



# Identification of *Ligularia* Herbs Using the Complete Chloroplast Genome as a Super-Barcode

Xinlian Chen<sup>1†</sup>, Jianguo Zhou<sup>1†</sup>, Yingxian Cui<sup>1</sup>, Yu Wang<sup>1</sup>, Baozhong Duan<sup>2</sup> and Hui Yao<sup>1\*</sup>

<sup>1</sup> Key Lab of Chinese Medicine Resources Conservation, State Administration of Traditional Chinese Medicine of the People's Republic of China, Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China, <sup>2</sup> College of Pharmaceutical Science, Dali University, Dali, China

## OPEN ACCESS

### Edited by:

Jiang Xu,  
China Academy of Chinese Medical  
Sciences, China

### Reviewed by:

Zhi Chao,  
Southern Medical University, China  
Eszter Virag,  
University of Pannonia, Hungary

### \*Correspondence:

Hui Yao  
scauyaoh@sina.com

<sup>†</sup>These authors have contributed  
equally to this work.

### Specialty section:

This article was submitted to  
Ethnopharmacology,  
a section of the journal  
Frontiers in Pharmacology

**Received:** 27 April 2018

**Accepted:** 08 June 2018

**Published:** 03 July 2018

### Citation:

Chen X, Zhou J, Cui Y, Wang Y,  
Duan B and Yao H (2018)  
Identification of *Ligularia* Herbs Using  
the Complete Chloroplast Genome as  
a Super-Barcode.

Front. Pharmacol. 9:695.  
doi: 10.3389/fphar.2018.00695

More than 30 *Ligularia* Cass. (Asteraceae) species have long been used in folk medicine in China. Morphological features and common DNA regions are both not ideal to identify *Ligularia* species. As some *Ligularia* species contain pyrrolizidine alkaloids, which are hazardous to human and animal health and are involved in metabolic toxification in the liver, it is important to find a better way to distinguish these species. Here, we report complete chloroplast (CP) genomes of six *Ligularia* species, *L. intermedia*, *L. jaluensis*, *L. mongolica*, *L. hodgsonii*, *L. veitchiana*, and *L. fischeri*, obtained through high-throughput Illumina sequencing technology. These CP genomes showed typical circular tetrameric structure and their sizes range from 151,118 to 151,253 bp. The GC content of each CP genome is 37.5%. Every CP genome contains 134 genes, including 87 protein-coding genes, 37 tRNA genes, eight rRNA genes, and two pseudogenes (*ycf1* and *rps19*). From the mVISTA, there were no potential coding or non-coding regions to distinguish these six *Ligularia* species, but the maximum likelihood tree of the six *Ligularia* species and other related species showed that the whole CP genome can be used as a super-barcode to identify these six *Ligularia* species. This study provides invaluable data for species identification, allowing for future studies on phylogenetic evolution and safe medical applications of *Ligularia*.

**Keywords:** *Ligularia* Cass., chloroplast genome, identification, super-barcode, Illumina sequencing

## INTRODUCTION

*Ligularia* Cass., belonging to the Senecioneae tribe of Asteraceae, comprises about 140 species of perennial herbs. These species are distributed in Asia and Europe, with a total of 123 species distributed in China, 89 of which are endemic (Liu and Illarionova, 1989). In China, *Ligularia* species are mainly distributed in mountainous areas in the southwest (Liu and Illarionova, 1989) and more than 30 *Ligularia* species have long been used in folk medicine (Wang, 2007). The roots, stems, leaves, and flowers of them contain various chemical compounds, such as

**Abbreviations:** IR, inverted repeat; LSC, large single-copy; ML, maximum likelihood; SSC, small single-copy; SSRs, simple sequence repeats.

sesquiterpenes (Wang, 2007; Shimizu et al., 2014; Saito et al., 2017) and alkaloids (Asada et al., 1981; Feng, 2016). They are used as herbal medicines for the treatment of bronchitis, coughing, pulmonary tuberculosis, and hemoptysis. These herbal medicines are usually used as substitutes for *Asteris Radix* et *Rhizoma* which originates from *Aster tataricus* L. and is recorded in the Chinese Pharmacopoeia (Lin and Liu, 1989; Chinese Pharmacopoeia Commission, 2015). Approximately, 3% of flowering plants (as many as 6,000 species), including *Ligularia* species (Smith and Culvenor, 1981; Stegelmeier et al., 1999), contain pyrrolizidine alkaloids (PAs). It has been reported that various *Ligularia* species contain PAs, including *L. japonica* (Asada et al., 1981), *L. wilsoniana* (Xiong et al., 2016), *L. duciformis*, *L. intermedia*, *L. hodgsonii*, and *L. veitchiana* (Pu et al., 2004). PAs are phytoalexins that function in plant defense systems against herbivores, insects, and plant pathogens. However, they are harmful to human and animal health (Jank and Rath, 2017; Martinello et al., 2017), as they are involved in metabolic toxicification in the liver caused by PA poisoning (Bull et al., 1968; Prakash et al., 1999). The German Federal Department of Health stated that the safe total daily dose of PA is less than 1 µg, and doctors do not allow continuous administration of drugs with PA for more than 6 weeks. In addition, all PA-containing products are banned in Australia (Wiedenfeld and Edgar, 2011).

*Ligularia* has been traditionally classified based on morphological structures, such as the arrangement of inflorescences, leaf shape, leaf veins, and phyllaries (Liu and Illarionova, 1989). Interspecific hybridization of *Ligularia* species is common and their morphological variation is complicated (Hanai et al., 2012; Yu et al., 2014; Saito et al., 2017), making it difficult to correctly identify species. Common DNA barcoding sequences (ITS, *matK*, *psbA-trnH*, and *rbcL*) are also not ideal for identifying *Ligularia* species (He and Pan, 2015). Recently, researchers have screened sequences from the whole chloroplast (CP) genome from numerous plant taxa, such as *Juglans* L. plants and bamboo (Zhang et al., 2011; Hu et al., 2016), or use CP genome as a super-barcode to distinguish species (Xia et al., 2016). The CP genome is highly conserved in plants regardless of the size, structure, or gene content (Tonti-Filippini et al., 2017), and the majority of the retained core genes are involved in the light reactions of photosynthesis or in functions related to transcription or translation (Sato et al., 1999). The CP genome map is a circular DNA molecule that includes a SSC region, a LSC region, and two inverted-repeat (IRa and IRb) regions (Sato et al., 1999). Several CP genomes from Asteraceae have previously been reported, including CP genomes from *Aster* (Choi and Park, 2015), *Ambrosia* (Nagy et al., 2017), *Carthamus* (Lu et al., 2015), and *Taraxacum* (Salih et al., 2017). However, only one CP genome from *Ligularia*, for *L. fischeri*, has previously been published (Lee et al., 2016). In this study, we report the CP genomes of six *Ligularia* species, *L. intermedia*, *L. jaluensis*, *L. mongolica*, *L. hodgsonii*, *L. veitchiana*, and *L. fischeri*, obtained through high-performance Illumina sequencing technology. Our aim is to use the CP genome as a super-barcode for the identification of *Ligularia* species to provide invaluable genetic information for future studies.

## MATERIALS AND METHODS

### Plant Materials and DNA Extraction

Fresh leaves of *L. intermedia* and *L. fischeri* were collected from Baishan City and Tonghua City, Jilin Province, China, respectively. Fresh leaves of *L. jaluensis* and *L. mongolica* were collected from Yanbian Korean Autonomous Prefecture, Jilin Province. These four species were identified by Prof. Junlin Yu from Tonghua Normal University, Jilin. Fresh leaves of *L. hodgsonii* and *L. veitchiana* were collected from Enshi Tujia and Miao Autonomous Prefecture, Hubei Province, and the Qinling Mountains, Shaanxi Province, respectively. These two samples were identified by Prof. Yulin Lin from the Institute of Medicinal Plant Development (IMPLAD), Chinese Academy of Medical Sciences (CAMS), and Peking Union Medical College (PUMC). The exact GPS coordinates for the collection locations of six *Ligularia* species are listed in Supplementary Table S1. Voucher specimens were deposited in the herbarium at IMPLAD. Collected fresh leaves were stored in a  $-80^{\circ}\text{C}$  freezer until further use. DNA extraction was performed using a DNeasy Plant Mini Kit (Qiagen Co., Germany) following the manufacturer's protocol.

