5-2021

Genome-Wide Association Studies of Antimicrobial Activity in Global Sorghum [*Sorghum bicolor* (L.) Moench]

Lindsay Kaye Shields
Clemson University, lindsay.shields4@gmail.com

GENOME-WIDE ASSOCIATION STUDIES OF ANTIMICROBIAL ACTIVITY IN GLOBAL SORGHUM [*SORGHUM BICOLOR* (L.) MOENCH]

A Thesis
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Master of Science
Plant and Environmental Sciences

---

by
Lindsay Kaye Shields
May 2021

Accepted by:
Dr. Stephen Kresovich, Committee Chair
Dr. Richard Boyles
Dr. William Bridges

ABSTRACT

Sorghum is a common feed grain globally with vast genetic and phytochemical diversity that may provide numerous health benefits, including its aptitude as an antimicrobial feed grain. This study highlights the antimicrobial potential of a collection of 384 diverse sorghum accessions against two prominent foodborne pathogens, *Clostridium perfringens* and *Salmonella enterica*. Following extensive screening, we determined that sorghum grain extract is more efficient at inhibiting *C. perfringens* than *S. enterica*. Antimicrobial activity observed against *C. perfringens* was not significantly correlated with either total phenols ($r = 0.12$) or tannin concentration ($r = 0.12$). Moreover, we mapped loci associated with antimicrobial activity to *C. perfringens* that are independent of loci associated with total phenols and tannins. The two most significant associations were determined to have an epistatic interaction and a total of 20 candidate genes were identified. By sequence homology studies we found the potential functions of these candidates to include plant stress response (*Sobic.002G083600*) and phenol metabolism regulation (*Sobic.010G222600).* Additionally, we noted no relationship between antimicrobial activity and either grain yield or composition. These results highlight significant heritable variation of antimicrobial activity in sorghum that may be useful for breeding to improve its value as a feed source by incorporating grain-based antibiotics in animal production.

ACKNOWLEDGMENTS

I would like to acknowledge that this work was supported by Agribusiness Center for Research and Entrepreneurship (ACRE), the United States Department of Agriculture's National Institute of Food and Agriculture graduate student fellowship grant (USDA- NIFA) (SC-2015-10844), and Robert and Lois Coker Endowment.

I would also like to acknowledge the members of the Kresovich group, namely, Dr. Lucas Boatwright and Kathleen Jordan, for their constant help and patience. Finally, I would like to thank my committee and those that have taken the time to revise and provide comments for this work.

TABLE OF CONTENTS

Page

Table of Contents (Continued) Page

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

Animal agriculture contributes to the global public health concern of antibiotic resistance, as the industry uses 73% of global distribution of clinical antibiotics to treat bacterial infection in animals raised for food (White et al., 2002; Matthew et al., 2007; Van Boeckel et al., 2019). Facilities with poor sanitation and loose veterinary regulation enable the spread of harmful pathogens and are overcome with copious use of antibiotics (Van Boeckel et al., 2019). *Clostridium perfringens* and *Salmonella enterica* are two prominent foodborne pathogens that regularly threaten the poultry industry and are traditionally treated with antibiotics. *C. perfringens* causes inflammation in the gut of broilers accounting for 1% of losses a day (Van Immerseel et al., 2005), whereas *S. enterica* is a threat to both human and animal health as it repeatedly contaminates animal products (Swartz, 2002; White et al., 2002). Prolific spread of these pathogens impairs the efficiency of the farm and if untreated, are detrimental to farmers. Antibiotic usage in agriculture could be reduced by implementing alternative solutions, such as incorporating feed products with antimicrobial properties into animal rations (Gyawali and Ibrahim, 2014).

Sorghum [*Sorghum bicolor* (L.) Moench] is a cereal grain used most commonly for feed, food, forage and bioenergy. Though its uses are diverse, sorghum grain is economically prominent as a staple food supply in Africa and Asia but is most prevalently used as a feed grain for livestock in the United States. Sorghum was first domesticated in the Horn of Africa, where subsequent migration events and adaptation of early domesticates led to the establishment of the five major genetically distinct races within the sorghum species (Harlan & De Wet, 1972; Brown et al., 2011). The racial structure within sorghum largely contributes to the genetic and phenotypic diversity within the species; however, because of its photoperiod sensitivity (it

requires a short daylength to flower and produce grain), much of the available germplasm in the U.S. National Plant Germplasm System has yet to be exploited for crop improvement in temperate regions. For instance, nutritional and health related traits have traditionally been underexploited in breeding programs although the genetic potential may exist. To identify the genes that underlie these quantitative traits, the Sorghum Association Panel (SAP) is a genetic resource designed to be used for association mapping studies. The SAP has been created from a diverse collection of accessions that represents the five major cultivated races, geographic centers of diversity, and important United States breeding lines (Casa et al., 2008).

Sorghum grains are phytochemically rich with phenols, which have beneficial antioxidant (Herald et al., 2012), anti-cancer (Hargrove, 2011), and anti-inflammatory properties (Burdette et al., 2010; Rhodes and Kresovich, 2016). Phenols are ubiquitously found throughout the plant kingdom as secondary metabolites produced in response to biotic and abiotic stresses. Because of their wide range in structural differences and diversity within the class, phenols contribute to several physiological processes and traits. One of the more prominent phenolic-based traits studied in plants is the antimicrobial effect (Cowan, 1999; Nitiemea et al., 2012; Alzoreky & Nakahara, 2001; Kil et al., 2009). Plant phenols exhibit antimicrobial activity against a number of bacteria, both Gram-negative and Gram-positive, and maintain these inhibitory effects *in vitro (*Nitiema et al., 2012). Furthermore, antimicrobial activity against foodborne pathogens has previously been observed in sorghum through evaluation of metabolite extractions (Kil et al. 2009). However, these tests were limited to a small collection of genotypes that were not representative of global diversity, and the underlying genetics and mechanisms of antimicrobial activity were not considered.

Importantly, not all phenol subclasses may be beneficial for use as a feed grain. Tannins are a broad class of phenols, divided into three subclasses, that are recognized for their antimicrobial effects (Scalbert, 1991). Specifically, the subclass of proanthocyanidins, or condensed tannins, are the most prominent in sorghum grain. Condensed tannins localized in the testa, a layer of tissue that is located between the pericarp and endosperm of a grain, which can, along with other polyphenols, give pigment to the testa layer (Earp and Rooney, 1982). The presence of condensed tannins in sorghum grain is modulated by two loci, B_1 and B_2 (Dykes & Rooney, 2005), whose underlying genes have now been identified as *Tannin1* (*Tan1*) and *Tannin2*, respectively (Wu et al., 2012; Wu et al., 2019).

Though tannins are found throughout the plant kingdom and are readily available antimicrobial agents, tannins are not a solution to the current problem of antibiotic resistance. Tannins bind to a variety of nutrients and impair digestion, thus reducing the bioavailability of essential nutrients and nutrient efficiency (Chung et al., 1998). For this reason, tannins have largely been eliminated in common cereal grains such as wheat (*Triticum aestivum* L.), corn (*Zea mays* L.), and rice (*Oryza sativa* L.). Sorghum, however, has maintained non-tannin and tannin cultivars over time because tannin types grown by African (Wu et al., 2019) and South American farmers provided protection against severe bird predation. Despite recent findings that suggest tannins in feed grain may replace antibiotics in poultry production as growth promoters (Redondo et al. 2014; Huang et al. 2018), the threshold at which tannins provide beneficial versus adverse effects for the animal are still unknown and therefore breeding for low tannin cultivars remains a priority.

Development of nutritionally efficient, natural food products with antimicrobial activity may combat the continuing rise in antibiotic resistance in animal agriculture. The genetic and

metabolite diversity maintained in sorghum germplasm provides support that it may be a promising candidate for this use. Therefore, the goals of this study were: (a) to characterize antimicrobial activity across a globally diverse collection of sorghum germplasm, and (b) to identify the genetic basis associated with antimicrobial activity that is unrelated to the anti-nutritional effects of tannins. From these experiments, we found significant variation of antimicrobial activity that exists in sorghum grain independent of antinutritional components such as tannins and total phenols. This research also highlights that antimicrobial effects did not have a negative impact on yield or grain macro- and micronutrient composition. Twenty potential candidate genes were identified through the results of the genome-wide association studies which may regulate antimicrobial activity, further supporting sorghums potential as an antimicrobial feed grain.

2. MATERIALS AND METHODS

Plant material and field design

The plant material evaluated was a subset of 384 accessions, representing the sorghum association panel (SAP) (Casa et al. 2008; Boyles et al. 2016) (https://www.ars-grin.gov/npgs/). Materials were grown and sampled during the 2017 field season. The SAP was planted in a randomized complete block design with two replications at the Clemson University Pee Dee Research and Education Center in Florence, South Carolina. Plots contained two rows, 6.1 m in length and spaced 0.726 m apart with an average planting density of approximately 62,350 plants ha^{-1}. Blocking decisions were based on both maturity and plant height, with full details described in Sapkota et al. 2020. Fields were irrigated when needed and adequate nutrients were supplied

to minimize abiotic stress. In detail, variable rates of N, P, and K fertilizer applications were applied prior to planting based on soil samples, followed by an application of 93 kg ha^{-1} of nitrogen 35 days after planting. Bicep II Magnum (S-metolachlor + atrazine; Syngenta) was applied prior to planting at 3.5 L ha^{-1}. Atrazine was subsequently applied at 4.7 L ha^{-1} post emergence. To control the sugarcane aphid population, a single application of 0.5 L ha^{-1} of Sivanto^{TM} Prime (Bayer CropScience) occurred 60 days after planting (Sapkota et al., 2020). Grain was collected from the primary panicle when the plant reached physiological maturity. Harvesting of grain from the secondary panicles, located on the tillers of the sorghum plant, was avoided to prevent confounding effects of maturity on grain composition. Harvested panicles were dried from 10-14 days in an electric dryer to a constant weight and subsequently threshed using a BT-14 belt thresher (Almaco; Nevada, IA). Maximum forced air was used when threshing to remove all glumes, foreign plant debris, and poorly filled or damaged grains.

Compositional analysis

Compositional data were collected from each genotype by near-infrared spectroscopy (NIRS), performed with a DA7250 NIR analyzer (Perten Instruments). Dried and threshed grained were ground to a particle size of 1 mm with a Cyclotec sample mill (FOSS; Hillerod, Denmark) and used to evenly fill a 43 mL Teflon dish. Ground samples, as opposed to whole-kernel samples, have been reported to get the most efficient measurements (de Alencar Figueirido et al. 2010). The Teflon dish containing the ground sample was gradually rotated during NIRS analysis for accurate sampling. NIRS data were recorded for 29 compositional traits for each sample, including key macronutrients such as starch, protein, and crude fat (ether extracted lipids). A full list of compositional traits can be found in supplemental data **(Supplemental F1**). Trait

calibrations were previously established using a subset of 100 samples in the SAP (Boyles et al. 2017). Harvested panicles were air dried to a constant moisture and hand threshed. The dried, ground samples were sent to Dairyland Laboratories, Inc (Arcadia, WI) and the Quality Assurance Laboratory in Murphy-Brown, LLC (Warshaw, NC) for wet chemistry. Calibration curves were established with a DA7250 NIR analyzer (Perten Instruments).

Extraction of metabolites

Following NIRS, metabolites were extracted from ground sorghum using an acetone extraction method described in Herald et al. (2012). For each sample, metabolites were extracted by adding 10 mL of 70% acetone to 0.5 g of representative ground grain and agitated for 2 h. Samples were stored at -20°C overnight. The next day, samples were centrifuged at 2970 x g, 10 mi, 4 °C and the supernatant was transferred to new tubes. An additional round of extraction was performed on the existing tissue by adding another 10 mL of 70% acetone. Samples were agitated for 10 minutes and centrifuged (2970 x g, 10 min, 4°C). The supernatant was collected and combined with the previously collected supernatant. Acetone was removed from the extract using nitrogen evaporation with a 96-well microtiter microvaps (Fisher Scientific; Pittsburgh, PA). For long-term storage, extracts were resuspended in 1 mL of DMSO and stored in the dark at -20°C.

Quantification of polyphenols and tannins

Total phenols were quantified using the Folin-Ciocalteu Assay (Singleton, Orthofer, and Lamuela-Raventos, 1999). A standard curve was established using gallic acid concentrations ranging from (12.5 - 400 μg/mL in 70% acetone), following the protocol outlined in Rhodes et al. (2017). In individual wells of a 96-well plate, 75 μL of Deionized water, 25 μL of Folin-

Ciocalteu reagent (diluted 1:1 with deionized water) was mixed with either 25 µL of extract, standard, or 70% acetone, and left for six minutes for the reaction to complete. Subsequently, 100 µL of 7.5% sodium carbonate was added to each well and mixed, covered and left in the dark for 90 minutes. Absorbance at 765 nm was measured using a Synergy H1 Multi-Mode Microplate Reader (BioTek Instruments, Inc.; Winooski, VT). Twenty-five microliters of 70% acetone were used as a control. Total phenol concentrations are reported in [gallic acid equivalent (GAE)/g] based on dry weight.

