Published November 6, 2015

Population Genetics and Structure of a Global Foxtail Millet Germplasm Collection

Hari D. Upadhyaya,* Mani Vetriventhan, Santosh P. Deshpande, Selvanayagam Sivasubramani, Jason G. Wallace, Edward S. Buckler, C. Tom Hash, and Punna Ramu

Abstract

Foxtail millet [*Setaria italica* (L.) P. Beauv.] is one among the most ancient crops of dryland agriculture. It is the second most important crop among millets grown for grains or forage. Foxtail millet germplasm resources provide reservoirs of novel alleles and genes for crop improvement that have remained mostly unexplored. We genotyped a set of 190 foxtail millet germplasm accessions (including 155 accessions of the foxtail millet core collection) using genotyping-by-sequencing (GBS) for rapid single nucleotide polymorphisms (SNP) characterization to study population genetics and structure, which enable allele mining through association mapping approaches. After filtering a total 350,000 raw SNPs identified across 190 germplasm accessions for minor allele frequency (MAF), coverage for samples and coverage for sites, we retained 181 accessions with 17,714 high-quality SNPs with ≥5% MAF. Genetic structure analyses revealed that foxtail millet germplasm accessions are structured along both on the basis of races and geographic origin, and the maximum proportion of variation was due to among individuals within populations. Accessions of race *indica* were less diverse and are highly differentiated from those of *maxima* and *moharia*. Genome-wide linkage disequilibrium (LD) analysis showed on an average LD extends up to ~150 kbp and varied with individual chromosomes. The utility of the data for performing genome-wide association studies (GWASs) was tested with plant pigmentation and days to flowering and identified significant marker–trait associations. This SNP data provides a foundation for exploration of foxtail millet diversity and for mining novel alleles and mapping genes for economically important traits.

FOXTAIL MILLET is one of the ancient C_4 annual crops of dryland agriculture, grown since 10,500 yr ago in China (Yang et al., 2012). It is the second most important millet (after pearl millet), distributed in warm and temperate regions of the world including Asia, Europe, America, Australia, and Africa and used as grain, forage, or bird feed. It is mainly grown in China, India, Korea and Japan for food consumption (Austin, 2006). The genus *Setaria* is closely related to several C_4 bioenergy grasses with large complicated genomes, such as switchgrass (*Panicum virgatum* L.), Napier grass (*Pennisetum purpureum* Schumach.), and pearl millet [*Pennisetum glaucum* (L.) R. Br.]. It serves as an experimental model species to explore plant architectural traits, evolutionary genomics, and physiological attributes of the C_4 panicoid crops because of its short generation time, small diploid genome and inbreeding nature (Doust et al., 2009; Lata et al., 2013; Muthamilarasan and Prasad, 2014), and the availability of reference genome information from two studies (Bennetzen et al., 2012; Zhang et al., 2012).

H.D. Upadhyaya, M. Vetriventhan, S.P. Deshpande, and S. Sivasubramani, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru 502 324, Telangana, India; H.D. Upadhyaya, Dep. of Agronomy, Kansas State Univ., Manhattan, KS 66506, USA, and Institute of Agriculture, Univ. of Western Australia, Crawley, WA 6009, Australia; C.T. Hash, ICRISAT Sahelian Center (ISC), BP 12404, Niamey, Niger; E.S. Buckler, USDA–ARS, Ithaca, NY 14853; J.G. Wallace, E.S. Buckler, and P. Ramu, Institute for Genomic Diversity, Cornell Univ., Ithaca, NY 14853; Current Address: J.G. Wallace, Dep. of Crop and Soil Sciences, The Univ. of Georgia, Athens, GA 30602. Received 14 July 2015. Accepted 24 Aug. 2015.*Corresponding author (H.Upadhyaya@cgiar.org).

Abbreviations: CV, cross-validation; GBS, genotyping-by-sequencing; GWAS, genome-wide association study; LD, linkage disequilibrium; MAF, minor allele frequency; PCA, principal component analysis; SNP, single nucleotide polymorphism; SP, subpopulation; SSR, simple-sequence repeat.

Published in The Plant Genome 8
doi: 10.3835/plantgenome2015.07.0054
© Crop Science Society of America
5585 Guilford Rd., Madison, WI 53711 USA
An open-access publication
All rights reserved.

Intergenomic analyses of foxtail millet, *Brachypodium* spp., rice (*Oryza sativa* L.), sorghum [*Sorghum bicolor* (L.) Moench], and maize (*Zea mays* L.) revealed highly conserved collinearity that shows strong evolutionary relationship among these grasses (Zhang et al., 2012).

Foxtail millet is valued for its drought tolerance (Qi et al., 2013) and short duration, and its grains are nutritionally superior to other cereals such as rice and wheat (*Triticum aestivum* L.) (Saleh et al., 2013; Upadhyaya et al., 2011). It has a low glycemic index and is thus an ideal food for people suffering from diabetes (Anju and Sarita 2010). Foxtail millet has one of the largest germplasm collections of >46,000 accessions conserved in genebanks globally (Dwivedi et al., 2012) and has large within-species racial diversity: three races (*moharia*, *maxima*, and *indica*) and 10 subraces (*aristata*, *fusiformis*, and *glabra* in *moharia*; *compacta*, *spongiosa*, and *assamense* in *maxima*; and *erecta*, *glabra*, *nana*, and *profusa* in *indica*) (Prasada Rao et al., 1987). Representative germplasm sets such as core and mini-core collections have been formed (Upadhyaya et al., 2008, 2011), and these function as important genetic resources for genomic studies. Understanding the genetic basis of traits of economic interest in foxtail millet is essential for improving yield and adaptability to various stresses. Association mapping is an effective approach for dissecting the genetic basis of complex traits in plants and has been used in various crops, including foxtail millet (Jia et al., 2013). Initially, simple-sequence repeat (SSR) markers have been used to identify genomic regions associated with agronomic traits in foxtail millet (Vetriventhan 2011; Gupta et al., 2014), however, numbers of SSR markers used in these studies were less and did not cover the entire genome. Recently, Jia et al. (2013) sequenced 916 diverse foxtail millet genotypes and through GWAS identified several SNPs associated with morphoagronomic traits. The ICRISAT foxtail millet core collection represents the diversity of the entire foxtail millet collection conserved at the ICRISAT genebank (Upadhyaya et al., 2008) and included potential sources for climate resilience traits, such as disease resistance (Sharma et al., 2014) and salinity tolerance (Krishnamurthy et al., 2014), and agronomic and grain nutritional traits (Upadhyaya et al., 2011). To implement GWAS using the foxtail millet core collection as an association mapping panel, there is a need for genome-wide scanning of sequence variation. Recent advances in using next-generation sequencing platforms for genotyping approaches such as GBS (Elshire et al., 2011) have enabled genome-wide scanning of large germplasm collections to discover SNPs for exploring genetic diversity, population structure, and LD at the species level. The cost of generating such a GBS-SNP data set is a tiny fraction of that required for whole-genome sequencing or even reasonably dense SSR fingerprinting. It helps in effective germplasm management and its enhanced use in crop improvement. Here, we characterized a diverse representative foxtail millet germplasm set, including the foxtail millet core collection, to investigate population genetics, structure, phylogenies and LD and use GWAS to identify genomic regions and genes underlying natural variation for plant pigmentation and flowering time.

Materials and Methods

Genotyping-by-Sequencing Library Preparation, Sequencing, Single Nucleotide Polymorphism Calling, and Annotation

The experimental materials consisted of 190 foxtail millet germplasm accessions including all 155 accessions of the core collection (Upadhyaya et al., 2008) plus four control cultivars. This germplasm set represents three races and 10 subraces of foxtail millet originating from 23 countries spread over Asia, Africa, the Americas, and Europe (Supplemental Table S1). DNA was isolated from leaves of each accession at 4- to 6-leaf stage using the modified hexadecyltrimethyl ammonium bromide (CTAB) protocol (Mace et al., 2003) from 12-d-old seedlings. Lyophilized DNA was sent to the Genomic Diversity Facility (Cornell University, Ithaca, NY, USA) for genotyping following the GBS approach (Elshire et al., 2011) with ApeKI restriction enzyme used for complexity reduction. The library was sequenced in 96-plex in two lanes of an Illumina HiSeq 2000. Further, the sequences were mapped to 'Yugu1' foxtail millet reference genome (Bennetzen et al., 2012) using Bowtie v2.2.3 (Langmead and Salzberg, 2012) and SNPs were called using TASSEL v4.3.10 GBS pipeline (Glaubitz et al., 2014). Sequence tags, 64-bp sequences that included a leading 4-bp C[T/A] GC signature from the cut site, were identified and only tags with at least 10 total reads were retained (to remove sequencing errors). Functional annotation of the SNPs was performed with the help of reference gene feature information using snpEff v3.6 (Cingolani et al., 2012).

Population Genetics and Structure Analyses

Evolutionary divergence of within and between races and between pairs of individuals was performed using MEGA 6 (Tamura et al., 2013) following maximum composite likelihood model (Tamura et al., 2004) with 1000 bootstrap iterations of all genotypes. Diversity statistics such as allele frequency, heterozygosity, and gene diversity were estimated using the software Arlequin version 3.5.1.2 (Excoffier and Lischer, 2010). An analysis of molecular variance (AMOVA) to estimate population differentiation among populations with 1000 permutations was performed using Arlequin version 3.5.1.2 (Excoffier and Lischer, 2010). Population level differentiation statistics (F_{ST}) were measured using Arlequin version 3.5.1.2 (Excoffier and Lischer, 2010). The F_{ST} ranges from 0 (no population subdivision, random mating occurrence, no genetic divergence within the population) to 1 (complete isolation or extreme division) with F_{ST} of up to 0.05 representing negligible genetic differentiation. Population level pairwise F_{ST} and Reynolds' genetic distance (Reynolds et al., 1983) estimates were calculated

for subpopulations (identified through ADMIXTURE program) and three foxtail millet races using Arlequin version 3.5.1.2 (Excoffier and Lischer, 2010).

