Genome-Wide Association Mapping of Resistance to the Sugarcane Aphid in *Sorghum bicolor*

Somashekhar Punnuri (punnuris@fvsu.edu)
Agriculture Research Station, 1005 State University Dr, Fort Valley State University, Fort Valley, GA 31030
Addissu Ayele
Agriculture Research Station, 1005 State University Dr, Fort Valley State University, Fort Valley, GA 31030
Karen Harris-Shultz
USDA-ARS, Crop Genetics and Breeding Research Unit, 115 Coastal Way, Tifton, GA 31793
Joseph Knoll
USDA-ARS, Crop Genetics and Breeding Research Unit, 115 Coastal Way, Tifton, GA 31793
Alisa Coffin
USDA-ARS, Southeast Watershed Research Laboratory, 2316 Rainwater Road, Tifton, GA 31793
Haile Tadesse
USDA-ARS, Southeast Watershed Research Laboratory, 2316 Rainwater Road, Tifton, GA 31793
Scott Armstrong
USDA-ARS, Wheat Peanut and Other Field Crops Research Unit, 1301 N. Western Road, Stillwater, OK 74075
Trahmad Wiggins
Agriculture Research Station, 1005 State University Dr, Fort Valley State University, Fort Valley, GA 31030
Hanxia Li
University of Georgia Institute of Bioinformatics, 120 Green Street, Athens, GA 30602
Scott Sattler
USDA-ARS, Wheat, Sorghum and Forage Research Unit, Lincoln, NE 68583
Jason Wallace
University of Georgia, Department of Crop & Soil Sciences, 120 Carlton Street, Athens GA 30602

Research Article

Keywords: Melanaphis sacchari (Zehntner), marker-trait associations, genotyping-by-sequencing

Posted Date: October 21st, 2021

DOI: https://doi.org/10.21203/rs.3.rs-982658/v1

Abstract

Since 2013, the sugarcane aphid (SCA), *Melanaphis sacchari* (Zehntner), has been a serious pest that hampers all types of sorghum production in the U.S. Our understanding of sugarcane aphid resistance in sorghum is limited to knowledge about a few genetic regions on chromosome SBI-06. In this study, a subset of the Sorghum Association Panel (SAP) was used along with some additional lines to identify genetic and genomic regions that confer sugarcane aphid resistance. SAP lines were grown in the field and visually evaluated for SCA resistance during the growing seasons of 2019 and 2020 in Tifton, GA. In 2020, the SAP accessions were also evaluated for SCA resistance in the field using drone-based high throughput phenotyping (HTP) and visual scoring under greenhouse conditions. Plant height and flowering time were also recorded in the field to confirm that our methods were sufficient for identifying known quantitative trait loci (QTL). This study combined phenotypic data from field-based visual ratings, reflectance data, and greenhouse evaluations to identify genome-wide associated (GWAS) marker-trait associations (MTA) using genotyping-by-sequencing (GBS) data. Several MTAs were identified for sugarcane aphid-related traits across the genome, with a few common markers that were consistently identified on SBI-08 and SBI-10 for aphid count and plant damage as well as for reflectance indices-based traits on SBI-02, SBI-03, and SBI-05. Candidate genes encoding leucine-rich repeats (LRR), Avr proteins, lipoxygenases (LOXs), calmodulins (CAM) dependent protein kinase, WRKY transcription factors, flavonoid biosynthesis genes, and 12-oxo-phytodienoic acid reductase are identified near SNPs that had significant associations with different SCA traits. In this study, plant height and flowering time-related genes were also identified. The total phenotypic variation explained by significant SNPs across SCA-scored traits, plant height, and flowering time ranged from 0 to 74%, while the heritability value ranged from 4 to 74%. These results supported the existing literature, and also revealed several new loci. Markers identified in this study will support marker-assisted breeding for sugarcane aphid resistance.

Introduction

Sorghum (*Sorghum bicolor* (L.) Moench) is the fifth most important cereal crop worldwide, and it is used for food, feed, fiber, and fuel[1,2]. Farmers in the United States use sorghum mainly as livestock feed and silage while the grain is also utilized for bird feed and to produce ethanol[3]. Sweet sorghum, with large amounts of free sugars in the stem juice, can potentially be used as a next-generation biofuel feedstock for ethanol and bioenergy production[4]. The United States is the world's largest producer and lead exporter of grain sorghum, having harvested 1.9 million hectares (4.7 million acres) and having exported 6.99 million metric tons (275 million bushels) in 2020. Importers of U.S. sorghum included Mexico, China, Japan, Ethiopia, and New Zealand[5]. Grain sorghum flour is also gluten-free and has potential nutritional and health benefits[6]. Its wide agro-climatic adaptation results from its high phenotypic diversity, which enabled its expansion from the Sahel desert into temperate and subtropical regions of Asia and the Americas[7].

Since 2013, sorghum farmers in the U.S. have suffered yield losses caused by a new invasive pest, the sugarcane aphid (SCA), *Melanaphis sacchari* (Zehntner)[8]. Although still unclear how this new super-clone came to the U.S.[9,10], it was first discovered near Beaumont, TX on grain sorghum[11]. By the end of 2013, sugarcane aphids had colonized sorghum across four states, including Texas, Louisiana, Mississippi, and Oklahoma[8]. Sorghum in these areas was so heavily infested that combines would clog during harvest due to the sticky honeydew that aphids had secreted on the plant material[12]. Currently, sugarcane aphids have spread throughout the U.S. and are found in all areas that grow sorghum.

Plant damage caused by sugarcane aphids includes decreased grain yield, reduced seed weight, lower grain quality, increased lodging[8], and even aborted panicle emergence. Integrated pest management practices such as seed treatments, varying planting time, use of insecticides, natural enemies, and host plant resistance are considered useful for the control of sugarcane aphids[8,13–16]. Currently there are only two pesticides labeled for the control of sugarcane aphids in sorghum, both with similar modes of action: Sivanto Prime (flupyradifurone; Bayer Crop Science) and Transform WG (sulfoxaflor; Corteva).

Host plant resistance plays a significant role in reducing sugarcane aphid population dynamics and can reduce plant damage and yield loss compared to susceptible hybrids[14,17]. Despite many sorghum inbred lines that have been identified to be resistant to sugarcane aphids[11,18–22], very few resistance genes have been identified. A single dominant nuclear resistance locus named *RMES1* (Resistance to *Melanaphis sacchari* 1) on SBI-06 was associated with sugarcane aphid resistance[23]. Mapping in a Henong16 (resistant) x BTx623 (susceptible) population localized sugarcane aphid resistance to a 126 kb region containing five predicted genes, three of which are leucine-rich repeat (LRR) proteins[24]. Wang *et al.*[25] later identified the same region as conferring resistance in a different mapping population (Tx2783 x BTx623). Knoll and Cuevas[26] identified the same *RMES1* region from a SC112-14 (resistant) x PI 609251 (susceptible) recombinant inbred line (RIL) population. Zhang *et al.*[27] identified four QTLs for sugarcane aphid resistance on SBI-06 from resistant source 407B. One of these QTLs, qtlMs-6.1, overlaps with the *RMES1* locus.

Assessing insect damage in a large field experiment can be very slow and laborious, and visual ratings may be subjective. The need for high-throughput phenotyping (HTP) has spurred advances in using image processing, robotics, and drone-based methods to assess insect resistance in crop plants. Outputs from remote sensing applications provide fast and accurate forecasting of targeted insect pests[28], pest damage[29–31], and pest monitoring[32–36]. While satellite-borne sensors have been useful to monitor broad regional and global cropping systems, multi- and hyperspectral sensors held at closer distances (from a few centimeters to a few hundred meters) provide extremely fine resolution data associated with plant responses to stressors[34, 37], including aphid stress[38]. In these cases, sensors may be borne on low-altitude aircraft, including unmanned aerial systems (UAS, also known as drones), mounted on fixed towers, tethered on cables, or held by hand. In the last several years, UAS borne sensors have become ubiquitous in agricultural research, presenting a booming frontier for research and data collection to answer research questions surrounding crop pest and disease detection and monitoring, as well as insecticide and fungicide applications worldwide. One complication for sugarcane aphid population estimation is that the aphids feed on the abaxial (underside) leaf surface. Therefore, the direct visualization of aphids with current high-throughput platforms may not be easy, even at close range.

The Sorghum Association Panel (SAP) is a collection of accessions that represent all the major cultivated races and U.S. sorghum breeding lines[39]. Our objective was to examine the hypothesis that there may be multiple genomic regions associated with aphid resistance other than the *RMES1* locus and other QTL identified on SBI-06. To identify new regions of the sorghum genome that confer resistance to the sugarcane aphid, we used 276 accessions from the existing SAP database plus 7 additional inbred lines with resistance to sugarcane aphid or other insects, to perform genome-wide association (GWAS) for sugarcane aphid resistance using different insect-related phenotyping traits such as the number of aphids, plant damage rating, and other physio-morphological traits. To facilitate the rapid assessment of plant damage across all accessions we used UAS multi-spectral imagery to corroborate the damage assessed through visual inspection.

Results

For the field trials, sugarcane aphid population number, damage, plant height, and flowering time were recorded for two years in 2019 and 2020. In 2019 and 2020, the first alate aphids (winged aphids) were observed on sorghum on July 11 and July14, respectively. Aphid population increased gradually in both years before the population collapsed in the second half of August each year. Aphid counts and damage ratings were started at peak of the aphid population growth, recorded as first ratings on August 14, 2019, and August 11, 2020, while the second ratings were taken on August 28, 2019 and August 18, 2020, respectively. The sugarcane aphid damage on the SAP panel was rated twice in each year and the ratings were averaged across both years (“D1” and “D2” for the first and second rating dates, respectively).

General statistical summary

Aphid damage on plants was rated using a 1-9 scale, with 1 describing plants that showed no aphid-induced plant damage and 9 describing plants that had heavy aphid infestation up the flag leaf and 80% of the leaves displayed aphid damage[40]. There were 5 resistant lines that scored between 1.0-3.0 and 12 susceptible lines that scored 8.0-9.0 scale (Supplementary Table S1-a). There were two susceptible checks (N98 & N109B) and two resistant checks (No. 5-Gambela and SC110). The average aphid damage score over two seasons for the susceptible checks ranged from 2.3 to 6.3, while for the resistant checks aphid damage rating ranged from 1.0-3.2 (Supplementary Table S1-b).

Most SAP accessions were susceptible to aphid damage, with a mean damage score of 3.8 for D1 and 5.9 for D2 (Supplementary Table S1-c). Aphid induced plant damage for D1 and D2 ranged from 1.0-7.0 and 2.5-9.0, respectively. The average aphid count over the two seasons for the first-rating dates (AC1) ranged from 48 to 1500 aphids leaf^{-1}, while the average of the second-rating dates aphid count (AC2) ranged from 0 to 1166 aphids leaf^{-1} (see Methods and abbreviations for details). For the SAP lines grown in the greenhouse assessed for aphid-induced plant damage (GHD) at the seedling stage, the average damage rating was 8.5 and the range varied from 5.0-9.0.

Remotely sensed imagery from a UAS-borne sensor was used to calculate three vegetation reflectance indices, namely, Normalized Difference Red-Edge (NDRE), Normalized Difference Vegetation Index (NDVI), and Soil Adjusted Vegetation Index (SAVI). Wall-to-wall coverage of the index values was used to score aphid damage more accurately across the entire field. Only the UAS-based data collected on August 25, 2020, was used for analysis as it had a high correlation to manual phenotyping and high heritability values. The NDRE value ranged from 0.12 (high aphid damage) to 0.51 (low or no aphid damage) with a mean of 0.28. The NDVI scores ranged from 0.16 to 0.78 with a mean of 0.46, while the SAVI score ranged from 0.07 to 0.41 with a mean of 0.16 (Supplementary Table S1-c).

The average plant height for the two growing seasons was 106 cm, ranging from 39 to 285 cm (Supplementary Table S1-c). Flowering time ranged from 41 to 97 days after planting, with an average of 59 days after planting.

Correlation analysis

The correlation between reflectance indices from 2020 and average aphid damage ratings from 2019 and 2020 were calculated. A significant and negative linear relationship was observed between aphid damage for the first-rating date in 2020 and UAS based reflectance indices with NDRE (r = -0.41), NDVI (r = -0.38), and SAVI (r = -0.35) (Supplementary Table S1-d). Similarly, the second-rating date for visual aphid damage in 2020 also showed a relatively strong negative correlation with NDVI (r = -0.48), SAVI (r = -0.50), and NDRE (r = -0.52). The average first-rating dates (D1) for aphid damage for 2019 and 2020 were also negatively correlated with NDVI (r = -0.43), SAVI (r = -0.41), and NDRE (r = -0.44). A relatively strong negative linear relationship was observed between the second-rating dates (D2) for aphid damage for 2019 and 2020 with NDVI (r = -0.55), SAVI (r = -0.56), and NDRE (r = -0.58). The correlation between reflectance indices and the first average aphids count (AC1) ratings showed a negative linear relationship, whereas the second aphid count ratings (AC2) showed a positive relationship. A relatively strong positive linear relationship was observed between AC1 and D1 (r = 0.62), as well as AC1 and D2 (r = 0.51) (Supplementary Table S1-d), suggesting that the higher the aphid population, the higher the aphid damage.

