MITOCHONDRIAL DNA PART B: RESOURCES
2025, VOL. 10, NO. 1, 32–36
https://doi.org/10.1080/23802359.2024.2444593

Taylor & Francis
Taylor & Francis Group

PLASTOME REPORT

OPEN ACCESS Check for updates

Characterization and phylogenetic analysis of the complete chloroplast genome of *Carpesium longifolium* F. H. Chen & C. M. Hu (Asteraceae, Inuleae)

Hui Chen[a,b], Tingyu Li[b], Xinyu Chen[b], Tianmeng Qu[b], Xinyi Zheng[b], Yueguang Wang[b], Liang Pang[c,d#], Xianhua Gu[e*] and Zhixi Fu[a,b,f*]

[a]Ministry of Education, Key Laboratory of Land Resources Evaluation and Monitoring in Southwest (Sichuan Normal University), Chengdu, China; [b]College of Life Sciences, Sichuan Normal University, Chengdu, China; [c]Sichuan Tianshengyuan Environmental Services Co., Ltd, Chengdu, China; [d]Sichuan Tianshengyuan Information Technology Co., Ltd, Chengdu, China; [e]Sichuan Leshan Ecological Environment Monitoring Center Station, Leshan, China; [f]Sustainable Development Research Center of Resources and Environment of Western Sichuan, Sichuan Normal University, Chengdu, China

ABSTRACT

In this study, we studied the complete chloroplast (cp) genome of *Carpesium longifolium* F. H. Chen & C. M. Hu 1974. The results indicated that the cp genome had a typical circular structure of 151,260 bp in length. A total of 127 genes were identified, including 36 tRNA genes, 8 rRNA genes and 83 protein-coding genes, with the GC content of 37.7%. Phylogenetic analysis revealed the systematic position of *C. longifolium* is sister to *C. cernuum* and *C. faberi*. For the identification and phylogenetics study of the genus, the chloroplast genome sequence of *C. longifolium* provides a useful genetic resource.

ARTICLE HISTORY
Received 29 May 2024
Accepted 15 December 2024

KEYWORDS
Carpesium longifolium; plastid genome; phylogenetic relationship

Introduction

Carpesium L. is a genus of the Asteraceae family with beneficial medicinal value. About 21 species of *Carpesium* are majorly distributed in the Eurasian continent (Li et al. 2011; Yang et al. 2014). Seventeen species and 3 variety species of *Carpesium* found in China, mainly distributed in the Southwest of China (Shi et al. 2022). The plants of genus *Carpesium*, such as *C. abrotanoides*, *C. cernuum*, and *C. divaricatum* has been widely used as a folk medicine in treating mumps, folliculitis, toothache, colds, and fever (Zhang et al. 2015; Yang 2016). *Carpesium longifolium* F. H. Chen & C. M. Hu 1974 is mainly distributed in Gansu, Guizhou, Hubei, Shaanxi, Sichuan, growing in moist forests, riversides and grasslands at an altitude of 600-2300 meters.

In comparison to nuclear and mitochondrial genomes, the chloroplast genome offers significant advantages for genetic and phylogenetic analyses, attributable to its stable genetic structure, relatively compact size, and moderate nucleotide substitutions (Jansen et al. 2007; Daniell et al. 2016). In many angiosperms, the cp genome is usually a quadruplex, consisting of 2 inverted repeat regions (IRs), a large single-copy region (LSC), and a small single-copy region (SSC), with the IR regions separating the LSC and SSC regions (Palmer 1985). To date, the chloroplast genomes of *C. longifolium* remains uncharacterized. In this study, we sequenced, assembled, and analyzed the complete chloroplast genome of *C. longifolium*, and investigated the phylogenetic relationship within the genus *Carpesium*.

