

The complete plastid genome of *Phoenix canariensis* Chabaud (Arecaceae) and phylogenetic analysis

Bowen Zou, Wanwan Long and Li He Yang Wu

College of Life Science, Jingtangshan University, Ji'an, Jiangxi, China

ABSTRACT

Phoenix canariensis Chabaud is a vital ornamental and widely planted in the urban landscape of China. In this study, we reported the complete chloroplast genome (cpDNA) of *P. canariensis*, which is 158,477 bp in length, including a large single copy region (LSC) of 86,189 bp, a small single copy region (SSC) of 17,704 bp, and a pair of inverted repeats (IR) regions of 27,292 bp inserted between LSC and SSC. 132 genes are encoded, including 86 protein-encoding genes, 8 ribosomal RNA genes, and 38 transfer RNA genes. The overall GC content of the chloroplast genome is 37.22%, wherein the corresponding values in the LSC, SSC, and IR regions are 35.3%, 30.79%, and 42.35%, respectively. Phylogenetic analysis showed that *P. canariensis* is sister to *P. dactylifera* with strong bootstrap support.

ARTICLE HISTORY

Received 9 September 2020
Accepted 8 November 2020

KEYWORDS

Arecaceae; *P. canariensis*;
plastid genome;
phylogenetic

Phoenix canariensis Chabaud, Arecaceae, is one of the date palm species that grows in countries around the Arabian Gulf. Genus *Phoenix* contains about 19 species, distributed in southern Africa and Asia (Adawy et al. 2002; Jain et al. 2011). *Phoenix canariensis* Chabaud is used in an ornamental landscapes around the world (Antonio et al. 2017). It is also present in other continents in areas where the winter is temperate (Rousseau et al. 1999). For example, it is widely planted in parks and streets of China (Lan et al. 2020). It hybridizes naturally with the *Phoenix dactylifera* (date palm), but *P. canariensis* is stronger and has bigger leaves (Antonio et al. 2017). The plastid genome is valuable in plant systematics research due to its highly conserved structures, uniparental inheritance, and haploid nature (Fu et al. 2016). Plastid genome have also been smartly engineered to confer useful agronomic traits and/or serve as bioreactors (Jin and Daniell 2015). As one of the world's most popular ornamental palms, the complete plastid genome of *P. canariensis* was sequenced, which will provide genomic and genetic sources for further research.

In this study, the complete plastome sequence of *P. canariensis* was reported and characterized. The fresh leaf sample of *P. canariensis* was acquired from one individual in Jingtangshan University (N 27°06'45.46", E 115°01'55.84"), Ji'an, Jiangxi Province of China. The voucher specimen was kept in Key Laboratory of Ecological Environment and Resource Utilization, Jingtangshan University (JGSU20190826). Total DNA of *P. canariensis* was extracted from the fresh, young leaves (about 1.5 g) with a modified CTAB method (Doyle and Doyle 1987). The DNA library was prepared with a TruSeq DNA Sample Prep Kit (Illumina, USA) according to the instructions of the manufacturer. A genomic

DNA library with an insert size of 400 bp was constructed and then sequenced by an Illumina HiSeq 2500 system. Approximately 5 Gb of sequences data were extracted from the total sequencing output and input into Organelle PBA (Soorni et al. 2017) to assemble the plastid genome. Annotation of the plastid genome was performed using the Dual Organellar GenoMe Annotator (DOGMA) online tool (Wyman et al. 2004) and Geneious v11.1.5 (Kearse et al. 2012) with *Washingtonia robusta* (GenBank: NC_029974.1) as the reference, then manually verified and corrected by comparison with *Phoenix dactylifera* (CM018784.1). Finally, we obtained a complete plastid genome of *P. canariensis* and submitted to GenBank with accession number (MT937173).

