Turkish Journal of Agriculture and Forestry

Volume 40 | Number 4 Article 15

1-1-2016

Genetic analyses of Turkish watermelons based on SRAP markers

METİN YAĞCIOĞLU

OSMAN GÜLŞEN

İLKNUR SOLMAZ

HALİT YETİŞİR

NEBAHAT SARI

Follow this and additional works at: https://journals.tubitak.gov.tr/agriculture

Part of the Agriculture Commons, and the Forest Sciences Commons

Recommended Citation

YAĞCIOĞLU, METİN; GÜLŞEN, OSMAN; SOLMAZ, İLKNUR; YETİŞİR, HALİT; and SARI, NEBAHAT (2016) "Genetic analyses of Turkish watermelons based on SRAP markers," *Turkish Journal of Agriculture and Forestry*: Vol. 40: No. 4, Article 15. https://doi.org/10.3906/tar-1602-32
Available at: https://journals.tubitak.gov.tr/agriculture/vol40/iss4/15

This Article is brought to you for free and open access by TÜBİTAK Academic Journals. It has been accepted for inclusion in Turkish Journal of Agriculture and Forestry by an authorized editor of TÜBİTAK Academic Journals. For more information, please contact academic.publications@tubitak.gov.tr.

Turkish Journal of Agriculture and Forestry

http://journals.tubitak.gov.tr/agriculture/

Research Article

Turk J Agric For
(2016) 40: 613-620
© TÜBİTAK
doi:10.3906/tar-1602-32

Genetic analyses of Turkish watermelons based on SRAP markers

Metin YAĞCIOĞLU[1], Osman GÜLŞEN[1,*], İlknur SOLMAZ[2], Halit YETİŞİR[1], Nebahat SARI[2]
[1]Department of Horticulture, Erciyes University, Melikgazi, Kayseri, Turkey
[2]Department of Horticulture, Çukurova University, Balcalı, Adana, Turkey

Received: 07.02.2016 • **Accepted/Published Online:** 08.05.2016 • **Final Version:** 14.06.2016

Abstract: Genetic studies of watermelons [*Citrullus lanatus* (Thunb.) Matsum. & Nakai] sampled from Turkey, the second leading producer, may provide valuable information for breeding and research programs. The objectives of this study were to estimate the level of genetic diversity, population structure, and optimum genome sampling size among the Turkish and several introduced watermelons. From the collection, 256 watermelon lines representing all watermelon-growing areas of Turkey along with two accessions of a related genus (*Praecitrullus fistulosus* (Stocks) Pangalo) and a few popular cultivars were genotyped using sequence-related amplified polymorphism (SRAP) markers. Twenty-seven primers generated 210 molecular markers for genetic analyses. The unweighted pair group method with arithmetic average analysis produced no clear-cut pattern, while principal component analysis indicated three subgroups among the lines studied. They had a narrow genetic base within *C. lanatus* var. *lanatus* and were mostly distinguished from each other. Model-based structure analysis indicated that the number of subpopulations in watermelons was four. Among the 258 lines, only 20% (51 lines) had 0.80 or greater membership coefficients to one subpopulation, and therefore were not admixed. The remaining 207 were admixed by at least two subpopulations. The similarity matrix of genome sampling size of 40 randomly selected markers was highly correlated (r = 0.915) with the matrix of simple matching coefficients based on 100 markers, implying that 40 markers were detected to be sufficient for this species. Overall, this study concluded that the majority of the Turkish watermelons were distinguished from each other, difficult to classify, and admixed, with narrow genetic variation.

Key words: Watermelon, estimation of genetic parameters, germplasm, identification of duplicate, sampling size

1. Introduction

Watermelon [*Citrullus lanatus* (Thunb.) Matsum. & Nakai] is one of the most important fruit crops and belongs to the family *Cucurbitaceae*. Almost 7% of the fruit and vegetable production area in the world is accounted for by watermelon. Turkey is the second largest watermelon (*C. lanatus*) producer with 4.044 × 10^6 t (http://faostat.fao.org/site/339/default.aspx). It is cultivated in all regions of the country, which has a diverse range of climates (Solmaz and Sari, 2009). Although it is not a center of origin for watermelons, many landraces emerged in different regions of Turkey from the Mediterranean to the Black Sea regions and from Hakkari bordering Iran to Edirne neighboring Bulgaria and Greece, as documented in the studies of Sari et al. (2008) and Solmaz and Sari (2009). Production is mainly based on F1 hybrid cultivars, but there are local types as well.

