MITOCHONDRIAL DNA PART B: RESOURCES
2024, VOL. 9, NO. 3, 371–375
https://doi.org/10.1080/23802359.2024.2329674

Taylor & Francis
Taylor & Francis Group

PLASTOME REPORT

OPEN ACCESS Check for updates

Elucidating phylogenetic relationships within the genus *Curcuma* through the comprehensive analysis of the chloroplast genome of *Curcuma viridiflora* Roxb. 1810 (Zingiberaceae)

Ling Lin[a,b]*, Zhigang Hao[a,c,d]*, Liying Zhou[a,e]*, Wenyi Liu[b], Nian Liu[e], Kebao Wang[f] and Ruizong Jia[a]

[a]Sanya Research Institution/Hainan Key Laboratory for Biosafety Monitoring and Molecular Breeding in Off-Season Reproduction Regions, Chinese Academy of Tropical Agriculture Sciences, Sanya, China; [b]Foshan Lianyi Biotechnology Co., Ltd., Foshan, China; [c]Sanya Institute of China Agricultural University, Sanya, China; [d]Hainan Seed Industry Laboratory, Sanya, China; [e]Department of Horticulture and Landscape Architecture, Zhongkai University of Agriculture and Engineering, Guangzhou, China; [f]Kenli Vocational Education Center, Dongying, China

ABSTRACT

Curcuma viridiflora Roxb., a plant species of significant pharmaceutical interest, has been the subject of limited chloroplast genomic research. In this study, we present the sequencing and assembly of the *C. viridiflora* chloroplast genome, which is characterized by a circular chromosome spanning 162,212 base pairs and a GC content of 36.20%. The genome encodes 87 protein-coding genes (PCGs), 38 transfer RNA (tRNA) genes, and eight ribosomal RNA (rRNA) genes. A phylogenetic analysis was conducted, incorporating eight related species, and based on the complete chloroplast genome and protein-coding DNA sequences of six related taxa within the genus. Outgroup species *Zingiber zerumbet* and *Zingiber officinale* were also included in the analysis. The results indicate a close relationship between *C. viridiflora* and *Curcuma phaeocaulis*, *Curcuma sichuanensis*, and *Curcuma yunnanensis*. This study provides the first chloroplast genome of *C. viridiflora*, thereby contributing a valuable genomic resource for future research on medicinal plants within the *Curcuma* genus.

ARTICLE HISTORY
Received 28 November 2023
Accepted 7 March 2024

KEYWORDS
Chloroplast genome; phylogenetic relationships; *Curcuma viridiflora*

Introduction

Curcuma viridiflora Roxb. 1810, native to Sumatra and other eastern islands, distributed in Indonesia, Malaysia, and Taiwan, China, belonging to genus *Curcuma* L. of the family Zingiberaceae, consists of about 80 species in the world, with 15 species mostly occurring in China (Roxburgh 1820; Valeton 1918; Wu 2000). Known for its diverse applications in culinary arts, traditional medicine, and ornamentation, the species, particularly its rhizomes and tubers, exhibits a wide array of pharmacologically relevant bioactivities – ranging from antitumorigenic to anti-fatigue properties – attributed primarily to curcuminoids and volatile oils (sesquiterpenes, diterpenes, etc.) (Wu et al. 2016).

The genus *Curcuma* occupied a pivotal role in the taxonomy of Zingiberaceae, especially a Chinese variant distinguished by its green flower, serves as a significant agricultural export, mainly to Japan. While extant research offers substantive insights into its bioactive potential, there exists a conspicuous research gap concerning the genetic study of its chloroplast genome. The present investigation addresses this deficiency by pioneering the assembly of the chloroplast genome of *C. viridiflora*, thus enriching the existing corpus of genomic data and providing a foundational resource for future research in comparative genomics and species identification within the *Curcuma* genus.

Materials and methods

The fresh leaves of *C. viridiflora* were meticulously collected from a designated location in Xiaoxiang Town, Gaoyao District, Zhaoqing, Guangdong Province, China (geographical coordinates: 23.1708 N, 112.4191 E, picture see Figure 1). A specimen was deposited at the South China Botanical Garden Herbarium (website: herbarium.scbg.cas.cn; Curator: Dr. Luoshixiao, email: luoshixiao@scbg.ac.cn) under the voucher number 1003828. Additional populations of *C. viridiflora* have been cultivated in Shishan Town, Nanhai District, Foshan, Guangdong Province, China, at the coordinates 23.1464 N, 113.0589 E.

