


A comparative analysis of complete plastid genomes from *Prangos fedtschenkoi* and *Prangos lipskyi* (Apiaceae)

Feruz U. Mustafina^{1,2}  | Dong-Keun Yi¹  | Kyung Choi¹ | Chang Ho Shin¹ | Komiljon Sh. Tojibaev² | Stephen R. Downie³

¹Division of Forest Biodiversity and Herbarium, Korea National Arboretum, Pocheon, Republic of Korea

²Institute of Botany, Uzbek Academy of Sciences, Tashkent, Republic of Uzbekistan

³Department of Plant Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA.

Correspondence

Feruz U. Mustafina, Institute of Botany, Uzbek Academy of Sciences, Tashkent, Republic of Uzbekistan.
Email: mustafinaferuz@yahoo.com

Funding information

Korea National Arboretum, Grant/Award Number: KNA-1-1-17, 15-2; Ministry of Innovation Development of the Republic of Uzbekistan, Grant/Award Number: VA-FA-F5010

Abstract

Prangos fedtschenkoi (Regel & Schmalh.) Korovin and *P. lipskyi* Korovin (Apiaceae) are rare plant species endemic to mountainous regions of Middle Asia. Both are edificators of biotic communities and valuable resource plants. The results of recent phylogenetic analyses place them in *Prangos* subgen. *Koelzella* (M. Hiroe) Lyskov & Pimenov and suggest they may possibly represent sister species. To aid in development of molecular markers useful for intraspecific phylogeographic and population-level genetic studies of these ecologically and economically important plants, we determined their complete plastid genome sequences and compared the results obtained to several previously published plastomes of Apiaceae. The plastomes of *P. fedtschenkoi* and *P. lipskyi* are typical of Apiaceae and most other higher plant plastid DNAs in their sizes (153,626 and 154,143 bp, respectively), structural organization, gene arrangement, and gene content (with 113 unique genes). A total of 49 and 48 short sequence repeat (SSR) loci of 10 bp or longer were detected in *P. fedtschenkoi* and *P. lipskyi* plastomes, respectively, representing 42–43 mononucleotides and 6 AT dinucleotides. Seven tandem repeats of 30 bp or longer with a sequence identity $\geq 90\%$ were identified in each plastome. Further comparisons revealed 319 polymorphic sites between the plastomes (IR, 21; LSC, 234; SSC, 64), representing 43.8% transitions (Ts), 56.1% transversions (Tv), and a Ts/Tv ratio of 0.78. Within genic regions, two indel events were observed in *rpoA* (6 and 51 bp) and *ycf1* (3 and 12 bp), and one in *ndhF* (6 bp). The most variable intergenic spacer region was that of *accD/psal*, with 21.1% nucleotide divergence. Each *Prangos* species possessed one of two separate inversions (either 5 bp in *ndhB* intron or 9 bp in *petB* intron), and these were predicted to form hairpin structures with flanking repeat sequences of 18 and 19 bp, respectively. Both species have also incorporated novel DNA in the LSC region adjacent to the LSC/IRa junction, and BLAST searches revealed it had a 100 bp match (86% sequence identity) to noncoding mitochondrial DNA. *Prangos*-specific primers were developed for the variable *accD/psal* intergenic spacer and preliminary PCR-surveys suggest that this region will be useful for future phylogeographic and population-level studies.

KEYWORDS

Apiaceae, chloroplast genome, mitochondrial DNA, plastid DNA, Umbelliferae

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2018 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd.

1 | INTRODUCTION

The genus *Prangos* Lindl. (Apiaceae, Umbelliferae) comprises some 45 species mostly endemic to Asia (Lyskov, Degtjareva, Samigullin, & Pimenov, 2017). They are typically herbaceous, xerophytic plants that perform important roles as edifiers of biotic communities. They are also of great economic importance, as many of its members are sources of phytocoumarins, ornamental plants, and fodder for cattle. The plants are diverse morphologically and in their fruit anatomy, resulting in ever-changing species delimitations and infrageneric classifications (Lyskov et al., 2017).

Prangos fedtschenkoi (Regel & Schmalh.) Korovin and *P. lipskyi* Korovin are rare species endemic to the mountainous regions of Middle Asia (Shishkin, 1950). *P. fedtschenkoi* is more common in the Pamir-Alay than the Tien Shan mountain systems, with several populations occurring in Kyrgyzstan and Tajikistan (Figure 1a). *P. lipskyi* is an endemic to a narrow geographic area in the Western Tien Shan Mountains and grows on talus slopes (Figure 1b). A recent revision of the genus based on morphological, carpological, and molecular evidence revealed that *P. fedtschenkoi* and *P. lipskyi*, along with three other species, unite as monophyletic in *Prangos* subgen. *Koelzella* (M. Hiroe) Lyskov & Pimenov (Lyskov et al., 2017). In the Bayesian inference tree of molecular data, these two species form two branches of a trichotomy (with the clade of *P. pabularia* Lindl. + *P. ornata* Kuzmina), suggesting that they may comprise monophyletic sister taxa in other trees. Other molecular systematic studies including *Prangos* species exist (e.g., Banasiak et al., 2013), however, none included representation of *P. fedtschenkoi* or *P. lipskyi*.

Significant declines of both *P. fedtschenkoi* and *P. lipskyi* populations have resulted from human activities, such as agriculture, construction, overgrazing, and collection for pharmaceutical and cosmetological purposes (CACILM, 2006; Limin & Zhang, 2012). *P. fedtschenkoi* is also intensively collected from the wild, as there are currently no plantations of this valuable resource plant. Only limited information is available on the nature of genetic variability in *P. fedtschenkoi* and *P. lipskyi* that can be used to inform conservation strategies, management, and species restoration. Recently, six populations of *P. fedtschenkoi* from Uzbekistan were studied

using 10 intersimple sequence repeat (ISSR) markers (Mustafina et al., 2017). A high level of genetic differentiation among populations was detected based on Nei's genetic diversity analysis and AMOVA that was likely caused by range fragmentation as a result of anthropogenic impacts. These factors, along with strong inbreeding, may have decreased within-population genetic diversity because of low gene flow.

To aid in the development of molecular markers useful for intraspecific phylogeographic and population-level genetic studies of these rare, but ecologically and economically important plant species, we determined their complete plastid genome sequences, performed a comparative analysis of these data, and compared the results obtained to several previously published plastomes of other Apiaceae species. While the plastomes are typical of Apiaceae and most other eudicot plastid DNAs in their structural organization, gene arrangement, and gene content, variable regions were revealed that will be useful as molecular markers in future studies.

2 | MATERIALS AND METHODS

2.1 | Plants materials and DNA extraction

Voucher specimens of field-collected individuals are as follows: *Prangos fedtschenkoi*, Baysun region, Surkhandarya province, Uzbekistan, 28 June 2015, *Turginov* 2310 (TASH) and *Prangos lipskyi*, Chatkal chain, Sari-Chelek Nature Reserve, Kyrgyzstan, 11 July 2015, *Lazkov* 266 (FRU). DNA extractions were conducted using the DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) and checked by electrophoresis in a 1.3% agarose gel, stained with ethidium bromide, and run with a DNA ladder (Thermo Scientific, Waltham, MA, USA). Concentration and quality of DNA were checked using a NanoDrop 2000 spectrophotometer (Thermo Scientific).

2.2 | Sequencing and plastid genome assembly

Plastid genome sequences were obtained using the Illumina HiSeq 2000 sequencing system and standard protocols. A total of 12,484,132 reads of *P. fedtschenkoi* and 10,849,852 reads of

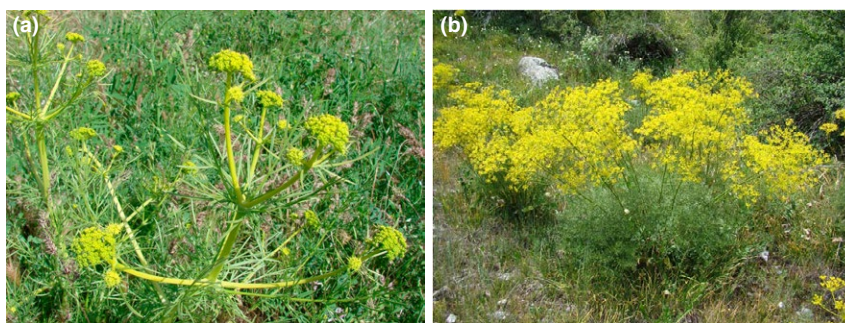


FIGURE 1 A. *Prangos fedtschenkoi*. Piedmont of the Nuratau range (Pamir-Alay mountain system), Forish district, Jizzakh region, Uzbekistan. Photo by N. Beshko (04 June 2009). B. *Prangos lipskyi*. Chatkal range (Western Tien Shan mountain system), Sary - Chelek Nature Reserve, Jalal - Abad Province, Kyrgyzstan. Photo by G. Lazkov (10 July 2015)

P. lipskyi were analyzed. The filtered sequences were assembled using Bowtie 2 v. 2.2.3 (Langmead & Salzberg, 2012) with *Daucus carota* subsp. *sativus* (domesticated carrot; GenBank accession number DQ898156) used as the reference sequence since its plastome has ancestral IR boundaries and no gene rearrangements (Ruhlman et al., 2006). A total of 187,581 reads of *P. fedtschenkoi* and 389,753 reads of *P. lipskyi* were mapped to the reference sequence, with an average coverage of 599.5× for *P. fedtschenkoi* and 1,086× for *P. lipskyi*.

