

Research Article

Phylogeny and Comparative Analysis for the Plastid Genomes of Five *Tulipa* (Liliaceae)

Juan Li,¹ Megan Price,² Dan-Mei Su,¹ Zhen Zhang,¹ Yan Yu,¹ Deng-Feng Xie,¹ Song-Dong Zhou ,¹ Xing-Jin He ,¹ and Xin-Fen Gao³

¹Key Laboratory of Bio-Resources and Eco-Environment of Ministry of Education, College of Life Sciences, Sichuan University, Chengdu, 610065 Sichuan, China

²Sichuan Key Laboratory of Conservation Biology on Endangered Wildlife, College of Life Sciences, Sichuan University, Chengdu, 610065 Sichuan, China

³CAS Key Laboratory of Mountain Ecological Restoration and Bioresource Utilization & Ecological Restoration and Biodiversity Conservation Key Laboratory of Sichuan Province, Chengdu Institute of Biology, Chinese Academy of Sciences, Chengdu, 610041 Sichuan, China

Correspondence should be addressed to Song-Dong Zhou; zsd@scu.edu.cn and Xing-Jin He; xjhe@scu.edu.cn

Received 4 November 2020; Accepted 9 June 2021; Published 19 June 2021

Academic Editor: Cao Deng

Copyright © 2021 Juan Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Species of *Tulipa* (Liliaceae) are of great horticultural importance and are distributed across Europe, North Africa, and Asia. The Tien Shan Mountain is one of the primary diversity centres of *Tulipa*, but the molecular studies of *Tulipa* species from this location are lacking. In our study, we assembled four *Tulipa* plastid genomes from the Tien Shan Mountains, *T. altaica*, *T. iliensis*, *T. patens*, and *T. thianschanica*, combined with the plastid genome of *T. sylvestris* to compare against other Liliaceae plastid genomes. We focussed on the species diversity and evolution of their plastid genomes. The five *Tulipa* plastid genomes proved highly similar in overall size (151,691–152,088 bp), structure, gene order, and content. With comparative analysis, we chose 7 mononucleotide SSRs from the *Tulipa* species that could be used in further population studies. Phylogenetic analyses based on 24 plastid genomes robustly supported the monophyly of *Tulipa* and the sister relationship between *Tulipa* and *Amana*, *Erythronium*. *T. iliensis*, *T. thianschanica*, and *T. altaica* were clustered together, and *T. patens* was clustered with *T. sylvestris*, with our results clearly demonstrating the relationships between these five *Tulipa* species. Our results provide a more comprehensive understanding of the phylogenomics and comparative genomics of *Tulipa*.

1. Introduction

Plastid, a unique semiautonomous organelle of green plants, serves as a protagonist in photosynthesis and carbon fixation and provides essential energy for plants [1, 2]. Plastid DNA has been widely used in evolutionary biology analyses because of its uniparental inheritance, relatively stable genome structure, and gene content [3–7]. A plastid genome has a quadripartite circular structure consisting of two copies of inverted repeat (IR) regions, a large single copy (LSC) region, and a small copy region (SCR) in most seed plants [8–10]. Comparative genomics of whole plastid genomes has been used to generate genetic markers for molecular

identification [11, 12]. Plastid genomic data provide new and robust insights into phylogenetic analysis and genetic variation detection of plants [13].

Tulips (*Tulipa* L.) are famous ornamental and cut flowers due to their beautiful and colorful corolla [14–16]. *Tulipa* is a member of Liliaceae sensu APG IV [17], subfamily Lilioideae, tribe Tulipeae Kostel [18, 19]; the tribe Tulipeae including *Gagea* Salisbury, *Amana* Honda, *Erythronium* L., and *Tulipa* [20] is sister to the tribe Lilieae [21]. The close relationship between *Gagea* and *Tulipa*, *Amana*, and *Erythronium* is generally accepted; however, the phylogenetic relationships among *Tulipa*, *Amana*, and *Erythronium* have remained controversial due to inconsistencies across studies

and lack of strong support [22, 23]. *Amana* used to be treated as a group in *Tulipa* [24–26] but is now generally accepted as a separate genus [27–29]. Several studies have supported a sister relationship between *Tulipa* and *Amana* [30, 31], whereas others clustered *Tulipa* and *Erythronium* together [22, 29]. A close relationship between *Amana* and *Erythronium* has been suggested; however, the *Tulipa* data used in these studies was limited [31–33].

Tulipa includes more than 100 species [34, 35] and are distributed across the temperate regions of Europe, North Africa, and Asia. The Tien Shan, Pamir-Alay, and Caucasus Mountains are considered the primary diversity centres of *Tulipa* [36, 37]. Thirteen species of *Tulipa* are recorded and described in China, and 11 species are distributed in the Xinjiang Uygur Autonomous Region [37], which is home to the Tien Shan Mountain. Several studies have utilised a few plastid regions and internal transcribed spacer makers to analyse the phylogeny and evolution of *Tulipa* in the Middle East [38, 39] and Europe [29]. However, the phylogeny and evolution of *Tulipa* at Tien Shan Mountain are limited. Two *Tulipa* plastid genomes have been published [40, 41], but genomics analysis for *Tulipa* is lacking. Therefore, comprehensive studies including the specimens from Tien Shan Mountain and accurate analysis of the plastid genome are required to enable further *Tulipa* molecular studies.

We collected four *Tulipa* species from Tien Shan Mountain and reported the complete plastid genome sequences of these four *Tulipa* species to address gaps in *Tulipa* phylogenetic research. Combining previously reported plastid genomes of *T. sylvestris* and other Liliaceae species, we performed comparative genomics and phylogenetic analyses. We aimed to resolve the phylogenetic relationships between five *Tulipa* species and other genera and characterize and compare the plastid genomes of *Tulipa* species to detect the genetic variation. Our conclusion will contribute to an understanding of *Tulipa* plastid phylogenomics and provide genetic resources for tulip research.

2. Materials and Methods

2.1. Plant Materials, DNA Extraction, and Sequencing. Fresh leaves of four *Tulipa* species, *T. altaica*, *T. iliensis*, *T. patens*, and *T. thianschanica*, were collected from Yuming county (Xinjiang Uygur Autonomous Region, China) and dried with silica gel, then stored at -80°C. Total genomic DNA was extracted from leaf material with a modified CTAB method [42] and then sequenced on an Illumina Novaseq2500 sequencer (Illumina, San Diego, CA, USA) by Biomarker Technologies, Inc. (Beijing, China).

2.2. Plastid Genome Assembly, Annotation, and Analysis. The plastid genomes were assembled using raw data by NOVO-Plasty 2.7.2 [43], and the plastid genome of *T. sylvestris* (MT261172) was selected for seed input and the reference sequence. Genome annotation and IR region search were processed by PGA [44]. Geneious R11 [45] was used on manual modifications to accurately confirm the start and stop codons and the exon-intron boundaries of genes based on comparison with other Liliaceae plastid genomes. The circu-

lar plastid genome map of *Tulipa* was drawn by the OGDRAW1 program [46]. The total GC content and GC content of each region (IR, LSC, SSC) were analysed by the program Geneious R11.

2.3. Contraction and Expansion of IRs and SSRs. Four plastid genomes of the tribe Tulipeae, *Amana edulis* (NC034707), *Erythronium japonicum* (MT261155), *Erythronium sibiricum* (NC035681), and *Gagea triflora* (MT261157), were downloaded from the GenBank for comparative analysis with five *Tulipa* species.

