Preliminary studies of genom-wide association mapping for some selected morphological characters of watermelons

Metin Yagcioglu^a, Osman Gulsen^{a,*}, Halit Yetisir^a, Ilknur Solmaz^b, Nebahat Sari^b

^a Department of Horticulture, Erciyes University, Melikgazi 38039, Kayseri, Turkey

^b Department of Horticulture, Çukurova University, Balcali 01330, Adana, Turkey

*Corresponding author: ogulsen@erciyes.edu.tr (O. Gulsen).

Phone: +90 352 207 6666 (ext. 38550)

Fax: +90 352 437 6902

ABSTRACT

Developing more efficient tools for watermelon (*Citrullus lanatus* (Thunb.) Matsum. & Nakai) breeding programs may greatly contribute to genetic improvement. The aims of this study were to investigate genome-wide association analysis (GWAS), and to develop regression models for some important watermelon traits such as yield, fruit weight, flesh color, total soluble solid, seed coat color, seed length and flowering time using 96 lines selected from 258 pure lines. An initial analysis was conducted to reduce sample size to 96 in order to avoid identical, highly similar and distant relatives causing low level of polymorphism and rare alleles. These 96 lines were morphologically and genetically characterized for several morphological characters under the eastern Mediterranean conditions, Adana, Turkey and PCR-based molecular markers. Associated markers were estimated by TASSEL program based on general linear model and a regression model because

each trait was established through backward regression procedure nested in the SAS program. Levels of variation explained by regression models were 41% with 6 markers for flesh color, 22% with 2 markers for fruit weight, 48% with 5 markers for total soluble solids, 21% with 6 markers for seed size, 9% with 4 markers for fruit flesh firmness, 39% with 6 markers for 50% flowering time. This first report of GWAS for the selected characteristics of watermelon indicated a great potential to increase the efficiency of breeding programs.

Keywords: *Cucurbitaceae*, *Citrullus lanatus*, association mapping, linkage disequilibrium, Bayesian analysis

1. Introduction

Watermelon (*Citrullus lanatus* (Thunb.) Matsum. and Nakai) belongs to the genus *Citrullus* Schrad. Ex Eckl. et Zeyh in *Cucurbitaceae* family (Whitaker and Bemis, 1976). The genus comprises of four known diploid ($n = 11$) species (Dane and Liu, 2007). Of these, *Citrullus lanatus*, the most commercially important species, has two botanical varieties: *C. lanatus* var. *lanatus* Schrad. Ex Eckl. et Zeyh and var. *citroides* (L.H. Bailey). The other members of the genus *Citrullus* are *C. colocynthis* (L.) Schrad, *C. ecirrhosus* and *C. rehmii* De Winter.

Watermelon has many characters of interest for consumers and producers. Most of them are quantitatively controlled, and the genetic mechanisms of some of these traits were reviewed by Wehner (2008). The molecular tools helping in understanding the genetic mechanism of these traits might be very critical for increasing the efficiency of breeding programs. Trait loci affecting the phenotype can be detected by using linkage analysis and linkage disequilibrium mapping or a combination of both (Pikkuhookana and Sillanpaa, 2014). Both procedures use crossovers or recombinations between two loci. However, they differ in types of populations and power of detecting the number of alleles in a given population. Linkage mapping employs populations derived from controlled hybridizations

such as F1, F2, backcross and recombinant inbred lines whereas linkage disequilibrium (LD) or association mapping (AM) uses populations with broad genetic background such as germplasm collections, elite lines and synthetic populations. Linkage mapping is simple and effective, and can be handled by a single software (i.e. Mapmaker, MapManager, JOINMAP). Association mapping, on the other hand, requires sophisticated statistical procedure to avoid spurious associations dealing with structured populations (Ball, 2007). More importantly, linkage mapping detects alleles only present in two parents of a specific mapping population while AM finds all associating alleles with trait interest in a given population.

