The complete chloroplast genome sequence of *Thunbergia erecta* (Benth.) T. Anders. (Acanthaceae)

Lili Tong**a, Lu Tianb,c, Xiaogang Xuc and Yao Chengb,c

aSchool of Horticulture & Landscape Architecture, Jinling Institute of Technology, Nanjing, China; bCo-Innovation Center for Sustainable Forestry in Southern China, College of Biology and the Environment, Key Laboratory of State Forestry and Grassland Administration on Subtropical Forest Biodiversity Conservation, Nanjing Forestry University, Nanjing, China; cState Environmental Protection Scientific Observation and Research Station for Ecology and Environment of Wuyi Mountains, Nanping, China

ABSTRACT

*Thunbergia erecta* (Benth.) T. Anders is an upright shrub species of Acanthaceae with great ecological and economical values. In this paper, we explored the complete chloroplast genome sequence of *T. erecta* using next generation sequencing to provide genomic resources that could help to promote Its conservation. The genome of *T. erecta* is 152,202 bp in length, containing a large single-copy region of 84,232 bp, and a small single-copy region of 17,656 bp. It encodes a total of 131 genes, including 8 rRNA genes, 37 tRNA genes, 84 protein-coding genes and 2 pseudo genes. The GC content of *T. erecta* genome is 38.47%. The phylogenetic analysis suggests that it was closely related to *Avicennia marina* in Acanthaceae. In addition, we found that *T. erecta* appeared late clade in the whole family while out-groups plants appeared even later. *T. erecta* is morphologically similar to other plants in Acanthaceae, but is genetically closed to the outgroup species.

*Thunbergia erecta* is a prominent species of Acanthaceae, around two meters tall, which is native to tropical West Africa and cultivated as ornamental plants around the world (Hu and Tsui 2002, 2011). It possesses high value for medical (Refay et al. 2021) and ornamental purposes and a good vertical greening material for scaffolds, flower fences and walls as well. However, little progress has been made to its complete chloroplast (cp) genome. In this work, we characterized the complete cp genome sequence of *T. erecta*, based on Illumina pair-end sequencing data, and deposited the sequence in GeneBank (MZ555773) to provide a valuable complete cp genome resource.

The fresh leaves were collected from Xishuangbanna Tropical Botanical Garden (**latitude** 21.6840 and **longitude** 101.4677) in Jinghong, Yunnan, China. A specimen was deposited at the herbarium of Nanjing Forestry University (contact person: Xuehong Ma; E-mail: xuehongma@njfu.edu.cn) under the voucher number NF2021038. According to the International Union for Conservation of Nature (IUCN) policy on endangered species research, the sample collection and the study was conducted with permission from Xishuangbanna Tropical Botanical Garden. The genomic DNA was extracted and then sequenced based on Illumina pair-end fragments sequencing data. By applying ultrasound to break DNA, the fragments of DNA were passivated, repaired and bonded and selected by agarose gel electrophoresis. The sample of genome sequencing library was formed by PCR amplification, which was carried out on Illumina Novaseq platform with PE150 reads by Nanjing Genepioneer Biotechnologies Inc. (Nanjing, China).

The original reading was filtered by fastp (version 0.20.0), and the clean data were assembled into chloroplast genome using SPAdes (Bankevich et al. 2012). There were no uncertain bases in the assembly results. Then, the reference sequence (Genebank accession number: NC050991) was used for quality control after assembly, and the assembled genome was annotated using CpgAVAS2 (Shi et al. 2019). The complete cp genome sequences of species were acquired necessary from NCBI.

