



The complete chloroplast genome sequence and phylogenetic analysis of *Nanhaia speciosa* (Fabaceae)

Yufei Xiao^a, Mimi Li^{b,c} , Lianxiang Zhong^a, Yufeng Qin^a, Xiaoning Zhang^a , Qiulan Wei^a, Zihai Qin^a, Ye Zhang^a and Bowen Chen^a

^aGuangxi Forestry Research Institute, Nanning, China; ^bInstitute of Botany, Jiangsu Province and Chinese Academy of Sciences, Nanjing, China; ^cThe Jiangsu Provincial Platform for Conservation and Utilization of Agricultural Germplasm, Nanjing, China

ABSTRACT

Nanhaia speciosa (Fabaceae) is a woody perennial vine used as an important traditionally Chinese medicine. In this study, the complete chloroplast genome of *Nanhaia speciosa* was sequenced and assembled. The chloroplast genome of *N. speciosa* was 132,551bp in length including only one copy of the inverted repeat (IR). It encoded a total of 110 genes, containing 76 protein-coding genes, 30 tRNA and 4 rRNA. The overall GC content was 34.1%. Phylogenetic analysis using a matrix of 69 protein-coding genes illustrated that *N. speciosa* is most closely related to *Wisteriopsis reticulata* of tribe Wisterieae.

ARTICLE HISTORY

Received 7 July 2021
Accepted 14 November 2021

KEYWORDS

Chloroplast genome;
Fabaceae; IR loss;
Leguminosae;
Nanhaia speciosa

Nanhaia speciosa (Champ. ex Benth.) J.Compton & Schrire (2019), commonly name Niudali, is a woody perennial vine belonging to pea family (Fabaceae) distributed mainly in south China and Vietnam (Wei and Pedley 2010). The tuberous roots are used as a traditionally Chinese medicine with hepatoprotection, cough suppressants & expectorant and anti-asthmatic, increasing immunity effects (Yao et al. 2016; 2021). The wild resource of *N. speciosa* have been sharply reduced in recent year by overexploitation. Furthermore, the taxonomy of *N. speciosa* has been controversial. It was previously treated as *Millettia speciosa* Champ. ex Benth and placed in the *Callerya* group of tribe Millettieae (Wei et al. 2010). Now it transferred to the new genus *Nanhaia* of tribe Wisterieae because of differ in its densely pubescent ovaries, larger flower, persistent of floral bracts and gibbosities (Compton et al. 2019). The morphological characters indicated that it is closely related to the genus *Wisteriopsis* (Compton et al. 2019). Therefore, the chloroplast genomic information of *Nanhaia speciosa* presented here is valuable for future classification, phylogenetic and evolution studies in the family Fabaceae.

The fresh mature and healthy leaves of *Nanhaia speciosa* was collected from Guangxi Forestry Research Institute (22°55'30"N, 108°21' E) and the voucher specimen was deposited at Guangxi Forestry Research Institute with number 20210317004 (contact: Bowen Chen, e-mail: bwchen_gfri@163.com). Total genomic DNA was extracted by modified CTAB (hexadecyltrimethylammonium bromide) method (Doyle and Doyle 1987). A paired-end sequencing library was generated using NEB Next® Ultra DNA Library Prep Kit for Illumina (NEB, USA) following manufacturer's instructions

with an insert size of 300 bp. We sequenced the library on an Illumina Hiseq X-ten platform (San Diego, USA) at Novogene Biotech Co., Ltd. (Beijing, China) and at least 2.49 GB of raw sequencing data were obtained.

High-quality reads were assembled into chloroplast genome using the software NOVOPlasty 4.3.1 (Dierckxsens et al. 2017) with *Callerya nitida* (MT120748) as a reference. The genome annotation was applied in GeSeq (Tillich et al. 2017) and manually adjusted the start/stop codons and/or intron/exon boundaries in Geneious 11.1.5 (Kearse et al., 2012). The final complete chloroplast genome sequence was submitted to GenBank databases (National Center for Biotechnology Information, NCBI) through Bankit (<https://submit.ncbi.nlm.nih.gov/about/bankit/>) with accession number MZ028462.

