Genetic Diversity and Population Structure of Collard Landraces and their Relationship to Other Brassica oleracea Crops

Sandra E. Pelc,* David M. Couillard, Zachary J. Stansell, and Mark W. Farnham

Abstract
Landraces have the potential to provide a reservoir of genetic diversity for crop improvement to combat the genetic erosion of the food supply. A landrace collection of the vitamin-rich specialty crop collard (Brassica oleracea L. var. viridis) was genetically characterized to assess its potential for improving the diverse crop varieties of B. oleracea. We used the Illumina 60K Brassica SNP BeadChip array with 52,157 single nucleotide polymorphisms (SNPs) to (i) clarify the relationship of collard to the most economically important B. oleracea crop types, (ii) evaluate genetic diversity and population structure of 75 collard landraces, and (iii) assess the potential of the collection for genome-wide association studies (GWAS) through characterization of genomic patterns of linkage disequilibrium. Confirming the collection as a valuable genetic resource, the collard landraces had twice the polymorphic markers (11,322 SNPs) and 10 times the variety-specific alleles (521 alleles) of the remaining crop types examined in this study. On average, linkage disequilibrium decayed to background levels within 600 kilobase (kb), allowing for sufficient coverage of the genome for GWAS using the physical positions of the 8273 SNPs polymorphic among the landraces. Although other relationships varied, the previous placement of collard with the cabbage family was confirmed through phylogenetic analysis and principal coordinates analysis (PCoA).

Plant breeders require access to a genetically diverse pool of germplasm to provide alleles for crop improvement to meet the demand for food security in the face of an increasing human population (Gepts, 2006). Genetic erosion of the food supply is occurring by both an overreliance on only a few major crop species, including corn (Zea mays L.), wheat (Triticum aestivum L.), and rice (Oryza sativa L.) (Harlan, 1975; Collins and Hawtin, 1999) and wide-scale adoption of a limited number of elite breeding lines and cultivars (Gepts, 2006). Previous studies have shown that crop landraces can harbor substantial genetic variation through both adaptation to local environmental conditions and farmer-specific selection for desirable agricultural traits based on local cultural preferences and customs (Elías et al., 2004; Pressoir and Berthaud, 2004; Adoukonou-Sagbadja et al., 2007; Pusadee et al., 2009). Indeed, the characterization and subsequent conservation of germplasm in the form of wild relatives, landraces, and underutilized crops is essential to ensure future availability of genetic resources for breeding efforts (Harlan, 1975; Hammer et al., 2001; Gepts, 2006; Mayes et al., 2012).

Collard (Diederichsen, 2001) is a specialty vegetable crop primarily consumed in the southeastern United States. This leafy-green cruciferous vegetable is nutrition dense with high levels of β-carotene, lutein, vitamin C, vitamin B9, and vitamin K1 (Farnham et al., 2012). While primarily a cool-weather plant, collard can be grown

USDA–ARS, U.S. Vegetable Lab., Charleston, SC 29414. Received 16 Apr. 2015. Accepted 25 June 2015. *Corresponding author (Sandra.Pelc@ars.usda.gov).

Abbreviations: GRIN, Germplasm Resource Information Network; GWAS, genome-wide association studies; IBD, isolation by distance; K, the true number of populations; kb, kilobase; LD, linkage disequilibrium; Mb, megabase; NPGS, National Plant Germplasm System; PCoA, principal coordinates analysis; RAPD, random amplified polymorphic DNA; SNP, single nucleotide polymorphism.
under a wide range of conditions with year-round production in some areas. Commercial production of collard is limited to two or three primary cultivars, leading to a severe reduction of the germplasm available for this crop than a few decades ago (Farnham et al., 2008).

Collard has been widely grown across the southeastern United States by home gardeners and seed savers since at least the early 1800s, which has led to the proliferation of locally adapted collard landraces that display a wide range of phenotypic variability for both morphological and phytochemical traits and pathogen resistance (Farnham et al., 2001, 2008, 2012; Stansell et al., 2015). Extensive collection efforts and the subsequent ex situ conservation by the National Plant Germplasm System (USDA–NPGS) have preserved this important crop resource (Farnham et al., 2008). With the exception of five landraces (Farnham, 1996), the genetic diversity harbored within this collection was never characterized and no studies have examined the population structure or linkage disequilibrium (LD) of the collection.

