

RESEARCH

Open Access



Exploring the phylogenetic framework and trait evolution of *Impatiens* through chloroplast genome analysis

Wen-Xiang Lan^{1†}, Qing Mo^{1†}, Meng-Meng Jin^{1†}, Yong-Hui Wen¹, Meng-Qing Yang¹, Hui Ma¹, Hai-Quan Huang^{1*} and Mei-Juan Huang^{1*}

Abstract

Background The genus *Impatiens*, which includes both annual and perennial herbs, holds considerable ornamental, economic, and medicinal value. However, it posed significant challenges for taxonomic and systematic reconstruction. This was largely attributed to its high intraspecific diversity and low interspecific variation in morphological characteristics. In this study, we sequenced samples from 12 *Impatiens* species native to China and assessed their phylogenetic resolution using the complete chloroplast genome, in conjunction with published samples of *Impatiens*. In addition, a comparative analysis of chloroplast genomes were conducted to explore the evolution of the chloroplast genome in *Impatiens*.

Results The chloroplast genomes of 12 *Impatiens* species exhibited high similarity to previously published samples in terms of genome size, gene content, and sequence. The chloroplast genome of *Impatiens* exhibited a typical four-part structure, with lengths ranging from 146,987 bp(*I. morsei*)- 152,872 bp(*I. jinpingensis*). Our results identified 10 mutant hotspot regions (*rps16*, *rps16-trnG*, *trnS-trnR*, and *rpoB-trnC*) that could serve as effective molecular markers for phylogenetic analyses and species identification within the *Impatiens*. Phylogenetic analyses supported the classification of *Impatiens* as a monophyletic taxon. The identified affinities supported the taxonomic classification of the subgenus *Clavicarpa* within the *Impatiens*, with subgenus *Clavicarpa* being the first taxon to diverge. In phylogenetic tree, the *Impatiens* was divided into eight distinct clades. The results of ancestral trait reconstruction suggested that the ancestral traits of *Impatiens* included a perennial life cycle, four sepals and three pollen grooves. However, the ancestral morphology regarding fruit shape, flower colour, and spacing length remained ambiguous.

Conclude Our study largely supported the family-level taxonomic treatment of *Impatiens* species in China and demonstrated the utility of whole chloroplast genome sequences for phylogenetic resolution. Comparative analysis of the chloroplast genomes of *Impatiens* facilitated the development of molecular markers. The results of ancestral trait reconstruction showed that the ancestor type of habit was perennial, the number of sepals was 4, and morphology and number of aperture was 3 colpus. The traits of capsule shape, flower colour, and spur length underwent

[†]Wen-Xiang Lan, Qing Mo and Meng-Meng Jin contributed equally to this work.

*Correspondence:

Hai-Quan Huang
haiquan_huang@swfu.edu.cn
Mei-Juan Huang
meijuanhuang@swfu.edu.cn

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



a complex evolutionary process. Our results provided data support for further studies and some important new insights into the evolution of the *Impatiens*.

Keywords *Impatiens*, Phylogeny, Chloroplast genome, Comparative analysis, Adaptive evolution

Backgrounds

Impatiens is a genus of annual or perennial herbs, which can be sparsely epiphytic or subshrubs, comprising over 1,000 species worldwide. Phylogenetic studies demonstrated that the *Impatiens* can be divided into two subgenera: Subgenus *Impatiens* and Subgenus *Clavicarpa* [1]. The classification of the *Impatiens* is particularly challenging due to the high intrageneric diversity of morphological characters. Hooker declared that the *Impatiens* is a very difficult taxon in plant taxonomy [2]. In recent years, numerous scholars identified new species of *Impatiens* in China, particularly in karst regions, which are recognized as biodiversity hotspots [3–6].

Fujihashi analyzed the phylogenetic relationships of *Impatiens* using combined chloroplast *rbcL* and *trnL-trnF* spacer regions [7]. However, the small sample size and the considerable distance of the outgroups resulted in a phylogenetic tree that did not adequately resolve the relationships within the genus. Yuan selected 112 plant species to establish a phylogenetic framework for the family Balsaminaceae, and demonstrated that both the family and the *Impatiens* are monophyletic [8]. This work addressed the issue of infrageneric classification. Additionally, Yu Shengxiang constructed a phylogenetic tree that integrated traits such as inflorescence type, fruit type, sepal number and carpel number, along with molecular data from *ITS*, *atpB-rbcL* and *trnL-F* regions. This analysis resulted in the division of *Impatiens* into two subgenera: Subgen. *Impatiens* and Subgen. *Clavicarpa*. Within Subgen. *Impatiens*, seven distinct groups were identified. Furthermore, the study found that *Impatiens* and *Hydrocera* formed a sister group, constituting a monophyletic clade [9].

Chloroplasts are semi-autonomous organelles responsible for photosynthesis and energy conversion in plants [10]. They contain their own DNA, which is inherited matrilineally in most angiosperms [11]. The chloroplast genome remained structurally stable throughout evolution, with mutation rates that are intermediate compared to those of mitochondrial and nuclear genomes [12, 13], suggesting a unique evolutionary trajectory [14]. Chloroplast genomic data was instrumental in resolving the phylogeny of *Delphinium* [15], elucidating the intergeneric relationships and the spatiotemporal evolutionary history of *Eriocaulon* (Eriocaulaceae) [16]. Liu discovered five hypervariable regions by using the chloroplast genome, which could serve as potential molecular markers for

Caragana [17]. RAN sequenced and compared *Tuberculata* and found that six mutation hot spot regions could serve as potential molecular markers. *Tuberculata* formed a monophyletic group and was divided into two evolutionarily independent branches, which confirmed the independence of this part [18]. Li identified five combined DNA regions that could serve as potential markers for future phylogenetic analysis and species identification of *Costaceae* plants [19]. QIU analyzed the chloroplast genome characteristics of seven *Impatiens* species and explored the affinities among 27 species of *Impatiens*. The results effectively clarified the relationship between Subgen. *Impatiens* and Subgen. *Clavicarpa* [20]. LUO examined the structure of chloroplast genomes in three ornamental *Impatiens* species, identified differentiation hotspots, and determined their phylogenetic positions [21]. Thus, the chloroplast genome served as an ideal model for studying genome evolution and provides molecular markers for resolving systematic affinities [22, 23]. Chloroplast sequences were among the first to be utilized for molecular evolution studies [24], and the differences in evolutionary rates among genes or lineages within chloroplast genomes garnered considerable attention [25].

In recent years, an increasing number of studies demonstrated that variations in the chloroplast genome provide valuable information for resolving phylogenetic relationships across multiple taxonomic levels, particularly in taxonomically complex groups [24, 26]. In this study, we assembled the complete chloroplast genomes of 12 samples of *Impatiens* and combined them with published samples available in GenBank. Additionally, we analyzed a majority of the chloroplast gene sequences present in GenBank. Our specific objectives were: (a) to compare the chloroplast genome structures within the *Impatiens*; (b) to identify mutational hotspot regions as potential chloroplast markers for species identification and phylogenetic analysis; (c) to utilize the complete chloroplast genome to infer and test phylogenetic relationships and trait evolution among *Impatiens* genera; (d) to incorporate chloroplast gene sequences from GenBank to elucidate deep relationships among *Impatiens* species worldwide. We revealed phylogenetic relationships among these species. 10 mutational hotspot regions were identified, which can be used as potential molecular markers. We found Some ancestral traits of *Impatiens* in this study. *Impatiens* undergone a complex evolution in

the course of its history, and this was linked to its adaptation to ecological environment.

Results

General features of *Impatiens* chloroplast genomes

The chloroplast genome of the *Impatiens* exhibited characteristics similar to those of other angiosperms. It was structured as a double-stranded circular tetrad, comprising a large single-copy region (LSC), a pair of inverted repeat regions (IRa and IRb), and a small single-copy region (SSC) (Fig. 1). The lengths of the chloroplast genomes across the 12 species analyzed ranged from 146,987 bp (*I. morsei*) to 152,872 bp (*I. jinpingensis*). The IR regions varied in length, measuring from

52,472 bp (*I. morsei*) to 51,044 bp (*I. duclouxii* var.), while the LSC region spanned from 83,611 bp (*I. gasterocheila*) to 81,527 bp (*I. morsei*). The SSC region showed a range from 17,991 bp (*I. gasterocheila*) to 12,988 bp (*I. morsei*). The GC content of the chloroplast genomes within the *Impatiens* was consistently around 37% (Table 1).

SSR polymorphisms and long repeat structure

A total of 1,020 simple sequence repeats (SSRs) were identified in the chloroplast genomes of 12 *Impatiens* species (Fig. 2a). The number of SSRs in these species ranged from 75 to 106, with an average of 93 (Fig. 2b). Mononucleotide repeats were the most prevalent, accounting for 92.16% of the total SSRs, followed by

