

Population genomics of sorghum (*Sorghum bicolor*) across diverse agroclimatic zones of Niger¹

Fanna Maina, Sophie Bouchet, Sandeep R. Marla, Zhenbin Hu, Jianan Wang, Aissata Mamadou, Magagi Abdou, Abdoul-Aziz Saïdou, and Geoffrey P. Morris

Abstract: Improving adaptation of staple crops in developing countries is important to ensure food security. In the West African country of Niger, the staple crop sorghum (*Sorghum bicolor*) is cultivated across diverse agroclimatic zones, but the genetic basis of local adaptation has not been described. The objectives of this study were to characterize the genomic diversity of sorghum from Niger and to identify genomic regions conferring local adaptation to agroclimatic zones and farmer preferences. We analyzed 516 Nigerien accessions for which local variety name, botanical race, and geographic origin were known. We discovered 144 299 single nucleotide polymorphisms (SNPs) using genotyping-by-sequencing (GBS). We performed discriminant analysis of principal components (DAPC), which identified six genetic groups, and performed a genome scan for loci with high discriminant loadings. The highest discriminant coefficients were on chromosome 9, near the putative ortholog of maize flowering time adaptation gene *Vgt1*. Next, we characterized differentiation among local varieties and used a genome scan of pairwise F_{ST} values to identify SNPs associated with specific local varieties. Comparison of varieties named for light- versus dark-grain identified differentiation near *Tannin1*, the major gene responsible for grain tannins. These findings could facilitate genomics-assisted breeding of locally adapted and farmer-preferred sorghum varieties for Niger.

Key words: sorghum, agroclimatic zones, genomics, local adaptation, genomics-enabled breeding, sub Saharan Africa.

Résumé : L'amélioration de cultures dans les pays en développement est importante afin d'assurer sécurité alimentaire. En Afrique de l'Ouest, particulièrement au Niger, le sorgho (*Sorghum bicolor*) est cultivé dans différentes zones agroclimatiques. Cependant les bases génétiques de leur adaptation locale sont peu décrites. Cette étude a pour objectifs de caractériser la structure génétique du sorgho au Niger et d'identifier les régions génomiques associées à l'adaptation locale aux zones agroclimatiques et aux préférences des agriculteurs. L'analyse par le génotypage par séquençage (GBS) de 516 accessions du Niger, dont le nom local, la race botanique et l'origine sont connus, nous a permis d'identifier 144 299 polymorphismes nucléotidiques (SNPs). La méthode d'analyse discriminante des composantes principales a identifié six clusters génétiques. Le balayage génomique des coefficients de discrimination a montré des locus aux coefficients élevés au niveau du chromosome 9, colocalisés avec le gene *Vgt1* responsable de la variation de la date de floraison. Aussi, nous avons caractérisé la différenciation des variétés locales. Le balayage génomique de F_{ST} entre les variétés locales Mota (grains blancs) et Jenjari (grains sombres) a identifié des locus près du *Tannin1*, le gène responsable des tanins. Ces résultats permettront de faciliter la sélection assistée par la génomique de variétés de sorgho localement adaptées et préférées en Niger.

Mots-clés : sorgho, zones agroclimatiques, génomique, adaptation locale, sélection génomique, Afrique subsaharienne.

Introduction

Sorghum (*Sorghum bicolor* L. Moench) is a major staple crop that is adapted to multiple agroclimatic zones of the world (Smith and Frederiksen 2000). In smallholder production systems of Africa, sorghum is used as food, forage, and building material, and cultivated under constraints such as drought, insects, weeds, and bird predation (National Research Council 1996; Smith and Frederiksen 2000). In the West African country of Niger, sorghum is cultivated across agroclimatic zones defined by the precipitation gradient (Saharan, Sahelian, Sahelo-soudanian, and Soudanian) (Mohamed

et al. 2002; Deu et al. 2010). Four botanical races (durra, caudatum, guinea, and bicolor), classified based on the morphology of the inflorescence and grain, are found distributed across the agroclimatic zones of Niger (Deu et al. 2010). Durra, caudatum, and their intermediates are grown in the Sahelian and Sahelo-soudanian zones (400–600 mm) while guinea is predominantly grown in the Soudanian zone (800 mm). Sorghum is grown by smallholder farmers of several ethnic groups including Hausa (predominantly in central Niger), Zarma (predominantly in western Niger), and Kanuri (predominantly in eastern Niger).

Received 23 June 2017. Accepted 26 January 2018.

Corresponding Editor: Mark Ungerer.

F. Maina. Department of Agronomy, Kansas State University, Manhattan, KS 66506, USA; Institut National de la Recherche Agronomique du Niger (INRAN) Niamey, Niger.

S. Bouchet. Institut National de la Recherche Agronomique (INRA), l'Université Clermont Auvergne (UCA), 63000 Clermont-Ferrand, France.

S.R. Marla, Z. Hu, J. Wang, and G.P. Morris. Department of Agronomy, Kansas State University, Manhattan, KS 66506, USA.

A. Mamadou. Institut National de la Recherche Agronomique du Niger (INRAN) Niamey, Niger.

M. Abdou. La Sahélienne Des Semences HALAL, Maradi, Niger.

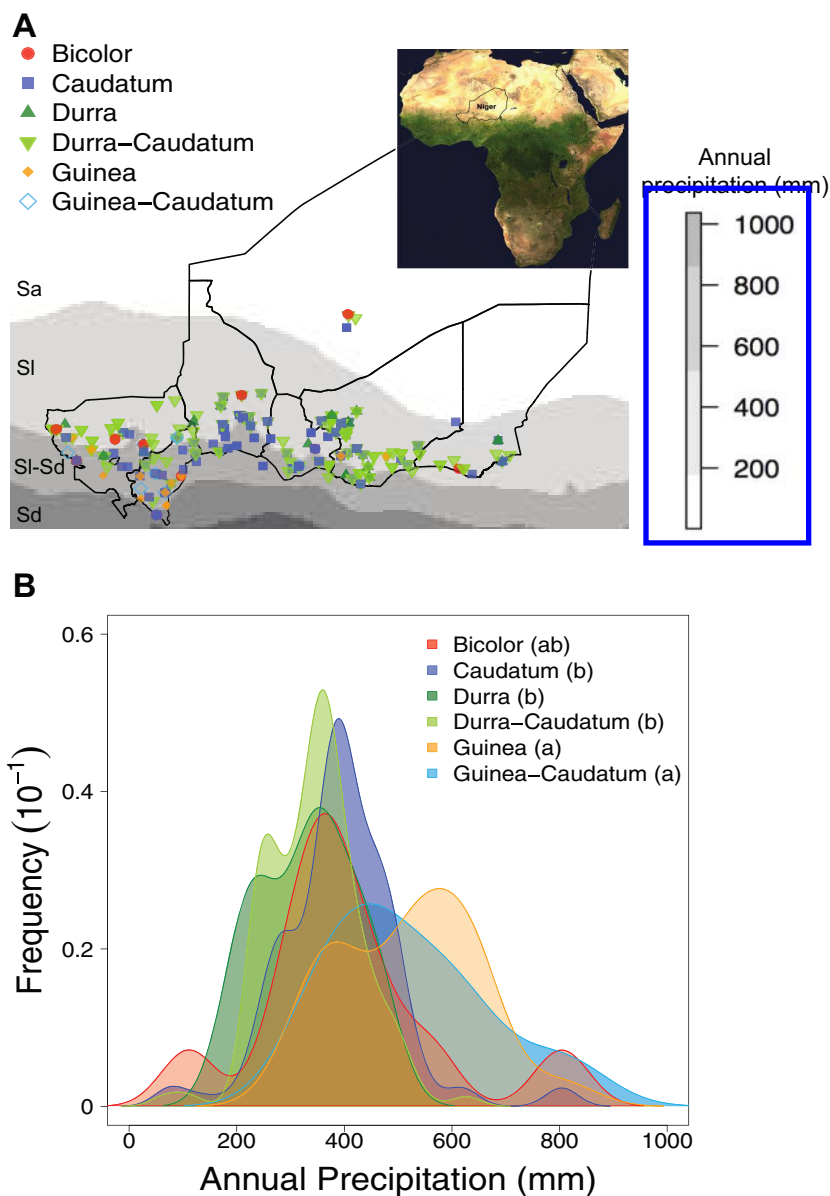
A.-A. Saïdou. Université Dan Dicko Dankoulodo de Maradi, Niger.

