



Heterogeneous Genetic Diversity Estimation of a Promising Domestication Medicinal Motherwort *Leonurus Cardiaca* Based on Chloroplast Genome Resources

Jiahui Sun^{1†}, Yiheng Wang^{1†}, Thomas Avery Garran¹, Ping Qiao^{1,2}, Mengli Wang¹, Qingjun Yuan^{1*}, Lanping Guo^{1*} and Luqi Huang^{1*}

¹National Resource Center for Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing, China,

²Academician workstation, Jiangxi University of Traditional Chinese Medicine, Nanchang, China

OPEN ACCESS

Edited by:

Meng Li,
Nanjing Agricultural University, China

Reviewed by:

Himanshu Sharma,
National Agri-Food Biotechnology
Institute, India
Mayank Kaashyap,
RMIT University, Australia

*Correspondence:

Qingjun Yuan
yuanqingjun@icmm.ac.cn
Lanping Guo
glp01@126.com
Luqi Huang
huangluqi01@126.com

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Plant Genomics,
a section of the journal
Frontiers in Genetics

Received: 05 June 2021

Accepted: 01 September 2021

Published: 15 September 2021

Citation:

Sun J, Wang Y, Garran TA, Qiao P,
Wang M, Yuan Q, Guo L and Huang L
(2021) Heterogeneous Genetic
Diversity Estimation of a Promising
Domestication Medicinal Motherwort
Leonurus Cardiaca Based on
Chloroplast Genome Resources.
Front. Genet. 12:721022.
doi: 10.3389/fgene.2021.721022

Leonurus cardiaca has a long history of use in western herbal medicine and is applied for the treatment of gynaecological conditions, anxiety, and heart diseases. Because of its botanical relationship to the primary Chinese species, *L. japonicus*, and extensive medical indications that go beyond the traditional indications for the Chinese species, it is a promising medicinal resource. Therefore, the features of genetic diversity and variability in the species have been prioritized. To explore these issues, we sequenced the chloroplast genomes of 22 accessions of *L. cardiaca* from different geographical locations worldwide using high-throughput sequencing. The results indicate that *L. cardiaca* has a typical quadripartite structure and range from 1,51,236 bp to 1,51,831 bp in size, forming eight haplotypes. The genomes all contain 114 distinct genes, including 80 protein-coding genes, 30 transfer RNA genes and four ribosomal RNA genes. Comparative analysis showed abundant diversity of single nucleotide polymorphisms (SNPs), indels, simple sequence repeats (SSRs) in 22 accessions. Codon usage showed highly similar results for *L. cardiaca* species. The phylogenetic and network analysis indicated 22 accessions forming four clades that were partly related to the geographical distribution. In summary, our study highlights the advantage of chloroplast genome with large data sets in intraspecific diversity evaluation and provides a new tool to facilitate medicinal plant conservation and domestication.

Keywords: traditional herbal medicine, new medicinal source, simple sequence repeat, INDEL, codon usage, intraspecific variation

INTRODUCTION

The *Leonurus* genus consists of herbaceous perennial plants that are widely distributed in Asia and Europe and naturalized in America and Africa (National Commission of Chinese Pharmacopoeia, 2015). One species that has long been used in Western herbal medicine and is commonly called “motherwort,” is *Leonurus cardiaca*. As the vernacular name indicates, it is applied for the treatment of gynaecological diseases. *L. cardiaca* is the typical species used medicinally of the *Leonurus* genus in Western countries, such as England, Poland, Bulgaria, Czech, the United States, etc. (Paweł et al.,

2014; Upton, 2018). *L. cardiaca*, like its Chinese cousin, is frequently used for regulating menstruation and treating other gynecological diseases. However, it is also commonly used to treat anxiety, sleeplessness, and heart diseases (Shikov et al., 2011; Pitschmann et al., 2017; Zhang et al., 2018; Garran et al., 2019; Garran, 2020). And, unlike its Chinese cousin, it is a perennial, and because of the harvesting methods used, it can be harvested for several years without need for tilling the soil, which may be of benefit to conservation of soil and water. Due to its wide clinical use and economic value, *L. cardiaca* is a meritorious medicinal resource that hosts promising values for domestication and cultivation in China.

Typical of many domesticated medicinal plants, the domestication of the wild plants may reduce genetic diversity, which can result in a genetic bottleneck (Doebley et al., 2006; Miller and Gross, 2011; Wang et al., 2020). Although it is impossible to know how much of the original wild genetic diversity was brought into the progenitor population, it is likely that, at least, some genetic diversity has been left behind. At the same time, the domestication of the wild plants are under the effects of artificial selection periods to meet human needs, which likely led to a decrease of genetic diversity. Therefore, understanding the population structure and genetic diversity of *L. cardiaca* is an important step for more in-depth investigation in order to avoid potentially serious genetic problems in the future.

Genetic diversity is the basis of evolutionary change, including all types of variation of single nucleotide polymorphisms (SNPs), insertion-deletion (Indel), and structural variation (Sv). Deterministic and stochastic forces, over millions of years, including natural selection, adaptation, and genetic drift have created abundant genetic diversity that generates considerable genotypic and phenotypic diversity. It is also the foundation of species, population, and individual diversity (Nevo, 2001). For medicinal plants, by all means of molecular features, genetic diversity is crucially important for species identification and to guide breeding population selection. Effective introduction, conservation, and utilization of medicinal plants also requires clarity of genetic diversity. And, considering that many important medicinal plants are facing severe threats due to increasing demand, overharvesting, and habitat loss, the determination of genetic diversity is urgently needed for better introduction, conservation, and utilization of medicinal plants.

Previously, several biomarkers have been used to evaluate the genetic diversity of *L. cardiaca*, such as amplified fragment length polymorphism (AFLP), random amplified polymorphic DNA (RAPD), inter-simple sequence repeats (ISSR), inter-retrotransposon amplified polymorphism (IRAP), and the inter-primer binding site (iPBS) (Khadiji-Khub and Soorni, 2014; Borna et al., 2017), indicating that a number of molecular variations exist among *L. cardiaca* populations. However, some advanced biomarkers, such as the chloroplast genome, can provide much more information for genetic diversity studies.

Chloroplasts are key organelles for photosynthesis and other biochemical pathways in plants. A chloroplast has its own genome. Determination of chloroplast genomes was once carried out by either isolation of chloroplasts or PCR

amplification of whole chloroplast genomes, both of which are laborious and time-consuming (Nock et al., 2011; Souza et al., 2019). With the advent of next-generation sequencing technologies, a chloroplast genome can be determined by sequencing total genomic DNA and *de novo* assembling whole chloroplast genomes (McPherson et al., 2013; Brozynska et al., 2014). The chloroplast genome usually has a typical circular quadripartite structure, including a small single copy region (SSC) and a large single copy region (LSC), which are separated by a pair of inverted repeat regions (IRa, IRb) and harbor 110 to 130 genes with sizes ranging from 120 to 165 kb (Ravi et al., 2008; Cheng et al., 2017). Exceptions have been found in Alismatales, Fabaceae, Geraniaceae, Aristolochiaceae, and many parasitic plants with abnormal genome sizes and structures (Csanad and Pal, 2014; Blazier et al., 2016; Ross et al., 2016; Sinn et al., 2018). Owing to rare recombination, small genome size, uniparental transmission, and moderate evolution rate, chloroplast genome research has been used extensively in different scientific fields. In molecular phylogeny, it can clearly reflect the relationship at different taxonomic levels, such as *Chaenomeles*, *Juglans*, *Coryloideae*, *Angelica*, and *Distylium*, and even difficult relationships within Fabaceae can be addressed (Dong et al., 2017; Hu et al., 2017; Dong et al., 2018; Hu et al., 2020; Sun et al., 2020; Zhang et al., 2020; Wang et al., 2021a; Dong et al., 2021). In phylogeographical analysis, the advantage of the chloroplast genome in non-recombination and uniparental inheritance can allow for successful estimation of divergence times and determine a biogeographic history (Zhang et al., 2017; Liu et al., 2018; Zhao et al., 2019; del Valle et al., 2019). Moreover, chloroplast genome markers, such as single nucleotide polymorphisms (SNPs) and simple sequence repeats (SSRs), can be used for population diversity estimations (Cao et al., 2018; Bautista et al., 2020; Ren et al., 2020).

Thus far, the chloroplast genome of *L. cardiaca* has not been reported. Since it is a new potential resource for domestication within China, we chose to use the chloroplast genome to infer the divergence within *L. cardiaca*. In this study, we sequenced 22 accessions of *L. cardiaca* chloroplast genomes from different geographical locations around the world using the Illumina HiSeq platform and compared heterogeneous divergence, such as SNPs, insertion/deletions, simple sequence repeats, and codon usage. The objectives of this study was 1) to determine the chloroplast genome of *L. cardiaca* and evaluate the intraspecific variation in this species and 2) to provide baseline data for genetic resources of *L. cardiaca*, including SNPs, SSRs, and indels, for genetic diversity assessment, to guide future domestication and conservation efforts.

MATERIALS AND METHODS

Plant Material, DNA Extraction and Sequencing

A total of 22 *L. cardiaca* accessions were collected from China, United States, and Europe (**Supplementary Table S1**) to represent the distribution of this species. Voucher specimens

were deposited at the herbarium of the Institute of Chinese Materia Medica (CMMI). Total genomic DNA was extracted from fresh leaves of a single individual using the Plant DNA Kit (D200-200, <http://www.gene-better.cn/>) from Genebetter Life Science Co., Ltd. and purified it using a Wizard DNA clean-up system (Promega, Madison, WI, United States). A paired-end library was constructed using a NEBNext Ultra™ DNA library prep kit and PE150 sequencing of 22 accessions was conducted on an Illumina HiSeq XTen platform at Novogene (Tianjin, China).