To construct a phylogenetic tree, pairwise distance matrices were generated for entire set and for individual races using MEGA 6 (Tamura et al., 2013) following maximum composite likelihood model with 1000 bootstrap iterations of all genotypes (Tamura et al., 2004). The resulting pairwise-distances matrices were formatted to import into DARwin software version 6 (Perrier and Jacquemoud-Collet, 2006) and were used to construct unweighted neighbor-joining phylogenetic tree. Principal component analysis (PCA) was performed using the R software packages SNPRelate and pca3d (Zheng et al., 2012; Weiner, 2014). Hierarchical population structure was estimated by using the ADMIXTURE program, a model-based estimation of ancestry in unrelated individuals using maximum-likelihood method (Alexander et al., 2009). ADMIXTURE implements a cross-validation (CV) feature that allows, together with the number of iterations to convergence, to determine the number of subpopulations (*k* values) that best fits the data. After choosing a subpopulation level, we considered 60, 70, and 80% inferred ancestry membership and individual accessions were assigned to subpopulation if they had at least 60% membership in that respective population following Wallace et al. (2015). The LD parameter r^2 was estimated among loci with TASSEL 5.0.1 (Bradbury et al., 2007). We examined pairwise LD values, analyzing all polymorphisms with MAF ≥5% and compared pairwise LD decay between polymorphisms assessed in the whole population. Genome-wide patterns of LD decay were assessed by measuring r^2 averaged in distance intervals at the whole-genome level and across nine foxtail millet chromosomes.

Genome-Wide Association Study

The foxtail millet core collection plus four control cultivars (n = 159) was used as an association mapping panel to dissect sequence variation associated with plant pigmentation (color of the plants in a plot recorded at flowering stage as green, pigmented, and deep purple), and days to 50% flowering (number of days from sowing to when 50% of plants in plot have started flowering) recorded during the year 2010 rainy season at ICRISAT, Patancheru, Telangana, India (17.53°N, 78.27°E, 545 m asl). The experiment was planted in an α-design with three replications; Best linear unbiased predictors for days to 50% flowering of each accession were used for GWAS. Genome-wide association studies were performed with a compressed mixed linear model (Zhang et al., 2010) implemented in the GAPIT R package (Lipka et al., 2012) and significant marker–trait associations were identified.

Table 1. Distribution of single nucleotide polymorphisms (SNPs) of each chromosome and percentage heterozygosity.

Chromosome	No. of SNP loci	Heterozygote	Start SNP locus	End SNP locus
		%	bp	
Entire genome	17,714	0.009	90,070	58,564,228
Chr 1	2068	0.009	90,070	42,062,941
Chr 2	2131	0.010	40,633	49,140,659
Chr 3	2273	0.008	5528	50,559,590
Chr 4	1291	0.012	24,259	40,203,116
Chr 5	1882	0.009	175,839	47,238,213
Chr 6	1663	0.010	46,423	35,962,247
Chr 7	2145	0.009	248,377	35,811,828
Chr 8	2131	0.011	55,218	40,614,315
Chr 9	2130	0.007	52,516	58,564,228

Results

Genome-Wide Single Nucleotide Polymorphism Variation

The GBS library preparation protocol, with ApeKI restriction enzyme, was used for complexity reduction and 366 million reads were generated, of which ~327 million (92%) passed the barcoding and quality thresholds in the TASSEL FastqToTagCountPlugin function. In total, 28.2 million unique tags were identified across all 190 accessions. A total of 1.5 million unique tags were retained after the filtering (with a threshold of minimum of 10 counts), out of which 91.35% tags aligned to the 'Yugu1' reference genome. From these, SNPs were called using TASSEL pipeline. Raw SNPs (over 0.35 million) were further filtered to obtain SNPs that had minimum 80% coverage across the samples, MAF ≥5%, and ≥25% coverage across the remaining sites. For the final data set with these filters, we were able to retain 181 accessions with 17,714 SNPs (Supplemental Table S1), with a range between 1291 SNPs on chromosome 4 to 2273 SNPs on chromosome 3 (Table 1). In general, the SNPs were distributed along the nine foxtail millet chromosomes but were concentrated in subtelomeric than pericentromeric regions (Fig. 1). Analysis of the position and distribution of each SNP locus on whole-genome level show that majority of SNPs (~40%) are within 1 kbp of adjacent SNPs (Supplemental Fig. S1). Further, using SnpEff (Cingolani et al., 2012), each SNP locus was annotated based on its genomic location to predict coding effects. It was found that 12% of SNP loci were located in exon regions (6.12% nonsynonymous; 5.77% synonymous), 23% in intergenic regions, and ~4% in intron regions (Fig. 2).

Population Genetics and Diversity

A total of 17,714 SNPs with MAF ≥5% on 181 accessions were used to study the population genetics and diversity of foxtail millet germplasm accessions. Average evolutionary divergence of individuals within races showed the maximum divergence within individuals of race *moharia*

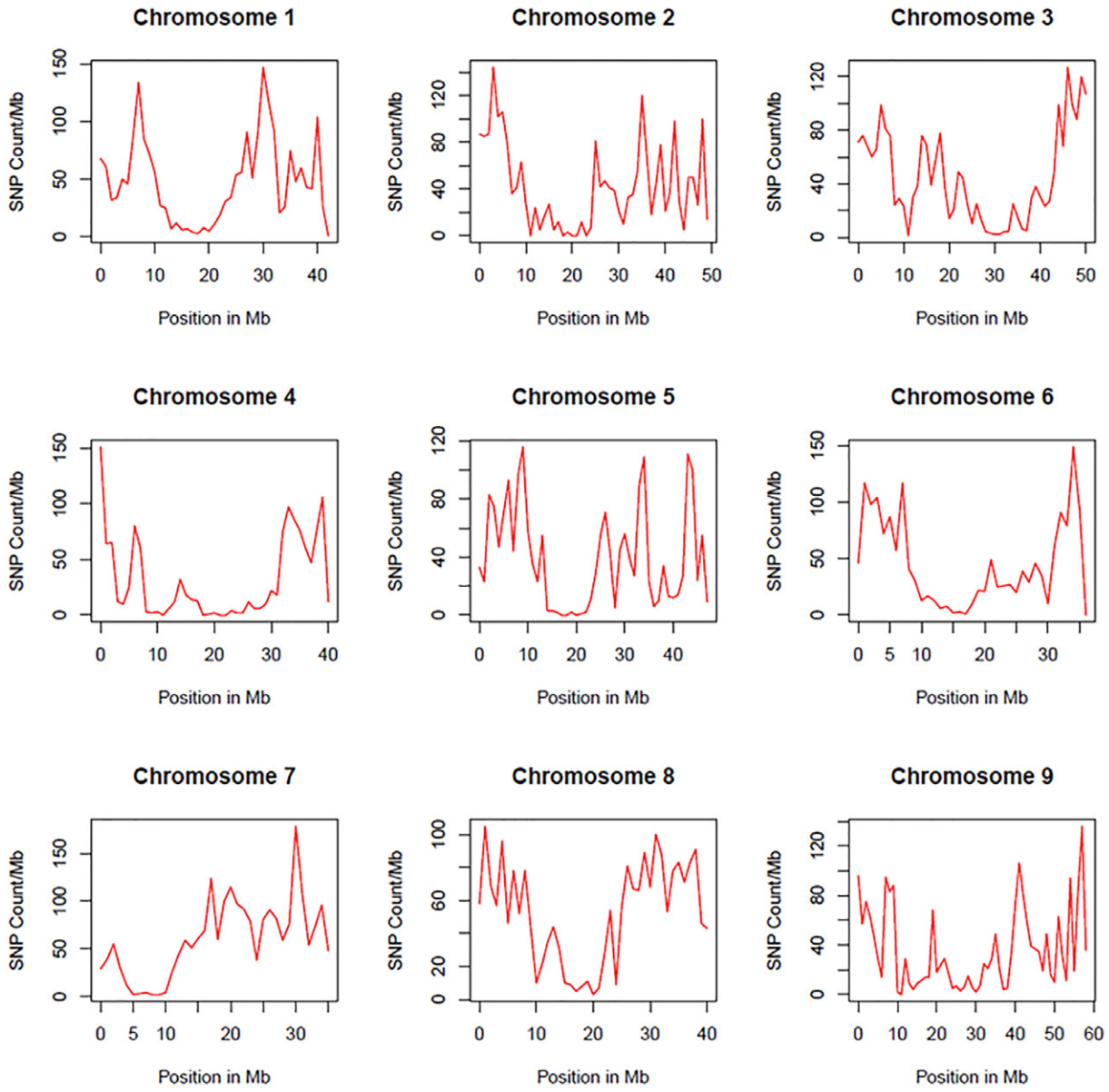

Figure 1. Distribution of single nucleotide polymorphism (SNP) loci across the nine foxtail millet chromosomes. The *x*-axis represents the physical location of the SNPs on each chromosome and *y*-axis denotes change per Mb.

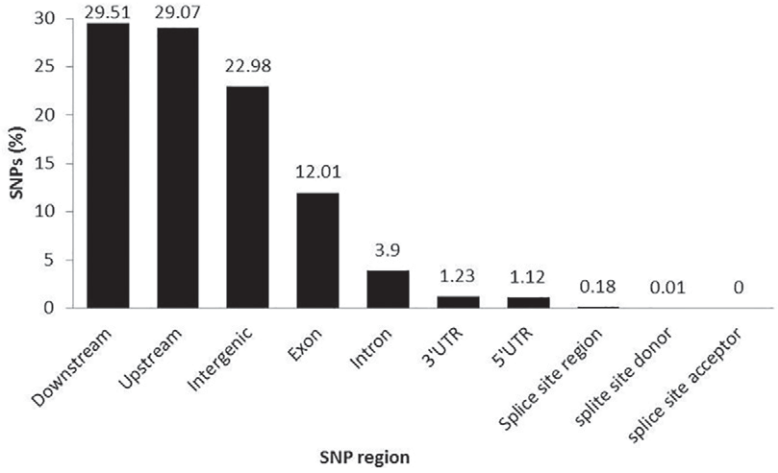

Figure 2. Single nucleotide polymorphism (SNP) annotation shows percentage SNP loci on different regions.

