



OPEN

# The complete chloroplast genome of critically endangered *Chimonobambusa hirtinoda* (Poaceae: *Chimonobambusa*) and phylogenetic analysis

Yanjiang Liu<sup>1,2</sup>, Xiao Zhu<sup>2</sup>, Mingli Wu<sup>1,2</sup>, Xue Xu<sup>1,2</sup>, Zhaoxia Dai<sup>3</sup>✉ & Guangqian Gou<sup>1,2</sup>

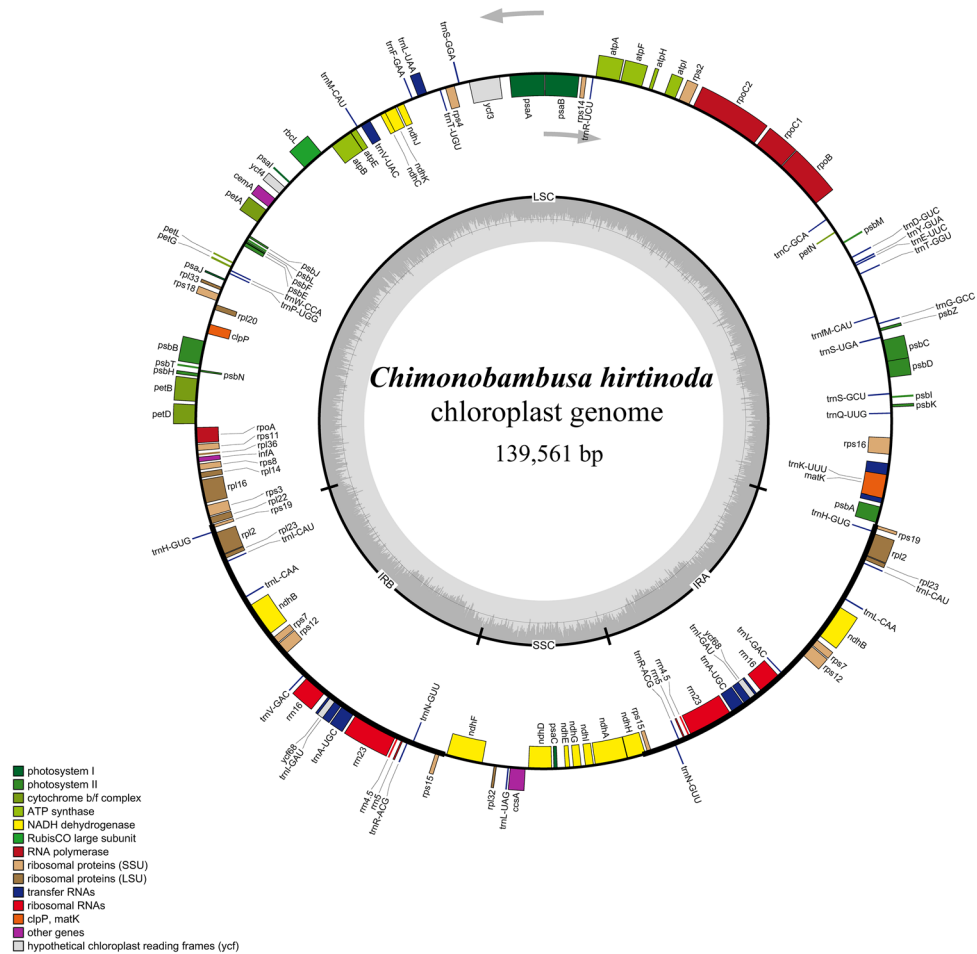
*Chimonobambusa hirtinoda*, a threatened species, is only naturally distributed in Doupeng Mountain, Duyun, Guizhou, China. Next-generation sequencing (NGS) is used to obtain the complete chloroplast (cp) genome sequence of *C. hirtinoda*. The sequence was assembled and analyzed for phylogenetic and evolutionary studies. Additionally, we compared the cp genome of *C. hirtinoda* with previously published *Chimonobambusa* species. The cp genome of *C. hirtinoda* has a total length of 139,561 bp and 38.90% GC content. This genome included a large single-copy (LSC) region of 83,166 bp, a small single-copy (SSC) region of 20,811 bp and a pair of inverted repeats of 21,792 bp each. We discovered 130 genes in the cp genome, including 85 protein-coding genes, 37 tRNA, and 8 rRNA genes. A total of 48 simple sequence repeats (SSRs) were detected. The A/U preference of the third nucleotide in the cp genome of *C. hirtinoda* was obtained by measuring the codon usage frequency of amino acids. Furthermore, phylogenetic analysis using complete cp sequences and *matK* gene revealed a genetic relationship within the *Chimonobambusa* genus. This study reported the chloroplast genome of the *C. hirtinoda*.

*Chimonobambusa* genus contains 37 species, 34 species distributed in China. Most of its bamboo shoots proliferate in autumn, which is delicious, nutritious and contain protein, amino acids and minerals<sup>1,2</sup>. Bamboo shoots are healthy and beneficial food for local people. Additionally, the bamboo is rich in cellulose pulp and is a high-quality raw material for papermaking, bamboo handicrafts, bamboo plywood, and craft furniture. As a result, *Chimonobambusa* has high economic, edible, and cultural value<sup>3-5</sup>.

*Chimonobambusa hirtinoda* was listed in the red catalogue by IUCN in 2007 and rated as a nationally endangered plant. *C. hirtinoda* is only naturally distributed in Doupeng Mountain, Duyun, Guizhou, China<sup>6,7</sup>. However, recently, with the tourism development in Doupeng Mountain and natural environmental losses, the habitat of *C. hirtinoda* has been damaged, and they are on the verge of extinction<sup>8</sup>. Thus, the natural resources of *C. hirtinoda* must be protected in all aspects. Most of the woody bamboos are primarily reproduced through rhizomes, with a random flowering period, and some species only bloom once in a lifetime<sup>9</sup>. The division of the Bambusoideae family is controversial because their classification and identification depend upon the morphological characteristics of vegetative organs. Consequently, their research from a morphological to molecular perspective is required to assess the classification and evolutionary relationship of bamboo species.

The origin of chloroplast (cp) is hypothesized to be obtained from cyanobacteria through endosymbiosis<sup>10</sup>. As the photosynthetic organelle of plant cells, chloroplast plays a key role in photosynthesis and has crucial implications for plant physiology and development<sup>11,12</sup>. The structure of the chloroplast genome is conservative, comprising a large single-copy (LSC) region, a small single-copy (SSC) region, and two inverted repeats (IRs) regions in opposite directions<sup>13</sup>. Therefore, chloroplast genomes have been widely used as DNA barcodes to identify species. Phylogenetic studies and chloroplast haplotypes are used to analyze the genetic diversity of species<sup>14-16</sup>.

<sup>1</sup>Key Laboratory of Plant Resource Conservation and Germplasm Innovation in Mountainous Region (Ministry of Education), College of Life Sciences/Institute of Agro-Bioengineering, Guizhou University, Guiyang 550025, Guizhou Province, China. <sup>2</sup>Bamboo Research Institute, Guizhou University, Guiyang 550025, China. <sup>3</sup>College of Forestry, Guizhou University, Guiyang 550025, China. ✉email: daizhaoxia1999@163.com



**Figure 1.** Chloroplast genome map of *C. hirtinoda*. Different colors represent different functional genes groups. Genes outside the circle indicate counterclockwise transcription, and genes inside the circle indicate clockwise transcription. The thick black line on the outer circle represents the two IR regions. The GC content is the dark gray area within the ring.

