

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

Hui Chen

Sichuan Normal University

Junjia Luo

Sichuan Normal University

Xiaofeng Liu

Sichuan Normal University

Tingyu Li

Sichuan Normal University

Tianmeng Qu

Sichuan Normal University

Bo Li

Sichuan Normal University

Zhixi Fu (✉ [fuzx2017@sicnu.edu.cn](mailto:fuzx2017@sicnu.edu.cn))

Sichuan Normal University

---

## Research Article

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

**Posted Date:** June 15th, 2023

**DOI:** <https://doi.org/10.21203/rs.3.rs-3049453/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

# 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 have sequenced and analyzed the cp genome of *D. benthamii*. The results show 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*). The plastome includes 87 simple repeats and 48 long repeats. The phylogenetic analysis reveals *D. benthamii* is the basal group of Astereae. Therefore, we confirm that the taxa of Astereae emerges from southern Africa and dispersed to other big continents.

The results of this study are a significant contribution to the field of genetics and species identification related to *D. benthamii*.

## 1. INTRODUCTION

The tribe Astereae is the second largest tribe of the Asteraceae family, consisting of approximately 225 genera and 3,100 species. Within China, there are a total of 29 genera and 237 species of the tribe, including 112 endemic species (112 species are endemic) (Brouillet et al. 2004; Funk et al. 2009; Panero et al. 2008; Fu et al. 2016; Noyes and Rieseberg, 1999; Brouillet et al. 2009). It inhabits worldwide except for Antarctica, with the highest diversity in temperate region. The genus of *Dichrocephala* is a small genus belonging to the tribe Astereae with 6 species, distributed across tropical Oceania, Africa, and Asia (Chen et al. 2011). *D. benthamii* is distributed across Gansu, Guangxi, Guizhou, Hubei, Sichuan, Xizang, Yunnan provinces, China (Bhutan, Cambodia, India, Laos, Nepal, Vietnam). The genus of *Dichrocephala* is mostly used as traditional Chinese medicine (Ai et al. 2013). Among them, *D. benthamii* is a common herb among the 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; Tian et al. 2013a; Tian et al. 2013b).

Chloroplast is the essential organelle that plays a critical role in providing energy for plant metabolism through photosynthesis. The plastome has traditionally been regarded as a single evolutionary unit, genes can be concatenated, which is useful in analyzing phylogenetic signals (Shaw et al. 2007). The chloroplast genome has been proven a valuable tool for gene discovery and genetic engineering of plants due to its unique features (Wang et al. 2023). The quadripartite structure, with a large single-copy region (LSC), a small single-copy region (SSC) and a pair of inverted repeats regions (IRa and IRb), is a defining characteristic of chloroplast genomes. The chloroplast genome typically ranges in size from 115–165 kb (Leseberg et al. 2009). The plastome genome is generally more conserved in size and structure than the nuclear genome, which undergoes frequent rearrangements during evolution (Smith et al. 2015; Wicke et al. 2011; Tonti-Filippini et al. 2017). The comparison of complete cp sequences is a powerful tool for exploring sequence variation and molecular evolutionary patterns in plants (Nie et al. 2012; Shen et al. 2017; Cai et al. 2006; Maier et al. 1995; Zhong et al. 2019). Based on these characteristics of the chloroplast genome, the cp genome further provides the opportunity for exploring deep relationships in the plant tree of life (Gitzendanner et al. 2018; Shahzadi et al. 2020; Parks et al. 2009; Jansen et al.; 2007). The cp genome has been proposed as a super-barcode for species identification (Chen et al. 2018; Krawczyk et al. 2018).

The availability of high-throughput sequencing technology has greatly accelerated the pace of research on chloroplast genome evolution. The analysis of chloroplast genomes has become a major research focus in the field of plant evolution and systematics (Zhang et al. 2019; Kumar et al. 2009; Vargas et al. 2017; Palazzesi et al. 2002). Previously, some sequences such as *rbcL* (Liu et al. 2004), *ndhF* (Liu et al. 2004), *matK* (Enke et al. 2008) have been studied. Some complete chloroplast genomes of the tribe Astereae have been reported (eg. *Crepidiastrum* Nakai (Do et al. 2019); *Nouelia* Franch. (Liu et al. 2022); *Cavea* W.W. Sm. & J. Small (Yu et al. 2022); *Aster* L. (Park et al. 2017; Ou et al. 2019; Wang et al. 2019); *Erigeron* L. (Zhou et al. 2019); *Saussurea* DC. (Zhnag et al. 2019); *Conyza* Less. (Zhang et al. 2019); *Heteroplexis* C.C.Chang (Duan et al. 2022). Besides, a LSC inversion from *trnE-UUC* to *rpoB* and an SSC inversion from *rps15* to *ndhF* were detected in some lineages of Asteraceae (Liu et al. 2012; Salih et al. 2017; Hu et al. 2020). To date, some sequences of *D. benthamii* have been reported *rbcL* (Saslis-Lagoudakis et al. 2012), *matK* (Gao et al. 2010), *rpoC* (Gao et al. 2010). However, the systematic and comparative study

of the mysterious *Dichrocephala* chloroplast genome has not yet been published in Genbank. Further study of the evolutionary history and status of *Dichrocephala* is not beneficial. Available genetic data for comparative genomic studies of *D. benthamii* and related genera is limited.

Over the past two decades, the field of molecular phylogenetics has made significant advances in the evolutionary relationships of the tribe Astereae (Fu et al. 2016; Noyes and Rieseberg, 1999; Brouillet et al. 2009; Fiz et al. 2002; Morgan et al. 1997; Noyes and Rieseberg, 2000; Li et al. 2012). While the tribe Astereae has a wide distribution, there are several genera within the tribe that are locally endemic to specific regions such as North America, South America, and East Asia. A molecular phylogenetic study by Noyes and Rieseberg (Noyes and Rieseberg, 1999) suggests that the Astereae originated in Africa. The studies of Funk (Funk et al. 2009) and Brouillet (Brouillet et al. 2009) were agreed with this view. The biogeographic history of Astereae suggests that there was a development in Africa during the Tertiary which was followed by repeated dispersals to eastern Asia and to other Southern Hemisphere continents, including South America, New Zealand, and Australia, either directly or via Antarctica (Brouillet et al. 2009). These dispersals contributed to the diversity and distribution of Astereae plants around the world today.

In the study, we sequenced, assembled and annotated the complete chloroplast genome of *D. benthamii*. The objectives of this study were to: 1) identify and characterize the chloroplast genome structure, sequence variation, mutation hotspot regions and duplicated regions throughout the plastid; 2) elucidate the phylogenetic relationships between *D. benthamii* and other Astereae; 3) determine the position of *D. benthamii* in the phylogenetic tree.

## 2. MATERIALS AND METHODS

### 2.1. Plant material, DNA extraction and sequencing

Fresh leaves of *D. benthamii* were collected from Yuexi county, Liangshan prefecture, Sichuan Province, China. The voucher specimen was collected and placed in the herbarium of the Sichuan Normal University, China (SCNU) (Contact: Zhixi Fu, fuzx2017@sicnu.edu.cn) under the voucher number: Junjia Luo 311. Total genomic DNA was isolated from silica-dried leaves 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 with the Illumina NovaSeq 6000 platform. The sequencing read length was 150 bp. The cp genome sequence of *D. benthamii* was uploaded to NCBI with the accession number ON751565.

### 2.2. Chloroplast genome assembly and annotation

SPAdes software (v3.10.1) is a commonly used software for genome assembly (Bankevich et al. 2012). The cp genome of *D. benthamii* was assembled using it with default parameters. The Bandage was then used to identify circular maps to assess the quality of the assembly (Wick et al. 2015). Subsequently, the results were annotated using PGA based on the reference chloroplast genome sequence of *Eschenbachia blinii* (H.Lév.) (NC 037605.1) (Qu et al. 2019). The annotation results were then checked using Geneious R11 (Kease et al. 2012). The complete cp genome map was constructed using OGDRAW (Greiner et al. 2019). Plastid information analysis was performed using Geneious R11.

### 2.3. Comparative genome analysis and molecular marker identification

IRscope is a bioinformatics tool used to visualize the The expansion and contractions of the cp genomes (Amiryousefi et al. 2018). Besides, as an online server, mVISTA compares DNA sequences from 6 species of tribe Astereae.

### 2.4. Repeat sequences and SSR analysis

REPuter is a bioinformatics program that is specifically designed for finding and classifying various types of repetitive DNA sequences in genomic sequences (Kurtz et al. 2001). The parameters were as follows: the hamming distance was three, the maximum computed repeats were 50 bp, and the minimal repeat size was 30 bp. Simple sequence repeats (SSR) or microsatellites in the plastomes were detected by Perl script MISA (Beier et al. 2010). 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.

## 2.4. Phylogenetic analysis

The sequence matrix was constructed from the NCBI database. This study downloads 28 of the complete chloroplast sequences and combines them with the sequenced complete chloroplast sequence of *D. benthamii*. *Achillea millefolium* L. and *Ajania pacifica* (Nakai) K. Bremer & Humphries (tribe Anthemideae) were included as outgroups. Phylogenetic relationship reconstruction analysis by Maximum Likelihood (ML) method uses RAxML (Stamatakis et al. 2014) based on the GTRGAMMA model on the CIPRES (Miller et al. 2010). The default settings were selected for other parameters. 1,000 bootstrap repetitions were performed to calculate the bootstrap values (BS) for each node of the phylogenetic tree.

## 3. RESULTS

### 3.1. Chloroplast genome organization of *Dichrocephala benthamii*

The complete cp genome of *D. benthamii* was 152,350 bp in size. The cp genome of *D. benthamii* is the quadripartite structure consisting of two IR regions, a LSC region, and a SSC region (Fig. 1). These regions were comprised of 24,982 bp, 84,136 bp, and 18,250 bp, respectively (Table 1). Overall GC content of the *D. benthamii* cp genome was 37.3%. The GC content of the SSC and LSC regions was 31.2% and 35.3%, respectively. The GC content of the IR region was 43.0% in *D. benthamii*. The IRs contained rRNA and tRNA genes, which contributed to the total GC content of the IR. *D. benthamii* possessed 134 unique genes, including 87 protein-coding genes, 38 tRNA genes, and 8 RNA genes. 7 protein-coding genes (*ndhB*, *rpl2*, *rpl23*, *rps12*, *rps7*, *ycf15*, *ycf2*), 7 tRNA genes (*trnA-UGC*, *trnI-CAU*, *trnI-GAU*, *trnL-CAA*, *trnN-GUU*, *trnR-ACG*, *trnV-GAC*), and all rRNA genes (*rrn16*, *rrn23*, *rrn4.5*, *rrn5*) are located at the IR regions. 15 genes (*ndhA*, *ndhB*, *petB*, *petD*, *atpF*, *rpl16*, *rpl2*, *rps16*, *rpoC*, *trnA-UGC*, *trnG*, *trnI-GAU*, *trnK-UUU*, *trnL-UAA*, *trnV-UAC*) contain one intron, while 3 genes (*rps12*, *ycf3*, *clpP*) possess two introns (Table 2).

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

Table 2  
List of genes found in *Dichrocephala benthamii*.

