A Genome-Wide Scan of Selective Sweeps and Association Mapping of Fruit Traits Using Microsatellite Markers in Watermelon


From the Department of Biology, Gus R. Douglass Institute, West Virginia State University, Institute, WV 25112-1000 (Reddy, Abburi, Saminathan, Cantrell, Vajja, Reddy, Tomason, and Nimmakayala); the U.S. Vegetable Laboratory, USDA, ARS, 2875 Savannah Highway, Charleston, SC 29414 (Levi); and the Department of Horticultural Science, North Carolina State University, Raleigh, NC 27695-7609 (Wehner).

*These authors contributed equally to the work.

Author correspondence to Umesh K. Reddy at the address above, or e-mail: ureddy@wvstateu.edu.

### Abstract

Our genetic diversity study uses microsatellites of known map position to estimate genome level population structure and linkage disequilibrium, and to identify genomic regions that have undergone selection during watermelon domestication and improvement. Thirty regions that showed evidence of selective sweep were scanned for the presence of candidate genes using the watermelon genome browser ([www.icugi.org](http://www.icugi.org)). We localized selective sweeps in intergenic regions, close to the promoters, and within the exons and introns of various genes. This study provided an evidence of convergent evolution for the presence of diverse ecotypes with special reference to American and European ecotypes. Our search for location of linked markers in the whole-genome draft sequence revealed that BVWS00358, a GA repeat microsatellite, is the GAGA type transcription factor located in the 5′ untranslated regions of a structure and insertion element that expresses a Cys₂His₂ Zinc finger motif, with presumed biological processes related to chitin response and transcriptional regulation. In addition, BVWS01708, an ATT repeat microsatellite, located in the promoter of a DTW domain-containing protein (Cla002761); and 2 other simple sequence repeats that association mapping link to fruit length and rind thickness.

### Subject areas:

Genomics and gene mapping; Quantitative genetics and Mendelian inheritance

### Key words:

domestication, heterozygosity, linkage disequilibrium, population structure, recombination rate

Watermelon [*Citrullus lanatus* (Thunb.) Matsum. and Nakai] is a specialty crop with a genome size of 425 Mb (*2n* = 2× = 22). The cultivated dessert watermelon flesh contains 87–90% water, a high sugar content (8–13%), and several important health-related compounds, including lycopene, citrulline, arginine, and glutathione ([Ren et al. 2012](#); [Nimmakayala et al. 2014a](#)). *Citrullus lanatus* subsp. *mosospermus* represent the “egusi” watermelon group, and have large edible seeds with a fleshy pericarp; *C. lanatus* subsp. *vulgatis* represent the red sweet dessert type ([Fursa 1972](#); [Jeffrey 2001](#)). Although genetic diversity within subsp. *vulgatis* is extremely narrow, the cultivars are phenotypically diverse in fruit shape, flesh texture and color, fruit weight, soluble solids, and rind thickness ([Levi 2004](#)).

Breeding watermelon of fairly similar backgrounds has resulted in a narrow bottleneck of genetic diversity ([Levi et al. 2001](#)). The molecular diversity of watermelon cultivars is reflected in the low number of single nucleotide polymorphisms (SNPs) between the American heirloom cultivar Charleston Gray and the Chinese elite line 97103, which show an average of 1 SNP every 1430 bp ([Guo et al. 2013](#)). Our previous study confirmed the narrow genetic diversity among American watermelon accessions ([Nimmakayala et al. 2014a](#)). One explanation for this narrow diversity is the...
founder effect created by a small number of accessions being brought to a continent or region as people travel (Dane and Liu 2007; Zoltán 2007).

Excavations in Egypt and Libya have indicated that northern Africa is a primary center of domestication for watermelon (Wasylikowa and van der Veen 2004). Watermelon may have entered Europe around 512 AD, when the Moors invaded the Iberian peninsula, or during the Crusades (Zoltán 2007). In India and China, watermelon was introduced around 800 and 1100 AD, respectively (Paris et al. 2013). Watermelon cultivars were introduced into the Americas after the second voyage of Columbus and during the time of the slave trade and colonization (Romão 2000; Zoltán 2007; Paris et al. 2013).