2.3 | Chloroplast gene annotation

Gene annotations were performed using BLAST (BLASTN, PHI-BLAST, BLASTX) and the program Dual Organellar Genome Annotator (DOGMA; Wyman, Jansen, & Boore, 2004). Gene nomenclature followed the Chloroplast Genome Database (<http://chloroplast.cbio.psu.edu>; Cui et al., 2006). Complete plastid genomes have been deposited in GenBank (*P. fedtschenkoi*, KY652265; *P. lipskyi*, KY652266) using Sequin v. 15.50 (6 January 2017). Circular genome maps were drawn using OrganellarGenomeDRAW (Lohse, Drechsel, Kahlau, & Bock, 2013), with nomenclature and arrangement of the plastome's major structural components following convention. Sequences of six Apiaceae plastomes, that is, *D. carota*, *Petroselinum crispum* (HM596073), *Coriandrum sativum* (KR002656), *Prangos trifida* (NC037852), and the two newly sequenced plastomes of *P. fedtschenkoi* and *P. lipskyi*, were compared to align IR single-copy boundary positions.

2.4 | Sequence analysis

AT content and codon usage were obtained using MEGA6 v. 6.06 (Tamura, Stecher, Peterson, Filipski, & Kumar, 2013). Plastid genomes were aligned using Mauve v. 2.3.1 (Darling, Mau, & Perna, 2010) and sequences of the protein-coding genes and intergenic spacer (IGS) regions were extracted. Nucleotide diversity and K_a/K_s values of protein-coding regions were analyzed using DnaSP v. 5.10.01 (Librado & Rozas, 2009) and MEGA6. Polymorphic sites (Ts and Tv) were analyzed by pairwise comparison of each protein-coding region, intergenic spacer, and intron using MUSCLE (Edgar, 2004) and the SNP finder program integrated into Geneious v. 10.3 (Kearse et al., 2012). To determine the origin of the novel plastid DNA fragments adjacent to IRA/LSC, these sequences were checked against the NCBI nucleotide DNA database using BLAST.

2.5 | Prangos-specific primers

Primers specific for the *accD/psal* IGS region in *Prangos* were developed with software Primer3 v. 2.3.7 (Untergasser et al., 2012) and synthesized by Macrogen, Korea. Three pairs of primers located in the spacer region between *accD* and *psal* were considered, with the goal of identifying those primer pairs yielding the best results. PCR was conducted with the following regime: initial denaturation at 94°C (1 min), followed by 35 cycles of denaturation at

94°C (30 s), annealing at 54°C (40 s), extension at 72°C (40 s), and a final extension at 72°C (45 min). PCR products of the *accD/psal* intergenic spacer region were obtained for 46 additional accessions belonging to 16 Middle Asian *Prangos* species, including those from geographically remote populations. *P. fedtschenkoi* was represented by 10 accessions (from Kyrgyzstan, Tajikistan, and Uzbekistan) and *P. lipskyi* was represented by three accessions (from Kyrgyzstan and Uzbekistan).

2.6 | Repeat and inversion analyses

Short sequence repeats (SSRs) were analyzed with Phobos v. 3.3.12 (Mayer, Leese, & Tollrian, 2010) using the default mismatch and a gap score of -5, with a size threshold ≥ 10 bp. Repeating sequences were analyzed by REPuter (Kurtz et al., 2001) and Tandem Repeats Finder v. 4.07b (Benson, 1999). Direct and inverted (palindromic) repeats were identified with a size ≥ 30 bp and a Hamming distance of 3 (limiting hits to sequence identity of $\geq 90\%$). Secondary structure was predicted by mFOLD (Zuker, 2003). To survey for the presence of identical tandem repeats and the two small inversions and their associated flanking repeat regions identified in the plastomes of *P. fedtschenkoi* and *P. lipskyi* in related taxa, the complete plastid genomes of 23 other Apiaceae species (represented by 25 accessions) and two Araliaceae species were downloaded from NCBI's public database. Sequence similarity searches and alignments were carried out using Clustal W (Thompson, Higgins, & Gibson, 1994). Repeating sequences were analyzed by REPuter, and the forms and stabilities of hairpin structures were evaluated by MFOLD.

3 | RESULTS

3.1 | General plastome features

The plastomes of *P. fedtschenkoi* and *P. lipskyi* are typical of most other nonmonocot angiosperm plastid DNAs in their structural organization, gene arrangement, and gene content (Figures 2 and 3). They have two large inverted repeats (IRs), separated by large single-copy (LSC) and small single-copy (SSC) regions. A comparison of their major structural features, to each other and to *P. trifida* and *D. carota*, is presented in Table 1. The plastomes of *P. fedtschenkoi* and *P. lipskyi* range in size from 153,626 bp in *P. fedtschenkoi* to 154,143 bp in *P. lipskyi*, with the latter having a substantially larger LSC region. Both plastomes share identical complements of genes, each with four unique rRNA genes, 30 unique tRNA genes, and 79 unique protein-coding genes. A list of unique genes present in both plastomes, as represented by *P. fedtschenkoi*, is presented in Table 2. Allowing for duplication of genes in the IR and open reading frames (ORFs), each plastome contains 134 complete, predicted coding regions. Comparing the *Prangos* plastome structural features to those of *D. carota* reveals major differences in sizes of all major structural components and total number of predicted coding regions (Table 1).

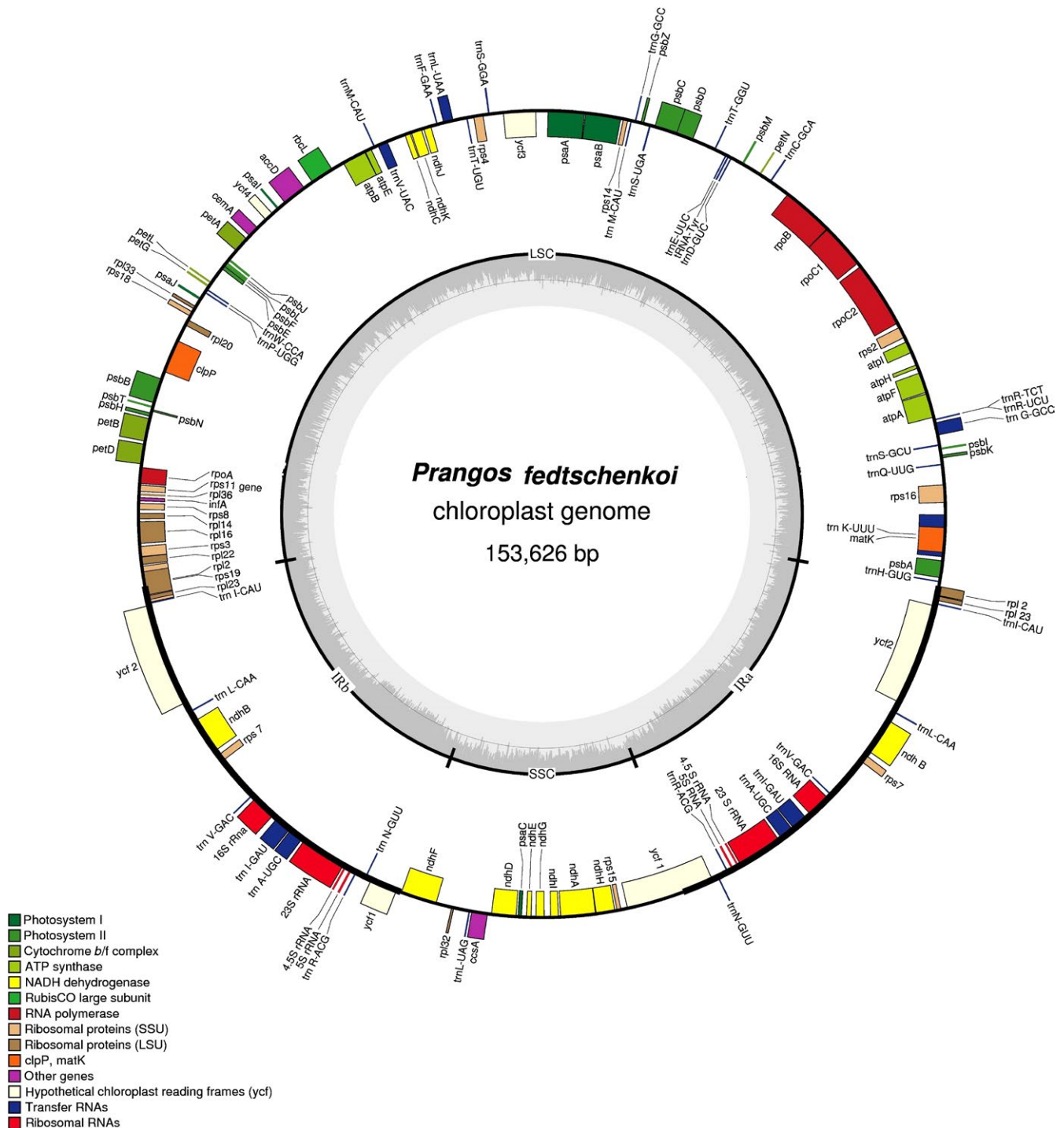


FIGURE 2 Circular plastome map of *Prangos fedtschenkoi*. Genes are classified into 14 groups according to their biological function and are shown by different colored boxes. Genes transcribed clockwise are shown inside of the circle; genes transcribed counter-clockwise are shown outside of the circle. The small single-copy (SSC) region is 17,494 bp in size, the large single-copy (LSC) region is 85,614 bp in size, and each inverted repeat (IRa, IRb) region is 25,259 bp in size. The internal gray circle indicates GC content and the thin circular line marks the 50% threshold. The nucleotide sequence of the *P. fedtschenkoi* chloroplast genome appears under the accession number KY652265 in the DDBJ/GenBank databases

In *P. fedtschenkoi*, 57.62% of the plastome consists of gene-coding regions (54.02% proteins and 3.6% RNAs). In *P. lipyskyi*, 57.61% of the plastome is gene-coding (54.01% proteins and 3.6% RNAs).

