

## The complete chloroplast genome sequence of horticultural plant, *Impatiens hawkeri* (Sect. Balsaminacea, *Impatiens*)

Chao Luo<sup>a</sup>, Wulue Huang<sup>a</sup>, Yang Li<sup>a</sup>, Zhixi Feng<sup>a</sup>, Jiapeng Zhu<sup>a</sup>, Yingli Liu<sup>a</sup>, Zhenkai Tong<sup>a</sup>, Yan Liang<sup>a</sup>, Haiquan Huang<sup>a,b</sup> and Meijuan Huang<sup>a</sup>

<sup>a</sup>College of Landscape Architecture and Horticultural Science, Southwest Forestry University, Kunming, China; <sup>b</sup>Research and Development Center of Landscape Plants and Horticulture Flowers, Southwest Forestry University, Kunming, China

### ABSTRACT

The complete chloroplast genome sequence of *Impatiens hawkeri*, a widely cultivated horticultural species in the world is 151,692 bp, with a typical quadripartite structure including a pair of inverted repeat (IRs, 25,584 bp) regions separated by a small single copy (SSC, 17,494 bp) region and a large single copy (LSC, 83,029 bp) region. The overall GC content of *I. hawkeri* plastid genome was 36.8%. The whole chloroplast genome contains 135 genes, including 89 protein-coding genes (PCGs), 38 transfer RNA genes (tRNAs), and 8 ribosomal RNA genes (rRNAs). Among these genes, 15 genes have one intron and 2 genes contain two introns. To investigate its evolution status, the phylogenetic tree based on APGIII reveal that there are close relationships to the same genus species *I. uliginosa* and *I. pinfanensis*.

### ARTICLE HISTORY

Received 4 November 2019  
Accepted 23 November 2019

### KEYWORDS

*Impatiens hawkeri*;  
chloroplast; genome  
sequence; horticultural  
plant

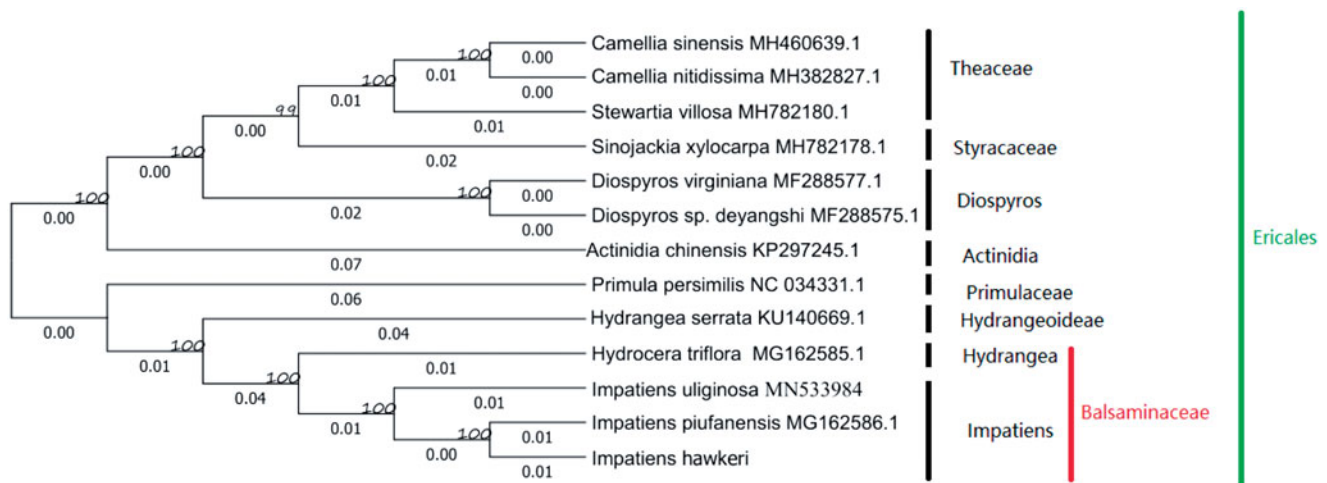
The largest angiosperm genus *Impatiens* including the approximately 1,000 known species, most live in the wild and commonly as understory woodland plants with 250 species in China (Grey-Wilson 1980; Yu et al. 2016). *Impatiens* stems and roots have been isolated and identified in traditional Chinese medicine (TCM) for treatment of a wider range of diseases and ailments, including lumbago, neuralgia, burns and scalds (Janssens et al. 2012; Mohammad et al. 2012; Li et al. 2015). *I. hawkeri* is an annual shade-tolerant flower used extensively in display gardens, landscape beds, container gardens and hanging baskets (Salgado et al. 2018). Though there are many pharmacology studies done on the plant, not much of molecular genetic informatics is currently available (Rahelivololona et al. 2018). For a better understanding of *I. hawkeri*, it is essential to reconstruct a phylogenetic tree of the *Impatiens* species based on high-throughput sequencing approaches (Luo et al. 2019).

