Published March 29, 2018

ORIGINAL RESEARCH

Genome-Wide Association Mapping of Anthracnose (*Colletotrichum sublineolum*) Resistance in the U.S. Sorghum Association Panel

Hugo E. Cuevas,* Louis K. Prom, Elizabeth A. Cooper, Joseph E. Knoll, and Xinzhi Ni

Abstract

The productivity and profitability of sorghum [*Sorghum bicolor* (L.) Moench] is reduced by susceptibility to fungal diseases, such as anthracnose (*Colletotrichum sublineolum* P. Henn.). A limited number of resistant accessions are present in the temperate-adapted germplasm; other exotic sources of resistance are not currently available for breeding programs. Among 335 accessions available to breeders from a previously genotyped sorghum association panel (SAP), we found that 75 were resistant to anthracnose. A phylogenetic analysis of these accessions showed high genetic diversity and multiple resistance sources. Genome-wide association scans (GWAS) were conducted using 268,289 single-nucleotide polymorphisms to identify loci associated with anthracnose resistance. Using logistic regressions for binary measures of resistance responses, we identified three loci within a region on chromosome 5 that have been previously associated with three sources of anthracnose resistance. A GWAS limited to Caudatum germplasm identified an association with a region on chromosome 1 and with the same previous region on chromosome 5. Candidate genes within these loci were related to R-gene families, signaling cascades, and transcriptional reprogramming, suggesting that the resistance response is controlled by multiple defense mechanisms. The strategic integration of exotic resistant germplasm into the SAP is needed to identify additional rare resistance alleles via GWAS.

Core Ideas

- The sorghum association panel includes multiple anthracnose-resistant accessions.
- High genetic diversity among resistant accessions is useful for breeding programs.
- Resistance was associated with pathogen recognition, signaling cascades, and transcriptional reprograming.
- Integration of exotic resistant germplasm can help to identify rare resistant alleles

SORGHUM IS THE FIFTH most important cereal after maize (*Zea mays* L.), wheat (*Triticum aestivum* L.), rice (*Oryza sativa* L.), and barley (*Hordeum vulgare* L.) (Food and Agriculture Organization, 2017). It currently serves as an important dietary staple for over 500 million people worldwide. Sorghum is also cultivated for animal consumption and is becoming an important source of grain and cellulosic-based ethanol. As a C_4 grass that experienced multiple redomestication processes in Africa (Smith and Frederiksen, 2000), sorghum is a highly diverse crop adapted to an array of environments, ranging from tropical to temperate regions, and is classified

H.E. Cuevas, USDA-ARS, Tropical Agriculture Research Station, 2200 Pedro Albizu Campos Avenue, Mayaguez, Puerto Rico 00680; L.K. Prom, USDA-ARS, Southern Plains Agriculture Research Center, 2881 F & B Road, College Station, TX 77845; E.A. Cooper, Advanced Plant Technology Program, Clemson Univ., Clemson, SC 29634; J.E. Knoll, X. Ni, USDA-ARS, Crop Genetics and Breeding Research, 115 Coastal Way, Tifton, GA 31793. Received 14 Nov. 2017. Accepted 8 Jan. 2018. *Corresponding author (hugo.cuevas@ars.usda.gov).

Abbreviations: Chr., chromosome; ECMLM, enriched compressed mixed linear model; GBS, genotyping-by-sequencing; GWAS, genome-wide association study; NPGS, National Plant Germplasm System; SAP, sorghum association panel; SNP, single-nucleotide polymorphism.

Plant Genome 11:170099
doi: 10.3835/plantgenome2017.11.0099

into five major races (Bicolor, Durra, Guinea, Caudatum, and Kafir) (Harlan and Dewet, 1972). These races are associated with particular environments and are differentiated by their inflorescence type (Snowden, 1936).

The United States is the leading sorghum-producing country, accounting for 10.98 million of the 85.2 million tons produced worldwide (Food and Agriculture Organization, 2017). Nevertheless, the productivity and profitability of sorghum is limited by several diseases. Anthracnose, caused by the fungal pathogen *C. sublineolum* in Kabat and Bubák (syn. *Colletotrichum graminicola* (Ces.) G. W. Wilson) is one of the most problematic diseases, causing yield losses of both grain and biomass of up to 50%. Moreover, the pathogen can persist for up to 18 mo in disease residue on or above the soil surface (Thakur and Mathur, 2000). Anthracnose is most prevalent in warm and humid sorghum production regions, where the acreage of sorghum is expanding most rapidly in the United States (USDA, 2014). Because most of the currently available commercial germplasm has been developed in areas with low anthracnose pressure, new resistant lines are needed to meet the demands of growers in these areas. The preferred strategy to control anthracnose is through the incorporation of resistance genes. Although several sources of resistance have been identified in exotic germplasm (Cuevas et al., 2015; Cuevas et al., 2014b; Erpelding and Prom, 2004; Prom et al., 2012a), a limited number are available in temperate-adapted germplasm (Cuevas et al., 2014a; Felderhoff et al., 2016; Patil et al., 2017; Perumal et al., 2009; Prom et al., 2016). The widespread use of just a few resistant lines, coupled with the rapid evolution of the fungal pathogen, creates a system that could be vulnerable to collapse, where even currently resistant lines can become susceptible to an evolved pathogen (Prom et al., 2012b).

The genetic architecture and mode of inheritance of anthracnose resistance are not well understood; studies have reported different numbers of underlying genes with different modes of action. Two closely linked genes with dominant effects were reported by Coleman and Stokes (1954), whereas Tenkouano (1993) identified a single gene with multiple alleles for anthracnose resistance. Moreover, segregation for disease response has been observed in progeny derived from a cross of two resistant inbred lines, indicating that the parents have different genetic sources of resistance (Mehta et al., 2005). Discrepancies between the results of inheritance studies might result from the use of different pathogen isolates, evaluation methods, and/or environment–host–pathogen interactions. In fact, many studies have demonstrated an abundance of different pathotypes of *C. sublineolum* in different regions (Casela et al., 1992; Marley et al., 2001; Pande et al., 1991; Valerio et al., 2005). Similarly, in the United States, Ali and Warren (1987), Cardwell et al. (1989), Moore et al. (2008), and Prom et al. (2012b) have all identified different pathotypes originally from different regions. These studies have resulted in the establishment of 18 sorghum lines with precisely defined pathotypes based on their disease response under greenhouse conditions [i.e., resistance against one particular pathotype (Prom et al., 2012b)].

Nevertheless, several recent studies have demonstrated that one main resistance locus can be effective across multiple pathotypes or environments. The anthracnose resistance gene present in 'SC748–5', which was initially found to control resistance against pathotypes from Texas, was mapped to the distal region of sorghum chromosome 5 (Burrell et al., 2015; Perumal et al., 2009). Likewise, Cuevas et al. (2014a) and Patil et al. (2017) found that the lines 'SC112–14' and 'SC414–12E' have a resistance locus at chromosome 5 that was not associated with previously mapped loci and provides resistance against pathotypes from Puerto Rico, Texas, Arkansas, and Georgia. Another main resistance locus against pathotypes from Florida, Georgia, and Texas was mapped to chromosome 9 using the cultivar 'Bk7' and the line 'SC155–14E' (Felderhoff et al., 2016; Patil et al., 2017). Therefore, genomic resources should be used to identify multiple resistance genes with the objective of breeding for increased resistance durability.

The USDA-ARS National Plant Germplasm System (NPGS) maintains a sorghum germplasm collection that includes >41,860 accessions from 114 countries. Today, it is the primary source of genetic diversity for sorghum breeding programs; however, the tropical exotic germplasm available in this collection cannot be integrated into U.S. sorghum breeding programs without introgressing dwarfing and early maturity genes into each line to enable plant survival in temperate regions (i.e., converted germplasm) (Thurber et al., 2013). Owing to the large size of this collection and the time and cost involved in the introgression procedure, it is imperative to identify the most valuable germplasm with novel alleles for breeding programs. In parallel, a priori knowledge of the genetic diversity and alleles that are already present in breeding programs is necessary to avoid the conversion of genetically similar germplasm. For instance, the NPGS collections from Mali, Uganda, South Africa, and Zimbabwe, among other areas, include anthracnose-resistant germplasm with potential value for breeding programs (Cuevas et al., 2014b, 2016; Erpelding, 2012; Prom et al., 2011) but the lack of genetic data for these lines limits their use. Therefore, the identification and study of resistance sources in germplasm that is already temperate-adapted is necessary for current breeding programs to produce disease-resistant lines rapidly. Additional sources from exotic germplasm can be incorporated by conversion or introgression.

The SAP, comprising 149 U.S. breeding lines and 228 lines adapted to temperate regions (i.e., converted lines), was assembled to capture the majority of genetic diversity present in sorghum breeding programs (Casa et al., 2008). Later, genotyping-by-sequencing (GBS) (Elshire et al., 2011)] was used to identify >250,000 single-nucleotide polymorphisms (SNPs) (Morris et al., 2013a), establishing a genomic resource for the genetic dissection of economically important traits. Multiple GWAS have been performed to determine the genetic basis of dhurrin content,

stalk rot, grain flavonoid and polyphenol concentrations, and yield with the SAP (Adeyanju et al., 2015; Boyles et al., 2016; Hayes et al., 2015; Morris et al., 2013b; Rhodes et al., 2014). However, despite the discovery of many SNP–trait associations in these GWAS, the power to detect loci with small effects or at a low frequency in the panel is still low. For this reason, many traits with more complex genomic architectures (i.e., many genes of small effect) still have a large portion of unexplained variation, despite this dense marker panel. Hence, the assessment of the SAP for economically important traits is essential to identify valuable accessions for breeding programs and to develop genomic resources for marker-assisted selection.

