BMC Plant Biology

RESEARCH Open Access

Historical biogeography and plastome evolution of Commelinaceae Mirb. (Commelinales) corroborate the East Gondwanan origins

Joonhyung Jung[1,2], Marco O. O. Pellegrini[3] and Joo-Hwan Kim[1*]

Abstract

Background Commelinaceae are a Pantropical monocot family comprising ca. 36 genera and 810 species. Due to the high morphological variation and limited sampling, the relationships within the family remain unclear. Previous studies, which relied on limited genetic loci or low-resolution sequencing, have not fully resolved these relationships. Additionally, there is a need to investigate the historical biogeography of Commelinaceae using genomic data. To address these issues, we assembled twelve Commelinaceae plastomes and inferred their biogeographical pathways using both newly generated and previously published data.

Results Our results indicate that all members of the subfamily Commelinoideae have shortened *acc*D and *rpo*A genes, which do not affect functionality. In contrast, members of the subf. Cartonematoideae possess intact forms of both genes. Phylogenetic analyses based on 79 plastid protein-coding genes provide much-needed insight into the placement of some key genera that had previously remained unplaced or were erroneously placed elsewhere within the family. Of these, our results support the monophyly of the subtribe Streptoliriinae, the non-monophyly of the subtr. Coleotrypinae, and the need to recognise a broader circumscription of *Aneilema* R.Br. (including *Rhopalephora* Hassk.). Our biogeographic analysis shows that the ancestor of Commelinaceae originated in the Australian region around 104.8 million years ago (Mya), thus revealing an East Gondwanan origin. We also found that four major dispersal events have shaped the intercontinental diversification of this family: (1) from the Australian to the Ethiopian region ca. 81.4 Mya; (2) from the Ethiopian to the Oriental + Palearctic region ca. 50.4 Mya; (3) from the Oriental + Palearctic to the Neotropical region ca. 50.4 Mya; and (4) from the Neotropical to the Ethiopian region ca. 32.3 Mya.

Conclusions We identified genetic events, such as inversions and IR expansion, which shed light on the genomic evolution of Commelinaceae. Our findings support taxonomic updates, and future work will integrate comprehensive phylogenomic and morphological datasets to deepen our understanding of morphological evolution and biogeographic history within Commelinaceae, establishing a robust foundation for systematic revisions in the family.

*Correspondence:
Joo-Hwan Kim
kimjh2009@gachon.ac.kr

Full list of author information is available at the end of the article

Clinical trial number
Not applicable.

Keywords Commelinid monocots, Comparative plastome genomics, Next-Generation sequencing, Phylogenetic analysis, Systematics, Biogeography

Introduction

Commelinaceae Mirb. (Commelinales) are a monocot family comprising 36 genera and 810 species distributed in the tropical and subtropical regions of the world [1–3]. Due to their ephemeral flowers opening during the mornings and their stems with mucilaginous secretion, they are commonly known as dayflowers or spiderworts. Members of this family are usually succulent herbs with internodes with a leaf-opposite line of uniseriate hairs, swollen internodes, bifacial leaves, closed and asymmetric leaf-sheaths, heterochlamydeous perianth, and the absence of a hypanthium and nectaries of any kind (Fig. 1) [1, 2, 4]. They also have highly branched thyrsoid inflorescences consisting of one to many-flowered cincinni, or, more rarely, the whole inflorescence can be reduced to a single flower [5]. Furthermore, Commelinaceae are anatomically well-circumscribed, characterized by a nodal vascular plexus and raphide canals [1, 4, 6].

Since the development of next-generation sequencing (NGS), several studies have been conducted for land plants to define the structure of their plastomes associated with age estimation and biogeographic studies [7, 8]. NGS enabled the extraction of low-copy genes or partial sequences that contain high-copy genes from raw data using genome skimming [9]. Researchers have prioritized using plastome sequences over the other two types of plant genomes due to their highly conserved characteristics and relatively small in size [10]. It has a typical quadripartite structure containing a large single copy (LSC) region, a small single copy (SSC) region, and pair of inverted repeat (IR) regions [11]. Several genomic events have been identified from complete plastome sequences, such as inversion, duplication, and pseudogenization, which are important for understanding their structural evolution [12, 13]. Recent NGS studies for Commelinaceae have focused on both plastid and nuclear genomes. Jung et al. [14] assembled 15 plastome sequences of the subfamily Commelinoideae and compared their structural variations. They confirmed the placement of historically phylogenetically dubious genus *Palisota* Rchb. Ex Endl. and monophyly of the subtribe Tradescantiinae s.l. at the genomic level, but highlighted limitations in understanding genome evolution due to restricted sampling for the subf. Cartonematoideae. Zuntini et al. [3] reconstructed a phylogenomic analysis for Commelinaceae based on 353 nuclear genes and a few selected plastid genes. They sequenced seven genera (*Aëtheolirion* Forman, *Dictyospermum* Wight, *Gibasoides* D.R.Hunt, *Matudanthus* D.R.Hunt, *Pseudoparis* H.Perrier, *Tapheocarpa* Cornan, and *Triceratella* Brenan) for the first time and proposed their placements within the family. They also elevated the tribe Triceratelleae to the subfamilial rank, subtr. Palisotinae to the tribal rank, and rejected subtribal divisions in the tr. Tradescantieae. Nonetheless, several questions related to plastome evolution in Commelinaceae remained unanswered by these authors.

Historical biogeography is essential for understanding the origins and evolution of modern biodiversity [15]. By studying the distribution history of organisms and the factors that have shaped these distributions over time, researchers can answer fascinating questions about the reasons behind current species distributions, the origins of a continent's flora, and the processes responsible for the spatial distribution of plants throughout history [16]. Biogeographic reconstructions can also inform conservation efforts and help predict how ecosystems may respond to climate change [16].

As part of our ongoing efforts to understand the genome evolution in Commelinaceae and clarify the group's phylogeny, we assembled twelve new plastomes for Commelinaceae, including the subf. Cartonematoideae. We also extracted 79 plastid protein-coding genes from plastomes and published sequence read archive (SRA) data. In this study, we aim to (1) explore the plastome structural evolution through analyses of sequence variation, gene content, and gene order; (2) define phylogenetic relationships within Commelinaceae solely based on plastomes; and (3) investigate the historical biogeography of Commelinaceae by reconstructing a dated phylogeny calibrated with fossils and estimate ancestral area reconstructions for continents.

Methods

Taxon sampling, DNA extraction, and sequencing

Fresh leaf material was collected in the field or botanical gardens and dried using silica gel for DNA extractions (Table 1). The sampling included one member from the subf. Cartonematoideae and eight from the subf. Commelinoideae, comprising two from the tr. Commelineae and six from the tr. Tradescantieae. We prepared the voucher specimens for all the samples and deposited

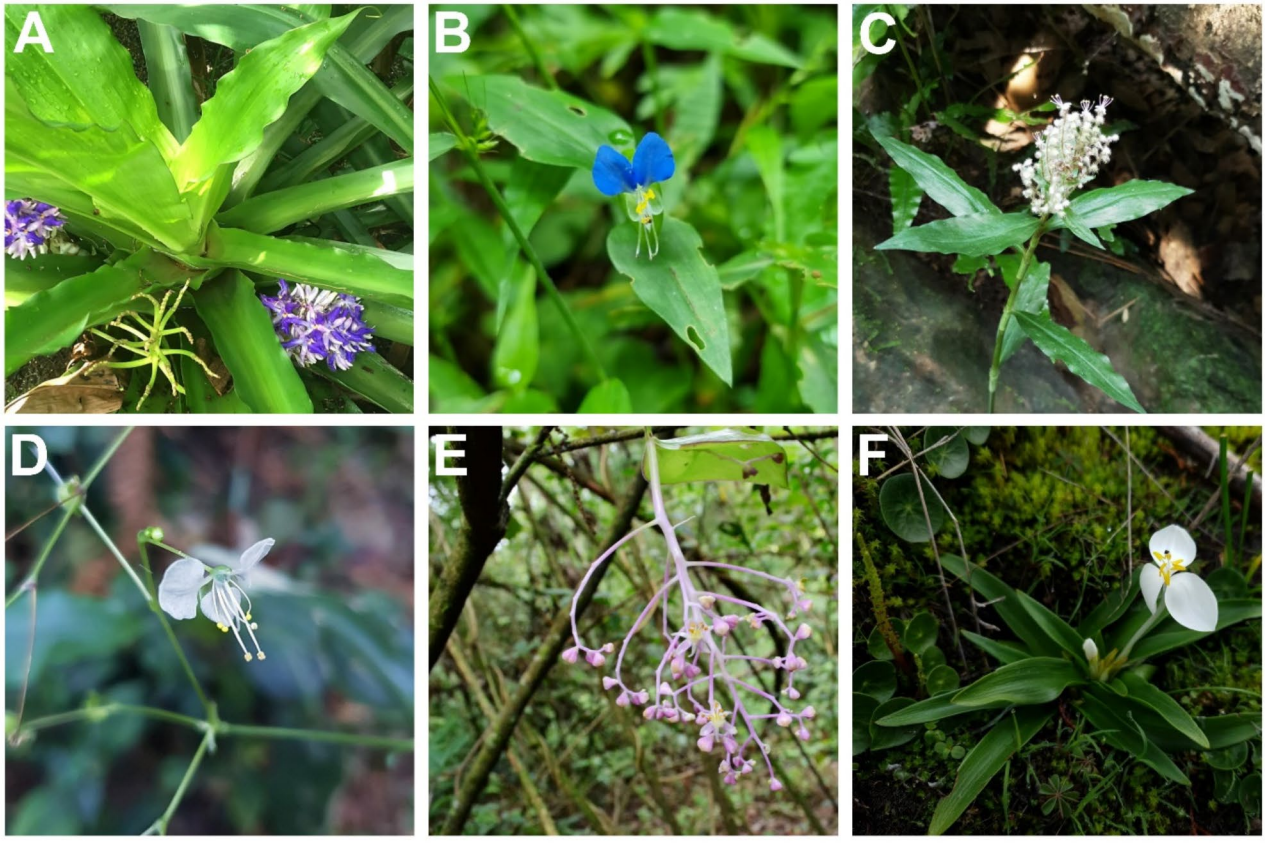

Fig. 1 Representative taxa of Commelinaceae: (**A**) *Cochliostema odoratissimum* Lem.; (**B**) *Commelina communis* L.; (**C**) *Floscopa scandens* Lour.; (**D**) *Rhopalephora scaberrima* (Blume) Faden; (**E**) *Spatholirion longifolium* (Gagnep.) Dunn; (**F**) *Weldenia candida* Schult.f

them in the Gachon University Herbarium (GCU) with their accession numbers. Total genomic DNA was extracted using a modified 2× CTAB method [17]. The concentration and optical density were measured using a spectrophotometer (Biospec-nano; Shimadzu) and assessed by agarose gel electrophoresis. We performed NGS on the nine taxa using the Illumina MiSeq sequencing system (Illumina, Seoul, Korea) (Table S1).

Plastome assembly and annotation

We downloaded three SRA datasets from the NCBI, including GenBank accession No. ERX5611015 for *Pollia condensata* C.B.Clarke, ERX8772748 for *Polyspatha paniculata* Benth., and ERX8772762 for *Commelina calandrinioides* (F.Muell.) Zuntini & Frankel (≡ *Tapheocarpa calandrinioides* (F.Muell.) Conran). Total reads were used for *de novo* assembly of a plastome using GetOrganelle toolkit [18]. To verify the complete plastome, we performed a 'map to reference' with raw reads and checked the coverage using the Geneious Prime 2024.0.5 [19]. Gene content and order were annotated using GeSeq [20]. All tRNAs were rechecked using tRNAScan-SE [21] with the default search mode. Illustrations of the plastomes were produced using CPGview [22].

Comparative plastome analyses

The genome structure, size, gene content, and order were compared (Table 1). The GC content was calculated and compared using the Geneious Prime 2024.0.5 [19]. To examine the structural variations at the generic level, we incorporated 18 previously published plastomes (Table S2). Whole plastome sequences were aligned using MUSCLE and visualized using the LAGAN mode in mVISTA [23]. For the mVISTA plot, we used annotated plastome sequences of *Hanguana malayana* (Jack) Merr. (Hanguanaceae Airy Shaw; GenBank accession No. KT312930) as a reference which has a sister-group relationship to Commelinaceae. The IR/LSC and IR/SSC boundaries of the 28 taxa were compared and illustrated using IRscope [24].