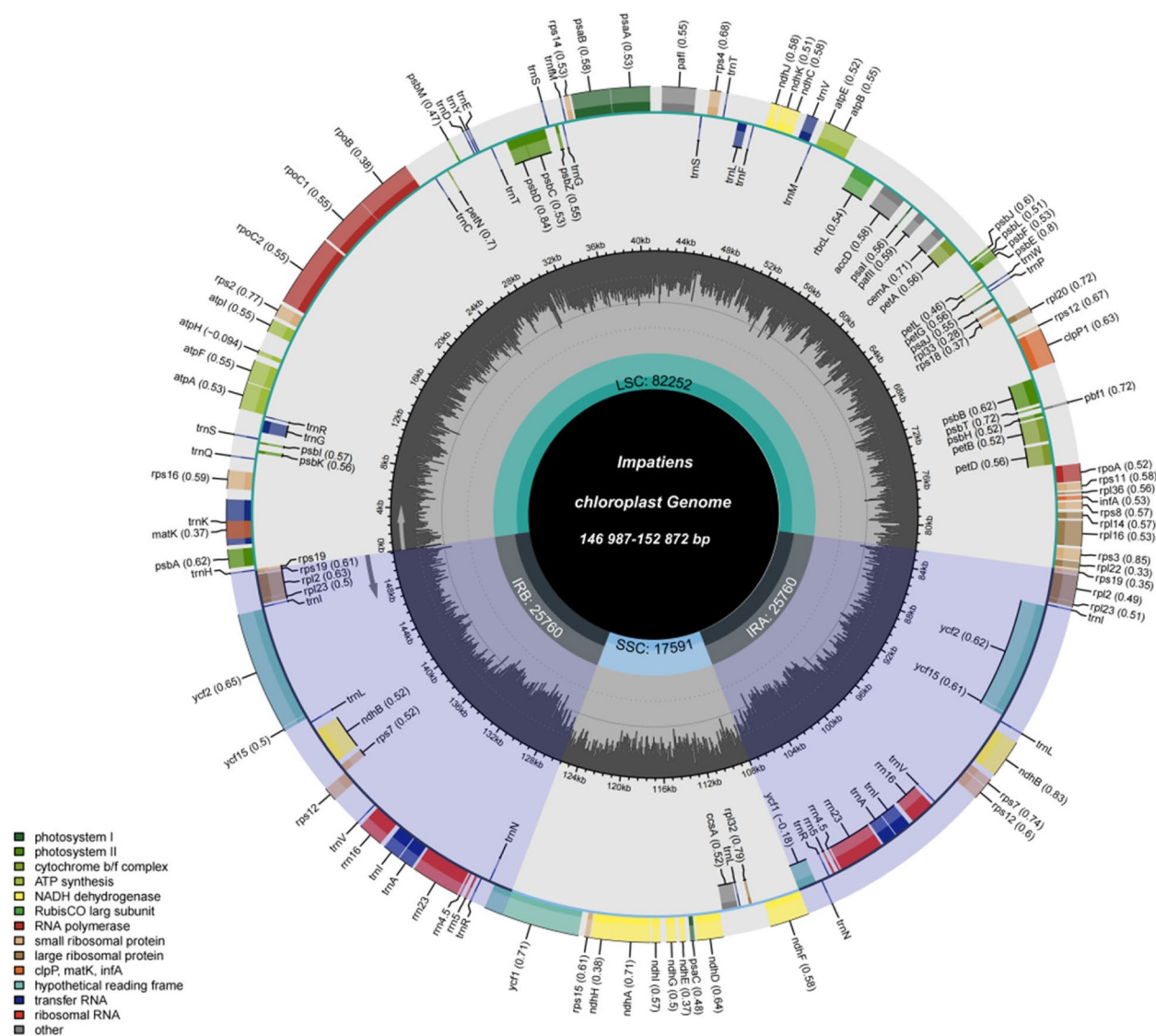
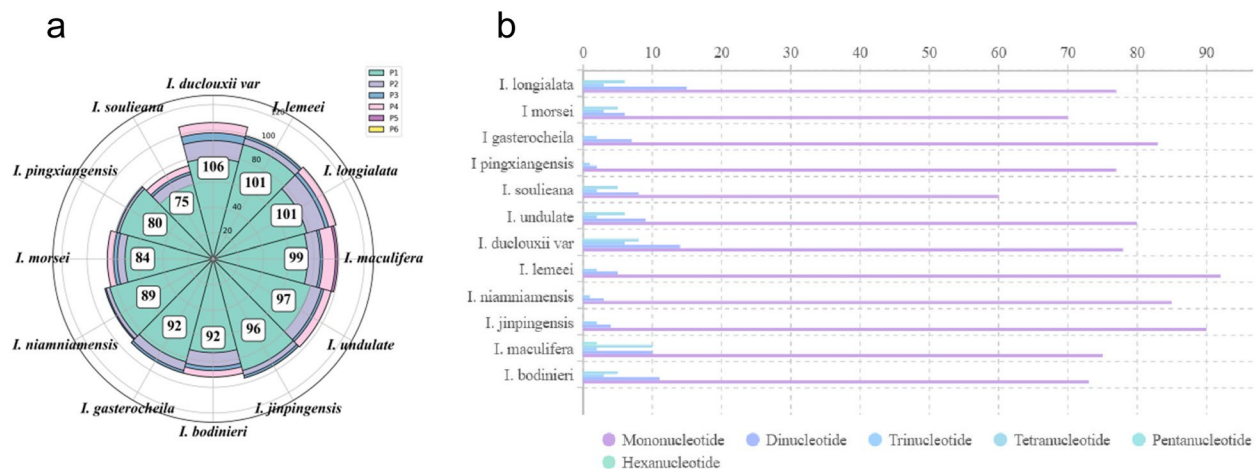


Fig. 1 Structural and gene map of the *Impatiens* chloroplast genomes

Table 1 Summary of the chloroplast genomes for 12 *Impatiens* samples

Species	Total length (bp)	LSC(bp)	IR(bp)	SSC(bp)	Total GC content (%)	Total Genes	Number of Total Genes	Number of tRNA Genes	Number of rRNA Genes
<i>I. bodinieri</i>	151,363	82,252	51,520	17,591	37%	116	82	30	4
<i>I. maculifera</i>	152,756	83,419	51,488	17,849	37%	115	81	30	4
<i>I. jinpingsensis</i>	152,872	83,946	52,062	17,279	37%	114	80	30	4
<i>I. niamniamensis</i>	152,602	83,544	51,940	17,568	37%	109	76	29	4
<i>I. lemeei</i>	151,643	82,728	51,532	17,843	37%	121	88	29	4
<i>I. duclouxii</i> var	151,607	83,081	51,044	17,482	37%	118	84	30	4
<i>I. undulate</i>	152,040	82,995	51,534	17,511	37%	119	85	30	4
<i>I. soulieana</i>	152,067	83,042	51,514	17,511	37%	119	85	30	4
<i>I. pingxiangensis</i>	152,689	83,375	51,514	17,800	37%	112	72	36	4
<i>I. gasterocheila</i>	152,664	83,611	51,062	17,991	37%	111	79	28	4
<i>I. morsei</i>	146,987	81,527	52,472	12,988	37%	115	80	30	4
<i>I. longialata</i>	151,568	82,593	51,476	17,499	37%	118	84	30	4

**Fig. 2** Numbers and types of SSRs in the 12 *Impatiens* chloroplast genomes. **a** Total number of the four repeat types per genome **b** Number of repeat

dinucleotide repeats at 9.22%, tetranucleotide repeats at 4.41%, and trinucleotide repeats at 2.84%. Pentanucleotide repeats were detected only twice in *I. maculifera*, while hexanucleotide repeats were not observed at all. Additionally, four types of long repeat sequences were detected: direct, inverted, complementary, and palindromic. We identified four types of long repeat sequences: direct repeats, inverted repeats, complementary repeats, and palindromic repeats (Fig. 3). The number of direct repeat sequences ranged from 7–11, while inverted repeat sequences were found in quantities of 0–2. Complementary repeat sequences were observed in 0–1 instance, and palindromic repeat sequences varied from 7–16. Among the species studied, *I. niamniamensis* exhibited the fewest occurrences of these sequences, with a total of 16, whereas *I. morsei* had the highest occurrence, totaling 28.

Impatiens chloroplast genome variation

A comparison of the boundaries of the inverted repeat (IR) and small single-copy (SSC) regions was conducted for 19 species of the *Impatiens* (Fig. 4). The *rps19* gene is typically located at the junction between the large single-copy (LSC) and inverted repeat b (IRb) regions (JLB), except in the case of *I. morsei*. In the chloroplast genomes of the 19 *Impatiens* species, these extensions lead to a reduction in the size of the LSC, which includes the N-terminal portion of *rps19*.

The mVISTA results showed the chloroplast genomes of the 19 species within the Balsaminaceae exhibited strong conservation, high collinearity, (Fig. 5) and significant homology, reflecting a high degree of similarity. However, certain differences were noted, particularly with varying mutation rates in the inverted repeat (IR), small single-copy (SSC), and large single-copy (LSC)

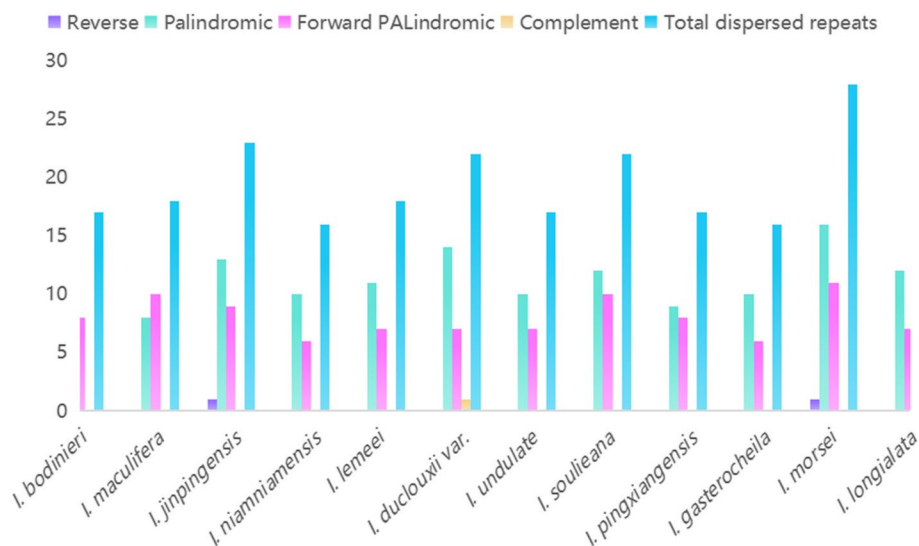


Fig. 3 Long-repeat sequences in the *Impatiens* chloroplast genomes

regions, with the IR region demonstrating greater conservation. Additionally, the coding regions were found to be more conserved than the non-coding regions. Nonetheless, a number of highly differentiated regions were identified in the intergenic spacers and coding genes, including *matK*, *psbK*, *petN*, *psbM*, *atpE*, *rbcl*, *accD*, *psaL*, *rpl16*, *rpoB*, *ndhB*, *ndhF*, *ycf1*, and, *ndhH*.

Using nucleotide polymorphism analysis, we found that the single-copy regions of the Large Single-Copy (LSC) and Small Single-Copy (SSC) sequences are highly differentiated, while the inverted repeat (IR) regions exhibit relatively low nucleotide polymorphism (Fig. 6a). The nucleotide diversity (P_i) values for the coding and intergenic regions within the *Impatiens* indicated that the intergenic regions have higher values than the coding regions, suggesting a greater degree of differentiation in the intergenic regions. The average nucleotide diversity (P_i) value across 19 species of *Impatiens* was 0.021057682. Specifically, the P_i values for the LSC region range from 0 to 91,785 bp, with values between 0.00063 and 0.0185; for the IR region, the P_i values range from 91,786 bp to 118,364 bp, with values between 0.00059 and 0.13189; and for the SSC region, the P_i values range from 118,365 bp to 137,534 bp, with values between 0.00465 and 0.12854 (Fig. 6b).

Molecular evolution of the *Impatiens* chloroplast genomes

The average K_a/K_s ratio (ω) of 19 protein-coding genes was calculated. Overall, 45 genes exhibited ω values less than 1 (Fig. 7a). In contrast, 34 genes had K_a/K_s values greater than 1 in at least one species, indicating the action of strong selective constraints or purifying selection. The

expression levels of photosynthesis-related genes, such as *accD*, *atpB*, *atpF*, *clpP*, *ndhC*, *ndhJ*, *psaA-B*, *psaI*, *psbB*, and *psbI*, were consistent across most tested species. The evolutionary rates of 11 genes related to transcription and translation (*rpl14*, *rps11*, *rpl16*, *rpl20*, *rpl33*, *rpoA*, *rpoB*, *rpoC1*, *rps19*, *rps3*, and *rps8*) were found to be lower than those of photosynthesis-related genes, suggesting that these genes were under relaxed selective constraints and may be experiencing weak positive selection (Fig. 7b-e).

