MEDDOCS Open Access Publisher

ISSN: 2637-7721

Journal of Plant Biology and Crop Research

Open Access | Research Article

Phylogenetic Analysis of *F3H* Genomic DNA Sequence of *Impatiens* Species in Yunnan-Guizhou Plateau

Li Linju[1]; Feng Zhixi[1]; Zheng Junjun[2]; Qu Suping[3]; Li Xinyi[1]; Shi wanlei[1]; Huang Haiquan[1]*; Huang Meijuan[1]*

[1]College 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.

[2]Maize Research Institute of Sichuan Agricultural University, Chengdu, Sichuan, 611134,China

[3]Flower Research Institute, Yunnan Academy of Agricultural Sciences, Kunming, Yunnan, 650205, China.

***Corresponding Author(s):**

Huang Haiquan & Huang Meijuan

College 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.
Email: haiquanl@163.com & xmhhq2001@163.com

Received: Jan 09, 2023

Accepted: Feb 02, 2023

Published Online: Feb 06, 2023

Journal: Journal of Plant Biology and Crop Research

Publisher: MedDocs Publishers LLC

Online edition: http://meddocsonline.org/

Keywords: *Impatiens*; F3H genomic DNA; coding region and intron sequence; Phylogenetic analysis; Yunnan-Guizhou plateau.

Abstract

In order to examine the genetic and evolutionary connections of *Impatiens*, the *F3H* genomic DNA sequences of 28 species of *Impatiens* were extracted by degenerate primers. The key findings are as follows: The sequence of the *F3H* gene CDS coding region is very conservative, although the sequences of the first and second introns are highly variable, and the *F3H* gene CDS coding region is mostly composed of base conversions. The base substitution and transversion of the CDS coding region sequence, the first intron sequence, the second intron sequence, and the combined analysis of two intron sequences did not reach saturation, enabling the sequence to be used for future phylogenetic analysis. Evolutionary research demonstrated that the phylogenetic tree based on CDS coding region sequence and two intron sequence joint analysis was roughly comparable with the usual classification of *Impatiens*. The *Impatiens* materials used in this investigation are limited, and the genetic link architecture of *Impatiens* must be expanded and improved. The *F3H* gene might be utilized as a novel molecular marker in conjunction with the trait index of conventional categorization. It may be used to evaluate the genetic connection of *Impatiens* and offer a better scientific basis for later *Impatiens* categorization.

Introduction

The Balsaminaceae family includes two genera: *Impatiens* and *H. trifloral*. *H. triflora* includes only one species, but *Impatiens* comprises around 1000 species [1-3]. There are currently 352 species of *Impatiens* recorded in China, of which approximately three-quarters are endemic to China [4]. The majority of these species are found in southwest and northwest China, particularly the southwest provinces. The species are the most abundant [5,6,7] and Yunnan has the highest population density in the country [8]. Due of its many plant species and intricate plant morphology, *Impatiens* has been studied using morphology [9,10], plant anatomy [11], and plant palynology [12,13,14] in order to determine its origin and evolution. With

Cite this article: Huang HQ, Huang MJ, Linju Li, Zhixi F, Junjun Z, at al. Phylogenetic analysis of *F3H* genomic DNA sequence of *Impatiens* species in Yunnan-Guizhou plateau. J Plant Biol Crop Res. 2023; 6(1): 1072.

the advancement of molecular systematics in *Impatiens*, we can observe that *nrITS* sequences and *cpDNA* sequences have been extensively examined [15,16,17,18], but few people pay attention to the development of its introns [19].

Introns were formerly regarded to be useless. Introns in the genome are currently regarded to put a considerable load on many cells and may have a number of negative impacts on gene expression [20,21]. However, other experts believe that introns provide more benefits throughout the evolution process, which may explain why introns have been conserved for so long [22]. There is still opportunity for investigation in the molecular systematics of *Impatiens*, both at home and abroad. The amino acid sequence, the first intron sequence, and the second intron sequence have high potential for demonstrating *Impatiens*' evolutionary connection.

This research examined the evolutionary connection between 28 species of *Impatiens* based on the *F3H* gene coding region and intron sequences in order to offer molecular phylogenetic data for phylogenetic analysis and the gathering and use of superior germplasm resources. It also has significant theoretical implications for the study of the evolution of other *Impatiens* plants.

Materials and Methods

Plant materials and DNA extraction

A total of 28 species (or varieties) of *Impatiens* were collected, and the sampling information statistics are shown in **Table 1**. Genomic DNA of 28 species of *Impatiens* was extracted according to the instructions of the BioTeke DNA extraction kit.

Table 1: The distribution of the part of Impatiens.

Number	Species	Locality	Plant height /cm	Distribution altitude /m
1	*I. uliginosa*	Laoyuhe Park	40-80	1453
2	*I. cyathiflora*	Dabao Mountain	30-60	2720
3	*I. racemosa* var.	Anlong County	80-90	1146
4	*I. delavayi*	Dali	30-50	3158
5	*I. corchorifolia*	Zhuanlong Town	30-50	2292
6	*I. cyanantha*	Jizu Mountain	30-54	2475
7	*I. arguta*	Ailao Mountain	30-50	1433
8	*I. chlorosepala*	Wangmo County	30-50	821
9	*I. napoensis*	Anlong County	40-50	1347
10	*I. siculifer*	Fanjing Mountain	40-50	530
11	*I. clavigera*	Wenshan City	20-30	1387
12	*I. ruiliensis*	Fugong County	40-60	1228
13	*I. stenosepala*	Fanjing Mountain	25-35	730
14	*I. corchorifolia* var.	Wenshan City	30-50	2100
15	*I. dicentra*	Fanjing Mountain	40-120	530
16	*Impatiens*	Greenhouse planting	20-30	1407
17	*I. pinetorum*	Gaoligong Mountain	30-60	2207
18	*I. rectangula*	Gaoligong Mountain	30-70	2700
19	Unknow1	Xundian	35-70	1347
20	*I. faberi*	Wenshan City	60-70	1350
21	*I. loulanensis*	Anlong County	80-120	1741
22	*I. aquatilis*	Suiyang County	20-30	1260
23	*I. tubulosa*	Fugong County	40-60	1228
24	*I. polyceras*	Xundian	20-75	2279
25	Unknow2	Xundian	30-50	2826
26	*I. radiata*	Xundian	40-50	2826
27	*I. guizhouensis*	Fanjing Mountain	20-30	870
28	*I. siculifer* var.	Fanjing Mountain	40-50	530

F3H genome cloning of some plants in *Impatiens*

Referring to the *F3H* gene sequence in *I. uliginosa* transcriptome and comparing it with the *F3H* gene sequence of other plants reported by NCBI, degenerate primers were designed for highly conserved regions including introns. The following primers: JB.*F3H*.F:GTKGCYTACAATRWATTCAGC and JB.*F3H*.R: CATYTTCCTC-YTRTACATYTC (K=G/T, Y=C/T, R=A/G, W=A/T).

