Characterization of the complete plastid genome of an endangered species *Fortunaria sinensis* (Hamamelidaceae)

Yong Xu*a,b, Tian-Wen Xia*a,b, Nan Zhao*a,b, Ting Li*a,b, Tong-Jian Liu*a,c and Hai-Fei Yan*a,c

*aKey Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, PR China; *bUniversity of Chinese Academy of Sciences, Beijing, PR China; *cGuangdong Provincial Key Laboratory of Applied Botany, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, PR China

**ABSTRACT**

*Fortunaria sinensis* Rehder & E.H.Wilson, the only species of monotypic genus *Fortunaria* (Hamamelidaceae), is a native rare and vulnerable deciduous tree in China. In this study, the complete plastid genome of *F. sinensis* was determined using the Illumina paired-end sequencing data. The complete plastid genome of *F. sinensis* is 1,59,441 bp in length, consisting of a pair of inverted repeats (IR) regions (26,268 bp), a large-single copy (LSC) region (88,124 bp), and a small-single copy (SSC) region (18,781 bp). The plastid genome encodes 112 unique genes, including 80 protein-coding genes, 28 tRNA genes, and 4 rRNA genes. The phylogenetic analysis demonstrates a close relationship between *F. sinensis* and *Sinowilsonia henryi*.

*Fortunaria sinensis* Rehder & E.H. Wilson is the only species of the monotypic genus *Fortunaria* in Hamamelidaceae. This species is an endemic medicinal plant in China, since its leaves contain bergenin, fortunarioside, and other medicinal ingredients (Wang 1996). In addition, *F. sinensis* is economically important because of its high oil content in seeds (Jia and Zhou 1987). It has been assessed as a vulnerable species in China (Qin et al. 2017). In this study, we characterized the complete plastid genome sequence of *F. sinensis* based on Illumina paired-end sequencing for further genetic studies, conservation, and utilization of the endangered and rare species.

Fresh leaves of *F. sinensis* were collected from Wuhan Botanical Garden and dried with silica gel. The voucher specimen (voucher number: wh271) was deposited in the Herbarium of South China Botanical Garden (IBSC). Total genomic DNA was extracted using the modified CTAB method (Doyle and Doyle 1987). The prepared Illumina paired-end library with an average insert size of 270 bp was sequenced on a Hiseq X Ten platform. After quality filtering and trimming, c. 3.0Gb clean data was assembled with a reference plastid genome of *Sinowilsonia henryi* Hemsfl. (GenBank accession number NC_036069) using NOVOPlasty (Dierckxsens et al. 2017). We remapped the clean reads to the resulting circular sequence in BWA version 0.7.17 (Li 2013), and manually identified and fixed single nucleotide and small structural errors using Geneious version R11.1.5 (Biomatters Ltd., Auckland, New Zealand). The annotation of the plastome was performed through the online programme GeSeq – Annotation of organellar genomes (Tillich et al. 2017) with necessary manual adjustment in Geneious.

The plastome of *F. sinensis* (GenBank accession number MK533616) is 159,441 bp in length. The genome shows a typical quadripartite structure, containing two copies of inverted repeat (IR) regions (26,268 bp), a large-single copy (LSC) region (88,124 bp), and a small-single copy (SSC) region (18,781 bp). Its overall GC content is 38.1%, while the corresponding values of the LSC, SSC, and IR regions are 36.3, 32.9, and 43.1%, respectively. A total of 112 unique genes were annotated, containing 80 protein-coding genes, 28 tRNA genes, and 4 rRNA genes (Table 1).

Seven published plastid genome sequences, belonging to five species in Hamamelidaceae and two outgroups (*Daphniphyllum oldhamii* and *Cercidiphyllum japonicum*), were downloaded from GenBank for inferring the phylogenetic position of *F. sinensis*. A total of 79 common protein-coding genes of all eight plastomes were aligned separately using MAFFT (Katoh and Standley 2013) and concatenated into a single matrix in Geneious. The maximum likelihood (ML) tree based on the combined matrix was reconstructed using RAxML version 8.2.9 (Stamatakis 2014) with a basis of 1000 bootstrap replicates (Figure 1). The result showed the species of Hamamelidaceae formed a monophyletic clade with 100% bootstrap value. As expected, *F. sinensis* and *S. henryi* were recovered as sister with high support (100%), which agreed with the phylogenetic result of Li and Bogle (2001).
Table 1. List of 112 unique genes in the plastome of Fortunearia sinensis.