Expectedly, genetic erosion is a main concern in watermelon breeding due to increasing pressure on farmers to use more productive and popular F1 hybrid cultivars. However, information on genetic diversity and population structure of watermelons from Turkey is scarce and the number of lines studied is limited (Solmaz and Sari, 2009; Uluturk et al., 2011). Watermelon has been cultivated in the Near East including Turkey for thousands of years (Wehner, 2008). Over the years, local landraces emerged in regions of Turkey with varying environmental constraints such as low-high temperatures, humidity, and pathogen pressure. Agronomical properties of these landraces such as biotic and abiotic stress responses, quality, and yield are unknown. These collections provide an important genetic base for breeding programs and require appropriate sampling and characterization. In general, genetic diversity among the cultivated watermelons is low (Uluturk et al., 2011), but intra- and interspecific crosses are possible to some degree, which may help broaden their genetic basis (Sain et al., 2002). The objectives of this study were to estimate the level of genetic diversity, population structure, and genome sampling size among Turkish watermelons along with a few introductions that are available in the germplasm.

* Correspondence: ogulsen@erciyes.edu.tr

2. Materials and methods

2.1. Plant materials

A total of 258 lines were used consisting of 249 cultivated watermelon *C. lanatus* var. *lanatus* landraces; seven wild forms of *C. lanatus* var. *citroides*, known as citron or preserving melon; and two *Praecitrullus fistulosus* lines as an outgroup. These plant materials and their DNA isolation were previously described by Ocal et al. (2014).

2.2. SRAP analyses

Twenty-seven sequence-related amplified polymorphism (SRAP) primer combinations determined to produce clear, polymorphic, and repeatable bands were applied to DNA samples of the 258 lines (Table). For amplifications, each 15 µL of PCR components consisted of 1.33 mM of each primer, 200 µM of each dNTP, 1.5 µL of 10X PCR buffer, 2.5 mM of $MgCl_2$, 7.62 µL of ddH_2O, 1 U of Taq polymerase, and 20 ng of template DNA. PCR cycling parameters were the same as reported by Gulsen et al. (2007). PCR products were separated on 2% agarose gel at 110 V for 5 to 6 h and visualized under UV light.

2.3. Statistical analyses

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, NY, USA). For estimating relationships among the lines, the similarity matrix (SM) based on the Dice similarity coefficient was used to construct an unweighted pair group method with arithmetic average (UPGMA) dendrogram to determine genetic relationships. Principal component analysis (PCA) was used to easily visualize the differences among the individuals and determine the optimum number of clusters. The EIGEN module was used to calculate eigenvalues and 2-dimensional plots based on the variance-covariance matrix, as suggested by NTSYS manual, calculated among the watermelon lines. With these matrices, the eigenvectors were computed using the EIGEN module, and finally a 2-dimensional graph was obtained with the PROJ module.

Population structure was estimated by using a model-based approach, the Bayesian method nested in STRUCTURE software (Pritchard et al., 2000). This program assigns individuals to subpopulations by calculating their membership coefficient. If a band was scored as present 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 subpopulations. 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. Membership coefficients produced by the STRUCTURE program were used to infer whether the lines are admixed by a number of subpopulations or pure. For all analyses with STRUCTURE, we used only a subset of markers that had less than 10% missing data and were uncorrelated or had loose correlation. These markers were detected by using the correlation option of the SIMINT module nested in the NTSYS software. As described in the manual, there is an informal pointer to detect the best K, in which values of log Pr (X/K) reach more or less plateaus after a major decrease. In this sort of situation where several values of K give similar estimates of log Pr (X/K), it seems that the smallest is often correct. We used this approach to estimate K. 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).