The extraction of total genomic DNA was executed utilizing the Blood and Cell Culture DNA Midi Kit (catalog no. 13343, Qiagen, New York, NY), strictly adhering to the manufacturer's protocol. The resultant DNA was employed in the creation of a DNA library featuring an insert size ranging

CONTACT Zhigang Hao 1067323960@qq.com; Ruizong Jia jiaruizong@itbb.org.cn Sanya Research Institution/Hainan Key Laboratory for Biosafety Monitoring and Molecular Breeding in Off-Season Reproduction Regions, Chinese Academy of Tropical Agriculture Sciences, Sanya 572011, China
*These authors contributed equally to this work.

Supplemental data for this article can be accessed online at https://doi.org/10.1080/23802359.2024.2329674.

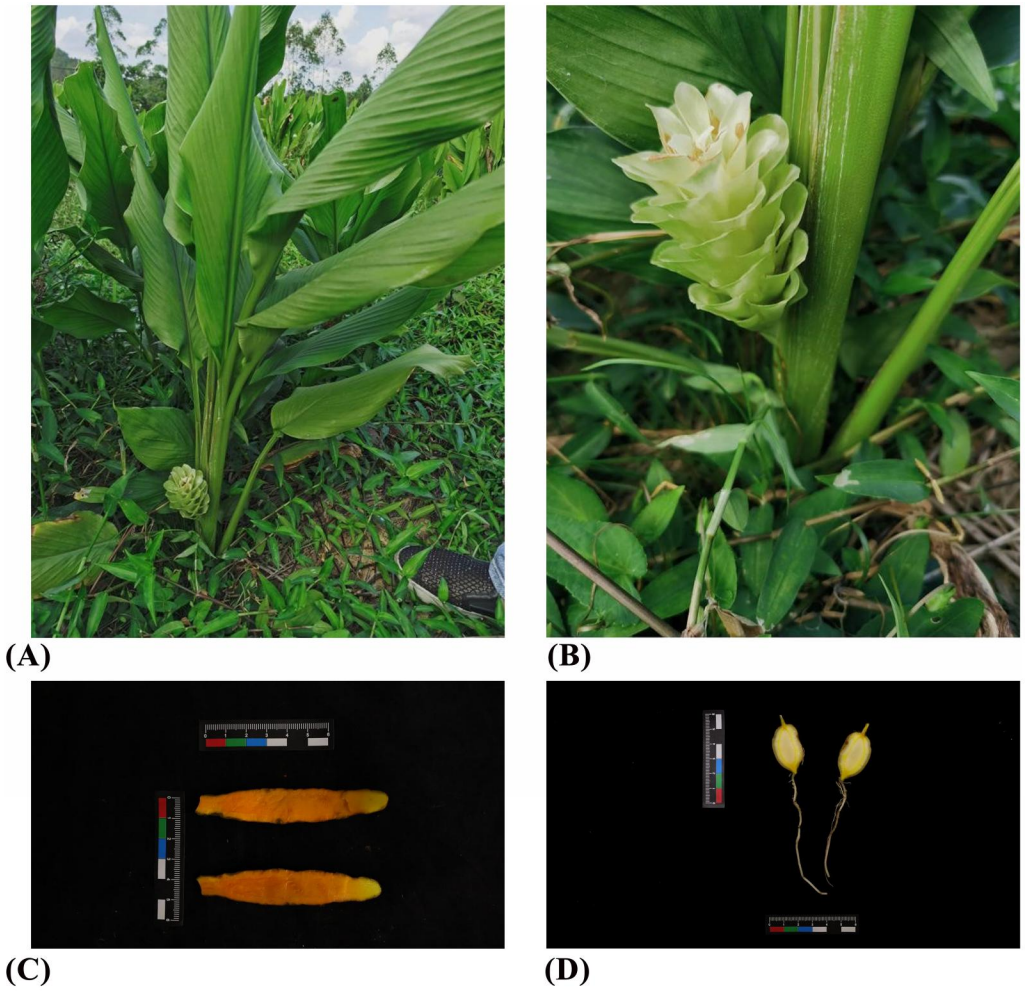

Figure 1. Photographs of *C. viridiflora* (these photographs were taken by Ling Lin and Prof. Nian Liu). (A) Plant panorama of *C. viridiflora*. (B) The flowers of *C. viridiflora*. (C) Rhizomes of *C. viridiflora*. (D) Tubers of *C. viridiflora*.

