CLASSIFICATION: BIOLOGICAL SCIENCES: Evolution

TITLE: Genomics of sorghum local adaptation to a parasitic plant

AUTHORS: Emily S. Bellis[1], Victoria L. DeLeo[1], Elizabeth A. Kelly[1], Claire M. Lorts[1], Germinal Rouhan[2], Andrew Budden[3], Govinal Badiger Bhaskara[4], Zhenbin Hu[5], Robert Muscarella[6], Thomas E. Juenger[4], Michael P. Timko[7], Geoffrey P. Morris[5], Claude W. dePamphilis[1], Jesse R. Lasky[1]

[1]Department of Biology, Pennsylvania State University, University Park, PA, 16802
[2]Institut Systématique Evolution Biodiversité (ISYEB), Muséum national d'Histoire naturelle, CNRS, Sorbonne Université, EPHE, 57 rue Cuvier, CP39, 75005 Paris, France
[3]Royal Botanic Gardens, Kew, Richmond TW9 3AB, UK
[4]Department of Integrative Biology, 2415 Speedway #C0930, University of Texas, Austin, TX, 78712
[5]Department of Agronomy, Kansas State University, Manhattan, Kansas, 66506
[6]Department of Plant Ecology and Evolution, Evolutionary Biology Centre, Uppsala University, SE-75236 Uppsala, Sweden
[7]Department of Biology, University of Virginia, Charlottesville, Virginia, 22904

CORRESPONDING AUTHOR: Emily S. Bellis, Department of Biology, Pennsylvania State University, University Park, PA, 16802, ezb336@psu.edu

ABSTRACT:

Host-parasite coevolution can maintain high levels of genetic diversity in traits involved in species interactions. In many systems, host traits exploited by parasites are constrained by use in other functions, leading to complex selective pressures across space and time. Here, we study genome-wide variation in the staple crop *Sorghum bicolor* (L.) Moench and its association with the parasitic weed *Striga hermonthica* (Delile) Benth., a major constraint to food security in many African countries. We hypothesize that sorghum landraces are subject to geographic selection mosaics within parasite-prone areas and selection against resistance where *S. hermonthica* is never found. Supporting this hypothesis, multiple independent loss-of-function alleles at sorghum *LOW GERMINATION STIMULANT 1* (*LGS1),* a locus known to impact resistance, are broadly distributed among African landraces and geographically associated with *S. hermonthica* occurrence, suggesting a role in local adaptation to parasite pressure. However, the low frequency of these alleles within *S. hermonthica*-prone regions and their absence elsewhere indicates potential trade-offs restricting their distribution. *LGS1* impacts stereochemistry of strigolactones, hormones controlling plant architecture, belowground signaling with other organisms, and abiotic stress tolerance. Supporting trade-offs, transcriptome profiling of nutrient-stressed roots revealed differential regulation of several strigolactone biosynthesis and signaling genes in *LGS1*-deficient sorghum compared to a susceptible line. Signatures of balancing selection surrounding *LGS1* and candidates from analysis of genome-wide associations with parasite distribution support long-term maintenance of diversity in parasite resistance genes. Our study of host resistance evolution across smallholder agroecosystems provides a valuable contrast to both industrial farming systems and natural communities.

KEYWORDS: species distribution modeling, environmental niche modeling, genotype-environment association analysis, Red Queen

SIGNIFICANCE STATEMENT:

Understanding co-evolution in crop-parasite systems is critical to management of myriad pests and pathogens confronting modern agriculture. In contrast to wild plant communities, parasites in agricultural ecosystems are usually expected to gain the upper hand in co-evolutionary 'arms races' due to limited genetic diversity of host crops in cultivation. Here, we develop a framework for studying associations between genome diversity in global landraces (traditional varieties) of the staple crop sorghum with the distribution of the devastating parasitic weed *Striga hermonthica.* We find long-term maintenance of diversity in genes related to parasite resistance, highlighting an important role of host adaptation for co-evolutionary dynamics in smallholder agroecosystems.

INTRODUCTION:

Host-parasite interactions can be powerful and dynamic selective forces maintaining genetic variation in natural populations (1). In wild plant pathosystems, long-term balancing selection often maintains diverse resistance alleles in host populations over evolutionary time (2–4). Resistance polymorphism can be maintained by negative frequency-dependent selection, which drives cycling of resistance and virulence alleles when rare alleles provide a selective advantage (i.e. fluctuating Red Queen dynamics *sensu* (5)) (2, 6). Costs of resistance can also generate allele frequency cycles and maintain diversity over gradients of parasite pressure (7).

In contrast to wild plant communities where fluctuating Red Queen dynamics have frequently been observed, low host diversity in agricultural settings is often assumed to permit runaway 'arms races' in fast-evolving parasites (4, 8). Relative to smallholder farms, however, industrial scale farming accounts for a fraction of global production for many food crops (9). The dynamics of host adaptation to parasites across diverse smallholder agricultural systems remains poorly known, despite relevance for identifying novel resistance alleles and managing crop genetic resources (e.g. preserving germplasm both *ex situ* and *in situ*; Jensen et al. 2012). Are co-evolutionary dynamics in smallholder farming systems more similar to natural plant pathosystems, where high connectivity among genetically diverse patches can help promote evolution of host resistance (11)?

85 Current approaches for identifying and studying the evolution of resistance
86 alleles often involve scoring large panels of diverse individuals (in genome-wide
87 association studies, GWAS) or many recombinant individuals deriving from controlled
88 crosses (in linkage mapping) and using DNA sequence information to identify genomic
89 regions associated with parasite susceptibility. These mapping studies have revealed
90 numerous insights to mechanisms of plant-pathogen dynamics (12), but require
91 extensive phenotyping and genotyping effort for adequate statistical power. To
92 understand the genetic basis of response to abiotic stress, another 'bottom-up'
93 approach is to leverage analysis of the extensive genetic variation present in landraces,
94 traditional crop or livestock varieties which have spread and subsequently adapted to
95 diverse agro-ecological environments following domestication (13–16). Compared to
96 modern improved varieties, which may have lost many resistances due to population
97 bottlenecks and selection for performance in optimal environments (17), landraces and
98 crop wild relatives may be a rich source of alleles conferring resistance to sympatric
99 parasites under a hypothesis of local adaptation to environment of origin. Furthermore,
100 landraces can be studied to test the hypothesis that a natural, putatively
101 environmentally-adapted allele identified from a limited set of experimental
102 environments and genetic backgrounds is indeed adaptive across a wide range of
103 similar environments and across diverse genetic backgrounds (14). Genotype-
104 environment association (GEA) analyses of georeferenced landraces have been a
105 powerful strategy for understanding the genetic basis of local adaptation to gradients of
106 abiotic stressors (14, 15). To our knowledge, similar approaches to investigate
107 adaptation to biotic stress at continent scale have yet to be exploited (but see (16) for
108 an example in Ugandan cattle).
109 In this study, we extend GEA to biotic gradients to evaluate the frequency and
110 distribution of alleles in genomes of sorghum that confer resistance to *Striga*
111 *hermonthica* (Delile) Benth., a root-parasitic weed native to Africa. *Sorghum bicolor* (L.)
112 Moench is the world's fifth most important crop and was domesticated from the wild
113 progenitor *Sorghum arundinaceum* (Desv.) Stapf in Africa more than 5000 years ago
114 (18). Sorghum is particularly important in arid and semiarid regions due to its drought
115 tolerance compared to maize and rice. While plant drought responses are influenced by

116 many processes, in recent years the role of strigolactones, plant hormones that regulate
117 shoot branching (19), root architecture (20) and response to abiotic stress (21), have
118 received particular attention. Strigolactones exuded into the rhizosphere, particularly
119 under nutrient limitation (22) and drought (23), promote interactions with beneficial
120 arbuscular mycorrhizal fungi (24) and are also required as germination stimulants for
121 many root-parasitic plants. In addition to sorghum, *S. hermonthica* attacks maize, rice,
122 millet, sugarcane, and wild grasses and is considered one of the greatest biotic threats
123 to food security in Africa, costing billions of dollars (USD) in crop losses annually
124 through yield loss and abandonment of fields (25).

125 Here, we evaluate the hypothesis that a geographical selection mosaic across
126 gradients of biotic interactions maintains genetic diversity in *S. hermonthica* resistance
127 in sorghum landraces (26, 27). To identify genomic signatures of host adaptation to
128 parasites, we first develop species distribution models for *S. hermonthica* and search for
129 statistical associations between sorghum genotype and prediction from the parasite
130 model at collection location of each sorghum landrace. We characterize diversity and
131 geographic distribution of loss-of-function alleles at the sorghum resistance gene *LOW*
132 *GERMINATION STIMULANT 1* (*LGS1*) (28). Loss of *LGS1,* a gene of unknown
133 function with a sulfotransferase domain, alters stereochemistry of the dominant
134 strigolactone in sorghum root exudates, from 5-deoxystrigol to orobanchol, which does
135 not strongly stimulate parasite germination (28, 29). Our ecological genetic analyses of
136 diverse sorghum landraces suggest that *LGS1* loss-of-function mutations are adaptive
137 across a large region of high *S. hermonthica* prevalence in Africa. However, we also
138 present evidence supporting potential trade-offs related to *LGS1* loss-of-function.
139 Because endogenous strigolactones are integral to host plant development, rhizosphere
140 communications, and stress response, we predicted that *LGS1* loss-of-function may
141 adversely impact fitness- or domestication-related traits through altered hormone
142 signaling and perception. We compare root transcriptomes of sorghum lines producing
143 high levels of 5-deoxystrigol or orobanchol to determine if structural variation in SLs
144 impacts expression of strigolactone biosynthesis and signaling pathway components.
145 After focused analyses on *LGS1,* we perform genome-wide tests of association with
146 parasite distribution. We investigate patterns of polymorphism surrounding candidate

147 resistance genes with evidence of locally-adaptive natural variation to determine
148 whether balancing selection has maintained diversity in *S. hermonthica* resistance over
149 evolutionary time.

150 **RESULTS:**

151 S. hermonthica *distribution model*

152 We predicted that host alleles affecting resistance or tolerance would be strongly
153 associated with the geographic distribution of parasites. To identify regions of likely *S.*
154 *hermonthica* occurrence in the absence of continent-wide surveys, we built MaxEnt
155 species distribution models (SDMs) (30). The optimal model showed good ability to
156 predict occurrence with an AUC value of 0.86 (Fig. 1). A high degree of overlap was
157 observed between models generated using all *S. hermonthica* records (*n*=1,050) and a
158 subset of 262 occurrences that were observed specifically in fields of sorghum
159 (Schoener's *D* = 0.82; *I* = 0.97). Annual rainfall and total soil N were the most
160 informative variables for predicting *S. hermonthica* occurrence (Table S1). Compared to
161 all cells in the study background, distributions of environmental values for locations with
162 high habitat suitability (HS) were generally restricted to a narrower range of intermediate
163 values of precipitation and soil quality (Fig. S1-2). Locations with the highest HS scores
164 exhibited mean annual rainfall ranging from ~500-1300 mm/year and soil nitrogen
165 ranging from ~400 – 1000 g/kg (10th-90th percentiles for all cells with HS >0.5, Table
166 S1). Soil clay content also contributed substantially to the sorghum-only model, and
167 clay content in locations with the highest HS scores ranged from 12-29% (10th-90th
168 percentile, HS > 0.5, all-occurrence model) or up to 36% (sorghum-only model, Table
169 S1).
170

171 LGS1 *associations with* S. hermonthica *occurrence*

172 We predicted that sorghum resistance alleles would be more common in
173 parasite-prone regions. Evaluating this prediction also allowed us to validate our SDM-
174 based genotype-environment association approach by characterizing associations
175 between *S. hermonthica* distribution and genetic variation at *LGS1*
176 (*Sobic.005G213600*), a known locus causing resistance to *Striga* spp. (28).

177 Using whole genome sequencing data (~25x coverage) from 143 sorghum
178 landraces, we found evidence for three naturally occurring mutations resulting in *LGS1*
179 loss-of-function (Fig. 2). Two ~30 kb deletions were identified between positions
180 69,958,377-69,986,892 (*n* = 5 accessions) and 69,981,502-70,011,149 (*n* = 4
181 accessions, Table S2; Fig. 2A). These two deletions appear to be identical to previously
182 described *lgs1-2* and *lgs1-3* (28), although breakpoint positions reported by our
183 structural variant caller differed slightly. No SNPs in a separate genotyping-by-
184 sequencing (GBS) dataset of >2000 sorghum landraces tagged *lgs1-2* or *lgs1-3* (Fig.
185 S3), and so we imputed large deletions identified from the WGS dataset to the GBS
186 dataset based on patterns of missing data (see Methods). Deletion call imputations
187 were validated by testing root exudate from a subset of sorghum accessions for their
188 ability to induce *S. hermonthica* germination (Fig. S4). We tested four genotypes with
189 likely deletion alleles, and these genotypes stimulated significantly fewer *Striga* seeds to
190 germinate compared to eight other genotypes that did not show strong evidence for
191 deletions (linear mixed-effects model with genotype random effect, deletion genotypes
192 stimulated germination of 8.43 fewer seeds out of 75 total, Wald 95% CI=1.49,15.36).
193 Two of the genotypes stimulated low germination (P721 and SC1439) and had all 14
194 GBS SNPs missing from *LGS1*, but SNP calls outside of the gene model were
195 inconsistent with *lgs1-2* or *lgs1-3* alleles, perhaps due to the presence of an additional
196 uncharacterized deletion allele.
197 In addition to *lgs1-2* and *lgs1-3*, we identified a previously unknown high-impact
198 two bp insertion resulting in a frameshift variant in the beginning of the *LGS1* coding
199 region (position 69986146, allele frequency in WGS dataset = 8%). The frameshift was
200 linked to a SNP genotyped in the GBS dataset (T/A at position 69,985,710; D' = 0.93, r^2
201 = 0.84, Fig. 2B). All nine accessions with the frameshift in the WGS dataset also shared
202 a 315 bp deletion (positions 69,984,268-69,984,583) overlapping 143 bp of the 3'
203 untranslated region in the 1580 bp second exon of *LGS1*.
204 Although our approach is not well-powered to detect heterozygous deletions or
205 small loss-of-function variants unlinked to SNPs in the GBS dataset, we find evidence
206 that at least three of six independent *LGS1* loss-of-function alleles characterized here
207 and elsewhere (28) are present at low to intermediate frequency. Among accessions

