Journal of Molecular Structure 1284 (2023) 135300

Contents lists available at ScienceDirect

Journal of Molecular Structure

journal homepage: www.elsevier.com/locate/molstr

Phylogenetic analysis of *Prunus* genus using nuclear and chloroplast gene markers as a bioorganic structure profiling

Shamiran Salih Abdulrahman[a], Sevgi Durna Daştan[b,*], Saleem Esmael Shahbaz[c], Zeliha Selamoglu[d]

[a] Zakho University, Faculty of Science, Department of Biology, Duhok, Iraq
[b] Sivas Cumhuriyet University, Faculty of Science, Department of Biology, Sivas, Turkey
[c] Duhok University, Faculty of Agriculture and Forestry, Duhok, Iraq
[d] Niğde Ömer Halisdemir University, Faculty of Medicine, Department of Medical Biology, Nigde, Turkey

ARTICLE INFO

Article history:
Received 24 September 2022
Revised 25 February 2023
Accepted 4 March 2023
Available online 9 March 2023

Keywords:
Chemical structure profiling
Functional chemistry
MatK
Nucleotide chemistry
Prunus

ABSTRACT

Prunus species are known as model species for the family *Rosaceae* and other woody plants with their small genome size and relatively small life cycle as woody plants. Also *Prunus* species are economically important plants in the subfamily A*mygdaloideae* of the *Rosaceae* have a controversial taxonomic history due to the lack of a phylogenetic framework. Therefore, mapping, classification, determination the genetic structure and characterization of genes as a spesific chemical and fuctional units of nucleic acids and making phylogenetic analysis in these species have important implications. The aim of this study to find any gene markers as a bioorganic – chemical structure profiling to determine genetic diversity of different *Prunus* species from Iraq by using nuclear and chloroplast DNA markers to evaluate the phylogeny of *Prunus.* A total of 12 species *Prunus* subgenus *amygdalus* were collected from Iraq. Ten gene region markers on both nuclear DNA and chloroplast DNA have been tested to evaluate the genetic variation at their molecular levels by using multiplex PCR with 10 gene region primers and sequence analysis for matK. The name of the gene regions using in this study are ITS (ITS1+ITS2+5.8S coding region), rbcL, matK, rpl16, trnL-trnF, atpB-rbcL, rp136-infA-rps8, trnK-rps16, pbsM-trnD, psbA-trnH. As a result, five different phylogenetic trees were obtained from the gene sequences of the samples and the base sequences were close to each other. The estimated value of the shape parameter for the discrete Gamma Distribution is 0.05. The nucleotide frequencies are $A = 29.41\%$, T/$U = 37.37\%$, C $= 17.52\%$, and $G = 15.67\%$. The nucleotide diversity of the sequence was low with the 0.004113 value. There were 14 different informative base site between the samples. The polymorphism observed was less in agreement with higher level of conservation of the gene regions. The *Prunus* species from Iraq region were grouped on two closed clusters on the phylogenetic trees.

1. Introduction

The *Prunus* belong to the family Rosaceae, which comprises over 100 different genera and 200–250 species distributed worldwide [1]. Flora of Iraq is rich in *Prunus* taxa, although poorly surveyed, 5 wild species of the subgenus Amygdalus are observed (*Prunus arabica, Prunus webbii, Prunus argentea, Prunus kotschyi,* and *Prunus carduchorum*) together with several other ill-defined infraspecific taxa (*Prunus argentea* var. *elaeagnifolia,* and *Prunus carduchorum* var. *glabra*) which form the bulk of food source for honey bees at the time of flowers scarcity in winter and in very early spring time. Many species of this subgenus are economically important as source of edible fruits, oil, timber and ornamentals [2]. In further, several species of *Prunus* are used as rootstocks. Very limited cases of hybridization have been recorded in flora of Iraq in isolated habitats, for instance, the hybrid *Prunus amygdalus* Batsch *x Prunus kotschyi* (Boiss. et Hoh.) Meikle and *Amydalus x sefinensis* Bornm.

Recent phylogenetic analyses of molecular data have shown *Prunus* s.l. to be monophyletic [3]. Moreover, in a recently presented monograph using morphological and molecular sequence data, Yazbek concluded that phylogenetic relationships of *Prunus* subgenus *amygdalus* is best circumscribed to include almonds and peaches [4]. The phylogenetic structure and geographic origin of *Prunus* was studied in some researches [5]. Many different factors,

* Corresponding author.
E-mail address: sdurna@cumhuriyet.edu.tr (S.D. Daştan).

https://doi.org/10.1016/j.molstruc.2023.135300
0022-2860/

such as population size, habitat area, and habitat connectivity, can influence genetic diversity. These factors primarily influence genetic diversity by influencing genetic drift, gene flow, and natural selection [6].

Heritable variation is required for microevolutionary changes and phylogenetic studies. In other words, the resulting characteristics should be transferred to generations. The other important feature of molecular structure data is that they allow the estimation of the separation time of species using the molecular clock. If there is information on the mutation rates for a particular species or taxon genome (mitochondria or nuclear genome) or for certain regions of the genome, it is possible to calculate the nucleotide divergences between the taxa compared and to estimate how long ago they differentiated. It is also possible to calculate the branch lengths of phylogenetic trees formed in this way and to estimate the evolutionary relationships between species temporarily. The chloroplast DNA gene sequences as a biomolecular marker are very important to distinguish plant species and to reveal phylogenetic affinity relationships at the species level or sub-species level. Most of these regions tested are plastid genome regions, such as MATK (Megakaryocyte-Associated Tyrosine Kinase), rbcL (ribulose-bisphosphate carboxylase gene), rpoB (encoding the β-subunit from the plastid genome), rpoC1 (encode Beta subunits of chloroplast) from coding regions and atpF–atpH, trnH–psbA and psbK–psbI regions (these genes are as potential plant DNA barcodes using) from non-coding regions. Nevertheless, the Internal Transcribed Spacers (ITS1 and ITS2) of nuclear gene regions are commonly used [7]. The matK region is one of the coding regions in the plastid genome which is the most diverse in plants [8]. As a result of BLAST analyses, the highest values in match success at the species level were found to be matK (99%), trnH-psbA (95%), and rbcLa (75%), respectively. While the highest rate (98%) was obtained in species identification performed using three loci (matK- trnH-psbA- rbcLa), it was followed by trnHpsbA/rbcLa (95%) and matK-rbcLa (92%) combinations, respectively [9].

The aims of the study are; 1. To use some gene regions of nuclear DNA and chloroplast DNA as abiochemical markers for the molecular and structural characterization and pofilling within and between 12 species of *Prunus* subgenus *amygdalus* according to the nucleotide structural polymorphisms in Iraq. 2. Comparing the *matK* gene sequence of these species to the sequences of same gene in the International Gene bank in NCBI to study the genetics relationship and to understand their biodiversity for taxonomy and identification. 3. To understand the evolutionary relationships of *Prunus* subgenus *amygdalus* among *Prunus* taxa.

