Comparative chloroplast genomics of Fritillaria (Liliaceae), inferences for phylogenetic relationships between Fritillaria and Lilium and genome evolution

Jiao Huang (8037763@qq.com)
Sichuan University https://orcid.org/0000-0002-7491-4679
Yan Yu
Sichuan university
Yan-Mei Liu
Sichuan University
Deng-Feng Xie
Sichuan University
Xing-Jin He
Sichuan University
Song-Dong Zhou
Sichuan University

Research article

Keywords: Fritillaria, Liliaceae, plastome genome, genome structure, phylogenomics

Posted Date: September 10th, 2019

DOI: https://doi.org/10.21203/rs.2.14263/v1

Abstract

Background Fritillaria is a genus consisting of about 140 species that has important medicinal and horticultural values. The monophyly of Fritillaria and phylogenetic relationships with Lilium were previously not fully resolved. The study involved the most comprehensive chloroplast genomes samples to date referring to Old and New World clades of Fritillaria as identified in earlier studies.

Results We reported and compared eleven newly sequenced whole-plastome sequences of Fritillaria as well as characterization of SSRs and repeat sequence. These 11 plastomes proved highly similar in overall size (151,652-152,434bp), genome structure, gene content and order; Comparing them with other species of Liliales (6 out of 10 families) indicated the same similarity but showed some structural variations due to the contraction or expansion of the IR regions out or into of adjacent single-copy regions. A/T mononucleotides, palindromic and forward repeats were the most common types. Six hypervariable regions (rps16 - trnQ, rbcL - accD, accD - psaI, psaJ - rpl33, petD - rpoA and rpl32 - trnL) were discovered based on 26 Fritillaria whole-plastomes to be potential molecular markers. 26 species of Fritillaria plastomes and 21 species of Lilium plastomes were combined in a phylogenomic study with 3 Cardiocrinum species as out groups. Fritillaria was monophyly with moderate support as sister to Lilium based on 64 protein-coding genes (CDS) and the interspecific relationship within subgenus Fritillaria has strong resolution. The topology recovered from the whole plastome and single-copy gene data sets was the same as for coding gene data, but weak support for monophyly of Fritillaria .

Conclusions The phylogenomic analysis reconstructed a topology that had some incongruences with previous studies. The six hypervariable regions can be used as candidate DNA barcodes for global genetic diversity detection of Fritillaria . The phylogenomic framework from this study can guide extensive genomic sampling to further discern the relationships among the Old and New World clades of Fritillaria and Lilium .

Background

Plastid genome of angiosperm have a typical quadripartite structure with a pair of inverted repeat (IR) regions separated by a small single-copy (SSC) region and a large single-copy (LSC) region [1, 2]. In contrast to mitochondrial and nuclear genomes, plastid genome is highly conserved and generally contains 110–130 distinct genes, ranging in size from 120–160 kb [3]. Although plastid genomes are reported as highly conserved sequence and structure in most angiosperms, they have been showed considerable variation in many taxa [2]. These structural alterations most typically involve contraction or extension of the IR region out or into of adjacent SC regions [1], the presence of large inversions or deletions [4, 5], and gene gain and loss events (including pseudogenization) [2, 6]. Some hotspots regions with single nucleotide polymorphisms could be identified and these regions may be used for species identification in terms of enough information [7]. Due to low rates of nucleotide substitutions, lack of recombination and uniparental inheritance, many plastid DNA sequences have been used for inferring plant phylogenies and population genetic analyses [8, 9], such as *matK, rbcL* and *trnH-psbA* [10].

Currently, whole chloroplast genomes have been increasingly used for phylogenetic analyses and inferring phylogeographic histories, as the advent and fast development of next-generation sequencing (NGS) technologies. They can provide plenty of variable sites among their entire size for phylogenetic analyses [3]. Thus, whole chloroplast genomes show the potential for resolving evolutionary relationships and have been employed to generate highly resolved phylogenies and genetic diversity, especially in unresolved relationship of some complex taxon or at low taxonomic levels [2, 11–13]. Because different regions of the whole chloroplast genomes differ with their evolutionary rates, it was preferable for phylogenomic analysis by partition the genome by regions or genes and the concatenated coding genes were by far most widely used [3, 14, 15].

The genus *Fritillaria* L. (Liliaceae) contains 130 to 140 species and is mostly distributed in the temperate regions of the Northern Hemisphere, with centers of speciation and diversity may in the Qinghai-Tiber Plateau (QTP), Irano-Turanian region and the Mediterranean Basin [16, 17]. The bulbs of *Fritillaria* have long been used as herbs in Asian countries, especially in traditional Chinese medicine, and are called “Bei-mu” in Chinese. The bulbs of *Fritillaria* are a botanical source for various pharmaceutically active components and have relieving cough, clearing heat, eliminating phlegm, analgesic and detoxifying effect, moreover, they have been used in some Chinese herbal formulations for treating cancers [18, 19]. As the chemical compounds and pharmacological effects greatly differ among *Fritillaria* species [18], it is crucial to identify accurately them based on medicinal purpose. For this purpose, different barcoding such as plastid DNA (*matK, rpl16, trnL-trnF)* and nuclear DNA internal transcribed spacer (ITS), single nucleotide polymorphisms (SNPs) and genomic inter-simple sequence repeats (ISSRs) have been performed for identifying *Fritillaria* species [18, 20]. *Fritillaria* comprises eight subgenera which are supported by molecular phylogenetic analyses and these subgenera are all monophyletic except subgenus *Fritillaria* [17, 21]. However, the combined sequence analyses of plastid (*matK, rpl16* and *rbcL)* have identified *Fritillaria* was paraphyletic and included the New World and Old World clades; *Lilium* was nested within *Fritillaria* but only with moderate support [17]. Moreover, the ITS analysis with limited species sampling resolved the monophyly of *Fritillaria* but only moderate support [17]. Day et al. attempted to adopt low-copy nuclear gene regions as barcoding, but none of the target regions experimented could be amplified across all species tested16]. These uncertainties about phylogenetic relationships within *Fritillaria* and related genus are most likely due to few genetic barcoding used in previous studies, even though confounding influences such as hybridization in the evolutionary processes cannot be excluded. So far, a total of fifteen plastid genomes of *Fritillaria* have been sequenced and are available on GenBank. These comparative analyses with different *Fritillaria* plastid genomes have provided a basic understanding of the plastid genome characteristics of this genus [22–25].

In this study, we report 11 species whole-plastome sequence data of *Fritillaria* through next-generation sequencing (NGS) and performed a genomic comparative analysis with 15 plastid genome sequences of *Fritillaria* available from Genebank. The specific purposes of this study were to: (1) comparative analysis for the plastome structural patterns and evolution of the eleven *Fritillaria* species; (2) identify highly divergent regions of all 26 *Fritillaria* plastomes as DNA barcodes for future population genetics and species identification; and (3) resolve the evolutionary relationship between two clades of *Fritillaria* and *Lilium.* Overall, this study should be helpful to further understand the plastome evolution and phylogenetic genomics of *Fritillaria.*

Results

Plastome features of *Fritillaria* and comparison with other Liliales taxa

1.02~2.24 G clean base of the eleven species of *Fritillaria* were obtained from Next Generation Sequencing. And the eleven plastomes ranged in size from 151,652bp (*Fritillaria yuzhongensis)* to 152,434bp (*Fritillaria maximowiczii)* (Table 1). All these plastomes showed the typical quadripartite structure akin to other seed plants, consisting of a pair of IRs (26,232–26,574bp) separated by the SSC regions (17,310–17,684bp) and the LSC (81,424–81,976bp). The whole GC content of these 11 plastomes was very similar (36.9–37.1%). The eleven *Fritillaria* plastid genomes contained about 131–136 predicted functional genes, of including 85–90 protein-coding genes, 38 transfer RNA (tRNA) genes, and 8 ribosomal RNA (rRNA) genes (Fig.1 and Table 1& S1). Six-eight protein-coding genes, eight tRNA genes, and all four rRNA genes were duplicated in the IR regions. Among these genes, two CDS (*clpP* and *ycf3)* possessed two introns, while ten CDS genes and six tRNA genes contained a single intron (Table S1). The gene *rps12* was a trans-spliced gene with the 5’ end exon located in the LSC region and the 3’ exon and intron located in the IR regions. Across the five *Fritillaria* plastomes (*F.crassicaulis, F.dajinensis, F.delavayi, F.sichuanica, F.unibracteata),*, four pseudogenes

(*ψinfA, ψycf1, ψycf15* and *ψycf68)* were identified. In these species, *ψycf15* and *ψycf68* were duplicated and the two copies were located in the IR_A and IR_B regions, respectively. The *rps19* gene at the LSC-IR_A border in *F.maximowiczii* was also identified as a pseudogene. In addition, The *F.przewalskii* and *F.yuzhongensis* plastid genome contained two pseudogenes (*ψinfA* and *ψycf1)*.. The genes *ycf68, ycf15* and *infA* contained several internal stop codons and indicated that they may be identified as pseudogenes [9, 22]. Pseudogene for *ycf1* was found at the SSC-IR_B junctions and has lost its protein-coding ability because of incomplete gene duplication. The similar phenomenon was also observed in the *rps19* gene at the LSC-IR_A border.