Several other botanical varieties of B. oleracea are economically important vegetable crops, with broccoli (B. oleracea L. var. italica Plenk) and cabbage (B. oleracea L. var. capitata L.) alone accounting for a yearly cash value of over $1 billion in the United States (USDA, 2010). The recently conserved collection of collard landraces is a potentially rich genetic resource for introgression of beneficial alleles into these crops. However, only two of the studies that have attempted to elucidate the relationships of the botanical varieties within B. oleracea have included collard in their analyses and both were limited by availability of markers (Song et al., 1988; Farnham, 1996).

The Brassica 60k Infinium SNP array (Illumina) was recently developed by an international consortium using genomic and transcriptomic sequencing data primarily from B. napus L. but also B. oleracea and B. rapa L. (Isobel Parkin, Agriculture and Agri-Food Canada, personal communication, 2014). Relatively recent historical hybridization events (likely during human cultivation only 10,000 years ago; Cheung et al., [2009]) between the diploid progenitor species B. oleracea (CC genome) and B. rapa (AA genome) formed the allotetraploid B. napus (AACC genome) (U.N., 1935). Therefore, the Brassica SNP array includes both A- and C-genome SNPs. Genetic mapping studies have shown very few marker rearrangements between the C-genomes of B. oleracea and B. napus, suggesting the genomes are nearly identical (Parkin et al., 1995; Brown et al., 2014). The 60k Brassica SNP array has been successfully used to map trait variation in B. napus (Hatzig et al., 2015; Liu et al., 2013; Qian et al., 2014; Zhang et al., 2014) and B. oleracea (Brown et al., 2014) to distinguish Brassica species (Mason et al., 2014, 2015) and to examine genetic diversity within the A- and C-genomes of several Brassica species, including B. oleracea (Mason et al., 2015).

In this study, we used the Brassica 60k Infinium SNP array to genotype representative cultivars of the most economically important crop varieties of B. oleracea, as well as the USDA–NPGS collection of collard landraces. Our objectives were to (i) assess the usefulness of the Brassica 60K chip for genotyping these crops, (ii) examine the relationship of collards to the other crop varieties within B. oleracea, (iii) characterize the genetic diversity of the collard landraces, and (iv) evaluate population structure and patterns of linkage disequilibrium to assess the potential for future GWAS in this collection.

**Materials and Methods**

**Plant Materials**

We examined B. oleracea germplasm diversity using two sets of genotypes. The first was a set of 79 accessions obtained from the USDA–NPGS that had been collected from farmers and home gardeners across the southeastern United States and were originally designated as collard (Farnham et al., 2008). The second was a diverse array of popular commercial cultivars representing the most important crops within B. oleracea, including four broccoli (var. italica), four cabbage (var. capitata), one Brussels sprouts (var. gemmifera), two cauliflower (var. botrytis), one kale (var. acephala), one Portuguese broccoli (var. costata), and three collard cultivars (var. viridis) (Supplemental Table S1). Seeds for each accession were planted in 200-cell speedling trays (Speedling Inc.) and maintained in a greenhouse under natural light. Leaf tissue was collected at the 2- to 3-true-leaf stage from 20 plants per accession, bulked, lyophilized, and stored in a −80°C freezer.

**DNA Preparation and Genotyping**

DNA was extracted from ground lyophilized tissue of each bulked sample using ChargeSwitch gDNA plant kits (Invitrogen). DNA extractions were quantified with a Qubit fluorometer (Invitrogen) and diluted to a concentration of 50 ng µL⁻¹. Samples were genotyped at 52,157 SNP loci using the Illumina 60K Brassica SNP BeadChip array (Illumina). Samples were prepared at the Hollings Cancer Center Genomics Core Facility at the Medical University of South Carolina (Charleston, SC) following Illumina protocols for custom iSelect bead chip sample hybridization and staining. BeadChip array fluorescence was imaged using an Illumina HiScanSQ. Genome Studio software (Illumina) was used for allele calling of each locus with a GenCall threshold of 0.15.

C-genome physical positions of SNPs polymorphic in this population were determined by BLAST search of the array probe sequences against the B. oleracea draft genome database, OboBase (http://ocri-genomics.org/bolbase) using an E-value threshold of 1 × 10⁻⁴ (Yu et al., 2013). Probe positions in the A-genome were provided by the manufacturer (Illumina).