Heritability

High broad-sense heritability values were observed for plant height (H^2 = 0.74) and days to anthesis (H^2 = 0.69) (Supplementary Table S1-e). Aphid damage ratings for D1 and D2 had moderate heritability values (H^2 = 0.31 and 0.34, respectively), whereas aphid damage ratings from the greenhouse experiment had lower heritability (H^2 = 0.22). Of the aphid traits analyzed, the aphid count data had the lowest heritability at H^2 = 0.16 for AC1 and 0.04 for AC2. Moderate to high heritability was observed for the drone data recorded on August 25, 2020. Heritability values for NDVI, NDRE, and SAVI for August 25, 2020, were 0.68, 0.56, and 0.36, respectively.

GWAS Marker-Trait Associations (MTA)

Genome-wide association was performed using two complementary methods. First, we performed single-marker regression with a generalized linear model (GLM), using genetic principal coordinates to correct for population structure in TASSEL[41]. Significant hits were identified by using 1000 permutations to establish empirical p-values, with empirical p≤0.01 considered as significant. In a complementary analysis, we used Fixed and random model Circulating Probability Unification (FarmCPU), a model-selection algorithm that includes both fixed and random effects components[42]. Preliminary analyses found individual runs of FarmCPU to be unstable, so the algorithm was rerun100 times with 90% subsampling to identify the most robust hits, scored by their “resample model inclusion probability” (RMIP)[43] (see Methods for details). The RMIP values of 0.05 or greater were considered significant[44]. Genomic regions identified by both methods—meaning low empirical p-values and high RMIPs—are high-confidence hits; those identified by only one are still significant but of lower confidence. For those traits that did not pass through significant thresholds and also showed weak signals in the above two methods, GWAS was also performed with FarmCPU in GAPIT package from R software without any iterations. GWAS analysis identified a total of 200 MTAs for different aphid resistance-related traits, flowering time, and plant height which are discussed below (Supplementary Table S2-a).

AC1-Aphid Count first rating dates and AC2-Aphid Count second rating dates

FarmCPU-RMIP model selection analysis identified 11 loci for AC1(**Fig. 1**, Supplementary Table S2-a). Three loci each were located on SBI-07 and SBI-09, while the remaining loci were distributed with a single SNP on SBI-01, SBI-02, SBI-06, SBI-08, and SBI-10. Among these markers, S8_11781182 showed a strong association (RMIP=0.41) with AC1. All SNPs that were significantly associated with AC1 explained 55% of total phenotypic variation (Supplementary Table S3). Two markers, S7_58778565 and S7_58839694 significantly associated with AC1, are located close (~479 kb) to the gene Sobic.007G151100 on SBI-07, which encodes a protein that is similar to 12-oxo-phytodienoic acid reductase, a gene expressed in response to sugarcane aphid attacks[45]. There were no significant associations identified for AC2 in either GLM or RMIP methods.

D1-First aphid damage ratings

FarmCPU-RMIP identified nine total SNPs with two SNPs each on SBI-07 and SBI-10 and a single SNP on SBI-02, SBI-04, SBI-05, SBI-06, and SBI-08 (**Fig. 1**, Supplementary Table S2-a). GLM analysis did not identify any significant MTAs for aphid damage D1 at permutation p-value ≤ 0.01. Among nine loci detected here using RMIP, the MTA on SBI-06 (S6_334458) showed the highest RMIP (13%) value (Supplementary Table S2-a). The total phenotypic variance explained by all SNPs associated with the first aphid damage rating dates-D1 was 32% (Supplementary Table S3).

D2-Second aphid damage ratings

Use of the RMIP method detected 14 SNPs and GLM method detected 6 SNPs associated with aphid damage D2 (**Fig. 1**; Supplementary Table S2-a). These 20 MTAs were detected on all chromosomes except SBI-02 and SBI-04, with the strongest associations on SBI-01, SBI-06, SBI-07, SBI-08, and a cluster of SNPs on the tail end of SBI-09 with higher RMIP and p values. The portion of phenotypic variance explained by significantly associated SNPs was 63% (Supplementary Table S3). GLM analysis detected one MTA on SBI-06 (S6_3050581) approximately 298 kb away from the *RMES1* locus. The S6_53836196 marker is located ~99 kb away from Sobic.006G184300 gene on SBI-06 that encodes a phloem protein 2 gene known to be induced due to sugarcane aphid infestation[46]. RMIP identified two SNPs, S9_56657642, and S9_56678671, which were < 1Mb away from WRKY transcription factor SbWRKY86, Sobic.009G238200 (S9_57628850 to 57630763), which was differentially expressed in a gene expression study[45]. The SNP S8_11781182 showed a strong association with damage D2 in both GLM and RMIP analyses.

Greenhouse evaluation of aphid damage-(GHD)

There were no significant associations found for aphid damage in the greenhouse, except on SBI-07 using the FarmCPU RMIP method (**Fig. 1**). The marker on SBI-07, S7_2494675 was identified with RMIP value= 0.09 (Supplementary Table 2-a).

GWAS analysis for reflectance indices to assess SCA damage (with and without flowering time as a covariate)

Analysis of UAS reflectance data was performed both with and without flowering time as a covariate in case plant maturity had a significant impact on the observed spectra.

A total of 6 SNPs were associated with NDRE when flowering time was used as a covariate in the model selection analysis. Three of these SNPs were located on SBI-06, while a single SNP was found on SBI-01, SBI-02, and SBI-05. No significant SNP associated with NDVI was detected when the flowering time was used as a covariate. A single SNP marker (S9_3054933) was significantly associated with SAVI when flowering time was used as a covariate (**Fig. 2**, Supplementary Table S2-a).

We also did analysis without using flowering time as a covariate, and the RMIP analysis identified 20 SNPs significantly associated with NDRE. Of these, four SNPs were found on SBI-05, three on SBI-03, SBI-06, and SBI-08, two on SBI-01, SBI-02, and SBI-10, and a single SNP on SBI-07 (Supplementary **Fig. S1** and Supplementary Table S2-a). There were four markers, namely, S1_5101469, S2_61431704, S6_58686833, and S6_60922873, that were significantly associated with NDRE and found in either case with or without using flowering time as a covariate and the rest 16 SNPs were new loci when flowering time was not used as covariate. Some of the important notable ones were S6_32034639 marker was found ~366 kb of the sugarcane aphid resistance QTL, qtlMs-6.2, located between 31,020,371 to 32,400,770 bp[27]. Among several loci identified, SNP S8_59192389 associated with NDRE is approximately10 kb away from the calmodulin, (CAM) gene, Sobic.008G159100 (SBI-08: 59203243 to 59204321), a differentially expressed gene family observed by Kiani & Szczepaniec[45].

A total of 8 SNPs were associated with NDVI without using flowering time as a covariate, with two SNPs on SBI-01, SBI-02, and SBI-05, while a single SNP each on SBI-03 and SBI-06 (Supplementary **Fig. S1**; Supplementary Table S2-a). The S6_3383155 showed a strong association (RMIP=0.23) with NDVI and was located 631 kb away from *RMES1* locus.

A single SNP marker (S9_3054933) associated with SAVI on SBI-09 was also found in either case with or without using flowering time as a covariate. NDRE associated markers explained 21-54% of the phenotypic variance, NDVI associated markers accounted for 41% of phenotypic variance and the single marker associated with SAVI accounted for 6% of the phenotypic variance (Supplementary Table S3). SNPs on SBI-05 at 3,230,535 bp and 3,282,948 bp were remarkably closer to each other (52.4 kb) and associated with NDRE and NDVI respectively. A common SNP on SBI-05 at 63,115,845 bp was associated with NDRE and NDVI.

Plant Height

FarmCPU-RMIP identified 12 SNPs with a RMIP ≥ 0.05, and these were located on all chromosomes except SBI-04, SBI-05 and SBI-10 (**Fig. 3**, Supplementary Table S2-a). The most SNPs were found on chromosome SBI-09 with 4 SNPs that colocalized to the *Dw1* locus (Sobic.009G229800) (19 kb to 131 kb away). There were two SNPs identified on SBI-06 that were close to the *Dw2* locus, about 863 kb to 1.58 Mb away. There was one SNP, S7_59614079 that was located within the *Dw3* locus and was 3 kb away from the start position for this gene. There were two SNPs on chromosome SBI-01 and single SNPs on chromosomes SBI-02, SBI-03 and SBI-08. The strongest association of SNP with plant height was identified on SBI-09 (S9_57172609; $p = 1.41 \times 10^{-15}$; RMIP=0.48). The GLM results identified 78 markers on chromosomes SBI-06 and SBI-09 that colocalized to the *Dw1* and *Dw2* loci with markers as close as 9kb (S9_57051085) to 1.5 Mb from these loci. Markers on other chromosomes were not detected. These SNPs explained 74 % of phenotypic variance (Supplementary Table S3). All

these loci were detected in previous studies compiled in Sorghum QTL Atlas by Mace *et al.*[47] except two locations. The two SNPs, S2_6406243 on SBI-02 and S8_45638928 on SBI-08 were found to be new loci associated with plant height in sorghum from this study.

Maturity/Flowering time

GLM analysis identified two SNPs that are located on SBI-03 in close proximity (S3_4423241, and S3_4423302) with high confidence (post-permutation p-value ≤ 0.01) (**Fig. 3**, Supplementary Table S2-a). FarmCPU model selection identified 17 SNPs across all chromosomes except SBI-05 and SBI-10, including a highly associated SNP at the end of SBI-03, a cluster of SNPs on SBI-06, and a strongly associated SNP at the end of SBI-09. These SNPs explained 63 % of the total phenotypic variance for maturity (Supplementary Table S3). One MTA on SBI-01 (S1_7584419) was ~1.5 Mb away from the known maturity gene *Ma5* (*PHYC*)[48, 49], while S6_40987299 was 0.7 Mb away from major maturity locus *Ma1* (*SbPRR37*)[50]. The other loci identified in this study are also found in the vicinity of QTLs reported in Sorghum QTL Atlas and from other studies[47, 51], with the ones on SBI-03 having the most significant associations. Some of the loci were not detected in previous studies compiled in the QTL atlas of Mace *et al.*[47] and are reported here as novel loci identified in this study as they were very far from the known maturity loci (Supplementary Table S2-a). There were 6 new loci and these were at least 10 Mb away from known maturity loci or at least 100-200 kb away from the known existing QTLs. There were two loci on SBI-01 (S1_30327933 and S1_30462377 that were 30 Mb away from *Ma3*), two loci on SBI-06 (S6_50581550 and S6_50582469, 10 Mb away from *Ma1*), and one single locus on SBI-02 (S2_2488178, 65 Mb away from *Ma2*) and on SBI-07 (S7_2559101).

FarmCPU analysis without iteration (single-run)

For traits that did not yield any MTAs by GLM or RMIP analysis, we noticed that running FarmCPU with default parameters (a single run with no resampling) still generated a few hits near known candidate genes. These hits were apparently not robust enough to be picked up in the resampling analysis, and we include them here for the sake of completion and under the assumption that they probably indicate real signal that simply fell below our statistical threshold (**Fig. S2** & Supplementary Table S2-a).

AC2 Aphid count for second-rating dates using FarmCPU single-run

Single-run FarmCPU analysis identified a total of 7 SNPs significantly associated with the second rating dates of aphid count (AC2); three SNPs closely (<100 kb apart) located on SBI-04, one SNP on SBI-05 and three on SBI-06 were associated with AC2. Three SNPs, namely, S6_3381010, S6_3381051, and S6_3381094, were remarkably close to each other and (**Fig. S2** & Supplementary Table S2-a) were approximately 629 kb away from the Sb06g001650 gene encoding for the *RMES1* locus.

Greenhouse evaluation of aphid damage (GHD) rating using FarmCPU single-run

A total of 13 SNPs with a significant association to GHD ratings were identified using FarmCPU single-run analysis. Three SNPs were detected on SBI-01, two each from SBI-02 SBI-03, and SBI-09, and single SNP each detected on SBI-04, SBI-05, SBI-06, and SBI-07 (**Fig. S2** & Supplementary Table S2-a). The total phenotypic variance explained by all SNPs of FarmCPU single-run for GHD rating was 30% (Supplementary Table S3). The RMIP and GLM analysis detected none or low phenotypic variation as only one significant SNP associated with GHD rating was identified (Supplementary Table S3). The S7_2494675 marker, significantly associated with GHD, was consistently found across RMIP and FarmCPU single-run data analyses. FarmCPU single-run analysis also detected a marker (S4_53265404 on SBI-04) associated with greenhouse aphid damage at 23 kb away from Sobic.004G180500 gene that encodes for 12-oxo-phytodienoic acid.

