Identification of evolutionary relationships and DNA markers in the medicinally important genus *Fritillaria* based on chloroplast genomics (#60657)

First submission

Guidance from your Editor

Please submit by **6 Jun 2021** for the benefit of the authors (and your $200 publishing discount) .

Structure and Criteria
Please read the 'Structure and Criteria' page for general guidance.

Custom checks
Make sure you include the custom checks shown below, in your review.

Author notes
Have you read the author notes on the guidance page?

Raw data check
Review the raw data.

Image check
Check that figures and images have not been inappropriately manipulated.

Privacy reminder: If uploading an annotated PDF, remove identifiable information to remain anonymous.

Files

Download and review all files from the materials page.

13 Figure file(s)
2 Table file(s)
1 Raw data file(s)

Custom checks

DNA data checks

- Have you checked the authors data deposition statement?
- Can you access the deposited data?
- Has the data been deposited correctly?
- Is the deposition information noted in the manuscript?

Structure and Criteria

Structure your review

The review form is divided into 5 sections. Please consider these when composing your review:

1. **BASIC REPORTING**
2. **EXPERIMENTAL DESIGN**
3. **VALIDITY OF THE FINDINGS**
4. General comments
5. Confidential notes to the editor

You can also annotate this PDF and upload it as part of your review

When ready submit online.

Editorial Criteria

Use these criteria points to structure your review. The full detailed editorial criteria is on your guidance page.

BASIC REPORTING

- Clear, unambiguous, professional English language used throughout.
- Intro & background to show context. Literature well referenced & relevant.
- Structure conforms to PeerJ standards, discipline norm, or improved for clarity.
- Figures are relevant, high quality, well labelled & described.
- Raw data supplied (see PeerJ policy).

EXPERIMENTAL DESIGN

- Original primary research within Scope of the journal.
- Research question well defined, relevant & meaningful. It is stated how the research fills an identified knowledge gap.
- Rigorous investigation performed to a high technical & ethical standard.
- Methods described with sufficient detail & information to replicate.

VALIDITY OF THE FINDINGS

- Impact and novelty not assessed. *Meaningful* replication encouraged where rationale & benefit to literature is clearly stated.
- All underlying data have been provided; they are robust, statistically sound, & controlled.
- Speculation is welcome, but should be identified as such.
- Conclusions are well stated, linked to original research question & limited to supporting results.

Standout reviewing tips

The best reviewers use these techniques

Tip	*Example*
Support criticisms with evidence from the text or from other sources	*Smith et al (J of Methodology, 2005, V3, pp 123) have shown that the analysis you use in Lines 241-250 is not the most appropriate for this situation. Please explain why you used this method.*
Give specific suggestions on how to improve the manuscript	*Your introduction needs more detail. I suggest that you improve the description at lines 57- 86 to provide more justification for your study (specifically, you should expand upon the knowledge gap being filled).*
Comment on language and grammar issues	*The English language should be improved to ensure that an international audience can clearly understand your text. Some examples where the language could be improved include lines 23, 77, 121, 128 – the current phrasing makes comprehension difficult. I suggest you have a colleague who is proficient in English and familiar with the subject matter review your manuscript, or contact a professional editing service.*
Organize by importance of the issues, and number your points	*1. Your most important issue* *2. The next most important item* *3. …* *4. The least important points*
Please provide constructive criticism, and avoid personal opinions	*I thank you for providing the raw data, however your supplemental files need more descriptive metadata identifiers to be useful to future readers. Although your results are compelling, the data analysis should be improved in the following ways: AA, BB, CC*
Comment on strengths (as well as weaknesses) of the manuscript	*I commend the authors for their extensive data set, compiled over many years of detailed fieldwork. In addition, the manuscript is clearly written in professional, unambiguous language. If there is a weakness, it is in the statistical analysis (as I have noted above) which should be improved upon before Acceptance.*

Identification of evolutionary relationships and DNA markers in the medicinally important genus *Fritillaria* based on chloroplast genomics

Tian Zhang [1], **Sipei Huang** [1], **Simin Song** [1], **Meng Zou** [1], **Tiechui Yang** [2], **Weiwei Wang** [1], **jiayu zhou** [Corresp., 1], **Hai Liao** [Corresp. 1]

[1] Southwest Jiaotong University, Chengdu, China

[2] Qinghai lvkang Biological Development Co., Ltd, Xining, China

Corresponding Authors: jiayu zhou, Hai Liao
Email address: spinezhou@home.swjtu.edu.cn, ddliaohai@home.swjtu.edu.cn

Fritillaria genus has attracted attention because of its medicinal and ornamental values. At least three reasons, including the accurate discrimination between various *Fritillaria* species, protection and sustainable development of rare *Fritillaria* resources as well as understanding of relationship of some perplexing species, have prompted to the necessary phylogenetic analysis and development of molecular markers of *Fritillaria* species. Here we obtained the complete chloroplast (CP) genomes of *F. unibracteata*, *F. przewalskii*, *F. delavayi* and *F. sinica* through Illumina sequencing followed by *de novo* assembly. The lengths of the genomes ranged from 151,076 in *F. unibracteata*) to 152,043 in *F. przewalskii*. Those CP genomes displayed a typical quadripartite structure, all including a pair of inverted repeats (26,078-26355 bp) separated by the large single-copy (81,383-81,804 bp) and small single-copy (17537-17569 bp) regions. *F. przewalskii*, *F. delavayi* and *F. sinica* equivalently encodes 133 unique genes consisting of 38 transfer RNA genes, 8 ribosomal RNA genes and 87 protein coding genes, whereas *F. unibracteata* contained 132 unique genes due to lack of *rps16* gene. Subsequently, comparative analysis of the complete CP genomes revealed that *ycf1*, *trnL*, *trnF*, *ndhD*, *trnN-trnR*, *trnE-trnT*, *trnN*, *psbM-trnD*, *atpI* and *rps19* might be used as molecular markers in taxonomic study due to their significant difference. Additionally, almost all the plastid protein coding genes were found to prefer ending with A/T. Based on the comprehensive CP genome data that were collected from 53 species in *Fritillaria* and *Lilium* genera, a phylogenomic study was carried out with 3 species in *Cardiocrinum* genus and 5 species in *Amana* genus as outgroups. *Fritillaria* genus and *Lilium* genus showed the closest relationship with a high support value, and the interspecific resolution within subgenus *Fritillaria* were much better than those of the phylogenetic trees based on the separate regions, including *matK*, *psbA-trnH* and *rpl16*. The geographical distribution pattern of the 11 medicinal species neatly

mapped on the phylogenetic relationship based on CP genomes. Furthermore, phylogenetic analysis based on CP genome was promising method to select potential novel medicinal resources to substitute current medicinal species that were on the verge of extinction. More importantly, the specie-specific molecular identification for *F. taipaiensis*, *F. unibracteata* and *F. cirrhosa*, were successfully developed, respectively.

Identification of evolutionary relationships and DNA markers in the medicinally important genus *Fritillaria* based on chloroplast genomics

Tian Zhang[1], Sipei Huang[1], Simin Song[1], Meng Zou[1], Tiechui Yang[2], Weiwei Wang[1], Jiayu Zhou[1,*], Hai Liao[1,*]

[1]*School of Life Science and Engineering, Southwest Jiaotong University, Chengdu, Sichuan, 610031, China*

[2]*Qinghai lvkang Biological Development Co., Ltd., Xining 810003, China*

Authors' email address:

Tian Zhang: 1215695297@qq.com; Sipei Huang: 691786509@qq.com; Simin Song: 1158216140@qq.com; Meng Zou: 815112542@qq.com; Tiechui Yang: 298237664@qq.com; Weiwei Wang: 1028694049@qq.com; Jiayu Zhou: spinezhou@home.swjtu.edu.cn; Hai Liao: ddliaohai@home.swjtu.edu.cn.

* For Correspondence

*Corresponding authors' information

Jiayu Zhou: spinezhou@home.swjtu.edu.cn;

Hai Liao: ddliaohai@home.swjtu.edu.cn

Abstract: *Fritillaria* genus has attracted attention because of its medicinal and ornamental values. At least three reasons, including the accurate discrimination between various *Fritillaria* species, protection and sustainable development of rare *Fritillaria* resources as well as understanding of relationship of some perplexing species, have prompted to the necessary phylogenetic analysis and development of molecular markers of *Fritillaria* species. Here we obtained the complete chloroplast (CP) genomes of *F. unibracteata*, *F. przewalskii*, *F. delavayi* and *F. sinica* through Illumina sequencing followed by *de novo* assembly. The lengths of the genomes ranged from 151,076 in *F. unibracteata*) to 152,043 in *F. przewalskii*. Those CP genomes displayed a typical quadripartite structure, all including a pair of inverted repeats (26,078-26355 bp) separated by the large single-copy (81,383-81,804 bp) and small single-copy (17537-17569 bp) regions. *F. przewalskii*, *F. delavayi* and *F. sinica* equivalently encodes 133 unique genes consisting of 38 transfer RNA genes, 8 ribosomal RNA genes and 87 protein coding genes, whereas *F. unibracteata* contained 132 unique genes due to lack of *rps16* gene. Subsequently, comparative analysis of the complete CP genomes revealed that *ycf1*, *trnL*, *trnF*, *ndhD*, *trnN-trnR*, *trnE-trnT*, *trnN*, *psbM-trnD*, *atpI* and *rps19* might be used as molecular markers in taxonomic study due to their significant difference. Additionally, almost all the plastid protein coding genes were found to prefer ending with A/T. Based on the comprehensive CP genome data that were collected from 53 species in *Fritillaria* and *Lilium* genera, a phylogenomic study was carried out with 3 species in *Cardiocrinum* genus and 5 species in *Amana* genus as outgroups. *Fritillaria* genus and *Lilium* genus showed the closest relationship with a high support value, and the interspecific resolution within subgenus *Fritillaria* were much better than those of the phylogenetic trees based on the separate regions, including *matK*, *psbA-trnH* and *rpl16*. The geographical distribution pattern of the 11 medicinal species neatly mapped on the phylogenetic relationship based on CP genomes. Furthermore, phylogenetic analysis based on CP genome was promising method to select potential novel medicinal resources to substitute current medicinal species that were on the verge of extinction. More importantly, the specie-specific molecular identification for *F. taipaiensis*, *F. unibracteata* and *F. cirrhosa*, were successfully developed, respectively.

Key words: Chloroplast genomics; *Fritillaria*; InDel; Phylogenetic relationship; Specie-specific identification

Abbreviations: CP, chloroplast; IR, inverted repeat; ITS, internal transcribed spacer; LSC, large single copy; RSCU, relative synonymous codon usage; SSC, small single copy; SSR, simple sequence repeats; TCM, traditional Chinese medicine

Introduction

The genus *Fritillaria* (*Liliaceae*), consisting of 140 known species, is widely distributed in Europe, Asia and North America (Huang et al 2018; Rix et al 2001). Based on the Flora of China, twenty-two species are distributed throughout most provinces in China, among which four diversity hotspots are constituted. *Fritillaria* species have attracted much attention because they are mostly used as medicinal materials in traditional Chinese medicine (TCM) and partly used as ornamental plants. The dried bulbs of some *Fritillaria* species, called as Bei-mu, were firstly introduced in "Book of Songs", the earliest masterpiece on Chinese classical realism, in 7th century BC. About 2000 years ago, the medicinal value of Bei-mu was introduced in "Sheng Nong's herbal classic", the earliest book on Chinese herbal medicine in the world, for the first time. The bulbs were known to contain pharmaceutical active steroidal alkaloids, and were used as medicine and food materials in many TCM prescriptions. In China, the production of medicinal preparations containing Bei-mu is a huge industry with an estimated value of more than 400 million dollars per year (Day et al 2014). Totally, it was reported by the Chinese Pharmacopeia 2020 that the bulbs from 11 species used in TCM were divided into five traditional Chinese medicines, including Chuan-Bei-mu, Yi-Bei-mu, Zhe-Bei-mu, Ping-Bei-mu and Hubei-Bei-mu. The bulbs of *F. cirrhosa*, *F. unibracteata*, *F. przewalskii*, *F. delavayi*, *F. taipaiensis* and *F. wabuensis*, all included in the complex group of *F. cirrhosa*, were called as Chuan-Bei-mu. The bulbs of *F. pallidiflora* and *F. walujewii*, which were from Xinjiang plain, were used as Yi-Bei-mu. The bulbs of *F. thunbergii*, *F. ussuriensis* and *F. hupehensis* were used as Zhe-Bei-Mu, Ping-Bei-mu and Hubei-Bei-mu, respectively. Each medicine has its own unique efficacy and bioactive compounds, and should be used separately for given clinical purposes in traditional prescription. For instance, Chuan-Bei-mu has been applied to treat cough caused by deficiency of the lung, asthenia of the viscera, and tidal fever. Ping-Bei-mu has been mainly used to treat cough due to exogenous dryness and deficiency of Yin (a kind of Meridians) (Park et al 2017). Zhe-Bei-mu was used to treat cough due to exogenous wind-heat or phlegm-fire stagnation. Additionally, it was reported that the various original species of Chuan-Bei-mu had different degrees of relieving cough and phlegm. However, although the bulbs of *Fritillaria* showed some degrees of difference in efficiency, various *Fritillaria* species were still used indiscriminately in clinical prescription due to their similar morphology and names. Furthermore, in recent years, the market price of Chuan-Bei-mu has increased considerably because of its scare wild resources and limited output from wild habitats, and hence unripe bulbs of other *Fritillaria* species have been added as adulterants imitating the original Chuan-Bei-mu. It was serious that the mixed use of various *Fritillaria* in medicine and food might cause quality uncertainty and safety risks, thus the importance of taxonomy identification on various *Fritillaria* species was highlighted.

The morphological traits of *Fritillaria* species, particularly the complex group of *F. cirrhosa* that were widely distributed in Hengduan mountains in China, were complex because of several highly variable characters including stem length, petal color, capsule winged or not, leaf curling; and scale number (Luo et al 1996). For instance, *F. cirrhosa* has stem length of less than 60 cm, yellow or yellowish green petals with 3 bracts per flower, curling leaf, narrowly winged capsules, and bulb with 2-3 scales. However, the mechanism

of the variation remains elusive and the morphologic traits, affected by ecological stress factors, thus are subjected to phenomena of parallel evolution. Nowadays, DNA-based classification occurred and has been applied in angiosperm phylogeny group because of their reliability and readable data (Yang et al 2016). Therefore, first of all, accurate identification (eg. using DNA markers) of *Fritillaria* species, especially the complex group of *F. cirrhosa*, was necessary to distinguish genuine medicinal plants and the medicines derived from them. Secondly, since the bulbs of some *Fritillaria* species showed great economic value in Asian countries and have long been used in TCM, the wild *Fritillaria* populations decreased sharply due to long-term excessive harvesting. To date, four species of Chuan-Bei-mu and eight species in Xinjiang plain have been classified as rare resources based on the list of rare endangered higher plants of China (Li et al 2018). DNA markers was helpful to understand accurately genetic diversity and structure of *Fritillaria* population, and thus provided scientific approach for conservative requirement. Thirdly, a better understanding of the relationships within the genus could be great significance for the medicinal use of *Fritillaria*. Identifying the closest relatives could point to additional species that might be analyzed for their potential medicinal value, which might in turn reduce pressure on those species that were currently faced with survival risk. . Finally, phylogenetic position of some medicinal species in *Fritillaria*, such as *F. pallidiflora*, *F. wabuensis* and *F. davidii*, remained elusive. *F. pallidiflora* was always considered a member of subgenus *Fritillaria* by Rix (2001), whereas Rønsted et al (2005) linked it to subgenus *Petillium* based on the molecular and morphological comparison. *F. wabuensis* was firstly discovered and nominated as a new species in *Fritillaria* genus (Tang and Yue 1983), while it was classified as variant of *F. crassicaulis* (Luo et al 1996) and *F. unibracteata* (Liu et al 2009), respectively. Therefore, more resolved molecular phylogenetic studies of *Fritillaria*, especially those medicinal species, were necessary to carry out for advanced understanding of relationships within this genus.