Category of Genes	Group of Genes	Name of Genes	Number of genes
Photosynthesis	Subunits of photosystem I	<i>psaA, psaB, psaC, psal, psaJ</i>	5
	Subunits of photosystem II	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbl, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ</i>	15
	Subunits of NADH dehydrogenase	<i>ndhA*, ndhB*(x2), ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i>	12
	Subunits of cytochrome b/f complex	<i>petA, petB*, petD*, petG, petL, petN</i>	6
	Subunits of ATP synthase	<i>atpA, atpB, atpE, atpF*, atpH, atpI</i>	6
	Large subunit of rubisco	<i>rbcl</i>	1
Self-replication	Proteins of large ribosomal subunit	<i>rpl14, rpl16, rpl2*(x2), rpl20, rpl22, rpl23(x2), rpl32, rpl33, rpl36</i>	11
	Proteins of small ribosomal subunit	<i>#rps19, rps11, rps12**(x2), rps14, rps15, rps16*, rps18, rps19, rps2, rps3, rps4, rps7(x2), rps8</i>	15
	Subunits of RNA polymerase	<i>rpoA, rpoB, rpoC1*, rpoC2</i>	4
	Ribosomal RNAs	<i>rrn16(x2), rrn23(x2), rrn4.5(x2), rrn5(x2)</i>	8
	Transfer RNAs	<i>trnA-UGC*(2), trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnG*, trnG-UCC, trnH-GUG, trnI-CAU(2), trnI-GAU*(2), trnK-UUU*, trnL-CAA(2), trnL-UAA, trnL-UAA*, trnL-UAG, trnM-CAU, trnN-GUU(2), trnP-UGG, trnQ-UUG, trnR-ACG(2), trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-UGU, trnV-GAC(2), trnV-UAC*, trnW-CCA, trnY-GUA, trnY-M-CAU</i>	38
Other genes	Maturase	<i>matK</i>	1
	Protease	<i>clpP**</i>	1
	Envelope membrane protein	<i>cemA</i>	1
	Acetyl-CoA carboxylase	<i>accD</i>	1
	c-type cytochrome synthesis gene	<i>ccsA</i>	1
	Translation initiation factor	<i>infA</i>	1

Notes: \*Gene contains one intron. \*\*Gene contains two introns. #: Pseudo gene. (x2) indicates the number of the repeat unit is 2.

Category of Genes	Group of Genes	Name of Genes	Number of genes
Genes of unknown function	Conserved hypothetical chloroplast ORF	#ycf1, ycf1, ycf15(2), ycf2(2), ycf3**, ycf4	8
Notes: *Gene contains one intron. **Gene contains two introns. #: Pseudo gene. (×2) indicates the number of the repeat unit is 2.			

## 3.2. Comparative genome analysis

It indicates that information about the evolution of size differences in the IR, LSC, and SSC regions among Astereae species (Table 1). It appears that the analysis of 6 Astereae species has revealed some divergence in the boundaries of their IR regions (Fig. 2).

Analysis of the IR boundaries of 6 Astereae species revealed some divergence. It appears that the comparison of the border junctions in the cp genomes has revealed some variations between two genes based on their positions. The genes such as *ycf1*, *rps19* were observed. The *rps19* gene is located at the LSC-IRb border regions, while they varied in sizes (17bp, 60bp); The *ycf1* gene spans the SSC-IRa junction. The *rpl12* gene is entirely in the IR region, 115bp away from the LSC; However, a more interesting phenomenon is that the pseudogene *ycf1* of *D. benthamii* spans the boundary between the SSC and IRa regions (6bp).

The mVISTA-based identity plot revealed the sequences shared a high degree of similarity. But there were a few variants. These variations are usually observed in the intergenic spacers (IGS) instead of coding-regions, which implies that coding regions are more conserved than non-coding regions (Fig. 3).

## 3.3. Sequence repeats

In this study, it is noted that there were 532 SSRs in 6 species of Astereae and the number of their SSRs is very similar (84–100) (Fig. 4A). Among them, the number of SSRs is the highest in *Aster tataricus* L.f. (100) and the lowest in *Aster ageratoides* Turcz. and *Heteropappus gouldii* (C.E.C.Fisch.) Grierson (84). In addition, 6 types of SSRs are detected, including mononucleotides, dinucleotides, trinucleotides, tetranucleotides, pentanucleotides, and hexanucleotides (Fig. 5). The most mononucleotide repeats (38.16%) are detected, followed by trinucleotide repeats (20.49%), dinucleotide repeats (17.86%), tetranucleotide repeats (16.17%), and pentanucleotide repeats (6.02%) and hexanucleotide repeats (1.32%). It is noteworthy that hexanucleotide repeats are not found in the chloroplast genome of *A. tataricus*. It indicated that a total of 278 tandem repeats (38–50 of each species) were identified using REPuter, with varying numbers of repeats of different types: 133 palindromic (P), 113 forward (F), 24 reverse (R), and 8 complement (C) (Fig. 4B). In Astereae, most of the repeats consisted of 30-34bp (59.00%) and 40-44bp (14.79%), 35-39bp (12.95%), followed by ≥ 50bp (7.55%) and 45-49bp (5.76%).

## 3.4. Codon usage

By calculating the RSCU values for all protein-coding genes in the chloroplast genomes of *D. benthamii*, a total of 30,600 codons were identified. 31 codons exhibited greater preference (RSCU > 1) (Fig. 6). Serine displayed no preferences (RSCU = 1), and the remaining codons were used least preferred. It's also noted that there were no extremely rare codons (RSCU < 0.1). Among the 20 amino acids, leucine (10.67%) accounts for the largest proportion and cysteine (1.12%) accounts for the smallest.

## 3.5. Phylogenetic analysis

27 complete chloroplast sequences from Asteraceae and two plastomes of outgroup (*Ajania pacifica*, *Achillea millefolium*) were used to infer the evolutionary relationships. The Maximum Likelihood (ML) tree (Fig. 7) resulting from the phylogenomic

analysis had many highly supported nodes (Brouillet et al. 2009). The phylogenomic analysis was used to identify several clades based on the chloroplast genome data. The tribe Astereae forms a monophyletic group. The Chinese genus *Nannoglottis* (African clade) is sister to all other Astereae, followed by *Llerasia caucana* (S.F.Blake) Cuatrec. (Paleo South American clade). It is sister to the remaining Astereae. The following clades are successively recovered along the spine of the maximum likelihood tree: 1) the Australasian clade, *D. benthamii* belongs to African-S Asian subtribe Grangeinae. *D. benthamii* is sister to the clade including *Aster ageratoides*, *Aster pekinensis* (Hance) Merr., *Aster tataricus*, *Heteropappus gouldii*, *Heteroplexis sericophylla* Y.L.Chen (BP = 100%). 2) North American clade. The subtribe Conyzinae includes *Erigeron annuus* (L.) Pers., *Erigeron canadensis* Ten. and *Erigeron breviscapus* (Vaniot) Hand. - Mazz. Conyzinae is sister to *Pityopsis falcata* (Pursh) Nutt., *Solidago decurrens* Lour. and *Symphotrichum subulatum* (Michx.) G.L.Nesom (BP = 100%); 3) South American clade, a clade (BP = 100%) comprises the remaining Astereae. The first clade includes *Diplostephium alveolatum* Cuatrec., *Diplostephium antioquense* Cuatrec., *Linochilus alveolatus* Cuatrec., *Linochilus antioquensis* (Cuatrec.) Saldivia & O.M.Vargas (BP = 100%). The second clade is composed of *Floscaldasia hypsophila* Cuatrec., *Hinterhubera ericoides* Wedd., *Blakiella bartsiifolia* (S.F.Blake) Cuatrec., *Laestadia muscicola* Wedd. (BP = 54%), while the last clade is made of the rest of the species of section South American lineages (BP = 100%).

## 4. DISCUSSION

### 4.1. Comparison of the chloroplast genome of *D. benthamii*

The complete chloroplast genome of *D. benthamii* is analyzed based on second-generation high-throughput sequencing technology and bioinformatics analysis. The chloroplast genome of *D. benthamii* is found to be a typical circular DNA molecule (152,350 bp) long. In the present study, 134 unique genes are annotated, including 87 protein-coding genes, 38 tRNA genes and 8 rRNA genes. Genes of unknown function, including the *ycf1*, *ycf2*, *ycf3*, *ycf4* and *ycf15* genes found in most angiosperms, are present in the *D. benthamii* genome. It revealed that the GC content of the chloroplast genome was similar between *D. benthamii* and 5 Astereae species, both at 37.26%-37.36%, and both the LSC region (35.15%-35.30%) and the SSC region (31.20%-31.33%) have lower GC content than the IR region (42.98%-43.06%). In addition, all eight rRNAs are located in the IR region, resulting in the high GC content of IR, while the lower GC content of SSC might be related to the presence of NADH (Shen et al. 2017; Shimada et al. 1991).

### 4.2. Comparative analysis

The IR regions are known to be highly conserved in the genome of chloroplasts. During evolution, the expansion and contraction of the IR, LSC, and SSC regions are common, which leads to variability in genome length (Kim et al. 2004). In the present study, the LSC/IR and SSC/IR boundaries of the chloroplast genomes of *D. benthamii* and five other Astereae species were highly similar. The *rps19* gene spanned the LSC/IRb boundary of *D. benthamii* and *ycf1* gene spanned the IRb/SSC boundary. It is similar to that of the species of Astereae (*D. benthamii*, *A. ageratoides*, *A. pekinensis*, *A. tataricus*, *H. gouldii*, *H. sericophylla*) boundary genes. It suggests that the IR region of *D. benthamii* has a highly conserved boundary with LSC/IRa. However, the IRa side of the LSC/IRa boundary in *D. benthamii* does not have the *rps19* gene. The *rps19* gene is also found in other Asteraceae (Lopes et al. 2018).

The mVISTA-based identity plot is a tool used to analyse conservation and variation of DNA sequences and gene synteny across multiple genomes. Then, it was used to study 6 different plastomes and identify regions of increased genetic variation. The sequence variation in the IRs region was significantly smaller than that in the LSC and SSC regions. The coding regions are more conserved over evolution than the non-coding regions (intergenic regions and some intronic sequence), which is a common observation in genomic studies. Variation patterns showed significant differences in different regions. It appeared that in general, there are more variation observed in the LSC and SSC regions of the sequence compared to the IR region. The result is consistent with patterns of chloroplast genomes of chloroplast genomes in other Asteraceae species (Nie et al. 2012; Hatmaker et al. 2020; Ma et al. 2020; Loeuille et al. 2021). The high variation observed in the LSC and SSC regions was primarily attributed to non-coding sequences. These non-coding sequences were primarily intergenic regions, followed by intronic sequences of certain genes. It is generally observed that non-coding sequences in Astereae evolve faster than protein-

coding genes. This pattern is consistent with what has been observed in other Asteraceae chloroplast genome sequences (Tonti-Filippini et al. 2017; Enke et al. 2008; Liu et al. 2012).

### 4.3. Sequence repeats

SSRs in the chloroplast genome are known for their high levels of sequence polymorphism, abundance and uniparental inheritance, making them a valuable tool for studying the genetics of plant populations. By analysing variation in SSRs, researchers can gain insights into the genetic diversity, structure and evolutionary history of plant populations (Du et al. 2012; Pauwels et al. 2012; Provan et al. 2001). In this study, 36 mononucleotide, 19 dinucleotide, 12 trinucleotide, 16 tetranucleotide, 2 pentanucleotide and 2 hexanucleotide SSRs were detected (Table 3). 80 of these SSRs were composed of A/T, AT/AT, AAT/AAT, AAAT/ATTT, AATT/AATT, AACT/AGTTT, AATAT/ ATATT. The chloroplast genome of *D. benthamii* contains a number of A or T repeats. However, this result is not surprising because this AT preference pattern is widely reported in many plant plastids (Somaratne et al. 2019; Zavala-Paez et al. 2020). SSRs were detected in different regions of the plastids of *D. benthamii*. SSRs were mainly distributed in the LSC (82.76%) relative to the IR (6.90%) and SSC regions (10.34%). SSR and tandem repeats are useful genetic markers for studying the genetic structure, diversity, phylogeny, and differentiation of Astereae and other Dichrocephala species.

