

RESEARCH

Open Access



Complete plastid genome structure of 13 Asian *Justicia* (Acanthaceae) species: comparative genomics and phylogenetic analyses

Zhengyang Niu^{1,2}, Zheli Lin^{1,3}, Yi Tong⁴, Xin Chen⁵ and Yunfei Deng^{1*}

Abstract

Background *Justicia* L. is the largest genus in Acanthaceae Juss. and widely distributed in tropical and subtropical regions of the world. Previous phylogenetic studies have proposed a general phylogenetic framework for *Justicia* based on several molecular markers. However, their studies were mainly focused on resolution of phylogenetic issues of *Justicia* in Africa, Australia and South America due to limited sampling from Asia. Additionally, although *Justicia* plants are of high medical and ornamental values, little research on its genetics was reported. Therefore, to improve the understanding of its genomic structure and relationships among Asian *Justicia* plants, we sequenced complete chloroplast (cp.) genomes of 12 Asian plants and combined with the previously published cp. genome of *Justicia leptostachya* Hemsl. for further comparative genomics and phylogenetic analyses.

Results All the cp. genomes exhibit a typical quadripartite structure without genomic rearrangement and gene loss. Their sizes range from 148,374 to 151,739 bp, including a large single copy (LSC, 81,434–83,676 bp), a small single copy (SSC, 16,833–17,507 bp) and two inverted repeats (IR, 24,947–25,549 bp). GC contents range from 38.1 to 38.4%. All the plastomes contain 114 genes, including 80 protein-coding genes, 30 tRNAs and 4 rRNAs. IR variation and repetitive sequences analyses both indicated that *Justicia grossa* C. B. Clarke is different from other *Justicia* species because its lengths of *ndhF* and *ycf1* in IRs are shorter than others and it is richest in SSRs and dispersed repeats. The *ycf1* gene was identified as the candidate DNA barcode for the genus *Justicia*. Our phylogenetic results showed that *Justicia* is a polyphyletic group, which is consistent with previous studies. Among them, *J. grossa* belongs to subtribe Tetramerinae of tribe Justicieae while the other *Justicia* members belong to subtribe Justiciinae. Therefore, based on morphological and molecular evidence, *J. grossa* should be undoubtedly recognized as a new genus. Interestingly, the evolutionary history of *Justicia* was discovered to be congruent with the morphology evolution.

Conclusion Our study not only elucidates basic features of *Justicia* whole plastomes, but also sheds light on interspecific relationships of Asian *Justicia* plants for the first time.

Keywords Asian *Justicia* plants, *Justicia Grossa*, Complete chloroplast genome, Comparative genomics, Phylogeny

*Correspondence:
Yunfei Deng
yfdeng@scbg.ac.cn

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



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

Introduction

Justicia L. is the largest and most taxonomically complex genus in the tribe Justicieae of Acanthaceae [1–8]. It comprises approximately 700 species widely distributed in the tropical and subtropical regions of the world [9, 10]. This genus is characterized by the 2-lipped corolla with the bilobed upper lip and trilobed lower lip, two bitheous stamens, usually one theca above the other and the lower one with a spur at the base [1, 5, 11, 12]. With approximately 150 species, the tropical and subtropical regions of Asia are one of the diversity centers of the genus [3, 5, 7, 13–18]. In Asia, many *Justicia* plants are widely cultivated for ornamental or medical values [3, 19–21]. For example, *Justicia adhatoda* L., *Justicia betonica* L., *Justicia grossa* C.B. Clarke and *Justicia latiflora* Hemsl. are commonly cultivated for ornamental in the gardens [21]. And *J. adhatoda*, *Justicia gendarussa* N. J. Burman and *Justicia procumbens* L. also have high medicinal value [20, 22, 23]. However, despite high economical values of *Justicia* plants, few reports on its genomics were available [24]. Therefore, to improve our understanding on plastid genomes of these economically important plants and provide useful genetic information for conservation and breeding of them in the future, it is necessary to carry out relevant genetic research.

In addition, due to extensive geographic distributions and high biological diversity, the infrageneric classification of *Justicia* has been controversial for a long time [1, 25–30]. Up to now, there are two divergent approaches in the generic delimitation of *Justicia*. One is to divide *Justicia* into several small segregated genera [25, 26, 31–33], and another is to adopt a broad sense of *Justicia* by dividing into several sections [1, 5, 27–30, 34–39]. In the former, Bremekamp separated *Justicia* s.l. into dozens of genera and published several new genera [26, 31, 32, 40–42]. His treatment was followed by some authors [33, 43–49]. In contrast, Graham [1] adopted a broad concept of the genus and reduced more than 70 names to the synonymies of *Justicia* and divided the genus into nine sections and seven subsections. Her treatment was widely accepted in the recently published flora works [3, 5, 39, 50–55].

However, recent phylogenetic studies indicated that *Justicia* s.l. is a polyphyletic group with its members randomly nested within other genera in tribe Justicieae [9–12], suggesting that previous classification of *Justicia* is problematic. Then, Kiel et al. [9] proposed to divide Justicieae into ten informal clades, of which nine for Old World (OW) species and one for New Old (NW) species. Although their suggestion on the classification of Justicieae has been most comprehensive until now, molecular data of *Justicia* in their studies were mainly based on samples collected from Africa, Australia and America, but few from Asia. For example, some Asian genera

separated from *Justicia*, e.g., *Calymmostachya* Bremek. [32], *Mananthes* Bremek. [26], *Plegmatolemma* Bremek. [32], were not involved in their analysis. Thus, more genetic resources of *Justicia* in Asia need to be supplemented for completion of the evolutionary history of *Justicia* in the future.

The genus *Justicia* from Asia has never been revised except for the regional revisionary works for some countries including China [3, 33], Bangladesh [56], Pakistan [57], Sri Lanka [58], etc. China is one of the diversity centers of *Justicia* in Asia. The most comprehensive works of *Justicia* from China were done by Hu and her colleagues [3, 33]. Hu [33] recognized 44 species in seven genera in Chinese edition of *Flora Reipublicae Popularis Sinica* to follow the narrow sense of generic delimitation proposed by Bremekamp [26, 31, 32]. But, nine years later, Hu et al. [3] adopted the broad sense of generic circumscription proposed by Graham [1] and recorded 43 species in English edition of *Flora of China*. Later, Deng et al. [11] found that *Justicia microdonta* W.W. Sm. is quite different from other *Justicia* plants by having two staminodes and two fertile stamens with both anther-thecae spurred at base, and might be a member of subtribe Graptophyllinae. Therefore, they established a new genus *Wuacanthus* Y.F. Deng et al. for this species. It is implied that the relationships among the remaining Asian species of *Justicia* s. l. are still poorly understood, and thus the further studies on the phylogenetic research among Asian plants is necessary.

The complete cp. genome is characterized by haploid inheritance, relatively small genome and low substitution rates compared with mitochondrial and nuclear genomes, and thus widely used in recent studies of plant phylogeny, phylogeography and population genetics [59–65]. Its molecular structure is highly conservative in most angiosperms, with a double-stranded circular structure divided into four regions, including a large single copy (LSC) region, a small single copy (SSC) region and a pair of inverted repeats (IRs) [66, 67]. It is typically 107–218 kb in genomic size in most land plants, encoding about 100–130 unique genes, mostly containing about 70–80 protein-coding genes, 28–32 tRNAs and 4 rRNAs [68, 69]. Besides, recent advances in sequencing technology and bioinformatic analysis tools have made the acquisition of complete cp. genomes both convenient and cost-effective [70]. Therefore, based on whole cp. genome data, more information sites could be accessible. Thus, our obtained variable sites from whole plastomes are sufficient than previous molecular markers in re-evaluation of the evolutionary histories of some difficult taxa, including some major clades in angiosperms, such as basal lamiid [71] and monocot [72], and other taxa below order, such as Orchidaceae Juss. [63], Ulmaceae Mirb. [73], subtribe Melocanninae Benth. (Poaceae Barnhart)

[62], *Horsfieldia* Willd. (Menispermaceae Juss.) [74] and *Oreocnide* Miq. (Urticaceae Juss.) [75]. Additionally, evolutionary rates of coding and non-coding regions of the plastomes are incongruous, suggesting great applicability to screen potential DNA barcodes at various taxonomic levels [76–78]. Therefore, there is no doubt that whole plastid genomes may provide critical insights into historically difficult relationships at different taxonomic levels. Moreover, some IR expansion and contraction events [79], genomic rearrangement [80], gene loss and pseudogenization [81, 82] have also attracted much attention due to their particularities. Thus, the whole cp. genome is definitely an efficient tool for species identification at the genomic level [83–87].

In our study, a total of 12 *Justicia* complete cp. genomes were newly sequenced and assembled, then combined with the previously published cp. genome *J. leptostachya* for further genome comparison analyses. This study aims to (i) understand basic features of *Justicia* plastomes, including genomic size, organization and gene compositions, (ii) find interspecific variation at the genomic structure level, (iii) identify some hypervariable regions and special repetitive sequences for species identification, and (iv) improve our understanding of phylogenetic relationships of these Asian *Justicia* plants, which is also useful to provide baseline information for further completion of the evolutionary history of *Justicia* in the future.

Results
Basic characteristics of *Justicia* complete chloroplast genomes and nrDNAs

A total of 3,774,489–22,941,320 unique reads were recruited from about 2 Gb clean reads for plastome *de novo* assemblies (Table S1). The average base-coverages of *Justicia* cp. genomes vary from 96X to 521X with 150 bp read length for each sample. The 13 *Justicia* cp. genomes sizes vary from 148,374 bp (*J. latiflora*) to 151,739 bp (*Justicia quadrifaria* (Nees) T. Anderson) and their overall GC content range from 38.1 to 38.4% (Table 1). All the cp. genomes exhibit the identical typical quadripartite structure, comprising of a large single copy region (LSC) from 81,434 bp to 83,676 bp, a small single copy region (SSC) from 16,833 bp to 17,507 bp and a pair of IR regions (IRa/IRb) from 24,947 bp to 25,549 bp.

Gene number, order and directions are consistent in the 13 *Justicia* cp. genomes (Fig. 1A). All the cp. genomes share 114 unique genes, containing 80 protein-coding genes, 30 tRNAs and 4 rRNAs (Table 1). According to its location, 62 are located in LSC region, 12 are in SSC region and 6 are in IR regions. As for gene categories, 61 genes are relevant to the gene expression, and 43 genes are associated with photosynthesis (Table 2). According to the sizes of all the protein-coding genes (Table S3), *ycf2* is the longest from 6723 bp (*J. gendarussa*) to 6780 bp (*J. grossa*, *J. demissa* N. H. Xia & Y. F. Deng, *J. mollissima* (Nees) Y. F. Deng & T. F. Daniel and *J. procumbens*), while *petN* is the shortest with 90 bp identical in all the plastomes. Of the 80 unique protein-coding

Table 1 General characteristics of 13 Asian *Justicia* complete chloroplast genomes

	<i>J. adhatoda</i>	<i>J. betonica</i>	<i>J. demissa</i>	<i>J. gendarussa</i>	<i>J. grossa</i>	<i>J. latiflora</i>	<i>J. leptostachya</i>
Accession number	MN848249	MN848244	MN885664	MN848252	MN848251	MN848246	MK389502
Total size (bp)	149,503	151,005	150,208	149,735	150,469	148,374	149,227
LSC length (bp)	82,600	82,809	82,326	82,373	82,536	81,434	82,114
SSC length (bp)	17,009	17,182	16,970	17,218	17,507	16,866	16,975
IR length (bp)	24,947	25,507	25,456	25,072	25,213	25,037	25,069
Number of Genes	114	114	114	114	114	114	114
Number of PCGs	80	80	80	80	80	80	80
Number of tRNAs	30	30	30	30	30	30	30
Number of rRNAs	4	4	4	4	4	4	4
Overall GC (%)	38.3	38.3	38.4	38.3	38.4	38.1	38.2
	<i>J. lianshanica</i>	<i>J. mollissima</i>	<i>J. patentiflora</i>	<i>J. procumbens</i>	<i>J. quadrifaria</i>	<i>J. vagabunda</i>	
Accession number	MN885665	MN848247	MN848248	MN848245	MN848243	MN848250	
Total size (bp)	148,574	150,513	149,018	150,471	151,739	151,247	
LSC length (bp)	81,776	82,811	82,031	82,426	83,676	83,343	
SSC length (bp)	16,868	17,010	16,833	16,947	16,999	17,040	
IR length (bp)	24,965	25,346	25,077	25,549	25,532	25,432	
Number of Genes	114	114	114	114	114	114	
Number of PCGs	80	80	80	80	80	80	
Number of tRNAs	30	30	30	30	30	30	
Number of rRNAs	4	4	4	4	4	4	
Overall GC (%)	38.1	38.2	38.1	38.3	38.2	38.3	

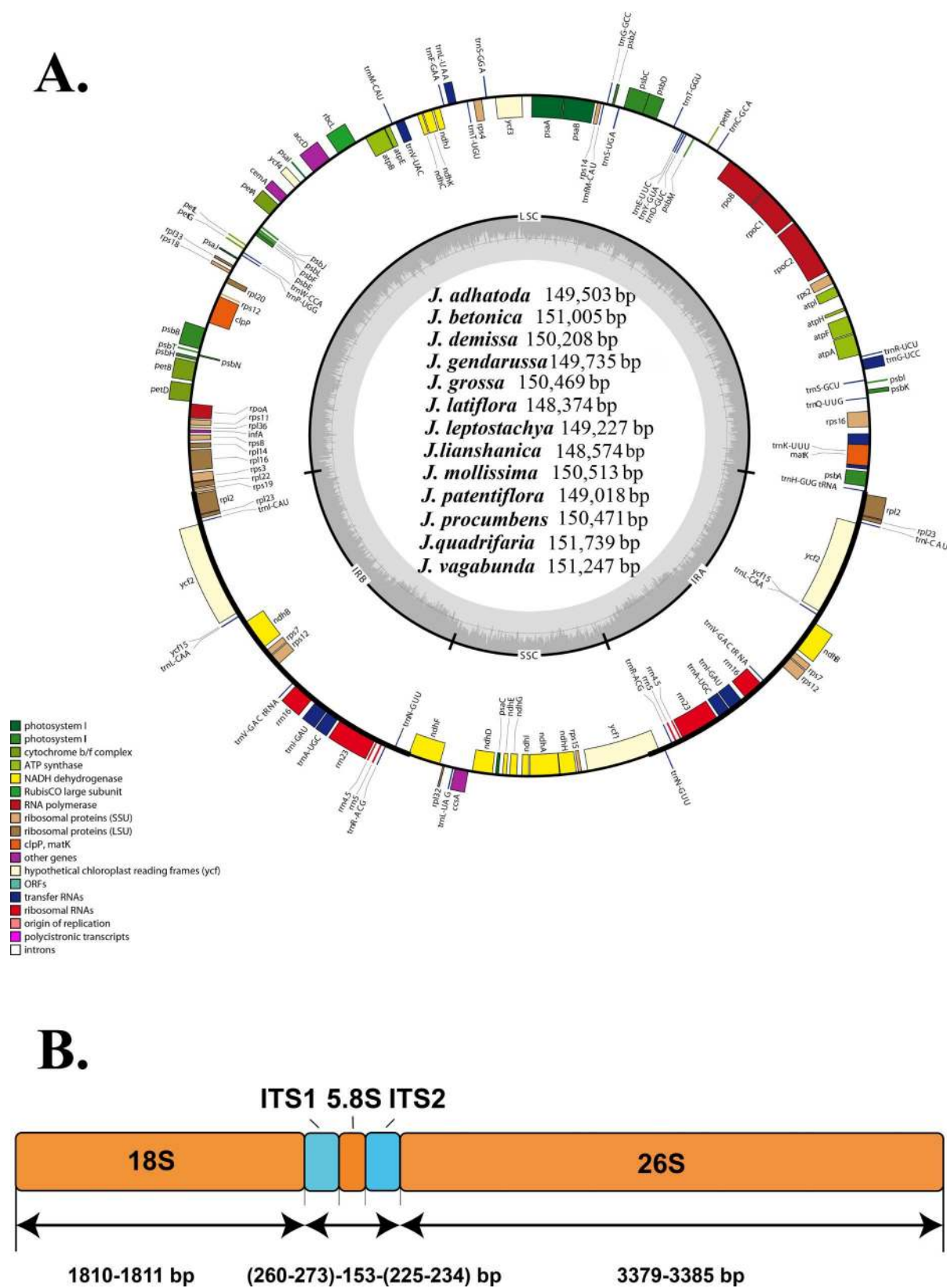


Table 2 Gene contents in the chloroplast genomes of 13 *Justicia* species

Genes category	Groups of genes	Name of genes
Self-replication	Ribosomal RNAs	<i>rrn4.5</i> (×2), <i>rrn5</i> (×2), <i>rrn16</i> (×2), <i>rrn23</i> (×2)
	Transfer RNAs	<i>trnA</i> -UGC*(×2), <i>trnC</i> -GCA, <i>trnD</i> -GUC, <i>trnE</i> -UUC, <i>trnF</i> -GAA, <i>trnM</i> -CAU, <i>trnG</i> -GCC, <i>trnG</i> -UCC*, <i>trnH</i> -GUG, <i>trnI</i> -CAU(×2), <i>trnI</i> -GAU*(×2), <i>trnK</i> -UUU*, <i>trnL</i> -CAA(×2), <i>trnL</i> -UAA*, <i>trnL</i> -UAG, <i>trnM</i> -CAU, <i>trnN</i> -GUU(×2), <i>trnP</i> -UGG, <i>trnQ</i> -UUG, <i>trnR</i> -ACG(×2), <i>trnR</i> -UCU, <i>trnS</i> -GCU, <i>trnS</i> -GGA, <i>trnS</i> -UGA, <i>trnT</i> -GGU, <i>trnT</i> -UGU, <i>trnV</i> -GAC(×2), <i>trnV</i> -UAC*, <i>trnW</i> -CCA, <i>trnY</i> -GUA
	Large ribosomal subunits	<i>rpl2</i> *, <i>rpl14</i> , <i>rpl16</i> *, <i>rpl20</i> , <i>rpl22</i> , <i>rpl23</i> (×2), <i>rpl32</i> , <i>rpl33</i> , <i>rpl36</i>
	Small ribosomal subunits	<i>rps2</i> , <i>rps3</i> , <i>rps4</i> , <i>rps7</i> (×2), <i>rps8</i> , <i>rps11</i> , <i>rps12</i> ***(×2) ^a , <i>rps14</i> , <i>rps15</i> , <i>rps16</i> *, <i>rps18</i> , <i>rps19</i>
	DNA-dependent RNA polymerase	<i>rpoA</i> , <i>rpoB</i> , <i>rpoC1</i> *, <i>rpoC2</i>
Phytosynthesis	Subunits of photosystem I	<i>psaA</i> , <i>psaB</i> , <i>psaC</i> , <i>psaI</i> , <i>psaJ</i>
	Subunits of Photosystem II	<i>psbA</i> , <i>psbB</i> , <i>psbC</i> , <i>psbD</i> , <i>psbE</i> , <i>psbF</i> , <i>psbH</i> , <i>psbI</i> , <i>psbJ</i> , <i>psbK</i> , <i>psbL</i> , <i>psbM</i> , <i>psbN</i> , <i>psbT</i> , <i>psbZ</i>
	Subunits of NADH dehydrogenase	<i>ndhA</i> *, <i>ndhB</i> *(×2), <i>ndhC</i> , <i>ndhD</i> , <i>ndhE</i> , <i>ndhF</i> , <i>ndhG</i> , <i>ndhH</i> , <i>ndhI</i> , <i>ndhJ</i> , <i>ndhK</i>
	Subunits of ATP synthase	<i>atpA</i> , <i>atpB</i> , <i>atpE</i> , <i>atpF</i> *, <i>atpH</i> , <i>atpI</i>
	Subunits of cytochrome	<i>petA</i> , <i>petB</i> *, <i>petD</i> *, <i>petG</i> , <i>petL</i> , <i>petN</i>
	Large subunit of Rubisco	<i>rbcl</i>
Other genes	Maturase	<i>matK</i>
	Protease	<i>clpP</i> **
	Subunit of acetyl-CoA	<i>accD</i>
	Envelope membrane protein	<i>cemA</i>
	C-type cytochrome synthesis gene	<i>ccsA</i>
	Translation initiation factor	<i>infA</i>
Function uncertain	Conserved open reading frames	<i>ycf1</i> , <i>ycf2</i> , <i>ycf3</i> **, <i>ycf4</i> , <i>ycf15</i> (×2)

