


## The complete chloroplast genome of *Wurfbainia neaurantiaca* (Zingiberaceae: Zingiberea) from Yunnan Province, China

Chunyong Yang<sup>a</sup> , Yanqian Wang<sup>a</sup>, Yanfang Wang<sup>a</sup>, Shuang Li<sup>a</sup>, Ge Li<sup>a</sup>, Zhen Yan<sup>a</sup>, Lixia Zhang<sup>a</sup> and Jianjun Qi<sup>b</sup>

<sup>a</sup>Yunnan Key Laboratory of Southern Medicinal Utilization, Yunnan Branch Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences and Peking Union Medical College, Jinghong, China; <sup>b</sup>Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

### ABSTRACT

*Wurfbainia neaurantiaca* is a medicinal plant endemic to Yunnan Province, China. In this study, its complete chloroplast genome was assembled and characterized. The total genome size of *W. neaurantiaca* was 158,484 bp in length, consisting of a large single-copy region (LSC), a small single-copy region (SSC) and two inverted repeat regions (IRs) with 88,605 bp, 15,285 bp and 29,822 bp, respectively. Its GC content was 36.08%. The chloroplast genome encoded 113 unique genes, including 79 protein-coding, 30 tRNA, and four rRNA genes. The result of the phylogenetic analysis indicated that *W. neaurantiaca* was related to *W. villosa* var. *xanthioides* and supported de Boer's classification that *W. compacta*, *W. longiligularis*, *W. neaurantiaca*, *W. villosa*, *W. villosa* var. *xanthioides* and *Amomum krervanh* belonged to the *Wurfbainia* Clade.

### ARTICLE HISTORY

Received 24 February 2021  
Accepted 30 May 2021

### KEYWORDS

Complete chloroplast genome; phylogenetic analysis; *Wurfbainia neaurantiaca*



*Wurfbainia neaurantiaca* (former Latin names: *Amomum aurantiacum* H. T. Tsai & S. W. Zhao, 1979, *Amomum neaurantiacum* T. L. Wu, K. Larsen & N. J. Turland, 2000) is a perennial medicinal herb classified in the family Zingiberaceae, which is endemic to the Yunnan Province of China (de Boer et al. 2018; Wu and Turland 2001). Its fruits are used as herbal medicine and have been used to treat gastrointestinal diseases (Ma et al. 2018). It is difficult to distinguish *W. neaurantiaca* from its related species such as *Wurfbainia villosa*, *W. villosa* var. *xanthioides* and *W. longiligularis* due to their similar morphological characters. In recent years, the chloroplast genome as a superbarcode has been shown to be an efficient method to distinguish species (Abdullah et al. 2019; Cui et al. 2019; Liang et al. 2019), and phylogenetic analysis of chloroplast genomes has been successfully used to identify *Wurfbainia* species (Cui et al. 2019). In this study, the chloroplast genome of *W. neaurantiaca* was assembled, annotated, characterized, and a phylogenetic analysis was performed to investigate its evolutionary relationship with other *Wurfbainia* species and confirm its phylogenetic position in the family location.

The plants of *W. neaurantiaca* were collected from Jinghong (100°52'44"E; 21°48'56"N), Yunnan Province, China. Voucher specimen (Y20180726015) and its DNA (DNA-Y20180726015) were deposited in the Herbaria and Medicinal Plant Cultivation Research Center of Yunnan Branch Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences and Peking Union Medical College,

respectively ([www.yn-implad.ac.cn](http://www.yn-implad.ac.cn), contacts: Dr. Haitao Li, email contact: [lhtxyl@126.com](mailto:lhtxyl@126.com)). The genomic DNA was extracted from the leaves using the DNeasy Plant Mini Kit (Qiagen, Dusseldorf, Germany). The quality of the DNA was assessed using a Nanodrop One<sup>C</sup> spectrophotometer (Thermo Scientific, Madison, USA) and determined by 1.5% agarose gel electrophoresis.