Optimum genome sampling size in watermelons was estimated according to Gulsen et al. (2011). Seven different data files of 5, 10, 20, 40, 60, 80, and 100 randomly selected markers were created. SMs were then produced based on Dice similarity coefficients for each data file. Later, a two-way Mantel test was used to estimate Mantel correlation coefficients between each pair of SMs using the MXCOMP option nested in NTSYS pc version 2.1. The least number of permutations for each run was 1000 and the normalized Mantel test was used for comparisons in this study.

3. Results

3.1. SRAP and UPGMA analysis

A total 27 SRAP primer combinations produced 210 markers (Table). EM8-ME9, EM10-ME10, EM1-ME10, EM7-ME2, and EM9-ME11 were the most productive with 10 or more bands per primer combination. Markers were dominant and primers were multilocus, targeting more than one locus as expected.

The similarity coefficients ranged from 0.16 to 0.98 when all lines were included, 0.46 to 0.98 when only *C. lanatus* var. *lanatus* and *C. lanatus* var. *citroides* were included, and 0.83 to 0.98 when only *C. lanatus* var. *lanatus* was included (Figure 1). Although most watermelon lines were discriminated from each other, several others were completely identical as follows: 162 and 185; 17 and 88; 62, 142, and 306; 87 and 244; 222 and 296; 241 and 242; 127 and 254; 98 and 180; 65 and 106; 271 and 273; 22 and 276; and 184 and 188 (data not shown). These pairs are likely to be derived from the same individual. The remaining 232 *C. lanatus* lines were distinct from each other. Of the 257 *C. lanatus* lines, 242 lines had a similarity range from 0.94 to 1.00. Four popular cultivars (36, Sugar Baby; 39, Crimson Sweet; 233, Calhoun Gray; and 235, Charleston Gray) used in this study were discriminated from the other lines but closely clustered with the Turkish watermelons (Figure 1).

Table. Primer pairs, number of polymorphic loci, allele sizes, and rates of polymorphic loci.

Primer pairs	#Polymorphic loci	Allele sizes (bp)	Polymorphism rate (%)
EM14-ME13	9	280–1500	100
EM5-ME2	8	250–1300	100
EM11-ME5	8	240–1300	80
EM3-ME4	7	280–1800	100
EM4-ME5	6	550–2300	100
EM5-ME5	4	350–1100	100
EM1-ME10	11	180–1850	90.9
EM12-ME13	8	250–900	100
EM7-ME2	10	150–1400	90
EM8-ME9	13	270–1500	76.9
EM14-ME5	5	290–1800	100
EM1-ME4	7	100–800	100
EM10-ME10	12	200–1000	100
EM9-ME11	10	120–1000	100
EM16-ME11	3	550–800	100
EM16-ME8	6	150–800	100
EM15-ME13	8	250–810	100
EM15-ME7	7	250–950	85.7
EM12-ME9	5	230–680	80
EM14-ME11	7	280–950	100
EM13-ME7	6	300–900	85.7
EM11-ME7	1	320	100
EM6-ME10	1	420	100
EM7-ME10	7	220–1000	71.4
EM2-ME10	2	800–900	100
EM11-ME10	7	200–1090	100
EM2-ME6	5	200–800	100

3.2. Principal component analysis

The PCA confirmed three distinct groups (Figure 2). Group A contained two lines of *P. fistulosus* (331 and 333) and group B had ten lines (24, 26, 151, 216, 234, 324, 327, 328, 351, and G41). Of these 10 lines in Group B, 26, 234, 324, 327, 328, 351, and G41 were listed as *C. lanatus* var. *citroides* and 24, 151, and 216 were listed as *C. lanatus* var. *lanatus* in our records. The largest group was C that included 246 *C. lanatus* var. *lanatus* lines. The first three eigenvalues explained 92% of the total genetic variation.

