Genetic diversity, population structure, and selection signature in Ethiopian sorghum [*Sorghum bicolor* L. (Moench)] germplasm

Zeleke Wondimu,[1] Hongxu Dong,[2] Andrew H. Paterson,[2] Walelign Worku,[3] and Kassahun Bantte[1],*

[1]College of Agriculture and Veterinary Medicine, Jimma University, P.O. Box 307, Jimma, Ethiopia
[2]Plant Genome Mapping Laboratory, University of Georgia, Athens, GA 30602, USA
[3]College of Agriculture, Hawassa University, P.O. Box 05, Hawassa, Ethiopia

*Corresponding author: Department of Horticulture and Plant Sciences, College of Agriculture and Veterinary Medicine, Jimma University, PO Box 307, Jimma, Ethiopia. Tel: +251-917-12 832801, E-mail: kassahunb@gmail.com

Abstract

Ethiopia, the probable center of origin and diversity for sorghum [*Sorghum bicolor* L. (Moench)] and with unique ecogeographic features, possesses a large number of sorghum landraces that have not been well studied. Increased knowledge of this diverse germplasm through large-scale genomic characterization may contribute for understanding of evolutionary biology, and adequate use of these valuable resources from the center of origin. In this study, we characterized genetic diversity, population structure and selection signature in 304 sorghum accessions collected from diverse sorghum growing regions of Ethiopia using genotyping-by-sequencing. We identified a total of 108,107 high-quality single-nucleotide polymorphism (SNPs) markers that were evenly distributed across the sorghum genome. The average gene diversity among accessions was high ($H_e = 0.29$). We detected a relatively low frequency of rare alleles (26%), highlighting the potential of this germplasm for subsequent allele mining studies through genome-wide association studies. Although we found no evidence of genetic differentiation among administrative regions ($F_{ST} = 0.02$, $P = 0.12$), population structure and cluster analyses showed clear differentiation among six Ethiopian sorghum populations ($F_{ST} = 0.28$, $P = 0.01$) adapting to different environments. Analysis of SNP differentiation between the identified genetic groups revealed a total of 40 genomic regions carrying signatures of selection. These regions harbored candidate genes potentially involved in a variety of biological processes, including abiotic stress tolerance, pathogen defense and reproduction. Overall, a high level of untapped diversity for sorghum improvement remains available in Ethiopia, with patterns of diversity consistent with divergent selection on a range of adaptive characteristics.

Keywords: genetic diversity; population structure; selection signature; sorghum

Introduction

Sorghum [*Sorghum bicolor* L. (Moench)], native to the dry regions of northeast Africa (Dahlberg and Wasylikowa 1996), is a major food crop in the arid and semi-arid regions of the world (Balota *et al.* 2008). It is a highly diverse crop that has experienced multiple domestication processes, resulting in five major races differentiated by inflorescence type (Harlan and de Wet 1972). Ethiopia, one of Vavilov's centers of origin for several crop species (Vavilov 1951), hosts wide genetic variability for sorghum; all races of sorghum and their corresponding intermediates are cultivated across the country's diverse agro-ecological zones and farming systems (Doggett 1988; Teshome *et al.* 1997; Ayana and Bekele 1998).

The wealth of genetic variability in the Ethiopian sorghum germplasm has already been noted worldwide as sources of desirable genes for sorghum improvement (Singh and Axtell 1973; Schertz 1977; Reddy *et al.* 2009). In addition, due to its unique ecogeographic features, Ethiopia possesses a large number of sorghum landraces in the gene bank as well as under subsistence agriculture. These landraces have evolved by the interaction between adaptation to a wide range of environments and selection imposed by farmers for traits enhancing agricultural productivity and performance, such as high yield, and resistance to biotic and abiotic stresses. Consequently, the genome of sorghum landraces might have experienced strong selection at genes controlling traits of agronomic and adaptive importance since domestication. Therefore, assessing genetic diversity, population structure, and selection signatures is meaningful from the perspectives of improving adequate use and conservation of these valuable resources, and may provide insights into evolutionary genomics.

Previously, genetic diversity of Ethiopian sorghum germplasm was studied using agro-morphological traits (Gebeyehu 1993; Teshome *et al.* 1997; Ayana and Bekele 1998, 2000; Desmae *et al.* 2016b). However, this approach may not give reliable estimates of genetic diversity as these traits are limited in number and subjected to strong environmental influences (van Beuningen and Busch 1997). Genetic diversity analyses have also been carried

out using various DNA marker techniques such as random amplified polymorphic DNA (Ayana *et al.* 2000), amplified fragment length polymorphisms (Geleta *et al.* 2006), simple sequence repeats (Cuevas and Prom 2013; Adugna 2014; Desmae *et al.* 2016a; Weerasooriya *et al.* 2016), and Inter-simple sequence repeats (Desmae 2007). Although these studies generated useful information that is relevant to both plant breeding and germplasm conservation efforts, they were either focused on samples collected from a limited geographic range (Geleta *et al.* 2006; Desmae *et al.* 2016a), or involved limited numbers of markers (Ayana *et al.* 2000; Cuevas and Prom 2013; Adugna 2014; Weerasooriya *et al.* 2016) that are too small to fully reflect the breadth of genetic diversity that exist in the country. As a result, detailed information on genetic diversity and population structure of cultivated sorghum using reliable marker systems, while indispensible, is lacking in the center of origin, Ethiopia.

Several studies on sorghum (Hamblin *et al.* 2004; Casa *et al.* 2005; Frere *et al.* 2011; Bouchet *et al.* 2012; Mace *et al.* 2013; Morris *et al.* 2013; Zhang *et al.* 2015; Campbell *et al.* 2016; Cuevas *et al.* 2017; Tao *et al.* 2017) have utilized selective sweep analysis to detect genomic regions and genes affected by natural and artificial selection. However, most of these studies had certain limitations, either they were based on limited genome coverage (Hamblin *et al.* 2004; Casa *et al.* 2005; Frere *et al.* 2011; Bouchet *et al.* 2012) or used sorghum germplasm that have gone through the sorghum conversion program (Morris *et al.* 2013; Zhang *et al.* 2015; Cuevas *et al.* 2017). Nevertheless, these converted sorghum lines (*i.e.*, short, early maturity, and photoperiod insensitive) that are adapted to temperate regions represent partial of the genetic diversity in breeding programs, the diversity underlying traits of economic and adaptive importance remains trapped within the tropical germplasm (Cuevas and Prom 2020). Thus, characterization of the Ethiopian germplasm at genome-wide scale based on patterns of nucleotide variation and selection signature will improve conservation efforts and its utilization in research and breeding programs.

Next-generation sequencing technologies have made important contributions to the development of new genotyping platforms. Genotyping-by-sequencing (GBS) is increasingly being used for profiling genome-wide nucleotide variation in many species (Elshire *et al.* 2011). The inherent characteristics of GBS including genome-wide molecular marker discovery, highly multiplexed genotyping, flexibility and low cost make it an excellent tool in genomic analysis of diverse populations, including genome-wide association studies and genomic signatures of selection (Deschamps *et al.* 2012; Poland and Rife 2012; Morris *et al.* 2013).