Corresponding author: Geoffrey P. Morris (email: gpmorris@k-state.edu).

¹This paper is part of a Special Issue entitled Ecological Genomics.

Copyright remains with the author(s) or their institution(s). Permission for reuse (free in most cases) can be obtained from [RightsLink](https://www.rightslink.com).

Fig. 1. (A) Map of Niger with precipitation gradient colored in grey. The points represent the georeferenced accessions and each color represents the botanical race of the accession. The precipitation gradient, decreasing from south to north, reflects the agroclimatic zones. The highest precipitation is in the southwestern part of the country. Agroclimatic zones key: Sa, Saharian; Sl, Sahelian; Sl-Sd, Sahelo-Soudanian; and Sd, Soudanian. (B) Density plot of the botanical races according to annual precipitation for each accession. Different lowercase letters (a, b) represent significant differences in annual precipitation (at the site of collection) among accessions from each botanical race.

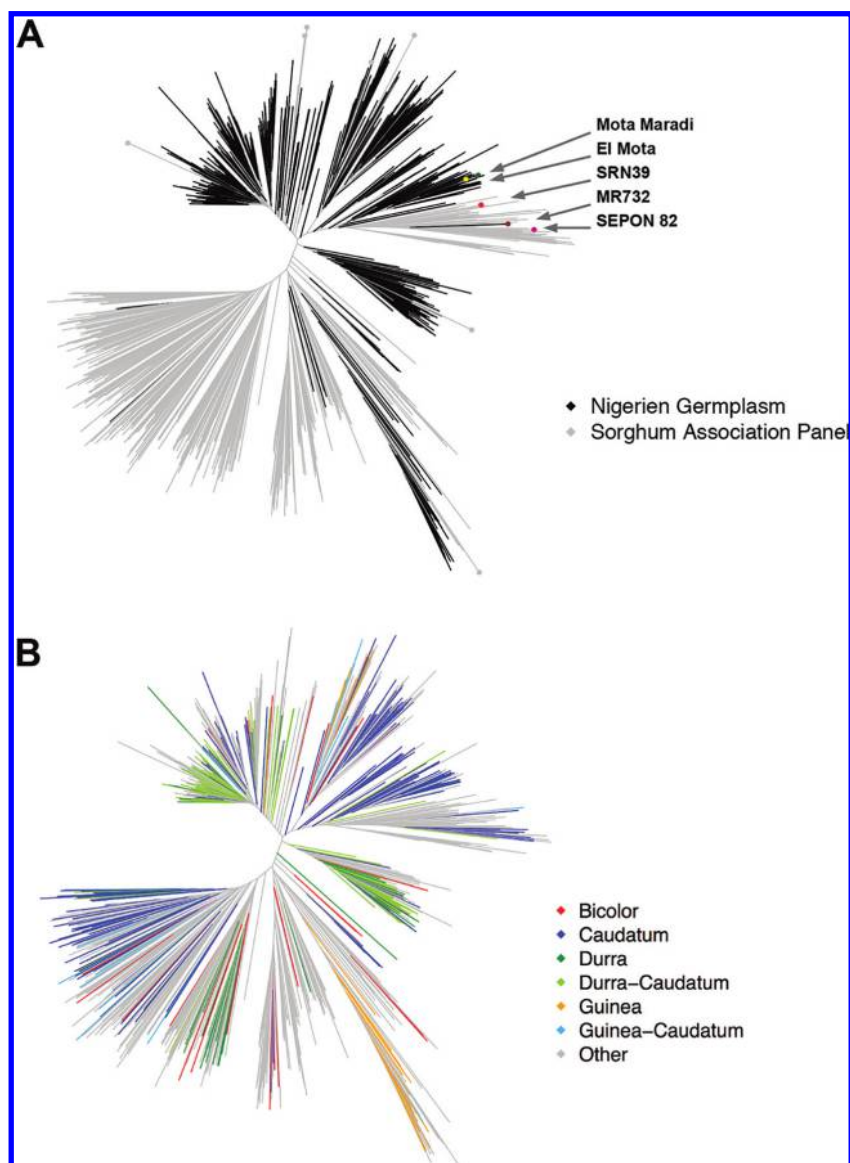


The sorghum breeding program at the National Institute of Agronomic Research of Niger (INRAN) has focused on improving local varieties by introgressing new traits conferring resistance to biotic (*Striga*) and abiotic (drought, soil fertility) stressors (Ejeta and Gressel 2007). However, the level of farmer's adoption is limited (5%–10%) due to the lack of varieties that combine improved yield and tolerance traits with locally preferred traits related to maturity and grain pigmentation (Bezançon et al. 2009; Walker and Alwang 2015). Marker-assisted selection (MAS) can accelerate the development of new varieties with improved yield and adaptation (Collard and Mackill 2008). Genotyping-by-sequencing (GBS) can accelerate the development of high-density markers for genetic diversity studies and MAS (Elshire et al. 2011). In sorghum, global and regional diversity have been studied using high-density GBS markers (Morris et al. 2013a; Leiser et al. 2014; Lasky et al. 2015). However, most studies of local diversity in African sorghum

were performed using low-density markers (Deu et al. 2010; Barro-Kondombo et al. 2010; Ng'uni et al. 2011). Strong population structure has been observed across agroclimatic zones in Niger (Deu et al. 2010). High-density GBS markers could help identify loci underlying local adaptation and farmer preference, and facilitate development of genomic tools for MAS.

Genome scans of diverse crop germplasm can identify loci under natural and human selection (Lasky et al. 2015; Romero Navarro et al. 2017). In Nigerien sorghum, we hypothesize that adaptation to agroclimatic zones has acted on loci controlling maturity and grain pigmentation. To characterize population structure and identify genomic regions associated with local adaptation and farmer preferences, we characterized 516 accessions of Nigerien sorghum germplasm at 144 299 SNPs. We performed the genome scans to identify genome regions associated with local adaptation and

Fig. 2. Comparison of the Nigerien and the global diversity (SAP) accessions. (A) Neighbor-joining tree of 872 accessions including both Nigerien and SAP accessions. The grey lines represent the SAP (356 accessions) and the black lines represent the Nigerien germplasm (516 accessions). The grey arrows represent the Nigerien accessions in SAP: Mota Maradi (PI656050), El Mota (PI656035), SRN39 (PI656027), MR732 (PI656051), and Sepon 82 (PI656024). The grey points represent the SAP accessions (PI656066, PI597972, PI597973, PI655981, PI533882, PI656077, PI656115) that are found within Nigerien clusters. (B) The same neighbor-joining tree of 872 accessions (including both Nigerien and SAP accessions) color-coded by botanical race.



farmer preferences that can be used in breeding for climate-resilient, locally preferred varieties.

Methods

Genetic materials

The collection is composed of 520 Nigerien accessions that were collected in 1976 (Borgel and Sequier 1977). This collection was obtained from United States National Plant Germplasm System (NPGS) along with passport data for each accession including local variety name, botanical race, latitude, and longitude coordinates. Precipitation maps were created using average annual precipitation from 1960–1990 obtained from WorldClim 1.4 using raster package in R (Hijmans et al. 2005; R Core Team 2013;

