The Complete Chloroplast Genome of *Coptis teeta* (Ranunculaceae), An Endangered Plant Species Endemic to the Eastern Himalaya

Ya-Fang Gao¹, Xiao-Li Liu¹, Guo-Dong Li¹⁺, Zi-Gang Qian¹⁺, Yong-Hong Zhang² and Ying-Ying Liu¹⁺,³

¹Faculty of Traditional Chinese Pharmacy, Yunnan University of Traditional Chinese Medicine, Kunming, P.R. China
²School of Life Sciences, Yunnan Normal University, Kunming, P.R. China
³Yunnan Institute for Food and Drug, Kunming, P.R. China

**Abstract**

*Coptis teeta* is an endemic and endangered medicinal plant from the Eastern Himalaya. It has been categorized by the International Union for Conservation of Nature (IUCN) as Endangered (EN). The whole chloroplast genome of *C. teeta* was sequenced based on next-generation sequencing (NGS) in present study. The circular chloroplast genome exhibits typical quadripartite regions with 154,280 bp in size, including two inverted repeat (IR, 24,583 bp) regions, one large single copy region (LSC) and one small single copy region (SSC) of 87,519 bp and 17,595 bp, respectively. The genome contains 125 genes, including 81 protein-coding genes (PCGs), 36 tRNA genes and 8 rRNA genes. Total GC content of *C. teeta* is 38.3%, while those of IR regions (43.3%) are higher than LSC (36.7%) and SSC (32.2%) regions. Forty-two forward and twenty-three reverted repeats were detected in cp genome of *C. teeta*. The genome was rich in SSRs and totally 62 SSRs were visualized. The phylogenetic tree showed that species from the Ranunculaceae formed a monophyletic clade and the intra-family topology was mainly focus on microsite requirements that cannot be met in other habitats. Owing to the over-exploitation, several anthropogenic factors, and environmental disruption, the wild population of *C. teeta* decreased rapidly in recent years [4]. *C. teeta* has been listed in IUCN Red List of Threatened Species (http://www.iucnredlist.org/) as endangered species with status “A2cd”. And it is also included in Category II of the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) [5]. Therefore, it is necessary to protect this endangered plant for its highly economic and ecological values, and for the conservation of biodiversity.

To date, few studies of this species have been performed due to lack of genomic data of *C. teeta*. The previous studies mainly focus on phylogenetic analysis and biogeographic pattern of Coptis by using two plastid and one nuclear markers including *psbA-trnH, trnl-trnF* and ITS, and six markers, including five plastid and one nuclear markers, respectively [6,7]. In this study, as a part of the genome sequencing project of *C. teeta*, we assemble and annotate its complete plastid genome and describing its characteristics.

**Materials and Methods**

**Plant material and DNA extraction**

Fresh leaves of *C. teeta* were collected from Gongshan County (27°73'E, 98°66'N), Yunnan province and voucher specimens were deposited in Yunnan University of Traditional Chinese Medicine. Total genomic DNA was extracted using the modified plant genome kit (Biotek, Beijing, China). DNA quality was detected by electrophoresis on 1% agarose gel (Figure 1) and 1 μL of DNA sample to test concentration using to the NanoDrop spectrophotometers (ThermoFisher...
Scientific, Wilmington, Delaware, USA), the result showed that its value is 62.6 ng/μL>50 ng/μL.

Figure 1 1% Agarose gel electrophoretic separation mapping of total DNA.

Genome sequencing, assembly and annotation
A sequence library was constructed and sequencing was performed using the Illumina HiSeq 2500-PE150 platform (Illumina, CA, USA). All raw reads were filtered by using NGSQC Toolkit_v2.3.3 with default parameters to obtain clean reads that has discard low quality regions [8]. The plastome was de novo assembled using GetOrganelle pipeline (https://github.com/Kinggerm/GetOrganelle). The complete plastome genome was annotated with the online annotation tool GeSeq (https://chlorobox.mpimp-golm.mpg.de/geseq.html) [9], using the published cp genome of C. chinensis (NCBI accession number: NC036485) as a reference sequence, then manual correction was performed with Geneious R11 software [10]. The plastid genome map was drawn using OGDRAW program (http://ogdraw.mpimp-golm.mpg.de/) [11]. The annotated cp genome of C. teeta has been deposited into GenBank with the Accession Number MH359096.

