Analysis of complete chloroplast genome sequences and insight into the phylogenetic relationships of Ferula L.

Lei Yang1,2,3, Ozodbek Abduraimov4, Komiljon Tojibaev4, Khabibullo Shomurodov4, Yuan-Ming Zhang1,3,5 and Wen-Jun Li1,3,5,6*

Abstract

**Background:** Ferula L. is one of the largest and most taxonomically complicated genera as well as being an important medicinal plant resource in the family Apiaceae. To investigate the plastome features and phylogenetic relationships of Ferula and its neighboring genera Soranthus Ledeb., Schumannia Kuntze., and Talassia Korovin, we sequenced 14 complete plastomes of 12 species.

**Results:** The size of the 14 complete chloroplast genomes ranged from 165,607 to 167,013 base pairs (bp) encoding 132 distinct genes (87 protein-coding, 37 tRNA, and 8 rRNA genes), and showed a typical quadripartite structure with a pair of inverted repeats (IR) regions. Based on comparative analysis, we found that the 14 plastomes were similar in codon usage, repeat sequence, simple sequence repeats (SSRs), and IR borders, and had significant collinearity. Based on our phylogenetic analyses, Soranthus, Schumannia, and Talassia should be considered synonymous with Ferula. Six highly divergent regions (rps16/trnQ-UUG, trnS-UGA/psbZ, psbH/petB, ycf1/ndhF, rpl32, and ycf1) were also detected, which may represent potential molecular markers, and combined with selective pressure analysis, the weak positive selection gene ccsA may be a discriminating DNA barcode for Ferula species.

**Conclusion:** Plastids contain abundant informative sites for resolving phylogenetic relationships. Combined with previous studies, we suggest that there is still much room for improvement in the classification of Ferula. Overall, our study provides new insights into the plastome evolution, phylogeny, and taxonomy of this genus.

**Keywords:** Ferula, Chloroplast genome, Comparative analysis, Phylogenetic relationships

Background

Ferula L., a perennial single- or multi-bearing herb in the family Apiaceae, contains approximately 170 species mainly distributed in the Mediterranean region of southern Europe, northern Africa, Iran, Afghanistan, Central Asia, Siberia, Russia, India, and Pakistan [1]. Some Ferula species can secrete aromatic resins that have the aroma of onions and garlic, and these aromatic resins have insecticidal and fatigue-reducing properties, can be used to treat stomach diseases, dyspepsia, and abdominal pain, and is a plant resource with potentially important medicinal value [2–5].

Due to the similar morphologies and wide distribution of its constituent species, Ferula is recognized as one of the most taxonomically complicated genera within the Apiaceae [5–8]. Ferula was originally divided into three sections, Euferrula Boiss., Peucedanoides Boiss., and Scorodosma Bunge [9], and later into four subgenera,
Fifty years later, Korovin [11] systematically divided the genus into six subgenera according to fruit, inflorescence, petals, and the number of vitta in fruits, namely *Scorodosma* (Bunge) Drude, *Soranthus* Ledeb., and *Euryangium* (Kaufm.) Drude [10]. Moreover, the relationship between *Ferula* and some neighboring genera has been debated frequently, especially in the cases of *Soranthus* Ledeb., *Schumannia* Kunzt., and *Talassia* Korovin. *Soranthus* was established as a monotypic genus by Ledebour [17], with *S. sibiricus* (Willd.) Kosso-Pol. considered a combination based on *S. sibiricus* as a monotypic genus by Ledebour [17], with *Ferula* subgenus. Later, Korovin [11] established the genus *Merwia* (B. Fedtsch.) Korovin, which was also listed within *Ferula* by Bunge [18], Drude [10], Safina and Pimenov [14], Piminov [1], and Tojibaev et al. [23], but not in the Flora of the Soviet Union [19], the Flora of China [20], or the Flora Xinjiangensis [21]. *Soranthus* was established as a later homonym of *Schumannia* Kunzt. [22]. *S. turcomanica* was also listed within *Ferula* by Bunge [18], Drude [10], Safina and Pimenov [14], Piminov [1], and Tojibaev et al. [23], but not in the Flora of the Soviet Union [19], the Flora of China [20], or the Flora Xinjiangensis [21]. *Talassia renardi* (Regel & Schmalt.) Korovin and *T. transiliensis* (Herder) Korovin, which were isolated from *Peucedanum transiliensis* Regel & Herder from the genus *Peucedanum* L. [24], were recorded in the Flora of Kazakhstan and subsequently transferred to *Ferula* by Pimenov [25] and admitted by Govaerts et al. [26]. However, *Talassia* has also been listed as an independent genus in some Chinese floras [20, 21]. In addition, some studies have suggested that *Schumannia* should be merged with *Soranthus* based on their fruit, pollen morphology, and serological investigations [27, 28]. Recently, some molecular phylogeny based on the relatively limited number of nrDNA and cpDNA sequences indicated that *Soranthus*, *Schumannia*, and *Talassia* were embedded in *Ferula*, but show low support values [7, 16, 29].

