scientific reports

Check for updates

OPEN

# Comparative and phylogenetic analysis of the complete chloroplast genome sequences of *Melanoseris cyanea* group

Qianqian Zhong, Zehuan Wang✉, Jiaju Xu, Li Yan & Qingwen Sun

***Melanoseris*, a diverse genus in the Lactucinae subtribe, has 21 species in China, with 13 being endemic. The high diversity of this genus presents taxonomic challenges, particularly in the *M. cyanea* group, where overlapping distributions and transitional morphological traits complicate classification. This study aims to analyze the chloroplast genomes of *Melanoseris*, with a focus on the *M. cyanea* group, to explore structural differences and phylogenetic relationships among these closely related species. We analyzed the chloroplast genomes of 16 *Melanoseris* samples, including 12 new genomes from the *M. cyanea* group. The genome sizes ranged from 152,255 to 152,558 bp and exhibited a typical quadripartite structure, with an average GC content of 37.7%. Each genome encodes 132 genes, including 87 protein-coding genes, 37 tRNAs, and 8 rRNAs. Repeat analysis identified 89 to 105 dispersed repeats, 24 to 28 tandem repeats, and 35 to 39 SSRs, with mononucleotide A/T repeats being the most common. Sequence alignment revealed that variable regions were mainly concentrated in the single-copy regions. Nucleotide diversity ranged from 0 to 0.00485, highlighting 10 mutation hotspot regions. Phylogenetic analysis showed a limited differentiation among species within the *M. cyanea* group. This research enhances our understanding of the genetic diversity of *Melanoseris*, laying the foundation for future taxonomic and phylogenetic studies.**

**Keywords** *Melanoseris cyanea* group, Chloroplast genome, Genomic structure, Phylogenetic analysis, Comparative analysis

*Melanoseris* Decne., a genus in the subtribe Lactucinae of the tribe Cichorieae within the Asteraceae family, was first described by Decaisne in 1843[1]. It was originally established for two Himalayan species: *M. lessertiana* (DC.) Decne. (the type species) and *M. lyrata* Decne. (now considered a synonymy of *M. lessertiana*). In 1846, Edgeworth[2] expanded the genus by adding five more Himalayan species: *M. cyanea* (D.Don) Edgew., *M. hastata* (DC.) Edgew., *M. paniculata* Edgew., *M. rapunculoides* (DC.) Edgew., and *M. saxatilis* Edgew. After this expansion, *Melanoseris* faded into obscurity until its revival in 2011 by N. Kilian[3] in the Flora of China. This revival was based on molecular phylogenetic studies[4] that revealed the inclusion of the new genera *Chaetoseris* Shih and *Stenoseris* Shih, which are distinguished from *Lactuca* L. and *Cicerbita* Wallr. from the Sino-Himalayan region[5]. In addition, several species previously classified under *Cephalorrhynchus* Boiss., *Cicerbita*, *Lactuca*, *Mulgedium* Cass., and *Prenanthes* L. were found to belong to the broader *Melanoseris* clade.

Recent advances in understanding the phylogeny, molecular dating, and biogeography of the genus have focused on the subtribe Lactucinae or *Lactuca* closely related genera through analyses of non-coding nuclear marker (nrITS, ETS), plastid DNA regions (*ndhF*, *petD*, *psbA-trnH*, *rpl32-trnL*, *trnL-F*, *trnQ-rps16*), and plastomes[6–9]. Phylogenetic studies of Sino-Himalayan species have revealed that the newly established genus *Parasyncalathium* J.W.Zhang, Boufford & H.Sun[10], with a single species, forms a basal lineage within the Chinese *Melanoseris* clade. This clade, in turn, forms a distinct lineage within the Lactucinae subtribe, showing a sister-group relationship with the *Lactuca* lineage[6]. Global sampling has further expanded the distribution of *Melanoseris*, incorporating African endemic *Lactuca* species (AE *Lactuca*). In this expanded phylogeny, *Melanoseris* forms a well-supported clade (Clade J), with AE *Lactuca* (Clade J1) taxa is sister to the other two clades: Clade J2 (comprising species primarily from SW and Central Asian mountain ranges) and Clade J3, the Sino-Himalayan–Southeast Asian clade[7]. The close relationship between *Melanoseris* and AE *Lactuca* species has also been supported by a recent plastome-based phylogeny, where these species form a separate clade sister

Department of Traditional Chinese Medicine Resources and Development, College of Pharmacy, Guizhou University of Traditional Chinese Medicine, Guiyang 550025, Guizhou, China. ✉email: wzh0442046020@126.com

to the core *Lactuca* lineage, with no representatives from Clade J2[9]. However, further taxon sampling, especially of Asian *Melanoseris* species, is needed to fully resolve the phylogenetic relationships between AE *Lactuca* and *Melanoseris*, as only *M. decipiens* (C.B.Clarke) N.Kilian & Ze H.Wang from Clade J3 has been analyzed, and no sequences from Clade J2 have been included.

*Melanoseris* is the largest genus in the Lactucinae subtribe in China, with about 70 species recognized globally, 21 of which are found in China[3,4,6,7,10–22]. The Sino-Himalayan species, which constitute the majority of Clade J3 in *Melanoseris*, are among the most taxonomically complex and variable members of the genus. Within this Sino-Himalayan clade, Chinese species (excluding *M. bracteata* Hook.f. & Thomson ex Hook.f.) form a well-supported subclade, and four terminal clades or species groups have been identified, including the *M. cyanea* group — a particularly challenging taxonomic unit due to its morphological variability[7].

The *Melanoseris cyanea* group, which is the focus of this study, shares morphological similarities with *M. cyanea* and consists of 11 species that were previously classified under Shih's new established genus *Chaetoseris*[23]. Among them are *C. beesiana* (Diels) Shih, *C. cyanea* (D.Don) Edgew., *C. hastata* (DC.) Shih, *C. hispida* Shih, *C. lutea* (Hand.-Mazz.) Shih, *C. lyriformis* Shih, *C. macrocephala* Shih, *C. pectiniformis* Shih, *C. sichuanensis* Shih, *C. teniana* (Beauverd) Shih, and *C. yunnanensis* Shih (Table S1). These species are primarily distinguished by their floret color, leaf morphology, and trichomes on inflorescence branches, peduncles and involucres (Fig. 1 and Table S2). The morphological variability of these traits has made it particularly challenging to resolve species boundaries within this group.

Widely distributed in southwestern China, the *Melanoseris cyanea* group exhibits considerable polymorphism due to significant morphological variation. This variation has often been interpreted as representing distinct species, further complicating its taxonomy. Following Shih's[23] systematic classification of Chinese Lactucinae species, Zhu[12] conducted a taxonomic review of Shih's newly established genera *Chaetoseris* and *Stenoseris*. Using extensive morphological statistical data from herbarium specimens and field observations, Zhu treated *C. lyriformis*, *C. pectiniformis* and *C. sichuanensis* as synonyms of *C. cyanea* (Table S1). She noted that the variation in pubescence — from sparse glandular hairs to dense glandular bristles — forms a continuous spectrum with numerous intermediate forms, indicating a gradient of morphological change. Based on these observations, Zhu concluded that *C. cyanea* is a polymorphic species, a view that is consistent with the findings of Mamgain & Rao[24] and Mamgain[25], who also recognized this species (as *Cicerbita cyanea* (D.Don) Beauverd) as highly polymorphic and exhibiting significant intraspecific variation.

In 2011, Kilian[3] revived the genus name *Melanoseris* in Flora of China and formally recombined *Chaetoseris cyanea* as *M. cyanea* (C.Shih) N.Kilian, treating *C. hastata* and *C. hispida* as synonyms (Table S1). Two years later, Wang[6], based on a thorough study of herbarium specimens, extensive field investigations, and molecular phylogenetic analysis, expanded the scope of synonyms for *M. cyanea*. Wang proposed that, with the exception of *C. macrocephala*, the remaining nine species should all be merged into *M. cyanea* as synonyms (Table S1). Due to limited availability of herbarium specimens and incomplete morphological data at the time, *C. macrocephala* was temporarily retained as a distinct species under the name *M. macrocephala* (C.Shih) N.Kilian & J.W.Zhang. Wang's analysis highlighted the lack of distinct morphological breaks between these species, coupled with their


**Fig. 1**. Images of different species from the *Melanoseris cyanea* group, illustrating their capitulum characteristics. (**a**) *Chaetoseris cyanea* (D.Don) Shih. (**b**) *C. hastata* (DC.) Shih. (**c**) *C. lyriformis* Shih. (**d**) *C. sichuanensis* Shih. (**e**) *C. hispida* Shih. (**f**) & (**g**) *C. lutea* (Hand.-Mazz.) Shih. (**h**) *C. macrocephala* Shih. Photographed by Zehuan Wang.

sympatric distribution and transitional morphological characteristics. These factors led Wang to suggest that there is frequent gene flow within the *M. cyanea* group. She proposed that the species in this group are closely related morphologically, and emphasized the need to address questions regarding their potential origin and whether they should all be treated as a single species.

Numerous studies have shown that complete plastomes, compared to non-coding regions of plastid and nrDNA, provide more reliable and valuable data for phylogenetic reconstruction of angiosperm groups[8,9,26–28]. This study aims to analyze the chloroplast genome structures of *Chaetoseris cyanea*, *C. hastata*, *C. hispida*, *C. lutea*, *C. lyriformis*, *C. sichuanensis*, *C. macrocephala*, and *C. yunnanensis*. It also seeks to assess their phylogenetic relationships within the *Melanoseris cyanea* group and clarify the evolutionary connections among these species. Additionally, this research provides essential foundational data and new insights for the evolutionary studies of the entire Lactucinae subtribe and the Asteraceae family. Ultimately, this will contribute to the refinement of the classification system within the Asteraceae family and offer scientific support for biodiversity conservation.

## Results

### Chloroplast genomic structure and characteristics analyses of the *Melanoseris cyanea* group

By sequencing and assembling the chloroplast genomes of 12 *Melanoseris cyanea* group samples, the study demonstrates that these genomes exhibit highly conserved structures (Fig. 2). Additionally, a statistical analysis of all 16 available *Melanoseris* chloroplast genomes revealed that their lengths range from 152,255 bp to 152,558 bp, with a typical quadripartite structure consisting of a pair of inverted repeat regions (IRa and IRb, 24,992 bp–25,043 bp), a large single-copy region (LSC, 83,662 bp–83,940 bp), and a small single-copy region (SSC, 18,518 bp–18,627 bp) (Table 1). The study also found that the total GC content of these species ranges from 37.67 to 37.70%, with the IR region having the highest GC content at 43.15%, which is significantly higher than that of the LSC and SSC regions.