The genomic DNA of 28 species (or varieties) of *Impatiens* was extracted as a template, and JB.*F3H*.F, JB.*F3H*.R is the primer for PCR amplification of *F3H* genomic DNA of different species (or varieties) of *Impatiens*. A typical amplification assay of 40 uL contained 3.2 uL of High Pure dNTPs (2.5 mM), 4.8 uL of 10×EasyTaq Buffer (+Mg^{2+}), 2.0 uL of each primer, 0.4 uL of EasyTaq DNA Polymerase, 2.0 uL of Template DNA and 25.6 uL of ddH_2O. The PCR reaction procedures are 95 ° C for 5min, 95 ° C for 50s, 52 ° C for 30s, 72 ° C for 1min for 10s, and 72 ° C for 10min. Take 8ul PCR product and 4ul 6* loading buffer to beat and mix, and use 1.2% gel electrophoresis to detect whether the PCR product is correct. The PCR amplification products whose target bands are consistent with the expected results

will be sequenced by Sheng gong Biotechnology Co., Ltd.

Sequence Analysis of 28 species of *Impatiens*

The *F3H* genomic DNA sequences of 28 species of *Impatiens* were analyzed by using the software DNAMAN multiple sequence alignment. The codon preference of its coding region, the content of GC and GC3s in the sequence were analyzed using the online software EMNOSS explore. In software MEGA 7.0, the base substitution of CDS coding region sequence is calculated by the Maximum Likelihood Estimate of Substitution Matrix function, the overall average genetic distance is calculated by the Computer Overall Mean Distance function, and the p distance, s distance and v distance of the coding region, the first intron sequence, the second intron sequence and the joint analysis of the two intron sequences of *F3H* gene are calculated by Computer Pairwise Distance. Drawing scatter plot with s distance and v distance as the vertical axis and p-distance as the horizontal axis, combined with Excel software. The phylogenetic tree of the coding region, the first intron sequence, the second intron sequence and the joint analysis of the two intron sequences of the *F3H* gene of 28 species of *Impatiens* is constructed by ML method in MRGA7.0 software.

Results

Cloning results of *F3H* genomic DNA of some *Impatiens*

The extracted DNA of 28 *Impatiens* species (or variations) was electrophoretically tested, and the bands were single and brilliant without primer dimer, suggesting specific amplification **(Figure S1, Figure S2, Figure S3)**.

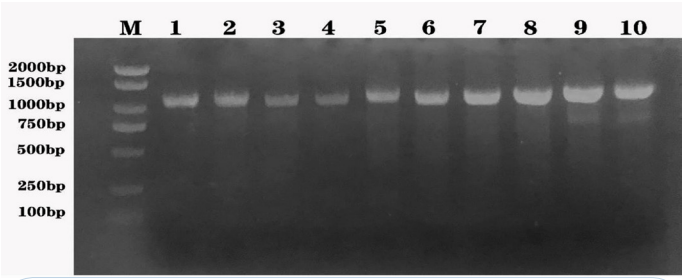

Figure S1: Electrophoretic detection of gDNA PCR amplification of *F3H* in *Impatiens.*

Note: M: DL 2000+ DNA Marker; 1: *I. uliginosa*; 2: *I. cyathiflora*; 3: *I. racemosa* var.*;* 4: *I. delavayi*; 5: *I. corchorifolia*; 6: *I. cyanantha*; 7: *I. arguta*; 8: *I. chlorosepala*; 9: I. napoensis; 10: *I. siculifer*;

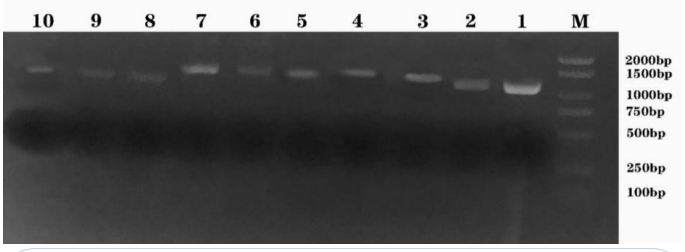

Figure S2: Electrophoretic detection of gDNA PCR amplification of *F3H* in *Impatiens.*

Note: M: DL 2000+ DNA Marker; 1: *I. clavigera*; 2: *I. ruiliensis*; 3: *I. stenosepala*; 4: *I. corchorifolia* var.; 5: *I. dicentra*; 6: *Impatien*; 7: *I. pinetorum*; 8: *I. rectangula*; 9: Unknown; 10: *I. faberi*;

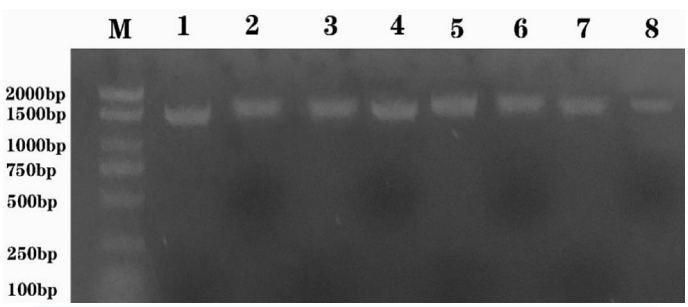

Figure S3: Electrophoretic detection of gDNA PCR amplification of *F3H* in *Impatiens.*

Note: M: DL 2000+ DNA Marker; 1: *I. loulanensis*; 2: *I. ruiliensis*; 3: *I. aquatilis;* 4: *I. polyceras*; 5: Unknown2; 6: *I. radiata*; 7: *I. guizhouensis*; 8: *I. siculifer* var.

