Australian Journal of Crop Science
AJCS
AJCS 5(10):1199-1209 (2011) ISSN:1835-2707

Molecular tagging of agronomic traits using simple sequence repeats: Informative markers for almond (*Prunus dulcis*) molecular breeding

S. Kadkhodaei[1,5,*], M. Khayyam Nekouei[1], M. Shahnazari[1], H. Etminani[1], A. Imani[2], M. Ghaderi-Zefrehei[3], M. Elahy[4], Arbakariya B. Ariff[5]

[1]Department of Genomics, Agricultural Biotechnology Research Institute of Iran (ABRII), Seed & Plant Improvement Campus, Mahdasht Rd., Karaj, Iran
[2]Departments of Horticulture, Seed and Plant Improvement Campus, Mahdasht Rd., Karaj, Iran
[3]Department of Animal Science, Faculty of Agriculture, University of Yasouj, Iran
[4]Faculty of Engineering and Science, School of Biomedical and Health Science, Victoria University, Melbourne, Australia
[5]Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia (UPM), Serdang 43400, Selangor, Malaysia

***Corresponding author: skad@abrii.ac.ir**

Abstract

Informative markers are the most applicable genetic information in breeding schemes. Association studies which soundly integrate molecular and morphological data are the best choice to find informative markers, particularly in crops that are limited to only one generation per year. Therefore, in the present research the associations between different morphological traits and highly polymorphic SSRs were studied to find possible informative markers for some morphological and/or agronomical traits in almond. In total, 39 morphological traits were recorded during two years among 53 almond genotypes/cultivars**.** Extracted almond genomic DNA was PCR-amplified using 9 pairs flanking SSRs sequences previously cloned and sequenced specifically for almond. For finding association between molecular markers and morphological traits and identification of possible informative markers, Pearson correlation and stepwise regression analysis were employed. The results revealed a significant correlation between the morphological traits and the studied microsatellite loci. A total of 141 positive markers out of 556 polymorphic bands were identified for different traits. For some of the morphological traits more than one informative marker was detected, which consequently finding their additive effects, degree of dominance and sum of the positive and negative effects need further analysis. These informative markers can be considered as postulated candidate markers for scanning the genome for related morphological (particularly agronomical) traits, mapping and finally marker assisted selection programs.

Keywords: Almond, association mapping, fruit trait, informative markers, marker assisted selection, microsatellite markers, molecular breeding, *Prunus dulcis*.
Abbreviations: Marker assisted selection (MAS), Polymorphism information content (PIC), Principal component analysis (PCA), Quantitative trait loci (QTL), Simple sequence repeat (SSR).

Introduction

Prunus & Almond

The genus *Prunus* in Prunoideae, a subfamily of Rosaceae, includes several economically important fruit tree species such as apricot (*P. armeniaca*), cherry (*P. avium*), peach (*P. persica*), plum (*P. mume*), and common almond (*P. dulcis* Mill. syn. *P. communis* Fritsch.). Almond was probably domesticated during the 3rd millennium BC. It has been suggested that this domestication could have taken place in Central Asia where wild almond trees, can still be found (Socias I Company, 1998). Twenty-six almond species form a distinct and easily identified taxonomic group in the world (Browicz and Zohary, 1996). In Iran, 21 almond species and 6 natural hybrids have been described (Khatamsaz, 1992). The basic chromosome number (x) of almonds (common and wild types) is 8 (2n=16) and its DNA contents is relatively small (0.54–0.67 pg DNA/2C) (Kadkhodaei, 2008; Dickson, 1992). Almond occupies a very peculiar place among fruit trees. Because of almond's tolerance to cold, drought and salinity, it is considered an important tree crop and is cultivated in different climatic regions of Iran. Almond culture is considered to be of great economic importance in Iran which is amongst the most important countries in the world producing *Prunus* fruits, as its ranking regarding to production of cherries, apricot and almond is the first, second and fourth, respectively. In 2007, worldwide annual production of almonds is exceeded 2065489 tons, which the contribution of Iran is 5.32% (110000 tons).

Table 1. Names and origins of the analyzed almond genotypes.

Code	Name	Origin	Code	Name	Origin	Code	Name	Origin
1	*Amygdalus elaegnifolia*	Iran	19	Talkh2	Iran	37	Hamedan20	Iran
2	*Amygdalus lycioides*	Iran	20	Yalda	Iran	38	Hamedan21	Iran
3	Arak51	Iran	21	Hamedan1	Iran	39	Hamedan22	Iran
4	Arak52.3	Iran	22	Hamedan4	Iran	40	Hamedan27	Iran
5	Arak54	Iran	23	Hamedan5	Iran	41	Hamedan29	Iran
6	Arak56	Iran	24	Hamedan6	Iran	42	Hamedan30	Iran
7	Arak58.4	Iran	25	Hamedan7	Iran	43	Hamedan55	Iran
8	Arak59	Iran	26	Hamedan8	Iran	44	Hamedan57	Iran
9	Rabee2	Iran	27	Hamedan9	Iran	45	Hamedan60	Iran
10	Sahand46.2	Iran	28	Hamedan10	Iran	46	Shiraz7	Iran
11	Shekofeh3	Iran	29	Hamedan11	Iran	47	Shiraz8	Iran
12	Tajeri12.3	Iran	30	Hamedan12	Iran	48	Shiraz17	Iran
13	Najafabad-MamaeiDobahreh	Iran	31	Hamedan13	Iran	49	Shiraz18	
14	Najafabad-Monagha	Iran	32	Hamedan14	Iran	50	Shiraz21	Iran
15	Najafabad-Rabee1	Iran	33	Hamedan15	Iran	51	NonPareil1	USA
16	Najafabad-Dobahreh	Iran	34	Hamedan16	Iran	52	Spain47.2	Spain
17	Shiraz-MonaghaShirazi	Iran	35	Hamedan17	Iran	53	Spain47.3	Spain
18	ShirazPoostnazok1	Iran	36	Hamedan18	Iran			

Almond harvesting area in Iran is about 172000 ha and annual yield in 2007 has been 639.5 kg/ha (FAOSTAT Data Sources). Almond cultivation in Iran has a long historical background, and because of its self-incompatibility nature there are many genotypes growing in different regions of the country. These genotypes include a vast range of diversity in many characteristics such as blooming time.

Molecular Markers, Molecular Breeding and Informative Markers

One of the major concerns of modern agriculture is the conservation and utilization of valuable genetic resources of crop plants. The need for correct identification applies to cultivars and accessions, independently of their mode of conservation, i.e. whether they are maintained in an *in situ* or *ex situ* field gene bank or an *in vitro* gene bank. Tools developed for the characterization for biodiversity may allow clarifications of synonyms and detection of the origin of species and cultivars. Characterization and determination of fruit tree cultivars are sometimes difficult using conventional methods. Since morphological markers are prone to equivocal interpretations and time consuming, molecular approaches should be implemented in cultivar identification and breeding programs.

