Characterization and comparative analysis of the plastome sequence from
Justicia ventricosa (Lamiales: Acanthaceae)

Junchen Zhou, Qing Du, Mei Jiang, Shengyu Liu, Liqiang Wang, Haimei Chen, Bin Wang and Chang Liu

ABSTRACT
Justicia ventricosa is a characteristic ethnic herb commonly used to treat Orthopedic pains. Here, to confirm its phylogenetic position and to develop molecular markers that can distinguish different Justicia species, we obtained and analyzed the plastome of Justicia ventricosa. The plastome was sequenced using the Illumina HiSeq sequencing platform, assembled with NOVOPlasty, and annotated with CPGAVAS2. The genome has a circular structure of 149,700 bp, containing a single-copy region of 82,324 bp, a small single-copy region of 17,260 bp, and two reverse repeat regions of 25,058 bp each. It encodes 112 unique genes, including 76 protein-coding genes, eight ribosomal RNA genes, and 28 transfer RNA genes. Twenty cis-splicing genes were found. In total, we predicted 19 microsatellite repeats and 13 tandem repeat sequences. For distributed repeats, four were palindromic repeats and five were direct repeats. To find the highly variable intergenic spacer (IGS) regions, we calculated the K2P distances for IGS regions from four Justicia species. The K2P values ranged from 6.11 to 57.82. The largest K2P distances were found for ccs–ndhD, petB–petD, psbK–psbl, and ycf4–cemA. Phylogenetic analysis results showed that J. ventricosa was most closely related to J. leptostachya. To determine how Justicia species adapt to the environment, we performed selection pressure analysis. Nine genes were found to have undergone positive selection. Lastly, we performed a genome-wise DNA barcode prediction, seven pairs of primers were found. The results provide valuable information that can be used for molecular marker development and bioprospecting in Justicia species.

Introduction
Justicia ventricosa Wallich belongs to the Asteraceae family. It is found in the South and Southwest regions of China, including Guangdong, Guangxi, Yunnan, Hainan provinces, and Hong Kong SAR, and it is also distributed in Vietnam to Thailand and Myanmar. It is usually found under the sparse forest or shrub near the village, which are both wild and cultivated (Khaing et al., 2020). Some Justicia species are valuable medicines (Lino et al. 1997; Kitadi et al. 2019). The whole plant of J. ventricosa can be used as a medicine, which can reinforce tendons and bones, dispel wind and dampness, and treat fracture, sprain, arthritis, chronic low back, and leg pain (Gomez et al. 1996).

The complete plastome can be a circular, linear, or polycyclic double-stranded DNA molecule, with the length mostly around 120–160 kb (Bock 2015). In most higher plants, the complete plastome is highly conserved and consists of four parts, including a small single-copy (SSC) region, a large single-copy (LSC) region, and two inverted repeats (IRA and IRB) regions. The complete plastome sequence can be used for better barcoding, understanding of phylogenetic relationship, understanding patterns of gene loss, understanding adaptive changes that optimize photosynthesis, and potential application in synthetic biology (Olejniczak et al. 2016). To take advantage of these benefits with a complete plastome sequence, we sequenced and assembled a complete plastome sequence of J. ventricosa and compared it with other Justicia species.

Materials and methods

Plant materials, DNA extraction
Fresh leaf samples of J. ventricosa were collected from the Guangxi Medical Botanical Garden, Nanning, Guangxi, China (Geospatial coordinates: N22.859968, E108.383475). The specimen was deposited at the Institute of Medicinal Plant Development with the specimen number implad201910236.
Total genomic DNA was extracted from the fresh leaf cells of *J. ventricosa* using the plant genomic DNA kit (Tiangen Biotech, Beijing, China).

**Plastome sequencing, assembly, and annotation**

We constructed the library with an insert size of 500 bp fragments. We sequenced the library using the Illumina HiSeq platform (Caporaso et al. 2012). The genome was assembled using NOVOPlasty with default parameters (Dierckxsens et al. 2017) and annotated with CPGAVAS2 web service (Shi et al. 2019). The final genome assembly and annotation results have been submitted to GenBank with the accession number MW580585.