**Data Analysis**

Marker summary statistics including allele frequencies and variety-specific alleles were determined using TASSEL v5.1.0 (Bradbury et al., 2007). Monomorphic SNPs and loci with more than 75% missing data within each
dataset (all *B. oleracea* crop types and only collards) were removed from further analysis. Pairwise genetic distance matrices using polymorphic, unlinked markers specific to each dataset (i.e., all 95 accessions; *B. oleracea* accessions, *N* = 91; collards, *N* = 78; and landraces, *N* = 75 [accession numbers explained in results and discussion]) were created in MEGA with the *p*-distance model (version 6; Tamura et al., 2013). Principal coordinates analysis was performed on the pairwise genetic distance matrices with the pcoa function of the ape package (Paradis et al., 2004) in R (R Development Core Team, 2014) to visualize genetic distances between accessions.

Pairwise LD was calculated in sliding windows of 50 markers using PowerMarker software (Version 3.25; Liu and Muse, 2005). Linkage disequilibrium decay across the genome was evaluated by fitting a second-degree LOESS smoothing line to the scatterplot of *r*² over physical distance in R (Chambers and Hastie, 1992; R Development Core Team, 2014). The point at which the loess curve reached a plateau was considered the background level of LD. Adjacent loci with an *r*² > 0.5 were considered linked, and one of each pair was removed to create a reduced dataset of unlinked loci for all subsequent relatedness analyses.

The Bayesian clustering algorithm of the program STRUCTURE v2.3.4 (http://pritchardlab.stanford.edu/structure.html) was run to cluster the botanical varieties (*N* = 95) into populations using the admixture model with correlated allele frequencies (Pritchard et al., 2000; Falush et al., 2003, 2007; Hubisz et al., 2009). The population numbers of *K* = 1 to *K* = 10 were tested 10 times each with an initial burn-in of 35,000 iterations, followed by 35,000 Markov Chain Monte Carlo repetitions. The STRUCTURE output was summarized and graphics produced using the program Cluster Markov Packager Across *K* (CLUMPAK; Kopelman et al., 2015). CLUMPAK is an online pipeline (http://clumpak.tau.ac.il/) that runs STRUCTURE output files through the software packages CLUMPP (Jakobsson and Rosenberg, 2007), DISTRACT (Rosenberg, 2004), as well as two methods of determining the best *K*. Through permutation, the software CLUMPP aligns multiple runs of clustering to produce the best match for each *K*-value. The CLUMPP parameters used were the LargeK Greedy algorithm, with random input order and 2000 repeats. The CLUMPP consensus membership coefficients for all individuals were then graphically displayed using DISTRACT (Rosenberg, 2004). The best *K*-value was calculated using the Evanno Δ*K* method (Evanno et al., 2005).

A pairwise genetic distance matrix of all 95 accessions was used to create an unrooted neighbor joining tree in MEGA (version 6; Tamura et al., 2013). Tree support was determined by the interior-branch test method using 10,000 bootstrap replications with branches of less than 50% confidence collapsed. The tree was transformed to a cladogram with FigTree v1.4.2 (Rambaut, 2014).

To test for isolation by distance (IBD; Wright, 1943) between the bulked collard landraces (*N* = 75), a Mantel test (Mantel, 1967) was implemented with 10,000 permutations (to assess significance) using the ade4 package (Dray and Dufour, 2007) in R (R Development Core Team, 2014). A pairwise genetic distance matrix was created in MEGA as described above for the collard landraces, *N* = 75 (Tamura et al., 2013). Latitude and longitude for collection locations were obtained from the Germplasm Resource Information Network (GRIN) of the USDA-ARS (Supplemental Table S1) and a geographic distance matrix was created with the R package sp. (Pebesma and Bivand, 2005; Bivand et al., 2013).

