Identification of Laportea bulbifera using the Complete Chloroplast Genome as a Super-Barcode

Wenting wang (wangwenttt@163.com)
China Academy of Chinese Medical Sciences Institute of Chinese Materia Medica

Xingwen Wang
China Academy of Chinese Medical Sciences Institute of Chinese Materia Medica

Yuhua Shi
China Academy of Chinese Medical Sciences Institute of Chinese Materia Medica

Qinggang yin
China Academy of Chinese Medical Sciences Institute of Chinese Materia Medica

Ranran Gao
China Academy of Chinese Medical Sciences Institute of Chinese Materia Medica

Mengyue Wang
China Academy of Chinese Medical Sciences Institute of Chinese Materia Medica

Li Xiang
China Academy of Chinese Medical Sciences Institute of Chinese Materia Medica

Lan Wu
China Academy of Chinese Medical Sciences Institute of Chinese Materia Medica

Research Article

Keywords: Laportea bulbifera, chloroplast genome, super-barcode, species identification, Miao medicine

Posted Date: July 13th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1793400/v1

Version of Record: A version of this preprint was published at Journal of Applied Genetics on January 12th, 2023. See the published version at https://doi.org/10.1007/s13353-022-00746-4.

Abstract

Laportea bulbifera, a Miao medicine grown in karst areas, has exerts a unique curative effect on skin itching in the elderly, with an annual sales of > 100 million Yuan. Owing to the shortage of resources and large morphological variations in *L. bulbifera*, it is difficult to identify the species correctly using only traditional methods, which seriously affects the safety of drug usage for patients. This study obtained the complete high-quality *L. bulbifera* chloroplast (cp) genome, using second- and third-generation high-throughput sequencing. The cp genome was 149,911 bp in length, with a typical quadripartite structure. A total of 127 genes were annotated, including 83 protein-coding genes, 36 tRNA genes, and 8 rRNA genes. There was an inverted small single copy (SSC) structure in the *L. bulbifera* cp genome, one large-scale rearrangement of ~ 39 kb excised in the SSC and IR regions. The complete cp genome sequence used as a super-barcode and the highly variable regions (*ycf1*, *matK*, and *ndhD*), especially *matK*, can be used as specific barcodes to accurately distinguish *L. bulbifera* from counterfeits and closely related species. This study is important for the identification of *L. bulbifera* and lays a theoretical foundation for elucidating the phylogenetic relationship of the species.

1. Introduction

There are 28 *Laportea* species in Urticaceae worldwide, of which 8 species and 3 subspecies are distributed in China. The plants of this genus display a high efficacy in treating conditions such as rheumatism, diabetes, traumatic injury, and skin pruritus. The most widely studied and applied plant in the genus is *Laportea bulbifera*, commonly known as *Honghuoma* and *Urtica dentata hand*. It is often used as a medicine with its roots and fresh or dried whole grass (Y.-r. Chen et al. 2019) to dispel wind and dampness, promote blood circulation, and relieve pain. *L. bulbifera* is rich in secondary metabolites, such as flavonoids, coumarins, phenylpropanoids, phenolic acids, and sterols (Han et al. 2018; Dan et al. 2019; Zhu et al. 2011; Tang et al. 2018; Z. Su et al. 2009; H.-J. HU 2020). The plant possesses a multitude of pharmacological activities, including anti-inflammatory (H-Y 2018; Lu et al. 2012), analgesic (Z. Su et al. 2009; Lin et al. 2012), immunosuppressive (J. Wang et al. 2013; M. Xiang et al. 2013; Hou et al. 2010), antirheumatoid arthritis (RA) (X. Luo et al. 2011; L. Xiang 2010; M Xiang et al. 2006; Yao 2008; Han et al. 2018), antioxidant (Y. Chen et al. 2020), hypoglycemic(J. Wang et al. 2013; Zengyu 2009), and lipid-lowering (Zengyu 2009; QingXin 2018) activities. *L. bulbifera* is mainly produced in Guizhou Province, China, and distributed sporadically in karst landform areas known for their dry, alkaline, high calcium, and sediment loss (Chunni Liu et al. 2021). Owing to the precise and excellent therapeutic effects, *L. bulbifera* has gradually entered the herbal medicine market. At present, Guizhou ethnic medicines that use *L. bulbifera* as the main raw material and have passed the national drug standards include “Runzao Zhiyang Capsule,” “Compound Shangfuning Cataplasmata,” and “Six Flavor Fushangning Tincture,” (H.-J. HU 2020). “Runzao Zhiyang Capsule” is used to treat skin pruritus caused by blood deficiency in the elderly in China owing to its unique clinical efficacy, lower recurrence rate, and fewer adverse reactions (Yanbo 2020). In recent years, its product sales have reached ~ 100 million Yuan. The product has also been rated as “National Key New Product,” “Guiyang City Famous Brand Product,” etc., meaning it

possesses a broad market and strong regional characteristics (Q. Sun et al. 2015). *L. bulbifera* has become one of the key ethnic medicinal materials cultivated and developed by the pillar industry of Guizhou ethnic medicine.

With the increased market demand for *L. bulbifera*, deterioration of its growth environment, and people's digging, its wild resources have been sharply reduced from year to year. Moreover, the morphological variations of *L. bulbifera* are complicated, and the species has homonyms and synonyms. These phenomena have led to the emergence of several fake products (*Urtica lobatifolia, Laportea cuspidata, Girardinia diversifolia, G. suborbiculata*, and closely related species) in the market, often mistakenly sold, purchased, and misused, which seriously affects the clinical medication safety of patients (Wei-na et al. 2013; Qing-wen et al. 2015; Q. Sun et al. 2015). Therefore, an accurate and effective identification method is urgently needed. Unfortunately, research on *L. bulbifera* is limited to resource distribution, few chemical components, and pharmacological effects. There is almost no molecular research and resources on the genome and molecular markers are in a state of extreme shortage, which is inconsistent with the widespread use of the plant in folk medicine and its market value.

The chloroplast (cp) is the most important semiautonomous organelle in plants. The cp genome contains rich genetic information and is considered an ideal model for studying gene evolution and phylogenetic relationships. The cp genome has a wide range of applications in the origin of medicinal plants, species identification, and cultivar breeding (Y. K. Kim et al. 2020; G. B. Kim et al. 2020; F. H. Wu et al. 2010). In recent years, the whole cp gene sequence has been used as a super-barcode to identify species (Krawczyk et al. 2018; Zhang et al. 2019), and the highly variable regions can also be used to identify species (Y. Hu et al. 2017; L. Wu et al. 2022). The National Center for Biotechnology Information (NCBI) has reported the complete cp genome sequence of some plants belonging to Urticaceae, which comprise 12 genera. However, the complete organelle genome resource of *Laportea*, which has excellent clinical efficacy, is scarce.

In this study the whole-genome sequence of *L. bulbifera* cp was obtained using second-generation sequencing (Illumina) and third-generation sequencing (Nanopore). This study revealed its structural characteristics and functions and analyzed the codon usage preference of the *L. bulbifera* cp genome. Combined with the cp whole-genome sequence of its closely related species, boundary, collinearity, whole-genome sequence comparison, and phylogenetic analyses were performed. The uniqueness of its structure was revealed, and highly differentiated regions were selected for phylogenetic analysis. The results provided valuable genetic information for the phylogeny, species identification, resource exploitation, and safe use of *L. bulbifera*.

2. Results

2.1. cp Genome Features of L. bulbifera

The *L. bulbifera* cp genome was 149,911 bp in length, with a typical quadripartite structure (Fig. 1), and contained a pair of IR regions (49842bp), a large single copy (LSC) region of 82,388 bp, and an SSC region of 17,681 bp. The total GC content in the cp genome was 36.80%. The GC content was 34.44% in the LSC region, 30.64% in the SSC region, and 42.86% in IRa and IRb regions (**Supplementary Table S1**).

A total of 127 genes were annotated in the cp genome, including 92 single-copy genes, and the remaining genes were duplicated in the reverse regions. The annotated genes included 83 protein-coding genes, 36 tRNA genes, and 8 rRNA genes. The IRa and IRb of the reverse repeat region included 7 protein-coding genes (*rpl2*, *rps7*, *rps12*, *ndhB*, *rpl23*, *ycf1*, and *ycf2*), 4 rRNA genes (*rrn4.5*, *rrn5*, *rrn16*, and *rrn23*), and 7 tRNA genes (*trnI-CAU*, *trnL-CAA*, *trnV-GAC*, *trnI-GAU*, *trnA-UGC*, *trnR-ACG*, and *trnN-GUU*). The SSC region included 1 tRNA gene (*trnL-UAG*) and 12 protein-coding genes (*ndhF*, *rpl32*, *ccsA*, *ndhD*, *psaC*, *ndhE*, *ndhG*, *ndhI*, *ndhA*, *ndhH*, *rps15*, and *ycf1*). One gene (*ycf1*) spanned the IRb and SSC regions. The remaining 21 tRNA genes and 60 protein-coding genes were located in the LSC region.

Twenty genes containing introns were present in the *L. bulbifera* cp genome. There were 8 tRNA genes (*trnK-UUU*, *trnG-UCC*, *trnL-UAA*, *trnV-UAC*, *trnI-GAU*, *trnA-UGC*, *trnA-UGC*, and *trnI-GAU*) and 12 protein-coding genes (*rps16*, *atpF*, *rpoC1*, *ycf3*, *clpP*, *petB*, *rpl16*, *rpl2*, *ndhB*, and *ndhA*), among which *ycf3* and *clpP* possessed two introns. These genes were divided into different categories according to their functions, such as photosynthetic system genes, RNA polymerase, ATP synthase, ribosomal protein, and NADH dehydrogenase (Table 1).

Table 1
Genes and their classification in cp genomes of *L.bulbifera*.

