Planta (2023) 257:45
https://doi.org/10.1007/s00425-023-04078-3

ORIGINAL ARTICLE

Plastome evolution and phylogenomics of *Impatiens* (Balsaminaceae)

He Qiu[1] · Zhi-Hui Zhang[1] · Mei-Zhen Wang[1] · Xin-Jie Jin[2] · Jie-Dong Lin[3] · Hans Peter Comes[4] · Jing-Xuan Chen[1] · Rui-Ning Cui[1] · Ru-Qing Duan[1] · Pan Li[1]

Received: 20 September 2022 / Accepted: 16 January 2023 / Published online: 25 January 2023

Abstract

Main conclusion **This study reported seven new plastomes from *Impatiens* and observed three highly variable regions for phylogeny and DNA barcoding, which resolved the relationships among sections of subgenus *Impatiens*.**

Abstract *Impatiens* L. (Balsaminaceae, Ericales) is one of the largest and most diverse genera of angiosperms, widely known for its taxonomic difficulty. In this study, we reevaluated the infrageneric relationships within the genus *Impatiens*, using complete plastome sequence data. Seven complete plastomes of *Impatiens* (representing 6 species) were newly sequenced and characterized along with 20 previously published plastomes of other *Impatiens* species, plus 2 plastomes of outgroups (*Hydrocera triflora*, Balsaminaceae; *Marcgravia coriacea*, Marcgraviaceae). The total size of these 29 plastomes ranged from 151,538 bp to 152,917 bp, except 2 samples of *Impatiens morsei*, which exhibited a shorter length and lost some genes encoding NADH dehydrogenase subunits. Moreover, the number of simple sequence repeats (SSRs) ranged from 51 to 113, and the number of long repeats from 17 to 26. In addition, three highly variable regions were identified (*trn*G-*GCC* (The previous one), *ndh*F–*rpl*32–*trn*L-*UGA*–*ccs*A, and *ycf*1). Our phylogenomic analysis based on 80 plastome-derived protein-coding genes strongly supported the monophyly of *Impatiens* and its two subgenera (*Clavicarpa* and *Impatiens*), and fully resolved relationships among the six (out of seven) sampled sections of subgenus *Impatiens*. Overall, the plastome DNA markers and phylogenetic results reported in this study will facilitate future identification, taxonomic and DNA barcoding studies in *Impatiens* as well as evolutionary studies in Balsaminaceae.

Keywords *Impatiens* · Balsaminaceae · Plastome · Comparative analysis · Phylogeny · Molecular markers

Communicated by Anastasios Melis.

✉ Pan Li
panli_zju@126.com

He Qiu
qiuhe2333@qq.com

Zhi-Hui Zhang
6253591zzh@gmail.com

Mei-Zhen Wang
21907104@zju.edu.cn

Xin-Jie Jin
xinjie_jin@yeah.net

Jie-Dong Lin
ljd067@163.com

Hans Peter Comes
hans-peter.comes@plus.ac.at

Jing-Xuan Chen
3200100364@zju.edu.cn

Rui-Ning Cui
1985984859@qq.com

Ru-Qing Duan
2690835668@qq.com

[1] Laboratory of Systematic and Evolutionary Botany and Biodiversity, College of Life Sciences, Zhejiang University, Hangzhou 310058, China

[2] College of Life and Environmental Science, Wenzhou University, Wenzhou 325035, China

[3] Zhangzhou Huaqiao Middle School, Zhangzhou 363030, China

[4] Department of Environment and Biodiversity, Salzburg University, 5020 Salzburg, Austria

Introduction

The Balsaminaceae A. Rich. are a family of dicotyledonous plants within the order of Ericales Bercht. and J. Presl, comprising two genera: the monotypic *Hydrocera* Blume from Southeast Asia (*Hydrocera triflora* (Linn.) Wight et Arn.) and *Impatiens* Riv. ex L. (Byng et al. 2016). The latter is one of the most diverse genera of angiosperms with more than 1000 species (Byng et al. 2016; Yu et al. 2016), which are mostly distributed in highland and mountainous regions of the Old World tropics and subtropics (Grey-Wilson 1980; Janssens et al. 2012). However, some species of *Impatiens* also occur in the temperate regions of Europe, North America and China (Janssens et al. 2006; Yu et al. 2016). Recently, numerous new species of the genus have been identified in China, especially in the karst areas, which are considered as biodiversity hotspots (Li et al. 2018; Wang et al. 2020, 2022a, b; Luo et al. 2021; Song et al. 2021). Several species of *Impatiens*, such as *I. balsamina* L., have their stems used in traditional Chinese medicine (TCM) to deal with fingernail inflammation, beriberi and onychomycosis, and their juices are commonly used to dye nails (Jiang et al. 2017; Kim et al. 2017). In addition, previous research showed that some metals such as cadmium and zinc are accumulated by *Impatiens* species (Li et al. 2018; Luo et al. 2021). Because of the unique and novel posture of its flowers like a flying phoenix, it is known as the orchid of dicotyledons and owns high ornamental value (Yin et al. 2018; Fu et al. 2022).

Impatiens was once considered as one of the most difficult group in angiosperm taxonomy (Hooker 1908). Morphologically, *Impatiens* is well known as difficult to classify due to its fragile flowers, semi-succulent stems and mainly fleshy leaves, which makes the separation and reconstruction of different parts laborious, and the production of well-dried herbarium specimens challengeable (Yu et al. 2016; Rahelivololona et al. 2018). Micromorphological characteristics, such as features of pollen and seeds, were included in some morphological studies, but rarely combined with phylogeny (Janssens et al. 2012; Cai et al. 2013; Yu et al. 2016). The infrageneric relationships within *Impatiens* are not well resolved, especially the division of subgenera and sections. Fujihashi et al. (2002) published the first molecular phylogeny of *Impatiens* based on combined plastid sequences of the *rbc*L gene and the *trn*L–F spacer region. Subsequently, Yuan et al. (2004) used internal transcribed spacer (ITS) sequences of nuclear ribosomal (nr) DNA to reconstruct the phylogeny of 111 species of *Impatiens* and divided 16 clades for the main lineage of *Impatiens*, yet the relationship among most of the clades could not be resolved with confident support, probably due to an insufficient number of informative sites (Fujihashi et al. 2002; Yuan et al. 2004; Luo et al. 2021). Furthermore, Janssens et al. (2006), using the plastid *atp*B–*rbc*L spacer sequence, produced a fine tuning of phylogenetic relationships which defined 15 clades among *Impatiens* (Janssens et al. 2006). More recently, Yu et al. (2016) presented a comprehensive phylogenetic analysis of *Impatiens* based on morphological characters combined with nuclear (ITS) and plastid (*atp*B–*rbc*L and *trn*L–F) sequence data; according to their results, *Impatiens* can be subdivided into two major clades, corresponding to subgenus *Clavicarpa* S.X. Yu and subgenus *Impatiens* Warb., with the latter comprising seven sections, among which some relationships are not resolved (Yu et al. 2016).

However, all the above studies were mainly based on morphology and/or limited molecular data. With the development of next-generation sequencing technologies, the acquisition of complete plastomes is nowadays possible. Recently, numerous studies have used whole-plastome sequence data to solve the intrageneric phylogenetic relationships (Amenu et al. 2022; Guo et al. 2022; Song et al. 2022; Yang et al. 2022). Along this line, Luo et al. (2021) based on plastome resolved the phylogeny of *Impatiens* under the framework of Ericales, but this research only included 11 *Impatiens* species and the resolution of relationship within the *Impatiens* subgenus and sections with limited support (Luo et al. 2021).

Here, we newly sequenced and assembled seven plastomes of *Impatiens*, representing 6 species, and downloaded previously published plastomes of 20 extra species of the genus, together covering both subgenera and 6 of the 7 sections (sensu (Yu et al. 2016)), plus 2 outgroups. Based on these 29 plastomes, we aimed to (1) compare and analyze the structural features of *Impatiens* plastomes; and (2) elucidate major relationships among subgenera and sections within the genus.

Materials and methods

Taxon sampling, DNA isolation, plastome sequencing and de novo assembly

Previously published plastomes of 20 species of *Impatiens* were downloaded from the National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov) (Table 1). In addition, we newly sequenced plastomes of seven samples, representing six *Impatiens* species (see Table 1 for GenBank accession numbers, including two accessions ('1' vs. '2') of *I. morsei*). All these new samples were collected from China, with voucher specimens deposited in the Herbarium of Zhejiang University (HZU), China (Table S1). Based on recent studies (Li et al. 2018;

