RESEARCH ARTICLE

Exploiting genetic variation from unadapted germplasm—An example from improvement of sorghum in Ethiopia

Techale Birhan[1] | Hongxu Dong[2,3] | Nezif Abajebel[1] | Misganu Wakjira[1] | Cornelia Lemke[2] | Vincent Vadez[4] | Andrew H. Paterson[2] | Kassahun Bantte[1]

[1]Department of Horticulture and Plant Science, Jimma University, Jimma, Ethiopia

[2]Plant Genome Mapping Laboratory, University of Georgia, Athens, Georgia, USA

[3]Department of Plant and Soil Sciences, Mississippi State University, Starkville, Mississippi, USA

[4]ICRISAT, Patancheru, Andhra Pradesh, India

Correspondence
Kassahun Bantte, College of Agriculture and Veterinary Medicine, Department of Horticulture and Plant Science, Jimma University, Jimma, Ethiopia.
Email: kassahunb@gmail.com

Hongxu Dong, Department of Plant and Soil Sciences, Mississippi State University, Starkville, MS, USA.
Email: hd552@msstate.edu

Funding information
U.S. Agency for International Development (USAID), Grant/Award Number: AID-OAA-A-13-00044

Societal Impact Statement

The productivity of sorghum in Ethiopia has been largely limited by rain-fed conditions because farmers tend to use local drought-tolerant but low-yielding landraces, as high-yielding and late-maturing landrace cultivars risk failure due to drought. Addressing such issues often requires a far-reaching approach to identify and incorporate new traits into a gene pool, followed by a period of selection to re-establish an overall adaptive phenotype. The sorghum backcross nested association mapping (BC-NAM) population developed in this study increases the genetic diversity available in Ethiopian elite adapted sorghum germplasm, providing new scope to improve food security in a region known for periodic devastating droughts.

Summary

- As the center of diversity for sorghum, *Sorghum bicolor* (L.) Moench, elite cultivars selected in Ethiopia are of central importance to sub-Saharan food security. Despite being presumably well adapted to their center of diversity, elite Ethiopian sorghums nonetheless experience constraints to productivity, for example, associated with shifting rainfall patterns associated with climate change.

- A sorghum backcross nested association mapping (BC-NAM) population developed by crossing 13 diverse lines pre-identified to have various drought resilience mechanisms with an Ethiopian elite cultivar, Teshale, was tested under three rain-fed environments in Ethiopia.

- Twenty-seven, 15, and 15 quantitative trait loci (QTLs) with predominantly small additive effects were identified for days to flowering, days to maturity, and plant height, respectively. Many associations detected in this study corresponded closely to known or candidate genes or previously mapped QTLs, supporting their validity.

- The expectation that genotypes such as Teshale from the center of diversity tend to have a history of strong balancing selection, with novel variations more likely to persist in small marginal populations, was strongly supported in that for these three traits, nearly equal numbers of alleles from the donor lines conferred increases and decreases in phenotype relative to the Teshale allele. Such rich

variation provides a foundation for selection to arrive at a new "adaptive peak," exemplifying the nature of efforts that may be necessary to adapt many crops to new climate extremes.

1 | INTRODUCTION

Rain-fed agriculture plays a dominant role in the world's semiarid and dry subhumid regions (Dunkelman et al., 2018; Rockström et al., 2010). In Ethiopia, rain-fed agriculture is the main source of livelihood for 89% of the population, while contributing 39% of the country's gross domestic product (CSA, 2014). Sorghum, *Sorghum bicolor* (L.) Moench, a short-day C_4 tropical grass native to Africa, is exceptional in its wide range of adaptation. It is the fifth most important cereal crop in the world with a total production of 58.7 million tons in 2020, with Ethiopia producing approximately 5 million tons, ranking third after the United States and Nigeria (FAOSTAT, 2022).

Ethiopia is the center of origin and diversity for sorghum and harbors rich genetic diversity. However, productivity of sorghum in Ethiopia has been largely limited by rain-fed conditions because farmers tend to use local drought-tolerant but low-yielding landraces, as high-yielding and late-maturing landrace cultivars risk failure due to the frequent occurrence of drought (Amelework et al., 2016; Derese et al., 2018). Other constraints including insect pest, farmland shortage, and poor soil fertility also hinder the productivity of sorghum. Moreover, genetic diversity within breeding programs decreases due to selection, small population size, genetic drift and other factors (Fu, 2015; Reif et al., 2005) and reaching outside of local breeding programs may be necessary to adapt many crops to new climate extremes. Therefore, improvement of locally adapted cultivars is crucial to food security in Ethiopia and other sub-Saharan African regions.

The introduction of new alleles from exotic germplasm has been exploited as a solution to improving elite cultivars (Holland, 2004). Numerous traits have been characterized in a wide range of sorghum germplasm (Cuevas & Prom, 2020; Upadhyaya et al., 2009; Vadez et al., 2011). Genomic diversity of sorghum has also been well characterized on a global scale (Hu et al., 2019; Lasky et al., 2015; Morris et al., 2013) and regional scale such as Ethiopian sorghum landraces (Girma et al., 2019; Menamo et al., 2021; Wondimu et al., 2021) and West African sorghum panel (Faye et al., 2021). Interspecific cross and backcross are common practices in introgressing desirable traits from exotic germplasm into adapted cultivars (Cox & Frey, 1984; Kong et al., 2020; Menkir et al., 1994; Piper & Kulakow, 1994). However, introducing substantial amounts of genetic material from exotic sources into locally adapted cultivars while maintaining their productivity is difficult (Holland, 2004; Jordan et al., 2011). For rain-fed agriculture in Ethiopia, adaptive traits including flowering time, maturity, and plant height are vital factors in maintaining sorghum production.

Flowering time plays a central role in plant adaptation to local environmental conditions, with local ideotypes ranging from short-day forms for the tropics to day-neutral rapidly flowering forms for high temperate latitudes with short growing seasons. Genetic improvement of sorghum and other cereal crops has also adjusted plant stature to meet needs ranging from provision of extensive biomass forage and thatch for building (Blümmel & Rao, 2006; Mathur et al., 2017; Murray et al., 2008; Tesso et al., 2008), to dwarf forms ideal for mechanical harvesting and to avoid lodging and other natural hazards. Indeed, the latter is exemplified by the success of the Green Revolution, in which grain yield increased substantially in rice (*Oryza sativa* L.) and wheat (*Triticum aestivum* L.) by the introduction of semidwarf traits (Hedden, 2003; Peng et al., 1999). Flowering time and plant height are quantitative in nature. Cultivated grain sorghum varieties typically flower between 45 and 120 days after planting under various day lengths and could range from 2 to 18 ft in height depending on the dwarfing genes they contain. Classical studies suggested that sorghum flowering and plant height are each controlled by at least four major loci (*Ma1-4* and *Dw1-4*, respectively) (Quinby, 1974). Additional maturity loci (*Ma5-6*) with large effects were reported by Rooney and Aydin (1999). The control of flowering and height are much more complex than suggested by classical genetics, as genetic linkage analyses have revealed numerous additional loci (Zhang et al., 2015), among which some show major effects under various genetic backgrounds. A new recessive dwarf mutation, *dw5*, was recently isolated from a mutagenized BTx623 mutant library, but its molecular function is yet to be studied (Chen et al., 2019).

In natural populations, genotypes at the center of diversity tend to be under strong balancing selection, with novel variations more likely to persist in small marginal populations conducive to intense selection and/or fixation by drift. Here, a genotype selected in and presumably well adapted to its center of diversity is crossed to each of 13 diverse lines from across the natural and introduced range of sorghum, chosen for their increased capacity to extract water from the soil or for their exceptional transpiration efficiency (Vadez et al., 2011), but also with diverse morphology and growth habit. We hypothesize that the introduction of new alleles from exotic germplasm into locally adapted cultivars will create substantial variations in adaptive traits and necessitate a new epoch of selections to reestablish the adaptive peak or reach a new one. Therefore, a backcross nested association mapping (BC-NAM) population such as was produced here, consisting of multiple families crossed to a common tester, allows one to catalogue allelic variants at numerous QTLs and

|

determine their contribution to phenotype and distribution across diverse germplasm (Yu et al., 2008). We evaluated this BC-NAM population under three natural environments that are prone to drought in Ethiopia, which afforded the opportunity to study adaptive traits under rain-fed environments. Joint-linkage and GWAS approaches were applied to map the genetic basis of flowering time, maturity, and plant height.

2 | MATERIALS AND METHODS

2.1 | Plant materials and population development

The 13 founder lines used for the population development were obtained from ICRISAT (Table 1). These 13 diverse founder lines were selected based on their diverse spectrum of drought responsiveness traits, especially in their water extraction ability and transpiration efficiency (Vadez et al., 2011). IS2205, IS14446, and IS23988 were characterized by excellent water extraction ability; IS3583, IS14556, IS16044,

IS16173, IS22325, IS10876, and IS15428 showed good transpiration efficiency; IS9911, IS14298, and IS32234 showed good harvest index (Vadez et al., 2011). The recurrent common parent, Teshale, is an Ethiopian variety of caudatum origin, which is highly preferred by the Ethiopian farmer for its grain quality and high yield. Seeds of Teshale were sourced from Melkassa Agricultural Research Center in Ethiopia. The sorghum BC-NAM population was developed at Jimma University using a nested design, crossing the common parent, Teshale, with the selected founder line and backcrossing the resulting F_1 to Teshale to produce BC_1F_1 families. Crossing was done by hand emasculation of normal bisexual florets (approximately 50 per panicle) of Teshale, transferring pollen from the founder lines to the stigma of the emasculated florets 3–5 days later. Following the BC_1F_1 generation, genotypes were continuously selfed to BC_1F_4 via single seed descent. No artificial selections were imposed during population development. Finally, the BC_1F_4 generation was used for genotyping and phenotyping. The populations were developed between 2013/14 to 2016/17. Below, when referring to an individual BC_1F_4 population, we will use the name of the alternate parent (e.g., the IS9911 population; Table 1).

