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] School of Life Science and Engineering, Southwest Jiaotong University, Chengdu, Sichuan, China

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

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

The genus *Fritillaria* has attracted great 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 phylogenetic analyses and development of molecular markers for *Fritillaria* species. Here we determined the complete chloroplast (CP) genomes for *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 to 26,355 bp) separated by the large single-copy (81,383 to 81,804 bp) and small single-copy (17,537 to 17,569 bp) regions. *Fritillaria przewalskii*, *F. delavayi*, and *F. sinica* equivalently encoded 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 absence of the *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* to be useful molecular markers in taxonomic studies owning to their interspecies variations. Based on the comprehensive CP genome data collected from 53 species in *Fritillaria* and *Lilium* genera, a phylogenomic study was carried out with three *Cardiocrinum* species and five *Amana* species as outgroups. The results of the phylogenetic analysis showed that *Fritillaria* was a sister to *Lilium*, and the interspecies relationships within subgenus *Fritillaria* were well resolved. Furthermore, phylogenetic analysis based on the CP genome was proved to be a promising method in selecting potential novel medicinal resources to substitute current medicinal species that are on the verge of extinction.

1 **Identification of evolutionary relationships and DNA markers in the medicinally important genus**

2 ***Fritillaria* based on chloroplast genomics**

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

4 *

5 *[1]School of Life Science and Engineering, Southwest Jiaotong University, Chengdu, Sichuan, China*

6 *[2]Qinghai lvkang Biological Development Co., Ltd., Xining, Qinghai, China*

7 Authors' email address:

8 Tian Zhang: 1215695297@qq.com; Sipei Huang: 691786509@qq.com; Simin Song: 1158216140@qq.com;

9 Meng Zou: 815112542@qq.com; Tiechui Yang: 298237664@qq.com; Weiwei Wang: 1028694049@qq.com;

10 Jiayu Zhou: spinezhou@home.swjtu.edu.cn; Hai Liao: ddliaohai@home.swjtu.edu.cn.

12 * For Correspondence

14 *Corresponding authors' information

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

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

19   **Abstract**: The genus *Fritillaria* has attracted great attention because of its medicinal and ornamental values.

20   At least three reasons, including the accurate discrimination between various *Fritillaria* species, protection and

21   sustainable development of rare *Fritillaria* resources as well as understanding of relationship of some

22   perplexing species, have prompted phylogenetic analyses and development of molecular markers for *Fritillaria*

23   species. Here we determined the complete chloroplast (CP) genomes for *F. unibracteata*, *F. przewalskii*, *F.*

24   *delavayi*, and *F. sinica* through Illumina sequencing, followed by *de novo* assembly. The lengths of the

25   genomes ranged from 151,076 in *F. unibracteata* to 152,043 in *F. przewalskii*. Those CP genomes displayed a

26   typical quadripartite structure, all including a pair of inverted repeats (26,078 to 26,355 bp) separated by the

27   large single-copy (81,383 to 81,804 bp) and small single-copy (17,537 to 17,569 bp) regions. *Fritillaria*

28   *przewalskii*, *F. delavayi*, and *F. sinica* equivalently encoded 133 unique genes consisting of 38 transfer RNA

29   genes, 8 ribosomal RNA genes, and 87 protein coding genes, whereas *F. unibracteata* contained 132 unique

30   genes due to absence of the *rps16* gene. Subsequently, comparative analysis of the complete CP genomes

31   revealed that *ycf1*, *trnL*, *trnF*, *ndhD*, *trnN-trnR*, *trnE-trnT*, *trnN*, *psbM-trnD*, *atpI*, and *rps19* to be useful

32   molecular markers in taxonomic studies owning to their interspecies variations. Based on the comprehensive

33   CP genome data collected from 53 species in *Fritillaria* and *Lilium* genera, a phylogenomic study was carried

34   out with three *Cardiocrinum* species and five *Amana* species as outgroups. The results of the phylogenetic

35   analysis showed that *Fritillaria* was a sister to *Lilium*, and the interspecies relationships within subgenus

36   *Fritillaria* were well resolved. Furthermore, phylogenetic analysis based on the CP genome was proved to be a

37   promising method in selecting potential novel medicinal resources to substitute current medicinal species that

38   are on the verge of extinction.

39

40   **Abbreviations:** CP, chloroplast; IR, inverted repeat; ITS, internal transcribed spacer; LSC, large single copy;

41   SSC, small single copy; SSR, simple sequence repeats

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 are diversity hotspots (Xinjiang Plain, East China Plain, Hengduan Mountains, and Northeast Plain). *Fritillaria* species have attracted much attention because they are widely used in traditional Chinese medicine and sometimes as ornamental plants. Dried bulbs from 11 species used singly or as components in traditional Chinese medicine preparations are recorded on Chinese Pharmacopeia (2020), and they are divided into five main concoctions, including Chuan-Bei-Mu (bulbs of the complex group of *F. cirrhosa*), Yi-Bei-Mu (bulbs of *F. palilidiflora* and *F. walujewii*), Zhe-Bei-Mu (bulbs of *F. thunbergii*), Ping-Bei-Mu (bulbs of *F. ussuriensis*), and Hubei-Bei-Mu (bulbs of *F. hupehensis*). Although the bulb of each original species has its own unique curative effect and bioactive compounds and should be used separately for given purposes in traditional prescription, various *Fritillaria* species are still used indiscriminately in clinical prescription due to their similar morphology and names. Particularly, the morphological traits in the group that includes *F. cirrhosa*, *F. unibracteata*, *F. przewalskii*, *F. delavayi*, *F. taipaiensis*, and *F. wabuensis*, are extremely complex due to several highly variable characteristics including stem length, petal color, capsule winged or not, leaf curling, and scale number (*Luo et al.*, *1996*). Therefore, it is vital to carry out taxonomic identification of various *Fritillaria* species.

Molecular systematics has been widely used to clarify angiosperm phylogeny (*Yang et al.*, *2016*). Firstly, accurate identification (e.g., using DNA markers) has been necessary to discriminate between the *Fritillaria* species and its adulterants. Secondly, since the bulbs of some *Fritillaria* species show great economic values in Asian countries (*Yeum et al.*, *2007*) and have long been used in traditional Chinese medicine, the wild *Fritillaria* populations have experienced a sharp decline due to long-term overharvesting. 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 in China (*Li et al.*, *2018*). DNA markers have been helpful in understanding accurately the genetic diversity and structure of *Fritillaria* population, and thus can be an effective scientific approach for conservation purpose. Thirdly, a better understanding of the relationships within the genus could be of great significance for the medicinal use of *Fritillaria*. Some of the phylogenetically close species might

70   be analyzed for their potential medicinal values to determine if they can be used as substitutes for species that

71   are currently rare. Finally, the phylogenetic positions of some medicinal *Fritillaria* species, such as *F.*

72   *pallidiflora*, *F. wabuensis*, and *F. davidii*, remain elusive. *Fritillaria pallidiflora* has always been considered a

73   member of the subgenus *Fritillaria* by *Rix* (*2001*), whereas *Rønsted et al* (*2005*) linked it to the subgenus

74   *Petillium* based on the results of molecular and morphological analyses. *Fritillaria wabuensis* was firstly

75   discovered and nominated as a new species in *Fritillaria* (*Tang & Yue*., *1983*), but later it was classified as a

76   variant of *F. crassicaulis* (*Luo et al.*, *1996*) and *F. unibracteata* (*Liu et al.*, *2009*), respectively. Therefore, well

77   resolved molecular phylogenies of *Fritillaria*, especially the medicinal species, are necessary.

78   Currently, the genus *Fritillaria* is divided into eight subgenera, including *Liliorhiza* (including species

79   mainly in North America), *Japonica* (including species mainly in Japan), *Fritillaria* (the biggest subgenus),

80   *Rhinopetalum*, *Petilium*, the monotypic *Davidii* (including only *F. davidii*), *Theresia* (only *F. persica*), and

81   *Korolkowia* (only *F. sewerezowi*), by *Rix* (*2001*). At present, despite the frequent usage of nuclear DNA

82   internal transcribed spacer (ITS) and several plastid genome regions (*trnL-trnF*, *matK*, *rbcL*, and *rpl16*) in the

83   classification of this genus, previous studies have found that these markers only provided weak phylogenetic

84   signal. *Rønsted et al* (*2005*), who contributed to the current understanding of evolutionary relationships within

85   *Fritillaria*, investigated the phylogenetic position of 37 *Fritillaria* species in detail using *matK*, *rpl16* intron,

86   and ITS. Consequently, *Fritillaria* was shown to be of two clades, with one clade mainly including species

87   from the North American subgenus *Liliorhiza* and the other clade from the seven remaining subgenera.