Repeats and simple sequence repeats (SSRs) analysis
REPuter [12] was used to find forward and reversed tandem repeats15 bp with minimum alignment score and maximum period size at 100 and 500, respectively. IMEx [13] was used to visualize the SSRs with the minimum repeat numbers set to 10, 5, 4, 3, 3 and 3 for mono-, di-, tri-, tetra-, penta- and hexa-nucleotides, respectively.

Phylogenetic analysis
The phylogenetic analysis was conducted based on 31 published chloroplast genomes to infer phylogenetic position of C. teeta within the family of Ranunculaceae. The cp genome of Nandina domestica (GenBank: DQ923117) was included as outgroup. The LSC, SSC and one IR region of the total 32 chloroplast genomes were aligned using MAFFT 7.308 [14]. The maximum likelihood (ML) tree was reconstructed by RAxML 8.2.11 [15] with the nucleotide substitution model of GTR+G and node support was estimated by means of bootstrap analysis with 1000 replicates.

Results and Discussion
Characteristics of chloroplast genome of C. teeta
The complete chloroplast genome of C. teeta is a circular DNA with 154,280 bp in length, comprising four subunits: one large single copy (LSC) (87,519 bp), one small single copy (SSC) (17,595 bp) and two inverted repeat regions (IRs) (24,583 bp for each) (Figure 2). The overall GC content was 38.3 %. The IR regions had a higher GC content (43.3%) than LSC (36.7%) and SSC regions (32.2%). That was caused by the high GC content of the four ribosomal RNA (rRNA) genes (55.5%) presented in the IR regions, similar to that of C. chinensis Franchet [16].

Figure 2 Plastome map of Coptis teeta. The darker gray in the inner circle corresponds to GC content, while the lighter gray corresponds to AT content.
Repeat and SSR analysis

For repeat structure analysis, 42 forward and 23 reverted repeats with minimal repeat size of 15 bp were detected in cp genome of *C. teeta* (Table 1). Most of these repeats were between 15 and 20 bp. The longest forward repeats were of 39 bp, one sequence of which located in the intergenic region between trnV-GAC and rps7 of inverted repeated regions (IR), the other sequence located in ycf3 of LSC. There are 31 repeats with two sequences started in the same region. Among them, 21 repeats located in the LSC region, 7 located in the IR regions, and 3 located in SSC region. Other 34 repeats with two sequences started in separated regions.

Table 1 Repeat sequences in *C. teeta* chloroplast genome.