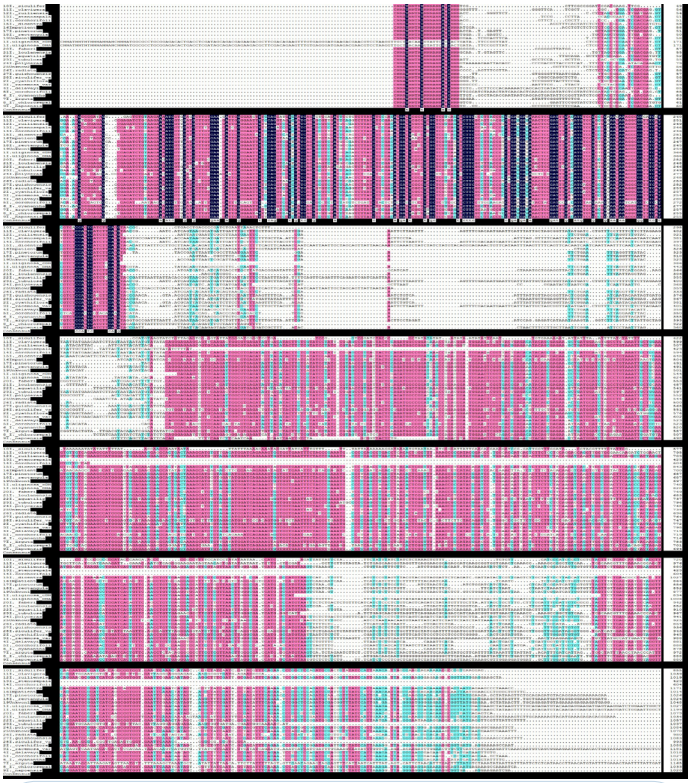

Figure S4: Homologous gDNA sequence alignment of F3H gene in 28 species Impatiens.

Analysis of *F3H* Genomic DNA Sequences of 28 *Impatiens*

The location and length of partial coding sections and introns of *F3H* gene fragments were identified by comparing the *F3H* genome sequences of *I. uliginosa* with other NCBI plants cloned in our laboratory. The *F3H* genomic DNA segments of 28 different *Impatiens* species were cloned. Except for *I. siculife*, *I. stenosepala*, *I. corchorifolia* var., and *I. tubulosa*, the *F3H* genomes of the remaining 24 *Impatiens* had three exons and two introns. In contrast, *I. tubulosa* had only one intron sequence. Only partial sequences of the *F3H* genome coding region were recovered from *I. siculife*, *I. stenosepala*, and *I. corchorifolia* var. **Table 2** shows the lengths of the isolated sequence fragments, the first intron sequence, and the second intron sequence.

Table 2: length of 28 species Impatiens *F3H* gene sequence.

Species	The part of gene/bp	Intron 1/bp	Intron 2/bp
I. uliginosa	1104	80	92
I. cyathiflora	849	138	79
I. racemosa var.	1067	170	100
I. delavayi	863	65	68
I. corchorifolia	827	73	75
I. cyanantha	863	65	68
I. arguta	1144	153	108
I. chlorosepala	843	75	93
I. napoensis	836	70	92
I. siculifer	833		
I. clavigera	745	170	102
I. ruiliensis	836	79	82
I. stenosepala	401		
I. corchorifolia var.	582		
I. dicentra	976	81	92
Impatiens	847	74	71
I. pinetorum	863	66	75
I. rectangula	863	68	71
Unkwon1	869	70	87
I. faberi	857	119	99
I. loulanensis	876	61	75
I. aquatilis	839	117	79
I. tubulosa	845	170	0
I. polyceras	866	97	99
Unknown2	840	118	90
I. radiata	824	69	65
I. guizhouensis	845	138	78
I. siculifer var.	775	129	67

Alignment and Analysis of *F3H* Gene of 28 *Impatiens*

The CDS coding region sequencing of *F3H* genomic gDNA from 28 *Impatiens* species was determined to be generally conservative. The *F3H* genomic sequence of most *Impatiens* plants is similar to that of other plants, with three exons and two introns. It was also discovered that the intronic sequence of *Impatiens*' *F3H* gene group varies widely, with the length of the first intron ranging from tens to hundreds of bp. Even *F3H* genomic sequences like *I. tubulosa* only have the first intron.

Codon bias analysis of CDS coding regions of *F3H* gene in 28 *Impatiens*

I. siculifer has the lowest CAI value for the *F3H* gene (0.653), whereas *I. napoensis* has the highest (0.723). The lowest ENc value was 49.314 for *I. chlorosepala*, while the maximum ENc value was 61.000 for four species of *I. siculifer* var., *I. clavigera*, *I. siculifer*, and *I. cyanantha*. The GC content ranges between 40 and 60%, with the GC3s level around 50% **(Table 3)**. The statistics of codon bias in the *F3H* gene coding region of 28 *Impatiens* species demonstrate that the GC content and GC3s content account for around half of the total number of codons. There is no obvious preference for *F3H* gene codons of 28 *Impatiens* species.

Table 3: Code bias analysis of dense CDS coding region of *F3H* gene of Impatiens.

Species	CAI	ENc	GC3s(%)	GC(%)
I. uliginosa	0.657	54.782	51.63	48.37
I. cyathiflora	0.660	57.044	57.60	51.00
I. racemosa var.	0.682	56.558	59.98	48.86
I. delavayi	0.684	59.750	41.81	48.32
I. corchorifolia	0.673	58.678	46.90	47.37
I. cyanantha	0.674	61.000	40.42	47.85
I. arguta	0.672	57.010	41.73	42.78
I. chlorosepala	0.692	49.341	38.79	44.72
I. napoensis	0.723	53.272	43.53	49.64
I. siculifer	0.653	61.000	41.88	41.52
I. clavigera	0.699	61.000	56.45	52.96
I. ruiliensis	0.703	60.334	51.44	49.88
I. stenosepala	0.688	60.334	44.36	42.86
I. corchorifolia var.	0.696	59.740	46.91	45.70
I. dicentra	0.676	53.740	50.77	42.26
Impatien	0.673	55.698	40.78	45.39
I. pinetorum	0.680	57.222	39.37	48.78
I. rectangula	0.671	55.065	54.36	48.90
Unknow1	0.686	57.171	40.48	49.37
I. faberi	0.710	51.300	47.31	56.27
I. loulanensis	0.692	56.376	54.45	46.92
I. aquatilis	0.675	56.119	52.51	49.03
I. tubulosa	0.706	59.993	56.23	51.36
I. polyceras	0.684	57.732	44.44	55.21
Unknown2	0.663	56.468	52.50	48.10
I. radiata	0.676	54.856	40.51	47.45
I. guizhouensis	0.701	59.487	57.61	51.63
I. siculifer var.	0.672	61.000	49.61	56.07

Analysis of base substitution and genetic distance of *F3H* gene in 28 species of *Impatiens*

It was found that most of the site variations were base substitutions and relatively few base transversions **(Table 4)**. The average genetic distance is 0.262. At the same time, the Computer Pairwise Distance was used to analyze the Pairwise Distance of the CDS coding region of the *F3H* gene in 28 *Impatiens* species. The most recent genetic distance between *I. delavayi* and *I. cyanantha* was 0.003, and the most distant genetic distance between *I. dicentra* and *I. corchorifolia* was 0.923 **(Figure 1)**.