3.3. Model-based cluster analysis

Since the STRUCTURE program requires uncorrelated markers, the 41 most loosely correlated markers were selected from the 210 markers by using the CORR option of the SIMINT module nested in the NTSYS program. These 41 markers were then used to perform structure analysis. We used an informal pointer to detect the optimum number of subpopulations as suggested by the STRUCTURE manual (Pritchard et al., 2000). The analysis indicated that there were four subpopulations among

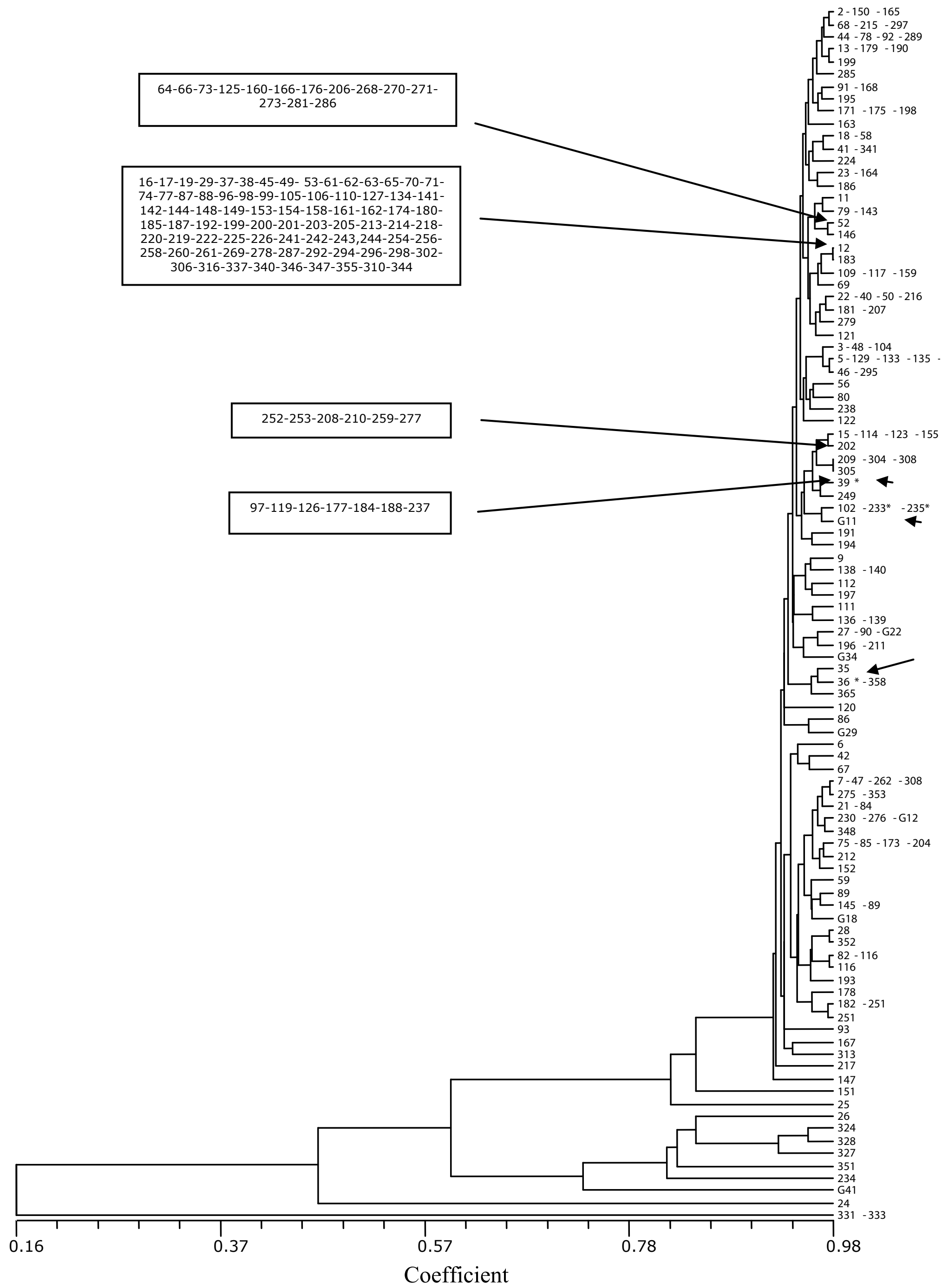

Figure 1. The UPGMA analysis based on Dice coefficients of 210 SRAP markers sampled from the 256 *Citrullus* and 2 *P. fistulosus* lines. Some popular cultivars, marked with an asterisk, are included: 36, Sugar Baby; 39, Crimson Sweet; 233, Calhoun Gray; and 235, Charleston Gray. They are indicated with arrows. The numbers in each box on the left are the accessions nested together with the genotypes pointed at by an arrow.

