T.C.
ERCİYES UNIVERSITY
SCHOOL OF NATURAL AND APLLIED SCIENCE
DEPARTMENT OF AGRICULTURAL BIOTECHNOLOGY

COMPLETE CHLOROPLAST GENOME SEQUENCE AND PHYLOGENETIC ANALYSIS OF COMMON APRICOTS (*Prunus armeniaca L.)* AND COMPARATIVE ANALYSIS WITH OTHER *Prunus* CHLOROPLAST GENOMES

Prepared by
Lungelo KHANYILE

Supervisor
Assoc. Prof. Dr. Kahraman GÜRCAN

Master Thesis

July 2023
KAYSERİ

T.C.
ERCİYES UNIVERSITY
SCHOOL OF NATURAL AND APLLIED SCIENCE
DEPARTMENT OF AGRICULTURAL BIOTECHNOLOGY

COMPLETE CHLOROPLAST GENOME SEQUENCE AND PHYLOGENETIC ANALYSIS OF COMMON APRICOTS (*Prunus armeniaca L.*) AND COMPARATIVE ANALYSIS WITH OTHER *Prunus* CHLOROPLAST GENOMES

Prepared by
Lungelo KHANYILE

Supervisor
Assoc. Prof. Dr. Kahraman GÜRCAN

This project was supported by The Scientific and Technological Research Council of Türkiye (Project number: 214O400 and 119O846)

June 2023
KAYSERİ

COMPLIANCE AND SCIENTIFIC ETHICS

I declare that all information in this study has been obtained in accordance with academic and ethical rules. At the same time, I declare that I have fully cited and referenced all materials and results that are not at the core of this work, as required by these rules and conduct.

Lungelo KHANYILE

İmza

COMPLIANCE WITH THE DIRECTIVE

The Master's Thesis titled **"Complete chloroplast genome sequence and phylogenetic analysis of common apricots (Prunus armeniaca L.) and comparative analysis with other Prunus chloroplast genomes."** has been prepared in accordance with Erciyes University Graduate Thesis Proposal and Thesis Writing Directive.

Prepared by
Lungelo KHANYILE
İmza

Supervisor.
Assoc. Prof. Dr. Kahraman GÜRCAN
İmza

Head of the Agricultural Biotechnology Department

Prof. Dr. Semih YILMAZ
İmza

ACKNOWLEDGEMENTS

First and foremost, I want to thank the almighty God for guiding and protecting me through my studies. Secondly, I want to thank me, I want to thank me for surviving in a foreign country, I want to thank me for sacrificing everything for this degree and not giving up.

I would also like to express my gratitude to my supervisor, Assoc. Prof. Dr. Kahraman GÜRCAN for supporting me through my degree. Even though at times we could not see eye to eye, his professionalism and father figure ensured the success of my studies. I would forever be grateful for that. I also thank YTB for funding my studies, without their support, I would not be here.

In addition, I want to thank my family for their endless support, thank you so much boKhanyile. I would also like to thank all the Africans in Kayseri for creating a warm environment, which made me to feel home. I cannot leave out my family away from home: Mandla Ndlangamandla, Bukelwa Maphanga, Lindokuhle Mkhatshwa, Bonginkhosi Dlamini, Lexie Dlamini, Bupe Kasanya, Gary Masuku, Chennai, and all the other Sourthen Africans.

Lastly, I want to thank my lab mates, Duygu, Merve, Izem, Ilyas, Dooa, Maryam, Samuel, Şuheda, Şeyda, Brian, Musab, Faduma, Sibel, Adil, and everyone else from GENKÖK for creating a warm environment for conducting this research.

Lungelo KHANYILE

July 2023, KAYSERİ

KAYILARININ (*Prunus armeniaca L.)* TAM KLOROPLAST GENOM DİZİSİ VE FİLOGENETİK ANALİZİ VE DİĞER *Prunus* KLOROPLAST GENOMLARI İLE KARŞILAŞTIRMALI ANALIZ

Lungelo KHANYILE

Erciyes Üniversitesi, Fen Bilimleri Enstitüsü
Yüksek Lisans Tezi, Haziran 2023
Danışman: Doç. Dr. Kahraman GÜRCAN

ÖZET

Bu çalışmada, kültür kaysısı (*Prunus armeniaca* L.*)* kloroplast (CP) genomunun mutasyon özelliklerini ve genom yapısını belirlemek amacıyla, toplam DNA'dan kloroplast nükleotid okumaları filtrelenip *de novo* birleştirilerek, 30 kaysı CP genomu elde edilmiştir. Kayısıların CP genomları, GenBank'tan alınan *Prunus* cinsine ait 30 CP genomu ile birlikte analiz edilmiştir. Tüm 52 CP genomunda, genomik karşılaştırma ve filogenetik analiz gerçekleştirilmiştir. 30 kayısı CP genomunun uzunlukları, 158 057 bp ile 158 089 bp arasında değişmekte olup, genomik yapının, gen organizasyonu ve sırasının korunduğu belirlenmiştir. İlginç bir şekilde, tüm kayısılar sadece üç haplotip ve yakın ilişki gösterirken, plum pox virüsüne dayanıklı "Zard" ve melezleri, Mandurya kayısı (*P. mandshurica* Maxim*)* ile gruplandırılmıştır. CP genomu filogenetik analizi kayısı türlerinin morfolojik gruplanmalarından farklı gruplandırmıştır.ı; kültür ve Mandurya ve Japon kaysısı (*P. mume* Sieb.) aynı grupta yer alırken Sibirya kaysısı (*P. sibirica* L.), Japon eriği (*P. cerasifera* Ehrh.) ile ayrı bir grup oluşturmuştur. Bu tez çalışmasında üretilen CP filogenomik sonuçlar, kayısı türlerinin filogenetik ilişkilerine yeni ve önemli perspektifler sunmakta ve *Prunus* evrimi üzerine gelecekteki karşılaştırmalı çalışmalar için bir çerçeve oluşturmaktadır.

Anahtar Kelimeler: *Prunus armeniaca*, Filogenomik, Kloroplast genomu, Maternal kalıtım

COMPLETE CHLOROPLAST GENOME SEQUENCE AND PHYLOGENETIC ANALYSIS OF COMMON APRICOTS (*Prunus armeniaca L.)* AND COMPARATIVE ANALYSIS WITH OTHER *Prunus* CHLOROPLAST GENOMES

Lungelo KHANYILE

Erciyes University, Graduate School of Natural and Applied Sciences
Master Thesis, June 2023
Supervisor: Assoc. Prof. Dr. Kahraman GÜRCAN

ABSTRACT

In this study, to better reveal the mutation characteristics and evolution patterns of the common apricot (*Prunus armeniaca* L.) chloroplast (CP) genome, the complete CP genomes from 30 apricot accessions were *de novo* assembled filtering from total DNA. The complete CP genomes of the apricots were analyzed together with 30 CP genomes from the *Prunus* genus retrieved from the GenBank. All 52 CP genomes were used for genomic comparison and phylogenetic inference. The lengths of the CP genome of the 30 apricot CP genomes ranged from 158 057 nt to 158 089 nt, exhibiting conservation of genomic structure, and gene organization and order. Interestingly, while all apricots showed only three haplotypes and close relation, "Zard" and its offsprings, known as common apricot and plum pox virus resistant, was grouped with *P. mandshurica.* Based on the CP genome evolution analysis, apricot species did not reflect their morphological systematics grouping; while common apricot, Manchurian apricot (*P. mandshurica* Maxim) and Japanese apricot (*P. mume*) were grouped as adjacent clades, Siberian apricot (*P. sibirica* L.) took distant position together with Japanese plum (*P. cerasifera* Ehrh.) on the phylogenetic tree. The CP phylogenomics results suggest new and important insights into phylogenetic relationships of the apricot species and serve as a framework for future comparative studies on *Prunus* evolution.

Keywords: *Prunus armeniaca*, Phylogenomics, Chloroplast genome, Maternal inheritance

TABLE OF CONTENTS

COMPLETE CHLOROPLAST GENOME SEQUENCE AND PHYLOGENETIC ANALYSIS OF COMMON APRICOTS (*Prunus armeniaca L.)* AND COMPARATIVE ANALYSIS WITH OTHER *Prunus* CHLOROPLAST GENOMES

ABBREVIATIONS

CP : Chloroplast

DNA : Deoxyribonucleic acid

SEO : Stark Early Orange

OB : Ordubat

tRNA : Transfer Ribonucleic Acid

rRNA : Ribosomal Ribonucleic Acid

SNP : Single Nucleotide Polymorphism

In : Insertions

Del : Deletions

LSC : Large Single Copy region

SSC : Small Single Copy region

IR : Inverted Repeats

RPS : Ribosomal Proteins

RSCU : Relative Synonymous Coding Usage

PPV : Plum Pox Virus

RFLP : Restriction Fragment Length Polymorphisms

RAPD : Random Amplified Polymorphic DNA

SSRs : Simple Sequence Repeats

AFLP : Amplified Fragment Length Polymorphism

OM : Optical Mapping

USA : United States of America

BWA : Burrows Wheel Aligner

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

DNA variations in CP genomes have been used since the 1980s to study the relationships of species because the CP provides a lot of important information that helps in the identification of species and phylogeny analyses because it has a low evolutionary rate, uniparental inheritance, a low rate of nucleotide replacement, almost neutral evolution, and no recombination [1], [2]. The first report on the CP genome of plants was published on Liverwort, and since then more complete CP genomes of plants have been sequenced [3]. In recent years, the sequencing of CP genomes has increased remarkably due to advanced sequencing techniques and reduced costs [4], [5]. As a result of the increased pace of CP genome sequencing, researchers have sequenced many angiosperms for phylogenomic analysis [6]–[8]. Among the angiosperms, Rosaceae is one of the most diverse families, and due to multiple rapid radiations and reticulation events, studying phylogenetic relationships in this family is a major challenge [9].

The *Rosaceae* family consists of over 120 genera and more than 3300 species of economic importance. To understand the diversity of the Rosaceae family, many phylogenetic studies have been conducted on several *Malus domestica* species, the *genus Firago*, *Rubus accidentalis* species, and *Prunus* species[10]–[14]. The Rosaceae family is divided into four subfamilies in accordance with their fruit type: *Prunoideae*, *Spiraeoideae*, *Maloideae*, and *Rosoideae* [15], [16]. *Prunus* is a diverse genus under the Prunoideae subfamily, including approximately 200 species, with most of them growing in the temperate zone and some in the tropical and subtropical regions. This genus is economically important due to its diverse uses as fruit (plums, peaches, apricots, cherries, and almonds), oil, timber, and ornamentals [17], [18]. The chloroplast DNA (cpDNA) has been used since the 1990s to study the phylogenetic relationship of *Prunus* species [1] and among the *Prunus* genus, the CP genome of the *Prunus persica* species was one of

the first completely sequenced genomes [19]. CP genomes of other *Prunus* species such as *Prunus yedoensis*, *Prunus serrulate*, *Prunus mume*, and *Prunus subhirtella* were then later sequenced using the *Prunus persica* CP genome as a reference [20], [21]. In the *Prunus* genus, the CP genome has been used to study phylogenetic relationships between species such as *Prunus cerasus*, *Prunus phaeosticta, Prunus dulcis*, *Prunus zhengheensis, Prunus mume,* and Tunisian *Prunus armeniaca* [15], [22]–[25]. Wang et al. [22] also studied the phylogenetic relationships of six almond species using CP genome comparative analysis. In addition, Coulibaly et al. [27] also did comparative studies among 10 Japanese apricot (*Prunus mume*) accessions using CP genomes. Phylogenomic studies have been an insurmountable problem in the *Prunus* genus, and in this research, the phylogenetic relationship study of the *Prunus armeniaca L.* accessions that are available around Türkiye, as well as other Prunus species will be conducted.

The common apricot*, Prunus armeniaca L.,* which is a diploid species with eight pairs of chromosomes (2n = 16), is economically important in many countries, and its phylogenetic studies are of high importance. The structure of the CP genome of *Prunus armeniaca* has been previously described in recent studies, and the size of the genome is estimated to be about 157,797 to 158,138 bp [15], [28]. The CP genome of *Prunus armeniaca* has also been used among *Prunus mume* and *Prunus salicina* for comparative analysis in recent studies [15]. Li et al. [29] also used the CP DNA and nuclear ribosomal sequences to investigate the phylogeography of *Prunus armeniaca* L. In a recent study, phylogenetic relationship analysis of Turkish *Prunus armeniaca L*. genotypes was conducted with the use of RAPD-PCR, ISSR-PCR, and cpDNA sequence analyses, but only a few accessions were used [30]. Based on our knowledge, no comparative analysis of *Prunus armeniaca* accessions has been done using CP genome analysis. Therefore, in this study, the phylogenetic analysis of *Prunus armeniaca* accessions will be studied using CP genomes to further understand the phylogenetic relationship among the *Prunus armeniaca* species.

The inheritance of CP genomes is also a very interesting topic of discussion because various studies have shown that the CP genome shows non-Mendelian inheritance properties [31]. Angiosperm CP genomes generally show maternal inheritance, while conifers sometimes show parental CP inheritance [32]–[34]. Even though maternal

inheritance is common in CP genomes, alternative inheritance patterns have been reported in some chloroplast studies, which is why in this study the maternal inheritance of the apricot CP genomes will be confirmed by including F1 hybrids in the study.

In this study, the complete CP genomes of 20 apricot accessions and 10 F1 hybrids were *de novo* assembled for phylogenetic analyses. To enhance the phylogenetic study, 16 CP genomes from the *Prunus* genus recruited from GenBank, were also included in this analysis. The objectives of this study were to 1) phylogenetically characterize the 30 assembled genomes and phylogenetically classify *Prunus* species; 2) confirm the chloroplast inheritance patterns of the *Prunus armeniaca* species; and 3) explore variations in the whole CP genomes, especially for “Zard” and “Hacihaliloglu”. The results of this study will serve as a reference for future genome-scale phylogenetic studies of *Prunus*.

1. CHAPTER

LITERATURE REVIEW

1.1. Taxonomic classification of apricots

Apricot belongs to the Rosaceae family under the subgenus Prunuphora Focke and genus *Prunus* L. [35]–[37]. It is classified under the Armeniaca (Lam) section [38]. The number of apricot species ranges from 3 to 12 depending on the classification system used, and all of them are diploids with eight pairs of chromosomes (2n = 16, x = 8) [38], [39]. The most recognised apricot species are: *P.brigantina* V., *P.mandshurica* (Maxim), *P.sibirica* L., *P.mume* (Sieb.), *P.holosericeae* Batal., and the important one in this study, *P.armeniaca* L. [38], [40]. These apricot species can be intercrossed in either direction, which makes the phylogenetic classification of apricots confusing [41]. The botanic classification of apricots is shown below (Table 1.1).

Table 1.1. Botanic classification of the apricot species

Category	**Class**
Kingdom:	Plantae
Clade:	Angiosperms
Clade:	Eudicots
Clade:	Rosids
Order:	Rosales
Family:	Rosaceae
Genus:	*Prunus*
Subgenus:	*Prunus*
Section:	*Prunus* sect. *Armeniaca*
Species:	*P. armeniaca*
Binomial name:	*Prunus armeniaca L.*

1.2. Origin and Domestication of apricots

1.2.1. Origin of apricots

There are 3 centres of origin for the apricot species: Chinese center, Central Asiatic center, and Near-Eastern center. The Chinese centre comprises central and western China with the inclusion of the lowlands in China [42]. The Central Asiatic centre includes countries such as Afghanistan, India, Uzbekistan, and Pakistan, while the Near-Eastern centre includes countries such as Türkiye and Iran [42], [43]. Apricots were then dispersed to the Middle East, Egypt, and North Africa in countries such as Algeria and later to Spain [44].

1.2.2. Domestication of apricots

Domestication of living plants is a significant source of evolutionary transitions in plants, and many plant traits have been modified during the process of plant domestication [39], [45], [46]. Apricots, like any other plant, have been domesticated for a long period of time, and previous studies showed that apricots were first domesticated in Central Asia 3000 years ago [42], [47], [48]. The domestication of apricots has spread worldwide, and Figure 1.1 demonstrates the domestication of apricots in different countries. Recently molecular data based on SSR markers and genome sequencing revealed highest levels of diversity in Central Asian and Chinese wild and cultivated apricots, confirming an origin in this region [46], [49], resulting from independent domestication events from distinct wild Central Asian populations, and with subsequent gene flow [39]. Most apricot species are domesticated for fruits, while some are domesticated for ornamental purposes [38]. The annual production of apricots is estimated to be around 4.1 million tonnes [50], [51]. Türkiye is the major producer of apricots, and other countries like Iran and Uzbekistan are also major producers [39].

Figure 1.1. Global domestication of apricots [52].

