

The complete chloroplast genome of *Prunus campanulata* ‘Fugui’ (Rosaceae)

Zheng-Feng Wang^{a,b,c} , Xue-Cheng Yang^d, Ming-Xuan Zheng^d and Qing Zhou^d

^aGuangdong Provincial Key Laboratory of Applied Botany, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, China;

^bSouthern Marine Science and Engineering Guangdong Laboratory (Guangzhou), Guangzhou, China; ^cCenter for Plant Ecology, Core Botanical Gardens, Chinese Academy of Sciences, Guangzhou, China; ^dCollege of Forestry and Landscape Architecture, South China Agricultural University, Guangzhou, China

ABSTRACT

Prunus campanulata ‘Fugui’ is newly bred cultivar. Here, we report its complete chloroplast genome. The length of the *P. campanulata* ‘Fugui’ chloroplast genome is 157,948 bp, with a large single-copy region of 85,948 bp, a small single-copy region of 19,128 bp and a pair of inverted repeat regions of 26,436 bp each. The genome contains 90 protein-coding genes, 65 transfer RNA genes and 9 ribosomal RNA genes. In addition, the genome contains 67 simple sequence repeats. Phylogenetic analysis revealed that *P. campanulata* ‘Fugui’ is genetically related to previously reported *P. campanulata*.

ARTICLE HISTORY

Received 15 October 2021

Accepted 12 June 2022

KEYWORDS

Chloroplast; genome assembly; next generation sequencing; *Prunus campanulata*

Prunus, a genus of flowering shrubs and trees in the rose family (Rosaceae). It includes more than 400 species worldwide, but only distribute in northern temperate regions. Many of *Prunus* species are economically important, including almond, apricot, cherry, peach and plum. Cherry trees have high ornamental values and renowned for their blossoms in all branches. *Prunus campanulata* (Maxim.) Yu et Li 1883 is one of most charming cherry trees, native to China, Japan and Vietnam. Due to low reproductive barrier, it has many cultivars by breeding. *Prunus campanulata* ‘Fugui’ is newly bred cultivar with pink double flower in China. It has stable morphological traits since cultivated and is considered as an acceptable *Prunus campanulata* cultivar in the market. Due to its beautiful flowers, it is highly demanded and now has been widely planted in southern China as ornamental plant (Supplemental Figure S1). However, due to unknown parentage breeding history, its origin is obscure. To better understanding the origin of *P. campanulata* ‘Fugui’ in the future, therefore, we report its complete chloroplast genome to provide a genomic resource for the phylogeny study.

Fresh leaves of *P. campanulata* ‘Fugui’ were collected from the South China Agricultural University (N23° 09′ 34″, E113° 20′ 23″). A voucher specimen was deposited at the Herbarium of South China Agricultural University (Ming-Xuan Zhen, zhengmx@scau.edu.cn) under the voucher number 32204. The genomic DNA of *P. campanulata* ‘Fugui’ was extracted by the CTAB (cetyltrimethylammonium bromide) method. Using a 2 × 150 bp paired-end sequencing strategy, the extracted DNA was sequenced using the Illumina HiSeq X Ten system (Illumina, San Diego, CA) and about 26 Gb

whole-genome sequencing reads were produced. The sequences were then used to assemble the chloroplast genome of *P. campanulata* ‘Fugui’ by NOVOPlasty 4.3.1 (Dierckxsens et al. 2017). The assembled genome was then annotated with CPGAVAS2 (Shi et al. 2019), GeSeq (Tillich et al. 2017) and PGA (Qu et al. 2019). The annotated genome is now available in GenBank under the accession number MZ727582. Maximum-likelihood phylogenetic analysis was performed in PhyloSuite 1.2.2 (Zhang et al. 2020) and RAxML 8.2.12 (Stamatakis 2014) with the concatenated protein sequences of 73 chloroplast coding genes for *P. campanulata* ‘Fugui’ and the other 12 species. Both *Malus baccata* and *Malus sieboldii* were used as outgroups. The simple sequence repeat (SSR) was identified with MISA-web (Beier et al. 2017). The substitution hotspots estimated by nucleotide diversity (P_i) in the three *P. campanulata* chloroplast genomes (Figure 1) were calculated using DnaSP v6.12.03 (Rozas et al., 2017) by a window size 500 bp and step size 250 bp.

The *P. campanulata* ‘Fugui’ chloroplast genome was 157,948 bp in length with the GC content of 36.72%. The genome showed a large single-copy region of 85,948 bp, a small single-copy region of 19,128 bp and two copies of inverted repeat regions of 26,436 bp each. After annotation, a total of 137 genes were identified in the *P. campanulata* ‘Fugui’ chloroplast genome, including 90 protein-coding genes, 65 transfer RNA genes and 9 ribosomal RNA genes. Phylogenetic analysis revealed that *P. campanulata* ‘Fugui’ was genetically related to previously reported *P. campanulata* (Figure 1).

By aligning with the other *P. campanulata* (MG827394 and MH491529) chloroplast genomes, 63 single nucleotide

CONTACT Qing Zhou  qingzhou@scau.edu.cn  South China Agricultural University, Guangzhou, 510642, Guangdong, China

 Supplemental data for this article is available online at <https://doi.org/10.1080/23802359.2022.2106796>.

© 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.

