹Gene with one intron in all species except for *H. pandurifolia*, *H. kingii*, *K. furfuracea*, *K. linifolia*, *K. conferta*, *K. globularia* and *K. cinerea*

²Gene with two intron in *H. amygdalina*, *K. elegans* and *M. yunnanensis*

³⁰Gene with copies, the number in the brackets was the number of copies and

⁴Gene only distributed in *K. conferta*, *K. globularia*, *K. cinerea* and *K. linifolia*

<https://doi.org/10.1371/journal.pone.0281042.t003>

psaA, *rbcL*, *rpoA*, *rpoC2*, *ycf1*, and *ycf2*) had dN/dS values greater than 1; among these 11 PCGs, the dN/dS ratio of the *psaA* gene in most species comparisons was greater than 1, while for the *rpoA* gene, the ratio was greater than 1 only in the comparison group of *Myristica* vs. *Knema* (Fig 4, S2 Table). Most of the PCGs had negative selection, and only a few PCGs had positive selection. The value of synonymous (dS) ranged from 0.00 to 0.68 (*psaA* gene) in all of the genes. Most PCGs, including *atpB*, *atpH*, *cemA*, *clpP*, *ndhK*, *petG*, *psaC*, *psbD*, *psbE*, *psbF*, *psbJ*, *psbK*, *psbL*, *psbM*, *psbT*, *psbZ*, *rpl2*, *rpl14*, *rpl23*, *rpl33*, *rpl36*, *rps11*, *rps15*, *rps16*, *rps19*, and *ycf3*, showed no nonsynonymous (dN) changes.

To clearly understand the nucleotide changes in the Myristicaceae chloroplast genome, we calculated the nucleotide diversity (Pi) values of the PCGs and intergenic spacers. In the coding region, the mean Pi in the PCGs was 0.00279 (ranging from 0 to 0.00886, Fig 5A). In total, 11 mutation sites (Pi > 0.005) were identified, including *matK*, *ndhD*, *ndhF*, *ndhG*, *psbL*, *rpl16*,

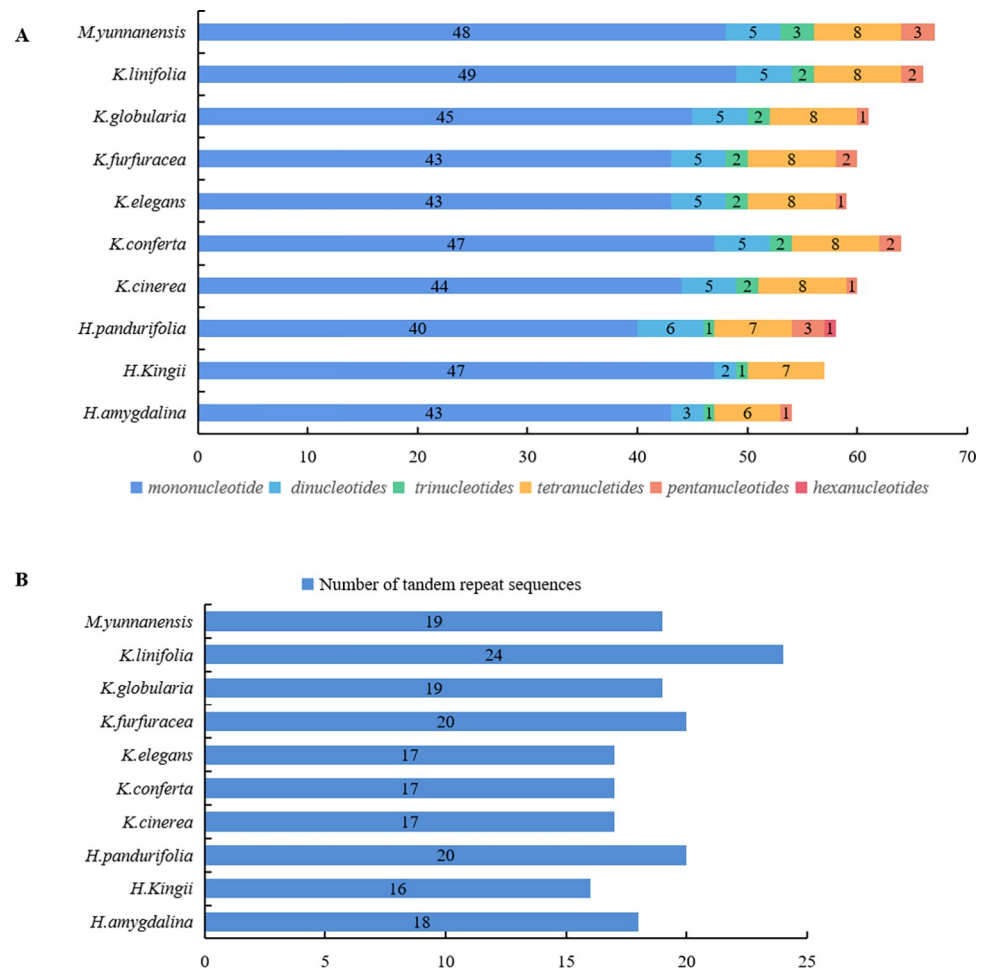


Fig 2. Types and numbers of repeat in the chloroplast genomes of ten cp genome of Myristicaceae. A, Numbers and types of SSRs; B, Numbers of Tandem repeat sequences.

<https://doi.org/10.1371/journal.pone.0281042.g002>

rpl32, *rpoA*, *rps3*, *rps19*, and *ycf1*. In intergenic spacers, the mean Pi value was 0.006933 (ranging from 0.00054 to 0.04744, Fig 5B), and there were 18 mutation sites ($Pi > 0.005$) in these spacers, including *accD~psaI*, *atpF~atpH*, *matK~rps16*, *ndhC~trnM-CAU*, *ndhF~rpl32*, *petA~psbJ*, *petN~psbM*, *psbE~petL*, *psbM~trnD-GUC*, *rpl20~rps12*, *rpl32~trnL-UAG*, *rpoB~trnC-GCA*, *rps16~trnQ-UUG*, *trnC-GCA~petN*, *trnE-UUC~trnT-GGU*, *trnN-GUU~trnR-ACG*, *trnT-UGU~trnF-GAA*, and *ycf3~trnS-GGA*. We also assessed the statistics of initiation codon and termination codon of 74 PCGs, the results showed that the initiation codons were mainly ATG, GTG, CAA, ACG, CCA, AAT, GGG, and AAC, and the termination codons included TGA, TAG, and TAA; among these, ATG was used as the initiation codon in 63 genes, and these termination codons were used at a similar frequency.

Genome comparison

To investigate the variation in chloroplast genome sequences, the 10 Myristicaceae species were compared with the mVISTA using the annotation sequence of *L. tulipifera*. The results indicated that the chloroplast genome sequences of Myristicaceae were relatively conserved, although a certain level of variation was detected. The pair of reverted repeat region was highly conserved than the LSC region and SSC region, and the PCGs were highly conserved than

Table 4. SSR site information of ten species of Myristicaceae.