The IR/SC junction regions with full annotations were compared and had the almost same relative positions among the eleven *Fritillaria* chloroplast genomes (Fig.2). Except for *F.anhuiensis,* all the LSC-IR_B junctions were located within the *rps19* gene, resulting in the IR_B region expanded by a part (28–140bp) toward the *rps19* gene. In *F.anhuiensis,* the junction was located at the intergenic spacer region (IGS) between *rps*19 and *trnH-GUG.* The SSC-IR_A borders in the eleven plastomes were located in the *ycf1* gene and part of this gene was duplicated from 1147–1230bp in the IR_B. In addition, the *ycf1* gene in four species (*F.anhuiensis, F.monantha, F.davidii* and *F.maximowiczii)* extended 24–122bp into the SSC region, which also had a 27–33bp overlap with *ndhF.* In five species (*F.crassicaulis, F.dajinensis, F.delevayi, F.sichuanica, F.unibracteata),*, the SSC-IR_B boundary positions were located at the junction between the pseudogene ψ*ycf1* and *ndhF.* However, for the other two species (*F.przewalskii* and *F.yuzhongensis),*, the distance between ψ*ycf1* and *ndhF* varied from 58bp (Fig.2). The LSC-IR_A border was located at the IGS between *trnH-GUG* and *psbA* in all but one species (*F.maximowiczii)*.. In *F.maximowiczii,* the junction was located in the ψ*rps19-psbA* spacer.

Comparison of the 11 *Fritillaria* plastomes with 10 other species of Liliales showed obvious expansion and contraction of the IRs in several families (Fig.S1). The LSC-IR_B boundary was located in the *rps19* gene in most Liliales species. However, this boundary shifted into the IGS between the *rpl22* and *rps19* genes in Campynemataceae and it expanded a part toward the *rpl22* gene in Smilacaceae. It further shifted into a part of *rps3* in Melanthiaceae (*Paris)*.. Long ψ*ycf1* fragment located at the IR_B varied from 2555 to 992bp, and the IR has expanded ~1.0–2.0kb at the SSC-IR_S border in Colchicaceae and Campynemataceae compared to the most Liliales species (see Table S2).Meanwhile, the IR region had a relative contraction in Melanthiaceae (*Veratrum)* (Fig.S1). Notably, too, the SSC region of Campynemataceae had a different orientation (inversion) relative to the other Liliales species.

Comparative genomic analysis and divergence hotspot regions

We investigated the comprehensive sequence divergence of the 11 *Fritillaria* plastid genomes using mVISTA with *F.cirrhosa* as the reference. The aligned sequences revealed high sequence similarities, and several regions of high sequence length polymorphism were revealed (Fig.S2). As expected, IR regions exhibited comparatively fewer sequence divergence than LSC and SSC regions. We identified 130 regions (18 introns, 53 intergenic spacers, and 59 coding regions) with more than 200 bp in length from the 26 *Fritillaria* plastid genomes. Of the 59 protein-coding regions (CDS), nucleotide variability (Pi) for each locus ranged from 0.00036 *(ndhB)* to 0.0112 (*rps19)* and the average of 0.00453. Thereby seven regions (*matK, rps16, accD, ycf4, rps19, psbH, ndhD)* had remarkably high values (Pi > 0.007; Fig.3A and Table S3). For the 71 noncoding (intergenic spacer and intron) regions, Pi values ranged from 0.0000 (*ycf4-cemA* and *trnV-rrn16)* to 0.05996 (*rps16-trnQ)* and the average of 0.01163. Six of those regions also exhibited considerable high values (Pi > 0.02; i.e. *rps16-trnQ, rbcL-accD, accD-psaI, psaJ-rpl33, petD-rpoA* and *rpl32-trnL;* see Fig.3B and Table S4). The results proved that the IR regions had more sequence conservation than the SC regions, and the average value of Pi in the non-coding regions was more than twice as much as in the coding regions. Specifically, *rbcL-accD, accD-psaI* and *rpl32-trnL* had the highest resolution and similar topology according to the NJ trees. These six

divergence hotspot regions in the noncoding regions should provide a useful resource for future phylogenetic and phylogeographic analyses as well as *Fritillaria* species identification.

SSRs analysis and repeat sequences

With MISA analysis, a total of 848 SSRs were identified across the 11 *Fritillaria* plastomes. Each plastome/species was found to contain the number of SSRs ranged from 72 (*F. delavayi)* to 86 (*F.monantha)*,, with 20 SSRs shared between all 11 plastomes (Table S5, Fig.4A). Six types of SSRs were detected and the overall length ranged from 10 to 23 bp (Fig.4A, table S5). Mono- (all A/T), di- (mostly AT/AT and AG/CT) and tetra-nucleotide (AAAG/CTTT, AAAT/ATTT and AATT/AATT) SSRs were present in each plastome. Tri-nucleotide (AAG/CTT) SSRs were present in nine plastomes except for *F.przewalskii* and *F.yuzhongensis* and (AAT/ATT) SSRs were present in eight plastomes except for *F.dajinensis, F.unibracteata* and *F.sichuanica.* Penta-nucleotide (AATAT/ATATT) SSRs were in seven plastomes except for *F.monantha, F.unibracteata, F. delavayi* and *F.maximowiczii.* By contrast, hexa-nucleotides (AATTAT/AATTAT) were only found in *F.anhuiensis* and *F.monantha* and penta- (AAAAT/ATTTT) and hexa-nucleotides (AGGGAT/ATCCCT) were only in *F.unibracteata.* Furthermore, penta- (AGAGG/CCTCT) nucleotides were only found in *F.davidii,* and di- (AC/GT) and penta-nucleotides (AAAGT/ACTTT) were only in *F.maximowiczii (*Fig.4B). Overall, mono-nucleotide SSRs were most abundant type (68.28%) showed among all 848 SSRs (Fig.4A). In the total 848 SSRs, most loci were situated in the intergenic spacer (IGS) regions, but some were found in *rpoC2, rpoB, accD, cemA, rpl22, ycf2, ndhD, ndhG, ndhH* and *ycf1*coding genes (Table S5). Moreover, most of the SSR loci were situated in the LSC region (75.8%), followed by the SSC region (19.2%) and a minimum in the IR regions (5.0%) (Table S5). In addition, we found 41 polymorphic SSRs between the eleven species of *Fritillaria* (excluding mononucleotide SSRs) (Table S6), which could be useful for further population genetic studies.

We identified the total of 471 repeats including 196 forward, 224 palindromic, 36 reverse and 15 complement in the eleven *Fritillaria* plastomes using REPUTER (Table S7). *F.maximowiczii* and *F.yuzhongensis* possessed the greatest total number of repeats (49 and 48), while *F.unibracteata* and *F.sichuanica* contained the fewest (39) (Fig.4C). Each *Fritillaria* plastid genome contained the number of large repeat sequences ranging from 30 to 57 and repeats with highest proportion ranging in size between 30 and 39 bp (Fig.4D). Repeats situated in homologous regions with the identical lengths were recognized as shared repeats [13, 26]. Under this criterion, there were18 repeats shared by all eleven *Fritillaria* species, 24 repeats shared by ten of the *Fritillaria* species (except for *F.maximowiczii)*,, as well as 27 repeats shared by nine of the *Fritillaria* species (except for *F.maximowiczii* and *F.davidii)*.. Additionally, *F.maximowiczii* owned the most unique repeats (23), whereas four species had no unique repeats, namely *F.monantha, F.sichuanica, F.przewalskii, F.crassicaulis* (Table S7).