In this study, we used a high-throughput GBS approach to generate whole-genome profiles and high-quality single-nucleotide polymorphism (SNP) markers in a collection of 304 sorghum accessions. The objectives of this study were to (1) assess the extent and patterns of genetic diversity among sorghum accessions collected from major sorghum growing regions of Ethiopia, (2) determine the population structure of the accessions, and explore their potential for future genome-wide association studies, and (3) identify genomic regions and genes potentially subjected to selection.

Materials and methods
Plant materials
A total of 304 sorghum accessions used in this study were collected from farmers' fields of major sorghum growing administrative regions of Ethiopia (see Supplementary File S1). Accessions

from regions with sample size <10 were included in adjacent regions to reduce bias due to small sample size. In doing so, six, two, and two accessions from Gambella, Afar, and Somali regions were, respectively, placed under the Southern Nations, Amhara, and Oromia regions. This reduced the seven regions from which the accessions were originally collected to four major sorghum producing regions (Amhara, Oromia, Tigray, and South Nations). These regions include a broad swath of the range of sorghum cultivation that account for 94% of the total sorghum production in the country (Central Statistical Agency, 2018). During collection readings of the coordinates and altitudes of the collection sites were recorded by a Global Positioning System (GPS) map 60CSx GPS (Garmin), which were overlaid on to the maps of Ethiopia (Figure 1).

DNA extraction and GBS
Prior to DNA extraction, the accessions were grown under field conditions and subjected to one cycle of controlled self-fertilization for purification. Leaf samples from a single representative plant per accession were collected from 15-day-old plants grown in small pots in a greenhouse. DNA was then extracted from lyophilized leaf tissues following a modified cetyltrimethyl ammonium bromide protocol (Mace *et al.* 2003). A total of four 96-plex GBS libraries were constructed and genotyped at the University of Georgia, Genomics and Bioinformatics Core Facility. The GBS procedure (Elshire *et al.* 2011) was implemented using the *Ape*KI enzyme system. In brief, each DNA sample was digested with *Ape*KI (recognition site: G|CWCG; New England Biolabs Inc., Ipswich, MA, USA), then ligated to a unique barcoded adapter. For each library, 96 samples were pooled, and fragments with 200–500 base pair (bp) in length were extracted from a 2% agarose gel after electrophoresis and purified using a Qiagen Gel Extraction Kit (Qiagen, Hilden, Germany). The purified DNA was PCR amplified using GoTaq Colorless Master Mix (Promega, Madison, WI, USA), and the PCR product was extracted as above to eliminate primer–dimers. All libraries were sequenced on a NextSeq platform (Illumina, San Diego, CA, USA) with 150 bp single-end reads.

SNP calling and quality control
SNP calling was performed using the TASSEL GBS pipeline (Bradbury *et al.* 2007) with the following parameters: kmer length of 100 bp, minimum quality score of 10, minimum call rate of 0.5, and minor allele frequency (MAF) of 0.01. Physical positions of generated SNPs were obtained based on alignment to the *Sorghum bicolor* reference genome v1.4 (Paterson *et al.* 2009). Missing data were imputed with Beagle V4.0 (Browning and Browning 2007).

Analysis of population structure
Two approaches were used to describe the population structure of the Ethiopian sorghum collection. First, hierarchical population structure was assessed with a model-based estimation of admixed ancestry using the ADMIXTURE program (Alexander *et al.* 2009). To determine the optimal number of subpopulations (K), ADMIXTURE was run with a fivefold cross-validation (CV) procedure for K ranging from 1 to 20, and the K value with the lowest CV error was selected (Alexander *et al.* 2009). Second, pairwise genetic distances among individuals were calculated using the Sokal and Michener dissimilarity index (Sokal and Michener 1958). The resulting distance matrix was then subjected to a clustering analysis using a Neighbor-Joining (NJ) tree with 1000 bootstraps as implemented in DARwin 6.0.14 (Perrier and Jacquemoud-Collet 2006). To further investigate the spatial pattern of genetic
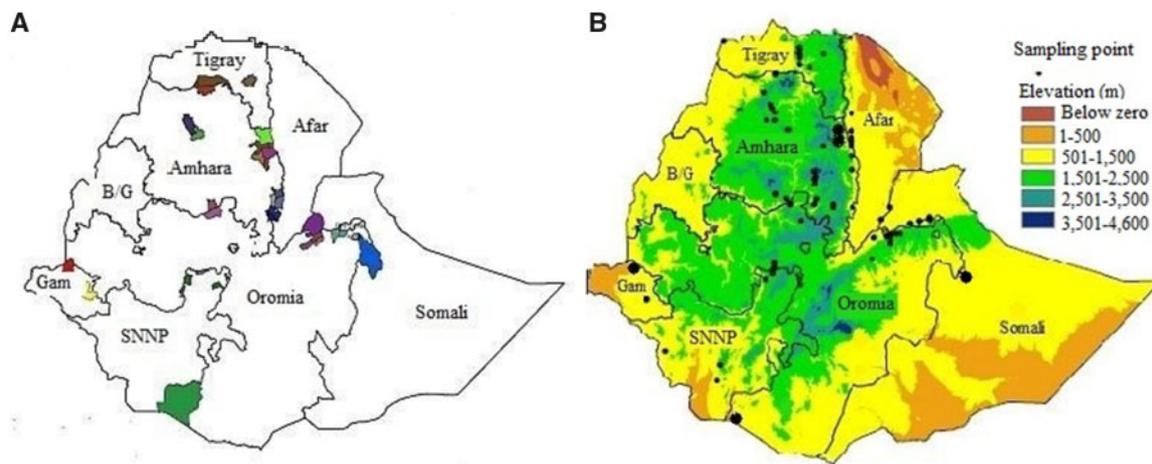
Figure 1 Distribution of Ethiopian sorghum accessions. (A) Geographic distribution of geo-referenced Ethiopian sorghum accessions. Color highlighted areas represent accession collection sites within regions. (B) Geographic distribution of the Ethiopian sorghum accessions along elevation gradient. B/G, Benishangul-Gumuz; Gam, Gambella; SNNP, Southern Nations.

diversity, R package tess3r was used to perform the spatial interpolation of ancestry coefficients structure onto the Ethiopian geographical map (Caye *et al.* 2016). Ancestry coefficients (*q*) estimated with ADMIXTURE program using the optimum number of subpopulations ($K = 6$) suggested by CV procedure (Alexander *et al.* 2009) were used to explore genome relatedness among 304 sorghum accessions to the stated locations of origin in Ethiopia.

Genetic diversity and population differentiation

For each SNP, the number and frequency of alleles was calculated using TASSEL 5.0 (Bradbury *et al.* 2007). To determine the extent of genetic diversity among individuals of the entire panel, effective number of alleles (N_E), observed heterozygosity (H_o), gene diversity (H_e, *i.e.*, expected heterozygosity) and polymorphism information content (PIC) were estimated using the allele frequencies of each SNP. The above genetic diversity estimates were also computed for pooled accessions within each administrative region and ADMIXTURE inferred subpopulation. However, a comparison of diversity estimates in big populations compared that in small populations could be largely biased by the different sample sizes. To account for differences in population size, we used a subsampling scheme by taking into account the required level of precision ($\alpha = 0.05$), the variances and average differences in allele frequencies between populations (Miaoulis and Michener 1976). This procedure indicated that a subsample size of 25 and 30 accessions randomly selected from each ADMIXTURE and regional population, respectively, would be appropriate to obtain unbiased estimates of the above genetic parameters (*i.e.*, N_E, H_o, H_e, and PIC) for each population. In addition, allelic richness (R_s) and number of private alleles for each population were computed with the rarefaction method, that adjusts for differences in sample sizes across populations (Hulbert 1971), using the PopGenReport package in R (Adamack and Gruber 2014).

