

RESEARCH

Open Access



# Comparative phylogenomic study of East Asian endemic genus, *Corchoropsis* Siebold & Zucc. (Malvaceae s.l.), based on complete plastome sequences

Joonhyung Jung<sup>1</sup>, Tao Deng<sup>2</sup>, Yu Gyeom Kim<sup>1</sup>, Changkyun Kim<sup>3</sup>, Hang Sun<sup>2\*</sup> and Joo-Hwan Kim<sup>1\*</sup>

## Abstract

**Background** Endemic plants are key to understanding the evolutionary history and enhancing biodiversity within their unique regions, while also offering significant economic potential. The East Asian endemic genus *Corchoropsis* Siebold & Zucc., classified within the subfamily Dombeyoideae of Malvaceae s.l., comprises three species.

**Results** This study characterizes the complete plastid genomes (plastomes) of *C. crenata* var. *crenata* Siebold & Zucc. and *C. crenata* var. *hupehensis* Pamp., which range from 160,093 to 160,724 bp. These genomes contain 78 plastid protein-coding genes, 30 tRNA, and four rRNA, except for one pseudogene, *infA*. A total of 316 molecular diagnostic characters (MDCs) specific to *Corchoropsis* were identified. In addition, 91 to 92 simple sequence repeats (SSRs) in *C. crenata* var. *crenata* and 75 in *C. crenata* var. *hupehensis* were found. Moreover, 49 long repeats were identified in both the Chinese *C. crenata* var. *crenata* and *C. crenata* var. *hupehensis*, while 52 were found in the South Korean *C. crenata* var. *crenata*. Our phylogenetic analyses, based on 78 plastid protein-coding genes, reveal nine subfamilies within the Malvaceae s.l. with high support values and confirm *Corchoropsis* as a member of Dombeyoideae. Molecular dating suggests that *Corchoropsis* originated in the Oligocene, and diverged during the Miocene, influenced by the climate shift at the Eocene–Oligocene boundary.

**Conclusions** The research explores the evolutionary relationships between nine subfamilies within the Malvaceae s.l. family, specifically identifying the position of the *Corchoropsis* in the Dombeyoideae. Utilizing plastome sequences and fossil data, the study establishes that *Corchoropsis* first appeared during the Eocene and experienced further evolutionary divergence during the Miocene, paralleling the evolutionary patterns observed in other East Asian endemic species.

**Keywords** *Corchoropsis*, Plastid genome (Plastome), Phylogenomic study, Molecular dating

\*Correspondence:

Hang Sun  
sunhang@mail.kib.ac.cn

Joo-Hwan Kim  
kimjh2009@gachon.ac.kr

<sup>1</sup>Department of Life Sciences, Gachon University, 1342, Seongnam-daero, Seongnam-si 13120, Republic of Korea

<sup>2</sup>State Key Laboratory of Plant Diversity and Specialty Crops, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, China

<sup>3</sup>Department of Island and Coast Biodiversity, Honam National Institute of Biological Resources, 99, Gohadoan-gil, Mokpo-si 58762, Republic of Korea



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

## Background

Biodiversity is an indicator that determines the range of evolutionary and ecological adaptations of species to specific environments [1]. Endemic plants, which are taxa that exclusively grow in specific regions and enhance the biodiversity of those regions, serve as crucial models for studying the evolutionary history of plants. Reports indicate that East Asia, including China, Korea, and Japan, is home to about 600 genera and more than 18,000 species of endemic plants across 31 families [2]. This number is notably higher than that of endemic plants found in the Northern Hemisphere, such as in North America and Europe [3, 4]. The rich diversity of endemic plants in East Asia is attributed to a combination of factors, including varied climatic and geographical changes during the Cenozoic era, a diversity of habitats, and the presence of numerous refugia during the ice ages [5, 6].

The genus *Corchoropsis* Siebold & Zucc., endemic to East Asia, comprises three annual species within the Malvaceae s.l. It is characterized by simple, alternately arranged leaves, solitary bisexual flowers with five yellow or white petals, 10–15 stamens, the presence of staminodes, a single pistil, and linear fruits [7, 8]. Originally, *Corchoropsis* was classified within the Tiliaceae by Siebold and Zuccarini in 1843. However, Takeda [9] reclassified it into the tribe Dombeyeae of the Sterculiaceae, based on its floral characteristics. In the current classification, *Corchoropsis* is placed within the subfamily Dombeyoideae of the Malvaceae s.l. based on its morphological traits [7]. In the initial molecular phylogenetic analyses based on three plastid regions, results indicated that *Corchoropsis* belongs to the Dombeyoideae and was well-resolved [10]. More recently, Dorr and Wurdack [11] proposed that the morphological similarities and the number of chromosomes shared between the two genera, *Corchoropsis* and *Paradombeya* Stapf, justify treating *Paradombeya* as a synonym of *Corchoropsis*. However, genomic information for *Corchoropsis* is not available, and comparative phylogenomic studies on *Corchoropsis* are lacking.

Of the three types of plant genomes, researchers have prioritised the usage of the plastome sequences due to them being highly conserved and their relatively small size [12]. Given the crucial role of chloroplasts in plant photosynthesis, the plastome has been extensively

utilized in recent studies for developing molecular markers, conducting phylogenetic analyses, estimating divergence times, and performing biogeographical analyses [13–15]. In this study, we completed the plastomes of *Corchoropsis* collected from China and South Korea. We compared their genomic structures and reconstructed phylogenomic relationships among related species. Finally, we estimated the divergence times of *Corchoropsis*, which will aid in understanding the evolutionary patterns of the Malvaceae s.l. and provide essential information for future research on East Asian endemic taxa.

## Methods

### Taxon sampling, DNA extraction, and plastome assembly

We collected whole individuals of *C. crenata* var. *crenata* Siebold & Zucc. in China (N 31°35'51" E 110°52'15", 857m) and South Korea (N 33°29'47" E 126°39'46", 157m), as well as *C. crenata* var. *hupehensis* Pamp. from South Korea (N 37°25'30" E 127°04'10", 114m) in the field. Plant collection did not require any specific permits. After collecting the samples, we prepared voucher specimens for each and stored them in the Gachon University Herbarium (GCU), assigning unique accession numbers (Table 1). All voucher specimens were identified by their morphological characters by the authors (Joonhyung Jung and Tao Deng). Total genomic DNA (gDNA) was extracted from fresh leaf material of each taxon employing a modified cetyltrimethylammonium bromide (CTAB) method [16]. This extracted gDNA was then used for next-generation sequencing (NGS) analysis on the Illumina Mi-seq platform (Illumina, Seoul, Korea). The raw sequencing data were then utilized for the de novo assembly of plastome sequences, facilitated by the GetOrganelle toolkit [17]. Subsequently, we conducted a 'map to reference' analysis of the plastome sequences to assess coverage, using the Geneious Prime 2023.1.1 program [18]. We annotated the gene content and sequence order with GeSeq [19]. All tRNAs were subjected to a secondary check using tRNAscan-SE web server (<http://lowelab.ucsc.edu/tRNAscan-SE/>) in its default search mode [20]. Finally, OGDraw [21] was employed to create visual representations of the complete plastome sequences.

**Table 1** Features of the LSC, SSC, and IR of plastomes in this study

Taxa	Length and G + C content				GenBank accession No.	Voucher (Accession)
	LSC bp (G + C%)	SSC bp (G + C%)	IR bp (G + C%)	Total bp (G + C%)		
<i>Corchoropsis crenata</i> Siebold & Zucc. (from China)	88,848 (34.5)	20,496 (30.5)	25,413 (43.1)	160,170 (36.7)	PP840627	Tianm-05
<i>Corchoropsis crenata</i> Siebold & Zucc. (from South Korea)	88,770 (34.5)	20,497 (30.5)	25,413 (43.1)	160,093 (36.7)	PP840628	TH200909001
<i>Corchoropsis crenata</i> var. <i>hupehensis</i> Pamp.	89,379 (34.4)	20,505 (30.6)	25,420 (43.1)	160,724 (36.6)	PP840629	JH211005001

### Comparative plastome analyses

The whole plastome sequences of seven Dombeyoideae, including *Corchoropsis*, and four Tilioideae taxa were aligned and visualized using the LAGAN mode in mVISTA [22, 23], with *Pityranthe trichosperma* (Merr.) Kubitzki (Brownlowioideae; GenBank accession No. ON813239) serves as the reference. Additionally, the nucleotide diversity (Pi) of each gene and non-coding regions among the seven Dombeyoideae taxa was examined using a sliding window size of 100 bp and a step size of 25 bp in DnaSP v6.0 [24]. The plastid protein-coding genes were also utilized to identify molecular diagnostic characters (MDCs) specific to *Corchoropsis* and to each subfamily, employing FastaChar v0.2.4 for the analysis [25].

### Repeat and codon usage analyses

The plastomes of *Corchoropsis* were analysed for simple sequence repeats (SSRs) using the MISA Perl script (MicroSatellite Identification Tool) [26]. The analysis established specific thresholds for minimum repeats: at least ten for mononucleotides, five for dinucleotides, four for trinucleotides, and three for tetra-, penta-, and hexanucleotide sequences. Concurrently, the REPuter tool [27] was applied to identify four types of sequence repetitions: forward, reverse, complementary, and palindromic, focusing on sequences at least 30 bp in length with a minimum of 90% similarity.

We employed DAMBE v7.3.11 [28] to determine the RSCU values across 78 plastid protein-coding genes of *Corchoropsis*.

