

Genome-wide Association Links Candidate Genes to Fruit Firmness, Fruit Flesh Color, Flowering Time, and Soluble Solid Content in Apricot (*Prunus Armeniaca* L.)

Filiz Ferik

Ege University: Ege Universitesi

Duygu Ates

Ege University: Ege Universitesi

Sezai Ercisli

Ataturk University: Ataturk Universitesi

Abdullah Erdogan

Malatya Apricot Research Institute

Emine Orhan

Ataturk University: Ataturk Universitesi

Bahattin Tanyolac (✉ bahattin.tanyolac@gmail.com)

Ege Universitesi <https://orcid.org/0000-0002-4368-0988>

Research Article

Keywords: association mapping, apricot, fruit firmness, fruit flesh color, flowering time, soluble solid content

Posted Date: September 3rd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-854114/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background Apricots originated from China, Central Asia and the Near East and arrived in Anatolia, and particularly in their second homeland of Malatya province in Turkey. Apricots are outstanding summer fruits, with their beautiful attractive color, delicious sweet taste, aroma and high vitamin and mineral content.

Methods and results In the current study, a total of 259 apricots genotypes from different geographical origins in Turkey were used. Significant variations were detected in fruit firmness (FF), fruit flesh color (FFC), flowering time (FT), and soluble solid content (SSC). A total of 11,532 SNPs based on DArT were developed and used in the analyses of population structure and association mapping (AM). According to the STRUCTURE (v.2.2) analysis, the apricot genotypes were divided into three groups. The mixed linear model with Q and K matrixes were used to detect the associations between the SNPs and four traits. A total of 131 SNPs were associated with FF, FFC, and SSC. No SNP marker was detected associated with FT.

Conclusion The results demonstrated that AM had high potential of revealing the markers associated with economically important traits in apricot. The SNPs identified in the study can be used in future breeding programs for marker-assisted selection in apricot.

Introduction

Rosaceae is one of the most important fruit tree family from temperate regions, including apple, peach, strawberry, plum, almond, pear, European plum, and sweet cherry which are economically important fruit species [1]. Apricot, *Prunus armeniaca* (Lam.), is also a member of this family and an important stone fruit with a total world production of about 3.8 million tons [2]. The major apricot producing countries are Turkey (730.000 tons), Uzbekistan (662.123 tons), and Iran (306.115 tons) in the world [2].

Various parameters affect fruit quality in apricot. Consumer preferences are based on fruit quality which refers to sensorial properties, such as appearance, texture, taste and aroma, high nutritional components, chemical components, functional properties, and mechanical characteristics [3]. The appearance of fruit is the main criterion for consumers, while a low level of sweetness and hard texture are undesired [3]. Thus, fruit firmness and fruit flesh color are important for consumer satisfaction [3]. Another important parameter affecting fruit quality is soluble solid content (SSC), which includes sugars, organic acids, proteins, minerals, lipids, amino acids, and vitamins, and it is the main criteria that determines the taste, flavor and nutritional value of the fruit [4]. In addition to being a delicious edible product, the fruit of apricot is also considered as functional due to its chemical ingredients [4]. The apricot fruit makes a significant contribution to human health through its phenolic compounds content with immune-stimulating, anti-inflammatory and antioxidants properties [4]. One of the important traits for apricot producers is flowering time, which is a substantial agronomic trait with an impact on fruit and seed growth in temperate fruit tree species. In cold regions, early flowering individuals are damaged by late

frost, while in warm regions, late flowering individuals could lead to some problems concerning leaf and flower bud break, resulting in a decrease in the amount of harvest [5]. In addition, early flowering species have economic value in terms of early market prices.

Complex traits including most of the fruit quality properties are controlled by interacting genes called quantitative traits [1]. Therefore, the clarification of transmission of complex traits is one of the main topic for agricultural sciences [6]. Quantitative trait locus (QTL) mapping is a prevalent method for the mapping of these kinds of traits and bases on biparental mapping populations [6]. Constructing a new cross population is a tedious, time consuming and expensive process. Considering the long generation and juvenile period of fruit trees, it is more difficult to apply QTL mapping [1]. To date, a certain number of QTL maps have been constructed for apricot, such as flowering time [7–8], resistance to sharka disease [9–12], fruit quality traits [13–14], tree architectural traits [14] and chilling requirements [15]. Association mapping (AM) is an alternative method to pedigree-based QTL mapping and uses natural populations, contrary to QTL mapping, to determine the correlations between phenotypes and genotypes [16]. AM also utilizes historical recombination and natural variation as a basis and provides high map resolution in shorter time due to no requirement of developing a new cross population [16]. This method has been previously employed for different fruit trees, such as peach [17–18], apple [19], almond [16], and apricot [15, 20].

The objective of this study was to investigate the associations between SNPs based on the diversity arrays technology (DArT) and the pomological traits of apricot, namely fruit firmness (FF), fruit flesh color (FFC), flowering time (FT), and SSC using 259 apricot genotypes.

Materials And Methods

Plant Material and DNA Isolation

A total of 259 apricots (*Prunus armeniaca* L.) genotypes, which were grown together at Malatya Apricot Research Institute in Turkey, were used in this study (Online resource 1). All genotypes had been planted with eight-meter spacing between and within the rows in the experimental station of Malatya Apricot Research Institute. Each genotype was planted in three replications, and all trees were 20 years old. Standard management practices concerning chemical fertilization, pruning, and disease control were being applied to the trees.

Young leaves were collected from each apricot genotype, cooled in liquid nitrogen, and stored at -80°C for future analyses. The leaf samples were ground into small pieces with a tissue lyser (Technogen Co., Turkey). DNA extraction was carried out with 0.1 g samples of each individual following the protocol described by Deshmukh omitted RNase application step [21]. Tris-EDTA (TE) buffer (100 µl) was used to dissolve the extracted DNA. For the purification and quantification assessment of the isolated DNA, 1% agarose gel and a spectrophotometer (NanoDrop ND 1000) were used, respectively. After confirmation, the DNA samples were stored at -80°C until they were used for SNP analyses.