from 200 to 400 base pairs, facilitated by the BGISEQ sequencing platform (BGI, Shenzhen, China). The raw data consisting of 5.80 GB of raw data, comprised a total of 19,200,779 reads. These reads were subsequently utilized in the assembly of a chloroplast genome, employing the NOVOPlasty (version 2.7.2) (Dierckxsens et al. 2017). Annotation of the assembled genome was conducted through the application of CPGAVAS2 (Shi et al. 2019). The authenticity of the genome sequence was rigorously corroborated through the alignment of all raw sequence reads against the assembled genomic framework, employing the BWA software (version 0.7.17) and SAMtools (version 1.9) (Rimmer et al. 2014), all within the computational milieu of the Genome Information System (GeIS; https://geis.infoboss.co.kr/).

For the elucidation of the phylogenomic affiliations within the genus *Curcuma*, a collection of eight complete chloroplast genome sequences was acquired from publicly accessible repositories, which also included two sequences from outgroup plant species. The common genes of nine chloroplast genomes (including *C. viridiflora*, and outgroup were *Zingiber zerumbet* and *Zingiber officinale*) were extracted and concatenated with PhyloSuite (Zhang et al. 2020). The concatenated sequences underwent multiple sequence alignment facilitated by the MAFFT algorithm (version 7.450) (Rozewicki et al. 2019). Subsequently, a phylogenetic tree was constructed, leveraging the aligned sequences and the maximum-likelihood (ML) algorithm as implemented in the IQtree software package (Trifinopoulos et al. 2016).

Results

The assembled genome exhibited average and minimum read mapping depths of ×2099 and ×540, respectively (Figure S1). A comprehensive circular map of the chloroplast genome, along with a schematic representation of the cis- and trans-splicing genes (Figure 2 and Figure S2), was visualized through CPGView software (Liu et al. 2023). The chloroplast genome of *C. viridiflora*, designated under the GenBank accession number PP234477.1, manifests a total length of 162,212 base pairs, accompanied by a GC content of 36.20% (Figure 2). Architecturally, this chloroplast genome adopts a quadripartite structure, comprising a pair of inverted repeats (IRs) spanning 29,737 bp, flanked by a large single-copy (LSC) region of 87,008 bp and a small single-copy (SSC) region of 15,730 bp. The genomic assembly encodes a total of 133 genes, among which 112 are unique. This