| Category                              | Groups of genes        | Names of genes |
|---------------------------------------|------------------------|----------------|
| Protein synthesis and DNA replication | Transfer RNAs          | trnA-UGCa<sup>a</sup> | trnC-ACA<sup>a</sup> | trnC-GCA | trnD-GUC |
|                                       |                        | trnE-UUC<sup>a</sup> | trnF-GAA | trnM-CAU | trnM-GCC |
|                                       |                        | trnH-UUG<sup>a</sup> | trnK-UUA<sup>a</sup> | trnP-UAG |
| Ribosomal RNAs                        | rrn4.5                 | rrn5           | rrn16 | rrn23 |
| Ribosomal protein small subunit       | rps2                   | rps3           | rps4 | rps7<sup>b</sup> |
| Ribosomal protein large subunit       | rpl2<sup>a</sup>       | rpl14          | rpl16 | rpl20 |
| Subunits of RNA polymerase            | rpoA                   | rpoB           | rpoC1<sup>a</sup> | rpoC2 |
| Photosynthesis                        | Photosystem I          | psaA           | psaB | psaC | psaL |
|                                       |                        | psaJ           | ycf3<sup>b</sup> |
| Photosystem II                       | psbA                   | psbB           | psbC | psbD |
|                                       | psbE                   | psbF           | psbH | psbI |
|                                       | psbJ                   | psbK           | psbL | psbM |
|                                       | psbN                   | psbT           | psbZ |
|                                       | petA                   | petB<sup>a</sup> | petD<sup>a</sup> | petG |
|                                       | atpA                   | atpB           | atpE | atpF<sup>a</sup> |
|                                       | ndhA<sup>a</sup>       | ndhB<sup>a</sup> | ndhC | ndhD |
|                                       | ndhE                   | ndhF           | ndhG | ndhH |
| Miscellaneous group                   | Large subunit Rubisco  | rbcL           | -    | -    |
|                                       | Translation initiation factor IF-1 | intA | -    | -    |
|                                       | Acetyl-CoA carboxylase | accD           | -    | -    |
|                                       | Cytochrome c biogenesis | cssA          | -    | -    |
|                                       | Maturase               | matK           | -    | -    |
|                                       | ATP-dependent protease | clpP<sup>b</sup> | -    | -    |
|                                       | Inner membrane protein | cemA           | -    | -    |
| Unknown function                      | Conserved open reading frames | ycf1 | ycf2 | ycf4 | ycf1S |

<sup>a</sup>Gene containing a single intron.
<sup>b</sup>Gene containing two introns.

Figure 1. Phylogenetic relationships of Hamamelidaceae inferred from maximum likelihood method based on 79 common protein-coding genes. Daphniphyllum oldhamii and Cercidiphyllum japonicum were used as outgroups. The node labels indicate the ML bootstrap (1000 replicates) support values.
Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This study was financially supported by the National Natural Science Foundation of China [Grant No. 31570210] and the International Partnership Program of Chinese Academy of Sciences [Grant No. GJHZ1620].

References

Dierckxsens N, Mardulyn P, Smits G. 2017. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. 45:e18.

Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull. 19:11–15.

Jia LZ, Zhou J. 1987. Fat-bearing plants in China. Beijing, China: Science Press.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30:772–780.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv. 1303.3997.

Li JH, Bogle AL. 2001. A new suprageneric classification system of the Hamamelidoideae based on morphology and sequences of nuclear and chloroplast DNA. Harv Papers Bot. 5:499–515.

Qin HN, Yang Y, Dong SY, He Q, Jia Y, Zhao LN, Yu SX, Liu B, Yan YH, et al. 2017. Threatened species list of China’s higher plants. Biodivers Sci. 25:696–744.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 30:1312–1313.

Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. 2017. GeSeq - versatile and accurate annotation of organelle genomes. Nucleic Acids Res. 45:W6–W11.

Wang JS. 1996. Hamamelidaceae denoted. Plants. 5:32–34.