Materials and methods

The species of *C. longifolium* was collected from Baisha River in Ya'an City, Sichuan Province, China (29.9859310 N, 103.0087944E) (Figure 1). The voucher specimen of Chenhui no. 188 has been deposited in the botanical herbarium of Sichuan Normal University (SCNU) (https://bio.sicnu.edu.cn/; contact person: Dr. Zhixi Fu, email: fuzx2017@sicnu.edu.cn). The species were harvested, and the fresh leaves were promptly preserved at −80 °C till examination. Total genomic DNA was extracted from fresh leaves using the CTAB DNA extraction protocol (Allen et al. 2006). The Illumina Paired-End DNA Library Kit (Illumina Inc. San Diego, CA, USA) was utilized to construct DNA libraries. Sequencing of the qualified library was performed on the Illumina NovaSeq 6000 platform (NovoGene Inc., Beijing, China), with a read length of 150 bp. Last, the Illumina Genome Analyzer (Hiseq 2000,

CONTACT Xianhua Gu 472678781@qq.com Sichuan Leshan Ecological Environment Monitoring Center Station, Leshan 614099, China; Zhixi Fu fuzx2017@sicnu.edu.cn Ministry of Education, Key Laboratory of Land Resources Evaluation and Monitoring in Southwest (Sichuan Normal University), Chengdu, China.

Supplemental data for this article can be accessed online at https://doi.org/10.1080/23802359.2024.2444593.

*Both authors contributed equally to this work.

#Present address: Xizang Shengyuan Environmental Engineering Co., Ltd, Tibet, China.

Figure 1. The photos of *Carpesium longifolium* were taken by Dechang, Meng without any copyright issues. Panel a shows a panorama and panel B shows a detail. The plant's coordinate is 29.9859310 N, 103.0087944E. It's a perennial herb 50 to 100 cm in height with base woody and upper branched and lanceolate leaves measuring 8-15 × 1.5-3 cm. Capitula spicately arranged, axillary or terminal, terminal one lanceolate bracteal leaves, involucre hemispheric 4, marginal florets 3- or 4-seriate, corolla tubular, disk florets tubular. Bloom in the July to September. This plant thrives in moist forests, riversides and grasslands at an altitude of 600-2300 meters.

