Plastome Evolution at the Edge: Structural Rearrangements, IR Expansion, and Gene Flux in Hypericaceae (Malpighiales)

Sivagami Jean Claude
Gachon University
Kyutae Park
Kyungpook National University
SeonJoo Park
sjpark01@ynu.ac.kr

Yeungnam University

Research Article

Keywords: Plastome evolution, Hypericaceae, Hypericum, Genome rearrangements, matK relocation, IR expansion

Posted Date: October 8th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-7295767/v1

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at BMC Plant Biology on December 23rd, 2025. See the published version at https://doi.org/10.1186/s12870-025-07888-7.

Abstract

Background

The evolutionary history of the *Hypericaceae* Juss. family remains poorly understood despite previous phylogenomic efforts. A prior study on *Hypericum ascyron* revealed exceptional plastome rearrangements and gene loss events, prompting questions about whether such genomic patterns are unique to *Hypericum* or reflect broader evolutionary trends within the family.

Results

To explore plastome evolution across *Hypericaceae*, we sequenced 14 complete chloroplast genomes representing seven genera from the three major tribes: *Hypericeae*, *Vismieae*, and *Cratoxyleae*, alongside two genera from the related family *Clusiaceae*. Comparative analyses revealed extensive variation in plastome structure and gene content, including lineage-specific rearrangements, inversions, and expansions of inverted repeat (IR) regions. Notably, species within *Vismieae* exhibited significantly expanded IR regions. In *Hypericum*, unique lineage-specific open reading frames (ORFs) were identified, with genes such as *accD* and *matK* relocated into or near the IR regions, likely driven by repeat-mediated recombination. Multiple independent losses of genes (*rpl23*, *rpl32*, *rps16*, *infA*, *ycf1*, *ycf2*) and introns were observed across the family, particularly in *Hypericeae*, often accompanying structural rearrangements. Additionally, *matK* was translocated from its typical position within the *trnK-UUU* intron into the IR region, a rare event in angiosperm plastomes. The protein-coding genes *accD* and *clpP* also showed domain-disrupting expansions, potentially impacting their functional roles.

Conclusions

Our results demonstrate that plastome evolution in *Hypericaceae* is highly dynamic, characterized by substantial structural plasticity, gene loss, and lineage-specific innovation. These findings provide new insights into plastome diversification across the family and lay the groundwork for further phylogenomic and evolutionary studies within Malpighiales.

Introduction

The order Malpighiales Juss ex Bercht & J. Presl, one of the most taxonomically and ecologically diverse clades of flowering plants, was first resolved through molecular phylogenetic studies by Chase et al. (1993). This reclassification reshaped the understanding of angiosperm relationships and highlighted the potential of molecular tools in resolving deep evolutionary lineages (Chase et al., 1993, Savolainen et al., 2000, Soltis et al., 2000). Among its major subclades, the clusioid clade comprising five families: Bonnetiaceae L. Beauvis Ex Nakai, Calophyllaceae J.Agardh, Clusiaceae Lindl., Hypericaceae Juss, and Podostemaceae Rich., represents a morphologically and biogeographically complex lineage with over

1,900 species spread across 94 genera (Ruhfel et al., 2011, Stevens, 2007a, Wurdack and Davis, 2009). Members of this clade are predominantly tropical, though certain genera extend into temperate and high-altitude regions, often adapting to extreme or specialized ecological niches such as montane forests and fast-flowing aquatic systems (Davis et al., 2005, Ruhfel et al., 2016). Within the clusioid clade, Hypericaceae stands out due to its taxonomic complexity and unevenly studied genera. Historically treated as a subfamily within Clusiaceae, Hypericaceae was elevated to family rank based on distinct morphological and molecular features (Kubitzki et al., 2007). It currently comprises approximately 518–700 species across nine genera and is taxonomically divided into three tribes: Hypericeae Choisy, Vismieae Vandelli, and Cratoxyleae Engl. APG III, 2009 (Crockett and Robson, 2011). The genus *Hypericum*, the most species-rich lineage with a cosmopolitan distribution, accounts for nearly 80% of the family's diversity. In contrast, the other genera, many of which are confined to tropical regions, have remained poorly sampled in molecular studies, leaving tribal and intergeneric relationships largely unresolved (Nürk et al., 2013, Stevens, 2007b).

Previous phylogenetic efforts in Hypericaceae have focused heavily on *Hypericum*, employing morphological data (Robson, 1977, Robson, 1981a, Robson, 1981b, Robson, 1985, Robson, 1990, Robson, 2006, Robson, 2010a, Robson, 2010b, Robson, 2012), ITS markers (Nurk and Blattner, 2010, Park and Kim, 2004), and plastid regions (e.g., *trn*L-*trn*F, *psb*A-*trn*H). These studies provided key insights into biogeographic patterns, such as the disjunction between New World and Old World clades, and proposed evolutionary scenarios involving ancient Gondwanan dispersal and multiple niche shifts (Meseguer et al., 2015, Nürk et al., 2015, Meseguer et al., 2014). However, these single- or low-copy gene datasets offered limited resolution at deeper nodes and often failed to resolve intertribal relationships, especially for Vismieae and Cratoxyleae.

Plastid genomes (plastomes) have emerged as powerful tools for resolving angiosperm relationships at both deep and shallow phylogenetic levels (Jansen et al., 2007). Their quadripartite structure, gene content, and substitution dynamics offer evolutionary signals that are often inaccessible through nuclear markers alone. Furthermore, increasing evidence suggests that plastome structural variation, such as inversions, IR boundary shifts, and gene loss, is not just phylogenetically informative but may also reflect lineage-specific ecological adaptations and genome evolution under selection (Guisinger et al., 2011a, Wicke et al., 2011). In this context, *H. ascyron*. was recently shown to possess one of the most highly rearranged plastomes among angiosperms, with large inversions, pseudogenization, and accelerated substitution rates (Claude et al., 2022). Whether this pattern is unique to *H. ascyron* or indicative of a broader trend across Hypericaceae remains unclear.

To address this question and improve resolution of phylogenetic relationships within Hypericaceae, we sequenced and analyzed complete chloroplast genomes from 14 species across all three tribes: *Hypericeae* (including *Hypericum*, *Triadenum* Raf., and *Thornea* Breedlove & E.M.McClint), *Vismieae* (*Harungana* Lam., *Vismia* Vand.), and Cratoxyleae (*Cratoxylum* Blume., *Eliea* (Lam.) Cambess.). Two species from Clusiaceae were included as outgroups. By combining comparative plastome analysis with phylogenomic reconstruction based on 67 plastid protein-coding genes, we aimed to:

1. Assess the extent and pattern of structural plastome variation across the family,
2. Investigate gene loss, pseudogenization, and possible functional relocation events and
3. Resolve tribal and infrageneric relationships in light of previous nuclear and morphological findings.

Our results provide new insight into plastome evolution in Hypericaceae and reveal lineage-specific trends in genome reorganization that align with ecological transitions and biogeographic structuring. These findings not only refine the phylogenetic backbone of the family but also shed light on how structural plastome dynamics correlate with environmental adaptation and diversification.

Results

Variation in Plastome Structure and Size across Hypericaceae Tribes

Plastome sizes in the tribe Hypericeae ranged from 138,163 bp in *Hypericum patulum* to 167,482 bp in *Triadenum fauriei* (Table 1). The large single-copy (LSC) region exhibited substantial variation, spanning from 103,150 bp in *H. patulum* to 85,948 bp in *H. calcicola*. The inverted repeat (IR) region reached its maximum length in *H. laxum* (32,823 bp) and its minimum in *H. patulum* (12,032 bp). The small single-copy (SSC) region showed a narrower range, from 10,423 bp in *H. laxum* to 12,421 bp in *T. calcicola*. In the Vismieae tribe, plastome sizes were generally larger, ranging from 166,590 bp in *Vismia baccifera* to 176,170 bp in *Harungana madagascariensis* (Fig. 1). The LSC region measured 67,585 bp in *V. baccifera* and 57,619 bp in *H. madagascariensis*, while the IR region was largest in *H. madagascariensis* (52,124 bp) and smallest in *V. baccifera* (42,583 bp). The SSC region in both species remained within the size range observed in Hypericeae. For the Cratoxyleae tribe, plastome sizes were more conserved, with a median length of approximately 156–157 kb. The LSC region measured 85,588 bp in *Eliea articulata* and 86,024 bp in *Cratoxylum maingayi*. The SSC regions ranged from 17 to 18 kb, and the IR regions ranged from 25 to 26 kb. These values are comparable to those found in the outgroup plastomes of *Mesua riparia* and *Rheedia edulis*. Across all three tribes, plastome sizes varied from 138,163 bp (*H. patulum*) to 176,170 bp (*H. madagascariensis*), reflecting both lineage-specific expansions and reductions in chloroplast genome architecture (Table 1).

Table 1

able 1. Summary of plastome features for 15 Hypericaceae species used in this study for phylogenetic analysis. The table includes total plastome size, lengths of the large single- copy (LSC), inverted repeat (IR), and small single-copy (SSC) regions, GC content, and the number of protein-coding sequences (CDS) and tRNA genes.

Species	Total	LSC	IR	SSC	GC %	CDS	tRNA
H. ascyron	162,286	97,542	26,846	11,052	37.40%	74	34
H. erectum	143,167	99,272	16,370	11,155	37.10%	72	34
H. chejuense	142,029	98,131	16,375	11,148	37.20%	73	34
H. patulum	138,163	103,150	12,032	10,949	38.10%	72	34
H. laxum	165,288	89,219	32,823	10,423	38.10%	73	34
H. japonicum	165,190	89,123	32,805	10,457	32.00%	75	34
T. fauriei	167,482	92,077	31,916	11,191	37.40%	75	34
T. calcicola	150,935	85,948	26,283	12,421	37.90%	75	34
H. madagascariensis	176,170	57,619	52,124	14,303	36.90%	75	35
H. rubescens	174,661	57,746	51,809	13,297	37.10%	75	35
V. baccifera	166,590	67,585	42,583	13,839	36.90%	75	35
E. articulata	156,123	85,588	25,640	19,254	36.40%	77	35
C. maingayi	157,279	86,024	26,155	18,945	36.30%	77	35
M. riparia	156,826	84,914	27,057	17,798	36.00%	77	35
R. edulis	158,160	86,384	27,010	17,756	35.80%	78	35

Plastome Size Variation in Hypericaceae and Related Malpighiales Lineages

To evaluate patterns of plastome size evolution within *Hypericaceae* and across the broader order Malpighiales, we compared plastome sizes among 381 species, including representatives from both focal and outgroup families (Fig. 2). The distribution of plastome sizes revealed substantial inter- and intra-lineage variability. Species within

Hypericum, including *H. ascyron*, *H. laxum*, and *H. patulum*, clustered within a relatively narrow and conserved size range. The genera *Triadenum* and *Eliea*, also members of Hypericaceae, showed a similar pattern. In contrast, taxa from the genera *Harungana* and *Vismia* exhibited significantly larger plastome sizes, appearing as outliers relative to the general distribution of Hypericaceae and most Malpighiales families. These enlarged plastomes may reflect lineage-specific structural expansions or

accumulation of non-coding sequences. At the lower end of the spectrum, *H. chejuense* and *H. erectum* displayed comparatively reduced plastome sizes within Hypericaceae, suggesting possible genome contraction events. Overall, the plastome size comparison highlights both the evolutionary conservation and structural divergence across Malpighiales, with *H. ascyron* and its close relatives occupying a distinct position within this landscape.

