



RESEARCH

Open Access



Plastome evolution and phylogenomic insights into the evolution of *Lysimachia* (Primulaceae: Myrsinoideae)

Tong-Jian Liu^{1,2}, Shu-Yan Zhang³, Lei Wei^{1,4}, Wei Lin³, Hai-Fei Yan^{1,2*} , Gang Hao³ and Xue-Jun Ge^{1,2} 

Abstract

Background *Lysimachia* L., the second largest genus within the subfamily Myrsinoideae of Primulaceae, comprises approximately 250 species worldwide. China is the species diversity center of *Lysimachia*, containing approximately 150 species. Despite advances in the backbone phylogeny of *Lysimachia*, species-level relationships remain poorly understood due to limited genomic information. This study analyzed 50 complete plastomes for 46 *Lysimachia* species. We aimed to identify the plastome structure features and hypervariable loci of *Lysimachia*. Additionally, the phylogenetic relationships and phylogenetic conflict signals in *Lysimachia* were examined.

Results These fifty plastomes within *Lysimachia* had the typical quadripartite structure, with lengths varying from 152,691 to 155,784 bp. Plastome size was positively correlated with IR and intron length. Thirteen highly variable regions in *Lysimachia* plastomes were identified. Additionally, *ndhB*, *petB* and *ycf2* were found to be under positive selection. Plastid ML trees and species tree strongly supported that *L. maritima* as sister to subg. *Palladia* + subg. *Lysimachia* (Christinae clade), while the nrDNA ML tree clearly placed *L. maritima* and subg. *Palladia* as a sister group.

Conclusions The structures of these plastomes of *Lysimachia* were generally conserved, but potential plastid markers and signatures of positive selection were detected. These genomic data provided new insights into the interspecific relationships of *Lysimachia*, including the cytonuclear discordance of the position of *L. maritima*, which may be the result of ghost introgression in the past. Our findings have established a basis for further exploration of the taxonomy, phylogeny and evolutionary history within *Lysimachia*.

Keywords Primulaceae, *Lysimachia*, Plastid genome, Phylogeny, Conflict signature

Background

Lysimachieae is the second largest tribe within the subfamily Myrsinoideae of Primulaceae, mainly distributed in temperate and subtropical climates of the Northern Hemisphere, although it is found worldwide [1]. The tribe is traditionally considered a monophyletic group with six genera, namely *Lysimachia* L., *Anagallis* L., *Trientalis* L., *Glaux* L., *Asterolinon* Hoffmanns. & Link, and *Pelletiera* A. St.-Hil. These genera were characterized by the pattern of capsule dehiscence, number of corolla lobes, and corolla color [2, 3]. While the morphological distinctiveness of the genera within the tribe is relatively easy to recognize, the morphological delimitation among the

*Correspondence:

Hai-Fei Yan

yanhaifei@scbg.ac.cn

¹ Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, China

² South China National Botanical Garden, Guangzhou 510650, China

³ College of Life Sciences, South China Agricultural University, Guangzhou, China

⁴ University of Chinese Academy of Sciences, Beijing, China



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

genera is not clear in nature. *Lysimachia* s.s. is the largest and most morphologically diverse genus in the tribe, with approximately 180 species, and is segregated into six subgenera: subg. *Idiophyton*, subg. *Heterostylandra*, subg. *Lysimachia*, subg. *Lysimachiopsis*, subg. *Naumburgia* and subg. *Palladia* [4–6]. However, the monophyly of *Lysimachia* s.s. has been questioned based on morphology and molecular data [2, 7–10]. For example, *Anagallis arvensis* L. is strikingly similar to *L. nemorum* L., *L. serpyllifolia* Schreb. and a few other *Lysimachia* species, being distinguished solely by color of the corolla and mode of capsule dehiscence [7, 9, 11]. Molecular phylogenetic results confirmed the above finding and even found that these satellite genera (*Anagallis*, *Trientalis*, *Glaux*, *Asterolinon*, and *Pelletiera*) were all embedded within *Lysimachia* s.s. [2, 7, 8, 10, 12–14]. Consequently, the most recent treatment of the tribe synonymized all of these genera with *Lysimachia*, and formed a newly circumscribed *Lysimachia* s.l. with approximately 250 species [11, 15]. Previous molecular phylogenetic studies further detected at least 11 main clades within *Lysimachia* s.l. [2, 8, 10, 16], which were broadly consistent with the former subgenera of *Lysimachia* s.s. and previously described genera, and recognized four new clades from subg. *Lysimachia* of *Lysimachia* s.s. (i.e., Christinae clade, Vulgaris clade, Nemorum clade and Andina clade) [10]. The newly circumscribed genus *Lysimachia* were used in this study.

China is widely recognized as the species diversity center of *Lysimachia*, with eight clades/subgenera (i.e., subg. *Idiophyton*, Christinae clade of subg. *Lysimachia*, Vulgaris clade of subg. *Lysimachia*, subg. *Palladia*, subg. *Naumburgia*, *Glaux* clade, *Trientalis* clade, and *Anagallis* clade) [5, 10]. The first three subgenera all exceed 35 species and have undergone rapid speciation since the Middle Miocene [10]. However, the species-level relationships within Chinese *Lysimachia* remain controversial. For instance, *Lysimachia maritima* (L.) Galasso, Banfi & Soldano (= *Glaux maritima* L.) was recovered as sister to subg. *Palladia* + subg. *Lysimachia* (Christinae clade) with weak support based on 10 plastid markers and one ITS [10]. If only based on ITS data, *L. maritima* was recovered as sister to subg. *Palladia* [12] or subg. *Lysimachia* (Christinae clade) [8]. These uncertainties in phylogenetic relationships within *Lysimachia* are likely due to limited genetic markers and/or taxon sampling employed in previous studies, rapid evolution history, incomplete lineage sorting (ILS), and hybridization. Additionally, the magnitude of infraspecific variation in many widespread species (e.g. *L. fortunei* Maxim.) or the status of taxonomic entities in some taxa (e.g. *L. coreana* Nakai) lack detailed studies.

The plastid is an essential organelle in plant cells, playing a crucial role in plant growth and development

[17]. The majority of complete plastomes in plants have a typical tetrad structure, consisting of a large single copy (LSC), a small single copy (SSC), and two copies of inverted repeats (IR_A and IR_B) [18]. Due to its conserved genome structure, moderate size, and stable gene content, the plastome is a powerful marker for elucidating complex evolutionary relationships at various taxonomic levels of plants [19–21]. The advancement of high-throughput sequencing technology and the reduced cost of sequencing have enabled the widespread use of plastome phylogeny in plants. However, only seven plastomes of *Lysimachia* have been reported and/or released on GenBank (Table S1) [22–26], leading to a lack of progress in the phylogenomics of *Lysimachia*.

In this study, we assembled the complete plastomes for 43 *Lysimachia* species and combined them with seven available *Lysimachia* plastomes in GenBank. The objectives of this study were to (1) identify the plastome structure variation and features of *Lysimachia*; (2) determine the hypervariable loci for the identification of *Lysimachia*; (3) reconstruct the phylogenetic tree to elucidate *Lysimachia*'s relationships; and (4) examine the phylogenetic conflict signals for some *Lysimachia* species. Ultimately, this study should contribute to a better understanding of the rapid evolution of *Lysimachia* based on plastome data.

Results

Basic characteristics of plastomes in *Lysimachia*

In this study, fifty whole plastomes of *Lysimachia* were analyzed (Table S2). The length of these plastomes ranged from 152,691 bp (*L. mauritiana* Lam.) to 155,784 bp (*L. capillipes* Hemsl.). All these plastomes contained the typical quadripartite structure, consisting of a pair of IRs (25,476 bp–26,251 bp) separated by the LSC region (83,676 bp–85,520 bp) and SSC region (17,844 bp–18,203 bp) (Fig. 1a; Table S2). The characteristics of these plastomes are conserved in terms of gene content and GC content (Table S2). The total GC content among these plastomes varied from 36.9% to 37.1%. Each plastome encoded 133 predicted functional genes, of which 19 were duplicated in the IR regions (Table S2). Among the unique genes, there were 80 protein-coding genes, four rRNA genes, and 30 tRNA genes. Within the IR, eight protein-coding genes, seven tRNA genes, and all four rRNA genes were completely duplicated. Additionally, 11 genes possessed a single intron (i.e., *atpF*, *petB*, *petD*, *ndhA*, *ndhB*, *rpoC1*, *rpl2*, *rps12*, *rps16*, *trnA-UGC* and *trnI-GAU*), and two genes (*ycf3* and *clpP*) contained two introns (Table S3).

