ORIGINAL RESEARCH

Ecology and Evolution
Open Access   WILEY

Genomic signatures of adaptation to Sahelian and Soudanian climates in sorghum landraces of Senegal

Jacques M. Faye[1] | Fanna Maina[1,2] | Zhenbin Hu[1] | Daniel Fonceka[3,4] | Ndiaga Cisse[3] | Geoffrey P. Morris[1] (iD)

[1]Department of Agronomy, Kansas State University, Manhattan, Kansas

[2]Institut National de la Recherche Agronomique du Niger, Niamey, Niger

[3]Centre d'Étude Régional pour l'Amélioration de l'Adaptation à la Sécheresse, Thiès, Sénégal

[4]CIRAD, UMR AGAP, Montpellier, France

Correspondence
Geoffrey P. Morris, Department of Agronomy, Kansas State University, Manhattan, KS.
Email: gpmorris@ksu.edu

Funding information
United States Agency for International Development, Grant/Award Number: AID-OAA-A-13-00047

Abstract

Uncovering the genomic basis of climate adaptation in traditional crop varieties can provide insight into plant evolution and facilitate breeding for climate resilience. In the African cereal sorghum (*Sorghum bicolor* L. [Moench]), the genomic basis of adaptation to the semiarid Sahelian zone versus the subhumid Soudanian zone is largely unknown. To address this issue, we characterized a large panel of 421 georeferenced sorghum landrace accessions from Senegal and adjacent locations at 213,916 single-nucleotide polymorphisms (SNPs) using genotyping-by-sequencing. Seven subpopulations distributed along the north-south precipitation gradient were identified. Redundancy analysis found that climate variables explained up to 8% of SNP variation, with climate collinear with space explaining most of this variation (6%). Genome scans of nucleotide diversity suggest positive selection on chromosome 2, 4, 5, 7, and 10 in durra sorghums, with successive adaptation during diffusion along the Sahel. Putative selective sweeps were identified, several of which colocalize with stay-green drought tolerance (*Stg*) loci, and a priori candidate genes for photoperiodic flowering and inflorescence morphology. Genome-wide association studies of photoperiod sensitivity and panicle compactness identified 35 and 13 associations that colocalize with a priori candidate genes, respectively. Climate-associated SNPs colocalize with *Stg3a*, *Stg1*, *Stg2*, and *Ma6* and have allelic distribution consistent with adaptation across Sahelian and Soudanian zones. Taken together, the findings suggest an oligogenic basis of adaptation to Sahelian versus Soudanian climates, underpinned by variation in conserved floral regulatory pathways and other systems that are less understood in cereals.

KEYWORDS
Africa, drought tolerance, local adaptation, selective sweeps, stay-green

1 | INTRODUCTION

Local adaptation is critical for survival of traditional crop varieties in stressful environments (Camus-Kulandaivelu et al., 2006; Xu et al., 2006). Smallholder farmers in developing countries are particularly vulnerable to environmental factors such as drought and heat stress limiting crop production (Morton, 2007). Climatic gradients in relation to precipitation are major drivers of adaptation in plants including traditional crop varieties (Fournier-Level et al., 2011; Lasky et al., 2015, 2012; Siepielski et al., 2017; Vigouroux et al., 2011). Adaptation to water-limited environments involves phenological, physiological, and morphological traits such as photoperiod sensitivity, delayed senescence, and inflorescence morphology (Blum, 2014). For instance, when growing seasons are shortened by end-of-season droughts, selection favors early maturity alleles to escape drought (Franks, Sim, & Weis, 2007; Kenney, McKay, Richards, & Juenger, 2014; Vigouroux et al., 2011). Identifying genetic polymorphisms underlying adaptive traits and their eco-geographic distributions is necessary to understand the genetic basis of local adaptation of landraces (Romero Navarro et al., 2017).

The patterns of genome-wide nucleotide polymorphisms provide insight into selective forces varying over time and space (Olsen et al., 2006; Slatkin, 2008). Recent studies in rice (Caicedo et al., 2007; Li, Li, Jia, Caicedo, & Olsen, 2017), tomato (Lin et al., 2014), and maize (Swarts et al., 2017) have shown that high genetic differentiation among populations reflects adaptation to specific agroclimatic zones. Population genomic approaches for identifying signatures of selection include decreased pairwise nucleotide diversity, composite likelihood ratio (CLR) analysis for selective sweeps, and genome–environment associations (GEA) (Fang et al., 2017; Fournier-Level et al., 2011; Lasky et al., 2015; Li et al., 2017; Lin et al., 2014). The CLR analysis in SweeD is relatively robust to demographic events because the method conservatively estimates the neutral site frequency spectrum (SFS) based on the observed data (Nielsen et al., 2005; Pavlidis, Živković, Stamatakis, & Alachiotis, 2013). Linear regression models and genome-wide association studies (GWAS) mixed models are common methods used for GEA, especially to investigate adaptation to environmental gradients (Rellstab, Gugerli, Eckert, Hancock, & Holderegger, 2015), and have been applied by several studies in plants and crop species (Fournier-Level et al., 2011; Lasky et al., 2015; Yoder et al., 2014). Redundancy analysis (RDA) provides an estimate of allelic variance explained by climatic factors based on multivariate linear regressions (Meirmans, 2015). Genome-wide association studies can provide high mapping resolution of adaptive traits in diverse populations (Cavanagh, Morell, Mackay, & Powell, 2008).

Sorghum (*Sorghum bicolor* L. [Moench]) is a staple food crop for smallholder farmers in semiarid regions worldwide. The modest genome size (~800 Mbp) of sorghum relative to other grass species (Paterson et al., 2009) makes it a tractable system for the genomic studies of local adaptation. Five botanical types (bicolor, durra, guinea, caudatum, and kafir) have been described (Harlan & De Wet, 1972). Durra types, known for their adaptation to arid zones,

are thought to have originated in Ethiopia before westward diffusion along the Sahel to West Africa and finally Senegal (Harlan & De Wet, 1972). Guinea types, known for their humid adaptation (Deu et al., 1994; Folkertsma, Rattunde, Chandra, Raju, & Hash, 2005), may reflect a second center of domestication in the humid savanna of West Africa (Deu et al., 1994; Doggett, ; Folkertsma et al., 2005). Inflorescence morphology is a major component of agroclimatic adaptation in sorghum and varies from loose panicle in guinea to compact panicle in durra sorghum (Brown et al., 2006). Most traditional sorghum varieties in West Africa are photoperiod sensitive such that grain maturation coincides with the end of the rainy season (Bhosale et al., 2012; Sanon et al., 2014). In U.S. sorghum, variation in flowering time is controlled by conserved cereal floral regulatory networks, including phytochromes (*Ma3/PhyB*, *Ma5/PhyC*), CCT-domain regulators (*Ma1/PRR37*, *SbEhd1*, *SbEhd2*), and florigens (*SbCN15/Hd3a*, *SbCN12*) (Mullet et al., 2010; Murphy et al., 2011). Several quantitative trait loci (QTL) (*Stg1–4*) confer stay-green (i.e., delayed leaf senescence) postflowering drought tolerance in lines derived from Ethiopian durra (Borrell et al., 2014; Harris et al., 2007; Kebede, Subudhi, Rosenow, & Nguyen, 2001; Tuinstra, Grote, Goldsbrough, & Ejeta, 1997), but it is not known whether these loci contribute to drought adaptation more widely across the Sahel.

Analyses of genetic diversity, linkage disequilibrium (LD), and GEA have provided an understanding of worldwide sorghum genetic structure across diverse agroclimatic regions (Bouchet et al., 2012; Lasky et al., 2015; Mace et al., 2013; Morris et al., 2013; Wang et al., 2013). However, the genomic basis of climate adaptation at a regional scale remains poorly understood. The variation of agroclimatic conditions in Senegal reflects the sub-Saharan climatic gradient with increasing annual precipitation from north to south across the Sahelian (~200–600 mm) and Soudanian zones (~600–1100 mm). A large panel of sorghum landraces was collected from these agroclimatic zones in Senegal in the 1970s (Clément & Houdiard, 1977). To better understand the genomic basis of Sahelian and Soudanian climate adaptation, we used genotyping-by-sequencing (GBS) to characterize genome-wide single-nucleotide polymorphism (SNP) in georeferenced and phenotyped Senegalese sorghum landraces. We characterized population structure of genomic diversity, identified signatures of selection, and mapped genetic polymorphisms associated with phenotype and climate. The findings suggest that climate has shaped genomic variation across Sahelian and Soudanian zones, with variation in floral regulatory pathways and other systems contributing to this adaptation.

2 | MATERIALS AND METHODS

2.1 | Plant materials

The Senegalese sorghum germplasm (SSG) used in the present study were obtained from the U.S. Department of Agriculture (USDA) Germplasm Resources Information Network (GRIN). These accessions (*n* = 341) were collected from various agro-ecological zones of Senegal in 1976 (Clément & Houdiard, 1977). Germplasm Resources

Information Network accessions from neighboring countries of Gambia (n = 60), which is surrounded by Senegal, and Mauritania (n = 15), which shares border along the Senegal River Valley, were also included in our panel. Six improved varieties (CE 151-262, CE 180-33, ISRA-S-621-B, 53-49, CE 260-12-1-1, IRAT 4) from the sorghum breeding program based at the Centre National de Recherche Agricole (CNRA) and two sorghum conversion lines, SC 1,067 (PI 576,432) and SC 417 (PI 533,861), were included. Information about the SSG including botanical race, geographic origin, local name, and ethno-linguistic group from which the landrace was collected are presented in Data S1. Assignment in "durra" group was from the GRIN genebank, based on a phenotypic assessment. To compare the SSG landraces with the global sorghum diversity, we reanalyzed available raw sequencing data of worldwide sorghum diversity panels (Morris et al., 2013), hereafter referred to as the global diversity panel (GDP). This data set included 582 lines from the sorghum mini core collection and the Generation Challenge Program reference set, and 178 lines from the sorghum association panel. The GDP includes accessions from Africa, Asia, and the Americas.

