





## Phylogenomic Analysis of *Dichrocephala benthamii* and Comparative Analysis within Tribe Astereae (Asteraceae)

Hui Chen<sup>1,2</sup>, Tingyu Li<sup>2</sup> , Xinyu Chen<sup>2</sup>, Xinyi Zheng<sup>2</sup>, Tianmeng Qu<sup>2</sup>, Bo Li<sup>3</sup> and Zhixi Fu<sup>1,2,4</sup> 

<sup>1</sup>Sichuan Normal University, Key Laboratory of Land Resources Evaluation and Monitoring in Southwest, Chengdu, Sichuan, China.

<sup>2</sup>Sichuan Normal University, College of Life Sciences, Chengdu, Sichuan, China.

<sup>3</sup>Sichuan Environmental Monitoring Center, Chengdu, Sichuan, China.

<sup>4</sup>Sichuan Normal University, Sustainable Development Research Center of Resources and Environment of Western Sichuan, Chengdu, Sichuan, China.

### Abstract

*Dichrocephala benthamii* C. B. Clarke has long been used as traditional Chinese medicine. However, the chloroplast (cp) genome of *D. benthamii* is poorly understood so far. In this study, we sequenced and analyzed the cp genome of *D. benthamii*. The results showed that the cp genome is 152,350 bp in length, with a pair of inverted repeat regions (IRa and IRb, each 24,982 bp), a large single-copy (LSC) region comprising 84,136 bp, and a small single-copy (SSC) region comprising 18,250 bp. The GC content of the cp genome was 37.3%. A total of 134 genes were identified, including 87 protein-coding genes (CDS), 38 tRNA genes, 8 rRNA genes, and 1 pseudogene (*ycf1*). Expansion or contraction of IR regions were detected in *D. benthamii* and other species of the tribe Astereae. Additionally, our analyses showed the types of sequence repeats and the highly variable regions discovered by analyzing the border regions, sequence divergence, and hot spots. The phylogenetic analysis revealed *D. benthamii* is the basal group of Astereae. The results of this study will be a significant contribution to the genetics and species identification related to *D. benthamii*.

**Keywords:** *Dichrocephala benthamii*, chloroplast genome, phylogenetic analysis, comparative analysis.

Received: December 08, 2024; Accepted: August 09, 2024.

The family Asteraceae contains about 1600-1700 genera and 26,000 species (Funk *et al.*, 2005). The plants in Asteraceae are characterized by the distinctive capitula. The tribe Astereae, a second largest tribe of Asteraceae, includes about 225 genera and 3,100 species, of which 29 genera and 237 species are native to China (112 species are endemic) (Brouillet *et al.*, 2009). The whole plant of *D. benthamii* is used medicinally as common herb among Dai nationality of China for the treatment of indigestion, common cold, fever in children, pneumonia and hepatitis (Song *et al.*, 2017). Previous research on *D. benthamii* has focused on medicinal and pharmacological studies (Song *et al.*, 2017).

The analysis of cp genomes has become a major research focus in plant evolution and systematics (Vargas *et al.*, 2017). To date, sequences of *D. benthamii* (*rbcL*, *matK*, and *rpoC*) have been reported (<https://www.ncbi.nlm.nih.gov>). However, available genetic data for comparative genomic studies of *D. benthamii* and related genera are limited. In the study, we sequenced, assembled and annotated the complete cp genome of *D. benthamii*. The objectives of this study were to: 1) identify and characterize the cp genome structure and sequence differentiation throughout the plastids; and 2) assess the phylogenetic relationships among *D. benthamii* and

other Asteraceae species, which may be useful for further speciation studies.

Fresh leaves of *D. benthamii* were collected from Yuexi county, Liangshan Prefecture, Sichuan Province, China (Figure S1). The voucher specimen was collected and placed in the herbarium of the Sichuan Normal University, China (SCNU) (Contact: Zhixi Fu, [fuzx2017@sicnu.edu.cn](mailto:fuzx2017@sicnu.edu.cn)) under the voucher number: Junjia Luo 311. Total genomic DNA was isolated using a modified CTAB method (Allen *et al.*, 2006). DNA libraries were constructed using the Illumina Paired-End DNA Library Kit. The qualified library was sequenced using the Illumina NovaSeq 6000 platform with a sequencing read length was 150 bp. The cp genome of *D. benthamii* was assembled using SPAdes software (v3.15.1) (Bankevich *et al.*, 2012). Subsequently, the results were annotated using PGA based on the reference cp genome sequence of *Eschenbachia blinii* (H.Lév.) (NC 037605.1) (Qu *et al.*, 2019). The cp genome sequence of *D. benthamii* was uploaded to NCBI with the accession number ON751565. The complete cp genome map was constructed using OGDRAW (Greiner *et al.*, 2019).