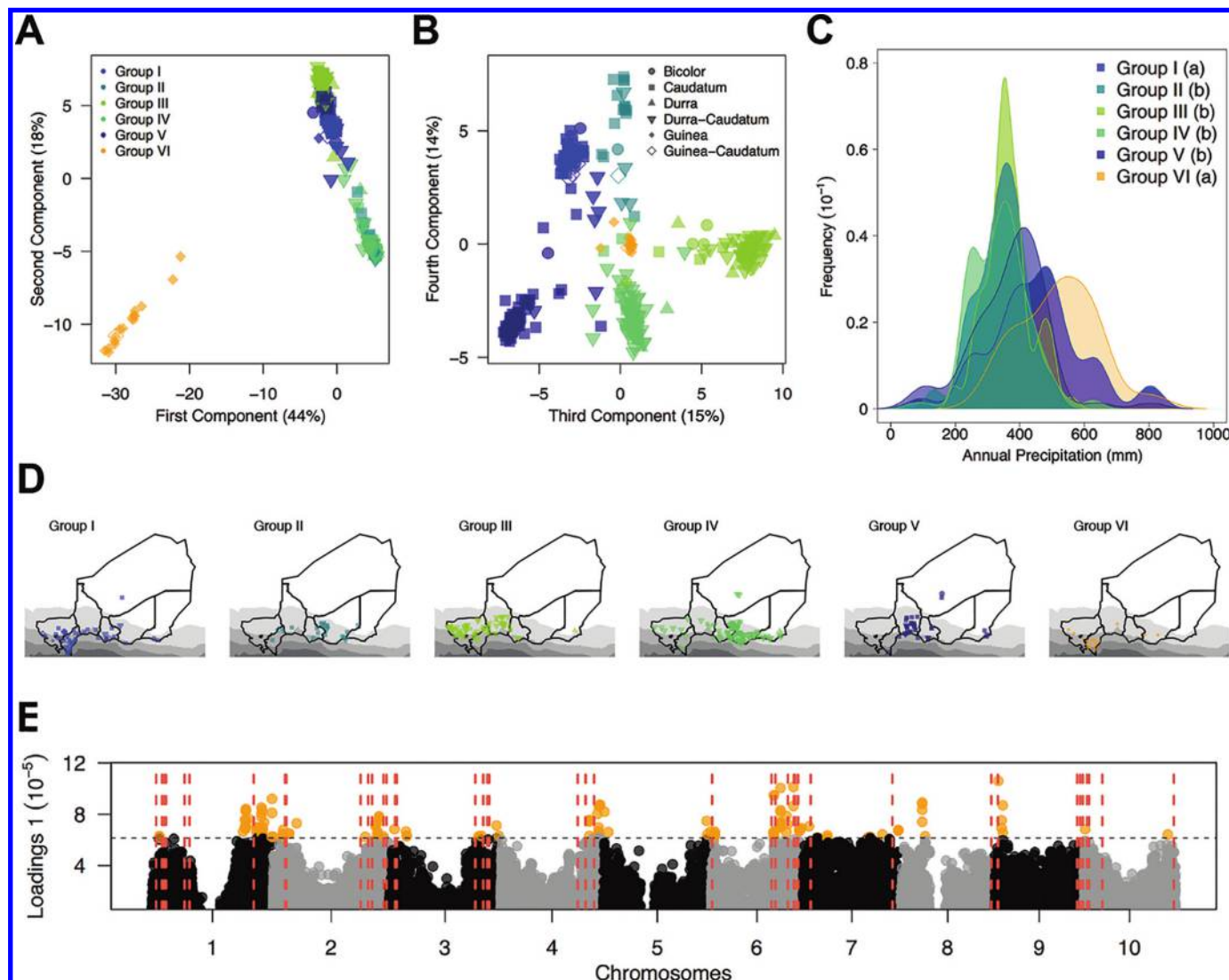
Hijmans 2016) (Fig. 1). Two weeks after planting in the greenhouse, 516 accessions successfully germinated (File S1²). Fresh leaf tissue (50 mg) was collected for each accession and transferred into a 96-deep well plate. Each plate contains 95 samples, with an empty well as the control. Tissue was lyophilized (Labconco Freeze Dryer) for two days and then grounded using a 96-well plate plant tissue grinder (Retsch Mixer Mill). Genomic DNA was extracted using the BioSprint 96 DNA Plant Kit (QIAGEN), quantified using Quant-iTTM PicoGreen® dsDNA Assay Kit, and normalized to 10 ng/μL.

Genotyping-by-sequencing (GBS)

GBS method (Elshire et al. 2011) was used for this study. DNA libraries were first digested using *ApeKI* restriction enzyme (NEB

²Supplementary data are available with the article through the journal Web site at <http://nrcresearchpress.com/doi/suppl/10.1139/gen-2017-0131>.

Fig. 3. Population structure of the Nigerien germplasm using discriminant analysis of principal components (DAPC). Scatter plot of the principal components, with first and second components (A) and third and fourth components (B) that identify six genetic groups. (C) Density plot of the genetic groups according to annual precipitation for each accession. Different lowercase letters (a, b) represent significant differences among groups. (D) Maps of Niger with the collection location of accessions for each genetic group. For A–C, botanical race for each accession is noted by shape and genetic group is noted by color. (E) Genome scan of the loadings from the first principal component. The x-axis represents the position of each SNP on the chromosomes and the y-axis the loadings values. The colored points represented highly differentiated genome regions (top 0.1% of SNPs). The dashed vertical lines represent a priori candidate genes.



R0643L) and ligated with the barcode and common adapters using T4 DNA ligase (NEB M0202L). Ligated libraries were pooled (96-plex libraries) before PCR amplification. Libraries were purified with QIAquick PCR purification kit and a Bioanalyzer (Agilent Technologies 2100) was used to determine the library size distribution. The 384-plex library was obtained after pooling four 96-plex libraries. Sequencing was performed using single end 100-cycle sequencing with Illumina HiSeq2500 (University of Kansas Medical Center). TASSEL3 GBS pipeline (Glaubitz et al. 2014) was used to call SNP genotypes. The tags were aligned with the sorghum reference genome version v3.0 (<https://phytozome.jgi.doe.gov>) using Burrows-Wheeler Aligner (Li and Durbin 2010), and SNPs were filtered at minor allele frequency (MAF) < 0.01. Missing genotypes were imputed with Beagle (Browning and Browning 2007) using default settings. For comparison with the Nigerien germplasm, published raw GBS sequence data for the global Sorghum Association Panel (SAP) ($n = 356$) were obtained (Morris et al. 2013a). We

then combined both raw sequences from fastq format and recalled the SNPs.

Genomic analysis

To determine genetic relatedness among accessions, neighbor-joining analysis was performed using ape package in R (Paradis et al. 2004; R Core Team 2013). Genetic groups were inferred using the discriminant analysis of principal component (DAPC) package Adegenet in R (Jombart and Ahmed 2011; R Core Team 2013) using the Bayesian Information Criterion (BIC), criterion = "goodfit", n.iter = 1e5, and n.start = 1000. To characterize SNP variation, we used VCFtools (Danecek et al. 2011). We estimated minor allele frequency, nucleotide diversity, Tajima's D , and F -statistics (Weir and Cockerham 1984) between botanical races, and local varieties.

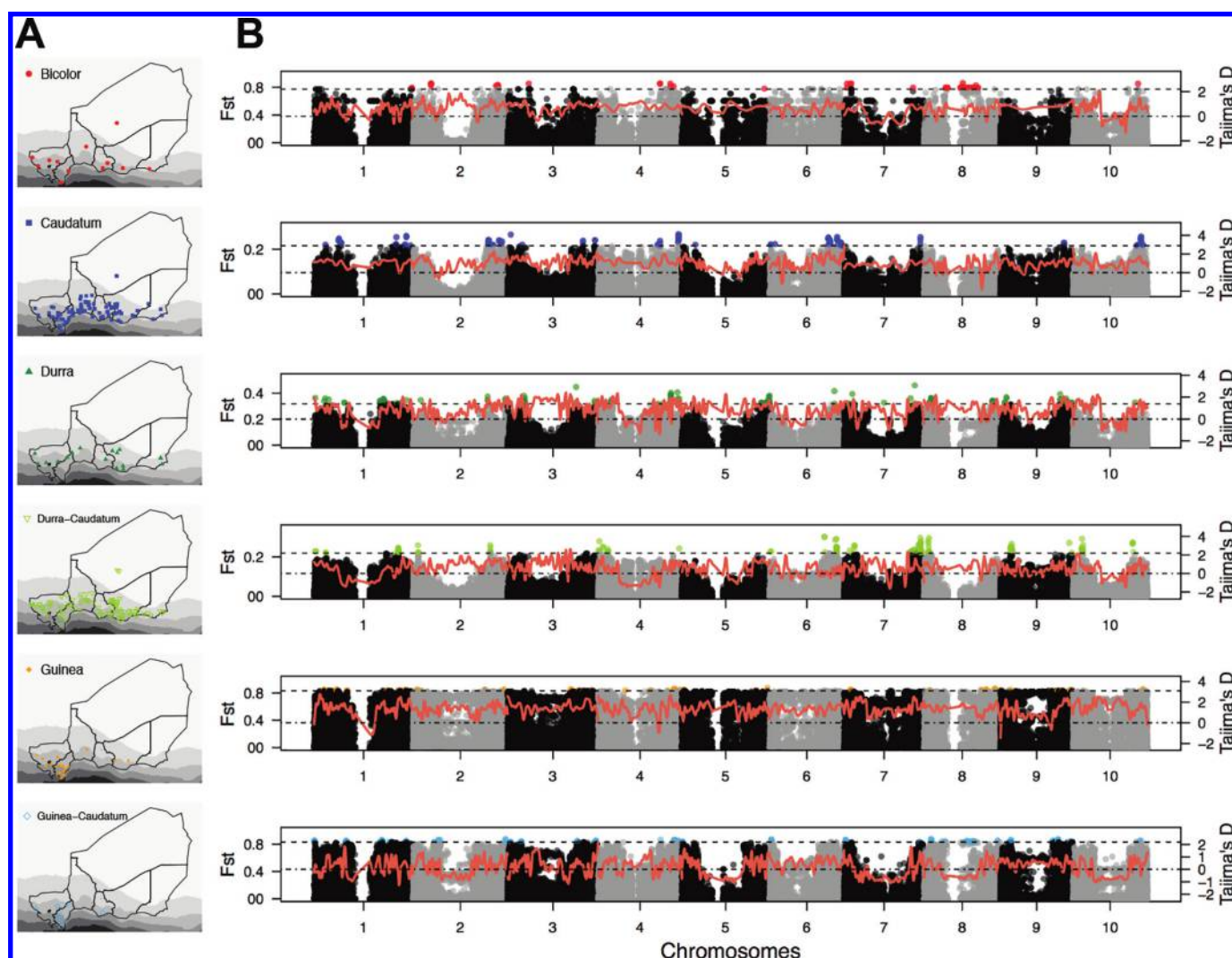
For comparison with genome-wide scans, we develop a list of a priori candidates genes for some traits (maturity, plant height, grain tannins, and stay-green) that have been implicated in agro-