Table 1 Summary of 29 complete plastomes used in this study

Taxa	Accession number	Length (bp)				GC (%)				Gene number			
		Plastome	LSC	IR	SSC	Plastome	LSC	IR	SSC	Full	PCG	tRNA	rRNA
Impatiens alpicola	NC_053940.1	151,366	82,245	25,708	17,705	36.9	34.6	43.1	29.4	104	69	31	4
I. arguta	**OP022554**	152,075	82,781	25,786	17,724	36.9	34.7	43.1	29.2	114	80	30	4
I. balsamina	NC_059942.1	152,271	83,497	25,249	18,276	36.7	34.3	43.2	29.3	114	81	30	4
I. chlorosepala 1	NC_059943.1	152,763	83,740	25,773	17,477	36.7	34.3	43.1	29.5	115	81	30	4
***I. chlorosepala* 2**	**ON186542**	152,917	83,691	25,746	17,737	36.7	34.3	43.1	29.4	114	80	30	4
I. conchibracteata	**OP022555**	152,142	82,889	25,731	17,793	36.8	34.5	43.1	29.3	114	80	30	4
I. cyanantha	NC_058204.1	152,391	83,284	25,653	17,801	36.8	34.5	43.1	29.6	115	81	30	4
I. davidii	NC_058801.1	152,214	83,128	25,634	17,818	36.9	34.6	43.1	29.4	108	79	25	4
I. fanjingshanica	NC_059944.1	151,538	82,542	25,726	17,547	36.9	34.6	43.1	29.4	115	81	30	4
I. glandulifera	NC_044718.1	152,260	83,261	25,631	17,737	36.8	34.5	43.1	29.4	108	80	29	4
I. guizhouensis	NC_059945.1	152,774	83,572	25,772	17,662	37.0	34.8	43.0	29.9	115	81	30	4
I. hawkeri	NC_048520.1	151,691	83,030	25,584	17,493	36.8	34.4	43.2	29.6	114	81	30	4
I. huangyanensis	**OP022556**	152,155	82,982	25,768	17,639	36.8	34.5	43.1	29.3	114	80	30	4
I. linearisepala	NC_059946.1	152,212	83,508	25,699	17,309	37.0	34.8	43.0	30.0	115	81	30	4
I. loulanensis	NC_059947	152,472	83,460	25,737	79,560	36.7	34.4	43.0	29.6	115	81	30	4
I. macrovexilla	NC_060668.1	152,437	83,331	25,865	17,376	36.8	34.5	43.0	29.2	114	80	30	4
I. macrovexilla var. *yaoshanensis*	NC_060669.1	152,286	83,212	25,881	17,312	36.8	34.5	43.0	29.3	114	80	30	4
I. mengtszeana	NC_058215.1	152,928	83,722	26,007	17,192	36.7	34.3	43.0	29.5	115	81	30	4
I. monticola	NC_058205.1	152,656	83,740	25,664	17,588	36.7	34.3	43.1	29.5	115	81	30	4
***I. morsei* 1**	**OP022557**	146,234	82,748	24,303	14,882	36.8	34.4	43.4	28.7	108	74	30	4
***I. morsei* 2**	**OP022558**	148,277	82,765	25,702	14,110	37.0	34.4	43.2	29.1	111	77	30	4
I. piufanensis	NC_037401.1	152,236	83,115	25,755	17,611	36.9	34.5	43.1	29.3	114	81	30	4
I. pritzelii	NC_047191.1	152,487	83,290	25,684	17,829	37.0	34.9	43.1	29.8	113	79	30	4
***I.* sp.**	**OP022559**	151,939	82,740	25,769	17,663	36.8	34.5	43.1	29.4	114	80	30	4
I. stenosepala	NC_059948.1	152,802	83,626	25,720	17,739	36.9	34.5	43.2	29.8	115	81	30	4
I. uliginosa	NC_059760.1	152,609	83,365	25,871	17,502	36.8	34.5	29.6	42.9	115	81	30	4
I. walleriana	NC_059949.1	151,953	82,906	25,710	17,627	36.8	34.4	43.2	29.4	114	81	30	4
Hydrocera triflora	NC_037400.1	154,189	84,865	25,622	18,082	36.9	34.7	43.1	29.9	115	80	31	4
Marcgravia coriacea	NC_041255	158,401	87,116	26,255	18,775	36.7	34.5	42.8	30.0	114	80	30	4

Newly generated sequences are in bold

Luo et al. 2021), plastomes of *Hydrocera triflora* (Linn.) Wight et Arn. (Balsaminaceae) and *Marcgravia coriacea* Vahl (Marcgraviaceae) were selected as outgroups, and also downloaded from NCBI. Balsaminaceae and Marcgraviaceae, together with Tetrameristaceae, are generally referred to as 'balsaminoid clade' within Ericales (Byng et al. 2016).

Total genomic DNA was extracted from approximately 10–20 mg silica-dried leaf material of each of the seven samples (six species), using Plant DNAzol Reagent (Life-Feng, Hangzhou, China) according to the manufacturer's protocol. The qualities and quantities of extracted DNA were determined using agarose gel electrophoresis and ultraviolet-microspectrophotometry. Approximately, 1 µg of extracted DNA with a concentration higher than 12.5 ng/µl was sent to the Beijing Genomics Institute (BGI, Wuhan, China) for whole-genome sequencing. Before sequencing, total DNA was sheared into fragments shorter than 800 bp. The qualities of obtained DNA fragments were evaluated using Agilent Bioanalyzer 2100 (Agilent Technologies) and the pooled library was run in a single lane on an Illumina HiSeq X10 platform to obtain raw reads with about 150 bp in length.

The high-quality reads were used for subsequent plastome assembly and annotation. According to the GETORGANELLE pipeline (Jin et al. 2020), we *de novo* assembled the plastome using *Impatiens fanjingshanica* Y.L. Chen (GenBank accession number: MW411294.1) as a seed, applying a k-mer gradient (– k 21, 45, 65, 85, 105) for automatic calling SPAdes v.3.13.1, Bowtie2 v.2.4.1 and BLAST v.2.12.0 (Bankevich et al. 2012), which first filtered low-quality reads and adapters before conducting the *de novo* assembly. Then, we checked the assembly results using BANDAGE v.0.8.1 to obtain a complete plastome (Wick et al. 2015).

Genome annotation

In order to generate the draft of plastome annotation, we used the 'Map to Reference' function with the parameter 'High Sensitivity' in GENEIOUS Prime v.2020.0.5 (Kearse et al. 2012), and the genome of *I. fanjingshanica* was used as the reference. We used the MAFFT plugin v1.4.0 (Katoh and Standley 2013) in GENEIOUS with default parameters to align the sequences with the reference, and then annotated them. The presence of start and stop codons of each protein-coding gene (PCG) was checked and adjusted manually, those genes with any premature termination codon, which may interrupt the translation of the original reading frame, were annotated as pseudogenes and excluded from the phylogenetic analyses (see below). Circular plastome maps were visualized using OGDRAW v.1.3.1 (Greiner et al. 2019).

Analysis of long sequence repeats

We used the online tool REPuter (Kurtz et al. 2001) to identify 4 kinds of sequence repeats (i.e., forward, reverse, complement, and palindromic) in the 27 *Impatiens* plastomes. The Hamming distance (i.e., the number of bit positions in which the two bits are different) was set to three, and the minimal repeat size was limited to 30, while the other settings were retained as default. In addition, we searched 27 *Impatiens* plastomes for simple sequence repeats (SSRs; or microsatellites), using the online tool MISA-web (Beier et al. 2017). The minimum numbers for mono-, di-, tri-, tetra-, penta-, and hexanucleotide repeats were set as 10, 5, 4, 3, 3, and 3, respectively.

Comparative analysis of plastome sequence divergence

We checked the gene order of all 29 plastomes (including outgroups), using the MAUVE plugin v.1.1.3 (Darling et al. 2004) in GENEIOUS. To study the contraction or expansion of inverted repeated (IR) regions in *Impatiens*, we used IRscope (Amiryousefi et al. 2018) and its perl script (https://www.github.com/xul962464/perl-IRscope) for visualization and manual checking of IR boundaries.

Levels of nucleotide diversity (*Pi*) were calculated using DnaSP v.6.12.03 (Rozas et al. 2017) with a step size of 200 bp and a window length of 800 bp, in order to explore highly variable regions across the 27 *Impatiens* plastomes. For three highly variable regions identified (see Results), we also used IQ-TREE v.2.0.3 (Minh et al. 2020) to assess their marker discriminatory power *inter-se* and in relation to the whole-plastome data under a best-fitting (automatically selected) model and using 1000 bootstrap replicates. Finally, for examining codon usage bias, i.e., the preference for certain synonymous codon, we calculated values of relative synonymous codon usage (RSCU) for a total of 80 PCGs in each of the 27 *Impatiens* plastomes, using the program codonW v.1.4.2 (http://www.codonw.sourceforge.net/culong.html#CodonW).

Phylogenetic analyses

For phylogenetic analyses, we used a total of 80 PCGs that were extracted from the 7 newly sequenced plastomes and all other available plastomes of *Impatiens* from NCBI (Table S1), plus those of the outgroups *H. triflora* and *M. coriacea*. For the alignment, we used the MAFFT plugin in GENEIOUS with default settings, and manually removed those regions with more than 50% missing data. Phylogenetic relationships were inferred using two methods. First, we performed a Maximum Likelihood (ML) analysis in IQ-TREE under the optimal substitution models, which was

automatically selected by the parameter -m MPF + MERGE. For Adequate sampling, 1000 bootstrap replicates were generated. The R package PHYTOOLS (Revell 2011) was used to generate 240 possible gene-by-codon position partition files (80 genes × three codon positions) and the partition files was applied using the parameter -p in IQ-TREE to exclude the influence of different codon variation rates on phylogenetic relationship inference. In addition, we performed a Bayesian Inference (BI) analysis in MRBAYES v.3.2.7 on the CIPRES Science Gateway (Miller et al. 2010), using 2 independent Markov Chain Monte Carlo (MCMC) runs, each consisting of 4 chains of 1,000,000 generations and sampled every 1000 generations. TRACER v.1.7.2 (Rambaut et al. 2018) was used to assess convergence of the MCMC parameters, ensuring their sufficient effective sample size (ESS) ≥ 200. The final tree with bootstrap support (BS) and posterior probability (PP) values was visualized using FIGTREE v.1.4.4 (Rambaut et al. 2018).