Tannin data were kindly provided by Dr. Davina Rhodes (Rhodes et al., 2014). To generate these data, the SAP was planted and harvested at Clemson University Pee Dee Research and Education Center in Florence, SC in 2013 and 2014, under the same field conditions and management described in Section 2.1. Tannin concentrations were collected using NIRS, for which 20g of whole grain samples were scanned with a FOSS XDS spectrometer (FOSS North America, Eden Prairie, MN, USA) at a wavelength range from 400 - 2500 nm. Each sample was measured in duplicate, based on dry weight, and the reported tannin concentration (mg CE/g) represented the mean of duplicates. The calibrations curves, software, and spectrometer used were all previously described in Dykes et al. (2014).

Average tannin concentration, for each year, was used in a t-test and Pearson's correlation. A t-test was performed to determine if the two years of tannin data were statistically different. Previous studies have found that the tannin trait had a high broad-sense heritability estimate (H^2 = 0.80).

Disc-diffusion antimicrobial assay

Antimicrobial susceptibility testing for sorghum grain extracts against *C. perfringens* (CP#6; Miller, Skinner, Sulakvekidze, Mathis, & Hofacre, 2010) and *S. enterica* (ATCC 30661) was performed using a disc-diffusion assay following Clinical and Laboratory Standards Institute guidelines (Clinical and Laboratory Standards Institute, 2012). Samples were prepared by saturating cotton discs 6-mm in diameter (Becton, Dickenson and Company) with 20 µL of extracts, twice, allowing sufficient time for the discs to dry in between saturations. All samples were prepared in triplicate. Each pathogen was appropriately prepared for the assay by following optimal culturing methods described by Chen and Jiang (2017) for *S. enterica* and Dharmasena and Jiang (2018) for *C. perfringens*. Cultures of each pathogen were washed twice and resuspended in 0.85% sterile saline to OD 0.5. Mueller-Hinton (*S. enterica*) and Brucella blood (*C. perfringens*) agar plates were inoculated and streaked to ensure an evenly distributed lawn of growth. Six saturated discs were placed equidistantly on the surface of the plate, ensuring the disc lay completely flat. Two additional discs were included in each assay plate: a positive control with 30 ug per disc of kanamycin (*S. enterica*) or tetracycline (*C. perfringens*), and a disc saturated with 30 ug per disc of DMSO as a negative control. Plates were inverted and incubated at 35°C for 16 to 18 h (*S. enterica*) or 24 h (*C. perfringens*). The diameter was measured to the nearest tenth of a millimeter for the zone of inhibition. The zone of inhibition for each sample was compared to the inhibition with a standard microorganism, *Escherichia coli* (ATCC 25922). Samples that had clear inhibition zones or zones with a diameter greater than 7 mm (**Figure 1D**) were considered to have a strong effect. However, samples that showed detectable inhibition yet maintained some bacterial growth in the inhibition zone were classified as having weak effect (**Figure 1A-C**).

Minimum inhibitory concentration assay

To confirm the results of the disc-diffusion assay, a subset of sorghum extracts were tested by using the microbroth dilution method following Clinical and Laboratory Standards Institute guidelines (Clinical and Laboratory Standards Institute, 2012). Briefly, Brucella broth and Mueller-Hinton broth were inoculated with *C. perfringens* and *S. enterica*, respectively, and incubated at 37°C with shaking anaerobically (*C. perfringens*) or aerobically (*S. enterica*) to establish logarithmic growth (Chen & Jiang, 2017; Dharmasena & Jiang, 2018). Following incubation, each culture was pelleted by centrifugation (3,500 x g for 5 min) and resuspended in 0.85% sterile saline solution to an OD of 0.1 measured at 625 nm. Samples were tested in triplicate using 96-well microplates, yielding final bacterial concentrations of 5×10^5 CFU/ml, and incubated overnight at 37°C. Following incubation, OD of each well were determined with a μQuant microplate spectrophotometer (BioTek Instruments, Inc.; Winooski, VT) at 625 nm. *Escherichia coli* (ATCC 25922) and *Clostridium difficile* (ATCC 700057) were used as quality control strains for *S. enterica* serotype *enteritidis* and *C. perfringens*, respectively. Minimum inhibition concentration (MIC) was defined as the lowest concentration needed to completely inhibit growth as compared with the controls.

Genomic analysis

The genotypes of the SAP were collected by means of genotype-by-sequencing (GBS) (Morris et al., 2013; Boyles et al., 2016). Raw sequence reads were aligned to the most current sorghum reference genome (BTx623 v3.1, https://phytozome.jgi.doe.gov) using Burrow-Wheelers aligner (Li and Durbin, 2010). SNP calling was done using the TASSEL 5.0 pipeline (Glaubitz et al.,

2014), and subsequent imputation was performed using the FILLINFindHaplotypesPlugin and FILLINImputationPlugin in TASSEL (Swarts et al., 2014). A total of 484,799 SNPs was generated. SNPs were filtered for minor allele frequency (>0.05), missing data (0.30) and Hardy-Weinberg equilibrium (0) using VCFtools (Danecek et al., 2011). Subsequently, the marker set was pruned, using PLINK (Purcell et al. 2007). The PLINK pruning method is based on variance inflation factor (VIF), which recursively removes SNPs above the VIF threshold (VIF =2) within a sliding 50 SNP window, shifting steps at every five SNPs. The VIF is defined as $1/(1-R^2)$, with R^2 being the multiple correlation coefficient for the SNP being tested against all other SNPs. Therefore, VIF accounts for the multicollinearity in the linear regression (Purcell et al., 2007). Filtering and pruning resulted in 99,126 SNP markers that were used for association mapping analysis.

GEMMA Version (v0.98, Zhou and Stephens, 2014) software was used to perform the genome-wide association study (GWAS) on antimicrobial activity against *C. perfringens* by implementing a univariate linear mixed model (LMM), which takes population stratification and sample structure into account. To genetically confirm antimicrobial activity's independence from *Tan1*, best linear unbiased predictors (BLUPs) were calculated for tannin concentrations and used as a covariate in a univariate linear mixed model. The use of BLUPs instead of mean values accounts for environmental variation in the tannin measurements across the two years. Manhattan and quantile-quantile plots were generated using R software CMplot (https://github.com/YinLiLin/R-CMplot). The Bonferroni-corrected significance threshold ($0.05/99{,}126$ SNPs $= 5.04 \times 10^{-7}$) was used to determine significant associations ($\alpha = 0.05$) in the Manhattan plot. Linkage disequilibrium (LD) was calculated locally within 1 MB of significantly

associated SNPs using PLINK. Linkage disequilibrium was considered to decay at $r^2 < 0.1$. Genes found within local LD of each SNP were identified using a custom script and considered a potential candidate gene for antimicrobial activity in sorghum grain. Broad-sense heritability was calculated by R package 'Heritability' (Kruijer et al., 2015). Marker-based narrow-sense heritability (h^2) was calculated using the relatedness matrix generated from GEMMA, which utilized the same SNPs as the GWAS. All scripts used for analysis are located at (https://github.com/lkshiel/ACRE)

Epistatic interaction analysis

We performed SNP-SNP interaction tests were performed on all SNPs. PLINK files generated for the LD analysis were used to test each SNP pair for epistasis using the following linear regression model implemented with PLINK --epistasis command and --set parameter.

$$Y = \beta_0 + \beta_1 g_A + \beta_2 g_B + \beta_3 g_A g_B$$

where Y represents the antimicrobial activity as measured in the disc-diffusion assay, g_A and g_B are allele counts for each inspected variant pair, and β coefficients 0-3 represent the intercept, effect of g_A, effect of g_B and the epistatic interaction between the variant pair, respectively. Frequencies of the two locus genotypes were also calculated using PLINK *twolocus function*. Analysis was performed on un-imputed data so as to not skew frequency of a particular genotype.

Statistical analysis

Pearson's correlation coefficient was determined for all relationships investigated in this study. Statistical significance was determined for p-values less than 0.05. Categorical values such as

'weak' were given numerical value of 3.55, following the method described in Billard & Diday (2000). This numerical value was determined by averaging the lowest measured inhibition zone (0 mm) and the lowest 'strong' value (7.1 mm). Folin-Ciocalteu samples that were measured above 420 [GAE/g] were capped at 420 [GAE/g] for analysis. The t-test was performed using the Python package *Pingouin* (Vallat, 2018). Tables were generated with the Python package *pandas* (McKinney, 2010) and figures were generated with the python packages *seaborn* and *matplotlib* (Waskom et al. 2014; Hunter, 2007).

3. RESULTS

Antimicrobial activity of sorghum grain extracts

Approximately half of the tested germplasm (188 accessions) were found to demonstrate antimicrobial activity, ranging from weak to strong effect (**Supplemental F2**). Weak activity was defined as having an indistinguishable inhibition zone while still showing inhibition; whereas, strong effect was defined as having a clear inhibition zone or a zone with a diameter greater than 7 mm. Using the disc-diffusion assay, we identified 103 accessions that had antimicrobial activity against *C. perfringens*. Of which, 37 of which had a strong effect with inhibition zones that ranged from 6.88 to 9.45 mm in diameter. Assays tested on *S. enterica* showed 119 accessions with antimicrobial activity, however, only a weak effect was observed across those genotypes.

Minimum inhibitory concentration analysis

Minimum inhibitory concentrations (MIC) were determined for a subset of 12 accessions that were selected on the basis of a range of activity levels observed in the disc diffusion assay,

including both high and low extremes. The average MIC values for *C. perfringens* were 0.911 mg/mL for the first replicate and 1.37 mg/mL for the second. However, the average MIC values for *S. enterica* were reported as 32.8 mg/mL in the first replication, and 24.8 mg/mL in the second (**Table 1**). The MIC values observed confirm the results of the disc-diffusion assay. We expected extracts with strong antimicrobial activity to require smaller concentration to inhibit growth than extracts with weaker or no activity. The MIC analysis of *S. enterica* revealed instances with high MIC values, but relatively strong antimicrobial activity. We would expect that genotypes demonstrating antimicrobial activity would have lower MIC values than genotypes that displayed no activity. High MIC values that were found to have weak antimicrobial activity in the disc-diffusion assay could be attributed to the diffusion patterns of the extract on the agar media, which impacts the role of these compounds in evaluations of antimicrobial potential (Alzoreky & Nakahara, 2003). However, because the initial characterization of *S. enterica* inhibition did not provide reliable measurements of antimicrobial activity, and because antimicrobial activity against *S. enterica* was found to be correlated with unfavorable traits (**Supplemental Figure S1, S2, Supplemental Table S1**), data regarding *S. enterica* assays were subsequently excluded from further analyses used to identify useful germplasm for crop improvement.

Quantification of total phenol concentration

Phenol concentrations were measured for each sample via a Folin-Ciocalteu phenolic assay. It was found that across samples, total phenol concentration (GAE/g) was highly variable across the diverse sorghum accessions, ranging from 5.38 GAE/g to 420 GAE/g, where 420 GAE/g was the highest value the instrument could measure on the basis of established standards

(**Supplemental F3**). Total phenol concentration across field replications were found to be highly correlated *(r* = 0.82) regardless of soil differences, pest pressure, and weather events. This is consistent with previous reports as total phenols were previously found to be highly heritable (H^2 = 0.82 (Pfeiffer & Rooney, 2016).

Effects of total phenols and tannin concentration on food-borne pathogens

The relationship between total phenol concentration and inhibition zone size from the disc diffusion assay was investigated to determine if higher total concentrations of phenols were correlated with greater antimicrobial activity. Samples showing positive inhibition of *C. perfringens* had striking variation in total phenol concentration, with a range of 6.91 GAE/g to >420 GAE/g. As a result, *C. perfringens* inhibition was not significantly correlated (*r* = - 0.12) with total phenol concentration at the 0.01 significance level (**Figure 2a**).

The results of the t-test showed that the 2 yr of tannin data were not statistically different (p-value = 0.34) and therefore provide support for using the 2013 and 2014 tannin data for our analysis with grain extracts from 2017 (**Supplemental Table S2**).

Additionally, the correlation between the average inhibition zone diameter (mm) for each accession and average tannin measurement across year was calculated. The inhibition zone diameter did not have a significant correlation (*r* = 0.12) with tannin measurements at the alpha = 0.01 significance level (**Figure 2b**).

Identification of antimicrobial germplasm with unpigmented testa

Testa presence was previously identified across the SAP by cutting a thin layer of the pericarp from each seed and examining testa pigmentation under a dissecting microscope (Rhodes et al., 2014). Pigmented testa may be indicative of the presence of condensed tannins, therefore investigation of the relationship between antimicrobial activity and pigmented testa is warranted. There were no observable differences between the distributions of accessions with unpigmented testa and pigmented testa across the *C. perfringens* inhibition zones (**Supplemental Figure S3**). The similar distributions and the low correlation between inhibition zone and tannin concentration suggests that antimicrobial activity could be achieved in the absence of tannins for *C. perfringens* (**Table 2**).

Selection of suitable accessions containing antimicrobial activity

To identify germplasm that maintain antimicrobial activity without condensed tannins, we rigorously filtered genotypes by phenotypic values. First, from the list of accessions that demonstrated antimicrobial activity, only the accessions that maintained strong antimicrobial activity across both replicates were considered. Next, accessions with detectable tannins were eliminated from consideration. Filtering resulted in five accessions characterized as having strong antimicrobial activity and little or no detectable tannin concentration (**Table 3**). Key agronomic traits, such as plant height (PH), days to maturity (DTM), and 1000-grain weight (TGW), were also evaluated to determine if that agronomic and compositional phenotypes of the identified accessions were undesirable for plant breeding. The agronomic and compositional traits for the accessions in the SAP were reported by Sapkota et al. (2020).