Several reports are present on the chloroplast genome of the Arundinariae in the Bambusoideae<sup>17</sup>, but limited data is available on the *Chimonobambusa* genus. Thus, this study reported the chloroplast genome of *C. hirtinoda*, including its gene content, codon usage, and its comparison with closed species. A phylogenetic relationship was constructed based on previously published cp genomes of Bambusoideae to clarify the taxonomic position of *C. hirtinoda*. These findings will provide valuable genetic resources for further research on the phylogenetic position of *C. hirtinoda* and investigate evolutionary relationships of the order Bambusoideae.

## Results

**Assembly and annotation of the chloroplast genomes.** Assembly resulted in a whole cp genome sequence of *C. hirtinoda* with a length of 139,561 bp (Fig. 1), consisting of 83,166 bp large single-copy region, 20,811 bp small single-copy regions, and two 21,792 bp IR regions, comprising the typical quadripartite structure of terrestrial plants. The cp genome of *C. hirtinoda* was annotated with 130 genes, including 85 protein-coding genes, 37 tRNA genes, and 8 rRNA genes (Table 1). Most of the 15 genes in the *C. hirtinoda* cp genome contain introns. Of these, 13 genes contain one intron (*atpF*, *ndhA*, *ndhB*, *petB*, *petD*, *rpl2*, *rpl16*, *rps16*, *trnA-UGC*, *trnI-GAU*, *trnK-UUU*, *trnL-UAA*, *trnV-UAC*) and only the gene *cyf3* includes two introns, and the gene *clpP* intron was deleted (Supplementary Table S1). The *rps12* gene contained two copies, and the three exons were spliced into a trans-splicing gene<sup>18</sup>.

The *accD*, *ycf1*, and *ycf2* genes were missing in the cp genome of *C. hirtinoda*, and the introns in the genes *clpP* and *rpoC1* were lost. This phenomenon is consistent with previous systematic evolutionary studies on the genome structure of plants in the Poaceae family<sup>19</sup>. The phenomenon of missing genes is reported in other plants<sup>20–23</sup>.

The total GC content in the *C. hirtinoda* cp genome was 38.90%, and the content for each of the four bases, A, T, G, and C, was 30.63%, 30.46%, 19.57%, and 19.33%, respectively (Table 2). The LSC region (36.98%) and SSC region (33.21%) exhibited much lower values than the IR region (44.23%), indicating a non-uniform distribution of the base contents in the cp genome, probably because of four rRNAs in the IR region, which in turn makes

Genome features	<i>C. hirtinoda</i>
Genome size (bp)	139,561
LSC size (bp)	83,166
SSC size (bp)	12,811
IR size (bp)	21,792
GC content (%)	38.9%
No. of genes	130
No. of PCGs	85
No. of tRNA	37
No. of rRNA	8

**Table 1.** Summary of the chloroplast genome of *C. hirtinoda*.

Region	Length (bp)	A (%)	T (%)	G (%)	C (%)	GC (%)
LSC	83,166	31.24	31.78	18.76	18.22	36.98
SSC	12,811	36.02	30.78	16.17	17.04	33.21
IRA	21,792	27.96	27.81	21.19	23.04	44.23
IRB	21,792	27.96	27.81	21.19	23.04	44.23
Total genome	139,561	30.63	30.46	19.57	19.33	38.90
CDS	60,531	29.63	30.85	21.20	18.3	39.53

**Table 2.** Base composition in the *C. hirtinoda* chloroplast genome.

the GC content higher in the IR region. These values were similar to cp genome results previously reported for some Poaceae plants<sup>24,25</sup>.

**Repeat sequences and codon analysis.** SSR consists of 10-bp-long base repeats and is widely used for exploring phylogenetic evolution and genetic diversity analysis<sup>26–29</sup>.

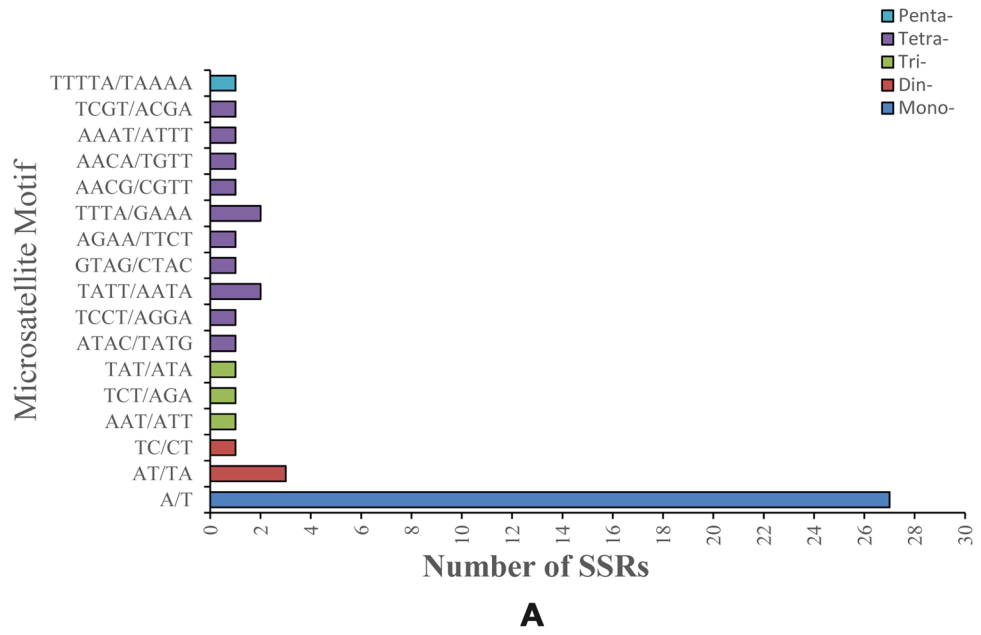
In total, 48 SSRs were detected in *C. hirtinoda*, including 27 mononucleotide versions, accounting for 56.25% of the total SSRs, primarily consisting of A or T. Additionally, four dinucleotide repeats consisting of AT/TA and TC/CT repeats, and 3 tri, 13 tetra, and 1 penta-repeats (Fig. 2A). From the SSRs distribution perspective, the majority (79%) of SSRs (38) were observed in the LSC area, whereas 6 SSRs in the IR region (13%) and 4 SSRs in the SSC region (8%) were discovered (Fig. 2B). Previous research suggests that the distribution of SSRs numbers in each region and the differences among locations in GC content are related to the expansion or contraction of the IR boundary<sup>30</sup>.

The REPuter program revealed that the cp genome of *C. hirtinoda* was identified with 61 repeats, consisting of 15 palindromic, 19 forward and no reverse and complement repeats (Fig. 3). We noticed that repeat analyses of three *Chimonobambusa* genus species exhibited 61–65 repeats, with only one reverse in *C. hejiangensis*. Most of the repeat lengths were between 30 and 100 bp, and the repeat sequences were located in either IR or LSC region<sup>31</sup> (Supplementary Table S2).

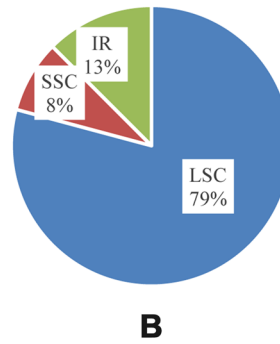
We identified 20,180 codons in the coding region of *C. hirtinoda* (Fig. 4, Supplementary Table S3). The codon AUU of Ile was the most used, and the TER of UAG was the least used codon (817 and 19), excluding the termination codons. Leu was the most encoded amino acid (2,170), and TER was the lowest (85). The Relative Synonymous Codon Usage (RSCU) value greater than 1.0 means a codon is used more frequently<sup>32</sup>. The RSCU values for 31 codons exceeded 1 in the *C. hirtinoda* cp genome, and of these, the third most frequent codon was A/U with 29 (93.55%), and the frequency of start codons AUG and UGG used demonstrated no bias (RSCU = 1).