Results

Features of the newly sequenced *Impatiens* plastomes

For the seven newly sequenced plastomes, representing six *Impatiens* species, the number of paired-end clean reads per sample ranged from 9,205,472 (*I. huangyanensis*) to 36,897,312 (*I. morsei* 2), with a coverage from 63.8 × (*I. huangyanensis*) to 190.9 × (*I. morsei* 1) (Table S1). The length of the plastomes varied from 146,234 to 152,917 bp. All seven plastomes possessed the typical angiosperm quadripartite structure (Fig. 1), with very similar lengths of the two IR regions (IRa/IRb; 24,946–26,388 bp), the large single-copy (LSC) region (81,459–82,982 bp), and the small single-copy (SSC) region (14,044–17,793 bp). Across these seven plastomes, the total GC content ranged from 36.7% (*I. chlorosepala* 2) to 37.0% (*I. morsei* 2), with the IR regions showing the highest GC content (42.9–43.1%), followed by the LSC region (34.3–35.7%) and the SSC region (29.1–29.9%) (Table 1). Almost all newly assembled plastomes contained 114 genes, including 80 PCGs, 30 tRNA genes, and 4 rRNA genes; the only exception was *I. morsei*, of which the 2 accessions analyzed were lacking 6 and 3 genes in the SSC region, respectively (Fig. 1; Table 1; Table 2).

Analysis of long sequence repeats

Across the seven newly assembled *Impatiens* plastomes, the total number of long repeat sequences identified ranged from 17 to 26 (Table S2). Among those, forward (F) repeats were most abundant (7–17), closely followed by palindromic (P) repeats (6–13), while reverse (R) and complement (C) repeats were rare (0–2 and 0–1, respectively; Fig. 2A). Overall, the length of repeats varied from 30 to 216 bp, yet most of them were 30–39 bp in length, while only a small part (4 out of 550) fell into the 50–59 bp category (Fig. 2B).

Among the 27 plastomes of *Impatiens*, the total number of SSRs ranged from 51 to 113, with mononucleotide repeats being the most abundant (Fig. 3A; Table S3). Notably, in all these 27 plastomes, the great majority of SSRs resided in the LSC region (Fig. 3B).

We used the 80 PCGs in all of 27 *Impatiens* plastomes, the number of leucine (Leu) was the most abundant, accounting for 10.29% (*I. morsei* 1) to 10.82% (*I. cyanantha*) of total selected amino acids (Table S4). Considering estimates of relative synonymous codon usage (RSCU; Fig. 4), we found that 30 codons had RSCU values > 1.00 (indicates that the actual frequency of the codon is higher than the theoretical frequency), and most of them (29) ended with A or U, excepting 'UUG', which codes for leucine.

Comparative plastome analysis

The visualization analysis of the alignment using MAUVE showed that the genomic order and orientation were highly conserved without observed rearrangements, except for slight variations in size and gene positioning (Fig. S1). Across 27 plastomes (i.e., 27 *Impatiens* ssp.), levels nucleotide diversity (*Pi*) ranged from 0.0005 to 0.0976, with an average of 0.0228 (Fig. 5; Table S5). Based on a cutoff *Pi* value ≥ 0.06, we identified three highly variable regions, including *trn*G-*GCC*, *ndh*F–*rpl*32–*trn*L-*UGA*–*ccs*A and *ycf*1 (Fig. 5; Table S5). The two identified variable regions located in the SSC region (i.e., *ndhF–rpl32–trn*L-*UGA–ccsA* and *ycf1*) had higher *Pi* values (and, respectively) than *trn*G-*GCC*, while the intergenic spacer between *ndh*F and *rpl*32 possessed the highest *Pi* value (0.0976) (Fig. 5; Table S5). In addition, it was obvious that the *Pi* value of the IR regions was lower than that of the SSC and LSC regions (Fig. 5). Based on the discriminatory power analysis in IQ-TREE (Fig S2–S5), *ndh*F–*rpl*32-*trn*L-*UGA*–*ccs*A and *ycf*1 exhibited the highest discriminatory power among the three highly variable regions and in relation to the whole-plastome data.

For the analysis of IR expansion and contraction (Fig. S2), we used all 27 plastomes of *Impatiens* with *Hydrocera triflora* (Balsaminaceae) and *Marcgravia coriacea* (Marcgraviaceae) as outgroups. All of the *Impatiens* plastomes had similar IR boundary features, with particular genes mainly found at the junctions (i.e., *rpl*22, *rps*19, *rpl*2, *rps3*, *ycf*1, *ndh*F, *trn*H, and *psb*A). However, in all analyzed species of *Impatiens*, the LSC–IRb junction was located in the *rps*19 gene (with 0–200 bp extending into the IRb region), while in *Marcgravia coriacea* this gene was fully embedded in the IRb region. Across all

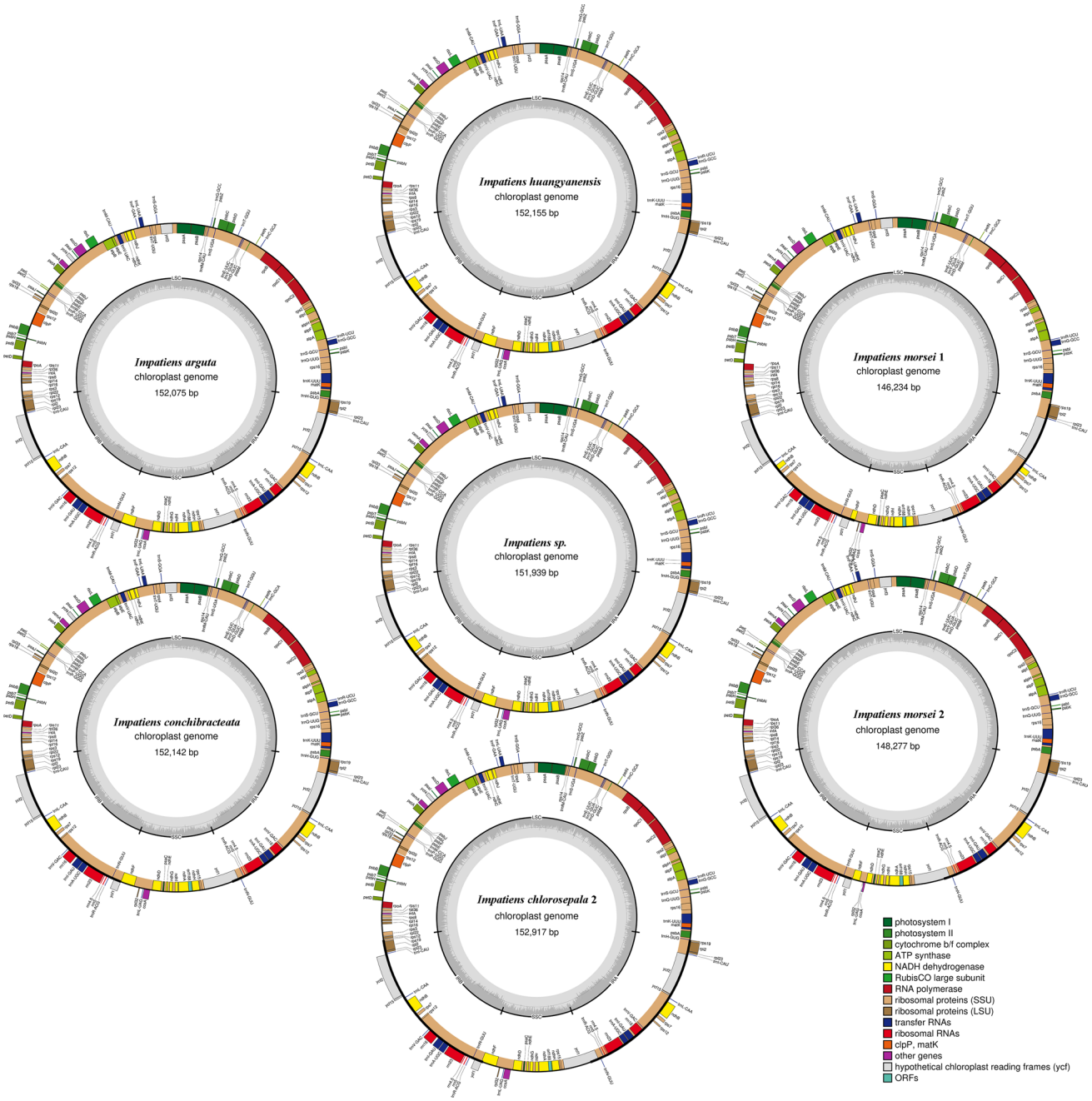

Fig. 1 Gene maps of seven newly assembled *Impatiens* plastomes (accession numbers: ON186542, OP022555, OP022556, OP022559, OP022554, OP022557, OP022558; see also Table 1). The genes located inside the circles are transcribed in a clockwise direction, while those outside the circle are transcribed counterclockwise. Genes are colored based on their functional groups as identified in the legend. The inner circle shows the quadripartite structure of the plastid genome. The gray ring marks the GC content, with the inner circle marking a 50% threshold

plastomes, the SSC-IRa junction was always located in the gene *ycf*1 (with 4221 to 4585 bp in SSC region and 1039 to 1688 bp in IRa region) except in *I. davidii*, in which the whole gene *ycf*1 was located in SSC, with 254 bp to the SSC-IRa junction.