Species	Type	Repeat unit	Start	End	Location	Type	Repeat unit	Start	End	Location
<i>H. amygdalina</i>	tetra	(TTTG)3	33689	33700	<i>atpH~atpI</i>	tetra	(ACAT)3	105492	105503	<i>rpl22</i>
	tetra	(TCTA)3	55980	55991	<i>psbC~trnS-UGA</i>	tetra	(CATT)3	145487	145498	<i>ycf1</i>
	tetra	(TTTC)3	69141	69152	<i>trnF-GAA~ndhJ</i>	penta	(AAATA)3	97116	97130	<i>psbH~petB</i>
	tetra	(AATG)3	83317	83328	<i>cemA</i>					
<i>H. Kingii</i>	tetra	(TTTG)3	59557	59568	<i>atpH~atpI</i>	tetra	(AATG)3	109157	109168	<i>cemA</i>
	tetra	(TCTA)3	81828	81839	<i>psbC~trnS-UGA</i>	tetra	(CTTT)3	129381	129392	<i>rpl16~rps3</i>
	tetra	(TAAC)3	93138	93149	<i>trnT-UGU~trnF-GAA</i>	tetra	(ACAT)3	131314	131325	<i>rpl22</i>
	tetra	(TTTC)3	95007	95018	<i>trnF-GAA~ndhJ</i>					
<i>H. pandurifolia</i>	hexa	(AATAAA)3	28500	28517	<i>matK~rps16</i>	tetra	(AATG)3	88391	88402	<i>cemA</i>
	tetra	(TTTG)3	38712	38723	<i>atpH~atpI</i>	tetra	(CTTT)3	108650	108661	<i>rpl16~rps3</i>
	penta	(TTGAT)3	52817	52831	<i>trnC-GCA~petN</i>	tetra	(ACAT)3	110583	110594	<i>rpl22</i>
	penta	(AATCA)3	52842	52856	<i>trnC-GCA~petN</i>	penta	(TATTT)3	142128	142142	<i>ndhD~psaC</i>
	tetra	(TCTA)3	61093	61104	<i>psbC~trnS-UGA</i>	tri	(AAT)4	148613	148624	<i>rps15~ycf1</i>
	tetra	(TAAC)3	72349	72360	<i>trnT-UGU~trnF-GAA</i>	tetra	(CATT)3	150631	150642	<i>ycf1</i>
<i>K. cinerea</i>	tetra	(TTTG)3	38736	38747	<i>atpH~atpI</i>	tetra	(AATG)3	88416	88427	<i>cemA</i>
	tetra	(TCTA)3	61082	61093	<i>psbC~trnS-UGA</i>	tetra	(CTTT)3	108655	108666	<i>rpl16~rps3</i>
	penta	(ATTCT)3	72088	72102	<i>trnT-UGU~trnF-GAA</i>	tetra	(ACAT)3	110591	110602	<i>rpl22</i>
	tetra	(TAAC)3	72424	72435	<i>trnT-UGU~trnF-GAA</i>	tri	(AAT)4	148593	148604	<i>rps15~ycf1</i>
	tetra	(TTTC)3	74263	74274	<i>trnF-GAA~ndhJ</i>	tetra	(CATT)3	150611	150622	<i>ycf1</i>
	tri	(ATA)4	81105	81116	<i>atpB~rbcl</i>					
<i>K. conferta</i>	penta	(ATTAA)3	37586	37600	<i>atpF~atpH</i>	tetra	(AATG)3	88455	88466	<i>cemA</i>
	tetra	(TTTG)3	38749	38760	<i>atpH~atpI</i>	tetra	(CTTT)3	108689	108700	<i>rpl16~rps3</i>
	tetra	(TCTA)3	61095	61106	<i>psbC~trnS-UGA</i>	tetra	(CATT)3	150670	150681	<i>ycf1</i>
	penta	(ATTCT)3	72091	72105	<i>trnT-UGU~trnF-GAA</i>	tetra	(ACAT)3	110624	110635	<i>rpl22</i>
	tetra	(TTTC)3	74287	74298	<i>trnF-GAA~ndhJ</i>	tetra	(TAAC)3	72449	72460	<i>trnT-UGU~trnF-GAA</i>
	tri	(ATA)4	81144	81155	<i>atpB~rbcl</i>	tri	(AAT)4	148652	148663	<i>rps15~ycf1</i>
<i>K. elegans</i>	tri	(AAT)4	13626	13637	<i>rps15~ycf1</i>	tetra	(TAAC)3	93147	93158	<i>trnT-UGU~trnF-GAA</i>
	tetra	(CATT)3	15644	15655	<i>ycf1</i>	tetra	(TTTC)3	94986	94997	<i>trnF-GAA~ndhJ</i>
	tetra	(TTTG)3	59459	59470	<i>atpH~atpI</i>	tri	(ATA)4	101828	101839	<i>atpB~rbcl</i>
	tetra	(TCTA)3	81805	81816	<i>psbC~trnS-UGA</i>	tetra	(AATG)3	109139	109150	<i>cemA</i>
	penta	(ATTCT)3	92811	92825	<i>trnT-UGU~trnF-GAA</i>	tetra	(CTTT)3	129379	129390	<i>rpl16~rps3</i>
<i>K. furfuracea</i>	tetra	(ACAT)3	131315	131326	<i>rpl22</i>	tetra	(ACAT)3	37646	37657	<i>rpl22</i>
	tetra	(TAAC)3	205	216	<i>trnT-UGU~trnF-GAA</i>	tri	(AAT)4	75258	75269	<i>rps15~ycf1</i>
	tetra	(TTTC)3	2044	2055	<i>trnF-GAA~ndhJ</i>	tetra	(CATT)3	77276	77287	<i>ycf1</i>
	tri	(ATA)4	8160	8171	<i>atpB~rbcl</i>	penta	(ATTAA)3	119856	119870	<i>atpF~atpH</i>
	tetra	(AATG)3	15471	15482	<i>cemA</i>	tetra	(TTTG)3	121023	121034	<i>atpH~atpI</i>
	tetra	(CTTT)3	35710	35721	<i>rpl16~rps3</i>	tetra	(TCTA)3	143371	143382	<i>psbC~trnS-UGA</i>
<i>K. globularia</i>	penta	(ATTCT)3	154374	154388	<i>trnT-UGU~trnF-GAA</i>	tetra	(TAAC)3	93129	93140	<i>trnT-UGU~trnF-GAA</i>
	tri	(AAT)4	13592	13603	<i>rps15~ycf1</i>	tetra	(TTTC)3	94968	94979	<i>trnF-GAA~ndhJ</i>
	tetra	(CATT)3	15610	15621	<i>ycf1</i>	tri	(ATA)4	101817	101828	<i>atpB~rbcl</i>
	penta	(ATTAA)3	58280	58294	<i>atpF~atpH</i>	tetra	(AATG)3	109128	109139	<i>cemA</i>
	tetra	(TTTG)3	59443	59454	<i>atpH~atpI</i>	tetra	(CTTT)3	129366	129377	<i>rpl16~rps3</i>
	tetra	(TCTA)3	81788	81799	<i>psbC~trnS-UGA</i>					
<i>K. linifolia</i>	tetra	(ACAT)3	131301	131312	<i>rpl22</i>	tetra	(TTTC)3	74378	74389	<i>trnF-GAA~ndhJ</i>
	penta	(ATTAA)3	37598	37612	<i>atpF~atpH</i>	tri	(ATA)4	81227	81238	<i>atpB~rbcl</i>
	tetra	(TTTG)3	38761	38772	<i>atpH~atpI</i>	tetra	(AATG)3	88538	88549	<i>cemA</i>
	tetra	(TCTA)3	61144	61155	<i>psbC~trnS-UGA</i>	tetra	(CTTT)3	108778	108789	<i>rpl16~rps3</i>
	penta	(ATTCT)3	72151	72165	<i>trnT-UGU~trnF-GAA</i>	tetra	(ACAT)3	110715	110726	<i>rpl22</i>
	tetra	(TAAC)3	72533	72544	<i>trnT-UGU~trnF-GAA</i>	tri	(AAT)4	148757	148768	<i>rps15~ycf1</i>