TABLE 1 Description of the sorghum backcross nested association mapping (BC-NAM) populations

Donor Name[a]	Country of origin[b]	Properties of donor parents[c]	Genetic similarity with Teshale[d]	No. of SNPs	Average percentage of Teshale genome	Pop. size in Kobo	Pop. size in Meiso	Pop. size in Sheraro[e]
IS10876	Nigeria	Transpiration efficiency	0.610	2592	75.0%	135	151	153
IS14298	South Africa	Harvest index	0.752	2809	70.7%	121	132	131
IS14446	Sudan	Water extraction ability	0.776	2425	72.5%	134	145	149
IS14556	Ethiopia	Transpiration efficiency	0.896	1888	88.5%	32	34	
IS15428	Cameroon	Transpiration efficiency	0.811	2963	79.9%	124	142	144
IS16044	Cameroon	Transpiration efficiency	0.723	1597	78.5%	36	37	
IS16173	Cameroon	Transpiration efficiency	0.706	1541	76.5%	98	111	112
IS2205	India	Water extraction ability	0.884	2283	75.5%	36	40	
IS22325	Botswana	Transpiration efficiency	0.753	2494	68.9%	101	120	120
IS23988	Yemen	Water extraction ability	0.770	2019	83.6%	34	42	36
IS32234	Yemen	Harvest index	0.854					
IS3583	Sudan	Transpiration efficiency	0.796	2285	83.8%	104	119	119
IS9911	Sudan	Harvest index	0.645	1932	78.3%	76	82	82

Abbreviation: SNP, single nucleotide polymorphism.

[a]Founder lines used in the BC-NAM population. Note that population IS32234 was discarded due to severe contamination.

[b]Source information is obtained from http://genebank.icrisat.org/IND/Passport?Crop=Sorghum.

[c]Properties of donor parents were characterized in (Vadez et al., 2011).

[d]Genetic similarity between donor lines and recurrent parent Teshale was calculated as the percentage of identical genotypes across raw single nucleotide polymorphism markers in respective population.

[e]Populations IS2205, IS14556, IS16044, and IS32234 were not included in Sheraro due to space constraint.

2.2 | Experimental design and trial management

Parental lines and BC_1F_4 lines were initially evaluated in 2017 at two drought-prone environments in Ethiopia: Kobo (12°09′N, 39°38′E), Amhrara, northern Ethiopia, and Meiso (09°14′N, 40°45′E), Oromia, eastern Ethiopia. Due to strong moisture stress during the sowing season (July 2017), large numbers of seeds failed to germinate at Kobo and Meiso fields, resulting in uneven planting density at these two locations. Therefore, a third field trial was arranged at Sheraro (14°23′N, 37°46′E), Tigray, northern Ethiopia, in 2018. These three locations represent major sorghum production regions in Ethiopia and were selected for their natural drought conditions (Figure 1). Irrigation was not available at these three locations and thus this BC-NAM population was challenged with rain-fed condition, with no well-watered controls. An alpha lattice design with two replications was used at each location. Seeds of each line were sown into one-row plots, with 75 cm between rows for a net plot size of 0.75 m × 4 m. Inorganic fertilizers DAP and Urea were added at the rates of 100 and 50 kg ha^{-1} as side dressing during sowing and 3 weeks after sowing, respectively. Thinning of seedlings was done 3 weeks after sowing, to 20 cm spacing between individual plants. Therefore, individual plots without plant loss would consist of 20 plants. However, 362 lines, 195 lines, and 2 lines lost one of two replicates at Kobo, Meiso, and Sheraro, respectively (Table S1). Natural drought conditions in Ethiopia (Figure 1) could have selected plants with local adaptation because some seedlings failed to germinate (K. Bantte, personal communications). At the field sites, weeding and pest control were carried out as needed.

Flowering and plant height traits were evaluated in this BC-NAM population. Days to flowering (DF) was defined as the number of days until 50% of plants per plot were in anthesis. Days to maturity (DM) was the number of days until 50% of plants per plot reached physiological maturity. Plant height (PH) was the mean of five representative plants per row, measured from the base to the tip of panicle after physiological maturity in centimeters.

2.3 | Phenotypic data analysis

Analysis of variance (ANOVA) was first conducted across all three environments to test significance of environment, family, genotype nested within family, family by environment interaction, and genotype nested within family by environment interaction. We compiled weather data including daily precipitation and minimum and maximum temperature from nearby weather stations during the calendar years of field trials (Figure 1). The cumulative precipitation before sowing (Jan–June) was 194.4 mm and 313.6 mm at Kobo and Meiso, respectively, whereas it was 401.2 mm at Sheraro. The lower soil moisture at Kobo and Meiso likely explained why many seeds failed to germinate compared to Sheraro. Given the distinct conditions across these three environments (Figure 1), best linear unbiased predictions (BLUPs) were estimated for each line within each environment using a mixed linear model implemented in the lme4 package (Bates et al., 2015). All model terms were treated as random effects except for grand mean in the following model:

$$Y_{ijk} = \mu + F_i + G(F)_{ij} + R_k + \varepsilon_{ijk}$$

where Y represents raw phenotypic data, μ is grand mean, F is the individual BC-NAM family, $G(F)$ is genotype nested within family, R is replication, and ε is random error. Pearson correlation coefficients between traits were calculated using line BLUPs. Broad-sense heritability was determined as the proportion of total phenotypic variance explained by the combined family and line terms using the equation:

$$H^2 = \frac{\sigma_F^2 + \sigma_{G(F)}^2}{\sigma_F^2 + \sigma_{G(F)}^2 + \sigma_\varepsilon^2/2}$$

where σ_F^2 is the variance explained by family term, $\sigma_{G(F)}^2$ is the variance explained by individual lines, and σ_ε^2 is the random error variance.

2.4 | Marker development and genomic analyses

Genomic DNA was extracted from freeze-dried leaves using a CTAB (cetyltrimethylammonium bromide) protocol (Doyle & Doyle, 1987). Twelve randomly selected DNA samples from each population were checked for genomic integrity on 2% agarose gels before library construction. DNA concentration of each sample was quantified using a Qubit Fluorometer dsDNA system (Invitrogen, Carlsbad, CA) and diluted to 20 ng/μl. Libraries were constructed using a PstI-MspI
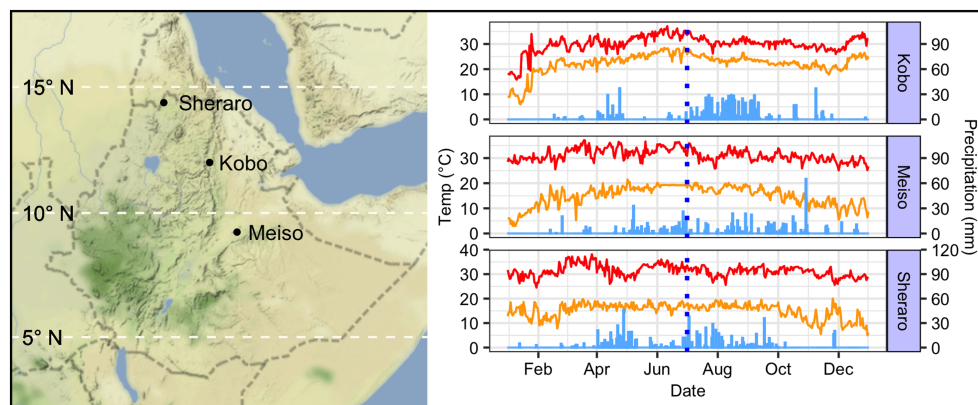

FIGURE 1 Geographic location and environmental condition of the three field trials in Ethiopia for the evaluation of the sorghum backcross nested association (BC-NAM) population. Red lines and orange lines represent daily maximum and minimum temperatures, respectively. Vertical solid blue bars represent daily precipitation, and vertical dashed lines represent the sowing time in July.

enzyme system (Poland et al., 2012) with modifications based on Clark et al. (2014). DNA samples were digested with the rare-cutting enzyme *Pst*I-HF (High-Fidelity; New England Biolabs Inc., Ipswich, MA, USA) and the common-cutting enzyme *Msp*I (New England Biolabs Inc., Ipswich, MA, USA), then ligated to a unique barcode adapter and a common adapter. A total of 192 samples (i.e., corresponding to 192 unique barcodes) was pooled in one library, and 200–500 bp fragments 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 $2 \times$ GoTaq Colorless Master Mix (Promega, Madison, WI, USA), and PCR product was extracted as above to eliminate primer-dimers. All GBS libraries were sequenced on a NextSeq500 (Illumina, San Diego, CA, USA) with 150 bp single-end reads at the University of Georgia Genomics and Bioinformatics Core. The 14 parents were replicated at least twice in order to improve coverage and to correctly call SNPs in progeny. SNP calling was performed with GBS v2 pipeline in TASSEL (Bradbury et al., 2007) using version 2.1 of the *Sorghum bicolor* genome (Goodstein et al., 2012; Paterson et al., 2009). To remove low-quality genotypic data, raw genotypes were filtered for tag coverage (tag found in >5% of taxa), minor allele frequency (MAF > 0.03), and single marker missing data (<0.8). Trimming SNPs with 5% missing data and trimming nearby (<100 bp) SNPs with identical genotypes yielded 4395 SNPs for further analysis.