Table 2. Allelic richness and diversity in foxtail millet core collection based on racial classification and population groups identified based on ADMIXTURE program.

Molecular diversity indices	Entire	Race			Subpopulation (SP) based on ADMIXTURE program					
		Indica	*Maxima*	*Moharia*	SP1	SP2	SP3	SP4	SP5	SP6
No. of individuals	181	116†	28	37	24	21	43	36	24	33
Avg. heterozygosity	0.01	0.01	0.02	0.01	0.01	0.02	0.02	0.02	0.02	0.01
Standard deviation	0.01	0.01	0.02	0.02	0.03	0.03	0.03	0.02	0.03	0.02
Avg. gene diversity	0.28	0.21	0.34	0.35	0.33	0.31	0.22	0.23	0.25	0.33
Standard deviation	0.133	0.17	0.13	0.13	0.16	0.16	0.17	0.17	0.16	0.15
Evolutionary divergence		0.204	0.472	0.498						

†ISe 1470, yet unclassified and showed 100% ancestry membership with the race *indica* through both ADMIXTURE and neighbor joining population structure analyses, and hence included under the race *indica*.

Table 3. Analysis of molecular variance (AMOVA) and associated *F*-statistics based on racial classification and those based on subgroup identified by ADMIXTURE program.

Source of variation	df	Variance components	Percentage variation
Based on three biological races			
Among races	2	637.28**	29.59
Among individuals within races	178	1436.37**	66.69
Within individuals	181	80.15**	3.72
Total	361	2153.80	
Based on Subpopulations identified through ADMIXTURE program			
Among populations	5	69.62**	40.76
Among individuals within populations	175	92.55**	54.19
Within individuals	181	8.62**	5.05
Total	361	170.79	
F-statistics for biological races F_{IS}, 0.95**; F_{ST}, 0.30**; F_{IT}, 0.96**			
F-statistics for subgroups F_{IS}, 0.91**; F_{ST}, 0.41**; F_{IT}, 0.95**			

** Significant at $P \leq 0.01$; significance tests 1000 permutations.

(0.498) followed by *maxima* (0.472), but it was comparatively low in *indica* accessions (0.204) (Table 2). In addition, evolutionary divergence between pair of accessions was estimated, which ranged from 0.001 (with four pairs of accessions) to 0.865 (ISe 1339 and ISe 1408) with an average of 0.391 (Supplemental Table S2). The average gene diversity and heterozygosity of the entire set (n = 181) were 0.28 and 0.01, respectively (Table 2). The results of AMOVA revealed the large proportion of the total genetic variation from individuals within races (66.69%, $P <$ 0.001) compared with among races (29.59%, $P < 0.001$) and within individuals (3.72%, $P < 0.001$) (Table 3). The measure of population differentiation, F_{ST}, among races was 0.30, indicating all the three races are distinct and greatly differentiated. Inbreeding coefficient (F_{IS}) among races was 0.95, and population specific F_{IS} for *indica* (0.94), *maxima* (0.95), and *moharia* (0.96) showed the true inbreeding nature of foxtail millet accessions. Population pairwise F_{ST} is among the most widely used measures of genetic differentiation and plays a critical role in evolutionary genetic studies. The pairwise F_{ST} between *indica* and *maxima* (0.308, $P < 0.001$) and *indica* and *moharia* (0.366, $P <$ 0.001) were moderate compared with that of *maxima* and *moharia* (0.083, $P < 0.05$) (Table 4). Coancestry coefficients or Reynolds' distance (Reynolds et al., 1983) were in agreement with estimates of F_{ST}, where the maximum divergence was observed to occurs between *indica* with *moharia* (0.455) and *maxima* (0.367), while between that of *maxima* and *moharia* was low (0.087).

Phylogenetics and Population Structure

Phylogenetic structure analysis on 181 accessions using 17,714 SNPs showed that accessions belonging to *indica* race were completely differentiated from those of *maxima* and *moharia*. Few accessions that were found to be admixtures and were not grouped with those of three races (Fig. 3a). These could be of genome admixtures or possible hybrids between *indica* with those of *maxima* or *moharia*. The phylogenetic structure of subraces belong to three races seemed highly promising (Fig. 3b–d), however, representation in some subraces were with only a few accessions: subraces *erecta* and *profusa* in *indica*, *assamense* and *spongiosa* in *maxima*, and *fusiformis* in *moharia* were represented by less than four accessions each (Fig. 3b–d). In a PCA, conducted to study population structure, the first three PCs explained 59% of variation, and accessions were structured according to biological races with some exceptions (Fig. 4). The hierarchical population structure was determined separately for the entire set (n = 181) and the core collection (n = 155) using the model-based ADMIXTURE program assuming k = 1 to 10 populations, without providing any information on population structure. A line graph using CV errors of each k was drawn (Fig. 5a,b). When the k-values (from 1 to 10) were plotted against corresponding CV values, the CV error reduced steadily up to k = 4 and plateaued afterward in both the entire set and the core collection (Fig. 5a,b). At k = 4, 13 accessions belonging to *maxima* were included in Subpopulation (SP) 1 along with 10 accessions of *moharia*. Subpopulation 2 represents maximum accessions belong to *moharia* (17 accessions). The SP3 and SP4 represent the accessions belonging to *indica* (Fig. 6a); however, we obtained the lowest CV error at k = 6, where accessions belonging to *indica* were further divided into four subpopulations while *maxima* and *moharia* grouped into two separate subpopulations. The CV error values between k = 6 and k = 7 were reduced minutely (0.00018); hence, we decided to use k = 6, as six subgroups also fits the population structure as expected with racial diversity based on phenotypic characterization.

Table 4. Population pairwise F_{ST} estimates (below diagonal) and Reynolds' distance (above diagonal; Reynolds et al., 1983) among races and six subpopulations (SP) identified through ADMIXTURE program.

	Indica	*Maxima*	*Moharia*		SP1	SP2	SP3	SP4	SP5	SP6
Indica	0	0.367	0.455	SP1	0	0.391	0.859	0.898	0.781	0.503
Maxima	0.308***	0	0.087	SP2	0.324***	0	0.932	0.961	0.822	0.536
Moharia	0.366***	0.083***	0	SP3	0.576***	0.606***	0	0.132	0.234	0.252
				SP4	0.593***	0.617***	0.123***	0	0.173	0.294
				SP5	0.542***	0.561***	0.209***	0.159***	0	0.216
				SP6	0.395***	0.415***	0.223***	0.255***	0.194***	0

