SUPPLEMENTARY FIGURES

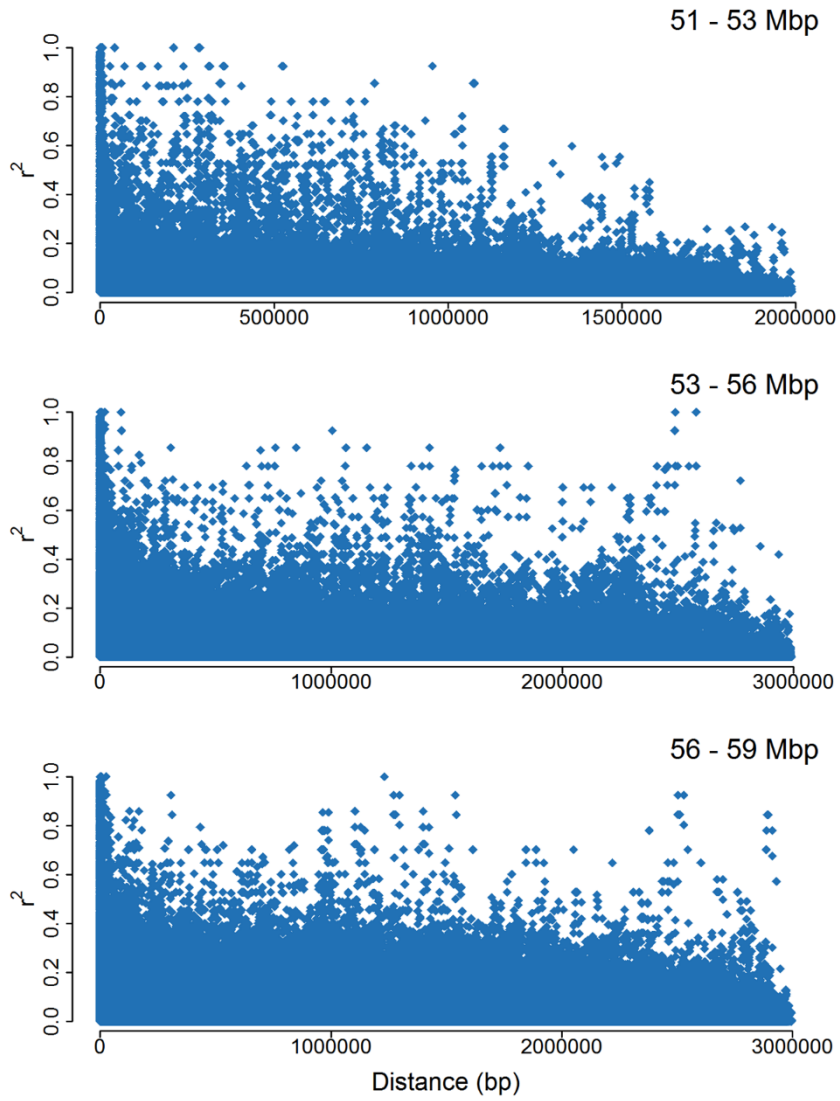

Fig. S1. Linkage disequilibrium (LD) at the end region of chromosome 9. The physical region where SNPs associated with Al tolerance and *SbMATE* expression were found (Fig. 2*F*) was assessed for LD. The squared allele-frequency correlations (1) (r^2) is shown as a function of physical distance between pairs of loci.

