

The complete chloroplast genome of *Nageia fleuryi* (Hickel) de Laub. (Podocarpaceae)

Xuelian Yang, Li Yan, Xia Wang, Yongfei Wu, Xiaojing Hu and Shanjun Tian

College of Agricultural, Guizhou University, Guiyang City, Guizhou Province, China

ABSTRACT

Nageia fleuryi (Hickel) de Laub. 1987 belongs to the genus *Nageia* in the family Podocarpaceae and is distributed throughout southeast Asia, including China, Vietnam, and Cambodia. It is a plant with high economic beneficial for food and construction industries. Here, we report on the complete chloroplast (cp) genome of *N. fleuryi* for the first time. The complete cp genome is similar to many gymnosperm plants, however, it lacks inverted repeat regions and does not possess a typical quadripartite structure. The complete cp genome is 133,870 bp in size and the overall guanine-cytosine (GC) content was found to be 37.27%. The total number of genes is 119, including 82 protein-coding genes, 33 tRNA genes, and 4 rRNA genes. Of these, 14 genes contain one intron, two genes contain two introns, and rps12 possessed a trans-splicing mechanism. Finally, the phylogenetic tree demonstrated that *N. fleuryi* is closely related to *Nageia nagi* (AB830885.1 and LC572156.1)

ARTICLE HISTORY

Received 25 January 2022
Accepted 18 June 2022

KEYWORDS

Nageia fleuryi; complete chloroplast genome; phylogenetic analysis; illumina; Podocarpaceae

Belonging to genus *Nageia* in the family Podocarpaceae, *Nageia fleuryi* (Hickel) de Laub. 1987 is a plant with high economic value as a wood and oil resource (Yongbin 2013), also used as a herbal medicine to treat rheumatism and backache (Chen 2007). *N. fleuryi* is distributed throughout Southeast Asia, including China, Vietnam, and Cambodia. To date, the complete chloroplast (cp) genome of *N. fleuryi* has not been reported; therefore, in this study, we sequenced and assembled the complete cp genome of *N. fleuryi* to analyze its phylogenetic relationship for the first time.

With permission, the samples were obtained from Guizhou Botanical Garden, Guiyang, Guizhou Province, China (N 26°37'20", E 106°43'29") and the voucher specimen was deposited at the Laboratory of College of Agriculture in Guizhou University, Guiyang (contact: Xuelian Yang, yx11299927812@outlook.com) under the voucher number CYZ20210701YX. Whole-genome DNA was extracted from 150 mg samples of fresh leaves following a modified CTAB protocol (Doyle 1991). The purified genomic DNA was sheared into c. 350 bp fragments to construct a paired-end (PE) library according to the Nextera XT sample preparation procedures (Illumina, San Diego, CA, USA). The PE reads of 150 bp were generated by a Novaseq 6000 sequencer (Illumina, San Diego, CA, USA). The raw data totaled 3.35 G, and the clean data totaled 3.34 G after quality control processing by NGS QC Toolkit (Patel et al. 2012), and yielded a 98-fold depth of coverage of the cp genome. The guanine-cytosine (GC) content of the clean data was 33.72%, the Q20-value was 97.78%, and the Q30 value was 93.15%, indicating a very high level of data quality for cp genome sequencing

and assembly results. High-quality reads were assembled into the cp genome using the *de novo* assembler SPAdes v.3.11.0 software (Bankevich et al. 2012). Finally, the PGA program (Qu et al. 2019) was used to annotate the cp genome, using the *Nageia nagi* (GenBank accession AB830885) cp genome as the reference. We selected 24 species from NCBI to construct a maximum likelihood (ML) tree. The 77 common protein-coding genes in each complete cp genome of 24 species were aligned with the genes in *N. fleuryi* using MAFFT 7.037 (Katoh and Standley 2013) with the FFT-NS-2 strategy. Then, model-finder var 1.6 was run to select the best-fit model and the TVM++F+I+G4 model was chosen (Kalyaanamoorthy et al. 2017). Finally, RAxML var 8.2.9 was used to construct a phylogenetic tree with 1,000 bootstraps based on the ML method (Alexandros 2014). This study adhered to National Wild Plant Protective Regulations and was approved by the College of Agriculture, Guizhou Province, China.

The complete cp genome of *N. fleuryi* (GenBank accession no. OL435123) differs from Angiospermae, including Asteraceae, Solanaceae, and Rutaceae, because it lacks inverted repeat regions; thus, it is not a typical quadripartite structure. However, it was found to be similar to the complete cp genome of many gymnosperm plants, for example *Metasequoia glyptostroboides* and *Cathaya argyrophylla* (Ching-Ping et al. 2010; Chen et al. 2015). *N. fleuryi* is 133,870 bp in size and the overall GC content was found to be 37.27%. The total number of genes was found to be 119, including 82 protein-coding genes, 33 tRNA genes, and four rRNA genes. Furthermore, 14 genes (*trnK-UUU*, *trnG-UCC*, *atpF*, *rpoC1*, *trnL-UAA*, *trnV-UAC*, *petB*, *petD*, *rpl16*, *rpl2*, *ndhB*,