Currently, the genus *Fritillaria* was divided into eight subgenera, including, *Liliorhiza* (including species mainly in North America), *Japonica* (including species mainly in Japan), *Fritillaria* (the biggest subgenera), *Rhinopetalum*, *Petilium*, and the monotypic *Davidii* (including only *F. davidii*), *Theresia* (only *F. persica*) and *Korolkowia* (only F. sewerezowi), by Rix (2001). At present, despite the frequent usage of nuclear DNA internal transcribed spacer (ITS) and several plastid genome regions (*trnL-trnF*, *matK*, *rbcL* and *rpl16*) in the classification of this genus, previous studies have found that these markers merely provided partial phylogenetic signal. In detail, Rønsted et al (2005), who primarily established the current understanding of evolutionary relationships with *Fritillaria*, investigated the phylogenetic position of 37 *Fritillaria* species using *matK*, *rpl16* intron and ITS. Consequently, *Fritillaria* genus formed two clades, in which one clade mainly included species from the North American subgenus *Liliorhiza* and the other clade comprised species from the seven remaining subgenera. In common with the result of Rønsted et al (2005), Khourang et al (2014) revealed that the subgenus *Fritillaria* was sister to subgenus *Rhinopetalum* on the basis of the phylogenetic tree of nine Iran species using the ITS and *trnL-trnF* regions. . However, Day et al (2014) supported that the largest subgenus (subgenus *Fritillaria*) appeared to be polyphyletic and formed two clades with *matK* and *rbcL*

sequences, in which one clade comprised taxa occurring mainly in Europe, the Middle East, Japan and North Africa, and the other clade comprised taxa distributing in China and Central Asia. In our previous research, various *Fritillaria* species from China were classified as North-China group and South-China group based on nrITS2 sequences, but 57.1 % species were not effectively resolved (Zheng et al 2019). Recently, Li et al (2014) presented high-quality chloroplast genome using single molecule real-time sequencing, and suggested that *rps19* gene varied greatest among various species. However, the noncoding regions showed higher extent of variability during evolution and were postulated to search candidate molecular markers. Therefore, it was proposed that genomics based on the entire chloroplast genome sequence might provide more sufficient phylogenetic signals and identify molecular markers with higher resolution (Xue et al 2019).

Chloroplasts play an essential role in sustaining life on earth (Watson et al 2018). In higher plants, the chloroplast (CP) genome ranges from 120 to 180 kb, with a conserved quadripartite structure consisting of two copies of a larger inverted repeat (IR), a large single copy (LSC) region and a small single copy (SSC) region (Guo et al 2017; Wicke et al 2011). The CP genome contains up to approximately 80 unique protein-coding genes, four ribosomal RNAs (rRNAs), and 30 transfer RNAs (tRNAs). To date, more than 800 sequenced chloroplast genomes from a variety of land plants have been deposited in the GenBank database (Daniell et al 2016). Recently, chloroplast genome assembly based on next-generation sequencing (NGS) technology has become more affordable and easier compared to the Sanger method. The CP genome has been extensively used for understanding phylogenetic relationships and discovering more effective molecular markers, some of which, such as *trnH-psbA*, *matK* and *rpl16*, were used as universal plant DNA barcodes (Bansala et al 2018; Vinnersten and Bremer 2001) To date, the availability of 23 *Fritillaria* CP genomes in GenBank enhanced our understanding of phylogenetic relationship and identification of molecular markers. Although previous reports (Chen et al 2019; Huang et al 2020; Park et al 2017) performed comparative analyses with *Fritillaria* CP genomes available on GenBank, specie-specific identification has not been reported and the phylogenetic place of some ambiguous species remained elusive. In the present study, the CP genomes of four *Fritillaria* species were obtained using the Illumina platform. The objectives of this study included (1) analyzing the global structural patterns of the four CP genomes and comparing them with the available 23 CP genomes of *Fritillaria*; (2) assessing the phylogenetic relationships of the 11 medicinal species used in TCM, by which to understand the phylogenetic position of some ambiguous species and find potential medicinal plants; and (3) discovering and verifying highly divergent DNA markers for specie-specific identification of *Fritillaria* species and for population genetics. To the best of our knowledge, we are the first to develop specie-specific molecular identification of *Fritillaria* species.

Materials and methods

Plant material

The fresh leaves of *F. unibracteata*, *F. przewalskii*, *F. delavayi* and *F. sinica* were collected from the Huzhu County (36°50′15″N, 101°57′06″E), Xining city, Qinghai Province, respectively. All samples were

immediately frozen in liquid nitrogen and stored at -80 °C until DNA extraction.

Chloroplast genome sequencing and assembly

Total genomic DNA was isolated from 100 mg of fresh leaves using a modified CTAB method. The DNA concentration (>50 ng μL-1) was measured using a NanoDrop spectrophotometer. The isolated DNA was fragmented into small pieces using sonication. After end reparation and A-tailing, the short DNA fragments were ligated with the Illumina paired-end adaptors. Based on gel electrophoresis, the suitable fragments were purified and selected as templates for next-step PCR amplification, so as to create the final DNA library. The quality and quantity of the DNA library were measured using the Agilent 2100 Bioanalyzer. Finally, the library was sequenced from both the 5′ and 3′ ends using Illumina NovaSeq6000 PE150 Sequencing platform (Illumina, CA, USA). By use of NGSQCToolkit v2.3.3, the raw reads were filtered to remove the linker sequence and low-quality reads defined as having more than 10% bases with Q-value <20, and thus high-quality clean reads were obtained. The clean reads were then assembled using SPAdes (Bankevich et al 2012) 3.10.1 (http://cab.spbu.ru/software/spades/) software with CP genome of *F. cirrhosa* as reference (NCBI accession number NC_024728.1). Finally, LSC/IR and SSC/IR junctions were further verified by Sanger sequencing. .

Genome annotation and sequence alignment

In order to predict putative gene function, the CDS, rRNA and tRNA genes were aligned using blast v2.2.25 (https://blast.ncbi.nlm.nih.gov/Blast.cgi), HMMER v3.1b2 (http://www.hmmer.org/) and aragorn v1.2.38 (http://130.235.244.92/ARAGORN/), respectively, with *E*-value of 10^{-5}. The OGDRAW (https://chlorobox.mpimp-golm.mpg.de/OGDraw.html) helped to make the chloroplast genome maps of *F. unibracteata*, *F. przewalskii*, *F. delavayi*, and *F. sinica*. The vmatch v2.3.0 (http: //www.vmatch.de/) could identify their scattered repetitive sequences. MISA v1.0 (MIcroSAtellite identification tool, http://pgrc.ipk-gatersleben.de/misa/misa.html) helped to analyze CPSSR. We also used CodonW to analyze codon usage bias. The mafft v7.310 was used to perform and InDel identification.

Chloroplast genome analysis by sliding window

After using the mafft to align the chloroplast genome sequences, BioEdit software was used to adjust the sequences manually. DanSP v6.0 was used to perform sliding window analysis (step size =200 bp and window length=600 bp) for nucleotide variability (Pi) in the whole chloroplast genome

Phylogenetic analysis

Firstly, the phylogenetic analysis was performed based on *matK*, *psbA-trnH* and *rpl16*, respectively, by use of Neighbor Joining and Maximum Likelihood methods. Then, the chloroplast genomes in the phylogenetic analysis included the 27 species of *Fritillaria*, 26 species of *Lilium*, 3 species of *Cardiocrinum* and 5 species of *Amana*. The chloroplast genome evolutionary tree was constructed by BLAST2OGMSA script (https://github.com/fenghen360/BLAST2OGMSA) and MEGA-X software. Firstly, multi-sequence alignment was conducted using BLAST tool of NCBI. Then, the initial alignment result was extracted by

BLAST2OGMSA script to obtain homologous blocks. It was reported that BLAST2OGMSA relied on ProgressiveMauve, a kind of anchored alignment algorithm, to determine where locally collinear blocks (LCBs) represented the landmarks among organelle genomes (such as chloroplast and mitochondrial genomes). The co-exist LCBs among all organelle genomes were extracted and prepared for the further phylogenetic tree construction. In this study, the conserved CDS genes, functional non-coding regions, and rRNA genes as well were combined by BLAST2OGMSA. Finally, the alignment data from BLAST2OGMSA was imported into MEGA-X software to construct the phylogenetic tree using the Neighbor-Joining method and the Maximum likelihood method respectively.

Specie-specific identification of *Fritillaria* species

In an attempt to identify InDels unique to various *Fritillaria* species, the Inserting/deleting regions were compared between genomes using the program Mafft using default settings. Compared with other *Fritillaria* species, a unique InDel with 137 bp deletion was found within intergenic space region, *accD-psaI*, of *F. taipaiensis* CP genome, making it a suitable target for developing specie-specific test for *F. taipaiensis*. The test for *F. taipaiensis* was performed by routine PCR using 5'-GCG AAC GAG TAT TTA GTT CAT C -3' as former primer and 5'-AGG GTT CTT TCA CTC CTT TCT -3' as reverse primer. The routine PCR were performed under the following conditions (Table 1). All samples were performed in triplicate.

Similarly, two InDels with 47 bp insertion and 6 bp deletion were also found within *trnG-GCC-trnR-UCU* and intron of *atpF* of *F. unibracteata* and *F. cirrhosa* CP genome respectively, were thus selected for development of real time PCR based marker. The specie-specific tests for *F. unibracteata* and *F. cirrhosa* were performed by Taqman MGB real time PCR, respectively. The former primer (5'-GCT ACC CGC TTA ATA CAT AC-3'), reverse primer (5'-CCG GAA CAG ATC GAA CAG -3') and the probe (5'-FAM-CCA TTG TCT AAT GGA AAA GA-MGB-3') were used for identification of *F. unibracteata*. The former primer (5'-GCT ACC CGC TTA ATA CAT AC-3'), reverse primer (5'-CCG GAA CAG ATC GAA CAG -3') and the probe (5'-FAM-CCA TTG TCT AAT GGA AAA GA-MGB-3') were used for specie-specific identification of *F. cirrhosa*. The TaqMan MGB real-time PCR was performed using 2×T5 Fast qPCR Mix (Qingke, China) with LightCycler 96 real-time fluorescence PCR instrument system (Roche) under the following conditions (Table 1). All samples were performed in triplicate.

Table 1

Results

Genome sequencing, assembly, and genome features

Based on a stringent quality control, a total of 23,755,399 to 26,831,529 paired-end reads were obtained, generating 7,126,619,700 to 8,049,458,700 clean bases data, from the four *Fritillaria* species. The resultant clean paired-end reads were then employed to assemble the chloroplast genome using CP genome of *F.*

cirrhosa as the reference. Totally, 471,385 to 652,632 mapping reads yielded average coverage of 934X to 1292X for each species, generating four full-length CP genomes that ranged from 151,076 in *F. unibracteata* to 152,043 in *F. przewalskii*. The CP genome contained identical structure including two IR regions (26,078-26355 bp each), which were separated by one LSC region (81,383-81,804 bp) and one SSC region (17537-17569 bp) (Fig 1 and Table 2).

A total of 133 genes were annotated, including 87 protein-coding genes (PCGs), 38 tRNA and 8 rRNA genes. The global gene order and content were identical in the four species, except *F. unibracteata* that was lack of *rps16* gene. 21 genes were duplicated in the CP genome, including 8 tRNA genes (*tRNA-ACG*, *tRNA-CAA*, *tRNA-CAU*, *tRNA-GAC*, *tRNA-GAU*, *tRNA-GUG*, *tRNA-GUU* and *tRNA-UGC*), 4 rRNA genes (*rrn16*, *rrn23*, *rrn4.5* and *rrn5*) and 9 PCGs (*rpl2*, *rpl23*, *rps7*, *rps12*, *ycf1*, *ycf2* and *ycf15*). There were 13 genes containing intron, among which *clpP* and *ycf3* had two introns, whereas the other 13 genes had one intron, including 6 tRNA genes (*tRNA-AAU*, *tRNA-CGA*, *tRNA-GAU*, *tRNA-UAA*, *tRNA-UGC* and *tRNA-UUU*) and 5 PCGs (*atpF*, *ndhA*, *ndhB*, *rpl2* and *rpoC1*). 8 introns (*tRNA-UUU, tRNA-CGA, atpF, rpoC1, tRNA-UAA, tRNA-AAU, ycf3 and clpP*) were located in the LSC region, one intron (*ndhA*) was located in the SSC region, four introns (*rpl2, ndhB, tRNA-GAU* and *tRNA-UGC*) were located in the IRa region, and two introns (*tRNA-UGC* and *tRNA-GAU*) were located in the IRb region (*rpl2* and *ndhB* were duplicated genes) (Table 3 and Table 4). Table 4 listed the 15 intron-containing genes in the chloroplast genome of *F. unibracteata*, and those of *F. przewalskii*, *F. delavayi*, and *F. sinica* were included in supplementary table1 2 and 3, respectively.

Contraction and expansion of IR regions, especially the boundary region, are important aspects of chloroplast genomes, which are the main reason of length variation in these genomes (Abdullah et al 2020). These 12 species have the same gene content and array in IR region, which is expanded in *rps19* and *ycf1* genes. The *rps19* gene in the 12 *Fritillaria* species crossed the LSC/IRa boundary and showed the same length of 279 bp which was similar to that of *Lilium superbum*, except that *F. cirrhosa* had *rps19* gene of 285 bp. In the LSC region, the length of *rps19* genes ranged from 250 to 268bp, whereas that of *rps19* genes in the IRa region varied from 11 to 35 bp. Besides, the *rps19* genes lost their protein-coding function due to incomplete gene duplication. The similar event was also observed in the *ycf1* genes at the IRa/SSC border. The *ycf1* genes were largely located in the IRa and extended 17-32 bp into the SSC region, whereas the *ycf1* gene in *F. taipaiensis* was fully located in the IRa region, 58 bp from the IRa/SSC boundary. In the SSC/IRb boundary of 12 species of *Fritillaria* subgenus, *ycf1* was a key gene and almost equally distributed (Fig 2). *Ycf1* has a SSC region of 4320 bp in *F. unibracteata* and *F. przewalskii*, but 4314 bp in *F. delavayi*, *F. sinica*, *F. cirrhosa* and *F. taipaiensis*, and also has an IRb region of 1230bp in all species. By comparing the LSC/IRa, IRa/SSC and SSC/IRb regions, it was found that there were significant length variation in SSC and IRa regions between the 12 *Fritillaria* species.

Table 2

Table 3

Table 4

Fig. 1

Fig. 2

Repeat sequence, simple sequence repeats (SSRs) and codon analysis

We used REPuter software to identify a large number of repeat sequences in the CP genome of *Fritillaria* (Table 5). The length of the repeat sequence distributed mainly from 15 to 20 bp and rarely from 21 to 38 bp among four *Fritillaria* species. The repeating sequences were divided into forward repeating and palindrome sequences (including reverse and complementary sequences). The number of repeating sequences from 15 to 20 bp of *F. unibracteata* and *F. przewalskii* were more than 487, while those of *F. delavayi* and *F. sinica* were less than 350, respectively (Fig 3). The number of repeat sequences in *F. przewalskii*, *F. unibracteata*, *F. sinica* and *F. delavayi* were 1200, 976, 656 and 425, respectively. Although repeat sequences from 21 to 38 bp were rare, several promising molecular markers were found. For instance, *F. unibracteata* had three forward repeating in repeat sequence at length of 23, 30 and 47 bp, respectively. *F. delavayi* also contained a palindrome in repeat sequence at length of 54 bp, and *F. przewalskii* contains two forward repeating sequences and a palindrome in repeat sequence at length of 23 bp.