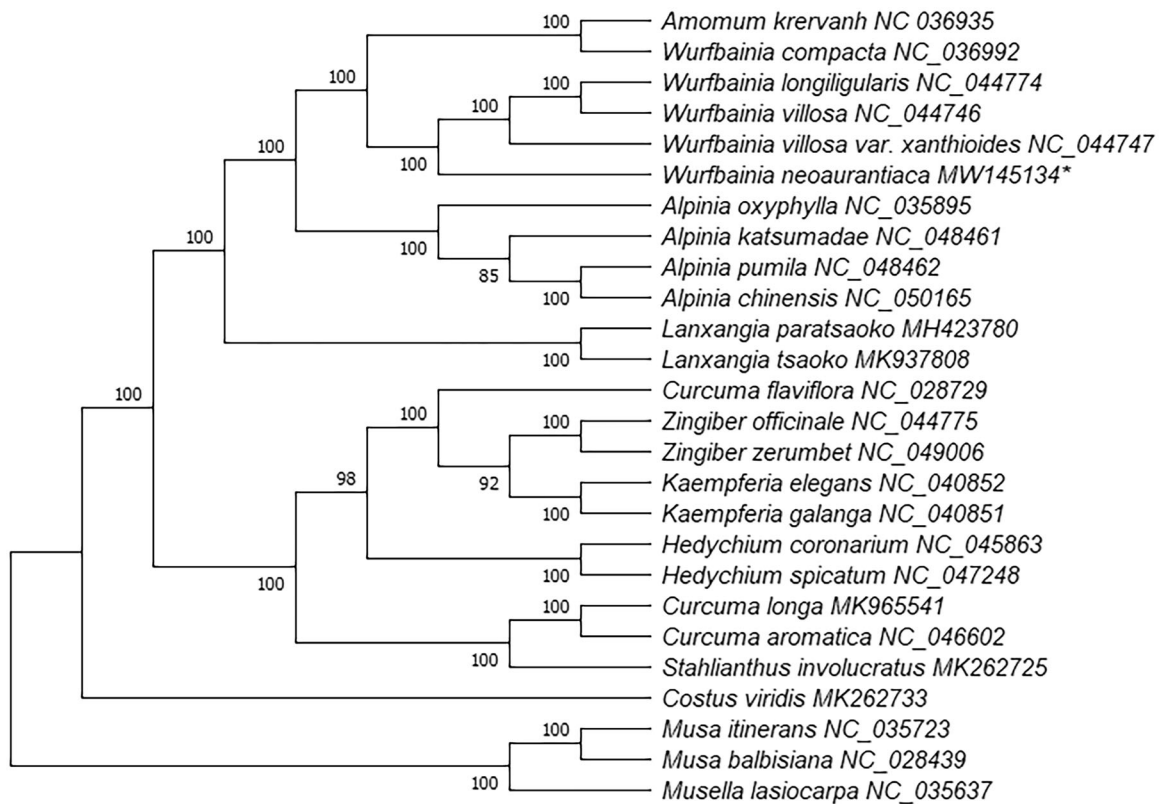
A paired-end library of 350 bp was constructed and sequenced on the NovaSeq system (Illumina, San Diego, USA). A total of 8.5 Gb reads were generated and used to assemble the chloroplast genome using the default settings in NOVOplasty 4.0 (Dierckxsens et al. 2017). Annotation of the chloroplast genome was executed using GeSeq (Tillich et al. 2017) and corrected manually.

The chloroplast genome sequence of *W. neaurantiaca* is 163,534 bp in length and exhibits a general quadripartite structure containing a large single-copy region (LSC), a small single-copy region (SSC) and two inverted repeat regions (IRs) with 88,605 bp, 15,285 bp and 29,822 bp, respectively. The GC content is 36.08%. The chloroplast genome of *W. neaurantiaca* encodes 113 unique genes, including 79 protein-coding, 30 tRNA and four rRNA genes, in which 10 genes (*atpF*, *ndhA*, *ndhB*, *petB*, *petD*, *rpoC1*, *rps12*, *rps16*, *rpl2*, *rpl16*) contain one intron, and two genes (*clpP* and *ycf3*) contain two introns. Relative synonymous codon usage (RSCU) was calculated with CodonW 1.4.4 (<https://sourceforge.net/projects/codonw/files/>) to reveal synonymous codon bias in the

**CONTACT** Lixia Zhang  [87050233@qq.com](mailto:87050233@qq.com)  Yunnan Key Laboratory of Southern Medicinal Utilization, Yunnan Branch Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences and Peking Union Medical College, Jinghong, China

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Figure 1.** A phylogenetic ML tree based on the complete chloroplast sequences of 23 Zingiberaceae species and three Musaceae species.

coding sequences. Except for *trnL*-CAA encoded by UUG and termination codon UAA, 28 amino acid codons with RSCU value > 1 ended with A/U. Simple sequence repeats (SSRs) were mined using MISA-web (<https://web-blast.ipk-gatersleben.de/misa/>), and up to 153 SSRs were detected in the whole chloroplast genome, including 94 mono-, 32 di-, five tri-, 16 tetra-, four penta-, and two hexa-nucleotide SSRs.