Distribution of Tandem Repeats Across Hypericaceae Plastomes

The comparative analysis of tandem repeat content revealed notable variability across Hypericaceae plastomes and outgroup species from Podostemaceae, Calophyllaceae, and Clusiaceae (Fig. 3). Species within Hypericaceae, particularly *T. japonicum*, *H. ascyron*, and *H. hookerianum*, exhibited a markedly higher number of tandem repeats, with over 90 repeats detected in some cases. These were predominantly within the 100–199 bp and 200–499 bp length classes. In contrast, outgroup taxa such as *M. riparia* and *R. edulis* displayed relatively low repeat abundance, primarily in the 50–199 bp range. Notably, several Hypericaceae species harbored long repeats (≥ 500 bp), which were either rare or absent in outgroup plastomes. The presence of longer and more numerous repeats in Hypericaceae suggests lineage-specific repeat proliferation, potentially contributing to the structural rearrangements and plastome plasticity observed in this family.

Structural Evolution of the Hypericaceae Plastome

Hypericeae Tribe

Comparison of plastomes from the Hypericeae tribe with *Mesua ferrea* used as a reference for the ancestral plastome structure revealed dynamic structural rearrangements, including multiple inversions, segmental rearrangements, and inverted repeat (IR) boundary shifts **(Supplementary Fig. 1**). Mauve alignment analysis identified 25 locally collinear blocks (LCBs), suggesting at least seven inversion events involving 21 breakpoints, including: *trnH–psbA*, *psbA–trnK*, *rps16–psbL*, *atpA–rps2*, *rpoC2–rpoC1*, *rpoC1–rpoB*, *petN–psbM*, *rps7–psbD*, *psbD–ndhC*, *ndhC–ndhJ*, *atpE–rbcL*, *accD*, *psaI–petA*, *clpP–psbJ*, *psbB–rps19*, *rpl2*, *petG*, *ndhG*, *ndhF*, *rpl32–rps15*, and *ycf1*. In total, we identified ten separate inversion events across the Hypericeae plastomes within LCB regions 2–11 and 17. Specifically:

- *H. ascyron* exhibited eight inversions,
- *H. chejuense* and *H. erectum* showed eleven inversions,
- *T. calcicola* displayed ten inversions,
- *H. patulum* showed seven inversions,
- *H. laxum* and *T. fauriei* had six inversions each.

The largest inversion identified across Hypericeae was LCB 8 (~ 14 kb), present in all species, and LCB 18 (~ 15 kb) in *H. ascyron*, *H. chejuense*, and *H. patulum*. Notably, one inversion located in the IR region was associated with IR expansion. A prominent structural rearrangement involved the relocation of the *matK* gene from the LSC region into the IR region, consistently observed across all species in the tribe. This rearrangement was accompanied by the apparent loss of the *trnK-UUU* gene, leaving only a 655 bp remnant of the 5' intron with 73.1% sequence identity (**Supplementary Fig. 5**).

Vismieae Tribe

The plastomes of the Vismieae tribe also demonstrated dynamic structural variation compared to *Mesua ferrea*, including six inversion events (LCBs 3, 4, 5, 7, 9, and 12), IR boundary shifts, and a notably large inversion spanning 54 kb (LCB 4). Mauve alignment revealed 12 LCBs corresponding to seven inversions and ten breakpoints, including: *trnH–psbA*, *trnK–rps16*, *trnG–trnQ*, *rbcL–trnR*, *petA–accD*, *psbJ–rps18*, *clpP–rpl20*, *psbB–trnN*, *rps15–ndhF*, and *ycf1* (**Supplementary Fig. 2).** Despite these rearrangements, a high degree of synteny was maintained among Vismieae plastomes. The IR regions were markedly expanded (~ 54 kb), distinct from other Hypericaceae tribes. As a consequence of this expansion, a partial duplication of the *ycf1* gene was detected only in Vismieae plastomes.

Cratoxyleae Tribe

Plastomes from the Cratoxyleae tribe exhibited a single large inversion spanning ~ 65 kb compared to *Mesua ferrea*. Mauve alignment identified four LCBs, suggesting two inversion events with three major breakpoints: *trnH–psbA*, *rbcL–trnK*, *accD–trnN*, and *ndhF–ycf1* (**Supplementary Fig. 3**). The two representative species—*Cratoxylum maingayi* and *Eliea articulata*, displayed identical syntenic block structures, indicating structural conservation within the tribe.

Ancestral Plastome Structural Evolution in Hypericaceae

Phylogenomic analysis suggests that the Cratoxyleae tribe, comprising the genera *Eliea* and *Cratoxylum*, represents the most ancestral lineage within the Hypericaceae family. A defining event in this lineage is a single large inversion spanning approximately 65 kb, representing the earliest major plastome structural rearrangement (Fig. 4). This inversion distinctly positions the *matK* gene within the large single-copy (LSC) region and the *accD* gene within the inverted repeat (IR) region, a configuration not found in the Hypericeae tribe, where a unique rearrangement consistently relocates *matK* into the IR region and positions *accD* within the LSC. This pattern is characterized by a *matK* relocation to the IR—is specific to the Hypericeae tribe and is not observed in the ancestral Cratoxyleae or in the structurally intermediate Vismieae tribe.

In Cratoxyleae plastomes, this arrangement contrasts with the canonical organization found in the core *Hypericum* group. Notably, 22 plastomes across Hypericaceae show shifts in gene order near the LSC–IR boundaries, particularly involving the repositioning of *trnH* and *rbcL*, often replacing the ancestral

location of *matK*. This rearranged configuration appears to be a defining structural signature of the Hypericeae lineage.

Another ancestral structural event occurred in the Vismieae tribe, which exhibits a ~ 54 kb inversion encompassing two large- locally collinear blocks (LCBs) (Fig. 4). This inversion appears to represent a secondary rearrangement relative to the Cratoxyleae event, with breakpoints near the *accD* and *trnK-UUU* genes. The Vismieae plastomes also display a pronounced IR expansion, extending into the *ycf1* region, resulting in a partial duplication of the *ycf1* gene in the opposing IR copy, a unique feature restricted to this lineage.

In contrast, plastomes from the Hypericeae tribe reveal extensive structural complexity, characterized by a greater number of LCBs and at least 15 breakpoints localized primarily within the LSC region (**Supplementary Figs. 1–3**). These rearrangements reflect multiple independent inversion events and IR boundary shifts, marking Hypericeae as the most structurally dynamic lineage among the three tribes. This pattern suggests a progressive accumulation of plastome complexity throughout the diversification of Hypericaceae.

IR Expansion and Contraction in the Hypericaceae Family

Structural variation in the IR boundaries of Hypericaceae plastomes, relative to the ancestral plastome of *Mesua ferrea*, reveals dynamic patterns of expansion and contraction, resulting in gene duplication and relocation events across lineages (Fig. 3; **Supplementary Fig. 4**). Within the Hypericeae tribe, significant IR contraction was observed at the LSC/IRA junction in *Hypericum ascyron* and *H. chejuense*, reducing the IR by approximately 11–13 kb and excluding a block of eight genes (from *rpl2* to *rps12*). An even more pronounced contraction (~ 17 kb) was detected in *H. erectum*, affecting the *rps11–rps7* region at the LSC/IRA/SSC boundaries. In contrast, species such as *H. laxum* and *T. japonicum* within Hypericeae, as well as all examined members of the Vismieae tribe, showed no evidence of IR contraction. Across the family, IR expansions at both the IRB/SSC and IRA/SSC boundaries were more consistently observed. These expansions led to the duplication of *ndhF* in most species across Hypericeae, Vismieae, and Cratoxyleae tribes. The only exception was *Thornea calcicola*, which retained a full copy and a partial copy of *ndhF* near the IRB/SSC boundary. In addition, a complete duplication of *ycf1* was detected only in *H. ascyron* and *T. japonicum*, suggesting lineage-specific expansions in these taxa (Fig. 4). A particularly notable case of large-scale IR expansion occurred in the Vismieae tribe, where the IR region extended from IRB/LSC to the *rpl2–psaJ* region, expanding to approximately 56 kb—nearly double the IR size compared to other tribes. This boundary shift resulted in the duplication of up to 30 genes, with no contraction events observed in this tribe. Similarly, the Cratoxyleae tribe showed no evidence of contraction but exhibited an expansion at the IRB/SSC junction, leading to the duplication of the entire *ycf1* gene on one side. A partial copy (~ 981 bp) of *ycf1* was found at the opposite IR boundary near the LSC/IRA junction (Fig. 4). These IR boundary dynamics of expansions, contractions, and gene duplications that demonstrate the extensive plastome structural evolution across the Hypericaceae family and highlight the lineage-specific mechanisms shaping IR architecture. Comparative analysis of

the *accD* gene across Hypericaceae plastomes revealed the presence of internal insertions comprising tandem repeat sequences, with notable variation in both length and position. Two species, *H. erectum* and *H. chejuense*, exhibited exceptionally long insertions of 1,468 and 1,484 *bp*, respectively, disrupting the conserved coding region of *acc*D. In contrast, most other species contained shorter insertions ranging from 79 bp to 459 *bp*. These repeat-derived expansions fractured the gene's conserved domains and contributed to significant structural divergence among lineages. The pattern suggests that repeat proliferation has played a key role in *acc*D evolution, potentially impacting gene function and driving plastome plasticity within Hypericaceae.