www.pnas.org/cgi/doi/10.1073/pnas.1808400115

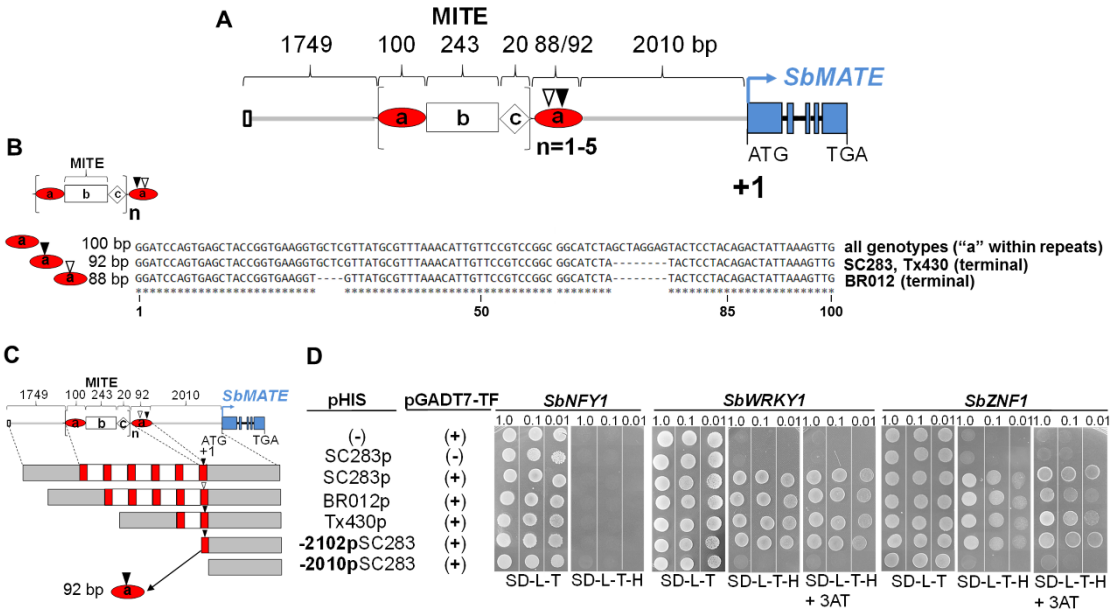

Fig. S2. SbWRKY1 and SbZNF1 activate *SbMATE* promoter alleles in yeast. (A) Structure of the SbMATE promoter region. The 243-bp MITE element (unit "b") is flanked by 100-bp (unit "a", 100-bp repeat) and 20-bp (unit "c") sequences. This MITE-containing a-b-c triplet (designated MITE repeats) is followed by a single copy of the 100-bp "a" unit with either an 8-bp deletion (depicted by an inverted black triangle, present in SC283 and Tx430) or a 12-bp deletion (inverted white triangle, present in BR012), which are the terminal 88- or 92-bp fragments depicted in (*B*). The number of identical a-b-c units (n), varies from 1 to 5 in different sorghum lines. (*C*) *SbMATE* promoter alleles (SC283p, BR012p and Tx430p) with 5, 4 and 1 repeats, respectively, and SC283 truncated constructs containing (-2102pSC283) or lacking (-2010pSC283) the 92-bp terminal were cloned into pHIS2.1. (*D*) Transactivation assays in yeast. The vector pGADT7 contained (+) either the nuclear transcription factor Y (SbNFY1, as a negative TF control), SbWRKY1 or SbZNF1 TFs (pGADT7-TF). Strains carrying empty plasmids (pGADT7(-) and pHIS(-)) are negative controls. Growth of co-transformed yeast in synthetic drop-out medium lacking leucine and tryptophan (SD-L-T) confirm the presence of both constructs. Transactivation (Y1H) was monitored on SD-L-T-H (H, histidine) with 10mM 3-AT. Transformant dilutions are shown at the top.