Phylogenetic analyses

Here, phylogenetic analyses based on complete plastome sequences, 80 PCG datasets were performed, the results were similar despite the position of *Impatiens arguta*, but

Table 2 Genes in the plastomes of *Impatiens* species

Group of genes	Name of genes
ATP synthase	*atpA*, *atpB*, *atpE*, *atpF*[a], *atpH*, *atpI*
NADH dehydrogenase	*ndhA* [ac], *ndhB* [acd] (×2), *ndhC*, *ndhD*[d], *ndhE*, *ndhF*[cd], *ndhG*[c], *ndhH*[c], *ndhI*, *ndhJ*, *ndhK*
Cytochrome b6/f complex	*petA*, *petB*[a], *petD*[a], *petG*, *petL*, *petN*
Photosystem I	*psaA*, *psaB*, *psaC*, *psaI*, *psaJ*
Photosystem II	*psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbH*, *psbI*, *psbJ*, *psbK*, *psbL*, *psbM*, *psbN*, *psbT*, *psbZ*
Rubisco	*rbcL*
Subunit of Acetyl-CoA-carboxylase	*accD*
c-type cytochrome synthesis gene	*ccsA*
Envelop membrane protein	*cemA*
Protease	*clpP*[b]
Translational initiation	*infA*
Maturase	*matK*
Large subunit of ribosome	*rpl2*[a] (×2), *rpl14*, *rpl16*[a], *rpl20*, *rpl22*, *rpl23* (×2), *rpl32*, *rpl33*, *rpl36*
DNA-dependent RNA polymerase	*rpoA*, *rpoB*, *rpoC1*[a], *rpoC2*
Small subunit of ribosome	*rps2*, *rps3*, *rps4*, *rps7* (×2), *rps8*, *rps11*, *rps12*[b] (×2), *rps14*, *rps15*, *rps16*[a], *rps18*, *rps19*
rRNA genes	*rrn4.5* (×2), *rrn5*(×2), *rrn16*(×2), *rrn23*(×2)
tRNA genes	*trnA-UGC* (×2), *trnC-GCA*, *trnD-GUC*, *trnE-UUC*, *trnF-GAA*, *trnfM-CAU*, *trnG-GCC*, *trnH-GUG*, *trnI-CAU* (×2), *trnI-GAU* (×2), *trnK-UUU*, *trnL-CAA* (×2), *trnL-UAA*, *trnL-UAG*, *trnM-CAU*, *trnN-GUU* (×2), *trnP-UGG*, *trnP-GGG*, *trnQ-UUG*, *trnR-ACG* (×2), *trnR-UCU*, *trnS-GCU*, *trnS-GGA*, *trnS-UGA*, *trnT-GGU*, *trnT-UGU*, *trnV-GAC* (×2), *trnV-UAC*, *trnW-CCA*, *trnY-GUA*
Unknown function	*ycf1*, *ycf2* (×2), *ycf3*[b], *ycf4*, *ycf15* (×2)

(×2) Duplicated genes located within the inverted repeat (IR) regions

[a]Genes containing a single intron

[b]Genes containing two introns

[c]Genes lost in *I. morsei* 1

[d]Genes lost in *I. morsei* 2

in the analysis based on complete plastome, this node was weakly supported (BS = 38) while in the analysis based on the 80 PCG dataset this node had stronger support (BS = 72, PP = 0.99) (Fig. 6; Fig. S3). The identical topologies of the ML and BI trees (Fig. 6) had strong internal support: only three nodes had BS values < 90, while all the remaining nodes had BS values of 100, and only two nodes had PP values < 1.0. The phylogenetic trees consistently revealed that *Impatiens* split into two major clades (subgenera *Clavicarpa* and *Impatiens*) with maximum support (BS = 100, PP = 1.0), and thus in line with previous studies (Yu et al. 2016; Luo et al. 2021). Based on the phylogenetic framework by Yu et al. (2016), the species selected here included six (out of seven) sections in subgen. *Impatiens*, these sections were all monophyletic and largely well resolved, despite conflicting positions between sect. *Impatiens* and sect. *Fasciculatae* (Yu et al. 2016).

Moreover, within the genus, we recovered the major split between subgenera *Clavicarpa* and *Impatiens*, and all six (out of seven) sections sampled for the latter subgenus (except sect. *Tuberosae*) were highly supported as monophyletic and well resolved. Considering sectional relationships, sects. *Semeiocardium* + *Racemosae* proved sister to a grade comprising sects. *Impatiens*, *Fasciculatae* and the sister group *Scorpioidae* + *Uniflorae*.

Discussion

Plastome features

Here, we compared 27 complete plastomes of *Impatiens* (25 species), including 7 newly assembled ones, along with those of *Hydrocera triflora* and *Marcgravia coriacea*, serving as outgroups. The total size of most plastomes ranged from 151,538 bp to 152,917 bp, excepting those of the two accessions of *Impatiens morsei* ('1' and '2'), which were of much smaller size (146,236 bp and 148,277 bp, respectively). The outgroup *Hydrocera triflora* has a plastome of 154,189 bp in length, compared with which the plastome length of *Impatiens* species were reduced (Li et al. 2018). Overall, these 29 plastomes contained between 108 and 115 genes, including 74–81 protein-coding genes (PCGs), 25–31 tRNA genes, and 4

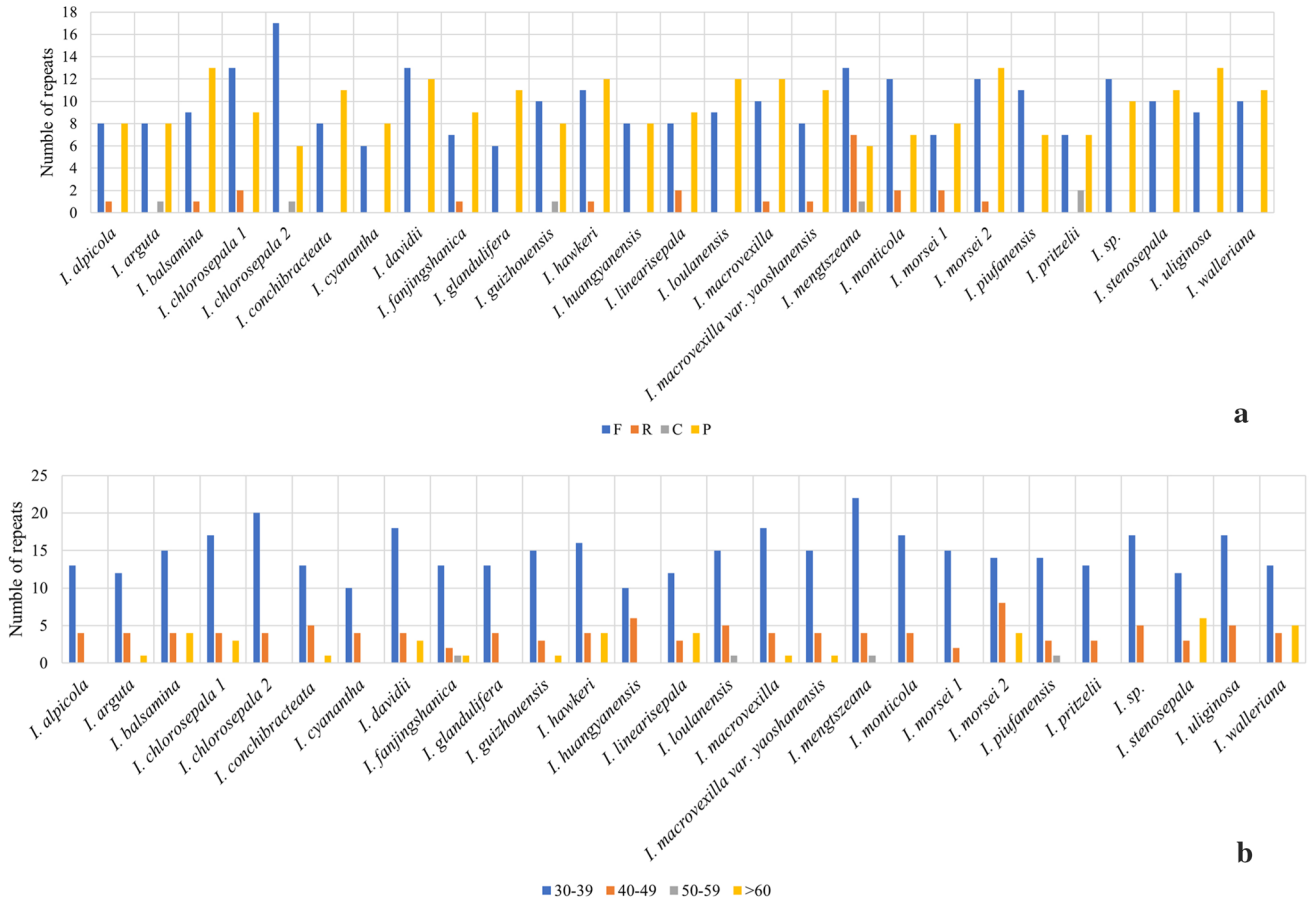

Fig. 2 Analyses of long repeat sequences in 27 complete plastomes of *Impatiens*. **A** Number of four kinds of repeats. *F* forward, *R* reverse, *C* complement, *P* palindromic. **B** Number of repeats within four different length categories (in bp)