Phylogenetic relationships

Phylogenetic trees were constructed using both maximum likelihood and maximum parsimony methods, based on a comprehensive chloroplast genome dataset that included 46 chloroplast genomes of the *Impatiens*, along with a single representative from the genus *Hydrocera*, specifically *Hydrocera triflora*, which served as the outgroup in our phylogenetic analysis. Phylogenetic trees constructed on the basis of different approaches had very high similarity across branches. The relationships among all major clades within the Balsaminaceae family are strongly supported, as illustrated in (Fig. 8). In the phylogenetic tree, the family Balsaminaceae was clearly differentiated into two primary clades: one representing the *Impatiens* and the other representing the *Hydrocera*. The monophyly of the two genera is further illustrated. Within the *Impatiens*, there was a further differentiation into two major clades. The first taxon to diverge within *Impatiens* was Subgenus *Clavicarpa*, which encompassed five distinct species: *I. maculifera*, *I. guizhouensis*, *I. pritzelii*, and



Fig. 4 Comparison of large single-copy region (LSC), small single-copy region (SSC), inverted repeat regions (IR) boundaries in the chloroplast genomes of 19 *Impatiens* species

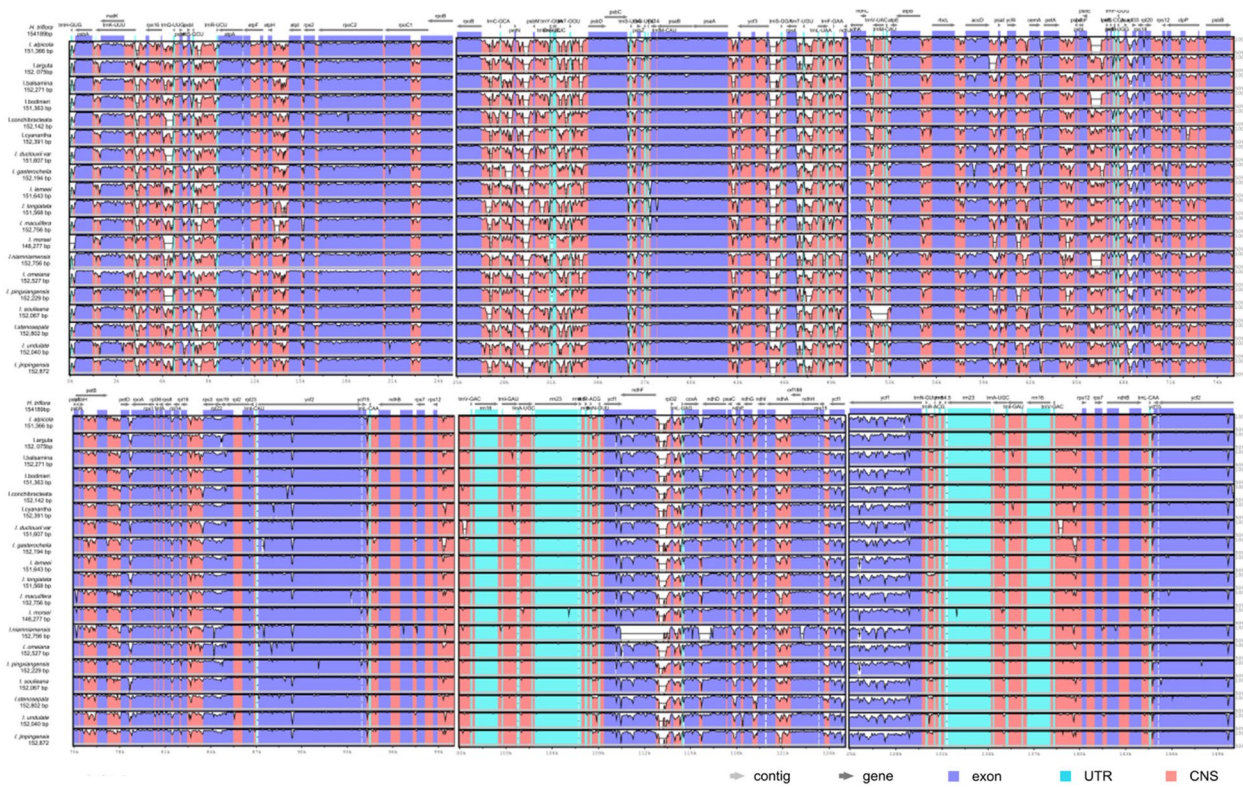


Fig. 5 Visualization of chloroplast genomes alignment of 19 *Balsaminaceae* species

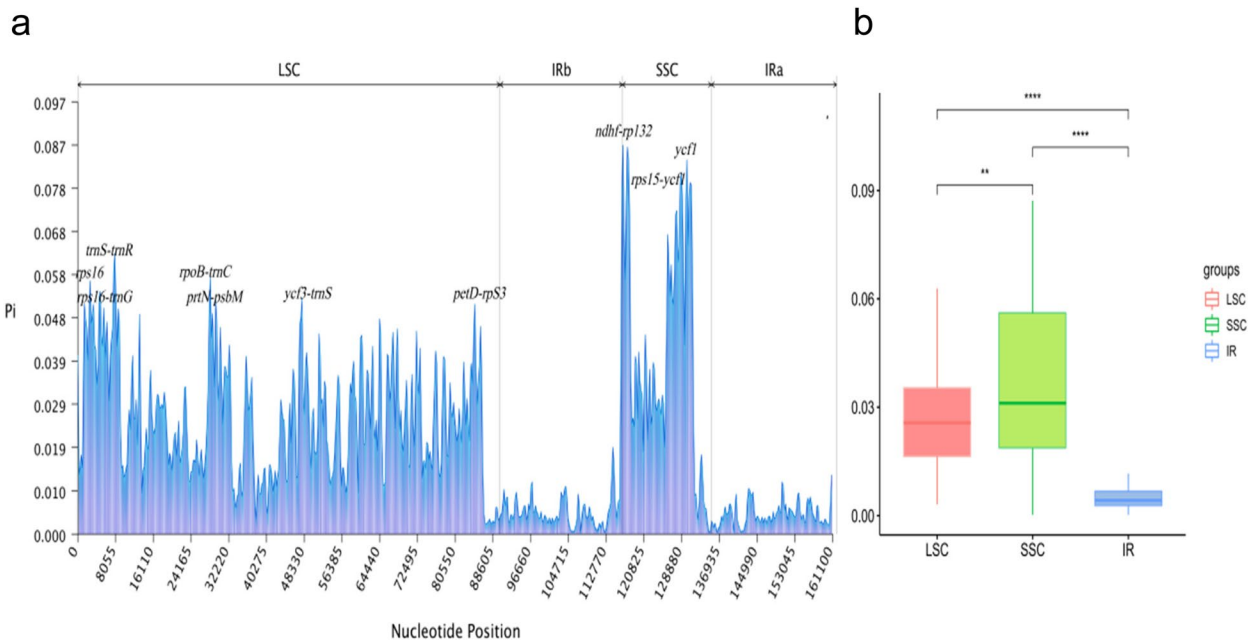


Fig. 6 The nucleotide diversity (π) values in the *Impatiens* chloroplast genomes. Window size: 800 bp, step size: 100 bp. **a** The π values of the windows. **b** Boxplots of π -value differences among the LSC, IR, and SSC regions

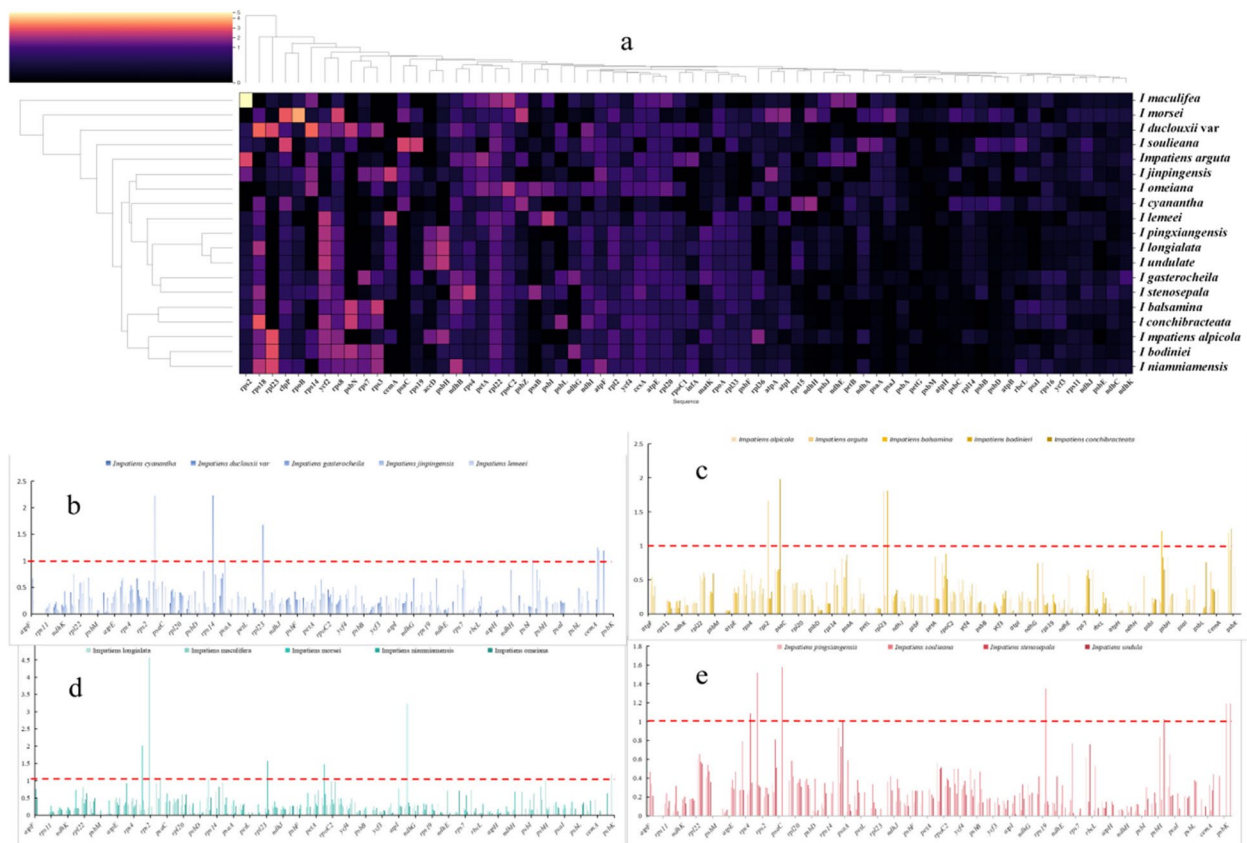


Fig. 7 Selective pressure analysis results. **a** Cluster heatmap showing the Ka/Ks values of chloroplast genomes from 19 species, using *Hydrocera triflora* as a reference, the Ka/Ks value varies between 0 and 2, corresponding to a colour range of blue to red. **b** A set of graphs displaying the Ka/Ks values of each gene

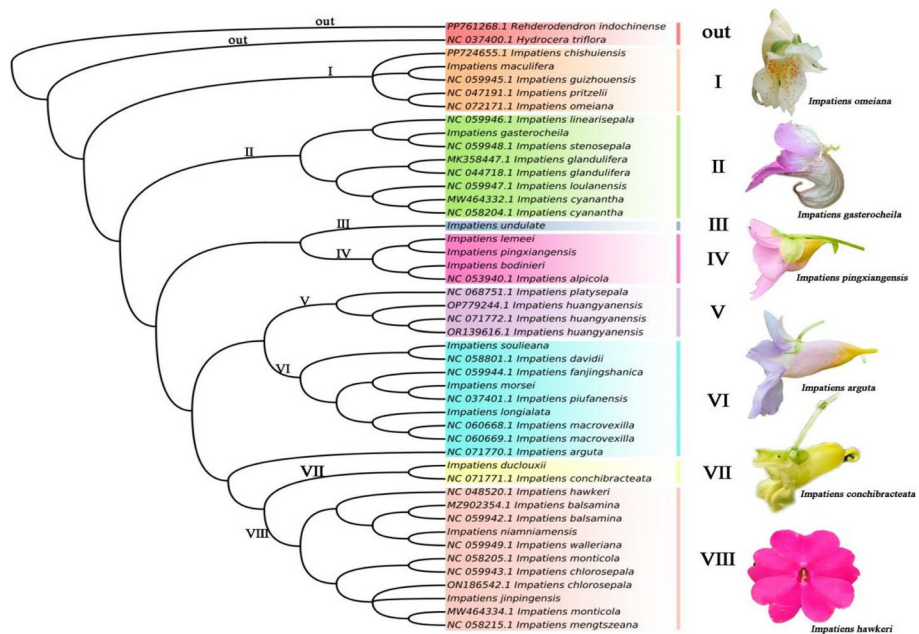


Fig. 8 Maximum Likelihood phylogenetic tree based on complete chloroplast genomes

I. omeiana. Following this, the second clade to diverge was Subgenus *Impatiens*, which subsequently differentiated into seven separate clades.