Genetic analysis of this valuable crop involves several critical steps that include accurate phenotyping for various fruit traits along with high-throughput genotyping and sequencing (Guo et al. 2013; Ren et al. 2012). Simple sequence repeats (SSRs), also known as microsatellites, are repeats of nucleotides consisting of a variable number of repeat units that produce codominant, multiallelic, reproducible bands upon amplification (Stägel et al. 2008; Parida et al. 2009; Cavagnaro et al. 2010; Sun et al. 2013; Nimmakayala et al. 2014b). SSR markers are transferrable across Citrullus spp. and can be used in distantly related taxa (Jarret et al. 1997). Microsatellites have been used for the construction of genetic maps in many plant species, and provide dependable landmarks throughout the genome (Córdoba et al. 2010; Cavagnaro et al. 2011; Ren et al. 2012). Hamblin et al. (2007) concluded that SSRs were more effective in clustering germplasm into populations than SNPs, by comparing analyses based on 89 SSRs to analyses based on 847 SNPs, covering 554 SNP haplotypes, in individuals from the same 259 inbred maize lines. An ancestry matrix (Q matrix) calculated using SSRs can be used along with SNP genotypes in an association mapping model to reduce spurious associations (Nimmakayala et al. 2014b).

Currently, the most genome-wide map for watermelon involved 698 SSRs, 219 insertions or deletions, and 36 structural variants, covering 800 cM of the genome with a mean marker interval of 0.8 cM (Ren et al. 2012). The map positions 234 watermelon genome sequence scaffolds, accounting for 93.5% of the assembled 353 Mb genome size (Ren et al. 2012). More recently this map was integrated with another 386 SNP markers that were previously mapped by Sandlin et al. (2012). The additional SNP markers further reduced the mean marker interval to 0.6 cM (Ren et al. 2014). Fifty-eight previously reported quantitative trait loci (QTL) for 12 traits (Sandlin et al. 2012) were integrated into the map, and 10 new QTLs for sugar content were identified (Ren et al. 2014). Notably, some of the traits for which QTLs mapped were Brix, fructose, sucrose, glucose, fruit weight, fruit length, fruit width, fruit shape index, rind thickness, 100 seed weight, seed length, seed width, seed oil percentage, and egusi loci. This map is useful for locating recombination cold spots and the distribution of segregation distortion. Linkage disequilibrium (LD) is nonrandom association of alleles across the chromosomes and therefore features recombination cold spots. An examination of population structure, genetic diversity at the molecular level, and LD distribution across various chromosomes allows for the identification of QTLs through the model-based association mapping of various traits (Nimmakayala et al. 2014b). A unified, mixed-model approach for association mapping combined with a population structure (Q) analysis is a dependable and robust system for identifying reliable QTLs (Yu and Buckler 2006; Zhao et al. 2007).

Our current genetic diversity study uses microsatellites of known map position to look for evidence of selection and to identify genomic regions that have undergone selection during watermelon’s domestication and improvement. We also estimate genome level population structure and LD across the linkage groups, and perform association mapping for fruit traits in the watermelon.

Materials and Methods

Our study used 96 watermelon accessions collected across a wide geographic sample of the world to maximize molecular diversity (Supplementary Table S1 online). The plants from individual accessions were grown in 3 randomized plot replicates during 2 seasons (2012 and 2013) with a row-to-plant spacing of 8 × 3 ft. Standard management practices were followed throughout the growing season. Fruit length and diameter, rind thickness, and soluble solids were recorded for 5 fruits in each replicate.

Marker Resources

We used 201 microsatellites, all with known positions on the watermelon genetic and physical map. Details related to the microsatellite map locations and positions obtained from Ren et al. (2012) are presented in Supplementary Tables S2 and S3 online.

DNA Isolation, PCR, and Resolution of SSR Alleles

Genomic DNA isolation was done with a plant DNA isolation kit (Qiagen cat.#69104). The resolution of SSR polymorphisms on a high-throughput DNA fragment analyzer (AdvanCE FS; Advanced Analytical Technologies, Ames, IA) was carried out as per Nimmakayala et al. (2014a).