To estimate the components of variance among and within populations, analysis of molecular variance (AMOVA) was performed as described in Excoffier *et al.* (1992) using the R package Hierfstat (Goudet 2005). To investigate population differentiation, pairwise fixation index (F_{ST}) among populations was estimated based on the method of Weir and Cockerham (1984) using the same package. Gene flow among populations was also estimated

using indirect method based on the number of migrants per generation (N_m) as $(1 - F_{ST})/4F_{ST}$ as described by Wright (1965).

Detection of F_{ST} outliers

To detect signatures of selection among ADMIXTURE subpopulations, F_{ST} outliers were detected based on SNPs with MAF > 5% using BayeScan 2.1 (Foll and Gaggiotti 2008). To reduce the identification of false positives, a 50,000-iteration burn-in period and thinning interval size of 10 were used. The prior odd threshold to identify F_{ST} outlier SNPs was determined using a false discovery rate (FDR) of 0.05 as implemented in the "plot_bayescan" function in R. Genes found within 100 kb of the genomic regions detected in the above test were also searched using the most recently annotated version of the sorghum genome v3.1 (www.phytozome. net). The distance 100 kb was based on the average genome-wide linkage disequilibrium decay of 100 kb (data not shown).

Data availability

Supplementary File S1 contains detailed descriptions of Ethiopian sorghum accessions, their regions of origin and geographic information. Supplementary File S2 contains SNP ID numbers, locations and SNP genotypes for all accessions and SNPs. Supplementary File S3 contains co-ancestry coefficient matrix of 304 Ethiopian sorghum accessions based on ADMIXTURE analysis at $K = 6$. Supplementary File S4 contains detailed description of F_{ST} outlier SNPs and candidate genes located in the vicinity of these SNPs based on BAYESCAN results for outlier prediction. Supplementary Figure S1 contains genomic distribution of 108,107 high-quality SNPs across the 10 sorghum chromosomes, and their corresponding density. Supplementary Table S1 contains AMOVA and F_{ST} test results for accessions by geographic/administrative region and ADMIXTURE subgroup analyses.

Supplementary material is available at figshare DOI: https://doi.org/10.25387/g3.12813224.

Results

SNP discovery

A total of 350,618,420 reads were generated after sequencing of GBS libraries from 304 sorghum accessions. After de-duplication and alignment of unique sequence tags to the reference sorghum genome v1.4 (Paterson *et al.* 2009), a total of 236,000 SNPs were

called using the GBS pipeline in TASSEL 5 (Bradbury *et al.* 2007). The quality control of SNP data (see 'Materials and Methods' for criteria) produced a total of 115,501 high-quality SNPs (see Supplementary File S2). Overall, 32.81% of the SNP calls were missing and imputed using Beagle V4.0 (Browning and Browning 2007). We further retained SNPs with MAF > 0.01 for downstream analysis.

A genome-wide SNP density plot (see Supplementary Figure S1) revealed that the highest number of these SNPs were physically mapped on chromosome 2 (12.09%, 13,075 SNPs). The highest and lowest marker densities were observed on chromosome 7 (7.41 kb) and chromosome 5 (5.31 kb), respectively, with an average marker density of 6.13 kb per chromosome. The identified SNPs were also categorized according to nucleotide substitutions as either transitions (A↔G or C↔T) or transversions (A↔C, C↔G, A↔T, and G↔T). Our analysis of transitions (Ts) and transversions (Tv) SNPs showed a Ts/Tv ratio of 1.7:1 (*i.e.*, 68,097/40,010;

Table 1), which is very close to the expected 2:1 ratio of neutral variants (Siol *et al.* 2010). The observed transition bias could be caused by a mutational bias due to intrinsic properties of DNA (*e.g.*, cytosine deamination) in plant genomes (Gaut *et al.* 2011). Although this result suggest that most of the single-nucleotide mutations observed in this study are nearly neutral, we expect that some SNPs are likely to be under selection, and thus may not susceptible to mutation bias.

Population structure

ADMIXTURE analysis using a fivefold CV for $K=1$ to $K=20$ indicated a steep decrease in CV error values until $K=6$ (Figure 2A). For example, CV error at $K=1$ was 0.57937, at $K=6$ was 0.41739, at $K=8$ was 0.41433, at $K=9$ was 0.41027, indicating that there is no steep decrease in CV error values after $K=6$. Given the modest population size in this study, we chose $K=6$ as an optimal number of subpopulations, referred to as subgroups, SG-I to SG-VI (Figure 2B). Although comparing the two methods showed that there were few accessions that clustered differently depending on the analysis method, overall the clustering pattern generated using a NJ tree (Figure 3A), also supported the possibility that the Ethiopian sorghum collection evaluated in this study has six ($K=6$) well-differentiated genetic groups and some admixtures. Therefore based on ADMIXTURE analysis, we assigned 234 (77%) accessions to one of the six subgroups with an ancestry membership coefficient probability of >0.60 ($q > 0.60$), whereas the remaining 23%