Many genetic linkage maps have been reported for several species of *Cucurbitaceae* family such as melon (Dogimont et al., 2000); cucumber (Park et al., 2000); and watermelon (Hawkins et al., 2001; Hashizume et al., 2003; Levi et al., 2006). Current status of mapping efforts on QTLs associating with several important fruit traits such as Brix, rind thickness, fruit length, seed oil percentage and 100 seed weight by using F2 and recombinant inbred line (RIL) populations were summarized by Reddy et al. (2014). Several studies on AM were reported for several major crops such as wheat (*Triticum aestivum* L.) (Tadesse et al., 2014); perennial ryegrass (*Lolium perenne* L.) (Aleliunas et al., 2015); pearl millet (*Pennisetum glaucum* (L.) R.Br.) (Kannan et al., 2014) and rapeseed (*Brassica napus* L.) (Cai et al., 2014). Linkage disequilibrium between loci in watermelons was previously reported (Ocal et al., 2014; Reddy et al., 2014). To date, however, there has been one report of marker-trait association through AM in the *Cucurbitaceae* family. Reddy et al. (2015) indicated utilization of AM based on 96 accessions sampled throughout the world and mapped four quantitative trait loci (QTL) for fruit length, width and weight. This approach has advantages over that derived from only two parents and may detect all alleles available in a given germplasm. Aims of this study were to make survey for GWAS and develop regression models for some selected characters of watermelon.

2. Materials and methods

2.1. Plant materials and DNA isolation

A total of 258 accessions used in this study were described by Ocal et al. (2014). The samples mainly included selfed (4 to 6 times) lines of the local genotypes. Seeds of each accession were sown in plastic multi-pots (4 x 4 x 4 cm) filled with peat:perlite (2:1) for germination and bulk of fresh true leaves from each accession were harvested into freezer bags, and kept at -20°C until DNA isolation. Total DNA was extracted from 30 mg tissue by using a modified CTAB DNA extraction procedure (Gulsen et al., 2010). DNA pellets were diluted with 300 µl of TE (10 mM Tris, 0.1 mM EDTA, pH 7.4). For PCR reactions, 10 ng/µl DNA templates were made using double-distilled water and checked on %2 agarose gels.

2.2. Molecular marker analysis

Sequence related amplified polymorphism (SRAP), random amplified polymorphic DNA (RAPD) and inter-simple-sequence repeats (ISSR) analyses were performed according to standard protocols. Two-hundred and eighth SRAP, 40 RAPD and 4 ISSR primers were tested for producing clear, polymorphic and repeatable bands. Selected primers were applied to the 258 lines. Each of 15 µl PCR components consisted of 0.66 mM of each of primers, 200 µM of each dNTPs, 1.5 µl of 10X PCR buffer, 2-2.5 mM of MgCl₂, ddH₂O, one unit of Taq polymerase and 20 ng of template DNA. PCR products of three marker systems were separated on 2-3% agarose gel at 110 V for 3-6 h and visualized under UV light. Peroxidase gene based markers were kindly provided by Ocal et al. (2014).

2.3. Selection of subpopulation for association analysis

The 258 accessions were subjected to initial screening to select individuals suitable for association mapping that requires non-identical, unique, but with low ratio of rare alleles. In

order to achieve this, all accessions were analyzed for the similarity. Each marker was visually scored as present (1) or absent (0) and data were analyzed with the Numerical Taxonomy Multivariate Analysis System (NTSYS-pc) version 2.1 software package (Exeter Software, Setauket, New York, USA) (Rohlf, 1993). For estimating relationships among the accessions, all polymorphic markers were used to produce simple matching coefficients. Then the similarity matrix was used to construct a dendrogram using the unweighted pair group method arithmetic average (UPGMA) to determine genetic relationships among the accessions studied. Finally, the number of plant materials was reduced to 96 by eliminating highly similar and distinct ones that likely cause low level of polymorphism or rare alleles.

2.4. Morphological characterization

Morphological characterization of the 96 accessions selected in previous step was performed in the conditions of Adana. Experimental and treatment designs were randomized block and unstructured design, respectively. Seeds were sown to 2:1 in perlite and when seedlings had come to the level with 2 true leaves, they have been planted under plastic tunnels with 2.0 m x 0.5 m spacing and 7 plants in each replication. The morphological traits investigated are as follows: fruit weight by averaging at least 3 fruits for each line, flesh color (orange, dark pink, yellow, light pink, pink, red, white, light red), fruit flesh firmness by averaging three fruits measured in 3 different points for fruit with 11 mm diameter penetrometer), the percentage of total soluble solid (TSS) after one slice from each three fruits in every lines have been squeezed and filtered, 3-5 drops were read in hand refractometer, the first flowering (beginning from sowing to that at least one female flower occurs in 50% of the plants), flowering time (beginning from sowing to when 50% of flowers in the plant were female), seed coat color (spotted black, black, cream, cream with black hilum, brown, milky brown, creamy pink, cream black edge, spotted cream, reddish brown, dark brown) and seed length

(mm) (measured with digital caliper), rind thickness, fruit diameter and height, presence of lines in fruit, fruit line width, ovary diameter and height measured by ruler or caliper. Ovary hairiness and existence of lines were visually detected. Distributions of phenotypes were drawn with Microsoft Excel version 2007.

