



OPEN

Association analysis for resistance to *Striga hermonthica* in diverse tropical maize inbred lines

A. E. Stanley^{1,2}, A. Menkir²✉, B. Ifie¹, A. A. Paterne², N. N. Unachukwu², S. Meseka², W. A. Mengesha², B. Bossey², O. Kwadwo¹, P. B. Tongoona¹, O. Oladejo², C. Sneller³ & M. Gedil²

Striga hermonthica is a widespread, destructive parasitic plant that causes substantial yield loss to maize productivity in sub-Saharan Africa. Under severe *Striga* infestation, yield losses can range from 60 to 100% resulting in abandonment of farmers' lands. Diverse methods have been proposed for *Striga* management; however, host plant resistance is considered the most effective and affordable to small-scale farmers. Thus, conducting a genome-wide association study to identify quantitative trait nucleotides controlling *S. hermonthica* resistance and mining of relevant candidate genes will expedite the improvement of *Striga* resistance breeding through marker-assisted breeding. For this study, 150 diverse maize inbred lines were evaluated under *Striga* infested and non-infested conditions for two years and genotyped using the genotyping-by-sequencing platform. Heritability estimates of *Striga* damage ratings, emerged *Striga* plants and grain yield, hereafter referred to as *Striga* resistance-related traits, were high under *Striga* infested condition. The mixed linear model (MLM) identified thirty SNPs associated with the three *Striga* resistance-related traits based on the multi-locus approaches (mrMLM, FASTmrMLM, FASTmrEMMA and pLARmEB). These SNPs explained up to 14% of the total phenotypic variation. Under non-infested condition, four SNPs were associated with grain yield, and these SNPs explained up to 17% of the total phenotypic variation. Gene annotation of significant SNPs identified candidate genes (Leucine-rich repeats, putative disease resistance protein and VQ proteins) with functions related to plant growth, development, and defense mechanisms. The marker-effect prediction was able to identify alleles responsible for predicting high yield and low *Striga* damage rating in the breeding panel. This study provides valuable insight for marker validation and deployment for *Striga* resistance breeding in maize.

Maize (*Zea mays* L.) is an important cereal that plays a crucial role in alleviating food insecurity in sub-Saharan Africa (SSA) due to its high yield potential, ease in processing and low cost¹. However, its production is constantly hampered by a plethora of biotic stresses, including the parasitic weed *Striga*. Among the numerous *Striga* species endemic to Africa, *Striga hermonthica* (Del.) Benth is the most destructive and widespread, affecting cereals, including maize and sorghum (*Sorghum bicolor* L.)². Yield losses attributed to *Striga* infestation in maize range from 60 to 100% under severe field infestation, especially in marginal production areas where smallholder farmers cannot afford high inputs and other control measures³.

Striga hermonthica is an obligate root hemiparasite, which depends on its host for survival, notwithstanding its photosynthetic capacity after emergence from the soil⁴. The parasite's lifecycle is intimately associated with its host to ensure its survival⁵. The interaction between the parasitic plant and its host is initiated immediately a chemical compound known as strigolactone is released from the host plant. This chemical compound stimulates the germination of *Striga* seeds. Once the *Striga* seeds germinate, it establishes a connection with the roots of its host, extracting water, carbon and essential nutrients for its growth. The parasitic plant inflicts more damage on its host underground before its emergence from the soil, and this damage is accentuated in areas affected by sub-optimal soil fertility and recurrent drought⁶. *Striga hermonthica* parasitism is characterized by chlorosis, firing of leaves around margins, wilting, stunting, poorly filled ears, and death under severe infestation⁷. *Striga* resistance is a complex quantitative trait controlled by multiple genes/polygenes, and it is highly affected by environmental factors⁸.

¹West Africa Centre for Crop Improvement, University of Ghana, Legon, Ghana. ²International Institute of Tropical Agriculture, Ibadan, Nigeria. ³Ohio Agriculture Research and Development Center, Ohio State University, Wooster, OH, USA. ✉email: a.menkir@cgiar.org

Source	DF	Yield (kg/ha)	<i>Striga</i> damage score at 8 WAP (1–9)	<i>Striga</i> damage score at 10 WAP (1–9)	Emerged <i>Striga</i> count at 8 WAP	Emerged <i>Striga</i> count at 10 WAP
Env (E)	3	104128033**	172.75 [§]	445.94 [§]	5792.54 [§]	39,237.32 [§]
Rep(Env)	4	2831059**	1.26 ns	4.05 ns	91.92 ns	301.57 ns
Blk(Rep*E)	112	524846***	1.11 [§]	1.33***	451.34 [§]	1112.03 [§]
Lines (L)	149	2559095 [§]	7.00 [§]	7.81 [§]	1528.28 [§]	2435.87 [§]
L x E	446	483792 [§]	1.18 [§]	1.40 [§]	268.44**	530.09**
Error	483	321643	0.63	0.83	204.33	416.2
CV (%)		35.95	19.62	15.45	61.34	58.53

Table 1. Mean squares from the combined ANOVA for traits recorded under *Striga* infestation across four environments. *, **, ***, [§]Significant at $p \leq 0.05$, $p \leq 0.01$, $p \leq 0.001$ and 0.0001 levels, respectively, ns = not significant. Blk, Block; Env, Environment; Rep, Replication; CV, Coefficient of Variation; WAP, weeks after planting.

Several control methods have been proposed for *Striga* management, including host plant resistance, cultural, chemical and manual control options. However, integrated *Striga* management approach is considered the most economical and affordable for small-scale farmers who cannot afford high inputs control options. The approach involves the combination of two or more control options. Host plant resistance is considered a cost effective, environmental feasible and affordable option for smallholder farmers. It is also an essential component of any successful integrated approach for controlling *Striga* parasitism. Several studies have shown progress in breeding for natural genetic resistance to *Striga* in maize⁵. In addition, extensive research has been done to map quantitative trait loci (QTLs) for *Striga* resistance in maize using molecular markers. QTL mapping and genome-wide association study (GWAS) are two methods widely used to discover genetic loci controlling complex traits. Quantitative trait loci (QTLs) associated with *Striga hermonthica* have been identified in maize using QTL mapping approach^{9–11}. Badu-Apraku et al.⁹ identified 12 QTLs associated with four *Striga* resistance/tolerance traits in maize and these QTLs explained 3.2 to 34.9% of the phenotypic variation. However, the QTL mapping approach have several limitations, for example, it has limited allelic diversity, and limited mapping resolution due to limited recombination events¹². On the other hand, GWAS explores ancestral recombination in natural genetically diverse population to dissect complex traits¹³. GWAS is an improvement over QTL mapping in that it improves the resolution of QTLs due to accumulated meiotic events and reduces the time taken in developing mapping populations¹⁴. GWAS is a powerful tool for detecting QTLs associated with important complex quantitative traits, as well as predicting or identifying causative genes¹⁵.

Many statistical models have been developed to improve the power of identifying QTNs with the GWAS approach. The single-locus mixed-linear model (MLM) is the most common method used for GWAS. The method uses several algorithms such as the compressed MLM¹⁶, enriched MLM¹⁷, however, all these models perform single dimensional genome scan and require multiple correction. These models also have major limitations in mapping QTNs with small effects. Wang et al.¹⁸ proposed a new model based on multi-locus random-SNP-effect MLM (mrMLM). The methods include polygenic-background control-based least angle regression plus empirical Bayes (pLARM EB), fast multi-locus random-SNP-effect efficient mixed model association (FASTmrEMMA), iterative-sure independence screening expectation-maximization (EM)-Bayesian LASSO (ISIS EM-BLASSO) and fast multi-locus random-SNP-effect mixed linear model (FASTmrMLM)^{18–22}. These methods can effectively detect small-effect QTNs and improve the efficiency and accuracy of GWAS. Recently, few studies have implemented the above GWAS methods to detect important loci controlling different traits in maize²³.

Genome-wide association study (GWAS) for *S. hermonthica* resistance has been conducted in maize. Adewale et al.²⁴ identified 24 SNPs that were significantly associated with four *Striga* resistance-related traits in early maturing maize inbred lines using compressed MLM, these lines captured only the genetic variation existing in the extra early and early maturing maize germplasm developed at IITA. However, genomic regions governing *Striga* resistance in many intermediate and late-maturing maize inbred lines with consistent expression of polygenic resistance to *S. hermonthica* have not been identified. This germplasm offers an excellent resource for discovering functional genes underlying the genetic variation in the *Striga* resistance-related traits. This study was thus conducted; to evaluate diverse intermediate and late maturing maize inbred line for *Striga* resistance under *Striga* infested and non-infested conditions and identify genomic regions and candidate genes related to *Striga* resistance.

Results

Phenotypic diversity. In the combined analyses of variance (ANOVA), environments and lines had significant ($p < 0.001$) effects on the three *Striga* resistance-related traits under *Striga* infested conditions and for grain yield under non-infested condition (Tables 1 and 2). The line x environment interaction mean squares were also significant for most of the traits measured under the two conditions.

Further assessment of the line x environment interaction using rank correlation analyses between pairs of environments found highly significant ($p < 0.0001$) correlations for the *Striga* resistance-related traits (Supplementary Table S1). The broad-sense heritability estimates for the *Striga* resistance-related traits varied from 81 to 85% (Supplementary Table S2).

Source	DF	Yield (kg/ha)
Env (E)	3	175723422.1 [§]
Rep(Env)	4	2428705***
Blk(Rep*Env)	112	1281736.9 [§]
Lines (L)	149	3287995 [§]
L x E	446	732612.7 [§]
Error	483	470085
CV (%)		32.73

Table 2. Mean squares from the combined ANOVA for traits recorded under non-infested conditions across four environments. *, **, ***, [§]Significant at $p \leq 0.05$, $p \leq 0.01$, $p \leq 0.001$ and 0.0001 levels, respectively, ns = not significant. Blk, Block; Env, Environment; Rep, Replication; CV, Coefficient of Variation; WAP, weeks after planting.