The IR/SC borders with full annotations were compared between the five *Tulipa* species and with the other four tribe Tulipeae species using the program IRscope (<https://irscope.shinyapps.io/irapp/>) [47]. Simple sequence repeats (SSRs) were detected using Perl script MISA [48] with the following minimum number (threshold) settings: 10, 5, 4, 3, 3, and 3 repeat units for mono-, di-, tri-, tetra-, penta-, and hexanucleotide SSRs, respectively.

2.4. Phylogenetic Analyses. To reconstruct phylogenetic relationships between the five *Tulipa* species and other Liliaceae species, a total of 20 plastid genome sequences were downloaded from GenBank, and *Smilax china* (Smilacaceae, HM536959) was selected as an outgroup. The alignment of 24 plastid genome sequences was generated by MAFFT v7.402 [49] with the default parameters set. The best-fit model selected by ModelFinder was GTR+G. A maximum likelihood (ML) tree was constructed using RAxML 8.0 [50] with 1000 bootstrap replicates.

2.5. Codon Usage Analysis and SNP Analyses. All 84 protein-coding sequences extracted from nine plastid genomes were used to analyse codon usage, which was undertaken with the CodonW v1.4.2 program (J. Peden, <http://codonw.sourceforge.net>). The plastid genome sequences of nine tribe Tulipeae species were used for SNP analyses. The alignment of all plastid genome sequences was generated by MAFFT v7.402. SNP analysis was conducted to determine the nucleotide diversity of the plastid genomes using DnaSP v5, with the following parameters: 200 bp of step size and 600 bp window length [51]. Results of the SNP analysis were illustrated using TBtools v1.087 software [52].

3. Results and Discussion

3.1. The Plastid Genomes of Five *Tulipa* Species. The total plastid genome sizes of the five *Tulipa* species ranged from 151691 bp (*T. altaica*) to 152088 bp (*T. patens*). All five plastid genomes showed the typical quadripartite structure (Figure 1) and, like other angiosperms, consisted of a pair of IR regions. The G+C content of the five species in whole genomes (36.6–36.7) and LSC (34.5–34.6), SSC (30.0–30.2) was nearly identical but in the IR regions was higher (42.0%). Details of genome features are given in Table 1. The annotated genome sequences of *T. altaica*, *T. iliensis*, *T. patens*, and *T. thianschanica* were deposited in the GenBank under the accession numbers MW077741, MW077740, MW077739, and MW077738, respectively (Table 1).

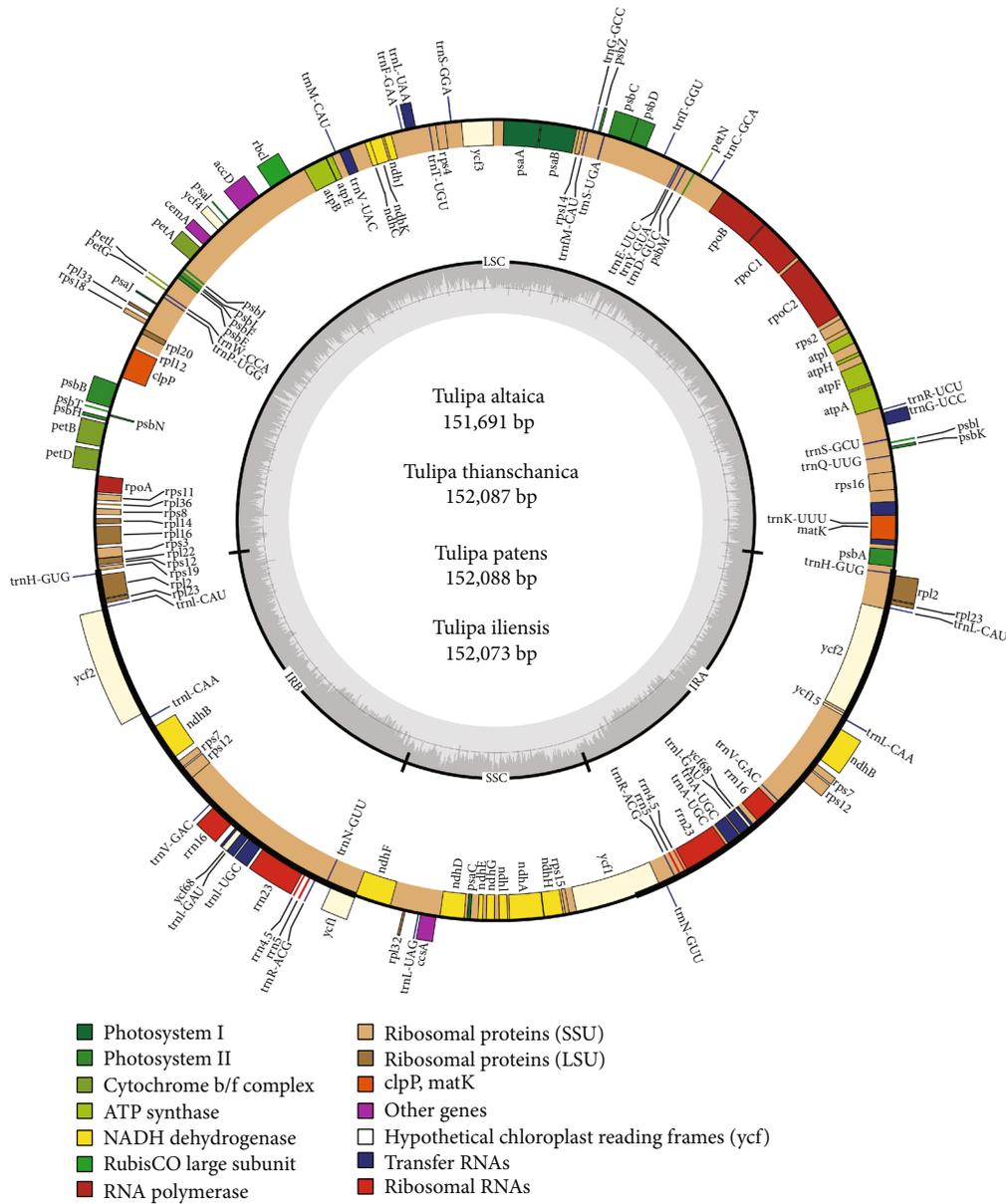


FIGURE 1: Plastid genome map of *T. altaica*, *T. iliensis*, *T. patens*, *T. sylvestris*, and *T. thianschanica*.

The plastid genomes of five *Tulipa* species contained 134 genes. Of these 134 genes, 112 genes were nonredundant including 78 protein-coding genes, 4 ribosomal RNA (rRNA) genes, and 30 transfer RNA (tRNA) genes, and four genes were pseudogenes (Table 2). The 134 genes had 18 duplicated genes located in the IR region, including six coding genes (*ndhB*, *rpl2*, *rpl23*, *rps7*, *rps12*, and *ycf2*), four rRNA genes (*rrn4.5*, *rrn5*, *rrn16*, and *rrn23*), and eight tRNA genes (*trnA-UGC*, *trnH-GUG*, *trnI-CAU*, *trnI-GAU*, *trnL-CAA*, *trnN-GUU*, *trnR-ACG*, and *trnV-GAC*).

Four pseudogenes (*ycf1*, *rps19*, and two *ycf68*) were found in the five plastid genomes (Table 2). The *rps19* and *ycf1* genes were located in the boundary area of the IR regions, and their protein-coding ability was lost due to partial gene duplication [6, 10, 40, 53]. Whether *ycf68* and *ycf15* genes lost abilities or occur as pseudogenes has already been discussed

in several studies [40, 53, 54]. In this study, the *ycf15* gene was not annotated due to its short length. The *infA* gene, which codes for translation initiation factor 1, was lost in all five *Tulipa* plastid genomes because of a missing base. The deletion of the *infA* gene also occurred in *Amana* and *Erythronium* [40], which were sister relationships with *Tulipa*, and many other seed plants, such as *Smilax* (Smilacaceae) [40] and *Alstroemeria* (Alstroemeriaceae) [33].