### Phylogenetic analyses

We downloaded 32 complete plastome sequences from NCBI, including one from the Byttnerioideae (*Melochia corchorifolia* L.), three from the Grewioideae (*Colona floribunda* (Kurz) Craib, *Grewia biloba* G.Don, and *Microcos paniculata* L.) to serve as outgroups. Additionally, for ingroups, we acquired four sequences from the Helicteroideae, seven from the Sterculioideae, two from the Bombacoideae, six from the Malvoideae, one from the Brownlowioideae, four from the Tilioideae, and four from the Dombeyoideae to cover all nine subfamilies in Malvaceae s.l. (Table S1). Then, we extracted 78 plastid protein-coding genes and aligned them using MUSCLE, as embedded in the Geneious Prime 2023.1.1 program. We performed Maximum Parsimony (MP), Maximum Likelihood (ML), and Bayesian Inference (BI) analyses to infer the phylogenetic relationships of *Corchoropsis*. The MP analysis was performed using PAUP\* v4.0a [29], with all characters considered equally important and unordered, and gaps treated as missing data. We carried out searches involving 1,000 random taxon addition replicates with tree-bisection-reconnection (TBR) branch swapping

in PAUP\*, permitting up to ten trees to be held at each step. To evaluate internal support, we executed bootstrap analyses, termed parsimony bootstrap percentages (PBP), with 1,000 pseudoreplicates, applying the same parameters. For ML analyses, we utilized the IQ-TREE web server (<http://iqtree.cibiv.univie.ac.at/>) [30], calculating the support value, indicated as mean bootstrap percentage (MBP) and SH-like approximate likelihood ratio test (SH-aLRT), using 1,000 ultrafast bootstrap replicates. Prior to the Bayesian Inference (BI) analysis, we identified the optimal substitution model using the Bayesian Information Criterion (BIC) in MEGA 11 (Table S2) [31]. BI was conducted with MrBayes v3.2.7 [32], initiating two simultaneous runs from random trees for at least 1,000,000 generations, and sampling one tree every 1,000 generations. We discarded 25% of the trees as burn-in samples, and the remainder were used to construct a 50% majority-rule consensus tree. The proportions of bifurcations in this consensus tree were shown as posterior probabilities (PP) to gauge the robustness of the BI tree. We also checked the effective sample size values (ESS) for model parameters to ensure they exceeded 200. Finally, the phylogenetic trees were refined using FigTree v1.4.4 [33].

### Molecular dating

In our study, divergence times for *Corchoropsis* were estimated using BEAST v1.10.4 [34], based on 78 plastid protein-coding genes. Throughout this process, the GTR+I+G model was implemented alongside a Birth and Death speciation tree prior and an uncorrelated lognormal model for molecular clock estimations [35, 36]. The analysis was conducted through a Markov chain Monte Carlo (MCMC) over 100 million generations, with parameter sampling every 1,000 generations. We removed the initial 10,000 (10%) trees as burn-in and utilized TreeAnnotator v1.10.4 to derive a maximum clade credibility tree from the remaining samples. This tree was produced considering a posterior probability threshold of 0.50 and the average node heights. Mean divergence times and 95% higher posterior density (HPD) intervals for these estimates were compiled using Tracer v1.7.2 and subsequently visualized with FigTree v1.4.4.

For our calibration accuracy, we employed five fossils: (1) Wood representing the stem node of the Grewioideae, identified as *Grewinium canalisum* (Bande & Srivastava) Srivastava & Guleria, was calibrated to an age of 64–67 million years ago (Mya) with a lognormal prior distribution (mean=1.5, standard deviation=0.15, and offset=64; C1) [37]; (2) Leaf corresponding to the stem node of *Sterculia*, specifically *S. washburnii* Berry, was assigned ages of 66–72 Mya with a lognormal prior distribution (mean=3, standard deviation=0.3, and offset=66; C2) [38]; (3) Pollen representing the stem node

of the Bombacoideae, from *Bombacacidites annae* (Van der Hammen) Germeraad, was dated to 56–66 Mya with a lognormal prior distribution (mean=5, standard deviation=1, and offset=56; C3) [39, 40]; (4) Leaf indicative of the stem node of the Eumalvoideae, identified as *Malvaciphyllum macondicus* M.Carvalho, was calibrated to 55.8–61.7 Mya with a lognormal prior distribution, (mean=3, standard deviation=0.3, and offset=55.8; C4) [41]; and (5) Pollen representing the stem node of the Tilioideae, from *Tillia* sp., was dated to 66–72 Mya with a lognormal prior distribution (mean=3, standard deviation=0.3, and offset=66; C5) [42].

## Results

### Plastome features of *Corchoropsis*

A total of 116,672 to 1,084,633 reads were assembled, accounting for 1.6–5.0% of the total 7,371,166 to 24,724,830 reads (Table S3). The plastid genome of *Corchoropsis* is characterized by a quadripartite structure, which consists of a large single copy (LSC) region (88,770–89,379 bp), a small single copy (SSC) region (20,496–20,505 bp), and two inverted repeats (IRs) regions (25,413–25,420 bp) (Fig. 1 and Table 1). A comparison of the plastome sequences from two individuals of *C. crenata* var. *crenata*, collected respectively from China and South Korea, revealed a difference of 164 bp (0.2%). Additionally, we identified 889 and 937 difference (0.9–1.0%) between *C. crenata* var. *crenata* and *C. crenata* var. *hupehensis*, respectively. The genetic composition of *Corchoropsis* was found to encompass 129 genes, among *infA* gene is identified as a pseudogene, and 17 genes are repeated in the IR regions (Table 2). These genes comprise nine protein-coding genes (*atpF*, *ndhA*, *ndhB*, *petB*, *petD*, *rpl2*, *rpl16*, *rpoC1*, and *rps16*) and five tRNAs (*trnK-UUU*, *trnG-UCC*, *trnL-UAA*, *trnI-GAU*, and *trnA-UGC*), each with a single intron. In addition, three protein-coding genes (*clpP1*, *rps12*, and *pbf1*) are noted for having two introns (Table 2). Notably, the *rps12* gene undergoes trans-splicing; its 5' exon is found in the LSC region, whereas the 3' exon along with an intron are located within the IR regions.

### Comparative plastome analyses

Based on mVISTA results, we found that the plastomes exhibited high similarity, particularly in the coding and IR regions, which were more conserved compared to the non-coding, LSC, and SSC regions (Fig. 2). We analysed nucleotide divergences of plastid protein-coding genes, tRNA, rRNA, and non-coding regions to elucidate variant characteristics among the seven Dombeyoideae taxa (Fig. 3 and Table S4). The nucleotide diversity (Pi) for each plastid protein-coding gene ranged from 0 (*pbf1*, *petL*, *petN*, *psaI*, and *rps7*) to 0.04202 (*rpl32*), with an average of 0.00962. In the tRNA and rRNA regions,

variations in only nine genes were observed, ranging from 0.0002 (*rrn23*) to 0.01299 (*trnS-GCU*). In the non-coding regions, two intergeneric spacers (*trnR-UCU-atpA* and *rpl22-rps19*) showed remarkably high values ( $P_i > 0.1$ ).

From the alignment data of 78 plastid protein-coding genes within members of the Malvaceae s.l., we identified 316 MDCs specific to *Corchoropsis* (Fig. 4 and Table S5). Notably, both two individuals of *C. crenata* var. *crenata* exhibit 122 unique MDCs, in comparison to *C. crenata* var. *hupehensis*. Within the members of the Malvaceae s.l., the individual from China exhibited nine MDCs, including six deletions in the *psbK* gene, whereas the individual from South Korea displayed ten MDCs, all characterized as single nucleotide polymorphisms (SNPs). Moreover, we calculated the number of MDCs unique to each genus and subfamily within the Malvaceae s.l., with the range for genera extending from 63 (*Tilia* L.) to 707 (*Melochia* L.), and for subfamilies from 45 (Helicteroideae) to 707 (Byttnerioideae).

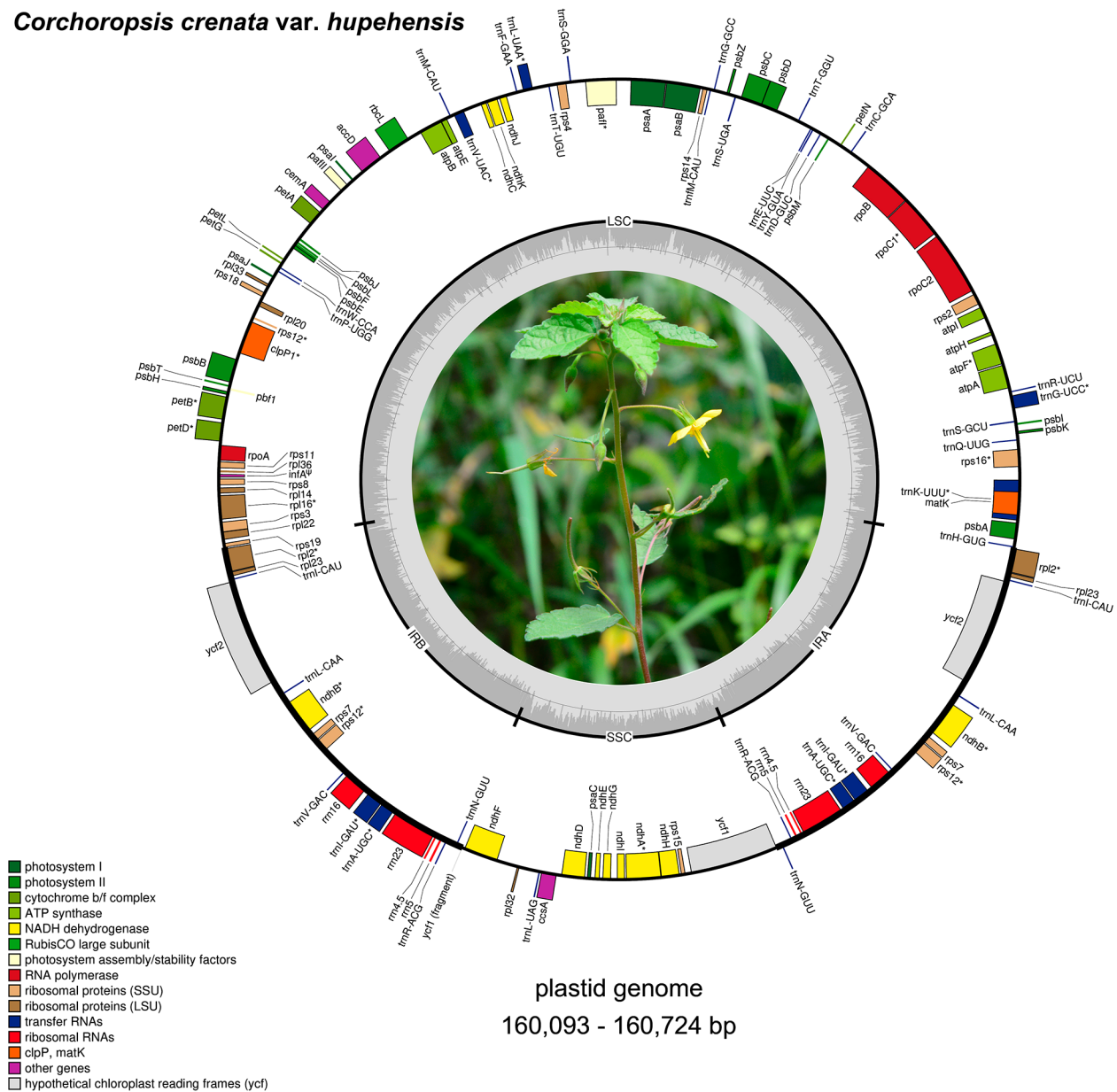
### Repeat and codon usage of *Corchoropsis*

In total, 91 to 92 SSRs were identified in *C. crenata* var. *crenata* and 75 in *C. crenata* var. *hupehensis* (Fig. 5). Both exhibited a high number of mono-nucleotide repeats. The number of dinucleotide repeats ranged from 13 in *C. crenata* var. *hupehensis* to 16–17 in *C. crenata* var. *crenata*. There were four instances of trinucleotide repeats. The number of tetranucleotide repeats varied, with eight observed in the *C. crenata* var. *crenata* individual from China and seven in both *C. crenata* var. *crenata* individual from South Korea and *C. crenata* var. *hupehensis*. Pentanucleotide repeats were less common, with two identified in *C. crenata* var. *crenata* and one in *C. crenata* var. *hupehensis*. The majority of SSRs consisted of the A/T motif, in contrast to the G/C motif, as detailed in Table S6.