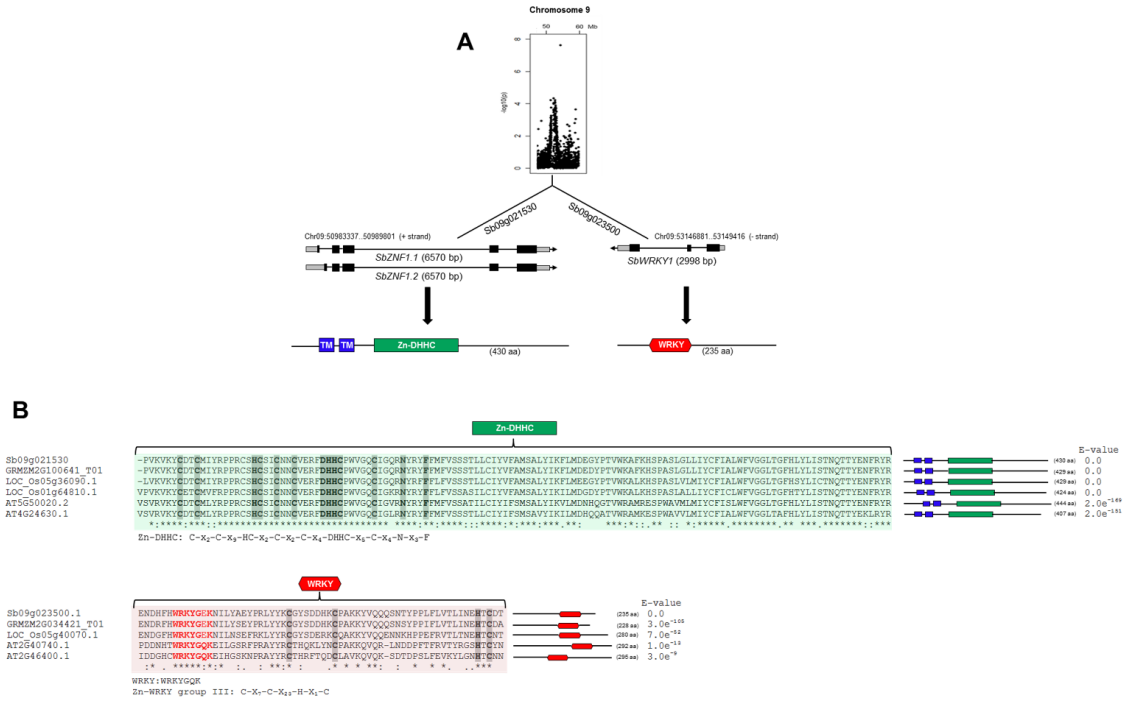

Fig. S3. *In silico* analysis of *SbWRKY1* and *SbZNF1*. (*A*) Physical positions of *SbZNF1* (*Sb09g021530*) and *SbWRK1* (*Sb09g023500*) in the context of the *SbMATE* expression QTL on chromosome 9 (Fig. 2*C*) and gene models based on *Sorghum bicolor* v3.1 (https://phytozome.jgi.doe.gov/pz/portal.html). This version predicts an isoform of *SbZNF1* (*SbZNF1.2*) in addition to *SbZNF1.1*, which is common to both v1.4 and v3.1 of the sorghum genome. Protein domains were predicted with Pfam (http://pfam.xfam.org/). The positions of the transmembrane (TM), Zn-DHHC and WRKY domains are shown. (*B*) Multiple alignment of the Zn-DHHC and WRKY domains was undertaken with Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/) and domain architectures of related proteins in maize (GRMZM), rice (LOC_Os) and Arabidopsis (AT) are shown. Expected values (*e-values*) were obtained based on a sequence similarity analysis using as query the complete protein sequences of SbWRKY1 and SbZNF1 in Phytozome. Alignment of the WRKY domain indicated that SbWRKY1, AtWRKY46 (AT2G46400) and the remaining proteins possess a single C2H-type zinc-finger motif, typical of WRKY Group III (2). The WRKY (2) and the Zn-DHHC consensus (Pfam, PFM01529) are shown. Identical (*) amino acid residues and conservation between groups of strongly (:) or weakly (.) similar properties are shown in the ClustalO output.

