

Characterization of chloroplast genome of *Eleusine coracana*, a highly adaptable cereal crop with high nutritional reputation

Li-Ying Feng and Li-Zhi Gao

Institution of Genomics and Bioinformatics, South China Agricultural University, Guangzhou, China

ABSTRACT

Eleusine coracana (L.) Gaertn. is a kind of highly adaptable cereal crop with a high nutritional value with the reputation of 'black pearl'. In this study, we sequenced, assembled and characterized the complete chloroplast genome of the grass species. The circular genome of *E. coracana* was 135,137 bp in length, which comprised two inverted repeat (IRa and IRb) regions of 20,919 bp in length separated by a large single copy (LSC) region of 80,663 bp and a small single copy (SSC) region of 12,636 bp. The total GC content of the *E. coracana* chloroplast genome was ~38.13%. A total of 108 functional genes were predicted, including 76 protein-coding genes, 28 tRNA genes, and four rRNA genes. Our phylogenomic analysis of all protein-coding genes further revealed that *E. coracana* is closely related to *Bouteloua curtipendula* and *B. gracilis*, and they are together positioned in the subfamily Chloridoideae clade of the grass family.

ARTICLE HISTORY

Received 2 June 2021
Accepted 14 August 2021

KEYWORDS

Eleusine coracana;
chloroplast genome;
Poaceae; phylogeny

Eleusine coracana (L.) Gaertn., an annual herbaceous grass, is a staple food crop in the semi-arid tropics. Its grains can be utilized to brew beer, and its straw can be used for the preparation and papermaking raw materials (Hanna 1995). In China, *E. coracana* has long served as a traditional Chinese medicine, such as the treatment of stomach disease. It is a good medicinal plant with high nutritional and antioxidant capacities (Kumar et al. 2016; Sharma et al. 2017). *E. coracana* is rich in calcium, fiber, and iron, with excellent malt quality but low glycemic index (GI), and thus, it is the preferred food for diabetic patients (Pradhan et al. 2010).

In this study, *E. coracana* plants were collected in the suburbs of Guangzhou City (22° 48' 13" N, 113° 33' 25" E), Guangdong Province, China. A specimen was deposited at SCAU (the herbarium of the College of Agriculture, South China Agricultural University <https://nxy.scau.edu.cn>, Li-Zhi Gao, SCAUgenomics@163.com), China, under the voucher number SCAU 2020180. The fresh and healthy leaves were sampled and then dried with a silica gel drying method. The chloroplast genomic DNA was extracted using the improved method (CTAB) (Shi et al. 2012). For the genome sequencing, ~3.4 Gb of Illumina sequencing data with 250 bp insert size were generated on the HiSeq X Ten platform. Using the closely related species, *E. coracana* (GenBank accession number: NC_030486) as a reference, the obtained clean reads were then assembled using GetOrganelle v1.5 (Jin et al. 2020). Subsequently, gene prediction and annotation were performed by CPGAVAS2 (Shi et al. 2019).

The complete chloroplast genome of *E. coracana* was 135,137 bp in size, comprising two inverted repeat regions (IRs) with a total of 41,838 bp in size, which are split by a large single copy (LSC) with 80,663 bp and small single copy (SSC) with 12,636 bp in length. The chloroplast genome contained 108 functional genes, including 76 protein-coding genes, 28 tRNAs, and four rRNAs, among which nine genes had introns. Most of the genes are single-copy genes. However, *ycf3* contained two introns in the *E. coracana* chloroplast genome, and an intron was observed in the nine protein-coding genes (*rpl2*, *rpl16*, *rps12*, *rps16*, *ndhA*, *ndhB*, *petB*, *petD*, *atpF*) and three tRNA genes. A total of 18 genes were repeated in the IR regions, including four rRNA genes, six protein-coding genes, and eight tRNA genes. The overall GC content of the *E. coracana* chloroplast genome was ~38.13% with the corresponding values of 36.09%, 32.34%, and 44.00% in the LSC, SSC, and IR regions, respectively.

To determine the phylogenetic position of *E. coracana* in the grass family, 34 grass chloroplast genomes together with *Cyperus rotundus* from Cyperaceae were separately downloaded from GenBank. Phylogenomic analysis was performed by incorporating the *E. coracana* chloroplast genome obtained in this study. All protein-coding gene sequences were aligned with MAFFT 7.409 (Katoh et al. 2002). Using *C. rotundus* as outgroup phylogenetic tree was reconstructed using the maximum-likelihood method using RAXML (Stamatakis 2014) based on 1000 bootstrap replicates. Our