2.2 | Genotyping-by-sequencing

Accessions of the SSG were grown in a glasshouse at Kansas State University. Leaf tissues from each accession were harvested from two weeks old seedlings (five seedlings pooled per accession), placed into 96-well plates, and dried in a lyophilizer for two days. Genomic DNA of SSG accessions was extracted from ~50 mg dried leaf tissue using the BioSprint robot with DNeasy Mini Kit (Qiagen) according to the manufacturer's instructions. DNA was quantified with PicoGreen and normalized to 10 ng/µl DNA for each sample. The GBS library was constructed using the restriction enzyme ApeKI for DNA digestion and 384-plex barcode ligation (4 × 96-plex) following the GBS protocol (Elshire et al., 2011). Digested DNA fragments were ligated to the barcode-adapters in a solution containing the 10× T4 DNA Ligase Reaction Buffer, ultrapure water, and T4 DNA Ligase (New England Biolabs), then cleaned using a QIAquick PCR purification kit (Qiagen). The adapter-ligated DNA fragments were amplified by polymerase chain reaction (PCR). The PCR-amplified DNA fragments were cleaned and quantified with PicoGreen. Four 96-plex libraries were pooled to form a 384-plex GBS library. GBS libraries were diluted into 20 µl at 4 nM for each library and analyzed by the Agilent 2100 Bioanalyzer for sequencing. GBS libraries were sequenced on Illumina HiSeq 2500 at the University of Kansas Medical Center.

2.3 | SNP calling

The SNP calling was done based on 1,208 samples including the accessions from the SSG panel and accessions from the GDP. Single-end sequence reads obtained from Illumina sequencing and raw sequencing data from the GDP were processed with the TASSEL 5 GBS v2 pipeline (Glaubitz et al., 2014). All unique sequence reads were trimmed to 64 bp, which was the default setting. The first step in the pipeline (GBSSeqToTagDBPlugin) allowed to collapse identical reads into tags using the key files of both SSG and GDP accessions.

Distinct tags were pulled and exported from the database in the FASTQ format using the TagExportToFastqPlugin for their alignment to the BTx623 sorghum reference genome v.3.1 (McCormick et al., 2018; Paterson et al., 2009). The alignment was performed with the Burrows–Wheeler Alignment (Li & Durbin, 2009) where the created SAM file was passed through the SAMToGBSdbPlugin to store the position information of aligned tags. The SNPs were called from the aligned tags. The DiscoverySNPCallerPlugin was used to identify SNPs from the aligned tags where minor allele frequencies (MAF) was set to 0.0001 and minimum locus coverage (mnLCov) was kept as the default setting of 0.1. For downstream population genomic analyses, SNPs with <20% missing data rate and MAF > 0.01 were retained. Monomorphic sites were removed and only biallelic sites were retained. Missing genotypes were imputed using Beagle v4.1 program (Browning & Browning, 2016). For the association mapping studies, the SNP data set was filtered for MAF > 0.05 to reduce the chance of observing false-positive associations.

2.4 | Population structure analysis

Principal components analysis (PCA) of SNP variation was performed using the snpgdsPCA function of the R package SNPRelate (Zheng et al., 2012). Neighbor-joining (NJ) analysis was performed using TASSEL 5 program, and the tree was visualized with the ape package in R (Paradis, Claude, & Strimmer, 2004). Bayesian model-based clustering in ADMIXTURE v1.23 (Alexander, Novembre, & Lange, 2009) was used to estimate the subpopulation membership/admixture for K = 2–20 subpopulations. To reduce SNP redundancy due to LD for the admixture analysis, genotypic data were LD-pruned with a window size of 50 SNPs, step size 10, and VIF threshold of 0.5 using the function indep in PLINK 1.9 (Purcell et al., 2007). Default settings of ADMIXTURE were used, and fivefold cross validation (CV) error with block bootstrap and 2,000 iterations was used to determine the optimum value of K. Each accession was assigned to subpopulation when the proportion of the coefficient of membership to subpopulation was >0.60. To determine the spatial genetic co-ancestry structure with respect to geography, we used the R package TESS3 (Caye, Deist, Martins, Michel, & François, 2016). Results were visualized using the R program (R Core Team, 2016).

2.5 | Linkage disequilibrium analysis

LD was characterized in the whole SSG and separately in the guinea and durra accessions. VCFtools (Danecek et al., 2011) was used to filter the genotypic data based on MAF > 0.05. The pairwise correlation coefficient (r^2) among SNPs was used to estimate LD using TASSEL 5 (Bradbury et al., 2007). LD decay, measured as the distance by which the r^2 decays to half its maximum value, was fit using the nonlinear least square (nls) function (Hill & Weir, 1988; Remington et al., 2001) in R program. The R package LDheatmap 0.99-4 (Shin, Blay, McNeney, & Graham, 2006) was used to determine and display the pairwise LD surrounding (50 kb region from both sides of the SNP) a SNP-environment variable association.

2.6 | Genome-wide nucleotide variation and genome scans

Minor allele frequencies and observed and expected heterozygosity for SNP markers were calculated using VCFtools program and R program (R Core Team, 2016). Pairwise genetic differentiation (F_{ST}) among subgroups defined based on eco-geography was estimated using the Weir and Cockerham method in VCFtools. F_{ST} values among subgroups obtained at $K = 7$ from the TESS3 program were calculated using the R package HierFstat (de Meeûs & Goudet, 2007). Pairwise genome-wide nucleotide diversity (π) and Tajima's D test statistics were calculated based on nonoverlapping sliding windows of 1 Mbp across the genome using VCFtools. Ratios of π were analyzed between guinea and durra accessions in the SSG (π_{guinea}/π_{durra}), and across putative prebottleneck and postbottleneck events ($\pi_{guinea}/\pi_{Ethiopia\ durra}$, $\pi_{Ethiopia\ durra}/\pi_{Niger\ and\ Mali\ durra}$, and $\pi_{Niger\ and\ Mali\ durra}/\pi_{Senegal\ durra}$). Selective sweeps were detected using the CLR method in SweeD program (Pavlidis et al., 2013). Each chromosome was divided into 5,000 grid points (nonoverlapping windows). The CLR windows with ≥8 SNPs (approximately 1 SNP per 2 kb) were retained during the analysis. The significance threshold representing the 95th percentile cutoff was determined based on 1,000 simulations.

2.7 | Genome-wide association studies (GWAS)

GWAS were carried out using mixed-linear models (MLM) in GAPIT in R (Lipka et al., 2012) with the three first principal components eigenvectors and kinship matrix. The Bonferroni correction at $\alpha = 0.05$ level was used to define the significance of association tests. SNPs were filtered at MAF > 0.05, yielding 145,235 SNPs. Phenotypic data were obtained from the GRIN database and treated as binary data for both photoperiod sensitivity (e.g., sensitive vs. insensitive) and panicle compactness (e.g., compact vs. open panicle). For GEA, both MLM and general linear models (GLM) were used. Nineteen WorldClim-derived bioclimatic variables (Hijmans, Cameron, Parra, Jones, & Jarvis, 2005) were used for genome–environment association tests. To identify environment-associated SNPs with the greatest significance among SNPs of the same genomic region, the multilocus mixed-model (MLMM) (Segura et al., 2012) was used to complement the GLM and MLM. In both MLM and MLMM, the first three principal components were included to account for population structure.

2.8 | A priori candidate genes

A list of a priori candidate genes for climate adaptation was defined from known sorghum genes, orthologs of cloned genes from rice and maize, and candidates from previous sorghum mapping studies (see Data S2 for candidate genes, gene functions, and references). A literature survey of sorghum orthologs of maize and rice genes that affect inflorescence architecture, flowering time, and drought tolerance was carried out. Inflorescence architecture candidate genes from a previous global GWAS (Morris et al., 2013), photoperiodic flowering time candidate genes from a previous study (Bhosale et al., 2012), and validated drought tolerance loci

(stay-green, Stg1–4) from (Borrell et al., 2014) were included. Genomic position of candidate genes was determined using Phytozome v12.1.6 (https://phytozome.jgi.doe.gov) (Goodstein et al., 2012).

2.9 | Redundancy analysis

RDA was performed using the R package vegan (Oksanen et al., 2017) for climatic factors, ethnicity, and space. Independent variables included nineteen climatic variables, space variables (latitude and longitude), and ethnicity variables. Ethnicity was coded as binary variable indicating the ethno-linguistic group of the farmer that contributed the landrace to the collection (Clément & Houdiard, 1977). Forward selection based on 1,000 permutations was performed for space (e.g., using polynomial coordinates), climate, and ethnicity variables to include only the meaningful variables for ordination. The total among-population genetic variance was partitioned into space, climate, ethnicity, and their overlapping fractions using 1,000 randomly selected SNP (MAF > 0.05). The significance of each variance fraction was tested with 1,000 permutations.

3 | RESULTS

3.1 | Genome-wide SNP variation in Senegalese sorghum

The Senegalese sorghum accessions included in this study originated from diverse agroclimatic zones (Figure 1a), agro-ecological regions (Figure S1a), and ethnic–linguistic groups (Figure S1b). Across 421 accessions, we identified 213,916 SNPs after filtering out SNPs with >20% missing data, MAF < 0.01, and retaining only biallelic SNPs. The SNP density was determined based on nonoverlapping windows of 1 Mb where SNPs were distributed across the genome with higher density in the subtelomeric regions (Figure S2a). The SNPs covered most of the genome with an average coverage of 1 SNP every 2 kb. The average observed and expected heterozygosity in the SSG were estimated at 0.05 and 0.23, respectively. The average pairwise nucleotide diversity (π) was 0.00054 in durra and 0.00060 in guinea accessions. The average pairwise LD (r^2) decreased from its initial value (~0.5) to 0.2 at 220 kb, 150 kb, and 81 kb in durra, guinea, and whole SSG, respectively (Figure S2b). LD decayed to background level (~0.1) at 880 kb in durra and 430 kb in guinea. The SSG had a lesser proportion of low frequency minor alleles (<5% MAF) and greater proportion of intermediate frequency minor alleles than the GDP, based on nonoverlapping window size of 1 Mb (Figure S2c). About 60% of SNPs were rare (MAF < 0.05).