The identification of multiple anthracnose-resistant sources in germplasm adapted to temperate regions is vital to assure long-term disease control in leading production regions. Owing to the diversity and preexisting genomic resources available for the SAP, we characterized the anthracnose resistance of this germplasm. Genome-wide association scans using both binary and quantitative data were used to identify resistance-related loci and strategies to detect rare resistance alleles based on frequency. These results provide critical information for the exploitation of these naturally resistant genetic sources and provide a basis for introgressing new sources from the NPGS collection.

Methods

Germplasm and Field Experiment

A total of 335 accessions from the SAP (Casa et al., 2008), consisting of 213 converted tropical sorghum, 37 breeding lines, and 85 varieties, were evaluated for anthracnose resistance (Supplemental Table S1). The SAP accessions and the susceptible ('BTx623', 'RTx430', and 'PI609251'; Prom et al., 2012b) and resistant (SC112–14; Cuevas et al., 2014a) controls were planted in research farms of the USDA-ARS Tropical Agriculture Research Station at Isabela and Mayaguez, PR, for 2 consecutive years (September 2013–January 2014 and April–August 2014) and at the Gibbs Farm of the University of Georgia at Tifton, GA (May–September 2017). At the three locations, a complete randomized design was used, with plots measuring 1.8 m and 4.0 m in length at Puerto Rico and Georgia, respectively, with 0.9 m between rows. A subset of 46 resistance accessions that enclose the genetic diversity of the SAP was evaluated for anthracnose resistance for two additional consecutive years (April–August 2015 and November 2015–March 2016). The subset, reference checks, and 27 susceptible accessions from the SAP were planted at Isabela, PR, in a randomized complete block design consisting of three blocks with plots 1.8 m in length with 0.9 m between rows in both years. Plants were maintained with standard management practices, and weeds were controlled by mechanical tillage and hand hoeing.

Anthracnose Response

The inoculation and disease assessment methods were similar to those described by Prom et al. (2009). Two fungal cultures were prepared with different isolates of *C. sublineolum* representing the pathotypes present at the Isabela and Mayaguez research farms. The five pathotypes of Isabela had been previously characterized (Prom et al., 2012b), whereas the three pathotypes from Mayaguez have not been characterized. These isolates were used to colonize sorghum seeds, which were placed into the sorghum leaf whorls 30 to 40 d after planting. The high natural disease pressure was sufficient to cause symptoms in Tifton, GA, in 2017. Disease assessment was performed before harvesting on a scale of 1 to 5 as follows: 1 = no symptoms or chlorotic flecks on leaves; 2 = hypersensitive reaction on inoculated leaves but no acervuli in the center; 3 = infected bottom leaves with acervuli; 4 = necrotic lesions with acervuli observed on bottom leaves and spreading to middle leaves; 5 = most leaves dead from infection, including infection on the flag leaf.

Phenotype Analysis

The anthracnose resistance responses for each year and location were combined and averaged to categorize accessions as resistant (<2.0) or susceptible (>2.0). A mixed linear model with locations treated as fixed effects and accessions as a random effect was used to estimate the broad-sense heritability with the following formula:

$$H^2 = \frac{\sigma_G^2}{\sigma_G^2 + \frac{\sigma_E^2}{L}}, \quad [1]$$

where G represents the accession, L is location, and E is an error term (Bernardo, 2002).

The χ^2 test was used to determine whether the frequency of resistant accessions in one population or sorghum race was higher or lower than expected on the basis of the frequencies within the SAP.

Genotype-by-Sequencing

Genotype information for this study is a community resource generated for the entire SAP (Morris et al., 2013a) and was improved by alignment to the most recent version of the sorghum references genome (version 3.1; www.phytozome.net, accessed 15 Feb. 2018) (Boyles et al., 2016). In total, 268,289 SNPs with a minor allele frequency of >0.05 were used for association analyses.

Population Structure

The optimal number of subpopulations within the SAP is estimated to be between four and five and the predicted population structure is closely associated with the different sorghum botanical races (Adeyanju et al., 2015; Boyles et al., 2016). Since the number of accessions used in this study was not identical to those presented in previous population structure analyses (Adeyanju et al., 2015; Boyles et al., 2016), the Bayesian model-based clustering method implemented in STRUCTURE version 2.1 (Pritchard et al., 2000) was used to assign the 335 accessions to clusters on the basis of genotype. A pruned subset of 11,609 unlinked SNPs ($r^2 < 0.10$) was

generated with PLINK (Purcell et al., 2007) and further thinned to 100 SNPs per chromosome with an average of one marker every 50 kb. The final subset of 1000 SNPs was used to test the a priori values of $k = 1$ to $k = 12$ using three independent runs for each k-value based on an admixture model with correlated frequencies, 20,000 burn-in iterations, and 30,000 Markov chain Monte Carlo sampling replicates. As expected, the optimal k-value based on the likelihood of the data was 5; therefore, ancestry membership coefficients for $k = 5$ were matched by permutation in CLUMPP (Jakobsson and Rosenberg, 2007). Accessions with an ancestry membership coefficient of <0.60 were considered to be admixed ($n = 159$) and the other four populations were named according to the sorghum botanical races: Caudatum ($n = 76$), Durra ($n = 33$), Kafir ($n = 47$), and Guinea ($n = 20$) (Supplemental Table S1 and Supplemental Fig. S1).

Association Analysis

Association analyses were completed with both quantitative (scores 1 to 5) and binary (resistant or susceptible) classifications of the anthracnose resistance response. The enriched compressed mixed linear model [ECMLM; (Li et al., 2014)] was implemented to analyze quantitative data via the Genome Association and Prediction Integrated Tool in R (Lipka et al., 2012). For the ECMLM, the population structure matrix (**Q**) was obtained from the first three principal components as calculated by the analysis of 11,609 unlinked SNPs in PLINK (Purcell et al., 2007) and the kinship matrix (**K**) was calculated as described by VanRaden (2008) and implemented in the Genome Association and Prediction Integrated Tool. A multivariate logistic regression model was fitted to the binary data with PLINK (Purcell et al., 2007). The five sorghum races and first three principal components were included as covariates in the logistic regression to control for population structure and family relatedness. Log quantile–quantile (QQ) p-value plots were examined to determine how well ECMLM and logistic models accounted for population structure and relatedness. Manhattan and QQ plots were visualized with the R package qqman (Turner, 2014).

The identification of some resistance loci might have been masked by overcorrection for the population structure in the SAP. Therefore, a second set of association analyses was performed using accessions only from the two largest sorghum race populations in the SAP, namely Caudatum ($n = 76$) and Kafir ($n = 47$). The ECMLM and multivariate logistic regressions were fitted as previously described. Log QQ p-value plots were examined to determine how well both the ECMLM and logistic models accounted for population structure and relatedness. Manhattan and QQ plots were visualized with the R package qqman (Turner, 2014).

Genome-Wide Association Scans' Significance Threshold

Empirical significance thresholds for ECMLM [$-\log_{10}$ (p-value) = 5.90] and logistic regression [$-\log_{10}$ (p-value) = 5.95] analyses were calculated with 1000 permutations for an experiment-wise error rate of $P = 0.05$. Likewise, the significance thresholds for the association analysis limited to the Caudatum sorghum race were also calculated with 1000 permutations for an experiment-wise error rate of $P = 0.05$ for both ECMLM [$-\log_{10}$ (p-value) = 4.66] and logistic regression [$-\text{Log}_{10}$ (p-value) = 4.22].

Targeted Resequencing

Next-generation DNA sequencing has relatively high error rates compared with of traditional dye-terminator sequencing (Fox et al., 2014), and the GBS method often requires genotype imputation to account for the large number of missing data points (Chan et al., 2016). Thus candidate SNPs identified by high-throughput marker screening require further validation. Therefore, the top gene candidates identified by GWAS were also partially sequenced in a subset of accessions via BigDye terminator chemistry (SeqWright Genomic Services; A NeoGenomics Company, Houston, TX) to verify and identify additional valuable SNPs for marker development. The primers were designed with Primer3 [Supplemental Table S2; (Koressaar and Remm, 2007)], and the sequence chromatograms were examined with SEQUENCHER (version 4.1; Gene Codes Corporation, Ann Arbor, MI).

Results

Anthracnose Resistance in the SAP

The majority of accessions in the SAP were susceptible to anthracnose, with a mean quantitative value of 3.51 (on the 1–5 scale) and resistance frequency (score < 2.0) of 0.26 (Table 1). The estimated broad-sense heritability was 0.80, indicating that the majority of observed variation was caused by genetic variation. In fact, the observations were not markedly different across years and locations. A total of 116 accessions showed different anthracnose resistance responses among different locations (i.e., accessions were resistant at one location but susceptible at another), whereas 75 and 144 were resistant and susceptible, respectively, at all locations. The subset of accessions evaluated for two additional consecutive years supported previous results: 40 accessions (87%) were anthracnose-resistant across multiple years of evaluation. Moreover, some of these anthracnose-resistant accessions (e.g., SC748, 'BTx378', and SC155) are also resistant against pathotypes from other locations (Patil et al., 2017; Prom et al., 2012b, 2016). In fact, cultivars with consistent resistance across years at a single location present similar resistance responses at another location (Burrell et al., 2015; Patil et al., 2017). Therefore, these 40 resistant accessions are expected to be broadly resistant against pathotypes from multiple locations.

The geographic origins of sorghum can influence the frequency of anthracnose resistance resulting from both natural and human selection. Originally, the SAP was classified into 10 populations corresponding to particular mixtures of sorghum races (Casa et al., 2008). Genotyping

Table 1. Frequency of the anthracnose resistance response in 335 accessions from the U.S. sorghum association panel.