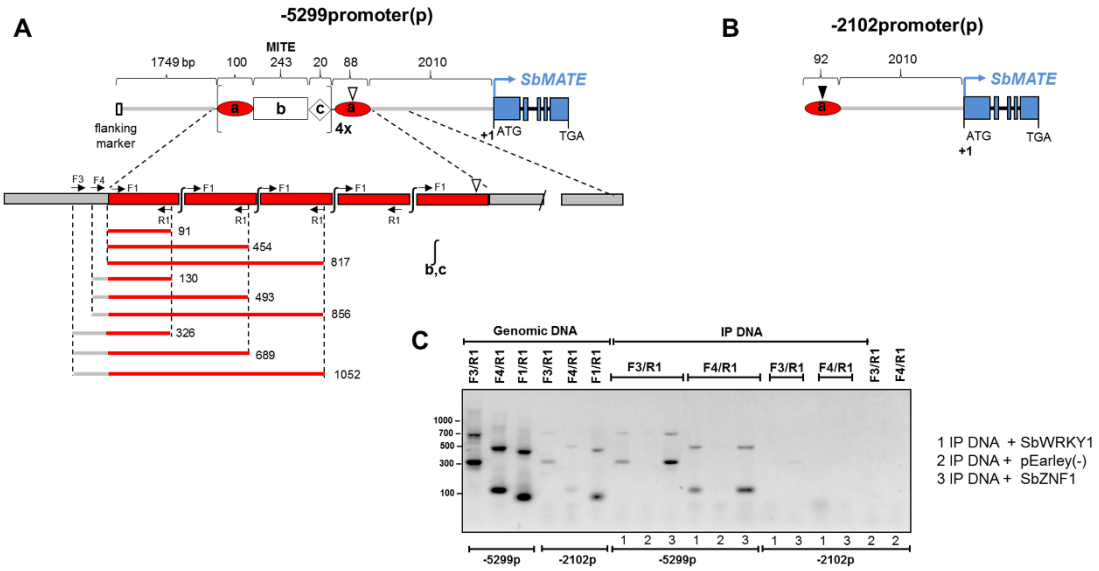

Fig. S4. Chromatin Immunoprecipitation (ChiP) assays show that SbWRKY1 and SbZNF1 bind to the 100-bp a repeat (within the MITE repeats). Schematic representation of the *SbMATE* promoters used in the ChIP assays: (*A*) promoter fragment from the BR012 line extending to position -5,299-bp [-5299promoter(p)] relative to the *SbMATE* start codon, and (*B*) SC283 promoter fragment extending to position -2,102-bp relative to the SbMATE start codon [-2102promoter(p)]. White and black triangles on the terminal “a” unit in BR012 and SC283 depict a 12-bp deletion in BR012 and 8-bp deletion in SC283, compared to the full 100-bp a unit within the MITE. Arrows represent primer positions whose sequences are in Table S2. The forward primers, F3 and F4, anneal to the unique 1749-bp fragment upstream of the 100-bp repeat whereas the R1 primer anneals within this 100-bp repeat. The expected sizes of each PCR-amplified fragments are shown as numbers in front of thick red and gray lines. (*C*) PCR analysis of immunoprecipitated (IP) DNA from Arabidopsis protoplasts transfected with vector pEarleyGate104 containing SbWRKY1 or SbZNF1 inserts, or the empty vector (minus sign). Genomic DNA from *A. thaliana* plants transformed with the promoter constructs was used as positive control.

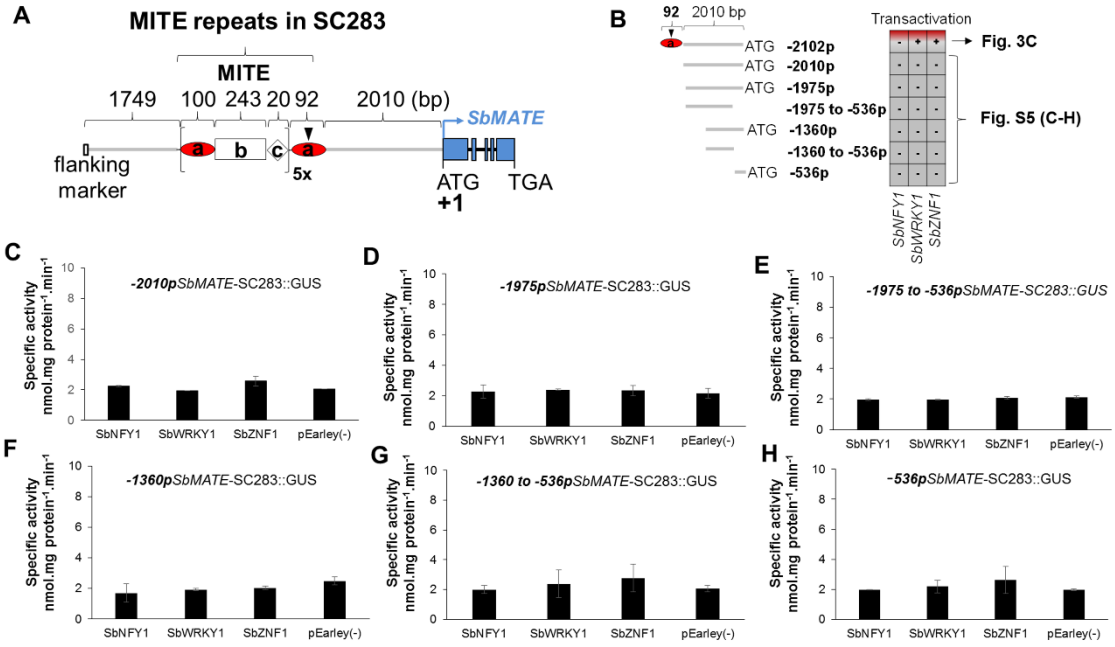

Fig. S5. Transactivation analysis of different truncated fragments of the *SbMATE* promoter by SbWRKY1 and SbZNF1. (*A*) Structure of the MITE repeats as described in Fig. 3. Protoplasts obtained from Arabidopsis were stably transformed with the GUS-fused promoter constructs and then electroporated with plasmids containing SbNFY1, SbWRKY1, SbZNF1, or the empty pEarleyGate104 (pEarley(-)) vector under the control of the S35 promoter. (*B*) Schematic representation of the transactivation results with a construct containing the terminal 92-bp fragment (data in Fig. 3*C*) or in truncated constructs excluding this terminal fragment (*C* – *H*). After 36 hours, β-glucuronidase activity was determined from protein extracts isolated from the transfected cells. The activities are shown as the means ± S.D. from three independent experiments, each of which was performed with triplicate samples. The transactivation activity between the empty vector and the TFs for each promoter:GUS construct is not statistically different. The *SbMATE* promoter of SC283 was used to construct promoter truncations.

