Genetic variability of watermelon accessions based on microsatellite markers

R.N.C. de S. Gama[1], C.A.F. Santos[2] and R. de C.S. Dias[2]

[1]Programa de Pós-Graduação em Rercursos Genéticos Vegetais,
Universidade Estadual de Feira de Santana, Feira de Santana, BA, Brasil
[2]Laboratório de Genética Vegetal, Embrapa Semiárido, Petrolina, PE, Brasil

Corresponding author: R.N.C. de S. Gama
E-mail: renata.natalia@hotmail.com

ABSTRACT. We analyzed the genetic variability of 40 watermelon accessions collected from 8 regions of Northeastern Brazil using microsatellite markers, in order to suggest strategies of conservation and utilization of genetic variability in this species. These accessions are not commercial cultivars. They were sampled in areas of traditional farmers that usually keep their own seeds for future plantings year after year. An UPGMA dendrogram was generated from a distance matrix of the Jaccard coefficient, based on 41 alleles of 13 microsatellite loci. Analysis of molecular variance was made by partitioning between and within geographical regions. The similarity coefficient between accessions ranged from 37 to 96%; the dendrogram gave a co-phenetic value of 0.80. The among population genetic variability was high ($\hat{\phi}_{ST}$ = 0.319). Specific clusters of accessions sampled in 3 regions of Maranhão were observed while the other 5 regions did not presented specific clusters by regions. We conclude that watermelon genetic variability is not uniformly dispersed in the regions analyzed, indicating that geographical barriers or edaphoclimatic conditions have limited open mating. We suggest sampling a greater number of

populations, so regional species diversity will be better represented and preserved in the germplasm bank.

INTRODUCTION

The watermelon (*Citrullus lanatus*) belongs to the Cucurbitaceae family and is native to Africa (Whitaker and Davis, 1962). It was introduced to Brazil in two phases, the first during the slave trade and the second when bred cultivars were introduced from breeding programs in the United States and Japan (Romão et al., 2008).

An important step in cultivar development is studying the genetic variability found in genetic resources. The use of genetic resources to create new varieties is important for obtaining higher yields and for the technological transformations required for modernization of agribusiness. It is a dynamic process, but requires continuous enrichment and characterization of the materials maintained in germplasm collections (Valls, 2007).

The Brazilian Northeast is a secondary diversity center for the watermelon, with great variability in several plant, fruit, and disease-resistance traits (Queiróz, 1993; Dias et al., 1996). The variability in this region served as the basis of an Active Cucurbitaceae Germplasm Bank (BAGC) for Northeastern Brazil in the 1980s, located at Embrapa Semiarid, Petrolina - PE (Queiróz, 1993). The Watermelon Germplasm Bank (BGCIA) is a component of the BAGC for northeastern Brazil and has about 870 conserved accessions (Silva et al., 2010) including commercial cultivars, samples from traditional agriculture, and some wild watermelon relatives. Many studies on the use of watermelon genetic resources have utilized accessions preserved in the BGCIA based on morphological descriptions, which have allowed great advances in describing genetic divergence (Queiróz, 1993; Dias et al., 1996; Silveira et al., 2009).

Molecular information can complement ecological, morphological, and agronomic information on genetic resources; increase the efficiency of collection processes; direct enrichment of the genetic base; help form and validate nuclear and study collections; reveal genetic diversity and purity; identify duplicate and redundant accessions; facilitate botanical and phylogenetic classification studies; subsidize parent selection, and help plan crossing and selection of genotypes with desired characteristics in breeding programs (Faleiro, 2007).

Silva et al. (2007) reported quantitative trait analysis of variance and wide genetic variability within and between the BGCIA accessions, highlighting the importance of this germplasm bank for watermelon breeding purposes. Capeloto et al. (2004) studied genetic divergence within and between 18 watermelon accessions collected in Maranhão State with 59 RAPD (random amplified polymorphic DNA) primers and concluded, based on clustering analysis, that there was considerable divergence among and between accessions.

