

## Research Article

# Complete Chloroplast Genome Sequence of *Coptis chinensis* Franch. and Its Evolutionary History

Yang He,<sup>1</sup> Hongtao Xiao,<sup>2</sup> Cao Deng,<sup>3</sup> Gang Fan,<sup>1</sup> Shishang Qin,<sup>3</sup> and Cheng Peng<sup>1</sup>

<sup>1</sup>College of Medical Technology, Chengdu University of Traditional Chinese Medicine, Chengdu 611137, China

<sup>2</sup>Department of Pharmacy, Sichuan Academy of Medical Sciences & Sichuan Provincial People's Hospital, School of Medicine, University of Electronic Science and Technology of China, Chengdu 610072, China

<sup>3</sup>Department of Bioinformatics, DNA Stories Bioinformatics Center, Chengdu 610000, China

Correspondence should be addressed to Cheng Peng; [tcmpengcheng@163.com](mailto:tcmpengcheng@163.com)

Received 10 February 2017; Accepted 23 May 2017; Published 18 June 2017

Academic Editor: Isabel Portugal

Copyright © 2017 Yang He et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The *Coptis chinensis* Franch. is an important medicinal plant from the Ranunculales. We used next generation sequencing technology to determine the complete chloroplast genome of *C. chinensis*. This genome is 155,484 bp long with 38.17% GC content. Two 26,758 bp long inverted repeats separated the genome into a typical quadripartite structure. The *C. chinensis* chloroplast genome consists of 128 gene loci, including eight rRNA gene loci, 28 tRNA gene loci, and 92 protein-coding gene loci. Most of the SSRs in *C. chinensis* are poly-A/T. The numbers of mononucleotide SSRs in *C. chinensis* and other Ranunculaceae species are fewer than those in Berberidaceae species, while the number of dinucleotide SSRs is greater than that in the Berberidaceae. *C. chinensis* diverged from other Ranunculaceae species an estimated 81 million years ago (Mya). The divergence between Ranunculaceae and Berberidaceae was ~111 Mya, while the Ranunculales and Magnoliaceae shared a common ancestor during the Jurassic, ~153 Mya. Position 104 of the *C. chinensis ndhG* protein was identified as a positively selected site, indicating possible selection for the photosystem-chlororespiration system in *C. chinensis*. In summary, the complete sequencing and annotation of the *C. chinensis* chloroplast genome will facilitate future studies on this important medicinal species.

## 1. Introduction

Chinese goldthread, *Coptis chinensis* Franch., is an important medicinal plant in the Ranunculaceae. *C. chinensis* is native to China and has been used in traditional Chinese medicine for centuries [1, 2]. The major active compounds of *C. chinensis* are protoberberine alkaloids [1], such as berberine, palmatine, jatrorrhizine, coptisine, columbamine, and epiberberine. These compounds have antiviral, anti-inflammatory, and antimicrobial activity, and they dispel dampness, remove toxicosis, and aid detoxification [3–6]. Despite the prominent roles of *C. chinensis* in medicine, understanding of its biology and evolution is limited due to a lack of genomic resources.

Chloroplast genomes in angiosperms are mostly circular DNA molecules ranging from 115 to 165 kb in length [7]. They exhibit a conserved quadripartite structure consisting of one large single copy (LSC) region, one small single copy (SSC)

region, and two copies of inverted repeats (IR). Due to their low levels of recombination and substitution rates compared to nuclear genomes, plant chloroplast genomes are valuable sources of genetic markers for phylogenetic analyses. Over 21 complete genomes of species within the Ranunculales have been sequenced and deposited in the NCBI database (as of August 2016), and these data can be used to study chloroplast genome evolution in the Ranunculales.

The improvement of NGS technologies allows the sequencing of entire chloroplast genomes cheaper [8] and has resulted in the extensive use of chloroplast genomes for molecular marker and molecular phylogenetic studies. In our study, we assembled the complete *C. chinensis* chloroplast genome sequenced using the sequencing data generated by the Illumina HiSeq platform. Genome annotation reported both the conserved and variable information of the *C. chinensis* genome compared to other Ranunculales species. The phylogeny and molecular dating analyses also deepen our

understanding of the evolutionary history of the Ranunculales order.