*** Significant at $P \leq 0.001$ (100 permutations).

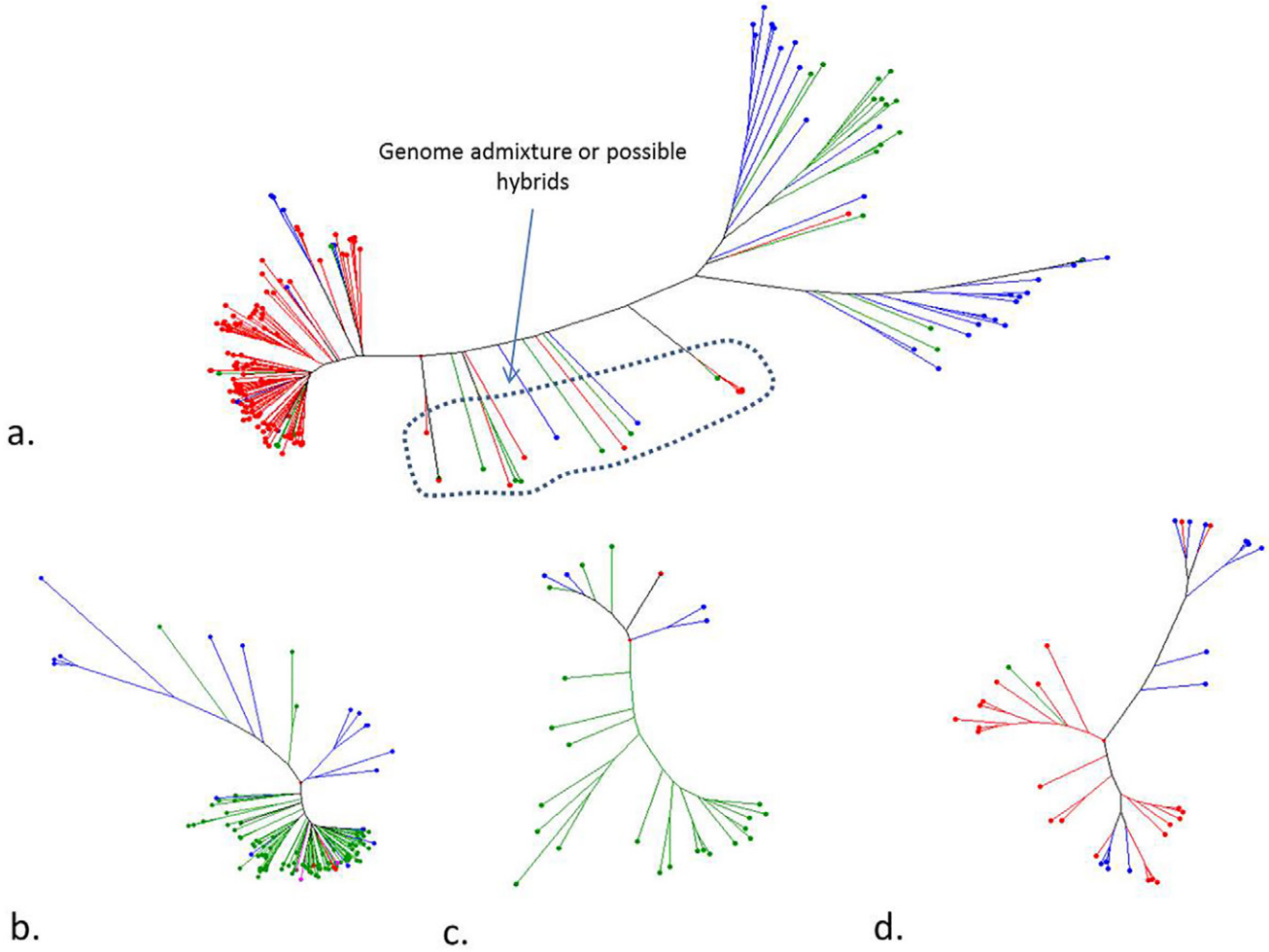

Figure 3. Phylogenetic structure of foxtail millet germplasm: (a) three races of foxtail millet: *indica*, red; *maxima*, green; *moharia*, blue; (b) four subraces of *indica*: *nana*, green; *glabra*, blue; *profusa*, purple; *erecta*, red; (c) three subraces of *maxima*: *assamense*, red; *spongiosa*, blue; *compacta*, green; and (d) three subraces of *moharia*: *aristata*, blue; *glabra*, red; *fusiformis*, green.

At $k = 7$, four *indica* accessions were extracted out from a larger *indica* group. This allowed the consideration of six subpopulation ($k = 6$) groups as SP1 to SP6 (Fig. 6b). Individual samples were assigned to subpopulations if they had ≥60% membership in that population. About 74% (134 accessions) of individuals had 60% membership in their respective subpopulation. Few accessions found admixture and were assigned to subpopulations in which they had maximum inferred ancestry. The 181 accessions used in this study, 115 accessions belong to *indica*, 28 to *maxima*, 37 to *moharia*, and one unclassified accession.

Among accessions of the entire germplasm set that were structured into six subpopulations, SP1 included accessions that belong to *maxima* (13 accessions) and *moharia* (10 accessions), while the majority of the accessions from *moharia* were in SP2 (17 accessions). Accessions belonging to the race *indica* grouped into four subpopulations (SP3, SP4, SP5, and SP6), where many accessions seem admixtures with <60% inferred ancestry mostly shared with/among subpopulations of *indica* (Supplemental Table S1). When we look into the percentage membership of individual accessions with >70 and 80%, we found 106

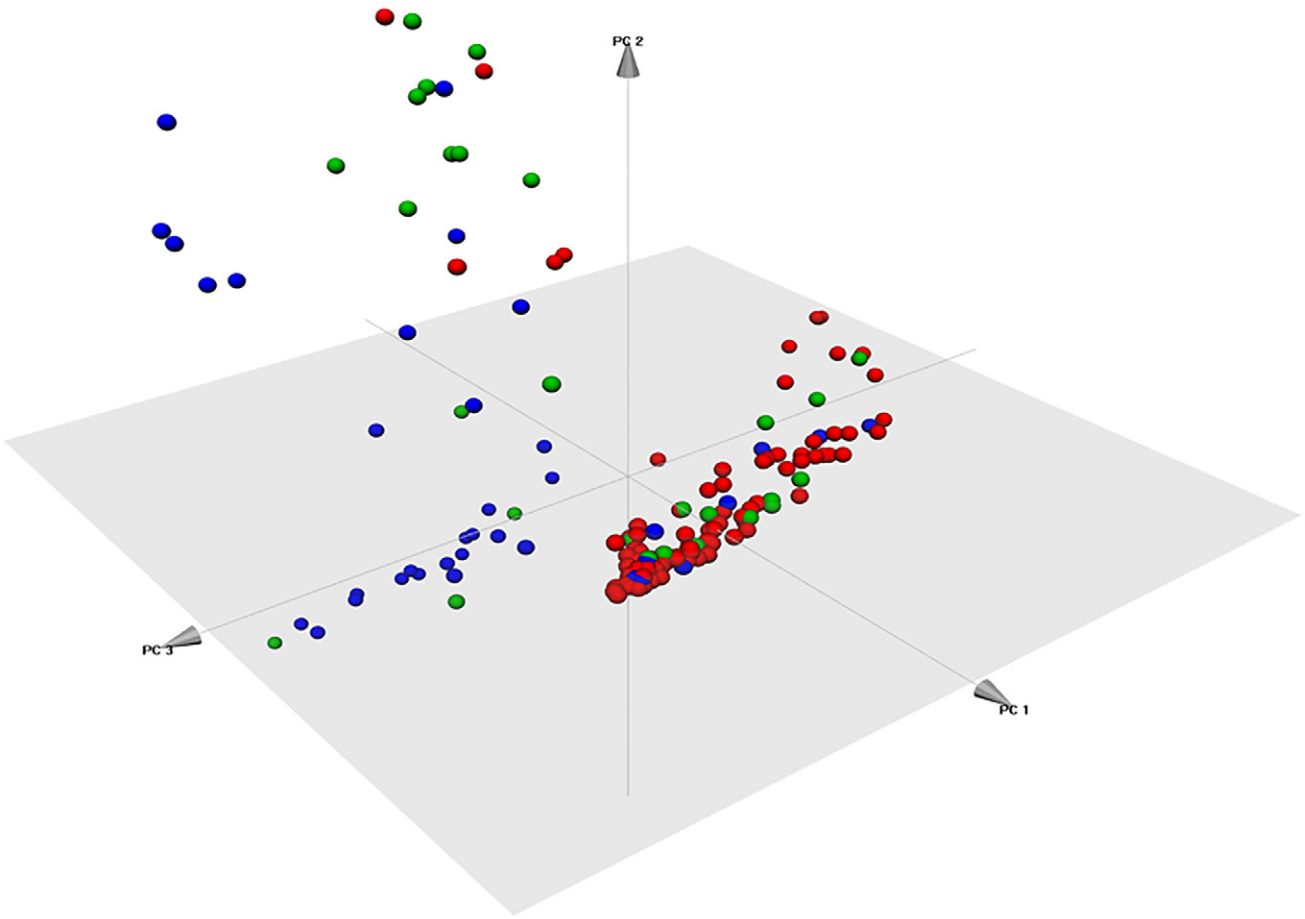

Figure 4. Principal component analysis of foxtail millet germplasm (181 accessions) and color coded according to races: *indica*, red; *maxima*, green; and *moharia*, blue.

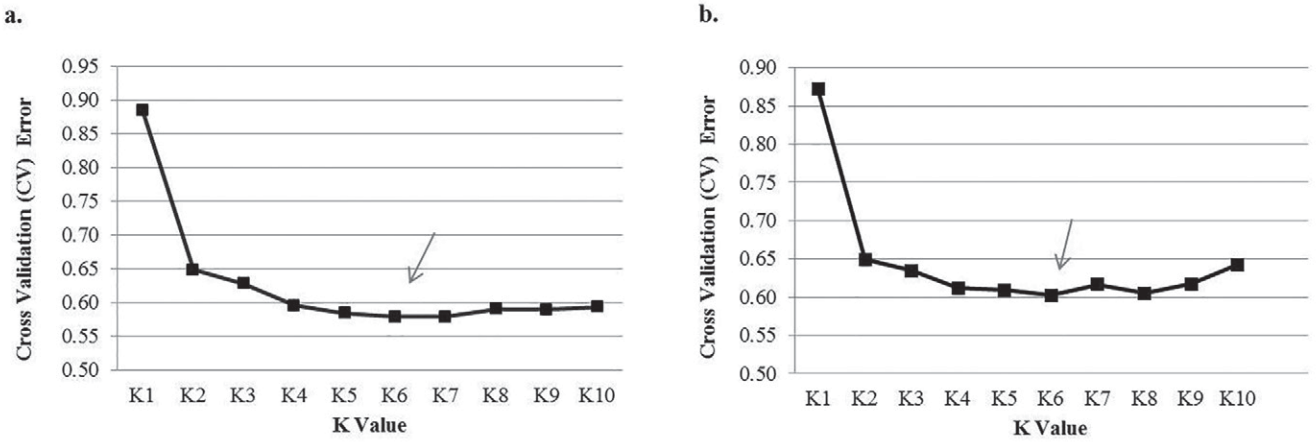

Figure 5. Rate of change in cross-validation (CV) error between successive *k*-values; *k*-values ranged from 1 to 10; (a) Entire population (*n* = 181); (b) Core collection (*n* = 155).

for former and 85 accessions for latter in their respective subpopulations (Supplemental Table S1). When we looked into phylogenetic consensus network, which was color coded on the basis of subpopulations detected via a model ADMIXTURE structure (Fig. 6c), we identified two clusters, wherein accessions in SP3 to SP6 mostly belongs to race *indica* and were highly differentiated from those of SP1 and SP2, wherein accessions belonging to *maxima* and *moharia* were present. The AMOVA revealed a large proportion of the total genetic variation resulting from differences among individuals within SPs (54.19%, $P < 0.001$) and among SPs (40.76%, $P < 0.001$), whereas only 5.05% of variation was within individuals (Table 3). Population pairwise estimates of F_{ST} were significantly higher ($P < 0.001$) between all SPs, which were in agreement with Reynolds' distance (Table 4).