IRscope is a bioinformatics tool used to visualize the expansions and contractions of cp genomes (Amiryousefi *et al.*, 2018). Additionally, as an online server, mVISTA was employed to compare DNA sequences of 6 species in tribe Astereae (Frazer *et al.*, 2004).

Simple sequence repeats (SSRs) in the plastomes were detected using the Perl script MISA (Beier *et al.*, 2017). The repeat units were set to 10 for mononucleotides, 5 for

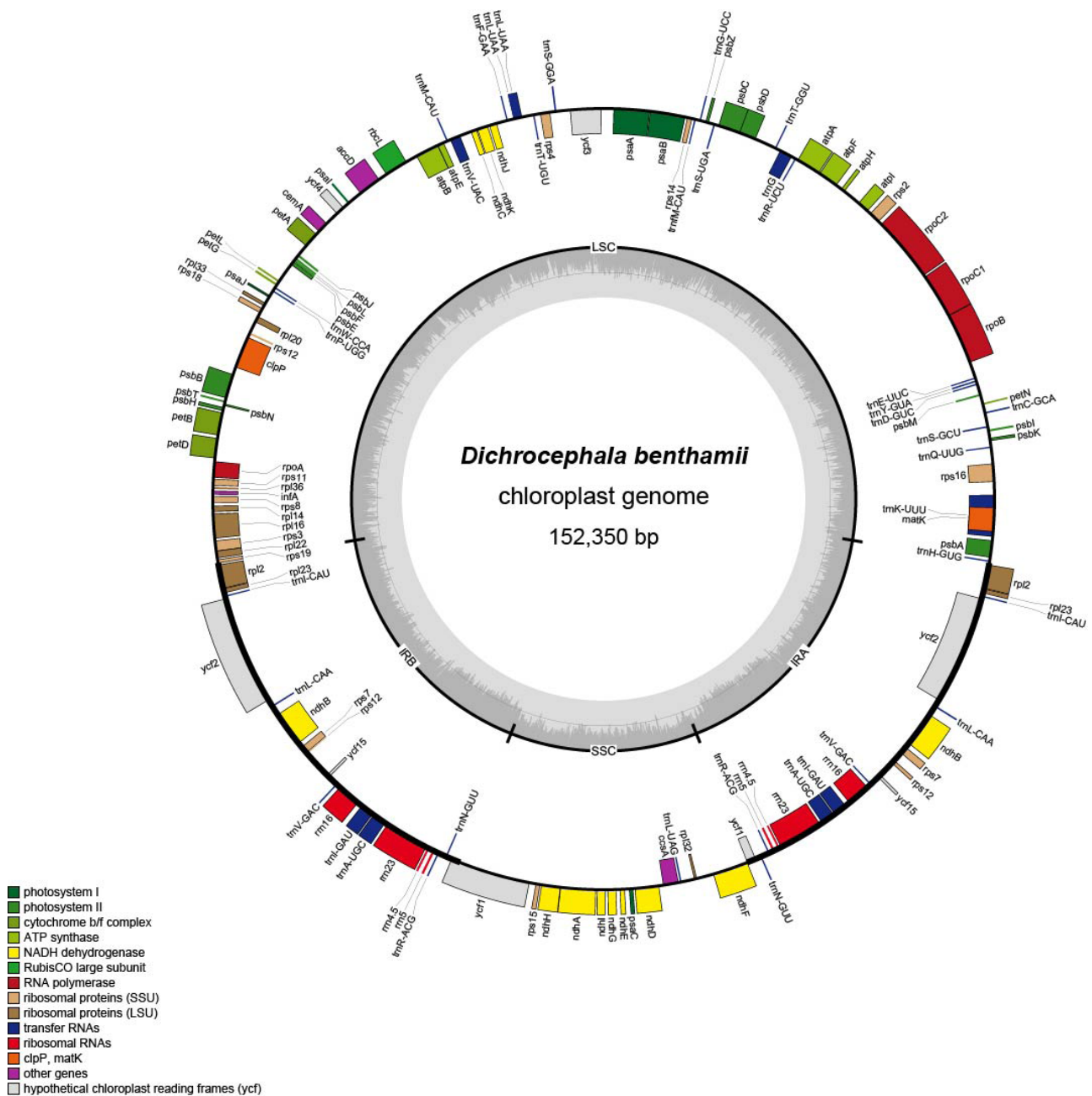
dinucleotides, 4 for trinucleotides and 3 for hexanucleotides, respectively.