Nucleotide diversity (*Pi*) and relative synonymous codon usage (RSCU) analysis

We examined the nucleotide diversity (*Pi*) of plastid protein-coding genes, tRNA genes, and rRNA genes, including 16 previously published plastome data, through the sliding window analysis using DnaSP v6.0 [25]. We used a window size of 100 base pairs (bp) with a step size of 25 bp for sequence divergence analysis. Relative synonymous codon usage (RSCU) was computed from 79

Table 1 Characteristics of the LSC, SSC, and IR regions of plastomes analyzed in this study

Taxa	Tribe	Subtribe	Length and G + C content				GenBank accession No.	Voucher
			LSC bp (G + C%)	SSC bp (G + C%)	IR bp (G + C%)	Total bp (G + C%)		
Tripogandra serrulata (Vahl) Handlos	Tradescantieae	Tradescantiinae	89,860 (33.2)	18,786 (30.1)	27,102 (42.3)	162,850 (35.9)	OP758347	JH211213001
Tinantia pringlei (S.Watson) Rohweder	Tradescantieae	Tradescantiinae	87,080 (33.8)	18,684 (30.0)	27,090 (42.6)	159,944 (36.3)	OP758353	XX-0-GENT-20080921
Cyanotis ciliata (Blume) Bakh.f.	Tradescantieae	Cyanotinae s.l.	95,469 (31.6)	19,847 (28.6)	27,240 (42.7)	169,796 (34.8)	OP758346	JH191109005
Cyanotis speciosa (L.f.) Hassk.	Tradescantieae	Cyanotinae s.l.	94,953 (32.5)	19,780 (28.7)	27,338 (42.6)	169,409 (35.3)	OP758350	JH210912001
Coleotrype natalensis C.B.Clarke	Tradescantieae	Cyanotinae s.l.	92,936 (33.0)	19,364 (29.9)	27,308 (42.4)	166,916 (35.7)	OP758349	JH220429001
Spatholirion longifolium (Gagnep.) Dunn	Tradescantieae	Streptoliriinae	91,751 (33.3)	19,126 (30.0)	27,367 (42.1)	165,611 (35.8)	OP758352	KC0390
Pollia condensata C.B.Clarke	Commelineae		90,221 (33.3)	18,897 (30.0)	27,646 (42.2)	164,410 (35.9)	BK062834	Chase 18154 (LCD: 1993–3114)
Polyspatha paniculata Benth.	Commelineae		87,281 (33.7)	18,165 (29.7)	27,803 (42.0)	161,052 (36.1)	BK062835	Mas 33 (K001275399)
Aneilema beniniense (P.Beauv.) Kunth	Commelineae		89,991 (33.1)	18,377 (29.6)	26,964 (42.4)	162,296 (35.8)	OP758345	Pellegrini218
Commelina calandrinioides (F.Muell.) Zuntini & Frankel	Commelineae		87,725 (33.0)	18,510 (29.0)	27,124 (42.3)	160,483 (35.7)	BK062833	Clarkson 9511 (K001393142)
Floscopa scandens Lour.	Commelineae		93,868 (32.1)	20,193 (28.1)	28,229 (41.5)	170,519 (34.7)	OP758351	JH191109023
Cartonema parviflorum Hassk.	Cartonemeae		88,582 (35.0)	18,962 (31.3)	26,850 (42.8)	161,244 (37.2)	OP758348	AQ827754

plastid protein-coding genes using the DAMBE program [26].

Phylogenetic analyses

We downloaded 22 published SRA datasets more from NCBI (18 from Commelinaceae, three from Haemodoraceae R.Br., and one from Philydraceae Link) (Table S3). Then we performed a "map to reference" using the Bowtie 2 program [27] with a closely related taxa as a reference sequence, with the highest sensitivity to isolate raw reads of plastid protein-coding genes. Plastid protein-coding genes were used exclusively because intron regions could not be reliably obtained from the SRA data. A total of 38 complete plastome sequences (including twelve new plastome sequences in this study) and 22 partial plastid protein-coding sequences were used for the phylogenetic analysis (Table S4). We aligned them using MUSCLE embedded in Geneious Prime 2024.0.5 [19]. In the dataset, 11 taxa from the remaining four Commelinales families, Hanguanaceae, Haemodoraceae, Pontederiaceae Kunth, and Philydraceae, were designated as outgroups based on recent phylogenetic relationships [3, 28]. Consequently, only four genera, each comprising one to five species, were not included in the dataset: *Dictyospermum*, *Matudanthus*, *Sauvallia* C.Wright ex Hassk., and *Stanfieldiella* Brenan. The SRA data of *Dictyospermum conspicuum* (Blume) J.K.Morton (GenBank accession No. ERX8772747), identified via the Basic Local Alignment Search Tool (BLAST) from NCBI, showed a considerable percentage similarity (96.8–97.7%) with the plastomes of *Pollia* Thunb. A pilot study indicated that it grouped with *Pollia*, hinting at a probable misidentification of the species, initially tagged as *Dictyospermum conspicuum*, but likely belonging to *Pollia*. Therefore, we chose not to include this data in our investigation. Nonetheless, a recent study defined the position of *Dictyospermum*, sister to a clade including *Aneilema* R.Br., *Commelina* L., *Pollia*, *Polyspatha* Benth., and *Rhopalephora* Hassk. with high support values [29].

We performed maximum parsimony (MP), maximum likelihood (ML), and Bayesian inference (BI) to define the relationships within Commelinaceae. MP analysis was carried out in PAUP* v4.0a [30] with all characters equally weighted and unordered. Gaps were treated as missing data. Searches of 1,000 random taxon addition replicates used tree-bisection-reconnection (TBR) branch swapping, and MulTrees permitted ten trees to be held at each step. Bootstrap analysis with 1,000 pseudoreplicates was conducted to examine internal support (parsimony bootstrap percentages; PBP) using the same parameters. The IQ-TREE web server (http://iqtree.cibiv.univie.ac.at/) was used to estimate the best substitution model for each gene [31]. We then performed ML searches based on the edge-linked partition models.

Mean bootstrap percentage (MBP) and SH-like approximate likelihood ratio (SH-aLRT) were calculated with 1,000 ultrafast bootstrap replicates [31]. MrBayes v3.2.7 [32] was used for the BI analyses. Two simultaneous runs were performed, starting with random trees for at least 1,000,000 generations. One tree was sampled every 1,000 generations. In total, 25% of the trees were discarded as burn-in samples. The remaining trees were used to construct a 50% majority-rule consensus tree, with the proportion bifurcations found in this consensus tree representing the posterior probability (PP) used to estimate the robustness of the BI tree. The effective sample size values (ESS) were then checked for model parameters (at least 200). Phylogenetic trees were edited using FigTree v1.4.4 [33].

Molecular dating

We estimated the divergence times of Commelinaceae using BEAST v1.10.4 [34] based on 79 plastid protein-coding genes. The BEAUti v1.10.4 interface was used to generate input files for BEAST, where we applied the GTR + I + G model, Yule speciation tree prior, and uncorrelated lognormal molecular clock model [35, 36]. The Markov chain Monte Carlo (MCMC) analysis was run for 100 million generations, sampling parameters every 1,000 generations. We discarded the first 10,000 (10%) trees as burn-in and summarized the remaining samples in a maximum clade credibility tree using TreeAnnotator v1.10.4 with a posterior probability limit of 0.50 and mean node heights. We obtained mean and 95% highest posterior densities (HPD) of age estimates from the combined outputs using tracer and visualized the results using Figtree v1.4.4 [33].

We calibrated the age using two fossils. One of them was from an extinct *Pollia* species [37], which represented the most recent common ancestor (MRCA) of *Pollia* and was calibrated with an lognormal distribution (mean = 1, standard deviation = 1, and offset = 11.6; F1). The other fossil was from a mid-Tertiary period amber called *Pseudhaplocricus hexandrus* Poinar & K.L.Chambers [38], which represented the MRCA of the Dichorisandrinae *sensu* Faden and Hunt [1] + Coleotrypinae + Cyanotinae + Tradescantiinae s.l. clade. It was calibrated with an uniform distribution (upper = 37.8 and lower = 23.03; F2) based on the discovery that the amber was redeposited in turbiditic sandstones from the Upper Eocene to Lower Miocene Mamey Group [39]. To improve the accuracy of our calibration process, we incorporated three additional secondary calibration points: (1) crown node of Commelinales was estimated to be 112 million years ago (Mya) with normal prior distribution (mean = 112 and standard deviation = 1; S1) [28, 40]; (2) crown node of Haemodoraceae + Pontederiaceae + Philydraceae was calibrated to be 103 Mya with normal prior distribution (mean = 103 and standard deviation = 3; S2) [40]; and (3) crown node of Pontederiaceae was constrained to be 31 Mya with normal prior distribution (mean = 31 and standard deviation = 1; S3) [28, 41].

Ancestral range reconstructions and diversification rates

In our research, we utilized both herbarium specimens and the Plants of the World Online (POWO) database to procure biogeographical data about Commelinaceae [42]. We divided the distribution into five distinct regions based on Wallace's biogeographic regions to confirm dispersal events between the continents: Australian, Oriental + Palearctic, Ethiopian, Nearctic, and Neotropical [43]. Due to the inherent uncertainties regarding the geographical origins of the samples, we categorized species based on their entire distribution range. For the reconstruction of ancestral regions and to discern the geographical diversification patterns within Commelinaceae, we used the Bayesian binary method (BBM) and BioGeoBEARS [44] as incorporated in the Reconstruct Ancestral State in Phylogenies (RASP) v4.4 [45]. We applied all post-burn-in trees from the BEAST analysis to the BBM. It was executed for two million generations, spanning across 10 chains each, and two simultaneous runs, implementing the fixed state frequencies model (Jukes-Cantor) along with equal rate variation across sites. With the assistance of RASP's Compute Condense feature, we obtained a consensus tree that we then used to map the ancestral distribution at each node. We limited the maximum number of ancestral areas to five. In BioGeoBEARS, the Dispersal-Extinction-Cladogenesis (DEC) model [46] was the best-fit model compared to the Dispersal-Vicariance Analysis (DIVA) model [47], and the Bayesian Inference of Historical Biogeography for Discrete Areas (BayArea) model [48] (Table S5).

Results

Plastome characteristics

Twelve plastome sequences were newly assembled in this study, with genome sizes ranging from 159,944 bp in *Tinantia pringlei* (S.Watson) Rohweder to 170,519 bp in *Floscopa scandens* Lour., and read counts ranging from 5.0–12.6 million for each taxon (Fig. 2, Tables 1, and Table S1). Like most angiosperms, the Commelinaceae encoded 79 plastid protein-coding genes, 30 tRNA genes, and four rRNA genes (Table S6). However, in the subf. Commelinoideae (the remaining 11 taxa), we identified small indels in the 5' end of *acc*D gene and the 3' end of *rpo*A gene, as in previous studies (Table S6) [14]. These alterations resulted in a reduction of the total length of these genes.