3.2. Inverted Repeats Contraction, Expansion, and SSR Analysis. The IR/SC boundary regions of the *Tulipa* plastid genomes were compared to the closely related plastid genomes, *Amana*, *Erythronium*, and *Gagea*. Typically, the lengths of IR regions are different among various plant species [53], while the lengths of *Tulipa* plastid genome IR regions were similar (26307 bp-26341 bp) but larger than

TABLE 1: Genome features of five *Tulipa* plastid genomes.

Region	<i>T. altaica</i>	<i>T. iliensis</i>	<i>T. patens</i>	<i>T. thianschanica</i>	<i>T. sylvestris</i>
GenBank numbers	MW077741	MW077740	MW077739	MW077738	MT261172
Genome size (bp)	151,691	152,073	152,088	152,087	151,940
LSC (bp)	81,806	82,252	82,262	82,317	81,958
SSC (bp)	17,213	17,175	17,204	17,156	17,524
IR (bp)	26,336	26,323	26,311	26,307	26,341
Number of total genes	134	134	134	134	134
Protein-coding genes	79	79	79	79	79
tRNAs	30	30	30	30	30
rRNAs	4	4	4	4	4
Total GC content (%)	36.7	36.6	36.6	36.6	36.7
LSC GC content (%)	34.6	34.6	34.6	34.5	34.6
SSC GC content (%)	30.0	30.1	30.1	30.1	30.2
IR GC content (%)	42.0	42.0	42.0	42.0	42.0

TABLE 2: List of genes encoded in five *Tulipa* species.

Category for genes	Group of genes	Name of genes
Self-replication	Transfer RNAs	<i>trnA-UGC</i> *, <i>trnC-GCA</i> , <i>trnD-GUC</i> , <i>trnE-UUC</i> , <i>trnF-GAA</i> , <i>trnFM-CAU</i> , <i>trnG-GCC</i> , <i>trnG-UCC</i> , <i>trnH-GUG</i> *, <i>trnI-CAU</i> *, <i>trnI-GAU</i> *, <i>trnK-UUU</i> , <i>trnL-CAA</i> *, <i>trnL-UAA</i> , <i>trnL-UAG</i> , <i>trnM-CAU</i> , <i>trnN-GUU</i> *, <i>trnP-UGG</i> , <i>trnQ-UUG</i> , <i>trnR-ACG</i> *, <i>trnR-UCU</i> , <i>trnS-GCU</i> , <i>trnS-GGA</i> , <i>trnS-UGA</i> , <i>trnT-GGU</i> , <i>trnT-UGU</i> , <i>trnV-GAC</i> *, <i>trnV-UAC</i> , <i>trnW-CCA</i> , <i>trnY-GUA</i>
	Ribosomal RNAs	<i>rrn4.5</i> *, <i>rrna5</i> *, <i>rrn16</i> *, <i>rrn23</i> *
	RNA polymerase	<i>rpoA</i> , <i>rpoB</i> , <i>rpoC1</i> , <i>rpoC2</i>
	Small subunit of ribosomal proteins (SSU)	<i>rps2</i> , <i>rps3</i> , <i>rps4</i> , <i>rps7</i> *, <i>rps8</i> , <i>rps11</i> , <i>rps12</i> *, <i>rps14</i> , <i>rps15</i> , <i>rps16</i> , <i>rps18</i> , <i>rps19</i> * (<i>rps19</i> , ψ <i>rps19</i>)
	Large subunit of ribosomal proteins (LSU)	<i>rpl2</i> *, <i>rpl14</i> , <i>rpl16</i> , <i>rpl20</i> , <i>rpl22</i> , <i>rpl23</i> *, <i>rpl32</i> , <i>rpl33</i> , <i>rpl36</i>
Genes for photosynthesis	Subunits of NADH-dehydrogenase	<i>ndhA</i> , <i>ndhB</i> *, <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 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 cytochrome b/f complex	<i>petA</i> , <i>petB</i> , <i>petD</i> , <i>petG</i> , <i>petL</i> , <i>petN</i>
	Subunits of ATP synthase	<i>atpA</i> , <i>atpB</i> , <i>atpE</i> , <i>atpF</i> , <i>atpH</i> , <i>atpI</i>
Other genes	Large subunit of rubisco	<i>rbcL</i>
	Protease	<i>clpP</i>
	Maturase	<i>matK</i>
	Subunit of acetyl-CoA-carboxylase	<i>accD</i>
	Envelope membrane protein	<i>cemA</i>
Genes of unknown function	C-type cytochrome synthesis gene	<i>ccsA</i>
	Hypothetical chloroplast reading frames	<i>ycf1</i> * (<i>ycf1</i> , ψ <i>ycf1</i>), <i>ycf2</i> *, <i>ycf3</i> , <i>ycf4</i> , ψ <i>ycf68</i> *

*Duplicated genes; ψ shows pseudogenes.

Amana (25633 bp), *Erythronium* (25765 bp and 26001 bp), and *Gagea* (25521 bp) plastid genome IR regions (Figure 2). Furthermore, the expansion and contraction at the IR regions were the primary cause of size variation in plastid genomes and played an important role in the evolution of

the genome [55–57]. After comparing the location and adjacent genes of IR regions between nine plastid genomes, we found that the gene number and order were conserved, but some distinct differences existed at the boundaries (Figure 2). The boundary of the LSC and IRb regions was