Pomological Evaluation

Forty fruits from each replication of genotype were randomly selected and harvested separately for each tree. Fruits were harvested at maturity stage based on appearance and taste of each genotype according to Gecer et al. (2020) [22]. For FFC measurements; Color intensity (Chroma) of the flesh (after peeling flesh) were measured with a Minolta chromometer (CR-400, Minolta). Tristimulus color analyzer calibrated to a white porcelain reference plate. The chroma $(a^*2 + b^*2)^{1/2}$ were determined by using a and b color values around the equatorial region in three different positions with an average of nine times for each apricot fruits and chroma were used for evaluation. For FT, observations were made by experts, and the first day of flowering was noted as FT. The juice of the 40 apricots was measured with a digital refractometer (Model RA-250HE Kyoto Electronics, Kyoto, Japan), and the SSC values were recorded in °Brix. FF was measured with an acoustic firmness sensor (Aweta BV, the Netherlands). These fruit traits were measured for two consecutive years (2016 and 2017).

Variance Analysis

To define the variations in FF, FFC, FT and SSC among the 259 genotypes over two years (2016 and 2017), an analysis of variance (ANOVA) was performed using TOTEMSTAT software [23] according to the significance level of $P \leq 0.01$. The analysis were applied with complete randomized blocked design (CRBD) with two factors (year and genotype) and three replications. The variations were determined according to year (Y), genotype (G), and Y x G interactions.

DArT Analysis

SNP analyses were performed at DArT PLT (Diversity Arrays Technology Pty. Ltd., Canberra, Australia). The *Pst*-*Mse*I method was preferred from four reducing complexity methods that were tested (data not presented). Digestion/ligation reaction were applied to DNA samples according to Kilian et al. [24], except one single *Pst*-compatible adaptor, two various adaptors were used to create two various restriction enzyme (RE) overhangs. The *Pst*-compatible adaptor with a sequencing primer sequence and staggered, the Illumina flowcell attachment sequence and a barcode region with different length was designed as analogue sequence with Elshire et al. [25]'s. The second adaptor includes the *Mse*I-compatible overhang sequence and flowcell attachment sequence like the first adaptor. In order to amplified *Pst*-*Mse*I fragments following PCR reaction conditions were used; 30 rounds 94°C for 1 min, followed by 29 cycles of 94°C for 20 sec, ramp 2.4°/sec to 58°C, 58°C for 30 sec, ramp 2.4°C/sec to 72°C, 72°C for 45 sec. then 72°C for 7 min. Following the PCR step, c-Bot (Illumina) bridge PCR was performed to equal amounts of PCR products of each and followed by sequencing on Illumina HiSeq2500. PCR products were sequenced in a single lane with 77 cycles. DArT analytical pipelines were used to process all generated sequences. FASTQ files were performed in the primary pipeline in order to filter the poor quality results. The barcode regions were subjected to more reliable selection criteria (\geq Phred pass score of 30). For marker calling

step, 2,000,000 sequences per barcode/sample were utilized. Lastly, high quality sequences were used to create the fastqcall files and these files fastqcall were applied to secondary pipeline for Presence/Absence Markers (PAM) calling algorithms and DArT P/L's proprietary SNP (DArTsoft14).

The polymorphism information content (PIC) values represent the discrimination power of the markers. The PIC values were calculated for each marker according to Lynch and Ritland (1999) with the following equation: $PIC=1-\sum p_i^2$, where p_i demonstrates the proportion of the population with the i^{th} allele [26]. A dendrogram tree was drawn with the R software with reference to Nei's genetic distance [27].

Genetic Variation Analysis

STRUCTURE software (v.2.3.4), which is based on Bayesian modelling, was used to determine the population structure of 259 apricot genotypes [28]. The software was run with a burn-in period of 10,000 and 100,000 Markov Chain Monte Carlo (MCMC) replications. Ten runs were performed for each number of populations (K), ranging from 1 to 10. The best number of subpopulations was determined with the Delta K (ΔK) value using STRUCTURE HARVESTER [29].

Association Mapping Analysis

TASSEL (v.5.2.3) software with a mixed linear model [MLM (K+Q) model] was used for the detection of the associations between DNA markers and pomological traits (FF, FFC, FT, and SSC) [29]. The relative kinship matrix, which shows the genetic relationships between the individuals, was calculated by TASSEL (v.5.0) based on the centered IBS method [30]. The Q matrix was obtained from STRUCTURE software at the $\Delta K = 3$ value. The associations between the SNP markers and pomological traits were visualized as Manhattan plots in R software with the "qqman" package. The designation of significant markers was performed in the same software with the false discovery rate (FDR) [31] and Bonferroni correction [32] being calculated separately for each pomological trait (FF, FFC, FT and SSC). Furthermore, the quantile-quantile (Q-Q) plots were visualized with the same software.

Identification of candidate genes

The sequences of the SNP markers associated with FF, FFC, and SSC were analyzed to determine the functions of the candidate genes using the Phytome (v12.1) database.

Results

Phenotypic Variation

In the present study, FF, FFC, FT, and SSC were measured for two years (2016-2017), and showed normal distribution (these distributions are based on average by accessions) (Online resource 2). These findings

indicated the importance of the genetic background of each genotype for the Prunus phenotyping profile. The minimum, maximum and mean values of each year showed high consistency, and no significant differences were observed between the years (2016 and 2017) for the mean values. The minimum, maximum and mean values of all phenotypic traits are presented in Table 1.