RAPD markers have the disadvantages of being dominant and difficult to reproduce (Esselink et al., 2003). In contrast, microsatellite markers, also called SSR (simple sequence repeats), have been the best markers for fingerprinting studies because of their polymorphic character, co-dominance, reliability, and reproducibility (Varshney et al., 2005). In spite of these benefits, SSR markers have not been used to analyze genetic variation in the BGCIA accessions.

Mujaju et al. (2010) revealed greater within-accession variability in wild Zimbabwean watermelon, based on molecular analysis of variance (AMOVA) and RAPD and SSR data

clustering. Studies by Jarret et al. (1997), Guerra-Sanz (2002), and Joobeur et al. (2006) revealed genetic variability in watermelon accessions from the formation of clusters with SSR molecular data. Cluster analyses are only indications of the existence of diversity among populations and the $\hat{\phi}_{ST}$ parameter, similar to the F coefficient obtained in AMOVA, can be used as a measure of population differentiation (Excoffier et al., 1992).

The objective of this study was to analyze the genetic variability of 40 accessions from 8 regions of the Brazilian Northeast and preserved in the BGCIA at Embrapa Semiarid, based on microsatellite markers to supply information for preservation strategies and use of the genetic variability of this species.

MATERIAL AND METHODS

Plant material

Watermelon accessions were selected from 8 regions of the Brazilian Northeast and five accessions were assessed per region (Table 1). Seeds, from the Cucurbitaceae Active Germplasm Bank for the Brazilian Northeast, sampled in areas of traditional farmers, were sown on polystyrene trays containing a commercial plant substrate based on vermiculite and plant ashes, and kept in a greenhouse. Leaf samples were collected from 5 plants after the appearance of the third definitive leaf to represent each accession. Samples were stored in a freezer at -80°C prior to DNA extraction.

Table 1. Sampling area of the accessions of the Watermelon Germplasm Bank (BGCIA) of Northeastern Brazilian genotyped with 13 microsatellite loci.

BGCIA code	Collection area	BGCIA code	Collection area
BGCIA 30	Xique-Xique - BA	BGCIA 357	Lagoa Grande - PE
BGCIA 31	Xique-Xique - BA	BGCIA 358	Lagoa Grande - PE
BGCIA 32	Xique-Xique - BA	BGCIA 359	Lagoa Grande - PE
BGCIA 33	Xique-Xique - BA	BGCIA 376	Lagoa Grande - PE
BGCIA 37	Xique-Xique - BA	BGCIA 375	Lagoa Grande - PE
BGCIA 51	Itaguaçu - BA	BGCIA 416	São Luís - MA
BGCIA 52	Itaguaçu - BA	BGCIA 417	São Luís - MA
BGCIA 53	Itaguaçu - BA	BGCIA 419	São Luís - MA
BGCIA 54	Itaguaçu - BA	BGCIA 422	São Luís - MA
BGCIA 55	Itaguaçu - BA	BGCIA 424	São Luís - MA
BGCIA 127	Juazeiro - BA	BGCIA 469	Colinas - MA
BGCIA 129	Juazeiro - BA	BGCIA 470	Colinas - MA
BGCIA 130	Juazeiro - BA	BGCIA 471	Colinas - MA
BGCIA 131	Juazeiro - BA	BGCIA 472	Colinas - MA
BGCIA 134	Juazeiro - BA	BGCIA 484	Colinas - MA
BGCIA 295	Jacobina - BA	BGCIA 511	Codó - MA
BGCIA 296	Jacobina - BA	BGCIA 525	Codó - MA
BGCIA 297	Jacobina - BA	BGCIA 552	Codó - MA
BGCIA 298	Jacobina - BA	BGCIA 554	Codó - MA
BGCIA 299	Jacobina - BA	BGCIA 559	Codó - MA

BA = Bahia, PE = Pernambuco, and MA = Maranhão States.

DNA extraction and quantification

DNA was extracted using the 2% CTAB method (Doyle and Doyle, 1990), modified to 7500 and 10,000 rpm, in the first and second centrifuges, respectively; 2% β-mercaptoethanol

and incubation at 60°C for 30 min. After adding Tris-EDTA buffer, the solution was treated with RNAse to remove co-extracted RNAs. DNA was quantified and its integrity verified by electrophoresis on 0.8% agarose gel. The genomic DNA was diluted to 50 ng/µL and stored at -20°C.

