

## The characteristics of the complete chloroplast genome for *Eucalyptus robusta* (Myrtaceae)

Yisen Gao<sup>a</sup>, Jiayi Zhang<sup>b,c</sup>, Haining Tang<sup>b,c</sup>, Nuoyi Liu<sup>b,c</sup>, Guiying Li<sup>d</sup> and Daran Yue<sup>c</sup>

<sup>a</sup>College of Computer Science and Technology, Changchun University of Science and Technology, Changchun, China; <sup>b</sup>College of Forestry, Hainan University, Haikou, China; <sup>c</sup>Key Laboratory of Genetics and Germplasm Innovation of Tropical Special Forest Trees and Ornamental Plants, Engineering Research Center of Rare and Precious Tree Species in Hainan Province, College of Forestry, Hainan University, Haikou, China; <sup>d</sup>Rubber Research Institute, Chinese Academy of Tropical Agricultural Science, Danzhou, China

### ABSTRACT

*Eucalyptus robusta* is a great tree of Myrtaceae. It is mainly distributed in Southeast provinces of China and Southeast Asian countries. There is no study on the genome of *E. robusta* far. Here we report the complete chloroplast genome of *E. robusta*, assembled from whole-genome high-throughput sequencing data, as a resource for future studies on the taxonomy and evolution of *E. robusta*. The complete chloroplast genome of *E. robusta* is 160,201 bp in length with a typical quadripartite structure, consisting of a large single-copy region (LSC, 88,905 bp), a single-copy region (SSC, 18,498 bp) and a pair of inverted repeats (IRs, 26,399 bp). It was predicted to contain a total of 128 genes, with an overall GC content of 36.86%. Phylogenetic analysis placed *E. robusta* closest to *Eucalyptus globulus*.

### ARTICLE HISTORY

Received 13 September 2021  
Accepted 6 November 2021

### KEYWORDS

*Eucalyptus robusta*;  
plastome; phylogeny;  
genome structure;  
Myrtoideae



*Eucalyptus robusta* is native to Australia, Indonesia, the Philippines and other countries. By 2019, the total area of *E. robusta* plantations in China has reached 580 hm<sup>2</sup>, mainly distributed in Guangdong, Hainan, Yunnan and other regions of China (Tang and Chen 2020). *Eucalyptus robusta* has become a strategic species of fast-growing and high-yield forest in south China. However, a recent study on populations of both varieties using nuclear microsatellite markers found low genetic differentiation between them, suggesting that the current taxonomic treatment may not hold (Song et al. 2016). Also, to our knowledge, there have been no studies on the genome of *E. robusta* up to now. To provide a rich genetic information and improve *E. robusta* molecular breeding in the future, we report and characterize the complete plastid genome sequence of *E. robusta* (GenBank accession number: MZ670598).

The mature and healthy leaves of a single individual of *E. robusta* was sampled from Geze village in Lingshui county, Hainan province (18.63667°N, 109.97111°E). The voucher specimen was deposited in the Herbarium of Guangxi Institute of Botany (accession number: IBK00397639; e-mail address: [karst@gxib.cn](mailto:karst@gxib.cn)). The total genomic DNA was extracted from silica gel dried leaves using a modified CTAB method (Doyle and Doyle 1987) and sequenced based on the Illumina pair-end technology. Approximately 11.68 Gb of paired-end (150 bp) sequence data, deposited in SRA database with Accession number SRR15315685 under the Bioproject NO. PRJNA751248 and

Biosample NO. SAMN20513345, was used in GetOrganelle (Jin et al. 2020) software to assemble the chloroplast genome. Annotation of the chloroplast genome was performed using the cpgeval and Geseq (Tillich et al. 2017), then manually verified and corrected by comparison with the annotation of *Eucalyptus globulus* (AY780259.1) as the reference.