2.5. Data Analysis

The UPGMA dendrogram was produced as described above. Principal component analysis (PCA) was used to estimate efficiency of selection of subpopulation in the previous step and easily visualize dispersion among the individuals on 2-D plot. The EIGEN module was used to calculate Eigen values and two-dimensional plots based on the variance-covariance matrix calculated between each two pairs of the 96 watermelons. Then PROJ module was used to construct 2-D graph of 258 accessions.

Population structure was analyzed using a model-based approach, Bayesian method by STRUCTURE software version 2.3.2 (Pritchard et al., 2000). Output of this analysis was used as covariate to avoid spurious linkages. Since this software was developed for codominant markers such as simple sequence repeat (SSR), the data file was modified based on suggestions by the manual. We designated if the band was scored as present, and we used (1,-9), unlike (2,-9), where -9 is the value used for missing data. Model-based cluster analysis was used to test whether $K = 1-10$, where K is the number of sub-populations. Several statistical procedures to allow a user to decide what value of K best fits the data were discussed by Barkley et al. (2006). In addition to these, there is informal pointer to detect the best K , in which values of $\log Pr(X/K)$ reach more or less plateaus after major decrease. In this sort of situation, where several values of K s give similar estimates of $\log Pr(X/K)$, it seems that the smallest is often correct according to the manual. We used this approach to estimate K . The authors state that while it may not be possible to know the true value of K ,

one should try to pick the smallest value of K that captures the major structure of the data (Pritchard et al., 2000). Admixture and independent allele frequencies models were used. For each population (K), 5000 iteration and 5000 burn-in period options were used. For each number of K from 1 to 10, five independent calculations were performed, and likelihood values obtained from these calculations were averaged for each K . Finally a graph was drawn by using Microsoft Office Excel 2007 for visual observation. Membership coefficients produced by the STRUCTURE program were used to infer whether the accessions are admixed by number of subpopulations or pure.

Association analyses between marker loci and phenotypic values were performed using the general linear model (GLM) functions in TASSEL stand-alone version 3.1 (Bradbury et al., 2007). To calculate associations between studied characters and polymorphic bands by TASSEL, three separate data files (trait, molecular marker data and population structure) were prepared. Initially, filtered structure file was created by selecting one subpopulation. Then, three files were joined by using JOIN option of TASSEL software. Two different p -values, F test- and permutation-based, were calculated between each marker and character, which provided significantly associating markers with the trait interest at alpha 0.05. Finally, backward regression was used to develop regression model to predict phenotype based on genotype by using SAS software. Backward regression starts with all candidate variables, testing the deletion of each variable using a chosen model comparison criterion, deleting the variable (if any) that improves the model the most by being deleted, and repeating this process until no further improvement is possible. Type I error was controlled by applying the 1,000 permutation test in TASSEL stand-alone version 3.1 (Bradbury et al., 2007) to control error associated with multiple comparisons and generate strong p -values for association analysis of polymorphic sites. In addition, marker loci with less than 10% occurrence for a given trait and remaining in the final model, was removed from the list and the regression analysis was rerun.

3. Results

3.1. Molecular marker analysis

A total of 55 primer combinations (40 SRAP, 11 RAPD and 4 ISSR) produced 210 scorable markers in this study. The number of bands per primer varied between 1 and 13 with the mean of 6.7. The ratio of polymorphism was 79%. The most productive pair was EM8-ME9 (13 bands). Data file of 150 peroxidase gene-based (POGP) markers were kindly provided by Ocal et al. (2014). A total of 360 markers were evaluated in further analyses in the current study. Markers were mostly dominant and primers were multilocus.