**Table 1** Percentage of transition and transversion SNPs identified using GBS

	Transition		Transversion			
	A/G	C/T	A/C	A/T	G/C	G/T
No. of allelic sites	34,166	33,931	9,321	7,879	13,310	9,500
% of allelic sites	31.60	31.40	8.60	7.30	12.31	8.78
Total	68,097		40,010			
Percentage	63.00		37.00			
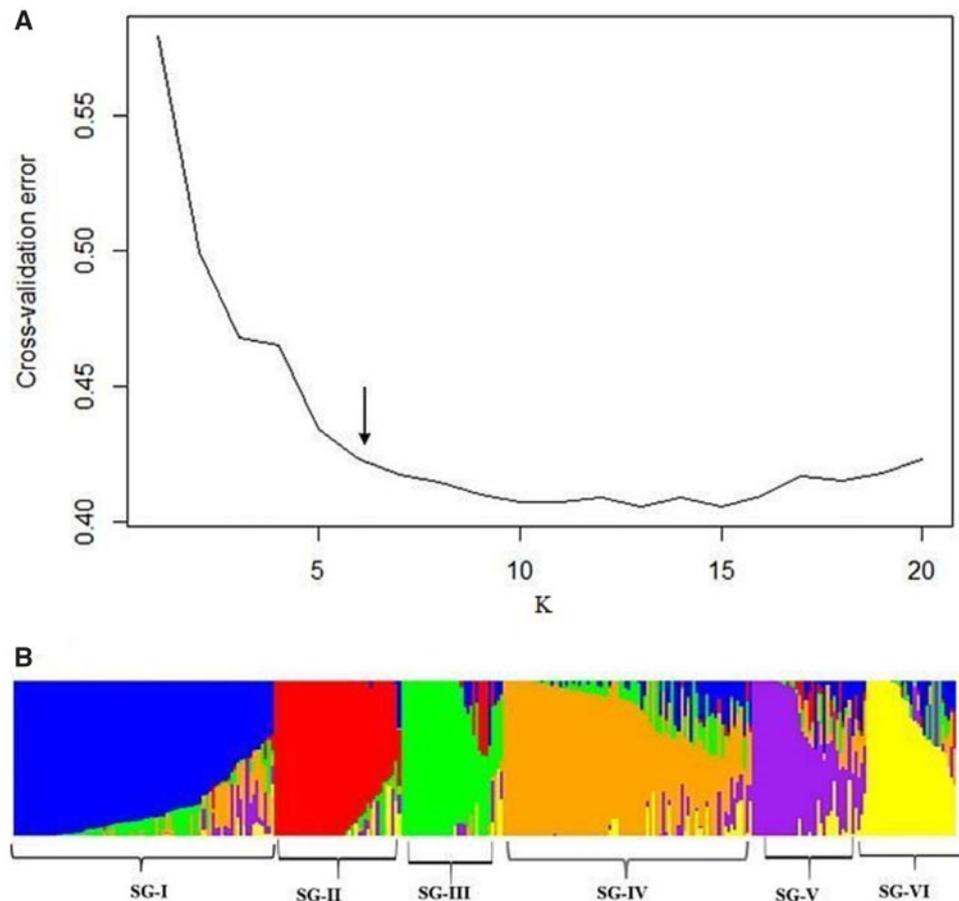
Figure 2 Population structure analysis of 304 Ethiopian sorghum accessions using 108,107 SNPs. (A) The CV error (y-axis) for *K* values from 1 to 20 (x-axis) decreased steeply until it reached 6 (arrow), suggesting an optimal number of subgroups at $K=6$. (B) Bar-plot describing the population structure estimated from ADMIXTURE analysis at $K=6$. Color-coding of Q-value bar plots is arbitrary. SG-I to SG-VI represents subgroups 1–6.
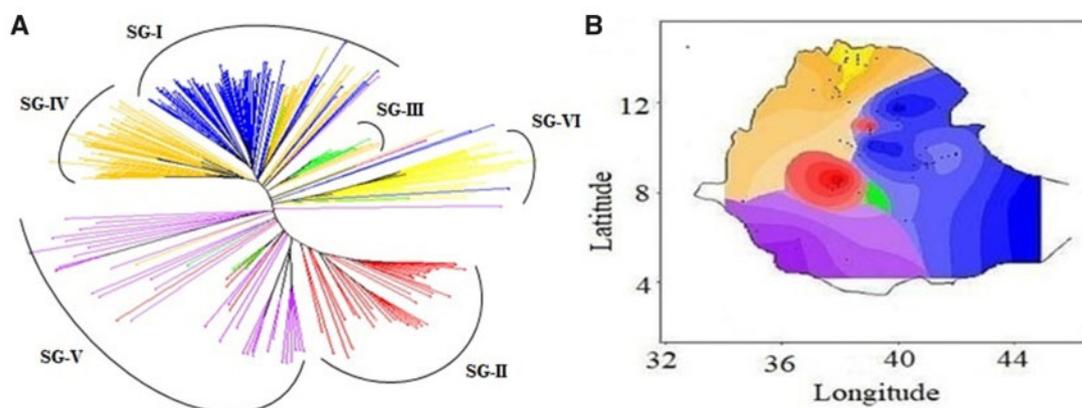
Figure 3 Genetic clustering of Ethiopia sorghum collection based on 108,107 SNPs. (A) NJ tree of 304 sorghum accessions from DARwin 6.0.14. Tree edges were color-coded based on predominant ancestry groups determined in ADMIXTURE. (B) Spatial population structure across the geographic distribution of the accessions, colored by ancestry coefficients for six ancestral populations ($K = 6$) determined in ADMIXTURE.
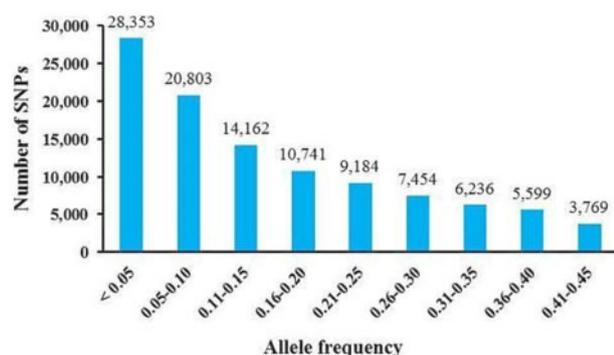
Figure 4 MAF and number of SNPs based on 304 sorghum accessions from Ethiopia.

showed evidence of mixed population ancestry (see Supplementary File S3).

According to Amede *et al.* (2015) eight agro-ecological zones (cool/humid, cool/subhumid, cool/semiarid, cool/arid, warm/humid, warm/subhumid, warm/semiarid, and warm/arid) have been identified in Ethiopia based on the Global 16 Class Classification System. Given the fact that sorghum is grown in all these agroecologies except warm/humid and cool/arid (Menamo *et al.* 2021), we hypothesized that genetic groups could reflect to population structure across the agro-ecological zones of Ethiopia. Consistent with this hypothesis, we observed strong geographic clustering when the accessions were mapped by group (Figure 3B), suggesting significant contribution of agro-ecological variation to ancestry, with individuals in specific subgroup found to co-locate in geographic regions. For instance, in SG-I (blue), the majority of the individuals were from eastern parts of Ethiopia (Figure 3B). In addition, individuals with high membership coefficients in SG-III (green), SG-V (purple) and SG-VI (yellow) showed strong clustering according to their geographic origin (central, western, and northern parts of the country, respectively). In contrast, SG-II which includes 41 accessions showed modest clustering according to geography, suggesting that the population structure could also be affected by other factors such as seed exchange and food preferences (Deu *et al.* 2010).

Genetic diversity and population differentiation

Based on the allele frequency distribution of this collection, 26% of the SNPs were rare (MAF < 0.05; Figure 4). For the four regional populations, we found a similar proportion of rare alleles, with values ranging from 15% to 20% in accessions collected from the Amhara and Tigray regions, respectively. Among the six ADMIXTURE subgroups, SG-III had the highest percentage (46%) of rare alleles, while in the remaining groups (SG-I, SG-II, IV, V, and VI), an average of 24% of the detected alleles were rare. The distribution of these rare alleles among populations could represent a recent admixture as a result of interpopulation gene flow (Memon *et al.* 2016). The correlation of a recent admixture and the distribution of rare alleles among populations, may in part explained by the fact that some alleles would have the opportunity to be reintroduced to the populations through recent introgression, and that these alleles will come to be spread fairly evenly in the populations in very small quantities. In addition, pairwise population comparisons showed that 76–91% of the SNPs common in at least one population were common among all regional populations (*i.e.*, 58,673 SNPs), whereas only 18–41% of such SNPs were common among all ADMIXTURE subgroups (*i.e.*, 12,400 SNPs).