88   Consistent with the result of *Rønsted et al* (*2005*), *Khourang et al* (*2014*) revealed that the subgenus *Fritillaria*

89   was a sister to the subgenus *Rhinopetalum* based on the phylogenetic tree constructed using the ITS and *trnL-*

90   *trnF* regions of nine Iranian species. However, *Day et al* (*2014*) showed that the largest subgenus (subgenus

91   *Fritillaria*) appeared to be polyphyletic and formed two clades using *matK* and *rbcL* sequences, with one clade

92   comprising taxa occurring mainly in Europe, the Middle East, Japan, and North Africa, and the other clade

93   comprised taxa distributing in China and Central Asia. In our previous research, various *Fritillaria* species

94   from China were classified as North China group and South China group based on the ITS2 sequences, but

95   57.1 % of those species were not effectively resolved (*Zheng et al.*, *2019*). Recently, *Li et al* (*2014*) presented

96   high-quality chloroplast genome using single molecule real-time sequencing, and suggested that *rps19* gene

97   varied the greatest among various species. However, the noncoding regions showed higher variability and were

98   potential effective molecular markers. Therefore, it was proposed that genomics based on the entire chloroplast

99   genome sequences might help identify molecular markers with higher resolution (*Xue et al.*, *2019*).

100 The chloroplast (CP) genome has been extensively used for understanding phylogenetic relationships and

101 discovering more effective molecular markers, some of which, such as *trnH-psbA, matK*, and *rpl16*, have been

102 used as universal plant DNA barcodes (*Bansala et al., 2018*; V*innersten & Bremer., 2001*). To date, there are

103 23 *Fritillaria* CP genomes that are available in GenBank and they can be used to enhance our understanding of

104 the phylogenetic relationships and to identify molecular markers. Although previous reports (*Bi et al., 2018*;

105 *Chen et al., 2019*; *Chen et al., 2020*; *Huang et al., 2020*; *Park et al., 2017*) performed comparative analyses

106 with *Fritillaria* CP genomes available in GenBank, but species-specific identification has not been reported

107 and the phylogenetic place of some ambiguous species remains elusive. At the initial stage of our study, the CP

108 genomes of three important medicinal species (*F. unibracteata, F. przewalskii*, and *F. delavayi*) and *F. sinica*

109 had not been reported before 2018. The increasing CP genomes may not only provide a better phylogenetic

110 analysis of this genus, but also promisingly promote the development of species-specific identification method

111 in the future. Therefore, the CP genomes of these *Fritillaria* species were determined using the Illumina

112 platform in the present study. The objectives of this study included (1) analyzing the global structural patterns

113 of the four CP genomes and comparing them with the available 23 CP genomes of *Fritillaria*; (2) assessing the

114 phylogenetic relationships of the 11 medicinal species used in traditional Chinese medicine, so as to

115 understand the phylogenetic position of some ambiguous species and find potential medicinal plants; and (3)

116 obtaining candidate DNA markers (repeat sequences, SSRs, divergent regions, and indels).

117

118 **Materials and methods**

119 **Plant material**

120 The fresh leaves of *F. unibracteata, F. przewalskii, F. delavayi* and *F. sinica* were collected from the

121 Huzhu County (36°50′15″N, 101°57′06″E), Xining City, Qinghai Province, respectively. The Huzhu County is

122 located in north Hengduan Mountains and east of the Qinghai-Tibetan Plateau. All samples were immediately

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

Chloroplast genome sequencing and assembly

124
125   Total genomic DNA was isolated from 100 mg of fresh leaves using a modified CTAB method. The DNA

126 concentration (>50 ng μL⁻¹) was measured using a NanoDrop spectrophotometer. The isolated DNA was

127 fragmented into small pieces using sonication. After end reparation and A-tailing, the short DNA fragments

128 were ligated with the Illumina paired-end adaptors. Based on gel electrophoresis, the suitable fragments were

129 purified and selected as templates for next-step PCR amplification to create the final DNA library. The quality

130 and quantity of the DNA library were measured using the Agilent 2100 Bioanalyzer. Finally, the library was

131 sequenced from both the 5' and 3' ends using Illumina NovaSeq6000 PE150 Sequencing platform (Illumina,

132 CA, USA). By use of NGSQCToolkit v2.3.3, the raw reads were filtered to remove the linker sequence and

133 low-quality reads defined as having more than 10% bases with Q-value <20, and thus high-quality clean reads

134 were obtained. The clean reads were then assembled using SPAdes (*Bankevich et al.*, *2012*) 3.10.1

135 (http://cab.spbu.ru/software/spades/) software with CP genome of *F. cirrhosa* as reference (NCBI accession

136 number NC_024728.1). Finally, LSC/IR and SSC/IR junctions were further verified by Sanger sequencing.

137

Genome annotation and sequence alignment

138
139   In order to predict putative gene function, the CDS, rRNA and tRNA genes were aligned using blast

140 v2.2.25 (https://blast.ncbi.nlm.nih.gov/Blast.cgi), HMMER v3.1b2 (http://www.hmmer.org/) and aragorn

141 v1.2.38 (http://130.235.244.92/ARAGORN/), respectively, with *E*-value of 10^{-5}. The OGDRAW

142 (https://chlorobox.mpimp-golm.mpg.de/OGDraw.html) helped to make the CP genome maps of *F.*

143 *unibracteata*, *F. przewalskii*, *F. delavayi*, and *F. sinica*.

144   The vmatch v2.3.0 (http: //www.vmatch.de/) could identify their scattered repetitive sequences (*Askitis &*

145 *Sinha.*, *2010*). MISA v1.0 (MIcroSAtellite identification tool, http://pgrc.ipk-gatersleben.de/misa/misa.html)

146 helped to analyze CPSSR. The mafft v7.310 was used to perform indel identification (*Katoh & Standley.*,

147 *2013*). After using the mafft to align the CP genome sequences, BioEdit software was used to adjust the

148 sequences manually (*Gupta et al.*, *2014*). DanSP v6.0 was used to perform sliding window analysis (step size

149 =200 bp and window length=600 bp) for nucleotide variability (pi) in the whole CP genome (*Rozas et al.*,

150 *2017*).

151

Phylogenetic analysis

The phylogenetic analysis was firstly performed based on *matK*, *psbA-trnH* and *rpl16*, respectively, by use of neighbor-joining (NJ) and maximum-likelihood (ML) methods. Then, the CP genomes in the phylogenetic analysis included the 27 *Fritillaria* species, 26 *Lilium* species, three *Cardiocrinum* species and five *Amana* species. The CP genome evolutionary tree was constructed by BLAST2OGMSA script (https://github.com/fenghen360/BLAST2OGMSA) (*Bi.*, *2018*) and MEGA-X software (*Kumar et al.*, *2018*). Firstly, multi-sequence alignment was conducted using BLAST tool of NCBI (*Johnson et al.*, *2008*), and 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 NJ and ML methods, respectively.

167

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 CP genome using the complete genome sequence of *F. cirrhosa* as the reference. Totally, 471,385 to 652,632 mapping reads yielded an average coverage of 934X to 1292X for each species, generating four near full-length CP genomes that ranged from 151,076 in *F. unibracteata* to 152,043 in *F. przewalskii*. The CP genomes contained identical structures, such as two IR regions (26,078 to 26,355 bp each) that were separated by a LSC region (81,383 to 81,804 bp) and a SSC region (17,537 to 17,569 bp) (Fig 1 and Table S1).

178   A total of 133 genes were annotated, including 87 protein-coding genes (PCGs), 38 tRNA, and 8 rRNA

179   genes. The global gene order and content were identical in the four species, except that *F. unibracteata* was

180   absent of the *rps16* gene. 21 genes were duplicated in the CP genome, including eight tRNA genes, four rRNA

181   genes, and nine PCGs. There were 13 genes containing introns, among which *clpP* and *ycf3* each had two

182   introns, whereas the other 13 genes each had one intron. Eight, one, four, and two introns were located in the

183   LSC, SSC, IRa, and IRb region, respectively (Table 1 and Table S2). Table S2 listed the 15 intron-containing

184   genes in the CP genome of *F. unibracteata*, and those of *F. przewalskii*, *F. delavayi*, and *F. sinica* were

185   included in Table S3, S4, and S5, respectively.

186   Four *Fritillaria* species showed high sequence similarity (> 90% identity). IR regions showed a lower

187   level of sequence divergence than LSC and SSC regions. Contraction and expansion of IR regions, especially

188   the boundary region, are important aspects of CP genomes, which are the main reason of length variation in

189   these genomes (*Abdullah et al.*, *2020*). As shown in Fig 2, these 12 *Fritillaria* species had the same gene

190   contents and arrays in IR regions that were expanded in *rps19* and *ycf1* genes. The *rps19* gene in the 12

191   *Fritillaria* species crossed the LSC/IRb boundary and showed the same length of 279 bp which was similar to

192   that of *Lilium superbum*, except that *F. cirrhosa* had *rps19* gene of 285 bp. In the LSC region, the length of

193   *rps19* gene ranged from 250 to 268 bp, whereas that of *rps19* gene in the IRb region varied from 11 to 35 bp.