The chloroplast genome of *N. speciosa* was 132,551bp in length including only one copy of the inverted repeat (IR). It encoded a total of 110 genes, containing 76 protein-coding genes, 30 tRNA and 4 rRNA. The overall G+C content was 34.1%. Only one gene (*ycf3*) had two introns, while six tRNA (*trnI*-GAU, *trnK*-UUU, *trnA*-UGC, *trnG*-UCC, *trnV*-UAC, *trnL*-UAA) and 10 protein-coding genes (*rps12*, *rpoC1*, *ndhA*, *ndhB*, *rpl2*, *rpl16*, *petB*, *atpF*, *clpP*, *petD*) had one intron. Three genes (*rps16*, *rpl22* and *infA*) were lost in *N. speciosa*, which was the same as that reported previously for the other legume species (Oyebanji et al. 2020).

A total of 16 cp genome sequences were used for the phylogenetic analysis with 15 plastome data downloaded from GenBank. We performed the alignment in software MAFFT 7.409 (Katoh and Standley 2013) using 69 protein-coding genes. *Glycyrrhiza glabra* (NC_024038) and *G. uralensis* (KU862308) were defined as outgroups. The phylogenetic tree

CONTACT Bowen Chen  bwchen_gfri@163.com  Guangxi Forestry Research Institute, Nanning, China

© 2022 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.