Across the 50 *Lysimachia* species, the *ycf1* gene overlapped in the SSC-IR_A junction region, ranging from 971 bp (*L. monelli* (L.) U.Manns & Anderb.) to 1,014 bp

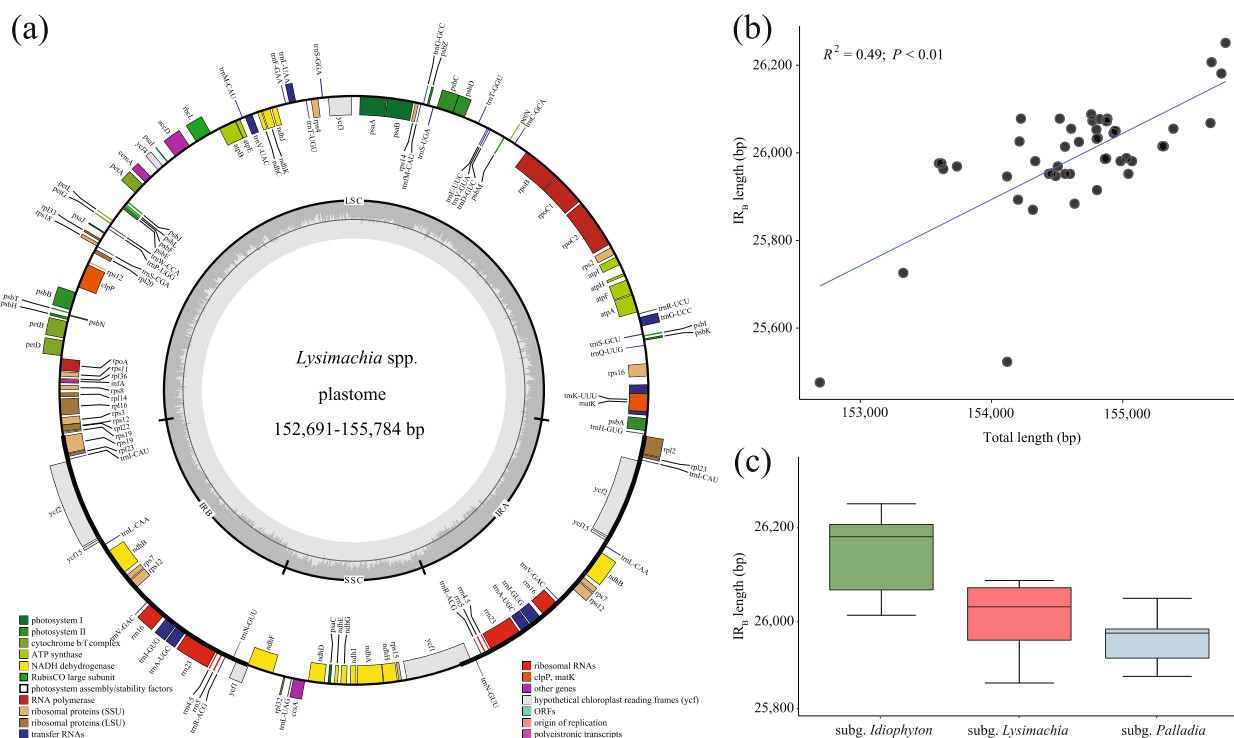


Fig. 1 The plastomes and their expansion/contraction of IR among *Lysimachia* species. **a** plasmome map of *Lysimachia*. The genes belonging to different functional groups are shown in different colors, which are shown on the bottom left. Genes inside and outside of the external circle are transcribed in clockwise and counterclockwise directions, respectively. The inner circle represents the quadripartite structure, with two copies of the inverted repeat (IR_A and IR_B), an LSC, and an SSC region in black with GC content in dark gray and AT content in light gray. **b** Correlation of IR length and *Lysimachia* plastome size. **c** Boxplot of the IR length in three subgenera of *Lysimachia*

(*L. klattiana* Hance) in the IR_A region with the exception of *L. stenosepala* Hemsl., whose *ycf1* gene (4,428 bp) was located entirely in the SSC region (Fig. S1). Additionally, a duplicated pseudogene *ycf1* was present in the IR_B region. The *trnH* gene, with a length of 68–75 bp, was conservatively distributed on the right side of the LSC-IR_A region. A significant correlation was observed between the size of the plastome and the IR length among *Lysimachia* species ($R^2=0.49$, $P=0.01$; Fig. 1b). Besides, a strong relationship between plastome sizes and intron length, while relationship between plastome sizes and IGS length was not significant (Fig. S2). The species were then divided into three subgenera for IR region analysis (Fig. 1c). Subg. *Palladia* was found to have a relatively small IR size (Fig. 1c), with the *rps19* gene of 150 bp entirely embedded into the LSC region, with a gap ranging from 51 bp (*L. pentapetala* Bunge) to 149 bp (*L. auriculata* Hemsl.) (Fig. S1). In contrast, the LSC-IR_B regions of subg. *Idiophyton*, subg. *Lysimachia* (i.e., clade I, referred to as the “Christinae clade” in text; see the phylogenetic results below) and subg. *Lysimachia* (i.e., clade IV, referred to as the “Vulgaris clade” in the text; see the phylogenetic results below) were located within the *rps19*

gene (237 bp), resulting in the presence of *rps19* duplication in the IR_A region. The IR_B/SSC boundary was relatively conserved in all *Lysimachia* taxa, with the *ndhF* gene crossing over to the IR_B region (Fig. S1).

Comparative genome analysis and mutation hotspot identification

No gene rearrangements were identified in *Lysimachia* based on Mauve alignment (Fig. S3), indicating that their genome structure was highly conserved. The mVISTA-based identity plot also exhibited a high degree of conservation in genome structure, gene order, and gene content across all *Lysimachia* plastomes (Fig. S4). Variability was higher in the LSC and SSC regions than in the two IR regions. Furthermore, more variable characteristics were observed in the noncoding regions than in the coding regions. Specifically, distinct sequence variations in intergenic regions included *rps16-trnQ-UUG*, *trnG-UCC-trnR-UCU*, *trnY-GUA-trnT-GGU*, *ndhC-trnV-UAC*, *petA-psbJ*, *rps12-rpl32* and *ccsA-ndhD*. The rRNA genes were highly conserved. Several protein-coding genes were highly conserved (e.g., *rpoB*, *psbD*, *psaA*,

psaB, *psbC*, *psbB* and *rpl23*), while others (e.g., *trnR*, *trnH* and *ycf1*) had slightly more variation.

Sliding window analysis across all *Lysimachia* plastomes revealed that the P_i values in the IR regions were more conserved than those in the single copy (SC) regions (Fig. 2). The P_i values ranged from 0 to 0.0464, with an average of 0.0113. The average P_i values of the LSC, SSC, and IR regions were 0.0144, 0.0210, and 0.0043, respectively. Highly variable hotspots ($P_i > 0.03$) were identified in *matK*, *rps16*, *petN*, *petN-psbM*, *psbM-trnD-GUC*, *accD*, *rpl36*, *rps3-rpl22*, *rpl22-rps19*, *ndhF-rpl32*, *ndhD*, *ndhD-psaC* and *ycf1*. Nucleotide diversity among the four clades (with more than two species) was also analyzed. In subg. *Palladia* (clade II), their P_i values varied from 0 to 0.0320, with an average value of 0.0063 (Fig. S5a). Intergenic regions (i.e., *trnK^{UUU}-rps16*, *petN-psbM* and *trnI^{UGU}-trnF^{GAA}*) and coding regions (i.e., *petA*, *rpl32* and *ycf1*) had relatively high sequence divergence. The average nucleotide diversity (P_i) of subg. *Idiophyton* (clade VI) was 0.0049 with a range of 0 to 0.0280, whereby their LSC, SSC, and IR regions had average P_i values of 0.0084, 0.0064, and 0.0014, respectively (Fig. S5b). For the Christinae clade of subg. *Lysimachia* (clade I), *rps3-rpl22* exhibited the greatest variation, with

a maximum value of 0.2757 (Fig. S5c). Other relatively highly variable regions (i.e., *trnK-UUU-rps16*, *rps16-trnQ-UUG*, *petN-psbM*, *accD*, *psab-petL*, *ndhF-rpl32*, *ndhD* and *ycf1*) were identified. The P_i values of the Vulgaris clade of subg. *Lysimachia* (clade IV) had a range of 0–0.0427, with an average value of 0.0089 (Fig. S5d). There were six highly variable regions ($P_i > 0.03$), namely, *trnH-GUG*, *trnS-GCU-trnR-UCU*, *rpl16*, *ndhD*, *psaC* and *ycf1*.

Selective pressure analysis

The ratio of nonsynonymous (d_N) to synonymous (d_S) mutations is a powerful tool in selection pressure analysis. A d_N/d_S (ω) ratio less than 1 signifies negative selection, while a ω ratio greater than 1 indicates positive selection. Protein-coding genes shorter than 300 bp were filtered out from this analysis. In *Lysimachia* (excluding *L. monelli* and *L. europaea* (L.) U. Manns & Anderb.), 52 protein-coding genes were identified, and the results showed that d_S values were greater than d_N values in most genes, indicating purifying selection (Fig. 3). However, the *ndhB* gene was under positive selection with a ratio of 1.6956 ($\omega > 1$) (Fig. 3).