Table 1

Minimum, maximum and mean values of all phenotypic traits

Trait	2016			2017		
	Min	Max	Mean	Min	Max	Mean
FF	0.1 N	9.60 N	1.86 N	0.03N	6.62 N	1.71 N
FFC	11.6°	46.72°	34.04°	13.99°	46.77°	34.12
FT	95 days	125 days	114.24 days	95 days	126 days	112 days
SSC	8.22°Brix	29.92°Brix	15.28°Brix	8.23°Brix	32.37°Brix	16.27°Brix

The mean values of four traits (FF, FFC, FT and SSC) only slightly differed between the two years (2016 and 2017). However, there were fourfold differences between the SSC and FFC values obtained from 2016 and 2017 (Table 1). FT ranged from 95 to 125 days with a mean value of 114.2 days in 2016, and it ranged from 95 to 126 days with a mean value of 112 days in 2017 (Table 1). FF varied between 0.1 N and 9.60 N in 2016 and 0.03 N and 6.62 N in 2017 (Table 1). There was a nearly 90-fold difference in the ranges obtained for FF from the two years. The individuals showing the highest and lowest average values are listed in online resource 3.

The results of the correlation analysis showed no significant correlation between the four traits (FF, FFC, FT and SSC) (Online resource 4). The results of ANOVA are presented in online resource 5. ANOVA demonstrated significant variations according to year, genotype, and year X genotype interactions for all apricot traits at the $P \leq 0.01$ significance level (Online resource 5).

Population Structure Analysis

A total of 24,864 SNP markers were generated from the DArT analysis (Online resource 6.), and after filtering the missing data [max 5% missing data, Marker Allele Frequency (MAF>0.5)], 11,532 high-quality SNP markers were obtained (Online resource 7.). Apricot genome was comprised of scaffolds and the detected SNPs located on scaffolds rather than chromosomes. The PIC value was 0.77, ranging between 0.05 and 0.99. These markers were assigned to the related scaffolds and were used in the STRUCTURE (v.2.2) analysis. This analysis was performed for K from 1 to 10, and the peak was observed at K = 3 according to the ΔK computation data. The STRUCTURE results showed that the 259 genotypes were divided into three main populations: namely POPI (red), POPII (green) and POPIII (blue) (Online resource 8).

All these genotypes were also further divided into three groups according to Nei's genetic distance analysis (Online resource 9). The first group consisted of Geno 185 (Nigde - Turkey) and Geno 186 (Malatya -Turkey), the second comprised Geno 38 (Siverek/Urfa – Turkey), Geno 230 (USA) and Geno 255 (Russia), and the third contained the remaining 254 genotypes. These results indicate that the genotypes used in this study were not clustered according to their geographical origin.

The expected heterozygosity and fixation index (F_{st}) are parameters that explain the heterozygosity level of a population. In this study, the expected heterozygosity was determined as 0.20 for Cluster 1, 0.06 for Cluster 2, and 0.11 for Cluster 3, with a mean value of 0.12. On the other hand, the F_{st} value varied between 0.14 and 0.81 with a mean value of 0.55, representing a high genetic variation level for the population.

AM Analysis

AM analyses were carried out for four pomological traits (FF, FFC, FT and SSC) using TASSEL (v.5.2.3) software and the MLM (Q+K) model in two consecutive years (2016 and 2017). These analyses detected a large number of associations related to the pomological traits. FDR and Bonferroni corrections were applied to eliminate the false positives among the associations. Eventually, 131 SNP markers were found to be associated with three traits (FF, FFC, and SSC). Among these associations, three, 57 and 71 SNPs were associated with FFC, FF and SSC respectively.

A total of 88 and 228 SNPs were associated with FF in 2016 and 2017, respectively (FDR correction applied, $-\log_{10} P \geq 2.90$ for 2016 and ≥ 2.78 for 2017), and 57 of these markers were common for both 2016 and 2017 (Online resource 10 and 11 and Fig. 1). Most of the significant SNPs for FF were detected in 2017 but not in 2016 (Online resource 10 and 11 and Fig. 1).

For FFC, three SNPs ($-\log_{10} P$ value is ≥ 3.27 , FDR correction applied) and 13 SNPs ($-\log_{10} P \geq 3.10$, FDR correction applied) were associated with FFC in 2016 and 2017, respectively, and three of these SNPs (SNP 4257, SNP 17194 and SNP 22875) was common for both years (Online resource 10 and Fig. 2).

The marker-trait association analysis for FT revealed that it was associated with 10 SNPs (FDR correction applied, $-\log_{10} P \geq 3.28$) in 2016 and 22 SNPs (FDR correction applied, $-\log_{10} P \geq 3.06$) in 2017. However, none of these SNPs was commonly seen in both years (Online resource 10 and Fig. 3).

For SSC, 167 SNPs ($-\log_{10} P \geq 2.82$, FDR correction applied) and 352 SNPs ($-\log_{10} P \geq 2.72$, FDR correction applied) were found related in 2016 and 2017, respectively. Of these SNPs, 71 were detected in both years (Fig. 4 and online resource 10 and 12). The P values presenting the significance level of the associations between the markers and pomological traits are given in Q-Q plots in online resource 13.