*gene containing one intron, **gene containing two introns, ^atrans-splinting gene, (×2) shows genes have two copies

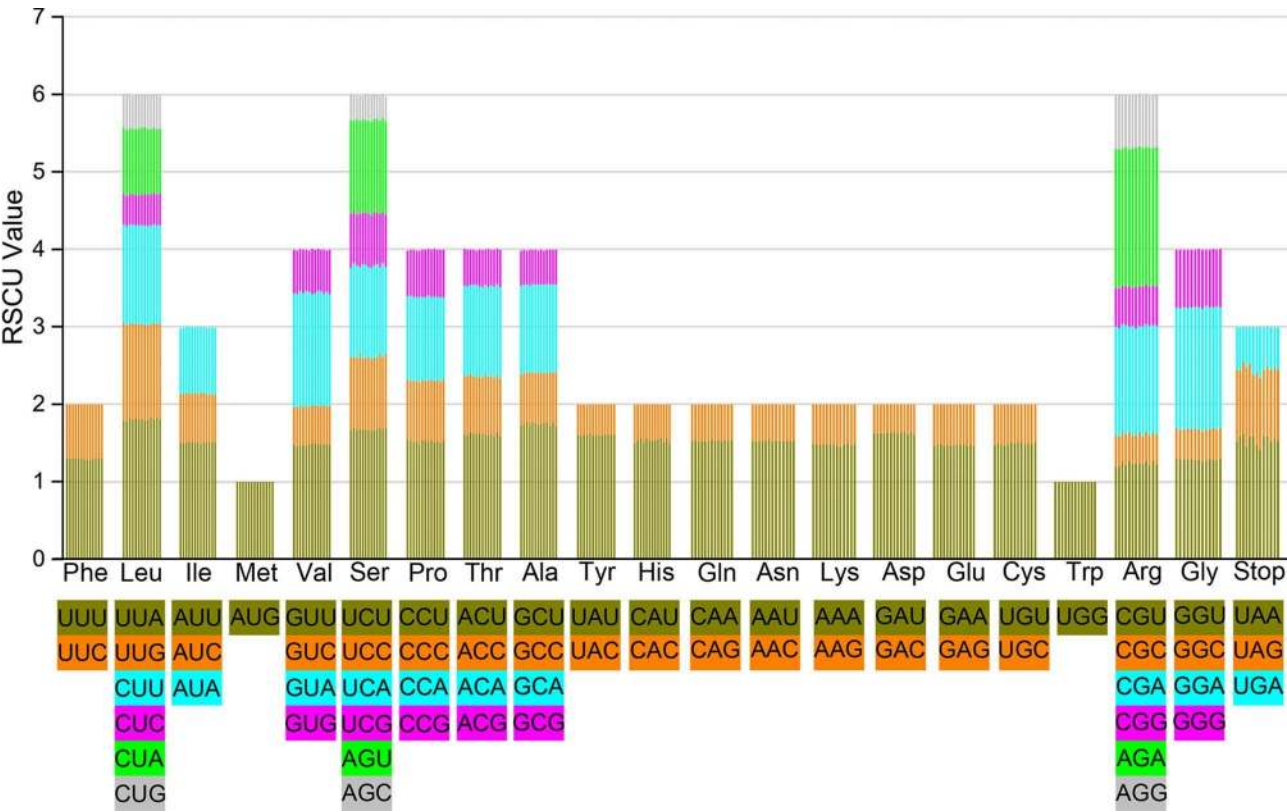


Fig. 2 Relative synonymous codon usage (RSCU) in all protein-coding genes of the 13 plastomes. The histogram from the left-hand side of each amino acid shows codon usage value within *Justicia* (From left to right: *J. quadrifaria*, *J. betonica*, *J. lianshanica*, *J. mollissima*, *J. patentiflora*, *J. adhatoda*, *J. vagabunda*, *J. grossa*, *J. gendarussa*, *J. leptostachya*, *J. procumbens*, *J. latiflora* and *J. demissa*). Codons encoding 20 amino acids and the stop codon are displayed in rectangular shapes of different colors

genes, 57 are identical in length among different species, while 23 are variable as such. In addition, 15 genes contain one intron (Table S4), including *atpF*, *ndhA*, *ndhB*, *petB*, *petD*, *rpl2*, *rpl16*, *rpoC1*, *rps12*, *rps16*, *trnA-UGC*, *trnG-UCC*, *trnI-GAU*, *trnK-UUU*, *trnL-UAA*, *trnV-UAC*, *clpP* and *ycf3*, while two genes (*clpP* and *ycf3*) have two introns. Among them, *ndhA*, *rpl2*, *rpoC1*, *rps16* and *ycf3* genes vary in size at the interspecific level.

Additionally, some particular genes in *Justicia* plastomes were also identified in our study. First, six genes were determined as partially overlapped genes, including *trnK-UUU/matK*, *atpB/atpE* and *psbC/psbD*. Secondly, *rps12* gene was identified as a trans-splicing gene with 5' exon located in LSC and 3' exon duplicated and distributed in two IR regions. Thirdly, the gene *ycf15* in *J. adhatoda* cp genome was found to be about half (63 bp) the length of the others (132 bp).

According to the codon usage bias analysis, all the protein-coding genes (77,985–78,681 bp) of the 13 *Justicia* cp. genomes, encoding 25,995–26,227 codons, were investigated (Fig. 2, Table S5). Our results showed that all *Justicia* species are similar in amino acid patterns. Among them, Leucine is encoded by the largest number of codons from 2801 to 2852, while Cysteine is the least with 69–79. Besides, a total of 28 codons are directly involved in tRNA synthesis. Most amino acids are encoded with at least two synonymous codons except methionine (Met) and tryptophan (Trp). A total of 77 protein-coding genes identified in *Justicia* cp. genomes started with an AUG codon, but *rps19* and *psbC* start with GUG while *ndhD* contains ACG instead.

For the tandemly repeated nrDNAs, our *de novo* assembly obtained 5,819 bp (*J. grossa*) to 5,846 bp (*J. patentiflora* Hemsl.) comprising 18 S (1,810–1,811 bp), 5.8 S (153 bp), and 26 S (3,379–3,385 bp) ribosomal RNA gene along with two internal transcribed spacer (ITS) I (260–273 bp) and II (225–234 bp) in the middle (Fig. 1B).

IR contraction and expansion

In our study, IR/SC junctions of cp. genomes of the 13 *Justicia* species and seven species of other genera in Acanthaceae were compared and visualized (Fig. 3). First, the gene *rps19* stretches across LSC and IRb regions of cp. genomes of all *Justicia* species and *Dicliptera acuminata* (Ruiz & Pav.) Juss., with 5' end of the *rps19* located in the IRbs (82–104 bp) and 3' end located in the LSCs (175–203 bp). Therefore, *rps19* gene creates a pseudogene of the 5' end of this gene (*Ψrps19*) in IRa. However, in *Rungia pectinata* (L.) Nees and *Ruellia brittoniana* Leonard, it is found that *rpl22* and *ycf2* gene rather than *rps19* gene span the junction of LSC/IRb borders. Meanwhile, the gene *rpl22* duplicates a pseudogene (*Ψrpl22*) in the border of LSC/IRa of *R. pectinata*, but *ycf2* gene in *R. brittoniana* not as such. Different from others,

genes *rpl22*, *rps19* and *ycf2* are closed to the junction of LSC/IRb in plastomes of *Clinacanthus nutans* (Burm.f.) Lindau, *Pseuderanthemum haikangense* C. Y. Wu et H. S. Lo, *Echinacanthus lofuensis* (H.Lév.) J.R.I.Wood and *Strobilanthes cusia* (Nees) Kuntze. Secondly, the tRNA genes *trnH-GUG* and *Ψrps19* are adjacent to the junctions of LSC/IRa in cp. genomes of *Justicia* and *D. acuminata*. However, the genes *rps19* in *P. haikangense* and *E. lofuensis* are duplicated due to this gene fully located in IRs. Additionally, *rpl2* and *ycf15* gene are adjacent to the LSC/IRa borders of *C. nutans* and *R. brittoniana* while (*Ψ*) *rps19* genes are adjacent to the same locations in others. Particularly, *psbA* was found to be a crossing gene within the LSC/IRa border of *S. cusia*. Thirdly, it is discovered that *ycf1* genes of plastomes of most genera span SSC/IRa border with the exception of *C. nutans* fully located in SSC region with 1,118 bp far away from the junction. Notably, the *ycf1* gene is only 576 bp in IRa region of *J. grossa* cp. genome, but about 800 bp in all of the other species. Fourthly, the pseudogene *Ψycf1* is a part of *ycf1* protein-coding gene copy with the 5' end located in the IRb region, with the sizes of 647 bp (*J. grossa*) to 848 bp (*E. lofuensis*). Meanwhile, most *Ψycf1* can cross SSC and IRb regions, but those of *Justicia lianshanica* (H.S.Lo) H.S.Lo, *D. acuminata*, *R. brittoniana* and *S. cusia* are fully located in IRb regions. Most *ndhF* genes are within the SSC/IRb borders with the exception of *C. nutans* fully located in SSC region. Notably, the length of *ndhF* gene in IRb of *J. grossa* plastome is shorter (35 bp) than that in other *Justicia* species (100–129 bp), but similar in length with *R. brittoniana* (37 bp) and *S. cusia* (44 bp).

Genome divergence comparison

To visualize hypervariable regions, multiple sequence alignments were implemented using the program mVISTA (Fig. 4). The divergence of non-coding regions (CNS) is greater than that of coding regions (CDS), while LSC and SSC regions are more variable than IR regions. According to global alignment, the most highly divergent regions in intergenic spacer are *rps16-trnQ*, *trnS-trnG*, *atpF-atpH*, *rpoB-trnC*, *trnE-psbD*, *psbZ-trnfM*, *rps4-trnT*, *trnF-ndhJ*, *ndhC-trnV*, *petA-psbJ*, *psbE-petL*, *rpl32-trnL* and *rps15-ycf1*, while divergent regions in coding regions are *atpF*, *rpl16* and *ycf1*.

A sliding window was used to compare hotspots regions among 13 *Justicia* species. 693 representative loci were divided into two groups, which are composed of two clades of staggered loci (Fig. 5A and B). The nucleotide diversity (P_i) value enormously ranges from 0 to 0.072, and the mean value is 0.0219. In general, P_i value of SC regions is significantly greater than that of IR regions. To exactly analyze interspecific variations, eight highly variable regions ($P_i > 0.06$) were identified, including *trnT-trnL* ($P_i=0.07944$), *ycf1* ($P_i=0.07521$),

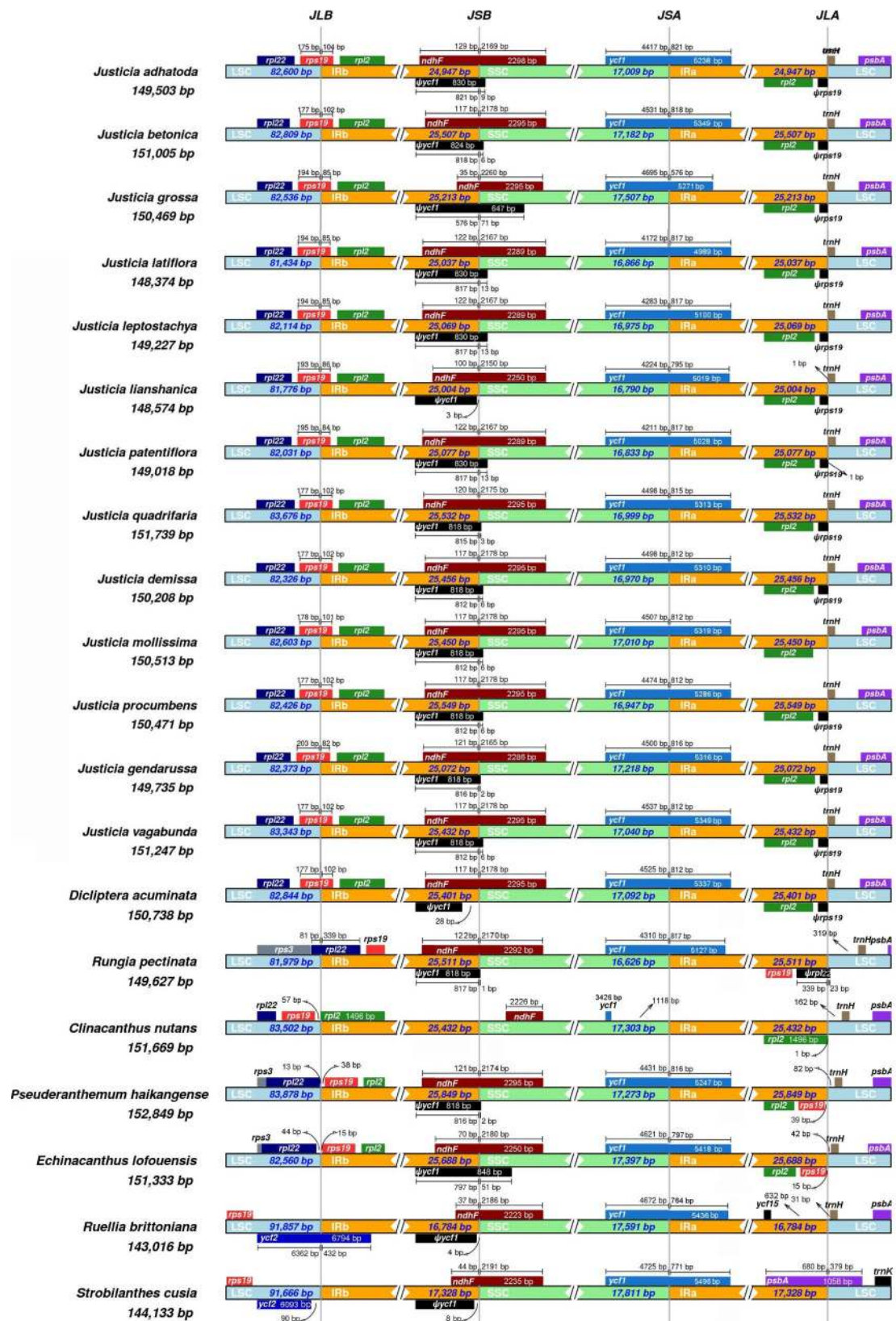


Fig. 3 Comparison of IR/SC boundary regions of the 13 *Justicia* species and seven species of other genera of Acanthaceae

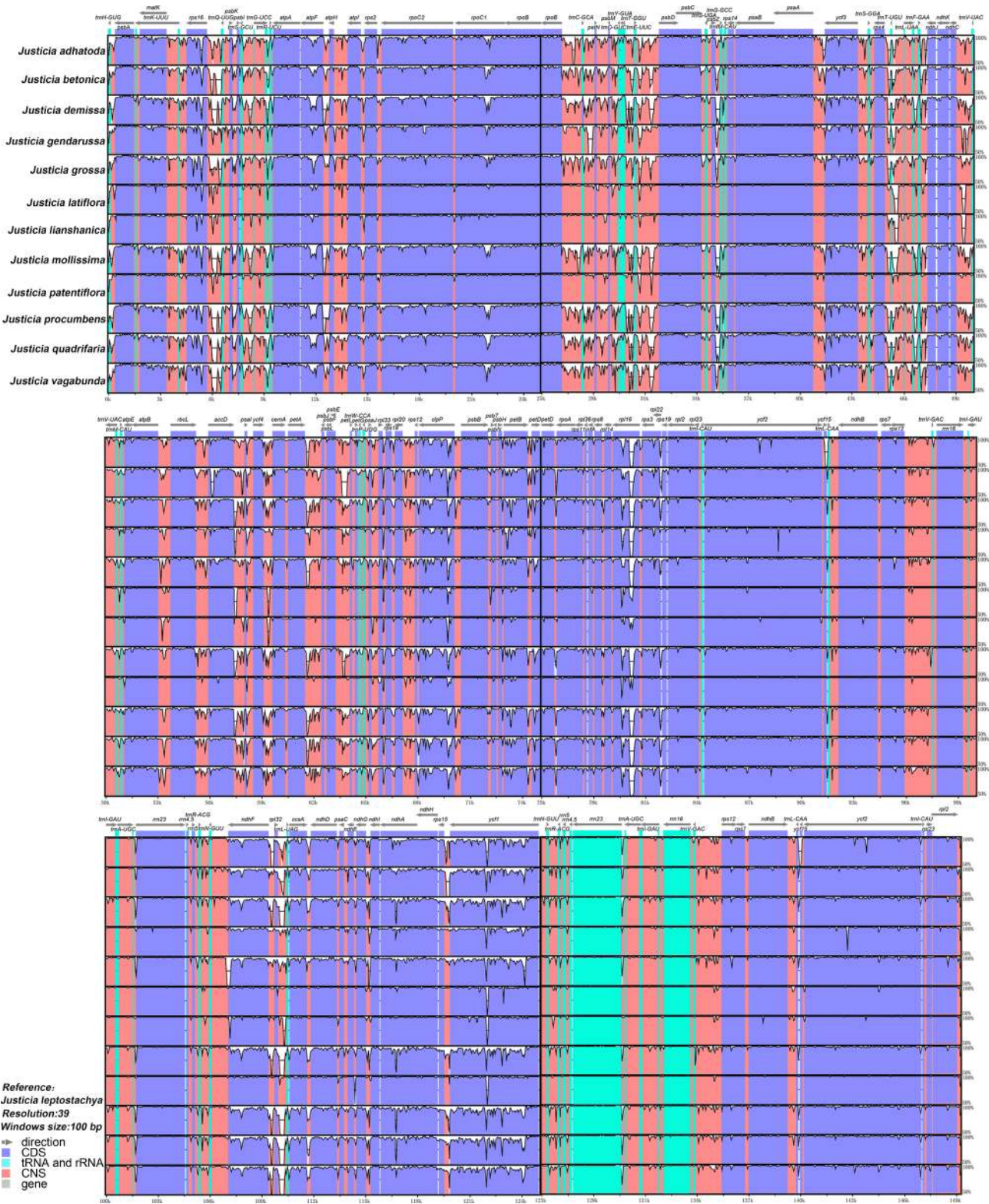


Fig. 4 Genome divergence plots of the 13 cp. genomes with *J. leptostachya* as a reference based on visualized global alignment. Grey arrows and thick black lines above the alignment indicate genes with their orientations and directions. Protein-coding regions (exon), non-coding regions (CNS) and untranslated regions (UTR) are marked in red, blue and green, respectively. A cut-off of 70% identity was used for the plots, and the Y-scale represents the percent identity from 50–100%