### Results and Discussion

#### A-Genome Contaminants

Multiple lines of evidence suggested that some of the samples of collard seeds might have been misidentified or mixed with other species of crops grown by the individual seed savers. First, a couple of seed savers described saving seed as “mixed greens” (GRIN). Second, all of the collard landraces were grown under standard field conditions at the US Vegetable Laboratory in Charleston, SC (Stansell et al., 2015), and a few plots had individual plants with the phenotypic appearance of either a turnip (*B. rapa*) or a rutabaga (*B. napus* L. subsp. *rapifera* Metzg.) (unpublished data, 2011). Finally, each of the 95 taxa was evaluated for the number of A-genome SNPs that successfully amplified to test for contamination with either *B. rapa* (*A*-genome) or rutabaga (*B. napus; *A-* and *C*-genomes). Because of the high degree of homology and synteny between the *A-* and *C*-genomes (Kaczmarek et al., 2009; Yu et al., 2013; Chalhoub et al., 2014), all of the *B. oleracea* (*C*-genome) samples were expected to amplify some of the A-genome markers, but four lines had elevated numbers. These four accessions (V049, V056, V063, and V108) amplified 3077 to 6476 more of the 24,015 A-genome SNPs than any of the other accessions without a corresponding loss in C-genome SNP hybridization, indicating they contained both the *A-* and *C*-genomes (Supplemental Fig. S1). Because the samples were bulked, it is unclear whether they were *B. napus* accessions (AACC) or a mixture of collard (CC) with either *B. rapa* (AA) or *B. napus* accessions. The putative A-genome contaminants not only amplified more SNPs than any other grouping but also had twice as many polymorphic markers and 18 times more unique alleles than the other varieties, which may indicate they were preserved as a mix of leafy green *Brassica* species (Table 1). In addition, these four putative *B. rapa* or *B. napus* accessions form the most basal lineages of the neighbor joining tree of all 95 taxa (Fig. 1). Population structure analyses also separated the putative contaminants from the remaining samples (Supplemental Fig. S2, S3). The four accessions were removed from all further analysis.

#### Diverse *Brassica oleracea* Germplasm

The SNPs on the *Brassica* 60k array were primarily developed with *B. napus* sequencing data causing strong ascertainment bias toward loci most common in *B. napus* accessions. Possible outcomes of this ascertainment bias
were low rates of array hybridization for *B. oleracea* accessions and underestimates of genetic diversity. However, all of the *Brassica* varieties included in this study successfully amplified greater than 44% of the SNPs on the BeadChip array (Supplemental Table S2, full data matrix) and each resulted in thousands to tens of thousands of polymorphic markers (Table 1), supporting the use of the Illumina *Brassica* 60k SNP array for genotyping a diverse set of *B. oleracea* germplasm. Within the *B. oleracea* crop varieties, the collard landraces had the highest number of polymorphic markers and variety-specific alleles.

Genetic distance between accessions of the *B. oleracea* germplasm set was visualized with PCoA of a genetic similarity matrix using 9125 unlinked, polymorphic markers (Fig. 2). The collard landraces formed one main cluster
with the cultivars of collard, cabbage, Brussels sprout, and Portuguese tronchuda cabbage (var. *costata*). The broccoli and cauliflower cultivars separated from the main cluster along the first coordinate (30.1% variance explained) with these two crop types separating from one another along the second coordinate (11.7% variance explained). The kale cultivar formed its own unique group.

**Phylogenetic Relationships**

The neighbor-joining cladogram revealed several interesting patterns of relatedness in the diverse *Brassica* germplasm examined (Fig. 1). With the exception of collards, taxa grouped by botanical variety. The putative *B. napus* or *B. rapa* contaminants formed the most basal lineages, followed by a few individual collard landraces and a mixed clade of collard landraces, the kale cultivar, and the Brussels sprout cultivar. V113 was most closely related to the Blue Knight kale cultivar in the phylogenetic tree and displayed a curly-leaved, red kale morphological appearance in field trials (unpublished data, 2011) and therefore may have been crossed with a kale at some point in its breeding history. While some of the collard landraces were dispersed throughout the tree, the majority grouped into one large clade with the cabbage cultivars as the most closely related accessions. About half the collard landraces are semiheading types described as cabbage collards or heading collards from the Carolinas (Supplemental Table S1) and almost all clustered together in the tree near the cabbage clade. The cabbage cultivars, the Portuguese tronchuda cabbage (var. *costata*), cauliflower, and broccoli cultivars, each formed a separate group but were part of the same larger clade. Although these relationships must be interpreted with caution because of the limited number of cultivars for each botanical variety, the placement of collard with cabbage agrees with the only previous studies that have examined the relationship of collard within *B. oleracea* (Song et al., 1988; Farnham, 1996). The results of Farnham (1996) identified similar relationships to those of the cladogram presented herein except Brussels sprouts clustered with cabbage and collard instead of a separate clade, which agrees with the results of the PCoA (Fig. 2).