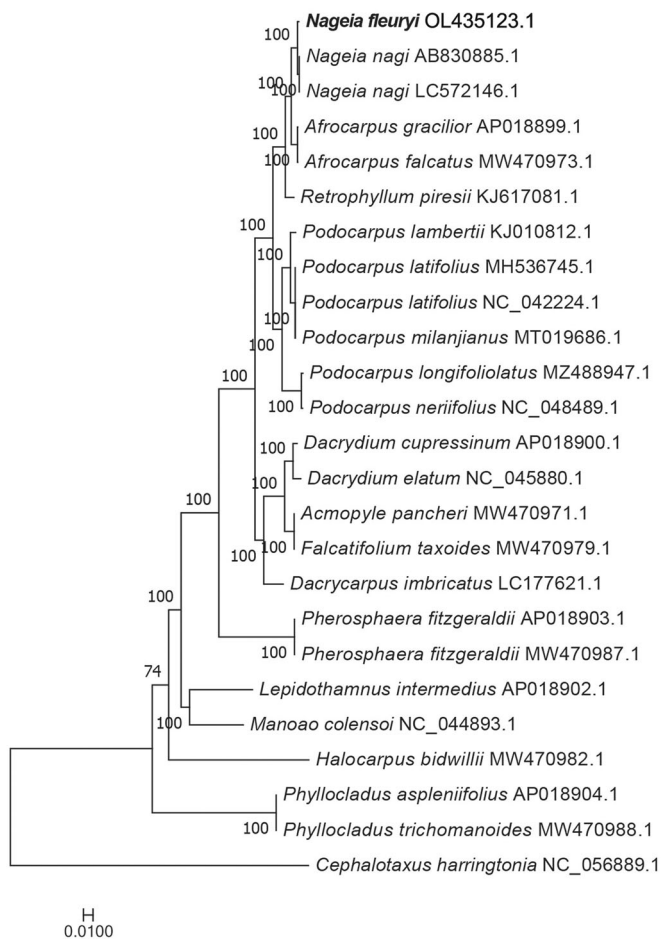


Figure 1. The maximum likelihood phylogenetic tree based on 77 homologous protein-coding genes in 25 species. *Note:* Numbers at the right of nodes represent the support value of 1,000 bootstrap replicates.

trnI-GAU, *trnA-UGC*, and *ndhA*) contain one intron, two genes (*clpP* and *ycf3*) contain two introns, and *rps12* possess a trans-splicing mechanism. The phylogenetic relationship analysis revealed *N. fleuryi* to be closely related to *N. nagi* (AB830885.1 and LC572156.1) (Figure 1). Moreover, the genera *Afrocarpus* and *Retrophyllum* exhibited a close relationship with the genus *Nageia* compared to the other genera. This study provides useful information to aid further studies on species belonging to the genus *Nageia*.

Author contributions

Xue-Lian Yang and Li Yan conceived and designed the research; Xue-Lian Yang, Xia Wang, Yong-Fei Wu, Li Yan, Shanjun Tian, and Xiang-Jing Hu collected the samples, performed the experiments and analyzed the data; Shanjun Tian and Xiang-Jing Hu wrote the draft of the paper and revised

the manuscript. All authors approved the final manuscript and agreed to be accountable for all aspects of the work.

Disclosure statement

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

Funding

This study was supported by: National Fund [31860225]; Qian Kehe Foundation [2019] 1408; Qian Kehe platform personnel [2018] 5781.

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. OL435123. The associated BioProject, SRA, and Bio-Sample numbers are PRJNA786333, SRR17134456, and SAMN23667491, respectively.

References

- Alexandros S. 2014. Raxml version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. (9):1312–1313.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, et al. 2012. Spades: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*. 19(5):455–477.
- Chen L. 2007. The early growth seasonal rhythm and the relationship with climate of different tree species for stand conversion in Dongguan ecological forest. *Guangdong Forest Sci Technol*. 23(1): 28–33.
- Chen J, Hao Z, Xu H, Yang L, Liu G, Sheng Y, Zheng C, Zheng W, Cheng T, Shi J, et al. 2015. The complete chloroplast genome sequence of the relict woody plant *Metasequoia glyptostroboides* Hu et Cheng. *Front Plant Sci*. 6:447.
- Ching-Ping L, Jen-Pan H, Chung-Shien W, Chih-Yao H, Shu-Miaw C. 2010. Comparative chloroplast genomics reveals the evolution of *Pinaceae* genera and subfamilies. *Genome Biol Evol*. 2(1):504–517.
- Doyle J. 1991. DNA protocols for plants: CTAB total DNA isolation. In: Hewitt GM, Johnston A, editors. *Molecular techniques in taxonomy*. Berlin: Springer-Verlag Press; p. 283–293.
- Kalyaanamoorthy S, Minh BQ, Wong TKF, Haeseler AV, Jermin LS. 2017. Modelfinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. 14(6):587–589.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 30(4):772–780.
- Patel RK, Mukesh J, Liu Z. 2012. Ngs qc toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One*. 7(2):e30619.
- Qu XJ, Moore MJ, Li DZ, Yi TS. 2019. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. *Plant Methods*. 15(1):1–12.
- Yongbin WU. 2013. Early growth of 13 species of broad-leaved trees in limestone area of north guangdong. *Guangdong Forest Sci Technol*.