Population structure based on 47 SSRs (Casa et al. 2008)					Population structure based on 1000 SNPs				
Subpopulations	n	R†	S†	p-value χ^2	Subpopulations	n	R†	S†	p-value χ^2
Caudatum	38	0.32	0.68	0.17	Caudatum	76	0.21	0.79	0.73
Caudatum–Kafir	32	0.09	0.91	0.08	Guinea	20	0.30	0.70	0.41
Caudatum–Bicolor	17	0.59	0.41	0.00	Durra	33	0.18	0.82	0.57
Zerazera–Caudatum	46	0.17	0.83	0.42	Kafir	47	0.11	0.89	0.06
Guinea–Caudatum–Bicolor (East Africa and India)	23	0.61	0.39	0.00	Admixed	159	0.26	0.74	0.21
Guinea–Caudatum (West Africa)	27	0.22	0.78	0.98					
Sudanense–Broomcorn–Guinea	14	0.21	0.79	0.94					
Milo–Feterita	30	0.03	0.97	0.01					
Durra	44	0.20	0.80	0.77					
Kafir	64	0.14	0.86	0.11					

† R, resistant to anthracnose; S, susceptible to anthracnose.

based on the most current alignment indicates that the optimal classification includes four main sorghum races and one large admixture group (Adeyanju et al., 2015; Boyles et al., 2016). Anthracnose-resistant accessions were found at a higher frequency in two populations (Caudatum–Bicolor and Guinea–Caudatum–Bicolor from East Africa and India) when the SAP was divided into 10 populations (Table 1). Highly genetically related accessions were grouped in these 10 populations; therefore, the high frequency of anthracnose resistance in two populations suggests that the genes in these accessions are identical by descent (i.e., the accessions share the same resistance genes). Nevertheless, when the SAP was divided into the four main sorghum races, the frequencies of anthracnose-resistant accessions were similar. Remarkably, some anthracnose-resistant accessions were present in each of the 10 populations or four sorghum races; in both analyses, Kafir germplasm was the most susceptible.

An adequate knowledge of the genetic diversity in anthracnose-resistant germplasm is necessary to use the diversity effectively in sorghum breeding programs. An unrooted neighbor-joining tree was generated to understand the genetic relationships among resistant accessions (Fig. 1; Supplemental Fig. S2). The anthracnose-resistant lines captured much of the genetic diversity present in the SAP as a whole; therefore, there does not seem to be a single subpopulation in which resistance evolved. In fact, the average genetic distance among resistant accessions was similar to that observed in the SAP (0.19 and 0.17, respectively). Moreover, resistant accessions represented 14 countries, including Ethiopia and Sudan (13 and 8 accessions, respectively), considered the center of origin for the crop (Smith and Frederiksen, 2000). Thirty-three accessions were classified into the four main sorghum races (Caudatum = 16; Durra = 6; Guinea = 6; Kafir = 5) and 42 were considered to be admixed (Table 2). Remarkably, admixed accessions were detected in all clades of sorghum races, indicating the possible combination of sources of resistance. Certainly, these subsets of resistant accessions indicate high genetic diversity for the selection of the most suitable germplasm for sorghum production regions with high anthracnose disease pressure.

Genome-Wide Association Analysis

Genome-wide association scans with ECMLM (quantitative data) could not associate genomic regions with anthracnose resistance. In contrast, logistic regression (binary categories) detected three loci at the distal region of chromosome (Chr.) 5 (Chr5: 65193948, 66491767, and 71578176) that were strongly associated with anthracnose resistance (Fig. 2). The discrepancy between the two analyses could be attributed to the fact that ECMLM accounts for variation associated with more minor-effect genes (i.e., tolerance); in contrast, logistic regression is limited to the effects of major resistance genes. Indeed, the genomic region on chromosome 5 has been previously associated with anthracnose resistance in three biparental inheritance studies (Burrell et al., 2015; Cuevas et al., 2014a; Patil et al., 2017; Perumal et al., 2009) that used different resistance sources. Both analyses had adequate control for population structure and familial relatedness based on the visual inspection of QQ plots (Supplemental Fig. S3).

A phylogenetic analysis showed that anthracnose resistance was not confined to a single race or geographic population. This suggests that the alleles that are important for resistance pre-date the diversification of sorghum into different races, but it is also possible that parallel mechanisms have evolved in each of the racial groups (or some combination of the two). To identify possible resistance genes particular to a single racial genetic background, ECMLM and logistic regression analyses were conducted within subpopulations or races where there were enough accessions to allow for adequate statistical power. The GWAS using logistic regression and limited to Caudatum individuals identified genomic regions on chromosomes 1 and 5 associated with anthracnose resistance (Fig. 3). In this particular analysis, we detected one locus on chromosome 5 (Chr5: 65194648) that was identified by logistic regression with the entire SAP. Based on this result, it is clear that a GWAS limited to a particular sorghum race can lead to the identification of loci that are spuriously masked by overcorrection for population structure in the analysis of a structured panel. Indeed, the lack of linkage disequilibrium among anthracnose resistance loci on chromosomes 1 and 5 ($r^2 < 0.1$) is evidence that

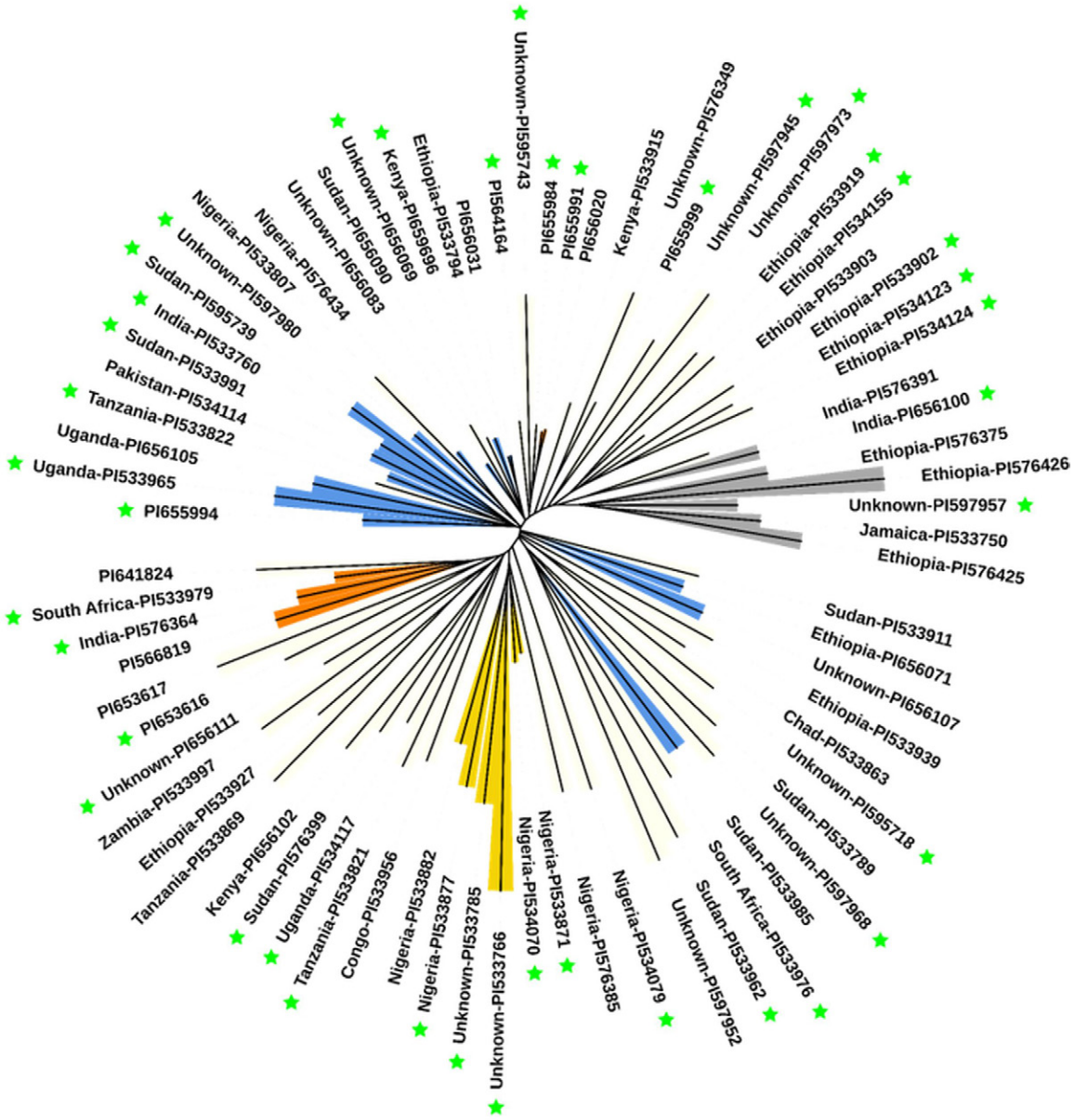

Fig. 1. Unrooted neighbor-joining tree of 75 anthracnose resistance accessions present in the sorghum association panel. (A) Each branch includes country of origin and PI and the colors represent the four sorghum races (blue = Caudatum; gray = Durra; yellow = Guinea; orange = Kafir; no color = admixture) based on a STRUCTURE analysis. Anthracnose resistant accessions evaluated for 4 consecutive years are labeled with a star.

they are completely independent and are not associated with genetic relatedness among resistant accessions. The identification of a novel locus on chromosome 1 is suggestive of at least some parallel race-specific mechanisms, whereas the repeated finding of an association with a region on chromosome 5 simultaneously shows that some alleles may have been present in the common ancestor.