Noncoding regions comprise approximately 42.4% of each plastome (11.3% intron and 31.1% IGS regions). The overall AT content in each *Prangos* species was 62.3% (Table 3). The AT content of the IR (57.1%)

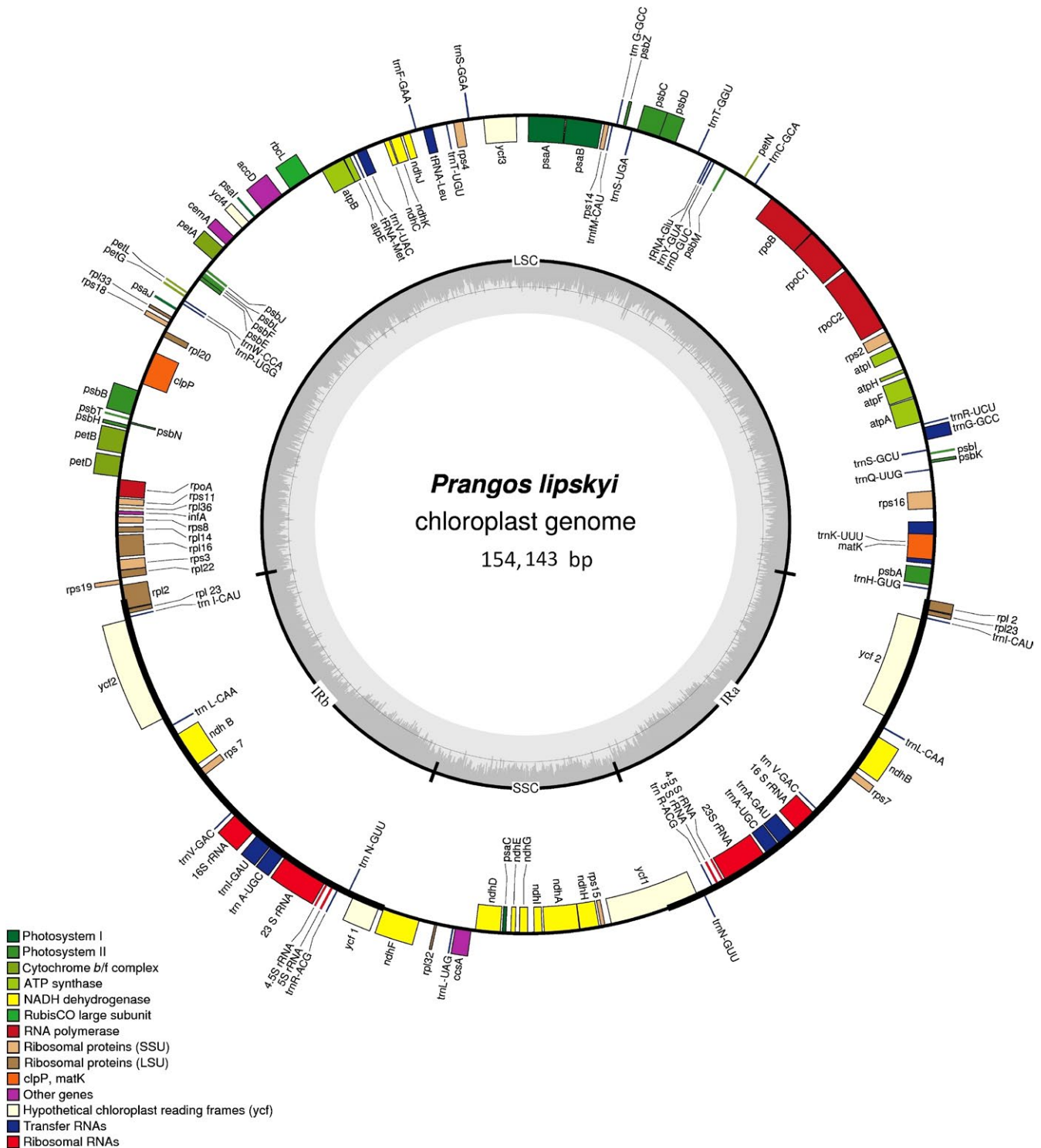


FIGURE 3 Circular plastome map of *Prangos lipskyi*. Genes are classified into 14 groups according to their biological function and are shown by different colored boxes. Genes transcribed clockwise are shown inside of the circle; genes transcribed counter-clockwise are shown outside of the circle. The small single-copy (SSC) region is 17,402 bp in size, the large single-copy (LSC) region is 86,131 bp in size, and each inverted repeat (IRA, IRB) region is 25,305 bp in size. The internal gray circle indicates GC content and the thin circular line marks the 50% threshold. The nucleotide sequence of the *P. lipskyi* chloroplast genome appears under the accession number KY652266 in the DDBJ/GenBank databases

is lower than that of both LSC (64.0%) and SSC (68.7%) regions. Within protein-coding regions of *P. fedtschenkoi*, the AT content is higher at the third codon position (63.0%), whereas in *P. lipskyi* the

AT content is lowest at the third codon position (61.7%). The codon usage pattern suggests that codons containing A/T in the third position are given preference in *P. fedtschenkoi* and *P. lipskyi* (Supporting

TABLE 1 Comparison of major structural features of the plastomes of *Prangos fedtschenkoi*, *Prangos lipskyi*, *Prangos trifida*, and *Daucus carota*

Feature	<i>P. fedtschenkoi</i>	<i>P. lipskyi</i>	<i>P. trifida</i>	<i>D. carota</i>
Entire plastome size (bp)	153,626	154,143	153,510	155,911
IR size (bp)	25,259	25,305	24,792	27,051
LSC region size (bp)	85,614	86,131	86,481	84,242
SSC region size (bp)	17,494	17,402	17,445	17,567
Number of coding regions	134	134	134	136
Number of genes	113	113	113	115
Number of protein-coding genes	79	79	79	81
Number of genes duplicated in the IR	21	21	21	21
Number of pseudo-genes	2 (<i>rpl2</i> , <i>ycf1</i>)	2 (<i>rpl2</i> , <i>ycf1</i>)	2 (<i>rpl2</i> , <i>ycf1</i>)	2 (<i>rps19</i> , <i>ycf1</i>)
Number of open reading frames	4	4	4	4
Number of tRNA genes	30	30	30	30
Number of rRNA genes	4	4	4	4
Number of genes with intron(s)	18	18	18	18

Information Tables S1 and S2), in accordance with the codon usage pattern of the majority of eudicot species.

3.2 | IR single-copy boundary positions

A comparison of IR single-copy boundary positions of the three *Prangos* species, *D. carota*, and two additional species of the family Apiaceae (*Petroselinum crispum*, parsley; *Coriandrum sativum*, coriander) is presented in Figure 4. In *Prangos* spp., the IRb/LSC boundary falls in gene *rpl2*, which results in *rpl2* pseudogenes of 567 bp in *P. fedtschenkoi* and 591 bp in *P. lipskyi* in IRa adjacent to the IRa/LSC junction. These *rpl2* pseudogenes have lost the second exon of the gene. A similar IRb/LSC boundary position in *rpl2* occurs in *P. trifida*

and *P. crispum*. *D. carota* has the ancestral IRb/LSC boundary position in or near *rps19*. A considerable contraction of the IR occurs in *C. sativum*. Here, IRa/LSC is located downstream of *trnK-UUU*, the positions of *trnH-GUG* and *psbA* are contained within the IR, and the IRb/LSC border falls downstream of *rps12* and *trnV-GAC*. In all species, the IRa/SSC boundary position occurs in *ycf1*, creating *ycf1* pseudogenes ranging between 1,674 bp (*P. fedtschenkoi*) and 1,939 bp (*P. trifida*) in IRb at the IRb/SSC boundary position.

3.3 | Plastid genome sequence divergence

The average sequence divergence between entire plastomes of *P. fedtschenkoi* and *P. lipskyi* was 0.6%. Average sequence divergences between *P. fedtschenkoi* and *D. carota* and between *P. lipskyi* and *D. carota* were 9.1% and 9.9%, respectively. The average divergence values of the IR regions between *P. fedtschenkoi* and *P. lipskyi*, between *P. fedtschenkoi* and *D. carota*, and between *P. lipskyi* and *D. carota* were 0.1%, 9.0%, and 9.5%, respectively, while divergence values of the LSC regions were 1.1%, 9.1%, and 11.4%, respectively, and divergence values of the SSC regions were 1.1%, 9.4%, and 9.3%, respectively.

3.4 | Coding region divergence

Detailed sequence comparisons of 83 protein-coding genes between *P. fedtschenkoi* and *P. lipskyi* are provided in Supporting Information Table S3. Average Ka/Ks ratios for these gene regions were 0.3988 across entire plastomes, 0.2857 for the IRs, and 0.2476 for the LSC and 0.6795 for the SSC regions. Sequence divergence values of each protein-coding gene region between *P. fedtschenkoi* and *P. lipskyi* varied from identity to 1.3% (reaching 0.1% in the IR, 1.3% in the LSC, and 0.7% in the SSC). Genes showing highest sequence divergence included *rpoA* (1.3%), *psbT* (1.0%), *psbI* (0.9%), and *ycf1* (0.7%). Genes *rpoA*, *ycf1*, and *ndhF* exhibited length difference in pairwise comparisons resulted from multiple indels and length variation (of 57, 15, and 6 bp, respectively).

