BMC Plant Biology

RESEARCH

Open Access

Comparative plastome analyses and phylogenetic insights of *Elatostema*

Yu-Hsin Tseng[1*], Han-Chun Chien[1] and Geng-Xi Zhu[1]

Abstract

Background *Elatostema*, one of the largest genera in Urticaceae, comprises approximately 570 species. The taxonomic delimitation of *Elatostema* and its closely related genera, *Elatostematoides* and *Procris* and the infrageneric classification of *Elatostema*, have historically been challenging. Previous studies have been limited by insufficient molecular data, hindering our understanding of species-level relationships and the evolution of plastid genomes in this group. To address these limitations, we assembled and analyzed a comprehensive plastome analysis of 42 species across *Elatostema* and its allied genera. Our study focused on plastome structure, sequence diversity, and phylogenetic relationships to elucidate the evolutionary history of these taxa.

Results Our findings reveal that *Elatostema* plastomes exhibit a typical quadripartite structure, with genome sizes ranging from 149,152 bp to 164,019 bp. Comparative analysis of plastome structures across *Elatostema* and its related genera indicates high conservation in genome size, structure, gene content, and inverted repeat boundary configuration. Our findings indicate a strong association between the length of small single-copy (SSC) regions and phylogenetic grouping within *Elatostema* and between *Elatostema*, *Elatostematoides* and *Procris*. The length variations in the *ndhF-rpl32*, *rpl32-trnL*, and *rps15*-SSC/IRa regions may account for this observed correlation, highlighting the utility of SSC sequences in resolving phylogenetic relationships within this genus. Furthermore, we identified seven highly variable regions with potential as DNA barcodes for species identification and phylogenetic analysis. Our phylogenomic analysis provides robust support for the taxonomic delimitation of *Elatostema s.l.* into three distinct genera: *Elatostema*, *Procris*, and *Elatostematoides*. We also reconfirm the infrageneric classification of *Elatostema* into four major clades.

Conclusions The utilization of plastome sequences has enabled a highly resolved phylogenetic framework, shedding light on the evolutionary history and speciation mechanism within *Elatostema*, particularly its species-rich core *Elatostema* clade. These findings provide a valuable foundation for future taxonomic revisions and evolutionary studies within this challenging plant group.

Keywords *Elatostema*, *Elatostematoides*, Plastome, *Procris*, Structural variation, Urticaceae

Background

Elatostema J.R. Forst. & G. Forst. (Urticaceae) is a group of approximately 570 species [1–3] of understory herbs, subshrubs, and shrubs. *Elatostema* is most common in areas of deep shade, particularly in dense forests, along streams, at/near waterfalls, and in caves, where they can become a dominant element of the forest understory. They are distributed in tropical and subtropical Africa, East and Southeast Asia, and Australasia, with centers of

*Correspondence:
Yu-Hsin Tseng
yuhsin.tseng@nchu.edu.tw
[1] Department of Life Sciences, National Chung Hsing University, Taichung 40227, Taiwan

species richness in Southeast Asia and Southwest China [1, 2].

The taxonomic delineation of *Elatostema* (abbreviated as *E.*) and its closely related taxa, *Elatostematoides* C.B. Rob. (abbreviated as *Eld.*), *Pellionia* Gaudich., and *Procris* Comm. ex Juss. have historically been challenging due to the large number of taxa and the reliance on homoplastic morphological characters for identification [1, 4–8]. These genera, collectively referred to as *Elatostema* sensu lato, have been the subject of considerable taxonomic debate since Weddell's initial revision in 1869 [9]. While numerous regional revisions and flora accounts have contributed to understanding these taxa, the inter- and infrageneric delimitations have been controversial for a long time. Major taxonomic treatments of *Elatostema s.l.* have generally fallen into three categories: 1. recognizing four distinct genera: *Elatostema*, *Elatostematoides*, *Pellionia*, and *Procris* [7–11]; 2. treating *Pellionia* and *Procris* as subgenera of *Elatostema* [12, 13]; 3. recognizing *Procris* as a distinct genus while treating *Pellionia* and *Elatostematoides* as subgenera of *Elatostema* [14–16].

In the study by Tseng et al. [3], phylogenetic analyses of molecular data were conducted to investigate the circumscription and relationships of *Elatostema s.l.* These analyses revealed three well-supported and morphologically distinct genera: *Elatostema*, *Elatostematoides*, and *Procris*. Within the newly defined *Elatostema*, four strongly supported clades were identified: core *Elatostema*, *Pellionia*, *Weddellia*, and an African *Elatostema*. Despite the challenges posed by homoplasy in morphological characters, combined character suites enabled the morphological diagnosis of all three genera. Although this phylogenetic study [3] has provided sufficient resolution at the inter- and infrageneric level of *Elatostema*, resolution at the species level within the four major clades of *Elatostema* has proven recalcitrant, probably due to the limited molecular markers. Therefore, further research with more extensive molecular data is necessary to clarify the intraspecific relationships, evolutionary history, and drivers of speciation within *Elatostema* [3].

Plastome structure has emerged as an indispensable tool for elucidating the evolutionary history of plants. Despite evolving at a rate approximately half that of nuclear DNA [17], plastome resources have been extensively utilized across various areas of plant science, including phylogenomics, evolutionary biology, comparative genomics, population genomics, phylogeography, and chloroplast genetic engineering [18–20]. The typical structure of a plastome is characterized by a circular molecule composed of two large rRNA-encoding inverted repeats (IRs) separated by a large single copy (LSC) region and a small single copy (SSC) region, forming a quadripartite structure [17]. The confluence of advancements in sequencing technologies, assembly techniques, and annotation tools has spurred a rapid accumulation of complete plastome sequences. However, *Elatostema*, *Elatostematoides* and *Procris* remain underrepresented in plastome research, with only a few being published and analyzed to date [21–26]. Detailed comparative genomics and plastome phylogenomics studies have yet to be undertaken for these taxa.

While plastome research in Urticaceae remains relatively limited, recent studies have demonstrated the significant potential of plastomes in this family. Comparative plastid genomics of four *Pilea* species (Urticaceae) identified eight hypervariable regions, providing promising molecular markers for species-level identification [27]. Furthermore, comparative plastome analyses of *Debregeasia* revealed nine genes (*clpP*, *ndhF*, *petB*, *psbA*, *psbK*, *rbcL*, *rpl23*, *ycf2*, and *ycf1*) that may have undergone positive selection, further suggesting the crucial role of the *ycf1* gene in speciation and habitat adaptation within *Debregeasia* [26]. Plastome phylogenomics of *Boehmeria s.l.* inferred the polyphyly of this genus, and plastome structural variation further corroborated these phylogenetic relationships [28]. Wu et al., [29] reconstructed the plastome phylogenomics of *Oreocnide*, confirming its monophyly and estimating its diversification time at approximately 6.06 Ma. Plastome sequences have also been applied in the tribe Urticeae of Urticaceae to resolve deep relationships within the tribe and evaluate and update existing classifications for Urticeae [30].

This study presents an analysis of 42 newly assembled and annotated plastomes from the Urticaceae family, encompassing 38 species of *Elatostema*, two species of *Elatostematoides*, and two of *Procris*. In our initial comparative analysis of plastome structure, we observed a distinctive pattern: the length of SSC appears to be correlated with the phylogenetic groupings. Given the critical role of the SSC region in plastome evolution, we further investigated the underlying causes of this pattern by examining whether specific regions—such as coding genes or intergenic spacers—contribute to the observed variation in SSC length. Based on these observation, the objectives in this study are to: (1) characterize and compare plastome structure and gene organization, with a specific focus on the relationship between SSC length variation and phylogenetic relatedness; (2) identify putative repeated sequences; (3) detect potential molecular markers to support phylogenetic analyses; and (4) reconstruct plastome-based phylogenomic relationships to elucidate the characteristics, structural diversity, and evolutionary history of *Elatostema*, *Elatostematoides*, and *Procris*.

Methods

Sampling

The samples in this study included 38 species of *Elatostema*, representing four major clades in *Elatostema* [3] (four of African *Elatostema*, 28 of core *Elatostema*, four of *Pellionia* and two of *Weddellia*). To compare the plastome structure to the allied genera of *Elatostema*, two species of *Elatostemtoides* and two of *Procris* were also included. Fresh samples were collected and deposited at HAST, IBK, TAIF, and TNM. Herbarium specimens were sampled from BM, K, MO, and KYO. In addition to the newly sequenced species, five complete plastome sequences of *Elatostema* (*E. dissectum, E. parvum, E. qinzhouense*, *E. scabrum* and *E. stewardii*; accession number: MK227819, OM761906, MW172521, OL800583, MZ292972, respectively) and the plastome sequence of *Procris crenata* (MW114890) were downloaded from GenBank for structure comparative, repeat sequence and phylogenomic analyses. For phylogenomic analysis, 30 species of Urticaceae from GenBank were chosen to represent each major Urticaceae clade [31] and three species of Moraceae, *Artocarpus altilis*, *Broussonetia papyrifera*, and *Morus mongolica*, were used as outgroups. The sample and voucher information are provided in Additional File 1.

DNA extraction, library preparation and sequencing

Total genomic DNA was extracted from fresh or dried leave samples following Li et al., [32]. The DNA concentration was measured by Qubit 3.0 Fluorometer (Thermo Scientific, Massachusetts, USA). Approximately 500–1 μg per DNA sample was sheared by Bioruptor Pico (Cosmobio Inc., Tokyo, Japan) into fragments of about 350 bp. Dual-indexed libraries were made using the NEBNext Ultra II DNA Library Prep Kit (New England BioLabs, Massachusetts, USA), following the 300–400 bp insert size protocol. Libraries for paired-end 150 bp sequencing were conducted using an Illumina NovaSeq X Plus platform to generate approximately 2 Gb of data per sample.

Plastome assembly and annotation

The quality of raw reads was assessed in FastQC v0.12.0 [33] and low-quality reads were filtered in Trimmomatic v0.39 [34]. De novo assembling of the plastome used GetOrganelle v1.7.3.s with default parameters [35]. The draft plastomes were verified assembly errors by using "Map to reference" in Geneious Prime v2024.0.4. Complete plastome assemblies were annotated using the GeSeq web tool [36]. The annotations were then manually reviewed and refined in Geneious Prime to ensure accurate identification of start and stop codons for each gene, using the plastome sequence of *Elatostema stewardii* [25] as a reference. Annotated genomes were visualized by created in OGDRAW v1.3.1 [37]. The length of open reading frames (ORFs) and coding potential of genes was verified using Geneious by translating the regions and predicting the corresponding CDS lengths.

Genome comparison and analysis

To compare the junction site of LSC-IRa/a and SSC-IRa/b, we used IRplus [38] to identify IR expansion or contraction within the plastome of *Elatostema*, *Elatostematoides* and *Procis* species. The nucleotide diversity (*pi*) of 38 newly generated plastomes *Elatostema* and five *Elatostema* plastomes from NCBI was evaluated by DnaSP v6.10 [39]. First, all plastome sequences were aligned by MAFFT v7.45 [40] and manually adjusted in Mesquite v3.7 [41]. To avoid potential biases in the nucleotide diversity analysis caused by a ~ 7,000 bp insertion located between the *ycf15* to *rrn16* RNA regions within IR observed in five species (see Results for a detailed explanation), this region was excluded from alignment prior to the following analysis. The sliding window analysis was performed with DnaSP v6.10 [39], utilizing a step size of 200 bp and a window length of 600 bp. The variable and parsimony-informative sites of potential DNA barcode were calculated by AMAS [42]. To further investigate the correlation between SSC length and phylogenetic relatedness, the SSC regions were divided into gene sequences and intergenic spacers based on their annotations. Given the variability in the length of the *ycf1* gene across different plastomes in SSC, the region extending from the end of the *rps15* gene to the SSC/IRa boundary was separately extracted from the SSC for comparison, hereafter referred to as the " *rps15* to SSC/IRa" region. Chi-square values ($\chi 2$), degrees of freedom (df), and significance (p) calculated with the ANOVA function of the R package 'car' [43] to test the correlation between SSC length and phylogenetic grouping.