Table 3  
SSR in the cp genome of *Dichrocephala benthamii*.

Repeat type (number)	Repeat Sequence	Number of repetitions														total
		3	4	5	6	7	8	9	10	11	12	13	14	15	16	
Mononucleotides (36)	A/T	-	-	-	-	-	-	-	19	6	5	3		1	2	36
Dinucleotides (19)	AT/AT	-	-	13	5	1										19
trinucleotide (12)	AAG/CTT	-	1	1												2
	AAT/AAT	8	2													10
Tetranucleotide (16)	AAAT/ATTT	12														12
	AATC/ATTG	1														1
	AATT/AATT	2														2
	AGAT/ATCT	1														1
Pentanucleotide (2)	AACT/AGTTT	1														1
	AATAT/ATATT	1														1
Hexanucleotide (2)	AATCCT/AGGATT	2														2

### 4.4. Codon usage bias

Differences in codon bias can affect the speed and accuracy of protein translation, as well as protein folding, stability, and function (Quax et al. 2015). This study found that 85 protein-coding genes in *D. benthamii* genome are encoded by 64 codons. Three are termination codons (UAA, UGA, UAG). Among the 20 amino acids, leucine accounts for the largest proportion and cysteine accounts for the smallest (Table 4). In other angiosperm cp genomes, the reported leucine and cysteines are also the most and least abundant amino acids (Sharp et al. 1987). The frequency of the codon AUU encoding isoleucine is the highest, but that of the codon UGC encoding cysteine is the lowest (Table 4). Relative synonymous codon usage (RSCU) is a commonly used statistic to measure codon usage bias. It compares the frequency of usage of a specific codon to the expected frequency of that codon. A value of RSCU greater than 1 indicates that the codon is used more frequently than expected, while a value less than 1 indicates it is used less frequently than expected. 31 preferred (RSCU > 1) synonymous codons are detected, indicating



that these codons are preferentially used in coding amino acids. Additionally, we found that only the codons encoding Serine amino acids have no bias (RSCU = 1); however, other codons have an obvious bias in *D. benthamii* (Fig. 7). Intriguingly, except UUG, all preferentially used codons end with A/U. This result agrees with that observed in other species. It shows that the dominant codon contains more A or U in codon selection. The high proportion of A/U is the driving force of deviation (Claude et al. 2020; Peng et al. 2020).

Table 4  
Codon usage in the chloroplast genomes of *Dichrocephala benthamii*.

Amino acid	Symbol	Codon	Count	RSCU	Amino acid	Symbol	Codon	Count	RSCU
*	Ter	UAA	48	1.7778	M	Met	AUU	1	0.0128
*	Ter	UAG	19	0.7037	M	Met	CUG	0	0
*	Ter	UGA	14	0.5185	M	Met	GUG	0	0
A	Ala	GCA	359	1.1089	M	Met	UUG	0	0
A	Ala	GCC	211	0.6517	N	Asn	AAC	239	0.443
A	Ala	GCG	147	0.4541	N	Asn	AAU	840	1.557
A	Ala	GCU	578	1.7853	P	Pro	CCA	275	1.1329
C	Cys	UGC	73	0.5703	P	Pro	CCC	184	0.758
C	Cys	UGU	183	1.4297	P	Pro	CCG	142	0.585
D	Asp	GAC	180	0.3987	P	Pro	CCU	370	1.5242
D	Asp	GAU	723	1.6013	Q	Gln	CAA	648	1.5521
E	Glu	GAA	881	1.5073	Q	Gln	CAG	187	0.4479
E	Glu	GAG	288	0.4927	R	Arg	AGA	422	1.8522
F	Phe	UUC	421	0.6502	R	Arg	AGG	140	0.6145
F	Phe	UUU	874	1.3498	R	Arg	CGA	303	1.3299
G	Gly	GGA	599	1.5041	R	Arg	CGC	89	0.3906
G	Gly	GGC	188	0.4721	R	Arg	CGG	98	0.4301
G	Gly	GGG	286	0.7181	R	Arg	CGU	315	1.3826
G	Gly	GGU	520	1.3057	S	Ser	AGC	98	0.3488
H	His	CAC	135	0.5153	S	Ser	AGU	354	1.2598
H	His	CAU	389	1.4847	S	Ser	UCA	342	1.2171
I	Ile	AUA	607	0.9435	S	Ser	UCC	246	0.8754
I	Ile	AUC	358	0.5565	S	Ser	UCG	135	0.4804
I	Ile	AUU	965	1.5	S	Ser	UCU	511	1.8185
K	Lys	AAA	895	1.4992	T	Thr	ACA	356	1.2255
K	Lys	AAG	299	0.5008	T	Thr	ACC	221	0.7608
L	Leu	CUA	316	0.778	T	Thr	ACG	115	0.3959
L	Leu	CUC	155	0.3816	T	Thr	ACU	470	1.6179
L	Leu	CUG	150	0.3693	V	Val	GUA	479	1.5218
L	Leu	CUU	524	1.2901	V	Val	GUC	154	0.4893
L	Leu	UUA	793	1.9524	V	Val	GUG	166	0.5274
L	Leu	UUG	499	1.2286	V	Val	GUU	460	1.4615
M	Met	AUA	0	0	W	Trp	UGG	403	1

Amino acid	Symbol	Codon	Count	RSCU	Amino acid	Symbol	Codon	Count	RSCU
M	Met	AUC	1	0.0128	Y	Tyr	UAC	152	0.3543
M	Met	AUG	547	6.9745	Y	Tyr	UAU	706	1.6457
Ref									
Ai TM, Zhang SR, Yang XW, Du LJ, Yan ZY (2013) <i>Dichrocephala</i> In Chen YL, Yang XW, Yu LZ, Cui Z (Eds) medicinal flora of China. Vol 10. Beijing, pp 362–365.									
Allen GC, Flores-Vergara MA, Krasnyanski S, Kumar S, 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, Poczai P (2018) IRscope: an online program to visualize the junction sites of chloroplast genomes. Bioinformatics 34:3030–3031.									
Anderberg A (1991) Taxonomy and phylogeny of the tribe Plucheeae (Asteraceae). Opera Botanica 104: 1–195.									
Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA (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, 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, Semple JC (2009) Astereae. In: Funk, Susanna VA, Stuessy A, Bayer RT (eds.) Systematics, evolution and biogeography of the Compositae. IAPT, Vienna, pp 449–490.									
Brouillet L, Urbatsch L, Roberts RP (2004) <i>Tonestus kingii</i> and <i>T. aberrans</i> are related to eurybia and the Machaerantherinae (Asteraceae: Astereae) based on nrDNA (ITS and ETS) data: reinstatement of herrickia and a new genus. triniteurybia. Sida 21: 889–900.									
Cai ZQ, Penaflor C, Kuehl JV, Leebens-Mack J, Carlson JE, dePamphilis CW, Boore JL, Jansen RK (2006) Complete plastid genome sequences of <i>Drimys</i> , <i>Liriodendron</i> , and <i>Piper</i> : implications for the phylogenetic relationships of magnoliids. BMC Evol. Biol 6: 77.									
Chen XL, Zhou JG, Cui YX, Wang Y, Duan BZ, Yao H (2018) Identification of <i>Ligularia</i> Herbs Using the Complete Chloroplast Genome as a Super-Barcode. Front. Pharmacol. 9:695.									
Chen YL, Liu SW, Liu Y, Yang QE, Nordenstam B, Illarionova ID, Jeffrey C, Koyama H, Vincent L (2011) <i>Dichrocephala</i> . In: Chen Yilin, Luc Brouillet (eds) Flora of China, Vols 20–21. Beijing, pp 550–551.									
Claude SJ, Park SJ (2020) <i>Aster spathulifolius</i> Maxim. a leaf transcriptome provides an overall functional characterization, discovery of SSR marker and phylogeny analysis. Plos One 15.									
Do HDK, Jung J, Hyun J, Yoon SJ, Lim C, Park K, Kim JH (2019) The newly developed single nucleotide polymorphism (SNP) markers for a potentially medicinal plant, <i>Crepidiastrum denticulatum</i> (Asteraceae), inferred from complete chloroplast genome data. Mol. Biol. Rep 46:3287–3297.									
Du QZ, Wang BW, Wei ZZ, Zhang DQ, Li BL (2012) Genetic Diversity and Population Structure of Chinese White Poplar ( <i>Populus tomentosa</i> ) Revealed by SSR Markers. J. Hered 103:853–862.									
Duan N, Deng LL, Zhang Y, Shi YC, Liu BB (2022) Comparative and phylogenetic analysis based on chloroplast genome of <i>Heteroplexis</i> (Compositae), a protected rare genus. BMC Plant Biology 22:1–10.									
Enke N, Gemeinholzer B (2008) Babcock revisited: new insights into generic delimitation and character evolution in <i>Crepis</i> L. (Compositae: Cichorieae) from ITS and <i>matK</i> sequence data. Taxon 57:756–768.									
Fiz O, Valcárcel V, Vargas P (2002) Phylogenetic position of Mediterranean Astereae and character evolution of daisies ( <i>Bellis</i> , Asteraceae) inferred from nrDNA ITS sequences. Mol. Phylogenet. Evol. 25:157–171.									
Fu ZX, Jiao BH, Nie B, Zhang GJ, Gao TG, China Phylogeny C (2016) A comprehensive generic-level phylogeny of the sunflower family: Implications for the systematics of Chinese Asteraceae. J Syst Evol 54:416–437. DOI 10.1111/jse.12216									
Funk VA, Anderberg AA, Baldwin BG, Bayer RJ, Bonifacino JM, Breitwieser I, Brouillet L, et al. (2009) Compositae metatrees: the next generation. In: Funk VA, Susanna A, Stuessy T, Bayer R (eds) Systematics, Evolution, and Biogeography of Compositae. IAPT, Vienna, pp 747–777.									