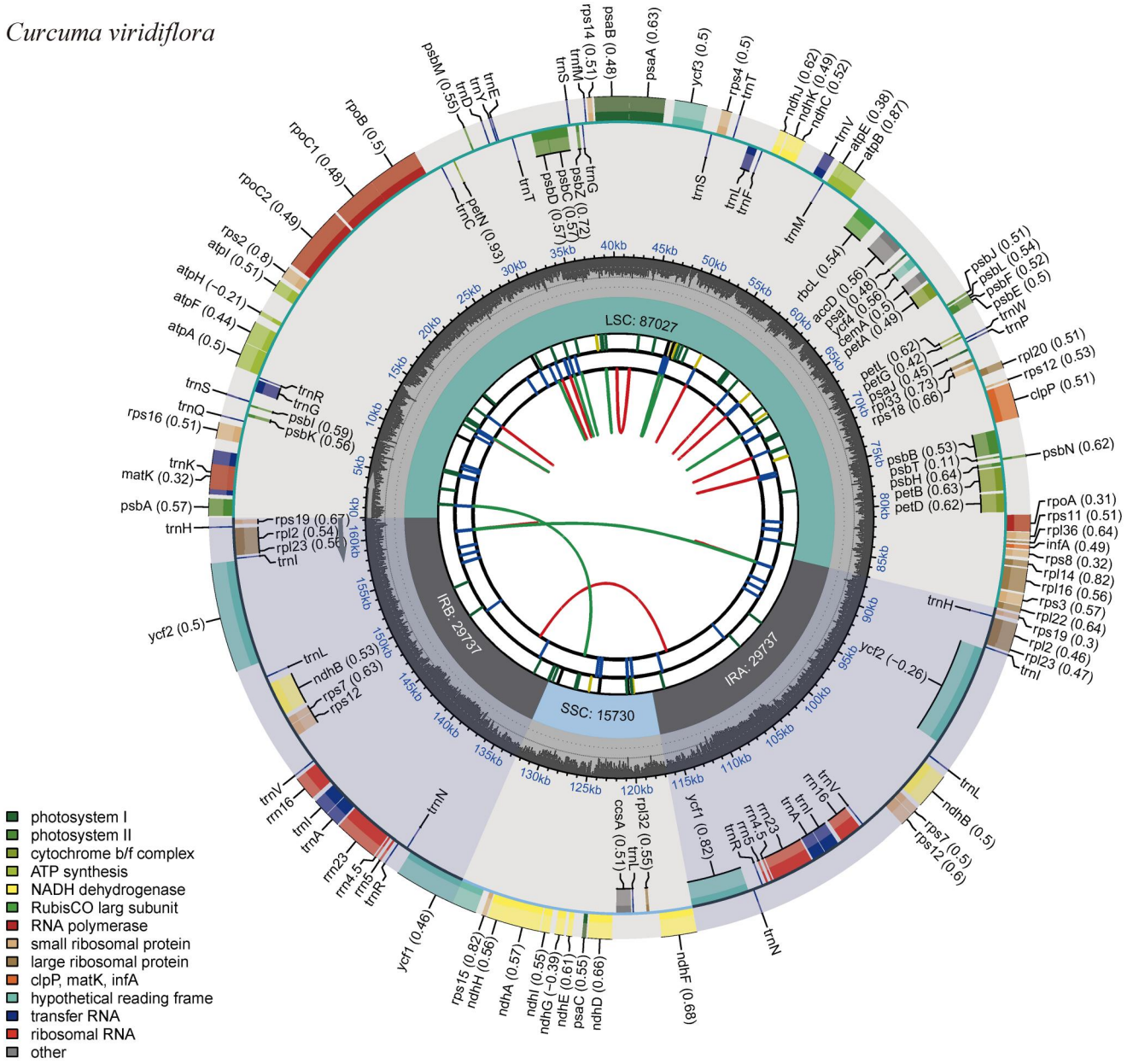

Figure 2. The circle map of chloroplast genome map of *C. viridiflora*. Distinctive colored boxes encircling the outer circle depict genes, with clockwise and counter-clockwise transcribed genes represented inside and outside the circle, respectively. The inner circle features a gray region indicating the GC content, while the quadripartite structure (LSC, SSC, IRA, and IRB) is illustrated on the inner circle accordingly.

includes 87 protein-coding genes (PCGs), eight ribosomal RNA (rRNA) genes, and 38 transfer RNA (tRNA) genes. Nineteen PCGs had one intron, and four had two introns. The phylogenetic analysis revealed that *C. viridiflora* formed a monophyletic clade together with *C. phaeocaulis*, *C. sichuanensis*, and *C. yunnanensis* (Figure 3). The chloroplast genome sequence of *C. viridiflora* is a valuable resource for genome evolution and taxonomy research of the genus *Curcuma*.

Discussion and conclusions

The present study marks the inaugural sequencing of the complete chloroplast genome of *C. viridiflora*, characterized by a length of 162,212 bp. In terms of both genome size and gene content, *C. viridiflora* aligns closely with extant chloroplast genomes within the genus *Curcuma* (Chen 2019; Shan et al. 2021). Our empirical findings hold substantive utility for the authentication of *C. viridiflora* and facilitate analyses related to genetic diversity and phylogenetic relationships within the genus *Curcuma*. Despite the richness of species within the *Curcuma* genus, the availability of sequenced chloroplast genomes in the NCBI database remains limited as of the most recent update on 15 October 2023. Consequently, to gain a more nuanced understanding of the evolutionary history of *C. viridiflora*, the acquisition of a broader spectrum of chloroplast sequences from various *Curcuma* species is imperative. The elucidation of the *C. viridiflora* chloroplast genome in this study constitutes a valuable resource, enriching the existing data pool for phylogenetic and molecular identification endeavors concerning *Curcuma* species.