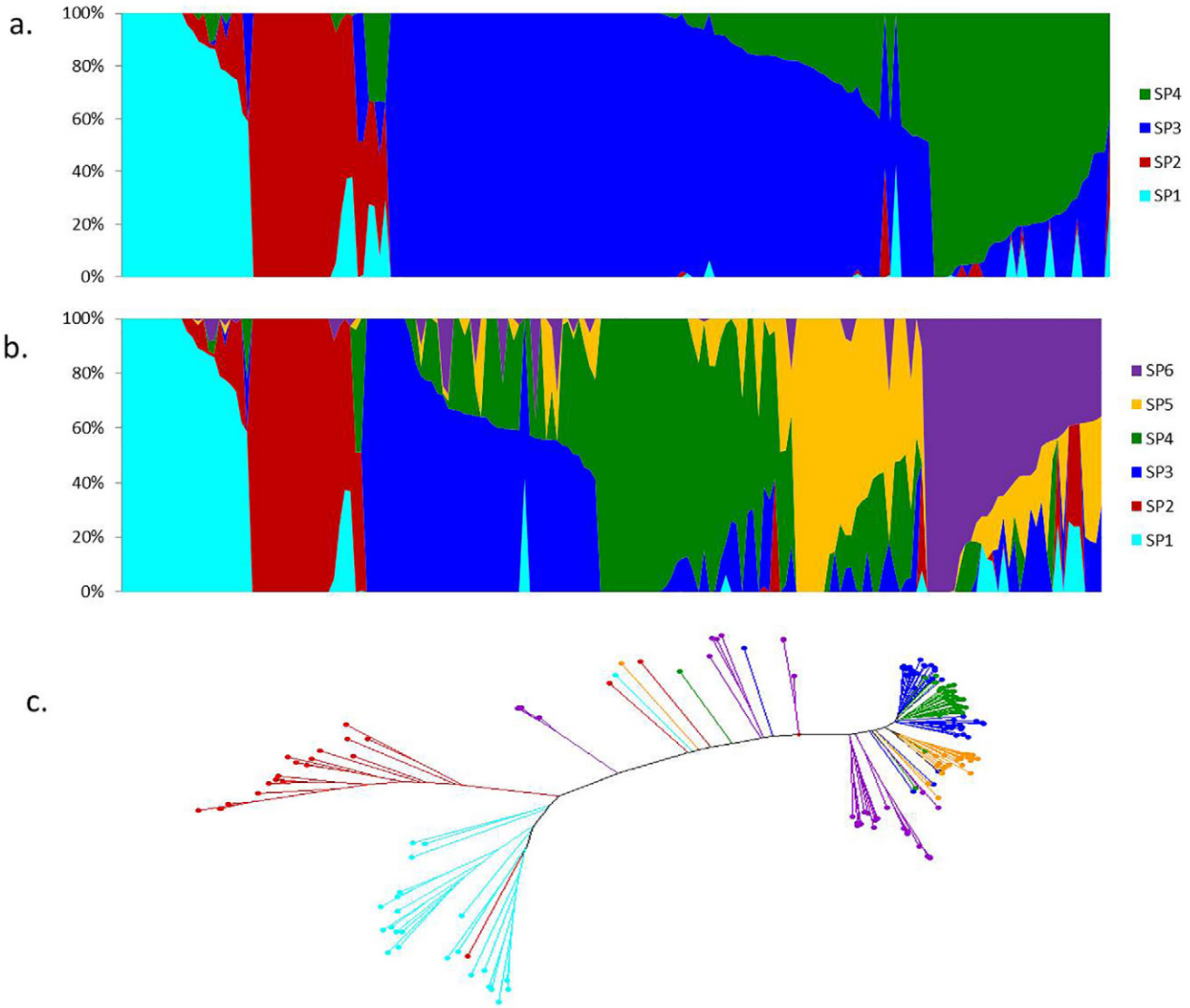

Figure 6. Population structure of foxtail millet germplasm characterized using 17,714 single nucleotide polymorphisms assessed through ADMIXTURE program; (a) k = 4, four subpopulations (SP): SP1, aqua; SP2, red; SP3, blue; SP4, dark green; (b) k = 6, six subpopulations; SP1, aqua; SP2, red; SP3, blue; SP4, dark green; SP5, orange; and SP6, purple; and (c) phylogenetic tree of foxtail millet germplasm accessions, color coded according to six subpopulations.

Genome-Wide Linkage Disequilibrium and Genome-Wide Association Study

The average extent of LD decay at the whole-genome level and for individual chromosomes was quantified to understand the pattern of LD decay. The level of LD was measured and squared allele-frequency correlations (r^2) were plotted against the distance between pairwise SNPs (Fig. 7). At the whole-genome level, average LD decays from its initial value of r^2 = 0.62 to a background level of r^2 = 0.20 (where it decayed to ~70% from the initial value) after 150 kb, and to 0.12 (where it decayed to ~80% from the initial value) at 400 kb. This varies with individual chromosome (Supplemental Table S3). The foxtail millet core collection plus four control cultivars (n = 159) were used to test the ability of identified SNP loci for detecting significant associations for days to flowering and pigmentation. Days to 50% flowering in the foxtail millet core collection varied from 33 to 104 d with the mean of 55 d after sowing, and mean days to flowering significantly differed among races (Vetriventhan, 2011). We performed genome-wide association analyses on these two traits and located the genomic regions that are associated with these traits (Table 5). These traits are highly investigated traits in many crop species, including maize, rice, and *Arabidopsis thaliana* (L.) Heynh. Genome-wide association studies on plant pigmentation revealed the most significant association between 7.2 to 7.3 Mbp of chromosome 4 (Table 5; Fig. 8). Jia et al. (2013) reported many marker–trait associations for bristle color, leaf sheath color, and pulvinus color between ~7.2 to 7.4 Mbp on chromosome 4 and found a coloration pathway gene *Si008089m.g* (a homolog of maize *C1*-colored aleurone1; Maize gene ID: GRMZMG005066) on chromosome 4 at 7.0 Mbp. We found the gene *Si006495m.g* located on chromosome

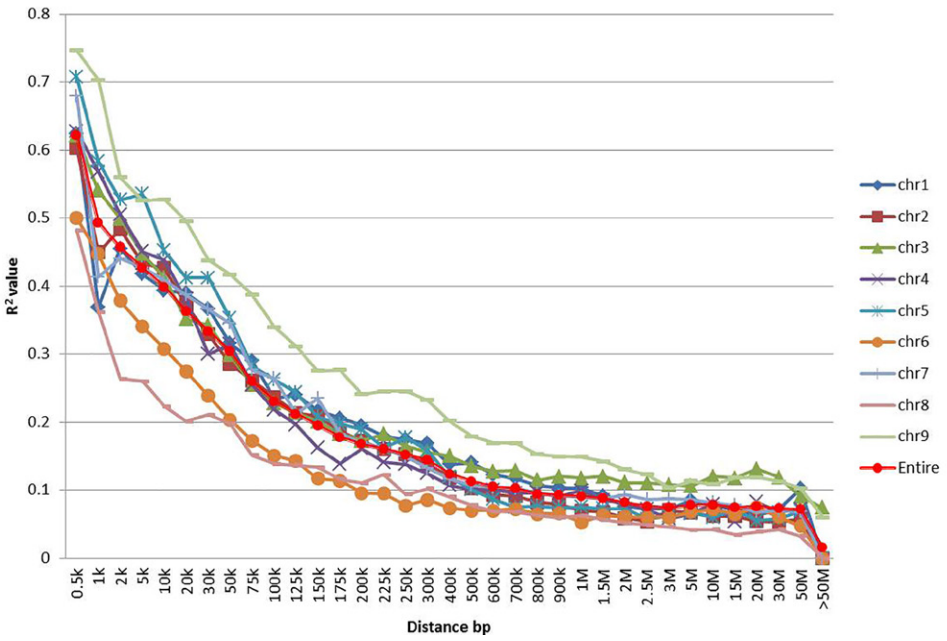

Figure 7. Genome-wide patterns of linkage disequilibrium decay as measured by r^2 averaged in distance intervals across nine foxtail millet chromosomes and averaged at whole-genome level.

Table 5. Significant marker traits associations detected for plant pigmentation and days to 50% flowering in the foxtail millet core collection.

Trait	Chromosome	Position	Major allele	Minor allele	Minor allele frequency	*P*-value	Candidate gene	Distance from candidate gene	Protein family (function)	Reference
Pigmentation	4	7291610	G	A	0.17	1.73×10^{-9}				
	4	7279106	G	A	0.18	2.06×10^{-9}				
	4	7313545	G	C	0.20	4.29×10^{-9}				
	4	7345322	G	A	0.19	1.36×10^{-8}	Si007984m.g	~8 kb downstream	Helix–loop–helix DNA-binding protein, putative (involved in anthocyanin biosynthetic pathway)	Heim et al., 2003; Himi et al., 2011; Nesi et al., 2000; Spelt et al., 2000
	4	7301853	T	C	0.24	1.15×10^{-7}				
	4	7284766	A	G	0.23	1.41×10^{-7}				
	4	7353388	A	T	0.21	1.50×10^{-7}	Si007984m.g	232 bases downstream	Helix–loop–helix DNA-binding protein, putative (involved in anthocyanin biosynthetic pathway)	Heim et al., 2003; Himi et al., 2011; Nesi et al., 2000; Spelt et al., 2000
	4	7228209	C	G	0.25	2.05×10^{-7}	Si0006495m.g	~32 kb upstream	Chalcone-flavanone isomerase (involved in flavonoid biosynthesis pathway)	Winkel-Shirley, 2001
Days to flowering	8	34187207	T	G	0.16	8.85×10^{-5}	Si028145m.g	3553 bases Downstream	Leucine-rich repeat receptor-like protein kinase (LRR-RLK) (involved in flowering time)	Zhou et al., 2004
	6	31867864	G	A	0.32	0.0001	Si015103m.g	~38 kb upstream	F-box domain (involve in growth and development of floral organ)	Samach et al., 1999; Ikeda et al., 2007; Hepworth et al., 2006
	6	31867469	T	A	0.34	0.0003	Si015103m.g	~38 kb upstream	F-box domain (involve in growth and development of floral organ)	Samach et al., 1999; Ikeda et al., 2007; Hepworth et al., 2006
	7	25380812	C	G	0.31	0.0003				
	6	34472720	G	A	0.21	0.0003	Si014861m.g	332 bases downstream	bZIP transcription factor (promote flowering)	Abe et al., 2005
	6	34568483	T	C	0.42	0.0003	Si015213m.g	In gene	Helix–loop–helix DNA-binding domain containing protein (involve in floral organ formation and development)	http://59.163.192.91/FmTFDb/tf.php?trans_fac_id=Si015213m
	6	34570454	A	G	0.44	0.0003	Si015213m.g	1891 bases upstream		