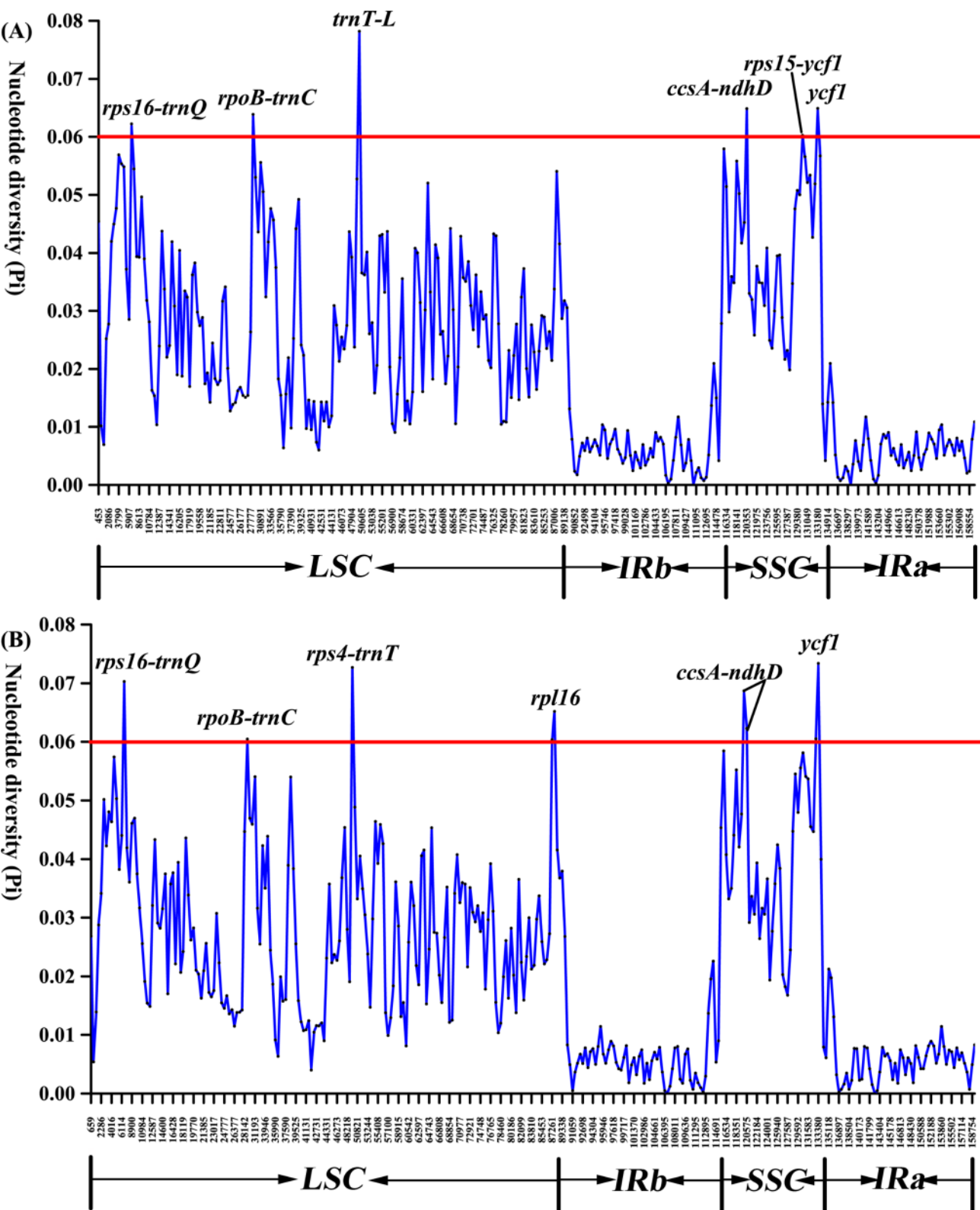


Fig. 5 Sliding window analysis of the 13 cp. genomes alignment. Window length: 600 bp; step size: 200 bp. X-axis: position of the midpoint of a window. Y-axis: nucleotide diversity of each window. **(A)** Nucleotide diversity of A-clade dataset; **(B)** Nucleotide diversity of B-clade dataset

rps4-trnT (Pi=0.07203), *rps16-trnQ* (Pi=0.06823), *ccsA-ndhD* (Pi=0.06671), *rpoB-trnC* (Pi=0.06387), *rpl16* (Pi=0.06316), and *rps15-ycf1* (Pi=0.06047). According to their locations, six of them are located in LSC region while two are in SSC.

Repetitive sequences analyses

As short tandem repeats with 1–6 nucleotides units, SSRs are widely dispersed in *Justicia* plastomes. The number of SSRs varies unevenly in 13 *Justicia* cp. genomes (Fig. 6, Table S6). Statistically, SSRs fluctuate within a range of 39–59, comprising 23–41 SSRs in LSC regions, 9–14 SSRs in SSC regions and 2–11 SSRs in IR regions (Fig. 6B). According to genomic regions, 8–16 SSRs were identified in coding genes, 23–37 SSRs in intergenic spacer and 4–9 SSRs in introns (Fig. 6C). Most SSRs were detected in LSC and intergenic spacer regions, whereas few SSRs were detected in IR regions and introns. The most abundant SSRs (59) were identified in *J. grossa*, while the others range from 39 (*J. adhatoda*) to 51 (*J. demissa*). For base contents of SSRs, all the *Justicia* cp. genomes are made up of 90% A/T and 10% C/G bases (Table S6). Among them, mononucleotide is the most abundant units and hexanucleotide was only identified in five species including *J. leptostachya*, *J. latiflora*, *J. quadrifaria*, *J. adhatoda* and *J. lianshanica*, of which *J. lianshanica* has the largest number (5) (Table 3; Fig. 6A). Notably, *J. grossa* has the largest number of mononucleotide (37) and tetranucleotide (13) repeats compared with other members of *Justicia* (Fig. 6A).

A total of 22–57 dispersed repeats were also defined in the 13 cp. genomes, including forward, palindromic, reverse and complement repeats (Fig. 7A). Among them, palindromic repeats are the richest in all the *Justicia* cp. genomes. Besides, the maximum number of dispersed repeats were detected in *J. grossa* (56) compared with others. In terms of repeat length, most dispersed repeats concentrate on lengths of 20–25 bp, with the exception of three species having dispersed repeats of over 50 bp, including *J. gendarussa* (1), *J. mollissima* (1) and *J. grossa* (3) (Fig. 7B).

Phylogenetic analysis

The whole cp. genome data matrix consisting of 62 sequences is 188,699 bp in total length. It is characterized by sequence divergence with 56,714 variable sites, including 38,522 parsimony informative sites and 18,192 singleton variable sites. The ITS data matrix made up of 63 sequences is 988 bp in total length. It is characterized by sequence divergence with 457 variable sites, including 312 parsimony informative sites and 145 singleton variable sites.

Because the reconstructed ML tree and BI tree share the same topology, we only showed the ML phylogram

with bootstrap (BS) and posterior probability (PP) values labeled near each node (Fig. 8). Our phylogenetic results indicated that phylogenetic relationships of the 13 *Justicia* species based on three datasets (WCG, PCG and ITS) exhibit identical tree topologies (Fig. 8, Fig. S1–S4).

According to our phylogenetic results, a robust phylogenetic framework for four subfamilies of Acanthaceae is as follows: (Nelsonioideae(Acanthoideae(Thunbergioideae+Avicennioideae))). Additionally, the stable framework of most tribes of Acanthaceae with the exception of Neuracanthaceae and Whitfieldieae is also exhibited, that is (Nelsonieae((Acanthaceae((Andrographideae+Barlerieae) (Justicieae+Ruellieae))) (Thunbergieae+Avicenniceae))) (Fig. 8). Importantly, all the nodes of subfamilies and tribes are strongly supported (BS=100, PP=1.0) in our study.

Additionally, our results also strongly support (BS=100, PP=1.0) that *Justicia* is a polyphyletic group and suggest to divide all sampled *Justicia* species in the present study into three informal clades—Clade I, II and III (Fig. 8). In Clade I, *J. grossa* (the type of *Justicia* sect. *Grossa*) is the earliest diverging species sister to the monospecific genus *Clinacanthus*. And both of them belong to subtribe Tetramerinae of tribe Justicieae. However, Clade II and III consist of all the remaining *Justicia* species and three other genera, which belong to subtribe Justiciinae of tribe Justicieae. Clade II includes a single sampled species of *Rungia* Nees and six sampled species of *Justicia*, including *J. gendarussa*, *J. ventricosa*, *J. lianshanica*, *J. latiflora*, *J. patentiflora* and *J. leptostachya*. Within this clade, *Rungia* is the earliest diverging genus. Then, *J. gendarussa* and *J. ventricosa* form a sister subclade with the remaining four species (BS=100, PP=1.0). Clade III is sister to Clade II with strong support values (BS=100, PP=1.0). This clade contains *Peristrophe japonica* (Thunb.) Bremek., five sampled species of *Dicliptera* Juss. and nine sampled species of *Justicia*. Within Clade III, *J. adhatoda* and *J. betonica* are prior diverging species and form two separate subclades. Then, the African species *J. flava* forms a subclade with four other Asian *Justicia* species, including *J. quadrifaria*, *J. demissa*, *J. procumbens* and *J. mollissima* (BS=100, PP=1.0). However, *Justicia vagabunda* Benoist is distantly related to other members of *Justicia* but sister to *Dicliptera* and *Peristrophe* Nees with strong support values (BS=100, PP=1.0).

Discussion

Basic features and genomic variation of *Justicia* cp. genomes

The complete cp. genome often tracks back maternal line inheritance in contrast to the nuclear genome [88]. Therefore, due to its highly conserved structure, limited sequence length and countable genes, it is widely used in recent studies of genetic variation, genomic

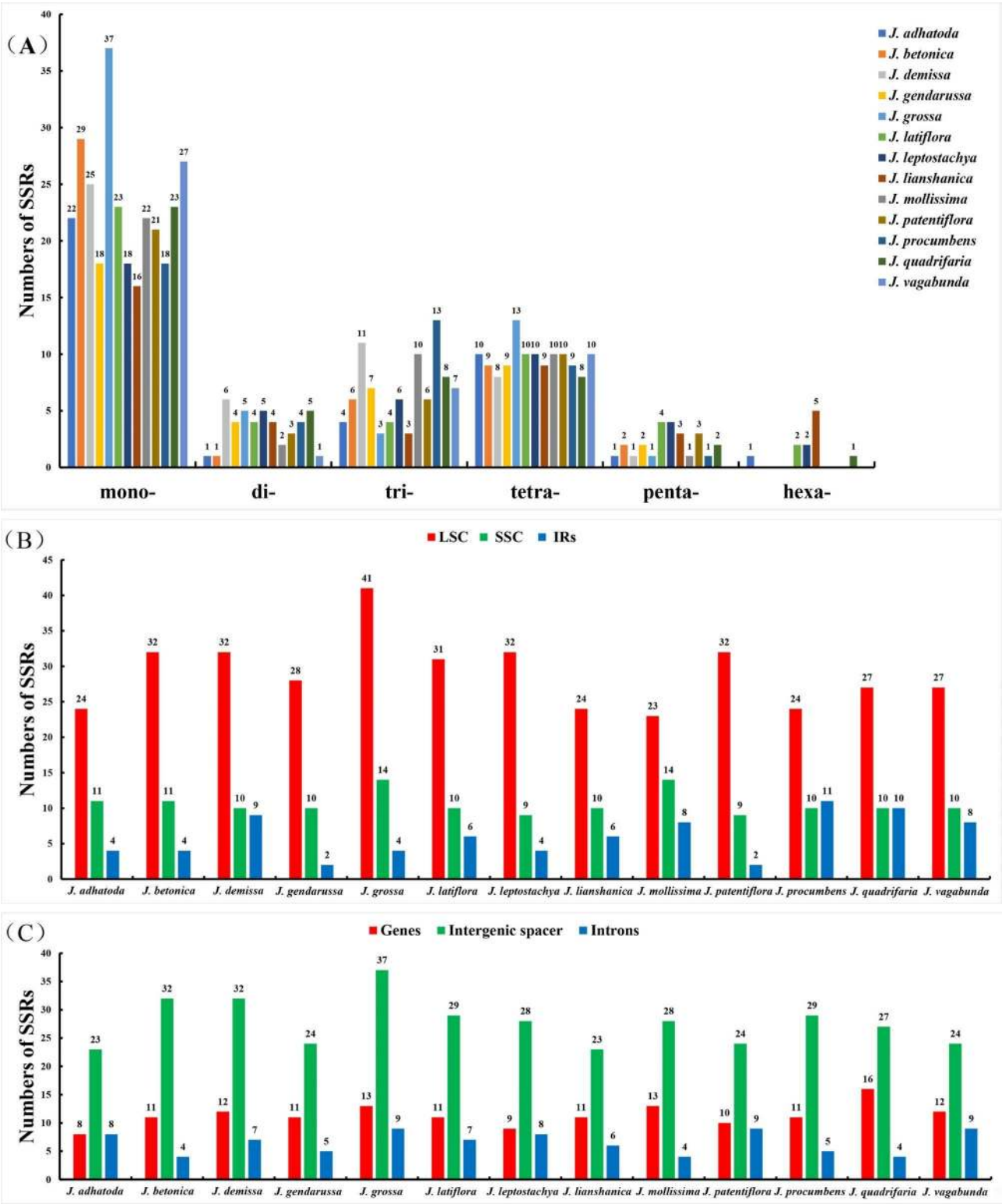


Fig. 6 Distribution of SSRs in the chloroplast genomes of 13 *Justicia* species. (A) Number of different SSRs types; (B) SSRs distribution in LSC, SSC and IR regions; (C) SSRs distribution between genes, intergenic spacer and introns