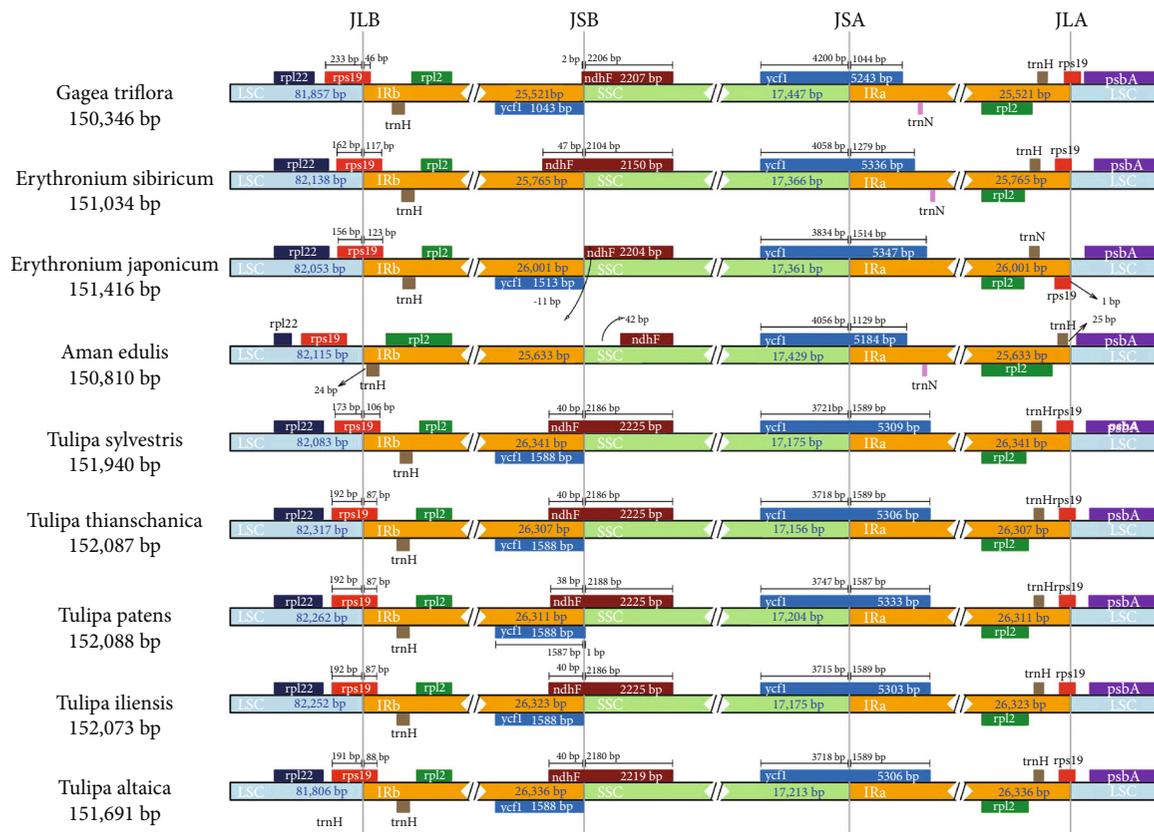


FIGURE 2: Comparison of the LSC, IR, and SSC junction positions among nine tribe Tulipeae plastid genomes. JLB: junction of LSC and IRb; JSB: junction of SSC and IRb; JSA: junction of SSC and IRa; JLA: junction of LSC and IRa.

located at *rps19* genes in eight plastid genomes and positioned at the noncoding region between *rps19* and *rpl2* genes in the *Amana edulis* plastid genome. The *ndhF* and *ycf1* genes traversed the regions of IRb/SSC and SSC/IRa, with 38/40 bp of the *ndhF* gene and 1587 bp/1589 bp of the *ycf1* gene located at the IR region in the *Tulipa* plastid genome. In general, the length and structure of IR regions were similar in *Tulipa* genomes but showed obvious differences with other tribe Tulipeae genomes.

Simple sequence repeats (SSRs) in the plastid genomes are suitable molecular makers and have been widely used in evolutionary and ecological studies due to their high variation [58–64]. Given that SSRs have high polymorphism at the species level and commonly show intraspecific variation, SSRs were used as important molecular markers to reconstruct phylogenetic relationships [65–67]. The results of SSR analysis of *Tulipa* and its close relatives are shown in Figure 3 and Table S1-S2. There were six categories of SSRs (mono-, di-, tri-, tetra-, penta-, and hexanucleotide repeats) found in the plastid genome of nine species, and the mononucleotide repeats were the most frequent (Figure 3(a), Table S1). The highest percentages of SSRs located in the LSC region were the maximum (69.44%-79.17%), and 4%-10% SSRs were distributed in IR regions (Figure 3(b), Table S2). A total of 41 types were detected in nine plastid genomes (Figure 3(c)), where bases T and A were the dominant elements. In this study, we manually chose 7 mononucleotide SSRs located in *trnK-rps16*, *psbK-I*, *accD-*

psaI, and *psaJ-rpl33* regions and *atpF*, *rpoC1*, and *petB* genes as effective polymorphic SSRs between *Tulipa* species based on three critical criteria outlined in previous research [40]. These 7 mononucleotide SSRs could be used in the further population studies of *Tulipa* (Table S3).

3.3. Codon Usage Bias and SNP Analyses. The results of codon usage frequency and relative synonymous codon usage (RSCU) from 84 protein-coding sequences from nine tribe Tulipeae species are presented in Figure 4 and Table S4. The 84 protein-coding sequences were similar across the nine species. The total codon number from the nine species plastid genomes ranged between 21170 (*Erythronium sibiricum*) and 21284 (*T. thianschanica*) (Table S4). Among all amino acids, leucine and cysteine were the most and the least frequent, on average, 2165 (10.19%) and 333 (1.57%), respectively (Figure 4 and Table S4). The third codon position occupied by the A or T base was the most common in all nine tribe Tulipeae species, which is also found in other plastid genomes in seed plants [68–71]. Codon usage bias was related to gene expression and differed between species [72, 73]. Our results of codon usage bias will be important for understanding the molecular evolution mechanisms of *Tulipa* and its relatives.

The single-nucleotide polymorphism (SNP) analyses of the alignment for the nine tribe Tulipeae plastid genomes showed that the IR regions were more conserved than the LSC and SSC regions, where the SNP number was low

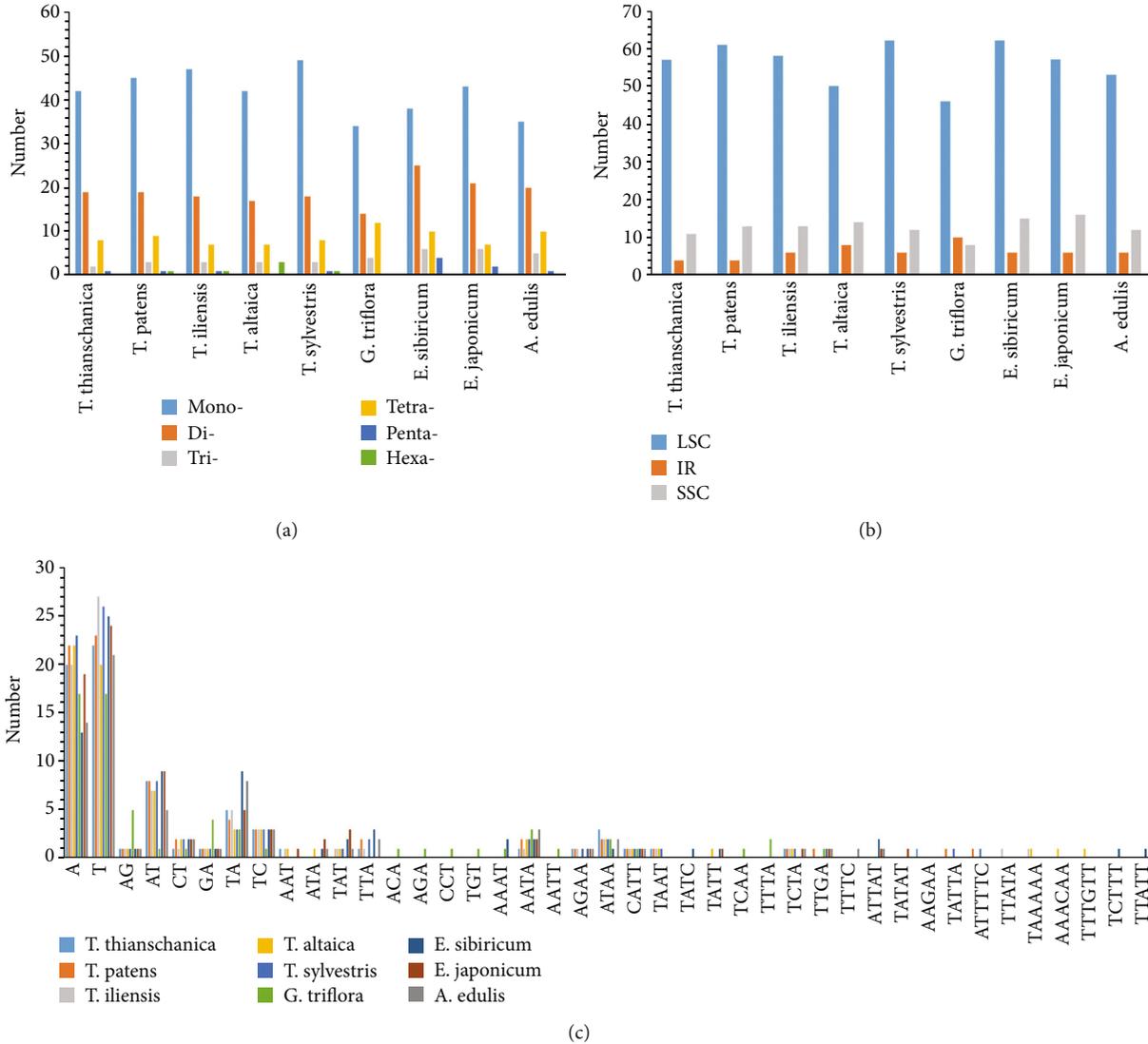


FIGURE 3: Analyses of simple sequence repeats (SSRs) in nine tribe Tulipeae plastid genomes: (a) numbers of different repeat types; (b) frequency of repeat types in LSC, SSC, and IR regions; (c) numbers of identified SSR motifs.