Table 1. Summary table of the genetic groups inferred by discriminant analysis of principal components (DAPC).

Group	Number	Botanical race							Kernel color				Main local names				
		B	C	D	G	CB	CD	CG	B	R	Y	W	M	J	B	A	S
I	95	4	32	—	7	12	4	8	15	2	17	46	2	2	1	7	3
II	74	1	15	—	—	44	4	—	62	—	2	2	16	1	3	—	
III	87	4	12	9	—	5	37	—	36	2	9	26	1	7	—	27	
IV	158	2	18	11	—	33	74	—	79	3	10	46	14	—	25	2	
V	77	1	49	—	—	3	4	—	30	2	3	28	1	24	—	—	
VI	25	—	—	—	16	—	—	1	5	—	12	2	—	—	—	13	

Note: Botanical race key: B, bicolor; C, caudatum; D, durra; G, guinea; CB, caudatum–bicolor; CD, caudatum–durra; CG, caudatum–guinea. Kernel color key: B, brown; R, red; Y, yellow; W, white. Main local names key: M, Mota; J, Jenjari; B, Babadia; A, Amo Koire; S, Sokombe.

Fig. 4. Genome scan for differentiation among botanical races. (A) Map of Niger with the georeferenced accessions for each botanical race. (B) F_{ST} and Tajima's D genome scans for the botanical races. The x -axis represents the chromosomes and the first y -axis the F_{ST} values. Each point represents the SNP marker. The top 0.1% was used to set the threshold. The colors represent the SNPs above the threshold comparing one botanical race compared to the five others. F_{ST} was calculated using Weir and Cockerham's method. The second y -axis represents the Tajima's D average of 100 kb window size.



climatic adaptation in sorghum and other cereals (File S2²) (Wu et al. 2012; Borrell et al. 2014; Lasky et al. 2015; Romero Navarro et al. 2017). Among the candidate genes, we refer to those identified based on synteny as orthologs (Schnable and Freeling 2011), while we refer to those based on high pairwise similarity in Phytozome as putative orthologs (<https://phytozome.jgi.doe.gov>). To identify SNP markers associated with environmental variables, we performed genome-wide association studies (GWAS) using the GAPIT regular mixed linear model (MLM) method

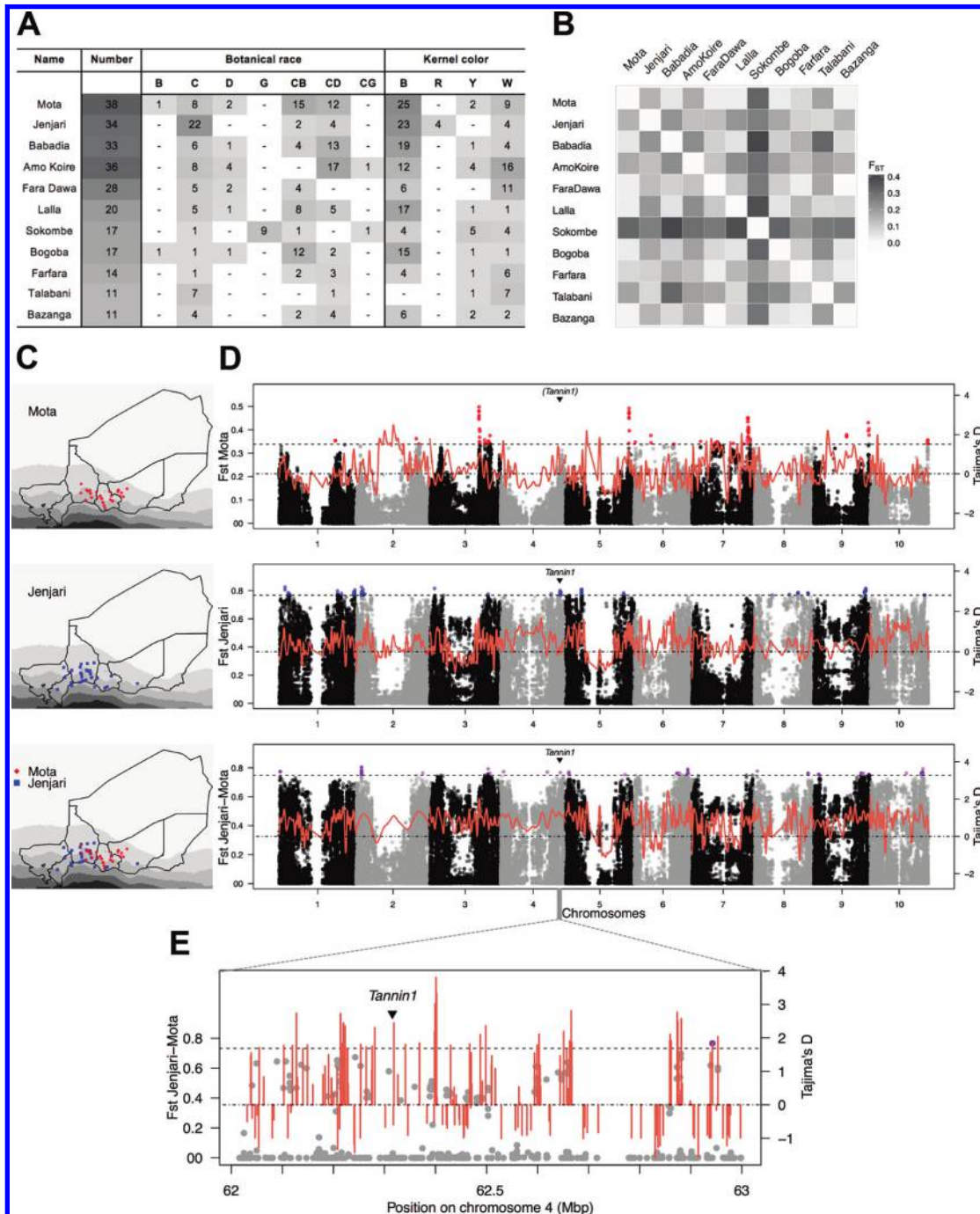
(Lipka et al. 2012). Annual precipitation and annual mean temperature for each accession were obtained from WorldClim 1.4 using raster package in R (Hijmans et al. 2005; R Core Team 2013).