Phylogenetic analyses

Both BI and ML methods based on the 64 common CDS shared among the 52 plastomes (28 *Fritillaria,* 21 *Lilium* and 3 *Cardiocrinum* species) generated almost identical topologies (Fig.5). *Lilium* was strongly supported as a monophyletic group (ML bootstrap support, BS = 100%, posterior probability, PP = 1). However, the monophyly of *Fritillaria* was moderately supported (BS = 81%, PP = 0.71). Within the Old World clade of *Fritillaria,* relationships between five subgenera (except subgenus *Korolkowia* and *Japonica)* were fully supported with generally high support values. The phylogenetic trees in this study also indicated subgenus *Fritillaria* was not monophyletic which divided into two clades, and *Fritillaria* clade B formed three monophyletic subclades (*F.pallidiflora, F. thunbergii* and *F.cirrhosa* subclades). This was accord with our previous phylogenetic analysis based on three plastid markers [17]. Within *Lilium,* two strongly supported lineages formed the backbone of the phylogeny, however, the monophyly of four sections (section

Sinomartagon, Martagon, Leucolirion and *Pseudolirium)* were not supported based on our limited samples. In addition, the topologies from the single-copy genes and the whole complete chloroplast genome sequences were similar to that from the CDS sequences, and most lineages possess high bootstrap values. Rather unexpectedly, however, the monophyly of *Fritillaria* was weakly supported (BS = 67%, PP = 0.52) based on 64 single-copy genes (Fig.S3) and not recovered based on whole genome sequences (Fig.S4).

Discussion

Comparison of *Fritillaria* plastomes and phylogenetically informative markers

Our results revealed that the whole plastome sequences newly obtained herein for 11 *Fritillaria* species only slightly vary in size (151,652bp–152,434bp) (Table1), and also largely similar in overall gene content, structure and order (Fig.1). The genome size in our study is accord with previously sequenced plastid genomes of *Fritillaria* [23–25]. Nonetheless, we have been able to document that the CDS genes vary in the plastomes of these eleven species, mainly there are particular (pseudo) genes: *ψycf15, ψycf68* and *ψinfA* in some species. The pseudogenization of these genes are also detected in other angiosperm plastid genomes [7, 9, 26, 27]. It is worth noting that there are only *F.crassicaulis, F.dajinensi*s, *F.delavayi, F. sichuanica, F. unibracteata* having *ψycf15* and *ψycf68* in our results, indicating a few common aspects in their evolutionary processes and functions. The absence or presence of *ψycf15* and *ψycf68* in other *Fritillaria* species has also been reported by Li et al. [25]. The same phenomenon has also been observed in *Ipomoea* [28]. Additionally, the functions of pseudogene *ycf68* and *ycf15* are equivocal in all kinds of land plants [28], for example in Podophylloideae, the *ycf15* gene was detected as non-functional because of the presence of interrupted sequence where large amounts of stop codons were located [2]. The *infA* gene, which codes for translation initiation factor 1 and aids to assemble the translation initiation complex [29, 30] was found to be a pseudogene in seven *Fritillaria* plastid genomes. However, it has been totally lost from the other four *Fritillaria* plastid genomes (Table S1). Similar phenomena have been also observed in other angiosperm plastid genomes, for instance those of *Primula* species, only the plastid genome of *Primula poissonii* contains the *ψ*infA [31]. Our study indicates that the pseudogenization of *infA* may be a synapomorphy for the *F.cirrhosa* subclade (Fig.5)

The 11 plastomes sequenced in our study are also largely similar in overall gene content and structure when compared with some previously published plastomes in Liliales [9, 26, 32–37]. However, the LSC-IR borders of Smilacaceae, Melanthiaceae (*Paris)* and Campynemataceae vary from those of *Fritillaria (*and other Liliales) by showing expansions (Fig.S1), which may assist to stabilize the structure of the entire plastome as well as the prevention of gene gain and gene loss phenomenon [2, 38–40]. By contrast, the SSC-IR boundaries of Melanthiaceae (*Veratrum)* feature contractions when compared to *Fritillaria* and other Liliales (Fig.S1). Our results included 6 out of 10 families in the order Liliales (Table S2) and are in accordance with those of Do et al. *[*33], who used 4 families in Liliaeles and Doet al.41], who used 5 families in Liliales. Wang et al. [42] suggested that *trnH-rps19* clusters were located in the IR/LSC junctions within Liliales taxa, but different patterns of junction were found in Liliales from the present study (Fig.S1). Furthermore, more studies covering all of 10 families should be performed to investigate the whole evolutionary trends of plastomes in Liliales. In general, the contraction and expansion of IR regions are relatively common evolutionary events in plants and have been used as evolutionary loci for phylogenetic relationships [1, 31, 43–45]. Li et al. reported that IR/LSC junctions expanded into *rps19* and suggested the events seem to be an ancenstral symplesiomorphy of Liliaceae [26]. Here, we also found the feature but there was no obvious phylogenetic implication and further more evidences were needed using sufficient genera of Liliaceae.

The cpDNA sequences from a variety of intergenic spacers (*trnL-trnF)* and genes (*matK, rbcL, rpl16* and *atpB)* have often been used to infer the phylogeny of *Fritillaria* [16, 17, 46–48]. From our results, these frequently used plastid barcodes, such as *atpB, rpl16* and *rbcL,* are among the relatively least informative genes (Fig.3). Therefore, the previous phylogenetic analysis based on these barcodes generated low resolution, especially for deep phylogenetic relationships with short internodes and fast rates, such as subgenus *Fritillaria.* By contrast, our sliding window analyses find six intergenic spacer regions with the highest Pi (>0.02) values from 26 *Fritillaria* species (Fig.3). We suggest these hotspot regions are valuable loci to understand the phylogenetic relationships for *Fritillaria* lineages (subgeneric levels) which have experienced rapid radiation.

Phylogenetic relationships and implications

Phylogenetic analyses of the 64 common plastid protein-coding genes generated a well-resolved phylogenetic tree (Fig.5). *Lilium* sisters to *Fritillaria,* and the monophyly of *Fritillaria* was recovered but only moderate support. Our study differs from the previous results which showed *Fritillaira* was paraphyletic and *Lilium* was nested within *Fritillaria* with moderate support based on three cpDNA fragments [17]. This difference may be attributed to our smaller sampling size. Notably, too, in our single-copy genes and the whole genome-derived phylogeny, the monophyly of *Fritillaria* was weakly supported (BS = 67%, PP = 0.52) (Fig.S3) and not recovered (Fig.S4). Thus, the relationship between the New World and the Old world clades of *Fritillaria* and *Lilium* still remains unsolved. But our results provide further evidence for relationships between the two clades of *Fritillaria* and *Lilium* from phylogenomic analyses based on the most complete plastid genomes sampling by far compared with those of the earlier study by Bi et al. [24] and Li et al. [25]. More extensive plastid genomic sampling should be needed for further resolution. The topologies of five subgenera of the Old World clade within *Fritillaria* were strongly confirmed and subgenus *Fritillaria* still was polyphyletic which has high resolution among species than in previous studies (Fig.5) [16, 17]. Thereby, our results also prove the plastid genome to be a good tool in resolving phylogenetically difficult taxa [13].

Conclusion

Our study is the first to report the eleven whole plastomes of *Fritillaria,* of which organization is described. The comparison of these plastid genomes among each other as well as with other species of Liliales showed high similarities in entire structure and content. The synteny of gene order for eleven *Fritillaria* plastomes is also rather conserved. However, these plastid genomes show some structural variations at the junctions of their four regions due to the expansion or contraction of the IRs, especially obvious in Liliales. The 6 noncoding-cpDNA regions are identified as the fastest evolving loci from the comparison of DNA sequence divergence among 26 *Fritillaria* plastomes. Therefore, these highly variable loci, and 471 repeat sequences from eleven *Fritillaria* plastoms can be used for future phylogenetic and phylogeographic analysis. Furthermore, the plastid genomes SSRs with rich diversity identified herein could be useful for further population genetic studies of *Fritillaria.*

This study demonstrated a new level of phylogenomic sampling, 52 species of Lilieae and identified *Fritillaria* was monophyletic with moderate resolution as sister to *Lilium* based on 64 protein-coding genes (CDS), which showed some incongruencies with previous studies. Future phylogenomic studies require more extensive taxonomic sampling, especially subgenus *Liliorhiza* of genus *Fritillaria* to discern the relationships among the Old and New World clades of *Fritillaria* and *Lilium.* In conclusion, our results will be valuable to understand of the evolutionary relationship between *Fritillaria* and *Lilium,* especially plastid gene evolution of *Fritillaria.*

Methods

Taxon sampling, DNA extraction and sequencing

We generated new plastome data for 11 species of *Fritillaria*. Fresh leaves of each species were collected in the field and dried with silica gel. All samples were obtained in China and field sampling was permitted by Natural Reserves in HeilongJiang, Sichuan, Yunnan, Gansu, Anhui and Zhejiang provinces. The vouchers were identified by Dr. Jiao Huang and were deposited in the Herbarium of Sichuan University (SZ); the voucher information is presented in Table S8. Total genomic DNAs were extracted from leaf material following the protocols using a plant genomic DNA kit (Tiangen Biotech, Beijing, China). The isolated genomic DNA was used to generate average 350 bp paired-end (PE) library according to the Illumina Hiseq platform (Illumina, San Diego, CA, USA), and sequenced by an Illumina genome analyser (Hiseq PE150).