**Comparative analysis of genome structure.** The nucleotide variability (Pi) values of the three cp genomes discovered in the *Chimonobambusa* genus species ranged from 0 to 0.021 with an average value of 0.000544, as demonstrated from DnaSP 5.10 software analysis. Five peaks were observed in the two single-copy regions, and the highest peak was present in the *trnT-trnE-trnY* region of the LSC region (Fig. 5). The Pi value for LSC and SSC is significantly higher than that of the IR region. In the IR region, highly different sequences were not observed, a highly conserved region. The sequences of these highly variable regions are reported in other plants during examinations for species identification, phylogenetic analysis, and population genetics research<sup>33–35</sup>.

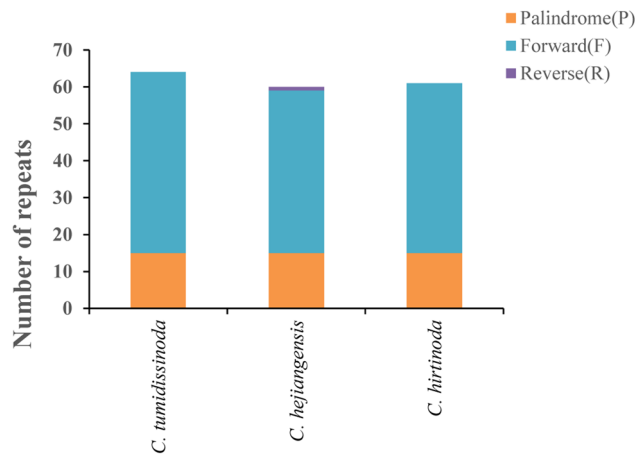
The structural information for the complete cp genomes among three *Chimonobambusa* genus species revealed that the sequences in most regions were conserved (Fig. 6). The LSC and SSC regions exhibit a remarkable degree of variation, higher than the IR region, and the non-coding region demonstrates higher variability than the coding region. In the non-coding areas, 7–9 k, 28–30 k, 36 k and other gene loci differed significantly.



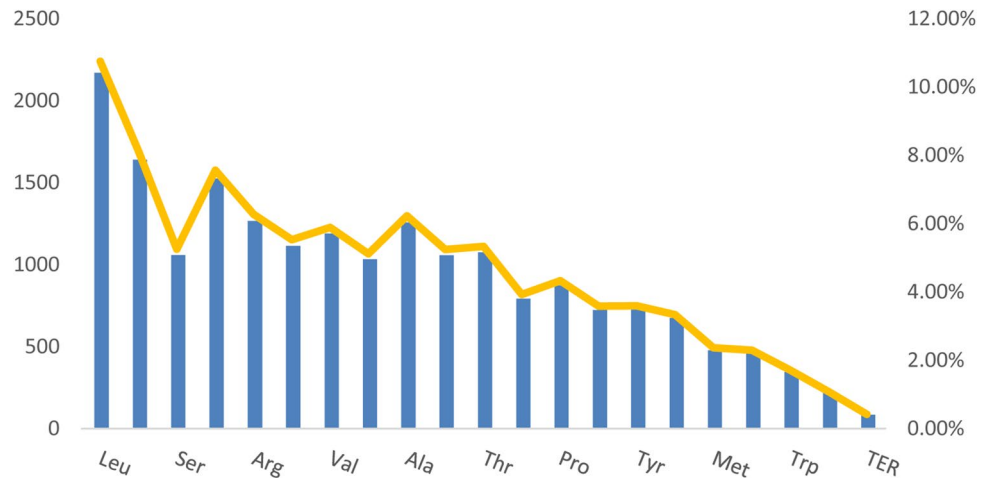
### Regional distribution



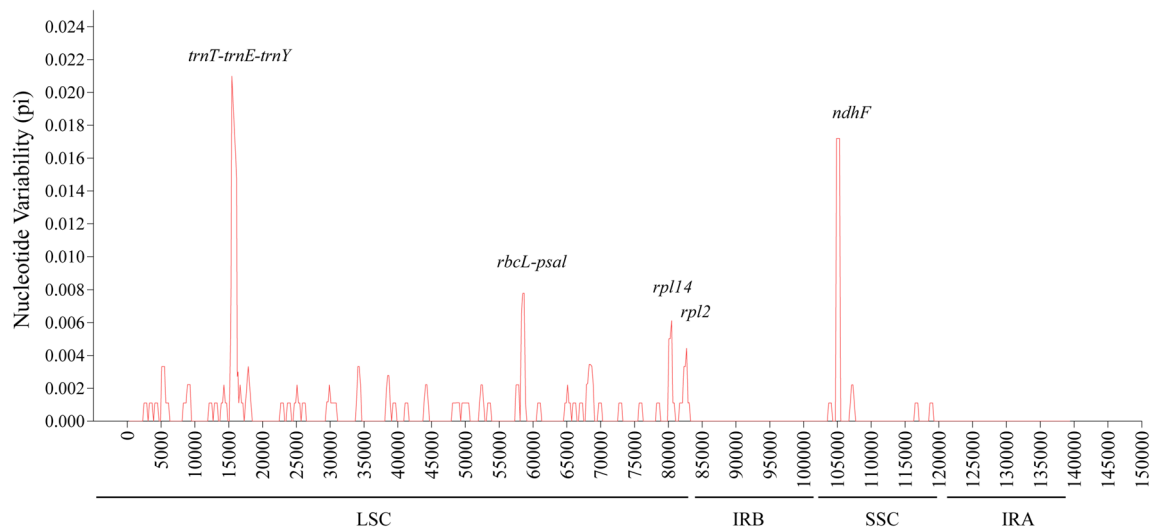
**Figure 2.** Analysis of simple sequence repeats in *C. hirtinoda* cp genome. (A) The percentage distribution of 45 SSRs in LSC, SSC, and IR regions. (B).



**Figure 3.** Information of chloroplast genome repeats of *Chimonobambusa* genus species.



**Figure 4.** Amino acid frequencies in *C. hirtinoda* cp genome protein coding sequences. The column diagrams indicate the number of amino acid codes, and the broken line indicates the proportion of amino acid codes.



**Figure 5.** Sliding window analysis of *Chimonobambusa* genus complete chloroplast genome sequences. X-axis: position of the midpoint of a window, Y-axis: nucleotide diversity of each window.