2. Materials and methods

2.1. Plant material

The aerial parts of the plant comprising leaves of *Prunus* subgenus *amygdalus* the field trips started early 2016 and continued during 2017 throughout the year from various localities of different provinces in Iraq. Field notes were recorded for each specimen together with latitude, longitude, and altitude using Global Positioning System (GPS) data in Table 1. Which include 12 species of *Prunus subgenus amygdalus* (*P. persica, P. dulicis, P. arabica, P. kotschyi, P. webbii, P. longispinosa, P. persica nectarine, P. argentea* var. *argentea, P. argentea* var. *elaeagnifolia, P. carduchorm* var. *carduchorm, P. carduchorm* var. *glabra, P. dulicis* var. *spontanea)* were used for molecular structure analyses. The plant leaves samples were collected for every plant to extraction of whole genomic DNA.

2.2. The homogenization of plant leaves

The plant samples from leaves were ground to powder in liquid nitrogen using a mortar and pestle. Allsheng-BioPrep-6 fully automatic tissue homogenizer was used. Plant powders were taken in to a 1.5 mL tube and 500 µL plant lysis buffer, 10 µL proteinaz K solution and silica beads were added later vortexed for 15 s and left for 60 min at 60 °C incubation. The samples were then centrifuged at 12.000 rpm at 4 °C for 15 min, the supernatant portion was transferred to the new 1.5 mL eppendorf tube by precipitating the tissue and cell particles.

2.3. Total genomic DNA isolation

DNA solutions were obtained by Fully Automatic DNA / RNA Extraction Device (Bioneer ExiPrep16 Plus). Bioneer Genomic DNA Mini Kit for plant samples was used for whole genomic DNA isolation of the samples. Isolation was performed using the kit protocol (Bioneer Plant Genomic DNA Extraction Kit) in accordance with the manufacturer's instructions. The determination of DNA quality and concentration in samples were performed using Nano drop spectrophotometry (DeNovix DS-11+) instrument.

2.4. Multiplex polymerase chain reaction (PCR) analysis

Bioneer Brand Primer sets (S-1001) were used. All the primer sequence data are given in the Table 2. PCR conditons were performed with the Bioneer K-2111 Multiplex PCR Premix kit. The Multiplex PCR Premix protocol we used recommends the addition of 50-100 ng of DNA to prepare 20 µL reaction volumes. These mixes contain all the material required for PCR conditions and the user uses it only by adding primer and sample DNA. Primers and sample DNA are added to ready reaction tubes, and the final sample volume is completed to 20 µL with PCR Grade water and it is studied. The ExiSpin Model Vortex-Mixer (Bioneer), which performs both vortex and spinning process in a programmed and automatic way, was used to ensure homogeneity before the tubes were loaded on the TurboCycler Model PCR (Blue-Ray**)** device. For seeing the DNA fragment, 2% agarose gel for DNA fragments between 100 and 2500 bp was prepared. Gel images were recorded with the Gel Imaging System (Syngene G: BOX F3) and analyzed with Syn gene Gene Tools Model Analysis software. The Molecular Weight calculation was performed using DNA ladder and the mapping (yes / no) table and dendogram map were generated. The degree of affinity of the samples whose molecular weight was calculated was determined as species and subspecies.

2.5. PCR application for sequence analysis

Amplification was performed for both gene regions by using ITS1-4 and *MatK* gene region primers by using multiplex PCR. The Optimus 96 G Gradient Thermal Cycler and Wizbio PCR Master Mix set were used for amplifications. As a result of the PCRs performed, sequence analysis was started with 12 samples, and the ExoSAP protocol was applied since a single band structure of gene regions was observed on the gel. The sequence was established bidirectionally with forward and reverse primer. Applied Biosystems Simpli-Amp and Thermal Cycler device were used for the sanger sequence method. Purification was performed by the sephadex gel filtration method.

2.6. Data analysis

Contig was created by combining raw sequence sequences obtained in (.ab1) format from sequence device software, and forward and reverse readings in Geneious Prime 2019 program. NCBI

Table 1
Details of specimens used for DNA analysis.

Taxa	Place of collection	Altitude	Longitude	Latitude
P. dulcis	Dohuk/Zakho	782m	37°16′50″ N	42°59′17"E
P. arabica	Dohuk/Zakho	570m	37°06′17"N	42°40′16"E
P. kotschyi	Duhok/Kara	1991m	37°00′55"N	43°20′45"E
P. webbii	Duhok/Kara	1540m	37°00′05"N	43°19′31"E
P. longispinosa	Suliamaniya/Han qul	1384m	35°19′09"N	45°04′02"E
P. persica	Dohuk/Zakho	637m	37°06′43"N	43°03′14"E
P. persica nectarina	Dohuk/Zakho	635m	37°08′39"N	43°05′15"E
P. dulcis var. *spontanea*	Dohuk/Zakho	776m	37°18′44′'N	42°57′18"E
P. argentea var. *argentea*	Duhok dam	723m	36°53′08"N	43°01′17"E
P. argentea var. *elaeagnifolia*	Duhok dam	812m	36°44′09"N	43°05′20"E
P. carduchorum var. *carduchorum*	Suliamaniya/Harman	929m	35°20′02"N	46°05′35"E
P. carduchorum var. *glabra*	Ahmed awa	887m	35°19′02"N	46°04′28"E

Table 2
Primer sets which have used in this study.

Gene Region	Primer sequence
ITS	ITS1 5'- TCC GTA GGT GAA CCT GCG G-3'
	ITS4 5'- TCC TCC GCT TAT TGA TAT GC-3'
RbcL	1F 5'- ATG TCA CCA CAA ACA GAA AC-3'
	724R 5'- TCG CAT GTA CCT GCA GTA GC-3'
MatK	F 5'- CGT ACA GTA CTT TTG TGT TTA CGA G-3'
	R 5'- ACC CAG TCC ATC TGG AAA TCT TGG TTC-3'
rpl16	F 5'- TGT TTA CGA AAT CTG GTT CT-3'
	R 5'AAC CAG ATT TCG TAA ACA AC-3'
TrnF	Tab 5' CGA AAT CGG TAG ACG CTA CG-3'
	TabF 5'ATT TGA ACT GGT GAC ACG AG-3'
atpB-	atpB-1 5'- ACATCK ART ACK GGA CCA ATA A-3'
rbcL	rbcL-1 5'- AAC AAC AGC TTT RAA TCC AA-3'
rp136-	rp136F 5'- CAC AAA TTT TAC GAA CGA AG-3'
infA-	rps8R 5'- TAA TGA CAG AYC GAG ARG CTC GAC-3'
trnK-	trnk 5'- TAC TCT ACC RTT GAG TTA GCA AC-3'
rps16	rps16modF 5'- AAA GGK GCT CAA CCT ACA RGA AC-3'
pbsM-	pbsM1 5'- GCG GTA GGA ACT AGA ATA AAT AG-3'
trnD	trnD 5'- GGG ATT GTA GTT CAA TTG GT-3'
TrnH	psbA 5'- GTT ATG CAT GAA CGT AAT GCT C-3'
	trnHGUG 5'- CGC GCA TGG TGG ATT CAC AAT CC-3'

