Insights into Phylogenetic Relationships and Genome Evolution of Subfamily Commelinoideae (Commelinaceae Mirb.) Inferred from Complete Chloroplast Genomes

Joonhyung Jung
Gachon University
Changkyun Kim
Gachon University
Joo-Hwan Kim (kimjh2009@gachon.ac.kr)
Gachon University

Research Article

Keywords: Commelinoideae, Commelinaceae, chloroplast genome, nucleotide diversity, phylogenomics

Posted Date: December 11th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-116972/v1

Version of Record: A version of this preprint was published at BMC Genomics on April 2nd, 2021. See the published version at https://doi.org/10.1186/s12864-021-07541-1.

Abstract

Background

Commelinaceae (Commelinales) comprise 41 genera and widely distributed in both the Old and New Worlds except Europe. The relationships among genera in this family have been suggested in several morphological and molecular studies. However, it is difficult to explain their relationships due to high morphological variations and low support values. Nowadays, many researchers are commonly using complete chloroplast genome data for inferring evolution of land plants. In this study, we completed 15 new chloroplast genome sequences of subfamily Commelinoideae using Mi-seq platform. We utilized genome data for the first time to reveal the structural variations and reconstruct the problematic positions of genera.

Results

All examined species of Commelinoideae have three pseudogenes (*acc*D, *rpo*A, and *ycf*15) and former two genes might be a synapomorphy within the Commelinales. Only four species in tribe Commelineae appear IR expansion which affected duplication of *rpl*22 gene. We identified inversions which range from approximately 3 to 15 kb from four taxa (*Murdannia*, *Streptolirion*, *Amischotolype*, and *Belosynapsis*). The phylogenetic analyse using 77 chloroplast protein coding genes with maximum parsimony, maximum likelihood, and the Bayesian inference suggest that *Palisota* connected with tribe Commelineae with high support values, differ from recent classification of Commelinaceae. Also, we resolved unclear position of Streptoliriinae and monophyly of Dichorisandrinae.

Conclusions

In this study, we provide detailed information of the 15 plastid genomes of Commelinaceae taxa. We identified characteristic pseudogenes and nucleotide diversity, which can be used for inferring evolutionary history about this family. Also, we need a further research to revise position of *Palisota* in recent classification.

Introduction

Commelinaceae Mirb., commonly known as dayflower and spiderwort group, are the largest family in the order Commelinales Mirb. ex Bercht. & J.Presl, which comprised four more families: Hanguanaceae, Haemodoraceae, Pontederiaceae, and Philydraceae [1, 2]. The Commelinaceae consist approximately 730 species of 41 genera and widely distributed in both the Old and New Worlds except Europe [2-4]. In this family, we are commonly using *Callisia* Loefl. and *Tradescantia* L. as an ornamental and *Commelina* L. for vegetables. The species of Commelinaceae is usually succulent and distinct with others by having closed sheathed leaves, raphide-canals and three celled glandular microhairs [3, 4]. Additionally, flowers of Commelinaceae species are mainly insect-pollinated or autogamous which have short blooming times and lack of nectar [5, 6]. The flowering unit (inflorescence) of Commelinaceae is single or compound, commonly panicle-like thyrses composed of several to many scorpioid-cymose (cincinii) branches, sometimes reduced to a single cincinnus or single flower [4, 7].

Previous classifications of Commelinaceae emphasized on floral and anatomical characters. In the first classification, Commelinaceae were divided into two tribes, Commelineae Meisner and Tradescantieae Meisner, based on number of stamens and their reproductivity [8]. Then, Bruckner [9] used flower symmetry and Pichon [10] used anatomical characters to exclude *Cartonema* from Commelinaceae. In 1966, 15 genera of Commelinaceae were defined by using various flower morphological characters [11]. In the recent classification, Commelinaceae were divided into two subfamilies, Cartonematoideae (Pichon) Faden ex G. C. Tucker and Commelinoideae Faden & D. R. Hunt, by existence of raphide-canals and glandular microhairs [4]. Cartonematoideae consists two genera (*Cartonema* R.Br. and *Triceratella* Brenan) whereas Commelinoideae includes 39 genera, which are divided into two tribes by palynological characters, Commelineae (Meisner) Faden & D. R. Hunt and Tradescantieae (Meisner). Faden & D. R. Hunt. The latter tribe was arranged into seven subtribes by morphological and cytological characters [4, 12]. However, it is difficult to interpret relationships among genera due to morphological variations. Morphological cladistic result was homoplasy and incongruent with recent classification [13]. To clarify relationships of Commelinaceae, several phylogenetic studies have been conducted [14-20]. In plastid *rbc*L phylogenetic analysis, *Cartonema* was in basal clade and both Commelineae and Tradescantieae were monophyletic except *Palisota* Rchb. which had low support values [15]. Additionally, plastid *ndh*F from previous research suggested that subtribe Tradescantiinae were paraphyletic whereas Thyrsantheminae and Dichorisandrinae were polyphyletic [16]. Combined data of nuclear 5S NTS and plastid *trn*L-F regions resulted in a well-supported relationship between Commelineae and Tradescantieae, however the positions of *Palisota* and *Spatholirion* Ridl. were ambiguous [17]. These confused relationships between genera require further research.

Chloroplast genome or plastid genome (cpDNA) is highly conserved and has a typical quadripartite structure containing a large single copy (LSC) and a small single copy (SSC) separated by two inverted repeats (IRs). The size of cpDNA ranges from 19,400 bp (*Cytinus hypocistis*) to 242,575 bp (*Pelargonium transvaalense*) and generally contains 120-130 genes, which performs important roles of photosynthesis, translation, and transcription [21, 22]. Raphid development of next-generation sequencing (NGS) enables many studies on completing plastid genomes with high quality of raw reads at low costs. Due to its conserved characteristics, chloroplast protein-coding genes were used to reconstruct the phylogenetic relationships in other monocot groups [23-25]. Also, these data are useful to infer biogeography, molecular evolution, and age estimation [26-28]. The aims of this study are to 1) explore genome evolution in Commelinoideae through analyses of sequence variation, and gene content and order; 2) find latent phylogenetically informative genes through high nucleotide diversity; 3) reconstruct the phylogenetic relationships among members of Commelinoideae with other monocot groups using 77 chloroplast protein-coding genes data, especially the relationships among seven subtribes of Tradescantieae.

Materials And Methods

Taxon sampling and DNA extraction

Fresh leaf samples were collected in the field and dried directly with silica gel in room temperature until extraction of DNA (Table 1). The samples covered four out of 14 genera in tribe Commelineae and 11 out of 25 genera which include six subtribes of tribe Tradescantieae. We prepared the voucher specimens for all used samples and deposited them in the Gachon University Herbarium (GCU) with the accession numbers. We used modified CTAB method to extract total DNA [44] and checked quality using spectrophotometer (Biospec-nano; Shimadzu) and assessed by agarose gel electrophoresis.

Genome sequencing, assembly, and annotation

Next-generation sequencing (NGS) was conducted using the Illumina MiSeq sequencing system (Illumina, Seoul, Korea). We imported NGS raw data and trimmed ends limited 5% error probability to remove poor quality of reads using Geneious prime 2020.1.2 [45]. Then, we performed 'map to reference' using *Hanguana malayana* chloroplast genome (GenBank accession = NC_029962.1) as a reference to isolate cpDNA reads. De novo assembly was implemented to reassemble reads using Geneious prime 2020.1.2 [45]. We used newly generated sequences as a reference to reassemble raw reads. We repeated this step until quadripartite structures were completed. Gaps were filled by Sanger sequencing using specific primers. Gene content and order were annotated using *Hanguana malayana* as a reference using 80% similarity to identify genes in Geneious. All tRNAs were checked by tRNAScan-SE [46] with default search mode. Illustration of plastomes were produced using OGDraw [47].

Comparative genome analysis

We compared genome structure, size, gene content across all 16 species including *Belosynapsis ciliata* (GenBank accession = MK133255.1) to cover lacking subtribe, Cyanotinae. The GC content was calculated and compared using Geneious. The whole chloroplast genome sequences of Commelinoideae species were aligned using MUSCLE embedded in Geneious and visualized using LAGAN mode in mVISTA [48, 49]. For the mVISTA plot, we used the annotated cpDNA of *Hanguana malayana* as a reference. We also examined the nucleotide diversity (Pi) of chloroplast protein coding genes, transfer RNA genes and ribosomal RNA genes among the 16 Commelinoideae species through a sliding window analysis using DnaSP v. 6.0 [50]. For the sequence divergence analysis, we applied the window size of 100 bp with a 25 bp step size. The IR and SC boundaries of the 16 Commelinoideae species were compared and illustrated using IRscope [51].

Phylogenetic analysis

A total of 42 chloroplast genome sequences (including 15 new chloroplast genomes of Commelinoideae) were used (Table S2). We extracted 77 protein coding genes and aligned using the MUSCLE embedded in Geneious prime 2020.1.2 [45]. For the data set, *Acorus calamus* (Acoraceae) was designated as an outgroup. We performed maximum parsimony (MP), maximum likelihood (ML), and Bayesian inference (BI) to infer relationships of Commelinoideae and related taxa. The MP analyses were carried out in PAUP* v4.0a [52] with all characters equally weighted and unordered. Gaps were treated by missing data. Searches of 1000 random taxon addition replicates used tree-bisection-reconnection (TBR) branch swapping and MulTrees permitted 10 trees to be held at each step. Bootstrap analyses (PBP, parsimony bootstrap percentages, 1000 pseudoreplicates) were conducted to examine internal support with the same parameters. We used jModelTest version 2.1.7 [53, 54] to find the best model with Akaike's information criterion (AIC) before running the ML and BI analyses. The GTR + I + G was the best model for the concatenated data sets. We used the IQ-TREE web server (http://iqtree.cibiv.univie.ac.at/) to make the ML searches [55]. Support value (MBP, mean bootstrap percentage) was calculated with 1000 replicates of ultrafast bootstrap [56]. MrBayes v3.2.7 [57] was used for BI analyses. Two simultaneous runs were performed starting from random trees for at least 1,000,000 generations. One tree was sampled every 1,000 generations. In total, 25% of 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 given as posterior probability (PP) to estimate robustness of the half of BI tree. Then, the effective sample size values (ESS) were checked for model parameters (at least 200). The phylogenetic trees were edited using FigTree v1.4.4 program [58].