rRNA genes. All of them exhibited the typical quadripartite structure of most angiosperm plastomes, including a pair of inverted repeats (IRs) as well as a large (LSC) and small (SSC) single-copy region (Guo et al. 2022; Song et al. 2022; Sun et al. 2022). The contraction and expansion of IR boundaries was previously proposed as the major reason for the reduced plastomes of *Impatiens* relative to *Hydrocera* (Luo et al. 2021). However, the extreme plastome downsizing in *I. morsei* is also due to the lack of genes encoding NADH dehydrogenase subunits (*ndh*A, *ndh*B (×2), *ndh*F, *ndh*G, *ndh*H in *I. morsei* 1; *ndh*B (×2), *ndh*D, *ndh*F in *I. morsei* 2), and which usually reside in the SSC region (except *ndh*B). Many previous studies have also observed the loss of *ndh* genes, demonstrating that they might be dispensable for some photoautotrophic plant species (e.g., orchids, *Genlisea* and *Selaginella*; Lin et al. 2017; Silva et al. 2018; Xu et al. 2018). In our case, the two sequenced *I. morsei* individuals are cultivated ones which have different phenotypes, and are probably from different wild populations (Fig. 7A and E). *Impatiens morsei* is such a polymorphic species (Fig. 7), which makes it understandable to have such variations of *ndh* genes loss within it. In addition, the GC content was unevenly distributed across the 29 plastomes, and generally higher in the IR regions than in the LSC and SSC regions, possibly because of the higher GC content of four rRNAs in the IRs (Shen et al. 2017; Zhang et al. 2022a, b).

Long repeat sequences are of great importance for inducing indels and identifying mutational hotspots (Ahmed et al. 2012; Ren et al. 2022). In this study, based on 27 *Impatiens* plastomes, we demonstrated that forward (F) and palindromic (P) repeats were far more abundant than reverse (R) or complement (C) repeats, and the lengths of all four types usually ranged from 30 to 39 bp, similar to recent studies (Li et al. 2018; Luo et al. 2021).

Simple sequence repeats (SSRs) are widely used as molecular markers due to their high variability and reproducibility (Zalapa et al. 2012; Ramzan et al. 2020). In this study, 4 to 6 types of SSRs were found across the 27 plastomes of *Impatiens*. These SSRs mainly found in the LSC region, while the fewest SSRs were in the IRs, and mononucleotides were the most dominant among all of the SSRs,

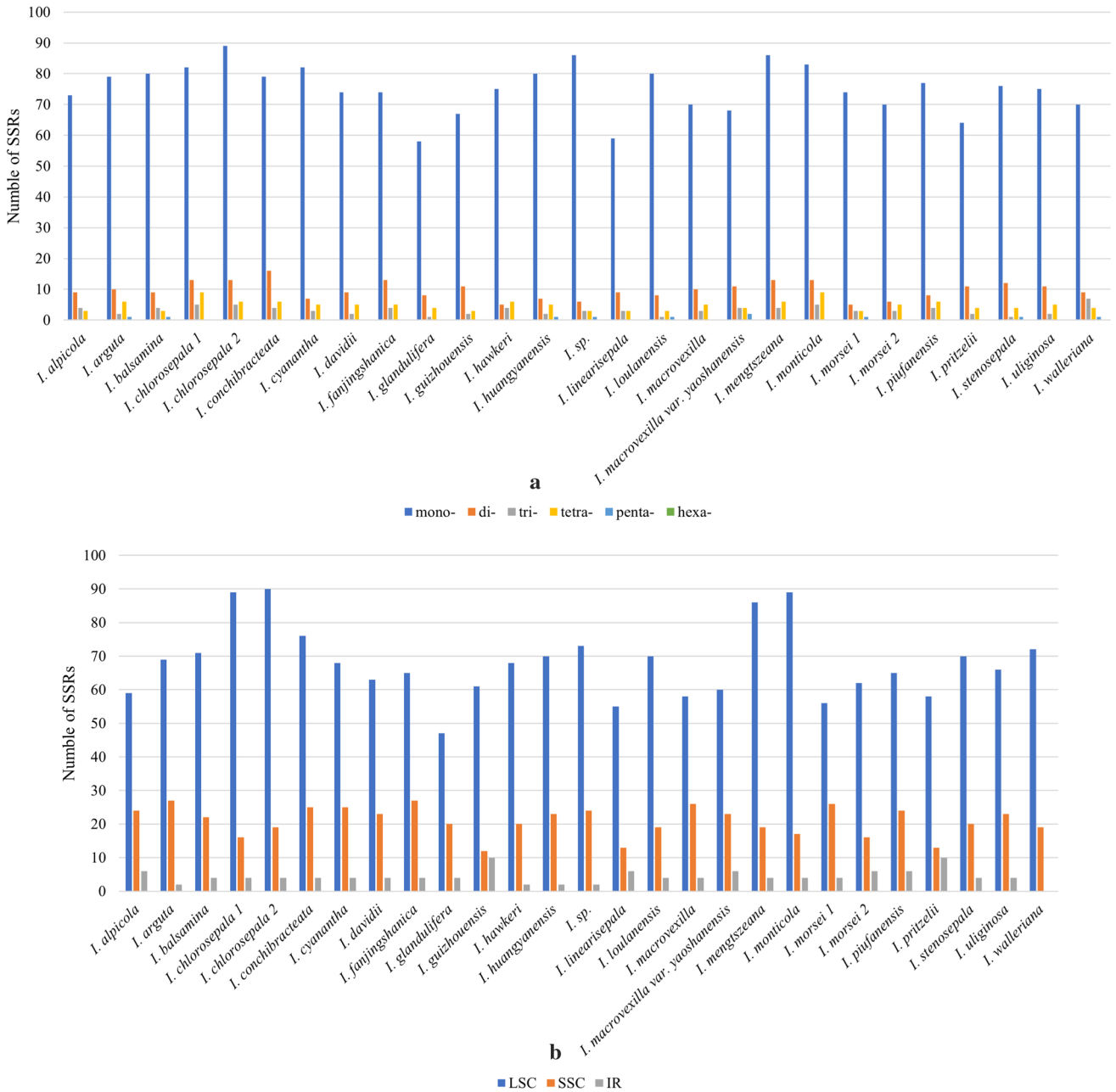

Fig. 3 Analysis of simple sequence repeats (SSRs) in 27 complete plastomes of *Impatiens*. **A** The number of different SSR types. **B** The overall number of SSRs detected in particular plastome regions. *LSC* large single-copy, *SSC* small single-copy, IRs inverted repeats

consistent with other studies (Li et al. 2018; Luo et al. 2021; Liu et al. 2022a, b; Yang et al. 2022).

In general, strong codon usage bias is often found in highly expressed protein encoding genes and thus usually correlates with mRNA and protein levels genome-wide (Zhou et al. 2013; Lyu and Liu 2020). The most abundant codons of *Impatiens* were those for leucine, consistent with numerous other studies (Somaratne et al. 2020; Ren et al. 2022; Zhang et al. 2022a, b). In this study, A/U-end codons were preferred in *Impatiens*, as 29 of 30 codons whose RSCU value more than 1.00 were ended with A or U, which might have contributed to the molecular evolution (Sharp and Wen-Hsiung 1986; Zhang et al. 2022a, b).

Identification of highly variable plastome regions

Recently, Luo et al. (2021) demonstrated that *trn*G-*GCC* (The previous one) and *ycf*1 are two highly variable plastid regions of potential use as DNA barcode markers in *Impatiens*. Here, based on 27 *Impatiens* plastomes, we

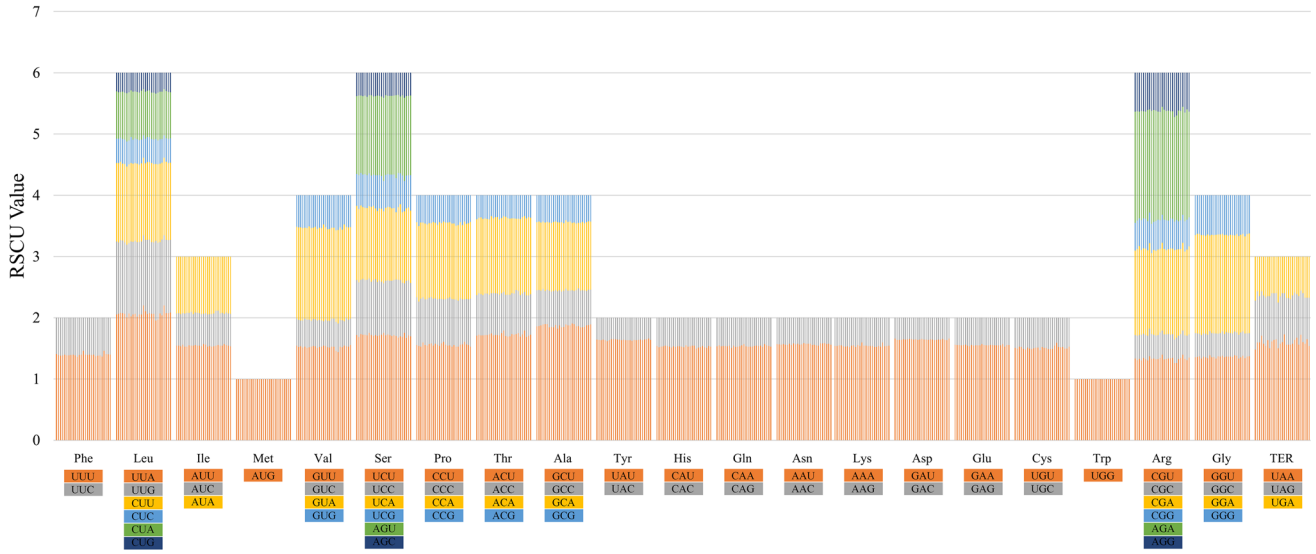

Fig. 4 Comparison of codon usage bias among the 80 protein-coding genes of 27 plastomes of *Impatiens*. The 27 histograms for each amino acid from left to right are: *I. alpicola*, *I. arguta*, *I. balsamina*, *I. cyanantha*, *I. chlorosepala* 1, *I. chlorosepala* 2, *I. conchibracteata*, *I. davidii*, *I. fanjingshanica*, *I. glandulifera*, *I. guizhouensis*, *I. hawkeri*, *I. huangyanensis*, *I. linearisepala*, *I. loulanensis*, *I. macrovexilla* var. *yaoshanensis*, *I. macrovexilla*, *I. mengtszeana*, *I. monticola*, *I. morsei* 1, *I. morsei* 2, *I. piufanensis*, *I. pritzelii*, *I.* sp., *I. stenosepala*, *I. uliginos* and *I. walleriana*. The height of histograms represents the value of relative synonymous codon usage (RSCU)