Table 3 The polymorphic SSRs among 13 *Justicia* chloroplast genomes

Type	<i>J. adhatoda/J. betonica/J. demissa/J. gendarussa/J. grossa/J. latiflora/J. leptostachya/J. lianshanica/J. mollissima/J. patentiflora/J. procumbens/J. quadrifaria/J. vagabunda</i>	Location	Regions
ATT	0/0/0/0/0/0/0/4/0/0/0/0	<i>trnH-GUG-psbA</i>	LSC
TTC	0/0/0/4/0/0/0/0/0/0/0/0	<i>trnH-GUG-psbA</i>	LSC
ATTAA	0/0/0/0/0/3/3/0/0/0/0/0	<i>trnK-UUU-rps16</i>	LSC
TATT	0/3/0/0/0/0/0/0/0/0/0/0	<i>trnK-UUU-rps16</i>	LSC
T	0/0/0/0/0/0/0/0/0/0/0/11	<i>trnK-UUU-rps16</i>	LSC
C	0/0/10/10/0/0/0/0/11/0/0/0/0	<i>trnK-UUU-rps16</i>	LSC
TTGAAA	3/0/0/0/0/0/0/0/0/0/0/0	<i>rps16 intron</i>	LSC
ATTG	3/0/3/3/0/0/0/3/0/3/3/3	<i>rps16 intron</i>	LSC
TTTC	0/0/0/0/0/0/0/0/0/0/0/3	<i>rps16 intron</i>	LSC
A	0/0/0/12/14/11/0/11/10/0/0/0/0	<i>rps16-trnQ-UUG</i>	LSC
AAT	0/0/0/0/0/0/0/0/0/0/5/0/0	<i>psbK-psbI</i>	LSC
AT	0/0/8/6/6/7/7/7/0/9/7/0/0	<i>psbK-psbI</i>	LSC
T	13/10/10/10/0/10/0/0/0/0/0/0	<i>psbK-psbI</i>	LSC
A	0/0/10/10/13/0/0/0/0/0/0/16/0	<i>psbI-trnS-GCU</i>	LSC
T	11/12/0/0/0/10/10/0/0/0/0/15/0	<i>trnR-UCU-atpA</i>	LSC
A	12/14/0/10/12/10/13/11/0/12/0/0/0	<i>atpF intron</i>	LSC
T	11/10/0/0/10/0/0/0/0/0/0/0/0	<i>atpF-atpH</i>	LSC
A	0/0/0/0/10/0/0/0/0/0/0/0/0	<i>atpF-atpH</i>	LSC
A	0/16/0/0/0/0/0/0/0/0/0/12/15	<i>atpH-atpI</i>	LSC
A	12/0/11/12/12/16/13/10/13/11/11/11	<i>rps2-rpoC2</i>	LSC
AATTCA	0/0/0/0/0/0/0/0/0/0/0/3/0	<i>rpoC2</i>	LSC
T	0/0/0/0/0/0/0/0/0/0/10/0/0/0	<i>rpoC2</i>	LSC
C	0/13/0/0/10/12/17/13/0/19/0/0/0	<i>rpoC1 intron</i>	LSC
ATT	0/0/0/0/0/0/0/0/0/0/0/4/0	<i>rpoB-trnC-GCA</i>	LSC
TATTAA	0/0/0/0/0/0/0/3/0/0/0/0/0	<i>trnC-GCA-petN</i>	LSC
A	0/0/0/0/0/10/0/0/0/12/0/0/0	<i>trnC-GCA-petN</i>	LSC
ATAG	0/3/0/3/3/0/0/0/0/0/0/0/0	<i>petN-psbM</i>	LSC
AT	0/0/6/0/0/0/0/0/0/0/6/0/0	<i>petN-psbM</i>	LSC
ATTT	3/0/0/0/0/0/0/0/0/0/0/0/3	<i>psbM-trnD-GUC</i>	LSC
TTTA	0/0/0/0/0/3/3/0/0/3/0/0/0	<i>psbM-trnD-GUC</i>	LSC
CAATA	3/3/0/0/3/3/3/0/0/3/0/3/0	<i>trnE-UCC-trnT-GGU</i>	LSC
T	10/17/0/0/10/0/0/0/0/0/0/0/10	<i>trnE-UCC-trnT-GGU</i>	LSC
T	11/0/10/10/0/11/10/0/0/10/13/11/0	<i>trnT-GGU-psbD</i>	LSC
ATTA	0/0/0/0/0/3/3/0/0/3/0/0/0	<i>trnT-GGU-psbD</i>	LSC
A	11/0/0/0/0/13/10/13/0/10/0/0/0	<i>psbZ-trnG-GCC</i>	LSC
T	0/0/0/0/11/0/0/0/0/0/0/0/0	<i>psbZ-trnG-GCC</i>	LSC
AT	0/0/0/0/0/0/0/0/8/0/0/0/0	<i>ycf3-trnS-GGA</i>	LSC
A	0/0/0/0/10/0/0/0/0/0/0/0/0	<i>ycf3-trnS-GGA</i>	LSC
TTA	0/4/0/0/0/0/0/0/0/0/0/0/0	<i>trnS-GGA-rps4</i>	LSC
TTTC	0/3/0/0/0/0/0/0/0/0/0/0/0	<i>rps4-trnT-UGU</i>	LSC
ATAG	0/0/0/4/0/0/0/0/0/0/0/0/0	<i>rps4-trnT-UGU</i>	LSC
AT	0/0/0/0/0/6/7/6/0/0/0/7/0	<i>rps4-trnT-UGU</i>	LSC
TTTC	0/0/0/0/0/3/3/0/3/0/0/0/0	<i>trnT-UGU-trnL-UAA</i>	LSC
AT	6/0/6/0/6/7/6/7/6/6/6/0/0	<i>trnT-UGU-trnL-UAA</i>	LSC
TA	0/0/0/6/0/0/0/0/0/0/8/0/7	<i>trnT-UGU-trnL-UAA</i>	LSC
GTTG	0/0/0/0/0/3/3/0/3/0/0/0/0	<i>trnF-GAA-ndhJ</i>	LSC
ATT	0/0/0/4/0/0/0/0/0/0/0/0/0	<i>trnF-GAA-ndhJ</i>	LSC
T	0/0/12/0/15/0/0/0/0/0/0/0/0	<i>trnF-GAA-ndhJ</i>	LSC
TTA	0/0/0/0/0/0/0/0/4/0/0/0/0	<i>ndhC-trnV-UAC</i>	LSC
T	14/10/10/10/17/10/10/0/0/14/0/10/10	<i>ndhC-trnV-UAC</i>	LSC
ATA	0/0/4/4/5/4/4/0/0/4/4/0/0	<i>atpB-rbcL</i>	LSC

Table 3 (continued)

Type	<i>J. adhatoda/J. betonica/J. demissa/J. gendarussa/ J. grossa/J. latiflora/J. leptostachya/J. lianshanica/ J. mollissima/J. patentiflora/J. procumbens/ J. quadrifaria/J. vagabunda</i>	Location	Regions
T	0/0/0/0/11/0/0/0/0/16/16/10	<i>atpB-rbcl</i>	LSC
A	0/0/0/0/12/0/0/11/0/0/0/0/0	<i>rbcl-accD</i>	LSC
ATTA	0/0/0/0/0/0/0/0/0/0/4/4/0	<i>accD-psaI</i>	LSC
TTAA	0/0/0/0/0/0/3/3/0/3/0/0/0	<i>accD-psaI</i>	LSC
T	0/12/11/11/0/10/0/0/12/0/12/12/11	<i>psaI-ycf4</i>	LSC
GAAA	0/0/0/0/3/0/0/0/0/0/0/0/0	<i>ycf4-cemA</i>	LSC
ATA	0/0/4/0/0/0/0/0/4/0/4/4/0	<i>psbE-petL</i>	LSC
A	0/0/0/0/0/0/0/0/0/0/0/0/11	<i>psbE-petL</i>	LSC
T	0/0/0/0/0/0/0/0/0/0/0/16/0	<i>psaI-rpl33</i>	LSC
T	0/0/0/12/14/19/10/12/0/12/0/0/0	<i>rpl20-rps12</i>	LSC
TTTC	0/0/0/3/3/3/3/0/3/0/0/0/0	<i>clpP intron</i>	LSC
AT	0/0/0/0/0/6/8/6/0/6/0/0/0	<i>clpP intron</i>	LSC
T	0/0/10/0/0/0/0/0/0/0/0/0/12	<i>clpP intron</i>	LSC
T	10/0/12/0/11/0/10/0/0/15/11/0/0	<i>petB intron</i>	LSC
TTTA	0/0/0/0/3/0/0/0/0/0/0/0/0	<i>petB-petD</i>	LSC
T	0/0/0/0/16/0/0/0/0/0/0/0/10	<i>rpoA</i>	LSC
AAAT	3/0/0/0/0/0/0/0/0/0/0/0/0	<i>rpl16 intron</i>	LSC
TTTC	0/3/0/0/0/0/0/0/0/0/0/0/0	<i>rpl16 intron</i>	LSC
AATA	0/0/0/3/0/0/0/0/0/0/0/0/0	<i>rpl16 intron</i>	LSC
TTA	0/0/0/0/0/0/0/0/0/0/0/0/4	<i>rpl16 intron</i>	LSC
T	0/10/0/0/12/0/0/0/0/0/0/0/10	<i>rpl16 intron</i>	LSC
AATAAG	0/0/0/0/0/0/0/3/0/0/0/0/0	<i>rps12-trnV-GAC</i>	IR
T	0/0/0/0/10/0/0/0/0/0/0/0/0	<i>rps12-trnV-GAC</i>	IR
TTTAA	0/0/0/3/0/3/3/3/0/3/0/0/0	<i>trnR-ACG-trnN-GUU</i>	IR
ATT	4/0/0/0/0/0/0/0/0/0/0/0/0	<i>trnR-ACG-trnN-GUU</i>	IR
T	13/18/15/0/10/0/0/0/16/0/15/11/11	<i>trnR-ACG-trnN-GUU</i>	IR
T	0/10/0/0/15/0/0/0/0/0/0/0/0	<i>ndhF</i>	SSC
T	0/0/0/0/11/0/0/0/0/0/0/0/0	<i>rpl32-trnL-UAG</i>	SSC
AATA	3/3/3/3/0/3/3/3/3/3/3/3/3	<i>ndhD</i>	SSC
A	10/10/0/0/10/10/11/0/10/0/0/12	<i>ndhD-psaC</i>	SSC
A	0/0/0/10/0/0/14/0/11/0/0/0/0	<i>ndhG</i>	SSC
AATC	3/3/3/0/3/0/0/0/3/0/3/0/3	<i>rps15-ycf1</i>	SSC
TTTG	0/0/0/0/0/0/0/0/3/0/0/3/0	<i>rps15-ycf1</i>	SSC
T	0/14/0/0/0/0/0/0/0/0/0/0/0	<i>rps15-ycf1</i>	SSC
AATT	3/0/3/3/3/3/3/3/3/3/3/3/3	<i>ycf1</i>	SSC
TTTC	0/0/0/0/3/0/0/0/0/0/0/0/0	<i>ycf1</i>	SSC
TTA	0/0/0/0/0/4/4/4/0/4/0/0/0	<i>ycf1</i>	SSC
TCT	8/5/6/4/10/6/5/6/6/6/5/6/6	<i>ycf1</i>	SSC
T	13/13/16/12/16/12/12/12/13/12/16/16/13	<i>ycf1</i>	SSC

evolution and phylogeny [59–65, 87–90]. In our study, the 13 *Justicia* cp. genomes exhibit a typical quadripartite structure, with two distinct single-copy regions separated by two inverted repeat (IR) regions (Fig. 1). All the cp. genomes are similar in genomic structure, gene composition and order (Tables 1 and 2; Fig. 3), which is consistent with other genera in Acanthaceae [91, 92]. Despite the overall conservation in structure, *Justicia* whole plastomes vary from 148,374 bp (*J. latiflora*) to 151,739 bp (*J. quadrifaria*) in size, falling within the

middle range (120–218 kb) of land plants [66]. Compared with those of previously reported genera in Acanthaceae, *Justicia* plastid genomes are generally smaller than *Barleria* L. (151,977–152,324 bp) [93, 94], *Echinacanthus* Nees (152,384–152,672 bp) [91], *Aphelandra knappiae* Wassh. (152,457 bp) [95], *P. haikangense* (152,849 bp) [96] and *Thunbergia erecta* Nees (152,202 bp) [97], but larger than *Strobilanthes* Blume (144,012–145,110 bp) [98–100]. Additionally, some *Justicia* plants also have similar genome sizes with its closely related genera

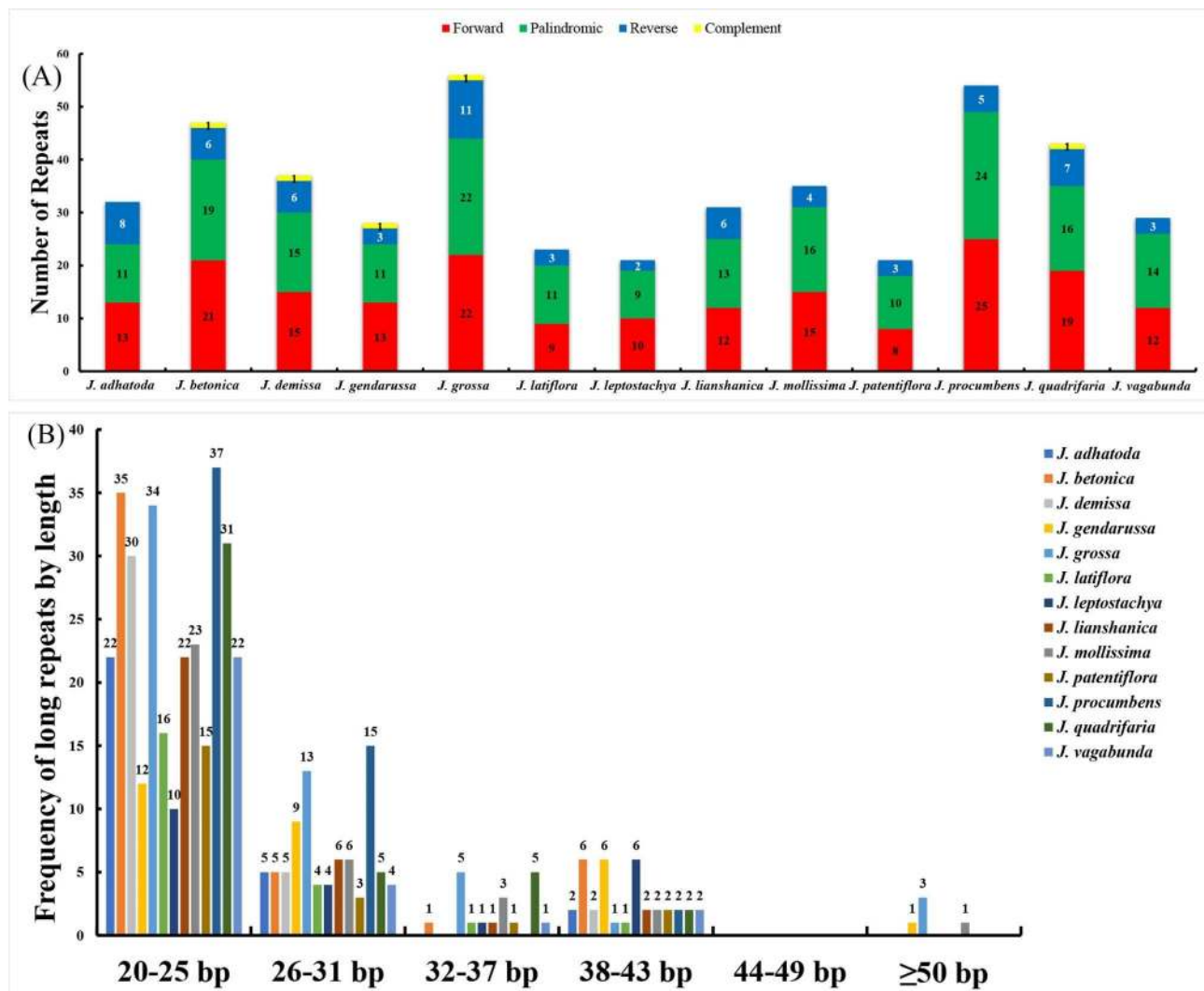


Fig. 7 Dispersed repeats of the 13 *Justicia* cp. genomes. (A) Number of four repeat types; (B) Frequency of long repeats by length

(Fig. 8). For example, the cp. genome size of *J. gendarussa* (149,735 bp) is close to that of *R. pectinata* (149,627 bp) [101], and the genome sizes of three species belonging to *Justicia* sect. *Rostellaria* (*J. demissa*, *J. mollissima* and *J. procumbens*) (~150 kb) are similar to that of *Dicliptera* [92]. It is indicated that the length of the cp. genome sequence is quite variable among different species within *Justicia*. Additionally, the LSC length extends quite large from 81,434 bp (*J. latiflora*) to 83,676 bp (*J. quadrifaria*), however, the IR length is less variable between 24,947 bp (*J. adhatoda*) to 25,549 bp (*J. procumbens*). The most considerable length of SSC region was recorded in *J. grossa* (17,507 bp), while the others are between 16,790 bp (*J. lianshanica*) to 17,218 bp (*J. gendarussa*) (Table 2). It is implied that greater sequence length divergence was observed in LSC and SSC, while fewer sequence differences were found in the two IR regions.

All the cp sequences consist of 114 unique genes, which is same as those in other genera of Acanthaceae, including *Aphelandra* R.Br., *Dicliptera* and *Rungia* [92, 95, 101], but fewer than those of *Barleria* (131) [94]. The codon usage bias of 20 amino acids among different species is similar (Fig. 2, Table S5), which is congruent with other angiosperms [65, 89]. It is revealed that most protein-coding genes are generally identical, but genes *accD*, *matK*, *ndhI*, *rpl22*, *rpl20*, *rpoA*, *rps16*, *rps18*, *ycf1* and *ycf2* of *J. grossa* are obviously different from those of other *Justicia* plants in length and base variation (Table S3), suggesting *J. grossa* is different from other *Justicia* plants in plastid genes. Regarding the initiation codon of *ndhD*, ACG is commonly used as an alternative to AUG in many land plants, but it can still be converted to a functional AUG initiator codon by RNA editing [102–104], probably depending on a molecular cofactor PPR protein CRR4 during transcription [105]. Additionally, GUG is

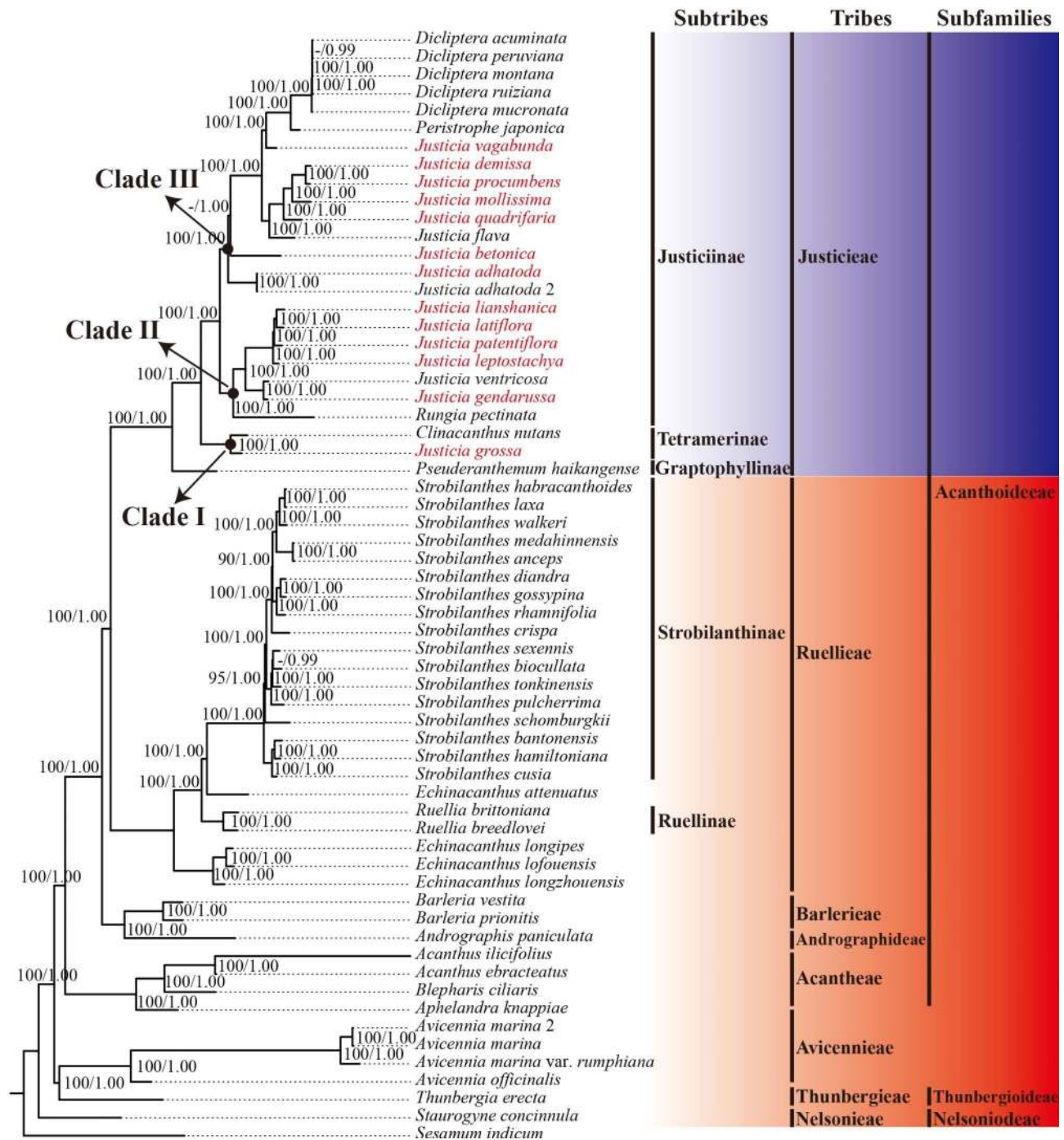


Fig. 8 Phylogenetic tree reconstruction for *Justicia* species and other genera of Acanthaceae based on whole chloroplast genome (WCG) by using Maximum likelihood (ML) and Bayesian inference (BI) methods. Different Asian sections of *Justicia* are marked with different colors. The tribe Justiciae are printed in blue while the other tribes of Acanthaceae are printed in red. Only bootstrap values (BS) $\geq 70\%$ and posterior probabilities (PP) ≥ 0.95 are indicated at each node

also reported as an initiation codon of *psbC* and *rps19* in other plants, such as *Thalictrum* L. (Ranunculaceae Juss.) [104], *Garcinia* L. (Clusiaceae Lindl.) [106], *Populus* L. (Salicaceae Mirb.) [107] and *Betula platyphylla* Suk. (Betulaceae Gray) [108], but these two genes cannot be edited back to AUG. However, recent studies suggested

that an important translation initiation signal, known as Shine-Dalgarno (SD) sequence, can interact with 3' end of the 16 S rRNA and facilitate translation initiation from the GUG [108, 109], which is responsible for expression of *psbC* and *rps19* in chloroplast. The *ycf15* gene is often duplicated in IR and annotated as the open reading frame

77 (ORF77), which belongs to protein families (Pfam) with accession PF10705 [110, 111]. In our study, *ycf15* gene is normally expressed in most *Justicia* plants, but pseudogenized in plastome of *J. adhatoda* due to its premature stop codons, which was doubted to be caused by gene degradation during RNA transcription [112, 113]. Meanwhile, this gene is different among other genera of Acanthaceae. For instance, it expresses under positive selection in *Dicliptera* [92], but acts as a pseudogene in *Echinacanthus* [91], or is even lost in *Barleria prionitis* L., *R. pectinata* and *S. cusia* [93, 98, 101]. Therefore, it is suggested that *ycf15* gene could be associated with plastome evolution of Acanthaceae, but its gene function remains to be further studied.