Statistical Analysis

Data were processed using statistical analysis of variance (ANOVA) models for all fruit traits. Structure 2.2 (Pritchard et al. 2000) was used to cluster individuals into K groups by estimating the posterior probability of the data for a given K, Pr(X|K). The number of populations (K) was determined using an admixture model with correlated alleles, and K = 2–10. Five independent runs of 100 000 Markov Chain Monte Carlo generations, after 100 000 generation burn-in periods, were used to estimate each value of K. The optimal K value was determined using the ad hoc statistic, ΔK (Evanno et al. 2005). The number of ΔKs in each dataset was
evaluated using $\Delta K$ values estimated with Structure Harvester (http://taylor0.Biology.ucla.edu/structureHarvester/), a program implementing the Evanno method for visualizing Structure output (Earl and vonHoldt 2012). Genetic distance estimates were calculated as described by Crossa and Franco (2004). The neighbor-joining algorithm was used to build a dendrogram based on genetic distances with the software MEGA 5.0 (Tamura et al. 2011). Genetic variance was partitioned among the watermelon groups identified by molecular diversity, and population structure was determined by analysis of molecular variance (AMOVA), using the Arlequin 2.0 program (Schneider 2000). Molecular genetic diversity was estimated using genetic diversity and heterozygosity ($h$). Estimation of $F_{ST}$ (the proportion of genetic variance in a subpopulation relative to the total population) and $F_{IS}$ (the inbreeding coefficient) was based on Wright's F-statistics (Weir and Cockerham 1984) with the program PopGene 1.31 (Yeh 1999). The Ewens–Watterson neutrality test (Watterson 1978) was performed using Manly’s algorithm (Manly 1985) as implemented in PopGene 1.31 to investigate neutrality across the linkage groups. The $P$ matrix for first 3 principal components was calculated from 795 SSR alleles with the TASSEL 3.0 package (http://www.maizegenetics.net) (Bradbury et al. 2007). LD was estimated as the correlation between all pairs of SSRs with 1000 permutations, also using TASSEL 3.0. The $Q$ matrix was adapted from the K-5 cluster for association mapping to control spurious results arising from population stratification. The generalized linear model (GLM) of TASSEL 3.0 was used for association mapping. The linked marker $P$ values obtained from our association study underwent false discovery rate (FDR) analysis (Benjamini and Hochberg 1995).

Results

Morphological Traits

The 96 accessions in this research included 90 sweet watermelons (var. vulgaris) and 6 egusi types (var. mucospermus), cultivated for their edible seeds rather than fruit. Egusi types are generally known for their hard rinds, firm flesh, and low total soluble solids, unlike the sweet watermelon cultivars (var. vulgaris). ANOVA results with various traits are presented in Supplementary Table S4 online.

Basic Properties of Microsatellite Alleles

Allele numbers for the 201 microsatellites varied across the accessions. There were 135, 94, and 128 usable loci, of which 100, 57, and 107 were polymorphic for accessions belonging to Africa, Asia, and the rest of the world (Europe, and North and South America), respectively. The total number of alleles amplified for the different microsatellites is shown in Supplementary Table S5. BVWS00117, located on chromosome 6, amplified 10 different alleles. A total of 795 alleles were amplified across the entire dataset. Three alleles per microsatellite were amplified, on average. Shannon indices ($SI$) are a measure of the level of polymorphism in microsatellites. The $SI$ of ours ranged from 0.06 SI for BVWS00309 to 1.36 SI for BVWS00117 (Supplementary Table S5 online).

Genetic Diversity

Molecular diversity analysis showed that mean genetic diversities were $0.26 \pm 0.13$, $0.24 \pm 0.12$, and $0.22 \pm 0.12$ for the accessions collected in Africa, Asia, and the rest of the world, respectively. The mean genetic diversity among the 90 vulgaris types and 6 mucospermus types was $0.343 \pm 0.12$. AMOVA results for the vulgaris and mucospermus types, combined and separately, are shown in Supplementary Tables S6a and S6b. Variance was lower between populations, rather than within populations, indicating wide introgression among all groups. Variance among vulgaris and mucospermus was 9.38, as compared with the groups from Asia, Africa, and the rest of the world, which showed a variance of 3.73. The overall $F_{ST}$ was 0.04 among the global collections studied.

