




Article

Comparative Analysis and Characterization of Ten Complete Chloroplast Genomes of *Eremurus* Species (Asphodelaceae)

Dilmurod Makhmudjanov ^{1,2,3,4} , Davlatali Abdullaev ⁵, Inom Juramurodov ^{1,2,3,4}, Shakhzodbek Tuychiev ³, Ziyoviddin Yusupov ⁵, Hang Sun ^{1,2} , Komiljon Tojibaev ^{3,*} and Tao Deng ^{1,2,*} 

- ¹ CAS Key Laboratory for Plant Diversity and Biogeography of East Asia, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, China; dilmurodxab@gmail.com (D.M.); ijuramurodov@mail.ru (I.J.); sunhang@mail.kib.ac.cn (H.S.)
- ² Yunnan International Joint Laboratory for Biodiversity of Central Asia, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, China
- ³ Flora of Uzbekistan Laboratory, Institute of Botany of the Academy of Sciences of the Republic of Uzbekistan, 32 Durmon Yuli St., Tashkent 100125, Uzbekistan; shakhzodbektuychiev@gmail.com
- ⁴ University of Chinese Academy of Sciences, Beijing 100049, China
- ⁵ International Joint Lab for Molecular Phylogeny and Biogeography, Institute of Botany, Academy Sciences of Uzbekistan, Tashkent 100125, Uzbekistan; davlatali.1991@mail.ru (D.A.); ziyo-nur87@mail.ru (Z.Y.)
- * Correspondence: ktojibaev@mail.ru (K.T.); dengtao@mail.kib.ac.cn (T.D.)

Abstract: *Eremurus*, a perennial rhizomatous mesophytic ornamental plant and one of the largest genera of the family Asphodelaceae, is distributed mainly in southwestern and central Asia. We sequenced the complete chloroplast genomes of ten species corresponding to all sections of the genus and analyzed their basic structure and evolutionary relationships. The cp genomes showed significant similarities in size, gene sequences, gene classes, and inverted repeat regions (IRs). The complete chloroplast genome of *Eremurus* has a typical tetrad structure, ranging in length from 153,782 bp (*E. lactiflorus*) to 155,482 bp (*E. aitchisonii*). The length of the large single-copy region (LSC) ranges from 84,005 bp (*E. lactiflorus*) to 84,711 bp (*E. robustus*), that of the small single-copy region (SSC) ranges from 16,727 bp (*E. soogdianus*) to 17,824 bp (*E. suworowii*), and that of the inverted repeat regions (IR) ranges from 26,484 bp (*E. lactiflorus*) to 26,597 bp (*E. inderiensis* and *E. soogdianus*). A total of 131 genes were detected, including 85 protein-coding genes, 8 rRNA genes, and 38 tRNA genes. In addition, we found seven common and eight unique SSRs in ten *Eremurus* species. Among the protein-coding genes, five highly variable genes (*ycf1*, *rps15*, *rps16*, and *rpl36*) with high Pi values were detected and showed potential as DNA barcodes for the genus. Three genes (*rps19*, *ycf1*, and *ndhB*) had positive Ka/Ks values. Codon usage patterns were very similar across species: 33 codons had relative synonymous codon usage values of more than one, of which three ended with G, and the remaining codons ended with A and U. Phylogenetic analyses using complete cp genomes and 81 protein-coding genes confirmed previous studies with the genus as well as subgenus *Eremurus* monophyletic and the subgenus *Henningia* paraphyletic.

Keywords: *Eremurus*; chloroplast genome; codon usage; comparative analysis; phylogeny; protein-coding genes



Citation: Makhmudjanov, D.; Abdullaev, D.; Juramurodov, I.; Tuychiev, S.; Yusupov, Z.; Sun, H.; Tojibaev, K.; Deng, T. Comparative Analysis and Characterization of Ten Complete Chloroplast Genomes of *Eremurus* Species (Asphodelaceae). *Forests* **2023**, *14*, 1709. <https://doi.org/10.3390/f14091709>

Academic Editor: Jianquan Liu

Received: 14 July 2023

Revised: 13 August 2023

Accepted: 15 August 2023

Published: 24 August 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Eremurus M.Bieb. is one of the largest genera of Asphodelaceae Juss. [1], with 45 to 50 [2] or 59 [3] species of perennial, rhizomatous mesophytic plants. The genus is distributed in loess slopes and arid to semiarid mountainous areas in Central Asia, China, India, Pakistan, Afghanistan, Iran, Iraq, Lebanon, the Caucasus area, and Turkey [4,5]. Species of *Eremurus* are important as ornamental plants [6] and are called “foxtail lily” or “desert candle” because of their large and colorful inflorescence spikes [1]. They are also used in industry for products such as bio-oil [7] and adhesives [8], and some species within

this genus are used as potential sources of drugs with antibacterial, anti-inflammatory, and antiprotozoal properties and have been traditionally utilized in medicine [9,10].

Eremurus differs from other closely related genera due to its leafless inflorescence with over 50 flowers and its rhizomatous rootstock [11]. The genus comprises two subgenera: *Eremurus* (1) is characterized by light brownish green or cream tubular or campanulate flowers with incurved tepals bearing three or five nerves on the underside and exerted filaments; on the other hand, species of the subgenus *Henningia* (2) have white, pink, or yellow rotate flowers, mostly with included filaments and tepals exhibiting one nerve on the underside [12]. While two cladistic phylogenetic studies based on the aforementioned key morphological characters [13,14] have been conducted, the genus has been subject to limited molecular studies [15–17]. Previous molecular studies using plastid (*trnL-F*) and nuclear ribosomal DNA (ITS) sequences have shown that the genus is monophyletic, but at the subgenus level, subgenus *Eremurus* Baker is monophyletic, whereas subgenus *Henningia* (Kar. & Kir.) Baker is paraphyletic based on *trnL-F* sequence data [17]. To enhance the phylogenetic relationships in *Eremurus*, complete cp genomes could be utilized. However, at present, only one species of *Eremurus* has had its cp genome published [18]. Comparative cp genomics can be used to identify important structural sequences and reveal evolutionary changes between genomes.

Our study provides structural and phylogenetic analyses of complete cp genome sequences of ten *Eremurus* species: *E. indieriensis* (M.Bieb.) Regel, *E. hissaricus* Vved., *E. iae* Vved., *E. regelii* Vved., *E. soogdianus* (Regel) Benth. and Hook.f., *E. aitchisonii* Baker, *E. albertii* Regel, *E. lactiflorus* O.Fedtsch., *E. luteus* Baker, and *E. suworowii* Regel. Our specific objectives were as follows: (1) to compare the chloroplast structures within *Eremurus*; (2) to identify potential DNA barcoding markers to identify *Eremurus* species by recognizing the regions of high variability; (3) to infer the phylogenetic relationships among *Eremurus* species.

2. Materials and Methods

2.1. Plant Materials

Species from each section of *Eremurus* were selected for comparative genomic analyses: *Eremurus indieriensis* (sect. *Ammolirion* (Kar & Kir) Boiss.), *E. hissaricus*, *E. iae*, *E. regelii*, *E. soogdianus* (sect. *Eremurus* (Baker) Wendelbo) from the *E. subg. Eremurus*, and *E. aitchisonii*, *E. albertii*, *E. lactiflorus*, *E. luteus*, *E. robustus*, *E. suworowii* from the *E. subg. Henningia*. All fresh materials were collected in Uzbekistan (Table S1), and their complete chloroplast (cp) genome sequences were generated (Figure 1). Herbarium specimens are stored in the National Herbarium of Uzbekistan (TASH) and Kunming Institute of Botany, Chinese Academy of Sciences (KUN).

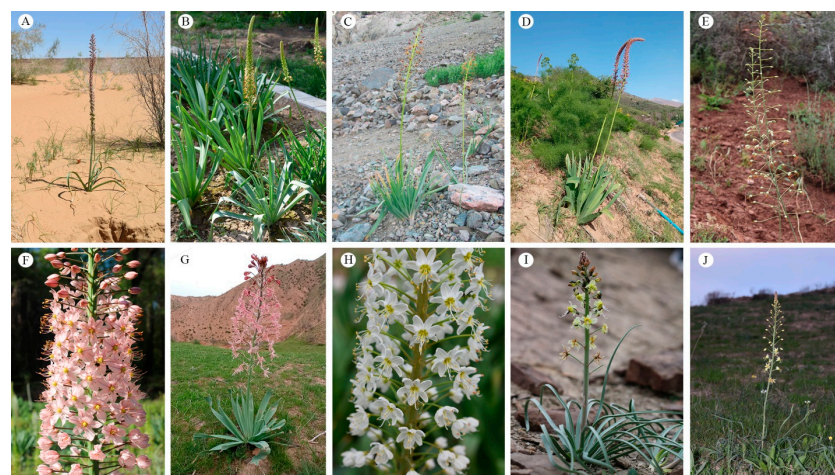


Figure 1. Species sequenced in this study. (A) *E. indieriensis*, (B) *E. hissaricus*, (C) *E. iae*, (D) *E. regelii*, (E) *E. soogdianus*, (F) *E. aitchisonii*, (G) *E. albertii*, (H) *E. lactiflorus*, (I) *E. luteus*, (J) *E. suworowii*. The photo of *E. suworowii* was taken by S. Pulatov.