Further analysis of longer repeats indicated a higher prevalence of forward and palindromic repeats over reverse and complementary ones across the *Corchoropsis*. Specifically, 49 long repeats were identified in both the *C. crenata* var. *crenata* individual from China and *C. crenata* var. *hupehensis*, and 52 in the *C. crenata* var. *crenata* individual from South Korea (Fig. 5). Only one complementary repeat was found in both *C. crenata* and its variety. The specific locations and the number of occurrences of these long repeats are detailed in Table S7.

An examination of 78 plastid protein-coding genes was conducted across *Corchoropsis* taxa to evaluate their relative synonymous codon usage (RSCU), excluding the stop codons UAA, UAG, and UGA (Fig. 6). There were slight differences in codon counts among the species, with *C. crenata* var. *crenata* presenting 22,729–22,731 codons and *C. crenata* var. *hupehensis* having a slightly

***Corchoropsis crenata* var. *crenata***  
***Corchoropsis crenata* var. *hupehensis***



**Fig. 1** Complete plastome map of *Corchoropsis* and their gene contents. The colored boxes represent conserved plastid genes. Genes located inside the circle are transcribed clockwise, while those outside the circle are transcribed counterclockwise. The grey bar graphs in the inner circle indicate the GC content of the plastome

higher figure of 22,732 (Table S8). Regarding amino acids, leucine (L) was identified as the most frequently occurring, constituting 10.47–10.49% of the total amino acids, whereas cysteine (C) was found to be the least common, representing only 1.14–1.15%.

**Phylogenetic analyses**

We conducted MP, ML, and BI analyses and observed consistent topologies across the phylogenetic trees, which strongly supported the monophyly of the nine

subfamilies of Malvaceae s.l. (Fig. 7). The sequence matrix encompassed 68,010 characters, with 57,869 (85.1%) being constant and 5,729 (8.4%) being parsimony informative. We derived the most parsimonious tree with a tree length of 14,303, consistency index (CI) of 0.802, and retention index (RI) of 0.837, as depicted in Fig. 7. Subsequently, two distinct clades were identified diverging with moderate support values: one comprising the Malvoideae and Bombacoideae, and the other including the Brownlowioideae, Dombeyoideae, and Tilioideae

**Table 2** Features of the LSC, SSC, and IR of *Corchoropsis*

Groups of genes		Names of genes	No.
RNA genes	Ribosomal RNAs	<i>rrn4.5</i> <sup>x2</sup> , <i>rrn5</i> <sup>x2</sup> , <i>rrn16</i> <sup>x2</sup> , <i>rrn23</i> <sup>x2</sup>	8
	Transfer RNAs	<i>trnH-GUG</i> , <i>trnK-UUU</i> <sup>a</sup> , <i>trnQ-UUG</i> , <i>trnS-GCU</i> , <i>trnG-UCC</i> <sup>a</sup> , <i>trnR-UCU</i> , <i>trnC-GCA</i> , <i>trnD-GUC</i> , <i>trnY-GUA</i> , <i>trnE-UUC</i> , <i>trnT-GGU</i> , <i>trnS-UGA</i> , <i>trnG-GCC</i> , <i>trnM-CAU</i> , <i>trnS-GGA</i> , <i>trnT-UGU</i> , <i>trnL-UAA</i> <sup>a</sup> , <i>trnF-GAA</i> , <i>trnV-UAC</i> <sup>a</sup> , <i>trnM-CAU</i> , <i>trnW-CCA</i> , <i>trnP-UGG</i> , <i>trnL-CAU</i> <sup>x2</sup> , <i>trnL-CAA</i> <sup>x2</sup> , <i>trnV-GAC</i> <sup>x2</sup> , <i>trnL-GAU</i> <sup>a</sup> <sup>x2</sup> , <i>trnA-UGC</i> <sup>a</sup> <sup>x2</sup> , <i>trnR-ACG</i> <sup>x2</sup> , <i>trnN-GUU</i> <sup>x2</sup> , <i>trnL-UAG</i>	37
Protein genes	Photosystem I	<i>psaA</i> , <i>psaB</i> , <i>psaC</i> , <i>psaI</i> , <i>psaJ</i>	5
	Photosystem II	<i>psbA</i> , <i>psbB</i> , <i>psbC</i> , <i>psbD</i> , <i>psbE</i> , <i>psbF</i> , <i>psbH</i> , <i>psbI</i> , <i>psbJ</i> , <i>psbK</i> , <i>psbL</i> , <i>psbM</i> , <i>psbT</i> , <i>psbZ</i>	14
	Photosystem I assembly	<i>pafl</i> <sup>b</sup> , <i>paflI</i>	2
	Photosystem biogenesis	<i>pbfl</i>	1
	Cytochrome	<i>petA</i> , <i>petB</i> <sup>a</sup> , <i>petD</i> <sup>a</sup> , <i>petG</i> , <i>petL</i> , <i>petN</i>	6
	ATP synthases	<i>atpA</i> , <i>atpB</i> , <i>atpE</i> , <i>atpF</i> <sup>a</sup> , <i>atpH</i> , <i>atpI</i>	6
	Large unit of Rubisco	<i>rbcl</i>	1
	NADH dehydrogenase	<i>ndhA</i> <sup>a</sup> , <i>ndhB</i> <sup>a</sup> <sup>x2</sup> , <i>ndhC</i> , <i>ndhD</i> , <i>ndhE</i> , <i>ndhF</i> , <i>ndhG</i> , <i>ndhH</i> , <i>ndhI</i> , <i>ndhJ</i> , <i>ndhK</i>	12
	ATP-dependent protease subunit P	<i>clpP1</i> <sup>b</sup>	1
	Envelope membrane protein	<i>cemA</i>	1
Ribosomal proteins	Large units of ribosome	<i>rpl2</i> <sup>a</sup> <sup>x2</sup> , <i>rpl14</i> , <i>rpl16</i> <sup>a</sup> , <i>rpl20</i> , <i>rpl22</i> , <i>rpl23</i> <sup>x2</sup> , <i>rpl32</i> , <i>rpl33</i> , <i>rpl36</i>	11
	Small units of ribosome	<i>rps2</i> , <i>rps3</i> , <i>rps4</i> , <i>rps7</i> <sup>x2</sup> , <i>rps8</i> , <i>rps11</i> , <i>rps12</i> <sup>x2</sup> , <i>rps14</i> , <i>rps15</i> , <i>rps16</i> <sup>a</sup> , <i>rps18</i> , <i>rps19</i>	14
Transcription/translation	RNA polymerase	<i>rpoA</i> , <i>rpoB</i> , <i>rpoC1</i> <sup>a</sup> , <i>rpoC2</i>	4
	Initiation factor	<i>infA</i> <sup>ψ</sup>	0
	Miscellaneous protein	<i>accD</i> , <i>ccsA</i> , <i>matK</i>	3
	Hypothetical proteins and conserved reading frames	<i>ycf1</i> , <i>ycf2</i> <sup>x2</sup>	3
Total			129

<sup>a</sup>: gene with one intron; <sup>b</sup>: gene with two introns; <sup>x2</sup>: duplicated gene; <sup>ψ</sup>: pseudogene

(PBP = 71/SH-aLRT = 78/MBP = 77/PP = 0.938). Within this framework, the subfamily Tilioideae was identified as the sister group to the Dombeyoideae, supported by high values (PBP = 86/SH-aLRT = 100/MBP = 100/PP = 1). In the Dombeyoideae, *Pterospermum* Schreb. formed the basal clade, and *Eriolaena* DC. was determined to be the sister to *Corchoropsis* (PBP = 100/SH-aLRT = 100/MBP = 100/PP = 1).