2.5 | Genomic analyses

Principal component analysis (PCA) was first performed within each of these 13 BC-NAM populations to identify putative contaminants. One to seven individuals within each population exhibited high levels of errant genotypes and could not cluster with their respective population; IS32234 was of very small size ($N = 25$) and grouped into two clusters (Figure S1), probably because of contamination, mistaken parental identity, or incorrect pollination. Therefore, 54 individuals (25 individuals of IS32234 and 29 putative contaminants from the other 12 populations) were excluded and 1171 BC_1F_4 lines were retained for all analyses. A composite PCA of the retained 1171 BC_1F_4 lines was then conducted. Recurrent parent allele frequencies, genome-wide heterozygosity, and SNP monomorphism rates were calculated in R with customized scripts. Linkage disequilibrium (LD) was calculated as squared allele frequency correlations (r^2) in VCFtools (Danecek et al., 2011). Decay of LD with distance in base pairs between sites was modeled using the nonlinear regression model of Hill and Weir (1988). Polymorphic markers within each population were used to estimate the percentage of common parent genome present in each of the derived BC_1F_4 lines. Whole genome mean, maximum, and minimum percentages of the common parent genotype were estimated for each population.

2.6 | Marker-trait association

To map QTL in the BC-NAM population, we used a joint-linkage (JL) model (Buckler et al., 2009; Tian et al., 2011) and GWAS

approach. In JL analysis, we only included eight large populations ($N > 80$, Table 1), removing IS14556, IS16044, IS2205, and IS23988 due to their small size. This decision was supported by the consideration that small population size would lead to reduced power in QTL detection, underestimation of QTL number, and overestimation of QTL effect (Vales et al., 2005; Yu et al., 2008). For JL mapping across the eight populations, a new SNP dataset was created to track parent-of-origin within each population. The common parent Teshale genotypes were set to 0, alternative parent genotypes were set to 1, and heterozygous genotypes were set to 0.5. Monomorphic SNPs within each family were set to missing, and missing data were imputed as the mean of the nearest flanking markers weighted by physical distance (Tian et al., 2011). Therefore, the result can be interpreted as the probability that a SNP comes from the donor parent, and adjacent SNPs are always in high linkage disequilibrium with each other in this dataset because it reflects only the meiosis that occurred during the creation of each BC_1F_4 population. Joint-linkage analyses were performed using the Stepwise Plugin of TASSEL 5 (Bradbury et al., 2007). SNP effects were nested within populations to reflect the potential for unique QTL allele effects within each population. Although it is unlikely that there is a unique allele for each population at every QTL, this nested model provides a statistical framework for modeling multiple alleles at any given QTL. Therefore, based on this model, multiple allelic effects, as opposed to only two, are reported for each QTL. Multi-parental mapping using the GWAS approach used the unmodified genotypic dataset of all 12 populations. In each approach, the population term was included as a fixed effect to account for the inherent structure in the BC-NAM lines. GWAS approach was conducted in R using the lm function. The critical difference between joint-linkage mapping and GWAS is that joint-linkage mapping relies on parent-of-origin information while GWAS only uses allele state information of markers.

All JL QTLs identified in this study were compared to the Sorghum QTL Atlas database described in (Mace et al., 2019), which collated the projected locations of ∼6000 QTL or GWAS loci from 150 publications in sorghum from 1995 to present. QTL comparison was conducted based on their projected physical locations on version 2.1 of the *S. bicolor* genome. QTLs for the same trait were declared as common QTLs if they showed overlapping confidence intervals. Some QTLs from maps of low resolution occasionally span whole chromosomes and were not considered for comparison. In addition, sorghum genes containing or directly adjacent to SNP associations were searched using BEDOPS (Neph et al., 2012) and standard UNIX scripts.

3 | RESULTS

3.1 | Genetic diversity and structure of the BC-NAM population

To evaluate the genetic diversity and structure of the BC-NAM population, we characterized the 1171 BC_1F_4 lines at 4395 high-quality

GBS SNPs, which corresponds to an average density of one SNP per 1.5 Mb. Among these 4395 SNPs, 3029 (68.9%) were located within genic regions (Data S1). Principal component analysis showed individuals of each population to be clearly clustered (Figure 2a). The first three principal components explained 7.7%, 5.3%, and 3.9% of the variance, respectively. The IS22325-derived population exhibited the most genetic difference with the other populations based on PC2, followed by IS14298 based on PC3 (Figure 2a). Genetic similarity between the common parent and each alternate parent was lowest with IS10876 (0.610) and highest with IS14556 (0.896), which led to variation in monomorphism across the genome within each population (Table 1).

The overall mean percentage of recurrent parent genome (PRPG) was about 76.3% for all the populations but varied considerably between populations, from 68.9% in IS22325 to 88.5% in IS14556 (Figure 2b, Table 1). Three populations including IS22325 (68.9%), IS14298 (70.7%), and IS14446 (72.5%) showed lower mean PRPG than the theoretical 75% (Figure 2b, Table 1). The highest PRPG in IS14556 echoes its highest genetic similarity with the common parent (0.896, Table 1). Although we did not impose artificial selections during population development, the natural drought conditions in Ethiopia could have selected plants with local adaptation because some seedlings failed to germinate (K. Bantte, personal communications), and thereby explained higher than expected PRPG in the other 10 populations (Figure 2b). Indeed, a common set of 58 BC-NAM lines lost one of two replicates at both Kobo and Sheraro (Table S1; i.e., presumably due to poor germination associated with moisture stress). The average PRPG in these 58 lines was 74%, compared to 76% in the remaining lines without severe plant loss (Data S2). This indicated that BC-NAM lines with poor germination enriched for exotic parent alleles. The minimum percentage of recurrent parent genome also varied between the populations from as low as 38.44% for one line from the IS22325 population to 70.33% for a line from the IS14556 population (Data S2). The theoretical range of PRPG after one generation of backcross without selection is 50–100%. Few individuals with PRPG lower than 50% were likely derived from F_1 seeds rather than BC_1s. These few individuals were expected to have minimal impact on association analyses given their overall consistent clustering within respective population (Figure 2a). The maximum percentages of recurrent parent genome varied from a high of 98.71% for a line from the IS22325 population to as low as 83.21% for a line from the IS22325 population (Figure 2b, Data S2).

Linkage disequilibrium (LD) decayed to 0.2 at ~260 kb in this BC-NAM population (Figure S2a), larger than the 100–150 kb in sorghum diversity panels (Hamblin et al., 2005; Morris et al., 2013) due to the backcross breeding scheme. One generation of backcross recovers 75% of the recurrent parent genome, resulting in long haplotypes of the recurrent parent being maintained across the genome in these BC_1F_4 lines. LD is of great importance for the design of association studies to identify the genetic basis of complex traits. Given that the genome length of sorghum is ~730 Mb (Paterson et al., 2009), a minimum of ~2800 markers (730/0.26) would provide an average of one marker per LD block in the present study. Therefore, the 4395 SNP markers here are expected to sample most genetic variation in this
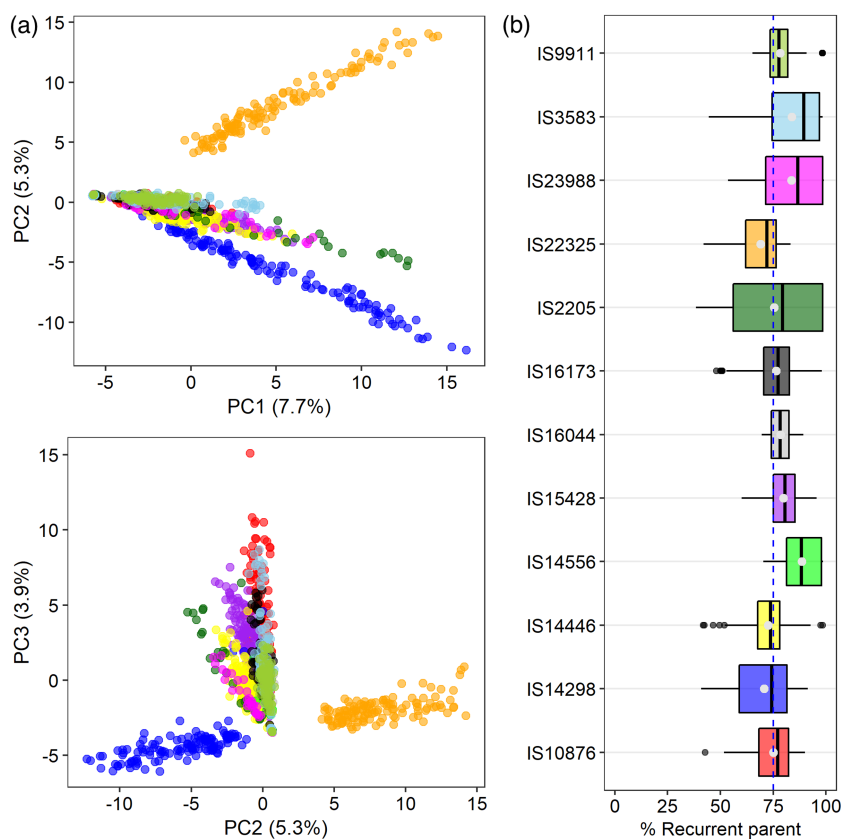

FIGURE 2 Genomic diversity of the sorghum backcross nested association (BC-NAM) population. (a) Principal component analysis across 1171 BC_1F_4 lines at 4395 SNPs. Variance explained by each principal component was shown in parenthesis. Sorghum BC_1F_4 lines were shown as dots and color-coded for each population; (b) boxplot distribution of the percentage of recurrent parent genome present in each population. The theoretical 75% value was indicated with a vertical dashed line. Dot inside each boxplot is mean, and the vertical line is median. Individual populations were color-coded as in (a).

BC-NAM population, with high power to detect marker-trait associations.