## 2. Materials and Methods

**2.1. Plant Material and Library Preparation.** *C. chinensis* was collected from Shizhu, Chongqing City, China. DNA extraction and library preparation used methods described by He et al. [8]. Fresh leaves were used to extract total chloroplast DNA with the Tiagen Plant Genomic DNA Kit (Beijing, China). 300-bp DNA fragments were obtained by breaking extracted genomic DNA using a Covaris M220 Focused-Ultrasonicator (Covaris, Woburn, MA, USA). NEBNext® Ultra™ DNA Library Prep Kit Illumina (New England, Biolabs, Ipswich, MA, USA) was used to construct a sequencing library according to the manual from the manufacturer.

**2.2. DNA Sequencing, Data Preprocessing, and Genome Assembly.** Cluster generation was performed using TruSeq PE Cluster Kit (Illumina, San Diego, CA, USA), and 2 × 100 bp reads were generated on an Illumina HiSeq 2500. FASTX-Toolkit (2016a) was used to remove the adaptor-contaminated reads, low-quality bases (quality scores <20 or ambiguous nucleotide) dominated reads, and short reads (<20 bp). The remaining reads were called “clean reads.” Velvet v1.2.07 [9] was used for the de novo assembly of these clean reads, with the parameters described by He et al. [10]. To determine the contig orders and orientations, the 43 Velvet contigs were then aligned to the *M. saniculifolia* chloroplast genome [11] (NCBI RefSeq accession NC\_012615.1, a species in the Ranunculaceae). Then, five pairs of primers linking adjacent contigs were designed and used to perform PCR amplification of the unassembled regions, and the PCR products were sequenced with Sanger method. Finally, using the Lasergene SeqMan program from DNASTAR (Madison, WI, USA), the Sanger reads, together with the Velvet contigs, were further assembled into high-quality complete chloroplast genome (NCBI GenBank accession: KY120323).

**2.3. Genome Annotation.** The *C. chinensis* genomes were annotated with the DOGMA (Dual Organellar GenoMe Annotator) [12], followed by being manually reviewed to remove duplicated annotations and checking for start and stop codons. The predicted genes were also BLASTed [13] to the nonredundant protein sequences database from the NCBI, the KEGG [14], and the COG [15] database. The graphical illustration of the circular plastome was drawn using the GenomeVx [16]. To compare the function of chloroplast proteins from Ranunculales species, we annotated these proteins from Supplementary Table S3 against COG [15] database with the method same as *C. chinensis* (see Supplementary Material available online at <https://doi.org/10.1155/2017/8201836>).

**2.4. SSR Identification.** MISA (MICroSATellite identification tool, 2016b) was used to identify simple sequence repeats (SSRs) in the *C. chinensis* chloroplast genome together with 23 other chloroplast genomes. The settings included the following: more than 10 repeats for mononucleotide SSRs, six

repeats for dinucleotide SSRs, five repeats for trinucleotide SSRs, five repeats for tetranucleotide SSRs, five repeats for pentanucleotide SSRs, and five repeats for hexanucleotide SSRs. Compound SSRs were defined as two SSRs with <100 nt interspace nucleotides.

**2.5. Phylogenetic Tree Reconstruction and Divergence Time Estimation.** The chloroplast genome annotation data from the species listed in Supplementary Table S3 were downloaded from NCBI. Then, genes existing in all 24 chloroplast genomes were exacted, and a total of 42 genes remained. Using the MUSCLE (version: v3.8.31, parameters: default) [17], the protein sequences from each gene were aligned. The CDS alignments were obtained by translating the corresponding protein alignments using PAL2NAL [18] and were further concatenated into a supermatrix. Using the CDS alignments dataset, the phylogenetic tree was reconstructed by the RAxML [19] with the GTR + I +  $\Gamma$  substitution model, and the divergence times were estimated by the MCMCTree program from the PAML4.7 package [20] following the methods described by He et al. [8]. 125 and 193 Mya were set as the lower and upper boundaries for the splitting of Magnoliaceae–Ranunculales clade [21].

**2.6. Identification of Positively Selected Genes (PSGs).** The CDS alignments of 42 genes were used for the identification of positively selected genes. The  $\omega$  (Ka/Ks) ratios of filtered reliable codons in 42 genes were calculated using the branch-site model of CODEML in PAML4.7a [20], setting *C. chinensis* as the foreground branch and the others as background branches. The null hypothesis was that  $\omega$  of each site was either equal to 1 or less than 1, while the alternative hypothesis allows  $\omega$  of particular sites on the foreground branch to be larger than 1. Then, likelihood ratio test (LRT) analyses were performed, and the *p* values were used to guide against violations of model assumptions. The branch was considered to have undergone positive selection if they showed a statistically significant LRT and positively selected sites on the branch were identified in the BEB analysis.