Gene Relocation and ORF Emergence in Hypericum Plastomes

Comparative analysis of plastid genomes among five *Hypericum* species (*H. ascyron*, *H.* chejuense, *H. erectum*, *H. laxum*, and *H. patulum*) revealed lineage-specific structural rearrangements associated with the relocation of *acc*D and *mat*K genes and the emergence of novel open reading frames (ORFs) (**Supplementary Fig. 5**). In *H. ascyron*, both *mat*K and *acc*D are translocated into or near the inverted repeat (IR) regions. This rearrangement is accompanied by the presence of lineage-specific ORFs, including ORF13 (between *acc*D and *mat*K) and ORF15 (downstream of *mat*K). These novel ORFs are absent in other species, suggesting they may have originated from IR-mediated recombination or structural rearrangement at the IR junctions. In *H. chejuense* and H. *erectum*, similar patterns are observed. While *mat*K remains conserved, ORF66/67 (*H. chejuense*) and ORF8/9 (*H. erectum*) are located downstream, suggesting independent ORF formation following structural rearrangement events. In *H. laxum*, *acc*D is placed upstream of ORF1, again implying that the disruption of plastome structure near *acc*D may promote ORF emergence. In contrast, *H. patulum* shows a markedly different organization. Both *acc*D and *mat*K are located adjacently within the large single-copy (LSC) region, and no novel ORFs are detected in their vicinity. This more conserved structure likely reflects the absence of recombination-prone IR junctions near these genes, supporting the hypothesis that relocation of coding genes into the IR and disruption of neighboring intergenic regions play a key role in the generation of novel ORFs in specific *Hypericum* lineages.

Parallel Structural Placement of matK and accD Genes

An unusual ~ 65 kb inversion in the basal clade of the Cratoxyleae tribe positions the *trnH–rbcL* region at the beginning of the large single-copy (LSC) region, placing *trnH* and *matK* in a canonical configuration typically observed in chloroplast genomes (Fig. 4). Structural comparisons across Cratoxyleae, Vismieae, and Hypericeae tribes revealed a pattern of parallel rearrangements affecting the relative positions of *matK* and *accD* genes, which may suggest signatures of biparental structural inheritance or convergent rearrangement events. In Cratoxyleae, the *matK* and *accD* genes are positioned adjacently in the LSC region (*matK:accD*), a configuration inferred to represent the ancestral state. In contrast, a second large inversion (~ 54 kb) in the Vismieae tribe results in the reversal of this order to *accD:matK*, positioned adjacent to the inverted repeat (IR) region, along with a substantial IR expansion. This modified *accD:matK* orientation is also retained in the basal genus of the Hypericeae tribe (*Triadenum*),

accompanied by the incorporation of *ycf1* into the IR, further supporting a shared structural signature between Vismieae and early diverging Hypericeae plastomes. Interestingly, *Thornea* and *H. patulum* retain the ancestral *matK:accD* configuration, suggesting structural conservation with the Cratoxyleae lineage. However, pronounced divergence is evident in the *Hypericum* core clade, particularly the Trigynobrathys section (*H. laxum*, *H. japonicum*), where multiple inversion events appear to have disrupted the original gene order. In this group, the repositioning of *accD* into the LSC and *matK* into the IR region is observed, except in *H. patulum*, which uniquely retains the ancestral *Cratoxyleae*-like arrangement. These findings indicate that the *matK–accD* gene pair serves as a phylogenetically informative marker reflecting structural divergence within Hypericaceae.

Selective Retention of matK–trnK-UUU

In plastomes of the Hypericeae tribe, substantial restructuring occurred around the *trnK-UUU* region. The *trnK-UUU* gene was lost in all examined species, and *matK* was retained as a free-standing gene, uncoupled from its usual intronic context. Despite the loss of intron, *matK* was preserved across all species, indicating strong functional conservation (**Supplementary Figs. 5 and 6**). The Hypericeae plastomes retained only five introns overall, with both the cis-spliced *rps12* and *trnK-UUU* introns being absent. Selection analysis showed dN/dS ratios > 1 for *matK* along the branches leading to Hypericaceae and related Clusioid lineages. Likelihood ratio tests (LRTs) supported the hypothesis that *matK* underwent positive selection, likely driven by structural reorganization and the functional compensation for the loss of *trnK-UUU* (Supplementary Figure). Furthermore, although most plastid genes showed conservation, some *rpo* genes (e.g., *rpoC1*) displayed intron loss and signs of positive selection. These genes encode the RNA polymerase type I enzyme, which is essential for plastid tRNA and mRNA synthesis. The adaptive evolution of this gene group may reflect compensatory mechanisms associated with plastome structural reconfiguration in Hypericaceae. Collectively, these findings suggest that the relocation and selection of *matK*, coupled with functional shifts in *rpo* genes, are key evolutionary responses to plastome rearrangement in Hypericaceae. A more comprehensive sampling across this family will be essential to fully elucidate the evolutionary mechanisms underlying elevated substitution rates and gene reorganization.

Gene and Intron Loss Patterns in Hypericaceae and Related Lineages

The plastomes of Hypericeae species contained 69–75 protein-coding genes, 29 tRNA genes, and four rRNA genes (Figs. 2–8; Table 1). Notably, the translation initiation factor (*inf*A), ribosomal protein S16 (*rps*16), and tRNA-Lys (*trn*K-UUU) genes were completely absent across all examined Hypericeae plastomes (**Supplementary Fig. 7**). Additionally, *rps*7 was found to be pesduogenized only in *H. ascyron*, *H. erectum*, and *H. patulum*, while *rpl*23 was pesduogenized in all Hypericeae species. The *rpl*32 gene also appeared to be pesduogenized in all Hypericeae members, likely due to frameshift mutations and internal stop codons. Complete or near- loss of *ycf*1 and *ycf*2 was detected in *H. erectum*, *H. chejuens*e, *H. patulum*, and *T. calcicola*. Intron loss was also widespread in Hypericeae plastomes. Specifically, all

species lacked both introns of the *clp*P gene, the second (cis-spliced) intron of *rps*12, the *rpo*C1 intron, and the second intron of *ycf*3. In the Cratoxyleae tribe, gene loss was limited to *inf*A and rps16, while intron losses mirrored those in Hypericeae, with absence of introns in *rps*12, *ycf*3, and the second intron of *clp*P. The Vismieae tribe also exhibited complete loss of *inf*A and *rps*16, and pseudogenization of *rpl*23, *rpl*32, and *rps*7, as well as *ycf*1 and *ycf*2 in all sampled species of *Harungana* and *Vismia*. Intron losses in Vismieae included *rps*12, *ycf*3, *clp*P, *rpo*C1, and *atp*F, suggesting extensive structural reduction across the plastomes of this lineage. In the related Clusiaceae family, plastome gene losses were more limited: *inf*A was absent in both *Mammea riparia* and *Rheedia edulis*, while *rps*16 was lost only in *M. riparia*. Both species exhibited intron loss in *ycf*3 but retained other intron-containing genes intact.

Phylogenetic Reconstruction of Hypericaceae and Related Malpighiales

To reconstruct evolutionary relationships within the Hypericaceae family and among related lineages in the order Malpighiales, a maximum likelihood (ML) approach was applied using a concatenated alignment of 59 plastid protein-coding genes (supermatrix length: 46,929 bp) across 29 genera, including 25 Hypericaceae species and 4 outgroup taxa (Fig. 4). The outgroup comprised representatives from the clusioid clade, including *Hypericum monogynum* and *Calophyllum cochinchinense* (Hypericaceae), *Tristicha trifaria* (Podostemaceae), *Mesua ferrea* (Calophyllaceae), and *Moronobea* and *Rheedia* species (Clusiaceae) (Fig. 5**).**

The ML analysis yielded an optimal tree topology with a log-likelihood of − 226352.109, while both ML and BI trees displayed highly congruent topologies. All major branches were strongly supported with bootstrap support (BS) = 100 and Bayesian posterior probability (PP) = 1.00 (Fig. 5). Phylogenetic analyses recovered the Cratoxyleae tribe as a monophyletic group and sister to a clade composed of Vismieae and Hypericeae, confirming tribe-level relationships within Hypericaceae. Within this framework, Vismieae was recovered as sister to Hypericeae, indicating a closer evolutionary affinity between these two tribes.

Among the *Hypericum* species:

- *H. laxum* and *H. japonicum* grouped closely with *Triadenum fauriei* and *T. japonicum*, forming a well-supported clade corresponding to the Trigynobrathys section (BS/PP = 100/1.00).
- *Thornea* species also clustered within this clade, confirming their close phylogenetic relationship to *Triadenum* and the Trigynobrathys section.
- *H. patulum* and *H. monogynam* formed a clade within the Ascyreia s.l. + campylosporous section, which was resolved as sister to the Roscyna section represented by *H. ascyron* (BS/PP = 100/1.00).
- In the core *Hypericum*, *H. erectum* and *H. chejuense* formed a distinct clade (BS/PP = 100/1.00), which was recovered as sister to the combined Ascyreia s.l., Roscyna, and campylosporous lineages.

Within Vismieae, *Harungana* and *Vismia* species were recovered as a well-supported monophyletic group. In Cratoxyleae, *Cratoxylum* and *Eliea* formed a strongly supported sister clade to the Hypericeae–Vismieae clade (BS/PP = 100/1.00), further confirming tribal boundaries. Additional ML phylogenies constructed using individual plastid gene sets (e.g., *atp*, *ndh*, *pet*, *psa*, *psb*, *rpl*, *rps*, *rpo*, *ycf*) yielded consistent topologies (Supplementary Fig. 1), though variation was observed in certain marker-based trees (e.g., *ccsA*, *clpP*, *pet*, *psb*, *ycf*), which showed slight topological shifts in relationships among tribes. Analysis of nuclear ITS sequences provided complementary evidence, particularly supporting the separation of *H. laxum* and *H. japonicum* from core *Hypericum*, instead of grouping them with *Triadenum*, thereby corroborating plastome-based results (**Supplementary Fig. 8**). Notably, the relationships among *Thornea*, *Triadenum*, and the Trigynobrathys section appeared controversial in certain gene trees, particularly those constructed using *rps*, *pet*, and *ycf* genes. Collectively, both coalescent-based and concatenated tree methods support a scenario in which: The New World *Hypericum* clades (e.g., *Thornea*, *Triadenum*, *Trigynobrathys*) diverge early, While both New and Old-World clades (e.g., *Androsaemum*, *Ascyria*, *Roscyna*, and the *Hypericum* core) form a monophyletic group sister to Vismieae. These results provide robust phylogenomic evidence for major taxonomic realignments within Hypericaceae and reinforce plastome structure and gene order as useful phylogenetic markers for resolving complex tribal and sectional relationships.

Elevated Substitution Rates in Hypericaceae

Analysis of substitution rates showed that plastid genes in Hypericeae experienced significantly elevated rates of molecular evolution in specific bordered disturbed genes among the tribe. Nonsynonymous (dN) and synonymous (dS) substitution rates were observed higher, respectively, than those observed in *Cratoxylum*. *H. ascyron* exhibited especially high substitution rates across most plastid genes, with dN and dS values are greater than in *Cratoxylum*, respectively. The genes *accD*, *clpP*, and *matK*, which also underwent structural modifications, showed markedly accelerated substitution rates compared to those in *Tristicha* (**Supplementary Fig. 10–15**). Branch-site models and RELAX analyses further revealed that several genes in Hypericaceae underwent episodic positive selection (**Supplementary Table 2–3**). The genes *accD*, *clpP*, *matK*, *cemA*, *psbN*, *rpl33*, *rps3*, *rps12*, *rpoA*, and *rpoC2* showed elevated dN/dS ratios along branches leading to *Hypericum*, *Cratoxylum*, and *Marathrum*. Likelihood ratio tests (LRTs) indicated significant positive selection ($p < 0.05$, Bonferroni-corrected) in branches leading to (*Hypericum* + *Cratoxylum* + *Marathrum*) for *clpP*, (*Hypericum* + *Cratoxylum*) for *accD*, *clpP*, and *matK*, and at the *Hypericum* terminal branch for *accD*, *clpP*, and *matK* (**Supplementary Fig. 10–16**).