The genetic structure and diversity of the BC-NAM population might have been affected by natural selection during population development, which can lead to decreased residual heterozygosity and segregation distortion. Heterozygosity rate in the BC-NAM population was 0.0606 (Figure S2b), slightly lower than the expected value for the BC_1F_4 generation (0.0625). The decreased heterozygosity also echoes the slightly higher percentage of recurrent parent genome (76% vs. 75%, Figure 2b). Across the genome 85.67% of markers exhibited heterozygosity ≤ 0.1 (Data S1), suggesting that natural selection had little effect overall. The frequency of alleles from the common parent, Teshale, was close to the neutral expectation (75%; Figure S2c), suggesting no overall selection for or against common parent alleles. Still, a small proportion of markers showed skewed segregation, for either the common parent (e.g., IS14556), or alternate parent (IS22325) allele (Figure S2c), suggesting selection at some loci. No clear difference was observed among families in terms of proportion of distorted markers, and skewed chromosome regions were generally specific to one or a few families.

3.2 | Phenotypic variation within and between families

We evaluated the BC-NAM population over three environments (Kobo, Meiso, and Sheraro) representing major sorghum cultivation zones in Ethiopia for days to flowering, days to maturity, and plant height (Figure 3). As shown in Figure 1, these three locations had similar temperature profiles, but different daily precipitation distributions. Kobo received a cumulative precipitation of 194.4 mm before sowing (Jan–June), whereas it was 313.6 mm at Meiso and 401.2 mm at Sheraro. In particular, from May to June, Kobo only received 32.6 mm precipitation, while Meiso and Sheraro received 250.2 mm and 167.2 mm, respectively (Figure 1). As a result, severe plant losses occurred at Kobo and Meiso, with 362 and 195 individuals lost in one of the two replicates, respectively (Table S1). In Sheraro, virtually all plants survived—only two individuals lost one replicate (Table S1). Given the precipitation data, plant losses at Kobo and Meiso were presumably caused by poor germination due to moisture stress (K. Bantte, personal communications). Multi-environment ANOVA confirmed strong environmental effect and G × E interactions (P < 2.2E−16; Table S2). Between G and G × E, mean squares of G were generally twice the magnitude of G × E for the three traits (Table S2). Thus, trait BLUPs were estimated within each environment and trait-marker association analyses were performed separately for each environment.

Plants in the least drought-stressed location, Sheraro, flowered and matured earliest while also being nearly twice the height of those at the other locations. Average flowering of the BC-NAM population occurred in Sheraro at 65 days after sowing, followed by Kobo at 78 days and Meiso at 85 days (Figure 3, Table S3), reaching maturity in Sheraro at 90 days, followed by Meiso at 123 days, and Kobo at
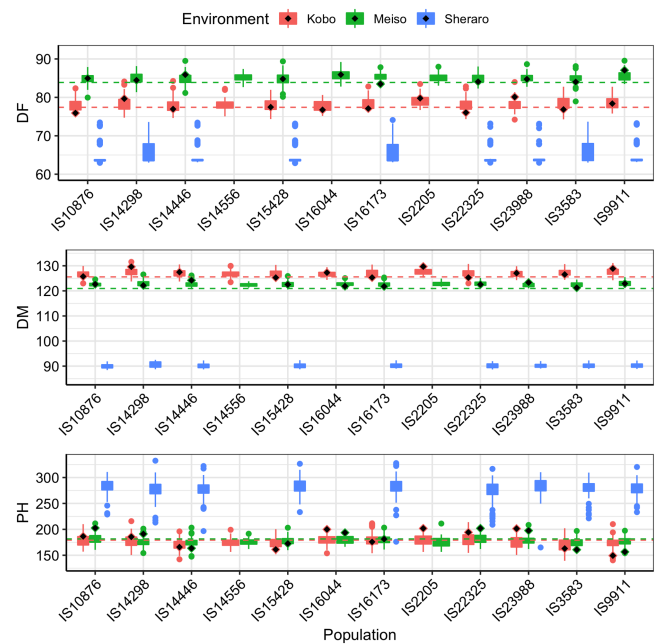

FIGURE 3 Phenotypic variation within each individual sorghum population at three environments (Kobo, Meiso, and Sheraro). Phenotypes were shown for DF (days to 50% flowering), DM (days to maturity), and PH (plant height). Inside each plot, the horizontal dashed lines represent trait means of the common parent Teshale colored based on environments, and the black diamonds represent trait means of donor parents. Note that parents were not included in Sheraro due to space constraints.

127 days. The BC-NAM populations were of similar average plant height at Kobo and Meiso, at 175.7 cm and 177.7 cm, respectively, but significantly taller at Sheraro with 279.6 cm. Trait distributions within each of the 12 populations at the three environments were similar, in terms of population mean and standard deviation (Figure 3, Data S3). Among the 12 donor parents and the recurrent parent Teshale, days to flowering ranged from 75.9 to 80.1 and 83.5 to 87.1 in Kobo and Meiso, respectively; and days to maturity ranged from 125.3 to 129.7 and 121.0 to 124.2 in Kobo and Meiso, respectively (Data S3). In contrast, plant height was more variable, parental lines ranging from 149.3 cm to 201.8 cm in Kobo and 156.5 cm to 202.6 cm in Meiso (Data S3). Due to the backcross breeding scheme, individual population means were generally close to those of the recurrent parent Teshale (Figure 3).

Broad-sense heritabilities of these three adaptive traits were generally higher in the least drought-stressed location, Sheraro (Table 2). Days to flowering heritability was high in Sheraro (0.71) but low in Kobo (0.33) and Meiso (0.30). Days to maturity heritabilities were consistently low across all three environments (0.25–0.34), which was probably due to the very limited phenotypic variance (Figure 3). Heritability of plant height followed a similar pattern as days to flowering, with the highest value observed in Sheraro (0.75), medium in Kobo (0.55), and relatively low in Meiso (0.39). Stress responses may have masked genetic potential and/or invoked different and more complex genetic controls than in favorable environments (Paterson

TABLE 2 Heritability and joint-linkage model power for each trait in the sorghum backcross nested association (BC-NAM) population

Trait	Kobo			Meiso			Sheraro		
	H^{2a}	No. QTL	Model R^b	H^2	No. QTL	Model R^2	H^2	No. QTL	Model R^2
DF	0.33	10	0.26 (79%)	0.30	10	0.25 (83%)	0.71	7	0.20 (28%)
DM	0.34	4	0.12 (35%)	0.27	5	0.12 (44%)	0.25	6	0.15 (60%)
PH	0.55	1	0.13 (24%)	0.39	8	0.20 (51%)	0.75	6	0.16 (21%)

Abbreviations: DF, days to first flowering; DM, days to maturity; PH, plant height; QTL, quantitative trait loci.
[a]Broad-sense heritability of BC-NAM population.
[b]Variance explained by joint-linkage model after fitting family term and detected QTL. Numbers in parentheses represent proportion of genetic variance explained by the joint-linkage model, which was calculated by dividing model R^2 by H^2.

et al., 2003). Moreover, medium to high positive correlations were observed between days to flowering and days to maturity (Kobo: 0.69; Meiso: 0.59; Sheraro: 0.33), whereas negligible or negative correlations were found between plant height and days to flowering (Kobo: 0.12; Meiso: −0.28; Sheraro: −0.03) and between plant height and days to maturity (Kobo: 0.06; Meiso: −0.19; Sheraro: 0.01).

3.3 | Genetic dissection of adaptive traits

Association analyses for the three measured traits revealed 27, 15, and 15 QTLs across the three environments using the joint-linkage (JL) model for days to flowering, days to maturity, and plant height, respectively (Tables 2 and S4). Among the 27 JL QTLs for days to flowering, 25 (92.6%) showed overlapping confidence interval with previous QTLs detected for the same trait from multiple studies in the Sorghum QTL Atlas database (Mace et al., 2019) (Data S4). Similarly, 12 of the 15 JL QTLs for days to maturity (80.0%) and 13 of the 15 JL QTLs (86.7%) for plant height overlapped with those found in previous studies (Data S4). Despite this validation of most QTLs, the total phenotypic variance explained by the final JL model was generally low, ranging from 0.20 for plant height to 0.26 for days to flowering in Kobo, from 0.12 for days to maturity to 0.15 for plant height in Meiso, and from 0.13 for days to flowering to 0.20 for days to maturity in Sheraro (Table 2). Heritability imposes an approximate upper limit to the R^2 of a QTL model (Yu et al., 2008). Therefore, this is not unexpected given the low to moderate heritabilities of these adaptive traits under the natural drought conditions in Ethiopia. By taking the broad-sense heritability into account, proportions of genetic variance explained by the final joint-linkage models of each trait ranged from 28% to 83% in Kobo, 35% to 60% in Meiso, and 21% to 51% in Sheraro (Table 2).

In order to leverage the investment in their analysis, we incorporated the four small populations (IS2205, IS14556, IS16044, and IS32234) into GWAS analyses. The GWAS model included the same fixed effect of family term as the JL model, but marker effects were not nested within family. A total of 43, 6, and 35 SNPs exceeded the 10E−3 threshold for days to flowering, days to maturity, and plant height across three environments, respectively, which corresponded to 19, 4, and 15 likelihood peaks (Figures 4, 5, and S3, Data S5). Despite the different statistical frameworks, there was generally high

correspondence between JL QTLs and GWAS signals (Figures 4, 5, and S3). Exact overlap of linkage mapping and GWAS is not expected, as linkage mapping tests markers within an individual population whereas GWAS tests marker effects across populations, with different strengths and weaknesses of each approach (Tian et al., 2011). Joint-linkage analysis produces many more small effects than GWAS analysis as an artifact of the model fitting process, which assigns a separate effect to all populations at each QTL. Moreover, the addition of four small populations in GWAS analyses may also confer discrepancies between these two methods.

