Assembly and comparative analysis of the complete mitochondrial genome of *Fritillaria ussuriensis* Maxim. (Liliales: Liliaceae), an endangered medicinal plant

Ping Xie[1†], Jingru Wu[2†], Mengyue Lu[1], Tongxin Tian[1], Dongmei Wang[1], Zhiwen Luo[1], Donghong Yang[2], Lili Li[1], Xuewen Yang[1], Decai Liu[1], Haitao Cheng[1], Jiaxin Tan[1], Hongsheng Yang[1*] and Dequan Zhu[3*]

Abstract

Background *Fritillaria ussuriensis* is an endangered medicinal plant known for its notable therapeutic properties. Unfortunately, its population has drastically declined due to the destruction of forest habitats. Thus, effectively protecting *F. ussuriensis* from extinction poses a significant challenge. A profound understanding of its genetic foundation is crucial. To date, research on the complete mitochondrial genome of *F. ussuriensis* has not yet been reported.

Results The complete mitochondrial genome of *F. ussuriensis* was sequenced and assembled by integrating PacBio and Illumina sequencing technologies, revealing 13 circular chromosomes totaling 737,569 bp with an average GC content of 45.41%. A total of 55 genes were annotated in this mitogenome, including 2 rRNA genes, 12 tRNA genes, and 41 PCGs. The mitochondrial genome of *F. ussuriensis* contained 192 SSRs and 4,027 dispersed repeats. In the PCGs of *F. ussuriensis* mitogenome, 90.00% of the RSCU values exceeding 1 exhibited a preference for A-ended or U-ended codons. In addition, 505 RNA editing sites were predicted across these PCGs. Selective pressure analysis suggested negative selection on most PCGs to preserve mitochondrial functionality, as the notable exception of the gene *nad3* showed positive selection. Comparison between the mitochondrial and chloroplast genomes of *F. ussuriensis* revealed 20 homologous fragments totaling 8,954 bp. Nucleotide diversity analysis revealed the variation among genes, and gene *atp9* was the most notable. Despite the conservation of GC content, mitogenome sizes varied significantly among six closely related species, and colinear analysis confirmed the lack of conservation in their genomic structures. Phylogenetic analysis indicated a close relationship between *F. ussuriensis* and *Lilium tsingtauense*.

Conclusions In this study, we sequenced and annotated the mitogenome of *F. ussuriensis* and compared it with the mitogenomes of other closely related species. In addition to genomic features and evolutionary position, this study also provides valuable genomic resources to further understand and utilize this medicinal plant.

Keywords *Fritillaria ussuriensis* Maxim., Mitochondrial genome, Repeat sequence, Comparative analysis, Phylogenetic analysis

†Ping Xie and Jingru Wu contributed equally to this work.

*Correspondence:
Hongsheng Yang
yhongsheng@126.com
Dequan Zhu
zdqoliver@163.com
Full list of author information is available at the end of the article

Xie *et al. BMC Genomics*   (2024) 25:773

Page 2 of 18

Introduction

Mitochondria not only convert biomass energy to chemical energy but also participate in metabolic processes associated with the degradation and synthesis of intracellular compounds [1]. According to the endosymbiotic theory, mitochondria originate from ancient endosymbiotic bacteria and possess semi-autonomous mitogenomes capable of encoding certain self-related proteins [2, 3]. During long-term evolution, mitogenomes have established stable regulatory mechanisms with the nuclear genomes [4]. Notably, angiosperm mitogenomes are predominantly maternally inherited, which eliminates the influence of the paternal line and facilitates studies on genetic mechanisms [5].

Compared to other eukaryotes, plant mitogenomes are more complex in terms of structure, number of repeat sequences, and size [6]. The sizes and structures vary largely among different plant species, which can be attributed to the genomic rearrangements involved in evolutionary processes [7]. Moreover, the sizes of plant mitogenomes are larger than those of other eukaryotes [8], and fluctuate greatly even within the same genus [9]. The mitogenomes of terrestrial plants vary significantly in size [10, 11], ranging from 66 kb to 11.3 Mb. A tremendous number of repetitive sequences, including simple sequence repeats (SSRs), tandem repeats, and dispersed repeats, have been discovered within plant mitogenomes and are believed to contribute to their structural variations [12]. Although circular molecules are commonly considered as the predominant form of plant mitochondrial genomes, other distinct structures are also observed, such as linear configurations, branched structures, and numerous smaller circular molecules [13]. For example, the complete mitogenome of *Coptis chinensis* can be mapped as six unequal circular molecules in length [14]. Notably, the plant mitochondrial genome not only varies in structure and size, but also is diverse and complex in gene content [15]. The number of genes typically ranges from 32 to 67 in most land plants [16]. Gene transfer events further complicate plant mitogenomes, as the transferred genes are primarily associated with the encoding of ribosomal proteins [17]. Generally, the transfer frequency of genes from chloroplast genomes to mitogenomes is higher in plants than in other eukaryotes [18]. In summary, these features, particularly the extensive genomic recombination, make the assembly and annotation of plant mitogenomes more challenging than those of other organelle genomes [19]. Currently, only about 500 complete plant mitogenomes have been uploaded to the National Center for Biotechnology Information (NCBI), which significantly lags behind the number of chloroplast genomes studied. The large repetitive regions in the mitochondrial genome cannot be accurately identified using next-generation sequencing, and effective resolution of this issue requires integration with third-generation sequencing [20]. Research on plant mitochondrial genomes has been continuously advancing due to significant breakthroughs in sequencing and assembly technologies.

Fritillaria ussuriensis Maxim., commonly known as "Ping Bei" in Chinese, is a perennial herbaceous plant belonging to *Fritillaria* of Liliaceae. It is mainly distributed in the lowland regions of northeast China, such as the Changbai Mountain range [21]. The dried bulbs of *F. ussuriensis* have long been recognized for their therapeutic properties in clearing heat, resolving toxins, and relieving cough and phlegm [22]. For thousands of years, it has been one of the most vital antitussive and expectorant drugs in China and other Asian countries [23]. In recent decades, various ingredients have been found in *F. ussuriensis* [24, 25], while the main bioactive components are steroidal alkaloids [26]. As a well-known medicinal material in the northeast provinces of China [27], *F. ussuriensis* was officially recorded in the 2005 edition of Chinese Pharmacopoeia [28]. There has been research on the chloroplast genome of *F. ussuriensis* [29], but none on its mitochondrial genome. In response to these circumstances, the data of *F. ussuriensis* mitogenome was obtained by integrating PacBio and Illumina sequencing technology. Based on these data, the mitogenome of *F. ussuriensis* was assembled and annotated. Then we conducted various analyses, including the assessment of codon usage, identification of the gene transfer from the chloroplast genome to the mitogenome, analysis of repeated sequences, identification of RNA editing sites, analysis of selective pressure, comparative genomics with closely related species and investigation of phylogenetic relationships. These results will help to better understand the structure and function of the *F. ussuriensis* mitogenome and provide valuable data for conservation biology, population genetics, and evolutionary studies of this species.

Results

Features of the *F. ussuriensis* mitogenome

Bandage was used to visualize the structure of contigs generated by the assembly result from GetOrganelle (Fig. 1A). The assembly had multiple nodes representing the contigs, with overlapping regions indicated by the connecting lines. Moreover, the presence of repetitive regions was identified through the deep coverage analysis. Subsequently, we mapped the PacBio long reads to the repetitive regions of the genome, which simplified the assembly into 13 circular contigs (Fig. 1B). These simplified contigs were also referred to as chromosomes in the previous study [30]. The Illumina data were mapped to
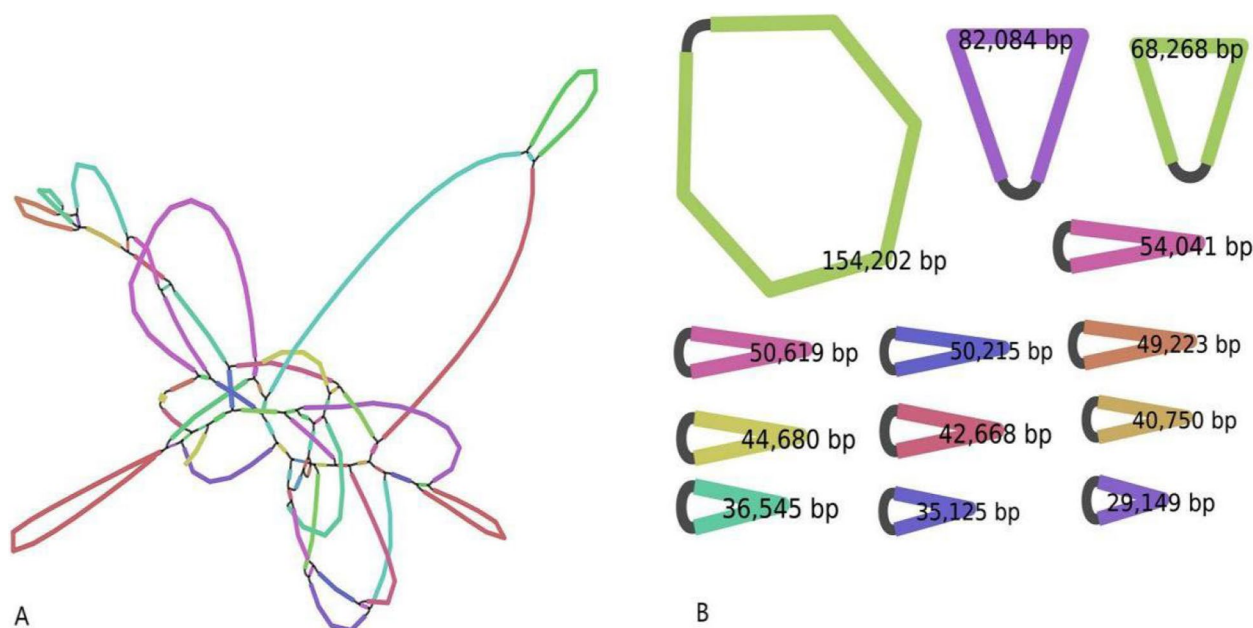
Fig. 1 Schematic diagram of hybrid assembly. **A** Graphical mitogenome assembly of *F. ussuriensis*. **B** 13 circular chromosomes obtained by solving repetitive regions based on PacBio long-reads

the assembled result, covering each base of the 13 chromosomes with an average depth of approximately 65-fold (Fig. S1). These results provide strong evidence for the accuracy of the mitochondrial genome assembly.

The assembled *F. ussuriensis* mitogenome had 13 chromosomes (Fig. 2), totaling 737,569 bp in length, with an average GC content of 45.41%. The detailed information of the length, GC content, and accession number for each chromosome was presented in Table 1. The size range of these chromosomes was from 29,149 bp to 154,202 bp. In the mitochondrial genome of *F. ussuriensis*, a total of 55 genes were annotated, comprising 41 protein-coding genes (PCGs), 12 transfer RNA (tRNA) genes, and 2 ribosomal RNA (rRNA) genes (Table S1). These PCGs included 9 NADH dehydrogenase genes, 5 ATP synthase genes (two copies of *atp9*), 4 cytochrome c biogenesis genes, 3 cytochrome c oxidase genes, 1 ubiquinol cytochrome c reductase gene, 1 transport membrane protein gene, 1 maturase gene, 11 small subunits of ribosome genes, 4 large ribosomal subunit genes, and 1 succinate dehydrogenase gene. Among these PCGs, genes *ccmFC*, *nad1*, *rps10*, *rps3*, and *rpl2* each had 1 intron, while *cox2*, *nad4*, *nad5*, and *nad7* each had 2 introns. In addition, we annotated 12 tRNA genes, including *trnC-GCA*, *trnD-GUC*, *trnQ-UUG*, *trnE-UUC*, *trnM-CAU*, *trnH-GUG*, *trnW-CCA*, *trnK-UUU*, *trnL-GAG*, *trnN-GUU*, *trnY-GUA*, and *trnfM-CAU*. Meanwhile, 2 rRNA genes (*rrn5* and *rrnS*) were also identified in this mitogenome.