Genome Assembly and Annotation

Contigs were *de novo* assembled from the high-quality paired-end reads by using the SPAdes 3.6.1 program (Kmer = 95) (Bankevich et al., 2012) after low-quality reads were filtered using Trimmomatic 0.39 (Bolger et al., 2014). The chloroplast genome sequence contigs was extracted directly from the initial assembly by performing a BLAST search (Altschul et al., 1990) using the closely related species *L. japonicus* (GenBank: NC038062) as the reference. The selected contigs were assembled using Sequencher 5.4.5 (Gene Codes Corporation, Ann Arbor, MI United States, <http://www.genecodes.com>). A double-check process, Geneious 8.1, was used to map all reads to the assembled chloroplast genome sequence to verify the assembly accuracy (Kearse et al., 2012). The complete chloroplast genome sequences were annotated with Plann (Huang and Cronk, 2015) by using *L. japonicus* (GenBank: NC038062) as the reference and were manually adjusted in Sequin. The circle chloroplast map was drawn using the online program OrganellarGenomeDRAW (OGDRAW) (Lohse et al., 2007). The genome features of different regions were calculated using Geneious 8.1.

Single Nucleotide Polymorphisms, Indels, and Divergent Hotspot Identification

A total of 22 chloroplast genomes were aligned using MAFFT online service in auto strategy (Kato and Standley, 2013; Kato et al., 2019) and manually adjusted by Se-al 2.0 (Rambaut, 2002); these sequences formed eight haplotypes. The SNPs and indels were calculated by DNasp (Librado and Rozas, 2009) and MEGA X (Kumar et al., 2018) in default parameter. The synonymous and nonsynonymous SNPs and nucleotide transitions and transversions were manually checked in MEGA. The nucleotide diversity of the chloroplast genome was calculated based on sliding window analysis using DnaSP v5.10 software with parameter settings of a 600-bp window length and 100-bp step length. The IR/SC boundary map of eight haplotypes was performed in IRscope (Amiryousefi et al., 2018).

Codons and Microstructural Events

All protein-coding genes were isolated using a Python script written by Wuping (<http://github.com/wpwupingwp/>). The variable mutation sites, parsimony information sites, and relative synonymous codon usage (RSCU) values were analysed using DNasp and MEGA X. The heatmap from all RSCUs of the chloroplast genome was carried out using TBtools (Chen et al., 2020).

The software REPuter was used to visualize the dispersed and palindromic repeats with the following parameters: Hamming distance = 3, repeat size ≥ 30 bp and at least 90% similarity (Stefan et al., 2001). Tandem repeats were identified using the Tandem Repeats Finder with default parameters (Benson, 1999). Simple sequence repeats (SSRs) were obtained using Genome-wide Microsatellite Analysing Tool Package (GMATA) software (Wang and Wang, 2016) with the search parameters set at >10 repeat units for mononucleotides, >5 repeat units for dinucleotides, >4 repeat units for trinucleotides, >3 repeat units for tetranucleotide, pentanucleotide, and hexanucleotide SSRs. The sequence of "USOR-2" was used as a standard reference to count the microstructural events.

Phylogenetic Inference and Network Analysis

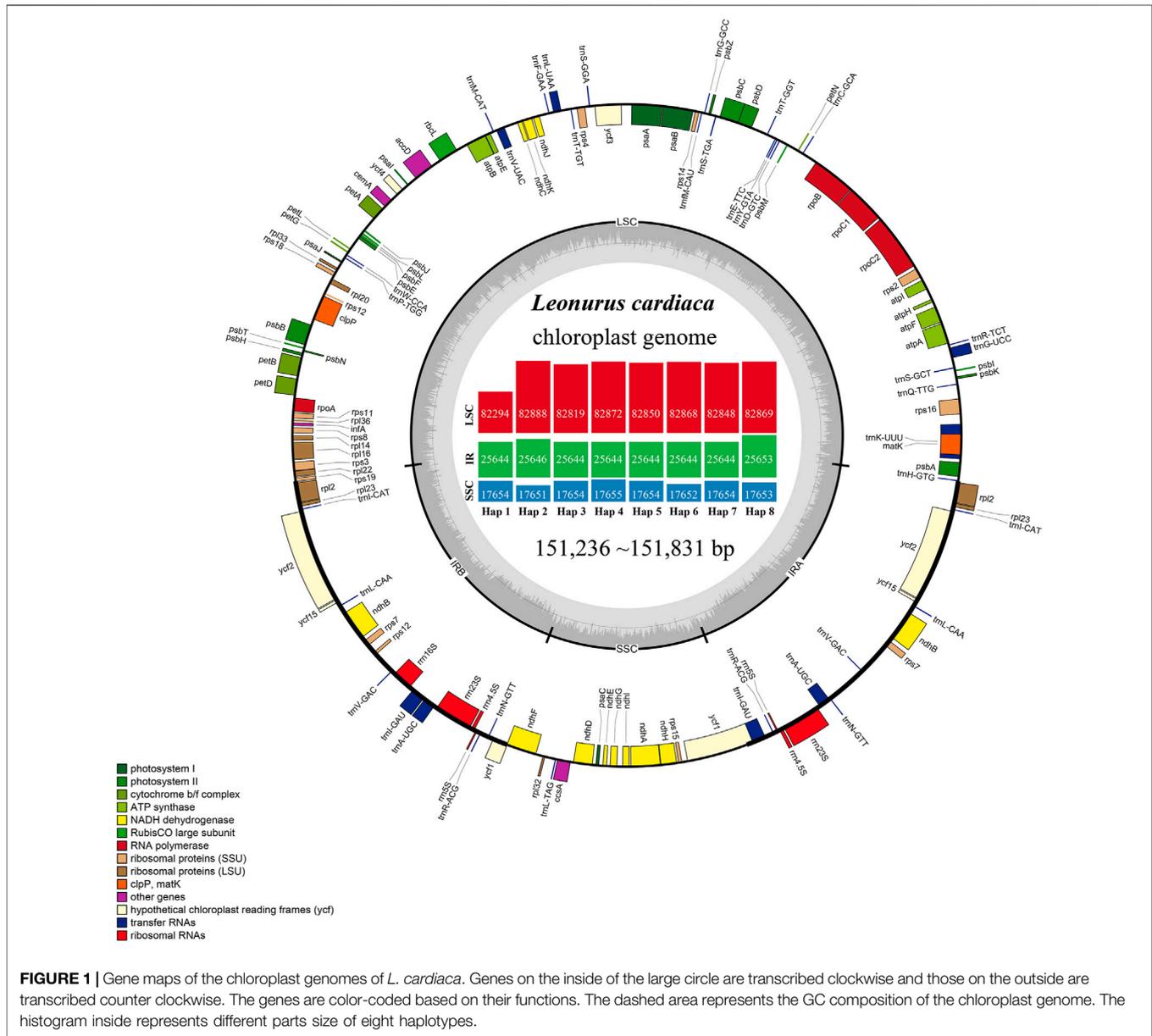
The optimal substitution mode was finalized using ModelFinder (Kalyaanamoorthy et al., 2017). Phylogenetic analysis was carried out using RAxML v.8.2 in the maximum likelihood (ML) method with the GTR + G model (Stamatakis, 2014). Node support values were determined with 500 rapid bootstrap replicates; bootstrap values branches ($<50\%$) were merged. Haplotype data was generated in DNasp and haplotype frequencies in populations were calculated by Arlequin v3.5.1.3 (Excoffier and Lischer, 2010), and a TCS network of 22 chloroplasts was generated by the network program PopArt v1.7 (Clement et al., 2000; Leigh and Bryant, 2015) using the haplotype data and populations haplotype frequencies data. *L. japonicus* (GenBank: NC038062) was used as an outgroup in both the phylogenetic analysis and TCS network.

RESULT

Characteristics of *L. cardiaca* Chloroplast Genome

The number of paired-end raw reads obtained by the Illumina HiSeq Xten system ranged from 16,330,634 to 31,010,620 for the 22 *L. cardiaca* accessions, clean reads of chloroplast genome ranging from 572,567 to 3,688,804 were extracted, yielding $566 \times$ to $3,659 \times$ chloroplast genome coverage (Supplementary Table S1). The accession numbers MZ274149 to MZ274170 of the complete chloroplast genome sequences were deposited in GenBank (Supplementary Table S1). The genome size ranged from 1,51,236 bp to 1,51,831 bp. All genome structures were extremely well conserved, and, as with most angiosperms, contained typical quadripartite structures with a pair of IR regions (25,644–25,653 bp), LSC regions (82,294–82,888 bp) and SSC regions (17,651–17,655 bp). The GC content of all the sequences was consistent in the LSC, SSC, and IR regions, accounting for 36.6, 32.2, and 43.4%, respectively. The main reason for the higher GC content of the IR regions was that the IR regions contained four high GC content rRNA genes. A total of 22 accessions of *L. cardiaca* forming eight haplotypes were obtained.

L. cardiaca contained 114 distinct functional genes, including 80 protein-coding genes, 30 tRNA genes, and four rRNA genes.



Eighteen genes were duplicated in the IR regions (**Figure 1; Table 1**). In addition, a total of 84 genes were located in LSC, including 62 protein-coding and 22 tRNA genes, while the SSC region harboured 11 protein-coding genes and one tRNA gene. Fourteen genes (*atpF*, *rpoC1*, *ndhB*, *petB*, *rpl2*, *ndhA*, *rps12*, *rps16*, *trnA-UGC*, *trnI-GAU*, *trnK-UUU*, *trnL-UAA*, *trnG-GCC* and *trnV-UAC*) contained a single intron, and two genes (*clpP* and *ycf3*) had two introns. The *trnK-UUU* gene had the largest intron, in which the *matK* gene was completely contained. The *rps12* gene is a trans-spliced gene with a 5'-end located in the LSC region and a 3' end located in the IR region. The border regions and adjacent genes of eight haplotypes were compared to analyze the expansion and contraction in junction regions (**Supplementary Figure S1**). The expansion and contraction at the IR/SC borders exhibited completely identical structures

among haplotypes. The IRb/LSC junction (JLB) occurred in the *rps19* gene, thus *rps19* had a 34 bp extension in the IRb region. The *ndhF* gene overlapped with the IRb/SSC junction (JSB) in 20 bp. The *ycf1* gene crossed over the IRa/SSC junction (JSA) in 1,084 bp, as well as in the JSB. In addition, IRa/LSC junction (JLA) extended into the *trnH-GUG* gene with only 1 bp.