MEGA v.7.0 was used to analyze the synonymous codon usage and the relative synonymous codon usage (RSCU) of the *D. benthamii* cp genome.

The 28 complete cp sequences were downloaded from NCBI and combined with the sequenced complete cp sequence of *D. benthamii*. *Achillea millefolium* L. and *Ajanía pacífica* (Nakai) K. Bremer & Humphries (tribe Anthemideae) were included as outgroups (Table S1). Phylogenetic relationship reconstruction analysis using Maximum Likelihood (ML) method was conducted with RAxML (Stamatakis, 2014)

based on the GTRGAMMA model on the CIPRES platform (Miller *et al.*, 2010). Default settings were used for other parameters. Bootstrap analysis with 1,000 replicates was performed to assess bootstrap values (BS) for each node of the phylogenetic tree.

The complete cp genome of *D. benthamii* was 152,350 bp in size. It exhibited the quadripartite structure consisting of two IR regions (24,982 bp each), a LSC region (84,136 bp), and a SSC region (18,250 bp) (Figure 1). The overall GC content of the *D. benthamii* cp genome was 37.3%, similar to that of the other 5 Astereae species (37.28% to 37.36%) (Table 1). The GC contents of the SSC regions, LSC region and SSC region



**Figure 1** - Gene map of the complete chloroplast genomes of *D. benthamii*. Annotated genes are colored according to functional categories whereby the genes outside the circle were transcribed clockwise, while the genes placed inside the circle were transcribed counterclockwise. The dark grey color in the inner circle represents GC content, whereas the light grey color corresponds to AT content.

among 6 Astereae species varied from 31.2% to 31.33%, from 35.15% to 35.3% and from 42.99% to 43.06%, respectively. The consistent GC content may play a role in maintaining genetic stability (Niu *et al.*, 2017). *D. benthamii* possessed 134 unique genes, comprising 87 protein-coding genes, 38 tRNA genes, and 8 rRNA genes. Seven tRNA genes and all rRNA genes were located in the IR regions, contributing to the high GC content of these regions. This phenomenon has been associated with the presence of NADH (Shen *et al.*, 2017). Fifteen genes contained one intron, while three genes possessed two introns (Table S2).