Chloroplasts are independent organelles in plant cells that have their own complete set of genomes and typically covalently closed circular DNA, which exists in cells as multiple copies [30]. The chloroplast genomes of higher plants have a highly conserved tetrads structure involving inverted repeat sequences (IRs) and large single-copy (LSC) and small single-copy (SSC) regions [31]. Chloroplast genomes are relatively conserved in terms of gene number and sequence in terrestrial plants [32]. The sizes of chloroplast genomes are generally within the range of 115–165 kb, and genome size variation is mainly affected by reverse repeat length variation. Additionally, chloroplast genomes usually exhibit uniparental inheritance and low nucleotide substitution rates [33]. At present, chloroplast genome sequences and nuclear genome sequences can be obtained using shallow whole genome sequencing technology. This is considered an effective means of improving the rate of species identification and has been developed as a tool for plant phylogenetic studies at different taxonomic levels [34–42]. For example, the complete plastomes and nrDNA sequences obtained based on shallow genome sequencing have greatly improved the species identification rate of *Rhododendron*, which is also difficult to classify [43]. Thus, the complete plastomes might insight into the phylogenetic relationships of *Ferula* and its neighboring genera.

Here, we used plastomes to infer the phylogenetic relationships between *Ferula* and its confused neighboring genera. Fourteen newly sequenced plastomes of *Ferula* (including *Soranthus*, *Schumannia*, and *Talassia*) were analyzed to (1) conduct comprehensive research on the *Ferula* chloroplast genome; (2) identify hotspot regions, microsatellite types, and comparative genomic divergence; (3) analyze the relationships
between *Ferula*, *Soranthus*, *Schumannia*, and *Talassia* based on their complete chloroplast genomes; and (4) serve as a reference for subsequent phylogenomic studies of the genus *Ferula*.

**Results**

**Chloroplast genome features**

The 14 complete cp genomes ranged from 165,607 to 167,013 bp. Newly sequenced *Ferula* chloroplast genome maps are shown in Fig. 1. All cp genomes possessed the typical quadripartite structure of angiosperms, consisting of a pair of inverted repeat regions (IRs: 31,392–31,880 bp) and a circular molecular structure (Fig. 1; Table 1). All 14 cp genomes possessed 133 distinct genes arranged in the same order, including 87 protein-coding genes, 37 tRNA genes, and eight rRNA genes. Of these, 14 protein-coding genes and eight tRNAs contained at least one intron. The genes were classified

---

*Fig. 1* Chloroplast genome maps for *Ferula* L. Genes on the inside of the circle are transcribed clockwise and those on the outside are transcribed counterclockwise. The darker gray inner circle corresponds to the GC content, whereas the lighter gray indicates the AT content. Different colors represent different functional genes.
into the following four groups based on their functions: (1) 74 self-replication genes; (2) 45 photosynthesis-related genes (in Rubisco, ATP synthase, Photosystem I, cytochrome b/f complex, photosystem II, and NADH dehydrogenase groups); and 13 other genes including (3) six genes with known functions (\textit{matK}, \textit{cemA}, \textit{accD}, \textit{ccsA}, \textit{infA}, and \textit{clpP}) and (4) seven genes with unknown functions (\textit{ycf1}(2), \textit{ycf2}(2), \textit{ycf3}, \textit{ycf4}, and \textit{ycf15}) (Table 2). The total GC content for 12 sequenced species was 37.8–38.0% (Table 1).

### Codon usage

The RSCU values of all codons are shown in Fig. 2 in the form of a heatmap; the red values indicate higher RSCU values, and the blue values indicate lower RSCU values. For the \textit{Ferula} species, the most commonly used transcription initiation codon was AUG, the most commonly used termination codon was UAA, and the initiation codon AUU only existed in \textit{F. olivacea}. Except for the initiation codon and termination codon, the most used transcription codon was UTA, and AGC showed the lowest RSCU values; the most abundant amino acid (AA) was leucine, while cysteine was the lowest frequency AA. Except for tryptophan, all AAs had more than one synonymous codon, and three AAs (leucine, serine, and arginine) had the most (six) synonymous codons. The use of one codon, UGG, showed no bias (RSCU = 1) (Table S2).

### Repeat structure analysis

Forward, palindromic, reverse, and complementary repeats were detected in 14 \textit{Ferula} plastomes. Except for IR repeats, 837 repeats were identified in total; the numbers of forward repeats (398) and palindromic repeats (421) were much higher than the complement repeats (7) and reverse repeats (11). Reverse and complementary repeats were missing in four samples (\textit{F. sibirica}, \textit{F. kelifi}, \textit{F. ovina}, and \textit{F. karelinii}). \textit{F. kelifi} contained the maximum number of repeats (94), whereas \textit{F. equisetacea} and \textit{F. olivacea} contained the least (46) (Table S3). A total of 1,061 SSRs were identified in the 14 species, six of which did not have pentanucleotides, and hexanucleotides were only found in \textit{F. olivacea}. Additionally, mononucleotides were most frequent followed by dinucleotides, tetranucleotides, trinucleotides, pentanucleotides, and hexanucleotides. \textit{F. transiliensis} contained the highest number of SSRs (82), whereas \textit{F. oopoda} contained the least (69). Poly (A/T) SSRs were typically most common, while poly (C/G) repeats were extremely rare (Table S4).