Chloroplast genome assembly, annotation and structural analyses

FastQC v0.11.7 was used to assess the quality of sequenced raw reads [49]. Then, we sieved the chloroplast genome related reads by mapping all raw reads to the chloroplast genome sequences downloaded from NCBI of the genera *Fritillaria* and *Lilium (*38 species). Contigs, assembled from all related reads using SOAPdenovo2 [50], were sorted and joined into a single-draft sequence with *Fritillaria cirrhosa* (KF769143), *Fritillaria thunbergii* (NC034368), *Fritillaria taipaiense (*KC543997) and *Lilium leucanhum* (NC035590) as reference species in the software Geneious v11.0.4 [51]. Some gaps in the assembled plastomes were corrected using Sanger sequencing. The primers were designed by Primer 5.0 (Premier, Canada). Primer synthesis and the sequencing of the polymerase chain reaction products were performed by Sangon Biotech (Shanghai, China). The amplifications and primers are shown (Table S9). Annotations of the complete plastomes were conducted using Geneious v11.0.4. The draft annotation was checked and edited manually following the reference genome to accurately confirm the start/stop codons and the exon/intron borders of genes. Furthermore, the schematic diagram of the circular plastome map was generated utilizing OGDRAW [52]. The annotated 11 plastome sequences are deposited in GenBank (accession numbers MK258138-MK258148), which were reported here for the first time. The gene order and structure of 11 *Fritillaria* plastomes were compared using Geneious v11.0.4. We used the plastome of *Fritillaira crassicaulis* as a representation further to compare with plastomes of other 10 Liliales species (see Table S2).

Genome comparative analysis and identification of hypervariable regions

Chloroplast genome comparisons across the eleven *Fritillaria* species was performed by mVISTA program (http://genome.lbl.gov/vista/mvista/submit.shtml)[53], using *F.cirrhosa (*KF769143) as the reference. For identifying hypervariable regions in *Fritillaria* and facilitate its utilization for future population genetic and species identification studies, multiple sequence alignments of 26 *Fritillaria* species (15 species were downloaded from NCBI; Table S10) were performed in MAFFT v.7 [54], and the software MEGA v.6 was used to adjust manually where necessary [55]. A sliding window analysis was conducted for the sequence alignment was subjected using DNASP v5.10 [56] to evaluate nucleotide diversity (Pi) including all protein-coding and noncoding (intron and intergenic spacer) regions. The extraction under the followed two criteria: (a) mutation site > 0; and (b) an aligned length >200 bp. Neighbor- Joining (NJ) trees were constructed by hypervariable markers in the noncoding regions using MEGA v.6 based on a k2Pdistance model.

Characterization of SSRs and repeat sequences

MISA perl script [57] was used to detect simple sequence repeat (SSR) loci of the 11 *Fritillaria* plastid genomes. The minimum numbers (thresholds) of repeats were 10, 5, 4, 3, 3 and 3 for mononucleotide, dinucleotides, trinucleotides, tetranucleotides, pentanucleotides, and hexanucleotides, respectively. The size and position of repeat sequences are assessed by REPuter [58], including inverted (palindromic), direct (forward), reverse and complement repeats. The constraints set for repeat identification were used: 1) 90% greater sequence identity; 2) hamming distance equal to 3; and 2) a minimum repeat size of 30bp.

Phylogenetic analyses

Phylogenetic analyses were performed for the 28 *Fritillaria* species (11 species sequenced here) and 21 *Lilium* species, using three *Cardiocrinum* species as outgroups based on previous studies [17] (Table S10). The analysis was performed base on an alignment of 64 protein-coding genes from the plastid genomes of the 52 species. The sequences of the 64 common CDS were extracted and aligned using MAFFT v.7 [54]. MEGA v.6 was used to adjust manually where necessary [55]. Topologies were constructed using both Bayesian inference (BI) and maximum likelihood (ML) methods. We used the Akaike Information Criterion (AIC) in JModeltest v.2.1.7 to determine the best-fitting models of nucleotide substitutions [59]. The GTR+I+G model were most suitable for both datasets. ML analyses were performed using RAxML-HPC2 [60] on XSEDE of the CIPRES Science Gateway [61], and statistical node supports were estimated via a bootstrap analysis. Bayesian inference (BI) analyses were conducted in MrBayes v3.2 [62]. Two independent Markov Chain Monte Carlo chains were run simultaneously for five million generations. The trees were sampled every 100 generations with the first 25% of calculated trees discarded as burn-in. The consensus tree was constructed from the remaining trees to estimate posterior probabilities (PPs). Additionally, the complete chloroplast genome sequences and the single-copy genes of the 52 species also were used to perform the phylogenetic analyses, respectively.

Abbreviations

AIC: Akaike Information Criterion; BI: Bayesian Inference; CDS: Protein-coding region; cpDNA: Chloroplast DNA; di-: Dinucleotides; IGS: Intergenic spacer region; IR: Inverted repeat; ISSRs: Inter-simple sequence repeats; ITS: Internal transcribed spacer; LSC: Large single-copy; ML: Maximum Likelihood; Mono-: Mononucleotides; NCBI: National Center for Biotechnology Information; NGS: Next-generation sequencing; NJ: Neighbor-Joining; penta-: Pentanucleotides; Pi: nucleotide variability; PPs: Posterior probabilities; QTP: Qinghai-Tiber Plateau; SC: Single-copy; SNPs: Single nucleotide polymorphisms; SSC: Small single-copy; SSRs: Simple-Sequence Repeats; tRNA: Transfer RNA; rRNA: Ribosomal RNA

Declarations

Acknowledgements

We are grateful to Jin-Bo Tan, Qun-Ying Xiao, Jun-Pei Chen, Chuan Xie, Yi-Qi Deng, Juan Li, Fang-Yu Jin, Hai-Ying Liu, Hao Li and Xin Yang for the help in the collection of the field materials and data treating.

Authors’ contributions

JH, X-JH and S-DZ conceived the ideas; JH and Y-ML contributed to the sampling; JH performed the experiment; JH, YY and D-FX analyzed the data. The manuscript was written by JH. X-JH and S-DZ revised the manuscript. All authors read and approved the final manuscript.

Funding

This work research and publication costs were supported by the National Natural Science Foundation of China (Grant numbers 31570198, 31872647), and the National Infrastructure of Natural Resources for Science and Technology (Grant No. 2005DKA21403-JK), the Science and Technology Basic Work (Grant No. 2013FY112100). The funding agencies had no role in the design of the experiment, analysis, and interpretation of data and in writing the manuscript.

Availability of data and materials

The datasets analysed during the current study are available in the NCBI Genbank repository (http://www.ncbi.nlm.nih.gov), and accession numbers can be found in Additional file 10: Table S10. The newly sequenced complete chloroplast genomes of the 11 *Fritillaria* species have been submitted to the NCBI Genbank repository, and accession number can be found in Additional file 10: Table S10.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

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

[2]College of Life Science, Leshan Normal University, Leshan 614000, Sichuan, China

References

1.Zhu A, Guo W, Gupta S, Fan W, Mower JP. Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates. New Phytologist. 2016; 209(4):1747–1756.

2.Ye WQ, Yap ZY, Li P, Peter HC, Qiu YX. Plastome organization, genome-based phylogeny and evolution of plastid genes in Podophylloideae (Berberidaceae). Molecular Phylogenetics and Evolution. 2018; 127: 978–987.

3.Fu CN, Li HT, Milne R, Zhang T, Ma PF, Yang J, Li DZ, Gao LM. Comparative analyses of plastid genomes from fourteen Cornales species: inferences for phylogenetic relationships and genome evolution. Bmc Genomics. 2017; 18(1):956.

4.Downie SR, Jansen RK. A Comparative Analysis of Whole Plastid Genomes from the Apiales: Expansion and Contraction of the Inverted Repeat, Mitochondrial to Plastid Transfer of DNA, and Identification of Highly Divergent Noncoding Regions. Systematic Botany. 2015; 40(1):336–351.

5.Hsu CY, Wu CS, Chaw SM. Birth of Four Chimeric Plastid Gene Clusters in Japanese Umbrella Pine. Genome Biology & Evolution. 2016; 8(6):1776–1784.

6.Sun Y, Moore MJ, Lin N, Adelalu KF, Meng A, Jian S, Yang L, Li J, Wang H. Complete plastome sequencing of both living species of Circaeasteraceae (Ranunculales) reveals unusual rearrangements and the loss of the ndh gene family. Bmc Genomics. 2017; 18(1):592.