Table 5

Fig. 3

Using MISA software, we also found 77, 76, 75 and 72SSRs of at least 10 bp in *F. przewalskii*, *F. sinica*, *F. unibracteata* and *F. delavayi,* respectively (Table 5, Fig 4). These SSRs were mainly located in the LSC region, followed by 50 SSRs in IR region, and a few SSRs in the SSC region. The single- and three-nucleotide SSRs were the majority detected in these *Fritillaria* species, the double-and four-nucleotide SSRs were the

minority detected and a few were five-nucleotide. Single- together with three-nucleotide repeats in *F. unibracteata*, *F. przewalskii*, *F. delavayi* and *F. sinica* accounted for 81.33%, 83.12%, 79.17%, and 81.58% of SSRs, respectively. The single-nucleotide SSR with eight to nine repeated units were the most abundant and accounted for 53.91% (Fig 4). The high variation of SSRs might provide abundant information for molecular marker studies and plant breeding.

Fig. 4

In addition to the standard initiator codon AUG, noncanonical start codons, such as GTG and TTG, have been detected in the CP genome of four *Fritillaria* species. Similar noncanonical start codons have been also found in *Paris Sect.Marmorata* (Gao et al 2018). The relative synonymous codon usage (RSCU) and bias analyses were carried out using the software CodonW. The protein-coding sequences of the *F. unibracteata*, *F. przewalskii*, *F. delavayi* and *F. sinica* chloroplast genomes consisted of 26,540, 26,591, 26,607 and 26,587 codons, respectively. Among them, AUG, UUA and AGA, encoding Met, Leu and Arg residues, respectively, were identified as the top 3 preferable codons according to RSCU values. As observed in CP genomes of most land plants, codon usage patterns of this section were likely driven by the bias towards the high A/T content (Kim and Lee 2004; Wolfe et al 1987) (Fig 5).

Fig. 5

Comparative Genomic Analysis and Divergence Hotspot Regions

A genome wide alignment revealed four *Fritillaria* species had high sequence similarity (>90% identity). IR regions showed a lower level of sequence divergence than LSC and SSC regions. Using slide window analysis, 18 regions were eventually extracted to calculate the nucleotide variability with Pi value ranging from 0.0104 (*rpl12*) to 0.0159 (*ycf1*). 10 most divergent hotspots were identified and thus might be utilized as potential molecular markers for future phylogenetic and phylogeographic analyses as well as species identification of genus *Fritillaria*. These hotspots included *ycf1*, *trnL*, *trnF*, *ndhD*, *trnN-trnR*, *trnE-trnT*, *trnN*, *psbM-trnD*, *atpI* and *rps19*. In addition, the other eight regions, such as *petN-psbM*, *trnT-psbD*, *psbD*, *trnG-trnG*，*rps18-rpl20* and *rpl2*, also constituted potential candidates (Fig 6).

Fig. 6

Phylogenetic tree on the basis of *matK*, *psbA-trnH* and *rpl16*, respectively

The result of ML tree was a bit better than that of NJ tree (Supplementary Fig 1), and thus the phylogenetic analyses were performed using Maximum Likelihood method. The *matK* matrix included 27 species of *Fritillaria*, 26 of *Lilium*, 5 of *Amana* and 3 of *Cardiocrinum*. On average 1524 bp of the *matK* region were collected from NCBI. In the ML tree, *Lilium* genus formed paraphyly and *Fritillaria* formed monophyly although the topology was only weakly supported (38 BP). In the *Fritillaria* genus, *F. maximowiczii* was sister to *F. wabuensis* (64 BP). *F. davidii* was sister to the remaining *Fritillaria* species (73 BP). *F. karelinii, F. ussuriensis* and *F. meleagroide* form a small monophyly with weak support (67 BP). The other species in the largest subgenus *Fritillaria* (including *F. eduardii*, a member of subgenus *Petilium*) formed a clade despite the poor support (10 BP). Totally, 17 out of 59 branches have bootstrap value of more than 70 BP (Fig 7A).

The *psbA-trnH* matrix included 28 species of *Frtitillaria*, 25 of *Lilium* and 5 of outgroup. On average, 309 bp of *psbA-trnH* was aligned. The *Lilium* and outgroup (*Cardiocrinum* and *Amana* genus) were polyphyletic and mixed together. Except *F. maximowiczii*, the remaining species in *Fritillaria* formed a moderately supported (78 BP) clade, which was composed of two subclades (40 BP and 23 BP, respectively). In total, only 9 out of 56 branches obtained bootstrap values of more than 70 BP (Fig 7B).

In *rpl16* matrix, 26 species of *Fritillaria*, 25 of *Lilium* and 8 of outgroup were included. On average, 1395 bp of *rpl16* was aligned. Except *F. wabuensis*, the other members of *Fritillaria* and 3 members of *Cardiocrinum* form a monophyletic clade but with poor support (7 BP). In the main clade of *Fritillaria*, four species from Huadong plain, including *F.hupehensis*, etc, form a weak clade (53 BP). 5 species from Xinjiang plain, such as *F. walujewii*, etc, formed a weak clade (63 BP). Similar to the results on the basis of *matK* and *psbA-trnH*, there was little support based on the *rpl16* analysis in which 6 out of 57 branches obtained bootstrap values of more than 70 BP (Fig 7C).

Fig. 7

Phylogenetic tree on the basis of CP genome

The CP genome matrix included the 27 species of *Fritillaria* genus and 26 of *Lilium* genus, with 3 of *Cardiocrinum* genus and 5 of *Amana* genus as outgroups. One average, (152,099) bp of the CP genome were aligned. The result of ML tree was similar to that of NJ tree (Supplementary Fig2). In the ML tree (Fig 8), the ingroup corresponding to *Fritillaria* and *Lilium* was strongly supported (100 BP), and sister to *Cardiocrinum* (100 BP). In this analysis, *Lilium* was monophyletic (100 BP) and was sister to *Fritillaria* genus. Furthermore, *Lilium* was nested with *Fritillaria* with higher bootstrap support (75 BP) than that (53 BP) of result of Day et al. *Fritillaria*, as the largest subgenus, was paraphyletic and majority of which fall in one strong supported Eurasain clade (A) except *F. maximowiczii* (subgenus *Liliorhiza*). Within the clade A, *F. davidii* appeared as successive sister taxa to the remaining Eurasian species (100 BP), which split into two well-supported clades.

Clade A1 grouped with the monotypic subgenus *Rhinopetalum* (*F. karelinii*) as sister to two species from subgenus *Fritillaria* (*F. ussuriensia* and *F. meleagroides*), which occurred in North region of China. The sister clade (A2) was composed of the remaining 22 species that could be classified into two subclades (100 BP). Subclade B1 contained subgenus *Theresia* (*F. persica*) and subgenus *Petilium* (*F. eduardii*), which occur in the Middle East and Central Asia, while subclade B2 comprising subgenus *Fritillaria* includes 15 species from South China and five species (*F. tortifolia*, *F. verticillata*, *F. yuminensis*, *F. pallidiflora*, and *F. walujewii*) from Xinjiang plain (Fig 9). The 11 most valuable species used in TCM were not included in monophyletic group, as *F. ussuriensis* was separated from the other 11 species. As a whole, the phylogenetic tree based on CP genome was highly supported, in which 54 out of 58 branches obtained bootstrap values more than 70 BP.

Fig. 8

Fig. 9

Specie-specific test for *F. taipaiensis*, *F. unibracteata* and *F. cirrhosa*

For specie-specific test for *F. taipaiensis*, the target intergenic space region, *accD*-psaI, was chosen due to a 137 bp deletion observed in *F. taipaiensis* relative to the other *Fritillaria* species, making it a suitable target for developing specie-specific test for *F. taipaiensis*. As shown in Fig 10A, after routine PCR and electrophoresis, *F. taipaiensis* revealed a unique DNA band at length of 302 bp with limit of detection at 0.239 ng/μL (Lane 4), whereas the other *Fritillaria* species showed DNA band at length of 439 bp.

For specie-specific test for *F. unibraceata* and *F. cirrhosa*, the Taqman MGB real time PCR were carried out because of the 47 bp insertion and 6 bp deletion, respectively. As shown in Fig 10B, only *F. unibraceata* showed positive result and other *Fritillaria* species showed negative results, confirming an analysis specificity of 100%. Furthermore, the reactivity was detected at the limit of 0.1543ng/μL. Similarly, *F. cirrhosa* showed unique positive result in Fig 10C with the limit of detection of 0.0145ng/μL (Supplementary Figure3).

Fig. 10

Discussion

The overall structure of CP genome

Fritillaria species have among the largest known genomes of diploid angiosperms with a 1*C* value (DNA content of the unreplicated haploid chromosome complement varying from 33.25 to 85.38 Gb (Kelly et al. 2015), the phylogenetic relationship and molecular marker based on whole nuclear genome was therefore

difficultly attainable. With the development of *De novo* (Illumina) sequencing technology, the CP genome assembly has become cost-affordable and easier compared with previous Sanger method. Moreover, *De novo* sequencing technology has been widely used in transcriptome assembly in order to identify the biosynthetic and regulatory genes in traditional medicine, such as *Ligusticum Chuanxiong* (Song et al 2015) and *Cassia obtusifolia* (Deng et al 2018). In this research, four new CP genomes of *Fritillaria* were obtained using *De novo* sequencing technology. The size in this research (from 151,076 to 152,043 bp) was in accordance with those of reported CP genomes, such as *F. ussuriensis* (151,524 bp), *F. taipaiensis* (151,693bp), *F.cirrhosa* (151,991bp), and etc. The four CP genomes contained similar genome structure, gene content and gene order that were typical of land plants. Compared with other three species, the number of *tRNA* and *rRNA* genes were identical, but the number of protein coding genes ranged from 77 to 78 due to the absence of *rps16* gene in *F. unibracteata*. The absence of *rps16* gene has also been observed in *Brassicaceae*, *Fabaceae* and *Populus* species (Jin et al 2019). The functional loss of *rps16* gene from the CP genome could be compensated by the mitochondrial and (or) nuclear-encoded *rps16* that could target chloroplast as well as mitochondria (Ueda et al 2008).

The highly conserved genomic structure and gene order as well as no rearrangement of the Fritillaria CP genomes has been observed. The 26 kb of IRs in the *Fritillaria* species was within the size range of most angiosperm CP genome (20-30 kb). However, the IRs of *Pelargonium hortorum* (75 kb) (Chumley et al 2006), *Melanthiaceae* (33 kb) exhibited expansions compared with those of *Fritillaria*, which might play role in stabilizing the structure of the entire plastid as well as the prevention of gene gain and gene loss phenomenon (Zhao et al 2019). Instead, short or loss of IR has also been observed in *Hordeum vulgare* (Fig 2), Geraniaceae (11 kb) (Guisinger et al 2011) and *Cryptomeria fortunei* (114 bp) (Hirao et al 2008) as well. These significant contractions and expansion of IR not only contributed to genome size variation, but have been used as evolutionary loci for phylogenetic relationships (Choi et al 2019). The IR/LSC boundaries in the *Fritillaria* and *Lilium* (*Lilium superbum*) CP genomes expanded into the *rps19* gene, which might be a characteristic CP genome structure of *Fritillaria* genus and its relative genus. Similar expansion was also observed in other taxa from family *Liliaceae*, including *Lilium* (Kim and Lee 2013), *Fritillaria* (Li et al 2014), and *Cardiocrinum* (Lu et al 2016). Li et al (2017) reported that the common location of IR/LSC junctions in *rps19* seemed to be an ancestral symplesiomorphy of *Liliaceae*. Here, the similar feature was also found that the whole *rps19* gene was contained inside the IR in *Smilax china*, *Oncidium gower* and *Allium Chinese*, while in *Hordeum vulgare*, *rps19* did not extend into the IR. The similar IR/LSC boundaries between *Fritillaria*, *Lilium* and *Cardiocrinum* showed there was obvious phylogenetic implication, as they were located closely in phylogenetic tree, and further evidence was needed using sufficient genera of *Liliaceae*.

A careful comparison between repeat sequence and SSR regions revealed the important differences between various *Fritillaria* species leading to establish specific markers for molecular identification. Gao et al (2018) found that frequent variation in repeat sequences played an important role in sequence rearrangement and variation in CP genomes. In this study, a large number of repeat sequences, mainly ranging from 15 to 20

bp, were detected in the chloroplast genomes of four *Fritillaria* species, consistent with the results of studies on the chloroplast genomes of *Cannabaceae* (Zhang et al 2018). In addition to the repeat sequences, the CPSSRs also represented important molecular markers, which were widely used in plant population genetics, polymorphism study, and phyletic evolution. For instance, SSRs have been used to explain the morphology and differentiation of species in the subgenus *Prunus* (Pervaiz et al 2015), and discriminate black spruce (*Picea mariana*) and red spruce (*P. rubens*) (Shi et al 2014). In the CP genomes of *F. unibracteata*, *F. przewalskii*, *F. delavayi* and *F. sinica*, the content of A/T repeats was far greater than that of G/C repeats, similar to the results of Xue (2019) and other studies (Melotto-Passarin et al 2011). Although several variable CP DNA markers, for instance *matK*, *rpl16*, *atpB* and *rbcL*, have been used in phylogenetic studies of *Fritillaria*, they showed small divergence (Pi value of 0.00717, 0.00571, 0.00391 and 0.00505, respectively) among 12 *Fritillaria* species. Based on the result of sliding window analysis, the 10 most divergent hotspots were identified with Pi value ranging from 0.0116 to 0.0159. These hotspots included *ycf1*, *trnL*, *trnF*, *ndhD*, *trnN*, *atpI* and *rps19* in the coding region, and *trnN-trnR*, *trnE-trnT and psbM-trnD* in Internal Gene Space. The highly variable *trnE-trnT* and gene *ycf1* have also been found by Li et al (2018), and gene *ycf1* has been regarded as the most promising plastid DNA barcode of land plants (Dong et al 2015).

The relationship within *Liliaceae*

Prior to the CP genomics, the phylogenetic analysis using three common DNA barcodes, including *matK*, *psbA-trnH* and *rpl16*, were performed. However, the phylogenetic trees were weakly supported with no more than 28.8% branches with bootstrap values of 70 BP. As compared to separate regions, the whole CP genome showed higher resolution than the former since no less than 93.1% branches with bootstrap value of 70 BP in Fig. The higher is the credibility of branch, the more concordant is the guiding value of the evolutionary analysis for the relationship. As suggested by Kress et al (2005) and Ng et al (2017), the whole CP genome was promising to act as super DNA barcode to resolve various *Fritillaria* species efficiently. Moreover, such phylogenetic tree based on whole CP genome data in this study showed similar topology, but with higher resolution, as previous studies (Rønsted et al 2005). Our results also indicated that *Fritillaria* and *Lilium* were evidently sisters, the closest relative being *Cardiocrinum* in a monophyletic genus (100 bp). Rix (2001) recognized 7 sections in genus *Lilium*, but based on our analysis, the infrageneric classification of *Lilium* were mixed throughout the clade as our sampling was more limited than previous studies (Liu et al 2018). More samples for the construction of phylogenetic tree (26 species in the current work, representing 23% of the genus) might further clarify the subgenus relationship in *Lilium* genus. Furthermore, our findings partly supported the recent classification of *Fritillaria* suggested by Rix (2001) and improved the resolution compared with previous studies. In this study, phylogenetic tree based on CP genome did not support genus *Fritillaria* as monophyletic. On the contrary, it was divided into two clades, one containing *F. maximowiczii* from subgenus *Liliorhiza* and the other containing species corresponding to other 5 subgenus, with higher bootstrap (100 BP) compared to 54 BP and 53 BP in the findings of Rønsted et al (2005) and Day et al (2014), respectively. By comparing separate analysis with combined analysis of plastid data, Rønsted et al (2005)

stated that increasing additional gene region would help to improve resolution and bootstrap. Consistently, the whole CP genome in the current study materially improved the results compared with that obtained by separate regions, such as *matK*, *trnH-psbA* and *rpl16*.

