

## The complete chloroplast genome sequence of *Gentianopsis grandis* (Harry Sm.) Ma (Gentianaceae) and phylogenetic analysis

Hailing Li<sup>a</sup>, Jingjing Zhao<sup>a,b</sup> and Nong Zhou<sup>a,b</sup>

<sup>a</sup>College of Food and Biology Engineering, Chongqing Three Gorges University, Chongqing, China; <sup>b</sup>The Chongqing Engineering Laboratory for Green Cultivation and Deep Processing of the Three Gorges Reservoir Area's Medicinal Herbs, Chongqing Three Gorges University, Chongqing, China

### ABSTRACT

The high-throughput sequencing technology was used to sequence and assemble the chloroplast genome of *Gentianopsis grandis*, and we analyzed its structural characteristics and phylogenetic relationships. The complete chloroplast genome of *G. grandis* was 151,271 bp in length, consisting of a large single copy (LSC) region of 82,572 bp and a small single copy (SSC) region of 17,907 bp, which were separated by a pair of inverted repeat regions (IRs) of 25,396 bp. The annotation contained a total of 114 unique genes, including 78 protein-coding genes, 30 tRNA genes, four rRNA genes, and two pseudogenes. The phylogenetic analysis indicated the genus *Gentianopsis* was closely related to *Halenia* and *Swertia*.

### ARTICLE HISTORY

Received 22 July 2020  
Accepted 21 June 2021

### KEYWORDS

*Gentianopsis grandis*;  
Gentianaceae; complete  
chloroplast genome;  
phylogenetic analysis

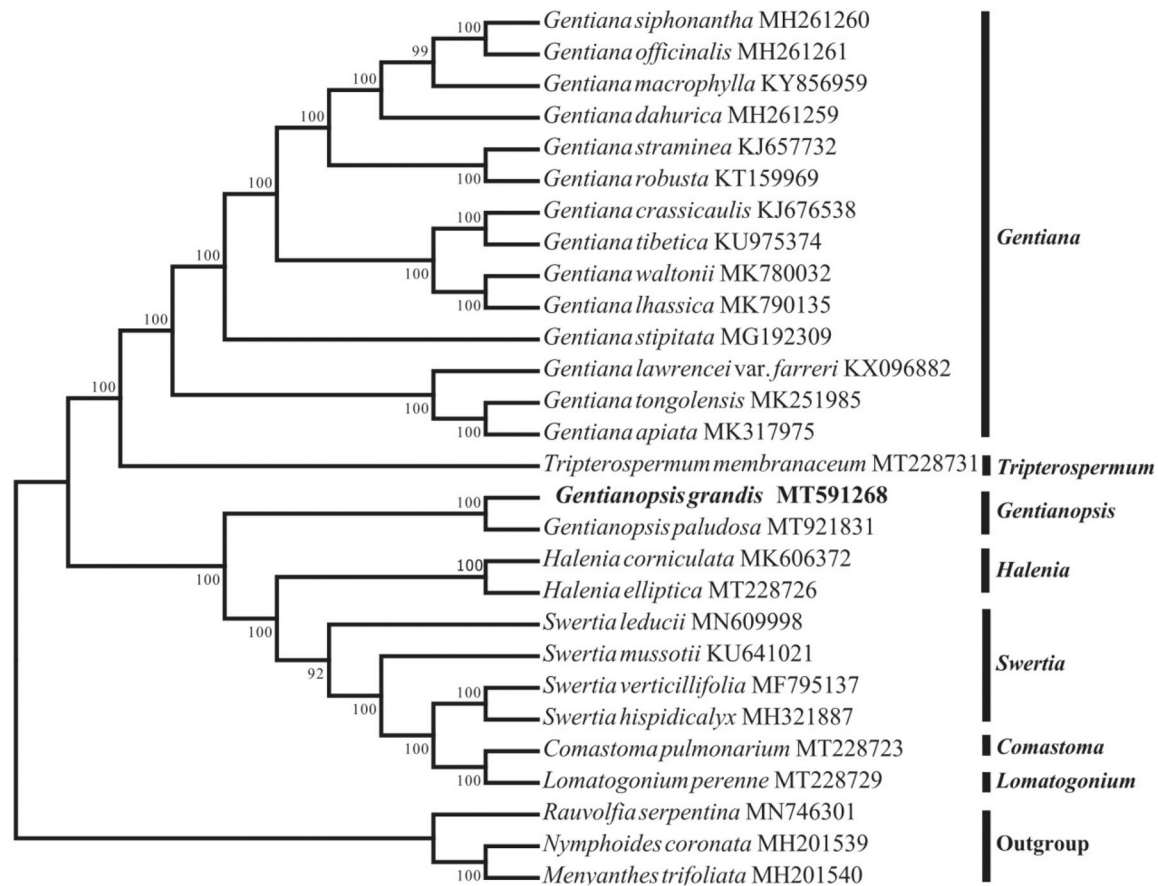
*Gentianopsis grandis* (Harry Sm.) Ma is an annual or biennial herb that belongs to genus *Gentianopsis*, family Gentianaceae. *Gentianopsis* comprises 24 species in the world, and there are five species in China. *G. grandis* is distributed in the northwest of Yunnan province and the southwest of Sichuan province in China (Chen et al. 1999). Some species of *Gentianopsis* were extensively used for clinical treatment of conjunctivitis, hepatitis, nephritis, gastroenteritis, dyspepsia, fever and influenza in China (Lu et al. 2017). Nevertheless, most studies on *Gentianopsis* and its related genus have focused on describing chemical compositions, pharmacological activity and quantitative analysis using high performance liquid chromatography (HPLC) methods. DNA barcode was used to identify *Gentianopsis paludosa* from adulterant species (Xue and Li 2011). And only the chloroplast sequence of *G. paludosa* was submitted to Genbank. There are less researches about molecular biology of *Gentianopsis*. Here, we reported and characterized the complete chloroplast genome sequence of *G. grandis*, which can be used to reveal its phylogenetic relationships with other species of Gentianaceae.