### Comparisons of border and sequence identity

Single-copy and inverted repeat borders were examined; \textit{F. kelifi} and \textit{F. equisetacea} harbored the longest (31,880 bp) and shortest (31,392 bp) IR regions, respectively. Among all 14 \textit{Ferula} species, \textit{rps19} is embedded in the LSC/IRb junction region and only 81 bp with the IRb overlap; \textit{ycf1} spans SSC/IRa and occupies a long section in both regions; and \textit{trnH} occurs in the LSC region and is only 5 bp away from IRa, except for \textit{F. sibirica} (11 bp). The variety of IRb/SSC is relatively high, most (or all) of which occur in the SSC region, and the overlap with the IRb region varied from -18 to 16 bp (Fig. 3). According to the sequence identity plots, the 14 sequences were almost identical in their genetic structure and showed a very high degree of identity.

### Table 1

| Sample ID | Species           | Genome size (bp) | GC content (%) | LSC (bp) | SSC (bp) | IR (bp) |
|-----------|-------------------|------------------|----------------|----------|----------|---------|
| L-6       | \textit{F. gigantea} | 166,222          | 37.9           | 85,383   | 17,563   | 31,638  |
| L-12      | \textit{F. equisetacea} | 165,607         | 37.9           | 85,231   | 17,592   | 31,392  |
| L-14      | \textit{F. sibirica} 1 | 166,648          | 37.9           | 85,346   | 17,632   | 31,835  |
| L-15      | \textit{F. litwinowiana} | 166,554         | 38             | 85,226   | 17,614   | 31,857  |
| L-23      | \textit{F. kelifi}    | 166,712          | 38             | 85,323   | 17,629   | 31,880  |
| L-29      | \textit{F. transiliensis} 1 | 166,547         | 38             | 85,306   | 17,599   | 31,821  |
| L-58      | \textit{F. renardii}  | 166,520          | 38             | 85,317   | 17,559   | 31,822  |
| L-59      | \textit{F. oopoda}   | 166,565          | 38             | 85,328   | 17,595   | 31,821  |
| L-60      | \textit{F. fedtschenkoana} | 166,445         | 38             | 85,205   | 17,568   | 31,836  |
| L-88      | \textit{F. ovina}    | 166,450          | 38             | 85,341   | 17,561   | 31,774  |
| L-101     | \textit{F. olivacea} | 167,013          | 37.8           | 85,598   | 17,687   | 31,864  |
| L-108     | \textit{F. transiliensis} 3 | 166,520         | 38             | 85,293   | 17,585   | 31,821  |
| L-109     | \textit{F. sibirica} 3 | 166,644          | 37.9           | 85,348   | 17,626   | 31,835  |
| L-111     | \textit{F. karelinii} | 166,037          | 37.9           | 84,839   | 17,592   | 31,803  |
conservation (Fig. 4). To determine divergent hotspots, nucleotide diversity (Pi) values were calculated (Fig. 5, Table S5), yielding a maximum value of 0.01019 in ycf1. The SSC area showed the maximum nucleotide diversity followed by the LSC region, and the IR regions had the lowest Pi value. Additionally, six highly divergent regions (> 0.006) were detected in the LSC region (rps16/trnQ-UUG, trnS-UGA/psbZ, psbH/petB), SSC region (ycf1/ndhF, rpl32, ycf1), and IR region (0).

We calculated the Ka/Ks ratios of the 79 common protein-coding genes to reveal selection patterns among the protein-coding genes. The Ka/Ks ratios of most of the genes were less than 0.5 or could not be computed because either the Ka or Ks value was zero; three genes (ccsA, ndhC, and ycf2) had values greater than 1; and the total Ka/Ks ratio of all genes was 0.5331 (Table S6). In addition, we found several annotation errors (ndhH and ccsA) in the previously reported sequences of F. sinkiangensis (MW411057).

**Phylogenetic analyses**

To determine the phylogenetic relationship of Soranthus Lede., Schumannia Kuntz., Talassia Korovin, and Ferula L., 25 chloroplast genomes were used to construct maximum likelihood (ML) and Bayesian inference (BI) phylogenetic trees. These included 10 samples of 10 Ferula species (including F. sinkiangensis, GenBank accession no. MW411057), two samples of Soranthus, two samples of Schumannia, one sample of Talassia, and nine other Apiaceae genera, i.e., Caucaulis L., Daucus L., Cuminum L., Anthriscus Pers., Aegopodium L., Cyclospormum Lag., Apium L., Cryptotaenia DC., and Oenanthe L., with an outgroup of Diplopanax stachyanthus Hand.-Mazz (Fig. 6).