(Continued)

Table 4. (Continued)

Species	Type	Repeat unit	Start	End	Location	Type	Repeat unit	Start	End	Location
<i>M. yunnanensis</i>	tetra	(CATT)3	150775	150786	<i>ycf1</i>	penta	(ATTCT)3	93002	93016	<i>trnT-UGU~trnF-GAA</i>
	tri	(AAT)4	13667	13678	<i>rps15~ycf1</i>	tetra	(TAAC)3	93334	93345	<i>trnT-UGU~trnF-GAA</i>
	tetra	(CATT)3	15685	15696	<i>ycf1</i>	tetra	(TTTC)3	95203	95214	<i>trnF-GAA~ndhJ</i>
	penta	(ATCCA)3	48367	48381	<i>matK~rps16</i>	tri	(ATA)4	102041	102052	<i>atpB~rbcL</i>
	penta	(GAATA)3	52032	52046	<i>trnS-GCU~trnR-UCU</i>	tetra	(AATG)3	109358	109369	<i>cemA</i>
	tetra	(TTTG)3	59658	59669	<i>atpH~atpI</i>	tetra	(CTTT)3	129630	129641	<i>rpl16~rps3</i>
	tetra	(TCTA)3	82020	82031	<i>psbC~trnS-UGA</i>	tetra	(ACAT)3	131556	131567	<i>rpl22</i>
	tri	(TAA)4	91499	91510	<i>ycf3~trnS-GGA</i>					

<https://doi.org/10.1371/journal.pone.0281042.t004>

non-coding sequences, particularly in intergenic regions. The intergenic spacer region featured a high level of variation within the genome, including *atpH~atpI*, *trnS-GGA~rps4*, *ndhC~trnM-CAU*, *petA~psbJ*, *trnP-UCG~psaJ*, *psbE~petL*, and *psbH~petB*, and the protein coding genes were *ycf2*, *rpoC1*, *clpP*, and *rps12* (Fig 6).

Previous studies have found that the plastome sequence was conserved in flowering plants [36], although, due to evolutionary events, changes occurred in the size and boundaries of individual replicates and reverse repeats [37, 38]. Comparison among the reverse repeat region, the LSC region, and the SSC region boundaries in the 10 species of Myristicaceae are presented in Fig 7. Most species exhibit some variation in the number of nucleotides in the boundaries of the LSC, IR, and SSC regions. Except *H. pandurifolia*, the studied Myristicaceae species had the same set of genes at the border: the *rpl22* and *trnH* genes were located in the LSC region, the *ndhF* and *trnR* genes were located in the SSC region, and the *rps19* and *rrn5* genes were found in the IR regions. However, for *H. pandurifolia*, the *trnH* gene was located in the IR regions, the *rps19* gene was located on the LSC/IRs border, and the positioning of the *ycf1* gene in the IRb/SSC border was observed only in the genome of *H. pandurifolia*.

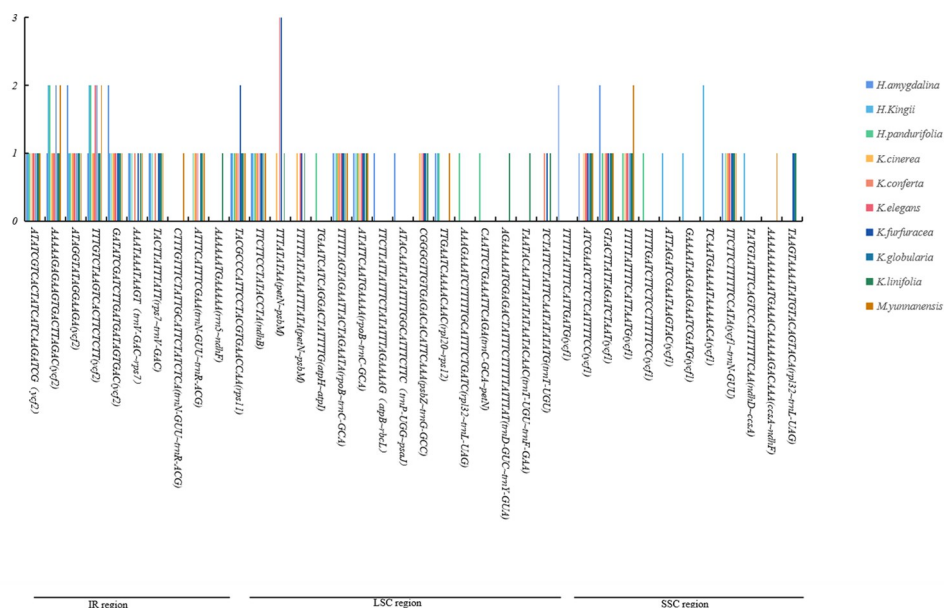


Fig 3. Details of tandem repeats sequence in 10 species and number of repeats in each species. The gene or intergenic regions in parentheses indicates the location of the tandem repeat. The ordinate represents the number of repetitions.