**Genome analysis**

The complete plastome sequences of *J. adhatoda* (NC_047476.1), *J. leptostachya* (NC_044668.1), *J. nutans* (NC_042162.1), and *J. pseudospicata* (NC_044862.1), *Silvianthus bracteatus* (NC_044974.1) were downloaded from the GenBank database for further analysis. The IR regions of the four *Justicia* species were analyzed using the IRscope online software. The genetic distance of the intergenic spacer (IGS) was calculated by the distmat program from EMBOSS (v6.3.1) and Kimura 2-parameter (K2P) evolutionary model (Rice et al. 2000; Casanellas et al. 2020). The selection pressure of protein-coding genes (PCGs) was analyzed using the aBSREL (adaptive branch-site random effects likelihood) model implemented in the Hyphy software (Smith et al. 2005).
We used ecoPrimers software to identify potential DNA barcoding markers to distinguish the four Justicia species (Riaz et al. 2011; Khan et al. 2020).

Phylogenetic analysis

We performed the phylogenetic analysis with the plastomes of five Justicia species and that of Silvianthus bracteatus as the outgroup taxon. The shared protein sequences of these species were extracted by PhyloSuite (v1.1.16) and aligned with MAFFT (v7.3.13) (Katoh et al. 2002; Zhang et al. 2020). We performed the phylogenetic analysis using the maximum-likelihood (ML) method implemented in IQ-TREE (v1.6.8) with the TVM + F + I + G4 nucleotide substitution mode (Nguyen et al. 2015). We assessed the reliability of the phylogenetic tree by bootstrap test with 1000 replications. We generated the final phylogenetic tree with the maximum-likelihood (ML) method implemented in MEAG7.0 (Kumar et al. 2016).

Results

Genome organization and compositions

The plastome sequence of J. ventricosa is a typical circular DNA molecule with a total length of 149,700 bp. A schematic representation of the plastome is shown in Figure 1. It has a conserved tetrad structure, including an LSC region, an SSC region, and a pair of IR regions, with length of 82,324 bp, 17,260 bp, and 25,058 bp, respectively. The total GC content in the plastome of J. ventricosa is 38.36%. The GC content of the IR region (43.36%) is higher than those of the SSC (32.67%) and LSC (36.52%) regions. The length of the coding sequence (CDS) in the plastome is 79,767 bp, accounting for 53.28% of the whole genome length. The size of the rRNA gene is 9410 bp, accounting for 6.29% of the entire genome length. The tRNA gene’s length is 2873 bp, accounting for 1.92% of the total genome length.

Gene content

The plastome of J. ventricosa encodes 112 unique genes, including 76 PCGs, 28 tRNA genes, and eight rRNA genes, which were made of two copies of rrm16S, rrm23S, rrm4.5S, and rrm5S genes. The gene structures of 20 spliced genes are shown in Figure 2. The black regions represent the exons. The arrow indicates the sense direction. The 20 spliced genes contain 12 PCGs and eight tRNA genes. All PCGs except ycf3 have only one intron. In contrast, the ycf3 gene contains two introns. Three genes, petB, petD, and rpl16, have small exons, consistent with those found in other plastomes (Table 1).

Repeat sequences

Microsatellite repeats are sequences consisting of multiple repeat units with length ranging from 1 to 6 nucleotides. In the plastome of J. ventricosa, there were 19 microsatellite repeats (A/T), followed by repeats made up with G/C, AT/AT, and AAG/CTT as repeat units (Table 2). Tandem repeat sequence refers to identical DNA segments lying one after the other in a sequence. It is similar to microsatellite sequence except that the repeat unit is usually $\geq 7$. We identified 13 tandem repeat sequences, whose total length ranged from 25 to 46 bp. The similarity between all repeat units is $\geq 90\%$ (Table 3). Dispersed repeats are distributed in the genome in a scattered manner. Dispersed repeats can be classified as palindromic repeats and direct repeats. With the threshold of $1e^{-4}$, nine dispersed repeats met the requirements. Four of them are palindrome repeats, with the lengths of the repeat units being 41, 39, 34, and 36 bps,
respectively. The other five are direct repeats. The lengths of the repeat units being 41, 39, 39, 35, and 32 bps, respectively (Table 4).