Blast scanning was performed after completing all editing processes (correction of bad readings). *MatK* gene sequence data belonging to the same species group were downloaded from the Gen Bank database so that the analysis would be more reliable. Alignment analysis was performed on a total of 30 sequences (NCBI+original sequences) in the Cluster W program. Neighbor-Joining (NJ) analysis, UPGMA tree, and distance matrix analyses were performed using various analysis programs in the Geneious program. 5 trees were drawn using the TBR (the Tree-Bisection-Regrafting) algorithm with 1000 replicates and a tree length of 37 steps using the MEGA X program for maximum parsimony analysis, and the most appropriate one of them was selected. Similarly, the base number of Parsimony Informative in the data set was determined using the MEGA X program. Segregating sites and nucleotide diversity were determined by the Tajima-Nei neutrality test. The evolutionary divergence matrix was created by synonymous and nonsynonymous distance. Gamma estimation values and nucleotide frequencies of sequences were determined by the "Gamma parameter test". JMODELTEST and MEGA programs were used to determine the best model to be used in maximum Likelihood analysis, and the GTR+G model with the lowest AIC value was selected. For maximum Likelihood analysis, the data in FASTA format were converted into PHYLIP format by the ALTER (http://www.sing-group.org/ALTER/) online tool, and the tree was drawn with 1000 bootstrap using the PHYML program (using the plugin in Geneious Prime). A tree was created in terms of sequence similarities as a result of PhylML.

3. Results and discussion

3.1. Multiplex PCR study results

Multiplex PCR study was conducted with ITS, rbcL, matK, rpl16, trnF, atpB-rbcL, rp136-infA-rps8, trnK-rps16, pbsM-trnD and trnH primers using the multiplex PCR kit, and it was run in 2% agarose gel under suitable conditions (Fig. 1). The fragmental structure of different gene regions of DNA organic compound observed in agarose gel image obtained was analyzed by the Genetools program. The bands obtained according to the PCR result of each plant sample were shown in agarose gel using DNA ladder (Maestrogen, AccuRuler 100 bp Plus DNA Ladder), and they were compared in the match (yes/no) table. The bands obtained by multiplex PCR method in plant samples together with Gene tools software were compared both in terms of base size and density. Using the Genetools software, the molecular weights were calculated according to the number and type of bands carried by each plant sample, and band densities (Fig. 1), and a dendrogram map was created (Fig. 2). Furthermore, the similarity matrix was obtained by comparing the band profiles obtained from the data, such as band density and number of bands, obtained by displaying the Multiplex PCR results of plant samples in the SynGene Gel imaging system in each sample (Fig. 3). According to the similarity matrix, each plant sample was individually compared with DNA ladder band profiles and converted to numerical similarity ratios.

3.2. Sequence analysis of matk gene region from chloroplast DNA

The fragment size of the *MatK* gene region is 886 bp (Fig. 4). By using the sequence data of *MATK* gene regions obtained from the study, analyses were performed using various phylogenetic programs to reveal the similarities and distances between the existing plant species. GenBank scanning was performed to strengthen our phylogenetic analysis since our sample group in the present study was generally small and almost every species was included in the study in a way to be represented by a single individual. The sequences related to these gene regions of the existing species in our study and literature data were first reviewed while performing GenBank scanning. However, *MATK* gene region data were not found apart from *Prunus dulcis (Prunus amygdalus), Prunus persica*, and *Prunus webbii* species. Therefore, the sequences of sequencing studies conducted with other species in the same species were taken from GenBank and involved in phylogenetic analyses along with the sequence data of the species in our study. Thus, they were compared with the species in this study and with other species of

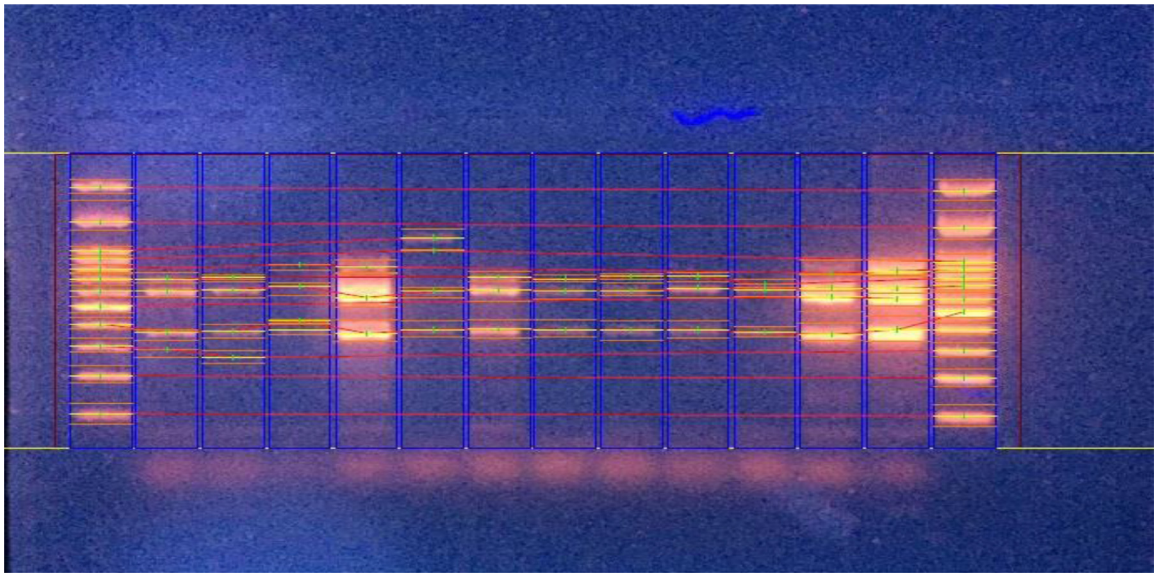

Fig. 1. Determination of the molecular weight of the DNA bands obtained by multiplex PCR in Genetools program for each sample according to the band density and number.