Single Nucleotide Polymorphisms, Indels, and Divergent Hotspot Identification

The 22 alignment sequences were 1,52,010 bp in length and contained 225 SNP mutation sites (IR region only counted one time), including 83 singleton variable sites and 142 parsimony informative sites, forming eight haplotypes with a haplotype polymorphism of 0.732 (**Table 2, Supplementary**

TABLE 1 | The basic chloroplast genome information of *L. cardiaca* species.

Category for genes	Group of genes	Name of genes	
Photosynthesis related genes	Rubisco	<i>rbcL</i>	
	PhotosystemI	<i>psaA, psaB, psaC, psal, psaJ</i>	
	Assembly/stability of photosystem I	<i>*ycf3, ycf4</i>	
	Photosystem II	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ</i>	
	ATP synthase	<i>atpA, atpB, atpE, *atpF, atpH, atpI</i>	
	Cytochrome b/f compelx	<i>petA, *petB, *petD, petG, petL, petN</i>	
	Cytochrome c synthesis	<i>ccsA</i>	
	NADPH dehydrogenase	<i>*ndhA, *ndhB, ndhC, ndhD, ndhE, ndhF,ndhG, ndhH, ndhI, ndhJ, ndhK</i>	
	Transcription and translation related genes	Transcription	<i>rpoA, rpoB, *rpoC1, rpoC2</i>
		Ribosomal proteins	<i>rps2, rps3, rps4, rps7, rps8, rps11, *rps12, rps14,rps15, *rps16, rps18, rps19,*rpl2, rpl14, *rpl16, rpl20, rpl22, rpl23, rpl32, rpl33,rpl36</i>
Translation initiation factor		<i>infA</i>	
RNA genes		Ribosomal RNA	<i>rrn5, rrn4.5, rrn16, rrn23</i>
	Transfer RNA	<i>*trnA-UGC, trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, *trnG-UCC, trnG-GCC, trnH-GUG, trnI-CAU, *trnI-GAU,*trnK-UUU, trnL-CAA, *trnL-UAA, trnL-UAG, trnM-CAUI, trnM-CAU, trnN-GUU, trnP-UGG, trnQ-UUG, trnR-ACG, trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-UGU, trnV-GAC, *trnV-UAC, trnW-CCA, trnY-GUA</i>	
	Other genes	RNA processing	<i>matK</i>
Carbon metabolism		<i>cemA</i>	
Fatty acid synthesis		<i>accD</i>	
Proteolysis		<i>*c1pP</i>	
Genes of unknown function	Conserved reading frames	<i>ycf1, ycf2, ycf15</i>	

Intron-containing genes are marked by asterisks (*).

TABLE 2 | Haplotype diversity and mutation of 22 *L. cardiaca* accessions.

Location	Length	Polymorphic sites	Singleton variable sites	Parsimony informative sites	Nucleotide diversity	Number of hap	Hap diversity
LSC	83,040	151	60	91	0.00048	8	0.732
IR	25,654	10	3	7	0.00014	6	0.71
SSC	17,662	64	20	44	0.00101	6	0.71
Total	152,010	225	83	142	0.00042	8	0.732

TABLE 3 | Summary of variants detected in *L. cardiaca* chloroplast genomes.

Variants	Exon	Spacer	Intron	Total	LSC density/kb	SSC density/kb	IR density/kb	Total density/kb
SNPs	99	114	12	225	1.82	3.62	0.39	1.48
Snon	55	—	—	—	—	—	—	—
Ssyn	44	—	—	—	—	—	—	—
Indels	3	39	7	49	0.48	0.28	0.16	0.32

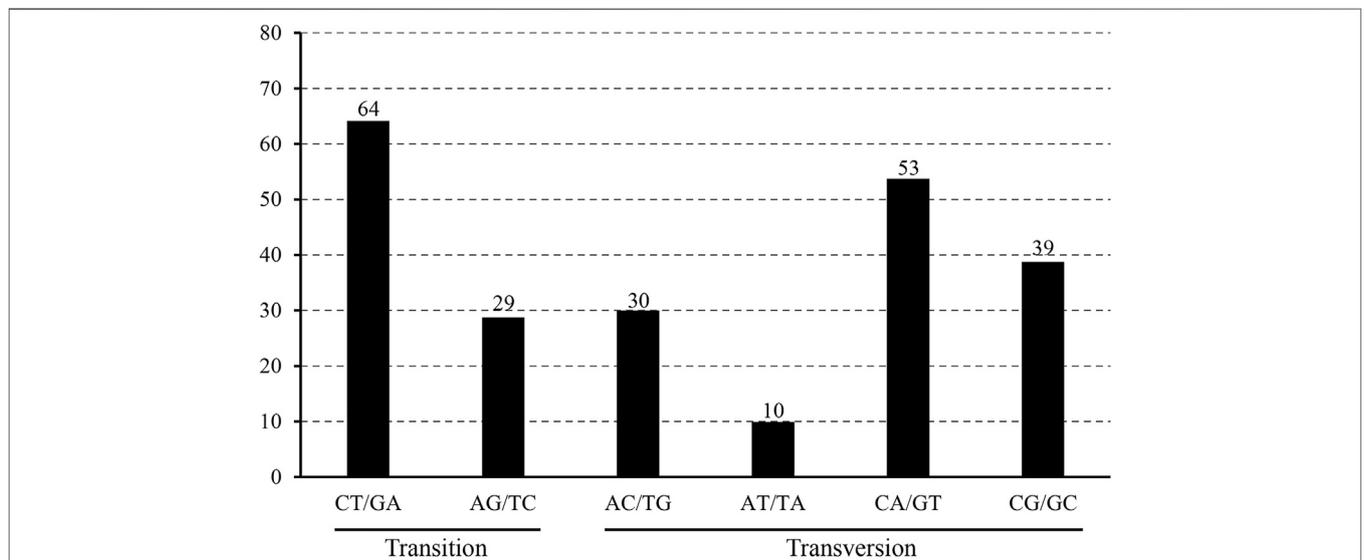
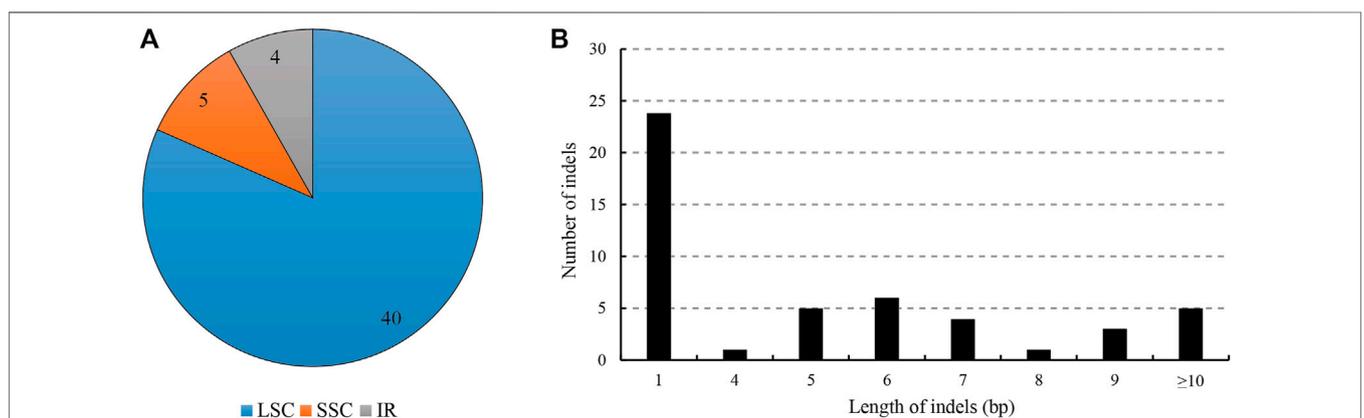
Table S1). The nucleotide diversity ranged from 0.00014 (IR region) to 0.00101 (SSC region) among the three different parts, and the overall nucleotide diversity was 0.00042. In the total 225 SNP mutations, 114 mutations were found in intergenic spacers with 99 located in exons and only 12 located in intron regions. The overall SNP density was 1.48 per kb (1.82 per kb in LSC, 3.62 per kb in SSC and 0.39 per kb) (**Table 3**). The 99 coding region SNPs were distributed in 34 different genes, meaning that some genes contained more than two SNPs, and 55 were nonsynonymous SNPs. Ten out of 34 genes contained more than 3 SNPs (*ycf1, rpoC2, ndhF, ndhH, matK, ndhD, psbA, ndhA, psaB* and *ycf2*) (**Table 4**). Among these ten highly SNP