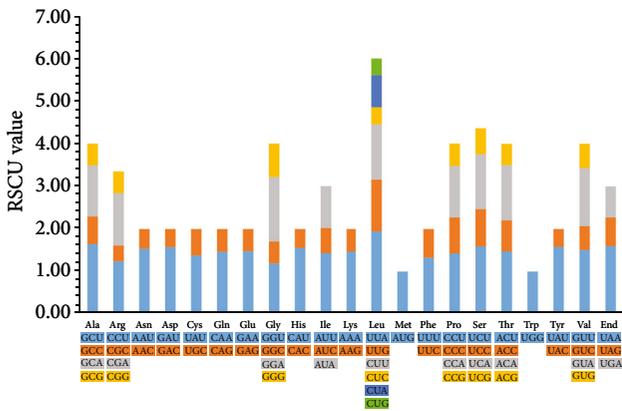


FIGURE 4: Codon content of 20 amino acids and stop codons in 84 coding genes of nine tribe Tulipeae plastid genomes. The color of the histogram corresponds to the color of codons.

(Figure 5), and was in agreement with previous reports for the angiosperm plastid genome [74, 75]. The noncoding regions were more variable than the coding-protein regions. In the sequence alignment of the nine tribe Tulipeae plastid genomes, six noncoding regions, *rps16-trnQ*, *trnE-trnT*, *accD-psaI*, *rpl32-trnL*, *rps15-ycf1*, and *rps4-trnT*, and *ycf1* and *ndhA* genes were highly variable.

3.4. Phylogenetic Analysis. The phylogenetic relationship of *Tulipa* and other Liliaceae species was reconstructed based on 24 plastid genomes, representing 14 genera (Figure 6). The maximum likelihood (ML) tree strongly supported *Tulipa* as a monophyletic genus that was sister to *Amana* and *Erythronium* (1/100%), which was the same as previous research [29, 40]. *Tulipa*, *Amana*, *Erythronium*, and *Gagea* formed a monophyletic clade (1/100%). The ML analytical result of *Tulipa* and other Liliaceae species based on plastid genomes was in accordance with APG IV [17]. Previous

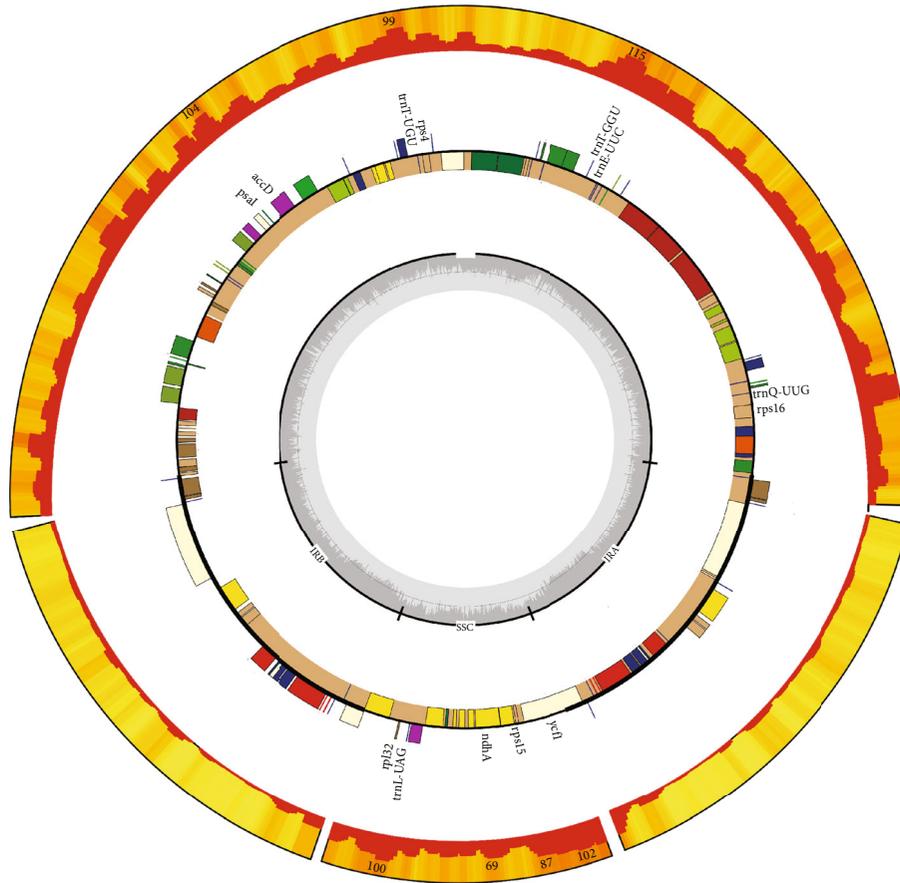


FIGURE 5: Statistics of the SNPs in the nine tribe Tulipeae plastid genomes. Red bars and heat map in the outer ring represent the SNP number. The locations and SNP numbers of eight highly variable regions (see text) are labeled.

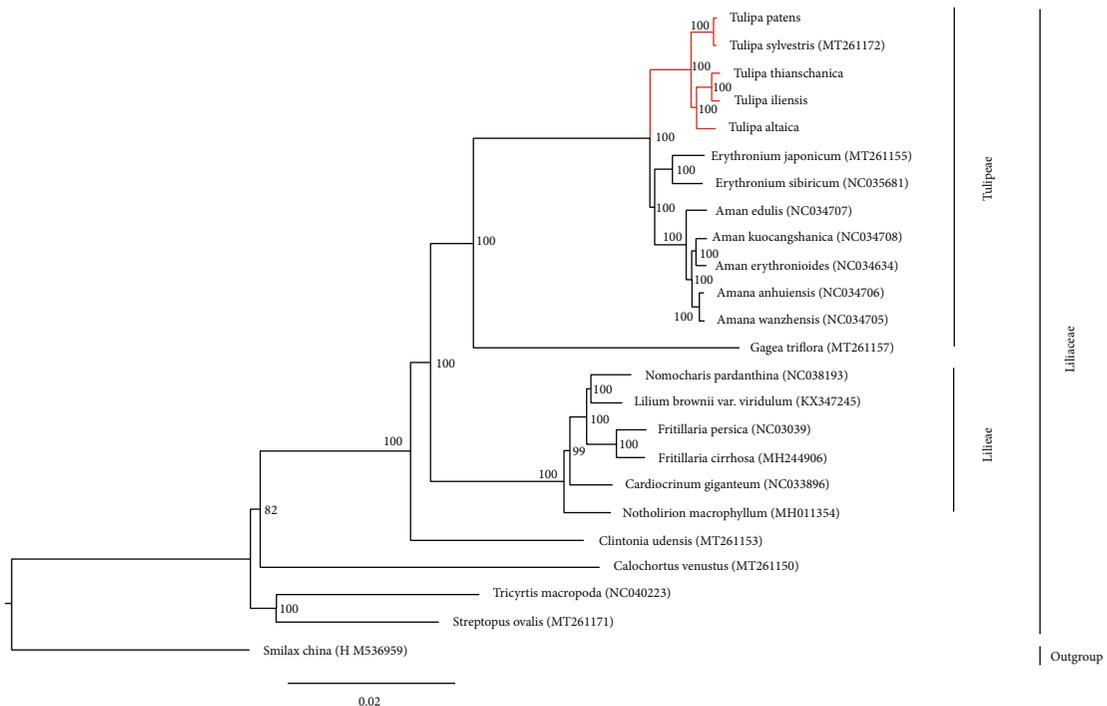


FIGURE 6: ML phylogenetic tree reconstruction containing the plastid genomes of 24 plants. *Smilax china* was set as the outgroup.

studies could not resolve the phylogenetic relationships of *Tulipa*, *Amana*, and *Erythronium* because of insufficient data. In our ML tree, inferred from 23 Liliaceae species' plastid genomes, *Amana* and *Erythronium* clustered together with strong support (1/100%), and *Tulipa* was a sister to the *Amana*/*Erythronium* clade. The phylogenomics reconstructed distinct relationships among *Tulipa* and other genera.