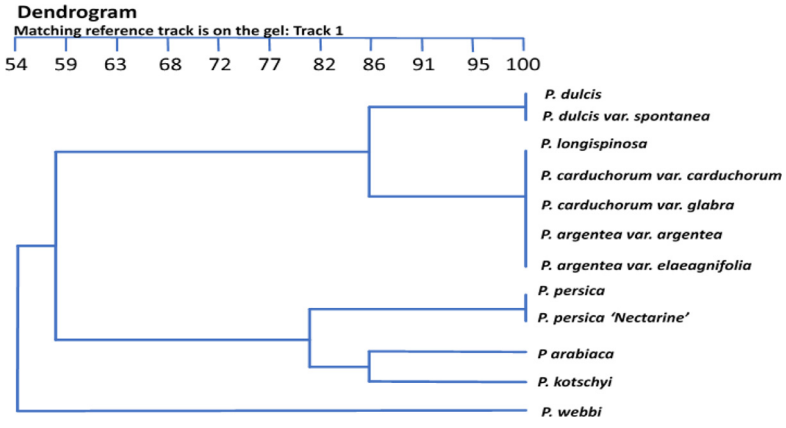

Fig. 2. The dendogram obtained by the molecular weight of the bands for 10 gene regions.

the same genus included in other studies in the literature. With this study, the sequence data of some species in our sample group will first be uploaded to GenBank. The sequence data and Accession numbers of other studies downloaded from GenBank are presented in the following table (Table 3).

Contig was created by combining raw sequences obtained in (.ab1) format after sequencing, and forward and reverse readings in Geneious Prime 2019 program. After all editing and trimming processes (correction of bad readings) were completed, the sequences that seemed to be longer in agarose gel electrophoresis were slightly shortened and NCBI Blast scanning was performed in this way. Alignment analysis was performed on a total of 30 sequences, including species-specific new sequences in this study along with the sequences obtained from the NCBI site, in the Clustal W program. Neighbour-Joining (NJ) analysis, UPGMA tree, and distance matrix analyses were performed using various analysis programs in the Geneious program. The NJ tree, the UPGMA tree, and the phyML tree (Fig. 5,6) obtained from the Neighbour-Joining (NJ) analyses of plant samples were found to be as follows. The branches with species in all trees were found to be generally similar. *P. persica* and *P. persica nectarina* were found on the same evolutionary branch in all tree drawings. Both species and how much part of the sequences of species are similar were shown in the phyML sequence similarity dendrogram created according to the sequence similarities of the species (Fig. 5).

While comparing the sequences of our 12 plant samples, 5 trees were drawn using the TBR (the Tree-Bisection-Regrafting) algorithm with 1000 replicates and a tree length of 37 steps using the MEGA X program for maximum parsimony analysis, and the most appropriate one of them was selected (Fig. 6). The evolutionary history was inferred using the Maximum Parsimony method. The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analyzed [10]. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The MP tree was obtained using the Tree-Bisection-Regrafting (TBR) algorithm [11]. This analysis involved 30 nucleotide sequences. There were a total of 915 positions in the final dataset. Evolutionary analyses were conducted in MEGA X [12].

Similarly, the Parsimony Informative base number in the data set was determined using the MEGA X program, and it was presented in the Fig. 5-6. In general, variations that could be used as informative were found in 14 different nucleotide sequences. Segregating sites and nucleotide diversity were determined by the Tajima-Nei neutrality test. According to the Tajima-Nei neutrality test, the nucleotide diversity of 30 sequence data entered into the

	Track 1	Track 2	Track 3	Track 4	Track 5	Track 6	Track 7	Track 8	Track 9	Track 10
Track 1	1,000	0,500	0,500	0,375	0,400	0,500	0,400	0,400	0,400	0,400
Track 2	0,500	1,000	1,000	0,500	0,571	0,500	0,857	0,857	0,857	0,857
Track 3	0,500	1,000	1,000	0,500	0,571	0,500	0,857	0,857	0,857	0,857
Track 4	0,375	0,500	0,500	1,000	0,857	0,500	0,571	0,571	0,571	0,571
Track 5	0,400	0,571	0,571	0,857	1,000	0,571	0,667	0,667	0,667	0,667
Track 6	0,500	0,500	0,500	0,500	0,571	1,000	0,571	0,571	0,571	0,571
Track 7	0,400	0,857	0,857	0,571	0,667	0,571	1,000	1,000	1,000	1,000
Track 8	0,400	0,857	0,857	0,571	0,667	0,571	1,000	1,000	1,000	1,000
Track 9	0,400	0,857	0,857	0,571	0,667	0,571	1,000	1,000	1,000	1,000
Track 10	0,400	0,857	0,857	0,571	0,667	0,571	1,000	1,000	1,000	1,000
Track 11	0,400	0,857	0,857	0,571	0,667	0,571	1,000	1,000	1,000	1,000
Track 12	0,500	0,500	0,500	0,750	0,857	0,500	0,571	0,571	0,571	0,571
Track 13	0,500	0,500	0,500	0,750	0,857	0,500	0,571	0,571	0,571	0,571
Track 14	0,917	0,500	0,500	0,375	0,400	0,375	0,400	0,400	0,400	0,400

	Track 11	Track 12	Track 13	Track 14
Track 1	0,400	0,500	0,500	0,917
Track 2	0,857	0,500	0,500	0,500
Track 3	0,857	0,500	0,500	0,500
Track 4	0,571	0,750	0,750	0,375
Track 5	0,667	0,857	0,857	0,400
Track 6	0,571	0,500	0,500	0,375
Track 7	1,000	0,571	0,571	0,400
Track 8	1,000	0,571	0,571	0,400
Track 9	1,000	0,571	0,571	0,400
Track 10	1,000	0,571	0,571	0,400
Track 11	1,000	0,571	0,571	0,400
Track 12	0,571	1,000	1,000	0,500
Track 13	0,571	1,000	1,000	0,500
Track 14	0,400	0,500	0,500	1,000

Track 2 : *P. dulcis*	**Track 8: *P. argentea var.elaengnifolia***
Track 3: *P.dulcis var. spontanea*	**Track 9: *P. logispinosa***
Track 4: *P. arabica*	**Track 10: *P. carduchorum var. carduchorum***
Track 5: *P. kotschyi*	**Track 11: *P. carduchorum var. glabra***
Track 6: *P. webbi*	**Track 12: *P. persia***
Track 7: *P. argentea var. argentea*	**Track 13: *P. persia 'Nectarine'***
Track 1 and Track 14 DNA Ladder	

Fig. 3. The similarity matrix generated by comparing the band profiles obtained by displaying the Multiplex PCR results of the plant samples on the SynGene Gel imaging system.

Mega X program was found to be 0.004113. The number of segregating sites determined according to each sequence data was 33, and the sample separation efficiency of each nucleotide was found to be 0.009104. The evolutionary divergence matrix was also created by synonymous and nonsynonymous distance in the Mega X program. The nucleotide differences between the samples were low. The similarity ratios between the sequences were found to be generally high between our 12 plant samples and also the other sequence data from gene bank. On the other hand, gamma estimation values and nucleotide frequencies of sequence data were determined by the "Gamma parameter test" in the Mega X program [13]. The estimated value of the shape parameter for the dis-

Table 3
The analysis programme output of Parsymony informative base numbers and base site in plant species for *MatK* gene sequence (There are 14 different informative base site).