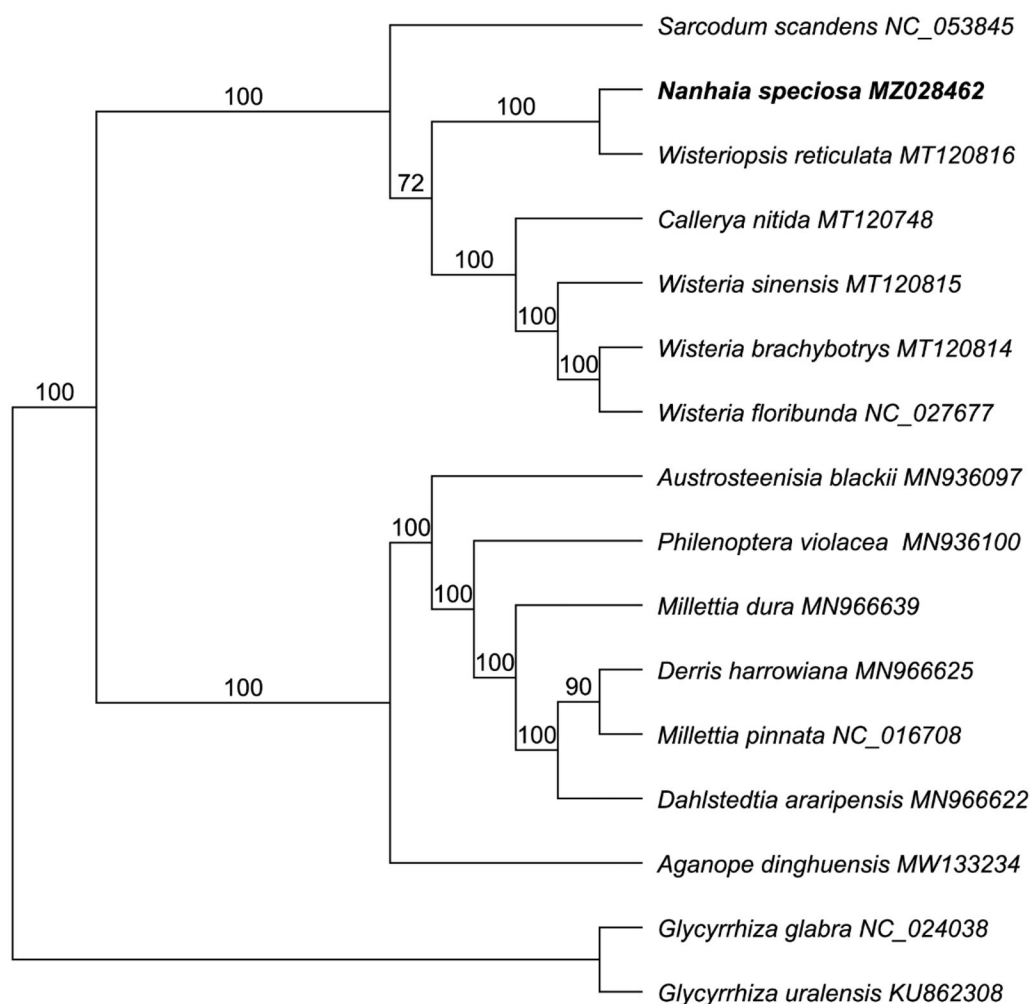


Figure 1. Maximum-likelihood (ML) tree *N. speciosa* and other 15 species. Numbers on the branch indicated the bootstrap values after 1000 replicates.

was constructed by the GTR + GAMMA nucleotide substitution model implemented in RAxML with 1000 bootstrap replicates (Stamatakis 2014). Based on the phylogenetic trees, most of nodes were well resolved with strongly bootstrap values of 72~100% that provided robust topological structure of the selected taxa in Fabaceae. The results illustrated that *N. speciosa* is sister to *Wisteriopsis reticulata* (MT120816) of tribe Wisterieae (Figure 1) with a 100% bootstrap value, which is consistent with previous studies using morphological traits and molecular fragments (Compton et al. 2019). The present study suggested that the chloroplast genome sequences will provide a theoretical basis for future studies of Fabaceae in the evolution and phylogenetic relationships.

Disclosure statement

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

Funding

This work was financial supported by the Open program of Guangxi Key Laboratory of Special Non-wood Forest Cultivation and Utilization (JB-20-02-01).

ORCID

Mimi Li  <http://orcid.org/0000-0002-7883-0325>
Xiaoning Zhang  <http://orcid.org/0000-0003-2105-0560>

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. MZ028462. The associated BioProject, SRA, and Bio-Sample numbers are PRJNA743413, SRR15032818 and SAMN20034749, respectively.

References

- Compton JA, Schrire BD, Könyves K, Forest F, Malakasi P, Mattapha S, Sirichamorn Y. 2019. The *Callerya* group redefined and Tribe Wisterieae (Fabaceae) emended based on morphology and data from nuclear and chloroplast DNA sequences. PK. 125:1–112.
- Dierckxsens N, Mardulyn P, Smits G. 2017. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. 45(4):e18.
- Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull. 19:11–15.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30(4):772–780.

- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28(12): 1647–1649.
- Oyebanji O, Zhang R, Chen S, Yi T. 2020. New insights into the plastome evolution of the Millettoid/Phaseoloid clade (Papilionoideae, Leguminosae). *Front Plant Sci.* 11:151.
- 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.
- Wei Z, Pedley L. 2010. *Callerya*. In: Wu ZY, Raven PH, Hong DY, editors. *Flora of China* 10. Beijing: Science Press; St Louis: Missouri Botanical Garden Press; p. 181–187.
- Wei Z, Chen D, Zhang D, Sun H, Pedley L. 2010. Tribe Millettieae. In: Wu ZY, Raven PH, Hong DY, editors. *Flora of China* 10 Fabaceae. Beijing: Science Press; St Louis: Missouri Botanical Garden Press; p. 165–196.
- Yao S, Bai L, Lan Z, Tang M, Zhai Y, Huang H, Wei R. 2016. Hairy root induction and polysaccharide production of medicinal plant *Callerya speciosa* Champ. *Plant Cell Tiss Organ Cult.* 126(1):177–186.
- Yao S, Lan Z, Huang R, Tan Y, Huang D, Gu J, Pan C. 2021. Hormonal and transcriptional analyses provides new insights into the molecular mechanisms underlying root thickening and isoflavonoid biosynthesis in *Callerya speciosa* (Champ. ex Benth.) Schot. *Sci Rep.* 11(1):9.