Turkish Journal of Agriculture and Forestry

Volume 48 | Number 5 Article 6

10-11-2024

Plastome analysis of cultivated apricots: genome structure, nucleotide diversity, and phylogeny

LUNGELO KHANYILE

KAHRAMAN GÜRCAN

TÜBİTAK

Turkish Journal of Agriculture and Forestry

http://journals.tubitak.gov.tr/agriculture/

Research Article

Turk J Agric For
(2024) 48: 692-702

doi:10.55730/1300-011X.3212

Plastome analysis of cultivated apricots: genome structure, nucleotide diversity, and phylogeny

Lungelo KHANYILE, Kahraman GÜRCAN*

Department of Agricultural Biotechnology, Genome and Stem Cell Center, Faculty of Agriculture, Erciyes University, Kayseri, Turkiye

Received: 02.09.2023 • **Accepted/Published Online:** 07.05.2024 • **Final Version:** 11.10.2024

Abstract: The genetic diversity and evolution of plastomes in *P. armeniaca* L. (common/cultivated apricot) have been poorly studied. Complete plastomes of 20 accessions and 10 F1 hybrids, identified as *P. armeniaca* L., were de novo assembled filtering from total DNA nt reads. These plastomes were subsequently analyzed with the *Prunus* genus accessions retrieved from the GenBank. The plastomes exhibited conservation of genomic structure, gene organization and order, with lengths ranging from 158,057 nt to 158,089 nt, displaying a narrow size variation. Except for "Zard" and its hybrids, the plastomes of apricot genotypes from different regions showed only three haplotypes, indicating narrow genetic diversity in cultivated *P. armeniaca* accessions. "Zard" and its hybrids exhibited the highest identity with *P. mandshurica* Maxim (Manchurian apricot), demonstrating the significance of plastome analysis for the accurate identification of apricots. Further, while the three apricot species, *P. armeniaca*, *P. mandshurica*, and *P. mume* Sieb. (Japanese apricot), were grouped as sister clades, *P. sibirica* L. (Siberian apricot) exhibited the highest nucleotide identity with *P. cerasifera* Ehrh. (Cherry plum), contrary to conventional morphological systematics of apricots. Additionally, the plastomes of hybrids obtained in this study supported maternal inheritance of the plastome. These results reveal the evolutionary relationship among apricots and will serve as a framework for future comparative studies on *Prunus* evolution.

Key words: *Prunus*, chloroplast, phylogenomic, high throughput sequencing, maternal inheritance

1. Introduction

The common apricot, *Prunus armeniaca* L., along with other economically important species such as almond (*Prunus dulcis* L.), peach (*Prunus persica* L.), plum (*Prunus domestica* L.), sweet (*Prunus avium* L.), and sour cherry (*Prunu cerasus* L.), is a species of the genus *Prunus* in the Rosaceae family. Additionally, many species, ranging from 77 (Rehder, 1954) to 150 (Sauer, 1993), including several species of minor importance as scions or rootstocks, are classified under the genus *Prunus*. Five other species—*P. brigantina* Vill. (Alpine plum), *P. holosericeae* Batal (Tibetan apricot), *P. armeniaca* L. (common apricot), *P. mandshurica* Maxim (Manchurian apricot), *P. sibirica* L. (Siberian apricot), *P. mume* Sieb. and Succ (Japanese apricot) —are also recognized as apricots within this genus. The basic chromosome number for apricots is X = 8, and they are all regular diploids (2n = 16). All apricots, including various species of the genus *Prunus*, can be intercrossed, which makes their classification confusing.

Currently, the common apricot is predominantly produced in Türkiye, followed by Uzbekistan, Iran, and Mediterranean countries. In particular, Türkiye is the leading producer of dried apricots, supplying approximately 67% of the global market.[1] However, it is believed that the apricots were introduced to Western Asia and then to Europe over the last couple of millennia (Ercisli, 2009). The center of origin of apricots is China and Central Asia (Zhebentyayeva et al., 2012). Indeed, recent research based on SSR markers and genome sequencing has revealed the highest levels of diversity in Central Asian and Chinese wild and cultivated apricots, confirming their origin in this region (Decroocq et al., 2016; Liu et al., 2019; Groppi et al., 2021).

Breeding programs in the USA and Europe focus on common objectives such as resistance to plum pox virus (PPV), fresh fruit marketability, self-compatibility, and the ripening period (Ledbetter, 2010; Krška and Vachůn, 2016). However, in Türkiye and other apricot producer countries including Uzbekistan, Iran, Pakistan,

[1] International Council for Nuts and Dried Fruits (2023). Nuts & Dried Fruits Statistical Yearbook 2022/2023 [online]. Website https://inc.nutfruit.org/wp-content/uploads/2023/05/Statistical-Yearbook-2022-2023.pdf [accessed 09 May 2024].

* Correspondence: kgurcan@erciyes.edu.tr

and Afghanistan, the breeding of cultivars that produce fruits suitable for drying and that can withstand late spring frosts is a priority. Temperate climate featuring late spring frost in these countries is a major limiting factor in apricot production, as it causes the flowers and young fruits to freeze. Undoubtedly, PPV resistance is also a major breeding objective, particularly in Türkiye, since the country is considered a diversity center for globally problematic PPV strains such as Dideron (Gürcan et al., 2020), Marcus (Gürcan et al., 2019a; Teber et al., 2023), Recombinant (Gürcan et al., 2021), and the PPV-Türkiye strain, which is mostly specific to the country (Teber et al., 2019). Understanding the source of apricot cultivars such as "Zard," which exhibit three important traits—late blooming, homozygous PPV resistance, and suitability for drying (Ledbetter, 2010; Gürcan et al., 2019b)—is crucial for breeding programs in temperate climate regions.

DNA variations in plastomes have been used since the 1980s to study species identification and phylogeny analyses due to their low evolutionary rate, nearly neutral evolution, lack of recombination, and uniparental inheritance (Daniell et al., 2016). While angiosperm plastomes generally exhibit maternal inheritance, some plants, such as conifers, display parental plastome inheritance (Greiner et al., 2015; Kormutak et al., 2018; Park et al., 2021). The increased plastome analysis due to advanced sequencing techniques and reduced costs result in a better understanding of angiosperms evolution (Li and Zheng, 2018; Wang et al., 2016; Zhong et al., 2011). In the *Prunus* genus, the plastome of the *P. persica* species was one of the first completely sequenced genomes (Jansen et al., 2011). Subsequently, plastomes of other *Prunus* species, such as *P. armeniaca*, *P. yedoensis*, *P. serrulate*, *P. mume*, *P. sibirica*, and *P. subhirtella*, have been sequenced and used to reveal phylogenetic relationships within *Prunus* (e.g., Xue et al., 2019; Wang et al., 2020; Dong et al., 2020; Huang et al., 2021; Wu et al., 2023).

Although the structure of the plastome of *P. armeniaca* has been previously described, the evolutionary dynamics of the plastome, including nt variation, gene conversation, and genome structure in the accessions of *P. armeniaca*, have not been reported. Recently, the plastomes of a significant number of accessions in *P. mume* were reported (Shi et al., 2020; Coulibaly et al., 2022; Huang et al., 2021). Genetic diversity assessments in apricots have primarily been conducted using nuclear DNA (e.g., Gürcan et al., 2015, 2016; Decroocq et al., 2016; Liu et al., 2019; Herrera et al., 2021; Sheikh et al., 2021). In contrast to nuclear DNA, maternally inherited plastome DNA offers an alternative opportunity to reveal evolutionary relationships, phylogeny, and diversity in apricots and the *Prunus* genus. In this study, the researchers focused on the genomic features, phylogeny, and maternal inheritance of plastome in apricots, as well as in other *Prunus* spp. retrieved from the GenBank.

2. Materials and methods

2.1. Plant materials, DNA extraction and sequencing

Twenty apricot accessions, including well-known international and local cultivars, as well as 10 open-pollinated apricot hybrids obtained in 2018 from Erciyes University in Kayseri, were used for a comprehensive plastome analysis in this study. Seven hybrids (HH F1-3, HH F1-4, HH F1-11, HH F1-16, HH F1-17, HH F1-18, and HH F1-19) were produced by crossing the cultivar Hacıhaliloğlu (HH) with Stark Early Orange (SEO). The remaining three hybrids (Zard F1-70, Zard F1-98, and Zard F1-146) were produced through the open pollination of the accession "Zard." Additionally, 22 plastomes from members of the *Prunus* genus obtained from the GenBank were included in this study for phylogenetic analysis. Source information for the accessions used in this study is listed in Table S1. For plastome sequencing, DNA was extracted from the fresh leaves using the DNeasy Plant Mini Kit (Qiagen, Valencia, CA, USA). High-quality DNA was then sequenced using various Illumina platforms by Macrogene, Inc. (Seoul, Korea). The quality of the raw reads was evaluated using the FastQC and FASTX toolkits.[2]

2.2. Plastome assembly

First, the raw reads were mapped to the nucleus genome of apricots (GenBank accession number GCA_903114435.1) to remove nuclear DNA. Subsequently, the unmatched reads were mapped to the apricot mitochondrial genome (NC_065228.1) to remove mitochondrial sequences. Burrows–Wheeler Aligner (BWA) software with default parameters was used for mapping (Li and Durbin, 2010). The remaining reads were then de novo assembled, and the top 100 longest contigs were mapped to the apricot reference plastome (NC_043901.1) to obtain a complete plastome assembly using Geneious Prime v2020.2.5[3] (Biomatters, New Zealand). Secondly, NOVOPlasty with default parameters was used to assemble the plastome (Dierckxsens et al., 2017). The NOVOPlasty assembly was then visualized using Geneious Prime and manually edited with the plastome assembled by the Geneious assembler to obtain a complete plastome without heterogeneity.