### Illumina Sequencing and Genome Assembly

Approximately 5–10 µg of high-quality DNA were used to build shotgun libraries with insert sizes of 500 bp and were sequenced in accordance with the protocol for Illumina HiSeq X technology. The total raw data of the six species were produced with 150 bp paired-end read lengths. The software Trimmomatic (Bolger et al., 2014) was employed to filter low-quality reads from the raw data. After filtering for quality sequences, the remaining clean reads were used to assemble the CP genome sequences. The CP sequences of all plants downloaded from the National Center for Biotechnology Information (NCBI) were used to create a reference database. Then, the clean sequences were mapped to the database and the mapped reads were extracted on the basis of coverage and similarity. The extracted reads were assembled into contigs using SOAPdenovo2 (Luo et al., 2012). The scaffold of the CP genome was constructed using SSPACE (Boetzer et al., 2011), and the gaps were filled using GapFiller (Boetzer and Pirovano, 2012).

### Validation, Annotation, and Sequence Submission

The accuracy of the assembly of the four boundaries (SSC, LSC, IRa, and IRb regions) of the CP sequences was confirmed through PCR and Sanger sequencing using validated primers (Supplementary Table S2). The thermocycler conditions for the PCR were as follows:  $94^{\circ}\text{C}$  for 5 min;  $94^{\circ}\text{C}$  for 30 s,  $56^{\circ}\text{C}$  for 30 s,  $72^{\circ}\text{C}$  for 1.5 min, and 32 cycles;  $72^{\circ}\text{C}$  for 10 min. The online programs Dual Organellar GenoMe Annotator (DOGMA) (Wyman et al., 2004) and CPGAVAS (Liu et al., 2012) were used for the initial annotation of the CP genomes of the six species, followed by manual correction. The complete data from the study were submitted to NCBI under the BioProject ID PRJNA400300

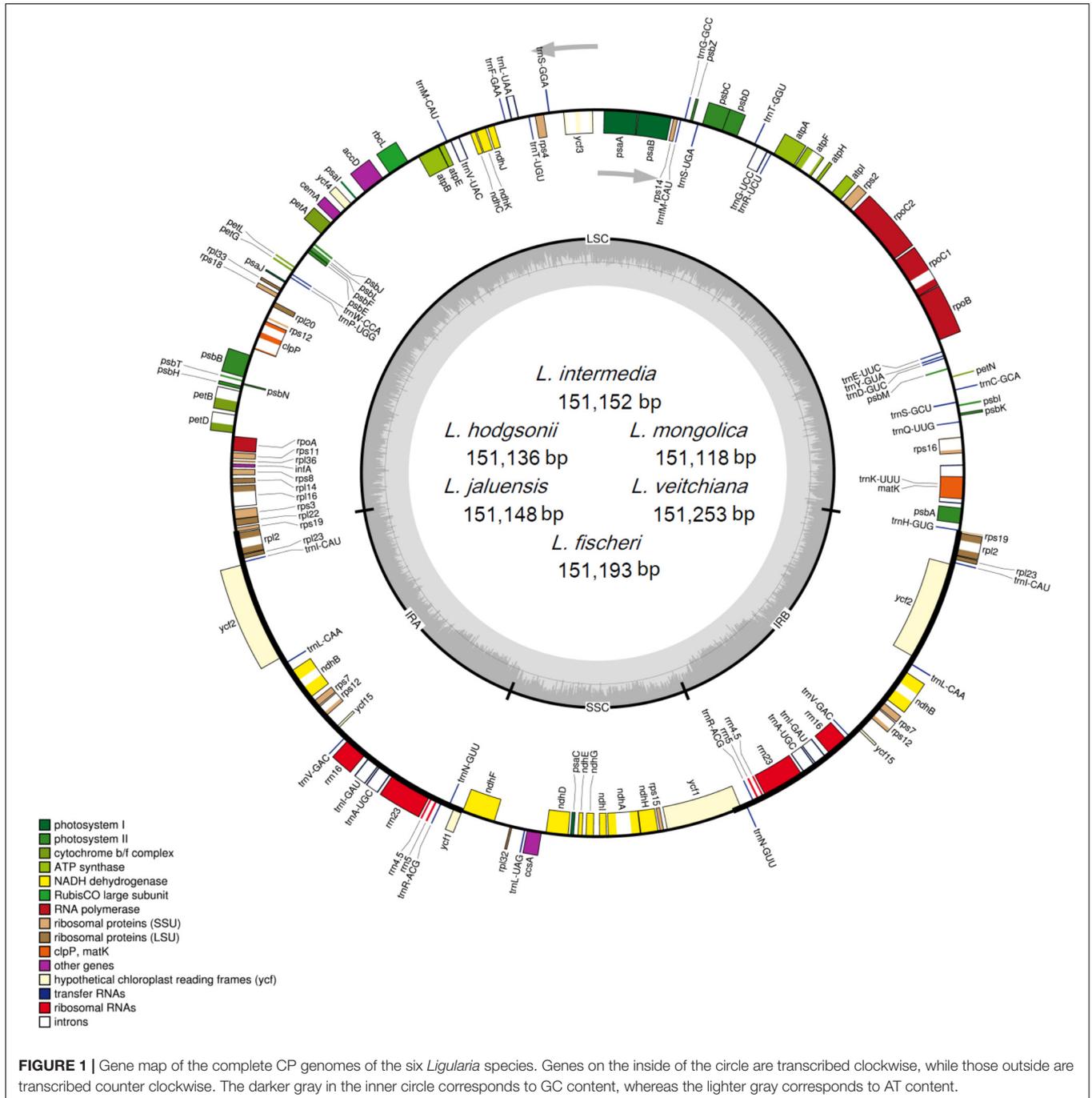
and BioSample ID SAMN07562669. The assembled complete CP genome sequences of the six *Ligularia* species were submitted to NCBI GenBank with the accession numbers MF539929-MF539933, and MG729822.

### Genome Structure Analysis

The software tRNAscan-SE (Schattner et al., 2005) and DOGMA (Wyman et al., 2004) were used to identify tRNA genes. Gene maps were generated using Organellar Genome DRAW v1.2 (Lohse et al., 2007) with default settings and then the gene maps

were checked manually. MEGA 6.0 was used to calculate the GC content (Tamura et al., 2013). REPuter (University of Bielefeld, Bielefeld, Germany) (Kurtz et al., 2001) was used to identify the size and location of repeat sequences in the CP genomes of the six *Ligularia* species. We used the MISA software Misa-Microsatellite Identification Tool, 2017<sup>1</sup> to detect SSRs with the parameter settings the same as those described in Li et al. (2013). All the repeated sequences were manual verified and excess data