208 with SNP calls in *LGS1*, 7.0% of accessions exhibited SNP calls consistent with
209 homozygous *lgs1-2* and *lgs1-3* in the GBS dataset (Supplemental Data File S1).
210 Homozygous deletion frequency in African landraces for the GBS dataset was 7.5%
211 compared to 5.3% in non-African landraces. The SNP tagging the frameshift is present
212 at an allele frequency of 15% in the GBS dataset and is found in 16% of African
213 landraces but only 7% of non-African landraces. *LGS1* loss-of-function alleles were
214 found in diverse races and regions for both datasets (Fig. S4, Table S2, Supplemental
215 Data File S1), suggesting that these mutations have had time to spread and that their
216 benefit is not strongly masked by epistasis. Most landrace accessions with *LGS1*
217 deletions and botanical race assignments from the GBS dataset were guinea (12
218 accessions), caudatum (8 accessions), or durra-caudatum hybrids (17 accessions,
219 Table S2, Supplemental Data File S1).
220 *LGS1* loss-of-function alleles were significantly more common among landraces
221 with high parasite HS scores (Fig. 2C-D). However, correlations with population
222 structure reduced power to detect associations with these resistance alleles after
223 accounting for kinship (Fig. S5). The median *S. hermonthica* HS score was 0.20 for
224 accessions homozygous for *lgs1-2,* 0.54 for accessions homozygous for *lgs1-3,* 0.25 for
225 accession with the frameshift, or 0.09 for accessions without evidence for *LGS1* loss-of-
226 function. The difference in *S. hermonthica* HS score between *lgs1-3* and *LGS1* intact
227 accessions was statistically significant before accounting for relatedness ($p < 0.001$,
228 Wilcoxon rank sum test vs. $p = 0.10$, MLM). Frameshift associations with HS were also
229 statistically significant prior to correction for relatedness ($p < 0.001$, Wilcoxon rank sum;
230 $p = 0.69$, MLM) and were stronger with the sorghum-only model compared to the all-
231 occurrence model ($p = 0.33$, MLM). We observed modest support for associations
232 between *lgs1-2* and *S. hermonthica* HS before correcting for relatedness ($p = 0.06$,
233 Wilcoxon rank sum test; $p = 0.53$, MLM), and stronger associations with the sorghum-
234 only model ($p = 0.26$, MLM).
235
236 *Upregulation of SL biosynthesis pathway genes in* an LGS1-*deficient line*
237 Changes in strigolactones (and indeed any hormone) are potentially pleiotropic
238 given the many known downstream functions of SLs, suggesting tradeoffs may be

239 associated with *LGS1* variation. We found multiple SL biosynthesis genes were up-
240 regulated in nutrient-stressed roots of a resistant sorghum line (SRN39), which
241 produces high levels of orobanchol, compared to a susceptible sorghum line (Shanqui
242 Red) that produces mainly 5-deoxystrigol (28), a dominant SL in sorghum root exudates
243 that strongly stimulates *S. hermonthica* germination (29). We first confirmed differences
244 in expression for genes located in the *LGS1* region deleted in SRN39. Two gene
245 models in the deletion region were differentially expressed with low expression in
246 Shanqui Red and no reads detected in SRN39 (*Sobic.005G213500*, $p = 0.001$;
247 *Sobic.005G213700*, $p = 0.008$). The two other gene models in the deleted region,
248 *Sobic.005213766* and *Sobic.005213832*, had no reads detected in either line in roots
249 under nutrient stress. *LGS1* was differentially expressed between lines, with low
250 expression in the susceptible line and no expression in SRN39 ($p = 0.03$).
251 In *Arabidopsis* and rice, the strigolactone biosynthesis pathway undergoes
252 negative feedback regulation, in which low strigolactone levels stimulate upregulation of
253 several biosynthesis genes (31, 32). We asked if low endogenous 5-deoxystrigol might
254 also influence SL biosynthesis gene expression in high-orobanchol producing sorghum
255 genotypes. Four of eleven expressed SL biosynthesis genes had higher transcript
256 levels in the *LGS1*-deficient sorghum line SRN39 compared to the susceptible line (Fig.
257 S6, Table S3). These included two genes with homology to *NSP1* (*Sobic.001G341400*,
258 $p < 0.001$ and *Sobic.002G372100*, $p = 0.03$), *CCD7* (*Sobic.006G170300*, $p < 0.001$),
259 and *LBO* (*Sobic.003G418000*, $p = 0.01$). Differences between lines in expression of
260 seven other SL biosynthesis genes were not statistically significant. Among 17 genes
261 with probable roles in SL perception and downstream signaling, *SMXL3*
262 (*Sobic.008G042800*) was differentially expressed between lines and was
263 downregulated in the *LGS1*-deficient line SRN39 ($p = 0.03$). This change in *SMXL3*
264 suggests possible downstream pleiotropy given that regulation of SL-dependent
265 processes such as shoot development and root growth depends on the ubiquitination
266 and degradation of SMXL repressor proteins (33, 34).
267 Overall, 2937 transcripts were differentially expressed in roots of the two lines
268 under nutrient stress. The majority of transcripts were more highly expressed in *LGS1*-
269 deficient line SRN39 compared to *Striga*-susceptible Shanqui Red (1290 downregulated

270 vs. 1647 upregulated, Supplemental Data File S2). No GO-term was overrepresented
271 among differentially expressed genes, but genes involved in response to a biotic
272 stimulus exhibited the highest enrichment score (GO: 0009607; uncorrected p = 0.002;
273 FDR corrected p = 1).
274
275 *Parasite-associated SNPs across the sorghum genome*
276 We performed genome-wide tests of association with predicted parasite HS
277 score. A scan of 317,294 SNPs across 2070 sorghum landraces revealed 97 loci
278 exhibiting significant associations with *S. hermonthica* distribution at a false discovery
279 rate of 5% (Fig. 3A, Table S4). Of SNPs exceeding the threshold for significance, 45
280 were present within 1 kb of a predicted gene model, and four were predicted to cause
281 an amino acid change (Table S4). Three outlier SNPs were in QTL previously
282 associated with *Striga* resistance (35) including one intron variant in the uncharacterized
283 gene model *Sobic.001G227800* (Table S4). Another outlier SNP was present in an
284 intron of *Sobic.007G090900*, a gene model with high homology to *SMAX1/D53*, which is
285 degraded in an SL-dependent manner to control downstream SL signaling and is
286 associated with tillering and height in rice (36). SNPs among those with the strongest
287 associations to parasite occurrence were also found in genes related to suberin and
288 wax ester biosynthesis (*Sobic.007G091200*) including phenylalanine ammonia-lyase
289 (*Sobic.006G148800*). Phenylalanine ammonia-lyase is highly upregulated in the
290 resistant rice line Nipponbare compared to a susceptible line during infection with *S.*
291 *hermonthica* (37) and is associated with increased lignin deposition and post-
292 attachment resistance (38).
293 Across all gene models tagged by SNPs in the genome-wide analysis, no GO
294 term met the threshold for significance after correction for multiple comparisons. The
295 strongest enrichment scores were in genes with GO terms related to cell wall
296 organization (GO:0071555; corrected p = 0.13, mean p score for 48 genes = 0.27), cell
297 wall (GO:0005618, corrected p = 0.17, mean p score for 70 genes = 0.29), and
298 pectinesterase activity (GO:0030599, corrected p = 0.22, mean p score for 39 genes =
299 0.26). The strongest SNP association to parasite occurrence in a pectinesterase gene
300 model was in *Sobic.002G138400* (SNP S2_21521798, Table S4). The allele associated

301 with parasite occurrence at S2_21521798 is not predicted to cause an amino acid
302 change, but was in strong linkage disequilibrium (r^2 >0.9) with SNPs up to 204.4 kb
303 upstream of the gene model, which encompasses a region of likely substantial structural
304 variation based on visual inspection of read alignments to the reference. Overall, SNPs
305 in genes related to SL biosynthesis and signaling (Table S3) showed a non-significant
306 enrichment for associations with *S. hermonthica* distribution (uncorrected p = 0.09).
307
308 *Signatures of selection in candidate regions*
309 We further investigated three candidate genes with polymorphism that exhibited
310 distinct geographic patterns and had known or potential roles in *S. hermonthica*
311 resistance. Elevated Tajima's D values can indicate an excess of shared polymorphism
312 across SNPs at a locus, expected for regions of the genome under balancing selection,
313 whereas strongly negative values can indicate an excess of low frequency
314 polymorphism, expected under either purifying or positive selection. Two 5 kb genomic
315 regions, spanning SNPs in *LGS1* (SNP S5_69986146 in gene model
316 *Sobic.G005G213600*) and a pectinesterase gene (SNP S2_21521798 in gene model
317 *Sobic.002G138400,* MAF=0.275) exhibited elevated values of Tajima's D compared to
318 1000 randomly sampled 5 kb windows containing or overlapping gene models (Fig. 4; p
319 = 0.06, *Sobic.G005G213600*; p = 0.02, *Sobic.002G138400*). Regions of elevated
320 Tajima's D were localized to relatively small windows centered on SNPs associated with
321 *Striga* habitat suitability, and larger window sizes produced weaker signals (data not
322 shown). We looked for evidence of trans-species polymorphism and found no reads
323 mapping to the pectinesterase gene and no evidence for the *LGS1* loss-of-function
324 alleles characterized in our study in previously sequenced accessions of *S. propinquum*
325 (Kunth) Hitchc. (n=2) and *S. arundinaceum* (as synonym *S. bicolor* subsp. *verticilliflorum*
326 (Steud.) de Wet ex Wiersema & J.Dahlb.) (n=2)(39).
327 We did not observe strong departures from the neutral expectation for the region
328 surrounding a gene with homology to *SMAX1* (gene model *Sobic.007G090900* tagged
329 by SNP S7_14459084, p=0.6). The minor allele at S7_14459084 was at low frequency
330 in the GBS dataset (MAF = 0.014) and most common in West Africa (Fig. 3C), which is
331 not well sampled in the WGS dataset. The signal of association with *S. hermonthica*

332 occurrence extended more than 7.5 Mb on Chromosome 7 (Fig. 3A), but we did not
333 observe evidence to suggest that S7_14459084 tags an incomplete or soft sweep in
334 either the GBS or WGS datasets according to the haplotype-based statistic nS_L (40).

335 **DISCUSSION:**

336 Pests, pathogens, and parasites threaten human health and food security in a
337 changing world but understanding mechanisms of resistance across diverse taxa
338 remains challenging. Here, we evaluate the hypothesis that geographic selection
339 mosaics maintain genetic diversity in host resistance alleles across gradients of parasite
340 occurrence in smallholder farming systems. We extended genotype-environment
341 association analyses to biotic environmental gradients using species distribution models
342 to model high resolution variation in parasite occurrence at continent scales. We report
343 strong associations with parasite occurrence for novel candidate resistance loci in the
344 sorghum-*Striga hermonthica* pathosystem and characterize diverse loss-of-function
345 mutations in the known sorghum gene *LGS1*. Geographic distribution of loss-of-
346 function alleles suggests that *LGS1*-conferred resistance is stable across environments
347 and genetic backgrounds. However, the intermediate frequency and paucity of *LGS1*
348 loss-of-function alleles outside of parasite-prone areas, combined with *LGS1*-associated
349 changes in sorghum strigolactone biosynthesis and endogenous signaling, suggest
350 there may be trade-offs associated with *LGS1* loss-of-function.

351 Our results support the hypothesis that spatial variation in selective pressures
352 controls geographic clines in the frequency of host resistance alleles. The patterns we
353 characterized are likely representative of long-term averaged conditions as opposed to
354 a snapshot of coevolution because our parasite distribution model used occurrence
355 records spanning more than 150 years, and the landraces we studied were collected
356 across the last several decades. Negative frequency-dependent selection and rapid
357 coevolutionary dynamics, for example, could decouple parasite abundance and the
358 frequency of host resistance alleles. Our approach may be helpful for identifying host
359 alleles whose resistance phenotype is conditional on local parasite genotypes (e.g. as
360 might be suggested by patterns of distribution for polymorphism in *SMAX1*, Fig. 3D).
361 Many parasites including *S. hermonthica* are highly genetically diverse, so that host

362 resistance phenotype differs depending on local parasite genotypes (41). This host by
363 parasite genotype interaction might obscure GWAS for host resistance depending on
364 the parasite genotype used (42). Combining knowledge of parasite population structure
365 with the spatial perspective of parasite-associated host genomic variation presented
366 here could facilitate complementary, inexpensive detection of genomic regions
367 contributing to resistance across diverse biotic environments.

368 Our study revealed evidence for locally-adaptive natural variation in genes
369 related to cell wall modification. Cell-wall modifying enzymes including pectinesterases
370 are highly expressed in the developing haustorium of parasitic plants and in the host-
371 parasite interface (43–46). Pectinesterases de-esterify pectin in plant cell walls, making
372 it accessible to other cell-wall degrading enzymes and loosening cell walls. However,
373 some studies have suggested that in the presence of Ca^{2+}, de-esterified pectin forms
374 egg-box structures and instead stiffens cell walls (47). Rigidification of sorghum cell
375 walls by their own pectinesterases (such as *Sobic.002G138400*, Table S4) or reduced
376 activity could help defend against parasitic invaders. Notably, Yang et al. (2015)
377 reported haustorium specific expression and positive selection on pectinesterase
378 inhibitors in parasitic plant lineages and a pectinesterase inhibitor showed exceptionally
379 high host species-associated expression in field populations of *S. hermonthica* (Lopez-
380 Perez et al., *in press*). Parasite pectinesterase inhibitors could interact with host
381 pectinesterases or help maintain integrity of parasite cell walls in the face of high
382 pectinesterase expression during haustorial invasion. The SNP in a pectinesterase
383 gene most strongly associated with *S. hermonthica* distribution does not code for an
384 amino acid change but is linked with other SNPs across a region spanning substantial
385 structural variation in the WGS dataset, potentially influencing sorghum pectinesterase
386 expression levels during haustorial invasion.

387 Our results also suggest that *LGS1* loss-of-function alleles may be adaptive in *S.*
388 *hermonthica*-prone regions, but that costs of resistance may limit their distribution.
389 Loss-of-function alleles are relatively uncommon but higher in frequency and broadly
390 distributed where parasites occur (Fig. 2C). Associations for the known locus *LGS1*
391 were not statistically significant after correcting for kinship, likely due to covariation with
392 genomic background (Fig. S4), which has been shown to substantially reduce power to

393 detect causal loci in locally-adapted sorghum landraces (14) and in simulations (48).
394 The diversity of loss-of-function variants reported here (Fig. 2A-B) and elsewhere (28),
395 their wide geographic distribution (Fig. 2C), and an excess of high frequency
396 polymorphism localized to *LGS1* (Fig. 4) are consistent with long-term maintenance of
397 *LGS1* diversity under balancing selection. The underlying evolutionary processes
398 remain unknown but could include negative frequency dependent selection or
399 spatiotemporally variable selection favoring different alleles in different environments,
400 depending on the relative costs of resistance (4, 49). Costs of resistance linked to SL
401 structural changes could include the ability to interact with AM fungi, increased
402 susceptibility to other *S. hermonthica* ecotypes more sensitive to orobanchol (e.g. *S.*
403 *hermonthica* that parasitize rice, an orobanchol-exuder), or impacts on endogenous
404 strigolactone signaling. Consistent with the last hypothesis, we find increased
405 expression of SL biosynthesis genes in a sorghum line with low levels of 5-deoxystrigol
406 (5DS), one of the primary SLs produced by sorghum (29, 50). This result could suggest
407 transcriptional feedback in response to deficits in endogenous 5DS in sorghum, as
408 previously observed in *Arabidopsis* and rice in response to low overall SL levels (31,
409 32). Feedback regulation could be mediated by repressors of SL signaling such as
410 SMXL3, a homolog of which was downregulated in the 5DS-deficient line (Table S3).
411 Costs of resistance may also differ among particular loss-of-function alleles, depending
412 on biological pathways influenced by other genes in deleted regions.
413 Another factor potentially contributing to natural variation in SL biosynthesis and
414 signaling is the role of human selection against *S. hermonthica* resistant varieties during
415 domestication and diversification of cultivated sorghum. While impacts of SL structural
416 variation on endogenous function remain poorly known, natural variation in levels of
417 strigolactone production correlates with increased tillering in rice (51, 52), a trait subject
418 to strong selection over the course of cereal domestication (53). One of the most well-
419 studied domestication genes, *Teosinte branched 1 (Tb1)*, controls branching and was a
420 major target of selection during maize domestication (54, 55). *Tb1* activity is decoupled
421 from strigolactone signaling in maize, enabling strigolactone-independent branching
422 regulation (56). However, in pearl millet, signatures of selection on *Tb1* orthologs are
423 weaker (57), consistent with observations that *Tb1* orthologs do not act independently of

424 strigolactones in other cereals (58). There is also evidence that *Tb1* regulates the
425 maize domestication gene *Teosinte glume architecture 1* (59), a homolog of which also
426 exhibited parasite-associated genetic variation in our study (Table S4).
427 Taken together, this study provides evidence of locally-adaptive natural variation
428 in sorghum parasite resistance genes across African smallholder farming systems. We
429 report long-term maintenance of diversity in known and novel candidates implicated in
430 pre- and post-attachment resistance to the parasitic plant *Striga hermonthica.* However,
431 the possibility of *LGS1*-driven tradeoffs or the existence of orobanchol-sensitive *S.*
432 *hermonthica* populations (for example, from rice) suggest potential pitfalls with
433 widespread deployment of the *LGS1* loss-of function allele in sorghum cultivation. Our
434 findings highlight the complexity of interacting abiotic, biotic, and human pressures
435 shaping genome polymorphism across environments in cultivated species.