Codon usage analysis

The codon usage in the mitochondrial PCGs of *F. ussuriensis* was analyzed using CodonW. The GC content varied among different codon positions, with an average of 43.42%. Specifically, the GC contents at positions GC1, GC2, and GC3 were 48.23%, 43.31%, and 38.72%, respectively. This uneven distribution of bases suggests the bias of codon composition. The effective number of codon (ENC) for these mitochondrial genes ranged from 39.65 to 60.00, with an average of 53.72 (Table S2). Among them, only 4 genes had an ENC value below 45, indicating that these codons in the mitochondrial genome had a relatively weak preference in usage. As depicted in Table S3 and Fig. 3, Alanine preferred GCU codon, displaying a relative synonymous codon usage (RSCU) value of 1.60, which was highest in these PCGs. A total of 30 codons were observed with RSCU values exceeding 1.00, indicating the relatively high frequency of usage of these codons. Among these codons, 27 codons had A or U at the third position, with 11 codons ending with A and 16 codons ending with U, accounting for 90.00% of the total. This finding suggested a preference for A-ended or U-ended codons in the *F. ussuriensis* mitochondrial genome.

Repeat sequence analysis

In the *F. ussuriensis* mitogenome, we identified a total of 192 SSRs (Fig. 4), and 103 of them consisted of tetrameric repeats, accounting for 53.65% of the count. Monomeric
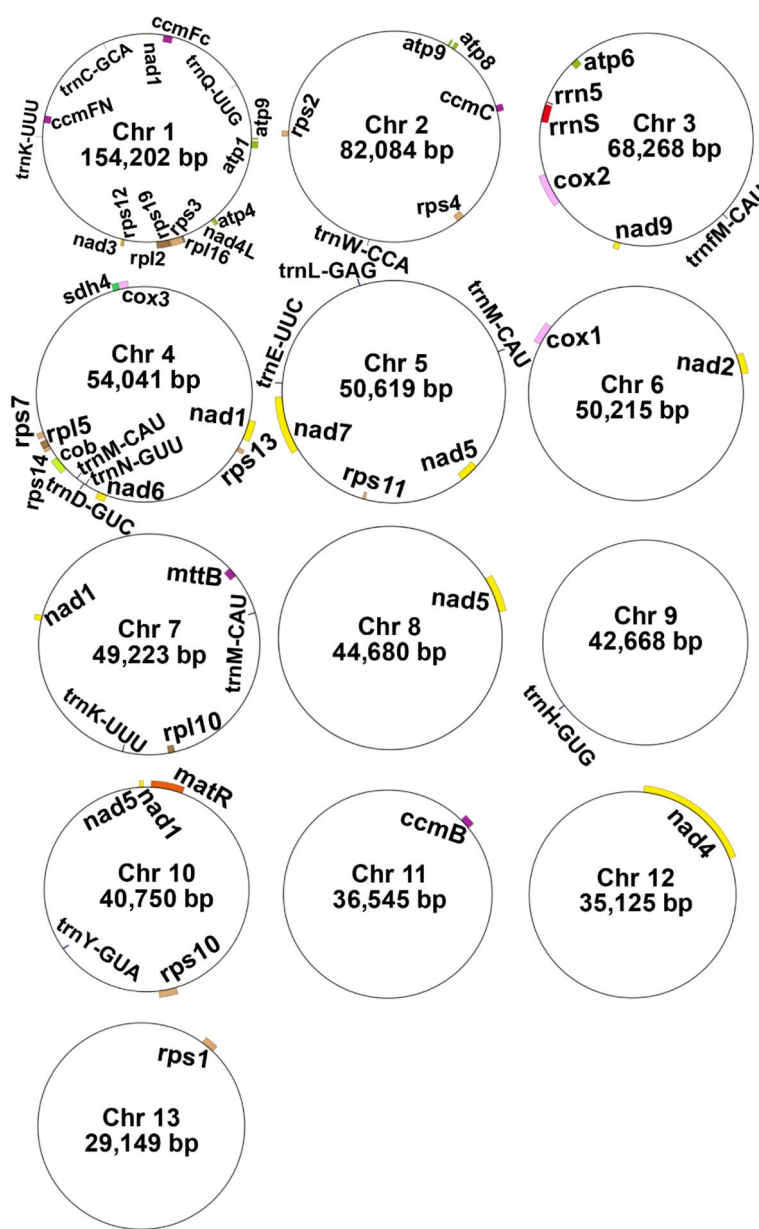
Fig. 2 The mitogenome maps of *F. ussuriensis*. The mitogenome consists of 13 circular chromosomes with different lengths and gene contents

and dimeric repeats had similar numbers, with 37 and 36, respectively, while there were only 13 trimeric repeats. The content of pentameric repeats was relatively low, with only 3 repeats (1.56%), and no hexameric repeats existed. In the 37 monomeric repeat sequences, 36 of them were repeat type A or T, accounting for 97.30% of the total.

As shown in Fig. 5A, a total of 4,270 dispersed repeat sequences were identified in the *F. ussuriensis* mitogenome, including 2,160 forward repeats, 32 reverse repeats, 2,061 palindromic repeats, and 17 complement

repeats. These repeat sequences varied in size ranging from 30 to 418 bp. Of them, 22 and 8 repeat sequences exceeded 100 and 200 bp in length, respectively. Totally 3,815 (70.66%) dispersed repeat sequences fell within the length of 30–34 bp (Fig. 5B). These dispersed repeat sequences had 139,575 bp in length, which accounted for 18.92% of the complete mitogenome sequence.

Prediction of RNA editing sites

In the mitogenome of *F. ussuriensis*, 505 RNA editing sites were identified, all of which were of the C to U type.

Xie *et al. BMC Genomics*   (2024) 25:773

Page 5 of 18

Table 1 The information of the *F. ussuriensis* mitogenome

Mitogenome	Length(bp)	GC content(%)	Accession number
Chromosome 1	154,202	45.66%	OR783162
Chromosome 2	82,084	44.58%	OR783163
Chromosome 3	68,268	46.53%	OR783164
Chromosome 4	54,041	45.25%	OR783165
Chromosome 5	50,619	45.52%	OR783166
Chromosome 6	50,215	44.53%	OR783167
Chromosome 7	49,223	45.07%	OR783168
Chromosome 8	44,680	46.81%	OR783169
Chromosome 9	42,668	44.90%	OR783170
Chromosome 10	40,750	45.74%	OR783171
Chromosome 11	36,545	44.67%	OR783172
Chromosome 12	35,125	45.87%	OR783173
Chromosome 13	29,149	44.60%	OR783174

The gene types were closely correlated with the frequency of RNA editing, with NADH dehydrogenase genes being relatively more prone to editing. Specifically, the gene *nad4* exhibited the highest number of editing sites (46 sites), while the ribosomal protein encoding gene *rps11* had the fewest (1 site), as shown in Fig. 6. Of these editing sites, 157 (31.09%) existed at the first base of the codon, while 332 (65.74%) appeared at the second base position. In addition, RNA editing not only changed the encoded amino acid but also led to the generation of stop codons, which prematurely terminated the coding process. This phenomenon was also observed in the coding genes *rps10*, *ccmFc*, and *atp9* of the *F. ussuriensis* mitogenome. Similarly, RNA editing also resulted in the generation of start codons for the genes *nad4L*, *nad1*, *cox2,* and *atp4* by converting ACG to AUG. The predicted results also showed that codon editing led to the highest tendency for conversion to leucine, accounting for 41.98% (212 sites) of the total conversion, and following that, 24.36% of the amino acids altered to phenylalanine (123 sites). Even with RNA editing, 44.56% of the amino acids retained their original hydrophilicity or hydrophobicity, while 45.54% changed from hydrophilicity to hydrophobicity, and 9.31% changed from hydrophobicity to hydrophilicity (Table S4).

DNA migration from chloroplast to mitochondria
We identified 20 homologous fragments between the chloroplast and mitochondrial genomes of *F. ussuriensis* (Fig. 7). These fragments had a total length of 8,954 bp, with individual size ranging from 65 bp to 1,917 bp, collectively accounting for 1.21% of the mitogenome (Table S5). In the evolutionary process, 5 genes (*ndhA*, *ndhC*, *psbD*, *rrn16*, and *rrn23*) lost their integrity through migration from the chloroplast genome to the mitogenome. Furthermore, 8 intact genes were also found in these homologous fragments (*rrn5*, *rps7*, *rps12*, *trnC-GCA*, *trnQ-UUG*, *trnH-GUG*, *trnfM-CAU*, and *trnW-CCA*), most of which were tRNA genes.

Phylogenetic analysis
To better understand the evolutionary status of *F. ussuriensis* in the monocots system, we performed a phylogenetic analysis along with the mitochondrial genomes of 24 published plant species. Among the selected plants, there were 12 species of Poales, 5 species of Alismatales, 4 species of Asparagales, and 1 species of Liliales, with *Liriodendron tulipifera* and *Magnolia officinalis* serving as the outgroup (Table S6). These plant mitogenomes were obtained from NCBI. Then a phylogenetic tree was constructed using the maximum likelihood analysis with the GTR + G model and 1,000 bootstrap replicates. This phylogenetic tree displayed the high bootstrap support values of more than 80 for most nodes (Fig. 8). Our analysis revealed a close phylogenetic relationship between *F. ussuriensis* and *Lilium tsingtauense*. The overall topology of the phylogenetic tree was congruent with the classification provided by the Angiosperm Phylogeny Group IV (APG IV) system.