Genomic regions/clusters commonly identified for aphid-related traits

Using FarmCPU and GLM analyses, overlapping genetic regions (within 500 kb) were consistently identified for aphid population number, aphid damage, and reflectance indices. These clusters found across different traits were sorted based on their chromosomal position (Supplementary Table S2-b). We found a few regions on SBI-02, SBI-03, SBI-05, SBI-08 and SBI-10 that were consistently found across different traits. The SNPs, S8_11781182 on SBI- 08, and S10_ 2507813 on SBI-10 were consistently identified and associated with AC1 and plant damage D2 respectively (Supplementary Table S2-a & b). It is also noteworthy to observe that SNP S8_11408106 associated with aphid damage D1 was 373 kb away from this common genomic region, S8_11781182 on SBI-08. Additionally, three common SNPs, S2_61431704,

S3_19558428, and S5_63115845 on SBI-02, SBI-03 and SBI-05 respectively were associated with the reflectance indices, NDRE and NDVI consistently. Another SNP associated with NDRE, S5_63115742 was remarkably close within 103 base pairs of consistently identified region S5_63115845 for NDRE and NDVI. This study identified two markers (S6_3050581 and S6_3383155) on SBI-06 that were associated with D2 and AC2 found in proximity (298-629 kb away) of *RMES1* locus (Sb06g001650). The markers associated with D1 (S6_179752) and AC1 (S6_334458) were 154 kb apart from each other and 2.3 Mb away from *RMES1* locus (Supplementary Table S2-b). The greenhouse GHD (S2_551409) and NDVI (S2_887719) associated SNPs on SBI-02 were closely located within 336 kb distance. The SNP marker associated with NDRE (S3_19929613) was within 371 kb from a marker, S3_19558428, consistently identified on SBI-03. There were three traits, GHD (S7_2494675), flowering (S7_2559101) and AC1(S7_2888228) that were within 64-393 kb which may show that flowering time and aphid resistance genes were found in a cluster on SBI-07. Aphid damage, D1(S8_11408106) was remarkably close (373 kb) to the common genomic region S8_11781182 that associated with damage D2 and AC1. The D1 SNP S8_11408106 is located 124 kb away from a gene, Sobic.008G075700, encoding LRR. SNPs associated with flowering time (S8_39703004) and NDRE (S8_39703090) were within 86 base pairs in another location. SNPs associated with plant height (S9_56530072) and D2 (S9_56705459) were within 174 kb on SBI-09. Plant height (S9_57804067) and flowering loci (S9_57914426) were identified within 109 kb on SBI-09.

Identification of differentially expressed known SCA genes within 200 kb of MTAs

There are two studies that have done in-depth analysis of differential gene expression in sorghum upon sugarcane aphid infestation[45, 46]. Our gene analysis was done exclusively to identify the candidate genes around MTAs identified in this study using genes expressed in these previously conducted SCA differential gene expression studies[45, 46]. The genes from these two studies were searched if they were found within 200 kb distance upstream and downstream of the markers identified in our study. The genes that were detected two times due to their association with closely linked markers were removed and only unique genes were retained.

Of the 9,291 unique genes identified by Kiani and Szczepaniec,[45] that were differentially expressed in resistant and susceptible lines after exposure to the sugarcane aphid, 561 (6%) unique genes were detected within 200 kb distance of the 74 markers associated with sugarcane aphid-related traits (Supplementary Table S4). Of these, eight genes encoding leucine-rich repeats (LRR) were found on SBI-05, SBI-06, and SBI-07. The SNP marker, S6_334458, associated with aphid damage (D1), is 5 kb away from Sobic.006G001900 encoding a LRR gene that was upregulated during aphid infestation in the previous study[45]. Four genes near NDRE associated SNPs on SBI-03, SBI-06 and SBI-08 were similar to lipoxygenase (LOXs), and two genes on SBI-05 and SBI-07 encoded flavonol biosynthesis genes (Supplementary Table S4). The NDVI marker, S1-7442679, was found within ~99 kb of the Sobic.001G0955 gene that encodes a *WRKY1* transcription factor.

Additionally, we pooled gene datasets obtained from Tetreault *et al.*[46] that were detected within a 200 kb distance of MTAs. We analyzed genes that were induced in both resistant and susceptible plants upon sugarcane aphid infestation using 16 gene-expression modules in this study. Out of 15074 genes evaluated, 911 (6%) unique genes were within 200 kb region of sugarcane aphid associated SNPs (Supplementary Table S5). The SNP markers associated with aphid damage D2, S6_3050581 is 147 kb away from Sobic.006G018900 encoding protein similar to avr9/Cf-9 rapidly elicited response protein. S2_61431704 associated with NDRE is 339 base pairs away from Sobic.002G222800 encoding protein similar to Putative Avr9 elicitor response protein. The analysis identified twenty-five LRR, four LOXs, four CAM dependent protein kinases, two WRKY transcription factors, and one flavonol 3-O-glucosyltransferase.

In total, twenty-nine LRR, four CAM proteins, four lipoxygenase (LOXs), two WRKY, and two flavanol biosynthesis genes were identified from the above two studies (all supplementary files of Kiani and Szczepaniec,[45] and 16 selected modules of Tetreault *et al.*[46] (Supplementary Table S4 and S5). When all the modules from the Tetreault *et al.*[46] study were used for candidate gene analysis, one 12-oxophytodienoic acid reductase gene (Sobic.004G180500) was found very close within 23 kb from S4_53265404. The Sobic.007G151100 encoding 12-oxophytodienoic acid reductase was also found within 483 kb to a marker, S7_58778565, associated with aphid count (AC1). We also found three LRR genes on SBI-05 and SBI-07, and one WRKY1 gene on SBI-01, which were commonly found in the studies by Kiani and Szczepaniec,[45] and Tetreault *et al.*[46].

We also analyzed the number of genes and gene frequency on individual sorghum chromosomes found near our SNPs associated with different aphid traits, by using these two datasets[45, 46]. The highest gene frequency distribution was observed on SBI-01 followed by SBI-06 (Supplementary **Fig. S3-a & c**). The majority of these genes were captured by markers associated with NDRE and D2 aphid traits (Supplementary **Fig. S3-b & d**).

The markers in the current study that showed more than 10 genes were considered as genic-rich regions as they may reveal a cluster of genes around them. Our result indicated that a total of 59 markers were associated with more than 10 genes found in the above two gene expression

data[45, 46]. These clusters ranged from 11-32 candidate genes for each of these markers within 200 kb that might have promising genic-rich regions/clusters related to sugarcane aphid resistance (Supplementary Table S6).

Candidate gene analysis near consistently identified markers

The candidate gene analysis was focused near common and consistent markers identified between different traits on SBI-02, SBI-03, SBI-05, SBI-08 and SBI-10 using previously known SCA genes[45, 46](Supplementary Table S4 and S5). A total of 25, 3, 1, 1, and 32 genes were found near these consistently identified markers, S2_61431704, S3_19558428, S5_63115742, S8_11781182 and S10_2507813, respectively. Some of the noted ones were Sobic.005G158000; oxidative stress response genes-similar to Cytochrome P450 71E1 found on SBI-05 and Sobic.008G077600- similar to Diacylglycerol kinase 1, found on SBI-08 as they were narrowed down to single genes from the previous gene expression studies around these markers. S2_61431704 marker was close to Avr9 elicitor response protein found in both studies. Among 32 genes found near S10_2507813, several LRR repeats were detected and one of them, Sobic.010G030000 was within 77 kb.

Discussion

Genetic and molecular basis of aphid resistance among agriculturally important crops has often been associated with the involvement of large-effect R genes, highlighting gene-for-gene Flor's hypothesis[52–55]. But often we also see the small effects of several other genes/QTLs that contribute to aphid resistance in sorghum[56]. Sugarcane aphid resistance is not completely understood owing to its recent emergence of pest status on sorghum in North America. Some studies have attempted to understand the molecular basis of this interaction in sorghum as monogenic, though it is believed to be oligogenic and polygenic[23–25, 56]. Most of the studies highlighted the importance of *RMES1* locus located on SBI-06 and the causative variants around this region to be the major resistance gene for sugarcane aphid resistance in sorghum depending on the source of resistance used in these studies[24, 25, 27, 56]. SCA resistance is monogenic (single locus) for RME6 ACK60 x RTx2783. However, in other sources it could be polygenic.

This study utilized 283 accessions from SAP population grown in the field and greenhouse to identify SNPs associated with sugarcane aphid resistance in sorghum. It combined phenotypic data from two years of field-based visual ratings with one year of reflectance data, and greenhouse evaluations to identify genome-wide associated (GWAS) marker-trait associations (MTA) using genotyping-by-sequencing (GBS) data. GWAS analysis identified a total of 200 markers for 8 different aphid resistance-related traits, flowering time, and plant height (Supplementary Table S2-b). MTAs were identified through GWAS studies that used three different methods such as RMIP in FarmCPU, GLM and single-run FarmCPU. These methods were robust in identifying high quality and high confidence GWAS-QTLs with stable genomic positions for MTAs as flowering time was used as a covariate in much of the analysis. Using flowering time as a covariate reduced any significant spurious associations with flowering time genes and separated them from most of the other MTAs. The significant associations were further explored to identify candidate genes that are known to influence sugarcane aphid resistance in sorghum. We noticed that the SNPs associated with the reflectance indices such as NDVI were not detected when flowering time was used as covariate. We also found that flowering time SNPs did not collocate with these reflectance indices indicating flowering time genes were different from NDVI SNPs. When flowering time was not used as covariate, NDVI identified significant MTA's that had aphid related biological significance. At this time, it is not clear if flowering morphology would impact these traits.

The GWAS analysis identified 96 markers related to sugarcane aphid resistance traits. After removing commonly identified duplicate markers associated with aphid-related traits, about 85 markers were at unique locations. Among these markers, UAS reflectance indices (NDRE, NDVI and SAVI) captured 28 markers, which accounted for 33% of the aphid related traits. The remaining 57 markers (67%) were captured by visually recorded SCA traits (D1, D2, AC1, AC2, and GHD). In the current study, the high throughput phenotyping has enhanced our ability to tag aphid resistance and plant health traits and also supported visual scoring with a few common markers. The phenotypic variance explained by these markers varied from 0-63% for all aphid resistance-related traits (Supplementary Table S3). Among the 96 MTAs (manual scoring, greenhouse, and drone) markers, S1_14448039 on SBI-01 explained the highest phenotypic variance for aphid damage D2. The S8_11781182 marker found on SBI-08 had higher RMIP and GLM permutation values for D2 traits. These markers could be potential indicators of metabolic pathways or hormonal signaling defense pathways for aphid resistance. A significant correlation was observed between UAS reflectance indices and visually recorded aphid traits rating indicating that both methods quantify the damage that occurred on sorghum plants due to SCA infestation. The negative correlation of NDVI, SAVI, and NDRE with aphid damage ratings indicated that plant stress increased with increasing plant injury. Higher values for NDVI, SAVI, or NDRE indicated lower stress levels, while lower NDVI, SAVI, or NDRE values were associated with higher plant stress due to aphid damage. This study confirms with previous studies[57, 58]. Further study may be needed to identify the best index to analyze the impact of the sugarcane aphid pests on crop yield and biomass.

Among aphid resistance traits, the maximum number of markers,16, was identified on SBI-06. Recent studies have specifically dissected regions around the *RMES1* locus and found several NBS-LRR containing genes that were associated with resistance gene products[25, 27, 56]. In our investigation, we were able to locate LRR motifs and Avr9 elicitor response protein within 200 kb region for markers associated with aphid damage and aphid count on SBI-06. Recently, Muleta *et al.*[56] also found regions on SBI-06 and other chromosomes for sugarcane aphid resistance and our study also found these regions to be involved with aphid damage, aphid count, greenhouse aphid damage evaluation and UAS-based reflectance vegetation indices. When we compared our markers with these *FSt* outlier markers for sugarcane aphid resistance, 34 markers were found to be in close proximity (within 500 kb) to 40 markers identified in Muleta *et al.*[56] (Supplementary Table S7). These markers were as close as 0 to 356 kb between the two studies and most of these markers came from SBI-06- and SBI-09.

There are only a few studies that have elucidated the changes in gene expression upon sugarcane aphid infestation in sorghum. In our study, we were able to detect 6% of known unique genes from these two datasets related to sugarcane aphid response using our MTAs identified. Our candidate gene search was limited to these two datasets from the existing two experiments, and our window to examine these genes was only 200 kb within significant markers[45, 46]. Therefore, we also looked at some potential candidate genes that were beyond 200 kb and some genes that may not be reported in these two experiments.

Kiani and Szczepaniec,[45] reported that 2-week-old and 6-week-old resistant and susceptible plants orchestrated several defense pathways to sugarcane aphids by induction of hormone-signaling pathways coding for secondary metabolites, glutathione metabolism, and plant-pathogen interaction. Also, genes associated with the metabolic process, biosynthetic process, macromolecule process, and oxidation-reduction were upregulated in the younger resistant plants as enriched GO terms. Genes related to the Ca^{2+} sensor group, which encodes calmodulin-like (CMLs) and calmodulin (CAM) proteins showed up-regulation in response to aphid injury regardless of sorghum genotype or developmental stage[45]. Our UAS-based result indicated that the marker identified on SBI-08, SNP S8_59192389 associated with NDRE is approximately 10 kb away from a gene that encodes a CAM family protein, Sobic.008G159100 (SBI-08: 59203243 to 59204321), a differentially expressed gene family observed by Kiani and Szczepaniec, [45]. In this study markers such as S1_27239199, S4_53265404, and S5_3230535, associated with greenhouse, NDRE, NDVI, and aphid count are located within 200 kb away from the genes encoding calmodulin-binding family proteins (Supplementary Table S4). These genes are expressed upon plant infestation by aphids[45, 46, 59].