| ID | Repeat Start 1 | Type | Size (bp) | Repeat Start 2 | E-Value | Region | Gene |
|----|----------------|------|-----------|----------------|---------|--------|------|
| 1  | 161            | F    | 15        | 94911          | 6.23    | IRb/LSC| ycf1; IGS |
| 2  | 1319           | F    | 15        | 146257         | 6.23    | IRb/SSC| ycf1; ndhA |
| 3  | 1686           | F    | 18        | 8903           | 0.0974  | IRb    | trnN-GUU; trnL-GAU |
| 4  | 1729           | F    | 15        | 22897          | 6.23    | IRb    | trnN-GUU; ycf2 |
| 5  | 2352           | F    | 15        | 34195          | 6.23    | IRb    | trnR-ACG; trnS-GCU |
| 6  | 2352           | F    | 15        | 61944          | 6.23    | IRb    | trnR-ACG; trnS-UGA |
| 7  | 3964           | F    | 15        | 7706           | 6.23    | IRb    | rRNA23; trnL-GAU |
| 8  | 6084           | F    | 15        | 91596          | 6.23    | IRb    | IGS |
| 9  | 6598           | F    | 19        | 7684           | 0.0244  | IRb    | trnA-UGC; trnL-GAU |
| 10 | 7003           | F    | 15        | 95317          | 6.23    | IRb    | trnA-UGC; IGS |
| 11 | 7017           | F    | 18        | 74997          | 0.0244  | IRb    | trnA-UGC; trnF-GAA |
| 12 | 8184           | F    | 18        | 8215           | 0.0974  | IRb    | IGS |
| 13 | 9475           | F    | 15        | 148599         | 6.23    | IRb    | rRNA16; ndhH |
| 14 | 10179          | F    | 17        | 57871          | 0.39    | IRb    | trnV-GAC; trnT-GGU |
| 15 | 11638          | F    | 16        | 38957          | 1.56    | IRb    | IGS |
| 16 | 11991          | F    | 39        | 70172          | 2.21E-14| IRb    | IGS; ycf3 |
| 17 | 19923          | F    | 16        | 20059          | 1.56    | IRb    | ycf2 |
| 18 | 20113          | F    | 16        | 71145          | 1.56    | IRb    | ycf2; IGS |
| 19 | 24047          | F    | 16        | 94901          | 1.56    | IRb    | ycf2; IGS |
| 20 | 30432          | F    | 20        | 62669          | 0.00609 | LSC    | IGS |
| 21 | 32229          | F    | 16        | 81798          | 1.56    | LSC    | IGS |
| 22 | 34192          | F    | 21        | 61941          | 0.0152  | LSC    | trnS-GCU; trnS-UGA |
| 23 | 35625          | F    | 19        | 62985          | 0.0244  | LSC    | trnG-GCC; trnG-UGU |
| 24 | 38000          | F    | 16        | 38051          | 1.56    | LSC    | atpF |
| 25 | 39908          | F    | 17        | 101565         | 0.39    | LSC    | IGS; petB |
| 26 | 45078          | F    | 16        | 94306          | 1.56    | LSC    | rpoC2; IGS |
| 27 | 46933          | F    | 16        | 98952          | 1.56    | LSC    | rpoC1; psbB |
| 28 | 54520          | F    | 16        | 102145         | 1.56    | LSC    | IGS; petB |
| 29 | 55783          | F    | 16        | 77512          | 1.56    | LSC    | IGS |
| 30 | 57577          | F    | 17        | 139324         | 0.39    | LSC/SSC| IGS |
| 31 | 58768          | F    | 16        | 92720          | 1.56    | LSC    | IGS |
| 32 | 62639          | F    | 17        | 71507          | 0.39    | LSC    | IGS |
cpSSRs markers are widely used to study the population genetics and evolutionary processes of wild plants [17,18]. There were totally 62 SSRs in cp genome of C. teeta, most of which were in LSC (Table 2). Among them, 31 (50.0%) were mononucleotide SSRs, fifteen (24.2%) were dinucleotide SSRs, six (9.7%) were tri-nucleotide SSRs, eight (12.9%) were tetra-nucleotide SSRs, one (0.2%) was penta-nucleotide SSR, and one (0.2%) was hexa-nucleotide SSRs. Only twelve SSRs were located in genes and the others were in intergenic regions. 30 (96.8%) of the mononucleotide SSRs belonged to the A/T type, which were consistent with the hypothesis that cpSSRs were generally composed of short polyadenine (poly A) or polythymine (poly T) repeats and rarely contained tandem
guanine (G) or cytosine (C) repeats. These cpSSR markers can be used in the conservation genetics of *C. teeta*.