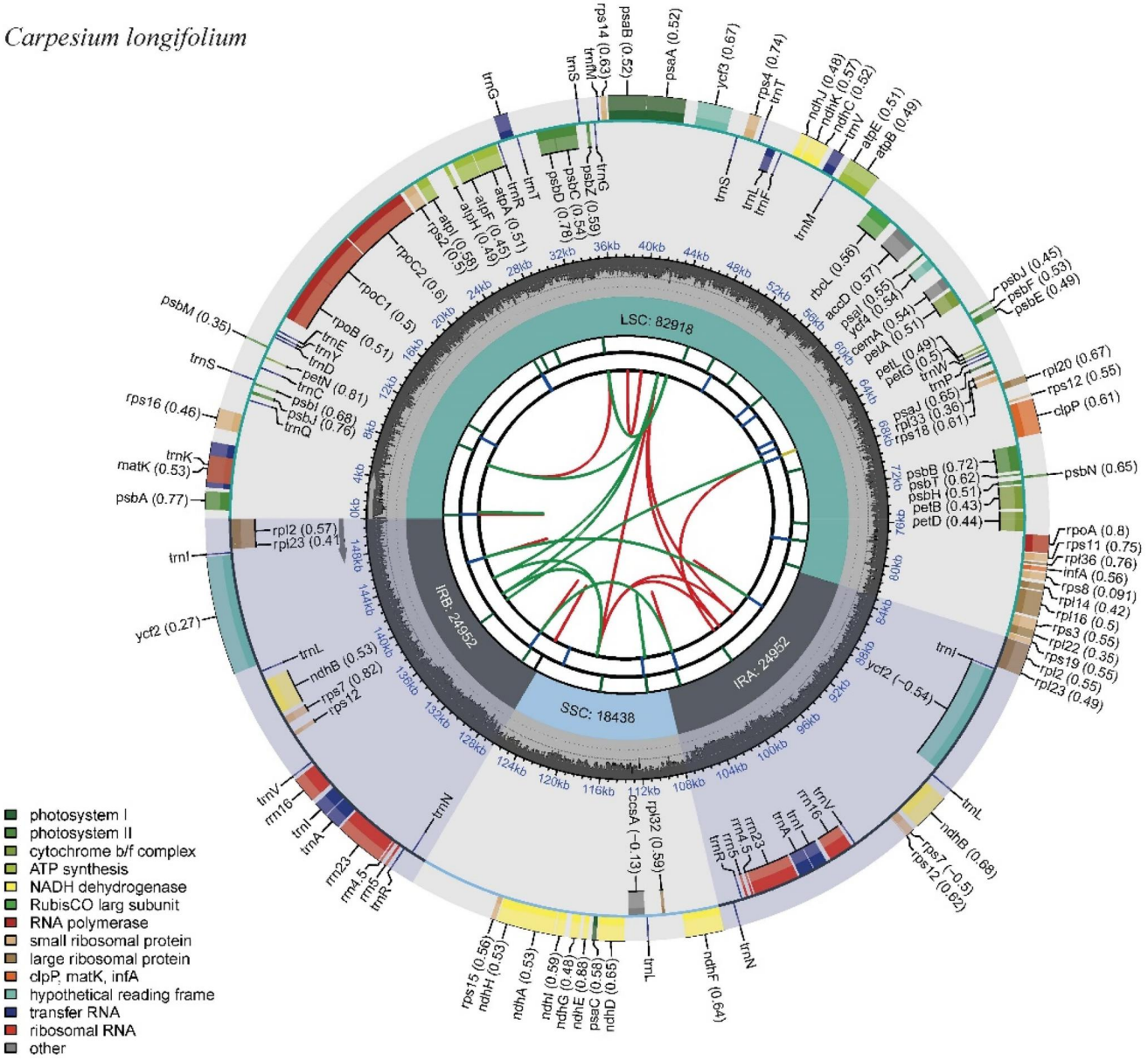

Figure 2. Circular gene map of the chloroplast genome of *Carpesium longifolium*. Genes drawn within the circle are transcribed clockwise, while those drawnoutside are transcribed counterclockwise. Genes are color-coded according to their functional groups. Inner circle represents GC content.