The fresh leaves of *I. hawkeri* were obtained from the nursery of Southwest Forestry University (Yunnan, China; Coordinates: 102°76'43"E, 25°06'15"N; Altitude: 1953.7 m). Total genomic DNA was extracted with the Omega Plant Genomic DNA Preps Kit. The certificate specimens of *I. hawkeri* samples were properly deposited at the herbarium of Southwest Forestry University (NO. SWFU-IBXJNY20180811) and DNA samples were properly stored in College of Landscape Architecture and Horticultural Science, Southwest Forestry University, Kunming, Yunnan, China. Total genomic DNA was used to generate libraries with an average insert size of 400 bp and sequences using the Illumina HiSeq X

platform. Approximately 7.68 GB of raw data were generated with 150 bp paired-end read lengths. The complete chloroplast genome used the software of the program Geneious R10 by the raw data with *I. pinfanensis* (GenBank\_MG162586.1) as the reference (Jin et al. 2018). Genome annotation was performed with the DOGMA (Wyman et al. 2004). The cpDNA sequence with complete annotation information was deposited at GenBank database under the accession number MN687854.

The plastome of *I. hawkeri* is 151,692 bp, a double stranded circular DNA with a pair of inverted repeats (IRA and IRB) of 25,584 bp, a large single copy (LSC) region of 83,029 bp and a small single copy (SSC) region of 17,494 bp. The junction between IRB and SSC is 50 bp within the *ndhF* gene. The junction between SSC and IRA is within the *ycf1* gene, which is 5,508 bp long, of which 4327 bp lies in the SSC. The overall GC content of the whole plastome was 36.8%, while the corresponding values of the LSC, SSC, and IR regions were 34.5%, 29.6%, and 43.2%, respectively. The complete chloroplast plastome annotated 136 genes, including 89 protein-coding genes (PCGs), 38 tRNA genes and 8 rRNA genes. Among these genes, 15 genes have 1 intron and 2 genes contain 2 introns.

To analyze the phylogenetic location of *I. hawkeri* 13 chloroplast genome sequences of Ericales were aligned by the MAFFT version 7 software (Kato and Standley 2013). A maximum likelihood method for phylogenetic analysis was made based on GTR + I + G model in the RAxML version 8 program with 1000 bootstrap replicates (Darrriba et al. 2012;



**Figure 1.** The ML phylogenetic tree for *I. hawkeri* based on 12 chloroplast genome sequences of Ericales. Numbers on the nodes are bootstrap values from 1000 replicates. Accession numbers: *Camellia sinensis* (MH460639.1), *Camellia nitidissima* (MH382827.1), *Stewartia villosa* (MH782180.1), *Diospyros virginiana* (MF288577.1), *Diospyros sp.deyangshi* (MF288575.1), *Sinojackia xylocarpa* (MH782178.1), *Actinidia chinensis* (KP297245), *Primula persimilis*(NC034331.1), *Hydrangea serrata*(KU140669.1), *Hydrocera triflora* (MG162585.1), *Impatiens uliginosa* (MN533984) and *Impatiens piufanensis* (MG162586.1).