3.2. Selection of subpopulation for association mapping

The number of accessions (258) was reduced to 96 based on their similarity levels as detected by the UPGMA analysis. The similarity coefficients among the 258 lines ranged from 0.41 to 1.00 with a mean of 0.70 (data not shown). The majority of the lines had high similarity values. The 2-D scattered plot indicated strong clustering pattern (Fig. 1A) among 258 accessions. Although some of the watermelon accessions were discriminated, several others were highly similar or completely identical as follows: 275 and 276. The highly similar or identical accessions were removed based on their similarity values. After removal, the similarity coefficients ranged from 0.63 to 0.92, in which the pattern of clustering largely disappeared (Fig. 1B). Seven wild forms of *C. lanatus* var. *citroides* (Bailey) Mansf.) known as citron or preserving melon and two accessions of a related species (*Praecitrullus fistulosus* (Stocks) Pangalo) used as out-group were also removed due to possible cause of rare alleles. This step provided a core collection of the 96 *C. lanatus* var. *lanatus* lines, which were well-dispersed on 2-D graph based on PCA (Fig. 1A and B).

Figure 1. 2-D distributions of the PCA of 258 (A) and, after removal of similar, identical and highly distant individuals, 96 (B) accessions selected to study marker-trait associations.

3.3. Structure analysis

For substructuring Bayesian analysis by STRUCTURE software, we used only a subset of 35 markers since the program requires uncorrelated or loosely correlated markers with less than 10% missing data. These 35 markers were detected by using correlation option of SIMINT module nested in NTSYS software (Figure not shown). Only one marker was selected from each subgroup at a similarity value 0.65 or higher. The Bayesian analysis detected four subpopulations. However, after removal of very similar and those causing rare alleles, only two subpopulations were detected among the 96 samples (Fig. 2). The table indicating membership coefficients of each individual to each of hypothetical subpopulation was formatted and used in TASSEL.

Figure 2. Population structure of 96 watermelon lines for $K = 2$ to 7. Clusters are separated by vertical lines with cluster colors indicating various ancestries.

3.4. Estimating associations of markers with traits and regression models for traits

Genome-wide associations for 18 watermelon characters with marker loci and regression analysis were also performed. Table 1 indicates the number of marker loci detected to be associated with trait by TASSEL, the number of loci remaining after backward regression and proportion of variation explained by the final regression model for each character by SAS

software. For the highest (48%) and lowest (9%) variations, explained by the models, were of TSS and fruit flesh firmness, respectively. For each character, the loci remaining in the final model were not correlated implying independence from the rest of the loci.

For TSS, distribution of lines was continuous, therefore it was intervalized (Fig. 3A). The most frequent observation was of 5.01-6.00% with 27 times and the highest TSS value was >10.01 observed in eight lines. This result indicated possible control by polygenes. The TASSEL software detected 16 marker loci for TSS (Table 1 and 2). BR removed 11 of the loci and five remained in the final model, which explained 48% of the total variation for TSS among the population (Table 1 and 2). For fruit weight, the largest number individuals, 25 and 26, were between 2000.1-2500 and 2500.1-3000 g, respectively. Overall distribution was bell-shaped (Fig. 3B). TASSEL program detected 5 loci and after BR, only two remained in the model that explained 22% of the variation (Table 1 and 2). These markers were Em12.Me13.500 and (GA)₈YG.1000.

Table 1. Summary of association analysis based on TASSEL and SAS software for 17 characters studied

Trait	# Marker loci by TASSEL ^a	# Loci after backward regression ^b	Variation explained by model (%)
Fruit weight	5	2	22
Flesh color	13	6	41
Fruit flesh firmness	6	4	9
Total soluble solid	16	5	48
First flowering time	5	3	37
flowering time	7	6	39
Seed coat color	10	6	29
Seed length	10	6	21
Rind thickness	7	3	16
Fruit diameter	6	5	17
Fruit height	6	2	44
Presence of lines in fruit	1	1	11
Fruit line width	7	5	16
Ovary diameter	9	5	21
Ovary height	6	5	31
Ovary hairiness	5	4	15
Total yield	7	6	16

^a Number of loci detected by TASSEL software

^b Number of loci remaining after backward regression performed with SAS software

For seed coat color, the GLM analysis detected 10 loci at alpha 0.05 or lower, and 6 of them remained in the final model, explaining 29% of the total variation (Table 1, 3 and Fig. 3C). For seed length 10 loci were detected to be significant by TASSEL, 6 of which remained in the final model, explaining 21% of the total variation. For seed length, distribution was continuous, initial analysis found 10 markers, 6 of which remained in regression model (Table 1, 3 and Fig. 3D). For existence of lines on fruit, only one locus was detected by TASSEL programme, which explained 11% of the variation (Table 1). For flesh color, thirteen loci were found by TASSEL and 6 remained in the final model. This explained 41% of the total variation.