Next, we investigated the genetic variation and structure of the SSG. The two first principal components explained 3.8% and 2.5% of SNP variation (Figure 1b). The accessions originated from the center formed one cluster, accessions from the south formed a second cluster, and accessions from the north formed a third cluster. The third cluster included durra accessions, caudatum accessions, a few guinea accessions from the north, and improved varieties. Neighbor-joining tree matched the PCA results and revealed that SSG durra accessions
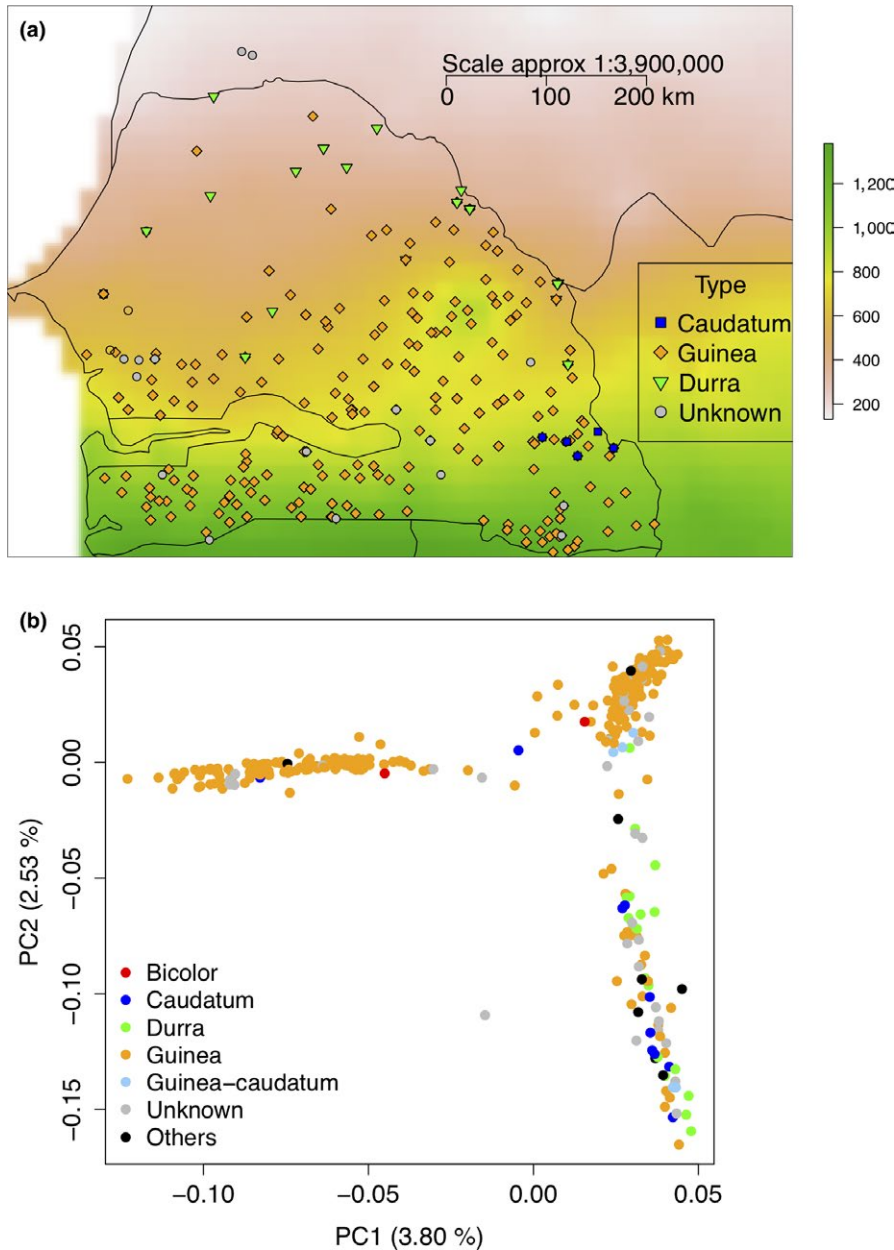
FIGURE 1 SNP variation in the Senegalese sorghums accessions. (a) Geographic distribution of the Senegalese sorghums accessions along precipitation gradient. The accessions are colored coded with respect to botanical race. The color background scale indicates the annual precipitation in millimeters with green color representing the highest precipitation of the Soudanian zone, pink representing lowest precipitation of the Sahelian zone, and yellow representing the zone of transition between Sahelian and Soudanian zones. (b) Scatterplot of the two first principal components explaining the genomic variation within the SSG collection

were closely related to the durra from Ethiopia and other West African countries (Figure S2d). Durra and guinea accessions within the SSG were genetically differentiated from each other. The SSG also clustered somewhat with respect to ethno-linguistic groups, which are nested within geographic origins of the accessions (Figure S2e).

3.2 | Model-based population structure and variance partitioning

To further characterize genetic structure and gene flow among groups, we used Bayesian model-based clustering. ADMIXTURE revealed a hierarchical genetic structure and high amount of gene flow among subpopulations (File S1). Cross validation error was minimized with $K = 7$ subpopulations (Figure S3). We investigated the spatial genetic co-ancestry in the SSG with TESS3 based on allele frequency

distribution and geographic origin. Seven optimum spatial genetic clusters ($K = 7$) were identified (Figure 2a). The TESS3 results matched the ADMIXTURE groups for different K values. Genetic differentiation among the subpopulations (including only samples with admixture rate ≥ 0.7) found at $K = 7$ from TESS3 results was determined using the F_{ST} analysis (Figure 2b). The durra accessions from the northern subpopulation (pop1) were more related to the improved varieties (pop4), based on F_{ST} analysis. Both pop1 and pop4 were distinct from central and southern subpopulations (pop2, 3, 5–7), which were mostly formed by guinea accessions where guineas in the center were differentiated from guineas in the south. F_{ST} of 0.185 and 0.052 were estimated between guinea and durra accessions in the SSG, and between SSG durra and GDP durra, respectively.

We used RDA to estimate the proportion on SNP variation explained by climate variation, ethno-linguistic origin, and space.
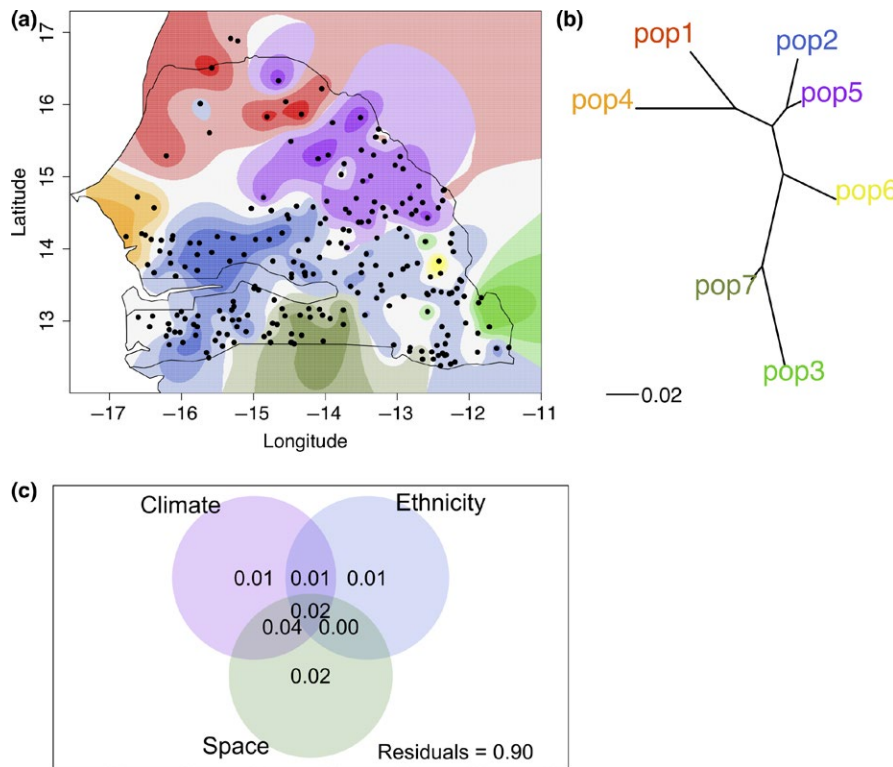
FIGURE 2 Spatial population structure and SNP variance partitioning in the Senegalese sorghum. (a) Spatial genetic co-ancestry structure of the accessions at $K = 7$. Each accession is represented by dot on the map and each color represents a genetic co-ancestry matrix. (b) The F_{ST} genetic differentiation among subpopulations at $K = 7$ ancestral groups from b; the color-coding matches that in a. (c) Among-population genetic variance at 1,000 randomly selected SNPs with MAF > 0.05 explained independently by climatic, space, and ethnicity variables

Climate and ethnicity explained up to 6% and 4% of SNP variance, respectively, including variance collinear with space ($p > 0.001$) (Figure 2c). After accounting for space, climate and ethnicity explain up to 2% of the variance, each. Climate collinear with space, the putative proportion of clinal adaptation, explained 6% of variance.

3.3 | Genome-wide patterns of nucleotide polymorphism

To identify genomic regions subject to selection, we compared genome-wide nucleotide polymorphism (π) between guinea (Soudanian) and durra (Sahelian) accessions within the SSG. Since guinea sorghums are generally more genetically diverse than durra sorghums, we used π_{guinea} in the numerator and π_{durra} in the denominator to identify low-diversity genomic regions in the durra genome. Nucleotide polymorphism was reduced in durra compared to guinea across most of the genome, with notably low π on pericentromeric regions of chromosome 2, 5, 7, and 10 (Figure 3a). For durra, 34 genomic regions (1 Mb windows) were identified as putative selected regions (top 5% cutoff > 1.96). A notable region of low π_{durra} on chromosome 1 colocalized with the *Ma3* photoperiodic flowering gene. Modestly lower π_{durra} was observed around *Stg1*, *Stg3a*, and *Stg3b*. Generally, negative values of Tajima's *D* were observed in durra, contrasting the positive values observed in guinea (Figure 3b).

To better understand the timing of putative selection events, we investigated ratios of nucleotide polymorphism across three putative genetic bottlenecks: (a) since the divergence of durra from its common ancestor with guinea types, (b) from Ethiopian durra (center of durra origin) to West African durra (Niger and Mali), (c) and from

West African durra to Senegalese durra (Figure 3c). We also characterized nucleotide polymorphism between all Sahelian durra against worldwide guinea (Figure S4). The π reduction in the pericentromeric regions of chromosome 4 occurred mainly in Ethiopian durra. The π reduction on pericentromeric regions of chromosomes 5 and 10 and subtelomeric region of chromosome 6 were common to all West African durra sorghums. The π reduction in the pericentromeric region of chromosome 2 was specific to the SSG durra.