## 3. Results and Discussions

**3.1. Genome Sequencing and Assembly.** We generated 2.13 GB pair end reads (2 × 100 bp) using the Illumina HiSeq 2500 platform (Illumina, San Diego, CA, USA). Clean reads were obtained by removing adaptors and low-quality read pairs. In total, we got 10,624,225 clean read pairs, and these clean reads were assembled into 43 contigs with N50 length of 47,033 bp using Velvet assembler (Table 1). To determine the orders and orientations, these Velvet assembled contigs were aligned to the *Megaleranthis saniculifolia* chloroplast genome [11] (Supplementary Table S1), and then gaps between two adjacent contigs were closed by Sanger reads (Supplementary Figure S1; primers are listed in Supplementary Table S2). The final complete *C. chinensis* chloroplast genome is comprised of 155,484 bp with guanine-cytosine content of 38.17% and falls within the range of the typical angiosperm chloroplast genome. By comparing it with the *M. saniculifolia* chloroplast

TABLE 1: Genome sequencing and assembly of the *C. chinensis* chloroplast genome.

Sequencing	Raw data	2.13 G
	Raw reads (pair)	10,640,000
	Read length (bp)	2 * 100
	Clean data	2.12 G
	Clean reads (pair)	10,624,225
Assembly	Total size	168,210 bp
	Contig num	43
	Average length	3,911 bp
	GC contents	38.77%
	N50 length	47,033 bp
	Min contig length	511 bp
Gap closing	Max contig length	64,014 bp
	Total size	155,484 bp
	Scaffold number	1
	GC contents	38.17%

genome, we confirmed the synteny and the absence of reversions or disorders in the genome.

**3.2. Genome Annotations.** As the general quadripartite structure found in plant chloroplast genomes, the *C. chinensis* chloroplast genome has two inverted repeated regions (IRA and IRb) of 26,758 bp in length, which split the circular genome into small single copy (SSC) and large single copy (LSC) region with 17,383 and 84,585 bp lengths, respectively. We found that the guanine-cytosine content in LSC and SSC regions (36.4% and 32.1%, respectively) is less than that in IR regions (43%). The relatively higher GC content in IR regions may be attributable to the transfer-RNA genes and ribosomal-RNA genes, which is consistent with the results from *Pogostemon cablin* [10].

The chloroplast genome of *C. chinensis* was predicted to consist of 128 gene loci, including 8 rRNA gene loci, 28 tRNA gene loci, and 92 protein-coding gene loci (Figure 1, Table 2). These gene loci contained 107 unique genes, including 80 protein-coding genes, 23 transfer-RNA genes, and 4 ribosomal-RNA genes. Each IR region contained five tRNA genes (including *trnI-CAT*, *trnL-CAA*, *trnV-GAC*, *trnR-ACG*, and *trnN-GTT*), nine protein-coding genes (ten loci), and all 4 rRNA genes. Extensions of the IR into the genes *rps19* and *ycf1* were identified (Figure 1) resulting in its pseudogenization due to incomplete duplication. There were 92 protein-coding gene loci, of which nine are duplicated (Table 2). The *ycf15* gene has four copies in the *C. chinensis* chloroplast genome, and each IR region has two copies. The *rps12* gene has three copies, and IRA, IRb, and LSC region each have one copy (Table 2).

We mapped the proteins to the NR, Clusters of Orthologous Groups (COG) [15], and Kyoto Encyclopedia of Genes and Genomes (KEGG) [14] database. A total of 76 proteins were aligned to homologous orthologs in the KEGG database; only 56 proteins could be assigned to COG orthologs (Supplementary Tables S5-S6). Homologs of all 92 proteins

except for two proteins from gene *ycf15* were identified in the NR database (Supplementary Table S4) showing high-quality annotation. Most of the proteins are involved in photosynthesis, energy metabolism, and ribosome-related functions, as indicated by annotations from the NR database and KEGG database. Consistent with other species from the same order, the COG classification of these proteins also mainly grouped into two groups: Category J (translation, ribosomal structure, and biogenesis) and Category C (energy production and conversion), which are in Supplementary Tables S7-S8.