2.3. Plastome annotation and phylogenetic analysis

The unique assembled genomes were annotated using GeSeq and CPGAVAS2 (Tillich et al., 2017; Shi et al., 2019). The tRNA genes were identified using tRNA-SE with default parameters (Lowe and Chan, 2016). Polymorphic

[2]Hannon GJ (2010). FASTX-Toolkit [online]. Website http://hannonlab.cshl.edu/fastx_toolkit [accessed 09 May 2024].

[3]Geneious Prime [online]. Website https://www.geneious.com/ [accessed 09 July 2024].

sites, including INDELS (insertions and deletions) and SNPs (single nucleotide (nt) polymorphisms), were identified using Geneious Prime with the "find variations/ SNPs tool" and DnaSP v5 (Librado and Rozas, 2009). Variations were also manually checked to select reliable SNPs and avoid software errors. Finally, maps of the circular plastome were drawn using OGDraw 1.3.1 (Greiner et al., 2019). Additionally, the gene codon usage was evaluated using CodonW v1.4.4. The CodonW v1.4.4 program was run on an Ubuntu terminal using a simple script (*codonw file_name.fasta -nomenu*).[4] The codon usage analysis was conducted following the protocol described by Zhang et al. (2023). The mVISTA program, applying the shuffle-Lagan model, was used to compare plastomes and to visualize the alignments with annotation information (Frazer et al., 2004). Additionally, to compare the contraction and expansion of the IR boundaries among LSC, SSC, IRa, and IRb, the IRscope online software (Amiryousefi et al., 2018) was used.

For the phylogenetic analysis, the plastomes were first aligned using the ClustalW2 version 2.1 algorithm on Ubuntu (Larkin et al., 2007). A gap opening penalty of 15 and a gap extension penalty of 6.66 were used for both pairwise and multiple alignments. Maximum likelihood (ML) analyses were conducted to reveal the phylogenetic relationships of the samples using MEGA11 (Tamura et al., 2021) with bootstrap repetition.

3. Results

3.1. Plastome assembly and characteristics

In this study, the DNA of 30 apricot samples was sequenced using Illumina sequencing technology, and plastomes were assembled. The lengths of the assembled novel plastomes ranged from 158,057 bp to 158,089 bp. The nucleotide (nt) comparisons of the 30 apricot plastomes, together with the GenBank accession NC_043901.1, revealed nt variations. The nt variations among plastomes obtained in this study were listed in Table S2. Pairwise nt identity analysis and variation of the novel plastomes, along with GenBank retrieved accessions, revealed that the plastomes of samples identified as *P. armeniaca* can be grouped into three haplotype groups in *P. armeniaca* and Zard group. The nt sequence of seven apricot samples—"M2249," "Harcot," "KZ1," "KZ44," "OB," "Sakıt," and "M2252"—was identical and classified as Haplotype 1. In this group, the plastome size was 158,057 nt in length and exhibited unique variations, including eight deletions, four insertions, and four SNPs specific to this group. The apricots, "Fracasso," "Hacıkaya," "Harlayne," "Hasanbey," "HH," HH F1-3, HH F1-4, HH F1-11, HH F1-16, HH F1-17, HH F1-18, HH F1-19, "Kabaaşı," "Kırmızı," "SEO," and Zhenzhuyou from GenBank (NC_043901.1) had identical plastomes and were classified as Haplotype 2. Haplotype 2 had only one deletion and one SNP variation specific to this group. Another four apricots (M2244, M2243, "Roxana," and "Zerdali") with identical plastomes were classified as Haplotype 3. This group exhibited two deletions and one SNP. The final group that consisted of the "Zard," and its offspring (Zard F1-70, Zard F1-98, and Zard F1-146) had 10 insertions, seven deletions, and 17 SNPs. The plastome characteristics of the four groups were presented in Table 1. "Zard" and its offspring were not classified as a haplotype in *P. armeniaca*, as this group shared higher nt similarity with another apricot species, the Manchurian apricot, while other *P. armeniaca* apricots exhibited close identity with each other within *P. armeniaca.* For three haplotypes of *P. armeniaca*, the overall nt diversity was 0.000021. The complete plastomes are available in NCBI[5] (Haplotype1, Haplotype 2, Haplotype 3, and Zard, with accession numbers OR504552, OR504551, OR504553, and OR504554, respectively).

The structural features, gene content, gene order, introns, and intergenic spacers of the apricot species are highly conserved and similar to those found in previously published plastomes of *Prunus* trees. The plastomes were divided into four regions: large single copy (LSC) regions, small single copy (SSC) regions, and two inverted repeat (IR) regions. The quadripartite structure of the plastomes consisted of an LSC region of 85,746–85,861 bp, an SSC region of 19,049–19,063 bp, and an IR region of 26,342–26,357 bp. The GC content of the plastomes was 36.7%. Coding regions accounted for approximately 52.7% of the plastomes. A total of 122 to 123 coding genes were identified, including 81 protein-coding genes, 33 tRNA genes (the "Zard" group had an additional *trnG-GCC* gene), and eight rRNA genes. The gene content of three haplotypes and the Zard group is listed in Table 2. The plastome map for "HH," representing *P. armeniaca*, is presented in Figure 1. Additionally, the plastome map for "Zard," representing *P. mandshurica*, is displayed in Figure S1.

3.2. Relative synonymous codon usage (RSCU) in apricot species

A total of eight unique plastomes of apricot species—one representative from each of the three haplotype groups and the Zard group, and one representative from other apricot species (*P. mandshurica, P. mume*, *P. sibirica,* and *P. zhengheensis*)—were further analyzed for RSCU. The RSCU of the selected samples was predicted using CodonW, based on protein-coding genes. The total number of codons ranged from 52,400 to 52,700 across all the plastomes (Table S3).

[4]Peden JF (2000). Analysis of codon usage [online]. Website https://codonw.sourceforge.net/ [accessed 11 March 2024].

[5]National Center for Biotechnology Information (NCBI) [online]. Website https://www.ncbi.nlm [accessed 07 July 2024].

Table 1. Structural information on the plastomes of three *P. armeniaca* haplotypes and the Zard group.

Group	Length (bp)	LSC	SSC	IRa	IRb	GC (%)	Total genes	tRNA genes	rRNA genes	PC genes	Number of codons
1*	158,057	86,269–158,057	112,659–131,668	131,669–158,057	86,270–112,658	36.7	122	33	8	81	52,685
2**	158,071	86,284–158,072	112,674–131,682	131,683–158,071	86,285–112,673	36.7	122	33	8	81	52,690
3***	158,058	86,271–158,058	112,661–131,669	131,670–158,058	86,272–112,660	36.7	122	33	8	81	52,686
4****	158,089	86,312–158,089	112,702–131,700	131,701–158,089	86,313–112,701	36.7	123	33	8	82	52,696

*1 Haplotype 1 accessions: Harcot, M2249, M2252, KZ1, KZ44, OB, and Sakıt.
**2 Haplotype 2: Fracasso, HH, Hacıkaya, Harlayne, Hasanbey, Kabaaşı, Kırmızı, HH F1-3, HH F1-4, HH F1-11, HH F1-16, HH F1-17, HH F1-18, HH F1-19, SEO, and "Zhenzhuyou" from the GenBank (NC_043901.1).
***3 Haplotype 3: M2243, M2244, Roxana, and Zerdali.
***4 Zard group: Zard, Zard F1-70, Zard F1-98, and Zard F1-146.

Table 2. Genes identified in the *P. armeniaca* plastomes, grouped by function: self-replication, photosynthesis, and unknown function.

Category	Group of genes	Name of genes
Self-replication	Large subunit of ribosomal proteins	*rpl2*, *14*, *16*, *20*, *22*, *23*, *32*, *33*, *36*
	Small subunit of ribosomal proteins	*rps2*, *3*, *4*, *7*, *8*, *11*, *12*, *14*, *15*, 16, *18*, *19*
	DNA-dependent RNA polymerase	*ndhB, rpoA*, *B*, *C1*, *C2*
	rRNA genes	*rrn16, 23, 4.5, 55*
	tRNA genes	*trnA-UGC, trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnfM-CAU,* *trnG-UCC,* ***trnG-GCU****, trnH-GUG, trnI-CAU, trnI-GAU, trnK-UUU, trnL-CAA, trnY-ATA* *trnL-UAA, trnL-UAG, trnfM-CAU, trnN-GUU, trnP-GGG, trnP-UGG,* *trnQ -UUG, trnR-ACG, trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA,* *trnT-GGU, trnT-UGU, trnV-GAC, trnV-UAC, trnW-CCA, trnY-GUA,*
Photosynthesis	Photosystem I	*psaA*, *B*, *C*, *I*, *J*, *M*
	Photosystem II	*psbA*, *B*, *C*, *D*, *E*, *F*, *H*, *I*, *J*, *K*, *L*, *M*, *N*, *T*, *Z*
	Maturase	*MatK*
	Protease	*ClpP*
	Cytochrome b6/f complex	*petA*, *B*, *D*, *G*, *L*, *N*
	ATP synthase	*atpA*, *B*, *E*, *F*, *H*, *I*
	Rubisco	*RbcL*
	Chlorophyll biosynthesis	*chlB*, *L*, *N*
Unknown	Conserved open reading frames	*ycf1*, *2*, *3*, *4, 15*