For flower type, 18 and 72 lines had hermaphrodite and non-hermaphrodite flowers (Figure not shown). The presence of hermaphrodite flower is controlled by a single gene (Wehner, 2008). Association analysis in TASSEL software detected 9 markers at alpha 0.05 or lower. These nine markers tested for regression model and only six remained in the model. The final model explained 37% of the variation for type of flower among the lines (Table 4).

For the first flowering time; it was found that individual number in 47-50 days interval is 8, individual number in 50.01-53 days interval is 52, individual number in 53.01-56 days interval is 30 and individual number in 56.01-59 days interval is 1 (Figure not shown). Distribution of first flowering time was continuous and probably controlled by polygenes. GLM analysis detected 13 marker loci at alpha level of 0.05 or lower. Backward regression removed 2 loci and 3 remained in the model at 10% level. The final model explained 37% of the variation for the first flowering time (Table 1).

Table 2. Results of backward regression analysis of total soluble solids (upper) and fruit weight (lower)

Variable	Parameter estimate	Standard error	Marker present/absent	# Observation	Mean $\pm$ Std. Dev
Total soluble solids (%)					
Intercept	5.52	0.33	-	-	-
EM3-ME4.1100	-0.69	0.24	0	54	6.2 $\pm$ 1.61
-			1	21	5.6 $\pm$ 1.12
OPS09.1800	-0.68	0.26	0	21	7.1 $\pm$ 1.62
-			1	70	5.67 $\pm$ 1.47
EM1-ME3.2000	0.95	0.27	0	18	5.03 $\pm$ 0.98
-			1	72	6.27 $\pm$ 1.65
OPE14.1200	0.97	0.23	0	55	5.32 $\pm$ 0.85
-			1	36	6.70 $\pm$ 1.67
Fruit weight (g)					
Intercept	3573.41	171.11	-	-	-
EM12-ME13.500	-677.90	171.55	0	22	3322 $\pm$ 981
-			1	68	2761 $\pm$ 756
(GA) ₈ YG.1000	-376.58	150.72	0	32	3150 $\pm$ 882
-			1	57	2760 $\pm$ 806

Table 3. Results of backward regression analysis of seed coat color (upper) and seed length (lower)

Variable	Parameter estimate	Standard error	Marker present/absent	#Observation	Mean±Std. Dev
Seed coat color					
Intercept	6.93	0.75	-	-	-
EM3-ME4-700	-2.37	0.77	0	23	6.26±4.67
-			1	63	3.9±2.95
EM5-ME13.410	2.88	0.85	0	72	3.94±2.98
-			1	19	6.31±4.78
EM12-ME1.1500	-2.44	0.69	0	45	5.6±4.3
-			1	44	3.31±2.13
Seed length (mm)					
Intercept	12.57	0.40	-	-	-
EM16-ME8.700	-1.08	0.50	0	75	12.85±1.46
-			1	13	11.44±1.62
EM9-ME1.720	0.83	0.43	0	19	11.87±1.96
-			1	70	12.86±1.51
EM14-ME12.510	-0.88	0.32	0	48	13.01±1.55
-			1	43	12.21±1.42

Figure 3. Distributions of watermelon characters: A) total soluble solids, B) fruit weight, C) seed coat pattern and color, D) seed length

4. Discussion

Ocal et al. (2014) and Reddy et al. (2014) reported existence of LD among chromosomes of watermelon, which is required for association analysis in plants. In addition, population structure and frequencies of allele distributions among the samples must be considered to avoid spurious linkages. Here, we performed association analysis for several watermelon characters with molecular markers based on conditions discussed above. We also developed regression models for every character to predict phenotypes of watermelons.

Selection of subpopulation for association analysis based on UPGMA analysis was highly efficient. This elimination provided a well set of accessions with very few rare alleles and good level of marker polymorphism. After removal, strong clustering pattern disappeared (Fig. 1A, before removal; Fig.1B, after removal of highly similar and distinct ones). The substructuring Bayesian analysis also indicated efficiency of selection of germplasm. By reducing the number, the number of subpopulations also decreased from four among the 258 samples to two among the 96 samples (Fig. 2).