To date, more than 20 species were grouped in subgenus *Liliorhiza* by Rix et al (2001), which were predominantly ranging into North American, but *F. maximowiczii* was located in northeastern China and southeastern Russia. In the previous study, *F. maximowiczii*, as ancestral specie, was sister to the remaining species of either subgenus *Liliorhiza* or the whole genus *Fritillaria*, reflecting the possible origin of genus *Fritillaria* in Asia (Rønsted et al 2005). The bulbs of species in subgenus *Liliorhiza* were formed by several imbricate scales, and were attached with numerous loosely rice-shaped bulbils, much resembling those of *Lilium* and hence the name. Due to the lack of CP genome data of *Korolkowia* and *Japonica*, the next clade included the remaining 5 subgenus, *Fritillaria*, *Petilium*, *Theresia*, *Rhinopetalum*, and *F. davidii* as sister group. *F. davidii* was located in western China (Sichuan province), and has previously been clustered in subgenus *Liliorhiza* due to its rice-shaped bulbils (Luo et al 1996). In consistence with our results, *F. davidii* was grouped into its own subgenus, *Davidii*, due to the lack of whorled stem leaves in this specie by Rix (2001). In our analysis, *F. davidii* showed more close relationship with species in the Eurasian clade (Fig 8). That independent evolution of rice-shaped bulbils in *F. davidii* and subgenus *Liliorhiza* might be possible cause, and further comparative developmental studies on the bulbils of *F. davidii* and subgenus *Liliorhiza* were required.

The relationship between evolution and genome size in this genus was particularly interesting, as this genus had genomes exceeding 1C=33 Gb, which was 200 times larger than the genome of the model plant *Arabidopsis thaliana*. Previously, *F. davidii* was regarded as the basal specie, based on the position of branch point, in Eurasian clade and reached the smallest genome size (33.25 Gb) in *Fritillaria*. Meanwhile, *F. maximowiczii* was represented by a basal type, and also contained small genome (33.54 Gb) the second in size only to *F. davidii*. It was recorded that the Japanese endemic species in subgenus *Japonica* had the largest genome for any diploid plant species, which reached over 85 Gb in size (85.38 Gb of *F. japonica* and 85.24 Gb of *F. koidzumiana*) (Kelly et al 2015). These giant genomes were not resulted from the recent massive duplication, but from the relatively low-abundance repeat-derived DNA. Therefore, we hypothesized that the genome size of species was linked to its evolutionary status, and that the species with similar genome size might have close relationship. For instance, the reported genome size of *F. gibbosa*, a member of subgenus *Rhinopetalum*, was approximately 42 Gb. Intriguingly, subgenus *Japonica* showed strongly close relationship with *Rhinopetalum*, but with remarkably divergent genome sizes (with 1C values in member of subgenus *Rhinopetalum* almost half those found in the Japanese species) (Leitch et al., 2007). Under among very closely related lineages, such a dramatic shift in the amount of genome might indicate differences in the underlying mechanisms that control the amplification and nuclear DNA. The linkage between relation and genome size might be attributed mainly to intra-subgenus level but not inter-subgenus level. The availability of an expanded phylogenetic hypothesis of interspecific relationships in *Fritillaria* will allow us to infer the direction and rate

of genome size change and provides the evolutionary background for ongoing comparative genomic studies (Kelly and Leitch, 2011; Kelly et al., 2012).

The relationship within subgenus *Fritillaria* and phylogenetic distribution of 11 medicinal species used in TCM

The subgenus *Fritillaria* appeared to be a paraphyletic group, similar to the results of Day et al (2014). Two species (*F. meleagroides* and *F. ussuriensis*) of subgenus *Fritillaria* clustered together and formed sister clade to *F. karelinii* of subgenus *Rhinopetalum*, similar to the results of Huang et al (2020), Khourang et al (2014) and Li et al (2018), possibly due to small sample size. Previously, Bakshi-Khaniki and Persson (1997) placed this subgenus *Rhinopetalum* on the basis of its deeply impressed nectaries. Subsequently, Rix (2001) supported genetic rank of this subgenus based on the unique raceme, which were clustered with even more than 10 flowers together and hence the name. In the recent studies, subgenus *Rhinopetalum* showed closer relationship with subgenus *Fritillaria* than subgenus *Theresia* and *Petilium*, but with low resolution (Day et al 2014; Rønsted et al 2005). However, according to our analyses, *F. karelinii* was nested in the subgenus *Fritillaria* with highly supporting (100 BP), and showed remote affinity with subgenus *Fritillaria*. The placement of *F. karelinii* was ambiguous, and it would be worthwhile to add more species of subgenus *Rhinopetalum* in future analysis for better understanding of the placement of this subgenus. *F. ussuriensis* and *F. meleagroides* were frequently considered as members of the large subgenus *Fritillaria* (Rix 2001). However, the two species does have some similarities with *F. karalinii* since all of them have small mastoid on filament, which was different from other species in Xinjiang plain with no mastoid. Meanwhile, such mastoid on filament was proposed to be a potential primitive feature, and our results partly supported this hypothesis because *F. karelinii* and *F. meleagroides* diverged early from other species from Xinjiang, such as *F. pallidiflora* etc. Therefore, the genetic rank of *F. meleagroides* and *F. ussuriensis* might be treated as an additional subgenus, or alternatively an expanded subgenus *Rhinopetalum* would add *F. meleagroides* and *F. ussuriensis*. Subgenus *Theresia* (*F. persica*) and *Petilium* (*F. eduardii*) had close relationship and formed monophyletic subclade B1, which was similar to the result of Day et al (2014) and Li et al (2018). Commonly, the species in subgenus *Petillium,* including *F. eduardii,* had tall stems (up to 100 cm), and were distributed in Central and Western Asia. The bulbs were much larger than those of most *Fritillaria* species. Like the species of subgenus *Petilium*, *F. persica*, a member of monotypic subgenus *Theresia*, have a sturdy stem (up to 100 cm) and big bulb second in size only to subgenus *Petilium*.

The other 20 species of subgenus *Fritillaria* formed a strongly supported clade (Clade B2, 100 BP), including two resolved subclades (Fig 8). As shown in Fig 8 and Fig 9, 5 species from Xinjiang plain were included in a strongly supported subclade (subclade C1), which was sister to subclade C2 containing the other 15 species from outside Xinjiang plain. This signified that the Xingjiang species had a close genetic relationship. The other 15 species in the subclade C2 were located in South China, in which four species, *F. monantha*, *F. anhuiensis*, *F. thunbergii* and *F. hupehensis*, were located in east China plain and included in subclade (100 BP), and 11 species in another subclade were distributed in Hengduan mountains (100 Bp).

Based on China pharmacopoeia 2020, the 11 important medicinal *Fritillaria* species were widely distributed in four hotspots, Xinjiang plain, northeastern China plain, east China plain and Hengduan Mountains. The Hengduan Mountains are located in the southeastern part of the Qinghai-Tibet plateau, running from North to South China. The former two regions constituted hotspots in North China, while the latter two regions constituted hotspots in South China. Interestingly, the eight species in the upper location of the clade originated from Xinjiang plain and northeastern China plain (*F. ussuriensis*), whereas the 12 species in the lower location distributed in East China and Hengduan Mountains region. Consequently, it was evident that the geographical distribution pattern of the 11 medicinal species as well as relative species neatly mapped on the phylogenetic tree, especially by plastid data (Rønsted et al 2005). Previous studies have demonstrated the geographical distribution pattern of *Fritillaria* species was significantly affected by various eco-environmental factors in environment, including temperature, moisture, altitude, light, etc. For instance, *F. cirrhosa* was distributed in Hengduan Mountains (ie, Sichuan, Qinghai and Yunnan province), with suitable parameters including high altitude (3200-4300 meters), ample precipitation (65-135 mm) and big diurnal temperature (13-17 °C) (Cunningham et al 2018). In Hengduan Mountains, high altitude might be an essential environmental constraint influencing the growth of *Fritillaria* species. To date, most *Fritillaria* species in Hengduan Mountains distributed areas with altitude no less than 1800 meters, whereas *Fritillaria* species growing areas in the other 3 hotspots did not exceed 1000 meters above sea level.

Early in 1987, *F. unibracteata*, *F. cirrhosa*, *F. prezewalskii* and *F. delavayi* were recorded as national third-class endangered medicinal plants of China, and were facing extinction in the wild due to the over-excavation, habitat compression, over grazing and expanding international herbal market (Konchar et al 2011). The discovery that some of the most important species used in China pharmacopoeia showed close relationship to widely cultivated members of subgenus *Fritillaria*, raised the possibility that the rare species were replaced by those widely cultivated species. Recent analyses have demonstrated that *F. crassicaulis*, showing closest relationship with *F. cirrhosa*, has been widely used as the substitution of *F. cirrhosa* by people of Naxi nationality and Tibetan since Ming/Qing Dynasty (Tang et al 1992). These findings not only supported that traditional knowledge could be effective in identifying plants with bioactive compounds, but also highlighted that phylogenetic trees based on CP genomes were promising method to select potential novel medicinal species. In future, a number of species within the subgenus *Fritillaria*, especially those showing close relationship to the important species in TCM, such as *F. sichuanica*, *F. dajinensis*, *F. yuzhongensi*s, *F. sinica* and *F. crassicaulis*, might be investigated whether these bulbs contain the same bioactive compounds that existed in the complex of *F. cirrhosa*.

The perplexing phylogenetic placement and specie-specific identification of *Fritillaria* species

It was evident that the subgenus *Fritillaria* was split into two distinct groups and was paraphyletic. The non-monophyletic trait of subgenus *Fritillaria* has also been observed by Rønsted et al (2005) and Day et al (2014). To date, there were at least four factors that contributed to the perplexing phylogenetic placement of *Fritillaria* species. Firstly, hybridization/introgression might contribute to the molecular phylogenetic non-

monphyly (Funk and Omland 2003). There were several reports of natural interspecific hybrids (e.g. *F. ussurinensis* (Ruan et al 2004) and *F. eduardii* (Wietsma et al 2001)), which might promote the splitting of different variants into separate. Convergent or parallel evolution of phenotypic traits might be another common cause of incongruence between morphological classification and the results of molecular phylogenetic analyses. For example, *F. davidii* had rice-shaped bulbils, resembling the morphological character of subgenus *Liliorhiza*, and used to be grouped in subgenus *Liliorhiza*. But based on our results, it was distantly related to genus *Liliorhiza* and was thus placed in subgenus *Davidii* as described by Rix (2001). On the other hand, such bulbils could be also observed in *F. ussuriensis*, *F. anhuiensis*, and *F. persica*, in which *F. anhuiensis* was suggested to be a member of subgenus *Liliorhiza*, but all of which were grouped together with *F. davidii* in Eurasian clade. It was suggested that rice-shaped bulbils have independently evolved in *F. davidii* and subgenus *Liliorhiza* due to geographic separation, followed by a loss in some species in Eurasian clade during evolution. In future, comparative developmental studies on the bulbils of *F. davidii* with subgenus *Liliorhiza*, *Fritillaria* and *Theresia* were required. The third crucial factor might be the morphological heterogeneity of some *Fritillaria* species. It was surprising to find that *F. delavayi* was so closely related to *F. crassicaulis* because they were obviously different in stem length. The stem lengths of *F. delavayi* and *F. crassicaulis* were usually 35 cm and more than 80 cm, respectively. One possible reason might be the biogeographic limit, as *F. delavayi* distributed in alpine zone of the Henduan Mountains with altitude of 3800-4700 m, while *F. crassicaulis* was located in semi-alpine zone with altitude of 3000-3400 m. The alpine zone of the Hengduan Mountains in Southwest China is generally characterized by low air temperatures, fresh gale, high air humidity and short periods of intense solar radiation. Plants in such habitats have developed a variety of adaptive morphologic characterizers, including short, hairy stems and leaves, even creeping like cushions, to cope with the hostile environmental conditions (Tsukaya and Tsuge 2001). The fourth possible factor is the sample size in the phylogenetic tree. The obvious case was from the result of Rønsted et al (2005), who found that *F. pallidiflora* was resolved solely within the *Korolkowia/Petilium/Theresia* clade. It was suggested that *F. pallidiflora* might either be grouped as a new subgenus or included subgenus *Petilium*. Our new results demonstrated that *F. pallidiflora* was clustered within subgenus *Fritillaria* and actually more closely related to *Petilium/Theresia*. The conflict in *F. pallidiflora* was likely to be solved by using whole CP genome instead of separate regions. In addition, some species with confused phylogenetic placement might be an explanation for the non-monophyly. For example, based on the CP genome, *F. unibraceata* was sister to *F. wabuensis* with divergence of 0.003, which was more than that between *F. sichuanica* and *F. dajinensis* (0.002). If *F. sichuanica* and *F. dajinensis* were given at rank of specie, based on the CP genome it was preferable to follow Tang and Yue (1983) and rank *F. wabuiensis* as specie than rank of variant. However, this result was merely based on CP genome, the accurate placement of *F. wabuensis* kept further evaluation by nuclear genome comparison although it was extremely difficult to obtain.

Thus, the incongruence between morphology and molecular phylogeny highlighted the specie-specific identification of *Fritillaria* species. However, there were hardly any established tests published so far show

deficiencies in specificity or sensitivity. There were at least three advantages, including specificity, sensitivity and efficiency, in this study. First, the routine and TaqMan MGB real time PCR were tested on 12 different species within subgenus *Fritillaria*. Only the target species tested positive in the assay. Second, the detection limit of method was 0.2391 ng/μL, 0.1543ng/μL and 0.0145ng/μL for *F. taipaiensis*, *F. unibraceata* and *F. cirrhosa*, respectively, which enabled to detect trace amount of samples. Third, the whole process in tests from DNA extraction to final result took as long as no more than 3 or 4 h, which supplied rapid detection and might develop standard detection reagents in future.

Conclusion

The four CP genome of species from subgenus *Fritillaria* provided support for taxonomic clarification, phylogenetic relationship and development of DNA markers. Compared with partial regions, such as *psbA-trnH*, *matK* and *rpl16*, the whole CP genome could increase resolution based on high bootstrap, the result of which supported for the monophyly of genus *Lilium*, *Amana* and *Cardiocrinum*, except that the largest genus *Fritillaria* was paraphyletic. The 11 members of subgenus *Fritillaria* that were used in TCM were split into two clusters since *F. ussuriensis* was clustered with *F. meleagroides* and *F. karelinii*. In addition, the high consistence between the location of species in phylogenetic tree and their geographical distribution provided evidence that environment substantially affected evolution of *Fritillaria* subgenus, and also highlighted the importance of CP genome in the evolutional analysis. The most important medicinal species, especially *F. cirrhosa* complex, were found to be close to species that were in widespread cultivation for medicinal and ornamental purposes. Excitingly, those closely related species from subgenus *Fritillaria* might be promising alternatives to balance the improving market and rare resources. Finally, this study developed specie-specific identification on *F. taipaiensis*, *F. cirrhosa* and *F. unibracteata*, respectively, the result of which might supply a standard process to conveniently designate and validate specific DNA markers for traditional medicine.