Discussion

In a previous study, we reported that the plastid genome of *H. ascyron* exhibited extensive structural disruption, including multiple large inversions, gene and intron loss, and significantly elevated substitution rates (Claude et al., 2022). These features positioned *H. ascyron* as one of the most highly rearranged plastomes among angiosperms and raised questions about whether such plastomes instability was species-specific or indicative of a broader evolutionary trend within Hypericaceae. The

present study addresses this by expanding the taxonomic scope to include representatives from all three tribes of Hypericaceae (Hypericeae, Vismieae, and Cratoxyleae) and conducting comparative analyses with outgroups from Clusiaceae. Our results clearly demonstrate that the patterns observed in *H. ascyron* are reflective of recurring evolutionary mechanisms operating across the family.

Plastome Structural Dynamics and Gene Evolution in Hypericaceae

Structural variation was widespread. Plastome sizes ranged from 138,163 bp (*H. patulum*) to 176,170 bp (*H. madagascariensis*), with much of this variation attributable to IR expansions and contractions. Members of the Vismieae tribe exhibited dramatic IR expansions exceeding 50 kb, leading to the duplication of over 20 genes. Conversely, significant IR contraction in *Hypericeae* resulted in the loss of multiple boundary-associated genes. These findings are paralleled in other lineages of Malpighiales, most notably in Passifloraceae, which exhibit IR loss and extensive structural reorganization (Cauz-Santos et al., 2020), and Podostemaceae, which display compact, gene-reduced plastomes with high levels of rearrangement (Bedoya et al., 2019). Such recurring patterns suggest that plastome instability may be an emergent trait in several clades of Malpighiales, especially those undergoing ecological specialization.

Gene loss and pseudogenization were most pronounced in the Hypericeae tribe, where *inf*A, *rpl*23, *rpl*32, and *rps*16 were absent in all species, while *rps*7, *ycf*1, and *ycf*2 were pesduogenized in several lineages. These losses likely reflect functional transfers to the nuclear genome, a process previously documented in other angiosperm families such as *Ranunculaceae* and Passifloraceae (Park et al., 2015, Shrestha et al., 2020). For example, *rpl*32 was functionally replaced by a nuclear-encoded SOD fusion in *Populus* (Ueda et al., 2007), and *rps*16 loss occurred repeatedly with the emergence of dual-targeting mechanisms (Ueda et al., 2008). In Geraniaceae, multiple plastid genes, including *clp*P, underwent loss or pseudogenization, often associated with nuclear transfer or coevolutionary constraints (Weng et al., 2016). These examples collectively highlight that gene loss in plastid genomes is not stochastic but frequently accompanied by functional compensation and structural rearrangement (Wicke et al., 2011, Jansen et al., 2007).

The consistent relocation of *matK* into the IR region and its decoupling from the *trn*K-UUU intron further suggest a functional repurposing under lineage-specific selective pressures. This inference is supported by our dN/dS analysis, which revealed elevated substitution rates in *mat*K across multiple branches in the Hypericeae tribe. Notably, in several species where *matK* has relocated, we observed the presence of adjacent partial or large open reading frames (ORFs), a feature absents in closely related species where *mat*K remains embedded within the *trn*K-UUU intron. These ORFs may represent truncated pseudogenes, novel gene fusions, or potentially co-evolving loci that have emerged in tandem with *mat*K displacement. Their consistent association with relocated *matK* implies a possible structural or regulatory role, although further functional characterization is needed. The relocation of *matK* or *accD* into IR regions, as seen in *Hypericum* species, may lead to structural disruption of neighboring intergenic regions and

generation of lineage-specific ORFs, a phenomenon documented in other taxa (Guisinger et al., 2011b, Wicke et al., 2011, Zhu et al., 2016). Underlying this process is the presence of repeat elements that including palindromic and inverted repeats at IR junctions, which are known hotspots for homologous or illegitimate recombination, facilitating genomic rearrangements and the creation of novel ORFs (Kolodner and Tewari, 1979, Day and Madesis, 2007, Odahara et al., 2015).

Evidence from previous studies reinforces the idea that *matK* is subject to adaptive evolution. Hao et al. (2009) found that *mat*K evolves under positive selection across multiple angiosperm lineages, particularly in domains critical to its maturase function(Hao et al., 2009). Furthermore, cases such as *Epifagus virginiana*, *Cuscuta*, and leptosporangiate ferns demonstrate that *matK* can function as a freestanding gene, independent of *trn*K-UUU, and still retain its splicing role (Kuo et al., 2008, Wicke et al., 2011). These findings align with our observations in Hypericaceae, suggesting that *matK* relocation is not a random genomic event but rather a potentially adaptive response accompanied by secondary structural or functional changes.

In addition, structural alterations in coding genes like *acc*D, *clp*P, and *rpoC*1 were associated with sequence divergence and loss of introns, a pattern consistent with accelerated substitution rates. These genes, essential for plastid function and organelle-nuclear communication, appear to be hotspots for adaptive evolution. Their modification, often via insertion or fragmentation, may be linked to selection acting on plastid performance in stress-prone habitats.

Phylogenetic Relationships and Tribal Evolution in Hypericaceae

Phylogenomic reconstruction based on 67 plastid protein-coding genes yielded a well-resolved tree that supports the monophyly of the three recognized tribes of Hypericaceae: Hypericeae, Vismieae, and Cratoxyleae. Our analyses resolved Cratoxyleae as the earliest diverging lineage and sister to a clade comprising Vismieae and Hypericeae. This branching pattern provides strong support for tribal distinctions and clarifies deep evolutionary splits within the family, which were previously ambiguous in studies based on nuclear ITS and plastid intergenic spacers (Nurk and Blattner, 2010, Nürk and Crockett, 2011, Park and Kim, 2004).

Cratoxyleae, with its conserved plastome structure and geographically restricted range in tropical Southeast Asia, likely represents an ancestral lineage within Hypericaceae. In contrast, Hypericeae and Vismieae, both show greater plastome rearrangements including IR expansion and gene loss that share a more recent common ancestor and reflect derived evolutionary trends. These patterns correspond with ecological transitions, as Hypericeae includes both tropical and temperate species, while *Vismieae* is largely tropical(Nurk and Blattner, 2010).

Within Hypericeae, our plastome phylogeny revealed unexpected relationships. *H. laxum* and *H. japonicum* form a strongly supported clade with *Triadenum* and *Thornea*, contradicting traditional genus-level assignments and indicating that these taxa may be better classified within a redefined *Hypericum*.

This finding aligns with earlier ITS and low-copy nuclear marker studies suggesting the paraphyly of core *Hypericum* and potential instances of incomplete lineage sorting or ancient hybridization (Meseguer et al., 2014, Norman, 2016). Morphological similarities among these species may therefore reflect retained ancestral traits or convergent evolution rather than deep taxonomic splits(Nürk et al., 2013).

Moreover, our results mirror geographic lineage structuring previously described in the genus. The *H. laxum–Triadenum* clade corresponds to the New World lineage (e.g., section *Trigynobrathys*), while species such as *H. erectum* and *H. patulum* group with Old World clades like *Ascyreia* and *Roscyna*. This East–West divergence has been attributed to climatic shifts during the Oligocene–Miocene, which drove altitudinal migrations, ecological niche shifts, and speciation events in *Hypericum* (Nürk et al., 2015, Meseguer et al., 2015). Our genome-scale phylogeny supports this biogeographic scenario and suggests that these divergent events were accompanied by plastome evolution and structural remodeling.

Our study provides the most robust plastome-based framework to date for Hypericaceae. It not only affirms tribal relationships but also highlights hidden paraphyly within *Hypericeae*, signals the need for genus-level taxonomic revision, and corroborates ecological and geographic hypotheses about the group's diversification. The concordance between plastome evolution, previous nuclear-based reconstructions, and biogeographic history reinforces a model of adaptive radiation in *Hypericum* driven by historical climate change and niche expansion across continents.

Environmental shifts may be a major driver of the plastome dynamism observed. Nürk et al. (2015) demonstrated that cold-tolerant *Hypericum* species radiated rapidly in montane habitats such as the Andes and East Asia, likely in response to Oligocene–Miocene climate changes. Structural genome evolution, including gene relocations, intron losses, and inversions may reflect adaptive responses to new ecological pressures, as similarly seen in *Geraniaceae* and *Silene* (Erixon and Oxelman, 2008, Park et al., 2017). Such plastome remodeling could facilitate transcriptional or regulatory flexibility under fluctuating environmental regimes.

In summary, this study provides the first family-wide plastome-based phylogenomic assessment of Hypericaceae, revealing extensive structural dynamism, recurrent gene loss, and adaptive rearrangements across its three major tribes. By expanding the scope beyond the previously analyzed *H. ascyron*, our data demonstrate that genome instability—manifested as IR expansion/contraction, gene relocation, pseudogenization, and intron loss—is not species-specific but a recurrent feature across the family, particularly within the *Hypericeae* and *Vismieae* lineages. The consistent relocation of *mat*K and its association with novel open reading frames, as well as elevated dN/dS ratios in *mat*K, *acc*D, and *clp*P, suggest that plastome remodeling is not a random consequence of genome decay, but rather reflects selective responses to ecological pressures. These structural changes are paralleled in other lineages of Malpighiales, such as Passifloraceae and Podostemaceae, where genome reconfiguration has also been linked to adaptation in specialized habitats. Phylogenetically, The placement of *Triadenum* and *Thornea* within the *Hypericum* clade calls for taxonomic revision and reflects historical processes such as hybridization and incomplete lineage sorting. These findings are consistent with previous nuclear and

morphological studies but now gain stronger support through genome-scale data. Biogeographically, our phylogeny reinforces the pattern of East–West lineage divergence within *Hypericum*, supporting a scenario of rapid radiation in response to Miocene climate shifts and montane expansion. The convergence of genomic rearrangements with ecological transitions suggests that plastome evolution played a significant role in enabling the widespread distribution and diversification of *Hypericum*. Altogether, this study not only provides a resolved phylogenetic framework for Hypericaceae but also advances our understanding of how plastid genome structure, gene content, and selection intersect with environmental history. These findings lay the groundwork for future investigations into plastome–nuclear interactions, functional consequences of gene loss, and the genomic basis of ecological adaptation in angiosperms.