### Table 2: Simple sequence repeats (SSRs) in the *C. teeta* chloroplast genome.

| ID | Repeat Motif | Length (bp) | Start | End | Region | Gene   |
|----|--------------|-------------|-------|-----|--------|--------|
| 1  | (T) 10       | 10          | 2610  | 2619| IRb    |        |
| 2  | (AATA) 3     | 12          | 11645 | 11656| IRb    |        |
| 3  | (ATCT) 3     | 12          | 29571 | 29582| LSC    | trnK-UUU|
| 4  | (A) 10       | 10          | 30435 | 30444| LSC    |        |
| 5  | (C) 11       | 11          | 31241 | 31251| LSC    | rps16  |
| 6  | (AT) 7       | 14          | 32232 | 32245| LSC    |        |
| 7  | (A) 11       | 11          | 33582 | 33592| LSC    |        |
| 8  | (T) 11       | 11          | 35187 | 35196| LSC    | trnG-UCC|
| 9  | (T) 10       | 10          | 35734 | 35743| LSC    |        |
| 10 | (CTGT) 3     | 12          | 36989 | 37000| LSC    | atpA   |
| 11 | (T) 10       | 10          | 40322 | 40331| LSC    |        |
| 12 | (T) 10       | 10          | 42352 | 42361| LSC    |        |
| 13 | (T) 10       | 10          | 44550 | 44563| LSC    | rpoC2  |
| 14 | (TA) 5       | 10          | 45302 | 45311| LSC    | rpoC2  |
| 15 | (TA) 5       | 10          | 45923 | 45932| LSC    | rpoC2  |
| 16 | (TA) 5       | 10          | 53640 | 53649| LSC    |        |
| 17 | (TA) 5       | 10          | 54985 | 54994| LSC    |        |
| 18 | (CTGT) 3     | 12          | 55785 | 55798| LSC    |        |
| 19 | (TA) 5       | 10          | 55828 | 55837| LSC    |        |
| 20 | (TA) 5       | 10          | 57582 | 57594| LSC    |        |
| 21 | (TTATA) 3    | 15          | 58077 | 58091| LSC    |        |
| 22 | (TA) 6       | 12          | 58573 | 58584| LSC    |        |
| 23 | (AAAG) 3     | 12          | 59161 | 59172| LSC    |        |
| 24 | (TA) 10      | 10          | 62672 | 62681| LSC    |        |
| 25 | (TA) 10      | 10          | 68971 | 68982| LSC    |        |
| 26 | (TA) 10      | 10          | 72877 | 72887| LSC    |        |
| 27 | (TA) 10      | 10          | 73196 | 73209| LSC    |        |
| 28 | (TA) 10      | 10          | 73741 | 73750| LSC    |        |
| 29 | (TA) 10      | 10          | 75418 | 75427| LSC    |        |
| 30 | (TA) 10      | 10          | 77514 | 77527| LSC    |        |
| 31 | (TA) 10      | 10          | 77525 | 77536| LSC    |        |
| 32 | (TA) 10      | 10          | 78836 | 78847| LSC    | trnV-UAC|
| 33 | (TA) 10      | 10          | 81391 | 81402| LSC    | atpB   |
| 34 | (TA) 10      | 10          | 81801 | 81814| LSC    |        |
| 35 | (TA) 10      | 10          | 86612 | 86623| LSC    |        |
The phylogenetic tree showed that species from the Ranunculaceae formed a monophyletic clade (Figure 3) and the intra-family topology was consistent with previous studies [6,16,19]. The result strongly supported C. teeta and its congeneric species, C. chinensis, as sister group with 100% bootstrap value. This newly reported chloroplast genome will provide new insights into phylogenetic studies within the Ranunculaceae and facilitate future conservation of C. teeta.

Conclusion

In this study, we reported and analyzed the first complete chloroplast genome of C. teeta, which are an endemic and endangered plant and a source for famous traditional Chinese medicine.

Figure 3 The plastome phylogeny of Ranunculaceae. Bootstraps were shown next to the node.
The circular chloroplast genome exhibits typical quadripartite regions with 154,280 bp in size, including two inverted repeat (IR, 24,583 bp) regions, one large single copy region (LSC) and one small single copy region (SSC) of 87,519 bp and 17,595 bp, respectively.