The chloroplast genome sequences of 23 species (including *W. neoaurantiaca*) from Zingiberaceae and three outgroup taxa from the Musaceae were aligned with MAFFT 7.307 (Kato and Standley 2013). A phylogenetic tree was constructed using RAxML 8.2.12 (Stamatakis 2014) with 1000 Bootstrap replicates and using the GTR + F + R2 model according to ModelFinder (Kalyanamoorthy et al. 2017). The result indicated that the selected species from the Zingiberaceae were clustered within a lineage distinct from the outgroup. Five *Wurfbainia* species and *Amomum krervanh* formed a monophyletic clade in the Zingiberaceae. *Wurfbainia neoaurantiaca* was fully resolved in a clade with *W. longiligularis*, *W. villosa* var. *xanthioides*, *W. villosa* and more closely related to *W. villosa* var. *xanthioides* (Figure 1). According to the recircumscription of the genus *Amomum*, *W. compacta*, *W. longiligularis*, *W. neoaurantiaca*, *W. villosa*, *W. villosa* var. *xanthioides* and *Amomum krervanh* were classified into the *Wurfbainia* Clade (de Boer et al. 2018). Our results confirm the classification and further illustrate the relationship among *W. neoaurantiaca* and the other five species.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Data availability

The genome sequence data that support the findings of this study are openly available in GenBank of NCBI at <https://www.ncbi.nlm.nih.gov/> under the accession no. MW145134. The associated BioProject, Bio-Sample and SRA numbers are PRJNA699991, SAMN17817191 and SRP304958, respectively.

## References

- Abdullah IS, Furrugh M, Zain A, Muhammad SM, Shahid W, Bushra M, Ibrar A, Mohammad TW. 2019. Comparative analyses of chloroplast genomes among three *Firmiana* species: Identification of mutational hotspots and phylogenetic relationship with other species of Malvaceae. *Plant Gene*. 19:100199.
- Cui YX, Chen XL, Nie LP, Sun W, Hu HY, Lin YL, Li HT, Zheng XL, Song JY, Yao H. 2019. Comparison and phylogenetic analysis of chloroplast genomes of three medicinal and edible *Amomum* species. *IJMS*. 20(16):4040.
- Cui YX, Zhou JG, Chen XL, Xu ZC, Wang Y, Sun W, Song JY, Yao H. 2019. Complete chloroplast genome and comparative analysis of three *Lycium* (Solanaceae) species with medicinal and edible properties. *Gene Rep*. 17:100464.
- de Boer H, Newman M, Poulsen AD, Droop AJ, Fér T, Hiên LTT, Hlavatá K, Lamxay V, Richardson JE, Steffen K, et al. 2018. Convergent morphology in *Alpinieae* (Zingiberaceae): Recircumscribing *Amomum* as a monophyletic genus. *Taxon*. 67(1):6–36.
- Dierckxnsens N, Mardulyn P, Smits G. 2017. NOVOPlasty: *de novo* assembly of organelle genomes from whole genome data. *Nucleic Acids Res*. 45(4):e18.

- Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jeremiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. 14(6):587–589.
- Katoh K, Standley D. 2013. MAFFT multiple sequence alignment software version 7 improvements in performance and usability. *Mol Biol E*. 30(4):772–780.
- Liang CL, Wang L, Lei J, Duan BZ, Ma WS, Xiao SM, Qi HJ, Wang Z, Shen XF, Guo S, et al. 2019. A comparative analysis of the chloroplast genomes of four *Salvia* medicinal plants. *Engineering*. 5(5): 907–915.
- Ma XJ, Zhang LX, Lin YF. 2018. *Dai medicine of China*. Vol 2(260). Beijing: People's. Medical Publishing House.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 30(9): 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(W1):W6–W11.
- Wu DL, Turland LN. 2001. Four new names in Chinese and Vietnamese Zingiberaceae. *Novon*. 10(1):90–91.