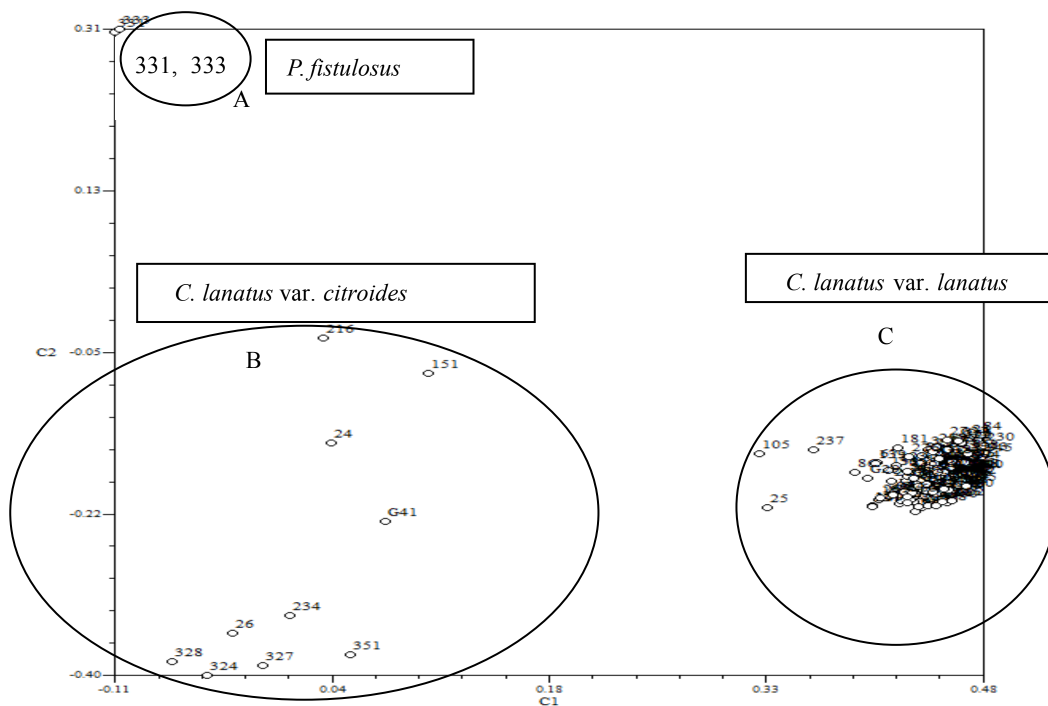

Figure 2. Two-dimensional graph of the PCA obtained by using 210 SRAP markers from the 256 *Citrullus* and 2 *P. fistulosus* lines.

the lines studied (Figure 3). The frequencies of the 258 lines in the four subpopulations were listed in decreasing order: subpopulation 3 (0.391), 4 (0.384), 1 (0.178), and 2 (0.047). Individuals with a membership coefficient of 0.80 or more to a subpopulation are considered as pure (Fukunaga, 2005), whereas individuals with lower membership coefficients are considered as hybrids. In this study 52 individuals had 0.80 and higher membership coefficients and therefore were likely to be pure and not admixed (Figure 3). The remaining 206 individuals had varying levels of membership coefficients and were likely to be admixed by at least two subpopulations. Of 52 nonadmixed lines, lines 43 and 9 belong to subpopulation 1 and 2, respectively.

3.4. Estimating optimum genome sampling size

Mantel correlation coefficients between 5 and 10, 10 and 20, 20 and 40, 40 and 60, 60 and 80, and 80 and 100 marker-based Dice SMs were 0.755, 0.697, 0.898, 0.931, 0.983, and 0.997, respectively (Figure 4). Matrices calculated with 5, 10, 20, 40, 60, and 80 markers correlated with 100 markers were 0.433, 0.504, 0.833, 0.915, 0.976, and 0.996, respectively. About 40 random markers achieved correlation values (r) higher than 0.9. Increasing the genome sampling size above 40 markers did not add significant information.