Grain yield under *Striga* infestation varied from 13 to 3299 kg/ha with an average of 1580 kg/ha, while grain yield under non-infested condition varied from 706 to 4171 kg/ha with an average of 2098 kg/ha (Supplementary Table S2). Relative to the non-infested condition, the resistant benchmark (9450) suffered 46% yield loss, whereas the susceptible benchmark (5057) suffered 77% yield loss indicating the inbred lines used in this study were exposed to severe *Striga* infection. In addition, 86 maize inbred lines supported significantly fewer emerged *Striga* plants at 8 and 10 weeks after planting (WAP). These lines, on average, suffered 19% yield reduction relative to the non-infested conditions, produced significantly higher grain yields than the resistant benchmark (9450), and were categorized as resistant lines. In contrast, 44 inbred lines supported as many *Striga* plants as the susceptible benchmark (5057) but produced significantly higher grain yields than the susceptible line, and were categorized as tolerant. The remaining 20 inbred lines, supported as many *Striga* plants as the susceptible benchmark and did not differ significantly from the susceptible line in their grain yields, and were regarded as susceptible. All paired traits showed statistically significant differences at p -value < 0.01 . A negative correlation was observed between grain yield and *Striga* damage rating and emerged *Striga* plants at 8 and 10 WAP. However, there was a positive correlation between *Striga* damage rating and emerged *Striga* plants at 8 and 10 WAP (Supplementary Fig. S1).

Genotyping. *Population structure and linkage disequilibrium.* For the genotypic analysis, 16,735 SNPs distributed across the ten maize chromosomes were identified after the quality control process. A minimum of 1208 SNPs (7.2%) were mapped on chromosome 10, whereas a maximum of 2532 SNPs (15.2%) were mapped on chromosome 1. The Admixture analysis using tenfold cross-validation from $k = 1$ to $k = 10$ showed a sharp elbow at $k = 3$, indicating the inbred lines can be grouped into three subgroups (Fig. 1A,B). The principal component analysis (PCA) also grouped the inbred lines into three subgroups and this is consistent with the Admixture results. A scree plot generated to visualize the fraction of variance represented by each of the ten principal components showed that two (PC1 and PC2) explained the largest proportion (42.3%) of the total variance (Fig. 1C). Furthermore, the phylogeny tree clustering also grouped the inbred lines into three distinct subgroups; 22, 27 and 101, derived from ZDIP, IWDS and Mixed groups, respectively (Fig. 1D). The assignment of the inbred lines into the three subgroup based on the phylogeny tree were in good agreement (98%) with those revealed by PCA. The inbred lines were grouped based on their genetic background/pedigree and maturity. The LD estimates (r^2) showed a slight increasing and then consistent pattern of LD decay was observed with an increase in the physical distance of SNP markers mapped on the 10 chromosomes (Supplementary Fig. S2). The average linkage disequilibrium decay varied from 2.73 kb on chromosome 6 to 3.68 kb on chromosome 8 (Supplementary Table S3).

Genome-wide association analysis. The GWAS multiple-locus models used in this study identified 30 significant SNPs that were significantly associated with the three *Striga* resistance-related traits. These SNPs were distributed on all maize chromosomes but chromosome 8. The highest number of SNPs was found on Chromosome 1, and the least on chromosomes 3, 6 and 7 (Table 3). The results of the Manhattan plot and the quantile-quantile plots revealed reasonable data adjustment and a few significant SNPs above the interval of the expected values of the null hypothesis (Fig. 2). This study employed four multi-locus methods, mrMLM, FASTmrMLM, FASTmrEMMA, and pLARmEB to perform comprehensive GWA mapping in our diversity panel. Among the four methods mrMLM identified the highest number of SNPs (19) while, FASTmrEMMA identified the least (5). In addition, FASTmrMLM identified the most codetected SNPs among the four GWAS models used. Four SNPs were associated with grain yield under *Striga* infested condition. These SNPs are located on three chromosomes, and each SNP explained between 3.21 to 13.36% of the phenotypic variation. One of the SNPs (S4_164335765) associated with grain yield was detected by two GWAS multi-locus methods (FASTmrMLM and mrMLM) and they explained 6.7 and 13.4% of the phenotypic variation. The LOD score of the significant SNPs ranged from 7.18 (S4_164335765) to 7.39 (S10_1784894).

Eleven SNPs were associated with *Striga* damage rating at 8 WAP. These SNPs are located on seven chromosomes and the proportion of phenotypic variation explained by each SNP ranged from 0.14 to 14.19%. The LOD values of the identified SNPs ranged from 6.06 (S4_160459526) to 11.80 (S10_68324912). Nine SNPs were associated with *Striga* damage rating at 10 WAP. These SNPs were located on five chromosomes, and they explained 4.13 to 11.60% of the phenotypic variation. Four SNPs (S1_284192573, S1_10576247, S2_99667127,

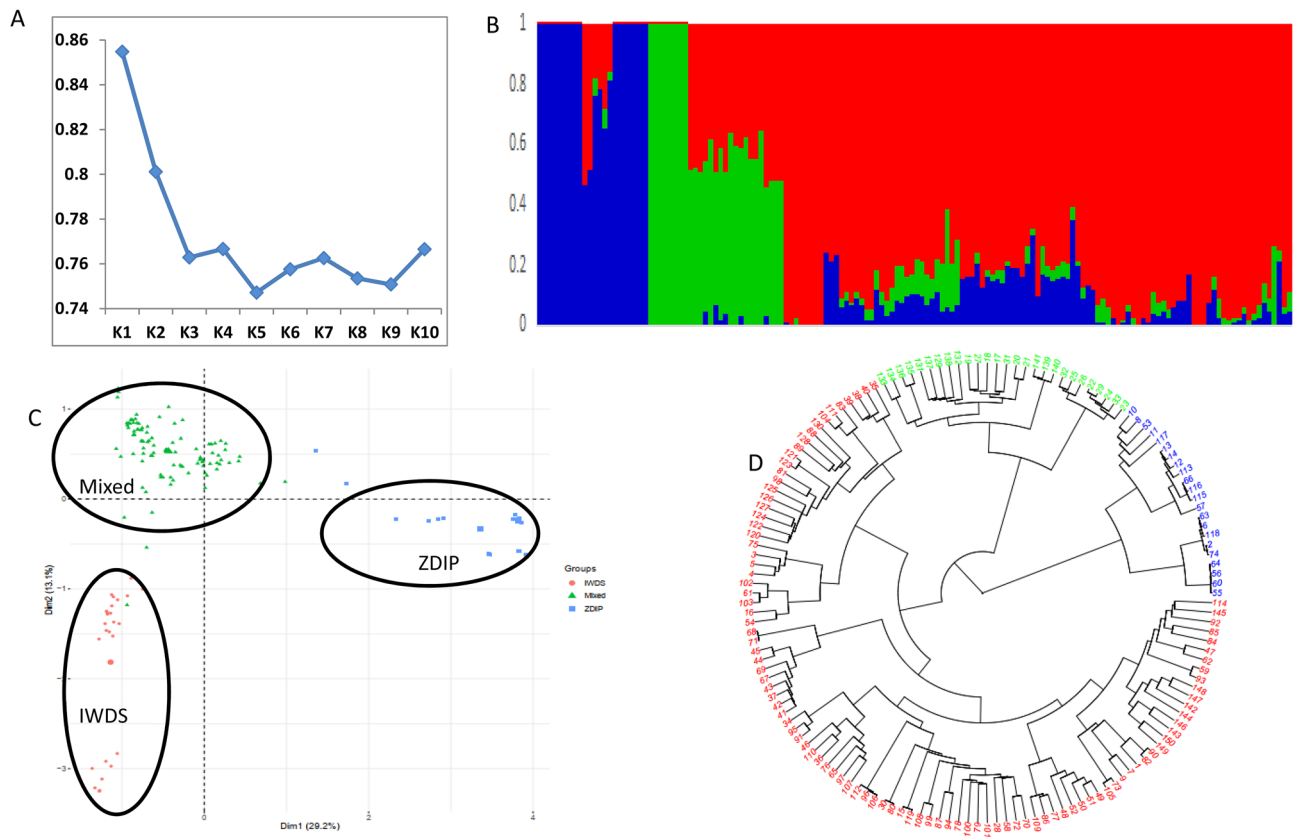


Figure 1. (A) Cross-validation plot showing the optimal number of clusters. (B) Population structure plot of the inbred lines ($k=3$). (C) Principal component analysis based on 150 maize inbred lines using the 16,735 SNP markers. (D) Phylogenetic tree showing the genetic relationship among 150 diverse maize inbred lines.

and S7_10795659) associated with *Striga* damage rating at 10 WAP were detected by two or more of the GWAS multi-locus (mrMLM, and FASTmrMLM, and FASTmrEMMA) methods.

The LOD values of the identified SNPs ranged from 6.02 (S2_190557148) to 13.76 (S2_188120710). In general, two SNPs (S2_160791711 and S10_25285761) were simultaneously associated with *Striga* damage rating at 8 and 10 WAP.

Eight SNPs were associated with emerged *Striga* plants at 8 WAP. These SNPs are located on five chromosomes and the proportion of phenotypic variation explained by each SNP ranged from 5.03 to 13.58%. Five SNPs (S2_208978140, S2_135038935, S5_148751913, S9_7727167, and S10_90133328) associated with emerged *Striga* plants at 8 WAP were detected by two or more of the GWAS multi-locus (mrMLM, FASTmrMLM, FASTmrEMMA, and pLARMEB) methods. The LOD values of the identified SNPs ranged from 6.14 (S1_298950342) to 9.28 (S10_125571525). Four SNPs were associated with emerged *Striga* damage plants at 10 WAP. These SNPs are located on three chromosomes and the proportion of phenotypic variation explained by each SNP ranged from 6.33 to 12.21%. Three SNPs (S2_135038935, S5_148751913, and S5_204969099) associated with emerged *Striga* plants at 10 WAP were detected by two GWAS multi-locus (mrMLM, and FASTmrMLM) methods. The LOD values of the identified SNPs ranged from 6.83 (S2_135038935) to 8.65 (S5_148751913). In general, three SNPs (S2_135038935, S5_148751913, and S10_12557152) were simultaneously associated with emerged *Striga* plants at 8 and 10 WAP.