<sup>1</sup><http://pgrc.ipk-gatersleben.de/misa/>



**TABLE 1** | Summary statistics for assembly of the six CP genomes of *Ligularia* species.

Species names	<i>L. intermedia</i>	<i>L. hodgsonii</i>	<i>L. jaluensis</i>	<i>L. mongolica</i>	<i>L. veitchiana</i>	<i>L. fischeri</i>
Raw reads	53,871,796	42,199,336	43,781,502	37,478,124	36,990,248	36,684,984
Clean reads	52,907,406	41,086,606	42,940,766	36,283,722	35,813,466	35,451,926
Mapped CP reads	434,480	623,388	516,254	378,838	278,424	361,650
Size (bp)	151,152	151,136	151,148	151,118	151,253	151,193
LSC length (bp)	83,258	83,253	83,263	83,244	83,330	83,301
SSC length (bp)	18,232	18,217	18,225	18,214	18,247	18,232
IR length (bp)	24,831	24,833	24,830	24,830	24,838	24,830
Coding (bp)	90,247	90,253	90,247	90,250	90,253	90,247
Non-coding (bp)	60,905	60,883	60,901	60,868	60,000	60,946

**TABLE 2** | List of genes found in the six CP genomes of *Ligularia* species.

No.	Group of genes	Gene names	Amount
1	Photosystem I	<i>psaA, psaB, psaC, psal, psaJ</i>	5
2	Photosystem II	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ</i>	15
3	Cytochrome b/f complex	<i>petA, petB*, petD*, petG, petL, petN</i>	6
4	ATP synthase	<i>atpA, atpB, atpE, atpF*, atpH, atpI</i>	6
5	NADH dehydrogenase	<i>ndhA*, ndhB* (×2), ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i>	11
6	RubisCO large subunit	<i>rbcL</i>	1
7	RNA polymerase	<i>rpoA, rpoB, rpoC1*, rpoC2</i>	4
8	Ribosomal proteins (SSU)	<i>rps2, rps3, rps4, rps7 (×2), rps8, rps11, rps12** (×2), rps14, rps15, rps16*, rps18, rps19</i>	12
9	Ribosomal proteins (LSU)	<i>rpl2* (×2), rpl14, rpl16*, rpl20, rpl22, rpl23 (×2), rpl32, rpl33, rpl36</i>	9
10	Other genes	<i>accD, clpP**, matK, ccsA, cemA, infA</i>	6
11	Proteins of unknown function	<i>ycf1, ycf2 (×2), ycf3**, ycf4, ycf15</i>	5
12	Transfer RNAs	37 tRNAs (6 contain an intron, 7 in the IRs)	
13	Ribosomal RNAs	<i>rrn4.5 (×2), rrn5 (×2), rrn16 (×2), rrn23 (×2)</i>	

One or two asterisks after genes indicate that gene contains one or two introns, respectively.

were removed. The distribution of codon usage was studied using CodonW with the relative synonymous codon usage (RSCU) ratio (Sharp and Li, 1987). The online program Predictive RNA Editor for Plants suite (Mower, 2009) with a cutoff value of 0.8 were used to predict RNA editing sites in the six CP genomes of *Ligularia* species.

## Phylogenetic Analysis

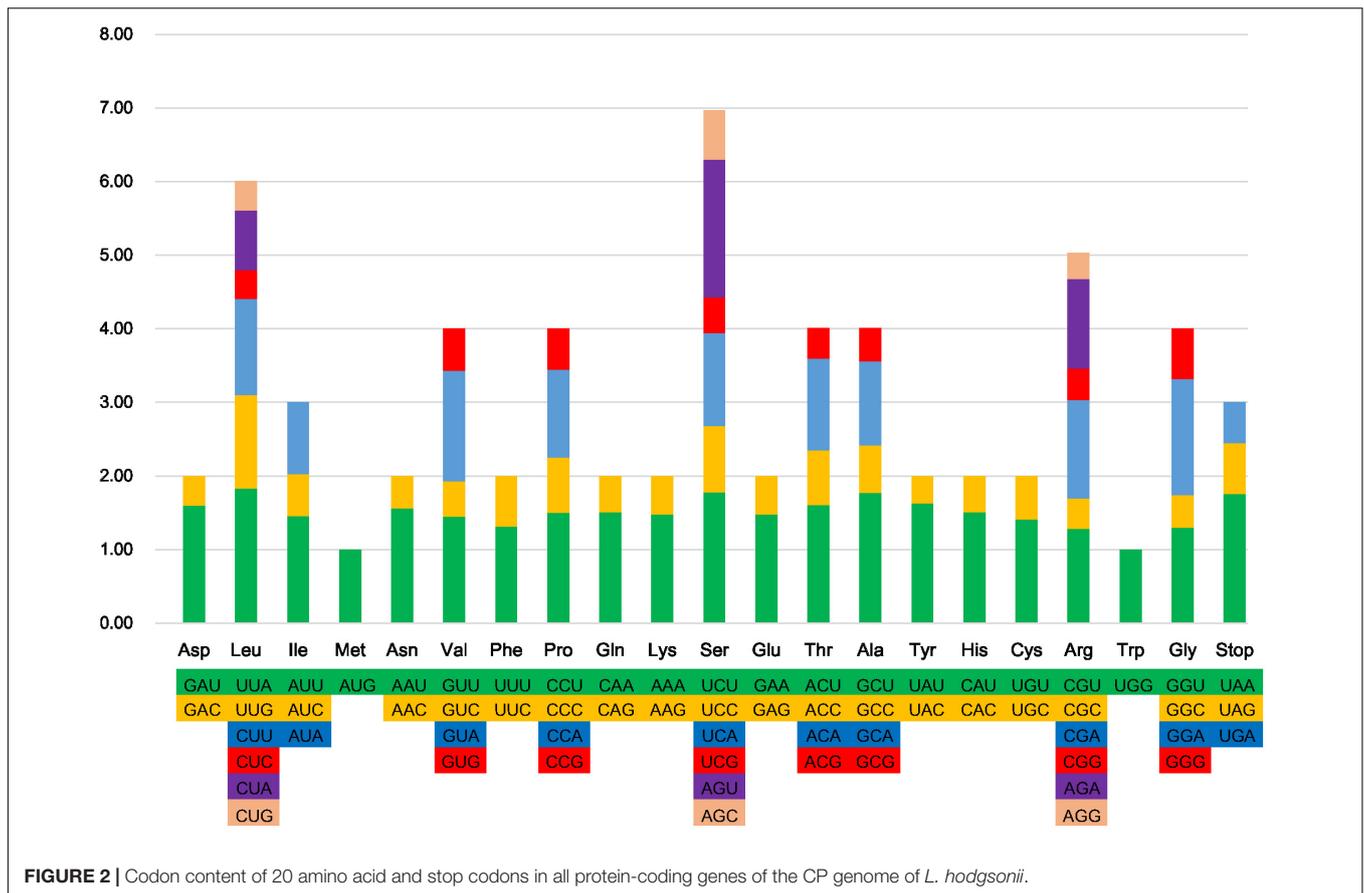
For identification purposes and to further phylogenetic research on this genus, we used mVISTA (Thompson et al., 1994) to compare six *Ligularia* species with *L. hodgsonii* as the reference genome. MEGA 6.0 was used to construct the phylogenetic tree with *Platycodon grandiflorus* and *Adenophora remotiflora* included as outgroups based on ML analysis. The details of the selected species (excluding the six *Ligularia* species) are presented in Supplementary Table S3.