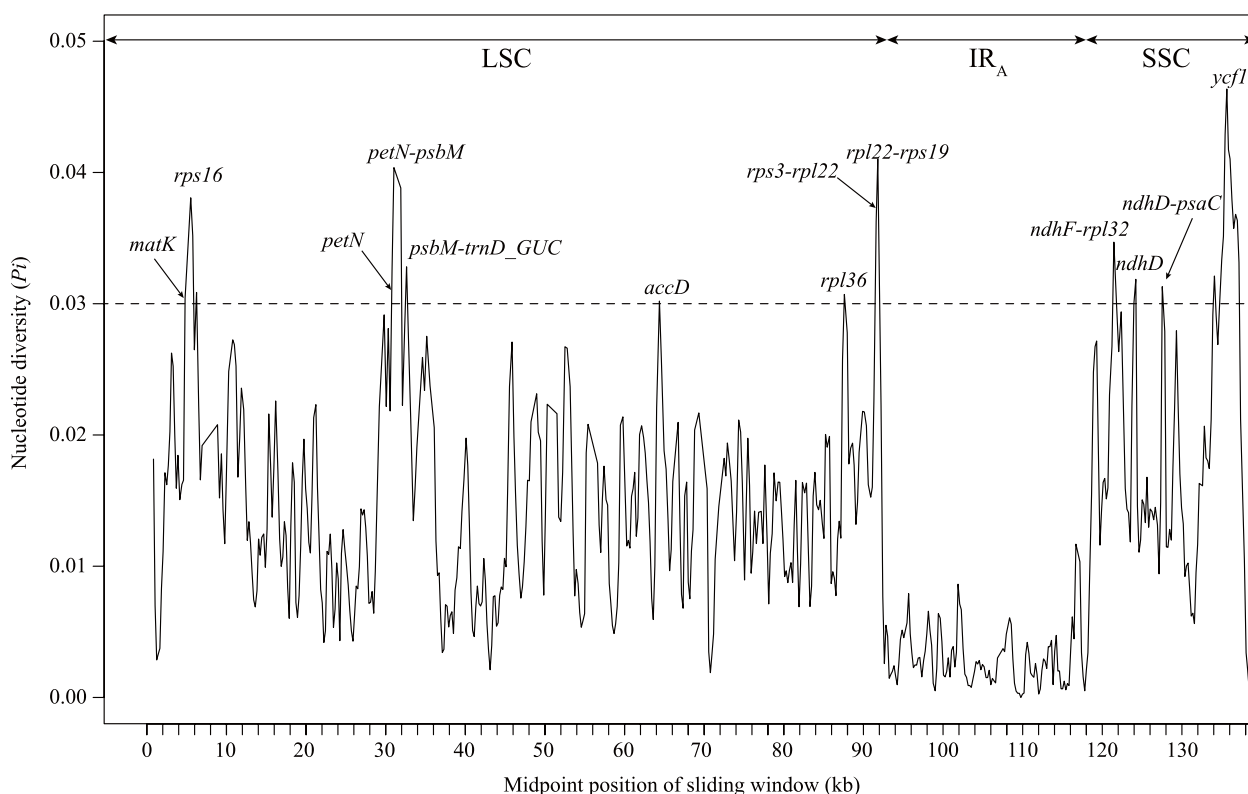


Fig. 2 Comparison of nucleotide diversity (P_i) values in *Lysimachia* plastomes. The X-axis shows the position of the midpoint of sliding windows (kb), while the Y-axis indicates the nucleotide diversity of each window

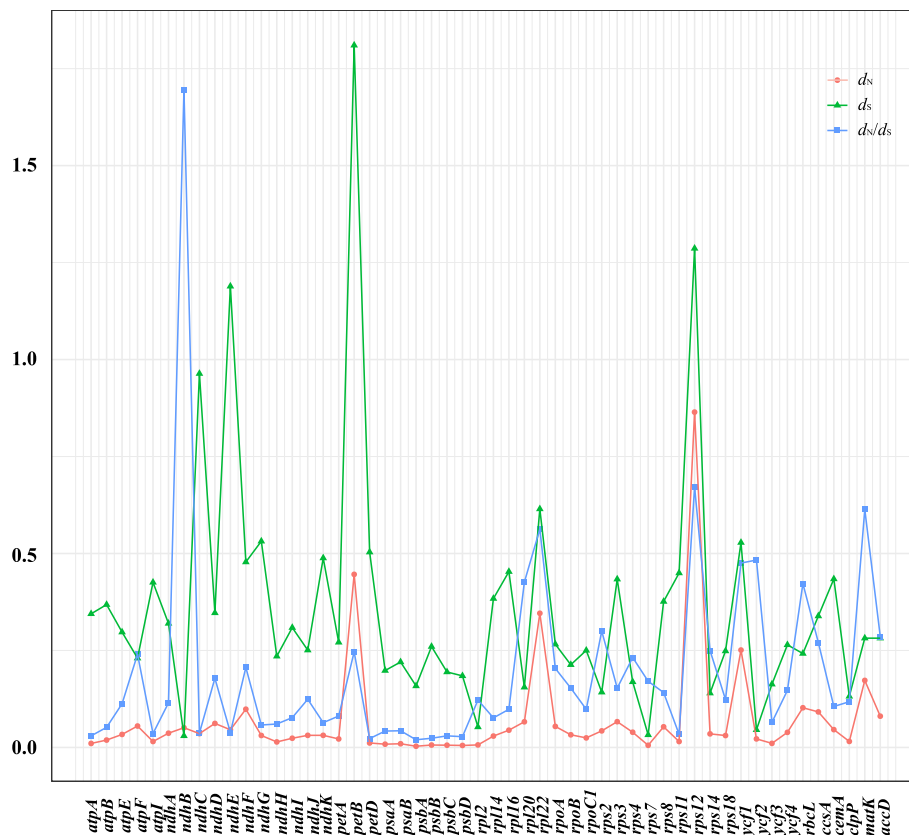


Fig. 3 Comparisons of d_N , d_S , and d_N/d_S of the protein-coding genes in *Lysimachia*

To further understand the adaptive evolution of *Lysimachia*, this analysis was applied to three main clades. The subg. *Lysimachia* (clade IV, Vulgaris clade) was excluded due to its few species (*L. vulgaris* L. and *L. davurica* Ledeb.). A total of 53 genes were used in subg. *Palladia*, displaying d_N values of 0–0.6511 and d_S values of 0.0104–2.1268 (Fig. S6a). The *petB* gene was the only one to present positive selection with a ω ratio of 1.0043. In subg. *Idiophyton*, the ω ratio of 49 protein-coding genes was less than 1, indicating purifying selection on these genes (Fig. S6b). The *ycf2* gene had the highest ω ratio of 0.9241 in subg. *Idiophyton*. Fifty-one shared protein-coding genes were analyzed in subg. *Lysimachia* (clade I, Christinae clade), and the results demonstrated that only the *ycf2* gene experienced positive selection with a ω ratio of 1.3690, while the other genes were under purifying selection (Fig. S6c).

Phylogenetic analysis based on plastomes and nuclear rDNA

In this study, the plastome sequences of 50 *Lysimachia* species were used for phylogenetic relationship reconstruction. Maximum likelihood topologies inferred from the concatenated datasets of whole plastome loci,

protein-coding genes (PCGs), intergenic spacers (IGS), and introns (T1-T4; Fig. 4a; Figs. S7-9) were largely similar, with only a few short internal branches having low bootstrap supports (e.g., the positions of *L. rubiginosa* Hemsl., *L. deltoidea* var. *cinerascens* Franch.) (Figs. S7-9). The MQSST species tree inferred by ASTRAL had a convergent topology compared to the ML trees (Fig. 4a; Fig. S10).

Lysimachia taxa were recovered as monophyletic with full support, and eight major clades in all four plastid ML trees and the ASTRAL species tree were fully supported (BS=100; LPP=1) (Fig. 4a; Fig. S10). Clade VIII contained *L. europaea* (belonging to the former genus *Trientalis*), which was resolved as sister to the remaining *Lysimachia* taxa, with clade VII (*L. monelli*, belonging to the former genus *Anagallis*) sister to clades I-VI. All datasets strongly supported subg. *Idiophyton* was formed as monophyletic group (clade VI), positioned as sister to clades I-V. Clades V (*L. coreana*) and IV (containing *L. davurica* and *L. vulgaris* both belong to subg. *Lysimachia*; Vulgaris clade) was placed as a successive sister to clades I-III. Clade I contained the remaining taxa of subg. *Lysimachia* (Christinae clade). Clade II contained the all species from subg. *Palladia*. Two individuals of *L. maritima*