Results

Chloroplast genome assembly and annotation

We completed 15 new plastid genomes in this study listed in Table 1 through 9 to 21 million raw reads for each species (Fig. S1, Table S1). A total of 16 plastid genomes, including *Belosynapsis ciliata*, exhibit the typical quadripartite structure containing LSC and SSC regions separated by two inverted repeats (Fig. 1). Plastid genome sequences of *Murdannia edulis* and *Belosynapsis ciliata* are over 170 kb in length whereas that of *Commelina communis* is 160,116 bp in length (Table 1). In addition, *Murdannia edulis* and *Belosynapsis ciliata* have the lowest GC content (34.5 %) whereas *Palisota barteri* has the highest GC content (36.2 %) (Table 1). The highest length difference was observed in LSC region about 8,801 bp between *Belosynapsis ciliata* and *Commelina communis*, GC content was in SSC region about 3.4 % between *Dichorisandra thyrsiflora* and *Murdannia edulis* (Table 1). Plastid genomes of Commelinoideae have 131 genes, of which 111 are unique and 20 are duplicated in the IR regions (Table 2), except *rpl*22 gene which was not duplicated in tribe Tradescantieae. There are 77 protein-coding genes (CDS), 30 transfer RNA (tRNA) genes and 4 ribosomal RNA (rRNA) genes in examined Commelinoideae taxa (Table 2). In these genes, three CDS (*rps*12, *clp*P, and *ycf*3) have two introns, nine CDS (*atp*F, *ndh*A, *ndh*B, *pet*B, *pet*D, *rpl*2, *rpl*16, *rpo*C1, and *rps*16) and six tRNA (*trn*K-UUU, *trn*G-UCC, *trn*L-UAA, *trn*V-UAC, *trn*I-GAU, and *trn*A-UGC) have one intron (Table 2). The *rps*12 gene was trans-spliced, which has 5' exon in LSC and 3' exon and intron in the IR regions. Three pseudogenes (*acc*D, *rpo*A, and *ycf*15) were identified from all Commelinoideae species, one (*ycf*15) of which was duplicated in the IR regions (Table 2). These three genes contained several internal stop codons due to insertions and deletions, thus are identified as pseudogenes. Also, we identified *ndh*B as pseudogene in two species (*Pollia japonica* and *Rhopalephora scaberrima*) in consequence of point mutation.

Comparative chloroplast genome structure and nucleotide diversity

The aligned data of whole plastid genomes showed high similarities in coding genes, and high variations in non-coding genes (Fig. 3). We found several genome structure variations among Commelinoideae species. *Murdannia edulis* and *Streptolirion volubile* had one inversion from *rbc*L to *psa*I intergenetic spacer (approximately 3 kb) and *pet*N to *trn*E-UUC (approximately 2.8 kb), respectively. *Amischotolype hispida* and *Belosynapsis ciliata* had two large inversions from *trn*V-UAC to *rbc*L and *psb*J to *pet*D about approximately 5 kb and 16 kb, respectively. The IR-SSC boundary was similar among species of Commelinoideae (Fig. 4). All plastid genomes have incomplete duplicated *ycf*1 gene in the IR_B-SSC junctions. We also found an expansion of IR regions in Commelineae species which resulted duplication of *rpl*22 genes (Fig. 4).

We analysed nucleotide divergences of CDS, tRNA, and rRNA to explain variant characteristics among the 16 Commelinoideae plastid genomes (Fig. 2, Table S3). Nucleotide diversity (Pi) for each CDS ranges from 0.00427 (*psb*L) to 0.09543 (*ycf*1) with an average of 0.03473. Nine CDS (*rps*3, *ndh*G, *ndh*D, *ccs*A, *rps*15, *rpl*32, *ndh*F, *mat*K, and *ycf*1) have remarkably high values (Pi > 0.05) and seven CDS (*psb*L, *rpl*23, *rps*19, *ndh*B, *rpl*2, *rps*7, *rps*12) have low values (Pi < 0.01; Fig. 2). Compared with Tradescantieae, Commelineae have relatively higher values in almost CDS (Fig. 2). In Tradescsantieae, however, the *rpl*22 gene has higher value (Pi = 0.04655) in comparison with Commelineae. In tRNA and rRNA, Pi values range from 0 (*trn*T-UGU, *trn*H-GUG, *trn*V-GAC, *trn*I-GAU) to 0.02697 (*trn*Q-UUG) with an average of 0.006. Commelineae have the highest value in the *trn*L-UAA (Pi = 0.02941) while Tradescantieae have no value in this gene. We tried to find latent phylogenetically informative genes for the Commelinoideae by checking individual CDS with high values (Pi > 0.045) and over 500 bp length. Ten CDS (*ndh*H, *rpo*C2, *ndh*A, *rps*3, *ndh*G, *ndh*D, *ccs*A, *ndh*F, *mat*K, and *ycf*1) were checked respectively with ML analysis and compared positions among 16 genera of Commelinoideae with Fig. 5. Total four CDS (*ndh*H, *rpo*C2, *mat*K, *ycf*1) have similar topology in Commelinoideae even though the other monocot groups were unclear.

Phylogenetic analysis

The aligned 77 chloroplast protein-coding genes had 65,481 characters, of which 16,380 were parsimony informative. The MP analysis produced single most-parsimonious tree (tree length = 72,586, CI = 0.488, RI = 0.626). The tree topologies from among MP, ML, and BI were found to be congruent with each other with 100% bootstrap (PBP, MBP) values and 1.00 Bayesian posterior probabilities (PP) supporting in almost all nodes except *Palisota* which was unresolved in MP analysis (not shown) (Fig. 5). The result suggested that *Palisota* was sister to the group consisting of the rest of Commelinoideae (Fig. 5). In Tradescantieae, Streptoliriinae was positioned at the basal node. Then, Dichorisandrinae divided into two clades ((*Dichorisandra*, *Siderasis*), (*Cochliostema*, *Geogenanthus*)) with relatively low support values in both MP and ML analysis (PBP = 74, MBP = 84, PP = 1) (Fig. 5). Among remain four subtribes, where two clades ((Coleotrypinae and Cyanotinae), (Tradescantiinae and Thyrsantheminae)) were formed with high support values (PBP = 100, MBP = 100, PP = 1), respecively (Fig. 5).

Discussion

Chloroplast genome structure

In this study, we completed 15 new plastid genomes of Commelinoideae taxa (Table 1). Plastid genomes have typical quadripartite structures, including LSC, SSC and two IR regions. Plastid genomes of Commelinoideae have variable total length and GC content. The LSC and SSC regions have relatively higher length and AT-content difference rather than IR region (Table 1). The functions of AT-rich sequences in the plastid genome were known as enhancing succeed of gene transfer by making stable transcripts [29]. However, AT-rich sequences caused structural variations like inversions by their weakness hydrogen bonding. In this study, we identified small to large inversions from four species (Fig. 3). There is one inversion in *Murdannia edulis* and *Streptolirion volubile*, whereas two inversions in *Amischotolype hispida* and *Belosynapsis ciliata* (Fig. 3). Inversions are known as common event of genome rearrangement and provide informative infrageneric relationships. In the previous research, inversions occurred by microhomology-driven recombination via short repeats and suggested monophyly of tribe Desmodieae in the Fabaceae [30]. Our result also suggests that both *Amischotolype* and *Belosynapsis* have two large inversions in same loci and formed a clade together which is sister to Dichorisandrinae (Fig. 5).

We identified an IR expansion in members of Commelineae (*Murdannia*, *Commelina*, *Pollia*, and *Rhopalephora*). Four species have one more *rpl*22 gene, which is duplicated in the terminal of IR regions (Fig. 4). Although IR expansion affected gene composition, the total length of IR region is similar among 16 Commelinoideae species. IR expansion and contraction are important events in several families. In Ranunculaceae, IR expansion was detected as a synapomorphy of the variation in tribe Anemoneae [31]. Likewise, IR expansion suggested more support for the relationship between the two subfamilies, Ehrhartoideae and Pooideae, in the Poaceae [32]. This event also may be phylogenetically informative in Commelinoideae due to only Commelineae species share this genome variation after diverged from *Palisota* in this study (Fig. 5).

Within Commelinoideae plastid genomes, three protein coding genes (*acc*D, *rpo*A, and *ycf*15) were found as pseudogenes (Fig. S2). The *ycf*15 gene has several abnormal stop codons caused by insertions and deletions (indel) of bases, which are similar with other monocots. We also identified that all examined species have indels at the front part of *acc*D gene (until 400 bp) and terminal part of *rpo*A gene (after 700 bp; Fig. S2). The *acc*D gene, encoding the beta-carboxyl transferase subunit of acetyl-CoA carboxylase, is in most flowering plants and synthesize fatty acid within the chloroplast. It was suggested as an essential gene that related with maintaining chloroplast structure [33]. However, it was reported as a gene loss or pseudogenization in Acoraceae, and Poaceae [34, 35]. Recent studies suggested that *acc*D gene was found in nuclear originated from chloroplasts in several eudicots [36, 37]. The *rpo*A gene, encoding the alpha subunit of RNA polymerase, is also in most flowering plants but recorded gene loss in the chloroplast genome of mosses [38]. One of species, *Physcomitrella patens* (Funariaceae), *rpo*A gene has transferred to the nuclear [39]. We need a further study whether these two genes transferred to the nuclear or not in the Commelinaceae. We identified that these pseudogened *acc*D and *rpo*A only appeared in the Commelinoideae among the Commelinales. It might be a specific character of gene composition in the Commelinales. We also found point mutated base in the third codon of *ndh*B gene in both *Pollia japonica* and *Rhopalephora scaberrima*, which formed a clade together in this study (Fig. 5).