Prunus longispinosa	G	T	A	T	T	G	G	C	G	A	A	C	C	-
P africana JQ024985	-	-	.	.	.	.	.	A	.	.	.	.	.	-
P angustifolia KJ773042	-	-	.	.	.	.	.	A	.	.	G	T	A	-
P armeniaca JF955812	T	C	.	.	.	.	.	A	.	.	.	.	A	-
P armeniaca JF955811	T	C	.	.	.	.	.	A	.	.	.	.	A	-
P armeniaca var.*armeniaca* KP089844	.	.	.	.	C	.	.	A	.	.	.	.	.	-
P davidiana JF955801	.	.	.	C	.	A	A	A	R	G	.	.	-	-
P domestica subsp. *insititia* KX229966	T	C	.	.	.	.	.	A	.	.	.	.	A	T
P dulcis HQ235078	.	.	.	.	.	.	.	.	.	.	.	.	.	-
P geniculata KJ773044	-	-	.	.	.	.	.	A	.	.	G	.	A	-
P mume JF955797	.	.	.	.	C	.	.	A	.	.	.	.	.	-
P nigra MG220723	.	.	.	.	.	.	.	A	.	.	G	T	A	-
P pensylvanica MG220935	.	.	.	C	.	.	.	A	.	.	.	.	.	T
P persica HQ235227	.	.	.	.	.	.	.	A	.	.	.	.	.	-
P persica nucipersica JF955829	.	.	.	.	.	A	A	A	R	.	.	.	-	-
P salicina KX238393	T	C	.	.	.	.	.	A	A	.	.	.	-	-
P salicina KX238395	T	C	.	.	.	.	A	A	A	.	.	.	-	-
P umbellata KJ773045	-	-	.	.	.	.	.	A	.	.	G	T	A	-
P webbii HQ235312	.	.	.	.	.	.	.	.	.	.	.	.	.	-
Prunus arabica	.	.	.	.	.	.	.	.	.	.	.	.	.	C
Prunus argentea var. *argentea*	.	.	G	.	.	.	.	A	.	.	.	.	.	-
Prunus argentea var. *elaeagnifolia*	.	.	.	.	.	.	.	.	.	.	.	.	.	-
Prunus carduchorum var. *carduchorum*	.	.	.	.	.	.	.	.	.	.	.	.	.	-
Prunus carduchorum var. *glabra*	.	.	.	.	.	.	.	.	.	.	.	.	.	C
Prunus dulcis	.	.	.	.	.	.	.	.	.	.	.	.	.	C
Prunus dulcis var spontanea	.	.	.	.	.	.	.	.	.	.	.	.	.	C
Prunus kotschyi	.	.	G	.	.	.	.	A	.	.	.	.	.	-
Prunus persica	.	.	.	.	.	.	.	A	.	G	.	.	.	C
Prunus persica nectarine	.	.	.	.	.	.	.	A	.	G	.	.	.	C
Prunus webbii	.	.	G	.	.	.	.	.	.	.	.	.	.	-

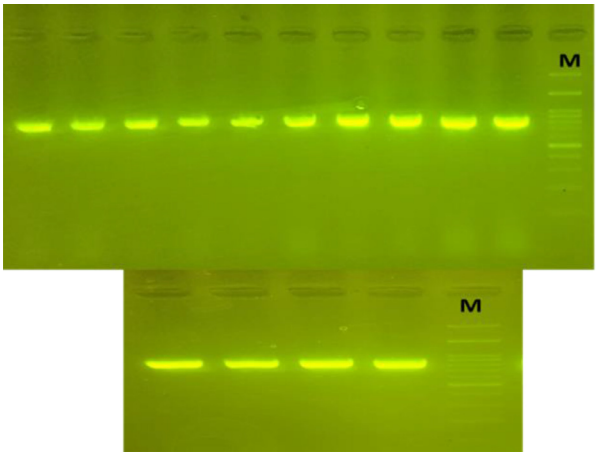

Fig. 4. Image of MATK gene region in agarose gel.

crete Gamma Distribution is 0.0500. Substitution pattern and rates were estimated under the General Time Reversible model (+G) [11]. Also, the nucleotide frequencies are $A = 29.41\%$, $T/U = 37.37\%$, $C = 17.52\%$, and $G = 15.67\%$.

Within the scope of this study, molecular genetic analyses were performed using 12 different plant samples including *P. persica, P. dulcis, P. arabica, P. kotschyi, P. webbii, P. longispinosa, P. persica nectarine, P. argentea var. argentea, P. argentea var. elaeagnifolia, P. carduchorm var. carduchorm, P. carduchorm var. glabra, P. dulcis var. spontanea* of the genus Prunus and subgenus *amygdalus* in Iraq. Using the total genomic DNA of 12 plant samples, some nucleus and chloroplast genes were amplified, and a sequencing study was conducted, and significant results were obtained about the phylogenies of these species. According to the results of the dendrogram obtained by comparing 10 different gene regions in terms of molecular size and quantity for the species investigated based on the multiplex PCR method, 12 *Prunus* samples collected from various regions in Iraq were generally gathered in 2 clusters. Accordingly, *P. kotschyi, P. persica, P. persica nectarina,* and *P. arabica* species were found in a cluster close to each other and also formed the second cluster on the dendrogram in other species except for *P. webbii*. Furthermore, *P. persica* and *P. persica nectarina* species in the first cluster were located on branches closer to each other, and it was found out that the nucleotide difference between them was very low and the band sequence similarities were very high according to multiplex PCR results in a way to support it. Indeed, it is observed in Fig. 2 that the band similarities of both species are 100% and their similarity values are 1.0 on the matrix (Fig. 3). On the other hand, while *P. kotschyi* and *P. arabica* species were also found on branches close to each other in the same cluster, they slightly differed from *P. persica* and *P. persica nectarina* species, but they were still in the same cluster (Fig. 2). In the second cluster on the dendrogram, *P. dulcis* species was on the same branch, while other species were on another sub-branch within this cluster (Fig. 2-3). *P. webbii* species was separated from other species much more differently. However, the similarity rate with almost all species was different from the same distance (Fig. 2-3).