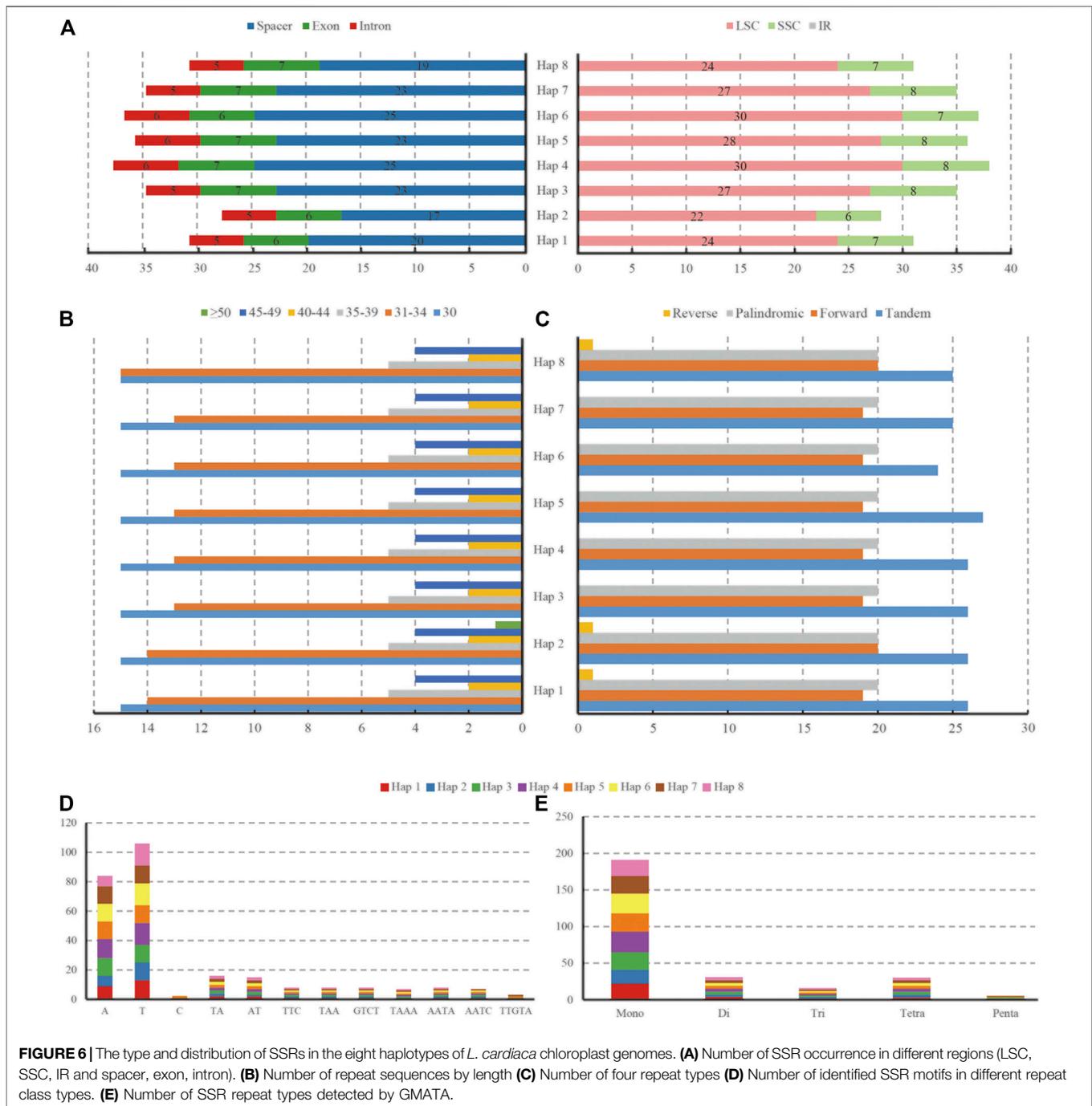
variable chloroplast coding genes, *ycf1* harboured a maximum of 21 SNPs, 17 were nonsynonymous SNPs and 4 synonymous SNPs, followed by nine SNPs in *rpoC2* and eight in *ndhF*. In addition, the highest SNP density of the coding genes occurred in *ndhH*, which contained 5.08 SNPs per kb. The amount and density of SNPs may indicate that for these 22 accessions, the coding genes *ycf1* and *ndhH* diverged markedly. The patterns of SNPs, 93 transitions (Ts) and 132 transversions (Tv) were counted, and the overall Ts:Tv rate was 0.705, indicating that it was in favor of transversions (**Figure 2**). The high frequency SNPs were C to T and G to A, and mutations from A to T and from T to A exhibited the lowest frequency.

TABLE 4 | Highly single nucleotide polymorphism variable chloroplast protein coding genes of *L. cardiaca*.

Gene	Snps	Non n	Syn n	Length bp	Density of SNPs/kb	Gene products
<i>ycf1</i>	21	17	4	5,570	3.77	ATP-binding cassette glutathione S-conjugate transporter
<i>rpoC2</i>	9	5	4	4,191	2.15	DNA-directed RNA polymerase subunit beta''
<i>ndhF</i>	8	2	6	2,223	3.6	NAD(P)H-quinone oxidoreductase subunit 5
<i>ndhH</i>	6	3	3	1,182	5.08	NAD(P)H-quinone oxidoreductase subunit 7
<i>matK</i>	5	4	1	1,536	3.26	Tyrosine-protein kinase
<i>ndhD</i>	4	1	3	1,410	2.84	NAD(P)H-quinone oxidoreductase subunit 4
<i>psaA</i>	4	0	4	2,253	1.78	Photosystem I P700 chlorophyll a apoprotein A1
<i>ndhA</i>	3	2	1	1,095	2.74	NAD(P)H-quinone oxidoreductase subunit 1
<i>psaB</i>	3	0	3	2,205	1.36	Photosystem I P700 chlorophyll a apoprotein A2
<i>ycf2</i>	3	3	0	6,867	0.44	Unknown

The listed gene contain three or more SNPs

**FIGURE 2** | The patterns of nucleotide substitution among the eight *L. cardiaca* haplotypes. Nucleotide substitution were divided into six types as indicated by the six non-strand-specific base-substitution types (i.e., numbers of considered G to A and C to T sites for each respective set of associated mutation types).**FIGURE 3** | Analyses of indels in the *L. cardiaca* chloroplast genomes. (A) Count of indel types and locations. (B) Number and size of indels in chloroplast genomes.



or U, except UUG. Moreover, for all codons, a bias in favor of purine at the third codon position was apparent, as reported by other previous studies.

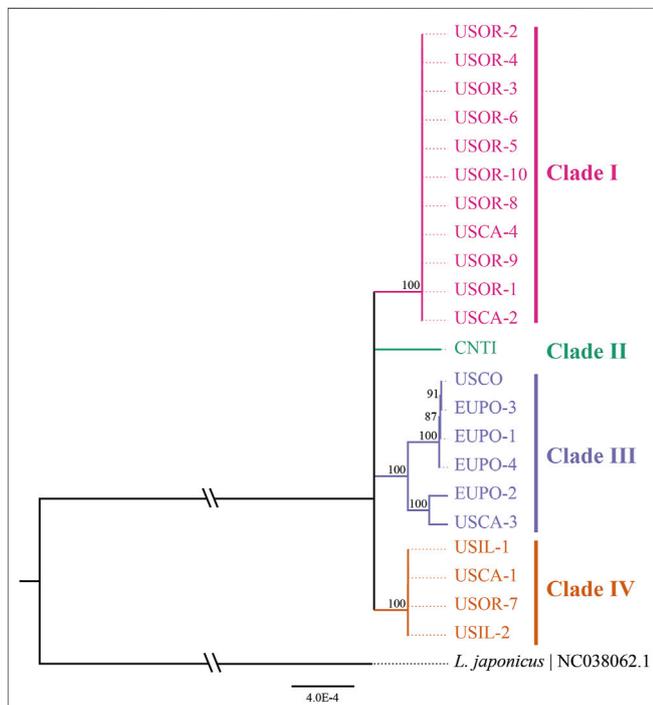
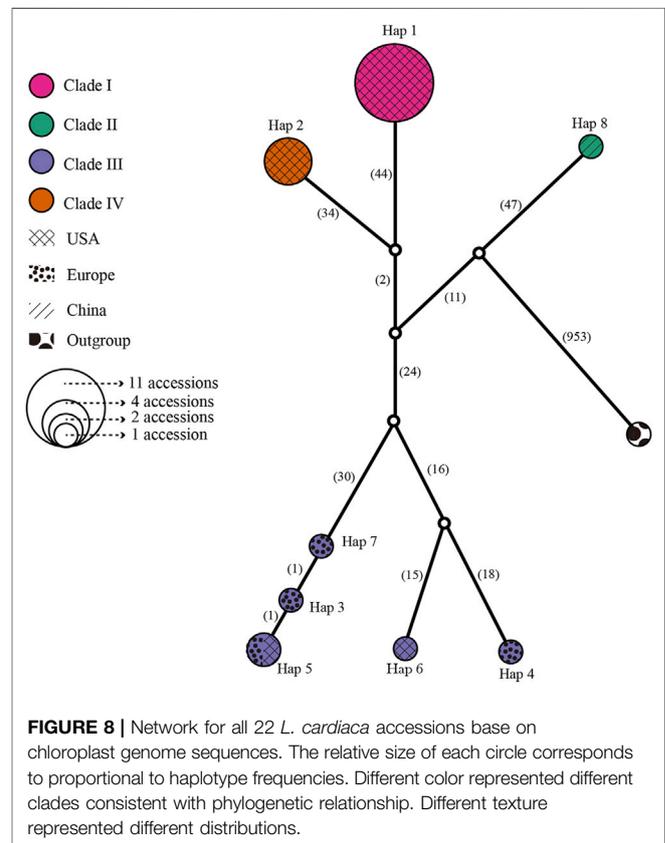
Repeat Sequences and Simple Sequence Repeats

We detected a total of 317 repeats with lengths of 30~52 bp in the forward, palindromic, and reverse regions of eight haplotypes and 39~41 repeats in each haplotype (Figure 6, Supplementary Table

S3). Specifically, the number of forward repeats ranged from 19 to 20, which was slightly less than that of palindromic repeats (20), and only one reverse repeat existed in Haps 1, 2, and 8. However, the complementary repeat was absent in any of the haplotypes. According to the range of length, we classified the repeats into six groups, as shown in Figures 6B,C. The most common repeat was 30 bp, and 84.5% of repeats were limited to 30~39 bp. Furthermore, 24~27 tandem repeats were detected in eight haplotypes. In addition, a total of 271 SSRs, mono-, di-, tri-, tetra-, and penta-, were detected by GMATA analysis. The

TABLE 5 | SSRs identified from in silico comparative analysis of the chloroplast genomes of eight haplotypes.