Genetic diversity parameters for the entire panel, regional populations and ADMIXTURE subgroups are summarized in Table 2. The individual SNP PIC values of the entire panel ranged between 0.09 and 0.37, with an average value of 0.24 across all polymorphic loci (Table 2). Gene diversity (H_e; *i.e.*, expected heterozygosity) ranged from 0.09 to 0.50, and its value average all loci was 0.29 (Table 2). The mean observed heterozygosity value ($H_o = 0.12$) of the entire panel was similar with that observed in previous studies of sorghum landraces (Dje *et al.* 2004; Cuevas *et al.* 2017). Among the four regions, the highest level of genetic diversity was observed in accessions collected from the Tigray region ($N_E = 1.53$, $H_e = 0.32$, PIC = 0.26), and the lowest in the Oromia region ($N_E = 1.42$, $H_e = 0.28$, PIC = 0.24). Allelic richness based on rarefaction was relatively higher in the Southern Nations ($R_S = 1.81$) and lower in the Oromia region ($R_S = 1.77$; Table 2). Considering the genetic diversity among the six ADMIXTURE subgroups, the highest level of genetic diversity ($N_E = 1.52$, $R_S = 1.77$, $H_e = 0.32$) was found in SG-V (Table 2). SG-VI had the second highest level of diversity ($R_S = 1.66$, $H_e = 0.31$) and harbored many private alleles ($N_{PA} = 1430$) compared with other subgroups (Table 2).

AMOVA showed that 98% of the total variation was found within regions, whereas 28.06% and 71.94% of the total variation was found among and within ADMIXTURE subgroups, respectively (see Supplementary Table S1). The number of migrants per

Table 2 Genetic diversity estimates of Ethiopian sorghum accessions at different population levels

Population	N^a	n	N_E	R_S	N_{PA}	H_o	H_e	PIC
Entire panel	304	—	1.46	—	—	0.12	0.29	0.24
Administrative regions								
Amhara	146	30	1.47	1.79	776	0.13	0.30	0.24
Oromia	77	30	1.42	1.77	230	0.11	0.28	0.24
Tigray	44	30	1.53	1.78	126	0.19	0.32	0.26
Southern nations	37	30	1.48	1.81	55	0.12	0.30	0.24
ADMIXTURE subgroups								
SG-I[b]	84	25	1.51	1.60	683	0.19	0.30	0.26
SG-II	41	25	1.50	1.50	500	0.18	0.29	0.25
SG-III	33	25	1.32	1.33	12	0.01	0.23	0.20
SG-IV	80	25	1.51	1.64	648	0.19	0.30	0.26
SG-V	37	25	1.52	1.77	637	0.15	0.32	0.26
SG-VI	29	25	1.50	1.66	1430	0.15	0.31	0.25

[a] N, number of individuals; n, subsample size; N_E, effective number of alleles; R_S, allelic richness; N_{PA}, number of private alleles; H_o, observed heterozygosity; H_e, expected heterozygosity; PIC, polymorphism information content.
[b] SG-I to SG-VI represents subgroups 1–6. Except for the entire panel, basic diversity statistics (N_E, H_o, H_e, and PIC) were estimated based on subsamples of equal size.

Table 3 Pair wise F_{ST} matrix, a measure of population divergence among the six ADMIXTURE subgroups

	SG-II	SG-III	SG-IV	SG-V	SG-VI
SG-I[a]	0.44	0.18	0.11	0.25	0.31
SG-II		0.48	0.41	0.21	0.23
SG-III			0.16	0.27	0.38
SG-IV				0.21	0.29
SG-V					0.23

[a] SG-I to SG-VI represents subgroups 1–6.

generation as indirect estimate of gene flow was also very high ($N_m = 12.25$) among the regions, leading to a low genetic differentiation between the regions. The pairwise fixation index (F_{ST}) among ADMIXTURE subgroups ranged from 0.11 to 0.48, indicating a relatively high level of genetic differentiation (Table 3) that resulted from a restricted gene flow ($N_m = 0.64$) among the populations.

Genomic signatures of selection

The identification of functional genomic regions that might be targets of selection provides information useful for the discovery of candidate genes of breeding importance in sorghum (Campbell et al. 2016). In this study, a total of 79,754 SNPs (MAF > 0.05) were tested for evidence of selection among the six ADMIXTURE subgroups using BayeScan v.2.1 (Foll and Gaggiotti 2008). This approach distinguishes between loci that diverged via random drift and those that diverged via selection. Among the 79,754 SNPs analyzed, 40 (FDR < 0.05) present evidence of selection among the six genetic groups according with BAYESCAN results (Figure 5; see Supplementary File S4). Among these 40 F_{ST} outlier SNPs, 38 were consistent with the evidence of diversifying selection ($\alpha > 0$) and two corresponded to balancing selection ($\alpha < 0$).

We also identified a total of 47 candidate genes in the vicinity of the genomic regions containing these F_{ST} outlier SNPs (see Supplementary File S4). These candidate genes represented different categories of biological processes, including regulation of biotic and abiotic stress tolerance (F-box proteins, MADS box transcription factor), signal transduction (similar to low temperature-responsive RNA-binding protein), plant cell wall synthesis (Glycosyl transferase 1) and ion transport.

Discussion

Ethiopia, the probable center of origin and diversity for sorghum and with unique ecogeographic features, possesses a large number of landraces in the gene bank as well as under subsistence agriculture. The germplasm from this region represents one of the most important sources of useful genes for sorghum improvement efforts around the world (Doggett 1988; Reddy et al. 2009; Adugna 2014). In addition to providing a broad sample of the diversity in sorghum, the genotypes included in this study are known to display agronomically important traits including drought tolerance (Wondimu et al. 2020). Therefore, the genomic characterization presented herein provides an advantageous starting point to make adequate use of these valuable resources, and could also be employed for the genomic dissection of important phenotypes in sorghum.

Genetic diversity and regional differentiation

In this study, a high-throughput GBS technology was used to explore genetic diversity, population structure, and selection signature in sorghum accessions collected across the center of origin and domestication, Ethiopia. Indeed, the lower frequency of rare alleles (Figure 4) observed in our study than in a previous GBS analysis of Ethiopian sorghum landraces (Girma et al. 2019), highlights the potential of this collection for subsequent allele mining studies through genome-wide association study. The level of diversity in this Ethiopian collection is higher than that observed in the global sorghum association panel (Maina et al. 2018), which confirms Doggett's long standing hypothesis that Ethiopia is not only part of the center of origin but also the center of diversity of sorghum (Doggett 1988). The diverse agro-ecological zones and farming systems where sorghum is grown in Ethiopia as well as the high level of gene flow between cultivated sorghum and its wild relatives, all seem to have contributed to the wide range of variation observed in this and previous studies (Snowden 1936; Stemler et al. 1977; Teshome et al. 1997; Ayana and Bekele 1998, 1999; Tesso et al. 2008) of Ethiopian sorghum germplasm. An intriguing hypothesis is that the richness of diversity in Ethiopia may facilitate selection for different allele combinations that result in particular suites of traits, providing rich genetic sources for sorghum improvement programs.