Compositional and yield data of the grain sorghum accessions

Near-infrared spectroscopy was used to measure 29 compositional traits spanning prevalent macronutrients across the 384 accessions with two replications. The quantitative variation in the grain compositional traits is evident in the descriptive statistics of each trait (**Supplemental Table S3**). Most importantly, the compositional data generated establishes that grain macro- and micronutrient composition were not compromised by the presence of antimicrobial activity. No significant relationships were found between individual compositional traits and antimicrobial activity (**Supplemental Table S3**). Additionally, to examine any potential trade-offs resulting from antimicrobial activity on grain yield components, we compared antimicrobial activity of grains to the grain number per panicle, 1000-grain weight and grain yield per primary panicle (Sapkota et al., 2020). No significant correlation was observed between antimicrobial activity against *C. perfringens* and any of the grain yield component traits (**Table S4**), suggesting that the presence of antimicrobial activity does not compromise yield. Additionally, we looked at the geographical and racial distribution of genotypes exhibiting antimicrobial activity against *C. perfringens* to identify any potential correlation with antimicrobial activity, though no relevant statistical inferences could be made concerning the role of race and/or origin.

Genome-wide association studies

To investigate the underlying genetics of antimicrobial activity against *C. perfringens* in sorghum grain, three GWAS were conducted. First, a univariate linear mixed model was used to map antimicrobial activity against *C. perfringens* (**Figure 3, Supplemental Figure S4**). However, since the effect of *Tan1* is so strong and may impact the GWAS, tannin BLUPs were used as a covariate for a second association analysis (**Figure 4, Supplemental Figure S5**). Comparisons between the two antimicrobial GWAS show that by adding tannin as a covariate to

the model, the same four SNPs on chromosomes 2, 4, and 10 were found to be significant across both models (**Supplemental Table S5**). The only difference was the singular association on chromosome 5 in the tannin covariate model, which failed to pass the Bonferroni-corrected significance threshold. To confirm that there was no statistical difference between the antimicrobial activity GWAS and tannin covariate GWAS, we performed a *t-test* comparing each SNP's Wald *p*-values from both GWAS. The *t*-test showed that the two models were not statistically different (**Supplemental Table S6**; *p* value = 0.227), and concerns regarding the effect of *Tan1* were disregarded. Additionally, tannin BLUPs were mapped as a genetic control and used to distinguish tannin peaks from novel peaks associated with antimicrobial activity (**Figure 5, Supplemental Figure S6**). The tannin GWAS contained a peak on chromosome 4, residing at ~62.3 MB, approximately 1 KB from the start position of *Tan1 (Sobic.004G280800;* Wu et al., 2012) (**Figure 5**). Significant associations found on chromosome 4 in the antimicrobial activity GWAS were located at ~64 MB and were not found in LD with *Tan1*. The associations on chromosome 2 (~ 8.9 MB) and chromosome 10 (~56 MB), however, are unique to antimicrobial activity and were the most significant loci.

Epistatic interactions

Epistatic interaction was only found between two of the four SNPs, S2_8924006 and S10_56476103 (p value = 1.93×10^{-6}), which were the two most significant SNPs in the antimicrobial activity association analysis. Joint and marginal counts, and frequencies of the two locus genotypes are shown in **Supplemental Table S7**. Moreover, we plotted the distributions of accessions with both, just one, and neither favorable allele in regard to inhibition zone. Accessions with missing data for at least one allele were removed, therefore n =184 accession

were used for this analysis. We observed that accessions with favorable alleles at both S2_8924006 (C/C) and S10_56476103 (T/T) had the larger median inhibition zone across four accessions (**Supplemental Figure S7)**. Accessions that had only the S10_56476103 allele maintained weak inhibition with a median inhibition zone at approximately 1 mm across six accessions. Whereas accessions that had only the S2_8924006 allele remained at no inhibitory effects with the exception of two outliers across ten accessions. Meanwhile, accessions that had neither of the favorable alleles largely had a no measured inhibitions across the 164 accessions with the exception of the 12 outliers which demonstrated a wide range of inhibition zones from 0 – 8.55 mm.

Heritability and potential candidate genes

On the basis of local LD estimates, 20 genes were identified around significant SNPs that putatively regulated antimicrobial activity (**Supplemental Table S8**). Six potential candidate genes were within the LD block containing the SNP on chromosome 2. Four of the candidate genes were within local LD of S4_64038743 marker on chromosome 4; six genes were within LD of the S4_64439967 marker. The remaining four potential candidate genes were found within local LD of the marker on chromosome 10. Marker-based narrow-sense heritability was calculated to be h^2 =0.55.

4. DISCUSSION

Food-borne pathogens that infect livestock during production and post-processing, impact the productivity and efficiency of the animal protein industry as well as human health. These food-borne pathogens are frequently controlled with antibiotics which, when used in excess, may

lead to the evolution of resistance in pathogen populations (McEwan and Fedorka-Cray, 2002). Antibiotic resistance is a global issue that may be reduced through the supplement of natural products with antibacterial agency (Gyawali and Ibrahim, 2014). Several plant species have been identified that contain secondary metabolites demonstrating antimicrobial activity (Cowan, 1999). However, these traits have yet to be integrated into feed grain breeding programs to produce improved cultivars that would minimize antibiotic usage in animal production. Our identification of loci significantly associated with antimicrobial activity against *C. perfringens* is the first step toward the incorporation of natural health-promoting compounds for sorghum improvement.

In this study, we tested the inhibitory properties of metabolite extracts from 384 diverse accessions of sorghum grain against two prominent foodborne pathogens, *C. perfringens* and *S. enterica.* Through a combination of disc-diffusion and microbroth-dilution assays, 188 unique lines were identified as having antimicrobial activity, constituting half of the experimental germplasm. The significant number of accessions with varying antimicrobial capabilities shows that vast genetic potential exists for antimicrobial activity in sorghum. Additionally, the results of the disc-diffusion assay, later confirmed with the MIC analysis on selected accessions, demonstrate that sorghum grain metabolite extract is a more effective antimicrobial agent against *C. perfringens* than against *S. enterica*.

The influence of total phenol concentration on antimicrobial effect against *C. perfringens* was investigated and found to be insignificant. The apparent lack of a relationship between *C. perfringens* inhibition and total phenol concentration suggests that total phenol concentration alone cannot predict the antimicrobial effect in sorghum. However, the metabolite responsible for inhibiting *C. perfringens* may be an individual subclass of phenols, such as flavan-3-ols,

which has previously been reported for its antimicrobial effect on *C. perfringens* (Daglia, 2012). Relationships between antimicrobial effects on *C. perfringens* and the subclass of phenols could not be determined precisley from the data collected. Moreover, total phenols were extracted using acetone, a method that is more prominently used to extract flavanols and other phenols with higher molecular weight (Dai and Mumper, 2010). Therefore, our correlation measures between total phenol and antimicrobial activity may not include effects from phenols with a lower molecular weight phenol.

Importantly, we noted that the presence of antimicrobial activity against *C. perfringens* was not significantly correlated with tannin concentration, and we subsequently we identified five sorghum accessions that upheld strong antimicrobial activity against *C. perfringens* that did not have a pigmented testa and quantifiable tannins. Condensed tannins are a long-established antimicrobial; however, their negative effects on nutrient efficiency have prevented their use in crop improvement (Scalbert, 1991). In the same regard, there is often a carbon utilization tradeoff between crop yield and metabolite production (Brown, 2002). However, our analysis showed there were no significant correlations found between antimicrobial activity and either compositional traits or yield.

This study used GWAS to identify novel genetic associations with sorghum grain antimicrobial activity against *C. perfringens*. We successfully identified significant associations that are independent of tannin and total phenols. As noted previously, antimicrobial activity has been associated with condensed tannins, which are regulated by duplicate recessive epistatic gene interaction between *Tan1* and *Tan2* located on chromosomes 4 and 2, respectively (Wu et al., 2012; Wu et al., 2019). Similarly, total phenols also have been associated with antimicrobial activity and were found to be associated with loci on chromosome 2 in Rhodes et al (2017).

However, the loci we identified on chromosomes 2 and 4 in the antimicrobial activity GWAS were different from the loci reported for *Tan1*, *Tan2*, and total phenols. Total phenol association was found at 7.5Mb (Rhodes et al. 2017) and *Tan2* (*Sobic.02G076600*) resides nearby at ~8.2 MB. Meanwhile, the antimicrobial activity SNP was located at ~8.9 MB. Moreover, the significant associations identified on chromosome 4 are located at ~64 MB while *Tan1* resides at ~62.3 MB (Wu et al., 2016). Loci significantly associated with antimicrobial activity were not found in LD with either *Tannin* genes or total phenols, further supporting that factors other than tannins or phenols facilitate antimicrobial activity in sorghum grain. Moreover, we accounted for the effect tannin might have on our association analysis by including a tannin covariate. However, the antimicrobial activity GWAS with and without tannin as a covariate were not statistically different, which supports the results from the correlation analysis and suggest that the effects of *Tan1* and *Tan2* did not impact our mapping *of C. perfringens* antimicrobial activity. Thus, we concluded that our significant loci were neither artifacts nor artificially inflated by tannin-related effects.

The two most significant associations positioned on chromosomes 2 and 10, were also determined to have an epistatic interaction. This epistatic interaction suggests that their multiplicative effects on antimicrobial activity arise from a nonlinear combination of allele presence, as shown through the results of the *twolocus function* analysis. Moreover, SNP interactions indicate that the loci may interact through intermediate loci in the metabolic pathways, which further confirms the genetic complexity of antimicrobial activity. Furthermore, mean antimicrobial activity was calculated among allele combinations, which demonstrated that accessions that were homozygous for favorable alleles at both S2_8924006 and S10_56476103 had the highest average inhibition zone (**Supplemental Figure S7**). This provides additional

evidence that S2_8924006 and S10_56476103 are important for regulating antimicrobial activity against *C. perfringens* in sorghum grain. There was substantial amount of missing data for SNP S10_56476103 preventing meaningful interpretation of the type of interaction occurring between the two loci. SNP markers were called from GBS data which are well-known for providing cost-effect method for genotyping, because this method is dependent on the frequency of the ApeK1 recognition site, some areas of the genome receive a low depth of coverage. This may explain the abundant missing data for S10_56476103 as well as its low minor allele frequency (MAF). Better sequence coverage to fill the missing SNP calls and subsequent analysis are needed to better understand how the interaction between these two loci regulate antimicrobial activity in sorghum.

From these genomic analyses, we were able to extract several potential candidate genes. The genes have yet to be characterized, however orthologs identified by OrthoDb (Kriventseva et al. 2018) were found in corn and rice, which may provide insights into the potential function of these genes. For instance, orthologs from maize and rice suggest that *Sobic.002G083200* is a zinc transport protein. Zinc ions have been found to contribute to antimicrobial activity and are commonly added to food as a preservative for their antimicrobial effect in food packaging in and material science (Espita et al.2012; Stanic et al., 2010). Moreover, gene *Sobic.002G083600*, was found to be an ortholog to *sid1*, which has been characterized to play a role in inflorescence architecture in maize (Chuck et al., 2008). However, similarity between *Sobic.002G083600* and *sid1* was due, in part, to the presence of the AP2/ERF domain which characterizes a family of transcription factors that has been found to be key regulators for various stress responses (Xie et al. 2019, Dietz et al. 2010, Mizoi, Shinozaki and Yamaguchi-Shinozaki, 2012). Similarly, *Sobic.010G222600* is an ortholog to *Z. mays TRAF34*, a transcription factor that has been found

involved in a gene regulatory network for phenolic metabolism (Yang et al. 2017). As such, *Sobic.010G222600* may also transcribe a TRAF transcription factor that regulates sorghum phenolic metabolism. The study by Yang et al. (2017) also identified a myriad of transcription factor families responsible for regulating phenolic metabolism, therefore the putative TRAF and AP2/ERF transcription factors may work jointly to regulating antimicrobial activitiy in sorghum.

Other potential candidate genes include two vacuolar iron transporters (VIT), which regulate and facilitates the accumulation of soluble sugars in the plant. Brenton et al. (2020) recently described a species-specific tandem duplication which resulted in the two vacuolar iron transporter genes accounting for higher sugar accumulation. Soluble sugars are known signaling modules for responses such as plant stress; specifically, sucrose has been linked to anthocyanin accumulation in *Arabidopsis thaliana* (L.) Heynh, as well as activation of pathogenesis related genes in rice and maize (Bolouri Modhaddam and Van den Ende, 2012; Solfanelli et al., 2006; Thibaud et al., 2004; Gomez-Aiza et al., 2007). Like other phenolic and flavonoid compounds, anthocyanins have antimicrobial effects (Cisowska et al.,2011). Higher sugar levels may amplify signaling for anthocyanin accumulation, resulting in antimicrobial activity. Further, soluble sugars, such as sucrose, control the expression of cyclins in *A. thaliana* (Riou- Khamlichi et al. 2000). *Sobic.004G305700*, and its maize and rice orthologs, contained a cyclin domain. Proteins containing cyclin domains are ubiquitous, regulating the cell cycle and in turn several biological processes across all life. Cyclins have been found to be involved in plant stress response (Kitsios and Doonan, 2011). Therefore, *Sobic.004G305700* may be a cyclin involved in a plant stress response. Sugar accumulation and signaling may play a role in regulating plant hormones that are responsible for stress responses such as antimicrobial activity. Further investigation of these potential candidate genes is needed to better understand their functional roles in sorghum's

antimicrobial activity. Importantly, although we were successful in mapping loci for antimicrobial activity, we emphasize that the antimicrobial effects assayed for mapping were specific to *C. perfringens* and therefore these loci may not be applicable with representative a broader range of pathogens. Further susceptibility testing on a wider and more diverse collection of bacteria is necessary to understand the genetic architecture regulating the antimicrobial potentials of sorghum grain.