DNA reaction amplification and resolution on polyacrylamide gels

Thirty-six microsatellite primers were used (Joobeur et al., 2006). The primer pairs were assessed using 4 cultivars to select those with the best polymorphic resolution on polyacrylamide gel. The DNA fragments were amplified by polymerase chain reaction (PCR) as described by Joobeur et al. (2004), with modifications: a final volume (12 µL) containing 50 ng DNA, 5 µM of each primer, 0.1 mM dNTP, 1.5 mM $MgCl_2$, 1X PCR buffer, and 0.75 U Taq DNA polymerase. Cycling conditions were as follows: a) a first cycle of 94°C for 2 min, followed by 30 cycles at 94°C for 15 s, 56°C for 30 s, and 72°C for 2 min, with a final cycle at 72°C for 30 min.

Amplification products were separated on 6% polyacrylamide gel, prepared on a 60-well sandwich-type glass plate. A pre-run of 30 min at 40 W was carried out before applying the PCR samples. A 2.5-µL aliquot of the denatured PCR product was electrophoresed on 6% polyacrylamide gel for approximately 3 h, at a constant 40-W power. The 50-bp molecular weight marker ladder (Fermentas) was loaded on the side edge of each gel. The gels were stained with silver nitrate as described by Creste et al. (2001).

Microsatellite data notation and analysis

Loci that presented easily interpreted polymorphisms were selected for genotyping of the 17 watermelon cultivars. The microsatellites were noted for allele presence (1) and absence (2) to construct a similarity matrix of the Jaccard index. The dendrogram with the same coefficient of similarity as the accessions was generated by the unweighted pair group method with arithmetic means grouping method. The phenogram fit was assessed by the cophenogram correlation; that is, the correlation between real- and graphically represented distances. The NTSYSpc: 2.1 version (Rophlf, 2000) computer application was used for these analyses. AMOVA was carried out by the total partitioning of components between and within populations (regions) using the squared distances according to Excoffier et al. (1992), using Excel Macro-GENALEX 6.2 (Peakall and Smouse, 2006). The significance of these estimates was obtained by the randomization method with 1000 permutations.

RESULTS

Only 13 of the 36 pairs of primers provided easily interpretable polymorphic amplifications: MCPI_04, MCPI_05, MCPI_07, MCPI_09, MCPI_11, MCPI_12, MCPI_14, MCPI_16, MCPI_21, MCPI_26, MCPI_29, MCPI_33, and MCPI_39.

Forty-one alleles were detected on these 13 microsatellites; the number of alleles per locus ranged from 2 to 5 (mean 3.15) in the 40 genotyped watermelon accessions.

PIC values, which reflect allele diversity, were not uniform among the tested microsatellite loci. The PIC mean was 0.391, with maximum and minimum values observed at loci MCPI_12 (0.683) and MCPI_14 (0.186), respectively (Table 2).

Table 2. Genetic parameters estimated for 13 microsatellites in 40 watermelon accessions of the Watermelon Germplasm Bank (BGCIA) of Northeastern Brazilian.

SSR	Major allele frequency	Genotype No.	Simple size (bp)	Allele No.	Gene diversity	Heterozygosity	PIC
MCPI_04	0.488	3	40	4	0.557	1.000	0.459
MCPI_05	0.700	9	40	5	0.477	0.275	0.445
MCPI_07	0.825	3	40	2	0.289	0.250	0.247
MCPI_09	0.775	6	40	3	0.366	0.275	0.326
MCPI_11	0.513	6	40	3	0.618	0.375	0.548
MCPI_12	0.321	9	40	5	0.731	0.410	0.683
MCPI_14	0.888	3	40	3	0.202	0.200	0.186
MCPI_16	0.718	6	40	4	0.439	0.410	0.392
MCPI_21	0.667	4	40	3	0.481	0.556	0.414
MCPI_26	0.654	3	40	2	0.453	0.385	0.350
MCPI_29	0.588	3	40	2	0.485	0.475	0.367
MCPI_33	0.500	4	40	3	0.512	0.375	0.393
MCPI_39	0.788	3	40	2	0.335	0.325	0.279
Mean	0.648	4.769	40	3.154	0.457	0.409	0.391

SSR = simple sequence repeats; PIC = polymorphism information content.