IR structure variation

Chloroplast genome structure is highly conserved across angiosperms [66, 67]. This is especially true for the IR regions, which is caused by low substitution rates and strict copy correction during repeat sequences replication [114]. The IR often ranges in size from 7 to 88 kb in angiosperms [115–117], with the extent of IR due largely to expansions and contractions at the SSC and LSC boundaries [114]. In our study, *Justicia* is different from other genera of Acanthaceae in three IR borders, i.e., LSC/IRa, LSC/IRb and SSC/IRb (Fig. 3). Compared with *Justicia*, significant IR expansion from IR to LSC was found in *E. lofouensis* and *R. pectinata*, and IR contraction with two directions of boundary shifts from IR to LSC and SSC was also detected in *A. knappiae*, *C. nutans* and *S. cusia*. The discrepancy of IR borders of plastomes within *Justicia*, however, only performs on SSC/IRa and SSC/IRb regions of *J. grossa* and other *Justicia* species. In detail, the lengths of *ndhF* and *ycf1* in IRs are much shorter than those of other *Justicia* species, while the lengths of these two genes located in SSC are much longer than others, which is thought to undergo IR contraction and cause an increase length of SSC region. Therefore, it is indicated that *J. grossa* is different from other *Justicia* species at the genomic structure level.

According to the statistics of cp. genome structure types of Lamiales [98], all the *Justicia* species belong to type II (a *rps19* pseudogene at the IR/LSC border). However, the plastome structure of *C. nutans*, with *ndhF* and *ycf1* boundary genes fully located in SSC, was not recorded before, thus it is firstly reported here. In our study, IR expansion and contraction events mainly contribute to genomic structure and sizes as well as gene composition variations among different genera of Acanthaceae, which is congruent with other plant lineages, including subfamily Commelinoideae (Commelinaceae Mirb.) [118], *Angelica* L. (Apiaceae Lindl.) [119], *Paphiopedilum* Pfitzer (Orchidaceae) [79] and *Balanites aegyptiaca* (L.) Delile (Zygophyllaceae R.Br.) [120]. It is

suggested that IR expansion and contraction events will provide useful references for further research on plastid genome rearrangement of angiosperms, with an emphasis on gene content and evolution of the IRs.

Potential molecular markers selection

Because the evolutionary rates of non-coding regions are faster than coding regions [76–78], LSC and SSC regions often exhibit higher sequence divergence than the IR regions in *Justicia* (Fig. 4), which is in accordance with other genera of Acanthaceae [91, 92, 94, 98]. Therefore, all of the mutational hotspots across the 13 *Justicia* complete cp. genomes were identified in single-copy regions (Fig. 5), of which six were intergenic spacer (*rps16-trnQ*, *rpoB-trnC*, *trnT-trnL*, *rps4-trnT*, *ccsA-ndhD* and *rps15-ycf1*), one was intron (*rpl16* intron) and one was protein coding gene (*ycf1*). The gene *ycf1* is a conservative homologous coding sequence with abundant variable sites [121–123]. Our phylogenetic topology based on *ycf1* gene is also generally identical with the cp. genome tree (Fig. S5). In addition, Dong et al. [121] also proposed that *ycf1* is the most promising plastid DNA barcode for land plants and plays an important role in genome evolution. Meanwhile, in some previous studies [123–125], *ycf1* gene has also been considered as an appreciated source to provide effective genetic information for phylogeny and species identification in breeding resources. Even one special concern for the use of *ycf1* as a barcode is the absence of *ycf1* in some taxa, such as Poaceae [121]. Therefore, this gene could be developed as a candidate DNA barcode for further phylogenetic reconstruction of *Justicia*. Compared with conserved coding regions, intergenic spacer and introns often show greater discrimination power at low taxonomic levels [126]. The *rpl16* intron and *trnT-trnL* have provided an effective molecular phylogeny in other plants, e.g., *Chusquea* Kunth (Poaceae) [127], *Echinochloa* P.Beauv. (Poaceae) and *Castanea* Mill. (Fagaceae Dumort.) [128, 129]. They were also proved to be a good resolution for phylogeny of Justicieae [9, 10]. Additionally, the five other non-coding regions have been proposed to be candidate DNA barcodes for phylogenetic research in other plant lineages, such as subfamily Dialioideae Azani et al. (Fabaceae Lindl.) [130], subfamily Zingiberoideae Hassk. (Zingiberaceae Martinov) [131], subfamily Allioideae Herb. (Amaryllidaceae J.St.-Hil.) [132], *Echinacanthus* (Acanthaceae) [91] and *Tetrastigma hemsleyanum* Diels & Gilg (Vitaceae Juss.) [133]. Therefore, it is believed that the eight mutational hotspots regions identified in our study could be potential molecular markers in *Justicia* phylogenetic studies. However, due to our results are only preliminary, more sampling and PCR amplification experiments for each primer of these barcodes should be carried out to test

whether they could be feasible in phylogenetic research of *Justicia* in the future.

Simple sequence repeats, SSRs, known as microsatellites, are short stretches of DNA containing repetitive sequences of 1–6 bp in length, have been the most frequently used genetic marker in species identification and population genetics [134], due to their co-dominant inheritance and high polymorphism [135]. SSRs are the same units with different repeat numbers located in the homologous regions and these regions are frequently used to identify variable species [92, 117, 136–138]. Therefore, cp. SSRs were identified in our study. As a result, repetitive sequences are significantly variable among different species (Figs. 6 and 7). Most cp. SSRs are located in intergenic spacer of LSC and SSC regions, with 61% in non-coding regions and only a small amount in protein-coding genes (25%) and introns (14%) (Table S6), which is consistent with other plants [91, 92, 139]. It is revealed that non-coding regions are more variable to screen valuable polymorphic SSRs [140–142]. Besides, cp. SSRs that are polymorphic within and among species can provide unique insights into species identification and their purities, particularly on those economically important plants [140]. Thus, a total of 91 polymorphic SSRs were identified here (Table 3). Due to the high similarity of universal DNA barcodes (*matK*, *rbcL*) among *Justicia* species (Table S4), our selected polymorphic SSRs can be effective genetic markers to identify these species. As the most common repeat unit, mononucleotide is mainly located in intergenic spacer and attributed to almost 90% A/T base richness (Table S6), which is in line with other plants [87, 141, 143]. Notably, apart from the highly variable hotspots region as mentioned above, *ycf1* is also detected as the most polymorphic gene with five different motifs (AATT, TTTC, TTA, TCT and T) in the 13 *Justicia* species (Table 3). Based on our results, it is believed that this gene is the most promising molecular marker for species identification in *Justicia* in the future. Importantly, based on our results of repetitive sequences analyses, it is indicated that *J. grossa* is quite different from other *Justicia* species owing to its richest SSRs and dispersed repeats among all the *Justicia* species, with an emphasis on the number of mononucleotide and dispersed repeats of over 50 bp (Figs. 6 and 7).

Potential reason for the low support values of ITS tree

The tree topology based on ITS sequence is generally similar with those based on whole chloroplast genome and 77 common cpCDS datasets, but the ITS tree has low support values whether based on ML or Bayes algorithm (BS<70, PP<0.95) (Fig. 8, Fig. S1–S4). In this case, the low support values are mainly attributed to the insufficiency of variable sites, though evolutionary rates of nuclear are faster than plastid. In our results,

the alignment of plastid genomes has much more variable sites in total (56,714) than ITS (457) (see Results part). Therefore, our ITS tree caused the sampling error, which means that in the process of substitution model selection, explaining too many parameters with too little data increases variance of estimable models [144, 145]. Anyway, phylogenetic analyses of too short sequences are more prone to result in sampling error than long sequences, simply because they contain less phylogenetic information [146].

Phylogenetic relationships of Asian *Justicia* plants

Recently, in the most comprehensive work of Graham [1], *Justicia* was divided into nine sections and seven subsections based on the combination of morphological characters of inflorescence, stamen, pollen, fruit and seed traits. In our study, we sampled 13 *Justicia* species from seven Asian *Justicia* sections. The phylogenetic results based on whole plastome, both 77 common protein-coding genes and ITS datasets (Fig. 8, Fig. S1–S4) suggest that *Justicia* s.l. is a polyphyletic group, which is supported in previous studies based on several molecular markers [8–12, 147].

In our results, Justicieae can be divided into three clades, i.e., Clade I, II and III (Fig. 8). Clade I contains two species, i.e., *J. grossa* and *C. nutans*, and might be assigned to subtribe Tetramerinae. *J. grossa* is isolated with other species in *Justicia* and forms the sister group with *C. nutans*. This result is also in accordance with previous phylogenetic studies using several molecular markers [9, 11]. *J. grossa* belongs to sect. *Grossa* B. Hansen. Sect. *Grossa* comprises three species from China, Vietnam, Laos, Thailand, Malaysia and Myanmar, and, morphologically, it is quite different from other *Justicia* plants in its bitheous anther having a solid, cusp-like spur at the base of each theca (Fig. S6), but other *Justicia* species only spurred on the lower theca [6, 8, 9, 11, 148]. Meanwhile, sect. *grossa* is also different from *Clinacanthus* in its bitheous anthers with both spurred thecae while *Clinacanthus* has muticous monotheous anthers [3, 6]. Therefore, *J. grossa* may be recognized as a new undescribed genus. However, the further phylogenetic research is necessary to determine the position of sect. *Grossa* since only one species was sampled in our study.

With the exception of *J. grossa*, all of the remaining Asian *Justicia* species may be assigned to subtribe Justiciinae and can be divided into two main clades, i.e., Clade II and III. Clade II contains *J. latiflora*, *J. lianshanica*, *J. leptostachya*, *J. patentiflora* and *J. gendarussa* together with *Rungia*. Clade III includes *J. adhatoda*, *J. betonica*, *J. demissa*, *J. mollissima*, *J. procumbens*, *J. quadrifaria* and *J. vagabunda* together with *Dicliptera*. In terms of morphology, those plants of Clade II have the fruits in which the placenta separated from the capsule wall but remain

attached at the apices causing them to rise up at dehiscence while the fruits not as such in Clade III [1, 9].

In Clade II, *J. gendarussa* is clustered with *J. ventricosa* and closely related to *Rungia* in terminal spike and elastic placenta when fruit dehiscence [3, 9]. Four species (*J. latiflora*, *J. lianshanica*, *J. leptostachya* and *J. patentiflora*) are clustered together in sharing the characters of elongated simple or rarely branched terminal spikes, narrow bracts subtending the small flowers or clusters of small flowers (Fig. 9), as well as 2-colporate pollen grains and rugulose seeds [26, 149, 150].

In Clade III, *J. adhatoda* and *J. betonica* were considered to be closely related by Graham [1], but differs from each other mainly in the flower number at each node of the spike and the bract shape [1, 30]. The former has the spikes with one flower per node and ovate-oblong bracts while the latter has the spikes with two flowers per node and white cordate bracts with green veins (Fig. 9). The next diverging species is *J. quadrifaria*, which is distributed in Asia and Africa and is characterized by the axillary cymose inflorescence, tiny subulate or triangular bracts and 5-partite calyx with equal segment (Fig. 9) [3, 5, 7, 29]. Next to diverge is the group including three species *J. demissa*, *J. mollissima* and *J. procumbens*, sharing the characters of short simple terminal spikes and 5-partite calyx with one extremely reduced segment and purplish red corolla (Fig. 9) [1, 3, 5]. In our analysis, *J. vagabunda* is the last diverging species and is sister to *Dicliptera*, but distantly related to other members of *Justicia* (Fig. 8). It differs from other sampled *Justicia* species in having axillary cymes and irregularly rounded-rugose tuberculate seeds. Besides, it is also easily distinguished from *Dicliptera* by its lower anther-theca spurred at base and the placenta not separate from the capsule wall while the anther and fruit not as such in the latter [3, 6, 151].

Interestingly, we discovered that the position and rachis internode of inflorescence of *Justicia* plants vary from terminal to axillary, spikes gradually shorten into cymes, seems to be a phenotype positively correlated with its evolutionary history. However, to fully resolve phylogenetic relationships of *Justicia*, more genetic resources and morphological evidence of *Justicieae* species from Africa, Australia and South America need to be combined with our Asian taxa for phylogenetic analyses in the future.

Conclusions

Our study sequenced 12 complete chloroplast genomes of Asian *Justicia* plants and combined with the previously published plastome of *J. leptostachya* for further comparative genomic analyses. The 13 *Justicia* cp. genomes are highly conserved in genome structures, organizations and gene contents. However, the gene *ycf15* was found to be a pseudogene in *J. adhatoda* while normally expressed in others. Four IR/SC junctions of plastomes

are generally identical within genus with the exception of *J. grossa*. Repetitive sequences are significantly variable at the interspecific level. A total of 91 polymorphic SSRs and the eight mutational hotspots were also identified. Among them, the gene *ycf1* is the most promising plastid DNA barcode for *Justicia* species identification and phylogenetic studies. Our phylogenetic results strongly supported that *Justicia* is polyphyletic and shed lights on the relationships among Asian *Justicia* plants for the first time. Interestingly, the evolutionary history of *Justicia* coincides with morphology of inflorescence position from terminal to axillary and spikes gradually shorten into cymes.

Additionally, it is noteworthy that *J. grossa* is different from other *Justicia* species in the following three aspects: (i) it is richest in SSRs and dispersed repeats compared with other *Justicia* species; (ii) its SC/IR boundary genes *ndhF* and *ycf1* located in IRs are much shorter than other *Justicia* species, while *Ψycf1* located in SSC is much longer than others; (iii) its systematic position is in subtribe Tetramerinae, which is distantly related to other members of *Justicia*. Therefore, combined with its morphology of bitheous anthers with both spurred thecae, *J. grossa* should be defined as a new genus. Our study may not only improve the understanding of plastomes of *Justicia* plants, but also provide more genetic information for further researches on the evolutionary history of *Justicia* in the future.

Materials and methods

Sampling, DNA extraction and sequencing

A total of 12 plants from seven Asian sections of *Justicia* were sampled in our study, followed by classification system of Graham [1] and Hansen [148], including *J. adhatoda*, *J. betonica*, *J. demissa*, *J. gendarussa*, *J. grossa*, *J. latiflora*, *J. lianshanica*, *J. mollissima*, *J. quadrifaria*, *J. patentiflora*, *J. procumbens* and *J. vagabunda*. Fresh and healthy leaves of these 12 *Justicia* plants were collected in the field, with sampling information listed in Table S1. Leaf samples were immediately dried with silica gel for further DNA extraction. All the voucher specimens were deposited in the Herbarium of South China Botanical Garden, Chinese Academy of Sciences (IBSC).

Total genomic DNA was isolated using the modified CTAB method [152]. The extracted genomic DNA was sent to the Beijing Genomics Institute (BGI) for qualification control by fluorometer (Qubit Fluorometer, Invitrogen). DNA samples of concentration up to standard ($\geq 1 \mu\text{g}$) were randomly sheared into fragments by Covaris M220 (Covaris, Woburn, MA). Insert size of 270 bp DNA fragments were enriched by PCR, and the paired-end (2×150 bp) libraries were constructed on the Illumina HiSeq 4000 platform. Finally, about 2 Gb genome skimming data were generated.