Results

A genome-wide map of SNP variation for Nigerien sorghum

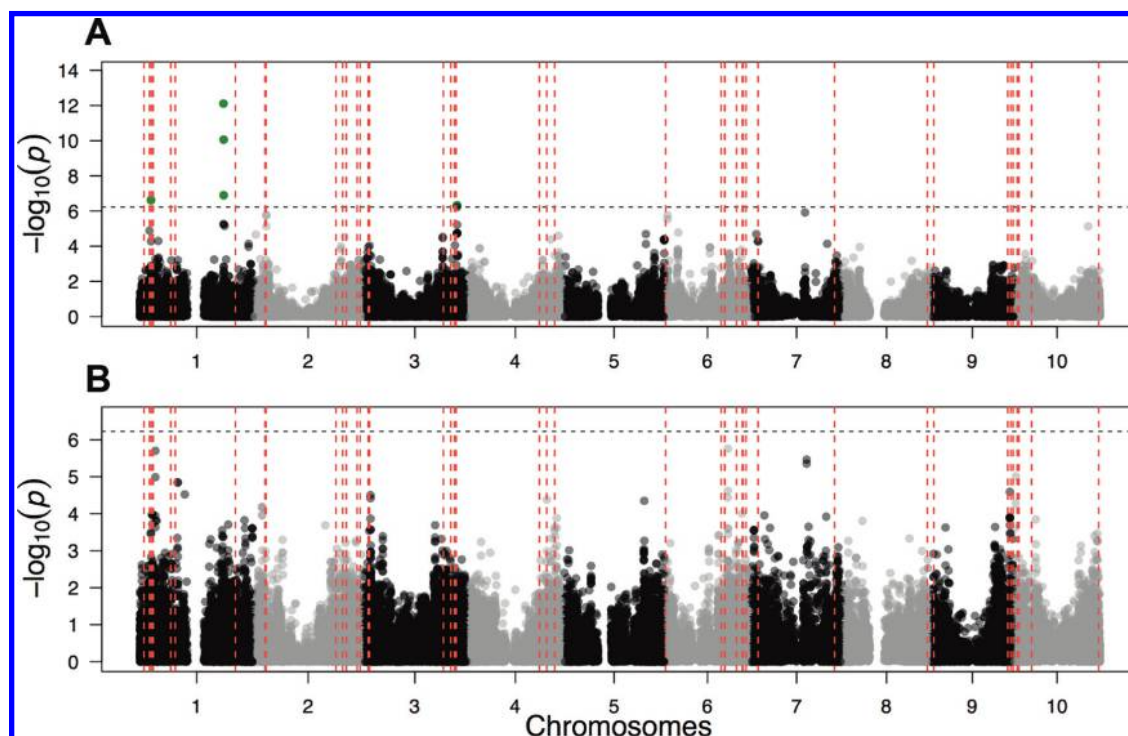
We discovered 158 019 SNPs in the Nigerien sorghum germplasm. After removing 13 720 monomorphic SNPs (8.7%), we

Fig. 5. Genomic differentiation among the varieties with different local varieties. (A) Summary table of the 11 most common local varieties with their number of accessions with local name, botanical races, and kernel colors number. The color gradient shows the scale of the occurrence for each column. (B) Heatmap of genetic differentiation (F_{ST}) values among local varieties. The F_{ST} varies from less differentiated (white) to highly differentiated (dark grey). Sokombe (guinea race) appeared the most differentiated among the local varieties. (C) Map of Niger with the georeferenced accessions for Mota and Jenjari. (D) F_{ST} and Tajima's D genome scans between Mota and the 10 local varieties, Jenjari and the 10 local varieties, and Mota and Jenjari. (E) F_{ST} and Tajima's D genome scans for the *Tannin 1* region between Mota and Jenjari. The x-axis represents the chromosomes and the first y-axis the F_{ST} values. Each point represents a SNP marker. The top 0.1% was used to set the threshold. The colors represent the SNPs above the threshold that differentiated the local varieties. F_{ST} was calculated using Weir and Cockerham's method. The second y-axis represents the Tajima's D average of 100 kb window size. Botanical race key: B, bicolor; C, caudatum; D, durra; G, guinea; CB, caudatum–bicolor; CD, caudatum–durra; CG, caudatum–guinea. Kernel color key: B, brown; R, red; Y, yellow; W, white.



Genome Downloaded from www.nrcresearchpress.com by 88.191.130.124 on 06/05/18 For personal use only.

Fig. 6. Genome-wide association studies of environmental variables. (A) Manhattan plot for annual mean temperature. (B) Manhattan plot for annual precipitation. The x-axis represents the chromosomes and the y-axis the $-\log_{10}(P\text{-values})$ for marker–trait association. Each point represents the SNP marker. The threshold is set based on the Bonferroni correction of the P -values. The dashed lines represent a priori candidate genes.



retained 144 299 SNPs. The average Tajima's D across all the germplasm is 0.36 and the nucleotide diversity (π) is 0.00010 (averaged over 1 kb windows) after removing the singletons. By contrast, the average Tajima's D of the Nigerien diversity alone is 0.51, and the nucleotide diversity (π) is 0.00046. To better understand the genetic structure of Nigerien germplasm relative to global sorghum diversity, diversity analysis and neighbor-joining analysis were conducted with 473 279 SNPs for the Nigerien germplasm and global diversity panel (SAP) (Casa et al. 2008; Morris et al. 2013a).

Nigerien germplasm forms several clusters relative to global diversity (Fig. 2A). The SAP accessions El Mota and Mota Maradi (local varieties from Niger in SAP) grouped with the Nigerien landraces germplasm. By contrast Sepon82, SRN39, and MR732 (non-Nigerien varieties/lines released in Niger) grouped with east African accessions. Figure 2B shows the same neighbor-joining tree color coded by botanical race. The caudatum race was composed of two clusters with one cluster mainly consisting of SAP accessions and the other with Nigerien germplasm. Neighbor-joining analysis of Nigerien germplasm without the SAP shows two durra-caudatum clusters, two caudatum clusters, and one guinea cluster (Fig. S1²).

Population structure

To characterize the genetic structure we performed DAPC, which identified six genetic groups (I–VI) (Figs. 3A, 3B). The genetic groups explained 99% of the SNP variation in five principal components (PCs) (Fig. S2²). The first and the second components, which capture 44% and 18% of variation, respectively, separate groups I–V from group VI (Fig. 3A). The third, fourth, and fifth components explained 15%, 14%, and 9% of the total variation, respectively, and separate groups I–V from each other (Fig. 3B).

We hypothesized that genetic groups reflect to population structure across the agroclimatic zones. We tested relationship between the genetic groups and annual precipitation using analysis of variance (ANOVA) and found a significantly higher annual

precipitation for groups I and VI versus groups II–V (P -value <0.05) (Fig. 3C). Consistent with this finding, we observed strong geographic clustering when the accessions were mapped by group (Fig. 3D). Genome scan of loadings on the first principal component identifies multiple genome regions that are highly differentiated (top 0.1% of SNPs) on each chromosome (Fig. 3E; File S3²). Some of these loading peaks colocalize (within 200 kb) with a priori candidate genes for maturity (*PhyB*, *GI*, *abph1*, and *Vgt1*) (Zhang and Yuan 2014).

To further characterize the genetic groups, we summarized kernel color, botanical race, and other traits for each group (Table 1). Group I (95 accessions) is mostly represented by guinea-caudatum and caudatum-bicolor races. Majority of the accessions in this group have a white kernel color with a predominance of Farfara (Hausa name), Amo Koire (Zarma name), and Gaberi Boul (Kanuri name), whose names indicate the white kernel color. Group II (74 accessions) is predominantly caudatum-bicolor accessions. The main kernel color in this group is brown, and Mota is the most frequent local variety. Group III (87 accessions) is predominantly durra-caudatum accessions with a mix of brown and white kernel color. Most frequent local variety in group III is Amo Koire (27 accessions of 87 accessions). Group IV (158 accessions) is predominantly durra-caudatum accessions. Babadia (25 accessions) (Hausa name) and Mota (14 accessions) are the frequent local varieties in group IV. Group V (77 accessions) is mostly represented by caudatum accessions from central Niger with a mix of brown and white kernel color. Jenjari (Hausa name) is the most common in group V. Group VI (25 accessions), which is the most differentiated from other groups, has most of the accessions from southwestern Niger (21 of 25). Sokombe (Zarma name, guinea botanical race) is the major local variety found in group IV.

Pairwise F_{ST} estimates among botanical races and local varieties

To better understand the genetic differentiation among botanical races, we performed F_{ST} genome scan for each botanical race

and evaluated the Tajima's D based on 100 kb window size for the botanical races (Fig. 4). We considered the six most common botanical races and intermediates (Fig. 4A). Durra, caudatum, and durra-caudatum accessions show low F_{ST} values across the genome with some distinct F_{ST} peaks. In contrast, guinea, guinea-caudatum, and bicolor have high F_{ST} throughout the genome (Fig. 4B). The overall Tajima's D averaged over the 100 kb windows is 0.83.