5. CONCLUSIONS

Antimicrobial activity was identified in half of the accessions within the SAP. Although the inhibitory activity of sorghum extracts was demonstrated across both food-borne pathogens in the study, it was determined that sorghum inhibits *C. perfringens* more efficiently than *S. enterica.* Further, heritability estimates show that this activity is under moderate genetic control, and strong activity for the selected accessions was found to be reliable across field replicates. Antimicrobial activity was found to be insignificantly correlated with condensed tannins and total phenols indicating that antimicrobial activity was not entirely dependent on the accumulation of these antinutrients. Additionally, antimicrobial activity did not negatively impact yield or other compositional traits. Novel associations found in the GWAS allowed for the identification of 20 potential candidate genes that may regulate antimicrobial activity. Using orthologs identified by OrthoDB, we characterized potential candidate genes that may be transcription factors known to regulate plant abiotic stress responses and phenolic metabolism. Subsequent studies are required to elucidate and validate the roles of the candidate genes as well as the putative metabolites that may be responsible for the observed antimicrobial activity to fully reveal the genetic and mechanistic basis of this phenotype. However, this initial phenotypic

and genetic evaluation of nearly 400 diverse sorghum accessions for antimicrobial potential provides valuable information on germplasm and genetic markers to facilitate the incorporation of natural antimicrobial activity via plant breeding for use in animal agriculture, which will provide a more healthy and sustainable feed grain for animal production systems.

6. REFERENCES

Alzoreky, N. S., & Nakahara, K. (2003). Antibacterial activity of extracts from some edible plants commonly consumed in Asia. *International Journal of Food Microbiology*, *80*(3), 223–230. DOI:10.1016/S0168-1605(02)00169-1

Billard, L., & Diday, E. (2000). Data analysis, classification and related methods. *Regression Analysis for Interval-Valued Data, Studies in Classification, Data Analysis, and Knowledge Organization*, 103-124.

Bolouri Moghaddam, M. R., & Van den Ende, W. (2012). Sugars and plant innate immunity. Journal of Experimental Botany, 63(11), 3989–3998.

Boyles, R. E., Cooper, E. A., Myers, M. T., Brenton, Z., Rauh, B. L., Morris, G. P., & Kresovich, S. (2016). Genome-Wide Association Studies of Grain Yield Components in Diverse Sorghum Germplasm. In The Plant Genome (Vol. 9, Issue 2, p. lantgenome2015.09.0091). https://doi.org/10.3835/plantgenome2015.09.0091

Boyles, Richard E., Brian K. Pfeiffer, Elizabeth A. Cooper, Bradley L. Rauh, Kelsey J. Zielinski, Matthew T. Myers, Zachary Brenton, William L. Rooney, and Stephen Kresovich. Genetic dissection of sorghum grain quality traits using diverse and segregating populations. *Theoretical and applied genetics* 130, no. 4 (2017): 697-716. DOI:10.1007/s00122-016-2844-6

Brenton, Z. W., Juengst, B. T., Cooper, E. A., Myers, M. T., Jordan, K. E., Dale, S. M., ... & Kresovich, S. (2020). Species-specific duplication event associated with elevated levels of nonstructural carbohydrates in Sorghum bicolor. *G3: Genes, Genomes, Genetics*, *10*(5), 1511-1520. DOI:10.1534/g3.119.400921

Brown, J. K. (2002). Yield penalties of disease resistance in crops. *Current opinion in plant biology*, *5*(4), 339-344. DOI:10.1016/S1369-5266(02)00270-4

Brown, P. J., Myles, S., & Kresovich, S. (2011). Genetic support for phenotype-based racial classification in sorghum. *Crop Science*, *51*(1), 224-230. DOI:10.2135/cropsci2010.03.0179

Burdette, A., Garner, P. L., Mayer, E. P., Hargrove, J. L., Hartle, D. K., & Greenspan, P. (2010). Anti-inflammatory activity of select sorghum (Sorghum bicolor) brans. *Journal of medicinal food,* 13(4), 879-887. DOI: 10.1089/jmf.2009.0147

Casa, A. M., Pressoir, G., Brown, P. J., Mitchell, S. E., Rooney, W. L., Tuinstra, M. R., ... & Kresovich, S. (2008). Community resources and strategies for association mapping in sorghum. *Crop science*, *48*(1), 30-40. DOI:10.2135/cropsci2007.02.0080

Chen, Z., & Jiang, X. (2017). Thermal resistance and gene expression of both desiccation-adapted and rehydrated Salmonella enterica serovar Typhimurium cells in aged broiler litter. *Applied and Environmental Microbiology*, *83*(12). DOI:10.1128/AEM.00367-17

Chuck, G., Meeley, R., & Hake, S. (2008). Floral meristem initiation and meristem cell fate are regulated by the maize AP2 genes ids1 and sid1. *Development*, *135*(18), 3013-3019. DOI: 10.1242/dev.024273

Chung, K. T., Wong, T. Y., Wei, C. I., Huang, Y. W., & Lin, Y. (1998). Tannins and human health: a review. *Critical reviews in food science and nutrition*, *38*(6), 421-464. DOI: 10.1080/10408699891274273

Cisowska, A., Wojnicz, D., & Hendrich, A. B. (2011). Anthocyanins as antimicrobial agents of natural plant origin. *Natural product communications*, *6*(1). DOI:10.1177/1934578X1100600136

CLSI. 2012. Methods for antimicrobial susceptibility testing of anaerobic bacteria, Approved standard, M11-A8-Eighth edition, Wayne, PA: Clinical and Laboratory Standard Institute.

Cowan, M. M. (1999). Plant products as antimicrobial agents. *Clinical microbiology reviews*, *12*(4), 564-582. DOI: 10.1128/CMR.12.4.564

Daglia, M. (2012). Polyphenols as antimicrobial agents. *Current opinion in biotechnology*, *23*(2), 174-181. DOI:10.1016/j.copbio.2011.08.007

Dai, J., & Mumper, R. J. (2010). Plant phenolics: extraction, analysis and their antioxidant and anticancer properties. *Molecules*, *15*(10), 7313-7352. DOI:10.3390/molecules15107313

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... & McVean, G. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156-2158. DOI:10.1093/bioinformatics/btr330

De Alencar Figueiredo, L. F., Sine, B., Chantereau, J., Mestres, C., Fliedel, G., Rami, J. F., ... & Courtois, B. (2010). Variability of grain quality in sorghum: association with polymorphism in Sh2, Bt2, SssI, Ae1, Wx and O2. *Theoretical and applied genetics*, *121*(6), 1171-1185.DOI:10.1007/s00122-010-1380-z

Dharmasena, M., & Jiang, X. (2018). Isolation of toxigenic Clostridium difficile from animal manure and composts being used as biological soil amendments. *Applied and environmental microbiology*, *84*(16). DOI: 10.1128/AEM.00738-18

Dietz, K. J., Vogel, M. O., & Viehhauser, A. (2010). AP2/EREBP transcription factors are part of gene regulatory networks and integrate metabolic, hormonal and environmental signals in stress acclimation and retrograde signaling. *Protoplasma*, *245*(1-4), 3-14. DOI:10.1007/s00709-010-0142-8)\

Dykes, L., Hoffmann Jr, L., Portillo-Rodriguez, O., Rooney, W. L., & Rooney, L. W. (2014). Prediction of total phenols, condensed tannins, and 3-deoxyanthocyanidins in sorghum grain using near-infrared (NIR) spectroscopy. *Journal of cereal science*, *60*(1), 138-142. DOI:0.1016/j.jcs.2014.02.002

Dykes, L., Rooney, L. W., Waniska, R. D., & Rooney, W. L. (2005). Phenolic compounds and antioxidant activity of sorghum grains of varying genotypes. *Journal of Agricultural and Food Chemistry*, *53*(17), 6813-6818. DOI:10.1021/jf050419e

Earp, C. F., & Rooney, L. W. (1982). Scanning electron microscopy of the pericarp and testa of several sorghum varieties. *Food Structure*, *1*(2), 3.

Espitia, P. J. P., Soares, N. D. F. F., dos Reis Coimbra, J. S., de Andrade, N. J., Cruz, R. S., & Medeiros, E. A. A. (2012). Zinc oxide nanoparticles: synthesis, antimicrobial activity and food packaging applications. *Food and bioprocess technology*, *5*(5), 1447-1464. DOI:10.1007/s11947-012-0797-6

Glaubitz, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire, R. J., Sun, Q., & Buckler, E. S. (2014). TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. PloS One, 9(2), e90346.

Gómez-Ariza, J., Campo, S., Rufat, M., Estopà, M., Messeguer, J., Segundo, B. S., & Coca, M. (2007). Sucrose-mediated priming of plant defense responses and broad-spectrum disease resistance by overexpression of the maize pathogenesis-related PRms protein in rice plants. Molecular Plant-Microbe Interactions: MPMI, 20(7), 832–842.

Gyawali, R., & Ibrahim, S. A. (2014). Natural products as antimicrobial agents. *Food control*, *46*, 412-429. DOI: 10.1016/j.foodcont.2014.05.047

Hargrove, J. L., Greenspan, P., Hartle, D. K., & Dowd, C. (2011). Inhibition of aromatase and α-amylase by flavonoids and proanthocyanidins from Sorghum bicolor bran extracts. *Journal of Medicinal Food*, *14*(7-8), 799-807. DOI:10.1089/jmf.2010.0143

Harlan, J. R., & De Wet, J. M. J. (1972). A simplified classification of cultivated sorghum 1. *Crop science*, *12*(2), 172-176. DOI:10.2135/cropsci1972.0011183X001200020005x

Herald, T. J., Gadgil, P., & Tilley, M. (2012). High-throughput micro plate assays for screening flavonoid content and DPPH-scavenging activity in sorghum bran and flour. *Journal of the Science of Food and Agriculture*, *92*(11), 2326-2331. DOI:10.1002/jsfa.5633

Huang, Q., Liu, X., Zhao, G., Hu, T., & Wang, Y. (2018). Potential and challenges of tannins as an alternative to in-feed antibiotics for farm animal production. *Animal Nutrition*, *4*(2), 137-150. DOI: 10.1016/j.aninu.2017.09.004

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science and Engineering*, *9*(3), 99–104. DOI:10.1109/MCSE.2007.55

Immerseel, F. V., Buck, J. D., Pasmans, F., Huyghebaert, G., Haesebrouck, F., & Ducatelle, R. (2004). Clostridium perfringens in poultry: an emerging threat for animal and public health. *Avian pathology*, *33*(6), 537-549. DOI:10.1080/03079450400013162

Kil, H. Y., Seong, E. S., Ghimire, B. K., Chung, I. M., Kwon, S. S., Goh, E. J., ... & Yu, C. Y. (2009). Antioxidant and antimicrobial activities of crude sorghum extract. *Food Chemistry*, *115*(4), 1234-1239. DOI: 10.1016/j.foodchem.2009.01.032

Kitsios, G., & Doonan, J. H. (2011). Cyclin dependent protein kinases and stress responses in plants. *Plant signaling & behavior*, *6*(2), 204-209. DOI:10.4161/psb.6.2.14835

Kriventseva, E. V., Kuznetsov, D., Tegenfeldt, F., Manni, M., Dias, R., Simão, F. A., & Zdobnov, E. M. (2019). OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. Nucleic Acids Research, 47(D1), D807–D811.

Kruijer, W., Boer, M. P., Malosetti, M., Flood, P. J., Engel, B., Kooke, R., Keurentjes, J. J. B., & van Eeuwijk, F. A. (2015). Marker-Based Estimation of Heritability in Immortal Populations. Genetics, 199(2), 379–398

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, *26*(5), 589–595. DOI:10.1093/bioinformatics/btp698

Mathew, A. G., Cissell, R., & Liamthong, S. (2007). Antibiotic resistance in bacteria associated with food animals: a United States perspective of livestock production. *Foodborne pathogens and disease*, *4*(2), 115-133. DOI:10.1089/fpd.2006.0066

McEwen, S. A., & Fedorka-Cray, P. J. (2002). Antimicrobial use and resistance in animals. Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America, 34 Suppl 3, S93–S106.

McKinney, W. (2010, June). Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference* (Vol. 445, pp. 51-56).