4. Discussion

Turkey is one of the main watermelon producers in the world and a considerable number of landraces exist in its different regions (Sari et al., 2008). In this study, we tried to elucidate the genetic structure of Turkish watermelons in relation to several known foreign watermelon cultivars, which indicated a low level of variation. This was consistent with the previous report of Uluturk et al. (2011). Solmaz and Sari (2009) indicated considerable diversity for fruit and seed characteristics among the Turkish watermelons. In general, morphological diversity usually reflects molecular diversity, probably due to small DNA changes that are difficult to detect with random genome sampling strategies and low resolution. Thus, morphological and molecular genetic diversity are not alternatives to each other and must be considered separately in germplasm characterization, as also suggested by Uluturk et al. (2011). In addition, it was difficult to classify *C. lanatus* var. *lanatus* lines from Turkey, which also was consistent with previous study of Turkish watermelons. Nuclear genome markers sufficiently differentiated among American watermelon cultivars with limited genetic diversity (Levi et al., 2004). Except for a few lines, similar findings were also observed among the Turkish watermelons. There could be several reasons for the low level of diversity among the watermelons in this study. First, introduction of watermelon to Turkey

Figure 3. Graphical presentation of membership coefficients of the 258 lines obtained from the STRUCTURE program by using 41 loosely correlated DNA markers. Each color indicates a putative subpopulation.

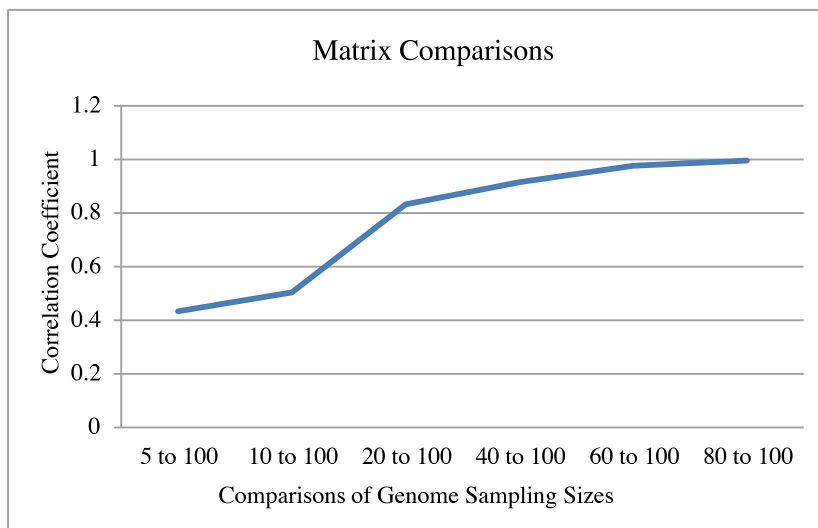

Figure 4. Changes of correlation coefficients with varying genome sampling sizes calculated by using the Mantel test.

is relatively recent, and therefore perhaps not enough time has passed for genetic diversification, as discussed by Navot and Zamir (1987). Second, selection pressure over the landraces is severe, which prevents genetic diversification, particularly for the traits that affect consumer preference. In nature, narrow diversity is likely to cause potential severe pest outbreaks and broadening genetic diversity may compensate for the severity of unexpected outbreaks. Hence, low genetic diversity indicates the need to broaden the genetic base of cultivated watermelon (Levi et al., 2001). Introgression of related species should be considered through sexual hybridization, somatic hybridization, genetic transformation, mutation breeding, and polyploidization to increase genetic diversity among watermelons. In addition, the Turkish watermelons were slightly different from four popular cultivars (36, Sugar Baby; 39, Crimson Sweet; 233, Calhoun Gray; and 235, Charleston Gray) (Figure 1). Geography-based clustering among the Turkish watermelons did not exist in this study, which was also consistent with previous reports of watermelons (Solmaz et al., 2010; Ocal et al., 2014; Wang et al., 2015).