Additionally, these 16 chloroplast genomes contain 132 genes, including 87 protein-coding genes, 37 tRNA genes, and 8 rRNA genes (Table 1). Among these species, 17 genes exist in two copies, comprising 7 protein-coding genes (*ndhB*, *rpl2*, *rpl23*, *rps12*, *rps7*, *ycf15*, *ycf2*), 6 tRNA genes (*trnL-CAA*, *trnV-GAC*, *trnI-GAU*, *trnA-UGC*, *trnR-ACG*, *trnN-GUU*), and 4 rRNA genes (*rrn5*, *rrn4.5*, *rrn23*, *rrn16*) (Table 2). Additionally, the genes *trnS-GCU* and *trnM-CAU* each have three copies. The arrangement of these 132 genes is completely homologous across all chloroplast genomes. Notably, 4 genes (*rps12* (×2), *ycf3*, and *clpP*) contain two introns, while 19 genes (*atpF*, *ndhA*, *ndhB* (×2), *petB*, *petD*, *rpl16*, *rpl2* (×2), *rps16*, *rpoC1*, *trnK*, *trnG-UCC*, *trnL-UAA*, *trnV-UAC*, *trnI-GAU* (×2), *trnA-UGC* (×2)) each contain one intron.

### Repeat sequence analysis

The study identified 89 to 105 dispersed repeat sequences across 16 chloroplast genomes, comprising 43 to 48 forward repeats (F), 34 to 46 palindromic repeats (P), 3 to 5 complementary repeats (C), and 6 to 17 reverse repeats (R) (Fig. 3a). Among these, the most prevalent repeat sequence lengths are 25 bp or fewer, as well as 26 to 28 bp (Fig. 3b). Longer repeat sequences exceeding 50 bp were only found in some samples and occurred at a low frequency. Forward and palindromic repeats are the most abundant types, with the majority of dispersed repeats predominantly located in the large single-copy (LSC) region (Fig. 3c).

The number of tandem repeats varies between 24 and 28 in the analysis of 16 chloroplast genomes, with *Melanoseris decipiens* NC_066793 having the fewest (Fig. 4a). Most tandem repeats are located in the LSC region, with 3 repeats in each of the IR regions. In the SSC region, only *M. decipiens* NC_066793 and *M. macrantha* PP525145 have 4 tandem repeats, while all other have 3 (Fig. 4b).

A total of 594 SSRs were identified across the 16 chloroplast genomes, with the number of SSRs ranging from 35 to 39. Mononucleotide repeats (T/A/C) were the most common type, ranging from 25 to 28, followed by dinucleotide repeats (AT/TA), totaling 4 to 5 (Fig. 5a). Overall, the SSRs are primarily found in the LSC region (Fig. 5b), with minimal differences in SSR types and quantities (Fig. 5c).

### IR contraction and expansion

The boundaries of the IR/SSC and IR/LSC regions exhibit a conserved pattern across the 16 chloroplast genomes. Among the 13 *Melanoseris cyanea* group samples, the IR region is consistently 25,032 bp, except in *Chaetoseris cyanea* M8-1, where it is 25,031 bp. The shortest IR region is in *M. decipiens* NC_066793 (24,992 bp), and the longest is in *M. macrantha* PP525145 (25,043 bp). In all samples, the LSC/IRb boundary is located within the *rps19* gene, with 60 bp in the IRb region, resulting in an incomplete *rps19* pseudogene in the IRa region. The IRb/SSC boundary falls within the *ycf1* gene, with 472 bp in the IRb region, leading to an incomplete *ycf1* pseudogene in the IRa region. Additionally, the *ndhF* and *trnH* genes are situated in the SSC and LSC regions, respectively, without extending into the IR region (Fig. 6).

### Comparative genomic analysis

Using mVISTA[29] software, a comparative map was generated to analyze the chloroplast genomes of 16 *Melanoseris* samples, using *Chaetoseris cyanea* M8-1 as the reference (Fig. 7). The analysis reveals that these chloroplast genomes are highly conserved across species, with regions of the conserved non-coding sequences (CNS) showing slightly more variation than the exonic regions. *M. macrantha* PP525145 exhibits a noticeable deletion between *trnH-GUG* and *psbA*, while *M. ciliata* PP525143 has one between *ycf3* and *trnS-GCU*. Deletions in the *rpl16–rps3*, *rrn5S–trnN-GUU*, *ycf1*, and *trnN-GUU–trnR-ACG* regions were observed in *M. decipiens* NC_066793, *M. ciliata* PP525143, and *M. macrantha* PP525145, whereas the *M. cyanea* group showed few or no deletions. Additional deletions occurred in multiple species between *trnK–rps16*, *atpI–atpH*, *trnT-GGU–psbD*, and *rbcL–accD*.

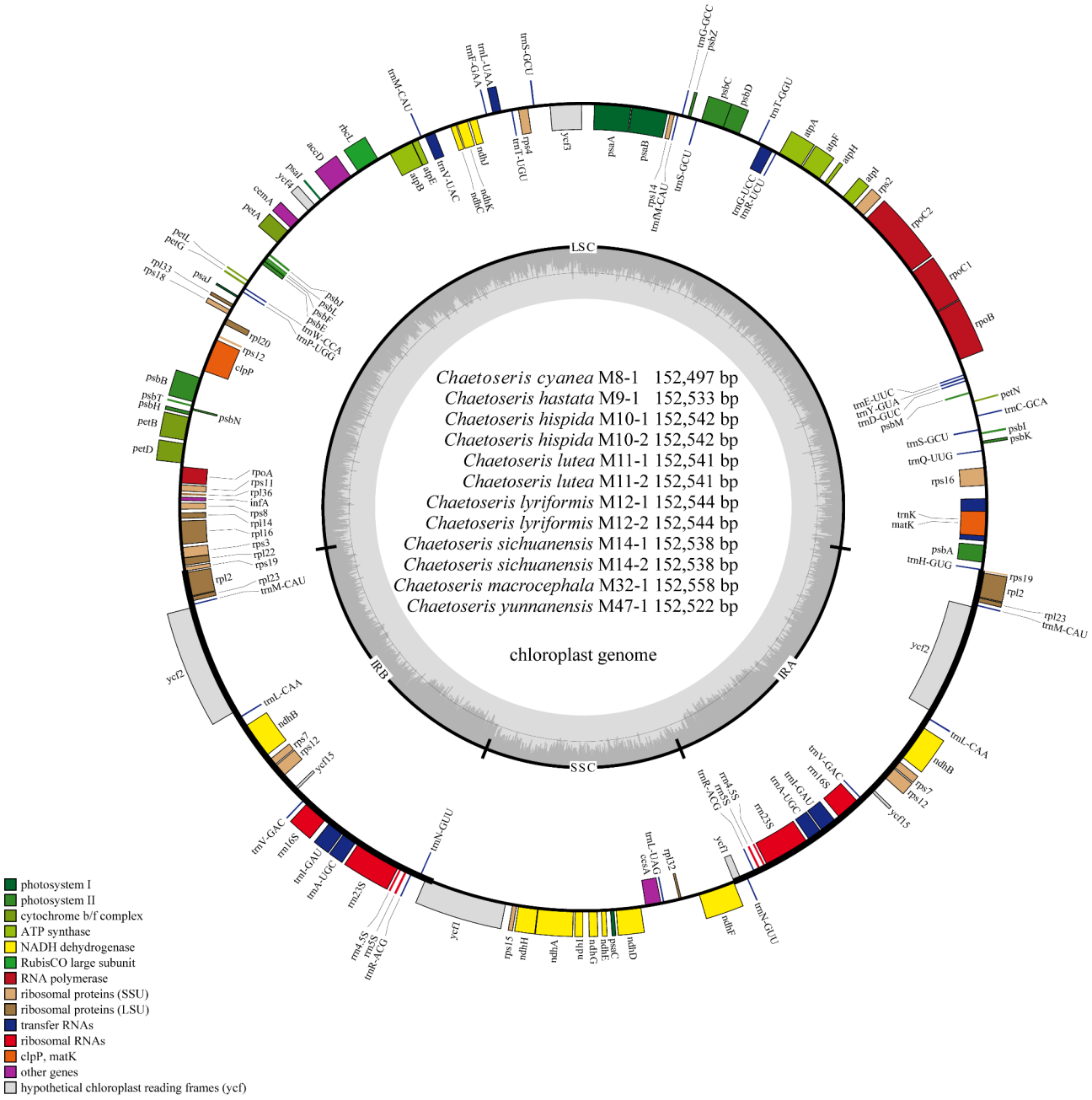


**Fig. 2**. Gene map of the chloroplast genome of the *Melanoseris cyanea* group. Genes shown outside the outer circle are transcribed clockwise, and those insides are transcribed counterclockwise. Genes are color-coded according to different functional groups. The darker gray in the inner circle indicates the GC content, and the lighter gray indicates the AT content. The inner circle also shows that the chloroplast genome contains two copies of inverted repeats (IRA and IRB), a large single-copy (LSC) region, and a small single-copy (SSC) region.

Further analysis using DnaSP v6[30] was performed to identify highly variable regions (Fig. 8). The findings corroborate the mVISTA analysis, indicating that the chloroplast genomes are highly conserved across species, with no significant regions of high variability detected. The two IR regions demonstrated remarkable conservation compared to the LSC and SSC regions. The highest Pi value, 0.00485, was observed in the *ycf1* gene within the SSC region. Additionally, the region between the *ycf1* and *rps15* genes in the SSC, as well as regions in the LSC such as *trnK*, *trnK–rps16*, *trnS-GCU–petN*, *petN–psbM*, *rps18–rpl20*, *clpP–psbB*, and *rpl16–rps3*, exhibited relatively high nucleotide diversity, with Pi values exceeding 0.002.