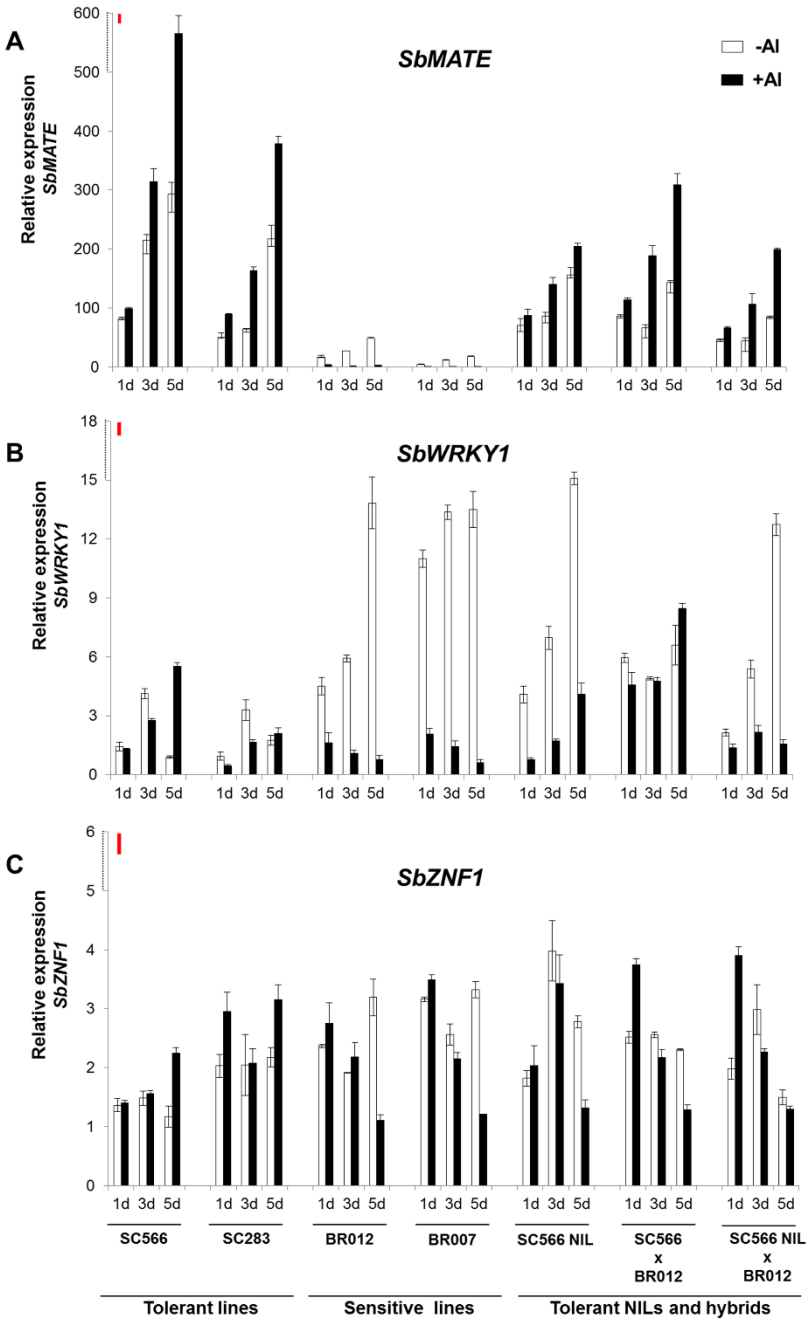
A
SbMATE
-Al
+Al
Relative expression SbMATE
600
500
400
300
200
100
0
1d 3d 5d 1d 3d 5d 1d 3d 5d 1d 3d 5d 1d 3d 5d 1d 3d 5d 1d 3d 5d
B
SbWRKY1
Relative expression SbWRKY1
18
15
12
9
6
3
0
1d 3d 5d 1d 3d 5d 1d 3d 5d 1d 3d 5d 1d 3d 5d 1d 3d 5d 1d 3d 5d
C
SbZNF1
Relative expression SbZNF1
6
5
4
3
2
1
0
1d 3d 5d 1d 3d 5d 1d 3d 5d 1d 3d 5d 1d 3d 5d 1d 3d 5d 1d 3d 5d
SC566
SC283
BR012
BR007
SC566 NIL
SC566 x BR012
SC566 NIL x BR012
Tolerant lines
Sensitive lines
Tolerant NILs and hybrids