Gene group	gene_name
ribosomal proteins (SSU)	*rps2,rps14,rps11,rps4,rps8,rps7,rps3,rps16,rps18,rps12,rps15*
cytochrome b/f complex	*petA,petG,petB,petD,petN,petL*
photosystem II	*psbT,psbJ,psbH,psbM,psbC,psbI,psbF,psbA,psbN,psbB,psbE,psbZ,psbD,psbK*
photosystem I	*psaB,psaC,psaA,psaJ*
Protease	*clpP*
RNA polymerase	*rpoA,rpoC2,rpoB,rpoC1*
hypothetical chloroplast reading frames(ycf)	*ycf1,ycf4,ycf3,ycf2*
ribosomal proteins(LSU)	*rpl16,rpl23,rpl14,rpl32,rpl36,rpl33,rpl2,rpl22,rpl20*
other genes	*ccsA,accD,cemA*
Maturase	*matK*
ribosomal RNAs	*rrn5S, rrn4.5S,rrn16S,rrn23S*
RubisCO large subunit	*rbcL*
transfer RNAs	*trnI-CAU,trnfM-CAU,trnE-UUC,trnM-CAU,trnH-GUG,trnR-UCU,trnG-UCC,trnL-UAA,trnS-UGA,trnW-CCA,trnN-GUU,trnV-UAC,trnS-GCU,trnA-UGC,trnT-UGU,trnI-GAU,trnL-CAA,trnK-UUU,trnV-GAC,trnD-GUC,trnL-UAG,trnY-GUA,trnR-ACG,trnQ-UUG,trnP-UGG,trnT-GGU,trnF-GAA*
NADH dehydrogenase	*ndhI,ndhA,ndhG,ndhB,ndhH,ndhD,ndhE,ndhJ,ndhK,ndhC,ndhF*
ATP synthase	*atpI,atpA,atpF,atpH,atpE,atpB*

2.2. SSR Polymorphisms and Long Repeat Sequence Analysis

SSRs are molecular markers of high variability within the same species that have been used in population genetics and polymorphism studies. Single nucleotide repeat and dinucleotide repeat were the main SSR markers in the *L. bulbifera* cp genome. Single nucleotide, dinucleotide, trinucleotide, and tetranucleotide accounted for 73.68%, 21.58%, 2.11%, and 2.11% of the total SSR markers, respectively. There was only

one hexanucleotide repeat and no pentanucleotide repeat. A/T, AA/TT, AT/AT, AAA/TTT, AAT/ATT, and AAAA/TTTT accounted for most of the cp genome, which indicates that short poly-A and poly-T structures were abundant in the cp genome, whereas poly-C and poly-G structures were relatively few.

A total of 134 SSR markers were found in the *L. bulbifera* cp genome, among which 88 SSR markers were located in the LSC region, 30 SSR markers were located in the SSC region, and 16 SSR markers were evenly distributed in the IRa and IRb regions. Among the 134 SSR fragments, 67 were located in the intergenic spacer region (X. Su et al. 2020), 16 were located in the intron region, and the remaining were located in the protein-coding region, including *matK*, *rps16*, *atpF*, *rpoC2*, *rpoB*, *psbC*, *rbcL*, *ycf4*, *cemA*, *petA*, *ycf2*, *ycf1*, *ndhD* and *ndhF*. The *ycf1* region contained the largest number of simple repeat sequences, with a total of 15 SSR fragments (**Supplementary Table S4**).

The long repeats of the cp genome were annotated using Reputer and included four types: forward repeat (F), reverse repeat (R), complement repeat (C), and palindromic repeat (P). These long repeats may promote cp genome rearrangement and enhance genetic diversity in the population. There were 13 forward repeats, 24 palindromic repeats, 11 reverse repeats, and 2 complement repeats in the cp genome.

2.3. Codon Usage

Codon usage bias is a crucial evolutionary feature of the genome. The RSCU of coding protein sequences was obtained using CodonW and plotted according to the calculated results (**Supplementary Fig.S1**). There were 64 codons in the sequence, among which 61 codons encoded 20 amino acids and the other 3 codons were stop codons. The most frequently used codon was UUA (854 times) encoding leucine, and the RSCU was 1.97; the least frequently used codon was UAC (176 times) encoding tyrosine, and the RSCU was 0.37. There were 30 codons with RSCU > 1, most ending in A/U, and 32 codons with RSCU < 1, most ending in G/C (**Supplementary Table S5**). The codon encoding methionine and tryptophan exhibited an RSCU of 1, which eliminated bias because there is only one codon for these two proteins.

2.4. Comparative Genome Analysis

To explore whether there were boundary contraction and expansion events in *L. bulbifera*, IRscope was used to compare the cp genome boundaries of Urticaceae (Fig. 2). The length of the IR region of the cp genome of Urticaceae was between 24,579 and 41,818 bp, and the length of the IR region of *L. bulbifera* was 24,889 bp. Owing to the movement of the IRa/SSC boundary, the *ycf1* gene in the *L. bulbifera* cp genome spanned 7 bp to the SSC region and the *ndhF* gene spanned 3 bp to the IRa region. Interestingly, the *ndhF* gene of most plants in Urticaceae was located in the SSC region at the boundary of SSC/IRb, whereas the *ndhF* gene of *L. bulbifera* was located at the boundary of SSC/IRa. There were two asymmetric *ycf1* genes in the *L. bulbifera* cp genome. Most plants of Urticaceae had only one longer *ycf1* gene located at the boundary of SSC/IRa, or there were two *ycf1* genes. The longer one was located at the boundary of SSC/IRa, and the shorter one was located at the boundary of SSC/IRb. However, the situation of the *L. bulbifera* cp genome was the opposite. There was a *ycf1* gene of 1080 bp at the boundary of SSC/IRa and a *ycf1* gene of 5474 bp at the boundary of SSC/IRb. Therefore, the *L. bulbifera* cp genome exhibits some unique structural variations compared with other plants in Urticaceae.

To explore the existence of gene rearrangements in the *L. bulbifera* cp genome, Mauve plugin in Geneious, with *Nicotiana* as the reference species, was used to compare the cp genomes of *L. bulbifera* and its relative species (*P. lanceolatum*, *C. pachystachya*, *U. lobatifolia*, *U. angustifolia*, and *U. fissa*; Fig. 3). A high degree of gene rearrangement was very common in the *L. bulbifera* cp genome and its relative species. There was a large rearrangement region of ~ 39 kb in *rps12-trnV-GAC*, which spanned the entire SSC area, IRa area, and part of the IRb area of the *L. bulbifera* cp genome. Similarly, among the species closely related to *L. bulbifera*, the rearrangement region of the *trnA-UGC-ycf1* fragment length was ~ 27 kb in *P. lanceolatum* and the rearrangement region of the *ycf1-ndhF* fragment length was ~ 20 kb in *C. pachystachya*. Moreover, unlike *L. bulbifera*, the three *Urtica* species (*U. lobatifolia*, *U. angustifolia*, and *U. fissa*) did not exhibit large rearrangements in the latter half of their cp genome, and they were generally conserved.

To further understand the differences in the DNA sequences of *L. bulbifera* and its related species, the mVISTA program was used to analyze the DNA sequences (Fig. 4). Although the basic structure of the cp genome was relatively conserved, some variations existed. In the cp genomes of six plants, the coding region was more conservative than the noncoding region. Combined with the Pi values (Fig. 5), four coding regions (*ndhD*, *ccsA*, *ycf1*, and *matK*) and four noncoding regions (*ndhI-ndhG*, *ndhD-ccsA*, *matK-rps16*, and *psbM-trnD-GCU*) were selected as hypervariable regions for subsequent analysis.

2.5. Phylogenetic Analysis

Based on the whole cp genome sequence, the phylogenetic tree of 33 Urticaceae species was constructed, with *M. alba* and *C. sativa* as the closely related outgroup (Fig. 6). The plants included 11 *Pilea* species, 1 *Droguetia* species, 2 *Elatostema* species, 6 *Debregeasia* species, 3 *Urtica* species, 1 *Pipturus* species, 1 *Poikilospermum* species, 1 *Gonostegia* species, 1 *Cecropia* species, 1 *Procris* species, 2 *Boehmeria* species, 1 *Pouzolzia* species, and 1 *Oreocnide* species. The results of phylogenetic trees constructed using the NJ and ML methods were consistent (**Supplementary Fig.S2**). All support degrees of the branches in the phylogenetic trees were 100%, except one with 91%, which indicated that the species relationship of Urticaceae constructed using the cp genome sequence possessed high credibility.

Phylogenetic analysis showed that Urticaceae was a polyphyletic group, in which the *Droguetia*, *Debregeasia*, *Pipturus*, *Gonostegia*, *Boehmeria, Pouzolzia*, and *Oreocnide* genera were clustered into one branch, whereas the *Pilea*, *Elatostema*, *Urtica*, *Poikilospermum*, *Cecropia*, *Procris*, and *Laportea* genera were clustered into another branch. Among them, 2 *Boehmeria* species, 11 *Pilea* species and 6 *Debregeasia* species were well clustered into one branch, which was consistent with the results of previous studies (R. Wang et al. 2020). The most closely related species of *L. bulbifera* was *P. lanceolatum and C. pachystachya.*

2.6. Species Identification using Super-barcode and DNA Barcode

All branches of the phylogenetic tree showed high support when the analysis was based on the complete chloroplast genome sequence, which meant that the chloroplast genome sequence could be used as a super-barcode to accurately identify *L. bulbifera* from counterfeits and closely related species.