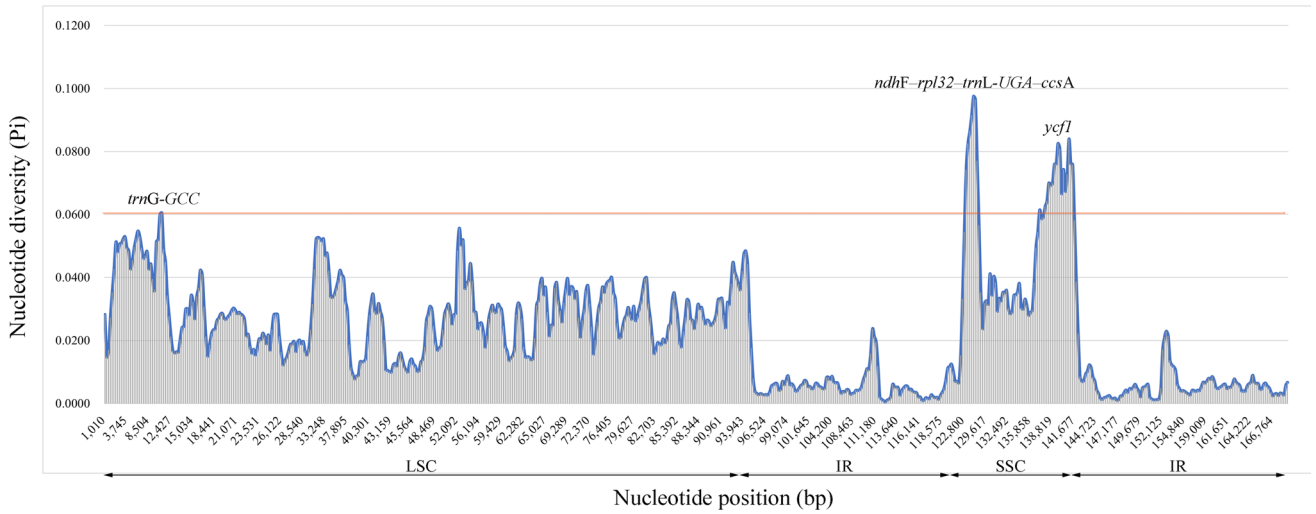

Fig. 5 Sliding window analysis of nucleotide diversity (*Pi*) based on the 27 complete plastomes of *Impatiens*, the red line in the middle means the cutoff value

additionally identified *ndh*F–*rpl*32–*trn*L-*UGA*–*ccs*A as a third highly variable region, based on a cutoff *Pi* value ≥ 0.06 (Fig. 5). However, phylogeny reconstructions based on these 3 regions showed that *ycf*1 recovered a similar phylogeny with 80 PCGs (Fig. S4, Fig. 6), while *ndh*F and *trn*G-GCC could not resolve the monophyly of the section *Semeiocardium* (Fig. S5–7). This might suggest that *ycf*1 has the greatest potential as DNA barcode marker for the genus *Impatiens*.

Phylogenetic relationships among sections of *Impatiens*

Impatiens was widely known as a taxonomically controversial genus morphologically especially owing to its fragile and complex floral structures (Chen 1978; Li et al. 2018; Luo et al. 2021). In most angiosperms, including *Impatiens*, the plastome is uniparentally inherited, and despite its lack of recombination and generally low mutation rate possesses

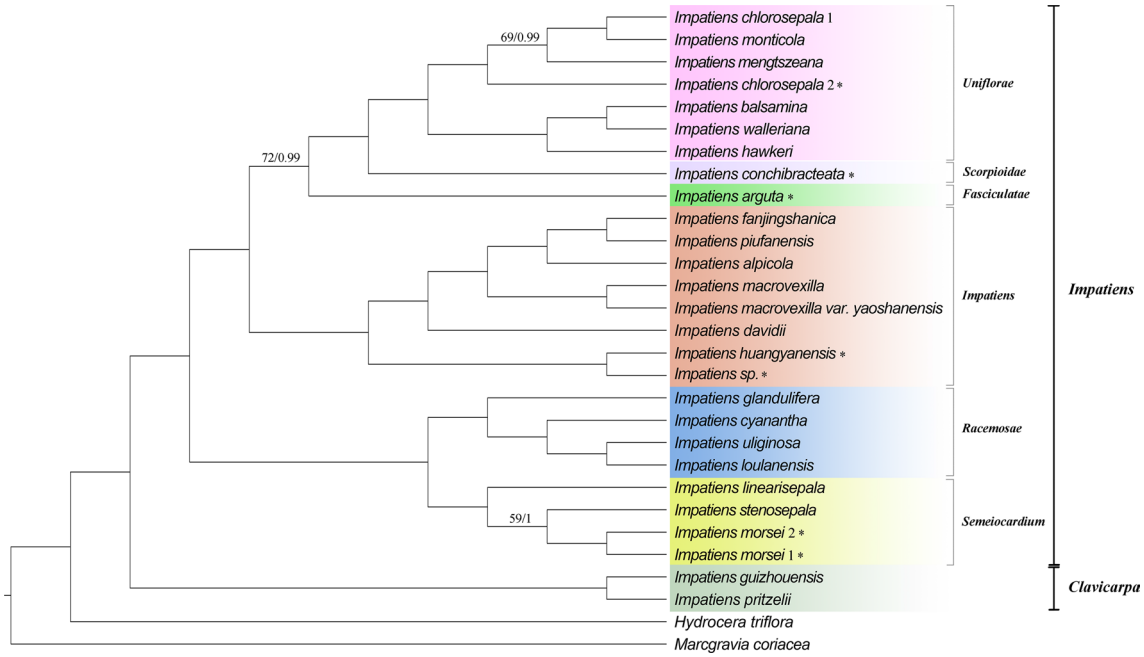

Fig. 6 Bayesian Inference (BI) tree of *Impatiens* (plus 2 outgroup species) based on 80 plastome-derived protein-coding genes (PCGs). Note that the Maximum Likelihood (ML) tree had the same topology. Bootstrap support and posterior probability values are shown above branches (BS/PP). Branches without values are supported by BS = 100 and PP = 1.0. Asterisks for *Impatiens* newly assembled in this study

highly variable gene (especially intergenic spacer) regions, with the potential to resolve phylogenetic relationships of families and even genera (Bock 2007; Shaw et al. 2014; Schneider et al. 2021; Liu et al. 2022a, b). However, previous molecular phylogenetic studies in *Impatiens* either used short DNA sequences or suffered from limited taxon sampling, resulting in restricted and/or conflicting conclusions (Fujihashi et al. 2002; Yuan et al. 2004; Janssens et al. 2006; Yu et al. 2016; Luo et al. 2021).

Therefore, we here performed phylogenomic (ML and BI) analyses of *Impatiens* based on 80 plastome-derived PCGs of 25 species (27 accessions), and with *Hydrocera triflora* and *Marcgravia coriacea* serving as outgroups (Fig. 6). This phylogeny recovered *Impatiens* as strongly supported sister clade (BS = 100, PP = 1.0) of the monotypic genus *Hydrocera*, in line with previous studies (Yu et al. 2016; Rahelivololona et al. 2018; Luo et al. 2021). Moreover, within the genus, we recovered the major split between subgenera *Clavicarpa* and *Impatiens*, and all six (out of seven) sections sampled for the latter subgenus (except that sect. *Tuberosae was not sampled*) were highly supported as monophyletic and well resolved. Considering sectional relationships, sects. *Semeiocardium* + *Racemosae* proved sister to a grade comprising sects. *Impatiens*, *Fasciculatae* and the sister group *Scorpioidae* + *Uniflorae*. We note that the sectional relationships of our tree topology (Fig. 6) are also broadly reflected in the comparative plastome analyses of Ericales by Luo et al. (2021), even though this study included only 11 species of *Impatiens* and only 4 sections of subgenus *Impatiens*. Overall, these relationships largely concur with the intrageneric classification of *Impatiens* by Yu et al. (2016), based on combined morphological and DNA sequence data (ITS; *atp*B–*rbc*L, *trn*L–F), except for a reversing position of sects. *Fasciculatae* and *Impatiens* (see Fig. 6), although the support for this node is not strong enough (BS = 72, PP = 0.99). Future studies involving target enrichment or RNA-seq might be able to fully resolve the backbone phylogeny of *Impatiens*.