**3.3. Identification of Simple Sequence Repeats (SSRs).** We identified perfect SSRs in the *C. chinensis* chloroplast genomes, as well as the chloroplast genomes of several other species in the Ranunculales. We found that both the numbers and types of chloroplast SSRs are variable in different species (Table 3 and Supplementary Tables S10-S11). The most abundant SSRs in all the species were mononucleotide type, with numbers varying from 16 to 71. Moreover, most mononucleotide types are comprised of polyadenine and polythymine, which is consistent with the results of other studies [10]. In addition, mononucleotide SSRs in *C. chinensis* and other species in Ranunculaceae family were fewer than those in species from Berberidaceae family, while dinucleotide type was relatively more common.

**3.4. Phylogenetic Tree Construction and Divergence Time Estimation.** To determine the evolutionary history of *C. chinensis* within the Ranunculales, we used 42 genes existing in all 24 chloroplast genomes, including 21 sequenced chloroplast genomes from species in the Ranunculales and two species from the Magnoliaceae as an outgroup. Phylogeny analysis shows that six Ranunculaceae species and 12 Berberidaceae plants comprise two unique clades, whereas the other four species are relatively divergent and ancestral in Ranunculales (Figure 2). Estimation of divergence times of these plants was performed using the MCMCTree program in the PAML4.7a package [20] (Figure 2), and all the times estimated matched well with the data deposited in TIMETREE, a public knowledge-base of divergence times among organisms, thereby confirming that the molecular clock dating strategy was reliable. *C. chinensis* is relatively ancestral in the Ranunculaceae and diverged from other Ranunculaceae plants about 81 million years ago (Mya). The divergence between Ranunculaceae and Berberidaceae was about 111 Mya, whereas Ranunculales and Magnoliaceae shared a common ancestor prior to divergence during the Jurassic period, around 153 Mya.

**3.5. Selection in the Goldthread Chloroplast Genome.** The extent to which the genes in the *C. chinensis* chloroplast genome have experienced selection is unknown. Therefore, *C. chinensis* genes and another 23 chloroplast genomes (Supplementary Table S3) were extracted and used to identify positively selected genes. The  $\omega$  ratio (dN/dS, namely, nonsynonymous substitution rate/synonymous substitution rate) is used to measure the natural selection acting on a