<https://doi.org/10.1371/journal.pone.0281042.g003>

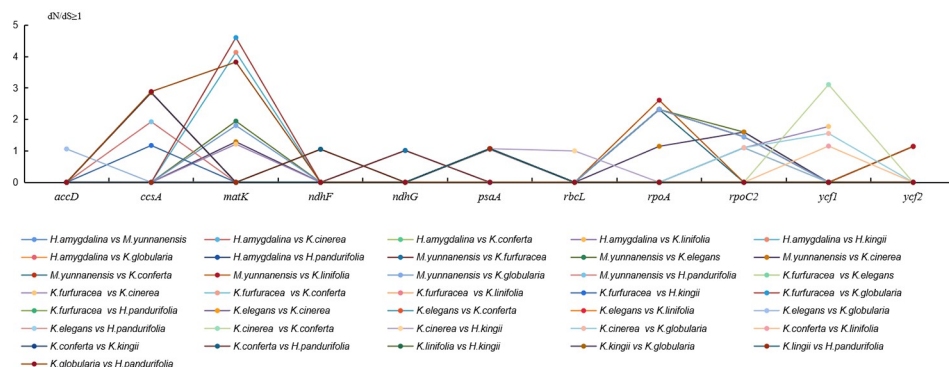


Fig 4. The genes of dN/dS > 1 from 10 cp genomes of Myristicaceae.

<https://doi.org/10.1371/journal.pone.0281042.g004>

Phylogenetic analysis

To determine the phylogenetic relationship of the 26 samples of Myristicaceae, phylogenetic trees were reconstructed using the chloroplast genome of those species using BI and ML, taking *L. tulipifera* as the outgroup. The results showed that BI and ML analyses were congruent with high-support PP, 1.0, and MP, 100, in most relationships (Fig 8). All species of Myristicaceae were clustered in one clade with high support. Both analysis methods showed that *Horsfieldia* genus formed a single clade except for *H. pandurifolia* with high bootstrap values of 100%/1.0. In contrast, all *H. pandurifolia* samples were separated from *Horsfieldia* genus and formed a clade with bootstrap values of 100%/1.0 and formed a sister group with genus *Knema* and *Myristica*.

Discussion

Myristicaceae plants are recognized as source plants of medical, wood, and oil products. *M. yunnanensis* is listed as a protected plant in China and is now considered endangered [39].

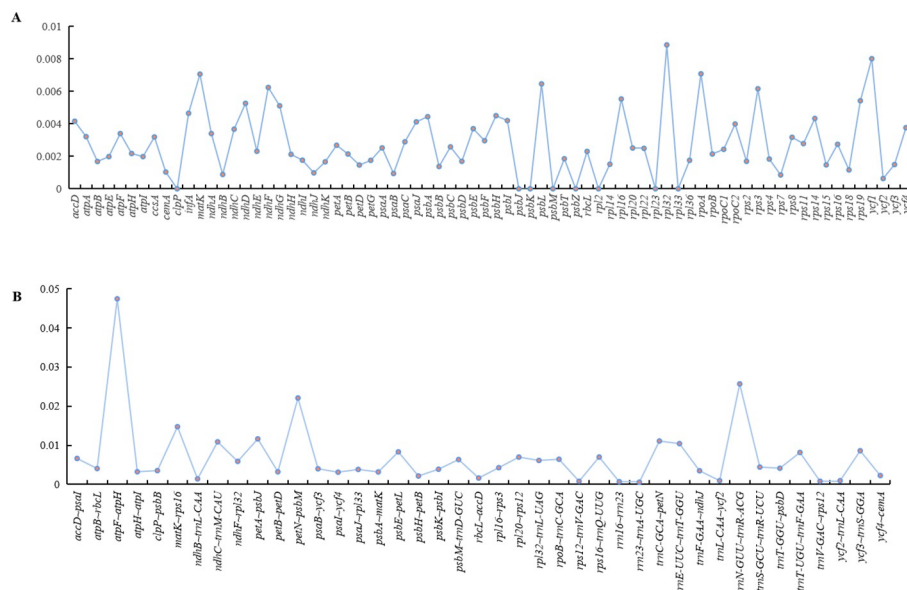


Fig 5. Nucleotide diversity (Pi) in the chloroplast genome of ten Myristicaceae species. (A:Pi value (range from 0–0.00886) of protein coding genes; B:Pi value (0.00054–0.04744) of intergenic regions genes).

<https://doi.org/10.1371/journal.pone.0281042.g005>

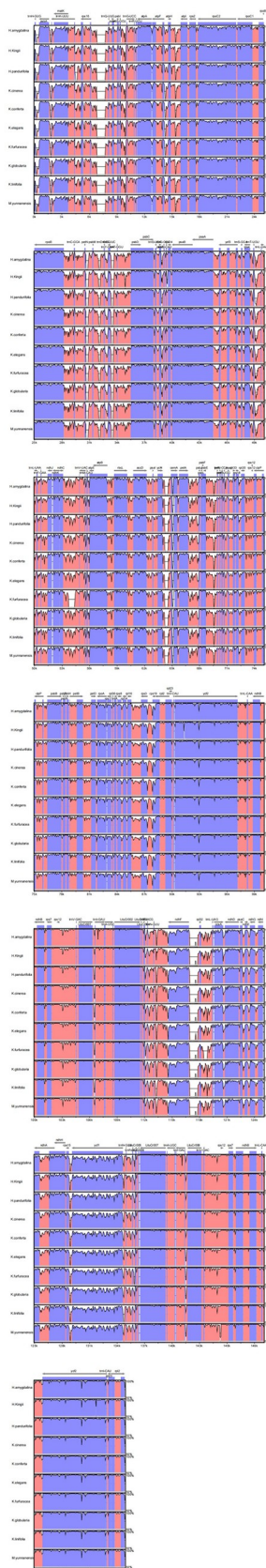


Fig 6. Sequences alignment of ten chloroplast genome in the Myristicaceae family performed with mVISTA using annotation of *L. tulipifera* as reference.

<https://doi.org/10.1371/journal.pone.0281042.g006>

Until recently, the reports of genome sequences in this family had only been published by the authors [40–47] and Cai [16]. In this study, we comprehensively analyzed the chloroplast genome of Myristicaceae species and reconstructed phylogenetic tree using 26 genome sequences. The results showed that all genomes exhibited a quadripartite structure, including LSC, small SSC, and a pair of IR regions (IRa and IRb), which consisted of 84–92 PCGs, 26–31 tRNAs, and 8 rRNAs, with genome sizes that ranged from 154,527 to 155,923 bp and with GC % contents that ranged from 39.18% to 39.24%. Compared to species from the same order Magnoliales, the organization and structure of these 10 chloroplast genomes were similar to that of the *L. tulipifera* [48].