Region	Position	Location	Forward sequence	Reverse sequence	SSR type	Product size
LSC	<i>matK-trnK(UUU)</i>	Spacer	CCGAATCGCTCTTTTGATTTTC	CCCAGTCTAACTAGACGTT	(A)11	175
LSC	<i>trnK(UUU)-rps16</i>	Spacer	AAGGATGTATAGATACAA	ATATTGACAACAGTGTATTG	(A)13	162
LSC	<i>rps16</i>	Intron	GATCATAAAAAGCCCACCCCTT	TGGGCCCATTTTGCATTT	(C)10	206
LSC	<i>psbK-psbI</i>	Spacer	CTATAGGGTCTCCACAATA	ACTACGGTATAAACGAAGAG	(T)10	199
LSC	<i>psbI-trnS(GCU)</i>	Spacer	GGCGAAATCCTGACGTGAAGAA	AATTGGGAGAGATGGCTGAGTGA	(A)10	220
LSC	<i>trnG(UCC)-trnR(UCU)</i>	Spacer	TTCGATTCGCCCTACCCGCTTATT	CGTCCAATAGGATTTGAACCTATAC	(T)10	242
LSC	<i>atpF</i>	Intron	CGATTCAATTTGGCTCTCGCGCTC	CTCGATATAGAACACTCATCTC	(T)10	296
LSC	<i>atpF</i>	Intron	ATTTGGCTCTCGCGCTCAA	GTCAGAAGAGTCCTCTAAGTA	(A)11	122
LSC	<i>atpF</i>	Intron	GAATGTAATGAGCCTATCCTCTTC	GGCATAGGTCGTCGATTCATCATTG	(A)10	170
LSC	<i>atpF-atpH</i>	Spacer	AACTCCCGGCAGATGGCCAGTGCCCAA	CCTACGGGAAAGGCTGATTTCGC	(T)14	285
LSC	<i>petN-psbM</i>	Spacer	GAATTTGGAATCTCATTGG	GGTCCGTCTTCCTCAGTATCCATC	(TTGTA)3	217
LSC	<i>ycf3</i>	Intron	CCAGAACCCTTAGGTGAACC	TGCTGGCTATGCCTAGATCGCGAA	(A)11	329
LSC	<i>trnF(GAA)-ndhJ</i>	Spacer	GGCACATGGTTAAATTTGGATGAGG	GTATCTGAGCCGAATCTTG	(TAAA)3	301
LSC	<i>ndhC-trnV(UAC)</i>	Spacer	CACCTAGGATCTTGCTAAAG	CCTAGCCAAGCACACCAAAC	(T)11	317
LSC	<i>petA-psbJ</i>	Spacer	CGGGTTTCCAGGTGTATC	CACCTCTTCTCGATCTTG	(AT)5	268
LSC	<i>rpl33-rps18</i>	Spacer	GTATGCTTATCCTTGAAGGGGG	CGCCTACGAAAAGATCGCTTGG	(A)12	170
LSC	<i>clpP</i>	Intron	CGCTGAACTGGACTTCCGA	ATTCCCTCACGCTTGCGCCA	(A)10	205
LSC	<i>petB</i>	intron	GGTAGTATCACACAAGAAGAG	TGAGATAGGCGAAAACCCCC	(T)16	193
LSC	<i>rpl14-rps3</i>	spacer	GCAATTAATTTGAATAGATATT	GAGAAATGACGAACAGGTGCT	(A)10	180
SSC	<i>ndhF</i>	exon	AGCCACATACGATGAAGAT	ATTTGAAGTATTTAGCAGA	(A)11	281
SSC	<i>rpl32-trnL(UAG)</i>	spacer	TTATCATTGCAAGGCGGG	CTCTTGTAAATCCTAC	(T)11	336
SSC	<i>rps15-ycf1</i>	Spacer	CAGATATGGATTTTACCG	TTGATCTCGAAATGAATCACC	(AATC)3	207

**FIGURE 7** | Phylogenetic relationships among 22 *L. cardiaca* accessions constructed from complete chloroplast genome sequences using maximum likelihood (ML). The ML topology is shown, with the ML bootstrap support value for each node. 22 accessions cluster to four clades show by different color.**FIGURE 8** | Network for all 22 *L. cardiaca* accessions base on chloroplast genome sequences. The relative size of each circle corresponds to proportional to haplotype frequencies. Different color represented different clades consistent with phylogenetic relationship. Different texture represented different distributions.

number of SSRs ranged from 28 (Hap 2) to 38 (Hap 4). Among these SSRs, there were 175 in the spacers, 53 in the exons, and 43 in the introns (212 in LSC and 59 in SSC, but none in IR regions)

(Figure 6A). The majority of SSRs were mononucleotide repeats (70.5%), and most of them were A or T repeats (19–28). The dinucleotides and tetranucleotides were almost equal at 11.4 and

11.1%, respectively. The lowest numbers of repeat trinucleotides and pentanucleotides were 5.9 and 1.1%, respectively (Figures 6D,E).

Among all the haplotypes, 22 SSR loci showed polymorphisms after *in silico* analysis, 14 in spacers, 7 in introns, and only one was located in exons. The intron of *atpF* contained three polymorphic loci; the most compared with the others (Table 5).

Interspecific Relationships According to Phylogenetic and Network Analysis

A median-joining network and phylogenetic analysis were carried out based on the 22 entire chloroplast genomes of *L. cardiaca*. The ML analysis strongly indicated significant divergence among 22 accessions, forming four clades with essential bootstrap support. In addition, samples from Tibet formed a clade separately, and clades I and IV were all from samples collected in the United States. However, clade III contained samples from different distributions from the United States and Europe (Figure 7). Regardless, the network result was largely comparable to the phylogenetic result. Additionally, four clades clustered by eight haplotypes, while Hap 1 contained 11 accessions, Hap 2 contained four accessions, and Hap 5 contained two accessions. The steps among haplotypes showed strong intraspecific variations (Figure 8).

DISCUSSION

Conserved Chloroplast Genome Structure of *L. cardiaca*

The chloroplast genomes of *L. cardiaca* were determined for the first time in this study. Among the 22 chloroplast genomes, the sizes range from 1,51,236 bp to 1,51,831 bp, forming eight haplotypes. The structure of chloroplast genomes was highly conserved in *L. cardiaca*, gene order and contents, while no large genome rearrangement or gene loss was detected. All genomes harbor 114 distinct genes, including 80 protein-coding genes, 30 tRNA genes, and four rRNA genes, which is consistent with other related species (Lukas and Novak, 2013; Qian et al., 2013; Cheon et al., 2018). The genes *accD*, *rpl32*, and *ycf2* may be absent in some species, and they also do not occur in *L. cardiaca* (Jansen et al., 2007; Oliver et al., 2010; Wicke et al., 2011). The total GC content from different distributions is highly consistent, while the genome size varied slightly but not significantly. In addition, variations, such as GC content, gene content, and genome size, also has the potential to enable different species/populations or even individuals to be distinguished. All wild and cultivated accessions share the same quadripartite structure, gene number, order and IR/SC junction, and have a similar chloroplast size. Collectively, there is a highly conserved structure in the *L. cardiaca* chloroplast genome.

Sequence Variability and Candidate DNA Barcodes of *L. cardiaca*

A total of 225 SNPs were detected in the 22 chloroplast genomes of *L. cardiaca*. According to the base frequency, the SNP mutations are likely to be from C to T or G to A.

Transversion from A to T or from T to A is very rare. Most SNPs are located in LSCs, followed by SSC and IR regions, as are other mutation types, such as SSRs and indels. This is highly relevant to the region size and sequence conservation level. The number of SNPs of *L. cardiaca* (225 SNPs in 22 accessions) is relatively high when compared with that of other species, such as *Jacobaea vulgaris* (32 SNPs in 17 accessions), *Brassica napus* (294 SNPs in 488 accessions), *Brachypodium distachyon* (298 SNPs in 53 accessions) and *Macadamia integrifolia* (407 SNPs in 63 accessions) (Leonie et al., 2011; Qiao et al., 2016; Sancho et al., 2017; Nock et al., 2019). With respect to LSC, SSC, and IR regions, the SSC region is more variable with 3.62 SNPs per kb than LSC with 1.82 SNPs per kb. This result is consistent with other studies such as *Ricinus communis* and *Macadamia integrifolia* (Nock et al., 2019; Muraguri et al., 2020). Due to evolutionary constraints and natural selection pressure, the coding regions are more conserved than the noncoding regions (Liu et al., 2019; Bautista et al., 2020; Zhang et al., 2021). The number of SNPs occurring in noncoding regions (126 SNPs) was greater than that in coding regions (99 SNPs). Among coding genes, the occurrence of nonsynonymous SNPs is more abundant than that of synonymous SNPs. As expected, this result can be deduced from the substitution pattern of SNPs because transversion tends to generate more nonsynonymous mutations. The *ycf1* and *ndhH* genes have much more SNPs than other genes. Approximately 80% of SNPs of *ycf1* are nonsynonymous and may cause the functional differentiation of *ycf1* across *L. cardiaca* populations worldwide. The *ycf1* gene could serve as a DNA barcode for *Leonurus* considering its better performance than the core plant barcodes *matK* and *rbcL*. This gene has been proposed as a candidate DNA barcode of angiosperms and is frequently applied in many species for identification and phylogenetic analysis (Dong et al., 2015). The *ndhH* gene is regarded as the best-performing marker in discrimination of grasses (Krawczyk et al., 2018). The noncoding regions usually have higher sequence variability compared to coding regions. The spacer of *trnT-psbD* and intron of *clpP* are two mutation hotspots though they are not as variable as *ycf1*. The spacer of *trnT-psbD* has been considered an effective locus for phylogenetic research at low taxonomic levels (Dong et al., 2012). Therefore, these four loci could be used as candidate DNA barcodes for species authentication in the *Leonurus* genus.