Next, we cataloged the 11 most common local varieties (with >10 accessions) (Fig. 5A) and characterized genetic differentiation (pairwise F_{ST}) among these varieties (Fig. 5B; File S3²). Sokombe, a guinea variety, is highly differentiated from all others. In contrast, the caudatum variety Mota is less differentiated from others, including Babadia, Bazanga, Bogoba, Fara Dawa, and Lalla. Based on these results, we investigated specific loci that may be involved in this differentiation, focusing on the two most common local varieties, Mota and Jenjari. Accessions of Mota and Jenjari both originate from central Niger, where annual precipitation is 400–600 mm (Fig. 5C).

We identified regions with high F_{ST} values (top 0.1% of SNPs) highlighting the differentiation between Mota versus all other local varieties, Jenjari versus all other local varieties, and Mota versus Jenjari (Fig. 5D; File S3²). We plotted the Tajima's D (100 kb windows) for the local varieties and identified genomic regions with high and low values (Fig. 5D). The F_{ST} estimates between Mota and other frequently observed local varieties are relatively less differentiated, but specific loci were found above the threshold with less differentiation across the genome. In contrast, Jenjari appears highly differentiated with a threshold F_{ST} value of 0.75. Mota and Jenjari have substantial differentiation across the genome, including near the major grain pigmentation gene *Tannin1* (Wu et al. 2012) with Tajima's D of 1.9 near the genomic region (Fig. 5E).

Genome-wide association studies for environmental variables

To further identify the genomic regions underlying local climate adaptation in the Nigerien germplasm, we performed genome-wide association scan for annual mean temperature and annual precipitation (Fig. 6). For annual mean temperature we found significant associations at two loci (5 SNPs) on chromosome 1 (*S1_7769638*, *S1_7769764*, *S1_59520997*, *S1_59654095*, *S1_59654013*) and one locus on chromosome 3 (*S3_67229110*) (Fig. 6A; File S3²). For annual precipitation, there were several association peaks but none exceeded the Bonferroni significance threshold (Fig. 6B).

Discussion

The structure of genomic diversity in Nigerien sorghum

The comparison of the Nigerien collection to the global sorghum association panel SAP (Fig. 2A) helps to characterize the diversity of released varieties in Niger relative to local landraces and global sorghum diversity. As expected, the landrace-derived Nigerien varieties in the SAP (El Mota and Mota Maradi) clustered with Nigerien caudatum landraces. MR732 (parent of the NAD1 hybrid), Sepon82, and SRN39 clustered with other germplasm from ICRISAT, along with East African caudatum accessions, which were major contributors to the ICRISAT breeding program (Vaidya et al. 1988; Bantilan et al. 2004). Together, these results demonstrate that the released sorghum varieties in Niger reflect much, but not all, of the genetic diversity of landraces in Niger. Future studies that include other released varieties (e.g., SSD-35, MDK, and S39) (Ministère de l'agriculture du Niger 2012) and

breeding lines will be needed to evaluate the existing diversity in the breeding program and prioritize efforts to increase genetic diversity (Jordan et al. 2011).

The neighbor-joining tree for the Nigerien germplasm (Fig. S1²) shows clusters separated by botanical races and geography. This is consistent with previous studies that have shown that sorghum populations in Niger, and in Africa more broadly, are structured according to farmer preferences and agroclimatic zones (Barnaud et al. 2007; Deu et al. 2010; Labeyrie et al. 2014). The DAPC analysis (Fig. 3) separates guinea (from the southwestern part of Niger, which receives the greatest annual precipitation) from the other groups. Surprisingly, we also observed two distinct groups of caudatum (groups I and V) and two distinct groups of durra-caudatum (groups III and IV). Both caudatum groups are found in central Niger, in apparent sympatry, yet are genetically distinct. The genetic differentiation may be due to the lack of gene flow between the groups, where group V, mostly composed of Jenjari variety (Fig. 5A; Table 1), is cultivated in distant fields compared to group I due to its tannin content which prevents bird predation in unattended fields (Doggett 1988). In addition, group V accessions (the dark-kernel caudatums) are more commonly found in high rainfall areas than group I accessions (light-kernel caudatums), consistent with natural selection favoring tannins in humid areas (Lasky et al. 2015).

Similarly, ethnic groups and geographical origin may influence population structure (Labeyrie et al. 2014; Westengen et al. 2014). The two durra-caudatum groups are geographically distant and have local varieties of Hausa and Zarma names. Group III is predominantly found in the western part of the country, with Amo Koire as local variety whereas Group IV is found in the central part with Babadia as local variety. In contrast, some local varieties are found in multiple groups suggesting that the population structure is affected by other factors such as seed exchange and food preferences (Deu et al. 2008, 2010).

Genomic regions underlying local adaptation and farmer preferences

To identify genomic regions associated with local adaptation and farmer preferences, we performed a variety of genome scans. The SNP *S9_2371665* (C/G) on chromosome 9 has the highest DAPC loading coefficient among all the SNPs (Fig. 3E). The minor allele frequency is 0.38 for the G allele compared to the C allele (0.62). Interestingly, this SNP is located ~200 kb from the putative ortholog of the maize *Vgt1* gene (Ducrocq et al. 2008), a flowering time gene associated with maize adaptation. Another SNP, *S6_51192692* (MAF = 40%) among the top 0.1% on chromosome 6, was found close to the *Sobic.006G151800* (~61 kb), the ortholog of the maize *abph1* gene for floral regulation (Jackson and Hake 1999; Giulini et al. 2004). This suggests that *Vgt1* and *abph1* are promising candidates for future studies of late-maturity adaptation in guinea sorghum of the Sudanian zone (Deu et al. 2008).

Early maturity is often favored in the Sahelian due to a short rainy season. For instance, Mota (which means “car” in multiple languages in Niger) is early maturing relative to other local landraces (Ejeta and Gressel 2007). The SNP on chromosome 9 (*S9_57517028*, $F_{ST} = 0.43$, MAF = 13%), identified from the F_{ST} genome scan between Mota and other local varieties (Fig. 5D), is colocalized with *SbFL9.1*, a major QTL underlying flowering time in global sorghum germplasm (Thurber et al. 2013; Bouchet et al. 2017).

Kernel color is another main characteristic that guides smallholder farmer's selection (Mekbib 2007). For instance, white grain sorghums have local varieties in different languages of Fara Dawa (Hausa), Farfara (Hausa), Gaberi Boul (Kanuri), and Amo Koire (Zarma) whereas red grain sorghums have local varieties of Jenjari (Hausa), Gaberi Kime (Kanuri), and Ja Dawa (Hausa) names. Kernel color classification from the gene bank demonstrates (Fig. 5A) that varieties named for red grain (Jenjari, for example) are

predominantly brown or red kernel (i.e., dark-grained). The major gene responsible for grain tannin content variation is *Tannin1* (*Sobic.004G280800*) on chromosome 4 (Wu et al. 2012; Morris et al. 2013b; Rhodes et al. 2014). For the genome scan between sympatric red/brown grain (Jenjari) and Mota varieties (Fig. 5D), a high F_{ST} SNP on chromosome 4 (*S4_62359202*, $F_{ST} = 0.47$, $MAF = 34\%$) is close to the *Tannin1* gene (~43 kb). Similarly, the F_{ST} genome scan between Jenjari and Fara Dawa (white-grained) showed highly differentiated SNPs near the *Tannin1* gene *S4_62389180* ($F_{ST} = 0.47$, $MAF = 13\%$), ~73 kb from *Tannin1* (Fig. S3²). Together, this suggests a history of balancing selection on *Tannin1* in Jenjari due to the farmer practice of planting dark-grained Jenjari in distant fields, where high tannin reduces bird predation. In this case, there is feedback of natural selection (from bird predation) and human selection (on grain color) acting on *Tannin1*.

To identify loci that may underlie adaptation to temperature and precipitation we also performed GWAS on these environmental variables. We do not find any genes in our a priori candidate list that are very near (<150 kb) to the significant SNPs. For annual mean temperature, the peak around 7.7 Mbp on chromosome 1 (*S1_7769638* and *S1_7769764*) is ~400 kb from an a priori candidate gene, the sorghum ortholog (*Sobic.001G140200*) of maize floral regulator *knotted1* (*KN1*) (Zhang and Yuan 2014) (Fig. 6A; File S2²). While there is no significant associations for annual precipitation, this is not surprising since there is low power to detect climate association with mixed models when population structure is confounded with climate variables (Lasky et al. 2015). Further phenotyping and trait mapping studies will be needed to confirm if some loci with signatures of selection control flowering time and kernel color in Nigerien germplasm.