We measured the nucleotide diversity of CDS, tRNA, and rRNA to identify the genetic divergence between 16 Commelinoideae plastid genomes. We found that the CDS in the IR regions have lower nucleotide diversity than that of the LSC and SSC regions (Fig. 2). This result has also been identified in the other monocots [40-42]. It may possibly be attributed to copy correction of the IR regions via gene conversion [43]. Especially, we can see this result in the *rpl*22 gene. Only Commelineae species have duplicated one due to IR expansion mentioned above while remain 12 taxa have one gene in the LSC or LSC-IR junction (Fig. 4). Difference of nucleotide diversity in this gene between Commelineae (Pi = 0.015) and Tradescantieae (Pi = 0.0466) is 0.0316. It might be phylogenetically useful information for Tradescantieae only.

Implication of phylogenomic study using plastomes data

In the first phylogenetic analysis of Commelinaceae based on plastid *rbc*L, they revealed a relationship of 32 species representing 30 genera of Commelinaceae [15]. Cartonematoideae was in a basal clade connected with Commelinoideae as a sister consisting of all remain species [15]. Except *Palisota*, Commelinoideae was divided into two tribes, Commelineae and Tradescantieae, with the low bootstrap support value due to insufficient information [15]. Although several phylogenetic studies were conducted, Commelinaceae still have unresolved relationships between genera. First, the position of *Palisota* had been problematic that 1) sister to all genera of Commelinoideae with high bootstrap values [15]; 2) support low bootstrap value with other Tradescantieae species [16], and; 3) belong to Tradescantieae as a basal group [19]. Second, Streptoliriinae was placed with Commelineae species in *trn*L-*trn*F analysis [17]. Third, subtribe Dichorisandrinae seemed polyphyletic in the previous researches [15, 16, 19]. These results are most likely due to limited taxon sampling and/or used few informative genetic markers. The aligned 77 chloroplast coding genes data in this study suggests improved relationship of each genera (Fig. 5). We identified that Commelinoideae divided into two clades, tribe Commelineae and Tradescantieae, with high support values (Fig. 5). However, *Palisota*, which belongs to Tradescantieae in recent classification [3], is connected with Commelineae species as basal group in this study (Fig. 5). Both ML and BI results are supported with high values even though unresolved in MP (data not shown). Compared with recent classification, it seems like that subsidiary cells in stomata and pollen exine are not key characters at least for *Palisota* [3]. In the Commelinaceae, fruits are commonly loculicidally dehiscent capsules while *Palisota*, *Pollia*, *Tapheocarpa*, and some *Aneilema* species have indehiscent type [3]. The latter three genera are groups of Commelineae in recent classification. Also, indehiscent fruit was distinctive character in previous research to place *Pollia* and *Palisota* as a same group [11]. Other four Commelineae species (*Murdannia*, (*Commelina*, (*Pollia*, *Rhopalephora*))) are connected with high support values, which have similar relationships with previous research [15]. Within Tradescantieae, Streptoliriinae was diverged in the first and Dichorisandrinae was divided into two clades with relatively low support values (PBP = 77/MBP = 84/PP = 1) (Fig. 5). After that, Coleotrypinae and Cyanotinae were diverged, which formed a sister with remain Thyrsantheminae and Tradescantiinae. Interestingly, the Asian and African subtribe Coleotrypinae and Cyanotinae were nested well within the New World subtribes (Fig. 5). This result is similar with previous research [15] and shows questions of biogeographic history.

Conclusion

Our study revealed genome structural characteristics, nucleotide diversity, improved relationships between genera using 15 newly complete chloroplast genomes of Commelinoideae. Compared with other Commelinales species, we found two characteristic pseudogenes in all members of Commelinoideae and this might be a synapomorphy within the Commelinales. Four genes (*ndh*H, *rpo*C2, *mat*K, *ycf*1) seem to provide phylogenetically useful information for Commelinoideae due to similar topology with Fig. 5. We also reconstruct the phylogenetic relationships using 77 chloroplast protein coding genes. Although we cannot explain of whole Commelinaceae due to loss of subfamily Cartonematoideae, we identified relationships of Commelinoideae taxa especially seven subtribes of Tradescantieae. One interesting result was that *Palisota* (Palisotinae) relates to Commelineae clade with high support values. This result is incongruent with the latest classification and we need a further research about that [3]. Also, we resolved the position of Streptoliriinae and monophyly of Dichorisandrinae. Future studies might use the information of chloroplast genomes that we provided in this study and make sure the evolutionary history of the Commelinaceae.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The 15 chloroplast genomes sequences we obtained from this study were archived in NCBI. The accession numbers are presented in Table 1.

Competing interests

The authors declare that they have no competing interests.

Funding

This work 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.

Authors' contributions

JJ performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft. CK authored or reviewed drafts of the paper, approved the final draft. JHK conceived and designed the experiments, contributed reagents / materials / analysis tools, authored or reviewed drafts of the paper, approved the final draft.

Acknowledgements

We would like to thank Gerardo A. Salazar at Universidad Nacional Autónoma de México (Mexico), Claudia T. Hornung-Leoni at Autonomous University of Hidalgo (Mexico), Manuel González Ledesma at Autonomous University of Hidalgo (Mexico), Kenneth M. Cameron at University of Wisconsin–Madison (United States of America), Chien-Ti Chao at National Taiwan Normal University (Taiwan), David Warmington at Cairns Botanic Gardens (Australia), Carlos Gustavo Espejo Zurita at Jardín Botánico Histórico La Concepción (Spain) for collecting and providing the plant material for this study.

References

1. Chase MW, Christenhusz M, Fay M, Byng J, Judd WS, Soltis D, Mabberley D, Sennikov A, Soltis PS, Stevens PF: **An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV**. *Botanical Journal of the Linnean Society* 2016, **181**(1):1-20.
2. Christenhusz MJM, Byng JW: **The number of known plants species in the world and its annual increase**. *Phytotaxa* 2016, **261**(3):201.
3. Faden RB, Hunt D: **The classification of the Commelinaceae**. *Taxon* 1991:19-31.
4. Faden RB: **Commelinaceae**. In: *Flowering Plants· Monocotyledons.* Springer; 1998: 109-128.
5. Owens S: **Self-incompatibility in the Commelinaceae**. *Annals of Botany* 1981, **47**(5):567-581.
6. Faden RB: **Floral attraction and floral hairs in the Commelinaceae**. *Annals of the Missouri Botanical Garden* 1992:46-52.
7. Panigo E, Ramos J, Lucero L, Perreta M, Vegetti A: **The inflorescence in Commelinaceae**. *Flora - Morphology, Distribution, Functional Ecology of Plants* 2011, **206**(4):294-299.
8. Meisner CJPvg: **CCLXI Commelinaceae**. 1842, **1**:406-407.
9. Bruckner G: **Beiträge zur anatomie morphologie und systematik der Commelinaceae**. 1926.
10. Pichon MJNS: **Sur les Commelinaces**. 1946, **12**:217-242.
11. Brenan JP: **The classification of Commelinaceae**. *Botanical Journal of the Linnean Society* 1966, **59**(380):349-370.
12. Hardy CR, Faden RB: **Plowmanianthus, a new genus of Commelinaceae with five new species from tropical America**. *Systematic Botany* 2004, **29**(2):316-333.
13. Evans TM, Faden RB, Simpson MG, Sytsma KJ: **Phylogenetic relationships in the Commelinaceae: IA cladistic analysis of morphological data**. *Systematic Botany* 2000:668-691.
14. Bergamo S: **A phylogenetic evaluation of Callisia Loefl.(Commelinaceae) based on molecular data**. uga; 2003.
15. Evans TM, Sytsma KJ, Faden RB, Givnish TJ: **Phylogenetic relationships in the Commelinaceae: II. A cladistic analysis of rbcL sequences and morphology**. *Systematic Botany* 2003:270-292.
16. Wade DJ, Evans TM, Faden RB: **Subtribal relationships in tribe Tradescantieae (Commelinaceae) based on molecular and morphological data**. *Aliso: A Journal of Systematic and Evolutionary Botany* 2006, **22**(1):520-526.
17. Burns JH, Faden RB, Steppan SJ: **Phylogenetic Studies in the Commelinaceae Subfamily Commelinoideae Inferred from Nuclear Ribosomal and Chloroplast DNA Sequences**. *Systematic Botany* 2011, **36**(2):268-276.
18. Zuiderveen GH, Evans TM, Faden RB: **A phylogenetic analysis of the African plant genus Palisota (family Commelinaceae) based on chloroplast DNA sequences**. 2011.
19. Hertweck KL, Pires JC: **Systematics and Evolution of Inflorescence Structure in the Tradescantia Alliance (Commelinaceae)**. *Systematic Botany* 2014, **39**(1):105-116.
20. Kelly SM, Evans TM: **A Phylogenetic Analysis of the African Plant Genus Aneilema (family Commelinaceae) based on Chloroplast DNA Sequences**. 2014.
21. DYER TAJTip: **The chloroplast genome: its nature and role in development**. 1984, **5**:23-69.
22. Sugiura MJPmb: **The chloroplast genome**. 1992, **19**(1):149-168.
23. Kim JH, Kim DK, Forest F, Fay MF, Chase MW: **Molecular phylogenetics of Ruscaceae sensu lato and related families (Asparagales) based on plastid and nuclear DNA sequences**. *Ann Bot* 2010, **106**(5):775-790.
24. Barrett CF, Baker WJ, Comer JR, Conran JG, Lahmeyer SC, Leebens-Mack JH, Li J, Lim GS, Mayfield-Jones DR, Perez LJNP: **Plastid genomes reveal support for deep phylogenetic relationships and extensive rate variation among palms and other commelinid monocots**. 2016, **209**(2):855-870.
25. Do HDK, Kim C, Chase MW, Kim JH: **Implications of plastome evolution in the true lilies (monocot order Liliales)**. *Mol Phylogenet Evol* 2020, **148**:106818.
26. Jones SS, Burke SV, Duvall MRJPs, evolution: **Phylogenomics, molecular evolution, and estimated ages of lineages from the deep phylogeny of Poaceae**. 2014, **300**(6):1421-1436.
27. Li Q-Q, Zhou S-D, Huang D-Q, He X-J, Wei X-QJAP: **Molecular phylogeny, divergence time estimates and historical biogeography within one of the world's largest monocot genera**. 2016, **8**.
28. Kim C, Kim S-C, Kim J-HJFips: **Historical biogeography of Melanthiaceae: a case of out-of-North America through the Bering land bridge**. 2019, **10**:396.
29. Stegemann S, Bock RJTpc: **Experimental reconstruction of functional gene transfer from the tobacco plastid genome to the nucleus**. 2006, **18**(11):2869-2878.