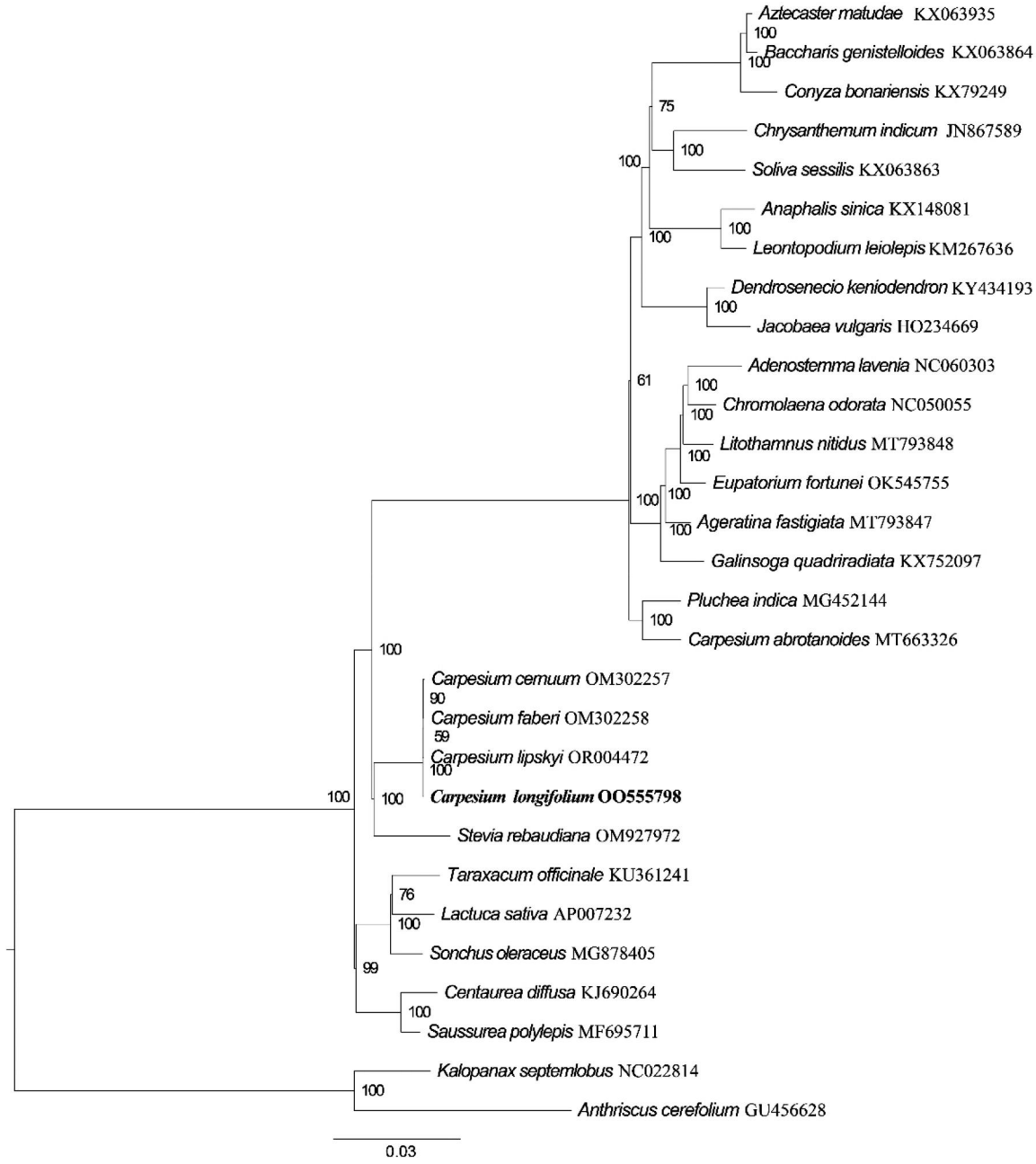

Figure 3. The best maximum likelihood (ML) phylogram inferred from 29 chloroplast genomes (bootstrap value are indicated on the branches) the specie marked in red is the newly sequenced species in this study. The bloded species is the newly sequenced species. The following sequences were used: *Anthriscus cerefolium* Hoffm. GU456628 (Downie and Jansen 2015), *Kalopanax septemlobus* (Thunb.) koidz. NC022814 (Li et al. 2013), *Litothamnus nitidus* (DC.) W.C.Holmes MT793848 (Thode et al. 2021), *Chromolaena odorata* (L.) R. M. King & H. Robinson NC050055 (Quan and Li 2020), *adenostemma lavenia* (L.) O. Kuntze NC060303 (unpublished), *ageratina fastigiata* (kunth) R.M.King & H.Rob. MT793847 (Thode et al. 2021), *Stevia rebaudiana* (bertoni) Bertoni OM927972 (unpublished), Pluchea indica (L.) less. MG452144 (Zhang et al. 2017), *Conyza bonariensis* (L.) cronq. KX792499 (unpublished), *Aztecaster matudae* (Rzed.) G.L.Nesom KX063935 (Vargas et al. 2017), *Baccharis genistelloides* (Lam.) Pers. KX063864 (Vargas et al. 2017), *soliva sessilis* ruiz & pav. KX063863 (Vargas et al. 2017), *Chrysanthemum indicum* L. JN867589 (unpublished), *Anaphalis sinica* hance KX148081 (unpublished), leontopodium leiolepis nakai KM267636 (unpublished), *Carpesium abrotanoides* L. MT663326 (unpublished), *Dendrosenecio keniodendron* (R.E.Fr. & T.C.E.Fr.) B.Nord. KY434193 (unpublished), *Jacobaea vulgaris* gaertn. HQ234669 (Doorduin et al. 2011), *Taraxacum officinale* F. H. Wigg. KU361241 (Kim et al. 2016), *Sonchus oleraceus* L. MG878405 (Hereward et al. 2018), *Lactuca sativa* L. AP007232 (unpublished), *Saussurea polylepis* Nakai MF695711 (Yun et al. 2017), *Centaurea diffusa* Lam. KJ690264 (unpublished), *Galinsoga quadriradiata* ruiz et pav. KX752097 (unpublished), *Eupatorium fortunei* turcz. OK545755 (unpublished), *Carpesium faberi* C. Winkl. OM302258 (unpublished), *Carpesium cernuum* L. OM302257 (Shi et al. 2022), *Carpesium lipskyi* C. Winkl. OR004472 (unpublished), *Carpesium longifolium* F. H. Chen & C. M. Hu OQ555798 (the sequenced species).

Illumina, San Diego, CA, USA) was used to get the original sequence.

In this study, SPAdes v3.10.1 software was employed to assemble high-quality data using default parameters (Bankevich et al. 2012). To assess the accuracy of assembly, this study used BWA (version v0.7.17) to generate depth of coverage (Figure S1) (minimum depth = 36 x; maximum depth = 3600; average depth = 168.59 x) (https://www.protocols.io/view/generating-sequencing-depth-and-coverage-map-for-o-4r3l27jkxg1y/v1). Subsequently, chloroplast genome sequences were annotated using PGA (Qu et al. 2019) with the results verified through Geneious R11 (Kearse et al. 2012) and manually adjusted as necessary. The annotation result was visualized using the CPGview program (Figure S2, S3) (Liu et al. 2023). The study

downloaded 28 chloroplast genome sequences from the GenBank database, which were subsequently imported into MAFFT v7.520 software (Katoh and Standley 2013) for multiple comparisons. Maximum likelihood (ML) analysis of phylogenetic connection reconstruction with RAxML (Stamatakis 2014) with GTRGAMMA model on the CIPRES (Miller et al. 2010). To determine the bootstrap values (BS) for each node in the phylogenetic tree, 1000 bootstrap repeats were conducted, with the remaining parameters set to their default settings.