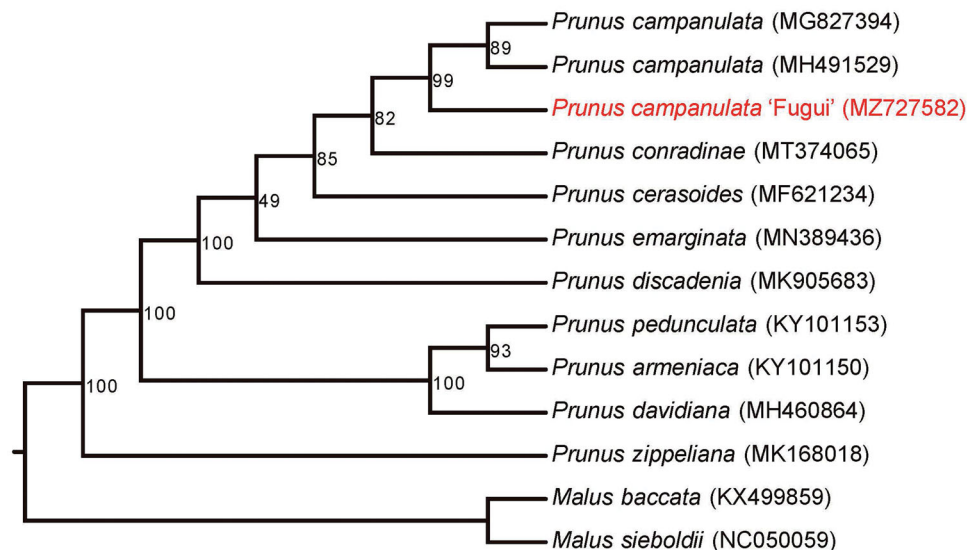


Figure 1. Maximum-likelihood phylogenetic tree for *Prunus campanulata* 'Fugui' and 12 additional species. The GenBank accession numbers are shown in parentheses. The numbers on the branches are bootstrap supported values.

polymorphisms (SNPs) and 42 insertion and deletions (indels) were identified among *P. campanulata* 'Fugui' chloroplast genome. Furthermore, a total of 67 simple sequence repeats (SSR) were discovered in *P. campanulata* 'Fugui' genome. These SSRs included 63 mononucleotides (A/T/G/C), 4 dinucleotides (TA/AT). Compared to the other two *P. campanulata* chloroplast genomes, that both displayed 66 mononucleotides and 4 dinucleotide type SSRs, *P. campanulata* 'Fugui' genome contained less mononucleotide SSRs than them, but the same dinucleotide SSRs to them. *Pi* values ranged from 0.0000 to 0.0120, among them three regions showed high value great than 0.005 within the sites of 42755–43259 bp, 78780–79563 bp and 109602–110351 bp, representing as hot-spots in the genome. These regions contained *rps2*, *ropC2*, *ndhK*, *ndhC* and *rpl14* genes.

Ethical approval

Experiments were performed in accordance with the recommendations of the Ethics Committee of South China Agricultural University. These policies were enacted according to the Chinese National Forestry and Grassland Administration for the plant collection protocols. The study species is not ethically sensitive and endangered.

Authors' contributions

Q.Z. designed the study and funded the genome sequencing; X.C.Y. and M.X.Z. collected the samples and molecular materials; Z.F.W., X.C.Y., M.X.Z. and Q.Z. performed the analyses; Z.F.W. drafted the manuscript; X.C.Y., M.X.Z. and Q.Z. revised the manuscript. All authors provided comments and final approval.

Disclosure statement

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

Funding

This work was funded by Forest Bureau of Guangdong Province [number 2016KJCX023]; Key Special Project for Introduced Talents Team of

Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou) under Grant [number GML2019ZD0408]; The project of Monitoring of Rare and Endemic Plants in Guangdong Province from Forest Resources Conservation Center of Guangdong Province; The Strategic Priority Research Program of the Chinese Academy of Sciences [XDA13020601].

ORCID

Zheng-Feng Wang  <http://orcid.org/0000-0002-7599-7891>

Data availability statement

The genome sequence data that support the finding of this study are openly available in GenBank of NCBI at [<https://www.ncbi.nlm.nih.gov>] (<https://www.ncbi.nlm.nih.gov>) under the accession number MZ727582 and is also accessible at <https://doi.org/10.13140/RG.2.2.16436.78722>. The associated BioProject, SRA, and Bio-Sample numbers for reads are PRJNA752956, SRR15371665 and SAMN20668218 respectively.

References

- Beier S, Thiel T, Münch T, Scholz U, Mascher M. 2017. MISA-web: a web server for microsatellite prediction. *Bioinformatics*. 33(16):2583–2585.
- Dierckxens N, Mardulyn P, Smits G. 2017. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res*. 45(4):e18.
- Qu XJ, Moore MJ, Li DZ, Yi TS. 2019. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. *Plant Methods*. 15:50.
- Rozas J, Ferrer-Mata A, Sanchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A. 2017. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol Biol Evol*. 34(12):3299–3302.
- Shi L, Chen H, Jiang M, Wang L, Wu X, Huang L, Liu C. 2019. CPGAVAS2, an integrated plastome sequence annotator and analyzer. *Nucleic Acids Res*. 47(W1):W65–W73.
- 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.
- Zhang D, Gao F, Jakovlić I, Zou H, Zhang J, Li WX, Wang GT. 2020a. PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Mol Ecol Resour*. 20(1):348–355.