Our results (see Supplementary Table S1) support previous reports of low level of regional differentiation for cultivated sorghum in Ethiopia (Ayana and Bekele 1998; Ayana et al. 2000; Desmae et al. 2016a,b). The lack of regional differentiation could be attributed, at least in part, to frequent gene flow as a consequence of extensive exchange of materials between farmers from these regions, which was also confirmed by the high rate of gene flow observed among the regions. An alternative or perhaps complementary explanation for the lack of regional differentiation is that these regions do not represent different agro-environmental conditions but political regions formed based on the federal system of Ethiopia. Overall, these results suggest that a single large random collection from the whole area would be adequate to capture and preserve most of the genetic variation present in Ethiopian sorghum germplasm. However, the high level of allelic diversity in population from the Southern Nations (Table 2) could be an indicator of the conservation status of its genetic diversity, thus additional collection from this region may be needed to
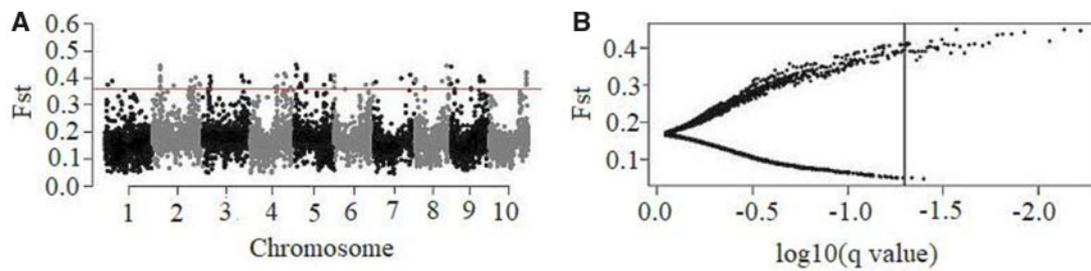
Figure 5 BAYESCAN results for the analysis of 79,754 SNPs among six Ethiopian sorghum subgroups for outlier prediction. (A) The distribution of F_{ST} values across the sorghum genome. The red horizontal line is a cutoff for top F_{ST} outlier SNPs. (B) Each F_{ST} value is plotted against the log10 of the corresponding q-value for outlier prediction. The vertical line indicates the threshold FDR (FDR = 0.05) value used to identify outlier SNPs represented on the right side of the line.

support and increase the genetic diversity of Ethiopian sorghum germplasm collection.

The structure of genomic diversity in Ethiopian sorghum

While regional differentiation is lacking, analysis of genotypic data revealed clear differentiation among six Ethiopian populations (Figures 2 and 3). Previous studies have shown that sorghum populations are structured according to botanical races and geography (Barnaud *et al.* 2007; Deu *et al.* 2010; Maina *et al.* 2018). Since the accessions used in the current study were not characterized for racial groups, it was not possible to relate the observed genetic structure with racial category. Of the different sorghum growing agro-ecological regions in Ethiopia, the wetter regions mostly represent the western parts of the country which receive high rainfall, and with rainfall rapidly decreasing to the east (Amede *et al.* 2015). Our spatial analysis (Figure 3B) separates populations in the SG-V (purple; mostly from the western parts of Ethiopia) from SG-I (blue; predominantly found in the eastern parts of the country), consistent with these two populations inhabiting contrasting environments, at least in terms of rainfall. Surprisingly, we also observed distinct geographic groupings between subgroups III (green) and IV (orange). SG-III is found in central Ethiopia, which is mostly characterized by cool/subhumid conditions (Amede *et al.* 2015), while SG-IV mainly from the northeastern parts of the country that is generally characterized by warm/semiarid conditions, providing additional insights into the patterns of ancestry resulting from adaptation to different agroecologies. In contrast, SG-II (red) is found in multiple groups along with populations from SG-I (blue), suggesting that the population structure may also be affected by other factors such as human activities including seed exchange and food preferences (Deu *et al.* 2010).

Overall, the observation that the six distinct genetic groups identified equaled the number of different agro-ecological zones (*i.e.*, cool/humid, cool/subhumid, cool/semiarid, warm/subhumid, warm/semiarid, and warm/arid) of Ethiopian sorghum may indicate a strong contribution of agro-ecological variation to genetic groupings observed in this study. Further studies that combine population analyses with environmental and phenotypic trait variables should provide a more complete understanding of sorghum genetic structure in Ethiopia and facilitate breeding of locally adapted sorghum varieties.

Signatures of selection

We expect higher genetic population differentiation for adaptive SNP than neutral SNP if adaptation to local environments is the principal source of genetic differentiation (Villemereuil and

Gaggiotti 2015). To identify genomic regions that may be under selection pressure, we used the F_{ST} outlier method (see '*Materials and Methods*' section). Of the 40 outlier SNPs identified, 31 were located at <100 kb from annotated genes (see Supplementary File S4), which provides the first support for their putative relevance.

For instance, the top F_{ST} outlier SNP (S5_3030678; F_{ST} = 0.45) with a signature of diversifying selection, was located at ~24.42 kb from *Sobic.005G033801*, a candidate gene which encodes flavonol 3-O-glucosyltransferase protein known to be involved in anthocyanin biosynthesis pathway (Holton and Cornish 1995). Anthocyanin accumulation appeared to be associated with grain pericarp color (Awika *et al.* 2004, 2005), and protection against bird predation (Xie *et al.* 2019) in sorghum. However, the presence of pericarp color is undesirable in sorghum grains for making *injera* (a traditional pancake like bread in Ethiopia) for human consumption (Ayana and Bekele 1998). The diversifying selection observed here thus could be the result of opposing selection pressures driven by humans and natural conditions. We found two additional candidate genes (*Sobic.005G041200* and *Sobic.007G135301*) located at 64.13 and 17.33 kb from S5_3724645 and S7_56063576 on chromosome 5 and 7, respectively. The *Sobic.005G041200* gene encodes F-box protein known to be associated with sorghum response to various stresses including drought (Johnson *et al.* 2014). While the *Sobic.007G135301* gene encodes MADS box protein which has been associated to smut resistance in sorghum (Girma *et al.* 2019) and maize (Wang *et al.* 2012). Another candidate gene (*Sobic.003G269600*), for which evidence of association with sorghum phenotypic diversity for plant height had been reported (Phuong *et al.* 2013), was identified near S3_60642776. In Ethiopia, sorghum grows in diverse agroecologies ranging from the hot and dry lowlands to high-altitude regions (Snowden 1936), where different environmental conditions favor different biotic and abiotic stress factors. Thus, the diversifying selections detected in this study are expected, as the type of selection acting on a gene can be different between populations depending on the environments.

Conclusion

This study reported a high level of genetic diversity and differentiation among Ethiopian sorghum accessions, which provides a great opportunity for developing new cultivars with desirable characteristics. Our results illustrate how populations adapting to different environments become structured genetically on small spatial scales. We found genomic regions of potential interest, with further large-scale phenotypic and geographic characterization can provide multiple lines of evidence for the putative importance of these particular loci in the genetic control of traits

of economic and adaptive importance in sorghum. Overall, this study contributes to the genomic resources available for sorghum improvement efforts around the world.

Literature cited

Adamack A.T. and Gruber B. 2014. PopGenReport: simplifying population genetic analyses in R. Methods in Ecology and Evolution. 5(4):384–387.

Adugna A. 2014. Analysis of in situ diversity and population structure in Ethiopian cultivated Sorghum landraces using phenotypic traits and SSR markers. SpringerPlus. 3:212–214. doi:10.1186/2193-1801-3-212.

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

Amede T, Auricht C, BoffaJ, Dixon J, Mallawaarachchi T, *et al.* 2015. The Evolving Farming and Pastoral Landscapes in Ethiopia: A Farming System Framework for Investment Planning and Priority Setting. Canberra: ACIAR.