1.2.2.1. Cultivation of apricot in Türkiye

Türkiye is divided into 9 agro-ecological regions, and has a diverse environment with mountains, rivers, lakes, and valleys [22]. The landscape and diverse climatic conditions make Türkiye to be a centre of origin and diversity for many fruits such as almonds and apples [22]. Türkiye produces a lot of stone fruits and is the leading producer of apricots globally, which is why the domestication of apricots in Türkiye is of interest. In Türkiye, apricots were first cultivated in Anatolia about 2000 years ago during the times of Alexander the Great, and thus Anatolia became another homeland for apricots [22], [53]. In Türkiye, apricots are grown throughout the country. Malatya is the major producer of apricots in Türkiye, and other cities like Erzincan, Aras, Sivas, Kayseri, and Niğde also produce significant amounts of apricots.

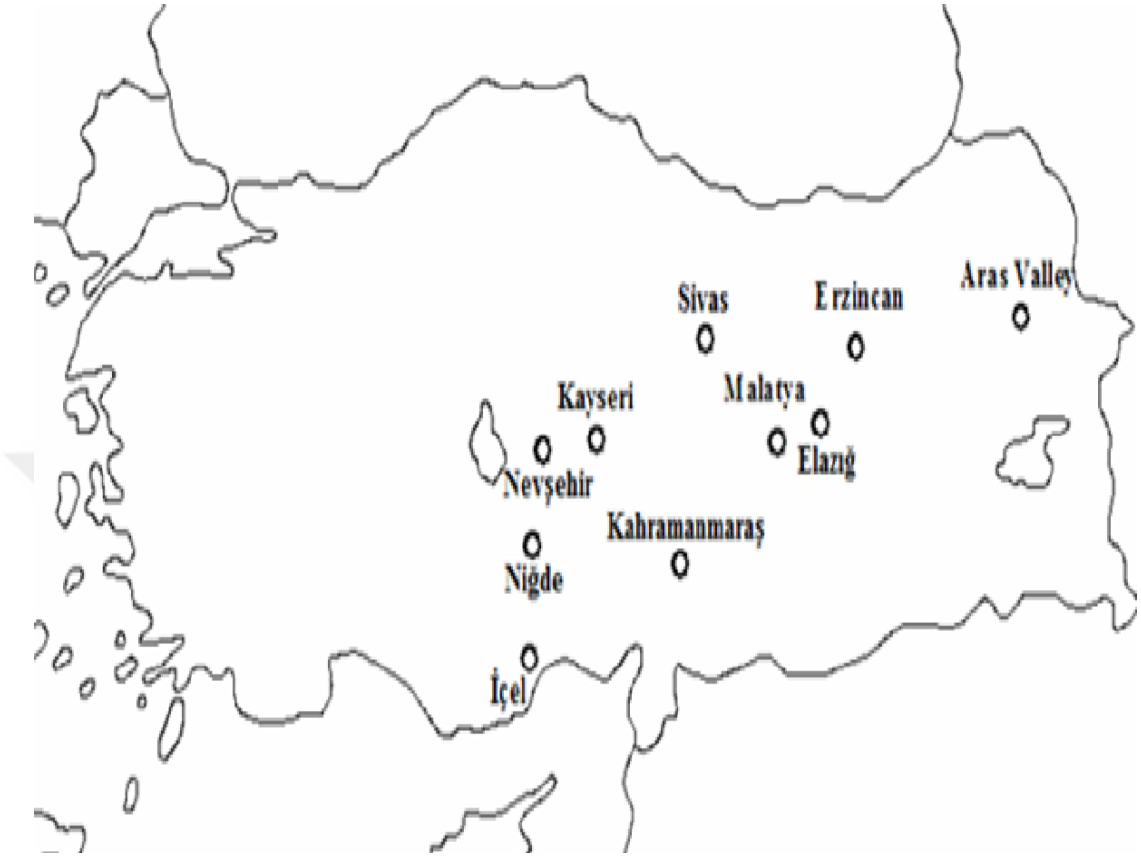

Figure 1.2. Regions of high apricot domestication in Türkiye [22]

1.3. Apricots breeding practices.

Apricot traits are qualitative because they show polygenic inheritance patterns [41]. Many breeders choose parents based on their phenotype characteristics for genetic improvement of the species [54]. In apricot breeding, intraspecific hybridisation is common, whereas interspecific breeding among *Prunus* species is usually used for rootstocks. Apricot breeding was initially done to improve fruit quality but nowadays it is primarily done to develop hybrids that are resistant to sharka disease, which is caused by Plum Pox Virus (PPV) [41], [55].

Sharka disease causes major losses in apricot production and as a result, breeders hybridising to develop resistant cultivars, thus increasing the number of apricot cultivars. Breeders have developed many molecular markers for apricots while in the

process of preventing PPV. Markers such as restriction fragment length polymorphisms (RFLPs), randomly amplified polymorphic DNAs (RAPDs), simple sequence repeats (SSRs), and amplified fragment length polymorphisms (AFLPs) have been developed [37], [56]–[58]. Many cultivars have been developed, thus increasing the need for phylogenetic studies to study the relationship between these cultivars.

1.4. Apricots cultivars in Türkiye

Türkiye, as a leading producer of apricots, has different apricot cultivars that are genetically different. Hacihaliloğlu is the most common Turkish apricot, and it is followed by Kabaasi. These two apricots are mostly grown in Malatya [59]. Other varieties such as Hasanbey, Cataloglu, Soganci, and Cologlu are also common in Malatya [59]. Tokaloglu, Zerdali, and Alyanak varieties are also found in other regions of Türkiye [60], [61]. The following apricot accessions are also found in Türkiye, but originated from other countries: Harleyne, Stack Early Orange (SEO), Zard, and Roxana.

These apricot varieties have different characteristics. Hacihaliloğlu is known to be sweet, and can be dried, but it is susceptible to diseases such as sharka disease [62], [63]. A study by [63] revealed that accessions such as SEO, Zard, Roxana, and Harlayne are resistant to sharka disease. These differences raise the need to study the phylogenetic relationships of these accessions so that breeders can plan breeding practices. When studying phylogenetic relationships in plants, it is better to use chloroplast genomes because they are conserved [64].

1.5. Phylogenetic studies in Apricots

The phylogeny problems in apricot have existed for a long period of time, and researchers are still trying to find solutions to this problem, while other researchers have already published their results regarding this problem. This problem is being tackled by many researchers using different approaches. [65] stated that morphological, biochemical, and DNA-based markers have been used to study variation in plants, and out of all these markers, DNA-based markers are the best for genetic variation studies and nowadays are

preferred for plant variety identification [30]. DNA-based markers include AFLP, RFLP, SSRs, and randomly amplified polymorphic DNA (RAPD) [30], [66].

AFLP markers have been successfully used to identify and classify cultivars in apricots [30]. In terms of variety identification, AFLP markers presented the best results when compared to markers such as SSRs and RFLPs [37], [67]. According to [30], AFLP markers and SSRs present a clear phylogenetic relationship among apricot varieties. [68] reported that RAPD markers are more advantageous because they require a low amount of DNA, give fast results, and present a higher polymorphism frequency. The CP DNA-based markers have presented good results, but that is not the same with nuclear DNA-based phylogenetic studies because they are restricted by their complexity. The nuclear DNA is not conservative and has high rates of mutation, which can negatively affect phylogenetic studies in plants [69], [70]. As a result, nowadays phylogenetic studies in plants are chloroplast-based because the chloroplast is more conservative and can give us unbiased phylogenetic results [2], [15], [71], [72]. Despite the high sequence conservation characteristics of the CP genome, there are small differences in gene content between different species, like the loss of the *ndh* gene in some species, and these small differences can be used to enhance phylogenetic studies [73].

1.6. Chloroplast genome

The chloroplast (CP) is a metabolically active, semi-autonomous organelle found in algae, cyanobacteria, and plants [74]. CP are intracellular organelles in plants that act as the main site for photosynthesis and adenosine triphosphate (ATP) transformation in plants [75]. In the process of photosynthesis, the CP is involved in the conversion of light energy into chemical materials that can be used in the synthesis of organic compounds [74]. It also plays a big role in the biosynthesis of nucleotides, lipids, starch, and amino acids [75], [76]. Many of the metabolites synthesized by the CP are important for coordinating communication between the different tissues of the plant, and they also play a major role in plant reactions to both biotic and abiotic stresses [72], [74]. The CP has an independent circular genome that is different from the main nuclear genome, and it is smaller than the nuclear genome because it has endosymbiotic gene transfer characteristics [77]. The CP genome consists of four main parts: protein-coding

sequences, LSC sequences, IR conserved sequences, and genome-wide sequences [78]. This genome usually consists of 110–130 genes, which regulate photosynthesis, transcription, translation, and other processes. These genes usually have sizes ranging from 120 to 160 kilobases in plants [4], [75]. In most flowering plants, the CP genome is usually maternally inherited, even though inheritance in some plants, such as conifers, is paternal [34], [79].

1.6.1. Chloroplast genome assembly

The CP genome, located within the chloroplasts of plant cells, plays a crucial role in photosynthesis and the production of energy for the cell. The CP genome is a small, circular, and double-stranded DNA molecule that encodes for essential photosynthetic and metabolic functions. The assembly of the CP genome is important for understanding the evolution and diversity of photosynthetic organisms as well as for the improvement of crop plants through genetic engineering [78], [80], [81]. In 1986, the first CP genome was assembled from tobacco and liverwort [82], [83]. Since then, more chloroplast genomes have been assembled for phylogenetic studies and comparative analyses in plants [81]. In 2020, about 3721 CP genomes from different plant species, including both aquatic and terrestrial plants, were deposited into the NCBI (National Centre for Biotechnology Information) organelle database.

Recent advancements in next-generation sequencing (NGS) technology have led to significant improvements in the assembly of chloroplast genomes. One of the most popular methods for CP genome assembly is the use of the Illumina platform, which generates millions of short reads that can be used to assemble the genome. However, the high cost and the need for a high coverage of the genome are some of the limitations of this method. Short reads also make it hard for bioinformaticians to assemble long contigs for a full presentation of the genome.

An alternative to Illumina is the use of long-read sequencing platforms such as PacBio and Oxford Nanopore. These platforms produce long reads that can span multiple introns and repeats, which enables the accurate assembly of the chloroplast genome. Illumina sequencing is sometimes preferred because it allows the use of rolling circle amplification products and thus giving a better assembly [84]. A study by Wang et al. [85] compared

the use of Illumina and PacBio for the assembly of chloroplast genomes and found that PacBio produced higher-quality assemblies with fewer errors and gaps. The longer reads enable the assembly of long contigs, thus giving a more accurate genome.

Another method for CP genome assembly is the use of a combination of NGS and Sanger sequencing. This approach, known as the "hybrid" method, involves generating short-read NGS data and then using Sanger sequencing to fill in any gaps or errors in the assembly. A study by Wu et al. [86] used this method to assemble the CP genomes of wild rice and found that it produced high-quality assemblies with a high level of accuracy.

Recently, a new approach called Optical Mapping (OM) has been used to assemble chloroplast genomes. OM is based on the use of high-resolution fluorescent imaging to create a physical map of the genome, which can be used to assemble the genome. OM produces a high-quality assembly with a high level of accuracy and can be used for structural variation detections [87].

Bioinformaticians have been trying to develop chloroplast assembling tools to improve the efficiency and accuracy of assembling CP genomes. A lot of tools have been developed specifically for organelle genomes, including the CP genome. Tools such as GetOrganelle, NOVOplasty, Genieous Prime Assembler, ORG.Asm, IOGA, and chloroExtractor have been used to assembly chloroplast genomes [88]–[92]. Even though a lot of tools have been developed, it is still very difficult to generate a complete CP genome using these tools [93]. The existence of different chloroplasts within a single individual plant results in a phenomenon called heteroplasmy, which complicates CP genome assembly [49], [94], [95].

The assembly of the CP genome is an important task for understanding the diversity and evolution of photosynthetic organisms. The use of long-read sequencing platforms such as PacBio and Oxford Nanopore, the hybrid method, and Optical Mapping (OM) are some of the best methods for assembling CP genomes.

It is also important to note that there are two main types of genome assembly: the reference-based approach and the *de novo* approach. In the reference-based approach, a reference genome of that species is used to guide the assembly of the new genome, while in *de novo* assembly, the genome is assembled without the use of reference genomes [96].

De novo assembly uses a process whereby individual sequence reads are merged to form long contiguous sequences called contigs, and the contigs share the same nucleotide sequence as the original DNA template from which the sequence reads were derived [96]. In this study, both methods were used, and the best CP genome was chosen.

1.6.2. Importance of CP genome assembly

In breeding, success is usually determined by genetic compatibility, and CP genomes serve as an important tool for identifying plant species that are closely related and therefore genetically compatible [81]. CP genomes help in evolutionary studies and systematic phylogenetic classification of plants [78], [80]. CP genomics also helps in other applications such as the prevention of biotic and abiotic diseases, the development of vaccines, and biopharmaceuticals in plants. This further helps in understanding important traits in plants that enhance breeding practices of closely related species to conserve those important traits. CP genomes can also be used for barcoding and meta-barcoding [97], [98]. In summary, an assembled CP genome can provide researchers with important information on plant molecular ecology, population genetics, evolution, and phylogeny. That is why this study is important in the study of apricots in Türkiye.

2. CHAPTER

MATERIALS AND METHODS

2.1. Plant materials

In this thesis, within the scope of TUBITAK supported projects 214O400 and 119O846, 30 apricot DNA reads produced by high throughput sequencing were studied. Among these 30 apricots, 20 apricot accessions (Table 2.1) were in Erciyes University (ERU) in Kayseri, and the remaining 10 apricot hybrids were produced in ERU (Table 2.2). Additionally, 30 *Prunus* CP genome were downloaded from NCBI (Table 2.3).

Table 2.1. Origin and features of the 20 apricot accessions subjected to high throughput sequencing within the scope of TUBITAK supported projects 214O400 and 119O846.

Accession *(P. armeniaca)*	**Origin**
Sakit	Türkiye
M2249	Türkiye
M2252	Türkiye
Ordubat	Türkiye
Hacikaya	Türkiye

Kabaasi	Türkiye
Harlayne	Canada rooted back to Central Asia
Zerdali	Türkiye
M2243	Türkiye
M2244	Türkiye
KZ1	Türkiye
KZ44	Türkiye
Roxana	Central Asia
Kirmizi	Türkiye
Hasanbey	Türkiye
SEO (Stark Early Orange)	North America rooted back to Central Asia
Hacihaliloğlu	Türkiye
Fracasco	Türkiye
Zard	North America rooted back to Central Asia
Harcot	North America

Table 3. The apricot hybrids produced at Erciyes University in Kayseri, Türkiye.

Hybrid (*P. armeniaca*)	**Origin**	**Parents (Maternal * Paternal)**
Hacihaliloglu F1 3	Türkiye	Hacihaliloglu * SEO
Hacihaliloglu F1 4	Türkiye	”
Hacihaliloglu F1 16	Türkiye	”
Hacihaliloglu F1 17	Türkiye	”
Hacihaliloglu F1 18	Türkiye	”
Hacihaliloglu F1 19	Türkiye	”
Hacihaliloglu F1 11	Türkiye	”
Zard F1 70	Türkiye	Zard F1 offspring
Zard F1 98	Türkiye	Zard F1 offspring
Zard F1 146	Türkiye	Zard F1 offspring

Table 4**.** The 30 *Prunus* CP genomes downloaded from NCBI with their accession numbers.

Name	GenBank accession	Country
P. dulcis	KY085904.1	USA
P. mume	KF765450.1	CHİNA
P. yedoensis	KU985054.1	South Korea
P. persica	HQ336405.1	USA
P. mandshurica R.	NC_068703.1	CHINA
P. mandshurica	MK905681.1	CHINA
P. armeniaca R.	NC_043901.1	USA
P. cerasifera	NC_068605.1	USA
P. fasciculata	NC_054253.1	USA
P. mahaleb	MT576896.1	CHINA
P. pedunculata	MG602257.1	CHINA
P. tomentosa	MT576919.1	CHINA
P. triloba	MH748555.1	CHINA
P. sibirica	NC_068704	CHINA
P. kanseusis	NC_023956.1	USA
P. mongolica	NC_037849.1	USA
P. tangutica	MZ145044.1	CHINA

P. serrulata	NC_066418.1	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

2.2. Chloroplast genome assembly

To get a high-quality assembly, high quality raw reads were *de novo* assembled using two different tools. Firstly, the Geneious assembler on Geneious Prime (https://www.geneious.com/) was used to *de novo* assembly the sequences. The raw reads were first mapped into the reference genome of apricots (GCA_903114435.1) and then it was assumed that the unused reads represents the organelle genomes. The unused reads

were further mapped to the apricot mitochondria reference genome (NC_065228.1) to remove mitochondria sequences. It was assumed that the remaining unused reads were related to the CP genome. The unused reads were then *de novo* assembled and the top 100 longest contigs were mapped to the apricot CP reference genome (NC043901.1) to get a full CP genome assembly (Figure 2.1). BWA (Burrows Wheeler Aligner) with default parameters was used for mapping [99]. In that way, complete CP genomes of the 30 apricot accessions were assembled from Geneious Prime.