Amino acid	Symbol	Codon	Count	RSCU	Amino acid	Symbol	Codon	Count	RSCU
Gao T, Yao H, Song JY, Zhu YJ, Liu C, Chen SL (2010) Evaluating the feasibility of using candidate DNA barcodes in discriminating species of the large Asteraceae family. BMC Evol. Biol 10.									
Gitzendanner MA, Soltis PS, Wong GKS, Ruhfel BR, Soltis DE (2018) Plastid phylogenomic analysis of green plants: A billion years of evolutionary history. Am. J. Bot 105:291–301.									
Greiner S, Lehwark P, Bock R (2019) OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Research 47: W59-W64.									
Hatmaker EA, Wadl PA, Rinehart TA, Carroll J, Lane TS, Trigiano RN, Staton ME, Schilling EE (2020) Complete chloroplast genome comparisons for <i>Pityopsis</i> (Asteraceae). Plos One 15: e0241391.									
Hu XJ (2020) Characterization of the complete chloroplast genome of <i>Aster subulatus</i> Michx. Mitochondrial DNA Part B 5:597–598.									
Jansen RK, Cai Z, Raubeson LA, Daniell H, Depamphilis CW, Leebens-Mack J, Muller KF, Guisinger-Bellian M, Haberle RC, Hansen AK, Chumley TW, Lee SB, Peery R, McNeal JR, Kuehl JV, Boore JL (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. 104:19369–19374.									
Kang SH, Lee HO, Shin MJ, Kim NN, Choi BS, Kumar M, Ali A, Lee SC, Kim CK (2018) The complete chloroplast genome sequence of <i>Coix lacryma-jobi</i> L. (Poaceae), a cereal and medicinal crop. Mitochondrial DNA Part B 3:982–983.									
Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A (2012) Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649.									
Kim KJ, Lee HL (2004) Complete chloroplast genome sequences from Korean ginseng ( <i>Panax schinseng</i> Nees) and comparative analysis of sequence evolution among 17 vascular plants. DNA Res 11:247–261.									
Krawczyk K, Nobis M, Myszczyński K, Klichowska E, Sawicki J (2018) Plastid super-barcodes as a tool for species discrimination in feather grasses (Poaceae: <i>Stipa</i> ). Sci. Rep 8:1924.									
Kumar S, Hahn FM, McMahan CM, Cornish K, Whalen MC (2009) Comparative analysis of the complete sequence of the plastid genome of <i>Parthenium argentatum</i> and identification of DNA barcodes to differentiate <i>Parthenium</i> species and lines. BMC Plant Biol 9:131.									
Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R (2001) REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Research 29:4633–4642.									
Leseberg CH, Duvall MR (2009) The complete chloroplast genome of coix lacryma-jobi and a comparative molecular evolutionary analysis of plastomes in cereals. J. Mol. Evol. 69: 311–318.									
Li WP, Yang FS, Jivkova T, Yin GS (2012) Phylogenetic relationships and generic delimitation of Eurasian Aster (Asteraceae: Astereae) inferred from ITS, ETS and <i>trnL-F</i> sequence data. Annals of Botany 109:1341–1357.									
Liu H, Trusty J, Oviedo R, Anderberg A, Francisco-Ortega J (2004) Molecular phylogenetics of the Caribbean genera <i>Rhodogeron</i> and <i>Sachsia</i> (Asteraceae). Int. J. Plant Sci. 165:209–217. DOI 10.1086/380746									
Liu J, Qi ZC, Zhao YP, Fu CX, Xiang QY (2012) Complete cpDNA genome sequence of Smilax china and phylogenetic placement of <i>Liliales</i> - Influences of gene partitions and taxon sampling. Mol. Phylogenet. Evol. 64:545–562.									
Liu X, Zhou BY, Yang HY, Li Y, Yang Q, Lu YZ, Gao Y (2018) Sequencing and Analysis of <i>Chrysanthemum carinatum</i> Schousb and <i>Kalimeris indica</i> . The Complete Chloroplast Genomes Reveal Two Inversions and <i>rbcl</i> as Barcoding of the Vegetable. Molecules 23:1358.									
Liu XF, Han MY, Chen J, Zhang X, Chen ZY, Tang Y, Qu TM, Huang CP, Yu SH, Fu ZX (2022) Characterization of the complete chloroplast genome sequence of Chinese endemic species of <i>Nouelia insignis</i> (Hyalideae, Asteraceae) and its phylogenetic implications. Mitochondrial DNA Part B 7:600–602.									
Loeuille B, Thode V, Siniscalchi C, Andrade S, Rossi M, 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.									
Lopes AD, Pacheco TG, Nimz T, Vieira LD, Guerra MP, Nodari RO, de Souza EM, Pedrosa FD, Rogalski M (2018) The complete plastome of macaw palm <i>Acrocomia aculeata</i> (Jacq.) Lodd. ex Mart. and extensive molecular analyses of the evolution of plastid genes in Arecaceae. Planta 247:1011–1030.									

Amino acid	Symbol	Codon	Count	RSCU	Amino acid	Symbol	Codon	Count	RSCU
Ma YP, Zhao L, Zhang WJ, Zhang YH, Xing X, Duan XX, Hu J, Harris A, Liu PL, Dai SL (2020) Origins of cultivars of <i>Chrysanthemum</i> —Evidence from the chloroplast genome and nuclear <i>LFY</i> gene. <i>J Syst Evol</i> 58:925–944.									
Maier RM, Neckermann K, Igloi GL, Kössel H (1995) Complete sequence of the maize chloroplast genome: gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. <i>J. Mol. Biol.</i> 251:614–628.									
Miller MA, Pfeiffer WT, Schwartz T (2010) Creating the CIPRES Science Gateway for Inference of Large Phylogenetic Trees. In Proceedings of the Gateway Computing Environments Workshop (GCE): 1–8.									
Morgan DR, Korn RL, Muggleston SL (2009) Insights into reticulate evolution in machaerantherinae (Asteraceae: Astereae): 5s ribosomal RNA spacer variation, estimating support for incongruence, and constructing reticulate phylogenies. <i>Am. J. Bot.</i> 96:920–932.									
Nie XJ, Lv SZ, Zhang YX, Du XH, Wang L, Biradar SS, Tan XF, Wan FH, Song WN (2012) Complete Chloroplast Genome Sequence of a Major Invasive Species, Crofton Weed ( <i>Ageratina adenophora</i> ). <i>Plos One</i> 7.									
Noyes RD (2000) Biogeographical and evolutionary insights on <i>Erigeron</i> and allies (Asteraceae) from ITS sequence data. <i>Plant Syst. Evol</i> 220:93–114									
Noyes RD, Rieseberg LH (1999) ITS sequence data support a single origin for North American Astereae (Asteraceae) and reflect deep geographic divisions in <i>Aster</i> s.l. <i>Am. J. Bot</i> 86(3): 398–412.									
Ou CZ, Feng YL, Hu YK, Tian XY, Fu ZX (2019) Characterization of the complete chloroplast genome sequence of <i>Aster hersileoides</i> (Asteraceae, Astereae) and its phylogenetic implications. <i>Mitochondrial DNA Part B</i> 4:985–986.									
Palazzesi L, Pellicer J, Barreda VD, Loeuille B, Mandel JR, Pokorny L, Siniscalchi CM, Telleria MC, Leitch IJ, Hidalgo O (2022) Asteraceae as a model system for evolutionary studies: from fossils to genomes. <i>Bot. J. Linn. Soc</i> 200:143–164.									
Panero JL, Funk VA (2008) The value of sampling anomalous taxa in phylogenetic studies: Major clades of the Asteraceae revealed. <i>Mol Phylogenet Evol</i> 47:757–782.									
Park J, Shim J, Won H, Lee J (2017) Plastid genome of <i>Aster altaicus</i> var. <i>uchiyamae</i> Kitam., an endanger species of Korean asterids. <i>Journal of Species Research</i> 2017 6: 76–90.									
Parks M, Cronn R, Liston A (2009) Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. <i>BMC Biol</i> 7.									
Pauwels M, Vekemans X, Gode C, Frerot H, Castric V, Saumitou-Laprade P (2012) Nuclear and chloroplast DNA phylogeography reveals vicariance among European populations of the model species for the study of metal tolerance, <i>Arabidopsis halleri</i> (Brassicaceae). <i>New Phytol</i> 193:916–928.									
Peng YJ, Cui XY, Tan MC, Hu L, Ruan HY, Shao YY, Song EL, Tang YJ (2020) Characterization of the complete chloroplast genome of <i>Chlorophytum comosum</i> (Liliaceae). <i>Mitochondrial DNA Part B</i> 5:1046–1047.									
Provan J, Pm. H, Powell W (2001) Chloroplast microsatellites: new tools for studies in plant ecology and evolution [Review]. <i>Trends Ecol. Evol</i> 16: 142–147.									
Qu XJ, Moore MJ, Li DZ, Yi TS (2019) PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. <i>Plant Methods</i> 15: 1–12.									
Quax TEF, Claassens NJ, Soll D, van der Oost J (2015) Codon Bias as a Means to Fine-Tune Gene Expression. <i>Mol. Cell</i> 59:149–161.									
Salih RHM, Majesky L, Schwarzacher T, Gornall R, Heslop-Harrison P (2017) Complete chloroplast genomes from apomictic <i>Taraxacum</i> (Asteraceae): Identity and variation between three microspecies. <i>Plos One</i> 12: e0168008.									
Saslis-Lagoudakis CH, Savolainen V, Williamson EM, Forest F, Wagstaff SJ, Baral SR, Watson MF, Pendry CA, Hawkins JA (2012) Phylogenies reveal predictive power of traditional medicine in bioprospecting. <i>Proc. Natl. Acad. Sci. U.S.A.</i> 109:15835–15840.									
Shahzadi I, Abdullah, Mehmood F, Ali Z, Ahmed I, Mirza B (2020) Chloroplast genome sequences of <i>Artemisia maritima</i> and <i>Artemisia absinthium</i> : Comparative analyses, mutational hotspots in genus <i>Artemisia</i> and phylogeny in family Asteraceae. <i>Genomics</i> 112:1454–1463.									
Sharp PM, Li WH (1987) The codon Adaptation Index—a measure of directional synonymous codon usage bias, and its potential applications. <i>Nucl Acids Res</i> 15:1281–1295									