Simple repeat sequence and long repeat characterization of *Elatostema* plastomes

Repeat sequences were analyzed to calculate the content of dispersed and tandem repeats in LSC, single IR and SSC regions. The simple sequence repeats (SSRs) of *Elatostema*, *Elatostematoides* and *Procris* were identified by a MISA perl script [44]. Thresholds for the minimum number of repeats were set as ten, five, and four repeat units for mono-, di- and trinucleotides SSRs, respectively and three repeat units for each tetra-, penta- and hexanucleotides SSRs. The results were visualized in the balloon plot generated in R v.4.0.1 using 'reshape' [45] and 'ggplot2' [46] packages. The identification of dispersed repeats (forward, reverse, complement, and palindromic) were analyzed using REPuter [47] with a 20 bp minimum repeat size, and 120 bp maximum computed repeated.

Phylogenomic analysis

A total of 43 plastome sequences, including 38 newly generated and five retrieved from NCBI, were used in the phylogenomic analysis. These sequences represented four major clades of *Elatostema*, two species of *Elatostematoides* and three species of *Procris* (two newly generated and *Procris crenata* from NCBI). Furthermore, 30 species of Urticaceae available from GenBank were chosen to represent four major Urticaceae clades (15 of Clade I, four of Clade II, 10 of Clade III and one of Clade IV) [31]. Three species of Moraceae, *Artocarpus altilis*, *Broussonetia papyrifera*, and *Morus mongolica*, were used as outgroups. The sequences were aligned using MAFFT v7.45 [40] and manually adjusted by Mesquite v3.7 [41]. The aligned sequence was divided into two datasets for phylogenetic analysis. The first, referred to as partitioned dataset, included genic region (including rRNA genes, tRNA genes and coding sequences (CDS)) and non-genic regions (including intron and spacers), which were treated as different partition. To minimize the potential bias from indels in non-genic regions, a second dataset consisting only of the CDS regions (referred to as CDS dataset) was also analyzed for comparison. Phylogenetic reconstructions were conducted using both maximum likelihood (ML) and Bayesian inference (BI) approached, implemented in IQ-TREE v.2.0.6 [48] and MrBayes v3.2.7a [49], respectively. For ML tree of partitioned dataset, ModelFinder in IQ-TREE as used to identify the best-fit partitioning scheme and substitution models, followed by 1,000 ultrafast bootstrap (UFBoot) replicates to assess node support. For BI analyses of the partitioned dataset, the best models fitting into MrBayes was evaluated by ModelFinder in IQ-TREE (using "-m TESTMERGEONLY -mset mrbayes" options). In MrBayes, the priors were set according to the results of IQ-TREE with models unlinked among each partition and set to run the Marko chain Monte Carlo (MCMC) simulation of 20,000,000 generations with two independent runs and four chains and sampling frequency of every 500 generations. The first 25% of sampled trees were discarded as burn-in and posterior probabilities (PP) were estimated from the remaining trees. The effective sample sizes (ESS) were also assessed for all parameters and statistics using Tracer v1.7.1 [50]. All ESS were obtained with values higher than 200, indicating that all parameters were sampled sufficiently for all chains to converge. For the CDS dataset, ML and BI analyses were performed using the same settings as for the partitioned dataset, but without applying further partitioning.

Results

Comparison of plastome structure and IR boundaries

All newly generated plastomes of *Elatostema* in this study exhibited the typical quadripartite structure as shown in Fig. 1. The genome sizes within the genus ranged from 149,152 bp (*E. monandrum*) to 164,019 bp (*E. oblongifoium*). Each plastome included a pair of inverted repeats (IRs) ranging from 24,436 bp to 31,509 bp, a large single-copy region (LSC) from 83,041 bp to 85,864 bp, and a small single-copy region (SSC) from 16,850 bp to 18,075 bp. A comparative analysis of plastome structures revealed no significant differences between *Elatostema*,

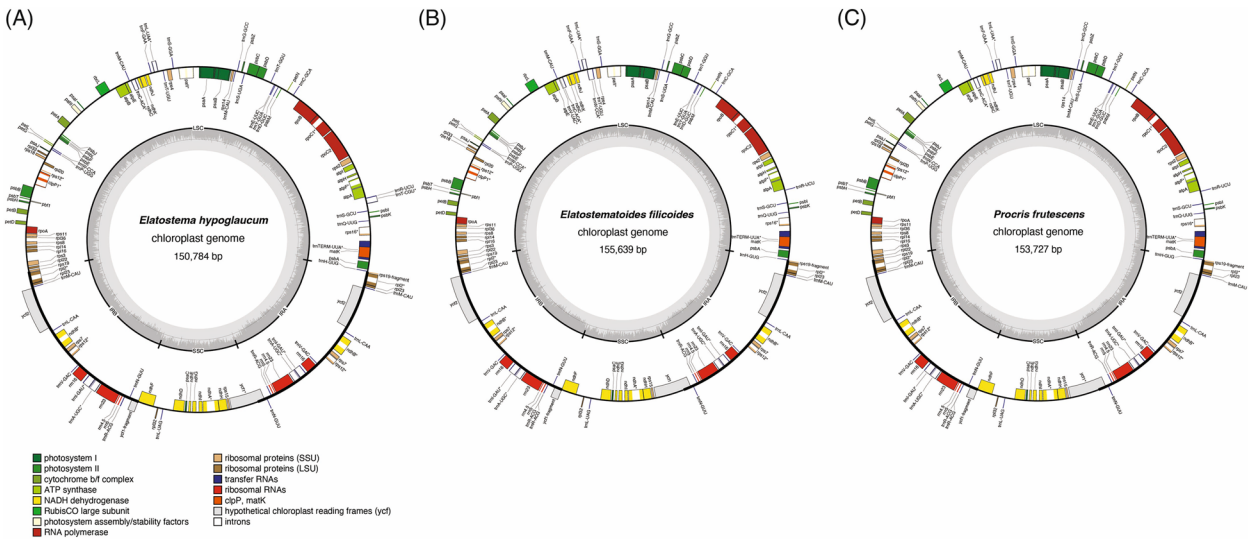

Fig. 1 Gene maps of plastid genomes for (**A**) *Elatostema hypoglacucum*, (**B**) *Elatostematoides filicoides*, and (**C**) *Procris frutescens*. The outer circles highlight the inverted repeats (IRs) with bold lines. The inner circles represent GC content across the genome, with lighter grey denoting AT content. The species name and genome size are displayed at the centre of each plot. Genes are colour-coded by functional group, and an asterisk (*) indicates genes containing introns

Elatostematoides, and *Procris* (Fig. 1). While the lengths of the plastome, LSC, IR, and GC content in *Elatostematoides* and *Procris* fell within the range observed in *Elatostema*, the SSCs and LSCs of these genera exhibited some variations (Table 1). The SSC regions of *Elatostematoides* and *Procris* were longer than those of *Elatostema*, ranging between 18,424 bp and 18,435 bp and 18,079 bp and 18,207 bp, respectively (explained below). In addition, the LSC of *Elatostematoides* (85,353 bp to 85,864 bp) was slightly larger than that of *Elatostema* and *Procris* (84,672 bp to 84,764 bp) (Table 1). The GC content across plastomes *Elatostema*, *Elatostematoides* and *Procris* was relatively consistent, ranging from 36.2% to 36.7% (Table 1).

The length of the SSC exhibited a strong correlation with the phylogenetic relatedness among the four clades in *Elatostema* ($\chi^2 = 22.124$, df = 3, $P < 0.001$) (Fig. 2). The core *Elatostema* clade displayed the shortest SSC lengths (16,850 bp in *E. tianeese* to 17,341 bp in *E. grandifolium*), followed by *Weddellia* clade (17,732 bp in *E. parvum* to 17,817 bp in *E. sinense* var. *longicornutum*, African *Elatostema* clade (17,864 bp in *E. madagascariense* and 17,894 bp in *E. goudotianum*) and *Pellionia* clade with the longest SSC lengths (18,018 bp of *E. radicans* and 18,075 bp of *E. scabrum*) (Fig. 2A, Table 1). While the lengths of all 12 genes and 8 out of 10 spacers (*ccsA-ndhD*, *ndhD-psaC*, *ndhE-ndhG*, *ndhG-ndhI*, *ndhH-rps15*, *ndhI-ndhA*, *psaC-ndhE*, *trnL-ccsA*) within the SSC were consistent across all samples (Fig. 2B, C) ($\chi^2 = 0.11062$, df = 3, $P = 0.9905$), *ndhF-rpl32* spacer, *rpl32-trnL* spacer and *rps15* to SSC/IRa region exhibited significant variation in sequence length, aligning with the phylogenetic relatedness (Fig. 2C, D) ($\chi^2 = 19.306$, df = 3, $P < 0.001$). The *ndhF-rpl32* spacer was, on average, longer in the *Weddellia* clade (average: 713.5 bp, 713 bp–714 bp) compared to the core *Elatostema* clade (average: 648.6 bp, 519 bp–758 bp), but still within the range of core *Elatostema* clade. The African *Elatostema* and *Pellionia* clades displayed slightly longer *ndhF-rpl32* spacers (833 bp–842 bp, 847 bp–880 bp, respectively). The *rpl32-trnL* spacer was shortest in the core *Elatostema* clade (546 bp–660 bp), followed by African *Elatostema* clade (818 bp to 823 bp), *Pellionia* clade (913 bp–931 bp) and *Weddellia* clade (921 bp to 943 bp) (Fig. 2C). Similarly, the end of *rps15* in SSC to SSC/IRa region was shortest in the core *Elatostema* clade (4,123 bp–4,040 bp), followed by *Weddellia* clade (4,146 bp–4,149 bp), *Pellionia* clade (4,491 bp–4497 bp) and African *Elatostema* clade (4,503 bp–4,515 bp) (Fig. 2D).

Among these three genera, *Elatostematoides* exhibited the longest SSC and *Elatostema* had the shortest (Fig. 2A). The length of SSC showed significant differences among *Elatostema*, *Elatostematoide* and *Procris* ($\chi^2 = 10.632$, df = 2, $P < 0.005$). Consistent with the findings in *Elatostema*, except for the three variable regions (*ndhF-rpl32* spacer, *rpl32-trnL* spacer, and *rps15* to SSC/IRa region) (Fig. 2C, D), all 12 genes and 8 of 10 spacers within the SSC displayed similar lengths across all samples (Fig. 2B) ($\chi^2 = 0.013298$, df = 2, $P = 0.9934$). In the three variable regions (*ndhF-rpl32* spacer, *rpl32-trnL* spacer and *rps15* to SSC/IRa region), *Elatostematoides* and *Procris* had a significantly longer sequence length (average length of 948.5 bp, 1112.5 bp and 4513.5 bp in *Elatostematoides* and 980.5 bp, 909.5 bp and 4,531 bp in *Procris*) compared to *Elatostema* (694.7 bp, 677.3 bp and 4071.6 bp) (Fig. 2B, C) ($\chi^2 = 8.962$, df = 2, $P = 0.0113$).