### Molecular dating

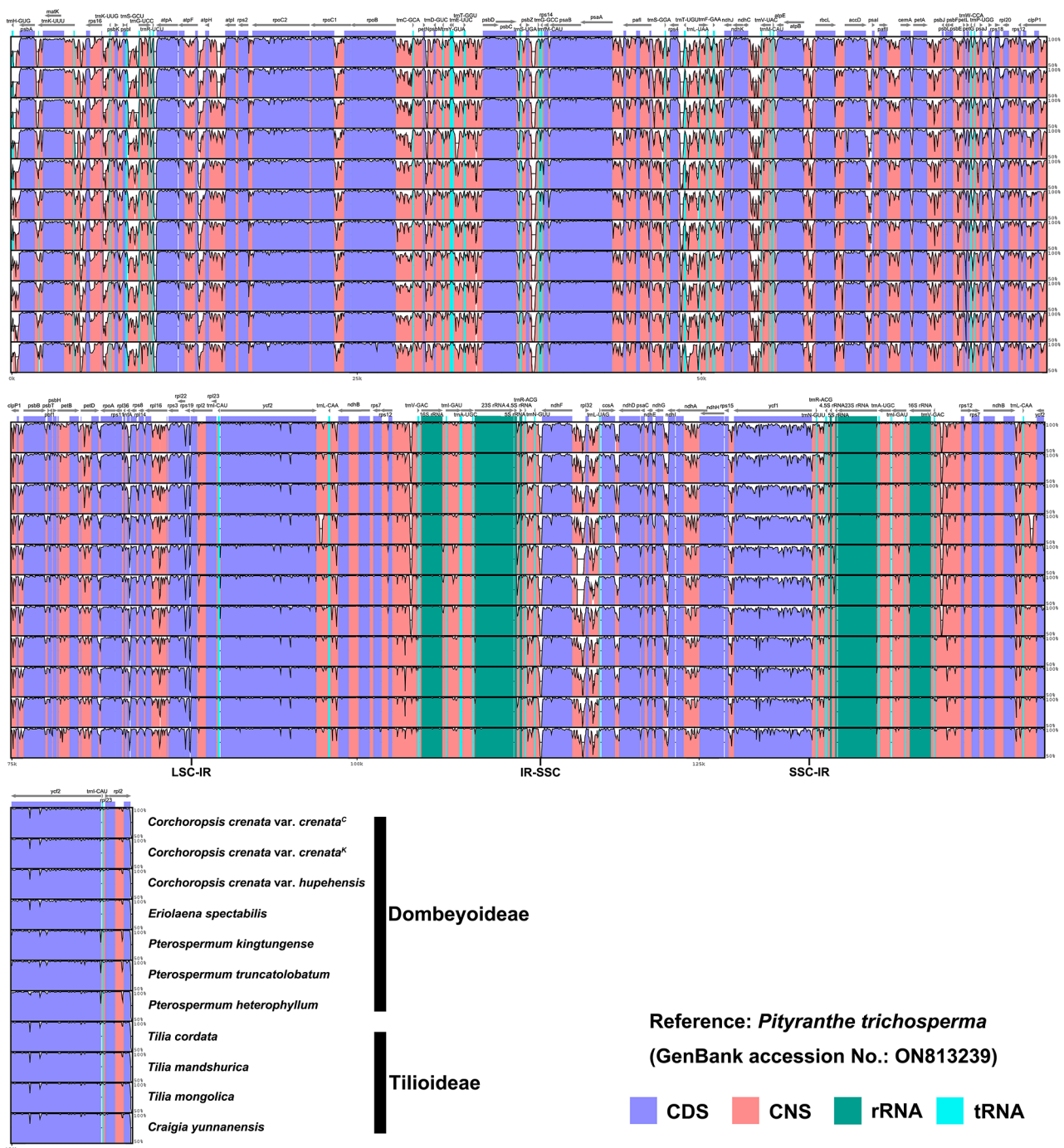
Mean divergence age estimates and their corresponding 95% HPD intervals for key phylogenetic nodes, derived from BEAST analysis, are presented in Fig. 8 and Table 3. This analysis indicates that the crown node of Malvaceae s.l. occurred during the Lower Cretaceous period, approximately 108.36 Mya, with a 95% HPD interval of 85.04–139.93 Mya (node 1). The Helicteroideae is estimated to have diverged around 96.59 Mya, with a 95% HPD interval of 81.05–119.44 Mya (node 2). The Sterculioideae is estimated to have diverged around 87.49 Mya, with a 95% HPD interval of 76.55–101.42 Mya (node 3). The Bombacoideae+Malvoideae clade is estimated to have diverged approximately 86.25 Mya, with a 95% HPD interval of 75.67–99.86 Mya (node 4), with further divergence occurring in the Paleocene. Within the Brownlowioideae+Tilioideae+Dombeyoideae clade (node 5),

Dombeyoideae+Tilioideae is diverged around 79.27 Mya, with a 95% HPD interval of 70.54–90.12 Mya (node 6). Additionally, within the Dombeyoideae, the genus *Corchoropsis* is estimated to have originated in the Oligocene, approximately 32.84 Mya, with a 95% HPD interval of 16.09–49.19 Mya (node 8).

## Discussion

### Complete plastomes of *Corchoropsis* and its comparison

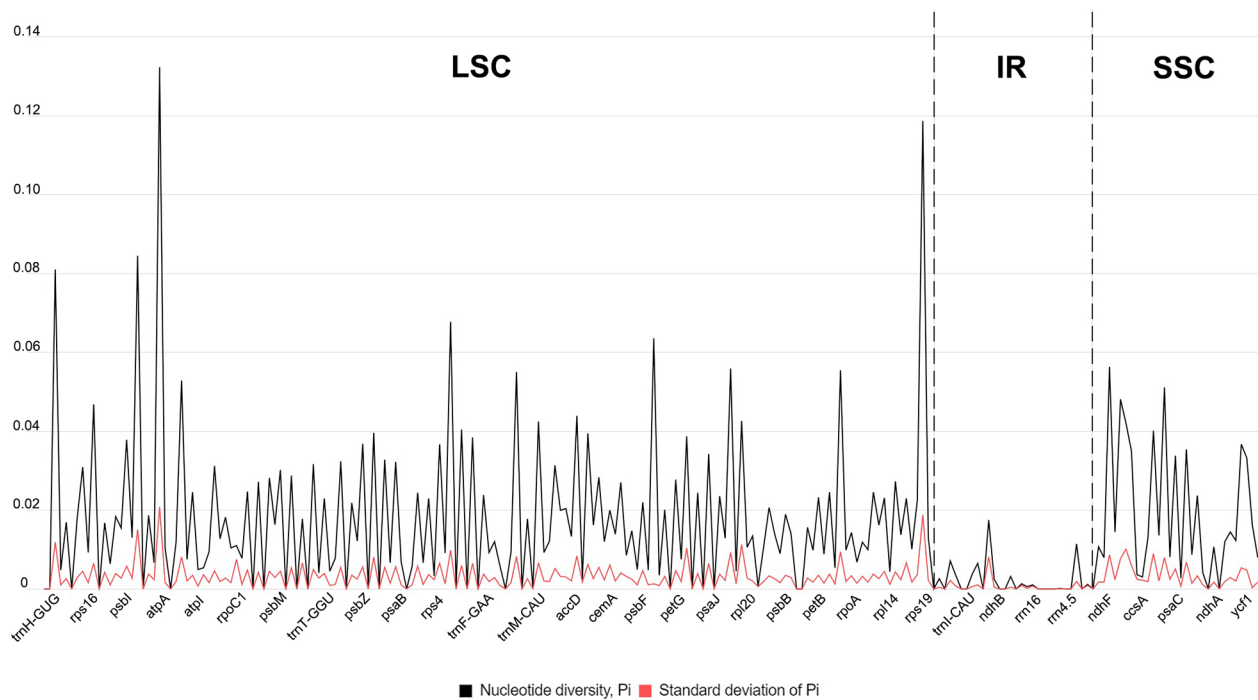
The plastome, well known for its photosynthetic functions, is highly conserved and small in size compared to the two other types of plant genomes [12]. With growing interest in biodiversity across various countries, genomic research focused on endemic plants has seen significant advancement [43, 44]. Here, we have completed plastome sequences of the East Asian endemic genus, *Corchoropsis*, noted for its highly conserved structure based on our comparative genomic analyses. It was observed that the *infA* gene, encoding translation initiation factor 1, has undergone pseudogenization, a common occurrence in many members of Malvaceae s.l. [45–47]. In the Dombeyoideae, *Pterospermum* also exhibits a pseudogenized *infA* gene. In contrast, *Eriolaena*, which is identified as the sister to *Corchoropsis* in this study, possesses an intact form of the gene, suggesting it is not a synapomorphic character (not shown).



**Fig. 2** Plots of percent sequence identity of the plastomes of seven Dombeyoideae and four Tilioideae taxa with *Pityranthe trichosperma* (Brownlowioideae; GenBank accession No. ON813239) as a reference. The percentage of sequence identities was estimated, and the plots were visualised in mVISTA. Superscript abbreviations; C : China individual, K : South Korea individual

Through mVISTA and nucleotide diversity analyses, we identified higher nucleotide variations in non-coding regions than in coding regions, a pattern found in most angiosperms [48, 49]. Two non-coding regions, *rpl22-rps19* and *trnR-UCU-atpA*, located in the LSC region, exhibit high diversity ( $P_i > 0.1$ , Fig. 3). Additionally, two

coding regions, *rpl32* and *ycf1*, located in the SSC region, also show high diversity ( $P_i > 0.03$ ). Generally, the IR region has lower diversity due to its importance in replication initiation, structural stability, and gene conservation, thus remaining well-conserved [50–52]. Despite its duplicated form affecting gene composition through IR



**Fig. 3** Nucleotide diversity (Pi) values of the plastomes in seven Dombeyoideae taxa

expansion and contraction, we identified no differences among seven members of the Dombeyoideae in this study.

Recent advancements in species identification have been achieved through the use of MDCs. These developments enhance traditional barcoding methods and focus on refining species identification across various taxonomic categories [25, 53]. We counted MDCs from 78 plastid protein-coding genes, which enhances the possibility of species identification due to their conservation for specific functions (Fig. 4 and Table S5). In *Corchoropsis*, a total of 316 MDCs were identified, including one independent insertion encoding 'IC'. Additionally, two independent deletions encoding 'NNHK' and 'FLN' were found. All these genetic variations were observed in the *ycf1* gene, which exhibits high diversity.

#### Repeat and codon usage of *Corchoropsis*

SSRs, highly regarded for their polymorphism, are particularly suited for phylogenetic and population genetic studies [54, 55]. In our study, the identified SSRs predominantly exhibited a high A/T content. Consequently, the majority of repeats in the *C. crenata* var. *crenata* individuals from China (67.03%), South Korea (67.39%), and *C. crenata* var. *hupehensis* (66.67%) were composed of A/T. Additionally, our study uncovered long repeats within *Corchoropsis*, showing slight variations across different types. The data on SSRs and long repeats collected here offer valuable insights for the selection of effective