Acknowledgments

This work was partially funded by National Natural Science Foundation of China (No. 31500276), Sichuan Science and Technology Program (No. 2018SZ0061 and 2021ZHFP0170).

Conflict of interest

We declare that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

Author contributions

Tian Zhang performed the experiments. Sipei Huang analyzed the data. Simin Song, Meng Zou, Weiwei Wang, Tiechui Yang contributed to the reagents/materials/analysis tool. Jiayu Zhou and Hai Liao wrote and were responsible for the paper.

Availability of data and materials

The chloroplast genomes generated during the current study were deposited in NCBI with accession number of MW849272 (*F. unibraceata*), MW849274 (*F. przewalskii*), MW849275 (*F. delavayi*) and MW849273 (*F. sinica*), respectively. All the raw Illumina data of *F. unibracteata*, *F. przewalskii*, *F. delavayi* and *F. sinica* have been deposited in the Sequence Read Archive (SRA) of the NCBI under accession numbers of SRR14454932, SRR14455034, SRR14454929 and SRR14455331, respectively.

References

Abdullah, Henriquez CL, et al. Complete chloroplast genomes of *Anthurium huixtlense* and *Pothos scandens* (*Pothoideae*, *Araceae*): unique inverted repeat expansion and contraction affect rate of evolution. J Mol Evol, 2020, 88(7): 562-574.

Bankevich A, Nurk S, Antipov D, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol, 2012, 19(5): 455-77.

Bansal S, Thakur S, Mangal M, et al. DNA barcoding for specific and sensitive detection of *Cuminum cyminum* adulteration in Bunium. Phytomedicine, 2018, 50(15): 178-183.

Chen Q, Wu XB, Zhang DQ. Phylogenetic analysis of *Fritillaria cirrhosa* D. Don and its closely related species based on complete chloroplast genomes. Peer J, 2019, 7: e7480.

Choi IS, Jansen R, Ruhlman T. Lost and found: return of the inverted repeat in the legume clade defined by its absence. Genome Biol Evol, 2019, 11(4): 1321-1333.

Chumley TW, Palmer JD, Mower JP, et al. The complete chloroplast genome sequence of *Pelargonium* x *hortorum*: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. Mol Biol Evol, 2006, 23(11): 2175-2190.

Cunningham AB, Brinckmann JA, Pei SJ, et al. High altitude species, high profits: Can the trade in wild harvested *Fritillaria cirrhosa* (*Liliaceae*) be sustained? J Ethnopharmacol, 2018, 223: 142-151.

Daniell H, Lin CS, Yu M, et al. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol, 2016, 17(1):, 134.

Day PD, Berger M, Hill L, et al. Evolutionary relationships in the medicinally important genus *Fritillaria* L. (*Liliaceae*). Mol Phylogenet Evol, 2014, 80: 11-19.

Deng Y, Zheng H, Yan Z, et al. Full-length transcriptome survey and expression analysis of *Cassia obtusifolia* to discover putative genes related to aurantio-obtusin biosynthesis, seed formation and development, and stress response. Int J Mol Sci, 2018, 19: 2476.

Dong W, Xu C, Li C. Ycf1, the most promising plastid DNA barcode of land plants. Sci Rep, 2015, 5: 8348.

Funk DJ, Omland KE. Species-level paraphyly and polyphyly: frequency, causes and consequences with insights from animal mitochondrial DNA. Ann Rev Ecol Evol Syst, 2003, 34: 397-423.

Gao X, Zhang X, Meng H, et al. Comparative chloroplast genomes of *Paris Sect. Marmorata*: insights into repeat regions and evolutionary implications. BMC Genomics, 2018, 19 (Suppl 10): 878.

Guisinger MM, Kuehl JV, Boore JL, et al. Extreme reconfiguration of plastid genomes in the angiosperm family *Geraniaceae*: rearrangements, repeats, and codon usage. Mol Biol Evol, 2011, 28 (1): 583-600.

Guo X, Liu J, Hao G, et al. Plastome phylogeny and early diversification of *Brassicaceae*. BMC Genomics, 2017, 18(1): 176.

Hirao T, Watanabe A, Kurita M, et al. Complete nucleotide sequence of the *Cryptomeria japonica* D. Don. chloroplast genome and comparative chloroplast genomics: diversified genomic structure of coniferous species. BMC Plant Biology, 2008, 8 (1): 70.

Huang J, Yang LQ, Yu Y, et al. Molecular phylogenetics and historical biogeography of the tribe *Lilieae* (*Liliaceae*): Bi-directional dispersal between biodiversity hotspots in Eurasia. Ann Bot, 2018, 122: 1245-1262.

Huang J, Yu Y, Liu YM, et al. Comparative chloroplast genomics of *Fritillaria* (*Liliaceae*), inferences for phylogenetic relationships between *Fritillaria* and *Lilium* and plastome evolution. Plants (Basel), 2020, 9(2): 133.

Jin DP, Choi IS, Choi BH. Plastid genome evolution in tribe *Desmodieae* (*Fabaceae*: *Papilionoideae*). PLoS One, 2019, 14(6): e0218743.

Kelly LJ, Leitch AR, Fay MF, et al. Why size really matters when sequencing plant genomes. Plant Ecol Divers, 2012, 5: 415-425.

Kelly LJ, Leitch IJ. Exploring giant plant genomes with next-generation sequencing technology. Chromosome Res, 2011, 19: 939-953.

Kelly LJ, Renny-Byfield S, Pellicer J, et al. Analysis of the giant genomes of *Fritillaria* (*Liliaceae*) indicates that a lack of DNA removal characterizes extreme expansions in genome size. New Phytol, 2015, 208(2): 596-607.

Khourang M, Babaei A, Sefidkon F, et al. Phylogenetic relationship in *Fritillaria* spp. of Iran inferred from ribosomal ITS and chloroplast *trnL-trnF* sequence data. Biochem Syst Ecol, 2014, 57: 451-457.

Kim KJ, Lee HL. Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants. DNA Res, 2004, 11: 247-261.

Konchar K, Li XL, Yang YP, et al. Phytochemical variation in *Fritillaria cirrhosa* D. Don (Chuan Bei Mu) in relation to plant reproductive stage and timing of harvest. Econ Bot, 2011, 65: 283-294.

Kress WJ, Wurdack KJ, Zimmer EA, et al. Use of DNA barcodes to identify flowering plants. P Natl Acad Sci USA, 2005, 102: 8369-8374.

Leitch IJ, Beaulieu JM, Cheung K, et al. Punctuated genome size evolution in *Liliaceae*. J Evol Biol, 2007, 20: 2296-2308.

Li P, Lu RS, Xu WQ, et al. Comparative genomics and phylogenomics of east Asian Tulips (*Amana*, *Liliaceae*). Front Plant Sci, 2017, 8: 451-463.

Li Q, Li Y, Song J, et al. High-accuracy *de novo* assembly and SNP detection of chloroplast genomes using a SMRT circular consensus sequencing strategy. New Phytol, 2014, 204(4): 1041-9.

Li Y, Zhang Z, Yang J, et al. Complete chloroplast genome of seven *Fritillaria* species, variable DNA markers identification and phylogenetic relationships within the genus. PLoS One, 2018, 13(3): e0194613.

Liu HY, Yu Y, Deng YQ, et al. The chloroplast genome of *Lilium henrici*: genome structure and comparative analysis. Molecules, 2018, 23(6): 1276.

Liu ZD, Wang S, Chen SC. A taxonomic note of *Fritillaria wabuensis* (Liliaceae). Acta Botanica Yunnanica (in Chinese), 2009, 31(2): 145.

Luo YB, Chen XQ. A revision of *Fritillaria* L. (*Liliaceae*) in the Hengduan mountains and adjacent regions, China (II). Acta Phytotaxonomica Sinica (in Chinese), 1996, 34(5): 547-553.

Melotto-Passarin DM, Tambarussi EV, Dressano KD, et al. Characterization of chloroplast DNA microsatellites from *Saccharum* spp and related species. Genet Mol Res, 2011, 10: 2024-2033.

Ng PK, Lin SM, Lim PE, et al. Complete chloroplast genome of *Gracilaria firma* (*Gracilariaceae*, *Rhodophyta*), with discussion on the use of chloroplast phylogenomics in the subclass *Rhodymeniophycidae*. BMC Genomics, 2017, 18(1): 40.

Park I, Kim WJ, Yeo SM, et al. The complete chloroplast genome sequences of *Fritillaria ussuriensis* Maxim. and *Fritillaria cirrhosa* D. Don, and comparative analysis with other *Fritillaria* species. Molecules, 2017, 22(6): 982.

Pervaiz T, Sun X, Zhang Y, et al. Association between chloroplast and mitochondrial DNA sequences in Chinese *Prunus genotypes* (*Prunus persica*, *Prunus domestica*, and *Prunus avium*). BMC Plant Biol, 2015, 15: 4.

Rix EM. Fritillaria: A revised classification together with an updated list of species. Publication of the Fritillaria Group of the Alpine Garden Society, UK, 2001.

Rønsted N, Law S, Thornton H, et al. Molecular phylogenetic evidence for the monophyly of *Fritillaria* and *Lilium* (*Liliaceae*; *Liliales*) and the infrageneric classification of *Fritillaria*. Mol Phylogenet Evol, 2005, 35(3): 509-27.

Ruan HL, Zhang YH, Pan XC, et al. Studies on the chemical constituents from culbs of hybridized Bulbus *Fritillariae ussuriensis*. Zhongguo Zhong Yao Za Zhi (in Chinese), 2004, 29(4): 331-334.

Shi YZ, Forneris N, Rajora OP. Highly informative single-copy nuclear microsatellite DNA markers developed using an AFLP-SSR approach in black spruce (*Picea mariana*) and red spruce (*P. rubens*). PLoS One, 2014, 9(8): e103789.

Song T, Liu ZB, Li JJ, et al. Comparative transcriptome of rhizome and leaf in *Ligusticum Chuanxiong*. Plant Syst Evol, 2015, 301: 2073-2085.

Tang SY, Yue SJ. Three new species of *Fritillaria* Linn. Acta Academiae Medicinae Sichuan (in Chinese), 1983, 14(4): 327-334.

Tang SY, Yue SJ. *Fritillaria* genus, Flora of Sichuan. Publication of Sichuan Ethnic Publishing House,

753 1992, 7: 55-82.
754 Tsukaya H, Tsuge T. Morphological adaptation of inflorescences in plants that develop at low
755 temperatures in early spring: the convergent evolution of “downy plants”. Plant Biol, 2001, 3: 536-543.
756 Ueda M, Nishikawa T, Fujimoto M, et al. Substitution of the gene for chloroplast RPS16 was assisted by
757 generation of a dual targeting signal. Mol Biol Evol, 2008, 25(8): 1566-75.
758 Vinnersten A, Bremer K. Age and biogeography of major clades in *Liliales*. Am J Bot, 2001, 88(9): 1695-
759 703.
760 Watson SJ, Sowden RG, Jarvis P. Abiotic stress-induced chloroplast proteome remodelling: a mechanistic
761 overview. J Exp Bot, 2018, 69(11): 2773-2781.
762 Wicke S, Schneeweiss GM, dePamphilis CW. The evolution of the plastid chromosome in land plants:
763 gene content, gene order, gene function. Plant Mol Biol, 2011, 76: 273-297.
764 Wietsma WA, van den Berg RG, van Scheepen J, et al. The nomenclatural history of *Fritillaria eduardii*
765 and the correct names of its varieties. TAXON, 2011, 60(6): 1754-1759.
766 Wolfe KH, Li WH, Sharp PM. Rates of nucleotide substitution vary greatly among plant mitochondrial,
767 chloroplast, and nuclear DNAs. Proc Natl Acad Sci USA, 1987, 84: 9054-9058.
768 Xue S, Shi T, Luo W, et al. Comparative analysis of the complete chloroplast genome among *Prunus*
769 *mume*, *P. armeniaca*, and *P. salicina*. Hortic Res-England, 2019, 6: 89.
770 Yang Y, Zhou T, Duan D, et al. Comparative analysis of the complete chloroplast genomes of five
771 Quercus species, Front Plant Sci, 2016, 7: 959.
772 Zhang HL, Jin JJ, Moore MJ, et al. Plastome characteristics of Cannabaceae. Plant Divers, 2018, 40(3):
773 127-137.
774 Zhao FY, Tao AE, Li Y. The complete chloroplast genome sequence of the medicinal plant *Paris*
775 *polyphylla* (Melanthiaceae). Mitochondrial DNA B Resour, 2019, 4(2): 3971-3972.
776 Zheng H, Deng KY, Chen AQ, et al. Molecular identification and genetic relationship of *Fritillaria*
777 *cirrhosa* and related species based on DNA barcode. Acta Pharmaceutica Sinica (in Chinese), 2019, 54(12):
778 2326-2334.

Figure 1

Fig 1. Chloroplast genome maps of *F. unibracteata*, *F. przewalskii*, *F. delavayi* and *F. sinica*.

Fig 1. Chloroplast genome maps of *F. unibracteata*, *F. przewalskii*, *F. delavayi* and *F. sinica*. Genes belonging to functional group are color-coded. The positive coding gene is located on the outside of the circle, and the reverse coding gene is located on the inside of the circle. The grey circle inside circle represents the GC content.

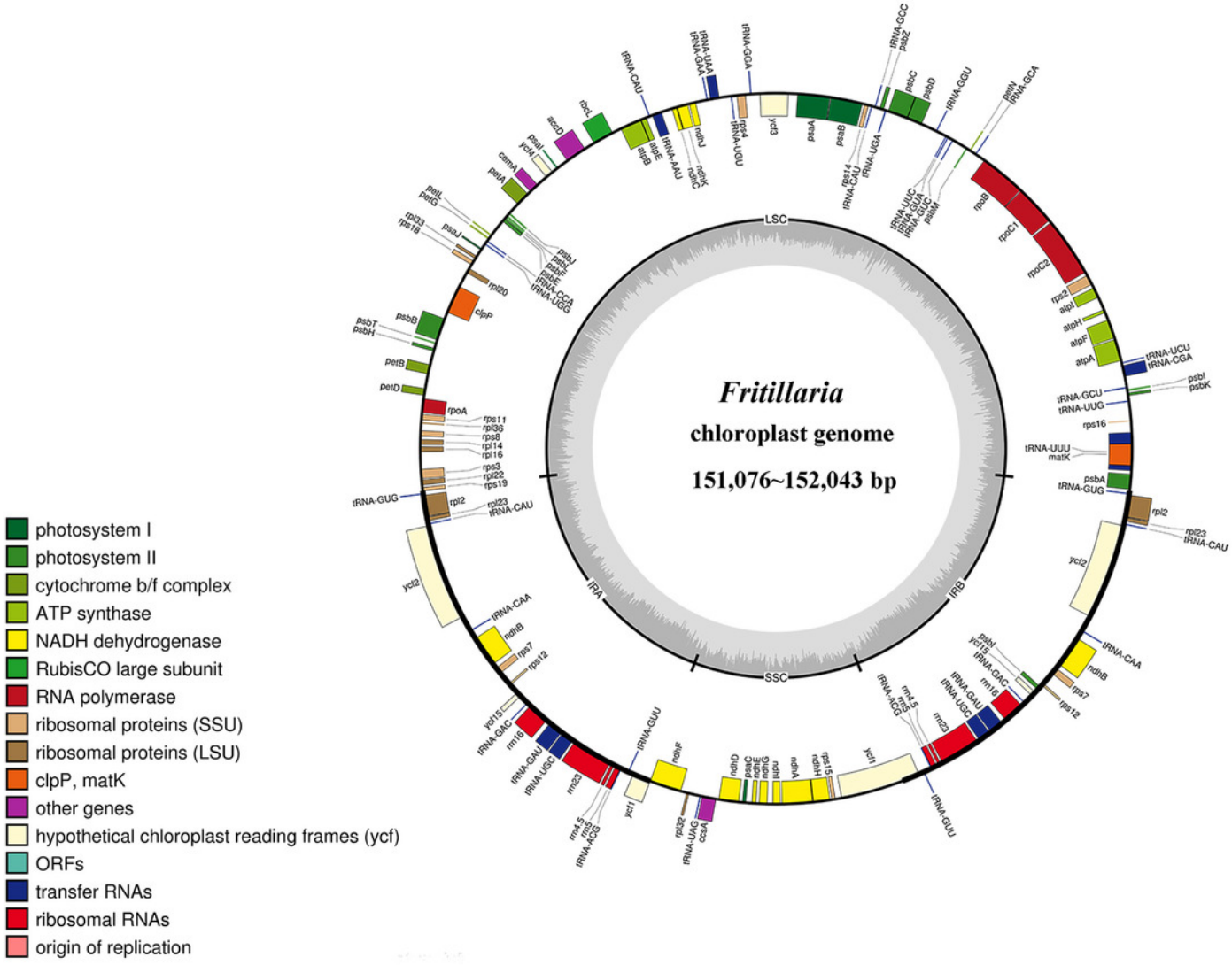


Figure 2

Comparison of LSC, IRs, and SSC junction positions among 17 CP genomes.