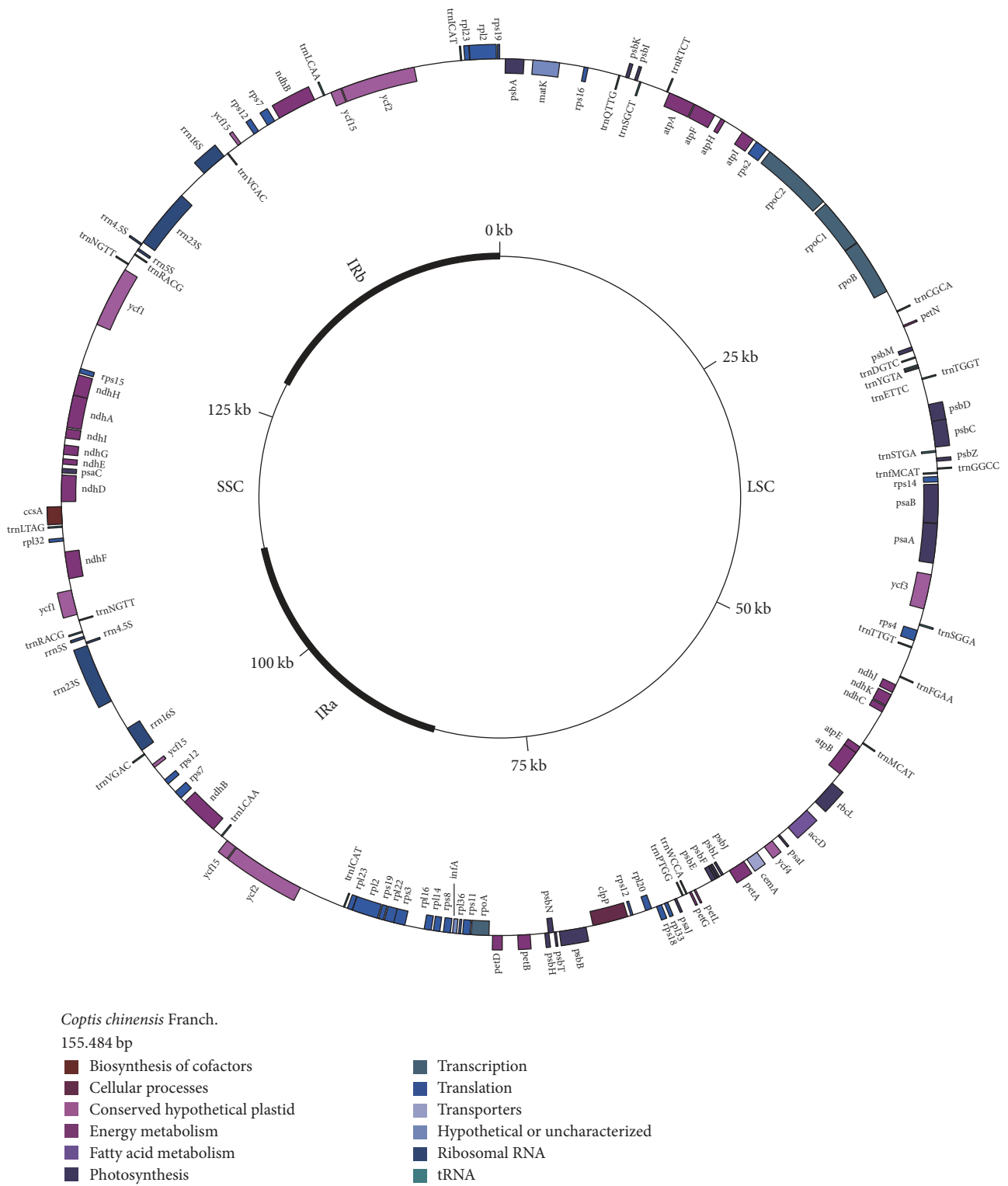


FIGURE 1: Genome schema of the *C. chinensis* chloroplast genome. Genes on the outer side the circle transcribe clockwise, while genes on the inner side transcribe counterclockwise. Genes from different functional groups are colored with different color.

TABLE 2: List of genes in the *C. chinensis* chloroplast genome. Numbers in the parentheses indicate the copy number in the genome.

Groups	Name of genes
Biosynthesis of cofactors	<i>ccsA</i>
Cellular processes	<i>clpP</i>
Conserved hypothetical plastid	<i>ycf1(2), ycf15(4), ycf2(2), ycf3, ycf4</i>
Energy metabolism	<i>atpA, atpB, atpE, atpF, atpH, atpI, ndhA, ndhB(2), ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK, petA, petB, petD, petG, petL, petN</i>
Fatty acid metabolism	<i>accD</i>
Hypothetical or uncharacterized	<i>infA, matK</i>
Photosynthesis	<i>psaA, psaB, psaC, psal, psaj, psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ, rbcL</i>
Ribosomal RNA	<i>rrn16S(2), rrn23S(2), rrn4.5S(2), rrn5S(2)</i>
Transcription	<i>rpoA, rpoB, rpoCl, rpoC2</i>
Translation	<i>rpl14, rpl16, rpl2(2), rpl20, rpl22, rpl23(2), rpl32, rpl33, rpl36, rps11, rps12(3), rps14, rps15, rps16, rps18, rps19(2), rps2, rps3, rps4, rps7(2), rps8</i>
Transporters	<i>cemA</i>
tRNA	<i>trnC-GCA, trnD-GTC, trnE-TTC, trnF-GAA, trnG-GCC, trnI-CAT(2), trnL-CAA(2), trnL-TAG, trnM-CAT, trnN-GTT(2), trnP-TGG, trnQ-TTG, trnR-ACG(2), trnR-TCT, trnS-GCT, trnS-GGA, trnS-TGA, trnT-GGT, trnT-TGT, trnV-GAC(2), trnW-CCA, trnY-GTA, trnfM-CAT</i>

TABLE 3: Statistics of chloroplast SSRs detected in 24 species. p1, mononucleotide SSRs; p2, dinucleotide SSRs; p3, trinucleotide SSRs; c, compound SSRs; A, adenine; G, guanine; T, thymine; C, cytosine.