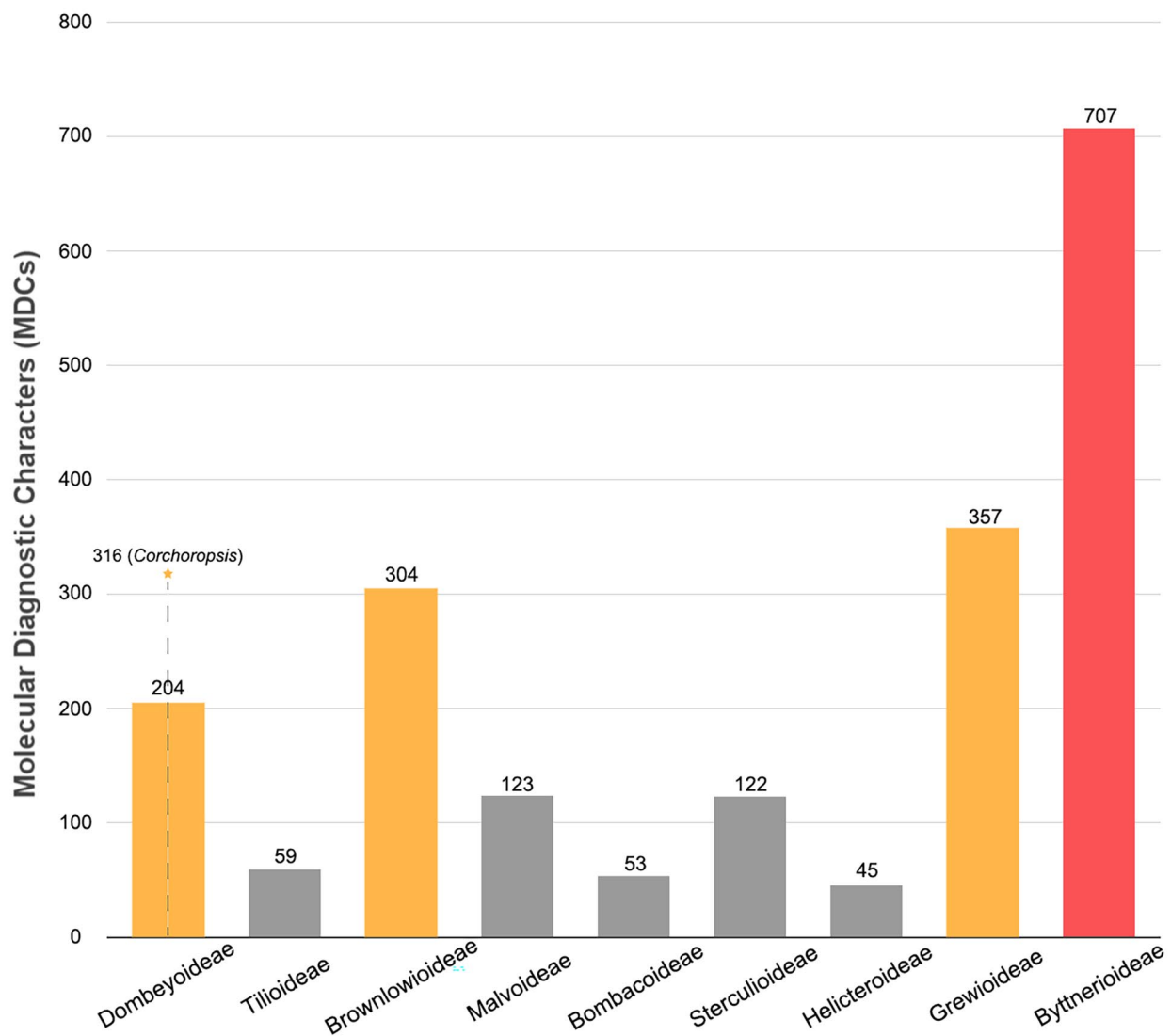
molecular markers for distinguishing *C. crenata* and its variety.

An examination of codon usage within plastid protein-coding genes can yield insights into mutation trends, the influences of selection, and genetic drift at the species level. Typically, most amino acids are encoded by two to six synonymous codons, with the exceptions of methionine (M) and tryptophan (W). Our analysis revealed that 29 codons exhibited elevated RSCU values (greater than 1), predominantly terminating in A or U. Conversely, codons with lower RSCU values (less than 1) frequently ended in G or C (Table S8). Notably, the codon AUU, which encodes for isoleucine (Ile), was found to be the most prevalently used, aligning with observations in other studies within the Malvaceae s.l. group [45, 56].

#### Phylogenetic relationships and divergence times of *Corchoropsis*

Historically, the Bombacaceae, Malvaceae s.s., Sterculiaceae, and Tiliaceae constituted the core Malvales, renowned for their close interrelationships [57, 58]. Phylogenetic analyses, employing morphological, anatomical, palynological, and chemical characteristics, have revealed that within these families, only the Malvaceae s.s. are monophyletic. The other three families demonstrate paraphyletic or polyphyletic relationships, led to the recommendation to unify these groups under a single family, Malvaceae s.l [59]. In previous studies on the sub-familiar relationships within the Malvaceae s.l., Alverson



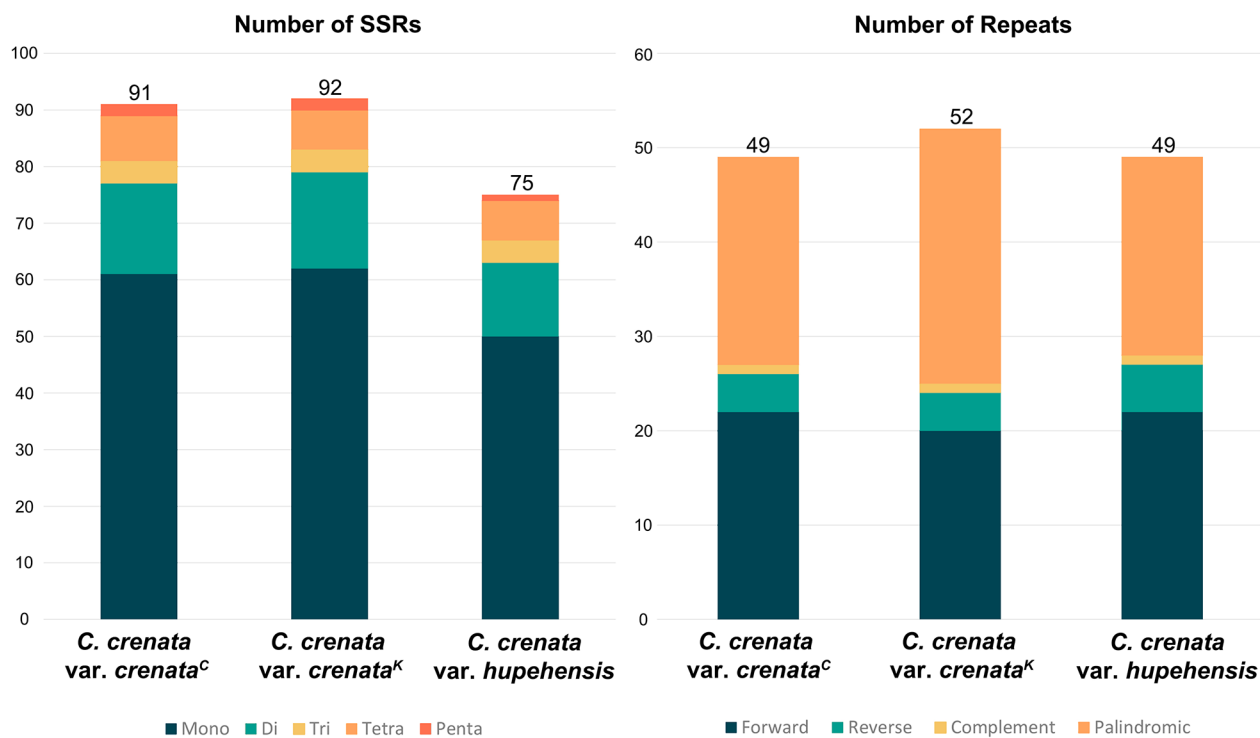


**Fig. 4** The number of MDCs in 78 plastid protein-coding genes across 35 Malvaceae s.l. taxa. The red bar represents more than 500 MDCs, the yellow bar shows 200 to 500 MDCs, and the grey bar indicates fewer than 200 MDCs among the subfamilies

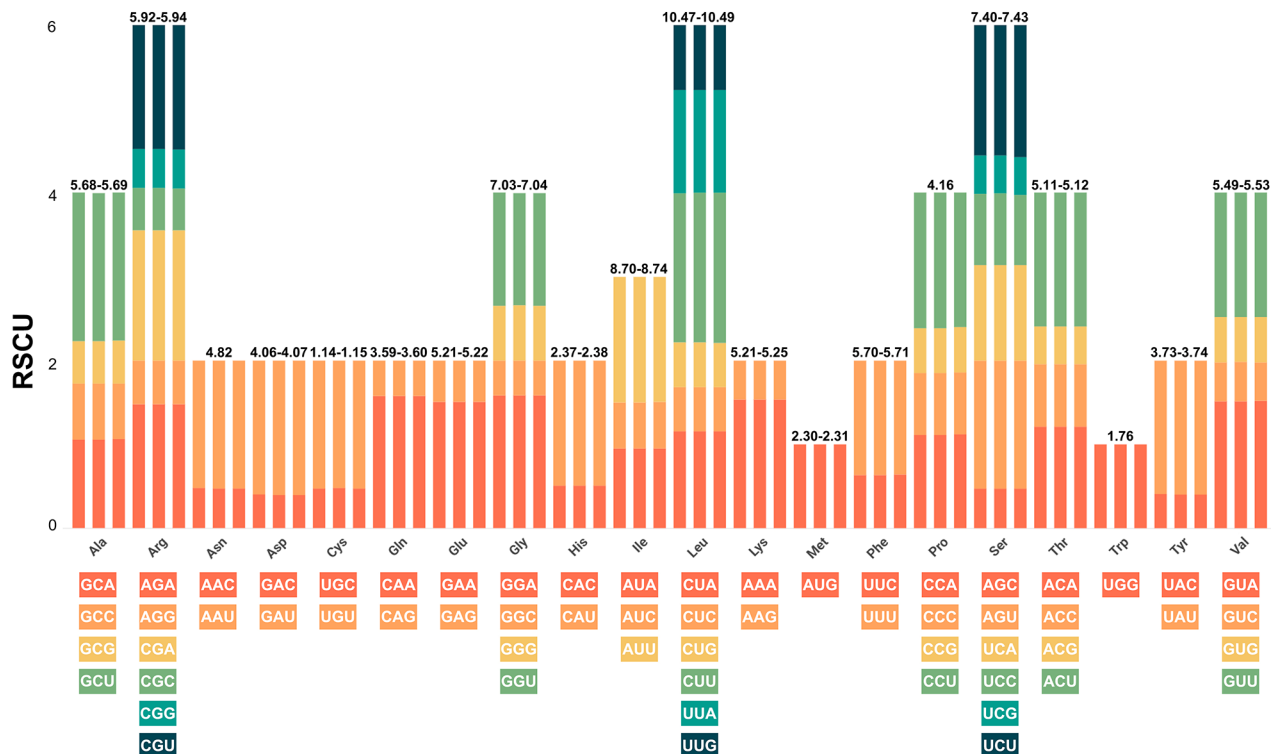
et al. [60] utilized the *ndhF* gene to group the Dombeyoideae with Tilioideae, revealing polytomy among the subfamilies. Hernandez-Gutierrez and Magallon [61] later reconstructed the Malvaceae s.l. using six plastid, one mitochondrial, and one nuclear region, suggesting that the Dombeyoideae is closely related to Brownlowioideae, although with low support. Conover et al. [62] examined the Malvaceae s.l. based on 67 plastid genes and identified the Dombeyoideae as sister to a group comprising the Bombacoideae, Malvoideae, Sterculioideae, and Tilioideae. Wang et al. [63] and Li et al. [64], focusing on plastome sequences, indicated that the Tilioideae as a sister group to the Dombeyoideae, notably excluding the Brownlowioideae from this relationship. Subsequently, the phylogenetic positions of nine subfamilies were clearly resolved based on plastome sequences with robust

support values, following the resolution of the plastome of the Brownlowioideae [56, 65]. Notably, the Dombeyoideae, which includes the genus *Corchoropsis*, emerged as a sister group to the Tilioideae, and these two subfamilies formed a cluster with the Brownlowioideae. Our phylogenetic analyses further validated the monophyly of the nine subfamilies of the Malvaceae s.l., corroborating these recent findings (Fig. 7) [56, 65].