The 5 clusters in the neighbor-joining tree contained mixtures of several African and Asian accessions (Figure 1). The branches to Sugar Baby (United States) and PI 379223 (a former Serbia and Montenegro accession) were the longest, compared with the rest, indicating occurrence of rapid genetic changes. The first, third, and fifth clusters were predominantly North American accessions with a mixture of some African and Asian types. The second cluster was mostly African types, with some mixture of Asian types. The fourth cluster contained accessions belonging to all types, which were evenly spread across the cluster. Population structure analysis based on model-based assumptions was used to estimate $K$-2 to $K$-10 clusters each with 3 iterations. The results were analyzed for mean ± SD LnP(K) and $\Delta K$ values as estimated with the Structure Harvester (Supplementary Figure S2). $K$-5 was the most appropriate cluster size for this population, with $\Delta K = 17.3$. A population structure ancestry chart using K-5 (Figure 2) shows accessions in pink and red that can be traced to Africa. Ancestry colored in green primarily occurred in Asia, whereas blue and yellow designate accessions with ancestors in the rest of the world (Figure 2). Sugar Baby, the US variety, grouped completely with yellow-colored ancestry accessions; and similarly PI 379223, the former Serbia and Montenegro accession, grouped with green-colored ancestry accessions. The distinctiveness of these varieties is corroborated in the neighbor-joining tree by the longer branches mentioned above. PI 508441 (Korea), PI 535947 (Cameroon), Sweet Princess (USA), PI 319212 (Egypt), PI 169261 (Turkey), and PI 357731 (another former Serbia and Montenegro accession) were grouped with completely red ancestry in our analysis.

Evidence of Selection Across the Genome

$F_{IS}$, $F_{ST}$, and $He$ (expected heterozygosity) were plotted across the physical maps of various chromosomes (Figure 3). Parts of the genome showed a pattern of higher $F_{IS}$ coupled with low $He$; this trend is considered a signature of selective sweep or purifying selection. Chromosomes 1, 2, 3, 4, 5, 6, 7,
9, 10, and 11 contained 4, 6, 2, 3, 4, 2, and 1 regions of selective sweep, respectively. Chromosome 10 showed a massive domain that had undergone neutral selection (low $F_{IS}$ and high $He$).

The 30 regions that we determined had undergone purifying selection were scanned for the presence of candidate genes using the watermelon genome browser. Selective sweep was noted in the intergenic regions, very close to the 5' UTR or promoter, and in the exons and introns of various genes. Important proteins identified included the ripening-related protein, basic helix-loop-helix protein, Cyclin D1, and cytochrome p450; playing varied roles in protein, nucleic acid, and ion binding, transcriptional regulation, and catalysis. Physical locations, gene IDs, and annotations of our results are presented in Table 1. Combining $F_{ST}$ test results with results from the Ewens–Watterson test enabled us to detect deviations from a neutral-equilibrium model: either as a deficit of genetic diversity relative to the scaffold size (below the curve of observed $F$), or as an excess of genetic diversity (above the curve) (Supplementary Figure S3).
Figure 2. Shared ancestry as revealed by population structure analysis across 96 diverse watermelon collections. Numbers correspond to the accessions in Supplementary Table S1. Numbers in parenthesis refers to (1) Africa; (2) Asia, and (3) Rest of the world (Europe, and North and South America).

Linkage Disequilibrium and Recombination Rate
A total 15 LD blocks of various sizes were identified across the 9 watermelon chromosomes. None of the blocks were located on chromosome 5 or 9. Genetic map distances in cM and the physical length of the various blocks are shown in Supplementary Table S7. Individual correlation values pertaining to marker associations are listed in Supplementary Table S8. Recombination rates estimated for the various LD blocks ranged from 0 to 7.02 (cM/Mb). In contrast, recombination rates at the ends of chromosomes ranged from 1.09 to 14.89, which indicates the presence of relative hot spots of recombination.

Association Mapping
Four linked markers were identified with the GLM approach. BVWS00358 and BVWS01708 were linked with fruit length across the seasons (Table 2). The BVWS01708 association was significant after our FDR correction during the 2013 season and still significant ($P = 0.01$) during the 2012 season. Additionally, BVWS00711 was associated with fruit width and length, but only during the 2013 season ($P = 0.02$, versus 0.05 for 2012). BVWS00681 was associated with rind thickness during the 2013 season ($P = 0.02$). Our search for the location of the linked markers in the watermelon whole-genome draft sequence showed that the BVWS00358, a GA repeat microsatellite, is a GAGA type transcription factor located in the 5' UTR of a structure and insertion element (Cla004026) that expresses a Cys$_2$His$_2$ (C2H2) Zinc finger motif, with presumed biological processes related to chitin response and transcriptional regulation. BVWS01708, an ATT repeat microsatellite, is located in the promoter of the DTW domain-containing protein (Cla002761). Another marker, BVWS00681 is located in the intergenic region between the genes Cla006870 (a nucleotide sugar transporter protein) and Cla06868 (a pentatricopeptide repeat-containing protein).