Fig 2. Comparison of LSC, IRs, and SSC junction positions among 17 CP genomes.

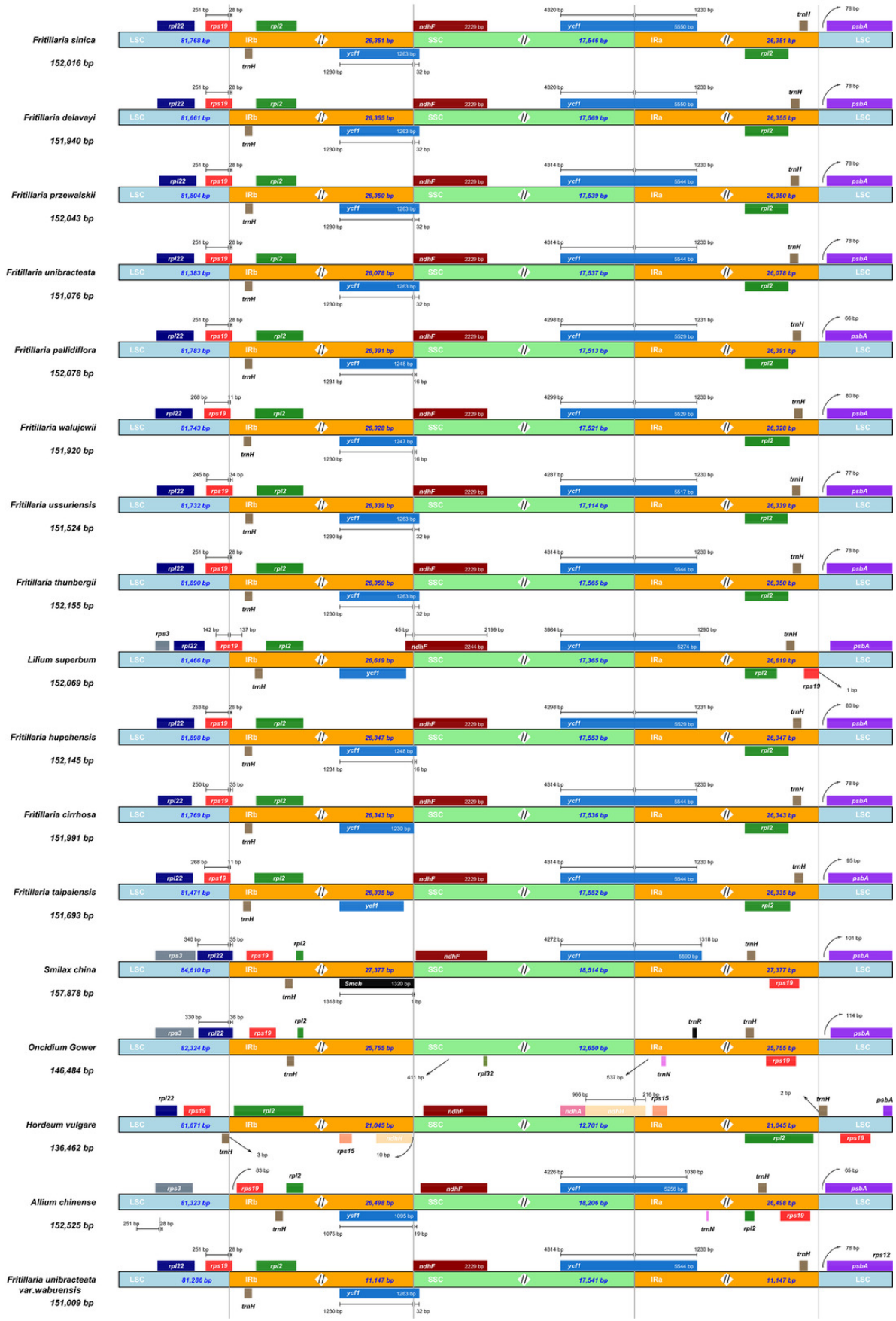

Fritillaria sinica
152,016 bp
Fritillaria delavayi
151,940 bp
Fritillaria przewalskii
152,043 bp
Fritillaria unibracteata
151,076 bp
Fritillaria pallidiflora
152,078 bp
Fritillaria walujewii
151,920 bp
Fritillaria ussuriensis
151,524 bp
Fritillaria thunbergii
152,155 bp
Lilium superbum
152,069 bp
Fritillaria hupehensis
152,145 bp
Fritillaria cirrhosa
151,991 bp
Fritillaria taipaiensis
151,693 bp
Smilax china
157,878 bp
Oncidium Gower
146,484 bp
Hordeum vulgare
136,462 bp
Allium chinense
152,525 bp
Fritillaria unibracteata var.wabuensis
151,009 bp
LSC
IRb
SSC
IRa
rpl22
rps19
rpl2
ndhF
ycf1
trnH
psbA
rps3
rps12
rps15
rpl32
trnR
trnN
Smch

Figure 3

Length distribution of repeat sequences in *Fritillaria* species.

Fig 3. Length distribution of repeat sequences in *Fritillaria* species. A *Fritillaria unibracteata* repeat sequence. B *Fritillaria przewalskii* repeat sequence. C *Fritillaria delavayi* repeat sequence. D *Fritillaria sinica* repeat sequence. Abscissa is the type of scattered repetition sequence, and ordinate is the number of scattered repetition sequence. D represents positive repetition, P represents palindrome repetition (including reverse and complementary).

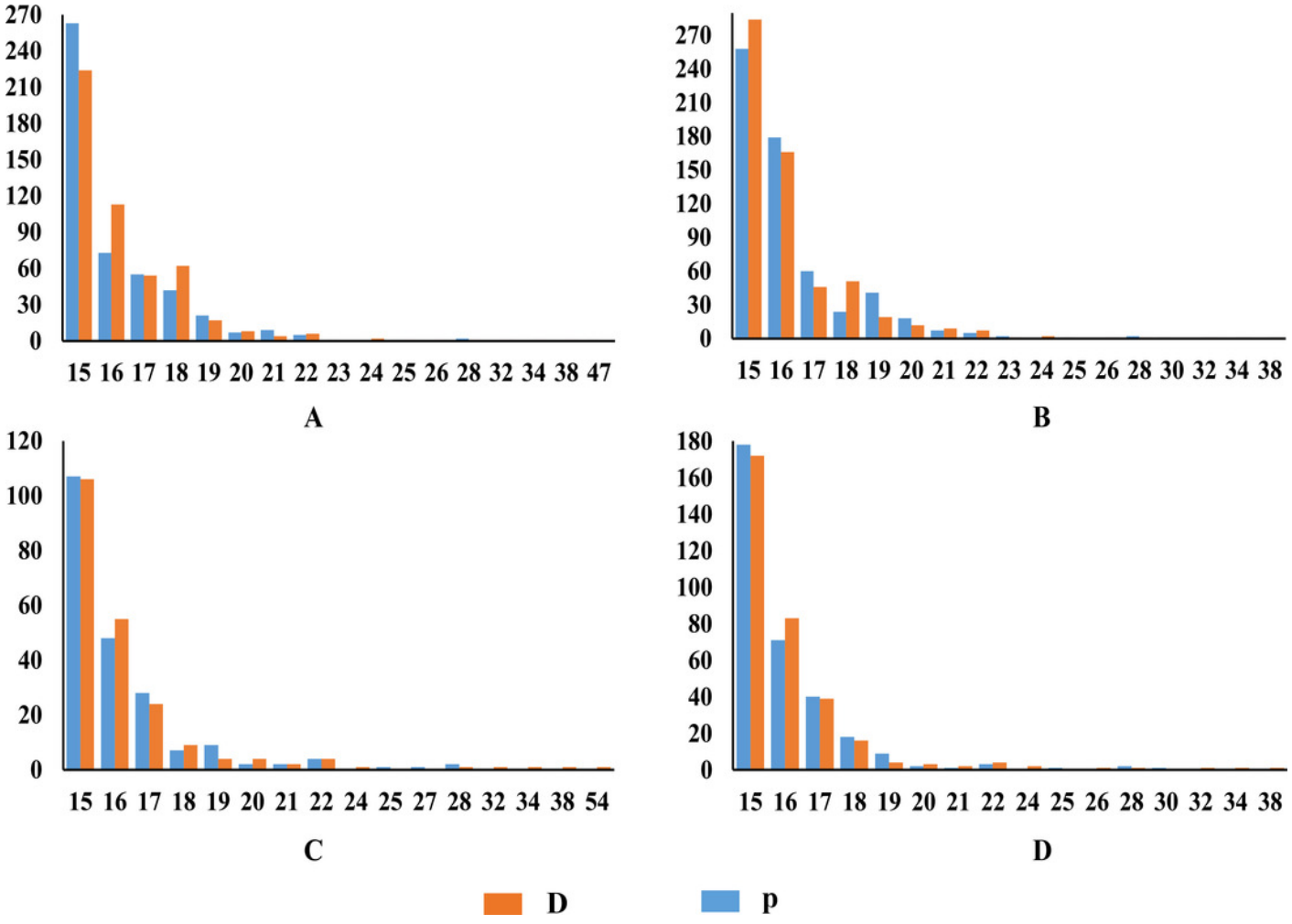


Figure 4

Analysis of simple repetitive sequences in four *Fritillaria* CP genomes.

Fig 4. Analysis of simple repetitive sequences in four *Fritillaria* CP genomes.

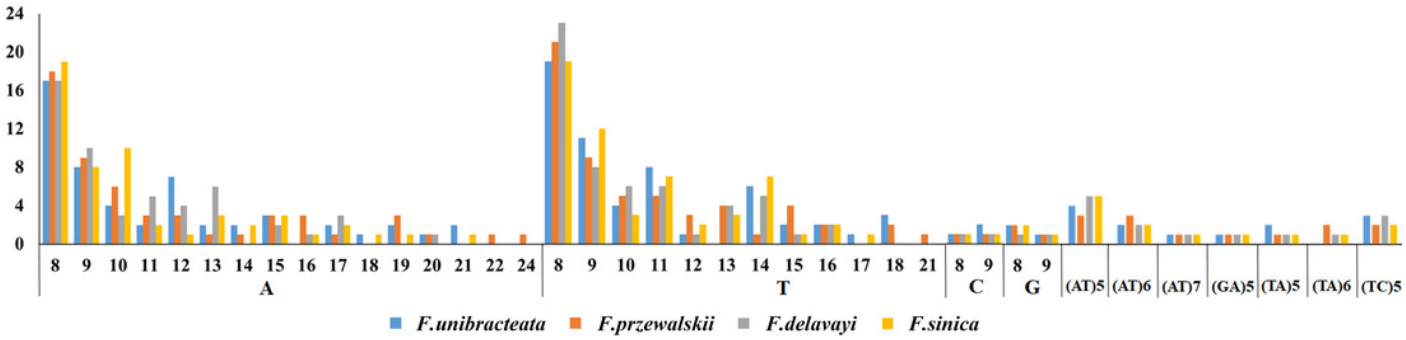


Figure 5

Histogram of RSCU analysis of four species of *Fritillaria.*

Fig 5. Histogram of RSCU analysis of four species of *Fritillaria*. The following block represented all codons encoding each amino acid, and the height of the upper column represents the sum of RSCU values of all codons.

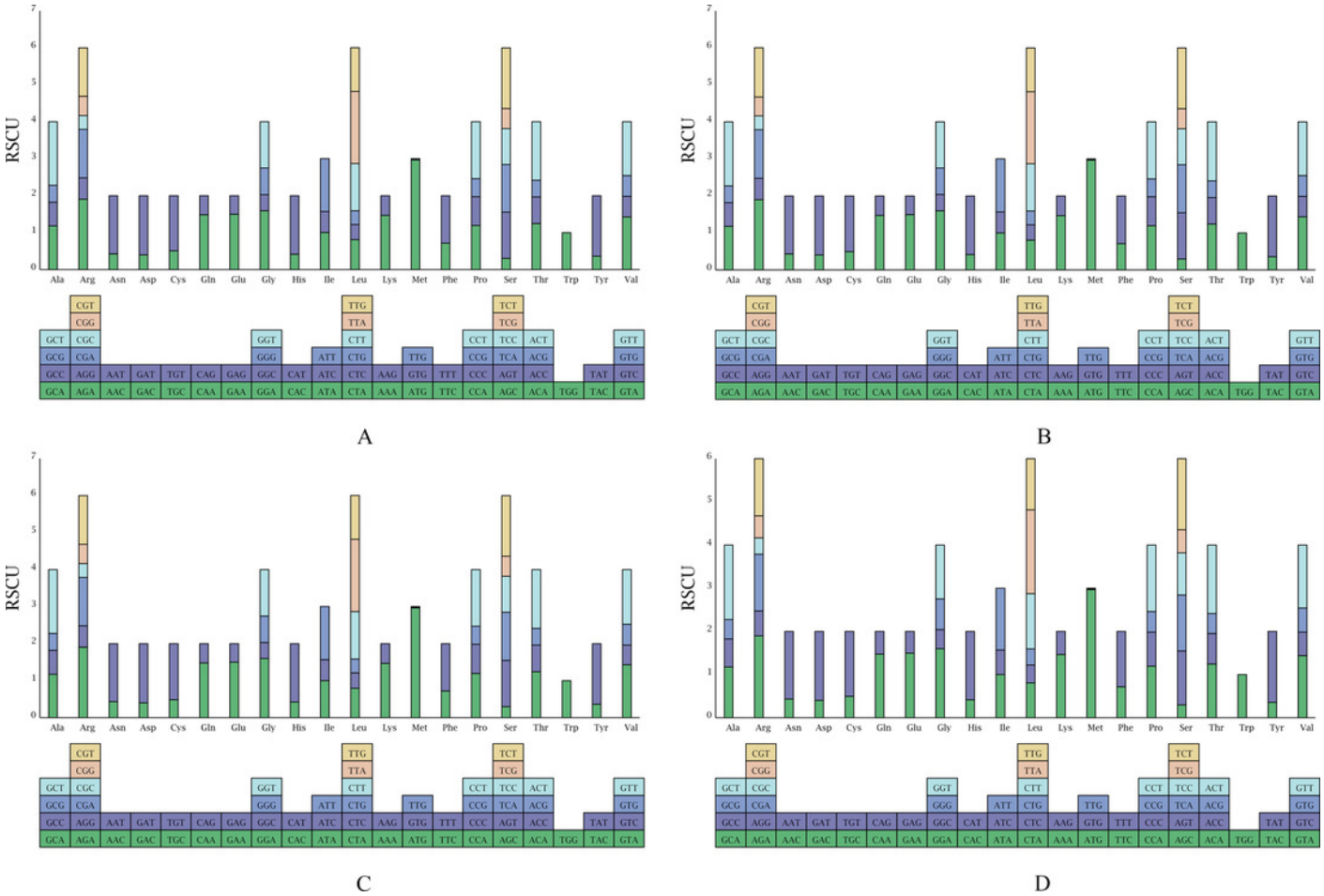


Figure 6

Sliding window analysis of the entire CP genome of 12 *Fritillaria* species.