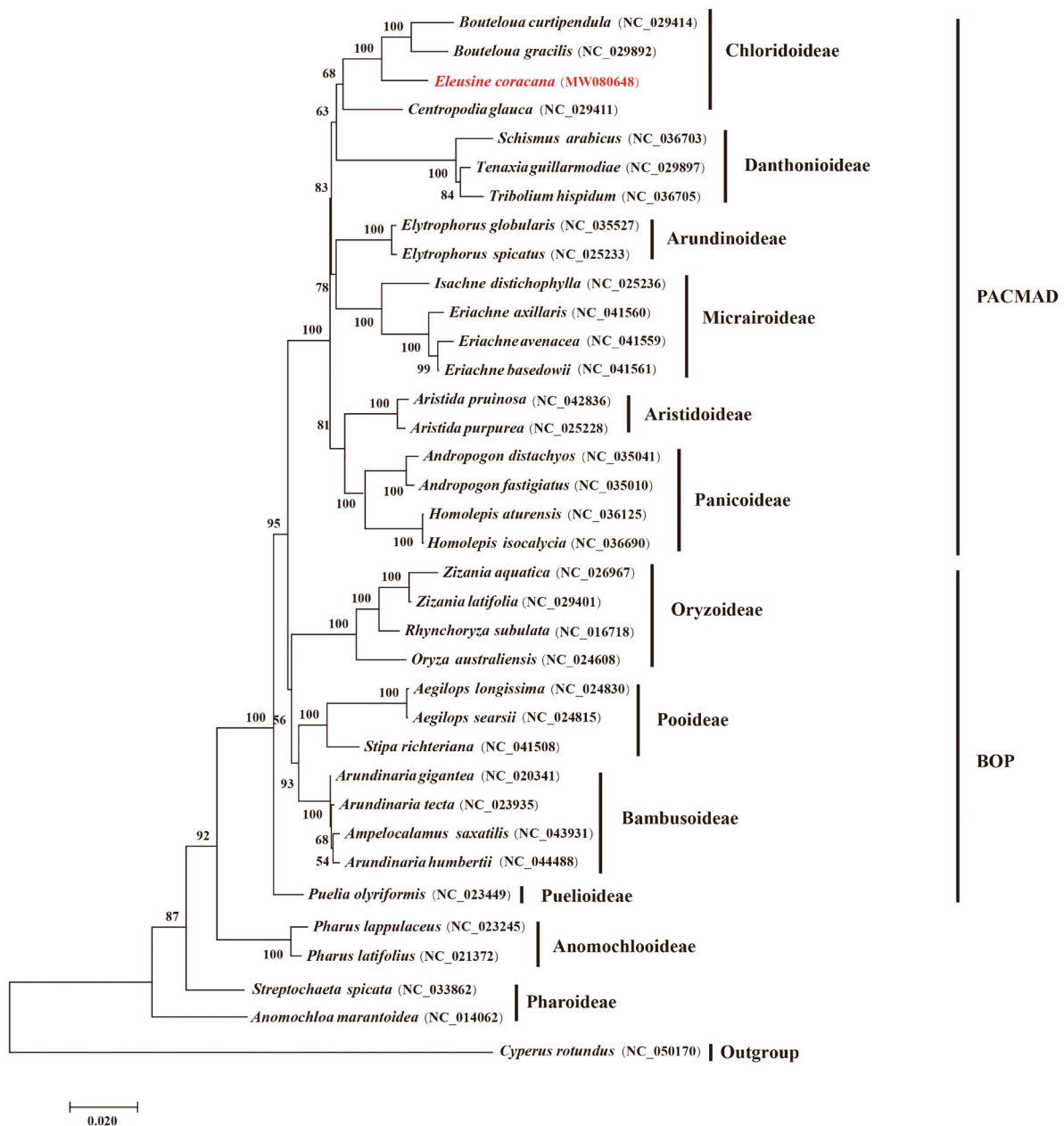


Figure 1. Maximum-likelihood phylogenetic tree based on all protein-coding genes of the 35 grass complete chloroplast genomes using *Cyperus rotundus* as out-group. Bootstraps values (1000 replicates) are shown at the nodes.

results indicated that the 35 examined grass species were evidently clustered into the 12 subfamilies of Poaceae with strong bootstrap supports (Figure 1). *E. coracana* is closely related to *Bouteloua curtipendula* and *B. gracilis* and is further grouped with the one other grass species from Chloridoideae with strong bootstrap supports.

Disclosure statement

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

Funding

This work was supported by South China Agricultural University under Startup Grant [Grant 201601103] to L. Z. Gao.

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 number MW080648. The associated BioProject, SRA, and Bio-Sample numbers are PRJNA667839, SRR15049124, and SAMN16386298, respectively.

References

- Hanna WW. 1995. Advances in small millets. *Field Crop Res.* 40(3):193.
- Jin J, Yu W, Yang J, Song Y, DePamphilis CW, Yi T, Li D. 2020. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol.* 21(1):241.
- Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30(14):3059–3066.

- Kumar A, Metwal M, Kaur S, Gupta AK, Puranik S, Singh S, Singh M, Gupta S, Babu BK, Sood S, et al. 2016. Nutraceutical value of finger millet [*Eleusine coracana* (L.) Gaertn.], and their improvement using omics approaches. *Front Plant Sci.* 7:934.
- Pradhan A, Nag SK, Patil SK. 2010. Dietary management of finger millet (*Eleusine coracana* L. Gaerth) controls diabetes. *Curr Sci.* 98(6):763–765.
- Sharma S, Sharma N, Handa S, Pathania S. 2017. Evaluation of health potential of nutritionally enriched Kodo millet (*Eleusine coracana*) grown in Himachal Pradesh, India. *Food Chem.* 214:162–168.
- Shi C, Hu N, Huang H, Gao J, Zhao Y, Gao L, Xu Y. 2012. An improved chloroplast DNA extraction procedure for whole plastid genome sequencing. *PLOS One.* 7(2):e31468.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 30(9):1312–1313.
- 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:W65–W73.