The ML and BI topologies were highly supported. Ten selected genera formed 10 monophyletic groups, all of which had support values of 100 or 1 in the ML and BI trees, respectively. Ferula was divided into three main lineages (A, B and C) with maximal support (PP = 1, BS ≥ 97%), and three genera (Soranthus meyeri,

| Table 2 | List of genes in the chloroplast genomes of the examined Ferula species |
|---------|---------------------------------------------------------------|
| **Category** | **Gene group** | **Gene name** |
| **Photosynthesis** | Subunits of photosystem I | psaA, psaB, psaC, psaI |
| | Subunits of photosystem II | psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ |
| | Subunits of NADH dehydrogenase | ndhA4, ndhB4, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK |
| | Subunits of cytochrome b/f complex | petA, petB4, petD, petG, petL, petN |
| | Subunits of ATP synthase | atpA, atpB, atpE, atpF3, atpH, atpl |
| | Large subunit of rubisco | rbcL |
| | Subunits photoschlorophyllide reductase | - |
| **Self-replication** | Proteins of large ribosomal subunit | rpl14, rpl164, rpl20, rpl22, rpl236, rpl32, rpl33, rpl36 |
| | Proteins of small ribosomal subunit | rps11, rps128, rps14, rps15, rps166, rps18, rps19, rps2, rps3, rps4, rps78, rps8 |
| | Subunits of RNA polymerase | rpoA, rpoB, rpoCl4, rpoC2 |
| | Ribosomal RNAs | rrn1612, rrn234, rrn455 |
| | Transfer RNAs | trnA-UGC, trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnG-GCC, trnG-UCC5, trnH-GUG, trnI-CAU3, trnK-GAA, trnL-UUA5, trnL-CAA6, trnL-UUG, trnM-CAU, trnN-GUU, trnP-UUG, trnQ-UUG, trnR-ACG4, trnR-UCU, trnS-GCU, trnS-UGA, trnT-GGU, trnT-UGU, trnV-GAC3, trnW-CCA, trnY-GUA, trnM-CAU |
| **Other genes** | Matrerase | matK |
| | Protease | clpP4 |
| | Envelope membrane protein | cemA |
| | Acetyl-CoA carboxylase | accD |
| | C-type cytochrome synthesis gene | ccsA |
| | Translation initiation factor | infA |
| | Other | - |
| **Genes of unknown function** | Conserved hypothetical chloroplast ORF | ycf1, ycf15, ycf2, ycf3, ycf4 |

**Notes:** GeneA: Gene with one intron
GeneB: Gene with two introns
GeneC: Gene: Pseudo gene
GeneD: Number of copies of multi-copy genes
Schumannia karelinii, Talassia transiliensis) were clustered into Ferula. Lineage A contained 11 Ferula species, S. meyeri, T. transiliensis, and S. karelinii. Within this lineage, S. sibirica and S. karelinii are sister species, and F. sinkiangensis and F. litwinowiana are sister species. Lineages B and C contained only F. equisetacea and F. olibacea, respectively. Moreover, Ferula and four genera of Apiaceae formed a monophyletic group.

Discussion
Comparison of Ferula plastid genome
Plastomes are considered an effective means used in taxonomic and evolutionary studies to assess evolutionary relationships and compare genome structure at different taxonomic levels [34–42]. Generally, the plastomes are highly conserved in genome structure, gene order, and gene content [32]. In this study, all 14 plastomes are divided into four regions consisting of an LSC (84,839–85,598 bp), an SSC (17,559–17,687 bp), and two IRs (31,392–31,880 bp). The comparative analysis of 14 complete plastomes showed great similarities in terms of genome length (165,607–167,013 bp), structure, IR/SC borders and GC content (37.8–38.0), the equal number of CDS, rRNA, and tRNA genes, and no rearrangement or a good collinearity relationship among them (Fig. 1; Table S1), indicated that the Ferula are relatively conserved.

Although the IR region is thought to be the most conserved region in the chloroplast genome, contraction and expansion of the IR region is common, and is the main reason for the variation in chloroplast genome size [44–46]. The junction of IRb/LSC located at ycf2 gene is defined as the type without any expansion or contraction [47]. In this study, we observed that 14 sequenced complete plastomes exhibited significant IR expansion (Fig. 3). All the species expanded into rps19 at the IRb/LSC junction region, contributing to rps19 fragment in the IRa/LSC region, and they also expanded into ycf1 at the IRb/SSC junction region, leading to an overlap between the ycf1 pseudo-gene and ndhF. This was consistent with previous studies, in which the pseudogenes ycf1 and rps19 were produced by contraction and expansion of the IR region in angiosperms [48–50].