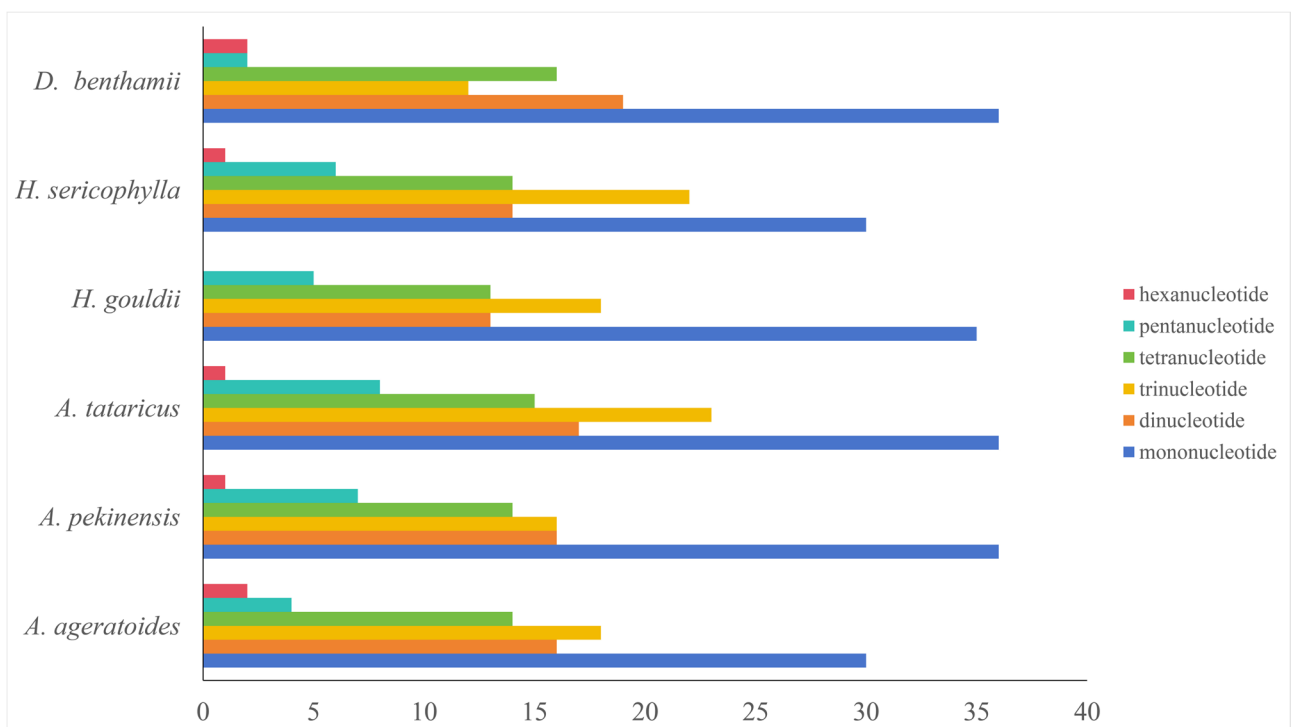
Analysis of the IR boundaries in six Astereae species revealed varying contractions and expansions of the IR, leading to variability in genome length (Kim and Lee, 2004) (Figure S2). The *rps19* was located at the LSC-IRb border regions, with variations in sizes (17bp to 62bp). The *ycf1* spanned the SSC-IRa junction. The *rpl12* was entirely within the IR region, located 115bp away from the LSC. However, the pseudogene *ycf1* of *D. benthamii* spanned the boundary between the SSC and IRa regions (6bp). The mVISTA-based

identity plot revealed high sequence similarity with a few variants. These variations were typically observed in the intergenic spacers (IGS) rather than coding-regions, suggesting that coding regions were more conserved than non-coding regions (Figure S3). Overall, more variations were observed in the LSC and SSC regions compared to the IR regions. This result is consistent with patterns observed in cp genomes of other Asteraceae species (Loeuille *et al.*, 2021). The high variation observed in the LSC and SSC regions was primarily attributed to non-coding sequences.

In this study, 532 SSRs were identified across six species of Astereae, with their counts being very similar (ranging from 84 to 100) (Figure 2). *Aster tataricus* L.f. exhibited the highest number of SSRs (100), while *Aster ageratoides* Turcz. and *Heteropappus gouldii* (C.E.C.Fisch.) Grierson had the lowest (84). The detected SSRs encompassed six types: mononucleotides (38.16%), dinucleotides (17.86%), trinucleotides (20.49%), tetranucleotides (16.17%), pentanucleotides (6.02%), and hexanucleotides (1.32%) (Figure 2). These SSRs could be utilized to explore the

**Table 1** – Summary of the complete chloroplast genomes of 6 species of Astereae.

| Species                          | Genome Size (bp) | LSC (bp) | IR (bp) | SSC (bp) | GC Content (%) |       |       |       |
|----------------------------------|------------------|----------|---------|----------|----------------|-------|-------|-------|
|                                  |                  |          |         |          | All            | LSC   | IR    | SSC   |
| <i>Dichrocephala benthamii</i>   | 152,350          | 84,136   | 24,982  | 18,250   | 37.3           | 35.25 | 42.98 | 31.2  |
| <i>Aster ageratoides</i>         | 153,071          | 84,896   | 24,953  | 18,269   | 37.28          | 35.15 | 43.06 | 31.33 |
| <i>Aster pekinensis</i>          | 152,815          | 84,530   | 25,033  | 18,219   | 37.3           | 35.22 | 42.99 | 31.33 |
| <i>Aster tataricus</i>           | 152,992          | 84,698   | 25,022  | 18,250   | 37.26          | 35.15 | 43.01 | 31.26 |
| <i>Heteropappus gouldii</i>      | 152,450          | 84,226   | 25,011  | 18,202   | 37.36          | 35.3  | 43.03 | 31.3  |
| <i>Heteropappus sericophylla</i> | 152,214          | 84,369   | 24,983  | 18,293   | 37.32          | 35.23 | 43.04 | 31.32 |