3.4 | Selective sweeps and colocalization of a priori candidate genes

Next, we used CLR to identify candidate selective sweeps for Sahelian adaptation in durra in the SSG. Composite likelihood ratio identified 47 candidate genomic regions (top 5% cutoff or CLR > 16.9) in durra (Figure 4a). We investigated if a priori candidate genes ($n = 64$) implicated in stay-green, flowering time, or inflorescence morphology colocalized with CLR outliers. Given that the candidate genes were identified a priori from the literature, a liberal cutoff of 1 Mb was used to define colocalization between CLR outlier regions and candidate genes. Sixteen out of 47 CLR outliers colocalized with candidate genes (Data S3). The photoperiodic flowering genes *Ma3*, *GI*, *CRY1*, and *ZFL1* and inflorescence architecture candidate genes *HAM3*, *Sbra2*, and *vt2* colocalized with CLR outliers. The stay-green loci *Stg3a* and *Stg3b* colocalized with outlier regions on subtelomeric regions of chromosome 2. We used CLR in guinea to identify candidate selective sweeps for Soudanian adaptation. The CLR identified 28 candidate genomic regions (CLR > 10.3) in guinea (Figure 4b). Eleven out of 28 CLR outliers colocalized

FIGURE 3 Genome-wide pattern of nucleotide diversity in durra accessions. Decrease in pairwise nucleotide diversity and Tajima's *D* test for nonoverlapping sliding windows of 1 Mbp across the genome. (a) Decreased pairwise nucleotide diversity in durra relative to guinea in the Senegalese sorghum. The horizontal dashed lines indicate the mean value (blue) and the top 5% (gray) of decreased nucleotide diversity. (b) Tajima's *D* test between durra (green) and guinea (red) accessions in Senegalese sorghum. (c) Positive selections between durra from Ethiopia and all guineas in the global diversity panel (blue), between Ethiopian durra and West African durra (green), and between West African durra and Senegalese durra (red)
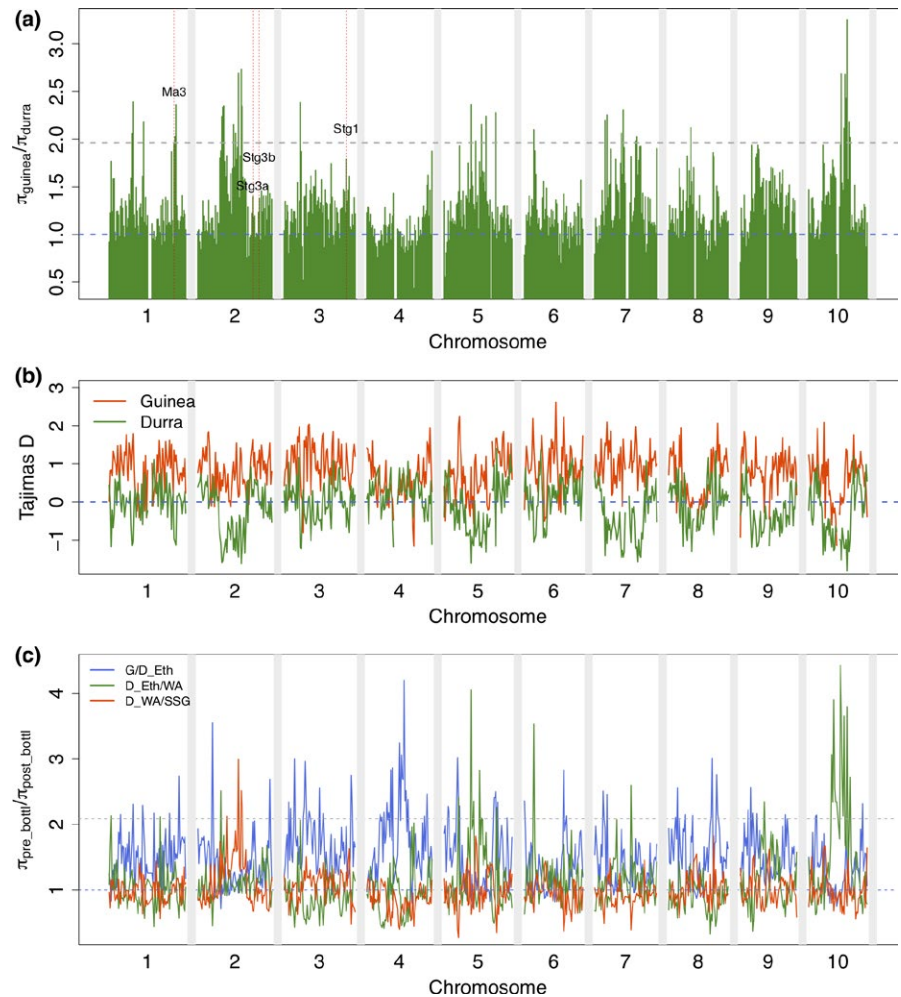
with candidate genes (Data S3). The photoperiodic flowering genes *PhyA*, *Hd1*, *SbCN2*, and *Ma6* colocalized with outlier regions. The stay-green locus *Stg1* colocalized with an outlier region on chromosome 3. The inflorescence morphology genes *IDS1*, *DFL2*, *Sbra3*, and *Dwarf8* colocalized with outlier regions.

## 3.5 | Genome-wide association studies of putative adaptive traits

To better characterize variation underlying putative adaptive traits, we mapped genotype–phenotype associations for photoperiodic flowering and inflorescence morphology. To reduce confounding effects of population structure, we also applied a regional mapping approach where durra accessions were excluded. In total, 445 and 178 significantly associated SNPs (Bonferroni *p*-value > 0.05) were identified for photoperiod sensitivity for the whole SSG and SSG without durra, respectively (Figure 5a and Figure S5a). Colocalization between associated SNPs and candidate genes was determined based on LD decay rate to background level ($r^2 = 0.1$) in durra (800 kb) and guinea (500 kb). Among the associated SNPs, 35 and 26 colocalized with photoperiodic flowering candidate genes for the whole SSG and SSG excluding durra, respectively. For panicle compactness, 48 and 124 significantly associated SNPs were found for the whole SSG

and SSG excluding durra, respectively (Figure 5b and Figure S5b). Among the associated SNPs, 13 SNPs colocalized with a priori candidate genes for inflorescence morphology.
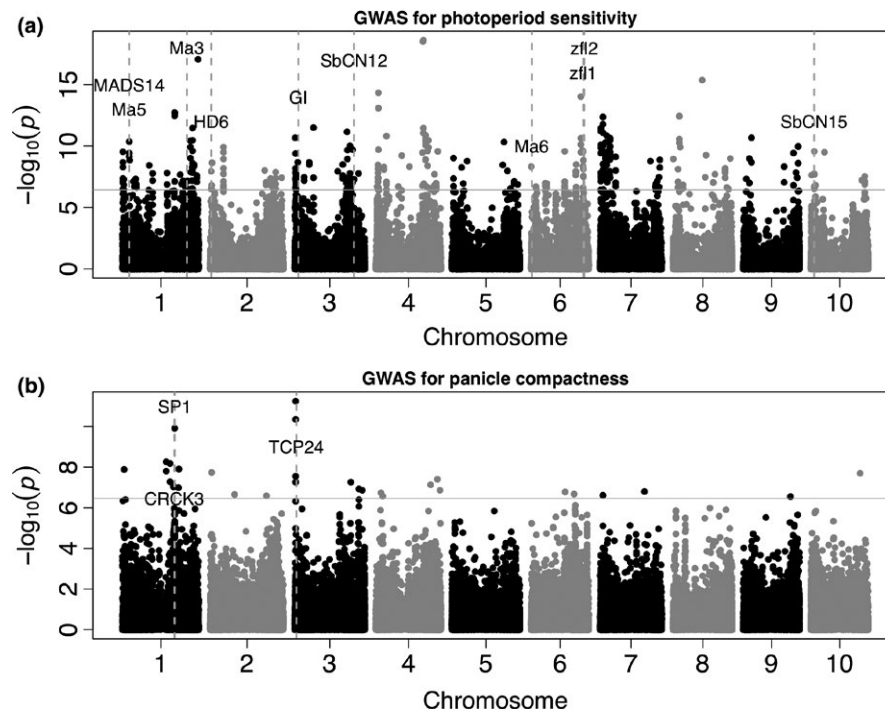
Photoperiod sensitivity-associated SNPs were found near floral regulators *Ma3*, *Ma5*, *Ma6*, *MADS14*, *GI*, *HD6*, *zfl1/2*, *Ehd2*, *SbCN12*, and *SbCN15* (Data S4). Most of these associations were observed whether or not durra were included. The association near *Ehd2* was only observed when durra accessions were excluded, while associations near *Ma6* and *HD6* were only observed when durra accessions were included. Eighteen of the highly significant (*p*-value > 10^{-10}) associations were not near any a priori candidate genes. For panicle compactness, significantly associated SNPs colocalized with *SP1*, *CRCK3*, *TCP24*, *DFL2*, *Sb-ra2*, *vt2*, and *rel2*. The SNP S1_55302939 (within the *SP1* gene) was significant in both GWAS approaches, while S1_55305415 (1 kb away from *SP1*) was only significant when using the whole SSG panel. Two of the highly significant (*p*-value > 10^{-10}) associations were not near a priori candidate genes.