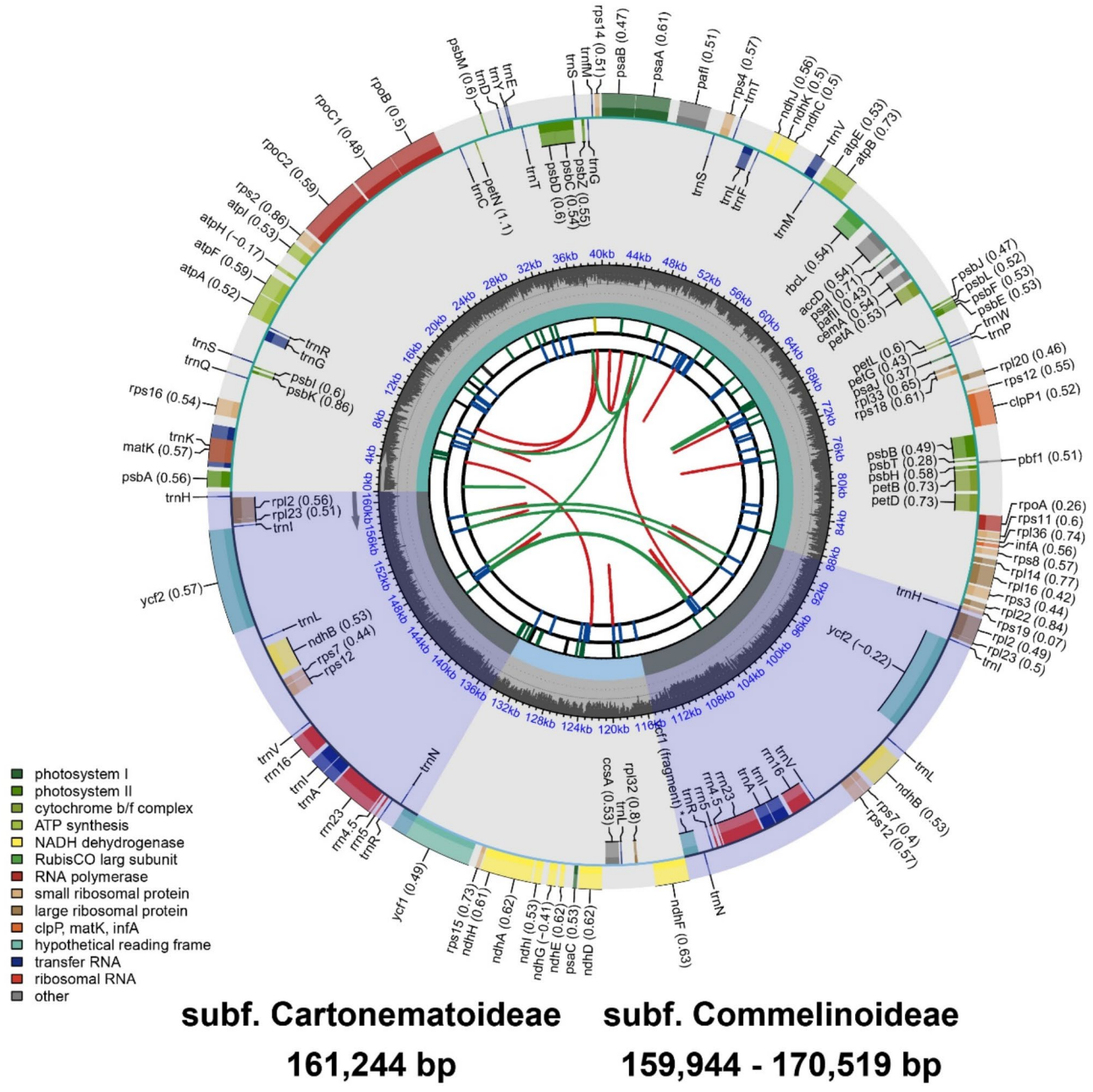

Fig. 2 Complete plastomes of Commelinaceae analyzed in this study. The first circle displays the distributed repeats, connected by red arcs (transcribed clockwise) and green arcs (transcribed counterclockwise), extending outward from the center. The second circle represents the tandem repeats, marked with short bars. The third circle shows the microsatellite sequences, also marked with short bars. The fourth circle indicates the size of the LSC and SSC regions. The fifth circle represents the IRa and IRb regions. The sixth circle illustrates the GC content within the plastome. The seventh circle categorizes genes with different colors based on their functional groups

Comparative plastome structure

To gain a better understanding of structural variations, we analyzed newly assembled plastomes alongside 18 previously published plastomes from Commelinaceae. The plastid protein-coding regions were conserved across the taxa, whereaes the non-coding regions were highly variable (Fig. 3). In this study, among the newly analyzed taxa, an inversion was found in *Floscopa scandens* and *Spatholirion longifolium* (Gagnep.) Dunn in the *rbc*L–*psa*I intergenic spacer (IGS; around 3.4 kb) and *trn*C-GCA–*trn*T-GGU IGS (around 4 kb), respectively. Two inversions were identified in *Cyanotis ciliata* (Blume) Bakh.f., *C. speciosa* (L.f.) Hassk., and *Coleotrype natalensis* C.B.Clarke, one in the *ndh*C–*acc*D IGS (around 8 kb) and another in the *pet*A–*rpo*A IGS (around 16 kb).

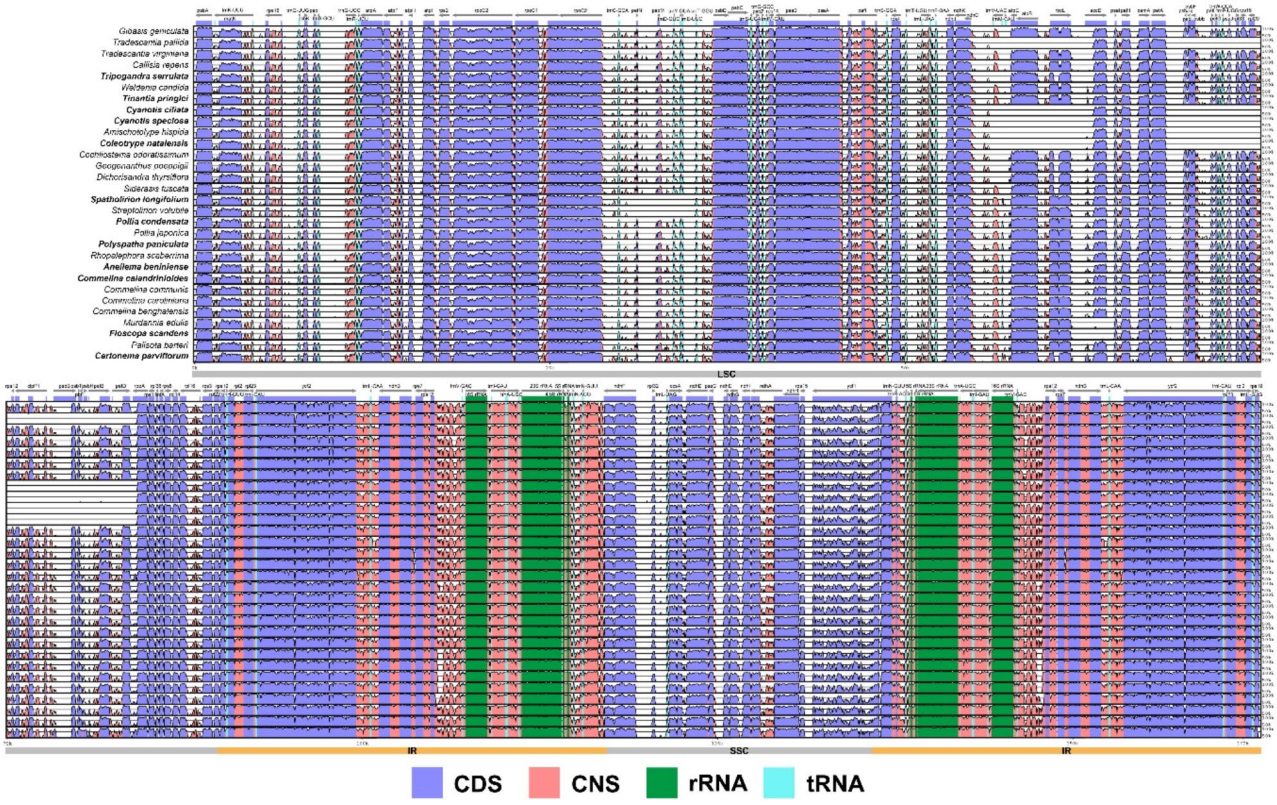

Fig. 3 Plots of percent sequence identity of the plastomes from 30 Commelinaceae taxa, using *Hanguana malayana* (Hanguanaceae) as the reference. The percentage of sequence identity was estimated and visualized using mVISTA. Species names highlighted in bold represent those sequenced in this study

The boundaries between the LSC/IRb (JLB), IRb/SSC (JSB), SSC/IRa (JSA), and IRa/LSC (JLA) were compared (Fig. 4). All the subf. Commelinoideae taxa have duplicated *rps*19 genes, whereas the subf. Cartonematoideae has only one gene due to IR contraction. Additionally, the taxa of the tr. Commelineae have duplicated *rpl*22 genes, with the exception of *Aneilema beniniense* (P.Beauv.) Kunth, due to IR expansion (Fig. 4).

Nucleotide diversity and relative synonymous codon usage of Commelinaceae

We calculated nucleotide diversity (*Pi*) for the 30 Commelinaceae taxa (Fig. 5 and Table S7). The *Pi* values for the protein-coding genes varied from 0.0046 (*rpl*23) to 0.1021 (*acc*D), averaging at 0.0392. As for tRNA and rRNA genes, their *Pi* values ranged from 0 (*trn*V-GAC and *trn*I-GAU) to 0.0326 (*trn*Q-UUG), with an overall average of 0.0071.

In total, 79 plastid protein-coding genes from 30 Commelinaceae accessions were used to measure relative synonymous codon usage (RSCU) (Fig. 6). We excluded stop codons UAA, UAG, and UGA. The total number of codons ranged from 22,408 in *Rhopalephora scaberrima* (Blume) Faden to 22,738 in *Cartonema parviflorum* Hassk. (Table S8). Leucine (L) was the most abundant amino acid (10.31–10.59%), whereas cysteine (C) was the least abundant amino acid (1.12–1.19%) (Table S8). For the codons, AUU, which encodes isoleucine (I), was the most used (953–998), and UGC, which encodes cysteine (C), was the least used (63–76).

Phylogenetic relationships

The alignment matrix of 79 plastid protein-coding genes for 49 ingroups and 11 outgroups showed 70,176 characters, of which 48,152 (68.62%) were constant and 15,885 (22.64%) were parsimony-informative. MP analysis produced the most parsimonious tree (tree length = 53,498; CI = 0.541; RI = 0.776) (Fig. S1). The topologies of the trees from MP, ML, and BI were congruent with over 90% parsimony bootstrap percentages (PBP), 95% SH-like approximate likelihood ratio (SH-aLRT), 95% mean bootstrap percentage (MBP), and 0.95 posterior probabilities (PP) in almost all nodes, except for *Palisota*, which was sister to the subf. Commelinoideae in the MP analysis (Fig. 7 and Fig. S1).

This study once again evidences the non-monophyletic relationships between the two genera, *Cartonema* R.Br. and *Triceratella* (PBP = 100%; SH-aLRT = 99.8%; MBP = 100%; PP = 1.00), which were classified under the subf. Cartonematoideae. The subf. Commelinoideae

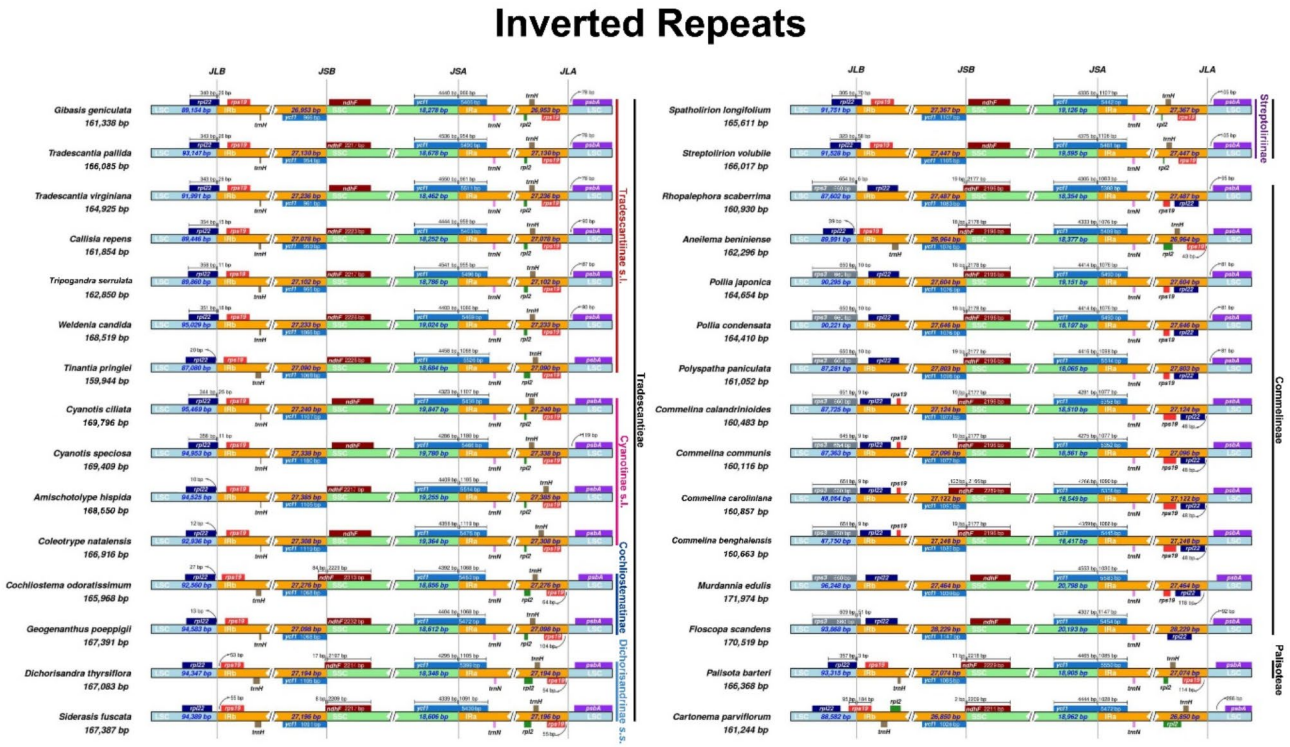

Fig. 4 Comparisons of the boundaries of the LSC, SSC, and IR regions among 30 Commelinaceae taxa