Results

The complete cp genome sequence of *C. longifolium* was 151,260 bp in length and had a typical quadripartite structure observed in most land plants (Figure 2). A pair of 24,952 bp IR regions (IRa and IRb) divided the 82,918 bp LSC and 18,438 bp SSC regions found in the cp genome. The GC content of the cp genome of *C. longifolium* was 37.7% overall GC content. A total of 127 unique genes were annotated, consisting of 83 protein-coding genes, 36 tRNA genes, and four rRNA genes. 11 genes including *rps16*, *rpoC1*, *atpF*, *petB*, *petD*, *rpl16*, *rpl2*, *ycf3*, *ndhB*, *clpP* and *ndhA* are cis-splicing genes; *ycf3* and *clpP* had two copies (Figure S2). The *rps12* is a trans-splicing gene (Figure S3).

The chloroplast genome sequences were used to performing the phylogenetic analyses. For the phylogenomic inference, a total of 27 complete Asteraceae plastome sequences and two outgroup plastomes (*Kalopanax septemlobus* Koidz. and *Anthriscus cerefolium* Hoffm.) were employed. Phylogenomic analysis showed that *C. longifolium* was sister to *C. lipskyi*, *C. cernuum* and *C. faberi* with a full support value (Figure 3). *Carpesium* and *Stevia rebaudiana* clustered into a clade supported by a bootstrap value of 100%.

Discussion and conclusions

This study examined the characteristics, composition, and structure of the chloroplast genome of *C. longifolium*. As a result, it was demonstrated that *C. longifolium* had the characteristic quadripartite structure prevalent in vascular plants. The chloroplast genomes are circular DNA molecules ranging in size from 120 to 170 kb, exhibiting highly conserved structures and organization in the majority of terrestrial plants (Wicke et al. 2011; Mehmood et al. 2020). *C. longifolium* showed relative conservation in genome size and gene content.

Previous studies on *Carpesium* classification used one to several molecular markers, such as ITS, ETS, *ndhF*, *trnL-F* and *trnH-psbA* sequences, suggesting that *Carpesium* has a polyphyletic nature (Englund et al. 2009; Nylinder and Anderberg 2015). The phylogenetic tree showed that the species of *Carpesium* clustered in a clade in phylogenetic evolution with 100% support values, closely related to genera such as *Stevia*. The classical taxonomic approach places *C. cernuum* and *C. lipskyi* in the Sect. *Carpesium* and *C. faberi* and *C. longifolium* in the Sect. *Abrotanoides* (Ling et al. 1985). However, this study found that *C. longifolium* was more closely related to *C. lipskyi*, deviating from the traditional morphological classification method. Further research is required to determine whether the traditional taxonomy accurately reflects the relationships among the species within this genus. There is a serious lack of available complete chloroplast genome sequence data. Consequently, additional studies on the complete chloroplast genome of this genus are essential to accurately analyze the affinities among its species.

Ethical approval

Carpesium longifolium does not belong to the national key protected wild plants category, and collecting it does not violate the Regulations of the People's Republic of China on Wild Plants Protection. According to Article 5 of the regulation, the state encourages and supports scientific research on wild plants and on-site and ex-situ protection of wild plants. No ethical approval or specific permission was needed in this research.

Author contributions

Z.F., X.G. and H.C. were involved in the conception and design. H.C. was involved in the drafting of the paper. T.L., X.C. and T.Q. completed assembly and annotation of the genomes. P.L., Y. W. and X.Z. analyzed genomes. Z.F. and X.G. revised the content. All authors agree to be accountable for all aspects of the work.

Disclosure statement

No potential conflict of interest was reported by the author(s). There is no conflict of interest between Sichuan Normal University and Sichuan Tianshengyuan Environmental Services Company.