Candidate Genes in Anthracnose Resistance Loci

Multiple candidate resistance genes were identified within each region of chromosomes 1 and 5 (Table 3). The first locus on chromosome 5 consisted of a 908-bp region with four SNPs (Chr5: 65193948, 65194648, 65194754, and 65194856) in linkage disequilibrium ($r^2 > 0.90$). These SNPs were located within the coding region of the putative gene *Sobic.005G172300* (F-box domain). The other two loci on chromosome 5 consisted of two SNPs (Chr5: 66491767 and 71578176) located with the coding region of the putative genes *Sobic.005G182400* (protein tyrosine kinase, leucine-rich repeat *N*-terminal domain, leucine-rich repeat) and *Sobic.005G228400* (oryzalide A biosynthesis). Resistance alleles were present in 42 of the resistant accessions (56%) and two accessions [PI533902 (SC6) and PI534124 (SC15)] carry resistant alleles of the three loci. Remarkably, the linkage disequilibrium among associated SNPs was <0.05, suggesting they were independent. The region is within the anthracnose resistance locus identified in SC112–14 (Cuevas et al., 2014a), SC414–12E (Patil

Table 2. Anthracnose-resistant accessions in the sorghum association panel.

		Populations				Populations	
Accession	Name	SNPs†	SSRs‡	Accession	Name	SNPs†	SSRs‡
PI533965	SC115	CAU§	1	PI533902	SC6	ADX	1
PI576434	SC1103	CAU	2	PI533903	SC17	ADX	1
PI655994	Wiley	CAU	3	PI533927	SC224	ADX	1
PI595739	SC1055	CAU	4	PI533869	SC283	ADX	1
PI533760	SC120	CAU	4	PI533997	SC465	ADX	1
PI656031	CE151_262	CAU	4	PI656102	SC59	ADX	1
PI533794	SC110	CAU	4	PI533956	SC623	ADX	1
PI533985	SC701	CAU	4	PI533915	SC79	ADX	1
PI656107	SC702	CAU	4	PI655999	Tx2784	ADX	2
PI533991	SC748	CAU	4	PI576349	SC942	ADX	2
PI597980	SC1345	CAU	9	PI595743	SC1201	ADX	3
PI656071	SC1019	CAU	9	PI641824	KansasOrange	ADX	3
PI656083	SC1451	CAU	9	PI576385	SC1070	ADX	3
PI656105	SC639	CAU	9	PI564164	RTx433	ADX	4
PI659696	SC720	CAU	10	PI595718	SC1246	ADX	5
PI534114	SC574	CAU	10	PI534155	SC155	ADX	5
PI655991	BTx378	KAF	3	PI576391	SC213	ADX	5
PI533979	SC628	KAF	3	PI655984	Chiltex	ADX	6
PI576364	SC782	KAF	3	PI533919	SC124	ADX	7
PI656020	BTx406	KAF	3	PI597968	SC1321	ADX	7
PI566819	Della	KAF	3	PI597973	SC1330	ADX	7
PI533766	SC265	GUI	8	PI656069	SC326_6	ADX	7
PI534070	SC279	GUI	8	PI597945	SC855	ADX	7
PI533785	SC299	GUI	8	PI656111	SC971	ADX	7
PI533877	SC396	GUI	8	PI534117	SC991	ADX	7
PI533871	SC566	GUI	8	PI653617	Keller	ADX	7
PI533882	SC399	GUI	8	PI656090	SC1494	ADX	7
PI597957	SC1158	DUR	5	PI533863	SC320	ADX	7
PI656100	SC500	DUR	5	PI534079	SC413	ADX	9
PI576375	SC1014	DUR	5	PI533962	SC60	ADX	9
PI576426	SC1033	DUR	5	PI533976	SC655	ADX	9
PI576425	SC1155	DUR	5	PI533807	SC223	ADX	9
PI533750	SC214	DUR	5	PI533939	SC557	ADX	9
PI534123	SC13	ADX	1	PI533789	SC57	ADX	9
PI534124	SC15	ADX	1	PI533911	SC58	ADX	9
PI533821	SC322	ADX	1	PI597952	SC738	ADX	9
PI576399	SC323	ADX	1	PI653616	Wray	ADX	10
PI533822	SC418	ADX	1				

† Population structure based on 1000 single-nucleotide polymorphisms (SNPs).

‡ Population structure based on 47 simple sequence repeats (SSRs) (Casa et al. 2008): 1 = Guinea–Caudatum–Bicolor (East Africa and India); 2 = Sudanense–Broomcorn–Guinea; 3 = Kafir; 4 = Zerazera–Caudatum; 5 = Durra; 6 = Milo–Feterita; 7 = Caudatum–Bicolor; 8 = Guinea–caudatum (West Africa); 9 = Caudatum; 10 = Caudatum–Kafir.

§ CAU, Caudatum; KAF, Kafir; GUI, Guinea; DUR, Durra; ADX, admixture group.

et al., 2017), and SC748–5 (Burrell et al., 2015; Perumal et al., 2009). The plant resistance response is associated with protein variants that involve the detection of pathogens, signaling cascades that activate the production of defense compounds, or both (Bent and Mackey, 2007). F-box proteins constitute one of the largest superfamilies in plants (Jain et al., 2007), including families associated with the regulation of cell death and defense responses in tobacco (*Nicotiana tabacum* L.) and tomato (*Solanum lycopersicum* L.) (van den Burg et al., 2008). Transmembrane proteins with an extracellular leucine-rich repeat and an intracellular protein kinase are involved in the initiation of plant disease resistance (Tang et al., 1996), and in numerous *R* genes shared among plant families (Godiard et al., 2003; Song et al., 1995; Scheer and Ryan, 2002). The diverse functions of candidate genes suggests that the resistance response is controlled by multiple defense mechanisms.

The loci on chromosome 1 consist of four SNPs: two (Chr1: 66554365, 66554507) located within the coding region of the putative gene *Sobic.001G377200* (glucuronosyl tranferase) and two (Chr.1: 66786128 and 66786951) located in the 3′ UTR of *Sobic.01G379300* (peroxidase 2 precursor) and the coding region of *Sobic.001G379400* (peroxidase). The linkage disequilibrium between this 230-kbp region (Chr1: 66554507 and 66786128) was <0.40, suggesting that the two loci were independent. Genes with similar functions to those located within this region are associated with disease resistance mechanisms, including glucuronosyl transferases (Umemura et al., 2009) and hypersensitive responses, including peroxidase (Levine et al., 1994), suggesting that these candidates deserve further study.

Allele Frequency Distribution

The frequency of anthracnose resistance alleles across the four major sorghum races provides an insight into the origin of the phenotype, which in turn helps us to identify the racial groups with the best potential as natural sources of resistance. According to the analysis, Caudatum and Guinea germplasm are the two main sources of anthracnose resistance in the SAP (Table 4). The frequency of resistance alleles for the two loci on chromosome 1 ranged from 0.61 to 0.73 in the SAP and they were nearly fixed within the Kafir, Durra, and Guinea germplasm. Therefore, trait associations for these loci were identified on the basis of the genetic variation present within Caudatum germplasm. The inconsistency between the high frequency of anthracnose resistance alleles and the observed resistant responses in Kafir, Durra, and Guinea germplasm could be attributed to differences in the decay of linkage disequilibrium among races (Morris et al., 2013a), other mutations within Caudatum germplasm that are in linkage disequilibrium with this SNP, and complex resistance-related molecular mechanisms requiring interactions with other proteins fixed in Caudatum germplasm [i.e., the influence of the genetic background (Cao et al., 2007)]. Likewise, the frequency of resistance alleles at loci on chromosome 5 varied among races. The frequency of resistant allele at the locus Chr.5: 65193948 was higher within Caudatum (0.13), indicating that this association was based on the genetic variation present within this germplasm. Trait associations for loci Chr.5: 66491767 and 71578176 were based on genetic variation present within Guinea germplasm and had resistant allele frequencies of 0.15 and 0.45, respectively. On the basis of these findings,