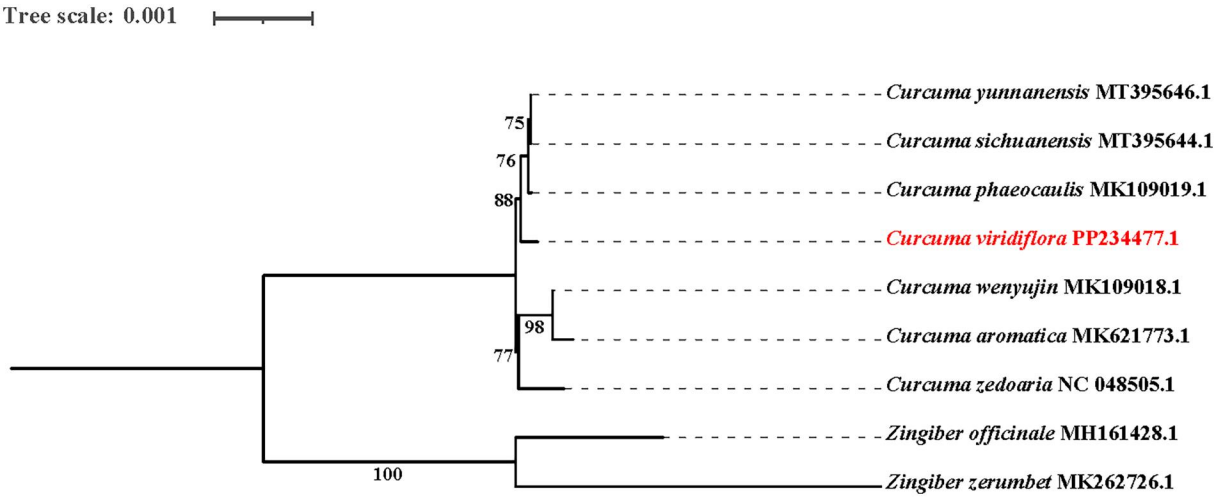

Figure 3. The maximum-likelihood phylogenetic tree of seven *Curcuma* species was constructed based on the CDS sequences extracted by IQ-TREE, with *Zingiber zerumbet* and *Zingiber officinale* added as outgroup. The phylogenetic tree was constructed using the maximum-likelihood method (ML) and bootstrap was performed 1000 times. The number on each branch indicates the boot support value. The following sequences were used: *Curcuma yunnanensis* MT395646.1 (Liang et al. 2020), *Curcuma sichuanensis* MT395644.1 (Liang et al. 2020), *Curcuma phaeocaulis* MK109019.1 (Kim et al. 2021), *Curcuma wenyujin* MK109018.1 (Kim et al. 2021), *Curcuma aromatica* MK621773.1 (Gui et al. 2020), *Curcuma zedoaria* NC 048505.1 (Li et al. 2019), *Zingiber officinale* MH161428.1 (Cui et al. 2019), and *Zingiber zerumbet* MK262726.1 (Li et al. 2019).

Author contributions

Ling Lin, Liying Zhou, and Ruizong Jia: drafting the work and revising it critically for important intellectual content. Zhigang Hao: analyzed and interpreted the data for the work. Nian Liu and Wenyi Liu contributed experimental materials and confirmed that the accuracy or integrity of any part of the work was appropriately investigated and resolved. Ling Lin, Kebao Wang, and Ruizong Jia: final check, revise, and approval of the version to be published. The authors agree to be accountable for all aspects of the work.

Ethical approval

The authors declare no ethical or legal violations when obtaining the study materials and performing the research. The species used in this study is not listed on the IUCN Red List, and the sample was legally collected by guidelines stipulated in national and international regulations. The materials were collected in a location not designated as a protected area in China.

Disclosure statement

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

Funding

The research was supported by the Project of Sanya Yazhou Bay Science and Technology City [Grant No. SCKJ-JYRC-2022-67, partially supported by No. ZDKJ202002, No. 323MS088, and No. ZDYF2022XDNY257]. This work also partially supported by Hainan Seed Industry Laboratory [Grant No.: B23C10004]; the Hainan Provincial Natural Science Foundation of China; the Hainan Major Science and Technology Project; the Hainan Province Major Research Project.