2.2. Sequencing, Assembly, and Annotation

The DP305 Plant Genomic DNA Kit (Tiangen, Beijing, China) was used to extract total genomic DNA from leaf material, following the manufacturer's protocol. The sequencing library was generated using the NEBNext[®] UltraTM DNA Library Prep Kit for Illumina (New England, USA, NEB, Catalog: E7370L) following the manufacturer's recommendations, and index codes were assigned to each sample. Briefly, the genomic DNA sample was fragmented to a size of 350 bp by sonication. Then, the DNA fragments were polished at the ends, A-tailed, and ligated to the full-length adapter for Illumina sequencing, followed by further PCR amplification. The PCR products were purified using the AMPure XP system (Beverly, MA, USA). Subsequently, the quality of the library was checked by Agilent 5400 System (Agilent, Santa Clara, CA, USA) and quantified by QPCR (1.5 nM). The qualified libraries were pooled and sequenced on Illumina platforms using the PE150 strategy from Novogene Bioinformatics Technology Co., Ltd. (Beijing, China), depending on the effective library concentration and the amount of data required.

The resulting clean reads were assembled using the GetOrganelle pipeline [19] with the optimized parameters “-F plant_cp -w 0.6 -o -R 20 -t 8 -k 75,95,115,127 &”. Gene annotation was performed in Geneious v.10.0.2, and *E. robustus* (accession number: NC046772) was set as the reference. Start and stop codons and intron/exon boundaries for protein-coding genes were manually checked [20].

2.3. Simple Sequence Repeats (SSRs)

The MicroSatellite (MISA) web tool was used for chloroplast simple sequence repeat (SSR) identification [21]. The search parameters for SSRs were configured to identify ideal mono-, di-, tri-, tetra-, penta-, and hexa-nucleotide patterns with at least 10, 5, 4, 3, 3, and 3 repeats, respectively. The REPuter program [22] was used to identify repeats: forward, reverse, palindrome, and complement sequences in cp genomes. The following settings were used to identify repeats: (1) hamming distance equal to 3; (2) minimal repeat size set to 30 bp; and (3) maximum calculated repeats set to 90 bp.

2.4. Comparative Analysis of Chloroplast Genomes

Physical maps of cp genomes were generated using OGDRAWv1.1 [23]. The program mVISTA in Shufu- LAGAN mode [24] was used to compare the complete cp genomes of the 10 *Eremurus* species, using the annotation of *E. robustus* as a reference (NC046772). After manual multiple alignments using the program MUSCLE [25] in the software MEGA X [26], coding regions were extracted to detect variable sites. The nucleotide variability (Pi) was calculated for the whole cp genome and protein-coding genes separately using DnaSP v. 6 software [27]. The window length was set to 800 bp and the step size to 200 bp. To determine whether protein-coding genes were under selection pressure, the synonymous (Ks) and nonsynonymous (Ka) substitution rates and ω -value ($\omega = K_a/K_s$) for shared protein-coding genes in ten *Eremurus* cp genomes were analyzed using DnaSP v. 6 software [27].

2.5. Codon Usage Bias Analysis

Coding sequences (CDS) found in chloroplast genomes were extracted manually one by one. Codon usage frequency analysis was performed for each species using the MEGA X [26]. The relative synonymous codon usage (RSCU) indicates whether a plastid gene is being favored, and codons with an RSCU value greater than 1 were considered high-frequency codons.

2.6. Phylogenetic Analysis

A total of 17, of which 11 cp were genomes of *Eremurus* and 6 cp were genomes of outgroups (*Aloe vera*, *A. maculata*, *Alloidendron pillansii*, *Xanthorrhoea preissii*, *Hemerocallis fulva* from Asphodelaceae and *Asparagus officinalis* from Asparagaceae), were used for phylogenetic analysis. Table 1 provides information about their NCBI accession numbers. Phylogenetic tree reconstruction was performed using the complete cp genomes and

protein-coding sequences, which were first aligned multiple times using MAFFT software v. 7 [28].

Table 1. Summary of chloroplast genome characteristics of 11 *Eremurus* and 6 outgroup species used for polygenetic analysis. New sequences are marked with an asterisk.

Species	Total Length (bp)	GC (%)	LSC Length (bp)	SSC Length (bp)	IR Length (bp)	Gene Number	Protein Coding Genes	tRNAs	rRNAs	GenBank Accession Numbers
<i>Eremurus inderiensis</i>	154,320	37.4	84,393	16,733	26,597	131	85	38	8	OL852091 *
<i>Eremurus hissaricus</i>	154,404	37.4	84,677	16,757	26,485	131	85	38	8	OL875065 *
<i>Eremurus iae</i>	154,808	37.4	84,578	17,046	26,592	131	85	38	8	OL875066 *
<i>Eremurus regelii</i>	154,443	37.4	84,353	16,900	26,590	131	85	38	8	OL875068 *
<i>Eremurus soogdianus</i>	154,311	37.4	84,390	16,727	26,597	131	85	38	8	OL875071 *
<i>Eremurus aitchisonii</i>	155,482	37.3	84,536	17,794	26,576	131	85	38	8	OL852090 *
<i>Eremurus albertii</i>	154,129	37.4	84,217	16,754	26,579	131	85	38	8	OL852089 *
<i>Eremurus lactiflorus</i>	153,782	37.4	84,005	16,809	26,484	131	85	38	8	OL875070 *
<i>Eremurus luteus</i>	155,439	37.3	84,510	17,781	26,574	131	85	38	8	OL852094 *
<i>Eremurus robustus</i>	155,647	37.3	84,711	17,786	26,575	131	85	38	8	NC046772
<i>Eremurus suworowii</i>	155,400	37.3	84,426	17,824	26,575	131	85	38	8	OL875060 *
<i>Aloe vera</i>	152,875	37.3	83,504	16,177	26,597	131	85	38	8	NC035506
<i>Aloe maculata</i>	153,175	37.6	83,567	15,892	26,858	131	85	38	8	NC035505
<i>Aloidendron pillansii</i>	154,094	37.6	84,002	16,952	26,570	131	84	38	8	NC044761
<i>Xanthorrhoea preissii</i>	158,116	37.9	86,080	18,256	26,850	132	86	38	8	NC035996
<i>Hemerocallis fulva</i>	155,855	37.4	84,607	18,508	26,370	133	87	38	8	NC041649
<i>Asparagus officinalis</i>	156,699	37.6	84,999	18,638	26,531	129	83	38	8	NC034777

Maximum likelihood (ML), Bayesian inference (BI), and maximum parsimony (MP) methods were used in this study to reconstruct phylogenetic trees. Nucleotide substitution models were statistically selected using jModelTest2 on XSEDE (www.phylo.org, accessed on 2 January 2020) using the Akaike information criterion (AIC). The GTR+I+G and TIM1+I+G models were selected as the best models for the protein-coding sequences and the complete cp genomes, respectively. For BI, we used MrBayes v. 3.2.7a [29] with 10 million generations, randomly sampling the trees every 1000 generations. In the latter analysis, after the first 25% of the trees were discarded as burn-in, a consensus tree with 50% majority rule was constructed from the remaining trees to estimate posterior probabilities (PP). ML trees were constructed with 1000 replicates for bootstrapping using RAXML v8.2.11 [30] via raxmlGUI 2.0.10 platform [31]. For MP analysis, we used PAUP* 4.0a169 [32]. The MP bootstrap analysis was performed with heuristic search, TBR branch-swapping, 1000 bootstrap replicates, random addition sequence with 10 replicates, and a maximum of 1000 saved trees per round.

3. Results

3.1. Chloroplast Genome Features of *Eremurus* Species

The complete cp genomes of *E. inderiensis*, *E. hissaricus*, *E. iae*, *E. regelii*, *E. soogdianus*, *E. aitchisonii*, *E. albertii*, *E. lactiflorus*, *E. luteus*, and *E. suworowii* were sequenced for this study. Their lengths ranged from 153,782 to 155,482 bp (Table 1). The cp genome of *E. robustus* was obtained from NCBI as a reference. All of the newly sequenced genomes exhibited the typical quadripartite structure of angiosperm chloroplasts, containing two pairs of IRs (26,484–26,597 bp) separated by LSC (large single copy, 84,005–84,711 bp) and SSC (small single copy, 16,727–17,824 bp) regions (Figure 2, Table 1). The GC (guanine+cytosine) content of the genomes of eleven species ranged from 37.3% to 37.4%. All genomes consisted of 131 genes, including 85 protein-coding genes, 8 rRNA genes, and 38 tRNA genes (Table 1).