Miller, R. W., Skinner, J., Sulakvelidze, A., Mathis, G. F., & Hofacre, C. L. (2010). Bacteriophage therapy for control of necrotic enteritis of broiler chickens experimentally infected with Clostridium perfringens. *Avian diseases*, *54*(1), 33-40. DOI:10.1637/8953-060509-Reg.1

Mizoi, J., Shinozaki, K., & Yamaguchi-Shinozaki, K. (2012). AP2/ERF family transcription factors in plant abiotic stress responses. Biochimica et Biophysica Acta, 1819(2), 86–96.

Mizoi, J., Kanazawa, N., Kidokoro, S., Takahashi, F., Qin, F., Morimoto, K., ... & Yamaguchi-Shinozaki, K. (2019). Heat-induced inhibition of phosphorylation of the stress-protective

transcription factor DREB2A promotes thermotolerance of Arabidopsis thaliana. *Journal of Biological Chemistry*, *294*(3), 902-917. DOI: 10.1074/jbc.RA118.002662

Morris, G. P., Rhodes, D. H., Brenton, Z., Ramu, P., Thayil, V. M., Deshpande, S., Hash, C. T., Acharya, C., Mitchell, S. E., Buckler, E. S., Yu, J., & Kresovich, S. (2013). Dissecting genome-wide association signals for loss-of-function phenotypes in sorghum flavonoid pigmentation traits. G3, 3(11), 2085–2094.

Nitiema, L. W., Savadogo, A., Simpore, J., Dianou, D., & Traore, A. S. (2012). In vitro antimicrobial activity of some phenolic compounds (coumarin and quercetin) against gastroenteritis bacterial strains. *Int J Microbiol Res*, *3*(3), 183-7. DOI: 10.5829/idosi.ijmr.2012.3.3.6414

Pfeiffer, B. K., & Rooney, W. L. (2016). Inheritance of pericarp color, nutritional quality, and grain composition traits in black sorghum. *Crop Science*, *56*(1), 164-172. DOI: 10.2135/cropsci2015.04.0224

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., ... & Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American journal of human genetics*, *81*(3), 559-575. DOI:10.1086/519795

Redondo, L. M., Chacana, P. A., Dominguez, J. E., & Fernandez Miyakawa, M. E. D. (2014). Perspectives in the use of tannins as alternative to antimicrobial growth promoter factors in poultry. *Frontiers in Microbiology*, *5*, 118. DOI:10.3389/fmicb.2014.00118

Rhodes, D. H., Hoffmann Jr, L., Rooney, W. L., Ramu, P., Morris, G. P., & Kresovich, S. (2014). Genome-wide association study of grain polyphenol concentrations in global sorghum [Sorghum bicolor (L.) Moench] germplasm. *Journal of agricultural and food chemistry*, *62*(45), 10916-10927. DOI:10.1021/jf503651t

Rhodes, D. H., & Kresovich, S. (2016). Sorghum [Sorghum bicolor (L.) Moench] genotypes with contrasting polyphenol compositions differentially modulate inflammatory cytokines in mouse macrophages. *Journal of Chemistry*, *2016*. DOI:10.1155/2016/9640869

Rhodes, D., Gadgil, P., Perumal, R., Tesso, T., & Herald, T. J. (2017). Natural variation and genome-wide association study of antioxidants in a diverse sorghum collection. *Cereal Chemistry*, *94*(2), 190-198. DOI:10.1094/CCHEM-03-16-0075-R

Riou-Khamlichi, C., Menges, M., Healy, J. S., & Murray, J. A. (2000). Sugar control of the plant cell cycle: differential regulation of Arabidopsis D-type cyclin gene expression. *Molecular and cellular biology*, *20*(13), 4513-4521. DOI: 10.1128/MCB.20.13.4513-4521.2000

Sapkota, S., Boyles, R., Cooper, E., Brenton, Z., Myers, M., & Kresovich, S. (2020). Impact of sorghum racial structure and diversity on genomic prediction of grain yield components. *Crop Science*, *60*(1), 132-148. DOI:10.1002/csc2.20060

Scalbert, A. (1991). Antimicrobial properties of tannins. *Phytochemistry*, *30*(12), 3875-3883. DOI:10.1016/0031-9422(91)83426-L

Singleton, V. L., Orthofer, R., & Lamuela-Raventós, R. M. (1999). [14] Analysis of total phenols and other oxidation substrates and antioxidants by means of folin-ciocalteu reagent. In Methods in Enzymology (Vol. 299, pp. 152–178). Academic Press.

Solfanelli, C., Poggi, A., Loreti, E., Alpi, A., & Perata, P. (2006). Sucrose-specific induction of the anthocyanin biosynthetic pathway in Arabidopsis. *Plant physiology*, *140*(2), 637-646. DOI: 10.1104/pp.105.072579

Stanić, V., Dimitrijević, S., Antić-Stanković, J., Mitrić, M., Jokić, B., Plećaš, I. B., & Raičević, S. (2010). Synthesis, characterization and antimicrobial activity of copper and zinc-doped hydroxyapatite nanopowders. *Applied Surface Science*, *256*(20), 6083-6089. DOI:10.1016/j.apsusc.2010.03.124

Swarts, K., Li, H., Romero Navarro, J. A., An, D., Romay, M. C., Hearne, S., ... & Buckler, E. S. (2014). Novel methods to optimize genotypic imputation for low-coverage, next-generation sequence data in crop plants. *The Plant Genome*, *7*(3), 1-12. DOI:10.3835/plantgenome2014.05.0023

Swartz, M. N. (2002). Human diseases caused by foodborne pathogens of animal origin. *Clinical Infectious Diseases*, *34*(Supplement_3), S111-S122. DOI:10.1086/340248

Thibaud, M.-C., Gineste, S., Nussaume, L., & Robaglia, C. (2004). Sucrose increases pathogenesis-related PR-2 gene expression in Arabidopsis thaliana through an SA-dependent but NPR1-independent signaling pathway. Plant Physiology and Biochemistry: PPB / Societe Francaise de Physiologie Vegetale, 42(1), 81–88.

Vallat, R. (2018). Pingouin: statistics in Python. *Journal of Open Source Software*, *3*(31), 1026. DOI: 10.21105/joss.01026

Van Boeckel, T. P., Pires, J., Silvester, R., Zhao, C., Song, J., Criscuolo, N. G., Gilbert, M., Bonhoeffer, S., & Laxminarayan, R. (2019). Global trends in antimicrobial resistance in animals in low- and middle-income countries. Science, 365(6459). https://doi.org/10.1126/science.aaw1944

Waskom, M., Botvinnik, O., Hobson, P., Cole, J. B., Halchenko, Y., Hoyer, S., … Allan, D. (2014, November 14). seabornLv0.5.0 (November 2014) (Version v0.5.0). Zenodo. DOI:10.528/zenodo.12710

White, D. G., Zhao, S., Simjee, S., Wagner, D. D., & McDermott, P. F. (2002). Antimicrobial resistance of foodborne pathogens. *Microbes and infection*, *4*(4), 405-412. DOI:10.1016/S1286-4579(02)01554-X

Wu, Y., Li, X., Xiang, W., Zhu, C., Lin, Z., Wu, Y., ... & Wang, M. L. (2012). Presence of tannins in sorghum grains is conditioned by different natural alleles of Tannin1. *Proceedings of the National Academy of Sciences*, *109*(26), 10281-10286. DOI:10.1073/pnas.1201700109

Wu, Y., Guo, T., Mu, Q., Wang, J., Li, X., Wu, Y., ... & Trick, H. N. (2019). Allelochemicals targeted to balance competing selections in African agroecosystems. *Nature Plants*, *5*(12), 1229-1236. DOI:10.1038/s41477-019-0563-0

Xie, Z., Nolan, T. M., Jiang, H., & Yin, Y. (2019). AP2/ERF transcription factor regulatory networks in hormone and abiotic stress responses in Arabidopsis. *Frontiers in plant science*, *10*, 228. DOI:10.3389/fpls.2019.00228

Yang, F., Li, W., Jiang, N., Yu, H., Morohashi, K., Ouma, W. Z., Morales-Mantilla, D. E., Gomez-Cano, F. A., Mukundi, E., Prada-Salcedo, L. D., Velazquez, R. A., Valentin, J., Mejía-Guerra, M. K., Gray, J., Doseff, A. I., & Grotewold, E. (2017). A Maize Gene Regulatory Network for Phenolic Metabolism. Molecular Plant, 10(3), 498–515.

Zhou, X., & Stephens, M. (2014). Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nature methods*, *11*(4), 407-409. DOI: 10.1038/nmeth.2848

FIGURES

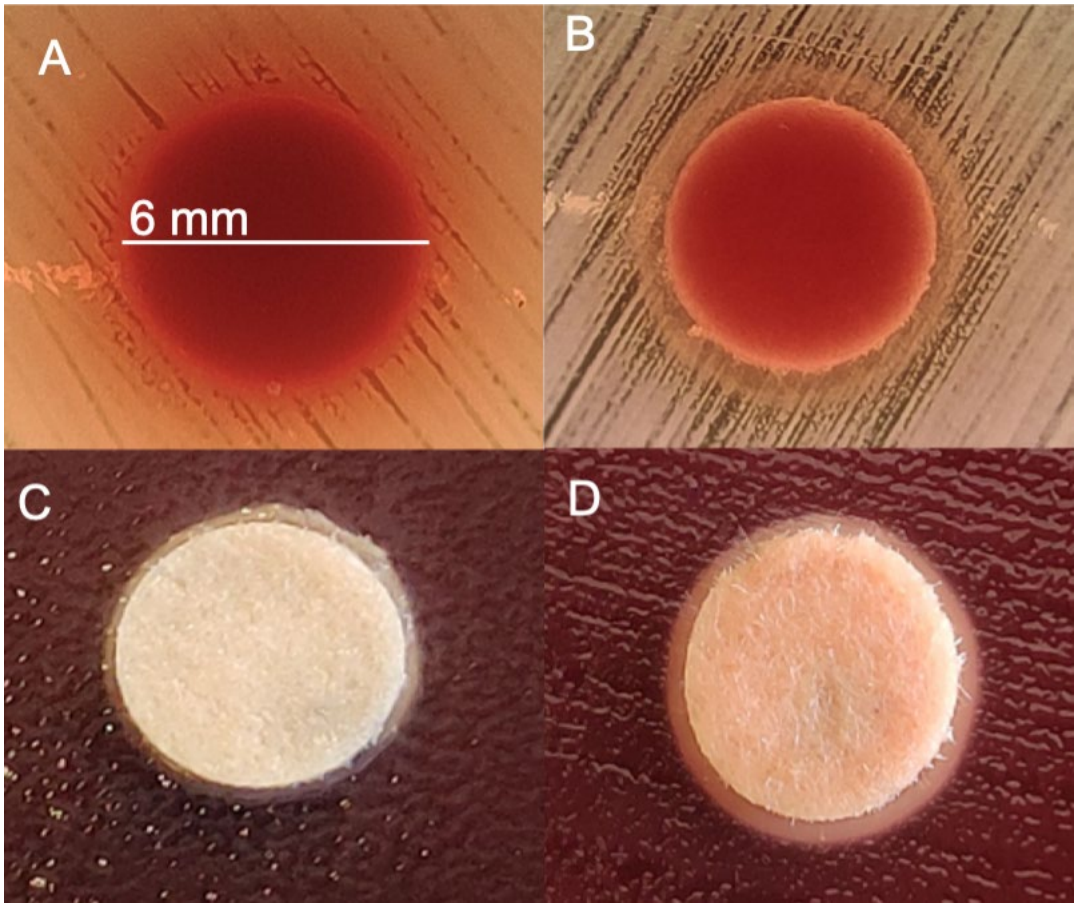

Figure 1. Range of antimicrobial activity observed from the disc-diffusion assay using a 6-mm cotton disc. A) and B) show weak antimicrobial activity against S. *enterica*. Although some inhibition can be visually observed, there is still presence of growth in inhibition zone therefore making it difficult to measure. C) is an example of weak antimicrobial activity against *C. perfringens* and D) shows strong antimicrobial activity against *C. perfringens*, which is identified as having clear inhibition zone boundaries or an inhibition zone greater than 7 mm.

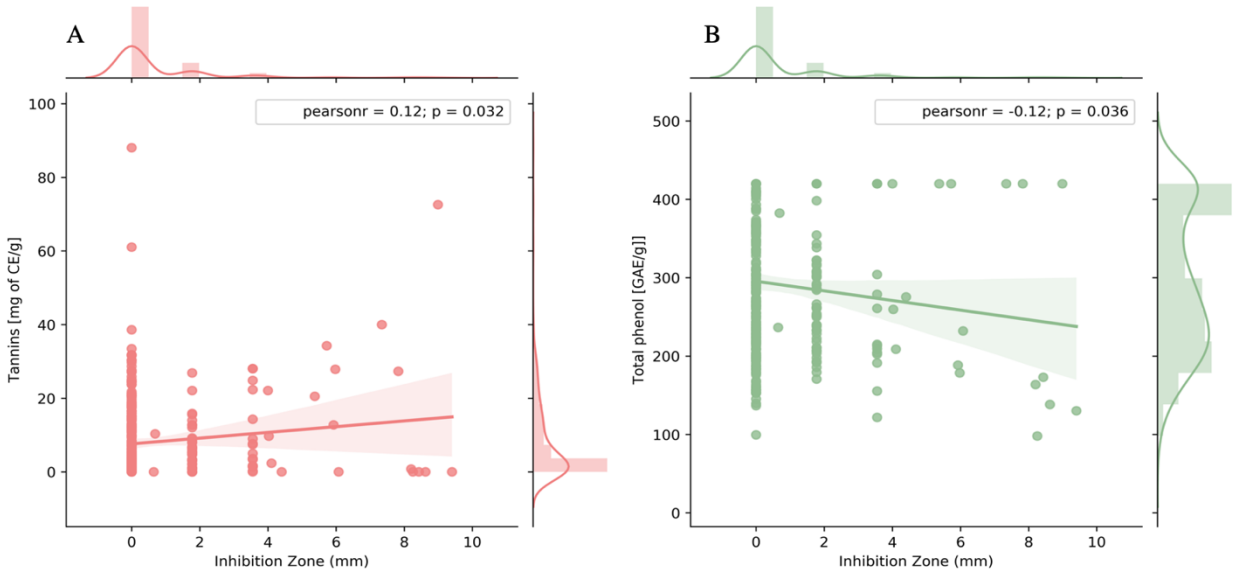

Figure 2. Relationships between *C. perfringens* inhibition zone (mm) and a) total phenol concentration *(r* = 0.12; p = 0.36) and b) tannin concentration (*r* = 0.12; p = 0.032).