According to the neighbor joining, UPGMA and PHYLML tree obtained from the sequence data, *Prunus* samples were separated from *Prunus* samples obtained from the NCBI data and settled in two clusters. While *P. persica* and *P. persica nectarina* samples were on the same branch, they were found on the branches close to each other in this cluster with the other *P. persica* sample downloaded from GenBank. *P. webbii* was located on a branch in the middle of the cluster where the Iraqi samples belong, but not out of all 12 samples, in a way to be slightly different from the dendrogram obtained from the nuclear DNA data. According to the NJ, PHYLML and UPGMA trees, *P. kotchyi* and *P. arabica* samples were located on the branches in different clusters and a little further away from each other. However, according to all 3 trees obtained from the se-

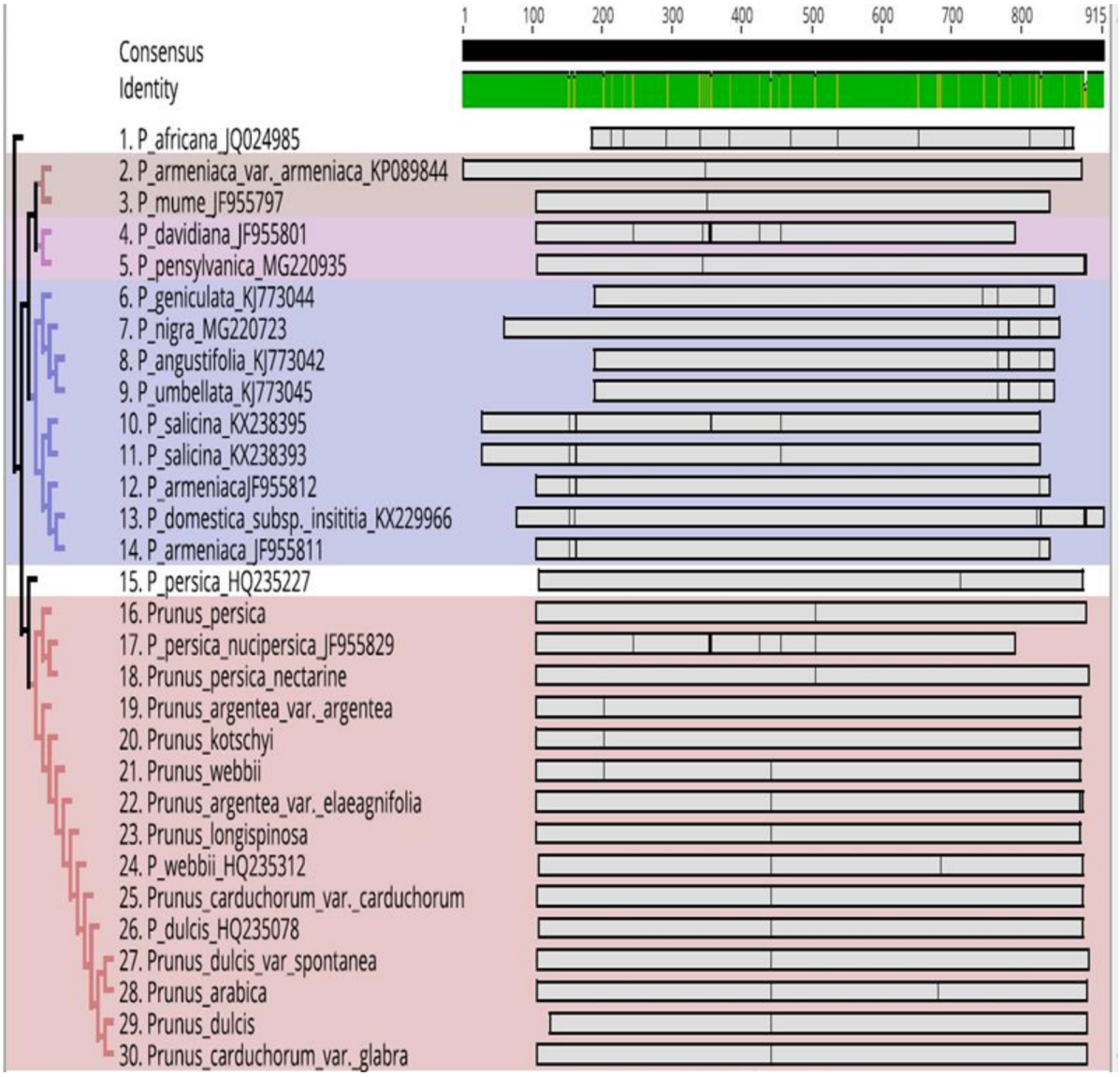

Fig. 5. PhyML sequence similarity dendogram between all plant species in this study according to the sequence data.

quence data, all samples obtained from Iraq region were branched to form two clusters. On the other hand, while a few of the species downloaded from the NCBI data were in the same cluster with Iraq samples, it was observed that these species were generally separated to form the third cluster. The supporting values of the branches where the species existed were found to be generally below 50%. It may be due to the fact that only one sample of each species was collected and that analyses were performed on this one sample (Figs. 2,3 and 5,6).

Furthermore, according to the PHYLML sequence similarity dendrogram obtained from the comparison of sequence similarities and differences, a high level of similarity was found in the sequences among all species. While it was observed that the sequence series of the species downloaded from the NCBI data and 12 samples collected from Iraq region were slightly different and distant from each other, the similarity rates of Iraq samples within themselves were found to be higher. According to the maximum parsimony analysis obtained from the sequence data of the species, it was observed that the samples collected from the Iraq region generally came together as one cluster, and in total, all species formed 3 clusters with different colors on the tree. Interestingly, *P. persica* and *P. persica nectarina* samples on this tree were placed in another cluster in a way to be on the branches very close to each other, away from Iraq and close to the species from the GenBank. While the supporting values on the branches were found to be generally less than 50% in the clusters with Iraq samples, the supporting values on the branches with species from the Genbank were found to be higher than 50%.

We used in this study both maximum likelihood and parsimony trees. Because nowadays, it has been suggested that maximum parsimony analyses of DNA data may not reproduce evolutionary history reliably, as parsimony always looks for the shortest route between sequences, and accepts the simplest explanation for the data. Maximum likelihood analysis uses probabilities of change from one character to another to calculate the likelihood for given phylogenetic tree. Unlike maximum parsimony, ML takes the possibility of multiple mutational events at the same site into account. And the trees from both analyses were supporting eachother with some small differencies.

There are almost 400 species within the genus *Prunus* [14]. The taxonomy and evolutionary relationships of members of *Prunus* genus have not been clear despite extensive investigation. The tax-