7.Zhou T, Wang J, Jia Y, Li W, Xu F, Wang X. Comparative Chloroplast Genome Analyses of Species in Gentiana section Cruciata (Gentianaceae) and the Development of Authentication Markers. International Journal of Molecular Sciences. 2018; 19(7):1962.

8.Shaw J, Lickey EB, Beck JT, Farmer SB, Wusheng L, et al. The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. American Journal of Botany. 2005; 92:142–166.

9.Lu RS, Li P, and Qiu YX. The complete chloroplast genomes of three *Cardiocrinum* (Liliaceae) species: comparative genomic and phylogenetic analyses. Frontiers in Plant Science. 2017; 7: 2054–2066.

10.Group CPW. A DNA barcode for land plants. Proceedings of the National Academy of Sciences of the United States of America. 2009; 106(31):12794–12797.

11.Hu H, Hu QJ, Al-Shehbaz IA, Luo X, Zeng TT, Guo XY, Liu JQ. Species Delimitation and interspecific relationships of the genus *Orychophragmus(*Brassicaceae) inferred from whole chloroplast genomes. Frontiers in Plant Science. 2016;7: 1826–1836.

12.Zhang SD, Jin JJ, Chen SY, Chase MW, Soltis DE, Li HT, Yang JB, Li DZ, Yi TS. Diversification of Rosaceae since the Late Cretaceous based on plastid phylogenomics. New Phytologist. 2017;214(3): 1355–1368.

13.Xu WQ, Losh J, Chen C, Pan L, Wang RH, Zhao YP, Qiu YX, Fu CX. Comparative genomics of figworts (*Scrophularia* Scrophulariaceae), with implications for the evolution of *Scrophularia* and Lamiales. Journal of Systematics & Evolution. 2018;

14.Xi ZX, Ruhfel BR, Schaefer H, Amorim AM, Sugumara M, et al. Phylogenomics and a posteriori data partitioning resolve the Cretaceous angiosperm radiation Malpighiales. Proceedings of the National Academy of Sciences of the United States of America. 2012; 109: 17519–17524.

15.Zhang N, Wen J, Zimmer EA. Another look at the phylogenetic position of the grape order Vitales: Chloroplast phylogenomics with an expanded sampling of key lineages. Molecular Phylogenetics & Evolution. 2016;101:216–223.

16.Day PD, Berger M, Hill L, Fay MF, Leitch AR, Leitch IJ, Kelly LJ. Evolutionary relationships in the medicinally important genus *Fritillaria* L. (Liliaceae). Molecular Phylogenetics & Evolution. 2014; 80:11–19.

17.Huang J, Yang LQ, Yu Y, Liu YM, Xie DF, et al. Molecular phylogenetics and historical biogeography of the tribe Lilieae (Liliaceae): bi-directional dispersal between biodiversity hotspots in Eurasia. Annals of botany. 2018; 122:1245–1262.

18.Hao DC, Xiao-Jie GU, Xiao PG. Phytochemical and biological research of *Fritillaria* medicinal resources. Chinese Journal of Natural Medicines. 2013; 11(4):330–344.

19.Wang D, Li Z, Zhang L, Atanasov AG, Wang S. Characterization of the Isosteroidal Alkaloid Chuanbeinone from Bulbus of *Fritillaria pallidiflora* as Novel Antitumor Agent In Vitro and In Vivo. Planta Medica. 2016; 82(03):195–204.

20.Altinordu F, Peruzzi L, Yu Y, He XJ. A tool for the analysis of chromosomes: KaryoType. Taxon. 2016;65(3): 586–592.

21.Rix EM. *Fritillaria:* A Revised Classification Together with an Updated List of Species. UK: the *Fritillaria* Group of the Alpine Garden Society Press;2001.

22.Li QS, Li Y, Song JY, Xu HB, Xu J, et al. High-accuracy de novo assembly and SNP detection of chloroplast genomes using a SMRT circular consensus sequencing strategy. New Phytologist. 2014;204:1041–1049.

23.Park I, Kim WJ, Yeo SM, Choi G, Kang YM, 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;22:982–997.

24.Bi Y, Zhang MF, Xue J, Dong R, Du YP, Zhang XH. Chloroplast genomic resources for phylogeny and DNA barcoding: a case study on*Fritillaria.* Scientific Reports. 2018; 8(1):1184–1196.

25.Li Y, Zhang Z, Yang J, and Lv GH. Complete chloroplast genome of seven *Fritillaria* species, variable DNA markers identification and phylogenetic relationships within the genus. Plos One. 2018; 13:e0194613.

26.Li P, Lu R-S, Xu WQ, Ohi-Toma T, Cai MQ, Qiu YX, Cameron KM, Fu CX. Comparative Genomics and Phylogenomics of East Asian Tulips (*Amana,* Liliaceae). Frontiers in Plant Science. 2017;8: 451–463.

27.Raman G, Park S. Analysis of the complete chloroplast genome of a medicinal plant, *Dianthus superbus* var. *longicalyncinus,* from a comparative genomics perspective. Plos One. 2015;10(10):e0141329.

28.Yan L, Lai X, Li X, Wei C, Tan X, Zhang Y. Analyses of the complete genome and gene expression of chloroplast of sweet potato [*Ipomoea batata]*.. Plos One.2015;10(4):e0124083.

29.Millen RS, Olmstead RG, Adams KL, Palmer JD, Lao NT, Heggie L, Kavanagh TA, Hibberd JM, Gray JC, Morden CW. Many parallel losses of infA from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. Plant Cell. 2001;13(3):645–658.

30.Wicke S, Schneeweiss GM, Depamphilis CW, Kai FM, Quandt D. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Molecular Biology. 2011;76(3–5):273–297.

31.Ren T, Yang Y, Zhou T, Liu ZL. Comparative plastid genomes of *Primula* species: sequence divergence and phylogenetic relationships. International Journal of Molecular Sciences. 2018;19(4):1050.

32.Liu J, Qi ZC, Zhao YP, Fu CX, Xiang QY. Complete cpDNA genome sequence of *Smilax china* and phylogenetic placement of Liliales—influences of gene partitions and taxon sampling. Molecular Phylogenetics & Evolution.2012;64(3):545–562.

33.Do HD, Kim JS, Kim JH. Comparative genomics of four Liliales families inferred from the complete chloroplast genome sequence of *Veratrum patulum* O. Loes. (Melanthiaceae). Gene. 2013; 530(2):229–235.

34.Kim JS, Kim JH. Comparative genome analysis and phylogenetic relationship of order Liliales insight from the complete plastid genome sequences of two Lilies (*Lilium longiflorum* and *Alstroemeria aurea*. Plos One. 2013; 8(6):e68180.

35.Mennes CB, Lam VKY, Rudall PJ, Lyon SP, Graham SW, Smets EF, Merckx VSFT. Ancient Gondwana break-up explains the distribution of the mycoheterotrophic family Corsiaceae (Liliales). Journal of Biogeography. 2015; 42(6):1123–1136.

36.Huang Y, Li X, Yang Z, Yang C, Yang J, Ji Y. Analysis of complete chloroplast genome sequences improves phylogenetic resolution in *Paris(*Melanthiaceae). Frontiers in Plant Science. 2016; 7(1): 1797.

37.Nguyen PAT, Kim JS, Kim JH. The complete chloroplast genome of colchicine plants (*Colchicum autumnale* L. and *Gloriosa superba* L.) and its application for identifying the genus. Planta. 2015; 242(1):223–237.

38.Palmer JD, Thompson WF. Chloroplast DNA rearrangements are more frequent when a large inverted repeat sequence is lost. Cell. 1982;29(2):537–550.

39.Xiao H, Jiang N, Schaffner E, Stockinger EJ, Knaap EVD. A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. Science. 2008; 319:1527–1530.

40.Yang M, Zhang XW, Liu GM, Yin YX, Chen KF, et al. The complete chloroplast genome sequence of date palm (*Phoenix dactylifera* L.). Plos One. 2010; 5:e12762.

41.Do HDK, Kim JH. A dynamic tandem repeat in monocotyledons inferred from a comparative analysis of chloroplast genomes in Melanthiaceae. Frontiers in Plant Science. 2017; 8:693.

42.Wang RJ, Cheng CL, Chang CC, Wu CL, Su TM, Chaw SM. Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. Bmc Evolutionary Biology. 2008; 8(1):36.

43.Plunkett GM, Downie SR. Expansion and contraction of the chloroplast inverted repeat in Apiaceae subfamily Apioideae. Systematic Botany. 2000; 25(4):648–667.

44.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 Research. 2005;11(4):247–261.

45.Dong WL, Wang RN, Zhang NY, Fan WB, Fang MF, Li ZH. Molecular evolution of chloroplast genomes of Orchid species: insights into phylogenetic relationship and adaptive evolution. International Journal of Molecular Sciences. 2018; 19(3):716.