Awika JM, Rooney LW, Waniska RD. 2004. Properties of 3-deoxyanthocyanins from sorghum. J Agric Food Chem. 52:4388–4394.

Awika JM, Rooney LW, Waniska RD. 2005. Anthocyanins from black sorghum and their antioxidant properties. Food Chem. 90: 293–301.

Ayana A, Bekele E. 1998. Geographical patterns of morphological variation in Sorghum germplasm from Ethiopia and Eritrea: qualitative characters. Hereditas. 129:195–205.

Ayana A, Bekele E. 1999. Multivariate analysis of morphological variation in sorghum germplasm from Ethiopia and Eritrea. Genet Resour Crop Evol. 46:273–284.

Ayana A, Bekele E. 2000. Geographical patterns of morphological variation in sorghum germplasm from Ethiopia and Eritrea: quantitative characters. Euphytica. 115:91–104.

Ayana A, Bryngelsson T, Bekele E. 2000. Genetic variation of Ethiopian and Eritrean sorghum germplasm assessed by random amplified polymorphic DNA (RAPD). Genet Resour Crop Evol. 47: 471–482.

Balota M, Payne WA, Rooney WL, Rosenow DT. 2008. Gas exchange and transpiration ratio in sorghum. Crop Sci. 48:2361–2371. doi: org/10.2131/cropsci.01.0051.

Barnaud A, Deu M, Garine E, McKey D, Joly HI. 2007. Local genetic diversity of sorghum in a village in northern Cameroon: structure and dynamics of landraces. Theor Appl Genet. 114:237–248. doi: 10.1007/s00122-006-0426-8. PMID:17089177.

Bouchet S, Pot D, Deu M, Rami JF, Billot C, *et al.* 2012. Genetic structure, linkage disequilibrium and signature of selection in sorghum: lessons from physically anchored DArT markers. PLoS One. 7:e33470.doi:10.1371/journal.pone.0033470.

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, *et al.* 2007. TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics. 23:2633–2635. doi: 10.1093/bioinformatics/btm308.

Browning SR, Browning BL. 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am J Hum Genet. 81:1084–1097. doi:10.1086/521987.

Campbell BC, Gilding EK, Mace ES, Tai S, Tao Y, *et al.* 2016. Domestication and the storage starch biosynthesis pathway: signatures of selection from a whole sorghum genome sequencing strategy. Plant Biotechnol J. 14:2240–2253. doi:10.1111/pbi.12578.

Casa AM, Mitchell SE, Hamblin MT, Sun H, Bowers JE, *et al.* 2005. Diversity and selection in sorghum: simultaneous analyses using simple sequence repeats. Theor Appl Genet. 111:23–30.

Caye K, Deist TM, Martins H, Michel O, François O. 2016. TESS3: fast inference of spatial population structure and genome scans for selection. Mol Ecol Resour. 16:540–548. doi:10.1111/1755-0998.12471

Central Statistical Agency. 2018. Agricultural sample survey report on area and production of major crops (private peasant holdings, Meher season). Vol I: September-December. In: Statistical Bulletin. Ethiopia: Addis Ababa.

Cuevas HE, Prom LK. 2013. Assessment of molecular diversity and population structure of the Ethiopian sorghum germplasm collection maintained by the USDA-ARS National Plant Germplasm Systems using SSR markers. Genet Resour Crop Evol. 60: 1817–1830.

Cuevas HE, Valentin GR, Hayes CM, Rooney WL, Hoffman L. 2017. Genomic characterization of a core set of the USDA-NPGS Ethiopian sorghum germplasm collection: implications for germplasm conservation, evaluation, and utilization in crop improvement. BMC Genomics. 18:108. doi:10.1186/s12864-016-3475-7.

Cuevas HE, Prom LK. 2020. Evaluation of genetic diversity, agronomic traits, and anthracnose resistance in the NPGS Sudan sorghum core collection. BMC Genomics. 21:88.doi.org/10.1186/s12864-020-6489-0.

Dahlberg JA, Wasylikowa K. 1996. Image and statistical analyses of early sorghum remains (8000 B.P.) from the Nabata Playa archaeological site in the Western Desert, southern Egypt. Veg Hist Archaeobot. 5:293–299.

Deschamps S, Llaca V, May GD. 2012. Genotyping-by-sequencing in plants. Biology (Basel). 1:460–483. doi: 10.3390/biology1030460.

Desmae H. 2007. Genetic diversity and variability in grain quality of sorghum landraces from North-Eastern Ethiopia [PhD thesis]. University of Queensland. Brisbane, Australia.

Desmae H, Jordan D, Godwin I. 2016a. DNA markers reveal genetic structure and localized diversity of Ethiopian sorghum landraces. Afr J Biotechnol. 15:2301–2311. doi:10.5897/AJB2016.1540.

Desmae H, Jordan D, Godwin I. 2016b. Geographic patterns of phenotypic diversity in sorghum landraces from North Eastern Ethiopia. Afr J Agric Res. 11:3111–3122.

Deu M, Sagnard F, Chantereau J, Calatayud C, Vigouroux Y, *et al.* 2010. Spatio-temporal dynamics of genetic diversity in Sorghum bicolor in Niger. Theor Appl Genet. 120:1301–1313. doi: 10.1007/s00122-009-1257-1.

Dje Y, Heuertz M, Ater M, Lefebvre C, Vekemans X. 2004. In situ estimation of outcrossing rate in sorghum landraces using microsatellite markers. Euphytica. 138:205–212.

Doggett H. 1988. Sorghum, 2nd ed. Longman, Burnt Mill, Harlow, Essex, England.

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, *et al.* 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One. 6:1–10. doi:10.1371/journal.pone.0019379.

Excoffier L, Smouse PE, Quattro JM. 1992. Analysis of molecular variance inferred frommetric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics. 131:479–491.

Frere CH, Prentis PJ, Gilding EK, Mudge AM, Cruickshank A, *et al.* 2011. Lack of low frequency variants masks patterns of non-neutral evolution following domestication. PLoS One. 6:e23041.

Foll M, Gaggiotti O. 2008. A genome-scan method to identify selected loci appropriate for both dominant and co-dominant markers: a Bayesian perspective. Genetics. 180:977–993.

Gaut B, Yang G, Takuno S, Luis E. 2011. The patterns and causes of variation in plant nucleotide substitution rates. Annu Rev Ecol Evol Syst. 42:245–266.

Gebeyehu G. 1993. Characterization and evaluation of sorghum germplasms collected from Gambella [MSc thesis]. Ethiopia: Alemaya University of Agriculture.

Geleta M, Labuschagne MT, Viljoen CD. 2006. Genetic diversity analysis in sorghum germplasm as estimated by AFLP, SSR and morpho-agronomical markers. Biodivers Conserv. 15:3251–3265.

Girma G, Nida H, Seyoum A, Mekonen M, Nega A, *et al.* 2019. A large-scale genome-wide association analyses of Ethiopian sorghum landrace collection reveal loci associated with important traits. Front Plant Sci. 10:691.doi:10.3389/fpls.2019.00691.

Goudet J. 2005. HIERFSTAT, a package for R to compute and test hierarchical F-statistics. Mol Ecol Notes. 5:184–186.