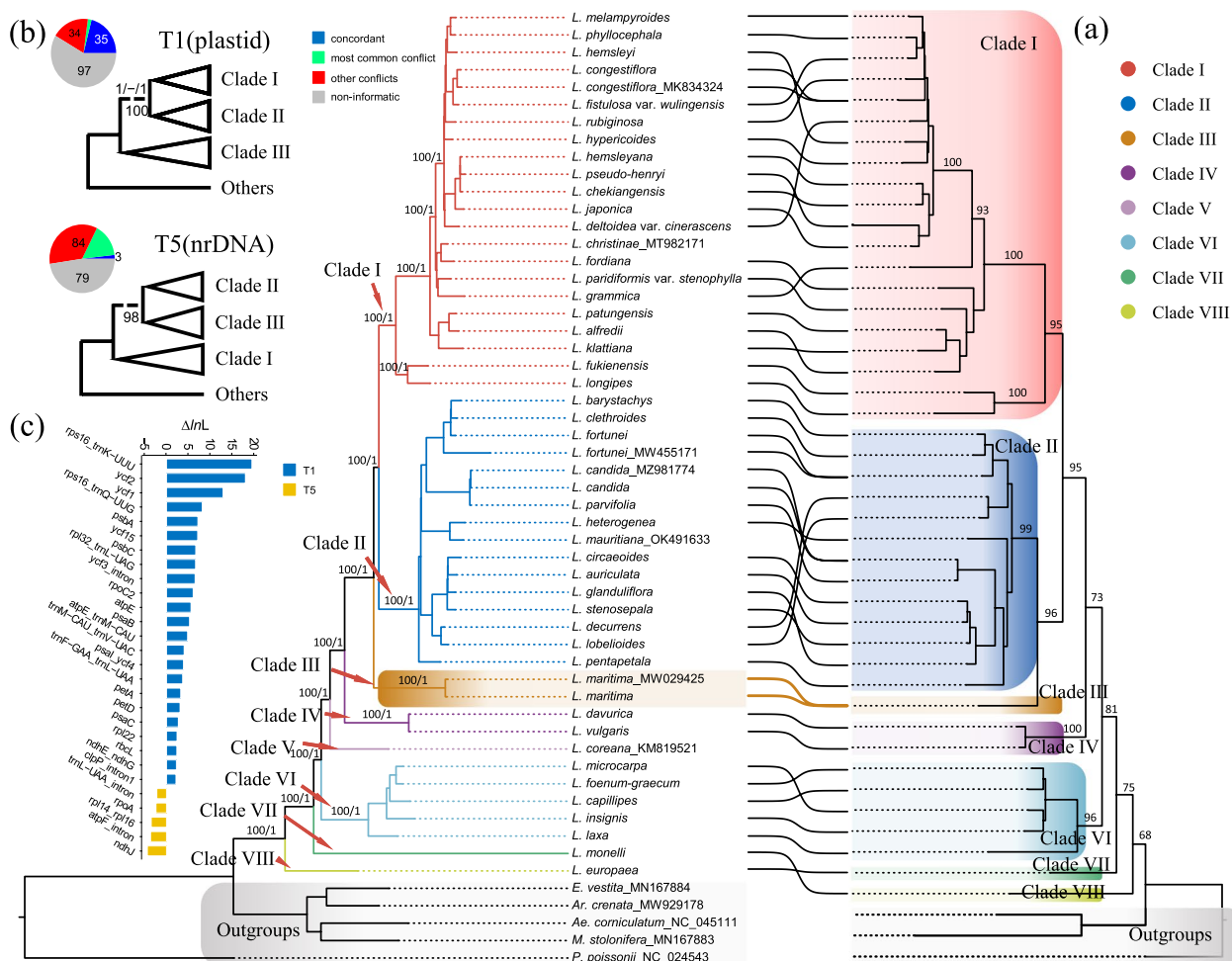


Fig. 4 Conflict signals between plastid and nrDNA phylogenies in *Lysimachia*. **a** The tanglegram of *Lysimachia* shows cytonuclear conflicts between the whole plastid ML tree (T1; left) and nrDNA ML tree (right). Eight major clades within *Lysimachia* are indicated by different colors, and the tangles among the group of *L. maritima* are highlighted. **b** Two hypotheses between two constraint phylogenies according to conflicting cytonuclear topologies using whole plastid datasets and the ratio of gene tree supports, conflicts and non-information. The focal branches we tested in two constraint topologies are shown with dotted lines. The bootstrap supports are shown above the focal branch, and three quartet metrics from quartet sampling analysis are shown below the focal branch. The colors in the legend indicate concordant, most common conflict, other conflicts, and non-informatics, respectively. **c** The difference in log-likelihoods ($\Delta \ln L$) of two hypotheses (T1 and T5) for each locus calculated in RAxML. The positive values of $\Delta \ln L$ indicate T1-supported phylogenetic signals, while negative values support the T2 hypothesis. Only the genes with significant phylogenetic signals ($\Delta \ln L > 2$) are shown on the bar plot. $\Delta \ln L$ values of T1 and T5 were showed with dark blue and yellow, respectively

clustered into clade III (belonging to the former monotypic genus *Glaux*). These plastid datasets strongly supported a sister relationship between clade I and clade II. Our results demonstrated that subg. *Palladia* (clade II) and subg. *Idiophyton* (clade VI) was monophyletic, while subg. *Lysimachia* (including clade I, clade IV and clade V) was paraphyletic (Fig. 4a; Fig. S10).

The nuclear rDNA (nrDNA) topology was largely congruent with that of trees based on plastomes (Fig. 4a). Upon comparison of phylogenetic trees, we identified several conflicting signals, which were mainly concentrated within clade I and II. These conflicting signals were

unstable, as they occur at positions with short branch lengths and low supports (Fig. 4a; Fig. S7-10). However, we observed a significant difference in the position of *L. maritima* between the two phylogenetic trees. The nrDNA phylogeny strongly supported (BS=96) a sister relationship of clade III (*L. maritima*) and clade II, generating an alternative hypothesis (Fig. 4a). To further investigate this positional conflict of clade III in *Lysimachia*, we tested the topology of five ML trees (T1–T5) using IQ-TREE v2.2.0. T5 was the constrained plastid tree, conforming to the alternative hypothesis of Clade III in nrDNA phylogeny (Fig. 4a). The AU, wKH and wSH tests

revealed that the whole plastid tree (T1), the PCG tree (T2) and the IGS tree (T3) all passed these tests, whereas the intron (T4) and constrained tree (T5) were statistically rejected (Table S4).

Incongruence assessment and phylogenetic signal heterogeneity

We identified incongruences in the phylogeny of the optimal plastid tree (T1) using quartet sampling and phyparts at both internal branch and terminal (rogue taxa) levels. Our quartet sampling revealed an average the quartet informativeness (QI) value of 0.95 for all branches, with 84.6% of them strongly supported (the quartet concordance (QC) > 0.2; Fig. S11). Low quartet concordance (QC < 0.2) was observed for short branches with low support values (BS and LPPs; Figs. S7–11). However, one group, containing *L. fortunei*, *L. barystachys* Bunge and *L. clethroides* Duby, showed lower quartet concordance (QC < 0.2) at long branches with high support (BS = 100; LPP = 1) (Figs. S7–11). Our quartet sampling analysis fully or strongly supports the relationships of Clade III (*L. maritima*) as sister to Clades I and II (Fig. S11). The phyparts analysis also showed that all eight clades (I–VIII) had the most concordant support at the gene tree level (Fig. S12), although two dominant alternative topologies (containing ca. 50% conflict gene trees; Fig. S12) were found at the ancestor node of clades I and II. Notably, 35 gene trees supported the relationship of Clade III (*L. maritima*) sister to Clades I and II, although 34 gene trees disagreed with this pattern. Therefore, both quartet sampling and phyparts analyses support Clade III (*L. maritima*) as a sister to Clades I and II.

The optimal plastid tree (T1) was selected and combined with the constrained tree (T5) to further test the relationship among clades I–III. We extracted 156 loci and evaluated the support of each locus for two topologies (T1 and T5). T1 was strongly supported by 23 loci (e.g., *rps16_trnK-UUU*, *ycf2*, *ycf1*, *rps16_trnQ-UUG*, *psbA*, *ycf15*, *psbC*, *rpl32_trnL-UAG*, *ycf3_intron*, *rpoC2*, *atpE*, *psaB*, *atpE_trnM-CAU*, *trnM-CAU_trnV-UAC*, *psaI_ycf4*, *trnF-GAA_trnL-UAA*, *petA*, *petD*, *psaC*, *rpl22*, *rbcL*, *ndhE_ndhG* and *clpP_intron1*; their absolute $\Delta \ln L > 2$). T5 was supported by only five loci (e.g., *trnL-UAA_intron*, *rpoA*, *rpl14_rpl16*, *atpF_intron* and *ndhJ*; their absolute $\Delta \ln L > 2$) (Figs. 4b, c; Table S5). To assess whether a small number of genes dominate the phylogenetic estimation, we removed the top five loci with the largest absolute $\Delta \ln L$ from the concatenated whole plastome supermatrix. The phylogenetic tree inferred without the top five plastid loci was concordant with the original plastid tree, indicating the robustness of the phylogenetic relationship among the three clades (clades I–III).

Discussion

Plastome structure variation and selective evolution in *Lysimachia*

The content and structure of genomes are closely linked to speciation and the ability of species to adapt to changing environments. In plants, the plastome plays a crucial role in growth and development, as it encodes key proteins involved in photosynthesis and other metabolic processes. However, the relationship between the evolutionary patterns of the plastome of *Lysimachia* and its rich species diversity and diverse habitats remains largely unknown due to the lack of sufficient plastomic data. To address this gap, we conducted a study in which we sequenced and assembled a large number of high-quality complete plastomes for this plant group, providing an unprecedented opportunity to conduct in-depth comparative genomics studies and attempt to unravel its evolutionary history.

Previous studies have shown that plastomes in *Lysimachia* possess a typical tetrad structure and are highly conserved in terms of genome size, structure, GC content, and gene composition [23–26]. Our study, based on 50 complete plastomes of *Lysimachia*, partially confirmed this finding. While GC content and gene composition were relatively conserved, plastome size in *Lysimachia* varied from 152,691 bp (*L. mauritiana*) to 155,784 bp (*L. capillipes*). Previous studies have identified three factors that contribute to the variation in plastome size, including expansion/contraction of the IR regions, gene loss and additional gene duplications outside of the IR, and the sizes of introns and intergenic spacer regions [27]. Our findings suggest that expansion/contraction of the IR regions is a major contributor to plastome sequence variation in *Lysimachia*, as supported by the positive relationship between genome size and IR length among species in this study (Fig. 1b). In contrast, gene loss and additional gene duplications outside of the IR do not appear to be major factors in *Lysimachia*, as their gene content and numbers are highly conserved (Table S2). Finally, we found that introns contribute to the variation in plastome size of *Lysimachia*, but intergenic spacer regions (IGS) do not (Fig. S2).

Our study revealed that plastome size has phylogenetic significance, with subg. *Idiophyton* having the largest plastome size, followed by subg. *Lysimachia*, while subg. *Palladia* has the most compact plastome (Fig. 1c; Table S2). Interestingly, *L. mauritiana*, a member of subg. *Palladia*, has the smallest plastome size (152,691 bp). This species is a sea perennial plant living along beaches and maritime rock crevices of East and Southeast Asia, Pacific Islands, and Indian Ocean Islands [1]. Generally, heterotrophic organisms tend to have smaller genome sizes due to the loss of functional genes in the plastome

[27]. However, our findings suggest that plastome size is associated with IR expansion/contraction and intron size, rather than overall genome complexity. The relationship between environmental stress factors and IR expansion/contraction and intron size requires further investigation.

To examine the adaptation of organisms to diverse environments, we analyzed whether coding genes are subject to selection using d_N/d_S (ω) ratios. Our results showed that *ndhB* ($\omega=1.6956$) in *Lysimachia*, *petB* ($\omega=1.0043$) in subg. *Palladia*, and *ycf2* ($\omega=1.3690$) in subg. *Lysimachia* (Christinae clade) were under positive selection, while all protein-coding genes in the plastomes of subg. *Idiophyton* were under negative selection with $\omega < 1$. The *ndhB* gene encodes a subunit of NADPH dehydrogenase, which plays a crucial role in photosynthetic electron transport and dark respiration of photosynthesis [28]. The *petB* gene encodes Cytochrome b6 of the Cyt b6/f complex, a subunit of the photosynthetic apparatus [29], while the *ycf2* gene encodes a 2-MD heteromeric AAA-ATPase complex, which is essential for plant viability and ATP production in chloroplasts in darkness [30]. All these positively selected genes are relevant to ecological adaptation to light preference. Most *Lysimachia* species are shade-tolerant herbs, but some taxa, such as *L. mauritiana*, *L. fortunei*, and *L. maritima*, prefer to grow in open areas and are adapted to sunny habitats [1]. These habitat shifts may have facilitated the divergence of the functional genes mentioned above. Although our study had limited samples, the detected signatures of adaptations to light environment shifts provide valuable clues for understanding the evolution of *Lysimachia* plastomes in response to diverse ecological habitats.