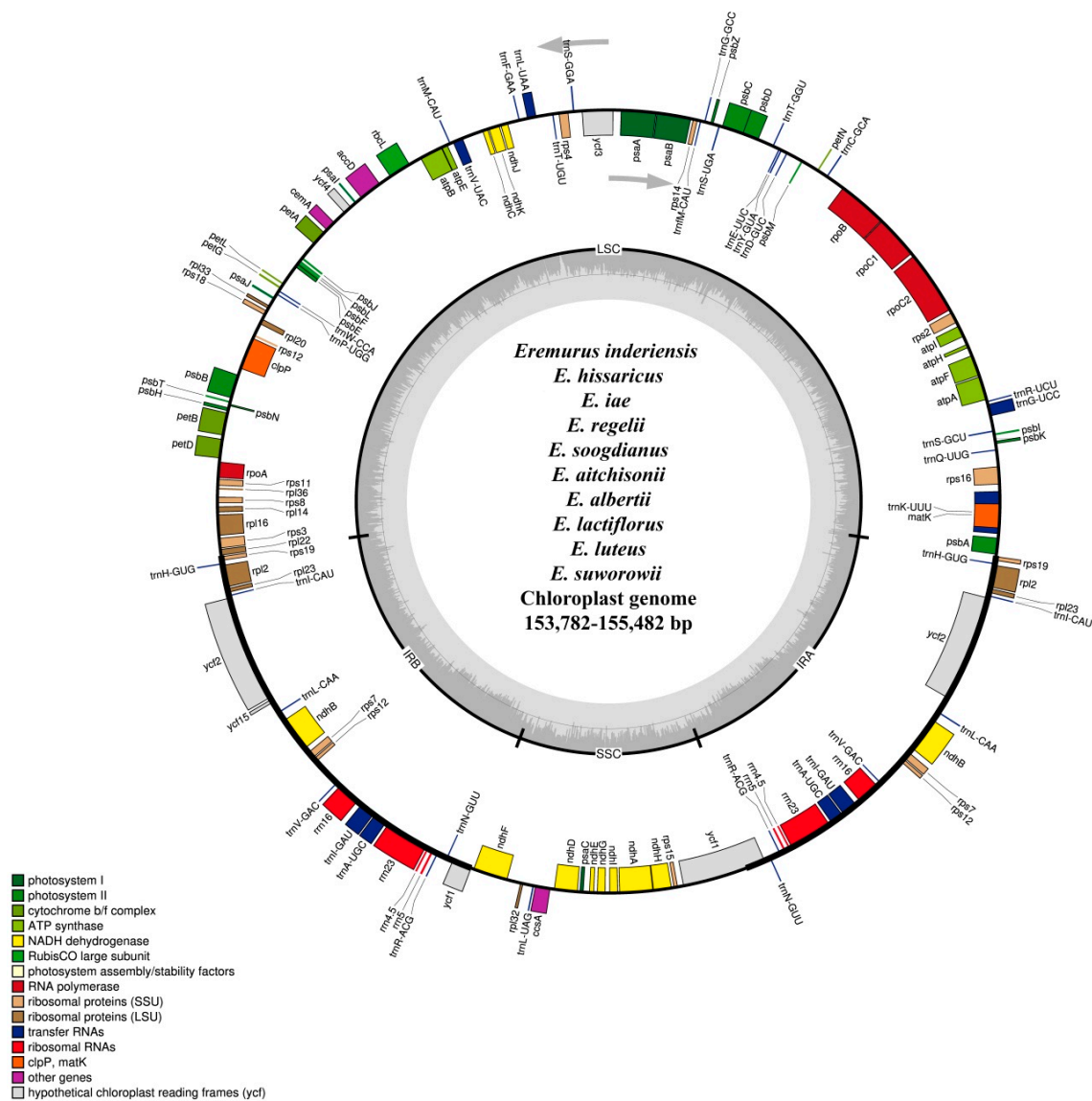


Figure 2. Chloroplast genome structure of ten *Eremurus* species. Genes shown outside the circles are transcribed clockwise, while those drawn inside are transcribed counterclockwise. Genes are color-coded according to their functional group.

Among the genes unique to *Eremurus*, 44 were related to photosynthesis, and 59 were related to self-replication (Table 2). A total of 18 introns occurred in the cp genomes of all *Eremurus* species in the genes *trnK-UUU*, *rps16*, *trnG-UCC*, *atpF*, *rpoC1*, *trnL-UAA*, *trnV-UAC*, *rps12*, *petB*, *petD*, *rpl16*, *rpl2*, *ndhB*, *trnI-GAU*, *trnA-UGC*, and *ndhA*; the genes *ycf3* and *clpP* each contained two introns (Tables 2 and S2). The *rps12* gene contained the largest intron, from 69,488 (*E. hissaricus*) to 70,651 (*E. suworowii*) bp.

Table 2. Genes in the CP genomes of eleven *Eremurus* species. ×2 indicates two gene copies. * and ** indicate genes that contain 1 and 2 introns, respectively. Ψ indicates a pseudogene.

Category of Genes	Group of Genes	Genes	Number of Genes
Genes for photosynthesis (44)	Subunits of photosystem I	<i>psaA, psaB, psaC, psaI, psaJ</i>	5
	Subunits of photosystem II	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ,</i>	15
	Subunits of ATP synthase	<i>atpA, atpB, atpE, atpF *, atpH, atpI</i>	6
	Subunits of NADH-dehydrogenase	<i>ndhA *, ndhB * (×2), ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i>	12
	Subunits of cytochrome b/f complex	<i>petA, petB *, petD *, petG, petL, petN</i>	6
	RubisCO large subunit	<i>rbcL</i>	1
Self-replication (59)	Large subunit of ribosome	<i>rpl2 * (×2), rpl14, rpl16 *, rpl20, rpl22, rpl23 * (×2) rpl32, rpl33, rpl36</i>	11
	Small subunit of ribosome	<i>rps2, rps3, rps4, rps7 (×2), rps8, rps11, rps12 * (×2), rps14, rps15, rps16 *, rps18, rps19 (×2)</i>	15
	RNA polymerase	<i>rpoA, rpoB, rpoC1 *, rpoC2</i>	4
	Ribosomal RNAs	<i>rrn4.5 (×2), rrn5 (×2), rrn16 (×2), rrn23 (×2)</i>	8
	tRNA genes	<i>trnA-UGC * (×2), trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnFM-CAU, trnG-GCC, trnG-UCC, trnH-GUG (×2), trnI-CAU (×2), trnI-GAU * (×2), trnK-UUU *, trnL-CAA (×2), trnL-UAA *, trnL-UAG, trnM-CAU, trnN-GUU (×2), trnP-UGG, trnQ-UUG, trnR-UCU (×2), trnR-ACG (×2), trnS-GGA, trnS-GCU, trnS-UGA, trnT-GGU, trnT-UGU, trnV-GAC (×2), trnV-UAC *, trnW-CCA, trnY-GUA</i>	39
Other genes (5)	Subunit of acetyl-CoA-carboxylase	<i>accD</i>	1
	c-type cytochrome synthesis gene	<i>ccsA</i>	1
	Envelop membrane protein	<i>cemA</i>	1
	Protease	<i>clpP **</i>	1
	Maturase	<i>matK</i>	1
Genes with unknown function (5)	hypothetical chloroplast reading frames (ycf)	<i>ycf1 (×2), ycf2 (×2), ycf3 **, ycf4, ycf15 Ψ</i>	7

3.2. Repeat Sequences and SSRs Analysis

A total of 82, 90, 86, 84, 84, 93, 79, 79, 91, and 87 SSRs were detected in the CP genomes of *E. inderiensis*, *E. hissaricus*, *E. iae*, *E. regelii*, *E. sogdianus*, *E. aitchisonii*, *E. albertii*, *E. lactiflorus*, *E. luteus*, and *E. suworowii*, respectively, with 64, 66, 65, 63, 66, 73, 60, 59, 69, and 67 mononucleotide SSRs; 10, 12, 11, 11, 10, 11, 10, 10, 12, and 11 dinucleotide SSRs; except *E. albertii* (2), nine species had 1 trinucleotide SSRs; 3, 3, 3, 3, 3, 5, 3, 5, 6, and 6 tetranucleotide SSRs; 3, 4, 3, 3, 3, 0, 2, 4, 1, and 1 pentanucleotide SSRs; and 1, 4, 3, 3, 1, 3, 2, 0, 2, and 1 hexanucleotide SSRs, respectively. Among the ten *Eremurus* cp genomes, the most abundant repeats were the mononucleotides from 59 (*E. lactiflorus*) to 79 (*E. aitchisonii*), and the most dominant SSR was A. The second most predominant SSR was the dinucleotides, especially AT, varying from seven to nine. AG were three, and trinucleotides were one in each species. A total of 40 repeats of tetranucleotides, varying from 3 (*E. inderiensis*, *E. hissaricus*, *E. iae*, *E. regelii*, *E. sogdianus*, and *E. albertii*) to 6 (*E. luteus* and *E. suworowii*), were identified among the ten *Eremurus* cp genomes. Our analysis revealed that only *E. aitchisonii* had no pentanucleotide repeats. Eight pentanucleotide repeats in the other nine *Eremurus* species varied from one (*E. luteus* and *E. suworowii*) to four (*E. hissaricus* and *E. lactiflorus*). As well, *E. lactiflorus* did not exhibit any pentanucleotide repeats, whereas the other species varied from one to four (Figure 3B,C).

Our study examined both common and unique SSRs in ten *Eremurus* species, which are listed in Tables S3 and S4. Our results showed that the majority of repeat units consisted of A and T, with rare occurrences of C or G, suggesting that the SSRs of different species have a clear preference for certain base types of repeat units. The common SSRs present in all ten species included A, AG, AT, AAT, AAAC, AAAT, and AATG. In addition, we found eight unique SSRs, including ATCC, AAATT, and ATATC in *E. lactiflorus*; AAAT

and AAATTC in *E. hissaricus*; and AATT and AAAAAT in *E. suworowii*. In addition, a single AAATTG SSR was detected in *E. aitchisonii*, while no unique SSRs were identified in *E. indierensis*, *E. iae*, *E. regelii*, *E. soogdianus*, *E. albertii*, and *E. luteus*.