**Figure 2** – Analysis of SSRs in 6 species of Astereae plastid genomes species.

genetic structure, diversity, phylogeny, and differentiation of Astereae and other *Dichrocephala* species. Additionally, 80 SSRs were identified in *D. benthamii*, comprising motifs such as A/T, AT/AT, AAT/AAT, AAAT/ATTT, AATT/AATT, AAAT/AGTTT, AATAT/ATATT (Table S3). The preference for AT-rich motifs is consistent with findings from many plant plastids (Zavala-Paez *et al.*, 2020).

By calculating the RSCU values for all protein-coding genes in the cp genomes of *D. benthamii*, a total of 30,600 codons were identified. Thirty-one types of codons exhibited greater preference (RSCU > 1) (Figure S4). Serine showed no preferences (RSCU = 1), while the remaining codons were less preferred. Notably, no codons were extremely rare (RSCU < 0.1). Among the 20 amino acids, leucine (10.67%) constituted the largest proportion, while cysteine (1.12%) accounted for the smallest (Table S4). In other angiosperm cp genomes, leucine and cysteines have been reported the most and least abundant amino acids, respectively (Sharp and Li, 1987). Intriguingly, with the exception of UUG, all preferentially used codons were ended with A/U. This result is consistent with observations in other Astereae species (Chen *et al.*, 2024). The high proportion of A/U is the major force of deviation (Claude and Park, 2020).

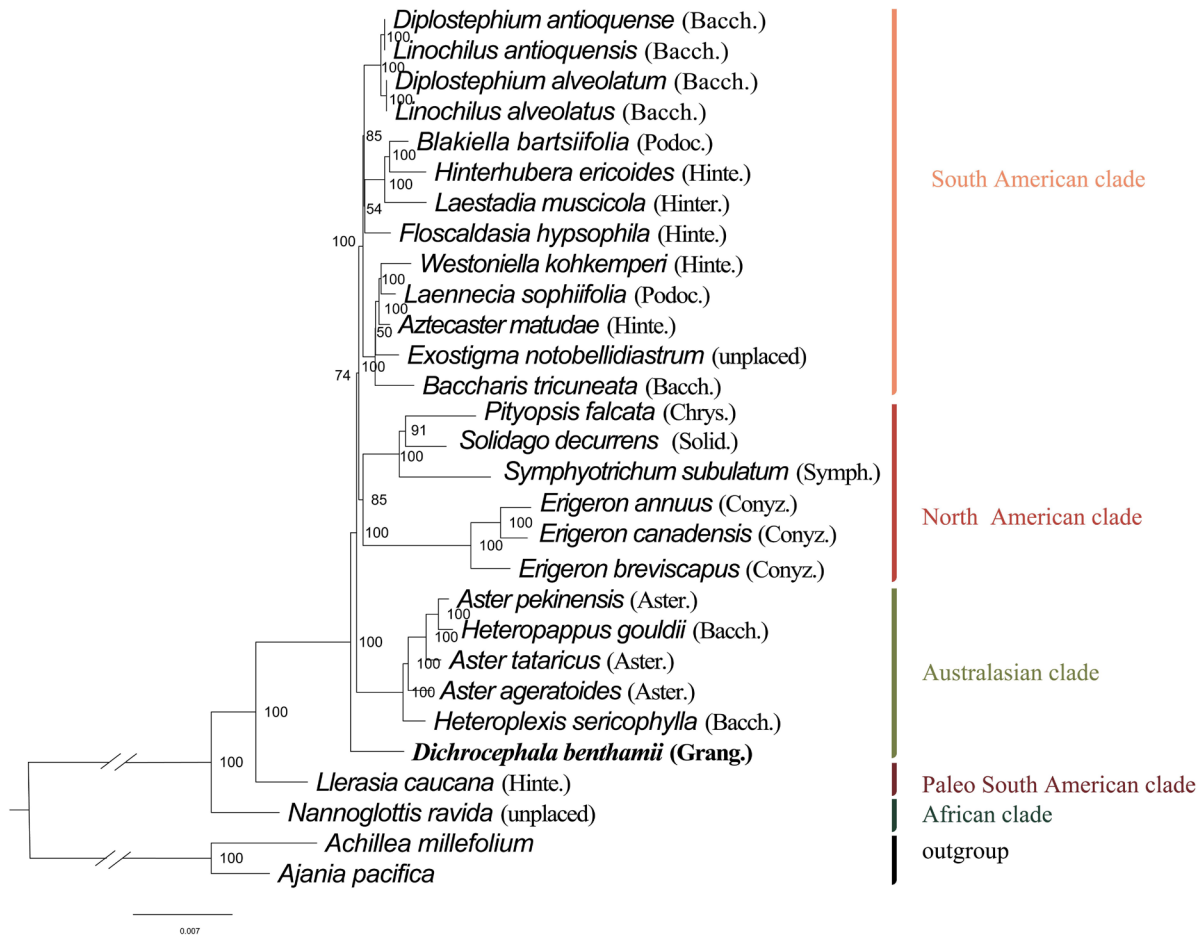
Phylogenomic analysis based on cp genome data identified several clades (Figure 3). The tribe Astereae was