Ancestral trait reconstruction

In this study, we utilized the chloroplast genome to reconstruct the phylogenetic relationships within the Balsaminaceae family. To analyze the ancestral morphology and evolutionary trends of *Impatiens*, we mapped six morphological characters—habit, capsule shape, flower colour, spur length, morphology and number of apertures on the constructed phylogenetic trees.

Habit

Perennial was considered to be the ancestral state. And perennial were inferred as a synapomorphy for branch II, IV and VIII. Branch I and VII evolved annual traits. Subgen. *Clavicarpa* (VIII) were all perennial in this study (Fig. 9a).

Capsule shape

The ancestral character state of the capsule shape was ambiguous in the last common ancestor of the *Impatiens*. The fusiform capsule shape was more conservative and was inferred to be the synapomorphy of branch IV and V. The fusiform fruit shape of branch VIII (Subgen. *Clavicarpa*) were almost clavate (Fig. 9b).

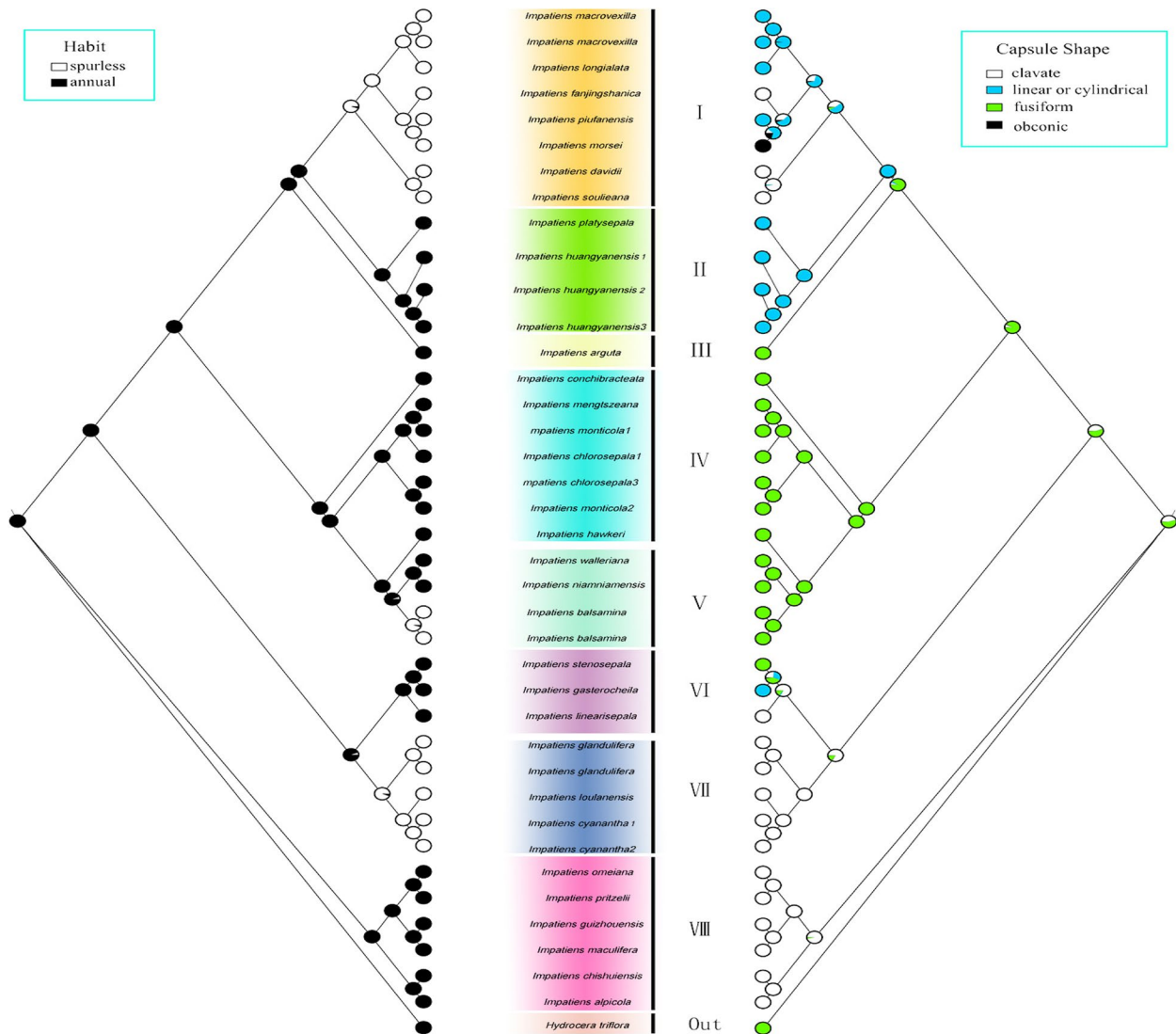


Fig. 9 a Ancestral reconstruction of the Habit in *Impatiens* using maximum parsimony (MP). b Ancestral reconstruction of the Capsule shape in *Impatiens* using maximum parsimony (MP)

Flower colour

The ancestral character state of flower colour were also ambiguous. Flower colour were found to have evolved so many times independently. The synapomorphies of the various branches were ambiguous. The different flower colours were scattered in different branches, and no clear regularity was found (Fig. 10a).

Spur length

We can found that the Spur length that less than or equal to 1 cm has undergone three independent evolutions, which appeared in branch I, IV and VI respectively. The evolution of spur from long to short was observed in the results (Fig. 10b).

Sepal number

Four sepals were recognized as the ancestral character state in the results of the character evolution. Two sepals acted as a synapomorphy in branch I, IV and V (Fig. 11a).

Morphology and number of aperture

A pollen Aperture number of 3 was considered to be the ancestral character state. All of branch VIII (Subgen. *Clavicarpa*) had 3-colpus pollen in the this study. Branch VIII was in a primitive state (Fig. 11b).

Discussion

The evolution of chloroplast genomes in the *Impatiens*

In this study, we utilized high-throughput sequencing technology to assemble the chloroplast genome

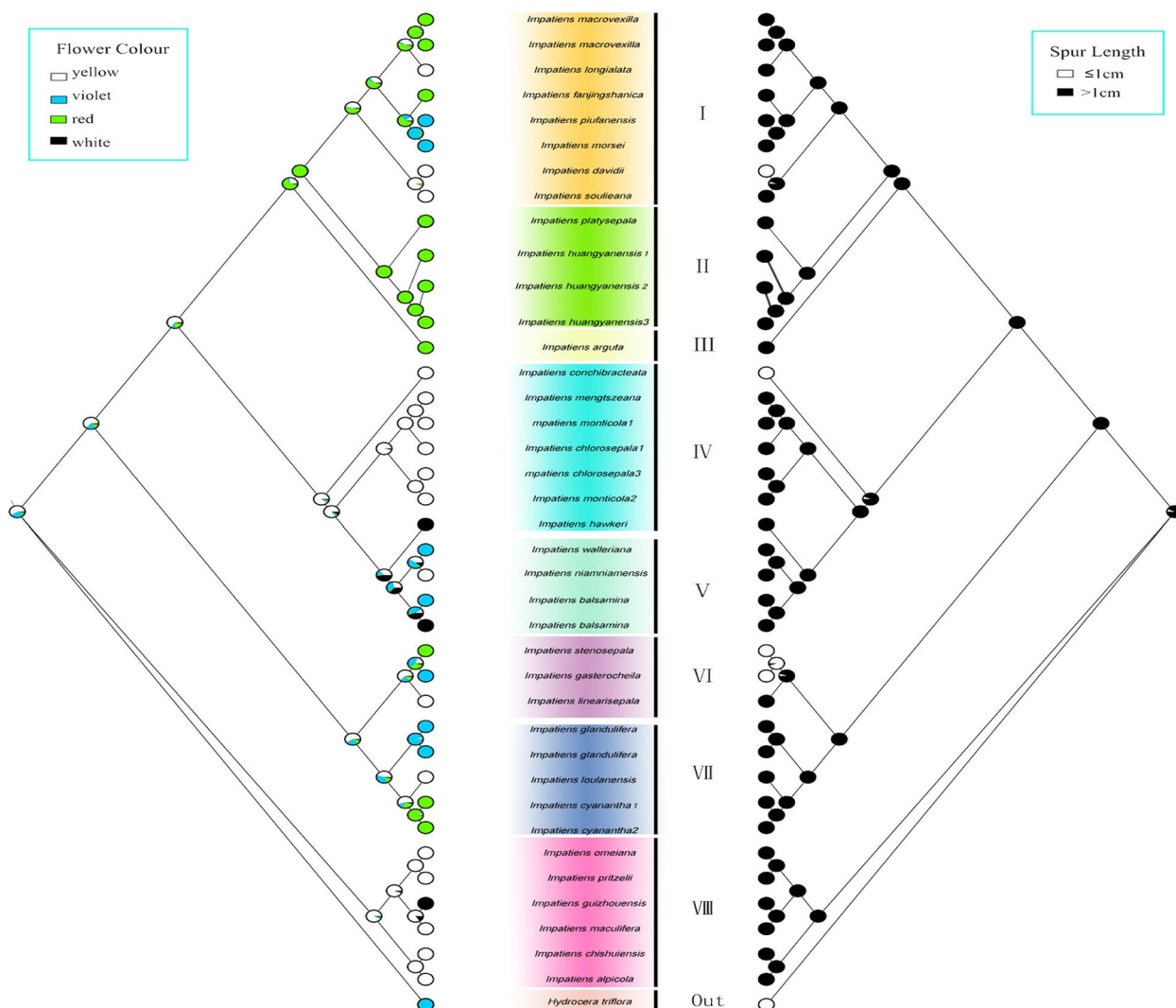


Fig. 10 a Ancestral reconstruction of the flower colour in *Impatiens* using maximum parsimony (MP). **b** Ancestral reconstruction of the spur length in *Impatiens* using maximum parsimony (MP)