Materials and Methods

Plant material and Next generation sequencing:

This study analyzed the plastid genomes of 14 species across the families Hypericaceae and Clusiaceae, including eight species from the tribe Hypericeae, three from Vismieae, and two from *Cratoxyleae*. Additionally, two species from Clusiaceae served as outgroups. Plant material used in this study was obtained either as leaf tissue or DNA from the Kew DNA Bank. Wild-collected samples from Korea were collected from common, non-protected species and therefore did not require special permits under Korean biodiversity regulations. Formal identification of the species was performed by SeonJoo Park, and voucher information is provided in **Supplementary Table 1**. Additional specimens were obtained from established DNA and herbaria, including the Missouri Botanical Garden, the Kew DNA Bank, and the University of Texas Herbarium, with voucher details available in **Supplementary Table 1**.This study did not involve the collection of any species listed in the Convention on the International Trade in Endangered Species of Wild Fauna and Flora (CITES). Fresh leaves were used for several species, including *H. erectum* Thunb., *H. patulum* Thunb., *H. chejuense* S.J.Park & K.J. Kim, *H. laxum* Blume., *H. japonicum* Thunp., and *Triadenum fauriei* R.Keller, with genomic DNA extracted using the GeneAll Plant SV Mini Kit (GeneAll Biotechnology, Seoul, Korea). Herbarium specimens leaves such as *Harungana madagascariensis* Lam. Ex Poir., *Vismia baccifera* (L.) Triana & Planch., *Eliea articulata* (Lam.), *Thornea calcicola* Standl. & Steyerm., *Rheedia edulis* (Seem.) Planch. & Triana., and *Moronobea riparia* Planch. & Triana. were processed using a modified 3× CTAB protocol with a 3-hour incubation, as described by Doyle and Doyle DNA extraction method (Allen et al., 2006). For *Cratoxylum maingayi* and *H. rubescens* (Oliv.) Byng & Christenh., DNA was sourced from the Kew DNA Bank. High-throughput sequencing was carried out on the Illumina HiSeq 2500 platform (Illumina Inc., San Diego, CA), generating approximately 6 Gb of 150 bp paired-end reads per sample from 550 bp insert libraries.

Assembly and genome size comparison

Chloroplast genome assembly was conducted using GetOrganelle v1.7.5.3 and Velvet v1.2.10, with k-mer sizes ranging from 97 to 147 optimized for genome coverage and accuracy (Jin et al., 2020, Zerbino

and Birney, 2008). Assembly quality was evaluated by mapping paired-end reads to the assembled plastomes using Bowtie2 v2.2.6(Langmead and Salzberg, 2012). The results were visually inspected and verified in Geneious R11.0.5 (https://www.geneious.com). To analyze genome size variation, 381 complete chloroplast genomes from the order Malpighiales were downloaded from the NCBI nucleotide database as of February 28, 2022. Genome size statistics and visualizations were generated in R v4.0.4 using the ggplot2 package(Wickham, 2016). All newly assembled plastome sequences have been deposited in the NCBI GenBank database under accession numbers [PQ010624–PQ010637].

Plastid gene annotation and Genome rearrangement

Gene annotation was initially performed in Geneious R11.0.5 using *Nicotiana tabacum* as the reference genome(Shinozaki et al., 1986). Gene boundaries were verified by BLASTN searches using NCBI-BLAST + v2.7.1(Camacho et al., 2009). Transfer RNA genes were annotated with tRNAscan-SE v2.0.3 (Chan and Lowe, 2019) and ARAGORN v1.2.38 (Laslett and Canback, 2004), and circular genome maps were constructed using OGDraw v1.3.1.(Greiner et al., 2019). Structural rearrangements were investigated by comparing species from each tribe of *Hypericaceae* with an outgroup species (*Mesua ferrea* L., Clusiaceae) using the progressive Mauve algorithm implemented in Geneious(Darling et al., 2010, Geneious).

Phylogenetic and substitution rate estimation

Phylogenetic reconstruction was based on 67 plastid protein-coding genes extracted from 29 plastomes. Each gene was aligned using MAFFT v7.450 with the G-INS-i strategy and concatenated into a supermatrix(Katoh and Standley, 2013). Phylogenetic inference was carried out using maximum likelihood in IQ-TREE v1.6.2 under the GTR + GAMMA + I model with 1,000 ultrafast bootstrap replicates(Minh et al., 2020), and Bayesian inference using MrBayes v3.3.7a(Ronquist et al., 2020). Coalescent-based species trees were generated using ASTRAL v5.7.8 with 67 gene trees derived from IQ-TREE, incorporating models selected by ModelFinder (Mirarab et al., 2014). Additionally, nuclear ITS regions were retrieved using GetOrganelle and aligned with MUSCLE(Edgar, 2004). ITS based phylogenies were reconstructed using IQ-TREE with the GTR + GAMMA + I model.

Selection analysis

To detect selection, nonsynonymous (dN) and synonymous (dS) substitution rates were estimated for each plastid gene using the CODEML program within the PAML v4.8 package under the F3×4 codon frequency model(Yang, 1997). Branch-site positive selection was tested using the adaptive Branch-Site Random Effects Likelihood (absREL) model in HyPhy v2.5 via the Datamonkey server, with Holm-Bonferroni correction for multiple testing(Pond et al., 2020). A total of 70 protein-coding genes were aligned using the Translation Align option in MAFFT, and likelihood ratio tests (LRTs) were conducted to evaluate variation in dN/dS ratios across branches. Visualization of substitution rates was performed in R v4.0.4 using ggplot2(Wickham, 2016). To further explore divergence in the inverted repeat (IR) regions, the *ycf1* and *ycf2* genes from 11 genera of Malpighiales, including a known pseudogene in *Passiflora*

edulis, were aligned using MAFFT. Phylogenetic trees were constructed using PhyML with 100 bootstrap replicates to assess evolutionary patterns associated with IR-specific gene dynamics(Guindon et al., 2010).

Declarations

Contributions

JCS performed the experiments, generated datasets and figures, and wrote the first draft of the manuscript. KTP contributed to the data assembled, performed analyses, and read/edited the manuscript. SJP contributed to project design and read/edited the manuscript. All authors read and approved the final draft of the manuscript.

Ethics declarations

Ethics approval and consent to participate

Experimental study on the plant, including collection of the material, comply with institutional, national, and international guidelines.

Clinical trial number

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Funding:

This work was supported by the 2024 Yeungnam University Grant (221A061009), South Korea.

Author Contribution

JCS performed the experiments, generated datasets and figures, and wrote the first draft of the manuscript. KTP contributed to the data assembled, performed analyses, and read/edited the

manuscript. SJP contributed to project design and read/edited the manuscript. All authors read and approved the final draft of the manuscript.All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

Data Availability

All newly generated plastome sequences have been deposited in GenBank under accession numbers [PQ010624–PQ010637]. Supporting data, including alignment files, gene order maps, and phylogenetic trees, are available in Supplementary Figures and Tables.

References

1. ALLEN GC, FLORES-VERGARA MA, KRASNYANSKI S, KUMAR S, THOMPSON WF. A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. Nat Protoc. 2006;1:2320–5.
2. BEDOYA AM, PHILBRICK RUHFELBR, & CT, ET AL. Plastid Genomes of Five Species of Riverweeds (Podostemaceae): Structural Organization and Comparative Analysis in Malpighiales. Front Plant Sci. 2019;10:1035.
3. CAMACHO C, COULOURIS G, AVAGYAN V, BEALER MANPAPADOPOULOSJ, K., MADDEN TL. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421.
4. CAUZ-SANTOS LA, DA COSTA ZP, CALLOT C, &, ET AL. A Repertory of Rearrangements and the Loss of an Inverted Repeat Region in Passiflora Chloroplast Genomes. Genome Biol Evol. 2020;12:1841–57.
5. CHAN PP, LOWE TM. 2019. tRNAscan-SE: Searching for tRNA Genes in Genomic Sequences. *Methods in Molecular Biology*, 1962, 1–14.
6. CHASE MW, SOLTIS DE, MORGAN OLMSTEADRG, MISHLER DLESDH, HILLS BDDUVALLMRPRICERA, H. G., QIU Y-L. 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene rbcL. Ann Mo Bot Gard, 528–80.
7. CLAUDE SJ, PARK S, PARK S. 2022. Gene loss, genome rearrangement, and accelerated substitution rates in plastid genome of (Hypericaceae). BMC Plant Biol, 22.
8. CROCKETT SL, ROBSON NK. Taxonomy and Chemotaxonomy of the Genus Hypericum. Med Aromat Plant Sci Biotechnol. 2011;5:1–13.
9. DARLING AE, MAU B, PERNA NT. progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss, and Rearrangement. PLoS ONE. 2010;5:e11147.

10. DAVIS CC, WEBB, C. O., WURDACK, K. J., JARAMILLO, C. A., DONOGHUE MJ. Explosive radiation of Malpighiales supports a mid-cretaceous origin of modern tropical rain forests. Am Nat. 2005;165:E36–65.
11. DAY A, MADESIS P. DNA replication, recombination, and repair in plastids. In: BOCK R, editor. Cell and Molecular Biology of Plastids. Berlin, Heidelberg: Springer; 2007. *In*:.
12. EDGAR RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32:1792–7.
13. ERIXON P, OXELMAN B. Whole genome duplication and gene loss in the evolution of the plastid clpP gene family in Silene. Mol Biol Evol. 2008;25:587–98.
14. GENEIOUS. progressiveMauve algorithm implemented in Geneious Prime (Mauve plugin).
15. 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.
16. GUINDON S, LEFORT DUFAYARDJ-F, ANISIMOVA V, HORDIJK M, W., GASCUEL O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 2010;59:307–21.
17. GUISINGER MM, KUEHL, J. V., BOORE, J. L., JANSEN RK. Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: rearrangements, repeats, and codon usage. Mol Biol Evol. 2011a;28:583–600.
18. GUISINGER MM, KUEHL, J. V., BOORE, J. L., JANSEN RK. Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: rearrangements, repeats, and codon usage. Genome Res. 2011b;21:734–45.
19. HAO D-C, CHEN MUJ, S.-L., XIAO P-G. Molecular evolution and positive Darwinian selection of the chloroplast maturase matK. J Plant Res. 2009;122:365–75.
20. JANSEN RK, CAI Z, RAUBESON LA, & ETAL. 2007. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proceedings of the National Academy of Sciences*, 104, 19369–19374.
21. JIN JJ, SONG YUWBYANGJB, Y., DEPAMPHILIS, C. W., YI, T. S., LI DZ. 2020. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol, 21.
22. KATOH K, STANDLEY DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772–80.
23. KOLODNER R, TEWARI KK. Inverted repeats in chloroplast DNA from higher plants. Proc Natl Acad Sci USA. 1979;76:41–5.
24. KUBITZKI K, BAYER C, STEVENS PF. Flowering plants: Eudicots ; Berberidopsidales, Buxales, Crossosomatales, Fabales p.p., Geraniales, Gunnerales, Myrtales p.p., Proteales, Saxifragales, Vitales, Zygophyllales, Clusiaceae Alliance, Passifloraceae Alliance, Dilleniaceae, Huaceae, Picramniaceae, Sabiaceae. Berlin; New York: Springer; 2007.