3.6 | Environment-SNP associations

We performed GEA to identify SNPs associated with climate variables (Data S4). Based on the GLM, GEA identified 560 SNPs significantly associated (Bonferroni-adjusted *p*-value > 0.05) with
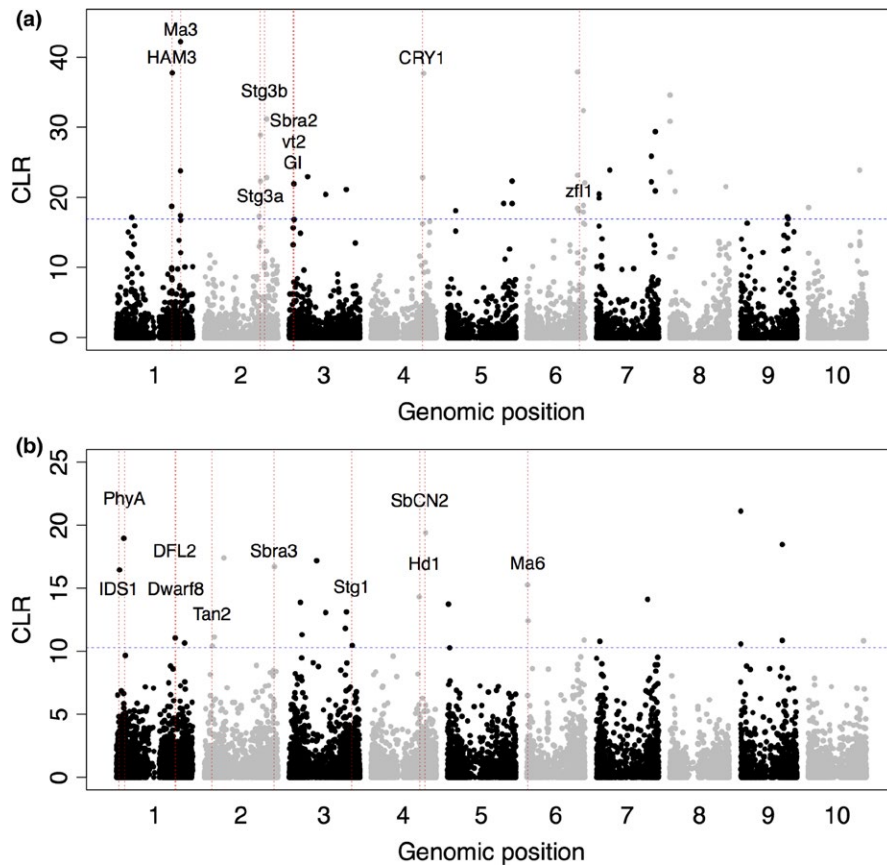
FIGURE 4 Genome-wide scan for selective sweeps in the Senegalese sorghum. Selective sweeps in the durra (a) and guinea (b) genomes. Each chromosome was divided into 5,000 grid points each corresponding to one dot. The *y*-axis represents the composite likelihood ratio (CLR) of each grid point. The vertical dashed lines indicate the colocalized candidate genes with genomic signatures. The horizontal dashed blue line represents the 95th percentile cutoff obtained from 1,000 simulations

FIGURE 5 GWAS of photoperiod sensitivity and panicle compactness. Manhattan plots of association tests using the Mixed-linear model for photoperiod sensitivity (a) and panicle compactness (b) for the whole Senegalese collection. The negative base 10 logarithm of the significance *p*-value (*y*-axis) of the SNP-phenotype association is plotted against the genomic position of each SNP on the chromosomes represented on the *x*-axis. The gray horizontal line indicates the significance threshold for the Bonferroni corrected *p*-value > 0.05. Candidate genes colocalizing with significantly associated SNPs are indicated

environment variables including precipitation of the driest quarter (Figure 6a), mean temperature of the warmest quarter (Figure S6a), and precipitation of the wettest quarter (Figure S6b). Associations for longitude variable were based on the MLM (Figure S6c) because GLM identified many associated SNPs. Multilocus mixed-model identified 16 significantly associated SNPs, including one overlapping SNP (S7_59683060) with the GLM, and 15 additional SNPs that were not identified by GLM or MLM (Table S1, Figure 6a, and Figure S6). Associated SNPs for precipitation of the driest quarter, such as S2_60708848 and S6_691400, identified by the MLMM, colocalized with the *Stg3a* locus and *Ma6* gene, respectively. The stay-green candidate loci (*Stg1–4*) colocalized with SNPs associated with mean temperature of the driest and warmest quarters, precipitation of the driest, warmest and wettest quarters, and longitude (Data S4). The SNP S1_7584419 identified by MLMM as associated with mean temperature of the warmest quarter colocalized with *Ma5* and *MADS14*, but at greater distance (>800 kb).

To determine the pairwise LD between the two SNPs colocalizing with *Stg3a* and *Ma6* and variation within these loci, we generated the LDheatmap of the 50 kb region surrounding each SNP (Figure 6b,c). Nearly complete LD ($r^2 > 0.9$) was found between S2_60708848 and other SNPs in the *Stg3a* locus. The SNP S6_691400 was in LD with two SNPs in *Ma6*. The genotypes carrying the minor allele at

S2_60708848 were distributed in the southern subhumid environments (Figure 6d). By contrast, genotypes carrying the minor allele at S6_691400 were distributed in the northern and dry environments (Figure 6e). The minor alleles at S3_67831630 (colocalized with *Stg1/SbPIN4*) and S3_57321183 (colocalized with *Stg2/SbPIN2*) were mostly found in durra landraces and few guinea landraces distributed in the dry areas of Senegal (Figure S7a,b).

4 | DISCUSSION

Genomic analysis of crop landraces can help determine the basis of local adaptation (Lasky et al., 2015; Li et al., 2017; Lin et al., 2014; Swarts et al., 2017). The aims of this study were to characterize factors shaping the genomic variation of Senegalese sorghum landraces, map genomic regions shaped by agroclimatic adaptation, and identify genes that could play a role in local adaptation.

4.1 | Factors shaping genomic variation of sorghum landraces

Population structure in the Senegalese landraces followed the north–south precipitation gradient. These regional-scale patterns

FIGURE 6 Genome–environment associations for precipitation. (a) SNP associations for "precipitation of the driest quarter" using the generalized-linear model (GLM). The red dots represent SNPs identified from the multilocus mixed-model (MLMM). Linkage disequilibrium displayed as heat map of coefficient of correlation r^2 in a 50 kb region around SNPs S2_60708848 (b) and S6_691400 (c) that colocalize with Stg3a and Ma6 loci in (a), respectively. Red asterisks on each heat map represent these SNPs and blue asterisks indicate the SNPs within Ma6. The color scale indicates the significance of r^2 values with black color indicating high r^2 values. Allelic map distribution at SNPs S2_60708848 (d) and S6_691400 (e) associated with precipitation of the driest quarter. The shape of the points indicates the botanical race of the accession and the color indicates the allele at the SNP with H being the heterozygous alleles
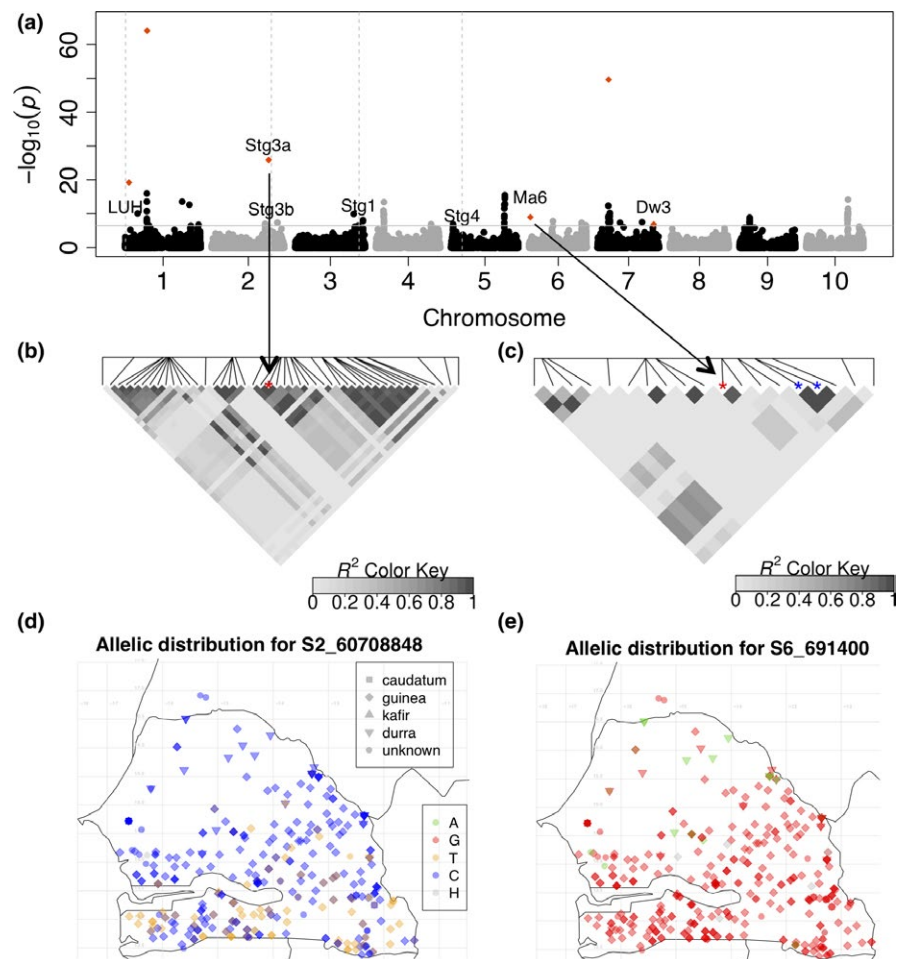
are in line with global patterns, where population structure is associated with precipitation-based agroclimatic zones (Lasky et al., 2015). Within Senegalese sorghums, guinea and durra clustered distinctly, consistent with global patterns of genetic differentiation (Morris et al., 2013; Sagnard et al., 2011). The relatively high proportion of variation explained by climate collinear with space suggests a role of clinal adaptation shaping variation, similar to recent findings in Nigerian and global sorghum germplasm (Lasky et al., 2015; Olatoye, Hu, Maina, & Morris, 2018). However, two guinea groups, from the center and south, clustered distinctly (Figure 1b and Figure S2d), suggesting possibly a specific genomic adaptation to the Soudano–Sahelian and the Soudanian agroclimatic, respectively.

The average pairwise nucleotide diversity, observed heterozygosity (data not shown), and the spatial and hierarchical genetic structure observed within guinea group (Figure 2a; File S1) is consistent with guinea being the most genetically diverse sorghum type (Deu et al., 1994; Folkertsma et al., 2005; Morris et al., 2013). Although the number of inferred subpopulations may not always correspond to the number of biological genetic groups (François & Durand, 2010; Meirmans, 2015), the spatial genetic co-ancestry structure analysis suggests the presence of untapped genetic diversity in the subpopulations in eastern Senegal (Figure 2a). The high estimated admixture coefficients among putative guinea subpopulations (File S1) could be due to gene flow among subpopulations or an effect of limited isolation-by-distance. The limited isolation-by-distance may occur because the geographic origin of the accessions is not broad (e.g., Senegal is not large) and there is any major geographic barrier that may create isolated subpopulations. There was little evidence of admixture between guinea and durra types, consistent with phenotype studies that rarely identify guinea–durra intermediates (Harlan & De Wet, 1972). Evidence of gene flow was mostly from guinea to durra (e.g., red subgroup at $K = 7$, File S1) and rarely from durra to guinea. The lower abundance of durra in this region may explain the limited gene flow between guinea and durra sorghums.