RSCU value is the ratio of specific codon usage frequency to desired frequency, which can eradicate the influence of amino acid composition on codon usage and promotes the detection of synonymous codons [51, 52]. Generally, the content of A/T was higher than that of G/C in plastomes codons and A/T is preferred in the third codon position [53], the bias also showed in the Ferula plastomes (Fig. 2). Leucine was encoded by 6 codons, the order of codon preference was UTA > CUT > UTG > CUA > CUC > CUG, which following previous studies [54, 55]. The analysis of RSCU can provide a basis for studying the specific mechanism of synonymous codon bias preference.
in different species, which plays a crucial role in molecular biology basis research [56, 57].

As a primary source of molecular markers, SSRs have been widely used in *Ferula* genetic diversity studies because of their high polymorphism rate and abundant variation at the species level [58, 59]. In our study, we identified 837 repeats (Table S3) and 1,061 SSRs (Table S4) in the 14 *Ferula* samples. In which, the single nucleotide and dinucleotide repeats were common, which is consistent with the results of previous studies [55, 60]. In general, during the evolutionary process of species, most repeated sequences in the genome are distributed in the

| Ferula oopoda | 166,555 bp | JLB  | JSB | JSA | JLA |
|---------------|-----------|------|-----|-----|-----|
| Ferula olivacea | 167,013 bp | JLB  | JSB | JSA | JLA |
| Ferula Illwinowiana | 165,554 bp | JLB  | JSB | JSA | JLA |
| Ferula kelii | 166,712 bp | JLB  | JSB | JSA | JLA |
| Ferula gigantea | 166,222 bp | JLB  | JSB | JSA | JLA |
| Ferula fedtschenkoana | 166,445 bp | JLB  | JSB | JSA | JLA |
| Ferula equisetacea | 165,607 bp | JLB  | JSB | JSA | JLA |
| Ferula ovina | 166,450 bp | JLB  | JSB | JSA | JLA |
| Ferula karelinii | 166,037 bp | JLB  | JSB | JSA | JLA |
| Ferula transiliensis | 166,520 bp | JLB  | JSB | JSA | JLA |
| Ferula transiliensis | 166,547 bp | JLB  | JSB | JSA | JLA |
| Ferula sibirica | 166,644 bp | JLB  | JSB | JSA | JLA |
| Ferula sibirica | 166,648 bp | JLB  | JSB | JSA | JLA |
| Ferula renardii | 166,520 bp | JLB  | JSB | JSA | JLA |

Fig. 3 Comparison of the border regions of the 14 studied *Ferula* plastomes.
non-coding region and retain as little genetic information as possible to improve its genetic efficiency. Therefore, repeat sequences play an important role in species evolution [61–63]. The repeats found in the 12 analyzed species indicate genetic variation among the Ferula species. In addition, we also observed that the poly (A/T) SSRs were typically most common, while poly (C/G) repeats were extremely rare. These results are consistent with those of a previous study and verify the hypothesis that cpSSRs generally consist of short polyadenine (polyA) or polythymine (polyT) repeats and rarely contain tandem guanine (G) or cytosine (C) repeats [64–66].

Divergent hotspots play a significant role in species identification and phylogenetic information. Moreover, IR regions often show lower sequence divergence than SSC and LSC regions [67], this probably due to higher mutation rates lead to rapid genome evolution compared to other regions [68]. In our study, this phenomenon was evident that the SSC area showed the maximum nucleotide diversity followed by the LSC region, and the IR regions had the lowest Pi value (Fig. 5, Table S5). And rps16/trnQ-UUG, trnS-UGA/psbZ, psbH/petB, ycf1/ndhF, rpl32, ycf1 were detected as the most divergent regions (Pi > 0.006) across all tested plastomes, suggesting that these variable loci can be used as important references and potential molecular markers for future studies on the evolution and diversity in Ferula. Generally, the Ka/Ks ratio is used to divide genes into positive selection, neutral evolution, and purification, with a limit of one [69]. Previously studies indicated that Ka/Ks ratios mostly are lower due to synonymous nucleotide substitutions rates that occur more often compared to nonsynonymous substitutions rates [70]. The genes with the highest Ka/Ks variability can be used as candidate barcodes to differentiate species and in the future applied to perform phylogenetic and phylogeographic
Fig. 5 Sliding window analysis of the newly sequenced chloroplast genomes of *Ferula* species

Fig. 6a Branch length diagram of the phylogenetic tree. b Phylogenetic tree of the 25 species inferred from maximum likelihood (ML) and Bayesian inference (BI) analyses based on the complete plastomes. The Shimodaira-Hasegawa-like support values approximate the likelihood ratio test (only *F. oopoda* and *F. gigantea* had SH-aLRT values below 80 in the terminal branch), and ultrafast bootstrap values (UFBS ≥ 95%, on the right) are shown on the branches. Green indicates two sequences of *S. meyeri* (*F. sibirica*), blue indicates one sequence of *S. karelinii* (*F. karelinii*), and red indicates two sequences of *T. transiliensis* (*F. transiliensis*).
analyses [71]. Our study suggests that 76 common protein-coding genes were under purifying selection, which indicates the typical evolutionary conservation of plant plastid genes [55, 72, 73], and three genes (ccsA, ndhC, ycf2) were under weak positive selection (Table S6), ycf2 have been proved to be pseudogenized in many studies [74] and ccsA was located in one of the most divergent regions, possibly as a discriminating DNA barcode for Ferula species.