In haplotypes 1, 2, and 3, the codons 52,685, 52,690, and 52,686 were encoded, respectively. *P. sibirica* had the highest number of codons (52,724) among the apricot plastomes, while haplotype 1 had the fewest (52,685). In this study, we found that there are 64 types of codons encoding 20 amino acids. Leucine (Leu), serine (Ser), and arginine (Arg) are each encoded by six codons; valine (Val), proline (Pro), threonine (Thr), alanine (Ala), and glycine (Gly) are each encoded by four codons; isoleucine (Ile) is encoded by three codons; while phenylalanine (Phe), tyrosine (Tyr), histidine (His), glutamine (Gln), asparagine (Asn), lysine (Lys), aspartic acid (Asp), glutamic acid (Glu), and cysteine (Cys) are each encoded by two codons; and the remaining methionine (Met) and tryptophan (Trp) are each encoded by only one codon. We also found that Leu is the most frequent amino acid, with a frequency range of

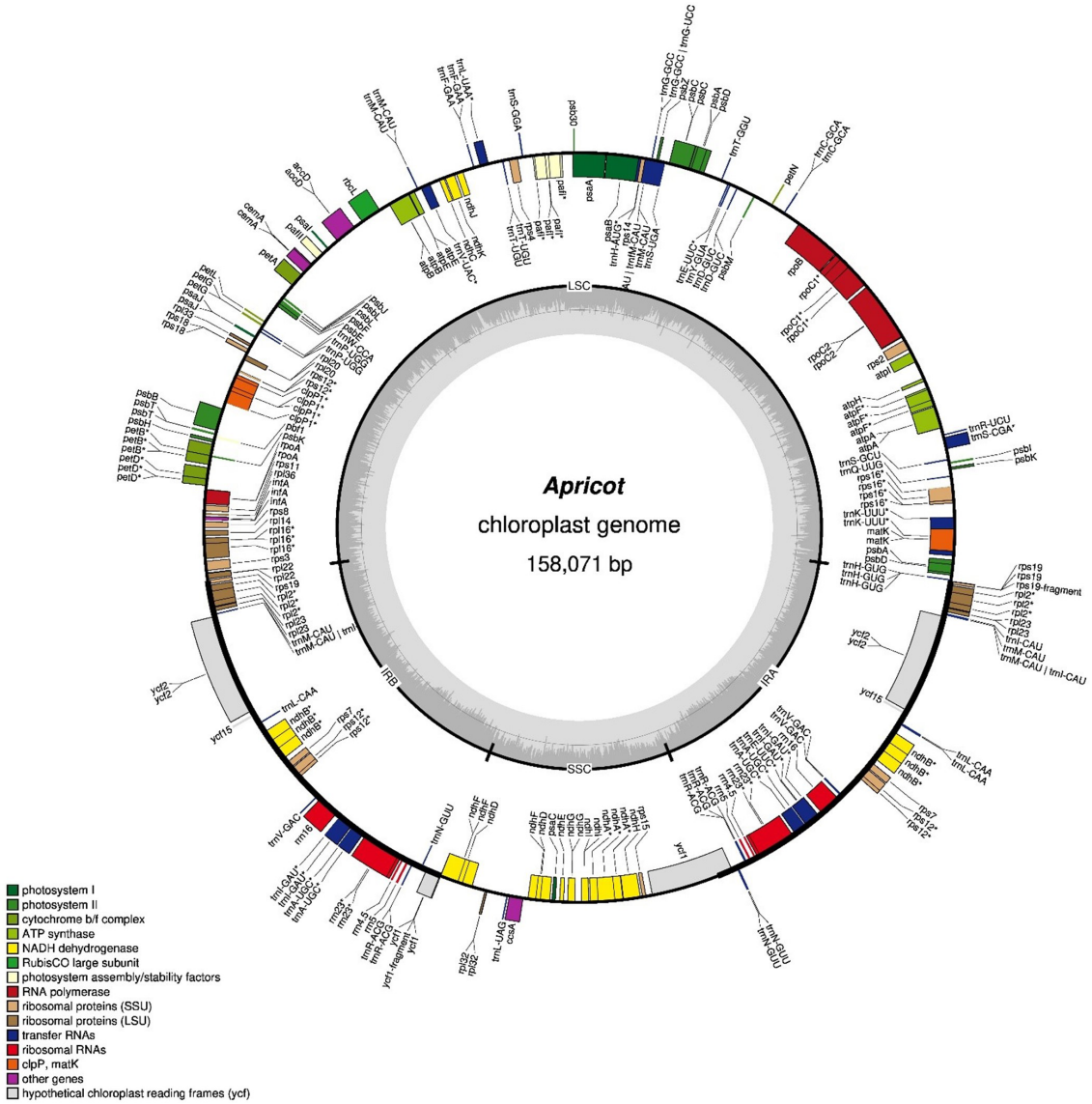

Figure 1. Plastome map of *P. armeniaca* showing the four regions of the plastome, (IRa, IRb, SSC, and LSC) with annotated genes.

9.80%–10.80%. Trp was the least frequent amino acid in plastome protein-coding genes, with a frequency range of 1.25%–1.35%. The highest RSCU among amino acids was observed in Arg with a value of 1.92, while other amino acids such as Leu and Ser had their highest RSCU values at 1.49 and 1.48, respectively. Met and Trp had an RSCU value of 1, indicating no codon usage bias for these two amino acids. Approximately 46% of the codons exhibited an RSCU value of less than 1, suggesting a negative codon usage bias for these codons.

3.3. Comparative analysis for gene order

To further analyze the potential divergence of the plastome sequences of the eight unique apricots (haplotypes 1, 2, 3, Zard group, *P. mandshurica, P. mume*, *P. Sibirica,* and *P. zhengheensis*), the mVISTA program was used. We depicted only 99 k length of the alignment of eight plastomes in Figure 2, as showing the comparison of entire plastomes reduced the resolution of the figure. The alignment indicated that there are no genome or gene rearrangements within the apricot plastomes. The plastomes were highly conserved, and no structural differences were observed.

3.4. Inverted repeat (IR) comparative analysis

IR comparative analysis of the apricot accessions revealed that haplotypes 1, 2, 3, Zard group, and *P. mandshurica* have the same genes in the IR region. These genes were of the same size in these plastomes, except for the *rpl2*

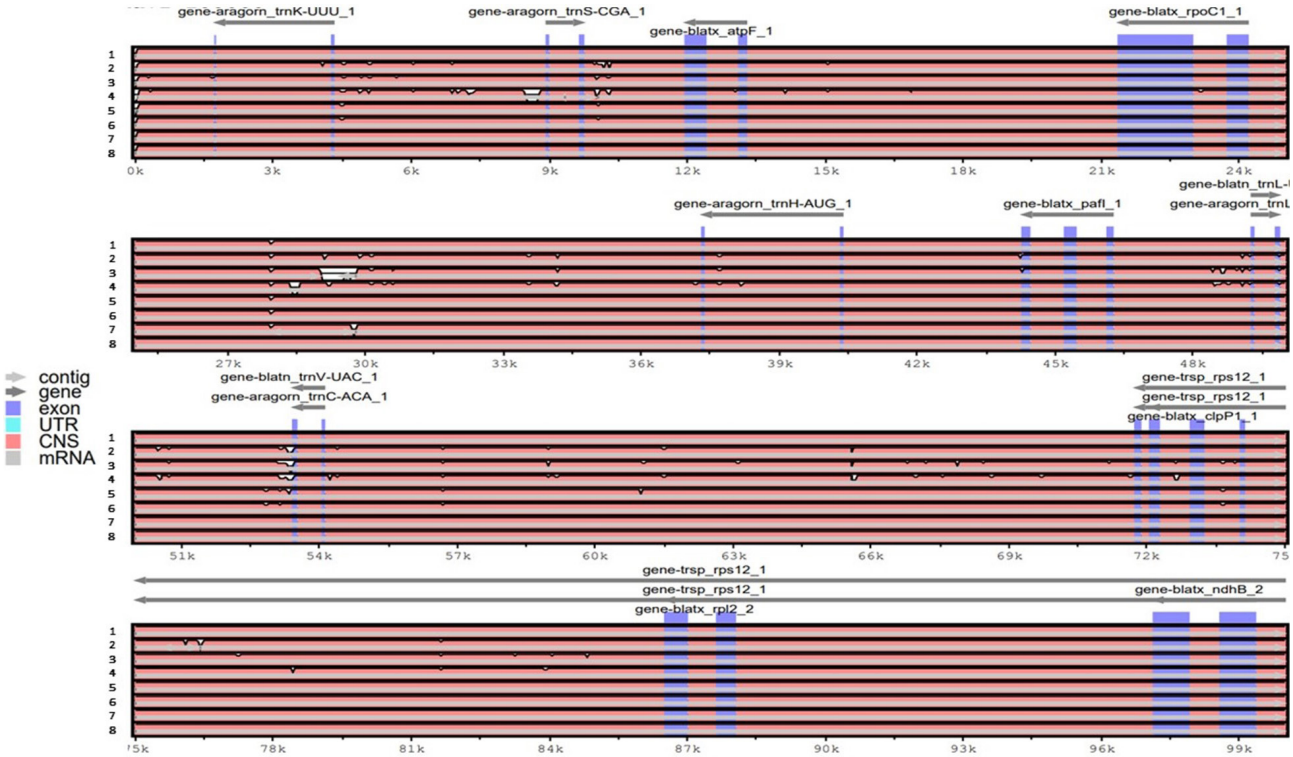

Figure 2. mVISTA alignment graph showing the aligned apricot plastomes; 1: Haplotype 1, 2: Haplotype 2, 3: Haplotype 3, 4: Zard group, 5: *P. mandshurica*, 6: *P. mume*, 7: *P. zhengheensis*, and 8: *P. sibirica*. This alignment shows the expressed genes in this region of the apricot plastomes, which are conserved.

gene in *P. mandshurica*, which was 13 bp, while in the other plastomes, it was 2 bp. It is also noteworthy that in these plastomes, the *ycf1* gene was found overlapping the JSA region, while in the other three apricot plastomes—*P. mume*, *P. zhengheensis*, and *P. sibirica*—the *ycf1* gene was absent in the JSA region (Figure 3).