3.5 | Intergenic spacer region divergence

Prangos plastomes contain 108 intergenic spacer regions longer than 10 bp in length (Supporting Information Table S4). Sequence divergence values of these IGS regions ranged from identity to 21.1% (reaching 6.9% in the IR, 21.1% in the LSC, and 19.9% in the SSC). The most divergent IGS regions are *accD/psaI* (21.1%), *psbZ/trnG-GCC* (13.2%), *rps16/trnQ-UUG* (8.5%), and *rps14/psaB* (7.3%) in the LSC, *rps19/rpl2* (6.9%) in the IR, and *ndhE/ndhG* (19.9%) in the SSC.

Small- and medium-sized insertion and deletion events prevailed over long length indels in IGS regions between the *P. fedtschenkoi* and *P. lipskyi* plastomes. The total length of all indel events varied from 7 bp in the IR, 119 bp in the SSC, to 641 bp in the LSC, totaling 767 bp. Medium-sized indel events (20–40 bp) occurred in *psbZ/trnG-GCC* (38 bp), *accD/psaI* (34 bp), and *petA/psbJ* (23 bp) in the LSC, and in

TABLE 2 Unique genes of the *Prangos fedtschenkoi* plastome

Group of gene	Name of gene	Number of genes
RNA genes		
Ribosomal RNAs	<i>rrn4.5</i> (x2), <i>rrn5</i> (x2), <i>rrn16</i> (x2), <i>rrn23</i> (x2)	4
Transfer RNAs	<i>trnA</i> -UGC(x2) ^a , <i>trnC</i> -GCA, <i>trnD</i> -GUC, <i>trnE</i> -UUC, <i>trnF</i> -GAA, <i>trnM</i> -CAU, <i>trnG</i> -GCC, <i>trnG</i> -GCC ^a , <i>trnH</i> -GUG, <i>trnI</i> -CAU (x2), <i>trnI</i> -GAU (x2) ^a , <i>trnK</i> -UUU ^a , <i>trnL</i> -CAA (x2), <i>trnL</i> -UAA ^a , <i>trnL</i> -UAG, <i>trnM</i> -CAU, <i>trnN</i> -GUU (x2), <i>trnP</i> -UGG, <i>trnQ</i> -UUG, <i>trnR</i> -ACG (x2), <i>trnR</i> -UCU, <i>trnS</i> -GCU, <i>trnS</i> -GGA, <i>trnS</i> -UGA, <i>trnT</i> -GGU, <i>trnT</i> -UGU, <i>trnV</i> -GAC(x2), <i>trnV</i> -UAC ^a , <i>trnW</i> -CCA, <i>trnY</i> -GUA	30
Protein genes		
Photosynthesis		
Photosystem I	<i>psaA</i> , <i>psaB</i> , <i>psaC</i> , <i>psal</i> , <i>psaJ</i> , <i>ycf3</i> ^b	5
Photosystem II	<i>psbA</i> , <i>psbB</i> , <i>psbC</i> , <i>psbD</i> , <i>psbE</i> , <i>psbF</i> , <i>psbH</i> , <i>psbl</i> , <i>psbJ</i> , <i>psbK</i> , <i>psbL</i> , <i>psbM</i> , <i>psbN</i> , <i>psbT</i> , <i>psbZ</i>	15
Cytochrome b/f complex	<i>petA</i> , <i>petB</i> ^a , <i>petD</i> ^a , <i>petG</i> , <i>petL</i> , <i>petN</i>	6
NADH-dehydrogenase	<i>ndhA</i> ^a , <i>ndhB</i> (x2) ^a , <i>ndhC</i> , <i>ndhD</i> , <i>ndhE</i> , <i>ndhF</i> , <i>ndhG</i> , <i>ndhH</i> , <i>ndhI</i> , <i>ndhJ</i> , <i>ndhK</i>	11
ATP synthase	<i>atpA</i> , <i>atpB</i> , <i>atpE</i> , <i>atpF</i> ^a , <i>atpH</i> , <i>atpI</i>	6
Large subunit of Rubisco	<i>rbcl</i>	1
ATP-dependent protease	<i>clp</i> ^b	1
Envelope membrane protein	<i>cemA</i>	1
Ribosomal proteins		
Large units	<i>rpl2</i> (x2, part) ^a , <i>rpl14</i> , <i>rpl16</i> ^a , <i>rpl20</i> , <i>rpl22</i> , <i>rpl23</i> (x2), <i>rpl32</i> , <i>rpl33</i> , <i>rpl36</i>	9
Small units	<i>rps2</i> , <i>rps3</i> , <i>rps4</i> , <i>rps7</i> (x2), <i>rps8</i> , <i>rps11</i> , <i>rps12</i> ^a (x2), <i>rps14</i> , <i>rps15</i> , <i>rps16</i> ^a , <i>rps18</i> , <i>rps19</i>	12
Transcription/translation		
DNA-dependent RNA polymerase	<i>rpoA</i> , <i>rpoB</i> , <i>rpoC1</i> ^a , <i>rpoC2</i>	4
Miscellaneous proteins	<i>accD</i> , <i>ccsA</i> , <i>infA</i> , <i>matK</i>	4/75
Hypothetical proteins and Conserved reading frame	<i>ycf1</i> (x2, part), <i>ycf2</i> (x2), <i>ycf4</i>	3/4
Total number of unique genes		113

^aGenes containing one intron. ^bGenes containing two introns.

ndhF/rpl32 (20 bp) in the SSC. Longer indel events were observed in *rps16/trnQ-UUG* (69 bp) and *accD/psal* (57 and 11 bp) in the LSC, and in *ndhE/ndhG* (50 bp) in the SSC.

3.6 | Intron region divergence

Detailed sequence comparisons among intron regions in *P. fedtschenkoi* and *P. lipskyi* are provided in Supporting Information Table S5. The most conserved intron, showing no variation, is that within 3'*rps12* of the IR. Intron diversity levels varied from identity to 7.4% (from identity to 0.9% in the IR, from 0.3% to 7.4% in the LSC, and

was 0.4% [*ndhA*] in the SSC). The most divergent intron regions were those in genes *trnL-UAA* (7.4%), *rpl16* (2.7%), *ycf3* (2.3%), and *clpP* (2.3%) in the LSC and *ndhB* (0.9%) in the IR. The length of all intron indel events totaled 118 bp. The majority of indels were less than 20 bp. One longer sized indel was located in the *trnL-UAA* intron (32 bp).

3.7 | Polymorphic sites

Data on polymorphic sites within the two *Prangos* plastomes are presented in Supporting Information Tables S3, S4 and S5. The total

TABLE 3 Comparison of nucleotide composition between *Prangos fedtschenkoi* and *Prangos lipskyi* plastomes

	Nucleotide composition of whole genome (%)						AT content (%)			AT content in codon regions (%)		
	T	C	A	G	AT	GC	LSC	IR	SSC	1	2	3
<i>P. fedtschenkoi</i>	31.4	19.2	30.8	18.5	62.3	37.7	64.0	57.1	68.7	61.6	61.9	63.0
<i>P. lipskyi</i>	31.4	19.2	30.8	18.5	62.3	37.7	64.0	57.1	68.7	62.9	62.9	61.7