Comparison of the *F. ussuriensis* mitogenome with other five closely related species
In order to delve deeper into the evolutionary characteristics of the mitochondrial genome of *F. ussuriensis*, we compared it with *L. tsingtauense*, and 4 species of Asparagales (*Allium cepa, Chlorophytum comosum, Asparagus officinalis*, and *Crocus sativus*). The GC content in the coding regions of these mitochondrial genomes ranged from 45.32% to 46.81%. Additionally, the GC1, GC2, and GC3 content of these mitochondrial genomes fell within the respective ranges of 48.08% to 49.76%, 43.31% to 43.98%, and 38.26% to 38.72%. The detailed values for each species were presented in Table 2. It was worth noting that while the GC content remained relatively consistent across these species, there was a significant variation in the number of genes. Among the species analyzed, a total of 25 PCGs were identified to be shared, including *atp1, atp4, atp6, atp8, atp9, nad1, nad2, nad3, nad4, nad4L, nad5, nad6, nad7, nad9, ccmB, ccmC, ccmFC, ccmFN, cob, cox1, cox2, cox3, matR, mttB,* and *rps12*. *C. comosum* had the highest number of mitochondrial genes with 62, while *A. cepa* had the lowest number of mitochondrial genes with 43. The selective pressure of the PCGs is determined by the the non-synonymous (Ka) and synonymous (Ks) substitution ratios (Ka/Ks): Ka/Ks > 1, positive or Darwinian selection; Ka/Ks = 1, neutral selection; and Ka/Ks < 1, negative or purifying selection [31]. These PCGs from the mitogenomes
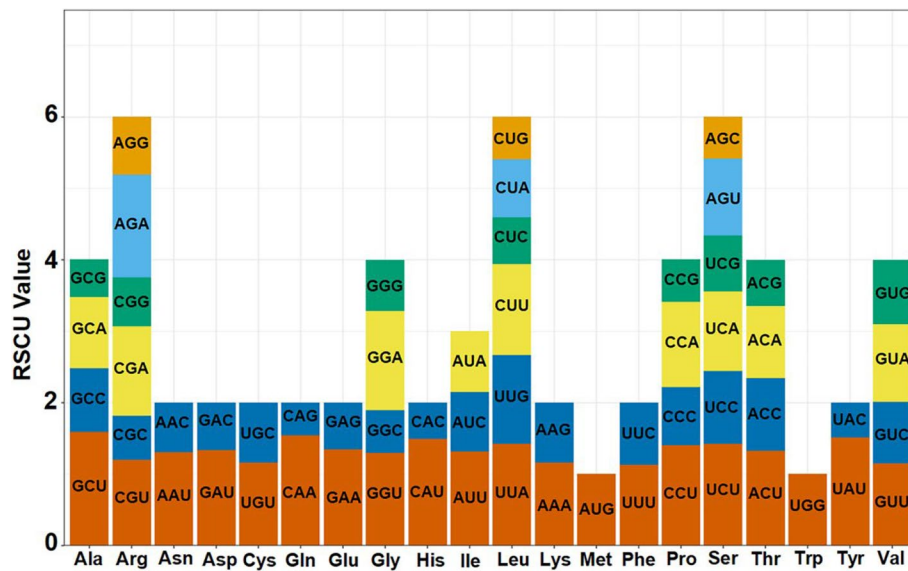
Xie *et al. BMC Genomics*   (2024) 25:773

Page 6 of 18

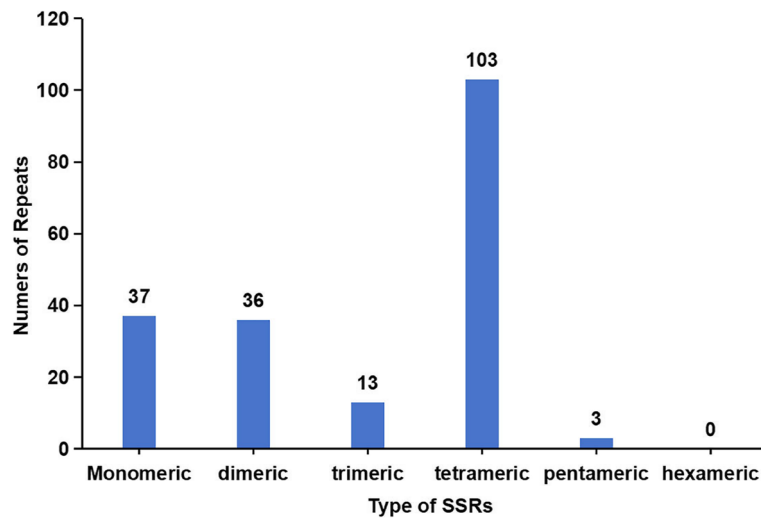**Fig. 3** Codon preference of *F. ussuriensis* mitogenome

Fig. 4 The SSRs identified in the *F. ussuriensis* mitogenome

of the six closely-related species were compared to determine the Ka/Ks ratio (Table 3). Genes *nad3* and *ccmB* exhibited a Ka/Ks value greater than 1, indicating that they underwent positive selection. The remaining PCGs all possessed the values of Ka/Ks under 1, accounting for 92.00% of the total, which suggested that these genes had undergone purifying selection to maintain the normal mitochondrial function.

Gene loss and multi-copy gene
The loss of PCGs is an important factor that leads to significant variation in the mitochondrial gene content of

angiosperms [17]. Reportedly, 41 PCGs were inferred to exist in the ancestral mitogenome of angiosperms [32]. Gene sdh3 was lost in the mitogenomes of all selected species (Fig. 9). Furthermore, 3 genes (*rpl2*, *rpl10*, and *sdh3*) were lost in the mitochondrial genomes of *A. cepa*, *C. comosum*, *A. officinalis*, *C. sativus*, and *L. tsingtauense*. The presence of multi-copy genes within plant mitochondrial genomes was widely observed [33]. Specifically, the mitogenome of *F. ussuriensis* contained two copies of *atp9*. Similarly, *A. cepa* and *C. sativus* mitochondrial genomes possessed two copies of *ccmFN* and *cox3*, respectively. In contrast, 3 genes (*nad6*, *nad9*, and *rps2*)

Xie *et al. BMC Genomics*   (2024) 25:773

Page 7 of 18
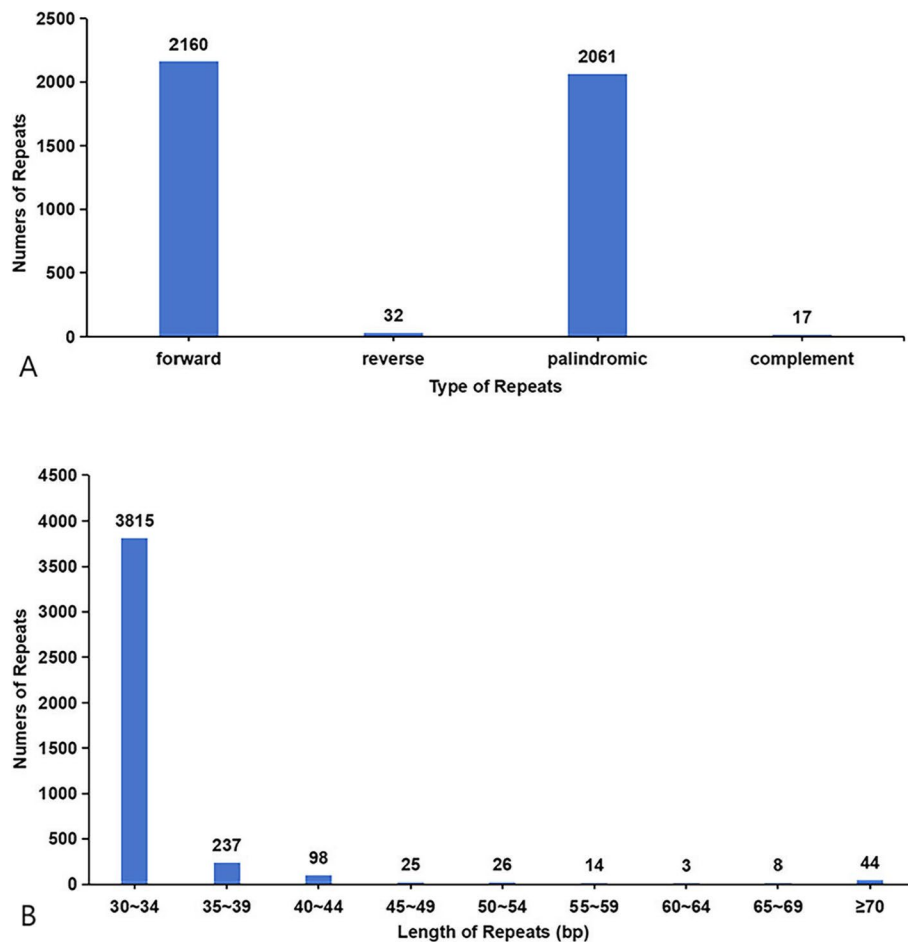
Fig. 5 The dispersed repeats identified in *F. ussuriensis* mitogenome. **A** Type of the dispersed repeats. **B** Distribution of the lengths of dispersed repeats

were found to possess two copies in the mitogenome of *L. tsingtauense*. Overall, *F. ussuriensis* had the relatively complete PCGs, while *A. cepa* lost all mitochondrial genes except gene *rps12* and 24 core PCGs.

Nucleotide diversity
Nucleotide diversity (Pi) serves as a measure for assessing the variation in nucleic acid sequences across different species, and regions with significant variability can be identified as potential molecular markers for population genetics research [34]. Figure 10 illustrated the nucleotide diversity of the 25 PCGs and 2 rRNA genes in the mitochondrial genomes of *F. ussuriensis*, *L. tsingtauense*, and 4 species of Asparagales. The Pi values for these genes ranged from 0.0175 to 0.1039, with most being below 0.1. Notably, the gene *atp9* exhibited the highest variability, with a Pi value of 0.1039. Additionally, the genes *cox3* (Pi = 0.0745), *cox2* (Pi = 0.0721), and *rps12* (Pi = 0.0715) demonstrated considerable

variability. In contrast, genes *nad1* (Pi = 0.0161) and *nad7* (Pi = 0.0187) were the most conserved PCGs. Additionally, the two rRNA genes (*rrn5* and *rrnS*) were also conserved, with Pi values of 0.0162 and 0.0114, respectively. Overall, the nucleotide diversity of the PCGs varied significantly among these species.