In *Tulipa*, *T. iliensis* and *T. thianschanica* were very close and were sisters to *T. altaica*. We also found that *T. patens* was sister to *T. sylvestris*. In previous research [26, 38], *Tulipa* was divided into four subgenera: *Clusianae*, *Tulipa*, *Eriostemones*, and *Orithyia*. *T. iliensis* and *T. altaica* belong to subgenus *Tulipa* [37], yet *T. iliensis*, *T. thianschanica*, and *T. altaica* have similar morphological characteristics, and our results confirmed this grouping. *T. patens* and *T. sylvestris* were within the subgenus *Eriostemones*, and *T. patens* was once treated as a varietas of *T. sylvestris* [37]. Our phylogenetic tree based on plastid genomes demonstrated clear relationships between the five *Tulipa*, which was in accordance with a previous classification [37], and found that *T. thianschanica* should belong to subgenus *Tulipa*. Our results provide a better understanding of the evolution and molecular biology of *Tulipa*.

4. Conclusions

In this study, the plastid genome sequences of four *Tulipa* species were reported. We described the comparative characteristics of nine tribe Tulipeae plastid genomes and the phylogenetic relationships of 23 Liliaceae plastid genomes. We found that *Tulipa* plastid genomes were highly similar in overall size, structure, IR/SC boundary, SSRs, and codon usage bias. The phylogenetic tree identified a clear sister relationship between *Tulipa* and *Amana*, *Erythronium* and clear relationships between the five *Tulipa* species. Our study supplements the molecular data of *Tulipa* and provides a better understanding of *Tulipa* plastid genome evolution.

Data Availability

The assembled plastid genome sequences of the four species were submitted to NCBI with the accession numbers MW077741 (*T. altaica*), MW077740 (*T. iliensis*), MW077739 (*T. patens*), and MW077738 (*T. thianschanica*). Users can download the data as a reference for research purposes only.

Conflicts of Interest

There was no conflict of interest.

Acknowledgments

We acknowledge Hai-Ying Liu, Yi-Qi Deng, Sheng-Bin Jia, and Qiu-Pin Jiang for their help in material collection. This work was supported by the National Natural Science Foundation of China (Grant Nos. 31872647 and 32070221); the National Specimen Information Infrastructure, Educational Specimen Sub-Platform (Grant No. 2005DKA21403-JK);

and the Fourth National Survey of Traditional Chinese Medicine Resources (Grant No. 2019PC002).

Supplementary Materials

Table S1: number of different SSR categories detected in nine species. Table S2: the frequency of identified SSRs in LSC, IR, and SSC of nine species. Table S3: seven polymorphic SSRs between *Tulipa* species. Table S4: the codon numbers of amino acids in nine plastid genomes. (*Supplementary Materials*)

References

- [1] S. Wicke, G. M. Schneeweiss, C. W. dePamphilis, K. F. Müller, and D. Quandt, "The evolution of the plastid chromosome in land plants: gene content, gene order, gene function," *Plant Molecular Biology*, vol. 76, no. 3-5, pp. 273-297, 2011.
- [2] H. Daniell, C. S. Lin, M. Yu, and W. J. Chang, "Chloroplast genomes: diversity, evolution, and applications in genetic engineering," *Genome Biology*, vol. 17, no. 1, p. 134, 2016.
- [3] M. Park, H. Park, H. Lee, B.-h. Lee, and J. Lee, "The Complete Plastome Sequence of an Antarctic Bryophyte *Sanionia uncinata* (Hedw.) Loeske," *International Journal of Molecular Sciences*, vol. 19, no. 3, p. 709, 2018.
- [4] W. P. Dong, H. Liu, C. Xu, Y. Zuo, Z. Chen, and S. Zhou, "A chloroplast genomic strategy for designing taxon specific DNA mini-barcodes: a case study on ginsengs," *BMC Genetics*, vol. 15, no. 1, 2014.
- [5] K. Nadachowska-Brzyska, C. Li, L. Smeds, G. Zhang, and H. Ellegren, "Temporal dynamics of avian populations during Pleistocene revealed by whole-genome sequences," *Current Biology*, vol. 25, no. 10, pp. 1375-1380, 2015.
- [6] D. F. Xie, Y. Yu, Y. Q. Deng et al., "Comparative analysis of the chloroplast genomes of the Chinese endemic genus *Urophyta* and their contribution to chloroplast phylogeny and adaptive evolution," *International Journal of Molecular Sciences*, vol. 19, no. 7, p. 1847, 2018.
- [7] H. Y. Liu, Y. Yu, Y. Q. Deng, J. Li, Z.-X. Huang, and S.-D. Zhou, "The chloroplast genome of *Lilium henrici*: genome structure and comparative analysis," *Molecules*, vol. 23, no. 6, p. 1276, 2018.
- [8] R. K. Jansen, L. A. Raubeson, J. L. Boore et al., "Methods for obtaining and analyzing whole chloroplast genome sequences," *Methods in Enzymology*, vol. 395, no. 20, pp. 348-384, 2005.
- [9] R. K. Jansen and T. A. Ruhlman, "Plastid genomes of seed plants," in *Advances in Photosynthesis and Respiration* vol. 35, R. V. Knoop, Ed., pp. 103-126, Springer, Dordrecht, The Netherlands, 2012.
- [10] X. L. Guo, H. Y. Zheng, P. Megan, S.-D. Zhou, and X.-J. He, "Phylogeny and comparative analysis of Chinese *Chamaesium* species revealed by the complete plastid genome," *Plants*, vol. 9, no. 8, p. 965, 2020.
- [11] W. Dong, J. Liu, J. Yu, L. Wang, and S. Zhou, "Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding," *PLoS One*, vol. 7, no. 4, p. e35071, 2012.
- [12] S. Ruben, P. C. Carlos, L. A. Diana et al., "Comparative plastome genomics and phylogenomics of *Brachypodium*: flowering time signatures, introgression and recombination in