Identification of Candidate Genes

A total number of 30 putative candidate genes were found to be related to the SNPs associated with FF and SSC (Online resource 14). SNPs which were associated with the FFC trait did not show similarity to

any of the putative candidate genes. For the SSC and FF of the apricots, the following proteins and enzymes related to putative candidate genes showed homology with SNPs (given in parentheses): putative 3,4-dihydroxy-2-butanone kinase (SNP526), putative leucine-rich repeat receptor-like protein kinase (SNP1023), pentatricopeptide repeat-containing protein (SNP1482), probable LRR receptor-like serine/threonine-protein kinase (SNP1494), vignain-like (SNP2823), transcription termination factor MTERF6 (SNP3309), LRR receptor-like serine/threonine-protein kinase GSO2 (SNP3842), WAT1-related protein At5g64700 (SNP4443 and SNP20435), cell division cycle 20.2, cofactor of APC complex-like (SNP4948), long-chain acyl-CoA synthetase8 (SNP5158), calcium-transporting ATPase 12 plasma membrane-type-like (SNP5444), probable LRR receptor-like serine/threonine-protein kinase At4g36180 (SNP13398), peroxidase 5 (SNP14227), UDP-glycosyltransferase TURAN (SNP15803), tRNA (guanine(37)-N1)-methyltransferase 1 (SNP16184), transcription factor TFIIIB component B" homolog (SNP16422), receptor like protein 30-like (SNP17431), putative disease resistance protein (SNP19126), putative pre-16S rRNA nuclease (SNP19568), DNA-directed RNA polymerase III subunit (SNP 20753), DEAD-box ATP-dependent RNA helicase 3 (SNP22678), putative pentatricopeptide repeat-containing protein (SNP23251), S locus F-box protein f (SLFf) gene, partial cds; and Sf-RNase and S haplotype-specific F-box protein f (SFBf) genes, complete cds (SNP23325), putative disease resistance RPP13-like protein 1 (SNP23732), TMV resistance protein N-like (SNP23833), putative disease resistance protein (SNP23953), LRR receptor-like serine/threonine-protein kinase GSO2 (SNP24483), and receptor-like protein 12 (SNP24611) (Online resource 14).

Discussion

Phenotypic Variation

Fruit quality parameters are of prime importance for both consumers and growers. Among these parameters, FT is the most widely studied physical attribute in apricot due to its economic importance related to early market prices [14]. In the present study, FT was detected between 95 and 126 days (Table 1). Among the genotypes studied, Geno115 was the earliest cultivar with 95 days. These genotypes can be used in further breeding studies for developing early and late genotypes. In previous studies, FT was reported to range from 59 to 84 days [8] and 55 to 78 days [14]. The variation that was detected in the present study for FT was greater compared to the literature and can be attributed to the different genetic origins of the cultivars.

SSC is one of the main criteria affecting fruit taste, and SSC value greater than 12°Brix indicates good gustative quality [3]. In the present study, SSC was measured between 8.22 and 32.36°Brix, most genotypes (243 genotypes) had a value over 12°Brix, and there was nearly a fourfold difference in the range (Table 1). In different studies, SSC was detected as 8.73 to 17.80°Brix [13], 10.6 to 16.2°Brix [10] and 6.2 to 19.5°Brix [14]. In our study, a larger variation was found in SSC compared to previous studies. This wide range of phenotypic discrepancy indicates the genotypic variation level of the apricot genotypes used in the study.

FFC and FF are two most important sensorial properties that affect consumer preferences at the purchase step. In the present study, FFC was measured between 11.6° and 46.77°, and there was nearly a fourfold difference in the range (Table 1). In previous studies, FFC was measured from 70° to 94.7° [10] and 67° to 100° [13]. In the current study, the range of FF was obtained as 0.03 to 9.60 N. In the literature, FF was reported to vary between 24.9 and 62.2 N [3] and 15 and 50 N [13]. Both FFC and FF ranges measured in the current study were quite different from those of previous studies.

Population Structure

In genetic mapping studies, associations between DNA markers and traits are affected by the type and number of markers [33]. The use of a high number of markers leads to high genome coverage, and therefore high-throughput systems gain importance. In the current study, DArT, a high-throughput system, was used and a total of 20,264 SNP markers were developed and after filtering 11,532 high quality SNP were used in further analyses. Forcada et al. [18] developed 8,144 SNPs on 94 peach genotypes. Comparing with this study, our sequencing results provide high genome coverage.

The STRUCTURE analysis was used for the identification of the population structure of the 259 apricots investigated in the current study. These genotypes were divided into three main populations (Online resource 4). Structure (Online resource 4) and dendrogram analysis (Online resource 9) show us these genotypes were not divided into populations according to their geographical origin. For example, the genotypes that originated from Turkey, Spain, Italy, Poland, Armenia, France, USA, and Hungary were included in the same (third) group (Online resource 9). The reason for this result could be the complex breeding history of these genotypes. In particular, the use of cultivars with different histories in introgression and intercrossing processes may have led to this situation [33]. In addition, humans move plants from one geographic realm to another, which results in confusion concerning the origin of the plants [33]. Similar to results from the current study, Li et al. [34] also found five different group in their study. Despite the fact that fewer accessions were used in their study compared to the current study, they generated more markers. They pointed out the reason for the high number of SNPs to be genetically varied wild genotypes [34]. In a previous AM study on apricot, 72 genotypes were used, and the genotypes were divided into two main groups [20]. The reason for the lower number of subpopulations in the current study could be the use of a population with a narrow genetic basis. Although the previous authors also selected the genotypes from different countries, they may have used those of the same origin.

In the present study, the mean F_{st} value was detected as 0.55 and the mean expected heterozygosity was 0.12, indicating the presence of a high genetic variation in the population structure. These findings also support the idea that DArT systems develop a large number of SNP markers distributed along the apricot genome. In previous studies, the F_{st} value was reported to be 0.51 [35] and 0.16 [36], and the expected heterozygosity value was 0.82 [49] and 0.29 [50]. These differences between previous studies in terms of

diversity may be due to the number and type of markers utilized, and genotyping being undertaken in distinct locations [37].