Gene order, content, and number of plastomes were generally identical in *Elatostema*, *Elatostematoides*, and *Procris* (Tables 1, 2 and Fig. 1). The plastomes assembled in this study contained 113 unique genes, including 79 unique protein-coding genes, 30 tRNA genes, and four rRNA genes, excluding five species in *Elatostema* (*E. huanjiangense*, *E. hechiense*, *E. gyrocephalum*, *E. oblongifolium*, and *E. malacotrichum*). These five species of core *Elatostema* clade underwent a deletion of the *ycf15* gene due to an approximately 7,000 bp insertion in IR region. In addition, significant variations of sequence length were observed in several genes in other species, including *matK*, *ndhF*, *rps3*, *ycf1* and *ycf2*. Compared to the typical *matK* gene length of 1530–1551 bp, *E. strigillosum* and *E. edule* had a relatively short truncated *matK* gene (621 and 717 bp, respectively). Similar patterns were also found in *ndhF* gene of *E. radicans*, *E. viridis* and *Eld. filicoides* (typical cases with 2,235–2,271 bp vs. expectations cases with 1569–1572 bp), *rps3* of *E. gyrocephalum* (639–648 bp vs. 324 bp), *ycf1* of *E. malacotrichum* and *E. scabrum* (4,587–5,511 bp vs. 2,883 and 3,600 bp) and *ycf2* of *E. strigillosum* (1041 bp vs. 5928–6771 bp). The detailed sequence variations of these genes are shown in Additional File 2. Within the IR regions, we identified 10 duplicated genes, including eight protein-coding genes (*ndhB*, *ndhF*, *rpl2*, *rpl23*, *rps19*, *rps7*, *ycf15*, *ycf2*), seven tRNA genes (*trnA-UGC*, *trnI-CAU*, *trnI-GAU*, *trnL-CAA* , *trnN-GUU*, *trnR-ACG*, *trnV-GAC*) and three rRNA gene (*rrn16*, *rrn23*, *rrn4.5*, *rrn5*). Of the 113 unique genes, gene functions and types in all newly assembled plastomes in this study are shown in Table 2.

The LSC/IRb and IRa/LSC boundaries were highly conserved in all newly assembled plastomes in *Elatostema*, *Elatostematoides* and *Procris* (Fig. 3). The LSC/IRb boundary was located within *rps19* gene, extending 112 bp to IRb, while the IRa/LSC boundary was adjacent to *trnH* in LSC, with 112 bp of *rps19* duplicated in the IRa (Fig. 3). Two distinct boundary types were identified for IRb/SSC and SSC/IRa junctions based on the sequence length and position of *ndhF* (Fig. 3). In Type

Table 1 Statistics of newly generated plastomes of *Elatostema*, *Elatostematoides* and *Procris* in this study

Genus / Species	Clade in *Elatostema*	Total length (bp)	LSC (bp)	SSC (bp)	IR (bp)	GC contect	Total reads	Number of reads mapped	% of read mapped	Average coverage	NCBI accession number
Elatostema											
E. goudotianum	African *Elatostema*	154,132	85,066	17,894	25,586	36.5%	14,606,089	262,391	1.80	228.866 ± 45.4854	PQ676460
E. incisum	African *Elatostema*	154,152	85,097	17,883	25,586	36.5%	18,351,401	302,878	1.65	249.896 ± 66.6245	PQ676461
E. madagascariense	African *Elatostema*	154,050	85,014	17,864	25,586	36.5%	19,173,014	263,378	1.37	212.807 ± 80.8444	PQ676462
E. orientale	African *Elatostema*	154,097	85,053	17,872	25,586	36.5%	46,065,484	774,433	1.68	670.646 ± 166.904	PQ676463
E. brachyodontum	core *Elatostema*	151,222	84,530	17,154	24,769	36.3%	22,177,759	442,531	2.00	365.799 ± 225.117	PQ676442
E. calcareum	core *Elatostema*	151,197	84,688	17,175	24,667	36.3%	14,288,885	328,666	2.30	274.598 ± 104.979	PQ676451
E. edule	core *Elatostema*	151,110	84,673	17,037	24,700	36.2%	20,352,666	1,047,958	5.15	975.361 ± 263.465	PQ776216
E. grande	core *Elatostema*	149,710	83,437	16,939	24,667	36.3%	23,047,370	1,175,810	5.10	944.912 ± 647.699	PQ676448
E. grandifolium	core *Elatostema*	149,742	83,529	17,341	24,436	36.3%	29,292,636	820,209	2.80	616.904 ± 446.142	PQ676449
E. gyrocephalum	core *Elatostema*	163,566	84,258	16,868	31,220	36.7%	15,666,556	305,321	1.95	228.313 ± 161.841	PQ676441
E. hechiense	core *Elatostema*	163,669	84,227	17,050	31,196	36.6%	15,152,390	1,031,900	6.81	898.785 ± 245.325	PQ676439
E. hezhouense	core *Elatostema*	150,624	84,052	17,118	24,727	36.3%	24,414,880	838,186	3.43	707.4 ± 232.202	PQ676445
E. huanjiangense	core *Elatostema*	162,829	84,414	16,917	30,749	36.6%	15,634,752	1,559,857	9.98	1375.57 ± 279.385	PQ676440
E. hypoglaucum	core *Elatostema*	150,784	84,395	16,951	24,719	36.4%	15,750,947	408,609	2.59	374.832 ± 107.513	PQ676437
E. laevissimum	core *Elatostema*	150,966	84,387	17,115	24,732	36.3%	34,709,067	2,573,232	7.41	2365.03 ± 554.246	PQ676457
E. lithoneurum	core *Elatostema*	150,601	84,265	17,012	24,662	36.3%	23,974,155	1,656,092	6.91	1413.11 ± 740.775	PQ676447
E. lutescens	core *Elatostema*	151,315	84,945	17,020	24,675	36.2%	17,776,909	910,678	5.12	765.898 ± 234.538	PQ676450
E. macintyrei	core *Elatostema*	150,808	84,292	17,158	24,679	36.4%	26,896,081	744,546	2.77	444.686 ± 1241.21	PQ676456
E. malacotrichum	core *Elatostema*	161,620	83,867	17,063	30,345	36.5%	16,156,497	460,044	2.85	363.739 ± 142.294	PQ676443
E. monandrum	core *Elatostema*	149,152	83,041	16,945	24,583	36.4%	37,308,535	2,713,549	7.27	2401.67 ± 1155.61	PQ676459
E. nasutum	core *Elatostema*	150,034	83,735	17,051	24,624	36.3%	18,740,015	709,883	3.79	624.918 ± 245.55	PQ676458
E. oblongifolium	core *Elatostema*	164,018	84,116	16,884	31,509	36.6%	15,435,826	493,980	3.20	409.16 ± 175.302	PQ676471
E. platyphyllum	core *Elatostema*	150,675	84,416	17,271	24,494	36.3%	10,178,771	380,567	3.74	339.988 ± 185.406	PQ676435
E. rugosum	core *Elatostema*	151,006	84,493	17,059	24,727	36.4%	22,917,736	548,903	2.40	441.067 ± 579.421	PQ676454
E. serra	core *Elatostema*	150,866	84,307	17,079	24,740	36.4%	21,986,067	632,316	2.88	554.169 ± 235.444	PQ676453
E. strigillosum	core *Elatostema*	150,623	84,348	16,911	24,682	36.2%	22,236,076	1,049,192	4.72	883.61 ± 432.672	PQ776217
E. subcoriaceum	core *Elatostema*	150,737	84,453	16,874	24,705	36.4%	14,338,300	354,568	2.47	322.682 ± 91.7588	PQ676472
E. suzukii	core *Elatostema*	150,817	84,428	16,987	24,701	36.3%	13,682,168	289,512	2.12	227.808 ± 216.003	PQ676452
E. tianeese	core *Elatostema*	150,527	84,301	16,850	24,688	36.3%	12,620,625	264,391	2.09	233.895 ± 82.3609	PQ676444
E. villosum	core *Elatostema*	150,656	84,373	16,917	24,683	36.2%	24,929,330	869,133	3.49	716.399 ± 358.489	PQ676438
E. welwitschii	core *Elatostema*	150,252	83,720	17,132	24,700	36.5%	26,769,513	1,269,419	4.74	907.425 ± 1447.58	PQ676455
E. yachense	core *Elatostema*	150,347	83,951	17,088	24,654	36.3%	19,934,022	861,946	4.32	765.217 ± 341.894	PQ676446
E. grijsii	*Pellionia*	154,282	85,120	18,062	25,550	36.4%	6,895,564	133,709	1.94	117.417 ± 47.1346	PQ676465

Table 1 (continued)

Genus / Species	Clade in *Elatostema*	Total length (bp)	LSC (bp)	SSC (bp)	IR (bp)	GC contect	Total reads	Number of reads mapped	% of read mapped	Average coverage	NCBI accession number
E. radicans	*Pellionia*	154,159	84,969	18,018	25,586	36.4%	19,647,562	303,648	1.55	195.119 ± 632.659	PQ676433
E. scabrum	*Pellionia*	154,251	85,004	18,075	25,586	36.4%	12,226,995	182,101	1.49	141.686 ± 155.942	PQ676436
E. viridis	*Pellionia*	154,249	85,092	18,057	25,550	36.4%	28,904,574	517,536	1.79	320.364 ± 786.235	PQ676464
E. parvum	*Weddellia*	152,198	83,986	17,732	25,240	36.4%	12,717,456	94,068	0.74	76.1447 ± 67.8084	PQ676434
E. sinense var. longicornutum	*Weddellia*	152,759	84,328	17,817	25,307	36.4%	15,453,820	152,713	0.99	103.68 ± 146.712	PQ676466
Elatostematoides											
Eld. filicoides	-	155,639	85,864	18,435	25,670	36.3%	16,326,936	288,549	1.77	222.744 ± 250.043	PQ676468
Eld. fruticulosa	-	155,075	85,353	18,424	25,649	36.3%	17,535,794	137,491	0.78	83.2957 ± 278.196	PQ676467
Procris											
P. frutescens	-	153,727	84,764	18,079	25,442	36.5%	14,922,106	284,946	1.91	248.749 ± 78.6042	PQ676469
P. montana	-	154,077	84,672	18,207	25,599	36.6%	32,779,209	1,599,954	4.88	1,351.28 ± 728.285	PQ676470