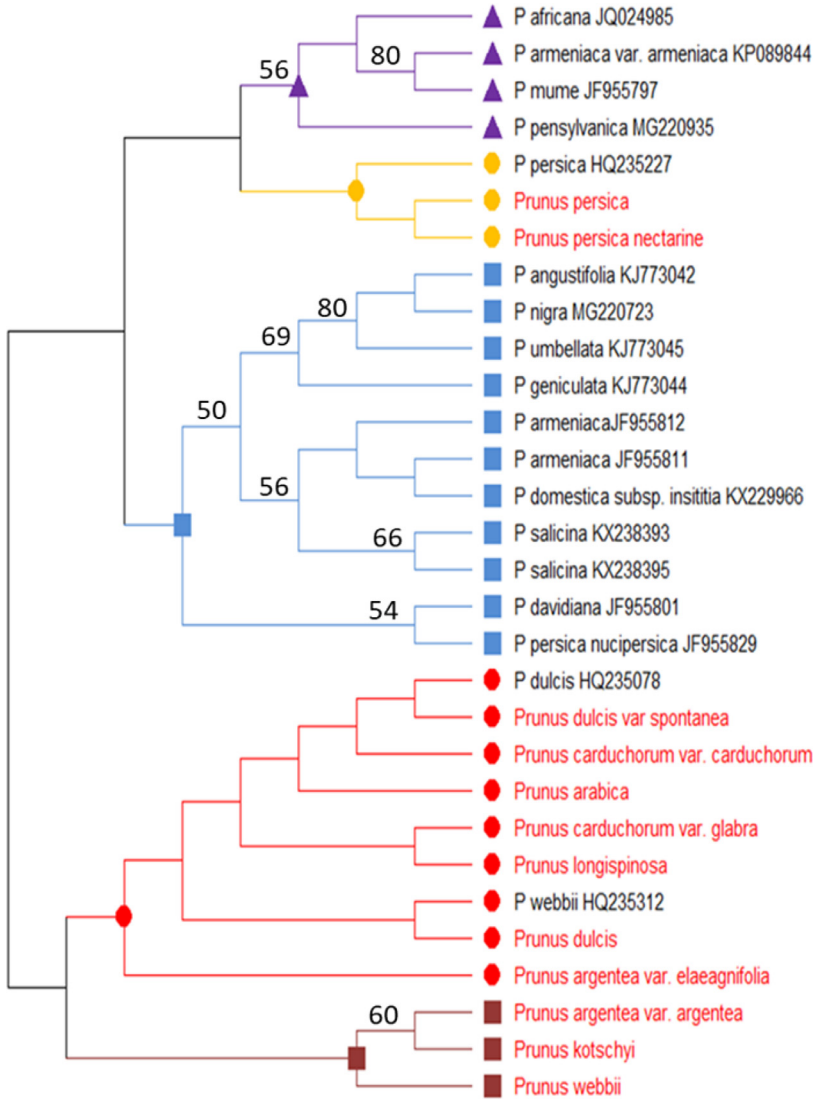

Fig. 6. Maximum Parsimony analysis of taxa. (The supporting values under 50% are not given on the branch).

onomy of *Prunus,* especially concerning the generic delimitation, has been controversial. The controversy on the classification of *Prunus* largely results from the lack of a phylogenetic studies. Generally, in the past some researchers mostly worked a few characters, especially fruit morphology, inflorescence type, and leaf duration in formulating the classification [2,15–17]. The character-based classification systems of *Prunus* s.l. and the Amygdaloideae need to be evaluated within a phylogenetic perspective. A few studies attempted to construct the evolutionary relationships of the genus [18–20]. The analysis by [19] examined the isozyme variation of 34 species belonging to three (subgenera *Amygdalus, Cerasus,* and *Prunus*) of the five subgenera [21].

Some researchers found that some registered Iranian almond cultivars appeared to be genetically related to Spanish and US commercial varieties, and concluded that this probably reflected gene flow between commercial varieties and local materials [22]. Also in some study it have reported that the Italian almond population show the largest level of mixed ancestry among western areas of Mediterranean basin [23]. Whatever the origin of this shared genetic variation, traditional almond varieties are completely unrelated to commercial varieties.

Wild *Prunus* species demonstrate high resistance to biotic and abiotic stress as well as provide valuable resources for breeding [24,25]. Controlled hybridization and selection programs are implemented to favourad vantageous traits in diverse ways, but it may also result in the loss of genetic diversity in most annual species [26]. Wild species are important sources of genetic diversity if they can be crossed with crop plants [27]. There are very few studies about the phenotypic diversity, distribution and morphological characterization about Iraq *Prunus* species related to the ecology and population structure of *Prunus* species. Description of such resources is important in order to identify areas of diversity, to utilize the genetic resources in conservation and breeding [25]. Worldwide, many markers have been used to identify the *Prunus* species genotypes such as isozymes [28–31], AFLP (Amplified fragment length polymorphism) [32], ISSR (Inter Simple Sequence Repeat) [33], RAPD (Random amplification of polymorphic DNA) [29] and SNP (Single-nucleotide polymorphism) [34], and Simple sequence repeat (SSR) or microsatellite markers.

In addition, in another research, a phylogenetic study in *Prunus* species in Afghanistan by using nuclear and chloroplast DNA markers was evaluated [35]. According to this research, the genetic diversity of different accessions of several wild almond species including *P. eleagnifolia, P. hausknechtii, P. scoparia* and *P. lycioides*, endemic to Irano-Afghan plate. In addition, five cultivated almond cultivars (Marta from Spain; Nonpareil and Mission from USA; Fer-

rangnes from France and Tuono from Italy) were included in the study. In 45 accessions, the number of alleles per locus in nuclear microsatellites ranged from 7 to 16, and expected heterozygosity varied between 0.54 and 0.93 with average PIC value of 0.81. It showed that they represent rather polymorphic species. In the case of chloroplast SSR, the polymorphism observed was much lesser in agreement with higher level of conservation of the chloroplast DNA. They obtained six logical clusters or groupings corresponding to the different subgenus and sections.

4. Conclusions

This is the first report about some *Prunus* species from Iraq also first report using and comparing molecular markers on *Prunus* species distributing in Iraq. Therefore, the results obtained this research are unique and valuable. Our research will guide the future studies. In this work, nuclear and chloroplast DNA regions are used to define the phylogenetic relationships among the 12 different taxa in *Prunus* genus. The actual degree of polymorphism and the genetic relationship in these *Prunus* samples are probably be underestimated with the several informative base changing. The samples are not so diverged with the each other and their sequence are very similar. So these species are closed according to their gene sequence, similar nucleotide base design, high similarity value and low distance value. But together with this result, the *Prunus t*axa from Iraq were collected generally in two cluster and separately clustered from *Prunus* species belonging to different countries downloaded from genbank data.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Data will be made available on request.

Acknowledgements

This study is extracted from the PhD thesis of Shamiran Salih Abdulrahman. Prof. Dr. Saleem Esmael Shahbaz was the supervisor of this PhD thesis. Prof. Dr. Zeliha Selamoglu was the second supervisor of this PhD thesis and Assoc. Prof. Dr. Sevgi Durna Daştan was the co-authors of this article and they supported this thesis work in all experiments, in all analysis, in making statistics, in making bioinformatics and writing skills. All co-authors provided support at all stages within this thesis period and the article.