Table 4: Nucleotide substitutions of 28 species *Impatiens F3H* gene CDS sequence.

	A	T/U	C	G
A		7.04	5.97	10.98
T/U	6.05		12.48	7.26
C	6.05	14.72		7.26
G	9.16	7.04	5.97	

Note: font-weight for Transition; italics for Transversion

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
1. 1 I. uliginosa																												
2. 2 I. cyathiflora	0.151																											
3. 3 I. racemosa var.	0.140	0.152																										
4. 4 I. delavayi	0.048	0.166	0.147																									
5. 5 I. corchorifolia	0.428	0.566	0.556	0.432																								
6. 6 I. cyanantha	0.051	0.170	0.151	0.003	0.438																							
7. 7 I. arguta	0.118	0.125	0.103	0.121	0.522	0.124																						
8. 8 I. chlorosepala	0.237	0.220	0.228	0.203	0.585	0.203	0.195																					
9. 9 I. napoensis	0.219	0.140	0.207	0.234	0.588	0.239	0.187	0.296																				
10. 10 I. siculifer	0.446	0.429	0.488	0.453	0.847	0.459	0.470	0.463	0.478																			
11. 11 I. clavigera	0.148	0.156	0.019	0.155	0.571	0.159	0.111	0.224	0.212	0.501																		
12. 12 I. ruiliensis	0.065	0.174	0.155	0.075	0.490	0.071	0.129	0.240	0.236	0.481	0.163																	
13. 13 I. stenosepala	0.403	0.353	0.334	0.410	0.806	0.416	0.288	0.445	0.377	0.514	0.340	0.430																
14. 14 I. corchorifolia var.	0.422	0.349	0.273	0.428	0.834	0.434	0.352	0.454	0.393	0.482	0.283	0.435	0.258															
15. 15 I. dicentra	0.490	0.439	0.439	0.448	0.923	0.454	0.444	0.331	0.488	0.506	0.444	0.495	0.360	0.325														
16. 16 Impatien	0.239	0.230	0.251	0.246	0.572	0.250	0.234	0.268	0.293	0.412	0.256	0.261	0.429	0.415	0.440													
17. 17 I. pinetorum	0.055	0.163	0.155	0.061	0.479	0.058	0.132	0.240	0.223	0.458	0.160	0.028	0.422	0.438	0.493	0.261												
18. 18 I. rectangula	0.068	0.178	0.162	0.061	0.468	0.058	0.132	0.232	0.224	0.452	0.167	0.051	0.412	0.443	0.487	0.269	0.028											
19. 19 Unknow1	0.068	0.170	0.163	0.075	0.503	0.071	0.140	0.245	0.244	0.481	0.168	0.041	0.439	0.455	0.506	0.270	0.022	0.038										
20. 20 I. faberi	0.113	0.106	0.120	0.120	0.501	0.124	0.095	0.214	0.162	0.468	0.128	0.132	0.368	0.387	0.477	0.212	0.131	0.138	0.139									
21. 21 I. loulanensis	0.075	0.198	0.178	0.078	0.482	0.075	0.158	0.250	0.262	0.500	0.179	0.065	0.450	0.471	0.506	0.288	0.045	0.048	0.054	0.154								
22. 22 I. aquatilis	0.162	0.178	0.227	0.173	0.499	0.178	0.193	0.219	0.257	0.427	0.223	0.193	0.425	0.450	0.455	0.229	0.182	0.182	0.198	0.180	0.218							
23. 23 I. tubulosa	0.148	0.167	0.019	0.151	0.544	0.155	0.118	0.250	0.224	0.507	0.035	0.167	0.359	0.297	0.468	0.261	0.163	0.163	0.164	0.132	0.179	0.240						
24. 24 I.polyceras	0.186	0.170	0.194	0.197	0.598	0.201	0.169	0.298	0.239	0.583	0.203	0.198	0.467	0.484	0.605	0.297	0.197	0.205	0.206	0.072	0.222	0.252	0.206					
25. 25 Unknow2	0.132	0.128	0.147	0.146	0.485	0.150	0.139	0.179	0.214	0.362	0.132	0.162	0.351	0.377	0.415	0.190	0.158	0.150	0.162	0.134	0.170	0.124	0.159	0.205				
26. 26 I.radiata	0.068	0.163	0.143	0.054	0.434	0.058	0.125	0.227	0.228	0.428	0.152	0.114	0.410	0.427	0.493	0.208	0.092	0.085	0.106	0.124	0.096	0.182	0.151	0.202	0.143			
27. 27 I. guizhouensis	0.117	0.029	0.117	0.131	0.530	0.135	0.092	0.211	0.129	0.441	0.121	0.139	0.351	0.352	0.436	0.221	0.128	0.142	0.135	0.074	0.162	0.162	0.132	0.142	0.113	0.128		
28. 28 I. siculifer var.	0.219	0.194	0.215	0.223	0.619	0.227	0.195	0.288	0.224	0.484	0.224	0.237	0.421	0.397	0.451	0.265	0.236	0.236	0.245	0.104	0.249	0.280	0.228	0.188	0.220	0.228	0.166	

Figure 1: Pairwise distance of 28 species *Impatiens F3H* gene CDS sequence.

Substitutional Saturation Analysis of *F3H* Gene of 28 *Impatiens* Species

It was found that the transformation genetic distance and transversion genetic distance of the coding region sequence, the first intron sequence, the second intron sequence and joint analysis of the two intron sequences of the *F3H* gene of 28 different *Impatiens* plants tended to increase linearly with the increase of the degree of difference among sequences **(Figure 2-5)**. To some extent, the base substitution is not saturated linear. The results further showed that the base substitution of the coding region sequence, the first intron sequence, the second intron sequence and joint analysis of the two intron sequences of the *F3H* gene had not reached the saturation point, which could be used in the subsequent phylogenetic analysis.