Secondly, NOVOPlasty with default parameters was used to assemble the CP genomes [91]. Firstly, the assembly was conducted using two different Kmer values, 33 and 39. Kmers are substrings of a specific length contained within a biological sequence. Kmer 33 gave a better assembly and then all the assemblies were conducted using Kmer 33. The first 100 nucleotides of the apricot CP reference genome were used as a seed when conducting the assembly. An example of the configuration text used in the assembly is shown in Figure 2.2. The first 2 longest contigs were assembled into a CP genome.

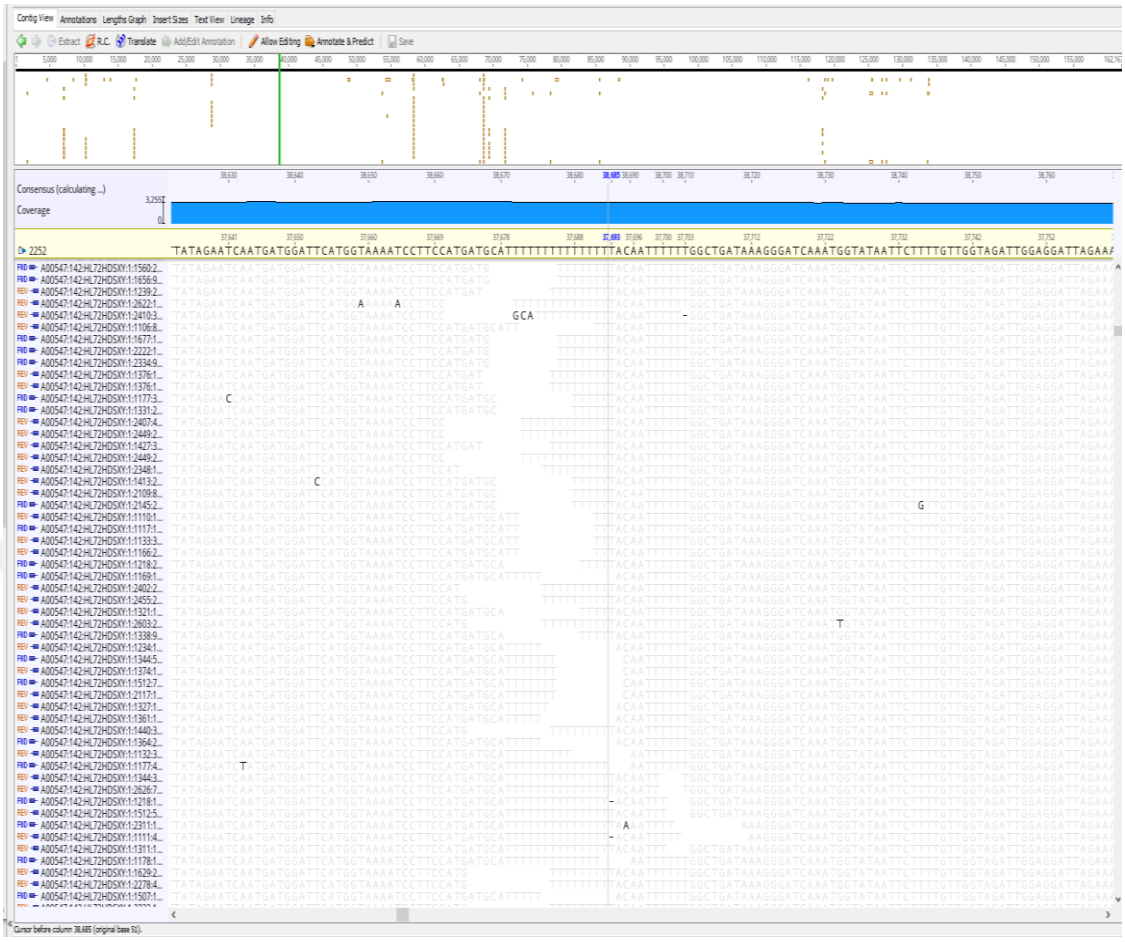

Figure 2.1. CP genome assembly in Geneious Prime Assembler, the long contigs were perfectly aligned to the reference genome.

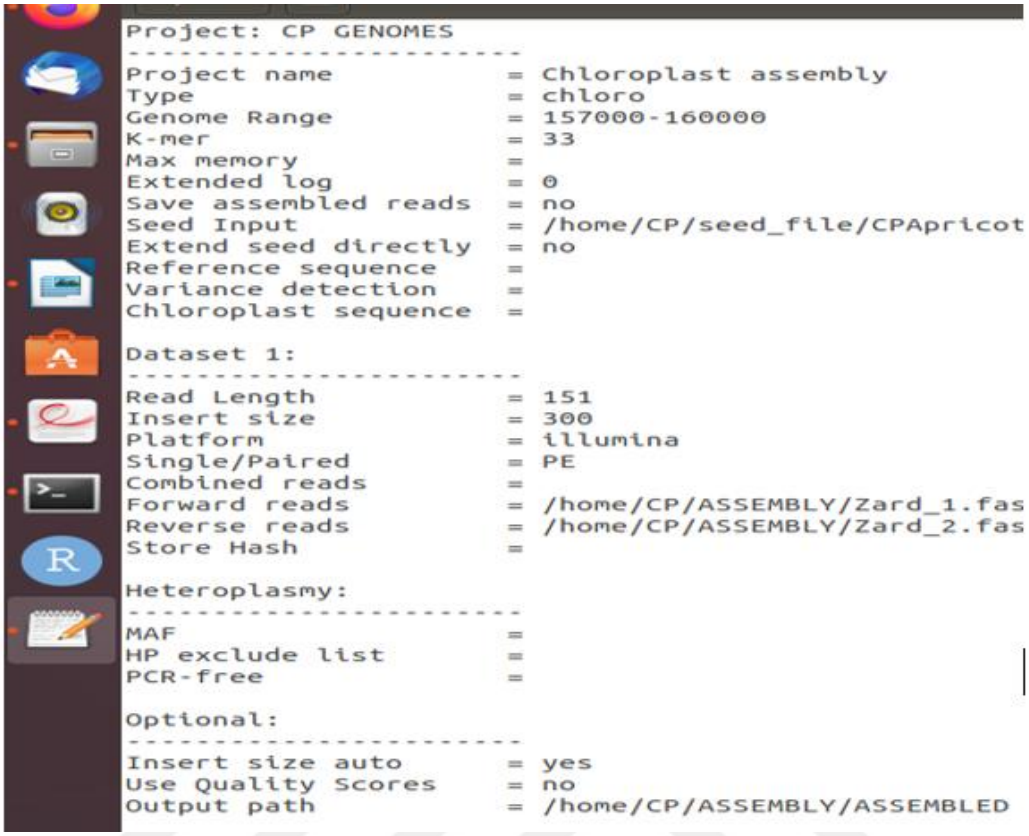

Figure 2.2. A configuration text used in the assembly of the “Zard” accession.

The 2 assemblies for each accession were compared and manually edited on Geneious Prime, and the best version of the CP genome was selected for further analyses.

2.3. CP genome annotation

The assembled genomes were then annotated using Geseq [100]. Geneious Prime was also used to annotate the CP genomes. In this study, tRNA genes were identified using tRNA-SE with default parameters [101], [102]. SNPs (Single Nucleotide Polymorphic) were identified using Geneious Prime and DnaSP4 v5 [103] (Figure 2.3). SNPs were manually checked to select reliable SNPs so that software errors can be avoided. A Python script was also used to check and validate variations among the genomes (Figure 2.4). Finally, maps of the circular chloroplast genome of “Zard”, and “Hacihaliloglu” were drawn using OGDraw 1.3.1 [104].

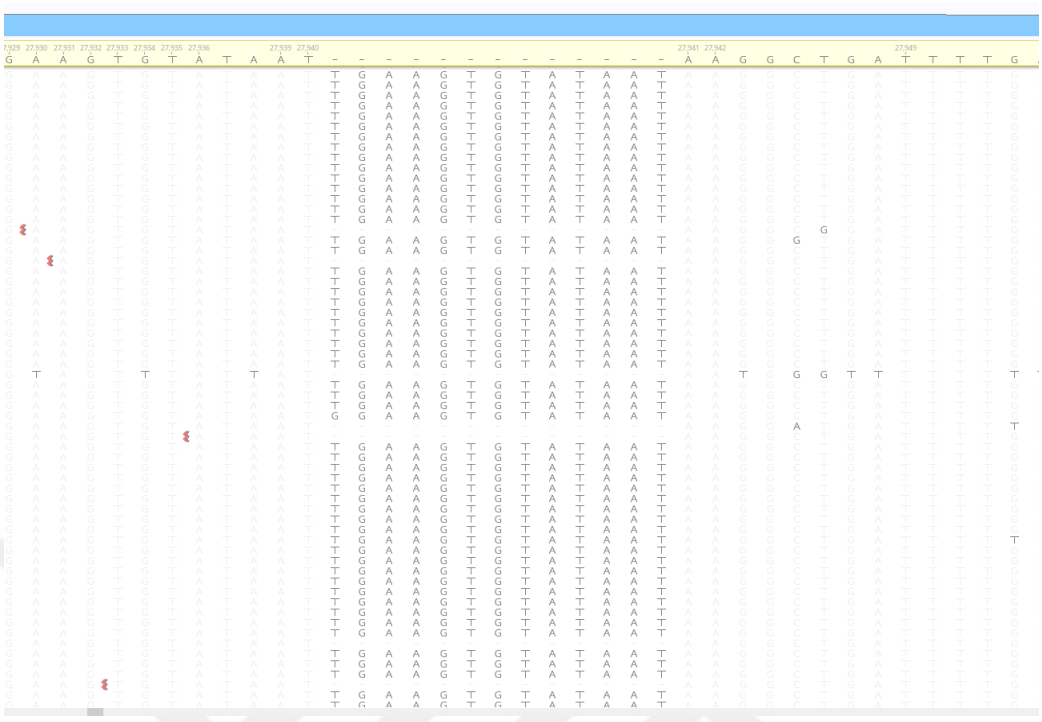

Figure 2.3. A 13 base insertion displayed on Geneious Prime, the contigs have this 13 bases while the reference genome does not.

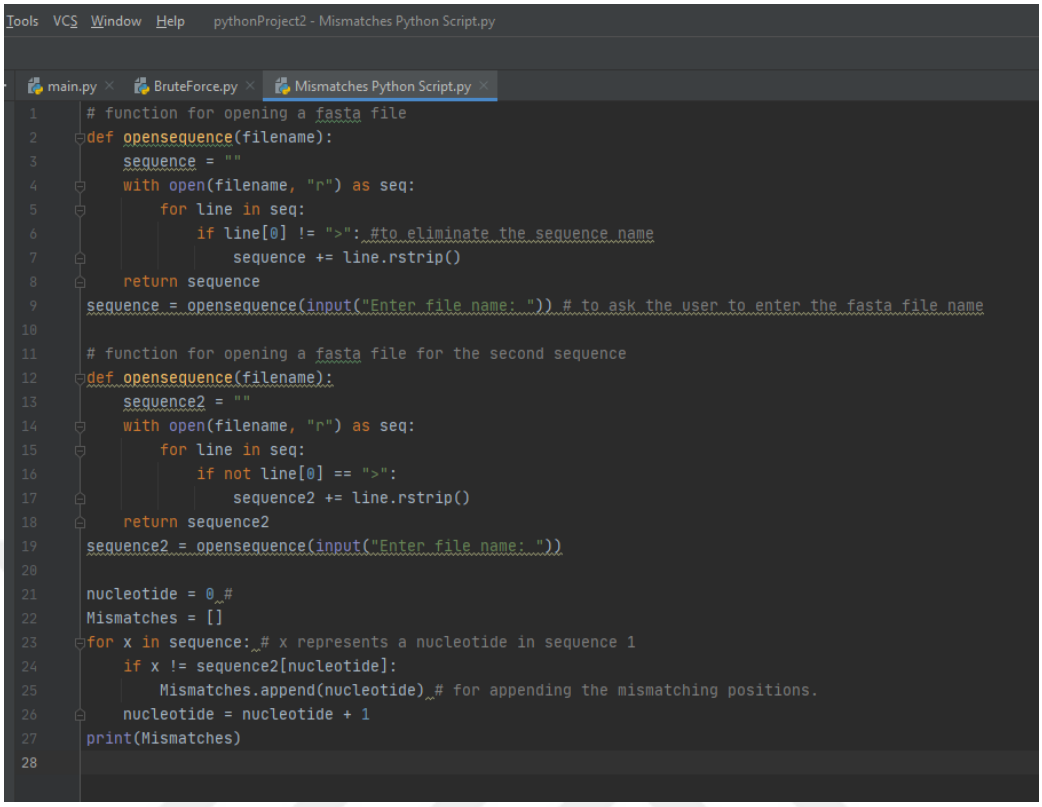

Figure 2.4. The python script that was used to validate variations among the genomes.

2.4. Relative Synonymous Codon Usage (RSCU)

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*) [105]. The RSCU analysis was conducted for all the CP genomes.

2.5. Comparative Analysis

The mVista program was used to compare apricot (*P. armeniaca, P. manshurica, P. sibirica, P. mume, and P. zhengheensis)* CP genomes using the shuffle-Lagan model [106]. The reference CP genome of apricots (NC_043901.1) was used for this comparison. Additionally, to compare the contraction and expansion of the IR boundaries among LSC, SSC, IRa, and IRb, the IRscope, software (https://irscope.shinyapps.io/irapp/) was used IR comparative analysis was first conducted on 9 apricot species and then on all the other CP genomes downloaded from NCBI [107].

2.6. Phylogenetic analyses

To show the phylogenetic relationship of the apricot accessions, phylogenetic trees were constructed. The 62 CP genomes were first aligned using the ClustalW algorithm on BioEdit v.7.2.5 (http://www.mbio.ncsu.edu/bioedit/bioedit.html) with default parameters [108]. To speed up the process, ClustalW was used on Linux (Ubuntu) user interface (Figure 2.5). The slow and accurate option was used when doing the alignment to ensure an accurate alignment. Maximum likelihood (ML) and Neighbor-joining analyses were conducted for the phylogenetic relationship analyses [109], [110]. Mega 11 was used to construct phylogenetic trees using maximum likelihood analysis and Neighbor-joining algorithms, a bootstrap repetition of 1000 was used for both analyses.

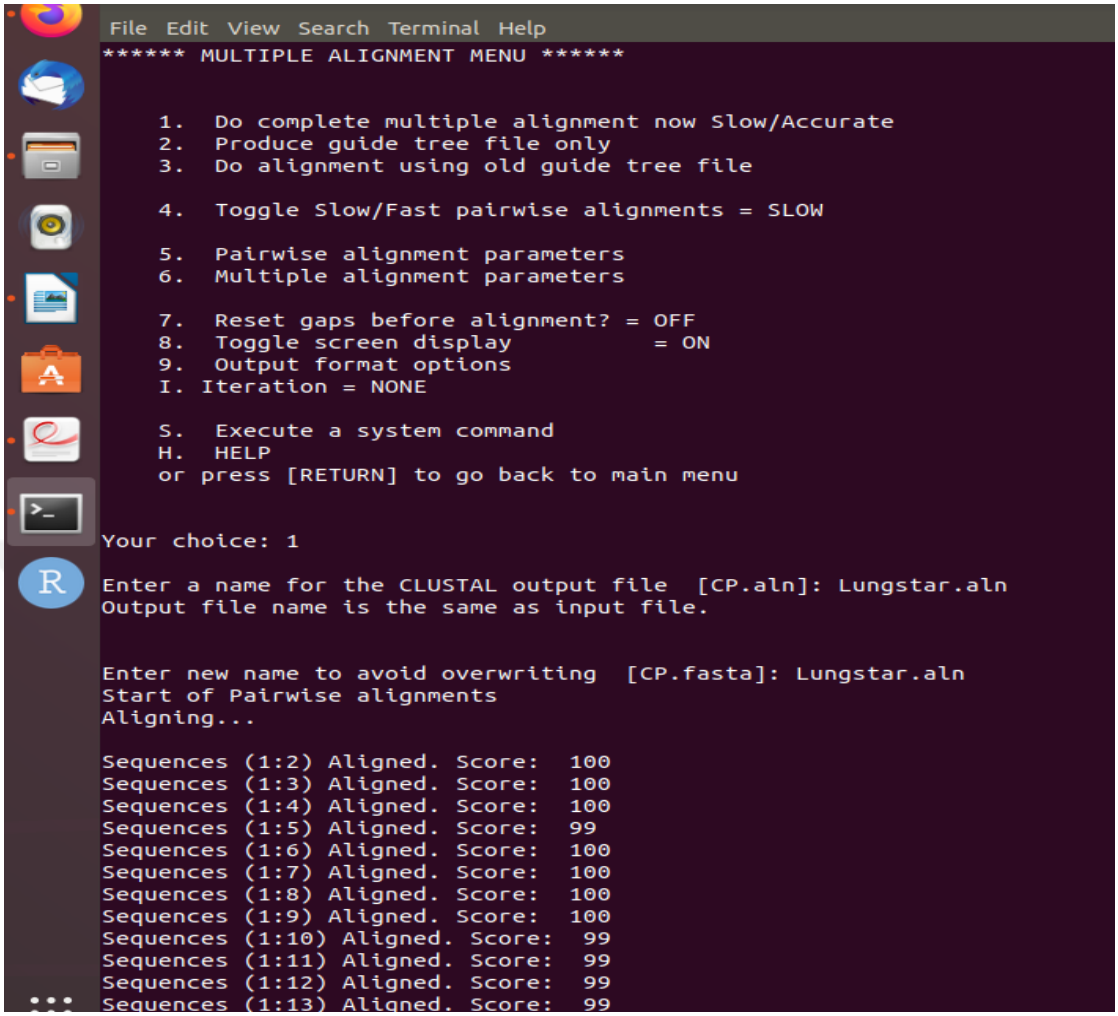

Figure 2.5. Alignment using ClustalW on Ubuntu 18.4.