Ethnicity of farmers has shaped genetic structure in several staple crops including maize and pearl millet (Naino Jika et al., 2017; Orozco-Ramírez, Ross-Ibarra, Santacruz-Varela, & Brush, 2016). Distribution and diffusion of ethnic groups in Senegal including the Wolof, Serer, Diola, and Fulani (Toucouleur, Peul, Peul Foulbe, and Peul Firdou) could have affected gene flow among landraces. Indeed, the ethno-linguistic origin of the accessions contributed to the genetic variance of the Senegalese sorghum (Figure 2c). Seed exchange among farmers of the same ethnic group may have contributed in shaping this genetic structure (Barnaud, Trigueros, McKey, & Joly, 2008; Orozco-Ramírez et al., 2016; Pressoir & Berthaud, 2004). Codiffusion of sorghum with human migration has been demonstrated at Africa-wide scale (Westengen et al., 2014) and at a regional scale in Kenya (Labeyrie, Thomas, Muthamia, & Leclerc, 2016). Durra sorghum in Senegal are grown mainly by the Fulani ethnic group, so the clustering of Senegalese durra with Ethiopian durra (Figure S2d) and low F_{ST} (0.052) suggest that durra sorghums moved with Fulani people from northeast Africa (Scheinfeldt, Soi, & Tishkoff, 2010).

4.2 | Genetic basis of Sahelian and Soudanian adaptation

Nucleotide polymorphism patterns can provide insight into loci underlying adaptation (Vitti, Grossman, & Sabeti, 2013). The reduction of nucleotide polymorphism observed throughout the durra genome (Figure 3a) could be resulted from the bottlenecks during its diffusion along the Sahelian zone. Because Ethiopia is known as the center of origin of durra, we investigated whether the reduced polymorphism in durra was common to all African durra or specific to the Senegalese durra. The results suggest selective sweeps across durra genomes as durra populations diffused along the Sahel (Figure 3c and Figure S4). Interestingly, putative selective sweeps on pericentromeric regions of chromosome 2 were specific to Senegalese durra. By contrast to durra, there was little reduction of nucleotide polymorphism in the guinea genome and predominantly positive values of Tajima's D test (Figure 3a,b), reflecting population structure or possible balancing selection (Vitti et al., 2013). Simulations with demographic models could be used for more robust genome scans. Unfortunately, the underlying population parameters (e.g., effective population size, migration rates) are poorly described in sorghum.

Photoperiodic flowering is a key factor underlying adaptation in tropical crops (Kloosterman et al., 2013). The colocalization of photoperiodic flowering candidate genes with putative selective sweeps and phenotypic and environment associations (Figures 4 and 5a; Table S1; Data S4) are consistent with a role of conserved cereal flowering pathways in sorghum climate adaptation. The rare allele at the SNP near *Ma6/Ghd7* (6 kb away) was present in durra genotypes distributed in the drier areas of the Sahelian zone characterized by short growing seasons and low rainfall (<400 mm per year) (Figure 6e). This rare allele may be associated with early maturity and thus suggesting a role in drought escape such that plants can rapidly cover their maturity cycle and produce seeds before the end of growing season.

Other photoperiod flowering regulators identified in U.S. sorghum, *PhyC* (*Ma5*), *PhyB* (*Ma3*), *PhyA*, and *Ma1* (*SbPRR37*), colocalized with phenotype-associated SNPs (Data S4) and putative selective sweeps in the Senegalese sorghum (Childs et al., 1997; Rooney & Aydin, 1999). Signatures of selection near *Ma3* in durra (π_{durra} and CLR; Figure 3a and Figure 4a) are consistent with signatures of positive selection in *Ma3* observed in global sorghum (Wang et al., 2015). The florigens *SbCN12* and *SbCN15* (ortholog of rice florigen *Hd3a*) found near photoperiod-associated SNPs are photoperiod-regulated activators of floral induction in U.S. sorghum (Murphy et al., 2011). Putative photoperiodic flowering regulators *SbCRY1* and *SbGI*, colocalizing with selective sweeps in durra and photoperiod-associated SNPs, were previously associated with photoperiodic flowering in regional West-Central African germplasm (Bhosale et al., 2012). Several of the above genes were associated with flowering time adaptation in maize landraces, including *PhyB*, *PhyC*, *PRR37*, and *ZFL1/2* (Romero Navarro et al., 2017).

Panicle compactness in sorghum is a function of the number and length of inflorescence branches and the number of aborted

spikelets (Brown et al., 2006). Several candidate genes from a previous GWAS of inflorescence branch length in global sorghum (Morris et al., 2013) colocalized with GWAS signals for panicle compactness and/or CLR outliers in the current study (*SP1*, *CRCK3/THE1*, *TCP24*, and *DFL2*). The minor alleles in/near *SP1* (S1_55302939, S1_55305415) observed in durra accessions and some guinea accessions (Figure S7c,d) suggests a rare variant in *SP1* could contribute to shorter inflorescence branches in some Senegalese sorghum.

The colocalization of selective sweeps and GEA (Data S4) with stay-green drought tolerance loci (Borrell et al., 2014) suggests a broader role for stay-green loci in Sahelian adaptation. A selective sweep and associated SNPs colocalized with the stay-green locus *Stg1/SbPIN4* in guinea sorghums, suggesting that this region may confer adaptation of some guinea accessions to the dry areas of Senegal. The rare allele of SNP S3_57321183, which colocalized with *SbPIN2*, was found in a few guinea sorghums (Figure S7b). One possibility is that severe droughts starting in the 1970s (Gautier, Denis, & Locatelli, 2016; Mbow, Mertz, Diouf, Rasmussen, & Reenberg, 2008) have favored the introgression of stay-green drought tolerance alleles into some guinea landraces. Genome scans comparing older landrace collections with recent collections may shed more light on whether more recent selection (e.g. 1970s–2000s) has occurred, as demonstrated in Sahelian pearl millet (Vigouroux et al., 2011).

4.3 | Prospects for genomic dissection and improvement of climate adaptation

Improving adaptation of staple crops to the Sahelian and Soudanian zones is critical for smallholder farmers and a major challenge for African plant breeders. Despite advances in genotyping platforms, genomic tools for crop adaptation in sub-Saharan countries remain lacking. This study generated substantial genomic resources (213,916 SNPs among which 145,235 SNPs have MAF > 0.05) representing high-quality markers useful for the genomic dissection of adaptive and complex traits. High rates of SNPs with low frequency minor alleles (about 60% of the data had MAF < 0.05) were detected. One possible explanation may be related to the fact that these accessions are mostly landraces grown in their center of origin; thus high number of rare polymorphisms might be segregating at intermediate frequency in the germplasm. In the USDA-NPGS Ethiopian sorghum collection, similar patterns of MAF were found where 60% of detected SNPs had MAF < 0.05 (Cuevas, Rosa-Valentin, Hayes, Rooney, & Hoffmann, 2017). Overall, the Senegalese sorghum landraces represent a useful genetic resource, harboring useful variation for maturity and inflorescence morphology, as well as resistant sources to grain mold and anthracnose (Cuevas, Prom, & Rosa-Valentin, 2018).

The moderate decay of LD observed within the germplasm (Figure S2b) is consistent with the predominance of inbreeding in sorghum (Hamblin et al., 2005). Studies in sorghum have found a comparable LD pattern, decaying to its background level at ~150 kb (Mace et al., 2013; Morris et al., 2013). The population structure of

the Senegalese sorghum landraces would be expected to increase spurious association and reduce the power of GWAS (Brachi, Morris, & Borevitz, 2011). Indeed, the number of associations for photoperiod sensitivity was reduced when applying the regional mapping approach excluding durra accessions, presumably due to fewer spurious associations. Future studies with West African multi-parent mapping populations could breakup confounding LD and improve power to detect climate-adaptive loci (Bouchet et al., 2017; McMullen et al., 2009).

The stay-green loci may be useful to improve for drought adaptation in the Sahel via marker-assisted selection. Circadian clock-related genes influence crop yield under abiotic stress (Bendix, Marshall, & Harmon, 2015) and photoperiodic flowering loci identified may contribute to early maturity and drought escape in the Sahel. Taken together, our findings suggest a complex oligogenic basis of adaptation to Sahelian versus Soudanian climate, underpinned by variation in conserved floral regulatory pathways and variation in other pathways that are more poorly understood. Whole-genome resequencing of African crop diversity for GWAS and genome scans could facilitate identification of causal variants in the molecular pathways that underlie climate adaptation.

CONFLICT OF INTEREST

None declared.

AUTHOR CONTRIBUTIONS

G.M., D.F., and N.C. conceived and managed the study. F.M., J.F, and Z.H. generated the SNP data set. J.F. analyzed the data. J.F. and G.M. wrote the manuscript.

DATA ACCESSIBILITY

The raw sequencing data generated in this study are available in National Center for Biotechnology Information under the BioProject accession number PRJNA433571. The SNP data set is available at Dryad Data Repository under accession https://doi.org/10.5061/dryad.32f5395.