Molecular markers help to distinguish labeling mistakes, identification of the genuine owner of the cultivar in question, routine identification of cultivars in nurseries and etc. Further, it simplifies the work in breeding programs by accelerating the breeding process due to allowing a selection before the first fruit crop, by tracking certain genes or genotypes among offspring of crosses (Lamier, 2005). Because of the greater speediness, efficiency and reproducibility, the use of molecular markers based on Polymerase Chain Reaction (PCR) has been the choice in genetic studies of plants and in preparing fingerprinting for many fruit species. These markers can be used for varietal control allowing a plant to be identified at any stage or vegetative cycle, and can solve cases involving plants with uncertain names and sources.

Microsattelites

Microsatellites or simple sequence repeat (SSR) are sequences of a few repeated and adjacent base-pairs well distributed over the eukaryote genome (Powell, 1996). PCR-based, SSRs markers are becoming the marker of choice for fingerprinting and genetic diversity studies for a wide range of plants (Gupta, 1996). Because of their nature of being highly polymorphic, abundance, and co-dominant inheritance, they are well-suited for the assessment of genetic variability within crop species, and of the genetic relationships among species.

In the case of *Prunus* species, primer pairs flanking SSRs have been cloned and sequenced in peach (Aranzana, 2002), apricot (Decroocq, 2003), cherry (Schueler, 2003) and almond (Testolin, 2004; Xu, 2004; Messina, 2004). These SSR markers have been used primarily for the molecular characterization and identification of cultivars in the related *Prunus* species. Also, SSRs developed from different *Prunus* species have highly potential of application for transportability across each other (Fathi, 2008; Shiran, 2007; Mnejja, 2005).

Informative markers

Informative markers are the most applicable and reliable genetic information for breeding purposes. These markers can be considered as an initial point to search the genome for the related traits.

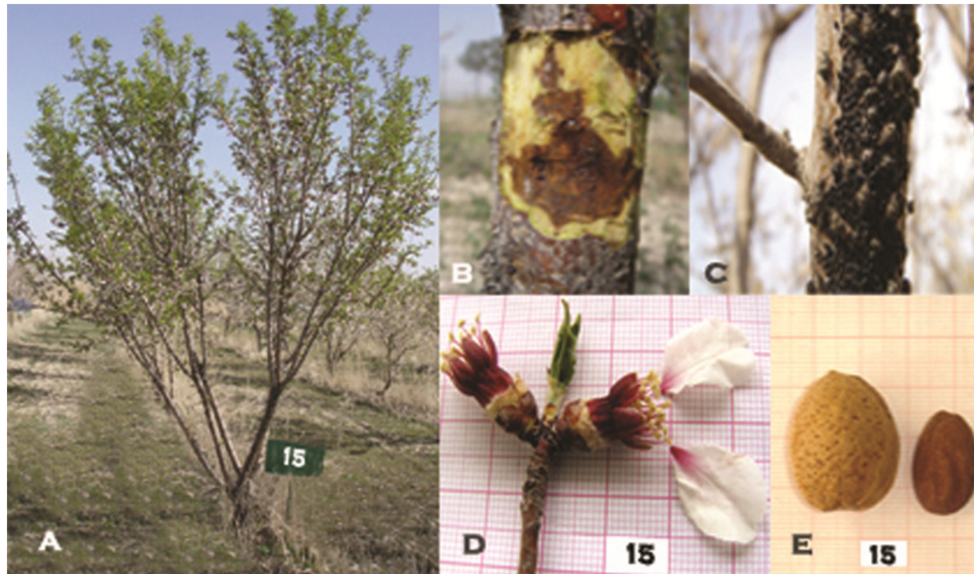

Figure 1. Some of the morphological traits evaluated for association studies. A: Growth habit, B: Bacterial Canker, C: Black Aphid, D: Flower characteristics, and E: Nut characteristics.

Running association analysis/mapping between molecular and morphological data is the best choice to find informative markers. The objective of association mapping in the present study is utilizing microsatellite markers (SSRs) in accompany with morphological traits to identify genetic markers that are assumed to as close as to the genes underlying agronomical traits of almond.

Results

Morphological traits

Cluster analysis. The almond genotypes clustered into two main groups: wild almonds and cultivated almonds. The cultivated almonds group divided by 4 subgroups included two, three, forty four and two genotypes, respectively. The wild almonds group contained genotypes which had the minimum average for most of the studied traits, but average of this group for traits such as ease of hulling (EH), shell color (SC), marking of outer shell (MS), leaves arrangement (LAr), leaf basal shape (LBS), kernel color (KC) and sensitivity to *Anarsia Lineatella* (the peach twig borer) was higher than the cultivated group.

Genotypes average in the first subgroup of cultivated almonds for some important traits such as kernel width (KW), thickness (KT) and weight (KWG), ease of hulling (EH), growth (tree) habit (GH), double flower in buds (DF), flower density (FD), bearing habit (BH), petiole length (PL), sensitivity to *Anarsia Lineatella* (peach twig borer), sensitivity to *Pterochloroides persica* (peach black aphid) and *Myzus persicae* (green peach aphid) was higher than the other subgroups. The second subgroup had the highest average for kernel length (KL), suture opening of the shell (SO), shriveling of kernel (SK), leaf length (LL), leaf width (LW), leaf area (LA) and leaf shape (LS), sensitivity to *Pseudomonas syringae* pv. syringae (Bacterial canker and blast of stone fruit trees) and duration of flowering (FDu).

The average of third subgroup for the most of studied traits was something in middle of the other subgroups but had the maximum value for nut length (NL), nut thickness (NT) and marking of outer shell (MS). The fourth subgroup showed the highest average for kernel width (KW), nut weight (NWG), Kernel/Nut W %, softness of shell (SS), kernel pubescence (KP), Initial flowering (F5) and completed flowering (F100). The third subgroup divided by 5 minor groups.