3. CHAPTER

RESULTS

3.1. Chloroplast Genome Assembly and Characteristics

In this study, DNA of 30 apricot samples were sequenced using Illumina sequencing technology and CP genomes were assembled. The assembled CP genomes' lengths ranged from 158 057 bp to 158 089 bp. The nucleotide comparison of 30 common apricot CP genomes together with the GenBank accession NC_043901.1 revealed that CP genomes can be grouped into four haplotype groups. Among each haplotype, a common single nucleotide change was observed. "M2249", "Harcot", "KZ1", "KZ44", "OB", "Sakit", and "M2252" were classified as Haplotype 1. In this group, 8 deletions, 4 insertions, and 4 SNPs were observed when compared to the NC_043901.1 CP genome. Haplotype 2 (Fracasco, Hacikaya, Harlayne, Hasanbey, HH, HH F1 (-3, -4, -11, -16, -17, -18, and -19), Kabaasi, Kirmizi, and SEO) had only 1 deletion and 1 SNP. In Haplotype 3 (M2244, M2243, Roxana, and Zerdali), 2 deletions and 1 SNP were observed. Haplotype 4, which consists of the Zard, Zard F1-70, -98, and -146, had 10 insertions, 7 deletions, and 17 SNPs (Table 3.1).

Table 5. SNP-InDels tables showing the different SNPs in different positions.

Haplotype 1: M2249, Harcot, KZ44, Ordubat, Sakit			
Nucleotide	**Type of variation**	**Position**	**Frequency**
T	DELETION	1651	96,90
TGAAGTGTATAAT	DELETION	27940	66,30

T--	DELETION	37693	75,90
T--	DELETION	52377	91,90
T--	DELETION	66206	88,80
GG--	INSERTION	66763	79,90
T--	DELETION	69730	64,80
T	INSERTION	73666	93,20
G-T	SNP	76147	98,90
T--	DELETION	83306	90,00
T	INSERTION	115568	88,50
A--	DELETION	121972	93,70
T	INSERTION	122331	76,90
G-T	SNP	123712	99,00
A-G	SNP	124312	98,90
T-G	SNP	130167	96,30
Summary			
Insertions	**4**		
Deletions	**8**		
SNPs	**4**		
CP genome length	**158057**		

Haplotype 2: Hacihaliloğlu F1 (3, 4, 11, 16, 17, 18, and 19), Fracasco, Hacihaliloğlu, SEO, Kirmizi, Kabaasi, Hasanbey, Harlayne, Hacikaya

Nucleotide	Type of variation	Position	Frequency
T--	DELETION	9959	91,90
C-T	SNP	56788	99,30
Summary			
Deletions	**1**		
SNPs	**1**		
CP genome length	**158071**		

Haplotype 3: Zerdali, Roxana, M2244, M2243

Nucleotide	Type of variation	Position	Frequency
T--	DELETION	9959	86,30
TGAAGTGTATAAT	DELETION	27940	63,40
C-T	SNP	56788	97,40
Summary			
Deletions	**2**		
SNPs	**1**		
CP genome length	**158058**		

Haplotype 4: Zard and P. mandshurica group: Zard, Zard 98, Zard 70, Zard 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	10054	70,60
T--	DELETION	12522	93,20
T	INSERTION	13402	87,80
C-T	SNP (Transition)	17066	99,50
T--	DELETION	37693	75,90
C-G	SNP (Transversion)	47415	97,60
G-T	SNP (Transversion)	47672	97,30
TCCATA	INSERTION	52833	98,10
ATAATA-----	DELETION	53158	83,30
ATTTCA	INSERTION	56658	97,60
C-T	SNP (Transition)	56788	99,80
CC	INSERTION	61068	91,60
T--	DELETION	66206	98,70
A-C	SNP (Transversion)	66812	92,00
G-A	SNP (Transition)	72116	99,00
TTTTTTT	INSERTION	73666	93,00
A-C	SNP (Transversion)	76910	98,30
T	INSERTION	77238	91,00
T--	DELETION	83306	90,00
A-G	SNP (Transition)	85987	99,00
C-T	SNP (Transition)	86095	99,70
G-A	SNP (Transition)	113201	99,50
G	INSERTION	115565	92,60
GTTTCTAATTTT	DELETION	115998	98,80
G-T	SNP (Transversion)	116553	98,80
G-A	SNP (Transition)	122249	99,00
T	INSERTION	122331	76,10
G-T	SNP (Transversion)	123712	99,50
A-G	SNP (Transversion)	124312	99,40
G-T	SNP	126252	99,10
T-A	SNP	127877	98,70
Summary			
Insertions	**10**		
Deletions	**7**		
SNPS	**17**		
CP genome length	**158089**		

In general, the genome features analyzed in this study showed that the gene content, gene order, introns, and intergenic spacers of the *Prunus* genomes are highly conserved and found to be similar to previously published CP genomes of *Prunus* trees. In brief, the CP genomes were divided into 4 parts: Large Single Copy (LSC) regions, Small Single Copy (SSC) region, and 2 Inverted Repeats (IR) regions. The quadripartite structure of the CP genomes consisted of an LSC region of 85 746–85861 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 CP genomes was 36.7%. Coding regions accounted for about 52.7% of the CP genomes. A total of 122 genes were identified in H1, H2, and H3, while, in H4, 123 genes were identified. These genes included 81 protein-coding genes in H1, H2, and H3, and 82 protein-coding genes in H4. A total of 33 tRNA genes, and 8 rRNA genes were identified in all the assembled CP genomes (Table 3.2). The *trnY-ATA* gene was identified as a unique gene in *P. armeniaca*. CP genome map for HH as representative for common apricots, and "Zard" which was found to be divergent than the other common apricots were represented in Figure 3.1 and Figure 3.2, respectively.

Table 6. Chloroplast genomes` structural information

H	Length (bp)	LSC	SSC	IRa	IRb	GC (%)	Total Genes	tRNA genes	rRNA genes	PC genes	# Codons
1	158057	86269 - 158057	112659 - 131668	131669 - 158057	86270 - 112658	36,7	122	33	8	81	52685
2	158071	86284 - 158072	112674 - 131682	131683 - 158071	86285 - 112673	36,7	122	33	8	81	52690
3	158058	86271 - 158058	112661 - 131669	131670 - 158058	86272 - 112660	36,7	122	33	8	81	52686
4	158089	86312 - 158089	112702 - 131700	131701 - 158089	86313 - 112701	36,7	123	33	8	82	52696

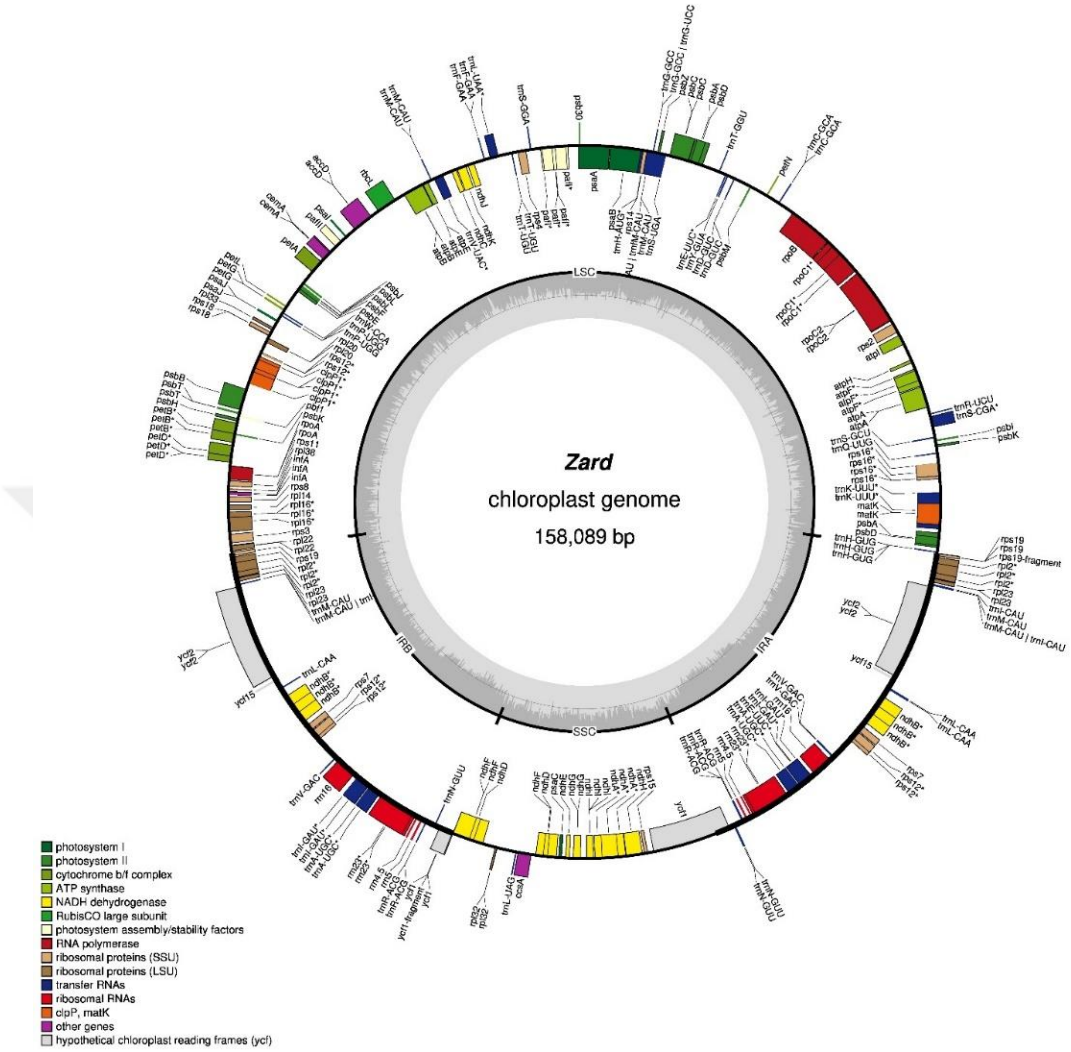

Figure 3.1. *P. armeniaca* (“Zard”) chloroplast genome map showing the four parts of the CP genome, IRA, IRB, SSC, and LSC. The annotated genes are also shown in this map.

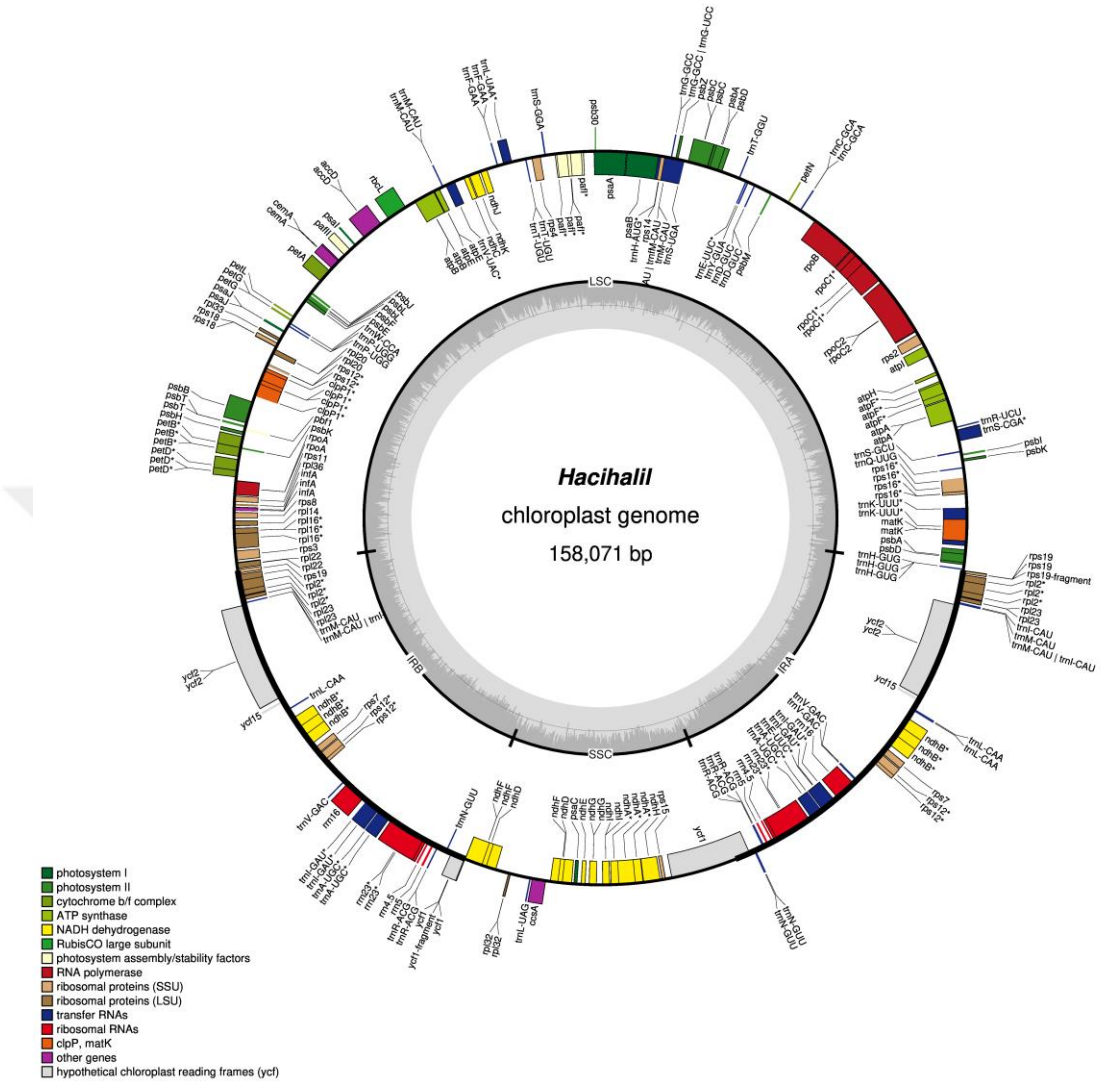

Figure 3.2**.** *P. armeniaca* (“Hacihaliloğlu”) chloroplast genome map showing the four parts of the CP genome, IRA, IRB, SSC, and LSC.

Table 7. Genes in the sequenced *P. armeniaca* CP genomes.

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	*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, trnH-GUG, trnI-CAU, trnI-GAU, trnK-UUU, trnL-CAA,* ***trnY-ATA*** *trnL-UAA, trnL-UAG, trnM-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*

3.3. Relative Synonymous Coding Usage (RSCU)

The RSCU of the apricot CP genomes was predicted using CodonW based on protein-coding genes. The total number of codons ranged from 52400 to 52700 for all the apricot CP genomes (Table S3). In haplotype 1, haplotype 2, and haplotype 3, 52685, 52690, and 52686 codons were encoded, respectively. The Zard (haplotype 4) accessions and *P. mandshurica* had the most codons (52696 codons) among the apricot CP genomes. The *P. tomentosa* species had the least number of codons (52416 codons). In this study, we found that there are 64 types of codons encoding 20 amino acids. Leucine (Leu), Serine (Ser), and Arginine (Arg) are encoded by 6 codons; Valine (Val), Proline (Pro), Threonine (Thr), Alanine (Ala), and Glycine (Gly) are encoded by 4 codons; 3 codons encode Isoleucine (Ile); 2 codons encode phenylalanine (Phe), Tyrosine (Tyr), Histidine (His), Glutamine (Gln); Asparagine (Asn); Lysine (Lys); aspartic acid (Asp); Glutamic acid (Glu); and Cysteine (Cys); and the remaining Methionine (Met) and Tryptophan (Trp) are encoded by only 1 codon. We also found out that Leucine (Leu) is the most frequent amino acid, with a frequency range of 9,80%–10,80%. Tryptophan (Trp) was the least frequent amino acid in CP genome protein-coding genes with a frequency range of 1,25%–1,35%. The highest RSCU among amino acids was observed in Arginine (Arg) with a value of 1,92, while other amino acids like Leucine (Leu) and Serine (Ser) had their highest RSCU values at 1,49 and 1,48, respectively. Methionine (Met) and Tryptophan (Trp) had a RSCU value of 1, which shows that there is no codon usage bias in these 2 amino acids. Around 46% of the codons showed a RSCU value of less than 1, meaning that there was a negative codon usage bias in these codons.