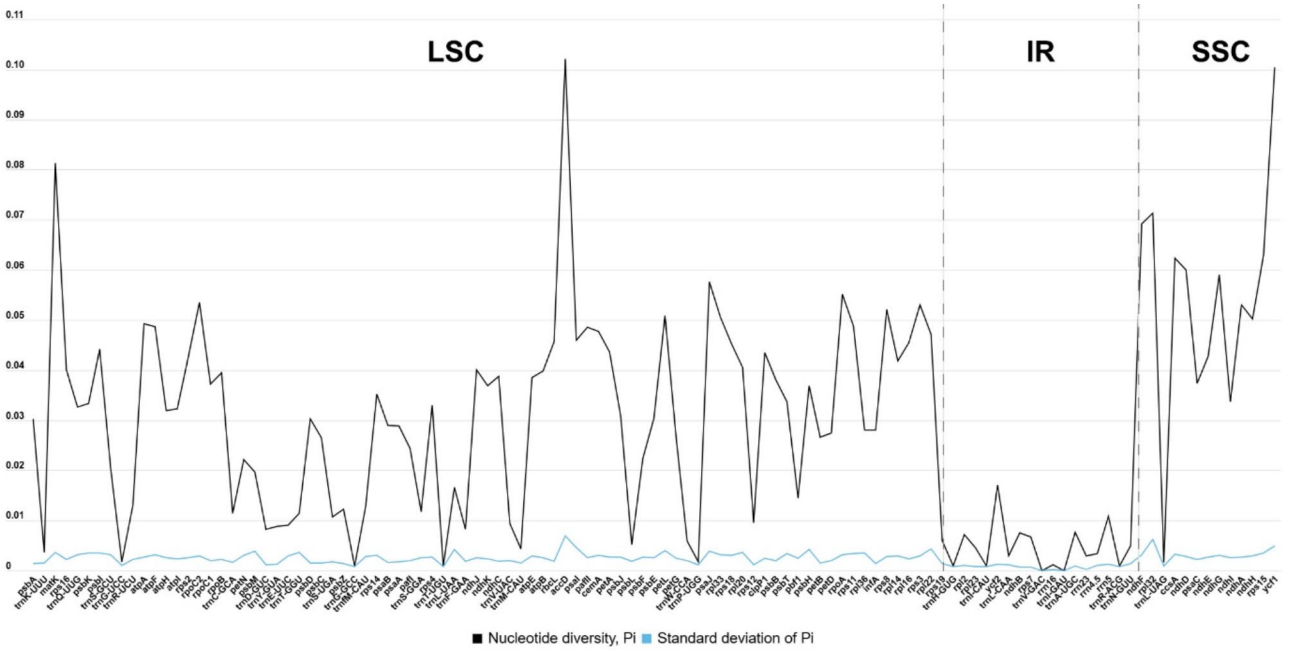

Fig. 5 Nucleotide diversity (*Pi*) based on 79 plastid protein-coding genes from 30 Commelinaceae taxa

was sister to the subf. Triceratelloideae (PBP = 95%; SH-aLRT = 100%; MBP = 100%; PP = 1.00) and further divided into two big clades representing tribes. In the ML and BI analyses, one clade comprised tr. Commelineae + Palisoteae and the other tr. Tradescantieae (SH-aLRT = 100%; MBP = 100%; PP = 1.00). In contrast, the MP analysis recovered a split between tr. Palisoteae and the tr. Commelineae + Tradescantieae (PBP = 100%). Within the tr. Commelineae, *Rhopalephora scaberrima* was found to be nested within the sampled taxa

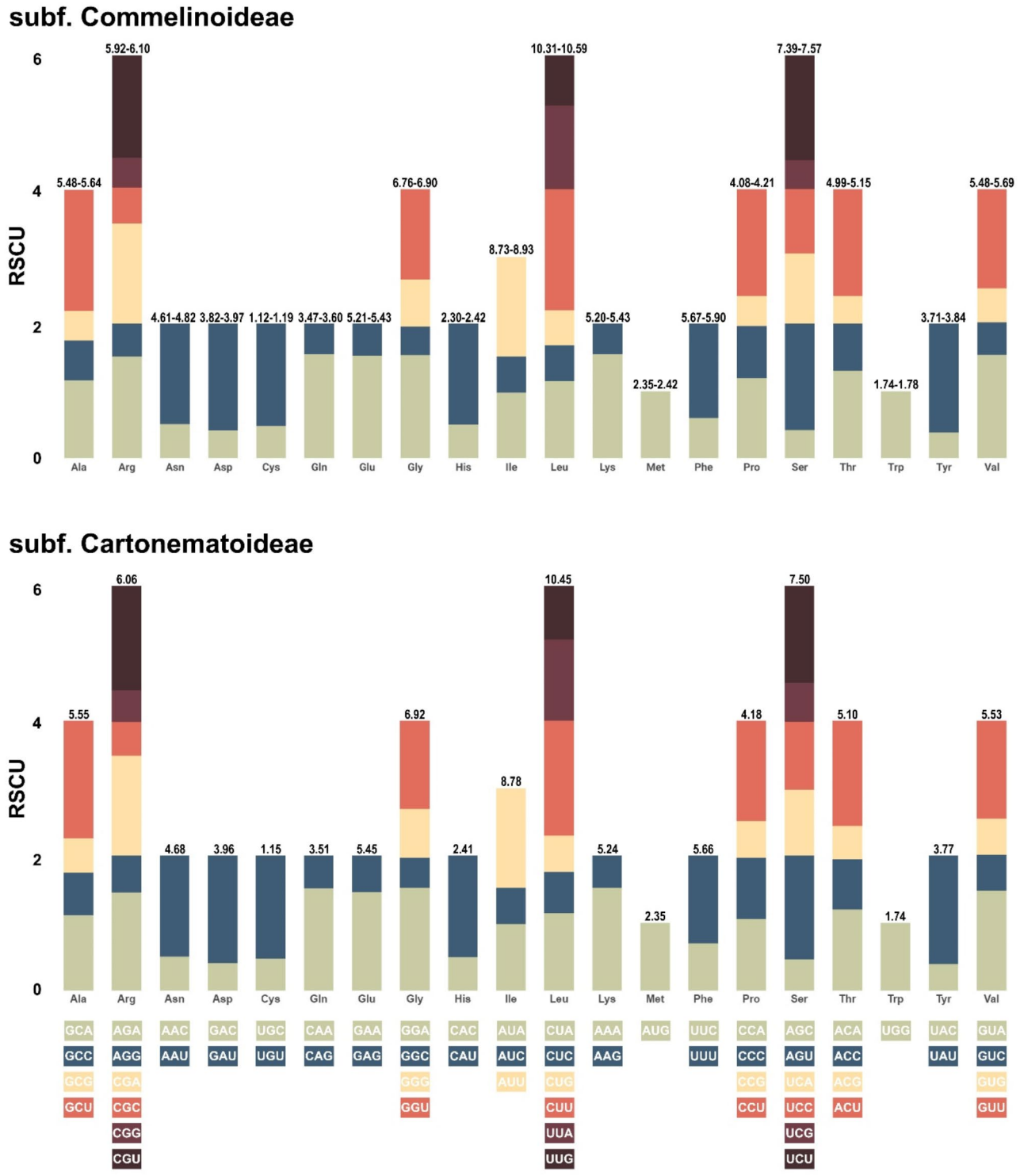

Fig. 6 Relative synonymous codon usage (RSCU) analysis of 20 amino acids across 79 plastid protein-coding genes from the complete plastomes of 30 Commelinaceae taxa. The values at the top of each stack represent the frequency of use for each amino acid

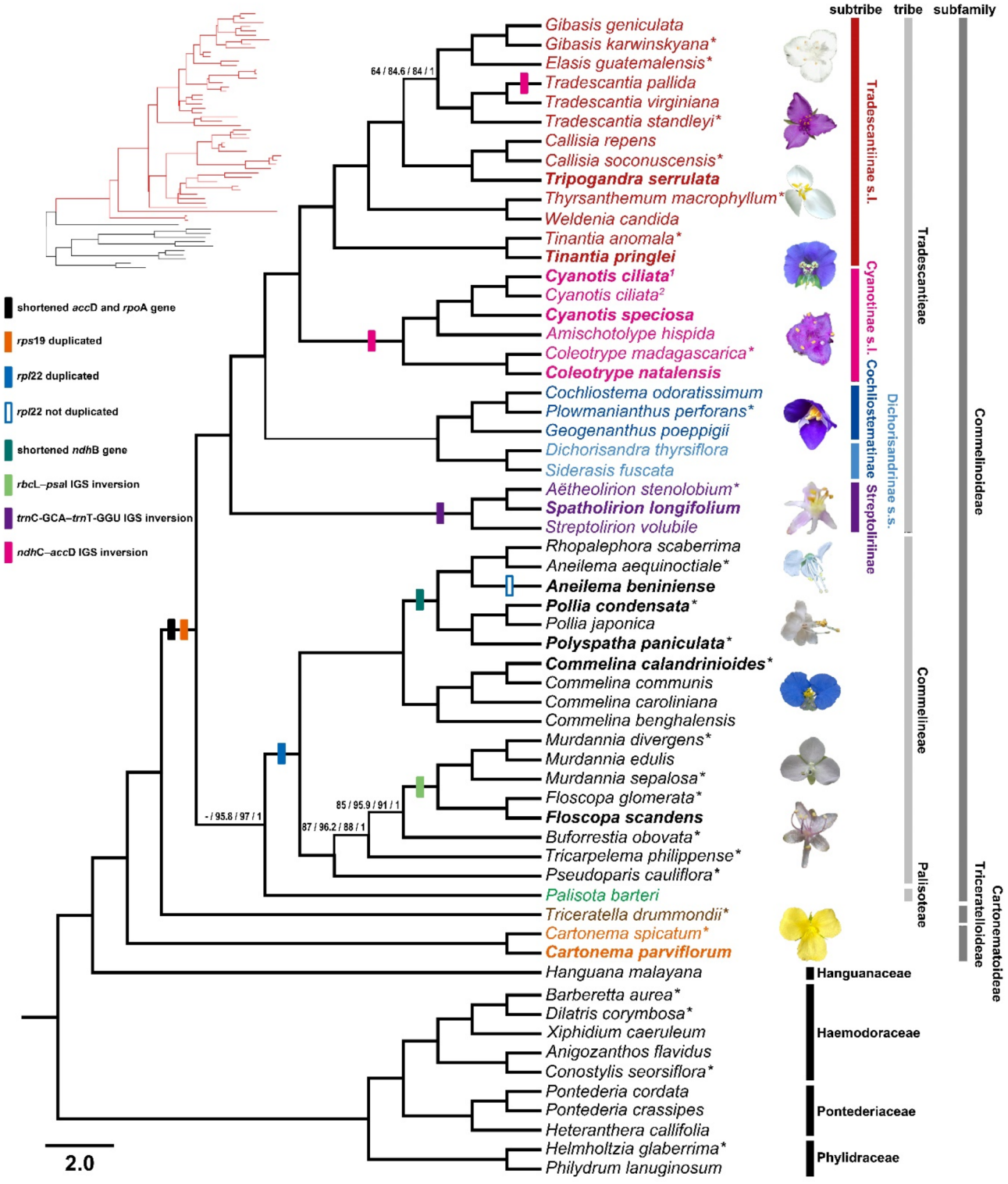

Fig. 7 The maximum likelihood (ML) tree of 60 taxa, inferred from 79 plastid protein-coding genes. Numbers indicate support values (PBP/SH-aLRT/MBP/PP). Only support values of PBP ≤ 90%, SH-aLRT ≤ 95%, MBP ≤ 95%, and PP ≤ 0.95 are shown, with node width reflecting these values. A dash represents incongruence between MP and ML/BI trees. Colored boxes indicate genomic events based on the complete plastomes. Species names in bold are newly sequenced in this study, while those marked with an asterisk (*) are based on SRA data

of *Aneilema*. In the tr. Tradescantieae, the subtr. Streptoliriinae diverged first (PBP = 100%; SH-aLRT = 100%; MBP = 100%; PP = 1.00). We also confirmed that the subtr. Dichorisandrinae *sensu* Faden and Hunt [1] was monophyletic (PBP = 96%; SH-aLRT = 96%; MBP = 96%; PP = 1.00), and the subtr. Coleotrypinae was non-monophyletic, as *Amischotolype* Hassk. clustered with *Cyanotis* D.Don (PBP = 99%; SH-aLRT = 99.4%; MBP = 100%; PP = 1.00) (Fig. 7).