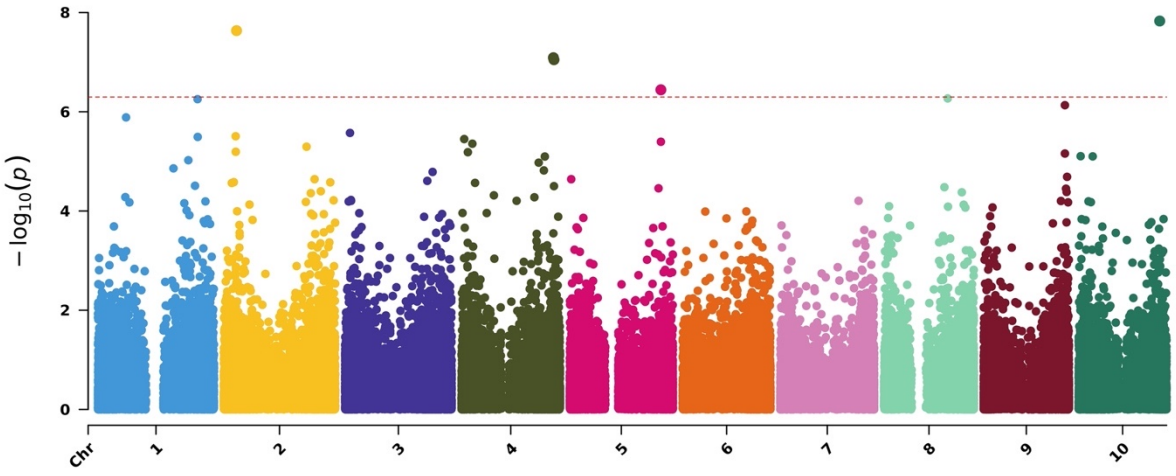

Figure 3. GWAS of antimicrobial activity in sorghum grain. An LMM was used for association analysis using 99,126 SNP markers. The y-axis (-$\log_{10}$ p-values) is plotted against the position on the chromosome (x-axis). The dashed line indicates the Bonferroni significance threshold. Regions with -$\log_{10}$ p-values above the threshold (dotted line) are candidates**.**

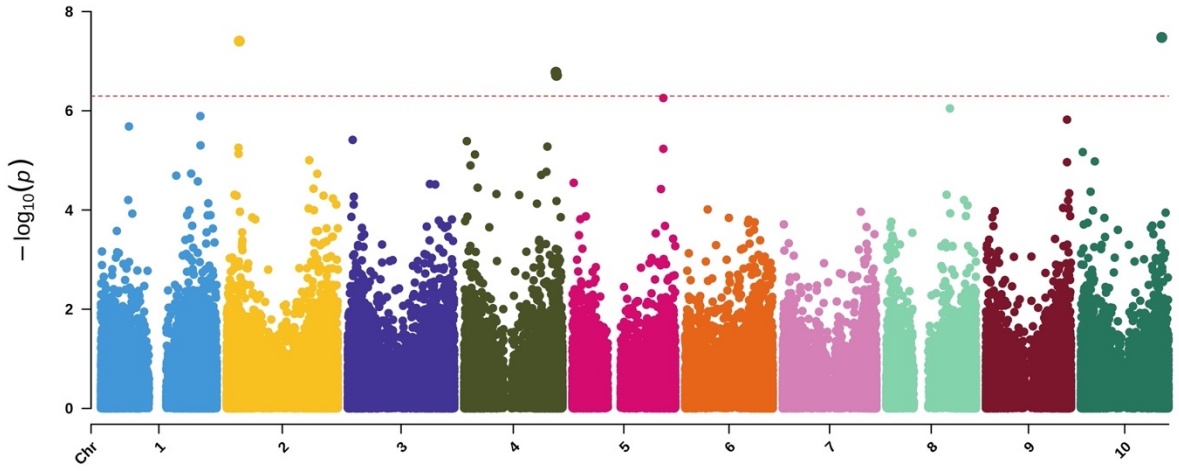

Figure 4. GWAS for antimicrobial activity with tannin as a covariate. Manhattan plot of association analysis using 99,126 SNP markers. The y-axis (-$\log_{10}$ p-values) is plotted against the position on the chromosome (x-axis). The dashed line indicates the Bonferroni significance threshold.

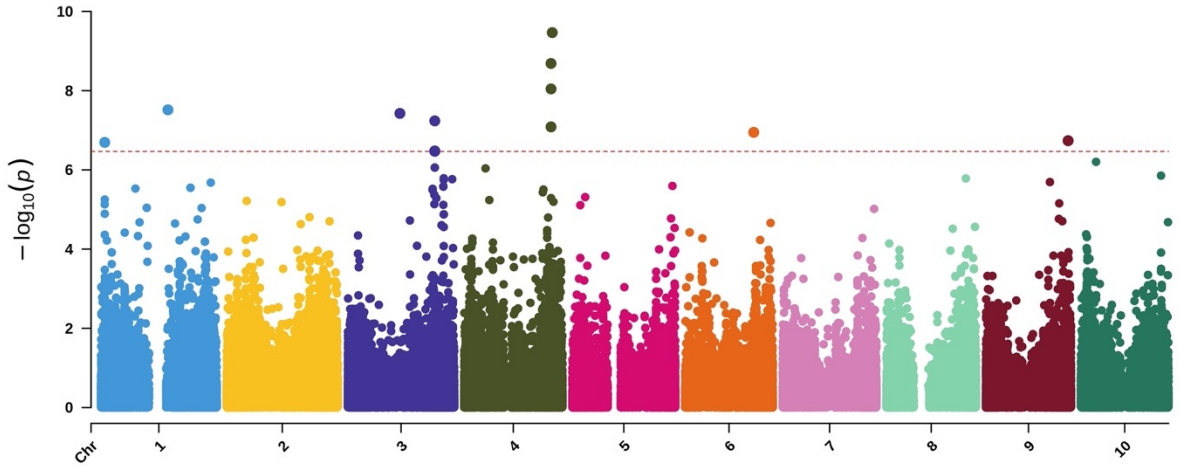

Figure 5. Manhattan plot based on the GWAS for the tannin trait, using tannin BLUPs. Association analysis included 146,280 SNP markers from 367 individuals. The y-axis (-$\log_{10}$ p-values) is plotted against the position on the chromosome (x-axis). The dashed line indicates the

Bonferroni significance threshold. The distinguished peak on chromosome 4 colocalizes with *Tan1* (~61-62 MB).

TABLES

Table 1. Minimum inhibitory concentrations (MIC) for *C. perfringens* and *S. enterica*

	C. perfringens				*S. enterica*			
Accession	MICr1 (mg/ml)	MICr2 (mg/ml)	IZr1 (mm)	Izr2 (mm)	MICr1 (mg/ml)	MICr2 (mg/ml)	Izr1 (mm)	Izr2 (mm)
PI607931	0.2	0.08	0	0	25.16	20.16	0	0
PI629059	1.14	0.23	0	0	36.35	29.49	0	0
PI576376	0.14	0.16	weak	weak	18.16	ND	0	0
PI597957	0.3	0.16	weak	8.6	18.91	20.16	0	0
PI533979	0.39	0.14	8.2	weak	49.92	9.27	weak	weak
PI576393	0.13	0.14	8.4	weak	8.56	17.59	weak	weak
PI533869	0.19	0.19	8.1	8.4	11.91	11.99	0	0
PI656003	0.18	0.11	weak	weak	46.1	58.28	weak	weak
PI656035	0.13	0.13	weak	weak	16.21	16.58	0	0
PI655995	0.1	0.09	8.53	9.45	50.71	44.29	weak	weak
PI642998	0.05	0.06	7.4	8.05	26.3	7.1	weak	weak
PI641836	7.5	12.54	7.9	weak	14.99	12.54	weak	weak

[a] MICr1, Minimum inhibitory concentration- replication 1; MICr2, Minimum inhibitory concentration - replication 2; Izr1, Inhibition zone - replication 1; Izr2, Inhibition zone- replication 2; ND= Non-detectable

Table 2. Summary of *C. perfringens* samples regarding testa and antimicrobial activity

	No activity	Weak activity	Strong activity
Unpigmented testa	299	51	19
Pigmented testa	227	32	15

Table 3. Accessions with antimicrobial properties and associated agronomic traits

PI	Common Name	IZ (mm)	PH (cm)	DTM	TGW (g)
PI533869	Msumbji SB 117	8.25	129	104	19.26
PI533871	M 1	8.43	118	101	29.78
PI533948	Nebraska 6350	8.63	79	96	27.72
PI534115	Akwu	8.20	78	113	31.20
PI597957	SC1158	6.08	76	105	22.65
			113*	106*	23.34*

[a]IZ, inhibition zones are reported as averages between replicates; PH, plant height; DTM, days to maturity; TGW, thousand grain weight.
[b]*mean phenotype value across the entire SAP

APPENDICES

SUPPLEMENTARY DATA

Supplemental data contains results of the MIC analysis, descriptive statistics and correlations of compositional traits, correlations analysis with S. enterica, race and origin distribution graphs, Manhattan plots from additional GWAS and their corresponding QQ plots, and the significance and MAF of the top associations from the antimicrobial activity GWAS. Supplementary data files to this article, and scripts used for analysis are available at https://github.com/lkshiel/ACRE.

SUPPLEMENTARY FIGURES AND TABLES

Table S1. Summary of *S. enterica* samples regarding testa and antimicrobial activity

	No activity	Antimicrobial activity
Unpigmented testa	350	19
Pigmented testa	109	165

Table S1 summarizes the distribution of samples having weak or no antimicrobial activity across unpigmented and pigmented testa. The majority of samples (350) show no antimicrobial activity and do not have a pigmented testa. Complemented by 165 samples that both do demonstrate antimicrobial activity and have a pigmented testa. The distribution of samples of these two traits follow expected relationships with germplasm that does not meet our criteria for breeding material. The 19 samples identified as having antimicrobial activity and unpigmented testa do meet the criteria of interest, however, only weak effect was observed and therefore was not considered any further.

Table S2. *t*-test between 2014 and 2013 tannin data for the SAP

	T	DF	P-val	CI (95%)	Cohen-d	BF10	Power
t-test	0.954	510	0.34	[-1.1 – 3.18]	0.084	0.153	0.159

[a]T, t-value; DF, degrees of freedom; P-val, p-value; CI, 95% confidence intervals of the difference in means; Cohen-d, Cohen d effect size; BF10, Bayes Factor of the alternative hypothesis; Power, achieved power of the test (= 1 – type II error)