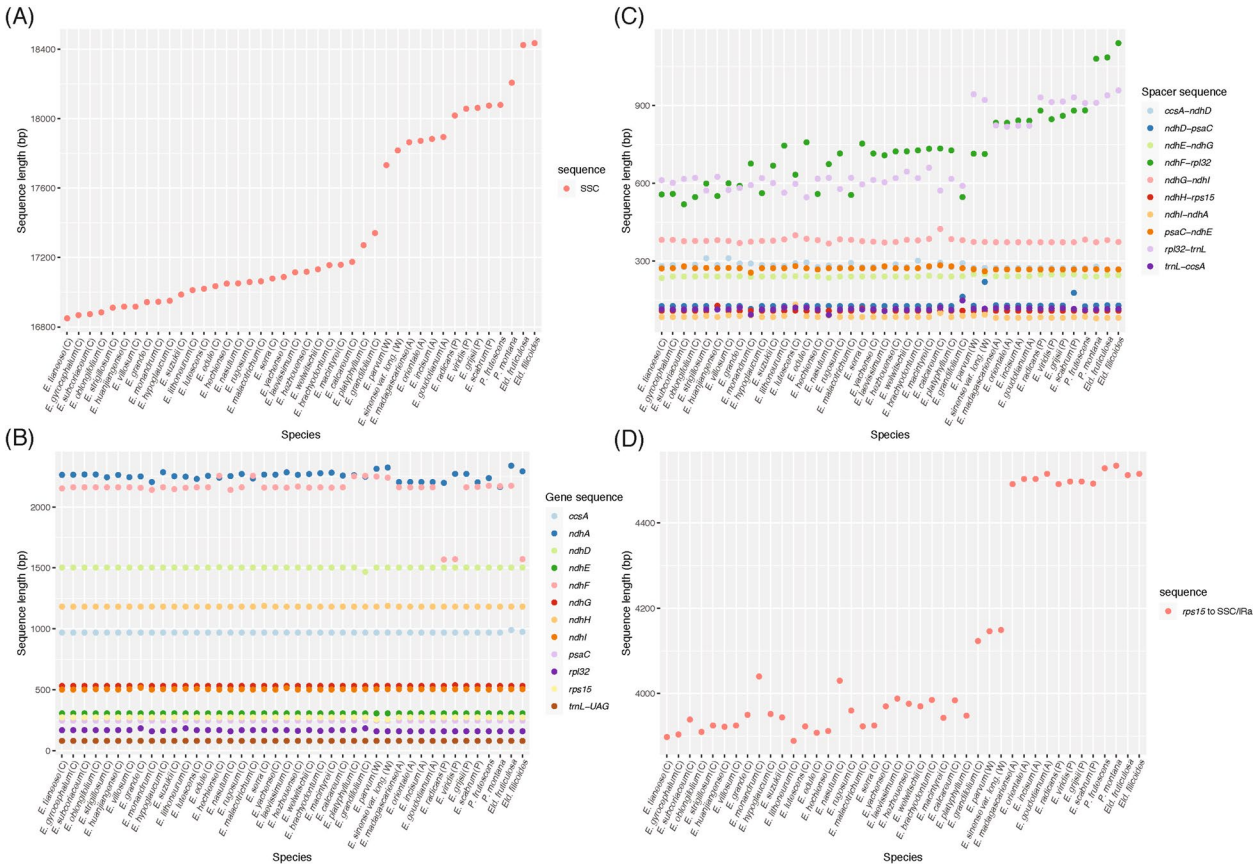

Fig. 2 Sequence length variation in small single copy (SSC) regions of newly generated plastomes from *Elatostema*, *Elatostematoides* and *Procris*. To explore the correlation between SSC length and phylogenetic relatedness, the SSC regions are analyzed as (**A**) whole sequence, (**B**) gene regions, (**C**) intergenic spacers and (**D**) *rps15* to SSC/IRa region. The letters in parenthesis following the species names correspond to the infrageneric classification of *Elatostema*: African *Elatostema* (A), core *Elatostema* (C), *Pellionia* (P) and *Wedellia* (W) clades

I, IRb/SSC boundary were within the *ndhF* gene, with a 93 bp–98 bp *ndhF* in IRb overlapping with *ycf1*. The SSC/IRa boundary was in the *ycf1* within *ndhF*, extending into IRa (Fig. 3A). Type II was characterized by the absence of *ndhF* extension into the IRb due to its reduced sequence length and variations (Fig. 3B). Among the 38 newly assembled plastomes of *Elatostema*, 30 species were classified as Type I and eight as Type II. Type I and Type II represented *Elatostematoides* species (Fig. 3A, B), while *Procris* species were exclusively Type I (Fig. 3A).

SSR and long repeat analyses of *Elatostema* and related genera plastomes

The number and type of SSRs are shown in Fig. 4. SSRs per species ranged from 68 in *E. orientale* to 107 in *E. lutescens* of *Elatostema*, from 86 in *Eld. fruticulosa* to 89 in *Eld. filicoides* of *Elatostematoides* and from 68 in *P. frutescens* to 83 in *P. crenata* of *Procris*. Mononucleotide repeats were the most prevalent SSR type across all three genera, with 2,954 identified in *Elatostema*, followed by bi-(600), tetra-(126), tri-(89), penta-(67) and hexanucleotide repeats (16). This pattern aligned with the distribution observed in *Elatostematoides* (65, 29, 11, 6, 4, 0) (Fig. 4). *Procris* exhibited a similar trend, except for a higher proportion of pentanucleotide repeats (8) than trinucleotide repeats (3). The repeat number in *Procris* was 122, 34, 9, 8, 3, 2, respectively. The mononucleotide repeat unit (A/T) was the most abundant SSRs across the three genera. In addition, several SSRs were clade-specific within *Elatostema* or each genus. For example, the AATG/ATTC repeat was also found in *Pellionia* clade of *Elatostema*, AAGT/ACTT only in *Elatostematoides*, the AAACT/AGTTT repeats in *Elatostematoides* and African *Elatostema* clade in *Elatostema*, and the AGAT/ATCT repeat in all *Elatostema* clades except the core *Elatostema* clade as well as *Elatostematoides* and *Procris*.

The total number of long repeats in each *Elatostema* plastome varied from 24 in *E. dissectum* to 187 in *E. sinense* var. *longicornutum* (Fig. 5). Except for *E. grande*, *E. hezhouense*, *E. lithoneurum* and *E. sinense* var.

Table 2 List of genes identified in *Elatostema*, *Elatostematoides* and *Procris* plastomes

Gene category	Gene names
Photosystem I	*psaA*, *psaB*, *psaC*, *psaI*, *psaJ*, *ycf3*
Photosystem II	*psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbH*, *psbI*, *psbJ*, *psbK*, *psbL*, *psbM*, *psbN*, *psbT*, *psbZ*
Cytochrome b/f complex	*petA*, *petB*, *petD*, *petG*, *petL*, *petN*
ATP synthase	*accD*, *atpA*, *atpB*, *atpE*, *atpF*, *atpH*, *atpI*
NADH dehydrogenase	*ndhA*, *ndhB*, *ndhC*, *ndhD*, *ndhE*, *ndhF*, *ndhG*, *ndhH*, *ndhI*, *ndhJ*, *ndhK*
Rubisco	*rbcL*
Ribosomal RNA genes	*rrn4.5*, *rrn5*, *rrn16*, *rrn23*
Transfer RNA genes	*trnA-UGC*, *trnC-GCA*, *trnD-GUC*, *trnE-UUC*, *trnF-GAA*, *trnfM-CAU*, *trnG-GCC*, *trnG-UCC* , *trnH-GUG*, *trnI-CAU*, *trnI-GAU*, *trnK-UUU*, *trnL-CAA*, *trnL-UAA*, *trnL-UAG*, *trnM-CAU* , *trnN-GUU*, *trnP-UGG*, *trnQ-UUG*, *trnR-ACG*, *trnR-UCU*, *trnS-GCU*, *trnS-GGA*, *trnS-UGA*, *trnT-GGU*, *trnT-UGU*, *trnV-GAC*, *trnV-UAC*, *trnW-CCA*, *trnY-GUA*
Small ribosomal subunit	*rps2*, *rps3*, *rps4*, *rps7*, *rps7*, *rps8*, *rps11*, *rps12*, *rps14*, *rps15*, *rps16*, *rps18*, *rps19*
Large ribosomal subunit	*rpl2*, *rpl14*, *rpl16*, *rpl20*, *rpl22*, *rpl23*, *rpl32*, *rpl33*, *rpl36*
RNA polymerase	*rpoA*, *rpoB*, *rpoC1*, *rpoC2*
Maturase, Protease	*matK*, *clpP*
Envelope membrane protein	*cemA*
Subunit of acetyl-CoA-carboxlyase	*accD*
Cytochrome c biogenesis protein	*ccsA*
Hypothetical chloroplast reading frames	*ycf1*, *ycf2*, *ycf4*, *ycf15*

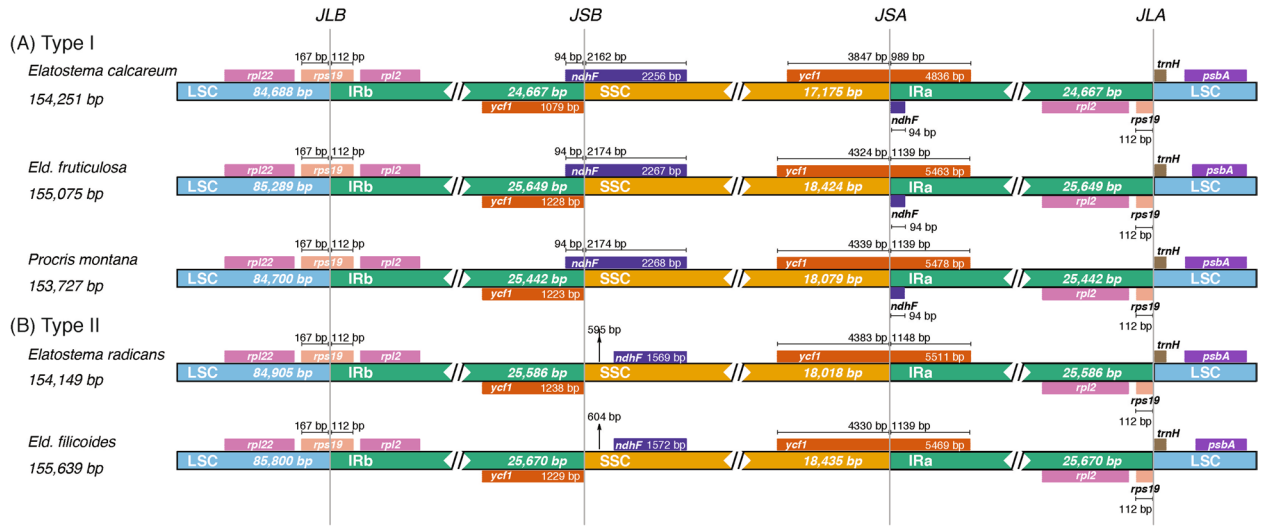

Fig. 3 Comparison of the borders of large single-copy (LSC), small single-copy (SSC), and inverted repeat (IR) regions among representative species of *Elatostema*, *Elatostematoides* and *Procris* species. (**A**) Type I: *E. calcareum*, *Eld. fruticulosa* and *Procris montana*, (**B**) Type II: *E. radicans* and *Eld. filicoides*

longicornutum with relatively abundant long repeats (93 to 193), more species have a close number of long repeats (23–55). In *Elatostematoides* and *Procris*, the total number of long repeats ranged from 32 in *Eld. fruticulose* to 41 in *Eld. filicoides* and from 23 in *P. frutescens* to 40 in *P. montana*. The majority of repeats in *Elatostema* species fell within the 30–39 bp length range (average number: 34.3), followed by the range class 40–49 bp (4.0), 50–59 bp (1.6), ≥ 70 bp (0.6), and 60–69 bp (0.3) (Fig. 5A). *Elatostematoides* and *Procris* species lacked repeats in the ≥ 70 bp and 60–69 bp range; the 30–39 bp range was the most abundant in these two genera (Fig. 5A). Regarding repeat types, forward repeats were the most prevalent in *Elatostema* plastomes (average number: 18.4), followed by palindromic (15.5), reverse (4.5), and complement repeats (3.7) (Fig. 5B). Similar pattern was also found in