Then, ML trees were established for the eight selected hypervariable regions, and the classification results were different from those based on the complete genome. Among the eight hypervariable candidate regions, *ycf1* exhibited support rates of > 90% for all branches except for 67% and 69% in two branches of the genus of *Pilea* (**Supplementary Fig.S3**). Furthermore, the support rates of *matK* and *ndhD* reached > 70% in the branch where *L. bulbifera* was located (**Supplementary Fig.S4, S5**), meaning that *ycf1*, *matK*, and *ndhD* can be used to identify *L. bulbifera* from counterfeits and closely related species. Then 42 *matK* sequences of *L. bulbifera*, counterfeits, and closely related species were used to established an ML tree. The results showed that the *matK* sequence could well identify *L. bulbifera* from counterfeits and closely related species (Fig. 7).

3. Discussion

3.1. cp Genome Structure Analysis

The length, GC content, and number of annotated genes of the *L. bulbifera* cp genome were similar to those of other reported plants belonging to Urticaceae, such as *Debregeasia squamata* (R. Wang et al. 2020)d *fissa*. In the codon preference analysis, if RSCU > 1, the codon usage frequency is relatively high; if RSCU < 1, the codon usage frequency is relatively low. The cp codon of *L. bulbifera* showed a high preference for A/U, which is also very common in the cp genome of higher plants (Li et al. 2019; H. Liu et al. 2020). Therefore, the basic composition and structure of the *L. bulbifera* cp genome are relatively conserved.

Although the overall structure of the cp genome is relatively conserved, significant changes often occur during plant evolution because of gene loss, gene rearrangement, and IR region expansion and loss (Chumley et al. 2006; J. Liu et al. 2021; Y. Y. Guo et al. 2021). The same is true for *L. bulbifera*. Upon comparing *L. bulbifera* and its related species, it was found that the IR region was more conserved than the LSC and SSC regions and the coding region was more conserved than the noncoding region, which is consistent with most angiosperms (Y.-Y. Guo et al. 2021; C. Luo et al. 2021). Based on genome comparison analysis, there was an inverted SSC fragment in the *L. bulbifera* cp genome structure. The copies of the pseudogenes *ycf1* and *ndhF* were located at the boundary of SSC/IRa, and the copy of the functional *ycf1* was located at the boundary of SSC/IRb. This phenomenon is relatively rare in plants. A search revealed that similar cases have been reported in *Piper cenocladum* (Magnoliids) (Cai et al. 2006), *Artemisia frigida* (Asteraceae) (Y. Liu et al. 2013), *Dioscorea elephantipes* (Dioscoreaceae) (Hansen et al. 2007), *Helianthus annuus* (Asteraceae), *Lactuca sativa* (Asteraceae) (Timme et al. 2007), *Artemisia annua* (Asteraceae) (Shen et al. 2017), and *Firmiana colorata* (Malvaceae) (Shahzadi et al. 2019). Although the gene sequences of plants are relatively conserved, special structures of other cp genomes have been reported, including expansion of IR (Y.-Y. Guo et al. 2021), inversion of LSC (Kumar et al. 2009),

and re-inversion of SSC (Y. Liu et al. 2013). These phenomena enrich our understanding of plant genetic diversity.

The cp genome rearrangement frequency was high, and the rearrangement range was relatively large and mainly concentrated in the SSC and IR regions, which is not common in the cp genome of plants. Upon comparing the locations of different fragments, a great difference was found in the continuous region of *rps12-ycf1* between *L. bulbifera* and three *Urtica* species. These highly different regions coincided with the unique rearrangement segments of these species. Hence, it was speculated that the rearrangement resulted in large structural differences in the corresponding regions of the cp genome. The genome rearrangement of large fragments might also be the reason for the inversion of the SSC fragments in the cp gene sequence compared with other Urticaceae species in the boundary analysis. *L. bulbifera* grow in special karst landform areas (the soil in this area is characterized by drought, high calcium, and rocky desertification) (C. Liu et al. 2021) and are known for their strong adaptability and great morphological variations. By adapting to the harsh environment, *L. bulbifera* might have gradually developed a special variation in its genome structure.

3.2. Identification and Phylogenetic Analysis

For the protection and rational exploitation and utilization of plant resources, it is very important to clarify the systematic classification and genetic relationship among plant families and genera. Urticaceae exhibits a high diversity in habits, life forms, and reproductive characteristics, and it is difficult to collect some species. Hence, the traditional morphological classification of Urticaceae has been ambiguous and controversial, and the taxonomic status of some genera has also been difficult to establish (Deng et al. 2013; Hadiah and Conn 2009). In recent years, the classification of Urticaceae has been studied using some gene fragments in the nucleus, cp, and mitochondrial genome, such as *ITS*, *18S*, *matR* (Z. Y. Wu et al. 2013), *rbcL*, *trnL-F* (Julisasi T. Hadiah A and B 2008), *matK*, *rpll4-rps8-infA-rpl36* (Z.-Y. Wu et al. 2013), and *nrITS* (Sun et al. 2015). However, few studies have used the cp whole-genome sequence to study the phylogenetic relationship of this family (R. Wang et al. 2020). The complete cp genome sequence of plants has been widely used to explore the phylogenetic relationship of species (Y. K. Kim et al. 2020; Zhao et al. 2021), and the results are more accurate and realistic than those obtained from DNA fragments alone. In this study, the phylogenetic trees were constructed based on the cp whole-genome sequence, and the phylogenetic relationship between the genera was similar to their predecessors (R. Wang et al. 2020). Because this is the first complete cp data resource of *Laportea*, this result directly established the taxonomic status of *Laportea* in Urticaceae. According to the available cp genome data of this family provided by the NCBI, the species most closely related to *L. bulbifera* is *P. lanceolatum* (*Poikilospermum*). *Poikilospermum* is controversial because its morphological characteristics are intermediate to those of Moraceae and Urticaceae (Kravtsova et al. 2020). This study provided molecular evidence that the genus belongs to Urticaceae.

Identifying the authenticity of traditional Chinese medicine is an important checkpoint in ensuring the quality of medicinal materials and a key link in ensuring the safety of clinical medications. In the case of *L. bulbifera*, there are various counterfeit products in the market, such as *U. fissa* and *U. lobatifolia* (Q.

Sun et al. 2015). DNA barcoding or super-barcoding could be used to distinguish these species (XW et al. 2015; S. Chen et al. 2010). The cp whole-genome sequences can be used as super-barcodes, and highly differentiated regions identified with comparative genomics can be used for DNA barcoding (S. Chen et al. 2010; Q. Chen et al. 2020). This study constructed phylogenetic trees based on the complete cp genome and selected eight highly variable regions. The complete cp genome exhibited high resolution and accuracy in both species identification and phylogenetic analysis and could be used as super-barcode for the species identification of *L. bulbifera*.

Among the highly variable regions, *ycf1* is the segment that has the most potential to become a DNA barcode for most genera of the family, and the support rates of *matK* and *ndhD* were > 70% in the branch where *L. bulbifera* was located. Thus, these three coding sequence fragments had the potential to identify *L. bulbifera*. However, in the phylogenetic tree constructed based on these three fragments, the accuracy of the phylogenetic relationship of the overall classification results was not high because our results suggest that plants of the same genus do not cluster well in phylogenetic trees based on DNA fragments. For example, *D. hekouensis* was not clustered with other plants from the same genus. It is worth noting that the *ycf1* fragment is also the gene fragment with the highest SSR enrichment, which confirms the previous view that these SSR-rich regions may be potential regions of mutation hotspots (George et al. 2015). Additionally, we collected another eight *matK* sequences from the *Laportea* and *Girardinia* genus for analysis, and *L. bulbifera* could be identified from all counterfeits and closely related species based on the barcode. However, we did not obtain more sequence information of *ycf1* and *ndhD* because it was difficult to collect plants from that genus. Based on existing information, *ycf1* and *ndhD* are able to distinguish some counterfeits and closely related species, but more data are required for further analysis. Therefore, the whole cp genome sequence of plants can be used not only as the super-barcode for species identification but also as the basis for determining the classification status in the phylogenetic analysis of *L. bulbifera*. The fragments of *matK*, *ycf1*, and *ndhD* can be used as DNA barcodes for the species identification of *L. bulbifera*, which can save time and cost, although it is not a wise choice for phylogenetic analysis.

4. Materials And Methods

4.1 Plant Material and DNA Extraction

L. bulbifera samples were collected on July 27, 2021, from Leigong Mountain, Qiandongnan Prefecture, Guizhou Province. Healthy and fresh leaves of the plants were separated and wrapped in tin foil, placed in liquid nitrogen, and stored at − 80°C for future use. The DNA of the leaves was extracted using the cetyltrimethylammonium bromide method (CTAB) (Porebski et al. 1997). Agarose gel electrophoresis was used to detect the quality of the DNA samples. The next sequencing step was performed only if the quality of the DNA was established.

4.2 DNA Sequencing, Assembly, and Annotation

The DNA samples were randomly interrupted with a Covaris ultrasonic analyzer, and the whole genome library was prepared using terminal repair, A-tail addition, sequencing connector addition, purification, polymerase chain reaction (PCR) amplification, and other steps. After the library construction was completed, Qubit 2.0 was used for the preliminary quantification and dilution, and Agilent 2100 was used to detect the inserted fragments of the library. After confirming that the inserted fragments met the expected size, quantitative PCR was used to accurately estimate the effective concentration of the library to ensure its quality. After the quality of the library was ascertained, different libraries were pooled into Flow cells according to the requirements of effective concentration and target data volume. Subsequent to cBOT clustering, sequencing was performed using the Illumina NovaSeq high-throughput sequencing platform.

Megaruptor was used to randomly interrupt the sample DNA, and magnetic beads were used to enrich and purify large DNA fragments and repair damaged DNA fragments. After purification, both ends of the DNA fragment were repaired and A was added. The SQK-LSK109 connection kit was utilized for connection response. Finally, Qubit was employed to accurately detect the established DNA library. After completing the construction of the library, a DNA library of a certain concentration and volume was added to the Flow cell, which was transferred to the Oxford Nanopore PromethION sequencer for real-time single-molecule sequencing.