Chloroplast Genome Structural Variations and Features of Codon Usage

Genome structural mutations are yet another form of information useful in revealing genetic diversity of species, population biology, or evolution. SSRs are the most common structural mutations, which cause insertions or deletions (McDonald et al., 2011; Ahmed et al., 2012; Yi et al., 2013; Abdullah et al., 2021). The most common SSRs in *L. cardiaca* chloroplast genomes are mononucleotides mainly composed of A or T and rarely G or C. There are significantly fewer Di-, Tetra-, Tri-, and Penta-nucleotide motif repeats. This phenomenon is similar to that of other studies in *Atractylodes* and *Chaenomeles* (Ahmed et al., 2012; Sinn et al., 2018; Hu et al., 2020; Sun et al.,

2020; Wang et al., 2021b). The SSR primers designed in this study for amplifying the 22 polymorphic loci are expected to facilitate evaluation of genetic diversity of *L. cardiaca*. There are two unique patterns in the relative synonymous codon usage (RSCU) and usage frequency based on eight haplotypes of protein-coding genes. First, all the high-frequency codons have a bias in favor of ending with A or U, but in low-frequency codons, they are biased towards G or C at the third codon position. Second, two start codons (UGG and AUG) have no bias yet. This also agrees with previous studies (Meng et al., 2018; Bautista et al., 2020; Ren et al., 2020).

Collectively, except for codon usage, SNPs, indels and SSRs vary within the *L. cardiaca* accessions and could be qualified genetic resources to evaluate the diversity among populations. Compared with other species as mentioned above (*Jacobaea vulgaris*, *Brassica napus*, etc.), *L. cardiaca* showed higher genetic diversity.

Tracing the Original Sources of Introduced *L. cardiaca*

The 22 genomes belong to eight haplotypes. These haplotypes form six clades. EUPO-1 (Poland, Hap 3) is a one-step mutant of EUPO-4 (Poland, Hap 7) and USCO (Colorado, Hap 5) is a one-step mutant of EUPO-1. *L. cardiaca* was naturalized from Europe to the United States. The nearly identical chloroplast genomes between EUPO-2 (Poland, Hap 4) and USCA-3 (California, Hap 6) and the Hap 5 containing accessions from Poland and Colorado could infer recent exchanges existed between the two continents. Four accessions from Europe contain their own unique haplotypes, indicating where the biodiversity center is. Close relationships among the genomes in Clade I and Clade IV indicate that the germplasm resources of *L. cardiaca* in United States have not been severely mixed. It is probably because *L. cardiaca* is not a widely popular herb and very limited seed exchange has occurred (Yuan et al., 2010; Wang et al., 2020). Due to very limited samples available from Europe, it is impossible to locate the places of origin of most samples, unlike the one collected in Tibet. Apparently, the chloroplast genome information does provide an opportunity to trace the origins of introduced or invasive plants.

Implication for Conservation of *L. cardiaca*

With the increasing demand for medical use, wild *L. cardiaca* resources are likely face over-exploitation. Bringing plants into cultivation as a crop is one solution. For the purpose of high yield and good quality, a seed bank for *L. cardiaca* needs to be constructed for easy access of the genetic resources during cultivar breeding. The genetic diversity revealed by the chloroplast genome information or the DNA barcodes proposed in this study could be used to guide the sampling processes and the subsequent evaluation of genetic integrity of this medicinal plant as was discussed by Yuan et al. (2010) for the Chinese medicinal plant *Scutellaria baicalensis* (Yuan et al., 2010).

In the period of seed bank construction, the chloroplast markers should be used to evaluate the overall genetic diversity and genetic structure of *L. cardiaca*. As has been shown in this study, the genetic diversity of *L. cardiaca* is highly structured and the seed bank to be built should contain as many variants as possible. During the maintenance stage of the seed bank, the molecular markers should

be used to monitor any loss of genetic diversity due to founder effect, genetic drift, stochastic event, or other factors.

CONCLUSION

Unambiguous genetic diversity and variability are not only prerequisites for the discovery of new medicinal resources but also a foundation of germplasm resource conservation and innovation. The genetic diversity of *L. cardiaca* was revealed with 22 chloroplast genomes. The biodiversity center is obviously in Europe because all four accessions have their own haplotypes. The molecular markers developed in this study could be used to guide the construction of germplasm resource banks either for conservation or cultivar breeding of this common medicinal plant. More extensive sampling in Europe is necessary for a better understanding of the overall genetic diversity and genetic structures of this species.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: Genbank (<https://www.ncbi.nlm.nih.gov/>), accession number (MZ274149 - MZ274170).

AUTHOR CONTRIBUTIONS

JS and YW did the data analysis and wrote the manuscript; TAG revised the manuscript. PQ and MW participated in the experiments; TAG and QY collected the study materials; QY, LG, and LH conceived and designed the research. The authors read and approved the final manuscript.

FUNDING

This work was financially supported by National Natural Science Foundation of China (No.81891014 and No.81874337) and the National Key Research and Development Program of China (2017YFC1703700: 2017YFC1703704). The funding agencies had no role in the design of the experiment, analysis, and interpretation of data and in writing the manuscript.

ACKNOWLEDGMENTS

The authors would like to thank professor Shiliang Zhou and Wenpan Dong for providing suggestion.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2021.721022/full#supplementary-material>