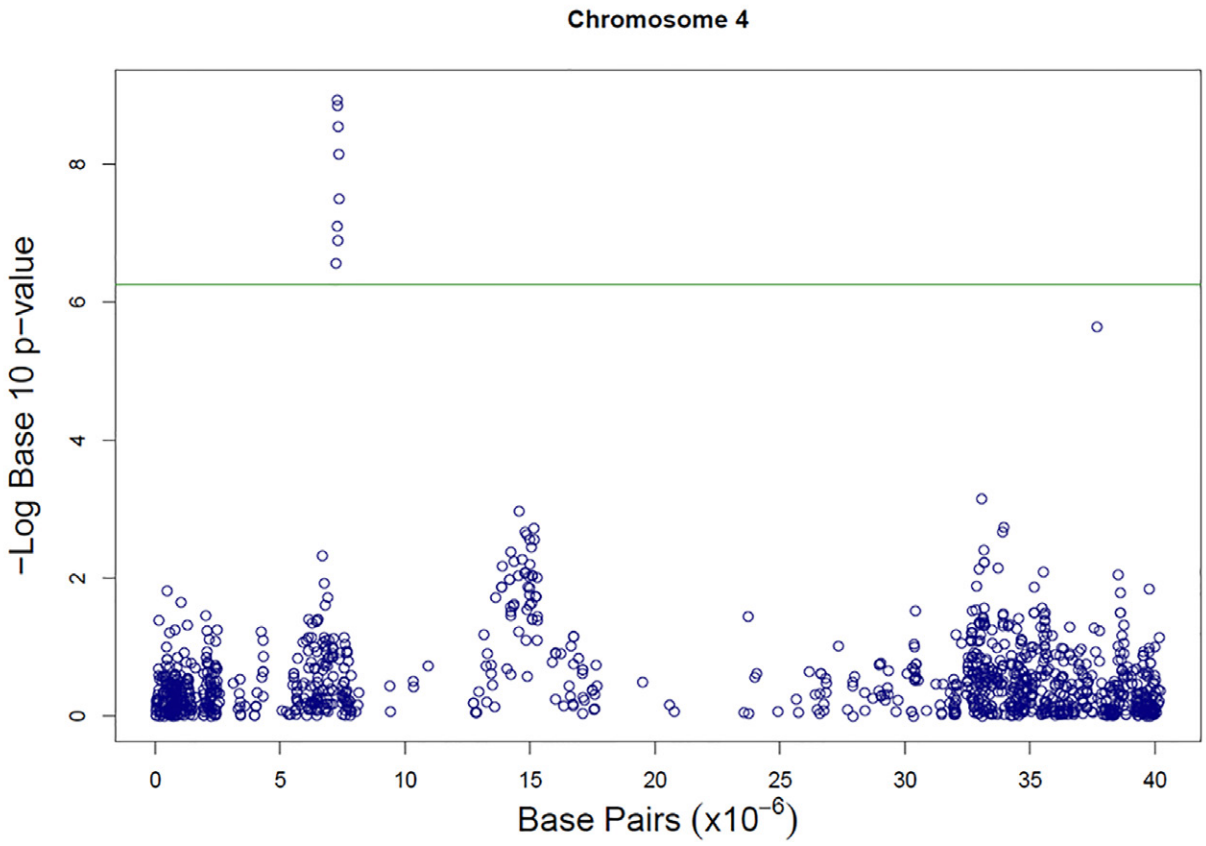

Figure 8. Manhattan plot depicting single nucleotide polymorphism (SNP) loci associated with plant pigmentation in foxtail millet. Each dot represents a SNP locus, with the *x*-axis showing genomic location (Mb) and *y*-axis showing association level.

4 (Position 7,189,103 to 7,195,068 bp), which encodes *Chalcone-flavanone isomerase*, a major enzyme for flavonoid biosynthesis pathway that leads to production of secondary metabolites including anthocyanins, which gives color to the plant organs. For days to 50% flowering, we found many SNP loci that showed significant associations on chromosome 6 (Table 5). Jia et al. (2013) also reported similar marker–trait association between 34.0 and 35.5 Mbp region on chromosome 6. The SNP locus located at 34,568,483 bp on chromosome 6 located on gene ID: *Si015213m.g* has a protein family called *MYC* type, *bHLH* domain, reported to have biological function of floral organ formation (GO: 0048449), regulation of flower development (GO: 0009909), leaf morphogenesis (GO: 0009965), etc. (http://59.163.192.91/FmTFDb/tf.php?trans_fac_id=Si015213m).

Discussion

Short life cycle, marginal land adaptability, biotic and abiotic stress tolerance, C4 photosynthetic pathway, wide architectural traits, small diploid genome, and inbreeding nature of foxtail millet are more desired valuable assets for the ever-warming and changing global climate. The ICRISAT genebank's germplasm collection of foxtail millet has the widest global representation and provides unlimited options for selection. Management of large numbers of germplasm had been facilitated by designing subsets like core and mini-core collections with the retention of large diversity in place (Upadhyaya et al., 2008, 2011). Advancements in sequencing and genotyping technologies have made it possible to genotype large germplasm collection, adding value to the germplasm conservation and use as demonstrated in foxtail and barnyard millets (*Echinochloa* spp.) (Jia et al., 2013; Wallace et al., 2015). Next-generation sequencing technologies support germplasm management and enhance use of germplasm resources in crop improvement program (van Treuren and van Hintum, 2014). They help in understanding the genetic diversity and population structure, allow effective germplasm management, access to a wider diversity of a crop's gene pool, and support researchers in selecting appropriate germplasm accessions for higher genetic gain in hybridization programs. In addition, sequence data of genebank accessions may be used to determine genetic composition and thereby help genebank curators to make more informed decisions about eliminating redundancies and identifying gaps to fill in with new acquisitions (van Treuren and van Hintum, 2014). Next-generation sequencing data could also be used to monitor the regeneration of accessions to ensure the maintenance of genetic integrity by comparing sequence data of samples before and after regeneration (van Hintun et al., 2007).

To improve the foxtail millet germplasm management and enhance its use and genomic research in foxtail

millet improvement, we identified the genome-wide SNP variation of a representative sample of the ICRISAT foxtail millet germplasm collection, including the entire core collection accessions (Upadhyaya et al., 2008). These germplasm accessions originate from 23 countries spread over Asia, Europe, America, and Africa. The GBS approach generated a total of 17,714 quality SNPs with MAF of ≥5% across 181 accessions. Evolutionary divergence between pairs of accessions has been provided for these 181 accessions (Supplemental Table S2), which will help in selecting diverse parents for hybridization programs. Although there were fewer accessions belonging to *moharia* and *maxima* than *indica*, we observed a maximum within-population diversity in both *moharia* and *maxima*. Similar findings were reported for this foxtail millet core collection characterized using SSR markers (Vetriventhan et al., 2012). This could possibly be due to a bottleneck or severe natural-selection sweep that the *indica* race faced during its separation from other natural populations. The estimates of gene flow and genetic differentiation within and between populations can be important and informative since they are related to the population's evolutionary history. We found low levels of shared alleles and significant genetic differentiation (F_{ST}) between *indica* with those of *maxima* and *moharia*, whereas a higher level of shared alleles and lower genetic differentiation (F_{ST}) between *maxima* and *moharia* were found. It shows that the *maxima* and *moharia* accessions have a shared evolutionary lineage and might have evolved from the same natural population, while *indica* had a divergent phylogenetic lineage, thus resulting in divergent population. Similar genetic structure was reported for this foxtail millet core collection characterized using SSR markers (Vetriventhan et al., 2012, 2014).

Prasada Rao et al. (1987) suggested three races in foxtail millet based on comparative morphology features, especially on the basis of variability in panicles structure of the foxtail millet accessions and our genetic structure analyses (phylogenetic tree, PCA, and a model-based ADMIXTURE) correlate well with the assigned racial structures. Accessions belonging to the *indica* race were highly diverged from those of *maxima* and *moharia*, as evidenced from shared allele frequency among races and F_{ST} statistics; however, the individual subraces did not clearly separate. This could be due to the smaller number of accessions studied from each subrace, making it difficult to separate them. Also, how well the morphology-based racial classifications correlates with genome-wide SNP variation depends on both the level of divergence between races and the coverage of sequence variations that actually differentiate the races. For example, Wallace et al. (2015) investigated the phylogenetic structure of barnyard millet through genome-wide SNPs variation and reported that the phylogenetic relationship does not correlate well with racial structure in barnyard millet. Racial classification (Prasada Rao et al., 1987) also encompasses geographical distribution and morphological features. Most of the *moharia* accessions originate from Europe and southern and western Asia; these include cultivars with five to 52 culms, each bearing several small, more or less erect, inflorescences. Accessions belonging to *maxima* mostly originate from eastern and southern Asia and are characterized by plants with one to eight, usually unbranched, culms that bear large inflorescences. Accessions belonging to *indica* originate from southern Asia and are characterized by plants with intermediate culm number (average 6.6) and inflorescence size between those of *moharia* and *maxima* (Prasada Rao et al., 1987).