Hamblin MT, Mitchell SE, White GM, Gallego J, Kukatla R, *et al.* 2004. Comparative population genetics of the panicoid grasses: sequence polymorphism, linkage disequilibrium and selection in a diverse sample of Sorghum bicolor. Genetics. 167:471–483.

Harlan JR, de Wet JMJ. 1972. A simplified classification of cultivated sorghum. Crop Sci. 12:172–176.

Holton TA, Cornish EC. 1995. Genetics and biochemistry of anthocyanin biosynthesis. Plant Cell. 7:1071–1083. doi:10.1105/tpc.7.7.1071.

Hulbert SH. 1971. The non-concept of species diversity: a critique and alternative parameters. Ecology. 52:577–586.

Johnson SM, Lim F-L, Finkler A, Fromm H, Slabas AR, *et al.* 2014. Transcriptomic analysis of Sorghum bicolor responding to combined heat and drought stress. BMC Genomics. 15:456.doi:10.1186/1471-2164-15-456.

Mace ES, Buhariwalla KK, Buhariwalla HK, Crouch JH. 2003. A high-throughput DNA extraction protocol for tropical molecular breeding programs. Plant Mol Biol Rep. 21:459–460. doi:10.1007/BF02772596.

Mace ES, Tai S, Gilding EK, Li Y, Prentis PJ, *et al.* 2013. Whole-genome sequencing reveals untapped genetic potential in Africa's indigenous cereal crop sorghum. Nat Commun. 4:2320.doi:10.1038/ncomms3320.

Maina F, Bouchet S, Marla SR, Hu Z, Wang J, *et al.* 2018. Population genomics of sorghum across diverse agro-climatic zones of Niger. Genome. 61:223–232. doi:10.1139/gen-2017-0131.

Menamo T, Kassahun B, Borrell AK, Jordan DR, Tao Y, *et al.* 2021. Genetic diversity of Ethiopian sorghum reveals signatures of climatic adaptation. Theor Appl Genet. 134:731–742. doi:10.1007/s00122-020-03727-5.

Memon S, Jia X, Gu L., Zhang X. 2016. Genomic variations and distinct evolutionary rate of rare alleles in Arabidopsis thaliana. BMC Evol BiolEvolutionary Biology. 16:25.doi:10.1186/s12862-016-0590-7.

Miaoulis G, Michener RD. 1976. An Introduction to Sampling. Dubuque, Iowa: Kendall/Hunt Publishing Company.

Morris GP, Ramu P, Deshpande SP, Hash CT, Shah T, *et al.* 2013. Population genomic and genome-wide association studies of agro-climatic traits in sorghum. Proc Natl Acad Sci U S A. 110:453–458. doi:org/10.1073/pnas.1215985110.

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, *et al.* 2009. The Sorghum bicolor genome and the diversification of grasses. Nature. 457:551–556.

Perrier X, Jacquemoud-Collet JP. 2006. DARwin Software. http://darwin.cirad.fr/darwin.

Phuong N, Stützel H, Uptmoor R. 2013. Quantitative trait loci associated to agronomic traits and yield components in a *Sorghum bicolor* L. Moench RIL population cultivated under pre-flowering drought and well-watered conditions. Agric Sci. 04:781–791. doi:10.4236/as.2013.412107.

Poland JA, Rife TW. 2012. Genotyping-by-sequencing for plant breeding and genetics. Plant Genome. 5:92–102. doi:10.3835/plantgenome2012.05.0005.

Reddy BVS, Ramesh S, Reddy PS, Kumar AA. 2009. Genetic enhancement for drought tolerance in sorghum. Plant Breed Rev. 31:189–222.

Schertz KF. 1977. Registration of A2Tx2753 and BTx2753 sorghum germplasm. Crop Sci. 17:983–983.

Singh R, Axtell JD. 1973. High lysine mutant gene (hl) that improves protein quality and biological value of grain sorghum. Crop Sci. 13:535–539.

Siol M, Wright SI, Barrett SC. 2010. The population genomics of plant adaptation. New Phytol. 188:313–332. doi:10.1111/j.1469-8137.2010.03401.x.

Snowden JD. 1936. The Cultivated Races of Sorghum. London: Adlard and Son.

Sokal RR, Michener CD. 1958. A statistical method for evaluating system relationship. Univ Kans Sci Bull. 38:1409–1430.

Stemler ABL, Harlan JR, de Wet JMJ. 1977. The sorghums of Ethiopia. Econ Bot. 31:446–460.

Tao Y, Mace ES, Tai S, Cruickshank A, Campbell BC, *et al.* 2017. Whole-genome analysis of candidate genes associated with seed size and weight in sorghum bicolor reveals signatures of artificial selection and insights into parallel domestication in cereal crops. Front Plant Sci. 8:1237. doi:10.3389/fpls.2017.01237.

Teshome A, Baum BR, Fahrig L, Torrance JK, Arnason TJ, *et al.* 1997. Sorghum landraces variation and classification in north Shewa and south Welo, Ethiopia. Euphytica. 97:255–263.

Tesso T, Kapran I, Grenier C, Snow A, Sweeney P, *et al.* 2008. The potential for crop to-wild gene flow in sorghum in Ethiopia and Niger: a geographic survey. Crop Sci. 48:1425–1431.

van Beuningen LT, Busch RH. 1997. Genetic diversity among North American spring wheat cultivars: Part III. Cluster analysis based on quantitative morphological traits. Crop Sci. 37:981–988.

Vavilov NI. 1951. The origin, variation, immunity and breeding of cultivated plants. Chronol Bot. 13:1–3.

Villemereuil P, Gaggiotti O. 2015. A new FST-based method to uncover local adaptation using environmental variables. Methods Ecol Evol. 6:1248–1258. doi:10.1111/2041-210X.12418

Wang M, Yan J, Zhao J, Song W, Zhang X, et al. 2012. Genome-wide association study (GWAS) of resistance to head smut in maize. Plant Sci. 196:125–131.

Weerasooriya DK, Maulana FR, Bandara AY, Tirfessa A, Mengistu G. 2016. Genetic diversity and population structure among sorghum germplasm collections from Western Ethiopia. Afr J Biotechnol. 15:1147–1158.

Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population-structure. Evolution. 38:1358–1370.

Wondimu Z, Bantte K, Paterson AH, Worku W. 2020. Agro-morphological diversity of Ethiopian sorghum [Sorghum bicolor (L.) Moench] landraces under water limited environments. Genet Resour Crop Evol 2149–2160. 67:doi:org/10.1007/s10722-020-00968-7.

Wright S. 1965. The interpretation of population structure by F-statistics with special regard to system of mating. Evolution. 19:395–420.

Xie P, Shi J, Tang S, Chen C, Khan A, et al. 2019. Control of bird feeding behavior by Tannin1 through modulating the biosynthesis of polyphenols and fatty acid-derived volatiles in Sorghum. Mol Plant. 12:1315–1324. doi:10.1016/j.molp.2019.08.004.

Zhang D, Kong W, Robertson J, Goff VH, Epps E, et al. 2015. Genetic analysis of inflorescence and plant height components in sorghum (Panicoidae) and comparative genetics with rice (Oryzoidae). BMC Plant Biol. 15:107.doi:10.1186/s12870-015-0477-6.

Communicating editor: J. Holland