Under non-infested conditions, four SNPs were associated with grain yield (Table 4). These SNPs are located on three chromosomes and the proportion of phenotypic variation explained by each SNP ranged from 5.63 to 17.40%. Furthermore, one of the SNPs on chromosome 1 (S4_189154251) was detected by two of the GWAS multi-locus methods (mrMLM, and FASTmrMLM). The LOD score of these SNPs ranged from 6.52 (S1_26517984) to 9.72 (S4_189154251).

Markers effect prediction. The frequencies and marker prediction effects of various alleles associated with the three *Striga* resistance-related traits are presented in Table 5. Two of the SNPs (S4_164335765 and S9_1994432) associated with grain yield under *Striga* infestation displayed high segregation among the inbred lines. For SNPs on chromosome 4, alleles AA and CA were associated with genotypes with higher grain yield, while alleles CC were associated with lower grain yield. For the second SNP on chromosome 9, alleles GG were associated with genotypes with higher grain yield, while alleles AA were associated with lower grain yield (Fig. 3). For *Striga* damage ratings at 8 and 10 WAP, three SNPs (S1_284192573, S4_160459526, and S5_5623740) displayed high segregation among the inbred lines. For two of the SNPs on chromosomes 1 and 4, alleles TT and TG were asso-

Trait name	SNP	Chr	Position (bp)	LOD	PVE (%)	Favorable allele	Methods
YLDIN_G	S4_164335765	4	164335765	7.18	13.36	A	1, 2
	S9_1994432	9	1994432	7.21	3.21	G	4
	S10_1784894	10	1784894	7.39	3.58	C	4
Striga damage rating at 8 WAP	S1_18512344	1	18512344	10.76	14.19	T	1
	S2_14081759	2	14081759	8.26	7.89	A	1
	S2_160791711	2	160791711	7.28	3.35	A	1
	S2_219240910	2	219240910	11.31	8.05	G	3
	S4_160459526	4	160459526	6.07	3.83	C	1
	S5_70442824	5	70442824	6.76	3.22	C	2
	S5_216138908	5	216138908	6.81	6.67	C	4
	S6_25428338	6	25428338	6.58	0.142	A	4
	S10_25285761	10	25285761	11.87	6.57	G	2
	S10_2743583	10	2743583	9.86	0.24	C	4
	S10_68324912	10	68324912	11.79	1.32	C	4
Striga damage rating at 10 WAP	S1_284192573	1	284192573	8.54	10.40	G	1, 2, 3
	S1_10576247	1	10576247	6.45	10.00	A	1, 2
	S2_190557148	2	190557148	6.02	4.13	T	1
	S2_99667127	2	99667127	7.82	4.82	A	1, 2
	S2_160791711	2	160791711	6.18	4.37	A	3
	S2_188120710	2	188120710	13.76	6.47	A	4
	S5_5623740	5	5623740	6.06	13.97	C	1
	S7_10795659	7	10795659	6.71	7.99	A	1, 2
	S10_25285761	10	25285761	6.56	5.04	G	3
Emergenced Striga plants at 8 WAP	S1_298950342	1	298950342	6.15	6.58	A	1
	S2_135038935	2	135038935	6.67	6.66	G	3, 4
	S2_208978140	2	208978140	6.82	6.97	C	3, 4
	S3_74335447	3	74335447	6.21	5.54	A	1
	S5_148751913	5	148751913	6.41	6.21	A	3, 4
	S9_7727167	9	7727167	8.32	11.51	G	1, 2
	S10_125571525	10	125571525	9.28	13.08	G	1, 2, 4
	S10_90133328	10	90133328	7.75	13.58	T	1, 2, 4
Emergenced Striga plants at 10 WAP	S2_135038935	2	135038935	6.83	12.22	G	1, 2
	S5_148751913	5	148751913	8.65	7.91	A	1, 2
	S5_204969099	5	204969099	6.93	10.22	T	1, 2
	S10_125571525	10	125571525	7.00	6.33	G	4

Table 3. Significant SNPs identified under *Striga* infested condition. Methods: Numbers 1 to 4 represents different GWAS methods: 1: mrMLM; 2: FASTmrMLM; 3: FASTmrEMMA; 4 pLARmEB Chr, Chromosome; YLDIN_G, Grain yield; WAP, weeks after planting.

ciated with high *Striga* damage ratings, while alleles GG and CC were associated with low *Striga* damage rating. For the SNP on chromosome 5, alleles AA and AC were associated with high *Striga* damage ratings, while alleles CC were associated with low *Striga* damage ratings at 8 and 10 WAP (Fig. 3). For emergenced *Striga* plants, four SNPs (S1_298950342, S3_74335447, and S5_204969099) displayed high segregation among the inbred lines. For the two SNPs on chromosomes 1 and 3, variants GG and AG supported the emergence of more *Striga* plants whereas, alleles AA supported little emergenced *Striga* plants. For the SNP on chromosome 5, variants CC supported the emergence of more *Striga* plants, while alleles AA and AC supported little emergenced *Striga* plants (Fig. 3).

Identifying putative genes. According to the genomic information of B73 Ref Gen_V4, thirty-one putative genes/proteins, including two uncharacterized proteins were found in the intervals adjacent to the significant SNPs detected for the three *Striga hermonthica* resistance-related traits (Table 6). Remarkably, most of the gene models identified encode transcription factors, disease resistance proteins, zinc-finger domain proteins, leucine-rich repeats protein kinase and some pathogenesis-related proteins. Most of the identified genes were located on chromosomes bins 1.10, 2.05, 2.06, 3.04, 6.01, 7.01, 9.01, 10.00, 10.01, and 10.03.

Five gene models were identified around three SNPs associated with grain yield under *Striga* infestation. Two gene models each were found on chromosomes 4 and 9 and they encode adenylyltransferase, sulfurtransferase (MOCS3 2), U-box domain-containing protein 39, and NLR family CARD domain-containing protein 3. The

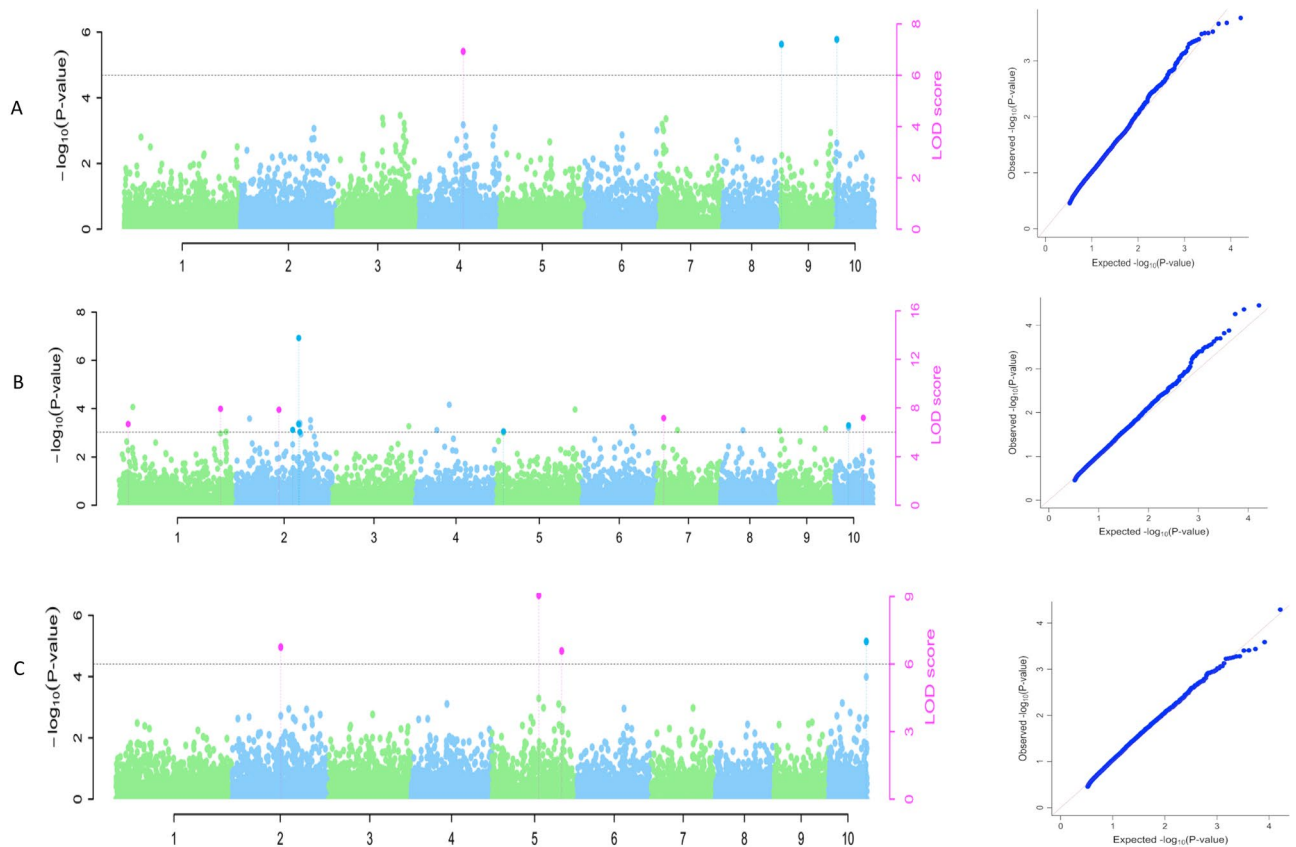


Figure 2. Manhattan plot indicating SNPs associated with (A) Grain yield, (B) *Striga* damage score at 10 WAP (C) Emerged *Striga* plants at 10 WAP. The graph refers to the quantile–quantile (Q–Q) plot of the P-values observed and expected from the association analysis under *Striga* infestation.