This was the first report of marker-trait associations based on non-segregating germplasm accessions in watermelons belonging to the family of *Cucurbitaceae*. This study indicated that marker loci sampled genome-wide significantly associated the traits of watermelons. This supported the findings of Ocal et al. (2014) and Reddy et al. (2014), who reported the existence of LD among watermelons based on peroxidase gene-based polymorphism and single nucleotide polymorphism (SNP) markers, respectively.

Seed characters such as seed size and seed coat color are important for acceptability of watermelon and may, to a certain extent, affect customer acceptance (Lou and Wehner, 2009). Out of 96 lines, 31 and 32 were black and creamy, and the rest were either patterned black or cream in seed coat (Fig. 3C). The GLM analysis detected 10 loci at alpha 0.05 or lower, and 6 of them remained in the final model, explaining 29% of the total variation (Table 1 and 3).

According to Lou and Wehner (2009), progenies of some specific crossings differed in seed coat color, implying that cytoplasmic organelles (chloroplast or mitochondria) might have control over seed coat color. Poole et al. (1941) reported that seed length was highly correlated with seed size and weight. They also reported that medium seed size was dominant over short and long seed size. The samples used in this study had seed sizes varying from 8.0 to 16 mm long, but the distribution was not bell shaped. Fifty-eight accessions (2/3rd) were in the range of 12.0 and 14 mm. In our study, for seed length, 10 loci were listed by TASSEL, 6 of which remained in the final model, explaining 21% of the total variation (Table 1 and 3). Apparently, seed length was under control by polygenes but few major and some other minor genes. This agreed with the results of Poole et al. (1941). A recent study found interaction of loci with years for this trait (Reddy et al., 2014).

Flesh color is another important trait of watermelons. Thirteen loci were found by TASSEL and 6 remained in the final model, which explained 41% of the total variation (Table 1). According to Lou and Wehner (2009), scarlet red flesh color was the major allele dominant over orange flesh color, but the different shades in the progenies indicates that there may be modifying genes as well as environmental effects. This agreed with the results of the current study.

For fruit weight, Lou and Wehner (2009) reported that the environmental variance was larger than the genetic variance, and both narrow- and broad-sense heritability were low to medium. We found five loci by TASSEL, but only two remained in the final model, explaining 22% of the total variation (Table 1 and 2). Probably, there are some other loci due to very low minor effects that were not detected with the environmental conditions in this study. Lou and Wehner suggested a level of 5% selection intensity for this trait. Thus, environmental conditions clearly play significant role for this trait. Reddy et al. (2014) found one marker for this trait explaining 10.6% of the total variation.

Existence and type of rind pattern on fruit is an important trait since consumers may have particular preference. Striped rind is more popular over some other rind patterns. We also investigated associations for this trait and found one marker that explained 11% of the total variation (Table 1). This is probably due to the high distance between marker and trait loci.

Total soluble solids are one of the most important characteristics of watermelon affecting consumer acceptance. In this study, distribution of lines was continuous, which indicated possible control by polygenes (Fig. 3A). Lou and Wehner (2009) reported that heritability of this trait was medium to high, and its repeatability over different environments was high. The variation explained by the markers detected was relatively high (48% of the total variation) for TSS among the population (Table 1 and 2). This study suggested that marker-assisted selection was possible for this trait.

For the first flowering time, distribution was continuous and probably controlled by polygenes. GLM analysis detected 5 marker loci at alpha level of 0.05 or lower. The BR removed 10 loci and 3 remained in the model at 10% level. The final model explained 37% of the variation for the first flowering time among the lines.

In this study, 96 accessions were selected from 258 watermelons. The samples included well-dispersed set of lines to avoid highly similar/identical lines and excluded relatively distant relatives causing rare alleles. In addition, few marker loci having less than 10% of occurrence for a given trait were removed from the final model. This probably increased power of association analyses and decreased frequency of spurious associations as discussed by Wongkaew et al. (2014). These may overcome limitations (i.e. small population size that likely miss some loci with minor effects and frequency of spurious linkages) of genome-wide associations. Bradbury et al. (2011) showed that population sizes of 300 were sufficient to detect QTL for traits with moderate to high heritability; however as the number of QTL increased, the larger effect QTL were detected but not the smaller effect QTL. Due to low

number of markers in this study, this might have happen. Spurious association found in a given population could be corrected by using statistical models accounting for population structure (Q matrix) to reduce the false positive associations (Yu et al., 2006). We accordingly used Q matrix as covariate for association analyses in TASSEL. Accordingly, this study indicated that association mapping offered great potential in watermelons, and some cautions such as selection of mapping population with fewer rare alleles and highly similar ones could enhance power of association analysis.