3.4. Comparative Alignment Analysis

The mVista program was used to align the 9 apricot CP genome sequences to compare genome divergence and sequence identity among the genomes. This alignment showed that there are no gene rearrangements within the apricot CP genomes. The apricot CP genomes were highly conserved, and no structural differences were observed.

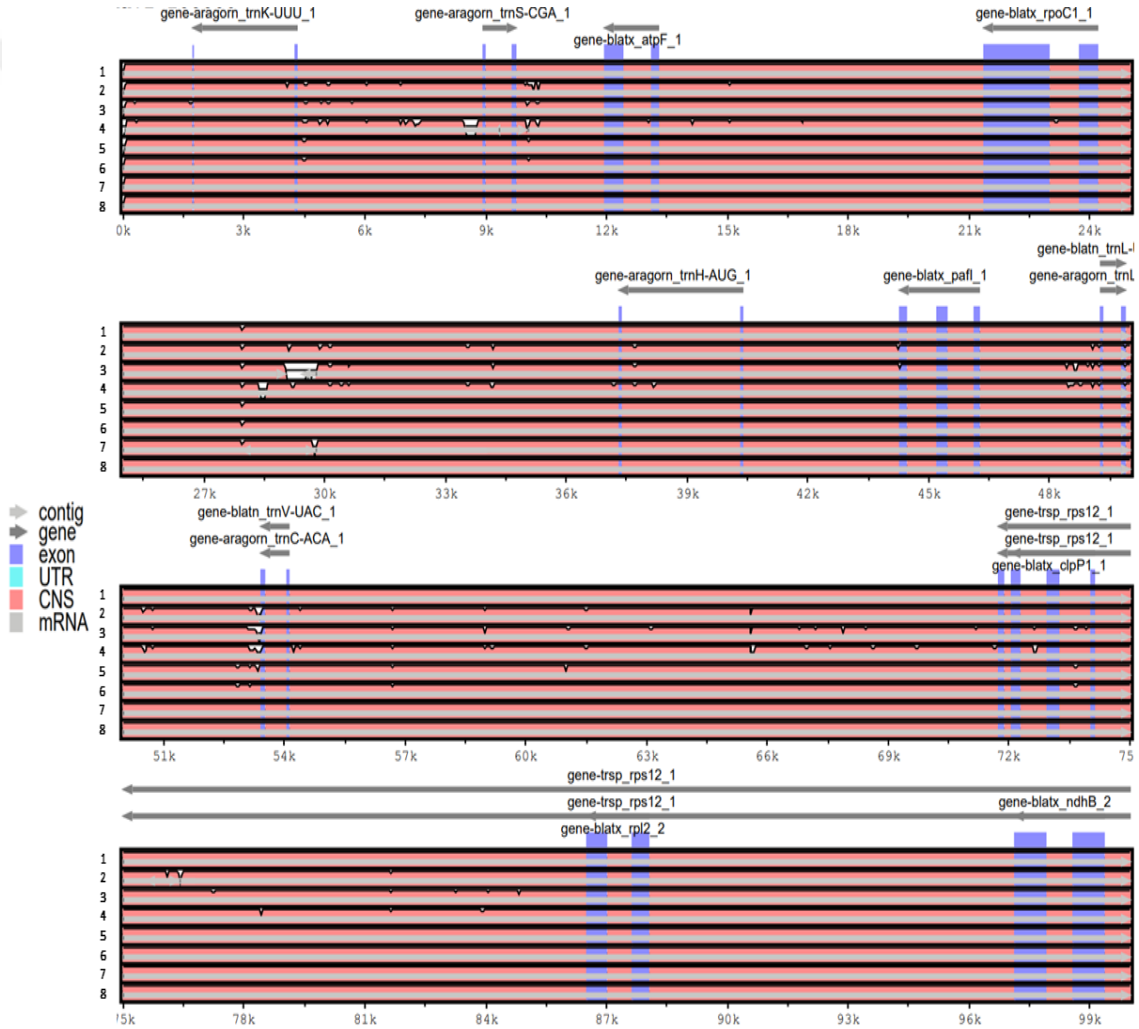

Figure 3.3. mVista alignment graph showing the aligned apricot CP genomes; 1: H2, 2: H3, 3: H4, 4: *P. mandshurica*, 5: *P. mume*, 6: *P. zhengheensis*, 7: *P. mume*, and 8: *P.*

sibirica. This alignment shows expressed genes in this region in the apricot CP genomes, and these genes are conserved.

3.5. IR comparative analysis

IR comparative analysis on the apricot accessions revealed that Haplotype 1, Haplotype 2, Haplotype 3, Haplotype 4, *P. mandshurica*, and the reference CP genome of *P. armeniaca* have the same genes in the IR region. The genes were highly conserved on these CP genomes. The genes were of the same size in these CP genomes except for the rpl2 gene on *P. mandshurica*, which was 13 bp while in the other CP genomes it is 2 bp. It is also noteworthy that in these CP genomes, the *ycf1* gene was found overlapping the JSA region, while on the other 3 apricot CP genomes; *P. mume*, *P. zhengheensis*, and *P. sibirica*, the *ycf1* gene was missing in the JSA region. (Figure 4).

An IR comparison among all the other *Prunus* species` genomes downloaded from NCBI revealed differences in IRa, IRb, LSC, and SSC lengths (*Figure S 5*). Most of the genes were conserved in these Prunus species except for the *rpl2* and the *ndhF* genes. The *rpl2* gene was missing in the IRb region of *P. tenella* while on the other CP genomes it was found. This gene was also longer in *P. padus* and *P. serotina* as compared when compared to the other *rbl2* genes in the other CP genomes. The *ndhF* gene was only missing in *P. serotina*, and it was not overlapping the JSB region in *P. subhirtella, P. yedoensis, P. tangutica* while in the other CP genomes it was overlapping this region.

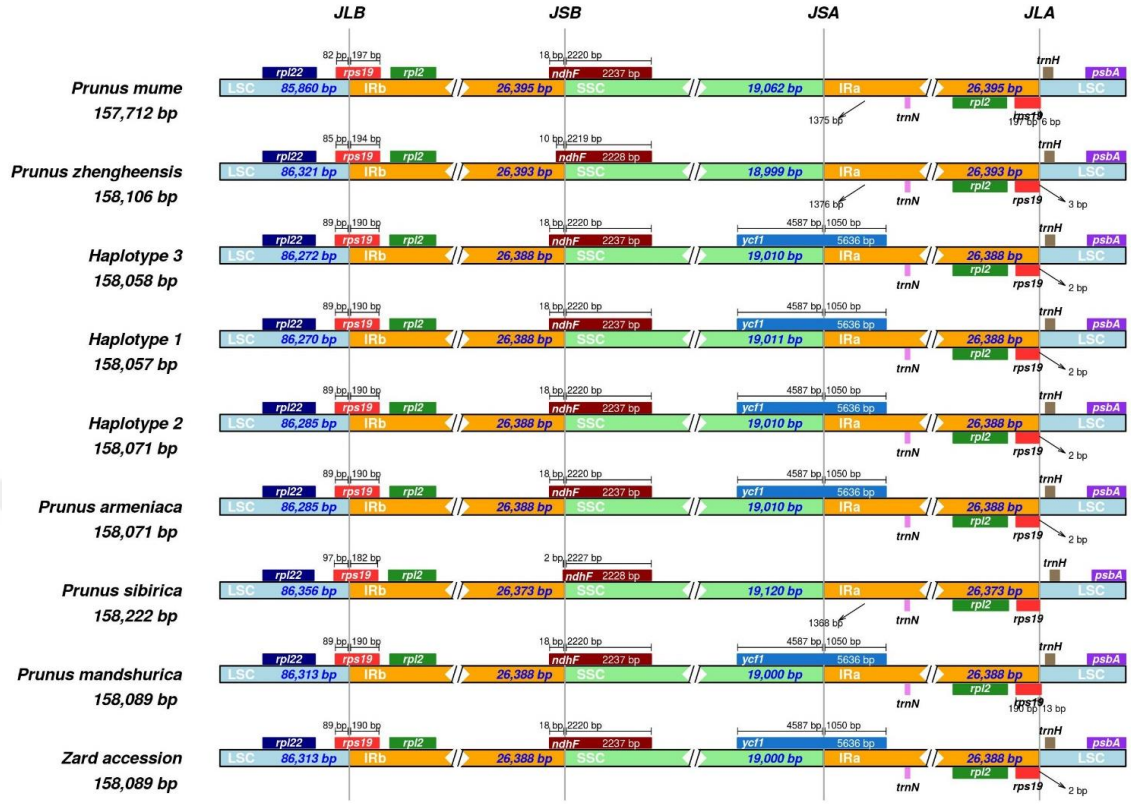

Figure 3.4. IR comparative analysis graph of the 9 apricot CP genomes, showing the different genes found in the IRa, IRb, SSC, and LSC regions and the length of these regions. Whereby JLB represents the junction between LSC and IRb, JSB represents the junction between IRb and SSC, JSA represents the junction between SSC and IRa, and JLA represents the junction between IRa and LSC.

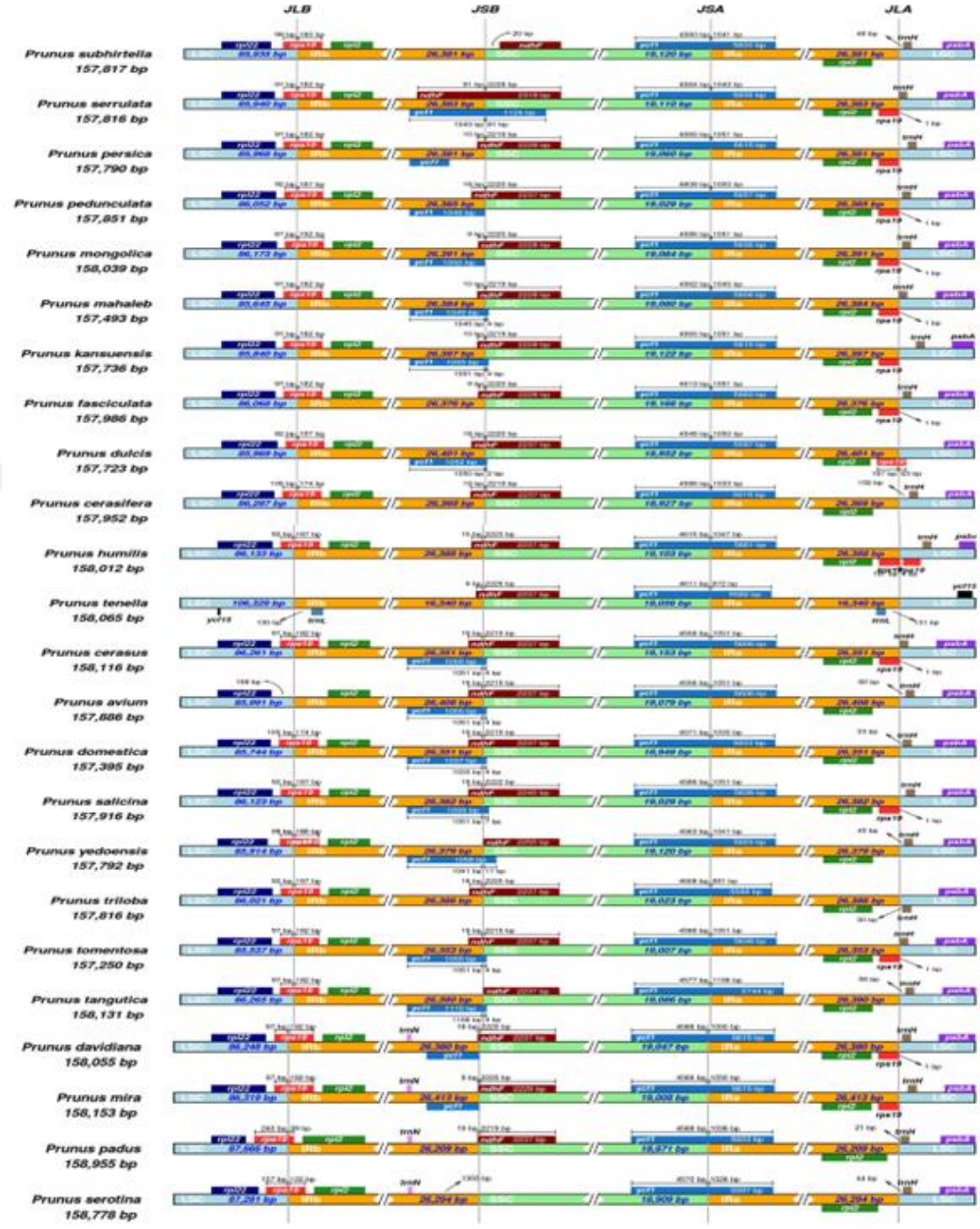

Figure 3.5. IR comparative analysis of all the *Prunus* species downloaded from NCBI except for apricots genomes.

3.7. Phylogenetic relationships

All CP genomes were used to construct maximum likelihood (ML) and neighbor-joining phylogenetic trees (Figure 3.6). Each haplotype was represented by 1 CP sequence and its offsprings. The phylogenetic trees suggest that the 3 *P. armeniaca* species: Haplotype 1, haplotype 2, and haplotype 3 are closely related. The "Zard" accession and its offsprings were identified as sister clades with *P. mandshurica*. These results showed that *P. armeniaca* is phylogenetically closer to *P. mume* and *P. zhengheensis* than it is to the other *Prunus* species. *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. triloba* and *P. pedunculata* are Chinese almonds. Surprisingly, *P. sibirica* is more closely related to *P. salicina*, *P. cerasifera*, and *P. domestica* than it is to *P. armeniaca*. *P. fasciculata* was positioned between *P. dulcis* and the cultivated cherry clade (*P. mahaleb, P. serrualta, P. subhirtella, and P. yedoensis*).

3.8. Inheritance of CP genomes

The maternal inheritance of CP genomes in the *Prunus* species was clearly validated as shown 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 their maternal CP genome of the "Zard" accession and they were phylogenetically classified under the same group. The "Hacihaliloğlu" F1 hybrids: HH F1-3, -4, -11, -16, -17, -18, -19 were also identical to their maternal parent. These results confirm the maternal inheritance of CP genomes in *Prunus armeniaca.*

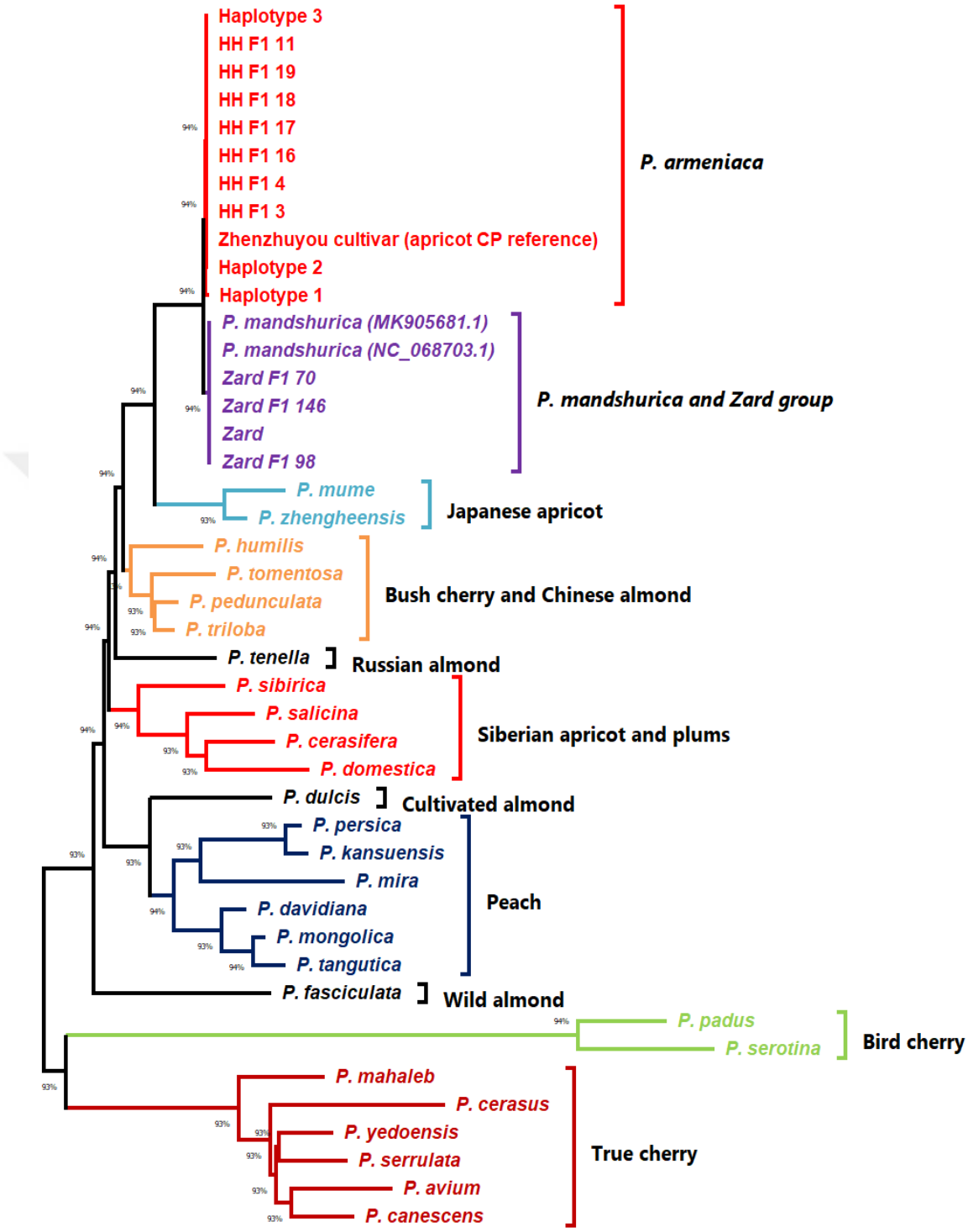

Figure 3.6. A Maximum-Likelihood Phylogenetic tree of the *Prunus* species showing the "Zard" accession in the same group as *P. mandshurica*.