## RESULTS AND DISCUSSION

### CP Genome Structure of Six *Ligularia* Species

The raw data from the six *Ligularia* species is 9.1 Gb for *L. intermedia*, 7.2 Gb for *L. hodgsonii*, 7.4 Gb for *L. jaluensis*, 6.4 Gb for *L. mongolica*, 6.3 Gb for *L. veitchiana*, and 6.2 Gb for

*L. fischeri*. The sizes of the six CP genomes range from 151,118 bp for *L. mongolica* to 151,253 bp for *L. veitchiana*, which are similar to other Asteraceae CP genomes (Liu et al., 2013; Salih et al., 2017; Wang et al., 2017; Zhang et al., 2017). The investigated genomes showed typical circular tetrameric structure, including an SSC region and an LSC region, separated by two IR regions (Figure 1). The corresponding lengths of the four regions from the six species are similar: the SSC lengths range from 18,214 to 18,247 bp, the LSC lengths range from 83,244 to 83,330 bp, and the IR lengths range from 24,830 to 24,838 bp (Table 1). The size of the previously published *L. fischeri* CP genome is 151,133 bp, and included an SSC region (18,233 bp), an LSC region (83,238 bp), and two IR regions (24,831 bp apart) (Lee et al., 2016). Our results showed that all six of the newly sequenced CP genomes have a GC content of 37.5%, which is lower than some Asteraceae species (Liu et al., 2013; Salih et al., 2017; Wang et al., 2017; Zhang et al., 2017). The GC content of four homologous regions of the six CP genomes is the same. However, the distribution of the GC content in each region is uneven. The GC content in the IR region is the largest (43.0%), followed by the LSC region (35.6%), and the region with the lowest GC content is the SSC region (30.7%). Our analysis showed that the high GC content in the IR region is attributed to four rRNA genes (*rrn16*, *rrn23*, *rrn4.5*, and *rrn5*). The AT content of the first, second, and third position of protein-coding genes in the six CP genomes are 54.5–54.6%, 61.9–62.0%,



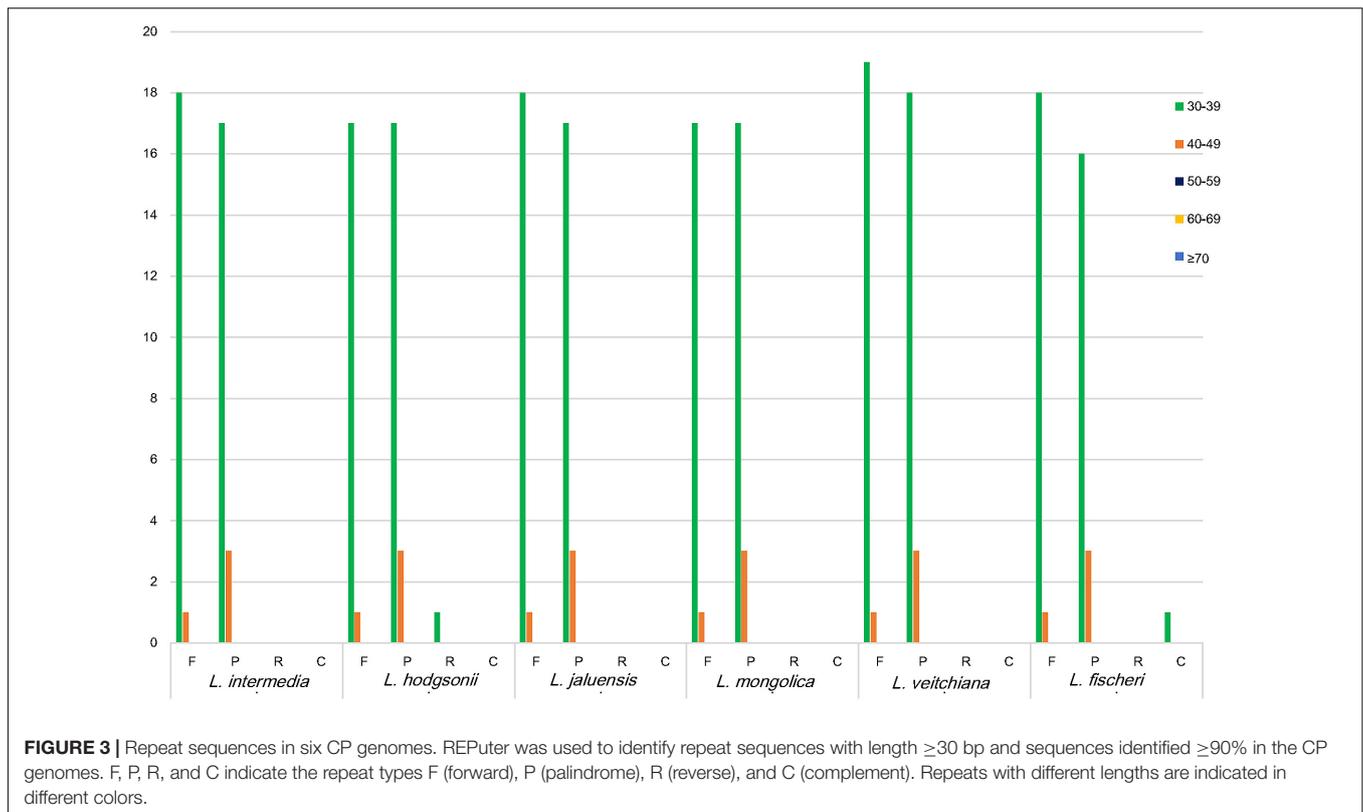
and 70.1%, respectively. The higher AT content in the third site has also been observed in other plants (Yi and Kim, 2012; He et al., 2017; Zhou et al., 2017) and is usually used to distinguish DNA of CP, nucleus, and mitochondria origin (Clegg et al., 1994).

Each of the six CP genomes contains 134 genes, including 87 protein-coding genes, 37 tRNA genes, eight rRNA genes, and two pseudogenes (*ycf1* and *rps19*; **Table 2**). Seven protein-coding genes (*ndhB*, *rp12*, *rp123*, *rps12*, *rps7*, *ycf15*, and *ycf2*), seven tRNAs (*trnA-UGC*, *trnI-CAU*, *trnI-GAU*, *trnL-CAA*, *trnN-GUU*, *trnR-ACG*, and *trnV-GAC*), and all of the rRNAs (*rrn16*, *rrn23*, *rrn4.5*, and *rrn5*) are duplicated in the IR regions, which is similar to *Artemisia annua* (Shen et al., 2017) and *Artemisia frigida* (Liu et al., 2013). The CP genomes of the six *Ligularia* species contain a small 3.4 kb inversion within a large 23 kb inversion in the LSC region, which is a unique feature in Asteraceae (Kim et al., 2005; Liu et al., 2013). The LSC region included 62 protein-coding genes and 22 tRNA genes. The SSC region included 11 protein-coding genes and one tRNA gene (*trnL-UAG*). The CP genomes of each of these six *Ligularia* species did not have an inverted SSC region, which has also been found in the CP genomes of *A. frigida* (Liu et al., 2013), *Scutellaria baicalensis* (Jiang et al., 2017), *Carthamus tinctorius* (Lu et al., 2015), and *Juglans L.* (Hu et al., 2016). In contrast, the SSC regions of *Helianthus annuus*, *Lactuca sativa* (Timme et al., 2007), and *Aster spathulifolius* (Choi and Park, 2015) are inverted. The functional *ycf1* copy is located in the IRb-SSC boundary and the pseudogene *ycf1* copy is located

in the IRa region. The functional *rps19* copy is on the boundary of LSC and IRa and the pseudogene *rps19* copy is on the IRb region. The coding region occupied 59.67–59.72% of the CP genomes of six *Ligularia* species, including protein-coding genes, tRNA genes, and rRNA genes. Meanwhile, non-coding regions, including introns, pseudogenes, and intergenic spacers occupied 40.28–40.33% of the CP genomes of the six *Ligularia* species.