Potential DNA markers for *Lysimachia* identification

Plastid markers have been a powerful tool for species identification and phylogenetic inferences in many phylogenetic studies. However, only a few plastid markers have been used in barcoding and phylogenetic inference in *Lysimachia* [2, 8, 10, 12, 31, 32]. In this study, 20 variable regions in *Lysimachia* were identified as potential DNA markers, including *rps16*, *rps16-trnQ-UUG*, *trnG-UCC-trnR-UUC*, *trnY-GUA-trnT-GGU*, *ndhC-trnV-UAC*, *petA-psbJ*, *rps12-rpl32*, *ccsA-ndhD*, *petN-psbM*, *psbM-trnD-GUC*, *accD*, *rpl36*, *rpl22-rps19*, *ndhD*, *ndhF-rpl32*, *ycf1*, *petN*, *matK*, *rps3-rpl22* and *ndhG-ndhI*. Of these, only two molecular markers (*rps16* and *matK*) have been used in previous studies [2, 8, 10, 31, 32]. Additionally, several highly variable regions (i.e., *rps16-trnQ-UUG*, *petN-psbM*, *accD*, *rpl22-rps19*, *ndhF-rpl32*, *ccsA-ndhD* and *ycf1*) were also found in the Myrsinaceae s.str clade in Primulaceae [26], suggesting that these markers can be used as important genetic resources for future studies on the evolution and diversity of Primulaceae. Notably,

several highly variable regions found in this study (i.e., *rps16*, *rps16-trnQ-UUG*, *ccsA-ndhD*, *petA-psbJ*, *psbM-trnD-GUC*, *ndhC*, *matK* and *rps3-rpl22*) were suggested as DNA barcodes in other plant groups, such as Zingiberaceae species, Coffeae alliance, Cinnamomeae, and *Clethra fargesii* Franch. and *Peucedanum* L. [22, 33–36]. Although these aforementioned markers have relatively large variability, the impact of gene-level heterogeneity on the phylogenetic reconstruction of plastids should be considered [37].

Phylogenetic relationships within *Lysimachia*

Lysimachia s.l., newly circumscribed to include all satellite genera (i.e., *Anagallis*, *Trientalis*, *Glaux*, *Asterolinon* and *Pelletiera*) [11, 15], contains approximately 250 species [10]. Previous studies have recovered eleven major clades of *Lysimachia* [2, 8, 10, 12], and its backbone phylogeny has been well resolved except for two nodes (i.e., the positions of the *L. maritima* and *Anagallis* clades; see Yan et al. [10]). In this study, eight major clades were recovered, which was concordant with the results found in several previous studies [2, 8, 10–12]. However, three other major clades were not included in this study. The plastome trees showed that most nodes in each clade were largely improved and strongly supported (Fig. 4a) when compared with those based on a few plastid markers (e.g., [10]). The weakly supported relationships were mainly in clade I and clade II (Figs. S7–10), which were likely caused by limited phylogenetic signals and extensive phylogenetic conflicts due to gene-level heterogeneity in *Lysimachia* plastomes (Figs. S11–12), suggesting that the two clades may have experienced a short radiation period from their most recent common ancestor (MRCA).

In this study, the phylogenetic positions of several *Lysimachia* species were of interest. *Lysimachia coreana*, an endemic taxon in Korea, belongs to clade V. This species was first described by Takenoshin Nakai in 1909 [38]. Morphologically, *L. coreana* is similar to *L. davurica*, as confirmed by palynotaxonomic data [39]. However, recent study has treated *L. coreana* as a synonym of *L. davurica* [40], suggesting that it should be classified under subg. *Lysimachia* (Vulgaris clade, IV clade). Our results, however, revealed that *L. coreana* forms a separate clade (clade V) and does not cluster with *L. davurica* and *L. vulgaris* of subg. *Lysimachia* (Vulgaris clade, clade IV) (Fig. 4). The plastome of *L. coreana* used in this study was obtained from GenBank [25], which made it difficult to determine whether the observed differences were due to the quality of the sequencing data or inherent species differences. If the latter is the case, it is appropriate to consider *L. coreana* as a distinct species.

Furthermore, it is imperative to conduct further investigations on the phylogenetic position of *L. fortunei*. This species, belonging to subgenus *Palladia*, is extensively distributed in central and southeastern China, Vietnam, the Korean peninsula, and Japan [1]. Despite its widespread distribution, there has been little controversy regarding the taxonomic classification of *L. fortunei* since its initial description in 1868 [1]. Morphologically, *L. fortunei* bears the closest resemblance to *Lysimachia chikungensis* Bail., but differs in its creeping rhizomes and relatively large leaves [1]. Unfortunately, we were unable to test the hypothesis of their close relationship due to the absence of *L. chikungensis* accessions in this study. Our findings indicate that two *L. fortunei* accessions did not cluster together (Fig. 4a, Figs. S7-10). This scenario may be attributed to three factors. Firstly, the wide range of habitats and ecological niches occupied by *L. fortunei* may have resulted in significant differentiation at the intraspecific taxonomic level, suggesting the possibility of cryptic lineages within this species. Secondly, misidentification may have occurred, as one accession (MW455171) was obtained directly from GenBank, and we were unable to examine its voucher specimens. Finally, hybridization/introgression is a common occurrence in plants [41], and chloroplast capture is frequently observed during hybridization/introgression events [42]. Therefore, hybridization/introgression may be a plausible explanation for the paraphyly of *L. fortunei* accessions. In conclusion, the taxonomic status of *L. fortunei* warrants further examination through the use of additional samples and genetic markers, such as low-copy nuclear markers, in future studies.

The phylogenetic position of *Lysimachia maritima* and its presumed origin history

Lysimachia maritima (formerly belonging to the monotypic genus *Glaux*; clade III) is undoubtedly a member of *Lysimachia*, as supported by molecular systematics [2, 15, 43]. However, its phylogenetic position remains contentious. Yan et al. [10] recovered *L. maritima* (clade III) as sister to subg. *Palladia* (clade II) + subg. *Lysimachia* (clade I) with weak support, based on ten plastid markers and ITS. When only ITS data was used, *L. maritima* was recovered as sister to either subg. *Palladia* [12] or subg. *Lysimachia* [8]. In the present study, plastid ML trees and species tree strongly supported *L. maritima* (clade III) as a sister to subg. *Palladia* (clade II) + subg. *Lysimachia* (clade I), while the nrDNA ML tree clearly placed *L. maritima* (clade III) and subg. *Palladia* (clade II) as a sister group (Fig. 4a; Figs. S7-10).

Plastid loci have traditionally been concatenated into a single dataset for phylogenetic inference due to the low levels of recombination in organelle genomes; therefore,

their different historical signals that lead to incongruence between nucleotide characters or sequence blocks have long been assumed to be negligible ([44], and references therein). However, recent studies [45–49] have challenged this view, although the sources of conflict within the plastome remain understudied and poorly understood [46, 47, 50], such as heterogeneity in molecular evolution, heteroplasmy, and recombination. Therefore, the cytonuclear discordance detected in this study was first examined within the plastome to better characterize the extent and sources of the conflict signature. Quartet sampling and phyparts analyses both support Clade III (*L. maritima*) as a sister to Clades I and II (Figs. S11-12). Furthermore, 28 of 156 plastid loci had phylogenetic signals at the controversial node (clades I-III), of which only five loci significantly supported ($\Delta\ln L > 2$) the opposite constrained tree (T5) (Fig. 4b, c; Table S5). This suggests that the original plastid relationship (clade III, (clade I, clade II)) is robust. It is plausible that heterogeneity in molecular evolution may have resulted in a few plastid fragments that have generated opposing topologies.

It is well-documented that inconsistency between plastid and nuclear gene trees is a common and widespread phenomenon in plants [51]. Previous studies have suggested that certain evolutionary events, such as gene replication, hybridization, and lineage sorting of ancestral polymorphisms, may be responsible for the incongruent topologies from the plastome and nuclear genome [42, 51, 52]. The pattern of longer internal edges followed by short external edges of the node of *L. maritima* (clade III) and subg. *Palladia* (clade II) in the nrDNA ML tree (Fig. 4a) is consistent with the pattern of a relatively constant rate of diversification [53], coupled with low ILS level due to reduced gene tree discordance and quartet incongruence (Figs. S11-12), suggesting the possibility of hybridization rather than incomplete lineage sorting (ILS) on the early diversification of the group (clades II-III).

Lysimachia maritima undoubtedly occupies an isolated position within *Lysimachia* due to its corolla loss and specialized adaptation to saline habitats. Morphologically, it is similar to *L. mauritiana* of subg. *Palladia* but lacks a corolla [2], suggesting a potential relationship between the two taxa. Our nuclear data strongly indicate that *L. maritima* is closely related to subg. *Palladia*, while the plastome suggests that the species is a more independent lineage. It is hypothesized that the plastome of *L. maritima* is derived from a lineage that is either extinct or has not been sampled (a “ghost lineage”; [54, 55]). Its hybrid origin may have resulted in adaptive introgression, which may have enabled *L. maritima* to acquire its unique morphological characteristics (such as the corolla lost) and habitat adaptation (i.e., typically found on saline soils,

such as beaches, muddy shallows, saline soils, and inland salt marshes). However, only one nuclear gene, nrDNA, was used for the construction of the nuclear gene tree in this study, so we cannot rule out the possibility that systematic errors in nrDNA itself caused this plastid-nuclear discordance, which should be tested further with more nuclear data.