194   Besides, the *rps19* genes lost their protein-coding function due to incomplete gene duplication. The similar

195   event was also observed in the *ycf1* gene at the IRb/SSC border. The *ycf1* gene was largely located in the IRb

196   and extended 16 to 32 bp into the SSC region, whereas the *ycf1* gene in *F. taipaiensis* was fully located in the

197   IRb region, 58 bp from the IRb/SSC boundary. In the SSC/IRa boundary of 12 *Fritillaria* species, *ycf1* was a

198   key gene and almost equally distributed. *Ycf1* gene had an SSC region of 4,320 bp in *F. unibracteata* and *F.*

199   *przewalskii*, but 4,314 bp in *F. delavayi*, *F. sinica*, *F. cirrhosa*, and *F. taipaiensis*, and also had an IRa region

200   of 1,230 bp in all species. By comparing the LSC/IRb, SSC/IRa, and IRa/LSC regions, it was found that there

201   were obvious differences in IRb/SSC regions among the 12 *Fritillaria* species. The *ycf1* gene in *F. taipaiensis*

202   and *F. cirrhosa* did not cross the IRb/SSC boundary, whereas those in the other *Fritillaria* species extended 16

203   to 32 bp into the SSC region, which resulted a 16 to 32 bp overlap with *ndhF* gene.

204

205

Table 1

206

Fig. 1

207

208

Fig. 2

209

210

211

212 **Repeat sequence, simple sequence repeats (SSRs), divergent regions and indels**

213   The length of the repeat sequence ranged mainly from 15 to 20 bp and rarely from 21 to 38 bp among four

214 *Fritillaria* species (Table S6). The repeating sequences were divided into forward repeating and palindrome

215 sequences (including reverse and complementary sequences). The numbers of repeating sequences ranging

216 from 15 to 20 bp of *F. unibracteata* and *F. przewalskii* were more than 487, while those of *F. delavayi* and *F.*

217 *sinica* were less than 350 (Fig S1). The numbers of repeat sequences in *F. przewalskii, F. unibracteata, F.*

218 *sinica,* and *F. delavayi* were 1,200, 976, 656, and 425, respectively. Although repeat sequences with length

219 ranging from 21 to 38 bp were rare, several promising molecular markers were found. For instance, *F.*

220 *unibracteata* had three forward repeats at lengths of 23, 30, and 47 bp, respectively. *Fritillaria delavayi* also

221 contained a palindromic repeat at a length of 54 bp, and *F. przewalskii* contained two forward repeats and a

222 palindromic repeat at a length of 23 bp.

223   We also found 77, 76, 75, and 72 SSRs of at least 10 bp in *F. przewalskii, F. sinica, F. unibracteata,* and

224 *F. delavayi,* respectively (Table S6, Fig 3). These SSRs were mainly located in the LSC region, followed by 50

225 SSRs in IR region, and a few SSRs in the SSC region. The single- and three-nucleotide SSRs were the

226 majority detected in these *Fritillaria* species, the double-and four-nucleotide SSRs were the minority detected,

227 and a few were five-nucleotide SSR. Single- and three-nucleotide repeats in *F. unibracteata, F. przewalskii, F.*

228 *delavayi,* and *F. sinica* together accounted for 81.33%, 83.12%, 79.17%, and 81.58% of SSRs, respectively.

229 The single-nucleotide SSR with eight to nine repeated units were the most abundant and accounted for 53.91%

230 of SSRs (Fig 3). The high variation in numbers of SSRs might provide abundant information for molecular

231   marker studies and plant breeding.

232

Fig. 3

233

234   Using slide window analysis, 16 regions were eventually extracted to calculate the nucleotide variability

235   with pi value ranging from 0.0104 (*rpl12*) to 0.0159 (*ycf1*). The top ten most divergent regions were identified

236   and thus might be used as potential molecular markers for future phylogenetic analysis and species

237   identification in genus *Fritillaria*. These regions included *ycf1*, *trnL*, *trnF*, *ndhD*, *trnN-trnR*, *trnE-trnT*, *trnN*,

238   *psbM-trnD*, *atpI*, and *rps19* (Fig S2). Due to its highest divergence, *ycf1* gene from 11 medicinal *Fritillaria*

239   species was used to test its usefulness as a promising molecular marker for species identification. Based on our

240   results, species-specific molecular markers for *F. ussurinensis*, *F. pallidiflora*, *F. taipaiensis*, *F. walujewii*, *F.*

241   *thunbergii*, *F. hupehensis*, *F. unibracteata*, and *F. delavayi* were found (Fig S3). *Fritillaria ussurinensis* (78

242   SNPs and nine indels) had the highest number of molecular markers, whereas *F. unibracteata* (two SNPs), *F.*

243   *wabuensis* (two SNPs), and *F. delavayi* (two SNPs and two indels) had the least number of molecular markers.

244   However, species-specific markers for *F. cirrhosa* and *F. przewalskii* were not found in the *ycf1* gene. For

245   these two species, other highly divergent regions may provide better information for species-specific

246   identification. Furthermore, due to its highest divergence, *ycf1* gene was used to construct phylogenetic tree in

247   the following section.

248   Also, three indels, including a 137-bp deletion within *accD-psaI*, a 47-bp insertion within *trnG-GCC-*

249   *trnR-UCU* and a 6-bp deletion within intron of *atpF*, were found in *F. taipaiensis*, *F. unibracteata*, and *F.*

250   *cirrhosa* CP genome, respectively. Inspringly, these indels provided a basis of species-specific identification

251   despite the fact that further experiments should be needed in the future.

252

253   **Phylogenetic tree on the basis of CP genomes**

254   Prior to the phylogenetic analysis based on the CP genomes, we attempted to construct phylogenetic trees

255   based on three common DNA barcodes from the CP genomes, including *matK*, *psbA-trnH*, and *rpl16*.

256   Moreover, the *ycf1* gene was also used to construct the phylogenetic tree. As a result, *matK*, *psbA-trnH*, and

257   *rpl16* obtained weakly supported trees, whereas phylogenetic tree based on *ycf1* gene was moderately

258 supported as more than 50% and 60% branches got bootstrap values of more than 90 BP using NJ (Fig S4) and

259 ML methods (Fig S5), respectively.

260   In addition, in comparison with four partial regions, the whole CP genomes obtained highly reliable

261 phylogenetic tree. The CP genome matrix included the 27 *Fritillaria* species and 26 *Lilium* species, with three

262 *Cardiocrinum* species and five *Amana* species as outgroups. On average, (152,099) bp of the CP genome were

263 aligned. The result of ML tree was similar to that of NJ tree (Fig S6). In the ML tree (Fig 4), the ingroup

264 corresponding to *Fritillaria* and *Lilium* was strongly supported (100 BP), and were sisters to *Cardiocrinum*. In

265 this analysis, *Lilium* was monophyletic (100 BP) and was a sister to *Fritillaria*. Furthermore, *Lilium* nested

266 with *Fritillaria* with moderate bootstrap support (75 BP) than that (53 BP) of previous report (*Day et al., 2014*).

267 *Fritillaria*, as the largest subgenus, was paraphyletic and majority of which fell into one strongly supported

268 Eurasain clade (A) except *F. maximowiczii* (subgenus *Liliorhiza*). Within the clade A, *F. davidii* appeared as

269 successive sister taxa to the remaining Eurasian species (100 BP), which split into two well-supported clades.

270 Clade A1 grouped with the monotypic subgenus *Rhinopetalum* (*F. karelinii*) as a sister to two species from

271 subgenus *Fritillaria* (*F. ussuriensia* and *F. meleagroides*), which occurred in North region of China. The sister

272 clade (A2) was composed of the remaining 22 species that could be classified into two subclades (100 BP).

273 Subclade B1 contained subgenus *Theresia* (*F. persica*) and subgenus *Petilium* (*F. eduardii*), which occurred in

274 the Middle East and Central Asia, while subclade B2 comprising subgenus *Fritillaria* included 15 species from

275 South China and five species (*F. tortifolia*, *F. verticillata*, *F. yuminensis*, *F. pallidiflora*, and *F. walujewii*)

276 from Xinjiang Plain (Fig 5). The 11 most valuable species used in traditional Chinese medicine were not in the

277 same monophyletic group, as *F. ussuriensis* was separated from the other 10 species. As a whole, the

278 phylogenetic tree based on the CP genome was highly supported, in which 91% (53 out of 58) branches

279 obtained bootstrap values of more than 90 BP.