The LD decay rate determines the mapping resolution of GWAS, interpretation of association peaks, and the transfer of alleles in marker-assisted selection (Jia et al., 2013; Morris et al., 2013). Foxtail millet is a self-pollinating species, expected to have a high level of LD (Wang et al., 2012). As expected, the genome-wide LD decay rate was ~150 kb on average, where the LD decays from its initial value of 0.62 to background (0.2), and it varies with individual chromosomes, which could lead to differential resolution in different chromosomes and regions of the genome in GWAS. The present study shows that about 40% of these SNP loci were within 1 kb of adjacent SNPs at the whole-genome level and were concentrated in subtelomeric regions. A similar LD decay rate was reported by Jia et al. (2013), where the genome-wide LD decay rate was ~100 kb on average. These decay rates in foxtail millet are similar to those in rice, another self-fertilizing species (Huang et al., 2010). To test the utility of this SNP data set, we performed GWAS on plant pigmentation and days to 50% flowering; both being highly investigated traits in many crop species including maize, rice and *Arabidopsis* (Nesi et al., 2000; Heim et al., 2003; Huang et al., 2012; Komeda, 2004). We identified genomic regions and SNP loci associated with plant pigmentation and flowering time. Putative candidate genes colocalized with significant SNPs, indicating that our SNPs dataset can be used for GWAS of many traits for foxtail millet improvement.

Conclusions

This study reports genome-wide SNP variation in a set of 181 foxtail millet germplasm including 155 accessions of the foxtail millet core collection. The SNP data set generated here could serve as a resource for further investigation on GWAS in foxtail millet. For example, the foxtail millet core collection included in this study has wide genetic variation for resistance to blast (Sharma et al., 2014), salinity tolerance (Krishnamurthy et al., 2014), and agronomic and grain nutritional traits (Upadhyaya et al., 2008, 2011). Associating sequence variation with such data will result in identification of novel alleles and genotypes for marker-assisted breeding to improve foxtail millet yield and adaptation to various stresses. This germplasm set is available to the global research community via the standard material transfer agreement (SMTA). As of now, we have distributed a total of 23 core and mini core-sets across 13 countries.

Supplemental Information Available

Supplemental information is available with the online version of this article.

Supplemental Figure S1. Distance between single nucleotide polymorphisms (SNPs) used in this study. The *x*-axis represents base pair distances between adjacent SNPs. The *y*-axis represents the total number of SNPs that fall within a specific range. The *z*-axis represents the percentage of total SNPs that fall within a specific range.

Acknowledgments

This work was supported by NSF grants DBI-0820619 and IOS-1238014, ICRISAT, and the USDA–ARS. This work has been undertaken as part of the CGIAR Research Program on Dryland Cereals.

References

Abe, M., Y. Kobayashi, S. Yamamoto, Y. Daimon, A. Yamaguchi, Y. Ikeda, H. Ichinoki, M. Notaguchi, K. Goto, and T. Araki. 2005. FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex. Science 309:1052–1056. doi:10.1126/science.1115983

Alexander, D.H., J. Novembre, and K. Lange. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19:1655–1664. doi:10.1101/gr.094052.109

Anju, T., and S. Sarita. 2010. Suitability of foxtail millet (*Setaria italica*) and barnyard millet (*Echinochloa frumentacea*) for development of low glycemic index biscuits. Malays. J. Nutr. 16:361–368.

Austin, D.F. 2006. Fox-tail Millets (*Setaria*:Poaceae): Abandoned food in two hemispheres. Econ. Bot. 60:143–158. doi:10.1663/0013-0001(2006)60[143:FMSPFI]2.0.CO;2

Bennetzen, J.L., J. Schmutz, H. Wang, R. Percifield, J. Hawkins, A.C. Pontaroli, M. Estep, L. Feng, J.N. Vaughn, J. Grimwood, J. Jenkins, K. Barry, E. Lindquist, U. Hellsten, S. Deshpande, X. Wang, X. Wu, T. Mitros, J. Triplett, X. Yang, C.Y. Ye, M. Mauro-Herrera, L. Wang, P. Li, M. Sharma, R. Sharma, P.C. Ronald, O. Panaud, E.A. Kellogg, T.P. Brutnell, A.N. Dust, G.A. Tuskan, D. Rokhsar, and K.M. Devos. 2012. Reference genome sequence of the model plant Setaria. Nat. Biotechnol. 30:555–561. doi:10.1038/nbt.2196

Bradbury, P.J., Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss, and E.S. Buckler. 2007. TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635. doi:10.1093/bioinformatics/btm308

Cingolani, P., A. Platts, L.L. Wang, M. Coon, T. Nguyen, L. Wang, S.J. Land, X. Lu, and D.M. Ruden. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w^{1118}; *iso-2*; *iso-3*. Fly (Austin) 6:80–92. doi:10.4161/fly.19695

Doust, A.N., E.A. Kellogg, K.M. Devos, and J.L. Bennetzen. 2009. Foxtail Millet : A sequence-driven grass model system. Plant Physiol. 149:137–141. doi:10.1104/pp.108.129627

Dwivedi, S., H.D. Upadhyaya, S. Senthilvel, C.T. Hash, K. Fukunaga, X. Diao, D. Santra, D. Baltensperger, and M. Prasad. 2012. Millets: Genetic and Genomic Resources. Plant Breed. Rev. 35:247–375.

Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, and S.E. Mitchell. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6:E19379. doi:10.1371/journal.pone.0019379

Excoffier, L., and H.E.L. Lischer. 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources 10:564–567. doi:10.1111/j.1755-0998.2010.02847.x

Glaubitz, J.C., T.M. Casstevens, F. Lu, J. Harriman, R.J. Elshire, Q. Sun, and E.S. Buckler. 2014. TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. PLoS ONE 9:E90346. doi:10.1371/journal.pone.0090346

Gupta, S., K. Kumari, M. Muthamilarasan, S.K. Parida, and M. Prasad. 2014. Population structure and association mapping of yield contributing agronomic traits in foxtail millet. Plant Cell Rep. 33:881–893. doi:10.1007/s00299-014-1564-0

Hepworth, S.R., J.E. Klenz, and G.W. Haughn. 2006. UFO in the Arabidopsis inflorescence apex is required for floral-meristem identity and bract suppression. Planta 223:769–778. doi:10.1007/s00425-005-0138-3

Heim, M.a., M. Jakoby, M. Werber, C. Martin, B. Weisshaar, and P.C. Bailey. 2003. The basic helix-loop-helix transcription factor family in plants: A genome-wide study of protein structure and functional diversity. Mol. Biol. Evol. 20:735–747. doi:10.1093/molbev/msg088

Himi, E., M. Maekawa, and K. Noda. 2011. Differential expression of three flavanone 3-hydroxylase genes in grains and coleoptiles of wheat. Int. J. Plant Genomics 2011:369460. doi:10.1155/2011/369460

Huang, X., X. Wei, T. Sang, Q. Zhao, Q. Feng, Y. Zhao, C. Li, C. Zhu, T. Lu, Z. Zhang, M. Li, D. Fan, Y. Guo, A. Wang, L. Wang, L. Deng, W. Li, Y. Lu, Q. Weng, K. Liu, T. Huang, T. Zhou, Y. Jing, W. Li, Z. Lin, E.S. Buckler, Q. Qian, Q.F. Zhang, J. Li, and B. Han. 2010. Genome-wide association studies of 14 agronomic traits in rice landraces. Nat. Genet. 42:961–967. doi:10.1038/ng.695

Huang, X., Y. Zhao, X. Wei, C. Li, A. Wang, Q. Zhao, W. Li, Y. Guo, L. Deng, C. Zhu, D. Fan, Y. Lu, Q. Weng, K. Liu, T. Zhou, Y. Jing, L. Si, G. Dong, T. Huang, T. Lu, Q. Feng, Q. Qian, J. Li, and B. Han. 2012. Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. Nat. Genet. 44:32–39. doi:10.1038/ng.1018

Ikeda, K., M. Ito, N. Nagasawa, J. Kyozuka, and Y. Nagato. 2007. Rice ABERRANT PANICLE ORGANIZATION 1, encoding an F-box protein, regulates meristem fate. Plant J. 51:1030–1040. doi:10.1111/j.1365-313X.2007.03200.x

Jia, G., X. Huang, H. Zhi, Y. Zhao, Q. Zhao, W. Li, Y. Chai, L. Yang, K. Liu, H. Lu, C. Zhu, Y. Lu, C. Zhou, D. Fan, Q. Weng, Y. Guo, T. Huang, L. Zhang, T. Lu, Q. Feng, H. Hao, H. Liu, P. Lu, N. Zhang, Y. Li, E. Guo, S. Wang, S. Wang, J. Liu, W. Zhang, G. Chen, B. Zhang, W. Li, Y. Wang, H. Li, B. Zhao, J. Li, X. Diao, and B. Han. 2013. A haplotype map of genomic variations and genome-wide association studies of agronomic traits in foxtail millet (*Setaria italica*). Nat. Genet. 45:957–961. doi:10.1038/ng.2673

Komeda, Y. 2004. Genetic regulation of time to flower in *Arabidopsis thaliana*. Annu. Rev. Plant Biol. 55:521–535. doi:10.1146/annurev.arplant.55.031903.141644

Krishnamurthy, L., H.D. Upadhyaya, C.L.L. Gowda, J. Kashiwagi, R. Purushothaman, S. Singh, and V. Vadez. 2014. Large variation for salinity tolerance in the core collection of foxtail millet (*Setaria italica* (L.) P. Beauv.) germplasm. Crop Pasture Sci. 65:353–361. doi:10.1071/CP13282

Langmead, B., and S. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9:357–359. doi:10.1038/nmeth.1923