ORCID

Ling Lin http://orcid.org/0000-0003-3615-1161

Data availability statement

The genome sequence data supporting the findings of this study can be publicly obtained at NCBI GenBank at https://www.ncbi.nlm.nih.gov with the accession number PP234477. Associated BioProject, SRA, and BioSample numbers are PRJNA1070241, SAMN39643034, and SRR27778106, respectively.

References

Chen S. 2019. Genetic and phylogenetic analysis of the complete genome for the herbal medicine plant of *Scutellaria baicalensis* from China. Mitochondrial DNA B Resour. 4(1):1683–1685. doi:10.1080/23802359.2019.1605859.

Cui Y, Nie L, Sun W, Xu Z, Wang Y, Yu J, Song J, Yao H. Comparative and phylogenetic analyses of ginger (*Zingiber officinale*) in the family Zingiberaceae based on the complete chloroplast genome. Plants. 2019;8(8):283. doi:10.3390/plants8080283.

Dierckxsens N, Mardulyn P, Smits G. 2017. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. 45(4):e18. doi:10.1093/nar/gkw955.

Gui L, Jiang S, Xie D, Yu L, Huang Y, Zhang Z, Liu Y. 2020. Analysis of complete chloroplast genomes of *Curcuma* and the contribution to phylogeny and adaptive evolution. Gene. 732:144355. doi:10.1016/j.gene.2020.144355.

Kim M-K, Lee WK, Choi YR, Kim J, Kang I, Kang J. 2021. The complete chloroplast genome sequence of three medicinal species; *Curcuma longa*, *Curcuma wenyujin*, and *Curcuma phaeocaulis* (Zingiberaceae). Mitochondrial DNA B Resour. 6(4):1363–1364. doi:10.1080/23802359.2020.1768917.

Li D-M, Zhao C-Y, Xu Y-C. 2019. Characterization and phylogenetic analysis of the complete chloroplast genome of *Curcuma longa* (Zingiberaceae). Mitochondrial DNA B Resour. 4(2):2974–2975. doi:10.1080/23802359.2019.1664343.

Liang H, Zhang Y, Deng J, Gao G, Ding C, Zhang L, Yang R. 2020. The complete chloroplast genome sequences of 14 *Curcuma* species: insights into genome evolution and phylogenetic relationships within Zingiberales. Front Genet. 11:802. doi:10.3389/fgene.2020.00802.

Liu S, Ni Y, Li J, Zhang X, Yang H, Chen H, Liu C. 2023. CPGView: a package for visualizing detailed chloroplast genome structures. Mol Ecol Resour. 23(3):694–704. doi:10.1111/1755-0998.13729.

Rimmer A, Phan H, Mathieson I, Iqbal Z, Twigg SRF, Wilkie AOM, McVean G, Lunter G. 2014. Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. Nat Genet. 46(8):912–918. doi:10.1038/ng.3036.

Roxburgh W. 1820. Flora Indica. Serampore: The Mission Press.

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. doi:10.1093/nar/gkz342.

Shan Y, Pei X, Yong S, Li J, Qin Q, Zeng S, Yu J. 2021. Analysis of the complete chloroplast genomes of *Scutellaria tsinyunensis* and *Scutellaria tuberifera* (Lamiaceae). Mitochondrial DNA B Resour. 6(9): 2672–2680. doi:10.1080/23802359.2021.1920491.

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. doi:10.1093/nar/gkz345.

Trifinopoulos J, Nguyen L-T, von Haeseler A, Minh BQ. 2016. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Research. 44(W1):W232–W235.

Valeton T. 1918. New notes on the Zingiberaceae of Java and Malaya Archipelago. *Curcuma*. Bull Jard Bot Buitenzor. 27:1–166.

Wu D, Liu N, Ye Y. 2016. The Zingiberaceous resources in China. Wuhan, China: Huazhong University of Science and Technology University Press; p. 143.

Wu TL. 2000. Zingiberaceae. Flora of China. Beijing; St. Louis: Sciences Press; Missouri Botanical Garden Press.

Zhang D, Gao F, Jakovlić I, Zou H, Zhang J, Li WX, Wang GT. 2020. PhyloSuite: an integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. Mol Ecol Resour. 20(1):348–355. doi:10.1111/1755-0998.13096.