Minor group one had the highest average for kernel weight (KWG), suture opening (SO), shell retention (SR), ease of hulling (EH), leaf area (LA), leaf length (LL) and leaf arrangement (LAr), and the genotypes of second minor group showed the highest average for traits like nut width (NW) and thickness (NT), kernel thickness (KT), softness of shell (SS), shriveling of kernel (SK), Kernel Pubescence (KP), growth habit, flower density (FD), sensitivity to *Anarsia Lineatella* (peach twig borer) and *Myzus persicae* (green peach aphid). The highest average in third minor group referred to nut weight (NWG), bearing habit (BH), and sensitivity to *Pterochloroides persica* (peach black aphid) as well as the fourth minor group for nut length (NL), Kernel/Nut W %, ease of hulling (EH), sensitivity to *Pseudomonas syringae* pv. syringae (Bacterial canker and blast of stone fruit trees) and fifth minor group for percentage of double kernel (DK), double flower in buds (DF) and Initial flowering (F5), completed flowering (F100) and duration of flowering (FDu).

PCA

In Principal component analysis (PCA), the first eight variables accounted for 62.23% of the total variability. The first component contains traits corresponding to leaf and sensitivity to *Myzus persicae* (green peach aphid); the second, nut width (NW), length (NL), kernel weight (KWG) and nut weight (NWG); and the third, kernel related traits, initial flowering (F5) and completed flowering (F100).

Table 2. Morphological traits used for almond characterization.

Trait	Code	Trait	Code
Kernel		Tree	
Kernel Length	KL	Growth habit	GH
Kernel Width	KW	Flower	
Kernel Thickness	KT	Bearing Habit	BH
Kernel Weight	KWG	Flower Density	FDe
Kernel/Nut W %	KNWP	Double Flower in buds	DF
Kernel Color	KC	Flower Color	FC
Double Kernel	DK	Initial Flowering	F5
Shriveling of Kernel	SK	Completed Flowering	F100
Kernel Pubescence	KP	Duration of Flowering	FDu
Leaf		Nut	
Petiole Length	PL	Nut Length	NL
Leaf Width	LW	Nut Width	NW
Leaf Length	LL	Nut Thickness	NT
Leaf Area	LA	Nut Weight	NWG
Leaf Basal Shape	LBS	Nut Shape	NS
Leaf Shape	LS	Nut Color	SC
Leaves Arrangement	LAr	Marking of outer Shell	MS
Pests and Diseases		Suture Opening	SO
Bacterial Canker	SR	Shell Retention	SR
Black Aphid	ABR	Ease of Hulling	EH
Twig Borer	SKR	Softness of Shell	SS
Green Aphid	AR		

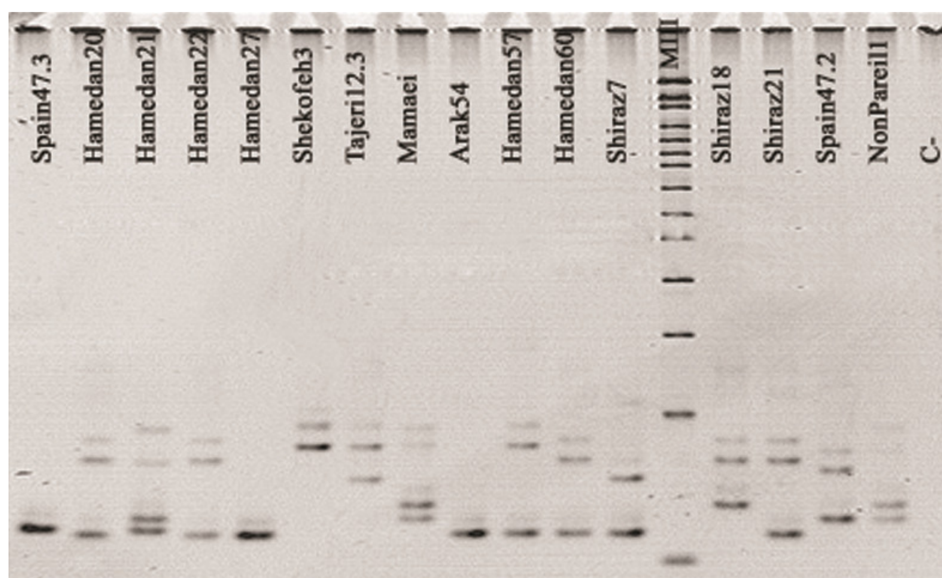

Figure 2. SSR alleles amplified in locus UDA-008 for some of the studied almond genotypes.

The fourth component stands for bearing habit (BH), nut thickness (NT) and shell retention (SR) in genotype characterization as well as in the fifth component which effect of suture opening (SO) and sensitivity to *Pterochloroides persica* (peach black aphid), in the sixth component flower density (FD), leaf basal shape and sensitivity to *Anarsia Lineatella* (peach twig borer) and in the seventh component shriveling of kernel (SK), shell retention (SR), kernel pubescence (KP), double flower in buds (DF) were highlighted, respectively.

Molecular markers

SSR polymorphism

Out of nine microsatellite primer pairs, 7 were polymorphic, which revealed sufficient alleles to characterize all genotypes. The number of alleles per locus varied from 8 (UDA022) to 17 (UDA002) with an average of 12.86 (Table 3).

Table 3. Almond SSR loci used in the present study and their characteristics.

SSR locus/ GeneBank no.	Sequence (5'-3')	Total No. of alleles	na[1]	Ne[2]	Ave_Het	PIC
UDA-002 BV102479	5'AAACGTGAGGTCTCACTCTCTC 5'-GCCATTTAAGGGTCTGGTCA	85	17	6.83	0.85	0.91
UDA-005 BV102480	5'CATCACACACAAACACAAATGC 5'-GCATTGTGCTCTTCATGGAC	119	15	4.73	0.79	0.93
UDA-008 BV102481	5'AGACGCTTTGCATACATACAAG 5'-TGCAGGAACTGGGATTAGAGA	135	13	6.56	0.85	0.93
UDA-009 BV102482	5'-AAAACATCTCTCTCCTCCATGC 5'-AGTTCTCTGGCAGCACAAGC	107	14	6.06	0.83	0.93
UDA-015 BV102474	5'-ACTCCATCGCTTGCATTTTC 5'-GCTCCGTGTGTGTTTGTGTG	81	9	2.48	0.60	0.69
UDA-022 BV102476	5'-GCCGTCTCATTTTCCCATTA 5'-GTGCGATGGAGGAGCACT	87	8	4.73	0.79	0.88
UDA-023 BV102477	5'-TTGCCGTGATACACTAACAACT 5'-ACCTGCCAAGTAAGTGCCTA	101	14	7.72	0.87	0.95
Mean		102	12.86	5.59	0.80	0.89
St. Dev			3.24	1.75	0.09	

[1] Number of alleles, [2] Number of effective alleles

The UDA-002 and UDA-005 markers were the most informative revealing 17 and 15 polymorphic alleles distributed among all genotypes, respectively, but considering PIC, UDA-023 was the most informative one (Table 3). According to the results, the average values for number of effective alleles (5.59), Shanon index (1.97), expected heterozygosity (0.8) and PIC (0.89, ranging from 0.69 to 0.95) were relatively high (Table 3). *Cluster analysis.* In the molecular dendrogram, the genotypes were divided into 2 major groups: group 1 included mainly 2 subgroups where a wild genotype *A.lycioides* and 4 genotypes comprised the first and second subgroups, respectively. Group 2 were divided into 4 subgroups which the genotypes with more Kernel/Nut W %, suture opening (SO), softness of shell (SS) and initial flowering (F5) and completed flowering (F100) comprised the first subgroup.