46.Rønsted N, Law S, Thornton H, Fay MF, Chase MW. Molecular phylogenetic evidence for the monophyly of *Fritillaria* and *Lilium* (Liliaceae; Liliales) and the infrageneric classification of *Fritillaria.* Molecular Phylogenetics & Evolution. 2005; 35(3):509–527.

47.Turktas M, Aslay M, Kaya E, Ertugrul F. Molecular characterization of phylogenetic relationships in *Fritillaria* species inferred from chloroplast trnL-trnF sequences. Turkish Journal of Biology. 2012; 36(5):552–560.

48.Kim JS, Hong JK, Chase MW, Fay MF, Kim JH. Familial relationships of the monocot order Liliales based on a molecular phylogenetic analysis using four plastid loci: matK, rbcL, atpB and atpF-H. Botanical Journal of the Linnean Society. 2013; 172(1):5–21.

49.Andrews, S. FastQC: A Quality Control Tool for High Throughput Sequence Data. 2015. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed 6 October 2011.

50.Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience. 2012; 1(1):18–18.

51.Kearse M, Moir R, Wilson A, Havas SS, Cheung M, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012. 28:1647–1649.

52.Lohse M, Drechsel O, Kahlau S, Bock R. Organellar Genome DRAW–a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. Nucleic Acids Research. 2013; 41:W575.

53.Frazer KA, Lior P, Alexander P, Rubin EM, Inna D. VISTA: computational tools for comparative genomics. Nucleic Acids Research. 2004.

54.Katoh K and Standley DM. MAFFT multiple sequence alignment software version 7:

improvements in performance and usability. Molecular Biology & Evolution. 2013; 30: 772–780.

55.Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. Molecular Biology and Evolution. 2013; 30:2725–2729.

56.Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009; 25(11):1451–1452.

57.Thiel T, Michalek W, Varshney RK, Graner A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (Hordeum vulgare L.). Theoretical and Applied Genetics. 2003; 106(3):411–422.

58.Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Research. 2001; 29(22):4633–4642.

59.Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. Nature Methods. 2012; 9(8):772–772.

60.Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014; 30(9):1312–1313.

61.Miller MA, Pfeiffer W, Schwartz T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: Proceedings of the Gateway Computing Environments Workshop (GCE), LA:New Orleans; 2010.p. 1–8.

62.Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 2003; 19(12):1572–1574.

Tables

Table 1 Summary of major characteristics of eleven *Fritillaria* plastomes, including aspects of genome size, G-C content, and gene number (per type and location).

Species	Genome size	LSC Length	IR Length	SSC Length	G-C	Number of genes				Number of CDS			
	(bp)	(bp)	(bp)	(bp)	(%)	Total	CDS	rRNAs	tRNAs	LSC	IR_A	SSC	IR_B
F.anhuiensis	152,196	81,950	26,340	17,566	36.9	131[18]	85[6]	8[4]	38[8]	60	7	11	7
F.crassicaulis	151,852	81,615	26,345	17,547	37.0	136[20]	90[8]	8[4]	38[8]	61(1)	9(3)	11	9(2)
F.dajinensis	151,971	81,728	26,351	17,541	37.0	136[20]	90[8]	8[4]	38[8]	61(1)	9(3)	11	9(2)
F.davidii	152,044	81,896	26,232	17,684	37.0	131[18]	85[6]	8[4]	38[8]	60	7	11	7
F.delavayi	151,948	81,683	26,357	17,551	36.9	136[20]	90[8]	8[4]	38[8]	61(1)	9(3)	11	9(2)
F.maximowiczii	152,434	81,976	26,574	17,310	37.1	132[18]	86[6]	8[4]	38[8]	61(1)	7	11	7
F.monantha	152,158	81,895	26,350	17,563	37.0	131[18]	85[6]	8[4]	38[8]	60	7	11	7
F.przewalskii	151,688	81,448	26,352	17,536	37.0	132[18]	86[6]	8[4]	38[8]	61(1)	7(1)	11	7
F.sichuanica	151,967	81,727	26,350	17,540	37.0	136[20]	90[8]	8[4]	38[8]	61(1)	9(3)	11	9(2)
F.unibracteata	151,971	81,730	26,350	17,541	36.9	136[20]	90[8]	8[4]	38[8]	61(1)	9(3)	11	9(2)
F.yuzhongensis	151,652	81,424	26,351	17,526	37.0	132[18]	86[6]	8[4]	38[8]	61(1)	7(1)	11	7

Abbreviations: CDS, protein-coding sequences/genes; LSC, large single-copy region; SSC, small single-copy region; IR, inverted repeat (A or B) regions. Numbers in brackets mean the number of duplicated genes, e.g. 131[18] means there were 131 total genes in the plastome, of which 18 were duplicated in the IRs. Numbers in parentheses mean the number of pseudogenes, e.g. 61(1) means there were 61 unique genes observed in LSC, of which one was pseudogene.

Figures

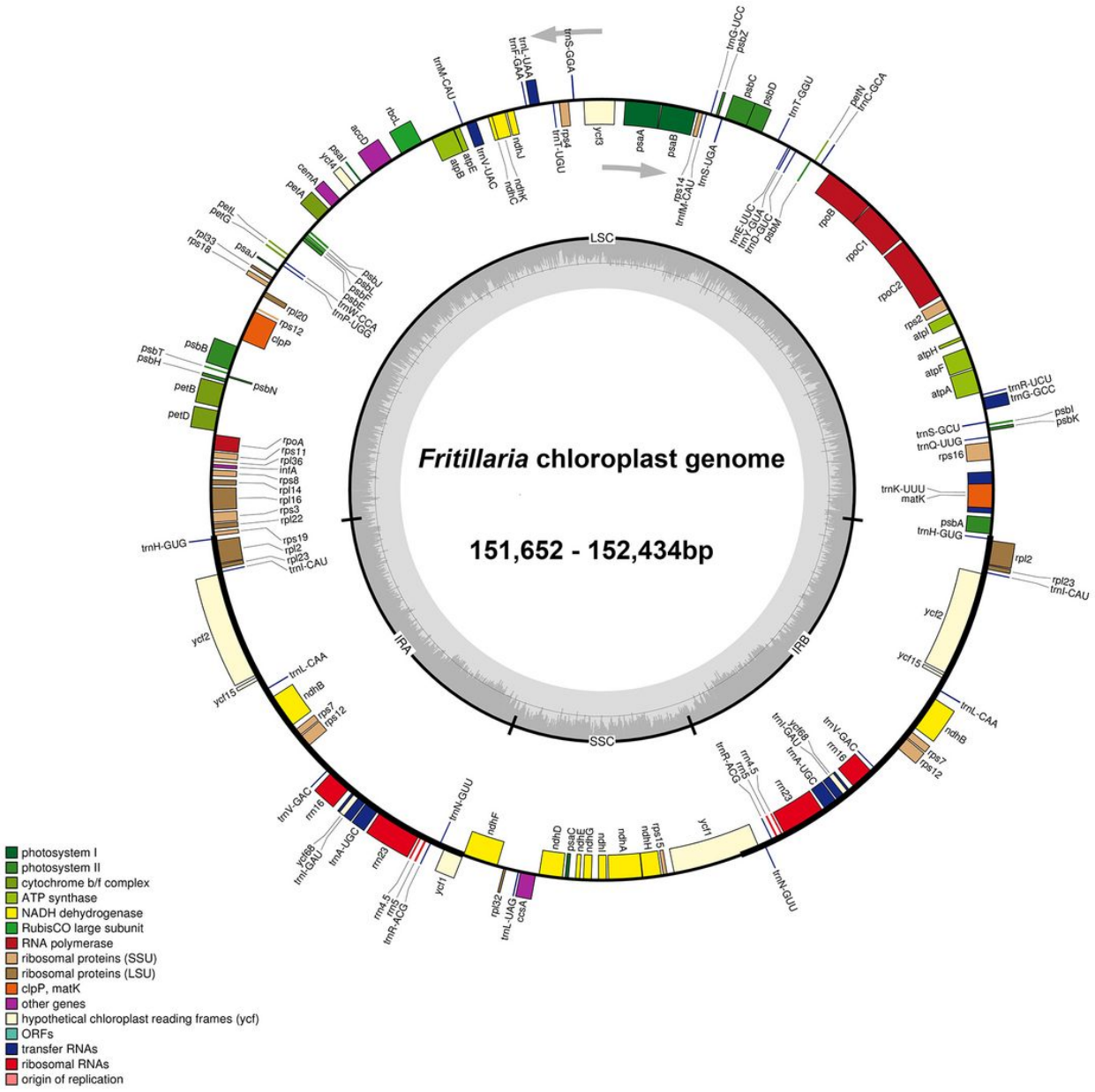

Figure 1

Circular gene map of Fritillaria species (Fritillaria anhuiensis, F. crassicaulis, F.dajinensis, F.davidii, F.delavayi, F.maximowiczii, F.monantha, F.przewalskii, F.sichuanica, F.unibracteata and F.yuzhongensis) chloroplast genomes. The genes inside and outside of the circle are transcribed in clockwise and counterclockwise directions, respectively. Genes belonging to different functional groups are shown in different colors. The dashed darker grey area in the inner circle indicates genome GC content, while the lighter grey area shows AT content. IR=inverted repeat; SSC=small single copy; LSC=large single copy.