25. KUO LY, LI FW, CHIOU WL, WANG CN. Conservation of selection on matK following an ancient loss of its flanking trnK intron. Gene. 2008;427:47–54.
26. LANGMEAD B, SALZBERG SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–U54.
27. LASLETT D, CANBACK B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res 1 ed.
28. MESEGUER AS, REE LOBOJM, R., BEERLING, D. J., SANMARTIN I. Integrating Fossils, Phylogenies, and Niche Models into Biogeography to Reveal Ancient Evolutionary History: The Case of (Hypericaceae). Syst Biol. 2015;64:215–32.
29. MESEGUER AS, SANMARTÍN I, MARCUSSEN T, PFEIL BE. 2014. Utility of low-copy nuclear markers in phylogenetic reconstruction of.
30. (L, Hypericaceae.). *Plant Systematics and Evolution*, 300, 1503–1514.
31. MINH BQ, SCHMIDT HA, CHERNOMOR O, SCHREMPF D, VON HAESELER WOODHAMSMD, A., LANFEAR R. IQ–TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020;37:1530–4.
32. MIRARAB S, REAZ R, BAYZID MS, ZIMMERMANN T, SWENSON, K. M., WARNOW T. ASTRAL: genome-scale coalescent-based species tree estimation. Bioinformatics. 2014;30:i541–8.
33. NORMAN K. then came molecular phylogenetics—Reactions to a monographic study of Hypericum (Hypericaceae). Phytotaxa. 2016;255:181–98.
34. NURK NM, BLATTNER FR. Cladistic analysis of morphological characters in Hypericum (Hypericaceae). Taxon. 2010;59:1495–507.
35. NÜRK NM, CROCKETT SL. Morphological and Phytochemical Diversity among Hypericum Species of the Mediterranean Basin. Med Aromatic Plant Sci Biotechnol. 2011;5:14–28.
36. NÜRK NM, CARINE MADRIÑÁNS, M. A., CHASE, M. W., BLATTNER FR. Molecular phylogenetics and morphological evolution of St. John's wort (Hypericum; Hypericaceae). Mol Phylogenet Evol. 2013;66:1–16.
37. NÜRK NM, URIBE-CONVERS S, TANK GEHRKEB, D. C., BLATTNER FR. Oligocene niche shift, Miocene diversification – cold tolerance and accelerated speciation rates in the St. John's Worts (Hypericum, Hypericaceae). BMC Evol Biol. 2015;15:80.
38. ODAHARA M, INOUYE T, FUJITA T, ENDO S, SHIKANAI T. Evidence for the involvement of plastid-encoded recombinase RecA in the maintenance of plastid genome stability in Marchantia polymorpha. Plant Cell Physiol. 2015;56:97–108.
39. PARK S, AN B, PARK S. Recurrent gene duplication in the angiosperm tribe Delphinieae (Ranunculaceae) inferred from intracellular gene transfer events and heteroplasmic mutations in the plastid matK gene. Sci Rep. 2017;7:10478.
40. PARK S, JANSEN, R. K., PARK S. Complete plastome sequence of Thalictrum coreanum and transfer of the rpl32 gene to the nucleus. BMC Plant Biol. 2015;15:40.

41. PARK SJ, KIM KJ. Molecular phylogeny of the genus Hypericum (Hypericaceae) from Korea and Japan: evidence from nuclear rDNA ITS sequence data. J Plant Biology. 2004;47:366–74.
42. POND SLK, MUSE FROSTSDW, WEAVER SV, LEK S, M., MURRELL, B., WERTHEIM, J. O., SMITH MD. HyPhy 2.5—A Customizable Platform for Evolutionary Hypothesis Testing Using Phylogenies. Mol Biol Evol. 2020;37:295–9.
43. ROBSON NKB. Studies in the genus Hypericum L. (Guttiferae).1. Infrageneric classification. Bull Br Mus (Nat Hist) Bot. 1977;5:293–355.
44. ROBSON NKB. Studies in the genus Hypericum L. (Guttiferae). London: British Museum, Natural History; 1981a.
45. ROBSON NKB. Studies in the genus Hypericum L. (Guttiferae).2. Characters of the genus. Bull Br Mus (Nat Hist) Bot. 1981b;8:55–226.
46. ROBSON NKB. Studies in the genus Hypericum L. (Guttiferae). 3. Sections 1. Campylosporus to 6a. Umbraculoides. Bull Br Mus (Nat Hist) Bot. 1985;12:163–325.
47. ROBSON NKB. Studies in the genus of Hypericum l. (Guttiferae). Bull Br Mus (Nat Hist) Bot. 1990;20:1–151.
48. ROBSON NKB. Studies in the genus Hypericum L. (Clusiaceae). Section 9. Hypericum sensu lato (part 3): subsection 1. Hypericum series 2. Senanensia, subsection 2. Erecta and section 9b. Graveolentia. Syst Biodivers. 2006;4:19–98.
49. ROBSON NKB. Studies in the genus Hypericum L (Hypericaceae) 5(2). Sections 17. Hirtella to 19. Coridium. Phytotaxa. 2010a;4:127–258.
50. ROBSON NKB. Studies in the genus Hypericum L. (Hypericaceae) 5(1). Sections 10. Olympia to 15/16. Crossophyllum. Phytotaxa. 2010b;4:5–126.
51. ROBSON NKB. Studies in the genus Hypericum L. (Hypericaceae) 9. Addenda, corrigenda, keys, lists and general discussion. Phytotaxa. 2012;72:1–111.
52. RONQUIST F, TESLENKO M, VAN DER MARK P, AYRES DL, DARLING A, HÖHNA S, LARGET, B., LIU, L., SUCHARD, M. A., HUELSENBECK JP. MrBayes 3.3.7a. Bayesian Inference of Phylogeny. MrBayes Development Team; 2020.
53. RUHFEL BR, BITTRICH V, GUSTAFSSON BOVECP, PHILBRICK MHG, XI CTRUTISHAUSERR, Z. X., DAVIS CC. Phylogeny of the Clusioid Clade (Malpighiales): Evidence from the Plastid and Mitochondrial Genomes. Am J Bot. 2011;98:306–25.
54. RUHFEL BR, PHILBRICK BOVECP, C. T., DAVIS CC. Dispersal largely explains the Gondwanan distribution of the ancient tropical clusioid plant clade. Am J Bot. 2016;103:1117–28.
55. SAVOLAINEN V, MORTON CHASEMWHOOTSB, BAYER CMSOLTISDE, BRUIJN CFAYMFDE, A. Y., SULLIVAN, S., QIU YL. 2000. Phylogenetics of flowering plants based on combined analysis of plastid atpB and rbcL gene sequences. *Syst Biol*, 49, 306 – 62.
56. SHINOZAKI K, OHME M, TANAKA M, WAKASUGI T, HAYASHIDA N, MATSUBAYASHI–SHINOZAKI K, ZAITA N, YAMAGUCHI–SHINOZAKI CHUNWONGSEJ, K., SUGIURA M. The complete nucleotide

sequence of the tobacco chloroplast genome: its gene organization and expression. EMBO J. 1986;5:2043–9.

57. SHRESTHA B, GILBERT, L. E., RUHLMAN, T. A., JANSEN RK. Rampant Nuclear Transfer and Substitutions of Plastid Genes in Passiflora. Genome Biol Evol. 2020;12:1313–29.
58. SOLTIS DE, SOLTIS PS, MORT CHASEMW, ALBACH ME, ZANIS DC, FAY MSAVOLAINENVHAHNWHHOOTSB, AXTELL MF, NIXON MSWENSENSMPRINCELMKRESSWJ, K. C., FARRIS JS. Angiosperm phylogeny inferred from 18S rDNA, rbcL, and atpB sequences. Bot J Linn Soc. 2000;133:381–461.
59. STEVENS PF. Clusiaceae-Guttiferae. In: KUBITZKI K, editor. Flowering Plants · Eudicots: Berberidopsidales, Buxales, Crossosomatales, Fabales p.p., Geraniales, Gunnerales, Myrtales p.p., Proteales, Saxifragales, Vitales, Zygophyllales, Clusiaceae Alliance, Passifloraceae Alliance, Dilleniaceae, Huaceae, Picramniaceae, Sabiaceae. Berlin, Heidelberg: Springer Berlin Heidelberg; 2007a. *In*:.
60. STEVENS PF. Hypericaceae. In: KUBITZKI K, editor. Flowering Plants · Eudicots: Berberidopsidales, Buxales, Crossosomatales, Fabales p.p., Geraniales, Gunnerales, Myrtales p.p., Proteales, Saxifragales, Vitales, Zygophyllales, Clusiaceae Alliance, Passifloraceae Alliance, Dilleniaceae, Huaceae, Picramniaceae, Sabiaceae. Berlin, Heidelberg: Springer Berlin Heidelberg; 2007b. *In*:.
61. UEDA M, FUJIMOTO M, ARIMURA S, MURATA J, TSUTSUMI N, KADOWAKI K. Loss of the rpl32 gene from the chloroplast genome and subsequent acquisition of a preexisting transit peptide within the nuclear gene in Populus. Gene. 2007;402:51–6.
62. UEDA M, NISHIKAWA T, FUJIMOTO M, & ETAL. Substitution of the gene for chloroplast RPS16 was assisted by generation of a dual targeting signal. Mol Biol Evol. 2008;25:1566–75.
63. WENG ML, RUHLMAN, T. A., JANSEN RK. Plastid–Nuclear Interaction and Accelerated Coevolution in Plastid Ribosomal Genes in Geraniaceae. Genome Biol Evol. 2016;8:1824–38.
64. WICKE S, SCHNEEWEISS GM, DEPAMPHILIS, C. W., MÜLLER, K. F., QUANDT D. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Mol Biol. 2011;76:273–97.
65. WICKHAM H. ggplot2: Elegant Graphics for Data Analysis. New York: Springer- New York; 2016.
66. WURDACK KJ, DAVIS CC. Malpighiales phylogenetics: Gaining ground on one of the most recalcitrant clades in the angiosperm tree of life. Am J Bot. 2009;96:1551–70.
67. YANG Z. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci (CABIOS). 1997;13:555–6.
68. ZERBINO DR, BIRNEY E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 2008;18:821–9.
69. ZHU A, GUO W, GUPTA, S., FAN, W., MOWER JP. Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates. New Phytol. 2016;209:1747–56.