Trait name	SNP	Chr	Position (bp)	LOD score	PVE (%)	Favorable allele	Methods
YLDUN_G	S1_14334790	1	14334790	7.0841	10.2684	C	1
	S1_26517984	1	26517984	6.5272	5.6321	G	2
	S4_189154251	4	189154251	9.7264	17.4015	A	1, 2
	S8_17462112	8	17462112	6.5008	7.0482	G	4

Table 4. Significant SNPs identified under non-infested condition. Methods: Numbers 1 to 4 represents different GWAS methods: 1: mrMLM; 2: FASTmrMLM; 4 pLARM EB Chr = Chromosome, YLDIN_G = Grain yield, WAP = weeks after planting.

remaining gene model on chromosome 10 encodes VQ proteins. These putative genes/proteins are mainly involved in developmental processes including responses to biotic and abiotic stress, seed development and photo-morphogenesis. The LD heat-map of significant SNPs (S9_1994432 and S10_1784894) identified on chromosomes 9 and 10 were highly correlated (0.5 to 0.8) with regions adjacent to the identified putative genes (U-box domain-containing protein 39, NLR family CARD domain-containing protein 3 and VQ proteins) (Supplementary Fig. S3).

Nineteen gene models were identified around seventeen SNPs associated with *Striga* damage ratings at 8 and 10 WAP. Two gene models each were associated with SNPs S10_25285761 and S10_2743583 located on chromosome 10. These gene models encode leucine-rich repeat extension-like protein, disease resistance protein RPM1, disease resistance RPP13-like protein1 and an uncharacterized protein. Other genes models associated with *Striga* damage ratings encode putative cytochrome P450 superfamily protein, xyloglucan endotransglycosylase, basic helix-loop-helix (bHLH7 and bHLH20) transcription factors, knotted related homeobox 5, ubiquitin-protein ligase, and zinc-finger domain proteins. Most of these genes/proteins identified are involved in different development and plant defense mechanism. Plant resistance genes allow plants recognize the presence of specific pathogens and initiate defense responses. The LD heat-map of significant SNPs (S5_70442824, S7_10795659 and S10_25285761) identified on chromosomes 5, 7 and 10 were highly correlated (0.5 to 0.8) with regions adjacent

Trait	Marker name	Allele1	Allele2	Sequence	Frequency	Adjusted probability	Adjusted significance
Grain yield	S4_164335765	AA	AC	AAAC	0.34	4.19E-01	ns
		AA	CC	AACC	0.47	4.47E-07	****
		AC	CC	ACCC	0.21	1.52E-01	ns
	Ch9_1994432	AA	GG	AAGG	1.00		***
Striga damage rating	Ch1_284192573	GG	GT	GGGT	0.42	2.46E-01	ns
		GG	TT	GGTT	0.48	1.23E-04	***
		GT	TT	GTTT	0.10	2.47E-01	ns
	Ch4_160459526	CC	TT	CCTT	1.00		**
	Ch5_5623740	AA	AC	AAAC	0.12	1.68E-01	ns
		AA	CC	AACC	0.49	5.25E-04	***
		AC	CC	ACCC	0.39	7.20E-02	ns
Ch5_216138908	AA	CC	AACC	1.00		***	
Emerged Striga plants	Ch1_298950342	AA	AG	AAAG	0.23	7.50E-02	ns
		AA	GG	AAGG	0.46	1.62E-05	****
		AG	GG	AGGG	0.31	6.10E-01	ns
	Ch3_74335447	AA	AG	AAAG	0.06	7.10E-02	ns
		AA	GG	AAGG	0.45	2.00E-03	**
		AG	GG	AGGG	0.49	2.45E-07	****
	Ch5_148751913	AA	AC	AAAC	0.36	3.29E-01	ns
		AA	CC	AACC	0.16	4.60E-02	*
		AC	CC	ACCC	0.48	4.60E-02	*
	Ch5_204969099	CC	CT	CCCT	0.13	7.04E-01	ns
		CC	TT	CCTT	0.48	4.56E-04	***
		TT	TC	TTTC	0.39	7.04E-01	ns

Table 5. Frequencies and marker prediction effects of various alleles associated with the *Striga* resistance-related traits.

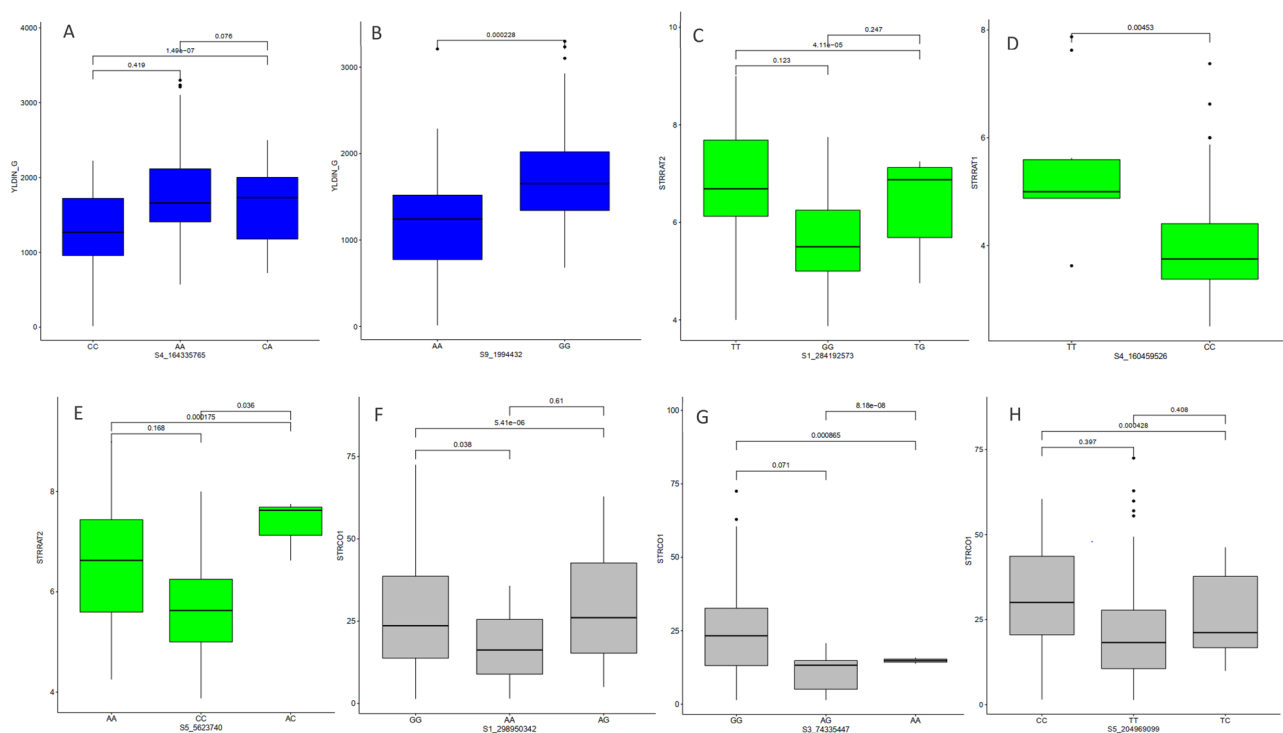


Figure 3. Allelic effects of haplotype blocks associated with Grain yield (A,B) blue colour, *Striga* damage ratings (C–E) green colour, emerged *Striga* plants (F,G) gray colour under *Striga* infestation.

Trait	SNP	Position (bp)	Gene ID	Putative Gene
Grain yield	S4_164335765	164335765	GRMZM2G157836; GRMZM5G881641	adenyltransferase and sulfurtransferase (MOCS3 2)
	S9_1994432	1994432	GRMZM2G406758; GRMZM2G110289	U-box domain-containing protein 39; NLR family CARD domain-containing protein 3
	S10_1784894	1784894	GRMZM2G180262	VQ
<i>Striga</i> damage score at 8 and 10 WAP	S1_18512344	18512344	GRMZM2G024099	Aspartyl protease AED3
	S1_284192573	284192573	GRMZM2G351582	ZPR1 zinc-finger domain protein (uaz7c01h10)
	S1_10576247	10576247	GRMZM2G028521	citrate transporter family protein (pco091082)
	S2_14081759	14081759	GRMZM2G092128	E3 ubiquitin-protein ligase PUB23
	S2_160791711	160791711	GRMZM2G102242	meiotic nuclear division protein 1 homolog
	S2_190557148	190557148	GRMZM2G414252	bHLH transcription factor (bHLH20)
	S2_99667127	99667127	GRMZM2G171830	Protein TIFY 10B
	S2_219240910	219240910	GRMZM2G162781	putative leucine-rich repeat protein kinase family protein
	S4_160459526	160459526	GRMZM2G081285	RING-H2 finger protein ATL1R
	S5_5623740	5623740	GRMZM2G112548	transcription factor JUNGBRUNNEN 1
	S5_216138908	216138908	GRMZM2G113418	glutaredoxin 2
	S5_70442824	70442824	GRMZM2G035073	putative cytochrome P450 superfamily protein
	S6_25428338	25428338	GRMZM5G832409	knotted related homeobox 5 (lg4b)
	S7_10795659	10795659	GRMZM2G162382	Transcription factor bHLH7
	S10_25285761	25285761	GRMZM2G300965; GRMZM2G300969	uncharacterized LOC100381459; leucine-rich repeat extensin-like protein 3
	S10_2743583	2743583	GRMZM5G873586; GRMZM2G356817	Disease resistance protein RPM1; putative disease resistance RPP13-like protein 1
	S10_68324912	68324912	GRMZM2G364748	Xyloglucan endotransglycosylase (gpm554)
Emerging <i>Striga</i> plants at 8 and 10 WAP	S1_298950342	298950342	GRMZM2G017470	Dof zinc finger protein
	S2_135038935	135038935	GRMZM2G088778	Protein ACCELERATED CELL DEATH 6
	S2_208978140	208978140	GRMZM2G179505	hydrolase/ protein serine/threonine phosphatase
	S3_74335447	74335447	GRMZM2G701566	uncharacterized LOC100382572
	S5_148751913	148751913	GRMZM2G129543	peroxidase 70
	S5_204969099	204969099	GRMZM5G823157	probable WRKY transcription factor 14
	S9_7727167	7727167	GRMZM2G033413	bZIP transcription factor 46
	S10_125571525	125571525	GRMZM2G006948	hapless 8
	S10_90133328	90133328	GRMZM2G063575	Pentatricopeptide repeat-containing protein

Table 6. Significant SNPs associated with the *Striga* resistance-related traits and putative genes identified for the 150 maize inbred lines.