Specific markers

The specific alleles, rare alleles and allelic genotypes (alleles and genotypes with frequency lower than 1%) were recorded among the germplasm (data not shown).

Genotype identification

Using the ABRIISTAT software, input data (allele sizes) were processed to produce data format for different analysis softwares such as PowerMarker3.25 (Liu, 2005), PopGene (Yeh and Boyle, 1997), NTSys (Rohlf, 2000), and also different statistical parameters related to genetic diversity. Another feature of ABRIISTAT software was development of a molecular identification key for the studied loci. We suggested a coding method for genotype identification using microsatellite markers.

Identification of informative markers

The results of stepwise regression analysis revealed a significant correlation/association between the morphological traits and some of the studied microsatellite loci. One or more informative marker(s) were identified for almost all of the studied traits.

Alleles with positive correlation with the related traits have been shown in Table 4. A total of 141 positive markers (markers with positive correlation) out of 556 polymorphic bands (135 alleles obtained from 9 SSR loci) were identified for different traits. In statistical analysis whereas no informative marker was detected in significant level of 1%, the correlation analysis was carried out in significant level of 5%, so it provided informative markers but with less significant lelvel.

The most important informative markers in 1% of significant level are listed below. Markers 006-2, 023-7 and 022-9 accounted for 64% of total variability of nut weight. 005-2, 005-10 and 015-3 were responsible for kernel weight with 92%, 002-8 for SS (Softness of Shell) with 53% and 005-2 for DK (Double Kernel), with 28% of the total variability. 005-2 and 023-15 for KWG (Kernel Weight), 006-2 and 002-15 for NWG (Nut Weight), 006-7 and 002-9 for BH (Bearing Habit) and 023-16, 008-9 and 005-16 for initial flowering accounted for 64%, 51%, 55% and 86 % of the total variability, respectively. Marker 023-16 for completed flowering and duration of flowering described 25% of the total variability. For DF (Double Flower in buds) and FC (Flower Color) characteristics, 20 and 22 alleles which accounted for the total variability were identified respectively. For some of the studied traits informative markers were detected in 5% of the significant level (Table 4).

For traits corresponding to nut and kernel size, loci UDA005 and UDA023 each one with 4 markers and UDA006, UDA008 and UDA022 each one with 2 markers showed a relation for expression of these traits more than the other loci. Concerning shell related traits, locus UDA002 revealed the highest correlation and in the case of nut characteristics and yield components, locus UDA005 had the most association.

Loci UDA008, UDA009, UDA002 with 5, 4 and 4 alleles respectively, showed a significant correlation with DF (double flower in buds). The locus UDA009 with 4 markers revealed an association with FC (flower color). Finally, about sensitivity to pests and diseases, locus UDA008 with 4 markers had the most correlation in this regard (Table 4).

Complete characterization of informative markers including accounting for the total variability, positive or negative correlation and their significance effect on a specific trait has been shown in Table 4.

Discussion

High diversity level in almond is due to gametophytic self-incompatibility nature of the species. On the other hand, the building up and generation of F2 populations for fruit trees particularly for almond- which is a cross pollinated species- is tedious and time-consuming task. Therefore, in the present investigation the association studies were carried out between different morphological traits (including phenological and agronomically important characteristics) and SSRs as highly polymorphic molecular markers to find possible informative markers. Identified informative markers can be served as main tools in marker-assisted selection program for some morphological and/or agronomical traits in almond. The findings showed positive correlations in some cases.

In cluster analysis according to the morphological traits, the first and fourth subgroups (Najafabad-Monagha, Najafabad-Dobahreh, Nonpariel1 and ShirazPoostnazok1) had the highest average for the most important yield related traits, so these subgroups contain the best genotypes regarding to the yield characteristics. Although considering geographical origin, the studied genotypes were from different regions of Iran (Hamadan, Shiraz, Najafabad and foreign countries), and comparing to molecular based cluster, the morphological dendrogram could not fit them correspondingly. This result agrees with previous studies (Kadkhodaei, 2011; Sorkheh, 2007) which can be mainly due to the different natures of the molecular and morphological markers, as molecular markers cover genome more extensively as they also contain non-coding regions, and on the other hand unlike morphological markers they have not exposed to artificial selection.

Principal component analysis (PCA) was performed to reduce the number of effective traits in group characterization. In this analysis, combination of the traits double kernel percentage (DK), duration of flowering (FDu) and sensitivity to *Pseudomonas syringae* pv. syringae (Bacterial canker and blast of stone fruit trees) in eighth component showed an important role in characterization of the almond genotypes.

In molecular analysis, a total of 9 SSRs primer pairs, previously isolated from almond (Testolin, 2004; Table 3), were used to amplify specific SSR loci from DNA of each almond genotype. Among these SSRs primers, UDA-006 and UDA-010 amplified monomorphic fragments and was withdrawn from further analysis. The SSRs markers used in this study had PIC values $\geq$ 0.7, the value which has been reported in previous studies to determine effectiveness of RAPD, SSR and AFLP markers for genetic discrimination of almond germplasms (Kadkhodaei, 2006; Sorkheh, 2007; Fathi, 2008; Kadkhodaei, 2010). The high percentage of polymorphic SSRs loci (89 %) detected in this study was consistent with previous studies (Testolin, 2004; Fathi, 2008; Kadkhodaei, 2010 and 2011). The number of alleles per locus varied from 8 (UDA022) to 17 (UDA002) with an average of 12.86 which is a little more than those same studies on other *Prunus* species (Aranzana, 2002; Decroocq, 2003) and also on almonds which were studied using peach SSRs (Fathi, 2008).

The relatively high values of different heterozygosity features in the studied population can be due to this fact that in this study we used almond specific SSRs rather than previous similar studies on almond (Fathi, 2008; Shiran, 2007). Also, to lesser extent-it could be due to utilizing a germplasm with less diversity among the genotypes (Mnejja, 2005; Xu, 2004; Testolin, 2004).

Of course, self-incompatibility behavior of almond species is the main reason for the higher polymorphism in almond rather than other *Prunus* species. Increasing the amount of Ne (and not necessarily the total number of alleles per locus) leaded to an increase in He and also indirectly in ability of the loci for separation of genotypes through increasing the number of allelic genotypes.