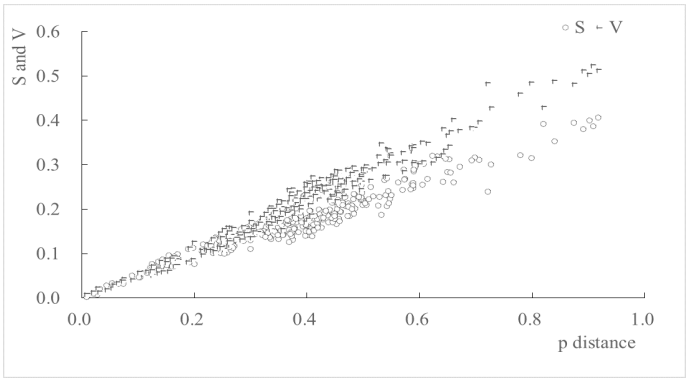

Figure 2: Saturation detection of variation of the CDS sequence of *F3H* gene in 28 species of *Impatiens*.

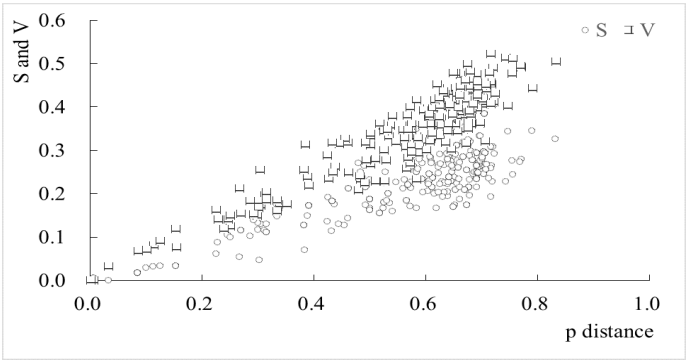

Figure 3: Saturation detection of variation of the first intron sequence of *F3H* gene in 28 species of *Impatiens.*

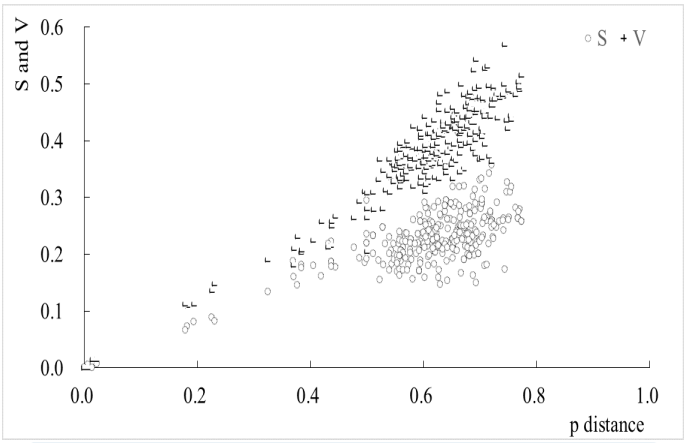

Figure 4: Saturation detection of variation of the second intron sequence of *F3H* gene in 28 species of *Impatiens.*

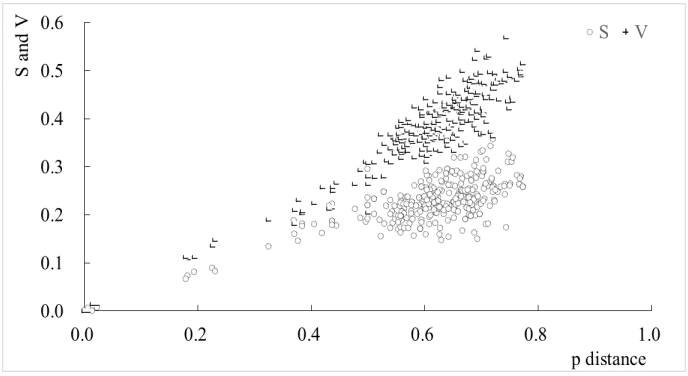

Figure 5: Saturation detection of variation of the two intron sequence joint analysis of F3H gene in 28 species of Impatiens.

Phylogenetic tree analysis based on *F3H* gene sequence

The phylogenetic tree of the *Impatiens F3H* gene coding region was constructed using MEGA 7.0 and the best nucleic acid substitution model K2+G **(Figure 6)**. *Impatiens* species are divided into four branches. Clade I includes *I. siculifer* var., *I. faberi*, and *I. polyceras*; the species on this branch are annual herbs with round flag petals and four germination holes in the pollen. Most species in Clade II are yellow perennial herbs with four germination pores and reticular pollen patterns with protuberances in their mesh, such as *I. arguta*, *I. racemosa* var., *I. clavigera*, and *I. tubulosa*. Clade III contains the following species: *I.*

corchorifolia, *I. delavayi*, *I. cyanantha*, *I. radiata*, *I. rectangula*, *I. loulanensis*, *I. uliginosa*, *I. ruiliensis*, *I. pinetorum*, and Unkwon1. The majority of the species on this branch are annual yellow herbs with two lateral sepals and four germination holes. Pollen's coarse reticular pattern is characterized by sparse granular protuberances in its mesh. The majority of Clade IV species are annual red herbs with two lateral sepals, suborbicular flag petals, sessile 2-lobed pterygotes, and pollen with four germinating pores, such as *I. corchorifolia* var., *I. dicentra*, *I. stenosepala*, *I. siculifer*, *I. napoensis*, *I. cyathiflora*, *I. guizhouensis*, Unknow2, *I. chlorosepala*, *Impatiens*, *I. aquatilis*.

MEGA 7.0 software chose the best nucleic acid substitution model T92, and a phylogenetic tree based on ML was generated for the first intron sequence, the second intron sequence, and the joint analysis of the two intron sequences of the *Impatiens F3H* gene **(Figure 7-9)**. Numerous data analysis results are found to be validated, with the development tree based on the joint analysis of two intron sequences being the most compatible with the traditional categorization of *Impatiens*. 25 species of *Impatiens* may be separated into four distinct branches. Clade I is comprised of six different species: *Impatiens*, *I. dicentra*, *I. guizhouensis*, *I. cyathiflora*, Unkwon 2, and *I. arguta*. The majority of species on this branch are annual herbs with round or oval petals. Clade II consists of annual herbs with four lateral sepals, flag petals that are predominantly orbicular or suborbicular, sessile, two-lobed wing petals, pollen with four germinating pores and granular processes in the mesh, and includes *I. napoensis*, *I. faberi*, *I. aquatilis*, *I. polyceras*, and *I. siculifer* var. Clade III consists of *I. cyanantha*, *I. delavayi*, *I. ruiliensis*, *I. pinetorum*, *I. rectangula*, Unkwon1, and *I. loulanensis*, all of which are annual herbs with four lateral sepals, pollen with four germination pores, and pollen with sparse granular processes in its reticulum. The majority of Clade IV species are annual plants with round or oval flag petal shapes and granular protuberances on the reticulum, including I. uliginosa, *I. racemosa* var., *I. corchorifolia*, *I. chlorosepala*, *I. clavigera*, *I. tubulosa*, and *I. radiata*.