Association Mapping Analysis

AM is a powerful technique which is based on the accumulation of genetic variability through evolution in natural populations to identify DNA markers based on the association between genetic markers and phenotypes [52]. AM analyses are used as a rapid and efficient alternative to linkage mapping analyses [16] for detecting the associations between traits and markers; therefore, these analyses are widely employed in the mapping of economically important traits in many crop species [18]. In the present study, K and Q matrixes were used to correct the population structure in the MLM (Q+K) model included in AM analyses. This model effectively eliminates possible false positives with random and fixed effects according to Henderson's notation [39]. In addition, FDR and Bonferroni corrections were applied to eliminate spurious associations [39]. To date, no AM study has been undertaken to reveal the associations between SNP markers and pomological traits (FF, FFC, FT and SSC). However, an association map was constructed by Mariette et al. [20] to identify the SNP markers that controlled resistance to plum pox virus in apricot. Furthermore, Olukolu [15] constructed an AM on the chilling requirements of apricot. Apart from these studies, there are only a limited number of association studies on the economically important traits of the other members of the Rosaceae family [16-18, 40]. In the present study, a total of 131 SNPs were found significantly associated with three pomological traits (FF, FFC and SSC) of the apricot genotypes via AM analyses (Online resource 10-12 and Fig. 1-4). Three SNPs (SNP 4257, SNP17194 and SNP 22875) were found associated with FFC and 57 SNPs were found associated with FF (Online resource 10-12 and Fig. 1-4). A total of 71 SNPs were associated with SSC in two consecutive years (2016 and 2017) (Online resource 12). In previous studies, Salazar et al. (2013) used control cross population of apricot and found QTLs in one linkage group for SSC and two linkage groups for FFC [13]. Socquet-Juglard et al. [14] also used a control cross population and found one linkage group was related to SSC. Campoy et al. [7] found one linkage group for FF. The number of significant markers identified in the present study was higher than previously reported [13-14] which may be related to the type of population investigated. Also, these previous studies used SSR markers, because of that it's difficult to compare these results. In mapping studies, natural populations provide higher genome coverage and mapping resolution with regard to wide genotypical variations [39]. Another reason for our high number of SNPs may be the use of DArT to produce the markers. This technology is known to produce a high number of SNP markers, and thus provide high genome coverage.

Identification of Candidate Genes

In the present study, 30 putative candidate genes showed homology with the sequences of SNPs associated with FF and SSC (Online resource 14). Among these, transcription termination factor MTERF6 plays an important role in plastid development on *Arabidopsis thaliana* [41]. WAT1-related protein is

located on cell wall and responsible for transmembrane transporter activity [42]. Long-chain acyl-CoA synthetase 8 is very important for lipid metabolism [43]. UDP-glycosyltransferase TURAN is one of the responsible enzyme for development of pollen tube. LRR receptor-like serine/threonine-protein kinase (GSO₂) and probable LRR receptor-like serine/threonine-protein kinase (GSO) together play a role in root growth and the growth of epidermal surface in embryos and cotyledons in *Arabidopsis thaliana* [44]. Putative pentatricopeptide repeat-containing protein is a member of the pentatricopeptide repeat (PPR) protein family and is involved in the organellar RNA metabolism [45]. Calcium-transporting ATPase 12, plasma membrane-type-like has a function in calcium-transporting ATPase activity and calmodulin binding [46]. tRNA (guanine(37)-N1)-methyltransferase 1 is responsible for the methylation of cytoplasmic and mitochondrial tRNAs in the N1 position of guanosine-37 *Arabidopsis thaliana* [47]. TMV resistance protein N-like provides resistance to tobacco mosaic virus in plants [48]. Putative 3,4-dihydroxy-2-butanone kinase plays a role in ATP binding [49]. Pentatricopeptide repeat-containing protein is a required protein for the intergenic processing between chloroplast *rsp7* and *ndhB* transcripts [50].

Conclusions

This is the first AM study that presents the associations between SNPs based on the DArT technology and economically important traits (FF, FFC, FT and SSC) in apricot. Large variations were determined for these four traits. The results of this study highlight the importance of using populations with wide variations in AM studies. AM revealed significant associations for FF, FFC and SSC. The SNPs identified in the study can be used in future breeding programs for marker-assisted selection in apricot. On the other hand, the genotypes with the shortest FT can be used as parents in developing earlier cultivars combined with other desirable traits.

Declarations

Funding

This research was funded by the Scientific and Technological Research Council of Turkey (TUBITAK) with the project number TUBITAK-2150445.

Conflicts of interest/Competing interests

The authors declare no competing interests.

Availability of data and material

not applicable

Code availability

not applicable

Authors' contributions

FF: Association map analysis, writing manuscript; DA: writing manuscript, data obtaining; SE: PI of the projects; AE: obtained phenotyping data, EO: statistical analysis, data obtaining; MBT: corresponding author, conceived and designed research.

Ethics approval

not applicable

Consent to participate

not applicable

Consent for publication

not applicable

References

1. Meneses C, Orellana A (2013) Using genomics to improve fruit quality. *Biol Res* 46(4):347–352
2. FAOSTAT [Online] Available: <http://www.fao.org/faostat/en/#data/QC>. [Accessed: 13-Apr-2021]
3. Ruiz D, Egea J (2008) Phenotypic diversity and relationships of fruit quality traits in apricot (*Prunus armeniaca* L.) germplasm. *Euphytica* 163(1):143–158
4. Hegedus A, Engel R, Abranko L et al (2010) Antioxidant and antiradical capacities in apricot (*Prunus armeniaca* L.) fruits: variations from genotypes, years, and analytical methods. *J Food Sci* 75(9)
5. Fan S, Bielenberg DG, Zhebentyayeva TN et al (2010) Mapping quantitative trait loci associated with chilling requirement, heat requirement and bloom date in peach (*Prunus persica*). *New Phytol* 185(4):917–930
6. Wurschum T (2012) Mapping QTL for agronomic traits in breeding populations. *Theor Appl Genet* 125(2):201–210
7. Campoy JA, Ruiz D, Egea J et al (2011) Inheritance of flowering time in apricot (*Prunus armeniaca* L.) and analysis of linked quantitative trait loci (QTLs) using simple sequence repeat (SSR) markers. *Plant Mol Biol Report* 29(2):404–410