Funding

This study was financially supported by the National Natural Science Foundation of China (No. 32000158, 31970225), the National Science & Technology Fundamental Resources Investigation Program of China (No. 2021XJKK0702), the Foundation of Sustainable Development Research Center of Resources and Environment of Western Sichuan, Sichuan Normal University (No. 2020CXZYHJZX03), Key Laboratory of Chemistry in Ethnic Medicinal Resources (Yunnan Minzu University), State Ethnic Affairs Commission & Ministry of Education (No. MZY2301) and Laboratory equipment research projects, Sichuan Normal University (No. SYJS20220014), and the research project of Xizang Shengyuan Environmental Engineering Co., Ltd 302 & Sichuan Tianshengyuan Environmental Services Co., Ltd (No. xzkyxm-2023-001)

Data availability statement

After uploading the data, the NCBI database releases two accession numbers of *C. longifolium*. These two accession numbers contain exactly the same sequence information and author. Data are available in the NCBI GenBank at https://www.ncbi.nlm.nih.gov (accession number: OQ555798). The associated BioProject, SRA, and BioSample numbers are PRJNA1097324, SRR28586436, SAMN40867589 respectively.

References

Allen GC, Flores-Vergara MA, Krasynanski S, Kumar S, Thompson WF. 2006. A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. Nat Protoc. 1(5):2320–2325. doi:10.1038/nprot.2006.384.

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. doi:10.1089/cmb.2012.0021.

Daniell H, Lin CS, Yu M, Chang WJ. 2016. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol. 17(1):134. doi:10.1186/s13059-016-1004-2.

Doorduin L, Gravendeel B, Lammers Y, Ariyurek Y, Chin-A-Woeng T, Vrieling K. 2011. The complete chloroplast genome of 17 individuals of pest species *Jacobaea vulgaris*: SNPs, microsatellites and barcoding markers for population and phylogenetic studies. DNA Res. 18(2):93–105. doi:10.1093/dnares/dsr002.

Downie SR, Jansen RK. 2015. A comparative analysis of whole plastid genomes from the Apiales: expansion and contraction of the inverted repeat, mitochondrial to plastid transfer of DNA, and identification of highly divergent noncoding regions. Syst Bot. 40(1):336–351. doi:10.1600/036364415X686620.

Englund M, Pornpongrungrueng P, Gustafsson MHG, Anderberg AA. 2009. Phylogenetic relationships and generic delimitation in Inuleae subtribe Inulinae (Asteraceae) based on ITS and cpDNA sequence data. Cladistics. 25(4):319–352. doi:10.1111/j.1096-0031.2009.00256.x.

Hereward JP, Werth JA, Thornby DF, Keenan M, Singh CB, Walter GH. 2018. Complete chloroplast genome of glyphosate resistant *Sonchus oleraceus* L. from Australia, with notes on the small single copy (SSC) region orientation. Mitochondrial DNA B Resour. 3(1):363–364. doi:10.1080/23802359.2018.1450682.

Jansen RK, Cai Z, Raubeson LA, Daniell H, Depamphilis CW, Leebens-Mack J, Müller KF, Guisinger-Bellian M, Haberle RC, Hansen AK, et al. 2007. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. Proc Natl Acad Sci U S A. 104(49):19369–19374. doi:10.1073/pnas.0709121104.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30(4):772–780. doi:10.1093/molbev/mst010.

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. doi:10.1093/bioinformatics/bts199.

Kim J-K, Park JY, Lee YS, Lee HO, Park H-S, Lee S-C, Kang JH, Lee TJ, Sung SH, Yang T-J, et al. 2016. The complete chloroplast genome sequence of the *Taraxacum officinale* F.H. Wigg (Asteraceae). Mitochondrial DNA B Resour. 1(1):228–229. doi:10.1080/23802359.2016.1155425.

Li R, Ma PF, Wen J, Yi TS. 2013. Complete sequencing of five Araliaceae chloroplast genomes and the phylogenetic implications. PLoS One. 8(10):e78568. doi:10.1371/journal.pone.0078568.

Ling R, Chen YL, Shi T. 1985. Flora of China. Vol. 75. Beijing: Science Press; p. 295–310.

Liu X, Luo J, Zhang M, Wang Q, Liu J, Wu D, Fu Z. 2023. Phylogenomic analysis of two species of *Parasenecio* and comparative analysis within tribe Senecioneae (Asteraceae). Diversity. 15(4):563. doi:10.3390/d15040563.

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.