found to form a monophyletic group. *D. benthamii* and other Asteraceae species clustered into a clade supported by a bootstrap value of 100%. *D. benthamii* was located at the base of the phylogenetic tree, which was in agreement with Brouillet *et al.* (2009). This study will fill a gap in the research of the cp genome of *Dichrocephala* plants and provide a wealth of information for the taxonomic study of this genus in Asteraceae.

This study firstly reported the cp genomes of *D. benthamii*. Comparative analyses revealed the genome's structure and composition, variable regions and SSR markers. Phylogenetic analysis indicated *D. benthamii* is at the base of the phylogenetic tree in Astereae. Thus, the complete cp genome of *D. benthamii* provides valuable genetic insights into this genus and establishes a foundation for investigating population evolution among Astereae species.

### Acknowledgements

The authors would like to thank the editor and anonymous reviewers for the constructive criticism of the original manuscript. This study was financially supported by the National Natural Science Foundation of China (No. 32000158), the National Science & Technology Fundamental Resources Investigation Program of China (No. 2021XJKK0702) and



**Figure 3** – Maximum likelihood tree of *D. benthamii* reconstructed based on 29 complete cp genome sequences. The species names in bold font represent our sequenced species plastomes. (the subtribe is marked by the Brouillet *et al.*, 2009; Bacch.: Baccharidinae; Podoc.: Podocominiae; Hinter.: Hinterhuberinae; Chrys.: Chrysopsidinae; Solid.: Solidagininae; Symph.: Symphyotrichinae; Conyz.: Conyzinae; Aster.: Asterinae; Grang.: Grangeinae; pentagram: Astereae originated in Africa).

the Foundation of Sustainable Development Research Center of Resources and Environment of Western Sichuan, Sichuan Normal University (No. 2020CXZYHJZX03).

### Conflict of Interest

The authors declare no conflict of interest.

### Author Contributions

HC, BL and ZF conceived and designed the research. HC and TL performed bioinformatic analyses. XC and XZ carried out wetlab experiments. TQ contributed to data interpretation. HC wrote the manuscript. BL and ZF revised the manuscript. All authors read and approved the manuscript. All authors have read and agreed to the published version of the manuscript.

### Data Availability

The cp genome sequence of *D. benthamii* was uploaded to NCBI with the accession number ON751565.