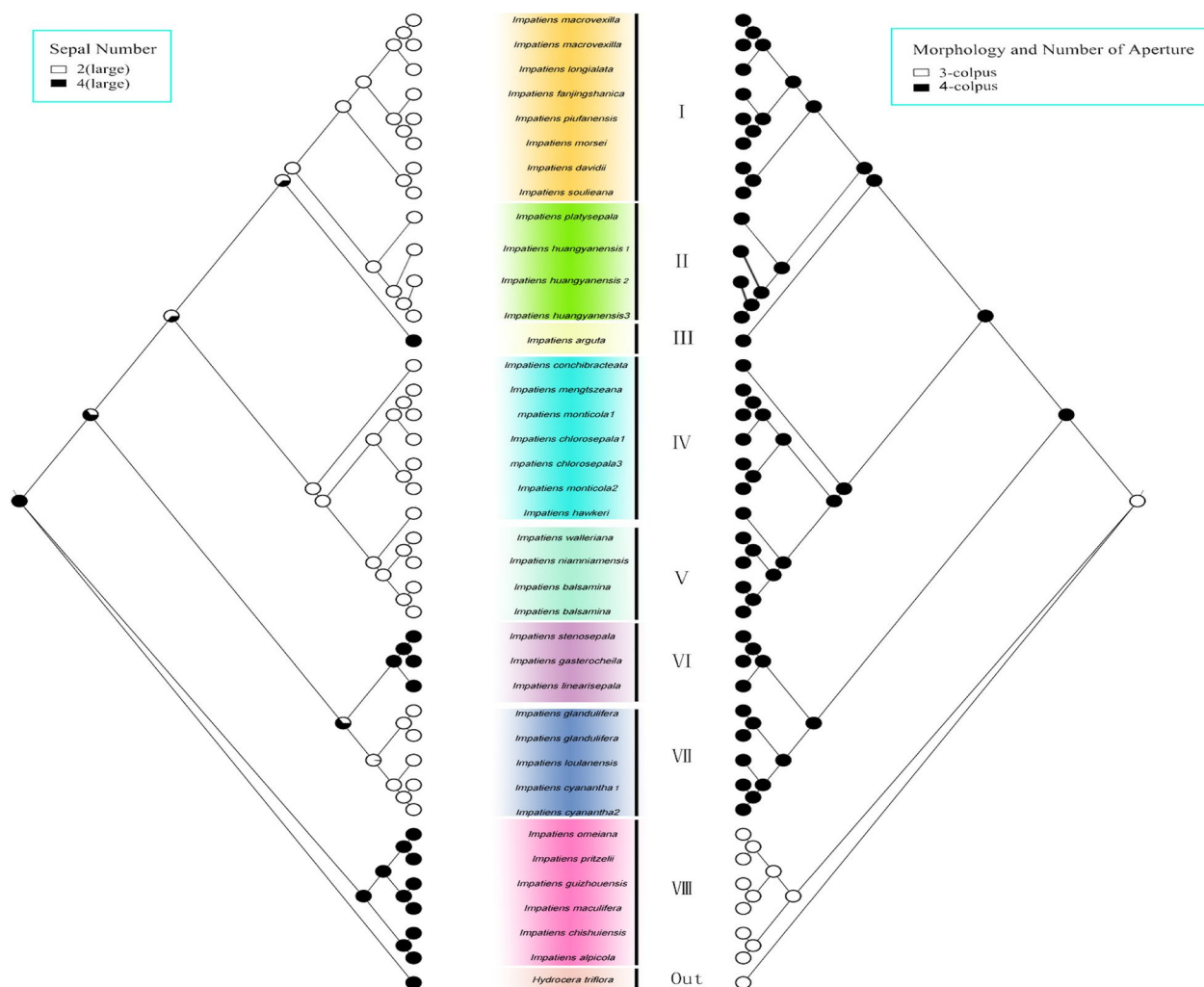


Fig. 11 **a** Ancestral reconstruction of the sepal number in *Impatiens* using maximum parsimony (MP). **b** Ancestral reconstruction of the Morphology and Number of Aperture in *Impatiens* using maximum parsimony (MP)

sequences of 12 species within the *Impatiens*. All chloroplast genomes of the *Impatiens* species exhibited a typical circular tetrad structure, consisting of a pair of inverted repeat regions along with large and small single-copy (SSC) regions (Fig. 1). In the Balsaminaceae, this structure was similar to the chloroplast genomes of other species [20, 21]. The tetrad structure of the chloroplast genome represented an adaptive feature that evolved over the long term in plants, contributing to genome stability, replication efficiency, and genetic diversity [27]. Notably, *I. morsei* exhibited a reduced total gene length, with no significant differences observed in the LSC and IR regions compared to other species in the *Impatiens*, primarily due to the contraction of the SSC region. While similar genomic contraction phenomena was documented in other plant species [28], no related studies were reported within the *Impatiens*. In the context of

plant genomic evolution, changes in genome size were often closely associated with environmental adaptation, gene duplication, gene loss, and other factors. For *I. morsei*, the contraction of the SSC region may be linked to its specific ecological adaptation strategies [29].

Through the comparison of genomes among different species within the *Impatiens*, we found that the genomes exhibit a relative conservativeness overall, but significant variations existed in specific regions. Firstly, the differences in genome size primarily arised from variations in the Large Single Copy (LSC) region of the chloroplast genome, as well as the contraction and expansion of the Inverted Repeat (IR) regions. These variations not only affected the size of the genome, but also may have significant impacts on the plant's adaptability and evolution [30, 31]. For instance, variations in the LSC region may be related to environmental adaptability, while changes in

the IR region could influence gene expression and function [32, 33]. Secondly, the chloroplast genomes of the *Impatiens* showed a preference for A/T bases, which was consistent with other species within the order Ericales. Relevant studies indicated that the A/T base preference of *Actinidia chinensis* is 62.8% [34], suggesting that the genomes of the *Impatiens* may be subjected to similar evolutionary pressures regarding base composition. This A/T preference may be closely related to factors such as ecological adaptability, genome stability, and genetic variation [35].

The mVISTA visualization analysis indicated that the non-coding region sequences of the chloroplast genomes in the *Impatiens* exhibited moderate differentiation. This suggested that the variation rate in non-coding regions was higher than that in coding regions, inferring that coding regions were more conserved. Specifically, there were higher variations in non-coding regions compared to coding regions, and higher variations in the LSC and SSC regions compared to the IR region. This was speculated to be related to the selection pressures experienced; lower selection pressure was associated with structural variations, while higher selection pressure contributed to the relative stability of the structure. In the *Impatiens* genome, the ka/Ks ratios of several coding genes, such as *rps2*, *rps14*, *rpl23*, *psbI*, and *psbK*, were greater than 1, reflecting evolutionary selection and indicating that these genes have undergone positive selection. Notably, most species in the *Impatiens* exhibited significant positive selection for the *rps123* gene. The *rpl32* gene, which played a key role in protein synthesis as a ribosomal protein, shown to be important under abiotic stress conditions in rice [36]. Additionally, studies in the family Euphorbiaceae found that the *rpl32* gene transferred from the chloroplast to the nuclear genome, acquiring new transport peptides in the process [37]. In summary, the significantly higher variation in the noncoding regions than in the coding regions, the significantly higher variation in the LSC and SSC regions than in the IR regions, and the synergistic evolution of several genes with plant nuclear genes in *Impatiens* may have contributed to the occurrence of positive selection.

Chloroplast markers for *Impatiens*

DNA sliding and mismatches, as well as unequal exchanges between sister chromatids during mitosis and meiosis, could lead to the formation of simple sequence repeats (SSRs) or tandem repeat sequences [38]. The number of SSR repeat sequences could influence gene regulation, transcription, and protein function, providing a source of both quantitative and qualitative variation [39]. In a study of 12 species within the *Impatiens*, Tetranucleotide repeat sequences were detected in seven

species, including *I. bodinieri*, *I. maculifera*, *I. duclouxii* var., *I. undulata*, and *I. soulieana*, among others. Additionally, pentanucleotide sequences were identified in *I. maculifera*. These tetranucleotide and pentanucleotide repeat sequences may play a role in the identification of these species. In fact, many studies employed this method for species identification [40], and related molecular markers were currently being developed. The diversity and distribution of SSRs were of significant importance for species adaptation and evolution, serving as genetic markers for studying the genetic structure and intraspecific variation within species [41, 42].

The classification, evolution and genetic development of plants were based on high mutation regions, simple sequence repeats (SSRs), and single nucleotide polymorphisms (SNPs) as molecular markers [43]. Chloroplast gene fragments such as *rbcl*, *rps16*, *atpB*, *ndhF*, *matK*, and *atpB-rbcL* were widely used as universal barcodes in classification and molecular systematics research [44, 45]. Currently, in the phylogenetic study of the *Impatiens*, a limited number of gene fragments, including *atpB-rbcL*, were insufficient to resolve the relationships among closely related species within the genus [46, 47]. By exploring the highly variable regions of the complete chloroplast genomes of 19 *Impatiens* species, we identified 10 gene fragments with a high degree of differentiation: *rps16*, *rps16-trnG*, *trnS-trnR*, *rpoB-trnC*, *trnN-psbM*, *ycf3-trnS*, *petD-rps3*, *ndhF-rpl32*, *rps15-ycf1*, and *ycf1*. These fragments were effectively utilized in phylogenetic analyses among wild *Impatiens* species. The diversity of germplasm resources in *Impatiens* and the complex evolutionary issues within the genus presented significant challenges for the systematic evolution and classification of these plants. This was particularly true for the morphological classification and identification of wild *Impatiens*, as well as for elucidating the phylogenetic relationships among species. Therefore, the highly mutated regions screened in this study to serve as potential molecular markers of *Impatiens* can help to provide certain scientific evidence for the identification of new species of *Impatiens*, molecular evolution, and phylogenetic studies.

Phylogenetic relationships in *Impatiens*

The results of the chloroplast genome phylogenetic tree derived from this study, based on different methods were highly similar. The Balsaminaceae family was identified as a monophyletic taxon. The outgroup, *Hydrocera triflora*, was positioned at the base of the branch and exhibits a sister-group relationship with the *Impatiens*. This was the same finding as in the study by Luo [5]. Most branches received high support, with 100% bootstrap values. All species of *Impatiens* formed a distinct monophyletic

clade. This was consistent with the findings of previous researchers [9, 20, 21, 48]. A total of forty-six species of *Impatiens* were classified into eight clades, and the phylogenetic tree exhibited very high self-expansion values and resolution. *Impatiens* and *Hydrocera* could be clearly differentiated, which clarified the interspecies phylogenetic relationships within the *Impatiens*.

In this study, all species of Subgen. *Clavicarpa* exhibited a pollen colpus count of three and a sepal count of four. It was observed that the morphological characteristics of these species align well with the phylogenetic tree. This taxon represented the original lineage of *Impatiens* [49] and was positioned at the base of the evolutionary tree. The majority of species within Subgen. *Impatiens* were endemic to China, particularly concentrated in the southwestern and southern regions [50]. Furthermore, the distribution of Subgen. *Clavicarpa* was geographically specific [51].