The PCA indicated three distinct groups (Figure 2). Of the lines studied, 26, 234, 324, 327, 328, 351, and G41 were listed as *C. lanatus* var. *citroides* and 24, 151, and 216 lines were listed as *C. lanatus* var. *lanatus* in our records. Lines 24 and 25 were sampled from Egypt, and both were placed intermediately between Group B and C. Accession 151 was morphologically distinct, grown as winter watermelon, and 216 was grown in the most western part of Turkey (near Bulgaria). Probably these two lines were admixed between *C. lanatus* var. *lanatus* and related species. The largest group was C that included 246 *C. lanatus* var. *lanatus* lines. The first three eigenvalues explained 92% of the total genetic variation, meaning that the PCA could be useful (Mohammadi and Prasanna, 2003). After removing two *P. fistulosus* lines, resolution among the var. *lanatus* lines was a little better. The *C. lanatus* var. *lanatus* lines were clustered into three subgroups (Figure 2). Probably *C. lanatus* previously hybridized with a few distantly related taxa such as *C. colocynthis*. The UPGMA (Figure 1) and PCA (Figure 2) resulted in similar results. For example, almost all genotypes in Group A and B in PCA were similarly placed in the UPGMA dendrogram.

The Bayesian analysis indicated four subpopulations among the samples based on the 41 most loosely correlated markers (Figure 3). Individuals with membership coefficients of 0.80 or more to a subpopulation are considered as pure (Fukunaga, 2005), whereas individuals with lower membership coefficients are considered as hybrids. In this study, 51 individuals had membership coefficients of 0.80 and higher, and therefore they are likely to be pure or not admixed (Figure 3). The remaining 207 individuals had varying levels of membership coefficient and were likely admixed by at least two subpopulations. Of the 51 nonadmixed lines, lines 41 and 10 belonged to subpopulation 1 and 2, respectively. In the third and fourth subpopulations, all lines were admixed by at least 2 to 4 subpopulations. For example, lines 150 and 237 were admixed by three subpopulations in the third subpopulation while 216 of the fourth subpopulation had membership in all four subpopulations. Introgressions among watermelons are expected because it has a monoicous flower structure, which favors cross-pollination by means of insects. Despite some similarities, the results of UPGMA (Figure 1) and PCA (Figure 2) were more similar than that of Bayesian analysis (Figure 3). For example, 331 and 333 were distantly placed by UPGMA and PCA but differently assigned by the Bayesian analysis, intermixing with some members of Group A of PCA. This is probably caused by essential differences in the parameters used by these approaches. This kind of information may help germplasm organization and directed hybridization in breeding programs.

Molecular marker systems and genome sampling size estimate different levels of associations (Gulsen et al., 2011). For example, data produced with SRAP, inter-simple sequence repeat (ISSR), random amplified polymorphic DNA (RAPD), and peroxidase gene-based polymorphism markers had a correlation value of 0.81 as verified by the Mantel test, but data of RAPD markers always had lower correlation values (about 0.7). Hence, SRAP-based estimation of relationships highly correlates with the results of most marker systems such as simple sequence repeat, amplified fragment length polymorphism, and ISSR (Wang et al., 2015). For estimating genome sampling size, one approach is to use a set that covers the genome. This was 20 loci for maize (*Zea mays* L.) (Mumm and Dudley, 1994). When numbers reach 50 to 100 loci, results are usually consistent with pedigree information. Here we tried to estimate the minimum genome sampling size for *Citrullus* lines (Figure 4). Correlation coefficients increased significantly with the number, then reached a plateau at around 40 to 60 markers. A size of 40 or 60 markers was found to provide a sufficient correlation coefficient (r) of higher than 0.90 with that of 100 markers; therefore, this size of genome sampling appeared to be sufficient in *Citrullus* lines because increasing the genome sampling size above 40 or 60 did not add significant information. It was 20 markers in *Cynodon* lines as reported by Gulsen et al. (2011). The number of polymorphic markers analyzed is important to detect true associations among taxa, and developing cost-efficient genome sampling strategies could be beneficial to watermelon breeding programs as well as other plant species. This approach may apply to other plant species with similar genome structure.