Raw sequencing data of second and third generations were filtered according to their respective criteria to obtain clean reads for subsequent analysis. Flye (version 2.8.3) (Kolmogorov et al. 2020) was used for genome splicing. Blastn (version: 2.2.30+; parameter: −evalue 1e-5) alignment was performed between the splicing results and *Poikilospermum lanceolatum* cp genome sequences (NC_056983.1), and the assembly results of candidate sequences were determined based on the alignment. Based on sequencing depth, read alignment, and comparison with related species, the cp genome linkage relationship was determined. Because the well-connected sequence contained gaps (including the N sequence), Gapcloser (version 1.12) was used to further fill the gaps to obtain the final splicing result. After obtaining the complete sequence, gene annotation was performed using the annotation tool CPGAVAS2 (http://47.96.249.172:16019/analyzer/annotate) (Shi et al. 2019) specially designed for cp. The circular cp genome map was generated using the OGdraw online tool (version1.3.1)(https://chlorobox.mpimp-golm.mpg.de/OGDraw) (Marc et al.). The assembled complete cp genome sequences of *L. bulbifera* was submitted to NCBI with the accession numbers ON320386.

4.3 Genome Structure Analysis

CodonW (Sharp and Li 1987) was used to calculate the relative synonymous codon usage (RSCU) of *L. bulbifera*. MISA (version 1.0) (Beier et al. 2017) was used for simple sequence repeat (SSR) detection of the cp gene. The parameters were set as follows: the minimum number of repeats of each corresponding unit size was 1−8, 2−4, 3−4, 4 − 3, 5 − 3, and 6 − 3. REPuter online tool (https://bibiserv.cebitec.uni-bielefeld.de/reputer) (Kurtz et al. 2001) were used to identify the interspersed long repeat fragments in the cp genome. The MISA output was processed using a Perl script to determine the specific location of SSR in the *L. bulbifera* cp genome.

4.4 Genome Comparison

The IRscope online tool (https://irscope.shinyapps.io/irapp/) (Amiryousefi et al. 2018) was used to compare the differences in the boundaries of the four regions of the cp genome of Urticaceae. The mauve plugin was installed in Geneious (version 9.0.2) (Kearse et al. 2012) to perform a collinearity analysis of the cp genomes of *L. bulbifera*, *P. lanceolatum*, *Cecropia pachystachya*, *U. fissa*, *U. lobatifolia*, and *U. angustifolia* by selecting the complete *Nicotiana tabacum* cp genome as the reference gene (Kyalo et al. 2018). The mVISTA online tool (Frazer et al. 2004) in the LAGAN mode was used to compare the whole cp genomes of *L. bulbifera*, *P. lanceolatum*, *C. pachystachya*, *U. fissa*, *U. lobatifolia*, and *U. angustifolia*, and the complete *P. lanceolatum* cp genome was selected as the reference gene. The nucleotide polymorphism (Pi) values were calculated using DnaSP (version 6.12.03) (Rozas et al. 2017).

4.5 Phylogenetic Analysis

The complete cp genome sequences of 32 Urticaceae species were downloaded from the NCBI website (Table 2). *Morus alba* and *Cannabis sativa* was selected as the closely related outgroup taxa (Tao 2015). MAFFT (Katoh and Standley 2013) was used to compare and clip the whole-genome sequences of 35 plants. Subsequently, MEGA7.0 (Tamura et al. 2007) was employed to construct a phylogenetic tree using the neighbor-joining (NJ) algorithm method. IQtree (Nguyen et al. 2014) was employed to construct a phylogenetic tree using the maximum likelihood (ML) method, and GTR + F + R3 was selected as the best model. Finally, FigTree (version 1.4.4.0) (Schneider et al. 2000) was utilized to beautify the phylogenetic tree.

Table 2
Cp genome information used for analysis

Species	Accession number	Length (bp)
Pilea peploides	NC_056133.1	150924
Pilea cavaleriei	NC_057226.1	151557
Droguetia iners	MN189960.1	149414
Elatostema laevissimum	MN189961.1	150244
Debregeasia hekouensis	MZ159965.1	155941
Pilea glauca	NC_054338.1	151210
Pilea peperomioides	NC_054339.1	152327
Pilea cadierei	NC_054343.1	152245
Urtica lobatifolia	NC_056314.1	146848
Debregeasia squamata	MN189959.1	156065
Debregeasia saeneb	MN189958.1	155790
Pipturus arborescens	MN189967.1	154069
Poikilospermum lanceolatum	NC_056983.1	153454
Boehmeria tomentosa	MN189945.1	154938
Pilea monilifera	NC_057227.1	150938
Gonostegia hirta	NC_053931.1	159086
Elatostema dissectum	NC_047192.1	150302
Cecropia pachystachya	NC_039763.1	153925
Pilea plataniflora	NC_056134.1	152143
Debregeasia orientalis	NC_041413.1	160283
Procris crenata	NC_056276.1	154124
Pilea pumila	NC_054345.1	152036
Debregeasia elliptica	MZ326656.1	155901
Urtica angustifolia	MZ145046.1	146679
Pilea microphylla	NC_054344.1	150019
Urtica fissa	MZ313540.1	146837
Pouzolzia sanguinea	MN189968.1	153715

Species	Accession number	Length (bp)
Pilea verrucosa	NC_054346.1	150969
Debregeasia longifolia	MN189952.1	155810
Pilea mollis	NC_054341.1	150587
Boehmeria nivea	MW057777.1	156056
Oreocnide frutescens	MN189965.1	156966
Morus alba	KU355276.1	159113
Cannabis sativa	NC_026562.1	153871

4.6 Species Identification using DNA Barcode

According to the phylogenetic tree, the cp genome sequence was evaluated for whether it could be used as a super-barcode to accurately identify *L. bulbifera* from counterfeits and closely related species. Combining the results of mVISTA and Pi values, eight potential DNA barcode fragments were screened for subsequent analysis. The ML phylogenetic trees of the hypervariable regions were constructed based on TVM + F + I + G4 (*ycf1*), TVM + F + G4 (*ccsA*), TVM + F + G4 (*matK*), TVM + F + G4 (*nahD*), TVM + F + R2 (*matK-rps16*), K3Pu + F + G4 (*ndhD-ccsA*), K3Pu + F + G4 (*ndhI-ndhG*), and TVM + F + G4 (*psbM-trnD-GUC*) models. Then, we collected another eight *matK* sequences of counterfeits and closely related species from the *Laportea* and *Girardinia* (Table 3) and established an ML tree for species identification based on the TVM + F + G4 models.

Table 3
Additional matK genome information used for analysis

Species	Family	accession number
Laportea cuspidata	*Laportea*	MH660267.1
Laportea canadensis	*Laportea*	AY257537.1
Laportea aestuans	*Laportea*	MK931216.1
Laportea lanceolata	*Laportea*	MK931214.1
Laportea interrupta	*Laportea*	MK931213.1
Laportea alatipes	*Laportea*	MK931210.1
Girardinia diversifolia	*Girardinia*	MH659810.1
Girardinia suborbiculata	*Girardinia*	MK931207.1

5. Conclusion

In this study, the *L. bulbifera* cp genome was sequenced using second-generation and third-generation sequencing technologies for the first time. There was an inverted SSC structure in the *L. bulbifera* cp genome, and a large fragment rearrangement occurred in this region. The ML tree results showed that the cp genome sequence could serve as a reliable basis for phylogenetic analysis and species identification and that the hypervariable regions *matK*, *ycf1*, and *ndhD* could be used as potential specific barcodes for the identification of *L. bulbifera*. This study elucidated the cp gene structure of the species, identified potential DNA barcode regions, and enriched the understanding of its genetic diversity. These findings could lay a strong foundation for medicinal quality control, molecular marker development, and phylogenetic analysis of *L. bulbifera*.

Declarations

Author Contributions: LX acquired the funding and conceived the research. WW and LX collected plant material. WW analysed the data. WW and XW contributed equally to this manuscript. LX revised the manuscript. LW, YS, QY, RG and MW provides guidance in the analysis and manuscript writing process. WW edited and approved the final manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China and Karst Science Research Center of Guizhou Province (U1812403-1), National Key R&D Program of China from the Ministry of Science and Technology of China (grant No. 2019YFC1711100), the CACMS Innovation Fund (CI2021A05103). National Natural Science Foundation of China (81891010, 81891013).

Data Availability Statement: The original contributions presented in the study are publicly available. This data can be found here: The chloroplast genomes assembled in this study were deposited in GenBank with the accession numbers ON320386.

Acknowledgments: I would like to thank my supervisor, LX, for her guidance during this research. I would also like to thank my tutors, LW, YS, QY, RG and MW, for their valuable guidance throughout my studies. I would particularly like to acknowledge my teammate, CL, GD, and TW for their wonderful collaboration and patient support.