| Species | Length | | | | GC% | | | | PCG | tRNA | rRNA | Total genes |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | cp. genome | LSC | SSC | IR | cp. genome | LSC | SSC | IR | | | | |
| *C. cyanea* M8-1 | 152,497 | 83,889 | 18,546 | 25,031 | 37.69 | 35.84 | 31.32 | 43.15 | 87 | 37 | 8 | 132 |
| *C. hastata* M9-1 | 152,533 | 83,923 | 18,546 | 25,032 | 37.69 | 35.84 | 31.34 | 43.15 | 87 | 37 | 8 | 132 |
| *C. hispida* M10-1 | 152,542 | 83,932 | 18,546 | 25,032 | 37.69 | 35.83 | 31.34 | 43.15 | 87 | 37 | 8 | 132 |
| *C. hispida* M10-2 | 152,515 | 83,905 | 18,546 | 25,032 | 37.69 | 35.84 | 31.34 | 43.15 | 87 | 37 | 8 | 132 |
| *C. lutea* M11-1 | 152,541 | 83,931 | 18,546 | 25,032 | 37.69 | 35.83 | 31.34 | 43.15 | 87 | 37 | 8 | 132 |
| *C. lutea* M11-2 | 152,508 | 83,884 | 18,560 | 25,032 | 37.69 | 35.84 | 31.33 | 43.15 | 87 | 37 | 8 | 132 |
| *C. lyriformis* M12-1 | 152,544 | 83,934 | 18,546 | 25,032 | 37.69 | 35.83 | 31.35 | 43.15 | 87 | 37 | 8 | 132 |
| *C. lyriformis* M12-2 | 152,530 | 83,920 | 18,546 | 25,032 | 37.69 | 35.84 | 31.33 | 43.15 | 87 | 37 | 8 | 132 |
| *C. sichuanensis* M14-1 | 152,538 | 83,928 | 18,546 | 25,032 | 37.69 | 35.84 | 31.34 | 43.15 | 87 | 37 | 8 | 132 |
| *C. sichuanensis* M14-2 | 152,550 | 83,940 | 18,546 | 25,032 | 37.69 | 35.83 | 31.34 | 43.15 | 87 | 37 | 8 | 132 |
| *C. macrocephala* M32-1 | 152,558 | 83,928 | 18,566 | 25,032 | 37.68 | 35.82 | 31.34 | 43.15 | 87 | 37 | 8 | 132 |
| *C. yunnanensis* M47-1 | 152,522 | 83,919 | 18,539 | 25,032 | 37.69 | 35.83 | 31.34 | 43.15 | 87 | 37 | 8 | 132 |
| *M. decipiens* NC_066793 | 152,474 | 83,863 | 18,627 | 24,992 | 37.67 | 35.85 | 31.24 | 43.12 | 87 | 37 | 8 | 132 |
| *M. ciliata* PP525143 | 152,255 | 83,681 | 18,518 | 25,028 | 37.7 | 35.85 | 31.33 | 43.15 | 87 | 37 | 8 | 132 |
| *M. cyanea* PP525144 | 152,542 | 83,932 | 18,546 | 25,032 | 37.69 | 35.83 | 31.35 | 43.15 | 87 | 37 | 8 | 132 |
| *M. macrantha* PP525145 | 152,290 | 83,662 | 18,542 | 25,043 | 37.7 | 35.88 | 31.29 | 43.12 | 87 | 37 | 8 | 132 |

**Table 1**. Features of the 16 Chloroplast genomes of *Melanoseris*.

| Category of genes | Group of genes | Name of genes | Number |
|---|---|---|---|
| Genes for photosynthesis | Subunits of ATP synthase | *atpA*, *atpB*, *atpE*, *atpF**, *atpH*, *atpI* | 6 |
| | Subunits of NADH-dehydrogenase | *ndhA**, *ndhB**(×2), *ndhC*, *ndhD*, *ndhE*, *ndhF*, *ndhG*, *ndhH*, *ndhI*, *ndhJ*, *ndhK* | 12 |
| | Subunits of cytochrome b/f complex | *petA*, *petB**, *petD**, *petG*, *petL*, *petN* | 6 |
| | Subunits of photosystem I | *psaA*, *psaB*, *psaC*, *psaI*, *psaJ* | 5 |
| | Subunits of photosystem II | *psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbH*, *psbI*, *psbJ*, *psbK*, *psbL*, *psbM*, *psbN*, *psbT*, *psbZ*, *ycf3*** | 16 |
| | Subunit of rubisco | *rbcL* | 1 |
| Self-replication | Large subunit of ribosome | *rpl14*, *rpl16**, *rpl2** (×2), *rpl20*, *rpl22*, *rpl23* (×2), *rpl32*, *rpl33*, *rpl36* | 11 |
| | Small subunit of ribosome | *rps11*, *rps12***(×2), *rps14*, *rps15*, *rps16**, *rps18*, *rps19*, *rps2*, *rps3*, *rps4*, *rps7* (×2), *rps8* | 14 |
| | DNA dependent RNA polymerase | *rpoA*, *rpoB*, *rpoC1**, *rpoC2* | 4 |
| | Ribosomal RNAs | *rrn16S* (×2), *rrn23S* (×2), *rrn4.5 S* (×2), *rrn5S* (×2) | 8 |
| | Transfer RNAs | *trnH-GUG*, *trnK**, *trnQ-UUG*, *trnS-GCU* (×3), *trnC-GCA*, *trnD-GUC*, *trnY-GUA*, *trnE-UUC*, *trnR-UCU*, *trnG-UCC**, *rnT-GGU*, *trnG-GCC*, *trnfM-CAU*, *trnT-UGU*, *trnL-UAA**, *trnF-GAA*, *trnV-UAC**, *trnM-CAU* (×3), *trnW-CCA*, *trnP-UGG*, *trnL-CAA* (×2), *trnV-GAC* (×2), *trnI-GAU** (×2), *trnA-UGC** (×2), *trnR-ACG* (×2), *trnN-GUU* (×2), *trnL-UAG* | 37 |
| Other genes | c-type cytochrom synthesis gene | *ccsA* | 1 |
| | Envelop membrane protein | *cemA* | 1 |
| | Maturase | *matK* | 1 |
| | Protease | *clpP*** | 1 |
| | Subunit of Acetyl-CoA-carboxylase | *accD* | 1 |
| | Translational initiation factor | *infA* | 1 |
| Unkown | Conserved open reading frames | *ycf1*, *ycf15* (×2), *ycf2* (×2), *ycf4* | 6 |

**Table 2**. List of genes identified in the 16 complete Chloroplast genomes of *Melanoseris*. Intron-containing genes are marked by asterisks (*), *gene with one intron; **gene with two introns.

## Phylogenetic analysis

A total of 6,042 parsimony-informative sites were obtained from the aligned 62 chloroplast genome sequences using MAFFT in Phylosuite v1.2.3[31]. The phylogenetic trees constructed using Maximum Likelihood (ML) and Bayesian Inference (BI) methods exhibited largely similar topologies (Fig. 9), with strong support for most nodes, although some discrepancies were observed at specific branches. Both the complete set of 16 *Melanoseris* sequences and the 13 sequences from the *M. cyanea* group formed a highly supported clade (BS = 100, PP = 1). This *Melanoseris* clade is sister to the AE *Lactuca* clade, and together, they form a sister group to the core *Lactuca* clade. Within the *M. cyanea* group, the overall topology of the IQ tree is slightly better than that of the BI tree. However, except for *Chaetoseris sichuanensis* (BS = 67, PP = 0.999), the other species with two samples did not form single clades and displayed variable grouping patterns with other species.

In addition, analysis of nrITS revealed 258 parsimony-informative sites identified from the aligned 60 sequences using MUSCLE[32]. The phylogenetic tree constructed using the same methods as described previously

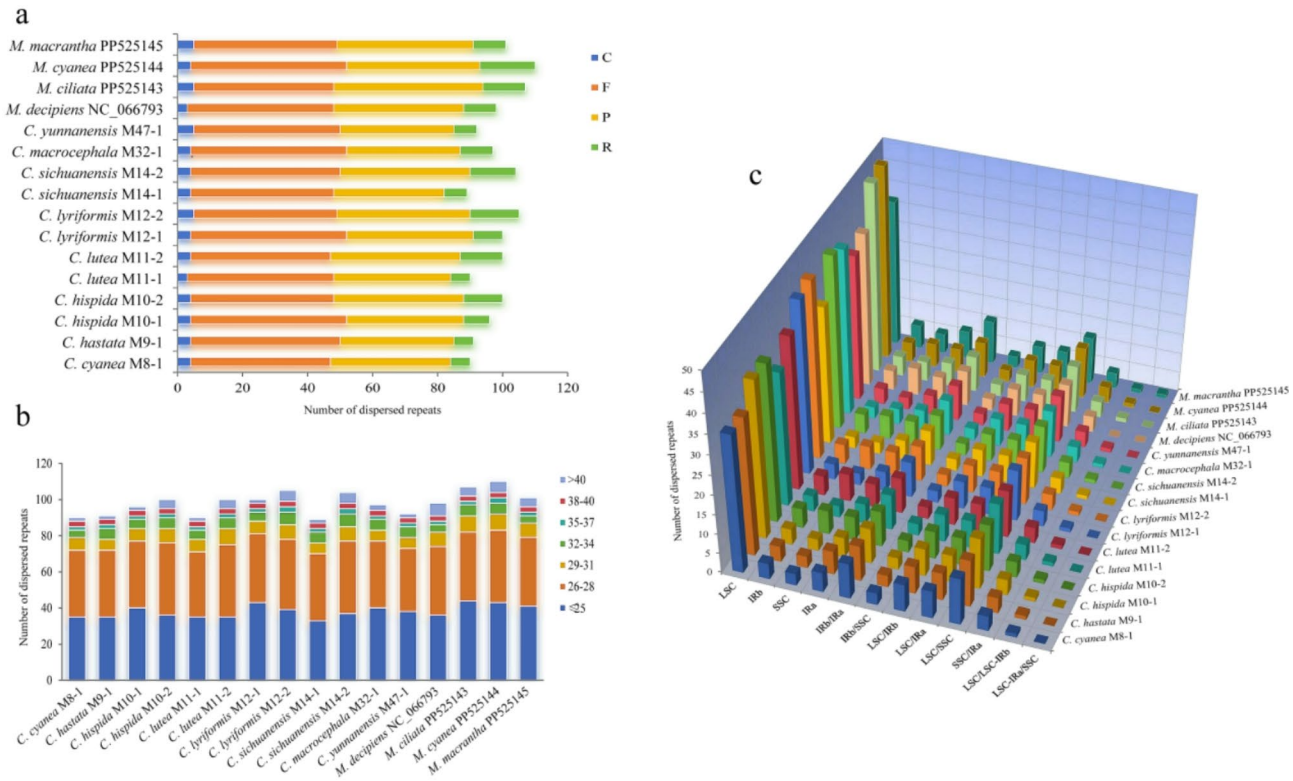


**Fig. 3**. Analysis of dispersed repeated sequences in the 16 cp. genomes of *Melanoseris*. (**a**) Statistics of four types of dispersed repeats sequences. "C" represents complementary repeats, "F" represents forward repeats, "P" represents palindromic repeats, and "R" represents reverse repeats. (**b**) Length statistics of dispersed repeated. (**c**) Distribution of dispersed repeats sequences.