The UDA-002, UDA-005 and UDA-023 markers which were the most informative markers according to the PIC and number of polymorphic alleles obtained can be considered for detecting almond genotypes and efficiently for further studies among almond populations and germplasms. All of the genotypes (except duplicates) could be distinguished using the studied SSRs loci, indicating their high level of polymorphism and performance in this regards. Also, we were able to detect some index genotypes using their specific SSRs profiles, which can be lead to rapid and direct detection of them using a low-cost and time-consuming method (data not shown). On the other hand, the specific markers (rare alleles) may be used efficiently in DNA fingerprinting and also breeding programs through marker assisted selection (MAS).

The suggested coding method in our study which includes two parts (the first for locus characteristics or Gene Bank No. and second for allele(s) size, for example UDA002-125/115) can be used easily as an international code to name different genotypes using molecular markers.

Regarding to the informative markers identified in the present study in terms of being positively correlated to morphological and horticulturally important traits, they can be considered as candidate markers for the related morphological (particularly agronomical) traits. For some of the morphological traits more than one informative marker was detected, which consequently finding their additive effects, degree of dominance and sum of the positive and negative effects need further analysis.

Marker assisted selection (MAS) is a very helpful strategy for increasing the effectiveness of selection gains. Knowledge provided by advances in molecular genetics promise faster and more efficient approaches to cultivar improvement. Early selection utilizing molecular markers allows accurate screening of seedlings several years before the mature plant traits can be evaluated in the field, makes possible the accumulation of different genes/QTLs for horticultural traits of interest, and shortens the number of generations to recover the desired genotype particularly after a cross with an exotic genotype or wild species (Morohoshi and Komamine, 2001).

Markers associated to main agronomic traits in almond have obtained through QTL mapping and Bulked Segregant Analysis (BSA).

The QTL mapped traits are flower color (Jáuregui, 1998), nematode resistance (Jáuregui, 1998; Bliss, 2002), shell hardness (Arús, 1999), anther color (Joobeur, 1998), blooming time (Ballester, 2001), kernel taste (Joobeur, 1998; Bliss, 2002), and self-incompatibility (Ballester, 1998, 2001; Arús, 1999; Bliss, 2002), and using BSA method some

Table 4. Detected informative markers for different morphological traits of almond, in 1 and 5% of significance level.

Significance level: 1 %

Trait	Marker	B (Regression Coefficient)	P value	Adjusted R2
NL	023-7	-0.75	0	0.537
NW	006-2	0.5	0.008	0.644
	023-7	-0.51	0.001	
	022-9	-0.4	0	
NT	006-5	-0.65	0	0.781
	005-9	0.59	0	
	008-19	-0.29	0.01	
KL	022-18	-0.6	0	0.746
	023-15	-0.54	0	
	023-16	-0.33	0.008	
KW	005-2	0.88	0	0.923
	005-10	-0.44	0	
	015-3	0.25	0.002	
KT	008-6	0.7	0	0.623
	005-16	-0.38	0.009	
MS	002-15	0.55	0.01	0.270
SO	009-17	1	.	1.000
	005-11	0.61	.	
	002-4	-0.61	.	
	005-8	0	.	
SS	002-8	0.75	0	0.536
SK	002-15	-0.54	0.002	0.509
	005-12	0.45	0.007	
KP	005-6	0.57	0.005	0.291
DK	005-2	0.56	0.007	0.280
KC	023-17	0.57	0.001	0.517
	006-2	-0.44	0.008	
KWG	005-2	0.9	0	0.646
	023-15	-0.62	0	
NWG	006-2	0.9	0	0.515
	002-15	-0.52	0.008	
KNWP	005-13	0.65	0	0.567
	005-2	0.48	0.003	
LA	022-17	0.54	0.009	0.252
PL	006-5	-0.74	0	0.530
LW	006-5	-0.63	0.001	0.372
LL	023-7	-0.6	0	0.557
	002-8	0.45	0.005	
LA	005-2	-0.61	.	1.000
	008-6	-0.61	.	
	009-9	-0.84	.	
	008-4	0.61	.	
	022-18	0	.	
	005-13	0	.	
DF	023-13	1.17	.	1.000
	002-7	-0.58	.	
	002-8	0.40	.	
	008-4	-0.29	.	
	008-8	0.00	.	
	023-16	0.00	.	
	008-7	0.00	.	
	009-12	0.00	.	
	006-6	0.00	.	
	009-9	0.00	.	
	008-9	0.00	.	
	009-13	0.00	.	
	006-2	0.00	.	
	002-10	0.00	.	
	006-1	0.00	.	
	015-3	0.00	.	
	022-14	0.00	.	
	009-8	0.00	.	
	002-15	0.00	.	
	015-8	0.00	.	
	008-19	0.00	.	
	005-14	0.00	.	
FC	005-13	-0.743	.	1.000
	005-11	-0.538	.	
	008-6	-0.538	.	
	002-7	0.00	.	
	022-18	0.00	.	
	023-13	0.00	.	
	015-10	0.00	.	
	022-14	0.00	.	
	009-10	0.00	.	
	009-13	0.00	.	
	015-3	0.00	.	
	023-19	0.00	.	
	005-10	0.00	.	
	022-17	0.00	.	
	015-8	0.00	.	
	008-11	0.00	.	
	009-12	0.00	.	
	009-17	0.00	.	
	005-6	0.00	.	
	023-11	0.00	.	
BH	006-7	0.68	0.000	0.549
	002-9	-0.50	0.003	
SR	008-11	0.77	0.000	1.000
	002-4	0.46	0.000	
	008-6	0.46	0.000	
	008-10	0.46	0.000	
ABR	009-11	0.69	0.000	0.456
AR	008-19	0.54	0.007	0.262
F5	023-16	0.37	0.001	0.863
	008-9	0.56	0.000	
	005-16	0.49	0.000	
F100	023-16	0.73	0.000	0.504
FD2	023-16	0.54	0.010	0.253