Fig. S6. Temporal expression of *SbMATE* and interacting transcription factors. (*A*) *SbMATE*, (*B*) *SbWRKY1* and (*C*) *SbZNF1* expression in SC566 (Al-tolerant), BR012 (Al-sensitive), SC566 - NIL (BR012 genetic background) and the hybrids BR012 x SC566-NIL and BR012 x SC566. All these genotypes possess 4 MITE repeats. The parents of our RIL population, SC283 (Al-tolerant) and BR007 (Al-sensitive), with 5 and 3 MITE copies, respectively, are included. The sorghum genotypes were grown without (-Al) or with {27} µM Al^{3+} (+Al) in nutrient solution at pH 4.0 for 1, 3 and 5 days and root apices (1 cm) were collected for RNA isolation. Error bars depicting standard deviation (n=3) are shown and least significant difference (Fisher´s LSD, α = 0.05) bars (in red) are drawn to scale for each gene (top of the Y-axis). A Pearson correlation analysis was undertaken with gene expression excluding SC283 and BR007, so that only genotypes with 4 MITE copies were used. *SbWRKY1* expression was significantly correlated with *SbMATE* expression (r = 0.32, *p* = 0.08).

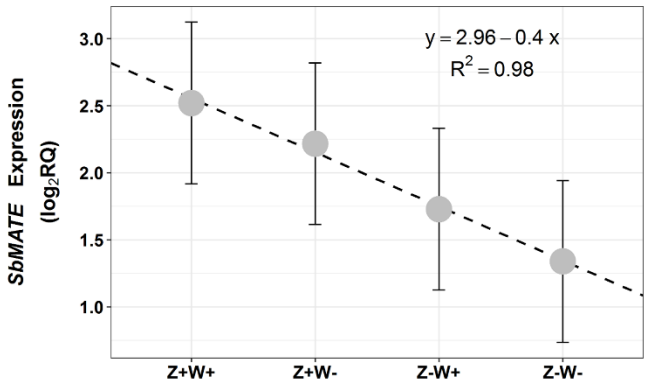

Fig. S7. Full linear regression model fit to *SbMATE* expression data for all *SbWRKY1*/*SbZNF1* haplotypes, which is shown as a linear trend-line in Fig. 4*C*. *SbWRKY1* and *SbZNF1* effect on *SbMATE* expression was estimated based on RILs homozygous for the SC283 (Al-tolerant) alleles at both TF loci (**Z+/W+**), for the BR007 (Al-sensitive) allele (**Z-/W-**)**,** or showing alternate TF alleles (**Z+/W-** and **Z-/W+**). The regression model was highly significant (α = 0.01). The intercept (2.96), the slope (|0.4|) and the coefficient of determination (R^2) are shown on top. The high R^2 value (0.98) indicates that haplotype variation for the TFs largely explain the variation in *SbMATE* expression. 90% confidence intervals for haplotype means are also shown. RQ: relative quantification using the ΔΔCt method (see methods).