Molecular dating, ancestral range reconstruction, and diversification rates

The distribution of Commelinaceae demonstrates a distinctive pattern across both the Old and New Worlds (Fig. 8). The mean divergence age estimates and 95% highest posterior densities (HPD) for critical nodes from the BEAST analysis are shown in Fig. 9; Table 2. The analysis result suggests that Commelinaceae diverged from other monocots in the Lower Cretaceous period, around 104.77 million years ago (Mya), with a 95% HPD range of 94.17–112.97 (node 1). The subf. Cartonematoideae diverged around 81.39 Mya, with a 95% HPD range of 68.53–96.07 Mya (node 2), while the subf. Triceratelloideae diverged around 74.07 Mya, with a 95% HPD range of 62.17–87.42 (node 3). Within the subf. Commelinoideae, the lineages leading to the tr. Commelineae + tr. Palisoteae and tr. Tradescantieae diverged around 59.01 Mya, with a 95% HPD range of 50.14–68.85 Mya (node 4). The subsequent split between the tr. Commelineae and tr. Palisoteae occurred around 57.45 Mya, with a 95% HPD range of 48.39–67.27 Mya (node 5). Within the tr. Tradescantieae, the subtr. Streptoliriinae originated around 50.44 Mya, with a 95% HPD range of 41.17–60.83 Mya (node 7). The Neotropical subtr. Dichorisandrinae *sensu* Faden and Hunt [1] diverged around 35.47 Mya with a 95% HPD range of 31.81–37.8 Mya (node 9) and subsequently separated around 34.18 Mya with a 95% HPD range of 30.26–37.40 Mya (node 10). The subtr. Cyanotinae s.l. (including the subtr. Coleotrypinae) and subtr. Tradescantiinae s.l., which exhibit a vicariance distribution between the Old and New Worlds, diverged approximately 32.26 Mya, with a 95% HPD range of 28.37–35.88 Mya (node 13).

The BBM-derived predictions of ancestral distributions for internal nodes within Commelinaceae are clear than those from BioGeoBEARS, so we have presented them

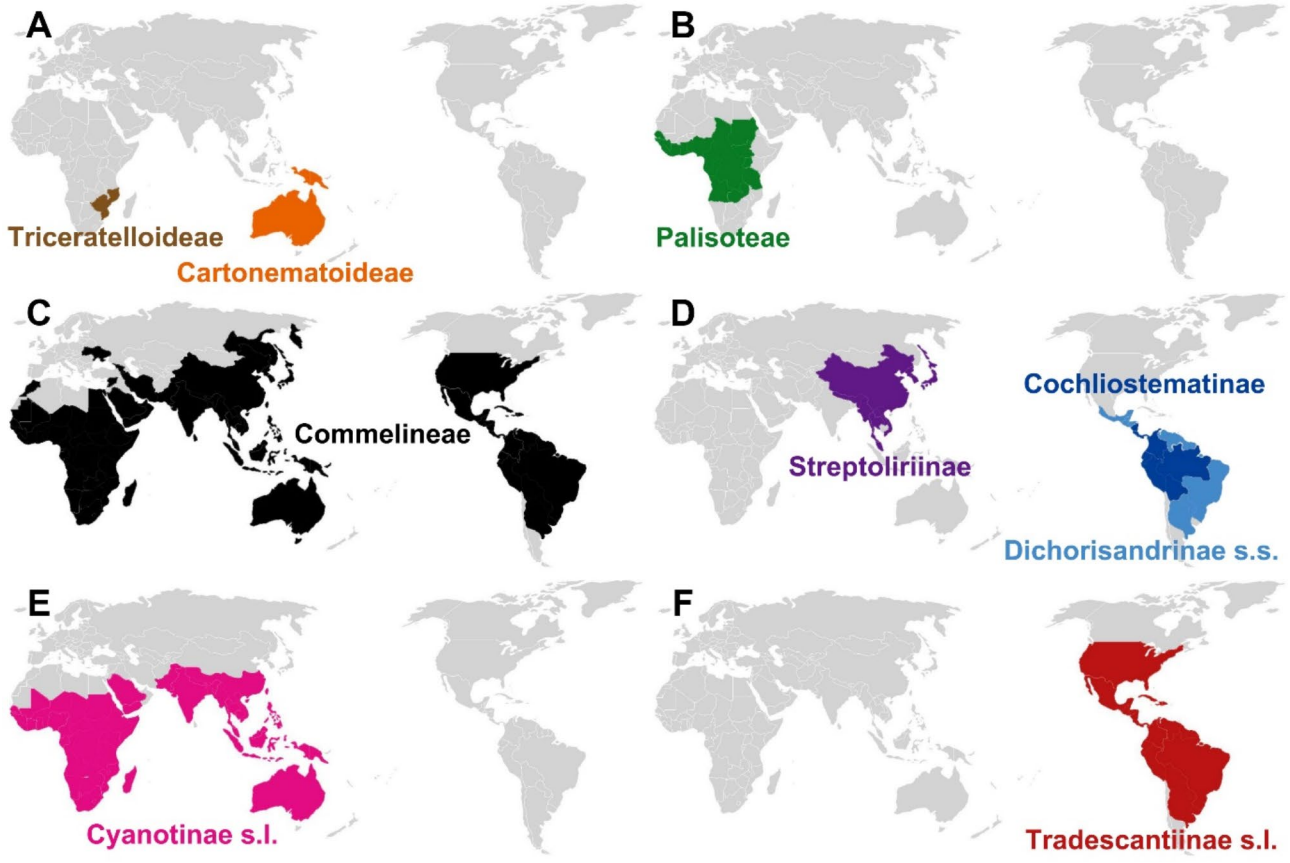

Fig. 8 The geographical distribution of Commelinaceae is illustrated, highlighting distinct regions: (**A**) represents the subf. Cartonematoideae and Triceratelloideae; (**B**–**F**) depict the subf. Commelinoideae with (**B**) tr. Palisoteae; (**C**) tr. Commelineae; (**D**–**F**) representing the tr. Tradescantieae with (**D**) subtr. Streptoliriinae, Cochliostematinae, and Dichorisandrinae s.s.; (**E**) subtr. Cyanotinae s.l.; and (**F**) subtr. Tradescantiinae s.l

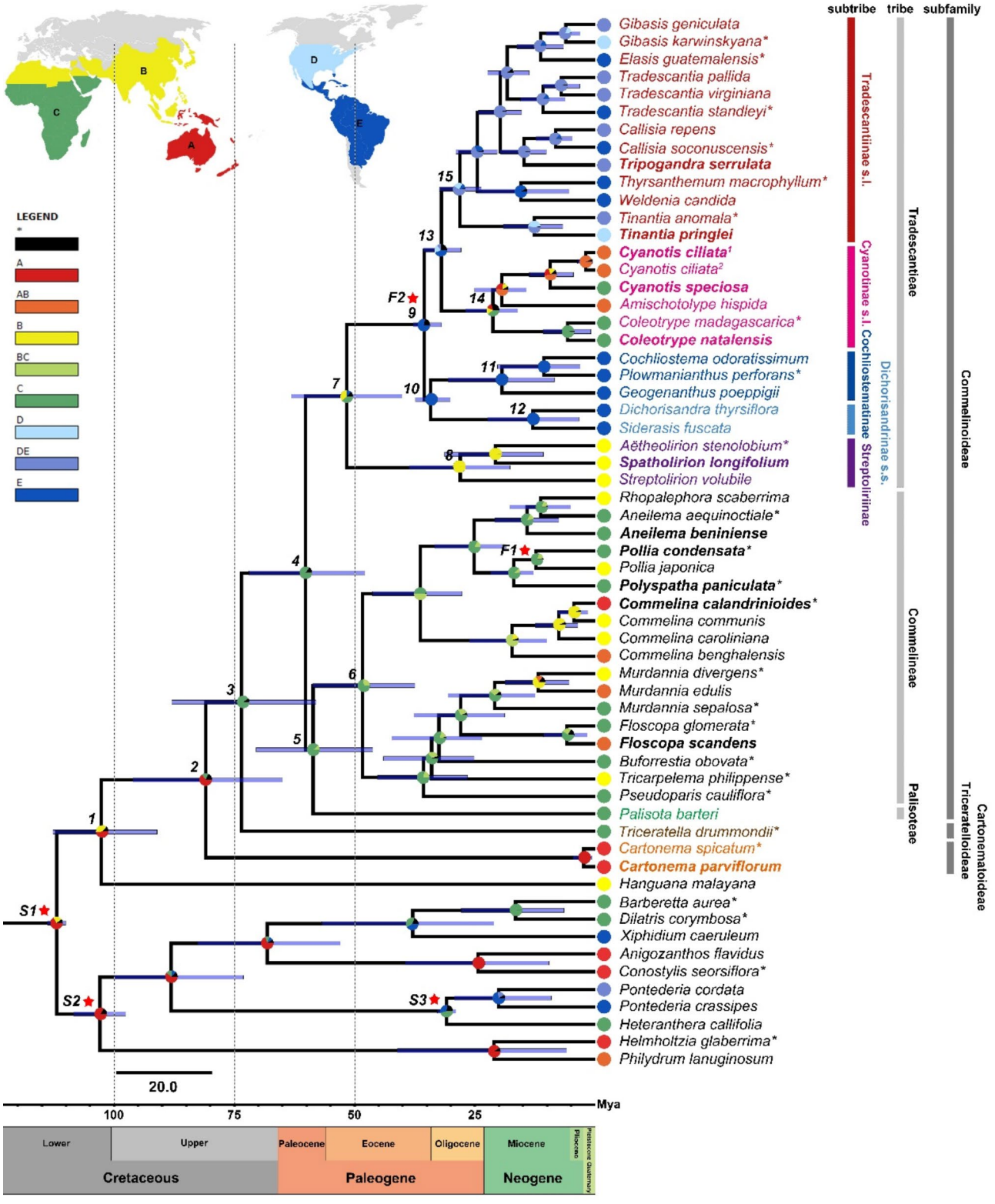

Fig. 9 The summary presents the Bayesian binary method (BBM) model for reconstructing ancestral areas in Commelinaceae, based on a reduced BEAST chronogram of 79 plastid protein-coding genes. The BBM ancestral area reconstructions are displayed for each clade, representing the highest likelihood scenarios. The biogeographic regions used in the BBM model are labeled as follows: (**A**) Australian region; (**B**) Oriental + Palearctic region; (**C**) Ethiopian region; (**D**) Nearctic region; and (**E**) Neotropical region. Red stars and numbers 1–15 indicate specific fossil data and key nodes, respectively. F1–F2, fossil calibration point; S1–S3, secondary calibration point

Table 2 The estimated age and reconstructed ancestral area of key nodes of Commelinaceae