ORCID

Geoffrey P. Morris iD https://orcid.org/0000-0002-3067-3359

REFERENCES

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*, 1655–1664. https://doi.org/10.1101/gr.094052.109

Barnaud, A., Trigueros, G., McKey, D., & Joly, H. I. (2008). High out-crossing rates in fields with mixed sorghum landraces: How are landraces maintained? *Heredity*, *101*, 445. https://doi.org/10.1038/hdy.2008.77

Bendix, C., Marshall, C. M., & Harmon, F. G. (2015). Circadian clock genes universally control key agricultural traits. *Molecular Plant*, *8*, 1135–1152. https://doi.org/10.1016/j.molp.2015.03.003

Bhosale, S. U., Stich, B., Rattunde, H. F. W., Weltzien, E., Haussmann, B. I., Hash, C. T., … Parzies, H. K. (2012). Association analysis of photoperiodic flowering time genes in west and central African sorghum [*Sorghum bicolor* (L.) Moench]. *BMC Plant Biology*, *12*, 32. https://doi.org/10.1186/1471-2229-12-32

Blum, A. (2014). Genomics for drought resistance – Getting down to earth. *Functional Plant Biology*, *41*, https://doi.org/10.1071/FP14018

Borrell, A. K., van Oosterom, E. J., Mullet, J. E., George-Jaeggli, B., Jordan, D. R., Klein, P. E., & Hammer, G. L. (2014). Stay-green alleles individually enhance grain yield in sorghum under drought by modifying canopy development and water uptake patterns. *New Phytologist*, *203*, 817–830. https://doi.org/10.1111/nph.12869

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

Bouchet, S., Pot, D., Deu, M., Rami, J.-F., Billot, C., Perrier, X., … Glaszmann, J.-C. (2012). Genetic structure, linkage disequilibrium and signature of selection in sorghum: Lessons from physically anchored DArT markers. *PLoS ONE*, *7*, e33470. https://doi.org/10.1371/journal.pone.0033470

Brachi, B., Morris, G. P., & Borevitz, J. O. (2011). Genome-wide association studies in plants: The missing heritability is in the field. *Genome Biology*, *12*, 232. https://doi.org/10.1186/gb-2011-12-10-232

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*, 2633–2635. https://doi.org/10.1093/bioinformatics/btm308

Browning, B. L., & Browning, S. R. (2016). Genotype imputation with millions of reference samples. *Am J Hum Genet*, *98*(1), 116–126. https://doi.org/10.1016/j.ajhg.2015.11.020

Brown, P. J., Klein, P. E., Bortiri, E., Acharya, C. B., Rooney, W. L., & Kresovich, S. (2006). Inheritance of inflorescence architecture in sorghum. *Theoretical and Applied Genetics*, *113*, 931–942. https://doi.org/10.1007/s00122-006-0352-9

Caicedo, A. L., Williamson, S. H., Hernandez, R. D., Boyko, A., Fledel-Alon, A., York, T. L., … Purugganan, M. D. (2007). Genome-wide patterns of nucleotide polymorphism in domesticated rice. *PLOS Genetics*, *3*, e163. https://doi.org/10.1371/journal.pgen.0030163

Camus-Kulandaivelu, L., Veyrieras, J.-B., Madur, D., Combes, V., Fourmann, M., Barraud, S., … Charcosset, A. (2006). Maize adaptation to temperate climate: Relationship between population structure and polymorphism in the Dwarf8 gene. *Genetics*, *172*, 2449–2463. https://doi.org/10.1534/genetics.105.048603

Cavanagh, C., Morell, M., Mackay, I., & Powell, W. (2008). From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. *Current Opinion in Plant Biology*, *11*(2), 215–221. https://doi.org/10.1016/j.pbi.2008.01.002.

Caye, K., Deist, T. M., Martins, H., Michel, O., & François, O. (2016). TESS3: Fast inference of spatial population structure and genome scans for selection. *Molecular Ecology Resources*, *16*, 540–548. https://doi.org/10.1111/1755-0998.12471

Childs, K. L., Miller, F. R., Cordonnier-Pratt, M. M., Pratt, L. H., Morgan, P. W., & Mullet, J. E. (1997). The sorghum photoperiod sensitivity gene, Ma3, encodes a phytochrome B. *Plant Physiology*, *113*, 611–619.

Clément, J.-C., & Houdiard, P. (1977). Prospection des Mils pénicillaires et Sorghos en Afrique de l'Ouest: Campagne 1976: Nigeria-Sénégal.

Cuevas, H. E., Prom, L. K., & Rosa-Valentin, G. (2018). Population structure of the NPGS Senegalese sorghum collection and its evaluation to identify new disease resistant genes. *PLoS ONE*, *13*, e0191877. https://doi.org/10.1371/journal.pone.0191877

Cuevas, H. E., Rosa-Valentin, G., Hayes, C. M., Rooney, W. L., & Hoffmann, 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. https://doi.org/10.1186/s12864-016-3475-7

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., … Group, 1000 Genomes Project Analysis (2011). The variant call format and VCFtools. *Bioinformatics*, *27*, 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

de Meeûs, T., & Goudet, J. (2007). A step-by-step tutorial to use HierFstat to analyse populations hierarchically structured at multiple levels. *Infection, Genetics and Evolution*, *7*, 731–735. https://doi.org/10.1016/j.meegid.2007.07.005

Deu, M., Gonzalez-de-Leon, D., Glaszmann, J.-C., Degremont, I., Chantereau, J., Lanaud, C., & Hamon, P. (1994). RFLP diversity in cultivated sorghum in relation to racial differentiation. *Theoretical and Applied Genetics*, *88*, 838–844. https://doi.org/10.1007/BF01253994

Doggett, H. (1988). *Sorghum*. Essex: Longman Scientific & Technical.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, *6*, e19379. https://doi.org/10.1371/journal.pone.0019379

Fang, L., Wang, Q., Hu, Y., Jia, Y., Chen, J., Liu, B., … Zhang, T. (2017). Genomic analyses in cotton identify signatures of selection and loci associated with fiber quality and yield traits. *Nature Genetics*, *49*, 1089–1098. https://doi.org/10.1038/ng.3887

Folkertsma, R. T., Rattunde, H. F. W., Chandra, S., Raju, G. S., & Hash, C. T. (2005). The pattern of genetic diversity of Guinea-race *Sorghum bicolor* (L.) Moench landraces as revealed with SSR markers. *Theoretical and Applied Genetics.*, *111*, 399–409. https://doi.org/10.1007/s00122-005-1949-0

Fournier-Level, A., Korte, A., Cooper, M. D., Nordborg, M., Schmitt, J., & Wilczek, A. M. (2011). A map of local adaptation in *Arabidopsis thaliana*. *Science*, *334*, 86–89. https://doi.org/10.1126/science.1209271

François, O., & Durand, E. (2010). Spatially explicit Bayesian clustering models in population genetics. *Molecular Ecology Resources*, *10*, 773–784. https://doi.org/10.1111/j.1755-0998.2010.02868.x

Franks, S. J., Sim, S., & Weis, A. E. (2007). Rapid evolution of flowering time by an annual plant in response to a climate fluctuation. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 1278–1282. https://doi.org/10.1073/pnas.0608379104

Gautier, D., Denis, D., & Locatelli, B. (2016). Impacts of drought and responses of rural populations in West Africa: A systematic review. *Wiley Interdisciplinary Reviews: Climate Change*, *7*, 666–681. https://doi.org/10.1002/wcc.411

Glaubitz, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire, R. J., Sun, Q., & Buckler, E. S. (2014). TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. *PLoS ONE*, *9*, e90346. https://doi.org/10.1371/journal.pone.0090346

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., ... Rokhsar, D. S. (2012). Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Research*, *40*, D1178–D1186. https://doi.org/10.1093/nar/gkr944

Hamblin, M. T., Fernandez, M. G. S., 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*, 1247–1256. https://doi.org/10.1534/genetics.105.041566

Harlan, J. R., & De Wet, J. J. M. (1972). A simplified classification of cultivated sorghum1. *Crop Science*, *12*, 172–176. https://doi.org/10.2135/cropsci1972.0011183X001200020005x

Harris, K., Subudhi, P. K., Borrell, A., Jordan, D., Rosenow, D., Nguyen, H., ... Mullet, J. (2007). Sorghum stay-green QTL individually reduce post-flowering drought-induced leaf senescence. *Journal of Experimental Botany*, *58*, 327–338. https://doi.org/10.1093/jxb/erl225

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, *25*, 1965–1978. https://doi.org/10.1002/joc.1276

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

Kebede, H., Subudhi, P. K., Rosenow, D. T., & Nguyen, H. T. (2001). Quantitative trait loci influencing drought tolerance in grain sorghum (*Sorghum bicolor* L. Moench). *Theoretical and Applied Genetics*, *103*, 266–276. https://doi.org/10.1007/s001220100541

Kenney, A. M., McKay, J. K., Richards, J. H., & Juenger, T. E. (2014). Direct and indirect selection on flowering time, water-use efficiency (WUE, δ13C), and WUE plasticity to drought in *Arabidopsis thaliana*. *Ecology and Evolution*, *4*, 4505–4521. https://doi.org/10.1002/ece3.1270

Kloosterman, B., Abelenda, J. A., Gomez, M. D. M. C., Oortwijn, M., de Boer, J. M., Kowitwanich, K., ... Bachem, C. W. B. (2013). Naturally occurring allele diversity allows potato cultivation in northern latitudes. *Nature*, *495*, 246–250. https://doi.org/10.1038/nature11912

Labeyrie, V., Thomas, M., Muthamia, Z. K., & Leclerc, C. (2016). Seed exchange networks, ethnicity, and sorghum diversity. *Proceedings of the National Academy of Sciences of the United States of America*, *113*, 98–103. https://doi.org/10.1073/pnas.1513238112

Lasky, J. R., Des Marais, D. L., McKay, J. K., Richards, J. H., Juenger, T. E., & Keitt, T. H. (2012). Characterizing genomic variation of *Arabidopsis thaliana*: The roles of geography and climate. *Molecular Ecology*, *21*, 5512–5529. https://doi.org/10.1111/j.1365-294X.2012.05709.x

Lasky, J. R., Upadhyaya, H. D., Ramu, P., Deshpande, S., Hash, C. T., Bonnette, J., ... Morris, G. P. (2015). Genome-environment associations in sorghum landraces predict adaptive traits. *Science Advances*, *1*, e1400218. https://doi.org/10.1126/sciadv.1400218

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, L.-F., Li, Y.-L., Jia, Y., Caicedo, A. L., & Olsen, K. M. (2017). Signatures of adaptation in the weedy rice genome. *Nature Genetics*, *49*, 811–814. https://doi.org/10.1038/ng.3825

Lin, T., Zhu, G., Zhang, J., Xu, X., Yu, Q., Zheng, Z., ... Huang, S. (2014). Genomic analyses provide insights into the history of tomato breeding. *Nature Genetics*, *46*, 1220–1226. https://doi.org/10.1038/ng.3117

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

Mace, E. S., Tai, S., Gilding, E. K., Li, Y., Prentis, P. J., Bian, L., ... Wang, J. (2013). Whole-genome sequencing reveals untapped genetic potential in Africa's indigenous cereal crop sorghum. *Nature Communications*, *4*, 2320. https://doi.org/10.1038/ncomms3320

Mbow, C., Mertz, O., Diouf, A., Rasmussen, K., & Reenberg, A. (2008). The history of environmental change and adaptation in eastern Saloum–Senegal—Driving forces and perceptions. *Global and Planetary Change, Climate Change and Desertification*, *64*, 210–221. https://doi.org/10.1016/j.gloplacha.2008.09.008

McCormick, R. F., Truong, S. K., Sreedasyam, A., Jenkins, J., Shu, S., Sims, D., ... Mullet, J. E. (2018). The *Sorghum bicolor* reference genome: Improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. *The Plant Journal*, *93*, 338–354. https://doi.org/10.1111/tpj.13781

McMullen, M. D., Kresovich, S., Villeda, H. S., Bradbury, P., Li, H., Sun, Q., ... Buckler, E. S. (2009). Genetic properties of the maize nested association mapping population. *Science*, *325*, 737–740. https://doi.org/10.1126/science.1174320

Meirmans, P. G. (2015). Seven common mistakes in population genetics and how to avoid them. *Molecular Ecology*, *24*, 3223–3231. https://doi.org/10.1111/mec.13243

Morris, G. P., Ramu, P., Deshpande, S. P., Hash, C. T., Shah, T., Upadhyaya, H. D., ... Kresovich, S. (2013). Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 453–458. https://doi.org/10.1073/pnas.1215985110

Morton, J. F. (2007). The impact of climate change on smallholder and subsistence agriculture. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 19680–19685. https://doi.org/10.1073/pnas.0701855104

Mullet, J. E., Rooney, W. L., Klein, P. E., Morishige, D., Murphy, R., & Brady, J. A. (2010). Discovery and utilization of sorghum genes (ma5/ma6). US20100024065 A1.