**IR structure analysis of four Justicia species**

IR boundary analysis of four complete plastomes of *Justicia* species showed that they had the same boundary structure (Figure 3). For the four species, the *rps19* genes were located in the border area of LSC and IRb. Its 5’ ends were located in the LSC region, and 3’ ends were located in the IRb region. In contrast, *ndhF* genes were located at the border area of IRb and SSC. Most of the *ycf1* genes were located in the border area of SSC and IRa. Besides, *rpl2*, *trnH* gene spacers were located in the IRa and LSC border area, respectively. Some of the *ycf1* genes were found in the border area of IRb and SSC in *J. pseudospicata* and *J. ventricosa*. A small fragment of the *rps19* gene was found in the border area of IRa and LSC.

| Table 2. Microsatellites in the *J. ventricosa* plastome. |
|-------------|-------------|
| **Type of repeat** | **Number of repeats** |
| A/T | 19 |
| C/G | 1 |
| AT/AT | 3 |
| AAG/CTT | 1 |

| Table 3. Tandem repeats in the *J. ventricosa* plastome. |
|-------------|----------------|---------------|
| **Start-end** | **Length of repeat (bp)** | **Copy number of repeat** | **The final length of repeat (bp)** | **Matching score of the repeat (%)** |
| 1716–1740 | 12 | 2.1 | 25 | 100 |
| 8226–8264 | 14 | 1.1 | 29 | 100 |
| 44,230–44,264 | 17 | 2.1 | 36 | 94 |
| 45,242–45,281 | 18 | 2.2 | 40 | 100 |
| 55,882–55,909 | 14 | 2.0 | 28 | 100 |
| 58,548–58,591 | 21 | 2.1 | 44 | 91 |
| 66,456–66,179 | 13 | 2.6 | 34 | 90 |
| 66,590–66,624 | 16 | 2.2 | 35 | 100 |
| 74,487–74,517 | 16 | 2.2 | 32 | 93 |
| 77,503–77,548 | 24 | 1.9 | 46 | 90 |
| 80,374–80,411 | 18 | 2.1 | 38 | 100 |
| 120,055–120,090 | 17 | 2.1 | 36 | 100 |
| 120,059–120,096 | 17 | 2.2 | 37 | 90 |

| Table 4. Interspersed repeats in the *J. ventricosa* plastome. |
|-------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| **Length of repeat unit I (bp)** | **Start of repeat unit I** | **Repeat type** | **Length of repeat unit II (bp)** | **Start of repeat unit II** | **E-value** |
| 41 | 96,120 | D | 41 | 116,676 | 1.30e-15 |
| 41 | 116,676 | P | 41 | 135,863 | 1.30e-15 |
| 39 | 42,685 | D | 39 | 96,122 | 2.09e-14 |
| 39 | 116,678 | D | 39 | 96,122 | 2.09e-14 |
| 39 | 42,685 | P | 39 | 135,863 | 2.09e-14 |
| 34 | 7733 | P | 34 | 44,145 | 2.18e-09 |
| 36 | 7733 | P | 36 | 74,919 | 7.57e-09 |
| 35 | 7733 | P | 35 | 74,919 | 7.57e-09 |
| 32 | 7733 | D | 32 | 34,565 | 4.58e-05 |
Hypervariable region identification

To find highly variable regions among the *Justicia* plastomes, we calculated the genetic distance of IGS regions using the K2P model. There were 58 IGS regions with K2P values ranging from 6.11 to 57.82. K2P values of *ccsA*-ndhD, *petB*-petD, *psbK*-psbI, and *ycf4*-cemA were higher, at 12.57, 10.73, 11.00, and 11.89, respectively (Figure 4). The variations of these IGS regions are large, suggesting these regions can be used to develop potential molecular markers for species discrimination.

Selective pressure analysis

Using the adaptive branch-site random effects likelihood model implemented in the HyPhy software, we analyzed the selection pressure of PCGs in *J. ventricosa*, *J. leptostachya*, and *J.*
adhatoda, and J. nutans. The genes of accD, ndhB, rpoA, rpoC1 were positively selected in species of J. ventricosa. For J. leptostachya, three genes of ndhF, rps12, ycf1 were positively selected. Moreover, the genes of ndhG, ycf2 were found positively selected in J. adhatoda and J. nutans, respectively. The optimized branch length, LRT, and p value varied in the range of 0.0079–0.0179, 8.1342–21.8014, 0–0.03, respectively. These genes may be related to the adaptation of J. ventricosa under diverse environments (Table 5).