Discussion
Our study provides comprehensive insight into the population genetic structure of sweet and unsweet cultivated forms of watermelon. Convergent evolution may have been responsible for the formation of American ecotypes such as Sugar Baby and PI 379223, 2 cultivars that do not appear to share significant ancestry with any of the African or Asian types. The distinctness of American and Asian ecotypes has been previously described in detail by Ren et al. (2012) and Guo et al. (2013). Phenotypic convergence suggests independent evolutionary lineages with parallel phenotypic evolution (the appearance of the same trait in closely related or potentially interbreeding lineages), and that this has occurred many times during crop domestication (Gross and Olsen 2010). We provide evidence for the presence of ecotype divergence at the molecular level, with special reference to American (Sugar Baby) and European ecotypes (PI 379223).

We investigated diversity in watermelon using the SSRs that have known genetic and physical locations in the genome to identify genomic regions with evidence of selection during watermelon domestication and improvement. Because these SSRs are located at regular intervals covering the whole length of various chromosomes, hitchhiking effects with the loci under selection are minimal, in contrast to what would happen were a similar analysis done with random microsatellites that do not have mapping information.

A limited number of watermelon accessions were introduced from the progenitor pools of Africa to the rest of the world. Therefore, much of the genetic diversity in the progenitor was probably left behind, which resulted in genetic bottlenecks everywhere watermelon was introduced.
Figure 3. Mapping of $F_{IS}$ against heterozygosity and $F_{ST}$ along the length of the physical map for various chromosomes. Red bars indicate sites of selective sweep.
<table>
<thead>
<tr>
<th>Chromosome no.</th>
<th>Marker</th>
<th>Physical position of gene</th>
<th>Strand</th>
<th>Gene ID</th>
<th>Gene name</th>
<th>Insertion position</th>
<th>GO biological process</th>
<th>GO molecular function</th>
<th>GO cellular component</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>BVWS02029</td>
<td>8 480 391..8 481 329</td>
<td>+</td>
<td>Cla008509</td>
<td>GNS1/SUR4 membrane family protein</td>
<td>Intron</td>
<td>—</td>
<td>Long chain fatty acid elongation systems</td>
<td>—</td>
</tr>
<tr>
<td>2</td>
<td>BVWS02398</td>
<td>31 561 416..31 565 029</td>
<td>—</td>
<td>Cla008724</td>
<td>Ripening-related protein grip22</td>
<td>Intron</td>
<td>—</td>
<td>Lipoprotein A</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>BVWS02343</td>
<td>32 923 309..32 924 254</td>
<td>+</td>
<td>Cla008579</td>
<td>Phospholipase A22</td>
<td>Exon</td>
<td>Lipid catabolic process</td>
<td>Calcium ion binding; phospholipase A2 activity</td>
<td>—</td>
</tr>
<tr>
<td>4</td>
<td>BVWS00405</td>
<td>5 942 099..5 945 562</td>
<td>+</td>
<td>Cla011671</td>
<td>GTP-binding family protein</td>
<td>Intron</td>
<td>Protein transport, regulation of growth</td>
<td>—</td>
<td>Cytosol, nucleus, ER</td>
</tr>
<tr>
<td></td>
<td>BVWS01427</td>
<td>10 205 343..10 211 318</td>
<td>+</td>
<td>Cla006847</td>
<td>Nop25 superfamily protein</td>
<td>Introns</td>
<td>—</td>
<td>—</td>
<td>Nucleolus</td>
</tr>
<tr>
<td>5</td>
<td>BVWS02422</td>
<td>6 367 196..6 367 822</td>
<td>+</td>
<td>Cla021804</td>
<td>bHLH-like transcription factor</td>
<td>Exon</td>
<td>Transcriptional regulation</td>
<td>Transcriptional regulation</td>
<td>Nucleus</td>
</tr>
<tr>
<td></td>
<td>BVWS01953</td>
<td>19 362 579..19 365 571</td>
<td>—</td>
<td>Cla001042</td>
<td>Shikimate kinase</td>
<td>Intron</td>
<td>Amino acid biosynthesis</td>
<td>Kinase activity, magnesium ion binding</td>
<td>Plastid, chloroplast stroma, cytoplasm</td>
</tr>
<tr>
<td>6</td>
<td>BVWS02256</td>
<td>6 687 405..6 694 182</td>
<td>—</td>
<td>Cla009404</td>
<td>Haloacid dehalogenase-like hydrolase</td>
<td>Intron</td>
<td>Metabolic process</td>
<td>Catalytic activity</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>BVWS02455</td>
<td>9 730 451..9 732 498</td>
<td>+</td>
<td>Cla010544</td>
<td>Cytochrome P450 CYP714B3</td>
<td>Intron</td>
<td>fatty acid metabolic process, oxidation-reduction</td>
<td>electron carrier, heme binding, iron ion binding, monooxygenase, oxygen binding</td>
<td>Membrane, ER, Vacuole</td>
</tr>
<tr>
<td>9</td>
<td>BVWS00309</td>
<td>3 409 466..3 412 876</td>
<td>—</td>
<td>Cla015221</td>
<td>Cytochrome P450</td>
<td>Introns</td>
<td>Callose deposition, root development, oxidation/reduction</td>
<td>Electron carrier, iron ion binding, heme binding, protein binding</td>
<td>Membrane, ER, mitochondria</td>
</tr>
<tr>
<td></td>
<td>BVWS02323</td>
<td>32 706 023..32 708 090</td>
<td>—</td>
<td>Cla003249</td>
<td>CYCD1;1 (CYCLIN D1;1); cyclin-dependent protein kinase regulator</td>
<td>Exon</td>
<td>Response to sucrose, brassinosteroid and cytokinin stimulus; cell cycle</td>
<td>Protein binding</td>
<td>Nucleus</td>
</tr>
</tbody>
</table>
Table 2 Level of significance, marker effect and location of associated alleles are revealed by association mapping