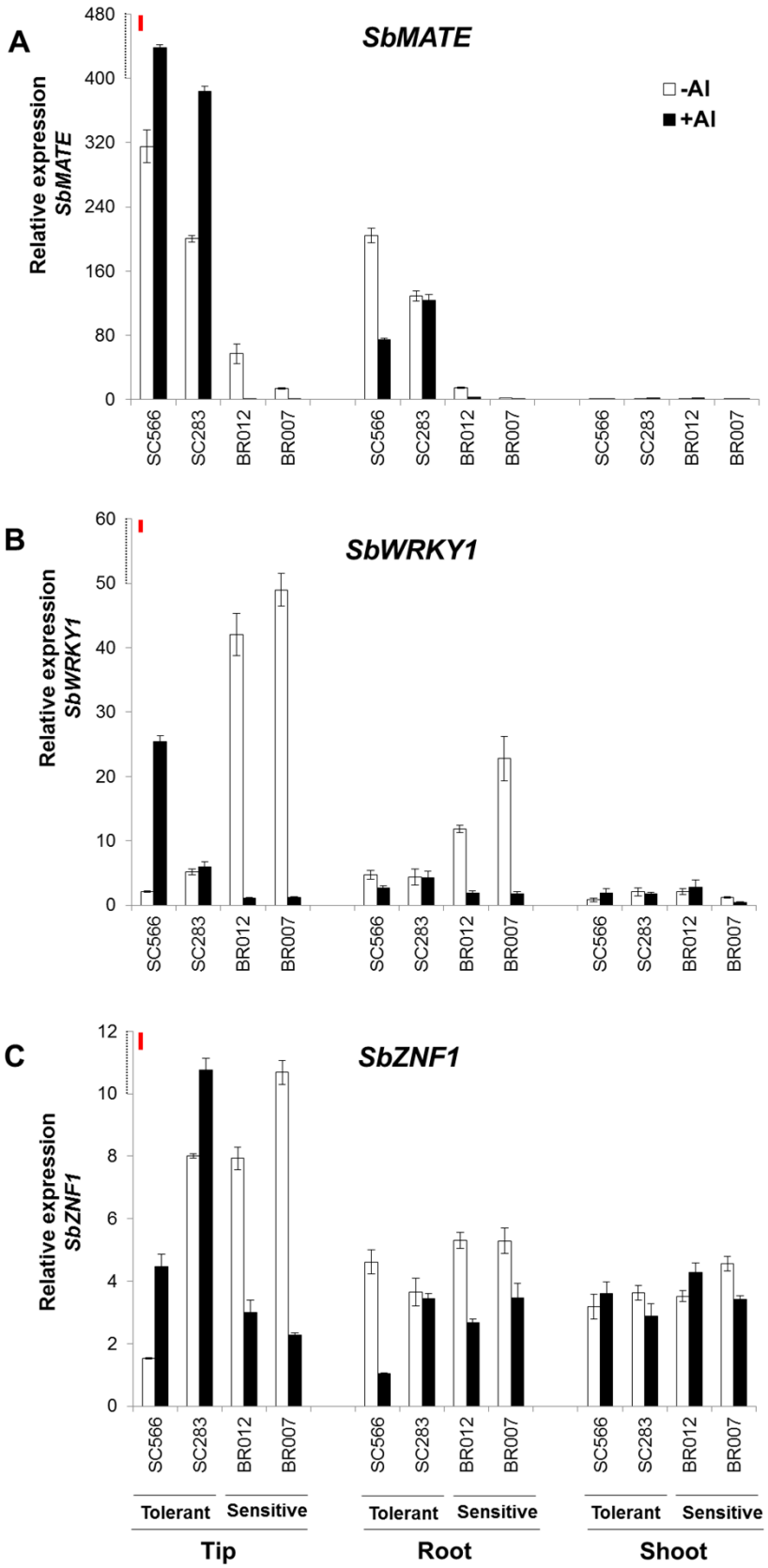
A
SbMATE
Relative expression SbMATE
-Al
+Al
480
400
320
240
160
80
0
SC566
SC283
BR012
BR007
B
SbWRKY1
Relative expression SbWRKY1
60
50
40
30
20
10
0
C
SbZNF1
Relative expression SbZNF1
12
10
8
6
4
2
0
Tolerant
Sensitive
Tip
Root
Shoot

Fig. S8. Tissue-specific expression of *SbMATE* and the interacting transcription factors. (*A*) *SbMATE*, (*B*) *SbWRKY1* and (*C*) *SbZNF1* expression in SC566 and SC283 (Al-tolerant) and the Al-sensitive lines, BR007 and BR012. The sorghum genotypes were grown without (-Al) or with {27} µM Al^{3+} (+Al) in nutrient solution at pH 4.0 for 5 days. The root tip (1 cm), a composite sample of the rest of the root system excluding the root apex, and shoots were collected for RNA isolation. Standard deviation bars of three replicates are shown and least significant difference (Fisher´s LSD, α = 0.05) bars (in red) are drawn to scale for each gene (top of the Y-axis).