Trait	SNP	Position	Gene ID	Putative Gene
Grain yield	S1_14334790	14334790	GRMZM2G424020, GRMZM2G464363	IQ domain-containing protein IQM5, uncharacterized protein LOC100277298
	S4_189154251	189154251	GRMZM2G174834, GRMZM2G174938	WR1 transcription factor 2, ENTH/VHS family protein
	S1_26517984	26517984	GRMZM2G098714	Replication protein A 70 kDa DNA-binding
	S8_17462112	17462112	GRMZM2G026015	Photosystem 1 reaction center subunit XI

Table 7. Quantitative trait nucleotides (QTNs) associated with grain yield under non-infested condition and putative genes identified for the 150 maize inbred lines.

to the identified putative genes (putative cytochrome P450 superfamily protein, Transcription factor bHLH7, uncharacterized LOC100381459 and leucine-rich repeat extensin-like protein 3) (Supplementary Fig. S4).

Nine gene models were identified around nine SNPs associated with emerging *Striga* plants at 8 and 10 WAP. These gene models encode Dof zinc finger protein, protein accelerated cell death 6, peroxidase 70, hapless, basic-domain leucine-zipper (bZIP46) and WRKY14 transcription factor. Transcription factors are usually involved in diverse plant processes including, growth, development and stress signaling. In addition, protein accelerated cell death is a positive regulator of programmed cell death and it is a mechanism used by plants for defense against pathogen infection. The LD heat-map of significant SNPs (S3_74335447 and S9_7727167) identified on chromosomes 3 and 9 were highly correlated (0.4 to 0.8) with regions adjacent to the putative genes (uncharacterized LOC100382572, bZIP transcription factor 46) (Supplementary Fig S5).

Under non-infested conditions, six gene models were identified around four SNPs associated with grain yield (Table 7). Most of the identified genes were located on chromosomes bins 1.02, 4.08 and 8.02. These gene models encode IQ domain-containing protein IQM5, uncharacterized protein LOC100277298, WRI1 transcription factor 2, ENTH/VHS family protein, Replication protein A 70 kDa DNA-binding and Photosystem I reaction center subunit XI. IQ-domain proteins are common in land plants and they are known for critical roles in host defense, cell shaping and drought resistance.

Discussion

The marked reduction in grain yield observed in the resistant and susceptible benchmark indicates the occurrence of severe parasite infestation across the test environments, eliciting significant differences in resistance or susceptibility reactions among the inbred lines. The diversity panel used in our study displayed considerable phenotypic variation for the three *Striga* resistance-related traits recorded under *Striga* infestation, and this is consistent with the findings in other studies²⁵. The significant line x environment interaction observed for traits measured under *Striga* infestation can be attributed to varying seasonal factors, soil pH, and nutrient levels²⁶. Also, the significant rank correlations among environments for the major *Striga* resistance-related traits indicates that the lines maintained consistent resistance or susceptibility reactions to *Striga* seeds collected from different locations and years. More than 55% of the lines evaluated in this study were resistant to *S. hermonthica*, and this is due to the severe selection pressure imposed by the breeders during the development of these inbred lines from diverse source populations.

Heritability estimates were high for the *Striga* resistance-related traits, indicating the predominant role of genetic factors for these traits. Traits with high heritability increase the power of detecting SNPs in an association panel and thus allow the identification of a true association between a marker and a putative gene¹⁵. The high heritability estimates observed for *Striga* damage ratings and emerged *Striga* plants in this study is consistent with the results reported by Najjar et al.²⁷ and Shayanowako et al.²⁵. These findings, however, differ from those of Badu-Apraku et al.²⁸, who recorded low heritability estimates ($h^2 < 50$) for emerged *Striga* plants and *Striga* damage ratings.

The efficiency of association mapping largely depends on population size and population structure, which infers the ancestry of lines based on their genotypic information²⁹. The diversity panel used in this study are inbred lines derived from broad-based populations containing tropical and temperate germplasm, backcrosses containing *Z. diploperennis* adapted to tropical environments, and some lines that are tolerant to drought³⁰. The two complementary approach used to infer the population structure grouped the inbred lines into three sub-populations based on their genetic backgrounds/pedigree and maturity information. It is worth noting that there was high agreement in the assignment of the inbred lines into the three subgroups based on the two approaches (Admixture and PCA). The phylogeny tree also grouped the inbred lines into three subgroups.

LD is an important factor that determines the power of marker-trait association analysis. In this study, more than 60% of the SNP pairs across the genome exhibit LD at $r^2 > 0.1$. In addition, the high r^2 value observed on chromosomes 4 and 8 in this study can be attributed to fewer recombination events on these chromosomes; this is consistent with the findings of Thirunavukkarasu et al.³¹ and Dinesh et al.³², who reported high r^2 value on chromosome 4 and 8 of maize. Reports have shown that the effectiveness of recombination is limited by the high level of homozygosity³³. In this study, faster decay of LD with increasing distance between markers was observed, which agrees with Doa et al.³⁴ and Dinesh et al.³².

Several studies have been conducted to dissect the genetic architecture of *Striga* resistance in maize, and many QTLs associated with *Striga* resistance have been detected using bi-parental populations^{9–11}. Badu-Apraku et al.¹⁰ identified 116 QTLs associated with four *Striga* resistance-related traits (grain yield, *Striga* damage ratings, ear aspect and emerged *Striga* plants) using bi-parental ($F_{2,3}$) population derived from two early maturing maize lines. In another study, Badu-Apraku et al.⁹ identified 14 QTLs that were associated with three *Striga* resistance-related traits (grain yield, ears per plant and *Striga* damage rating at 10 WAP). However, QTL mapping only exploits only a small fraction of available genetic diversity and exhibits limited capacity to detect polygenic resistance¹². However, only a few GWAS has been conducted to identify genomic regions associated with *Striga hermonthica* resistance in maize. Genome-wide association study exhibits high mapping resolution and abundant genetic variation due to the high ancestral recombination events in natural populations³⁵. Thus, it has been identified as a useful tool for detecting QTNs associated with complex quantitative traits, as well as predicting or identifying causative genes¹⁵.

Different statistical models have been used for GWAS, the multi-locus model exhibits a higher distinctive power and a lower false-positive rate for detecting QTNs compared with the single-locus GWAS model^{18,36}. The adjustments of single-locus GWAS model improves its detection accuracy to an extent, however, the multiple-testing correction (Bonferroni correction) of significance thresholds for single-locus model is too strict. This leads to the exclusion of important loci, especially when large experimental errors occur in field trials of crop genetics. To solve this problem, the application of multi-locus mrMLM is essential. Previous study on GWAS have used the single-locus model to identify quantitative trait nucleotides (QTNs) controlling *Striga* resistance in maize. Adewale et al.²⁴ identified 24 SNPs associated with four *Striga* resistance-related traits in early maturing maize inbred lines using single-locus GWAS model. Further annotation analysis identified three putative genes that explained 9 to 42% of the phenotypic variation. The high phenotypic variation explained can be attributed to the single-locus GWAS model used. However, the genomic regions identified by Adewale et al.²⁴ differs from those discovered in this study.

In this study, the four multi-locus GWAS models used identified thirty significant SNPs associated the three *Striga* resistance-related traits. This study is the first to use multi-locus GWAS model to identify SNPs associated with *S. hermonthica* resistance in maize. Comparing our GWAS results with previous studies on *S. hermonthica*

resistance in maize^{9–11}, there were no similar genomic regions detected, however, additional genomic regions associated with *Striga hermonthica* resistance were identified. The annotation analysis identified gene models with potential involvement in plant growth, development, and defense mechanism. Intriguingly, some of the gene models identified encode transcription factors (TFs) including WRKY14, basic helix-loop-helix (bHLH7), bHLH20, basic-domain leucine-zipper (bZIP46), JUNGBRUNNEN 1 and zinc finger proteins. Transcription factors (TFs) are critical regulators of gene expression in all living organisms. They are involved in plant development, cell signaling, and plant defense response.

Studies have shown that most WRKY TFs respond to pathogen attack and act as both positive and negative regulators in complex defense response network³⁷. Studies have associated WRKY TFs with *S. hermonthica* resistance mechanism in rice. Swarbrick et al.³⁸ reported the up-regulation of genes encoding WRKY TFs in the roots of Nipponbare, a rice variety with resistance to *S. hermonthica*. Also, Mutuku et al.³⁹ reported the knock-down of WRKY45 (WRKY45-kd) by RNA interference in rice plants resulted in susceptibility to *S. hermonthica* infestation. bHLH is another TF, they are commonly expressed in response to drought stress and they have been reported in rice⁴⁰. The bHLH family TFs in *Populus*, PebHLH35 from *Populus euphratica*, has been reported as an essential gene in response to drought by regulating stomatal development and photosynthesis in Arabidopsis⁴¹.