The cp genome of *C. teeta* was rich in SSRs, which will be informative sources for developing new molecular markers to evaluate genetic diversity and provide effective strategies for conservation of this species. The phylogenetic analysis showed that *C. teeta* and *C. chinensis* form one clade as sister group. This information will be useful on phylogenetic analysis of genus *Coptis* and will also enhance our understanding on the evolutionary relationships among Ranunculaceae.

**Acknowledgments**

This work was supported by the National Natural Science Foundation of China (Grant No. 81560613), Special subsidies for public health services of TCM “the national survey of TCM resources” (DSS, MOF. No 66/2017), and the Key laboratory training program in Yunnan (2017DG006).

**References**

1. Huang J, Long C (2007) *Coptis teeta*-based agroforestry system and its conservation potential: A case study from Northwest Yunnan. AMBIO 36: 344.
2. Pandit MK, Babu CR (1999) Synaptic mutation associated with gametic sterility and population divergence in *Coptis teeta* (Ranunculaceae). Bot J Linn Soc 133: 526.
3. Wang WQ (2016) A review on pharmacologic effects of effective ingredients in Huanglian. Clin J Chinese Med 26: 147-148.
4. Pandit MK, Babu CR (1998) Biology and conservation of *Coptis teeta* Wall: An endemic and endangered medicinal herb of Eastern Himalaya. Envir Conserv 25: 262.
5. UNEP-WCMC (Comps) (2014) Checklist of CITES species. CITES Secretariat, Geneva, Switzerland, and UNEP-WCMC, Cambridge, United Kingdom.
6. Xiang KL, Wu SD, Yu SX, Liu Y, Florian J, et al. (2016) The first comprehensive phylogeny of *Coptis* (Ranunculaceae) and its implications for character evolution and classification. PLoS ONE 11: e0153127.
7. Xiang KL, Andrey SE, Xiang XG, Florian J, Wang W (2018) Biogeography of *Coptis salisb.* (Ranunculales, Ranunculaceae,Coptidoideae), an Eastern Asian and North American genus. BMC Evol Biol 18: 74.
8. Patel RK, Jain M (2012) NGS QC toolkit: A toolkit for quality control of next generation sequencing data. PloS ONE 7: e30619.
9. Tillich M, Lehwark P, Pellicer T, Ulbricht-Jones ES, Fischer A, et al. (2017) GeSeq - versatile and accurate annotation of organelle genomes. Nucleic Acids Res 45: W6-W11.
10. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, et al. (2012) Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28: 1647-1649.
11. Lohse M, Drechsel O, Kahla S, Bock R (2013) Organellar Genome-DRAW a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. Nucleic Acids Res 41: 575-580.
12. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, et al. (2001) REPuter: The manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res 29: 4633-4642.
13. Mudunuri SB, Nagarajaram HA (2007) IMEx: Imperfect microsatellite extractor. Bioinformatics 23: 1181-1187.
14. Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol Biol Evol 30: 772-780.
15. Stamatakis A, Hoover P, Rougemont J (2008) A rapid bootstrap algorithm for the RaxML web servers. Syst Biol 57: 758-771.
16. He Y, Xiao HT, Deng C, Fan G, Qin SS, et al. (2017) Complete chloroplast genome sequence of *Coptis chinensis* Franch and its evolutionary history. BioMed Res Int 1-7.
17. Provan J (2009) Novel chloroplast microsatellites reveal cytoplasmic variation in *Arabidopsis thaliana*. Mol Ecol 9: 2183-2185.
18. Flannery ML, Mitchell FJ, Coyne S (2006) Plastid genome characterization in Brassica and Brassicaceae using a new set of nine SSRs. Theor Appl Genet 113: 1221-1231.
19. Liu HJ, Xie L (2016) Advances in molecular phylogenetics of Ranunculaceae. Acta Bot Boreali-Occidentalia Sinica 36: 1916-191.