Non-locus specific markers were sampled from the watermelon genome and used for GWAS for 18 important watermelon characters including TSS, flesh color and seed size. The results indicated existence of significant associations between marker loci and selected traits. One criticism may be the type of markers. The marker types used in this study were mainly dominant. This sort of markers provide less resolution power compared to codominant markers such as SNP and SSR markers. However, the Mantel test verified that similarity matrices of most known markers such as ISSR, SRAP, POGP, and SSR markers highly correlated ($r > 0.8$) as discussed by Gulsen et al. (2009). The results of this study provided a valuable preliminary survey of GWAS with some selected traits of watermelons, and also contributed the information for further research to explore potential of GWAS for complex traits such as TSS and seed size in watermelon. In the future, locus specific such as SNP and SSR marker-based regression models should be developed for MAS-based breeding programs of watermelon. This would significantly increase efficiency since morphological characterization is more time-consuming and less efficient.

Acknowledgments

This study was financially supported by the section of Scientific Research Unit of

Erciyes University [FBY-12-3838], and Turkish Scientific Council of Turkey [110 O 104].

Some of the plant materials were kindly provided by Aegean Agricultural Research Institute (AARI), USDA-USA, INRA-France and Dr. Csaba ZSAMOSI, Budapest Corvinus University.

References

- Aleliunas, A., Jonaviciene, K., Statkeviciute, G., Vaitiekunaite, D., Kemesyte, V., Lubberstedt, T., Brazauskas, G., 2015. Association of single nucleotide polymorphisms in LpIRI1 gene with freezing tolerance traits in perennial ryegrass. *Euphytica* 204, 523-534.
- Ball, R.D., 2007. Statistical Analysis and Experimental Design. In: Association Mapping In Plants (eds. Oraguzie NC, Rikkerink EHA, Gardiner SE, and Silva HND). Springer Science, Business LLC. P. 133-196.
- Barkley, N.A., Roose, M.L., Krueger, R.R., Federici, C.T., 2006. Assessing genetic diversity and population structure in a citrus germplasm collection utilizing simple sequence repeat markers (SSRs). *Theor Applied Genet* 112, 1519-1531.
- Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y., Buckler, E.S., 2007. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23, 2633-2635.
- Bradbury, P.J., Parker, T., Hamblin, M.T., Jannink, J.L., 2011. Assessment of power and false discovery in genome-wide association studies using the Barley CAP germplasm. *Crop Sci* 51, 52-59.
- Cai, D., Xiao, Y., Yang, W., Ye, W., Wang, B., Younas, M., Wu, J., Liu, K., 2014. Association mapping of six yield-related traits in rapeseed (*Brassica napus* L.). *Theor Applied Genet* 127, 85-96.
- Dane, F., Liu, J., 2007 Diversity and origin of cultivated and citron type watermelon (*Citrullus lanatus*). *Genet Resour Crop Evol* 54, 1255-1265.