- recently diverged ecotypes,” *New Phytologist*, vol. 218, no. 4, pp. 1631–1644, 2017.
- [13] T. L. Hong, T. S. Yi, L. M. Gao et al., “Origin of angiosperms and the puzzle of the Jurassic gap,” *Nature plants*, vol. 5, no. 5, pp. 461–470, 2019.
- [14] C. M. Van, D. Kerckhoffs, and T. J. Van, “Interspecific crosses in the genus *Tulipa* L.: identification of pre-fertilization barriers,” *Sexual Plant Reproduction*, vol. 10, no. 2, pp. 116–123, 1997.
- [15] R. Juodkaite, A. Baliuniene, J. R. Naujalis, M. Navalinskienė, and M. Samuitienė, “Selection and presentation of tulip (*Tulipa* L.) species and cultivars to the Lithuanian plant genetic resources,” *Biologia*, vol. 54, no. 2, pp. 139–146, 2008.
- [16] S. J. Van, “Classified list and international register of plant genetic resources,” *Biologia*, vol. 54, pp. 139–146, 1996.
- [17] The Angiosperm Phylogeny Group, “Angiosperm Phylogeny Group: An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV,” *Botanical Journal of the Linnean Society*, vol. 181, no. 1, pp. 1–20, 2016.
- [18] M. N. Tamura, “Liliaceae,” in *Flowering Plants Monocotyledons vol. 3*, K. Kubitzki, Ed., pp. 343–353, Springer, Verlag, Berlin, 1998.
- [19] P. Wendelbo and K. H. Rechinger, *Gagea, Flora Iranica vol. 165*, K. H. Rechinger, Ed., Akademische Druck- u. Verlagsanstalt, Graz, Austria, 1990.
- [20] P. Angela, H. Orte, P. Jens, H. Alexander, and P. Lorenzo, “A pre-Miocene Irano-Turanian cradle: Origin and diversification of the species-rich monocot genus *Gagea* (Liliaceae),” *Ecology and Evolution*, vol. 9, no. 10, pp. 5870–5890, 2019.
- [21] J. S. Kim and J. H. Kim, “Updated molecular phylogenetic analysis, dating and biogeographical history of the lily family (Liliaceae: Liliales),” *Botanical Journal of the Linnean Society*, vol. 187, no. 4, pp. 579–593, 2018.
- [22] G. A. Allen, D. E. Soltis, and P. S. Soltis, “Phylogeny and biogeography of *Erythronium* (Liliaceae) inferred from chloroplast *matK* and nuclear *rDNA* ITS sequences,” *Systematic Botany*, vol. 28, no. 3, pp. 512–523, 2003.
- [23] G. Petersen, O. Seberg, and J. I. Davis, “Phylogeny of the Liliales (monocotyledons) with special emphasis on data partition congruence and RNA editing,” *Cladistics*, vol. 29, no. 3, pp. 274–295, 2013.
- [24] J. V. Eijk, L. V. Raamsdonk, W. Eikelboom, and R. J. Bino, “Interspecific crosses between *Tulipa gesneriana* cultivars and wild *Tulipa* species: a survey,” *Sexual Plant Reproduction*, vol. 4, no. 1, pp. 1–5, 1991.
- [25] M. W. van Rossum, M. Alberda, and L. H. van der Plas, “Tulipaline and tuliposide in cultured explants of tulip bulb scales,” *Phytochemistry*, vol. 49, no. 3, pp. 723–729, 1998.
- [26] B. J. Zonneveld, “The systematic value of nuclear genome size for “all” species of *Tulipa* L. (Liliaceae),” *Plant Systematics and Evolution*, vol. 281, no. 1–4, pp. 217–245, 2009.
- [27] D. Y. Tan, X. R. Li, and D. Y. Hong, “Neotypification and additional description of *Amana anhuiensis* (X.S.Shen) D.Y.Tan & D.Y.Hong (Liliaceae) from Anhui, China,” *Acta Botanica Boreali-Occidentalia Sinica*, vol. 28, pp. 393–395, 2008.
- [28] B. X. Han, K. Zhang, and L. Q. Huang, “*Amana wanzhensis* (Liliaceae), a new species from Anhui, China,” *Phytotaxa*, vol. 177, no. 2, pp. 118–124, 2014.
- [29] M. J. M. Christenhusz, R. Govaerts, J. C. David et al., “Tiptoe through the tulips - cultural history, molecular phylogenetics and classification of *Tulipa* (Liliaceae),” *Botanical Journal of the Linnean Society*, vol. 172, no. 3, pp. 280–328, 2013.
- [30] K. Hayashi and H. Kawano, “Molecular systematics of *Lilium* and allied genera (Liliaceae): phylogenetic relationships among *Lilium* and related genera based on the *rbcl* and *matK* gene sequence data,” *Plant Species Biology*, vol. 15, no. 1, pp. 73–93, 2000.
- [31] M. Zarrei, P. Wilkin, M. F. Fay, M. J. Ingrouille, S. Zarre, and M. W. Chase, “Molecular systematics of *Gagea* and *Lloydia* (Liliaceae Liliales): implications of analyses of nuclear ribosomal and plastid DNA sequences for infrageneric classification,” *Annals of Botany*, vol. 104, no. 1, pp. 125–142, 2009.
- [32] N. Rønsted, S. Law, H. Thornton, M. F. Fay, and M. W. Chase, “Molecular phylogenetic evidence for the monophyly of *Fritillaria* and *Lilium* (Liliaceae; Liliales) and the infrageneric classification of *Fritillaria*,” *Molecular Phylogenetics and Evolution*, vol. 35, no. 3, pp. 509–527, 2005.
- [33] J. S. Kim and J. H. Kim, “Comparative genome analysis and phylogenetic relationship of order Liliales insight from the complete plastid genome sequences of two Lilies (*Lilium longiflorum* and *Alstroemeria aurea*),” *PLoS One*, vol. 8, no. 6, p. e68180, 2013.
- [34] A. D. Hall, *The genus Tulipa*, W. W. Smith, Ed., The Royal Horticultural Society, London, Britain, 1940.
- [35] A. Marasek and K. Okazaki, “Analysis of introgression of the *Tulipa fosteriana* genome into *Tulipa gesneriana* using GISH and FISH,” *Euphytica*, vol. 160, no. 2, pp. 217–230, 2008.
- [36] Z. P. Botschantzeva, *Tulips: taxonomy, morphology, cytology, phylogeography, and physiology*, B. J. Samuel, Ed., Balkema, Rotterdam, the Netherlands, 1982.
- [37] X. Chen and V. M. Helen, *Tulipa Linnaeus, Flora of China*, B. J. Samuel, Z. Y. Wu, and P. H. Raven, Eds., Science Press, Beijing, China, 2000.
- [38] T. Mine, K. Ozge, B. B. T. Metin et al., “Molecular phylogenetic analysis of *Tulipa* (Liliaceae) based on noncoding plastid and nuclear DNA sequences with an emphasis on Turkey,” *Botanical Journal of the Linnean Society*, vol. 172, no. 3, pp. 270–279, 2013.
- [39] A. Davoud, B. Alireza, R. N. Mohammad, and K. Mahmoud, “Biodiversity status of *Tulipa* (Liliaceae) in Iran inferred from molecular characterization,” *Horticulture, Environment, and Biotechnology*, vol. 61, no. 3, pp. 559–567, 2020.
- [40] P. Li, R. S. Lu, W. Q. Xu et al., “Comparative genomics and phylogenomics of east Asian Tulips (*Amana*, Liliaceae),” *Frontiers in Plant Science*, vol. 8, p. 451, 2017.
- [41] H. D. K. Do, C. Kim, M. W. Chase, and J. H. Kim, “Implications of plastome evolution in the true lilies (monocot order Liliales),” *Molecular Phylogenetics and Evolution*, vol. 148, p. 106818, 2020.
- [42] J. B. Yang, D. Z. Li, and H. T. Li, “Highly effective sequencing whole chloroplast genomes of angiosperms by nine novel universal primer pairs,” *Molecular Ecology Resources*, vol. 14, no. 5, pp. 1024–1031, 2014.
- [43] N. Dierckxsens, P. Mardulyn, and G. Smits, “NOVOPlasty: de novo assembly of organelle genomes from whole genome data,” *Nucleic Acids Research*, vol. 45, no. 4, p. gkw955, 2017.
- [44] X. J. Qu, M. J. Moore, D. Z. Li, and T. S. Yi, “PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes,” *Plant Methods*, vol. 15, no. 1, p. 50, 2019.
- [45] M. Kearse, R. Moir, A. Wilson et al., “Geneious basic: an integrated and extendable desktop software platform for the