Node	Description	Age estimate		Ancestral and reconstruction	
		Mean (Mya)	95% HPD (Mya)	BBM (%)	BioGeoBEARS (%)
1	Commelinaceae + Hanguanaceae	104.77	94.17–112.97	A (47.56) B (39.01)	C (24.30) B (16.59) A (14.97) BC (13.88) AC (12.05)
2	Commelinaceae	81.39	68.53–96.07	A (64.75) C (16.81)	AC (39.16) C (23.43) BC (13.29)
3	Triceratelloideae	74.07	62.17–87.42	C (90.05)	C (62.93) BC (20.05)
4	Commelinoideae	59.01	50.14–68.85	C (83.85)	BC (44.64) C (42.30)
5	Commelineae + Palisoteae	57.45	48.39–67.27	C (87.58) BC (11.20)	C (80.64) BC (14.95)
6	Commelineae	47.09	37.33–56.28	C (73.30) BC (26.02)	C (70.13) BC (25.78)
7	Tradescantieae	50.44	41.17–60.83	C (38.17) B (37.47) E (11.38)	BC (31.40) B (28.68) C (18.60) BE (10.05)
8	Streptoliriinae	25.95	13.57–37.39	B (99.06)	B (83.00) BC (12.61)
9	Cochliostematinae + Dichorisandrinae s.s. + Cyanotinae s.l.+ Tradescantiinae s.l.	35.47	31.81–37.8	E (78.03)	CE (57.44) BE (34.91)
10	Cochliostematinae + Dichorisandrinae s.s.	34.18	30.26–37.40	E (98.68)	E (99.53)
11	Cochliostematinae	20.65	10.79–31.90	E (99.68)	E (99.14)
12	Dichorisandrinae s.s.	13.71	3.37–24.99	E (99.71)	E (99.01)
13	Cyanotinae s.l.+ Tradescantiinae s.l.	32.26	28.37–35.88	E (44.67) D (20.58) DE (9.89)	CE (59.91) BE (35.45)
14	Cyanotinae s.l.	21.94	16.08–28.12	C (40.67) B (15.49) A (14.93)	BC (49.99) C (25.79) AC (19.93)
15	Tradescantiinae s.l.	28.42	24.30–32.46	DE (57.89) D (25.96) E (13.70)	E (91.98)

in Fig. 9; Table 2. The Australian region (A) is proposed as the most likely place of origin for Commelinaceae and subf. Triceratelloideae, with probabilities of 47.56% and 64.75%, respectively (nodes 1 and 2). Following the Australian region, the BBM reconstruction suggests that the Ethiopian region (C) is the probable origin of the subf. Commelinoideae (node 3), with a probability of 90.05%. Within the tr. Commelineae + Palisoteae, the Ethiopian region is identified as the most likely area where significant taxonomic differentiation occurred (node 5). In the tr. Tradescantieae, both the subtr. Streptoliriinae and Dichorisandrinae *sensu* Faden and Hunt [1] exhibited long-distance dispersal from the Ethiopian region, with the subtr. Streptoliriinae migrating to the Oriental + Palearctic region (B) and the subtr. Dichorisandrinae to the Neotropical region (E). A transition from the Neotropical region back to the Ethiopian region is observed in the crown node of subtr. Cyanotinae s.l. + Tradescantiinae s.l.

Discussion

Plastome structure of Commelinaceae and comparative analyses

Following recent studies, we annotated the *ycf*3 and *ycf*4 genes, which encode the assembly factors for the photosystem I complex, as the *paf*I and *paf*II gene, respectively, and *psb*N gene, which encodes the photosystem biogenesis factor 1, as *pbf*1 gene [49, 50]. Compared with published plastomes of Commelinaceae, the genus *Cartonema*, the first lineage to diverge within Commelinaceae (Fig. 7), encodes 79 plastid protein-coding genes similar to most angiosperms. However, in the subf. Commelinoideae, we identified small indels in the *acc*D and *rpo*A genes, which could potentially impact their functions. The *acc*D gene encodes the beta-carboxyltransferase subunit of acetyl-CoA carboxylase, essential for cell survival and leaf development. Therefore, knocking it down limits cellular lipids affecting seed development and chloroplast division [51–53]. The *rpo*A gene encodes the alpha subunit of RNA polymerase, which is also essential for chloroplast development and expression of most genes related to photosynthesis [54]. We checked whether these were pseudogenes by verifying the existence of an open reading frame (ORF) with a conserved domain, using the NCBI Conserved Domains Database (CDD) following the method outlined by Lu et al. [55]. We found that, despite a delayed start or the loss of the terminal region in the ORF, both *acc*D and *rpo*A genes still preserve a complete functional domain with a start codon. A recent study demonstrated parallels between the swift structural changes in the *acc*D gene and the differing ORF lengths observed in the Caprifoliaceae Juss. s.l [56], and those found in Commelinaceae. A previous study mentioned the pseudogene *ycf*15 in Commelinaceae [14], however, it is disabled and lost in many angiosperms with no function, so we excluded it from this study [57]. Also, termination codon was detected in the front part of the *ndh*B gene in *Pollia japonica* Thunb. and *Rhopalephora scaberrima* [14]. We verified that these taxa maintain a complete functional domain with an ORF, thereby conclusively determining that it is not a pseudogene. In this study, we identified the same event in the plastome of *Aneilema beniniense*, *Pollia condensata*, and *Polyspatha paniculata*, which form a clade together (Fig. 7).

Several inversions were detected in the studied taxa of Commelinaceae, which are common events in numerous plant plastomes [13]. Additionally, the subtr. Streptoliriinae and Cyanotinae s.l. were found to exhibit independent inversions. This event also occurs independently within *Tradescantia* L. (Fig. 3). Gene duplication is an important event in genome evolution. The expansion and contraction of IR can affect the total number of plastid protein-coding genes, with some of these events being phylogenetically informative [58, 59]. All taxa within the subf. Commelinoideae have duplicated *rps*19 genes, while the subf. Cartonematoideae has only one. The sister group of Commelinaceae, *Hanguana malayana* (Hanguanaceae), contains a duplicated *rps*19 gene (Fig. S2). This likely suggests the loss of one copy in *Cartonema* and that a duplication event occurred in the common ancestor of Commelinaceae + Hanguanaceae. Among the three tribes in the subf. Commelinoideae, only the taxa of the tr. Commelineae, except for *Aneilema beniniense*, exhibit duplicated copies of the *rpl*22 gene.

Regarding the genetic divergence, we found that IR regions had lower values than LSC and SSC regions due to their conserved structure (Fig. 5). Codon usage in plastid-protein coding genes is important for inferring the mutation rates, selection pressures, and genetic drift in different taxa. The most commonly used codon was AUU in Commelinaceae, encoding isoleucine (Ile), which is similar to other non-grass monocots [60].

Insights into Commelinaceae phylogenetic relationships

Several molecular phylogenetic studies have been conducted to clarify the ambiguous relationships within the infrafamilial ranks of Commelinaceae [2, 3, 14, 29, 61–64]. Analysis using the plastidial *rbc*L marker places *Cartonema* in a basal clade and supports the monophyly of the tr. Commelineae and Tradescantieae, though the placement of *Palisota* remains weakly supported [61]. In contrast, the plastidial *ndh*F marker suggests paraphyly in the subtr. Tradescantiinae and polyphyly in the subtr. Thyrsantheminae and Dichorisandrinae [62]. The combined analysis of nuclear 5 S non-transcribed spacer (NTS) and plastid *trn*L–F regions indicates a strong relationship between the tr. Commelineae and Tradescantieae, yet the positions of *Palisota* and *Spatholirion* Ridl. remain ambiguous [63]. In a recent study, Zuntini et al. [3] used target sequence capture to recover both plastid and 353 nuclear genes. For their phylogenetic analysis, they excluded two plastid genes, *ycf*1 and *ycf*2, citing "unreliable alignments" as the reason. In the present study, we aligned all plastid protein-coding genes and obtained strongly supported results (Fig. 7).

Genus *Cartonema* diverged first in Commelinaceae and was not recovered as sister to *Triceratella* (Fig. 7), as suggested by Faden and Hunt [1] based on morphological characters. This result further supports the non-monophyly of the subf. Cartonematoideae *sensu* Faden and Hunt [1], as also recovered by Zuntini et al. [3]. Nonetheless, we used the plastidial data for *Triceratella* generated by the aforementioned study, which has nearly 50% (i.e., 48.9%) of missing data in our dataset. Therefore, this placement is still tentative, and the recognition of the new subf. Triceratelloideae appears premature, as supported by the incongruent topologies generated by Zuntini et al. [3], which either support or refute the monophyly of the subf. Cartonematoideae.

Genus *Palisota* has been traditionally included in the tr. Tradescantieae [1]. However, our results place it as sister to the tr. Commelineae in ML and BI analyses (SH-aLRT = 100%; MBP = 100%; PP = 1.00), or as sister to the subf. Commelinoideae in the MP analysis (PBP = 100%). The phylogenetic placement of *Palisota* has remained elusive until recently [61, 62, 65], but its placement as sister to the tr. Commelineae is now widely supported [3, 14, 29, 64, 66]. Despite our MP analysis supporting *Palisota* as sister to the tr. Commelineae and Tradescantieae, it does so with average statistical support. Alternatively, our ML and BI results strongly and unambiguously place *Palisota* as sister to the tr. Commelineae, added to the genus's unique morphology (roots with mucilage canals, leaves with branched and rugose macrohairs, staminodes lacking antherodes, pollen grains dimorphic polliniferous stamens in the same whorl, sulcal membrane ornamentation almost indistinguishable from the tectum, berries, and seedling collar a thick ring [2]), support the recognition of the tr. Palisoteae.

In the tr. Commelineae, *Rhopalephora scaberrima* was found to be nested within *Aneilema* with high support values (PBP = 100%; SH-aLRT = 100%; MBP = 100%; PP = 1.00) (Fig. 7). This result is consistent with previous studies [3, 29, 61], supporting the inclusion of *Rhopalephora* within a broader circumscription of *Aneilema*. Lee et al. [29] subdivided the tr. Commelineae into two subtribes, Commelininae and Murdanniinae, based on the presence of hooked hairs. Our results also support the subtribal relationships within Commelineae, with the exception of two genera, *Dictyospermum* and *Stanfieldiella*, due to the lack of sampling. Currently, while the hair characteristic is used as a diagnostic feature for classifying the two subtribes, its reliability remains somewhat suspect, and additional distinguishing features need to be identified in the future.

Currently, five subtribes (i.e., Streptoliriinae, Coleotrypinae, Cyanotinae, Dichorisandrinae *sensu* Faden and Hunt [1], and Tradescantiinae s.l.) are accepted within the tr. Tradescantieae based on their morphological characters [1, 67]. Subtr. Streptoliriinae, characterized by a climbing habit, was recovered by us as a basal branch of the tr. Tradescantieae with high support values

(PBP = 100%; SH-aLRT = 100%; MBP = 100%; PP = 1.00) (Fig. 7). Although the dataset comprising five nuclear genes suggested *Aëtheolirion* as sister to the subf. Commelinoideae, analysis of whole plastid protein-coding genes identified it as sister to *Spatholirion*, with both as sister to *Streptolirion* Edgew [3]. Our findings also support the relationships among these genera, aligning with morphological evidence and corroborating the monophyly of the subtr. Streptoliriinae, as proposed by Faden and Hunt [1]. Based exclusively on morphological characters, Pellegrini [2] recovered *Streptolirion* as sister to *Aëtheolirion*, supported by characteristics such as a climbing habit, leaf-blades with posterior divisions, axillary inflorescences, and basal 1(−2) cincinni subtended by bracteose cincinnus bracts. For instance, the relationship recovered by us is potentially supported by inflorescences not arranged into a synflorescence, inflorescences always leaf-opposite, and not perforating the subtending leaf-sheath.

Subtr. Dichorisandrinae *sensu* Faden and Hunt [1] comprises five genera (*Cochliostema* Lem., *Dichorisandra* J.C.Mikan, *Geogenanthus* Ule, *Plowmanianthus* Faden & C.R.Hardy, and *Siderasis* Raf.) distributed in the Neotropical regions [1, 68]. In this study, it was recovered as monophyletic with high support values (PBP = 96%; SH-aLRT = 96%; MBP = 96%; PP = 1.00) (Fig. 7). This result contrasts with findings based on partial plastid genes and recent concatenated plastid gene datasets, which indicate that *Dichorisandra* is most closely related to *Siderasis* or *Geogenanthus*, with the remaining genera forming a separate clade [3, 61, 62, 64]. However, studies based on 353 nuclear genes and three partial plastid genes do support the monophyly of this group [3, 29]. Subtr. Dichorisandrinae *sensu* Faden and Hunt [1] is not morphologically cohesive, as there are no shared characters among the five genera except for the number of chromosomes [2, 68]. Subtr. Cochliostematinae was proposed by Lee et al. [29] based on the morphological characters and discussion presented by Pellegrini and Faden [68]. This proposal aimed to address the inconsistent monophyly of the subtr. Dichorisandrinae *sensu* Faden and Hunt [1]. Despite our results indicating a monophyletic relationship, we believe that recognizing subtr. Cochliostematinae, as recommended by Pellegrini and Faden [68], represents the most parsimonious and least systematically disruptive approach.