Amino acid	Symbol	Codon	Count	RSCU	Amino acid	Symbol	Codon	Count	RSCU
Shaw J, Lickey EB, Schilling EE, Small RL (2007) Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: The tortoise and the hare III. <i>Am. J. Bot</i> 94:275–288.									
Shen XF, Wu ML, Liao BS, Liu ZX, Bai R, Xiao SM, Li XW, Zhang BL, Xu J, Chen SL (2017) Complete Chloroplast Genome Sequence and Phylogenetic Analysis of the Medicinal Plant <i>Artemisia annua</i> . <i>Molecules</i> 22.									
Shimada H, Sugiura M (1991) Fine structural features of the chloroplast genome: comparison of the sequenced chloroplast genomes. <i>Nucleic acids research</i> 19:983–995									
Smith DR (2015) Mutation Rates in Plastid Genomes: They Are Lower than You Might Think. <i>Genome Biol. Evol</i> 7:1227–1234.									
Somaratne Y, Guan DL, Wang WQ, Zhao L, Xu SQ (2019) Complete chloroplast genome sequence of <i>Xanthium sibiricum</i> provides useful DNA barcodes for future species identification and phylogeny. <i>Plant Syst. Evol</i> 305:949–960.									
Song B, Si JG, Yu M, Tian XH, Ding G, Zou ZM (2017) Megastigmane glucosides isolated from <i>Dichrocephala benthamii</i> . <i>Chin. J. Nat. Med.</i> 15:288–291.									
Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. <i>Bioinformatics</i> 30:1312–1313.									
Tian XH, Ding G, Peng CZ, Hu YB, Li L, Chen H, Zou ZM (2013a) Sinapyl Alcohol Derivatives from the Lipo-soluble Part of <i>Dichrocephala benthamii</i> C. B. Clarke. <i>Molecules</i> 18:1720–1727.									
Tian XH, Li L, Hu YB, Zhang HW, Liu YT, Chen H, Ding G, Zou ZM (2013b) Dichrocephones A and B, two cytotoxic sesquiterpenoids with the unique 3.3.3 propellane nucleus skeleton from <i>Dichrocephala benthamii</i> . <i>Rsc Advances</i> 3:7880–7883.									
Tonti-Filippini J, Nevill PG, Dixon K, Small I (2017) What can we do with 1000 plastid genomes? <i>Plant J</i> 90:808–818.									
Vargas OM, Ortiz EM, Simpson BB (2017) Conflicting phylogenomic signals reveal a pattern of reticulate evolution in a recent high-Andean diversification (Asteraceae: Astereae: <i>Diplostephium</i> ). <i>New Phytol</i> 214:1736–1750.									
Wang Q, Wang XH, Yang XC, Fu LF, Gong QJ, Fu ZX (2019) The complete chloroplast genome of <i>Aster hypoleucus</i> (Asteraceae: Astereae): an endemic species from China. <i>Mitochondrial DNA Part B</i> 4:2647–2648.									
Wang, M.C.; Li, R.; Zhao, Q. Multi-Omics Techniques for Forest Plant Breeding. Preprints.org 2023, 2023040549. <a href="https://doi.org/10.20944/preprints202304.0549.v1">https://doi.org/10.20944/preprints202304.0549.v1</a> .									
Wick RR, Schultz MB, Justin Z, Holt KE Bandage: interactive visualization of de novo genome assemblies. <i>Bioinformatics</i> 31:3350–3352									
Wicke S, Schneeweiss GM, dePamphilis CW, Muller KF, Quandt D (2011) The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. <i>Plant Mol. Biol</i> 76:273–297.									
Xia Y, Hu ZG, Li XW, Wang P, Zhang XQ, Li Q, Lu CL (2016) The complete chloroplast genome sequence of <i>Chrysanthemum indicum</i> . <i>Mitochondrial DNA Part A</i> 27:4668–4669.									
Yu SH, Yang XC, Tian XY, Liu XF, Lu X, Huang CP, Fu ZX (2022) The complete chloroplast genome sequence of the monotypic and enigmatic genus <i>Cavea</i> (tribe Gymnarrheneae) and a comparison with other species in Asteraceae. <i>J. Genet</i> 101. 20.									
Zavala-Paez M, Vieira LD, de Baura VA, Balsanelli E, de Souza EM, Cevallos MC, Chase MW, Smidt ED (2020) Comparative Plastid Genomics of Neotropical <i>Bulbophyllum</i> (Orchidaceae; Epidendroideae). <i>Front Plant Sci</i> 11.									
Zhang X, Deng T, Moore MJ, Ji YH, Lin N, Zhang HJ, Meng AP, Wang HC, Sun YX, Sun H (2019) Plastome phylogenomics of <i>Saussurea</i> (Asteraceae: Cardueae). <i>BMC Plant Biol</i> 19:290.									
Zhang X, Jiang PP, Fan SJ (2021) Characterization of the complete plastome of <i>Aster pekinensis</i> (Asteraceae), a perennial herb. <i>Mitochondrial DNA Part B</i> 6:1064–1065.									
Zhang ZY, Jiang XB, Chen Y, Zhu P, Li L, Zeng YL, Tang TM (2019) Characterization of the complete chloroplast genome sequence of <i>Conyza canadensis</i> and its phylogenetic implications. <i>Mitochondrial DNA Part B</i> 4:2028–2030.									
Zheng TR, Liu MY, Wang T, Chen H (2017) The complete chloroplast genome sequence of medicinal plant <i>Conyza blinii</i> H. Lev. <i>Mitochondrial DNA Part B</i> 2:50–51.									

Amino acid	Symbol	Codon	Count	RSCU	Amino acid	Symbol	Codon	Count	RSCU
Zhong Q, Yang S, Sun X, Wang L, Li Y (2019) The complete chloroplast genome of the Jerusalem artichoke ( <i>Helianthus tuberosus</i> L.) and an adaptive evolutionary analysis of the <i>ycf2</i> gene. Peerj 7: e7596.									
Zhou JY, Li JJ, Peng SB, An XM (2022) Characterization of the complete chloroplast genome of the invasive plant <i>Erigeron annuus</i> (L.) Pers. (Asterales: Asteraceae). Mitochondrial DNA Part B 7:188–190.									

## 4.5. Phylogenomic analysis

It is interesting that the species of Astereae are traditional Chinese herbal medicines. However, the species identification and phylogeny of this tribe have been more controversial due to abundant species and intricate interspecific variation. It has been shown that chloroplast genomes are often used as DNA barcodes for species identification and phylogenetic relationship studies. To further define the phylogenetic position of *D. benthamii* in Astereae. To determine the phylogenetic position and evolutionary relationships in Asteraceae, the complete cp genome sequences of 29 reported Astereae were selected to construct a maximum likelihood (ML) phylogenetic tree, and two species of tribe Anthemideae (*Achillea millefolium*, *Ajania pacifica*) were considered as outgroups.

The genus of *Nannoglottis* is one of the earliest diverging lineages of Asteraceae, suggesting that the tribe originated in Africa and then underwent a long dispersal or drift to eastern Asia (Liu et al. 2012). Our phylogeny has led to the similar conclusion. *D. benthamii* belongs to the Grangeinae, the latest diverging African lineage, which is widely distributed in Africa and southern Asia. This suggests that the tribe Astereae spread from eastern Asia to Indo-China Peninsula and migrated from high to low altitudes, vialed the Qinghai-Tibet Plateau (Brouillet et al. 2009; Anderberg et al. 1991). We confirmed that Astereae emerged in southern Africa and dispersed early to other continents.

## 5. Conclusion

This study firstly reported the CP genomes from *D. benthamii*. Compared to the cp genomes of five related Astereae species, the genome structure and composition of *D. benthamii* are similar. Different variable regions and SSR markers were unmasked. Also, the gene content arrangements and order were very much conserved. Furthermore, 87 SSRs were found, which can be used for studies on population genetics and genetic breeding of *Dichrocephala*. Finally, a phylogenetic tree was constructed based on the complete chloroplast genomes of 27 Astereae species and two Anthemideae species, confirming that *D. benthamii* had the closest relationship with Baccharidinae and Asterinae from the molecular viewpoint and revealed the position of *D. benthamii* in Astereae. It was found that *D. benthamii* is at the base of the phylogenetic tree. It also confirms that the tribe Astereae underwent a long dispersal or drift to eastern Asia to the Indo-China peninsula and migrated from high to low altitudes. Thus, the complete cp genome of *D. benthamii* provides valuable genetic information for this genus and lays a foundation for identifying and studying population evolution in Astereae species.

## Declarations

**Author Contributions:** B.L. and Z.F. conceived and designed the research. B.L. and Z.F. performed bioinformatic analyses. X.L., T.L. and J.L. carried out wetlab experiments. T.X. contributed to data interpretation. H.C., B.L. and Z.F. wrote the manuscript. D.W. and Z.F. re-vised the manuscript. All authors read and approved the manuscript. All authors have read and agreed to the published version of the manuscript.

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

**Institutional Review Board Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

**Acknowledgments:** The authors would like to thank the editor and anonymous reviewers for the constructive criticism of the original manuscript.

## References

1. Ai TM, Zhang SR, Yang XW, Du LJ, Yan ZY (2013) *Dichrocephala* In Chen YL, Yang XW, Yu LZ, Cui Z (Eds) medicinal flora of China. Vol 10. Beijing, pp 362-365.
2. Allen GC, Flores-Vergara MA, Krasnyanski S, Kumar S, Thompson WF (2006) A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. *Nature Protocols* 1:2320-2325.
3. Amiryousefi A, Hyvonen J, Poczai P (2018) IRscope: an online program to visualize the junction sites of chloroplast genomes. *Bioinformatics* 34:3030-3031.
4. Anderberg A (1991) Taxonomy and phylogeny of the tribe Plucheeae (Asteraceae). *Opera Botanica* 104: 1–195.
5. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyskin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA (2012) SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J.comput.Biol* 19:455-477.
6. Beier S, Thiel T, Munch T, Scholz U, Mascher M (2017) MISA-web: a web server for microsatellite prediction. *Bioinformatics* 33:2583-2585.
7. Brouillet L, Lowrey TK, Urbatsch L, Karaman-Castro V, Sancho G, Wagstaff S, Semple JC (2009) Astereae. In: Funk, Susanna VA, Stuessy A, Bayer RT (eds.) *Systematics, evolution and biogeography of the Compositae*. IAPT, Vienna, pp 449-490.
8. Brouillet L, Urbatsch L, Roberts RP (2004) *Tonestus kingii* and *T. Aberrans* are related to eurybia and the Machaerantherinae (Asteraceae: Astereae) based on nrDNA (ITS and ETS) data: reinstatement of herrickia and a new genus. *triniteurybia*. *Sida* 21: 889-900.
9. Cai ZQ, Penafior C, Kuehl JV, Leebens-Mack J, Carlson JE, dePamphilis CW, Boore JL, Jansen RK (2006) Complete plastid genome sequences of *Drimys*, *Liriodendron*, and *Piper*: implications for the phylogenetic relationships of magnoliids. *BMC Evol. Biol* 6: 77.
10. Chen XL, Zhou JG, Cui YX, Wang Y, Duan BZ, Yao H (2018) Identification of *Ligularia* Herbs Using the Complete Chloroplast Genome as a Super-Barcode. *Front. Pharmacol.* 9:695.
11. Chen YL, Liu SW, Liu Y, Yang QE, Nordenstam B, Illarionova ID, Jeffrey C, Koyama H, Vincent L (2011) *Dichrocephala*. In: Chen Yilin, Luc Brouillet (eds) *Flora of China*, Vols 20-21. Beijing, pp 550–551.
12. Claude SJ, Park SJ (2020) *Aster spathulifolius* Maxim. a leaf transcriptome provides an overall functional characterization, discovery of SSR marker and phylogeny analysis. *Plos One* 15.
13. Do HDK, Jung J, Hyun J, Yoon SJ, Lim C, Park K, Kim JH (2019) The newly developed single nucleotide polymorphism (SNP) markers for a potentially medicinal plant, *Crepidiastrum denticulatum* (Asteraceae), inferred from complete chloroplast genome data. *Mol. Biol. Rep* 46:3287-3297.
14. Du QZ, Wang BW, Wei ZZ, Zhang DQ, Li BL (2012) Genetic Diversity and Population Structure of Chinese White Poplar (*Populus tomentosa*) Revealed by SSR Markers. *J. Hered* 103:853-862.
15. Duan N, Deng LL, Zhang Y, Shi YC, Liu BB (2022) Comparative and phylogenetic analysis based on chloroplast genome of *Heteroplexis* (Compositae), a protected rare genus. *Bmc Plant Biology* 22:1-10.
16. Enke N, Gemeinholzer B (2008) Babcock revisited: new insights into generic delimitation and character evolution in *Crepis* L. (Compositae: Cichorieae) from ITS and *matK* sequence data. *Taxon* 57:756-768.
17. Fiz O, Valcárcel V, Vargas P (2002) Phylogenetic position of Mediterranean Astereae and character evolution of daisies (*Bellis*, Asteraceae) inferred from nrDNA ITS sequences. *Mol. Phylogenet. Evol.* 25:157–171.
18. Fu ZX, Jiao BH, Nie B, Zhang GJ, Gao TG, China Phylogeny C (2016) A comprehensive generic-level phylogeny of the sunflower family: Implications for the systematics of Chinese Asteraceae. *J Syst Evol* 54:416-437. DOI 10.1111/jse.12216