436 **METHODS:**

437 *Species distribution models*

438 Genome-environment association approaches identify putatively locally adaptive
439 genetic loci where allelic variation is strongly associated with home environments (60).
440 To employ this approach with biotic gradients, we required information on local parasite
441 pressure for each sorghum landrace. We used species distribution models (SDMs) to
442 estimate habitat suitability of *Striga hermonthica* at the location of each georeferenced
443 sorghum landrace, under the assumption that modeled habitat suitability scores are a
444 reasonable proxy of parasite success averaged over the long term and in comparison
445 with sites where the parasite never occurs.
446 *S. hermonthica* SDMs were constructed with Maxent, a machine learning tool for
447 predicting habitat suitability for a species of interest given a set of environmental
448 variables and presence-only data (30). We compiled 1369 occurrence records for *S.*
449 *hermonthica* records downloaded from the Global Biodiversity Information Facility
450 (www.gbif.org), newly digitized records for specimens housed in the collections of the
451 Royal Botanic Gardens Kew, the National Museum of Natural History in Paris, the
452 French Agricultural Research Centre for International Development, the University of
453 Montpellier, and the Botanical Garden of Lyon, and observations from published studies

454 (61–76) (Supplemental Data File S3). Records within 0.01 degree (~1 km) of another
455 observation were excluded to reduce sampling bias. To characterize the background
456 environment across the study extent, we sampled 10,000 points at random from a 500
457 km radius surrounding locations of 1,050 occurrences in the final dataset. We also
458 created 'sorghum-only' models based on a subset of *S. hermonthica* records (*n* = 262)
459 that were annotated as occurring specifically on sorghum.
460 Environmental variables were chosen based on prior knowledge of the ecology of
461 *S. hermonthica* (61). Bioclimatic and topographic variables (annual rainfall, mean
462 temperature of the wettest quarter, isothermality, potential evapotranspiration [PET],
463 and topographic wetness index) were obtained from CHELSA (77) and ENVIREM
464 datasets (78). Soil variables (clay content, nitrogen, and phosphorus) were based on
465 continental and global-scale soil property maps (79, 80). We explored additional
466 ecologically relevant variables but did not include them in the final model due to high
467 correlation across the study background with annual rainfall (correlated with soil PH,
468 aluminum, and precipitation seasonality) or soil clay content (correlated with sand
469 fraction) as indicated by Pearson coefficients (|r| >0.7).
470 SDMs were implemented and evaluated with ENMeval, using the
471 'checkerboard2' method for data partitioning, which is designed to reduce spatial
472 autocorrelation between testing and training records (81). The distribution model with
473 the lowest ΔAICc was selected for further comparisons with genome variation in
474 sorghum. Two niche overlap statistics, Schoener's *D* (Schoener 1968) and the *I*
475 similarity statistic (83) were calculated for the all-occurrence and sorghum-only models
476 using the R package *dismo* (84).
477
478 *Sorghum* LGS1 *loss-of-function alleles*
479 Fine-scale natural variation in sorghum *LGS1* was characterized using whole
480 genome sequencing (WGS) data from a set of 143 georeferenced landraces from the
481 sorghum bioenergy association panel (BAP) (85). The BAP includes both sweet and
482 biomass sorghum types, accessions from the five major sorghum botanical races
483 (durra, caudatum, bicolor, guinea, and kafir), and accessions from Africa, Asia, and the
484 Americas. The BAP accessions were sequenced to approximately 25x coverage and

genotyped as part of the TERRA-REF project (www.terraref.org). This whole genome sequencing derived dataset is referred to throughout the manuscript as the 'WGS dataset' to distinguish it from the GBS dataset used for the GEA.

We characterized three loss-of-function alleles in *LGS1* using data from the WGS dataset. Frameshift and nonsense mutations were identified using *SnpEff* v4.3t for SNP calls and small indels in *Sobic.005G213600* available from TERRA-REF (86). High impact variants were manually checked to remove those near the 3' end of the coding region and two linked variants that were present in the same lines and did not cause a frameshift when combined. To characterize large deletion variants, we aligned quality trimmed reads to the *Sorghum bicolor* v3.1 reference genome (DOE-JGI, http://phytozome.jgi.doe.gov/) with BWA MEM v0.7.17 (87). Duplicates were removed with SAMBLASTER v0.1.24 (88) and structural variants were called for each landrace with LUMPY v0.2.13 (89). SVTYPER v0.6.0 was used to call genotypes for structural variants ≤ 1 Mb spanning *Sobic.005G213600* (90).

Following characterization of *LGS1* deletion breakpoints using the WGS dataset, we imputed deletion calls to the GBS dataset. We considered the *LGS1* region to be deleted if at least one SNP was called in the 5 kb region flanking positions of deletion breakpoints, but all data were missing between breakpoints. We considered the LGS1 region to be present if at least one SNP was called within the *Sobic.005G213600* gene model. Fifteen low-coverage samples, with missing data extending 5 kb into flanking regions, were excluded.

Experimental validation of LGS1 *deletion and frameshift alleles*

LGS1 loss-of-function alleles were validated by testing 12 accessions from the Sorghum Association Panel (SAP) (91) for their ability to stimulate *S. hermonthica* germination (Table S5). Three accessions were previously reported as resistant to *S. hermonthica*, two accessions were susceptible, and seven accessions had unknown resistance (28, 92). Root exudates were harvested 43 days after planting (see Supplemental Materials & Methods for a detailed description of plant growth conditions). Germination trials were conducted using seed of *Striga hermonthica* collected on sorghum by Steven Runo (Kenyatta U) in western Kenya and assayed in the USDA

516 quarantine lab at UVa. Seeds were surface sterilized with 0.5% bleach, rinsed three
517 times with sterile distilled water, and 75 seeds were transferred with 500uL diH_2O to 12-
518 well microtiter plates. After 10 days of preconditioning in the dark at 30°C, 2.25 mL of
519 fresh root exudate was applied to each well. Exudate from five biological replicates
520 (sorghum individuals) per sorghum genotype, with three technical replicates (wells) per
521 biological replicate, were tested. GR24 at two concentrations, 1 ppm and 0.1 ppm, and
522 water were used as positive and negative controls. Germinated seeds were counted
523 under stereomicroscope after 66 hours incubation. We tested significance of deletion
524 alleles using linear mixed effects models in the R package lme4 (93), where deletion
525 was a fixed effect and genotypes were random effects.
526
527 *LGS1 and expression of genes in strigolactone synthesis and signaling pathways*
528 To assess the impact of putatively locally adaptive variation at *LGS1*, we studied
529 root transcriptomes of two lines segregating for presence/absence. We generated
530 TagSeq libraries using root tissue from five replicate individuals each of sorghum lines
531 Shanqui Red (PI 656025) and SRN39 (PI 656027), grown under nutrient deficient
532 conditions. Shanqui Red is susceptible to *S. hermonthica*, whereas SRN39 is resistant
533 as a result of a ~34 kb deletion spanning *LGS1* and four adjacent genes (28). Five
534 replicates of each line were grown under identical conditions as for root exudate
535 collection. Twenty-five days after planting, all potting components were carefully
536 washed from roots, and 2-5 ml of the fine root tissue was sampled, placed into labeled
537 vials and stored on ice for 1 day during shipment to U of Texas, where samples were
538 stored at -80°C until RNA extraction.
539 RNA extractions were performed with TRIzol Reagent after grinding root tissues
540 in liquid nitrogen using mortar and pestle. We used the 3'-TagSeq approach (94) with
541 several modifications to construct cDNA libraries. This method focuses on the 3' end of
542 transcripts enriched in a size range of 400-500 bp fragments.
543 cDNA libraries were sequenced on a single lane of an Illumina Hiseq-2500 analyzer at
544 the Genomic Sequencing and Analysis Facility at the University of Texas at Austin. We
545 recovered between 3 to 5 million raw single-end 150 bp reads per sample and
546 compared expression differences between SRN39 and Shanqui Red using the TagSeq

547 v2.0 pipeline (https://github.com/Eli-Meyer/TagSeq_utilities/). A detailed description of
548 the library preparation method and bioinformatic analysis is given in Supplemental
549 Materials & Methods.
550 Count data were analyzed in DESeq2 with FDR correction (α=0.05) (95).
551 Enrichment analysis was performed as for genome-wide association analysis, except
552 only the set of gene models with non-zero expression were used as the background.
553
554 *Genome-environment associations*
555 We performed a genome-wide scan for SNPs in the sorghum genome strongly
556 associated with values of habitat suitability estimated by our *S. hermonthica* distribution
557 model. Sorghum genotypic information was extracted from a public dataset of
558 accessions genotyped using genotyping-by-sequencing (GBS) (14, 96–98). This
559 dataset comprises a diverse set of worldwide accessions including germplasm from the
560 SAP (91), the Mini-Core Collection (99), and the Bioenergy Association Panel (BAP)
561 (85). Beagle 4.1 was used to impute missing data based on the Li and Stephens (2003)
562 haplotype frequency model (100). The average missing rate in the non-imputed dataset
563 is 0.39 (98). After excluding sorghum accessions with missing coordinates and SNPs
564 with minor allele frequency less than 0.01, the dataset, hereafter referred to as the 'GBS
565 dataset', included 1547 African landraces among 2070 georeferenced accessions total
566 genotyped at 317,294 SNPs. At each location of a georeferenced accession in the GBS
567 dataset, we extracted logistic output from the *S. hermonthica* distribution model
568 (Supplemental Data File S4-5) as the 'phenotype'. To account for regions where
569 predicted habitat suitability is high but *S. hermonthica* has not been recorded, we
570 cropped model predictions to within 200 km from any occurrence record and set values
571 outside of this range to zero to derive for each grid cell an *S. hermonthica* occurrence
572 score ranging from zero to one; more than half of sorghum accessions are from
573 locations with parasite HS scores greater than zero (Fig. 1B). Genome-wide
574 associations for each SNP with *S. hermonthica* occurrence were computed for the GBS
575 dataset using a mixed linear model (MLM) fit with GEMMA v0.94 (101). To take into
576 account relatedness among individuals, we used a centered kinship matrix (-gk 1)
577 generated from all 317,294 SNPs before calculation of p score statistics (-lmm 3). P

578 scores were adjusted for multiple comparisons using the Benjamini and Hochberg
579 (1995) procedure (FDR = 0.05). To visualize genomic regions previously implicated in
580 resistance to *S. hermonthica,* locations of QTL from Haussman et al. (2004) in the *S.*
581 *bicolor* v3.0 genome were downloaded from the Sorghum QTL Atlas
582 (http://aussorgm.org.au) (102). We tested for associations with *LGS1* loss-of-function
583 mutations using the same procedure and kinship matrix as for the genome-wide
584 association analysis.
585 We identified gene functions enriched for associations with parasite distribution
586 using the gene score resampling method in ErmineJ (103). This method places higher
587 value on gene scores than their relative ranking and does not require choice of a
588 threshold for significance. For each gene model, we used the lowest p score from
589 GEMMA of any SNP within 1 kb of gene model boundaries, and enrichment analyses
590 were performed using the mean of all gene scores in the gene set. Gene sets were
591 created using GO terms for all gene models in the *Sorghum bicolor* v3.0 genome
592 (annotation version 3.1). We also created a custom gene set comprising 30 gene
593 models implicated in strigolactone biosynthesis and signaling (Table S3). This gene set
594 included nine gene models annotated as belonging to Phytozome pathway Sbicolor
595 PWY-7101 (*D27*, *CCD7*, and *CCD8* homologs), and all sorghum gene models
596 annotated with best BLASTP hits to Arabidopsis *MAX1*, *LBO* (104), *AtD14, SMAX1,*
597 *SMXL3/4/5/6/7, NSP1,* and *NSP2* (105). Enrichment analysis was performed with
598 200,000 iterations, excluding gene sets with less than 5 or more than 200 genes.
599
600 *Signatures of selection in candidate genomic regions*
601 We predicted functional impacts for candidate SNPs of interest and performed
602 scans for selection in 1 Mb regions surrounding each focal SNP. Linkage disequilibrium
603 between sites was determined with vcftools v0.1.15 (--geno-r2 parameter). To identify
604 regions under balancing selection among a subset of African landraces, we used
605 Tajima's D calculated with vcftools in non-overlapping 5 kb windows, excluding SNPs
606 with more than 70% missing data. *P*-values for candidate regions under selection were
607 calculated based on the empirical distribution of Tajima's D for 1000 randomly sampled

608 5 kb windows that overlapped or fully encompassed gene models. We searched for
609 sweeps using the nS_L statistic with selscan v1.2.0a (106).
610
611 **Acknowledgments**
612 We thank the many collectors, volunteers, and herbarium curators who made this work
613 possible and are particularly grateful to Marie-Hélène Weech and the staff of the Royal
614 Botanic Gardens Kew and the Muséum national d'Histoire naturelle. We thank Steven
615 Runo for providing *S. hermonthica* seeds and Alice MacQueen for comments that
616 improved the manuscript. Whole genome sequence data used here is from the TERRA
617 REF experiment, funded by the Advanced Research Projects Agency-Energy (ARPA-
618 E), U.S. Department of Energy, under Award Number DE-AR0000594. This material is
619 based on work supported by a National Science Foundation Postdoctoral Research
620 Fellowship in Biology to ESB under Grant No. 1711950. The views and opinions of
621 authors expressed herein do not necessarily state or reflect those of the United States
622 Government or any agency thereof.
623

FIGURES:

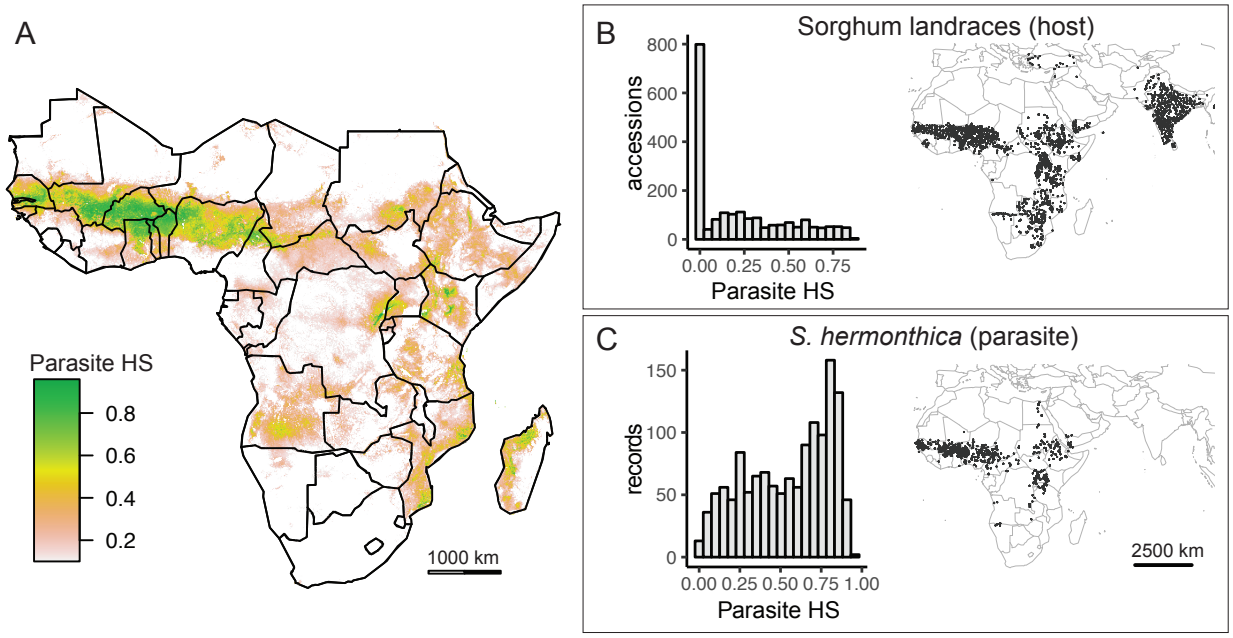


Fig. 1: A subset of global sorghum landraces are from parasite-prone areas. A) *Striga hermonthica* habitat suitability (HS) scores across Africa based on MaxEnt species distribution model. B) Geographic distribution and frequency histogram of HS scores at locations of 2070 georeferenced and genotyped sorghum landrace accessions. To account for areas with high HS where parasites have not been recorded, HS for accessions more than 200 km from any *S. hermonthica* record was set to zero. C) Geographic distribution and frequency histogram of HS scores at locations of 1369 *S. hermonthica* occurrence records.

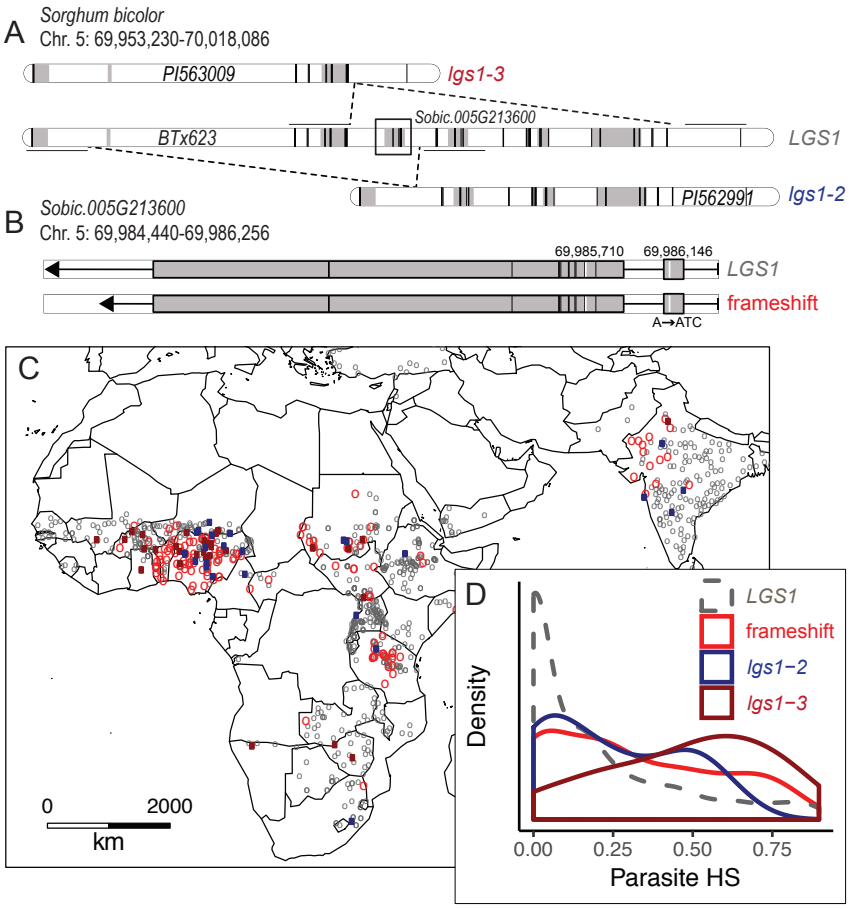


Fig. 2. *LGS1* loss-of-function alleles are broadly distributed within parasite-prone regions. Schematic of large deletion variants (A) and frameshift mutation (B) impacting sorghum *LGS1*, a known locus involved in resistance to *Striga* spp. Grey shading indicates position of gene models (A) or coding regions (B). Position of SNPs in the GBS dataset is indicated by vertical black bars, and horizontal black lines denote 5 kb flanking regions used to impute deletion calls. In B, vertical white bars show the frameshift mutation (position 69,986,146) and the SNP at position 69,985,710 that tags the frameshift in the GBS dataset. C) Geographic distribution of *LGS1* alleles in sorghum landraces. D) Distribution of parasite habitat suitability (HS) scores at locations of sorghum accessions with *lgs1-2* (*n*=25), *lgs1-3* (*n*=34), frameshift (*n*=131), or intact *LGS1* (*n*=785).

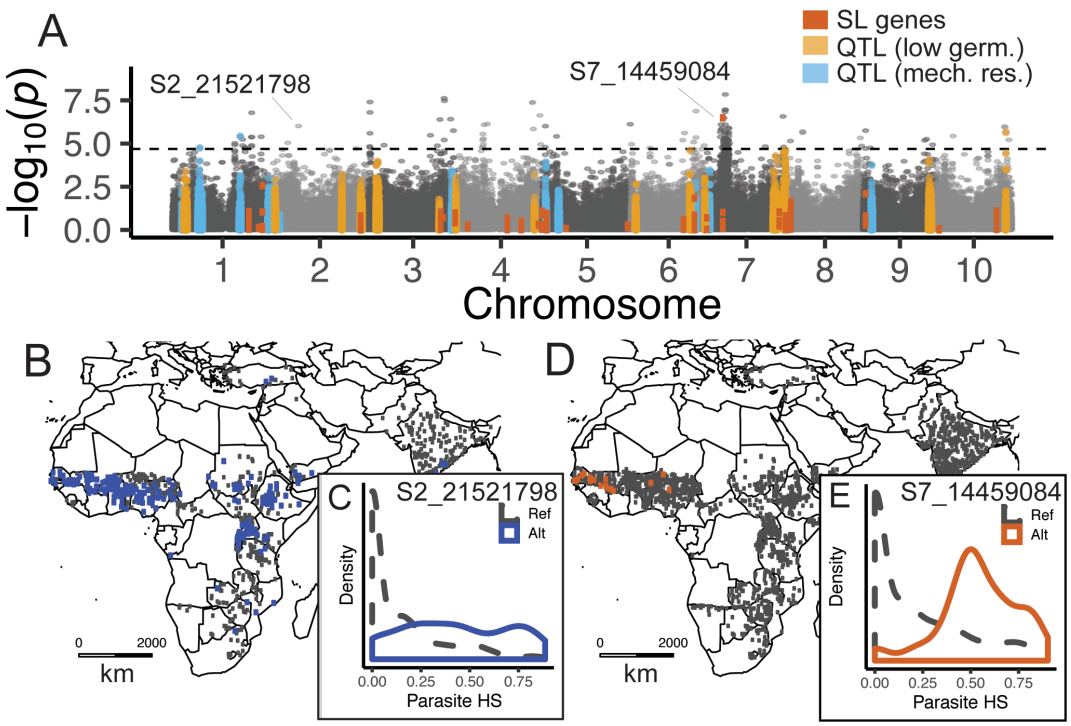
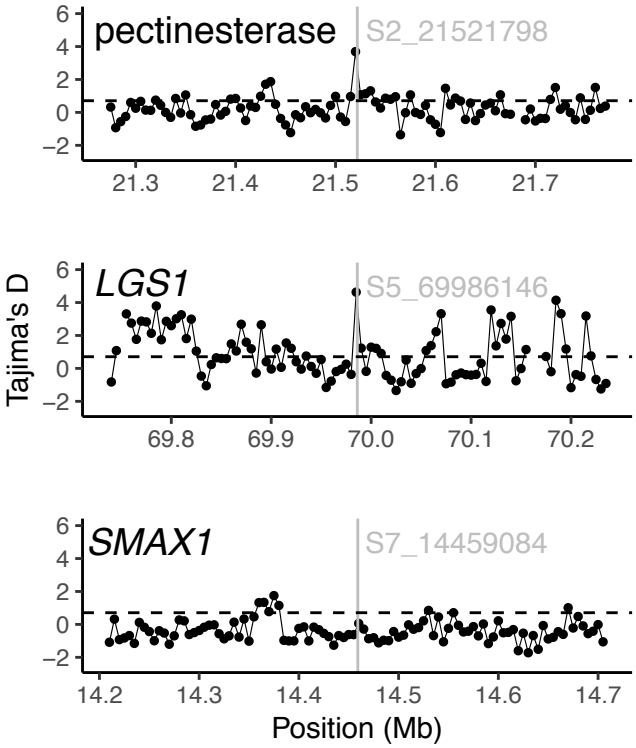


Fig. 3: Sorghum genome-wide associations with parasite distribution implicate cell-wall and strigolactone signaling genes. A) Genome-wide association with parasite habitat suitability (HS) score, based on 317,294 SNPs with minor allele frequency (MAF) > 0.01 in 2070 sorghum landraces. SNPs in genomic regions linked to strigolactone (SL) biosynthesis/signaling (red), mechanical resistance (light blue), or low *Striga hermonthica* germination (orange) are indicated. The dashed line represents significance threshold at a false discovery rate of 5%. B) Geographic distribution of reference and alternate alleles for a SNP (S2_21521798) in a pectinesterase gene (MAF = 0.275). C) Distribution of parasite habitat suitability scores for sorghum accessions segregating for S2_21521798. D) Geographic distribution of reference and alternate alleles for a SNP (S7_14459084) in a gene homologous to SMAX1 (MAF = 0.014). E) Distribution of parasite HS scores for sorghum accessions segregating for S7_14459084.



664

665 **Fig. 4. Signatures of balancing selection surround parasite-associated SNPs in a**
666 **pectinesterase gene (S2_21521798) and *LGS1* (S5_69986146).** The region
667 surrounding a SNP in a gene homologous to *SMAX1* (S7_14459084, lower panel) is
668 also shown. Tajima's D was calculated in non-overlapping windows of 5 kb using data
669 from the WGS dataset subset to 143 African landraces. The dashed horizontal line
670 indicates the median value for 1000 randomly selected 5 kb windows overlapping or
671 encompassing gene models. Vertical grey lines indicate positions of the CDS location
672 for the gene containing the focal SNP.

673 **REFERENCES:**

674 1. Haldane JBS (1949) Disease and Evolution. *Ricerca Scientifica* 19:68–76.

675 2. Stahl EA, Dwyer G, Mauricio R, Kreitman M, Bergelson J (1999) Dynamics of disease
676 resistance polymorphism at the Rpm1 locus of Arabidopsis. *Nature* 400(6745):667–671.

677 3. Tack AJM, Thrall PH, Barrett LG, Burdon JJ, Laine A-L (2012) Variation in infectivity and
678 aggressiveness in space and time in wild host-pathogen systems: causes and
679 consequences. *Journal of Evolutionary Biology* 25(10):1918–1936.

680 4. Karasov TL, Horton MW, Bergelson J (2014) Genomic variability as a driver of plant–
681 pathogen coevolution? *Current Opinion in Plant Biology* 18:24–30.

682 5. Brockhurst MA, et al. (2014) Running with the Red Queen: the role of biotic conflicts in
683 evolution. *Proceedings of the Royal Society B: Biological Sciences* 281(1797):20141382.

684 6. Bergelson J, Dwyer G, Emerson JJ (2001) Models and Data on Plant-Enemy Coevolution.
685 *Annual Review of Genetics* 35(1):469–499.

686 7. Agrawal A, Lively CM (2002) Infection genetics: gene-for-gene versus matching- alleles
687 models and all points in between. *Evolutionary Ecology Research* 4:79–90.

688 8. Möller M, Stukenbrock EH (2017) Evolution and genome architecture in fungal plant
689 pathogens. *Nature Reviews Microbiology* 15(12):756–771.

690 9. Samberg LH, Gerber JS, Ramankutty N, Herrero M, West PC (2016) Subnational
691 distribution of average farm size and smallholder contributions to global food production.
692 *Environ Res Lett* 11(12):124010.

693 10. Jensen HR, Dreiseitl A, Sadiki M, Schoen DJ (2012) The Red Queen and the seed bank:
694 pathogen resistance of *ex situ* and *in situ* conserved barley. *Evolutionary Applications*
695 5(4):353–367.

696 11. Jousimo J, et al. (2014) Ecological and evolutionary effects of fragmentation on infectious
697 disease dynamics. *Science* 344(6189):1289–1293.

698 12. Bartoli C, Roux F (2017) Genome-Wide Association Studies In Plant Pathosystems:
699 Toward an Ecological Genomics Approach. *Front Plant Sci* 8.
700 doi:10.3389/fpls.2017.00763.

701 13. Meyer RS, Purugganan MD (2013) Evolution of crop species: genetics of domestication
702 and diversification. *Nature Reviews Genetics* 14(12):840–852.

703 14. Lasky JR, et al. (2015) Genome-environment associations in sorghum landraces predict
704 adaptive traits. *Science Advances* 1(6):e1400218.

705 15. Romero Navarro JA, et al. (2017) A study of allelic diversity underlying flowering-time
706 adaptation in maize landraces. *Nature Genetics* 49(3):476–480.

16. Vajana E, et al. (2018) Combining Landscape Genomics and Ecological Modelling to Investigate Local Adaptation of Indigenous Ugandan Cattle to East Coast Fever. *Front Genet* 9. doi:10.3389/fgene.2018.00385.

17. Gaillard MDP, Glauser G, Robert CAM, Turlings TCJ (2018) Fine-tuning the “plant domestication-reduced defense” hypothesis: specialist vs generalist herbivores. *New Phytol* 217(1):355–366.

18. Winchell F, Stevens CJ, Murphy C, Champion L, Fuller DorianQ (2017) Evidence for Sorghum Domestication in Fourth Millennium BC Eastern Sudan: Spikelet Morphology from Ceramic Impressions of the Butana Group. *Current Anthropology* 58(5):673–683.

19. Gomez-Roldan V, et al. (2008) Strigolactone inhibition of shoot branching. *Nature* 455(7210):189–194.

20. Ruyter-Spira C, et al. (2011) Physiological Effects of the Synthetic Strigolactone Analog GR24 on Root System Architecture in Arabidopsis: Another Belowground Role for Strigolactones? *Plant Physiology* 155(2):721–734.