Significance level: 5 %

Trait	Marker	B (Regression Coefficient)	P value	Adjusted R2
EH	022-14	-0.45	0.03	0.160
SC	002-10	-0.48	0.01	0.370
	005-10	0.42	0.02	
SR	022-14	0.45	0.03	0.170
FD	015-10	-0.84	0	1.000
	006-1	-0.69	0	
	008-6	-0.37	0	
	002-9	0.52	0	
	005-8	0.75	0	
	002-4	-0.37	0	
	006-8	-0.52	0	
	006-2	0.37	0	
LS	002-7	0.91	0	1.000
	005-2	0.46	0	
GH	002-14	-0.83	0	1.000
	002-11	-0.36	0	
	008-13	-0.45	0	
	015-10	0.43	0	
	002-8	-0.27	0	
	023-11	0.41	0	
	008-12	0.27	0	
	009-9	-0.27	0	
	005-11	0.29	0	
	008-18	-0.1	0	
	015-8	-0.16	0	
FD1	015-10	-0.84	0	1.000
	006-1	-0.69	0	
	008-6	-0.37	0	
	002-9	0.52	0	
	005-8	0.75	0	
	002-4	-0.37	0	
	006-8	-0.52	0	
	006-2	0.37	0	
SKR	022-18	-0.64	0	1.000
	002-7	-0.93	0	
	023-15	-0.86	0	
	005-11	-0.46	0	
	005-2	0.46	0	
	005-13	0.64	0	

specific fruit characters in peach×almond crosses (Warburton, 1996), self-incompatibility and a gene conferring delayed blooming (Ballester, 2001; Kole, 2007) were mapped.

Also there are some other molecular techniques including comparative mapping (Dirlewanger, 2004) and candidate gene approaches (Le Dantec, 2010; Horn, 2005; Silva, 2005) which can be considered in this regard.

A very promising application of MAS in almond is the manipulation of self-incompatibility. Amplification of specific S-alleles using appropriately designed primers is being routinely used for the identification of cross incompatibility groupings for current almond cultivars and for efficiently breeding self-compatibility into new cultivars allowing earlier and more accurate selection of the most common self-incompatibility or self-compatibility alleles (López, 2004). Using the same methodology we developed a multiplex PCR strategy for identification of the most important Iranian almond cultivars (Kadkhodaei, 2008b). The developed molecular key efficiently enabled discrimination of the cultivars at the least time and most precise method.

Association mapping, also known as “association genetics,” “association studies,” and “linkage disequilibrium mapping”– although the latter term is also used to reflect studies detecting associations among loci. The general characteristics of association studies involve the use of unstructured or loosely structured populations (unlike QTL mapping) that are both phenotypically and genotypically characterized to detect statistical associations between genetic polymorphisms and heritable trait variation. Polymorphisms chosen for screening could come from whole genome scans, selectively chosen but phenotypically neutral sequences, or preselected candidate genes. Association genetics shares much in common with the field of what is commonly known as QTL mapping. Both attempt – via statistical inference – to detect co-segregation of polymorphic genetic markers with genes underpinning trait variation. The reasons of utilizing association genetics are: (1) the higher resolution afforded by use of unstructured populations allows the intriguing possibility of identifying the genes – or even the specific nucleotides underpinning trait variation, (2) the opportunity to use molecular markers to enhance rates of genetic gain, including the utilization of specific genes from non-elite germplasm in a more directed and efficient manner than was hitherto possible. The key advantages of association tests include their speed, because mapping populations may not be necessary, particularly in crops that are limited to no more than one generation per year. The other advantage is high resolution. In general, association genetics approaches may be more suited to organisms with little or no pedigree information; large effective population sizes resulting in less differentiation in trait values and little or no structure in the population (same as the studied population); populations with rich allelic diversity (including almond with 29 S-alleles, for instance), moderate to high nucleotide diversity; and traits with little or no selection history and controlled by many loci with small effects, and low frequency older alleles. On the other hand, linkage based fine mapping methods may be more efficient for marker assisted breeding in inbred crops than in some out-breeding perennial species.

Despite the recent advances in the application of the newer biotechnologies, almond, as well as other tree crops, lags behind the progress typically observed for annual crops. This is, in large part, the consequence of the inherent difficulties in doing genetic studies on such large-sized and long generation-time plants. However, these inherent obstacles to traditional breeding make the opportunities with the new technologies much more revolutionary when applied to tree crops. Almond is currently well positioned to be a leader in this effort (Gradziel, 2009).

Regarding to selection of the best germplasm for application of interest, the knowledge and information about syntenic relationships between different species would help the researchers in this case. For example, in the highly syntenic genus *Prunus* where the species with high (peach) and low (apricot, almond, cherry) Linkage disequilibrium conservation exists, peach collections maybe adequate for whole genome scans to find gene or QTL positions and apricot, almond or cherry populations maybe more helpful for the validation of candidate genes (Dirlewanger, 2009).

Materials and methods

Plant material

Fifty-three originally diversed almond genotypes distributed throughout Iran in accompany with some foreign cultivars and their hybrids (Iran-Froeign) were used in this study. The genotypes maintained at Shahid-Fozveh experimental orchard located at the Agriculture and Natural Resources Research Center in Isfahan, Iran (Table 1).

Morphological traits

A total of 39 (Seventeen quantitative and 22 qualitative) traits were recorded during 2006 and 2007 for 53 almond genotypes/cultivars (Table 2, Fig. 1) according to almond descriptor (Gulcan, 1985).

Molecular characterization

DNA extraction

Genomic DNA extractions were performed as previously described (Kadkhodaei, 2005). The purified total DNA was quantified by gel electrophoresis, and its quality verified by spectrophotometry. DNA samples were stored at 4°C.

Microsatellite analysis. Extracted almond genomic DNA was PCR-amplified using 9 pair flanking SSRs sequences previously cloned and sequenced specifically in almond (Testolin, 2004). Details of the microsatellites analyzed and their origin are given in Table 3. As a parameter for the selection of some of these microsatellites, we have used the genetic informativeness and the technical aspects of the analysis. The microsatellites used in this study, have not been extensively studied for almonds in earlier studies, and almost all of the previous studies on almonds have been carried out using peach isolated SSRs.

Amplification was conducted in a total volume of 15 µl with 10 ng DNA, 0.05 µM of both specific primers, 0.2mM of 10 mM dNTP mix, 0.1 U Taq polymerase, 1.5 mM $MgCl_2$, and 1×PCR buffer (100 mM Tris HCl, pH 8.3, 500 mM KCl). The amplification program consisted of 3 min at 94 °C, 34 cycles of 45 s at 94 °C, 45 s at the appropriate annealing temperature and 45 s at 72 °C, followed by a 5-min extension at 72 °C. Five micro-liters of the PCR product were separated on a 1% agarose gel to check the PCR amplification and determine approximated size of the amplified fragments.

Then electrophoresis of PCR products was carried out on a 6-10 % polyacrylamid gel (depended on allele ranges in each locus) in 0.5 X TBE.

The gels were run for 3 h at constant power (120 W) in ATTO Electrophoresis system. A 50-bp DNA Ladder (Fermentas) was used as a molecular size standard in each gel (Fig. 2).