In the phylogenetic tree constructed based on the *F3H* gene coding region sequence and the joint analysis of two intron sequences, we can clearly observe that *I. cyathiflora* and *I. guizhouensis*, *I. siculifer* var. and *I. polyceras*m, *I. racemose* var., *I. clavigera* and *I. tubulosa* always clustered in one branch with high support rate. The *Impatiens* floral organs on a single branch all belong to the same color scheme and have comparable morphological traits, such as big flower size, form of flag and lip, kind of pollen germination groove, surface ornamentation, etc. According to the results of the phylogenetic tree based on the *F3H* genome, flower size, flag petal, wing petal, flag petal, and other flower morphological traits are the most essential foundation for the interspecific categorization of the genus *Impatiens*. In contrast, the phylogenetic tree based on the first and second intron sequences revealed that the majority of species were intermingled to varied degrees in the evolutionary tree, which was consistent with the results of the sequence saturation analysis. Therefore, the coding region sequence and the joint analysis sequence of two introns of *F3H* gene are more suitable for developmental tree analysis.

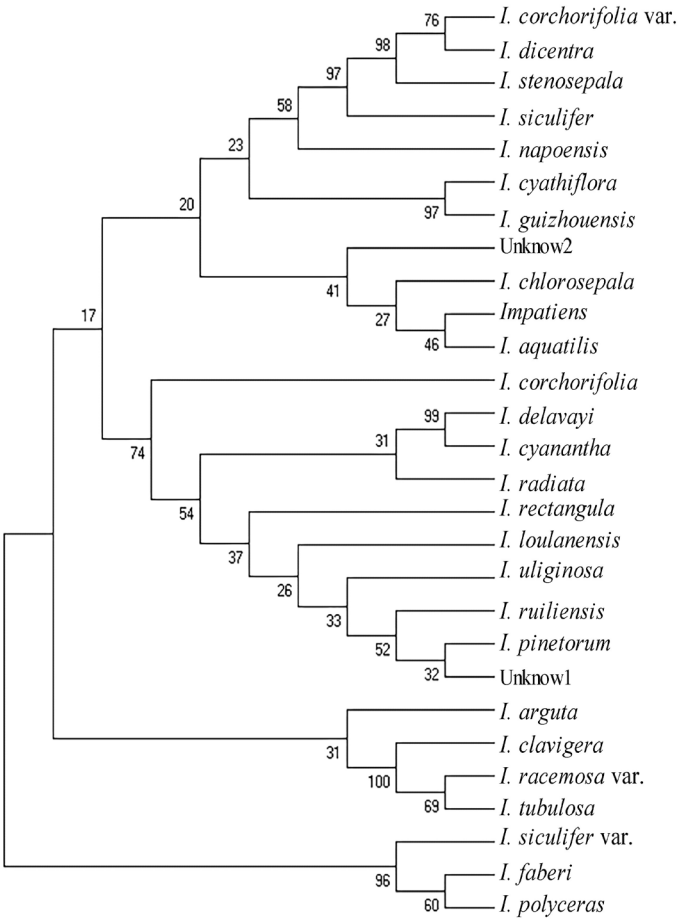

Figure 6: Phylogenetic tree based on 28 species varieties Impatiens of CDS sequence of F3H gene (ML).

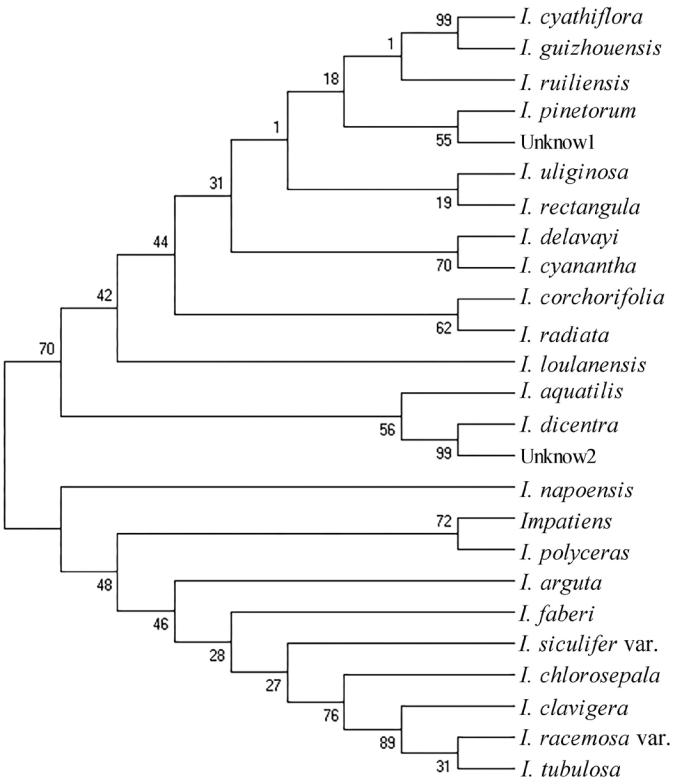

Figure 7: Phylogenetic tree based on 25 species varieties *Impatiens* of Intron1 sequence of *F3H* gene (ML).