Fig 6. Sliding window analysis of the entire CP genome of 12 *Fritillaria* species. (Window length: 600 bp, step size: 200 bp). X-axis indicated the position of the midpoint of a window. Y-axis indicated the nucleotide diversity of each window.

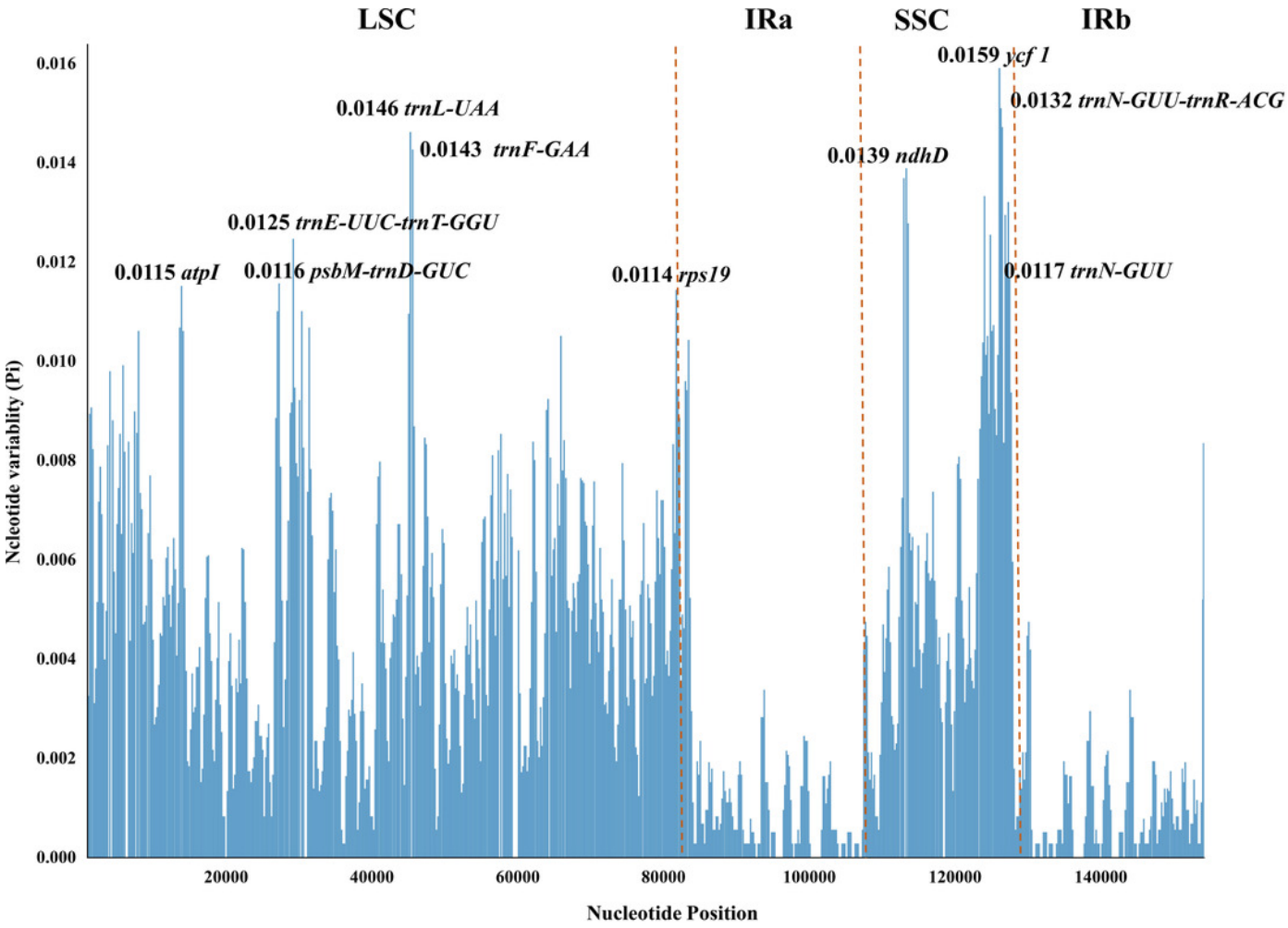


Figure 7

Phylogenetic analysis of *matK*, *psbA-trnH* and *rpl16* sequences inferred from Maximum Likelihood tree.

Fig 7. Phylogenetic analysis of *matK*, *psbA-trnH* and *rpl16* sequences inferred from Maximum Likelihood tree. Numbers above nodes are supporting values with ML bootstrap values. A: phylogenetic analysis of *matK* sequence; B: phylogenetic analysis of *psbA-trnH* sequence; C: phylogenetic analysis of *rpl16* sequence.

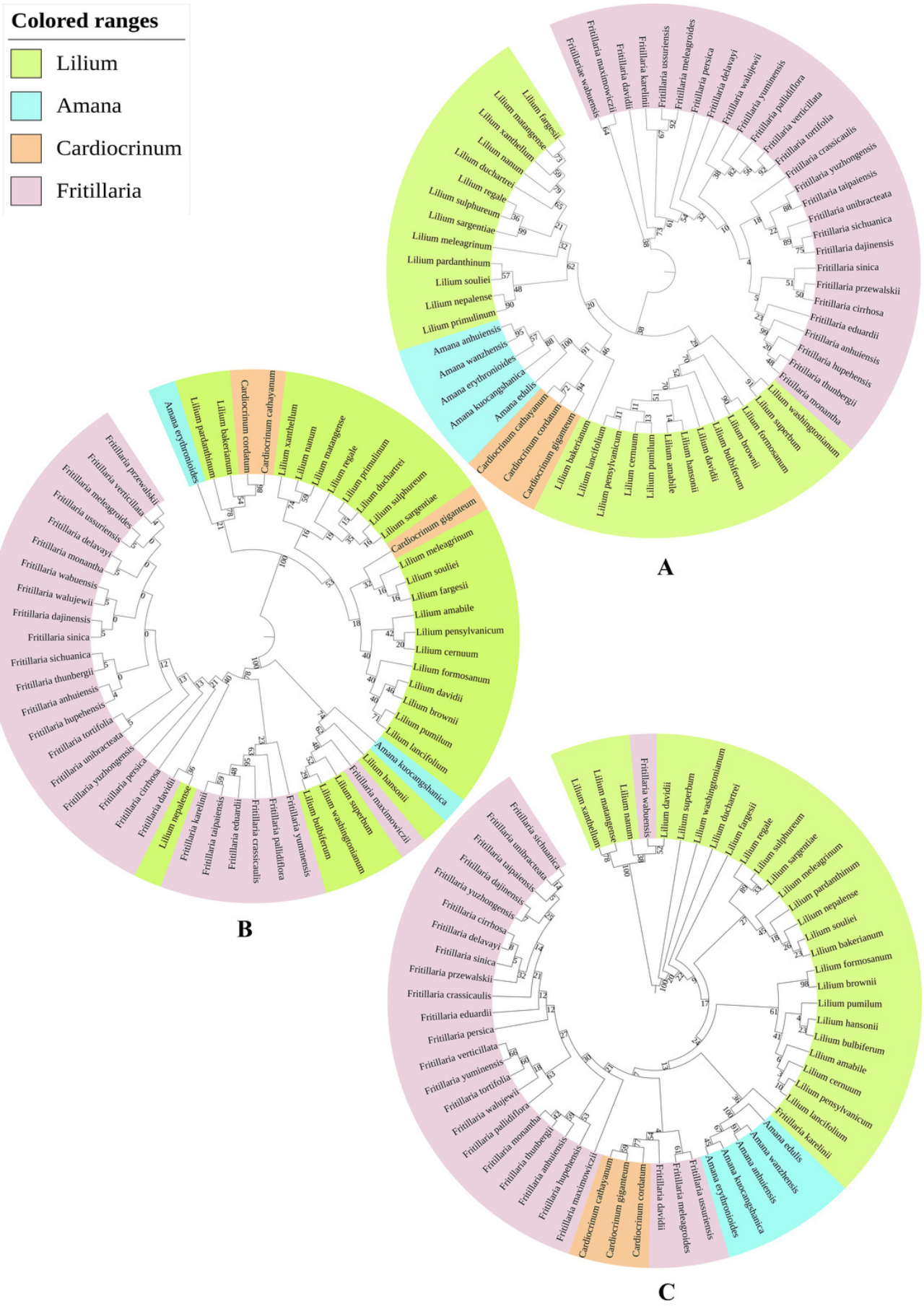

Colored ranges
Lilium
Amana
Cardiocrinum
Fritillaria
A
B
C

Figure 8

Phylogenetic relationship of 61 species inferred from Maximum Likelihood tree.

Fig 8. Phylogenetic relationship of 61 species inferred from Maximum Likelihood tree. Numbers above nodes are supporting values with ML bootstrap values.

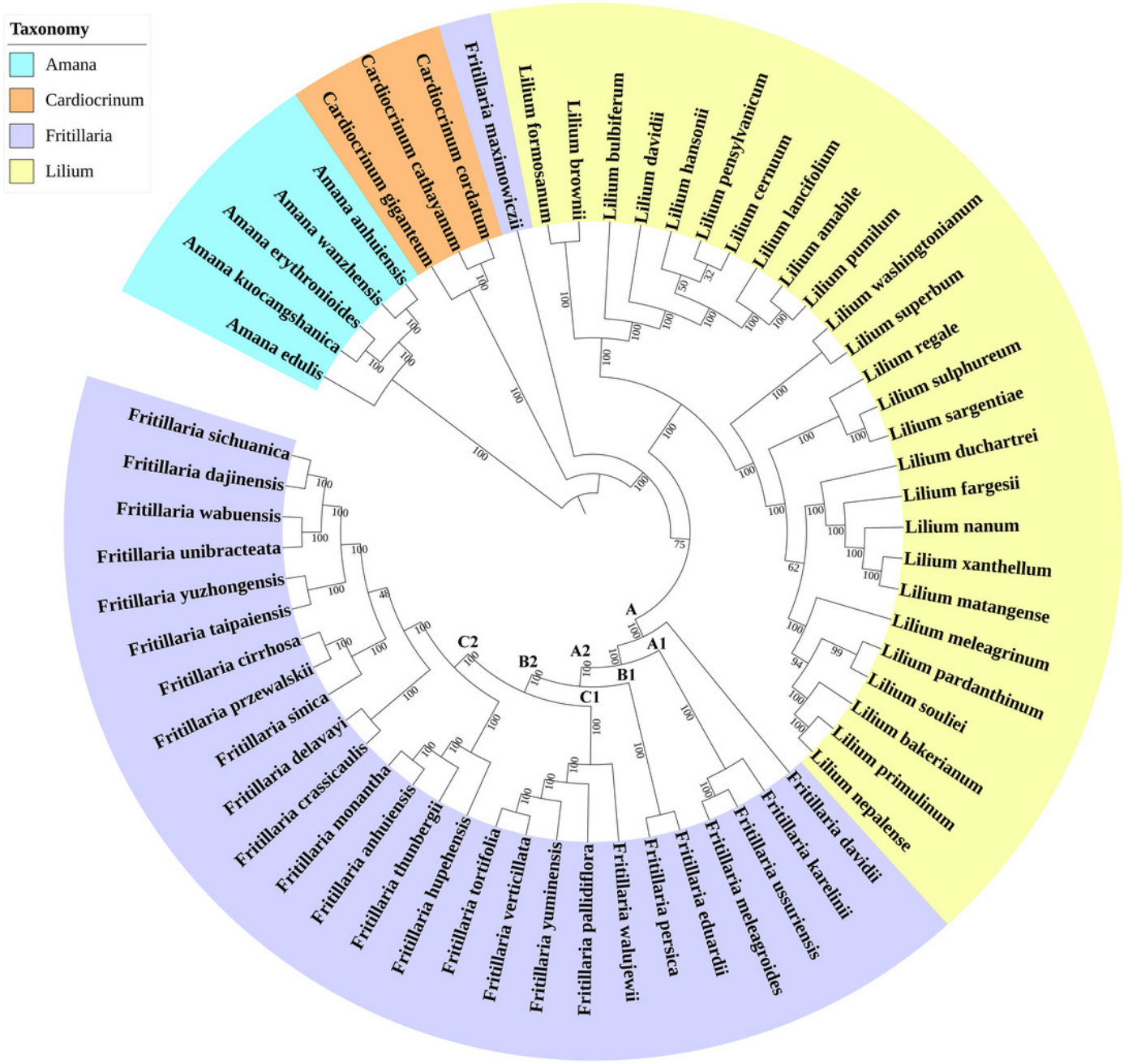


Figure 9

Distribution of 11 medicinal *Fritillaria* species.

Fig 9. Distribution of 11 medicinal *Fritillaria* species. The distribution area of each species is drawn according to the literatures and voucher specimens (http://www.cvh.ac.cn/). Photos of representative living plants of seven *Fritillaria* species Topographic data digital elevation modeling (DEM) data were required from the USGS website (https://glovis.usgs.gov/app?tour) with a 90-m spatial resolution grid.