Data Analysis

Morphological data

The Gower general similarity coefficient (Gower, 1971) was used in cluster analysis of morphological traits. Unweighted Pair-Group Method using Arithmetic average (UPGMA) analysis was performed with the program MVSP (version 3.13b, Kovach Computing Services, Anglesy, Wales, UK). The software statistical analysis system (SAS, 1998 V.8) was used for Principal Components Analysis (PCA) to define Eigenvalues and Eigenvectors and also for comparison of the mean of groups to define effective traits in separation of the groups.

Molecular Data

Allele size was measured with UVDoc 99.02 analysis software (UVI Tech, Cambridge, UK) by manual editing to increase accuracy. This procedure carried out in order of two times to exclude wrong scorings. In order to analysis of the large amount of data, ABRIISTAT30 software developed which facilitated manipulation of huge amount of data (Kadkhodaei, 2008c). PowerMarker3.25 software package (Liu, 2005) was used to generate a similarity matrix based on Nei coefficient (Nei, 1983) and to produce a dendrogram using UPGMA method. Different genetic diversity parameters calculated as follow: *Heterozygosity*- (the proportion of heterozygous individuals in the population) at a single locus, *Gene diversity* - often referred to as expected heterozygosity-is defined as the probability that two randomly chosen alleles from the population are different, and *Polymorphism information content* (PIC) (Botstein, 1980).

Genotype identification

A molecular identification key was developed using ABRIISTAT software (Kadkhodaei, 2008c) for all of the genotype.

The correlation analysis between polymorphic SSRs markers and morphological traits

After separated analysis of molecular and morphological datasets, the correlation/association between them was calculated (Table 4). A total of 135 markers, obtained from 9 specific loci, were used in molecular analysis. For association mapping between molecular markers and morphological traits and for identification of possible informative markers, pearson correlation analysis and a stepwise regression analysis were done respectively. In both analysis, levels of significance 0.05> and 0.01> was employed using SPSS v.16 software (SPSS Inc. Headquarters, 233 S. Wacker Drive, Chicago, Illinois 60606). Morphological traits (quantitative characteristic) and molecular data (SSRs alleles scored as ‘1’ for presence and ‘0’ for absence) were treated as dependent and independent variables in the stepwise regression analysis, respectively. UPGMA dendrogram for both sets of data (molecular and morphological) was constructed separately. To ensure the usefulness of tested SSRs loci, different polymorphism related parameters such as average heterozygosity, PIC value, average effective number of alleles, etc. were calculated in a primary study, and the results indicated the suitability of them for later analysis.

Conclusion

A major goal of this research consists of developing faster and better methods to link genotypic information to desirable phenotypic ones. Selection by means of molecular markers is particularly useful in fruit, nut, and other tree crops with a long juvenile period, when the expression of the gene is recessive or the evaluation of the character is difficult. So, association studies attract an especially high interest in fruit and nut tree genetic studies, because they do not require the creation of large segregating populations, that takes a long time and is expensive. In summary, the results of this investigation suggest that molecular marker-based estimation of gene effects controlling traits of interest can provide a useful and precise means for the establishment of appropriate breeding methods and preliminary information for accurate QTL mapping experiments.

Acknowledgments

This research was supported by the Agricultural Biotechnology Research Institute of Iran (ABRII). Isfahan Research Center for Agriculture and Natural Resources (H. A. Daneshvar, A. Nasr, H. Samavatian, B. Vafania, E. Aroon) kindly provided samples and information about cultivars indicated in this study.

References

Aranzana MJ, Garcia-Mas J, Carbo J and Arus P (2002) Development and variability analysis of microsatellite markers in peach. Plant Breed 121: 87–92.

Arús P, Ballester J, Jáuregui B, Joobeur T, Truco MJ, deVicente MC (1999) The European Prunus mapping project: update of marker development in almond. Acta Hort 484:331–336.

Ballester J, Boskovic R, Batlle I, Arús P, Vargas F, deVicente MC (1998) Location of the self-incompatibility gene on the almond linkage map. Plant Breed 117:69–72.

Ballester J, Sociasi Company R, Arús P, de Vicente MC (2001) Genetic mapping of a major gene delaying blooming time in almond. Plant Breed 120: 268–270.

Bliss FA, Arulsekar S, Foolad MR, Becerra AM, Gillen A, Warburton ML, Dandekar AM, Kocsisne GM, Mydin KK (2002) An expanded genetic linkage map of Prunus based on an interspecific cross between almond and peach. Genome 45:520–529.

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32: 314-331.

Decroocq V, Favé MG, Hagen L, Bordenave L, Decroocq S (2003) Development and transferability of apricot and grape EST microsatellite markers across taxa. Theor Appl Genet 106: 912-922.

Dickson EE, Arumuganathan K, Kresovich S, Doyle JJ (1992) Nuclear DNA content variation within the Rosaceae. Am J Bot 79:1081–1086.

Dirlewanger E, Graziano E, Joobeur T, Garriga-Caldere F, Cosson P, Howad W, Arus P (2004b) Comparative mapping and marker assisted selection in Rosaceae fruit crops. Proc Natl Acad Sci 101:9891–9896.

Dirlewanger E, Denoyes-Rothan B, Yamamoto T, Chagn D (2009) Genomics tools across Rosaceae species. In: K.M. Folta, S.E. Gardiner (Eds.), Genetics and Genomics of Rosaceae, Plant Genetics and Genomics: Crops and Models 6. Springer Science.

Fathi A, Ghareyazi B, Haghnazari A, Ghaffari MR, Pirseyedi M, Kadkhodaei S, Naghavi MR, Mardi M (2008) Assessment of the genetic diversity of almond (*Prunus dulcis*) using microsatellite markers and morphological traits. IJB 6(2):98-106.

Gower JC (1971) A general coefficient of similarity and some of its properties. Biometrics 27:857-874.

Gradziel TM (2009) Almond *(Prunus dulcis)* Breeding. In: S.M. Jain, P.M. Priyadarshan (Eds.), Breeding Plantation Tree Crops: Temperate Species. Springer Science+Business Media, LLC.

Gulcan R (1985) Descriptor list for almond (*Prunus amygdalus*) revised. IBPGR Secretariat, Rome.

Gupta PK, Balyan HS, Sharma PC, Ramesh B (1996) Microsatellites in plants: a new class of molecular markers. Curr Sci India 70: 45-54.

Horn R, Lecouls AC, Callahan A, Dandekar A, Garay L, Mc Cord P, Howad W, Cha H, Verde I, Main D, Jung S, Georgi L, Forrest S, Mook J, Zhebentyayeva T, Yu Y, Kim HR, Jesudurai C, Sosinski B, Arus P, Baird V, Parfitt D, Reighard G, Scorza R, Tomkins J, Wing R, Abbott AG (2005) Candidate gene database and transcript map for peach, a model species for fruit trees. Theor Appl Genet 110:1419–1428.