**Collard Landraces and Cultivars**

Of the 52,157 SNPs on the array, 48% (25,161 SNPs) failed to amplify in any of the collard landraces (*N* = 75), which was expected because approximately half of the probes were specific to the A-genome. Forty-two percent of the remaining SNPs were polymorphic (Table 1). After removing markers with >75% missing data, the final set of polymorphic markers in the collard landrace collection was 11,204 SNPs. There were 521 alleles that were unique to the collard landraces (i.e., 521 markers that were monomorphic if the landraces were excluded), while only nine
alleles were specific to the collard cultivars (Table 1). After accounting for differences in sample representation for the two groups, the landraces still harbored a substantial source of unique alleles (6.9 per landrace, on average) as compared with the cultivars (three per cultivar, on average). This finding differs from a previous study of collard genetic diversity using random amplified polymorphic DNA (RAPD) markers, which reported equal numbers (1.6 per sample, on average) of cultivar- and landrace-specific alleles after accounting for sample size (Farnham, 1996). This difference is not surprising given that estimates of genetic diversity and relatedness have been shown to vary based on the marker system used (Powell et al., 1996; Würschum et al., 2013; Rostami et al., 2015).

Genetic diversity within and between the collard landraces and cultivars was characterized with a pairwise allele-sharing matrix of 8464 polymorphic, unlinked SNPs (Supplemental Table S3). While the collard cultivars were very similar to the landraces (average similarity = 0.962), all of the landraces were genetically distinct with a minimum similarity of 0.865. On average, ‘Champion’ shared the most alleles with the landraces (0.977). The average similarity of ‘Top Bunch’ and ‘Hi-Crop’ to the landraces was 0.960 and 0.949, respectively. Although two of the collard cultivars, Top Bunch and Champion, shared most of their alleles (0.985), the third, Hi-Crop, was markedly different from either, with a similarity score of 0.904 and 0.927, respectively. These similarity scores are much higher than previously reported, with an average similarity within the cultivars of 0.83 and between the cultivars and landraces of 0.82 (Farnham, 1996). The disparate results of the two studies can again be explained by differences in experimental design, such as marker systems (RAPDs vs. SNPs), marker numbers (209 vs. 8464) and sample size (18 vs. 78 collards).

Collard Landraces
Genetic Diversity

The landraces exhibited a high degree of genetic similarity with an average score of 0.986. V112 was the exception, with inclusion in 65 of the 82 pairwise comparisons of less than 0.95 similarity, including the minimum of 0.883. V086 was also notably different, with membership in 12 of the 82 least-similar comparisons. Both V112 and V086 should be prioritized in NPGS seed bank collections to preserve the wide genetic base provided by these collard landraces. Three pairs of landraces were genetically identical (V079/V080, V093/V098, and V106/107), revealing redundancy that could be removed from the NPGS collard collection. The landraces within each identical pair were collected within 50 km of one another. Despite this geographic relationship, the genetic and geographic distance matrices were not correlated, showing a lack of IBD in the landraces (p-value = 0.144; r = 0.087). Parness et al. (2004) found a lack of IBD in barley (Hordeum vulgare L.) landraces as a result of gene flow primarily occurring through seed exchange between farmers, who can have unexpected long-distance relationships. Additionally, microenvironmental selection pressures were shown to affect genetic diversity and population structure of common bean (Phaseolus vulgaris L.) landraces (Tiranti and Negri, 2007).

In addition to genetic diversity between landraces, average accession heterozygosity was higher in the landraces (0.33; range = 0.09–0.50) than the cultivars (0.20), suggesting intra-accession variation in the bulked landrace samples (Supplemental Table S4). Farnham (1996) found the average intra-accession genetic similarity of 20 V062 individuals to their bulked sample (genetic similarity = 0.81) to be equal to the average interaccession genetic similarity of V062 to 18 different collard accessions. In addition, unique RAPD bands were found in some of the V062 individuals not found in any of the other collard samples.

Population Structure

The population structure of 78 collards (75 landraces plus three cultivars) was assessed using a set of 8464 unlinked markers with the program STRUCTURE and by PCoA analysis. STRUCTURE results for K = 3 (Evanno K) divided the collards into three main groups (Fig. 3). The first consisted of five North Carolina landraces, all described as cabbage collards, with membership coefficients in two clusters (excluding values <0.01; aqua and pink in Fig. 3). The second group consisted of all three
collard cultivars (Champion, Top Bunch, and Hi-Crop) and 16 collar landraces with membership in only two clusters (excluding values <0.01; purple and pink in Fig. 3). The landraces of this group vary phenotypically (leaf texture, color, and heading habit) and geographically with representation from all five states included in this study. The remaining 54 landraces of the third group were admixed with membership in all three clusters. The North Carolina landraces had substantially higher membership in the third cluster (aqua in Fig. 3) with an average of 20.2% as compared to less than 4% for the other geographic groups (Supplemental Table S5). Although there is no clear biological explanation for this pattern, state-specific cultural practices may have led to this distinction.