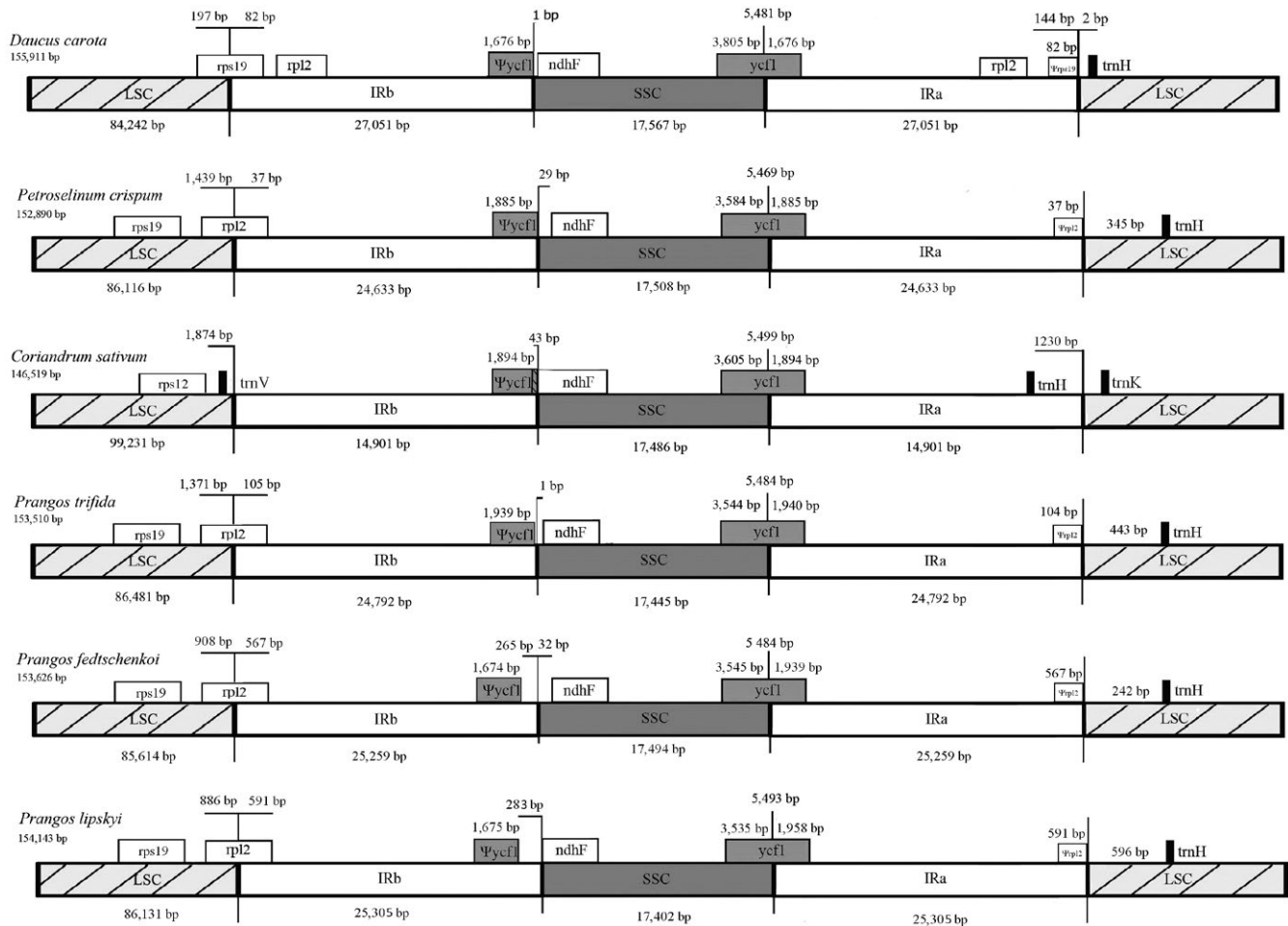


FIGURE 4 Schematic comparison of the large single-copy (LSC), inverted repeat (IRa, IRb), and small single-copy (SSC) regions among six Apiaceae plastid genomes: *Daucus carota*, *Petroselinum crispum*, *Coriandrum sativum*, *Prangos trifida*, *Prangos fedtschenkoi*, and *Prangos lipskyi*. Genes overlapping or flanking IR single-copy junctions are identified (their sizes do not correlate with their actual length). Pseudogenes are denoted by Ψ (i.e., Ψycf1, Ψrps19, Ψrpl2)

number of polymorphic sites in the protein-coding regions was 121, including seven in the IR, 68 in the LSC, and 46 in the SSC. Of these, 51.2% are Ts and 48.8% are Tv. The total number of the polymorphic sites in the intergenic spacer regions was 145, including five in IR, 124 in LSC, and 16 in SSC. The total number of the polymorphic sites in all introns is 53, including nine in the IR, 42 in the LSC, and two in the SSC. Overall, the total number of polymorphic sites in all regions of the plastome is 319, including 21 in the IR, 234 in LSC, and 64 in SSC (representing 43.8% Ts and 56.1% Tv).

3.8 | Development of *Prangos*-specific primers

In pairwise comparisons of sequences *P. fedtschenkoi* and *P. lipskyi* the *accD/psal* IGS region was deemed most variable, with a level of nucleotide divergence of 21.1%. This region also includes five indel events, of 6, 11, 11, 34, and 57 bp (Supporting Information Table S4). With these high levels of nucleotide and length variation, the region was considered further for future use in low-level taxonomic and population studies. Of the three pairs of *Prangos*-specific primers considered (Table 4), the *accD/psal_3F* and *accD/psal_3R* primer pair

produced the strongest, single product in PCR amplifications of 46 additional accessions belonging to 16 *Prangos* species from Middle Asia, including those from geographically remote populations (Mustafina et al., unpublished data). Comparisons of protein-coding genes, intergenic spacer regions, and introns among *P. fedtschenkoi*, *P. lipskyi*, and *P. trifida* plastomes provided evidence of high sequence variability within *accD/psal* (16.5%), *rps16/trnQ-UUG* (16.5%), and *psbZ/trnG-GCC* (11.4%) IGS regions (Supporting Information Tables S6, S7 and S8). In the *P. trifida* plastome, the *petA/psbJ* IGS region was also highly variable (26.9%); however, the high variability of this region is a result of a unique 315 bp insertion.

3.9 | Insertions of novel DNA

Unique, noncoding DNA of 242 bp in *P. fedtschenkoi*, 596 bp in *P. lipskyi*, and 443 bp in *P. trifida* occurs between gene *trnH-GUG* (LSC) and the IRa/LSC junction (Figure 4). BLAST searches of these novel DNA sequences did not show any significant similarity to any other known plastid DNA sequence. Instead, the BLAST results revealed small, significant matches to mitochondrial DNA

Primer	Sequence	Binding region	Tan (°C)	GC (%)
accD/psal_1F	5'-TGGGAGATATCATTATTGCC-3'	59,651–59,670	54.3	40.0
accD/psal_1R	5'-GCAATGGCTTCTTTATTCT-3'	60,544–60,563	52.3	35.0
accD/psal_2F	5'-CGCTTTCTTTCTTTGAATC-3'	59,844–59,863	54.3	40.0
accD/psal_2R	5'-GTAGGCTTAGTATTTCCGG-3'	60,520–60,538	55.2	47.3
accD/psal_3F	5'-TTAATCGTACCACGTAATCC-3'	59,788–59,807	54.3	40.0
accD/psal_3R	5'-GCAATGGCTTCTTTATTCT-3'	60,544–60,563	52.3	35.0

Note. Binding region: accD/psal spacer region within Mauve alignment of *P. fedtschenkoi* and *P. lipskyi* plastomes; Tan: annealing temperature.

(specifically, to an IGS region adjacent to mitochondrial gene cytochrome b [*cob*] in *D. carota*). The results of a BLAST search of the 242 bp novel insert in *P. fedtschenkoi* are shown in Table 5. Both *P. fedtschenkoi* and *P. lipskyi* plastomes showed a 100 bp match having 86% sequence similarity to this mtDNA region. The *P. trifida* plastome showed a 102 bp match having 88.2% sequence similarity to this same region. Alignment of these *Prangos* novel DNA fragments showing similarities to *D. carota* mtDNA (i.e., positions 147–242 bp in *P. fedtschenkoi*, 501–596 bp in *P. lipskyi*, and 348–443 bp in *P. trifida*) with a 101 bp fragment of the IGS region adjacent to *D. carota* mitochondrial gene *cob* showed 85 identical sites across the four sequences compared (Supporting Information Figure S1).

3.10 | Repeat analysis

Repeat analysis identified 49 short motif SSRs in *P. fedtschenkoi* and 48 SSRs in *P. lipskyi* plastomes having repeat lengths ≥ 10 bp (Table 6). In *P. fedtschenkoi*, 43 of these were mononucleotides and six were AT dinucleotides, and in *P. lipskyi* 42 were mononucleotides and six were AT dinucleotides. Repeat lengths of A and T classes of mononucleotides were 10–17 bp and 10–14 bp, respectively. In *P. lipskyi*, a poly-C nucleotide repeat of 11 bp was found in the *trnA*-UGC intron (IRa) and a poly-G nucleotide repeat of 11 bp occurred in the *trnA*-UGC intron (IRb). AT dinucleotide repeats with lengths of 10–14 bp were found in *P. fedtschenkoi* and *P. lipskyi*, with a repeat number of 6.

The plastomes of *P. fedtschenkoi* and *P. lipskyi* each had seven large tandem repeats, with repeat size ≥ 30 bp and sequence identity $>90\%$ (Tables 7 and 8). Repeat units are repeated 2–4 times, with the majority occurring in the LSC region. Among the ten uniquely

occurring large tandem repeats identified in *Prangos*, most were palindromic dispersed. Four tandem repeats have identical features and are located within the same genomic regions in both plastomes; two of these occur in IGS regions and two others occur in coding sequence regions. A survey for these 10 *Prangos* large tandem repeats in 27 other accessions of Apiaceae and Araliaceae resulted in three (i.e., repeat nos. 4, 5, and 6 in *P. fedtschenkoi* [Table 7] and repeat nos. 3, 5, and 6 in *P. lipskyi* [Table 8]) being widely distributed (Supporting Information Table S9); one of these occurs in the *petN/psbM* IGS region, and the other two occur in coding sequence regions. An additional repeat occurring in both *psbM/trnD*-GUC and *trnE-UUC/trnT*-GGU IGS regions was shared by *P. fedtschenkoi*, *P. lipskyi*, *P. trifida*, and *Crithmum maritimum*. *P. fedtschenkoi* and *P. trifida* share a repeat within the *trnL*-UAA intron.

3.11 | Stem-loop hairpin structures

Each *Prangos* species possessed one of two small inversions, of either 5 or 9 bp (Figure 5). The first occurs in the *P. lipskyi ndhB* intron in the IR, thus there are two copies of this inversion present; each copy has repeat sequences of 18 bp at its ends. The second inversion occurs in the *P. fedtschenkoi petB* intron and has repeat sequences of 19 bp at its ends. All regions form distinct stem-loop hairpin structures, with the sequences of the loop regions flip-flopped in each *Prangos* species.

Results of a survey for the presence of these two small inversions and their associated flanking inverted repeats in other Apiaceae and Araliaceae accessions are presented in Supporting Information Tables S10 and S11. The 5 bp inversion of *P. lipskyi*, represented by sequence ACCCA, is also detected in *P. trifida*, while similar sequences (e.g., ACTCA) are present in *Anthriscus*, *Carum*, *Coriandrum*, and Araliaceae.