Colinear analysis
To investigate the relationships among the mitochondrial genomes of *F. ussuriensis*, *L. tsingtauense*, and four other Asparagales species, we utilized the BLASTN program to analyze homologous genes and their sequence arrangements. The connection ribbon between the two mitogenomes represented a highly homologous collinear block (Fig. 11). However, these mitogenomes exhibited limited collinearity, with many regions lacking homology among the species. Additionally, the arrangement order of collinear blocks in these mitochondrial genomes showed inconsistencies. These findings suggested substantial genomic rearrangements among the mitochondria
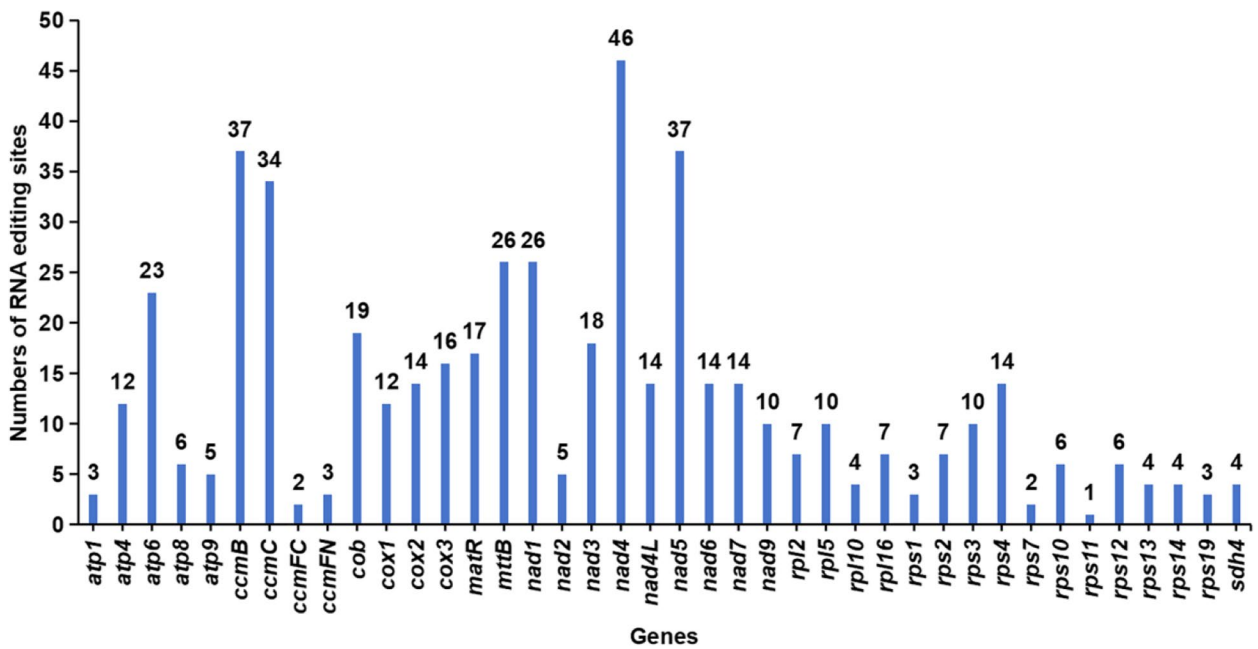
Fig. 6 The distribution of RNA editing sites among PCGs in the *F. ussuriensis* mitogenome
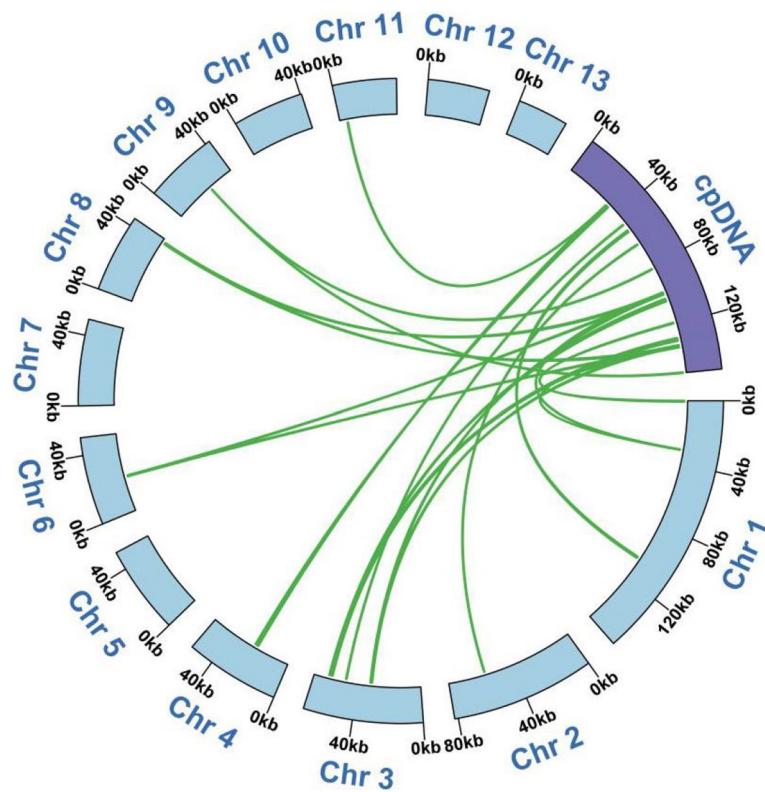
Fig. 7 The 13 blue arcs represent the mitochondrial genome of *F. ussuriensis*, and the purple arc represents its chloroplast genome. The genome fragments corresponding to the green connecting lines between arcs are homologous fragments

Xie *et al. BMC Genomics*   (2024) 25:773

Page 9 of 18
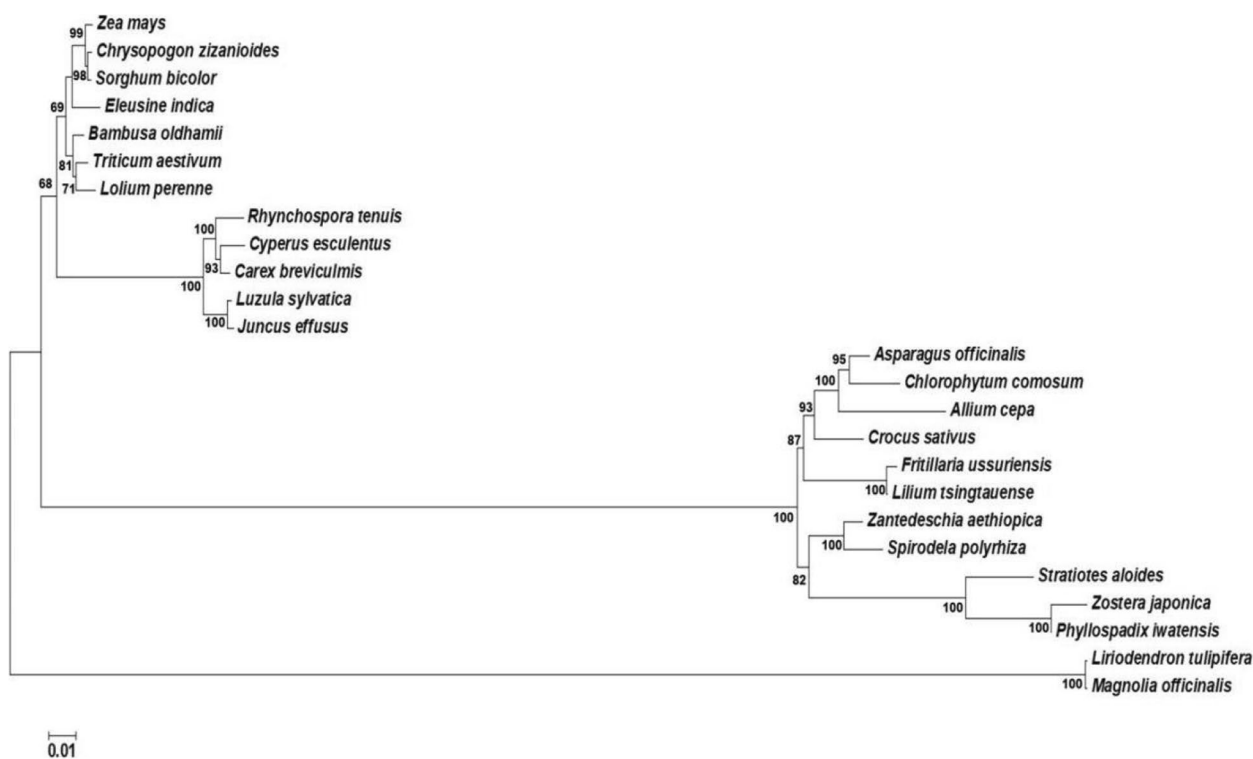
Fig. 8 The phylogenetic relationships of *F. ussuriensis* with other 24 plant species. The tree was constructed based on the DNA sequences of 24 conserved mitochondrial PCGs (*atp1*, *atp4*, *atp6*, *atp8*, a*tp9*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *nad7*, *nad9*, *ccmB*, *ccmC*, *ccmFC*, *ccmFN*, *cob*, *cox1*, *cox2*, *cox3*, *matR*, and *mttB*)

of *F. ussuriensis* and the other 5 species, indicating a lack of conservation in the genomic structure of these mitogenomes.

Discussion

Mitochondria are considered as the powerhouse that generates the energy required for the life processes of plants [35]. Research over the years has demonstrated the plant mitochondrial genome as a dynamic entity in evolution, displaying vast diversity among the different species [36]. Since the initial endosymbiotic event, the plant mitochondrial genomes have undergone significant changes [37], resulting in a series of important features that make them more complicated than those of animals [13]. Investigations have revealed that the reported sizes of most plant mitogenomes range between 200 and 800 kb [32]. The size and GC content are the primary factors for assessing species [38]. Despite significant variations in genome sizes, the GC contents are relatively consistent across species, supporting the notion that GC content has remained conserved throughout the evolution of higher plants [5, 39]. The number of genes within mitochondrial genomes varies among plant species, but the genes essential for the main respiratory functions and energy synthesis are highly conserved [40, 41]. This phenomenon may contribute to the maintenance of mitochondrial function and normal life activities.

Although plant mitochondrial genomes have been commonly reported as circular structures similar to those of animals [42, 43], many studies have revealed that their actual structures can also have multiple branched, linear, or mixed forms of genomic structures [44, 45]. Multi-circular structures were also identified in the mitochondrial genomes of various species, encompassing ferns, basal angiosperms, monocots, and eudicots. For instance, the mitochondrial genome of *Gelsemium elegans* comprised two circular chromosomes [46], and a similar phenomenon was also observed in the *Psilotum nudum* mitogenome [47]. In contrast, the mitochondrial genome of *Amborella trichopoda* was assembled into five circular chromosomes, with sizes ranging from 118.7 kb to 3,179.3 kb [48]. In the majority of eudicots, the chromosome count of mitochondrial genomes was generally less than five [49], while the mitogenome of *Gastrodia elata*, a monocot plant, comprised 19 chromosomes [50]. The assembled mitochondrial genome of *F. ussuriensis* consisted of 13 circular chromosomes, each of which contained genes. In plant mitochondrial genomes with multiple chromosomes, some chromosomes may lack functional genes. For instance, among

Table 2 General characteristics of the mitochondrial genomes of *F. ussuriensis*, *L. tsingtauense*, and 4 species of Asparagales

| Gene	*A. cepa*	*C. comosum*	*A. officinalis*	*C. sativus*	*L. tsingtauense*	*F. ussuriensis*
Accession number	NC_030100.1	MW411187.1	NC_053642.1	OL804177.1	OP973783.1-OP973810.1	OR783162-OR783174
Size (bp)	316,363	950,316	492,062	525,348	1,125,108	737,569
GC (%)	45.32	45.65	45.90	46.81	46.17	45.41
GC1(%)	49.13	49.76	48.08	49.52	48.74	48.23
GC2(%)	43.96	43.98	43.57	43.66	43.37	43.31
GC3(%)	38.57	38.26	38.45	38.56	38.63	38.72
Genes	43	62	59	50	56	55
tRNAs	15	21	17	14	15	12
rRNAs	2	5	6	4	3	2
PCGs	26	36	36	32	38	41

the 58 chromosomes of the *Silene noctiflora* mitogenome, 20 were discovered to contain no functional genes [51]. These results indicate that plant mitochondrial genomes are diverse and complex in terms of structure, size, and gene content.