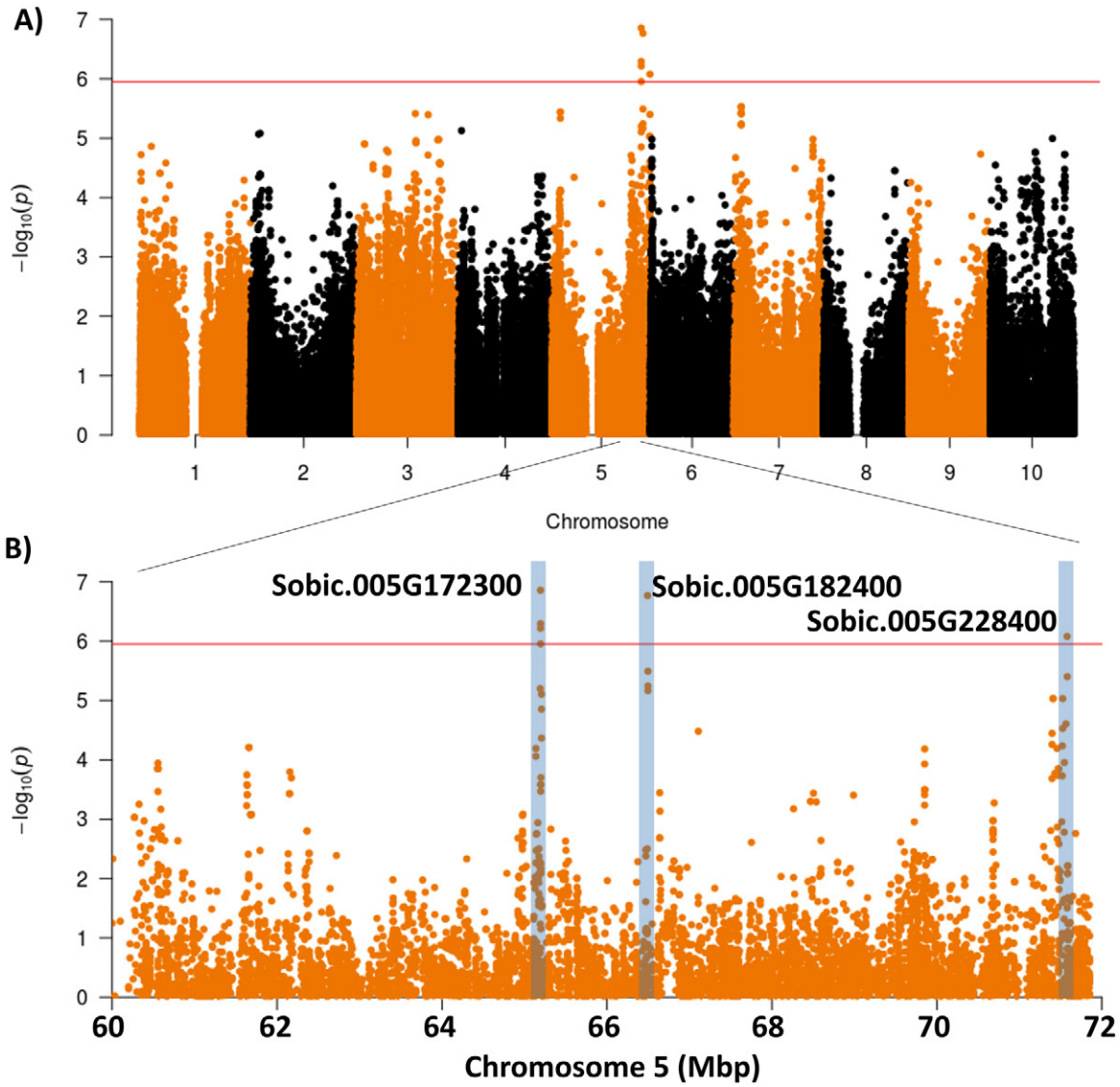

Fig. 2. Genome-wide association analysis for anthracnose resistance in the sorghum association panel. (A) Manhattan plot for logistic regressions based on a case–control analysis (i.e., resistant or susceptible) and the significance threshold ($p < 0.05$) based on 1000 permutations. (B) Anthracnose resistance loci present in chromosome 5.

there are clearly other important genes and combinations of alleles that determine anthracnose resistance within the SAP, but the individual effects of these genes and alleles are either too small to detect by GWAS or the frequency is too low within the SAP to be discernable.

Resequencing of Candidate Genes

The validation of SNPs by a resequencing analysis is imperative prior to the design of low-cost polymerase chain reaction-based genotyping assays used in marker-assisted selection. Hence, gene candidates within the two anthracnose resistance loci identified by logistic regression were partially sequenced in a subset of 74 Caudatum accessions. For the locus at chromosome 1, we resequenced exon 4 of *Sobic.01G379300* (peroxidase 2 precursor) and exon 2 of *Sobic. 001G379400* (peroxidase), both flanking the associated SNP Chr.1: 66786128 ($p = 2.57 \times 10^{-5}$). In *Sobic.01G379300*, we detected six SNPs ([validating one SNP, Chr.1: 66786039, from the GBS analysis) that encode two protein variants, neither of which was in linkage disequilibrium with SNP Chr.1: 66786128 ($r^2 < 0.1$). By contrast, by *Sobic.01G379400* resequencing, we detected six SNPs [validating five SNPs from the GBS analysis: Chr.1: 66786951, 66787125, 66787129, 66787138, and 66787228] encoding three protein variants in linkage disequilibrium with SNP Chr.1: 66786128 ($r^2 > 0.70$). *Sobic.01G379400* definitely deserves further functional genomics analysis and validation in biparental segregating populations derived from resistant Caudatum germplasm. The limited recombination events present in breeding populations and the proximity of the SNPs from both genes (1.4 kbp apart) imply that they could be equally effective for marker-assisted selection. By partial sequencing of *Sobic.05G172300* (F-box domain) including the upstream (100 bp) and coding region [1126 bp (81%)], we detected four nonsynonymous SNPs

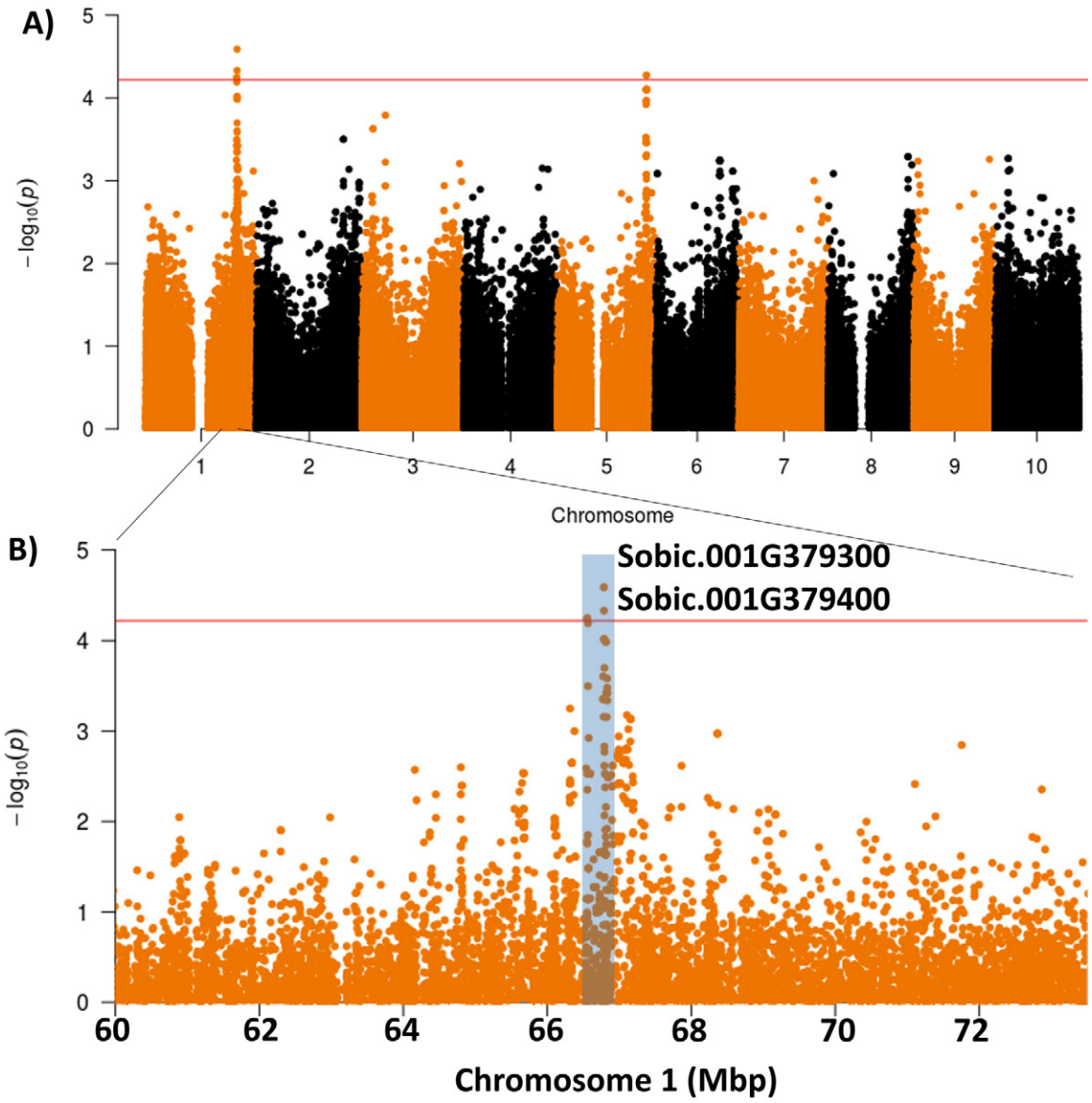

Fig. 3. Genome-wide association analysis for anthracnose resistance limited to Caudatum sorghum germplasm present in the association panel. (A) Manhattan plot for logistic regressions based on a case–control analysis (i.e., resistant or susceptible) and the significance threshold ($p < 0.05$) based on 1000 permutations. (B) Anthracnose resistance locus present in chromosome 1.

(validating two SNPs, Chr.5: 65194648 and 65194754) encoding three protein variants. Remarkably, one variant was only present in a subset of resistant accessions, whereas the other two were found in both resistant and susceptible accessions. This genomic region is evidently associated with anthracnose resistance (Cuevas et al., 2014a; Patil et al., 2017); however, on the basis of these SNPs, it is still unclear if *Sobic.05G172300* is the causal gene. Understanding the molecular mechanism of these resistance loci will be key for the targeted mining of new resistant alleles from the exotic germplasm collection.

Discussion

The SAP is a valuable germplasm collection with a publicly available genomic resource that is widely used for the genomic dissection of multiple agronomics traits (Adeyanju et al., 2015; Boyles et al., 2016; Morris et al., 2013a, 2013b; Rhodes et al., 2014; Zhang et al., 2015). Indeed, its genetic diversity is considered to be the basis of sorghum breeding programs and should be used as a reference during the assessment of the NPGS exotic germplasm collection. Therefore, the anthracnose-resistant accessions and genetic loci identified herein, together with other previously identified resistance loci (Burrell et al., 2015; Felderhoff et al., 2016; Patil et al., 2017), are available sources for U.S. breeding programs. The introgression of new sources of resistance from the NPGS germplasm collection must be based on SAP genetic diversity to avoid the duplication of similar sources of resistance (i.e., identical by descent).