In the subgenus *Impatiens*, the present study revealed complex morphological variation, accompanied by strong support values in its molecular phylogenetic tree. The subgenus *Impatiens* was primarily concentrated in southwestern China, particularly in Yunnan, Guizhou, and Sichuan, with a few species extending into central and northern China, Thailand, and other regions [52]. Notably, we found that the phylogenetic trees constructed using different methods clustered the following three cultivars—*I. hawkeri*, *I. walleriana*, and *I. balsamina*—into a distinct clade. These three cultivars occupied a relatively unique evolutionary position, which was consistent with the findings of Luo [53]. Furthermore, cultivated species exhibited a clear evolutionary trend compared to wild species, demonstrating significant genetic divergence and high levels of internal resolution. In this research, compelling molecular evidence was provided that the chloroplast genome could be effectively utilized for phylogenetic and taxonomic studies within or among species of *Impatiens*. Due to the large number of species in the *Impatiens*, relatively few chloroplast genomes of *Impatiens* were available. Although our phylogenetic tree resolved the major branching relationships within *Impatiens*, future studies will require more chloroplast genome data from *Impatiens* to adequately address the phylogenetic problem of *Impatiens*.

Ancestral trait reconstruction and ecological adaptation

This study was the first to trace six traits of *Impatiens* using chloroplast genomic data to investigate the ancestral character state. The results indicated that the traits of *Impatiens* were highly differentiated and exhibited significant diversity. However, there was a strong consistency of traits among species within the same clade. Our findings suggested that the traits of *Impatiens* may have

undergone a complex and variable evolutionary history, which was closely related to their wide distribution and ability to adapt to a variety of environments. Notably, the subgenus *Clavicarpa* may be diverged from the *Impatiens* lineage earlier, and formed a separate clade.

In this study, most of the species were perennials, while fewer evolved into annuals, such as those in branches I, V, and VII. Research was shown that perennials are better equipped to survive in resource-poor environments, whereas annuals tend to grow and reproduce more rapidly [54]. The ancestral character state of capsule shape remained ambiguous and exhibited significant homogeneity. Clavate and fusiform capsule shapes may represent more primitive forms. However, Song demonstrated that the ellipsoidal capsule shape was the ancestral state for the *Impatiens* [55]. This conclusion required further analysis based on a larger sample size.

Flower colour played a crucial role in the propagation of insectivorous plants and holds significant ecological value [56–58]. Ancestral trait reconstruction results suggested that the evolution of flower colour was a complex process and no clear pattern was found in this evolutionary process (Fig. 10a). In this study, we found that the majority of *Impatiens* species exhibited yellow flowers. The yellow petals' UV absorption pattern enhances visibility to many flower visitors [58–60], thereby facilitating pollination [61]. Through extensive experimental observations, we discovered that most *Impatiens* species possess floral spots on the petals near the pistil. This sharp colour contrast effectively served as a nectar guide for flower visitors [58, 62]. In *Turnera*, a similar phenomenon was observed [63]. We concluded that flower colour evolution was related to selection pressures on *Impatiens* during their evolutionary journey.

In *Impatiens*, the spur is a nectar-storing structure [64]. Spur lengths were predominantly greater than 1 cm, as indicated by the results of the ancestral character state analysis (Fig. 10b). Xiao's research suggested that the evolution of floral structures in *Impatiens* was influenced by the selection pressures exerted by pollinators. Pollinators play an irreplaceable role in the evolutionary history of floral structures [65]. The length of the spur is correlated with the mouthparts and body size of pollinators. Bees and moths with longer mouthparts are the primary pollinators of *Impatiens* in subtropical and tropical regions [66, 67]. Notably, 71% of the studied species exhibited bee pollination syndrome [68]. Interestingly, some species with very short or even absent spurs also were identified within *Impatiens*, leading researchers to speculate that their pollinators may be flies [69]. Furthermore, some studies shown that the spur (which stores nectar) in *I. uliginosa* did not significantly reduce the frequency of pollinator visits [70]. We speculated that the presence

or absence of a spur may not directly influence the frequency of pollinator visits.

In this study, the ancestral state of 4 sepals was considered, which was consistent with the findings of Yu [3]. The VIII branch (Subgen. *Clavycarpa*) all possessed 4 sepals and exhibited 3-colpus pollen. In the *Impatiens*, the number of pollen colpus has evolved from three to four [71]. Notably, the presence of three pollen colpus and four sepals was regarded as a relatively primitive state within *Impatiens* [72]. This all provides further data to support the idea that the Subgen. *Clavycarpa* is more primitive than the Subgen. *Impatiens*.

The results of ancestral trait reconstruction indicated that the individual traits of *Impatiens* exhibited different evolutionary rates. Song also demonstrated that mosaic evolution occurs within *Impatiens* [69]. The more rapid evolution observed in the subgenus *Impatiens* is associated with its wide distribution and strong adaptability. The subgenus *Clavycarpa* represented an early differentiated clade of *Impatiens* [49]. The subgenus *Clavycarpa* was discussed as the more primitive subgenus. At the same time, the subgenus *Clavycarpa* was primarily distributed in southwestern China Geographically. And this region was situated at one of the five major centers of *Impatiens* distribution worldwide [50, 73, 74]. So it was likely that southwestern China was one of the original centers of diversification for the *Impatiens*. The development of traits resulted from a combination of intrinsic plant characteristics and environmental factors. There was a clear evolutionary relationship between plants and their pollinators. Further analyses regarding the evolution of traits in *Impatiens* should incorporate additional samples and molecular datasets.

Conclusion

In this study, we sequenced and assembled the complete chloroplast genome sequences of 12 samples representing eight sections of species in the *Impatiens*. Comparative genomics studies by adding published samples indicated that the chloroplast genome of the *Impatiens* was relatively conserved, with the emergence of 10 mutational hotspot regions that could serve as potentially variable molecular markers for inferring physiological ontogenetic relationships and species identification. Phylogenetic analyses based on chloroplast genomes supported some of our previous results in taxonomic treatment studies using morphological characters. The world's *Impatiens* species were divided into two subgenus, and the *Impatiens* was further differentiated into eight sections. By mapping six morphological traits of the *Impatiens* into the phylogenetic tree, the life type of perennial, the number four sepals, and threes pollen colpus were probably the primitive traits of the *Impatiens*,

whereas the ancestral morphology of capsule shape, flower colour, and length of the flower spur were ambiguous. Overall, this study showed that the entire chloroplast genome sequence provided another perspectives on *Impatiens*'s genetic diversity and evolutionary history, which can contribute to resolve the phylogenetic relationships of this difficult-to-describe genus.

Methods

Sample collection and sequencing

In this study, we collected nine species of *Impatiens* from natural habitats in accordance with local and national regulations, and the collected samples were identified by Prof. Haiquan Huang of the College of Horticulture and Landscape Architecture, Southwest Forestry University, and preserved in the laboratory of the College of Horticulture and Landscape Architecture of the Southwest Forestry University (25°06'N, 102°76'E); the remaining three materials were obtained from the Laboratory of the College of Horticulture and Landscape Architecture of the Southwest Forestry University (25°06'N, 102°76'E).

The materials utilized in this study consisted of fresh leaves collected from Yunnan, Sichuan, and other regions. To better reflect the systematic position of the *Impatiens*, we selected eight groups of species based on their distribution locations and existing classifications from the "Flora of China" [73]. Additionally, we downloaded 31 published species of the *Impatiens* from the GenBank database, resulting in a total of 43 samples. Genomic DNA was extracted from 12 different samples of *Impatiens* using the CTAB method [75]. After confirming the quality of the genomic DNA, we performed sequencing on the chloroplast genomes of these 12 *Impatiens* species using the Illumina NovaSeq 6000 platform. The raw data were processed to remove adapter sequences and paired-end reads containing more than 10% of their length as N, as well as single-end reads with low-quality bases (Q ≤ 5) exceeding 50% of their length. Ultimately, clean sequencing data were obtained.

Chloroplast genome assembly and annotation

Using GetOrganelle (version 1.7.5.0) [76], we assembled the chloroplast genomes of 13 species within the *Impatiens*, applying the default parameters to obtain complete circular chloroplast genome sequences. The assembled fasta format files were subsequently submitted to the online annotation tool Cpgavas2 (<https://www.herbalgenomics.org/cpgavas2>) [77], and manual corrections to the annotations were performed using Geneious (R9.0.2) [78]. The physical maps of the chloroplast genomes were generated using OGDRAW (v1.3.1) [79]. All newly sequenced and annotated complete chloroplast genomes

have been deposited in GenBank, with the corresponding accession numbers listed in Table S1.

Repetitive sequences in chloroplast genomes analysis

Using the online software MISA (<https://webblast.ipkgatersleben.de/misa/index.php>) [80], we analyzed simple repeat sequences in the chloroplast genomes of 18 species of *Impatiens*. The parameters for the repeat units, ranging from one to six nucleotides, were set as follows: 10, 6, 4, 3, 3, and 3, with a minimum distance of 100 bp between two SSRs. Additionally, we utilized the online software REPuter (<https://bibiserv.cebitec.uni-bielefeld.de/reputer>) [81] to analyze dispersed repeat sequences in the cpDNA of the same 18 species of *Impatiens*. The parameters for this analysis included a minimum repeat sequence length of 30 bp, a Hamming distance of 3, and a sequence identity threshold of 90%.

Complete chloroplast analysis

Using *H. triflora* as a reference, we studied the genomic sequence specificity differences among 19 species of *Impatiens*. We employed the online tool mVISTA (<http://genome.lbl.gov/vista/mvista/submit.shtml>) [82] with the Shuffle-LAGAN model to compare variations in the coding regions, non-coding regions, introns, and exons of the chloroplast genomes within the *Impatiens*. This approach allowed us to visually reflect the similarities and differences among the species. To calculate the nucleotide diversity (Π) values for the 19 species, we utilized DnaSP V6 [83] with a sliding window of 600 bp and a step size of 200 bp. The results were visualized using Excel, and diversity hotspots corresponding to specific genes were annotated. Additionally, we employed IRScope [84] to illustrate the contraction and expansion of the IR/SC region boundaries.

Selective pressure analysis

Coding sequences (CDS) and protein sequences were extracted from the chloroplast genomes of 19 species within the *Impatiens*. Using BLASTN (v2.14.0+), the protein sequences were compared to reference protein sequences to identify the best matches, thereby obtaining homologous protein sequences. Subsequently, the homologous protein sequences were automatically aligned using MAFFT (v7.310) [85]. The aligned protein sequences were then mapped back to the coding sequences to obtain the aligned CDS. Based on the MLWL method, the K_a and K_s values were calculated using K_a/K_s -Calculator3.