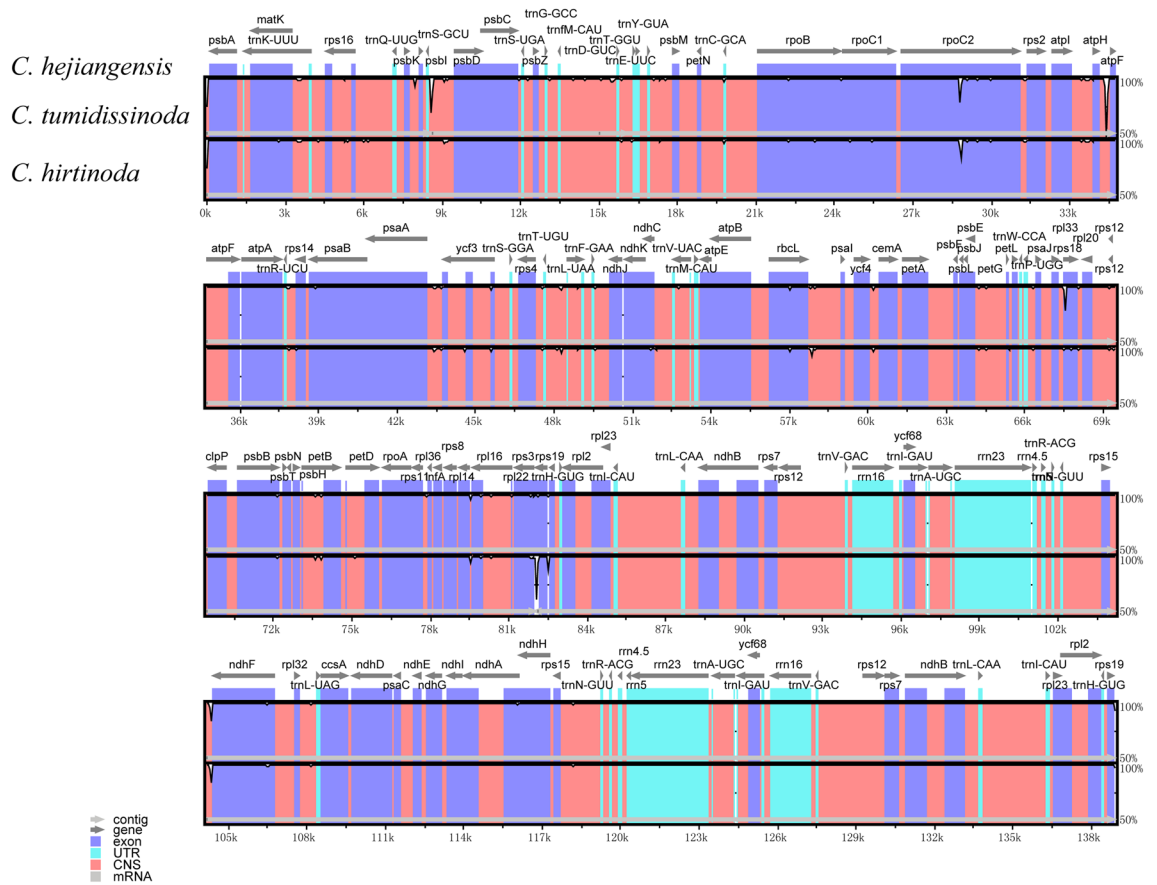
Genes *rpoC2*, *rps19*, *ndhJ* and other regions differ in the protein-coding region. However, the agreement between the tRNA and rRNA regions is 100%. A similar phenomenon has also been reported by others<sup>36</sup>.

**IR contraction and expansion in the chloroplast genome.** Due to the unique circular structure of the cp genome, there are four junctions between the LSC/IRB/SSC/IRA regions. During species evolution, the stability of the two IR regions sequences was ensured by the IR region of the chloroplast genome expanding and contracting to some degree, and this adjustment is the primary reason for chloroplast genome length variation<sup>37,38</sup>.

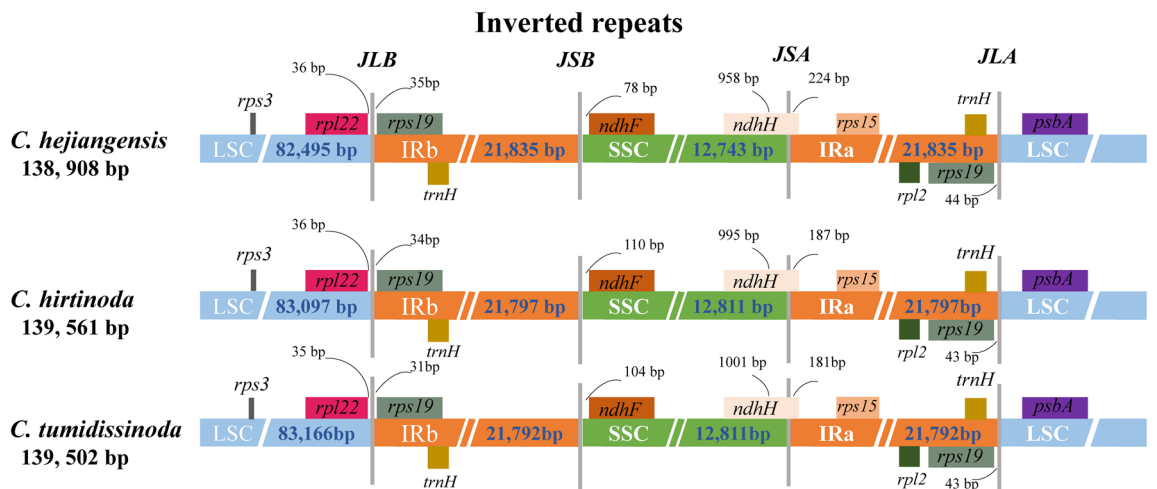
The variations at IR/SC boundary regions in the three *Chimonobambusa* genus chloroplast genomes were highly similar in the organization, gene content, and gene order. The size of IR ranges from 21,797 bp (*C. tumidissinoda*) to 21,835 bp (*C. hejiangensis*). The *ndhH* gene spans the SSC/IRa boundary, and this gene extended 181–224 bp into the IRa region for all three *Chimonobambusa* genus. The gene *rps19* was extended from the IRb to the LSC region with a 31–35 bp gap. The *rpl12* gene was located in the LSC region of all genomes, varied from 35–36 bp apart from the LSC/IRb (Fig. 7).

Three chloroplast genomes of the *Chimonobambusa* genus were compared using the Mauve alignment. The results showed that all sequences show perfect synteny conservation with no inversion or rearrangements (Fig. 8).

**Phylogenetic analysis.** We performed a phylogenetic analysis using the complete chloroplast genomes and *matK* gene reflecting the phylogenetic position of *C. hirtinoda*. The maximum likelihood (ML) analysis

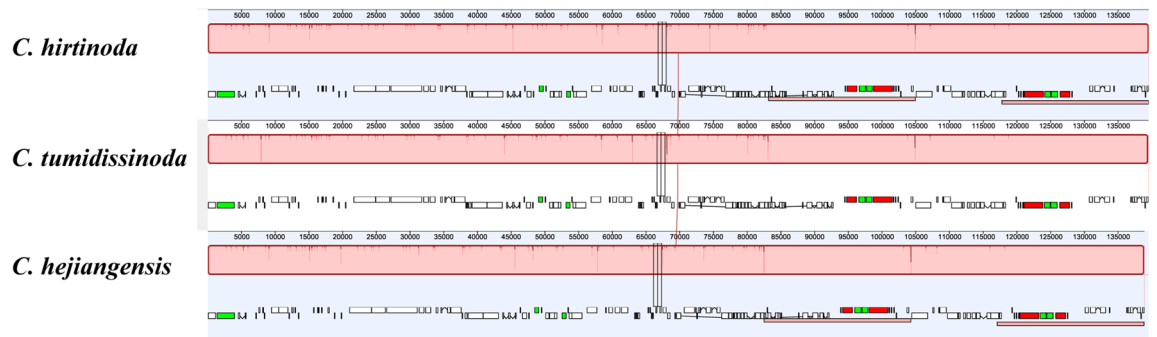


**Figure 6.** Visualization of genome alignment of three species chloroplast genome sequences using *Chimonobambusa hejiangensis* as reference. The vertical scale shows the percent of identity, ranging from 50 to 100%. The horizontal axis shows the coordinates within the cp genome. Those are some colors represents protein coding, intron, mRNA and conserved non-coding sequence, respectively.