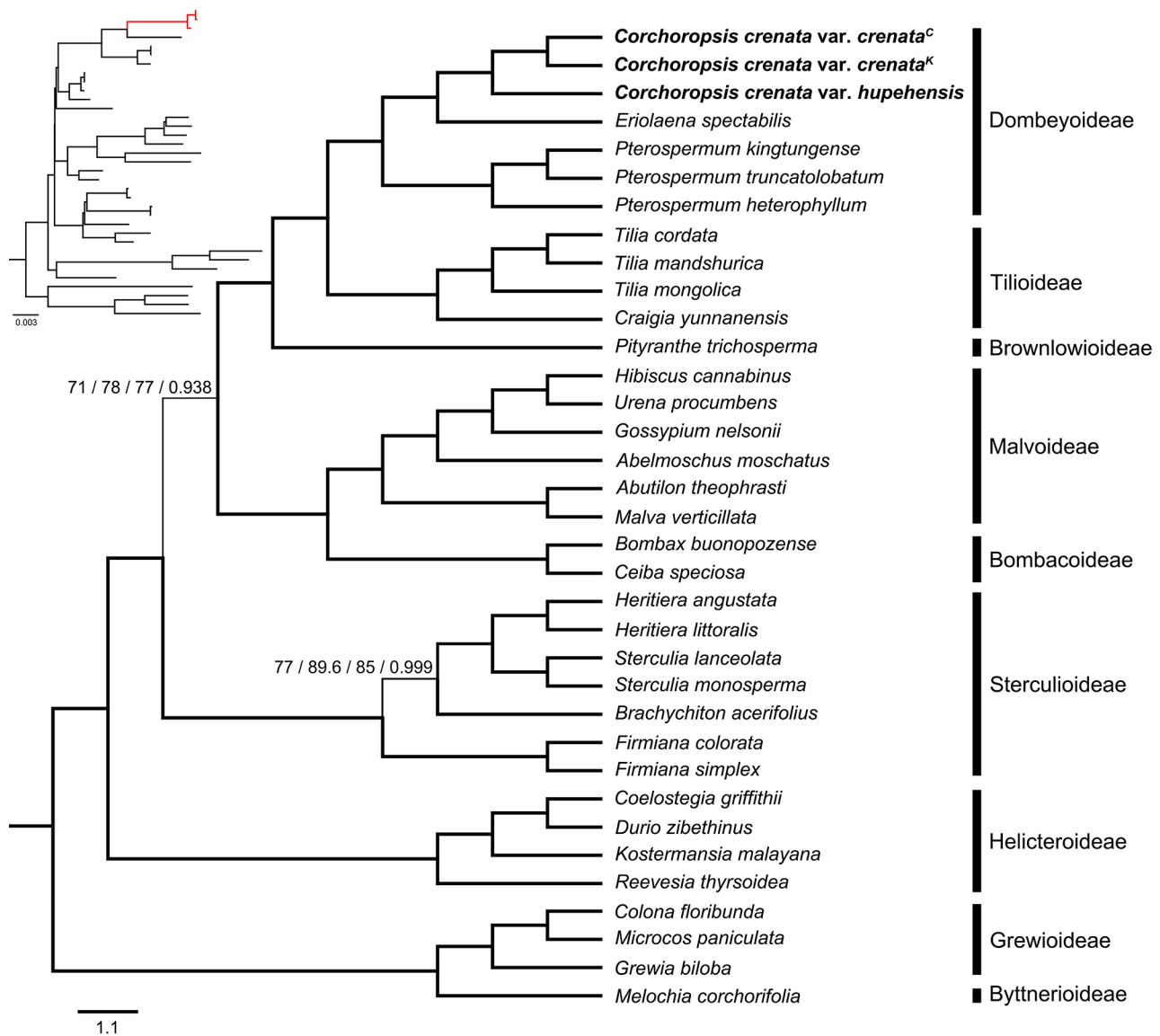
Currently, *Corchoropsis* shows a close association with other Asian genera like *Eriolaena* and *Pentapetes* L. within the Dombeyoideae of the Malvaceae s.l., a relationship supported by morphological characteristics [7]. Within the genus, *C. crenata* var. *hupehensis* is distinguished from *C. crenata* var. *crenata* by its glabrous ovary and capsule, featuring a red stigma, whereas the latter features a yellow stigma. The distinctive color of



**Fig. 5** Types and numbers of SSRs and long repeats in the plastomes of *Corchoropsis*. The number at the top of each stack represents the total count of both SSRs and long repeats



**Fig. 6** Relative synonymous codon usage (RSCU) analysis of 20 amino acids in protein-coding genes of the complete plastomes of *Corchoropsis*. The values at the top of each stack indicate the frequency of use of each amino acid. The order of each three columns from left to right is *C. crenata* var. *crenata*<sup>C</sup>, *C. crenata* var. *crenata*<sup>K</sup>, and *C. crenata* var. *hupehensis*



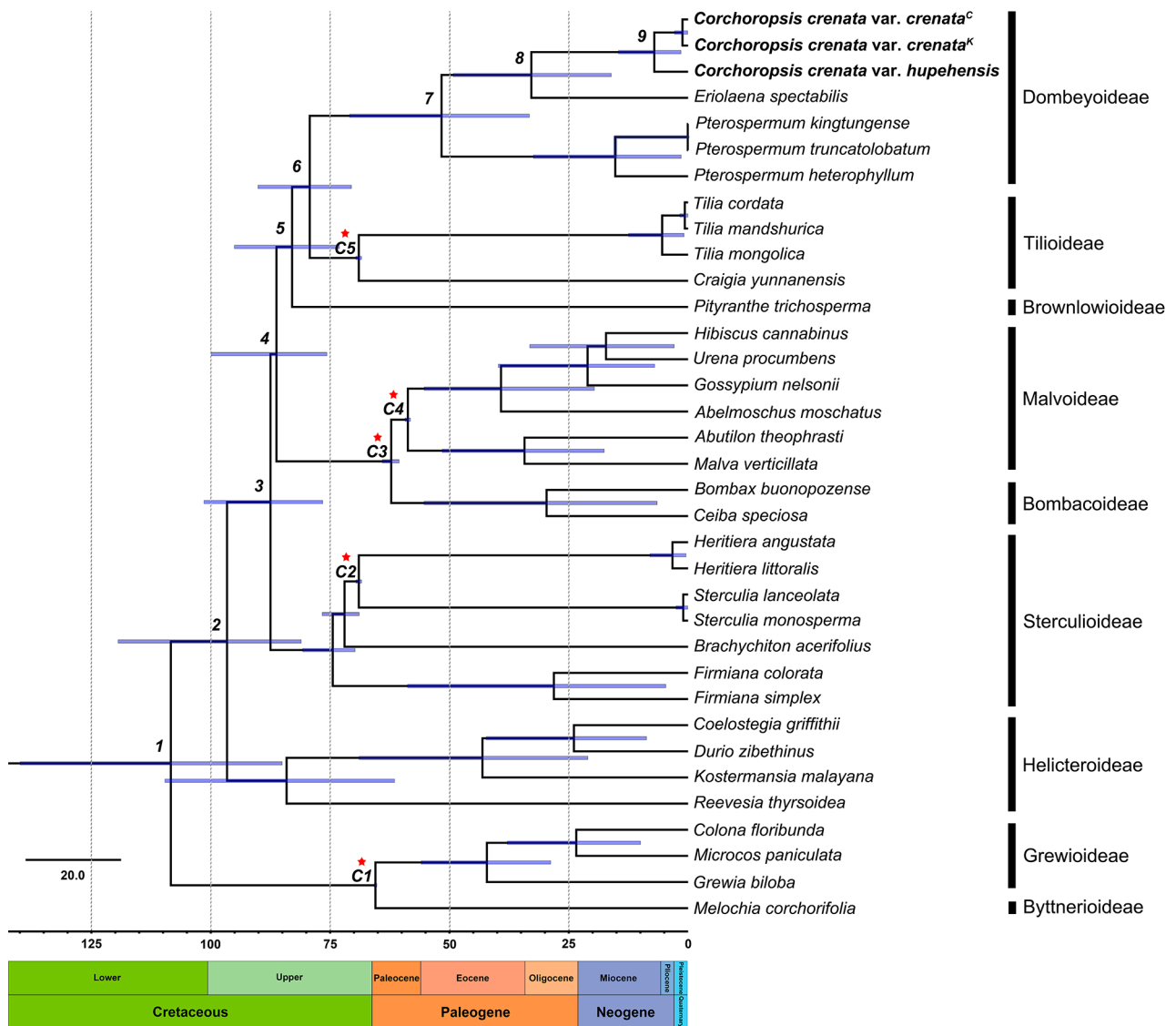
**Fig. 7** The Maximum Likelihood (ML) tree derived from 78 plastid protein-coding genes includes 35 Malvaceae s.l. taxa. Numbers indicate support values, represented as parsimony bootstrap percentages (PBP)/SH-aLRT support/mean bootstrap percentages (MBP)/posterior probability (PP). Only support values with PBP  $\leq$  85%, MBP  $\leq$  95%, and PP  $\leq$  0.95 are displayed. The bold names indicate genomes obtained in this study, while the boxes represent the subfamily within the Malvaceae s.l.

the stigma is influenced by the apocarotenoids crocetin and crocin, which are products of the oxidative cleavage of zeaxanthin and play a significant role in attracting pollinators [66, 67]. Both having the same distribution suggests that this may be an example of adaptive evolution and may necessitate a discussion regarding recognition as a separate species.