Tetreault *et al.*[46] found in general that aphid infestation induced global gene expression changes in the susceptible line that reflected elevated levels of herbivory-related stress and cessation of plant growth through deployment of plant defense genes such as nucleotide-binding-site leucine-rich-repeats (NBS-LRR), peroxidases, glutathione S-transferases, laccases and several phloem proteins, wound responsive genes and WRKY transcription factors. We found S6_53836196 that was very close to phloem proteins found in the above study[46]. We found at least 29 LRR regions that were all associated with aphid resistance traits (Supplementary Table S4 and S5). Most of the LRRs found in this study were on SBI-05, SBI-06, and SBI-07. There was one LRR family containing protein on SBI-08, Sobic.008G075700 which was ~500 kb away from S8_11781182 that was consistently found between GLM and RMIP with higher thresholds. This region from S8_11408106 to S8_11781182 (~ 373 kb) was consistently found for aphid damage D1, D2 and aphid count AC1 traits and the same region was also in close proximity (127 to 245 kb) to markers found by Muleta *et al.*[56]. The S6_53836196 marker associated with D2 was located ~99 kb away from Sobic.006G184300 gene on SBI-06 that encodes a phloem protein 2 gene and this gene showed an increased gene expression due to sugarcane aphid infestation[46].

Zhu-Salzman *et al.*[60] reported that treatments of sorghum (*Sorghum bicolor*) seedlings with methyl jasmonate resulted in a reduced number of greenbugs aphids (*Schizaphis graminum*) compared with the control plants, suggesting the significance of jasmonic acid (JA)-pathway-mediated defense in sorghum against aphids. Grover *et al.*[61] found the involvement of phytohormone 12-oxo-phytodienoic acid (enzyme,12-oxophytodienoate reductase), a precursor involved in jasmonic acid (JA) biosynthesis pathway as the main volatile component involved in aphid resistance in the sorghum SCA-tolerant SC35 plants. In the present investigation, we found two locations, S4_53265404 (GHD) was ~29 kb away, and S7_58778565 (D1 and AC1) was 479 kb away from this gene, 12-oxo-phytodienoic acid reductase, Sobic.007G151100. This enzyme, 12-oxo-phytodienoate reductase, is involved in recognizing aphid attacks and induces JA hormonal response. The early recorded traits such as AC1 and D1 (immediately after aphids were observed) and GHD rating response reveals early onset response and the markers, S4_53265404 and S7_58778565 associated with these traits (GHD, D1 and AC1) tagged to 12-oxo-phytodienoic acid suggest that these genes are the first line of defense when aphids attack.

Lipoxygenase (LOXs) genes are known to play a significant role in aphid resistance in sorghum and other crops[62–64]. Shrestha *et al.*[65] performed a genome-wide analysis of the sorghum lipoxygenases (LOXs) genes and identified the expression of nine LOXs genes. In this study, we found four lipoxygenase genes (Sobic.003G385500, Sobic.003G385900, Sobic.006G248300, and Sobic.008G157900) within 200 kb of S3_69633709, S6_58686833, and S8_59192389 markers on SBI-03, SBI-06, and SBI-08. Among nine 9-LOXs found by Shrestha *et al.*[65], *SbLOX3* (Sobic.003G385500) and *SbLOX4* (Sobic.003G385900) were found to be duplicated genes and induced in early response to

sugarcane aphids. The marker, S3_69633709, associated with NDRE, was found within 163 kb and 193 kb of *SbLOX3* and *SbLOX4*, respectively. *SBLOX9* was found to play an important role in resistant plants in early response to sugarcane aphid[65]. There were a few markers close to *SbLOX9* on SBI-06 but they did not pass significant thresholds set in GLM and/or RMIP methods. The LOX genes identified in our study were also reported by Kiani and Szczepaniec,[45] and Tetreault *et al*.[46] (Supplementary Table S4 and S5).

In a recent study, the combination of whole- genome resequencing and transcriptome analysis showed that differential gene expressions (DEGs) from the resistant source were enriched in 'DNA replication' and 'flavonoid biosynthesis'[27]. In our study, two flavonoid biosynthesis genes (Sobic.005G033801 and Sobic.007G059200) were identified within 200 kb of S5_3230535 and S7_6208175 markers on SBI-05 and SBI-07. Sobic.010G091100, located 408 kb away from S10_8350074 marker encodes flavonol-3-O-glycoside-7-O-glucosyltransferase 1 (Supplementary Table S4 and S5). These genes were reported in previous experiments[45, 46, 27]. The S6_32034639 marker associated with NDRE was found approximately 366 kb of the aphid resistant QTL, qtlMs-6.2, located between 31,020,371 to 32,400,770 bp[27].

The use of FarmCPU and GLM were robust and were able to identify previously identified dwarfing and maturity genes. Of the 200 total MTAs, 85 markers were associated with plant height, and 19 markers were associated with flowering time. Most of these loci were found on SBI-06 and SBI-09 and are discussed here in the context of existing literature on plant height and flowering. Of the four major dwarfing genes controlling plant height in sorghum designated as *Dw1* (Sobic.009g229800: 57,038,652-57,041,166 bp), *Dw2* (Sobic.006G067700: 42803036-42807520), *Dw3* (Sobic.007g023730: 59,822,089-59,830,560 bp) and *Dw4* (SBI-04: position unknown), only *dw1, dw2*, and *dw3* were detected in the current study (Supplementary Table S2-a)[66–71]. Morris *et al*.[72] also detected *dw1*, *dw2*, and *dw3* in a collection of 971 sorghum accessions (which included the SAP) using GWAS. Similarly, flowering time is controlled by many genes in sorghum of which the maturity genes *Ma1-Ma6* are the best characterized[73–77]. Of these six loci, *Ma1* is known to be the major locus controlling maturity and has greatest sensitivity to LD (Long Day) conditions and is closely linked to *Ma6* and *Dw2* on SBI-06[50]. In our study, we found several flowering time QTLs/genes across seven chromosomes. These were located on SBI-01, SBI-02, SBI-03, SBI-06, SBI-07, SBI-08, SBI-09. Of these, a marker linked to *Ma1* on SBI-06 was detected in this study within close proximity (Supplementary Table S2-a)[50]. Mace *et al*.[78] in a study of the Sorghum BC-NAM population, revealed some weak QTL on SBI-03, none on SBI-06, and a strong one on SBI-09 in the same place as the strong RMIP-identified MTAs in this study. The majority of these loci agree with the existing literature on flowering time genes in sorghum[47]. The identification of height and maturity loci to known regions from existing literature supported that our methods used in this study were robust and sufficient to resolve GWAS QTLs. This study has also identified six new genomic locations associated with flowering time genes that did not show in other studies to the best of our knowledge.

In this study, SNP markers on SBI-08 at 11,781,182 bp and SBI-10 at 2,507,813 bp were consistently identified as being associated with aphid count and aphid damage and SNP markers S2_61431704, S3_19558428, and S5_63115845 for reflectance indices-based traits were identified consistently on SBI-02, SBI-03, and SBI-05 respectively. The candidate gene search around these markers showed more on LRR genes, Avr9 elicitor response protein, kinases and oxidative stress response genes. These pleiotropic regions may harbor potential candidate genes for SCA resistance.

SAP represented in this study has shown 5 highly resistant lines and 12 highly susceptible lines and these accessions may represent potential sources for SCA resistance among 283 lines used. Most of the studies conducted to-date for SCA resistance have limited number of parental sources of resistance and reveal more on *RMES1* locus. This study is the first to assess SCA phenotyping in the field for two years using a large panel. This study reveals that SCA resistance may be controlled by different resistance genes based on the resistant source.

Conclusion

This study combined phenotypic evaluations from eight different traits conducted in the field and greenhouse to assess sugarcane aphid resistance in sorghum. In this research, UAS-based reflectance data was shown to be a valuable HTP tool for the rapid assessment of sugarcane aphid damage. The use of GWAS using the SAP population rated for sugarcane aphid resistance traits identified the *RMES1* locus and its surrounding cluster of genes but also new regions on SBI-02, SBI-03, SBI-05, SBI-08 and SBI-10 that were consistently found across different traits. The methods employed in GWAS were robust in stabilizing the genomic positions of MTA. Furthermore, leveraging existing gene expression data on sugarcane aphid resistance allowed us to identify candidate genes that are closely located within the significant markers identified in this study. Moreover, some of these genes were known to be involved in aphid resistance. This work has provided several markers that could be useful in marker-assisted selection for developing sugarcane aphid resistant varieties. Functional validation and positional cloning are needed to prove their involvement in sugarcane aphid resistance. Also, in the future, understanding of positive and negative epistatic interaction among major QTLs is needed to unravel the genetic architecture of sorghum for sugarcane aphid resistance breeding. This study has identified five resistant (1-3 score) lines from the SAP that could be valuable for future breeding. Resistant sources that contain these new sugarcane aphid resistance regions could be utilized for developing sugarcane aphid resistant varieties of sorghum.

Materials And Methods

Plant Materials

All methods were performed in accordance with the relevant guidelines/regulations/legislation.

Plant materials consisted of 276 lines from the SAP[39] and seven additional lines that have known insect resistance/susceptibility traits. The full list of plant accessions used in this experiment and their source is in Supplementary Table S8.

SAP accessions (283 in 2019 and 267 in 2020) were planted in a randomized complete block design with two replications in Tifton, GA. Three panicles in each plot of the first replicate in 2019 were bagged to produce seed for the next year. The 16 fewer lines in 2020 were due to a lack of seed from severe aphid damage in 2019 preventing seed set. Planting dates were on June 14, 2019, and on June 9, 2020. Each accession was planted in a one-row plot 4 m long with 0.9 m between rows.

Phenotyping

Visual Scoring

Sugarcane aphid population number, aphid-induced plant damage, and plant growth-stage were visually evaluated for each plot during the peak sugarcane aphid infestation and then evaluated for the same traits two weeks later. Evaluations in 2019 occurred on Aug 14 and Aug 28, and in 2020 on Aug 11 and Aug 18. Three plants per plot were chosen at random and aphid number was estimated from the second leaf from the top and lowest green leaf from the bottom. Counts were recorded using a scale of A to F, where A<25, B=25-49, C=50-99; D=100-499, E=500-1000, and F>1000 aphids[79]. Damage and growth stage were rated based on the entire plot. Plant damage was rated using a 1-9 scale where 1 = no leaf damage, 2 = 1-20% of the leaves are damaged; 3 = 21-30%, 4 = 31-40%, 5 = 41-50%, 6 = 51-60%, 7 = 61-70%, 8 = 71-80%, and 9 = greater than 80% of the leaves are damaged [40]. The growth stage of the sorghum plants was classified in five stages as vegetative, boot, flowering, soft dough, and hard dough stage.

Agronomic Traits

For each plot, flowering date, plant height, and stand count (the number of plants per plot) were recorded. Flowering date was evaluated three times each week, and a plot was scored as flowering on the date when half the plot has released pollen. Plant height (measured from the soil to the top of the panicle) was measured at maturity for three plants per plot.

High throughput phenotyping using UAS

The remote sensing datasets used in this study consisted of multispectral imagery data collected over sorghum field trials in Tifton, GA, during six field visits (six flight dates) from July through September 2020. Only the data from Aug. 25 were used in this study as it coincided with visual scoring dates and had higher correlation (Supplementary Table S1-d). The imagery was collected with a Micasense RedEdge 5-band sensor using a DJI Matrice M100 UAS. Spectral bands collected were blue (B, 475 +/-20nm), green (G, 560 +/- 20nm), red (R, 668 +/- 10nm), red edge (RE, 717 +/- 10nm), and near infrared (NIR, 842 +/- 40nm). To georectify the image mosaics, ground control points (GCPs) were established at the beginning of the study and marked with a temporary landmark, using a Trimble Geo7X global navigation satellite system receiver with real-time kinematic correction (i.e., RTK), providing +/- 2 cm horizontal accuracy at the coordinate location. Prior to each flight, the GCPs were targeted, and calibration images were collected using Micasense calibration protocols. Pix4DMapper software (v. 4.6.4; Pix4D SA, Switzerland) was used to produce orthorectified mosaics, photogrammetric point clouds and surface elevation models, georeferenced to within a satisfactory root mean square error (RMSE). The final spatial resolution of the mosaicked images was 2.64 cm. The UAS was flown with an iPad Mini using Pix4DCapture flight control software, and flight parameters were ~38 meters altitude, at 4.92 m s^{-1} (11 mi h^{-1}), resulting in 80% side and front overlap in the RedEdge camera imagery.

The three vegetation indices calculated for this study were Normalized Difference Vegetation Index (NDVI), Normalized Difference Red-Edge (NDRE), and Soil Adjusted Vegetation Index (SAVI). The formulae for NDVI, NDRE and SAVI areas follows:

NDVI = (NIR – R)/(NIR + R)

NDRE = (NIR – RE)/(NIR + RE)

SAVI = (1.5*(NIR-R))/(NIR+R+0.5)

ArcGIS software (Esri ArcMap v. 10.7.1, advanced) Spatial Analyst toolbox was used for index calculation and geospatial processing. Mean and standard deviation values were calculated using ArcMAP zonal statistics (table) tool.