Flowering time is one of the most important adaptive traits in grasses. The JL model detected 10, 10, and 7 JL QTLs for days to flowering at Kobo, Meiso, and Sheraro, respectively (Tables 2 and S4, Figure 4), explaining 79%, 83%, and 28% of genetic variance (Table 2). Several QTLs were consistently detected across three environments near known sorghum maturity genes (Figure 4). The most significant QTL for days to flowering was detected at the putative *Ma6* gene (*CONSTANS*-like 4; Sobic.006G004400) on chromosome 6. In the JL analysis, the QTL peaks for days to flowering were 96 kb (S06_769807), 339 kb (S06_1013548), and 1.7 Mb (S06_2410807) from a candidate *Ma6* gene (*SbGHD7*; Murphy et al., 2014) in Meiso, Kobo, and Sheraro, respectively (Table S4, Figure 4). The GWAS model also consistently detected associations adjacent to *Ma6* in Kobo (S06_1015768), Meiso (S06_769807), and Sheraro (S06_769807) at 342 kb, 96 kb, and 96 kb away from *SbGHD7*, respectively (Figure 4, Data S5). One JL QTL peak (S09_55566776) in Meiso was detected about 588 kb downstream of the *SbCN8* gene (Sobic.009G199900), which encodes phosphatidylethanolamine-binding protein (PEBP) and is an ortholog of maize *ZCN8* and rice *OsFTL10*. A marginally significant GWAS association (S09_52569648) was also detected near *SbCN8* (Figure 4b, Data S4). In Meiso, the GWAS model also detected an association peak (S03_58246694) near *SbCN12* but the JL model did not detect signals near this region (Figure 4b). The *LHY* gene (LATE ELONGATED HYPOCOTYL; Sobic.007G047400) is 97 kb from a JL QTL (S07_4611519) in Sheraro on chromosome 7 (Figure 4c, Data S4) and 1.9 Mb from a JL QTL in Kobo (S07_2790815) that was (Figure 4a, Table S4). Strong associations with flowering were also detected on sorghum chromosome 1 at Kobo and Sheraro, which were relatively distant (more than 3 Mb) from the *Ma5* gene (Sobic.001G087100). Associations on

FIGURE 4 Marker-trait associations for sorghum days to flowering in (a) Kobo, (b) Meiso, and (c) Sheraro. Each panel shows associations detected in joint-linkage (top) and genome-wide association study (bottom) models. In joint-linkage analysis, parental allelic effects were color coded, with blue represents positive effect and orange represents negative effect. Candidate genes were shown in green vertical lines and annotated with gene names.
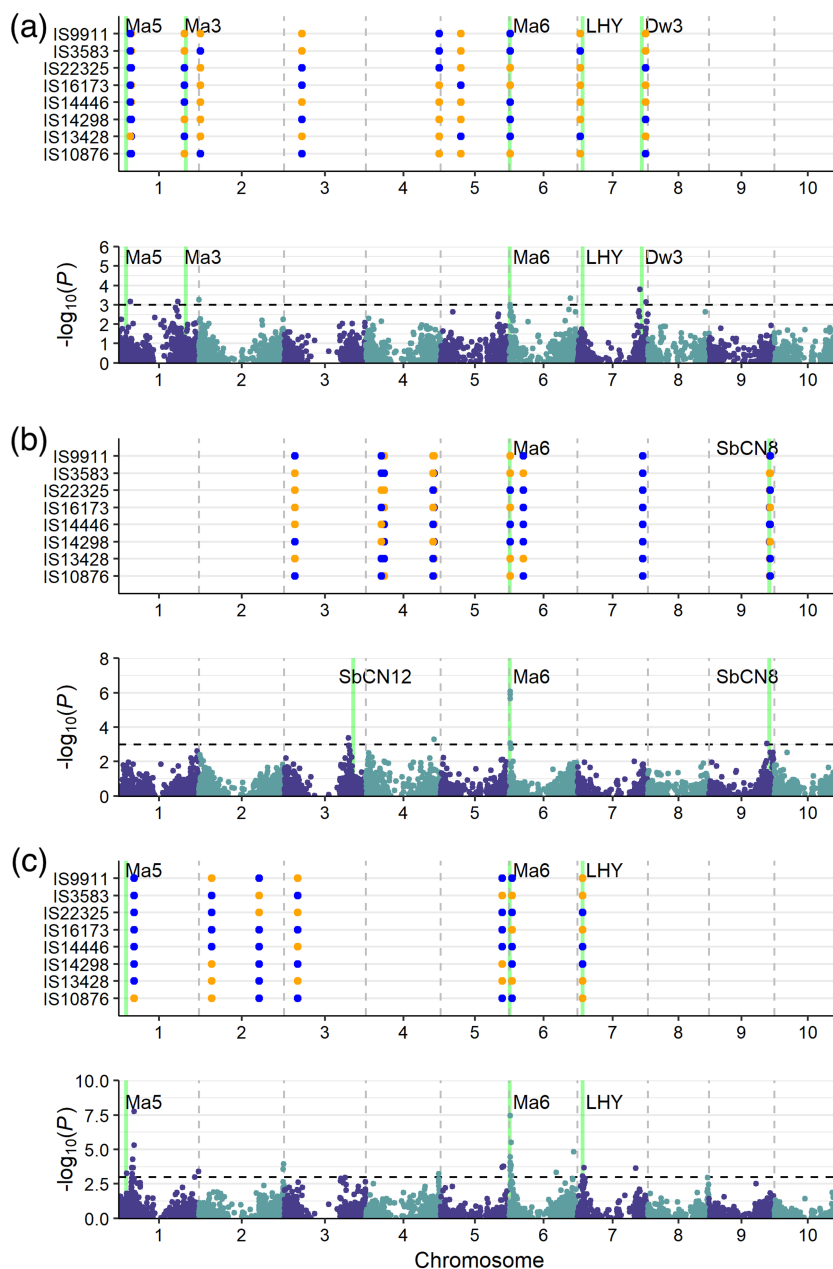
chromosomes 2, 3, and 4 (Figure 4, Table S4, Data S4 and S5) may represent novel genes.

For days to maturity, the JL model detected four, five, and six QTLs at Kobo, Meiso, and Sheraro, respectively, explaining 35%, 44%, and 60% of genetic variation (Table 2), while the GWAS model detected two, one, and one weak peaks (Figure S3). Days to maturity QTLs were mostly not near canonical maturity genes, except for one GWAS hit (S06_1015768) in Meiso near Ma6 (Figure S3, Data S5). The majority of GWAS hits for days to maturity were only marginally significant (Figure S3, Data S5), perhaps due to the very limited phenotypic variation for this trait (Figure 3).

For plant height, the JL model detected one, eight, and six QTLs at Kobo, Meiso, and Sheraro, respectively, explaining 24%, 51%, and 21% genetic variation (Figure 5, Tables 2 and S4). The GWAS model

detected two, eight, and five peaks at these three environments (Figure 5, Data S5). Several JL QTLs and GWAS hits were adjacent to known sorghum dwarfing candidate genes. One GWAS hit (S06_48457872) from Kobo was approximately 6 Mb from the *Dw2* candidate gene (Sobic.006G067700) (Data S5), which is suggested to encode a protein kinase (Hilley et al., 2017). One JL QTL (S04_63248157) and GWAS hit (S04_63653942) at Sheraro were detected in the vicinity of the *Dw4* candidate gene, thought to be near 66.7 Mb on chromosome 4 (Li et al., 2015). An additional GWAS hit (S09_49635018) was among the most significant associations for plant height in Sheraro but was far from the *Dw1* candidate gene (Sobic.009G229800).

QTL allele effects that deviate from the prediction of parental phenotypic values indicate opportunities for selecting "transgressive"
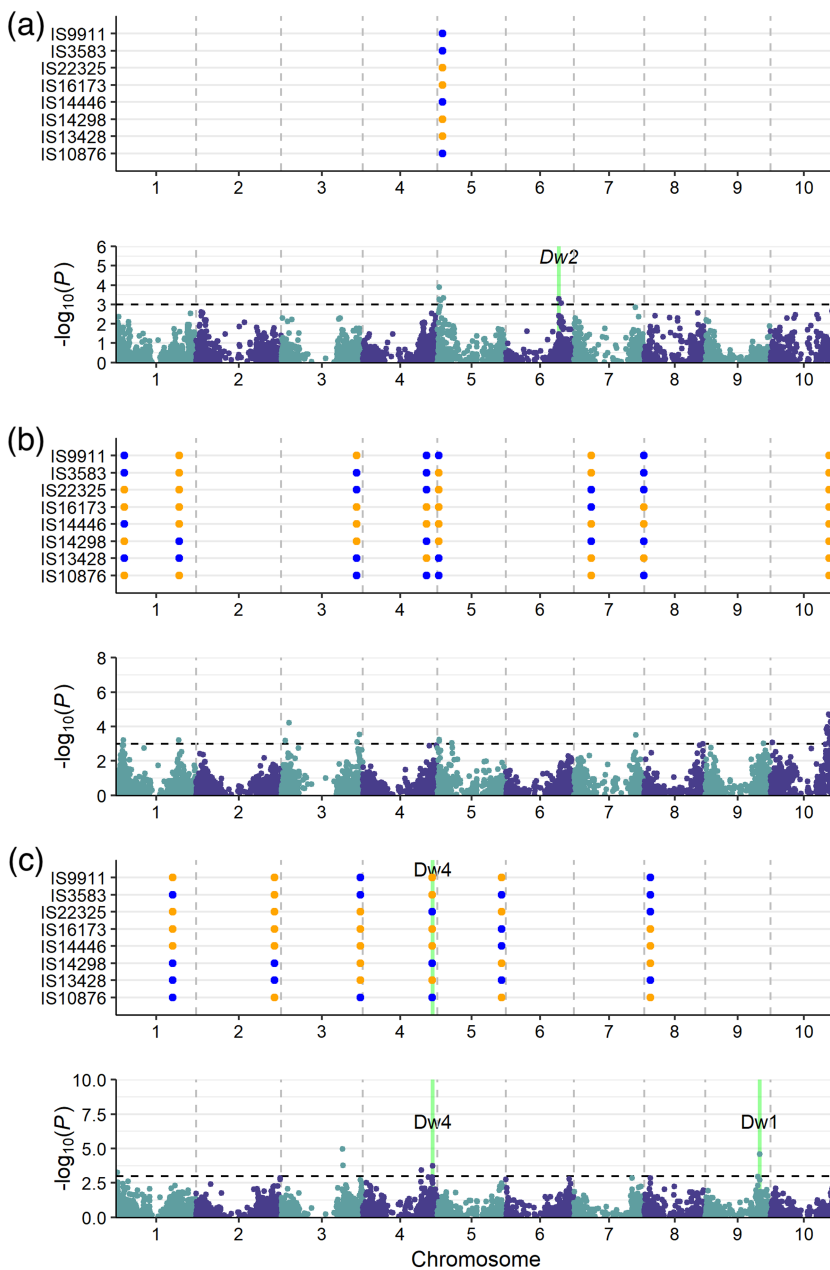
FIGURE 5 Marker-trait associations for sorghum plant height in (a) Kobo, (b) Meiso, and (c) Sheraro. Each panel shows associations detected in joint-linkage (top) and genome-wide association study (bottom) models. In joint-linkage analysis, parental allelic effects were color coded, with blue represents positive effect and orange represents negative effect. Candidate genes were shown in green vertical lines and annotated with gene names.