Genome-wide association scanning requires large sample sizes to identify significant phenotypic associations with rare alleles (Korte and Farlow, 2013). The frequency of anthracnose resistance in the SAP was moderate

Table 3. Summary statistics for logistic genome-wide association analyses for anthracnose resistance responses in the sorghum association panel (SAP) and Caudatum germplasm. The logistic regression was based on a case–control analysis (i.e., resistant or susceptible).

Logistic	SNP genomic position	*p*-value	*n*	Allele	Frequency†	OR‡	Gene§	Annotated function§
SAP	Chr.5: 65193948¶	1.39×10^{-7}	335	C	0.11	2.87	*Sobic.005G172300*	F-box domain
				T	0.89	–		
SAP	Chr.5: 66491767	1.71×10^{-7}	335	T	0.07	3.66	*Sobic.005G182400*	Protein tyrosine kinase, leucine-rich repeat
				A	0.93	–		*N*-terminal domain, leucine-rich repeat
SAP	Chr.5: 71578176	8.34×10^{-7}	335	C	0.15	2.36	*Sobic.005G228400*	Oryzalide A biosynthesis
				G	0.85	–		
Caudatum	Chr.1: 66554507	5.66×10^{-5}	76	C	0.17	5.63	*Sobic.001G377200*	Glucuronosyl transferases
				G	0.83	–		
Caudatum	Chr.1: 66786128	2.57×10^{-5}	76	G	0.28	5.57	*Sobic.001G379400*	Peroxidase
				C	0.72	–		

† Frequency of alleles.

‡ The odds ratio (OR) was computed by logistic regression in PLINK.

§ Gene candidates and annotated function, based on the sorghum reference genome version 3.1.

¶ Chr., chromosome.

Table 4. Allele frequency distribution of significant single nucleotide polymorphisms (SNPs) for anthracnose-resistant associations in the sorghum association panel (SAP) and the four main sorghum races, Caudatum (CAU), Kafir (KAF), Durra (DUR), and Guinea (GUI).

			Allele frequencies %				
SNP genomic position	GWAS†	Resistance allele	SAP	CAU	KAF	DUR	GUI
Chr.1: 66554507‡	Caudatum	C	61	17	95	100	55
Chr.1: 66786128	Caudatum	G	73	28	88	97	95
Chr.5: 65193948	SAP	C	11	13	2	2	5
Chr.5: 66491767	SAP	T	7	1	0	0	15
Chr.5: 71578176	SAP	C	15	6	0	12	45

† The genome-wide association study (GWAS) was based on a case–control analysis (i.e., resistant or susceptible).

‡ Chr., chromosome.

(26%); however, it is evident that this resistant response is controlled by multiple loci at low frequencies, since the associations we identified could not explain the variation in resistance observed in the panel. Despite the high genetic diversity present in the SAP, the majority of the germplasm consisted of Caudatum accessions, a sorghum race characterized by a high yield potential (Snowden, 1936). Hence, resistance sources originally from other sorghum races are unlikely to be detected in the SAP. In fact, we obtained very similar results for the GWAS limited to Caudatum germplasm ($n = 76$) and the GWAS including the whole SAP ($n = 335$). Germplasm originally from Ethiopia, Sudan, and Nigeria comprised 36% of the resistant accessions. Moreover, we did not detect the resistant loci present in SC115 [PI 534155; Patil et al., (2017)] located on chromosome 9, although the accession showed a resistant response. Indeed, increasing the number of resistant accessions from other sorghum races and countries is needed to detect these and other resistant loci present in the SAP. In this regard, exotic anthracnose-resistant germplasm from different African countries (Cuevas et al., 2014b; Cuevas et al., 2016; Erpelding, 2011; Erpelding and Prom, 2004; Prom et al., 2011, 2012a) could be strategically selected, genotyped, and integrated into the SAP.

Following its domestication in northern and central Africa, sorghum has undergone significant diversification, as it has spread to new and diverse environments, leading to the formation of the distinct races observed today. Genome-wide association scans using correlations between phenotypes and environmental factors have identified genomic signatures of adaptation for particular regions (Lasky et al., 2015). The Guinea sorghum race, domesticated in humid regions of West and Central Africa, is associated with resistance to biotic stress (Smith and Frederiksen, 2000) but is the most under-represented race in the SAP. Indeed, the loose and low-yield panicles that characterize this race make it the least attractive from a commercial standpoint. Certainly, the SAP encompasses the wide range of sorghum genetic diversity that has been exploited in breeding programs but it is not representative of the vast genetic diversity present in the NPGS collection. The NPGS sorghum germplasm collection includes 3044 accessions classified as the Guinea race originating from more than 20 countries, which exhibit a high frequency of anthracnose resistance (Cuevas et al., 2014b; Erpelding, 2011, 2012; Erpelding and Prom, 2004). The expected assembly of a representative subset of Guinea germplasm by low-coverage resequencing combined with phenotypic characterization will be valuable for the discovery of new sources of disease resistance currently absent in the SAP. In fact, the genomic characterization of the NPGS Ethiopian germplasm belonging to the Durra race showed that a single race has adequate genetic diversity and population structure for GWAS (Cuevas et al., 2017). As resequencing costs continue to decline and genotype imputation becomes more efficient, the genetic diversity in the NPGS sorghum collection will feasibly be integrated into breeding programs.

The anthracnose-resistant accessions present in the SAP could be used to achieve a more immediate but still durable anthracnose resistance response in sorghum production regions. The field population dynamics and genetic diversity of *C. sublineolum* are not well understood (Prom et al., 2012b) and the continued use of limited sources of resistance increases the risk that the pathogen will evolve to bypass the current resistance (McDonald and Linde, 2002). Here, we identified 75 resistant accessions in the SAP, including the lines SC748 and SC155, whose major resistance locus has been mapped (Patil et al., 2017; Perumal et al., 2009). Moreover, three resistant accessions ('SC1103', 'SC265', and 'SC1345') are parental lines in sorghum nested association mapping populations (Yu et al., 2013). Phylogenetic analysis revealed high genetic divergence among these different resistance sources, with four located on distant branches and two (SC1103 and SC1345) sharing a branch on the unrooted tree (Fig. 3). Therefore, a representative subset of 10 to 15 resistant accessions could be assembled on the basis of a phylogenetic analysis and used by breeders as multiple sources of resistance. Despite the unknown inheritance mechanisms and genomic architecture underlying the majority of these anthracnose-resistant responses, the derived resistant varieties and hybrids can still be used in combination with regional and temporal deployment strategies. Different combinations of resistance-related genes deployed either within a single variety or in different varieties can be systematically used to reduce the size of pathogen populations and generate patterns of disruptive selection that increase durable resistance (Burdon et al., 2014).

The molecular mechanism of anthracnose resistance is not well understood. Candidate genes within identified QTL regions include the NB-ARC (nucleotide-binding adaptor shared by APAF-1, certain *R* gene products and CED-4) and nucleotide-binding leucine-rich repeat genes, pathogenesis-related genes, cell death genes, and genes related to the biosynthesis of defense compounds (Biruma et al., 2012; Burrell et al., 2015; Cuevas et al., 2014a; Felderhoff et al., 2016; Patil et al., 2017; Upadhyaya et al., 2013). The family-based approach used in previous QTL studies enabled the identification of extended genomic regions with high numbers of candidate genes. By contrast, the large number of recombination events in an association panel such as the SAP allows for the fine mapping of casual genes or linkage disequilibrium blocks with significantly fewer candidate genes (Zhu et al., 2008). In this study, we found low linkage disequilibrium between neighboring anthracnose resistance loci, and the associated SNPs were either within or in close proximity to individual candidate genes. Typically, genomic regions with resistance genes evolve at fast rates by recombination and transposable elements to compete with the rapid evolution of pathogens (Bergelson et al., 2001). Somewhat surprisingly, the candidate genes within the resistant loci on chromosomes 1 and 5 belong to R-gene families and to other aspects of the immune response, suggesting that resistance in sorghum is complex. The plant immune system is divided into two groups of proteins: (i) transmembrane pattern recognition receptors, and (ii) proteins that act largely within the cell, typically encoded by R-genes (Jones and Dangl, 2006). Microarray analyses have identified genes that are upregulated in response to pathogens and have led to the elucidation of signal transduction pathways related to defense reactions (Michelmore, 2003). Although very little is known about the signaling events, transcription factors in the WRKY and TGA families are associated with local and systemic changes in gene expression (Eulgem, 2005). Previous results (Felderhoff et al., 2016; Patil et al., 2017; Upadhyaya et al., 2013), as well as those of the present study, strongly suggest that anthracnose resistance in sorghum is mainly controlled by polymorphisms in genes involved in these signaling cascades and transcriptional reprograming, instead of being limited to the recognition of pathotype-associated molecular patterns.

Conclusions

Owing to its overall genetic diversity, the SAP is an important source for anthracnose-resistant lines for sorghum breeding programs; however, our results show that this resource may be limited to the detection of alleles shared among races or alleles particular to Caudatum. In fact, the extensive Caudatum germplasm seems to limit the identification of sources of resistance originally from other sorghum races, which are represented at low frequencies within the panel. We did, however, successfully identify several novel candidate loci, both across the entire panel and within Caudatum alone. Anthracnose resistance loci on chromosomes 1 and 5 could be incorporated into elite susceptible germplasm to potentially increase resistance. Consistent with other studies of sorghum, we found that the molecular mechanism underlying anthracnose resistance in this crop is mainly determined by genes associated with upstream functions in resistance pathways, although further studies are needed to determine the exact roles of these genes. This assessment of anthracnose resistance is essential for selecting lines within the SAP to use in immediate breeding efforts and for strategically mining the NPGS sorghum collection to identify new sources of resistance in the near future.