The variation in mitogenome sizes can be primarily attributed to the insertion or loss of genetic sequences and the increase in repetitive elements [49]. The repetitive sequences in the mitogenomes are closely related to intermolecular recombination [52], which can cause

Table 3 Ka/Ks values of 25 PCGs of *F. ussuriensis* versus those of *L. tsingtauense* and 4 species of Asparagales

Gene	Ka/Ks (*F. Ussuriensis* vs *A. cepa*)	Ka/Ks (*F. Ussuriensis* vs *C. comosum*)	Ka/Ks (*F. Ussuriensis* vs *A. officinalis*)	Ka/Ks (*F. Ussuriensis* vs *C. sativus*)	Ka/Ks (*F. Ussuriensis* vs *L. tsingtauense*)
atp1	0.1581	0.0988	0.1876	0.1266	0.1524
atp4	0.5114	0.4340	0.3991	0.6750	0.3677
atp6	0.3302	0.1372	0.2079	0.2255	0.0462
atp8	0.3123	0.4527	0.2752	0.6362	0.2125
atp9	0.2401	0.0878	0.0749	0.2265	0.0332
ccmB	1.5492	1.4993	1.5252	1.6355	1.0418
ccmC	0.7319	0.4725	0.5331	*	0.3304
ccmFC	0.4453	0.4223	0.5052	0.6102	0.4117
ccmFN	0.6479	0.5726	0.5377	0.5145	0.2306
nad1	0.1869	*	0.1869	0.1869	0.1759
nad2	0.8824	0.8941	0.5442	0.6807	0.4954
nad3	2.4811	1.0939	2.0363	1.1185	1.2927
nad4	0.1721	0.2119	0.5192	0.6185	0.3455
nad4L	0.0447	0.0510	0.0409	0.0689	0.0425
nad5	*	0.4280	0.2797	0.2526	0.3354
nad6	0.1942	0.1859	0.4674	0.1537	0.2326
nad7	0.6221	0.5017	*	0.4911	0.0728
nad9	*	0.3041	0.3057	0.3935	0.3018
rps12	*	0.5085	0.4228	0.4639	0.4421
cob	0.4412	0.5147	0.2106	0.4145	0.2195
cox1	0.7125	0.9022	0.7746	0.6333	0.5147
cox2	0.3334	*	0.2156	0.5887	0.1914
cox3	0.5731	*	0.4412	0.5183	0.2768
matR	0.7254	0.6232	*	0.4194	0.3214
matB	0.6533	0.5493	0.3149	0.5155	0.1352

When Ks = 0, the value cannot be calculated, it was represented by*
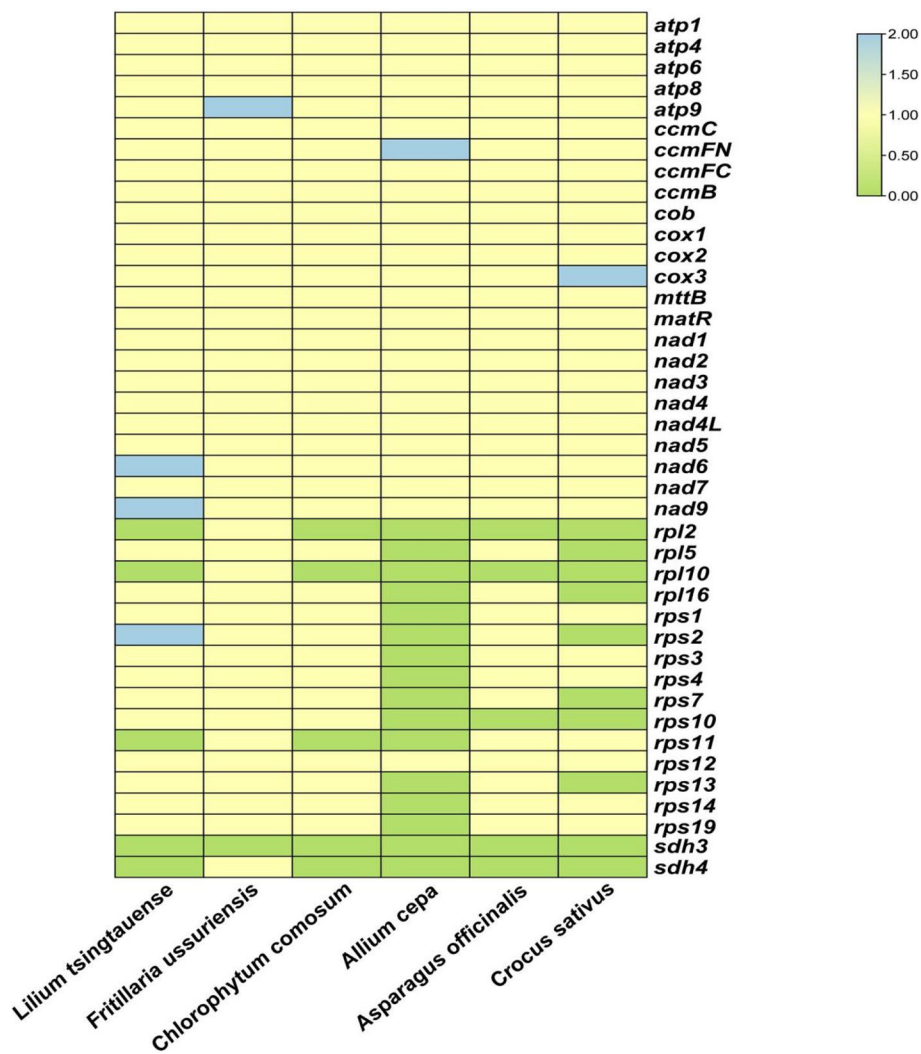
Fig. 9 The PCGs distribution of *F. ussuriensis*, *L. tsingtauense*, and 4 species of Asparagales. Green boxes indicate that the gene is not present in the mitogenomes. Yellow and blue boxes indicate that one and two copies exist in the particular mitochondrial genomes, respectively

drastic changes in the size and structure of the genomes [53]. Previous studies indicated that SSRs containing A and T bases were more likely to occur in the genomes of chloroplasts and mitochondria because the two hydrogen bonds that connected them were easier to break [54, 55]. Specifically, repetitive elements ranging from 0.5 kb to 120 kb were identified as key contributors to the notable size variance observed in several mitochondrial genomes of *Zea mays* [56]. In the *F. ussuriensis* mitogenome, the repeat sequences were less than 500 bp. A similar observation was made in the mitochondrial genomes of four *Populus* species, where the lengths of repeat sequences did not exceed 350 bp [57]. This phenomenon may contribute to the stability of size and structure of the plant mitochondrial genomes.

The frequent transfer of foreign DNA among chloroplasts, mitochondria, and nuclei also affects the size of the plant mitochondrial genomes [58]. For instance, numerous DNA sequences in the mitochondrial genome of *Malus domestica* are highly similar to the nuclear DNA sequences [59]. These sequences are also considered to be the driving force of the mitogenome expansion. Similarly, in the mitogenome of *Vitis vinifera*, a total of 30 chloroplast fragments with a combined length of 68,237 bp were discovered, accounting for 8.80% of its mitochondrial genome [60]. In this study, the majority of the sequences transferred from the chloroplast genome to the mitogenome were tRNA genes, which were consistent with similar transfer processes observed in other angiosperms [4]. The transfer of sequences between these
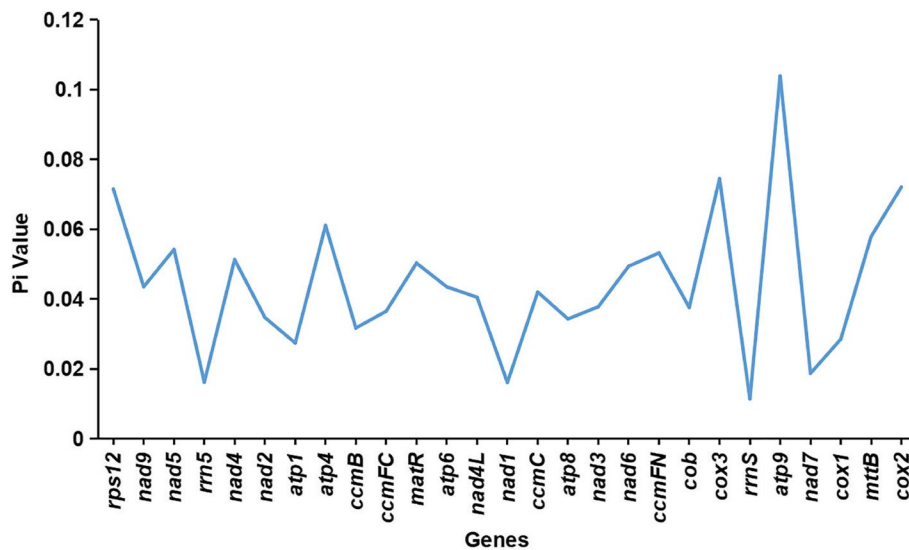
Fig. 10 Nucleotide diversity (Pi) among the mitochondrial genomes of *F. ussuriensis*, *L. tsingtauense*, and 4 species of Asparagales

two genomes was generally considered unidirectional [61], and related studies showed that sequences transferred from the chloroplast could compensate for tRNA genes lost during mitochondrial evolution [62]. The expression of tRNA genes from the chloroplast genome in the mitochondrial genome was demonstrated [63], and these genes participated in the transport of relevant amino acids to maintain normal life activities [64]. Analysis of sequence transfer provides valuable insights into

DNA transfer events between different genomes within cells [65], which contribute to a deeper understanding of the evolutionary process of eukaryotes.

RSCU reflects the ratio of the actual codon usage frequency to the theoretical frequency [66]. RSCU exceeding 1 indicates that the frequency of actual codon usage is higher than that of other synonymous codons [67]. Overall, there is a strong A/T bias at the third codon position in the mitochondrial genome of *F. ussuriensis*.
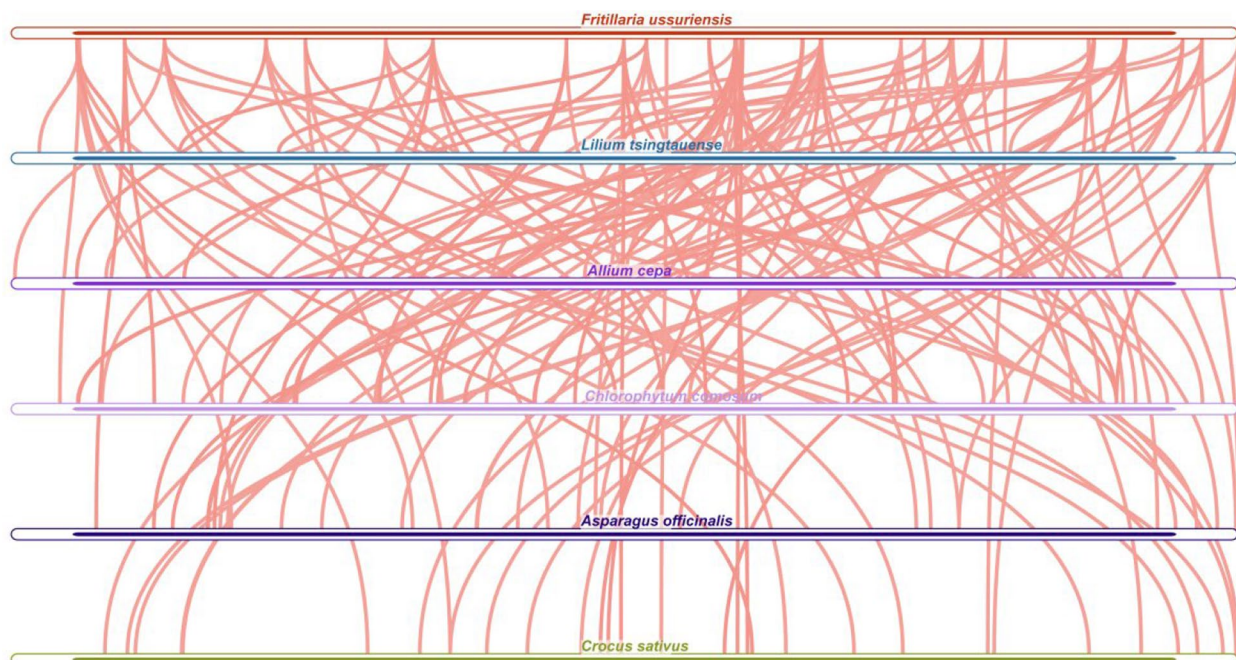
Fig. 11 Collinear analysis of the mitogenomes of *F. ussuriensis*, *L. tsingtauense*, and 4 species of Asparagales. The pink arcs indicate homologous regions of these mitochondrial genomes

Xie *et al. BMC Genomics*   (2024) 25:773

Page 13 of 18

This phenomenon was also found in other plant mitochondrial genomes [68, 69], which was considered to be the result of the long-term evolutionary process of plants adapting to the environment [70].