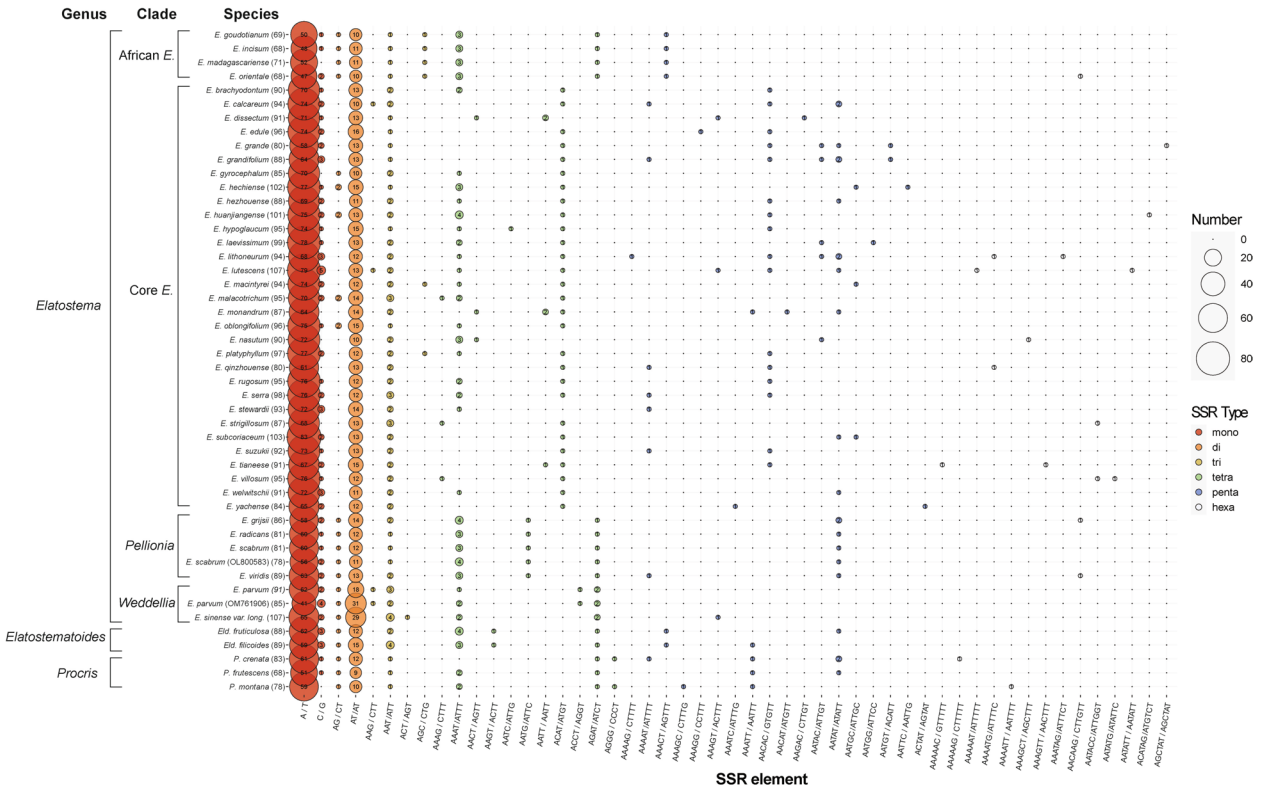

Fig. 4 Balloon plot showing the distribution of simple sequence repeat (SSR) patterns across *Elatostema*, *Elatostematoides*, and *Procris* plastomes (excluding one IR region). The plot illustrates the detailed counts of each SSR repeat unit pattern for each plastome. The total number of SSRs identified in each plastome is indicated in parentheses following the species name. Abbreviated clade names: African E.: African *Elatostema* clade; core E.: core *Elatostema* clade

Elatostematoides and *Procris*, forward repeats were the most abundant type in (21.5 and 14, respectively), by palindromic (11/7), reverse (2.5/3), and complement (1.5/1) repeats.

Sequence divergence analyses

The sliding window analysis of the 43 *Elatostema* plastomes revealed a nucleotide diversity (*pi*) ranging from 0.00008 to 0.17431 (Fig. 6). Severn regions (*ccsA-ndhD*, *ndhF*, *rpl32-trnL*, *rpoB-petN*, *rps16-trnQ*, *trnK-rps16* and *ycf1*) were identified as variable regions based on the windows with 95th percentile *pi* values, which were determined to be 0.0393 in this study. Among these regions, the *ycf1* gene exhibited the highest proportion of variable and parsimony-informative sites (31.5% and 25.3%), followed by the *ccsA-ndhD* spacer (29.3% and 23.6%) and the *trnK-rps16* spacer (21.6% and 13.2%) (see Table 3).

Phylogenetic analyses

The total length of the plastome alignment, excluding one IR, was 185,451 bp, of which 59,279 bp were variable sites (32%) and 39,204 bp were parsimony-informative sites (21.1%). The substitution models and best-fit partition schemes estimated by ModelFinder for IQ-TREE [48] and MrBayes [49] are summarized in Additional File 3. Phylogenetic analyses based on the partitioned dataset and CDS-only dataset, inferred using ML and BI methods, showed identical and highly supported topologies (Fig. 7, Additional File 4). Urticaceae was confirmed as a monophyletic group (ultrafast bootstrap value (UFBoot) = 100; posterior probability (PP) = 1 in both datasets) (Fig. 7, Additional File 4). Consistent with the findings of Wu et al. [31], Urticaceae was divided into four major clades, Clade I, II, III, and IV (designated in Wu et al. [31]). Clades I and IV (UFBoot = 100, PP = 1 in both datasets), as well as Clades II and III (UFBoot = 100, PP = 1 in both datasets), formed sister groups, respectively. The genera *Elatostema*, *Elatostematoides* and *Procris* formed a strongly supported monophyletic group (UFBoot = 100, PP = 1 in both datasets), with *Procris* sister to the remaining taxa. Three well-supported clades were identified within this clade, corresponding to genera delimitation previously proposed [3] (UFBoot = 100, PP = 1 in both datasets). *Elatostema* was further divided into four major clades, African *Elatostema*, core *Elatostema*, *Pellionia* and *Weddellia* clades, as previously

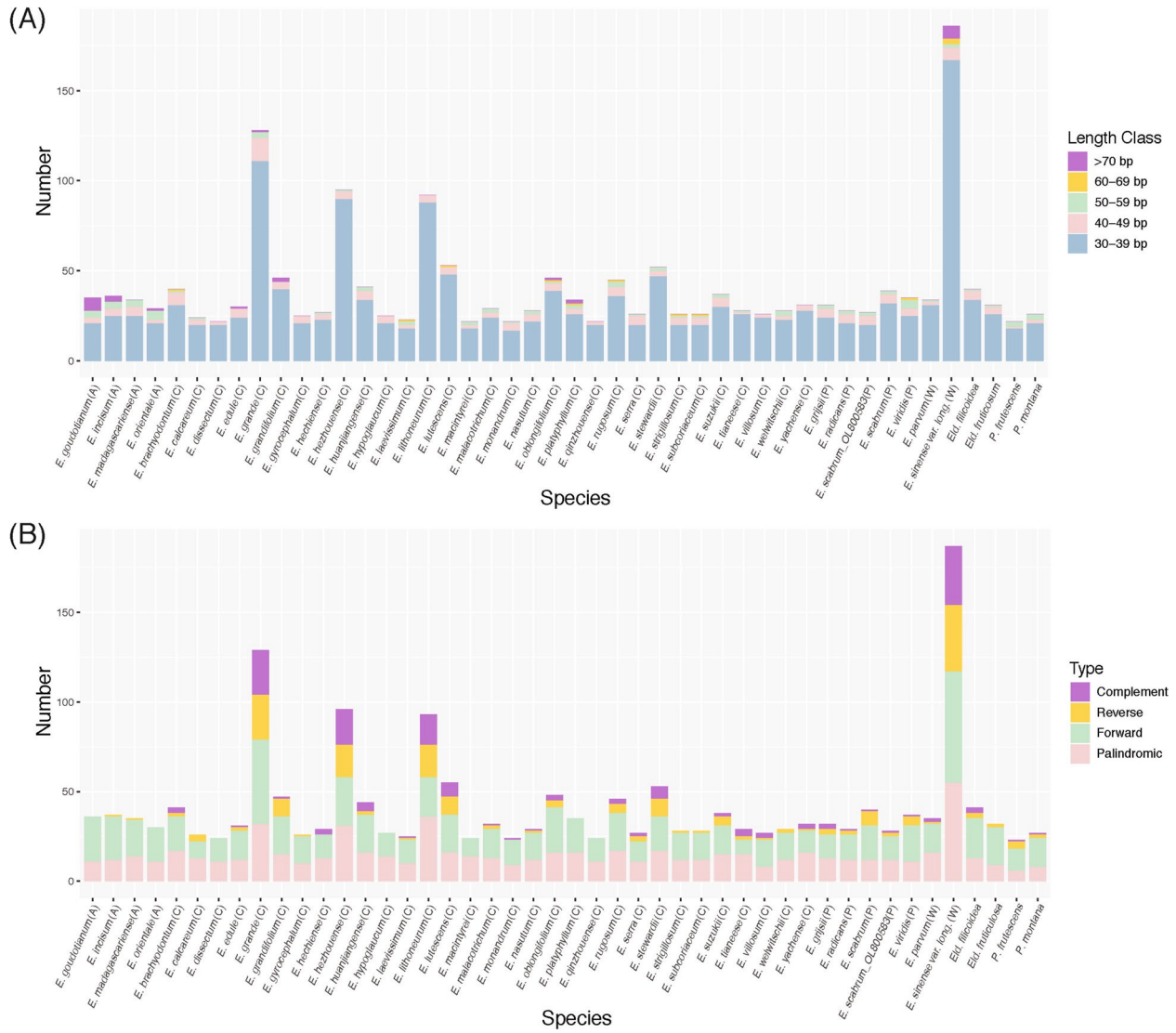

Fig. 5 Statistics of long repeat content in *Elatostema*, *Elatostematoides* and *Procris* plastomes (excluding one IR region). (**A**) Length distribution of long repeats categorized into five length classes. (**B**) The number of four types of long repeats identified in each plastome. The letters in parenthesis following the species names correspond to the infrageneric classification of *Elatostema*: African *Elatostema* (A), core *Elatostema* (C), *Pellionia* (P) and *Wedellia* (W) clades

defined by Tseng et al., [3]. These clades were also consistently resolved with full support values in ML and BI analyses (UFBoot = 100, PP = 1 in both datasets).

Discussion

Plastome conservation in *Elatostema*, *Elatostematoides*, and *Procris*

This study presents the first comprehensive plastome analysis of *Elatostema*, *Elatostematoides*, and *Procris*, comprising 42 newly assembled plastomes. Our findings reveal a high degree of conservation in plastome structure, genome size, and gene content within these genera. Comparative analysis with other Urticaceae species demonstrates that the sequence length of *Elatostema*, *Elatostematoides*, and *Procris* plastomes (149,152–164,019 bp) falls within the range observed in *Boehmeria* (142,627–170,958 bp) [28], *Debregeasia* (155,743–156,065 bp) [26], *Oreocnide* (156,663– 157,464 bp) [29], *Pilea* (150,398–152,327 bp) [27], and the tribe Urticeae (145,419–161,930 bp) [30]. Moreover, the range of plastome sizes in these three genera is remarkably close to the median (154,953 bp) of angiosperms [51]. We further observed that the GC content of *Elatostema*, *Elatostematoides*, and *Procris* (36.2% to 36.7%) is consistent with the