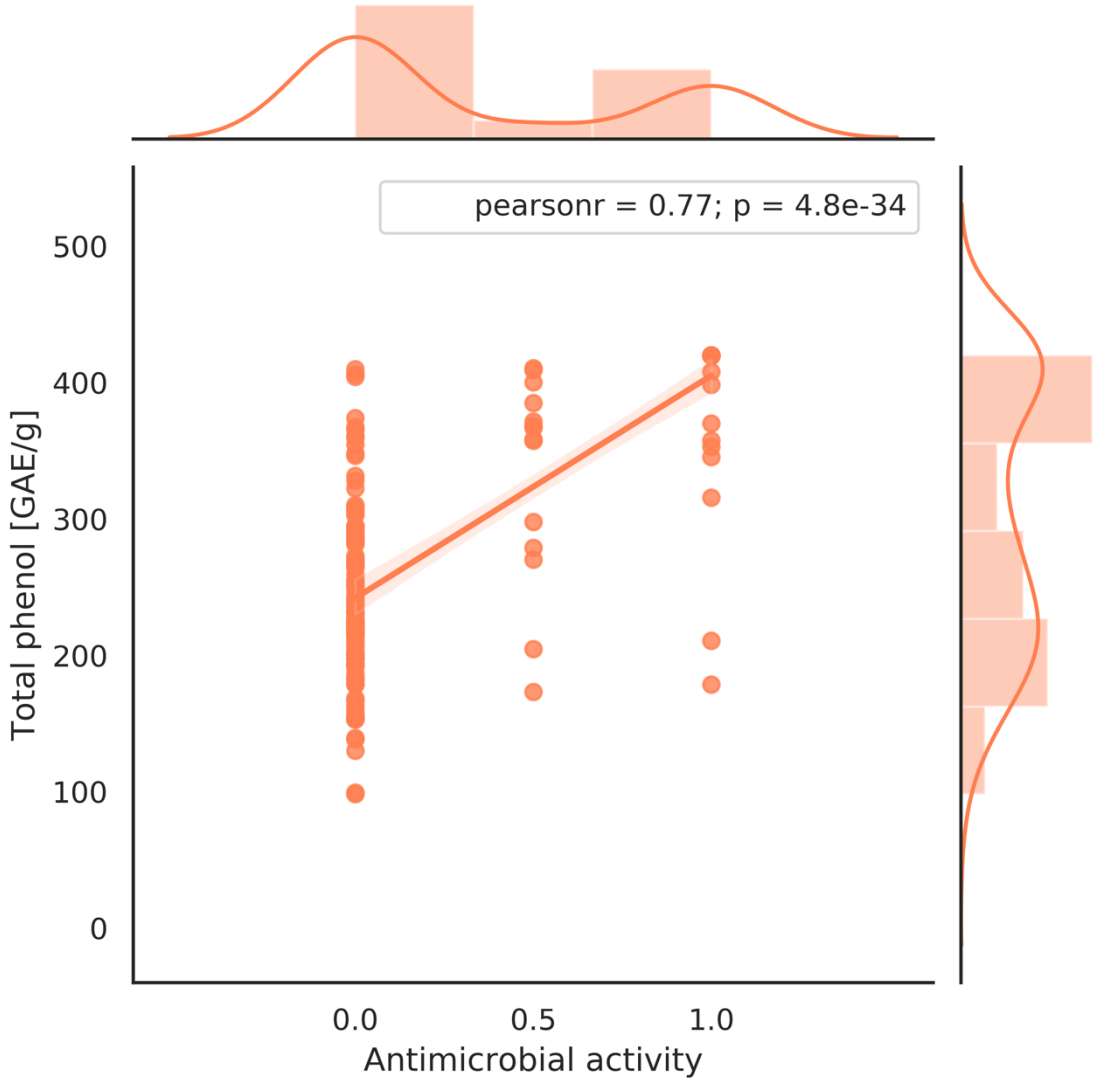

Figure S1. Relationship between antimicrobial activity against *S. enterica* (0 = no activity, 1= weak activity) and total phenol concentration. Pearson's r = 0.77

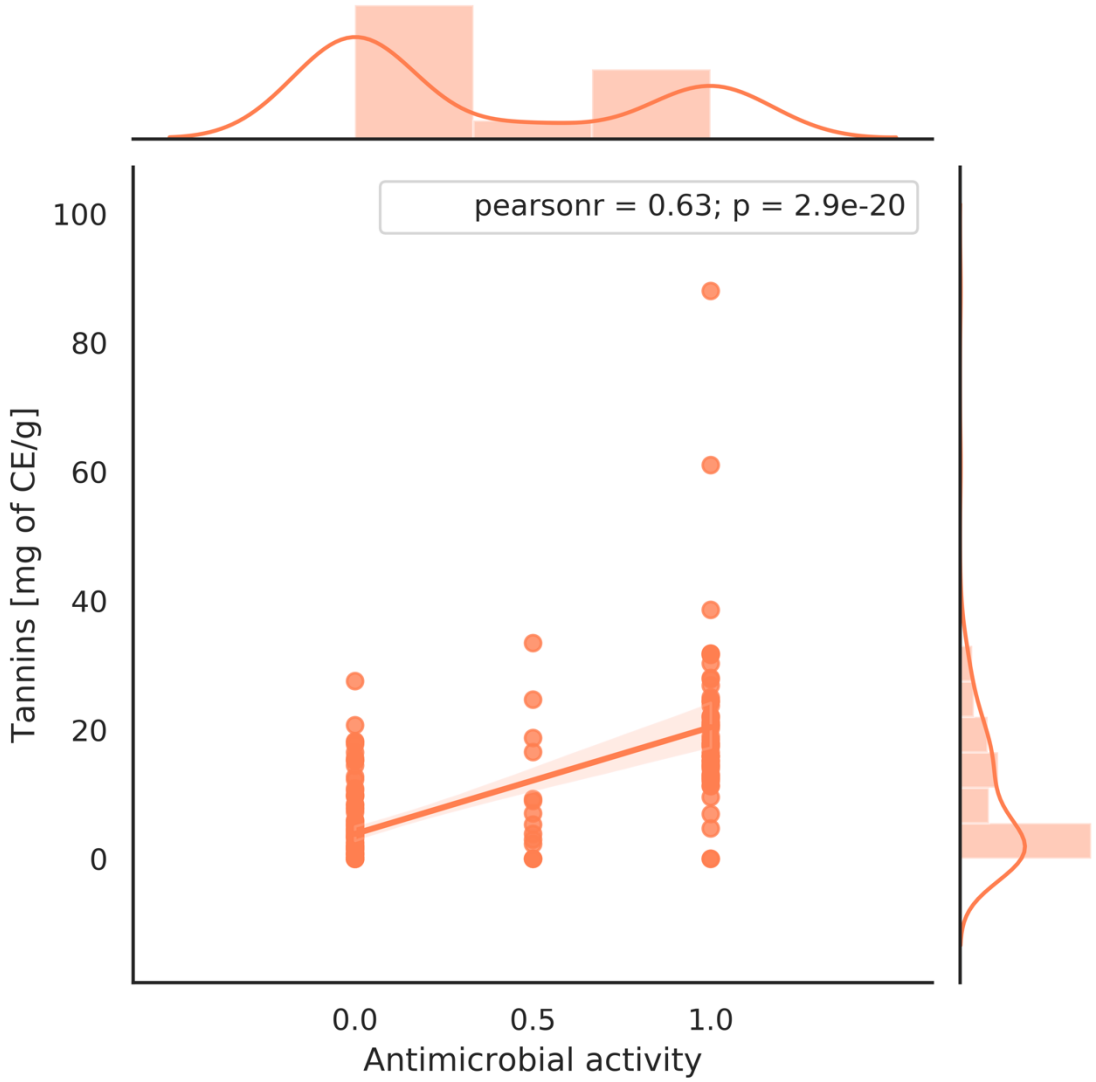

Figure S2. Relationship between antimicrobial activity against *S. enterica* (0 = no activity, 1 = weak activity) and tannin content. $r = 0.63$

Total phenol (**Figure S1**) and tannin content (**Figure S2)** were found to correlate with antimicrobial activity. Findings suggest that the majority of accessions do not meet the criteria required to identify sorghum germplasm that can be used for its potential as an antimicrobial in feed grain.

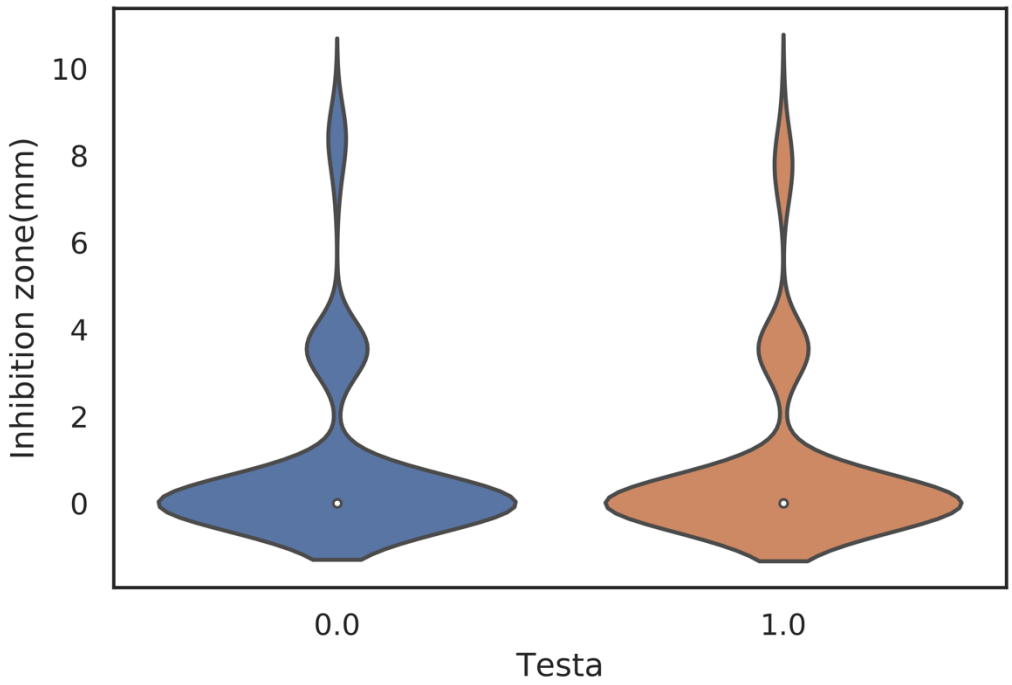

Figure S3. Distribution of accession with unpigmented testa (0) and pigmented testa (1) as it relates to inhibition zone size (mm) measured from *C. perfringens* disc-diffusion assay.

Table S3. Summary statistics for compositional, yield, and metabolite data of sorghum grain

	Minimum	Maximum	Average
ADF	3.03	9.39	5.32
Amylopectin to Starch	75	102.33	86.33
Amylopectin to Total Dry Matter	40.78	67.33	55.72
Amylose to Starch	0.83	23.46	13.46
Ash	1.38	1.93	1.64
BTU	7146.63	8006.36	7434.27
Cal	3922.98	4371.81	4079.62
Calcium	0.01	0.02	0.01
Copper	2.97	5.08	4.02
Dry matter	86.63	89.27	87.85
Fat	0.64	6.04	2.23
Iron	15.75	68.06	39.99
IVSD	38.6	56.06	46.23
KCal	1797.73	2004.51	1870.32
Lead	0.03	0.77	0.22
Magnesium	1094.48	2258.23	1537.04
Manganese	9.28	18.49	14.02
Moisture	10.73	13.37	12.15
Moisture for Ash	10.29	12.35	11.68
Moisture for Fat	9.81	12.59	11.64
NDF	4.5	23.68	9.23
Nitrogen	0.91	2.6	1.8
Nitrogen mg	4.53	13.46	9.22
Phosphorus	0.28	0.46	0.36
Prolamin	3.64	7.07	5.31
Protein	6.27	16.43	11.38
Sodium	0	0	0
Starch	54.97	76.54	68.49
Zinc	8.8	27.87	19.01
GNP	79	3445	1085
TGW (g)	5.73	56.2	23.3
YPP	0.47	97.9	25.2
Tannins (mg CE/g)	0	88.1	8.24
Total phenols [GAE/g]	6.9	420	292

[a] ADF, acid detergent fiber; BTU, British thermal unit; Cal, calorie; IVSD, in vitro starch. Disappearance/digestion; Kcal, kilocalorie; NDF, neutral detergent fiber; GNP, grain number panicle; TWG, thousand grain weight; YPP, yield primary panicle

Table S3 provides summary statistics of all 29 composition traits measured using NIRS. Correlation values and significant testing was determined to evaluate the relationships between antimicrobial activity against *C. perfringens* and each composition trait. All traits, with the exception of dry matter, moisture, and moisture for ash, were found to have insignificant correlations (**Table S4**). Due to the inevitable increase in the error during multivariate testing, a false discovery rate (FDR) correction was applied (q). Upon the FDR correction, all correlations between composition traits and antimicrobial activity were found insignificant (**Table S4**).