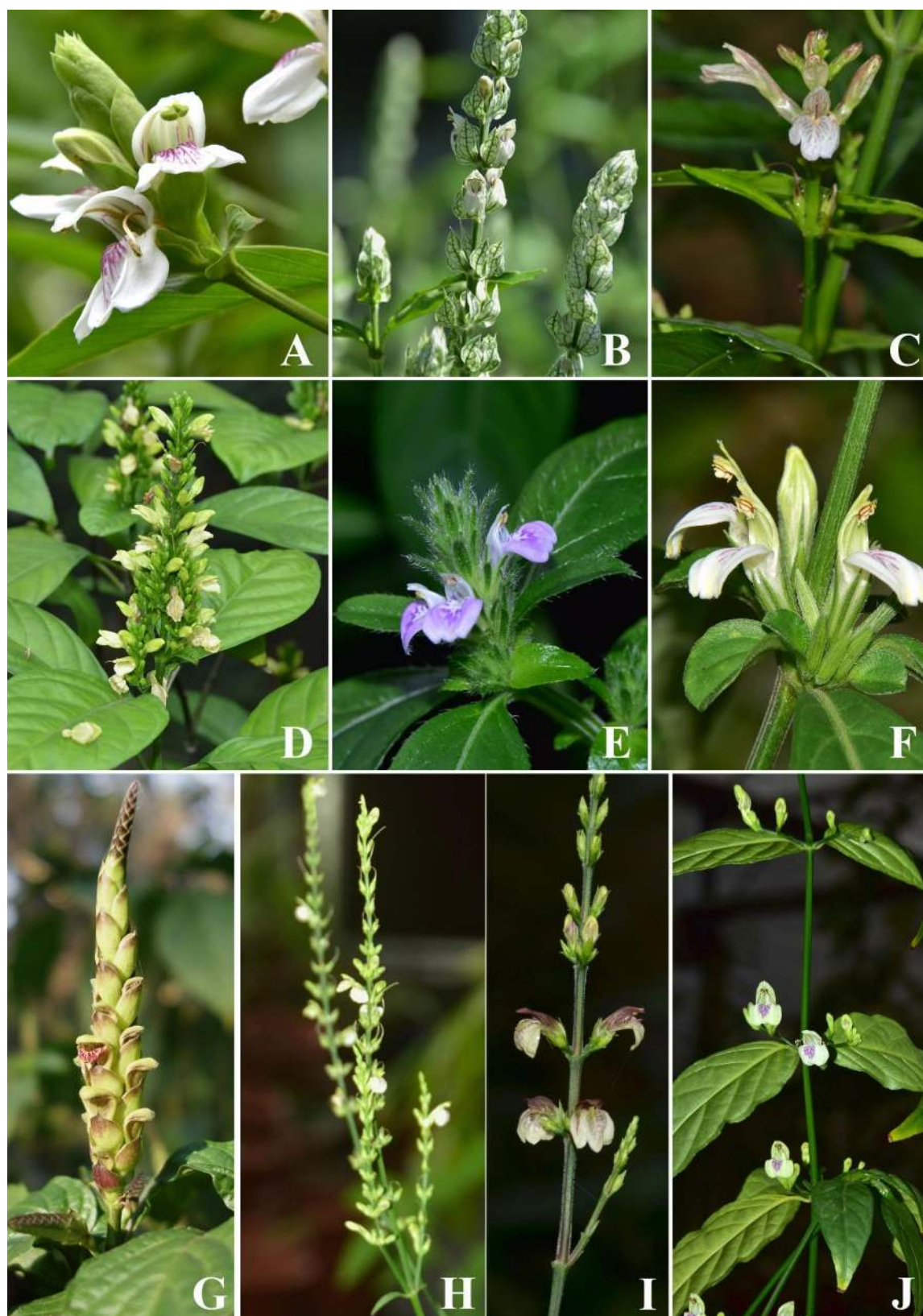


Fig. 9 Morphological characters of ten representative Asian *Justicia* plants. (A). *J. adhatoda* L.; (B). *J. betonica* L.; (C). *J. gendarussa* N. J. Burman; (D). *J. grossa* C. B. Clarke; (E). *J. procumbens* L.; (F). *J. quadrifaria* (Nees) T. Anderson; (G). *J. latiflora* Hemsl.; (H). *J. leptostachya* Hemsl.; (I). *J. lianshanica* (H. S. Lo) H. S. Lo; (J). *J. vagabunda* Benoist

Assembly and annotation of whole plastome and nrDNA

To improve assembly accuracy and efficiency, Trimmomatic v 0.39 was first employed to filter out unpaired and low-depth reads from clean data using default parameters [153]. The filtered clean reads were utilized to *de novo* assemble complete cp. genomes using GetOrganelle v 1.6.2 pipeline [154]. To obtain complete cp. genomes and nrDNA sequence, six k-mer values, including 21, 45, 65, 85, 105, 125, were set for plastid contigs connection. Subsequently, the filtered plastid reads were transferred to Bandage [155] software for visualization processing. Two opposite plastid sequences exported from Bandage were aligned with the reference sequence *Andropogon paniculata* (GenBank accession no. KF150644), and one that matched the reference was screened on the annotation of PGA software [156] and the Annotation of Organellar Genomes (GeSeq) [157]. The final annotations of plastomes and nrDNA sequences were manually corrected in Geneious Prime v 9.1.4 [158]. The whole cp. genomes and nrDNA maps were drawn by using OGDRAW v 1.3.1 (<https://chlorobox.mpimp-golm.mpg.de/>) with default settings [159].

Genome divergences comparison and codon usage analyses

The complete cp. genome of *J. leptostachya* was combined with newly assembled 12 cp. genomes in our study for further comparative genomic analyses. Whole plastomes of the 13 *Justicia* species and seven species of other genera in Acanthaceae were combined to visualize IR expansion and contraction by using IRscope online software (<https://irscope.shinyapps.io/irapp/>) [160]. Besides, the 13 *Justicia* plastomes were aligned and globally viewed using the online mVISTA program [161] (<https://genome.lbl.gov/vista/index.shtml>) in Shuffle-LAGAN mode [162], with the annotation of *J. leptostachya* as the reference. To evaluate nucleotide diversity (Pi), MAFFT v 7.450 [163] was operated to align the 13 *Justicia* cp. genomes. Then, Pi value was implemented based on a sliding window by Dnasp v 5.0 [164], with step size of 200 bp and window length of 800 bp. Relative synonymous codon usage (RSCU) in all the protein-coding sequences of 13 *Justicia* plants were calculated using CodonW v 1.4.2 software with default parameters [165].

Repetitive sequences analyses

Dispersed repeats among the 13 *Justicia* cp. genomes were identified with four directions (forward, reverse, palindromic, and complement) using the online REPuter program (<https://bibiserv.cebitec.uni-bielefeld.de/reputer>) [166], with the maximum computed repeats number of 100 and the minimal repeat size of 20 bp. The program MISA [167] was employed to obtain multiple short tandem repeats, including mononucleotide

(mono-), dinucleotide (di-), trinucleotide (tri-), tetranucleotide (tetra-), pentanucleotide (penta-), and hexanucleotide (hexa-) SSRs, with corresponding minimum repeat units set as 10, 6, 3, 3, 3, 3. Tandem repeats were also identified using Tandem Repeats Finder v 4.09 [168].

Phylogenetic analysis

Three datasets containing whole chloroplast genome (WCG), plastid protein-coding genes (PCG) and internal transcribed spacer (ITS) were designed for phylogenetic analysis based on two different algorithms including Maximum Likelihood (ML) and Bayesian Inference (BI). For WCG tree, a total of 62 samples were utilized for phylogenetic tree reconstruction, comprising 12 newly sequenced *Justicia* cp. genomes in our study, three previously published *Justicia* cp. genomes and 46 cp. genomes of other genera belonging to Acanthaceae from GenBank. *Sesamum indicum* L. (JN637766) was selected as the outgroup species because it belongs to the family Pedaliaceae R.Br., which is most closely related to Acanthaceae based on APG IV (<https://www.mobot.org/MOBOT/research/APweb/>). For PCG tree, with the exclusion of *psbA*, *rpl2* and *ycf15* gene due to lacking in some genera, a total of 77 common protein-coding genes were extracted from whole plastomes by using a python script 'get_annotated_regions_from_gb.py' (<https://github.com/Kinggerm/PersonalUtilities/>). Gblocks v 0.91b [169] was further employed to trim each gene matrix. The parameters are set as allowing up to half of the samples to have missing data and at least 87 minimum sequence length per gene matrix. For ITS tree, a total of 63 samples were utilized for phylogenetic inference, including 13 *Justicia* ITS sequences extracted from our nrDNA data by Geneious Prime and 50 previously published ITS sequences of Acanthaceae from GenBank. *Strobilanthes cusia* (Nees) Kuntze was set as the outgroup for the ITS tree. All the GenBank accession numbers of cp. genomes and ITS sequences used for our phylogenetic analyses were listed in Table S2.

Then, the three datasets were aligned by using MAFFT and the test for nucleotide substitution saturations was implemented in DAMBE v 7.2.133 referring to Xia's method [170], with a significance threshold of $I_{ss} < I_{ss.c}$ and $p\text{-value} < 0.05$. ML analyses were conducted by RAxML v 8.0.0 [171], with the best-fit parameter settings as rapid bootstrap algorithm and GTRGAMMAI model recommended by jModelTest v 2.1.6 [172]. The number of 12,345 was specified as the random seed of parsimony tree inference with 1000 replicates performed. BI analyses were operated by using MrBayes v3.2.2 [173], with the best-fit model selected as SYM+G inferred from MrModeltest v 2.3 [174]. Rates of variations across sites were trimmed as gamma. For each analysis, two simultaneous runs of four Monte Carlo Markov Chains (three

heated and one cold) were run for six million generations with a random tree as the starting point and saving trees every 1000 generations. After rejecting the first 25% burn-in samples, the optimized topology with posterior probabilities (PP) > 0.95 was generated. Finally, the phylogenetic results were visualized with FigTree v 1.4.3 (<http://tree.bio.ed.ac.uk/software/figtree/>).

Abbreviations

BI	Bayesian inference
BS	Bootstrap
CDS	Coding regions
CNS	Non-coding regions
cp	Chloroplast
HIV	Human immunodeficiency virus
IR	Inverted repeat
ITS	Internal transcribed spacer
LSC	Large single-copy region
ML	Maximum likelihood
nrDNA	Nuclear ribosome DNA
NW	New Word
ORF77	Open reading frame 77
OW	Old Word
PCG	Plastid protein-coding genes
Pfam	Protein families
Pi	Nucleotide diversity
PP	Posterior probability
RSCU	Relative synonymous codon usage values
SD	Shine-Dalgarno
SSC	Small single-copy region
SSRs	Simple sequence repeats
UTR	Untranslated regions
WCG	Whole chloroplast genome

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12870-023-04532-0>.

Additional file 1: Table S1. Collection and assembly information of 12 *Justicia* species

Additional file 2: Table S2. All the GenBank accession numbers used for phylogenetic analyses utilized in our study

Additional file 3: Table S3. Gene sizes of all the protein-coding genes of 13 *Justicia* chloroplast genomes

Additional file 4: Table S4. Genes with introns in the 13 *Justicia* chloroplast genomes, including the exon and intron lengths

Additional file 5: Table S5. Codon usage bias of 20 amino acids within 13 *Justicia* chloroplast genomes

Additional file 6: Table S6. Quantity statistics of SSRs of the 13 *Justicia* chloroplast genomes

Additional file 7: Figure S1. ML phylogram for 62 taxa of Acanthaceae based on 77 common protein-coding genes

Additional file 8: Figure S2. BI phylogram for 62 taxa of Acanthaceae based on 77 common protein-coding genes

Additional file 9: Figure S3. ML phylogram for 63 taxa of Acanthaceae based on ITS sequence

Additional file 10: Figure S4. BI phylogram for 63 taxa of Acanthaceae based on ITS sequence

Additional file 11: Figure S5. Phylogenetic reconstruction for *Justicia* species and other genera of Acanthaceae based on *ycf1* gene

Additional file 12: Figure S6. Morphology of inflorescence and anther of

Justicia grossa

Acknowledgements

We would like to thank Dr. Sunan Huang for her help in DNA extraction and data analyses. We thank Mrs. Caixia Peng for providing the images. And we are also very grateful to Dr. Sean W. Graham (Department of Botany, University of British Columbia) for his valuable comments during revision.

Authors' contributions

X.C. and Y.F.D. design the experiments, revised the manuscript and made funding acquisition; Z.Y.N. performed the experiments, analyzed the data and wrote the manuscript. Z.L.L. and Y.T. conducted an investigation and made funding acquisition. All the authors read and approved the manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (Grants No. 31970208, 31900182 and 31700166), the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDA13020500), and Southeast Asia Biodiversity Research Institute, the Chinese Academy of Sciences (Y4ZK111B01).

Data Availability

All the voucher specimens were deposited in the Herbarium of South China Botanical Garden, Chinese Academy of Sciences (IBSC), and their sampling information is listed in Table S1. All the newly sequenced 12 cp. genomes in this study are available in National Center for Biotechnology Information (NCBI) (<https://www.ncbi.nlm.nih.gov>), with accession numbers: MN848243–MN848252 and MN885664–MN885665 (Table 1). All the newly sequenced 13 nrDNAs in this study are available in NCBI with accession numbers: OQ785888–OQ785900. All the GenBank accession numbers of previously published sequences used for phylogenetic analyses in our study can be found in Table S2.

Declarations

Ethics approval and consent to participate

No specific permits were required for the collection of specimens for this study. The field work, collection of plant materials and molecular experiments were carried out in compliance with the relevant laws of China. All plant materials were identified by Yunfei Deng.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, China

²University of Chinese Academy of Sciences, Beijing 100049, China

³School of Biology and Agriculture, Shaoguan University, Shaoguan, Guangdong 512005, China

⁴School of Chinese Materia Medica Medical, Guangzhou University of Chinese Medicine, Guangzhou 510006, China

⁵Co-Innovation Center for Sustainable Forestry in Southern China, College of Life Sciences, Nanjing Forestry University, Nanjing, Jiangsu 210037, China

Received: 30 April 2023 / Accepted: 17 October 2023

Published online: 15 November 2023

References

- Graham VAW. Delimitation and infra-generic classification of *Justicia* (Acanthaceae). Kew Bull. 1988;43(4):551–624. <https://doi.org/10.2307/4129957>