Species	Total	All	p1				All	p2		p3	c
			(A)n	(C)n	(G)n	(T)n		(AT)n	(TA)n		
<i>E. pseudowushanense</i>	68	61	26	1	1	33	2	1	1	0	5
<i>E. lishihchenii</i>	73	67	26	1	1	39	2	1	1	0	4
<i>E. sagittatum</i>	63	55	22	1	1	31	2	1	1	0	6
<i>E. dolichostemon</i>	65	59	25	1	1	32	2	1	1	0	4
<i>E. acuminatum</i>	69	62	28	1	1	32	2	1	1	0	5
<i>E. koreanum</i>	68	60	26	1	1	32	2	2	0	0	6
<i>S. hexandrum</i>	33	27	10	0	1	16	2	2	0	0	4
<i>Nandina domestica</i>	52	48	19	0	0	29	1	0	1	0	3
<i>G. microrrhynchum</i>	71	59	26	1	0	32	1	0	1	1	10
<i>B. amurensis</i>	72	62	29	0	0	33	2	1	0	2	6
<i>B. koreana</i>	73	66	30	0	0	36	2	1	0	2	3
<i>B. bealei</i>	75	71	39	0	0	32	1	1	0	2	1
<i>R. macranthus</i>	32	28	13	0	0	15	3	3	0	1	0
<i>Clematis terniflora</i>	50	39	18	0	1	20	4	0	4	1	6
<i>Aconitum chiisanense</i>	34	21	9	2	0	10	10	4	6	0	3
<i>T. coreanum</i>	50	44	19	2	0	23	4	0	4	1	1
<i>M. saniculifolia</i>	38	36	10	0	0	26	2	2	0	0	0
<i>Coptis chinensis</i>	47	38	16	0	0	22	5	3	2	0	4
<i>Stephania japonica</i>	53	42	17	2	0	23	8	3	5	0	3
<i>Akebia trifoliata</i>	40	38	20	2	1	15	0	0	0	0	1
<i>P. somniferum</i>	16	16	8	0	0	8	0	0	0	0	0
<i>Euptelea pleiosperma</i>	65	57	27	0	0	30	0	0	0	0	8
<i>L. chinense</i>	49	41	13	2	0	26	3	2	1	1	4
<i>L. tulipifera</i>	56	47	18	2	0	27	3	2	1	1	5

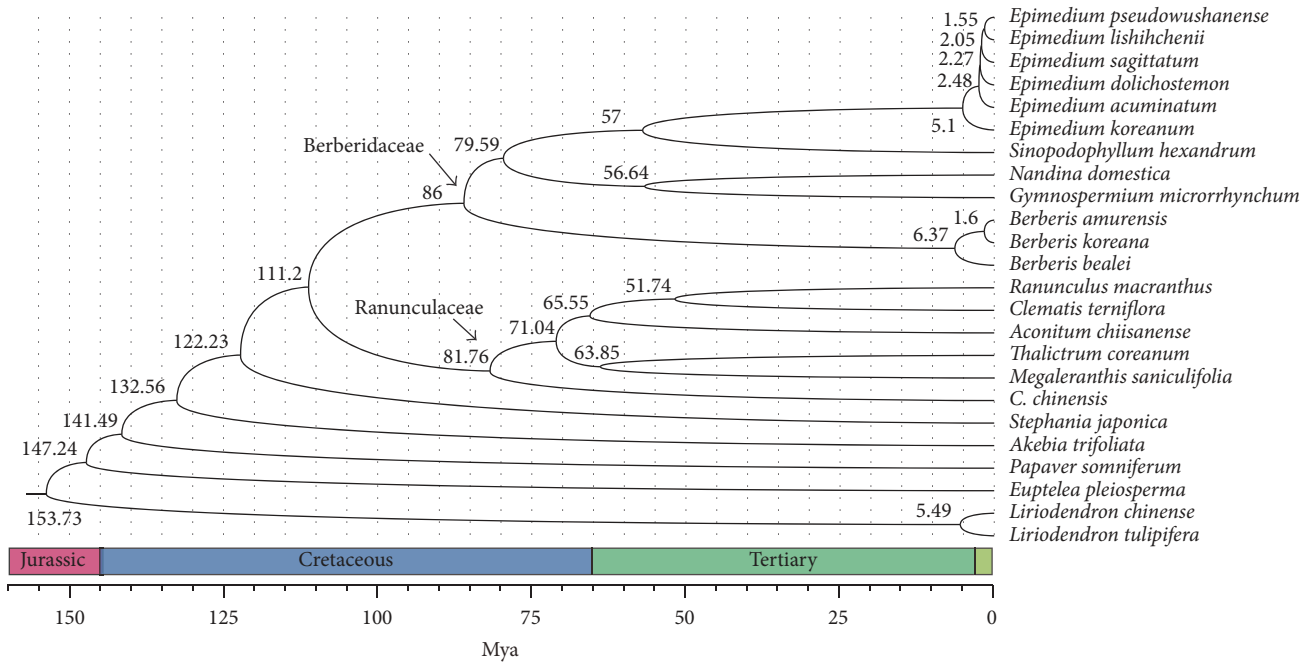


FIGURE 2: Phylogenetic tree and estimated divergence time based on chloroplast genomes from Ranunculales and Magnoliaceae.