Conflict of Interest: We declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Amiryousefi A, Hyvönen J, Poczai P (2018) IRscope: an online program to visualize the junction sites of chloroplast genomes. Bioinformatics 34(17):3030–3031. doi:10.1093/bioinformatics/bty220
2. Beier S, Thiel T, Münch T, Scholz U, Mascher M (2017) MISA-web: a web server for microsatellite prediction. Bioinformatics 33(16):2583–2585. doi:10.1093/bioinformatics/btx198

3. Cai Z, Penaflor C, Kuehl JV, Leebens-Mack J, Carlson JE, DePamphilis CW et al (2006) Complete plastid genome sequences of Drimys, Liriodendron, and Piper: implications for the phylogenetic relationships of magnoliids. BMC Evol Biol 6(1):1–20
4. Chen Q, Wu X, Zhang D (2020) Comparison of the abilities of universal, super, and specific DNA barcodes to discriminate among the original species of Fritillariae cirrhosae bulbus and its adulterants. PLoS ONE 15(2):e0229181. doi:10.1371/journal.pone.0229181
5. Chen S, Hui Y, Han J, Liu C, Song J, Shi L et al (2010) Validation of the ITS2 Region as a Novel DNA Barcode for Identifying Medicinal Plant Species.Plos One, 5(1), e8613
6. Chen Y-r, Xu W-f, Sun Q-w (2019) Research progress of Laportea bulbifera.Chinese Journal of Experimental Traditional Medical Formulae,214–220
7. Chen Y, Zou S, Xu W, Sun Q, Yun L (2020) Spectrum–effect relationship of antioxidant and anti-inflammatory activities of Laportea bulbifera based on multivariate statistical analysis.Biomedical Chromatography, *34*(2)
8. Chumley TW, Palmer JD, Mower JP, Fourcade HM, Calie PJ, Boore JL et al (2006) The Complete Chloroplast Genome Sequence of Pelargonium × hortorum: Organization and Evolution of the Largest and Most Highly Rearranged Chloroplast Genome of Land Plants. Mol Biol Evol 23(11):2175–2190. doi:10.1093/molbev/msl089
9. Dan W, Juan T, Ying L, Jing L, Si-ying C, Zi-peng G et al (2019) Simultaneous determination of 11 constituents in Laportea bulbifera of Miao medicine by UPLC-ESI-MS. Chin J Pharm Anal 39(08):1425–1432 (in Chinese)
10. Deng T, Kim C, Zhang DG, Zhang JW, Li ZM, Nie ZL et al (2013) Zhengyia shennongensis: A new bulbiliferous genus and species of the nettle family (Urticaceae) from central China exhibiting parallel evolution of the bulbil trait. *Taxon*
11. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I (2004) VISTA: computational tools for comparative genomics. Nucleic Acids Res 32(suppl2):W273–W279. doi:10.1093/nar/gkh458
12. George B, Bhatt BS, Awasthi M, George B, Singh AK (2015) Comparative analysis of microsatellites in chloroplast genomes of lower and higher plants. Curr Genet 61(4):665–677. doi:10.1007/s00294-015-0495-9
13. Guo Y-Y, Yang J-X, Bai M-Z, Zhang G-Q, Liu Z-J (2021) The chloroplast genome evolution of Venus slipper (Paphiopedilum): IR expansion, SSC contraction, and highly rearranged SSC regions. BMC Plant Biol 21(1):1–14
14. Guo YY, Yang JX, Bai MZ, Zhang GQ, Liu ZJ (2021) The chloroplast genome evolution of Venus slipper (Paphiopedilum): IR expansion, SSC contraction, and highly rearranged SSC regions. *BMC Plant Biology*
15. H-Y H (2018) Preliminary study on the basis and metabolism of anti - inflammatory substance of Laportea bulbifera. Beijing University Of Chinese Medicine
16. Hadiah JT, Conn BJ (2009) Usefulness of morphological characters for infrageneric classification of Elatostema (Urticaceae). Blumea J Plant Taxonomy Plant Geogr 54(1):181–191

17. Han H, Suo Y, Liu X, Wu Y, Dai Y, Ni Y et al (2018) Screening of active components of urtica dentata hand by RAW264. 7 anti-inflammatory cell model and chemical constituents. Global Tradit Chin Med 11:651–655
18. Hansen DR, Dastidar SG, Cai Z, Penaflor C, Kuehl JV, Boore JL et al (2007) Phylogenetic and evolutionary implications of complete chloroplast genome sequences of four early-diverging angiosperms: Buxus (Buxaceae), Chloranthus (Chloranthaceae), Dioscorea (Dioscoreaceae), and Illicium (Schisandraceae). Mol Phylogenet Evol 45(2):547–563
19. Hou W-R, Su Z-Q, Pi H-F, Yao G-M, Zhang P, Luo X et al (2010) Immunosuppressive constituents from Urtica dentata Hand. J Asian Nat Prod Res 12(8):707–713
20. HU H-J (2020) Study on quality control of Miao medicine Laportea bulbifera based on fingerprint analysis and quantitative analysis of multi-components.Chinese Traditional and Herbal Drugs,4325–4330
21. Hu Y, Woeste KE, Zhao P (2017) Completion of the chloroplast genomes of five Chinese Juglans and their contribution to chloroplast phylogeny. *Frontiers in Plant Science, 7*, 1955
22. Julisasi T, Hadiah A, Barry BC, Conn J, B CJQ (2008) Infra-familial phylogeny of Urticaceae, using chloroplast sequence data. Aust Syst Bot 21(5):375–385
23. Katoh K, Standley DM (2013) MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol 30(4):772–780. doi:10.1093/molbev/mst010
24. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S et al (2012) Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28(12):1647–1649. doi:10.1093/bioinformatics/bts199
25. Kim GB, Lim CE, Kim JS, Kim K, Mun JH (2020) Comparative chloroplast genome analysis of Artemisia (Asteraceae) in East Asia: Insights into evolutionary divergence and phylogenomic implications.BMC Genomics, *21*(1)
26. Kim YK, Jo S, Cheon SH, Joo MJ, Kim KJ (2020) Plastome Evolution and Phylogeny of Orchidaceae, With 24 New Sequences.Frontiers in Plant Science, *11*
27. Kolmogorov M, Bickhart DM, Behsaz B, Gurevich A, Rayko M, Shin SB et al (2020) metaFlye: scalable long-read metagenome assembly using repeat graphs. Nat Methods 17(11):1103–1110. doi:10.1038/s41592-020-00971-x
28. Kravtsova T, Moore A, Wilmot-Dear C, Zhinkina N (2020) Comparative carpological study of Poikilospermum (Urticaceae) in relation to taxonomy. Kew Bull 75(1):1–34
29. Krawczyk K, Nobis M, Myszczyński K, Klichowska E, Sawicki J (2018) Plastid super-barcodes as a tool for species discrimination in feather grasses (Poaceae: Stipa). Sci Rep 8(1):1–10
30. Kumar S, Hahn FM, McMahan CM, Cornish K, Whalen MC (2009) Comparative analysis of the complete sequence of the plastid genome of Parthenium argentatum and identification of DNA barcodes to differentiate Parthenium species and lines. BMC Plant Biol 9(1):1–12

31. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R (2001) REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res 29(22):4633–4642. doi:10.1093/nar/29.22.4633
32. Kyalo CM, Gichira AW, Li ZZ, Saina JK, Itambo M, Hu GW et al (2018) Characterization and Comparative Analysis of the Complete Chloroplast Genome of the Critically Endangered Species Streptocarpus teitensis (Gesneriaceae). *BioMed Research International,2018,(2018-3-25), 2018*, 1–11
33. Li G, Pan Z, Gao S, He Y, Xia Q, Jin Y et al (2019) Analysis of synonymous codon usage of chloroplast genome in Porphyra umbilicalis. Genes & Genomics 41(10):1173–1181
34. Lin M, Bing l, Zhu Z, Haiyan Z, Shufang L, Xiaosheng Y (2012) Study on analgesic pharmacological action of alcohol extract of Ethnomedicine Laportea bulbifera. J Guiyang Univ Chin Med 34(01):24–26 (in Chinese)
35. Liu C, Huang Y, Wu F, Liu W, Liang Y (2021) Plant adaptability in karst regions. J Plant Res 134(5):1–18
36. Liu C, Huang Y, Wu F, Liu W, Ning Y, Huang Z et al (2021) Plant adaptability in karst regions. J Plant Res 134(5):889–906
37. Liu H, Lu Y, Lan B, Xu J (2020) Codon usage by chloroplast gene is bias in Hemiptelea davidii. J Genet 99(1):1–11
38. Liu J, Jiang M, Chen H, Liu Y, Wu W (2021) Comparative genome analysis revealed gene inversions, boundary expansion and contraction, and gene loss in Stemona sessilifolia (Miq.) Miq. chloroplast genome.Plos One
39. Liu Y, Huo N, Dong L, Wang Y, Zhang S, Young HA et al (2013) Complete chloroplast genome sequences of Mongolia medicine Artemisia frigida and phylogenetic relationships with other plants.Plos One, 8(2), e57533
40. Lu J-L, Li W-J, Hou W-R, Lan Y, Zhou H, Yin L-J et al (2012) Study on effect of total coumarins from Urtica dentata on dextran sulfate sodium-induced colitis in mice. Zhongguo Zhong yao za zhi = Zhongguo Zhongyao Zazhi = China Journal of Chinese Materia Medica 37(21):3316–3320
41. Luo C, Huang W, Sun H, Yer H, Li X, Li Y et al (2021) Comparative chloroplast genome analysis of Impatiens species (Balsaminaceae) in the karst area of China: insights into genome evolution and phylogenomic implications. BMC Genomics 22(1):1–18
42. Luo X, Li LL, Zhang SS, Lu JL, Ying Z, Zhang HY et al (2011) Therapeutic effects of total coumarins from Urtica dentata Hand on collagen-induced arthritis in Balb/c mice. J Ethnopharmacol 138(2):523–529
43. Marc L, Oliver D, Sabine K, Ralph B OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets.Nucleic Acids Research(W1),W575
44. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ (2014) IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol Biol Evol 32(1):268–274. doi:10.1093/molbev/msu300