In addition, the Ring zinc-finger domain superfamily proteins are the most significant TFs known for their finger-like structure and ability to bind to zinc. These proteins have been reported in plants such as wheat (*Triticum aestivum*), soybean (*Glycine max*), and rice (*Oryza sativa*)⁴². Cao et al.⁴³ indicated that Ring zinc-finger domain superfamily proteins are involved in resistance to blast fungus infection in rice. The DNA binding with one finger proteins (dofs) also regulate the expression of genes involved in plant development and defense processes⁴⁴. In maize, ZmDof1 has been isolated and connected with C4 photosynthesis⁴⁵, which makes it thrive more than the C3 plants under warmer harsh climates because they are known to be drought resilient. In maize, there are no information on the roles of most of these TFs in *S. hermonthica* resistance; thus, further transcriptomic study will give a better understanding on the role of these TFs in *S. hermonthica* resistance in maize.

Plants generally lack specific cells to defend themselves against attack, but they possess the necessary components for detecting invasion and building up defense response. Xyloglucan endotransglucosylase/hydrolases (XTHs) are cell wall enzymes that are able to graft xyloglucan chains to oligosaccharides⁴⁶. One of its functions in plants is defense reaction against parasitic plants⁴⁷. In tomato, xyloglucan endotransglycosylase/hydrolase (XTH) plays a major role in defense reactions against plant parasite *Cuscuta reflexa*⁴⁷.

Plants have evolved a series of mechanisms to resist pathogens infection. Most plant disease resistance (R) genes contain nucleotide-binding site (NBS) and leucine-rich repeat (LRR) domains. NBS domains could bind and hydrolyze ATP or GTP, while LRR domains are critical for the formation of protein–protein interactions⁴⁸. NBS-LRR proteins have been suggested as the largest class of known R proteins that can either directly or indirectly recognize the presence of pathogens⁴⁹. R gene proteins are involved in pathogen detection and disease resistance⁵⁰. The Recognition of *Peronospora Parasitica* 13-like (*RPP13*-like) genes also play important roles in the resistance of various plant diseases including the downy mildew caused by *Peronospora parasitica*. In Arabidopsis, the *RPP13*-Nd, cloned from an ecotype (Niederzenz (Nd-1)), was characterized to resist the infection of various isolates of *P. parasitica*⁵¹.

Leucine-rich repeat receptor-like protein kinases (LRR-RLKs) are the largest group of receptor-like kinases in plants and play vital roles in development and defense-related processes including cell proliferation, hormone perception, host-specific defense response, wounding response and symbiosis⁵². In Arabidopsis two LRR were identified to regulate cell death and innate immunity⁵³. According to Yuan et al.⁵⁴, LRR-RLK can positively regulate plant biotic resistance but negatively regulate plant abiotic tolerance in Arabidopsis. Interestingly, several RLKs were found to possess dual or multiple roles during plant growth and development. For example, ERECTA is involved in both plant development and pathogen defense responses⁵⁵.

U-box proteins significantly contribute to the ability of plants to respond to diverse environmental stresses, due to plant immobility⁵⁶. The ubiquitination pathway regulates growth, development, and stress responses in plants, and the U-box protein family of ubiquitin ligases plays important role in this pathway. In higher plants, U-box-ARM proteins are associated with regulation of cell death and defense⁵⁷. In addition, VQ proteins regulate diverse developmental processes, including responses to biotic and abiotic stresses, seed development, and photomorphogenesis⁵⁸. Members of the VQ family either play a positive or negative role in plant immune response. Plants with loss of function of VQ23 lack resistance to both *Botrytis cinerea* and *Pseudomonas syringae*. While lines, which overexpresses VQ23, showed reduced disease symptoms upon infection with either pathogen⁵⁸.

Cytochrome P450 superfamily proteins were also associated with emerged *Striga* plants, they are often involved in phytoalexin synthesis and the scavenging of toxins. Plants utilize a wide array of cytochrome P450 monooxygenases (P450s) in biosynthetic and detoxification pathways. Several genes encoding P450s were observed to be highly up-regulated during the resistance response to *S. hermonthica* in rice³⁸. From this study, it was observed that the genomic regions controlling grain yield under *Striga* infestation differs from the genomic region controlling grain yield under non-infested condition. Genomic regions identified to be associated with plant defense mechanisms will be developed into kompetitive allele-specific PCR (KASP) genotyping assay and this will be validated in independent populations to improve *Striga* resistance breeding in maize before deployment for use in marker-assisted selection. Identified SNPs will also help expedite the use of molecular markers in *Striga* resistance improvement through the use of marker-assisted backcrossing (MABC) to advance the effectiveness of breeding for superior and desirable qualities but susceptible to *Striga* infestation.

In conclusion, most of the significant SNPs discovered in this study encode genes associated with plant defense mechanism. Most of the QTLs identified have not been documented in maize, indicating they are novel and are addition to the already identified QTLs for *Striga* resistance in maize from other studies. QTNs identified in this study can be potentially used to expedite the use of marker-assisted selection (MAS) in breeding for durable

Source population	Genetic backgrounds of inbred source population	No of lines evaluated
ZDIP	Inbred lines derived from a backcross (BC4) containing a <i>Zea diploperennis</i> accession as a donor parent plus lines derived from bi-parental crosses involving one parent derived from the same BC4	39
TZLC	Lines derived from a late-maturing composite formed by crossing TZB-SR with seven field resistant maize inbred lines against <i>S. hermonthica</i> plus lines derived from bi-parental crosses involving one parent derived from the same source populations	44
TZEC	Lines developed from an early maturing composite formed by crossing TZESR-W C3 with eight field resistant maize inbred lines against <i>S. hermonthica</i> plus lines derived from bi-parental crosses involving one parent derived from the same composite	13
IWDS	Lines extracted from a synthetic formed from medium maturing white inbred lines and improved for resistance to <i>Striga</i> and drought plus lines derived from bi-parental crosses involving one parent derived from these synthetic	30
MIXED	Lines derived from diverse source populations plus tolerant lines extensively used as donors of field resistance to form resistance source populations	24

Table 8. List of source populations for the 150 inbred lines used in this study.

resistance to *S. hermonthica* in maize. The chromosomal regions controlling the *Striga* resistance-related traits can also be exploited for selection and effective pyramiding of favorable alleles in *Striga* improvement. Since most of the maize lines used in this study were developed at IITA and have a diverse response to *Striga* infestation, this study will contribute to molecular-marker based transformation of *Striga* resistance breeding in maize.

Materials and methods

Genetic materials. A diversity panel of 150 maize inbred lines used in this study was developed by the Maize breeding program of the International Institute of Tropical Agriculture (IITA-Ibadan). The maize inbred lines in this panel were at S_{7,9} stages of inbreeding and had varying reactions to *S. hermonthica* (Supplementary Table S4). Summaries of the genetic backgrounds of source populations of these inbred lines are provided in Table 8. Ten inbred lines with either known resistance (9450), tolerance (5012, 1393, 1368, 4001, 9030, 9071, KU1414-SR, and MMB90) and susceptibility (5057) reactions to *S. hermonthica* were included as benchmarks to assess the performance of the 150 lines.

Phenotypic evaluation and trait measurements. The inbred lines were evaluated under *Striga* infested and non-infested conditions at Abuja (9° 15' N, 7° 20' E; 490 m asl) and Mokwa (9° 21' N, 5° 10' E; 210 m asl) in Nigeria during the main rainy seasons of 2017 and 2018. The experiment was laid out in a 15 X 10 alpha lattice design with two replications planted in a crisscross arrangement. Each experimental unit was planted in adjacent infested and non-infested strips, located opposite each other and separated by a 1.5 m alley. An inbred line was planted within each strip in a 4 m long row, with 0.75 m inter-row spacing and 0.25 m intra-row spacing. Ethylene gas was sprayed two weeks before planting to induce suicidal germination of *Striga* seeds in the soil.

Two maize seeds were planted in a 6 cm deep hole inoculated with 8.5 g of sand mixed with *Striga* seeds. The sand-*Striga* mixture contains approximately 3000 germinable *Striga* seeds. The *Striga* seeds used in this study were collected from sorghum field from the previous planting season in Mokwa and Abuja with farmers' consent before usage. Two weeks after planting, maize plants were thinned to one plant per hill to attain a population density of 53,333 plant/ha. Fertilizer was applied at the rate of 30 kg/ha of nitrogen, 60 kg/ha each of phosphorus and potassium at planting and an additional 30 kg/ha nitrogen was applied four weeks later. Weeds other than *Striga* were removed from plots manually throughout the planting season. Data were taken under both infested and non-infested conditions, except for *Striga* damage score and *Striga* emergence, recorded only under *Striga* infestation. Data recorded under the two environments included plant height, ear aspect, and grain yield (Supplementary Table S5). Ears were collected separately from each line and shelled to estimate per cent moisture in the grain. Grain yield was then calculated from grain weight adjusted for 15% moisture. This study is geared towards improving IITA maize breeding, and it complies with the country's local and national regulations.

Data analysis. *Phenotypic analysis.* Analysis of variance combined across the four year-location, which were hereafter referred to as environments, were computed for all traits measured under *Striga* infested and non-infested conditions based on a mixed-model analysis with restricted maximum likelihood procedure in SAS version 9.4⁵⁹. In this analysis, genotypes were considered fixed while all other factors were random.

Separate analyses were conducted for traits measured under infested and non-infested conditions. The mixed model analysis generated the best linear unbiased estimates (BLUEs), the variance components and broad-sense heritability estimates. In addition, Spearman rank correlations between pairs of environments were computed for the *Striga* resistance-related traits using SAS version 9.4. Also, correlation analysis among the different traits was performed using R software, and results were displayed as heat map.