Greenhouse Assessment of Plant Damage-GHD

Sugarcane aphids were obtained from a single parthenogenic female originally collected from infested grain sorghum in Matagorda County TX in August 2013. These sugarcane aphids were phenotyped and genotyped and determined to be invasive biotype in the U.S., which belongs to the Multilocus lineage (MLL)-F[80]. Sugarcane aphids were reared on Tx7000 seedlings in 4.4-L pots fitted with 45-cm-tall x 16 cm diameter cylinders of Lexan™ (SABIC Polymershapes, Tulsa, OK) sleeve cages and ventilated using organdy cloth to cover the top of the cylinder. Plants and aphids were maintained in the greenhouse at temperatures between 21 °C and 31 °C under natural greenhouse light supplemented by two T-8 fluorescent lights. Sugarcane aphid colonies were transferred to new seedling plants every two weeks.

The SAP sorghum entries were evaluated for resistance to sugarcane aphids using a free-choice flat screen trial where seedlings were infested and rated for aphid damage, plant height, number of true leaves, and chlorophyll content. Included for comparison was the known SCA-resistant sorghum Tx2783, as well as known susceptible hybrids KS 585, and Tx7000[11, 81]. The entire SAP collection was evaluated in batches of 30 SAP lines until all the SAP lines were evaluated. For each batch, four flats (Growers Supply, Dyersville, IA 52042) consisting of plastic trays with 128 individual cells were planted with 10 replications of a set of 30 SAP entries plus the known resistant and susceptible checks in a randomized complete block design. When the seedlings reached the 2 leaf-stage they were heavily infested with sugarcane aphids by cutting infested Tx7000 (above) at the soil surface and placing plants down and across each row within the flat[82]. After roughly two weeks (when the susceptible checks KS 585 and Tx7000 were 90-100% dead), the measurements of plant height, number of true leaves, chlorophyll content, and damage ratings were recorded. Damage ratings were conducted using a 1-9 scale; where 1 is a completely healthy plant with no necrotic tissue; 2 represents 1-5% chlorotic tissue; 3, 5-20%; 4, 21-35%; 5, 36-50%; 6, 51-65%; 7, 66-80%; 8, 81-95%; and 9, 95-100% or dead[83, 84]. Chlorophyll content was measured using SPAD-502 chlorophyll meter (Minolta, Ramsey, NJ 07466) where chlorophyll type *a* and *b* are summed for total content[85]. The top three leaves from each sorghum entry were measured for chlorophyll content and averaged for each entry. Each sorghum entry was cut at the soil surface so plant height could be measured (cm) and the number of leaves could be counted for both infested and non-infested entries. The variables of damage rating, plant height, number of leaves, and chlorophyll content were averaged across the 10 replications and the mean was used in the GWAS analysis.

SNP Data

Genotype data from the SAP was obtained from Morris *et al.*[72] and Hu *et al.*[86] which generated SNPs using Genotyping-by-Sequencing (GBS). For the seven additional accessions included for this study, DNA was obtained from each line using a GeneJET Plant Genomic Purification Mini Kit (Thermo Fisher Scientific, Waltham, MA) and was shipped to the University of Minnesota for GBS processing, using *Ape*KI restriction enzyme with no size selection. Raw GBS reads were aligned to the *S. bicolor* v3.1 genome downloaded from Phytozome (https://phytozome.jgi.doe.gov/) using bwa version 0.7.12-r1039 (command “bwa aln”)[87]. The SAP genotypes[86] were used to determine genome sites to call SNPs at, and SNPs were called using bcftools version 1.9 (commands “bcftools mpileup” with max depth of 1000 and minimum base quality of 20, followed by “bcftools call –multiallelic-caller” with the option to skip calling indels)[88]. After standardizing names between the datasets, the two datasets were combined, filtered for only biallelic sites, and imputed with BEAGLE 4.1 (21Jan17 release)[89]. Code for SNP joining is available on GitHub at https://github.com/wallacelab/paper-punnuri-aphids-2021.

Data analysis

Phenotypic data analysis

The sugarcane aphid plant damages were evaluated by averaging the first aphid damage ratings (D1) recorded in each year (Aug 14, 2019 and Aug 11, 2020). The second visually recorded aphid damage (D2) was the average of the second rating on Aug 28, 2019, and Aug 18, 2020. Similarly, aphid counts (AC1 and AC2) were estimated by averaging data collected on the same days, which correspond to 61 and 75 days after planting (DAP) in 2019, and 63 and 70 DAP in 2020. Flowering time and plant height were similarly averaged over both available years. UAS-based vegetative indices (NDVI, NDRE, SAVI) were only available for one year.

The frequency distributions of each trait were visualized using histograms in JMP Pro (JMP®, Version <2013>. SAS Institute Inc., Cary, NC, 1989–2021).

Plant height, aphid damage, and aphid count traits were log (x+1) transformed to achieve normality and homoscedasticity following Shapiro–Wilk, and Brown–Forsythe, and Levene tests[90]. Correlation analysis was performed in JMP using pairwise method to evaluate the relationship between different traits of interest.

Variance components were estimated by restricted maximum likelihood (REML) using an all-random model in the VARCOMP procedure of SAS v. 9.4 (SAS Institute, Cary, NC). Variance component estimates were used to calculate broad sense heritability (H^2) of quantitative traits, including sugarcane aphid damage, plant height and flowering time, by dividing genetic variance by the total phenotypic variance according to the following formulas.

Broad-sense heritability

across years $H^2 = \sigma^2_g / (\sigma^2_g + \sigma^2_{gy} + \varepsilon)$

single year $H^2 = \sigma^2_g / (\sigma^2_g + \varepsilon)$

where σ^2_g is the genotypic variance, σ^2_{gy} is the genotype by year variance, and ε is the residual error variance. Heritability values for UAS-based reflectance indices for specific date, Aug 25, 2020 is reported as it coincided with visual scoring dates and had a higher correlation than the other dates (Supplementary Table S1-e).

Genome-wide Association Analysis (GWAS)

From 281 Sorghum Association Panel (SAP) accessions, 458,313 SNPs were obtained of which 191,738 SNPs with a MAF ≥ 0.05 were selected and used for GWAS analysis. There were two accessions that did not have GBS data from our collection of 283 lines. Analyzed traits consisted of visual field assessment of sugarcane aphid damage scores (D1 and D2), estimated aphid count (AC1 and AC2), greenhouse visually scored aphid damage (GHD), UAS-based reflectance data recorded on Aug 25, 2020, for Normalized Difference Vegetation Index (NDVI), Normalized Difference Red Edge (NDRE) and Soil Adjusted Vegetation Index (SAVI), plant height (PH), and flowering time (FL) performed using Fixed and random model Circulating Probability Unification (FarmCPU). For traits that were not normally distributed such as aphid count and damage scores, plant height, and flowering time, log transformation was applied.

To control the population structure, the first three principal components (PCs) were used as covariates for all traits. In addition, flowering time was used as a covariate for aphid damage, aphid counts and UAS-based vegetative indices to remove associations due to plant maturity, and nSPAD readings were used as covariates for greenhouse visually scored aphid damage.

GWAS with TASSEL-GLM 1000

GWAS was performed in TASSEL v5.2.73[41] using a generalized linear model (GLM) with 1000 permutations to determine statistical significance. (The ability to rapidly perform permutations was the reason for choosing GLM over a mixed linear model with a kinship matrix. See the section on Model Selection, below, for a complementary GWAS analysis that includes a kinship matrix.) Population structure was

controlled by including the first three genetic principal coordinates as covariates, calculated using the distance matrix and MDS functions in TASSEL.

GWAS with FarmCPU-Model selection RMIP-method

To complement the GLM analysis above, model-selection GWAS was performed with FarmCPU[42] as part of the GAPIT3[91,92] software package for R. The same genotype file was used for all traits, and the model included three genetic principal coordinates and a kinship matrix calculated using the VanRaden method[93] with the function GAPIT.kinship.VanRaden().

Early tests found that the FarmCPU results, like many model selection algorithms[94], were unstable when six samples were accidentally excluded from the analysis and generated very different lists of significant loci. To overcome this instability, we performed model selection in two steps. First, a null distribution of results was created by randomly permuting the genotypes relative to the phenotypes 100 times. Kinship was not permuted to maintain the covariance with phenotypes, and a random 10% of samples were excluded each permutation to match expected power with the following resampling step. The results from this step formed a null distribution of p-values and allowed us to determine which results are more significant than due to chance.

The second step consisted of running the actual (non-permuted) genotypes 100 times, each time randomly excluding 10% of samples. Resampling allowed us to generate Resample Model Inclusion Probabilities (RMIPs)[43], a measure of how stable each genomic association is. SNPs with RMIP scores above 0.05 (meaning 5 out of 100 models) were considered significant[44].

Combined Manhattan plots of GLM and FarmCPU results were created using the R packages ggplot2 v3.3.3[95], ggpubr v0.4.0[96], gridExtra v2.3[97], and cowplot v1.1.1[98]. All code for the FarmCPU analysis is available on GitHub at https://github.com/wallacelab/paper-punnuri-aphids-2021.

Phenotypic Variance Explained

To estimate the variance explained by each marker (or set of markers), we used GAPIT3[91,92] to calculate a kinship matrix using the VanRaden method[93] and the prcomp() function in R to derive genetic principal components (PCs). For each phenotype, we fit a model including the first three PCs, the kinship matrix, and any covariate phenotypes (e.g., flowering time), as with the GWAS analyses. The residuals from that model were then fit in a second linear regression with the SNP(s) of interest to determine the fraction of residual variance explained in the form of the model R^2. Code for this analysis is available on GitHub at https://github.com/wallacelab/paper-punnuri-aphids-2021.

GWAS with single-run FarmCPU

For those traits that did not pass-through significant thresholds and also showed weak signals in the above two methods, GWAS was also performed with FarmCPU in GAPIT package from R software without any iterations. The results from this analysis are displayed in the supplementary files and were only performed in the case of aphid count second-dates rating (AC2) and greenhouse evaluation data as they showed genes relevant to sugarcane aphid herbivory on sorghum. The results need to be further validated as no iterations were conducted in this method.

Candidate gene search from previous gene expression studies

Candidate genes were also selected based on data available on differentially expressed genes (DEGs) between resistant and susceptible sorghum from two previous experiments[45, 46]. There were sixteen modules from Tetreault *et al.*[46] that were used in identifying potential candidate genes for sugarcane aphid infestation. These module numbers were 1, 2 ,4, 5, 6, 7, 9, 10, 12, 14, 15, 16, 17, 18, 19, and 23 which were known to be involved in sugarcane aphid infestation response. DEGs within 200 kb of a MTA (either GLM or FarmCPU resampling) were considered potential candidate genes. Gene distances and annotations were based on the *Sorghum bicolor* genome v3.1.1 available on Phytozome (https://phytozome.jgi.doe.gov/).

The code used throughout this article is available at the GitHub repository:

Graphics and figures were generated in R, and the scripts used to do so are available on GitHub at https://github.com/wallacelab/paper-punnuri-aphids-2021.

Abbreviations

AC1-First rating dates- estimate of aphid counts per leaf (The average of estimated aphid counts recorded on Aug 14, 2019, and Aug 11, 2020)

AC2-Second rating dates- estimate of aphid counts per leaf (The average of estimates of aphid counts recorded on Aug 28, 2019, and Aug 18, 2020)

CAM–Calmodulin

D1-First rating dates- visually recorded aphid damage (The average aphid damage rating recorded on Aug 14, 2019, and Aug 11, 2020)

D2-Second rating dates- visually recorded aphid damage (The average aphid damage rating recorded on Aug 28, 2019, and Aug 18, 2020)

DAP–Days After Planting

FarmCPU- Fixed and random model Circulating Probability Unification

FL-Flowering

GBS–Genotyping-by-Sequencing

GCPs-Ground Control Points

GHD-Greenhouse Damage

GLM-General linear model

GWAS–Genome-Wide Association Study

HTP–High-throughput Phenotyping

LOX–Lipoxygenase

LRR–Leucine-Rich Repeat

MTA–Marker-Trait Association

NDVI-Normalized Difference Vegetation Index

NDRE-Normalized Difference Red Edge

NIR–Near Infrared

PCs–Principal Components

PH-Plant Height

QTL–Quantitative Trait Locus/loci

R-Red

RE – Red edge

RMIP-Resample Model Inclusion Probability

SAP–Sorghum Association Panel

SAVI-Soil Adjusted Vegetation Index

SBI-*Sorghum bicolor* chromosome

SCA–Sugarcane Aphid

SNP–Single Nucleotide Polymorphism

UAS–Unmanned Aerial System

Declarations

Acknowledgements

The authors would like to thank Trevor Perla, Michael Purvis, and Tyler Bailey for technical support, as well as Barb Driskel, Hannah Baker and Cheyenne Edwards for running the greenhouse phenotyping trials. We would like to thank the USDA NIFA Capacity Building grant program, award no 2019-38821-29057 for funding support. We would like to thank USDA, Agricultural Research Service, National Plant Germplasm System and GRIN resources for providing SAP seeds continuously throughout the study period. Also, we would like to thank Dr. Xinzhi Ni for providing seeds for the first year of the SAP planting. The research was also supported in part by the Non-Assistance Cooperative Agreement 58-6048-6-016, which was funded by the United States Department of Agriculture – Agricultural Research Service with the research project entitled “Areawide Pest Management of Invasive Sugarcane Aphid in Grain Sorghum (62048-21220-018-00D).”