The initial molecular phylogenetic analysis of *Corchoropsis*, utilizing three plastid protein-coding genes, substantiated its placement within the Dombeyoideae and its close relationship with related genera [10]. In this study, plastome sequences have confirmed *Corchoropsis* as a member of the Dombeyoideae, however, there is still a

lack of genomic data for this subfamily to clarify inter-generic relationships. Additionally, further studies are needed to examine the recent taxonomic synonymization of members from *Paradombeya* with *Corchoropsis* [11], although recent study supported their monophyly with high support values based on six molecular markers [68].

Numerous studies have been conducted to estimate the divergence times of the Malvaceae s.l. and its members, with fossil data suggesting that the crown node age of the Malvaceae s.l. ranges from 70.7 to 110.47 Mya [61, 63, 65, 68, 69]. These results varied due to differences in taxon sampling, the fossils used, and the tree prior model, which are sensitive factors for the analyses. Our



**Fig. 8** Chronogram of Malvaceae s.l. based on 78 plastid protein-coding genes estimated from BEAST. The nodes represent posterior mean ages, with blue bars indicating 95% HPD intervals. Numbers 1–9 correspond to nodes of interest (Table 3). Nodes labeled C1–C5 with red star are calibration points used in the analysis. The geological epoch is displayed below the tree

**Table 3** The estimated age of interested nodes of Malvaceae s.l.

Node	Description	Age estimate	
		Mean (Mya)	95% HPD (Mya)
1	Malvaceae s.l. (crown node)	108.36	85.04–139.93
2	Helicteroideae (stem node)	96.59	81.05–119.44
3	Sterculioideae (stem node)	87.49	76.55–101.42
4	Bombacoideae + Malvoideae (stem node)	86.25	75.67–99.86
5	Brownlowioideae + Tilioideae + Dombeyoideae (crown node)	82.96	73.07–95.05
6	Tilioideae + Dombeyoideae (crown node)	79.27	70.54–90.12
7	Dombeyoideae (crown node)	51.65	33.26–70.89
8	<i>Corchoropsis</i> (stem node)	32.84	16.09–49.19
9	<i>Corchoropsis</i> (crown node)	7.09	1.49–14.57

results support the findings of Cvetković et al. [65], setting the oldest fossil, *Bombacoxylon langstoni* Wheeler & Lehman, for the crown node of Malvales [70]. Our analysis, based on plastome sequences and clearly defined relationships among the nine subfamilies, indicated that the stem ages of the subfamilies range between 62.20 Mya (Bombacoideae and Malvoideae) and 96.59 Mya (Helicteroideae). These ages are slightly higher than previous reports [61, 63, 65, 68, 69], however, the corresponding geological periods and the phylogenetic relationships align with Cvetković et al. [65]. Notably, their study incorporated the Dipterocarpaceae taxa into their data matrix, and the use of different taxa from each subfamily could influence the discrepancies observed in molecular dating. Our study suggests that the Dombeyoideae originated in the Upper Cretaceous (79.27 Mya) and further diverged in the Eocene (51.65 Mya). The crown node age of the Dombeyoideae is similar to that suggested by Skema et al. [68], which involved a broad ancestral range encompassing major areas including Asia, Africa, and Madagascar.

The Eocene–Oligocene boundary, around 34 Mya, had a significant impact on global biodiversity due to notable climatic shifts [71–73]. These changes led to increased dispersal events, especially pronounced during the Miocene, which in turn facilitated the diversification and emergence of new genera [74]. Specifically, within the Dombeyoideae, *Corchoropsis crenata* originated in this boundary and its variety further diversified during the Miocene. Other East Asian endemic angiosperms, especially *Dobinea* Buch.-Ham. ex D. Don (Anacardiaceae) and *Chimonanthus* Lindl. (Calycarthaceae), originated in the Eocene–Oligocene and further diversified in the Miocene, supporting speciation during that period [75, 76].

## Conclusions

Endemic plants, characterized by their restricted distribution, are crucial for biodiversity conservation and provide an essential foundation for investigating phylogenetic relationships, biogeographical histories, and genetic diversity. This study offers insights into the plastomes of the genus *Corchoropsis*, which is endemic to East Asia. The analysis of whole plastome sequences is instrumental in understanding structural variations, providing superbarcoding information, elucidating phylogenetic relationships, and estimating divergence times. We have gathered fundamental genomic data, especially regarding MDCs, SSRs, and long repeats, which are invaluable for future studies in superbarcoding, population genetics, and phylogenetics. Our research delineates the phylogenetic relationships among nine subfamilies of the Malvaceae s.l. and clarifies the phylogenetic position of *Corchoropsis* within the Dombeyoideae. Based on plastome sequences and fossil data, we have determined that *Corchoropsis*

originated in the Eocene and further diverged in the Miocene, similar to other East Asian endemic angiosperms.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-024-10725-0>.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Supplementary Material 4

Supplementary Material 5

Supplementary Material 6

Supplementary Material 7

Supplementary Material 8

## Author contributions

JJ and TD have contributed equally to this work. They performed the experiments, analysed the data, prepared figures and tables, and wrote the initial draft. CK and YGK collected the plant materials, designed the species sampling, and co-wrote the manuscript. HS and JHK designed the experiments and revised the manuscript. All authors agree with the content of the manuscript. All authors have contributed to the manuscript and approved the submitted version.

## Funding

This work was supported under the framework of international cooperation program managed by the National Research Foundation of Korea (NRF-2020K2A9A2A06069516), the Key Projects of the Joint Fund of the National Natural Science Foundation of China (U23A20149), the National Natural Science Foundation of China (32322006), and the Second Tibetan Plateau Scientific Expedition and Research (STEP) program (2019QZKK0502).

## Data availability

The three plastome sequences we obtained from this study were archived in NCBI. The accession numbers are presented in Table 1 (PP840627-PP840629).

## Declarations

### Ethics approval and consent to participate

The study including plant samples complies with relevant institutional, national, and international guidelines and legislation. No specific permits were required for plant collection. The study did not require ethical approval or consent, as no endangered or protected plant species were involved.

### Consent for publication

Not applicable.

### Competing interests

The authors declare no competing interests.

Received: 23 May 2024 / Accepted: 20 August 2024

Published online: 12 September 2024

## References

1. Myers N, Mittermeier RA, Mittermeier CG, Da Fonseca GA, Kent J. Biodiversity hotspots for conservation priorities. *Nature*. 2000;403(6772):853–8.
2. Manchester SR, CHEN ZD, LU AM, Uemura K. Eastern Asian endemic seed plant genera and their paleogeographic history throughout the Northern Hemisphere. *J Syst Evol*. 2009;47(1):1–42.
3. Huang JH, Chen JH, Ying JS, Ma KP. Features and distribution patterns of Chinese endemic seed plant species. *J Syst Evol*. 2011;49(2):81–94.