30. Jin D-P, Choi I-S, Choi B-HJPo: **Plastid genome evolution in tribe Desmodieae (Fabaceae: Papilionoideae)**. 2019, **14**(6):e0218743.
31. He J, Yao M, Lyu R-D, Lin L-L, Liu H-J, Pei L-Y, Yan S-X, Xie L, Cheng JJSr: **Structural variation of the complete chloroplast genome and plastid phylogenomics of the genus Asteropyrum (Ranunculaceae)**. 2019, **9**(1):1-13.
32. Guisinger MM, Chumley TW, Kuehl JV, Boore JL, Jansen RKJJoME: **Implications of the plastid genome sequence of Typha (Typhaceae, Poales) for understanding genome evolution in Poaceae**. 2010, **70**(2):149-166.
33. Kode V, Mudd EA, Iamtham S, Day AJTpj: **The tobacco plastid accD gene is essential and is required for leaf development**. 2005, **44**(2):237-244.
34. Goremykin VV, Holland B, Hirsch-Ernst KI, Hellwig FHJMb, evolution: **Analysis of Acorus calamus chloroplast genome and its phylogenetic implications**. 2005, **22**(9):1813-1822.
35. Harris ME, Meyer G, Vandergon T, Vandergon VOJPMBR: **Loss of the acetyl-CoA carboxylase (accD) gene in Poales**. 2013, **31**(1):21-31.
36. Rousseau-Gueutin M, Huang X, Higginson E, Ayliffe M, Day A, Timmis JNJPp: **Potential functional replacement of the plastidic acetyl-CoA carboxylase subunit (accD) gene by recent transfers to the nucleus in some angiosperm lineages**. 2013, **161**(4):1918-1929.
37. Li J, Gao L, Chen S, Tao K, Su Y, Wang TJSr: **Evolution of short inverted repeat in cupressophytes, transfer of accD to nucleus in Sciadopitys verticillata and phylogenetic position of Sciadopityaceae**. 2016, **6**(1):1-12.
38. Goffinet B, Wickett NJ, Shaw AJ, Cox CJJT: **Phylogenetic significance of the rpoA loss in the chloroplast genome of mosses**. 2005, **54**(2):353-360.
39. Sugiura C, Kobayashi Y, Aoki S, Sugita C, Sugita MJNAR: **Complete chloroplast DNA sequence of the moss Physcomitrella patens: evidence for the loss and relocation of rpoA from the chloroplast to the nucleus**. 2003, **31**(18):5324-5331.
40. Lee SR, Kim K, Lee BY, Lim CE: **Complete chloroplast genomes of all six Hosta species occurring in Korea: molecular structures, comparative, and phylogenetic analyses**. *BMC Genomics* 2019, **20**(1):833.
41. Huang J, Yu Y, Liu YM, Xie DF, He XJ, Zhou SD: **Comparative Chloroplast Genomics of Fritillaria (Liliaceae), Inferences for Phylogenetic Relationships between Fritillaria and Lilium and Plastome Evolution**. *Plants (Basel)* 2020, **9**(2).
42. Smidt EC, Paez MZ, Vieira LDN, Viruel J, de Baura VA, Balsanelli E, de Souza EM, Chase MW: **Characterization of sequence variability hotspots in Cranichideae plastomes (Orchidaceae, Orchidoideae)**. *PLoS One* 2020, **15**(1):e0227991.
43. Khakhlova O, Bock R: **Elimination of deleterious mutations in plastid genomes by gene conversion**. *The Plant Journal* 2006, **46**(1):85-94.
44. Doyle J, Doyle JJPB: **CTAB DNA extraction in plants**. 1987, **19**:11-15.
45. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran CJB: **Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data**. 2012, **28**(12):1647-1649.
46. Chan PP, Lowe TM: **tRNAscan-SE: searching for tRNA genes in genomic sequences**. In: *Gene Prediction*. Springer; 2019: 1-14.
47. Greiner S, Lehwark P, Bock RJNAR: **OrganellarGenomeDRAW (OGDRAW) version 1.3. 1: expanded toolkit for the graphical visualization of organellar genomes**. 2019, **47**(W1):W59-W64.
48. Brudno M, Do CB, Cooper GM, Kim MF, Davydov E, Green ED, Sidow A, Batzoglou S, research NCSPJG: **LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA**. 2003, **13**(4):721-731.
49. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak IJNar: **VISTA: computational tools for comparative genomics**. 2004, **32**(suppl_2):W273-W279.
50. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia AJMb, evolution: **DnaSP 6: DNA sequence polymorphism analysis of large data sets**. 2017, **34**(12):3299-3302.
51. Amiryousefi A, Hyvönen J, Poczai PJB: **IRscope: an online program to visualize the junction sites of chloroplast genomes**. 2018, **34**(17):3030-3031.
52. Swofford DJPaupS, Sunderland: **PAUP* 4.0 b. 4a**. 2000.
53. Guindon S, Gascuel O: **A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. Systematic Biology52: 696-704**. In.; 2003.
54. Darriba D, Taboada G, Doallo R, Posada DJJNM: **2: More models, new heuristics and high-performance computing**. 2012, **9**:772.
55. Trifinopoulos J, Nguyen L-T, von Haeseler A, Minh BQJNar: **W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis**. 2016, **44**(W1):W232-W235.
56. Stamatakis A, Hoover P, Rougemont JJSb: **A rapid bootstrap algorithm for the RAxML web servers**. 2008, **57**(5):758-771.
57. Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JPJSb: **MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space**. 2012, **61**(3):539-542.
58. Rambaut A, FigTree V: **1.4. 4**. In.; 2018.

Tables

Table 1 Comparison of the features of plastomes from 16 genera of Commelinaceae.

Taxa	Tribe	Subtribe	Length and G+C content				GenBank accession number	Voucher
			LSC bp (G+C%)	SSC bp (G+C%)	IR bp (G+C%)	Total bp (G+C%)		
Gibasis geniculata	Tradescantieae	Tradescantiinae	89,154(33.3)	18,278(30.5)	26,953(42.5)	161,338(36.1)	This study	JH200402001
Tradescantia virginiana	Tradescantieae	Tradescantiinae	91,991(32.7)	18,462(30.2)	27,236(42.3)	164,925(35.6)	This study	JH170813001
Callisia repens	Tradescantieae	Tradescantiinae	89,446(33.2)	18,252(30.3)	27,078(42.5)	161,854(36.0)	This study	Jardín Botánico Histórico La Concepción
Weldenia candida	Tradescantieae	Thyrsantheminae	95,029(32.6)	19,024(30.3)	27,233(42.6)	168,519(35.5)	This study	JH190730001
Amischotolype hispida	Tradescantieae	Coleotrypinae	94,525(32.9)	19,255(30.4)	27,385(42.4)	168,550(35.7)	This study	JH191109002
Belosynapsis ciliata	Tradescantieae	Cyanotinae	96,164(31.3)	20,224(28.0)	27,241(42.6)	170,870(34.5)	MK133255.1	
Cochliostema odoratissimum	Tradescantieae	Dichorisandrinae	92,560(33.2)	18,856(30.4)	27,276(42.5)	165,968(35.9)	This study	Cairns Botanic Gardens
Geogenanthus poeppigii	Tradescantieae	Dichorisandrinae	94,583(32.8)	18,612(30.7)	27,098(42.5)	167,391(35.7)	This study	JH190803001
Dichorisandra thyrsiflora	Tradescantieae	Dichorisandrinae	94,347(32.9)	18,348(31.1)	27,194(42.6)	167,083(35.8)	This study	JH190616001
Siderasis fuscata	Tradescantieae	Dichorisandrinae	94,389(32.9)	18,606(31.0)	27,196(42.6)	167,387(35.8)	This study	XX-0-GENT-19822394
Streptolirion volubile	Tradescantieae	Streptoliriinae	91,528(33.1)	19,595(29.3)	27,447(42.0)	166,017(35.6)	This study	JH180919003
Palisota barteri	Tradescantieae	Palisotinae	93,315(33.5)	18,905(30.8)	27,074(42.7)	166,368(36.2)	This study	JH190222001
Pollia japonica	Commelineae		90,295(33.2)	19,151(29.7)	27,604(42.2)	164,654(35.8)	This study	JH180805001
Rhopalephora scaberrima	Commelineae		87,602(33.2)	18,354(29.5)	27,487(42.1)	160,930(35.8)	This study	JH191109014
Commelina communis	Commelineae		87,363(33.0)	18,561(29.1)	27,096(42.3)	160,116(35.7)	This study	JH180709001
Murdannia edulis	Commelineae		96,248(31.4)	20,798(27.7)	27,464(42.1)	171,974(34.4)	This study	JH191110010

Table 2 Gene composition within chloroplast genomes of Commelinaceae species.