For data analysis, Python (v3.9.6) and the pandas library (v2.2.2) were utilized to read the K_a/K_s values of multiple genes from Excel files. The normality of the K_a/K_s values for each gene was assessed using the Shapiro–Wilk test

(`scipy.stats.shapiro`, v1.9.1). A p -value greater than the significance level ($\alpha=0.05$) indicated that the data followed a normal distribution. For normally distributed data, a one-sample t -test (`scipy.stats.ttest_1samp`) was conducted to determine whether the average K_a/K_s value significantly differed from 1. In contrast, for non-normally distributed data, the Mann–Whitney U test (`scipy.stats.mannwhitneyu`) was employed to evaluate whether the K_a/K_s values significantly differed from 1. To estimate confidence intervals, the t -distribution (`scipy.stats.t.ppf`) was used to calculate the 95% confidence interval for normally distributed data, while the interquartile range (25th and 75th percentiles) served as a non-parametric confidence interval for non-normally distributed data. In the results analysis, a p -value less than the significance level ($\alpha=0.05$) was considered significant. If the t -test statistic was greater than zero and the result was significant, or if the Mann–Whitney U test indicated that the average K_a/K_s value was greater than 1 and the result was significant, it was inferred that there was positive selection pressure. Conversely, if the t -test statistic was less than zero and the result was significant, or if the Mann–Whitney U test indicated that the average K_a/K_s value was less than 1 and the result was significant, it was inferred that there was purifying selection pressure. If no significance was found, no inference regarding selection pressure could be made.

Phylogeny and ancestral trait reconstruction analysis

This study employed two methods for constructing the phylogenetic tree: Maximum Likelihood (ML) and Maximum Parsimony (MP). A total of 12 chloroplast genomes were assembled, and an additional 34 sequences were retrieved from the NCBI database, resulting in a total of 46 sequences used for phylogenetic tree construction. The CPGANA-toolkit was employed to standardize the orientation of the sequence regions by adjusting the SSC direction and starting point. Multiple sequence alignment was conducted using the MAFFT (v7.310) [85] software, and TBtools [86] was utilized to trim the alignment results. For the Maximum Likelihood analysis, the IQ-TREE [87] software was utilized to identify the best-fitting DNA/protein model, which was determined to be GTR+ F + R3. The Maximum Parsimony analysis was performed using MEG [88] software, employing the Subtree-Pruning-Regrafting (SPR) search method, with an initial tree of 20 and a search level set to 2. Bootstrap values were calculated based on 1000 iterations.

Select six morphological traits used in the lower classification of the *Impatiens* from different classification systems to trace their evolutionary history [50] Habit, Capsule shape, Flower colour, Spur length, Sepal number, Morphology and Number of Aperture. Relevant literature

was reviewed to determine the character states for each species (references). Detailed coding for each species can be found in Table S3. We tracked the evolution of the six traits using the threshold model in maximum parsimony in Mesquite v3.61.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12870-024-05964-y>.

Supplementary Material 1.

Acknowledgements

We thank the editor and the reviewers for their helpful remarks that improved this article.

Authors' contributions

LWX, MQ and JMM performed the experiments and analyzed the data. They performed the experiments and analyzed the data and wrote the main manuscript text. LWX, MQ, JMM, WYH, YMQ and MH, participated in the experiment. HHQ and HMJ supervised and revised the manuscript. All authors read and approved the manuscript.

Funding

This study was supported by the Key Project of Yunnan Provincial Agricultural Joint Special Program (202301BD070001-011), National Natural Science Foundation of China (grant number 32060364, 32060366), the Project of High-level Introduction talents in Yunnan Province, and First-rate Discipline Landscape Architecture Construction Project of Yunnan Province, China.

Data availability

New sequenced and other published chloroplast genome sequences can be found in GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>), Its accession number is shown in Table S1 in the Supplementary Material.

Declarations

Ethics approval and consent to participate

The collecting of all samples in this study followed the Regulations on the Protection of Wild Plants of China, the IUCN Policy Statement on Research Involving Species at Risk of Extinction and the Convention on the Trade in Endangered Species of Wild Fauna and Flora. All methods were carried out in accordance with relevant guidelines and regulations.

Consent for publications

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹College of Landscape Architecture and Horticulture Sciences, Southwest Research Center for Engineering Technology of Landscape Architecture (State Forestry and Grassland Administration), Yunnan Engineering Research Center for Functional Flower Resources and Industrialization, Research and Development Center of Landscape Plants and Horticulture Flowers, Southwest Forestry University, Kunming, Yunnan 650224, China.

Received: 27 October 2024 Accepted: 11 December 2024

Published online: 20 December 2024

References

- Chase MW, Christenhusz MJM, Fay MF, Byng JW, Judd WS, Soltis DE, et al. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical J Linnean Soc.* 2016;181(1).
- Hook JD. Les especes du genre '*Impatiens*' dans l'herbier du Museum de Paris. *Nouv. Arch. Mus. Nat. Hist. Paris, Ser.* 1908;4(10):233–272.
- Li ZZ, Saina JK, Gichira AW, Kyalo CM, Wang QF, Chen JM. Comparative Genomics of the Balsaminaceae Sister Genera *Hydrocera triflora* and *Impatiens pinfanensis*. *Int J Mol Sci.* 2018;19(1):319.
- Wang JS, Lu YF, Xu YL, Jin SH, Jin XF. *Impatiens wuyiensis* (Balsaminaceae), a new species from Fujian of Southeast China, based on morphological and molecular evidences. *Bot Stud.* 2020;61(1):29–29.
- Luo C, Huang WL, Sun HY, Yer H, Li XY, Li Y, et al. Comparative chloroplast genome analysis of *Impatiens* species (Balsaminaceae) in the karst area of China: insights into genome evolution and phylogenomic implications. *BMC Genomics.* 2021;22(1).
- Song YX, Peng S, Cong YY, Zheng YM. *Impatiens rapiformis*, a new species of *Impatiens* with root tuber from Yunnan, China. *Nordic Journal of Botany.* 2021;39(5).
- Fujihashi H, Akiyama S, Ohba H. Origin and relationships of the Sino-Himalayan *Impatiens* (Balsaminaceae) based on molecular phylogenetic analysis, chromosome numbers and gross morphology. 2002.
- Yuan YM, Song Y, Geuten K, Rahelivololona E, Wohlhauser S, Fischer E, et al. Phylogeny and biogeography of Balsaminaceae inferred from *ITS* sequence date. *Taxon.* 2004;53(2):391–403.
- Yu SX, Steven B. Janssens, Zhu XY, Magnus L, Gao TG, Wang W. Phylogeny of *Impatiens* (Balsaminaceae): integrating molecular and morphological evidence into a new classification. *Cladistics.* 2016;32(2).
- Xu XM, Huang HC, Lin SQ, Zhou LW, Yi YC, Lin EW, et al. Twelve newly assembled jasmine chloroplast genomes: unveiling genomic diversity, phylogenetic relationships and evolutionary patterns among *Oleaceae* and *Jasminum* species. *BMC Plant Biology.* 2024;24(1).
- Liu XF, Luo JJ, Chen H, Li TY, Qu TM, Tang M, et al. Comparative analysis of complete chloroplast genomes of *Synotis* species (Asteraceae, Senecioneae) for identification and phylogenetic analysis. *BMC Genomics.* 2024;25(1).
- Corriveau JL, Coleman AW. Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperm species. *Am J Bot.* 1988;75(10):1443–58.
- Zhang Q, Liu Y, Sodmergen. Examination of the cytoplasmic DNA in male reproductive cells to determine the potential for cytoplasmic inheritance in 295 angiosperm species. *Plant and Cell Physiology.* 2003;44(9):941–951.
- Perry AS, Wolfe KH. Nucleotide substitution rates in legume chloroplast DNA depend on the presence of the inverted repeat. *J Mol Evol.* 2002;55:501–8.
- Song C, Zhu J, Li H. Complete chloroplast genomes of eight *Delphinium* taxa (Ranunculaceae) endemic to njiang, China: insights into genome structure, comparative analysis, and phylogenetic relationships. *BMC Plant Biol.* 2024;24(1):600.
- Li EZ, Liu KJ, Deng RY, Gao YW, Liu XY, Dong WP, et al. Insights into the phylogeny and chloroplast genome evolution of *Eriocaulon* (Eriocaulaceae). *BMC plant Biology.* 2023;23(1):32–32.
- Li EL, Li HY, Li JX, Li J, Hu N, Sun J, et al. Chloroplast genomes of *Caragana tibetica* and *Caragana turkestanica*: structures and comparative analysis. *BMC Plant Biol.* 2024;24(1):254–254.
- Ran ZH, Li Z, Xiao X, An MT, Yan C. Complete chloroplast genomes of 13 species of sect. *Tuberculata* Chang (*Camellia* L.): genomic features, comparative analysis, and phylogenetic relationships. *BMC Genomics.* 2024;25(1):108–108.
- Li MD, Pan GY, Liu LH, Yu B, Huang D, Zhu GF. Thirteen complete chloroplast genomes of the costaceae family: insights into genome structure, selective pressure and phylogenetic relationships. *BMC Genomics.* 2024;25(1):68–68.
- Qiu H, Zhang ZH, Wang MZ, Jin XJ, Lin JD, Comes HP, et al. Plastome evolution and phylogenomics of *Impatiens* (Balsaminaceae). *Planta.* 2023;257(2).
- Luo C, Li Y, Budhathoki R, Shi JY, Yer H, Li XY, et al. Complete chloroplast genomes of *Impatiens cyanantha* and *Impatiens monticola*: Insights into genome structures, mutational hotspots, comparative and phylogenetic analysis with its congeneric species. *PLoS one.* 2021;16(4).
- Dong WP, Xu C, Wen J, Zhou SL. Evolutionary directions of single nucleotide substitutions and structural mutations in the chloroplast genomes of the family Calycanthaceae. *BMC Evol Biol.* 2020;20(1):1–12.