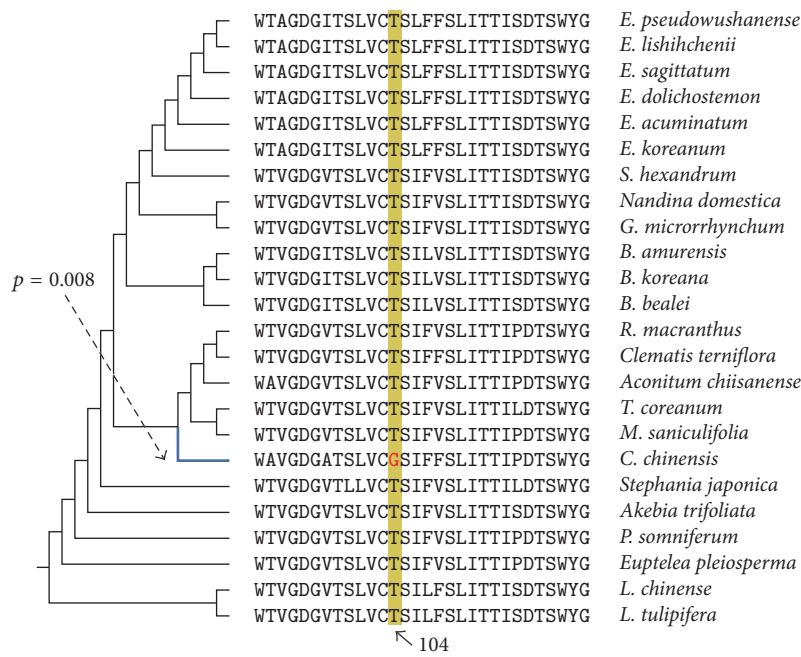


FIGURE 3: Multisequence alignments and positive selection of the *ndhG* gene. Blue branch is the foreground branch used in the branch-site test implemented, while yellow background color indicates the positively selected site (position 104 based on the *C. chinensis ndhG* protein sequence).

gene. To detect potential positive selection affecting selected sites along *C. chinensis* lineages, the branch-site model implemented in PAML [20] was applied (Figure 3). The results suggest that the *ndhG* (NADH dehydrogenase subunit 6) evolved under positive selection in the *C. chinensis* lineage (Supplementary Table S3). The test statistic ( $2\Delta\ln$ ) of *ndhG* gene was 5.79, and the *p* value was 0.008. BEB analysis

revealed the position 104 of this protein as positively selected in *C. chinensis*, with posterior probabilities of 0.994. The *ndhG* is one of the 11 NADH dehydrogenase genes, and the *ndhG* subunit is associated with nuclear-encoded subunits to form the NADH dehydrogenase-like complex in angiosperm chloroplasts. This protein complex associates with photosystem I and then forms a supercomplex, which mediates

cyclic electron transport [22], produces ATP to balance the ATP/NADPH ratio, and facilitates chlororespiration [23]. Therefore, the selection values identified in *C. chinensis* indicate positive selection for elements of the photosystem-chlororespiration system.

### Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

### Authors' Contributions

Yang He and Hongtao Xiao contributed equally to this work.

### Acknowledgments

This research was funded by the National Natural Science Foundation of China (Grant no. 81303168), the Science & Technology Program of Sichuan Province (Grant nos. 2015JQ0027 and 2015ZR0160), and the Science and Technology Special Fund of Sichuan Provincial Administration of Traditional Chinese Medicine (Grant no. 2016Q060).