References

[1] V. Shulaev, S.S. Korban, B. Sosinski, A.G. Abbott, H.S. Aldwinckle, K.M Folta, R.E. Veilleux, Multiple models Rosaceae genomics, Plant Physiol. 147 (2008) 985–1003.
[2] S. Lee, J. Wen, A phylogenetic analysis of Prunus and the Amygdaloideae (Rosaceae) using ITS sequences of nuclear ribosomal DNA, Am. J. Bot. 88 (2001) 150–160.
[3] E. Bortiri, S.H. Oh, J. Jiang, S. Baggett, A. Granger, C. Weeks, D.E. Parfitt, Phylogeny and systematics of Prunus (Rosaceae) as determined by sequence analysis of ITS and the chloroplast trnL-trnF spacer DNA, Syst. Bot. 26 (2001) 797–808.
[4] M.M. Yazbek, Systematics of Prunus Subgenus Amygdalus Monograph and Phylogeny, Cornell University, New York, USA, 2010.
[5] S.W. Chin, J. Shaw, R. Haberle, J. Wen, D. Potter, Diversification of almonds, peaches, plums and cherries–molecular systematics and biogeographic history of Prunus (Rosaceae), Mol. Phylogenet. Evol. 76 (2014) 34–48.
[6] R. Leimu, P. Mutikainen, J. Koricheva, M. Fischer, How general are positive relationships between plant population size, fitness and genetic variation? J. Ecol. 94 (2006) 942–952.
[7] E. Filiz, I. Koc, In silico chloroplast SSRs mining of Olea species, Biodiversitas J. Biol. Divers. 13 (2012) 114–117.
[8] K.W. Hilu, H. Liang, The matK gene: sequence variation and application in plant systematics, Am. J. Bot. 84 (1997) 830–839.
[9] W.J. Kress, D.L. Erickson, F.A. Jones, N.G. Swenson, R. Perez, O. Sanjur, E. Bermingham, Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama, Proc. Natl Acad. Sci. 106 (2009) 18627–18632.
[10] J. Felsenstein, Confidence limits on phylogenies: an approach using the bootstrap, Evolution (N Y) 39 (1985) 783–791.
[11] M. Nei, S. Kumar, Molecular Evolution and Phylogenetics, Oxford University Press, Inc, New York, 2000.
[12] S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, MEGA X: molecular evolutionary genetics analysis across computing platforms, Mol. Biol. Evol. 35 (2018) 1547–1549.
[13] K. Tamura, M. Nei, S. Kumar, Prospects for inferring very large phylogenies by using the neighbor-joining method, Proc. Natl Acad. Sci. 101 (2004) 11030–11035.
[14] S.L. Downey, A.F. Lezzoni, Polymorphic DNA markers in black cherry (Prunus serotina) are identified using sequences from sweet cherry, peach, and sour cherry, J. Am. Soc. Hortic. Sci. 125 (2000) 76–80.
[15] B.G. Baldwin, M.J. Sanderson, J.M. Porter, M.F. Wojciechowski, C.S. Campbell, M.J. Donoghue, The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny, Ann. Missouri Botanical Garden 82 (1995) 247–277.
[16] D.M. Hillis, M.T. Dixon, Ribosomal DNA: molecular evolution and phylogenetic inference, Q. Rev. Biol. 66 (1991) 411–453.
[17] R. McVaugh, A revision of the North American black cherries (Prunus serotina Ehrh., and relatives), Brittonia 7 (1951) 279–315.
[18] M.L. Badenes, D.E. Parfitt, Phylogenetic relationships of cultivated Prunus species from an analysis of chloroplast DNA variation, Theor. Appl. Genet. 90 (1995) 1035–1041.
[19] B.D. Mowrey, D.J. Werner, Phylogenetic relationships among species of Prunus as inferred by isozyme markers, Theor. Appl. Genet. 80 (1990) 129–133.
[20] C. Uematsu, T. Sasakuma, Y. Ogihara, Phylogenetic relationships in the stone fruit group of Prunus as revealed by restriction fragment analysis of chloroplast DNA, Japanese J. Genetics 66 (1991) 59–69.
[21] A. Rehder, Manual of Cultivated Trees and Shrubs Hardy in North America, The MacMillan Company, New York, 1940.
[22] A. Fathi, B. Gharahyazi, N.A. Hagh, M.R. Ghafari, S.M. Pirseyedi, S. Kadkhodaei, … M. Mardi, Assessment of the genetic diversity of almond (Prunus dulcis) using microsatellite markers and morphological traits, Iran. J. Biotechnol. 6 (2008) 98–106.
[23] M. Delplancke, N. Alvarez, L. Benoit, A. Espindola, H. I Joly, S. Neuenschwander, N Arrigo, Evolutionary history of almond tree domestication in the Mediterranean basin, Mol. Ecol. 22 (4) (2013) 1092–1104.
[24] T.M. Gradziel, P. Martínez-Gómez, F. Dicenta, D.E. Kester, The utilization of related Prunus species for almond variety improvement, J. Am. Pomological Soc. 55 (2001) 105–108.
[25] K. Sorkheh, R. Azimkhani, N. Mehri, M.H. Chaleshtori, J. Halász, S. Ercisli, G.C. Koubouris, Interactive effects of temperature and genotype on almond (Prunus dulcis L.) pollen germination and tube length, Sci. Hortic. 227 (2018) 162–168.
[26] B.L. Gross, K.M. Olsen, Genetic perspectives on crop domestication, Trends Plant Sci. 15 (2010) 529–537.
[27] R. Hajjar, T. Hodgkin, The use of wild relatives in crop improvement: a survey of developments in the last 20 years, Euphytica 156 (2007) 1–13.
[28] S. Arulsekar, D.E. Parfitt, D.E. Kester, Comparison of isozyme variability in peach and almond cultivars, J. Hered. 77 (1986) 272–274.
[29] H. Gouta, E. Ksia, N. Zoghlami, M. Zarrouk, A. Mliki, Genetic diversity and phylogenetic relationships among Tunisian almond cultivars revealed by RAPD markers, J. Hortic. Sci. Biotechnol. 83 (2008) 707–712.
[30] R. Hauagge, D.E. Kester, S. Arulsekar, D.E. Parfitt, L. Liu, Isozyme variation among California almond, J. Am. Soc. Hortic. Sci. 112 (1987) 693–698.
[31] M.A. Viruel, R. Messeguer, M.C. De Vicente, J. Garcia-Mas, P. Puigdomenech, F. Vargas, P. Arús, A linkage map with RFLP and isozyme markers for almond, Theor. Appl. Genet. 91 (1995) 964–971.
[32] S. Martins, F. Simões, J. Matos, A.P. Silva, V. Carnide, Genetic relationship among wild, landraces and cultivars of hazelnut (Corylus avellana) from Portugal revealed through ISSR and AFLP markers, Plant Systematics and Evolution 300 (2014) 1035–1046.
[33] M. Martins, R. Tenreiro, M.M. Oliveira, Genetic relatedness of Portuguese almond cultivars assessed by RAPD and ISSR markers, Plant Cell Rep. 22 (2003) 71–78.
[34] S.B. Wu, I. Tavassolian, G. Rabiei, P. Hunt, M. Wirthensohn, J.P. Gibson, … M. Sedgley, Mapping SNP-anchored genes using high-resolution melting analysis in almond, Mol. Genet. Genomics 282 (2009) 273–281.
[35] M. Zeinalabedini, K. Majourhat, M. Khayam-Nekoui, V. Grigorian, M. Torchi, F. Dicenta, P. Martínez-Gómez, Comparison of the use of morphological, protein and DNA markers in the genetic characterization Iranian wild Prunus species, Sci. Hortic. (2008) 116.