Lata, C., S. Gupta, and M. Prasad. 2013. Foxtail millet: A model crop for genetic and genomic studies in bioenergy grasses. Crit. Rev. Biotechnol. 33:328–343. doi:10.3109/07388551.2012.716809

Lipka, A.E., F. Tian, Q. Wang, J. Peiffer, M. Li, P.J. Bradbury, M.A. Gore, E.S. Buckler, and Z. Zhang. 2012. GAPIT: Genome association and prediction integrated tool. Bioinformatics 28:2397–2399. doi:10.1093/bioinformatics/bts444

Mace, E.S., H.K. Buhariwalla, and J.H. Crouch. 2003. A high-throughput DNA extraction protocol for tropical molecular breeding programs. Plant Mol. Biol. Rep. 21:459a–459h. doi:10.1007/BF02772596

Morris, G.P., P. Ramu, S.P. Deshpande, C.T. Hash, T. Shah, H.D. Upadhyaya, O. Riera-Lizarazu, P.J. Brown, C.B. Acharya, S.E. Mitchell, J. Harriman, J.C. Glaubitz, E.S. Buckler, and S. Kresovich. 2013. Population genomic and genome-wide association studies of agroclimatic traits in sorghum. Proc. Natl. Acad. Sci. USA 110:453–458. doi:10.1073/pnas.1215985110

Muthamilarasan, M., and M. Prasad. 2014. Advances in *Setaria* genomics for genetic improvement of cereals and bioenergy grasses. Theor. Appl. Genet. 128:1–14. doi:10.1007/s00122-014-2399-3

Nesi, N., I. Debeaujon, C. Jond, G. Pelletier, M. Caboche, and L. Lepiniec. 2000. The TT8 gene encodes a basic helix-loop-helix domain protein required for expression of DFR and BAN genes in *Arabidopsis* siliques. Plant Cell 12:1863–1878. doi:10.1105/tpc.12.10.1863

Perrier, X., and J.P. Jacquemoud-Collet. 2006. DARwin software, version 6. Cirad, Montpellier, France. http://darwin.cirad.fr/ .

Prasada Rao, K.E., J.M.J. de Wet, D.K. Brink, and M.H. Mengesha. 1987. Intraspecific variation and systematics of cultivated *Setaria italica*, foxtail millet (Poaceae). Econ. Bot. 41:108–116. doi:10.1007/BF02859358

Qi, X., S. Xie, Y. Liu, F. Yi, and J. Yu. 2013. Genome-wide annotation of genes and noncoding RNAs of foxtail millet in response to simulated drought stress by deep sequencing. Plant Mol. Biol. 83:459–473. doi:10.1007/s11103-013-0104-6

Reynolds, J., B.S. Weir, and C.C. Cockerham. 1983. Estimation for the coancestry coefficient: Basis for a short-term genetic distance. Genetics 105:767–779.

Saleh, A.S.M., Q. Zhang, J. Chen, and Q. Shen. 2013. Millet grains: nutritional quality, processing, and potential health benefits. Compr. Rev. Food Sci. Food Saf. 12:281–295. doi:10.1111/1541-4337.12012

Samach, A., J.E. Klenz, S.E. Kohalmi, E. Risseeuw, G.W. Haughn, and W.L. Crosby. 1999. The UNUSUAL FLORAL ORGANS gene of *Arabidopsis thaliana* is an F-box protein required for normal patterning and growth in the floral meristem. Plant J. 20:433–445. doi:10.1046/j.1365-313x.1999.00617.x

Sharma, R., A.G. Girish, H.D. Upadhyaya, P. Humayun, T.K. Babu, V.P. Rao, and R.P. Thakur. 2014. Identification of blast resistance in a core collection of foxtail millet germplasm. Plant Dis. 98:519–524. doi:10.1094/PDIS-06-13-0593-RE

Spelt, C., F. Quattrocchio, J.N.M. Mol, and R. Koes. 2000. *Anthocyanin1* of petunia encodes a basic helix–loop–helix protein that directly activates transcription of structural anthocyanin genes. Plant Cell 12:1619–1631. doi:10.1105/tpc.12.9.1619

Tamura, K., M. Nei, and S. Kumar. 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc. Natl. Acad. Sci. USA 101:11030–11035. doi:10.1073/pnas.0404206101

Tamura, K., G. Stecher, D. Peterson, A. Filipski, and S. Kumar. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol. Biol. Evol. 30:2725–2729. doi:10.1093/molbev/mst197

Upadhyaya, H.D., R.P.S. Pundir, C.L.L. Gowda, V.G. Reddy, and S. Singh. 2008. Establishing a core collection of foxtail millet to enhance the utilization of germplasm of an underutilized crop. Plant Genet. Resour. Charact. Util. 7:177–184. doi:10.1017/S1479262108178042

Upadhyaya, H.D., C.R. Ravishankar, Y. Narasimhudu, N.D.R.K. Sarma, S.K. Singh, S.K. Varshney, V.G. Reddy, S. Singh, H.K. Parzies, S.L. Dwivedi, H.L. Nadaf, K.L. Sahrawat, and C.L.L. Gowda. 2011. Identification of trait-specific germplasm and developing a mini core collection for efficient use of foxtail millet genetic resources in crop improvement. F. Crop. Res. 124:459–467. doi:10.1016/j.fcr.2011.08.004

van Hintum, T.J.L., C.C.M. van de Wiel, D.L. Visser, R. van Treuren, and B. Vosman. 2007. The distribution of genetic diversity in a *Brassica oleracea* gene bank collection related to the effects on diversity of regeneration, as measured with AFLPs. Theor. Appl. Genet. 114:777–786. doi:10.1007/s00122-006-0456-2

van Treuren, R., and T.J.L. van Hintum. 2014. Next-generation gene-banking: Plant genetic resources management and utilization in the sequencing era. Plant Genet. Resour. 12:298–307. doi:10.1017/S1479262114000082

Vetriventhan, M. 2011. Phenotypic and genetic diversity in the foxtail millet (*Setaria italica* (L.) P. Beauv.) core collection. Ph.D. thesis. Agricultural College and Research Institute, Tamil Nadu Agricultural University, Madurai, Tamil Nadu, India.

Vetriventhan, M., H.D. Upadhyaya, C.R. Anandakumar, S. Senthilvel, H.K. Parzies, A. Bharathi, R.K. Varshney, and C.L.L. Gowda. 2012. Assessing genetic diversity, allelic richness and genetic relationship among races in ICRISAT foxtail millet core collection. Plant Genet. Resour. Charact. Util. 10:214–223. doi:10.1017/S1479262112000287

Vetriventhan, M., H.D. Upadhyaya, C.R. Anandakumar, S. Senthilvel, R.K. Varshney, and H.K. Parzies. 2014. Population structure and linkage disequilibrium of ICRISAT foxtail millet (*Setaria italica* (L.) P. Beauv.) core collection. Euphytica 196:423–435. doi:10.1007/s10681-013-1044-6

Wallace, J.G., H.D. Upadhyaya, M. Vetriventhan, E.S. Buckler, C. Tom Hash, and P. Ramu. 2015. The genetic makeup of a global barnyard millet germplasm collection. Plant Gen. 8:7. doi:10.3835/plantgenome2014.10.0067

Wang, C., G. Jia, H. Zhi, Z. Niu, Y. Chai, W. Li, Y. Wang, H. Li, P. Lu, B. Zhao, and X. Diao. 2012. Genetic Diversity and Population Structure of Chinese Foxtail Millet [*Setaria italica* (L.) Beauv.] Landraces. G3: Genes, Genomes, Genet. 2:769–777.

Weiner, J. 2014. pca3d: Three dimensional PCA plots. R package version 0.3. http://CRAN.R-project.org/package=pca3d (accessed 13 Feb. 2015).

Winkel-Shirley, B. 2001. Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. Plant Physiol. 126:485–493. doi:10.1104/pp.126.2.485

Yang, X., Z. Wan, L. Perry, H. Lu, Q. Wang, C. Zhao, J. Li, F. Xie, J. Yu, T. Cui, T. Wang, M. Li, and Q. Ge. 2012. Early millet use in northern China. Proc. Natl. Acad. Sci. USA 109:3726–3730. doi:10.1073/pnas.1115430109

Zhang, Z., E. Ersoz, C.-Q. Lai, R.J. Todhunter, H.K. Tiwari, M.A. Gore, P.J. Bradbury, J. Yu, D.K. Arnett, J.M. Ordovas, and E.S. Buckler. 2010. Mixed linear model approach adapted for genome-wide association studies. Nat. Genet. 42:355–360. doi:10.1038/ng.546

Zhang, Y., J. Wang, W. Wang, P. Zeng, C. Han, Z. Quan, X. Peng, H. Li, Q. Xia, J. Wang, T. Guo, Y. Li, G. Zhang, H. Li, P. Huang, Z. Yue, X. Yang, R. Cao, C. Wang, J. Hu, H. Xiang, M. Xie, C. Liu, N. Li, S. Cheng, D. Zhan, R. Wang, S. Pan, Q. Chen, J. Wang, X. Xu, Z. Zhao, Y. Tao, X. Liu, C. Bian, Q. Shi, and Y. Cai. 2012. Genome sequence of foxtail millet (*Setaria italica*) provides insights into grass evolution and biofuel potential. Nat. Biotechnol. 30:549–554. doi:10.1038/nbt.2195

Zheng, X., D. Levine, J. Shen, S.M. Gogarten, C. Laurie, and B.S. Weir. 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics 28:3326–3328. doi:10.1093/bioinformatics/bts606

Zhou, A., H. Wang, J.C. Walker, and J. Li. 2004. BRL1, a leucine-rich repeat receptor-like protein kinase, is functionally redundant with BRI1 in regulating *Arabidopsis* brassinosteroid signaling. Plant J. 40:399–409. doi:10.1111/j.1365-313X.2004.02214.x