Author contributions

S.P., K.H.S., J.K., and J.W. conceptualized and designed the experiment for this manuscript.

S.P., A.A., K.H.S. J.K., and T.W. collected data from the field plots.

S.P., A.A., K.H.S., J.K., H.L., S.S., and J.W. performed the data analysis and interpretation.

A.C., and H.T. flew drones over the test plots and analyzed the resulting data.

S.A. grew the SAP collection in greenhouse, inoculated the SAP with sugarcane aphids, and collected all greenhouse data.

S.P. acquired funding.

Corresponding author

Somashekhar Punnuri

Ethics declarations and Competing interests

The authors declare no competing interests.

References

1. Doggett, H. *Sorghum. Harlow, Essex* (Longman Scientific & Technical, England, 1988).
2. Venkateswaran, K., Elangovan, M. & Sivaraj, Natarajan, S. “Origin, Domestication and Diffusion of Sorghum bicolor,” Breeding Sorghum for Diverse End Uses. in Aruna, C., Visarada, K.B.R.S., Bhat, B. V., Tonapi, V. A. Eds. (Cambridge, United Kingdom:Woodhead Publishing),15–31(2019).

3. Dahlberg, J. The role of sorghum in renewables and biofuels. In Humana Press (Ed.), Sorghum Methodos Protoc, 1, 269–277(2019). https://doi.org/10.1007/978-1-4939-9039-9_19.
4. Ratnavathi, C., Chakravarthy, S. K., Komala, V. V., Chavan, U. D. & Patil, J. V. Sweet sorghum as feedstock for biofuel production: a review. *Sugar Tech*, **13**, 399–407 https://doi.org/10.1007/s12355-011-0112-2 (2011).
5. Sorghum Checkoff. 2021. Sorghum Checkoff Annual Report. https://www.sorghumcheckoff.com/news-and media/newsroom/2021/02/25/2020-annual report/; last accessed 10 March 2021.
6. Xiong, Y., Zhang, P., Warner, R. D. & Fang, Z. Sorghum grain: from genotype, nutrition, and phenolic profile to its health benefits and food applications. *Compr. Rev. Food Sci. Food Saf*, **18**, 2025–2046 https://doi.org/10.1111/1541-4337.12506 (2019).
7. Ziska, L. H. *et al.* Food security and climate change: on the potential to adapt global crop production by active selection to rising atmospheric carbon dioxide. Proceedings of the Royal Society B, 279, 4097–4105(2012). https://doi.org/10.1098/rspb.2012.1005.
8. Bowling, R. D. *et al.* Sugarcane aphid (Hemiptera: Aphididae): A new pest on sorghum in North America. *J. Integr. Pest Manag*, **7**, 1–13 https://doi.org/10.1093/jipm/pmw011 (2016a).
9. Harris-Shultz, K. *et al.* Microsatellite markers reveal a predominant sugarcane aphid (Homoptera: Aphididae) clone is found on sorghum in seven states and one territory of the USA. *Crop Sci*, **57**, 2064–2072 https://doi.org/10.2135/cropsci2016.12.1010 (2017).
10. Nibouche, S. *et al.* Invasion of sorghum in the Americas by a new sugarcane aphid (Melanaphis sacchari) superclone. *PloS ONE*, **13**, e0196124 https://doi.org/10.1371/journal.pone.0196124 (2018).
11. Armstrong, J. S. *et al.* Sugarcane aphid (Hemiptera: Aphididae): host range and sorghum resistance including cross-resistance from greenbug sources. *J. Econ. Entomol*, **108**, 576–582 https://doi.org/10.1093/jee/tou065 (2015).
12. Villanueva, R. T. *et al.* Sugarcane aphid: a new pest of sorghum. Texas A&M Agrilife Extension. Ento-035, CollegeStation, TX. 4 p(2014).
13. Jones, N., Brown, S., Williams, S., Emfinger, K. & Kerns, D. Efficacy of neonicotinoid seed treatments against sugarcane aphid in grain sorghum. *Arthropod Manag. Tests*, **40**, https://doi.org/10.1093/amt/tsv139 (2015).
14. Brewer, M. J. *et al.* Sugarcane aphid population growth, plant injury, and natural enemies on selected grain sorghum hybrids in Texas and Louisiana. *J. Econ. Entomol*, **110**, 2109–2118 https://doi.org/http://doi10.1093/jee/tox204 (2017).
15. Haar, P. *et al.* Evaluation of tactics for management of sugarcane aphid (Hemiptera: Aphididae) in grain sorghum. *Journal of Economic Entomology*, **112**, 2719–2730 https://doi.org/10.1093/jee/toab013 (2019).
16. Harris-Shultz, K., Knoll, J., Punnuri, S., Niland, E. & Ni, X. Evaluation of strains of Beauveria bassiana and Isaria fumosorosea to control sugarcane aphids on grain sorghum. *Agrosyst Geosci Environ*, **3**, e20047 https://doi.org/10.1002/agg2.20047 (2020).
17. Szczepaniec, A. Interactive effects of crop variety, insecticide seed treatment, and planting date on population dynamics of sugarcane aphid (Melanaphis sacchari) and their predators in late-colonized sorghum. *Crop Prot*, **109**, 72–79 https://doi.org/10.1016/j.cropro.2018.03.002 (2018).
18. Armstrong, J. S., Paudyal, S., Limaje, A., Elliott, N. & Hoback, W. Plant resistance in sorghums to the sugarcane aphid (Hemiptera: Aphididae). *J. Entomol. Sci*, **53**, 478–485 https://doi.org/10.18474/JES17-106.1 (2018).
19. Hayes, C. M. *et al.* Registration of R. LBK1 and R. LBK2 sorghum germplasm with resistance to the sugarcane aphid [Melanaphis sacchari (Zehntner)]. *J. Plant Regist*, **13**, 91–95 https://doi.org/10.3198/jpr2018.06.0038crg (2019).
20. Limaje, A. *et al.* Antibiosis and tolerance discovered in USDA-ARS sorghums resistant to the sugarcane aphid (Hemiptera: Aphididae). *J. Entomol. Sci*, **53**, 230–241 https://doi.org/10.18474/JES17-70.1 (2018).
21. Peterson, G. C., Armstrong, J. S., Pendleton, B. B., Stelter, M. & Brewer, M. J. Registration of RTx3410 through RTx3428 sorghum germplasm resistant to sugarcane aphid [Melanaphis sacchari(Zehntner)]. *J. Plant Regist*, **12**, 391–398 https://doi.org/10.3198/jpr2018.02.0007crg (2018).
22. Paudyal, S. *et al.* Categories of resistance to sugarcane aphid (Hemiptera: Aphididae) among sorghum genotypes. J. Econ. Entomol. 112, 1932-1940(2019a). https://doi.org/10.1093/jee/toz077.
23. Chang, J. *et al.* Analysis of the resistance gene to the sorghum aphid, Melanaphis sacchari, with SSR markers in Sorghum bicolor. *Acta Prataculturae Sin*, **115**, 113–118 https://doi.org/10.1016/S2095-3119(12)60101-4 (2006).
24. Wang, F. *et al.* Efficient and fine mapping of RMES1 conferring resistance to sorghum aphid Melanaphis sacchari. *Mol Breed*, **31**, 777–784 https://doi.org/10.1007/s11032-012-9832-6 (2013).
25. Wang, B. *et al.* Pan-genome analysis in sorghum highlights the extent of genomic variation and sugarcane aphid resistance genes. *Preprint at bioRXiv*, https://doi.org/10.1101/2021.01.03.424980 (2021).
26. Knoll, J. E. & Cuevas, H. E. Genome Mapping of Sugarcane Aphid Resistance in Sorghum Line SC112-14 [Abstract]. ASA, CSSA and SSSA International Annual Meetings, San Antonio, TX(2019).

27. Zhang, J. *et al.* Whole-genome resequencing and transcriptome analysis provide insights on aphid-resistant quantitative trait loci/genes in Sorghum bicolor., **140**, 618–629 https://doi.org/10.1111/pbr.12946 (2021a).
28. Abd El-Ghany, N. M., El-Aziz, A., Marei, S. S. & S. E., & A review: application of remote sensing as a promising strategy for insect pests and diseases management. *ESPR*, **27**, 33503–33515 https://doi.org/10.1007/s11356-020-09517-2 (2020). http://doI
29. Hall, R. J., Castilla, G., White, J. C., Cooke, B. J. & Skakun, R. S. Remote sensing of forest pest damage: a review and lessons learned from a Canadian perspective. *Can. Entomol*, **148**, S296–S356 https://doi.org/10.4039/tce.2016.11 (2016).
30. Lan, Y., Huang, Y., Martin, D. E. & Hoffmann, W. C. Development of an airborne remote sensing system for crop pest management: system integration and verification. *Applied Engineering in Agriculture*, **25**, 607–615 https://doi.org/http://doi 10.13031/2013.27458 (2009).
31. Stone, C. & Mohammed, C. Application of remote sensing technologies for assessing planted forests damaged by insect pests and fungal pathogens: a review. *Current Forestry Reports*, **3**, 75–92 https://doi.org/10.1007/s40725-017-0056-1 (2017). https://DOI
32. Bhattarai, G. P., Schmid, R. B. & McCornack, B. P. Remote sensing data to detect Hessian Fly infestation in commercial wheat fields. *Sci. Rep*, **9**, 6109 https://doi.org/10.1038/s41598-019-42620-0 (2019).
33. Samseemoung, G., Jayasuriya, H. P. W. & Soni, P. Oil palm pest infestation monitoring and evaluation by helicopter-mounted, low altitude remote sensing platform. *J. Appl. Remote Sens*, **5**, https://doi.org/10.1117/1.3609843 (2011).
34. Singh, D., Sao, R. & Singh, K. P. A remote sensing assessment of pest infestation on sorghum. *Advances in Space Research*, **39**, 155–163 https://doi.org/10.1016/j.asr.2006.02.025 (2007).
35. Yue, J. W., Lei, T. J., Li, C. C. & Zhu, J. Q. The application of unmanned aerial vehicle remote sensing in quickly monitoring crop pests. *Intelligent Automation and Soft Computing*, **18**, 1043–1052 https://doi.org/10.1080/10798587.2008.10643309 (2012).
36. Zhang, J. C. *et al.* Monitoring plant diseases and pests through remote sensing technology: A review. *Computers and Electronics in Agriculture*, **165**, 04943 https://doi.org/10.1016/j.compag.2019.104943 (2019).
37. Malthus, T. J. & Madeira, A. C. High-Resolution spectroradiometry-spectral reflectance of field bean-leaves infected by Botrytis-Fabae. *Remote Sens. Environ.*, **45** (1), 107–116 https://doi.org/10.1016/0034-4257(93)90086-d (1993). https://do
38. Stanton, C. *et al.* Unmanned aircraft system-derived crop height and normalized difference vegetation index metrics for sorghum yield and aphid stress assessment. *Journal of Applied Remote Sensing*, **11**, 026035 https://doi.org/10.1117/1.JRS.11.026035 (2017).
39. Casa, A. M. *et al.* Community resources and strategies for association mapping in sorghum. *Crop Sci*, **48**, 30–40 https://doi.org/10.2135/cropsci2007.02.0080 (2008).
40. Sharma, H. C. Munghate, R. S. & Sharma, S. P. Phenotyping for resistance to the sugarcane aphid Melanaphis sacchari (Hemiptera: Aphididae) in Sorghum bicolor (Poaceae). *Int. J. of Trop. Insect Sci*, **33**, 227–238 https://doi.org/10.1017/S1742758413000271 (2013).
41. Bradbury, P. J. *et al.* TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, **23**, 2633–2635 https://doi.org/10.1093/bioinformatics/btm308 (2007).
42. Liu, X., Huang, M., Fan, B., Buckler, E. S. & Zhang Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLOS Genet*, **12**, e1005767 https://doi.org/10.1371/journal.pgen.1005767 (2016).
43. Valdar, W., Holmes, C. C., Mott, R. & Flint, J. Mapping in structured populations by resample model averaging. *Genetics*, **182**, 1263–1277 https://doi.org/10.1534/genetics.109.100727 (2009).
44. Wallace, J. G. *et al.* Association mapping across numerous traits reveals patterns of functional variation in maize. *PLOS Genet*, **10**, e1004845 https://doi.org/10.1371/journal.pgen.1004845 (2014).
45. Kiani, M. & Szczepaniec, A. Effects of sugarcane aphid herbivory on transcriptional responses of resistant and susceptible sorghum. *BMC Genom*, **19**, 774 https://doi.org/10.1186/s12864-018-5095-x (2018).
46. Tetreault, H. M. *et al.* Global responses of resistant and susceptible sorghum (Sorghum bicolor) to sugarcane aphid (Melanaphis sacchari). *Front Plant Sci*, **10**, 145–145 https://doi.org/10.3389/fpls.2019.00145 (2019).
47. Mace, E. *et al.* The Sorghum QTL Atlas: a powerful tool for trait dissection, comparative genomics and crop improvement. *Theor Appl Genet*, **132**, 751–766 https://doi.org/10.1007/s00122-018-3212-5 (2019).
48. Yang, S. *et al.* Sorghum phytochrome B inhibits flowering in long days by activating expression of SbPRR37 and SbGHD7, repressors of SbEHD1, SbCN8 and SbCN12. PLoS ONE 9, e105352(2014). https://doi.org/10.1371/journal.pone.0105352.
49. Faye, J. M. *et al.* A genomics resource for genetics, physiology, and breeding of West African sorghum. *Plant Genome*, **14**, e20075 https://doi.org/10.1002/tpg2.20075 (2021).
50. Murphy, R. L. *et al.* Coincident light and clock regulation of pseudo response regulator protein 37 (PRR37) controls photoperiodic flowering in sorghum. Proc. Natl. Acad. Sci. USA. 108,16,469–16,474(2011). doi:10.1073/pnas.1106212108.
51. Mural, R. *et al.* Meta-analysis identifies pleiotropic loci controlling phenotypic trade-offs in sorghum. *Genetics. 218*, https://doi.org/10.1093/genetics/iyab087 (2020).