Figures

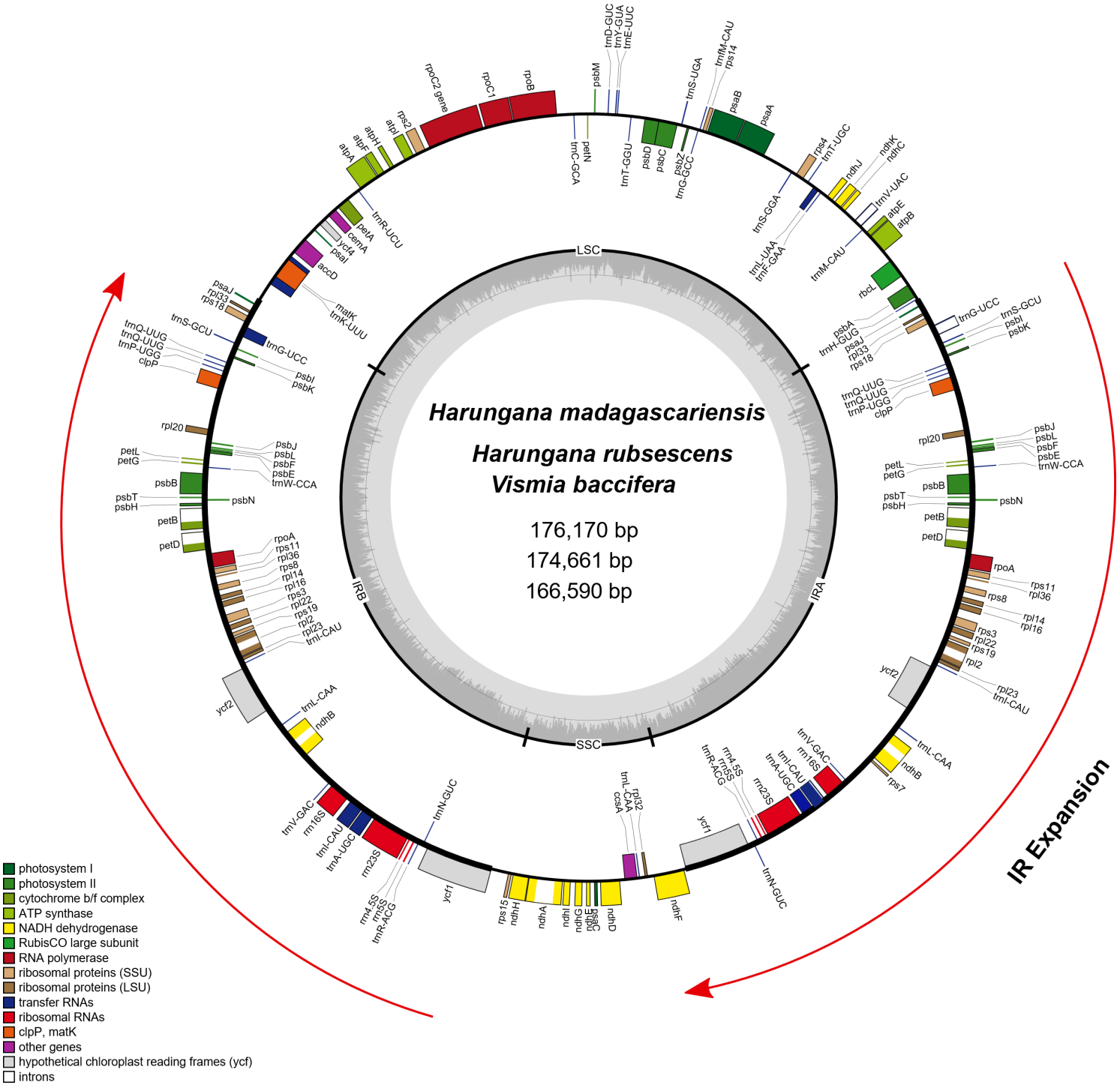

Figure 1

Chloroplast genome map of *Vismia and Harungana* (Vismieae) showing an expanded inverted repeat (IR). The circular plastome structure of *Vismia baccifera* displays a notable expansion of the IR region to ~50kb, leading to extensive duplication of genes beyond the typical IR boundary. Genes are color-coded by functional category, with arrows indicating transcriptional direction. LSC, SSC, and IR regions are labeled, and the inner grey circle indicates GC content.

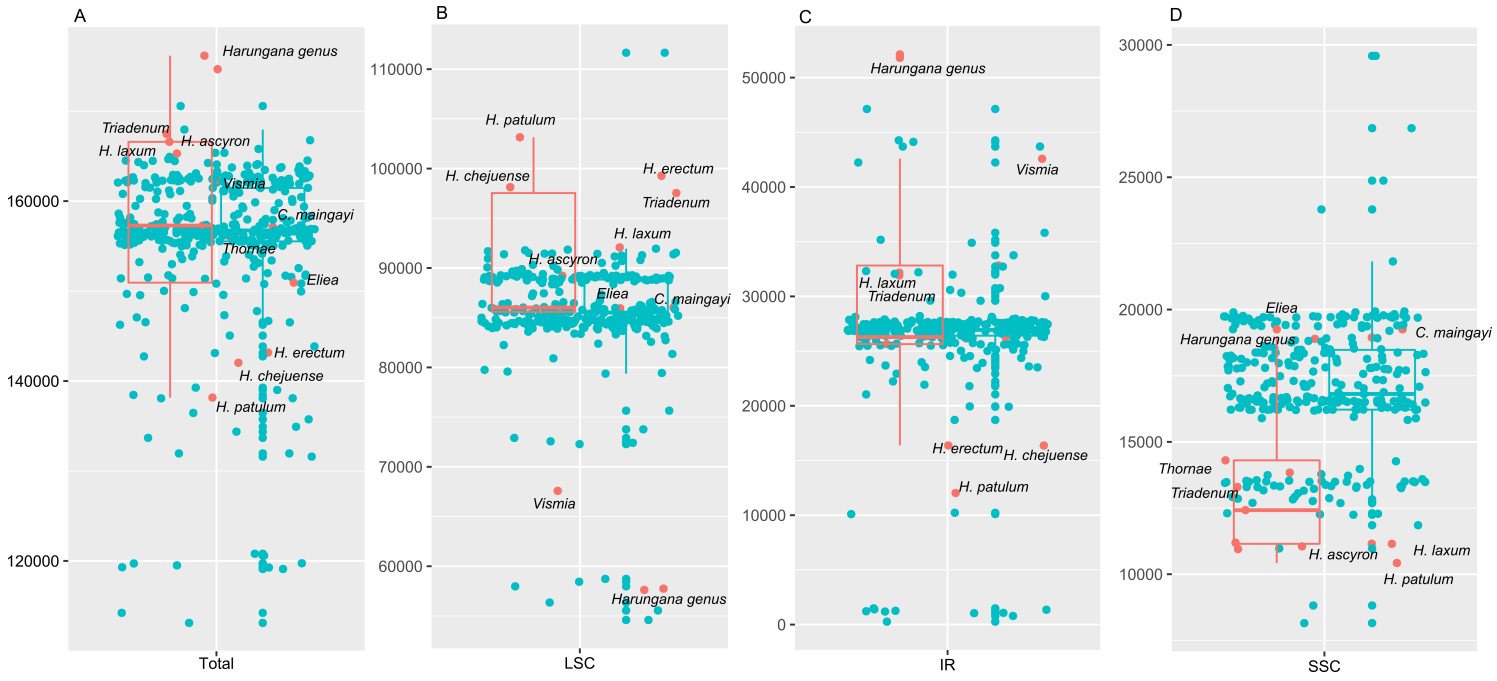

Figure 2

Comparison of plastome sizes across 381 species from Hypericaceae (red) and other families within the order Malpighiales (blue). Each panel represents a different comparative subset or phylogenetic grouping. Red dots and error bars highlight species and genera within Hypericaceae, while blue points represent plastome sizes from other Malpighiales lineages. Notable species and genera are labeled for reference.

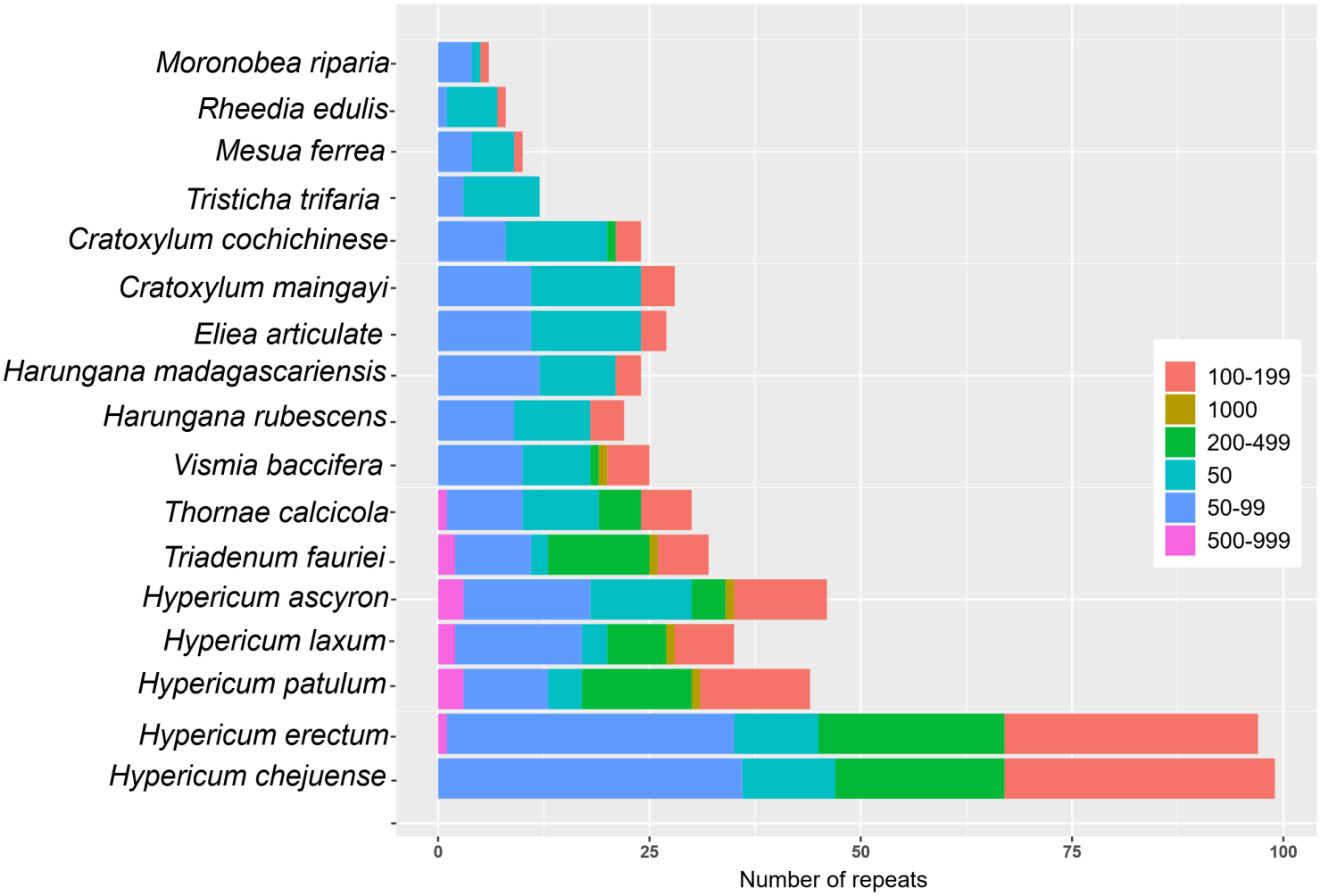

Figure 3

Distribution of tandem repeats in plastomes of Hypericaceae species compared to outgroup species from Podostemaceae, Calophyllaceae, and Clusiaceae. The bar plot shows the number of tandem repeats grouped by repeat length categories. Colors indicate different repeat size ranges, with Hypericaceae species generally exhibiting a higher abundance and broader size range of repeats than outgroups.