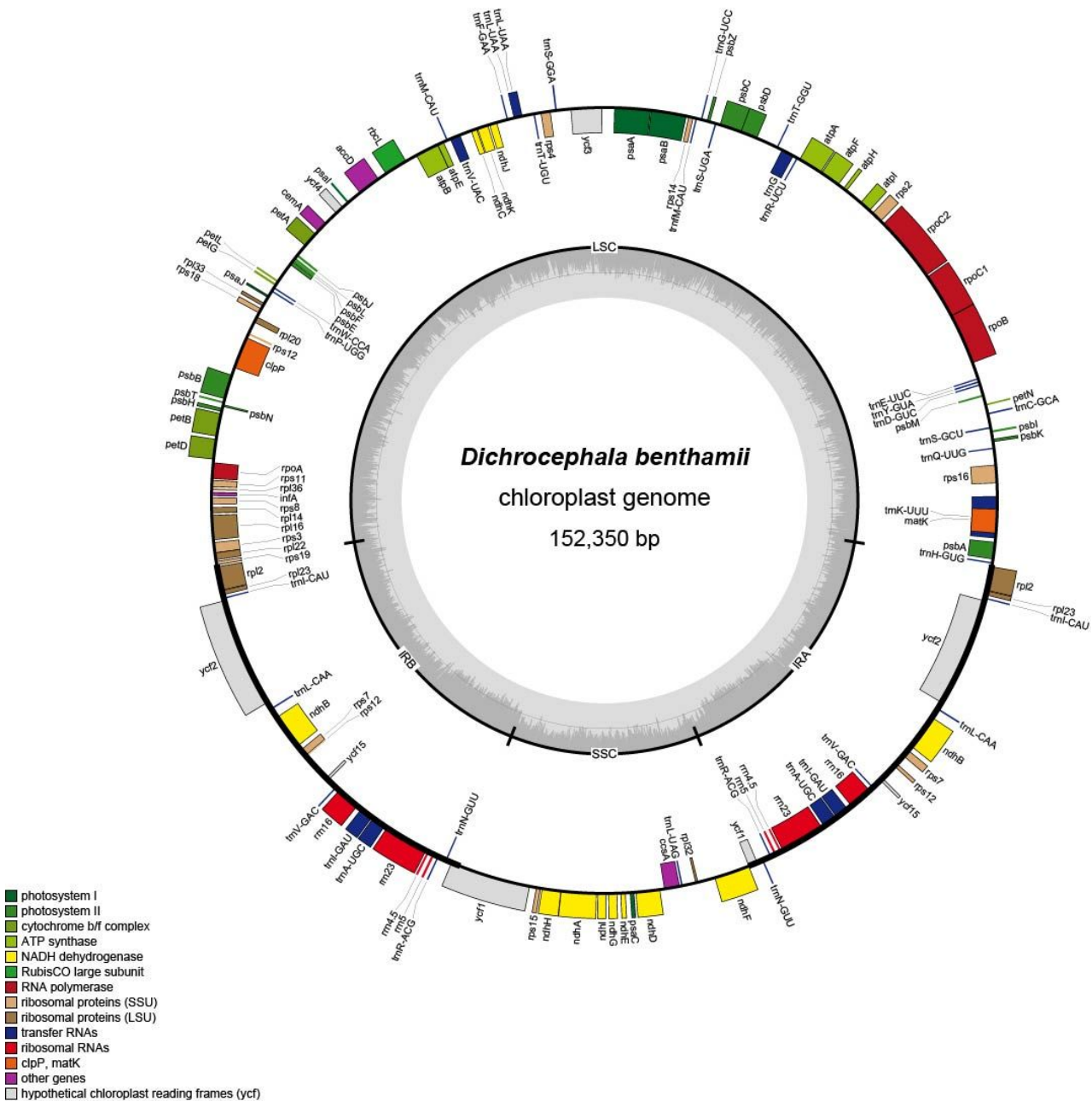
19. Funk VA, Anderberg AA, Baldwin BG, Bayer RJ, Bonifacino JM, Breitwieser I, Brouillet L, et al. (2009) Compositae metatrees: the next generation. In: Funk VA, Susanna A, Stuessy T, Bayer R (eds) Systematics, Evolution, and Biogeography of Compositae. IAPT, Vienna, pp 747–777.
20. Gao T, Yao H, Song JY, Zhu YJ, Liu C, Chen SL (2010) Evaluating the feasibility of using candidate DNA barcodes in discriminating species of the large Asteraceae family. BMC Evol. Biol 10.
21. Gitzendanner MA, Soltis PS, Wong GKS, Ruhfel BR, Soltis DE (2018) Plastid phylogenomic analysis of green plants: A billion years of evolutionary history. Am. J. Bot 105:291-301.
22. Greiner S, Lehwark P, Bock R (2019) OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Research 47: W59-W64.
23. Hatmaker EA, Wadl PA, Rinehart TA, Carroll J, Lane TS, Trigiano RN, Staton ME, Schilling EE (2020) Complete chloroplast genome comparisons for *Pityopsis* (Asteraceae). Plos One 15: e0241391.
24. Hu XJ (2020) Characterization of the complete chloroplast genome of *Aster subulatus* Michx. Mitochondrial DNA Part B 5:597-598.
25. Jansen RK, Cai Z, Raubeson LA, Daniell H, Depamphilis CW, Leebens-Mack J, Muller KF, Guisinger-Bellian M, Haberle RC, Hansen AK, Chumley TW, Lee SB, Peery R, McNeal JR, Kuehl JV, Boore JL (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. 104:19369-19374.
26. Kang SH, Lee HO, Shin MJ, Kim NN, Choi BS, Kumar M, Ali A, Lee SC, Kim CK (2018) The complete chloroplast genome sequence of *Coix lacryma-jobi* L. (Poaceae), a cereal and medicinal crop. Mitochondrial DNA Part B 3:982-983.
27. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A (2012) Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647-1649.
28. Kim KJ, 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.
29. Krawczyk K, Nobis M, Myszczyński K, Klichowska E, Sawicki J (2018) Plastid super-barcodes as a tool for species discrimination in feather grasses (Poaceae: *Stipa*). Sci. Rep 8:1924.
30. Kumar S, Hahn FM, McMahan CM, Cornish K, Whalen MC (2009) Comparative analysis of the complete sequence of the plastid genome of *Parthenium argentatum* and identification of DNA barcodes to differentiate *Parthenium* species and lines. BMC Plant Biol 9:131.
31. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R (2001) REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Research 29:4633-4642.
32. Leseberg CH, Duvall MR (2009) The complete chloroplast genome of coix lacryma-jobi and a comparative molecular evolutionary analysis of plastomes in cereals. J. Mol. Evol. 69: 311–318.
33. Li WP, Yang FS, Jivkova T, Yin GS (2012) Phylogenetic relationships and generic delimitation of Eurasian Aster (Asteraceae: Astereae) inferred from ITS, ETS and *trnL-F* sequence data. Annals of Botany 109:1341-1357.
34. Liu H, Trusty J, Oviedo R, Anderberg A, Francisco-Ortega J (2004) Molecular phylogenetics of the Caribbean genera *Rhodogeron* and *Sachsia* (Asteraceae). Int. J. Plant Sci. 165:209-217. DOI 10.1086/380746
35. Liu J, Qi ZC, Zhao YP, Fu CX, Xiang QY (2012) Complete cpDNA genome sequence of *Smilax china* and phylogenetic placement of *Liliales* - Influences of gene partitions and taxon sampling. Mol. Phylogenet. Evol. 64:545-562.
36. Liu X, Zhou BY, Yang HY, Li Y, Yang Q, Lu YZ, Gao Y (2018) Sequencing and Analysis of *Chrysanthemum carinatum* Schousb and *Kalimeris indica*. The Complete Chloroplast Genomes Reveal Two Inversions and *rbcl* as Barcoding of the Vegetable. Molecules 23:1358.
37. Liu XF, Han MY, Chen J, Zhang X, Chen ZY, Tang Y, Qu TM, Huang CP, Yu SH, Fu ZX (2022) Characterization of the complete chloroplast genome sequence of Chinese endemic species of *Nouelia insignis* (Hyalideae, Asteraceae) and its phylogenetic implications. Mitochondrial DNA Part B 7:600-602.

38. Loeuille B, Thode V, Siniscalchi C, Andrade S, Rossi M, 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.
39. Lopes AD, Pacheco TG, Nimz T, Vieira LD, Guerra MP, Nodari RO, de Souza EM, Pedrosa FD, Rogalski M (2018) The complete plastome of macaw palm *Acrocomia aculeata* (Jacq.) Lodd. ex Mart. and extensive molecular analyses of the evolution of plastid genes in Arecaceae. *Planta* 247:1011-1030.
40. Ma YP, Zhao L, Zhang WJ, Zhang YH, Xing X, Duan XX, Hu J, Harris A, Liu PL, Dai SL (2020) Origins of cultivars of *Chrysanthemum*—Evidence from the chloroplast genome and nuclear *LFY* gene. *J Syst Evol* 58:925-944.
41. Maier RM, Neckermann K, Igloi GL, Kössel H (1995) Complete sequence of the maize chloroplast genome: gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. *J. Mol. Biol.* 251:614-628.
42. Miller MA, Pfeiffer WT, Schwartz T (2010) Creating the CIPRES Science Gateway for Inference of Large Phylogenetic Trees. In *Proceedings of the Gateway Computing Environments Workshop (GCE)*: 1-8.
43. Morgan DR, Korn RL, Muggleston SL (2009) Insights into reticulate evolution in machaerantherinae (Asteraceae: Astereae): 5s ribosomal RNA spacer variation, estimating support for incongruence, and constructing reticulate phylogenies. *Am. J. Bot.* 96:920-932.
44. Nie XJ, Lv SZ, Zhang YX, Du XH, Wang L, Biradar SS, Tan XF, Wan FH, Song WN (2012) Complete Chloroplast Genome Sequence of a Major Invasive Species, Crofton Weed (*Ageratina adenophora*). *Plos One* 7.
45. Noyes RD (2000) Biogeographical and evolutionary insights on *Erigeron* and allies (Asteraceae) from ITS sequence data. *Plant Syst. Evol* 220:93-114
46. Noyes RD, Rieseberg LH (1999) ITS sequence data support a single origin for North American Astereae (Asteraceae) and reflect deep geographic divisions in *Asters*. *Am. J. Bot* 86(3): 398-412.
47. Ou CZ, Feng YL, Hu YK, Tian XY, Fu ZX (2019) Characterization of the complete chloroplast genome sequence of *Aster hersileoides* (Asteraceae, Astereae) and its phylogenetic implications. *Mitochondrial DNA Part B* 4:985-986.
48. Palazzesi L, Pellicer J, Barreda VD, Loeuille B, Mandel JR, Pokorny L, Siniscalchi CM, Telleria MC, Leitch IJ, Hidalgo O (2022) Asteraceae as a model system for evolutionary studies: from fossils to genomes. *Bot. J. Linn. Soc* 200:143-164.
49. Panero JL, Funk VA (2008) The value of sampling anomalous taxa in phylogenetic studies: Major clades of the Asteraceae revealed. *Mol Phylogenet Evol* 47:757-782.
50. Park J, Shim J, Won H, Lee J (2017) Plastid genome of *Aster altaicus* var. *uchiyamae* Kitam., an endanger species of Korean asterids. *Journal of Species Research* 2017 6: 76-90.
51. Parks M, Cronn R, Liston A (2009) Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC Biol* 7.
52. Pauwels M, Vekemans X, Gode C, Frerot H, Castric V, Saumitou-Laprade P (2012) Nuclear and chloroplast DNA phylogeography reveals vicariance among European populations of the model species for the study of metal tolerance, *Arabidopsis halleri* (Brassicaceae). *New Phytol* 193:916-928.
53. Peng YJ, Cui XY, Tan MC, Hu L, Ruan HY, Shao YY, Song EL, Tang YJ (2020) Characterization of the complete chloroplast genome of *Chlorophytum comosum* (Liliaceae). *Mitochondrial DNA Part B* 5:1046-1047.
54. Provan J, Pm. H, Powell W (2001) Chloroplast microsatellites: new tools for studies in plant ecology and evolution [Review]. *Trends Ecol. Evol* 16: 142-147.
55. 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: 1-12.
56. Quax TEF, Claassens NJ, Soll D, van der Oost J (2015) Codon Bias as a Means to Fine-Tune Gene Expression. *Mol. Cell* 59:149-161.
57. Salih RHM, Majesky L, Schwarzacher T, Gornall R, Heslop-Harrison P (2017) Complete chloroplast genomes from apomictic *Taraxacum* (Asteraceae): Identity and variation between three microspecies. *Plos One* 12: e0168008.