## Codon Usage and RNA Editing Sites

All protein-coding genes in the six *Ligularia* CP genomes are composed of 26,136–26,138 codons. The most and least universal amino acids of the CP genomes of the six *Ligularia* species are leucine (10.8%) and cysteine (1.1%), respectively (**Figure 2**). This is also similar to the CP genome from artichoke (Curci et al., 2015). However, the most universal amino acid from *A. frigida* is isoleucine (Liu et al., 2013). The most and the least abundant amino acids in the *Taraxacum obtusifrons* and *Taraxacum amplus* CP genomes are serine and methionine (Salih et al., 2017), respectively. **Figure 2** shows that with the increase of specific amino acid codes the RSCU increases accordingly. Most of the amino acid codons have preferences, except for methionine and tryptophan. Potential RNA editing sites were predicted for 35 genes from the CP genomes of the six *Ligularia* species. Forty-eight RNA editing sites were identified. S to L of amino acid change appeared most frequently, while R to W and T to I occurred least. Each corresponding gene from



**TABLE 3 |** The SSR types of the six CP genomes of *Ligularia* species.

SSR type	Repeat unit	Amount					
		<i>L. intermedia</i>	<i>L. hodgsonii</i>	<i>L. jaluensis</i>	<i>L. mongolica</i>	<i>L. veitchiana</i>	<i>L. fischeri</i>
Mono	A/T	30	29	30	30	37	31
Di	AT/AT	7	6	7	6	7	7
Tri	AAG/CTT	2	2	2	2	2	2
Tri	AAT/ATT	2	2	2	2	3	2
Tri	ATC/ATG	1	1	1	1	1	1
Tetra	AAAG/CTTT	1	1	1	1	1	1
Tetra	AAAT/ATTT	6	6	6	6	6	6
Tetra	AACT/AGTT	1	1	1	1	1	1
Tetra	AATC/ATTG	1	1	1	1	1	1
Tetra	AATT/AATT	1	1	1	1	1	1
Tetra	AGAT/ATCT	1	1	1	1	1	1

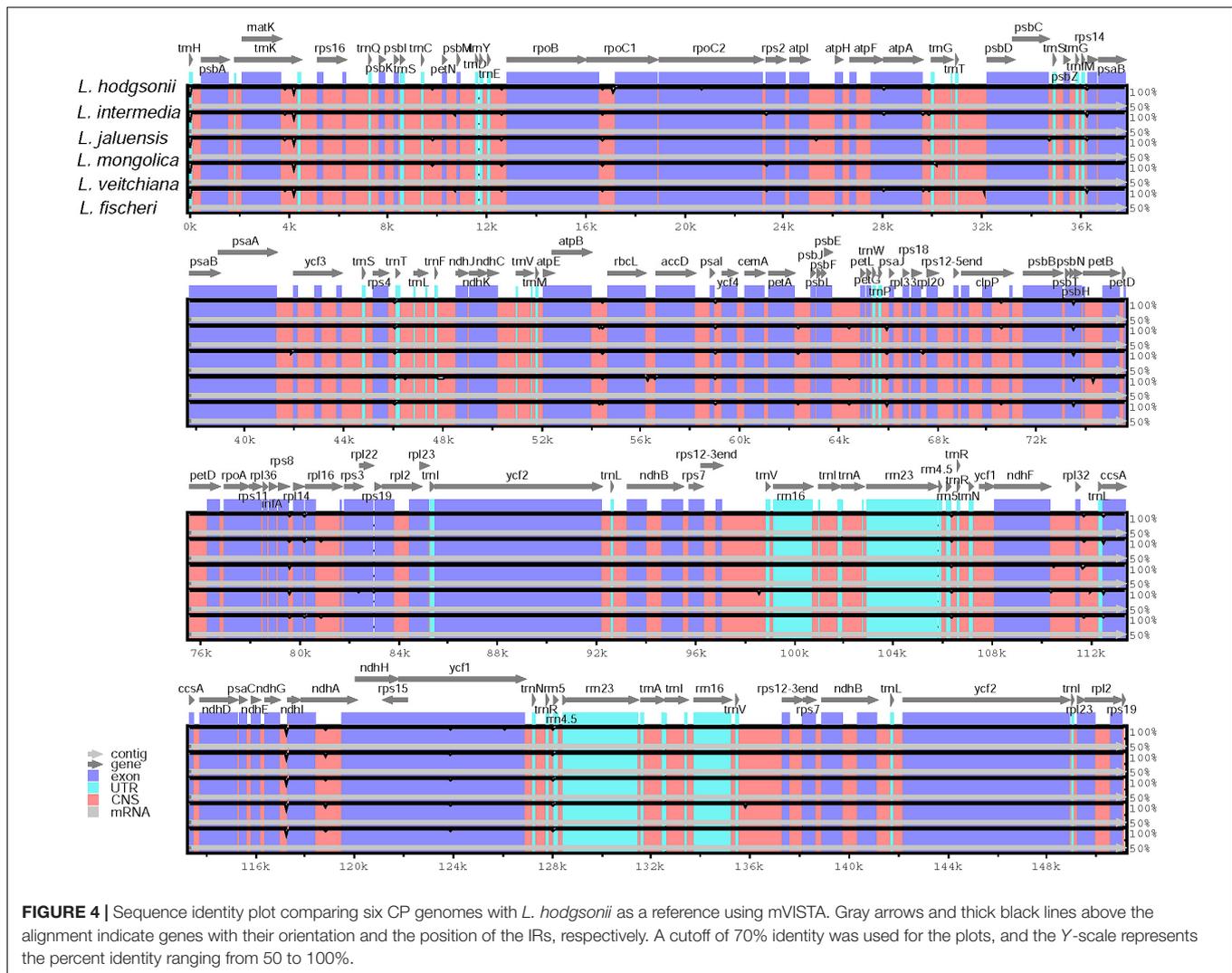
the RNA editing sites of the six *Ligularia* species is at the same nucleotide position (Supplementary Table S4).

A total of 18 genes containing introns, including 12 protein-coding genes (*atpF*, *clpP*, *ndhA*, *ndhB*, *petB*, *petD*, *rpl16*, *rpl2*, *rpoC1*, *rps12*, *rps16*, and *ycf3*), and six tRNA genes (*trnA-UGC*, *trnG-UCC*, *trnI-GAU*, *trnK-UUU*, *trnL-UAA*, and *trnV-UAC*; Supplementary Table S5), were identified in this study. Nine protein-coding genes contain only one intron and three protein-coding genes (*clpP*, *rps12*, and *ycf3*) contain two introns. All six tRNAs contain only one intron. *TrnK-UUU* has the longest intron (2,556 bp), which contains *matK*. The *clpP* gene and *ycf3* gene are both located in the LSC region. The *rps12* gene is a

special trans-splicing gene with the 5' exon located in the LSC region, but the 3' exon located in the IR region. This condition exists in many species, such as *A. frigida* (Liu et al., 2013), artichoke (Curci et al., 2015), and *Aster spathulifolius* (Choi and Park, 2015).

### Long Repeats and SSRs in the CP Genomes From the Six *Ligularia* Species

Repeat sequences, which are related to plastome organization (Salih et al., 2017), are mostly distributed in intergenic regions and intron regions, and only a small fraction is present in the