The heterozygosity mean was 0.409; the highest value was at MCPI_04 (1.00), the lowest at MCPI_14 (0.200) (Table 2).

The correlation between the co-phenotypic and similarity distance matrices was 0.80, indicating that the dendrogram (Figure 1) was a good fit in the watermelon accession cluster with 41 alleles of the 13 microsatellite loci. Between-accession similarity ranged from 37 to 96%. The 41 alleles of the 13 microsatellite loci were sufficient to separate all 40 watermelon accessions. The BGCIA 297 and BGCIA 299 accessions collected in Jacobina, BA, were the most similar (96%).
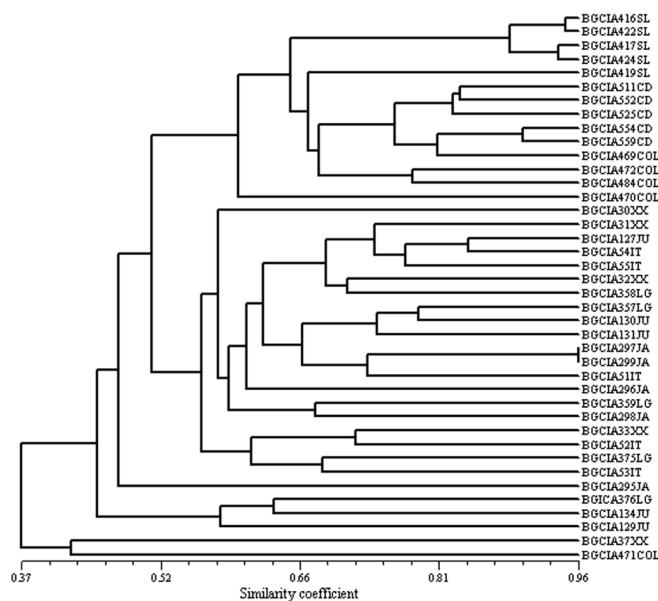

Figure 1. UPGMA dendrogram of the coefficient of similarity of 40 watermelon accessions of the Watermelon Germplasm Bank (BGCIA) of Northeastern Brazilian genotyped for 41 alleles of 13 microsatellite loci. Cophenotypic correlation (r) = 0.80. SL = São Luís; XX = Xique-Xique; LG = Lagoa Grande; IT = Itaguaçu; JU = Juazeiro; JA = Jacobina; CD = Codó; COL = Colinas.

Specific clustering was observed for the 3 regions of Maranhão (São Luís, Codó and Colinas). The accessions from these regions were placed continuously on the dendrogram with similarity between 61 and 94%. Although great similarity was observed for some accession pairs in other regions, no specific groupings were observed by sample location (Figure 1). The estimate of variation among ecoregions was high at 0.319 ($\hat{\phi}_{ST}$), suggesting the watermelon has restricted flow, less than one migrant per generation ($N_m = 0.534$) (Table 3).

Table 3. Analysis of molecular variance in 40 watermelon accessions sampled in 8 regions and belonging to the Watermelon Germplasm Bank (BGCIA) of Northeastern Brazilian.

Sources of variation	d.f.	SS	MS	Total variation (%)	Statistic $\hat{\phi}_{ST}$	P	N_m
Among populations	7	99.85	14.264	32	$\hat{\phi}_{ST} = 0.319$	<0.001	0.534
Within populations	32	136.6	4.269	68	$1-\hat{\phi}_{ST} = 0.681$	<0.001	-
Total	39	236.45	-	100	-	-	-

d.f. = degrees of freedom; SS = sum of squares; MS = mean squares; P = probability based on 1000 permutations; $N_m = [(1 - \hat{\phi}_{ST}) / (4\hat{\phi}_{ST})]$.

DISCUSSION

The number of microsatellite markers capable of detecting polymorphisms was lower than that reported by Joobeur et al. (2006), who assessed the capacity to detect polymorphisms with 36 primer pairs and reported that all were polymorphic in 8 watermelon accessions, including 5 *C. lanatus* var. *lanatus* and 3 *C. lanatus* var. *citroides* accessions. This decrease in the number of SSR polymorphic markers may have been because we assessed only watermelon accessions belonging to the *C. lanatus* var. *lanatus* subspecies.