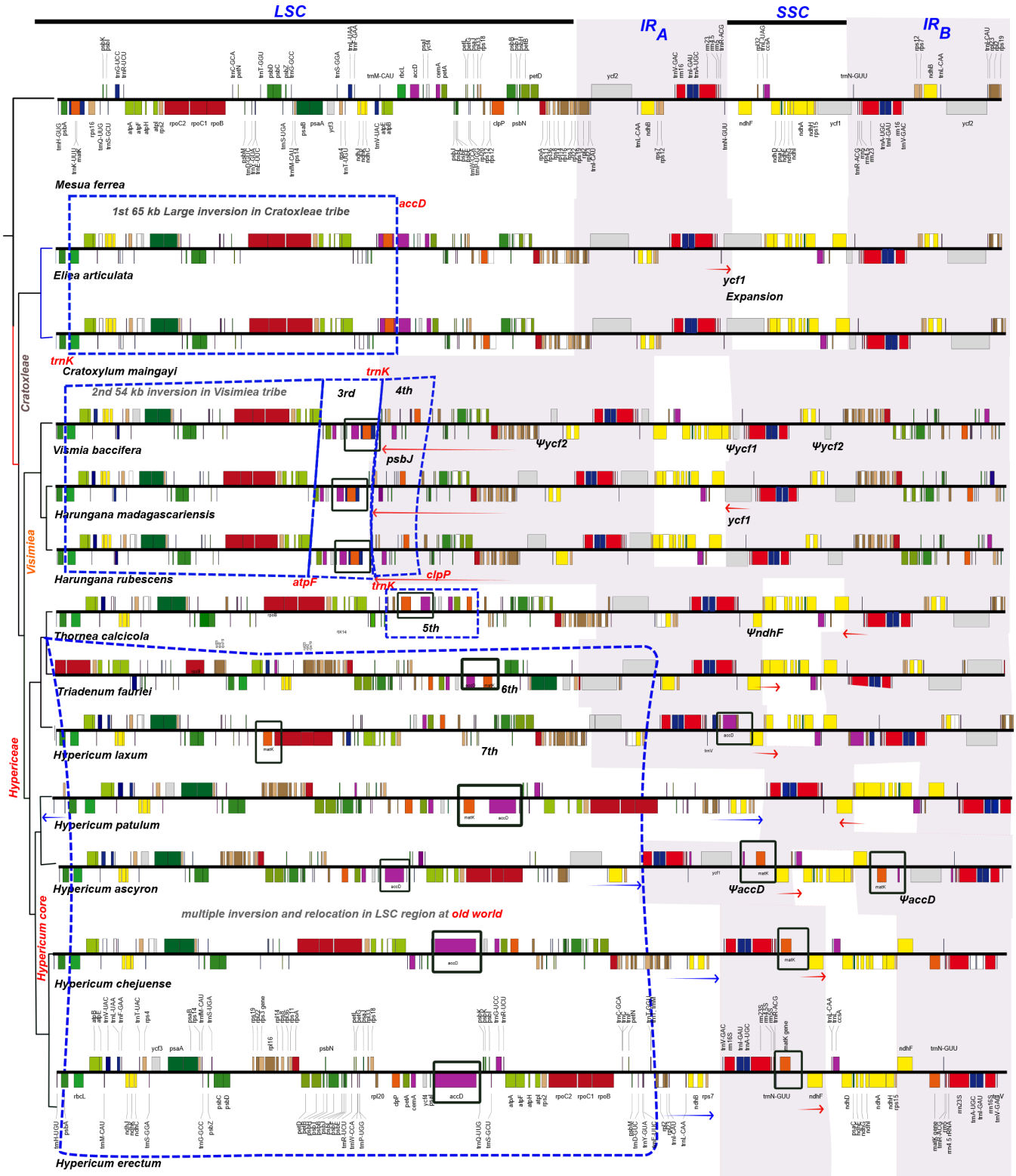

Figure 4

Ancestral reconstruction of plastome structural variation across Hypericaceae tribes compared to *Mesua* (Calophyllaceae) as the outgroup. The diagram illustrates major evolutionary changes, including inverted repeat (IR) boundary shifts, gene relocations, and multiple inversions. Key events are annotated by colored arrows and boxes, with major inversions labeled (e.g., 65 kb and 54 kb) and lineage-specific

rearrangements highlighted. Notable changes in the *accD*, *matK*, *clpP*, and *ycf* gene regions are shown across Cratoxyleae, Vismieae, Hypericeae, and *Hypericum* core lineages.

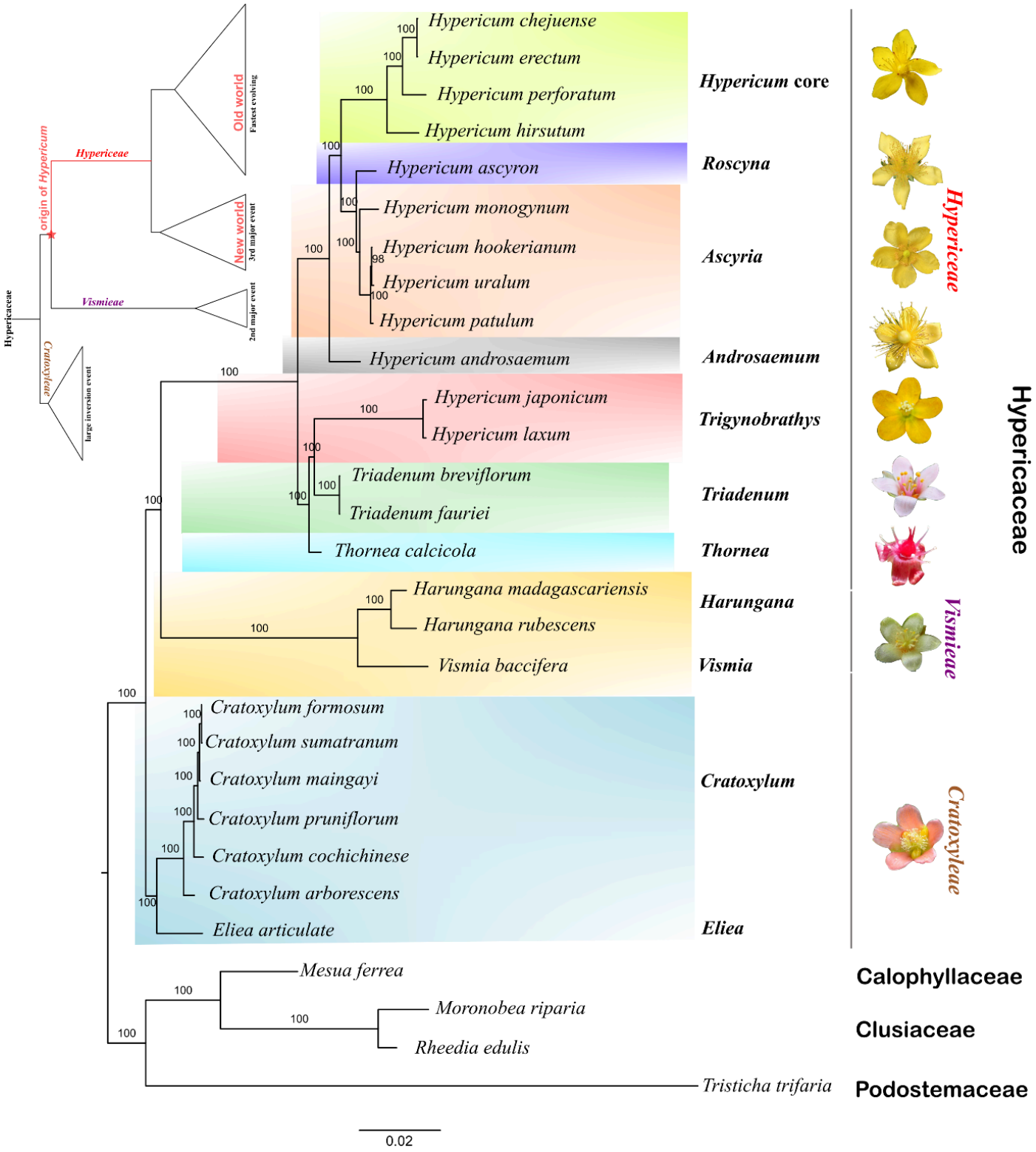

Figure 5

Maximum likelihood phylogeny of Hypericaceae based on 79 plastid protein-coding genes. The tree was reconstructed using concatenated nucleotide sequences of 79 plastid protein-coding genes from 29 Hypericaceae species and one outgroup (Calophyllaceae, Clusiaceae and Podostemaceae). The tree

resolves three major tribes of Hypericaceae: Hypericeae, Vismieae, and Cratoxyleae, each indicated with colored blocks and corresponding floral images. Bootstrap support values from 1,000 replicates are shown at the nodes. Species of *Hypericum* are further classified into “old” and “new” world lineages. The genus *Hypericum* exhibits multiple distinct clades, including *H. ascyron*, which forms a separate lineage within the new world clade. Representative floral images are shown for each tribal group. The tree is rooted with Clusiaceae and other family members, and branch lengths are proportional to nucleotide substitutions per site.

Supplementary Files

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

- SupplementaryTableFigure.pdf