58. Saslis-Lagoudakis CH, Savolainen V, Williamson EM, Forest F, Wagstaff SJ, Baral SR, Watson MF, Pendry CA, Hawkins JA (2012) Phylogenies reveal predictive power of traditional medicine in bioprospecting. *Proc. Natl. Acad. Sci. U.S.A.* 109:15835-15840.
59. Shahzadi I, Abdullah, Mehmood F, Ali Z, Ahmed I, Mirza B (2020) Chloroplast genome sequences of *Artemisia maritima* and *Artemisia absinthium*: Comparative analyses, mutational hotspots in genus *Artemisia* and phylogeny in family Asteraceae. *Genomics* 112:1454-1463.
60. Sharp PM, Li WH (1987) The codon Adaptation Index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucl Acids Res* 15:1281-1295
61. Shaw J, Lickey EB, Schilling EE, Small RL (2007) Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: The tortoise and the hare III. *Am. J. Bot* 94:275-288.
62. Shen XF, Wu ML, Liao BS, Liu ZX, Bai R, Xiao SM, Li XW, Zhang BL, Xu J, Chen SL (2017) Complete Chloroplast Genome Sequence and Phylogenetic Analysis of the Medicinal Plant *Artemisia annua*. *Molecules* 22.
63. Shimada H, Sugiura M (1991) Fine structural features of the chloroplast genome: comparison of the sequenced chloroplast genomes. *Nucleic acids research* 19:983-995
64. Smith DR (2015) Mutation Rates in Plastid Genomes: They Are Lower than You Might Think. *Genome Biol. Evol* 7:1227-1234.
65. Somaratne Y, Guan DL, Wang WQ, Zhao L, Xu SQ (2019) Complete chloroplast genome sequence of *Xanthium sibiricum* provides useful DNA barcodes for future species identification and phylogeny. *Plant Syst. Evol* 305:949-960.
66. Song B, Si JG, Yu M, Tian XH, Ding G, Zou ZM (2017) Megastigmane glucosides isolated from *Dichrocephala benthamii*. *Chin. J. Nat. Med.* 15:288-291.
67. Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312-1313.
68. Tian XH, Ding G, Peng CZ, Hu YB, Li L, Chen H, Zou ZM (2013a) Sinapyl Alcohol Derivatives from the Lipo-soluble Part of *Dichrocephala benthamii* C. B. Clarke. *Molecules* 18:1720-1727.
69. Tian XH, Li L, Hu YB, Zhang HW, Liu YT, Chen H, Ding G, Zou ZM (2013b) Dichrocephones A and B, two cytotoxic sesquiterpenoids with the unique 3.3.3 propellane nucleus skeleton from *Dichrocephala benthamii*. *Rsc Advances* 3:7880-7883.
70. Tonti-Filippini J, Nevill PG, Dixon K, Small I (2017) What can we do with 1000 plastid genomes? *Plant J* 90:808-818.
71. 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* 214:1736-1750.
72. Wang Q, Wang XH, Yang XC, Fu LF, Gong QJ, Fu ZX (2019) The complete chloroplast genome of *Aster hypoleucus* (Asteraceae: Astereae): an endemic species from China. *Mitochondrial DNA Part B* 4:2647-2648.
73. Wang, M.C.; Li, R.; Zhao, Q. Multi-Omics Techniques for Forest Plant Breeding. *Preprints.org* 2023, 2023040549. <https://doi.org/10.20944/preprints202304.0549.v1>.
74. Wick RR, Schultz MB, Justin Z, Holt KE Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics* 31:3350-3352
75. Wicke S, Schneeweiss GM, dePamphilis CW, Muller KF, Quandt D (2011) The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol. Biol* 76:273-297.
76. Xia Y, Hu ZG, Li XW, Wang P, Zhang XQ, Li Q, Lu CL (2016) The complete chloroplast genome sequence of *Chrysanthemum indicum*. *Mitochondrial DNA Part A* 27:4668-4669.
77. Yu SH, Yang XC, Tian XY, Liu XF, Lu X, Huang CP, Fu ZX (2022) The complete chloroplast genome sequence of the monotypic and enigmatic genus *Cavea* (tribe Gymnarrheneae) and a comparison with other species in Asteraceae. *J. Genet* 101. 20.
78. Zavala-Paez M, Vieira LD, de Baura VA, Balsanelli E, de Souza EM, Cevallos MC, Chase MW, Smidt ED (2020) Comparative Plastid Genomics of Neotropical *Bulbophyllum* (Orchidaceae; Epidendroideae). *Front Plant Sci* 11.

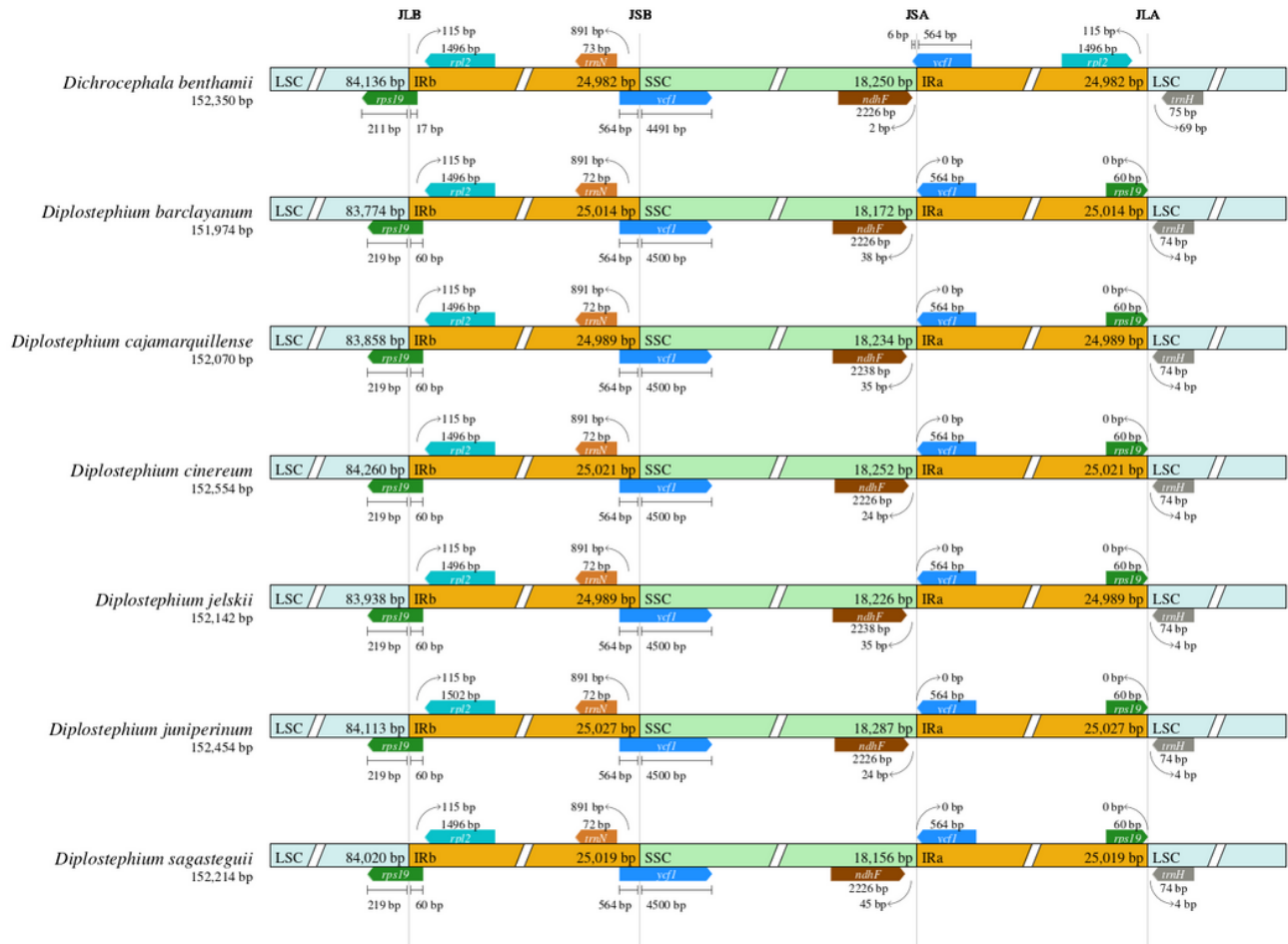
79. Zhang X, Deng T, Moore MJ, Ji YH, Lin N, Zhang HJ, Meng AP, Wang HC, Sun YX, Sun H (2019) Plastome phylogenomics of *Saussurea* (Asteraceae: Cardueae). *BMC Plant Biol* 19:290.
80. Zhang X, Jiang PP, Fan SJ (2021) Characterization of the complete plastome of *Aster pekinensis* (Asteraceae), a perennial herb. *Mitochondrial DNA Part B* 6:1064-1065.
81. Zhang ZY, Jiang XB, Chen Y, Zhu P, Li L, Zeng YL, Tang TM (2019) Characterization of the complete chloroplast genome sequence of *Conyza canadensis* and its phylogenetic implications. *Mitochondrial DNA Part B* 4:2028-2030.
82. Zheng TR, Liu MY, Wang T, Chen H (2017) The complete chloroplast genome sequence of medicinal plant *Conyza blinii* H. Lev. *Mitochondrial DNA Part B* 2:50-51.
83. Zhong Q, Yang S, Sun X, Wang L, Li Y (2019) The complete chloroplast genome of the Jerusalem artichoke (*Helianthus tuberosus* L.) and an adaptive evolutionary analysis of the *ycf2* gene. *PeerJ* 7: e7596.
84. Zhou JY, Li JJ, Peng SB, An XM (2022) Characterization of the complete chloroplast genome of the invasive plant *Erigeron annuus* (L.) Pers. (Asterales: Asteraceae). *Mitochondrial DNA Part B* 7:188-190.