Supplemental Information

Supplemental Fig. S1: Hierarchical organization of the genetic relatedness of 340 accessions from the U.S. Sorghum Association Panel based on the population structure analysis of 1000 unlinked SNPs.

Supplemental Fig. S2: Unrooted neighbor-joining tree of 335 SAP accessions, where green branches represent anthracnose-resistant accessions.

Supplemental Fig. S3: Log quantile–quantile (QQ) *p*-value plots for the genome-wide association analysis for anthracnose resistance response in the sorghum association panel. (A) Plots based on logistic regressions using the binary scores of SAP; (B) plots based on logistic regressions of the binary scores of Caudatum germplasm.

Supplemental Table S1: Anthracnose resistance response of the U.S. sorghum association panel.

Supplemental Table S2: Primer sequences, temperature of annealing (Ta), and SNP positions for the partial sequencing of three putative candidate genes associated with anthracnose resistance in sorghum.

Conflict of Interest Disclosure

The authors declare that there is no conflict of interest.

Acknowledgments

This research project was funded by USDA-ARS Current Research Information System project 6090-21000-053-00-D and the Department of Energy Office of Science, Biological and Environmental Research, grant no. DE-SC0014171.

References

Adeyanju, A., Little, C., Yu, J. M., and Tesso, T. 2015. Genome-wide association study on resistance to stalk rot diseases in grain sorghum. G3 (Bethesda) 5, 1165–1175.

Ali, M.E.K., and H.L. Warren. 1987. Physiological races of *Colletotrichum graminicola* on sorghum. Plant Dis. 71:402–404. doi:10.1094/PD-71-0402

Bent, A.F., and D. Mackey. 2017. Elicitors, effectors, and R genes: The new paradigm and a lifetime supply of questions. Annu. Rev. Phytopathol. 45:399–436.

Bergelson, J., M. Kreitman, E.A. Stahl, and D.C. Tian. 2001. Evolutionary dynamics of plant R-genes. Science 292:2281–2285. doi:10.1126/science.1061337

Bernardo, R. 2002. Breeding for quantitative traits in plants. Stemma Press, Woodbury, MN.

Biruma, M., T. Martin, I. Fridborg, P. Okori, and C. Dixelius. 2012. Two loci in sorghum with NB-LRR encoding genes confer resistance to *Colletotrichum sublineolum*. Theor. Appl. Genet. 124:1005–1015. doi:10.1007/s00122-011-1764-8

Boyles, R.E., E.A. Cooper, M.T. Myers, Z. Brenton, B.L. Rauh, G.P. Morris, et al. 2016. Genome-wide association studies of grain yield components in diverse sorghum germplasm. Plant Genome 9. doi:10.3835/plantgenome2015.09.0091

Burdon, J.J., L.G. Barrett, G. Rebetzke, and P.H. Thrall. 2014. Guiding deployment of resistance in cereals using evolutionary principles. Evol. Appl. 7:609–624. doi:10.1111/eva.12175

Burrell, A.M., A. Sharma, N.Y. Patil, S.D. Collins, W.F. Anderson, W.L. Rooney, et al. 2015. Sequencing of an anthracnose-resistant sorghum genotype and mapping of a major QTL reveal strong candidate genes for anthracnose resistance. Crop Sci. 55:790–799. doi:10.2135/cropsci2014.06.0430

Cao, Y.L., X.H. Ding, M. Cai, J. Zhao, Y.J. Lin, X.H. Li, et al. 2007. Expression pattern of a rice disease resistance gene *Xa3/Xa26* is differentially regulated by the genetic backgrounds and developmental stages that influence its function. Genetics 177:523–533. doi:10.1534/genetics.107.075176

Cardwell, K.F., P.R. Hepperly, and R.A. Frederiksen. 1989. Pathotypes of *Colletotrichum graminicola* and seed transmission of sorghum anthracnose. Plant Dis. 73:255–257. doi:10.1094/PD-73-0255

Casa, A.M., G. Pressoir, P.J. Brown, S.E. Mitchell, W.L. Rooney, M.R. Tuinstra, et al. 2008. Community resources and strategies for association mapping in sorghum. Crop Sci. 48:30–40. doi:10.2135/cropsci2007.02.0080

Casela, C.R., A.S. Ferreira, and R.E. Schaffert. 1992. Physiological races of *Colletotrichum graminicola* in Brazil. In: W.A.J. Milliano, R.A. Frederiksen, and G.D. Bengston, editors, Sorghum and millets diseases: A second world review. ICRISAT, Patancheru, India. p. 209–212.

Chan, A.W., M.T. Hamblin, and J.L. Jannink. 2016. Evaluating imputation algorithms for low-depth genotyping-by-sequencing (GBS) data. PLoS One 11(8): e0160733. doi:10.1371/journal.pone.0160733

Coleman, O.H., and I.E. Stokes. 1954. The inheritance of resistance to stalk red rot in sorghum. Agron. J. 46:61–63. doi:10.2134/agronj1954.00021962004600020002x

Cuevas, H.E., L.K. Prom, and J. Erpelding. 2015. Tapping the US sweet sorghum collection to identify biofuel germplasm. Sugar Tech 17:428–438. doi:10.1007/s12355-014-0349-7

Cuevas, H.E., L.K. Prom, and J.E. Erpelding. 2014a. Inheritance and molecular mapping of anthracnose resistance genes present in sorghum line SC112-14. Mol. Breed. 34:1943–1953. doi:10.1007/s11032-014-0151-y

Cuevas, H.E., L.K. Prom, J.E. Erpelding, and V. Brotons. 2014b. Assessments of genetic diversity and anthracnose disease response among Zimbabwe sorghum germplasm. Plant Breed. 133:234–242. doi:10.1111/pbr.12133

Cuevas, H.E., L.K. Prom, T. Isakeit, and G. Radwan. 2016. Assessment of sorghum germplasm from Burkina Faso and South Africa to identify new sources of resistance to grain mold and anthracnose. Crop Prot. 79:43–50. doi:10.1016/j.cropro.2015.10.007

Cuevas, H.E., G. Rosa-Valentin, C.M. Hayes, L.W. Rooney, and L. Hoffmann. 2017. Genomic characterization of a core set of the USDA-NPGS Ethiopian sorghum germplasm collection: Implications for germplasm conservation, evaluation, and utilization in crop improvement. BMC Genomics 18:108. doi:10.1186/s12864-016-3475-7

Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, et al. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6(5):e19379. doi:10.1371/journal.pone.0019379

Erpelding, J. 2011. Anthracnose field evaluation of sorghum germplasm from Botswana. Plant Prot. Sci. 47:149–156.

Erpelding, J. 2012. Anthracnose resistance in sorghum germplasm from the Segou region of Mali. J. Crop Improv. 26:397–414. doi:10.1080/15427528.2011.650295

Erpelding, J.E., and L.K. Prom. 2004. Evaluation of Malian sorghum germplasm for resistance against anthracnose. Plant Pathol. J. 3:65–71. doi:10.3923/ppj.2004.65.71

Eulgem, T. 2005. Regulation of the *Arabidopsis* defense transcriptome. Trends Plant Sci. 10:71–78. doi:10.1016/j.tplants.2004.12.006

Food and Agriculture Organization. 2017. FAOSTAT. Food and Agriculture Organization. http://faostat.fao.org/site/339/default.aspx (accessed 15 Feb. 2018).

Felderhoff, T. J., McIntyre, L. M., Saballos, A., and Vermerris, W. 2016. Using genotyping by sequencing to map two novel anthracnose resistance loci in Sorghum bicolor. G3 (Bethesda) 6:1935–1946.

Fox, E. J., Reid-Bayliss, K. S., Emond, M. J., and Loeb, L. A. 2014. Accuracy of next generation sequencing platforms. Next Gener. Seq. Appl. 1:1000106.

Godiard, L., L. Sauviac, K.U. Torii, O. Grenon, B. Mangin, N.H. Grimsley, et al. 2003. Erecta, an LRR receptor-like kinase protein controlling development pleiotropically affects resistance to bacterial wilt. Plant J. 36 353–365

Harlan, J.R., and J.M.J. Dewet. 1972. A simplified classification of cultivated sorghum. Crop Sci. 12:172–176. doi:10.2135/cropsci1972.0011183X001200020005x

Hayes, C.M., G.B. Burow, P.J. Brown, C. Thurber, Z.G. Xin, and J.J. Burke. 2015. Natural variation in synthesis and catabolism genes influences dhurrin content in sorghum. Plant Genome 8(2). doi:10.3835/plantgenome2014.09.0048

Jain, M., A. Nijhawan, R. Arora, P. Agarwal, S. Ray, P. Sharma, et al. 2007. F-box proteins in rice. Genome-wide analysis, classification, temporal and spatial gene expression during panicle and seed development, and regulation by light and abiotic stress. Plant Physiol. 143:1467–1483. doi:10.1104/pp.106.091900

Jakobsson, M., and N.A. Rosenberg. 2007. CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801–1806. doi:10.1093/bioinformatics/btm233

Jones, J.D.G., and J.L. Dangl. 2006. The plant immune system. Nature 444:323–329. doi:10.1038/nature05286

Koressaar, T., and M. Remm. 2007. Enhancements and modifications of primer design program Primer3. Bioinformatics 23:1289–1291. doi:10.1093/bioinformatics/btm091