Murphy, R. L., Klein, R. R., Morishige, D. T., Brady, J. A., Rooney, W. L., Miller, F. R., ... Mullet, J. E. (2011). Coincident light and clock regulation of pseudoresponse regulator protein 37 (PRR37) controls photoperiodic flowering in sorghum. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 16469–16474. https://doi.org/10.1073/pnas.1106212108

Naino Jika, A. K., Dussert, Y., Raimond, C., Garine, E., Luxereau, A., Takvorian, N., ... Robert, T. (2017). Unexpected pattern of pearl millet genetic diversity among ethno-linguistic groups in the Lake Chad Basin. *Heredity*, *118*, 491–502. https://doi.org/10.1038/hdy.2016.128

Nielsen, R., Williamson, S., Kim, Y., Hubisz, M. J., Clark, A. G., & Bustamante, C. (2005). Genomic scans for selective sweeps using SNP data. *Genome Research*, *15*, 1566–1575. https://doi.org/10.1101/gr.4252305

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., ...Wagner, H. (2017). vegan. Community Ecology Package.

Olatoye, M. O., Hu, Z., Maina, F., & Morris, G. P. (2018). Genomic signatures of adaptation to a precipitation gradient in Nigerian sorghum. *G3: Genes, Genomes, Genetics*, *8*(10), 3269–3281. https://doi.org/10.1534/g3.118.200551

Olsen, K. M., Caicedo, A. L., Polato, N., McClung, A., McCouch, S., & Purugganan, M. D. (2006). Selection under domestication: Evidence for a sweep in the rice waxy genomic region. *Genetics*, *173*, 975–983. https://doi.org/10.1534/genetics.106.056473

Orozco-Ramírez, Q., Ross-Ibarra, J., Santacruz-Varela, A., & Brush, S. (2016). Maize diversity associated with social origin and environmental variation in Southern Mexico. *Heredity*, *116*, 477–484. https://doi.org/10.1038/hdy.2016.10

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*, *20*, 289–290.

Paterson, A. H., Bowers, J. E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., ... Rokhsar, D. S. (2009). The *Sorghum bicolor* genome and the diversification of grasses. *Nature*, *457*, 551–556. https://doi.org/10.1038/nature07723

Pavlidis, P., Živković, D., Stamatakis, A., & Alachiotis, N. (2013). SweeD: Likelihood-based detection of selective sweeps in thousands of

genomes. _Molecular Biology and Evolution_, 30, 2224–2234. https://doi.org/10.1093/molbev/mst112

Pressoir, G., & Berthaud, J. (2004). Patterns of population structure in maize landraces from the Central Valleys of Oaxaca in Mexico. _Heredity_, 92, 88. https://doi.org/10.1038/sj.hdy.6800387

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., … Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. _The American Journal of Human Genetics_, 81, 559–575. https://doi.org/10.1086/519795

R Core Team. (2016). _A language and environment for statistical computing._ Vienna, Austria: R Foundation for Statistical Computing, 2015.

Rellstab, C., Gugerli, F., Eckert, A. J., Hancock, A. M., & Holderegger, R. (2015). A practical guide to environmental association analysis in landscape genomics. _Molecular Ecology_, 24, 4348–4370. https://doi.org/10.1111/mec.13322

Remington, D. L., Thornsberry, J. M., Matsuoka, Y., Wilson, L. M., Whitt, S. R., Doebley, J., … Buckler, E. S. (2001). Structure of linkage disequilibrium and phenotypic associations in the maize genome. _Proceedings of the National Academy of Sciences of the United States of America_, 98, 11479–11484. https://doi.org/10.1073/pnas.201394398

Romero Navarro, J. A., Willcox, M., Burgueño, J., Romay, C., Swarts, K., Trachsel, S., … Buckler, E. S. (2017). A study of allelic diversity underlying flowering-time adaptation in maize landraces. _Nature Genetics_, 49, 476–480. https://doi.org/10.1038/ng.3784

Rooney, W. L., & Aydin, S. (1999). Genetic Control of a Photoperiod-Sensitive Response in _Sorghum bicolor_ (L.). Moench. _Crop Science_, 39, 397. https://doi.org/10.2135/cropsci1999.0011183X003900020000 16x

Sagnard, F., Deu, M., Dembélé, D., Leblois, R., Touré, L., Diakité, M., … Traoré, P. C. S. (2011). Genetic diversity, structure, gene flow and evolutionary relationships within the _Sorghum bicolor_ wild-weedy-crop complex in a western African region. _Theoretical and Applied Genetics_, 123, 1231–1246. https://doi.org/10.1007/s00122-011-1662-0

Sanon, M., Hoogenboom, G., Traoré, S. B., Sarr, B., Garcia, A. G. Y., Somé, L., & Roncoli, C. (2014). Photoperiod sensitivity of local millet and sorghum varieties in West Africa. _NJAS - Wageningen Journal of Life Sciences_, 68, 29–39. https://doi.org/10.1016/j.njas.2013.11.004

Scheinfeldt, L. B., Soi, S., & Tishkoff, S. A. (2010). Working toward a synthesis of archaeological, linguistic, and genetic data for inferring African population history. _Proceedings of the National Academy of Sciences of the United States of America_, 107, 8931–8938. https://doi.org/10.1073/pnas.1002563107

Segura, V., Vilhjálmsson, B. J., Platt, A., Korte, A., Seren, Ü., Long, Q., & Nordborg, M. (2012). An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. _Nature Genetics_, 44, 825–830. https://doi.org/10.1038/ng.2314

Shin, J.-H., Blay, S., McNeney, B., & Graham, J. (2006). LDheatmap: An R function for graphical display of pairwise linkage disequilibria between single nucleotide polymorphisms. _Journal of Statistical Software_, 16, 1–9. https://doi.org/10.18637/jss.v016.c03

Siepielski, A. M., Morrissey, M. B., Buoro, M., Carlson, S. M., Caruso, C. M., Clegg, S. M., … MacColl, A. D. C. (2017). Precipitation drives global variation in natural selection. _Science_, 355, 959–962. https://doi.org/10.1126/science.aag2773

Slatkin, M. (2008). Linkage disequilibrium—Understanding the evolutionary past and mapping the medical future. _Nature Reviews Genetics_, 9, 477–485. https://doi.org/10.1038/nrg2361

Swarts, K., Gutaker, R. M., Benz, B., Blake, M., Bukowski, R., Holland, J., … Burbano, H. A. (2017). Genomic estimation of complex traits reveals ancient maize adaptation to temperate North America. _Science_, 357, 512–515. https://doi.org/10.1126/science.aam9425

Tuinstra, M. R., Grote, E. M., Goldsbrough, P. B., & Ejeta, G. (1997). Genetic analysis of post-flowering drought tolerance and components of grain development in _Sorghum bicolor_ (L.) Moench. _Molecular Breeding_, 3, 439–448. https://doi.org/10.1023/A:1009673126345

Vigouroux, Y., Mariac, C., Mita, S. D., Pham, J.-L., Gérard, B., Kapran, I., … Bezançon, G. (2011). Selection for earlier flowering crop associated with climatic variations in the Sahel. _PLoS ONE_, 6, e19563. https://doi.org/10.1371/journal.pone.0019563

Vitti, J. J., Grossman, S. R., & Sabeti, P. C. (2013). Detecting natural selection in genomic data. _Annual Review of Genetics_, 47, 97–120. https://doi.org/10.1146/annurev-genet-111212-133526

Wang, Y., Tan, L., Fu, Y., Zhu, Z., Liu, F., Sun, C., & Cai, H. (2015). Molecular evolution of the sorghum maturity gene Ma3. _PLoS ONE_, 10, e0124435. https://doi.org/10.1371/journal.pone.0124435

Wang, Y.-H., Upadhyaya, H. D., Burrell, A. M., Sahraeian, S. M. E., Klein, R. R., & Klein, P. E. (2013). Genetic structure and linkage disequilibrium in a diverse, representative collection of the C4 model plant, _Sorghum bicolor_. _G3: Genes, Genomes, Genetics_, 3, 783–793. https://doi.org/10.1534/g3.112.004861

Westengen, O. T., Okongo, M. A., Onek, L., Berg, T., Upadhyaya, H., Birkeland, S., … Brysting, A. K. (2014). Ethnolinguistic structuring of sorghum genetic diversity in Africa and the role of local seed systems. _Proceedings of the National Academy of Sciences of the United States of America_, 111, 14100–14105. https://doi.org/10.1073/pnas.1401646111

Xu, K., Xu, X., Fukao, T., Canlas, P., Maghirang-Rodriguez, R., Heuer, S., … Mackill, D. J. (2006). _Sub1A_ is an ethylene-response-factor-like gene that confers submergence tolerance to rice. _Nature_, 442, 705–708. https://doi.org/10.1038/nature04920

Yoder, J. B., Stanton-Geddes, J., Zhou, P., Briskine, R., Young, N. D., & Tiffin, P. (2014). Genomic signature of adaptation to climate in _Medicago truncatula_. _Genetics_, 196, 1263–1275. https://doi.org/10.1534/genetics.113.159319

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

SUPPORTING INFORMATION

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