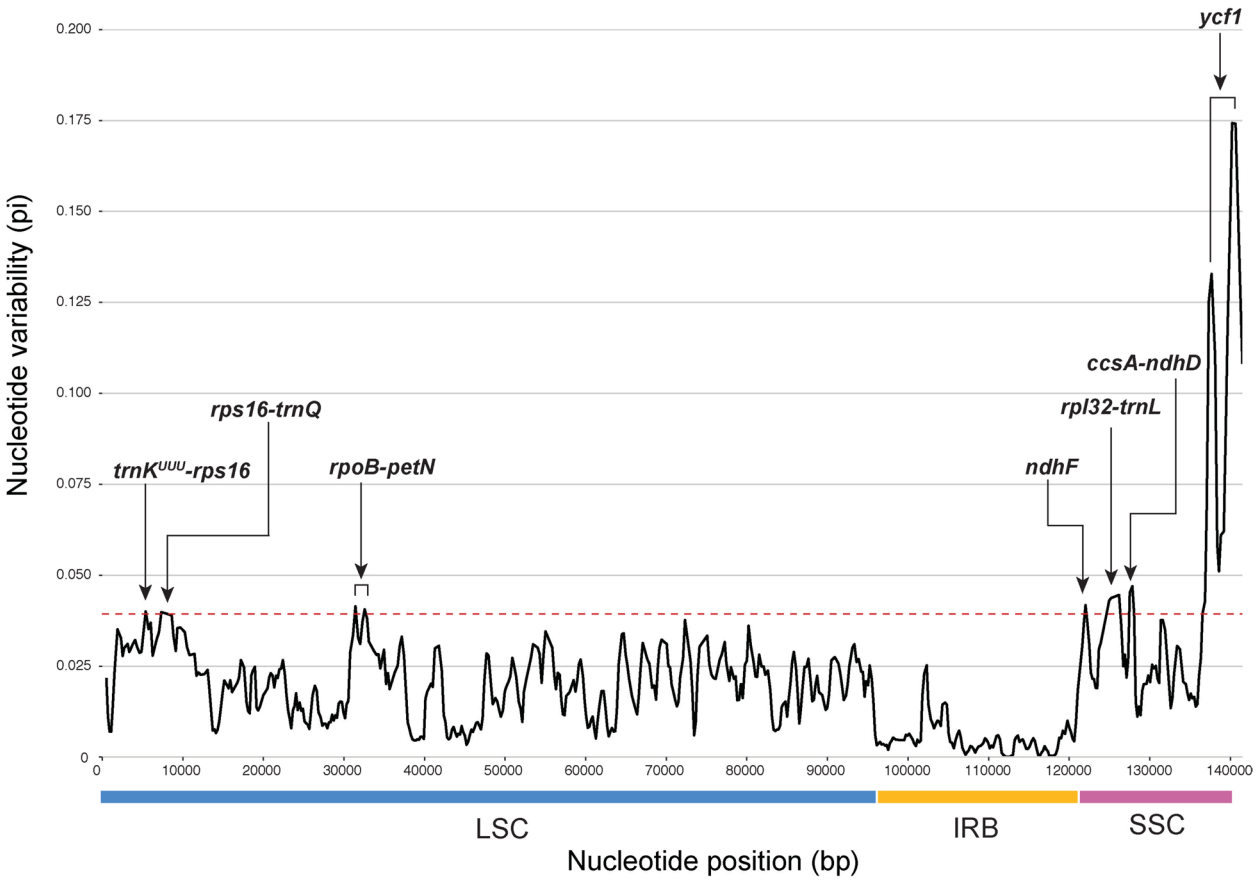

Fig. 6 Sliding-window analysis of plastomes from 39 newly generated and 5 NCBI-available *Elatostema* species (excluding one IR region). The red dashed line indicates the nucleotide variability (*pi*) value threshold (*pi* = 0.0393), representing the top 5% of all nucleotide diversity values in the dataset. Regions with pi values higher than the red dashed line were marked as variable regions

Table 3 Statistics of potential DNA barcode sequence in *Elatostema*

Marker	Length range (bp)	GC content (%)	Number of variable site (bp)	Proportion of variable site (%)	Parsimony-informative site (bp)	Proportion of parsimony-informative site (%)
ccsA-ndhD	273–311	27.8	103	29.3	83	23.6
ndhF	2170–2262	30.8	370	16.3	242	10.6
rpl32-trnL	714–1102	22.8	266	19.1	166	11.9
rpoB-petN	1992–2117	29.0	462	18.4	297	11.8
rps16-trnQ	927–988	33.7	199	18.3	123	11.3
trnK-rps16	671–802	22.9	229	21.6	140	13.2
ycf1	3753–4314	26.1	1638	31.5	1317	25.3

(See figure on next page.)

Fig. 7 Phylogenomic tree of Urticaceae with three Moraceae species as outgroups, inferred from the coding sequences (CDS) of plastome sequences reconstructed by Maximum Likelihood analysis. The tree was constructed with IQ-TREE. Node with full support (ultrafast bootstrap values of 100 and Bayesian inference posterior probabilities of 1) are unlabeled. Ultrafast bootstrap values below 95 and posterior probabilities below than 1 are shown. Major clades of Urticaceae and *Elatostema* are annotated following the classifications proposed by Wu et al., [31] and Tseng et al., [3], respectively

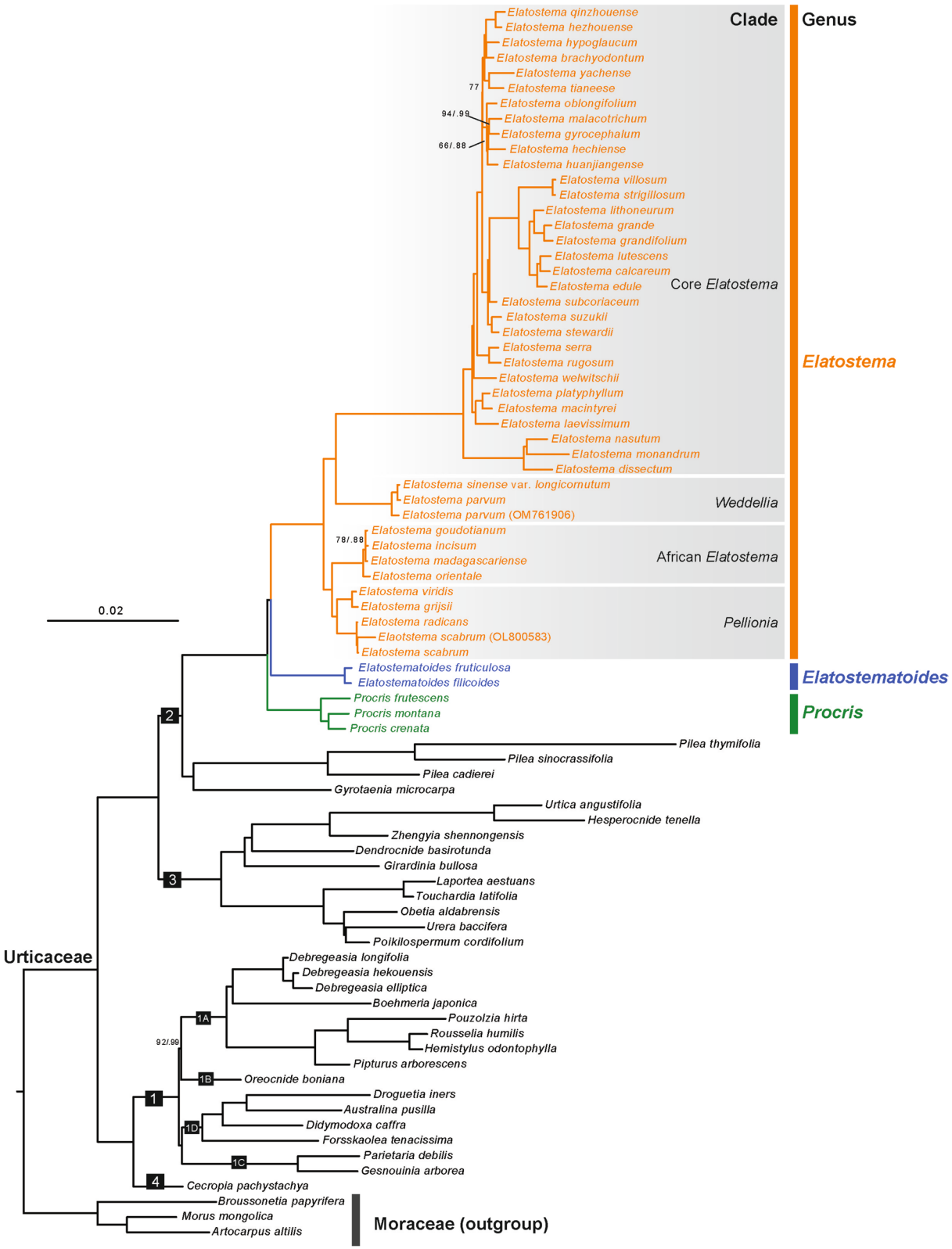

Fig. 7 (See legend on previous page.)

GC content range reported for other Urticaceae plastomes, including *Boehmeria* (35.32–36.41%), *Debregeasia* (36.2–36.9%), *Oreocnide* (36.2%), *Pilea* (36.35–36.69%), and the tribe Urticeae (36.3–37.2%) [30]. Furthermore, the gene number in *Elatostema*, *Elatostematoides* and *Procris* is also congruent with other species in Urticaceae [26–29] and identical to the gene number of *Pilea* (79 unique protein-coding genes, 30 tRNA genes, and four rRNA genes), a genus also belonging to Clade II of Urticaceae [31]. In summary, these results suggest that plastome structure, genome size, and gene content are conserved within *Elatostema*, *Elatostematoides*, and *Procris* and exhibit remarkable consistency compared to other Urticaceae species. Besides, we observed significant variations in sequence length across several genes in *Elatostema*, including *matK*, *ndhF*, *rps3*, *ycf1*, and *ycf2*. Some of these genes lack complete ORFs compared to their typical counterparts. Further validation, such as RNA-seq or qRT-PCR, is required to determine whether these truncated genes remain functional.

The plastomes of *Elatostema*, *Elatostematoides*, and *Procris* exhibit a high degree of conservation in their LSC/IRb and IRa/LSC boundaries, which are consistently located within the *rps19* gene and between *rps19* and *trnH*, respectively. These boundary configurations are comparable to those observed in other Urticaceae genera, *Debregeasia* [28], *Pilea* [27], *Urera* and *Zhengyia* [30], and in most non-monocot angiosperms [52]. The IRb/SSC and SSC/IRa boundaries in *Elatostema* are also highly conserved, with Type I being the predominant configuration, characterized by the overlap of the *ycf1* gene with *ndhF* and as found in *Pilea* [27]. Type II is correlated with the sequence length of *ndhF* and *ycf1*. While no strong correlation was observed between boundary type and phylogenetic grouping, all species of the *Weddellia* clade exclusively exhibited Type II. These findings suggest that *Elatostema* lineages have not undergone significant expansion or contraction of its IR regions. More comprehensive studies are required to investigate the potential impact of *ycf1* and *ndhF* overlap on gene expression and to explore the evolutionary implications of the observed boundary variations.

ycf15 loss in a distinct clade

Comparative plastome analyses of *Elatostema* reveal a distinct clade comprising five species (*E. huanjiangense*, *E. hechiense*, *E. huanjiangense*, *E. malacotrichum* and *E. oblongifolium*) characterized by a unique pattern of loss of *ycf15*. These species, forming a monophyletic group within the core *Elatostema* clade, exhibit larger plastome size than other *Elatostema* species, primarily due to an approximately 7,000 bp insertion within the IR region. This insertion further results in the complete loss of the *ycf15* gene which occurred before the divergence of the five-species clade. The functional significance of *ycf15* remains enigmatic and controversial [53]. Transcriptomic analyses of *Camellia* have indicated that *ycf15* is transcribed as a part of precursor polycistronic transcript encompassing *ycf2*, *ycf15* and antisense *trnL*-CAA [54]. While *ycf15* has been retained and potentially functional in some taxa, such as *Hydrocotyle* [55] and *Camellia* [54], it has been lost or pseudogenized in others, such as *Thottea* and *Verhuellia* [56], and *Tillandsia* [57]. The complete loss of *ycf15* has not been previously reported in the Urticaceae. Further studies are necessary to elucidate the functional roles of *ycf15* in this family, and the correlation between these genetic modifications and the unique characteristics of the five-species clade.

SSC length variations and their phylogenetic implications in *Elatostema*

The SSC region typically ranges from 16 to 27 kb in most land plants [58]. Its size can be significantly influenced by IR expansion, leading to a reduction or even the complete loss of the small SSC region, consequently altering the overall plastome size [59, 60]. The SSC length is generally conservative within the family and genus but can exhibit interspecific variation. For example, in Arecaceae [61], SSC length varies across species between 13,768 and 18,380 bp, but no significant differences have been detected across subfamilies. A similar pattern is observed in a hemiparasitic species, *Pedicularis ishidoyanum*, where an extensive IR expansion results in a drastically reduced SSC region [59]. This phenomenon, however, is absent in other *Pedicularis* species, such as *P. resupinatum*, *P. spicatum*, and *P. verticillatum* [62–64], suggesting independent IR expansion events across different lineages.