45. Porebski S, Bailey LG, Baum BR (1997) Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. Plant Mol biology Report 15(1):8–15
46. Qing-wen S, Wen-fen X, Wei-na Q, Shun-zhi H, Bo W, Sheng-hua W (2015) Morphological Variation of Germplasm ï¼²esources of Miao Medicine Hong Hema. Seed 34(04):59–63
47. QingXin S (2018) Study on hypidemic mechanism of Honghema and its pharmacodynamic substances. Hubei University Of Traditional Chinese Medicine
48. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE et al (2017) DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol Biol Evol 34(12):3299–3302
49. Schneider S, Roessli D, Excoffier L (2000) Arlequin: a software for population genetics data analysis. User Man ver 2:2496–2497
50. Shahzadi I, Mehmood F, Ali Z, Malik MS, Waseem S, Mirza B et al (2019) Comparative analyses of chloroplast genomes among three Firmiana species: Identification of mutational hotspots and phylogenetic relationship with other species of Malvaceae. Plant Gene 19:100199
51. Sharp PM, Li W-H (1987) The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res 15(3):1281–1295
52. Shen X, Wu M, Liao B, Liu Z, Bai R, Xiao S et al (2017) Complete chloroplast genome sequence and phylogenetic analysis of the medicinal plant Artemisia annua. Molecules 22(8):1330
53. Shi L, Chen H, Jiang M, Wang L, Wu X, Huang L et al (2019) CPGAVAS2, an integrated plastome sequence annotator and analyzer. Nucleic Acids Res 47(W1):W65–W73. doi:10.1093/nar/gkz345
54. Su X, Zhao J, Wang Z (2020) The complete chloroplast genome of Hordeum brevisubulatum. Mitochondrial DNA Part B 5(3):3006–3007
55. Su Z, Zhao Z, Xie S, Hou W-r, Tao E, Xiang M (2009) Effects of analgesia, anti-inflammation and immunosuppression of acetic ether extract of Chinese medicine honghuoma. Chin Pharmacol Bull (Chinese) 25(4):559–560
56. Sun H, Deng, Tao, Chase M et al (2015) Generic phylogeny and character evolution in Urticeae (Urticaceae) inferred from nuclear and plastid DNA regions. *Taxon*
57. Sun Q, Xu W, Qi W, Bai C, Wei S (2015) Study on the crude drug identification of honghema and Laportea Gaudich. J Chin Med Mater 38(09):1862–1867 (in Chinese)
58. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. Mol Biol Evol 24(8):1596–1599. doi:10.1093/molbev/msm092
59. Tang J, Wu D, Chen S, Li Y, Li J, Gong Z et al (2018) Identification of chemical compositions in Laportea bulbifera by UPLC-ESI-Q-TOF-MS. Chin J Exp Tradit Med Formul 24:67–72
60. Tao D (2015) Origin and diversification of important character taxa in the East Asian flora. Yunnan University
61. Timme RE, Kuehl JV, Boore JL, Jansen RK (2007) A comparative analysis of the Lactuca and Helianthus (Asteraceae) plastid genomes: identification of divergent regions and categorization of

shared repeats. Am J Bot 94(3):302–312

62. Wang J, Lu J, Lan Y, Zhou H, Li W, Xiang M (2013) Total coumarins from Urtica dentata Hand prevent murine autoimmune diabetes via suppression of the TLR4-signaling pathways. J Ethnopharmacol 146(1):379–392
63. Wang R, Milne RI, Du XY, Liu J, Wu ZY (2020) Characteristics and Mutational Hotspots of Plastomes in Debregeasia (Urticaceae).Frontiers in Genetics, *11*
64. Wei-na Q, Wen-fen X, Qing-wen S, Shun-zhi H, Sheng-hua W (2013) Study on Laportea Plants. Guizhou Sci 31(05):61–64 (in Chinese)
65. Wu FH, Chan MT, Liao DC, Hsu CT, Lee YW, Daniell H et al (2010) Complete chloroplast genome of Oncidium Gower Ramsey and evaluation of molecular markers for identification and breeding in Oncidiinae. BMC Plant Biol 10(1):68
66. Wu L, Nie L, Guo S, Wang Q, Wu Z, Lin Y et al (2022) Identification of Medicinal Bidens Plants for Quality Control Based on Organelle Genomes. *Frontiers in pharmacology, 13*
67. Wu Z-Y, Monro AK, Milne RI, Wang H, Yi T-S, Liu J et al (2013) Molecular phylogeny of the nettle family (Urticaceae) inferred from multiple loci of three genomes and extensive generic sampling. Mol Phylogenet Evol 69(3):814–827. doi:https://doi.org/10.1016/j.ympev.2013.06.022
68. Wu ZY, Monro AK, Milne RI, Wang H, Li DZ (2013) Molecular phylogeny of the nettle family (Urticaceae) inferred from multiple loci of three genomes and extensive generic sampling. Mol Phylogenet Evol 69:814–827
69. Xiang L (2010) Anti-inflmatory effect and Immune-tolerance mechanism of Hongmasu. Huazhong University of Science and Technology
70. Xiang M, Jingli LU, Zhang C, Lan Y, Zhou H, Xiping LI et al (2013) Identification and quantification of total coumarins from Urtica dentata Hand and its roles in promoting immune tolerance via TLR4-mediated dendritic cell immaturation. J Agricultural Chem Soc Japan 77(6):1200–1206
71. Xiang M, Tao E, Ghu T (2006) Preliminary study on peventing effects of the active component of Honghuoma on CIA model. Chin J Hosp Pharm 26(10):1201
72. Yang XW, Henry RJ, Rossetto W et al (2015) Plant DNA barcoding: from gene to genome. *BIOL REV, 2015,90(1)*(-), 157–166
73. Yanbo C (2020) Observation on the effect of ibastin tablets, runzao antipruritus capsule and calamine lotion in the treatment of senile pruritus. Contemp Med 26:171–172 (in Chinese)
74. Yao Y (2008) Study on the therapeutic effect and mechanism of effective parts of honghuoma on rheumatoid arthritis. Huazhong University of Science and Technology
75. Zengyu Z (2009) Study on preparation of total flavonoids of honghuoma and its effect on T2DM-IR. Huazhong University of Science and Technology
76. Zhang Z, Zhang Y, Song M, Guan Y, Ma X (2019) Species identification of Dracaena using the complete chloroplast genome as a super-barcode.Frontiers in pharmacology,1441

77. Zhao F, Chen Y, Salmaki Y, Drew BT, Xiang CL (2021) An updated tribal classification of Lamiaceae based on plastome phylogenomics.BMC Biology, *19*

78. Zhu Z, Ma L, Zhu H, Yang X, Hao X (2011) Studies on the chemical constituents of Laportea bulbifera. Zhong yao cai = Zhongyaocai = Journal of Chinese Medicinal Materials 34(2):223–225

Figures

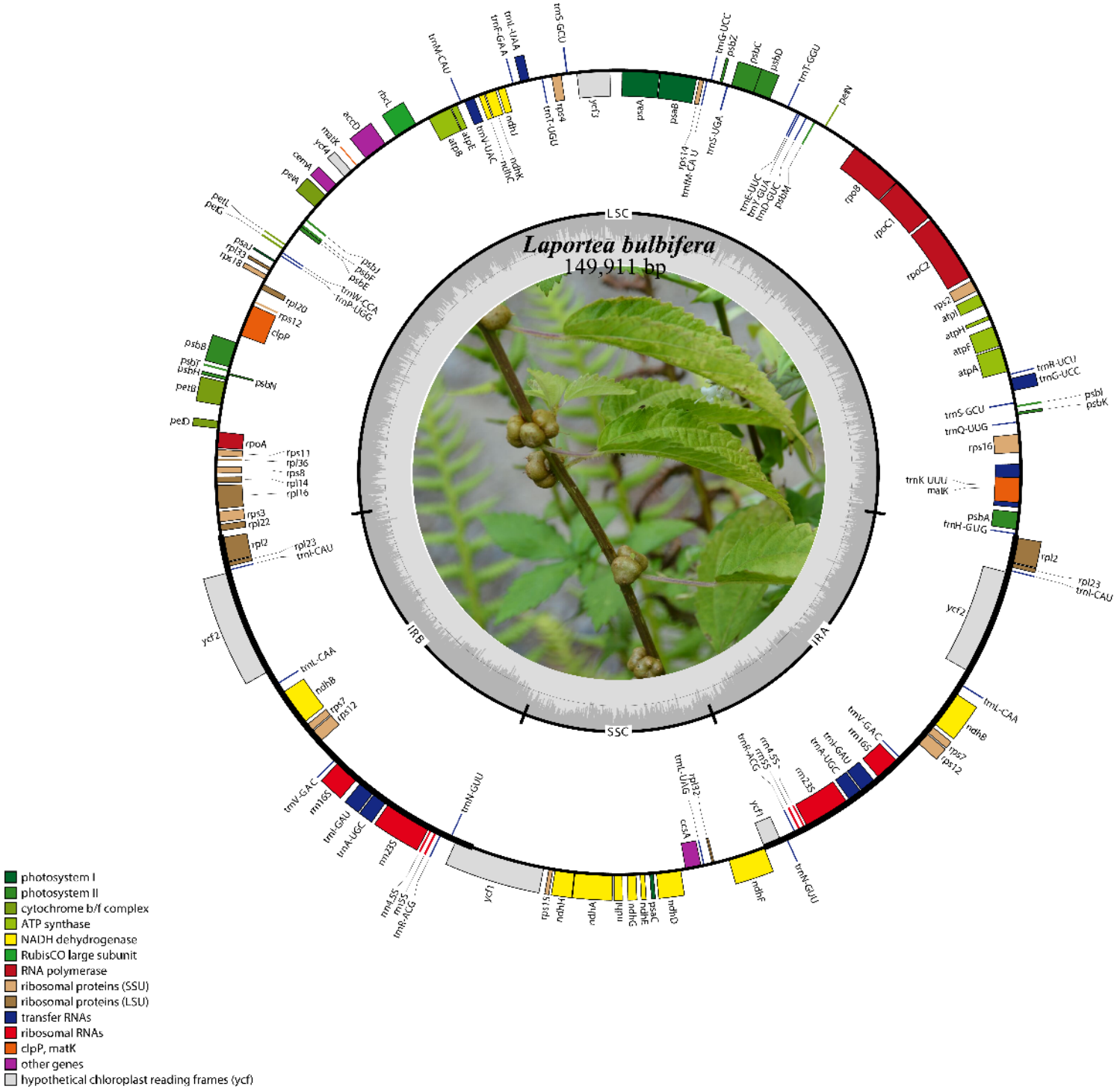

Figure 1

Gene maps of the cp genomes of *L. bulbifera*. Genes inside and outside the circle were transcribed clockwise and counterclockwise, respectively. Genes with different functions are characterized with different color bars.