Conclusions

In this study, the plastomes of 50 *Lysimachia* taxa were comparatively analyzed. Generally, the structures of these plastomes were conserved, although contraction and expansion of the IR regions was observed to be associated with plastome sequence variation. Thirteen hotspot regions were identified within *Lysimachia* plastomes, which could potentially serve as plastid markers for the identification of *Lysimachia* species. Additionally, signatures of positive selection were detected in *ndhB*, *petB* and *ycf2* in *Lysimachia*, subg. *Palladia*, and subg. *Lysimachia*, respectively, suggesting that these genes may be involved in adaptation and speciation processes in *Lysimachia*. These genomic data provided new insights into the interspecific relationships of *Lysimachia*, including the identification of a cytonuclear discordance of the position of *L. maritima*, which may be the result of ghost introgression in the past. Our findings have established a basis for further exploration of the taxonomy, phylogeny and evolutionary history within *Lysimachia*.

Materials and methods

Sampling, DNA extraction, and sequencing

A total of 50 *Lysimachia* individuals, representing 46 species, were included in the present study. Forty-three of these *Lysimachia* individuals were newly sampled, with fresh and healthy leaves collected in the field and dried with silica gel for DNA extraction. Voucher specimens were formally identified by Prof. Chi-Ming Hu, Gang Hao and Hai-Fei Yan and deposited at the Herbarium of South China Botanical Garden (IBSC). Additionally, seven *Lysimachia* plastomes were obtained from GenBank (i.e., *L. coreana*, *L. congestiflora* Hemsl., *L. christinae* Hance, *L. maritima*, *L. fortunei*, *L. candida* Lindl., and *L. mauritiana*). Although the sampling ratio of the genus was low (c. 18%), it represented all main clades in China identified by previous studies, except subg. *Naumburgia* [2, 8, 10]. Furthermore, the majority of our sampling (c. 93%) belonged to three highly diversified subgenera of *Lysimachia* s.s. in East Asia [10], namely 24, 14, and 5 species from subgenera *Lysimachia*, *Palladia*, and *Idiophyton*, respectively, each accounting for approximately 32%, 23%, and 8% of the above subgenera. Plastomes of five representatives of Primulaceae (i.e., *Ardisia crenata* Sims., *Myrsine stolonifera* (Koidzumi) E. Walker,

Aegiceras corniculatum (L.) Blanco, *Embelia vestita* Roxb. and *Primula poissonii* Franch.) were downloaded from GenBank and used as outgroups for plastid phylogenetic analyses. For nrDNA phylogenetic inferences, nrDNA sequences of *Ar. japonica* (Thunb.) Blume, *Ae. corniculatum* and *P. veris* L. were obtained from GenBank as outgroups. The sampling details can be found in Table S1.

DNA extraction, sequencing, assembly and annotation

Total genomic DNA was extracted from leaf materials using a modified cetyltrimethylammonium bromide (CTAB) method [56]. Genomic DNA was randomly fragmented into 350-bp fragments for constructing pair-end libraries and sequenced on MGISEQ-2000 (MGI, Shenzhen, China) at the Beijing Genomics Institution (BGI, Wuhan, China). Approximately 2 Gb of genome skimming data were generated for each sample. Paired-end sequence reads were trimmed to remove low-quality reads and adapter sequences using Trimmomatic v0.36 [57] before assembly. Plastomes were assembled using the GetOrangelle v1.7.7.0 [58] with default parameters. The completed assembled plastome sequences were checked and adjusted manually in Geneious v11.0.3 [59] and were annotated by the online annotation program GeSeq v2.03 [60]. A graphical circular plastome map representing *Lysimachia* was drawn using OGDRAW v1.3.1 [61, 62].

To confirm whether phylogenetic conflicts between the plastid and nuclear genomes in *Lysimachia* existed, the complete nrDNAs of 40 taxa were assembled based on the above genome skimming data by GetOrangelle with the parameters set as R=15 and K=21, 45, 65, 85, 105 and annotated and checked in Geneious. The whole nrDNA is composed of the intergenic spacer (IGS), the small-subunit (SSU) ribosomal RNA (rRNA) gene, the internal transcribed spacer 1 (ITS1), the 5.8S rRNA gene, the internal transcribed spacer 2 (ITS2), and the large-subunit (LSU) rRNA gene. The annotated plastomes and nrDNA sequences were deposited in GenBank with accession numbers (Table S1).

Comparative analyses, mutation hotspot identification, and substitution rate estimations

The boundary information between IRs and SSC/LSC was analyzed in Geneious. The relationship between the plastome size and IR length among species was examined using least squares linear regression in R 4.2.2 [63]. The full interspecific sequence divergences of *Lysimachia* plastomes were visualized using the mVISTA program under the Shuffle-LAGAN model [64], with the *L. fortunei* plastome (GenBank accession no.: MW455171) as a reference. The *Lysimachia* plastomes were further

aligned using MAUVE v1.1.3 [65] to verify gene orders and structure rearrangements. To identify mutation hotspots, nucleotide diversity (P_i) for all protein-coding and noncoding regions was calculated with DnaSP v6 [66] using the sliding window method (window length: 500 bp; step size: 200 bp).

The analysis of evolutionary rate was conducted along the phylogenetic tree of *Lysimachia* for each plastid protein-coding gene. The substitution ratio of nonsynonymous (d_N), synonymous (d_S), and ω (d_N/d_S) values of each protein-coding gene was calculated by a Site Model (M0) using EasyCodeML v1.4 [67]. The ω value is an indicator of selective pressure of the protein-coding genes, and $\omega > 1$, $\omega = 1$, and $\omega < 1$ indicate positive, neutral, and negative selection, respectively. The codon alignment used for this analysis was generated by using the MUSCLE (Codons) option [68] in MEGA 11 [69]. The fasta format sequences were converted to "PML" format under the option of the Convert Sequence format in PhyloSuite v1.2.2 [70].

Phylogenetic analyses and conflicting phylogenetic signal test

Fifty-five plastomes were extracted and aligned for all annotated loci, including coding and noncoding regions. Sites with > 80% gaps were trimmed, and loci with alignment lengths greater than 100 bp were retained. Eighty protein-coding genes (PCGs), 92 intergenic spacers (IGSs) and 18 introns were extracted from all plastomes for phylogenetic inferences. Alignments of individual coding or noncoding regions were generated using MAFFT v.7.4 [71] with a default setting. The concatenated alignments for four multigene datasets were generated by using the AMAS package [72]. RAxML v.8 [73] was used for plastid phylogenetic analyses under the GTRGAMMA model with the "autoMRE" option, which generates bootstrap replicates until the support values converge.

The multiple species coalescent model was used to infer phylogenetic relationships, which accounts for genealogic heterogeneity and allows the assessment of ancient hybridization/introgression and ILS [74]. Gene trees of the 190 plastid loci inferred by RAxML with the GTRGAMMA model and autoMRE rapid bootstraps. Poorly supported branches (i.e., bootstrap support (BS) < 50%) in each gene tree were collapsed. Maximum Quartet Support Species Tree (MQSST) analyses were performed under the coalescent model with plastid gene trees in ASTRAL-III v.5.6.3 [75]. Internal branch supports of the species tree were evaluated with local posterior probabilities (LPPs; [76]) by ASTRAL and with the quartet sampling (QS) method based on 1000 replicates per branch [77]. The quartet sampling

developed four metrics to evaluate the quartet concordance (QC), differential (QD), and informativeness (QI) for each branch and fidelity (QF) scores for each terminal in the given phylogeny. Phyparts [78] was used to assess whether gene tree topologies were concordant or conflicting with the species tree. For each node on the species tree, the ratio of concordant and conflicting or no support loci was visualized by a pie chart.

We conducted topology selection analyses with five maximum likelihood (ML) trees using the whole plastid dataset. The weighted Kishino-Hasegawa test (wKH test), weighted Shimodaira-Hasegawa test (wSH test) and Approximately Unbiased test (AU test) were used to examine the significant difference between the five ML tree topologies (the whole plastid tree (T1), PCG tree (T2), IGS tree (T3), intron tree (T4) and constrained plastid tree conforming to the hypothesis of Clade III in nrDNA phylogeny (T5)) in IQ-TREE [79]. The best-fit model of the five datasets according to Bayesian Information Criterion (BIC) was chosen as TVM+I+I+R5. Furthermore, the cytonuclear discordances between the plastome and nrDNA were examined. To detect the conflicting signals of two given hypotheses inside the plastome, we compared the log-likelihood scores between the constrained and unconstrained treatments following the method of Shen et al. [80]. We obtained the constrained tree using the option "-g" in RAxML and computed per locus log-likelihood values for the two trees with the option "-f g" in RAxML. The $\ln L$ differences between two hypotheses for each locus were ranked and plotted by the ggplot2 package [81] in R. An absolute value of $\ln L$ difference greater than 2 ($\Delta \ln L > 2$) was considered a strong signal for the support of the main hypothesis. We also removed the top five loci in the $\Delta \ln L$ ranking list and reconstructed the ML tree with the reduced plastid dataset to test the robustness of the plastid phylogeny.

Abbreviations

bp	Base pair
CDS	Coding sequence
DNA	Deoxyribonucleic acid
Gb	Gigabases
GC	Guanine-cytosine
kb	Kilobase
ML	Maximum likelihood
MRCA	The most recent common ancestor
RNA	Ribonucleic acid

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12870-023-04363-z>.

Additional file 1: Fig. S1. Comparison of LSC, IRs, and SSC junction positions among *Lysimachia* plastomes.

Additional file 2: Fig. S2. Correlation of intron and IGS length and *Lysimachia* plastome size. (a) Correlation of intron length and *Lysimachia* plastome size. (b) Correlation of IGS length and *Lysimachia* plastome size.

Additional file 3: Fig. S3. Mauve alignment of 50 representative *Lysimachia* plastomes.

Additional file 4: Fig. S4. Sequence identity plots of plastomes of *Lysimachia* by mVISTA using *Lysimachia fortunei* as the reference. The top line shows the orientation of genes. A cutoff of 70% identity was used for the plots, and the Y-scale represents the percentage identity ranging from 50 to 100%.