2. Wasshausen DC. New species and new combinations of *Justicia* (Acanthaceae) from the Venezuelan Guayana. Novon. 1992;2(1):62–80. <https://doi.org/10.2307/3391611>
3. Hu JQ, Deng YF, Daniel TF. In: Wu ZY, Raven PH, Hong DY, editors. *Justicia* in Flora of China. St. Louis: Science Press, Beijing and Missouri Botanical Garden Press; 2011. pp. 380–430.
4. Mabberley DJ. Mabberley's Plant-Book: a portable dictionary of plants, their classification and uses. 4th ed. Cambridge: Cambridge University Press; 2017. pp. 1–1120.
5. Deng YF. Acanthaceae. In: Hong DY, editor. Flora of Pan-Himalaya 46. Beijing: Science Press; 2020. pp. 39–443.
6. Deng YF, Gao CM. Acanthaceae. In: Li DZ, editor. The families and Genera of Chinese Vascular Plants 3. Beijing: Science Press; 2020. pp. 1968–93.
7. Tong Y, Deng YF. Taxonomic revision of *Justicia* sect. *Harnieria* (Acanthaceae) from Philippines. Phytotaxa. 2021;483:190–210. <https://doi.org/10.11646/phytotaxa.483.3.1>
8. Manzitto-Tripp EA, Darbyshire I, Daniel TF, Kiel CA, McDade LA. Revised classification of Acanthaceae and worldwide dichotomous keys. Taxon. 2022;71(1):103–53. <https://doi.org/10.1002/tax.12600>
9. Kiel CA, Daniel TF, Darbyshire I, McDade LA. Unraveling relationships in the morphologically diverse and taxonomically challenging justicioid lineage (Acanthaceae: Justicieae). Taxon. 2017;66(3):645–74. <https://doi.org/10.12705/663.8>
10. Kiel CA, Daniel TF, McDade LA. Phylogenetics of New World 'justicioids' (Justicieae: Acanthaceae): major lineages, morphological patterns, and widespread incongruence with classification. Syst Bot. 2018;43(2):459–84. <https://doi.org/10.1600/036364418x697201>
11. Deng YF, Gao CM, Xia NH, Peng H, Wuacanthus. (Acanthaceae), a new Chinese endemic genus segregated from *Justicia* (Acanthaceae). Plant Divers. 2016;38(6):312–21. <https://doi.org/10.1016/j.pld.2016.11.010>
12. McDade LA, Daniel TF, Masta SE, Riley KM. Phylogenetic relationships within the tribe Justicieae (Acanthaceae): evidence from molecular sequences, morphology, and cytology. Ann Mo Bot Gard. 2000;87(4):435–58. <https://doi.org/10.2307/2666140>
13. Turner IM. A catalogue of the vascular plants of Malaya. Gard Bull Singapore. 1995;47(1):1–346.
14. Cramer LH. Acanthaceae. In: Dassanayake MD, Clayton WD, editors. A revised handbook to the flora of Ceylon. Volume 12. Rotterdam: A. Balkema; 1998. pp. 1–140.
15. Ho PH. Cáyco Vietnam: an illustrated flora of Vietnam. Volume 3. Ho Chi Minh City: Nha Xuat Ban Tre; 2000.
16. Kress WJ, Filippis RA, Farr E, Kyi DYY. A checklist of the trees, shrubs, herbs, and climbers of Myanmar. Contr U S Nat Herb. 2003;45:1–590.
17. Newman M, Ketphanh S, Svengsuksa B, Thomas P, Sengdala K, Lamxay V, et al. A checklist of the vascular plants of Lao PDR. Edinburgh: Royal Botanic Garden Edinburgh Press; 2007. pp. 1–394.
18. Karthikeyan S, Sanjappa M, Moorthy S. Flowering plants of India, vol. 1. Dicotyledons (Acanthaceae–Avicenniaceae). Kolkata: Botanical Survey of India; 2009. pp. 1–366.
19. Staples GW, Herbst DR. A tropical garden flora. Hawaii: Bishop Museum Press; 2005. pp. 1–908.
20. Liu B, Yang YF, Liu HB, Xie ZT, Li Q, Deng M, et al. Screening for cytotoxic chemical constituents from *Justicia procumbens* by HPLC–DAD–ESI–MS and NMR. BMC Chem. 2018;12:6. <https://doi.org/10.1186/s13065-018-0371-z>
21. Peng CX, Tang WX, He KH. Ex Situ Flora of China, Acanthaceae. Beijing: China Forestry Publishing House; 2021. pp. 1–504. (In Chinese).
22. Bharath KR, Suryanarayana B. Ethnomedicinal recipes for respiratory and bronchial Diseases from tribals of Sriharikota island, Andhra Pradesh. Ethnobotanical Leaflets. 2008;12:896–911.
23. Zhang HJ, Rumschlag–Booms E, Guan YF, Wang DY, Liu KL, Li WF, et al. Correction to potent inhibitor of drug-resistant HIV–1 strains identified from the medicinal plant *Justicia gendarussa*. J Nat Prod. 2017;80(8):2390. <https://doi.org/10.1021/acs.jnatprod.7b00562>
24. Yaradua SS, Alzahrani DA, Albokhary EJ, Abba A, Bello A. Complete chloroplast genome sequence of *Justicia Flava*: genome comparative analysis and phylogenetic relationships among Acanthaceae. Biomed Res Int. 2019;4370258. <https://doi.org/10.1155/2019/4370258>
25. Nees VECG. Acanthaceae. de Candolle AP, editor. Prodrum Systematis Naturalis Regni Vegetabilis, Vol. 2. Paris: Treuttel & Würtz; 1847. pp. 46–519.
26. Bremekamp CEB. Notes on the Acanthaceae of Java. Verhandelingen Der Koninklijke Nederlandsche Akademie Van Wetenschappen Afdeling Natuurkunde Sectie. 1948;vol 2(452):1–78.
27. Anderson T. An enumeration of the Indian species of Acanthaceae. Bot J Linn Soc. 1867;9(40):455–526. <https://doi.org/10.1111/J.1095-8339.1867.TB01308.X>
28. Benthham G. Acanthaceae. In: Benthham G, Hooker JD, editors. Genera Plantarum. Volume 2. Reeve: London, England; 1876. pp. 1060–122.
29. Clarke CB. Acanthaceae. In: Hooker JD, editor. Flora of British India. Reeve: London, England; 1885. p. 387–558.
30. Lindau G. Acanthaceae. In: Engler A, editor. Die Natiirlichen Pflanzenfamilien tomus IV(3b). Britannica: Leipzig, Germany; 1895. pp. 274–354.
31. Bremekamp CEB. Delimitation and subdivision of the Acanthaceae. Bull Bot Surv India. 1965a;7:21–30.
32. Bremekamp CEB. Studies in the Flora of Thailand no. 32: Scrophulariaceae–Nelsonieae, Acanthaceae and Thunbergiaceae. Dansk Botanisk Arkiv. 1965b;23:195–224.
33. Hu CC, Tsui HP. Acanthaceae. Flora Reipublicae Popularis Sinicae, Tomus 70. Beijing: Science Press; 2002. pp. 1–350. (In Chinese).
34. Stearn WT. Taxonomic and nomenclatural notes on Jamaican gamopetalous plants. J Arnold Arbor Harv Univ. 1971;52(4):614–48. <https://doi.org/10.5962/bhl.part.9118>
35. Wasshausen DC. Acanthaceae. Flora of the Guianas: Series A: Phanerogams, fasc. 23. London: Royal Botanic Gardens, Kew; 2006. 1–141.
36. Wasshausen DC. Flora of Ecuador, volume 89, part 179: Acanthaceae. Gothenburg: Department of Biological and Environmental Sciences, University of Gothenburg; 2013. pp. 1–328.
37. Vollesen K. Flora of Tropical East Africa: Acanthaceae, Part 1. London: Royal Botanic Gardens, Kew; 2008. pp. 1–288.
38. Vollesen K, Flora Zambesiaca. Acanthaceae (part 1). Volume 8. London: Royal Botanic Gardens, Kew; 2013. 5.
39. Wasshausen DC, Wood JRI. Acanthaceae of Bolivia. Contr U S Natl Herb. 2004;49:1–52.
40. Bremekamp CEB. Remarks on the position of some Australian Acanthaceae. Acta Bot Neerl. 1962;11:195–200. <https://doi.org/10.1111/j.1438-8677.1962.tb00084.x>
41. Bremekamp CEB. Studies in the Flora of Thailand no. 35: Scrophulariaceae–Nelsonieae, Acanthaceae and Thunbergiaceae. Dansk Bot Arkiv. 1966;23:273–9.
42. Maguire B. Plant explorations in Guiana in 1944, chiefly to the Tafelberg and the Kaieteur Plateau–VI. Bull Torrey Bot Club. 1948;75(6):633–71. <https://doi.org/10.2307/2482346>
43. Santapau H. The Acanthaceae of Bombay. Univ Bombay Bot Mem. 1951;2:1–128.
44. Bakhuizen VDBRC. Flora of Java. Volume 2. Netherlands: N.V.P. Noordhoff, Groningen; 1965. pp. 1–713.
45. Hsieh CF, Huang TC. The Acanthaceous plants of Taiwan. Taiwanica. 1974;19:19–57. <https://doi.org/10.6165/TAI.1974.19.19>
46. Wu ZY. A checklist of seed plants in Yunnan, Tomus II. Kunming: Yunnan People's Publishing House; 1985. pp. 1071–2259. (In Chinese).
47. Wu ZY. The areal-types of Chinese genera of seed plants. Acta Bot Yunnanica. 1991;13(Suppl IV):1–139. (In Chinese).
48. Wu ZY, Lu AM, Tang YC, Chen ZD, Li DZ. The families and genera of angiosperms in China: a comprehensive analysis. Beijing: Science Press; 2003. pp. 1–1206. (In Chinese).
49. Hu JQ, Fu XP. Acanthaceae. In: Fu LK, editor. Higher plants of China. Volume 10. Qingdao: Qingdao Publishing House; 2005. pp. 329–416. (In Chinese).
50. Daniel TF. *Justicia Masiaca* (Acanthaceae), a new species from northwestern Mexico. Brittonia. 1995;47(4):408–13. <https://doi.org/10.2307/2807570>
51. Immelman KL. In: Baden K, Balkwill FM, Norris G, Immelman KL, Manning JC, Munday J, editors. *Justicia* in Flora of Southern Africa. South Africa: National Botanical Institute Press; 1995. pp. 18–46.
52. Ezcurra C. The genus *Justicia* (Acanthaceae) in the southern region of South America. Ann Mo Bot Gard. 2002;89(2):225–80. <https://doi.org/10.2307/3298565>
53. Darbyshire I, Vollesen K, Ensermu K. In: Beentje HJ, editor. *Justicia* in Flora of Tropical East Africa. London: Royal Botanic Gardens, Kew; 2010. pp. 495–601.
54. Wasshausen DC. *Justicia* L. In: Persson C, Stahl B, editors. Flora of Ecuador No.89. Gothenburg University Press; 2013. pp. 117–88.
55. Timberlake JR, Diniz MA, Simpson DP. *Justicia* L. in Flora Zambesiaca. In: Timberlake JR, Martins ES, editors. Chicago University Press. 2015;8(6):162–224.
56. Begum. Acanthaceae. In: Sarma, editor. Encyclopedia of flora and fauna of Bangladesh. Volume 6. Dhaka: Asiatic Society of Bangladesh; 2008. pp. 1–75.
57. Malik KA, Ghafoor A. Flora of Pakistan, No. 188, Acanthaceae. Islamabad: National Herbarium, Pakistan Agriculture Research Council; 1988.

58. Cramer LH. Acanthaceae. In: Dassanayake, M. D & Clayton, W. D, editors, A revised handbook to the Flora of Ceylon 12. A. A. Balkema, Rotterdam; 1998. p. 1–140.
59. Zhai W, Duan XS, Zhang R, Guo CC, Li L, Xu GX, et al. Mol Phylogenet Evol. 2019;135:12–21. <https://doi.org/10.1016/j.ympev.2019.02.024>. Chloroplast genomic data provide new and robust insights into the phylogeny and evolution of the Ranunculaceae.
60. Li EZ, Liu KJ, Deng RY, Gao YW, Liu XY, Dong WP, et al. Insights into the phylogeny and chloroplast genome evolution of *Eriocaulon* (Eriocaulaceae). BMC Plant Biol. 2023;23:32. <https://doi.org/10.1186/s12870-023-04034-z>
61. Gong L, Ding XX, Guan W, Zhang DC, Zhang J, Bai JQ, et al. Comparative chloroplast genome analyses of *Amomum*: insights into evolutionary history and species identification. BMC Plant Biol. 2022;22:520. <https://doi.org/10.1186/s12870-022-03898-x>
62. Zhou MY, Liu JX, Ma PF, Yang JB, Li DZ. Plastid phylogenomics shed light on intergeneric relationships and spatiotemporal evolutionary history of Melocanninae (Poaceae: Bambusoideae). J Syst Evol. 2022;60(3):640–52. <https://doi.org/10.1111/jse.12843>
63. Klimpert NJ, Mayer JLS, Sarzi DS, Prosdociimi F, Pinheiro F, Graham SW. Phylogenomics and plastome evolution of a Brazilian mycoheterotrophic orchid, *Pogoniopsis Schenkii*. Am J Bot. 2022;109(12):2030–50. <https://doi.org/10.1002/ajb2.16084>
64. Abdalla HAM, Wanga VO, Mkala EM, Amenu SG, Amar MH, Chen L, et al. Comparative genomics analysis of endangered wild Egyptian *Moringa peregrina* (Forssk.) Fiori plastome, with implications for the evolution of Brassicales order. Front Genet. 2023;14:1131644. <https://doi.org/10.3389/fgene.2023.1131644>
65. Tang CQ, Chen X, Deng YF, Geng LY, Ma JH, Wei XY. Complete chloroplast genomes of *Sorbus Sensus Stricto* (Rosaceae): comparative analyses and phylogenetic relationships. BMC Plant Biol. 2022;22:495. <https://doi.org/10.1186/s12870-022-03858-5>
66. Daniell H, Lin CS, Yu M, Chang WJ. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol. 2016;17(1):134. <https://doi.org/10.1186/s13059-016-1004-2>
67. Xu S, Teng K, Zhang H, Gao K, Wu J, Duan L, et al. Chloroplast genomes of four *Carex* species: long repetitive sequences trigger dramatic changes in chloroplast genome structure. Front Plant Sci. 2023;14:1100876. <https://doi.org/10.3389/fpls.2023.1100876>
68. Smith DR, Keeling PJ. Mitochondrial and plastid genome architecture: reoccurring themes, but significant differences at the extremes. Proc Natl Acad Sci USA. 2015;112(33):10177–84. <https://doi.org/10.1073/pnas.1422049112>
69. Mower JP, Vickrey TL. Structural diversity among plastid genomes of land plants. Adv Bot Res. 2018;85:263–92. <https://doi.org/10.1016/bs.abr.2017.11.013>
70. Li HT, Luo Y, Gan L, Ma PF, Gao LM, Yang JB, et al. Plastid phylogenomic insights into relationships of all flowering plant families. BMC Biol. 2021;19:232. <https://doi.org/10.1186/s12915-021-01166-2>
71. Stull GW, Dunod SR, Soltis DE, Soltis PS. Resolving basal lamiid phylogeny and the circumscription of Icacinaeaceae with a plastome-scale data set. Am J Bot. 2015;102:1794–813. <https://doi.org/10.3732/ajb.1500298>
72. Givnish TJ, Zuluaga A, Spalink D, Gomez MS, Lam VKY, Saarela JM, et al. Monocot plastid phylogenomics, timeline, net rates of species diversification, the power of multi-gene analyses, and a functional model for the origin of monocots. Am J Bot. 2018;105(11):1888–910. <https://doi.org/10.1002/ajb2.1178>
73. Zhang ML, Wang L, Lei Y, Sanderson SC. Cenozoic evolutionary history of *Zelkova* (Ulmaceae), evidenced from ITS, *trnL-trnF*, *psbA-trnH*, and *rbcl*. Tree Genet Genomes. 2017;13, 1–10. <https://doi.org/10.1007/s11295-017-1182-4>
74. Cai CN, Ma H, Ci XQ, Conran JG, Li J. Comparative phylogenetic analyses of Chinese *Horsfieldia* (Myricaceae) using complete chloroplast genome sequences. J Syst Evol. 2021;59(3):504–14. <https://doi.org/10.1111/jse.12556>
75. Wu ZY, Milne RI, Liu J, Slik F, Yu Y, Luo YH, et al. Phylogenomics and evolutionary history of *Oreocnide* (Urticaceae) shed light on recent geological and climatic events in SE Asia. Mol Phylogenet Evol. 2022;175:107555. <https://doi.org/10.1016/j.ympev.2022.107555>
76. Zhang ZL, Zhang Y, Song MF, Guan YH, Ma XJ. Species identification of *Dracaena* using the complete chloroplast genome as a super-barcode. Front Pharmacol. 2019;10:1441. <https://doi.org/10.3389/fphar.2019.01441>
77. Wang J, Fu CN, Mo ZQ, Möller M, Yang JB, Zhang ZR, Li DZ, et al. Testing the complete plastome for species discrimination, cryptic species discovery and phylogenetic resolution in *Cephalotaxus* (Cephalotaxaceae). Front Plant Sci. 2022;13:768810. <https://doi.org/10.3389/fpls.2022.768810>
78. Yu XQ, Jiang YZ, Folk RA, Zhao JL, Fu CN, Fang L, et al. Species discrimination in *Schima* (Theaceae): next-generation super-barcodes meet evolutionary complexity. Mol Ecol Resour. 2022;22(8):3161–75. <https://doi.org/10.1111/1755-0998.13683>
79. Guo YY, Yang JX, Bai MZ, Zhang GQ, Liu ZJ. The chloroplast genome evolution of Venus slipper (*Paphiopedilum*): IR expansion, SSC contraction, and highly rearranged SSC regions. BMC Plant Biol. 2021;21:248. <https://doi.org/10.1186/s12870-021-03053-y>
80. Claude SJ, Park S, Park S. Gene loss, genome rearrangement, and accelerated substitution rates in plastid genome of *Hypericum ascyron* (Hypericaceae). BMC Plant Biol. 2022;22(1):1–12. <https://doi.org/10.1186/s12870-022-03515-x>
81. Shi WB, Song WC, Liu J, Shi C, Wang S. Comparative chloroplast genome analysis of *Citrus* (Rutaceae) species: insights into genomic characterization, phylogenetic relationships, and discrimination of subgenera. Sci Hortic. 2023;313(1):111909. <https://doi.org/10.1016/j.scienta.2023.111909>
82. Li X, Yang JB, Wang H, Song Y, Corlett RT, Yao X, et al. Plastid NDH pseudo-genization and gene loss in a recently derived lineage from the largest hemiparasitic plant genus *Pedicularis* (Orobanchaceae). Plant Cell Physiol. 2021;62(6):971–84. <https://doi.org/10.1093/pcp/pcab074>
83. Song F, Li T, Burgess KS, Feng Y, Ge XJ. Complete plastome sequencing resolves taxonomic relationships among species of *Calligonum* L. (Polygonaceae) in China. BMC Plant Biol. 2020;20:261. <https://doi.org/10.1186/s12870-020-02466-5>
84. Kim SC, Lee JW, Choi BK. Seven complete chloroplast genomes from *Symplocos*: genome organization and comparative analysis. Forests. 2021;12(5):608. <https://doi.org/10.3390/f12050608>
85. Nanjala C, Wanga VO, Odago W, Mutinda ES, Waswa EN, Oulo MA, et al. Plastome structure of 8 *Calanthe* s.l. species (Orchidaceae): comparative genomics, phylogenetic analysis. BMC Plant Biol. 2022;22:387. <https://doi.org/10.1186/s12870-022-03736-0>
86. Jin GZ, Li WJ, Song F, Yang L, Wen ZB, Feng Y. Comparative analysis of complete *Artemisia* subgenus *Seriphidium* (Asteraceae: Anthemideae) chloroplast genomes: insights into structural divergence and phylogenetic relationships. BMC Plant Biol. 2023;23:136. <https://doi.org/10.1186/s12870-023-04113-1>
87. Zhang DC, Tu JJ, Ding XX, Guan W, Gong L, Qiu XH, et al. Analysis of the chloroplast genome and phylogenetic evolution of *Bidens pilosa*. BMC Genomics. 2023;24:113. <https://doi.org/10.1186/s12864-023-09195-7>
88. Fahrenkrog AM, Matsumoto GO, Toth K, Jokipii-Lukkari S, Salo HM, Häggman H, et al. Chloroplast genome assemblies and comparative analyses of commercially important *Vaccinium* berry crops. Sci Rep. 2022;12:21600. <https://doi.org/10.1038/s41598-022-25434-5>
89. Wan T, Qiao BX, Zhou J, Shao KS, Pan LY, An F, et al. Evolutionary and phylogenetic analyses of 11 *Cerasus* species based on the complete chloroplast genome. Front Plant Sci. 2023;14:1070600. <https://doi.org/10.3389/fpls.2023.1070600>
90. Han M, Xu XL, Xiong Y, Wei HK, Yao KJ, Huang TT, et al. Genome-wide survey and expression analyses of hexokinase family in Poplar (*Populus trichocarpa*). Plants. 2022;11:2025. <https://doi.org/10.3390/plants11152025>
91. Gao CM, Deng YF, Wang J. The complete chloroplast genomes of *Echinacanthus* Species (Acanthaceae): phylogenetic relationships, adaptive evolution, and screening of molecular markers. Front Plant Sci. 2019;9:1989. <https://doi.org/10.3389/fpls.2018.01989>
92. Huang SN, Ge XJ, Cano A, Salazar BGM, Deng YF. Comparative analysis of chloroplast genomes for five *Dicliptera* species (Acanthaceae): molecular structure, phylogenetic relationships, and adaptive evolution. PeerJ. 2020;8:e8450. <https://doi.org/10.7717/peerj.8450>
93. Alzahrani DA, Yaradua SS, Albokhari EJ, Abba A. Complete chloroplast genome sequence of *Barleria prionitis*, comparative chloroplast genomics and phylogenetic relationships among Acanthoideae. BMC Genomics. 2020;21:393. <https://doi.org/10.1186/s12864-020-06798-2>
94. Kaewdaungdee S, Sudmoon R, Tanee T, Lee SY, Chaveerach A. Chloroplast genome analysis for genetic information and authentication in five *Barleria* species. Genes. 2022;13:1705. <https://doi.org/10.3390/genes13101705>
95. Huang SN, Deng YF, Ge XJ. The complete chloroplast genome of *Aphelandra knappiae* (Acanthaceae). Mitochondrial DNA Part B-Resources. 2019;4(1):273–4. <https://doi.org/10.1080/23802359.2018.1541718>
96. Xu G, Wang HX, Zhu ZX, Wang HF. Complete plastome sequence of *Pseuderanthemum haikangense* C.Y. Wu & H.S. Lo (Acanthaceae): a medicinal plant in South China. Mitochondrial DNA Part B-Resources. 2020;5(3):3197–3198. <https://doi.org/10.1080/23802359.2020.1810158>
97. Tong LL, Xu XG, Cheng Y. The complete chloroplast genome sequence of *Thunbergia erecta* (Benth.) T. Anders. (Acanthaceae). Mitochondrial DNA Part