Our results show that *Amischotolype* s.l. (including *Porandra* D.Y.Hong) and *Cyanotis* s.l. (including *Belosynapsis* Hassk.) form a clade that is sister to *Coleotrype* C.B.Clarke, making the subtr. Coleotrypinae non-monophyletic. Pellegrini [2] recovered the same relationship based on morphological characters, supporting the reduction of the subtr. Coleotrypinae to a synonym of the Cyanotiinae. However, analyses based on nuclear genes revealed that *Cyanotis* s.l. occupies a basal position among the genera [3], highlighting conflicting phylogenetic results between plastome and nuclear datasets. Nonetheless, the phylogenetic relationships among them are supported by genome evolution, specifically by two large inversions in plastome sequences shared by these three genera. Subtr. Coleotrypinae was proposed by Faden and Hunt [1], not based on a single unique morphological character, but rather on a combination of character states. Faden and Beentje [69] noted that *Coleotrype* has arillate seeds, a feature shared with *Amischotolype* s.l., whereas *Cyanotis* s.l. lacks this trait, highlighting the need to identify a synapomorphy for these groups. Zuntini et al. [3] also proposed that no subtribal divisions should be recognized within the tr. Tradescantieae, citing the unclear and unstable relationships observed in previous studies.

Finally, subtr. Tradescantiinae s.l. was recovered as monophyletic (PBP = 100%; SH-aLRT = 100%; MBP = 100%; PP = 1.00), in accordance with the decision by Pellegrini [67] to reduce the subtr. Thyrsantheminae to a synonym of the subtr. Tradescantiinae. The relationships within the subtr. Tradescantiinae s.l. were supported by high support values, except for *Gibasis* Raf. + *Elasis* D.R.Hunt with *Tradescantia* (PBP = 64%; SH-aLRT = 84.6%; MBP = 84%; PP = 1.00) (Fig. 7). Previous studies have shown that the relationships within the core subtr. Tradescantiinae still require further investigation [2, 3, 14, 29, 61–64, 67]. For instance, earlier molecular studies were limited by insufficient sampling [61, 62], the number of markers employed [29, 61–64], or had issues with the reliability of the taxonomic determinations or the quality of the generated sequences [3, 64]. At present, our results align with the generic boundaries proposed in recent studies [2, 67, 70]. In the case of *Callisia* Loefl., both molecular and morphological studies have demonstrated its paraphyletic relationship with *Tripogandra* Raf [29, 61, 64, 67, 71]. All other lineages recently lumped to form the morphologically elusive *Callisia sensu* Christenhusz et al. [72] based on two references (Evans et al. [61] and Hertweck and Pires [64]), (i.e., *Tripogandra* and the remaining sections of *Callisia* accepted by Hunt [73]), still require genomic data with sufficient sampling to clarify their relationships.

Historical biogeography of Commelinaceae

In our study, we used two fossils and three estimated divergence times as calibration points, which were derived from genomic data [28, 40, 41]. Other studies that calibrated phylogenies of Commelinid have reported similar divergence time estimates [74, 75]. Our research proposes that Commelinaceae likely appeared around 104.77 Mya. Though this estimate aligns with the timeline suggested by Givnish et al. [40] and Zuntini et al.

[28], it significantly predates the time of emergence suggested by both Hertweck et al. [76] and Pellegrini [2] by over 20 Mya. Given the extensive species sampling and markers used, our study provides a reliable estimation of the divergence timeline for Commelinaceae.

Our biogeographical studies suggest that four critical long-distance migrations significantly impacted the evolutionary development of Commelinaceae. These events took place well after the split of Gondwana (Fig. 9). These shifts involved the movement of plant species across various continents, leading to the formation of several subfamilies, tribes, and subtribes within Commelinaceae. It appears that a noteworthy shift from the Australian to the Ethiopian region led to the formation of the subf. Triceratelloideae.

The division of the subf. Triceratelloideae is essential for understanding the genesis of the subf. Commelinoideae and the long-distance dispersal event from the Australian to the Ethiopian region during the Upper Cretaceous era. Concurrently, the Australian and the Neotropical regions were indirectly linked through Antarctica, while the globe was undergoing a period of elevated temperatures due to an increase in atmospheric CO_2 levels [77]. Our findings from the RASP analysis suggest a reconstructed geographic route with a probability of 58.0%, denoted as 'Australian (A)→Australian (A) Ethiopian (C)→ Australian (A)|Ethiopian (C)'. This indicates that the ancestral lineage dispersed from the Australian region to the Ethiopian region, leading to vicariance among the resulting offspring. This analysis aligns with the Eastern Gondwanan origin of Commelinaceae, as suggested by previous studies [61, 78].

The early Paleocene era, approximately 60 Mya, was a pivotal period for the subf. Commelinoideae. During this time, two main groups were formed: the tr. Commelineae + Palisoteae and the tr. Tradescantieae. Around nine Mya later, the common ancestor of the tr. Tradescantieae migrated from the Ethiopian to the Oriental + Palearctic region. This marked the first known appearance in Eurasia and led to the creation of the common ancestors of the subtr. Streptoliriinae. The third significant spread of species appears to have originated in the Oriental + Palearctic region and moved overland to the Neotropical region. These migration notably correlated with the development of the remaining four subtribes. During the Paleocene–Eocene Thermal Maximum (PETM), which lasted until around 56 Mya, the earth's surface experienced warming due to increased levels of atmospheric CO_2 [79]. In contrast, cooling trends during the Oligocene may have promoted southward migration within these regions, leading to the current disjunct distribution between East Asia and North America. This environmental shift caused many species to move, rapidly evolve, and adjust their feeding habits [79], as supported by several previous studies. A comparable spread of species was observed in the common ancestor of the tr. Caladieae (Araceae Juss.), which commenced in East Asia and moved to South America around 58 Mya [80]. Likewise, Givnish et al. [81] discovered that a substantial group associated with the Pleurothallids (Orchidaceae Juss.) originated in Southeast Asia, with significant migrations from Asia occurring around 32 Mya. Finally, the subtr. Cyanotinae s.l. seems to have migrated from the Neotropical to the Ethiopian region roughly 32 Mya, via long-distance dispersal.

During the Cenozoic epochs, the Old and New Worlds were linked by two key plant migration routes: the North Atlantic Land Bridge (NALB) and the Bering Land Bridge (BLB) [82, 83]. Despite their varying periods of activity, these pathways profoundly influenced the development of contemporary Northern Hemisphere flora. We posit that the BLB was the most probable route for plant migrations from the Oriental + Palearctic regions to the Neotropical region during the Paleocene–Eocene periods. Modern biogeographical studies suggest that the colonization process from East Asia to North America via the BLB may have begun around 50 Mya, as seen in the case of *Clintonia* Raf. + *Medeola* Gronov. ex L. (Liliaceae Juss.) [84]. Conversely, the migration from South America to Southeast Asia through the BLB could have commenced nearly 46 Mya, as suggested by the genus *Paphiopedilum* Pfitzer (Orchidaceae) [81]. Similarly, the family Melanthiaceae Batsch ex Borkh. experienced seven distinct migration events during the Oligocene and Miocene–Pliocene epochs via the BLB [85].

Conclusions

Whole plastome sequences capture genetic events, such as pseudogenisation, inversions, and IR contraction/expansion, which help infer genomic evolution. Our phylogenetic results, well-resolved using SRA data, support previous taxonomic and systematic changes, while also proposing new revisions based on our findings and morphological characteristics. Systematic changes need to be made based on a complete understanding of the studied group, also taking into consideration different datasets and methodologies used by previous authors. Aside from greatly contributing to understanding the genomic evolution of Commelinaceae, our present study also represents the first step for a robust, stable, and integrated updated classification for the family. The biogeographic history of Commelinaceae was revealed to be more complex than previously hypothesized. The East Gondwanan origin of the family was corroborated, with the MRCA of Commelinaceae arose in the Australian region ca. 104.77 Mya, when Gondwana was just beginning to breakup. Additionally, four major intercontinental dispersal events were recovered from 81 to 32 Mya. This work contributes

to the foundational understanding of the biogeographical history of Commelinaceae across five major continents, excluding Europe, and will support future studies on dispersal events at the generic level. Furthermore, research focusing on a genus with fewer than five species within Commelinaceae will provide insights into their evolutionary pathway and aid in conservation efforts.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-025-06504-y.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Supplementary Material 4

Supplementary Material 5

Supplementary Material 6

Supplementary Material 7

Supplementary Material 8

Supplementary Material 9

Supplementary Material 10

Acknowledgements

The authors would like to thank Benny Bytebier and Syd Ramdhani at the Bews Herbarium (Republic of South Africa), Chien-Ti Chao at National Taiwan Normal University (Taiwan), and Gillian Brown at Queensland Herbarium (Australia) for collecting and providing the plant material for this study; and Hoang Dang Khoa Do at Nguyen Tat at Thanh University (Vietnam) for their valuable comments on an early version of the manuscript.

Author contributions

JJ collected the plant materials, performed the experiments, analysed the data, prepared figures and tables, and wrote the initial draft. MOOP designed the experiments, collected the plant materials, designed the sampling, and co-wrote the manuscript. 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 study was supported by the Gachon University Research Fund of 2019 (GCU-2019-0821) and the National Research Foundation of Korea (NRF) Grant Fund (NRF-2017R1D1A1B06029326).

Data availability

The twelve plastome sequences we obtained from this study were archived in NCBI. The accession numbers are presented in Table 1 (OP758345–OP758353 and BK062833–BK062835).

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.

Author details

[1]Department of Life Sciences, Gachon University, Seongnam-si, Republic of Korea
[2]Division of Forest Biodiversity, Korea National Arboretum, Pocheon-si, Republic of Korea
[3]Royal Botanic Gardens, Kew, Richmond, UK

Received: 8 November 2024 / Accepted: 2 April 2025
Published online: 25 April 2025

References

1. Faden RB, Hunt D. The classification of the Commelinaceae. Taxon.1991;40:19–31.
2. Pellegrini MOO. Systematics of Commelinales focusing on Neotropical lineages. Doctoral dissertation. Instituto de Biociências São Paulo, Brazil; 2019.
3. Zuntini AR, Frankel LP, Pokorny L, Forest F, Baker WJ. A comprehensive phylogenomic study of the monocot order commelinales, with a new classification of Commelinaceae. Am J Bot. 2021;108(7):1066–86.
4. Faden RB. Commelinaceae. In: Flowering Plants· Monocotyledons. Springer; 1998: 109–128.
5. Panigo E, Ramos J, Lucero L, Perreta M, Vegetti A. The inflorescence in Commelinaceae. Flora - Morphology Distribution Funct Ecol Plants. 2011;206(4):294–9.
6. Vita RSB, Menezes NL, Pellegrini MOO, Melo-de-Pinna GFA. A new interpretation on vascular architecture of the cauline system in Commelinaceae (Commelinales). PLoS ONE. 2019;14(6):e0218383.
7. Yang J, Kim S-H, Gil H-Y, Choi H-J, Kim S-C. New insights into the phylogenetic relationships among wild onions (*Allium*, Amaryllidaceae), with special emphasis on the subgenera *Anguinum* and *Rhizirideum*, as revealed by plastomes. Front Plant Sci. 2023;14:1124277.
8. Xue B, Huang E, Zhao G, Wei R, Song Z, Zhang X, Yao G. 'Out of Africa' origin of the pantropical staghorn fern genus *Platycerium* (Polypodiaceae) supported by plastid phylogenomics and biogeographical analysis. Ann Botany. 2024;133(5–6):697–710.
9. Straub SC, Parks M, Weitemier K, Fishbein M, Cronn RC, Liston A. Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. Am J Bot. 2012;99(2):349–64.
10. Dobrogojski J, Adamiec M, Luciński R. The chloroplast genome: a review. Acta Physiol Plant 2020, 42(6).
11. Sugiura M. The chloroplast genome. Plant Mol Biol. 1992;19(1):149–68.
12. He J, Yao M, Lyu R-D, Lin L-L, Liu H-J, Pei L-Y, Yan S-X, Xie L, Cheng J. Structural variation of the complete chloroplast genome and plastid phylogenomics of the genus *Asteropyrum* (Ranunculaceae). Sci Rep. 2019;9(1):1–13.
13. Jin D-P, Choi I-S, Choi B-H. Plastid genome evolution in tribe Desmodieae (Fabaceae: Papilionoideae). PLoS ONE. 2019;14(6):e0218743.
14. Jung J, Kim C, Kim J-H. Insights into phylogenetic relationships and genome evolution of subfamily Commelinoideae (Commelinaceae Mirb.) inferred from complete chloroplast genomes. BMC Genomics. 2021;22(1):1–12.
15. Craw RC, Grehan JR, Heads MJ. Panbiogeography: tracking the history of life. Oxford University Press, 1999.
16. Sanmartín I. Event-based biogeography: integrating patterns, processes, and time. Biogeogr Chang World 2007:135–59.
17. Doyle J, Doyle J. CTAB DNA extraction in plants. Phytochemical Bull. 1987;19:11–5.
18. 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):1–31.
19. 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.
20. 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.
21. Chan PP, Lowe TM. tRNAscan-SE: searching for tRNA genes in genomic sequences. In: Gene Prediction. Springer; 2019: 1–14.