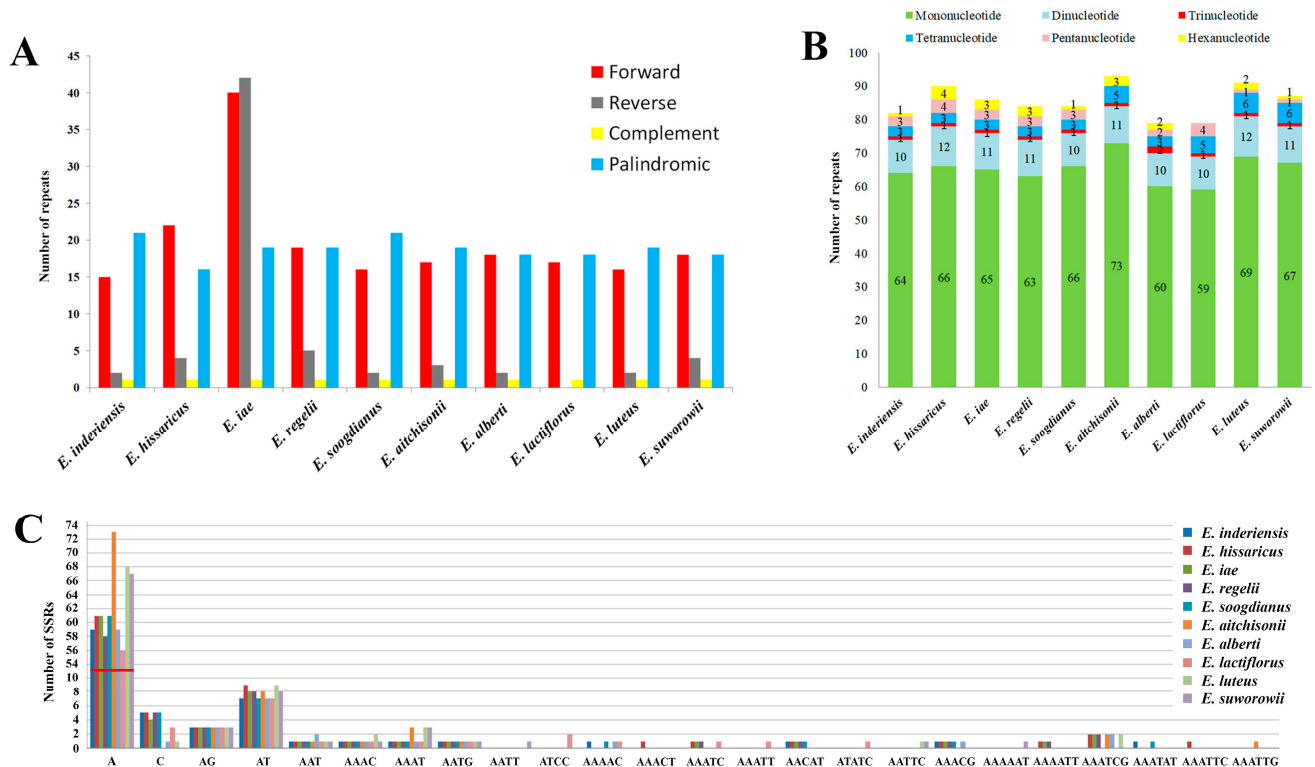


Figure 3. Chloroplast genome features of ten *Eremurus* species. SSR distribution (A); long repetitive sequences (B); type of SSRs (C).

In this study, we found many repeat regions, including forward, reverse, palindromic, and complementary repeats (Figure 3A). Among the ten studied *Eremurus* species, the longest repetitive sequences were detected in the cp genome of *E. iae*, which had 102 repetitive sequences with a length of 29 bp or less. In contrast, the smallest repetitive sequences were found in the cp genome of *E. lactiflorus*, which had 36 scattered repetitive sequences with a length not exceeding 12 bp. The length of the largest forward and reverse repeats was 40 bp and 42 bp, respectively, in the *E. iae* cp genome, while the largest palindromic repeats were 21 bp in the *E. nderiensis* and *E. soogdianus* cp genomes. Equal numbers of complement repeats were detected in all ten species. In addition, the reverse repeat was not found in the cp genome of *E. lactiflorus*.

3.3. Comparative Genomic Divergence and Hotspot Regions

We calculated nucleotide diversity (Pi) to estimate levels of interspecific sequence divergence across the genome (Figure 4A,B). The highest variations (Pi > 0.01) were mainly concentrated in the SSC regions, between 125,000 bp and 135,000 bp. Across protein-coding genes, *ycf1* (0.00961), *rps15* (0.00626), *rpl36* (0.00585), and *rps16* (0.00569) had the highest variability (Pi > 0.0055), and the *rpl2* gene had the lowest (0.00026). Values of Pi were less than 0.001 in 39.29% of the protein-coding genes and were 0.001–0.002 in 27.05%. Only 37.55% of protein-coding genes had Pi > 0.002 (Table S5).

The cp genome sequences of 11 *Eremurus* species were compared using mVISTA software, and their alignments were visualized with annotation data (Figure 5). After this visualization analysis, differences occurred between the sequences in *accD*, *AtpF*, *ndhA*, *ndhB*, *ycf1*, and *ycf2* genes from coding regions and mainly in noncoding intergenic regions.

The encoded gene classes and the alignments of most coding regions of the ten *Eremurus* species were highly congruent.

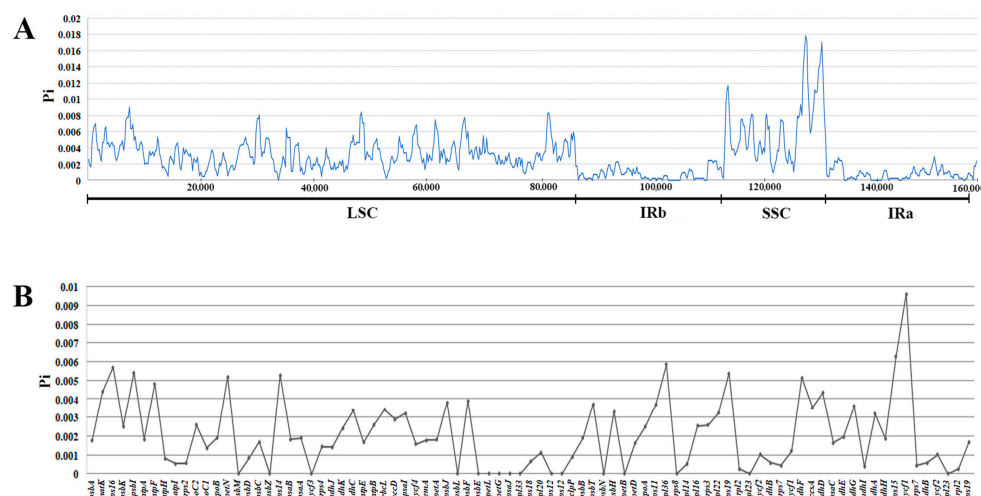


Figure 4. Nucleotide diversity (Pi) in whole chloroplast genomes (A); and in protein-coding genes (B) calculated using 85 shared genes in ten *Eremurus* species.

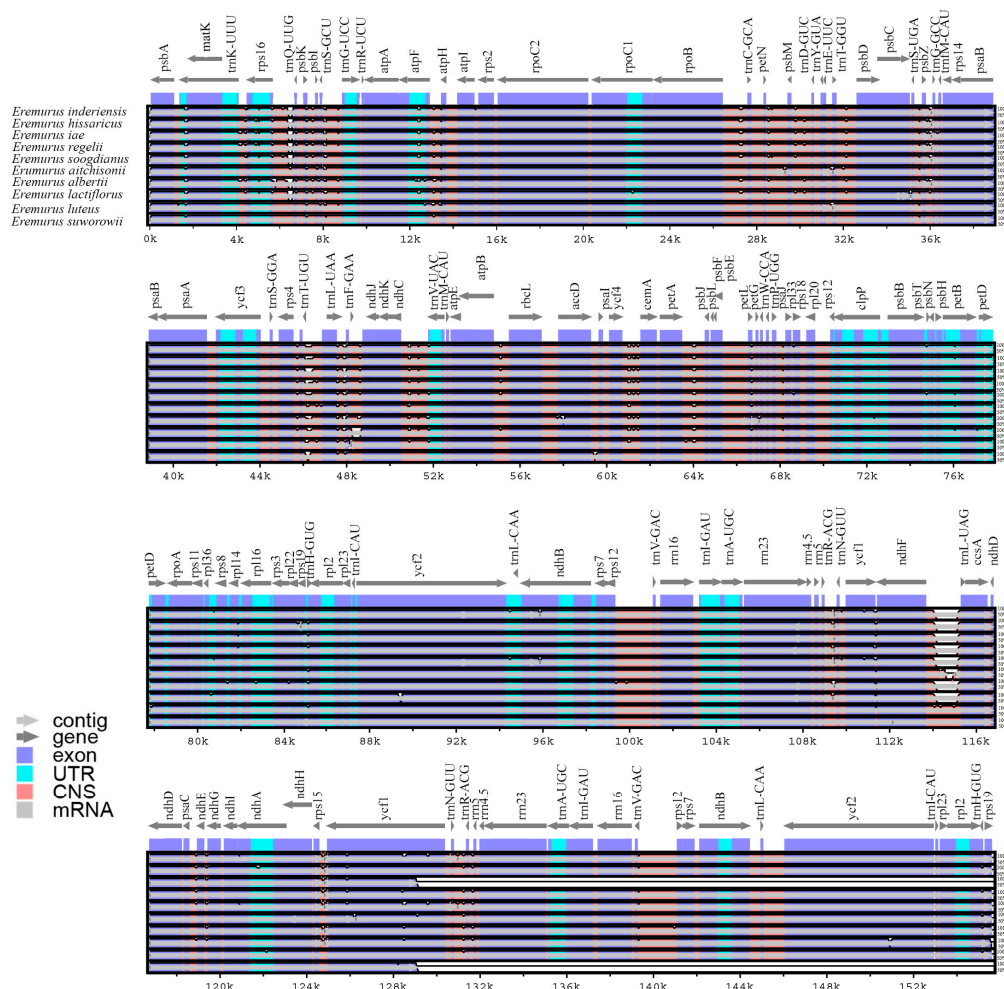


Figure 5. Comparison of the ten *Eremurus* chloroplast genome sequences using mVISTA. The genes are represented above. Genome regions are color-coded in the legend. The range of sequence similarity is presented in percentage (%).