21. Ha CV, et al. (2014) Positive regulatory role of strigolactone in plant responses to drought and salt stress. *Proc Natl Acad Sci U S A* 111(2):851–856.

22. Yoneyama K, et al. (2007) Nitrogen deficiency as well as phosphorus deficiency in sorghum promotes the production and exudation of 5-deoxystrigol, the host recognition signal for arbuscular mycorrhizal fungi and root parasites. *Planta* 227(1):125–132.

23. Ruiz-Lozano JM, et al. (2016) Arbuscular mycorrhizal symbiosis induces strigolactone biosynthesis under drought and improves drought tolerance in lettuce and tomato. *Plant, Cell & Environment* 39(2):441–452.

24. Akiyama K, Matsuzaki K, Hayashi H (2005) Plant sesquiterpenes induce hyphal branching in arbuscular mycorrhizal fungi. *Nature* 435:824–827.

25. Spallek T, Mutuku M, Shirasu K (2013) The genus *Striga*: a witch profile. *Molecular Plant Pathology* 14(9):861–869.

26. Laine A-L (2004) Resistance variation within and among host populations in a plant–pathogen metapopulation: implications for regional pathogen dynamics. *Journal of Ecology* 92(6):990–1000.

27. Thompson JN (1999) Specific Hypotheses on the Geographic Mosaic of Coevolution. *The American Naturalist* 153(S5):S1–S14.

28. Gobena D, et al. (2017) Mutation in sorghum LOW GERMINATION STIMULANT 1 alters strigolactones and causes *Striga* resistance. *PNAS* 114(17):4471–4476.

29. Mohemed N, et al. (2018) Genetic variation in *Sorghum bicolor* strigolactones and their role in resistance against *Striga hermonthica*. *J Exp Bot* 69(9):2415–2430.

30. Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190(3–4):231–259.

744 31. Arite T, et al. (2007) DWARF10, an RMS1/MAX4/DAD1 ortholog, controls lateral bud
745 outgrowth in rice. *Plant J* 51(6):1019–1029.

746 32. Mashiguchi K, et al. (2009) Feedback-Regulation of Strigolactone Biosynthetic Genes and
747 Strigolactone-Regulated Genes in Arabidopsis. *Bioscience, Biotechnology, and*
748 *Biochemistry* 73(11):2460–2465.

749 33. Wallner E-S, et al. (2017) Strigolactone- and Karrikin-Independent SMXL Proteins Are
750 Central Regulators of Phloem Formation. *Current Biology* 27(8):1241–1247.

751 34. Wang L, et al. (2015) Strigolactone Signaling in Arabidopsis Regulates Shoot Development
752 by Targeting D53-Like SMXL Repressor Proteins for Ubiquitination and Degradation.
753 *Plant Cell* 27(11):3128–3142.

754 35. Haussmann BIG, et al. (2004) Genomic regions influencing resistance to the parasitic
755 weed *Striga hermonthica* in two recombinant inbred populations of sorghum. *Theor Appl*
756 *Genet* 109(5):1005–1016.

757 36. Jiang L, et al. (2013) DWARF 53 acts as a repressor of strigolactone signalling in rice.
758 *Nature* 504(7480):401–405.

759 37. Swarbrick PJ, et al. (2008) Global patterns of gene expression in rice cultivars undergoing
760 a susceptible or resistant interaction with the parasitic plant *Striga hermonthica*. *New*
761 *Phytol* 179(2):515–529.

762 38. Mutuku JM, et al. (2019) The structural integrity of lignin is crucial for resistance against
763 *Striga hermonthica* parasitism in rice. *Plant Physiology*:pp.01133.2018.

764 39. Mace ES, et al. (2013) Whole-genome sequencing reveals untapped genetic potential in
765 Africa's indigenous cereal crop sorghum. *Nature Communications* 4:2320.

766 40. Ferrer-Admetlla A, Liang M, Korneliussen T, Nielsen R (2014) On Detecting Incomplete
767 Soft or Hard Selective Sweeps Using Haplotype Structure. *Mol Biol Evol* 31(5):1275–
768 1291.

769 41. Soltis NE, et al. (2019) Interactions of Tomato and *Botrytis cinerea* Genetic Diversity:
770 Parsing the Contributions of Host Differentiation, Domestication, and Pathogen Variation.
771 *The Plant Cell* 31(2):502–519.

772 42. MacPherson A, Otto SP, Nuismer SL (2018) Keeping Pace with the Red Queen:
773 Identifying the Genetic Basis of Susceptibility to Infectious Disease. *Genetics* 208(2):779–
774 789.

775 43. Honaas LA, et al. (2013) Functional genomics of a generalist parasitic plant: Laser
776 microdissection of host-parasite interface reveals host-specific patterns of parasite gene
777 expression. *BMC Plant Biology* 13(1):9.

778 44. Yang Z, et al. (2015) Comparative Transcriptome Analyses Reveal Core Parasitism Genes
779 and Suggest Gene Duplication and Repurposing as Sources of Structural Novelty. *Mol*
780 *Biol Evol* 32(3):767–790.

781 45. Sun G, et al. (2018) Large-scale gene losses underlie the genome evolution of parasitic
782 plant *Cuscuta australis*. *Nature Communications* 9(1):2683.

783 46. Losner-Goshen D, Portnoy VH, Mayer AM, Joel DM (1998) Pectolytic Activity by the
784 Haustorium of the Parasitic Plant *Orobanche* L. (Orobanchaceae) in Host Roots. *Ann Bot*
785 81(2):319–326.

786 47. Hocq L, Pelloux J, Lefebvre V (2017) Connecting Homogalacturonan-Type Pectin
787 Remodeling to Acid Growth. *Trends in Plant Science* 22(1):20–29.

788 48. Forester BR, Lasky JR, Wagner HH, Urban DL (2018) Comparing methods for detecting
789 multilocus adaptation with multivariate genotype–environment associations. *Molecular*
790 *Ecology* 27(9):2215–2233.

791 49. Brisson D (2018) Negative Frequency-Dependent Selection Is Frequently Confounding.
792 *Front Ecol Evol* 6. doi:10.3389/fevo.2018.00010.

793 50. Awad AA, et al. (2006) Characterization of Strigolactones, Germination Stimulants for the
794 Root Parasitic Plants *Striga* and *Orobanche*, Produced by Maize, Millet and Sorghum.
795 *Plant Growth Regul* 48(3):221.

796 51. Jamil M, Charnikhova T, Houshyani B, van Ast A, Bouwmeester HJ (2012) Genetic
797 variation in strigolactone production and tillering in rice and its effect on *Striga*
798 *hermonthica* infection. *Planta* 235(3):473–484.

799 52. Cardoso C, et al. (2014) Natural variation of rice strigolactone biosynthesis is associated
800 with the deletion of two MAX1 orthologs. *PNAS* 111(6):2379–2384.

801 53. Doust A (2007) Architectural evolution and its implications for domestication in grasses.
802 *Ann Bot* 100(5):941–950.

803 54. Studer A, Zhao Q, Ross-Ibarra J, Doebley J (2011) Identification of a functional transposon
804 insertion in the maize domestication gene *tb1*. *Nature Genetics* 43(11):1160–1163.

805 55. Wang RL, Stec A, Hey J, Lukens L, Doebley J (1999) The limits of selection during maize
806 domestication. *Nature* 398(6724):236–239.

807 56. Guan JC, et al. (2012) Diverse Roles of Strigolactone Signaling in Maize Architecture and
808 the Uncoupling of a Branching-Specific Subnetwork. *Plant Physiology* 160(3):1303–1317.

809 57. Remigereau M-S, et al. (2011) Cereal Domestication and Evolution of Branching: Evidence
810 for Soft Selection in the Tb1 Orthologue of Pearl Millet (*Pennisetum glaucum* [L.] R. Br.).
811 *PLOS ONE* 6(7):e22404.

812 58. Minakuchi K, et al. (2010) FINE CULM1 (FC1) Works Downstream of Strigolactones to
813 Inhibit the Outgrowth of Axillary Buds in Rice. *Plant Cell Physiol* 51(7):1127–1135.

814 59. Studer AJ, Wang H, Doebley JF (2017) Selection During Maize Domestication Targeted a
815 Gene Network Controlling Plant and Inflorescence Architecture. *Genetics* 207(2):755–
816 765.

817 60. Rellstab C, Gugerli F, Eckert AJ, Hancock AM, Holderegger R (2015) A practical guide to
818 environmental association analysis in landscape genomics. *Molecular Ecology*
819 24(17):4348–4370.

820 61. Mohamed KI, Musselman LJ, Riches CR (2001) The Genus *Striga* (Scrophulariaceae) in
821 Africa. *Annals of the Missouri Botanical Garden* 88(1):60–103.

822 62. Gethi JG, Smith ME, Mitchell SE, Kresovich S (2005) Genetic diversity of *Striga*
823 *hermonthica* and *Striga asiatica* populations in Kenya. *Weed Research* 45(1):64–73.

824 63. Olivier A, Glaszmann J-C, Lanaud C, Leroux GD (1998) Population structure, genetic
825 diversity and host specificity of the parasitic weed *Striga hermonthica* (Scrophulariaceae)
826 in Sahel. *Pl Syst Evol* 209(1):33–45.

827 64. Bozkurt ML, Muth P, Parzies HK, Haussmann BIG (2015) Genetic diversity of East and
828 West African *Striga hermonthica* populations and virulence effects on a contrasting set of
829 sorghum cultivars. *Weed Research* 55(1):71–81.

830 65. Unachukwu NN, et al. (2017) Genetic diversity and population structure of *Striga*
831 *hermonthica* populations from Kenya and Nigeria. *Weed Research* 57(5):293–302.

832 66. Estep MC, et al. (2011) Genetic Diversity of a Parasitic Weed, *Striga hermonthica*, on
833 Sorghum and Pearl Millet in Mali. *Tropical Plant Biol* 4(2):91–98.

834 67. Samejima H, Babiker AG, Mustafa A, Sugimoto Y (2016) Identification of *Striga*
835 *hermonthica*-Resistant Upland Rice Varieties in Sudan and Their Resistance Phenotypes.
836 *Front Plant Sci* 7:634.

837 68. Mohemed N, et al. (2016) Evaluation of field resistance to Striga hermonthica (Del.) Benth.
838 in Sorghum bicolor (L.) Moench. The relationship with strigolactones. *Pest Manag Sci*
839 72(11):2082–2090.

840 69. Tarr SAJ (1961) Witchweed (*Striga hermonthica*) on rain-grown pearl millet in nitrogen-
841 deficient sandy soil of the central Sudan. *Annals of Applied Biology* 49(2):347–349.

842 70. Yohannes T, et al. (2015) Marker-assisted introgression improves *Striga resistance* in an
843 Eritrean Farmer-Preferred Sorghum Variety. *Field Crops Research* 173:22–29.

844 71. Joel KA, Runo S, Muchugi A (2018) Genetic diversity and virulence of *Striga hermonthica*
845 from Kenya and Uganda on selected sorghum varieties. *1* 10(2):111–120.

846 72. Mrema E, Shimelis H, Laing M, Bucheyeki T (2017) Screening of sorghum genotypes for
847 resistance to *Striga hermonthica* and *S. asiatica* and compatibility with *Fusarium*
848 *oxysporum* f.sp. *strigae*. *Acta Agriculturae Scandinavica, Section B — Soil & Plant*
849 *Science* 67(5):395–404.

850 73. Abate M, Hussien T, Bayu W, Reda F (2017) Diversity in root traits of sorghum genotypes
851 in response to *Striga hermonthica* infestation. *Weed Research* 57(5):303–313.

852 74. Bebawi FF (1981) Intraspecific Physiological Variants of *Striga hermonthica*. *Experimental*
853 *Agriculture* 17(4):419–423.

854 75. Welsh AB, Mohamed KI (2011) Genetic Diversity of *Striga hermonthica* Populations in
855 Ethiopia: Evaluating the Role of Geography and Host Specificity in Shaping Population
856 Structure. *International Journal of Plant Sciences* 172(6):773–782.

857 76. Gbèhounou G, Pieterse AH, Verkleij JAC (2000) Endogenously induced secondary
858 dormancy in seeds of *Striga hermonthica*. *Weed Science* 48(5):561–566.

859 77. Karger DN, et al. (2017) Climatologies at high resolution for the earth's land surface areas.
860 *Scientific Data* 4:170122.

861 78. Title PO, Bemmels JB (2018) ENVIREM: an expanded set of bioclimatic and topographic
862 variables increases flexibility and improves performance of ecological niche modeling.
863 *Ecography* 41(2):291–307.

864 79. Hengl T, et al. (2017) SoilGrids250m: Global gridded soil information based on machine
865 learning. *PLOS ONE* 12(2):e0169748.

866 80. Hengl T, et al. (2017) Soil nutrient maps of Sub-Saharan Africa: assessment of soil nutrient
867 content at 250 m spatial resolution using machine learning. *Nutr Cycl Agroecosyst*
868 109(1):77–102.

869 81. Muscarella R, et al. (2014) ENMeval: An R package for conducting spatially independent
870 evaluations and estimating optimal model complexity for Maxent ecological niche models.
871 *Methods in Ecology and Evolution* 5(11):1198–1205.

872 82. Schoener TW (1968) The Anolis Lizards of Bimini: Resource Partitioning in a Complex
873 Fauna. *Ecology* 49(4):704–726.

874 83. Warren DL, Glor RE, Turelli M (2008) Environmental niche equivalency versus
875 conservatism: quantitative approaches to niche evolution. *Evolution* 62(11):2868–2883.

876 84. Hijmans RJ, Phillips S, Elith JL and J (2017) *dismo: Species Distribution Modeling*
877 Available at: https://CRAN.R-project.org/package=dismo [Accessed March 28, 2019].

878 85. Brenton ZW, et al. (2016) A Genomic Resource for the Development, Improvement, and
879 Exploitation of Sorghum for Bioenergy. *Genetics* 204(1):21–33.

880 86. Cingolani P, et al. (2012) A program for annotating and predicting the effects of single
881 nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster*
882 *strain* w^{1118} ; iso-2; iso-3. *Fly* 6(2):80–92.

883 87. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler
884 transform. *Bioinformatics* 25(14):1754–1760.

885 88. Faust GG, Hall IM (2014) SAMBLASTER: fast duplicate marking and structural variant
886 read extraction. *Bioinformatics* 30(17):2503–2505.

887 89. Layer RM, Chiang C, Quinlan AR, Hall IM (2014) LUMPY: a probabilistic framework for
888 structural variant discovery. *Genome Biology* 15(6):R84.

90. Chiang C, et al. (2015) SpeedSeq: ultra-fast personal genome analysis and interpretation. *Nature Methods* 12(10):966–968.

91. Casa AM, et al. (2008) Community Resources and Strategies for Association Mapping in Sorghum. *Crop Science* 48(1):30–40.

92. Hess DE, Ejeta G, Butler LG (1992) Selecting sorghum genotypes expressing a quantitative biosynthetic trait that confers resistance to Striga. *Phytochemistry* 31(2):493–497.

93. Bates D, Mächler M, Bolker B, Walker S (2015) Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67(1):1–48.

94. Meyer E, Aglyamova GV, Matz MV (2011) Profiling gene expression responses of coral larvae (Acropora millepora) to elevated temperature and settlement inducers using a novel RNA-Seq procedure. *Molecular Ecology* 20(17):3599–3616.

95. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* 15(12):550.

96. Elshire RJ, et al. (2011) A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLOS ONE* 6(5):e19379.

97. Morris GP, et al. (2013) Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proceedings of the National Academy of Sciences* 110(2):453–458.

98. Hu Z, Olatoye MO, Marla S, Morris GP (2019) An Integrated Genotyping-by-Sequencing Polymorphism Map for Over 10,000 Sorghum Genotypes. *The Plant Genome* 12(1):0.

99. Upadhyaya HD, et al. (2009) Developing a Mini Core Collection of Sorghum for Diversified Utilization of Germplasm. *Crop Science* 49(5):1769–1780.

100. Browning BL, Browning SR (2016) Genotype Imputation with Millions of Reference Samples. *Am J Hum Genet* 98(1):116–126.

101. Zhou X, Stephens M (2012) Genome-wide efficient mixed-model analysis for association studies. *Nat Genet* 44(7):821–824.

102. Mace E, et al. (2018) The Sorghum QTL Atlas: a powerful tool for trait dissection, comparative genomics and crop improvement. *Theor Appl Genet*. doi:10.1007/s00122-018-3212-5.

103. Gillis J, Mistry M, Pavlidis P (2010) Gene function analysis in complex data sets using ErmineJ. *Nat Protoc* 5(6):1148–1159.

104. Brewer PB, et al. (2016) LATERAL BRANCHING OXIDOREDUCTASE acts in the final stages of strigolactone biosynthesis in Arabidopsis. *Proc Natl Acad Sci USA* 113(22):6301–6306.

105. Liu W, et al. (2011) Strigolactone biosynthesis in *Medicago truncatula* and rice requires the symbiotic GRAS-type transcription factors NSP1 and NSP2. *Plant Cell* 23(10):3853–3865.

106. Szpiech ZA, Hernandez RD (2014) selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. *Mol Biol Evol* 31(10):2824–2827.

SUPPLEMENTAL DATA FILES:

Supplemental Data File S1: Sorghum accessions from the GBS dataset with predicted *LGS1* loss-of-function alleles. Botanical race is coded as 'NA' if accession was not found in master list of accession annotations. Accessions with missing data at SNP S5_69985710 tagging the frameshift are not included in this list

Supplemental Data File S2: Differentially expressed genes between nutrient-stressed roots of *Striga*-susceptible sorghum line Shanqui Red and *Striga*-resistant line SRN39.

Supplemental Data File S3: *Striga hermonthica* occurrence records used for species distribution models presented in this study.

Supplemental Data File S4: Raster file of MaxEnt logistic output for all-occurrence *S. hermonthica* species distribution model.

Supplemental Data File S5: Raster file of MaxEnt logistic output for sorghum-only *S. hermonthica* species distribution model.

SUPPLEMENTAL MATERIAL & METHODS:

Collection of sorghum root exudates.

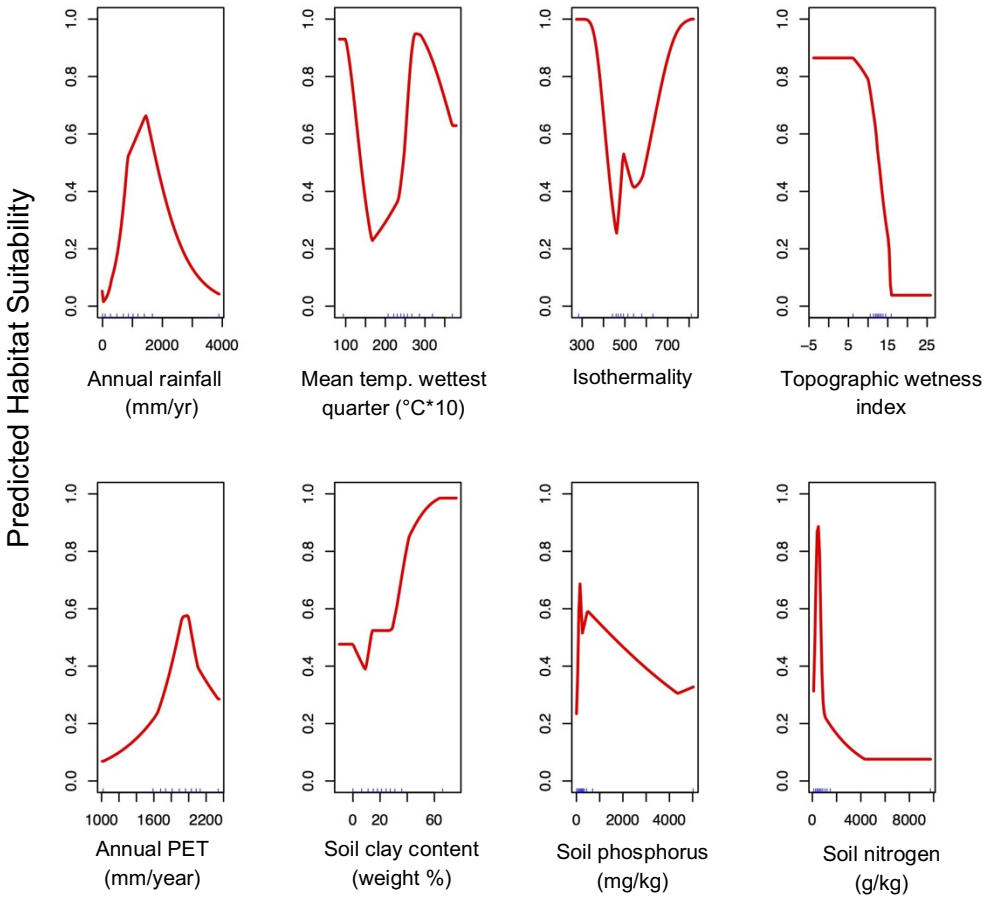
Accessions were grown individually in two-gallon pots with 70% potting mix (Premier pro-mix PGX) and 30% medium commercial grade sand in a greenhouse at 85°F during the day and 75°F at night with a 16-hour photoperiod. Pots were fertilized with 1x strength Miracle-Gro (Miracle-Gro® Water Soluble All Purpose Plant Food, The Scotts Company, LLC., Marysville, OH) once at fourteen days after planting. Forty-three days after planting, potting components were carefully washed from roots and whole plants were placed in separate flasks with a 1:5 ratio of root:DI water (v/v). Flasks were sealed

956 with parafilm to prevent evaporation and placed into darkness at room temperature.
957 After 48 hours, root exudate from each plant was centrifuged at 9,000 rpm for 10
958 minutes before removing supernatant for germination assays.
959
960 **Root RNA extraction and DNase treatment.**
961 For RNA extraction, root tissues were ground in liquid nitrogen using mortar and pestle.
962 Ground tissue powder was homogenized using an equal volume (W/V) of TRIzol™
963 reagent (Invitrogen, Cat#15596018) and 1 mL homogenate was used in RNA extraction.
964 RNA extraction generally followed TRIzol™ Reagent user guide. In brief, 200 µL of
965 Chloroform:Isoamyl alcohol 24:1(Sigma, C0549) was added to 1 mL homogenate and
966 incubated for 10 min on rotator mixer at room temperature. The homogenate was
967 centrifuged at 12000 RCF for 15 min at 4°C. The clear upper aqueous phase was
968 transferred to new 1.5 mL Eppendorf tubes and combined with an equal volume (V/V) of
969 isopropanol to precipitate nucleic acid. Samples were mixed well by inverting the tubes
970 several times and centrifuged as described above to obtain RNA pellet. The
971 supernatant was discarded and the pellet washed with 75% alcohol. The pellet was air
972 dried for 10 min at room temperature and suspended in 30 µL of 10 mM Tris-Cl (pH
973 8.0). RNA samples were left overnight in 4°C refrigerator to dissolve the pellet. Further,
974 RNA samples were treated with DNase I (Ambion^{TM}, AM2222) to remove contaminating
975 DNA. In brief, 25 µL of RNA was incubated with 2 units of DNase I and 3 µL of 10X
976 DNaseI buffer for 30 minutes at 37°C using a water bath. The nucleic acid was re-
977 precipitated, air dried and dissolved in 25 µL of 10 mM Tris-Cl (pH 8.0). The RNA was
978 quantified using NanoDrop® ND-1000 Spectrophotometer and ~100 ng of RNA was
979 analyzed by 1% gel electrophoresis to verify that the RNA was intact and free of
980 genomic DNA contamination.
981
982 **Library preparation and sequencing.**
983 We used 3'-TagSeq approach (Meyer *et al.,* 2011) with several modifications to
984 construct cDNA libraries. This method focuses on 3' end of transcripts enriched in a size
985 range of 400-500 bp fragments. Briefly, 1 µg of total RNA from each samples were
986 prepared in a volume of 10 µL and incubated with 8 µL of degradation buffer (0.5mM
987 dNTP mix, 1 mM DTT, 1X first strand buffer, 1 µM 3ILL-30TV oligo-dT primer to aim 3'
988 ends of cDNA) at 70°C for 16 min to achieve desire range of RNA fragments (300-600
989 bp). The entire fragmented RNA was further used in first strand cDNA (FS cDNA)
990 synthesis with RNA oligo primer (S-ILL-swMW) and SMARTScribe Reverse
991 Transcriptase (Clonetech). Then, cDNA was amplified with 16 cycles using Titanium
992 Taq Polymerase, 3ILL and 5ILL primers, and 10 µL of the FS cDNA as template. The
993 amplified cDNA products were purified using NucleoFast PCR Clean-up kit (Machery-
994 Nagel) and quantified using Qubit dsDNA High sensitivity Assay Kit (Invitrogen) on
995 Qubit® 2.0 Fluorometer.

996
997 Further, 50 ng of purified cDNA from each sample was individually barcoded with four
998 cycles of barcoding PCR. We used Illumina specific barcodes (ILL-BC) and multiplexed
999 those using Illumina TruSeq universal adapters (TrusSeq_Un) for use on the Illumina
1000 platform. An equal volume of barcoded cDNA libraries was pooled together and purified
1001 using PureLink Quick Gel Extraction and PCR Purification Combo kit (Invitrogen). The
1002 purified cDNA library pool was separated on 1.5% agarose in 1X TBE buffer (pH 8.2) to
1003 excise the fragment in the 400-500 bp size range. The excised gel fragment was
1004 purified using a gel extraction kit as mentioned above. The resulting library pool was
1005 loaded into a single lane of an Illumina Hiseq-2500 analyzer at the Genomic
1006 Sequencing and Analysis Facility at the University of Texas at Austin. We recovered
1007 between 3 to 5 million raw single-end 100 bp reads per sample. All primers used in the
1008 library preparation are given below.
1009
1010 **Bioinformatic analysis of TagSeq data.**
1011 Briefly, we excluded reads containing more than 20 bp with Phred quality scores <20,
1012 homopolymer runs longer than 30 bp, or at least one 12-mer match to Illumina adaptor
1013 sequences. Reads were trimmed to remove non-template bases introduced during
1014 library prep, identified based on occurrence of a 'GGG' motif within the first 10 bp.
1015 Reads were then mapped to the set of 34,211 primary transcripts from the *S. bicolor*
1016 v3.1 genome with SHRiMP v.2.2.3 (David *et al.* 2011) and we kept only unique
1017 alignments and those with more than 40 bp matching the reference.
1018
1019 **Oligos used in library preparations**
1020 **TruSeq_Un1:** AAT GAT ACG GCG ACC ACC GAG ATC TAC ACA TCA CGA CAC
1021 TCT TTC CCT ACA CGA CGC TCT TCC GAT CT
1022 **TruSeq_Un2:** AAT GAT ACG GCG ACC ACC GAG ATC TAC ACA CTT GAA CAC
1023 TCT TTC CCT ACA CGA CGC TCT TCC GAT CT
1024 **ILL-BC** (8 six-mer barcodes, here shown as NNNNNN):
1025 CAAGCAGAAGACGGCATACGAGATNNNNNNGTGACTGGAGTTCAGACGTGTGCTC
1026 TTCCGATC
1027 **S-III-swMW:**
1028 rArCrCrCrCrArUrGrGrGrGrCrUrArCrArCrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrNrNMW
1029 rGrGrG
1030 **3ILL-30TV:** ACGTGTGCTCTTCCGATCTAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTV
1031 **5ILL:** CTACACGACGCTCTTCCGATCT
1032
1033 **References**
1034 David, M., Dzamba, M., Lister, E., Ilie, L., Brudno, M. (2011). SHRiMP2: Sensitive yet
1035 Practical Short Read Mapping. *Bioinformatics.* 27, 1011-1012.

1036
1037 Meyer, E., Aglyamova, G. & Matz, M. (2011). Profiling gene expression responses of
1038 coral larvae (Acropora millepora) to elevated temperature and settlement inducers using
1039 a novel RNA-Seq procedure. *Mol. Ecol.* 20, 3599–3616.
1040

1041 **Supplemental Figures:**



1042

1043

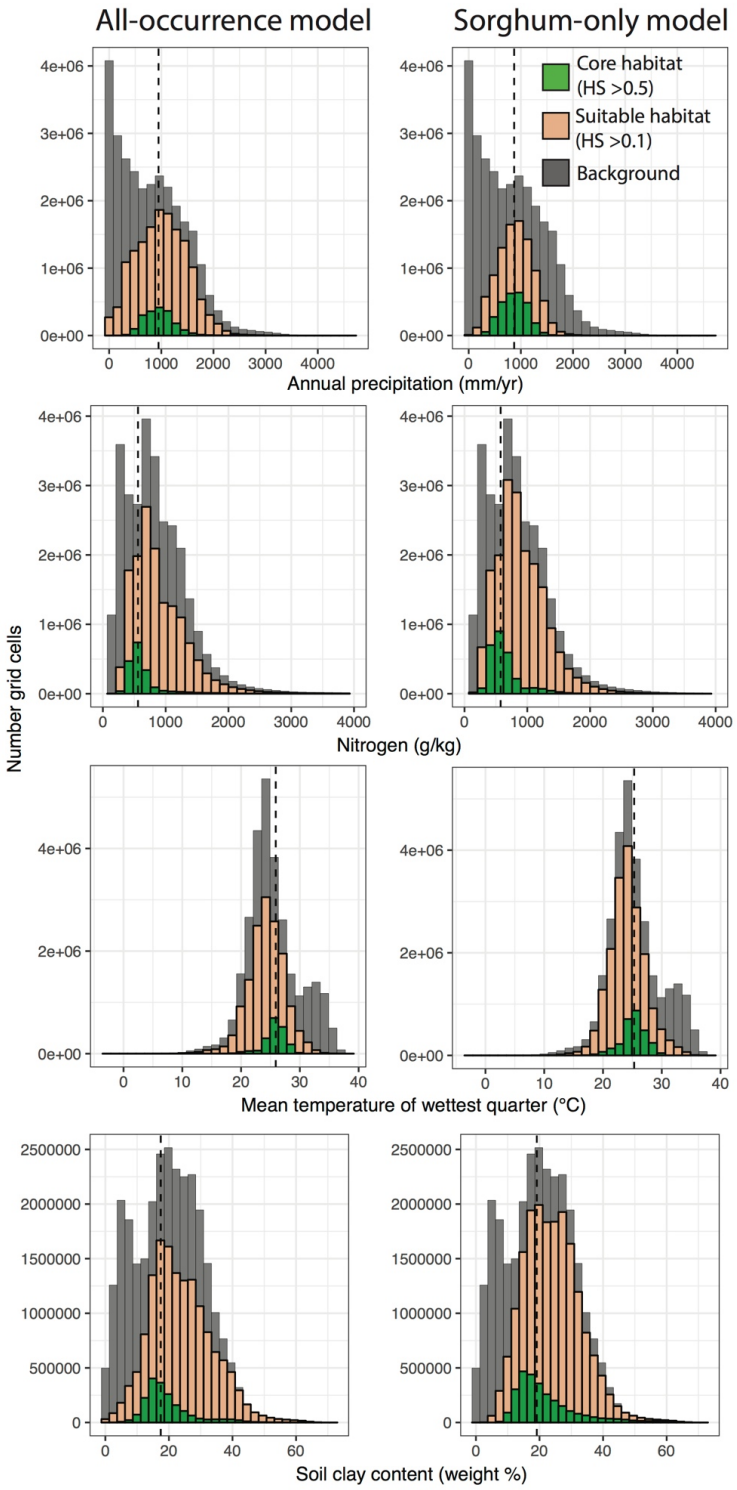
1044 **Supplemental Figure S1.** Response curves for 8 environmental variables in all