Groups of genes		Names of genes	No.
RNA genes	Ribosomal RNAs	*rrn*4.5 [X2], *rrn*5 [X2], *rrn*16 [X2], *rrn*23 [X2]	8
	Transfer RNAs	*trnK*-UUU [a], *trnQ*-UUG, *trnS*-GCU, *trnG*-UCC [a], *trnR*-UCU, *trnC*-GCA, *trnD*-GUC, *trnY*-GUA, *trnE*-UUC, *trnT*-GGU, *trnS*-UGA, *trnG*-GCC, *trnfM*-CAU, *trnS*-GGA, *trnT*-UGU, *trnL*-UAA [a], *trnF*-GAA, *trnV*-UAC[a], *trnM*-CAU, *trnW*-CCA, *trnP*-UGG, *trnH*-GUG [X2], *trnI*-CAU [X2], *trnL*-CAA [X2], *trnV*-GAC [X2], *trnI*-GAU [a X2], *trnA*-UGC [a X2], *trnR*-ACG [X2], *trnN*-GUU [X2], *trnL*-UAG	38
Protein genes	Photosystem I	*psa*A, *psa*B, *psa*C, *psa*I, *psa*J	5
	Photosystem II	*psb*A, *psb*B, *psb*C, *psb*D, *psb*E, *psb*F, *psb*H, *psb*I, *psb*J, *psb*K, *psb*L, *psb*M, *psb*N, *psb*T, *psb*Z	15
	Cytochrome	*pet*A, *pet*B [a], *pet*D [a], *pet*G, *pet*L, *pet*N	6
	ATP synthases	*atp*A, *atp*B, *atp*E, *atp*F [a], *atp*H, *atp*I	6
	Large unit of Rubisco	*rbc*L	1
	NADH dehydrogenase	*ndh*A [a], *ndh*B [a X2], *ndh*C, *ndh*D, *ndh*E, *ndh*F, *ndh*G, *ndh*H, *ndh*I, *ndh*J, *ndh*K	12
	ATP-dependent protease subunit P	*clp*P [b]	1
	Envelope membrane protein	*cem*A	1
Ribosomal proteins	Large units of ribosome	*rpl*2 [a X2], *rpl*14, *rpl*16 [a], *rpl*20, *rpl*22 [X2], *rpl*23 [X2], *rpl*32, *rpl*33, *rpl*36	12
	Small units of ribosome	*rps*2, *rps*3, *rps*4, *rps*7 [X2], *rps*8, *rps*11, *rps*12 [X2], *rps*14, *rps*15, *rps*16 [a], *rps*18, *rps*19 [X2]	15
Transcription/ translation	RNA polymerase	*rpo*A[Ψ], *rpo*B, *rpo*C1 [a], *rpo*C2	3
	Initiation factor	*inf*A	1
	Miscellaneous protein	*acc*D[Ψ], *ccs*A, *mat*K	2
	Hypothetical proteins and conserved reading frames	*ycf*1, *ycf*2 [X2], *ycf*3 [b], *ycf*4, *ycf*15[Ψ]	5
Total			131

a: gene with one intron; b: gene with two introns; X2: duplicated gene; Ψ: pseudogene

Figures

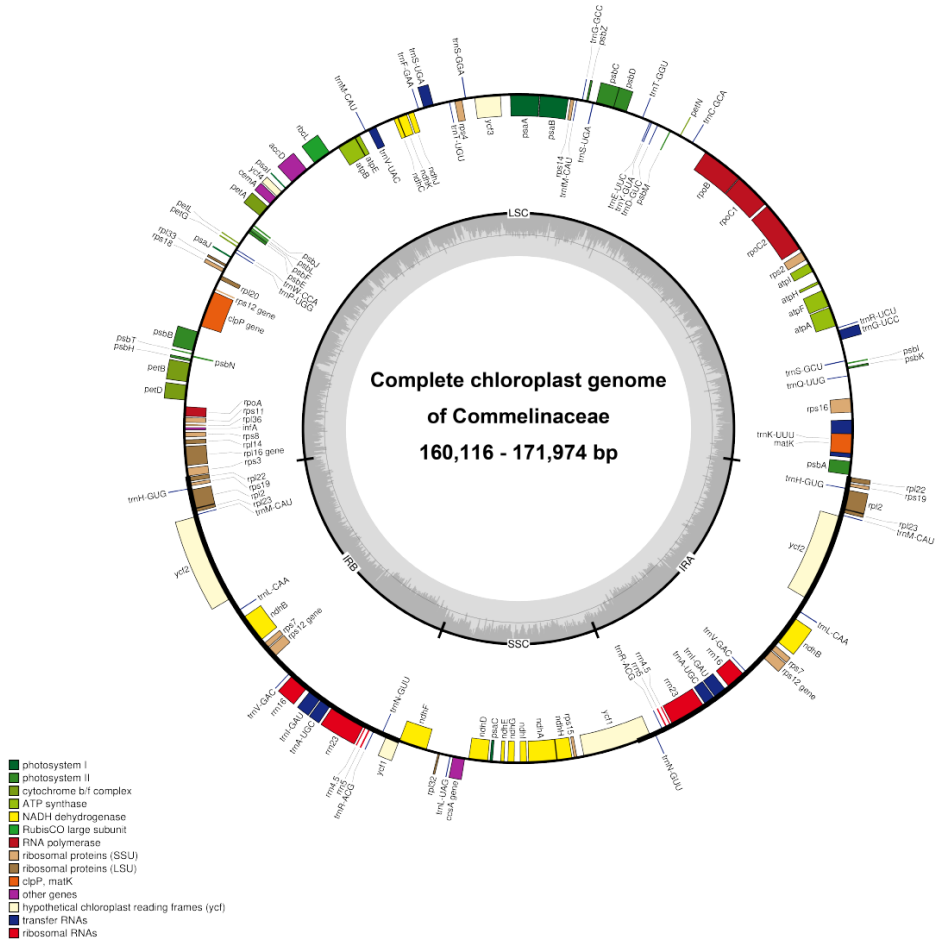

Figure 1

Representative chloroplast genome of Commelinaceae. The colored boxes represent conserved chloroplast genes. Genes shown inside the circle are transcribed clockwise, whereas genes outside the circle are transcribed counter-clockwise. The small grey bar graphs inner circle shows the GC contents.

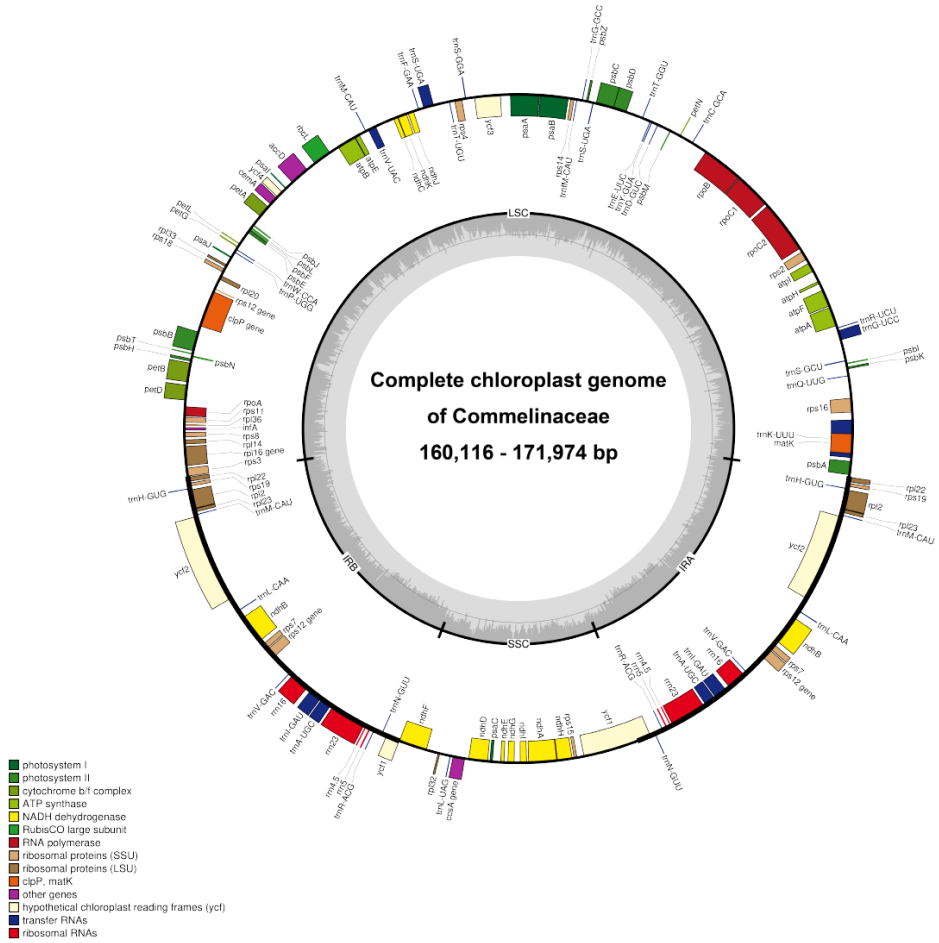

Figure 1

Representative chloroplast genome of Commelinaceae. The colored boxes represent conserved chloroplast genes. Genes shown inside the circle are transcribed clockwise, whereas genes outside the circle are transcribed counter-clockwise. The small grey bar graphs inner circle shows the GC contents.