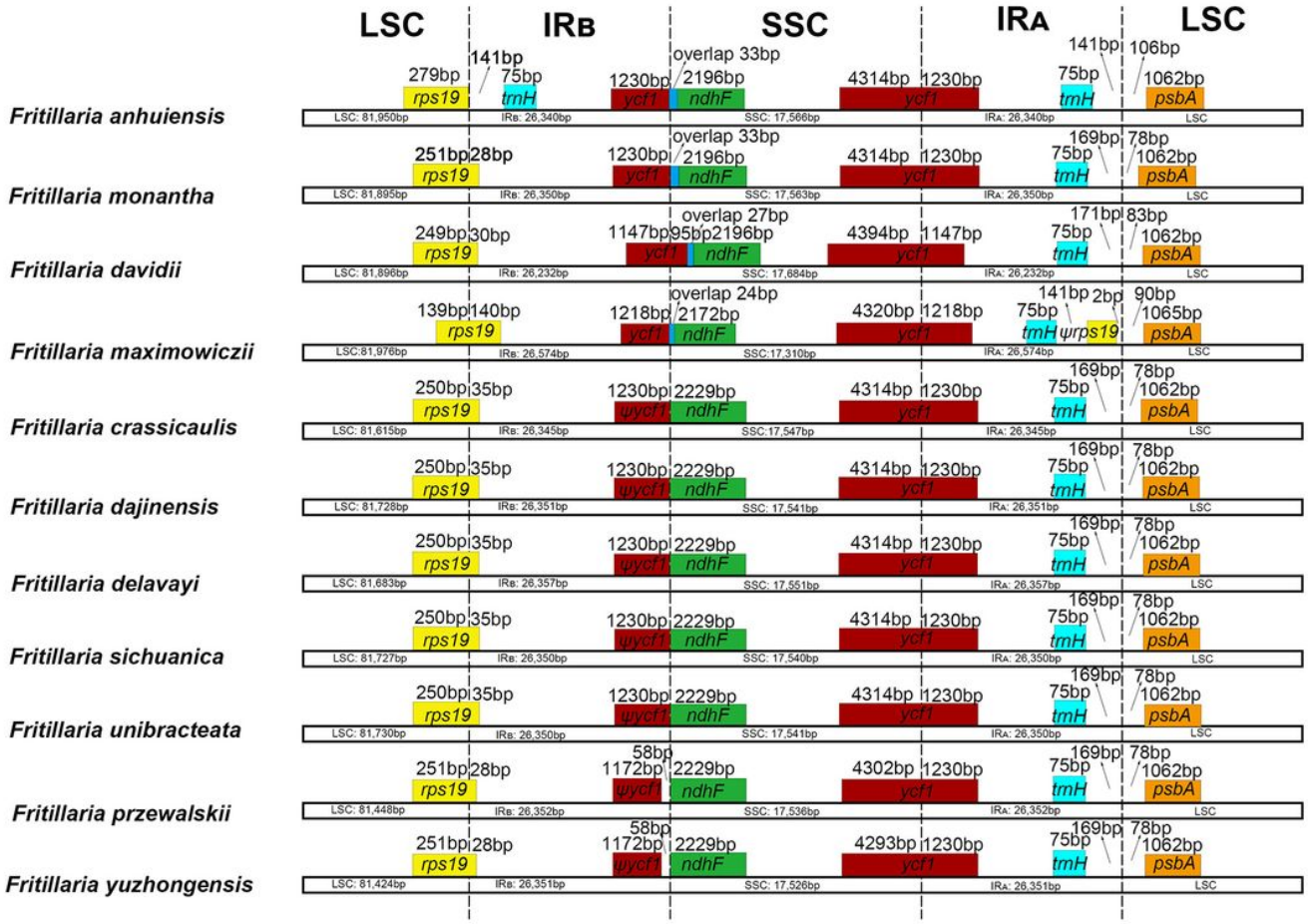

Figure 2

Comparison of LSC, IRs, and SSC junction positions among eleven Fritillaria chloroplast genomes.

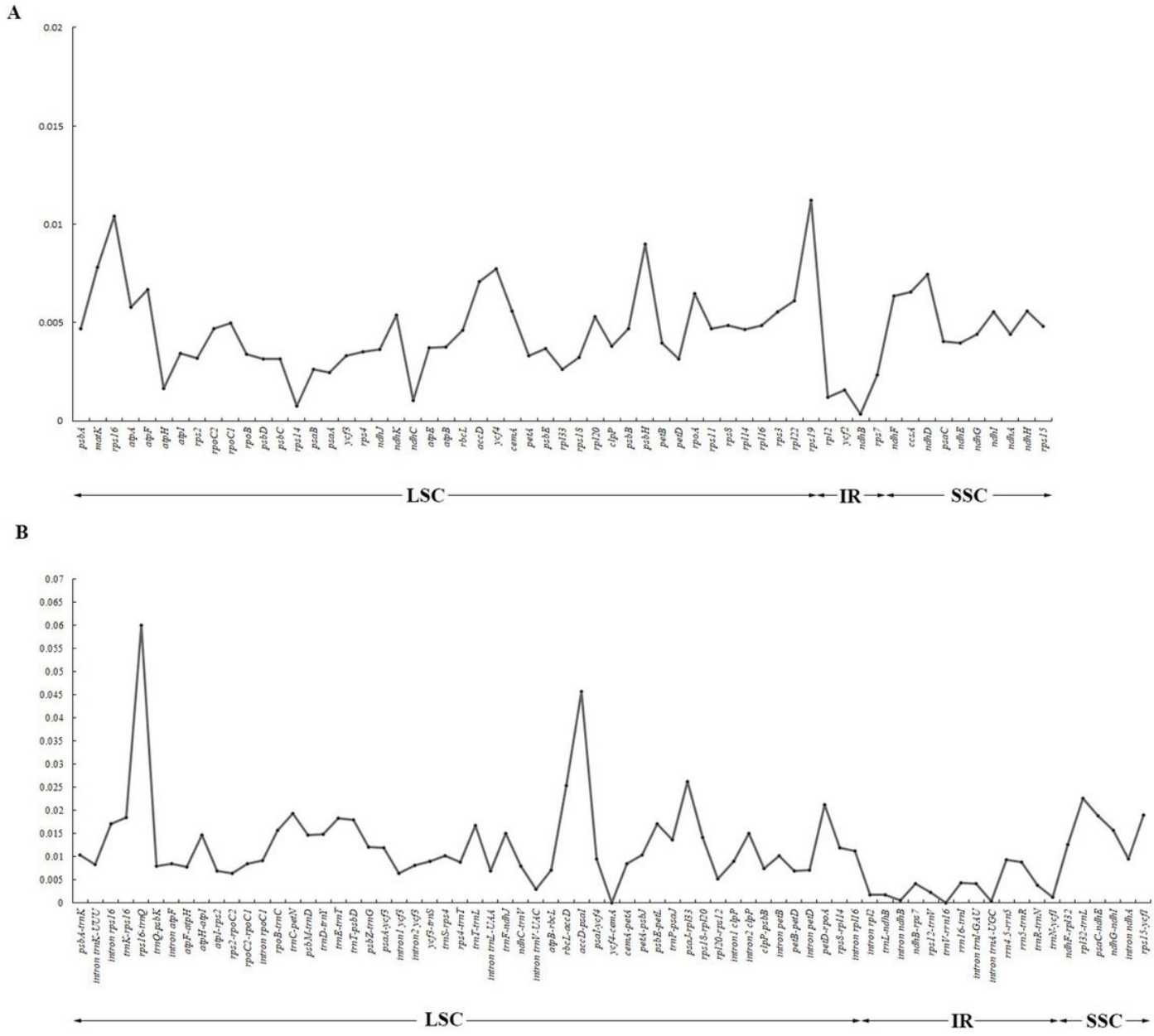

Figure 3

Comparison of nucleotide variability (Pi) values in 26 Fritillaria plastomes. (A) Pi values among protein-coding genes (CDS); (B) Pi values among intergenic spacer (IGS) regions.