RNA editing is a post-transcriptional process occurring commonly in the plant mitochondrial genome, and can lead to a huge diversity of gene sequences [71]. This process can modify the relevant genetic information at the mRNA level, and thus contributes to improving protein folding [72]. Reportedly, 517 RNA editing sites among 34 genes and 441 RNA editing sites across 36 genes were identified in the mitochondrial genomes of *Bupleurum chinense* [73] and *Arabidopsis thaliana* [74], respectively. Moreover, nearly 50.00% of RNA editing sites were found at the second codon [75]. The hydrophobic characteristics of more than half of the amino acids were altered [76]. Furthermore, the hydrophobicity of amino acids was closely related to protein folding and secondary structure formation [77]. The editing of amino acids into stop codons can prematurely terminate protein synthesis, as also reported in other plants [78, 79]. Despite the limitations of the existing prediction methods and other factors leading to low average coverage [80], the identification of RNA editing sites still provided valuable information to predict gene functions.

In this study, we compared the mitogenome of *F. ussuriensis* with those of *L. tsingtauense*, and 4 species of Asparagales to learn more about its structure and organization. The comparisons revealed that these mitogenomes had undergone substantial rearrangements, exhibiting an exceptional lack of conservation in structure across these species, which might serve as a fundamental driving force behind the evolution and diversification of plant mitogenomes [81]. The analysis of nucleotide diversity revealed that the rRNA genes were relatively conserved, a finding consistent with similar reports in the previous study [82]. The selection pressure on PCGs was usually estimated by the ratio of Ka/Ks, which helped deepen the understanding of the evolution of plant mitogenomes [31]. The phenomenon of relatively conserved PCGs in the mitochondrial genomes is also common in related studies of other plants [73, 83]. In natural selection, the Ka/Ks value of most genes in plant mitochondrial genomes was smaller than 1 in order to remove the deleterious mutations and maintain mitochondrial functions [84]. The maximum likelihood analysis established a preliminary evolutionary position of *F. ussuriensis* at the mitochondrial genome level and indicated a close affinity between *F. ussuriensis* and *L. tsingtauense*. Our findings underscored the utility of organelle genomic data in clarifying plant phylogenetic relationships, advancing the development of molecular markers, and fostering studies on genetic evolution.

Conclusions

By a combined strategy of Illumina and PacBio sequencing technologies, high-quality sequencing data of the *F. ussuriensis* mitogenome were obtained. Subsequently, we assembled and annotated this genome with a total length of 737,569 bp, which consisted of 13 circular chromosomes. Within this genome, 55 genes were annotated, including 41 PCGs, 12 tRNA genes, and 2 rRNA genes. Codon usage analysis suggested a preference for A-ended or U-ended codons in this mitogenome. We identified 192 SSRs and 4,270 dispersed repeat sequences. The majority of the transferred sequences from the chloroplast genome to the mitogenome of *F. ussuriensis* were tRNA genes. 505 RNA editing sites existed in the PCGs of this mitogenome. Based on the phylogenetic analysis of the mitogenomes of *F. ussuriensis* and 24 other plants, the evolutionary status of this plant was clarified, which was basically consistent with the traditional classification results. Additionally, Ka/Ks analysis, nucleotide diversity analysis, and comparative analysis of genomic features were performed to provide a more comprehensive understanding of mitogenome evolution in these closely related species. In summary, our analyses provide important information for biological research on *F. ussuriensis* and support further determination of evolutionary relationships within Liliaceae.

Materials and methods

Plant materials, DNA extraction and sequencing

The etiolated seedlings of *F. ussuriensis* that were covered with soil were collected from the medicinal plant plantation in Tieli City, Heilongjiang Province, China (46°58′28″ N, 127°57′7″ E). The sample was deposited at the herbarium of Jiamusi University (Jiamusi, China), and its accession number was PBM202301. Density gradient centrifugation was utilized to isolate the mitochondria from the leaves of these etiolated seedlings [85], further using DNase I for eliminating the DNA pollutants of other genomes [86]. After the above procedures, DNA was extracted from the purified mitochondria. We used the Express Template Prep Kit 2.0 of PacBio to convert approximately 2 μg of mitochondrial DNA into SMRTbell libraries [15], and then used PacBio Sequel II platform for sequencing [87]. Covaris M220 system was utilized to sonicate roughly 1 μg of mitochondrial DNA into fragments of approximately 500 bp in length [88], and then we used the TIANgel Midi Purification Kit to purify the processed fragments [89]. NEBNext® Ultra™ DNA Library Prep Kit was used to construct the libraries, and then the Illumina NovaSeq 6000 platform was utilized for their sequencing [30].

Xie *et al. BMC Genomics*   (2024) 25:773

Page 14 of 18

Assembly and annotation of the mitogenome

The mitogenome of *F. ussuriensis* was assembled using a hybrid strategy that combined Illumina and PacBio reads. Initially, FastQC (http://www.bioinformatics.babra ham.ac.uk/projects/fastqc/) was used to check the raw short reads [79], and then they were trimmed to produce clean reads by Trimmomatic (v0.36) [90]. GetOrganelle (v1.6.4) was used to build contigs from the short clean reads with default parameters [91]. The long reads were used to de novo assemble the *F. ussuriensis* mitogenome by the Canu (v2.1.1) [92], and then we aligned them with the plant mitogenome database to extract potential mitochondrial contigs. We mapped the short reads to the extracted contigs, retaining all mapped reads by BWA (v0.7.17) and SAMTools (v0.1.19) [93, 94]. The hybrid assembly of PacBio long reads and Illumina short reads was performed through SPAdes (v3.9.0) with multiple k-mer parameters [95]. Subsequently, we mapped the long PacBio reads onto the repeated sequence areas of the assembled mitochondrial scaffold to resolve these regions. Contigs generated by SPAdes were imported into Bandage to visualize and analyze the assembly [96]. All 24 highly conserved core genes were included in our assembly results [30], and all mitochondrial fragments were fully extended. Moreover, the assembled mitogenomes were aligned with the raw reads, and polished through Pilon (v1.18) for correcting the errors [97]. Finally, we obtained 13 simplified circular contigs by this assembly strategy, representing the complete mitogenome of *F. ussuriensis.*

Based on the mitochondrial genomes of *A. cepa* (NC_030100.1), *C. comosum* (MW411187.1), *A. officinalis* (NC_053642.1), and *L. tsingtauense* (OP973783.1-OP973810.1) from GenBank, the mitogenome of *F. ussuriensis* was annotated by GeSeq [98]. Moreover, tRNAscan-SE and BLASTN were utilized to annotate tRNA and rRNA [99, 100], respectively. Then, we checked and adjusted the errors from the annotated results through Apollo (v1.11.8) [101]. Finally, the genome map was generated by OGDRAW program (https://chlor obox.mpimp-golm.mpg.de/OGDraw.html) [102].

Analysis of repeat sequences

SSRs were identified with MISA (https://webblast.ipk-gatersleben.de/misa/) [103], while the minimum numbers of mono-, di-, tri-, tetra-, penta-, and hexa-nucleotides were 10, 5, 4, 3, 3, and 3, respectively. The minimum threshold between SSRs was set to be 100 bp. Moreover, the forward, reverse, palindromic, and complementary repeat sequences were identified by REPuter (https://bibiserv.cebitec.uni-bielefeld.de/reputer/) with a minimum length of 30 bp and a hamming distance of 3 [104].

Analysis of codon composition

The codon composition and preference of the mitochondrial genome of *F. ussuriensis* were analyzed through CodonW (v1.4.4) [105]. It was noteworthy that the codon preference analysis included determining the GC content, codon numbers per gene, RSCU value, and ENC.

Prediction of RNA editing sites

The RNA-editing sites of the PCGs in *F. ussuriensis* mitogenome were predicted using Deepred-mt [80]. This tool has high accuracy through the use of a convolutional neural network model. The results with the probability above 0.9 were retained.

Selective pressure analysis

The mitochondrial PCGs of *A. cepa*, *C. comosum*, *A. officinalis*, *C. sativus*, and *L. tsingtauense* were used for analyzing the Ka/Ks values of the PCGs in *F. ussuriensis* mitogenome. The mitochondrial PCGs of the selected species and *F. ussuriensis* were aligned using the Mega (v7.0) [106], and the Ka/Ks values were determined by the KaKs_Calculator (v2.0) [107].

Homologous fragments between chloroplast genome and mitochondrial genome

Initially, the chloroplast genome sequences of *F. ussuriensis* were downloaded from NCBI. Homologous fragments between the chloroplast and mitochondrial genomes of *F. ussuriensis* were determined using BLASTN with the following parameters: -word_size 9, -gapopen 5, -gapextend 2, and -evalue 1e-5 [100]. Then the identified fragments were annotated using GeSeq [98], and the results were visualized on Circos package (v0.69) [108].

Phylogenetic analysis

The mitogenomes of 24 other plants were downloaded from the NCBI database to analyze the evolutionary position of *F. ussuriensis* among these species. PhyloSuite (v.1.2.2) was used for extracting the common PCGs among the mitogenomes of the selected species sequences [109], and MAFFT (v7.471) was utilized for aligning the extracted sequences [110]. Based on the maximum likelihood (ML) method, the phylogenetic tree was constructed by Mega (v7.0) with the GTR+G model and 1000 bootstrap replicates [106]. Furthermore, the heatmap of PCGs distribution was created using TBtools (v1.18) [111].

Nucleotide diversity (pi) analysis

The MAFFT software (v7.471) was employed for the global comparison of homologous gene sequences

Xie *et al. BMC Genomics*   (2024) 25:773

Page 15 of 18

across diverse species [110], while DnaSP (v5.0) was utilized to calculate the Pi value corresponding to each gene [112].

Colinear analysis

The mitochondrial genomes of *F. ussuriensis*, *L. tsing-tauense*, and 4 species of Asparagales were utilized to perform a colinear analysis by sequence similarity using the BLASTN software. The following parameters were applied: -evalue 1e-5, -word_size 9, -gapopen 5, -gapex-tend 2, -reward 2, -penalty -3 [100]. TBtools was used to generate a multiple synteny plot [111].

Abbreviations
SSRs   Simple sequence repeats
NCBI   National Center for Biotechnology Information
RSCU   Relative synonymous codon usage
PCGs   Protein-coding genes
tRNA   Transfer RNA
rRNA   Ribosomal RNA
ENC   Effective number of codon
Ka/Ks   Nonsynonymous-to-synonymous substitution ratio

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-024-10680-w.

> Supplementary Material 1.
>
> Supplementary Material 2.

Authors' contributions
HY and DZ conceived and designed the research. PX worked on genome assembly, performed the data analysis and wrote the original manuscript. JW, ML, TT and DY analyzed the data. DW, ZL, JT and DL contributed to the plant sample collection. LL, XY and HC improved the manuscript. All authors have read and agreed to the published version of the manuscript.

Availability of data and materials
The raw sequencing data for the Illumina and PacBio platforms and the mitogenome sequences have been deposited in NCBI (https://www.ncbi.nlm.nih.gov/) with accession numbers PRJNA1034960, SAMN38080166, SRR26678951, SRR26678952, and OR783162-OR783174, respectively.

Declarations

Ethics approval and consent to participate
All plant materials were collected following national and international standards and local laws and regulations. No specific permission is required to collect all samples described in this study.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]College of Biology and Agriculture, Jiamusi University, Jiamusi 154007, Heilongjiang, China. [2]Affiliated Stomatological Hospital, Jiamusi University, Jiamusi 154002, Heilongjiang, China. [3]School of Chinese Ethnic Medicine, Guizhou Minzu University, Guiyang 550025, Guizhou, China.