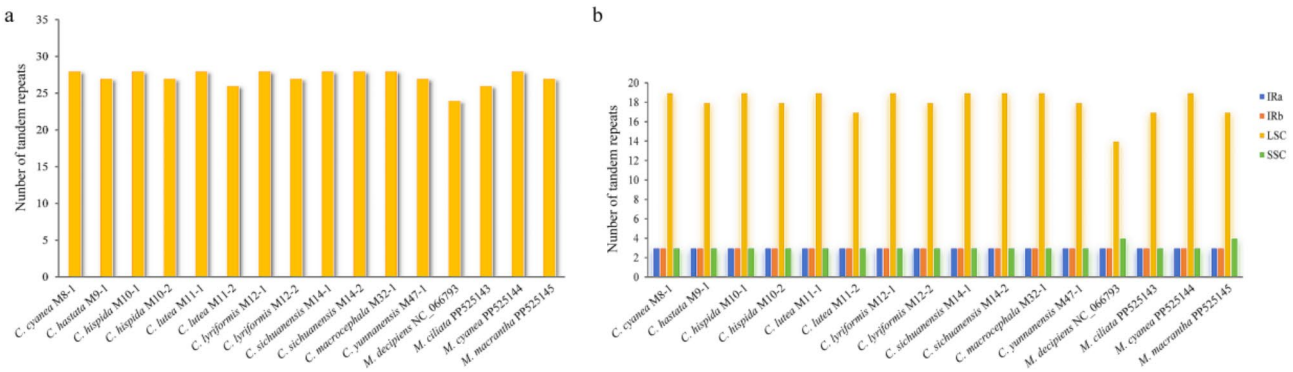


**Fig. 4**. Analysis of tandem repeated sequences in the 16 cp. genomes of *Melanoseris*. (**a**) Number of tandem repeats sequences. (**b**) Distribution of tandem repeats sequences.

showed a topology similar to that derived from the chloroplast genome data (Supplementary Fig. S1). The 13 sequences from the *Melanoseris cyanea* group formed a highly supported clade (BS = 97, PP = 1). However, the phylogenetic relationships among species within the group remain unresolved (BS < 60, PP < 0.6), as none of the four species with two sequences clustered together, mirroring the results obtained in the chloroplast analysis.

## Discussion

### Genome structure and comparative genomic analysis

The chloroplast genome length and GC content of 16 *Melanoseris* samples are very close to the average values for land plants (151 kb and 36.3%)[33]. Typical chloroplast traits, such as uneven GC content distribution and the trans-splicing of the *rps12* gene, were observed, with no unique structural variations identified[34–36]. All 16 genomes contain 132 genes and exhibit identical gene order, with consistent gene locations near the boundaries of different regions, supporting the structural conservation of *Melanoseris* chloroplast genomes. Additionally,

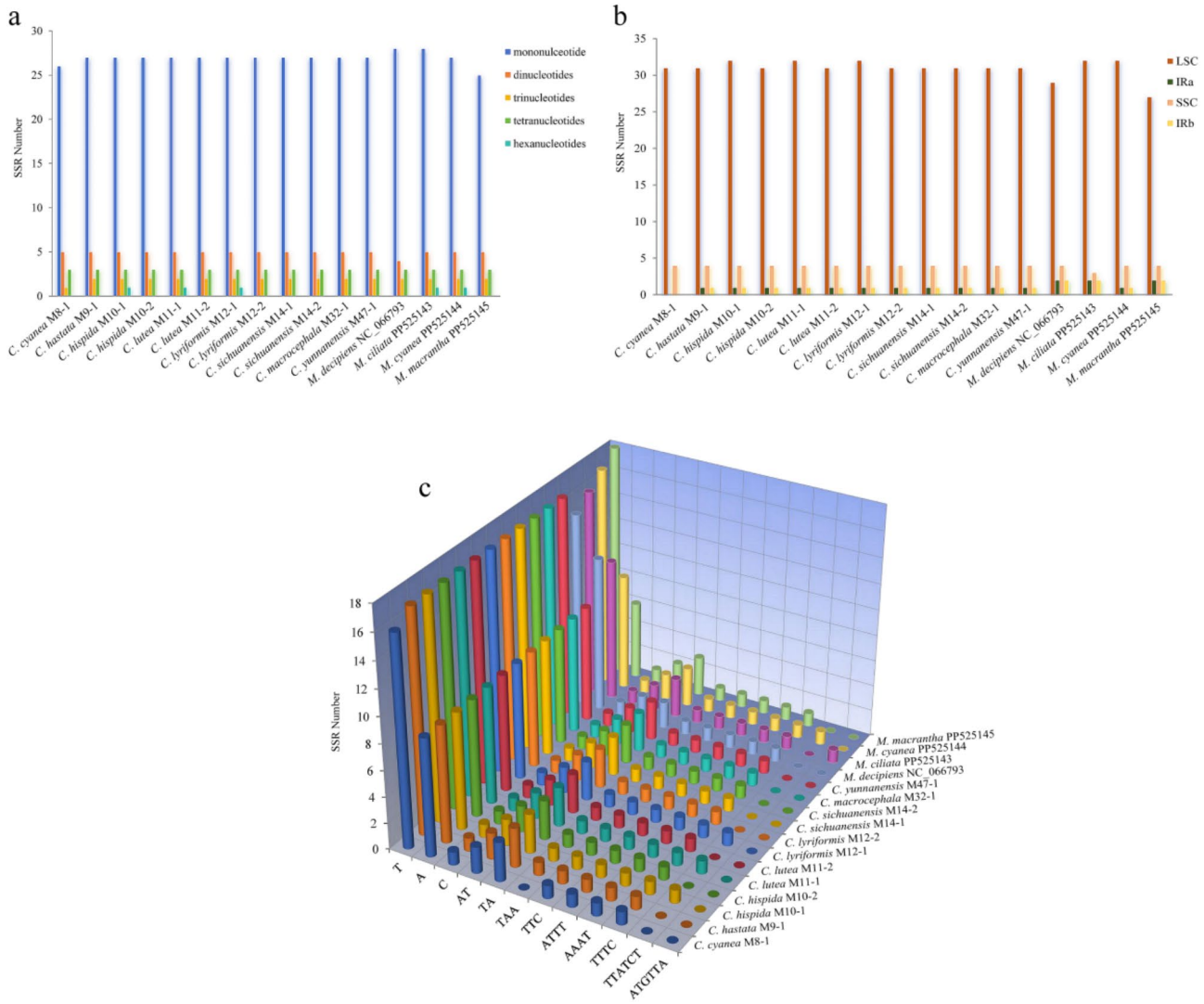


**Fig. 5**. Analysis of SSRs in the 16 cp. genomes of *Melanoseris*. (**a**) Number of different SSR types. (**b**) Distribution of SSRs in the LSC, SSC, and IR regions. (**c**) Frequency of SSRs in different repeat classes.

the IR region across these 16 genomes showed minimal variation (≤ 51 bp), indicating that, unlike in other species[33,37], size variation within the *M. cyanea* group is clearly not driven by changes in the IR region.

The mVISTA and DnaSP analyses also exhibit a high degree of conservation among the chloroplast genomes of the 16 *Melanoseris* samples, with the two single-copy regions showing greater potential for variation than the IR region. This finding is consistent with studies of other plant lineages[38–40], further highlighting the importance of these two single-copy regions in plant evolution. The higher GC content in the IR region may help reduce the extent of variation, highlighting the critical role of GC content in maintaining sequence stability, as reported in previous studies[41,42].

The DnaSP analysis indicates that the 16 chloroplast genomes of *Melanoseris* contain significantly more variable sites and exhibit higher nucleotide polymorphism compared to the 13 genomes of the *M. cyanea* group (Fig. 8 and Supplementary Fig. S2). Additionally, the results from both datasets consistently demonstrate that the *ycf1* gene has the highest π (Pi) value (Fig. 8 and Supplementary Fig. S2), which aligns with numerous previous studies[43–45] and suggests that it may be serve as a valuable DNA barcode for *Melanoseris*. As the second-largest gene in the chloroplast genome, *ycf1* spans approximately 5100 bp in length across the SSC and IR regions in the 16 *Melanoseris* samples. This structural characteristic, which allows the gene to traverse different genomic areas, increases its susceptibility to the evolutionary influences of these distinct regions. Furthermore, the precise functions of *ycf1* remain unclear, and some studies indicate that the selection pressure on this gene is relatively low[43,46,47]. This broad and poorly defined functional characteristic results in reduced functional constraints during evolution, potentially contributing to its status as a highly variable region.

In this study, three types of repeats were analyzed: dispersed repeats, tandem repeats, and simple sequence repeats (SSR). All *Melanoseris* samples contained all three types, with dispersed repeats being the most abundant, followed by SSRs and tandem repeats. The most common SSRs were (A/T)n sequences, consistent with the

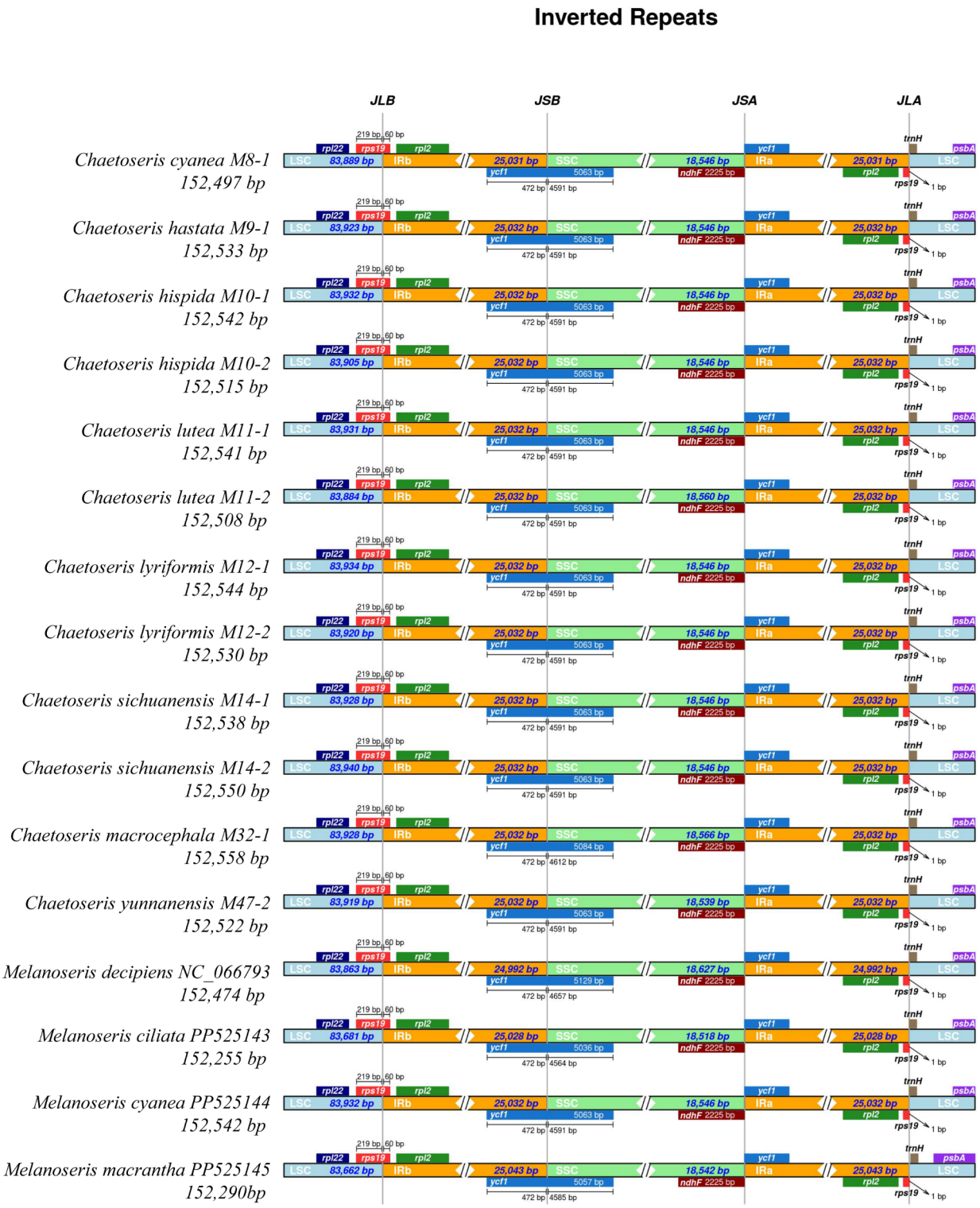


**Fig. 6**. Comparison of LSC, IR, and SSC junction positions among the 16 *Melanoseris* chloroplast genomes. JLB represents the LSC/IRb junction, JSB represents the SSC/IRb junction, JSA represents the SSC/IRa junction, and JLA represents the LSC/IRa junction.