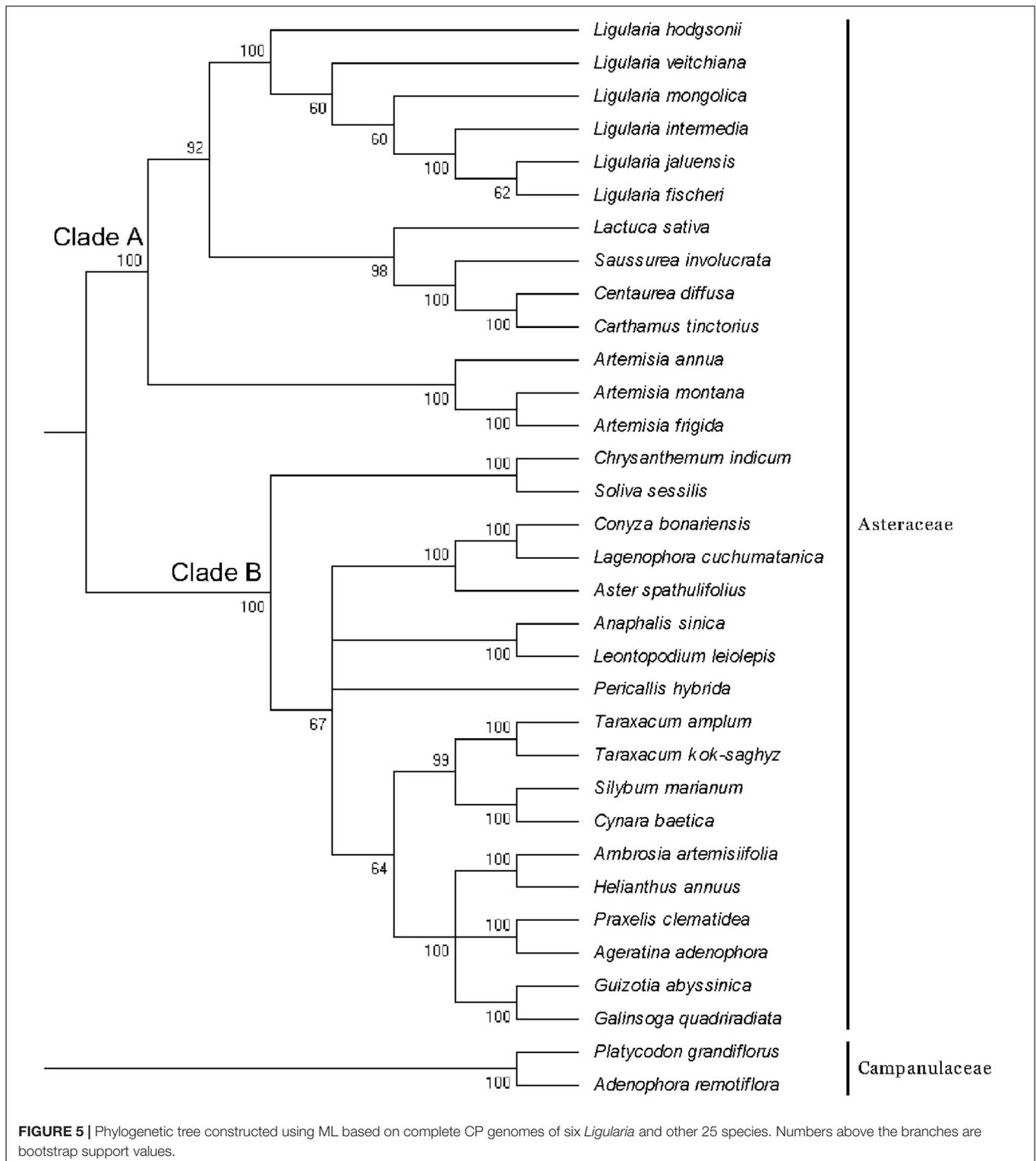
genetic region. Four types of long repeats were observed in the CP genomes of the six *Ligularia* species, including forward, palindromic, reverse, and complement repeats (Figure 3). The length of the repeat unit ranged from 30 to 48 bp. *Ligularia intermedia* and *L. jaluensis* both had 19 forward and 20 palindromic repeats. *Ligularia hodgsonii* had the following repeats: 18 forward, 20 palindromic, and one reverse. *Ligularia mongolica* had 18 forward and 20 palindromic repeats. *Ligularia veitchiana* had 20 forward and 21 palindromic repeats. *Ligularia fischeri* had the following repeats: 19 forward, 19 palindromic, and one complement. The long repeat sequences observed in the CP genomes of the six *Ligularia* species, with *L. hodgsonii* as the reference, are presented in Supplementary Table S6.

Simple sequence repeats, also called microsatellites, exist widely in the genome, and the sequences consist of one to six nucleotide repeat units (Powell et al., 1995). SSRs are widely used in studies on species identification, population genetics, and phylogenetic studies based on polymorphisms (Yang et al., 2011; Jiao et al., 2012; Xue et al., 2012). Four types of SSRs were found in the CP genomes from the six

*Ligularia* species: mononucleotide (56.6–60.7%), dinucleotide (11.5–13.2%), trinucleotide (9.3–9.8%), and tetranucleotide (18.0–21.6%); the SSRs were mainly distributed in the non-coding region of the LSC and SSC. Of all these SSRs, the number of mononucleotide SSRs (A/T) is the largest, ranging from 29 in *L. hodgsonii* to 37 in *L. veitchiana*, enriching A and T in the CP genomes. The next most common SSR is dinucleotide (AT/AT), six dinucleotide SSRs in CP genomes of *L. hodgsonii* and *L. mongolica* and seven dinucleotide SSRs in other four CP genomes. All of the CP genomes from the six species have two trinucleotide AAG/CTT SSRs, one ATC/ATG trinucleotide SSR, and 11 tetranucleotide SSRs (Table 3). The CP genome of *L. veitchiana* has three AAT/ATT trinucleotide SSRs, while the other five species only have two trinucleotide SSRs.

## Identification and Phylogenetic Analysis of *Ligularia* Species

The CP genomes from the six *Ligularia* species are highly similar. Among the few variations, non-coding regions exhibited higher



levels of variability than the coding regions. The largest change in gene length occurred in pseudogene *ycf1*, with 5,097 bp in *L. mongolica*, 5,100 bp in *L. hodgsonii* and *L. veitchiana*, and 5,094 bp in the other three species. This difference led to a divergence in the length of the coding regions of the six species.

The IR regions of the six CP genomes are conservative regardless of the number and order of the genes. Previous research screened highly variable region from CP genomes as the potential DNA barcodes for authenticating species, such as *Dioscorea* (Ma et al., 2018) and *Fritillaria* species (Li et al., 2016).

Sequence homology was investigated compared with the reference CP genome from *L. hodgsonii* using the mVISTA software (Figure 4). Our results showed high similarity among all sequences. Differences were observed in the intergenic regions of *matK-trnK* and *ndhG-ndhI* (Figure 4). There was only one variable site in the *matK-trnK* region and five variable sites in *ndhG-ndhI* region, but this is not enough to distinguish among the six *Ligularia* species. Because of the highly conservative sequences, the structure, and size of the CP genomes of *Ligularia* species, no obvious hypervariable region was screened. Thus, the complete CP genomes were considered to distinguish *Ligularia* species.

In addition to the six CP genomes sequenced in this study, 25 other CP genomes from Asteraceae were chosen to construct the phylogenetic tree, and *P. grandiflorus* and *A. remotiflora* (Campanulaceae) were included as outgroups (Figure 5). In the ML tree, we identified two main clades (clade A and B) excluding outgroup species. Six species of *Ligularia* were a monophyly with well-supported (100%). The support values in clade A were not less than 60%, and *L. fischeri* and *L. jaluensis* have a close relationship. *Ligularia* is most closely related to *L. sativa*, *Saussurea involucreta*, *Centaurea diffusa*, and *Carthamus tinctorius*. The results showed that the CP genomes can be used to identify the six *Ligularia* species.

## CONCLUSION

This study reported the CP genomes from six *Ligularia* species, and the structure and composition of the CP genomes are highly

similar. Like most Asteraceae species, the CP genomes of the six *Ligularia* species had a small 3.4 kb inversion within a large 23 kb inversion in the LSC region. The ML tree showed that the CP genome can be used to identify the six *Ligularia* species and is expected to become a super-barcode for the identification of *Ligularia* species.

## AUTHOR CONTRIBUTIONS

HY conceived the study and acquired the funding. XC, YW, and BD collected samples and conducted the experiment. JZ and YC performed the genome assembly and analysis on the data. XC and JZ wrote the manuscript. All authors have read and approved the final manuscript.