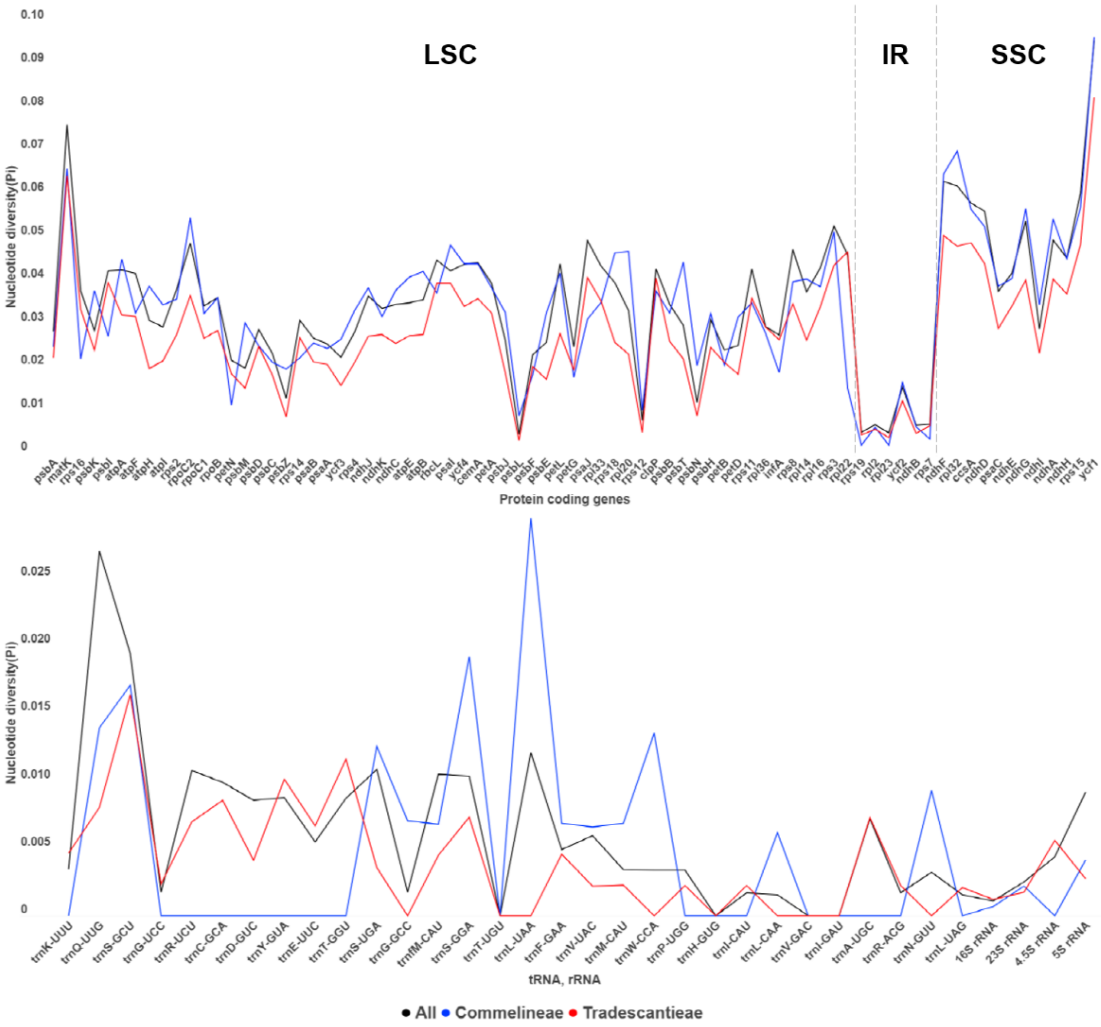

Figure 2

Nucleotide diversity (Pi) values in protein coding genes, tRNA, and rRNA in 16 Commelinaceae species. The dashed lines are the borders of LSC, IR and SSC regions.

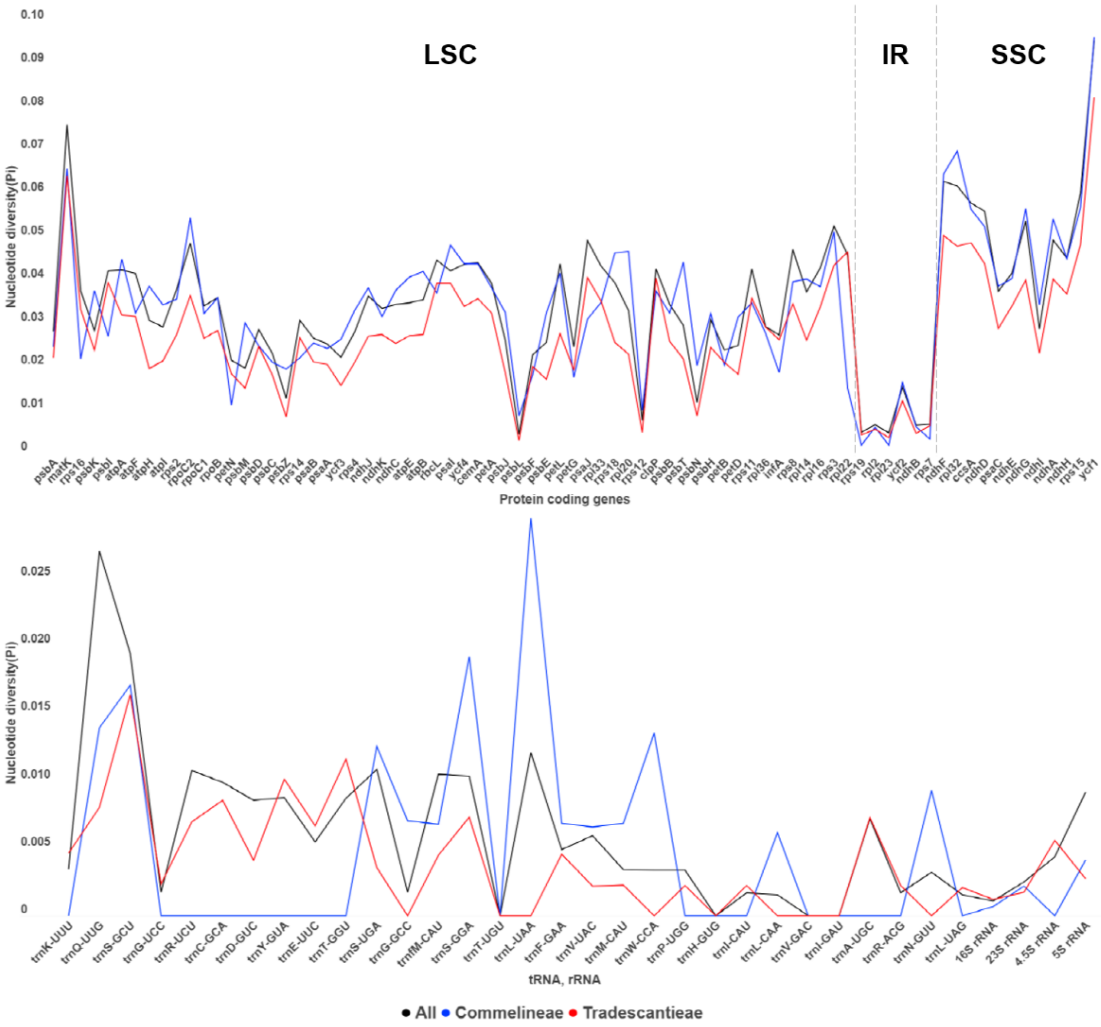

Figure 2

Nucleotide diversity (Pi) values in protein coding genes, tRNA, and rRNA in 16 Commelinaceae species. The dashed lines are the borders of LSC, IR and SSC regions.

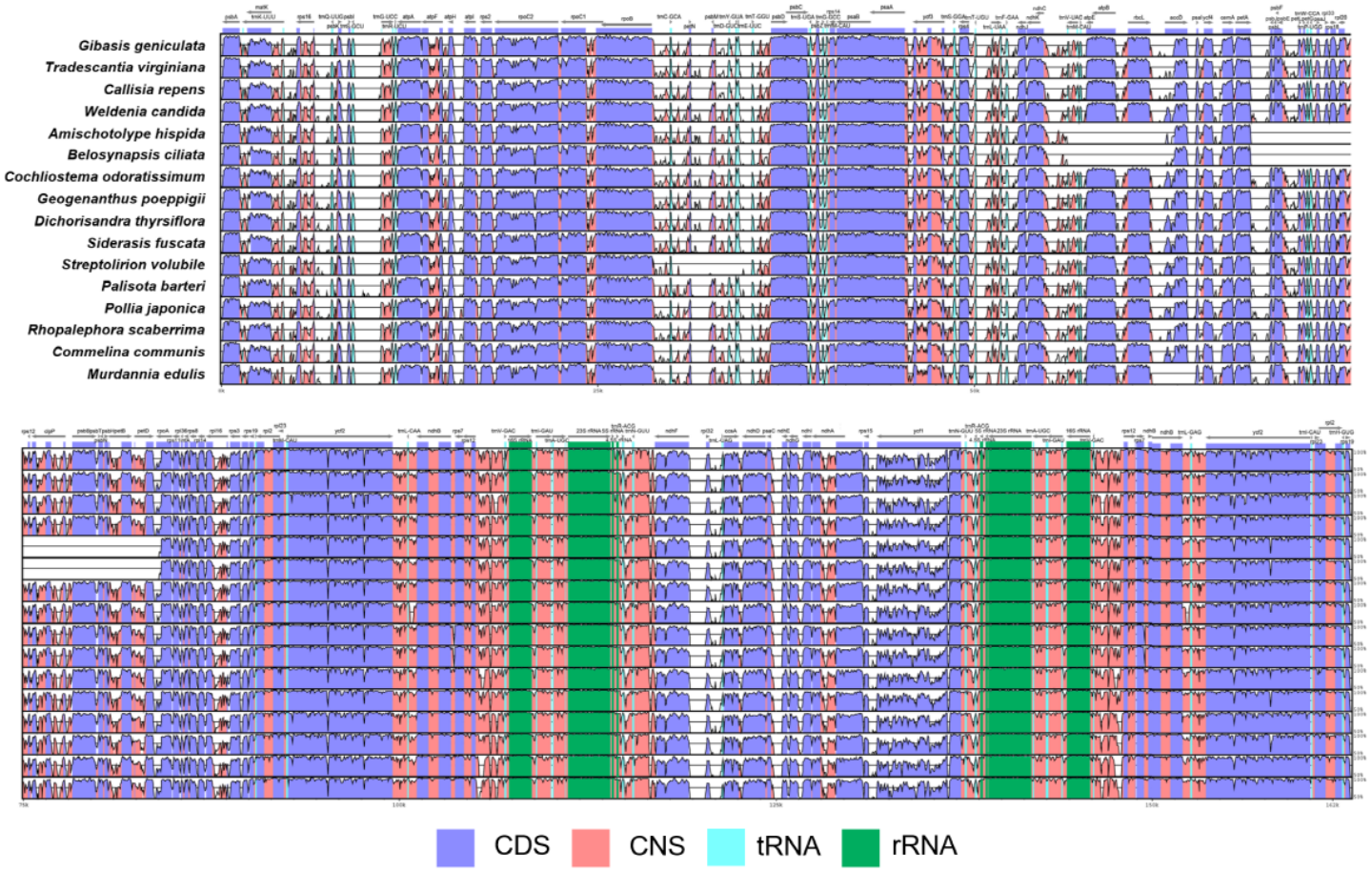

Figure 3

Plots of percent sequence identity of the chloroplast genomes of 16 Commelinaceae species with Hanguana malayana as a reference. The percentage of sequence identities were estimated and the plots were visualized in mVISTA.