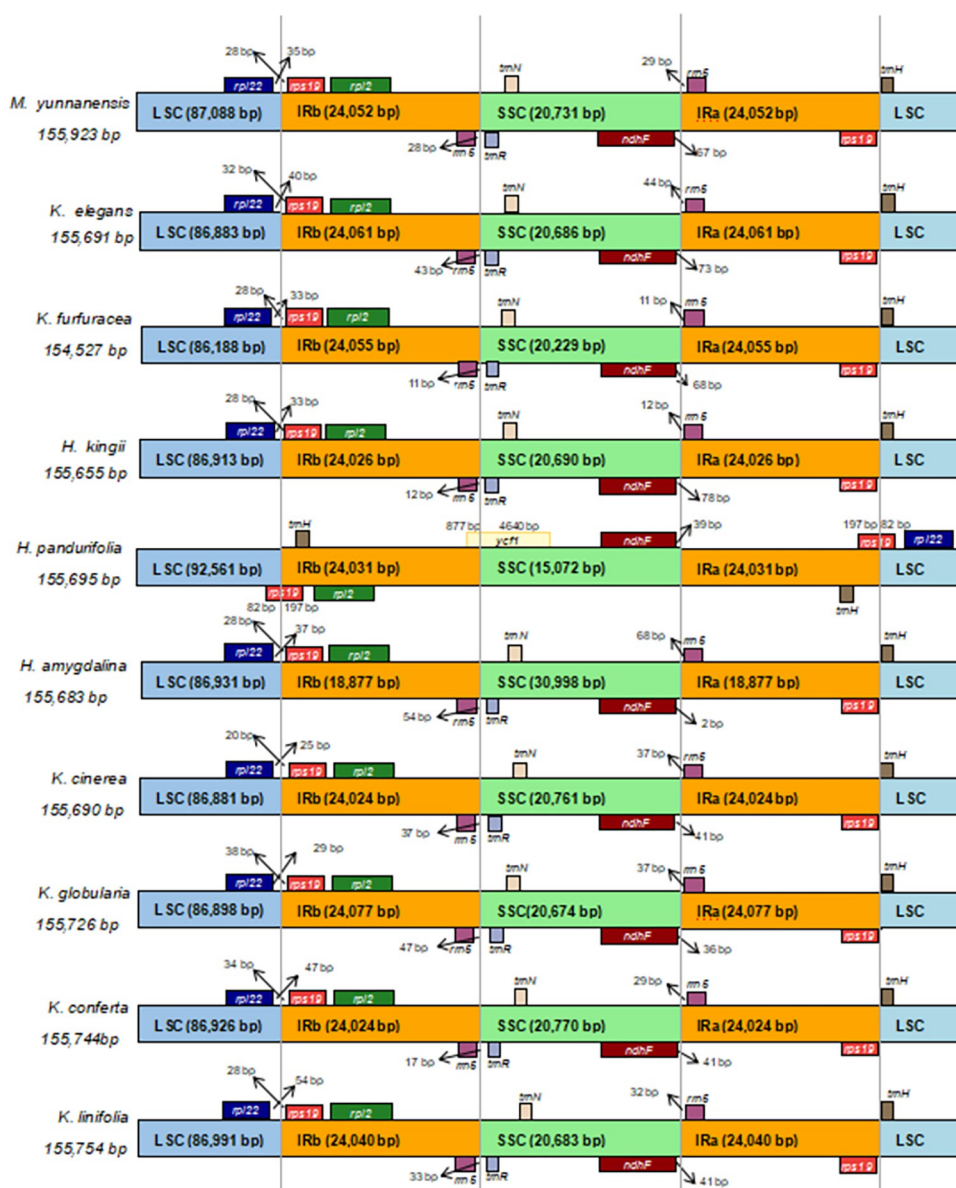


Fig 7. Comparison of the borders of the IR, SSC and LSC regions among ten chloroplast genome of Myristicaceae.

<https://doi.org/10.1371/journal.pone.0281042.g007>

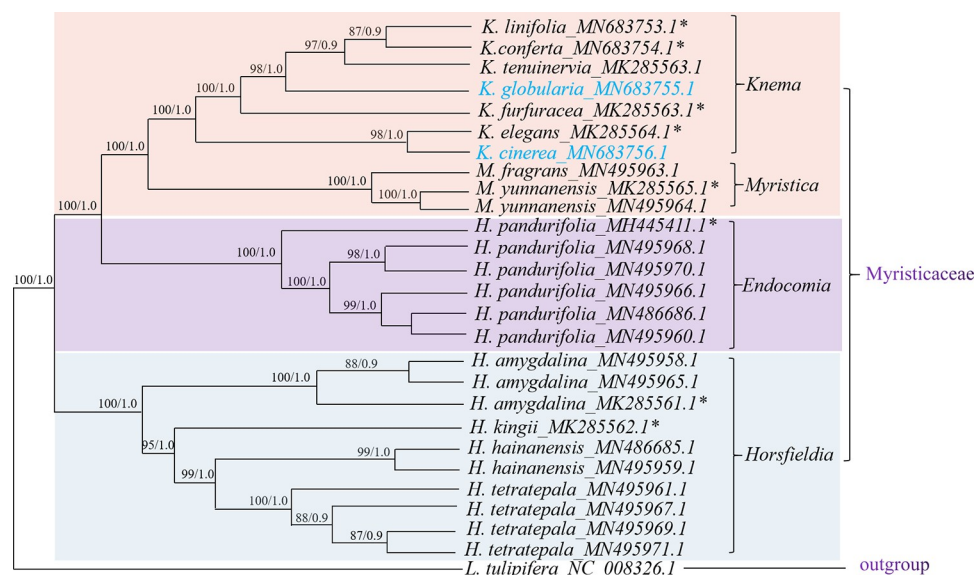


Fig 8. The phylogenetic tree based on complete chloroplast genome sequences of Myristicaceae with ML and BI method. Support values are bootstrap values (>50%, before slash) and posterior probability (>0.5, after slash), respectively. The species with blue font indicates the two newly sequenced in this study; "*" indicates that the data were derived from the authors.

<https://doi.org/10.1371/journal.pone.0281042.g008>

Genome repeats play an indispensable role in gene expression, transcriptional regulation, chromosome construction, genomic structural variation, expansion, and rearrangement [49–51]. Similarly, we analyzed the repeat sequences of Myristicaceae, and the results are consistent with previous studies that most SSRs were located in intergenic spacers regions, followed by coding regions, and most TRSs were located in *ycf* genes and non-coding regions [19, 52–54]. The cpSSR has been widely used in phylogenetic evaluation and population genetics [55], and it is a very effective marker as well [56, 57]. In the chloroplast genome of Myristicaceae, mononucleotide repeats made up over 79%, and over 95% of the mononucleotide consisted of A or T bases, and the majority of SSRs were found in the SSC and LSC regions, the proportion of mononucleotide repeats and the base composition were similar to previous research [53, 55].