AT bias commonly observed in plant chloroplast genomes[34,48–55]. Our analysis also showed that within the *M. cyanea* group, the numbers of these three types of repeats varied slightly between samples from different counties, even within the same species. This suggests no clear trend in repeat variation, either between or within species. The variability could be due to the diverse and harsh environmental conditions across counties in the Sino-Himalayan region, contributing to genetic differences among local populations.

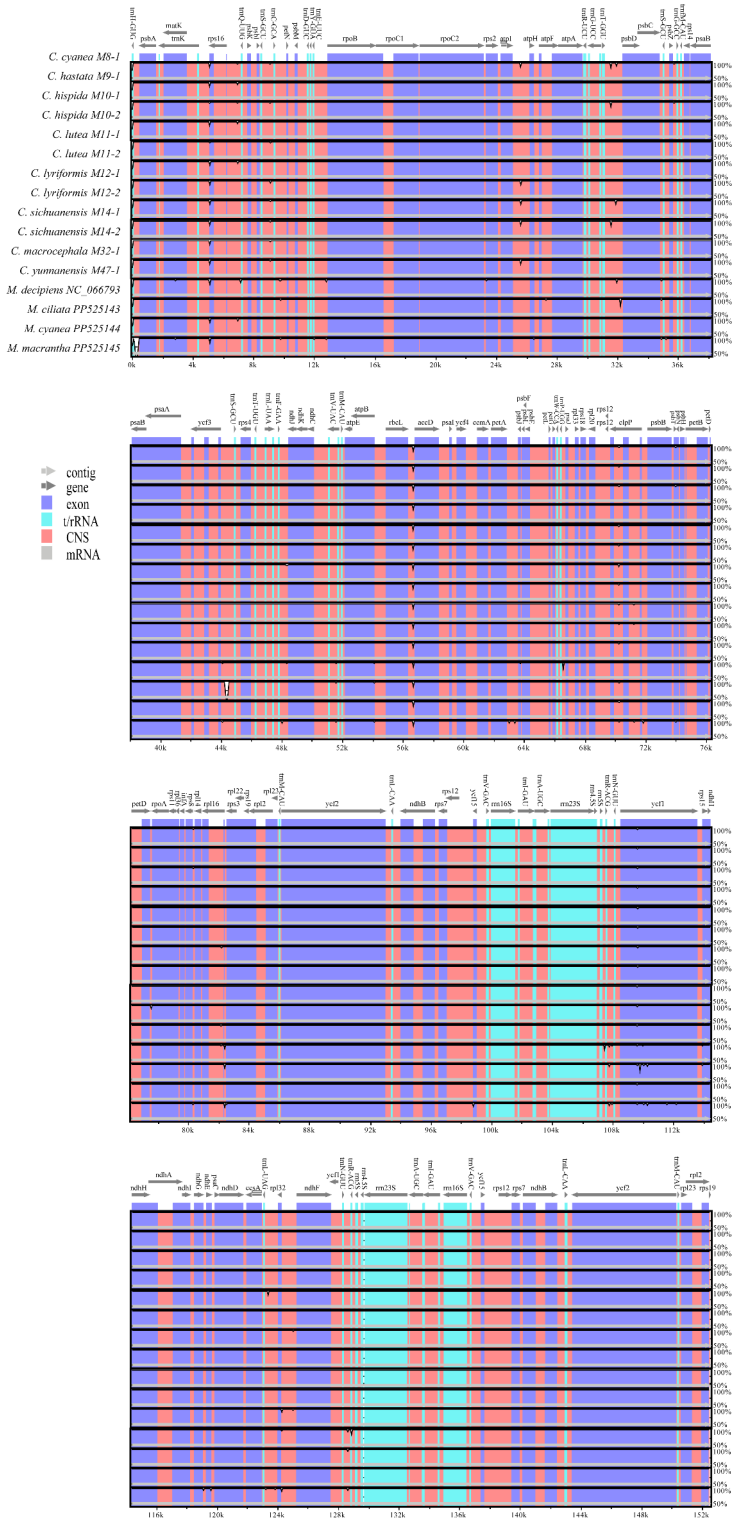


**Fig. 7**. Visualization of the genome alignment of the chloroplast genomes of 16 *Melanoseris* samples, using *Chaetoseris cyanea* M8-1 as a reference, generated by mVISTA. The X-axis represents the coordinates of the chloroplast genome, while the Y-axis represents the different species. Sequence similarity of aligned regions is displayed as horizontal bars, with similarity percentage ranging from 50–100%.

Genomically, all three types of repeats were primarily concentrated in the LSC region of the *Melanoseris* chloroplast genome. Previous studies have shown that repeat types and quantities in chloroplast genomes contribute to the generation of insertions, deletions, and substitution mutations, playing a crucial role in genome evolution and diversity[56,57]. For instance, short tandem repeat sequences may function as cis-regulatory

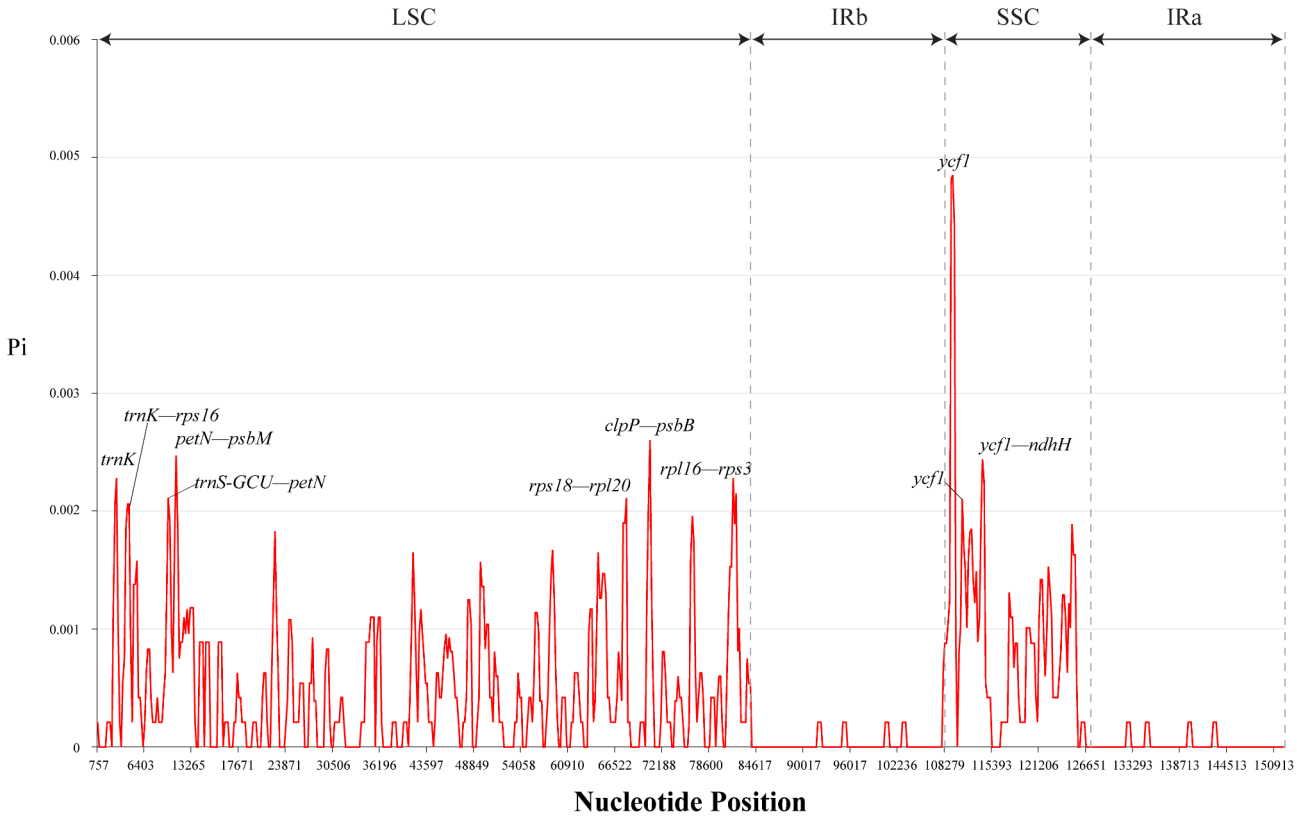


**Fig. 8**. Nucleotide diversity (Pi) values within a 600 bp sliding window of the whole chloroplast genomes of 16 *Melanoseris* samples. Genes with high Pi values (over 0.0002) are annotated.

elements, interacting with transcription factors to regulate gene expression[58]. The LSC region, which contains genes critical for essential processes like photosynthesis and energy metabolism, likely benefits from this repeat distribution pattern. This suggests that the LSC region plays a key role in genomic structure and function, with repeat sequence similarity contributing to functional redundancy and enhancing genome robustness.

## Phylogenetic analysis and taxonomic implications

To investigate the phylogenetic relationships within the *Melanoseris cyanea* group, we applied maximum likelihood (ML) and Bayesian inference (BI) methods using complete chloroplast genome and nrITS data. Both methods and datasets revealed generally consistent topologies (Fig. 9 and Supplementary Fig. S1), with the exception of a minor difference within the AE *Lactuca* clades, aligning closely with Wei's previous study[9]. This suggests that the addition of more *Melanoseris* sequences from Asia did not significantly alter the overall evolutionary relationships. In both the chloroplast genome and nrITS sequence phylogenetic trees, the 13 *M. cyanea* sequences formed highly supported clades (chloroplast genome: BS = 100, PP = 1; nrITS: BS = 97, PP = 1), consistent with earlier findings[6,7]. These results indicate that despite the use of complete chloroplast genomic data, the genetic differences among species within the *M. cyanea* group are minimal, with no clear interspecific boundaries. These results, to some extent, support the hypothesis that the eight species of the *M. cyanea* group may actually be a single species.