Jáuregui B (1998) Localización de marcadores moleculares ligados a caracteres agronómicos en un cruzamiento interespecífico almendro × melocotonero. Tesis Doctoral. Universitat de Barcelona, Barcelona, Spain.

Joobeur T, Viruel MA, deVicente MC, Jáuregui B, Ballester J, Dettori MT, Verde I, Truco MJ, Messeguer R, Batlle I, Quarta R, Dirlewanger E, Arús P (1998) Construction of a saturated linkage map for Prunus using an almond×Peach F2 progeny. Theor Appl Genet 97: 1034–1041.

Kadkhodaei S, Shahnazari M, Nekouei MK, Ghasemi M, Etminani H, Imani A, Ariff AB (2011) A comparative study of morphological and molecular diversity analysis among cultivated almonds (*Prunus dulcis*). Aust J Crop Sci 5(1): 82-91.

Kadkhodaei S, Elahy M, Nekouei M K, Imani A, Shahnazari M, Mardi M, Javanmard A, Ariff AB (2010) A panel of cultivate specific marker based on polymorphisms at microsatellite markers for Iranian cultivated almonds (*Prunus dulcis*). Aust J Crop Sci 4(9): 730-736.

Kadkhodaei S, Arabha A, Ebrahimi M (2008) Karyotype analysis of some Iranian wild almonds. 10th Iranian Genetics Congress. Tehran, Iran.

Kadkhodaei S, Khayyam M, Elahi M (2008c) ABRIISTAT 30: molecular data analysis software for genetic diversity studies. 18th International EUCARPIA General Congress. Valencia, Spain.

Kadkhodaei S, Tabaei-Aghdaei R (2005) A simple and inexpensive method for nucleic acid isolation from plants with high polysaccharides and polyphenolic compounds: development of DNA extraction procedure in case of *Amygdalus* spp. IV International Symposium on Pistachios and Almonds. Tehran, Iran.

Kadkhodaei S, Khayam Nekouei M, Elahi M (2008b) Development of a molecular identification key for the most important Iranian almond cultivars using DNA fingerprinting methods. 18th International EUCARPIA General Congress. Modern Variety Breeding for. Present and Future Needs. Valencia, Spain.

Kadkhodaei S, Aghdaei SRT, Grigorian V, Moghadam M, Hashemi SMM (2006) A study on genetic variation among some wild almond species using RAPD markers; in Javanshah, A., E. Facelli, M. Wirthensohn (eds): Acta Hort 726:93-98.

Khatamsaz M (1992) Rosaceae. In: Flora of Iran. Research Institute of Forests and Rangelands, pp.1–352.

Kole C (2007) Genome Mapping and Molecular Breeding in Plants, Volume4. Fruits and Nuts. Springer-Verlag Berlin Heidelberg.

Le Dantec L, Cardinet G, Bonet J, Fouché M, Boudehri K (2010) Development and mapping of peach candidate genes involved in fruit quality and their transferability and potential use in other Rosaceae species. Tree Genet Genomes, 6(6):995-1012.

Liu K and Muse S (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics, 21(9):2128-2129.

López M, Mnejja M, Rovira M, Collins G, Vargas FJ, Arús P and Batlle I (2004) Self-incompatibility genotypes in almond re-evaluated by PCR, stylar ribonucleases, sequencing analysis and controlled pollinations. Theor Appl Genet 109(5):954-964.

Messina R, Lain O, Marrazzo MT, Huang WG, Cipriani G, Testolin R (2004) Isolation of microsatellites from almond and apricot genomic libraries and testing for their transportability across different *Prunus* species. XIth Eucarpia Symp on Fruit Breed & Genetics. Acta Hort 663.

Mnejja M, Garcia-Mas J, Howad W, Arús P (2005) Development and transportability across *Prunus* species of 42 polymorphic almond microsatellites. Mol Ecol Notes 5(3):531-535.

Morohoshi N, Komamine A (2001) Molecular breeding of woody plants; Proc Intl Wood Biotech Symp (IWBS), Progress in Biotechnology 18, ELSEVIER.

Nei M, Takezaki N (1983) Estimation of genetic distances and phylogenetic trees from DNA anlysis. Proc 5th World Cong Genet Appl Livstock Prod 21:405-412.

Powell W, Machray GC, Provan J (1996) Polymorphism revealed by simple sequence repeats. Trends Plant Sci 1:215-222.

Rohlf FJ (2000) NTSYS-pc: Numerical Taxonomy and Multivariate Analysis System. Version 2.1. Exceter Sofware, New York, USA.

Schueler S, Tusch A, Schuster M, Ziegenhagen B (2003) Characterization of microsatellites in wild and sweet cherry–markers for individual identification and reproductive processes. Genome 46:95-102.

Shiran B, Amirbakhtiar N, Kiani S, Mohammadi SH, Sayed-Tabatabaei BT, Moradi H (2007) Molecular characterization and genetic relationship among almond cultivars assessed by RAPD and SSR markers. Sci Hortic 111:280-292.

Silva C, Garcia-Mas J, Sánchez AM, Arús P, Oliveira MM (2005) Looking into flowering time in almond (*Prunus dulcis* (Mill) D. A. Webb): the candidate gene approach. Theor Appl Genet 110(5):959-68.

Socias I Company R (1998) Almond genetics at a turning point: the almond example. Theor Appl Genet 96:588-601.

Sorkheh K, Shiran B, Gradziel TM, Epperson B, Martínez-Gómez P, Asadi E (2007) Amplified fragment length polymorphism as a tool for molecular characterization of almond germplasm: genetic diversity among cultivated genotypes and related wild species of almond, and its relationships with agronomic traits. Euphytica 156(3):327-344.

Testolin R, Messina R, Lain O, Marrazzo MT, Huang WG, Cipriani G (2004) Microsatellites isolated in almond from an AC-repeat enriched library. Mol Ecol Notes 4:459-461.

Warburton ML, Becerra-Velásquez VL, Goffreda JC, Bliss FA (1996) Utility of RAPD markers in identifying genetic linkages to genes of economic interest in peach. Theor Appl Genet 93:920-925.

Xu Y, Ma RC, Xie H, Liu JT, Cao MQ (2004) Development of SSR markers for the phylogenetic analysis of almond trees from China and the Mediterranean region. Genome 47:1091-1104.

Yeh FC, Boyle TJB (1997) Popgene Version 1.32. Microsoft Windows-based software for population genetic analysis. pp. 135.