4. Chung GY, Jang H-D, Chang KS, Choi HJ, Kim Y-S, Kim H-J, Son D-C. A checklist of endemic plants on the Korean Peninsula II. *Korean J Plant Taxonomy*. 2023;53(2):79–101.
5. Milne RI, Abbott RJ. The origin and evolution of Tertiary relict floras. 2002.
6. Qiu Y-X, Fu C-X, Comes HP. Plant molecular phylogeography in China and adjacent regions: tracing the genetic imprints of quaternary climate and environmental change in the world's most diverse temperate flora. *Mol Phylogenet Evol*. 2011;59(1):225–44.
7. Bayer C, Kubitzki K. *Malvaceae. Flowering Plants: Dicotyledons: Malvales, Caprales and Non-betain Caryophyllales*. Springer; 2003. pp. 225–311.
8. Tang Y. The systematic position of *Corchoropsis* Sieb. & Zucc. *Cathaya*. 1992;4:131–50.
9. Takeda H. The genus *Corchoropsis*. *Bull Misc Inf Kew* 1912, 365.
10. Won H. Phylogenetic position of *Corchoropsis* Siebold & Zucc. (Malvaceae s.l.) inferred from plastid DNA sequences. *J Plant Biology*. 2009;52:411–6.
11. Dorr LJ, Wurdack KJ. Indo-Asian *Eriolaena* expanded to include two Malagasy genera, and other generic realignments based on molecular phylogenetics of Dombeyoideae (Malvaceae). *Taxon*. 2021;70(1):99–126.
12. Dobrogowski J, Adamiec M, Luciński R. The chloroplast genome: a review. *Acta Physiol Plant*. 2020;42(6):98.
13. Kim T-H, Kim J-H. Molecular phylogeny and historical biogeography of *Goodyeria* R. Br. (Orchidaceae): a case of the vicariance between East Asia and North America. *Front Plant Sci*. 2022;13:850170.
14. Feng Z, Zheng Y, Jiang Y, Pei J, Huang L. Phylogenetic relationships, selective pressure and molecular markers development of six species in subfamily Polygonoideae based on complete chloroplast genomes. *Sci Rep*. 2024;14(1):9783.
15. Li J, Du Y, Xie L, Jin X, Zhang Z, Yang M. Comparative plastome genomics and phylogenetic relationships of the genus *Trollius*. *Front Plant Sci*. 2023;14:1293091.
16. Doyle J. DNA protocols for plants. *Molecular techniques in taxonomy*. Springer; 1991. pp. 283–93.
17. Jin J-J, Yu W-B, Yang J-B, Song Y, DePamphilis CW, Yi T-S, Li D-Z. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol*. 2020;21:1–31.
18. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. 2012;28(12):1647–9.
19. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. GeSeq—versatile and accurate annotation of organelle genomes. *Nucleic Acids Res*. 2017;45(W1):W6–11.
20. Chan PP, Lowe TM. tRNAscan-SE: searching for tRNA genes in genomic sequences. *Gene Prediction: Methods Protocols* 2019:1–14.
21. Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organelle genomes. *Nucleic Acids Res*. 2019;47(W1):W59–64.
22. Brudno M, Do CB, Cooper GM, Kim MF, Davydov E, Green ED, Sidow A, Batzoglou S, Program NCS. LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. *Genome Res*. 2003;13(4):721–31.
23. Mayor C, Brudno M, Schwartz JR, Poliakov A, Rubin EM, Frazer KA, Pachter LS, Dubchak I. VISTA: visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics*. 2000;16(11):1046–7.
24. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol Biol Evol*. 2017;34(12):3299–302.
25. Merckelbach LM, Borges LM. Make every species count: fastachar software for rapid determination of molecular diagnostic characters to describe species. *Mol Ecol Resour*. 2020;20(6):1761–8.
26. Beier S, Thiel T, Münch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. *Bioinformatics*. 2017;33(16):2583–5.
27. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res*. 2001;29(22):4633–42.
28. Xia X. DAMBE7: New and improved tools for data analysis in molecular biology and evolution. *Mol Biol Evol*. 2018;35(6):1550–2.
29. Swofford DL. Phylogenetic analysis using parsimony (\* and other methods). Version. 2002;4:b10.
30. Trifunopoulos J, Nguyen L-T, von Haeseler A, Minh BQ. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res*. 2016;44(W1):W232–5.
31. Tamura K, Stecher G, Kumar S. MEGA11: molecular evolutionary genetics analysis version 11. *Mol Biol Evol*. 2021;38(7):3022–7.
32. Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. *Syst Biol*. 2012;61(3):539–42.
33. Rambaut A. FigTree, a graphical viewer of phylogenetic trees (Version 1.4.4). Institute of evolutionary biology, University of Edinburgh; 2018.
34. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol*. 2018;4(1):vey016.
35. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. *PLoS Biol*. 2006;4(5):e88.
36. Gernhard T. Using birth-death model on trees. *J Theor Biol*. 2008;253(769–778):1066.
37. Wheeler EA, Srivastava R, Manchester SR, Baas P. Surprisingly modern latest cretaceous–earliest paleocene woods of India. *IAWA J*. 2017;38(4):456–542.
38. Berry EW. Tertiary flora from the Rio Pichileufu, Argentina. Volume 12. *Geological Society of America*; 1938.
39. Jaramillo CA, Dilcher DL. Middle Paleogene palynology of Central Colombia, South America: a study of pollen and spores from tropical latitudes; Middle Paleogene palynology of Central Colombia, South America: a study of pollen and spores from tropical latitudes. *Palaeontographica Abteilung B*. 2001;258(4–6):87–259.
40. Graham A. Late Cretaceous and Cenozoic history of Latin American vegetation and terrestrial environments; 2010.
41. Carvalho MR, Herrera FA, Jaramillo CA, Wing SL, Callejas R. Paleocene Malvaceae from northern South America and their biogeographical implications. *Am J Bot*. 2011;98(8):1337–55.
42. Rouse GE, Hopkins W, Piel K. Palynology of some late Cretaceous and early Tertiary deposits in British Columbia and adjacent Alberta. 1970.
43. Koepfli K-P, Paten B, Scientists GKCo, O'Brien SJ. The genome 10K project: a way forward. *Annu Rev Anim Biosci*. 2015;3(1):57–111.
44. Ghazal H, Adam Y, Idrissi Azami A, Sehli S, Nyarko HN, Chaouni B, Olasehinde G, Isewon I, Adebijoyi M, Ajani O. Plant genomics in Africa: present and prospects. *Plant J*. 2021;107(1):21–36.
45. Mehmood F, Shahzadi I, Waseem S, Mirza B, Ahmed I, Waheed MT. Chloroplast genome of *Hibiscus rosa-sinensis* (Malvaceae): comparative analyses and identification of mutational hotspots. *Genomics*. 2020;112(1):581–91.
46. Yang JY, Lee W, Pak J-H, Kim S-C. Complete chloroplast genome of Ulleung Island endemic basswood, *Tilia insularis* (Malvaceae), in Korea. *Mitochondrial DNA Part B*. 2018;3(2):605–6.
47. Lu Q, Luo W. The complete chloroplast genome of two *Firmiana* species and comparative analysis with other related species. *Genetica*. 2022;150(6):395–405.
48. Kim T-H, Ha Y-H, Setoguchi H, Choi K, Kim S-C, Kim H-J. First Record of Comparative Plastid Genome Analysis and phylogenetic relationships among *Corylopsis* Siebold & Zucc. (Hamamelidaceae). *Genes*. 2024;15(3):380.
49. Hu K, Sun X-Q, Chen M, Lu R-S. Low-coverage whole genome sequencing of eleven species/subspecies in *Dioscorea* sect. *Stenophora* (Dioscoreaceae): comparative plastome analyses, molecular markers development and phylogenetic inference. *Front Plant Sci*. 2023;14:1196176.
50. Palmer JD, Thompson WF. Chloroplast DNA rearrangements are more frequent when a large inverted repeat sequence is lost. *Cell*. 1982;29(2):537–50.
51. Heinhorst S, Cannon GC. DNA replication in chloroplasts. *J Cell Sci*. 1993;104(1):1–9.
52. Maréchal A, Brisson N. Recombination and the maintenance of plant organelle genome stability. *New Phytol*. 2010;186(2):299–317.
53. Jiang Y, Yang J, Folk RA, Zhao J, Liu J, He Z, Peng H, Yang S, Xiang C, Yu X. Species delimitation of tea plants (*Camellia* sect. *Thea*) based on super-barcodes. *BMC Plant Biol*. 2024;24(1):181.
54. Yang H, Li X, Liu D, Chen X, Li F, Qi X, Luo Z, Wang C. Genetic diversity and population structure of the endangered medicinal plant *Phellodendron amurense* in China revealed by SSR markers. *Biochem Syst Ecol*. 2016;66:286–92.
55. Zhang Z, Xie W, Zhao Y, Zhang J, Wang N, Ntakirutimana F, Yan J, Wang Y. EST-SSR marker development based on RNA-sequencing of *E. sibiricus* and its application for phylogenetic relationships analysis of seventeen *Elymus* species. *BMC plant biology* 2019, 19:1–18.
56. Wu M, He L, Ma G, Zhang K, Yang H, Yang X. The complete chloroplast genome of *Diplodiscus trichospermus* and phylogenetic position of Brownlowioideae within Malvaceae. *BMC Genomics*. 2023;24(1):571.

57. Bentham G. Notes on Malvaceae and Sterculiaceae. *Bot J Linn Soc.* 1862;6(23):97–123.
58. Cronquist A. An integrated system of classification of flowering plants. Columbia university; 1981.
59. Judd WS, Manchester SR. Circumscription of Malvaceae (Malvales) as determined by a preliminary cladistic analysis of morphological, anatomical, palynological, and chemical characters. *Brittonia.* 1997;49:384–405.
60. Alverson WS, Whitlock BA, Nyffeler R, Bayer C, Baum DA. Phylogeny of the core Malvales: evidence from *ndhF* sequence data. *Am J Bot.* 1999;86(10):1474–86.
61. Hernandez-Gutierrez R, Magallon S. The timing of Malvales evolution: incorporating its extensive fossil record to inform about lineage diversification. *Mol Phylogenet Evol.* 2019;140:106606.
62. Conover JL, Karimi N, Stenz N, Ané C, Grover CE, Skema C, Tate JA, Wolff K, Logan SA, Wendel JF. A Malvaceae mystery: a mallow maelstrom of genome duplications and maybe misleading methods? *J Integr Plant Biol.* 2019;61(1):12–31.
63. Wang J-H, Moore MJ, Wang H, Zhu Z-X, Wang H-F. Plastome evolution and phylogenetic relationships among Malvaceae subfamilies. *Gene.* 2021;765:145103.
64. Li R, Cai J, Yang J, Zhang Z, Li D, Yu W. Plastid phylogenomics resolving phylogenetic placement and genera phylogeny of Sterculioideae (Malvaceae sl). 2022.
65. Cvetković T, Arecas-Berzain F, Hinsinger DD, Thomas DC, Wieringa JJ, Ganesan SK, Strijk JS. Phylogenomics resolves deep subfamilial relationships in Malvaceae sl. *G3* 2021, 11(7):jkab136.
66. Castillo R, Fernández J-A, Gómez-Gómez L. Implications of carotenoid biosynthetic genes in apocarotenoid formation during the stigma development of *Crocus sativus* and its closer relatives. *Plant Physiol.* 2005;139(2):674–89.
67. Lv Y, Gao P, Liu S, Fang X, Zhang T, Liu T, Amanullah S, Wang X, Luan F. Genetic mapping and QTL analysis of stigma color in melon (*Cucumis melo* L). *Front Plant Sci.* 2022;13:865082.
68. Skema C, Jourdain-Fievet L, Dubuisson J-Y, Le Péchon T. Out of Madagascar, repeatedly: the phylogenetics and biogeography of Dombeyoideae (Malvaceae sl). *Mol Phylogenet Evol.* 2023;182:107687.
69. Richardson JE, Whitlock BA, Meerow AW, Madriñán S. The age of chocolate: a diversification history of *Theobroma* and Malvaceae. *Front Ecol Evol.* 2015;3:120.
70. Wheeler E, Lehman T. Late cretaceous woody dicots from the Aguja and Javelina Formations, Big Bend National Park, Texas, USA. *Iawa J.* 2000;21(1):83–120.
71. Bowen GJ. When the world turned cold. *Nature.* 2007;445(7128):607–8.
72. Katz ME, Miller KG, Wright JD, Wade BS, Browning JV, Cramer BS, Rosenthal Y. Stepwise transition from the Eocene greenhouse to the Oligocene icehouse. *Nat Geosci.* 2008;1(5):329–34.
73. Liu Z, Pagani M, Zinniker D, DeConto R, Huber M, Brinkhuis H, Shah SR, Leckie RM, Pearson A. Global cooling during the Eocene–Oligocene climate transition. *Science.* 2009;323(5918):1187–90.
74. Buerki S, Forest F, Stadler T, Alvarez N. The abrupt climate change at the eocene–oligocene boundary and the emergence of South-East Asia triggered the spread of sapindaceous lineages. *Ann Botany.* 2013;112(1):151–60.
75. Liu C, Yang J, Jin L, Wang S, Yang Z, Ji Y. Plastome phylogenomics of the east Asian endemic genus *Dobinea*. *Plant Divers.* 2021;43(1):35–42.
76. Jamal A, Wen J, Ma Z-Y, Ahmed I, Abdullah, Chen L-Q, Nie Z-L, Liu X-Q. Comparative chloroplast genome analyses of the winter-blooming eastern Asian endemic genus *Chimonanthus* (Calycanthaceae) with implications for its phylogeny and diversification. *Front Genet.* 2021;12:709996.

## Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.