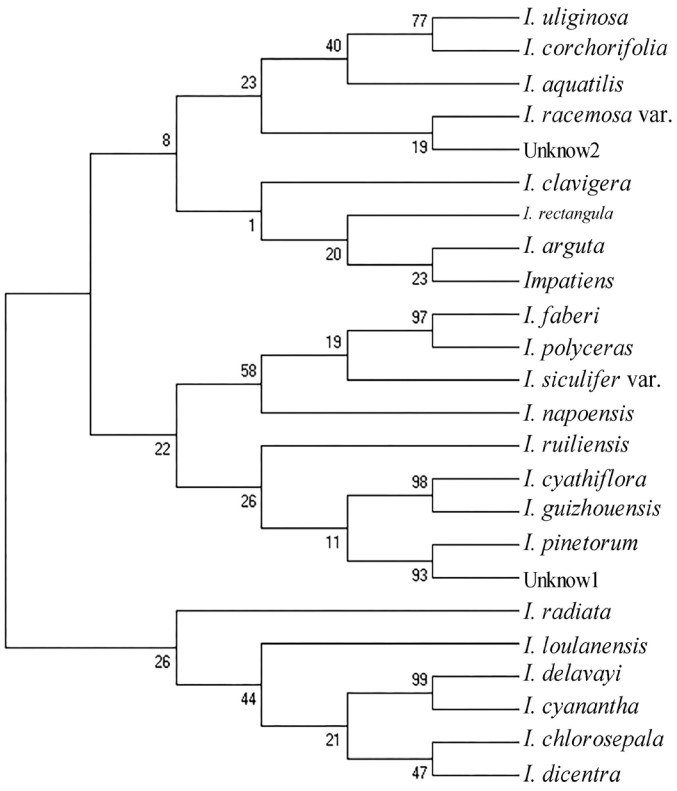

Figure 8: Phylogenetic tree based on 24 species varieties *Impatiens*. of Intron2 sequence of *F3H* gene (ML).

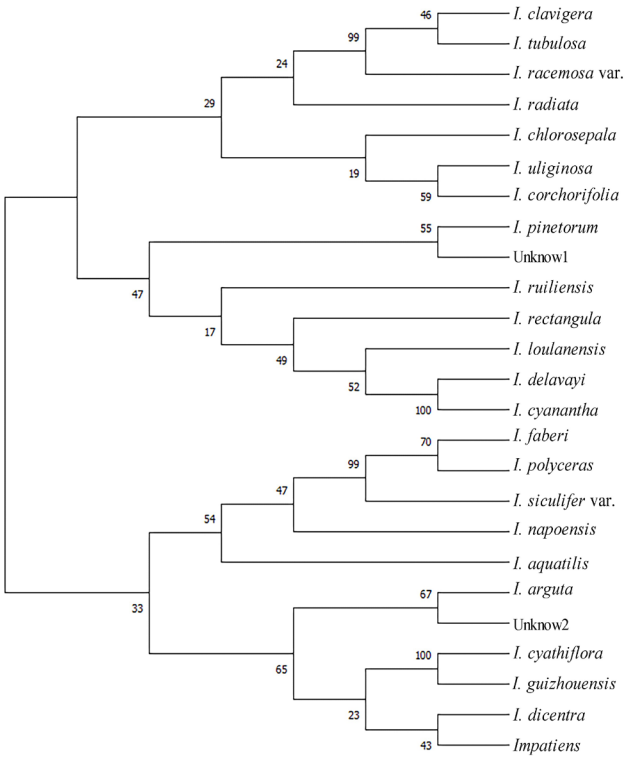

Figure 9: Combined analysis with phylogenetic tree based on 25 species varieties *Impatiens.* of Intron1 and Intron2 sequence of *F3H* gene (ML).

Discussion

DNA sequences comprise coding area sequences and non-coding region sequences. The majority of research on systematic evolution exclude intronic sequences, compare only exonic regions, and deduce their roles from mRNA [23,24,25,26,27]. There is little question, however, that introns play a significant role in genome evolution [28]. In this work, the CDS coding region, the first intron, and the second intron of the *F3H* genome of 28 species of *Impatiens* were thoroughly investigated using the MEGA 7.0 software. The phylogenetic tree was generated using the Maximum Likelihood (ML) approach employing the CDS coding region, the first intron, the second intron, and the combined analysis of the two introns sequences from the *F3H* genome.

According to the complete research, the phylogenetic tree based on the *F3H* gene in *Impatiens* plants varied significantly. The majority of phylogenetic trees based on the coding area sequences and the combined analysis of two intron sequences were mostly compatible with the conventional categorization of *Impatiens* plants. In eastern Yunnan, Zhao (2017) created a phylogenetic tree of *Impatiens* L. Based on the DEF gene fragment, *I. dicentra*, *I. uliginosa*, and *I. napoensis* were shown to be grouped into one branch, however the *F3H* gene fragment-based phylogenetic tree revealed that they were distantly related. The phylogenetic tree reveals that *I. corchorifolia* and *I. corchorifolia* var. are on separate branches. It is possible that a segment of the *F3H* genome of *I. corchorifolia* var. is missing, preventing it from displaying all of its genetic information. [29] investigated the phylogenetic development of *Impatiens* resources in Guizhou Province and discovered that *I. loulanensis* and *I. cyanantha* are always found together. *I. loulanensis* and *I. cyanantha* are grouped together in this research because they share comparable morphological traits, such as big and colorful blooms, unusual floral patterns, and round petals. *I. cyanantha* and *I. uliginosa* always congregate, which is compatible with the findings of [30,31]. Based on a thorough examination of the four evolutionary trees, it is discovered that the internal support rate of the branches on the system's branching diagram is high and low, and that there is a phenomenon of mixed classification of some species, which may be due to factors such as base substitution or subversion, or because the cloned genome fragments are not complete full-length sequences, the subsequent biological information may be in complete. Non-developmental information may also influence the creation of a phylogenetic tree [32]. However, it is apparent that the *F3H* genome gives strong support for *Impatiens* phylogenetic analyses. The results of the *F3H* genome's phylogenetic tree may be paired with *atpB-rbcL*, *trnL-F*, *ITS*, and other genes to refer to morphological traits, which can be used as a reference for *Impatiens* categorization in the future.

Declarations

Acknowledgments

Thanks to Southwest Forestry University for providing experimental instruments and equipment for this experiment, and also thanks to Teacher Huang Haiquan and Huang Meijuan for his patient guidance during the experiment.

Funding

This study was financially supported by the National Natural Science Foundation of China (32060364, 32060366, 31860230), Major Science and Technology Projects in Yunnan Province (202102AE090052), and Doctoral Tutor Team for Genetic Improvement and High-efficient Propagation of Landscape Plants in Yunnan Province.

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Author Contributions

The study's inception and design were contributed to by all authors. Li Linju, Feng Zhixi, Zheng Junjun, Qu Suping, Li Xinyi, Shi wanlei, Huang Haiquan, and Huang Meijuan were in charge of material preparation, data collecting, and analysis. Li Linju wrote the initial draft of the text, and other contributors provided feedback on prior drafts. The final text was reviewed and approved by all writers.