To our knowledge, no studies have demonstrated a correlation between SSC length and phylogenetic relatedness within and between genera. This study presents the first evidence of such a correlation in *Elatostema*, especially within infrageneric and intergeneric classification of *Elatostema.* This correlation mainly arises from the variations in specific fragments, including the *ndhF-rpl32* spacer, the *rpl32-trnL* spacer and the region of *rps15*-SSC/IRa boundary. While the SSC region exhibits limited variations in other Urticaceae genera [26–28], it does not correlate with subclade divisions, such as in *Boehmeria s.l.* [28]. To further validate the taxonomic utility of these regions, more comprehensive samplings of *Elatostema* and related genera is necessary.

Repeat variations and their taxonomic and evolutionary significance in *Elatostema* plastomes

Simple sequence repeats (SSRs) are ubiquitous in plastomes and serve as valuable tools for investigating evolutionary processes, population genetics, and genome polymorphism [65, 66]. Most mutations detected in spontaneous plastome mutants have been associated with repetitive DNA sequences [67], highlighting the crucial role of tandem repeats in driving plastome variation among closely related species [68].

In this study, we identified several SSR elements that show clade-specific patterns, suggesting their potential utility in taxonomic and evolutionary studies within *Elatostema*, *Elatostematoides*, and *Procris*. For example, AATG/ATTC found only in the *Pellionia* clade; AAGT/ACTT only in *Elatostematoides*; AAACT/AGTTT present in *Elatostematoides* and the African *Elatostema* clade; and the absence of the AGAT/ATCT element in the core *Elatostema* clade. These group-specific SSRs highlight their potential as lineage-diagnostic markers. Given the observed phylogenetic specificity of these SSRs and the limited molecular resources available for these taxa, the repeats identified here represent promising molecular markers for species delimitation, population genetics, and biogeographical studies. Further research is warranted to explore the underlying mechanisms by which these SSRs contribute to plastome evolution and to assess their broader applicability in taxonomic resolution across Urticaceae.

Elatostema species typically exhibit long repeat ranging from 22 to 55 per species. However, four specific species— *E. lithoneurum* (93), *E. hezhouense* (96), *E. grande* (129) and *E. sinense* var. *longicornutum* (187)—display a significantly higher number of long repeats, particularly within the 30–39 bp range. For example, approximately 90% (89.8%, 168 out of 187) of long repeats in *E. sinense* var. *longicornutum* fall within this range, with 124 being 30 or 31 bp long. Previous studies have indicated a positive correlation between dispersed repeats and genomic rearrangements [69–71], suggesting that long repeats (dispersed repeats) can drive of plastome rearrangement in land plants. However, contrasting findings have been reported in *Begonia* [72] and *Daucus* [73], where no such correlation was observed. Our analysis of *Elatostema* plastomes reveals a high degree of conservation in gene order and content, and no rearrangements are detected in four species with relatively more long repeats. This suggests there may not be a direct correlation between the abundance and type of repeats and the propensity for genomic rearrangements in *Elatostema*.

Identification of highly variable regions as DNA barcodes

Despite significant advancements in high-throughput sequencing technologies, the generation and analysis of large-scale plastome data for species-rich genera, such as *Elatostema* or extensive population sampling, remains a resource-intensive task. To address this challenge and facilitate efficient species identification, taxonomic revisions, inter- and intraspecific studies, and evolution studies, the development of phylogenetically informative markers is crucial [74, 75]. In this study, we identified seven highly variable regions (*ccsA-ndhD*, *ndhF*, *rpl32-trnL*, *rpoB-petN*, *rps16-trnQ*, *trnK-rps16*, and *ycf1*) as potential DNA barcodes for *Elatostema*. This finding is similar to previous studies in Urticaceae that have highlighted the *ycf1* gene as a highly variable locus with significant taxonomic utility [26, 27, 30]. Moreover, a study by Dong et al. [76] further supports this view, recommending *ycf1* as a suitable plastid barcode for land plants, corroborating our results.

Phylogenomic insights and taxonomic refinement in *Elatostema* and allied genera

The relationship between the four major clades and genera phylogenetic relatedness within the clade is also congruent with the classification of Wu et al. [31]. The full support values of these major clades in this study suggest that plastome sequence could be the significant taxonomic utility in the Urticaceae family. Our phylogenomic analysis of the plastome further supports the proposed taxonomic delimitation of *Elatostema s.l.* into three distinct genera: *Elatostema*, *Procris*, and *Elatostematoides*, as suggested by Tseng et al. [3]. Consistent with our findings and Tseng et al. [3], *Elatostematoides* is a sister to *Elatostema*, while *Procris* is sister to the remaining taxa. In addition, the plastome phylogenetic tree supports the infrageneric classification of *Elatostema* into four major clades (African *Elatostema*, core *Elatostema*, *Pellionia*, *Weddellia*). However, a conflict arises between our current plastome analysis and the results of Tseng et al. [3]: the basal lineage of *Elatostema* is identified as a clade comprising *Pellionia* and African *Elatostema*, rather than the *Weddellia* clade. Tseng et al. [3] utilized a combined dataset of one nuclear (nrITS) and two plastid markers (*psbA-trnH* and *psbM-trnD*), with nrITS contributing over 60% of the parsimony-informative sites. Although no significant conflict was detected between the nuclear and plastid trees (only considering the branches with bootstrap support > 80) [3], the lower bootstrap support values for the relationships among major clades in the plastid trees [3] suggest a potential reliance on the nuclear signal. To further elucidate these phylogenetic relationships and resolve the observed incongruence,

a more comprehensive sampling strategy is necessary, includes increasing the number of nuclear markers for *Elatostema* and expanding the plastome samplings for African *Elatostema*, *Pellionia* and *Weddellia* clades.

Elatostema comprises over 500 species [3], exhibiting a notable difference in diversification compared to its sister genera, *Procris* and *Elatostematoides*, despite their similar ecological preferences. The evolutionary mechanisms underlying this disparity in species richness remain unclear. Previous phylogenetic studies [3, 5] based on a limited set of molecular markers have yielded low-resolution topologies, with weak bootstrap support for key clades—particularly within the core *Elatostema* group—due to insufficient phylogenetic signal. In this study, we utilize complete plastome sequences to reconstruct a well-resolved phylogeny, providing a robust framework for future investigations into the speciation processes, biogeographical patterns, and evolutionary history of *Elatostema*.

Conclusion

This study presents the first comprehensive plastome analysis of *Elatostema*, *Elatostematoides*, and *Procris*, involving 42 newly assembled plastomes. Our findings demonstrate a high degree of conservation in plastome structure, genome size, and gene content within these genera. The plastome sizes and GC content of *Elatostema*, *Elatostematoides*, and *Procris* are consistent with the range observed in other Urticaceae species. The study reveals a significant correlation between the length of the small single-copy (SSC) region and phylogenetic relationships within *Elatostema* and between *Elatostema* and related genera. Furthermore, seven highly variable regions (*ccsA-ndhD*, *ndhF*, *rpl32-trnL*, *rpoB-petN*, *rps16-trnQ*, *trnK-rps16*, and *ycf1*) were identified as potential DNA barcodes for *Elatostema*. Among these regions, the *ycf1* gene exhibits the highest proportion of variable and parsimony-informative sites. Phylogenomic analyses further support the delimitation of *Elatostema s.l.* into three distinct genera: *Elatostema*, *Procris*, and *Elatostematoides*. Within *Elatostema*, our results reconfirm four major clades of this genus: African *Elatostema*, core *Elatostema*, *Pellionia*, and *Weddellia*. In conclusion, this study provides a highly resolved phylogenetic framework based on plastome sequences, offering valuable insights into the speciation, biogeography, and evolutionary history of *Elatostema*.

Abbreviations

BI	Bayesian inference
UFBoot	Ultrafast bootstrap value
E	*Elatostema*
Eld	*Elatostematoides*
IR	Inverted repeat
LSC	Larger single copy
ML	Maximum likelihood
PP	Posterior probability
SSC	Small single copy

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-025-06569-9.

Additional file 1. Species were used in this study.

Additional file 2. Sequence variations of *matK*, *ndhF*, *rps3*, *ycf1* and *ycf2* in *Elatostema*.

Additonal file 3. The nucleotide substitution models evaluated by ModelFinder using the Bayesian information criterion (BIC) for IQ-TREE and MrBayes analyses in two datasets.

Additional file 4. Phylogenomic tree of Urticaceae with three Moraceae species as outgroups, inferred from partitioned dataset of plastome sequences reconstructed by Maximum Likelihood analysis. The tree was constructed with IQ-TREE. Node with full support (ultrafast bootstrap values of 100 and Bayesian inference posterior probabilities of 1) are unlabeled. Ultrafast bootstrap values below 95 and posterior probabilities below than 1 are shown.

Acknowledgements

We thank the curators and staff of A, BM, IBK, KYO, MO, TAI, TAIF, and TNM herbaria for providing plant materials. We are grateful to Chia-Lun Hsieh for assisting with the fieldwork.

Authors' contributions

YHT conceptualized the study, collected the samples, conducted the experiments, and drafted the manuscript. HCC conducted the experiments. GXZ collected the samples and conducted the experiments. All authors read and approved the final manuscript.

Funding

The study was supported by the National Science and Technology Council, Taiwan to YHT (MOST 111–2621-B-005–001-MY3).

Data availability

All plastomes generated in this study are available in NCBI (https://www.ncbi.nlm.nih.gov, with accession numbers as shown in Table 1).

Declarations

Ethics approval and consent to participate

The plant materials used in the study were collected under permission. The collection and use of plant materials comply with relevant institutional, national, and international guidelines and legislation. This article does not contain any studies with human participants or animals and does not involve any endangered or protected species.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 4 January 2025 Accepted: 16 April 2025
Published online: 25 April 2025

References

1. Wang WT. Nova classification specierum sinicarum *Elatostematis* (Urticaceae). In: Fu DZ, editor. Paper collection of W T Wang. Beijing: Higher education press; 2012. p. 1016–178.