References

1. Osiewacz HD, Brust D, Hamann A, Kunstmann B, Luce K, Müller-Ohldach M, et al. Mitochondrial pathways governing stress resistance, life, and death in the fungal aging model Podospora anserina. Ann N Y Acad Sci. 2010;1197(1):54–66.
2. Roger AJ, Muñoz-Gómez SA, Kamikawa R. The origin and diversification of mitochondria. Curr Biol. 2017;27(21):1177–92.
3. Burki F. Mitochondrial evolution: going, going, gone. Curr Biol. 2016;26(10):410–2.
4. Han F, Qu Y, Chen Y, Xu LA, Bi C. Assembly and comparative analysis of the complete mitochondrial genome of *Salix wilsonii* using PacBio HiFi sequencing. Front Plant Sci. 2022;13:1031769.
5. Cheng Y, He X, Priyadarshani SVGN, Wang Y, Ye L, Shi C, et al. Assembly and comparative analysis of the complete mitochondrial genome of *Suaeda glauca*. BMC Genomics. 2021;22(1):1–15.
6. Kan SL, Shen TT, Gong P, Ran JH, Wang XQ. The complete mitochondrial genome of *Taxus cuspidata* (Taxaceae): eight proteincoding genes have transferred to the nuclear genome. BMC Evol Biol. 2020;20:10–26.
7. Maréchal A, Brisson N. Recombination and the maintenance of plant organelle genome stability. New Phytol. 2010;186(2):299–317.
8. Li J, Li J, Ma Y, Kou L, Wei J, Wang W. The complete mitochondrial genome of okra (*Abelmoschus esculentus*): Using nanopore long reads to investigate gene transfer from chloroplast genomes and rearrangements of mitochondrial DNA molecules. BMC Genomics. 2022;23(1):481.
9. Alverson AJ, Wei X, Rice DW, Stern DB, Barry K, Palmer JD. Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). Mol Biol Evol. 2010;27(6):1436–48.
10. Sloan DB, Alverson AJ, Chuckalovcak JP, Wu M, McCauley DE, Palmer JD, et al. Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. Plos Biol. 2012;10(1):e1001241.
11. Skippington E, Barkman TJ, Rice DW, Palmer JD. Miniaturized mitogenome of the parasitic plant *Viscum scurruloideum* is extremely divergent and dynamic and has lost all nad genes. Proc Natl Acad Sci. 2015;112(27):3515–24.
12. Gualberto JM, Mileshina D, Wallet C, Niazi AK, Weber-Lotfi F, Dietrich A. The plant mitochondrial genome: dynamics and maintenance. Biochimie. 2014;100:107–20.
13. Kozik A, Rowan BA, Lavelle D, Berke L, Schranz ME, Michelmore RW, et al. The alternative reality of plant mitochondrial DNA: One ring does not rule them all. Plos Genet. 2019;15(8):e1008373.
14. Zhong F, Ke W, Li Y, Chen X, Zhou T, Xu B, et al. Comprehensive analysis of the complete mitochondrial genomes of three *Coptis* species (*C chinensis*, *C deltoidea* and *C omeiensis*): the important medicinal plants in China. Front Plant Sci. 2023;14:1166420.
15. Yu R, Sun C, Zhong Y, Liu Y, Sanchez-Puerta MV, Mower JP, et al. The minicircular and extremely heteroplasmic mitogenome of the holo-parasitic plant *Rhopalocnemis phalloides*. Curr Biol. 2022;32(2):470–9.
16. Ma Q, Wang Y, Li S, Wen J, Zhu L, Yan K, et al. Assembly and comparative analysis of the first complete mitochondrial genome of *Acer truncatum* Bunge: A woody oil-tree species producing nervonic acid. BMC Plant Biol. 2022;22(1):1–17.

Xie *et al. BMC Genomics*   (2024) 25:773

Page 16 of 18

17. Adams KL, Palmer JD. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. Mol Phylogenet Evol. 2003;29(3):380–95.

18. Choi KS, Park S. Complete plastid and mitochondrial genomes of *Aeginetia indica* reveal intracellular gene transfer (IGT), horizontal gene transfer (HGT), and cytoplasmic male sterility (CMS). Int J Mol Sci. 2021;22(11):6143.

19. Cole LW, Guo W, Mower JP, Palmer JD. High and variable rates of repeat-mediated mitochondrial genome rearrangement in a genus of plants. Mol Biol Evol. 2018;35(11):2773–85.

20. Fertet A, Graindorge S, Koechler S, de Boer GJ, Guilloteau-Fonteny E, Gualberto JM. Sequence of the mitochondrial genome of *Lactuca virosa* suggests an unexpected role in *Lactuca sativa*'s evolution. Front Plant Sci. 2021;12:697136.

21. Day PD, Berger M, Hill L, Fay MF, Leitch AR, Leitch IJ, et al. Evolutionary relationships in the medicinally important genus *Fritillaria* L(Liliaceae). Mol Phylogenet Evol. 2014;80:11–9.

22. Jiao N, Song X, Song R, Yin D, Deng X. Diversity and structure of the microbial community in rhizosphere soil of *Fritillaria ussuriensis* at different health levels. PeerJ. 2022;10: e12778.

23. Li SL, Lin G, Chan SW, Li P. Determination of the major isosteroidal alkaloids in bulbs of *Fritillaria* by high-performance liquid chromatography coupled with evaporative light scattering detection. J Chromatogr A. 2001;909(2):207–14.

24. Gao WY, Fan L, Paek KY. Ultrastructure of amyloplasts and intercellular transport of old and new scales in *Fritillaria ussuriensis*. J Plant Biol. 1999;42:117–23.

25. Liu C, Chang J, Zhang L, Zhang J, Li S. Purification and antioxidant activity of a polysaccharide from bulbs of *Fritillaria ussuriensis* Maxim. Int J Biol Macromol. 2012;50(4):1075–80.

26. Pae HO, Oh H, Choi BM, Oh GS, Paik SG, Jeong S, et al. Differentiation-inducing effects of verticinone, an isosteroidal alkaloid isolated from the bulbus of *Fritillaria ussuriensis*, on human promyelocytic leukemia HL-60 cells. Biol Pharm Bull. 2002;25(11):1409–11.

27. Song J, Lu SZ, Liu LH, Wang M, Liu TX. Analysis of mechanophysical properties of *Fritillaria Ussuriensis* Maxim and designing of grade sieve machine. Adv Mater Res. 2013;764:165–8.

28. Yang ZD, Duan DZ. A new alkaloid from *Fritillaria ussuriensis* Maxim. Fitoterapia. 2012;83(1):137–41.

29. Lin Y, Wang XE, Yang X. The complete chloroplast genome and phylogenetic analysis of *Fritillaria ussuriensis* (Liliaceae: Fritillaria). Mitochondrial Dna B. 2019;4(1):2061–2.

30. Shan Y, Li J, Zhang X, Yu J. The complete mitochondrial genome of *Amorphophallus albus* and development of molecular markers for five *Amorphophallus* species based on mitochondrial DNA. Front Plant Sci. 2023;14:1180417.

31. Zhang Z, Li J, Zhao XQ, Wang J, Wong GKS, Yu J. KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. Genom Proteom Bioinf. 2006;4(4):259–63.

32. Guo W, Grewe F, Fan W, Young GJ, Knoop V, Palmer JD, et al. *Ginkgo* and *Welwitschia* mitogenomes reveal extreme contrasts in gymnosperm. Mol Biol Evol. 2016;33(6):1448–60.

33. Fang B, Li J, Zhao Q, Liang Y, Yu J. Assembly of the complete mitochondrial genome of Chinese plum (*Prunus salicina*): characterization of genome recombination and RNA editing sites. Genes. 2021;12(12):1970.

34. Bi Y, Zhang M, Xue J, Dong R, Du Y, Zhang X, et al. Chloroplast genomic resources for phylogeny and DNA barcoding: a case study on *Fritillaria*. Sci Rep-Uk. 2018;8(1):1184.

35. Liberatore KL, Dukowic-Schulze S, Miller ME, Chen C, Kianian SF. The role of mitochondria in plant development and stress tolerance. Free Radical Bio Med. 2016;100:238–56.

36. Petersen G, Anderson B, Braun HP, Meyer EH, Møller IM. Mitochondria in parasitic plants. Mitochondrion. 2020;52:173–82.

37. Sloan DB, Warren JM, Williams AM, Wu Z, Abdel-Ghany SE, Chicco AJ, et al. Cytonuclear integration and co-evolution. Nat Rev Genet. 2018;19(10):635–48.

38. Niu Y, Lu Y, Song W, He X, Liu Z, Zheng C, et al. Assembly and comparative analysis of the complete mitochondrial genome of three Macadamia species (*M integrifolia*, *M ternifolia* and *M tetraphylla*). PLoS ONE. 2022;17(5):e0263545.

39. Wang X, Zhang R, Yun Q, Xu Y, Zhao G, Liu J, et al. Comprehensive analysis of complete mitochondrial genome of *Sapindus mukorossi* Gaertn: an important industrial oil tree species in China. Ind Crop Prod. 2021;174: 114210.

40. Li J, Xu Y, Shan Y, Pei X, Yong S, Liu C, et al. Assembly of the complete mitochondrial genome of an endemic plant, *Scutellaria tsinyunensis*, revealed the existence of two conformations generated by a repeat-mediated recombination. Planta. 2021;254:1–16.

41. Gui S, Wu Z, Zhang H, Zheng Y, Zhu Z, Liang D, et al. The mitochondrial genome map of *Nelumbo nucifera* reveals ancient evolutionary features. Sci Rep-Uk. 2016;6(1):30158.

42. Baeza JA, León FJGD. Are we there yet? Benchmarking low-coverage nanopore long-read sequencing for the assembling of mitochondrial genomes using the vulnerable silky shark *Carcarhinus falciformis*. BMC Genomics. 2022;23(1):1–18.

43. Ren Z, Niu X, Lv T, Wang Y, Caraballo-Ortiz MA, Su X. The complete mitochondrial genome of *Panthera pardus* (Felidae: Pantheriinae), a first-class national-protected wild animal from China. Conserv Genet Resour. 2019;11:389–92.

44. Sloan DB. One ring to rule them all? Genome sequencing provides new insights into the 'master circle'model of plant mitochondrial DNA structure. New Phytol. 2013;200(4):978–85.

45. Jackman SD, Coombe L, Warren RL, Kirk H, Trinh E, MacLeod T, et al. Complete mitochondrial genome of a gymnosperm, Sitka spruce (*Picea sitchensis*), indicates a complex physical structure. Genome Biol Evol. 2020;12(7):1174–9.

46. You C, Cui T, Zhang C, Zang S, Su Y, Que Y. Assembly of the Complete Mitochondrial Genome of *Gelsemium elegans* Revealed the Existence of Homologous Conformations Generated by a Repeat Mediated Recombination. Int J Mol Sci. 2022;24(1):527.

47. Guo W, Zhu A, Fan W, Mower JP. Complete mitochondrial genomes from the ferns *Ophioglossum californicum* and *Psilotum nudum* are highly repetitive with the largest organellar introns. New Phytol. 2017;213(1):391–403.

48. Rice DW, Alverson AJ, Richardson AO, Young GJ, Sanchez-Puerta MV, Munzinger J, et al. Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm *Amborella*. Science. 2013;342(6165):1468–73.