- B-Resources. 2022;7(11):1952–1954. <https://doi.org/10.1080/23802359.2022.2140574>
98. Chen HM, Shao JJ, Zhang H, Jiang M, Huang LF, Zhang Z, et al. Sequencing and analysis of *Strobilanthes cusia* (Nees) kuntze chloroplast genome revealed the rare simultaneous contraction and expansion of the inverted repeat region in angiosperm. *Front Plant Sci.* 2018;9:324. <https://doi.org/10.3389/fpls.2018.00324>
 99. Wang Q, Huang SN, Chen X, Deng YF. The complete chloroplast genome of *Strobilanthes biocollata* (Acanthaceae). *Mitochondrial DNA Part B-Resources.* 2021;6(6):1668–9. <https://doi.org/10.1080/23802359.2021.1927868>
 100. Wang GQ, Liu YH, Zheng YZ, Gu XF, Lu CL. The complete chloroplast genome of *Strobilanthes crispus* (Acanthaceae). *Mitochondrial DNA Part B-Resources.* 2022;7(8):1553–4. <https://doi.org/10.1080/23802359.2022.2111979>
 101. Lin ZL, Huang SN, Deng YF. The complete chloroplast genome of *Rungia pectinata* (Acanthaceae). *Mitochondrial DNA Part B-Resources.* 2019;4(2):2736–7. <https://doi.org/10.1080/23802359.2019.1644216>
 102. Hoch B, Maier RM, Appel K, Igloi GL, Kossel H. Editing of a chloroplast mRNA by creation of an initiation codon. *Nature.* 1991;353(6340):178–80. <https://doi.org/10.1038/353178a0>
 103. Li J, Cullis C. Comparative analysis of 84 chloroplast genomes of *Tylosema esculentum* reveals two distinct cytotypes. *Front Plant Sci.* 2023;13:1025408. <https://doi.org/10.3389/fpls.2022.1025408>
 104. Xiang KL, Mao W, Peng HW, Erst AS, Yang YX, He WC, et al. Organization, phylogenetic marker exploitation, and gene evolution in the plastome of *Thalictrum* (Ranunculaceae). *Front Plant Sci.* 2022;13:897843. <https://doi.org/10.3389/fpls.2022.897843>
 105. Knoop V, C-to-U. U-to-C: RNA editing in plant organelles and beyond. *J Exp Bot.* 2022. <https://doi.org/10.1093/jxb/erac488>. erac488.
 106. Wee CC, Muhammad NAN, Subbiah VK, Arita M, Nakamura Y, Goh HH. Plastomes of *Garcinia mangostana* L. and comparative analysis with other *Garcinia* species. *Plants.* 2023;12:930. <https://doi.org/10.3390/plants12040930>
 107. Zong D, Zhou A, Zhang Y, Zou X, Li D, Duan A, et al. Characterization of the complete chloroplast genomes of five *Populus* species from the western Sichuan plateau, southwest China: comparative and phylogenetic analyses. *PeerJ.* 2019;7:e6386. <https://doi.org/10.7717/peerj.6386>
 108. Kuroda H, Suzuki H, Kusumegi T, Hirose T, Yukawa Y, Sugiura M. Translation of *psbC* mRNAs starts from the downstream GUG, not the upstream AUG, and requires the extended Shine-Dalgarno sequence in Tobacco chloroplasts. *Plant Cell Physiol.* 2007;48(9):1374–8. <https://doi.org/10.1093/pcp/pcm097>
 109. Scharff LB, Ehrnthal M, Janowski M, Childs LH, Hasse C, Gremmels J, et al. Shine-Dalgarno sequences play an essential role in the translation of plastid mRNAs in Tobacco. *Plant Cell.* 2017;29:3085–101. <https://doi.org/10.1105/tpc.17.00524>
 110. Dobrogowski J, Adamiec M, Luciński R. The chloroplast genome: a review. *Acta Physiol Plant.* 2020;42:98. <https://doi.org/10.1007/s11738-020-03089-x>
 111. Sara EG, Jaina M, Alex B, Sean RE, Aurelien L, Simon CP, et al. The pfam protein families database in 2019. *Nucleic Acids Res.* 2018;47(D1):D427–32. <https://doi.org/10.1093/nar/gky995>
 112. Raubeson LA, Peery R, Chumley TW, Dziubek C, Fourcade HM, Boore JL, et al. Comparative chloroplast genomics: analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. *BMC Genomics.* 2007;8:174. <https://doi.org/10.1186/1471-2164-8-174>
 113. Steane DA. Complete nucleotide sequence of the chloroplast genome from the tasmanian blue gum, *Eucalyptus globulus* (Myrtaceae). *DNA Res.* 2005;12(3):215–20. <https://doi.org/10.1093/dnares/dsi006>
 114. Hansen DR, Dastidar SG, Cai Z, Penafior C, Kuehl JV, Boore JL, et al. Phylogenetic and evolutionary implications of complete chloroplast genome sequences of four early-diverging angiosperms: *Buxus* (Buxaceae), *Chloranthus* (Chloranthaceae), *Dioscorea* (Dioscoreaceae), and *Illicium* (Schisandraceae). *Mol Phylogenet Evol.* 2007;45(2):547–63. <https://doi.org/10.1016/j.ympev.2007.06.004>
 115. Jansen RK, Ruhlman TA. Plastid genomes of seed plants. *Genomics of chloroplasts and mitochondria.* Dordrecht: Springer; 2012. 103–26.
 116. Ruhlman TA, Jansen RK. The plastid genomes of flowering plants. In: Maliga P, editor. *Chloroplast biotechnology: methods and protocols.* Methods in molecular biology. New York: Springer; Humana Press; 2014. pp. 3–38.
 117. Moghaddam M, Ohta A, Shimizu M, Terauchi R, Kazempour-Osaloo S. The complete chloroplast genome of *Onobrychis gaubae* (Fabaceae-Papilionoideae): comparative analysis with related IR-lacking clade species. *BMC Plant Biol.* 2022;22:75. <https://doi.org/10.1186/s12870-022-03465-4>
 118. Jung J, Kim C, Kim JH. Insights into phylogenetic relationships and genome evolution of subfamily Commelinoideae (Commelinaceae Mirb.) Inferred from complete chloroplast genomes. *BMC Genomics.* 2021;22:231. <https://doi.org/10.1186/s12864-021-07541-1>
 119. Wang ML, Wang X, Sun JH, Wang YH, Ge Y, Dong WP, et al. Phylogenomic and evolutionary dynamics of inverted repeats across *Angelica* Plastomes. *BMC Plant Biol.* 2021;21:26. <https://doi.org/10.1186/s12870-020-02801-w>
 120. AL-Juhani WS, Alharbi SA, Aboud MMA, Ajohani AY. Complete chloroplast genome of the desert date (*Balanites aegyptiaca* (L.) Del. comparative analysis, and phylogenetic relationships among the members of Zygophyllaceae. *BMC Genomics.* 2022;23:626. <https://doi.org/10.1186/s12864-022-08850-9>
 121. Dong WP, Xu C, Li CH, Sun JH, Zuo YJ, Shi S, et al. *ycf1*, the most promising plastid DNA barcode of land plants. *Sci Rep.* 2015;5:8348. <https://doi.org/10.1038/srep08348>
 122. Dong WP, Liu J, Yu J, Wang L, Zhou SL. Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. *PLoS ONE.* 2012;7:e35071. <https://doi.org/10.1371/journal.pone.0035071>
 123. Kim GB, Lim CE, Kim JS, Kim K, Lee JH, Yu HJ, et al. Comparative chloroplast genome analysis of *Artemisia* (Asteraceae) in East Asia: insights into evolutionary divergence and phylogenomic implications. *BMC Genomics.* 2020;21:415. <https://doi.org/10.1186/s12864-020-06812-7>
 124. Li HL, Xiao WJ, Tong T, Li YL, Zhang M, Lin XX, et al. The specific DNA barcodes based on chloroplast genes for species identification of Orchidaceae plants. *Sci Rep.* 2021;11:1424. <https://doi.org/10.1038/s41598-021-81087-w>
 125. Bi Y, Zhang MF, Xue J, Dong R, Du YP, Zhang XH. Chloroplast genomic resources for phylogeny and DNA barcoding: a case study on *Fritillaria*. *Sci Rep.* 2018;8:1184. <https://doi.org/10.1038/s41598-018-19591-9>
 126. Wang YH, Wang S, Liu YL, Yuan QJ, Sun JH, Guo LP. Chloroplast genome variation and phylogenetic relationships of *atractylodes* species. *BMC Genomics.* 2021;22:103. <https://doi.org/10.1186/s12864-021-07394-8>
 127. Kelchner SA, Clark LG. Molecular evolution and phylogenetic utility of the chloroplast *rpl16* intron in *Chusquea* and the Bambusoideae (Poaceae). *Mol Phylogenet Evol.* 1997;8(3):385–97. <https://doi.org/10.1006/mpev.1997.0432>
 128. Yamaguchi H, Utano A, Yasuda K, Yano A, Soejima A. A molecular phylogeny of wild and cultivated *Echinochloa* in East Asia inferred from non-coding region sequences of *trnT-L-F*. *Weed Biol Manage.* 2005;5(4):210–8. <https://doi.org/10.1111/j.1445-6664.2005.00185.x>
 129. Lang P, Dane F, Kubisiak TL. Phylogeny of *Castanea* (Fagaceae) based on chloroplast *trnT-L-F* sequence data. *Tree Genet Genomes.* 2006;2(3):132–9. <https://doi.org/10.1007/s11295-006-0036-2>
 130. Bai HR, Oyeibanji O, Zhang R, Yi TS. Plastid phylogenomic insights into the evolution of subfamily Dialioideae (Leguminosae). *Plant Divers.* 2021;43(1):27–34. <https://doi.org/10.1016/j.pld.2020.06.008>
 131. Li DM, Li J, Wang DR, Xu YC, Zhu GF. Molecular evolution of chloroplast genomes in subfamily Zingiberioideae (Zingiberaceae). *BMC Plant Biol.* 2021;21:558. <https://doi.org/10.1186/s12870-021-03315-9>
 132. Namjung J, Do HDK, Kim C, Choi HJ, Kim JH. Complete chloroplast genomes shed light on phylogenetic relationships, divergence time, and biogeography of Alliioideae (Amaryllidaceae). *Sci Rep.* 2021;11:3262. <https://doi.org/10.1038/s41598-021-82692-5>
 133. Dong SJ, Zhou MJ, Zhu JX, Wang QR, Cheng RB. The complete chloroplast genomes of *Tetrastigma hemsleyanum* (Vitaceae) from different regions of China: molecular structure, comparative analysis and development of DNA barcodes for its geographical origin discrimination. *BMC Genomics.* 2022;23:620. <https://doi.org/10.1186/s12864-022-08755-7>
 134. Wang HY, Yang BM, Wang H, Xiao HX. Impact of different numbers of microsatellite markers on population genetic results using SLAF-seq data for *Rhododendron* species. *Sci Rep.* 2021;11:8597. <https://doi.org/10.1038/s41598-021-87945-x>
 135. Liu LX, Wang YW, He PZ, Li P, Lee JK, Soltis DE et al. Chloroplast genome analyses and genomic resource development for epilithic sister genera *Oreolithrope* and *Mukdenia* (Saxifragaceae), using genome skimming data. *BMC Genomics.* 2018;19:235. <https://doi.org/10.1186/s12864-018-4633-x>
 136. Diethard T, Manfred R. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Res.* 1984;12:4127–38. <https://doi.org/10.1093/nar/12.10.4127>
 137. Jurka J, Pethiyagoda C. Simple repetitive DNA sequences from primates: compilation and analysis. *J Mol Evol.* 1995;40:120–6. <https://doi.org/10.1007/BF00167107>
 138. George B, Bhatt BS, Awasthi M, George B, Singh AK. Comparative analysis of microsatellites in chloroplast genomes of lower and higher plants. *Curr Genet.* 2015;61:665–77. <https://doi.org/10.1007/s00294-015-0495-9>
 139. Jayaswal K, Sharma H, Bhandawat A, Sagar R, Jayaswal D, Kumar A, et al. Chloroplast derived SSRs reveals genetic relationships in domesticated

- alliums and wild relatives. Genet Resour Crop Evol. 2022;69:363–72. <https://doi.org/10.1007/s10722-021-01235-z>
140. Ebert D, Peakall R. Chloroplast simple sequence repeats (cpSSRs): technical resources and recommendations for expanding cpSSR discovery and applications to a wide array of plant species. Mol Ecol Resour. 2009;9(3):673–90. <https://doi.org/10.1111/j.1755-0998.2008.02319.x>
 141. Yang Z, Zhao T, Ma Q, Liang L, Wang G. Comparative genomics and phylogenetic analysis revealed the chloroplast genome variation and interspecific relationships of *Corylus* (Betulaceae) species. Front Plant Sci. 2018;9:927. <https://doi.org/10.3389/fpls.2018.00927>
 142. Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch S. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Res. 2001;11(8):1441–52. <https://genome.cshlp.org/content/11/8/1441>
 143. Souza UJd, Nunes R, Targueta CP, Diniz-Filho JAF, Telles MPC. The complete chloroplast genome of *Stryphnodendron adstringens* (Leguminosae-Caesalpinioideae): comparative analysis with related mimosoid species. Sci Rep. 2019;9:14206. <https://doi.org/10.1038/s41598-019-50620-3>
 144. Posada D, Buckley TR. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and bayesian approaches over likelihood ratio tests. Syst Biol. 2004;53(5):793–808. <https://doi.org/10.1080/10635150490522304>
 145. Rasmussen MD, Kellis M. Accurate gene-tree reconstruction by learning gene- and species-specific substitution rates across multiple complete genomes. Genome Res. 2007;17(12):1932–42. <https://doi.org/10.1101/gr.7105007>
 146. Betancur R, Naylor GJP, Ortí G. Conserved genes, sampling error, and phylogenomic inference. Syst Biol. 2014;63(2):257–62. <https://doi.org/10.1093/sysbio/syt073>
 147. McDade LA, Daniel TF, Kiel CA, Borg AJ. Phylogenetic placement, delimitation, and relationships among genera of the enigmatic Nelsonioideae (Lamiales: Acanthaceae). Taxon. 2018;61:637–51. <https://doi.org/10.1002/tax.613012>
 148. Hansen B. *Justicia* sect. *Grossa* sect. *Nordic* J Bot. 1987;7(5):505–9. <https://doi.org/10.1111/j.1756-1051.1987.tb02017.x>. Acanthaceae.
 149. Hu CC, Tsui HP, Xi YZ, Zhang YL. Pollen morphology of one genus in Lepidagathideae, two in Andrographideae and eight in Justicieae (Acanthaceae) from China. Acta Phytotaxonomica Sinica. 2005;43(2):151–62.
 150. Tang HM. 2011. Seed morphology of Acanthaceae and taxonomy of *Justicia* sect. *Mananthes* from China. A thesis submitted to Graduate University of Chinese Academy of Sciences. (In Chinese).
 151. Benoist R. Acanthaceae. Lecomte, Flore générale de l'Indo Chine 4. Paris: Masson; 1935.
 152. Li J, Wang S, Jing Y, Wang L, Zhou S. A modified CTAB protocol for plant DNA extraction. Chin Bull Bot. 2013;48(1):72–8. <https://doi.org/10.3724/SPJ.1259.2013.00072>
 153. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):1–7. <https://doi.org/10.1093/bioinformatics/btu170>
 154. Jin JJ, Yu WB, Yang JB, Song Y, Claude W, Yi TS, et al. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol. 2020;21:241. <https://doi.org/10.1186/s13059-020-02154-5>
 155. Wick RR, Schultz MB, Justin Z, Holt KE. Bandage: interactive visualization of de novo genome assemblies. Bioinformatics. 2015;31(20):3350–2. <https://doi.org/10.1093/bioinformatics/btv383>
 156. Qu XJ, Moore MJ, Li DZ, Yi TS. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. Plant Methods. 2019;15(1):50. <https://doi.org/10.1186/s13007-019-0435-7>
 157. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, et al. GeSeq—versatile and accurate annotation of organelle genomes. Nucleic Acids Res. 2017;45(W1):W6–W11. <https://doi.org/10.1093/nar/gkx391>
 158. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28(12):1647–9. <https://doi.org/10.1093/bioinformatics/bts199>
 159. Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res. 2019;47(W1):W59–W64. <https://doi.org/10.1093/nar/gkz238>
 160. Amirouyefi A, Hyvönen J, Poccai P. IRscope: an online program to visualize the junction sites of chloroplast genomes. Bioinformatics. 2018;34(17):3030–1. <https://doi.org/10.1093/bioinformatics/bty220>
 161. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. Nucleic Acids Res. 2004;32(suppl2):W273–9. <https://doi.org/10.1093/nar/gkh458>
 162. Brudno M, Do CB, Cooper GM, Kim MF, Davydov E, Green ED, et al. NISC Comparative Sequencing Program. LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. Genome Res. 2003;13(4):721–31. <https://doi.org/10.1101/gr.926603>
 163. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–80. <https://doi.org/10.1093/molbev/mst010>
 164. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009;25(11):1451–2. <https://doi.org/10.1093/bioinformatics/btp187>
 165. Peden JF. Analysis of codon usage. Univ Nottm. 2000;90(1):73–4. <https://doi.org/10.1006/expr.1997.4185>
 166. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res. 2001;29(22):4633–42. <https://doi.org/10.1093/nar/29.22.4633>
 167. Beier S, Thiel T, Muench T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. Bioinformatics. 2017;33(16):2583–5. <https://doi.org/10.1093/bioinformatics/btx198>
 168. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999;27(2):573–80. <https://doi.org/10.1093/nar/27.2.573>
 169. Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst Biol. 2007;56(4):564–77. <https://doi.org/10.1080/10635150701472164>
 170. Xia XH, Lemey P. Assessing substitution saturation with DAMBE. The phylogenetic handbook, a practical approach to DNA and protein phylogeny; 2009. 612–26. <https://doi.org/10.1017/CBO9780511819049.022>
 171. Stamatakis A. Bioinformatics. 2014;30(9):1312–3. <https://doi.org/10.1093/bioinformatics/btu033>. RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies.
 172. Darriba D, Taboada GL, Doallo R, Posada D. Jmodeltest 2: more models, new heuristics and parallel computing. Nat Methods. 2012;9(8):772. <https://doi.org/10.1038/nmeth.2109>
 173. Ronquist F, Teslenko M, Mark Pvd, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 2012;61(3):539–42. <https://doi.org/10.1093/sysbio/sys029>
 174. Nylander JAA. MrModeltest v2. Program distributed by the author. Evolutionary Biology Centre. Uppsala University; 2004.

Publisher's Note

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