An IR comparison among the genomes of all other *Prunus* species downloaded from NCBI revealed differences in the lengths of IRa, IRb, LSC, and SSC (Figure S2). Most of the genes were conserved across these *Prunus* species, except for the *rpl2* and the *ndhF* genes. The *rpl2* gene was absent in the IRb region of *P. tenella*, while it was present in the other plastomes. This gene was also longer in *P. padus* and *P. serotina* compared to the *rbl2* genes in the other plastomes. The *ndhF* gene was absent only in *P. serotina* and did not overlap with the JSB region in *P. subhirtella, P. yedoensis,* and *P. tangutica*, while in the other plastomes, it did overlap with this region.

3.5. Phylogenetic relationships

ML phylogenetic tree analysis was conducted using the plastomes of the eight unique apricots (haplotypes 1, 2, 3, Zard group, *P. mandshurica, P. mume*, *P. sibirica*, and *P. zhengheensis*) and other *Prunus* species (Figure 4). The phylogenetic trees suggest that the three *P. armeniaca* species, haplotypes 1, 2, and 3, are closely related. The "Zard" accession and its offspring were placed in the same clade as *P. mandshurica*. *P. armeniaca* is phylogenetically sister to *P. mandshurica. The s*ister clades of *P. armeniaca* and *P. mandshurica* were adjacent to other apricot species *P. mume* and *P. zhengheensis* clades. However, the Siberian apricot, *P. sibirica*, is more closely related to plums (*P. cerasifera, P. domestica*, and *P. salicina*) than to other apricot species. When focusing on the other *Prunus* species, *P. fasciculata* was positioned between *P. dulcis* and the cultivated cherry clade (*P. mahaleb, P. serrualta, P. subhirtella,* and *P. yedoensis*). *P. pedunculata*, *P. triloba*, *P. humilis,* and *P. tomentosa* were also classified into one clade, even though *P. humilis* and *P. tomentosa* are bush cherries, while *P. pedunculata* and *P. triloba* are Chinese almonds.

3.6. Inheritance of plastomes

The maternal inheritance of plastomes in the *Prunus* species was confirmed by the similarity of the F1 hybrids to their maternal parents. The "Zard" F1 hybrids—Zard F1-98, Zard F1-70, and Zard F1-146—were identical to the maternal plastome of the "Zard" accession. Similarly, the "HH" F1 hybrids—HH F1-3, HH F1-4, HH F1-11, HH F1-16, HH F1-17, HH F1-18, HH F1-19—were identical to their maternal parent "HH," supporting the maternal inheritance of plastome in *P. armeniaca.*

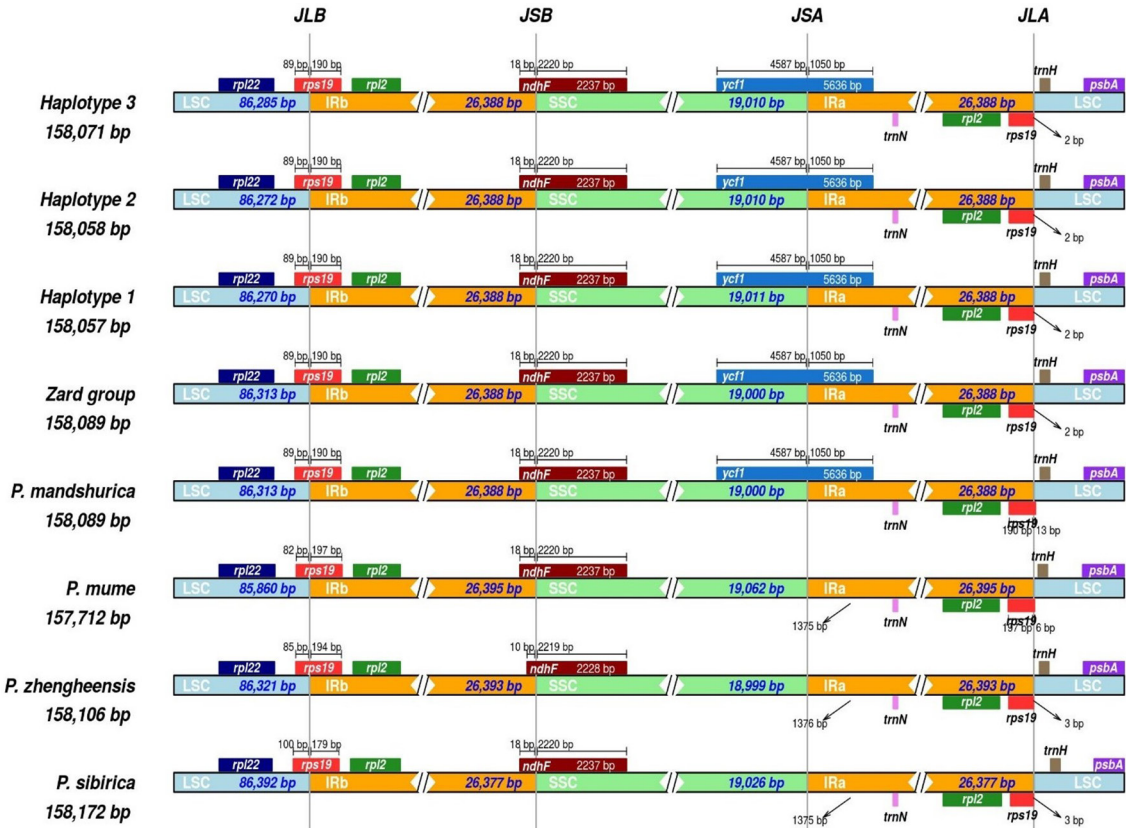

Figure 3. IR comparative analysis graph of the eight apricot plastomes, showing the gene distribution and region lengths in the IRa, IRb, SSC, and LSC regions. JLB represents the junction between LSC and IRb, JSB represents the junction between IRb and SSC, JSA represents the junction between SSC and IRa, and JLA represents the junction between IRa and LSC.

4. Discussion

The objective of the present study was to investigate the genomic structure, genetic diversity, and molecular phylogeny of the plastomes in *P. armeniaca* L. For this purpose, 20 plastomes of apricots were assembled and characterized through genome annotation and comparative studies. Additionally, plastomes of 10 F1 seedlings of apricots were obtained to investigate the parental origin of the plastome.

The lengths of the novel plastomes are closer to those of *Prunus* plastomes reported in previous studies (Xue et al., 2019; Wang et al., 2020; Huang et al., 2021). The novel plastomes had a GC content of 36.7%, which is typical of angiosperms, as they generally have a low GC content (Daniell et al., 2016; Zhang et al., 2017). The number of genes (122-123) of the plastomes of apricots sequenced in this study is similar to that of other *Prunus* species such as *P. persica (131), P. mume (132),* and *P. tomentosa (135)* (Chen et al., 2015; Wu et al., 2017; Xue et al., 2019). The *rps19* gene, which is common in apricots according to this study, is also reported to be common in *Prunus* species (Shen et al., 2016; Xue et al., 2019). RSCU analysis in this study showed that among all the amino acids, Leu is the most frequent amino acid (9.80%–10.80%), while Trp is the least frequent amino acid (1.25%–1.35%) in the novel apricots. Previous studies on *P. mume, P. armeniaca*, and *P. salicina* also show that Leu is the most frequently encoded amino acid, while Trp is the least frequently encoded amino acid. Additionally, cysteine appeared the least in the *P. zhengheensis* plastome (Xue et al., 2019; Huang et al., 2021). Overall, the analysis of the plastomes of 30 novel apricots revealed that genomic features, genomic structure, and gene organization and order are highly conserved among apricots and are consistent with those in plastomes of other *Prunus* species.

The apricot samples from germplasm collections in various countries were investigated using SSR and SNP markers, revealing high genetic diversity among the apricots. The apricot accessions used in this study were also examined by SSR and SNP markers in the previous research. The results indicated that apricots tended to group in the phylogenetic analysis according to their response to plum pox virus (PPV) infection. This suggests that PPV-resistant apricots have a common geographic