progeny with values that exceed those of the more extreme parent. For days to flowering, 27 JL QTLs detected in the eight large populations (Tables 2 and S4), permit estimation of a total of 216 (i.e., 27 × 8) QTL allelic effects—among these, 102 were negative (i.e., with the donor allele conferring earlier flowering than Teshale) and 114 were positive (i.e., with the donor allele conferring later flowering than Teshale) (Figure 6, Data S4). QTL allelic effects ranged from −7.4 to 5.1 days. At 26 of these 27 QTLs, both positive and negative alleles from donor parents were observed (Figure 4, Data S4), except for one QTL (S07_59174992) detected in Meiso at which all donor parents contributed positive effect alleles (Figure 4b, Data S4). On the other hand, across the 27 days to flowering QTLs, each donor parent contributed at least one positive and one negative allele(s). For days to maturity, the 120 QTL allelic effects of 15 JL QTLs (Tables 2 and S4)
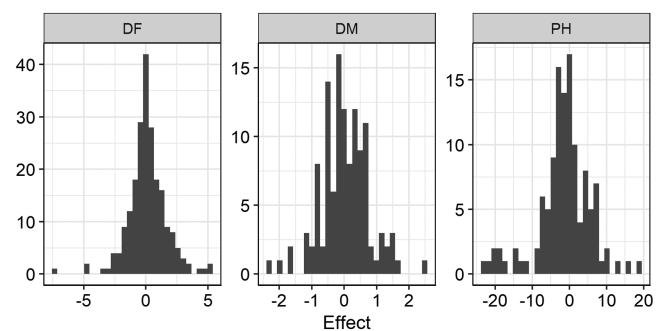

FIGURE 6 Distributions of allelic effects of joint-linkage quantitative trait loci detected in a sorghum backcross association mapping (BC-NAM) population. Inside each plot, x-axis represents allelic effects and y-axis represents frequency. DF, days to 50% flowering; DM, days to maturity; PH, plant height

included 61 negative (i.e., early maturity) and 59 positive effects (i.e., late maturity), ranging from −2.3 to 2.5 days, again with both positive and negative allelic effects at each QTL and from each donor (Figure S3, Data S4). For 120 QTL allelic effects estimated for the 15 JL plant height QTLs with a range from −23.8 cm to 18.1 cm (Figure 6, Data S4), 50 were positive (i.e., tall stature) and 70 negative (i.e., short stature). Similar to the other two traits, all plant height QTLs contained both positive and negative effects except for one QTL (S10_53361382) in Meiso, where all donor parents contributed negative effect alleles.

4 | DISCUSSION

Ethiopia is one among many sub-Saharan African countries that are extremely vulnerable to the impacts of climate change (Birara et al., 2018; Conway & Schipper, 2011). A backcross nested association mapping population made by crossing 13 diverse sorghum lines to an elite cultivar (Teshale) bred for the center of diversity (Ethiopia) provides insight into the biogeography of trait variation. The backcross nested breeding design combines practical breeding efforts for introgression of new alleles into adapted germplasm with statistical power to dissect quantitative traits (Buckler et al., 2009; Jordan et al., 2011; Yu et al., 2008). Employing donor lines chosen for divergent drought defense responses (Vadez et al., 2011), this BC-NAM population also increases the genetic diversity available in Ethiopian elite adapted sorghum germplasm, providing new scope to improve food security of societies dependent upon this crop in a region known for periodic devastating droughts.

Multiple genomic properties attest to the usefulness of this sorghum BC-NAM population (Figures 2 and S2, Table 1). Principal component analysis displayed clear population structure, indicating minimal cross contamination among families (Figure 2a). Indeed, most individual populations clustered together except for IS22325 and IS14298. This was likely due to the backcross breeding scheme used in this study. Molecular marker analysis indicated retention of an average of 76.3% of the recurrent parent genome in this BC-NAM population, close to the expected 75% (Figure 2b). In an Australian sorghum BC-NAM population, similar genome composition of the derived lines was reported, with an overall mean of 78% of the recurrent parent genome (Jordan et al., 2011). Both studies demonstrated the effectiveness of this relatively simple and practical breeding approach for introgression of new alleles into adapted crop genotypes.

Performance of the derived populations at three environments supported our hypothesis that the introduction of new alleles from exotic donors into adapted cultivars would confer substantial variations in adaptive traits (Figure 3, Data S3). Although population means were generally close to the recurrent parent Teshale, large numbers of individuals exhibited strong deviation in three adaptive traits relative to Teshale. For example, days to flowering of Teshale was 77.4 in Kobo, whereas it ranged from 76.7 to 83.5 in the population derived from IS2205 (mean DF = 79.8) (Data S3). Flowering time has major impacts on crop performance. Under rain-fed environments, relatively

small differences in flowering time potentially result in very large difference in grain yield (Hammer, 2006). Evaluation of Ethiopian sorghum landraces suggested that genotypes with reduced days to maturity were preferred to escape post-flowering moisture stress (Derese et al., 2018). Similarly for plant height, substantial phenotypic variations and extreme phenotypes were observed. Adaptive performance of these derived populations also exhibited significant differences among three environments as evidenced by ANOVA (Table S2), particularly between Sheraro and the other two sites (Figure 3), which was closely related to the precipitation profiles (Figure 1). Therefore, to develop improved sorghum cultivars for rain-fed agriculture in Ethiopia, adaptive traits need to be optimized for different locales. Indeed, realization of this goal could be challenging as evidenced in maize (Ewing et al., 2019) and bean (MacQueen et al., 2022). This underscores the selection of alleles that confer adaptation to local environments.

This BC-NAM population was effective in dissecting quantitative traits. Our study identified 27, 15, and 15 QTLs for days to flowering, days to maturity, and plant height, respectively (Tables 2 and S4). Both the present study and another of different germplasm (Mace et al., 2013) found that genetic control of flowering time in sorghum is substantially more complex than classical genetics was able to resolve (Quinby, 1974), involving a relatively large number of loci with small effects, as we have suggested (Zhang et al., 2015). Plant height QTLs, albeit detected in the vicinity of classical dwarf genes, were generally clearly separated from the candidate genes (Table S4, Figure 5). This could be due to several reasons including low marker density associated with the PstI-MspI GBS protocol, limited polymorphisms in the backcross breeding scheme, strong influence of drought stress on height, or that currently suspected candidate genes are not causal. A degree of validation is provided by the observation that many of the detected QTLs in these two BC-NAM populations overlapped with those detected by multiple bi-parental mapping studies (Data S4). Compared with gene-resolution mapping in other sorghum populations (Bouchet et al., 2017; Hu et al., 2019), identification of desirable chromosomal segments for introgression is perhaps more important than detecting individual genes in this BC-NAM population.

The ability to resolve many QTLs together with the ability to sample more allelic diversity than bi-parental populations reveals the spectrum of allele effects in the study population, in comparison to those of the common parent. In this case, the common parent, Teshale, is strategically chosen in that it was bred in and selected for a target environment near the species center of diversity—while the other parents were selected from a broad sampling of germplasm based on their diverse spectrum of drought responsiveness traits (Vadez et al., 2011). The finding that for all three measured traits, nearly equal numbers of alleles conferred increases and decreases in phenotype relative to the Teshale allele (Figure 6, Data S4), is consistent with the notion that Teshale is well adapted to the center of diversity for this particular gene pool, presumably with a history of balancing selection, while the 13 exotic sorghum lines from locales widely distributed across the natural and introduced range sample smaller marginal populations in which novel alleles may be more likely

to persist due to the effects of selection and genetic drift. Much of the variation in these three traits appears to be due to many loci of small to modest effects, with a few loci of large effects (Figure 6, Data S4). This aligns with results from maize (Buckler et al., 2009) and sorghum (Zhang et al., 2015).

This work exemplifies the nature of efforts that may be necessary to adapt many crops to new climate extremes, with the introduction of novel or extreme traits from exotic germplasm necessitating a new epoch of selection to re-establish an adaptive peak or reach a new one. Selection response for quantitative traits is determined by genetic variance, heritability, and selection intensity (Falconer & Mackay, 1996). Rich variation reflected by mixtures of "positive" and "negative" QTL alleles for all traits provides a foundation for selection, with new diversity from the diverse donor lines complementing the adaptive phenotype of Teshale. The finding in our companion paper (Dong et al., 2022) that correlations of plot-based grain yield with days to flowering (−0.20 to −0.42); and plant height (0.14–0.39) exemplify scope for the sorts of adjustments that may be needed to re-establish locally adaptive phenotypes. Indeed, with the enormous altitudinal variation of a country such as Ethiopia, somewhat different lines may be needed for different locales.