280

Fig. 4

281

282

Fig. 5

283

284

285   **Discussion**

286   **The overall structure of the CP genome**

287   With the rapid development of *de novo* (Illumina) sequencing technology, the sequencing of CP genome

288   is now cost-affordable and is much easier compared with the previous Sanger method. Moreover, *de novo*

289   sequencing technology has been widely used in transcriptome assembly for identifying the biosynthetic and

290   regulatory genes in traditional Chinese medicine, such as *Ligusticum chuanxiong* (*Song et al., 2015*) and

291   *Cassia obtusifolia* (*Deng et al., 2018*). Here, four new CP genomes of *Fritillaria* were obtained using *de novo*

292   sequencing technology. The CP genome sizes ranged from 151,076 to 152,043 bp, which were in accordance

293   with those of reported CP genomes, such as *F. ussuriensis* (151,524 bp), *F. taipaiensis* (151,693 bp), and

294   *F.cirrhosa* (151,991 bp). The four CP genomes contained similar genome structures, gene contents, and gene

295   order, which were typical for land plants. Compared with the other three species, the number of *tRNA* and

296   *rRNA* genes were the same, but the number of protein coding genes ranged from 77 to 78 due to the absence of

297   the *rps16* gene in *F. unibracteata*. The absence of *rps16* gene has also been observed in *Brassicaceae*,

298   *Fabaceae*, and *Populus* species (*Jin et al., 2019*). The functional loss of the *rps16* gene from the CP genome

299   could be compensated by the mitochondrial and (or) nuclear-encoded *rps16* gene that could target chloroplast

300   as well as mitochondria (*Ueda et al., 2008*).

301   The highly conserved genomic structure and gene order as well as no rearrangement of the *Fritillaria* CP

302   genomes have been observed in previous reports. The 26 kb of IRs in the *Fritillaria* species was within the size

303   range of most angiosperm CP genomes (20 to 30 kb). The IR/LSC boundaries in the *Fritillaria* and *Lilium*

304   (*Lilium superbum*) CP genomes expanded into the *rps19* gene, which might be a characteristic CP genome

305   structure of *Fritillaria* and its relative genus. Similar expansion was also observed in other taxa from family

306   *Liliaceae*, including *Lilium* (*Kim & Lee., 2004*), *Fritillaria* (*Li et al., 2014*), and *Cardiocrinum* (*Liu et al.,*

307   *2018*). *Li et al* (*2017*) reported that the common location of IR/LSC junctions in the *rps19* gene seemed to be

308   an ancestral symplesiomorphy of *Liliaceae*. Here, similar feature was observed again, and the whole *rps19*

309   gene was located inside the IR in *Smilax china*, *Oncidium gower*, and *Allium chinense*, while in *Hordeum*

310   *vulgare*, the *rps19* gene did not extend into the IR (Fig 2). The similar IRb/LSC boundaries among *Fritillaria*,

311   *Lilium*, and *Cardiocrinum* implicated that these genera were closely related, which coincided with the

312   phylogenetic result based on CP genomes (Fig 4).

313   A careful comparison between repeat sequence and SSR regions revealed significant differences between

314   various *Fritillaria* species, leading to the establishment of specific markers for molecular identification. In this

315   study, a large number of repeat sequences, mainly with lengths ranging from 15 to 20 bp, were detected in the

316   CP genomes of four *Fritillaria* species, consistent with the results on the CP genomes of *Cannabaceae* (*Zhang*

317   *et al.*, *2018*). SSRs have been used for the study of population genetics because of their high variability (*Asaf*

318   *et al.*, *2016*). The high ratio of SSRs in LSC region was also observed in *F. sichuanica* (*Chen et al.*, *2019*). In

319   the CP genomes of *F. unibracteata*, *F. przewalskii*, *F. delavayi*, and *F. sinica*, the contents of A/T repeats were

320   much higher than those of G/C repeats, similar to the results of *Xue et al* (*2019*) and other studies (*Melotto-*

321   *Passarin et al.*, *2011*). Although several variable CP DNA markers, for instance, *matK*, *rpl16*, *atpB*, and *rbcL*,

322   have been used in phylogenetic studies of *Fritillaria*, but they showed small divergence (pi value of 0.00717,

323   0.00571, 0.00391, and 0.00505, respectively) among the 12 *Fritillaria* species. Based on the result of sliding

324   window analysis, the top ten divergent regions were identified with pi value ranging from 0.0116 to 0.0159.

325   These highly divergent regions included *ycf1*, *trnL*, *trnF*, *ndhD*, *trnN*, *atpI*, and *rps19* in the coding region, and

326   *trnN-trnR*, *trnE-trnT*, and *psbM-trnD* in intergenic region. The molecular markers found in the *ycf1* gene

327   indicated that the highly divergent regions provided plentiful information for species-specific identification in

328   the future (Fig S3). Compared with the three common DNA barcodes, the *ycf1* gene generated a more reliable

329   phylogenetic tree and it thus confirmed that highly divergent region was potential molecular markers for future

330   phylogenetic analysis. The highly variable *trnE-trnT* and *ycf1* gene have also been reported by *Li et al* (*2018*),

331   and the *ycf1* gene has been proposed as the most promising plastid DNA barcode of land plants (*Dong et al.*,

332   *2015*).

333

334   **The phylogenetic analysis of medicinal genus *Fritillaria***

335   Compared with partial sequences, the whole CP genome showed higher resolution with more than 91%

336   branches having bootstrap value of 90 BP (Fig. 4). This result was consistent with a previous report (*Ronsted*

337   *et al.*, *2005*) that increasing additional gene regions would help to improve the resolution. As suggested by

338   *Kress et al* (*2005*) and *Ng et al* (*2017*), the whole CP genome was promising as a super DNA barcode to

339   resolve various *Fritillaria* species efficiently. Furthermore, our findings indicated that *Fritillaria* and *Lilium*

340   were evidently sisters, with the closest relative being *Cardiocrinum* in a monophyletic genus (100 BP), similar

341   with the result of *Chen et al* (*2019*). Such phylogenetic tree based on whole CP genome in this study showed

342   similar topology with the previous study (*Rønsted et al.*, *2005*), but with higher resolution. Specially, genus

343   *Fritillaria* was indicated as paraphyletic with higher bootstrap (100 BP) compared to 54 BP and 53 BP in the

344   findings of *Rønsted et al* (*2005*) and *Day et al* (*2014*), respectively.

345   The subgenus *Fritillaria* also appeared to be a paraphyletic group, similar to the results of *Day et al*

346   (*2014*). One important medicinal *Fritillaria* species, *F. ussuriensis*, clustered with *F. meleagroides* and formed

347   a sister clade to *F. karelinii* of subgenus *Rhinopetalum*, similar to the results of *Huang et al* (*2020*), *Khourang*

348   *et al* (*2014*), and *Li et al* (*2018*). *Fritillaria ussuriensis* and *F. meleagroides* were frequently considered as

349   members of the large subgenus *Fritillaria* (*Rix 2001*). However, the two species do have some similarities with

350   *F. karalinii* as both of them have small mastoid on filament, which are different from other species in Xinjiang

351   Plain with no mastoid. Similar conflict between molecules and morphology was also observed in other taxa

352   (*Anand et al.*, *2016*). Meanwhile, such mastoid on filament was proposed to be a potential primitive feature,

353   and our results partly supported this hypothesis because *F. karelinii* and *F. meleagroides* diverged early from

354   other medicinal species from Xinjiang Plain, such as *F. pallidiflora* and *F. walujewii*. Subgenus *Theresia* (*F.*

355   *persica*) and *Petilium* (*F. eduardii*) had close relationship and formed monophyletic subclade B1, which was

356   similar to the results of *Day et al* (*2014*) and *Li et al* (*2018*).

357   As shown in Fig 4 and Fig 5, five species from Xinjiang Plain were included in a strongly supported

358   subclade C1 (Fig 4), which was a sister to subclade C2 containing the other 15 species from outside Xinjiang

359   Plain. This signified that the Xinjiang species had a close genetic relationship. All the four species that were

360   distributed in East China Plain, including *F. monantha*, *F. anhuiensis*, *F. thunbergii*, and *F. hupehensis,* nested

361   in a supported subclade (100 BP). The remaining 11 species, including the complex group of *F. cirrhosa*, in

362   another subclade were distributed in Hengduan Mountains (100 BP). The 11 important medicinal *Fritillaria*

363   species were widely distributed in four hotspots, namely Xinjiang Plain, Northeastern China Plain, East China

364   Plain, and Hengduan Mountains. The former two regions constituted hotspots in North China, while the latter

365   two regions constituted hotspots in South China. Interestingly, the eight species in the upper location of the

366 clade originated from Xinjiang Plain and Northeastern China Plain (*F. ussuriensis*), whereas the 12 species in

367 the lower location distributed in East China Plain and Hengduan Mountains region. Consequently, the

368 geographical distribution pattern of the 11 medicinal species appeared to map on the phylogenetic tree,

369 especially by plastid data (*Rønsted et al.*, *2005*). Similar result was also reported by *Li et al* (*2018*), and thus

370 the investigation on the correlation between distribution pattern and phylogenetic relationship was needed in

371 the future.