4. CHAPTER

DISCUSSIONS AND CONCLUSIONS

4.1. Discussions

In this study, 30 CP genomes were assembled for phylogenetic comparative studies. The assembled genomes had lengths ranging from 158 057 to 158 089 bp. The lengths of the CP genomes are closer to the lengths of *Prunus* CP genomes obtained in previous studies [15], [23], [70]. Our CP genomes had a GC content of 36,7% and this is because angiosperms generally have a low GC content [72], [121]. The similarity of the CP genomes to previous genomes increases confidence in the quality of the CP genomes.

Upon assembling the CP genomes, they were annotated, and 111-112 unique genes, 8 rRNA genes, and 31 to 33 tRNA genes were found. Previous studies on Prunus species CP genomes reported similar results, as they reported around 133 genes (110 unique genes), including about 94 protein-coding, about31 tRNA, and 8 rRNA genes [15]. To be more specific, *P. persica, P. mume, and P. tomentosa* have 131, 132, and 135 genes, respectively [15], [122], [123]. A study on six almond species also yielded similar results [70]. Among the annotated genes, *rps19* was common, and the *trnY-ATA* gene was identified as a unique gene. The *rps19* gene is common in *Prunus* species, and the *trnY-ATA* gene was also identified as a unique gene in other previous studies [15], [124]. These results from the previous studies further validate these CP genomes.

RSCU analysis revealed that different species encode a different number of codons, and this is because codon preferences for amino acids in different species are different [125]. RSCU analysis showed that among all the amino acids, leucine (Leu) is the most frequent

amino acid with a frequency range of 9,80% - 10,80% while tryptophan (Trp) was the least frequent amino acid with a frequency range of 1,25% - 1,35%. Previous studies on *P. mume, P. armeniaca*, and *P. salicina* also show that leucine is the most encoded amino acid while tryptophan is the least encoded amino acid [15].

When comparing the CP genome alignments using mVista, no significant difference was observed; the genes were more conserved, while the IR comparison on IRscope showed significant differences. The repeat lengths for *P. mandshurica* and the "Zard" accession were similar, and that shows that they are closely related to each other. Most of the genes were conserved in the IR regions of the CP genomes, while the *ndhF* gene was found at one end of the SSC region in *P. yedoensis,* while in the other CP genomes it was found overlapping in the JSB region. This shows that *P. yedoensis* is divergent from *P. armeniaca*. Overall, the *ycf1* and *ndhF* genes were the only genes that were not conserved in the IR regions; the other genes were highly conserved.

The identified SNPs and InDels are intraspecific variations and have not been observed in other *Prunus* species. Even though intraspecific variations have not been reported in *Prunus* species before our study, intraspecific variations have been reported in other plant species such as *Ricinus communis* L. And *Sorghum bicolor* [126], [127]. Further molecular biological studies need to be done to validate these intraspecific variations.

The phylogenetic analysis showed that *P. mume* is more closely related to *P. armeniaca L*. than all the other *Prunus* species. These results are in congruence with other previous results on *Prunus* species [15], [128]. However, *P. sibirica*, a Siberian apricot was not closely related to *P. armeniaca* as compared to wild cherries and plums. Dong et al. [129] also reported that *P. armeniaca* is closely related to plums than it is to *P. sibirica*, and that is an interesting evolution problem. The phylogenetic tree further revealed that the "Zard" accession is in the same group as *P. mandshurica,* and that might mean that "Zard" is better classified as *P. mandshurica* than *P. armeniaca L*. Both *P. mandshurica* and "Zard" originated in Central Asia, so further molecular biological studies must be conducted to verify the phylogenetic classification of these two species [46], [130].

The maternal inheritance of the CP genome was clearly validated based on these results because of the similarities of the hybrids' CP genomes to their maternal parents` CP genomes. Evidently, the three "Zard" F1 hybrids ("Zard F1 98", "Zard F1 146", and "Zard F1 70") were identical to the "Zard" accession in the phylogenetic tree, and all the "Hacihaliloğlu" F1 hybrids ("Hacihaliloğlu F1 3", "Hacihaliloğlu F1 4", "Hacihaliloğlu F1 11", "Hacihaliloğlu F1 16", "Hacihaliloğlu F1 17", "Hacihaliloğlu F1 18", and "Hacihaliloğlu F1 19") were also identical to the CP genome of "Hacihaliloğlu". These results are congruent with the general inheritance properties shown by angiosperms [31], [33], [34].

4.2. Conclusions

In this study, the CP genomes of 30 apricots were *de novo* assembled and compared with 22 previously published *Prunus* species CP genomes. The complete CP genomes of the *Prunus* species were annotated, and coding genes were identified. The gene order and genome structure of *P. armeniaca* are highly conserved and only show a few variations when compared to other *Prunus* species. No major structural differences in the CP genomes of *Prunus armeniaca* accessions were observed. The CP genome sequences of the 30 apricots obtained in this work are highly conserved, and only the "Zard" accession shows major differences. "Zard" has a different number of polymorphic sites as compared to the other assembled CP genomes, and in the phylogenetic tree, it is in the same clade as *P. mandshurica*. These results further validated that *P. mume* and *P. mandshurica* is closely related to *P. armeniaca,* while Siberian apricot, *P. sibirica,* is not closely related to *P. armeniaca.* The maternal inheritance of the CP genome in the *Prunus* species was validated in this study as all the F1 hybrids were identical to their maternal parents` CP genomes. These results expand the researchers' perspective on apricot plant diversity and promote an understanding of the evolutionary relationship among *Prunus* species.

4.3. Recommendations

In this study, we were able to phylogenetically classify *P. armeniaca* accessions and *Prunus* species, but we recommend further molecular biological analysis to be conducted to validate the classification of the “Zard” accession and *P. mandshurica* species. These results show that there are less variations within the *P. armeniaca* species as compared to other *Prunus* species. This shows that variation in the *P. armeniaca* species must be improved through breeding practices. The “Zard” accession which is homozygous resistant to PPV was grouped with *P. mandshurica* and thus, *P. mandshurica* can be PPV resistance source. Additionally, considering that common apricots pose narrow genetic diversity, it is recommended to cross other apricot species with common apricot to improve agricultural traits of common apricots.

REFERENCES

[1] M. L. Badenes and D. E. Parfitt, "Phylogenetic relationships of cultivated Prunus species from an analysis of chloroplast DNA variation," *Theor. Appl. Genet.*, vol. 90, no. 7–8, pp. 1035–1041, Jun. 1995, doi: 10.1007/BF00222918/METRICS.

[2] N. Chevigny, D. Schatz-Daas, F. Lotfi, and J. M. Gualberto, "DNA Repair and the Stability of the Plant Mitochondrial Genome," *Int. J. Mol. Sci. 2020, Vol. 21, Page 328*, vol. 21, no. 1, p. 328, Jan. 2020, doi: 10.3390/IJMS21010328.

[3] K. Ohyama *et al.*, "Chloroplast gene organization deduced from complete sequence of liverwort Marchantia polymorpha chloroplast DNA," *Nat. 1986 3226079*, vol. 322, no. 6079, pp. 572–574, 1986, doi: 10.1038/322572a0.

[4] R. K. Jansen *et al.*, "Methods for Obtaining and Analyzing Whole Chloroplast Genome Sequences," *Methods Enzymol.*, vol. 395, pp. 348–384, Jan. 2005, doi: 10.1016/S0076-6879(05)95020-9.

[5] M. L. Metzker, "Emerging technologies in DNA sequencing," *Genome Res.*, vol. 15, no. 12, pp. 1767–1776, Dec. 2005, doi: 10.1101/GR.3770505.

[6] B. Zhong *et al.*, "Systematic Error in Seed Plant Phylogenomics," *Genome Biol. Evol.*, vol. 3, no. 1, pp. 1340–1348, Jan. 2011, doi: 10.1093/GBE/EVR105.

[7] Z. Xi *et al.*, "Phylogenomics and a posteriori data partitioning resolve the Cretaceous angiosperm radiation Malpighiales," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 109, no. 43, pp. 17519–17524, Oct. 2012, doi: 10.1073/PNAS.1205818109/SUPPL_FILE/SAPP.PDF.

[8] C. F. Barrett, C. D. Specht, J. Leebens-Mack, D. W. Stevenson, W. B. Zomlefer, and J. I. Davis, "Resolving ancient radiations: can complete plastid gene sets elucidate deep relationships among the tropical gingers (Zingiberales)?," *Ann. Bot.*, vol. 113, no. 1, pp. 119–133, Jan. 2014, doi: 10.1093/AOB/MCT264.

[9] C. S. Campbell, R. C. Evans, D. R. Morgan, T. A. Dickinson, and M. P. Arsenault, "Phylogeny of subtribe Pyrinae (formerly the Maloideae, Rosaceae): Limited resolution of a complex evolutionary history," *Plant Syst. Evol.*, vol. 266, no. 1–2, pp. 119–145, Jul. 2007, doi: 10.1007/S00606-007-0545-Y/METRICS.

[10] R. Velasco *et al.*, "The genome of the domesticated apple (Malus × domestica Borkh.)," *Nat. Genet. 2010 4210*, vol. 42, no. 10, pp. 833–839, Aug. 2010, doi: 10.1038/ng.654.

[11] S. V. Nikiforova, D. Cavalieri, R. Velasco, and V. Goremykin, "Phylogenetic Analysis of 47 Chloroplast Genomes Clarifies the Contribution of Wild Species to the Domesticated Apple Maternal Line," *Mol. Biol. Evol.*, vol. 30, no. 8, pp. 1751–1760, Aug. 2013, doi: 10.1093/MOLBEV/MST092.

[12] H. Hirakawa *et al.*, "Dissection of the Octoploid Strawberry Genome by Deep Sequencing of the Genomes of Fragaria Species," *DNA Res.*, vol. 21, no. 2, pp. 169–181, Apr. 2014, doi: 10.1093/DNARES/DST049.

[13] R. VanBuren *et al.*, "The genome of black raspberry (Rubus occidentalis)," *Plant J.*, vol. 87, no. 6, pp. 535–547, Sep. 2016, doi: 10.1111/TPJ.13215.

[14] Y. Ding *et al.*, "Phylogenic study of Lemnoideae (duckweeds) through complete chloroplast genomes for eight accessions," *PeerJ*, vol. 2017, no. 12, p. e4186, Dec. 2017, doi: 10.7717/PEERJ.4186/SUPP-8.

[15] S. Xue *et al.*, "Comparative analysis of the complete chloroplast genome among Prunus mume, P. armeniaca, and P. salicina," *Hortic. Res.*, vol. 6, no. 1, Dec. 2019, doi: 10.1038/S41438-019-0171-1/41964996/41438_2019_ARTICLE_171.PDF.

[16] S. D. Zhang *et al.*, "Diversification of Rosaceae since the Late Cretaceous based on plastid phylogenomics," *New Phytol.*, vol. 214, no. 3, pp. 1355–1367, May 2017, doi: 10.1111/NPH.14461.

[17] S. Lee and J. Wen, "A phylogenetic analysis of Prunus and the Amygdaloideae (Rosaceae) using ITS sequences of nuclear ribosomal DNA," *Am. J. Bot.*, vol. 88,

no. 1, pp. 150–160, 2001, doi: 10.2307/2657135.

[18] X. Zhang, T. Zhou, N. Kanwal, Y. Zhao, G. Bai, and G. Zhao, "Completion of eight gynostemma BL. (cucurbitaceae) chloroplast genomes: Characterization, comparative analysis, and phylogenetic relationships," *Front. Plant Sci.*, vol. 8, no. September, pp. 1–13, 2017, doi: 10.3389/fpls.2017.01583.

[19] R. K. Jansen, C. Saski, S. B. Lee, A. K. Hansen, and H. Daniell, "Complete Plastid Genome Sequences of Three Rosids (Castanea, Prunus, Theobroma): Evidence for At Least Two Independent Transfers of rpl22 to the Nucleus," *Mol. Biol. Evol.*, vol. 28, no. 1, pp. 835–847, Jan. 2011, doi: 10.1093/MOLBEV/MSQ261.

[20] M. S. Cho, C. H. Cho, S. Y. Kim, H. S. Yoon, and S. C. Kim, "Complete chloroplast genome of Prunus yedoensis Matsum.(Rosaceae), wild and endemic flowering cherry on Jeju Island, Korea," *https://doi.org/10.3109/19401736.2015.1079840*, vol. 27, no. 5, pp. 3652–3654, Sep. 2015, doi: 10.3109/19401736.2015.1079840.

[21] S. Wang, C. W. Gao, and L. Z. Gao, "Plastid genome sequence of an ornamental and editable fruit tree of Rosaceae, Prunus mume," *https://doi.org/10.3109/19401736.2015.1089546*, vol. 27, no. 6, pp. 4407–4408, Nov. 2015, doi: 10.3109/19401736.2015.1089546.

[22] S. Ercisli, "Apricot culture in Türkiye," *Sci. Res. Essays*, vol. 4, no. 8, pp. 715–719, 2009.

[23] X. Huang *et al.*, "The chloroplast genome of Prunus zhengheensis: Genome comparative and phylogenetic relationships analysis," *Gene*, vol. 793, p. 145751, Aug. 2021, doi: 10.1016/J.GENE.2021.145751.

[24] L. Shi *et al.*, "CPGAVAS2, an integrated plastome sequence annotator and analyzer," *Nucleic Acids Res.*, vol. 47, no. W1, pp. W65–W73, Jul. 2019, doi: 10.1093/NAR/GKZ345.

[25] J. Q. Wu, Y. Wang, P. Sun, Z. S. Sun, and J. S. Shen, "The complete chloroplast

genome of Prunus phaeosticta (Hance) Maxim. (Rosaceae) and its phylogenetic implications," *http://www.tandfonline.com/action/authorSubmission?journalCode=tmdn20&page=instructions*, vol. 8, no. 1, pp. 136–140, 2023, doi: 10.1080/23802359.2022.2163841.

[26] W. Wang *et al.*, "comparative and phylogenetic Analyses of the complete chloroplast Genomes of Six Almond Species (Prunus spp. L.)," doi: 10.1038/s41598-020-67264-3.

[27] D. Coulibaly *et al.*, "Comparative Analysis of Complete Chloroplast Genome and Phenotypic Characteristics of Japanese Apricot Accessions," *Horticulturae*, vol. 8, no. 9, p. 794, Sep. 2022, doi: 10.3390/HORTICULTURAE8090794/S1.

[28] S. Turan, S. Teber, O. Canbulat, and K. Gürcana, "The complete chloroplast genome of Prunus armeniaca," *Acta Hortic.*, vol. 1290, pp. 41–46, 2020, doi: 10.17660/ACTAHORTIC.2020.1290.8.