Genotyping and filtering. Genomic DNA was extracted from young leaf samples of the 150 maize inbred lines using the modified cetyltrimethylammonium bromide (CTAB) protocol⁶⁰. Purified DNA was sent to Elshire facility in New Zealand for genotyping-by-sequencing (GBS)⁶¹. Genomic DNA was digested with the restriction enzyme (*ApeK1*), and genotyping-by-sequencing (GBS) of the libraries were constructed in 96-plex and sequenced on Illumina HiSeq2500 following manufacturer's protocol. Raw flow cell output was processed to genotype calls using the TASSEL-GBS pipeline. Reads and tags found in each sequencing result were aligned with the *Zea mays* L. genome reference, version AGPV3 (B73 Ref Gen v4 assembly). SNPs with minor allele

frequency (MAF) of <0.05 and missing rate of >10% were excluded from the genotyping dataset using PLINK 1.9 beta⁶².

Population structure and linkage disequilibrium. To explore the genetic relationship among the inbred lines, principal component analysis (PCA) was conducted using factorMiner package in R⁶³. The pairwise genetic distance was calculated through identity-by-state (IBS), and the phylogenetic tree was generated using analysis of phylogenetics and evolution (ape) R package⁶⁴. The population stratification among the inbred lines was assessed using Admixture software⁶⁵. The method uses maximum likelihood estimation on data from many loci to estimate individual ancestries among the inbred lines. The analysis was performed using a cross-validation error (k) varying from 2 to 10. The most appropriate k -value selected exhibit low cross-validation error compared to other k -values. LD among markers was calculated using PLINK software. The window size for LD calculation was set based on the number of SNPs located in the genome. Pairwise linkage disequilibrium was measured using the squared allele frequency correlations, according to Weir⁶⁶, and assessed by calculating r^2 for pairs of SNP loci.

Marker-trait association analysis. All the phenotypic and genotypic information from the 150 diverse maize inbred lines were used to detect SNPs using four of the GWAS multi-locus models, multi-locus random-SNP-effect MLM (mrMLM), fast multi-locus random-SNP-effect mixed linear model (FASTmrMLM), fast multi-locus random-SNP-effect efficient mixed model association (FASTmrEMMA) and polygenic-background-control-based least angle regression empirical Bayes (pLARmEB), implemented in mrMLM v4.0 (<https://cran.r-project.org/web/packages/mrMLM.GUI/index.html>). The unified parameter settings for the four methods were as follows; the Q + K model was used, where the population structure value Q was calculated by Admixture software⁶⁵ and the kinship value K was analyzed by the “mrMLM” software package. The limit of detection (LOD) score was set to 6 for robust QTNs for all measured trait. The Manhattan and QQ plots for GWAS were displayed using the R package CMplot.

Gene annotation. SNPs detected for *Striga hermonthica* resistance-related traits by the four mrMLM methods were mapped to the maize reference genome B73 RefGen_V4 to identify associated candidate genes. The genes corresponding to each QTN was determined in MaizeGDB according to its physical position. The functional annotations of the candidate genes were predicted in NCBI. The Pairwise LD estimates in the region of interest for significantly associated markers were investigated using Haploview 4.2. Finally, LD plotting was done based on base pairs (bp) distance, using “ggplot2” package in R⁶⁷.

Marker effect prediction and variants comparison. Variants (ref/alt) associated with significant SNPs were identified using *rstatix* package implemented in R, and their effect were compared using ANOVA $p < 0.05$. The nature of the SNP marker (favorable and unfavorable) was defined based on the direct contribution to the traits using *rstatix* and visualized using ggplot2.

Data availability

Phenotypic data presented are within this document and the genotypic data can be provided upon request.

Received: 2 July 2021; Accepted: 18 November 2021

Published online: 17 December 2021

References

- Nyaga, C. *et al.* Genome-wide analyses and prediction of resistance to MLN in large tropical maize germplasm. *Genes* **11**(1), 16. <https://doi.org/10.3390/genes11010016> (2020).
- Menkir, A., Badu-Apraku, B., Yallou, C. G., Kamara, A. Y. & Ejeta, G. Breeding maize for broad-based resistance to *Striga hermonthica*. In Integrating new technologies for Striga control: Towards ending the witch-hunt, 99–114. https://doi.org/10.1142/9789812771506_0008 (2007).
- Menkir, A., Chikoye, D. & Lum, F. Incorporating an herbicide resistance gene into tropical maize with inherent polygenic resistance to control *Striga hermonthica* (Del) Benth. *Plant Breed.* **129**(4), 385–392. <https://doi.org/10.1111/j.1439-0523.2009.01687.x> (2010).
- Pageau, K., Simier, P., Nault, N., Robins, R. & Fer, A. Carbon dependency of the hemiparasite *Striga hermonthica* on *Sorghum bicolor* determined by carbon isotopic and gas exchange analyses. *Funct. Plant Biol.* **25**(6), 695–700. <https://doi.org/10.1071/PP98047> (1998).
- Rich, P. J. & Ejeta, G. Towards effective resistance to *Striga* in African maize. *Plant Signal. Behav.* **3**(9), 618–621. <https://doi.org/10.4161/psb.3.9.5750> (2008).
- Gedil, M. & Menkir, A. An integrated molecular and conventional breeding scheme for enhancing genetic gain in maize in Africa. *Front. Plant Sci.* **10**, 1430. <https://doi.org/10.3389/fpls.2019.01430> (2019).
- Watling, J. R. & Press, M. C. Impacts of infection by parasitic angiosperms on host photosynthesis. *Plant Biol.* **3**(3), 244–250. <https://doi.org/10.1055/s-2001-15195> (2001).
- Berner, D., Winslow, M., Awad, A., Cardwell, K., Raj, D. & Kim, S. *Striga* research methods: A manual (1997).
- Amusan, I. O. Mechanisms and quantitative trait loci for *Striga hermonthica* resistance in maize (*Zea mays* L.) inbred line. PhD thesis at Prudue University (2010).
- Badu-Apraku, B., Adewale, S., Paterne, A., Gedil, M. & Asiedu, R. Identification of QTLs controlling resistance/tolerance to *Striga hermonthica* in an extra-early maturing yellow maize population. *Agronomy* **10**(8), 1168. <https://doi.org/10.3390/agronomy10081168> (2020).
- Badu-Apraku, B. *et al.* Identification of QTLs for grain yield and other traits in tropical maize under *Striga* infestation. *PloSone* **15**(9), e0239205. <https://doi.org/10.1371/journal.pone.0239205> (2020).
- Kraakman, A. T., Niks, R. E., Van den Berg, P. M., Stam, P. & Van Eeuwijk, F. A. Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. *Genetics* **168**(1), 435–446 (2004).