Li X-W, Weng L, Gao X, Zhao Y, Pang F, Liu J-H, Zhang H-F, Hu J-F. 2011. Antiproliferative and apoptotic sesquiterpene lactones from *Carpesium faberi*. Bioorg Med Chem Lett. 21(1):366–372. doi:10.1016/j.bmcl.2010.10.138.

Mehmood F, Shahzadi I, Ahmed I, Waheed MT, Mirza B, Abdullah. 2020. Characterization of *Withania somnifera* chloroplast genome and its comparison with other selected species of Solanaceae. Genomics. 112(2):1522–1530. doi:10.1016/j.ygeno.2019.08.024.

Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES science gateway for inference of large phylogenetic trees. Proceedings of the Gateway Computing Environments Workshop (GCE) New Orleans; p. 1–8.

Nylinder S, Anderberg AA. 2015. Phylogeny of the Inuleae (Asteraceae) with special emphasis on the Inuleae-Plucheinae. Taxon. 64(1):110–130. doi:10.12705/641.22.

Palmer JD. 1985. Comparative organization of chloroplast genomes. Annu Rev Genet. 19(1):325–354. doi:10.1146/annurev.ge.19.120185.001545.

Qu XJ, Moore MJ, Li DZ, Yi TS. 2019. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. Plant Methods. 15(1):50. doi:10.1186/s13007-019-0435-7.

Quan G, Li J. 2020. The complete chloroplast genome of *Chromolaena odorata* (Compositae), a noxious invasive weed to South China. Mitochondrial DNA B. 5(2):1510–1511. doi:10.1080/23802359.2020.1742217.

Shi X, Xu W, Wan M, Sun Q, Chen Q, Zhao C, Sun K, Shu Y. 2022. Comparative analysis of chloroplast genomes of three medicinal *Carpesium* species: genome structures and phylogenetic relationships. PLoS One. 17(8):e0272563. doi:10.1371/journal.pone.0272563.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 30(9):1312–1313. doi:10.1093/bioinformatics/btu033.

Thode VA, Oliveira CT, Loeuille B, Siniscalchi CM, Pirani JR. 2021. Comparative analyses of *Mikania* (Asteraceae: Eupatorieae) plastomes and impact of data partitioning and inference methods on phylogenetic relationships. Sci Rep. 11(1):13267. doi:10.1038/s41598-021-92727-6.

Vargas OM, Ortiz EM, Simpson BB. 2017. Conflicting phylogenomic signals reveal a pattern of reticulate evolution in a recent high-Andean diversification (Asteraceae: Astereae: *Diplostephium*). New Phytol. 214(4):1736–1750. doi:10.1111/nph.14530.

Wicke S, Schneeweiss GM, Depamphilis CW, Müller KF, Quandt D. 2011. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Mol Biol. 76(3–5):273–297. doi:10.1007/s11103-011-9762-4.

Yang YX. 2016. Studies on sesquiterpene lactones from *Carpesium faberi*. Zhongguo Zhong Yao Za Zhi. 41(11):2105–2111. doi:10.4268/cjcmm20161121.

Yang Y-X, Shan L, Liu Q-X, Shen Y-H, Zhang J-P, Ye J, Xu X-K, Li H-L, Zhang W-D. 2014. Carpedilactones A-D, four new isomeric sesquiterpene lactone dimers with potent cytotoxicity from *Carpesium faberi*. Org Lett. 16(16):4216–4219. doi:10.1021/ol501923k.

Yun SA, Gil HY, Kim SC. 2017. The complete chloroplast genome sequence of *Saussurea polylepis* (Asteraceae), a vulnerable endemic species of Korea. Mitochondrial DNA B Resour. 2(2):650–651. doi:10.1080/23802359.2017.1375881.

Zhang J-P, Wang G-W, Tian X-H, Yang Y-X, Liu Q-X, Chen L-P, Li H-L, Zhang W-D. 2015. The genus *Carpesium*: a review of its ethnopharmacology, phytochemistry and pharmacology. J Ethnopharmacol. 163:173–191. doi:10.1016/j.jep.2015.01.027.

Zhang Y, Zhang J, Yang Y, Liu Q. 2017. Complete chloroplast genome of *Pluchea indica* (L.) Less. (Asteraceae) and its phylogenetic analysis. Mitochondrial DNA B Resour. 2(2):918–919. doi:10.1080/23802359.2017.1413299.