Phylogenetic analysis

Based on the 42 common protein sequences of five Justicia species and Silvianthus bracteatus, we constructed the phylogenetic tree (Figure 5). As expected, the J. ventricosa, J. leptostachya, J. adhatoda, J. nutans, and J. pseudospicata were clustered into one branch. The genetic relationship between J. ventricosa and J. leptostachya, J. adhatoda, and J. pseudospicata were the closest, respectively. In contrast, S. bracteatus was distantly related to the Justicia species. The bootstrap scores of all branches containing Justicia species are 100%, supporting the reliability of the tree (Figure 5).

Discussion

In this study, we obtained and analyzed the plastome of J. ventricosa for the first time. Overall, the plastome contains highly conserved features. Some worthy features are the followings. IR boundary analysis of the plastomes of four

---

Table 5. Selection pressure analysis.

| Species          | Gene | Position | Optimized branch length | LRT    | p Value |
|------------------|------|----------|-------------------------|--------|---------|
| J. ventricosa    | accD | 56,118–57,590 | 0.0079                  | 21.3902 | .0000   |
|                  | ndhB | 94,315–92,304 | 10.4026                 | .0095  |
|                  | rpoA | 76,068–77,081 | 10.6976                 | .0082  |
|                  | rpoC1| 23,060–20,234 | 21.8014                 | .0000  |
| J. leptostachya  | ndhF | 107,062–109,350 | 0.0129                  | 10.5454 | .0089   |
|                  | rps12| 67,827–67,940 | 8.6738                  | .0228  |
|                  | ycf1 | 95,117–95,899  | 14.9044                 | .0010  |
| J. adhatoda      | ndhG | 114,627–115,157 | 0.0094                  | 14.7286 | .0011   |
| J. nutans        | ycf2 | 85,633–92,400  | 0.0179                  | 8.1342  | .0300   |

Table 6. Primers for amplifying DNA barcodes to distinguish the four Justicia species: J. ventricosa, J. leptostachya, J. nutans, and J. adhatoda.

| No. | Forward primer sequences | Reverse primer sequences |
|-----|--------------------------|--------------------------|
| 1   | AACTGTCCAAAGATGAAAATCTTT | GGATATACAAAAAGACCCAGCTCT |
| 2   | ATGGATCTGGCTTCCGAAACGTTCGATG | AAGTAAGAAACGGGAAATTAAAT |
| 3   | AGAATTCGCCACAAGAATACAGTGAATAGT | TCAGATTTTATGACAAAGCTCAG |
| 4   | CAGGGTTTTAAGCCGAAAAAGAAAAAGAGAG | TCTTACACCCCTGACATCAGAATGAG |
| 5   | AATATCAGCCTTCCTCCGTTAGGTTATTT | TATCAGACAAAAATGAAAAAGGTAACAG |
| 6   | ATGTCTCATTTATTATTACTTATTATTATT | TTCAAGAAATGTTCTCAGCAGGGAAGGACATCAG |
| 7   | TCCTGCTGAGAATCAGAATCTCTGGTAGA | ATTTTCCCTCGACATCCAGGAGAAGTTGGAATAGTA |

Figure 5. The phylogenetic tree of five Justicia species of Acanthaceae and Silvianthus bracteatus selected as the outgroup. The tree was constructed using the maximum-likelihood method. Bootstrap support values were calculated from 1000 replicates.
Justicia species shows that the rps19 gene is located at the IRB-LSC boundary. Most of the rps19 gene sequences are in the LSC region. In contrast, the ycf1 gene is located in the IRB-SSC border. Besides, the K2P values for four IGS regions: cccA-ndhF (12.57), petB-petD (10.73), psbK-psbl (11.00), and ycf4-cemA (11.89), are high. Highly variable regions can be used as DNA barcodes for species identification. We also identified seven regions that can be used to distinguish the four species: J. ventricosa, J. leptostachya, J. nutans, and J. adhatoda. We also performed selective pressure analysis to identify the different genes in the Justicia species, which may play an essential role for Justicia species to adapt to varied environments (Chen and Blanchette 2007). The phylogenetic tree results showed that the five Justicia species are closely related, as would be expected. Overall, there is limited information on the phylogeny of the Justicia species (Yaradua et al. 2019). Therefore, further studies on the plastomes of this genus are needed in future studies.

Disclosure statement
No potential conflict of interest was reported by the authors.