Stamatakis 2014). The phylogenetic tree was divided into two parts, which is in agreement with previously studies (Li et al. 2018), and reveal that there are close relationships to the same genus species *I.uliginosa* (MN533984), *I. piufanensis* (MG162586.1) and *Hydrocera triflora* (MG162585.1) (Figure 1).

## Disclosure statement

The authors report no conflict of interest.

## Funding

This work was carried out with support of the National Natural Science Foundation of China [31560228; 31860230], Key Research and Development Plan Program of Yunnan Province [2018BB013], Young and Middle-aged Academic and Technical Leadership Training Project of Yunnan [2015HB046; 2018HB024], and Program for Innovative Research Team (in Science and Technology) in University of Yunnan Province.

## References

- Darriba D, Taboada GL, Doallo R, Posada D. 2012. More models, new heuristics and parallel computing. *Nat Methods*. 9(8):772.
- Grey-Wilson C. 1980. *Impatiens* of Africa. Rotterdam: Balkema.
- Janssens SB, Wilson SY, Yuan YM, Nagels A, Smets EF, Huysmans S. 2012. A total evidence approach using palynological characters to infer the complex evolutionary history of the Asian *Impatiens* (Balsaminaceae). *Taxon*. 61(2):355–367.
- Jin JJ, Yu WB, Yang JB, Song Y, Yi TS, Li DZ. 2018. A simple and fast pipeline for de novo assemble of a complete circular chloroplast genome using genome skimming data. *BioRxiv*. 19(1):pii: E1007.
- Katoh K, Standley DM. 2013. Improvements in performance and usability. *Mol Biol Evol*. 30(4):772–780.
- Li Q, Zhang XS, Cao JQ, Guo ZH, Lou YT, Ding M, Zhao YQ. 2015. Depside derivatives with anti-hepatic fibrosis and anti-diabetic activities from *Impatiens balsamina* L. flowers. *Fitoterapia*. 105:234–239.
- Li ZZ, Josphat KS, Andrew WG, Cornelius MK, Wang QF, Chen JM. 2018. Comparative Genomics of the Balsaminaceae Sister Genera *Hydrocera triflora* and *Impatiens pinfanensis*. *Int J Mol Sci*. 19(1):pii: E319.
- Luo C, Huang W, Zhu J, Feng Z, Liu Y, Li Y, Li X, Huang H, Huang M. 2019. The complete chloroplast genome of *Impatiens uliginosa* Franch., an endemic species in southwest China. *Mitochondrial DNA Part B*. 4(2):3846–3847.
- Mohammad ZI, Nazmun N, Saleha A, Md SR. 2012. Antinociceptive activity of methanol extract of flowers of *Impatiens balsamina*. *J Ethnopharmacol*. 142:804–810.
- Rahelivololona EM, Fischer E, Janssens SB, Razafimandimbison SG. 2018. Phylogeny, infrageneric classification and species delimitation in the Malagasy *Impatiens* (Balsaminaceae). *PhytoKeys*. 110:51–67.
- Salgado SC, LeBlanc N, Ismaiel E, Rivera Y, Warfield CY, Crouch JA. 2018. Genetic variation of the pathogen causing impatiens downy mildew predating and including 21st century epidemics on *Impatiens walleriana*. *Plant Disease*. 102(12):2411–2420.
- Stamatakis A. 2014. A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 30(9):1312–1313.
- Wyman SK, Jansen RK, Boore JL. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics*. 20(17):3252–3255.
- Yu S-X, Janssens SB, Zhu X-Y, Liden M, Gao T-G, Wang W. 2016. Phylogeny of *Impatiens* (Balsaminaceae): integrating molecular and morphological evidence into a new classification. *Cladistics*. 32(2): 179–197.