<table>
<thead>
<tr>
<th>Year</th>
<th>Trait</th>
<th>Locus</th>
<th>Chromosome number</th>
<th>Genetic distance (cM)</th>
<th>Physical distance (Mb)</th>
<th>GLM P value</th>
<th>R²</th>
<th>Allelic sizes (bp)</th>
<th>Allelic effect</th>
<th>No. of genotypes with type of allele</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013</td>
<td>FL</td>
<td>BVWS00358</td>
<td>7</td>
<td>9.7</td>
<td>3.93</td>
<td>0.0303235</td>
<td>0.1018875</td>
<td>177:177</td>
<td>-7.371135503</td>
<td>16</td>
</tr>
<tr>
<td>2012</td>
<td>FL</td>
<td>BVWS00358</td>
<td>4</td>
<td>0.0</td>
<td>0.329</td>
<td>0.0118073</td>
<td>0.155797</td>
<td>278:278</td>
<td>-0.2166411</td>
<td>16</td>
</tr>
<tr>
<td>2012</td>
<td>FW</td>
<td>BVWS00711</td>
<td>10</td>
<td>24.9</td>
<td>6.99</td>
<td>0.0290696</td>
<td>0.1067577</td>
<td>175:175</td>
<td>-1.428738633</td>
<td>16</td>
</tr>
<tr>
<td>2013</td>
<td>RT</td>
<td>BVWS00681</td>
<td>2</td>
<td>40.5</td>
<td>18.39</td>
<td>0.0189544</td>
<td>0.1146288</td>
<td>281:281</td>
<td>-0.2166411</td>
<td>25</td>
</tr>
</tbody>
</table>

FL, fruit length; FW, fruit width; RT, rind thickness.