The relationships between Soranthus, Schumannia, Talassia and Ferula

Based on the anatomical morphological characteristics of sclerosing cell layers in the mesocarp, the genera Soranthus, Schumannia and Talassia have been proposed to be located under the genus Ferula [25, 75], all of which are recognized in the Flora of China [76]. It is easily distinguished Ferula from Soranthus and Schumannia by gross morphology and inflorescence structure, combined with the presence of luteolin 7-glycosides in the leaves, that seems reasonable to combine the two genera into Soranthus [77]. Also, Talassia tends to be incorporated into Ferula because insignificant morphological differences, although a large extent similarity between T. transiliensis and F. conocaula in the spectrum of leaf flavonoids [77]. Through a comparative study of plant external morphology, fruit anatomy, and pollen morphology, Qin and Shen [27] suggest that Talassia should be an independent genus and agreed to combine the other two monotypic genera. However, the above four genera have been suggested to merge into one genus according to the presence or absence of comumarsins [78]. Recently, the molecular phylogeny of Ferula constructed Kurzyna-Młynik et al. [7] and Panahi et al. [16, 29] based on nrDNA ITS and cpDNA sequences (the rps16 intron, the rpoC1 intron and the rpoB-trnC) indicated that Soranthus, Schumannia, and Talassia were embedded in Ferula with low support values. In our study, 15 sequences (including S. meyeri, S. karelinitii and T. transiliensis) covered all of the branches except the subgenera Ferula (including section Ferula and section Stenocarpa) according to the latest Ferula phylogenetic tree [16]. Our results show that all those three species representing the genera Soranthus, Schumannia and Talassia were embedded in Ferula based on plastidic trees with high bootstrap values (Fig. 6). The species S. meyeri and S. karelinitii were clustered into section Soranthus (PP = 1, BS = 100%), and T. transiliensis and F. renardii clustered into section Glaucocephalum (PP = 1, BS = 100%), which was coincident with Panahi [16] while with higher support values. Therefore, we support the standpoint of sinking Soranthus Ledeb., Schumannia Kuntz., Talassia Korovin into synonymy of Ferula L.

Plastomes might provide new insight on phylogenetic relationships in Ferula

As one of a complex taxonomic genus within Apiaceae, the system of Ferula is paid attention at the morphological and molecular levels [7, 9–11, 13–16]. All those efforts on taxonomic systems have contributed greatly to understanding of the genus Ferula. Kurzyna-Młynik et al. [7] published the first molecular phylogeny for Ferula to solve the relationship among Dorema, Ferula and Leutea, in which nrDNA ITS sequences were used to construct a phylogenetic tree revising Dorema and Leutea to Ferula and transferring Ferula to Scandiceae from Peucedanaceae. Later, nrDNA ITS sequences and three fragments of cpDNA (the rps16 intron, the rpoC1 intron and the rpoB-trnC) were used to explore the relationship among the three genera, and it was found that Dorema was incorporated into Ferula and Leutea independently [16, 29]. Although these results provide an important foundation for the identification and classification of Ferula species, all previous studies have been based on relatively short sequences with low support values owing to the relatively limited number of nuclear/chloroplast genes. In addition, nrDNA and plastid DNA are highly incongruent, and intense reticulate evolution in Ferula means that proposing an unambiguous hierarchical classification system is almost impossible [16]. Furthermore, many species of Ferula have not been specifically addressed, and many only broadly grouped into branches.

Notably, studies based on plastomes can provide new insights into the phylogenetic relationships between species. For example, Clerodendranthus spicatus is closely related to two Lamiales species, Tectona grandis L.f. and Glechoma longituba (Nakai) Kuprian. [79]; Juglandaceae is monophyletic, and Caryya cathayensis Sarg. is a sister to C. kweichowensis Kuang & A.M.Lu and C. illinoiensis (Wangenh.) K.Koch [66]; and Fagus longipetiolata Seemen and F. engleriina Seemen ex Diels form a close relationship [41]. Here, we performed phylogenetic analyses for Ferula and other genera of Apiaceae using complete plastomes, and we recognized Ferula as a monophyletic group with the integration of Soranthus, Schumannia, Talassia (PP = 1, BS = 100%). Within Ferula, we recovered three main lineages in agreement with Panahi et al. [16], who proposed a new classification based on morphological characteristics and sequence data (nrDNA ITS sequences and three cpDNA fragments). This classification divides Ferula into four subgenera and 10 sections. In addition, Caucaulis, Daucus, Cuminum, and Anthriscus were all typical of Scandiceae and formed a monophyletic system with Ferula. This provides strong evidence and support for the transfer of Ferula from the Peucedanaceae to the Scandiceae [7]. However, we also observed some differences. When added into Panahi
et al.’s phylogenetic tree, *F. sinkiangensis* was clustered into the *Scorodosma* branch with the sister species *F. kelifi*. Based on our results, *F. sinkiangensis* is separated from *F. kelifi*, being clustered with *F. litwinowiana* in the *Merwia* branch. Further research is needed to confirm this phenomenon. Overall, our work demonstrates that plastome studies can provide highly useful information for future phylogenetic, taxonomic, and evolutionary studies of *Ferula*.