Korte, A., and A. Farlow. 2013. The advantages and limitations of trait analysis with GWAS: A review. Plant Methods 9:29. doi:10.1186/1746-4811-9-29

Lasky, J.R., H.D. Upadhyaya, P. Ramu, S. Deshpande, C.T. Hash, J. Bonnette, et al. 2015. Genome–environment associations in sorghum landraces predict adaptive traits. Sci. Adv. 1:e1400218. doi:10.1126/sciadv.1400218

Levine, A., R. Tenhaken, R. Dixon, and C. Lamb. 1994. H_2O_2 from the oxidative burst orchestrates the plant hypersensitive disease resistance response. Cell 79:583–593. doi:10.1016/0092-8674(94)90544-4

Li, M., X.L. Liu, P. Bradbury, J.M. Yu, Y.M. Zhang, R.J. Todhunter, et al. 2014. Enrichment of statistical power for genome-wide association studies. BMC Biol. 12:73. doi:10.1186/s12915-014-0073-5

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

Marley, P.S., R.P. Thakur, and O. Ajayi. 2001. Variation among foliar isolates of *Colletotrichum sublineolum* of sorghum in Nigeria. Field Crops Res. 69:133–142. doi:10.1016/S0378-4290(00)00128-3

McDonald, B.A., and C. Linde. 2002. Pathogen population genetics, evolutionary potential, and durable resistance. Annu. Rev. Phytopathol. 40:349–379. doi:10.1146/annurev.phyto.40.120501.101443

Mehta, P.J., C.C. Wiltse, W.L. Rooney, S.D. Collins, R.A. Frederiksen, D.E. Hess, et al. 2005. Classification and inheritance of genetic resistance to anthracnose in sorghum. Field Crops Res. 93:1–9. doi:10.1016/j.fcr.2004.09.001

Michelmore, R.W. 2003. The impact zone: Genomics and breeding for durable disease resistance. Curr. Opin. Plant Biol. 6:397–404. doi:10.1016/S1369-5266(03)00067-0

Moore, J.W., M. Ditmore, and D.O. TeBeest. 2008. Pathotypes of *Colletotrichum sublineolum* in Arkansas. Plant Dis. 92:1415–1420. doi:10.1094/PDIS-92-10-1415

Morris, G.P., P. Ramu, S.P. Deshpande, C.T. Hash, T. Shah, H.D. Upadhyaya, et al. 2013a. Population genomic and genome-wide association studies of agroclimatic traits in sorghum. Proc. Natl. Acad. Sci. USA 110:453–458. doi:10.1073/pnas.1215985110

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 (Bethesda):3:2085–2094.

Pande, S., L.K. Mughogho, R. Bandyopadhyay, and R.I. Karunakar. 1991. Variation in pathogenicity and cultural characteristics of sorghum isolates of *Colletotrichum graminicola* in India. Plant Dis. 75:778–783. doi:10.1094/PD-75-0778

Patil, N.Y., R.R. Klein, C.L. Williams, S. Delroy Collins, J.E. Knoll, A.M. Burrell, et al. 2017. Quantitative trait loci associated with anthracnose resistance in sorghum. Crop Sci. 57:877–890. doi:10.2135/cropsci2016.09.0793

Perumal, R., M.A. Menz, P.J. Mehta, S. Katile, L.A. Gutierrez-Rojas, R.R. Klein, et al. 2009. Molecular mapping of *Cg1*, a gene for resistance to anthracnose (*Colletotrichum sublineolum*) in sorghum. Euphytica 165:597–606. doi:10.1007/s10681-008-9791-5

Pritchard, J.K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.

Prom, L.K., J. Erpelding, R. Perumal, T. Isakeit, and H.E. Cuevas. 2012a. Response of sorghum accessions from four African countries against *Colletotrichum sublineolum*, causal agent of sorghum anthracnose. Am. J. Plant Sci. 3:125–129. doi:10.4236/ajps.2012.31014

Prom, L.K., T. Isakeit, R. Perumal, J.E. Erpelding, W. Rooney, et al. 2011. Evaluation of the Ugandan sorghum accessions for grain mold and anthracnose resistance. Crop Prot. 30:566–571. doi:10.1016/j.cropro.2010.12.025

Prom, L.K., R. Perumal, H.E. Cuevas, G. Radwan, S. Katile, T. Isakeit, and C. Magill. 2016. Assessing the vulnerability of sorghum converted lines to anthracnose and downy mildew infection. J. Agric. Crops 2:101–106.

Prom, L.K., R. Perumal, S.R. Erattaimuthu, C.R. Little, E.G. No, J.E. Erpelding, et al. 2012b. Genetic diversity and pathotype determination of *Colletotrichum sublineolum* isolates causing anthracnose in sorghum. Eur. J. Plant Pathol. 133:671–685. doi:10.1007/s10658-012-9946-z

Prom. L.K., R. Perumal, J.E. Erpelding, T. Isakeit, N. Montes-Garcia, and C.W. Magill. 2009. A pictorial technique for mass screening of sorghum germplasm for anthracnose (Colletotrichum sublineolum) resistance. Open Agric. J. 3:20–25. doi:10.2174/1874331500903010020.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M.A.R. Ferreira, D. Bender, et al. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81:559–575. doi:10.1086/519795

Rhodes, D.H., L. Hoffmann, W.L. Rooney, P. Ramu, G.P. Morris, and S. Kresovich. 2014. Genome-wide association study of grain polyphenol concentrations in global sorghum [*Sorghum bicolor* (L.) Moench] germplasm. J. Agric. Food Chem. 62:10916–10927. doi:10.1021/jf503651t

Scheer, J.M., and C.A.J. Ryan. 2002. The systemin receptor SR160 from *Lycopersicon peruvianum* is a member of the LRR receptor kinase family. Proc Natl Acad Sci USA 99: 9585–9590.

Smith, C.W., and R.A. Frederiksen. 2000. Sorghum: Origin, history, technology and production. John Wiley & Sons, New York.

Snowden, J.D. 1936. The cultivated races of sorghum. Adlard and Son, London.

Song, W.Y., G.L. Wang, L.L. Chen, H.S. Kim, L.Y. Pi, T. Holsten, et al. 1995. A receptor kinase-like protein encoded by the rice disease resistance gene *Xa21*. Science 270:1804–1806.

Tang, X., R.D. Frederick, J. Zhou, D.A. Halterman, Y. Jia, and G.B. Martin. 1996. Initiation of plant disease resistance by physical interaction of AvrPto and Pto kinase. Science 274 2060–2063.

Tenkouano, A. 1993. Genetic and ontogenic analysis of anthracnose resistance in *Sorghum bicolor* (L.) Moench. Ph.D. diss., Texas A&M Univ., College Station, TX.

Thakur, R.P., and K. Mathur. 2000. Anthracnose. In: R.A. Frederiksen and G.N. Odvody, editors, Compendium of sorghum diseases. The American Phytopathology Society, St. Paul, MN. p. 10–12.

Thurber, C.S., J.M. Ma, R.H. Higgins, and P.J. Brown. 2013. Retrospective genomic analysis of sorghum adaptation to temperate-zone grain production. Genome Biol. 14(6):R68. doi:10.1186/gb-2013-14-6-r68.

Turner, S.D. 2014. qqman: An R package for visualizing GWAS results using Q-Q and Manhattan plots. bioRxiv. https://www.biorxiv.org/content/early/2014/05/14/005165 (accessed 15 Feb. 2018).

Umemura, K., J. Satou, M. Iwata, N. Uozumi, J. Koga, T. Kawano, et al. 2009. Contribution of salicylic acid glucosyltransferase, OsSGT1, to chemically induced disease resistance in rice plants. Plant J. 57:463–472. doi:10.1111/j.1365-313X.2008.03697.x

Upadhyaya, H.D., Y.H. Wang, R. Sharma, and S. Sharma. 2013. Identification of genetic markers linked to anthracnose resistance in sorghum using association analysis. Theor. Appl. Genet. 126:1649–1657. doi:10.1007/s00122-013-2081-1

USDA. 2014. US Census of Agriculture. USDA. https://agcensus.usda.gov/Publications/2012/ (accessed 2 Mar. 2018).

Valèrio, H.M., M.A. Resende, R.C.B. Weikert-Oliveira, and C.R. Casela. 2005. Virulence and molecular diversity in *Colletotrichum graminicola* from Brazil. Mycopathologia 159:449–459. doi:10.1007/s11046-005-0373-y

van den Burg, H.A., D.I. Tsitsigiannis, O. Rowland, J. Lo, G. Rallapalli, D. MacLean, et al. 2008. The F-box protein ACRE189/ACIF1 regulates cell death and defense responses activated during pathogen recognition in tobacco and tomato. Plant Cell 20:697–719. doi:10.1105/tpc.107.056978

VanRaden, P.M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91:4414–4423. doi:10.3168/jds.2007-0980

Yu, J., M.T. Hamblin, and M.R. Tuinstra. 2013. Association genetics strategies and resources. In: A.H. Paterson, editor, Genomics of the Saccharinae. Springer, New York. p. 187–203. doi:10.1007/978-1-4419-5947-8_9

Zhang, D., Li, J. P., Compton, R. O., Robertson, J., Goff, V. H., Epps, E., et al. (2015). Comparative genetics of seed size traits in divergent cereal lineages represented by sorghum (Panicoidae) and rice (Oryzoidae). G3 (Bethesda) 5:1117–1128.

Zhu, C.S., M. Gore, E.S. Buckler, and J.M. Yu. 2008. Status and prospects of association mapping in plants. Plant Genome 1:5–20. doi:10.3835/plantgenome2008.02.0089