Moreover, with each generation during the breeding processes, only a few seeds from the best plants formed the next generation, and humans will experience a greater loss of diversity, because of narrow genetic diversity. Limited population size can cause loss of genetic diversity, even when the gene function and the phenotype of interest are yet unknown (Doebley et al. 2006). However, genes that influence important traits favored by humans will influence the diversity levels of neutral genes that are not linked to agronomic traits in domesticated organisms. Therefore, genetic bottlenecks due to conscious selection or gene flow from wild or weedy relatives can increase genetic diversity. A non-shattering phenotype in grain crops is an example of balancing selection for maintaining an adequate balance between harvesting and threshing (Chaisnam et al. 2010). Another important factor is that individuals that carry the most neutral alleles contribute the most alleles, whereas the most important genes are not linked to agronomic traits in domesticated organisms.
several genomic regions under selective sweep previously described by Guo et al. (2013). Vigouroux et al. (2002) identified 15 SSRs that exhibited evidence for selection in maize using a similar research approach as ours, and concluded that non-neutral SSRs are good candidates for agriculturally important genes. Our study used microsatellite alleles and entire world collections to characterize thirty genomic regions under selective sweep. Because microsatellite alleles are mutational hot spots, using them for characterizing genomic regions undergoing selection may provide additional information, compared with characterizing genome-wide selection using SNPs.

A major challenge remaining is how to distinguish true signals from those due to genetic drift caused by narrow genetic diversity (Tang et al. 2007; Granka et al. 2012; Qanbari et al. 2012). A possible solution is to estimate selective sweeps by analyzing diverse populations collected from various geographic areas with diverse marker sets, hypothesizing that true signals generated by selection would overlap across the populations (Qanbari et al. 2012). We are currently undertaking this, along with the validation of results obtained from larger populations of sweet watermelons.

It is still unclear whether crop plant recombination rates are positively associated with genetic diversity, despite considerable debate (Bauer et al. 2013). Bauer et al. (2013) characterized the recombination rate across the maize genome as highly variable. Recombination hot spots have been hypothesized to positively associate with genetic diversity, at least as demonstrated in human genetic diversity (Spencer et al. 2006). In this study, we noted that the recombination rate varies among and within chromosomes. In LD blocks, the recombination rate is quite low compared with the rest of the chromosomes. Recombination rates observed in the current research are comparable to those previously estimated (Ren et al. 2012).

Individual marker allele frequencies are very critical in association mapping studies. All the linked markers in our current research have an allele frequency of 0.15 or more. Out of 96 accessions a causative allele needs to be present in at least 12 individuals or more to infer association. Our current research identified important genes containing microsatellite motifs in UTR and promoter regions exhibiting associations with fruit length. BVWS01708 and BVWS00358 both showed linkage to fruit length. BVWS01708 is located near the promoter TATA box of a DTW domain-containing protein. BVWS00358 is a GAGA-type transcription factor in the 5’ UTR of a gene that expresses a C₂H₂ Zinc finger motif. Interestingly, these microsatellites containing GA and TA type repeats are specific to GAGA and TATA motifs in the watermelon genome. Furthermore, the typical American ecotype is characterized by elongated fruit length (oval shape), compared with the typical African and Asian ecotypes (round shape), so these 2 microsatellites, BVWS00358 and BVWS01708, might have important roles in ecotype differentiation. None of the QTLs mapped in the current study could be validated with the previously mapped QTLs by Sandlin et al. (2012). This may be due to the mapping population used in Sandlin et al. (2012) research, which were derived from wide crosses.

Analysis of sweet watermelon genome population structure, recombination rate, and selective sweeps will be immensely useful in the design of association mapping studies. The extensive resequencing of several sweet watermelon accessions around the world, along with wild-type counterparts, facilitated by the rapid progress in sequencing and analysis tools currently underway, will further facilitate generating genome scans for selective sweeps, and thereby, for mining novel alleles to improve horticulturally important traits.

Supplementary Material
Supplementary material can be found at http://www.jhered.oxfordjournals.org/.

Funding
USDA-NIFA (2013-38821-21453), NSF-EPSCOR#1003907, Gus R. Douglass Institute and NIH grant number P20RR016477.

Acknowledgments
We declare that we have no conflict of interest. The authors thank GRDI graduate assistantships for L. Abbari. The authors are grateful to R. Jarret, Plant Genetic Resources Conservation Unit, USDA-ARS (Griffin, GA) for providing the seeds of germplasm accessions.

References
Citrullus lanatus) accessions. Genetic

tant traits in watermelon (Citrullus lanatus) populations and quantitative trait loci associated with economically impor