52. Thompson, G. A. & Goggin, F. L. Transcriptomics and functional genomics of plant defence induction by phloem-feeding insects. *J Exp Bot*, **57**, 755–766 https://doi.org/10.1093/jxb/erj135 (2006).
53. Smith, C. M. & Boyko, E. V. The molecular bases of plant resistance and defense responses to aphid feeding: current status. *Entomol Exp Appl*, **122**, 1–16 https://doi.org/10.1111/j.1570-7458.2006.00503.x (2007).
54. Howe, G. A. & Jander, G. Plant immunity to insect herbivores. *Annu Rev Plant Biol*, **59**, 41–66 https://doi.org/10.1146/annurev.arplant.59.032607.092825 (2008).
55. Dogimont, C., Bendahmane, A., Chovelon, V. & Boissot N. Host plant resistance to aphids in cultivated crops: Genetic and molecular bases, and interactions with aphid populations. *C.R. Biol*, **333**, 566–573 https://doi.org/10.1016/j.crvi.2010.04.003(2010 (2010).
56. Muleta, K. T. & e *t al*. The recent evolutionary rescue of a staple crop depended on over half a century of global germplasm exchange. *Preprint at bioRxiv* (2021). https://doi.org/10.1101/2021.05.11.443651.
57. Elliott, N. C. *et al*. NDVI to detect sugarcane aphid injury to grain sorghum. *J. Econ. Entomol*, **108**, 1452–1455 https://doi.org/10.1093/jee/tov080 (2015).
58. Zhang, J. *et al.* Assessing spatio-temporal patterns of sugarcane aphid (Hemiptera: Aphididae) infestations on silage sorghum yield using unmanned aerial systems (UAS). *Crop Prot*, **146**, 105681 https://doi.org/10.1016/j.cropro.2021.105681 (2021b).
59. Liang *et al.* Identification of differentially expressed genes related to aphid resistance in cucumber (Cucumis sativus L.). *Sci Rep*, **5**, 9645 https://doi.org/10.1038/srep09645 (2015).
60. Zhu-Salzman, K., Salzman, R. A., Ahn, J. E. & Koiwa, H. Transcriptional regulation of sorghum defense determinants against a phloem-feeding aphid. *Plant Physiol*, **134**, 420–443 https://doi.org/10.1104/pp.103.028324 (2004).
61. Grover, S., Agpawa, E., Sarath, G., Sattler, S. E. & Louis, J. Interplay of phytohormones facilitate sorghum tolerance to aphids. *Plant Mol. Biol*, https://doi.org/10.1007/s11103-020-01083-y (2020).
62. Dugas, D. V. *et al.* Functional annotation of the transcriptome of Sorghum bicolor in response to osmotic stress and abscisic acid. *BMC Genomics*, **12**, 514 https://doi.org/10.1186/1471-2164-12-514 (2011).
63. Yan, L. *et al.* Role of tomato lipoxygenase D in wound-induced jasmonate biosynthesis and plant immunity to insect herbivores. *PLoS Genet*, **9**, e1003964 https://doi.org/10.1371/journal.pgen.1003964 (2013).
64. Woldemariam, M. G., Ahern, K., Jander, G. & Tzin, V. A role for 9-lipoxygenases in maize defense against insect herbivory. *Plant Signaling Behavior*, **13**, 4709–4723 https://doi.org/10.1080/15592324.2017.1422462 (2018).
65. Shrestha, K., Pant, S. & Huang, Y. Genome-wide identification and classification of Lipoxygenase gene family and their roles in sorghum-aphid interaction. *Plant Mol Biol*, **105**, 527–541 https://doi.org/10.1007/s11103-020-01107-7 (2021).
66. Multani, D. S. *et al.* Loss of an MDR transporter in compact stalks of maize br2 and sorghum dw3 mutants., **302**, 81–84 https://doi.org/10.1126/science.1086072 (2003).
67. Brown, P. J., Rooney, W. L., Franks, C. & Kresovich, S. Efficient mapping of plant height quantitative trait loci in a sorghum association population with introgressed dwarfing genes. *Genetics*, **180**, 629–637 https://doi.org/10.1534/genetics.108.092239 (2008). http://
68. Mace, E. S. & Jordan, D. R. Location of major effect genes in sorghum [Sorghum bicolor (L.) Moench]. *Theor. Appl Genet*, **121**, 1339–1356 https://doi.org/10.1007/s00122-010-1392-8 (2010).
69. Hilley, J., Truong, S., Olson, S., Morishige, D. & Mullet, J. Identification of Dw1, a regulator of sorghum stem internode length. *PLoS One*, **11**, e0151271 https://doi.org/10.1371/journal.pone.0151271 (2016).
70. Hilley, J. L. *et al.* Sorghum Dw2 encodes a protein kinase regulator of stem internode length. *Sci Rep*, **7**, 4616 https://doi.org/10.1038/s41598-017-04609-5 (2017).
71. Boyles, R. E., Brenton, Z. W. & Kresovich, S. Genetic and genomic resources of sorghum to connect genotype with phenotype in contrasting environments. *Plant J*, **97**, 19–39 https://doi.org/10.1111/tpj.14113 (2019).
72. Morris, G. P. *et al.* Population genomics and GWAS in sorghum. *PNAS*, **110**, 453–458 https://doi.org/10.1073/pnas.1215985110 (2013).
73. Quinby, J. R. & Karper, R. E. Inheritance of three genes that influence time of floral initiation and maturity date in milo. *Agron J*, **37**, 916–936 (1945).
74. Quinby, J. R. Fourth maturity gene locus in sorghum. *Crop Sci*, **6**, 516–518 https://doi.org/10.2135/cropsci1966.0011183X000600060005x (1966).
75. Lin, Y. R., Schertz, K. F. & Paterson, A. H. Comparative analysis of QTLs affecting plant height and maturity across the Poaceae, in reference to an interspecific sorghum population. *Genetics*, **141**, 391–411 https://doi.org/10.1093/genetics/141.1.391 (1995).
76. Rooney, W. L. & Aydin, S. Genetic control of a photoperiod-sensitive response in Sorghum bicolor (L.) Moench. Crop Sci 39, 397–400(1999). **https://doi.org/10.2135/cropsci1999.0011183X0039000200016x**.

77. Morgan, P. W. & Finlayson, S. A. Physiology and genetics of maturity and height. In: Smith, W.C., Frederiksen, R.A. (Eds.), Sorghum: Origin, History, Technology and Production.Wiley, New York, NY,pp. 227–259, (2000).
78. Mace, E. S., Hunt, C. H. & Jordan, D. R. Supermodels: sorghum and maize provide mutual insight into the genetics of flowering time. *Theor Appl Genet*, **126**, 1377–1395 https://doi.org/10.1007/s00122-013-2059-z (2013).
79. Bowling *et al.* Scouting sugarcane aphids in South, Central, and West Texas. Texas A&M AgriLife Extension. ENTO-043, 2 p. College Station, TX(2016b).
80. Paudyal, S. *et al.* Evidence of host plant specialization among the U.S. sugarcane aphid (Hemiptera: Aphididae) genotypes. *Trends in Entomol*, **15**, 47–58 (2019b).
81. Armstrong, J. S. *et al.* Resistance to sugarcane aphid Melanaphis sacchari, (Hemiptera: Aphididae) in forage and grain sorghums. *J. Econ. Entomol*, **110**, 259–265 https://doi.org/10.1093/jee/tow261 (2017).
82. Starks, K. J. & Burton, R. L. Greenbugs: determining biotypes, culturing, and screening for plant resistance with notes on rearing parasitoids. *USDA Tech. Bull*, **1556**, 1–12 https://doi.org/http://10.22004/ag.econ.158078 (1997).
83. Webster, J. A., Baker, C. A. & Porter, D. R. Detection and mechanisms of Russian wheat aphid (Homoptera: Aphididae) resistance in barley. *J. Econ. Entomol*, **84**, 699–673 https://doi.org/10.1093/jee/84.2.669 (1991).
84. Burd, J. D. & Porter, D. R. Biotypic diversity in greenbug (Hemiptera: Aphididae): characterizing new virulence and host associations. *J. Econ. Entomol*, **99**, 959–965 https://doi.org/10.1093/jee/99.3.959 (2006).
85. Markwell, J., Osterman, J. C. & Mitchell, J. L. Calibration of the Minolta SPAD-502 leaf chlorophyll meter. *Photosynth. Res*, **46**, 467–472 https://doi.org/10.1093/genetics/141.1.391 (1995).
86. Hu, Z., Olatoye, M. O., Marla, S. & Morris, G. P. An integrated genotyping-by-sequencing polymorphism map for over 10,000 sorghum genotypes. *Plant Genome*, **12**, 180044 https://doi.org/10.3835/plantgenome2018.06.0044 (2019).
87. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows Wheeler transform. *Bioinformatics*, **25**, 1754–1760 https://doi.org/10.1186/s12864-018-5095-x (2009).
88. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data., **27**, 2987–2993 https://doi.org/10.1093/bioinformatics/btr509 (2011).
89. Browning, B. L. & Browning, S. R. Genotype imputation with millions of reference samples. *Am. J. Hum Genet*, **98**, 116–126 https://doi.org/10.1016/j.ajhg.2015.11.020 (2016).
90. Razali, N. M. Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. Stat. Modeling and Anal. 2, 21-33(2011). https:doi: 10.5395/rde.2012.37.4.245.
91. Lipka, A. E. *et al.* GAPIT: Genome association and prediction integrated tool., **28**, 2397–2399 https://doi.org/10.1093/bioinformatics/bts444 (2012).
92. Wang, J., Zhang, Z. G. A. P. I. T. & Version Boosting power and accuracy for genomic association and prediction. *Genom. Proteom. Bioinform*, **3**, https://doi.org/10.1016/j.gpb.2021.08.005 (2021).
93. VanRaden, P. M. Efficient methods to compute genomic predictions. *J. Dairy Sci*, **91**, 4414–4423 https://doi.org/10.3168/jds.2007-0980 (2008).
94. Breiman, L. Heuristics of instability and stabilization in model selection. *Ann. Statist*, **24**, 2350–2383 https://doi.org/10.1214/aos/1032181158 (1996).
95. Wickham, H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4(2016). http://ggplot2.org.
96. Kassambara, A. ggpubr: 'ggplot2' Based Publication Ready Plots. R package version 0.4.0(2020). https://CRAN.R-project.org/package=ggpubr.
97. Auguie, B. & GridExtra Miscellaneous Functions for "Grid" Graphics. R package version 2.3(2017). https://CRAN.R-project.org/package=gridExtra.
98. Wilke, C. O. & Cowplot Streamlined Plot Theme and Plot Annotations for 'ggplot2'. R package version 1.1.1(2020). https://CRAN.R-project.org/package=cowplot.