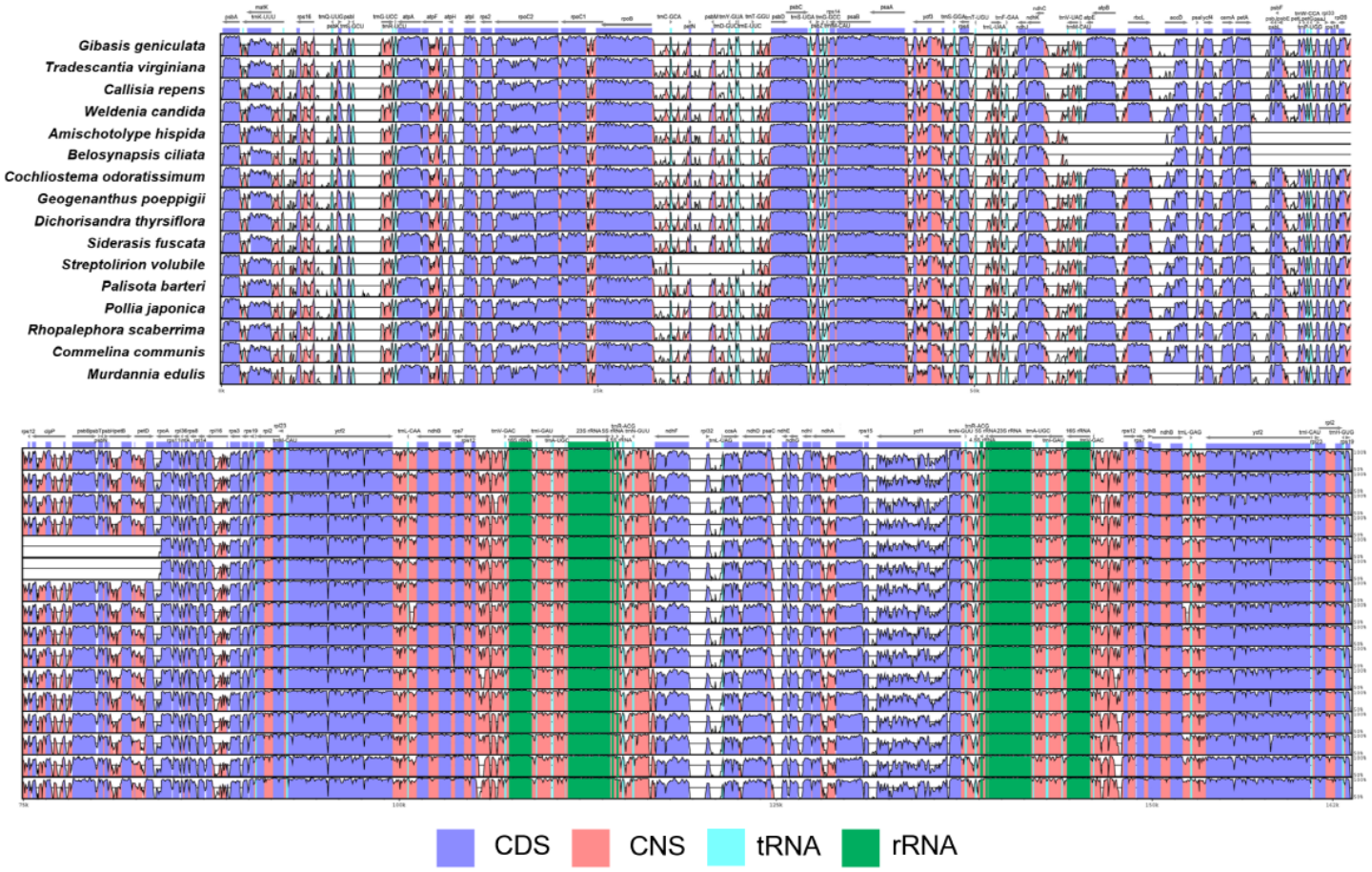

Figure 3

Plots of percent sequence identity of the chloroplast genomes of 16 Commelinaceae species with Hanguana malayana as a reference. The percentage of sequence identities were estimated and the plots were visualized in mVISTA.

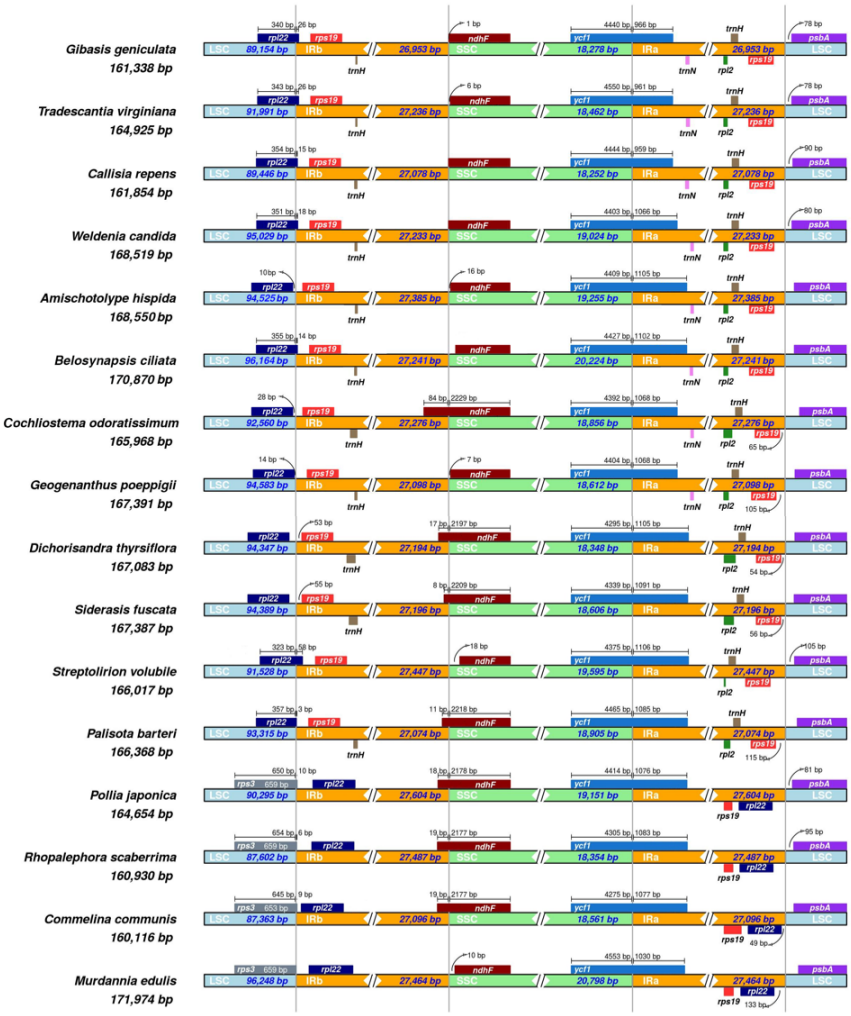

Figure 4

Comparisons of LSC, SSC, and IR regions boundaries between 16 Commelinaceae species.