1045 occurrence *S. hermonthica* distribution model. Curves represent the change in predicted

1046 habitat suitability (y-axis) for each variable with all other variables held at a constant

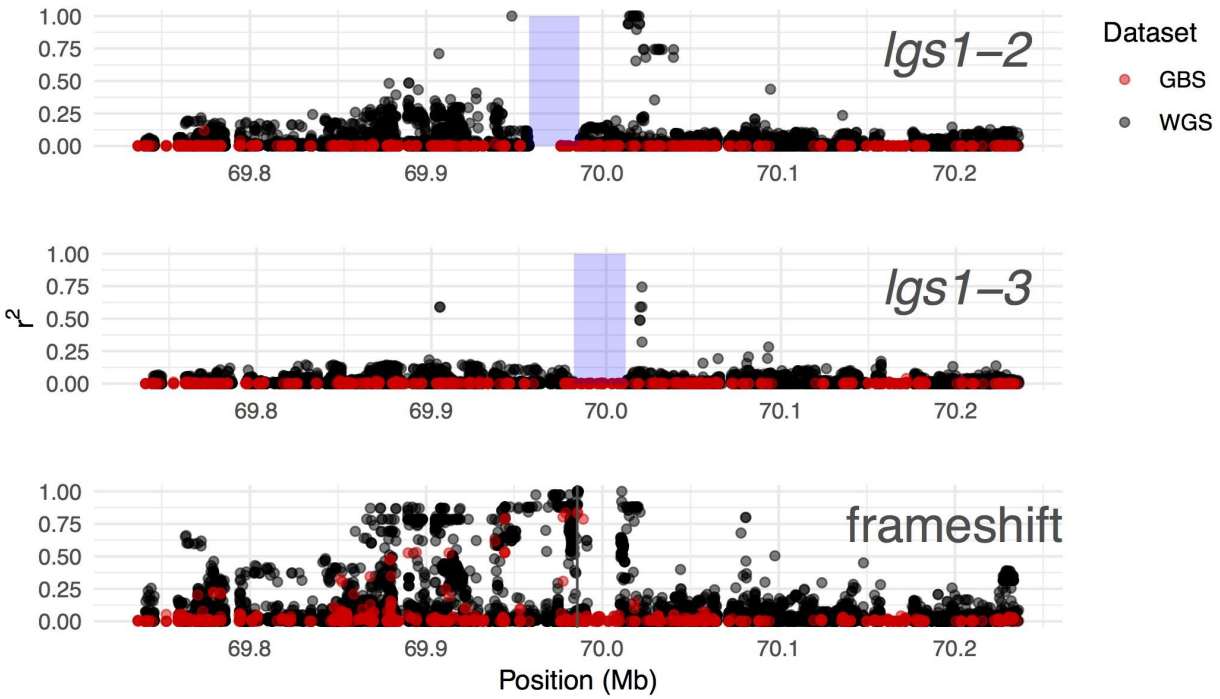
1047 value.

1048



Supplemental Figure S2. Distribution of environmental values for four variables with permutation importance greater than 10% for the all-occurrence or sorghum-only *S. hermonthica* SDMs. Distributions are shown for all grids cells exceeding habitat suitability (HS) scores of 0.5 or 0.1, or for all grid cells in the background.

1054

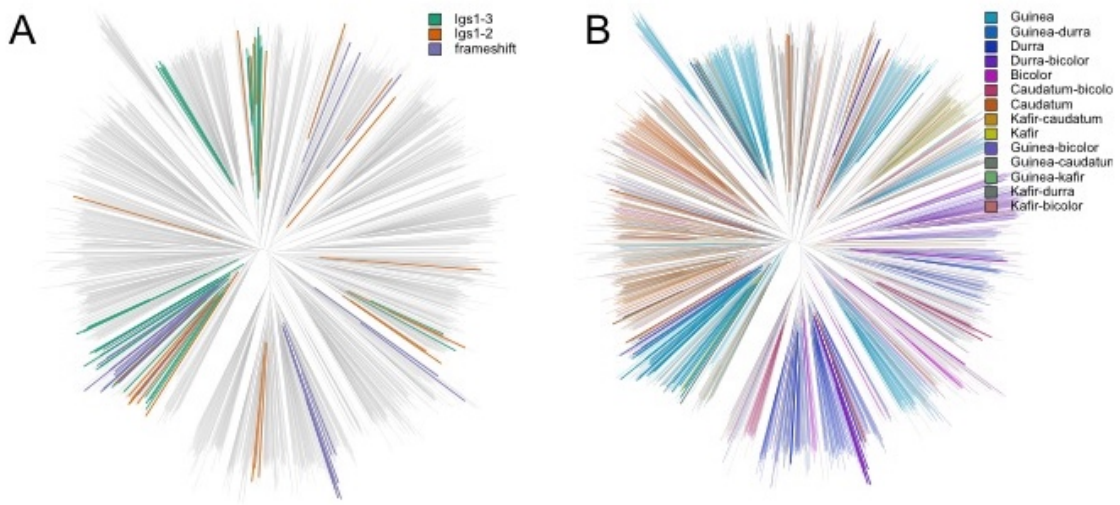


1055

1056 **Supplemental Figure S3.** Linkage disequilibrium with large deletion variants and the
1057 frameshift mutation at position 69,986,146 on *S. bicolor* Chromosome 5. Shaded boxes
1058 indicate positions of deletions (*lgs1-2*: 69,958,377-69,986,892*; lgs1-3*: 69,981,502-
1059 70,011,149) and vertical line indicates position of SNP tagging the frameshift mutation
1060 in the GBS dataset. For the WGS dataset, we excluded genotypes with depth of
1061 coverage less than 5x and SNPs with >70% missing data. Due to large amounts of
1062 missing data in the GBS dataset, r^2 was calculated based on imputed SNPs. An 18.8kb
1063 deletion beginning at position 69,991,579 also occurs in accessions with the frameshift,
1064 leading to missing data in this region for the non-imputed (WGS) dataset.

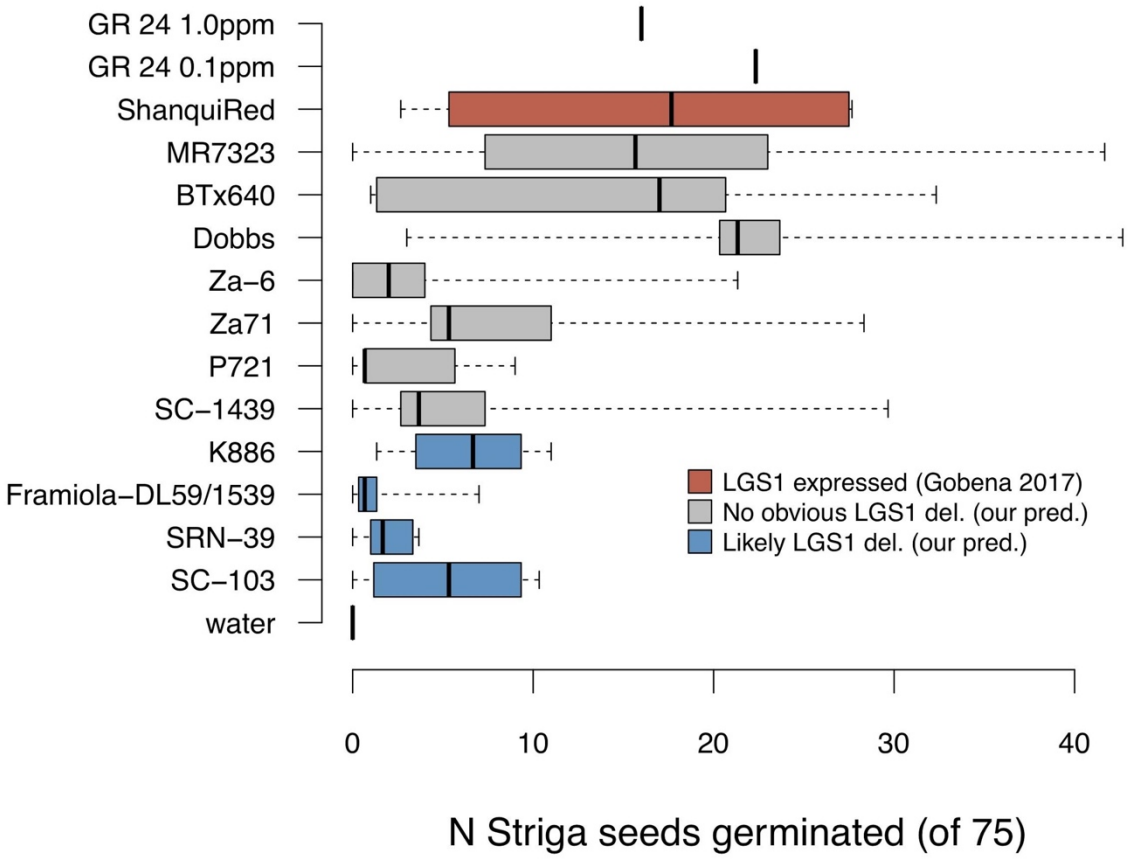
1065

1066



Supplemental Figure S4. Unrooted neighbor joining tree for *S. bicolor* chromosome 5 based on 2070 georeferenced individuals and 5434 SNPs with MAF >1%, after removing SNPs with >30% missing genotype calls. Branches are color-coded according to LGS1 deletion allele (A) or botanical race (B).

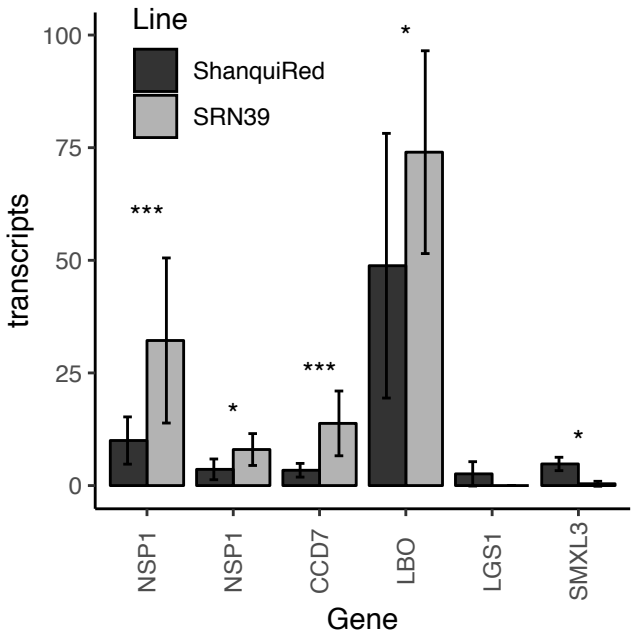
1073



1074

1075 **Supplemental Figure S5.** Germination of *S. hermonthica* in response to root exudates
1076 isolated from 12 sorghum varieties. Boxes show interquartile range and median,
1077 whiskers show maximum and minimum. Predictions for LGS1 deletions are based on
1078 genotypes from the GBS dataset.

1079



1080

1081 **Supplemental Figure S6. Strigolactone biosynthesis and signaling pathway genes**
1082 **differentially regulated between nutrient-stressed roots of *LGS1*-deficient**
1083 **sorghum line SRN39 and *LGS1*-intact line Shanqui Red.** Mean and standard
1084 deviation of read counts from 5 biological replicates per line are shown. Asterisks
1085 indicate significance following FDR correction. ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$.

1086

1087 **Table S1.** Environmental predictors for species distribution models. Core ranges are
1088 given as the 10th and 90th percentiles for grid cell values within the study extent (18W
1089 to 62E; 37S to 38N) with habitat suitability score >0.5 (all-occurrence and sorghum-only
1090 models) or for all grid cells within the full study extent. Median values are shown in
1091 parentheses. Models showed good discrimination ability with AUC values of 0.86 (all-
1092 occurrence model) or 0.85 (sorghum-only model). Both models included linear,
1093 quadratic, hinge, and product feature classes with a regularization multiplier of 1 for the
1094 all-occurrence model or 4 for the sorghum-only model. PI: Permutation importance

Variable	Units	Source	PI, % (all-occurrence)	PI, % (sorghum-only)	Core range (all-occurrence	Core range (sorghum-only)	Core range (full extent)
Annual rainfall (Bio12)	mm/year	CHELSA	35.6	28.7	608-1307 (947)	543-1183 (871)	42-1665 (737)
Total nitrogen at 0-30 cm depth (NTO)	g/kg	AfSoilGrids250m	24.3	28.8	411-932 (559)	399-977 (575)	257-1449 (766)
Clay fraction at 5 cm depth (CLYPPT)	weight %	SoilGrids250m	4.6	24.0	12-29 (17)	13-36 (19)	5-34 (20)
Mean temperature of wettest quarter (Bio8)	°C	CHELSA	15.4	4.0	23.8-27.8 (25.9)	22.5-27.8 (25.3)	20.6-32.3 (24.7)
Isothermality (Bio3)	-	CHELSA	8.1	4.8	464-583 (485)	464-573 (486)	433-622 (491)
Topographic wetness index (topoWet)	-	ENVIREM	6.5	6.6	11.1-13.5 (12.5)	10.5-13.8 (12.5)	10.4-14.6 (12.6)
Potential evapotranspiration (annualPET)	mm/year	ENVIREM	4.0	2.9	1774-2044 (1937)	1789-2090 (1958)	1540-2126 (1821)
Total phosphorus at 0-30 cm depth	mg/kg	AfSoilGrids250m	1.6	0.2	117-426 (192)	119-465 (211)	63-653 (264)