- Dogimont, C., Leconte, L., Périn, C., Thabuis, A., Lecoq, H. and Pitrat, M., 2000. Identification of QTLs contributing to resistance to different strains of cucumber mosaic cucumovirus in melon. *Acta Hort.* 510, 391-398.
- Gulsen, O., Sever-Mutlu, S., Mutlu, N., Tuna, M., Karaguzel, O., Shearman, R.C., Riordan, T.P., Heng-Moss, T.M., 2009. Polyploidy creates higher diversity among *Cynodon* accessions as assessed by molecular markers. *Theor Applied Genet* 118: 1309–1319.
- Gulsen, O., Kaymak, S., Ozongun, S., Uzun, A., 2010. Genetic analysis of Turkish apple germplasm using peroxidase gene-based markers. *Scientia Hort* 125: 368-373.
- Hashizume, T., Skimamoto, I., Hirai, M., 2003. Construction of a linkage and QTL analysis of horticultural traits for watermelon [*Citrullus lanatus* (Thunb.) Matsum & Nakai] using RAPD, RFLP and ISSR markers. *Theor Applied Genet* 106: 779-785.
- Hawkins, L.K., Dane, F., Kubisiak, T.L., Rhode, B.B., Jarret, R.L., 2001. Linkage mapping in A watermelon population segregating for *Fusarium* wilt resistance. *J Amer Soc Hort Sci* 126: 344-350.
- Kannan, B., Senapathy, S., Raj, A.G.B., Chandra, S., Muthiah, A., Dhanapal, A.P., Hash, C.T., 2014. Association analysis of SSR markers with phenology, grain, and stover-yield related traits in pearl millet (*Pennisetum glaucum* (L.) R. Br.) *Scientific World J* 2014: 1-14.
- Levi, A., Thomas, C.E., Trebitsh, T., Salman, A., King, J., Karalius, J., Newman, M., Reddy, O.U.K., Xu, Y., Zhang, X., 2006. An extended linkage map for watermelon based on SRAP, AFLP, SSR, ISSR, and RAPD markers *J Amer Soc Hort Sci* 131: 393-402.
- Lou, L., Wehner, T.C., 2009. Chapter one: Genetic control of seed characters in watermelon- A review. In: *Inheritance of fruit characteristics in watermelon [Citrullus lanatus* (Thunb.) Matsum. & Nakai]. MSc Thesis, North Carolina State University, NC, USA

- Ocal, N., Akbulut, M., Gulsen, O., Yetişir, H., Solmaz, I., Sari, N., 2014. Genetic diversity, population structure and linkage disequilibrium among watermelons based on peroxidase gene markers. *Scientia Horti* 176: 151-161.
- Park, Y.H., Sensoy, S., Wye, C., 2000. A genetic map of cucumber composed of RAPDs, RFLPs, AFLPs, and loci conditioning resistance to papaya ring spot and zucchini yellow mosaic viruses. *Genome* 43,1003-1010.
- Pikkuhookana, P., Sillanpää, M.J., 2014. Combined linkage disequilibrium and linkage mapping: Bayesian multilocus approach. *Heredity* 112, 351-360.
- Poole, C.F., Grimball, P.C., and Porter, D.R., 1941. Inheritance of seed characters in watermelon. *Journal of Agricultural Research* 63,433-456.
- Pritchard, K.J., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. *Genetics* 155, 945-959.
- Reddy, U.K., Nimmakayala, P., Levi, A., Abburi, V.L., Saminathan, T., Tomason, Y.R., Vajja, G., Reddy, R., Abburi, L., Wehner, T.C., Ronin, Y., Karol, A., 2014. High-resolution genetic map for understanding the effect of genome-wide recombination rate on nucleotide diversity in watermelon. *Genes Genomes Genetics* 4, 2219-2230.
- Reddy, U.K., Abburi, L., Abburi, V.L., Saminathan, T., Cantrell, R., Vajja, V.G., Reddy, R., Tomason, Y.R., Levi, A., Wehner, T.C., Nimmakayala, P., 2015. A Genome-wide scan of selective sweeps and association mapping of fruit traits using microsatellite markers in watermelon. *J. Heredity* 106, 166-176.
- Rohlf, J.F., 2000) NTSYS-pc: numerical taxonomy and multivariate analysis system. Exeter Software, Setauket, New York.
- Tadesse, W., Ogbonnaya, F.C., Jighly, A., Nazari, K., Rajaram, S., Baum, M., 2014. Association mapping of resistance to yellow rust in winter wheat cultivars and elite genotypes. *Crop Sci* 54, 607-616.

- Wehner, T.C., 2008. Watermelon. In: J Prohens, F Nuez (Eds.), Handbook of Plant Breeding; Vegetables I: Astaraceae, Brassicaceae, Chenopodiaceae, and Cucurbitaceae, Springer Science Business LLC, New York, NY, pp. 381-418.
- Whitaker, T.W., and W.P. Bemis, 1976 Cucurbits, pp. 64–69 in Evolution of crop plants, edited by N. W. Simmonds Longman, London.
- Wongkaew, A., Phumichai, C., Chunwongse, J., Jampatong, S., Grudloyma, P., Pulam, T., Doungchan, W., 2014. Detection of candidate R genes and single nucleotide polymorphisms for downy mildew resistance in maize inbred lines by association analysis. *Euphytica* 197, 109-118.
- Yu, J., Buckler, E.S., 2006. Genetic association mapping and genome organization of maize. *Curr Opin Biotechnol* 17,155-160.