Conclusions

In this study, we investigated 27 complete plastomes of *Impatiens* species and reached important conclusions, which could provide insights into the plastome structure and phylogeny of this genus. The plastomes exhibited the typical quadripartite structure and showed highly similar sizes, GC contents, gene orders and functions. However, *Impatiens morsei* exhibited a smaller plastome, largely due to the loss of several genes (mostly in the SSC region) encoding NADH dehydrogenase subunits. Three highly variable regions were identified that can be further used as DNA barcode markers. In addition, our phylogenomic analysis of 80 plastome-derived protein-coding genes

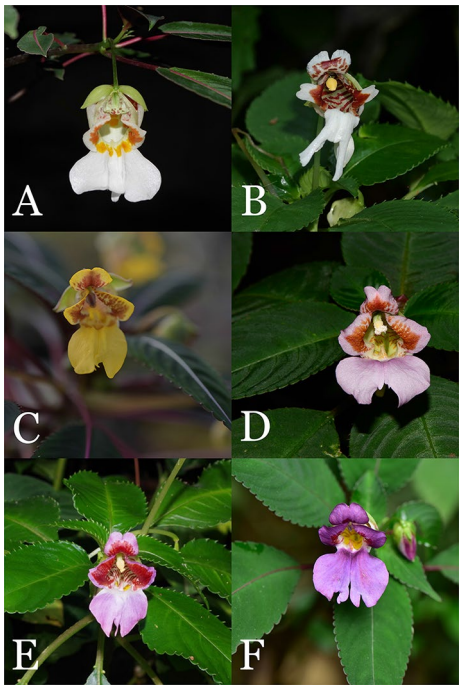

Fig. 7 Polymorphism in *Impatiens morsei*. **A** Cultivated individual (*Impatiens morsei* 1 in Fig. 6) from Guangxi Institute of Botany, China; **B** Wild individual from Daxin County, China; **C** Cultivated individual from Shanghai Chenshan Botanical Garden, originally collected from Tiandeng County, China; **D** Cultivated individual from Kunming Institute of Botany, China; **E** Cultivated individual (*Impatiens morsei* 2 in Fig. 6) from Guangxi Institute of Botany, originally collected from Napo County, China; **F** Wild individual from Jingxi County, China. Photo credits: **A** and **E**, Pan Li; **B**, Jun Li; **C**, Binjie Ge; **D** and **F**, Xinxin Zhu

(PCGs) better resolved major subgeneric and sectional relationships within *Impatiens* than previous studies using either 'traditional' DNA sequence information or comparative plastome data. Overall, the results reported in this study should facilitate future identification, taxonomic and DNA barcoding studies in *Impatiens* as well as evolutionary studies in Balsaminaceae.

Author contribution statement PL and JDL conducted the sampling. HQ, ZHZ, MZW, XJJ, RNC and RQD analyzed the data. JXC prepared the photo plate. HQ, PL, HPC and XJJ wrote the manuscript. ZHZ, MZW, JDL, JXC, RNC and RQD revised the manuscript. All the authors approved the final manuscript.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00425-023-04078-3.

Acknowledgements We thank Dr. Zhe-Chen Qi, Yu Feng, Rui-Sen Lu, Mr. Zi-Han Liu and Hai-Lei Zheng for helping with the fieldwork. This research was supported by the National Natural Science Foundation of China (Grant No. 31970225).

Data availability The 7 plastome sequence data generated in this study are available in GenBank of the National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/nuccore) under the access numbers: ON186542, OP022554–OP022559.

Declarations

Conflict of interest The authors declare no conflict of interest.

Ethical approval Not applicable.

Consent to participate All the authors consent to participate this study.

Consent for publication All the authors consent for this publication.

References

Ahmed I, Biggs PJ, Matthews PJ, Collins LJ, Hendy MD, Lockhart PJ (2012) Mutational dynamics of aroid chloroplast genomes. Genome Biol Evol 4:1316–1323. https://doi.org/10.1093/gbe/evs110

Amenu SG, Wei N, Wu L, Oyebanji O, Hu GW, Zhou YD, Wang QF (2022) Phylogenomic and comparative analyses of Coffeeae alliance (Rubiaceae): deep insights into phylogenetic relationships and plastome evolution. BMC Plant Biol 22:88. https://doi.org/10.1186/s12870-022-03480-5

Amiryousefi A, Hyvönen J, Poczai P (2018) IRscope: an online program to visualize the junction sites of chloroplast genomes. Bioinformatics 34:3030–3031. https://doi.org/10.1093/bioinformatics/bty220

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021

Beier S, Thiel T, Münch T, Scholz U, Mascher M (2017) MISA-web: a web server for microsatellite prediction. Bioinformatics 33:2583–2585. https://doi.org/10.1093/bioinformatics/btx198

Bock R (2007). In: Bock R (ed) Structure, function, and inheritance of plastid genomes. Springer, Berlin, pp 29–63

Byng JW, Chase MW, Christenhusz M, Fay MF, Judd WS, Mabberley DJ, Sennikov AN, Soltis DE, Soltis PS, Stevens PF, Briggs B, Brockington S, Chautems A, Clark JC, Conran J, Haston E, Moller M, Moore M, Olmstead R, Perret M, Skog L, Smith J, Tank D, Vorontsova M, Weber A, Angiosperm PG (2016) An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG IV. Bot J Linn Soc 181:1–20. https://doi.org/10.1111/boj.12385

Cai X, Yi R, Zhuang Y, Cong Y, Kuang R, Liu K (2013) Seed coat micromorphology characteristics of *Impatiens* L. and its systematic significance. Acta Horticulturae Sinica 40:1337–1348

Chen YL (1978) Notulae de genere *Impatiens* L. florae Sinicae. Acta Phytotax Sin 16(2):36–55

Darling AC, Mau B, Blattner FR, Perna NT (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res 14:1394–1403. https://doi.org/10.1101/gr.2289704

Fu CH, Chen XY, Li TJ, Zhou AP (2022) Characterizing the complete chloroplast genome of the *Impatiens davidii* (Balsaminaceae).