1095
1096

1097 Table S2. *LGS1* loss-of-function mutations for African landraces in the WGS dataset

Accession	Origin	Race	SV type	SV size (bp)	Start	GT
PI221651	Nigeria	Guinea	Insertion	2	69986146	1/1
PI221651	Nigeria	Guinea	Deletion	315	69984268	1/1
PI329338	Ethiopia	Caudatum-bicolor	Insertion	2	69986146	1/1
PI562781	Mali	Guinea	Deletion	29647	69981502	1/1
PI562971	Nigeria	Other	Deletion	28515	69958377	1/1
PI562981	Nigeria	Other	Deletion	28515	69958377	1/1
PI562982	Nigeria	Guinea	Deletion	28515	69958377	1/1
PI562990	Nigeria	Other	Insertion	2	69986146	1/1
PI562990	Nigeria	Other	Deletion	315	69984268	1/1
PI562991	Nigeria	Other	Deletion	28515	69958377	1/1
PI562994	Nigeria	Other	Deletion	28515	69958377	1/1
PI562998	Nigeria	Other	Insertion	2	69986146	1/1
PI563002	Nigeria	Other	Insertion	2	69986146	1/1
PI563009	Nigeria	Guinea-caudatum	Deletion	29647	69981502	1/1
PI563020	Nigeria	Other	Deletion	29647	69981502	1/1
PI563021	Nigeria	Unknown	Insertion	2	69986146	1/1
PI563021	Nigeria	Unknown	Deletion	315	69984268	0/1
PI563022	Nigeria	Other	Insertion	2	69986146	1/1
PI563022	Nigeria	Other	Deletion	315	69984268	1/1
PI585406	Nigeria	Guinea	Insertion	2	69986146	1/1
PI585406	Nigeria	Guinea	Deletion	315	69984268	1/1
PI585448	Ghana	Guinea	Insertion	2	69986146	1/1
PI585448	Ghana	Guinea	Deletion	315	69984268	1/1
PI585448	Ghana	Guinea	Deletion	29647	69981502	0/1
PI585452	Ghana	Guinea	Insertion	2	69986146	0/1
PI585452	Ghana	Guinea	Deletion	315	69984268	0/1
PI585467	Ghana	Guinea	Insertion	2	69986146	1/1
PI585467	Ghana	Guinea	Deletion	315	69984268	1/1
PI585966	Togo	Guinea	Insertion	2	69986146	1/1
PI585966	Togo	Guinea	Deletion	315	69984268	1/1

1098

1099

1100 **Table S3:** Sorghum genes related to strigolactone biosynthesis or signalling. The
1101 number of high impact variants with minor allele frequency (MAF) >5% in the WGS
1102 dataset are shown. Gene models with less than 5 SNPs were excluded from Tajima's D
1103 calculations. Expression for each gene model in root tissue of sorghum lines Shanqui
1104 Red and SRN39 are shown as the mean count for reads mapped, with asterisks
1105 indicating significant differences between lines. **adjusted *p* <0.001; *adjusted *p* <0.05

Homolog	Gene model	Function	Chr.	Pos.	Tajima's D	High impact variants (MAF >5%)	Expression
NSP1	Sobic.001G341400	GRAS-family transcription factor	1	62,864,836-62,867,833	NA	0	21.3**
NSP2	Sobic.001G428600	GRAS-family transcription factor	1	70,751,978-70,753,672	-0.51	0	3.0
D14	Sobic.001G465100	α/β-hydrolase	1	73,888,745-73,895,024	0.96	0	253
NSP1	Sobic.002G372100	GRAS-family transcription factor	2	73,049,658-73,053,841	-0.90	0	5.9*
CCD8	Sobic.002G380550	carlactone synthase	2	73,655,874-73,656,496	-0.84	1	0
MAX1	Sobic.003G269500	cytochrome P450	3	60,624,004-60,628,867	1.28	0	38.2
MAX1	Sobic.003G269600	cytochrome P450	3	60,633,944-60,636,136	0.92	1	73.4
CCD8	Sobic.003G293600	carlactone synthase	3	62,602,964-62,607,023	-0.23	0	24.0
LBO	Sobic.003G418000	oxidoreductase	3	72,382,455-72,384,461	3.80	0	60.6*
MAX1	Sobic.004G095500	cytochrome P450	4	8,304,449-8,307,025	2.39	0	0.4
SMXL5	Sobic.004G139900	Class I Clp ATPase	4	40,568,128-40,571,634	-0.07	0	1.5
SMXL3	Sobic.004	Class I Clp	4	51,922,148-	-0.19	2	1.9

	G168600	ATPase		51,925,516			
SMAX1	Sobic.004 G325000	Class I Clp ATPase	4	66,043,348-66,044,531	0.54	0	57.0
SMXL7	Sobic.005 G002400	Class I Clp ATPase	5	190,488-194,282	-0.64	0	3.9
CCD8	Sobic.005 G002500	carlactone synthase	5	202,329-204,425	-1.41	4	0.26
SMXL3	Sobic.005 G043000	Class I Clp ATPase	5	4,061,603-4,065,225[1]	0.74	3	24.3
CCD8	Sobic.005 G105700	carlactone synthase	5	19,797,305-19,799,526	-2.15	0	1.4
D27	Sobic.005 G168200	β-carotene isomerase	5	64,672,109-64,677,619	-0.88	0	3.0
LGS1	Sobic.005 G213600	sulfotransferase	5	69,984,440-69,986,256	0.19	3	1.1
SMXL3	Sobic.006 G074900	Class I Clp ATPase	6	43,746,994-43,751,171	-0.13	0	4.5
CCD7	Sobic.006 G170300	β-carotene cleaving dioxygenase	6	52,713,305-52,716,037[2]	0.47	2	9.6**
D27	Sobic.007 G016600	β-carotene isomerase	7	1,433,383-1,435,073	-1.13	0	0
SMAX1	Sobic.007 G090900	Class I Clp ATPase	7	14,458,102-14,461,202	2.31	0	80.6
CCD8	Sobic.007 G170300	carlactone synthase	7	60,503,029-60,506,362	-1.03	0	17.2
SMXL7	Sobic.008 G002400	Class I Clp ATPase	8	219,391-225,260	0.26	0	245.2
SMXL3	Sobic.008 G042800	Class I Clp ATPase	8	4,207,967-4,210,364	NA	0	2.2*
NSP2	Sobic.008 G046700	GRAS-family transcription	8	4,591,222-4,596,201	-1.58	0	0.4

[1] Overlaps with QTL05[1] (Yohannes et al. 2015)
[2] Overlaps with QTL06 (Yohannes et al. 2015)

		factor					
D27	Sobic.009G030800	β-carotene isomerase	9	2,741,604-2,744,934	NA	0	13.9
MAX2	Sobic.010G043000	F-box protein	10	3,312,894-3,316,043	NA	0	222.1
MAX1	Sobic.010G170400	cytochrome P450	10	50,177,361-50,180,730	NA	0	33.7

D27: DWARF27; CCD7: CAROTENOID CLEAVING DIOXYGENASE 7; CCD8: CAROTENOID CLEAVING DIOXYGENASE 8; MAX1: MORE AXILLARY BRANCHES 1; NSP1: NODULATION SIGNALING PATHWAY 1; NSP2: NODULATION SIGNALING PATHWAY 2; LBO: LATERAL BRANCHING OXIDOREDUCTASE; LGS1: LOW GERMINATION STIMULANT 1; SMAX1: SUPPRESSOR OF MAX2-1; SMXL7: SMAX1-LIKE 7; SMXL5: SMAX1-LIKE 5; SMXL3: SMAX1-LIKE 3

1113 Table S4. Sorghum SNPs showing significant genome-wide associations with *S.*
1114 *hermonthica* distribution (FDR correction at α=0.05). SNPs within 1kb of gene models
1115 are annotated and are ordered according to genomic location. AF: reference allele
1116 frequency; AdjP: P-value after FDR adjustment. For gene models with more than one
1117 SNP, the highest scoring position is shown. SNPs predicted to result in an amino acid
1118 change are shown in bold.

Chr	Position	AF	AdjP	Gene_model	Annotation
1	21828170[3]	0.292	0.046	Sobic.001G227800	Uncharacterized conserved protein, putative, expressed
1	50469815	0.035	0.037	NA	NA
1	54864458[4]	0.013	0.024	NA	NA
1	56800905	0.043	0.028	NA	NA
1	58808651	0.214	0.046	NA	NA
1	58837130	0.14	0.045	Sobic.001G304500	Os10g0551200 protein; GRAS domain family
1	64142276	0.052	0.024	NA	NA
1	64164495	0.054	0.004	Sobic.001G352600	Expressed protein
1	71929093	0.036	0.024	Sobic.001G441500	NA
2	12723171	0.028	0.035	NA	NA
2	21521798	0.275	0.013	Sobic.002G138400	Pectinesterase 14-related
2	68524689	0.018	0.049	Sobic.002G311400	Putative glucosyltransferase
3	1842885	0.011	0.045	NA	NA
3	2025341	0.012	0.002	Sobic.003G023500	Os06g0116300 protein; putative syntaxin
3	13445493	0.026	0.041	Sobic.003G139000	Os01g0290600 protein; putative L-cysteine desulfhydrase 2
3	51292869	0.012	0.023	NA	NA
3	56946323	0.055	0.043	Sobic.003G229700	glycerol-3-phosphate acyltransferase
3	59075308	0.018	0.008	Sobic.003G252200	HB1, ASXL, restriction endonuclease HTH domain (HARE-HTH)
3	62473154	0.077	0.002	NA	NA
3	62493603	0.093	0.028	Sobic.003G292500	ATP-dependent protease La (LON) domain-containing protein-like
3	62795897	0.012	0.002	Sobic.003G295900	Putative heat shock factor
3	67782723	0.012	0.008	Sobic.003G360200	WUSCHEL-related homeobox 9

[3] QTL from Haussman et al (2004) N13/E36-1
[4] QTL from Haussman et al (2004) N13/E36-1

3	70103330	0.055	0.017	Sobic.003G389700	Os01g0898800 protein
4	19466351	0.011	0.045	NA	NA
4	19724561	0.011	0.028	NA	NA
4	19735015	0.011	0.020	Sobic.004G132200	NA
4	20033536	0.014	0.040	NA	NA
4	20035072	0.012	0.027	NA	NA
4	20782140	0.012	0.012	NA	NA
4	23221234	0.041	0.012	NA	NA
4	60099891	0.012	0.002	Sobic.004G254500	Leafy cotyledon1
4	60767527	0.012	0.015	NA	NA
4	66105213	0.013	0.006	Sobic.004G325832	NA
4	66766354	0.03	0.040	NA	NA
5	1950501	0.087	0.045	NA	NA
5	1976589	0.086	0.013	Sobic.005G021400	NA
5	11089407	0.062	0.024	NA	NA
5	71122163	0.011	0.012	Sobic.005G224300	similar to O-methyltransferase
6	487441	0.086	0.040	Sobic.006G002900	similar to OSJNBa0094O15.15 protein; zinc/RING finger domain
6	15445897	0.011	0.046	NA	NA
6	42246344	0.129	0.028	NA	NA
6	43653802	0.022	0.018	Sobic.006G074300	similar to OSIGBa0092M08.2 protein; non-specific lipid transfer protein d6
6	51034850	0.028	0.033	NA	NA
6	51042152	0.079	0.050	Sobic.006G148800	Phenylalanine ammonia-lyase
6	52891272	0.087	0.004	NA	NA
6	59662162	0.012	0.042	Sobic.006G261700	similar to OSJNBb0004A17.4; Lactoylglutathione lyase
7	11575066	0.014	0.018	NA	NA
7	12044293	0.014	0.016	NA	NA
7	12094853	0.013	0.012	NA	NA
7	12891407	0.019	0.045	NA	NA
7	12895894	0.01	0.015	NA	NA
7	14193502	0.018	0.045	NA	NA
7	14220105	0.018	0.034	NA	NA
7	14279216	0.015	0.028	NA	NA
7	14459084[5]	0.014	0.007	Sobic.007G090900	ATP dependent CLP protease; SMAX1

[5] SL signaling (SMAX1 ortholog)

7	14942059	0.018	0.026	Sobic.007G091200	similar to H0117D06-OSIGBa0088B06.9 protein; Long-chain-alcohol O-fatty-acyltransferase
7	15219432	0.021	0.040	Sobic.007G091400	diacylglycerol kinase (ATP) (dgkA, DGK)
7	15254559	0.013	0.004	NA	NA
7	15325565	0.017	0.045	NA	NA
7	15359204	0.013	0.024	NA	NA
7	15532403	0.013	0.012	NA	NA
7	15534152	0.013	0.004	NA	NA
7	15554515	0.012	0.002	Sobic.007G092000	CCT/B-box zinc finger protein
7	15653673	0.019	0.046	Sobic.007G092200	aspartate carbamoyltransferase
7	15849206	0.014	0.027	NA	NA
7	15901805	0.017	0.026	NA	NA
7	15929687	0.019	0.018	Sobic.007G092600	Receptor-like kinase Xa21-binding protein 3-like
7	16293907	0.014	0.027	NA	NA
7	17082700	0.014	0.044	NA	NA
7	17092629	0.017	0.040	Sobic.007G093900	N-terminal acetyltransferase
7	17302379	0.019	0.045	Sobic.007G094600	Cytochrome P450 protein
7	17831891	0.019	0.040	Sobic.007G096200	Putative uncharacterized protein
7	17882047	0.018	0.028	NA	NA
7	17960172	0.018	0.018	Sobic.007G096800	Putative uncharacterized protein
7	18275486	0.012	0.018	NA	NA
7	18279010	0.013	0.015	NA	NA
7	18300164	0.012	0.024	NA	NA
7	18408217	0.012	0.015	NA	NA
7	19041599	0.011	0.037	Sobic.007G097400	sin-like protein conserved region containing protein
7	48195744	0.017	0.045	NA	NA
7	62605107	0.013	0.006	Sobic.007G193500	similar to Teosinte glume architecture 1
7	63109499	0.023	0.045	NA	NA
8	3177383	0.024	0.029	Sobic.008G034400	ATP-dependent RNA helicase, putative, expressed
8	13233629	0.016	0.018	NA	NA
8	60143889	0.012	0.026	Sobic.008G167300	Leucine-rich repeat-containing protein
8	61305329	0.014	0.043	Sobic.008G179600	similar to Expressed protein
9	1792519	0.011	0.017	Sobic.009G019900	NA

9	5744893	0.226	0.046	Sobic.009G056600	Putative uncharacterized protein
9	49976645	0.014	0.028	Sobic.009G142400	Putative gibberellin 20-oxidase
9	51474569	0.110	0.028	NA	NA
9	56421520	0.014	0.040	NA	NA
10	13677824	0.014	0.024	NA	NA
10	55681432	0.020	0.013	Sobic.010G213900	Receptor protein kinase PERK1-like
10	56006575	0.014	0.014	Sobic.010G216800	Putative uncharacterized protein
10	56761438[6]	0.026	0.018	NA	NA
10	59272319	0.016	0.028	Sobic.010G253700	NA

[6] QTL from Haussman et al (2004) IS9830/E36-1

1122 Table S5. Accessions from the Sorghum Association Panel used in germination study.

Accession	Name	Origin	Reported Resistance	*LGS1* allele
PI 656027	SRN39	-	Resistant	Deletion
PI 533752	SC103 (54.K.94)	-	Resistant	Deletion
PI 533972	Dobbs	Uganda	Resistant	Intact
PI 656025	Shanqui Red	China	Susceptible	Intact
PI 656055	P721	US	Susceptible	Intact*
PI 656051	MR7323	Niger	Unknown	Intact
PI 642791	BTx640	US	Unknown	Intact
PI 534088	Za6	Nigeria	Unknown	Intact
PI 534092	Za71	Nigeria	Unknown	Intact
PI 533976	Framiola DL59/1539	South Africa	Unknown	Deletion
PI 656040	K886	-	Unknown	Deletion
PI 656081	SC1439	-	Unknown	Intact*

1123 *No data in coding region, although SNP calls inconsistent with breakpoints for *lgs1-2* or
1124 *lgs1-3*.
1125