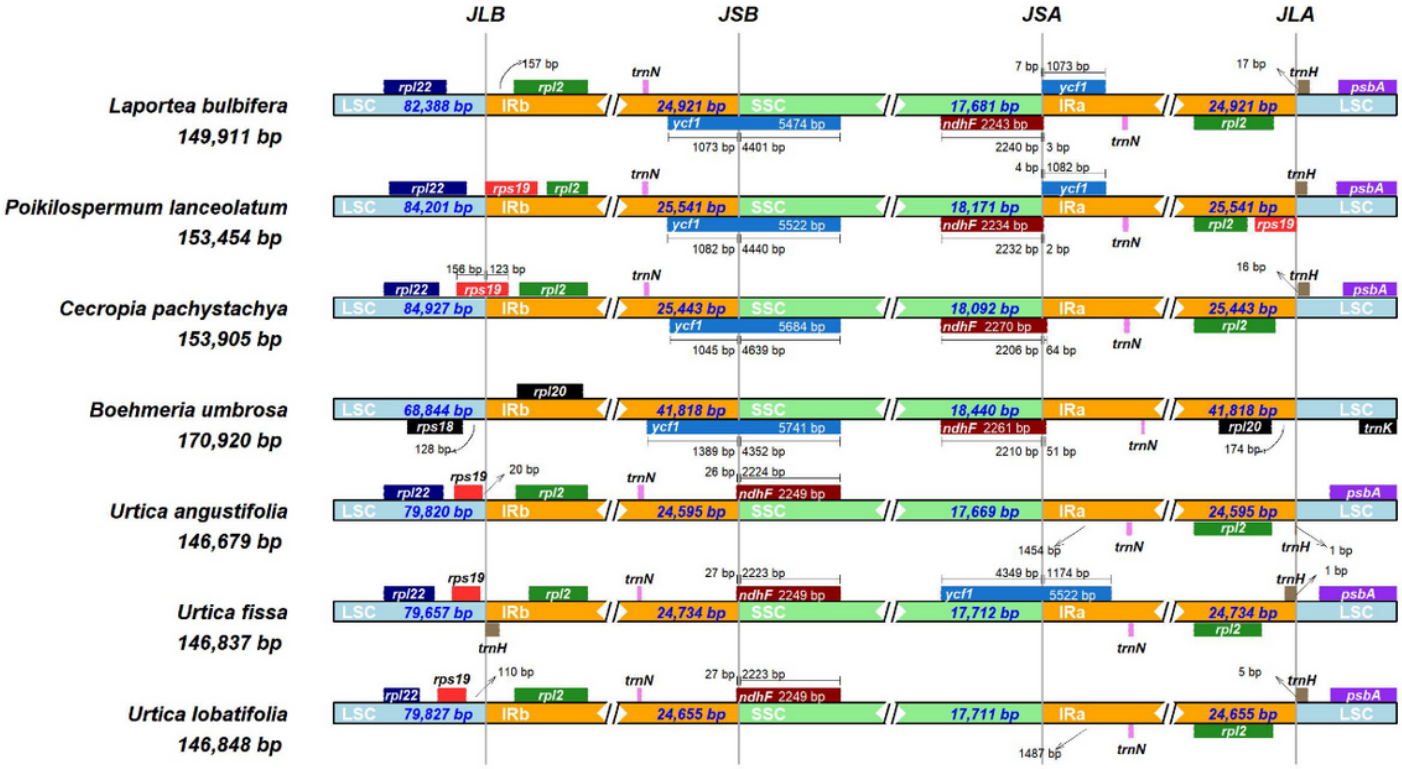

Figure 2

Comparison of IR, LSC and SSC regions among Urticaceae. The numbers above, below or adjacent to genes represent the length of genes or the distances from the front or end of genes to the boundary sites. It should be pointed out that the figure features are not to scale.

Figure 3

Extent of the gene rearrangements of *L. bulbifera* and its relative species (with *Nicotiana* as the reference species) using the MAUVE. Locally collinear blocks of the sequences are colour-coded and connected by lines.

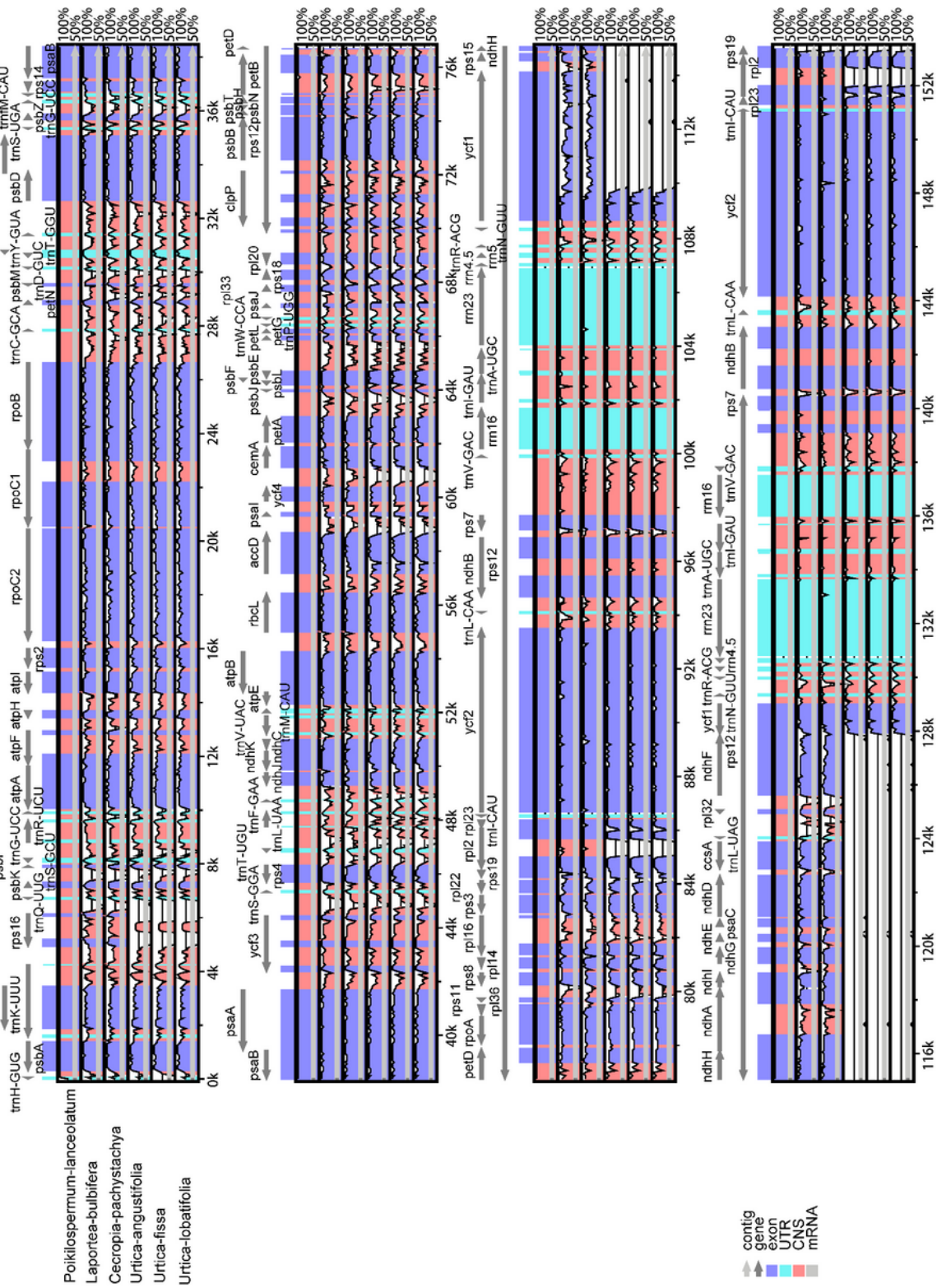

Figure 4

Comparison of the 6 cp genomes using mVISTA program. The gray arrows indicate the orientations and positions of genes. Untranslated, conserved non-coding and coding regions were characterized by sky-blue, red and blue blocks, respectively.