372   Early in 1987, *F. unibracteata*, *F. cirrhosa*, *F. prezewalskii*, and *F. delavayi* were recorded as national

373 third-class endangered medicinal plants in China (*Konchar et al.*, *2011*). The most important medicinal species

374 showed close relationship to widely cultivated members of subgenus *Fritillaria*, which raised the possibility of

375 the rare species being replaced by those widely cultivated species. Recent analyses have demonstrated that *F.*

376 *crassicaulis*, showing closest relationship with *F. cirrhosa*, has been widely used as the substitution of *F.*

377 *cirrhosa* by people of Naxi nationality and Tibetan since Ming/Qing Dynasty (*Tang & Xue.*, *1992*). These

378 findings highlighted those phylogenetic trees based on the CP genomes were promising in selecting potential

379 novel medicinal species. In the future, those showing close relationship to the important species in traditional

380 Chinese medicine, such as *F. sichuanica*, *F. dajinensis*, *F. yuzhongensi*s, *F. sinica*, and *F. crassicaulis*, should

381 be investigated to determine if these bulbs contain the same bioactive compounds found in the complex group

382 of *F. cirrhosa*.

383

384 **The phylogenetic placement of some *Fritillaria* species**

385   The non-monophyletic trait of subgenus *Fritillaria* indicated the incongruence in classification among

386 some species, similar to the reports by *Rønsted et al* (*2005*) and *Day et al* (*2014*). Although *F. ussurinensis*

387 was regarded as a member of subgenus *Fritillaria*, its splitting from other members of subgenus *Fritillaria* has

388 also been observed by *Chen et al* (*2019*) and *Huang et al* (*2020*). There were several reports of natural

389 interspecific hybrids (e.g., *F. ussurinensis* (*Ruan et al.*, *2004*) and *F. eduardii* (*Wietsma et al.*, *2011*)), which

390 might promote the molecular phylogenetic non-monophyly (*Funk & Omland.*, *2003*). Secondly, *F. davidii* had

391 rice-shaped bulbils, resembling the morphological character of subgenus *Liliorhiza*, and used to be grouped in

392 subgenus *Liliorhiza*. But based on our results, it was distantly related to subgenus *Liliorhiza* and was thus

393   placed in subgenus *Davidii* as described by *Rix* (*2001*). It was suggested that rice-shaped bulbils have

394   independently evolved in *F. davidii* and subgenus *Liliorhiza* due to geographic separation, followed by a loss

395   in some species in Eurasian clade during evolution (*Rønsted et al.*, *2005*). Thirdly, *Rønsted et al* (*2005*) found

396   that *F. pallidiflora* was resolved solely within the *Korolkowia/Petilium/Theresia* clade by combining plasmid

397   *rpl16* and *matK* sequences. Our results demonstrated that *F. pallidiflora* clustered within subgenus *Fritillaria*

398   and was more closely related to *Petilium/Theresia*. The conflict in *F. pallidiflora* was likely to be solved by

399   using whole CP genome instead of separate regions. In addition, based on the CP genome, *F. unibraceata* was

400   a sister to *F. wabuensis* with a divergence of 0.003, which was more than that between *F. sichuanica* and *F.*

401   *dajinensis* (0.002). If *F. sichuanica* and *F. dajinensis* were given at rank of species, it was preferable to follow

402   *Tang & Yue* (*1983*) and to rank *F. wabuiensis* as species instead of rank of variant. However, this result was

403   merely based on the CP genome, the accurate placement of *F. wabuensis* will be kept for further evaluation by

404   nuclear genome comparison although it is extremely difficult to obtain.

405

406   **Conclusion**

407   The CP genomes of the four *Fritillaria* species were useful resources for taxonomic clarification,

408   determination of phylogenetic relationship and development of DNA markers. The phylogenetic tree based on

409   the whole CP genome was reliable since 91% branches obtained bootstrap values of more than 90 BP, and the

410   result supported the monophyly of genus *Lilium*, *Amana* and *Cardiocrinum*, except that the largest genus

411   *Fritillaria* was paraphyletic. The 11 members of subgenus *Fritillaria* that were used in traditional Chinese

412   medicine were split into two clusters since *F. ussuriensis* clustered with *F. meleagroides* and *F. karelinii*. In

413   addition, the phylogenetic tree appeared to reflect a geographic distribution pattern of subgenus *Fritillaria*, and

414   also highlighted the importance of the CP genome in the evolutionary analysis. The most important medicinal

415   species, especially the complex group of *F. cirrhosa*, were found to be close to species that were in widespread

416   cultivation for medicinal and ornamental purposes. Excitingly, those closely related species from subgenus

417   *Fritillaria* might be promising alternatives to balance the improving market and rare resources. Finally, this

418   study provided comprehensive molecular markers that might be valuable for future establishment of species-

419   specific identification in *Fritillaria*.

420

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, Mehmood F, Carlsen MM, Islam M, Waheed MT, Poczai P, Croat TB, Ahmed I. 2020. Complete chloroplast genomes of *Anthurium huixtlense* and *Pothos scandens* (*Pothoideae, Araceae*): unique inverted repeat expansion and contraction affect rate of evolution. *J Mol Evol* **88(7):** 562-574. DOI: 10.1007/s00239-020-09958-w

Anand KK, Jena SN, Chaudhary LB, Singh M. 2016. Conflict between morphological and molecular data: a case study of *Ficus krishnae* (*Moraceae*). *Phytotaxa* **247:** 143-147. DOI: https://doi.org/10.11646/phytotaxa.247.2.7.

Angen Ø, Johannesen TB, Petersen RF, Uldum SA, Schnee C. 2021. Development of a species-specific real-time PCR test for *Chlamydia psittaci* and its employment in the investigation of zoonotic transmission from racing pigeons in Denmark. *Diagn Microbiol Infect Dis* **100(2):** 115341. DOI: 10.1016/j.diagmicrobio.2021.115341.

Asaf S, Khan AL, Khan AR, Waqas M, Kang SM, Khan MA, Lee SM, Lee IJ. 2016. Complete chloroplast genome of *Nicotiana otophora* and its comparison with related species. *Front Plant Sci* **7:** 843. DOI: 10.3389/fpls.2016.00843.

Askitis N, Sinha R. 2010. RepMaestro: scalable repeat detection on disk-based genome sequences. *Bioinformatics* **26(19):** 2368-74. DOI: 10.1093/bioinformatics/btq433.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA.

447   **2012.** SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J*

448   *Comput Biol* **19(5):** 455-77. DOI: 10.1089/cmb.2012.0021.

449   **Bansal S, Thakur S, Mangal M, Mangal AK, Gupta RK. 2018.** DNA barcoding for specific and sensitive

450   detection of *Cuminum cyminum* adulteration in Bunium. *Phytomedicine* **50(15):** 178-183. DOI:

451   10.1016/j.phymed.2018.04.023.

452   **Bi GQ**. **2018**. BLAST2OGMSA. *Available at https://github.com/fenghen360/BLAST2OGMSA*.

453   **Bi Y, Zhang MF, Xue J, Dong R, Du YP, Zhang XH. 2018.** Chloroplast genomic resources for phylogeny

454   and DNA barcoding: a case study on *Fritillaria*. *Sci Rep* **8:** 1184. DOI: 10.1038/s41598-018-19591-9.

455   **Chen Q, Wu XB, Zhang DQ. 2019.** Phylogenetic analysis of *Fritillaria cirrhosa* D. Don and its closely

456   related species based on complete chloroplast genomes. *Peer J* **7:** e7480. DOI: 10.7717/peerj.7480.

457   **Chen Q, Wu XB, Zhang DQ. 2020.** Comparison of the abilities of universal, super, and specific DNA

458   barcodes to discriminate among the original species of *Fritillariae cirrhosae* bulbus and its adulterants.

459   *PLoS One* **15(2):** e0229181. DOI: 10.1371/journal.pone.0229181.

460   **Day PD, Berger M, Hill L, Fay MF, Leitch AR, Leitch IJ, Kelly LJ. 2014.** Evolutionary relationships in the

461   medicinally important genus *Fritillaria* L. (*Liliaceae*). *Mol Phylogenet Evol* **80:** 11-19.   DOI:

462   10.1016/j.ympev.2014.07.024.

463   **Deng Y, Zheng H, Yan Z, Liao D, Li C, Zhou J, Liao H. 2018.** Full-length transcriptome survey and

464   expression analysis of *Cassia obtusifolia* to discover putative genes related to aurantio-obtusin

465   biosynthesis, seed formation and development, and stress response. *Int J Mol Sci* **19:** 2476. DOI:

466   10.3390/ijms19092476.

467   **Dong W, Xu C, Li C, Sun J, Zuo Y, Shi S, Cheng T, Guo J, Zhou S. 2015.** Ycf1, the most promising plastid

468   DNA barcode of land plants. *Sci Rep* **5:** 8348. DOI: 10.1038/srep08348.

469   **Funk DJ, Omland KE. 2003.** Species-level paraphyly and polyphyly: frequency, causes and consequences

470   with insights from animal mitochondrial DNA. *Ann Rev Ecol Evol Syst* **34:** 397-423.

471   **Gupta SK, Padmanabhan BR, Diene SM, Lopez-Rojas R, Kempf M, Landraud L, Rolain JM. 2014.**

472   ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes.

473   *Antimicrob Agents Chemother* **58(1):** 212-20. DOI: 10.1128/AAC.01310-13.