Implications for genomics-enabled breeding for smallholder farmers

Genomic regions that are highly differentiated among local varieties and botanical races may be useful in marker-assisted selection for local adaptation. For instance, when a local variety is crossed with a non-adapted trait-donor line, markers could be used to screen progeny lines carrying a larger number of alleles for local adaptation and (or) farmer preference. Local varieties have been used as recurrent parent in INRAN breeding program (Kapran et al. 2007). For example, Mota Maradi, a selection of Mota developed by INRAN, is the second most widely cultivated improved variety in Niger (3.6% of production area) (Walker and Alwang 2015). The F_{ST} estimates between Mota and other local varieties (Fig. 5D) reveal regions on the genome that are differentiated on chromosomes 3, 5, 7, and 9. When improving Mota, those regions can be selected to keep its preferred traits while introgressing new alleles from the donor line.

For some botanical races it is more challenging to prioritize genomic regions for selection. For instance, the F_{ST} for guinea and guinea-caudatum shows high differentiation across almost the entire genome (Fig. 4B), consistent with the population differentiation of guinea sorghums relative to all other botanical races (Figs. 2B, 3A). Thus, it might be easier to select markers across the genome with durra, caudatum, and their intermediates where few genome regions are differentiated (Fig. 4B). Although the F_{ST} genome scan for guinea races did not reveal specific loci under selection, the DAPC loadings (Fig. 3E), Tajima's *D* (Fig. 4), or environmental association scans (Fig. 6) may identify loci that control guinea traits, which could be useful for improving sorghum in humid regions. Follow-up studies that combine population analyses with quantitative trait analyses should provide a more complete understanding of sorghum adaptation in Niger and facilitate breeding of locally adapted and farmer-preferred varieties.

Data availability

Raw sequencing data are available in the NCBI Sequence Read Archive under project accession SRP132525. Imputed hapmap file

available in Dryad Digital Repository (doi:10.5061/dryad.5n2bs6r). Please contact corresponding author for availability.

Acknowledgements

This paper is made possible by the support of the American People provided to the Feed the Future Innovation Lab for Collaborative Research on Sorghum and Millet through the United States Agency for International Development (USAID) under Cooperative Agreement No. AID-OAA-A-13-00047. The contents are the sole responsibility of the authors and do not necessarily reflect the views of USAID or the United States Government. We thank two anonymous reviewers for suggestions that improved the manuscript. Contribution 18-329-J from the Kansas Agricultural Experiment Station.

References

- Bantilan, M., Deb, U., Gowda, C., Reddy, B., Obilana, A., and Evenson, R. (Editors). 2004. Sorghum genetic enhancement: research process, dissemination and impacts. International Crops Research Institute for the Semi-Arid Tropics, Patancheru 502324, Andhra Pradesh, India.
- Barnaud, A., Deu, M., Garine, E., McKey, D., and Joly, H.I. 2007. Local genetic diversity of sorghum in a village in northern Cameroon: structure and dynamics of landraces. *Theor. Appl. Genet.* **114**(2): 237–248. doi:10.1007/s00122-006-0426-8. PMID:17089177.
- Barro-Kondombo, C., Sagnard, F., Chantereau, J., Deu, M., vom Brocke, K., Durand, P., et al. 2010. Genetic structure among sorghum landraces as revealed by morphological variation and microsatellite markers in three agroclimatic regions of Burkina Faso. *Theor. Appl. Genet.* **120**(8): 1511–1523. doi:10.1007/s00122-010-1272-2. PMID:20180097.
- Bezançon, G., Pham, J.-L., Deu, M., Vigouroux, Y., Sagnard, F., Mariac, C., et al. 2009. Changes in the diversity and geographic distribution of cultivated millet (*Pennisetum glaucum* (L.) R. Br.) and sorghum (*Sorghum bicolor* (L.) Moench) varieties in Niger between 1976 and 2003. *Genet. Resour. Crop Evol.* **56**(2): 223–236. doi:10.1007/s10722-008-9357-3.
- Borgel, A., and Sequier, J. 1977. Prospection des mils pénicillaires et sorghos en Afrique de l'Ouest. Campagne 1976 - Niger. IRD. Available from http://horizon.documentation.ird.fr/exl-doc/pleins_textes/divers11-02/01656.pdf.
- Borrell, A.K., Mullet, J.E., George-Jaeggli, B., van Oosterom, E.J., Hammer, G.L., Klein, P.E., and Jordan, D.R. 2014. Drought adaptation of stay-green sorghum is associated with canopy development, leaf anatomy, root growth, and water uptake. *J. Exp. Bot.* **65**(21): 6251–6263. doi:10.1093/jxb/eru232. PMID:25381433.
- Bouchet, S., Olatoye, M.O., Marla, S.R., Perumal, R., Tesso, T., Yu, J., et al. 2017. Increased power to dissect adaptive traits in global sorghum diversity using a nested association mapping population. *Genetics*, **206**(2): 573–585. doi:10.1534/genetics.116.198499. PMID:28592497.
- Browning, S.R., and Browning, B.L. 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* **81**: 1084–1097. doi:10.1086/521987.
- Casa, A.M., Pressoir, G., Brown, P.J., Mitchell, S.E., Rooney, W.L., Tuinstra, M.R., et al. 2008. Community resources and strategies for association mapping in sorghum. *Crop Sci.* **48**(1): 30–40. doi:10.2135/cropsci2007.02.0080.
- Collard, B.C., and Mackill, D.J. 2008. Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philos. Trans. R. Soc. B Biol. Sci.* **363**(1491): 557–572. doi:10.1098/rstb.2007.2170.
- Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., et al. 2011. The Variant Call Format and VCFtools. *Bioinformatics*, **27**(15): 2156–2158. doi:10.1093/bioinformatics/btr330. PMID:21653522.
- Deu, M., Sagnard, F., Chantereau, J., Calatayud, C., Héroult, D., Mariac, C., et al. 2008. Niger-wide assessment of in situ sorghum genetic diversity with microsatellite markers. *Theor. Appl. Genet.* **116**(7): 903–913. doi:10.1007/s00122-008-0721-7.
- Deu, M., Sagnard, F., Chantereau, J., Calatayud, C., Vigouroux, Y., Pham, J.L., et al. 2010. Spatio-temporal dynamics of genetic diversity in *Sorghum bicolor* in Niger. *Theor. Appl. Genet.* **120**(7): 1301–1313. doi:10.1007/s00122-009-1257-1.
- Doggett, H. 1988. Sorghum. 2nd ed. Available from <http://www.wiley.com/WileyCDA/WileyTitle/productCd-0582463459.html> [accessed 22 June 2017].
- Ducrocq, S., Madur, D., Veyrieras, J.-B., Camus-Kulandaivelu, L., Kloiber-Maitz, M., Prestler, T., et al. 2008. Key impact of Vgt1 on flowering time adaptation in maize: evidence from association mapping and ecogeographical information. *Genetics*, **178**(4): 2433–2437. doi:10.1534/genetics.107.084830. PMID:18430961.
- Ejeta, G., and Gressel, J. 2007. Integrating new technologies for *Striga* control: towards ending the witch-hunt. *World Scientific*, doi:10.1142/6470.
- Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., and Mitchell, S.E. 2011. A robust, simple Genotyping-by-Sequencing (GBS) approach for high diversity species. *PLoS ONE*, **6**(5): e19379. doi:10.1371/journal.pone.0019379. PMID:21573248.
- Giulini, A., Wang, J., and Jackson, D. 2004. Control of phyllotaxy by the cytokinin-