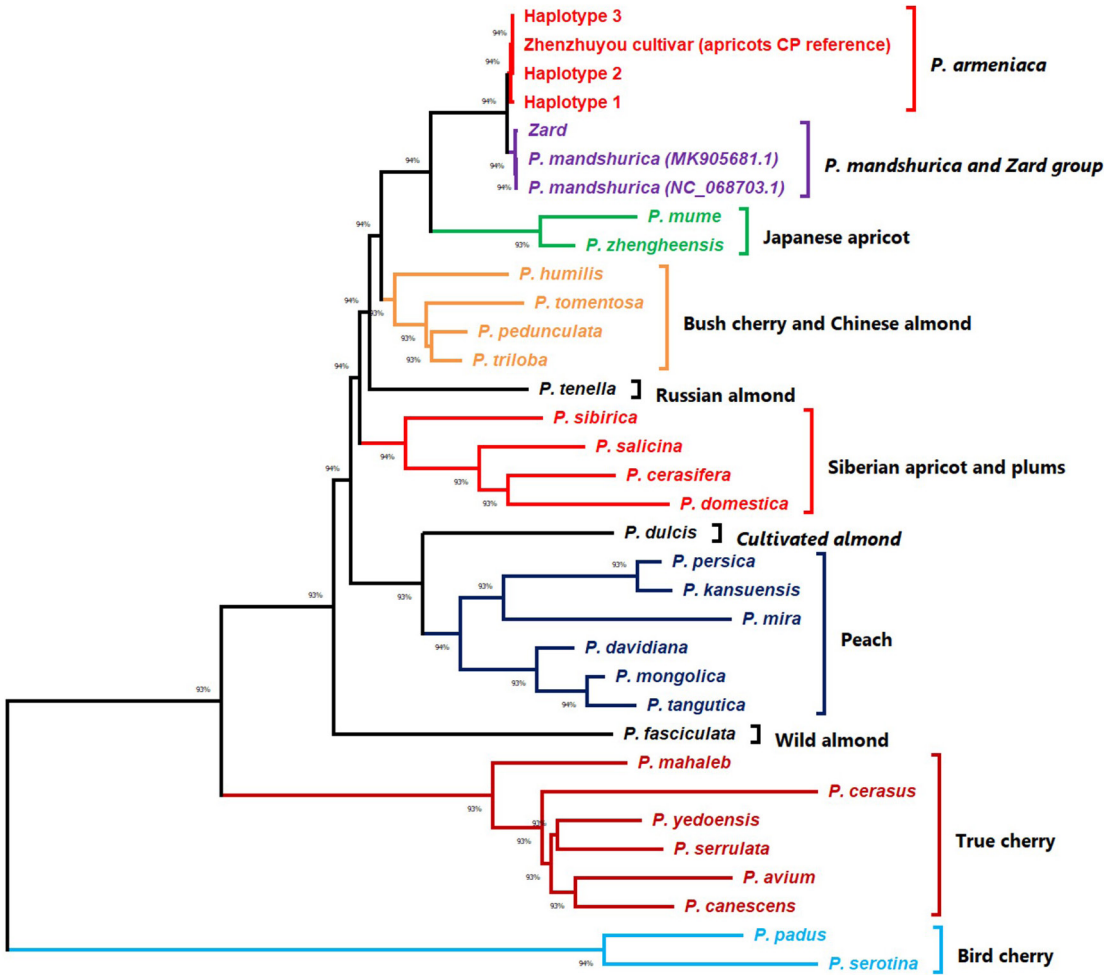

Figure 4. Maximum-likelihood phylogenetic tree of the *Prunus* species, showing the "Zard" accession in the same clade as *P. mandshurica*. The Hacıhaliloğlu F1 offspring were identical to Haplotype 3, while Zard F1 offspring were identical to their maternal parent, Zard.

origin (Zhebentyayeva et al., 2008; Gürcan et al., 2015, 2016; Decroocq et al., 2016). On the contrary, in this study, plastome analysis did not distinguish PPV resistant cultivated apricots (SEO, Harlayne, Kırmızı, M2243, and M224) from others, indicating that plastome analysis is less effective than nuclear DNA in distinguishing apricots based on their origin. However, the apricot "Zard," which was not distinguished from common apricots by nuclear DNA markers in the previous studies, was clearly distinguished from the remaining apricots and identified as Manchurian apricot by plastome analysis, highlighting the significance of plastome analysis in species discrimination among apricots. This result clearly explains why "Zard" has numerous unique features, such as late blooming, frost tolerance, high soluble solids in its fruits, upright growth, and higher survival ability under abiotic and biotic stress conditions (Ledbetter, 2010, personal observation of Kahraman Gürcan). Furthermore, unlike *P. armeniaca* accessions, which are typically heterozygous resistant to PVP, "Zard" is one of the rare apricot varieties that is homozygous resistant to PPV (Gürcan et al., 2019b).

With the exception of *P. mume*, multiple plastomes within a single species in *Prunus* are scarce, and this limits evolutionary studies at the plastome level for these species. In the species *P. mume*, Huang et al. (2022) reported the complete plastome sequencing of 146 accessions from different geographic locations in China, revealing 41 haplotypes. In this population, nt diversity (0.00035) was found to be almost 17 times higher than the nt diversity (0.000021) found in *P. armeniaca*. In another study, the sizes of plastomes among 10 *P. mume* accessions ranged from 157,915 to 158,150, exhibiting greater size variation (235 nt) compared to the close genome sizes observed in over 20 accessions of *P. armeniaca* (13 nt). In brief, compared to the *P. mume* species, the cultivated *P. armeniaca* exhibited narrow genetic diversity, as evidenced by similar genome sizes, fewer haplotypes and nt variations, and limited phylogenetic branching.

The maternal inheritance of the plastome was validated based on these results due to the similarities between the hybrids' plastomes and those of their maternal parents. Evidently, the three "Zard" F1 hybrids were identical to the "Zard" accession, and all the seven "Hacıhaliloğlu" F1 hybrids were also identical to the plastome of "Hacıhaliloğlu." These results align with the general inheritance patterns observed in angiosperms (Greiner et al., 2015; Kormutak et al., 2018).

In conclusion, the plastomes of 30 apricots were de novo assembled and analyzed phylogenetically with 22 previously published plastomes of *Prunus* species. The complete plastomes of the *Prunus* species were annotated, and coding genes were identified. The gene order and genome structure of *P. armeniaca* are highly conserved, with only a few variations compared to other *Prunus* species. No major structural differences were observed among the plastomes of *P. armeniaca* accessions. The plastomes of the 30 apricots obtained in this study are highly conserved, and three haplotypes were identified in *P. armeniaca*, indicating narrow genetic diversity in this species. Highly limited phylogenetic branching supported narrow genetic diversity in cultivated apricots. Only the "Zard" and its offspring exhibited greater nt differences and were placed in the same clade as *P. mandshurica*. Among the apricot species, *P. mume*, *P. zhengheensis*, and *P. mandshurica* are closely related to *P. armeniaca*, while the Siberian apricot, *P. sibirica*, is not closely related to *P. armeniaca*. The maternal inheritance of the plastome in the *Prunus* species was validated in this study, as all the F1 hybrids were identical to their maternal parents' plastome. These results expand the researchers' perspectives on apricot plant diversity and promote an understanding of the evolutionary relationship between apricot and *Prunus* species.

Acknowledgment
This study was supported by The Scientific and Technological Research Council of Türkiye (TÜBİTAK) (project numbers 214O400 and 119O846).

References

Amiryousefi Ali, Hyvönen J, Poczai P (2018). IRscope: an online program to visualize the junction sites of chloroplast genomes. Bioinformatics 34 (17): 3030-3031. https://doi.org/10.1093/bioinformatics/bty220

Chen T, Chen Q, Luo Y, Huang ZL, Zhang J et al. (2015). Phylogeography of Chinese cherry (*Prunus pseudocerasus* Lindl.) inferred from chloroplast and nuclear DNA: insights into evolutionary patterns and demographic history. Plant Biology 17 (4): 787-797. https://doi.org/10.1111/plb.12294

Coulibaly D, Huang X, Ting S, Iqbal S, Ni Z et al. (2022). Comparative analysis of complete chloroplast genome and phenotypic characteristics of Japanese apricot accessions. Horticulturae 8 (9): 794. https://doi.org/10.3390/horticulturae8090794

Daniell H, Lin CS, Yu M, Chang WJ (2016). Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biology 17: 1-29. https://doi.org/10.1186/S13059-016-1004-2

Decroocq S, Cornille A, Tricon D, Babayeva S, Chague A et al. (2016). New insights into the history of domesticated and wild apricots and its contribution to *Plum pox virus* resistance. Molecular Ecology 25 (19): 4712-4729. https://doi.org/10.1111/MEC.13772

Dierckxsens N, Mardulyn P, Smits G (2017). NOVOPlasty: *de novo* assembly of organelle genomes from whole genome data. Nucleic Acids Research 45 (4): e18. https://doi.org/10.1093/NAR/GKW955

Dong S, Zhang X, Sun Y, Xu H, Zhang H et al. (2020). The complete chloroplast genome sequence of *Prunus sibirica*. Mitochondrial DNA Part B 5 (1): 581-582. https://doi.org/10.1080/23802359.2019.1710589

Ercisli S (2009). Apricot culture in Turkey. Science Research Essays 4 (8): 715-719.

Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I (2004). VISTA: computational tools for comparative genomics. Nucleic Acids Research 32 (Suppl 2): W273-W279. https://doi.org/10.1093/NAR/GKH458

Greiner S, Lehwark P, Bock R (2019). OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Research 47 (W1): W59-W64. https://doi.org/10.1093/NAR/GKZ238

Greiner S, Sobanski J, Bock R (2015). Why are most organelle genomes transmitted maternally? Bioessays 37 (1): 80-94. https://doi.org/10.1002/BIES.201400110

Groppi A, Liu S, Cornille A, Decroocq S, Bui QT et al. (2021). Population genomics of apricots unravels domestication history and adaptive events. Nature Communications 12 (1): 3956. https://doi.org/10.1038/s41467-021-24283-6

Gürcan K, Çetinsağ N, Pınar H, Macit T (2019b). Molecular and biological assessment reveals sources of resistance to *Plum pox virus* - Turkey strain in Turkish apricot (*Prunus armeniaca*) germplasm. Scientia horticulturae 252: 348-353. https://doi.org/10.1016/j.scienta.2019.04.003

Gürcan K, Öcal N, Yılmaz KU, Ullah S, Erdoğan A et al. (2015). Evaluation of Turkish apricot germplasm using SSR markers: genetic diversity assessment and search for Plum pox virus resistance alleles. Scientia Horticulturae 193: 155-164. https://doi.org/10.1016/j.scienta.2015.07.012

Gürcan K, Teber S, Akbulut M, Çağlayan K (2021). Genetic diversity and a long evolutionary history of plum pox virus strain rec in Turkey. European Journal of Plant Pathology 161: 453-461. https://doi.org/10.1007/s10658-021-02336-2

Gürcan K, Teber S, Candresse T (2020). Genetic analysis suggests a long and largely isolated evolutionary history of plum pox virus strain D in Turkey. Plant Pathology 69 (2): 370-378. https://doi.org/10.1111/ppa.13115

Gürcan K, Teber S, Çağlayan K (2019a). Further investigation of a genetically divergent group of plum pox virus-M strain in Turkey. Journal of Plant Pathology 101: 385-391. https://doi.org/10.1007/s42161-018-0187-7

Gürcan K, Teber S, Ercisli S, Yilmaz KU (2016). Genotyping by sequencing (GBS) in apricots and genetic diversity assessment with GBS-derived single-nucleotide polymorphisms (SNPs). Biochemical Genetics 54: 854-885. https://doi.org/10.1007/s10528-016-9762-9

Herrera S, Hormaza JI, Lora J, Ylla G, Rodrigo J (2021). Molecular characterization of genetic diversity in apricot cultivars: current situation and future perspectives. Agronomy 11 (9): 1714.