Table S4. Correlation and significance of compositional, yield, and metabolite traits to antimicrobial activitiy

	r value	p value	q value
ADF	0	0.989	0.989
Amylopectin to Starch	-0.048	0.181	0.59
Amylopectin to Total Dry Matter	0.01	0.784	0.895
Amylose to Starch	0.041	0.259	0.716
Ash	-0.008	0.829	0.904
BTU	0.047	0.193	0.59
Cal	0.037	0.301	0.773
Calcium	-0.059	0.103	0.59
Copper	0.029	0.417	0.877
Dry matter	0.078	0.03	0.361
Fat	0.023	0.514	0.877
Iron	0.019	0.598	0.895
IVSD	0.022	0.538	0.877
KCal	0.064	0.077	0.59
Lead	0.021	0.56	0.877
Magnesium	-0.023	0.526	0.877
Manganese	0.011	0.763	0.895
Moisture	-0.078	0.03	0.361
Moisture for Ash	-0.093	0.01	0.354
Moisture for Fat	0.011	0.752	0.895
NDF	-0.012	0.729	0.895
Nitrogen	0.009	0.795	0.895
Nitrogen mg	0.015	0.686	0.895
Phosphorus	-0.061	0.088	0.59
Prolamin	-0.025	0.491	0.877
Protein	-0.006	0.858	0.908
Sodium	0	0	0
Starch	0.014	0.697	0.895
Zinc	0.004	0.906	0.932
GNP	0.00308	0.9353	-
TGW (g)	0.00796	0.8334	-
YPP	0.00201	0.9577	-
Tannins (mg CE/g)	0.12	0.032	-
Total phenols [GAE/g]	-0.12	0.036	-

[a] ADF, acid detergent fiber; BTU, British thermal unit; Cal, calorie; IVSD, in vitro starch. Disappearance/digestion; Kcal, kilocalorie; NDF, neutral detergent fiber; GNP, grain number panicle; TWG, thousand grain weight; YPP, yield primary panicle

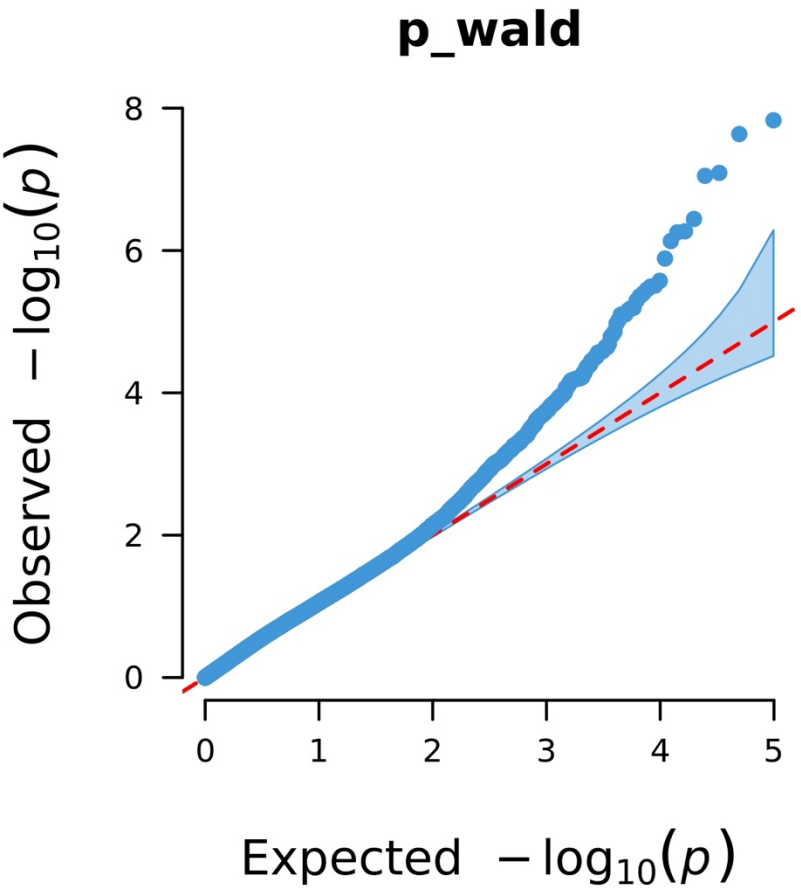

Figure S4. QQ plot for LMM of antimicrobial activity.

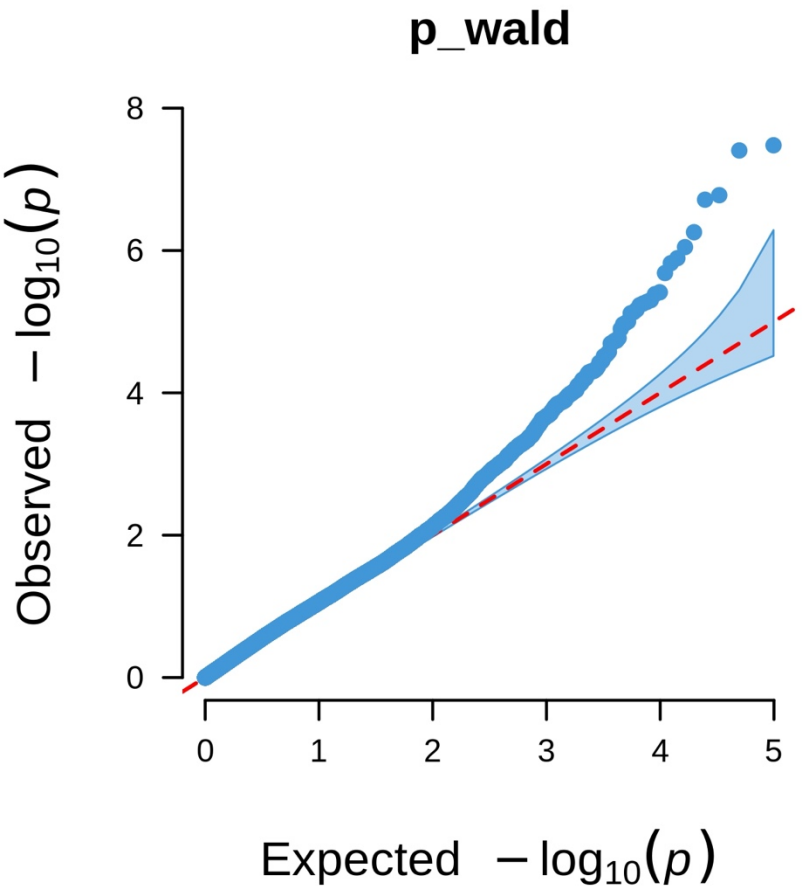

Figure S5. QQ plot for LMM with tannin covariate of antimicrobial activity.

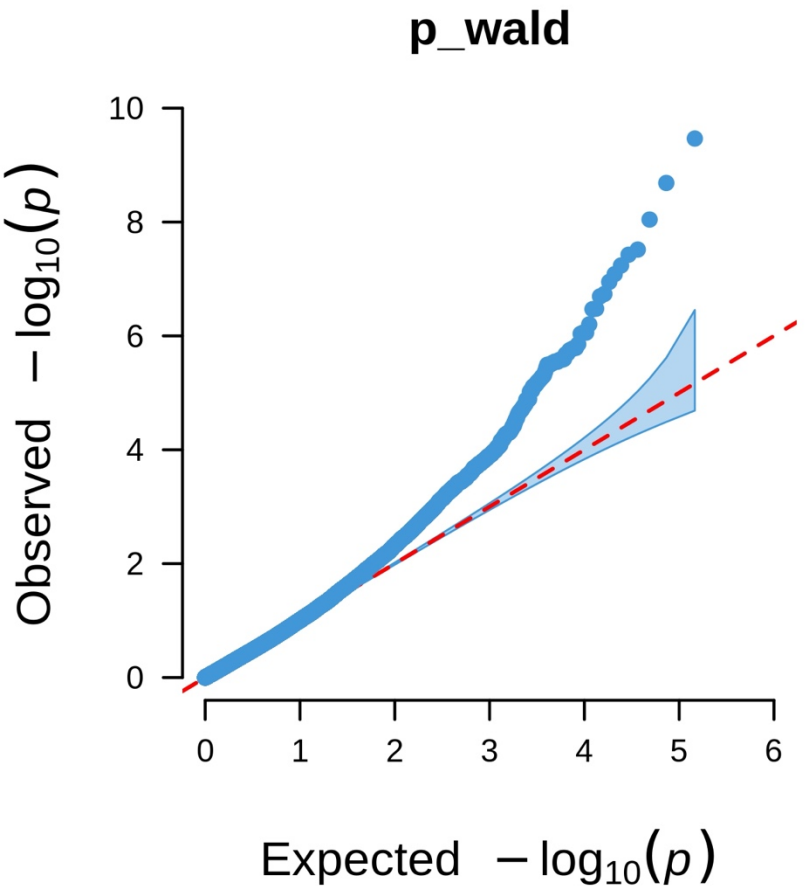

Figure S6. QQ plot for LMM of tannin trait.

Table S5. Top four significant associations for antimicrobial activity GWAS. P-Wald value was calculated and adjusted by the Bonferroni correction.

Chromosome	SNP	Position	allele1	allele0	MAF	p_wald
2	S2_8924006	8924006	C	T	0.06	2.31E-08
4	S4_64038743	64038743	T	G	0.051	8.13E-08
4	S4_64439967	64439967	C	A	0.15	8.92E-08
10	S10_56476103	56476103	T	G	0.055	1.48E-08

Table S6. T-test comparing Wald's p-values for SNPs in antimicrobial activity and tannin covariate GWAS

	T	DF	Tail	P-Val	CI 95%	Cohen-D	BF10	Power
T-test	-1.21	198239	Two-sided	0.227	[-0, 0]	0.00543	0.011	0.227

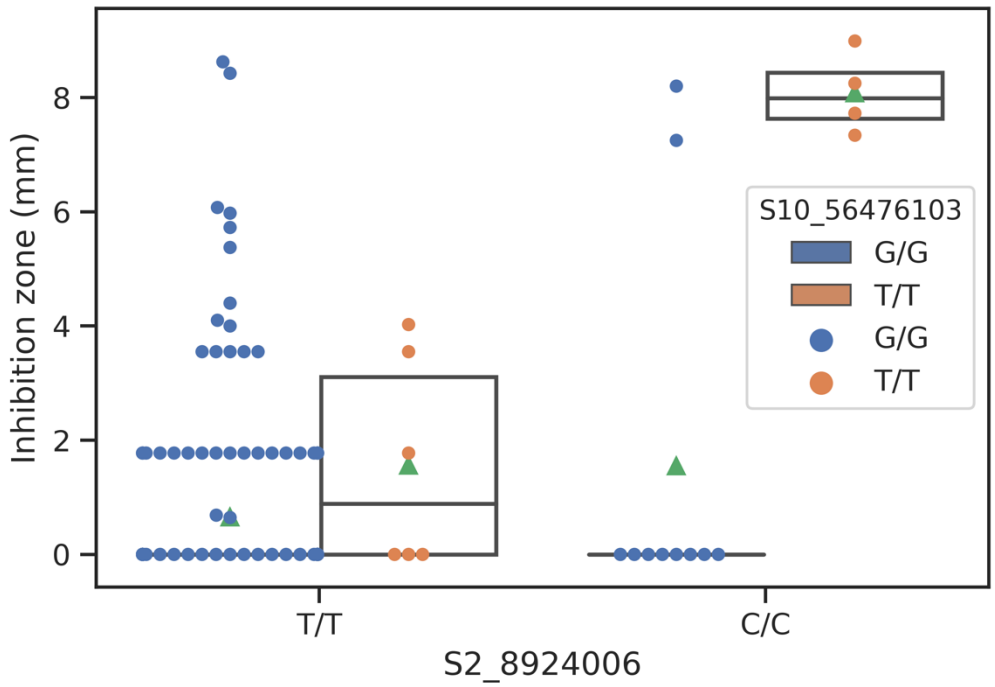

Figure S7. Distribution and density of accessions that have both favorable alleles at SNPs S2_8924006 and S10_56476103 (C/C + T/T), only S2_8934006 (C/C + G/G), only S10_56476103 (T/T + T/T), and neither favorable allele (T/T + G/G), across inhibition zone (mm). Triangles represent the mean inhibition zones for each genotype.

Table S8. Local LD blocks for each significant SNP and their potential candidate genes with descriptions and orthologs

SNP	Chr	Position	LD_start	LD_end	Genes	Description	Zm_orthologs	Os_orthologs
S2_8924006	2	8924006	8887681	8924006	Sobic.002G083200	contains ZIP domain	Zip/Zinc iron transport family protein	ZIP8 - Zinc transporter 8
					Sobic.002G083266	contains cytochrome b5	Cytochrome b5 domain-containing protein RLF	Flavohemoprotein b5/b5R-like
					Sobic.002G083400	contains Alfin and PHD domains	ALF19 - Alfin-like transcription factor, pHD finger protein	PHD finger protein ALFIN-LIKE 2
					Sobic.002G083500	F-box	uncharacterized	uncharacterized
					Sobic.002G083600	contains 2 AP2	sid11; ids1-AP2-EREBP transcription factor	Putative indeterminate spikelet 1
					Sobic.002G083332	uncharacterized	uncharacterized	uncharacterized
S4_64038743	4	64038743	64029018	64038743	Sobic.004G301400	contains ATG13	atg13a - Autophagy-related protein 13a	uncharacterized
					Sobic.004G301500	contains 2 VIT1 domains	contains a VIT1 domain	Vacuolar iron transporter homolog 2
					Sobic.004G301600	contains 2 VIT1 domains	contains a VIT1 domain	Vacuolar iron transporter homolog 2
					Sobic.004G301650	contains 2 VIT1 domains/transmembrane region	NA	Vacuolar iron transporter homolog 4
S4_64439967	4	64439967	64406892	64451499	Sobic.004G305700	contains cyclin and cyclin_C	umc1775- Cyclin-D2-1; Belongs to the cyclin family	Cyclin-D4-1 ; Belongs to the cyclin family. Cyclin D subfamily
					Sobic.004G305800	CAAD	Protein CURVATURE THYLAKOID 1D chloroplastic	contains CAAD
					Sobic.004G305900	DNAJ	Chaperone protein dnaJ	DNAJ heat shock N-terminal domain-containing protein-like
					Sobic.004G306100	uncharacterized	uncharacterized	uncharacterized
					Sobic.004G305733	uncharacterized	uncharacterized	uncharacterized
					Sobic.004G305766	uncharacterized	uncharacterized	uncharacterized
S10_56476103	10	56476103	56475132	56498633	Sobic.010G222600	contains BTB	TRAF transcription factor, Speckle-type POZ protein	Zinc finger POZ domain protein-like
					Sobic.010G222666	uncharacterized	uncharacterized	uncharacterized
					Sobic.010G222732	uncharacterized	uncharacterized	uncharacterized
					Sobic.010G222800	uncharacterized	uncharacterized	uncharacterized