**Figure 7.** Comparison of LSC, SSC and IR boundaries of chloroplast genomes among the three *Chimonobambusa* species. The LSC, SSC and IRs regions are represented with different colors. JLB, JSB, JSA and JLA represent the connecting sites between the corresponding regions of the genome, respectively. Genes are showed by boxes.





**Figure 8.** The chloroplast genomes of three *Chimonobambusa* species rearranged by the software MAUVE. Locally collinear blocks (LCBs) are represented by the same color blocks connected by lines. The vertical line indicates the degree of conservatism among position. The small red bar represents rRNA.

based on the complete chloroplast genomes indicated seven nodes with entirely branch support (100% bootstrap value). However, the three *Chimonobambusa* genera exhibited a moderate relationship due to fewer samples used, supporting that *C. hirtinoda* is closely related to *C. tumidissinoda* with a 62% bootstrap value more than *C. hejiangensis*. A phylogenetic tree based on the *matK* gene revealed that *Chimonobambusa* species clustered in one branch was consistent with the phylogenetic tree constructed by the complete cp genome tree (Fig. 9). The results show that the whole chloroplast genome identified related species better than the former, consistent with the previous study<sup>39</sup>.

## Discussion

In the current study, *Chimonobambusa* genus exhibited a typical circular tetrad structure, similar to most species. The flowering cycle in bamboo is long and unfixed because of its specific characteristics, and the morphological reproductive and nutritional traits are often difficult to identify because of environmental changes. Therefore, their classification based on morphological details is controversial, leading to unreliable systematic research results<sup>40</sup>. Molecular biology and sequencing technology are more practical methods for their classification. The analysis of plant genetic diversity, species formation, and genetic differentiation is also studied<sup>41</sup>. Therefore, mutation sites observed in the whole chloroplast genome can be used as super bar codes to study the phylogeny and molecular taxonomy of bamboo species.

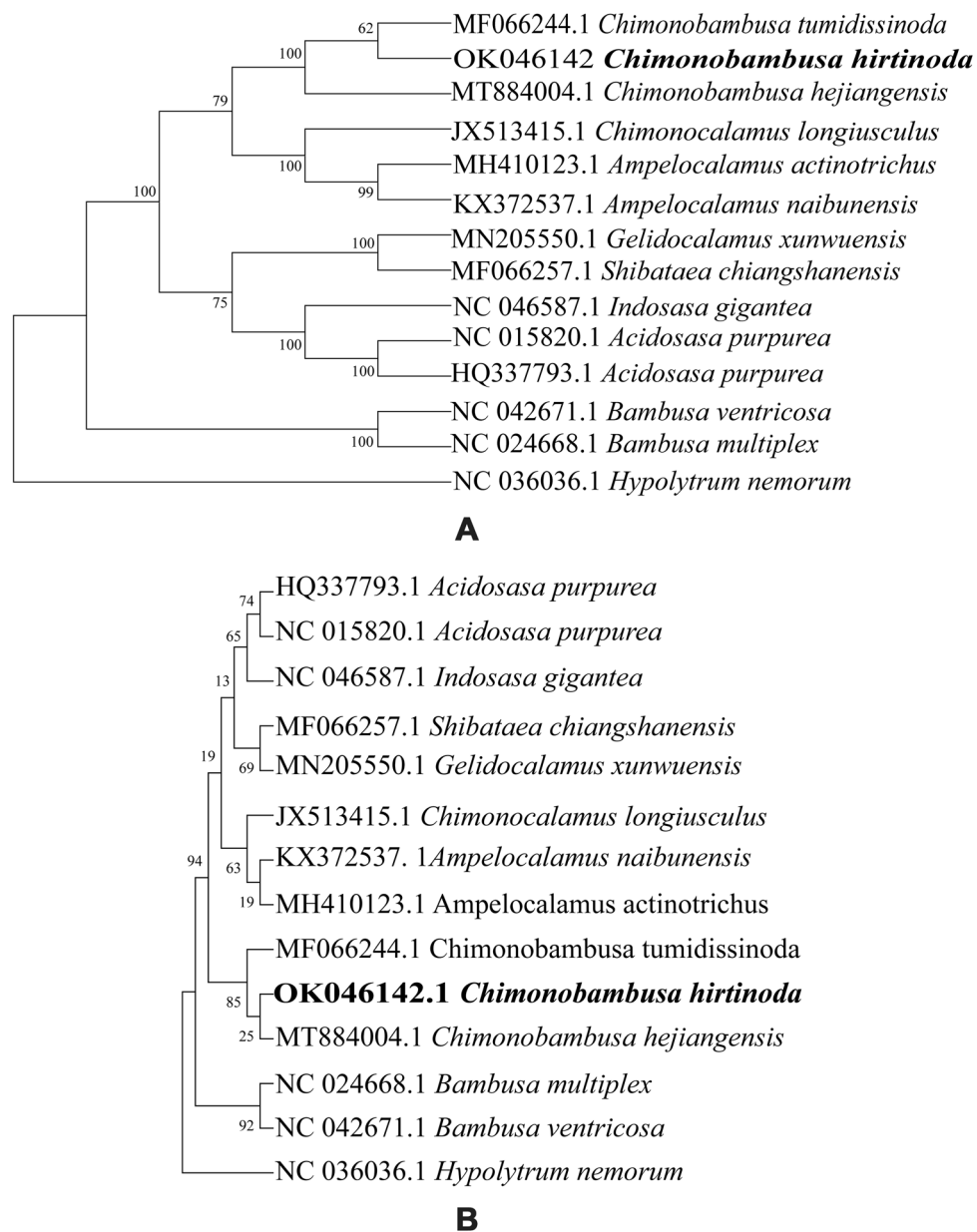
Genetic diversity is one of the crucial indicators to measure the degree of variation. The higher the degree of genetic diversity, the higher the genetic diversity and abundant genetic resources in the population. Genetic diversity is helpful to improve the adaptability of the species to climate change and historical events and provide scientific and effective strategies for the protection and management of germplasm resources of endangered species. Huang et al. used 16 SSR primers to explore the population genetic diversity and genetic differentiation of the endangered plant *Camellia chekiangoleosa*, and the species conservation strategy was formulated based on the research results<sup>42</sup>.

## Conclusions

The current study primarily explored the chloroplast genome of *C. hirtinoda* and compared it with related species within *Chimonobambusa* genus. These data provide valuable genetic information that advances the genetic research on *Chimonobambusa*. A phenomenon of genes loss was discovered by successfully assembling, annotating, and analyzing the whole chloroplast genome sequence of *C. hirtinoda*. The loss is associated with the rapid evolution of the Poaceae species and the extensive rearrangements of chloroplast structures during the evolutionary process. The acquisition of these data, particularly in terms of SSRs, will enhance the study of the phylogenetic relationships of *Chimonobambusa* plants, their cp genome variation, and gene function.

IR region is the most conserved region of the chloroplast genome. The expansion and contraction of the IR region impact understanding of the evolution of plant population development. The IR region showed no significant difference in the three *Chimonobambusa* species in our current research. In addition, a comparative analysis of the *Chimonobambusa* species revealed that coding regions of the cp genome are more conservative than non-coding regions. Such a change in genetic structure can reflect a relationship with the changes in species, but the mechanism that generates such variations and the subsequent results require further study.