3.4. Functional Gene Selection

The synonymous (Ks) and nonsynonymous (Ka) substitution rates of the ten species in *Eremurus* ranged from 0.0000 to 0.025 (*rpl36*) and from 0.0000 to 0.0099 (*ycf1*), respectively. Among the 85 total protein-coding genes, the Ka/Ks values of *rps19* (1.423), *ycf1* (1.128), and *ndhB* (1.074) were under positive selection ($\omega > 1$) (Figure 6 and Table S5). In addition, the ω -values ranged from less than 1 to 0.1 for 34.18% of the total protein-coding genes, whereas they were 0.095–0.01 for 10.59%. The remaining total protein-coding genes had no ω -values (Table S5).

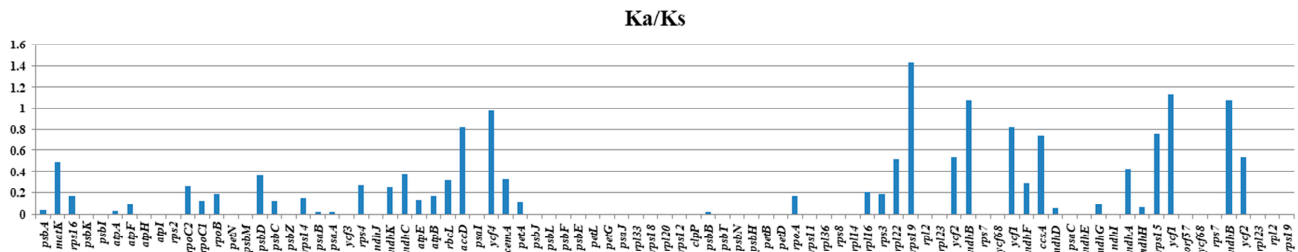


Figure 6. The nonsynonymous/synonymous substitution rates (Ka/Ks) calculated using 85 shared genes in ten *Eremurus* species.

3.5. Codon Usage

All 64 codons encoding 20 amino acids were detected (Figure 7). Two codons (AUU and AAA) occurred most frequently (>1000). The total number of codons detected ranged from 26,494 to 26,663 in *E. lactiflorus* and *E. iae*, respectively, while the number in *E. aitchisoni*, *E. regelii*, *E. hissaricus*, *E. suworowii*, *E. luteus*, *E. albertii*, *E. nderiensis*, and *E. soogdianus* was 26,655, 26,561, 26,560, 26,560, 26,558, 26,545, 26,528, and 26,528, respectively. The most common amino acid was Leucine (Leu), varying from 10.445% (2772) to 10.51% (2791). The frequency of cysteine (Cys) was the lowest at only 1.25% (333)–1.21% (323). Most codons demonstrated preferences except for AUG (Met) and UGG (Trp). Arginine encoding AGA (1.93) and AGC encoding Serine (≈ 0.29) had the highest and lowest RSCU values, respectively (Figure 7). A total of 33 codons had RSCU values > 1, of which three ended with G (UUG, AUG, and UGG), and 30 codons ended with A or U. The codons with an RSCU value less than 1 typically ended in C or G, except for UGA (stop codon), CUA, and AUA.

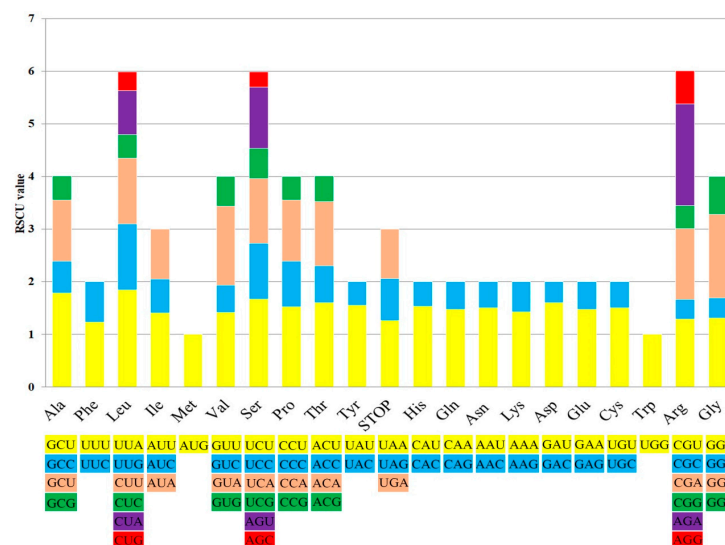


Figure 7. Identified codon contents for 20 amino acids and stop codons in all protein-coding genes of the cp genome of *Eremurus*.

3.6. Phylogenetic Analysis

As previously mentioned, our phylogenetic analysis included cp genome data from eleven species of *Eremurus* and six outgroups. The phylogenetic trees constructed based on 81 protein-coding genes and complete cp genome sequences by ML (Figure 8A,B), BI, and MP (Figure S1A,B) methods yielded quite similar topologies. The species of *Hemerocallis* and *Xanthorrhoea* (*H. fulva* and *X. preissii*, subfamilies Hemerocallidoideae and Xanthorrhoeoideae, respectively, Asphodelaceae) are sister to the rest. *Aloe* and *Aloidendron* (subfamily Asphodeloideae, Asphodelaceae) are sister to *Eremurus*. Thus, our results confirmed Naderi's studies [13] by revealing the monophyly of the genus/subgenus *Eremurus* and paraphyly of the subgenus *Henningia*. The genus is divided into two clades, the first containing only species of sect. *Hennigia* (*E. robustus*, *E. suworowii*, *E. luteus*, and *E. aitchisonii*) and the second containing species of all three sections (*E. albertii* and *E. lactiflorus* (sect. *Henningia*), *E. inderiensis* (sect. *Ammolirion*), *E. soogdianus*, *E. hissaricus*, *E. iae*, and *E. regelii* (sect. *Eremurus*)) (Figures 8 and S1). Subgenus *Eremurus* is well supported in the protein-coding analysis, with ML, BI, and MP support values of 74%, 0.99, and 78%, respectively, and has even higher support in the complete cp genome analysis, with support values of 96%, 1, and 97%, respectively. *E. regelii* and *E. iae* (sect. *Eremurus*), which are Central Asian endemics, clustered together with weak to moderate support (ML = 53%, BI = 0.68, and MP = 71%) based on protein-coding genes and with high support of 97% (ML), 1 (BI), and 99% (MP) based on the complete cp genome sequences.

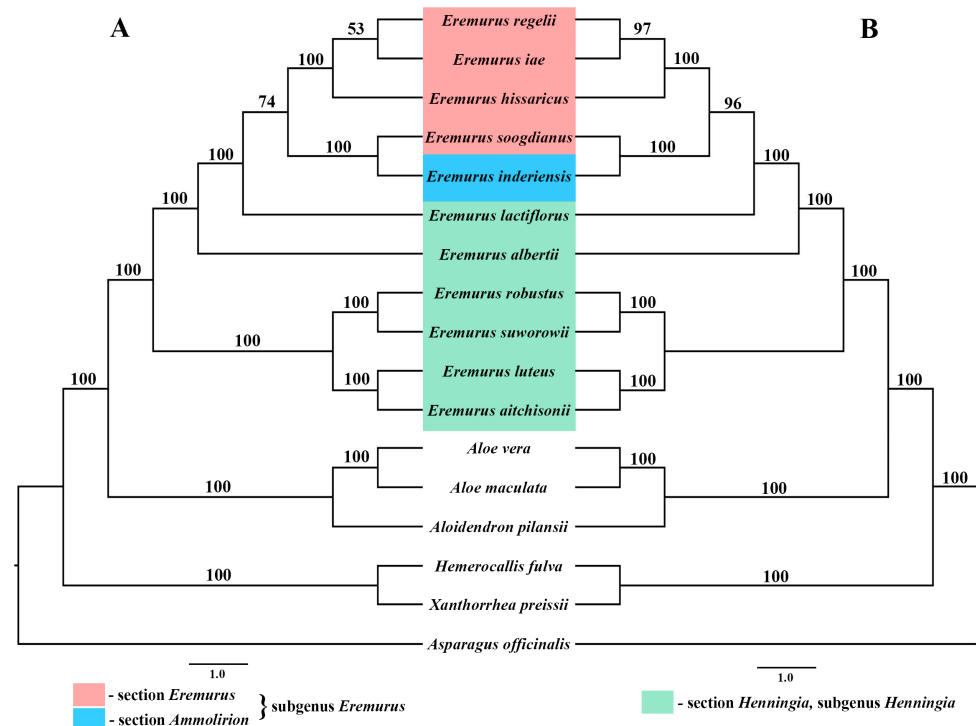


Figure 8. Phylogenetic trees of 17 species, including 11 *Eremurus* species using ML analysis based on 81 protein-coding genes (A) and complete cp genomes (B).