49. Wu ZQ, Liao XZ, Zhang XN, Tembrock LR, Broz A. Genomic architectural variation of plant mitochondria-A review of multichromosomal structuring. J Syst Evol. 2022;60(1):160–8.

50. Yuan Y, Jin X, Liu J, Zhao X, Zhou J, Wang X, et al. The *Gastrodia elata* genome provides insights into plant adaptation to heterotrophy. Nat Commun. 2018;9(1):1615.

51. Wu Z, Cuthbert JM, Taylor DR, Sloan DB. The massive mitochondrial genome of the angiosperm *Silene noctiflora* is evolving by gain or loss of entire chromosomes. Proc Natl Acad Sci. 2015;112(33):10185–91.

52. Liu D, Guo H, Zhu J, Qu K, Chen Y, Guo Y, et al. Complex physical structure of complete mitochondrial genome of *Quercus acutissima* (Fagaceae): A significant energy plant. Genes. 2022;13(8):1321.

53. Odahara M, Nakamura K, Sekine Y, Oshima T. Ultra-deep sequencing reveals dramatic alteration of organellar genomes in *Physcomitrella patens* due to biased asymmetric recombination. Commun Biol. 2021;4(1):633.

54. Kuang DY, Wu H, Wang YL, Gao LM, Zhang SZ, Lu L. Complete chloroplast genome sequence of *Magnolia kwangsiensis* (Magnoliaceae): implication for DNA barcoding and population genetics. Genome. 2011;54(8):663–73.

55. Qian J, Song J, Gao H, Zhu Y, Xu J, Pang X, et al. The complete chloroplast genome sequence of the medicinal plant *Salvia miltiorrhiza*. PLoS ONE. 2013;8(2):e57607.

56. Allen JO, Fauron CM, Minx P, Roark L, Oddiraju S, Lin GN, et al. Comparisons among two fertile and three male-sterile mitochondrial genomes of maize. Genetics. 2007;177(2):1173–92.

57. Bi C, Qu Y, Hou J, Wu K, Ye N, Yin T. Deciphering the multi-chromosomal mitochondrial genome of *Populus simonii*. Front Plant Sci. 2022;13:914635.

58. Bergthorsson U, Adams KL, Thomason B, Palmer JD. Widespread horizontal transfer of mitochondrial genes in flowering plants. Nature. 2003;424(6945):197–201.

Xie *et al. BMC Genomics*   (2024) 25:773

Page 17 of 18

59. Goremykin VV, Lockhart PJ, Viola R, Velasco R. The mitochondrial genome of *Malus domestica* and the import-driven hypothesis of mitochondrial genome expansion in seed plants. Plant J. 2012;71(4):615–26.

60. Goremykin VV, Salamini F, Velasco R, Viola R. Mitochondrial DNA of *Vitis vinifera* and the issue of rampant horizontal gene transfer. Mol Biol Evol. 2009;26(1):99–110.

61. Chen Z, Zhao N, Li S, Grover CE, Nie H, Wendel JF, et al. Plant mitochondrial genome evolution and cytoplasmic male sterility. Crit Rev Plant Sci. 2017;36(1):55–69.

62. Woodson JD, Chory J. Coordination of gene expression between organellar and nuclear genomes. Crit Rev Plant Sci. 2008;9(5):383–95.

63. Miyata SI, Nakazono M, Hirai A. Transcription of plastid-derived tRNA genes in rice mitochondria. Curr Genet. 1998;34:216–20.

64. Li Y, Gu M, Liu X, Lin J, Jiang H, Song H, et al. Sequencing and analysis of the complete mitochondrial genomes of *Toona sinensis* and *Toona ciliata* reveal evolutionary features of *Toona*. BMC Genomics. 2023;24(1):1–11.

65. Leister D, Kleine T. Role of intercompartmental DNA transfer in producing genetic diversity. Int Rev Cel Mol Bio. 2011;291:73–114.

66. Tang D, Huang S, Quan C, Huang Y, Miao J, Wei F. Mitochondrial genome characteristics and phylogenetic analysis of the medicinal and edible plant *Mesona chinensis* Benth. Front Genet. 2023;13:1056389.

67. Sharp PM, Tuohy TM, Mosurski KR. Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. Nucleic Acids Res. 1986;14(13):5125–43.

68. Shidhi PR, Biju VC, Anu S, Vipin CL, Deelip KR, Achuthsankar SN. Genome characterization, comparison and phylogenetic analysis of complete mitochondrial genome of *Evolvulus alsinoides* reveals highly rearranged gene order in solanales. Life-Basel. 2021;11(8):769.

69. Niu Y, Gao C, Liu J. Complete mitochondrial genomes of three *Mangifera* species, their genomic structure and gene transfer from chloroplast genomes. BMC Genomics. 2022;23(1):1–8.

70. Romero H, Zavala A, Musto H. Codon usage in *Chlamydia trachomatis* is the result of strand-specific mutational biases and a complex pattern of selective forces. Nucleic Acids Res. 2000;28(10):2084–90.

71. Lai C, Wang J, Kan S, Zhang S, Li P, Reeve WG, et al. Comparative analysis of mitochondrial genomes of *Broussonetia* spp. (Moraceae) reveals heterogeneity in structure, synteny, intercellular gene transfer, and RNA editing. Front Plant Sci. 2022;13:1052151.

72. Bi C, Paterson AH, Wang X, Xu Y, Wu D, Qu Y, et al. Analysis of the complete mitochondrial genome sequence of the diploid cotton *Gossypium raimondii* by comparative genomics approaches. Biomed Res Int. 2016;2016(1):5040598.

73. Qiao Y, Zhang X, Li Z, Song Y, Sun Z. Assembly and comparative analysis of the complete mitochondrial genome of *Bupleurum chinense* DC. BMC Genomics. 2022;23(1):1–17.

74. Unseld M, Marienfeld JR, Brandt P, Brennicke A. The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. Nat genet. 1997;15(1):57–61.

75. Wang Y, Chen S, Chen J, Chen C, Lin X, Peng H, et al. Characterization and phylogenetic analysis of the complete mitochondrial genome sequence of *Photinia serratifolia*. Sci Rep-Uk. 2023;13(1):770.

76. Xu Y, Dong Y, Cheng W, Wu K, Gao H, Liu L, et al. Characterization and phylogenetic analysis of the complete mitochondrial genome sequence of *Diospyros oleifera*, the first representative from the family Ebenaceae. Heliyon. 2022;8(7):e09870.

77. Li Z, Liu Z, Zhong W, Huang M, Wu N, Xie Y, et al. Large-scale identification of human protein function using topological features of interaction network. Sci Rep-Uk. 2016;6(1):37179.

78. Wang S, Li D, Yao X, Song Q, Wang Z, Zhang Q, et al. Evolution and diversification of kiwifruit mitogenomes through extensive whole-genome rearrangement and mosaic loss of intergenic sequences in a highly variable region. Genome Biol Evol. 2019;11(4):1192–206.

79. Wei L, Liu TJ, Hao G, Ge XJ, Yan HF. Comparative analyses of three complete *Primula* mitogenomes with insights into mitogenome size variation in Ericales. BMC Genomics. 2022;23(1):1–16.

80. Edera AA, Small I, Milone DH, Sanchez-Puerta MV. Deepred-Mt: Deep representation learning for predicting C-to-U RNA editing in plant mitochondria. Comput Biol Med. 2021;136: 104682.

81. Liu G, Cao D, Li S, Su A, Geng J, Grover CE, et al. The complete mitochondrial genome of *Gossypium hirsutum* and evolutionary analysis of higher plant mitochondrial genomes. PLoS ONE. 2013;8(8): e69476.

82. Zhou P, Zhang Q, Li F, Huang J, Zhang M. Assembly and comparative analysis of the complete mitochondrial genome of *Ilex metabaptista* (Aquifoliaceae), a Chinese endemic species with a narrow distribution. BMC Plant Biol. 2023;23(1):393.

83. Bi C, Lu N, Xu Y, He C, Lu Z. Characterization and analysis of the mitochondrial genome of common bean (*Phaseolus vulgaris*) by comparative genomic approaches. Int J Mol Sci. 2020;21(11):3778.

84. Shtolz N, Mishmar D. The mitochondrial genome–on selective constraints and signatures at the organism, cell, and single mitochondrion levels. Front Ecol Evol. 2019;7:342.

85. Tanaka Y, Tsuda M, Yasumoto K, Yamagishi H, Terachi T. A complete mitochondrial genome sequence of Ogura-type male-sterile cytoplasm and its comparative analysis with that of normal cytoplasm in radish (*Raphanus sativus* L.). BMC Genomics. 2012;13:1–12.

86. Suck D. DNA recognition by DNase I. J Mol Recognit. 1994;7(2):65–70.

87. Hao J, Liang Y, Su Y, Wang T. The complete mitochondrial genome of *Ophioglossum vulgatum* L. is with highly repetitive sequences: intergenomic fragment transfer and phylogenetic analysis. Genes. 2022;13(7):1287.

88. Lai X, Zhang Z, Zhang Z, Liu S, Bai C, Chen Z, et al. Integrated microbiome-metabolome-genome axis data of Laiwu and Lulai pigs. Sci Data. 2023;10(1):280.

89. Jin X, Wang R, Wei T, Tang D, Xu T. Complete mitochondrial genome sequence of *Tridentiger bifasciatus* and *Tridentiger barbatus* (Perciformes, Gobiidae): a mitogenomic perspective on the phylogenetic relationships of Gobiidae. Mol Biol Rep. 2015;42:253–65.

90. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20.

91. Jin JJ, Yu WB, Yang JB, Song Y, DePamphilis CW, Yi TS, et al. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol. 2020;21:1–31.

92. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 2017;27(5):722–36.

93. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754–60.

94. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078–9.

95. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19(5):455–77.

96. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of de novo genome assemblies. Bioinformatics. 2015;31(20):3350–2.

97. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS ONE. 2014;9(11):e112963.

98. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, et al. GeSeq-versatile and accurate annotation of organelle genomes. Nucleic Acids Res. 2017;45(W1):6–11.

99. Lowe TM, Chan PP. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res. 2016;44(W1):54–7.

100. Chen Y, Ye W, Zhang Y, Xu Y. High speed BLASTN: an accelerated MegaBLAST search tool. Nucleic Acids Res. 2015;43(16):7762–8.

101. Lewis SE, Searle SMJ, Harris N, Gibson M, Iyer V, Richter J, et al. Apollo: a sequence annotation editor. Genome Biol. 2002;3:1–14.

102. Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 13 1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res. 2019;47(W1):59–64.

103. Beier S, Thiel T, Münch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. Bioinformatics. 2017;33(16):2583–5.

104. 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 Res. 2001;29(22):4633–42.

105. Ma QP, Li C, Wang J, Wang Y, Ding ZT. Analysis of synonymous codon usage in *FAD7* genes from different plant species. Genet Mol Res. 2015;14:1414–22.

106. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 70 for bigger datasets. Mol Biol Evol. 2016;33(7):1870–4.
107. Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. Genom Proteom Bioinf. 2010;8(1):77–80.
108. Zhang H, Meltzer P, Davis S. RCircos: an R package for Circos 2D track plots. BMC Bioinformatics. 2013;14:1–5.
109. Zhang D, Gao F, Jakovlić I, Zou H, Zhang J, Li WX, et al. PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. Mol Ecol Resour. 2020;20(1):348–55.
110. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–80.
111. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, et al. TBtools: an integrative toolkit developed for interactive analyses of big biological data. Mol Plant. 2020;13(8):1194–202.
112. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009;25(11):1451–2.

Publisher's Note