The mean allele number was consistent with reports from Guerra-Sanz (2002), Joobeur et al. (2006), and Kwon et al. (2010).

PIC and heterozygosity represent variability (Weir, 1996); where there is greater variability, with greater heterozygosis and PIC, microsatellites are best for diversity studies. The PIC and heterozygosity values reported in this study indicate that the chosen microsatellites have great differentiation power.

The similarity values reported here reflect the high genetic variability of the studied populations. According to Silva et al. (2006), producer selection made tends to fix different types in certain regions, as observed in the accessions from Maranhão, with the appearance of specific groups.

The estimate of variation among ecoregions (Table 3) was high. High $\hat{\phi}_{ST}$ values for *Jatropha curcas*, a naturally occurring species of the Brazilian Semiarid region, were reported by Santos et al. (2010). Djè et al. (2010) analyzed 80 individuals from 4 African watermelon accessions using SSR markers and AMOVA and found low between-accession variation ($\hat{\phi}_{ST} = 0.09$) and high within-accession variation ($\hat{\phi}_{ST} = 0.91$). This low variability may have been due to the small number of accessions analyzed by these authors.

Although watermelon is an open pollinating species, genetic flow between populations was small, probably because of human disturbance in the study areas. Traditional watermelon cropping usually occurs in small areas where the endogamy process is favored; thus, genetic drift can lead to greater divergence between plantations in the same region (Queiróz, 1993).

Artificial selection by producers within a region can also increase variation, with the producers selecting their own seeds for future planting, according to their own criteria (Queiróz, 2004). Furthermore, exchange of seeds between producers when their fields are situated very close to each other, along with the fact that watermelon is an open-pollinating plant, increase the possibility of genetic recombination between selections, widening the genetic variability even in samples collected from the same location (Romão, 2000; Capeloto et al., 2004).

Dendrogram results (Figure 1) and AMOVA (Table 3) suggested that watermelon genetic variability is not uniformly dispersed in the regions analyzed, indicating that geographical barriers or edaphoclimatic conditions have limited open mating. In this scenario, we suggest sampling a greater number of populations, so regional species diversity will be better represented and preserved in the germplasm bank.

ACKNOWLEDGMENTS

REFERENCES

Capeloto A, Uneda SHT and Mauro AO (2004). Caracterização Molecular Entre e Dentro de Acessos de Melancia Através de RAPD - PCR. Available at [http://www.revista.inf.br/agro05/artigos/artigo01.pdf]. Accessed May 6, 2011.

Creste S, Tulmann Neto A and Figueira A (2001). Detection of single sequence repeat polymorphisms in denaturing polyacrylamide sequencing gels by silver staining. *Plant Mol. Biol. Rep.* 19: 299-306.

Dias R de CS, Queiróz MA and Menezes M (1996). Fontes de resistência em melancia a *Didymella bryonia*e. *Hortic. Bras.* 14: 15-18.

Djè Y, Tahi CG, Zoro BI AI, Baudoin JP, et al. (2010). Use of ISSR markers to assess genetic diversity of African edible seeded *Citrullus lanatus* landraces. *Sci. Hort.* 124: 159-164.

Doyle JJ and Doyle JL (1990). Isolation of plant DNA from fresh tissue. *Focus* 12: 13-15.

Esselink GD, Smulders MJ and Vosman B (2003). Identification of cut rose (*Rosa hybrida*) and rootstock varieties using robust sequence tagged microsatellite site markers. *Theor. Appl. Genet.* 106: 277-286.

Excoffier L, Smouse PE and Quattro JM (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131: 479-491.

Faleiro FG (2007). Marcadores Genético-Moleculares Aplicados a Programas de Conservação e Uso de Recursos Genéticos. Embrapa Cerrados, Planaltina.

Guerra-Sanz JM (2002). *Citrullus* simple sequence repeats markers from sequence databases. *Mol. Ecol. Notes* 2: 223-225.