4. Discussion

This study is the first comparative analysis of complete cp genome sequences in *Eremurus*. The sizes of the ten cp genomes sequenced ranged from 153,782 bp (*E. lactiflorus*) to 155,482 bp (*E. aitchisonii*). It is worth noting that many related genera with similar cp genome sizes to *Eremurus* have been reported in recent years [33–37]. The cp genomes of ten *Eremurus* species showed high similarity with regard to genome size, gene sequences, gene classes, and the IR region. Their GC content was also similar, which is an important indicator of species affinity, according to Tamura et al. [38].

Introns are recognized as being central to the regulation of gene expression in plants and animals [39–41]. In the present study, 16 genes with one intron and two genes (*ycf3* and *clpP*) with two introns were identified in each of the cp genomes of the ten studied *Eremurus* species. Most of the 18 identified genes have a high similarity in the structure of introns. However, a structural change was detected in the intron of the *atpF* (764–800), *clpP* (645–665, 806–811), *ndhA* (1058–1074), *petB* (787–794), *petD* (685–693), *rpoC1* (736–744), *rps12(1)* 28,440–28,585, *rps12(2)* (69,488–70,651), *rps16* (886–892), *trnK-UUU* (2613–2615), *trnL-UAA* (501–503), *ycf3* (755–763, 726–730). The *rps12* gene in the *Eremurus* cp genome was observed to be harboring the largest intron (69,488–70,651). Further experimental work on the role of introns in *Eremurus* should be very important and interesting as the effect or relationship between gene expression and short or long introns has not been studied for *Eremurus*.

Chloroplast SSRs are often used as fingerprinting markers in studies of phylogenetic relationships, population genetics, and species identification [42,43]. In *Eremurus*, 82 (*E. inderiensis*) to 93 (*E. aitchisonii*) SSRs were found across species. Several studies found that mononucleotide repeats are dominant among SSRs in the cp genome, where A/T bases account for the majority [44–46]. *Eremurus* shows the same pattern. The SSRs found here may be useful in future analyses of genetic diversity in *Eremurus*. In particular, the unique hexanucleotide SSRs identified in *E. aitchisonii* (AAATTG); penta- and hexanucleotide SSRs identified in *E. hissaricus* (AAACT and AAATTC, respectively); tetra- (AATT) and hexanucleotide (AAAAAT) SSRs identified in *E. suworowii*; and tetra- (ATCC) and two pentanucleotide (AAATT and ATATC) SSRs identified in *E. lactiflorus* have potential for future use in species identification and assessment of population genetic diversity. The unique SSRs were absent in the cp genome of *E. inderiensis*, *E. iae*, *E. regelii*, *E. soogdianus*, *E. albertii*, and *E. luteus*. In addition, it is well known that repeat sequences play a significant role in cp genome rearrangement, recombination, gene duplication, deletion, and gene expression [47–50], as well as being responsible for substitutions and indels [51]. We identified 36 (*E. lactiflorus*) to 102 (*E. iae*) repeat sequences among the ten *Eremurus* cp genomes analyzed, with forward and palindromic repeats being the most common in *E. inderiensis* (sect. *Ammolirion*), *E. hissaricus*, *E. regelii*, *E. soogdianus* (sect. *Eremurus*), *E. aitchisonii*, *E. albertii*, *E. lactiflorus*, *E. luteus*, and *E. suworowii* (sect. *Henningia*) whereas reverse repeats were the most abundant in *E. iae* (sect. *Eremurus*). Notably, species of section *Henningia* had fewer repeat regions (36–41) compared to section *Eremurus* (40–102). Additional research focused on repeat sequences in section *Henningia* is recommended.

Highly variable DNA barcodes are essential for species identification, resource conservation, and phylogenetic analyses [52–54]. The cp genome length varied among species, with the *E. lactiflorus* genome (153,782 bp) being the longest and that of *E. aitchisonii* (155,482 bp) the shortest. There was significant similarity in the content and order of genes. Noncoding regions exhibited a higher level of sequence variation compared to other regions. The analysis of nucleotide diversity identified that the SSC region has the highest level of variation, whereas the IR region exhibits the lowest degree of variation. The P_i values of the coding regions indicate that *ycf1*, *rps15*, *rps16*, and *rpl36* exhibit high levels of variation.

The value of K_a/K_s is an indicator of selective pressure and molecular adaptation. In this study, only three genes (*rps19*, *ycf1*, and *ndhB*) underwent positive selection with $K_a/K_s > 1$. *ycf1* exhibited high values of both P_i and K_a/K_s ratios, suggesting that its evolution has been important in *Eremurus* and can be a potential molecular marker for future studies. Previous studies have reported *ycf1* to be highly variable in flowering plants [55,56] and crucial for plant viability [57]. It may prove useful in future barcoding studies in *Eremurus*.

In the cp genomes of *Eremurus*, the 85 protein-coding genes encoded 26,494 to 26,663 codons, comparable to that of the genus *Iris*, in which cp genes encode 26,169–26,353 codons [58]. The codon usage patterns in *Eremurus* species indicated a notable level of conservation in their cp genomes. Like *Iris* [58], *Amomum* [59], *Panax* [60], *Dipterygium*, and *Cleome* [61], and many other species, the conservation of codon usage

patterns in *Eremurus* species' cp genomes was evident. A noteworthy observation was that the RSCU value of a single amino acid exhibited a positive correlation with the number of codons that encoded it. Furthermore, it was observed that 27 commonly utilized codons terminated with A/U, which could potentially be linked to the significant proportion of A/T present in cp genomes [62].

Our phylogenetic analysis based on complete cp genomes and protein-coding genes confirmed previous studies based on morphological cladistic analyses [13,14] as well as *trnL-F* and nuclear (ITS) data determining the monophyly of *Eremurus* [17]. We also found that the subgenus *Eremurus* is monophyletic while the subgenus *Henningia* is paraphyletic because *E. albertii* and *E. lactiflorus* are sister to subgenus *Eremurus*. Further studies with more species, particularly from sections *Eremurus* and *Ammolirion*, are necessary to confirm this outcome, but we note that the monophyly of section *Eremurus* is supported morphologically by shared characteristics of sections *Eremurus* and *Ammolirion*, including campanulate or tubular flowers, inward curved tepals and abaxial 3–5-nerved tepals and filaments longer than perianth. By contrast, subgenus *Henningia* has subrotate flowers, filaments shorter than perianth, and one-nerved tepals. Further research on morphological variation among species of the genus is needed. Additional studies with more sampling are currently being conducted by the authors to confirm the above phylogenetic results.

5. Conclusions

Our study is the first research work to investigate the genome characteristics of the genus *Eremurus*. We sequenced, assembled, and annotated the cp genome of *E. indierensis*, *E. hissaricus*, *E. iae*, *E. regelii*, *E. soogdianus*, *E. aitchisonii*, *E. albertii*, *E. lactiflorus*, *E. luteus*, and *E. suworowii* using high-throughput technology. Our study is based on cp genome data from a total of eleven *Eremurus* species, including one previously published *E. robustus*. The cp genomes of all ten *Eremurus* species analyzed contained 131 genes, including 85 protein-coding genes, 8 rRNA genes, and 38 tRNA genes. We identified between 79 and 91 microsatellites and 36 to 102 pairs of repeat sequences among the ten *Eremurus* species cp genomes. In addition, we identified seven common SSRs and eight unique SSRs in the studied *Eremurus* species. Furthermore, we detected highly variable regions in the *ycf1*, *rps15*, *rps16*, and *rpl36* protein-coding genes. Out of all the genes studied, only *rps19*, *ycf1*, and *ndhB* have a positive Ka/Ks value. Remarkably, the *ycf1* gene stands out with both a high Pi and Ka/Ks value. These repeat motifs and highly variable genes could be used for evolutionary studies, phylogenetic relationships, and plant population genetics and species identification. Our phylogenetic reconstructions using the complete cp genome and protein-coding genes confirmed the monophyly of *Eremurus*. The subgenus *Eremurus* is monophyletic, whereas the subgenus *Henningia* is paraphyletic. However, further studies with more species, especially with the sections *Eremurus* and *Ammolirion*, are needed to confirm this result and understand the biogeography of the genus.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/f14091709/s1>. Figure S1. Phylogenetic trees of 17 species, including 11 *Eremurus* species using BI and MP analysis based on 81 protein-coding genes (A) and complete cp genomes (B). Table S1. Information on sampling localities and voucher specimens of ten newly sequenced *Eremurus* species. Table S2. Location and length of intron-containing genes in ten *Eremurus* species. Table S3. Common simple sequence repeats (SSRs) in the chloroplast genome (cpDNA) of 10 *Eremurus* species. Table S4. Unique simple sequence repeats (SSRs) in the chloroplast genome (cpDNA) of 10 *Eremurus* species. Table S5. Individual characteristics of 85 protein-coding genes.