474   **Huang J, Yang LQ, Yu Y, Liu YM, Xie DF, Li J, He XJ, Zhou SD. 2018.** Molecular phylogenetics and

475   historical biogeography of the tribe *Lilieae* (*Liliaceae*): Bi-directional dispersal between biodiversity

476   hotspots in Eurasia. *Ann Bot* **122:** 1245-1262. DOI: 10.1093/aob/mcy138.

477   **Huang J, Yu Y, Liu YM, Xie DF, He XJ, Zhou SD. 2020.** Comparative chloroplast genomics of *Fritillaria*

478   (*Liliaceae*), inferences for phylogenetic relationships between *Fritillaria* and *Lilium* and plastome

479   evolution. *Plants (Basel)* **9(2):** 133. DOI: 10.3390/plants9020133.

480   **Jin DP, Choi IS, Choi BH. 2019.** Plastid genome evolution in tribe *Desmodieae* (*Fabaceae*: *Papilionoideae*).

481   *PLoS One* **14(6):** e0218743. DOI: 10.1371/journal.pone.0218743.

482   **Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. 2008**. NCBI BLAST: a

483   better web interface. *Nucleic Acids Res* **36:** W5-9. DOI: 10.1093/nar/gkn201.

484   **Katoh K, Standley DM. 2013.** MAFFT multiple sequence alignment software version 7: improvements in

485   performance and usability. *Mol Biol Evol* **30(4):** 772-80. DOI: 10.1093/molbev/mst010.

486   **Khourang M, Babaei A, Sefidkon F, Naghavi MR, Asgari D, Potter D. 2014.** Phylogenetic relationship in

487   *Fritillaria* spp. of Iran inferred from ribosomal ITS and chloroplast *trnL-trnF* sequence data. *Biochem*

488   *Syst Ecol* **57:** 451-457. DOI: 10.1016/j.bse.2014.10.001.

489   **Kim KJ, Lee HL. 2004.** Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng*

490   Nees) and comparative analysis of sequence evolution among 17 vascular plants. *DNA Res* **11:** 247-261.

491   DOI: 10.1093/dnares/11.4.247.

492   **Konchar K, Li XL, Yang YP, Emshwiller E. 2011.** Phytochemical variation in *Fritillaria cirrhosa* D. Don

493   (Chuan Bei Mu) in relation to plant reproductive stage and timing of harvest. *Econ Bot* **65:** 283-294. DOI:

494   10.1007/s12231-011-9170-3.

495   **Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH. 2005.** Use of DNA barcodes to identify

496   flowering plants. *P Natl Acad Sci USA* **102:** 8369-8374. DOI: 10.1073/pnas.0503123102.

497   **Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018.** MEGA X: molecular evolutionary genetics analysis

498   across computing platforms. *Mol Biol Evol* **35(6):** 1547-1549. DOI: 10.1093/molbev/msy096.

499   **Li P, Lu RS, Xu WQ, Ohi-Toma T, Cai MQ, Qiu YX, Cameron KM, Fu CX. 2017.** Comparative

500   genomics and phylogenomics of east Asian Tulips (*Amana*, *Liliaceae*). *Front Plant Sci* **8:** 451-463. DOI:

501   10.3389/fpls.2017.00451.

502 **Li Q, Li Y, Song J, Xu H, Xu J, Zhu Y, Li X, Gao H, Dong L, Qian J, Sun C, Chen S. 2014.** High-

503   accuracy *de novo* assembly and SNP detection of chloroplast genomes using a SMRT circular consensus

504   sequencing strategy. *New Phytol* **204(4):** 1041-9. DOI: 10.1111/nph.12966.

505 **Li Y, Zhang Z, Yang J, Lv G. 2018.** Complete chloroplast genome of seven *Fritillaria* species, variable DNA

506   markers identification and phylogenetic relationships within the genus. *PLoS One* **13(3):** e0194613. DOI:

507   10.1371/journal.pone.0194613.

508 **Liu HY, Yu Y, Deng YQ, Li J, Huang ZX, Zhou SD. 2018.** The chloroplast genome of *Lilium henrici*:

509   genome structure and comparative analysis. *Molecules* **23(6):** 1276. DOI: 10.3390/molecules23061276.

510 **Liu ZD, Wang S, Chen SC. 2009.** A taxonomic note of *Fritillaria wabuensis* (*Liliaceae*). *Acta Botanica*

511   *Yunnanica (in Chinese)* **31(2):** 145.

512 **Luo YB, Chen XQ. 1996.** A revision of *Fritillaria* L. (*Liliaceae*) in the Hengduan Mountains and adjacent

513   regions, China (II). *Acta Phytotaxonomica Sinica (in Chinese)* **34(5):** 547-553.

514 **Melotto-Passarin DM, Tambarussi EV, Dressano K, De Martin VF, Carrer H. 2011.** Characterization of

515   chloroplast DNA microsatellites from *Saccharum* spp and related species. *Genet Mol Res* **10:** 2024-2033.

516   DOI: 10.4238/vol10-3gmr1019.

517 **Ng PK, Lin SM, Lim PE, Liu LC, Chen CM, Pai TW. 2017.** Complete chloroplast genome of *Gracilaria*

518   *firma* (*Gracilariaceae*, *Rhodophyta*), with discussion on the use of chloroplast phylogenomics in the

519   subclass *Rhodymeniophycidae*. *BMC Genomics* **18(1):** 40. DOI: 10.1186/s12864-016-3453-0.

520 **Park I, Kim WJ, Yeo SM, Choi G, Kang YM, Piao R, Moon BC. 2017.** The complete chloroplast genome

521   sequences of *Fritillaria ussuriensis* Maxim. and *Fritillaria cirrhosa* D. Don, and comparative analysis

522   with other *Fritillaria* species. *Molecules*, **22(6):** 982. DOI: 10.3390/molecules22060982.

523 **Rix EM. 2001.** Fritillaria: A revised classification together with an updated list of species. *Publication of the*

524   *Fritillaria Group of the Alpine Garden Society* **UK**.

525 **Rønsted N, Law S, Thornton H, Fay MF, Chase MW. 2005.** Molecular phylogenetic evidence for the

526   monophyly of *Fritillaria* and *Lilium* (*Liliaceae*; *Liliales*) and the infrageneric classification of *Fritillaria*.

527   *Mol Phylogenet Evol* **35(3):** 509-27.   DOI: 10.1016/j.ympev.2004.12.023.

528 **Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-**

529 **Gracia A. 2017.** DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol Biol Evol*

530 **34(12):** 3299-3302. DOI: 10.1093/molbev/msx248.

531 **Ruan HL, Zhang YH, Pan XC, Dong T, Wu JZ. 2004.** Studies on the chemical constituents from culbs of

532 hybridized Bulbus *Fritillariae ussuriensis*. *Zhongguo Zhong Yao Za Zhi (in Chinese)* **29(4):** 331-334.

533 **Song T, Liu ZB, Li JJ, Zhu QK, Tan R, Chen JS, Zhou YU, Liao H. 2015.** Comparative transcriptome of

534 rhizome and leaf in *Ligusticum Chuanxiong*. *Plant Syst Evol* **301:** 2073-2085.

535 **Tang SY, Yue SJ. 1983.** Three new species of *Fritillaria* Linn. *Acta Academiae Medicinae Sichuan (in*

536 *Chinese)* **14(4):** 327-334.

537 **Tang SY, Yue SJ. 1992.** *Fritillaria* genus, Flora of Sichuan. *Publication of Sichuan Ethnic Publishing House*

538 **7:** 55-82.

539 **Ueda M, Nishikawa T, Fujimoto M, Takanashi H, Arimura S, Tsutsumi N, Kadowaki K. 2008.**

540 Substitution of the gene for chloroplast RPS16 was assisted by generation of a dual targeting signal. *Mol*

541 *Biol Evol* **25(8):** 1566-75. DOI: 10.1093/molbev/msn102.

542 **Vinnersten A, Bremer K. 2001.** Age and biogeography of major clades in *Liliales*. *Am J Bot* **88(9):** 1695-703.

543 **Wietsma WA, van den Berg RG, van Scheepen J, Wieringa JJ 2011.** The nomenclatural history of

544 *Fritillaria eduardii* and the correct names of its varieties. *TAXON* **60(6):** 1754-1759. DOI:

545 10.1002/tax.606018.

546 **Xue S, Shi T, Luo W, Ni X, Iqbal S, Ni Z, Huang X, Yao D, Shen Z, Gao Z. 2019.** Comparative analysis of

547 the complete chloroplast genome among *Prunus mume*, *P. armeniaca*, and *P. salicina*. *Hortic Res-*

548 *England* **6:** 89. DOI: 10.1038/s41438-019-0171-1.

549 **Yang Y, Zhou T, Duan D, Yang J, Feng L, Zhao G. 2016.** Comparative analysis of the complete chloroplast

550 genomes of five Quercus species. *Front Plant Sci* **7:** 959. DOI: 10.3389/fpls.2016.00959.