## FUNDING

This work was supported by CAMS Innovation Fund for Medical Sciences (CIFMS) (No. 2016-I2M-3-016) and Major Scientific and Technological Special Project for “Significant New Drugs Creation” (No. 2014ZX09304307001).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fphar.2018.00695/full#supplementary-material>

## REFERENCES

- Asada, Y., Furuya, T., and Murakami, N. (1981). Pyrrolizidine alkaloids from *Ligularia japonica*. *Planta Med.* 42, 202–203. doi: 10.1055/s-2007-971628
- Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D., and Pirovano, W. (2011). Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27, 578–579. doi: 10.1093/bioinformatics/btq683
- Boetzer, M., and Pirovano, W. (2012). Toward almost closed genomes with GapFiller. *Genome Biol.* 13:R56. doi: 10.1186/gb-2012-13-6-r56
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Bull, L. B., Culvenor, C. C. J., and Dick, A. T. (1968). *The Pyrrolizidine Alkaloids*. Amsterdam: North Holland Publishing Co.
- Chinese Pharmacopoeia Commission (2015). *Pharmacopoeia of the Peoples Republic of China*. China Medical. Ottawa: Science Press, 342–343.
- Choi, K. S., and Park, S. (2015). The complete chloroplast genome sequence of *Aster spathulifolius* (Asteraceae); genomic features and relationship with Asteraceae. *Gene* 572, 214–221. doi: 10.1016/j.gene.2015.07.020
- Clegg, M. T., Gaut, B. S., Learn, G. H., and Morton, B. R. (1994). Rates and patterns of chloroplast DNA evolution. *Proc. Natl. Acad. Sci. U.S.A.* 91, 6795–6801. doi: 10.1073/pnas.91.15.6795
- Curci, P. L., De, P. D., Danzi, D., Vendramin, G. G., and Sonnante, G. (2015). Complete chloroplast genome of the multifunctional crop globe artichoke and comparison with other Asteraceae. *PLoS One* 10:e0120589. doi: 10.1371/journal.pone.0120589
- Feng, K. (2016). *Analysis of Alkaloid Component and Toxicity Test in the Aerial Parts of Ligularia rumicifolia*. Ph.D. thesis, Northwest A&F University, Yangling.
- Hanai, R., Yamada, H., Suzuki, Y., Nagano, H., Kawahara, T., Yu, J. J., et al. (2012). Chemical constituents of *Ligularia nelumbifolia* and *L. subspicata* hybrid collected in Shangrila County, Yunnan province of China. *Nat. Prod. Commun.* 7, 1565–1568.
- He, L., Qian, J., Li, X., Sun, Z., Xu, X., and Chen, S. (2017). Complete chloroplast genome of medicinal plant *Lonicera japonica*: genome rearrangement, intron gain and loss, and implications for phylogenetic studies. *Molecules* 22, 249–260. doi: 10.3390/molecules22020249
- He, W., and Pan, Y. (2015). Study on the DNA barcoding of genus *Ligularia* Cass. (Asteraceae). *Plant Divers.* 37, 693–703.
- Hu, Y., Woeste, K. E., and Zhao, P. (2016). 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
- Jank, B., and Rath, J. (2017). The risk of pyrrolizidine alkaloids in human food and animal feed. *Trends Plant Sci.* 22, 191–193. doi: 10.1016/j.tplants.2017.01.002
- Jiang, D., Zhao, Z., Zhang, T., Zhong, W., Liu, C., Yuan, Q., et al. (2017). The chloroplast genome sequence of *Scutellaria baicalensis* provides insight into intraspecific and interspecific chloroplast genome diversity in *Scutellaria*. *Genes* 8:E227. doi: 10.3390/genes8090227
- Jiao, Y., Jia, H., Li, X., Chai, M., Jia, H., Chen, Z., et al. (2012). Development of simple sequence repeat (SSR) markers from a genome survey of Chinese bayberry (*Myrica rubra*). *BMC Genomics* 13:201. doi: 10.1186/1471-2164-13-201
- Kim, K. J., Choi, K. S., and Jansen, R. K. (2005). Two chloroplast DNA inversions originated simultaneously during the early evolution of the sunflower family (Asteraceae). *Mol. Biol. Evol.* 22, 1783–1792. doi: 10.1093/molbev/msi174
- Kurtz, S., Choudhuri, J. V., Ohlebusch, E., Schleiermacher, C., Stoye, J., and Giegerich, R. (2001). Reputer: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* 29, 4633–4642. doi: 10.1093/nar/29.22.4633
- Lee, J., Lee, H., Lee, S. C., Sang, H. S., Kang, J. H., Lee, T. J., et al. (2016). The complete chloroplast genome sequence of *Ligularia fischeri* (Ledeb.) Turcz. (Asteraceae). *Mitochondrial DNA B* 1, 4–5. doi: 10.1080/23802359.2015.1137793