Accession	Species	Location	Length of match (bp)	Percent similarity
AY007816	<i>Daucus carota</i>	cytochrome b (<i>cob</i>); ORF25 (orf25)	100	86
AY007821	<i>Daucus carota</i>	ATPase8 (<i>ATP8</i>); cytochrome b (<i>cob</i>)	100	86
JQ248574	<i>Daucus carota</i> subsp. <i>sativus</i>	cytochrome b (<i>cob</i>); ORF25 (orf25)	100	86

TABLE 5 Results of BLAST searches of the nucleotide database (as of 3 October 2018) querying the 242 bp novel insertion sequence in *Prangos fedtschenkoi*. Only hits with lengths ≥ 100 bp and a percent similarity at least 86% are shown

TABLE 6 Features of short sequence repeats (SSRs) with repeat length ≥ 10 bp in *Prangos fedtschenkoi* and *Prangos lipskyi* plastomes. Repeats include both copies of the IR

Unit size	Repeat length, bp	Repeat number	
		<i>P. fedtschenkoi</i>	<i>P. lipskyi</i>
A	10–17	20	17
T	10–14	23	23
C	11	–	1
G	11	–	1
AT	10–14	6	6

In contrast, the sequence TGGGT of *P. fedtschenkoi*, or TGAGT with 1 bp change, occurs in 20 other accessions, including *Bupleurum*, the most basal taxon within Apiaceae subfamily Apioideae. The 9 bp inversion of *P. fedtschenkoi*, with sequence ACGACAAGA, also occurs in *Anethum graveolens*, *Carum carvi*, and *Foeniculum vulgare*, and with 1 bp change (ACGAAAAGA) in *Hansenia* spp. and *Tiedemannia filiformis*. In contrast, the loop sequence of *P. lipskyi* and *P. trifida*, TCTTGTCGT, and similar sequence TCTTTTCGT, occurs in multiple other taxa. While the congeners of *Angelica* (4 spp.), *Bupleurum* (2 spp.), and *Hansenia* (4 spp.) are each represented by the same loop sequence for each of the two inversions, these loop regions are actually quite variable, with multiple motifs represented.

4 | DISCUSSION

The results of this study have multiple, important implications. First, entire plastome sequences were determined for two endemic species of *Prangos* from Middle Asia: *P. fedtschenkoi*, a valuable resource plant; and *P. lipskyi*, a narrow endemic species of the Western Tien Shan Mountains. *Prangos* is treated in the *Cachrys* clade of Apiaceae based on molecular phylogenetic study (Downie, Spalik, Katz-Downie, & Reduron, 2010). A comparative analysis of these plastomes identified new loci, specifically the *accD/psal* IGS region, that offer levels of variation higher than those regions commonly employed in molecular phylogenetic studies of the family. Second, the *Prangos*-specific primers developed and surveyed for the *accD/psal* IGS region will be useful for future phylogeographic and population-level studies. Similarly, the unique short repeated regions described for each plastome will be helpful to understand the genetic potential of each population, which is important in the development of ex situ conservation strategies. Third, the incorporation of novel DNA at the LSC/IRa junction suggestive of mtDNA adds to the growing body of evidence showing introgression of mtDNA into the angiosperm plastome. Fourth, the presence of two different inversions in sister species, each in one species but not the other, provides another marker that can be surveyed in a greater sampling of the group. However, similar inversions can occur in parallel in more distantly related taxa, thus reducing their phylogenetic utility.

4.1 | Identification of variable loci

Direct comparisons of 83 protein-coding regions, 113 IGS regions, and 18 introns from three complete *Prangos* plastomes resulted in the identification of seven loci having high nucleotide and length variability: *accD/psal*, *psbZ/trnG*-GCC, *rps16/trnQ*-UUG, *rps14/psaB*, *ndhE/ndhG*, *petA/psbJ*, and the *trnL*-UAA intron. Shaw et al. (2005) have discussed that the ideal plastid region for phylogenetic investigation should combine high variability with fragments of approximately 700–1,500 bp in size that will permit ease of PCR amplification. Using these criteria, the regions most suitable for phylogenetic inference and population-level analyses in *Prangos* are *accD/psal*, *psbZ/trnG*, *rps16/trnQ*, and the *trnL* intron. Within Apiaceae, the plastid loci most frequently employed in molecular phylogenetic studies to date have included the *rpl16*, *rpoC1*, and *rps16* introns and the *trnH/psbA*, *trnD/trnT*, and *trnF/trnL/trnT* IGS regions (Downie & Jansen, 2015). In *Prangos*, *accD/psal*, and *psbZ/trnG* are more variable and better suited for intraspecific and population-level studies than these other loci. Indeed, the *accD/psal* IGS region was listed among the most variable loci identified from the comparative analysis of five plastid genomes of Apiaceae (Downie & Jansen, 2015). In studies of other families, these aforementioned loci have previously been considered useful in resolving relationships. The *accD/psal* region was used in studies of *Pyrus* (Rosaceae), *Dendrochilum* (Orchidaceae), and *Castanea* (Fagaceae; Shaw, Lickey, Shilling, & Small, 2007). The *psbZ/trnG* region was found most divergent in the analysis of 13 *Gossypium* plastid genomes (Malvaceae; Xu et al., 2012). The *rps16/trnQ* region included many phylogenetically informative indels in *Carphephorus* (Asteraceae), *Gratiola* (Plantaginaceae), *Prunus* (Rosaceae), and *Trillium* (Melanthiaceae; Shaw et al., 2007), and the *trnL* intron resolved relationships in Gentianaceae (Yuan et al., 2003).

4.2 | Novel DNA

The three *Prangos* species examined herein have each incorporated novel DNA into the LSC region adjacent to the LSC/IRa junction, with BLAST searches revealing 100–102 bp matches having $\geq 86\%$ sequence identity to an IGS region adjacent to *Daucus* mitochondrial gene *cob*. The introgression of mtDNA into the plastid genome is not without precedence in Apiaceae, for previous studies have also reported the presence of mtDNA in the plastomes of some Apiaceae species (Downie & Jansen, 2015; Goremykin, Salamini, Velasco, & Viola, 2009; Iorizzo, Grzebelus, et al., 2012a; Iorizzo, Senalik, et al., 2012b; Peery, Downie, Jansen, & Raubeson, 2011). In *P. crispum*, for example, 345 bp of novel noncoding DNA has been incorporated into the LSC region adjacent to the LSC/IRb boundary and BLAST searches querying this insert resulted in hits of a 122 bp region to angiosperm mtDNA (Downie & Jansen, 2015). A comparison of IRa/LSC boundaries from 34 species of Apiaceae revealed that those species with their LSC/IRb boundaries within *rpl2* have insertions suggestive of mtDNA, ranging in size from 40 to 447 bp, between IRa and *trnH*-GUC (Peery, 2015).

TABLE 7 Features of large tandem repeat loci (≥ 30 bp) in the *Prangos fedtschenkoi* plastome

Repeat #	Repeat unit	Repeat number	Size (bp)	Repeat type	Location
1	ATTGACGAGCTACAGCACTCGCACCCTATTAAACGCAACTAAAAGAATTATT	2	100	Forward	IGS (<i>ndhE</i> - <i>ndhG</i>)
2	AAAAGGAAAAGATGATGGATGTAATCTTATTGAATCTGTCTG	2	78	Palindromic dispersed	IGS (<i>psbM</i> - <i>trnD</i> -GUC; <i>trnE</i> -UUC - <i>trnT</i> -GGU)
3	ATACGTATGTATATAC	2	32	Forward	Intron (<i>trnL</i> -UAA)
4	GAGGATATTGACTAGTGAGGATATTGATGCTAGTGA	4	76 (2x)	Palindromic dispersed	CDS (<i>ycf2</i>)
5	ACGAAAAGAGAGGGATTGGAACCCCTCGGTA	2	60	Palindromic dispersed	CDS (<i>trnS</i> -GCU; <i>trnS</i> -GGA)
6	GTAAGAAAAGAAATAT	2	30	Palindromic	IGS (<i>petN</i> - <i>psbM</i>)
7	ATATTCATAAAGTAATGATA	2	40	Forward	IGS (<i>ndhF</i> - <i>rp132</i>)

Notes. The bold characters represent the shared tandem repeats with *P. lipskyi* chloroplast genome. CDS: coding DNA sequence; IGS: intergenic sequence.

TABLE 8 Features of large tandem repeat loci (≥ 30 bp) in the *Prangos lipskyi* plastome

Repeat #	Repeat unit	Repeat number	Size (bp)	Repeat type	Location
1	ATCATATAAATACAAAAGATTATATTCATAATTCATTATCAT	2	40	Palindromic dispersed	IGS (<i>psbZ</i> - <i>trnG</i> -GCC; <i>psaC</i> - <i>ndhE</i>)
2	AAAAGGAAAAGATGATGGATGTAATCTTATTGAATCTGTCTG	2	78	Palindromic dispersed	IGS (<i>psbM</i> - <i>trnD</i> -GUC; <i>trnE</i> -UUC - <i>trnT</i> -GGU)
3	GAGGATATTGACTAGTGAGGATATTGATGCTAGTGA	4	76 (2x)	Palindromic dispersed	CDS (<i>ycf2</i>)
4	ATGATATATGCTTTTGACCTTCTATACTCACTTAG	2	72	Forward dispersed	IGS (<i>accD</i> - <i>psaI</i>)
5	ACGAAAAGAGAGGGATTGGAACCCCTCGGTA	2	30	Palindromic dispersed	CDS (<i>trnS</i> -GCU; <i>trnS</i> -GGA)
6	GTAAGAAAAGAAATAT	2	30	Palindromic	IGS (<i>petN</i> - <i>psbM</i>)
7	ATGATAAAAAATGGACATTATGA	2	46	Forward	IGS (<i>petA</i> - <i>psbJ</i>)

Notes. The bold characters represent the shared tandem repeats with *P. fedtschenkoi* chloroplast genome. CDS: coding DNA sequence; IGS: intergenic sequence.