Jarret RL, Merrick LC, Holms T, Evans J, et al. (1997). Simple sequence repeats in watermelon (*Citrullus lanatus* (Thunb.) Matsum. & Nakai). *Genome* 40: 433-441.

Joobeur T, King JJ, Nolin SJ, Thomas CE, et al. (2004). The *Fusarium* wilt resistance locus Fom-2 of melon contains a single resistance gene with complex features. *Plant J.* 39: 283-297.

Joobeur T, Gusmini G, Zhang X, Levi A, et al. (2006). Construction of a watermelon BAC library and identification of SSRs anchored to melon or *Arabidopsis* genomes. *Theor. Appl. Genet.* 112: 1553-1562.

Kwon YS, Oh YH, Yi SI, Kim HY, et al. (2010). Informative SSR markers for commercial variety discrimination in watermelon (*Citrullus lanatus*). *Genes Genomics* 32: 115-122.

Mujaju C, Sehic J, Werlemark G, Garkava-Gustavsson L, et al. (2010). Genetic diversity in watermelon (*Citrullus lanatus*) landraces from Zimbabwe revealed by RAPD and SSR markers. *Hereditas* 147: 142-153.

Peakall R and Smouse PE (2006). GENALEX 6: Genetic analysis in excel. Population genetic software for teaching and research. *Mol. Ecol. Notes* 6: 288-295.

Queiróz MA (1993). Potencial do germoplasma de Cucurbitáceas no Nordeste Brasileiro. *Hortic. Bras.* 11: 7-9.

Queiróz MA (2004). Germplasm of Cucurbitaceae in Brazil. *Crop Breed. Appl. Biotechnol.* 4: 377-383.

Romão RL (2000). Northeast Brazil: a secondary center of diversity fo watermelon (*Citrullus lanatus*). *Genet. Resour. Crop Evol.* 47: 207-213.

e44
444444444

Romão RL, Assis JGA and Queiróz MA (2008). Melancia. In: Origem e Evolução de Plantas Cultivadas (Barbieri RL and Stumpf ERT, eds.). Embrapa Informação Tecnológica, Brasília, 555-573.

Rophlf (2000). NTSYSpc Numerical Taxonomy and Multivariate Analysis System. Version 2.10m. Exeter Software, Setauket.

Santos CAF, Drumond MA, Rodrigues MA and Evangelista MRV (2010). Genetic similarity of *Jatropha curcas* accessions based on AFLP markers. *Crop Breed. Appl. Biotechnol.* 10: 364-369.

Silva AF, Santos CAF, Araújo FP, Lima Neto FP, et al (2010). Recursos Genéticos Vegetais Conservados na Embrapa Semiárido. In: Semiárido Brasileiro: Pesquisa: Pesquisa, Desenvolvimento e Inovação (Sá IB and Silva PCG, eds.). Embrapa Semiárido, Petrolina, 274-315.

Silva ML, Queiróz MA de, Ferreira MAJF and Buso GSC (2006). Caracterização morfológica e molecular de acessos de melancia. *Hortic. Bras.* 24: 405-409.

Silva ML, Queiróz MA de, Ferreira MAJF and Aragão CA (2007). Variabilidade genética de acessos de melancia coletados em três regiões do estado da Bahia. *Caatinga* 20: 93-100.

Silveira LM, Queiróz MA, Lima JAA, Nascimento AKQ, et al. (2009). Levantamento sorológico de vírus em espécies de cucurbitáceas na região do submédio São Francisco, Brasil. *Trop. Plant Pathol.* 34: 123-126.

Valls JFM (2007). Caracterização de Recursos Genéticos Vegetais. In: Recursos Genéticos Vegetais (Nass LL, ed.). Embrapa Recursos Genéticos e Biotecnologia, Brasília, 282-305.

Varshney RK, Graner A and Sorrells ME (2005). Genic microsatellite markers in plants: features and applications. *Trends Biotechnol.* 23: 48-55.

Weir BS (1996). Genetic Data Analysis II: Methods for Discreate Population Genetic Data. Sinauer Associates, Sunderland.

Whitaker TW and Davis GN (1962). Cucurbits: Botany, Cultivation and Utilization. Interscience Publishers, New York.