REFERENCES

- Abdullah, Henriquez, C. L., Croat, T. B., Poczai, P., and Ahmed, I. (2021). Mutational Dynamics of Aroid Chloroplast Genomes II. *Front. Genet.* 11, 610838.
- Ahmed, I., Biggs, P. J., Matthews, P. J., Collins, L. J., Hendy, M. D., and Lockhart, P. J. (2012). Mutational Dynamics of Aroid Chloroplast Genomes. *Genome Biol. Evol.* 4 (12), 1316–1323. doi:10.1093/gbe/evs110
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic Local Alignment Search Tool. *J. Mol. Biol.* 215 (3), 403–410. doi:10.1016/s0022-2836(05)80360-2
- Amiryousefi, A., Hyvönen, J., and 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
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: A New Genome Assembly Algorithm and its Applications to Single-Cell Sequencing. *J. Comput. Biol.* 19 (5), 455–477. doi:10.1089/cmb.2012.0021
- Bautista, M. A. C., Zheng, Y., Hu, Z., Deng, Y., and Chen, T. (2020). Comparative Analysis of Complete Chloroplast Genome Sequences of Wild and Cultivated *Bougainvillea* (Nyctaginaceae). *Plants* 9 (12), 1671. doi:10.3390/plants9121671
- Benson, G. (1999). Tandem Repeats Finder: a Program to Analyze DNA Sequences. *Nucleic Acids Res.* 27 (2), 573–580. doi:10.1093/nar/27.2.573
- Blazier, J. C., Jansen, R. K., Mower, J. P., Govindu, M., Zhang, J., Weng, M.-L., et al. (2016). Variable Presence of the Inverted Repeat and Plastome Stability in *Erodium*. *Ann. Bot.* 117 (7), 1209–1220. doi:10.1093/aob/mcw065
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* 30 (15), 2114–2120. doi:10.1093/bioinformatics/btu170
- Borna, F., Luo, S., Ahmad, N. M., Nazeri, V., Shokrpour, M., and Trethowan, R. (2017). Genetic Diversity in Populations of the Medicinal Plant *Leonurus Cardiaca* L. Revealed by Inter-primer Binding Site (iPBS) Markers. *Genet. Resour. Crop Evol.* 64 (3), 479–492. doi:10.1007/s10722-016-0373-4
- Brozynska, M., Furtado, A., and Henry, R. J. (2014). Direct Chloroplast Sequencing: Comparison of Sequencing Platforms and Analysis Tools for Whole Chloroplast Barcoding. *PLoS One* 9 (10), e110387. doi:10.1371/journal.pone.0110387
- Cao, J., Jiang, D., Zhao, Z., Yuan, S., Zhang, Y., Zhang, T., et al. (2018). Development of Chloroplast Genomic Resources in Chinese Yam (*Dioscorea Polystachya*). *Biomed. Res. Int.* 2018, 6293847. doi:10.1155/2018/6293847
- Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., et al. (2020). TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Mol. Plant* 13 (8), 1194–1202. doi:10.1016/j.molp.2020.06.009
- Cheng, H., Li, J., Zhang, H., Cai, B., Gao, Z., Qiao, Y., et al. (2017). The Complete Chloroplast Genome Sequence of Strawberry (*Fragaria × ananassa* Duch.) and Comparison with Related Species of Rosaceae. *PeerJ* 5 (10), e3919. doi:10.7717/peerj.3919
- Cheon, K.-S., Jeong, I.-S., Kim, K.-H., Lee, M.-H., Lee, T.-H., Lee, J.-H., et al. (2018). Comparative SNP Analysis of Chloroplast Genomes and 45S nrDNAs Reveals Genetic Diversity of *Perilla* Species. *Plant Breed. Biotech.* 6 (2), 125–139. doi:10.9787/pbb.2018.6.2.125
- Clement, M., Posada, D., and Crandall, K. A. (2000). TCS: a Computer Program to Estimate Gene Genealogies. *Mol. Ecol.* 9 (10), 1657–1659. doi:10.1046/j.1365-294x.2000.01020.x
- Csanad, G., and Pal, M. (2014). Two Distinct Plastid Genome Configurations and Unprecedented Intraspecies Length Variation in the accD Coding Region in *Medicago Truncatula*. *DNA Res.* 21 (4), 417. doi:10.1093/dnares/dsu007
- del Valle, J. C., Casimiro-Soriguer, I., Buide, M. L., Narbona, E., and Whittall, J. B. (2019). Whole Plastome Sequencing within *Silene* Section *Psammophilae* Reveals Mainland Hybridization and Divergence with the Balearic Island Populations. *Front. Plant Sci.* 10, 1466. doi:10.3389/fpls.2019.01466
- Doebley, J. F., Gaut, B. S., and Smith, B. D. (2006). The Molecular Genetics of Crop Domestication. *Cell* 127 (7), 1309–1321. doi:10.1016/j.cell.2006.12.006
- Dong, W., Liu, Y., Yu, J., Wang, L., and Zhou, S. (2012). Highly Variable Chloroplast Markers for Evaluating Plant Phylogeny at Low Taxonomic Levels and for DNA Barcoding. *PLoS One* 7 (4), e35071. doi:10.1371/journal.pone.0035071
- Dong, W., Liu, Y., Xu, C., Gao, Y., Yuan, Q., Suo, Z., et al. (2021). Chloroplast Phylogenomic Insights into the Evolution of *Distylium* (Hamamelidaceae). *BMC Genom.* 22, 293. doi:10.1186/s12864-021-07590-6
- Dong, W., Xu, C., Li, C., Sun, J., Zuo, Y., Shi, S., et al. (2015). ycf1, the Most Promising Plastid DNA Barcode of Land Plants. *Sci. Rep.* 5, 8348. doi:10.1038/srep08348
- Dong, W., Xu, C., Li, W., Xie, X., Lu, Y., Liu, Y., et al. (2017). Phylogenetic Resolution in *Juglans* Based on Complete Chloroplast Genomes and Nuclear DNA Sequences. *Front. Plant Sci.* 8, 1148. doi:10.3389/fpls.2017.01148
- Dong, W., Xu, C., Wu, P., Cheng, T., Yu, J., Zhou, S., et al. (2018). Resolving the Systematic Positions of Enigmatic Taxa: Manipulating the Chloroplast Genome Data of Saxifragales. *Mol. Phylogenet. Evol.* 126, 321–330. doi:10.1016/j.jmpev.2018.04.033
- Excoffier, L., and Lischer, H. E. L. (2010). Arlequin Suite Ver 3.5: a New Series of Programs to Perform Population Genetics Analyses under Linux and Windows. *Mol. Ecol. Resour.* 10 (3), 564–567. doi:10.1111/j.1755-0998.2010.02847.x
- Garran, T. A. (2020). A Comparative Study of *Leonurus Cardiaca* and *Leonurus Japonicus*. Doctoral Dissertation. China Academy of Chinese Medical Sciences.
- Garran, T. A., Ji, R., Chen, J.-L., Xie, D., Guo, L., Huang, L.-Q., et al. (2019). Elucidation of Metabolite Isomers of *Leonurus Japonicus* and *Leonurus Cardiaca* Using Discriminating Metabolite Isomerism Strategy Based on Ultra-high Performance Liquid Chromatography Tandem Quadrupole Time-Of-Flight Mass Spectrometry. *J. Chromatogr. A* 1598, 141–153. doi:10.1016/j.chroma.2019.03.059
- Hu, G., Cheng, L., Huang, W., Cao, Q., Zhou, L., Jia, W., et al. (2020). Chloroplast Genomes of Seven Species of Coryloideae (Betulaceae): Structures and Comparative Analysis. *Genome* 63 (7), 337–348. doi:10.1139/gen-2019-0153
- Hu, Y., Woeste, K. E., and Zhao, P. (2017). Completion of the Chloroplast Genomes of Five Chinese *Juglans* and Their Contribution to Chloroplast Phylogeny. *Front. Plant Sci.* 7, 1955. doi:10.3389/fpls.2016.01955
- Huang, D. I., and Cronk, Q. C. (2015). Plann: A Command-Line Application for Annotating Plastome Sequences. *Appl. Plant Sci.* 3 (8), 1500026. doi:10.3732/apps.1500026
- Jansen, R. K., Cai, Z., Raubeson, L. A., Daniell, H., Depamphilis, C. W., Leebens-Mack, J., et al. (2007). Analysis of 81 Genes from 64 Plastid Genomes Resolves Relationships in Angiosperms and Identifies Genome-Scale Evolutionary Patterns. *Proc. Natl. Acad. Sci.* 104 (49), 19369–19374. doi:10.1073/pnas.0709121104
- Kalyanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., and Jermini, L. S. (2017). ModelFinder: Fast Model Selection for Accurate Phylogenetic Estimates. *Nat. Methods* 14 (6), 587–589. doi:10.1038/nmeth.4285
- Katoh, K., Rozewicki, J., and Yamada, K. D. (2019). MAFFT Online Service: Multiple Sequence Alignment, Interactive Sequence Choice and Visualization. *Brief Bioinform.* 20 (4), 1160–1166. doi:10.1093/bib/bbx108
- Katoh, K., and Standley, D. M. (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
- 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
- Khadivi-Khub, A., and Soorni, A. (2014). Comprehensive Genetic Discrimination of *Leonurus Cardiaca* Populations by AFLP, ISSR, RAPD and IRAP Molecular Markers. *Mol. Biol. Rep.* 41 (6), 4007–4016. doi:10.1007/s11033-014-3269-4
- Krawczyk, K., Nobis, M., Myszczynski, K., Klichowska, E., and Sawicki, J. (2018). Plastid Super-barcodes as a Tool for Species Discrimination in Feather Grasses (Poaceae: *Stipa*). *Sci. Rep.* 8 (1), 1924. doi:10.1038/s41598-018-20399-w
- Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., and Mega, X. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* 35 (6), 1547–1549. doi:10.1093/molbev/msy096
- Leigh, J. W., and Bryant, D. (2015). PopART: Full-Feature Software for Haplotype Network Construction. *Methods Ecol. Evol.* 6 (9), 1110. doi:10.1111/2041-210x.12410
- Leonie, D., Barbara, G., Youri, L., Yavuz, A., Thomas, C., and Klaas, V. (2011). The Complete Chloroplast Genome of 17 Individuals of Pest Species *Jacobaea Vulgaris*: SNPs, Microsatellites and Barcoding Markers for Population and Phylogenetic Studies. *DNA Res.* 18 (2), 93–105. doi:10.1093/dnares/dsr002