The Poaceae family is generally divided into two large evolutionary branches (BEP and PACCMAD), among which the Bambusoideae, Pooideae, and Oryzoideae belong to the BEP branch. Panicoideae, Arundinoideae, Chloridoideae, Aristidoideae, Arundinoideae, and Micraioideae belong to the PACCMAD branch. Here, the complete chloroplast genomes based phylogenetic tree (Fig. 9A) revealed high bootstrap support values (only three values less than 80%), and these species can be polymerized into two clades and an outgroup. The genus *Bambusa* constitutes an isolated evolutionary branch, becoming a monophyletic group, a conclusion that is consistent with the previous reports<sup>43</sup>. The two phylogenetic trees revealed that *C. hirtinoda*, *C. hejiangensis* and *C. tumidissinoda* formed a closely related group. In the evolutionary subclade of the second branch, the genus *Ampelocalamus* and *C. longiusculus* have a very close relationship.



**Figure 9.** Maximum likelihood phylogenetic tree based on the complete chloroplast genomes (A) and *matK* gene (B).

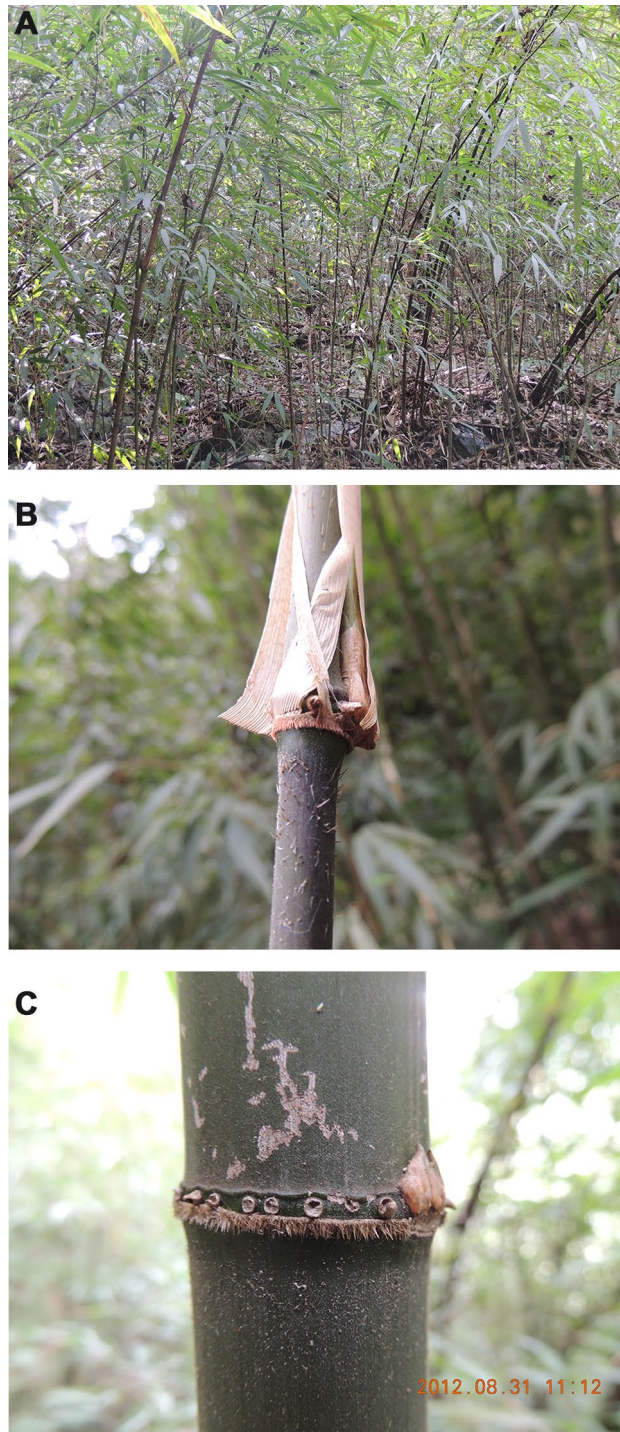
## Materials and methods

**Plant materials.** Four to six fresh leaves of *C. hirtinoda* from their native habitats, Doupeng Mountain, Guizhou Province, China (N 26°22'32.55", E 107°22'9", altitude 1074.93 m), were collected. The fresh leaves were dried with silica gel and stored at The Natural Museum of Guizhou University (accession number: GACP) for further DNA extraction (Fig. 10). The species identification was performed by Professor Guangqian Gou, the director of the Bamboo Research Institute in the college of life sciences, Guizhou University.

The collection and experiments of plant materials have complied with relevant guidelines and regulations of the Doupeng Mountain virgin forest nature reserve.

**DNA extraction, Chloroplast genome sequencing.** Total genomic DNAs were extracted from the sample using the TIANGEN DNA extraction kit (TIANGEN BIOTECH CO., Beijing, China), and the DNA concentration was detected using spectrophotometry. Total DNA quality was detected using 1% agarose gel electrophoresis. All the DNA obtained from *C. hirtinoda* was sent to BGI (Wuhan, China. <https://www.genomics.cn>), and the total DNA was sequenced using an Illumina sequencer with an HiSeq2500 system, with the library type selected to be the De Novo Sequencing  $\leq 800$  bp conventional library.





**Figure 10.** Morphological characteristic of *C. hirtinoda*. (A) The wild habitat of *C. hirtinoda*; (B) nodal ridges; (C) rings of root thorns.

**Genome assembly and annotation.** De novo assembly of the *C. hirtinoda* cp genome was performed in a GetOrganelle pipeline<sup>44</sup> (<https://github.com/Kinggerm/GetOrganelle>). Using the complete chloroplast genome of *C. hejiangensis* (MT884004) as a reference sequence, PGA (<https://github.com/quxiaojian/PGA>) annotated the chloroplast genes of *C. hirtinoda*, followed by manual correction using the software Generous 10.0.5<sup>45</sup>. The corrected sequence was uploaded to NCBI. The GenBank accession number of *C. hirtinoda* was OK046142.

Using the online software package Organellar Genome DRAW (OGDraw)<sup>46</sup> (<http://ogdraw.mpimp-golm.mpg.de/index.shtml>), a physical map of the chloroplast group was created.

**Structural of the *C. hirtinoda* cp genome.** A simple repeat sequence (SSR) is also called a microsatellite. SSR can be detected using the identification tools MISA<sup>47</sup> and REPuter<sup>48</sup>. The number for the repeat parameter was set to at least 10, 5, 4, 3, 3, and 3 repeat units, from mononucleotide to hexanucleotide. The codon bias for the chloroplast genome was analyzed using the software package CodonW1.4.2 (<http://downloads.fyxm.net/CodonW-76666.html>).

**Sequence divergence.** Determining the nucleotide diversity of the whole cp genome can make the identification of related species more accurate and help solve similar problems in the phylogenetic research<sup>49,50</sup>. To compare the differences, three species of *Chimonobambusa* were selected, using *C. hejiangensis* as a reference sequence. The software package MAFFT<sup>51</sup> was used to compare the whole cp genomes of the three species, with the comparison results manually truncated at both ends. Then the software package DnaSP5.10 was used to calculate the Pi values among species sequences<sup>51</sup>. The sliding window was set to 600, and the step size was 200. The online program mVISTA<sup>52</sup> was used to compare three species, using *C. hejiangensis* annotation as the reference. The software MAUVE<sup>53</sup> provided rearrangements of those gene sequences.