Huang X, Ni Z, Shi T, Tao R, Yang Q et al. (2022). Novel insights into the dissemination route of Japanese apricot (*Prunus mume* Sieb. et Zucc.) based on genomics. The Plant Journal 110 (4): 1182-1197. https://doi.org/10.1111/tpj.15731

Huang X, Tan W, Li F, Liao R, Guo Z et al. (2021). The chloroplast genome of *Prunus zhengheensis*: genome comparative and phylogenetic relationships analysis. Gene 793: 145751. https://doi.org/10.1016/j.gene.2021.145751

Jansen RK, Saski C, Lee SB, Hansen AK, Daniell H (2011). Complete plastid genome sequences of three rosids (*Castanea, Prunus, Theobroma*): evidence for at least two independent transfers of rpl22 to the nucleus. Molecular Biology and Evolution 28 (1): 835-847. https://doi.org/10.1093/MOLBEV/MSQ261

Kormutak A, Galgoci M, Sukenikova D, Bolecek P, Libantova J et al. (2018). Maternal inheritance of chloroplast DNA in *Pinus mugo* Turra: a case study of *Pinus mugo* × *Pinus sylvestris* crossing. Plant Systematics and Evolution 304: 71-76. https://doi.org/10.1007/s00606-017-1449-0

Krška, B, Vachůn, Z. (2016). Apricot breeding at the Faculty of Horticulture in Lednice. Agronomy 6 (2): 27. https://doi.org/10.3390/agronomy6020027

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA et al. (2007). Clustal W and Clustal X version 2.0. Bioinformatics 23 (21): 2947-2948. https://doi.org/10.1093/BIOINFORMATICS/BTM404

Ledbetter CA (2010). Apricot breeding in North America: current status and future prospects. Acta Horticulturae 862: 85-92. https://doi.org/10.17660/ActaHortic.2010.862.12

Li B, Zheng Y (2018). Dynamic evolution and phylogenomic analysis of the chloroplast genome in Schisandraceae. Scientific Reports 8 (1): 9285. https://doi.org/10.1038/s41598-018-27453-7

Li H, Durbin R (2010). Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics 26 (5): 589-595. https://doi.org/10.1093/BIOINFORMATICS/BTP698

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

Liu S, Cornille A, Decroocq S, Tricon D, Chague A et al. (2019). The complex evolutionary history of apricots: species divergence, gene flow, and multiple domestication events. Molecular ecology 28 (24): 5299-5314. https://doi.org/10.1111/MEC.15296

Lowe TM, Chan PP (2016). tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Research 44 (W1): W54-W57. https://doi.org/10.1093/NAR/GKW413

Park HS, Lee WK, Lee SC, Lee HO, Joh HJ et al. (2021). Inheritance of chloroplast and mitochondrial genomes in cucumber revealed by four reciprocal F1 hybrid combinations. Scientific Reports 11 (1): 2506. https://doi.org/10.1038/s41598-021-81988-w

Rehder A (1954). Manual of Cultivated Trees and Shrubs. 2nd ed. Portland, OR, USA: Dioscorides Press.

Sauer JD (1993). Historical Geography of Crop Plants. Boca Raton, FL, USA: CRC Press.

Sheikh ZN, Sharma V, Shah RA, Raina S, Aljabri M et al. (2021). Elucidating genetic diversity in apricot (*Prunus armeniaca* L.) cultivated in the North-Western Himalayan Provinces of India using SSR markers. Plants 10 (12): 2668. https://doi.org/10.3390/plants10122668

Shen L, Guan Q, Amin A, Zhu W, Li M et al. (2016). Complete plastid genome of *Eriobotrya japonica* (Thunb.) Lindl and comparative analysis in Rosaceae. SpringerPlus 5 (1): 1-4. https://doi.org/10.1186/s40064-016-3702-3

Shi L, Chen H, Jiang M, Wang L, Wu X et al. (2019). PLASTOMEGAVAS2, an integrated plastome sequence annotator and analyzer. Nucleic Acids Research 47 (W1): W65-W73. https://doi.org/10.1093/NAR/GKZ345

Shi T, Luo W, Li H, Huang X, Ni Z et al. (2020). Association between blooming time and climatic adaptation in *Prunus mume*. Ecology and Evolution 10 (1): 292-306. https://doi.org/10.1002/ece3.5894

Tamura K, Stecher G, Kumar S (2021). MEGA11: molecular evolutionary genetics analysis version 11. Molecular Biology and Evolution 38 (7): 3022-3027. https://doi.org/10.1093/MOLBEV/MSAB120

Teber S, Ceylan A, Gürcan K, Candresse T, Ulubaş Serçe Ç et al. (2019). Genetic diversity and molecular epidemiology of the T strain of *Plum pox virus*. Plant Pathology 68 (4): 755-763. https://doi.org/10.1111/ppa.12974

Teber S, Gürcan K, Brevet M, Palmisano F, Candresse T et al. (2023). High genetic diversity suggests a long evolutionary history of plum pox virus-M in Turkey. Plant Pathology 72 (2): 300-311. https://doi.org/10.1111/ppa.13663

Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A et al. (2017). GeSeq – versatile and accurate annotation of organelle genomes. Nucleic Acids Research 45 (W1): W6-W11. https://doi.org/10.1093/NAR/GKX391

Wang L, Wuyun TN, Du H, Wang D, Cao D (2016). Complete chloroplast genome sequences of *Eucommia ulmoides*: genome structure and evolution. Tree Genetics & Genomes 12: 1-5. https://doi.org/10.1007/s11295-016-0970-6

Wang W, Yang T, Wang HL, Li ZJ, Ni JW et al. (2020). Comparative and phylogenetic analyses of the complete chloroplast genomes of six almond species (*Prunus spp*. L.). Scientific Reports 10 (1): 10137. https://doi.org/10.1038/s41598-020-67264-3

Wu B, Gao L, Gao J, Xu Y, Liu H et al. (2017). Genome-wide identification, expression patterns, and functional analysis of UDP glycosyltransferase family in peach (*Prunus persica* L. Batsch). Frontiers in Plant Science 8: 389. https://doi.org/10.3389/FPLS.2017.00389/BIBTEX

Wu JQ, Wang Y, Sun P, Sun ZS, Shen JS (2023). The complete chloroplast genome of *Prunus phaeosticta* (Hance) Maxim. (Rosaceae) and its phylogenetic implications. Mitochondrial DNA Part B 8 (1): 136-40. https://doi.org/10.1080/23802359.2022.2163841

Xue S, Shi T, Luo W, Ni X, Iqbal S et al. (2019). Comparative analysis of the complete chloroplast genome among *Prunus mume*, *P. armeniaca*, and *P. salicina*. Horticulture Research 6: 89. https://doi.org/10.1038/s41438-019-0171-1

Zhang D, Tu J, Ding X, Guan W, Gong L et al. (2023). Analysis of the chloroplast genome and phylogenetic evolution of *Bidens pilosa*. BMC Genomics 24 (1): 1-3. https://doi.org/10.1186/s12864-023-09195-7

Zhang QP, Liu WS, Liu N, Zhang YP, Xu M (2017). Allelic variation of simple sequence repeats markers linked to PPV resistance in Chinese apricot. Horticultural Science 44 (1): 6-13. https://doi.org/10.17221/278/2015-HORTSCI

Zhebentyayeva T, Ledbetter C, Burgos L, Llácer G (2012). Apricot. In: Badenes M, Byrne D (editors). Fruit Breeding. Handbook of Plant Breeding. Volume 8. New York, NY, USA: Springer Science and Business Media, pp. 415-458. https://doi.org/10.1007/978-1-4419-0763-9_12

Zhebentyayeva TN, Reighard GL, Lalli D, Gorina VM, Krška B et al. (2008). Origin of resistance to *plum pox virus* in apricot: what new AFLP and targeted SSR data analyses. Tree Genetics & Genomes 4: 403-417. https://doi.org/10.1007/s11295-007-0119-8

Zhong B, Deusch O, Goremykin VV, Penny D, Biggs PJ et al. (2011). Systematic error in seed plant phylogenomics. Genome Biology and Evolution 3: 1340-1348. https://doi.org/10.1093/GBE/EVR105

Supplementary materials

Table S1. The list of apricots and other species of *Prunus* genus used for chloroplast genome analysis.

a) Twenty common apricot (*P. armeniaca*) accessions from Erciyes University and used for de nova chloroplast genome assembly.

Accession	Origin
Sakıt	Türkiye
M2249	Türkiye
M2252	Türkiye
Ordubat (OB)	Türkiye
Hacıkaya	Türkiye
Kabaaşı	Türkiye
Harlayne	USA
Zerdali	Türkiye
M2243	Türkiye
M2244	Türkiye
KZ1	Türkiye
KZ44	Türkiye
Roxana	Afghanistan
Kırmızı	Türkiye
Hasanbey	Türkiye
Stark Early Orange (SEO)	North America
Hacıhaliloğlu	Türkiye
Fracasso	Italy
Zard	USA
Harcot	Canada

b) Ten F1 apricots (*P. armeniaca*) from Erciyes University and used for de nova chloroplast genome assembly.