The variation in the chloroplast genome size was a result of the contraction and expansion of the reverse repeats (IRs) [58]. This contraction and expansion were observed in the chloroplast genome sequences of Myristicaceae. The size of the IRs ranged from 18,877 bp (*H. amygdalina*) to 24,077 bp (*K. globularia*). Despite the similar lengths of the IR regions of *H. pandurifolia* in relation to other Myristicaceae species, with the exception of *H. amygdalina*, some level of expansion and contraction were observed. Due to the different positions of the genes *rps19*, *ycf1*, and *trnH*, three types of variation in the border of IR-SC region appeared among these species, as a result of the contraction and expansion of reverse repeats. Type I occurred in *H. pandurifolia*, where parts of the genes *rps19* and *ycf1* were located in the IR region, and other parts were located in the LSC and SSC regions, respectively. Type II occurred in all other species, except *H. pandurifolia*, that the two *rps19* were located in the IR regions. Type III was found in all the studied species, except *H. pandurifolia*, that *trnH* was located in the LSC region whereas *trnH* was located in IR regions of *H. pandurifolia*. All genomes had *ndhF* in the SSC region, *rpl2* in the IR region, and *rpl22* in the LSC region.

dN, dS, and dN/dS were calculated to evaluate sequence diversity and purifying selection in species evolution. The results indicated low sequence diversity in most genes. dN/dS analysis showed that most protein coding genes faced negative selection (dN/dS < 1), only 11 genes

(*accD*, *ccsA*, *matK*, *ndhF*, *ndhG*, *psaA*, *rbcL*, *rpoA*, *rpoC2*, *ycf1*, and *ycf2*) were positively selected, and the genes with positively selected in this study have been reported in *Barleria prionitis* [18] and *Rheum* species [59]. Nucleotide diversity can be used to measure mutations in the population and can also be used to estimate evolutionary relationships [60]. In recent years, DNA barcoding has been considered as a reliable tool for resolving phylogenetic relationships and species authentication [53, 61]. In this study, we aligned the chloroplast genomes and found 11 and 18 mutation sites in the protein coding region and the intergenic spacer, respectively. These mutation hotspots provided valuable markers for identification of species as well as resolving phylogenetic relationships in the family.

Comparative genome analysis using mVISTA showed that the genomes are relatively conserved with minor variation, which mainly occurred in non-coding regions. Through the results of alignment, there were no considerable rearrangement being detected in the chloroplast genome.

The genome sequences were the effective resource for inferring phylogenetic relationships among species [62]. The phylogenetic relationship among the species of Myristicaceae is controversial, particularly *H. pandurifolia* [8, 12, 16]. In this study, the phylogenetic relationships of 26 Myristicaceae samples were inferred using ML and Bayesian methods. The genus *Knema* was clustered with the genus *Myristica*, and same-genus species were clustered in the same clade with a high support value, this was consistent with previously reported results which *Knema* species was found to share sister relationship with *Myristica* species [63]. In *Horsfieldia*, *H. kingii*, *H. hainanensis*, *H. tetratelpala* and *H. amygdalina* were clustered together, however, *H. pandurifolia* was separated from the four *Horsfieldia* species forming a single clade, same as the result of Amplified Fragment Length Polymorphism (AFLP) [8]. For the morphological observation, the aril's color of *H. pandurifolia* was recorded as orange in Flora of China, whereas bright red as observed by Wu [8]; the seed apex of other *Horsfieldia* species were round, whereas *H. pandurifolia* was pointed, same as *E. macrocoma* spp. *prainii*; the testa color of *H. pandurifolia*, same as *E. macrocoma*, was variegated, whereas other *Horsfieldia* species' testa were brown [8, 13]. Based on the results of fatty acid research, Wu believed that *H. pandurifolia* should be separated from genus *Horsfieldia* [8]. In summary, molecular, morphological, and fatty acid data show that *H. pandurifolia* should be treated as a genus [8, 9].

In current study, the chloroplast genome sequences that used to analysis the phylogenetic relationship included all species except for *Myristica cagayanensis* Merr. And *Myristica simiarum* A.DC. distributed in China, which can better explain the phylogenetic relationships of the Chinese species of this family. This study suggests taxonomical revision that *H. pandurifolia* should be separated from the genus *Horsfieldia* and placed in the genus *Endocomia* (1984) proposed by W.J.de Wilde [9]. In addition, morphological observations of genus *Knema* by the authors showed that the flowers and leaves of *K. cinerea* and *K. elegans* are similar. Based on the phylogenetic tree, *K. furfuracea* and *K. linifolia* had a relatively close phylogenetic relationship, as the same result occurred in both *K. cinerea* and *K. elegans*. We will do further research in the future on the phylogenetic relationship of genus *Knema*.

Conclusion

In this study, we sequenced two chloroplast genomes of *Knema* species, *K. globularia* and *K. cinerea*, and compared them with those of eight reported species. The size, gene content and the structure of Myristicaceae chloroplast genomes were similar, and comparative analysis revealed no gene inversion or relocation among chloroplast genome. We identified eleven highly variable sites, with seven intergenic spacer region and four PCGs, that could be explored to create valuable genetic markers for species authentication and phylogeny in Myristicaceae.

Through the comparison of the genomes, we identified 11 genes and 18 intergenic spacer regions were positively selected which can be used to analyze the population genetic structure of Myristicaceae. We comprehensively analyzed the phylogenetic relationship of Myristicaceae using 26 samples from 3 genera distributed in China. We suggesting taxonomical revision that *H. pandurifolia* should be separated from *Horsfieldia* and placed in the genus *Endocomia*.

Supporting information

S1 Table. Mononucleotide and dinucleotides SSR site information of ten species of Myristicaceae.

(XLSX)

S2 Table. The synonymous (dS) and dN/dS ratio values of 74 protein coding genes from ten cp genomes of Myristicaceae.

(XLSX)

Acknowledgments

We are grateful to thank Nangunhe National Nature Reserve and Xishuangbanna National Nature Reserve for the help of collection material.

Author Contributions

Data curation: Changli Mao.

Formal analysis: Yu Wu.

Funding acquisition: Yu Wu.

Investigation: Changli Mao, Fengliang Zhang, Tian Yang.

Methodology: Changli Mao, Xiaoqin Li.

Project administration: Yu Wu.

Resources: Fengliang Zhang, Xiaoqin Li, Tian Yang, Qi Zhao, Yu Wu.

Writing – original draft: Changli Mao.

Writing – review & editing: Changli Mao, Yu Wu.

References

1. Wu ZY, Raven PH, Hong DY. Flora of China (Vol. 7). Beijing: Science Press; 2008.
2. Xu YL, Wu Y, Yi XQ, Yang XL, Cai NH, Zhang KY, et al. Variation analysis on the seed traits and seed oil content of tree *Horsfieldia pandurifolia* in Yunnan. *Journal of Anhui Agriculture Science*. 2011; 39(6): 3426–3428.
3. Xu YL, Cai NH, Wu Y, Duan AA. Fatty acid composition of several plants of *Horsfieldia*. *China Oil and Fats*. 2012; 37(5): 80–82.
4. Mao CL, Zhang FL, Yang XL, Xu YL, Duan AA, Wu Y, et al. Variation tendency for percentages of main fatty acids from *Horsfieldia pandurifolia* seed. *Journal Southwest University (Natural Science Edition)*. 2017; 39(1): 76–82.
5. Lu LZ. Production and application of fatty acid in China. *Fine and Specialty Chemicals*. 2007; 15(1): 24–28.
6. Wu Y, Mao CL, Zhang FL, Zeng JS, He MY. Fatty acid composition of three species of *Knema*. *Tropical Agriculture Science and Technology*. 2015; 38(3): 28–29,41.
7. Mei WL, Ni W, Hua Y, Chen CX. Flavonoids from *Knema globularia*. *Natural Product Research and Development*. 2002; 14(5): 26–28.