CONFLICT OF INTEREST

Authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

AHP and KB jointly developed and led this project. TB and KB developed the populations, TB, NA, MW, and KB conducted field trials and phenotypic data collection. HD and CL made GBS libraries. HD performed data analysis and wrote the draft manuscript. VV examined data analysis. All authors commented and reviewed the manuscript.

DATA AVAILABILITY STATEMENT

Sequencing data are available in the NCBI Sequence Read Archive under BioProject ID PRJNA687679. Data analysis scripts have been deposited to GitHub (https://github.com/hxdong-genetics/Ethiopian-Sorghum-BCNAM). Please contact the corresponding author for other data.

ORCID

Techale Birhan https://orcid.org/0000-0002-7045-5209
Hongxu Dong https://orcid.org/0000-0003-4028-3597
Vincent Vadez https://orcid.org/0000-0003-2014-0281
Andrew H. Paterson https://orcid.org/0000-0003-2159-0487
Kassahun Bantte https://orcid.org/0000-0003-4970-145X

REFERENCES

Amelework, B. A., Shimelis, H. A., Tongoona, P., Mengistu, F., Laing, M. D., & Ayele, D. G. (2016). Sorghum production systems and constraints, and coping strategies under drought-prone agro-ecologies of Ethiopia. *South African Journal of Plant and Soil*, *33*(3), 207–217. https://doi.org/10.1080/02571862.2016.1143043

Bates, D., Mächler, M., Bolker, B., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Birara, H., Pandey, R. P., & Mishra, S. K. (2018). Trend and variability analysis of rainfall and temperature in the tana basin region, Ethiopia. *Journal of Water and Climate Change*, *9*(3), 555–569. https://doi.org/10.2166/wcc.2018.080

Blümmel, M., & Rao, P. P. (2006). Economic value of sorghum Stover traded as fodder for urban and peri-urban dairy production in Hyderabad, India. *International Sorghum and Millets Newsletter*, *47*, 97–100.

Bouchet, S., Olatoye, M. O., Marla, S. R., Perumal, R., Tesso, T., Yu, J., Tuinstra, M., & Morris, G. P. (2017). Increased power to dissect adaptive traits in global sorghum diversity using a nested association mapping population. *Genetics*, *206*(2), 573–585. https://doi.org/10.1534/genetics.116.198499

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*, *23*(19), 2633–2635. https://doi.org/10.1093/bioinformatics/btm308

Buckler, E. S., Holland, J. B., Bradbury, P. J., Acharya, C. B., Brown, P. J., Browne, C., Ersoz, E., Flint-Garcia, S., Garcia, A., Glaubitz, J. C., Goodman, M. M., Harjes, C., Guill, K., Kroon, D. E., Larsson, S., Lepak, N. K., Li, H., Mitchell, S. E., Pressoir, G., … McMullen, M. D. (2009). The genetic architecture of maize flowering time. *Science*, *325*(5941), 714–718.

Chen, J., Xin, Z., & Laza, H. (2019). Registration of BTx623 dw5, a new sorghum dwarf mutant. *Journal of Plant Registrations*, *13*(2), 254–257. https://doi.org/10.3198/jpr2018.09.0058crgs

Clark, L. V., Brummer, J. E., Głowacka, K., Hall, M. C., Heo, K., Peng, J., Yamada, T., Yoo, J. H., & Sacks, E. J. (2014). A footprint of past climate change on the diversity and population structure of *Miscanthus sinensis*. *Annals of Botany*, *114*(1), 97–107. https://doi.org/10.1093/aob/mcu084

Conway, D., & Schipper, E. L. F. (2011). Adaptation to climate change in Africa: Challenges and opportunities identified from Ethiopia. *Global Environmental Change*, *21*(1), 227–237. https://doi.org/10.1016/j.gloenvcha.2010.07.013

Cox, T. S., & Frey, K. J. (1984). Genetic variation for grain yield and related traits in sorghum introgression populations. *Theoretical and Applied Genetics*, *68*(1–2), 145–153. https://doi.org/10.1007/BF00252331

CSA. (2014). Volume IV: Report on Land Utilization (Private Peasant Holdings, Meher Season), Addis Abeda.

Cuevas, H. E., & Prom, L. K. (2020). Evaluation of genetic diversity, agronomic traits, and anthracnose resistance in the NPGS Sudan sorghum core collection. *BMC Genomics*, *21*(1), 1–15. https://doi.org/10.1186/s12864-020-6489-0

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Derese, S. A., Shimelis, H., Mwadzingeni, L., & Laing, M. (2018). Agromorphological characterisation and selection of sorghum landraces. *Acta Agriculturae Scandinavica Section B: Soil and Plant Science*, *68*(7), 585–595. https://doi.org/10.1080/09064710.2018.1448884

Dong, H., Birhan, T., Abajebel, N., Wakjira, M., Mitiku, T., Lemke, C., Vadez, V., Paterson, A. H., & Bantte, K. (2022). Natural variation further increases resilience of sorghum bred for chronically drought-prone environments. *Journal of Experimental Botany*. https://doi.org/10.1093/jxb/erac217 PMID: In press.

Doyle, J. J., & Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, *19*(1), 11–15.

Dunkelman, A., Kerr, M., & Swatuk, L. A. (2018). The new green revolution: Enhancing rainfed agriculture for food and nutrition security in Eastern Africa. In L. A. Swatuk & C. Cash (Eds.), *Water, energy, food and people across the global south: 'The Nexus' in an era of climate* (pp. 305–324). Springer International Publishing. https://doi.org/10.1007/978-3-319-64024-2_12

Ewing, P. M., Runck, B. C., Kono, T. Y. J., & Kantar, M. B. (2019). The home field advantage of modern plant breeding. *PLoS ONE*, *14*(12), e0227079. https://doi.org/10.1371/journal.pone.0227079

Falconer, D. S., & Mackay, T. F. C. (1996). *Introduction to quantitative genetics* (4th ed.). Pearson.

FAOSTAT. (2022). *Database of agricultural production*. Food and Agriculture Organization of the United Nations, Rome, Italy.

Faye, J. M., Maina, F., Akata, E. A., Sine, B., Diatta, C., Mamadou, A., Marla, S., Bouchet, S., Teme, N., Rami, J. F., Fonceka, D., Cisse, N., & Morris, G. P. (2021). A genomics resource for genetics, physiology, and breeding of West African sorghum. *Plant Genome*, *14*(2), e20075. https://doi.org/10.1002/tpg2.20075

Fu, Y. B. (2015). Understanding crop genetic diversity under modern plant breeding. *Theoretical and Applied Genetics*, *128*(11), 2131–2142. https://doi.org/10.1007/s00122-015-2585-y

Girma, G., Nida, H., Seyoum, A., Mekonen, M., Nega, A., Lule, D., Dessalegn, K., Bekele, A., Gebreyohannes, A., Adeyanju, A., Tirfessa, A., Ayana, G., Taddese, T., Mekbib, F., Belete, K., Tesso, T., Ejeta, G., & Mengiste, T. (2019). A large-scale genome-wide association analyses of Ethiopian sorghum landrace collection reveal loci associated with important traits. *Frontiers in Plant Science*, *10*, 691. https://doi.org/10.3389/fpls.2019.00691

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., & Rokhsar, D. S. (2012). Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Research*, *40*(D1), 1178–1186. https://doi.org/10.1093/nar/gkr944

Hamblin, M. T., Salas Fernandez, M. G., Casa, A. M., Mitchell, S. E., Paterson, A. H., & Kresovich, S. (2005). Equilibrium processes cannot explain high levels of short- and medium-range linkage disequilibrium in the domesticated grass *Sorghum bicolor*. *Genetics*, *171*(3), 1247–1256. https://doi.org/10.1534/genetics.105.041566

Hammer, G. L. (2006). Pathways to prosperity: Breaking the yield barrier in sorghum. *Agricultural Science*, *19*(2), 16–22.

Hedden, P. (2003). The genes of the green revolution. *Trends in Genetics*, *19*(1), 5–9. https://doi.org/10.1016/S0168-9525(02)00009-4

Hill, W. G., & Weir, B. S. (1988). Variances and covariances of squared linkage disequilibria in finite populations. *Theoretical Population Biology*, *33*(1), 54–78. https://doi.org/10.1016/0040-5809(88)90004-4

Hilley, J. L., Weers, B. D., Truong, S. K., McCormick, R. F., Mattison, A. J., McKinley, B. A., Morishige, D. T., & Mullet, J. E. (2017). Sorghum *Dw2* encodes a protein kinase regulator of stem internode length. *Scientific Reports*, *7*(1), 1–13. https://doi.org/10.1038/s41598-017-04609-5

Holland, J. B. (2004). Breeding: Incorporation of exotic germplasm. In *Encyclopedia of plant and crop science* (pp. 222–224). https://doi.org/10.1081/E-EPCS-120010536

Hu, Z., Olatoye, M. O., Marla, S., & Morris, G. P. (2019). An integrated genotyping-by-sequencing polymorphism map for over 10,000 Sorghum genotypes. *The Plant Genome*, *12*(1), 180044. https://doi.org/10.3835/plantgenome2018.06.0044

Jordan, D. R., Mace, E. S., Cruickshank, A. W., Hunt, C. H., & Henzell, R. G. (2011). Exploring and exploiting genetic variation from unadapted sorghum germplasm in a breeding program. *Crop Science*, *51*(4), 1444–1457. https://doi.org/10.2135/cropsci2010.06.0326

Kong, W., Nabukalu, P., Cox, T. S., Goff, V. H., Pierce, G. J., Lemke, C., Robertson, J. S., Compton, R., Tang, H., & Paterson, A. H. (2020). Transmission genetics of a *Sorghum bicolor × S. halepense* backcross populations. *Frontiers in Plant Science*, *11*, 467. https://doi.org/10.3389/fpls.2020.00467