Name	Parents (maternal * paternal)
HH 3	HH * SEO
HH 4	HH * SEO
HH 16	HH * SEO
HH 17	HH * SEO
HH 18	HH * SEO
HH 19	HH * SEO
HH 11	HH * SEO
Zard 70	Zard * Unknown
Zard 98	Zard * Unknown
Zard 146	Zard * Unknown

c) Twenty-two *Prunus* spp. CP genomes retrieved from NCBI.

Name	GenBank accession	Country
P. armeniaca	NC_043901.1	USA
P. cerasifera	NC_068605.1	USA
P. dulcis	KY085904.1	USA
P. fasciculata	NC_054253.1	USA
P. kanseusis	NC_023956.1	USA
P. mahaleb	MT576896.1	China
P. mandshurica	NC_068703.1	China
P. mandshurica	MK905681.1	China
P. mongolica	NC_037849.1	USA
P. pendunculata	MG602257.1	USA
P. persica	HQ336405.1	USA
P. serrulata	NC_066418.1	USA
P. sibirica	NC_068704	China
P. tangutica	MZ145044.1	China
P. tomentosa	MT576919.1	China
P. triloba	MH748555.1	China
P. yedoensis	KU985054.1	South Korea
P. dulcis	KY085904.1	USA
P. salicina	NC_047442.1	China
P. domestica	NC_050959.1	China
P. avium	NC_044701.1	USA
P. cerasus	NC_066420.1	USA
P. zhengheensis	NC_062793.1	China
P. tenella	NC_044965.1	China
P. humilis	NC_035880.1	China
P. canescens	MK816299.1	USA
P. serotina	NC_036133.1	China
P. padus	NC_026982.1	USA
P. mira	NC_040125.1	USA
P. davidiana	NC_039735.1	USA

Table S2. Variations in CP genomes: SNPs, insertions, and deletions across 31 apricots varieties examined in this study.

Haplotype 1: Harcot, KZ1, KZ44, M2249, Ordubat, Sakıt, and M2252.

Nucleotide	Type of variation	Position	Frequency
T	Deletion	1651	96.90
TGAAGTGTATAAT	Deletion	27,940	66.30
T--	Deletion	37,693	75.90
T--	Deletion	52,377	91.90
T--	Deletion	66,206	88.80
GG--	Insertion	66,763	79.90
T--	Deletion	69,730	64.80
T	Insertion	73,666	93.20
G-T	SNP (transversion)	76,147	98.90
T--	Deletion	83,306	90.00
T	Insertion	115,568	88.50
A--	Deletion	121,972	93.70
T	Insertion	122,331	76.90

Table S2. (Continued.)

G-T	SNP (transversion)	123,712	99.00
A-G	SNP (transition)	124,312	98.90
T-G	SNP (transversion)	130,167	96.30
Summary			
Deletions	8		
Insertions	4		
SNPs	4		
CP genome length	158,057		

Haplotype 2: Fracasso, Hacıkaya, Harlayne, Hasanbey, HH, HH F1 (-3, -4, -11, -16, -17, -18, and -19), Kabaaşı, Kırmızı, and SEO.

Nucleotide	Type of variation	Position	Frequency
T--	Deletion	9959	91.90
C-T	SNP (transition)	56,788	99.30
Summary			
Deletions	1		
SNPs	1		
CP genome length	158,071		

Haplotype 3: M2244, M2243, Roxana, and Zerdali.

Nucleotide	Type of variation	Position	Frequency
T--	Deletion	9959	86.30
TGAAGTGTATAAT	Deletion	27,940	63.40
C-T	SNP (transition)	56,788	97.40
Summary			
Deletions	2		
SNPs	1		
CP genome length	158,058		

Haplotype 4: Zard, Zard F1-70, Zard F1-98, and Zard F1-146.

Nucleotide	Type of variation	Position	Frequency
ATATTTAA	Insertion	4500	75.80
T-G	SNP (transversion)	8125	97.70
T--	Deletion	9959	92.80
TTAAATA	Insertion	10,054	70.60
T--	Deletion	12,522	93.20
T	Insertion	13,402	87.80
C-T	SNP (transition)	17,066	99.50
T--	Deletion	37,693	75.90
C-G	SNP (transversion)	47,415	97.60
G-T	SNP (transversion)	47,672	97.30
TCCATA	Insertion	52,833	98.10
ATAATA-----	Deletion	53,158	83.30
ATTTCA	Insertion	56,658	97.60
C-T	SNP (transition)	56,788	99.80
CC	Insertion	61,068	91.60
T--	Deletion	66,206	98.70
A-C	SNP (transversion)	66,812	92.00
G-A	SNP (transition)	72,116	99.00

Table S2. (Continued.)

TTTTTTT	Insertion	73,666	93.00
A-C	SNP (transversion)	76,910	98.30
T	Insertion	77,238	91.00
T--	Deletion	83,306	90.00
A-G	SNP (transition)	85,987	99.00
C-T	SNP (transition)	86,095	99.70
G-A	SNP (transition)	113,201	99.50
G	Insertion	115,565	92.60
GTTTCTAATTTT	Deletion	115,998	98.80
G-T	SNP (transversion)	116,553	98.80
G-A	SNP (transition)	122,249	99.00
T	Insertion	122,331	76.10
G-T	SNP (transversion)	123,712	99.50
A-G	SNP (transition)	124,312	99.40
G-T	SNP (transversion)	126,252	99.10
T-A	SNP (transversion)	127,877	98.70

Summary	
Insertions	10
Deletions	7
SNPs	17
CP genome length	158,089

Table S3. RSCU analysis results for all the eight apricot CP genomes.