The complete cp genome of *T. erecta* was 152,202 bp in length, containing a large single-copy region (LSC) of 84,232 bp, a small single-copy region (SSC) of 17,656 bp, and a pair of inverted repeat regions (IRs) of 25,157 bp. The new sequence possesses 131 genes, including 84 protein-coding genes (78 CDS species), 37 tRNA genes (30 tRNA species), 8 rRNA genes (4 rRNA species) and 2 pseudo genes. Among them, 6 protein-coding genes (ndhB, rpl2, rpl23, rps12, rps7 and ycf2), 7 tRNA genes (trnA-UGC, trnL-CAU, trnL-GAU, trnL-CAU, trnN-GUU, trnR-ACG and trnV-GAC), and 4 distinct rRNA genes (4.5S, 5S, 16S and 23S) were duplicated, apart from...
that, most of genes occurred in a single copy. A total of 10 protein-coding genes (atpF, ndhA, ndhB, petB, petD, rpl16, rpl2, rpoC1, rps12 and rps16) contained 1 intron while the other 2 genes (clpP, ycf3) had 2 introns each. The overall GC content of cp genome was 38.47%, while the corresponding values of the LSC, SSC and IR regions were 36.57%, 33.15% and 43.52%, respectively (Figure 1).

To reveal the phylogenetic evolution of *T. erecta*, we constructed a ML phylogenetic tree based on 16 cp genomes from Acanthaceae and 3 cp genomes as outgroups from 2 taxa (Pedaliaceae, Linderniaceae). The sequence alignment by MAFFT (Rozewicki et al. 2019), IQTREE (Garg and Biju 2021) was used to perform maximum Likelihood (ML) tree with the TVM + F + R4 model. The bootstrap method was used to test the reliability of phylogeny with 1000 replicates.

The phylogenetic analysis suggests that *T. erecta* was closely related to *Avicennia marina* in Acanthaceae. In addition, *T. erecta* was more likely appeared in late differentiation stage in the whole family succession while outgroups species did even later. *T. erecta* is morphologically similar to other species in Acanthaceae, but is genetically closed to those in outgroup. As the increasing of sample collection sequences of this taxa and the deepening of research, its status will become clearer.

**Author contributions**

Among the members of the author group, Lili Tong, Xiaogang Xu and Lu Tian contributed to substantial conception or design of the work; Lu Tian and Yao Cheng were in charge of acquisition, analysis, and interpretation of data for the work; Xiaogang Xu and Lili Tong, contributed to manuscript preparation (Drafting the work or revising it critically for important intellectual content); Lili Tong and Xiaogang Xu contributed to final approval of the version to be published. All authors agree on the final version and to be accountable for all aspects of the work.
Disclosure statement

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

Funding

This work was supported by the [The Biodiversity Investigation, Observation and Assessment Program of Ministry of Ecology and Environment of China #1] under Grant [number 20191226]; [The Special Fiscal Funds for Repair and Purchase in National Public Institutions #2] under Grant [number 2010002002]; and [Jiangsu Forestry Science and Technology Innovation and Extension Program #3] under Grant [number LYKJ[2018]13].

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] (https://www.ncbi.nlm.nih.gov/) under the accession no. MZ555773. The associated BioProject, SRA, and Bio-Sample numbers are PRJNA745545, SRR15100132, and SAMN20169050 respectively.

References

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 19(5):455–477.

Garg S, Biju SD. 2021. DNA barcoding and systematic review of Minervaryan Frogs (Dicroglossidae: Minervarya) of Peninsular India: resolution of a taxonomic conundrum with description of a new species. Asian Herpetol Res. 12(04):345–378.

Hu J-Q, Deng Y-F, John RIW, Thomas FD. 2011. Acanthaceae. In: Hu C-C, editor. Flora of China, Vols. 19 (Acanthaceae). Beijing: Science Press; St. Louis: Missouri Botanic Garden press; p. 377–379.

Hu C-C, Tsui H-P. 2002. Acanthaceae. In: Hu C-C, editor. Flora Reipublicaen Popularis Sinicae, vols. 70 (Acanthaceae). Beijing: Science Press, p. 24.

Refaey MS, Abdelhamid RA, Elimam H, Elshaier YAMM, Ali AA, Orabi MAA. 2021. Bioactive constituents from Thunbergia erecta as potential anticholinesterase and anti-ageing agents: Experimental and in silico studies. Bioorg Chem. 108:104643.

Rozewicki J, Li S, Amada KM, Standley DM, Katoh K. 2019. MAFFT-DASH: integrated protein sequence and structural alignment. Nucleic Acids Res. 47(W1):W5–W10.

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.