551 **Yeum HS, Lee YC, Kim SH, Roh SS, Lee JC, Seo YB. 2007.** *Fritillaria cirrhosa*, *Anemarrhena*

552 *asphodeloides*, Lee-Mo-Tang and cyclosporine a inhibit ovalbumin-induced eosinophil accumulation and

553 Th2-mediated bronchial hyperresponsiveness in a murine model of asthma. *Basic Clin Pharmacol Toxicol*

554 **100(3):** 205-13. DOI: 10.1111/j.1742-7843.2007.00043. x.

555   **Zhang HL, Jin JJ, Moore MJ, Yi TS, Li DZ. 2018.** Plastome characteristics of Cannabaceae. *Plant Divers*

556   **40(3):** 127-137. DOI: 10.1016/j.pld.2018.04.003.

557   **Zheng H, Deng KY, Chen AQ, Fu SB, Zhou D, Wang WW, Ni DM, Ren YY, Zhou JY, Liao H. 2019.**

558   Molecular identification and genetic relationship of *Fritillaria cirrhosa* and related species based on DNA

559   barcode. *Acta Pharmaceutica Sinica (in Chinese)* **54(12):** 2326-2334.

560

561

Figure 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.

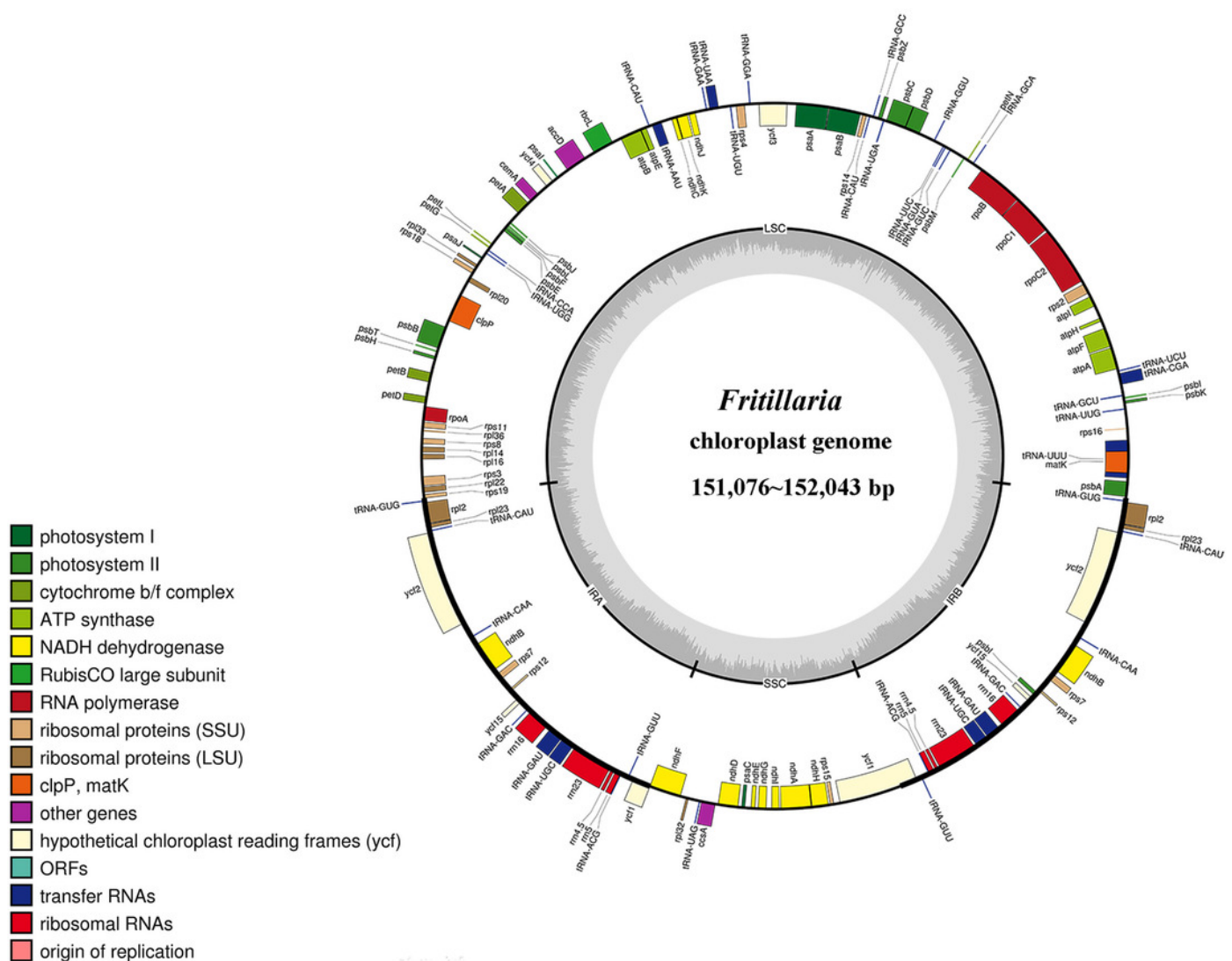



Figure 2

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.