**Phylogenetic analyses.** For 14 sequences of complete chloroplast genome sequences and *matK* gene of Bambusoideae species and *Hypolytrum nemorum* (Cyperaceae: *Hypolytrum*) was selected as an outgroup for the construction of the phylogenetic tree to identify the taxonomic position of the *C. hirtinoda*. All sequences were aligned using the tool MAFFT, and the maximum likelihood (ML) phylogenetic tree was constructed using the software package MEGA-X<sup>54</sup>, and the bootstrap replicates parameter was set to 1000.

**Specimen collection statement.** The collection of fresh leaves obtained the permission of the nature reserve.

### Data availability

The complete chloroplast sequence generated and analyzed during the current study is available in GenBank (<https://www.ncbi.nlm.nih.gov>, accession numbers are described in the text).

Received: 26 October 2021; Accepted: 23 May 2022

Published online: 10 June 2022

### References

- Liu, Y. J. *et al.* Nutrition components in shoots of different provenance of *Chimonobambusa hejiangensis* and *C. pachystachys*. *J. Zhejiang For. Sci. Technol.* **32**(04), 37–42 (2012).
- Gou, G. Q. *et al.* Nutrient analysis on bamboo shoots of three species in *Chimonobambusa*. *China Veg.* **16**, 79–81 (2010).
- Gao, Q. *et al.* Nitrogen self-doped activated carbons with narrow pore size distribution from bamboo shoot shells. *Colloids Surf. A Physicochem. Eng. Asp.* **629**, 0927–7757. <https://doi.org/10.1016/j.colsurfa.2021.127408> (2021).
- Zhou, Q. L. Effects of three kinds of economic forests on soil physical and chemical properties in Alpine Region. *Tibe J. Agric. Sci.* **42**, 51–53 (2021).
- Chen, G. J. *et al.* Characterization of physicochemical properties and antioxidant activity of polysaccharides from shoot residues of bamboo (*Chimonobambusa quadrangularis*): Effect of drying procedures. *Food Chem.* **292**, 281–293. <https://doi.org/10.1016/j.foodchem.2019.04.060> (2019).
- Editorial Committee of Chinese Flora. Chinese Academy of Sciences. Flora of China, Vol. 9 333 (Beijing, Science Press, 1996).
- Gou, S. & Xie, Y. *Red List of Chinese Species* Vol. 1 (Higher Education Press, Beijing, 2004).
- Su, C. H., Zhu, S. X., Zhang, H. & Luo, G. Research on rhythm of shooting and growth for critically endangered *Chimonobambusa hirtinoda*. *J. Fujian For. Sci. Tech.* **43**(03), 153–156. <https://doi.org/10.13428/j.cnki.fjlk.2016.03.031> (2016).
- de Carvalho, A. L. *et al.* Bamboo-dominated forests of the southwest Amazon: Detection, spatial extent, life cycle length and flowering waves. *PLoS ONE* **8**(1), e54852. <https://doi.org/10.1371/journal.pone.0054852> (2013).
- Zhu, R. M. *et al.* Arabidopsis Chloroplast protein for Growth and Fertility1 (CGF1) and CGF2 are essential for chloroplast development and female gametogenesis. *BMC Plant Biol.* **20**(1), 172. <https://doi.org/10.1186/s12870-020-02393-5> (2020).
- Bhattacharya, O., Ortiz, I. & Walling, L. L. Methodology: An optimized, high-yield tomato leaf chloroplast isolation and stroma extraction protocol for proteomics analyses and identification of chloroplast co-localizing proteins. *Plant Methods* **16**, 131. <https://doi.org/10.1186/s13007-020-00667-5> (2020).
- Boudreau, E. *et al.* The Nac2 gene of *Chlamydomonas* encodes a chloroplast TPR-like protein involved in psbD mRNA stability. *EMBO J.* **19**(13), 3366–3376. <https://doi.org/10.1093/emboj/19.13.3366> (2000).
- Paudel, Y. P. *et al.* Chloroplast genome analysis of *Chrysotila dentata*. *Gene* **804**, 145871. <https://doi.org/10.1016/j.gene.2021.145871> (2021).
- Wu, L. *et al.* Comparative and phylogenetic analyses of the chloroplast genomes of species of Paeoniaceae. *Sci. Rep.* **11**, 14643. <https://doi.org/10.1038/s41598-021-94137-0> (2021).
- Kim, S. & Suh, Y. Phylogeny of Magnoliaceae based on ten chloroplast DNA regions. *J. Plant Biol.* **56**, 290–305. <https://doi.org/10.1007/s12374-013-0111-9> (2013).
- Zhang, T. T. *et al.* Evolutionary analysis of chloroplast tRNA of Gymnosperm revealed the novel structural variation and evolutionary aspect. *PeerJ* **8**, e10312. <https://doi.org/10.7717/peerj.10312> (2020).
- Choi, K. S., Chung, M. G. & Park, S. The complete chloroplast genome sequences of three veronicaeae species (plantaginaceae): Comparative analysis and highly divergent regions. *Front. Plant Sci.* **7**, 355. <https://doi.org/10.3389/fpls.2016.00355> (2016).
- Androsiuk, P. *et al.* The complete chloroplast genome of *Colobanthus apetalus* (Labill.) Druce: Genome organization and comparison with related species. *PeerJ* **6**, e4723. <https://doi.org/10.7717/peerj.4723> (2018).
- Su, Q. *et al.* The complete chloroplast genomes of seventeen *Aegilops tauschii*: Genome comparative analysis and phylogenetic inference. *PeerJ* **8**, e8678. <https://doi.org/10.7717/peerj.8678> (2020).
- Lubna. *et al.* Complete chloroplast genome characterization of oxalis corniculata and its comparison with related species from family oxalidaceae. *Plants* **9**(8), 928. <https://doi.org/10.3390/plants9080928> (2020).
- Herdenberger, F., Pillay, D. T. & Steinmetz, A. Sequence of the trnH gene and the inverted repeat structure deletion site of the broad bean chloroplast genome. *Nucleic Acids Res.* **18**(5), 1297 (1990).