In general, the phylogenetic relationships among the eight species within the *Melanoseris cyanea* group did not correspond to the key traits used by Shih to distinguish them (Fig. 9 and Supplementary Fig. S1). For example, Shih[23] distinguished *Chaetoseris lutea* and *C. yunnanensis* from other species based on their yellow or white florets. However, our phylogenetic analysis showed that the two *C. lutea* sequences did not cluster together, nor did the two species themselves. Similarly, Shih's[23] distinctions among *C. cyanea*, *C. hastata*, *C. hispida*, *C. lyriformis*, and *C. sichuanensis* — based on leaf division and hair characteristics (Table S2) — were not supported by our tree. The nine sequences of these species fell into several distinct lineages within the *M. cyanea* group, with no clear species relationships (Fig. 9 and Supplementary Fig. S1). The only exceptions were *C. sichuanensis*, which clustered together in the chloroplast genomic tree (BS = 67, PP = 0.999), and *C. cyanea*, which clustered in the nrITS tree (BS = 63, PP < 0.6).

Field observations revealed that when the main stem of a *Melanoseris* plant is damaged, the leaves that develop on side branches often undergo significant changes, with newly formed leaves exhibiting varying degrees of dissection. This variability in leaf morphology suggests that relying solely on these characteristics for species classification is unreliable and undermines the stability of such classifications. In fact, some scholars have already noted the variation in floret color, leaf shape, and pubescence within the *Melanoseris* genus. For instance, Zhu[11,12] treated *C. lyriformis*, *C. pectiniformis*, and *C. sichuanensis* as synonyms of *C. cyanea*, while


**Fig. 9**. Phylogenetic cladogram of *Melanoseris* and related genera of the Lactucinae subtribe based on complete chloroplast genomes. (**a**) Phylogenetic tree with branch lengths. (**b**) Bayesian phylogram of the subtribe Lactucinae and the main morphological traits of *Melanoseris*, with support values shown as maximum likelihood bootstrap (BS) (below the branches) and Bayesian posterior probability (PP) (above the branches). BS values less than 60 and PP values less than 0.6 are not displayed on the tree. The 12 newly sequenced entries are labeled with species names and numbers, while the remaining entries are labeled with species names and NCBI accession numbers.

Kilian[3] considered *C. lutea* a synonym of *M. yunnanensis* (formerly *C. yunnanensis*), both indicating that the leaf morphology of some species can range from undivided to bipinnate.

Additionally, Zhu[11,12] classified *Chaetoseris lutea* as a variety of *C. cyanea*, suggesting that yellow or white florets might represent a variation of bluish florets. During field surveys, we observed that plants with yellowish-white florets were often found near those with purplish-blue florets, indicating that the yellowish-white form may be a sporadic variation of the purplish-blue one. This further supports the idea that distinguishing *C. lutea* and *C. yunnanensis* from *Melanoseris cyanea* based solely on floret color is not appropriate. Based on these observations, Wang[6] proposed that, except for *C. macrocephala*, the remaining nine species should be considered synonyms of *M. cyanea* (Table S1).

As for *Melanoseris macrocephala*, it has consistently been regarded as a separate species due to the incomplete morphological characteristics observed in the type specimens. However, during fieldwork in Nielamu County, the type locality of *M. macrocephala*, we observed that the capitulum characteristics matched those of *M. macrocephala* (involucre measuring 2 cm in length and 1 cm in diameter, with 40 ligulate florets), while the leaf morphology closely resembled the variations within the *M. cyanea* population. Additionally, analyses of chloroplast genome structure and phylogeny revealed that *M. macrocephala* is part of the *M. cyanea* group,

| Species | Code | Location | Collection no. | GenBank accession | |
|---|---|---|---|---|---|
| | | | | cpDNA | nrITS |
| *Chaetoseris cyanea* (D. Don) Shih | M8-1 | Jilong County, Shigatse Prefecture, Xizang Autonomous Region, China | WZH201708_120 | PQ586435 | PV110879 |
| *Chaetoseris hastata* (Wall. ex DC.) Shih | M9-1 | Jingdong Yi Autonomous County, Pu'er City, Yunnan Province, China | wzh112 | PQ586436 | PV110880 |
| *Chaetoseris hispida* Shih | M10-1 | Huize County, Qujing City, Yunnan Province, China | WZH0059 | PQ586437 | PV110881 |
| *Chaetoseris hispida* Shih | M10-2 | Eryuan County, Dali Bai Autonomous Prefecture, Yunnan Province, China | D640 | PQ586438 | PV110882 |
| *Chaetoseris lutea* (Hand.-Mazz.) Shih | M11-1 | Huize County, Qujing City, Yunnan Province, China | WZH0071 | PQ586439 | PV110883 |
| *Chaetoseris lutea* (Hand.-Mazz.) Shih | M11-2 | Heqing County, Dali Bai Autonomous Prefecture, Yunnan Province, China | wzh104 | PQ586440 | PV110884 |
| *Chaetoseris lyriformis* Shih | M12-1 | Jilong County, Shigatse Prefecture, Xizang Autonomous Region, China | WZH201708_104 | PQ586441 | PV110885 |
| *Chaetoseris lyriformis* Shih | M12-2 | Lang County, Linzhi City, Xizang Autonomous Region, China | 1,209,049 | PQ586442 | PV110886 |
| *Chaetoseris sichuanensis* Shih | M14-1 | Zhanyi County, Qujing City, Yunnan Province, China | 4597 | PQ586443 | PV110887 |
| *Chaetoseris sichuanensis* Shih | M14-2 | Ninglang County, Lijiang City, Yunnan Province, China | WZH20181003 | PQ586444 | PV110888 |
| *Chaetoseris macrocephala* Shih | M32-1 | Nielamu County, Shigatse City, Xizang Autonomous Region, China | WZH201708_131 | PQ586445 | PV110890 |
| *Chaetoseris yunnanensis* Shih | M47-1 | Bomi County, Linzhi City, Xizang Autonomous Region, China | 1,209,094 | PQ586446 | PV110889 |

**Table 3**. Information on the 12 new samples from the *Melanoseris cyanea* group.

showing no significant differences from other species, which indicates a close relationship with other members of the *M. cyanea* group.

Former research[7] indicates that the divergence time of *M. cyanea* and its sister clades is less than 2 mya, suggesting a relatively recent origin for the *M. cyanea* group, which may have undergone rapid radiation evolution. The co-occurrence of different species within their distribution areas, along with the emergence of various new hybrid forms[6], reveals that there is still extensive gene flow between these species. Based on extensive field observations, studies of type specimens, chloroplast genome characteristics, and phylogenetic analysis, we recommend merging the 11 species within the *M. cyanea* group into a single species, *M. cyanea* Edgew.

This study offers valuable insights into the *M. cyanea* group; however, it has notable limitations. The narrow focus on this specific group restricts our ability to draw comprehensive conclusions about the entire *Melanoseris* genus. Including more species could lead to variations in the genomic characteristics observed. Additionally, the geographic bias toward samples predominantly from the Sino-Himalayan region may obscure the full genetic diversity and evolutionary history of *Melanoseris* species in other regions, such as Africa and Central Asia.

While phylogenetic analyses based on chloroplast genomes and nrITS sequences have been conducted, the evolutionary relationships within the *M. cyanea* group remain unclear due to the limited number of informative loci examined. Furthermore, we must exercise caution when interpreting the detailed phylogenetic relationships among species, as the neutral theory[59] and the molecular clock hypothesis[60], which have historically guided phylogenetic methodologies, have been challenged in numerous studies[58,61–73]. To achieve a more comprehensive and reliable understanding of the genetic diversity and evolutionary processes within *Melanoseris*, future research should expand sample sizes by incorporating additional species and diverse populations. Furthermore, emphasizing nuclear genomic data and employing alternative methodologies will be crucial for unraveling the intricate relationships within this genus.

## Conclusion

This study presents the first systematic report on the chloroplast genomes of plants within the *Melanoseris* genus, offering an in-depth exploration of their structural variations, comparative genomic analyses, and phylogenetic relationships. Through a comprehensive examination of the chloroplast genomes of various species within the *M. cyanea* group, the results revealed a high sequence similarity with very few variation sites. Based on a detailed comparative analysis of genomic structure and morphological characteristics, this study supports the modern taxonomic classification that groups these eight species as a single species. These findings deepen our understanding of the genetic diversity of the *M. cyanea* group, providing crucial data for its taxonomic and phylogenetic studies. They also lay a foundation for future research on the genetic composition and evolutionary processes of the *Melanoseris* genus.

## Materials and methods

### Materials collection, DNA extraction, and sequencing

The 12 plant samples of the *Melanoseris cyanea* group in this study were fresh leaves collected from the wild and preserved using silica gel (Table 3). These specimens were identified by Dr. Zehuan Wang at the Guizhou University of Traditional Chinese Medicine, based on the identification keys and morphological descriptions in the Flora Reipublicae Popularis Sinicae (FRPS, Flora of China, Chinese edition). The voucher specimens have been deposited in the herbarium of the Kunming Institute of Botany, Chinese Academy of Sciences (KUN). All necessary permits for accessing the protected areas were obtained from local governments and National Park Services. This research was conducted in full compliance with the relevant laws and regulations of China. The experimental materials were sent to the Molecular Biology Experiment Center, Germplasm Bank of Wild Species, Kunming, China for sequencing. DNA extraction, library preparation, and deep genome sequencing were conducted, with the library sequenced on the DNBSEQ-T7 platform, yielding 150 bp paired-end reads and approximately 20 Gb of raw reads.

## Chloroplast genome assembly and annotation

Before assembly, fastp v0.23.2[74] was used to remove sequences containing adapters and low-quality reads. Assembly was performed using GetOrganelle v1.7.5[75], with the published chloroplast genome of *Melanoseris decipiens* (NC_066793) as the reference. Based on this reference genome, the starting position and orientation of the chloroplast assembly sequence, as well as the potential partition structure of the chloroplast (LSC/IR/SSC), were determined to obtain the final chloroplast genome sequence. The sequence was annotated using CpGAVAS 2[76] and GeSeq[77], with manual correction of the positions of start and stop codons and the boundaries between exons and introns using Geneious Prime 2023.0.4[78]. Finally, the physical map of the chloroplast genome was created using OrganellarGenomeDRAW (https://chlorobox.mpimp-golm.mpg.de/OGDraw.html)[79].

## Analysis of repeat sequences

This study used REPuter (https://bibiserv.cebitec.uni-bielefeld.de/reputer)[80] to identify direct (forward), inverted (palindromic), complementary, and reverse repeat elements in the chloroplast genome, with a maximum repeat length of 5000 bp, a Hamming distance of 3, and a minimum repeat length of 25 bp. Additionally, MISA (https://webblast.ipk-gatersleben.de/misa/index.php)[81] was employed to detect simple sequence repeats (SSRs) in the complete chloroplast genome. The detection thresholds for mononucleotide, dinucleotide, trinucleotide, tetranucleotide, pentanucleotide, and hexanucleotide SSRs were set to 10, 5, 4, 3, 3, and 3, respectively. Tandem repeats were identified using the TRF software[82], with repeat unit sizes greater than or equal to 7.