Lasky, J. R., Upadhyaya, H. D., Ramu, P., Deshpande, S., Hash, C. T., Bonnette, J., Juenger, T. E., Hyma, K., Acharya, C., Mitchell, S. E., Buckler, E. S., Brenton, Z., Kresovich, S., & Morris, G. P. (2015). Genome-environment associations in sorghum landraces predict adaptive traits. *Science Advances*, *1*(6), e1400218. https://doi.org/10.1126/sciadv.1400218

Li, X., Li, X., Fridman, E., Tesso, T. T., & Yu, J. (2015). Dissecting repulsion linkage in the dwarfing gene Dw3 region for sorghum plant height provides insights into heterosis. *Proceedings of the National Academy of Sciences*, *112*(38), 11823–11828. https://doi.org/10.1073/pnas.1509229112

Mace, E., Innes, D., Hunt, C., Wang, X., Tao, Y., Baxter, J., Hassall, M., Hathorn, A., & Jordan, D. (2019). The Sorghum QTL atlas: A powerful tool for trait dissection, comparative genomics and crop improvement. *Theoretical and Applied Genetics*, *132*(3), 751–766. https://doi.org/10.1007/s00122-018-3212-5

Mace, E. S., Hunt, C. H., & Jordan, D. R. (2013). Supermodels: Sorghum and maize provide mutual insight into the genetics of flowering time. *Theoretical and Applied Genetics*, *126*(5), 1377–1395. https://doi.org/10.1007/s00122-013-2059-z

MacQueen, A. H., Khoury, C. K., Miklas, P., McClean, P. E., Osorno, J. M., Runck, B. C., White, J. W., Kantar, M. B., Ewing, P. M., & Ewing, P. M. (2022). Local to continental-scale variation in fitness and heritability in common bean. *Crop Science*, *62*(2), 767–779. https://doi.org/10.1002/csc2.20694

Mathur, S., Umakanth, A. V., Tonapi, V. A., Sharma, R., & Sharma, M. K. (2017). Sweet sorghum as biofuel feedstock: Recent advances and available resources. *Biotechnology for Biofuels and Bioproducts*, *10*, 146. https://doi.org/10.1186/s13068-017-0834-9

Menamo, T., Kassahun, B., Borrell, A. K., Jordan, D. R., Tao, Y., Hunt, C., & Mace, E. (2021). Genetic diversity of Ethiopian sorghum reveals signatures of climatic adaptation. *Theoretical and Applied Genetics*, *134*(2), 731–742. https://doi.org/10.1007/s00122-020-03727-5

Menkir, A., Bramel-Cox, P. J., & Witt, M. D. (1994). Selection for agronomically acceptable inbred lines in adapted x exotic sorghum backcross populations. *Crop Science*, *34*(4), 1084–1089. https://doi.org/10.2135/cropsci1994.0011183X003400040046x

Morris, G. P., Ramu, P., Deshpande, S. P., Hash, C. T., Shah, T., Upadhyaya, H. D., Riera-Lizarazu, O., Brown, P. J., Acharya, C. B., Mitchell, S. E., Harriman, J., Glaubitz, J. C., Buckler, E. S., & Kresovich, S. (2013). Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proceedings of the National Academy of Sciences*, *110*(2), 453–458. https://doi.org/10.1073/pnas.1215985110

Murphy, R. L., Morishige, D. T., Brady, J. A., Brady, J. A., Rooney, W. L., Yang, S., Klein, P. E., & Mullet, J. E. (2014). Ghd7 (Ma6) represses sorghum flowering in long days: Ghd7 alleles enhance biomass accumulation and grain production. *The Plant Genome*, *7*(2), 1–10.

Murray, S. C., Sharma, A., Rooney, W. L., Klein, P. E., Mullet, J. E., Mitchell, S. E., & Kresovich, S. (2008). Genetic improvement of sorghum as a biofuel feedstock: I. QTL for stem sugar and grain nonstructural carbohydrates. *Crop Science*, *48*(6), 2165–2179.

Neph, S., Kuehn, M. S., Reynolds, A. P., Haugen, E., Thurman, R. E., Johnson, A. K., Rynes, E., Maurano, M. T., Vierstra, J., Thomas, S., Sandstrom, R., Humbert, R., & Stamatoyannopoulos, J. A. (2012). BEDOPS: High-performance genomic feature operations. *Bioinformatics*, *28*(14), 1919–1920. https://doi.org/10.1093/bioinformatics/bts277

Paterson, A. H., Bowers, J. E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., Haberer, G., Hellsten, U., Mitros, T., Poliakov, A., Schmutz, J., Spannagl, M., Tang, H., Wang, X., Wicker, T., Bharti, A. K., Chapman, J., Feltus, F. A., Gowik, U., ... Rokhsar, D. S. (2009). The *Sorghum bicolor* genome and the diversification of grasses. *Nature*, *457*(7229), 551–556.

Paterson, A. H., Saranga, Y., Menz, M., Jiang, C. X., & Wright, R. J. (2003). QTL analysis of genotype x environment interactions affecting cotton fiber quality. *Theoretical and Applied Genetics*, *106*(3), 384–396. https://doi.org/10.1007/s00122-002-1025-y

Peng, J., Richards, D. E., Hartley, N. M., Murphy, G. P., Devos, K. M., Flintham, J. E., Beales, J., Fish, L. J., Worland, A. J., Pelica, F., Sudhakar, D., Christou, P., Snape, J. W., Gale, M. D., & Harberd, N. P. (1999). "Green revolution" genes encode mutant gibberellin response modulators. *Nature*, *400*(6741), 256–261.

Piper, J. K., & Kulakow, P. A. (1994). Seed yield and biomass allocation in *Sorghum bicolor* and F1 and backcross generations of *S. bicolor* x *S. halepense* hybrids. *Canadian Journal of Botany*, *72*(4), 468–474.

Poland, J. A., Brown, P. J., Sorrells, M. E., & Jannink, J. L. (2012). Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE*, *7*(2), e32253. https://doi.org/10.1371/journal.pone.0032253

Quinby, J. R. (1974). The genetic control of flowering and growth in sorghum. *Advances in Agronomy*, *25*(C), 125–162. https://doi.org/10.1016/S0065-2113(08)60780-4

Reif, J. C., Zhang, P., Dreisigacker, S., Warburton, M. L., van Ginkel, M., Hoisington, D., Bohn, M., & Melchinger, A. E. (2005). Wheat genetic diversity trends during domestication and breeding. *Theoretical and Applied Genetics*, *110*(5), 859–864. https://doi.org/10.1007/s00122-004-1881-8

Rockström, J., Karlberg, L., Wani, S. P., Barron, J., Hatibu, N., Oweis, T., Bruggeman, A., Farahani, J., & Qiang, Z. (2010). Managing water in rainfed agriculture-the need for a paradigm shift. *Agricultural Water Management*, *97*(4), 543–550. https://doi.org/10.1016/j.agwat.2009.09.009

Rooney, W. L., & Aydin, S. (1999). Genetic control of a photoperiod-sensitive response in *Sorghum bicolor* (L.) Moench. *Crop Science*, *39*(2), 397–400. https://doi.org/10.2135/cropsci1999.0011183X0039000200016x

Tesso, T., Kapran, I., Grenier, C., Snow, A., Sweeney, P., Pedersen, J., Marx, D., Bothma, G., & Ejeta, G. (2008). The potential for crop-to-wild gene flow in sorghum in Ethiopia and Niger: A geographic survey. *Crop Science*, *48*(4), 1425–1431. https://doi.org/10.2135/cropsci2007.08.0441

Tian, F., Bradbury, P. J., Brown, P. J., Hung, H., Sun, Q., Flint-Garcia, S., Rocheford, T. R., McMullen, M. D., Holland, J. B., & Buckler, E. S. (2011). Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nature Genetics*, *43*(2), 159–162. https://doi.org/10.1038/ng.746

Upadhyaya, H. D., Pundir, R. P. S., Dwivedi, S. L., Gowda, C. L. L., Reddy, V. G., & Singh, S. (2009). Developing a mini core collection of sorghum for diversified utilization of germplasm. *Crop Science*, *49*(5), 1769–1780. https://doi.org/10.2135/cropsci2009.01.0014

Vadez, V., Krishnamurthy, L., Hash, C. T., Upadhyaya, H. D., & Borrell, A. K. (2011). Yield, transpiration efficiency, and water-use variations and their interrelationships in the sorghum reference collection. *Crop and Pasture Science*, *62*(8), 645–655. https://doi.org/10.1071/CP11007

Vales, M. I., Schön, C. C., Capettini, F., Chen, X. M., Corey, A. E., Mather, D. E., Mundt, C. C., Richardson, K. L., Sandoval-Islas, J. S., Utz, H. F., & Hayes, P. M. (2005). Effect of population size on the estimation of QTL: A test using resistance to barley stripe rust. *Theoretical and Applied Genetics*, *111*(7), 1260–1270. https://doi.org/10.1007/s00122-005-0043-y

Wondimu, Z., Dong, H., Paterson, A., Worku, W., & Bantte, K. (2021). Genetic diversity, population structure and selection signature in Ethiopian Sorghum [*Sorghum bicolor* L. (Moench)] germplasm. *G3*, *11*(6), jkab087. https://doi.org/10.1093/g3journal/jkab087

Yu, J., Holland, J. B., McMullen, M. D., & Buckler, E. S. (2008). Genetic design and statistical power of nested association mapping in maize. *Genetics*, *178*(1), 539–551. https://doi.org/10.1534/genetics.107.074245

Zhang, D., Kong, W., Robertson, J., Goff, V. H., Epps, E., Kerr, A., Mills, G., Cromwell, J., Lugin, Y., Phillips, C., & Paterson, A. H. (2015). Genetic analysis of inflorescence and plant height components in sorghum (Panicoidae) and comparative genetics with rice (Oryzoidae). *BMC Plant Biology*, *15*(1), 1–15. https://doi.org/10.1186/s12870-015-0477-6

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.