**Conclusion**

We obtained 14 complete cp genome sequences from 12 *Ferula* species (including *Soranthus*, *Schumannnia*, and *Talassia*) and compared them based on genome structure, gene content, and gene sequences. Some hotspots in the LSC and SSC regions were identified, which may provide useful markers for phylogenetic analysis. Notably, the Gene *ccsA* can be used as a DNA barcode for *Ferula* species. Our phylogenetic analysis showed a tight connection between *Soranthus* Ledeb., *Schumannnia* Kuntz., *Talassia* Korov., and *Ferula* L., indicating that treatment as separate genera is unreasonable. Instead, their phylogenetic relationship, which is now well resolved, strongly supports that they can be considered synonymous with *Ferula*. This new genomic information not only contributes to the better development and utilization of *Ferula* but also provides a basis for further understanding the evolutionary, genetic, and phylogenetic relationships of this important genera.

**Materials and methods**

**Plant materials and DNA extraction**

Fourteen samples were collected from the field and herbaria (Table S1). Of these, five specimens were taken from the specimen museum of the Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences (XJBI), one was obtained from the Komarov Botanical Institute of RAS (LE), five specimens were taken from the National Herbarium of Uzbekistan (TASH), and three were collected from the field in Tajikistan. Leaf samples were dried in silica gel and stored at -20 °C for DNA extraction. DNA extraction was performed using a plant genome extraction kit (DP320) from Tiangen Biochemical Technology (Beijing) according to the manufacturer’s instructions.

**DNA sequencing and genome assembly and annotation**

The extracted DNA was sent to a sequencing company for automatic sequencing using the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs) [80]. DNA extracts were quantified and sheared into approximately 500 base pair (bp) fragments for library construction using standard protocols (NEBNext Ultra IITMDNA Library Prep Kit for Illumina). Paired-end sequencing from both ends of 150 bp fragments was performed on the Illumina HiSeq X Ten platform at the Molecular Biology Experiment Center, Germplasm Bank of Wild Species in Southwest China, to generate no less than 2 GB data for each individual.

The paired-end reads were filtered using the GetOrganelle pipeline (https://github.com/king germ/GetOrganelle) to obtain plastid-like reads [81] and then assembled using SPAdes version 3.10 [82]. A complete circular assembly graph was checked and further extracted using Bandage version 0.8.1 [83]. The genomes were automatically annotated using CpGAVAS [84], PGA (https://github.com/quxiaojian/PGA), and then manually adjusted using Geneious version 9.1.7 [85]. The chloroplast sequences generated in this study have been submitted to GenBank (Table S1). Circular genome maps of all 14 plastomes were also obtained using the Organellar Genome DRAW (OGDRAW) tool [86].

**Codons, repeat sequences, and simple sequences repeat analysis**

The protein-coding genes were extracted for codon analysis. The final dataset included 86 protein-coding genes from each species. Codon usage and relative synonymous codon usage (RSCU) values were calculated using JSHY-Cloud (http://cloud.genepioneer.com:9929). A heatmap of all the RSCU values of the 14 plastomes was produced using ClustVis [87]. Using the parameters of a Hamming distance of 3, a minimum repeat size of 30 bp, and a maximum repeat size of 5,000 bp, REPuter was used to identify the size and location of four types of repeat sequences (i.e., forward, palindromic, reverse, and complement) [88]. Simple sequence repeats (SSRs) were detected using the online MISA software (http://pgrc.ipk.gatersleben.de/misa/misa.html) with minimum repeat number settings of 10, 5, 4, 3, 3, and 3 for mononucleotides, dinucleotides, trinucleotides, tetrannucleotides, pentanucleotides, and hexanucleotides, respectively.

**Genome comparison with other *Ferula* species and selective pressure analysis**

Sequence divergence among the 14 chloroplast (cp) genomes was compared using Mafft (version 7.0) [89], IRScope (https://irscope.shinyapps.io/irapp/) and Mauve [90]. DnaSP [91] was used to calculate nucleotide divergence values using the sliding window method, with a window length of 800 bp and a step size of 200 bp. Selective pressures were analyzed for 79 common protein-coding genes among 15 *Ferula* species (including one published plastome). The ratio of nonsynonymous to
synonymous nucleotide substitution rates (Ka/Ks) was calculated using DnaSP.