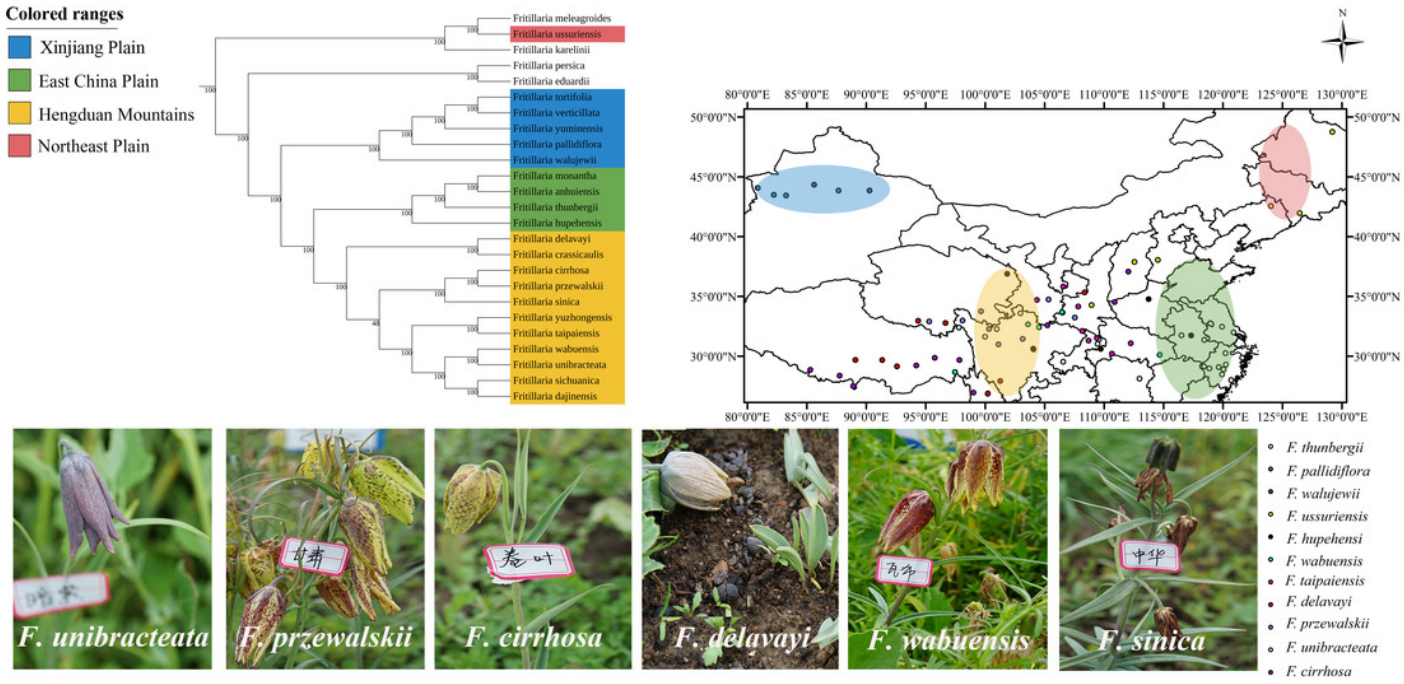
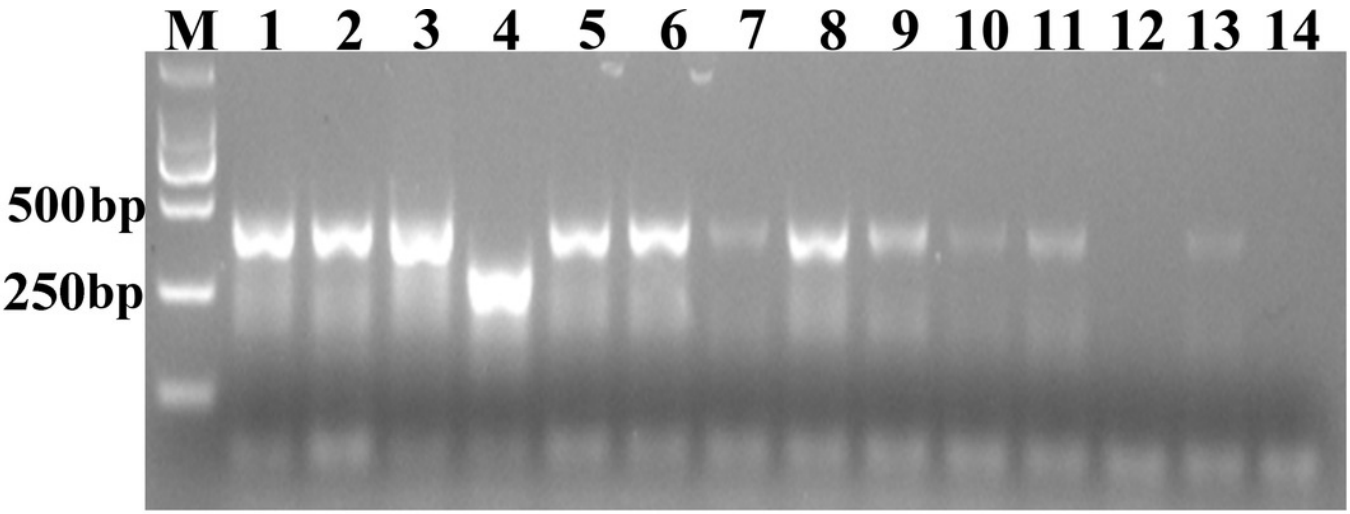


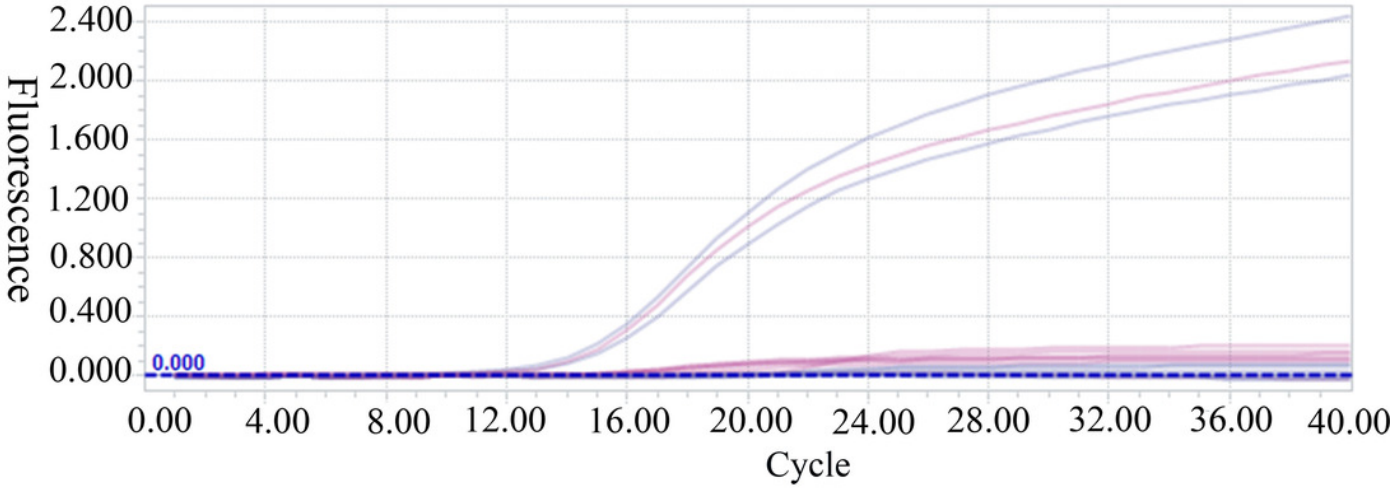
Figure 10

The specie-specific tests for *T. taipaiensis*, *F. unibraceata* and *F. cirrhosa*, respectively.

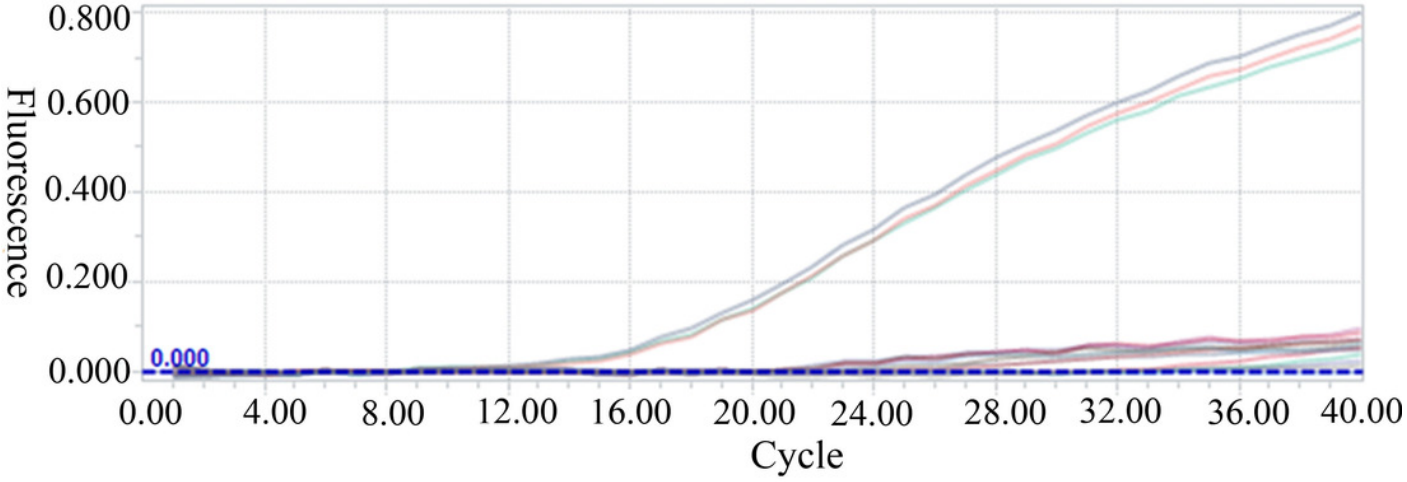
Fig 10. The specie-specific tests for *T. taipaiensis*, *F. unibraceata* and *F. cirrhosa*, respectively. A: specie-specific test for *T. taipaiensis* based on routine PCR. Lane “M” indicates nuclear acid maker. Lane 1 to 14 indicated *F. unibraceata*, *F. delavayi*, *F. cirrhosa*, *F. taipaiensis*, F. *przewalskii*, *F. wabuensis*, *F. sinica*, *F. cirrhosa* var.*ecirrhosa*, *F. pallidiflora*, *F. mellea*, *F. thunbergii*, *F. ussuriensis*, *F. walujewii* and *F. hupehensis*, respectively. B: specie-specific test for *F. unibraceata* based on TaqMan MGB real time PCR. C: specie-specific test for *F. cirrhosa* based on TaqMan MGB real time PCR.


M 1 2 3 4 5 6 7 8 9 10 11 12 13 14
500bp
250bp

A


Fluorescence
2.400
2.000
1.600
1.200
0.800
0.400
0.000
0.000
0.00 4.00 8.00 12.00 16.00 20.00 24.00 28.00 32.00 36.00 40.00
Cycle

B


Fluorescence
0.800
0.600
0.400
0.200
0.000
0.000
0.00 4.00 8.00 12.00 16.00 20.00 24.00 28.00 32.00 36.00 40.00
Cycle

C

Table 1(on next page)

table 1-5

Table 1 The PCR reaction system

System or parameter	Routine PCR	TaqMan MGB real-time PCR
Total volume	25 μL	20 μL
Positive and reverse primer(10 μM)	1.0 μL respectively	1.0 μL respectively
2×Master Mix	12.5 μL	-
2×T5 Fast qPCR Mix (Probe)	-	10 μL
DNA template	2.0 μL	2.0 μL
ddH_2O	8.5 μL	5 μL
TaqMan-MGB Probe (10 μM)	-	1.0 μL
Pre-denaturation	95 °C for 10min, 1 cycle	95 °C for 2 min, 1 cycle
Denaturation	95 °C for 30 s	95 °C for 10 s
Annealing	58 °C for 30 s	60 °C for 60 s, 40 cycles
Extension	72 °C for 50 s, 35 cycles	-
Final extension	72 °C for 7min.	-

Table 2 Summary statistics for the assembly of four *Fritillaria* species chloroplast genomes

Genome features	***F.unibracteata***	***F.przewalskii***	***F.delavayi***	***F.sinica***
Genome size (bp)	151,076	152,043	151,940	152,016
LSC size (bp)	81,383	81,804	81,661	81,768
SSC size (bp)	17,537	17,539	17,569	17,546
IR size (bp)	26,078	26,350	26,355	26,351
Number of genes	132 (109)	133 (110)	133 (110)	133 (110)
Protein genes [unique]	86 (77)	87 (78)	87 (78)	87 (78)
tRNA genes [unique]	38 (28)	38 (28)	38 (28)	38 (28)
rRNA genes [unique]	8 (4)	8 (4)	8 (4)	8 (4)
Duplicated genes in IR	39	39	39	39
GC content (%)	36.96	36.94	36.96	36.95
GC content in LSC (%)	34.79	34.77	34.80	34.79
GC content in SSC (%)	30.42	30.44	30.39	30.45
GC content in IR (%)	42.55	42.46	42.49	42.47
Total reads	23,755,399	26,831,529	25,258,295	26,585,105
Aligned paired-end reads	546,756	652,632	511,467	471,385
Assembled reads	149,891	150,858	150,755	150,831
Average organelle coverage	1081.3173	1291.608	1013.4826	933.6238
Average insert size (bp)	322.99	331.55	341.68	336.64

7 Table 3 List of annotated genes in four CP genomes

Category	Group of gene	Name of gene
Photosynthetic	Subunits of photosystem I	*psaA, psaB, psaC, psaI,psaJ*
	Submits of photosystem II	*psbA, psbB, psbC, psbD, psbE, psbF, psbH psbI(*2), psbJ, psbK, psbL, psbM, psbT, psbZ*
	Subunits of NADH dehydrogenase	*ndhA, ndhB(*2), ndhC, ndhD, ndhE, ndhF, ndhG, ndhH,ndhI, ndhJ, ndhK*
	Subunits of cytochrome b/f complex	*petA, petB, petD, petG,petL,petN*
	Subunits of ATP synthase	*atpA,atpB,atpE,atpF,atpH,atpI*
	Large subunit of rubisco	*rbcL*
Self-replication	Proteins of large ribosomal subunit	*rpl2(*2), rpl14, rpl16, rpl20, rpl22, rpl23(*2), rpl32, rpl33, rpl36*
	Proteins of small ribosomal subunit	*rps2, rps3, rps4, rps7(*2), rps8, rps11, rps12(*2), rps14, rps15, rps16*, rps18, rps19*
	Subunits of RNA paramerase	*rpoA, rpoB, rpoC1, rpoC2*
	Ribosomal RNAs	*rrn23s(*2), rrn16s(*2), rrn5s(*2),rrn4.5s(*2),*
	Transfer RNAs	*tRNA-UUU, tRNA-UUG, tRNA-UUC, tRNA-UGU, tRNA-UGG, tRNA-UGC(*2), tRNA-UGA, tRNA-UCU, tRNA-UAG, tRNA-UAA, tRNA-GUU(*2), tRNA-GUG(*2), tRNA-GUC, tRNA-GUA, tRNA-GGU, tRNA-GGA, tRNA-GCU, tRNA-GCC, tRNA-GCA, tRNA-GAU(*2), tRNA-GAC(*2), tRNA-GAA, tRNA-CGA, tRNA-CCA, tRNA-CAU(*4), tRNA-CAA(*2), tRNA-ACG(*2), tRNA-AAU*
Biosynthesis	Maturase	*matK*
	Protease	*clpP*
	Envelope membrane protein	*cemA*

	Acetyl-CoA carboxylase	*accD*
	c-type cytochrome synthesis gene	*ccsa*
Unknown function	Conserved hypothetical chloroplast	*ycf1(*2), ycf2(*2), ycf3, ycf4, ycf15(*2)*

8

10 Table 4 Information on 15 intron-containing genes in the chloroplast genome of *F.unibracteata*.

Gene	Location	Exon I (bp)	Intron I (bp)	Exon II (bp)	Intron II (bp)	Exon III (bp)
ycf3	LSC	124	740	230	709	159
clpP	LSC	71	771	294	588	159
tRNA-UUU	LSC	38	2558	36		
tRNA-CGA	LSC	32	666	60		
atpF	LSC	160	769	410		
rpoC1	LSC	432	777	1623		
tRNA-UAA	LSC	35	533	50		
tRNA-AAU	LSC	34	585	60		
rpl2	IRA	394	672	428		
ndhB	IRA	775	648	758		
tRNA-GAU	IRA	36	916	57		
tRNA-UGC	IRA	37	810	36		
ndhA	SSC	553	1037	539		
tRNA-UGC	IRB	38	808	37		
tRNA-GAU	IRB	37	914	58		

11

Table 5 Summary of repeat sequences and SSRs in four *Fritillari*a species.

Species	***F.unibracteata***	***F.przewalskii***	***F.delavayi***	***F.sinica***
SSR loci (N)	75	77	72	76
P1[a]loci (N)	27	29	23	27
P2[b]loci (N)	6	7	7	7
P3[c]loci (N)	34	35	34	35
P4 [d]loci (N)	8	6	7	6
P5 [e]loci (N)	0	0	1	1
Total number	212	212	211	212
LSC	125	124	122	124
SSC	37	38	39	38
IR	50	50	50	50

[a]single-nucleotide SSRs, [b]double-nucleotide SSRs, [c]three-nucleotide SSRs, [d]four-nucleotide SSRs, [e]five-nucleotide SSRs