Complete chloroplast genome sequence of Solanum erianthum: genome structure and genomic resources

Dan Li and Zhan-Lin Liu

Life Science College, Northwest University, Xi’an, China

ABSTRACT
Solanum erianthum is known for its valuable medicinal properties. In this study, we report its complete chloroplast genome. The chloroplast genome size is 156,343 bp, including a LSC region of 86,855 bp, a SSC region of 18,608 bp and two inverted repeats (IR). The complete chloroplast genome includes a total of 128 unique genes with 83 protein-coding sequences, 37 tRNA and 8 rRNA genes. The results showed that S. erianthum was the most closely related to S. violaceum.

Solanum erianthum is belonged to the genus Solanum (Solanaceae), which is widespread in tropical Asia, Oceania and South America. And it’s because of its wide distribution, S. erianthum have ten synonym such as S. adulterinum Buch.-Ham. ex Wall., S. ambycalyx Dunal, S. eriocalyx Dunal and so on (POWO 2021). In China, it mainly distributed in Fujian, Guangdong, Guizhi, Hainan, Sichuan, Taiwan, Xizang and Yunnan provinces of China (Zhang et al. 1994). It’s a kind of ethnic medicinal plant in China, also in India, Mexico, Spain and other countries, with main treatment is stomachache, abdominal pain, ventilation, etc. (You et al. 2018). Genomics and modern biotechnology can be used to study the molecular mechanisms of medicinal ingredients. In this study, we sequenced and analyzed the chloroplast genome of S. erianthum, aiming to provide its chloroplast genome structure and valuable genomic resources.

The samples is collected from Zhaotong city, Yunnan Province of China (28°15′10.24″N, 103°47′1.30″E), and the specimens are deposited in the Herbarium of Kunming Institute of Botany, Chinese Academy of Sciences (Xiang963). Total genomic DNA was extracted to construct a library for sequencing by using Illumina sequencing methods at the Beijing Novogene Bioinformatics Technology Center. The clean data of S. erianthum was assembled using NOVOPlasty v3.1 (Dierckxsens et al. 2017). Assembled plastid genome annotation was conducted using Geseq in Chlorobox web service (Tillich et al. 2017), with manual corrections for start and stop codons using Geneious v.9.0.2 (Kearse et al. 2012). The annotated plastid genome sequence was submitted to GenBank (MW420931).

In this study, the chloroplast genome size is 156,343 bp, including a LSC region of 86,855 bp, a SSC region of 18,608 bp and two inverted repeats (IR). The GC content is 37.7%. This complete chloroplast genome includes a total of 128 unique genes including 83 protein-coding sequences, 37 tRNA and 8 rRNA genes.

To obtain insight explore the position of S. erianthum within Solanaceae, phylogenetic analyses were performed using maximum likelihood (ML) in RAxML8.0 with 1000 bootstrap replicates (Stamatakis 2014) (Figure 1). Based on previous studies, the ML phylogenetic reconstructed based on whole chloroplast genomes from 62 Solanum plants and one sample Eucommia ulmoides as outgroup plants (Li et al. 2021). ML analysis showed that S. erianthum was the most closely related species to S. violaceum with 100% bootstrap values. The result is similar to the previous studies, can fully proved the classification status of S. erianthum. And the ML phylogenetic show Solanum with two clades, corresponded to the morphologically hairy and glabrous (Zhang et al. 1994). So, the complete chloroplast genome of S. erianthum will provide a valuable genomic basis for further studies to reconstruct its phylogeny, assess
genetic variation, and develop conservation strategies when appropriate.

Disclosure statement
No potential conflict of interest was reported by the authors.

Data availability statement
The genome sequence data that support the findings of this study is openly available in GenBank of NCBI at https://www.ncbi.nlm.nih.gov/ under the accession no. MW420931. The associated BioProject, SRA, and Bio-Sample numbers are PRJNA689102, SRR13341520 and SAMN17199541.

Funding
This study was supported by National Natural Science Foundation of China [32070354, 31870321].

ORCID
Zhan-Lin Liu http://orcid.org/0000-0002-5735-3750

Figure 1. Maximum-likelihood phylogenetic tree based on the chloroplast genome sequences. Numbers at nodes represent bootstrap support values.
References

Dierckxsens N, Mardulyn P, Smits G. 2017. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. 45(4):e18.

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.

Li S, Wang Y-P, Zhao Y-F, Zhang J-Y, Zhang J-Y, Ma H-R, Yue Y, Du C-Y, Zhao C-B, Han Y-Z. 2021. Characterization of the complete chloroplast genome of the Solanum tuberosum L. cv. Favorita (solanaceae). Mitochondrial DNA Part B. 6(3):909–911.

Plants of the World Online (POWO). 2021. Facilitated by the Royal Botanic Gardens, Kew. Published on the Internet. http://www.plantsoftheworldonline.org/

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. 45:6–11.

You YB, Peng XZ, Yan L, Xue KH. 2018. Research progress of the national medicinal Solanum erianthum. J Chinese Med Mater. 41(2):497–502. Chinese.

Zhang ZY, Lu AM, William GD. 1994. Flora of China. Beijing: Science Press. Vol. 17. p. 300–332.