Phylogenetic analysis
We used 25 complete plastome sequences to infer the phylogenetic relationships of Ferula. After comparison with Mafft, Trimal [92], and Phylosuite [93] were used to trim areas with poor quality. The phylogenetic tree was then constructed using RaxML-HPC v.8 [94] and the maximum likelihood method with 1,000 replicates and the GTR+gamma model. After screening for the best model using jModelTest2 [95], MrBayes 3.2.7a [96] was used to construct a Bayes tree, and the selected models for the complete plastome sequences in BI analyses were TPM1uf+1+G, and iTOl [97] and FigTree 1.4.2 [98] were used to construct the phylogenetic tree.

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1186/s12864-022-08868-z.

Additional file 1: Table S1. All the materials used in this article.
Additional file 2: Table S2. Codon usage and relative synonymous codon usage (RSCU) values of protein-coding genes of the 14 Ferula plastomes.
Additional file 3: Table S3. Distribution of repeat sequences in the 14 studied Ferula plastomes.
Additional file 4: Table S4. Distribution of simple sequence repeats (SSRs) in the 14 studied Ferula plastomes.
Additional file 5: Table S5. Nucleotide variability (Pi) of Ferula species.
Additional file 6: Table S6. Non-synonymous to synonymous nucleotide substitution rates (Ka/Ks) of Ferula species.

Acknowledgements
We thank Yu Wenbin, Liu Rong, Xin Yaxuan and Li Jinyue in Xishuangbanna Tropical Botanical Garden of Chinese Academy of Sciences for their assistance to assemble and annotation sequences. Our sincere thanks also go to the anonymous reviewers for their comments and suggestions.

Statement
The materials used in this article are not related to plant protection and are within the limits of national laws. And we have permission from The Komarov Tropical Botanical Garden of Chinese Academy of Sciences for their assistance to enter and collect plant specimens. This research was carried out in compliance with the relevant laws and all methods were performed in accordance with the relevant guidelines and regulations.

Competing interests
The authors declare that they have no competing interests.

Author details
1 State Key Laboratory of Desert and Oasis Ecology, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, No.818 South Beijing Road, Urumqi 830011, China. 2 Xinjiang Key Lab of Conservation and Utilization of Plant Gene Resources, No.818 South Beijing Road, Urumqi 830011, China. 3 College of Resources and Environment, University of Chinese Academy of Sciences, Shiingshan District, No.19(A) Yuyuan Road, Beijing 100049, China. 4 Institute of Botany, Uzbekistan Academy of Sciences, No.32 Durmon Yuli Street, Tashkent, Uzbekistan 100125. 5 Sino-Tajikistan Joint Laboratory for Conservation and Utilization of Biological Resources, No.818 South Beijing Road, Urumqi 830011, China. 6 The Specimen Museum of Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi 830011, China.

Received: 30 May 2022   Accepted: 30 August 2022

References
1. Pimenov MG, Leonov MV. The genera of the Umbelliferae: a nomenclator. Kew: Royal Botanic Gardens; 1993.
2. Sina AA. Kanon vrachebnoy nauki (Canon). In: Zaxidov TZ, editor. Book. 1. Tashkent: Academy of Sciences of the UzSSR; 1954. p. 549.
3. Shen GM. Chinese herbal medicine series: Ferula. Urumqi: people's publishing house; 1986.
4. Nazari ZE, Iranshahi M. Biologically active sesquiterpene coumarins from Ferula species. Phytother Res. 2011;25(3):315–23.
5. Ajani Y, Ajani A, Cordes JM, Watson MF, Downie SR. Phylogenetic analysis of nrDNA ITS sequences reveals relationships within five groups of Iranian Apioideae subfamily Apioideae. Taxon. 2008;57(2):383–401.
6. Ajani V, Ajani A, Cordes JM, Watson MF, Downie SR. Phylogenetic analysis of nrDNA ITS sequences reveals relationships within five groups of Iranian Apioideae subfamily Apioideae. Taxon. 2008;57(2):383–401.
7. Kurzyna-Młynik R, Oskolski AA, Downie SR, Kopacz R, Wojewódzka A, Spalik K. Phylogenetic position of the genus Ferula (Apioideae) and its placement in tribe Scandiceae as inferred from nrDNA ITS sequence variation. Plant Syst Evol. 2008;274(1–2):47–66.
8. Safina LK, Ostroumova TA, Pimenov MG. Carpolls of the species of Ferula subgen. Mervai(Umbelliferae-Apioideae) and some taxonomic implications. Nord J Bot. 2015;33(2):141–6.
9. Boisser PE. Flora orientalis sive, enumeratio plantarum in Oriente a Graecia et Aegypto ad Indiae fines hucusque observatae 2. Genève, Basel & Lyon: H.Georg; 1872. p. 1159.
10. Drude CGO. Umbelliferae. In: Engler A, Prantl K, editors. Die natürlichen Pflanzenfamilien, vol. 3. Leipzig: Druck von Wilhelm Engelmann; 1898. p. 63–250.
11. Korovin EP. Generis Ferula (Tourn.). L. monographia illustrata. Tashkent: Academiae Scientiarum UzRSS, 1947. p. 91.