Mitochondrial DNA B 7:466–467. https://doi.org/10.1080/23802359.2022.2048211
Fujihashi H, Akiyama S, Ohba H (2002) Origin and relationships of the Sino-Himalayan *Impatiens* (Balsaminaceae) based on molecular phylogenetic analysis, chromosome numbers and gross morphology. J Jpn Bot 77:284–295
Greiner S, Lehwark P, Bock R (2019) OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res 47:W59–W64. https://doi.org/10.1093/nar/gkz238
Grey-Wilson C (1980) *Impatiens* in Papuasia: studies in Balsaminaceae: I. Kew Bull 34:661
Guo X, Qu X, Zhang X, Fan S (2022) Comparative and phylogenetic analysis of complete plastomes among aristidoideae species (Poaceae). Biology 11:63. https://doi.org/10.3390/biology11010063
Hooker JD (1908) Les espèces du genre "*Impatiens*" dans l'herbier du Muséum de Paris. Nouvelles Archives Du Muséum D'histoire Naturelle 4 10:233-272: 39
Janssens S, Geuten K, Yuan YM, Song Y, Kupfer P, Smets E (2006) Phylogenetics of impatiens and *Hydrocera* (Balsaminaceae) using chloroplast atpB-rbcL Spacer sequences. Syst Bot 31:171–180. https://doi.org/10.1600/036364406775971796
Janssens SB, Smets EF, Vrijdaghs A (2012) Floral development of Hydrocera and Impatiens reveals evolutionary trends in the most early diverged lineages of the Balsaminaceae. Ann Bot-London 109:1285–1296. https://doi.org/10.1093/aob/mcs065
Jiang H, Zhuang Z, Hou B, Shi B, Shu C, Chen L, Shi G, Zhang W (2017) Adverse effects of hydroalcoholic extracts and the major components in the stems of *Impatiens balsamina* L. on *Caenorhabditis elegans*. Evid-Based Compl Alt 2017:1–10. https://doi.org/10.1155/2017/4245830
Jin JJ, Yu WB, Yang JB, Song Y, DePamphilis CW, Yi TS, Li DZ (2020) GetOrganelle: a fast and versatile toolkit for accurate *de novo* assembly of organelle genomes. Genome Biol 21:241. https://doi.org/10.1186/s13059-020-02154-5
Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780. https://doi.org/10.1093/molbev/mst010
Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A (2012) Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649. https://doi.org/10.1093/bioinformatics/bts199
Kim CS, Bae M, Oh J, Subedi L, Suh WS, Choi SZ, Son MW, Kim SY, Choi SU, Oh DC, Lee KR (2017) Anti-neurodegenerative biflavonoid glycosides from *Impatiens balsamina*. J Nat Prod 80:471–478. https://doi.org/10.1021/acs.jnatprod.6b00981
Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R (2001) REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res 29:4633–4642. https://doi.org/10.1093/nar/29.22.4633
Li Z, Saina J, Gichira A, Kyalo C, Wang Q, Chen J (2018) Comparative genomics of the Balsaminaceae sister genera *Hydrocera triflora* and *Impatiens pinfanensis*. Int J Mol Sci 19:319. https://doi.org/10.3390/ijms19010319
Lin CS, Chen J, Chiu CC, Hsiao H, Yang CJ, Jin XH, Leebens-Mack J, de Pamphilis CW, Huang YT, Yang LH, Chang WJ, Kui L, Wong G, Hu JM, Wang W, Shih MC (2017) Concomitant loss of NDH complex-related genes within chloroplast and nuclear genomes in some orchids. Plant J 90:994–1006. https://doi.org/10.1111/tpj.13525
Liu C, Lei J, Jiang Q, Zhou S, He X (2022a) The complete plastomes of seven Peucedanum plants: comparative and phylogenetic analyses for the *Peucedanum* genus. BMC Plant Biol 22:116. https://doi.org/10.1186/s12870-022-03488-x
Liu J, Lindstrom AJ, Gong X (2022b) Towards the plastome evolution and phylogeny of *Cycas* L. (Cycadaceae): molecular-morphology discordance and gene tree space analysis. BMC Plant Biol 22:101. https://doi.org/10.1186/s12870-022-03491-2
Luo C, Huang W, Sun H, Yer H, Li X, Li Y, Yan B, Wang Q, Wen Y, Huang M, Huang H (2021) Comparative chloroplast genome analysis of *Impatiens* species (Balsaminaceae) in the karst area of China: insights into genome evolution and phylogenomic implications. BMC Genom 22:571. https://doi.org/10.1186/s12864-021-07807-8
Lyu XL, Liu Y (2020) Nonoptimal codon usage is critical for protein structure and function of the master general amino acid control regulator CPC-1. Mbio 11:e2605–e2620. https://doi.org/10.1128/mBio.02605-20
Miller MA, Pfeiffer W, Schwartz T (2010) Creating the CIPRES science gateway for inference of large phylogenetic trees. 2010 gateway computing environments workshop (GCE). Ieee, New York, pp 1–8
Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R (2020) IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol 37:1530–1534. https://doi.org/10.1093/molbev/msaa015
Rahelivololona EM, Fischer E, Janssens SB, Sylvain GR (2018) Phylogeny, infrageneric classification and species delimitation in the Malagasy *Impatiens* (Balsaminaceae). Phytokeys. https://doi.org/10.3897/phytokeys.110.28216
Rambaut A, Drummond AJ, Dong X, Guy B, Suchard MA (2018) Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. Syst Biol 5:901–904
Ramzan M, Sarwar S, Kauser N, Saba R, Hussain I, Shah AA, Aslam MN, Alkahtani J, Alwahibi MS (2020) Assessment of Inter simple sequence repeat (ISSR) and simple sequence repeat (SSR) markers to reveal genetic diversity among *Tamarix* ecotypes. J King Saud Univ Sci 32:3437–3446. https://doi.org/10.1016/j.jksus.2020.10.003
Ren J, Tian J, Jiang H, Zhu X, Mutie FM, Wanga VO, Ding S, Yang J, Dong X, Chen L, Cai X, Hu G (2022) Comparative and phylogenetic analysis based on the chloroplast genome of *Colebrookea subtilis* (Tratt.) seidel, a protected rare species of monotypic genus. Front Plant Sci 13:828467. https://doi.org/10.3389/fpls.2022.828467
Revell LJ (2011) Phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol Evol 3:217–223
Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A (2017) DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol Biol Evol 34:3299–3302. https://doi.org/10.1093/molbev/msx248
Schneider JV, Paule J, Jungcurt T, Cardoso D, Amorim AM, Berberich T, Zizka G (2021) Resolving recalcitrant clades in the pantropical ochnaceae: insights from comparative phylogenomics of plastome and nuclear genomic data derived from targeted sequencing. Front Plant Sci. https://doi.org/10.3389/fpls.2021.638650
Sharp PM, Wen-Hsiung L (1986) Codon usage in regulatory genes in *Escherichia coli* does not reflect selection for 'rare' codons. Nucleic Acids Res 19:7737–7749
Shaw J, Shafer HL, Leonard OR, Kovach MJ, Schorr M, Morris AB (2014) Chloroplast DNA sequence utility for the lowest phylogenetic and phylogeographic inferences in angiosperms: the tortoise and the hare IV. Am J Bot 101:1987–2004
Shen XF, Wu ML, Liao BS, Liu ZX, Bai R, Xiao SM, Li XW, Zhang BL, Xu J, Chen SL (2017) Complete chloroplast genome sequence and phylogenetic analysis of the medicinal plant *Artemisia annua*. Molecules 22:1330. https://doi.org/10.3390/molecules22081330

Silva SR, Michael TP, Meer EJ, Pinheiro DG, Varani AM, Miranda VFO (2018) Comparative genomic analysis of *Genlisea* (corkscrew plants—Lentibulariaceae) chloroplast genomes reveals an increasing loss of the ndh genes. PLoS ONE 13:e190321. https://doi.org/10.1371/journal.pone.0190321

Somaratne Y, Guan DL, Wang WQ, Zhao L, Xu SQ (2020) The complete chloroplast genomes of two *Lespedeza* species: insights into codon usage bias, RNA editing sites, and phylogenetic relationships in Desmodieae (Fabaceae: Papilionoideae). Plants Basel. https://doi.org/10.3390/plants9010051

Song YX, Peng S, Cong YY, Zheng YM (2021) *Impatiens rapiformis*, a new species of Impatiens with root tuber from Yunnan. China Nord J Bot 39:3151. https://doi.org/10.1111/njb.03151

Song Y, Zhao W, Xu J, Li M, Zhang Y (2022) Chloroplast genome evolution and species identification of *Styrax* (Styracaceae). Biomed Res Int 2022:1–13. https://doi.org/10.1155/2022/5364094

Sun Y, Zou P, Jiang N, Fang Y, Liu G (2022) Comparative analysis of the complete chloroplast genomes of nine *Paphiopedilum* species. Front Genet 12:772415. https://doi.org/10.3389/fgene.2021.772415

Wang J, Lu Y, Xu Y, Jin S, Jin X (2020) *Impatiens wuyiensis* (Balsaminaceae), a new species from Fujian of Southeast China, based on morphological and molecular evidences. Bot Stud 61:29. https://doi.org/10.1186/s40529-020-00306-1

Wang Z, Wu PP, Liu CC, Guo K, Yu SX (2022a) *Impatiens nushanensis* (Balsaminaceae), a new species from Yunnan, China. Phytotaxa 545:186–196. https://doi.org/10.11646/phytotaxa.545.2.7

Wang ZW, Wang Q, Xu RH, Zhang Y, Li XC (2022b) *Impatiens chenmoui* (Balsaminaceae) a new species from southern Yunnan China. PhytoKeys 214:83–95. https://doi.org/10.3897/phytokeys.214.94898

Wick RR, Schultz MB, Zobel J, Holt KE (2015) Bandage: interactive visualization of de novo genome assemblies. Bioinformatics 31:3350–3352. https://doi.org/10.1093/bioinformatics/btv383

Xu Z, Xin T, Bartels D, Li Y, Gu W, Yao H, Liu S, Yu H, Pu X, Zhou J, Xu J, Xi C, Lei H, Song J, Chen S (2018) Genome analysis of the ancient tracheophyte *Selaginella tamariscina* reveals evolutionary features relevant to the acquisition of desiccation tolerance. Mol Plant 11:983–994. https://doi.org/10.1016/j.molp.2018.05.003

Yang J, Hu G, Hu G (2022) Comparative genomics and phylogenetic relationships of two endemic and endangered species (*Handeliodendron bodinieri* and *Eurycorymbus cavaleriei*) of two monotypic genera within Sapindales. BMC Genom 23:27. https://doi.org/10.1186/s12864-021-08259-w

Yin J, Cai X, Liu Y, Li S (2018) AHP-based ornamental value evaluation of sect. *Calcareimontana* (*Impatiens* L). Northern Horticulture 2018(22):93–97. (in Chinese with English abstract)

Yu SX, Janssens SB, Zhu XY, Liden M, Gao TG, Wang W (2016) Phylogeny of *Impatiens* (Balsaminaceae): integrating molecular and morphological evidence into a new classification. Cladistics 32:179–197. https://doi.org/10.1111/cla.12119

Yuan YM, Song Y, Geuten K, Rahelivololona E, Wohlhauser S, Fischer E, Smets E, Kupfer P (2004) Phylogeny and biogeography of Balsaminaceae inferred from ITS sequences. Taxon 53:391–403. https://doi.org/10.2307/4135617

Zalapa JE, Cuevas H, Zhu HY, Steffan S, Senalik D, Zeldin E, McCown B, Harbut R, Simon P (2012) Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. Am J Bot 99:193–208. https://doi.org/10.3732/ajb.1100394

Zhang Y, Han L, Yang C, Yin Z, Tian X, Qian Z, Li G (2022a) Comparative chloroplast genome analysis of medicinally important *Veratrum* (Melanthiaceae) in China: insights into genomic characterization and phylogenetic relationships. Plant Divers 44:70–82. https://doi.org/10.1016/j.pld.2021.05.004

Zhang Z, Tao M, Shan X, Pan Y, Sun C, Song L, Pei X, Jing Z, Dai Z (2022b) Characterization of the complete chloroplast genome of *Brassica oleracea* var. *italica* and phylogenetic relationships in Brassicaceae. PLoS ONE 17:e263310. https://doi.org/10.1371/journal.pone.0263310

Zhou M, Guo J, Cha J, Chae M, Chen S, Barral JM, Sachs MS, Liu Y (2013) Non-optimal codon usage affects expression, structure and function of clock protein FRQ. Nature 495:111–115

Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH (“Springer Nature”).
Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users (“Users”), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use (“Terms”). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.
These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.
We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.
While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval , sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.
These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.
To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.
Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.
If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com