Acknowledgments
This study was funded by the Scientific Research Unit of Erciyes University (FBY-12-3838). The authors thank the Turkish Gene Bank Aegean Agricultural Research Institute (AARI), USDA-USA, INRA-France, and Dr Csaba Zsamosi of Corvinus University of Budapest for providing seeds of some lines.

References

Fukunaga K, Hill J, Vigoroux Y, Matsuoka Y, Sanchez GJ, Liu K, Buckler ES, Doebley J (2005). Genetic diversity and population structure of Teosinte. Genetics 169: 2241-2254.

Gulsen O, Karagul S, Abak K (2007). Diversity and relationships among Turkish okra germplasm by SRAP and phenotypic marker polymorphism. Biol Bratis 62: 41-45.

Gulsen O, Sever-Mutlu S, Mutlu N, Karaguzel O, Hocagil MM (2011). Estimating optimum number of marker loci for genetic analyses in *Cynodon* accessions. Biochem Syst Ecol 39: 906-909.

Levi A, Thomas CE, Newman ML, Reddy O, Zhang X (2004). ISSR and AFLP markers sufficiently differentiated among the American watermelon cultivars with limited genetic diversity. J Amer Soc Hort Sci 129: 553-558.

Levi A, Thomas CE, Wehner TC, Zhang X (2001). Low genetic diversity indicates the need to broaden the genetic base of cultivated watermelon. Hortic Sci 36: 1096-1101.

Mohammadi SA, Prasanna BM (2003). Analysis of genetic diversity in crop plants-salient statistical tools and considerations. Crop Sci 43: 1235-1248.

Mumm RH, Dudley JW (1994). A classification of 148 U.S. maize inbreds: I. Cluster analysis based on RFLPs. Crop Sci 34: 842-851.

Navot N, Zamir D (1987). Isozyme and seed protein phylogeny of genus *Citrullus* (*Cucurbitaceae*). Plant Syst Evol 156: 61-67.

Ocal N, Akbulut M, Gulsen O, Yetisir H, Solmaz I, Sari N (2014). Genetic diversity, population structure and linkage disequilibrium among watermelons based on peroxidase gene markers. Sci Hortic-Amsterdam 176: 151-161.

Pritchard KJ, Stephens M, Donelly P (2000). Inference of population structure using multilocus genotype data. Genetics 155: 945-959.

Sain RS, Joshi P, Sastry EVD (2002). Cytogenetic analysis of interspecific hybrids in genus *Citrullus* (*Cucurbitaceae*). Euphytica 128: 205-210.

Sari N, Tan A, Yanmaz R, Yetisir H, Balkaya A, Solmaz I, Aykas L (2008). General status of cucurbit genetic resources in Turkey. In: Pitrat M, editor. Proceedings of IX. EUCARPIA Meeting on Genetics and Breeding of *Cucurbitaceae*. Paris, France: INRA, pp. 21-32.

Solmaz I, Sari N (2009). Characterization of watermelon (*Citrullus lanatus*) accessions collected from Turkey for morphological traits. Genet Res Crop Evol 56: 173-188.

Solmaz I, Sari N, Aka-Kacar Y, Yalcin Y (2010). The genetic characterization of Turkish watermelon (*Citrullus lanatus*) accessions using RAPD markers. Genet Resour Crop Evol 57: 763-771.

Uluturk ZI, Frary A, Doganlar S (2011). Determination of genetic diversity in watermelon [*Citrullus lanatus* (Thunb.) Matsum & Nakai] germplasm. Austral J Crop Sci 5: 1832-1836.

Wang P, Li Q, Hu J, Su Y (2015). Comparative analysis of genetic diversity among Chinese watermelon germplasms using SSR and SRAP markers, and implications for future genetic improvement. Turk J Agric For 39: 322-331.

Wehner T (2008). Watermelon. In: Prohens J, Nuez F, editors. Handbook of Plant Breeding, Vegetables I. New York, NY, USA: Springer, p. 418.