[29] W.-W. Li, L.-Q. Liu, Q.-P. Zhang, W.-Q. Zhou, G.-Q. Fan, and & Kang Liao, "Phylogeography of Prunus armeniaca L. revealed by chloroplast DNA and nuclear ribosomal sequences," *Sci. Reports* |, vol. 11, p. 13623, 123AD, doi: 10.1038/s41598-021-93050-w.

[30] E. Sevindik, Z. T. Murathan, and M. Sevindik, "Molecular Genetic Diversity of Prunus armeniaca L. (Rosaceae) Genotypes by RAPD, ISSR-PCR, and Chloroplast DNA (cpDNA) trnL-F Sequences," *https://doi.org/10.1080/15538362.2020.1828223*, vol. 20, no. S3, pp. S1652–S1661, 2020, doi: 10.1080/15538362.2020.1828223.

[31] H.-S. Park *et al.*, "Inheritance of chloroplast and mitochondrial genomes in cucumber revealed by four reciprocal F 1 hybrid combinations," *Sci. Reports* |, vol. 11, p. 2506, 123AD, doi: 10.1038/s41598-021-81988-w.

[32] E. Jablonka, "The evolution of the peculiarities of mammalian sex chromosomes:

an epigenetic view," *BioEssays*, vol. 26, no. 12, pp. 1327–1332, Dec. 2004, doi: 10.1002/BIES.20140.

[33] S. Greiner, J. Sobanski, and R. Bock, "Why are most organelle genomes transmitted maternally?," *BioEssays*, vol. 37, no. 1, pp. 80–94, Jan. 2015, doi: 10.1002/BIES.201400110.

[34] A. Kormutak, M. Galgoci, D. Sukenikova, P. Bolecek, J. Libantova, and D. Gőmőry, "Maternal inheritance of chloroplast DNA in Pinus mugo Turra: a case study of Pinus mugo × Pinus sylvestris crossing," *Plant Syst. Evol.*, vol. 304, no. 1, pp. 71–76, Jan. 2018, doi: 10.1007/S00606-017-1449-0/FIGURES/3.

[35] "1. Layne, R. E. C., Bailey, C. H., & Hough, L. F.... - Google Akademik." https://scholar.google.com/scholar?hl=tr&as_sdt=0%2C5&q=1.%09Layne%2C+R.+E.+C.%2C+Bailey%2C+C.+H.%2C+%26+Hough%2C+L.+F.+%281996%29.+Apricots%3A+79-111.+Fruit+Breeding%2C+1.&btnG= (accessed May 09, 2023).

[36] R. Ahmad, D. Potter, and S. M. Southwick, "Identification and characterization of plum and pluot cultivars by microsatellite markers," *J. Hortic. Sci. Biotechnol.*, vol. 79, no. 1, pp. 164–169, 2004, doi: 10.1080/14620316.2004.11511743.

[37] A. Pedryc, S. Ruthner, R. Hermán, B. Krska, A. Hegedus, and J. Halász, "Genetic diversity of apricot revealed by a set of SSR markers from linkage group G1," *Sci. Hortic. (Amsterdam).*, vol. 121, no. 1, pp. 19–26, Jun. 2009, doi: 10.1016/J.SCIENTA.2009.01.014.

[38] L. B. Tatyana Zhebentyayeva , Craig Ledbetter and and G. Llácer, *Fruit breeding, apricot*, no. October 2015. 2012.

[39] A. Groppi *et al.*, "Population genomics of apricots unravels domestication history and adaptive events," *Nat. Commun. 2021 121*, vol. 12, no. 1, pp. 1–16, Jun. 2021, doi: 10.1038/s41467-021-24283-6.

[40] E. Bortiri *et al.*, "Phylogeny and Systematics of Prunus (Rosaceae) as Determined

by Sequence Analysis of ITS and the Chloroplast trnL-trnF Spacer DNA Granger , Clay Weeks , Megan Buckingham , Daniel Potter and Dan E . Parfitt Published by : American Society of Plant Taxono," *Syst. Bot.*, vol. 26, no. 4, 2001.

[41] T. Zhebentyayeva, C. Ledbetter, L. Burgos, and G. Llácer, "Apricot," *Fruit Breed.*, pp. 415–458, Jan. 2012, doi: 10.1007/978-1-4419-0763-9_12/TABLES/3.

[42] N. I. Vavilov, "Phytogeographic basis of plant breeding.," *Orig. Var. Immun. Breed. Cultiv. plants.*, 1951.

[43] P. Zhukovsky, "Cultivated plant species and their relatives." Kolos, 1971.

[44] B. Khadari, L. Krichen, P. Lambert, M. Marrakchi, and J. M. Audergon, "Genetic structure in Tunisian apricot, Prunus armeniaca L., populations propagated by grafting: A signature of bottleneck effects and ancient propagation by seedlings," *Genet. Resour. Crop Evol.*, vol. 53, no. 4, pp. 811–819, Nov. 2006, doi: 10.1007/S10722-004-5721-0/METRICS.

[45] B. L. Gross and K. M. Olsen, "Genetic perspectives on crop domestication," *Trends Plant Sci.*, vol. 15, no. 9, pp. 529–537, Sep. 2010, doi: 10.1016/J.TPLANTS.2010.05.008.

[46] S. Liu *et al.*, "The complex evolutionary history of apricots: Species divergence, gene flow and multiple domestication events," *Mol. Ecol.*, vol. 28, no. 24, pp. 5299–5314, Dec. 2019, doi: 10.1111/MEC.15296.

[47] K. F. Kostina, "The use of varietal resources of apricots for breeding.," *Tr. nikit. bot. Sad. (Trans. Nikita bot. Gdn.).*, vol. 40, pp. 45–63, 1960.

[48] H. Bourguiba *et al.*, "Loss of genetic diversity as a signature of apricot domestication and diffusion into the Mediterranean Basin," *BMC Plant Biol.*, vol. 12, no. 1, pp. 1–17, Apr. 2012, doi: 10.1186/1471-2229-12-49/TABLES/6.

[49] J. Chat, S. Decroocq, V. Decroocq, and R. J. Petit, "A Case of Chloroplast Heteroplasmy in Kiwifruit (Actinidia deliciosa) That Is Not Transmitted During

Sexual Reproduction," *J. Hered.*, vol. 93, no. 4, pp. 293–300, Jul. 2002, doi: 10.1093/JHERED/93.4.293.

[50] FAO, "Apricots datastat," *FAO Departments and Offices*, 2020. .

[51] H. Canton, "Food and Agriculture Organization of the United Nations—FAO," *Eur. Dir. Int. Organ. 2021*, pp. 297–305, Jul. 2021, doi: 10.4324/9781003179900-41.

[52] T. N. Zhebentyayeva, G. L. Reighard, D. Lalli, V. M. Gorina, B. Krška, and A. G. Abbott, "Origin of resistance to plum pox virus in apricot: what new AFLP and targeted SSR data analyses tell," *Tree Genet. Genomes*, vol. 4, no. 3, pp. 403–417, 2008.

[53] K. Yilmaz, K. G.-G. D. in Plants, and undefined 2012, "Genetic diversity in apricot," *books.google.com*, Accessed: May 17, 2023. [Online]. Available: https://books.google.com/books?hl=tr&lr=&id=sAOaDwAAQBAJ&oi=fnd&pg=PA249&dq=Yılmaz+and+Kahraman+apricot&ots=OcUwIENJzp&sig=gr_BPvOeey2ogVs0dOqQrfJjPmA.

[54] J. Couranjou, "Genetic studies of 11 quantitative characters in apricot," *Sci. Hortic. (Amsterdam).*, vol. 61, no. 1–2, pp. 61–75, Feb. 1995, doi: 10.1016/0304-4238(94)00727-W.

[55] M. L. Badenes and G. Llácer, "Breeding for resistance: breeding for Plum pox virus resistant apricots (Prunus armeniaca L.) in Spain," *EPPO Bull.*, vol. 36, no. 2, pp. 323–326, Aug. 2006, doi: 10.1111/J.1365-2338.2006.01007.X.

[56] F. Geuna, M. Toschi, and D. Bassi, "The use of AFLP markers for cultivar identification in apricot," *Plant Breed.*, vol. 122, no. 6, pp. 526–531, Dec. 2003, doi: 10.1111/J.1439-0523.2003.00897.X.

[57] M. Baránek, J. Raddová, and M. Pidra, "Comparative analysis of genetic diversity in Prunus L. as revealed by RAPD and SSR markers," *Sci. Hortic. (Amsterdam).*, vol. 108, no. 3, pp. 253–259, May 2006, doi: 10.1016/J.SCIENTA.2006.01.023.

[58] J. I. Hormaza, H. Yamane, and J. Rodrigo, “Apricot,” vol. 4, 2007.

[59] B. M. Asma, “Breeding Program For Plum Pox Virus Resitance in Türkiye : Preliminary Results,” 2008.

[60] E. B. Akin, I. Karabulut, and A. Topcu, “Some compositional properties of main Malatya apricot (Prunus armeniaca L.) varieties,” *Food Chem.*, vol. 107, no. 2, pp. 939–948, Mar. 2008, doi: 10.1016/J.FOODCHEM.2007.08.052.

[61] H. Haciseferoğullari, I. Gezer, M. M. Özcan, and B. MuratAsma, “Post-harvest chemical and physical–mechanical properties of some apricot varieties cultivated in Türkiye,” *J. Food Eng.*, vol. 79, no. 1, pp. 364–373, Mar. 2007, doi: 10.1016/J.JFOODENG.2006.02.003.

[62] M. Altindag *et al.*, “Biological control of brown rot (Moniliana laxa Ehr.) on apricot (Prunus armeniaca L. cv. Hacıhaliloğlu) by Bacillus, Burkholdria, and Pseudomonas application under in vitro and in vivo conditions,” *Biol. Control*, vol. 38, no. 3, pp. 369–372, Sep. 2006, doi: 10.1016/J.BIOCONTROL.2006.04.015.

[63] K. Gürcan, S. Teber, and K. Ça, “Further investigation of a genetically divergent group of plum pox virus-M strain in Türkiye,” pp. 385–391, 2019.

[64] P. Uthaipaisanwong *et al.*, “Characterization of the chloroplast genome sequence of oil palm (Elaeis guineensis Jacq.),” *Gene*, vol. 500, no. 2, pp. 172–180, Jun. 2012, doi: 10.1016/J.GENE.2012.03.061.

[65] S. Hossein Jafari, A. Sepehry, H. Soltanloo, and A. A. Karimian, “Genetic differentiation between bitter and sweet asafetida plants using ISSR markers,” *Mol. Biol. Rep.*, vol. 46, no. 1, pp. 1069–1078, Feb. 2019, doi: 10.1007/S11033-018-4565-1/TABLES/8.

[66] A. E. Strand, J. Leebens-Mack, and B. G. Milligan, “Nuclear DNA-based markers for plant evolutionary biology,” *Mol. Ecol.*, vol. 6, no. 2, pp. 113–118, Feb. 1997, doi: 10.1046/J.1365-294X.1997.00153.X.

[67] J. I. Hormaza, H. Yamane, and J. Rodrigo, “Apricot,” *Fruits Nuts*, pp. 171–187, Mar. 2007, doi: 10.1007/978-3-540-34533-6_7.

[68] J. A. Ruiz-Chután, J. Salava, D. Janovská, J. Žiarovská, M. Kalousová, and E. Fernández, “Assessment of genetic diversity in sorghum bicolor using RAPD markers,” *Genetika*, vol. 51, no. 3, pp. 789–803, 2019, doi: 10.2298/GENSR1903789R.

[69] C. E. Olsen *et al.*, “Glucosinolate diversity within a phylogenetic framework of the tribe Cardamineae (Brassicaceae) unraveled with HPLC-MS/MS and NMR-based analytical distinction of 70 desulfoglucosinolates,” *Phytochemistry*, vol. 132, pp. 33–56, Dec. 2016, doi: 10.1016/J.PHYTOCHEM.2016.09.013.

[70] W. Wang *et al.*, “comparative and phylogenetic Analyses of the complete chloroplast Genomes of Six Almond Species (Prunus spp. L.),” doi: 10.1038/s41598-020-67264-3.

[71] S. V. Nikiforova, D. Cavalieri, R. Velasco, and V. Goremykin, “Phylogenetic Analysis of 47 Chloroplast Genomes Clarifies the Contribution of Wild Species to the Domesticated Apple Maternal Line,” *Mol. Biol. Evol.*, vol. 30, no. 8, pp. 1751–1760, Aug. 2013, doi: 10.1093/MOLBEV/MST092.

[72] H. Daniell, C. S. Lin, M. Yu, and W. J. Chang, “Chloroplast genomes: diversity, evolution, and applications in genetic engineering,” *Genome Biol. 2016 171*, vol. 17, no. 1, pp. 1–29, Jun. 2016, doi: 10.1186/S13059-016-1004-2.

[73] P. G. Nevill *et al.*, “Plastome-Wide Rearrangements and Gene Losses in Carnivorous Droseraceae,” *Genome Biol. Evol.*, vol. 11, no. 2, pp. 472–485, Feb. 2019, doi: 10.1093/GBE/EVZ005.

[74] J. Dobrogojski, M. Adamiec, and R. Luciński, “The chloroplast genome: a review,” *Acta Physiol. Plant.*, vol. 42, no. 6, pp. 1–13, Jun. 2020, doi: 10.1007/S11738-020-03089-X/TABLES/3.

[75] M. Sugiura, “The chloroplast genome.,” *Essays Biochem.*, vol. 30, pp. 49–57, Jan.

1995, doi: 10.1007/978-94-011-2656-4_10.

[76] H. E. Neuhaus and M. J. Emes, “NONPHOTOSYNTHETIC METABOLISM IN PLASTIDS,” *https://doi.org/10.1146/annurev.arplant.51.1.111*, vol. 51, pp. 111–140, Nov. 2003, doi: 10.1146/ANNUREV.ARPLANT.51.1.111.

[77] D. C. Price *et al.*, “Cyanophora paradoxa genome elucidates origin of photosynthesis in algae and plants,” *Science (80-.).*, vol. 335, no. 6070, pp. 843–847, Feb. 2012, doi: 10.1126/SCIENCE.1213561/SUPPL_FILE/PRICESOM.PDF.

[78] L. Gao, Y. J. Su, and T. Wang, “Plastid genome sequencing, comparative genomics, and phylogenomics: Current status and prospects,” *J. Syst. Evol.*, vol. 48, no. 2, pp. 77–93, Mar. 2010, doi: 10.1111/J.1759-6831.2010.00071.X.

[79] S. H. Strauss, J. D. Palmer, G. T. Howe, and A. H. Doerksen, “Chloroplast genomes of two conifers lack a large inverted repeat and are extensively rearranged.,” *Proc. Natl. Acad. Sci.*, vol. 85, no. 11, pp. 3898–3902, Jun. 1988, doi: 10.1073/PNAS.85.11.3898.

[80] X. Yao, P. Tang, Z. Li, D. Li, Y. Liu, and H. Huang, “The First Complete Chloroplast Genome Sequences in Actinidiaceae: Genome Structure and Comparative Analysis,” *PLoS One*, vol. 10, no. 6, p. e0129347, Jun. 2015, doi: 10.1371/JOURNAL.PONE.0129347.

[81] Y. Song *et al.*, “Evolutionary Comparisons of the Chloroplast Genome in Lauraceae and Insights into Loss Events in the Magnoliids,” *Genome Biol. Evol.*, vol. 9, no. 9, pp. 2354–2364, Sep. 2017, doi: 10.1093/GBE/EVX180.

[82] K. Shinozaki *et al.*, “The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression,” *EMBO J.*, vol. 5, no. 9, pp. 2043–2049, Sep. 1986, doi: 10.1002/J.1460-2075.1986.TB04464.X.

[83] K. Ohyama, H. Fukuzawa, T. Kohchi, H. Shirai, T. S.- Nature, and undefined 1986, “Chloroplast gene organization deduced from complete sequence of

liverwort Marchantia polymorpha chloroplast DNA," *nature.com*, Accessed: May 11, 2023. [Online]. Available: https://www.nature.com/articles/322572a0.

[84] R. A. Atherton, B. J. McComish, L. D. Shepherd, L. A. Berry, N. W. Albert, and P. J. Lockhart, "Whole genome sequencing of enriched chloroplast DNA using the Illumina GAII platform," *Plant Methods 2010 61*, vol. 6, no. 1, pp. 1–6, Sep. 2010, doi: 10.1186/1746-4811-6-22.

[85] W. Wang, M. Schalamun, A. Morales-Suarez, D. Kainer, B. Schwessinger, and R. Lanfear, "Assembly of chloroplast genomes with long- and short-read data: A comparison of approaches using Eucalyptus pauciflora as a test case," *BMC Genomics*, vol. 19, no. 1, pp. 1–15, Dec. 2018, doi: 10.1186/S12864-018-5348-8/FIGURES/3.