- organization and analysis of sequence data," *Bioinformatics*, vol. 28, no. 12, pp. 1647–1649, 2012.
- [46] M. Lohse, O. Drechsel, S. Kahlau, and R. Bock, "Organellar-GenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets," *Nucleic Acids Research*, vol. 41, no. W1, pp. W575–W581, 2013.
- [47] A. Amiryousefi, J. Hyvönen, and P. Poczai, "Irscope: an online program to visualize the junction sites of chloroplast genomes," *Bioinformatics*, vol. 34, no. 17, pp. 3030–3031, 2018.
- [48] T. Thiel, W. Michalek, R. K. Varshney, and A. Graner, "Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.)," *Theoretical and Applied Genetics*, vol. 106, no. 3, pp. 411–422, 2003.
- [49] K. Katoh and D. M. Standley, "MAFFT multiple sequence alignment software version 7: improvements in performance and usability," *Molecular Biology and Evolution*, vol. 30, no. 4, pp. 772–780, 2013.
- [50] A. Stamatakis, "RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models," *Bioinformatics*, vol. 22, no. 21, pp. 2688–2690, 2006.
- [51] P. Librado and J. Rozas, "DnaSP v5: a software for comprehensive analysis of DNA polymorphism data," *Bioinformatics*, vol. 25, no. 11, pp. 1451–1452, 2009.
- [52] C. J. Chen, H. Chen, Y. Zhang et al., "TBtools: an integrative toolkit developed for interactive analyses of big biological data," *Molecular Plant*, vol. 13, no. 8, pp. 1194–1202, 2020.
- [53] R. S. Lu, P. Li, and Y. X. Qiu, "The complete chloroplast genomes of three *Cardiocrinum* (Liliaceae) species: comparative genomic and phylogenetic analyses," *Frontiers in Plant Science*, vol. 7, p. 2054, 2017.
- [54] L. A. Raubeson, R. Peery, T. W. Chumley et al., "Comparative chloroplast genomics: analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*," *BMC Genomics*, vol. 8, p. 174, 2007.
- [55] R. J. Wang, C. L. Cheng, C. C. Chang, C.-L. Wu, T.-M. Su, and S.-M. Chaw, "Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots," *BMC Evolutionary Biology*, vol. 8, no. 1, p. 36, 2008.
- [56] M. Yang, X. Zhang, G. Liu et al., "The complete chloroplast genome sequence of date palm (*Phoenix dactylifera* L.)," *PLoS One*, vol. 5, no. 9, p. e12762, 2010.
- [57] K. J. Kim and H. L. Lee, "Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants," *DNA Research*, vol. 11, no. 4, pp. 247–261, 2004.
- [58] M. Parks, R. Cronn, and A. Liston, "Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes," *BMC Biology*, vol. 7, no. 1, p. 84, 2009.
- [59] G. L. Wheeler, H. E. Dorman, A. T. Buchanan, L. Challagundla, and L. E. Wallace, "A review of the prevalence, utility, and caveats of using chloroplast simple sequence repeats for studies of plant biology," *Applications in Plant Sciences*, vol. 2, no. 12, p. 1400059, 2014.
- [60] D. Ebert and R. Peakall, "Chloroplast simple sequence repeats (cpSSRs): technical resources and recommendations for expanding cpSSR discovery and applications to a wide array of plant species," *Molecular Ecology Resources*, vol. 9, no. 3, pp. 673–690, 2009.
- [61] A. Yang, J. Zhang, X. Yao, and H. Huang, "Chloroplast microsatellite markers in *Liriodendron tulipifera* (Magnoliaceae) and cross-species amplification in *L. chinense*," *American Journal of Botany*, vol. 98, no. 5, pp. e123–E126, 2011.
- [62] W. Dong, C. Xu, D. Li et al., "Comparative analysis of the complete chloroplast genome sequences in psammophytic *Haloxylon* species (Amaranthaceae)," *Peer J*, vol. 4, no. 2, p. e2699, 2016.
- [63] S. Kaur, P. S. Panesar, M. B. Bera, and V. Kaur, "Simple sequence repeat markers in genetic divergence and marker-assisted selection of rice cultivars: a review," *Critical Reviews in Food and Nutrition*, vol. 55, no. 1, pp. 41–49, 2015.
- [64] Y. Yang, T. Zhou, D. Duan, J. Yang, L. Feng, and G. Zhao, "Comparative analysis of the complete chloroplast genomes of five *Quercus* species," *Frontiers in Plant Science*, vol. 7, p. 959, 2016.
- [65] W. Powell, M. Morgante, R. McDevitt, G. G. Vendramin, and J. A. Rafalski, "Polymorphic simple sequence repeat regions in chloroplast genomes—applications to the population genetics of pines," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 92, no. 17, pp. 7759–7763, 1995.
- [66] J. Provan, G. Corbett, J. W. McNicol, and W. Powell, "Chloroplast DNA variability in wild and cultivated rice (*Oryza* spp.) revealed by polymorphic chloroplast simple sequence repeats," *Genome*, vol. 40, no. 1, pp. 104–110, 1997.
- [67] M. Pauwels, X. Vekemans, C. Gode, H. Frérot, V. Castric, and P. Saumitou-Laprade, "Nuclear and chloroplast DNA phylogeography reveals vicariance among European populations of the model species for the study of metal tolerance, *Arabidopsis halleri* (Brassicaceae)," *New Phytologist*, vol. 193, no. 4, pp. 916–928, 2012.
- [68] M. T. Clegg, B. S. Gaut, G. H. Learn, and B. R. Morton, "Rates and patterns of chloroplast DNA evolution," *Proceedings of the National Academy of Sciences*, vol. 91, no. 15, pp. 6795–6801, 1994.
- [69] Q. Liu and Q. Xue, "Codon usage in the chloroplast genome of rice (*Oryza sativa* L. ssp. *japonica*)," *Acta Agronomica Sinica*, vol. 30, no. 12, pp. 1220–1224, 2004.
- [70] M. Zhou, W. Long, and X. Li, "Analysis of synonymous codon usage in chloroplast genome of *Populus alba*," *Journal of Forestry Research*, vol. 19, no. 4, pp. 293–297, 2008.
- [71] S. Tangphatsornruang, D. Sangsrakru, J. Chanprasert et al., "The chloroplast genome sequence of mungbean (*Vigna radiata*) determined by high-throughput pyrosequencing: structural organization and phylogenetic relationships," *DNA Research*, vol. 17, no. 1, pp. 11–22, 2010.
- [72] Y. Du, Y. Bi, X. Chen, F. Yang, J. Xue, and X. Zhang, "The complete chloroplast genome of *Lilium cernuum*: genome structure and evolution," *Conservation Genetics Resources*, vol. 8, no. 4, pp. 375–378, 2016.
- [73] B. R. Morton, "Selection on the codon bias of chloroplast and cyanelle genes in different plant and algal lineages," *Journal of Molecular Evolution*, vol. 46, no. 4, pp. 449–459, 1998.
- [74] J. Meng, X. Li, H. T. Li, J. Yang, H. Wang, and J. He, "Comparative analysis of the complete chloroplast genomes of four *Aconitum* medicinal species," *Molecules*, vol. 23, no. 5, p. 1015, 2018.
- [75] C. L. Liang, L. Wang, J. Le et al., "A comparative analysis of the chloroplast genomes of four *Salvia* medicinal plants," *Engineering*, vol. 5, no. 5, pp. 907–915, 2019.