8. Wu Y, Mao CL, Zhang FL, Yang XL, Zeng JS, Duan AA. Taxonomic Position of *Horsfieldia pandurifolia* Hu (Myristicaceae). *Bulletin of Botanical Research*. 2015; 35(5): 652–659.
9. Wilde De. *Endocomia*, a new genus of Myristicaceae. *Blumea*. 1984; 30(1): 173–196.
10. Sauquet H. Androecium diversity and evolution in Myristicaceae (Magnoliales), with a description of a new Malagasy genus, *Doyleanthus* gen.nov. *American Journal of Botany*. 2003; 90(9): 1293–1305.
11. Sauquet H, Doyle JA, Scharaschkin T, Borsch T, Hilu KW, Chatrou LW, et al. Phylogenetic analysis of Magnoliales and Myristicaceae based on multiple data sets: implications for character evolution. *Botanical Journal of the Linnean Society*. 2003; 142: 125–186.
12. Ye M. Taxonomy of Myristicaceae in China. Thesis, *South China Agricultural University*. 2004. Available from: <https://d.wanfangdata.com.cn/thesis/Y665357>
13. Wu Y, Duan AA, Mao CL, Zhang FL, Li XQ, Xu YL, et al. Taxonomical position and population genetic diversity of peculiar oil tree *Horsfieldia pandurifolia* Hu. China Agricultural Science and Technology Press; 2019.
14. Hu HH. Notulae ad floram silvaticam sinensem (I). *Acta phytotaxonomica sinica*. 1963; 8(3): 197–201.
15. Yunnan Institute of Botany. Flora Yunnanica: Volume 1. Beijing: Science Press; 1977.
16. Cai CN, Ma H, Ci XQ, Conran JG, Li J. Comparative phylogenetic analyses of Chinese *Horsfieldia* (Myristicaceae) using complete chloroplast genome sequences. *Journal of Systematics and Evolution*. 2021; 59(3): 504–514.
17. Wu W, Zheng YL, Chen L, Wei YM, Yan ZH, Yang RW. PCR-RFLP analysis of cpDNA and mtDNA in the genus *Houttuynia* in some areas of China. *Hereditas*, 2005; 142: 24–32.
18. Alzahrani DA, Yaradua SS, Albokhari EJ, Abba A. Complete chloroplast genome sequence of *Barleria prionitis*, comparative chloroplast genomics and phylogenetic relationships among Acanthoideae. *BMC Genomics*. 2020; 21(1): 393.
19. Deng Y, Luo Y, He Y, Qin X, Li C, Deng X. Complete chloroplast genome of *Michelia shiluensis* and a comparative analysis with four Magnoliaceae species. *Forests*. 2020; 11(3): 267.
20. Liang H, Zhang Y, Deng JB, Gao G, Ding CB, Zhang L, et al. The complete chloroplast genome sequences of 14 *Curcuma* species insights into genome evolution and phylogenetic relationships within zingiberales. *Frontiers in Genetics*. 2020; 11: 802.
21. Wyman SK, Jansen RK, Boore JL. Organellar genomes with DOGMA. *Bioinformatics*. 2004; 20(17): 3252–3255.
22. Conant GC, Wolfe KH. GenomeVx: simple web-based creation of editable circular chromosome maps. *Bioinformatics*. 2008; 24(6): 861–862. <https://doi.org/10.1093/bioinformatics/btm598> PMID: 18227121
23. Thiel T, Michalek W, Varshney R, Graner A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *TAG Theoretical and Applied Genetics*. 2003; 106: 411–422.
24. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Research*. 2001; 29: 4633–42. <https://doi.org/10.1093/nar/29.22.4633> PMID: 11713313
25. Kelly AF, Lior P, Alexander P, Edward MR, Inna D. VISTA: computational tools for comparative genomics. *Nucleic Acids Research*. 2004; 32(2): W273–W279. Available from: <https://genome.lbl.gov/vista/mvista/submit.shtml>
26. Darling AE, Mau B, Perna NT. progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. *PLoS ONE*. 2010; 5(6): e11147. <https://doi.org/10.1371/journal.pone.0011147> PMID: 20593022
27. Amiryousssefi A, Hyövnén J, Poczei P. IRscope: an online program to visualize the junction sites of chloroplast genomes. *Bioinformatics*. 2018; 34(17): 3030–3031. Available from: <https://irscope.shinyapps.io/irapp> <https://doi.org/10.1093/bioinformatics/bty220> PMID: 29659705
28. Zhang D, Gao FL, Jakovlić I, Zou H, Zhang J, Li WX, et al. PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Molecular Ecology Resources*. 2020; 20(1): 348–355. <https://doi.org/10.1111/1755-0998.13096> PMID: 31599058
29. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*. 2013; 30(4): 772–780. <https://doi.org/10.1093/molbev/mst010> PMID: 23329690
30. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, et al. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution*. 2017; 34(12): 3299–3302. <https://doi.org/10.1093/molbev/msx248> PMID: 29029172