### References

- [1] S. Kamath, M. Skeels, and A. Pai, "Significant differences in alkaloid content of *Coptis chinensis* (Huanglian), from its related American species," *Chinese Medicine*, vol. 4, article no. 1749, p. 17, 2009.
- [2] K.-L. Xiang, S.-D. Wu, S.-X. Yu et al., "The first comprehensive phylogeny of *Coptis* (Ranunculaceae) and its implications for character evolution and classification," *PLoS ONE*, vol. 11, no. 4, Article ID e0153127, 2016.
- [3] Q. Zhang, X.-L.-L. Piao, X.-S.-S. Piao et al., "Preventive effect of *Coptis chinensis* and berberine on intestinal injury in rats challenged with lipopolysaccharides," *Food and Chemical Toxicology*, vol. 49, no. 1, pp. 61–69, 2011.
- [4] L. Cui, M. Liu, X. Chang et al., "The inhibiting effect of the *Coptis chinensis* polysaccharide on the type II diabetic mice," *Biomedicine Pharmacotherapy*, vol. 81, pp. 111–119, 2016.
- [5] T. Friedemann, Y. Ying, W. Wang et al., "Neuroprotective Effect of *Coptis chinensis* in MPP + and MPTP-Induced Parkinson's Disease Models," *American Journal of Chinese Medicine*, vol. 44, no. 5, pp. 907–925, 2016.
- [6] E. Kim, S. Ahn, H.-I. Rhee, and D.-C. Lee, "Coptis chinensis Franch. extract up-regulate type I helper T-cell cytokine through MAPK activation in MOLT-4 T cell," *Journal of Ethnopharmacology*, vol. 189, pp. 126–131, 2016.
- [7] R. J. Henry, *Plant Diversity and Evolution: Genotypic and Phenotypic Variation in Higher Plants*, CABI Publishing, 2005.
- [8] Y. He, H. Xiao, C. Deng, L. Xiong, J. Yang, and C. Peng, "The complete chloroplast genome sequences of the medicinal plant *Pogostemon Cablin*," *International Journal of Molecular Sciences*, vol. 17, no. 6, article no. 820, 2016.
- [9] D. R. Zerbino and E. Birney, "Velvet: algorithms for de novo short read assembly using de Bruijn graphs," *Genome Research*, vol. 18, no. 5, pp. 821–829, 2008.
- [10] Y. He, H. Xiao, C. Deng, L. Xiong, H. Nie, and C. Peng, "Survey of the genome of *Pogostemon cablin* provides insights into its evolutionary history and sesquiterpenoid biosynthesis," *Scientific Reports*, vol. 6, Article ID 26405, 2016.
- [11] Y.-K. Kim, C.-W. Park, and K.-J. Kim, "Complete chloroplast DNA sequence from a Korean endemic genus, *Megaleranthus saniculifolia*, and its evolutionary implications," *Molecules and Cells*, vol. 27, no. 3, pp. 365–381, 2009.
- [12] S. K. Wyman, R. K. Jansen, and J. L. Boore, "Automatic annotation of organellar genomes with DOGMA," *Bioinformatics*, vol. 20, no. 17, pp. 3252–3255, 2004.
- [13] C. Camacho, G. Coulouris, V. Avagyan et al., "BLAST+: architecture and applications," *BMC Bioinformatics*, vol. 10, article 421, 2009.
- [14] M. Kanehisa, S. Goto, Y. Sato, M. Furumichi, and M. Tanabe, "KEGG for integration and interpretation of large-scale molecular data sets," *Nucleic Acids Research*, vol. 40, no. 1, pp. D109–D114, 2012.
- [15] R.-L. Tatusov, N.-D. Fedorova, J.-D. Jackson et al., "The COG database: an updated version includes eukaryotes," *BMC Bioinformatics*, vol. 4, no. 1, p. 41, 2003.
- [16] G. C. Conant and K. H. Wolfe, "GenomeVx: Simple web-based creation of editable circular chromosome maps," *Bioinformatics*, vol. 24, no. 6, pp. 861–862, 2008.
- [17] R. C. Edgar, "MUSCLE: multiple sequence alignment with high accuracy and high throughput," *Nucleic Acids Research*, vol. 32, no. 5, pp. 1792–1797, 2004.
- [18] M. Suyama, D. Torrents, and P. Bork, "PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments," *Nucleic Acids Research*, vol. 34, pp. W609–W612, 2006.
- [19] A. Stamatakis, "RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies," *Bioinformatics*, vol. 30, no. 9, pp. 1312–1313, 2014.
- [20] Z. Yang, "PAML 4: phylogenetic analysis by maximum likelihood," *Molecular Biology and Evolution*, vol. 24, no. 8, pp. 1586–1591, 2007.
- [21] M. D. Crisp and L. G. Cook, "Cenozoic extinctions account for the low diversity of extant gymnosperms compared with angiosperms," *New Phytologist*, vol. 192, no. 4, pp. 997–1009, 2011.
- [22] Y. Munekage, M. Hashimoto, C. Miyake et al., "Cyclic electron flow around photosystem I is essential for photosynthesis," *Nature*, vol. 429, no. 6991, pp. 579–582, 2004.
- [23] G. Peltier and L. Cournac, "Chlororespiration," *Annual Review of Plant Biology*, vol. 53, pp. 523–550, 2002.