Additional file 5: Fig. S5. Comparison of nucleotide diversity (Pi) values in plastomes of four main *Lysimachia* clades. (a) The subg. *Palladia* (clade III). (b) The subg. *Idiophyton* (clade VI). (c) The subg. *Lysimachia* (clade I). (d) The subg. *Lysimachia* (clade IV).

Additional file 6: Fig. S6. Comparisons of d_N , d_S , and d_N/d_S of the protein-coding genes in three main subg. *Lysimachia* clades. (a) The 53 protein-coding genes in subg. *Palladia*. (b) The 49 protein-coding genes in subg. *Idiophyton*. (c) The 51 protein-coding genes in subg. *Lysimachia* (Christinae clade).

Additional file 7: Fig. S7. The maximum likelihood tree (T2) of *Lysimachia* reconstructed based on PCGs. The bootstrap support values are indicated along the branches.

Additional file 8: Fig. S8. The maximum likelihood tree (T3) of *Lysimachia* reconstructed based on IGS. The bootstrap support values are indicated along the branches.

Additional file 9: Fig. S9. The maximum likelihood tree (T4) of *Lysimachia* reconstructed based on introns. The bootstrap support values are indicated along the branches.

Additional file 10: Fig. S10. The maximum quartet support species tree (MQSST) inferred by ASTRAL based on plastid regions. Local posterior probabilities (LPPs) are indicated along the branches.

Additional file 11: Fig. S11. The results of quartet sampling analysis are based on the ASTRAL species tree using the whole plastid dataset. The three metrics are shown on each branch following the order of QC/QD/QI scores (QC: quartet concordance; QD: quartet differentiation; QI: quartet information). The QI score is a nonlinear elevation metric for quartet supports/conflicts. Heatmap of branches by Quartet Concordance (QC) scores for internal branches: dark green (QC > 0.2), light green (0.2 ≥ QC > 0), and dark orange (QC < -0.05).

Additional file 12: Fig. S12. Pie charts of the ratio of gene support, conflict, and non-informative gene loci on each node of an ASTRAL maximum quartet support species tree (MQSST). The blue sector of the pie chart indicates the gene support for the shown topology. The green sector of the pie chart indicates the most common conflicting topologies. The red sector of the pie chart indicates all other supported conflicting topologies. The gray sector has no support for two conflict testing topologies. The number in parentheses represents the most common conflict (green) rather than all other gene tree conflicts (red).

Additional file 13: Table S1. Samples, vouchers, and GenBank accessions used in this study. **Table S2.** The plastome features of *Lysimachia* taxa.

Table S3. Genes annotated of *Lysimachia* plastomes. **Table S4.** Topology tests for ML trees generated by various plastid datasets and constraint tree with the clade III sister to clade II using whole plastid dataset on RAXML.

Table S5. The difference in log-likelihoods of 156 plastid loci between the two hypotheses T1 and T5.

Acknowledgements

Many thanks to Prof. Chi-Ming Hu for specimen identification. We are grateful to Chenshan Herbarium, Xing Zhong, Xing-Xing Zhu, Yun-Fei Deng, Si-Rong Yi, Pan Li, Wei-Bin Xu, Li-Na Dong, Rong Li, Guo-Sheng He and Yong-Jie Guo for providing plant materials. We also thank Yu-Yin Zhou, Tian-Wen Xiao and Lu Jin for helping with the DNA extraction and data analyses.

Authors' contributions

TJL and HFY conceived this study. SYZ, TJL and LW performed the analyses. WL and TJL conducted the experimental study. SYZ, TJL and HFY wrote the draft manuscript. XJG and GH contributed feedback and editing of the manuscript. All authors read and approved the final manuscript.

Funding

This work was financially supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDB31000000), the National Natural Foundation of China (Grant No. 31870192) and the Biological Resources Programme, Chinese Academy of Sciences (KFJ-BRP-017–104).

Availability of data and materials

All sequences used in this study are available from the National Center for Biotechnology Information (NCBI) (see Table S1).

Declarations

Ethics approval and consent to participate

This study's material collections and experimental research followed the relevant institutional, national, and international guidelines and legislation. No specific permissions or licenses were needed.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 21 February 2023 Accepted: 23 June 2023

Published online: 14 July 2023

References

- Hu CM, Kelso S. Primulaceae. In: Flora of China. Edited by Wu CY, Raven PH. vol. 15. St Louis, USA/Beijing, China: Missouri Botanical Garden Press/Science Press; 1996: 39–189.
- Anderberg AA, Manns U, Källersjö M. Phylogeny and floral evolution of the Lysimachieae (Ericales, Myrsinaceae): evidence from ndhF sequence data. *Willdenowia*. 2007;37(2):407–21.
- Ståhl B, Anderberg AA. Myrsinaceae. In: Flowering plants · dicotyledons: Celastrales, Oxalidales, Rosales, Cornales, Ericales. Edited by Kubitzki K. Berlin, Heidelberg: Springer Berlin Heidelberg; 2004:266–81.
- Bennell AP, Hu CM. Pollen morphology and taxonomy of *Lysimachia*. *Notes Royal Botanic Garden Edinburgh*. 1983;40:425–58.
- Chen FH, Hu CM. Taxonomic and phytogeographic studies on Chinese species of *Lysimachia*. *Acta Phytotaxon Sin*. 1979;17:21–56.
- Hu CM. On the geographical distribution of the primulaceae. *J Trop Subtrop Bot*. 1994;2:1–14.
- Anderberg AA, Stahl B. Phylogenetic interrelationships in the order primulales, with special emphasis on the family circumscriptions. *Can J Bot*. 1995;73(11):1699–730.
- Hao G, Yuan YM, Hu CM, Ge XJ, Zhao NX. Molecular phylogeny of *Lysimachia* (Myrsinaceae) based on chloroplast trnL-F and nuclear ribosomal ITS sequences. *Mol Phylogenet Evol*. 2004;31(1):323–39.
- Källersjö M, Bergqvist G, Anderberg AA. Generic realignment in primuloid families of the Ericales s.l.: A phylogenetic analysis based on DNA sequences from three chloroplast genes and morphology. *Am J Bot*. 2000;87(9):1325–41.
- Yan HF, Zhang CY, Anderberg AA, Hao G, Ge XJ, Wiens JJ. What explains high plant richness in east asia? time and diversification in the tribe Lysimachieae (Primulaceae). *New Phytol*. 2018;219(1):436–48.
- Manns U, Anderberg AA. New combinations and names in *Lysimachia* (Myrsinaceae) for species of *Anagallis*. *Pelletiera Trientalis Willdenowia*. 2009;39(1):49–54.
- Zhang CY. Biogeography of the tribe Lysimachieae (Primulaceae) and the evaluation of DNA barcoding in closely related groups of *Lysimachia*. Guangzhou: University of Chinese Academy of Sciences; 2012.