Author Contributions: Conceptualization, D.M.; methodology, D.A., I.J. and S.T.; data analysis, D.A., I.J. and S.T.; investigation, K.T., T.D. and H.S.; writing—original draft preparation, D.M.; collection, Z.Y.; writing—review and editing, H.S., K.T. and T.D.; visualization, D.M. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by grants from the state research project “Taxonomic revision of polymorphic plant families of the flora of Uzbekistan” (FZ-20200929321) and the State Programs

for 2021–2025 years “Grid mapping of the flora of Uzbekistan” and the “Tree of life: monocots of Uzbekistan” of the Institute of Botany of the Academy of Sciences of the Republic of Uzbekistan, the National Natural Science Foundation of China (32170215), the International Partnership Program of Chinese Academy of Sciences (151853KYSB20180009), Yunnan Young & Elite Talents Project (YNWR-QNBJ-2019-033), the Ten Thousand Talents Program of Yunnan Province (202005AB160005) and the Chinese Academy of Sciences “Light of West China” Program.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: We declare that we have no conflict of interest.

References

- Hadizadeh, H.; Bahri, B.A.; Qi, P.; Wilde, H.D.; Devos, K.M. Intra-and interspecific diversity analyses in the genus *Eremurus* in Iran using genotyping-by-sequencing reveal geographic population structure. *Hortic. Res.* **2020**, *7*, 30. [\[CrossRef\]](#) [\[PubMed\]](#)
- Li, W.; Tojibaev, K.S.; Hisoriev, H.; Shomurodov, K.F.; Luo, M.; Feng, Y.; Ma, K. Mapping Asia Plants: Current status of floristic information for Central Asian flora. *Glob. Ecol. Conserv.* **2020**, *24*, e01220. [\[CrossRef\]](#)
- Eker, I. *Eremurus* M.Bieb. In *The Illustrated Flora of Turkey Web Version*; Guner, A., Kandemir, A., Menemen, Y., Yıldırım, H., Aslan, S., Eksi, G., Guner, I., Cimen, A., Sen, F., Eds.; ANG Foundation Nezahat Gökyiğit Botanik Bahçesi Publications: Istanbul, Türkiye, 2020; pp. 1–9.
- Wendelbo, P.; Furse, P. *Eremurus* of South West Asia. In *Lily Year Book*; Royal Horticultural Society: London, UK, 1969; Volume 32, pp. 56–69.
- Xinqi, C.; Turland, N. *Eremurus*. In *Flora of China*; Wu, Z., Raven, P., Eds.; Science Press and Missouri Botanical Garden Press: Beijing, China; St. Louis, MO, USA, 2000; Volume 24, pp. 159–160.
- Kamenetsky, R.; Rabinowitch, E. Flowering response of *Eremurus* to post-harvest temperatures. *Sci. Hortic.* **1999**, *79*, 75–86. [\[CrossRef\]](#)
- Aysu, T.; Demirbaş, A.; Bengü, A.Ş.; Küçük, M.M. Evaluation of *Eremurus spectabilis* for production of bio-oils with supercritical solvents. *Process Saf. Environ. Prot.* **2015**, *94*, 339–349. [\[CrossRef\]](#)
- Eghtedarnejad, N.; Mansouri, H.R. Building wooden panels glued with a combination of natural adhesive of tannin/*Eremurus* root (syrish). *Eur. J. Wood Wood Prod.* **2016**, *74*, 269–272. [\[CrossRef\]](#)
- Gaggeri, R.; Rossi, D.; Mahmood, K.; Gozzini, D.; Mannucci, B.; Corana, F.; Daglia, M.; Avanzini, A.; Mantelli, M.; Martino, E. Towards elucidating *Eremurus* root remedy: Chemical profiling and preliminary biological investigations of *Eremurus persicus* and *Eremurus spectabilis* root ethanolic extracts. *J. Med. Plants Res.* **2015**, *9*, 1038–1048. [\[CrossRef\]](#)
- Rossi, D.; Ahmed, K.M.; Gaggeri, R.; Della Volpe, S.; Maggi, L.; Mazzeo, G.; Longhi, G.; Abbate, S.; Corana, F.; Martino, E. (R)-(–)-Aloesaponol III 8-methyl ether from *Eremurus persicus*: A novel compound against leishmaniasis. *Molecules* **2017**, *22*, 519. [\[CrossRef\]](#) [\[PubMed\]](#)
- Fedtschenko, B. *Eremurus* M.Bieb. In *Flora of the USSR*; Komarov, V., Ed.; Academy of Sciences of the Soviet Union: Leningrad, Russia, 1935; Volume 4, pp. 37–52.
- Wendelbo, P. *Asphodeloideae: Asphodelus, Asphodeline & Eremurus*. In *Flora Iranica*; Rechinger, K., Ed.; Akademik Druck-u Verlagsanstalt: Graz, Austria, 1982; Volume 151, pp. 3–31.
- Naderi, S.K.; Kazempour, O.S.; Zareei, M. Phylogeny of the genus *Eremurus* (Asphodelaceae) based on morphological characters in the Flora Iranica area. *Iran. J. Bot.* **2009**, *15*, 7–35.
- Makhmudjanov, D.; Juramurodov, I.; Kurbonalieva, M.; Yusupov, Z.; Dekhkonov, D.; Deng, T.; Tojibaev, S.K.; Sun, H. Genus *Eremurus* (Asphodelaceae) in the flora of Uzbekistan. *Plant Divers. Cen. As.* **2022**, *2*, 82–127. [\[CrossRef\]](#)
- Chase, M.W.; De Bruijn, A.Y.; Cox, A.V.; Reeves, G.; Rudall, P.J.; Johnson, M.A.; Eguiarte, L.E. Phylogenetics of Asphodelaceae (Asparagales): An analysis of plastid *rbcL* and *trnL-F* DNA sequences. *Ann. Bot.* **2000**, *86*, 935–951. [\[CrossRef\]](#)
- Devey, D.S.; Leitch, I.; Pires, J.C.; Pillon, Y.; Chase, M.W. Systematics of Xanthorrhoeaceae sensu lato, with an emphasis on *Bulbine*. *Aliso A J. Syst. Florist. Bot.* **2006**, *22*, 345–351. [\[CrossRef\]](#)
- Safar, K.N.; Osaloo, S.K.; Assadi, M.; Zarrei, M.; Mozaffar, M.K. Phylogenetic analysis of *Eremurus*, *Asphodelus*, and *Asphodeline* (Xanthorrhoeaceae-Asphodeloideae) inferred from plastid *trnL-F* and nrDNA ITS sequences. *Biochem. Syst. Ecol.* **2014**, *56*, 32–39. [\[CrossRef\]](#)
- Makhmudjanov, D.; Yusupov, Z.; Abdullaev, D.; Deng, T.; Tojibaev, K.; Sun, H. The complete chloroplast genome of *Eremurus robustus* (Asphodelaceae). *Mitochondrial DNA B Resour.* **2019**, *4*, 3366–3367. [\[CrossRef\]](#) [\[PubMed\]](#)
- Jin, J.J.; Yu, W.B.; Yang, J.B.; Song, Y.; DePamphilis, C.W.; Yi, T.S.; Li, D.Z. GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol.* **2020**, *21*, 241. [\[CrossRef\]](#) [\[PubMed\]](#)
- Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, *28*, 1647–1649. [\[CrossRef\]](#) [\[PubMed\]](#)
- Beier, S.; Thiel, T.; Münch, T.; Scholz, U.; Mascher, M. MISA-web: A web server for microsatellite prediction. *Bioinformatics* **2017**, *33*, 2583–2585. [\[CrossRef\]](#) [\[PubMed\]](#)