## Comparative genomic analyses

The expansion and contraction of the IR regions in 16 chloroplast genome sequences of *Melanoseris* (comprising 12 newly sequenced genomes from the *M. cyanea* group and 4 downloaded *Melanoseris* genomes) were investigated using the online tool IRscope (https://irscope.shinyapps.io/irapp/)[83]. Comparative analysis of these 16 chloroplast genomes was performed using the Shuffle-LAGAN mode in the mVISTA program (https://genome.lbl.gov/vista/index.shtml)[29], with mVISTA plots generated to display sequence conservation profiles. Additionally, after sequence alignment using MAFFT v7.471[84], sliding window analysis was conducted using DnaSP v6[30] to determine nucleotide diversity (Pi) across the entire chloroplast genome. The sliding window length was set to 600 bp, with a step size of 200 bp.

## Phylogenetic analysis

To investigate the phylogenetic relationships within the *Melanoseris cyanea* group, we conducted a phylogenetic reconstruction using complete chloroplast genome data and nrITS regions from the entire Lactucinae subtribe. To minimize the impact of missing taxa and different outgroups on phylogenetic topologies, we selected the outgroups used in the latest phylogenomic research[9] and performed representative sampling of all 9 genera within Lactucinae, excluding *Lihengia* Y.S.Chen & R.Ke. Given the close relationships among the AE *Lactuca* clade, core *Lactuca* clade, and *Melanoseris* clade in previous studies[7–9], we sampled slightly more from these clades and downloaded all four available plastomes for *Melanoseris*.

We obtained complete chloroplast genome sequences for 47 ingroup and 3 outgroup taxa from the National Center for Biotechnology Information (NCBI), as well as nrITS sequences for 45 ingroup and 3 outgroup taxa. To reconstruct the phylogenetic relationships of *Melanoseris* within the Lactucinae subtribe, we utilized these sequences, along with 12 newly generated chloroplast genomes specific to the *M. cyanea* group and 12 new nrITS sequences extracted from deep genome sequencing data using Phyloherb v1.1.3[85], referencing *M. leptantha* (KF739607), *M. cyanea* (KF485623), and *Chaetoseris grandiflora* (HQ436191).

Before constructing the phylogenetic trees, all inversions were manually corrected using Geneious Prime 2023.0.4[79] to ensure consistent gene and nucleotide sequences. Phylogenetic analysis was conducted based on the complete chloroplast genome sequences. The sequences were aligned using MAFFT in Phylosuite v1.2.3[31] with default parameters. Phylogenetic relationships were reconstructed using both Maximum Likelihood (ML) and Bayesian Inference (BI) methods. The appropriate models were selected using ModelFinder in Phylosuite v1.2.3[31], based on the Akaike Information Criterion (AIC) with default parameters. The ML tree was constructed using IQ-TREE in Phylosuite v1.2.3[31] employing the selected IQ-TREE model and performing 1,000 standard bootstrap replicates to assess branch support. The BI tree was generated using MrBayes in Phylosuite v1.2.3[31], running 5,000,000 generations with default parameters and the selected MrBayes model. Finally, both trees were visualized and edited using FigTree v.1.4.4[86]. For the phylogenetic analysis based on nrITS sequences, the sequences were aligned using MUSCLE[32], while the other methods were consistent with those described above.

## Data availability

The data supporting the findings of this study are openly available in GenBank (NCBI) at https://www.ncbi.nlm.nih.gov, with accession numbers PQ586435–PQ586446 for chloroplast genomes and PV110879–PV110890 for nrITS sequences.

Received: 14 December 2024; Accepted: 20 March 2025
Published online: 27 March 2025

## References

1. Jacquemont, V. Voyage dens I' inde par Jacquemont pendant les Annes 1828a, 1832. In *Botanique, plantae rariores, quae in India orientali collegit. V Jacquemont*, **4**, 363 (Paris, 1843).
2. Edgeworth, M. P. Descriptions of some unpublished species of plants from North-Western India. *Trans. Linn. Soc. Lond.* **20**, 23–91 (1846).

3. Shih, C. & Kilian, N. *Melanoseris*. In *Flora of China, vols. 20–21* (ed. Wu, Z. Y., Raven, P. H., Hong, D. Y.), 217–226 (Science Press & St. Louis: Missouri Botanical Garden, 2011).
4. Kilian, N., Gemeinholzer, B. & Lack, H. W. Tribe Cichorieae. In *Systematics, evolution and biogeography of the Compositae* (ed. Funk, V. A., Susanna, A., Stuessy, T., Bayer, R.) 343–383 (IAPT, 2009).
5. Shih, C. On circumscription of the genus *Cicerbita* Wallr., and two new genera of compositae from Sino-Himalaya region. *Acta Phytotax Sin*. **29**, 394–417 (1991).
6. Wang, Z. H., Peng, H. & Kilian, N. Molecular phylogeny of the *Lactuca* alliance (Cichorieae subtribe lactucinae, Asteraceae) with focus on their Chinese center of diversity detects potential events of reticulation and Chloroplast capture. *PloS ONE*. **8**, e82692 (2013).
7. Kilian, N. et al. Sub-Paratethyan centred middle to late miocene diversification of the lactucinae (Cichorieae, Compositae) inferred from phylogeny, divergence-date estimates and biogeography. *Taxon* **66**, 675–703 (2017).
8. Wei, Z., Zhu, S. X., Van den Berg, R. G., Bakker, F. T. & Schranz, M. E. Phylogenetic relationships within *Lactuca* L.(Asteraceae), including African species, based on Chloroplast DNA sequence comparisons. *Genet. Resour. Crop Evol.* **64** (1), 55–71 (2017).
9. Ran, C. et al. Plastome-based phylogeny and biogeography of *Lactuca* L.(Asteraceae) support revised lettuce gene pool categories. *Front. Plant. Sci.* **13**, 978417 (2022).
10. Zhang, J. W. et al. Cichorieae): A new genus endemic to the Himalaya-Hengduan Mountains. *Taxon* **60**, 1678–1684 (2011).
11. Zhu, S. X. *Systematics of Chaetoseris and Stenoseris (Compositae–Lactuceae)*, 176 (Institute of Botany, the Chinese Academy of Science, 2004).
12. Zhu, S. X., Qin, H. N. & Shih, C. A new synonym in *Chaetoseris* (Compositae). *Acta Phytotax Sin*. **42**, 268–270 (2004).
13. Zhu, S. X., Qin, H. N. & Shih, C. Achene wall anatomy and surface sculpturing of *Lactuca* L. and related genera (Compositae: Lactuceae) with notes on their systematic significance. *J. Integr. Plant. Biol.* **48** (4), 390–399 (2006).
14. Wang, H., Wortley, A. H. & Blackmore, S. Pollen morphology of crepidinae and lactucinae (Asteraceae: Cichorieae) and its systematic significance. *Grana* **48** (3), 160–178 (2009).
15. Deng, T., Meng, Y., Sun, H. & Long, Z. Chromosome counts and karyotypes in chaetoseris and Stenoseris (Asteraceae–Cichorieae) from the Hengduan mountains of SW China. *J. Syst. Evol.* **49** (4), 339–346 (2011).
16. Wang, Z. H., Kilian, N. & Peng, H. Notes on *Melanoseris Lessertiana* (Lactucinae, Asteraceae) and morphologically allied species in the Pan-Himalayan region. *Plant. Divers. Resour.* **37**, 401–406 (2015).
17. Wang, Z. H., Kilian, N., Chen, Y. P. & Peng, H. *Sinoseris* (Crepidinae, cichorieae, Asteraceae), a new genus of three species endemic to China, one of them new to science. *Willdenowia* **50** (1), 91–110 (2020).
18. Abid, R., Qaiser, M., Ghafoor, A. & Genus *Melanoseris* Asteraceae (V)-Cichorieae. *Flora of Pakistan* **222**, 178–202 (2017).
19. Ghafoor, A., Qaiser, M. & Abid, R. Notes on cichoreae (Asteraceae) from Pakistan and Kashmir: some additions and corrections. *Pak J. Bot.* **49** (4), 1324–1325 (2017).
20. Yin, Z. J. At El. A new species of *Melanoseris* (Lactucinae, cichorieae, Asteraceae) from SW Xizang, China, based on morphological and molecular data. *Phytotaxa* **357** (3), 189–197 (2018).
21. Zhong, Q. Q., Wang, Z. H., Xu, J. J. & Sun, Q. W. *Melanoseris kangdingensis* (Lactucinae, cichorieae, Asteraceae), a new species reported from western Sichuan, China. *Phytokeys* **236**, 29–37 (2023).
22. Xu, J. J. at el. *Melanoseris penghuana* (Lactucinae, Cichorieae, Asteraceae), a new species from North-central Yunnan, China. *PhytoKeys* **238**, 1–10 (2024).
23. Shih, C. *Flora Reipublicae Popularis Sinicae, Tomus* , Vol. 80, 276–279 (Science Press, 1997).
24. Mamgain, S. K. & Rao, R. R. Cichorieae. In *Flora of India*, 226–345 (Botanical Survey of India, 1995).
25. Mamgain, S. K. Diversity, ecology and distribution of Indian lactuceae (Asteraceae). *Taiwania* **43** (2), 155–163 (1998).
26. Zhao, D. N., Ren, Y. & Zhang, J. Q. Conservation and innovation: plastome evolution during rapid radiation of Rhodiola on the qinghai-Xizangan plateau. *Mol. Phylogenet Evol.* **144**, 106713 (2020).
27. Su, C. at el. Chloroplast phylogenomics and character evolution of eastern Asian *Astragalus* (Leguminosae): Tackling the phylogenetic structure of the largest genus of flowering plants in Asia. *Mol. Phylogenet. Evol*. 156, 107025; (2021).
28. Wen, J. at el. Backbone phylogeny and evolution of Apioideae (Apiaceae): New insights from phylogenomic analyses of plastome data. *Mol. Phylogenet. Evol.* **161**, 107183 (2021).
29. Frazer, K. A., Pachter, L., Poliakov, A. & Rubin, E. M. Dubchak, I. VISTA: computational tools for comparative genomics. *Nucleic Acids Res.* **32** (suppl_2), W273–W279 (2004).
30. Rozas, J. et al. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* **34**, 3299–3302 (2017).
31. Xiang, C. Y. et al. Using phylosuite for molecular phylogeny and tree-based analyses. *Imeta* **1**, e87 (2023).
32. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32** (5), 1792–1797 (2004).
33. Guo, Y. Y., Yang, J. X., Li, H. K. & Zhao, H. S. Chloroplast genomes of two species of *Cypripedium*: expanded genome size and proliferation of AT-biased repeat sequences. *Front. Plant. Sci.* **12**, 142 (2021).
34. Wang, X. M., Zhou, T., Bai, G. Q. & Zhao, Y. M. Complete Chloroplast genome sequence of *Fagopyrum dibotrys*: genome features, comparative analysis and phylogenetic relationships. *Sci. Rep.* **8**, 12379 (2018).
35. Thode, V. A. & Lohmann, L. G. Comparative Chloroplast genomics at low taxonomic levels: a case study using *Amphilophium* (Bignonieae, Bignoniaceae). *Front. Plant. Sci.* **10**, 796 (2019).
36. Lu, G. Y. et al. An integrated study of violae herba (*Viola philippica*) and five adulterants by morphology, chemical compositions and Chloroplast genomes: insights into its certified plant origin. *Chin. Med.* **17** (1), 32 (2022).
37. Hansen, D. R. et al. Phylogenetic and evolutionary implications of complete Chloroplast genome sequences of four early-diverging angiosperms: *Buxus* (Buxaceae), *Chloranthus* (Chloranthaceae), *Dioscorea* (Dioscoreaceae), and *Illicium*. *Mol. Biol. Evol.* **45**, 547–563 (2007).
38. Gu, C., Tembrock, L. R., Johnson, N. G., Simmons, M. P. & Wu, Z. The complete plastid genome of *Lagerstroemia Fauriei* and loss of *rpl2* intron from *Lagerstroemia* (Lythraceae). *PloS One*. **11**, e0150752 (2016).
39. Xu, C. et al. Comparative analysis of six *Lagerstroemia* complete Chloroplast genomes. *Front. Plant. Sci.* **8**, 15 (2017).
40. Alzahrani, D. A., Yaradua, S. S., Albokhari, E. J. & Abba, A. Complete Chloroplast genome sequence of *Barleria prionitis*, comparative Chloroplast genomics and phylogenetic relationships among acanthoideae. *BMC Genom.* **21**, 393 (2020).
41. Necsulea, A. & Lobry, J. R. A new method for assessing the effect of replication on DNA base composition asymmetry. *Mol. Biol. Evol.* **24**, 2169–2179 (2007).
42. Kim, H. T., Pak, J. H. & Kim, U. S. The complete Chloroplast genome sequence of *Crepidiastrum lanceolatum* (Asteraceae). *Mitochondrial DNA Part. B*. **4** (1), 1404–1405 (2019).
43. Dong, W. P. *Chloroplast Genome Evolution of Calycanthaceae and Highiy Variable Chloroplast Markers Development of Flowering Plants*, 35 (Northeast Forestry University, 2012).
44. Li, C. C. et al. Insights into Chloroplast genome evolution in rutaceae through population genomics. *Hortic. Adv.* **2** (1), 1–14 (2024).
45. Jia, M. et al. The pan-plastome of *Hemerocallis citrina* reveals new insights into the genetic diversity and cultivation history of an economically important food plant. *BMC Plant. Biol.* **24** (1), 44 (2024).
46. Barnard-Kubow, K. B., Sloan, D. B. & Galloway, L. F. Correlation between sequence divergence and polymorphism reveals similar evolutionary mechanisms acting across multiple timescales in a rapidly evolving plastid genome. *BMC Evol. Biol.* **14**, 1–10 (2014).