22. Liu S, Ni Y, Li J, Zhang X, Yang H, Chen H, Liu C. CPGView: a package for visualizing detailed chloroplast genome structures. Mol Ecol Resour. 2023;23(3):694–704.
23. Brudno M, Do CB, Cooper GM, Kim MF, Davydov E, Green ED, Sidow A, Batzoglou S. LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. Genome Res. 2003;13(4):721–31.
24. Amiryousefi A, Hyvönen J, Poczai P. IRscope: an online program to visualize the junction sites of chloroplast genomes. Bioinformatics. 2018;34(17):3030–1.
25. 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.
26. Xia X, Xie Z. DAMBE: software package for data analysis in molecular biology and evolution. J Hered. 2001;92(4):371–3.
27. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357–9.
28. Zuntini AR, Carruthers T, Maurin O, Bailey PC, Leempoel K, Brewer GE, Epitawalage N, Françoso E, Gallego-Paramo B, McGinnie C. Phylogenomics and the rise of the angiosperms. Nature 2024;629(8013):1–8.
29. Lee C-K, Fuse S, Poopath M, Pooma R, Tamura MN. Phylogenetics and infrafamilial classification of Commelinaceae (Commelinales). Bot J Linn Soc. 2022;198(2):117–30.
30. Swofford D. PAUP* 4.0 b. 4a. Phylogenetic analysis using parsimony*(and other methods). Sinauer, Sunderland 2000.
31. Trifinopoulos 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.
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–Tree Figure Drawing Tool Version v. 1.4. 4. Institute of Evolutionary Biology, University of Edinburgh: 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, Hartmann K, Steel M. Stochastic properties of generalised Yule models, with biodiversity applications. J Math Biol. 2008;57:713–35.
37. Jacobs BF, Kabuye CH. An extinct species of *Pollia* Thunberg (Commelinaceae) from the Miocene Ngorora Formation, Kenya. Rev Palaeobot Palynol. 1989;59(1–4):67–76.
38. Poinar GO Jr, Chambers KL. *Pseudhaplocricus hexandrus* gen. et sp. nov. (Commelinaceae) in Mid-Tertiary Dominican amber. J Bot Res Inst Tex 2015;9:353–9.
39. Draper G, Mann P, Lewis JF. Hispaniola. Caribb Geology: Introduction 1994:129–50.
40. Givnish TJ, Zuluaga A, Spalink D, Soto Gomez M, Lam VK, Saarela JM, Sass C, Iles WJ, De Sousa DJL, Leebens-Mack J. Monocot plastid phylogenomics, timeline, net rates of species diversification, the power of multi-gene analyses, and a functional model for the origin of monocots. Am J Bot. 2018;105(11):1888–910.
41. Li ZZ, Gichira AW, Muchuku JK, Li W, Wang GX, Chen JM. Plastid phylogenomics and biogeography of the genus *Monochoria* (Pontederiaceae). J Syst Evol. 2021;59(5):1027–39.
42. POWO. Plants of the world online. Facilitated by the Royal Botanical Gardens, Kew. Published on the Internet. In.; 2024.
43. Wallace A. The Geographical Distribution of Animals. Vol. I & II.,(Harper and Brothers: New York, NY, USA). 1876.
44. Matzke NJ. BioGeoBEARS: BioGeography with Bayesian (and likelihood) evolutionary analysis in R Scripts. *R package, version 02* 2013, 1:2013.
45. Yu Y, Blair C, He X. RASP 4: ancestral state reconstruction tool for multiple genes and characters. Mol Biol Evol. 2020;37(2):604–6.
46. Ree RH, Smith SA. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. Syst Biol. 2008;57(1):4–14.
47. Ronquist F. Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography. Syst Biol. 1997;46(1):195–203.
48. Landis MJ, Matzke NJ, Moore BR, Huelsenbeck JP. Bayesian analysis of biogeography when the number of areas is large. Syst Biol. 2013;62(6):789–804.
49. Wicke S, Schneeweiss GM, Depamphilis CW, Müller KF, Quandt D. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Mol Biol. 2011;76(3):273–97.
50. Krech K, Fu HY, Thiele W, Ruf S, Schöttler MA, Bock R. Reverse genetics in complex multigene operons by co-transformation of the plastid genome and its application to the open reading frame previously designated PsbN. Plant J. 2013;75(6):1062–74.
51. Kode V, Mudd EA, Iamtham S, Day A. The tobacco plastid *acc*D gene is essential and is required for leaf development. Plant J. 2005;44(2):237–44.
52. Bock R. Structure, function, and inheritance of plastid genomes. In: Cell and molecular biology of plastids. Springer; 2007: 29–63.
53. Caroca R, Howell KA, Malinova I, Burgos A, Tiller N, Pellizzer T, Annunziata MG, Hasse C, Ruf S, Karcher D. Knockdown of the plastid-encoded acetyl-CoA carboxylase gene uncovers functions in metabolism and development. Plant Physiol. 2021;185(3):1091–110.
54. Zhang Y, Cui Y-L, Zhang X-L, Yu Q-B, Wang X, Yuan X-B, Qin X-M, He X-F, Huang C, Yang Z-N. A nuclear-encoded protein, mTERF6, mediates transcription termination of *rpo*A polycistron for plastid-encoded RNA polymerase-dependent chloroplast gene expression and chloroplast development. Sci Rep. 2018;8(1):1–12.
55. Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Marchler GH, Song JS. CDD/SPARCLE: the conserved domain database in 2020. Nucleic Acids Res. 2020;48(D1):D265–8.
56. Park S, Jun M, Park S, Park S. Lineage-specific variation in IR boundary shift events, inversions, and substitution rates among Caprifoliaceae sl (Dipsacales) plastomes. Int J Mol Sci. 2021;22(19):10485.
57. Shi C, Liu Y, Huang H, Xia E-H, Zhang H-B, Gao L-Z. Contradiction between plastid gene transcription and function due to complex posttranscriptional splicing: an exemplary study of *ycf*15 function and evolution in angiosperms. PLoS ONE. 2013;8(3):e59620.
58. Liu H, He J, Ding C, Lyu R, Pei L, Cheng J, Xie L. Comparative analysis of complete chloroplast genomes of *Anemoclema*, *Anemone*, *Pulsatilla*, and *Hepatica* revealing structural variations among genera in tribe Anemoneae (Ranunculaceae). *Frontiers in Plant Science* 2018, 9:1097.
59. Guisinger MM, Chumley TW, Kuehl JV, Boore JL, Jansen RK. Implications of the plastid genome sequence of *Typha* (Typhaceae, Poales) for Understanding genome evolution in Poaceae. J Mol Evol. 2010;70(2):149–66.
60. Mazumdar P, Binti Othman R, Mebus K, Ramakrishnan N, Ann Harikrishna J. Codon usage and codon pair patterns in non-grass monocot genomes. Ann Botany. 2017;120(6):893–909.
61. Evans TM, Sytsma KJ, Faden RB, Givnish TJ. Phylogenetic relationships in the Commelinaceae: II. A cladistic analysis of *rbc*L sequences and morphology. Syst Bot. 2003;23:270–92.
62. Wade DJ, Evans TM, Faden RB. Subtribal relationships in tribe Tradescantieae (Commelinaceae) based on molecular and morphological data. Aliso: J Syst Evolutionary Bot. 2006;22(1):520–6.
63. Burns JH, Faden RB, Steppan SJ. Phylogenetic studies in the Commelinaceae subfamily Commelinoideae inferred from nuclear ribosomal and chloroplast DNA sequences. Syst Bot. 2011;36(2):268–76.
64. Hertweck KL, Pires JC. Systematics and evolution of inflorescence structure in the *Tradescantia* alliance (Commelinaceae). Syst Bot. 2014;39(1):105–16.
65. Zuiderveen GH, Evans TM, Faden RB. A phylogenetic analysis of the African plant genus *Palisota* (family Commelinaceae) based on chloroplast DNA sequences. Honors project. Grand Valley State University, USA; 2011.
66. Crum AH. Phylogenetic Analysis of *Palisota* (Commelinaceae) Using Chloroplast and Nuclear Regions. Master thesis. Grand Valley State University, USA; 2019.
67. Pellegrini MOO. Morphological phylogeny of *Tradescantia* L. (Commelinaceae) sheds light on a new infrageneric classification for the genus and novelties on the systematics of subtribe Tradescantiinae. PhytoKeys. 2017;89:11–72.
68. Pellegrini MOO, Faden RB. Recircumscription and taxonomic revision of *Siderasis*, with comments on the systematics of subtribe Dichorisandrinae (Commelinaceae). PhytoKeys 2017;83:1–41.
69. Faden R, Beentje HJ. Flora of tropical East Africa: Commelinaceae. Royal Botanic Gardens, Kew; 2012.
70. Pellegrini MOO, Espejo-Serna A. Recircumscription and synopsis of *Thyrsanthemum* and *Weldenia* (Commelinaceae), two narrow endemic genera from Mesoamerica. Kew Bull. 2021;76(2):269–86.
71. Bergamo S. A phylogenetic evaluation of *Callisia* Loefl. (Commelinaceae) based on molecular data. Doctoral dissertation. University of Georgia, USA; 2003.

72. Christenhusz MM, Fay M, Byng JW. The global flora: special edition: GLOVAP nomenclature part 1. Volume 4. Plant Gateway Ltd.; 2018.
73. Hunt D. Amplification of *Callisia* Loefl.: American Commelinaceae: XV. Kew Bull 1986:407–12.
74. Eguchi S, Tamura MN. Evolutionary timescale of monocots determined by the fossilized birth-death model using a large number of fossil records. Evolution. 2016;70(5):1136–44.
75. Bell CD, Soltis DE, Soltis PS. The age and diversification of the angiosperms re-revisited. Am J Bot. 2010;97(8):1296–303.
76. Hertweck KL, Kinney MS, Stuart SA, Maurin O, Mathews S, Chase MW, Gandolfo MA, Pires JC. Phylogenetics, divergence times and diversification from three genomic partitions in monocots. Bot J Linn Soc. 2015;178(3):375–93.
77. Pross J, Contreras L, Bijl PK, Greenwood DR, Bohaty SM, Schouten S, Bendle JA, Röhl U, Tauxe L, Raine JI. Persistent near-tropical warmth on the Antarctic continent during the early eocene epoch. Nature. 2012;488(7409):73–7.
78. Givnish TJ, Evans TM, Pires JC, Sytsma KJ. Polyphyly and convergent morphological evolution in Commelinales and Commelinidae: evidence from *rbc*L sequence data. Mol Phylogenet Evol. 1999;12(3):360–85.
79. McInerney FA, Wing SL. The Paleocene-Eocene thermal maximum: A perturbation of carbon cycle, climate, and biosphere with implications for the future. Annu Rev Earth Planet Sci. 2011;39:489–516.
80. Nauheimer L, Metzler D, Renner SS. Global history of the ancient monocot family Araceae inferred with models accounting for past continental positions and previous ranges based on fossils. New Phytol. 2012;195(4):938–50.
81. Givnish TJ, Spalink D, Ames M, Lyon SP, Hunter SJ, Zuluaga A, Doucette A, Caro GG, McDaniel J, Clements MA. Orchid historical biogeography, diversification, Antarctica and the paradox of Orchid dispersal. J Biogeogr. 2016;43(10):1905–16.
82. Tiffney BH. The Eocene North Atlantic land bridge: its importance in Tertiary and modern phytogeography of the Northern Hemisphere. J Arnold Arboretum. 1985;66(2):243–73.
83. Wen J, Nie ZL, Ickert-Bond SM. Intercontinental disjunctions between eastern Asia and western North America in vascular plants highlight the biogeographic importance of the Bering land bridge from late Cretaceous to Neogene. J Syst Evol. 2016;54(5):469–90.
84. Givnish TJ, Zuluaga A, Marques I, Lam VK, Gomez MS, Iles WJ, Ames M, Spalink D, Moeller JR, Briggs BG. Phylogenomics and historical biogeography of the monocot order Liliales: out of Australia and through Antarctica. Cladistics. 2016;32(6):581–605.
85. Kim C, Kim S-C, Kim J-H. Historical biogeography of Melanthiaceae: a case of out-of-North America through the Bering land bridge. Front Plant Sci. 2019;10:396.

Publisher's note