8. Dirlewanger E, Quero-Garcia J, Heredity LLD et al (2012) Comparison of the genetic determinism of two key phenological traits, flowering and maturity dates, in three *Prunus* species: peach, apricot and sweet cherry. *Heredity* 109:280–292
9. Lambert P, Dicenta F, Rubio M, Audergon JM (2007) QTL analysis of resistance to sharka disease in the apricot (*Prunus armeniaca* L.) 'Polonais' x 'Stark Early Orange' F₁ progeny. *Tree Genet Genomes* 3(4):299–309
10. Soriano JM, Vera-Ruis EM, Martinez-Calvo J et al (2008) Identification and mapping of a locus conferring plum pox virus resistance in two apricot-improved linkage maps. *Tree Genet Genomes* 4(3):391–402
11. Marandel G, Salava J, Abbott A et al (2009) Quantitative trait loci meta-analysis of plum pox virus resistance in apricot (*Prunus armeniaca* L.): New insights on the organization and the identification of genomic resistance factors. *Mol Plant Pathol* 10(3):347–360
12. Dondini L, Lain O, Vendramin V (2011) Identification of QTL for resistance to plum pox virus strains M and D in Lito and Harcot apricot cultivars. *Mol Breed* 27(3):289–299
13. Salazar JA, Ruiz D, Egea J, Martínez-Gómez P (2013) Transmission of fruit quality traits in apricot (*Prunus armeniaca* L.) and analysis of linked quantitative trait loci (QTLs) using simple sequence repeat (SSR) markers. *Plant Mol Biol Report* 31(6):1506–1517
14. Socquet-Juglard D, Christen D, Devènes G et al (2013) Mapping architectural, phenological, and fruit quality QTLs in apricot. *Plant Mol Biol Report* 31(2):387–397
15. Olukolu B (2010) The genetics of chilling requirement in apricot (*Prunus armeniaca* L.). All Dissertations 537
16. Font i Forcada C, Velasco L, Socias i Company R, Fernández i Martí Á (2015) Association mapping for kernel phytosterol content in almond. *Front Plant Sci* 6:530
17. Cao K, Wang L, Zhu G et al (2012) Genetic diversity, linkage disequilibrium, and association mapping analyses of peach (*Prunus persica*) landraces in China. *Tree Genet Genomes* 8(5):975–990
18. Font I, Forcada C, Guajardo V, Chin-Wo SR, Moreno M (2019) Association mapping analysis for fruit quality traits in *Prunus persica* using SNP markers. *Front Plant Sci* 9:2005
19. Cevik V, Ryder CD, Popovich A et al (2010) A FRUITFULL-like gene is associated with genetic variation for fruit flesh firmness in apple (*Malus domestica* Borkh.). *Tree Genet Genomes* 6(2):271–279
20. Mariette S, Tai FWJ, Roch G et al (2016) Genome-wide association links candidate genes to resistance to plum pox virus in apricot (*Prunus armeniaca*). *New Phytol* 209(2):773–784
21. Deshmukh V, Thakare P, Chaudhari US et al (2007) A simple method for isolation of genomic DNA from fresh and dry leaves of *Terminalia arjuna* (Roxb.) Wight and Arnot. *Molecular Biology and Genetics* 10(3)
22. Gecer MK, Kan T, Gundogdu M et al (2020) Physicochemical characteristics of wild and cultivated apricots (*Prunus armeniaca* L.) from Aras valley in Turkey. *Genet Resour Crop Evol* 67:935–945
23. Acikgoz N, Ilker E, Gokcol A (2004) Computer evaluations of biological research. Ege University Press

24. Kilian A, Wenzl P, Huttner E et al (2012) Diversity arrays technology: A generic genome profiling technology on open platforms. *Methods Mol Biol* 888:67–89
25. Elshire RJ, Glaubitz JC, Sun Q et al (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6(5):19379
26. Lynch M, Ritland K (1999) Estimation of pairwise relatedness with molecular markers. *Genetics* 152(4):1753–1766
27. Nei M (1972) Genetic distance between populations. *Am Nat* 106(949):283–292
28. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
29. Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol Ecol* 14(8):2611–2620
30. Bradbury PJ, Zhang Z, Kroon DE et al (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23(19):2633–2635
31. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B* 57(1):289–300
32. Bonferroni CE (1935) Il calcolo delle assicurazioni su gruppi di teste. In *Studi in Onore del Professore Salvatore Ortu Carbon* 13–60
33. Mather D, Hayes P, Chalmers K et al (2004) Use of SSR marker data to study linkage disequilibrium and population structure in *Hordeum vulgare*: Prospects for association mapping in barley. *Czech Journal of Genetics and Plant Breeding 9th International Barley Genetics Symposium, 20–26 June, 2004 / Ing. Jaroslav Spunar and Jarmila Janikova (eds.):302–307*
34. Li W, Liu L, Wang Y et al (2020) Genetic diversity, population structure, and relationships of apricot (*Prunus*) based on restriction site-associated DNA sequencing. *Horticulture Research* 7:69
35. Khan MA, Maghuly F, Borroto-Fernandez EG et al (2008) Genetic diversity and population structure of apricot (*Prunus armeniaca* L.) from Northern Pakistan using simple sequence repeats. *Silvae Genetica* 57(1):6
36. Tian-Ming H, Xue-Sen C, Zheng X et al (2007) Using SSR markers to determine the population genetic structure of wild apricot (*Prunus armeniaca* L.) in the Ily Valley of West China. *Genet Resour Crop Evol* 54(3):563–572
37. Ates D, Ascioğul TK, Nemli S et al (2018) Association mapping of days to flowering in common bean (*Phaseolus vulgaris* L.) revealed by DArT markers. *Mol Breed* 38(9):1–14
38. Ozkuru E, Ates D, Nemli S et al (2019) Genome-wide association studies of molybdenum and selenium concentrations in *C. arietinum* and *C. reticulatum* seeds. *Mol Breed* 39:46
39. Cappa EP, El-Kassaby YA, Garcia MN et al (2013) Impacts of population structure and analytical models in genome-wide association studies of complex traits in forest trees: A case study in *Eucalyptus globulus*. *PLoS One* 8(11):e81267