Earlier studies in both cassava (*Manihot esculenta* Crantz) and maize landraces have shown that farmer seed selection can be strongly influenced by local cultural preferences (Elias et al., 2004; Pressoir and Berthaud, 2004).

In the PCoA analysis, the collards form one main cluster with eight samples deviating from the group (Fig. 4). This included all three collar cultivars, two of the North Carolina cabbage collar landraces from first group of the STRUCTURE run (V086 and V030), two phenotypically kale-like curly collar landraces (V112 and V113), and a Mississippi collar (V111) that almost completely overlaps the cultivar Champion. The upper-left quadrant contained only North Carolina landraces, again genetically differentiating this group from the other landraces.

**Linkage Disequilibrium**

Despite the high genome-wide levels of similarity, this set of collar landraces has displayed extensive variation in important agronomic and nutritional traits (Farnham et al., 2001, 2012; Stansell et al., 2015) and could, therefore, have potential for elucidating the genetic basis of these traits through GWAS. The extent of LD in a collection of germplasm determines both the required marker density for GWAS in that population and the genetic resolution of the resulting associated causal regions (Myles et al., 2009; Bergelson and Roux, 2010; Weigel 2012). On average, LD in the collar landraces (*N* = 75) decays rapidly to $r^2 < 0.5$ by 5 kb, followed by a gradual decline to background levels ($r^2 < 0.1$) by 600 kb (Fig. 5), which was higher than expected for an outcrossing species. For example, LD decays within 2 kb in maize (Remington et al., 2001; Tenaillon et al., 2001) and 15 kb in grapevine (*Vitis vinifera* L.) (Myles et al., 2010). Linkage disequilibrium decay of 600 kb is more comparable to levels reported in self-pollinated crops, such as soybean [*Glycine max* (L.) Merr.] (90–500 kb; Hyten et al., 2007), rice (75–500 kb; Mather et al., 2007), and sorghum [*Sorghum bicolor* (L.) Moench] (150 kb; Morris et al., 2013). This could be a result of the high relatedness found within the landraces and the controlled breeding practices used to keep the landraces as distinct populations. This idea is supported by previous work in barley that showed a successive increase in the rate of LD decay from elite breeding lines and cultivars (212...
With the current estimated genome size of 630 Mb, LD decay at 600 kb would break the genome into 1050 haplotype blocks. If the 8273 mapped SNPs (polymorphic in the collard landraces) were evenly distributed across the genome, each haplotype block would have an average of seven to eight SNPs. However, average LD in the collard landraces varies greatly both between chromosomes and along the length of each chromosome, with extended blocks primarily in regions likely to contain centromeres and the very tips of the telomeres, which have lower levels of historical recombination events (Fig. 6). Additionally, the mapped SNPs are not evenly distributed, with adjacent SNPs ranging from 3 bp to 1.5 megabase (Mb) apart and an average distance of 47 kb (Fig. 6). Given the lower rate of LD decay, the marker density in this study would provide the level of coverage required for future GWAS. However, we found both uneven patterns of LD across the chromosomes and uneven distribution of markers; therefore, we expect the resolution of GWAS using this population and set of markers would vary from genic to Mb-levels.

**Conclusions**

The collard landraces had a substantially higher number of both polymorphic markers and unique alleles than collard cultivars and other *B. oleracea* crops, supporting their role as an important source of genetic diversity for both the maintenance of a broad genetic base for the leafy vegetable collards and introgression of unique alleles into other *B. oleracea* crops. The diversity results presented here can be used for efficient management of germplasm conservation systems, allowing for the preservation of this important genetic resource. For example, one of each of the three pairs of identical landraces can be removed from the system to eliminate redundancy. Additionally, the most divergent landraces, as shown by the PCoA analysis, should be given priority in conservation efforts to ensure preservation of the greatest possible extent of genetic diversity of the collard landrace collection.
Acknowledgments
The sequence data saved at www.OCRI-genomics.org as a database, Bolbase, were produced by BoGSC.

References


Figure 6. Linkage disequilibrium ($r^2$) averaged across 1-megabase (Mb) intervals for each chromosome in the collard landraces ($N = 75$). The red points at the bottom of each line graph represent the physical position of the 8235 single nucleotide polymorphisms used to assess linkage disequilibrium.