- Li, X., Gao, H., Wang, Y., Song, J., Henry, R., Wu, H., et al. (2013). Complete chloroplast genome sequence of *Magnolia grandiflora* and comparative analysis with related species. *Sci. China Life Sci.* 56, 189–198. doi: 10.1007/s11427-012-4430-8
- Li, Y., Yao, H., Song, J., Ren, F., Li, X., and Sun, C. (2016). Screening *Fritillaria* genus-specific DNA barcodes based on complete chloroplast genome sequences. *Mod. Tradit. Chin. Med. Mater. Med. World Sci. Technol.* 18, 24–28. doi: 10.11842/wst.2016.01.004
- Lin, R., and Liu, S. (1989). *Flora Reipublicae Popularis Sinicae*. Ottawa: Science Press, 4–13.
- Liu, C., Shi, L., Zhu, Y., Chen, H., Zhang, J., Lin, X., et al. (2012). CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences. *BMC Genomics* 13:715. doi: 10.1186/1471-2164-13-715
- Liu, S., and Illarionova, I. D. (1989). *Flora of China*. Ottawa: Science Press, 20–21.
- Liu, Y., Huo, N., Dong, L., Wang, Y., Zhang, S., Young, H. A., et al. (2013). Complete chloroplast genome sequences of Mongolia medicine *Artemisia frigida* and phylogenetic relationships with other plants. *PLoS One* 8:e57533. doi: 10.1371/journal.pone.0057533
- Lohse, M., Drechsel, O., and Bock, R. (2007). OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr. Genet.* 52, 267–274. doi: 10.1007/s00294-007-0161-y
- Lu, C., Shen, Q., Yang, J., Wang, B., and Song, C. (2015). The complete chloroplast genome sequence of safflower (*Carthamus tinctorius* L.). *Mitochondrial DNA* 27, 1–3. doi: 10.3109/19401736.2015.1018217
- Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., et al. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* 1, 18–23. doi: 10.1186/2047-217X-1-18
- Ma, S., Zhou, J., Li, Y., Chen, X., Wu, M., Sun, W., et al. (2018). Study on complete chloroplast genome of *Dioscorea oppositifolia* and *D. collettii* and screening specific DNA barcodes (in Chinese). *Sci. China Life Sci.* 2018, 48–59. doi: 10.1360/N052017-00160
- Martinello, M., Borin, A., Stella, R., Bovo, D., Biancotto, G., Gallina, A., et al. (2017). Development and validation of a QuEChERS method coupled to liquid chromatography and high resolution mass spectrometry to determine pyrrolizidine and tropane alkaloids in honey. *Food Chem* 295–302. doi: 10.1016/j.foodchem.2017.04.186
- Misa-Microsatellite Identification Tool (2017). Available at: <http://pgrc.ipk-gatersleben.de/misa/> [accessed 16 September 2017]
- Mower, J. P. (2009). The PREP suite: predictive RNA editors for plant mitochondrial genes, chloroplast genes and user-defined alignments. *Nucleic Acids Res.* 37, 253–259. doi: 10.1093/nar/gkp337
- Nagy, E., Hegedűs, G., Taller, J., Kutasy, B., and Virág, E. (2017). Illumina sequencing of the chloroplast genome of common ragweed (*Ambrosia artemisiifolia* L.). *Data Brief.* 15, 606–611. doi: 10.1016/j.dib.2017.10.009
- Powell, W., Morgante, M., Mcdevitt, R., Vendramin, G. G., and Rafalski, J. A. (1995). Polymorphic simple sequence repeat regions in chloroplast genomes: applications to the population genetics of pines. *Proc. Natl. Acad. Sci. U.S.A.* 92, 7759–7763. doi: 10.1073/pnas.92.17.7759
- Prakash, A. S., Pereira, T. N., Reilly, P. E., and Seawright, A. A. (1999). Pyrrolizidine alkaloids in human diet. *Mutat. Res.* 443, 53–67. doi: 10.1016/S1383-5742(99)00010-1
- Pu, S., Xu, D., Zhang, M., Zhou, H., Wang, Z., and Yu, G. (2004). Detection of hepatotoxic pyrrolizidine alkaloids in *Ligularia* Cass. with LC/MSn. *Acta Pharm. Sin.* 39, 831–835. doi: 10.3321/j.issn:0513-4870.2004.10.014
- Saito, Y., Ichihara, M., Takashima, Y., Okamoto, Y., Gong, X., Hanai, R., et al. (2017). Chemical constituents of hybrids of *Ligularia cyathiceps* and *L. lamarum*/L. *subspicata* collected in china: structures of subspicatins M, N, O<sub>1</sub>, and O<sub>2</sub>, and related compounds. *Phytochemistry* 140, 69–76. doi: 10.1016/j.phytochem.2017.04.015
- Salih, R. H. M., Majeský, L., Schwarzacher, T., Gornall, R., and Heslop-Harrison, P. (2017). Complete chloroplast genomes from apomictic *Taraxacum* (Asteraceae): identity and variation between three microspecies. *PLoS One* 12:e0168008. doi: 10.1371/journal.pone.0168008
- Sato, S., Nakamura, Y., Kaneko, T., Asamizu, E., and Tabata, S. (1999). Complete structure of the chloroplast genome of *Arabidopsis thaliana*. *DNA Res.* 6, 283–290. doi: 10.1093/dnares/6.5.283
- Schattner, P., Brooks, A. N., and Lowe, T. M. (2005). The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.* 33, 686–689. doi: 10.1093/nar/gki366
- Sharp, P. M., and Li, W. H. (1987). The codon Adaptation Index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* 15, 1281–1295. doi: 10.1093/nar/15.3.1281
- 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, 1330–1343. doi: 10.3390/molecules22081330
- Shimizu, A., Suzuki, Y., Hanai, R., Okamoto, Y., Tori, M., Gong, X., et al. (2014). Chemical and genetic similarity and diversity of *Ligularia anoleuca* and *L. fischeri* collected in the Hengduan Mountains of China. *Phytochemistry* 102, 137–144. doi: 10.1016/j.phytochem.2014.03.019
- Smith, L. W., and Culvenor, C. C. J. (1981). Plant sources of hepatotoxic pyrrolizidine alkaloids. *J. Nat. Prod.* 44, 129–152. doi: 10.1021/np50014a001
- Stegelmeier, B. L., Edgar, J. A., Colegate, S. M., Gardner, D. L., Schoch, T. K., Coulombe, R. A., et al. (1999). Pyrrolizidine alkaloids plants, metabolism and toxicity. *J. Nat. Toxins* 8, 95–116.
- Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197
- Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680. doi: 10.1093/nar/22.22.4673
- Timme, R., Kuehl, J., Boore, J., and Jansen, R. (2007). A comparative analysis of divergent regions and categorization of shared repeats. *Am. J. Bot.* 94, 302–312. doi: 10.3732/ajb.94.3.302
- Tonti-Filippini, J., Nevill, P. G., Dixon, K., and Small, I. (2017). What can we do with 1000 plastid genomes? *Plant J.* 90, 808–818. doi: 10.1111/tpj.13491
- Wang, Q. (2007). *Cytotoxic Chemical Constituents of Three Ligularia Medicinal Plants*. Ph. D. thesis, Fudan University, Shanghai.
- Wang, X., Zhou, Z., Liu, G., and Qian, Z. (2017). Characterization of the complete chloroplast genome of the invasive weed *Galinsoga quadriradiata*, (Asterales: Asteraceae). *Conserv. Genet. Resour.* 10, 89–92. doi: 10.1007/s12686-017-0771-8
- Wiedenfeld, H., and Edgar, J. (2011). Toxicity of pyrrolizidine alkaloids to humans and ruminants. *Phytochem. Rev.* 10, 137–151. doi: 10.1007/s11101-010-9174-0
- Wyman, S. K., Jansen, R. K., and Boore, J. L. (2004). Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20, 3252–3255. doi: 10.1093/bioinformatics/bth352
- Xia, Y., Hu, Z., Li, X., Wang, P., Zhang, X., Li, Q., et al. (2016). The complete chloroplast genome sequence of *Chrysanthemum indicum*. *Mitochondrial DNA A* 27, 4668–4669. doi: 10.3109/19401736.2015.1106494
- Xiong, A., Yan, A. L., Bi, C. W., Lam, K. Y., Chan, G. K., Lau, K. K., et al. (2016). Clivorine, an otonecine pyrrolizidine alkaloid from *Ligularia* species, impairs neuronal differentiation via NGF-induced signaling pathway in cultured PC12 cells. *Phytomedicine* 23, 931–938. doi: 10.1016/j.phymed.2016.06.006
- Xue, J., Wang, S., and Zhou, S. (2012). Polymorphic chloroplast microsatellite loci in *Nelumbo* (Nelumbonaceae). *Am. J. Bot.* 99, 240–244. doi: 10.3732/ajb.1100547
- Yang, A., Zhang, J., Yao, X., and Huang, H. (2011). Chloroplast microsatellite markers in *Liriodendron tulipifera* (Magnoliaceae) and cross-species amplification in *L. chinense*. *Am. J. Bot.* 98, e123–e126. doi: 10.3732/ajb.100532
- Yi, D. K., and Kim, K. J. (2012). Complete chloroplast genome sequences of important oilseed crop *Sesamum indicum* L. *PLoS One* 7:e35872. doi: 10.1371/journal.pone.0035872

- Yu, J., Kuroda, C., and Gong, X. (2014). Natural hybridization and introgression between *Ligularia cymbulifera* and *L. tongolensis* (Asteraceae, Senecioneae) in four different locations. *PLoS One* 9:e115167. doi: 10.1371/journal.pone.0115167
- Zhang, Y., Iaffaldano, B. J., Zhuang, X., Cardina, J., and Cornish, K. (2017). Chloroplast genome resources and molecular markers differentiate rubber dandelion species from weedy relatives. *BMC Plant Biol.* 17:34. doi: 10.1186/s12870-016-0967-1
- Zhang, Y., Ma, P., and Li, D. (2011). High-throughput sequencing of six bamboo chloroplast genomes: phylogenetic implications for temperate woody bamboos (Poaceae: Bambusoideae). *PLoS One* 6:e20596. doi: 10.1371/journal.pone.0020596
- Zhou, J., Chen, X., Cui, Y., Sun, W., Li, Y., Wang, Y., et al. (2017). Molecular structure and phylogenetic analyses of complete chloroplast genomes of two *Aristolochia* medicinal species. *Int. J. Mol. Sci.* 18, 1839–1853. doi: 10.3390/ijms18091839
- Conflict of Interest Statement:** 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.
- Copyright © 2018 Chen, Zhou, Cui, Wang, Duan and Yao. 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.