22. Li, Y. *et al.* Gene losses and partial deletion of small single-copy regions of the chloroplast genomes of two hemiparasitic *Taxillus* species. *Sci. Rep.* **7**(1), 12834. <https://doi.org/10.1038/s41598-017-13401-4> (2017).
23. Nie, L. *et al.* Gene losses and variations in chloroplast genome of parasitic plant *macrosolea* and phylogenetic relationships within santalales. *Int. J. Mol. Sci.* **20**(22), 5812. <https://doi.org/10.3390/ijms20225812> (2019).
24. Ogihara, Y. *et al.* Structural features of a wheat plastome as revealed by complete sequencing of chloroplast DNA. *Mol. Genet. Genomics* **266**(5), 740–746. <https://doi.org/10.1007/s00438-001-0606-9> (2002).
25. Maier, R. M., Neckermann, K., Igloi, G. L. & Kössel, H. Complete sequence of the maize chloroplast genome: Gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. *J. Mol. Biol.* **251**(5), 614–628. <https://doi.org/10.1006/jmbi.1995.0460> (1995).
26. Huang, C. J. *et al.* Isolation and characterization of SSR and EST-SSR loci in *Chamaecyparis formosensis* (Cupressaceae). *Appl. Plant Sci.* **6**(8), e01175. <https://doi.org/10.1002/aps3.1175> (2018).
27. Wen, M. *et al.* Development of EST-SSR and genomic-SSR markers to assess genetic diversity in *Jatropha Curcas* L. *BMC Res. Notes* **3**, 42. <https://doi.org/10.1186/1756-0500-3-42> (2010).
28. Zhang, M., Mao, W., Zhang, G. & Wu, F. Development and characterization of polymorphic EST-SSR and genomic SSR markers for Tibetan annual wild barley. *PLoS ONE* **9**(4), e94881. <https://doi.org/10.1371/journal.pone.0094881> (2014).
29. Parthiban, S., Govindaraj, P. & Senthilkumar, S. Comparison of relative efficiency of genomic SSR and EST-SSR markers in estimating genetic diversity in sugarcane. *3 Biotech* **8**(3), 144. <https://doi.org/10.1007/s13205-018-1172-8> (2018).
30. Wei, X. K. *et al.* Complete chloroplast genomes of *Achnatherum inebrians* and comparative analyses with related species from Poaceae. *FEBS Open Bio* **11**(6), 1704–1718. <https://doi.org/10.1002/2211-5463.13170> (2021).
31. Saina, J. K., Li, Z. Z., Gichira, A. W. & Liao, Y. Y. The complete chloroplast genome sequence of tree of heaven (*Ailanthus altissima* (Mill.) (Sapindales: Simaroubaceae), an important pantropical tree. *Int. J. Mol. Sci.* **19**(4), 929. <https://doi.org/10.3390/ijms19040929> (2019).
32. Sarkar, I. *et al.* *Turdoides affinis* mitogenome reveals the translational efficiency and importance of NADH dehydrogenase complex-I in the Leiothrichidae family. *Sci. Rep.* **10**(1), 16202. <https://doi.org/10.1038/s41598-020-72674-4> (2020).
33. Bi, Y. *et al.* Chloroplast genomic resources for phylogeny and DNA barcoding: A case study on *Fritillaria*. *Sci. Rep.* **8**(1), 1184. <https://doi.org/10.1038/s41598-018-19591-9> (2018).
34. Li, Z. *et al.* The complete chloroplast genome sequence of tung tree (*Vernicia fordii*): Organization and phylogenetic relationships with other angiosperms. *Sci. Rep.* **7**(1), 1869. <https://doi.org/10.1038/s41598-017-02076-6> (2017).
35. Souza, U. J. B. *et al.* The complete chloroplast genome of *Stryphnodendron adstringens* (Leguminosae - Caesalpinioideae): Comparative analysis with related Mimosoid species. *Sci. Rep.* **9**(1), 14206. <https://doi.org/10.1038/s41598-019-50620-3> (2019).
36. Hand, M. L., Spangenberg, G. C., Forster, J. W. & Cogan, N. O. Plastome sequence determination and comparative analysis for members of the *Lolium-Festuca* grass species complex. *G3 (Bethesda)* **3**(4), 607–616. <https://doi.org/10.1534/g3.112.005264> (2013).
37. Pang, X. *et al.* Species identification of oaks (*Quercus* L., Fagaceae) from gene to genome. *Int. J. Mol. Sci.* **20**(23), 5940 (2019).
38. Zhang, X. *et al.* Comparative analyses of chloroplast genomes of cucurbitaceae species: Lights into selective pressures and phylogenetic relationships. *Molecules* **23**(9), 2165 (2018).
39. Nguyen, H. Q. *et al.* Complete chloroplast genome of novel *Adrinandra megaphylla* Hu species: Molecular structure, comparative and phylogenetic analysis. *Sci. Rep.* **11**, 11731. <https://doi.org/10.1038/s41598-021-91071-z> (2020).
40. Das, M., Bhattacharya, S., Basak, J. & Pal, A. Phylogenetic relationships among the bamboo species as revealed by morphological characters and polymorphism analyses. *Biol. Plant* **51**, 667–672. <https://doi.org/10.1007/s10535-007-0140-7> (2007).
41. Zhang, X., Chen, S. Y., Chen, P. & Liang, H. The complete chloroplast genome of *Chusquea culeou* (Poaceae: Bambusoideae: Bambuseae). *Mitochondrial DNA B* **4**(1), 91–92. <https://doi.org/10.1080/23802359.2018.1536479> (2019).
42. Huang, B. *et al.* Population genetic structure analysis reveals significant differentiation of the endemic species *Camellia chekiangoleosa* Hu. with a narrow geographic range. *Forests* **13**(2), 234. <https://doi.org/10.3390/f13020234> (2022).
43. Castiglione, S. Genetic diversity and differentiation of eleven medicago species from campania region revealed by nuclear and chloroplast microsatellites markers. *Genes* **13**(1), 97. <https://doi.org/10.3390/genes13010097> (2021).
44. Jin J. J. *et al.* GetOrganelle: A simple and fast pipeline for de novo assembly of a complete circular chloroplast genome using genome skimming data. *BioRxiv* <https://doi.org/10.1101/256479> (2018).
45. Kearse, M. *et al.* Geneious basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**(12), 1647–1649 (2012).
46. Lohse, M., Drechsel, O. & Bock, R. Organellar Genome DRAW (OGDRAW): A tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr. Genet.* **52**(5–6), 267–274. <https://doi.org/10.1007/s00294-007-0161-y> (2007).
47. Thiel, T., Michalek, W., Varshney, R. K. & Graner, A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor. App. Genet.* **106**(3), 411–422 (2003).
48. Kurtz, S. *et al.* REPuter: The manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* **29**(22), 4633–4642 (2001).
49. Collevatti, R. G., Terribile, L. C., Rabelo, S. G. & Lima-Ribeiro, M. S. Relaxed random walk model coupled with ecological niche modeling unravel the dispersal dynamics of a Neotropical savanna tree species in the deeper Quaternary. *Front. Plant Sci.* **6**, 653. <https://doi.org/10.3389/fpls.2015.00653> (2015).
50. Du, Y. P. *et al.* Complete chloroplast genome sequences of *Lilium*: Insights into evolutionary dynamics and phylogenetic analyses. *Sci. Rep.* **7**(1), 5751. <https://doi.org/10.1038/s41598-017-06210-2> (2017).
51. Katoh, K., Rozewicki, J. & Yamada, K. D. MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* **20**(4), 1160–1166. <https://doi.org/10.1093/bib/bbx108> (2019).
52. Librado, P. & Rozas, J. Dnasp v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451–1452. <https://doi.org/10.1093/bioinformatics/btp187> (2009).
53. Frazer, K. A., Pachter, L., Poliakov, A., Rubin, E. M. & Dubchak, I. mVISTA: Computational tools for comparative genomics. *Nucleic Acids Res.* **32**, W273–W279 (2004).
54. Kurtz, S. *et al.* Versatile and open software for comparing large genomes. *Genome Biol.* **5**, R12. <https://doi.org/10.1186/gb-2004-5-2-r12> (2004).

## Acknowledgements

This study was supported by The key Project of Guizhou Provincial Department of Education (Qian-Jiao-He KY [2019]062).

## Author contributions

Y.L. conceived and designed the experiments, performed the experiments, analyzed the data, provided reagents/materials/analysis tools, prepared figures and/or tables, wrote and reviewed drafts of the manuscript, and approved the final draft. X.Z. analyzed the data, contributed reagents/materials/analysis tools, authored and reviewed drafts of the manuscript, and approved the final draft. M.W. and X.X. prepared figures and/or tables,

authored and reviewed drafts of the manuscript, and approved the final draft. Z.D. and G.G. conceived and designed the experiments, authored or reviewed drafts of the manuscript, and approved the final draft.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-13204-2>.

**Correspondence** and requests for materials should be addressed to Z.D.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022