13. Anderberg AA, Stahl B, Kallersjo M. Maesaceae, a new primuloid family in the order Ericales s.l. *Taxon*. 2000;49(2):183–7.
14. Martins L, Oberprieler C, Hellwig FH. A phylogenetic analysis of Primulaceae s.l. Based on internal transcribed spacer (ITS) DNA sequence data. *Plant Syst Evol*. 2003;237(1–2):75–85.
15. Banfi E, Galasso G, Soladano A. Notes on systematics and taxonomy for the Italian vascular flora. 1. *Atti Soc Ital Sci Nat Mus Civ Stor Nat Milano*. 2005;146:219–44.
16. Manns U, Anderberg AA. Biogeography of “tropical Anagallis” (Myrsinaceae) inferred from nuclear and plastid DNA sequence data. *J Biogeogr*. 2011;38(5):950–61.
17. Allen JF, Raven JA, Howe CJ, Barbrook AC, Koumandou VL, Nisbet RER, Symington HA, Wightman TF. Evolution of the chloroplast genome. *Philos Trans R Soc Lond B: Biol Sci*. 2003;358(1429):99–107.
18. Wicke S, Schneeweiss GM, dePamphilis CW, Muller KF, Quandt D. The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function. *Plant Mol Biol*. 2011;76(3–5):273–97.
19. Gitzendanner MA, Soltis PS, Yi TS, Li DZ, Soltis DE. Plastome phylogenetics: 30 years of inferences into plant evolution. *Adv Bot Res*. 2018;85:293–313.
20. Li HT, Yi TS, Gao LM, Ma PF, Zhang T, Yang JB, Gitzendanner MA, Fritsch PW, Cai J, Luo Y, et al. Origin of angiosperms and the puzzle of the Jurassic gap. *Nat Plants*. 2019;5(5):461–70.
21. Ruhfel BR, Gitzendanner MA, Soltis PS, Soltis DE, Burleigh JG. From algae to angiosperms—inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. *BMC Evol Biol*. 2014;14:23.
22. Li DM, Li J, Wang DR, Xu YC, Zhu GF. Molecular evolution of chloroplast genomes in subfamily Zingiberaceae (Zingiberaceae). *BMC Plant Biol*. 2021;21(1):558.
23. Lee Y, Yun N, Kang J, Choi S, Paik JH. The complete chloroplast genome of the medicinal plant *Lysimachia mauritiana* (Lamarck, 1792). *Mitochondrial DNA B: Resour*. 2022;7(3):554–5.
24. Li HL, Cheng XL, Chen Y, Tan FL. Complete plastome sequence of *Lysimachia congestiflora* Hemsl. A medicinal and ornamental species in southern China. *Mitochondrial DNA B*. 2019;4(2):2316–7.
25. Son O, Park SJ. Complete chloroplast genome sequence of *Lysimachia coreana* (Primulaceae). *Mitochondrial DNA A*. 2016;27(3):2263–5.
26. Yan XK, Liu TJ, Yuan X, Xu Y, Yan HF, Hao G. Chloroplast genomes and comparative analyses among thirteen taxa within Myrsinaceae s.str. Clade (Myrsinoideae, Primulaceae). *Int J Mol Sci*. 2019;20(18):4534.
27. Jansen RK, Ruhlman TA. Plastid genomes of seed plants. In: *Genomics of chloroplasts and mitochondria*. Edited by Bock R, Knoop V. Dordrecht: Springer Netherlands; 2012: 103–26.
28. Ogawa T. A gene homologous to the subunit-2 gene of NADH dehydrogenase is essential to inorganic carbon transport of *Synechocystis* pcc6803. *P Natl Acad Sci USA*. 1991;88(10):4275–9.
29. Baumgartner BJ, Rapp JC, Mullet JE. Plastid genes encoding the transcription translation apparatus are differentially transcribed early in barley (*Hordeum vulgare*) chloroplast development: evidence for selective stabilization of psbA messenger-RNA. *Plant Physiol*. 1993;101(3):781–91.
30. Kikuchi S, Asakura Y, Imai M, Nakahira Y, Kotani Y, Hashiguchi Y, Nakai Y, Takafuji K, Bedard J, Hirabayashi-Ishioaka Y, et al. A ycf2-ftsH heteromeric AAA-ATPase complex is required for chloroplast protein import. *Plant Cell*. 2018;30(11):2677–703.
31. Oh IC, Schonenberger J, Motley TJ, Myrenas M, Anderberg AA. Phylogenetic relationships among endemic Hawaiian *Lysimachia* (Ericales: Primulaceae): insights from nuclear and chloroplast DNA sequence data. *Pac Sci*. 2013;67(2):237–51.
32. Zhang CY, Wang FY, Yan HF, Hao G, Hu CM, Ge XJ. Testing DNA barcoding in closely related groups of *Lysimachia* L. (Myrsinaceae). *Mol Ecol Resour*. 2012;12(1):98–108.
33. Amenu SG, Wei N, Wu L, Oyebanji O, Hu GW, Zhou YD, Wang QF. Phylogenomic and comparative analyses of coffeeae alliance (Rubiaceae): deep insights into phylogenetic relationships and plastome evolution. *BMC Plant Biol*. 2022;22(1):1–3.
34. Ding SX, Dong X, Yang JX, Guo CC, Cao BB, Guo Y, Hu GW. Complete chloroplast genome of *Clethra fargesii* Franch., an original sympetalous plant from central China: Comparative analysis, adaptive evolution, and phylogenetic relationships. *Forests*. 2021;12(4):441.
35. Liu CK, Lei JQ, Jiang QP, Zhou SD, He XJ. The complete plastomes of seven Peucedanum plants: comparative and phylogenetic analyses for the Peucedanum genus. *BMC Plant Biol*. 2022;22(1):101.
36. Xiao TW, Ge XJ. Plastome structure, phylogenomics, and divergence times of tribe Cinnamomeae (Lauraceae). *BMC Genomics*. 2022;23(1):642.
37. Bevan RB, Bryant D, Lang BF. Accounting for gene rate heterogeneity in phylogenetic inference. *Syst Biol*. 2007;56(2):194–205.
38. Nakai T. Aliquot novae plantae ex Asia orientale. *Bot Mag*. 1909;23:106.
39. Kim YR, Tae KH, Sim JK, Ko SCJKoPT. A palynotaxonomic study on the genus *Lysimachia* in Korea. *Korean J Plant Taxon*. 1993;23(2):43–56.
40. Chang C-S, Kim H, Chang KJK. Provisional checklist of vascular plants for the Korea peninsula flora (kpf). *Pajo: Designpost*; 2014.
41. Soltis PS, Soltis DE. The role of hybridization in plant speciation. *Annu Rev Plant Biol*. 2009;60:561–88.
42. Rieseberg LH, Soltis DE. Phylogenetic consequences of cytoplasmic gene flow in plants. *Evol Trend Plant*. 1991;5(1):65–84.
43. Alegro A, Segota V, Koletic N, Vukovic N, Vilovic T, Rimac A. *Glaux maritima* L. (Primulaceae), a new plant species in SE Europe. *Acta Bot Croat*. 2019;78(1):95–8.
44. Doyle JJ. Defining coalescent genes: Theory meets practice in organelle phylogenomics. *Syst Biol*. 2022;71(2):476–89.
45. Walker JF, Walker-Hale N, Vargas OM, Larson DA, Stull GW. Characterizing gene tree conflict in plastome-inferred phylogenies. *PeerJ*. 2019;7:e7747.
46. Xiao TW, Xu Y, Jin L, Liu TJ, Yan HF, Ge XJ. Conflicting phylogenetic signals in plastomes of the tribe Laureae (Lauraceae). *PeerJ*. 2020;8:e10155.
47. Yang YY, Qu XJ, Zhang R, Stull GW, Yi TS. Plastid phylogenomic analyses of Fagales reveal signatures of conflict and ancient chloroplast capture. *Mol Phylogenet Evol*. 2021;163:107232.
48. Zhang MH, Xiang QP, Zhang XC. Plastid phylogenomic analyses of the Selaginella sanguinolenta group (Selaginellaceae) reveal conflict signatures resulting from sequence types, outlier genes, and pervasive RNA editing. *Mol Phylogenet Evol*. 2022;173:107507.
49. Zhang R, Wang YH, Jin JJ, Stull GW, Bruneau A, Cardoso D, De Queiroz LP, Moore MJ, Zhang SD, Chen SY, et al. Exploration of plastid phylogenomic conflict yields new insights into the deep relationships of Leguminosae. *Syst Biol*. 2020;69(4):613–22.
50. Sullivan AR, Schifftaler B, Thompson SL, Street NR, Wang XR. Interspecific plastome recombination reflects ancient reticulate evolution in *Picea* (Pinaceae). *Mol Biol Evol*. 2017;34(7):1689–701.
51. Rose JP, Toledo CAP, Lemmon EM, Lemmon AR, Sytma KJ. Out of sight, out of mind: Widespread nuclear and plastid-nuclear discordance in the flowering plant genus *Polemonium* (Polemoniaceae) suggests widespread historical gene flow despite limited nuclear signal. *Syst Biol*. 2021;70(1):162–80.
52. Pelsner PB, Kennedy AH, Tepe EJ, Shidler JB, Nordenstam B, Kadereit JW, Watson LE. Patterns and causes of incongruence between plastid and nuclear Senecioneae (Asteraceae) phylogenies. *Am J Bot*. 2010;97(5):856–73.
53. Whitfield JB, Lockhart PJ. Deciphering ancient rapid radiations. *Trends Ecol Evol*. 2007;22(5):258–65.
54. Ottenburghs J. Ghost introgression: spooky gene flow in the distant past. *Bioessays*. 2020;42(6):2000012.
55. Zhang DZ, Tang LF, Cheng YL, Hao Y, Xiong Y, Song G, Qu YH, Rheindt FE, Alstrom P, Jia CX, et al. “Ghost introgression” as a cause of deep mitochondrial divergence in a bird species complex. *Mol Biol Evol*. 2019;36(11):2375–86.
56. Doyle JJ, Doyle JL. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull*. 1987;19:11–5.
57. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114–20.
58. 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(1):1–11.
59. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, et al. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. 2012;28(12):1647–9.
60. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. Geseq - versatile and accurate annotation of organelle genomes. *Nucleic Acids Res*. 2017;45(W1):W6–11.
61. Lohse M, Drechsel O, Bock R. Organellergenomedraw (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr Genet*. 2007;52(5–6):267–74.

62. 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(W1):W59–64.
63. R Core Team. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing; 2022.
64. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. Vista: computational tools for comparative genomics. *Nucleic Acids Res.* 2004;32:W273–9.
65. Darling ACE, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 2004;14(7):1394–403.
66. Rozas J, Ferrer-Mata A, Sanchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sanchez-Gracia A. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol Biol Evol.* 2017;34(12):3299–302.
67. Gao FL, Chen CJ, Arab DA, Du ZG, He YH, Ho SYW. Easycoedeml: a visual tool for analysis of selection using codeml. *Ecol Evol.* 2019;9(7):3891–8.
68. Edgar RC. Muscle: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004;32(5):1792–7.
69. Tamura K, Stecher G, Kumar S. MEGA11 molecular evolutionary genetics analysis version 11. *Mol Biol Evol.* 2021;38(7):3022–7.
70. Zhang D, Gao FL, Jakovlic I, Zou H, Zhang J, Li WX, Wang GT. Phylousuite: an integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Mol Ecol Resour.* 2020;20(1):348–55.
71. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol.* 2013;30(4):772–80.
72. Borowiec ML. AMAS: a fast tool for alignment manipulation and computing of summary statistics. *Peerj.* 2016;4:e1660.
73. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 2014;30(9):1312–3.
74. Degnan JH, Rosenberg NA. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends Ecol Evol.* 2009;24(6):332–40.
75. Zhang C, Rabiee M, Sayyari E, Mirarab S. ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics.* 2018;19:15–30.
76. Sayyari E, Mirarab S. Fast coalescent-based computation of local branch support from quartet frequencies. *Mol Biol Evol.* 2016;33(7):1654–68.
77. Pease JB, Brown JW, Walker JF, Hinchliff CE, Smith SA. Quartet sampling distinguishes lack of support from conflicting support in the green plant tree of life. *Am J Bot.* 2018;105(3):385–403.
78. Smith SA, Moore MJ, Brown JW, Yang Y. Analysis of phylogenomic datasets reveals conflict, concordance, and gene duplications with examples from animals and plants. *BMC Evol Biol.* 2015;15:1–5.
79. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-tree: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 2015;32(1):268–74.
80. Shen XX, Hittinger CT, Rokas A. Contentious relationships in phylogenomic studies can be driven by a handful of genes. *Nat Ecol Evol.* 2017;1(5):0126.
81. Wickham H. *Ggplot2: Elegant graphics for data analysis.* New York: Springer; 2016.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