Fresh and clean leaf materials of *G. grandis* were sampled from Songpan county, Sichuan province, China (30° 54' 58.52" N, 103° 26' 39.06" E), and the voucher specimen was deposited at the Herbarium of Medicinal Plants and Crude Drugs of the College of Pharmacy, Dali University (ZDQ17145). The total genomic DNA was extracted via the modified CTAB method (Doyle 1987; Yang et al. 2014). Genome sequencing was performed using Illumina HiSeq 2500 (Novogene, Tianjin, China) platform with the pair-end

(2 × 300bp) library. About 3.61 Gb of raw reads with 10,222,390 paired-end reads were obtained from high-throughput sequencing. Raw data was filtered using Trimmomatic v.0.32 with default settings (Bolger et al. 2014). Subsequently, the trimmed reads were assembled into circular contigs using GetOrganelle (Jin et al. 2020) with *Swertia verticillifolia* (MF795137) as the reference. Finally, the complete chloroplast genome sequences of *G. grandis* were annotated in Geneious (version 11.0.2; <https://www.geneious.com>). The annotated chloroplast genome was submitted to the GenBank with an accession number MT591268.

The complete chloroplast genome of *G. grandis* is 151,271 bp in length with an overall GC content of 37.9%, exhibiting a typical quadripartite structure with a large single copy-region (LSC) of 82,572 bp, a small single copy-region (SSC) of 17,907 bp, and a pair of inverted repeats (IRa and IRb) of 25,396 bp. A total of 114 unique genes were annotated, including 78 protein-coding genes, 30 tRNA genes, four rRNA genes, and two pseudogenes (*rps 16* and *inf A*), of which 18 genes were duplicated in the inverted repeat regions. Among the annotated genes, 14 genes (*trnK-UUU*, *ndhA*, *trnI-GAU*, *trnA-UGC*, *petD*, *rpoC1*, *rpl16*, *atpF*, *petB*, *ndhB*, *trnG-UCC*, *rpl2*, *trnV-UAC* and *trnL-UAA*) contained one intron, whereas another two genes (*ycf3* and *clpP*) possessed two introns.

To investigate the phylogenetic position of *G. grandis*, 28 cp genomes were downloaded from the NCBI database and aligned using MAFFT version 7.149 (Katoh and Standley 2013). A neighbor-joining (NJ) tree was constructed using the MEGA v.7.0.26 (Kumar et al. 2016) with 1000 bootstrap



**Figure 1.** Neighbor-joining (NJ) phylogenetic tree of 25 species within the family Gentianaceae based on the complete chloroplast sequences using *Rauvolfia serpentina*, *Nymphoides coronata* and *Menyanthes trifoliata* as an outgroup.

replicates and three species (*Rauvolfia serpentina* MN746301, *Nymphoides coronata* MH201539 and *Menyanthes trifoliata* MH201540) were selected as an outgroup. The results revealed that the genus *Gentianopsis* was closely related to *Halenia* and *Swertia* (Figure 1). In summary, this study accurately revealed the phylogenetic position of *G. grandis* in the Gentianaceae, which would be beneficial to further phylogenetic studies on the related species or genera in the family.

### Disclosure statement

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

### Funding

This study was supported by National Natural Science Foundation of China [81260622] and Chongqing Science and Technology Commission [cstc2018jcyjAX0267].

### Data availability statement

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. MT591268. The associated BioProject, SRA, and BioSample numbers are PRJNA717403, SRR14076012, and SAMN18499221, respectively.

### References

- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 30(15):2114–2120.
- Chen TQ, Lang KY, Lin Q, Li R. 1999. *Gentianopsis* (Gentianaceae). In: Li PQ, Fu LG, Hong T, editors. Higher plants China. Vol 9. Qingdao (China): Qingdao Publishing House; p. 60–62.
- Doyle J. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull*. 19:11–15.
- Jin JJ, Yu WB, Yang JB, Song Y, dePamphilis CM, Yi TS, Li DZ. 2020. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol*. 21(1):241.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 30(4):772–780.
- Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol*. 33(7):1870–1874.
- Lu NH, Zhao HQ, Jing M, Liu X, Ren CZ, Liu XF, Liu JJ, Zhang YX. 2017. The pharmacodynamic active components study of Tibetan medicine *Gentianopsis paludosa* on ulcerative colitis fibrosis. *Int Immunopharmacol*. 46:163–169.
- Xue CY, Li DZ. 2011. Use of DNA barcode sensu iato to identify traditional Tibetan medicinal plant *Gentianopsis paludosa* (Gentianaceae). *J Syst Evol*. 49(3):267–270.
- Yang JB, Li DZ, Li HT. 2014. Highly effective sequencing whole chloroplast genomes of angiosperms by nine novel universal primer pairs. *Mol Ecol Resour*. 14(5):1024–1031.