22. Kurtz, S.; Choudhuri, J.V.; Ohlebusch, E.; Schleiermacher, C.; Stoye, J.; Giegerich, R. REPuter: The manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* **2001**, *29*, 4633–4642. [\[CrossRef\]](#)
23. Lohse, M.; Drechsel, O.; Bock, R. OrganellarGenomeDRAW (OGDRAW): A tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr. Genet.* **2007**, *52*, 267–274. [\[CrossRef\]](#)
24. Frazer, K.A.; Pachter, L.; Poliakov, A.; Rubin, E.M.; Dubchak, I. VISTA: Computational tools for comparative genomics. *Nucleic Acids Res.* **2004**, *32*, W273–W279. [\[CrossRef\]](#)
25. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [\[CrossRef\]](#)
26. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Rozas, J.; Ferrer-Mata, J.C.; Sanchez-DelBarrio, P.; Librado, P.; Guirao-Rico, S.E. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Mol. Biol. Evol.* **2017**, *34*, 3299–3302. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Ronquist, F.; Teslenko, M.; Van der Mark, P.; Ayres, D.; Darling, A. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **2012**, *61*, 539–542. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313. [\[CrossRef\]](#)
31. Edler, D.; Klein, J.; Antonelli, A.; Silvestro, D. raxmlGUI 2.0: A graphical interface and toolkit for phylogenetic analyses using RAxML. *Methods Ecol. Evol.* **2020**, *12*, 373–377. [\[CrossRef\]](#)
32. Swofford, D.L. *PAUP: Phylogenetic Analysis Using Parsimony, Version 4.0 b10*; Sinauer Associates: Sunderland, MA, USA, 2002.
33. Zhang, X.; Lang, L.; Shang, X.; Wang, Z.; Jiang, L.; Pei, X.; Lu, J.; Li, D.; Yang, J. The complete chloroplast genome sequence of *Hemerocallis minor* (Asphodelaceae). *Mitochondrial DNA B Resour.* **2022**, *7*, 1227–1228. [\[CrossRef\]](#)
34. Ou, X.; Liu, G.; Wu, L.-H. The complete chloroplast genome of *Hemerocallis citrina* (Asphodelaceae), an ornamental and medicinal plant. *Mitochondrial DNA B Resour.* **2020**, *5*, 1109–1110. [\[CrossRef\]](#)
35. Lee, J.; Lim, J.-S.; Kim, S.-Y.; Chun, H.S.; Lee, D.; Nah, G. The complete chloroplast genome of *Hemerocallis fulva*. *Mitochondrial DNA B Resour.* **2019**, *4*, 2199–2200. [\[CrossRef\]](#)
36. Ren, J.J.; Wang, J.; Lee, K.K.; Deng, H.; Xue, H.; Zhang, N.; Zhao, J.C.; Cao, T.; Cui, C.L.; Zhang, X.H. The complete chloroplast genome of *Aloe vera* from China as a Chinese herb. *Mitochondrial DNA B Resour.* **2020**, *5*, 1092–1093. [\[CrossRef\]](#)
37. Malakasi, P.; Bellot, S.; Dee, R.; Grace, O.M. Museomics clarifies the classification of *Aloidendron* (Asphodelaceae), the iconic African tree aloes. *Front. Plant Sci.* **2019**, *10*, 1227. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Tamura, K.; Peterson, D.; Peterson, N.; Stecher, G.; Nei, M.; Kumar, S. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **2011**, *28*, 2731–2739. [\[CrossRef\]](#)
39. Callis, J.; Fromm, M.; Walbot, V. Introns increase gene expression in cultured maize cells. *Genes Dev.* **1987**, *1*, 1183–1200. [\[CrossRef\]](#) [\[PubMed\]](#)
40. Emami, S.; Arumainayagam, D.; Korf, I.; Rose, A.B. The effects of a stimulating intron on the expression of heterologous genes in *A. rabidopsis thaliana*. *Plant Biotechnol. J.* **2013**, *11*, 555–563. [\[CrossRef\]](#)
41. Choi, T.; Huang, M.; Gorman, C.; Jaenisch, R. A generic intron increases gene expression in transgenic mice. *Mol. Cell. Biol.* **1991**, *11*, 3070–3074. [\[CrossRef\]](#)
42. Olmstead, R.G.; Palmer, J.D. Chloroplast DNA systematics: A review of methods and data analysis. *Am. J. Bot.* **1994**, *81*, 1205–1224. [\[CrossRef\]](#)
43. Saski, C.; Lee, S.-B.; Daniell, H.; Wood, T.C.; Tomkins, J.; Kim, H.G.; Jansen, R.K. Complete chloroplast genome sequence of *Glycine max* and comparative analyses with other legume genomes. *Plant Mol. Biol.* **2005**, *59*, 309–322. [\[CrossRef\]](#)
44. Ellegren, H. Microsatellites: Simple sequences with complex evolution. *Nat. Rev. Genet.* **2004**, *5*, 435–445. [\[CrossRef\]](#)
45. George, B.; Bhatt, B.S.; Awasthi, M.; George, B.; Singh, A.K. Comparative analysis of microsatellites in chloroplast genomes of lower and higher plants. *Curr. Genet.* **2015**, *61*, 665–677. [\[CrossRef\]](#)
46. Ren, F.; Wang, L.; Li, Y.; Zhuo, W.; Xu, Z.; Guo, H.; Liu, Y.; Gao, R.; Song, J. Highly variable chloroplast genome from two endangered Papaveraceae lithophytes *Corydalis tomentella* and *Corydalis saxicola*. *Ecol. Evol.* **2021**, *11*, 4158–4171. [\[CrossRef\]](#)
47. Gemayel, R.; Vences, M.D.; Legendre, M.; Verstrepen, K.J. Variable tandem repeats accelerate evolution of coding and regulatory sequences. *Annu. Rev. Genet.* **2010**, *44*, 445–477. [\[CrossRef\]](#) [\[PubMed\]](#)
48. Do, H.D.K.; Kim, J.S.; Kim, J.-H. A *trnL*-CAU triplication event in the complete chloroplast genome of *Paris verticillata* M. Bieb. (Melanthiaceae, Liliales). *Genome Biol. Evol.* **2014**, *6*, 1699–1706. [\[CrossRef\]](#) [\[PubMed\]](#)
49. Vieira, L.d.N.; Faoro, H.; Rogalski, M.; Fraga, H.P.d.F.; Cardoso, R.L.A.; de Souza, E.M.; de Oliveira Pedrosa, F.; Nodari, R.O.; Guerra, M.P. The complete chloroplast genome sequence of *Podocarpus lambertii*: Genome structure, evolutionary aspects, gene content and SSR detection. *PLoS ONE* **2014**, *9*, e90618. [\[CrossRef\]](#) [\[PubMed\]](#)
50. Li, B.; Zheng, Y. Dynamic evolution and phylogenomic analysis of the chloroplast genome in Schisandraceae. *Sci. Rep.* **2018**, *8*, 9285. [\[CrossRef\]](#) [\[PubMed\]](#)

51. Yi, X.; Gao, L.; Wang, B.; Su, Y.-J.; Wang, T. The complete chloroplast genome sequence of *Cephalotaxus oliveri* (Cephalotaxaceae): Evolutionary comparison of *Cephalotaxus* chloroplast DNAs and insights into the loss of inverted repeat copies in gymnosperms. *Genome Biol. Evol.* **2013**, *5*, 688–698. [[CrossRef](#)] [[PubMed](#)]
52. Gregory, T.R. DNA barcoding does not compete with taxonomy. *Nature* **2005**, *434*, 1067. [[CrossRef](#)] [[PubMed](#)]
53. Liu, X.; Chang, E.-M.; Liu, J.-F.; Huang, Y.-N.; Wang, Y.; Yao, N.; Jiang, Z.-P. Complete chloroplast genome sequence and phylogenetic analysis of *Quercus bawanglingensis* Huang, Li et Xing, a vulnerable oak tree in China. *Forests* **2019**, *10*, 587. [[CrossRef](#)]
54. Bringloe, T.T.; Saunders, G.W. DNA barcoding of the marine macroalgae from Nome, Alaska (Northern Bering Sea) reveals many trans-Arctic species. *Polar Biol.* **2019**, *42*, 851–864. [[CrossRef](#)]
55. Dong, W.; Liu, J.; Yu, J.; Wang, L.; Zhou, S. Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. *PLoS ONE* **2012**, *7*, e35071. [[CrossRef](#)]
56. Amar, M.H. *ycf1-ndhF* genes, the most promising plastid genomic barcode, sheds light on phylogeny at low taxonomic levels in *Prunus persica*. *J. Genet. Eng. Biotechnol.* **2020**, *18*, 42. [[CrossRef](#)]
57. Kikuchi, S.; Bédard, J.; Hirano, M.; Hirabayashi, Y.; Oishi, M.; Imai, M.; Takase, M.; Ide, T.; Nakai, M. Uncovering the protein translocon at the chloroplast inner envelope membrane. *Science* **2013**, *339*, 571–574. [[CrossRef](#)]
58. Feng, J.-L.; Wu, L.-W.; Wang, Q.; Pan, Y.-J.; Li, B.-L.; Lin, Y.-L.; Yao, H. Comparison Analysis Based on Complete Chloroplast Genomes and Insights into Plastid Phylogenomic of Four *Iris* Species. *BioMed Res. Int.* **2022**, *2022*, 2194021. [[CrossRef](#)] [[PubMed](#)]
59. Yang, L.; Feng, C.; Cai, M.-M.; Chen, J.-H.; Ding, P. Complete chloroplast genome sequence of *Amomum villosum* and comparative analysis with other Zingiberaceae plants. *Chin. Herb. Med.* **2020**, *12*, 375–383. [[CrossRef](#)] [[PubMed](#)]
60. Kim, K.J.; Lee, H.L. Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants. *DNA Res.* **2004**, *11*, 247–261. [[CrossRef](#)] [[PubMed](#)]
61. Alzahrani, D.; Albokhari, E.; Yaradua, S.; Abba, A. Complete chloroplast genome sequences of *Dipterygium glaucum* and *Cleome chrysantha* and other *Cleomaceae* Species, comparative analysis and phylogenetic relationships. *Saudi J. Biol. Sci.* **2021**, *28*, 2476–2490. [[CrossRef](#)] [[PubMed](#)]
62. Eguiluz, M.; Rodrigues, N.F.; Guzman, F.; Yuyama, P.; Margis, R. Evolution. The chloroplast genome sequence from *Eugenia uniflora*, a Myrtaceae from Neotropics. *Plant Syst. Evol.* **2017**, *303*, 1199–1212. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.