40. Dhanapal AP, Crisosto CH (2013) Association genetics of chilling injury susceptibility in peach (*Prunus persica* (L.) Batsch) across multiple years. *3 Biotech* 3(6):481–490
41. Romani I, Manavski N, Morosettiet A et al (2010) A Member of the Arabidopsis mitochondrial transcription termination factor family is required for maturation of chloroplast transfer RNA ile (GAU). *Plant Physiol* 169(1):627–646
42. Kovi M, Amdahl H, Alsheikh M et al (2017) De novo and reference transcriptome assembly of transcripts expressed during flowering provide insight into seed setting in tetraploid red clover. *Sci Rep* 7:44383
43. Zhao L, Haslam T, Sonntag A et al (2019) Functional overlap of long-chain acyl-coa synthetases in Arabidopsis. *Plant Cell Physiol* 60(5):1041–1054
44. Racolta A, Bryan AC, Tax FE (2014) The receptor-like kinases GS01 and GS02 together regulate root growth in arabidopsis through control of cell division and cell fate specification. *Dev Dyn* 243(2):257–278
45. Williams PM, Barkan A (2003) A chloroplast-localized PPR protein required for plastid ribosome accumulation. *Plant J* 36(5):675–686
46. Song P, Chen X, Wu B (2016) Identification for soybean host factors interacting with P3N-PIPO protein of Soybean mosaic virus. *Acta Physiol Plant* 38(6)
47. Lee C, Kramer G, Graham DE et al (2007) Yeast mitochondrial initiator tRNA is methylated at guanosine 37 by the Trm5-encoded tRNA (guanine-N1-)-methyltransferase. *J Biol Chem* 282(38):27744–27753
48. Hehl R, Faurie E, Hesselbach J et al (1999) TMV resistance gene N homologues are linked to *Synchytrium endobioticum* resistance in potato. *Theor Appl Genet* 98:379–386
49. Herz S, Kis K, Bacher A et al (2002) A tomato enzyme catalyzing the phosphorylation of 3, 4-dihydroxy-2-butanone. *Phytochemistry* 60(1):3–11
50. Barkan A, Small I (2014) Pentatricopeptide Repeat Proteins in Plants. *Annu Rev Plant Biol* 65(1):415–442

Figures

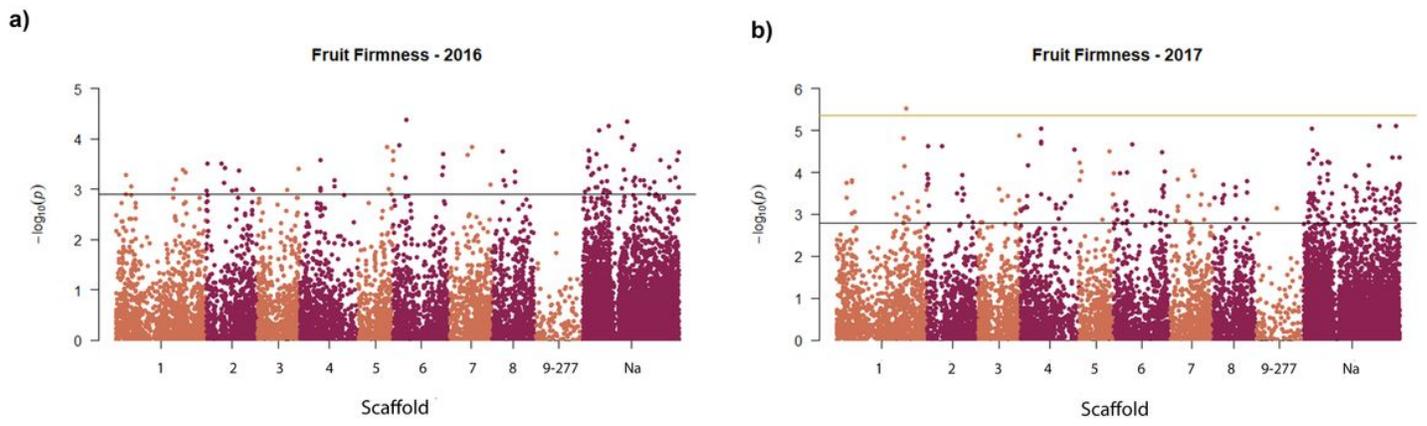


Figure 1

Manhattan plots of fruit firmness for the two years (2016 – 2017)