31. Felsenstein J. Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Biology*. 1978; 27(4): 401–410.
32. Ronquist F, Teslenko M, Mark PVD, Ayres L, Darling A, Höhna S, et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*. 2012; 61(3): 539–542. <https://doi.org/10.1093/sysbio/sys029> PMID: 22357727
33. Gemayel R, Vences MD, Legendre M, Verstrepen KJ. Variable tandem repeats accelerate evolution of coding and regulatory sequences. *Annual Review of Genetics*. 2010; 44: 445–477. <https://doi.org/10.1146/annurev-genet-072610-155046> PMID: 20809801
34. Dashnow H, Tan S, Das D, Eastel S, Oshlack A. Genotyping microsatellites in next-generation sequencing data. *BMC Bioinformatics*. 2015; 16(Suppl 2): A5.
35. Michal C, Meyza K, Chybicki I, Dzialuk A, Litkowiec M, Burczyk J. Chloroplast microsatellites as a tool for phylogeographic studies: the case of white oaks in Poland. *iForest—Biogeosciences and Forestry*. 2015; 8(6): 765–771.
36. Philippe H, Delsuc F, Brinkmann H, Lartillot N. Phylogenomics. *Annual Review Ecology Evolution and Systematics*. 2005; 36: 541–62.
37. Raubeson LA, Peery R, Chumley TW, Dziubek C, Fourcade HM, Boorem JL, et al. Comparative chloroplast genomics: Analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. *BMC Genomics*. 2007; 8: 174–201.
38. Wang RJ, Cheng CL, Chang CC, Wu CL, Su TM, Chaw SM. Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. *BMC Evolutionary Biology*. 2008; 8: 36–50. <https://doi.org/10.1186/1471-2148-8-36> PMID: 18237435
39. State Key Laboratory of Systematic and Evolutionary Botany (LSEB), IB, CAS. Information System of Chinese Rare and Endangered Plants (ISCREP). Available from: <https://www.plantplus.cn/rep/protlist>
40. Mao CL, Li XQ, Zhang FL, Yang T, Wu Y. The complete chloroplast genome sequence of *Knema elegans* (Myristicaceae). *Mitochondrial DNA Part B: Resources*. 2020; 5(1): 729–730.
41. Mao CL, Zhang FL, Li XQ, Yang T, Liu J, Wu Y. The complete chloroplast genome sequence of *Horsfieldia pandurifolia* (Myristicaceae). *Mitochondrial DNA Part B: Resources*. 2019; 4(1): 949–950.
42. Mao CL, Zhang FL, Yang T, Li XQ, Wu Y. The complete chloroplast genome sequence of *Myristica yunnanensis* (Myristicaceae). *Mitochondrial DNA Part B: Resources*. 2019; 4(1): 1871–1872.
43. Li XQ, Mao CL, Zhang FL, Yang T, Zhao Q, Wu Y. The complete chloroplast genome sequence of *Horsfieldia Kingii* (Myristicaceae). *Mitochondrial DNA Part B: Resources*. 2019; 4(2): 4184–4185.
44. Li XQ, Zhang FL, Mao CL, Yang T, Zhao Q, Wu Y. The complete chloroplast genome sequence of *Knema linifolia* (Myristicaceae). *Mitochondrial DNA Part B: Resources*. 2020; 5(3): 2918–2919.
45. Zhang FL, Mao CL, Li XQ, Yang T, Wu Y. The complete chloroplast genome sequence of *Horsfieldia amygdalina* (Myristicaceae). *Mitochondrial DNA Part B: Resources*. 2019; 4(2): 3923–3924.
46. Zhang FL, Yang T, Mao CL, Li XQ, Zhao Q, Wu Y. The complete chloroplast genome sequence of *Knema conferta* (Myristicaceae). *Mitochondrial DNA Part B: Resources*. 2020; 5(3): 3748–3749.
47. Yang T, Mao CL, Li XQ, Zhang FL, Zhao Q, Wu Y. The complete chloroplast genome sequence of *Knema furfuracea* (Myristicaceae). *Mitochondrial DNA Part B: Resources*. 2020; 5(1): 325–326.
48. Park J, Kim Y, Kwon W, Xi H, Kwon M. The complete chloroplast genome of tulip tree, *Liriodendron tulipifera* L. (Magnoliaceae): intra-species variation on mitochondrial genome. *Mitochondrial DNA Part B: Resources*. 2019; 4(1): 1308–1309.
49. Ai DY. Significance of repeat sequences in genomes. *Chemistry of Life*. 2008; 28(3): 343–345.
50. Maréchal A, Brisson N. Recombination and the maintenance of plant organelle genome stability. *New Phytologist*. 2010; 186(2): 299–317. <https://doi.org/10.1111/j.1469-8137.2010.03195.x> PMID: 20180912
51. Wiche S, Schneeweiss GM, Depamphilis CW, Müller KF, Quandt D. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Molecular Biology*. 2011; 76(3–5): 273–297. <https://doi.org/10.1007/s11103-011-9762-4> PMID: 21424877
52. Yang Y, Zhou T, Duan D, Yang J, Feng L, Zhao G. Comparative analysis of the complete chloroplast genomes of five *Quercus* species. *Front Plant Science*. 2016; 7: 959. <https://doi.org/10.3389/fpls.2016.00959> PMID: 27446185
53. Zhou T, Chen C, Wei Y, Chang Y, Bai G, Li Z, et al. Comparative transcriptome and chloroplast genome analyses of two related *Dipteronia* species. *Frontier in Plant Science*. 2016; 7: 1512.
54. Lu L, Li X, Hao Z, Yang L, Zhang J, Peng Y, et al. Phylogenetic studies and comparative chloroplast genome analyses elucidate the basal position of halophyte *Nitraria sibirica* (Nitrariaceae) in the Sapindales. *Mitochondrial DNA Part A: DNA Mapping, Sequencing, and Analysis*. 2018; 29(5): 745–755.

55. Ebert D, Peakall R. Chloroplast simple sequence repeats (cpSSRs): Technical resources and recommendations for expanding cpSSR discovery and applications to a wide array of plant species. *Molecular Ecology Resources*. 2009; 9(3): 673–690. <https://doi.org/10.1111/j.1755-0998.2008.02319.x> PMID: 21564725
56. Mohammad PN, Shabaniyan N, Khadivi A, Rahmani MS, Emami A. Genetic structure of gall oak (*Quercus infectoria*) characterized by nuclear and chloroplast SSR markers. *Tree Genetics and Genomes*. 2017; 13: 70.
57. Zeng JM, Chen XJ, Wu XF, Jiao FC, Xiao BG, Li YP, et al. Genetic diversity analysis of genus *Nicotiana* based on SSR markers in chloroplast genome and mitochondria genome. *Acta Tabacaria Sinica*. 2016; 22(4): 89–97.
58. Jane A, Barry WC, Ellis M. The role of insertions/deletions in the evolution of the intergenic region between psbA and trnH in the chloroplast genome. *Current Genetics*. 1988; 14: 137–146. <https://doi.org/10.1007/BF00569337> PMID: 3180272
59. Zhou T, Zhu H, Wang J, Xu Y, Xu F, Wang X. Complete chloroplast genome sequence determination of *Rheum* species and comparative chloroplast genomics for the members of Rumiceae. *Plant Cell Reports*. 2020; 39(6): 811–824.
60. Yu N, Jensen-Seaman MI, Chemnick L, Ryder O, Li WH. Nucleotide diversity in gorillas. *Genetics*. 2004; 166(3): 1375–1383. <https://doi.org/10.1534/genetics.166.3.1375> PMID: 15082556
61. Xu JH, Liu QX, Hu W, Want T, Xue Q, Messing J. Dynamics of chloroplast genome in green plants. *Genomics*. 2015; 106: 221–231.
62. Tong W, Kin TS, Park YJ. Rice chloroplast genome variation architecture and phylogenetic dissection in diverse *Oryza* species assessed by whole genome resequencing. *Rice*. 2016; 9: 57.
63. Swetha VP, Sheeja TE, Sasikumar B. DNA barcoding to resolve phylogenetic relationship in *Myristica* spp. *Journal of Spices and Aromatic Crops*. 2019; 28(2): 131–140.