- inducible response regulator homologue ABPHYL1. *Nature*, **430**(7003): 1031–1034. doi:10.1038/nature02778. PMID:15329722.
- Glaubitz, J.C., Casstevens, T.M., Lu, F., Harriman, J., Elshire, R.J., Sun, Q., and Buckler, E.S. 2014. TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. *PLoS ONE*, **9**(2): e90346. doi:10.1371/journal.pone.0090346. PMID:24587335.
- Hijmans, R.J. 2016. raster: geographic analysis and modeling with raster data. Available from <http://CRAN.R-project.org/package=raster>.
- Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., and Jarvis, A. 2005. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25**(15): 1965–1978. doi:10.1002/joc.1276.
- Jackson, D., and Hake, S. 1999. Control of phyllotaxy in maize by the *abphyl1* gene. *Development*, **126**(2): 315–323. Available from <http://dev.biologists.org/content/126/2/315> [accessed 14 April 2017].
- Jombart, T., and Ahmed, I. 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, **27**(21): 3070–3071. doi:10.1093/bioinformatics/btr521. PMID:21926124.
- Jordan, D.R., Mace, E.S., Cruickshank, A.W., Hunt, C.H., and Henzell, R.G. 2011. Exploring and exploiting genetic variation from unadapted sorghum germplasm in a breeding Program. *Crop Sci.* **51**(4): 1444–1457. doi:10.2135/cropsci2010.06.0326.
- Kapran, I., Grenier, C., and Ejeta, G. 2007. Introgression of genes for *Striga* resistance into African landraces of sorghum. Integrating New Technol. *Striga Control End.* Witch-Hunt. pp. 129–141. Available from https://books.google.com/books?hl=en&lr=&id=Cfd0DQAQBAJ&oi=fnd&pg=PA129&dq=ejeta+kapran+sm39&ots=pTEW6lgMIB&sig=q7aG_e4Pyy_Vf8UfaqhWUVdfx_g.
- Labeirie, V., Deu, M., Barnaud, A., Calatayud, C., Buiron, M., Wambugu, P., et al. 2014. Influence of ethnolinguistic diversity on the sorghum genetic patterns in subsistence farming systems in eastern Kenya. *PLoS ONE*, **9**(3): e92178. doi:10.1371/journal.pone.0092178. PMID:24637745.
- Lasky, J.R., Upadhyaya, H.D., Ramu, P., Deshpande, S., Hash, C.T., Bonnette, J., et al. 2015. Genome-environment associations in sorghum landraces predict adaptive traits. *Sci. Adv.* **1**(6): e1400218. doi:10.1126/sciadv.1400218. PMID:26601206.
- Leiser, W.L., Rattunde, H.F.W., Weltzien, E., Cisse, N., Abdou, M., Diallo, A., et al. 2014. Two in one sweep: aluminum tolerance and grain yield in P-limited soils are associated to the same genomic region in West African sorghum. *BMC Plant Biol.* **14**: 206. doi:10.1186/s12870-014-0206-6. PMID:25112843.
- Li, H., and Durbin, R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxf. Engl.)*, **26**(5): 589–595. doi:10.1093/bioinformatics/btp698.
- Lipka, A.E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P.J., et al. 2012. GAPIT: genome association and prediction integrated tool. *Bioinformatics (Oxf. Engl.)*, **28**(18): 2397–2399. doi:10.1093/bioinformatics/bts444.
- Mekbib, F. 2007. Infra-specific folk taxonomy in sorghum (*Sorghum bicolor* (L.) Moench) in Ethiopia: folk nomenclature, classification, and criteria. *J. Ethnobiol. Ethnomed.* **3**: 38. doi:10.1186/1746-4269-3-38. PMID:18162135.
- Ministère de l'agriculture du Niger. 2012. Catalogue national des espèces et variétés végétales du Niger. République du Niger, Ministère de l'agriculture. Available from http://www.fao.org/fileadmin/user_upload/spid/docs/Niger/CatalogueNationaldesEspecesetVarietesVegetales-Niger.pdf.
- Mohamed, A.B., van Duivenbooden, N., and Abdoussallam, S. 2002. Impact of climate change on agricultural production in the Sahel – Part 1. Methodological approach and case study for millet in Niger. *Clim. Change*, **54**(3): 327–348. doi:10.1023/A:1016189605188.
- Morris, G.P., Ramu, P., Deshpande, S.P., Hash, C.T., Shah, T., Upadhyaya, H.D., et al. 2013a. Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proc. Natl. Acad. Sci.* **110**(2): 453–458. doi:10.1073/pnas.1215985110. PMID:23267105.
- Morris, G.P., Rhodes, D.H., Brenton, Z., Ramu, P., Thayil, V.M., Deshpande, S., et al. 2013b. Dissecting genome-wide association signals for loss-of-function phenotypes in sorghum flavonoid pigmentation traits. *G3 Genes Genomes Genetics*, **3**(11): 2085–2094. doi:10.1534/g3.113.008417. PMID:24048646.
- National Research Council. 1996. Lost crops of Africa. Vol. I: Grains. Available from <https://www.nap.edu/read/2305/chapter/3> [accessed 27 April 2017].
- Ng'uni, D., Geleta, M., and Bryngelsson, T. 2011. Genetic diversity in sorghum (*Sorghum bicolor* (L.) Moench) accessions of Zambia as revealed by simple sequence repeats (SSR). *Hereditas*, **148**(2): 52–62. doi:10.1111/j.1601-5223.2011.02208.x. PMID:21561449.
- Paradis, E., Claude, J., and Strimmer, K. 2004. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*, **20**(2): 289–290. doi:10.1093/bioinformatics/btg412. PMID:14734327.
- R Core Team. 2013. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from <http://www.R-project.org/>.
- Rhodes, D.H., Hoffmann, L., Rooney, W.L., Ramu, P., Morris, G.P., and Kresovich, S. 2014. Genome-wide association study of grain polyphenol concentrations in global sorghum [*Sorghum bicolor* (L.) Moench] germplasm. *J. Agric. Food Chem.* **62**(45): 10916–10927. doi:10.1021/jf503651t. PMID:25272193.
- Romero Navarro, J.A., Willcox, M., Burguenio, J., Romay, C., Swarts, K., Trachsel, S., et al. 2017. A study of allelic diversity underlying flowering-time adaptation in maize landraces. *Nat. Genet.* **49**(3): 476–480. doi:10.1038/ng.3784. PMID:28166212.
- Schnable, J.C., and Freeling, M. 2011. Genes identified by visible mutant phenotypes show increased bias toward one of two subgenomes of maize. *PLoS ONE*, **6**(3): e17855. doi:10.1371/journal.pone.0017855. PMID:21423772.
- Smith, C.W., and Frederiksen, R.A. 2000. Sorghum, origin, history, technology and production. In *Wiley Series in Crop Science*. Texas A & M University.
- Thurber, C.S., Ma, J.M., Higgins, R.H., and Brown, P.J. 2013. Retrospective genomic analysis of sorghum adaptation to temperate-zone grain production. *Genome Biol.* **14**: R68. doi:10.1186/gb-2013-14-6-r68. PMID:23803286.
- Vaidya, P.K., Raghavender, B., and Mukuru, S.Z. 1988. *Striga* research in sorghum at ICRISAT Center. IITA, Ibadan, Nigeria. pp. 1–24. Available from <http://oar.icrisat.org/7677/> [accessed 29 May 2017].
- Walker, T.S., and Alwang, J. 2015. Crop improvement, adoption and impact of improved varieties in food crops in sub-Saharan Africa. CABI.
- Weir, B.S., and Cockerham, C.C. 1984. Estimating *F*-statistics for the analysis of population structure. *Evolution*, **38**(6): 1358–1370. doi:10.2307/2408641. PMID:28563791.
- Westengen, O.T., Okongo, M.A., Onek, L., Berg, T., Upadhyaya, H., Birkeland, S., et al. 2014. Ethnolinguistic structuring of sorghum genetic diversity in Africa and the role of local seed systems. *Proc. Natl. Acad. Sci.* **111**(39): 14100–14105. doi:10.1073/pnas.1401646111. PMID:25225391.
- Wu, Y., Li, X., Xiang, W., Zhu, C., Lin, Z., Wu, Y., et al. 2012. Presence of tannins in sorghum grains is conditioned by different natural alleles of *Tannin1*. *Proc. Natl. Acad. Sci.* **109**(26): 10281–10286. doi:10.1073/pnas.1201700109. PMID:22699509.
- Zhang, D., and Yuan, Z. 2014. Molecular control of grass inflorescence development. *Annu. Rev. Plant Biol.* **65**(1): 553–578. doi:10.1146/annurev-arplant-050213-040104. PMID:24471834.

This article has been cited by:

1. Theodore J. Morgan, Michael A. Herman, Loretta C. Johnson, Bradley J.C.S. Olson, Mark C. Ungerer. 2018. Ecological Genomics: genes in ecology and ecology in genes. *Genome* **61**:4, v-vii. [[Citation](#)] [[Full Text](#)] [[PDF](#)] [[PDF Plus](#)]