2. Wei YG, Monro AK, Wang WT. Additions to the Flora of China: seven new species of *Elatostema* (Urticaceae) from the karst landscapes of Guangxi and Yunnan. Phytotaxa. 2011;29:1–27.
3. Tseng YH, Monro AK, Wei YG, Hu JM. Molecular phylogeny and morphology of *Elatostema* s. l. (Urticaceae): Implications for inter- and infrageneric classifications. Mol Phylogenet Evol. 2019;132:251–64.
4. Hadiah JT, Conn BJ. Usefulness of morphological characters for infrageneric classification of *Elatostema* (Urticaceae). Blumea. 2009;54:181–91.
5. Hadiah JT, Quinn CJ, Conn BJ. Phylogeny of *Elatostema* (Urticaceae) using chloroplast DNA data. Telopea. 2003;10:235–46.
6. Hadiah JT, Conn BJ, Quinn CJ. Infra-familial phylogeny of Urticaceae, using chloroplast sequence data. Aust Syst Bot. 2008;21:375–85.
7. Wang WT. Classificatio specierum sinensium *Pellionae* (Urticaceae). Bull Bot Lab N E Forest. 1980;6:45–66.
8. Wang WT. Classificatio specierum sinicarum *Elatostematis* (Urticaceae). Bull Bot Lab N E Forest. 1980;7:1–96.
9. Weddell HA: Urticaceae. In: *Prodomus Systematis naturalis regni vegetabilis.* Edited by A. DC, vol. 16. Masson, Paris; 1869. p.32–235.
10. Robinson CB. Philippine Urticaceae. Philipp J Sci. 1910;5:465–542.
11. Yang YP, Shih BL, Liu HY. A revision of *Elatostema* (Urticaceae) of Taiwan. Bot Bull Acad Sinica. 1995;36:259–79.
12. Hallier H. Neue und bemerkenswerte pflanzen aus dem malaiisch-papuanischen inselmeer. Ann Jard Bot Buitenzorg. 1896;13:276–326.
13. Winkler H. Die urticaceen Papuasiens. Beiträge zur Flora von Papuasien. 1922;8:501–608.
14. Schröter H. Monographie der Gattung *Procris* l. Feddes Repert Spec Nov Regni Veg. 1938;45:179–92.
15. Schröter H, Winkler H. Monographie der gattung *Elatostema s. l.*: Allgemeiner teil. Repert Spec Nov Regni Veg. 1935;83:1–71.
16. Schröter H, Winkler H. Monographie der gattung *Elatostema s. l.*: Spezieller teil. Repert Spec Nov Regni Veg. 1936;83:1–237.
17. 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;76:273–97.
18. Twyford AD, Ness RW. Strategies for complete plastid genome sequencing. Mol Ecol Resour. 2017;17:858–68.
19. Daniell H, Jin SX, Zhu XG, Gitzendanner MA, Soltis DE, Soltis PS. Green giant-a tiny chloroplast genome with mighty power to produce high-value proteins: history and phylogeny. Plant Biotechnol J. 2021;19:430–47.
20. Qu XJ, Zou D, Zhang RY, Stull GW, Yi TS. Progress, challenge and prospect of plant plastome annotation. Front Plant Sci. 2023;14:1166140.
21. Fu LF, Monro AK, Yang TG, Wen F, Pan B, Xin ZB, Zhang ZX, Wei YG. *Elatostema qinzhouense* (Urticaceae), a new species from limestone karst in Guangxi, China. PeerJ. 2021;9:e11148.
22. Fu LF, Zhang ZX. Complete chloroplast genome sequence of *Procris crenata* C.B.Rob (Urticaceae). Mitochondrial DNA B. 2021;6:458–9.
23. Yang XL, Yan L, Wang X, Wu YF, Hu XJ. Characterization of the complete chloroplast genome of *Pellionia scabra* (Urticaceae). Mitochondrial DNA B. 2022;7:1732–3.
24. Fu LF, Xin ZB, Wen F, Li S, Wei YG. Complete chloroplast genome sequence of *Elatostema dissectum* (Urticaceae). Mitochondrial DNA B. 2019;4:838–9.
25. Yang LH, Feng YQ, Ding L, Tong LM, Qi ZC, Yan XL. The complete chloroplast genome sequence of *Elatostema stewardii* Merr. (Urticaceae). Mitochondrial DNA B. 2022;7:1494–6.
26. Wang RN, Milne RI, Du XY, Liu J, Wu ZY. Characteristics and mutational hotspots of plastomes in *Debregeasia* (Urticaceae). Front Genet. 2020;11:729.
27. Li JL, Tang JM, Zeng SY, Han F, Yuan J, Yu J. Comparative plastid genomics of four *Pilea* (Urticaceae) species: insight into interspecific plastid genome diversity in *Pilea*. BMC Plant Biol. 2021;21:25.
28. Zhan M, Xue L, Zhou JJ, Zhang Q, Qin XM, Liao XW, Wu L, Monro AK, Fu LF. Polyphyly of *Boehmeria* (Urticaceae) congruent with plastome structural variation. Front Plant Sci. 2024;15:1297499.
29. Wu ZY, Milne RI, Liu J, Slik F, Yu Y, Luo YH, Monro AK, Wang WT, Wang H, Kessler PJA, et al. Phylogenomics and evolutionary history of *Oreocnide* (Urticaceae) shed light on recent geological and climatic events in SE Asia. Mol Phylogenet Evol. 2022;175:107555.
30. Ogoma CA, Liu J, Stull GW, Wambulwa MC, Oyebanji O, Milne RI, Monro AK, Zhao Y, Li DZ, Wu ZY. Deep insights into the plastome evolution and phylogenetic relationships of the tribe Urticeae (Family Urticaceae). Front Plant Sci. 2022;13:870949.
31. Wu ZY, Monro AK, Milne RI, Wang H, Yi TS, Liu J, Li DZ. Molecular phylogeny of the nettle family (Urticaceae) inferred from multiple loci of three genomes and extensive generic sampling. Mol Phylogenet Evol. 2013;69:814–27.
32. Li DK, Huang CL, Tian JB, Wang YK, Wang YQ. Extraction ways of high qualitiful DNA from *Z. jujuba* Mill. Mol Plant Breed. 2005;3:579–83.
33. Andrews, S. FastQC: A quality control tool for high throughput sequence data [online]. Available online at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/. 2010.
34. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20.
35. Jin JJ, Yu WB, Yang JB, Song Y, dePamphilis CW, Yi TS, Li DZ. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol. 2020;21:241.
36. 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:W6–11.
37. Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res. 2019;47:W59–64.
38. Diez Menendez C, Poczai P, Williams B, Myllys L, Amiryousefi A. IRplus: An augmented tool to detect inverted repeats in plastid genomes. Genome Biol Evol. 2023;15:evad177.
39. 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;34:3299–302.
40. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772–80.
41. Maddison WP, Maddison DR: Mesquite: a modular system for evolutionary analysis., version 3.03 ed. 2015.
42. Borowiec ML. AMAS: a fast tool for alignment manipulation and computing of summary statistics. PeerJ. 2016;4:1660.
43. Fox J, Weisberg S. An R companion to applied regression. 2nd ed. Thousand Oaks, CA: Sage; 2011.
44. Beier S, Thiel T, Munch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. Bioinformatics. 2017;33:2583–5.
45. Wickham H. Reshaping data with the reshape package. J Stat Softw. 2007;21:1–20.
46. Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag; 2016.
47. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res. 2001;29:4633–42.
48. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ, Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020;37(5):1530–4.
49. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna 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;61:539–42.
50. Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. Systematic Biol. 2018;67:901–4.
51. Park S, An B, Park S. Reconfiguration of the plastid genome in *Lamprocapnos spectabilis*: IR boundary shifting, inversion, and intraspecific variation. Sci Rep. 2018;8:13568.
52. Wang RJ, Cheng CL, Chang CC, Wu CL, Su TM, Chaw SM. Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. BMC Evol Biol. 2008;8:36.
53. Steane DA. Complete nucleotide sequence of the chloroplast genome from the Tasmanian blue gum, *Eucalyptus globulus* (Myrtaceae). DNA Res. 2005;12:215–20.
54. Shi C, Liu Y, Huang H, Xia EH, Zhang HB, Gao LZ. Contradiction between plastid gene transcription and function due to complex posttranscriptional splicing: an exemplary study of *ycf15* function and evolution in angiosperms. PLoS ONE. 2013;8:e59620.

55. Wen J, Wu BC, Li HM, Zhou W, Song CF. Plastome structure and phylogenetic relationships of genus *Hydrocotyle* (Apiales): provide insights into the plastome evolution of *Hydrocotyle*. BMC Plant Biol. 2024;24:778.
56. Jost M, Wanke S. A comparative analysis of plastome evolution in autotrophic Piperales. Am J Bot. 2024;111:e16300.
57. Vera-Paz SI, Diaz DDDC, Jost M, Wanke S, Rossado AJ, Hernandez-Gutierrez R, Salazar GA, Magallon S, Gouda EJ, Ramirez-Morillo IM, et al. New plastome structural rearrangements discovered in core Tillandsioideae (Bromeliaceae) support recently adopted taxonomy. Front Plant Sci. 2022;13:924922.
58. Chumley TW, Palmer JD, Mower JP, Fourcade HM, Calie PJ, Boore JL, Jansen RK. The complete chloroplast genome sequence of *Pelargonium* x *hortorum*: Organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. Mol Biol Evol. 2006;23:2175–90.
59. Cho WB, Choi BH, Kim JH, Lee DH, Lee JH. Complete plastome sequencing reveals an extremely diminished SSC region in hemiparasitic *Pedicularis ishidoyana* (Orobanchaceae). Ann Bot Fenn. 2018;55:171–83.
60. Miao YJ, Chen HM, Xu WQ, Yang QQ, Liu C, Huang LF. Structural mutations of small single copy (SSC) region in the plastid genomes of five *Cistanche* species and inter-species identification. BMC Plant Biol. 2022;22:412.
61. Chen DJ, Landis JB, Wang HX, Sun QH, Wang Q, Wang HF. Plastome structure, phylogenomic analyses and molecular dating of Arecaceae. Front Plant Sci. 2022;13:960588.
62. Robart BW, Gladys C, Frank T, Kilpatrick S. Phylogeny and biogeography of North American and Asian *Pedicularis* (Orobanchaceae). Syst Bot. 2015;40:229–58.
63. Tkach N, Ree RH, Kuss P, Röser M, Hoffmann MH. High mountain origin, phylogenetics, evolution, and niche conservatism of arctic lineages in the hemiparasitic genus *Pedicularis* (Orobanchaceae). Mol Phylogenet Evol. 2014;76:75–92.
64. Ree RH. Phylogeny and the evolution of floral diversity in *Pedicularis* (Orobanchaceae). Int J Plant Sci. 2005;166:595–613.
65. Ebert D, Peakall R. Chloroplast simple sequence repeats (cpSSRs): technical resources and recommendations for expanding cpSSR discovery and applications to a wide array of plant species. Mol Ecol Resour. 2009;9:673–90.
66. Pauwels M, Vekemans X, Godé C, Frérot H, Castric V, Saumitou-Laprade P. Nuclear and chloroplast DNA phylogeography reveals vicariance among European populations of the model species for the study of metal tolerance, *Arabidopsis halleri* (Brassicaceae). New Phytol. 2012;193:916–28.
67. Massouh A, Schubert J, Yaneva-Roder L, Ulbricht-Jones ES, Zupok A, Johnson MTJ, Wright SI, Pellizzer T, Sobanski J, Bock R, et al. Spontaneous chloroplast mutants mostly occur by replication slippage and show a biased pattern in the plastome of *Oenothera*. Plant Cell. 2016;28:911–29.
68. Li ZH, Ma X, Wang DY, Li YX, Wang CW, Jin XH. Evolution of plastid genomes of *Holcoglossum* (Orchidaceae) with recent radiation. BMC Evol Biol. 2019;19:63.
69. Weng ML, Blazier JC, Govindu M, Jansen RK. Reconstruction of the ancestral plastid genome in Geraniaceae reveals a correlation between genome rearrangements, repeats, and nucleotide substitution rates. Mol Biol Evol. 2014;31:645–59.
70. Lee HL, Jansen RK, Chumley TW, Kim KJ. Gene relocations within chloroplast genomes of *Jasminum* and *Menodora* (Oleaceae) are due to multiple, overlapping inversions. Mol Biol Evol. 2007;24:1161–80.
71. Guisinger MM, Kuehl JV, Boore JL, Jansen RK. Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: rearrangements, repeats, and codon usage. Mol Biol Evol. 2011;28:583–600.
72. Tseng YH, Hsieh CL, Campos-Domínguez L, Hu AQ, Chang CC, Hsu YT, Kidner CA, Hughes M, Moonlight PW, Hung CH, et al. Insights into the evolution of the chloroplast genome and the phylogeny of *Begonia*. Edinb J Bot. 2022;79:Article408.
73. Ruhlman TA, Jansen RK. The plastid genomes of flowering plants. In: Maliga P, editor. Chloroplast biotechnology. Totowa, NJ, USA: Humana Press; 2014. p. 3–38.
74. Hsieh CL, Xu WB, Chung KF. Plastomes of limestone karst gesneriad genera *Petrocodon* and *Primulina*, and the comparative plastid phylogenomics of Gesneriaceae. Sci Rep. 2022;12:15800.
75. Gregory TR. DNA barcoding does not compete with taxonomy. Nature. 2005;434:1067.
76. Dong WP, Xu C, Li CH, Sun JH, Zuo YJ, Shi S, Cheng T, Guo JJ, Zhou SL. *ycf1*, the most promising plastid DNA barcode of land plants. Sci Rep. 2015;5:8348.

Publisher's Note