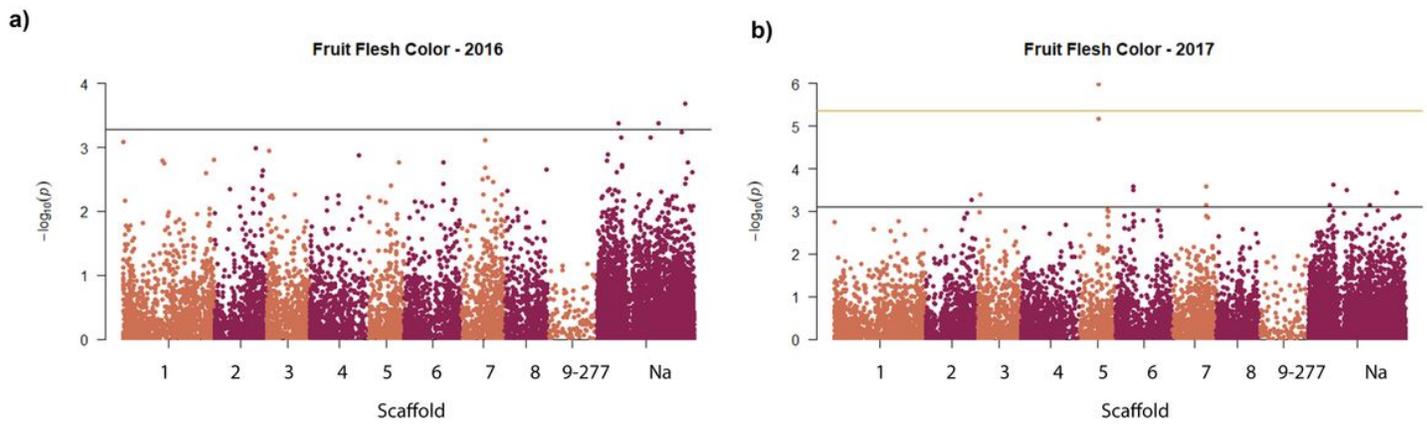


Figure 2

Manhattan plots of fruit flesh color for the two years (2016 – 2017)

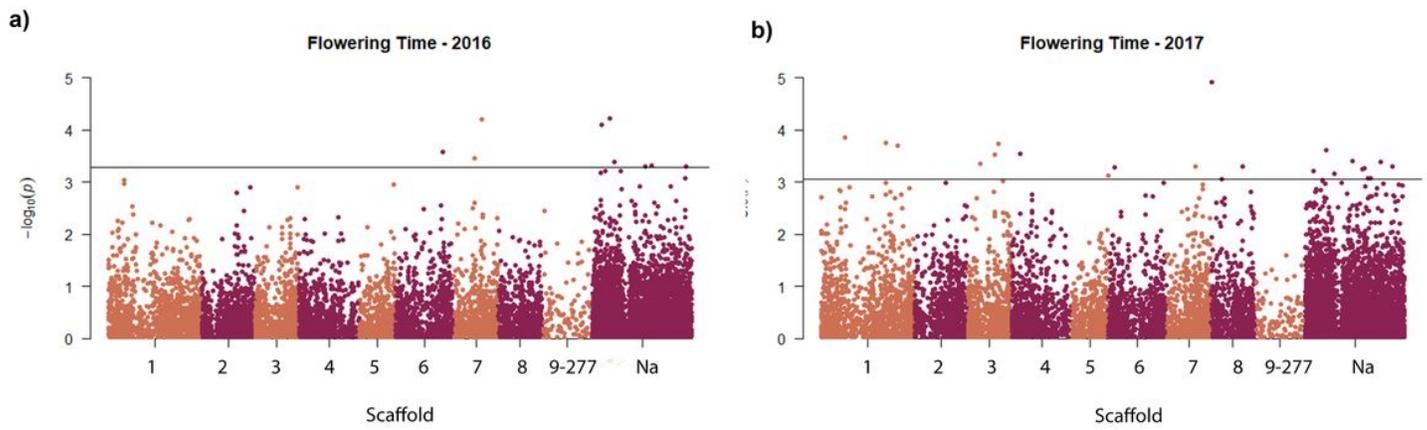


Figure 3

Manhattan plots of flowering time for the two years (2016 – 2017)

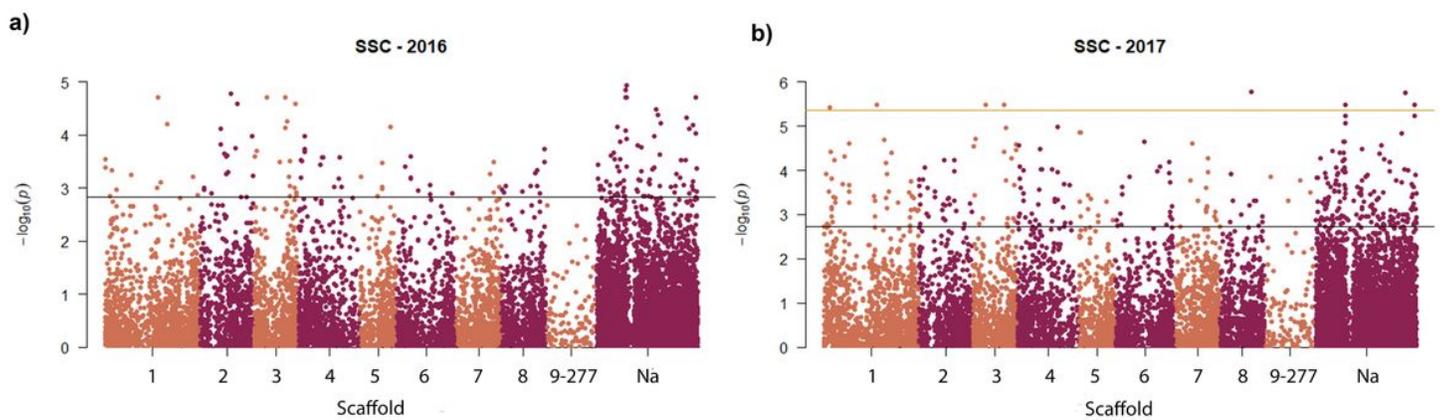


Figure 4

Manhattan plots of solid soluble content for the two years (2016 – 2017)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [ESM1.docx](#)
- [ESM2.jpg](#)
- [ESM3.docx](#)
- [ESM4.docx](#)
- [ESM5.docx](#)

- ESM6.xlsx
- ESM7.xlsx
- ESM8.jpg
- ESM9.jpg
- ESM10.docx
- ESM11.docx
- ESM12.docx
- ESM13.jpg
- ESM14.docx