13. Gómez, G., Álvarez, M. F. & Mosquera, T. Association mapping, a method to detect quantitative trait loci: statistical bases. *Agronomía Colombiana* **29**(3), 367–376 (2011).
14. Darvishzadeh, R. Population structure, linkage disequilibrium and association mapping for morphological traits in sunflower (*Helianthus annuus* L.). *Biotechnol. Biotechnol. Equip.* **30**(2), 236–246 (2016).
15. Brachi, B., Morris, G. P. & Borevitz, J. O. Genome-wide association studies in plants: the missing heritability is in the field. *Genome Biol.* **12**(10), 1–8. <https://doi.org/10.1186/gb-2011-12-10-232> (2011).
16. Zhang, Z. *et al.* Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.* **42**(4), 355–360. <https://doi.org/10.1038/ng.546> (2010).
17. Li, M. *et al.* Enrichment of statistical power for genome-wide association studies. *BMC Biol.* **12**(1), 1–10. <https://doi.org/10.1186/s1291> (2014).
18. Wang, S. B. *et al.* Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology. *Sci. Rep.* **6**(1), 1–10. <https://doi.org/10.1038/srep19444> (2016).
19. Tamba, C. L., Ni, Y. L. & Zhang, Y. M. Iterative sure independence screening EM-Bayesian LASSO algorithm for multi-locus genome-wide association studies. *PLoS Comput. Biol.* **13**(1), e1005357. <https://doi.org/10.1371/journal.pcbi.1005357> (2017).
20. Zhang, J. *et al.* pLARmEB: integration of least angle regression with empirical Bayes for multilocus genome-wide association studies. *Heredity* **118**(6), 517–524. <https://doi.org/10.1038/hdy.2017.8> (2017).
21. Wen, Y. J. *et al.* Methodological implementation of mixed linear models in multi-locus genome-wide association studies. *Brief. Bioinform.* **19**(4), 700–712. <https://doi.org/10.1093/bib/bbw145> (2018).
22. Tamba, C. L. & Zhang, Y. M. A fast mrMLM algorithm for multi-locus genome-wide association studies. *Biorxiv* <https://doi.org/10.1101/341784> (2018).
23. Xu, Y. *et al.* Genome-wide association mapping of starch pasting properties in maize using single-locus and multi-locus models. *Front. Plant Sci.* **9**, 1311 (2018).
24. Adewale, S. A. *et al.* Genome-wide association study of Striga resistance in early maturing white tropical maize inbred lines. *BMC Plant Biol.* **20**, 1–16. <https://doi.org/10.1186/s12870-020-02360-0> (2020).
25. Shayanowako, A. I. T., Shimelis, H., Laing, M. D. & Mwadzingeni, L. Genetic diversity of maize genotypes with variable resistance to *Striga asiatica* based on SSR markers. *Cereal Res. Commun.* **46**(4), 668–678. <https://doi.org/10.1556/0806.46.2018.044> (2018).
26. Menkir, A. & Meseka, S. Genetic improvement in resistance to Striga in tropical maize hybrids. *Crop Sci.* **59**(6), 2484–2497. <https://doi.org/10.2135/cropsci2018.12.0749> (2019).
27. Najjar, Z. A. *et al.* Genotypic and morphological diversity analysis in high altitude maize (*Zea mays* L.) inbreds under Himalayan temperate ecologies. *Maydica* **63**(1), 7 (2018).
28. Badu-Apraku, B., Fakorede, M. A. B. & Lum, A. F. Evaluation of experimental varieties from recurrent selection for Striga resistance in two extra-early maize populations in the savannas of West and Central Africa. *Exp. Agric.* **43**, 183–200. <https://doi.org/10.1017/S0014479706004601> (2007).
29. Lawson, D. J. & Falush, D. Population identification using genetic data. *Annu. Rev. Genomics Hum. Genet.* **13**, 337–361 (2012).
30. Ewansha, S. U., Menkir, A. & Tofa, A. I. Agronomic response of drought-tolerant and Striga-resistant maize cultivars to nitrogen fertilization in the Nigerian Guinea savannas. *Maydica* **57**(2), 114–120 (2012).
31. Thirunavukkarasu, N. *et al.* Unraveling the genetic architecture of subtropical maize (*Zea mays* L.) lines to assess their utility in breeding programs. *BMC Genomics* **14**(1), 877. <https://doi.org/10.1186/1471-2164-14-877> (2013).
32. Dinesh, A. *et al.* Genetic diversity, linkage disequilibrium and population structure among CIMMYT maize inbred lines, selected for heat tolerance study. *Maydica* **61**(3), 1–7 (2016).
33. Zhu, C., Gore, M., Buckler, E. S. & Yu, J. Status and prospects of association mapping in plants. *Plant Genome* <https://doi.org/10.3835/plantgenome2008.02.0089> (2008).
34. Dao, A., Sanou, J., Mitchell, S. E., Gracen, V. & Danquah, E. Y. Genetic diversity among INERA maize inbred lines with single nucleotide polymorphism (SNP) markers and their relationship with CIMMYT, IITA, and temperate lines. *BMC Genet.* **15**(1), 1–14. <https://doi.org/10.1186/s12863-014-0127-2> (2014).
35. Ersoz, E. S., Yu, J. & Buckler, E. S. Applications of linkage disequilibrium and association mapping in maize. *Biotechnol. Agric. For.* **63**, 173–195 (2009).
36. Segura, V. *et al.* An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nat. Genet.* **44**(7), 825–830. <https://doi.org/10.1038/ng.2314> (2012).
37. Chandran, N. K., Sriram, S., Prakash, T. & Budhwar, R. Transcriptome changes in resistant and susceptible rose in response to powdery mildew. *J. Phytopathol.* **169**(9), 556–569 (2021).
38. Swarbrick, P. J. *et al.* Global patterns of gene expression in rice cultivars undergoing a susceptible or resistant interaction with the parasitic plant *Striga hermonthica*. *New Phytol.* **179**(2), 515–529. <https://doi.org/10.1111/j.1469-8137.2008.02484.x> (2008).
39. Mutuku, J. M. *et al.* The WRKY45-dependent signaling pathway is required for resistance against *Striga hermonthica* parasitism. *Plant Physiol.* **168**(3), 1152–1163. <https://doi.org/10.1104/pp.114.256404> (2015).
40. Moumeni, A. *et al.* Comparative analysis of root transcriptome profiles of two pairs of drought-tolerant and susceptible rice near-isogenic lines under different drought stress. *BMC Plant Biol.* **11**(1), 1–17. <https://doi.org/10.1186/1471-2229-11-174> (2011).
41. Dong, Y. *et al.* A novel bHLH transcription factor PebHLH35 from *Populus euphratica* confers drought tolerance through regulating stomatal development, photosynthesis and growth in Arabidopsis. *Biochem. Biophys. Res. Commun.* **450**(1), 453–458. <https://doi.org/10.1016/j.bbrc.2014.05.139> (2014).
42. Han, G. *et al.* C2H2 zinc finger proteins: master regulators of abiotic stress responses in plants. *Front. Plant Sci.* **11**, 115 (2020).
43. Cao, H. *et al.* Characterization of 47 Cys2-His2 zinc finger proteins required for the development and pathogenicity of the rice blast fungus *Magnaporthe oryzae*. *New Phytol.* **211**(3), 1035–1051. <https://doi.org/10.1111/nph.13948> (2016).
44. Weng, L. *et al.* The zinc finger transcription factor SlZFP2 negatively regulates abscisic acid biosynthesis and fruit ripening in tomato. *Plant Physiol.* **167**(3), 931–949. <https://doi.org/10.1104/pp.114.255174> (2015).
45. Yanagisawa, S. & Izui, K. Molecular cloning of two DNA-binding proteins of maize that are structurally different but interact with the same sequence motif. *J. Biol. Chem.* **268**(21), 16028–16036. [https://doi.org/10.1016/S0021-9258\(18\)82353-5](https://doi.org/10.1016/S0021-9258(18)82353-5) (1993).
46. Maris, A. *et al.* Differences in enzymic properties of five recombinant xyloglucan endotransglucosylase/hydrolase (XTH) proteins of Arabidopsis thaliana. *J. Exp. Bot.* **62**(1), 261–271 (2011).
47. Albert, M., Werner, M., Proksch, P., Fry, S. C. & Kaldenhoff, R. The cell wall-modifying xyloglucan endotransglucosylase/hydrolase LeXTH1 is expressed during the defence reaction of tomato against the plant parasite *Cuscuta reflexa*. *Plant Biol.* **6**(04), 402–407 (2004).
48. Bittner-Eddy, P. D., Crute, I. R., Holub, E. B. & Beynon, J. L. RPP13 is a simple locus in Arabidopsis thaliana for alleles that specify downy mildew resistance to different avirulence determinants in *Peronospora parasitica*. *Plant J.* **21**(2), 177–188 (2000).
49. Boyes, D. C., Nam, J. & Dangl, J. L. The Arabidopsis thaliana RPM1 disease resistance gene product is a peripheral plasma membrane protein that is degraded coincident with the hypersensitive response. *Proc. Natl. Acad. Sci.* **95**(26), 15849–15854 (1998).
50. Rairdan, G. J. & Moffett, P. Distinct domains in the ARC region of the potato resistance protein Rx mediate LRR binding and inhibition of activation. *Plant Cell* **18**(8), 2082–2093 (2006).
51. Bittner-Eddy, P. *et al.* Genetic and physical mapping of the RPP13 locus, in Arabidopsis, responsible for specific recognition of several *Peronospora parasitica* (downy mildew) isolates. *Mol. Plant Microbe Interact.* **12**(9), 792–802 (1999).

52. Li, J. *et al.* Specific ER quality control components required for biogenesis of the plant innate immune receptor EFR. *Proc. Natl. Acad. Sci.* **106**(37), 15973–15978 (2009).
53. Gao, M. *et al.* Regulation of cell death and innate immunity by two receptor-like kinases in Arabidopsis. *Cell Host Microbe* **6**(1), 34–44 (2009).
54. Yuan, N. *et al.* STRESS INDUCED FACTOR 2, a leucine-rich repeat kinase regulates basal plant pathogen defense. *Plant Physiol.* **176**(4), 3062–3080 (2018).
55. Godiard, L. *et al.* ERECTA, an LRR receptor-like kinase protein controlling development pleiotropically affects resistance to bacterial wilt. *Plant J.* **36**(3), 353–365. <https://doi.org/10.1046/j.1365-313X.2003.01877.x> (2003).
56. Patterson, C. (2002) A new gun in town: the U box is a ubiquitin ligase domain. *Sci. Signal.* **116**, pe4 (2002).
57. Yang, C. W. *et al.* The E3 ubiquitin ligase activity of Arabidopsis PLANT U-BOX17 and its functional tobacco homolog ACRE276 are required for cell death and defense. *Plant Cell* **18**(4), 1084–1098. <https://doi.org/10.1105/tpc.105.039198> (2006).
58. Jing, Y. & Lin, R. The VQ motif-containing protein family of plant-specific transcriptional regulators. *Plant Physiol.* **169**(1), 371–378 (2015).
59. Vargas, M. *et al.* META: a suite of SAS programs to analyze multi-environment breeding trials. *Agron. J.* **105**(1), 11–19. <https://doi.org/10.2134/agronj2012.0016> (2013).
60. Saghai-Marouf, M. A., Soliman, K. M., Jorgensen, R. A. & Allard, R. W. L. Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. *Proc. Natl. Acad. Sci.* **81**(24), 8014–8018 (1984).
61. Elshire, R. J. *et al.* A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* **6**(5), e19379. <https://doi.org/10.1371/journal.pone.0019379> (2011).
62. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Human Genet.* **81**(3), 559–575 (2007).
63. Lê, S., Josse, J. & Husson, F. FactoMineR: an R package for multivariate analysis. *J. Stat. Softw.* **25**(1), 1–18 (2008).
64. Paradis, E., Claude, J. & Strimmer, K. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**(2), 289–290 (2004).
65. Alexander, D. H. & Lange, K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinform.* **12**(1), 1–6 (2011).
66. Weir, B. S. Inferences about linkage disequilibrium. *Biometrics* **35**, 235–254 (1979).
67. Wickham, H., Chang, W. & Wickham, M. H. Package 'ggplot2'. *Create Elegant Data Visualisations Using the Grammar of Graphics. Version, 2*(1), 1–189 (2016)

Acknowledgements

Stanley, Adekemi E. acknowledges funding from Deutscher Akademischer Austauschdienst (DAAD) and the Africa Centres of Excellence for Development Impact (ACE) to pursue PhD in Plant Breeding at WACCI, University of Ghana, Legon, Accra. This study was funded by the Bill and Melinda Gates Foundation through the AGG project (Accelerating Genetic Gains in Maize and Wheat for Improved Livelihoods; B&MGF Investment ID INV-003439), and the CGIAR Research Program MAIZE. The CGIAR Research Program MAIZE received W1&W2 support from the Governments of Australia, Belgium, Canada, China, France, India, Japan, Korea, Mexico, Netherlands, New Zealand, Norway, Sweden, Switzerland, United Kingdom, and the United States, as well as the World Bank.

Author contributions

A.M., M.G., B.I., P.T., and K.O. designed the research presented here, S.M., W.M., A.S. executed the field research. P.A.A, N.U., S.C., O.O. and A.S. conducted the analyses. M.G., N.U., P.A.A. and A.S. managed the genotyping. A.S. wrote the first draft and all authors read and edited the subsequent versions of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-03566-4>.

Correspondence and requests for materials should be addressed to A.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021