47. Zhao, Y. et al. The complete Chloroplast genome provides insight into the evolution and polymorphism of *Panax ginseng*. *Front. Plant. Sci.* **5**, 696 (2015).
48. Huang, H., Shi, C., Liu, Y., Mao, S. Y. & Gao, L. Z. Thirteen *Camellia* Chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships. *BMC Evol. Biol.* **14**, 151–151 (2014).
49. Li, B., Lin, F. R., Huang, P., Guo, W. Y. & Zheng, Y. Q. Development of nuclear SSR and Chloroplast genome markers in diverse *Liriodendron Chinense* germplasm based on low-coverage whole genome sequencing. *Biol. Res.* **53**, 1–12 (2020).
50. Peng, J. Y. et al. Newly reported Chloroplast genome of *Sinosenecio Albonervius* Y.Liu & Q.E.Yang and comparative analyses with other *Sinosenecio* species. *BMC Genom.* **23** (1), 639 (2022).
51. Li, J. et al. Comparative analysis of the Chloroplast genomes of eight *Piper* species and insights into the utilization of structural variation in phylogenetic analysis. *Front. Genet.* **13**, 925252 (2022).
52. Wu, L. W. et al. Comparative and phylogenetic analyses of the Chloroplast genomes of species of Paeoniaceae. *Sci. Rep.* **11** (1), 14643 (2021).
53. Li, Y. T. et al. Comparative analyses of *Euonymus* Chloroplast genomes: genetic structure, screening for loci with suitable polymorphism, positive selection genes, and phylogenetic relationships within celastrineae. *Front. Plant. Sci.* **11**, 593984 (2021).
54. Long, L. X. et al. Complete Chloroplast genomes and comparative analysis of *Ligustrum* species. *Sci. Rep.* **13** (1), 212 (2023).
55. Xu, J., Liu, C., Song, Y. & Li, M. F. Comparative analysis of the Chloroplast genome for four *Pennisetum* species: molecular structure and phylogenetic relationships. *Front. Genet.* **12**, 687844 (2021).
56. McDonald, M. J., Wang, W. C., Huang, H. D. & Leu, J. Y. Clusters of nucleotide substitutions and insertion/deletion mutations are associated with repeat sequences. *PLoS Biol.* **9**, e1000622 (2011).
57. Kawata, M., Harada, T., Shimamoto, Y., Oono, K. & Takaiwa, F. Short inverted repeats function as hotspots of intermolecular recombination giving rise to oligomers of deleted plastid DNAs (ptDNAs). *Curr. Genet.* **31**, 179–184 (1997).
58. Horton, C. A. et al. Short tandem repeats bind transcription factors to tune eukaryotic gene expression. *Science* **381** (6664), eadd1250 (2023).
59. Kimura, M. Evolutionary rate at the molecular level. *Nature* **217** (5129), 624–626 (1968).
60. Zuckerkandl, E. & Pauling, L. Molecular disease, evolution, and genetic heterogeneity. In *Horizons in Biochemistry* (ed. Kasha, M., Pullman, B.), 97–166 (Academic Press, 1962).
61. Huang, S. The genetic equidistance result of molecular evolution is independent of mutation rates. *Nat. Precedings* 1–1 (2008).
62. Huang, S. Inverse relationship between genetic diversity and epigenetic complexity. *Nat. Precedings*. **1**, 1–1 (2009).
63. Huang, S. Primate phylogeny: molecular evidence for a pongid clade excluding humans and a prosimian clade containing tarsiers. *Sci. China Life Sci.* **55**, 709–725 (2012).
64. Pulquerio, M. J. & Nichols, R. A. Dates from the molecular clock: how wrong can we be? *Trends Ecol. Evol.* **22**, 180–184 (2007).
65. Laird, C. D., McConaughy, B. L. & McCarthy, B. J. Rate of fixation of nucleotide substitutions in evolution. *Nature* **224**, 149–154 (1969).
66. Jukes, T. H. & Holmquist, R. Evolutionary clock: nonconstancy of rate in different species. *Science* **177**, 530–532 (1972).
67. Goodman, M., Moore, G. W., Barnabas, J. & Matsuda, G. The phylogeny of human globin genes investigated by the maximum parsimony method. *J. Mol. Evol.* **3**, 1–48 (1974).
68. Langley, C. H. & Fitch, W. M. An examination of the constancy of the rate of molecular evolution. *J. Mol. Evol.* **3**, 161–177 (1974).
69. Gago, S., Elena, S. F., Flores, R. & Sanjuán, R. Extremely high mutation rate of a hammerhead viroid. *Science* **323**, 1308 (2009).
70. Hu, T. B. et al. The genetic equidistance result: misreading by the molecular clock and neutral theory and reinterpretation nearly half of a century later. *Sci. China Life Sci.* **56**, 254–261 (2013).
71. Huang, S. New thoughts on an old riddle: what determines genetic diversity within and between species? *Genomics* **108** (1), 3–10 (2016).
72. Lake, N. J. et al. Quantifying constraint in the human mitochondrial genome. *Nature*, 1–8 (2024).
73. Lynch, M., Wei, W., Ye, Z. & Pfrender, M. The genome-wide signature of short-term Temporal selection. *PNAS* **121** (28), e2307107121 (2024).
74. Chen, S. F., Zhou, Y. Q., Chen, Y. R. & Gu, J. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
75. Jin, J. J. at el. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol.* **21**, 241 (2020).
76. Shi, L. C. at el. CPGAVAS2, an integrated plastome sequence annotator and analyzer. *Nucleic Acids Res.* **47**(W1), W65–W73 (2019).
77. Tillich, M. at el. GeSeq - versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* **45**, W6–11 (2017).
78. Kearse, M. et al. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28** (12), 1647–1649 (2012).
79. Lohse, M., Drechsel, O. & Bock, R. OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr. Genet.* **52**, 267–274 (2007).
80. Kurtz, S. & Schleiermacher, C. REPuter: fast computation of maximal repeats in complete genomes. *Bioinformatics (Oxford England)*. **15** (5), 426–427 (1999).
81. Beier, S., Thiel, T., Münch, T., Scholz, U. & Mascher, M. MISA-web: a web server for microsatellite prediction. *Bioinformatics* **33** (16), 2583–2585 (2017).
82. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27** (2), 573–580 (1999).
83. Amiryousefi, A., Hyvönen, J. & Poczai, P. IRscope: an online program to visualize the junction sites of Chloroplast genomes. *Bioinformatics* **34** (17), 3030–3031 (2018).
84. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30** (4), 772–780 (2013).
85. Cai, L., Zhang, H., Davis, C. C. & PhyloHerb A high-throughput phylogenomic pipeline for processing genome skimming data. *Appl. Plant. Sci.* **10**(3), e11475 (2022).
86. Rambaut, A. *FigTree, version 1.4.4*.. http://tree.bio.ed.ac.uk/software/figtree/ (Accessed 31 August 2024) (2018).


## Acknowledgements

We sincerely thank the Molecular Biology Experiment Center, Germplasm Bank of Wild Species, Kunming, China, for their invaluable support and resources, which greatly contributed to the success of this research. We would also like to express our gratitude to the editors and two anonymous reviewers for their valuable comments, which significantly helped improve this paper. This study was supported by the National Natural Science Foundation of China (grant no. 32260051).

## Author contributions

Z.H.W. and Q.Q.Z. conceived and designed the research. Z.H.W., Q.Q.Z and J.J.X. collected the materials. Z.H.W., Q.Q.Z. and J.J.X. performed bioinformatic analyses. Q.Q.Z. and L.Y. prepared all the figures and tables. Z.H.W. and Q.Q.Z. wrote the manuscript. Z.H.W. and Q.W.S. revised the manuscript. All authors contributed to

the article and approved the submitted version. All authors reviewed the manuscript.

## Declarations

## Competing interests
The authors declare no competing interests.

## Additional information
**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-95325-y.

**Correspondence** and requests for materials should be addressed to Z.W.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.