The chloroplast genome of *P. canariensis* shows a typical quadripartite structure of 158,477 bp in full length, consisting of a large single copy region (LSC) of 86,189 bp, a small single copy region (SSC) of 17,704 bp, and a pair of inverted repeats (IR) regions of 27,292 bp inserted between LSC and SSC. The total GC content of plastome is 37.22%, with the corresponding values of 35.3%, 30.79%, and 42.35% in the LSC, SSC, and IR regions, respectively. The chloroplast genome comprises 132 genes, including 86 protein-coding genes, 8 ribosomal RNA genes, and 38 transfer RNA genes. 18 genes occurred in the IR region have two copies, including 6 protein-coding genes (*rpl2*, *rpl23*, *rps7*, *rps19*, *ndhB*, and *ycf2*), 8 tRNA genes (*trnH-GUG*, *trnM-CAU*, *trnL-CAA*, *trnV-GAC*, *trnI-GAU*, *trnA-UGC*, *trnR-ACG*, *trnN-GUU*), and 4 rRNA genes (*rrn16*, *rrn23*, *rrn4.5*, *rrn5*). There are 111 unique genes, among which 16 genes have one intron, and two genes, *ycf3* and *clp P*, have two introns.

To analyze the phylogenetic position of *P. canariensis*, the maximum likelihood phylogenetic tree was generated based