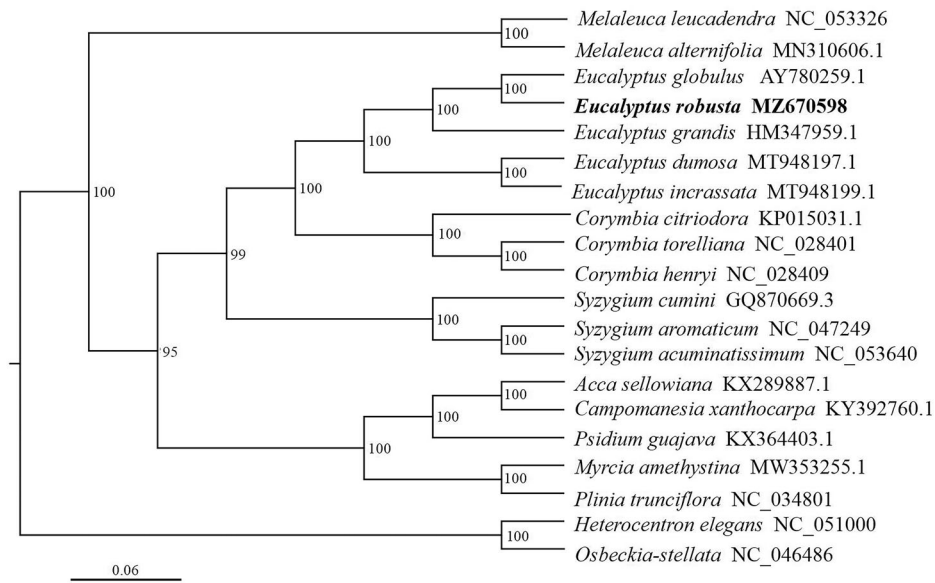
The complete chloroplast genome sequence of *E. robusta* obtained in this study was 16,201 bp in length, with a small single-copy (SSC) region of 18,498 bp, a large single-copy (LSC) region of 88,905 bp, separated by two inverted repeat (IR) regions of 26,399 bp each. The overall G/C content in the plastome of *E. robusta* is 36.86%. It was predicted to contain 128 genes, including eight rRNA genes, 37 tRNA genes, and 83 protein-coding genes.

To investigate the relationship between *Eucalyptus* and other genera within the family Myrtoideae, a phylogenetic tree was constructed. We used RAxML (Stamatakis 2006) with 1000 bootstraps under the GTRGAMMAI substitution model to reconstruct a maximum likelihood (ML) phylogeny of 20 published complete plastomes of Myrtoideae, using *Osbeckia stellate* (GenBank accession NC\_046486) and *Heterocentron elegans* (GenBank accession NC\_051000) (Melastomataceae) as outgroups. According to the phylogenetic topologies, *E. robusta* was closely related to *E. globulus* (GenBank accession AY780259). Most nodes in the plastome ML trees were strongly supported (Figure 1). The complete plastome sequence of *E. robusta* will

**CONTACT** Daran Yue  [1977151736@qq.com](mailto:1977151736@qq.com)  Key Laboratory of Genetics and Germplasm Innovation of Tropical Special Forest Trees and Ornamental Plants, Engineering Research Center of Rare and Precious Tree Species in Hainan Province, College of Forestry, Hainan University, Haikou, China

© 2021 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 based on 20 complete chloroplast genomes. The number on each node indicates the bootstrap value.

provides a useful resource for the conservation genetics of this species as well as for the phylogenetic studies for Myrtoideae.

### Disclosure statement

The authors declare that there is no conflict of interest regarding the publication of this article. The authors alone are responsible for the content and writing of the paper.

### Funding

This work was supported by the Scientific Research Startup Fund Project of Hainan University [KYQD (ZR) 1986] and the Natural Science Foundation of Hainan Province [319QN168].

### 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/> with the accession number is MZ670598. Raw sequencing reads used in

this study have been deposited in the SRA database of NCBI under accession number SRR15315685. The associated 'BioProject,' and 'Bio-Sample' numbers are PRJNA751248, and SAMN20513345 respectively.

### References

- Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull.* 19:11–15.
- Jin J-J, Yu W-B, Yang J-B, Song Y, dePamphilis CW, Yi T-S, Li D-Z. 2020. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol.* 21(1):241.
- Song Z, Zhang M, Li F, Weng Q, Zhou C, Li M, Li J, Huang H, Mo X, Gan S. 2016. Genome scans for divergent selection in natural populations of the widespread hardwood species *Eucalyptus grandis* (Myrtaceae) using microsatellites. *Sci Rep.* 6(1):34941.
- Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics.* 22(21):2688–2690.
- Tang QH, Chen YP. 2020. Status quo and prospect of wood processing and utilization of eucalyptus in China and abroad. *J Chin Wood Based Panel.* 27(6):18–21.
- 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):W6–11.