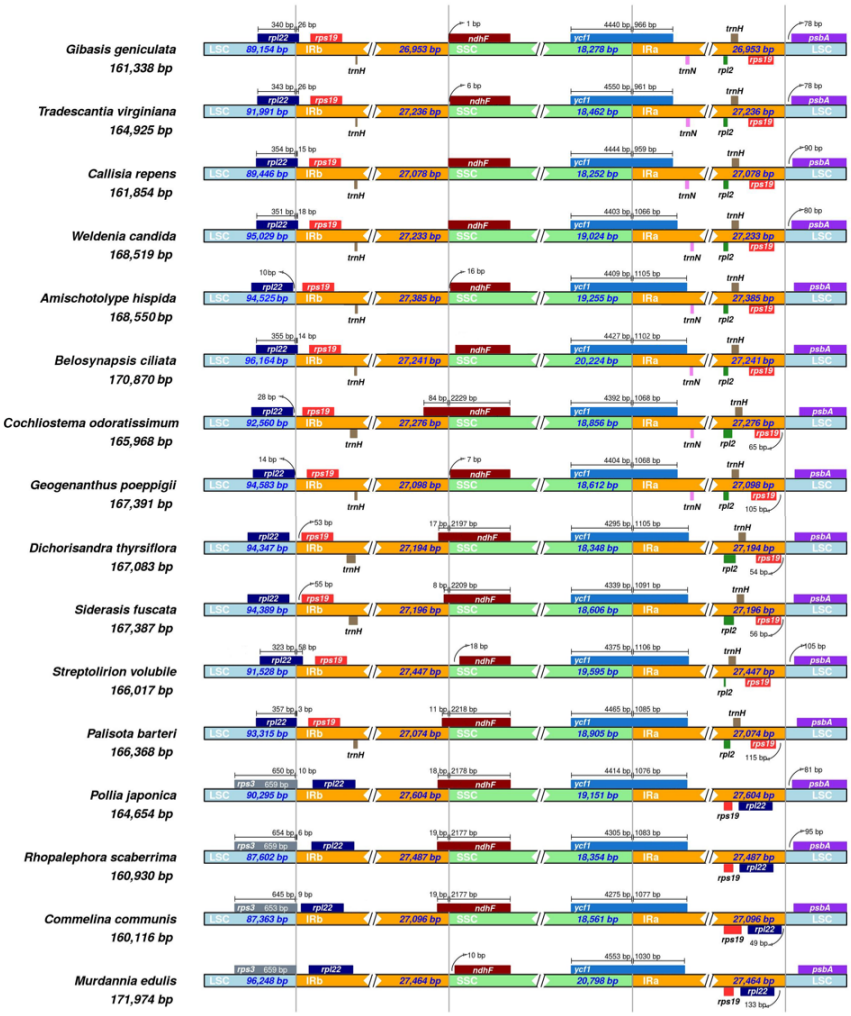

Figure 4

Comparisons of LSC, SSC, and IR regions boundaries between 16 Commelinaceae species.

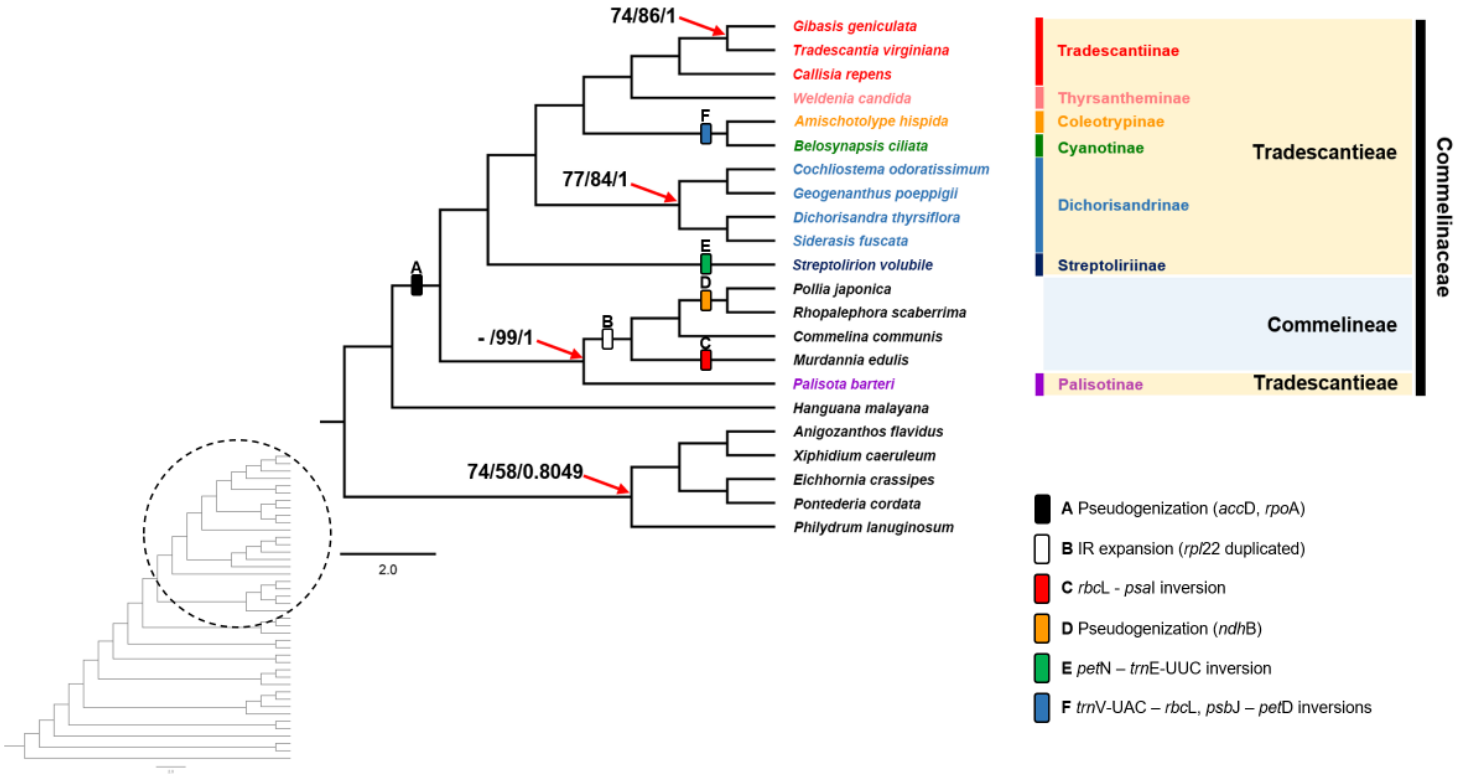

Figure 5

The Maximum Likelihood tree of 42 monocots inferred from 77 chloroplast protein coding genes. Numbers indicate support (maximum parsimony bootstrap (PBP)/maximum likelihood bootstrap (MBP)/posterior probability (PP)). Only support under PBP = 90/MBP = 100/PP = 1.00 is shown. The dashes "-" indicate incongruence between MP and ML/BI trees.

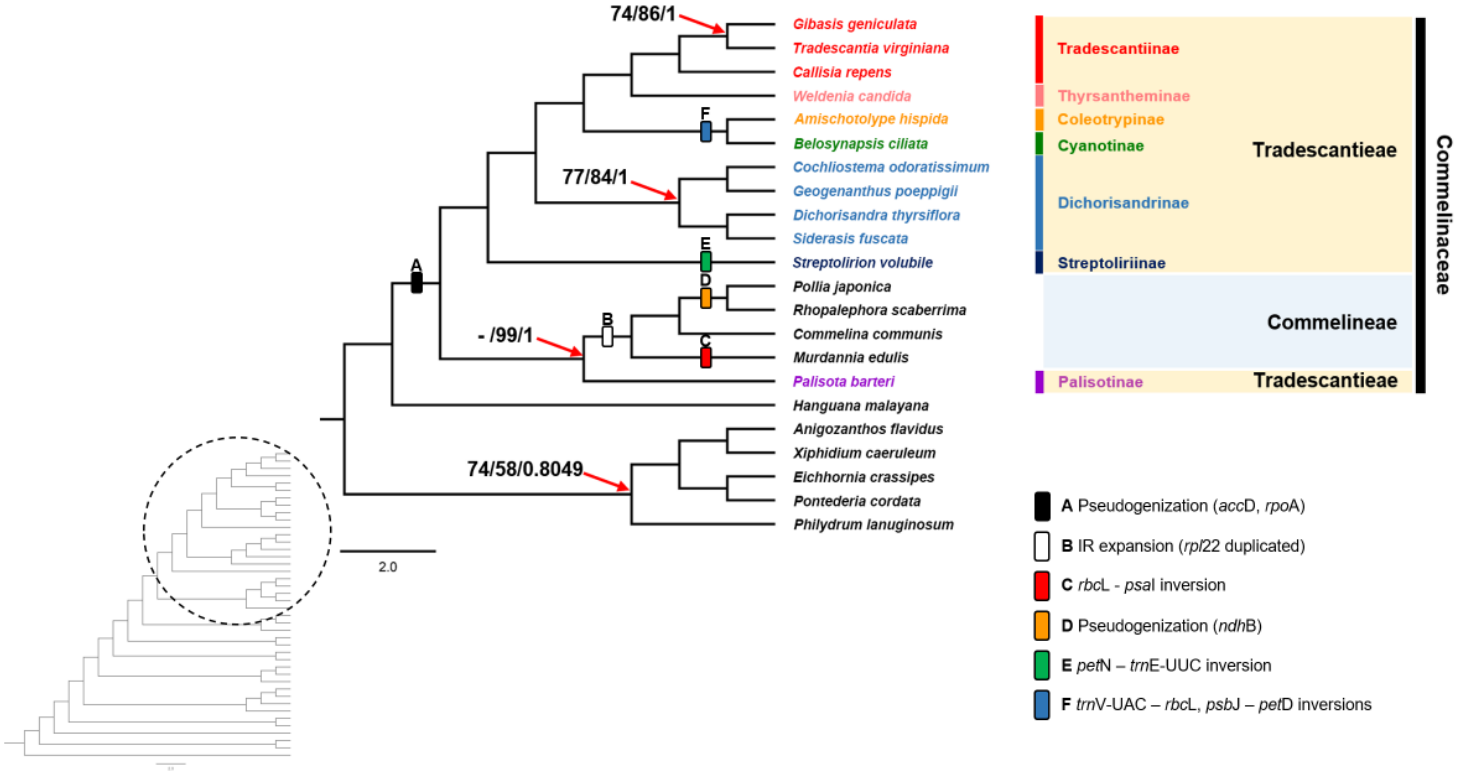

Figure 5

The Maximum Likelihood tree of 42 monocots inferred from 77 chloroplast protein coding genes. Numbers indicate support (maximum parsimony bootstrap (PBP)/maximum likelihood bootstrap (MBP)/posterior probability (PP)). Only support under PBP = 90/MBP = 100/PP = 1.00 is shown. The dashes "-" indicate incongruence between MP and ML/BI trees.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- FigureS1.png
- FigureS1.png
- FigureS2.png
- FigureS2.png
- TableS1.xlsx
- TableS1.xlsx
- TableS2.xlsx
- TableS2.xlsx
- TableS3.xlsx
- TableS3.xlsx