			UUC	1332	0.74		UCC	867	1.13		UAC	716	0.82		UGC	449	0.78
		Leu	UUA	1388	1.48		UCA	966	1.26	TER	UAA	1325	1.31	TER	UGA	934	0.93
			UUG	1097	1.17		UCG	557	0.72		UAG	768	0.76	Trp	UGG	671	1
			CUU	1127	1.2	Pro	CCU	651	1.06	His	CAU	997	1.44	Arg	CGU	410	0.77
			CUC	657	0.7		CCC	596	0.97		CAC	390	0.56		CGC	271	0.51
			CUA	865	0.92		CCA	766	1.25	Glu	CAA	1050	1.31		CGA	541	1.02
Haplotype 1	52685		CUG	510	0.54		CCG	438	0.71		CAG	557	0.69		CGG	422	0.79
		Ile	AUU	1815	1.21	Thr	ACU	635	1.11	Asn	AAU	1841	1.4	Ser	AGU	648	0.84
			AUC	1003	0.67		ACC	581	1.01		AAC	788	0.6		AGC	430	0.56
			AUA	1670	1.12		ACA	696	1.21	Lys	AAA	2163	1.38	Arg	AGA	958	1.8
		Met	AUG	898	1		ACG	381	0.66		AAG	967	0.62		AGG	586	1.1
		Val	GUU	815	1.35	Ala	GCU	480	1.3	Asp	GAU	1102	1.43	Gly	GGU	527	0.96
			GUC	395	0.66		GCC	360	0.98		GAC	441	0.57		GGC	351	0.64
			GUA	777	1.29		GCA	399	1.08	Glu	GAA	1317	1.38		GGA	756	1.38
			GUG	423	0.7		GCG	237	0.64		GAG	593	0.62		GGG	558	1.02
Species	Codon No.	Amino Acid	Codons	No.	RSCU	Amino Acid	Codons	No.	RSCU	Amino Acid	Codons	No.	RSCU	Amino Acid	Codons	No.	RSCU
		Phe	UUU	2356	1.26	Ser	UCU	1068	1.41	Tyr	UAU	1703	1.36	Cys	UGU	723	1.24
			UUC	1387	0.74		UCC	847	1.12		UAC	800	0.64		UGC	442	0.76
		Leu	UUA	1256	1.37		UCA	893	1.18	TER	UAA	1382	1.32	TER	UGA	968	0.92
			UUG	1092	1.2		UCG	575	0.76		UAG	796	0.76	Trp	UGG	714	1
			CUU	1153	1.26	Pro	CCU	633	1.04	His	CAU	967	1.4	Arg	CGU	395	0.74
			CUC	695	0.76		CCC	610	1		CAC	418	0.6		CGC	266	0.5
			CUA	764	0.84		CCA	739	1.21	Glu	CAA	1053	1.32		CGA	549	1.03
			CUG	521	0.57		CCG	451	0.74		CAG	545	0.68		CGG	421	0.79
Haplotype 2	52690																
		Ile	AUU	1871	1.22	Thr	ACU	621	1.09	Asn	AAU	1847	1.41	Ser	AGU	635	0.84
			AUC	1099	0.72		ACC	582	1.02		AAC	782	0.59		AGC	512	0.68
			AUA	1618	1.06		ACA	693	1.21	Lys	AAA	2143	1.34	Arg	AGA	950	1.78
		Met	AUG	886	1		ACG	390	0.68		AAG	1057	0.66		AGG	629	1.18
		Val	GUU	767	1.32	Ala	GCU	442	1.26	Asp	GAU	1062	1.42	Gly	GGU	510	0.97
			GUC	418	0.72		GCC	336	0.96		GAC	436	0.58		GGC	339	0.64
			GUA	718	1.24		GCA	390	1.12	Glu	GAA	1251	1.34		GGA	705	1.34
			GUG	418	0.72		GCG	231	0.66		GAG	615	0.66		GGG	555	1.05
Species	Codon No.	Amino Acid	Codons	No.	RSCU	Amino Acid	Codons	No.	RSCU	Amino Acid	Codons	No.	RSCU	Amino Acid	Codons	No.	RSCU
		Phe	UUU	2367	1.23	Ser	UCU	1135	1.43	Tyr	UAU	1589	1.34	Cys	UGU	766	1.26
			UUC	1490	0.77		UCC	875	1.1		UAC	788	0.66		UGC	450	0.74
		Leu	UUA	1248	1.44		UCA	920	1.16	TER	UAA	1323	1.31	TER	UGA	976	0.96
			UUG	1045	1.2		UCG	606	0.76		UAG	740	0.73	Trp	UGG	700	1
			CUU	1033	1.19	Pro	CCU	649	1.08	His	CAU	921	1.38	Arg	CGU	406	0.73
			CUC	639	0.74		CCC	623	1.04		CAC	411	0.62		CGC	275	0.49
			CUA	781	0.9		CCA	733	1.22	Glu	CAA	1078	1.41		CGA	566	1.02
			CUG	459	0.53		CCG	391	0.65		CAG	454	0.59		CGG	408	0.73
Haplotype 3	5686																
		Ile	AUU	1848	1.21	Thr	ACU	672	1.14	Asn	AAU	1858	1.4	Ser	AGU	756	0.95
			AUC	1095	0.72		ACC	578	0.98		AAC	802	0.6		AGC	478	0.6
			AUA	1644	1.08		ACA	724	1.23	Lys	AAA	2192	1.41	Arg	AGA	1064	1.91
		Met	AUG	900	1		ACG	376	0.64		AAG	923	0.59		AGG	626	1.12
		Val	GUU	798	1.39	Ala	GCU	471	1.24	Asp	GAU	1040	1.44	Gly	GGU	598	1.05
			GUC	403	0.7		GCC	350	0.92		GAC	401	0.56		GGC	365	0.64
			GUA	738	1.29		GCA	461	1.21	Glu	GAA	1205	1.36		GGA	758	1.33
			GUG	356	0.62		GCG	236	0.62		GAG	567	0.64		GGG	558	0.98
Species	Codon No.	Amino Acid	Codons	No.	RSCU	Amino Acid	Codons	No.	RSCU	Amino Acid	Codons	No.	RSCU	Amino Acid	Codons	No.	RSCU
		Phe	UUU	2343	1.26	Ser	UCU	1131	1.48	Tyr	UAU	1726	1.39	Cys	UGU	732	1.28
			UUC	1366	0.74		UCC	823	1.08		UAC	759	0.61		UGC	416	0.72
		Leu	UUA	1200	1.36		UCA	939	1.23	TER	UAA	1390	1.31	TER	UGA	1001	0.94
			UUG	1017	1.15		UCG	593	0.78		UAG	788	0.74	Trp	UGG	659	1
			CUU	1151	1.3	Pro	CCU	684	1.09	His	CAU	1019	1.42	Arg	CGU	358	0.67
			CUC	661	0.75		CCC	603	0.96		CAC	413	0.58		CGC	274	0.51
P. mandshuri	52696		CUA	766	0.87		CCA	758	1.21	Gln	CAA	1074	1.36		CGA	560	1.05
Zard group	52696		CUG	503	0.57		CCG	460	0.73		CAG	510	0.64		CGG	406	0.76
		Ile	AUU	1902	1.25	Thr	ACU	679	1.18	Asn	AAU	1818	1.4	Ser	AGU	639	0.84
			AUC	1066	0.7		ACC	552	0.96		AAC	786	0.6		AGC	458	0.6
			AUA	1587	1.05		ACA	646	1.13	Lys	AAA	2174	1.38	Arg	AGA	991	1.86
		Met	AUG	866	1		ACG	419	0.73		AAG	987	0.62		AGG	606	1.14
		Val	GUU	830	1.36	Ala	GCU	434	1.19	Asp	GAU	1039	1.42	Gly	GGU	541	0.99
			GUC	426	0.7		GCC	344	0.94		GAC	423	0.58		GGC	334	0.61
			GUA	750	1.22		GCA	429	1.17	Glu	GAA	1264	1.34		GGA	758	1.39
			GUG	443	0.72		GCG	255	0.7		GAG	617	0.66		GGG	550	1.01
Species	Codon No.	Amino Acid	Codons	No.	RSCU	Amino Acid	Codons	No.	RSCU	Amino Acid	Codons	No.	RSCU	Amino Acid	Codons	No.	RSCU
		Phe	UUU	2380	1.24	Ser	UCU	1129	1.44	Tyr	UAU	1585	1.34	Cys	UGA	737	1.24
			UUC	1453	0.76		UCC	878	1.12		UAC	786	0.66		UGC	451	0.76
		Leu	UUA	1213	1.4		UCA	968	1.24	TER	UAA	1298	1.27	TER	UGA	1004	0.98
			UUG	1032	1.19		UCG	598	0.76		UAG	766	0.75	Trp	UGG	705	1
			CUU	1064	1.23	Pro	CCU	640	1.07	His	CAU	987	1.44	Arg	CGU	374	0.68
			CUC	667	0.77		CCC	589	0.99		CAC	388	0.56		CGC	255	0.47
			CUA	777	0.9		CCA	748	1.26	Gln	CAA	1051	1.4		CGA	587	1.07
P. mume	52750		CUG	433	0.5		CCG	405	0.68		CAG	450	0.6		CGG	423	0.77
P. zhenghens is																	
		Ile	AUU	1913	1.24	Thr	ACU	655	1.11	Asn	AAU	1825	1.4	Ser	AGU	655	0.84
			AUC	1057	0.69		ACC	596	1.01		AAC	787	0.6		AGC	470	0.6
			AUA	1656	1.07		ACA	712	1.21	Lys	AAA	2164	1.37	Arg	AGA	1007	1.84
		Met	AUG	832	1		ACG	388	0.66		AAG	984	0.63		AGG	632	1.16
		Val	GUU	824	1.36	Ala	GCU	448	1.2	Asp	GAU	1045	1.43	Gly	GGU	539	0.98
			GUC	431	0.71		GCC	359	0.96						GGC	351	0.64
			GUA	759	1.25		GCA	430	1.15						GGA	770	1.4
			GUG	406	0.67		GCG	261	0.7						GGG	546	0.99
Species	Codon No.	Amino Acid	Codons	No.	RSCU	Amino Acid	Codons	No.	RSCU	Amino Acid	Codons	No.	RSCU	Amino Acid	Codons	No.	RSCU
		Phe	UUU	2364	1.29	Ser	UCU	1168	1.48	Tyr	UAU	1558	1.4	Cys	UGA	741	1.25
			UUC	1299	0.71		UCC	862	1.09		UAC	673	0.6		UGC	445	0.75
		Leu	UUA	1317	1.45		UCA	979	1.24	TER	UAA	1003	1.3	TER	UGA	953	0.95
			UUG	1049	1.15		UCG	598	0.76		UAG	740	0.74	Trp	UGG	666	1
			CUU	1135	1.25	Pro	CCU	650	1.05	His	CAU	970	1.45	Arg	CGU	407	0.76
			CUC	645	0.71		CCC	606	0.98		CAC	366	0.55		CGC	271	0.5
			CUA	847	0.93		CCA	807	1.3	Gln	CAA	1039	1.34		CGA	562	1.04
P. sibirica	52724		CUG	458	0.5		CCG	411	0.66		CAG	515	0.66		CGG	442	0.82
		Ile	AUU	1914	1.24	Thr	ACU	696	1.16	Asn	AAU	1859	1.41	Ser	AGU	658	0.83
			AUC	991	0.64		ACC	600	1		AAC	772	0.59		AGC	477	0.6
			AUA	1710	1.11		ACA	714	1.19	Lys	AAA	2139	1.38	Arg	AGA	902	1.81
		Met	AUG	876	1		ACG	380	0.64		AAG	957	0.62		AGG	574	1.07
		Val	GUU	824	1.38	Ala	GCU	493	1.31	Asp	GAU	1134	1.46	Gly	GGU	541	0.97
			GUC	389	0.65		GCC	360	0.96			421	0.54		GGC	359	0.64
			GUA	773	1.29		GCA	418	1.11			1339	1.4		GGA	765	1.37
			GUG	406	0.68		GCG	230	0.61			573	0.6		GGG	562	1.01

Figure S1. Chloroplast map of Zard.

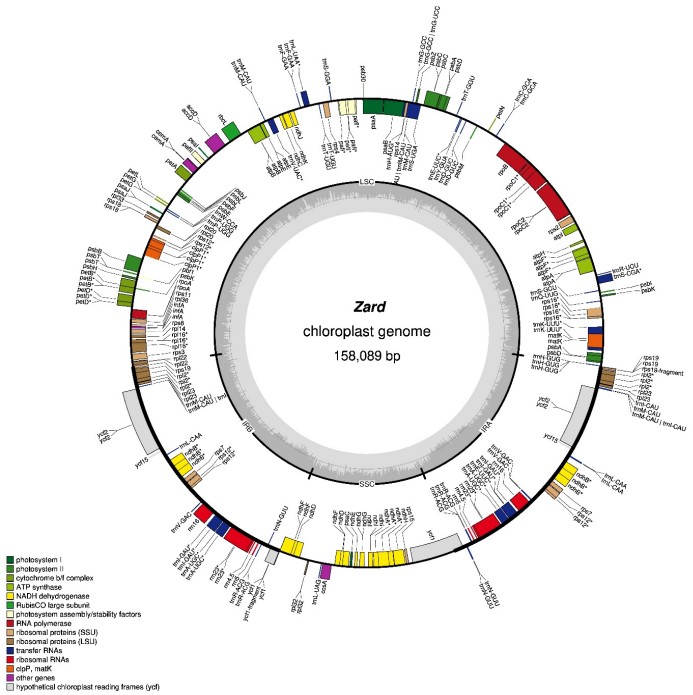

Figure S2. IR comparative analysis of all the *Prunus* species (excluding apricot genomes) from NCBI.