Figures

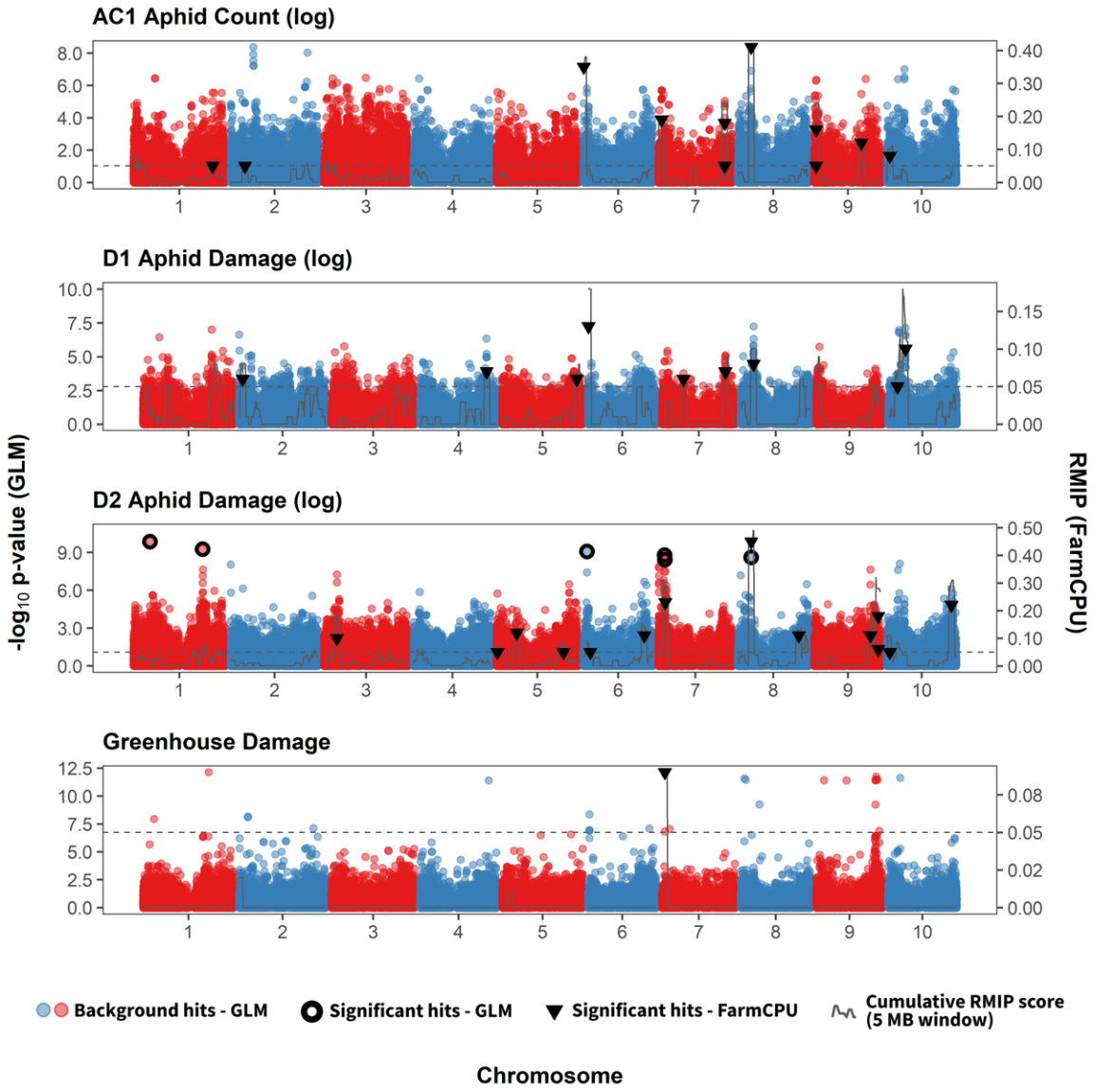

Figure 1

Genome-wide association for aphid-related traits. Red and blue markers indicate associations from a generalized linear model (GLM), plotted according to their physical location and raw -log10 p-value; SNPs with significant empirical p-values (≤0.01) are circled. The Resample Model Inclusion Probability (RMIP) from 100 FarmCPU iterations is plotted on the right axis, with significant individual SNPs (RMIP ≥ 0.05) as triangles. The total RMIP in a 5-Mb sliding window is shown as a gray line to show where several weak loci indicate a significant but poorly resolved locus (such as on SBI-09). D1, average score on the first rating dates across both years; D2, average score on the second rating dates across both years.

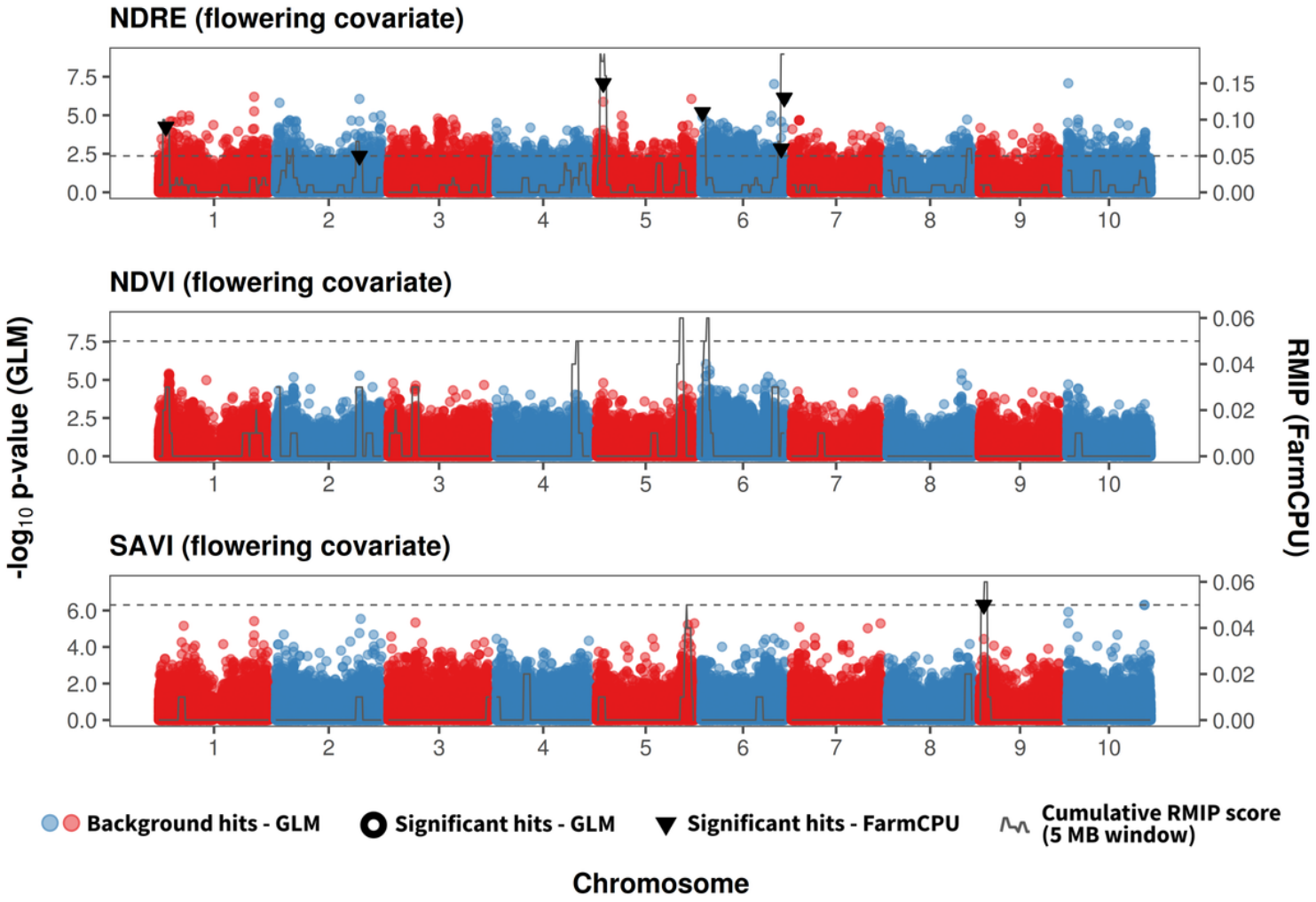

Figure 2

Genome-wide association for Unmanned Aircraft Systems-based indices. Data was taken on Aug 25, 2020 at Tifton, GA. Red and blue markers indicate associations from a generalized linear model (GLM), plotting according to their physical location and -log10 (raw) p-value; SNPs with significant empirical p-values (≤0.01) are circled. The Resample Model Inclusion Probability (RMIP) from 100 FarmCPU iterations is plotted on the right axis, with significant individual SNPs (RMIP ≥ 0.05) as triangles. The total RMIP in a 5-Mb sliding window is shown as a gray line to show where several weak loci indicate a significant but poorly resolved locus (such as on chromosomes SBI-04, SBI-05, and SBI-06 for NDVI).

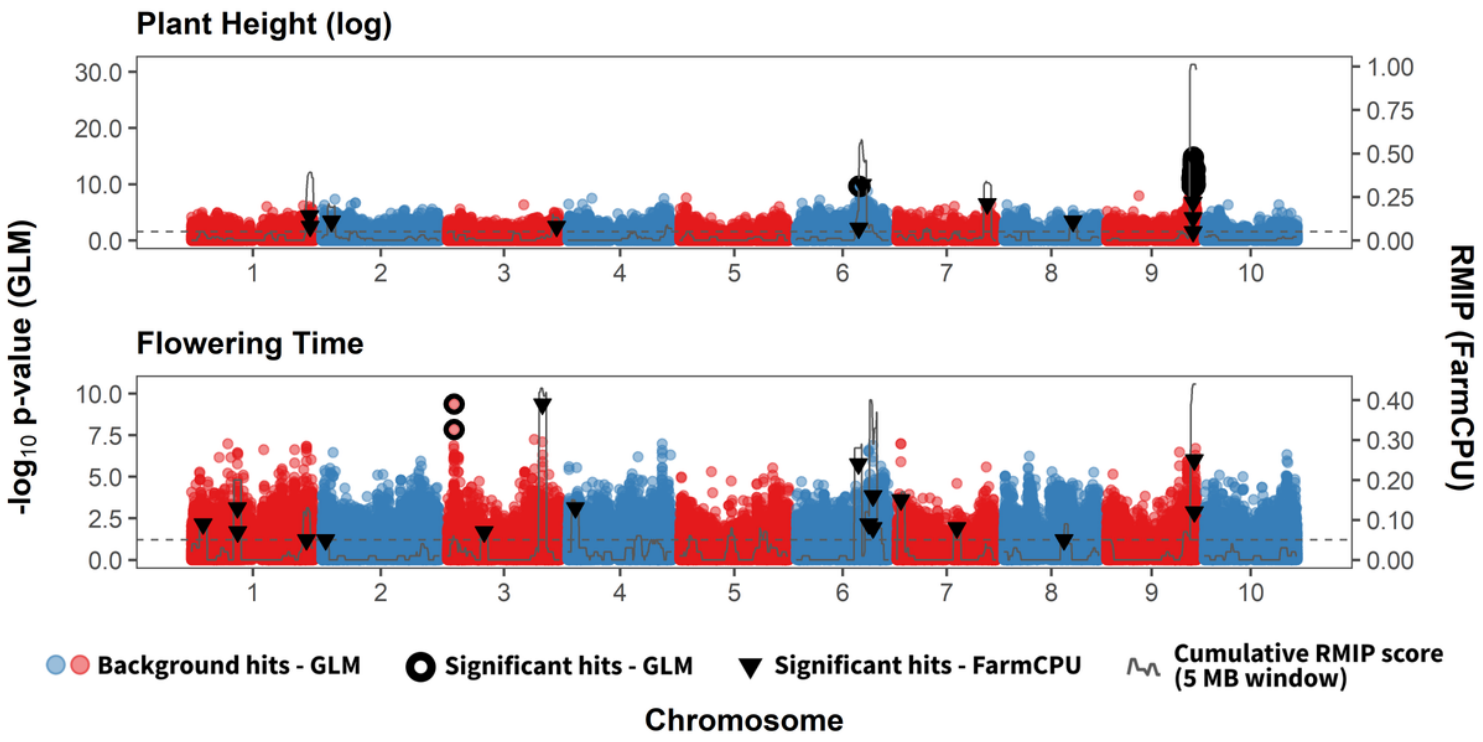

Figure 3

Genome-wide association of 281 Sorghum Association Panel accessions using 191,738 SNP markers. Red and blue markers indicate associations from a generalized linear model (GLM), plotting according to their physical location and -log10 (raw) p-value; SNPs with significant empirical p-values (≤0.01) are circled. The Resample Model Inclusion Probability (RMIP) from 100 FarmCPU iterations is plotted on the right axis, with significant individual SNPs (RMIP ≥ 0.05) as triangles. The total RMIP in a 5-Mb sliding window is shown as a gray line to show where several weak loci indicate a significant but poorly resolved locus (such as on SBI-09).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTableS1Selectedresistantandsusceptiblelines.docx
- SupplementaryTableS2SummaryofSignificantSNPsafterRMIPandGLManalysis.xlsx
- SupplementaryTableS3VarianceexplainedbysignificantSNPsrsquaredexplainedbySNPsbygroup.xlsx
- SupplementaryTableS4ListsofuniquegenesidentifiedfromKianiandSzczepanie2018.xlsx
- SupplementaryTableS5Listsofuniquegenesidentifiedwithin200kbofMTATetreaultH.M.etal.2019..xlsx
- SupplementaryTableS6AphidresistancegenomicregionidentifiedforKianiandSzczepaniecs2018andTetreaultetal.2019of16modules.xlsx
- SupplementaryTableS7SummaryofmarkersoverlapwithMuletaetal.2021.xlsx
- SupplementaryTableS8AcombinedSAPAccessionsofPlantinglistfortheyears2019and2020.xlsx
- SupplementaryFig.S1UASbasedrefelectanceManhattanplotsnocovariates.docx
- SupplementaryFig.S2GWASforFarmCPUsinglerunAnalysis.docx
- SupplementaryFig.S3GenefrequencyidentifiedonanindividualchromosomeandnearSNPsassociatedwithaphidresistancerelatedtraits2.docx