Funding
This work was supported by the National Science & Technology Fundamental Resources Investigation Program of China [2018FY100705], The National Mega-Project for Innovative Drugs of China [2019ZX09735-002], Chinese Academy of Medical Sciences, Innovation Funds for Medical Sciences (CIFMS) [2016-I2M-3-016, 2017-I2M-1-013], National Science Foundation Funds [81872966], Key Laboratory for Tibet Plateau Phytochemistry of Qinghai Province [2020-ZJ-40], the Key Scientific Research Projects of Hunan Education Department (No. 20A467), and Hunan Technological Innovation Guidance Project (2018SK52001).

ORCID
Chang Liu http://orcid.org/0000-0003-3879-7302

Data availability statement
The data that support the findings of this study are openly available in the GenBank database at https://www.ncbi.nlm.nih.gov with the accession number is MWS80585. The associated BioProject, SRA, and BioSample numbers are PRJNA717803, SRR14085608, and SAMN18511406, respectively.

References
Bock R. 2015. Engineering plastid genomes: methods, tools, and applications in basic research and biotechnology. Annu Rev Plant Biol. 66:211–241.
Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM, Betley J, Fraser L, Bauer M. 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J. 6(8):1621–1624.
Casanellas M, Fernández-Sánchez J, Roca-Lacostena J. 2020. Embeddability and rate identifiability of Kimura 2-parameter matrices. J Math Biol. 80(4):995–1019.
Chen H, Blanchette M. 2007. Detecting non-coding selective pressure in coding regions. BMC Evol Biol. 7(Suppl.1):S9–S13.
Dierksxens N, Mardulyn P, Smits G. 2017. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. 45(4):e18.
Gomez J, Reyes R, Aguilar I. 1996. Chemistry and pharmacology of selected Asian and American medicinal species of Justicia. In V.K. Gupta (Ed) Bioactive Phytochemicals: “Perspectives for Modern Medicine”, 1st ed.; p. 457–475, New Delhi: Daya Publishing House.
Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30(14):3059–3066.
Khang WW, Yee YM, Yee NN. 2020. Morphological and anatomical studies on Justicia adhatoda L. and Justicia ventricosa Wall. Myanmar Korea Conf Res J. 3:1088–1097.
Khan I, Shinwari ZK, Zahra NB, Jan SA, Shinwari S, Najeebullah S. 2020. DNA barcoding and molecular systematics of selected species of family Acanthaceae. Pak J Bot. 52(1):205–212.
Kitadi JM, Lengbiye EM, Gbolo BZ, Inkoto CL, Muanayishay CL, Lufulanbo GL, Tshibangu DS, Tshilanda DD, Mbala BM, Mpiana P. 2019. Justicia secunda Vahl species: phytochemistry, pharmacology and future directions: a mini-review. Disc Phytomed. 6(4):157–171.
Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genomics analysis version 7.0 for bigger datasets. Mol Biol Evol. 33(7):1870–1874.
Lino C, Taveira M, Viana G, Matos F. 1997. Analgesic and antiinflammatory activities of Justicia pectoralis Jacq and its main constituents: coumarin and umbelliferone. Phytother Res. 11(3):211–215.
Nguyen L-T, Schmidt HA, Von HA, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 32(1):268–274.
Olejniczak SA, Lojewska E, Kowalczyk T, Sakowicz T. 2016. Chloroplasts: state of research and practical applications of plastome sequencing. Planta. 244(3):517–527.
Riaz T, Shehzad W, Viari A, Pompanon F, Taberlet P, Coissac E. 2011. ecoPrimers: inference of new DNA barcode markers from whole genome sequence analysis. Nucleic Acids Res. 39(21):e145.
Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European molecular biology open software suite. Trends Genet. 16(6):276–277.
Shi L, Chen H, Jiang M, Wang L, Wu X, Huang L, Liu C. 2019. CPGAVAS2, an integrated plastome sequence annotator and analyzer. Nucleic Acids Res. 47(W1):W65–W73.
Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Kosakovský Pond SL. 2015. Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. Mol Biol Evol. 32(5):1342–1353.
Yaradua SS, Alzahrani DA, Albokhary EJ, Abba A, Bello A. 2019. Complete chloroplast genome sequence of Justicia flava: genome comparative analysis and phylogenetic relationships among Acanthaceae. BioMed Res Int. 2019:1–17.
Zhang D, Gao F, Jakovlíc I, Zou H, Zhang J, Li WX, Wang GT. 2020. PhyloSuite: an integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. Mol Ecol Resour. 20(1):348–355.