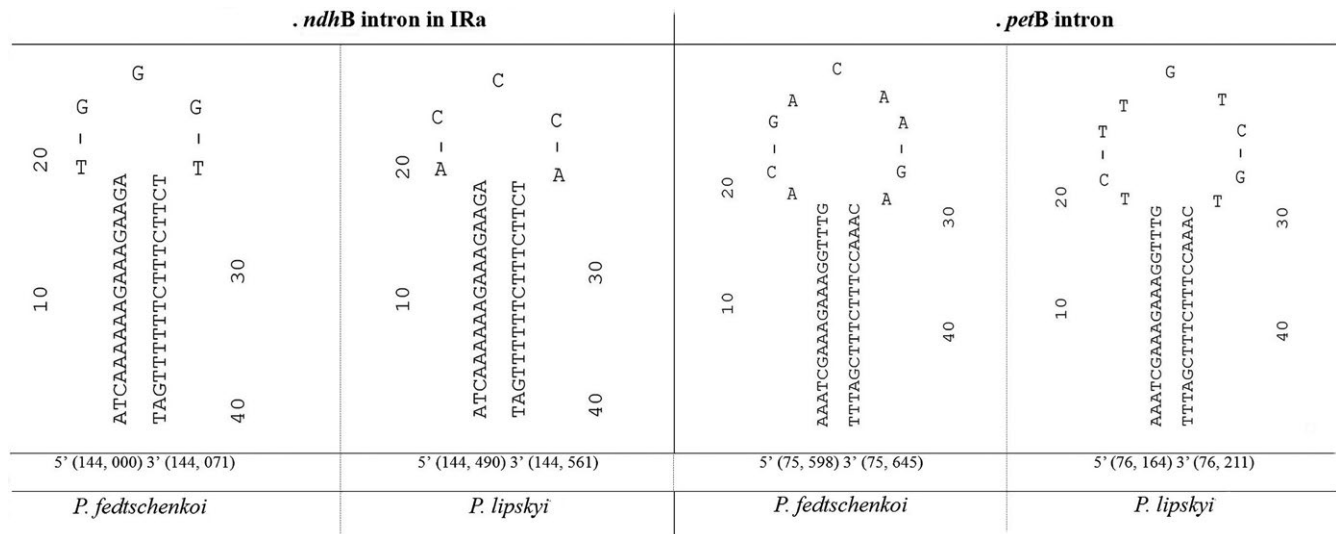


FIGURE 5 Short inversion mutations and associated secondary structure between *Prangos fedtschenkoi* and *Prangos lipskyi* plastomes. Inversions are located within intron regions of *ndhB* (IRb, IRa) and *petB*. Free energy of the secondary structures: (a) dG = -16.42 kcal/mol; dG = -16.72 kcal/mol; (b) dG = -19.37 kcal/mol, dG = -19.98 kcal/mol

Nevertheless, we interpret these results cautiously, for in the absence of complete mitochondrial genomes from *Prangos* and other related Apiaceae, the mitochondrial provenance of this novel DNA in *Prangos* is not absolute.

4.3 | Utility of plastome SSRs and tandem repeat units

In plastomes of *D. carota* and other Apiaceae species (Peery, 2015; Ruhlman et al., 2006), the number of SSRs reported previously ranged from 44 to 55. In *Prangos* plastomes, the number of mono- and dinucleotide repeats is similar, varying between 48 (*P. lipskyi*) and 49 (*P. fedtschenkoi*). SSRs are widely used molecular markers in plant population and phylogenetic studies due to their high mutation rates and high levels of polymorphism among individuals of a population. Another advantage is their ease of use and low-cost detection using PCR (Hoshino, Bravo, Nobile, & Morelli, 2012; Wang, Barkley, & Jenkins, 2009). To date, the availability of complete plastome sequences in public databases has reduced the economic costs of obtaining these data and increased their use in inter- and intraspecific phylogenetic research, population study, and conservation activities (Wang et al., 2009). In addition, taxon-specific primers, such as those generated herein for the *Prangos accD/psal* highly variable IGS region, will be useful to assess genetic differentiation among and within its populations.

Short sequence repeats that arise by chance in DNA sequences along with mutational changes can be presumably expanded by slipped-strand mispairing events into longer tandem repeats (Levinson & Gutman, 1987). Transposon-mediated insertions along with replication slippage were suggested as being responsible for generating direct and inverted long repeats (Hoshino et al., 2012; Palmer, 1991). Ten long tandem repeats were identified in *Prangos*,

four of which are located within the same genomic regions and identical in both plastomes. The three others, namely two in *ycf2* and one in *trnS*, were reported previously by Ruhlman et al. (2006) for *D. carota*. These repeats plus one palindromic tandem repeat in the *petN/psbM* IGS region occur in the same locations and are shared by most of the Apiaceae and Araliaceae species surveyed.

4.4 | Inversions

The shared possession of the same inversion is usually viewed as reliable evidence of common ancestry, although the presence of base substitutions within the region may cause misinterpretation (Kim & Lee, 2005). Large plastid inversions, as examples, have been suggested to be highly reliable phylogenetic markers (Graham, Reeves, Burns, & Olmstead, 2000; Jansen & Palmer, 1987; Raubeson & Jansen, 1992a, 1992b). In contrast, the utility of small inversions in phylogenetic inference is poor, because they are reported as often being homoplastic (Graham et al., 2000). In this study, we identified two small inversions in *Prangos*, both occurring in loop regions of *ndhB* and *petB* introns that are homoplastic within Apiaceae.

Comparisons of sequences comprising these 5- and 9-bp inversions in *ndhB* and *petB* introns with 23 other species (16 genera) of Apiaceae (representing a diversity of tribes and other major clades within subfamily Apioideae; Downie et al., 2010) and 2 genera of Araliaceae revealed both identical direct and inverted motifs of these two small regions. In many additional taxa, 1–2 bp substitutions within these regions were also apparent. Both flip-flop oriented forms of these sequences were revealed in the three *Prangos* species examined herein, whereas congeners of other genera shared identical forms of each sequence. The flip-flop orientation of loop sequences in the plastid *trnL/trnF* region was also described for nine *Jasminum* species (Oleaceae; Kim & Lee, 2005).

Two different accessions of *J. elegans* and single accessions of each of the other eight species were used in phylogenetic analyses. The accessions of *J. elegans* had different orientations of a loop sequence, with one orientation occurring in five other *Jasminum* species and the other orientation found in the remaining species. These sequence data may confound phylogenetic relationships unless one is cognizant that an inversion has taken place. Our future studies will survey for the presence of these inverted regions in other *Prangos* species.

In summary, the complete *Prangos* plastome sequences reported herein enhance the genomic information available for *Prangos* and will contribute to further studies of germplasm diversity, phylogeny, and phylogeography of these ecologically and economically important species. These data provide a valuable source of markers for future research at low taxonomic level and population genetics with further implementation in ex situ conservation strategies and restoration/regeneration programs.

ACKNOWLEDGMENTS

This research was conducted within the framework of the "Central Asia Green Road" project between the Korea National Arboretum (KNA-1-1-17, 15-2) of the Korea Forest Service (Republic of Korea) and the Institute of Botany of the Academy of Sciences (Republic of Uzbekistan). Funding for the project "Systematics of dicotyledonous plants of the natural flora of Uzbekistan" is also acknowledged (VA-FA-F5010; Republic of Uzbekistan). We thank Deborah S. Katz-Downie for valuable comments on the manuscript.

CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTION

FUM and KCh conceived and designated the study; DKYi and FUM conducted the molecular and computational analyses; SRD and FUM also performed analyses, interpreted the results, and wrote the manuscript; and ChHSh and KShT provided the necessary facilities, equipment, and chemicals, and helped with the field research.

DATA ACCESSIBILITY

Complete plastid genomes: Genbank accessions KY652265 (*P. fedtschenkoi*) and KY652266 (*P. lipskyi*).