[86] B. Wu, L. Li, T. Qiu, X. Zhang, and S. Cui, "Cytosolic APX2 is a pleiotropic protein involved in H2O2 homeostasis, chloroplast protection, plant architecture and fertility maintenance," *Plant Cell Rep.*, vol. 37, no. 6, pp. 833–848, Jun. 2018, doi: 10.1007/S00299-018-2272-Y/FIGURES/8.

[87] Y. Yuan, "Applications of Optical Mapping for Plant Genome Assembly and Structural Variation Detection," *Methods Mol. Biol.*, vol. 2443, pp. 245–257, 2022, doi: 10.1007/978-1-0716-2067-0_13.

[88] M. J. Ankenbrand, "ETH Library chloroExtractor: extraction and assembly of the chloroplast genome from whole genome shotgun data," doi: 10.3929/ethz-b-000376785.

[89] F. T. Bakker *et al.*, "Herbarium genomics: plastome sequence assembly from a range of herbarium specimens using an Iterative Organelle Genome Assembly pipeline," *Biol. J. Linn. Soc.*, vol. 117, no. 1, pp. 33–43, Jan. 2016, doi: 10.1111/BIJ.12642.

[90] E. Coissac, P. M. Hollingsworth, S. Lavergne, and P. Taberlet, "From barcodes to genomes: extending the concept of DNA barcoding," *Mol. Ecol.*, vol. 25, no. 7, pp.

1423–1428, Apr. 2016, doi: 10.1111/MEC.13549.

[91] N. Dierckxsens, P. Mardulyn, and G. Smits, “NOVOPlasty: de novo assembly of organelle genomes from whole genome data,” *Nucleic Acids Res.*, vol. 45, no. 4, pp. e18–e18, Feb. 2017, doi: 10.1093/NAR/GKW955.

[92] J.-J. Jin *et al.*, “GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes,” *bioRxiv*, p. 256479, Oct. 2019, doi: 10.1101/256479.

[93] J. A. Freudenthal, S. Pfaff, N. Terhoeven, A. Korte, M. J. Ankenbrand, and F. Förster, “A systematic comparison of chloroplast genome assembly tools,” *Genome Biol.*, vol. 21, no. 1, pp. 1–21, Sep. 2020, doi: 10.1186/S13059-020-02153-6/FIGURES/10.

[94] J. L. Corriveau and A. W. Coleman, “RAPID SCREENING METHOD TO DETECT POTENTIAL BIPARENTAL INHERITANCE OF PLASTID DNA AND RESULTS FOR OVER 200 ANGIOSPERM SPECIES,” *Am. J. Bot.*, vol. 75, no. 10, pp. 1443–1458, Oct. 1988, doi: 10.1002/J.1537-2197.1988.TB11219.X.

[95] N. Scarcelli *et al.*, “Intra-individual polymorphism in chloroplasts from NGS data: where does it come from and how to handle it?,” *Mol. Ecol. Resour.*, vol. 16, no. 2, pp. 434–445, Mar. 2016, doi: 10.1111/1755-0998.12462.

[96] K. Paszkiewicz and D. J. Studholme, “De novo assembly of short sequence reads,” *Brief. Bioinform.*, vol. 11, no. 5, pp. 457–472, Sep. 2010, doi: 10.1093/BIB/BBQ020.

[97] N. De Vere, T. C. G. Rich, S. A. Trinder, and C. Long, “Dna barcoding for plants,” *Methods Mol. Biol.*, vol. 1245, pp. 101–118, 2015, doi: 10.1007/978-1-4939-1966-6_8/COVER.

[98] K. L. Bell *et al.*, “Pollen DNA barcoding: current applications and future prospects1,” *https://doi.org/10.1139/gen-2015-0200*, vol. 59, no. 9, pp. 629–640, Mar. 2016, doi: 10.1139/GEN-2015-0200.

[99] H. Li and R. Durbin, “Fast and accurate long-read alignment with Burrows–Wheeler transform,” *Bioinformatics*, vol. 26, no. 5, pp. 589–595, Mar. 2010, doi: 10.1093/BIOINFORMATICS/BTP698.

[100] M. Tillich *et al.*, “GeSeq – versatile and accurate annotation of organelle genomes,” *Nucleic Acids Res.*, vol. 45, no. W1, pp. W6–W11, Jul. 2017, doi: 10.1093/NAR/GKX391.

[101] T. M. Lowe and P. P. Chan, “tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes,” *Nucleic Acids Res.*, vol. 44, no. W1, pp. W54–W57, Jul. 2016, doi: 10.1093/NAR/GKW413.

[102] N. Nayyar *et al.*, “Population structure and genetic diversity of invasive Fall Armyworm after 2 years of introduction in India,” *Sci. Rep.*, vol. 11, no. 1, pp. 1–12, 2021, doi: 10.1038/s41598-021-87414-5.

[103] P. Librado and J. Rozas, “DnaSP v5: a software for comprehensive analysis of DNA polymorphism data,” *Bioinformatics*, vol. 25, no. 11, pp. 1451–1452, Jun. 2009, doi: 10.1093/BIOINFORMATICS/BTP187.

[104] S. Greiner, P. Lehwark, and R. Bock, “OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes,” *Nucleic Acids Res.*, vol. 47, no. W1, pp. W59–W64, Jul. 2019, doi: 10.1093/NAR/GKZ238.

[105] J. F. Peden, “Analysis of Codon Usage.”

[106] K. A. Frazer, L. Pachter, A. Poliakov, E. M. Rubin, and I. Dubchak, “VISTA: computational tools for comparative genomics,” *Nucleic Acids Res.*, vol. 32, no. suppl_2, pp. W273–W279, Jul. 2004, doi: 10.1093/NAR/GKH458.

[107] A. Amiryousefi, J. Hyvönen, and P. Poczai, “IRscope: an online program to visualize the junction sites of chloroplast genomes,” *Bioinformatics*, vol. 34, no. 17, pp. 3030–3031, Sep. 2018, doi: 10.1093/BIOINFORMATICS/BTY220.

[108] M. A. Larkin *et al.*, “Clustal W and Clustal X version 2.0,” *Bioinformatics*, vol. 23, no. 21, pp. 2947–2948, Nov. 2007, doi: 10.1093/BIOINFORMATICS/BTM404.

[109] F. Ronquist and J. P. Huelsenbeck, “MrBayes 3: Bayesian phylogenetic inference under mixed models,” *Bioinformatics*, vol. 19, no. 12, pp. 1572–1574, Aug. 2003, doi: 10.1093/BIOINFORMATICS/BTG180.

[110] N. Saitou and M. Nei, “The neighbor-joining method: a new method for reconstructing phylogenetic trees.,” *Mol. Biol. Evol.*, vol. 4, no. 4, pp. 406–425, Jul. 1987, doi: 10.1093/OXFORDJOURNALS.MOLBEV.A040454.

[111] Y. Bao, F. Xu, and S. M. Shimeld, “Phylogenetics of Lophotrochozoan bHLH Genes and the Evolution of Lineage-Specific Gene Duplicates,” *Genome Biol. Evol.*, vol. 9, no. 4, pp. 869–886, Apr. 2017, doi: 10.1093/GBE/EVX047.

[112] P. Mazumdar, R. Binti Othman, K. Mebus, N. Ramakrishnan, and J. Ann Harikrishna, “Codon usage and codon pair patterns in non-grass monocot genomes,” *Ann. Bot.*, vol. 120, no. 6, pp. 893–909, Nov. 2017, doi: 10.1093/AOB/MCX112.

[113] S. T. Parvathy, V. Udayasuriyan, and V. Bhadana, “Codon usage bias,” *Mol. Biol. Reports 2021 491*, vol. 49, no. 1, pp. 539–565, Nov. 2021, doi: 10.1007/S11033-021-06749-4.

[114] D. Zhang *et al.*, “Analysis of the chloroplast genome and phylogenetic evolution of Bidens pilosa,” *BMC Genomics 2023 241*, vol. 24, no. 1, pp. 1–13, Mar. 2023, doi: 10.1186/S12864-023-09195-7.

[115] L. Gong *et al.*, “Comparative chloroplast genome analyses of Amomum: insights into evolutionary history and species identification,” *BMC Plant Biol.*, vol. 22, no. 1, pp. 1–17, Dec. 2022, doi: 10.1186/S12870-022-03898-X/FIGURES/8.

[116] H. Jiang *et al.*, “Comparative and phylogenetic analyses of six Kenya Polystachya (Orchidaceae) species based on the complete chloroplast genome sequences,”

BMC Plant Biol. 2022 221, vol. 22, no. 1, pp. 1–21, Apr. 2022, doi: 10.1186/S12870-022-03529-5.

[117] C. H. Galeano, A. C. Fernández, M. Gómez, and M. W. Blair, "Single strand conformation polymorphism based SNP and Indel markers for genetic mapping and synteny analysis of common bean (Phaseolus vulgaris L.)," *BMC Genomics*, vol. 10, no. 1, pp. 1–14, Dec. 2009, doi: 10.1186/1471-2164-10-629/FIGURES/5.

[118] P. H. Van Tienderen, A. A. De Haan, C. G. Dan Der Linden, and B. Vosman, "Biodiversity assessment using markers for ecologically important traits," *Trends Ecol. Evol.*, vol. 17, no. 12, pp. 577–582, Dec. 2002, doi: 10.1016/S0169-5347(02)02624-1.

[119] D. Edwards, J. W. Forster, D. Chagné, and J. Batley, "What Are SNPs?," *Assoc. Mapp. Plants*, pp. 41–52, 2007, doi: 10.1007/978-0-387-36011-9_3.

[120] J. B. Yang, S. X. Yang, H. T. Li, J. Yang, and D. Z. Li, "Comparative Chloroplast Genomes of Camellia Species," *PLoS One*, vol. 8, no. 8, p. e73053, Aug. 2013, doi: 10.1371/JOURNAL.PONE.0073053.

[121] X. Zhang, T. Zhou, N. Kanwal, Y. Zhao, G. Bai, and G. Zhao, "Completion of eight gynostemma BL. (cucurbitaceae) chloroplast genomes: Characterization, comparative analysis, and phylogenetic relationships," *Front. Plant Sci.*, vol. 8, p. 1583, Sep. 2017, doi: 10.3389/FPLS.2017.01583/BIBTEX.

[122] B. Wu *et al.*, "Genome-wide identification, expression patterns, and functional analysis of UDP glycosyltransferase family in peach (Prunus persica L. Batsch)," *Front. Plant Sci.*, vol. 8, p. 389, Mar. 2017, doi: 10.3389/FPLS.2017.00389/BIBTEX.

[123] T. Chen *et al.*, "Phylogeography of Chinese cherry (Prunus pseudocerasus Lindl.) inferred from chloroplast and nuclear DNA: insights into evolutionary patterns and demographic history," *Plant Biol.*, vol. 17, no. 4, pp. 787–797, Jul. 2015, doi: 10.1111/PLB.12294.

[124] L. Shen *et al.*, “Complete plastid genome of Eriobotrya japonica (Thunb.) Lindl and comparative analysis in Rosaceae,” *Springerplus*, vol. 5, no. 1, pp. 1–14, Dec. 2016, doi: 10.1186/S40064-016-3702-3/FIGURES/6.

[125] H. M. W. Salim and A. R. O. Cavalcanti, “Factors influencing codon usage bias in genomes,” *J. Braz. Chem. Soc.*, vol. 19, no. 2, pp. 257–262, 2008, doi: 10.1590/S0103-50532008000200008.

[126] S. Muraguri *et al.*, “Intraspecific variation within Castor bean (Ricinus communis L.) based on chloroplast genomes,” *Ind. Crops Prod.*, vol. 155, p. 112779, Nov. 2020, doi: 10.1016/J.INDCROP.2020.112779.

[127] Y. Song, Y. Chen, J. Lv, J. Xu, S. Zhu, and M. Li, “Comparative Chloroplast Genomes of Sorghum Species: Sequence Divergence and Phylogenetic Relationships,” *Biomed Res. Int.*, vol. 2019, 2019, doi: 10.1155/2019/5046958.

[128] H. Katayama and C. Uematsu, “Structural analysis of chloroplast DNA in Prunus (Rosaceae): Evolution, genetic diversity and unequal mutations,” *Theor. Appl. Genet.*, vol. 111, no. 7, pp. 1430–1439, Nov. 2005, doi: 10.1007/S00122-005-0075-3/FIGURES/5.

[129] S. Dong *et al.*, “The complete chloroplast genome sequence of Prunus sibirica,” *http://www.tandfonline.com/action/authorSubmission?journalCode=tmdn20&page=instructions*, vol. 5, no. 1, pp. 581–582, 2020, doi: 10.1080/23802359.2019.1710589.

[130] C.A. Ledbetter, “APRICOT BREEDING IN NORTH AMERICA: CURRENT STATUS AND FUTURE PROSPECTS,” *Acta Hortic.*, no. 862, pp. 85–92, May 2010, doi: 10.17660/ACTAHORTIC.2010.862.12.

SUPPLEMENTARY

Supplementary 1. RSCU table showing some accessions used in our study with their respective RSCU value, number of codons, and the amino acids encoded.

Species	Codons	Amino Acid	Codon	No.	RSCU	Amino Acid	Codon	No.	RSCU	Amino Acid	Codon	No.	RSCU	Amino Acid	Codon	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	442,00	0,76
Kabaasi	**52690**	Leu	UUA	1256	1,37		UCA	893	1,18	TER	UAA	1382	1,32	TER	UGA	968	0,92
Melez 19	**52690**		UUG	1092	1,2		UCG	575	0,76		UAG	796	0,76	Trp	UGG	714	1,00
Melez 18	**52690**													-			
Melez 17	**52690**		CUU	1153	1,26	Pro	CCU	633	1,04	His	CAU	967	1,40	Arg	CGU	395	0,74
Melez 16	**52690**		CUC	695	0,76		CCC	610	1		CAC	418	0,6		CGC	266,00	0,50
Melez 4	**52690**		CUA	764	0,84		CCA	739	1,21	Gln	CAA	1053	1,32		CGA	549	1,03
Melez 3	**52690**		CUG	521	0,57		CCG	451	0,74		CAG	545	0,68		CGG	421,00	0,79
Haci	**52690**													-			
HaciKa	**52690**	Ile	AUU	1871	1,22	Thr	ACU	621	1,09	Asn	AAU	1847	1,41	Ser	AGU	635	0,84
Harlayn	**52690**		AUC	1099	0,72		ACC	582	1,02		AAC	782	0,59		AGC	512,00	0,68

Hasanb	**52690**		AUA	1618	1,06		ACA	693	1,21	Lys	AAA	2143	1,34	Arg	AGA	950	1,78
SEO	**52690**	Met	AUG	886	1		ACG	390	0,68		AAG	1057	0,66		AGG	629,00	1,18
Kirmizi	**52690**																
P. armeniacaZ		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	339,00	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	555,00	1,05
Species	**Codons**	**Amino Acid**	**Codon**	**No.**	**RSCU**	**Amino Acid**	**Codon**	**No.**	**RSCU**	**Amino Acid**	**Codon**	**No.**	**RSCU**	**Amino Acid**	**Codon**	**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	450,00	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	1,00
			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	275,00	0,49
			CUA	781	0,9		CCA	733	1,22	Gln	CAA	1078	1,41		CGA	566	1,02
Roxana	**52686**		CUG	459	0,53		CCG	391	0,65		CAG	454	0,59		CGG	408,00	0,73

M2243	**52686**														
M2244	**52686**	Ile	AUU	1848	1,21	Thr	ACU	672	1,14	Asn	AAU	1858	1,40	Ser	AGU	756	0,95
Zerdali	**52686**		AUC	1095	0,72		ACC	578	0,98		AAC	802	0,6		AGC	478,00	0,60
			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	626,00	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	365,00	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	558,00	0,98

CURRICULUM VITAE

PERSONAL INFORMATION

Name and Surname: Lungelo Khanyile

Nationalitv: Swazi

EDUCATION

Degree	Institution	Graduation Date
Masters`degree	Erciyes University, Agricultural Biotechnology	2023
Bachelors`degree	University of Eswatini, Agricultural Biosystems Engineering	2018

İŞ DENEYİMLERİ

Year	Institution	Task
2018-2019	National Maize Corporation	Supervisor
2017-2018	Royal Eswatini Sugar Corporation	Trainee

LANGUAGES

English (Native)

Zulu (Native)

Turkish (C1)

PUBLICATIONS

1. Kılınçer, İ., Khanyile, L., & Gürcan, K. 2023. Population structure of sumac (Rhus coriaria L.) from Türkiye based on transcriptome-developed SSR marker. *Genetic Resources and Crop Evolution*, *70*(4), 1197-1213.
2. Khanyile, L., Kılınçer, İ., Gürcan, K., 2022. Transcriptpme analysis and Identifcation of SSRs (Simple Sequence Repeats) on Sumac using simple bioinformatics techniques. Ömer Halisci University, 4. Bitki Islahı ve Genetiği Öğrenci Kongresi Niğde, Özet kitapçığı ss 2.