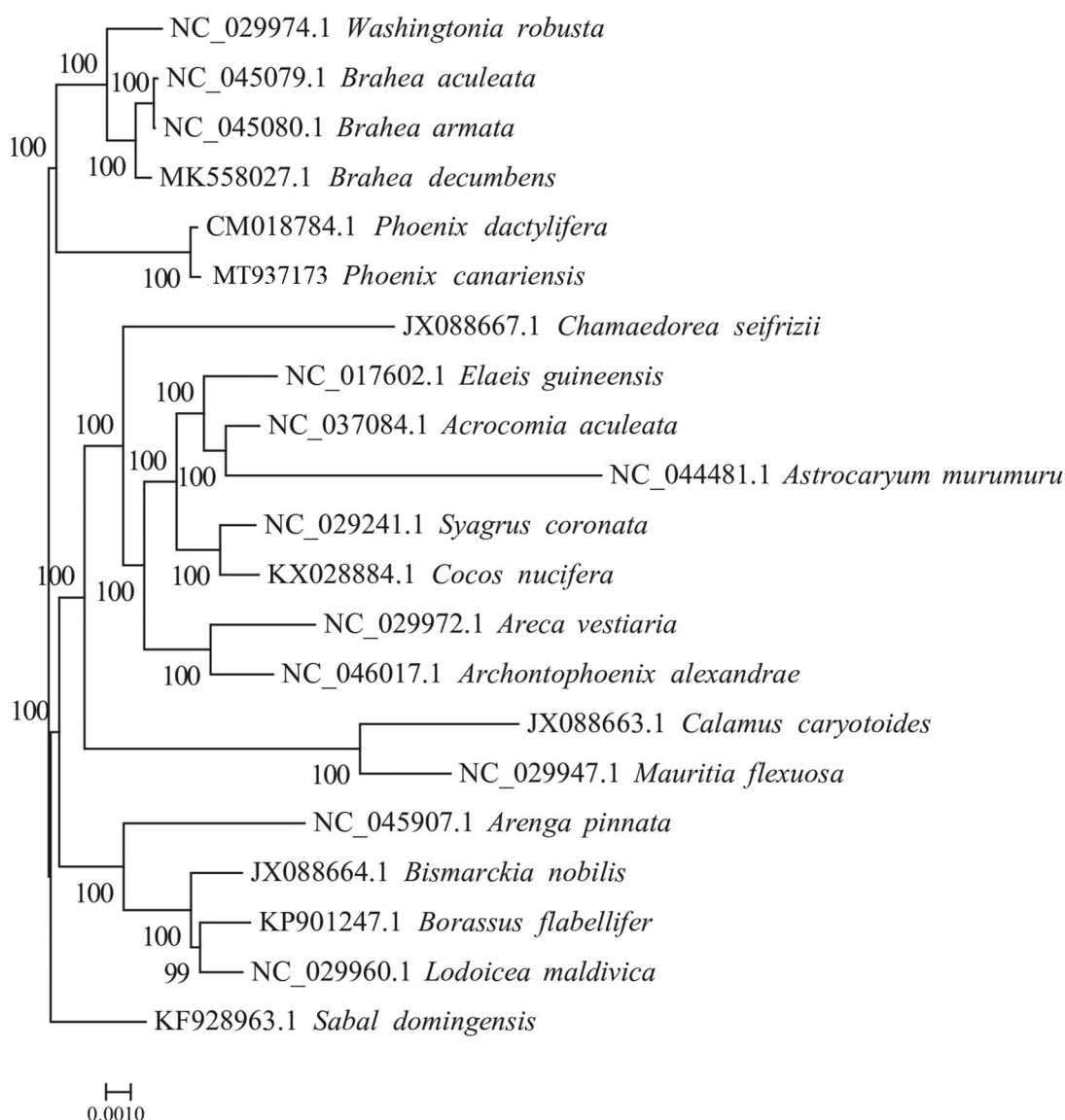


Figure 1. The maximum-likelihood (ML) tree based on the 20 representative plastid genomes of the Areaceae. Numbers near the nodes mean bootstrap support value.

on the plastid genome of *P. canariensis* and other 19 species of the Areaceae. Alignment was conducted using MAFFT v7.307 (Kato and Standley, 2013). The phylogenetic tree was built using RAxML (Stamatakis 2014) with bootstrap set to 1000. *Sabal domingensis* (KF928963.1) served as the outgroup. The maximum likelihood (ML) phylogenetic tree suggested that *P. canariensis* is closely to *P. dactylifera* (Figure 1). The completion and characterization of the complete plastid genome sequence in this study provided helpful molecular resource for future phylogenetic and evolutionary studies of the valuable tree *P. canariensis*.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research was supported by the Science and Technology Project of the Education Department of Jiangxi Province [GJJ190544] in China.

Data availability statement

The data that support the findings of this study are openly available in GenBank of NCBI at <https://www.ncbi.nlm.nih.gov>, reference number MT937173.

References

- Adawy SS, Hussein EH, Ei-Khishin D, Moharam H, Ei-Itriby HA. 2002. Genetic variability studies and molecular fingerprinting of some Egyptian date palm (*Phoenix dactylifera* L.) cultivars:II. RAPD and ISSR profiling. Arab J Biotechnol. 5:225–236.
- Antonio FG, Manuel FV, Clara EFG, Teresa GO, Maria TFG. 2017. Potential use of *Phoenix canariensis* biomass in binderless particleboards at low temperature and pressure. BioResoures. 12(3):6698–6712.
- Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure from small quantities of fresh leaf tissues. Phytochem Bull. 19:11–15.
- Fu PC, Zhang YZ, Geng HM, Chen SL. 2016. The complete chloroplast genome sequence of *Gentiana lawrencei* var. *farreri* (Gentianaceae) and comparative analysis with its congeneric species. PeerJ. 4:e2540
- Jain SM, Ai-Khayri JM, Johnson DV. 2011. Date Palm Biotechnology. Springer Science and Business Media.

- Jin S, Daniell H. 2015. The engineered chloroplast genome just got smarter. *Trends Plant Sci.* 20(10):622–640.
- 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, et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics.* 28(12):1647–1649.
- Lan CZ, Dai YL, Gan L, Yao JA, Yang XJ. 2020. First report of *Fusarium oxysporum* f. sp. *canariensis* causing Fusarium wilt on Canary Island date palm (*Phoenix canariensis* Hort. *J Plant Pathol.* 102(2):569–570.
- Rousseau M, Monfort S, Ferry M. 1999. In vitro vegetative propagation of the Canary Islands Palm (*Phoenix canariensis*). *Acta Hortic.* 486:155–158.
- Soorni A, Haak D, Zaitlin D, Bombarely A. 2017. Organelle_PBA, a pipeline for assembling chloroplast and mitochondrial genomes from PacBio DNA sequencing data. *BMC Genomics.* 18(1):49.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 30(9):1312–1313.
- Wyman SK, Jansen RK, Boore JL. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics.* 20(17):3252–3255.