ORCID

Feruz U. Mustafina  <https://orcid.org/0000-0003-4195-3353>

Dong-Keun Yi  <https://orcid.org/0000-0002-3988-323X>

REFERENCES

- Banasiak, L., Piwczynski, M., Ulinski, T., Downie, S. R., Watson, M. F., Shaky, B., & Spalik, K. (2013). Dispersal patterns in space and time: A case study of Apiaceae subfamily Apioideae. *Journal of Biogeography*, 40(7), 1324–1335. <https://doi.org/10.1111/jbi.12071>
- Benson, G. (1999). Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Research*, 27(2), 573–580. <https://doi.org/10.1093/nar/27.2.573>
- CACILM (2006). *Central Asian Countries Initiative for Land Management (CACILM) Multicountry Partnership Framework Support Project*. Retrieved from <https://www.adb.org/projects> (January, 2017).
- Cui, L., Veeraghavan, N., Richter, A., Wall, K., Jansen, R. K., Leebens-Mack, J., ... dePamphilis, C. W. (2006). ChloroplastDB: The chloroplast genome database. *Nucleic Acids Research*, 34(Database, issue), D692–D696. <https://doi.org/10.1093/nar/gkj055>
- Darling, A. E., Mau, B., & Perna, N. T. (2010). ProgressiveMauve: Multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE*, 5(6), e11147. <https://doi.org/10.1371/journal.pone.0011147>
- Downie, S. R., & Jansen, R. K. (2015). A comparative analysis of whole plastid genomes from the Apiales: Expansion and contraction of the inverted repeat, mitochondrial to plastid transfer of DNA, and identification of highly divergent noncoding regions. *Systematic Botany*, 40(1), 336–351. <https://doi.org/10.1600/036364415X686620>
- Downie, S. R., Spalik, K., Katz-Downie, D. S., & Reduron, J. P. (2010). Major clades within Apiaceae subfamily Apioideae as inferred by phylogenetic analysis of nrDNA ITS sequences. *Plant Diversity and Evolution*, 128(1–2), 111–136. <https://doi.org/10.1127/1869-6155/2010/0128-0005>
- Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5), 1792–1797. <https://doi.org/10.1093/nar/gkh340>
- Goremykin, V. V., Salamini, F., Velasco, F., & Viola, R. (2009). Mitochondrial DNA of *Vitis vinifera* and the issue of rampant horizontal gene transfer. *Molecular Biology and Evolution*, 26(1), 99–110. <https://doi.org/10.1093/molbev/msn226>
- Graham, S. W., Reeves, P. A., Burns, A. C. E., & Olmstead, R. (2000). Microstructural changes in noncoding chloroplast DNA: Interpretation, evolution, and utility of indels and inversions in basal angiosperm phylogenetic inference. *International Journal of Plant Sciences*, 161(S6), S83–S96. <https://doi.org/10.1086/317583>
- Hoshino, A. A., Bravo, J. P., Nobile, P. M., & Morelli, K. A. (2012). Microsatellites as tools for genetic diversity analysis, genetic diversity in microorganisms. In M. Caliskan (Ed.), *Genetic diversity in microorganisms* (pp. 149–170). Rijeka, Croatia: InTech. <https://doi.org/10.5772/35363>
- Iorizzo, M., Grzebelus, D., Senalik, D., Szklarczyk, M., Spooner, D., & Simon, P. (2012a). Against the traffic: The first evidence for mitochondrial DNA transfer into the plastid genome. *Mobile Genetic Elements*, 2(6), 261–266. <https://doi.org/10.4161/mge.23088>
- Iorizzo, M., Senalik, D., Szklarczyk, M., Grzebelus, D., Spooner, D., & Simon, P. (2012b). De novo assembly of the carrot mitochondrial genome using next generation sequencing of whole genomic DNA provides first evidence of DNA transfer into an angiosperm plastid genome. *BMC Plant Biology*, 12(1), 61. <https://doi.org/10.1186/1471-2229-12-61>
- Jansen, R. K., & Palmer, J. D. (1987). A chloroplast DNA inversion marks an ancient evolutionary split in the sunflower family (Asteraceae). *Proceedings of the National Academy of Sciences of the United States of America*, 84(16), 5818–5822. <https://doi.org/10.1073/pnas.84.16.5818>
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., ... Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28(12), 1647–1649. <https://doi.org/10.1093/bioinformatics/bts199>

- Kim, K. J., & Lee, H. L. (2005). Widespread occurrence of small inversions in the chloroplast genomes of land plants. *Molecules and Cells*, 19(1), 104–113.
- Kurtz, S., Choudhuri, J. V., Ohlebusch, E., Schleiermacher, C., Stoye, J., & Giegerich, R. (2001). REPuter: The manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Research*, 29(22), 4633–4642. <https://doi.org/10.1093/nar/29.22.4633>
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), 357–359. <https://doi.org/10.1038/nmeth.1923>
- Levinson, G., & Gutman, G. A. (1987). Slipped-Strand mispairing: A major mechanism for DNA sequence evolution. *Molecular Biology and Evolution*, 4(3), 203–221.
- Librado, P., & Rozas, J. (2009). DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, 25(11), 1451–1452. <https://doi.org/10.1093/bioinformatics/btp187>
- Limin, H., & Zhang, D. (2012). Engaging with land users: The first steps on a long road. In V. R. Squires (Ed.), *Rangeland stewardship in Central Asia: Balancing improved livelihoods, biodiversity conservation and land protection* (pp. 333–356). Dordrecht, The Netherlands: Springer. https://doi.org/10.1007/978-94-007-5367-9_14
- Lohse, M., Drechsel, O., Kahlau, S., & Bock, R. (2013). OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Research*, 41(Web Server issue), W575–W581. <https://doi.org/10.1093/nar/gkt289>
- Lyskov, D. F., Degtjareva, G. V., Samigullin, T. H., & Pimenov, M. G. (2017). The revision of *Prangos* subsections *Koelzella* and *Fedtschenkoana* (Apiaceae) with some notes to phylogeny and biogeography of the genus: Molecular and morphological evidences. *Plant Systematics and Evolution*, 303(7), 815–826. <https://doi.org/10.1007/s00606-017-1412-0>
- Mayer, C., Leese, F., & Tollrian, R. (2010). Genome-wide analysis of tandem repeats in *Daphnia pulex* – A comparative approach. *BMC Genomics*, 11(1), 277. <https://doi.org/10.1186/1471-2164-11-277>
- Mustafina, F. U., Kim, E. H., Son, S. W., Turginov, O. T., Chang, K. S., & Choi, K. (2017). Assessment of genetic diversity of *Prangos fedtschenkoi* (Apiaceae) and its conservation status based on ISSR markers. *Korean Journal of Plant Taxonomy*, 47(1), 11–22. <https://doi.org/10.11110/kjpt.2017.47.1.11>
- Palmer, J. D. (1991). Plastid chromosomes: Structure and evolution. In L. Bogorad, & I. K. Vasil (Eds.), *Cell culture and somatic cell genetics of plants* (Vol. 7A, pp. 5–53). San Diego, CA: Academic Press.
- Peery, R. M. (2015). *Understanding angiosperm genome interaction and evolution: From sacred lotus (Nelumbo nucifera) and the carrot family (Apiaceae)*. PhD thesis, University of Illinois at Urbana-Champaign, USA. Retrieved from <http://hdl.handle.net/2142/78479> (June, 2017).
- Peery, R., Downie, S. R., Jansen, R. K., & Raubeson, L. A. (2011). Apiaceae organellar genomes. In *Proceedings of the 18th International Botanical Congress* (p. 481). Melbourne, Australia.
- Raubeson, L. A., & Jansen, R. K. (1992a). A rare chloroplast-DNA structural mutation is shared by all conifers. *Biochemical Systematics and Ecology*, 20(1), 17–24. [https://doi.org/10.1016/0305-1978\(92\)90067-N](https://doi.org/10.1016/0305-1978(92)90067-N)
- Raubeson, L. A., & Jansen, R. K. (1992b). Chloroplast DNA evidence on the ancient evolutionary split in vascular land plants. *Science*, 255(5052), 1697–1699. <https://doi.org/10.1126/science.255.5052.1697>
- Ruhlman, T., Lee, S.-B., Jansen, R. K., Hostetler, J. B., Tallon, L. J., Town, C. D., & Daniell, H. (2006). Complete plastid genome sequence of *Daucus carota*: Implications for biotechnology and phylogeny of angiosperms. *BMC Genomics*, 7, 222. <https://doi.org/10.1186/1471-2164-7-222>
- Shaw, J., Lickey, E. B., Beck, J. T., Farmer, S. B., Liu, W., Miller, J., ... Small, R. L. (2005). The tortoise and the hare II: Relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany*, 92(1), 142–166. <https://doi.org/10.3732/ajb.92.1.142>
- Shaw, J., Lickey, E. B., Shilling, E. E., & Small, R. L. (2007). Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: The tortoise and the hare III. *American Journal of Botany*, 94(3), 275–288. <https://doi.org/10.3732/ajb.94.3.275>
- Shishkin, B. K. (1950). *Prangos* Lindl. In B. K. Shishkin (Ed.), *Flora of the USSR* (Vol. 16, pp. 263–273). Moskva & Leningrad, U. S. S. R.: Izdatel'stvo Akademii Nauk SSSR.
- Tamura, K., Stecher, G., Peterson, D., Filipi, A., & Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, 30(12), 2725–2729. <https://doi.org/10.1093/molbev/mst197>
- Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22(22), 4673–4680. <https://doi.org/10.1093/nar/22.22.4673>
- Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3—New capabilities and interfaces. *Nucleic Acids Research*, 40(15), e115. <https://doi.org/10.1093/nar/gks596>
- Wang, M. L., Barkley, N. A., & Jenkins, T. M. (2009). Microsatellite markers in plants and insects. Part I: Applications of biotechnology. *Genes, Genomes and Genomics*, 3(1), 54–67.
- Wyman, S. K., Jansen, R. K., & Boore, J. L. (2004). Automatic annotation of organellar genomes with DOGMA. *Bioinformatics*, 20(17), 3252–3255. <https://doi.org/10.1093/bioinformatics/bth352>
- Xu, Q., Xiong, G., Li, P., He, F., Huang, Y. i., Wang, K., ... Hua, J. (2012). Analysis of complete nucleotide sequences of 12 *Gossypium* chloroplast genomes: Origin and evolution of allotetraploids. *PLoS ONE*, 8(5), e37128. <https://doi.org/10.1371/journal.pone.0037128>
- Yuan, Y. M., Wohlhauser, S., Moller, M., Chassot, P. H., Mansion, G., Grant, J., ... Klackenberg, J. (2003). Monophyly and relationships of the tribe *Exaceae* (Gentianaceae) inferred from nuclear ribosomal and chloroplast DNA sequences. *Molecular Phylogenetics and Evolution*, 28(3), 500–517. [https://doi.org/10.1016/s1055-7903\(03\)00068-x](https://doi.org/10.1016/s1055-7903(03)00068-x)
- Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*, 31(13), 3406–3415. <https://doi.org/10.1093/nar/gkg595>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Mustafina FU, Yi D-K, Choi K, Shin CH, Tojibaev KS, Downie SR. A comparative analysis of complete plastid genomes from *Prangos fedtschenkoi* and *Prangos lipskyi* (Apiaceae). *Ecol Evol*. 2019;9:364–377. <https://doi.org/10.1002/ece3.4753>