### References

- Allen GC, Flores-Vergara MA, Krasnyanski S, Kumar S and Thompson WF (2006) A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. *Nature Protocols* 1:2320–2325.
- Amiryousefi A, Hyvonen J and Poczar P (2018) IRscope: An online program to visualize the junction sites of chloroplast genomes. *Bioinformatics* 34:3030–3031.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD *et al.* (2012) SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477.
- Beier S, Thiel T, Munch T, Scholz U and Mascher M (2017) MISA-web: A web server for microsatellite prediction. *Bioinformatics* 33:2583–2585.
- Brouillet L, Lowrey TK, Urbatsch L, Karaman-Castro V, Sancho G, Wagstaff S and Semple JC (2009) Astereae. In: Funk SVA, Stuessy A, Bayer RT (eds) *Systematics, evolution and biogeography of the Compositae*. IAPT, Vienna, pp 449–490.
- Claude SJ and Park SJ (2020) *Aster spathulifolius* Maxim. a leaf transcriptome provides an overall functional characterization, discovery of SSR marker and phylogeny analysis. *PLoS One* 15:e0244132.
- Chen H, Li TY, Chen XY, Qu TM, Zheng XY, Luo J, Li B, Zhang GJ and Fu ZX (2024) Insights into comparative genomics, structural features, and phylogenetic relationship of species from Eurasian *Aster* and its related genera (Asteraceae: Astereae) based on complete chloroplast genome. *Front Plant Sci* 15:1367132.
- Frazer KA, Pachter L, Poliakov A, Rubin EM and Dubchak I (2004) VISTA: Computational tools for comparative genomics. *Nucleic Acids Res* 32:W273–W279.
- Funk VA, Bayer RJ, Keeley S, Chan R, Watson L, Gemeinholzer B, Schilling E, Panero JL, Baldwin BG, Garcia-Jacas N, *et al.* (2005) Everywhere but Antarctica: Using a supertree to understand the diversity and distribution of the Compositae. *Biologiske Skrifter* 55:343–374.
- Greiner S, Lehwark P and Bock R (2019) OrganellarGenomeDRAW (OGDRAW) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res* 47:W59–W64.
- Kim KJ and Lee HL (2004) Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants. *DNA Res* 11:247–261.
- Loeuille B, Thode V, Siniscalchi C, Andrade S, Rossi M and Pirani JR (2021) Extremely low nucleotide diversity among thirty-six new chloroplast genome sequences from *Aldama* (Heliantheae, Asteraceae) and comparative chloroplast genomics analyses with closely related genera. *PeerJ* 9:e10886.
- Miller MA, Pfeiffer WT and Schwartz T (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: *Proceedings of the Gateway Computing Environments Workshop (GCE)*, New Orleans, p. 1–8.
- Niu ZT, Xue QY and Wang H (2017) Mutational biases and GC-biased gene conversion affect GC content in the plastomes of *Dendrobium* genus. *Int J Mol Sci* 18:2307.
- Qu XJ, Moore MJ, Li DZ and Yi TS (2019) PGA: A software package for rapid, accurate, and flexible batch annotation of plastomes. *Plant Methods* 15:50.
- Sharp PM and Li WH (1987) The codon Adaptation Index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res* 15:1281–1295.
- Shen XF, Wu ML, Liao BS, Liu ZX, Bai R, Xiao SM, Li XW, Zhang BL, Xu J and Chen SL (2017) Complete chloroplast genome sequence and phylogenetic analysis of the medicinal plant *Artemisia annua*. *Molecules* 22:1330.
- Song B, Si JG, Yu M, Tian XH, Ding G and Zou ZM (2017) Megastigmene glucosides isolated from *Dichrocephala benthamii*. *Chin J Nat Med* 15:288–291.
- Stamatakis A (2014) RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313.
- Vargas OM, Ortiz EM and Simpson BB (2017) Conflicting phylogenomic signals reveal a pattern of reticulate evolution in a recent high-Andean diversification (Asteraceae: Astereae: *Diplostephium*). *New Phytol* 214:1736–1750.
- Zavala-Paez M, Vieira LD, de Baura VA, Balsanelli E, de Souza EM, Cevallos MC, Chase MW and Smidt ED (2020) Comparative plastid genomics of Neotropical *Bulbophyllum* (Orchidaceae; Epidendroideae). *Front Plant Sci* 11:799.

### Supplementary material

The following online material is available for this article:

Table S1 – information on the 28 species used in the study.

Table S2 – List of genes found in *D. benthamii*.

Table S3 – SSR in the cp genome of *D. benthamii*.

Table S4 – Codon usage in the chloroplast genomes of *D. benthamii*.

Figure S1 – The photos of *D. benthamii* were taken by Xudong, Ma without any copyright issues.

Figure S2 – Comparison for border positions of LSC, SSC, and IR regions among the 6 species of tribe Astereae.

Figure S3 – The cp genomes of 6 species of Astereae were compared using the mVista program with *D. benthamii* as the reference.

Figure S4 – Codon content of 20 amino acid and stop codons in all protein-coding genes of the CP genome of *D. benthamii*.

Associate Editor: Rogerio Margis

License information: This is an open-access article distributed under the terms of the Creative Commons Attribution License (type CC-BY), which permits unrestricted use, distribution and reproduction in any medium, provided the original article is properly cited