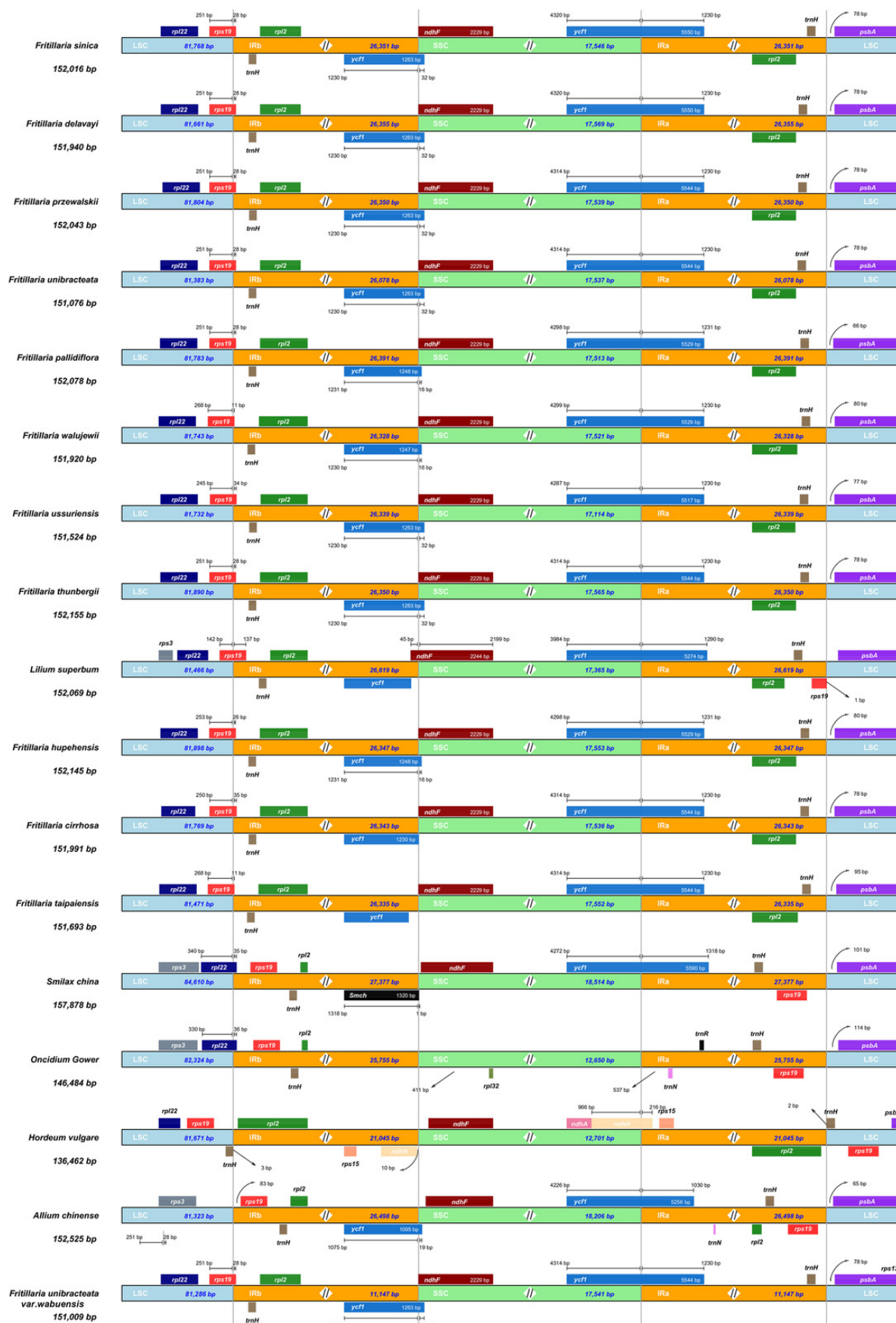



Figure 3

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

Fig 3. Analysis of simple repetitive sequences in four *Fritillaria* CP genomes

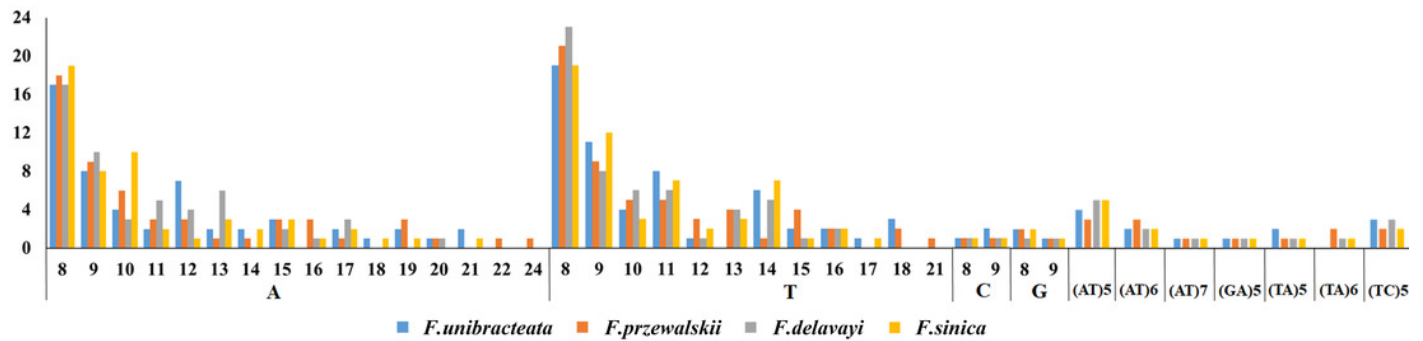



Figure 4

Phylogenetic relationship of 61 species inferred from Maximum Likelihood tree.

Fig 4. Phylogenetic relationship of 61 species inferred from Maximum Likelihood tree.
Numbers above nodes are supporting values with ML bootstrap values.

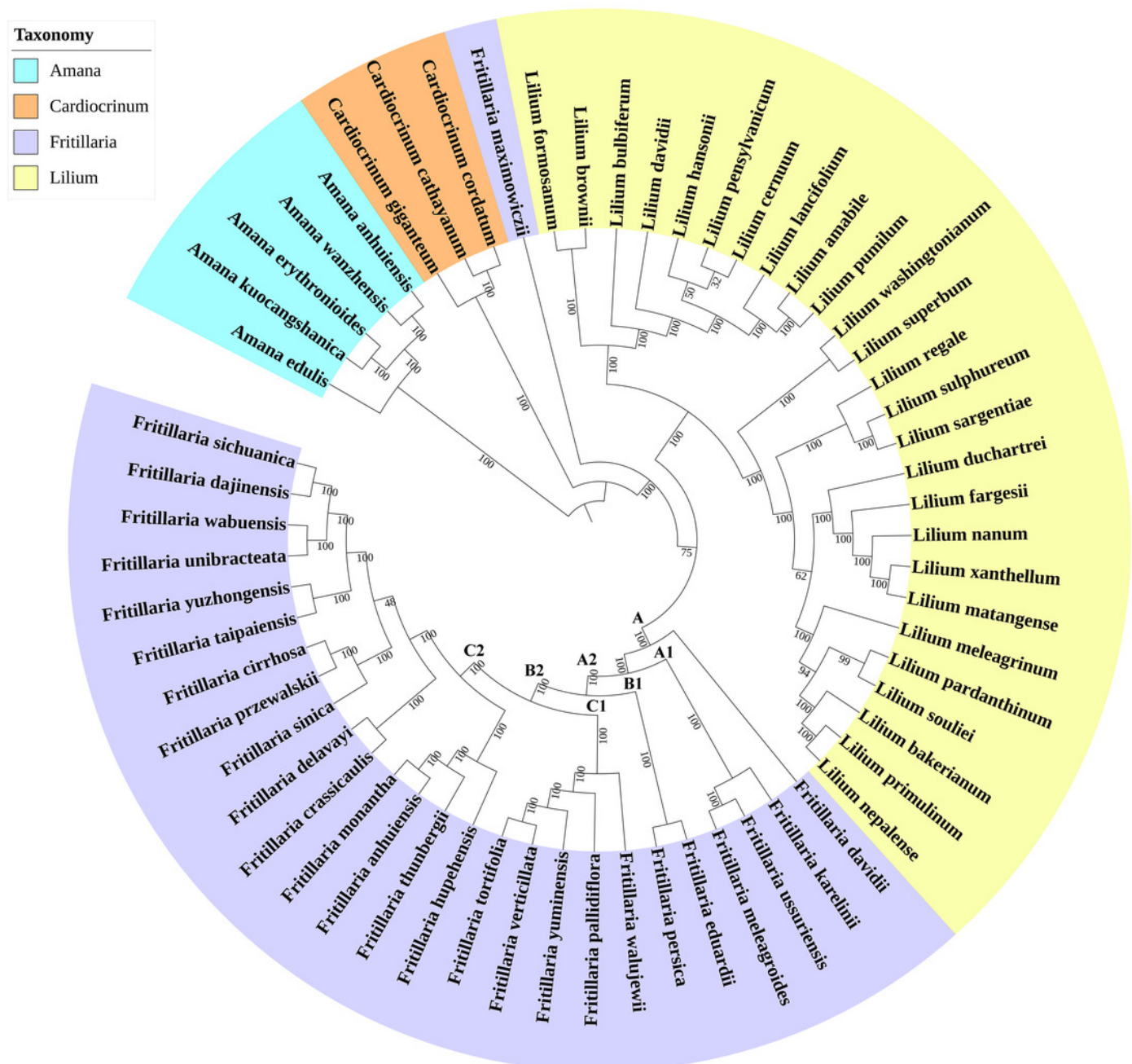
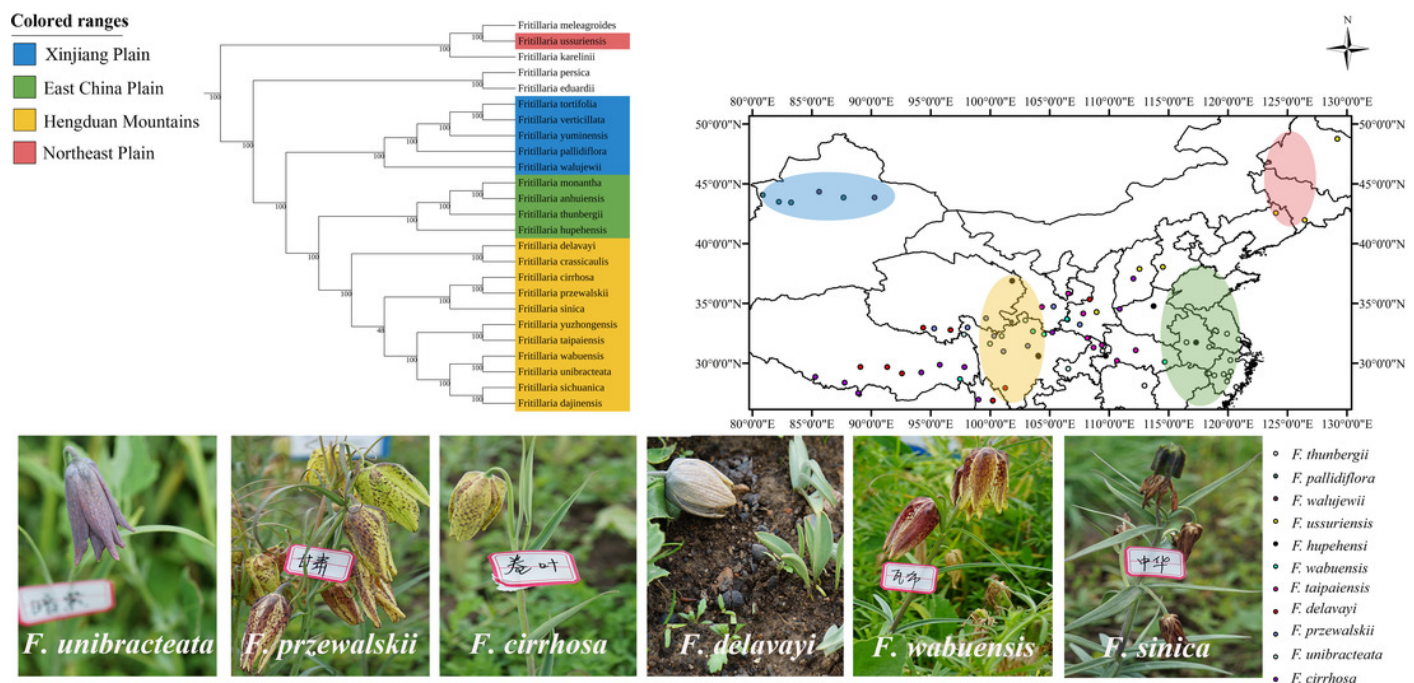



Figure 5

Distribution of 11 medicinal *Fritillaria* species in China.

Fig 5. Distribution of 11 medicinal *Fritillaria* species in China. 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.

List of annotated genes in four CP genomes

1   Table 1 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*
	Subunits 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 polymerase	*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)*