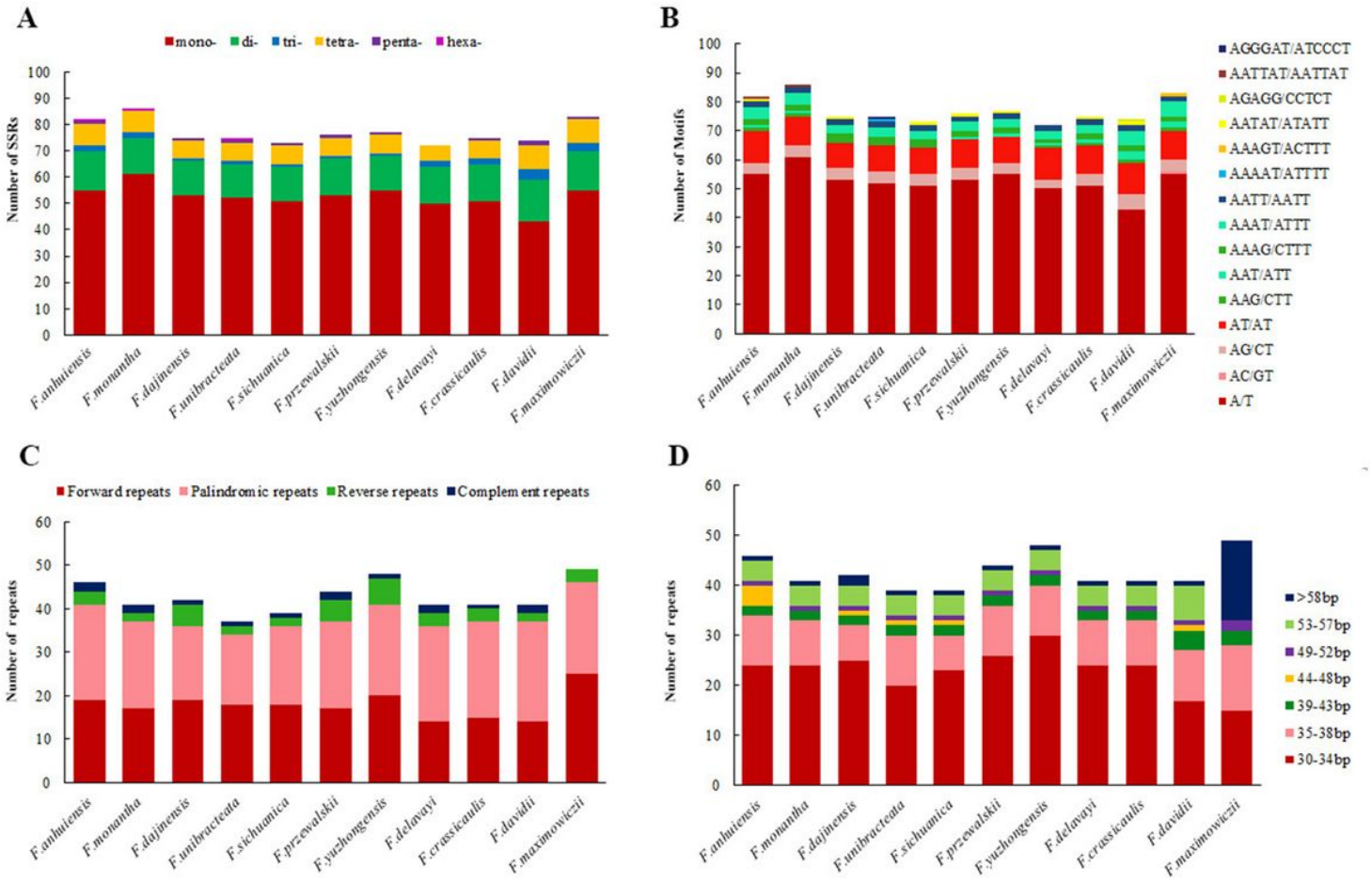

Figure 4

Analysis of different repeats in eleven chloroplast genomes of Fritillaria. (A) Number of different simple sequence repeat (SSR) types detected; (B) total numbers of different SSR motifs; (C) the number of four repeats types; (D) frequency of repeats by length.

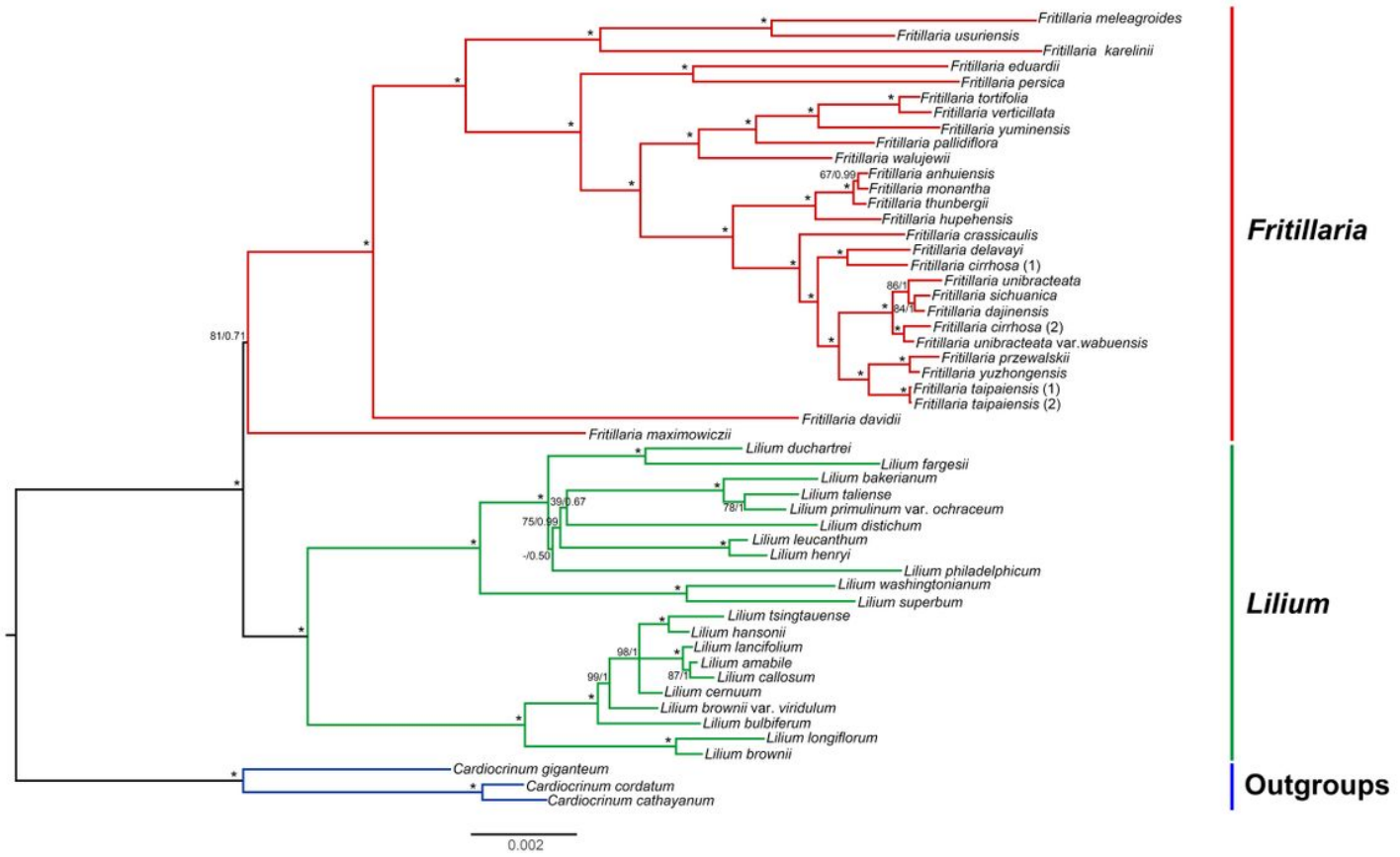

Figure 5

Phylogenetic relationship of the 52 species inferred from ML and BI analyses based on 64 shared protein-coding genes. Support values marker above the branches follow the order BS (bootstrap support)/pp (posterior probability), and "*" indicates 100% support values in both ML and BI trees. Accessions from different genera are written using different colors.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplement1.xlsx
- supplement2.xlsx
- supplement3.docx
- supplement4.doc
- supplement5.docx
- supplement6.xlsx
- supplement7.docx
- supplement8.xlsx
- supplement9.jpg
- supplement10.xlsx
- supplement11.xlsx

- supplement12.jpg
- supplement13.jpg
- supplement14.jpg