- Librado, P., and Rozas, J. (2009). DnaSP V5: a Software for Comprehensive Analysis of DNA Polymorphism Data. *Bioinformatics* 25 (11), 1451–1452. doi:10.1093/bioinformatics/btp187
- Liu, H., Su, Z., Yu, S., Liu, J., Yin, X., Zhang, G., et al. (2019). Genome Comparison Reveals Mutation Hotspots in the Chloroplast Genome and Phylogenetic Relationships of *Ormosia* Species. *Biomed. Res. Int.* 2019, 7265030. doi:10.1155/2019/7265030
- Liu, L., Wang, Y., He, P., Li, P., Lee, J., Soltis, D. E., et al. (2018). Chloroplast Genome Analyses and Genomic Resource Development for Epilithic Sister Genera *Oresitrophe* and *Mukdenia* (Saxifragaceae), Using Genome Skimming Data. *BMC Genom.* 19 (1), 235. doi:10.1186/s12864-018-4633-x
- Lohse, M., Drechsel, O., Bock, R., and OrganellarGenomeDRAW (OGDRAW) (2007). OrganellarGenomeDRAW (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. doi:10.1007/s00294-007-0161-y
- Lukas, B., and Novak, J. (2013). The Complete Chloroplast Genome of *Origanum Vulgare* L. (Lamiaceae). *Gene* 528 (2), 163–169. doi:10.1016/j.gene.2013.07.026
- McDonald, M. J., Wang, W.-C., Huang, H.-D., and Leu, J.-Y. (2011). Clusters of Nucleotide Substitutions and Insertion/deletion Mutations Are Associated with Repeat Sequences. *Plos Biol.* 9 (6), e1000622. doi:10.1371/journal.pbio.1000622
- McPherson, H., van der Merwe, M., Delaney, S. K., Edwards, M. A., Henry, R. J., McIntosh, E., et al. (2013). Capturing Chloroplast Variation for Molecular Ecology Studies: a Simple Next Generation Sequencing Approach Applied to a Rainforest Tree. *BMC Ecol.* 13, 8. doi:10.1186/1472-6785-13-8
- Meng, J., Li, X., Li, H., Yang, J., Wang, H., and He, J. (2018). Comparative Analysis of the Complete Chloroplast Genomes of Four *Aconitum* Medicinal Species. *Molecules* 23 (5), 1015. doi:10.3390/molecules23051015
- Miller, A. J., and Gross, B. L. (2011). From forest to Field: Perennial Fruit Crop Domestication. *Am. J. Bot.* 98 (9), 1389–1414. doi:10.3732/ajb.1000522
- Muraguri, S., Xu, W., Chapman, M., Muchugi, A., Oluwaniyi, A., Oyeibanji, O., et al. (2020). Intraspecific Variation within Castor Bean (*Ricinus communis* L.) Based on Chloroplast Genomes. *Ind. Crops Prod.* 155, 112779. doi:10.1016/j.indcrop.2020.112779
- National Commission of Chinese Pharmacopoeia (2015). *Pharmacopoeia of the People's Republic of China*. Beijing: China Medical Science Press.
- Nevo, E. (2001). "Genetic Diversity," in *Encyclopedia of Biodiversity*. Beijing: Elsevier Inc., Vol. 22, 195–213. doi:10.1016/b0-12-226865-2/00137-1
- Nock, C. J., Hardner, C. M., Montenegro, J. D., Ahmad Termizi, A. A., Hayashi, S., Playford, J., et al. (2019). Wild Origins of Macadamia Domestication Identified through Intraspecific Chloroplast Genome Sequencing. *Front. Plant Sci.* 10, 334. doi:10.3389/fpls.2019.00334
- Nock, C. J., Waters, D. L. E., Edwards, M. A., Bowen, S. G., Rice, N., Cordeiro, G. M., et al. (2011). Chloroplast Genome Sequences from Total DNA for Plant Identification. *Plant Biotechnol. J.* 9 (3), 328–333. doi:10.1111/j.1467-7652.2010.00558.x
- Oliver, M. J., Murdock, A. G., Mishler, B. D., Kuehl, J. V., Boore, J. L., Mandoli, D. F., et al. (2010). Chloroplast Genome Sequence of the moss *Tortula Ruralis*: Gene Content, Polymorphism, and Structural Arrangement Relative to Other green Plant Chloroplast Genomes. *BMC Genom.* 11, 143. doi:10.1186/1471-2164-11-143
- Paweł, M., Barbara, G., Jolanta, M., and Andrzej, J. (2014). Joachimiak: Taxonomic Individuality of *Leonurus Cardiaca* and *Leonurus Quinquelobatus* in View of Morphological and Molecular Studies. *Plant Syst. Evol.* 300, 255–261. doi:10.1007/s00606-013-0878-7
- Pitschmann, A., Waschulin, C., Sykora, C., Purevsuren, S., and Glasl, S. (2017). Microscopic and Phytochemical Comparison of the Three *Leonurus* Species *L. Cardiaca*, *L. Japonicus*, and *L. Sibiricus*. *Planta Med.* 83, 1233–1241. doi:10.1055/s-0043-118034
- Qian, J., Song, J., Gao, H., Zhu, Y., Xu, J., Pang, X., et al. (2013). The Complete Chloroplast Genome Sequence of the Medicinal Plant *Salvia Miltiorrhiza*. *PLoS One* 8 (2), e57607. doi:10.1371/journal.pone.0057607
- Qiao, J., Cai, M., Yan, G., Wang, N., Li, F., Chen, B., et al. (2016). High-throughput Multiplex cpDNA Resequencing Clarifies the Genetic Diversity and Genetic Relationships Among Brassica Napus, Brassica Rapa and Brassica oleracea. *Plant Biotechnol. J.* 14, 409–418. doi:10.1111/pbi.12395
- Rambaut, A. (2002). Se-AL: Sequence Alignment Editor. Version 2.0a11. Available at: <http://tree.bio.ed.ac.uk/software/seal/>.
- Ravi, V., Khurana, J. P., and Tyagi, A. K. P. (2008). An Update on Chloroplast Genomes. *Plant Syst. Evol.* (271), 101–122. doi:10.1007/s00606-007-0608-0
- Ren, T., Li, Z.-X., Xie, D.-F., Gui, L.-J., Peng, C., Wen, J., et al. (2020). Plastomes of Eight *Ligusticum* Species: Characterization, Genome Evolution, and Phylogenetic Relationships. *BMC Plant Biol.* 20 (1), 519. doi:10.1186/s12870-020-02696-7
- Ross, T. G., Barrett, C. F., Soto Gomez, M., Lam, V. K. Y., Henriquez, C. L., Les, D. H., et al. (2016). Plastid Phylogenomics and Molecular Evolution of Alismatales. *Cladistics* 32 (2), 160–178. doi:10.1111/cla.12133
- Sancho, R., Cantalapiedra, C. P., López-Alvarez, D., Gordon, S. P., Vogel, J. P., Catalán, P., et al. (2017). Comparative Plastome Genomics and Phylogenomics of Brachypodium: Flowering Time Signatures, Introgression and Recombination in Recently Diverged Ecotypes. *New Phytol.* 218 (4), 1631–1644. doi:10.1111/nph.14926
- Shikov, A. N., Pozharitskaya, O. N., Makarov, V. G., Demchenko, D. V., and Shikh, E. V. (2011). Effect of *Leonurus Cardiaca* Oil Extract in Patients with Arterial Hypertension Accompanied by Anxiety and Sleep Disorders. *Phytother. Res.* 25 (4), 540–543. doi:10.1002/ptr.3292
- Sinn, B. T., Sedmak, D. D., Kelly, L. M., and Freudenstein, J. V. (2018). Total Duplication of the Small Single Copy Region in the Angiosperm Plastome: Rearrangement and Inverted Repeat Instability in *Asarum*. *Am. J. Bot.* 105 (1), 71–84. doi:10.1002/ajb2.1001
- Souza, U. J. B. d., Nunes, R., Targueta, C. P., Diniz-Filho, J. A. F., and Telles, M. P. d. C. (2019). The Complete Chloroplast Genome of *Stryphnodendron Adstringens* (Leguminosae - Caesalpinioideae): Comparative Analysis with Related Mimosoid Species. *Sci. Rep.* 9 (1), 14206. doi:10.1038/s41598-019-50620-3
- Stamatakis, A. (2014). RAXML Version 8: a Tool for Phylogenetic Analysis and post-analysis of Large Phylogenies. *Bioinformatics* 30 (9), 1312–1313. doi:10.1093/bioinformatics/btu033
- Stefan, K., Choudhuri, J. V., Enno, O., Chris, S., Jens, S., and Robert, G. (2001). REPputer: the Manifold Applications of Repeat Analysis on a Genomic Scale. *Nucleic Acids Res.* (22), 4633–4642. doi:10.1093/nar/29.22.4633
- Sun, J., Wang, Y., Liu, Y., Xu, C., Yuan, Q., Guo, L., et al. (2020). Evolutionary and Phylogenetic Aspects of the Chloroplast Genome of *Chaenomeles* Species. *Sci. Rep.* 10 (1), 11466. doi:10.1038/s41598-020-67943-1
- Upton, R. (2018). *Leonurus Cardiaca* L., *Leonurus Quinquelobatus* Gilb. *Monograph and Therapeutic Compendium*. Beijing: American Herbal Pharmacopoeia, 1.
- Wang, M., Wang, X., Sun, J., Wang, Y., Ge, Y., Dong, W., et al. (2021). Phylogenomic and Evolutionary Dynamics of Inverted Repeats across *Angelica* Plastomes. *BMC Plant Biol.* 21 (1), 26. doi:10.1186/s12870-020-02801-w
- Wang, X., Liu, X.-Q., Ko, Y.-Z., Jin, X.-L., Sun, J.-H., Zhao, Z.-Y., et al. (2020). Genetic Diversity and Phylogeography of the Important Medical Herb, Cultivated Huang-Lian Populations, and the Wild Relatives *Coptis* Species in China. *Front. Genet.* 11, 708. doi:10.3389/fgene.2020.00708
- Wang, X., and Wang, L. (2016). GMATA: An Integrated Software Package for Genome-Scale SSR Mining, Marker Development and Viewing. *Front. Plant Sci.* 7, 1350. doi:10.3389/fpls.2016.01350
- Wang, Y., Wang, S., Liu, Y., Yuan, Q., Sun, J., and Guo, L. (2021). Chloroplast Genome Variation and Phylogenetic Relationships of *Atractylodes* Species. *BMC Genom.* 22 (1), 103. doi:10.1186/s12864-021-07394-8
- Wicke, S., Schneeweiss, G. M., dePamphilis, C. W., Müller, K. F., and Quandt, D. (2011). The Evolution of the Plastid Chromosome in Land Plants: Gene Content, Gene Order, Gene Function. *Plant Mol. Biol.* 76 (3-5), 273–297. doi:10.1007/s11103-011-9762-4
- Yi, X., Gao, L., Wang, B., Su, Y.-J., and Wang, T. (2013). The Complete Chloroplast Genome Sequence of *Cephalotaxus Oliveri* (Cephalotaxaceae): Evolutionary Comparison of *Cephalotaxus* Chloroplast DNAs and Insights into the Loss of Inverted Repeat Copies in Gymnosperms. *Genome Biol. Evol.* 5 (4), 688–698. doi:10.1093/gbe/evt042
- Yuan, Q. J., Zhang, Z. Y., Hu, J., Guo, L. P., Shao, A. J., and Huang, L. Q. (2010). Impacts of Recent Cultivation on Genetic Diversity Pattern of a Medicinal Plant, *Scutellaria Baicalensis* (Lamiaceae). *BMC Genet.* 11 (1), 29–13. doi:10.1186/1471-2156-11-29
- Zhang, R. H., Liu, Z. K., Yang, D. S., Zhang, X. J., Sun, H. D., and Xiao, W. L. (2018). Phytochemistry and Pharmacology of the Genus *Leonurus*: The Herb to Benefit

- the Mothers and More. *Phytochemistry* 147, 167–183. doi:10.1016/s0031-9422(18)30015-3
- Zhang, R., Wang, Y.-H., Jin, J.-J., Stull, G. W., Bruneau, A., Cardoso, D., et al. (2020). Exploration of Plastid Phylogenomic Conflict Yields New Insights into the Deep Relationships of Leguminosae. *Syst. Biol.* 69 (4), 613–622. doi:10.1093/sysbio/syaa013
- Zhang, S. D., Jin, J. J., Chen, S. Y., Chase, M. W., Soltis, D. E., Li, H. T., et al. (2017). Diversification of Rosaceae since the Late Cretaceous Based on Plastid Phylogenomics. *New Phytol.* 214 (3), 1355–1367. doi:10.1111/nph.14461
- Zhang, X.-F., Landis, J. B., Wang, H.-X., Zhu, Z.-X., and Wang, H.-F. (2021). Comparative Analysis of Chloroplast Genome Structure and Molecular Dating in *Myrtales*. *BMC Plant Biol.* 21 (1), 219. doi:10.1186/s12870-021-02985-9
- Zhao, T., Wang, G., Ma, Q., Liang, L., and Yang, Z. (2019). Multilocus Data Reveal Deep Phylogenetic Relationships and Intercontinental Biogeography of the Eurasian-North American Genus *Corylus* (Betulaceae). *Mol. Phylogenet. Evol.* 142, 106658. doi:10.1016/j.ympev.2019.106658
- Conflict of Interest:** The authors 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.
- Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.
- Copyright © 2021 Sun, Wang, Garran, Qiao, Wang, Yuan, Guo and Huang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.