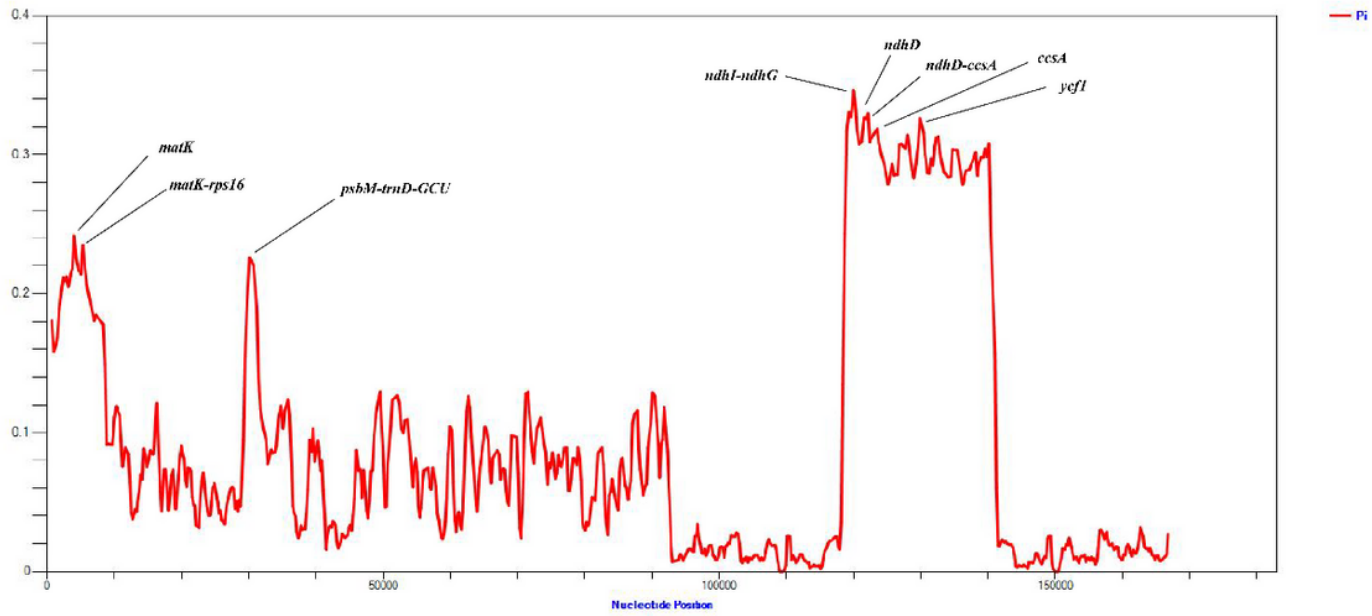

Figure 5

Nucleotide diversity of various shared regions in the cp genomes of *L. bulbifera* and its relative species.

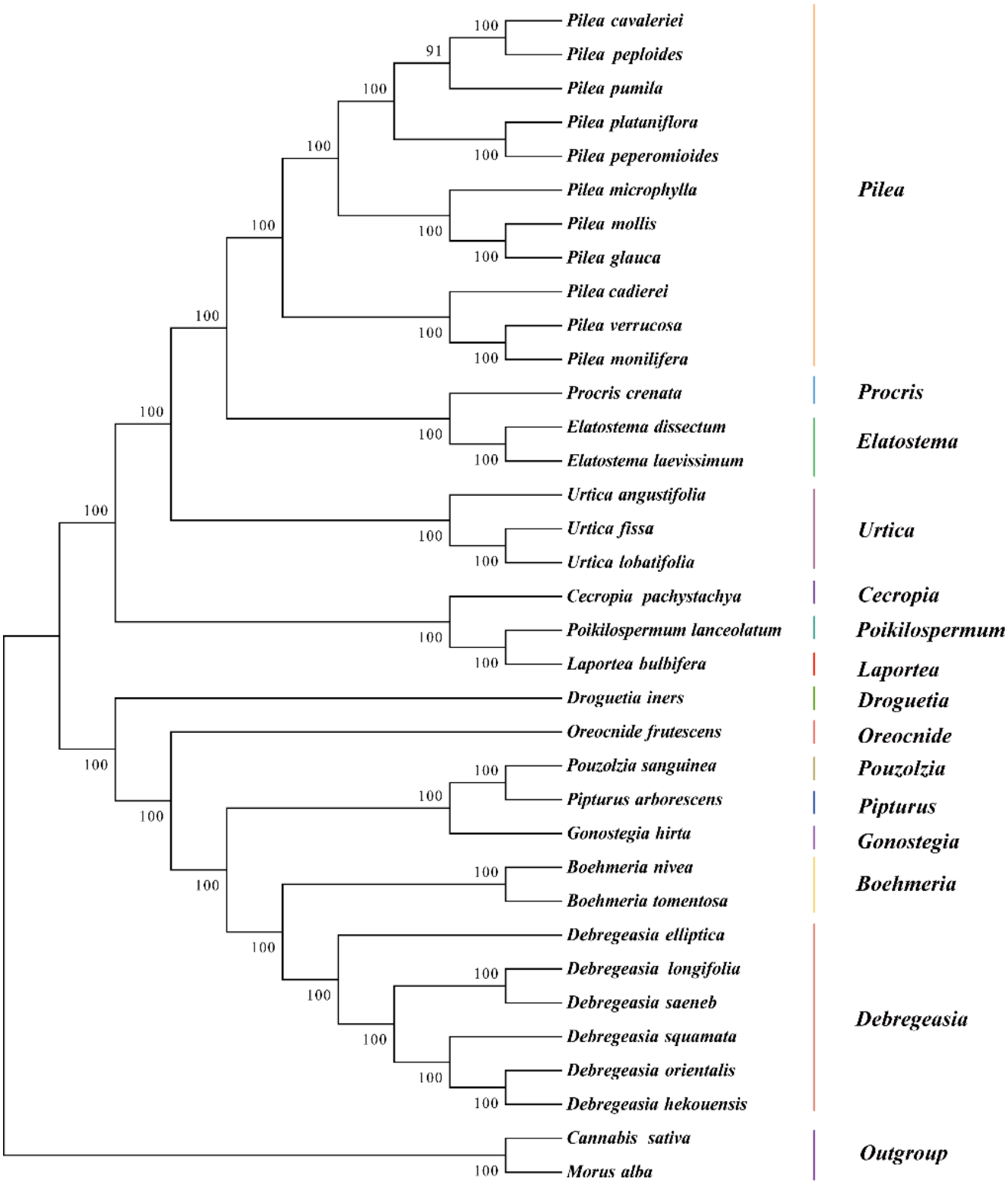

Figure 6

Phylogenetic relationships of 33 species based on complete cp genomes. The bootstrap values were based on 1000 replicates and are shown next to the branches. Different colors represent different genera.

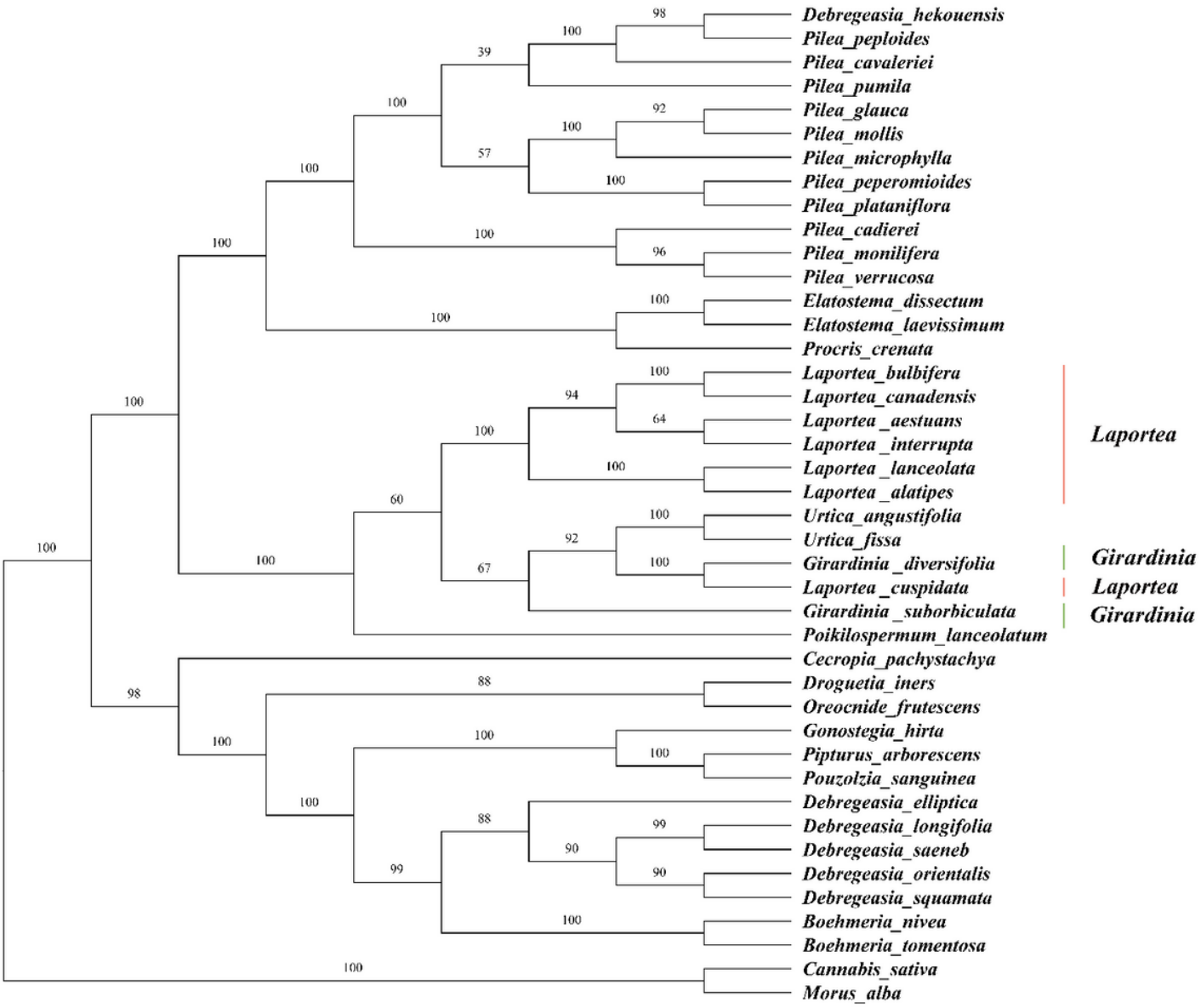

Figure 7

Phylogenetic relationships by maximum likelihood (ML) analyses based on *matK*. The bootstrap values are shown next to the branches. Different colors represent different genera.

Supplementary Files

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

- 0623SupplementaryMaterial.docx