23. Dong WP, Sun JH, Liu YL, Xu C, Wang YH, Suo ZL, et al. Phylogenomic relationships and species identification of the olive genus *Olea* (Oleaceae). *J Syst Evol.* 2021;60(6):1263–80.
24. Li L, Hu YF, He M, Zhang B, Wu W, Cai PM, et al. Comparative chloroplast genomes: insights into the evolution of the chloroplast genome of *Camellia sinensis* and the phylogeny of *Camellia*. *BMC Genomics.* 2021;22(1).
25. Dong WP, Li EZ, Liu YL, Xu C, Wang YS, Liu KJ, et al. Phylogenomic approaches untangle early divergences and complex diversifications of the *olive* plant family. *BMC Biol.* 2022;20(1):92–92.
26. M TC, B SG, G HL, Jr BR. Rates and patterns of chloroplast DNA evolution. *Proc Nat Acad Sci.* 1994;91(15):6795–6801.
27. Zhu JQ, Huang Y, Chai WG, Xia PG. Decoding the Chloroplast Genome of *Tetragium* (Vitaceae): Variations and Phylogenetic Selection Insights. *Int J Mol Sci.* 2024;25(15).
28. Kim YK, Cheon SH, Hong JR, Kim KJ. Evolutionary Patterns of the Chloroplast Genome in *Vanilloid Orchids* (Vanilloideae, Orchidaceae). *Int J Mol Sci.* 2023;24(4).
29. Guo YY, Yang JX, Bai MZ, Zhang GQ, Liu ZJ. The chloroplast genome evolution of Venus slipper (*Paphiopedilum*): IR expansion, SSC contraction, and highly rearranged SSC regions. *BMC Plant Biol.* 2021;21(1):248–248.
30. Bureš P, Elliott LT, Veselý P, Šmarda P, Forest F, Leitch IJ, et al. The global distribution of angiosperm genome size is shaped by climate. *New Phytol.* 2024;242(2):744–59.
31. Peng Y, Yang JX, Leitch IJ, Guignard MS, Seabloom EW, Cao D, et al. Plant genome size modulates grassland community responses to multi-nutrient additions. *New Phytol.* 2022;236(6):2091–102.
32. Wang XJ, Dong WP, Zhou SL. The Evolution Path of *Medicago* in China based on the Chloroplast Genome Analysis. *Acta Ecol Sin.* 2022;42(15):6125–36.
33. Weng ML, Ruhlman TA, Jansen RK. Expansion of inverted repeat does not decrease substitution rates in *Pelargonium* plastid genomes. *New Phytol.* 2017;214(2):842–51.
34. Han L, Xia L, Chong S, Hong LL, Zhe XL, Yuan G, et al. Chloroplast Genome Comparison and Phylogenetic Analysis of the Commercial Variety *Actinidia chinensis* 'Hongyang'. *Genes.* 2023;14(12).
35. Gao HH, Liang YH, Yao MZ, Zhang PF, Liu YL. Development of Genome-Wide SSR Molecular Markers and Analysis of Population Genetic Diversity in *Astragalus mongolicus*. *Acta Agrestia Sinica.* 2024;27(12):1–14.
36. Boakewaa AG, Badu-Apraku B, Akromah R, Garcia-Oliveira AL, Awuku FJ, Gedil M. Genetic diversity and population structure of early-maturing tropical maize inbred lines using SNP markers. *PLoS ONE.* 2019;14(4):e0214810.
37. Mukhopadhyay P, Reddy MK, Singla-Pareek SL, Sopory SK. Transcriptional downregulation of rice *rpl32* gene under abiotic stress is associated with removal of transcription factors within the promoter region. *PLoS one.* 2011;6(11).
38. Minoru U, Masaru F, Shin-ichi A, Jin M, Nobuhiro T, Koh-ichi 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(1).
39. Ellen ML, Gao ZY, Susanne P, Laure S, Adam A, Oliver V, et al. Multiple Instances of Ancient Balancing Selection Shared Between Humans and Chimpanzees. *Science.* 2013;339(6127).
40. Kashi Y, King DG. Simple sequence repeats as advantageous mutators in evolution. *Trends Genet.* 2006;22(5):253–9.
41. Pauwels M, Vekemans X, Godé C, Frérot H, Castric V, Laprade PS. Nuclear and chloroplast DNA phylogeography reveals vicariance among European populations of the model species for the study of metal tolerance, *Arabidopsis halleri* (Brassicaceae). *New Phytol.* 2012;193(4):916–28.
42. Moe AM, Weiblen GD. Development and characterization of microsatellite loci in dioecious figs (*Ficus*, Moraceae). *Am J Bot.* 2011;98(2):e25–7.
43. Wang XT, Zhang YJ, Qiao L, Chen B. Comparative analyses of simple sequence repeats (SSRs) in 23 mosquito species genomes: Identification, characterization and distribution (Diptera: Culicidae). *Insect science.* 2019;26(4):607–19.
44. Zhang TX, Li MM, Zhu XY, Li SX, Guo MY, Guo CH, et al. Comparative Chloroplast Genomes Analysis Provided Adaptive Evolution Insights in *Medicago ruthenica*. *Int J Mol Sci.* 2024;25(16):8689.
45. Wang G, Ren Y, Su YY, Zhang H, Li JF, Han JP. Molecular marker development and phylogenetic analysis of *Aconitum* species based on chloroplast genomes. *Ind Crops Prod.* 2024;221: 119386.
46. Shimizu T. A comment on the limestone flora of Thailand, special reference to *Impatiens*. *Acta Phytotaxa Geobotanica.* 1979;30(1):180–8.
47. Richardson JE, Fay MF, Cronk QB. A phylogenetic analysis of Rhamnaceae using *rbcL* and *trnL-F* plastid DNA sequences. *Am J Bot.* 2000;87:1309–24.
48. Elisette MR, Eberhard F, Steven BJ, Sylvain GR. Phylogeny, infrageneric classification and species delimitation in the Malagasy *Impatiens* (Balsaminaceae). *PhytoKeys.* 2018;110(110).
49. Xia CY. Phylogeny study and taxonomic revision of *Impatiens* subg. *Clavicarpa*: Southwest University; 2020.
50. Yu SX. Balsaminaceae of China. Peking University Press. 2012.
51. BHASKAR V, RAZI BA. A new kind of exine sculpturing in *Impatiens* L. (Balsaminaceae) from south India. *Curr Sci.* 1973;42:510–512.
52. Zeng L, Yan RY, Zhang M, Xu WB, Zhang LJ, Yu SX. Taxonomic significance of the pollen morphology of Subg. *Clavicarpa* (*Impatiens*, Balsaminaceae). *Guihaia.* 2016;36(10):1245–1252.
53. Luo C, Huang WL, Yer HY, Kamuda T, Li XY, Li Y, et al. Complete Chloroplast Genomes and Comparative Analyses of Three Ornamental *Impatiens* Species. *Frontiers in Genetics.* 2022; 13.
54. Paleo LG, Pastor AP, Rajnoch G, Ravetta DA. Mechanisms of nitrogen conservation at the leaf-level in annual and perennial desert forbs: implications for perennial crops domestication. *Flora.* 2019;252:62–8.
55. Song YX, Peng S, Mutie FM, Jiang H, Ren J, Cong YY, et al. Evolution and Taxonomic Significance of Seed Micromorphology in *Impatiens* (Balsaminaceae). *Front Plant Sci.* 2022;13:835943–835943.
56. Fenster CB, Armbruster WS, Wilson P, et al. Pollination syndromes and floral specialization. *Annu Rev Ecol Evol Syst.* 2004;35(1):375–403.
57. Gong Y, Huang S. On methodology of foraging behavior of pollinating insects. *Biodiversity Science.* 2007;15(6):576.
58. Lunau K, Wacht S, Chittka L. Colour choices of naive bumble bees and their implications for colour perception. *J Comp Physiol A.* 1996;178:477–89.
59. Briscoe AD, Chittka L. The evolution of color vision in insects. *Annu Rev Entomol.* 2001;46(1):471–510.
60. Papiorek S, Junker RR, Alves-Dos-Santos I, Melo GAR, Amaral-Neto LP. Bees, birds and yellow flowers: pollinator-dependent convergent evolution of UV patterns. *Plant Biol (Stuttg).* 2016;18(1):46–55.
61. María JG, Francisco P, Peter CK. The role of pollinator diversity in the evolution of corolla-shape integration in a pollination-generalist plant clade. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences.* 2014;369(1649):20130257.
62. Li BZ. Pollination ecology and adaptation research of four *Impatiens* L. (Balsaminaceae). Master's Dissertation. Guizhou: Guizhou Normal University. 2022.
63. Lamarck R, Patrícia LR, Peter KE, Alessandro R. A brainstorm on the systematics of *Turnera* (Turneraceae, Malpighiales) caused by insights from molecular phylogenetics and morphological evolution. *Molecular Phylogenetics and Evolution.* 2019;137.
64. Li Y, Huang WL, Li XY, Zhang YD, Meng DC, Wei CM, et al. The cellular and molecular basis of the spur development in *Impatiens uliginosa*. *Horticulture research.* 2024;11(3).
65. Xiao LX. The pollination biology of four species of *Impatiens* L. Master's Dissertation. Hunan: Hunan Normal University. 2009.
66. Saroj R, Pornpimon T, Ruth JC, Erik FS, Timotheüs N. Floral specialization for different pollinators and divergent use of the same pollinator among co-occurring *Impatiens* species (Balsaminaceae) from Southeast Asia. *Botanical Journal of the Linnean Society.* 2016;181(4).
67. Tang YF, Fang Y, Liu CQ, Lu QB, Hu XH. The long spur of *Impatiens macrovexilla* may reflect adaptation to diurnal hawkmoth pollinators despite diversity of floral visitors. *Flora.* 2020;266: 151599.
68. Ruchisansakun S, Mertens A, Janssens SB, Smets EF, Timotheüs VDN. Evolution of pollination syndromes and corolla symmetry in Balsaminaceae reconstructed using phylogenetic comparative analyses. *Annals of botany.* 2020;127(2).
69. Song YX, Hu T, Peng S, Cong YY, Hu GW. Palynological and macroscopic characters evidence infer the evolutionary history and insight into pollination adaptation in *Impatiens* (Balsaminaceae). *Journal of Systematics and Evolution.* 2023;62(3).

70. Li B. Z. Pollination ecology and adaptation Research of four *Impatiens* L.(Balsaminaceae). Guizhou Normal University. 2022.
71. Janssens S, Lens F, Dressler S, Geuten K, Smets E, Vinckier S. Palynological variation in Balsaminoid Ericales. II. Balsaminaceae, Tetrameristaceae, Pellicieraceae and general conclusions. *Annals of botany*. 2005;96(6).
72. Lu YQ. Pollen morphology of *Impatiens* L. (Balsaminaceae) and its taxonomic implication. *Acta Phytotaxonomica Sinica*. 1991;29(4):352–357.
73. Chen Y. L. Flora of China: vol. 47, fascicle 2. Science Press. 2001;49–212.
74. Chen Y. Geographic distribution patterns of wild *Impatiens* L. plants and their dominant environmental factors in China. Guizhou University. 2023.
75. Doyle JJ, Egan AN. Dating the origins of polyploidy events. *New Phytol*. 2010;186(1):73–85.
76. Jin JJ, Yu WB, Yang JB, Song Y, dePamphilis Claude W, Yi TS, et al. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome biology*. 2020;21(1):31–36.
77. Shi LC, Chen HM, Jiang M, Wang LQ, Wu X, Huang LF, et al. CPGAVAS2 an integrated plastome sequence annotator and analyzer. *Nucleic Acids Res*. 2019;47:67–73.
78. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* (Oxford, England). 2012;28(12).
79. Stephan G, Pascal L, Ralph B. Organellar Genome DRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic acids research*. 2019;47(W1):W59–W64.
80. Thiel T, Michalek W, Varshney R K, Graner A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik*. 2003;106(3).
81. Kurtz S, Choudhuri J V, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic acids research*. 2001;29(22).
82. Inna D, V DR. VISTA family of computational tools for comparative analysis of DNA sequences and whole genomes. *Methods in molecular biology* (Clifton, N.J.). 2006;33869–89.
83. Rozas J, Sánchez DB, Messeguer X, Rozas RD. DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* (Oxford, England). 2003;19(18).
84. Ali A, Jaakko H, Peter P. IRscope: an online program to visualize the junction sites of chloroplast genomes *Bioinformatics* (Oxford, England). 2018;34(17):3030–3031.
85. Kazutaka K, M D S. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*. 2013;30(4):772–80.
86. Chen CJ, Chen H, Zhang Y, Thomas HR, Frank MH, He YH, et al. TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Mol Plant*. 2020;13(8):1194–202.
87. Nguyen LT, Schmidt HA, Arndt H, Quang MB. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 2015;32(1):268–74.
88. Koichiro T, Glen S, Sudhir K. MEGA11: Molecular Evolutionary Genetics Analysis version 11. *Mol Biol Evol*. 2021;38(7):3022–7.

Publisher's Note

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