Data Availability Statements

1. The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.
2. All data generated or analysed during this study are included in this published article (and its supplementary information files).
3. The datasets generated during and/or analysed during the current study are not publicly available due to [REASON(S) WHY DATA ARE NOT PUBLIC] but are available from the corresponding author on reasonable request.
4. Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

References

1. Janssens SB, Viaene T, Huysmans S, Smets EF, Geuten KP. Selection on length mutations after frameshift can explain the origin and retention of AP3/DEF like paralogues in Impatiens. Journal of Molecular Evolution. 2008; 66: 424-435.
2. Janssens SB, Knox EB, Dessein S, Smets EF () Impatiens msisimwanensis (Balsaminaceae): Description, pollen morphology and phylogenetic position of a new East African species. South African Journal of Botany. 2009a; 75: 104-109.
3. Janssens SB, Knox EB, Huysmans S, Smets EF, Merckx VSFT. Rapid radiation of Impatiens (Balsaminaceae) during Pliocene and Pleistocene: Result of a global climate change. Molecular Phylogenetics and Evolution. 2009b; 52: 806-824.
4. Yu JH, Zhang WD, Qin F, Xia CY, Qin Y, et al. Impatiens yunlingensis (Balsaminaceae), a new species from Yunnan, China. PhytoKeys. 2022; 212: 13-27.
5. Chen Yilin. Studies on the genus Impatiens L. in china. 1978; 16: 36-55.
6. Cong YY, Xiang YL, Liu KM. Impatiens quadriloba (Balsaminaceae), a new species from Sichuan, China. Nordic Journal of Botany. 2010; 28: 309-312.
7. Ren LY. Study on wild Impatiens resources and community species diversity in Southern Guizhou. Guizhou University. 2022.
8. Yuan TH, Li MJ, Ren LY, Huang RX, Chen Y, et al. A dataset on the diversity and geographical distributions of wild Impatiens in China. Biodiversity Science. 2022; 30: 118-122.
9. Xia CY. Phylogeny study and taxonomic revision of Impatiens subg. Clavicarpa. Southwest University. 2020.
10. Song YX, Peng S, Mutie FM, Jiang H, Ren J, et al. Evolution and taxonomic significance of seed micromorphology in Impatiens (Balsaminaceae). Frontiers in Plant Science. 2022; 13: 835943.
11. Yang YY, Zhang Q, Chen T, Wu H, Huang HQ, et al. Investigation and analysis on germplasm resources of Impatiens L. in the southwest of Guizhou. Advances in ornamental horticulture of China. 2017; 35-41.
12. Janssens SB, Wilson YS, Yuan YM, Nagels A, Smets EF, et al. A total evidence approach using palynological characters to infer the complex evolutionary history of the Asian Impatiens (Balsaminaceae). Taxon. 2012; 61: 355-367
13. Zeng L. The phylogeny of subgenus Clavicarpa (Impatiens). Shanxi Normal University. 2016.
14. Wang Q, Yang XD, Xia CY, LI J, YU SX. Pollen Morphology of Impatiens sect. Impatiens (Balsaminaceae) Acta Horticulturae Sinica. 2020; 47: 893-906.
15. Yu SX, Janssens SB, Zhu XY, Lidén M, Gao TG, et al. Phylogeny of Impatiens (Balsaminaceae): integrating molecular and morphological evidence into a new classification. Cladistics. 2015; 32: 179-197
16. Shimizu T. A comment on the limestone flora of Thailand, with special reference to Impatiens. Acta Phytotaxonomica et Geobotanica. 1979; 30: 180-187.
17. Janssens SB, Lens F, Dressler S, Geuten K, Smets E, Vinckier S. Palynological variation in Balsaminoid Ericales. II. Balsaminaceae, Tetrameristacee, Pellicieraceae, Pellicieraceae and general conclusions. Annals of Botany. 2005; 96: 1061-1073.
18. Yuan YM, Song YI, Geuten K, Rahelivololona E, Wohlhauser S, et al. Phylogeny and biogeography of Balsaminaceae inferred from ITS sequence date. Taxon. 2004; 53: 391-403.
19. Volkmar U, Smets EF, Lenz H, Janssens SB. Intron evolution in a phylogenetic perspective: Divergent trends in the two copies of the duplicated DEF gene in Impatiens L. (Balsaminaceae). Journal of Systematics and Evolution. 2014; 52: 134-148.
20. Chorev M, Carmel L. The function of introns. Frontiers in Genetics. 2012; 3.
21. Jo BS, Choi SS. Introns: the functional benefits of introns in genomes. Genomics & informatics. 2015; 13: 112-118.
22. Poverennaya IV, Roytberg MA. Spliceosomal introns: features, functions, and evolution. Biochemistry (Moscow). 2020; 85: 725-734.
23. Davis JC, Petrov DA. Preferential duplication of conserved proteins in eukaryotic genomes. PLOS Biology. 2004; 2: E55.
24. Brunet FG, Crollius HR, Paris M, Aury JM, Gibert P, et al. Gene loss and evolutionary rates following whole genome duplication in teleost fishes. Molecular Biology and Evolution. 2006; 23: 1808-1816.
25. Semon M, Wolfe KH. Preferential sub functionalization of slow-evolving genes after allopolyploidization in Xenopus laevis. Proceedings of the National Academy of Sciences of the United States of America. 2008; 105: 8333-8338.
26. Geng SF, Zhao YL, Tang LC, Zhang RZ, Sun MH, et al. Molecular evolution of two duplicated CDPK genes CPK7 and CPK12 in grass species: A case study in wheat (Triticum aestivum L.). Gene. 2011; 475: 94-103.
27. Yang HL, Liu YJ, Wang CL, Zeng QY. Molecular evolution of trehalose-6-phosphate synthase (TPS) gene family in Populus, Arabidopsis and rice. PLOS One. 2012; 7: e42438.
28. Zhang SH, Qu LH. Genome Evolution and the Evolution of Genes in Introns. Acta Scientiarum Naturalium Universitatis Sunyatsen. 1999; 38: 49-53.

29. Yang YY. Studies on Genetic Diversity of Impatiens L. Germplasm Resources in Guizhou. Southwest Forestry University. 2018.

30. Kuang RP. Study on the Flora in Guizhou/southern Sichuan and Phylogenetic problems of Impatiens L. Hunan Normal University. 2015.

31. Yu SX. Revision of Impatiens L. (Balsaminaceae) in Guangxi, with a study on the phylogeny of the genus in China. The institute of botany, Chinese academy of sciences. 2008.

32. Rivest SA, Austen EJ, Forrest JRK. Foliage affects colour preference in bumblebees (Bombus Impatiens): a test in a three-dimensional artificial environment. Evolutionary Ecology. 2017; 31: 435-446.