## Figures



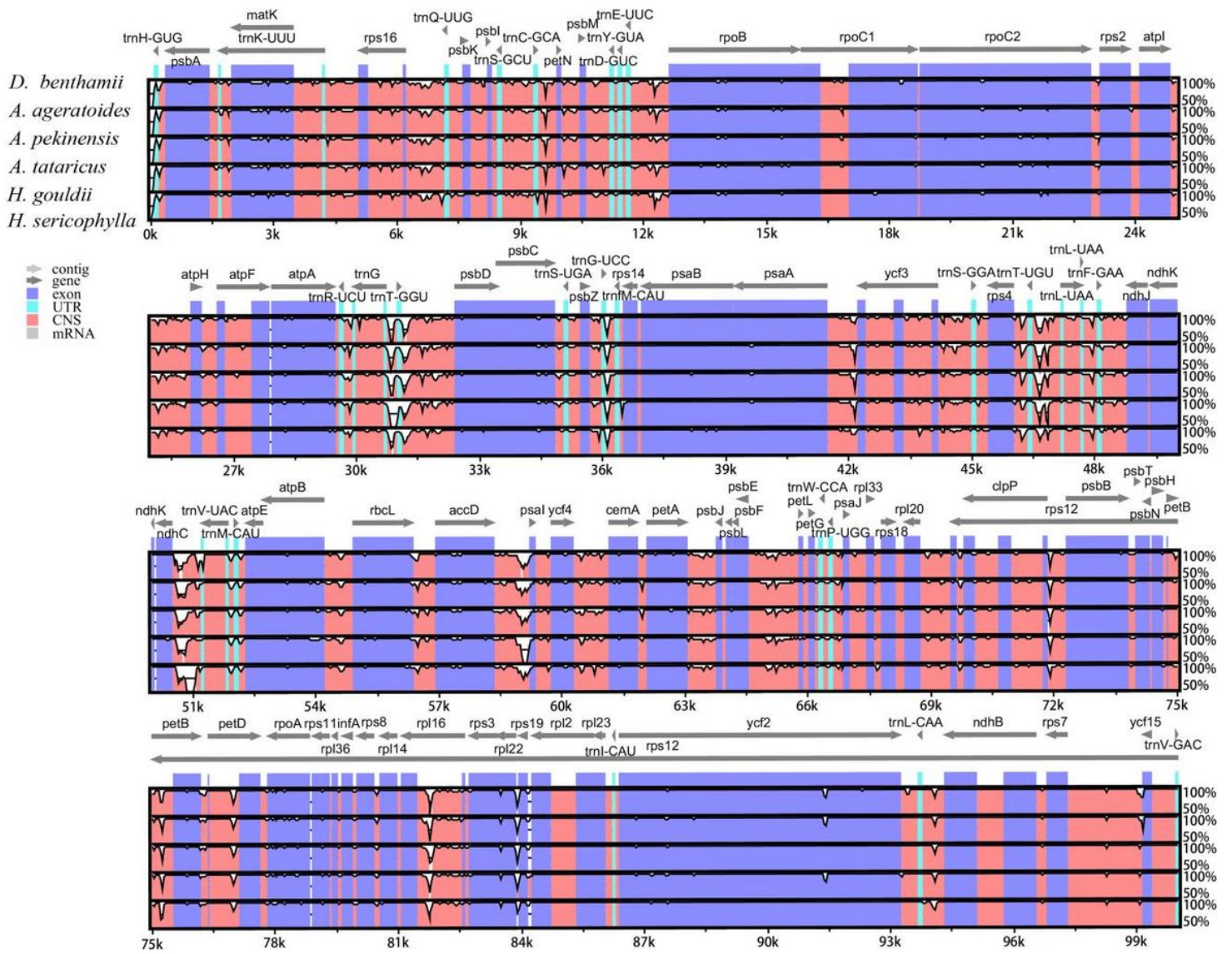
**Figure 1**

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.



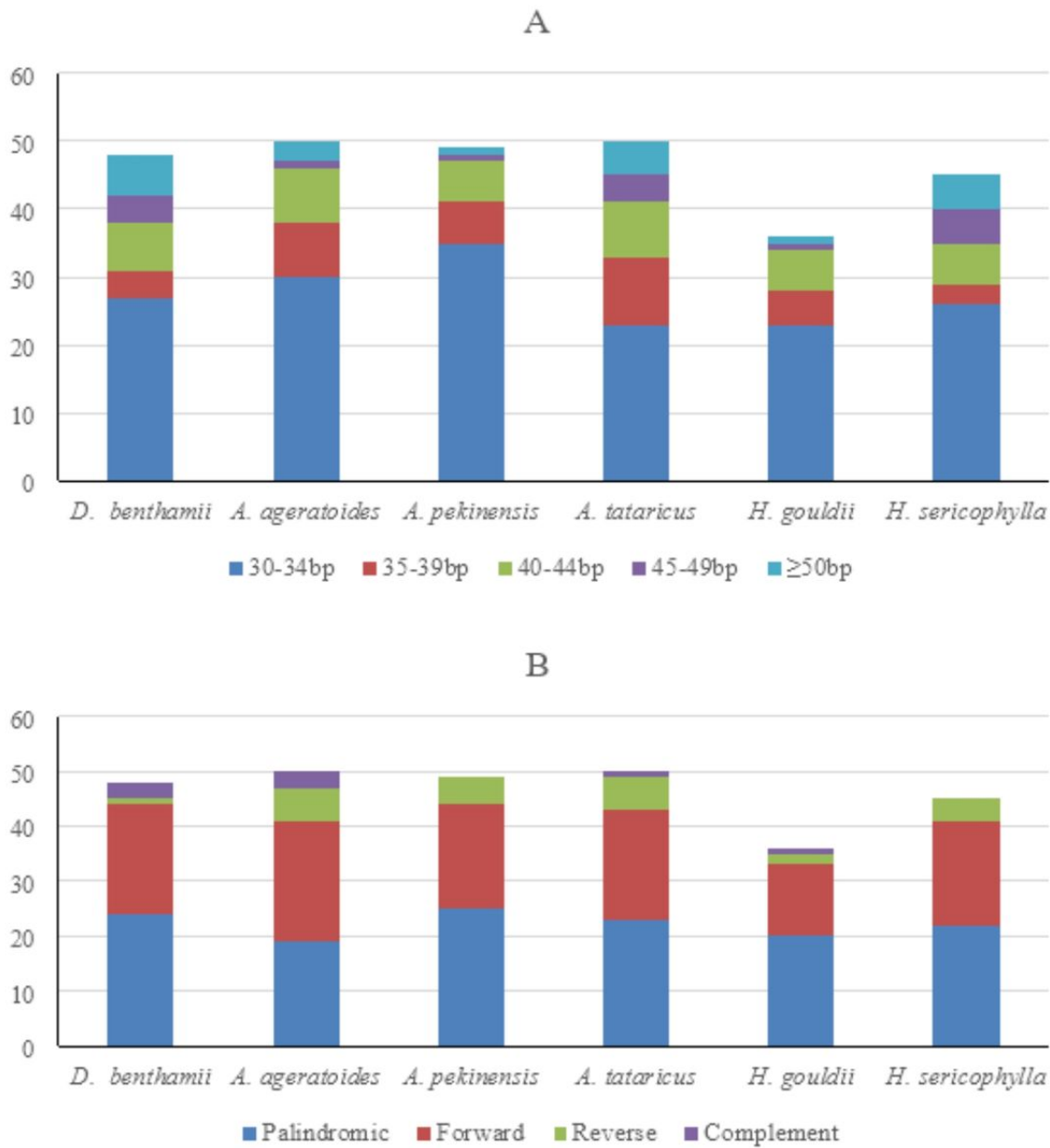
**Figure 2**

Comparison for border positions of LSC, SSC, and IR regions among the 6 species of tribe Astereae. The boxes denote genes, and the gap between the genes and the boundaries is indicated by the number of bases unless the gene coincides with the boundary. Extensions of genes are shown above the boxes.



**Figure 3**

The cp genomes of 6 species of Astereae were compared using the mVista program with *D. benthamii* as the reference. The arrow indicates the position and orientation of the gene. The Y-axis represents the percentage of identity among genome sequences (50%–100%).



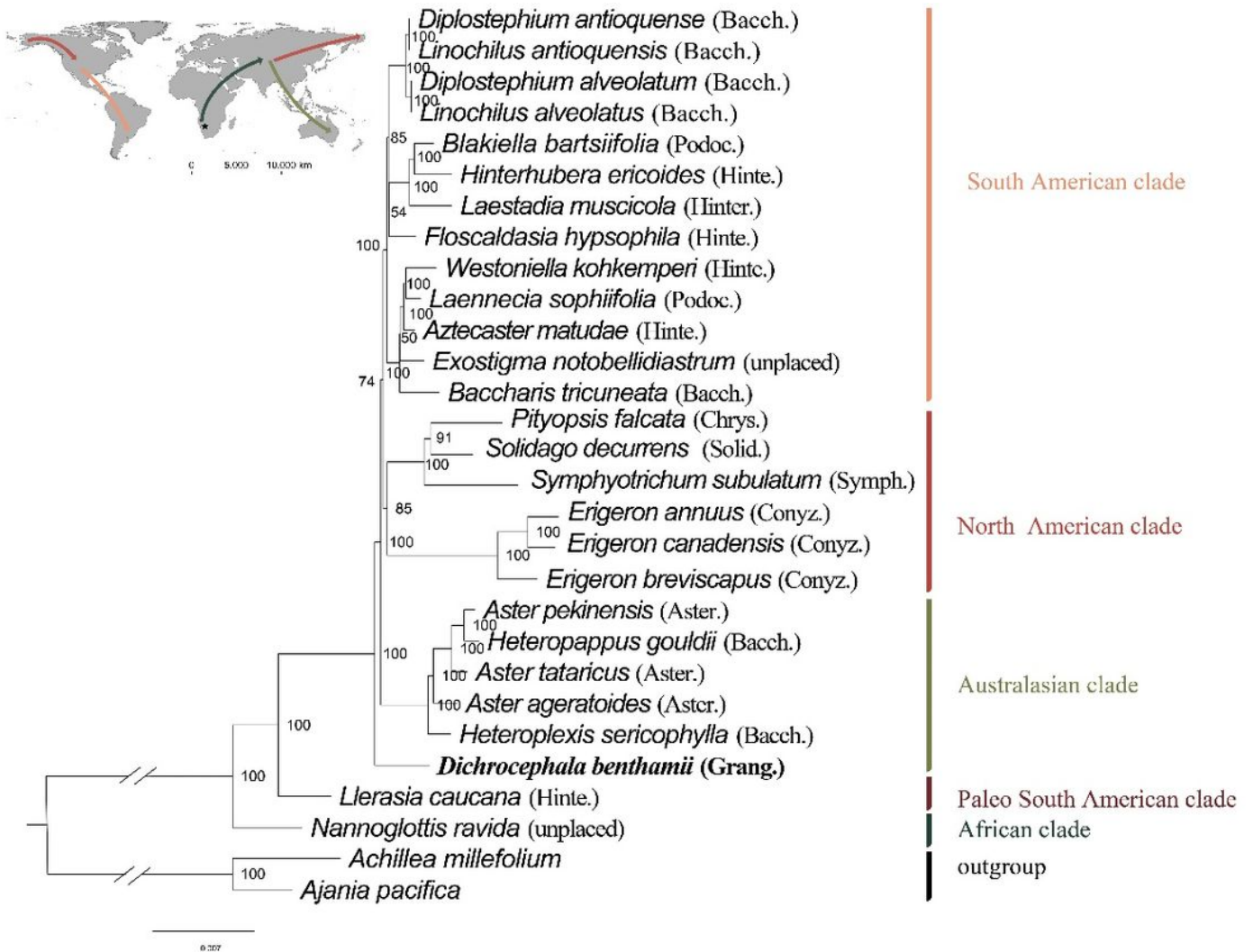
**Figure 4**

Analysis of interspersed repeat sequences in the 6 species of Astereae chloroplast genomes. (A) Frequency of repeat sequences by length; (B) Frequency of repeat types.





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



**Figure 7**

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.: Podocomaenae; Hinte.: Hinterhuberinae; Chrys.: Chrysopsidinae; Solid.: Solidagininae; Symph.: Symphyotrichinae; Conyz.: Conyzinae; Aster.: Asterinae; Grang.: Grangeinae; pentagram: Astereae originated in Africa.)