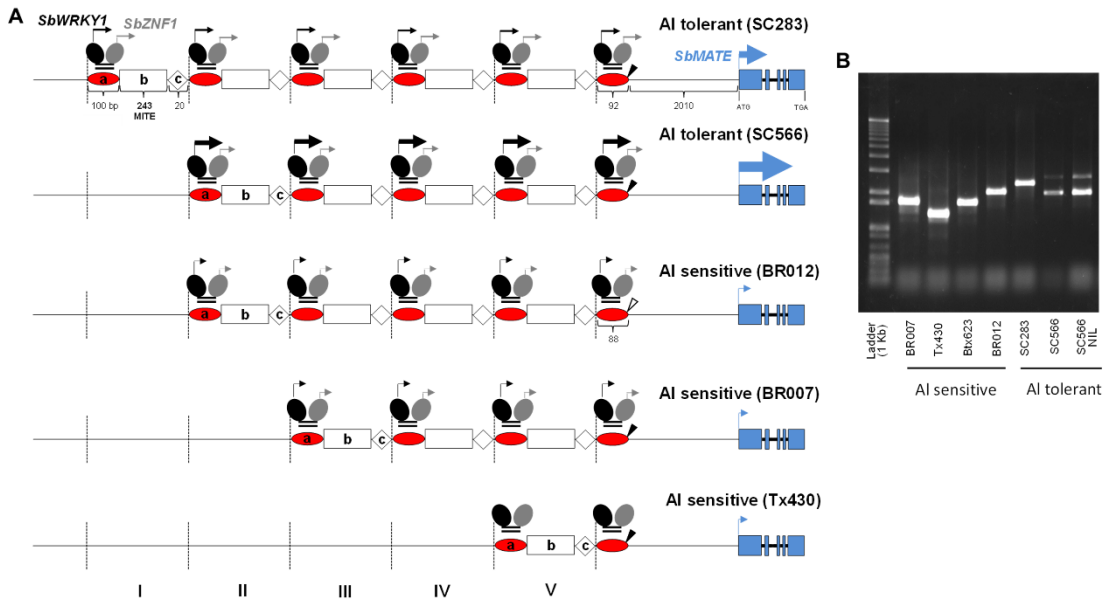

Fig. S9. *SbMATE* is regulated by *cis*/*trans* interactions. (*A*) The general structure MITE region is depicted: the 243-bp MITE element (unit "b") is flanked by 100-bp (unit "a", 100-bp repeat) and 20-bp (unit "c") sequences, and the number of identical a-b-c triplets varies in different promoter alleles. The MITE-containing a-b-c triplets (MITE repeats) is terminated by a single 100-bp "a" unit with either an 8-bp deletion (depicted by an inverted black triangle, present in SC283 and Tx430) or a 12-bp deletion (inverted white triangle, present in BR012), which results in the 88- or 92-bp terminal fragments. SbWRKY1 and SbZNF1 protein binding is represented by black (SbWRKY1) and gray (SbZNF1) ovals above the binding fragment (red). The size of the arrows above genes is proportional to relative expression values following a 5 day Al exposure for *SbWRKY1* and *SbZNF1* in Fig. 4*A* and *SbMATE* in Fig. S6. (*B*) Amplification profile of the MITE repeats in Al-tolerant and Al-sensitive lines (3). Our model depicts enhanced recruitment of SbWRKY1 and SbZNF1 to the *SbMATE* promoter region in Al-tolerant compared to Al-sensitive lines due to increased binding site abundance in the Al-tolerant lines, as well as reduced TF expression in Al-sensitive backgrounds, reducing *SbMATE* expression.

SUPPLEMENTARY METHODS

Genetic Stocks

Near-isogenic lines (NILs) were developed by introgressing the *SbMATE* allele from the Al-tolerant lines, SC566 and SC283, into the genetic background of the Al-sensitive line, BR012, using marker-assisted selection as previously detailed (4). A BR007 (Al-sensitive) x SC283 (Al-tolerant) RIL population (396 F9:10), was used for Al tolerance and *SbMATE* expression (191 RILs) QTL mapping. A 254-member sorghum association panel (3, 5) was used for GWAS.

In the BR012 x SC566 hybrid, *SbMATE* and the remainder of the genome are heterozygous. The SC566-NIL (4) is homozygous for the *SbMATE* allele of SC566 in the Al-sensitive BR012 background. The near-isogenic hybrid, BR012 x SC566-NIL is heterozygous for *SbMATE* within a homozygous, Al-sensitive BR012 genetic background.

Al Tolerance in Hydroponics

Al tolerance was assessed based on root growth inhibition elicited by {27} µM Al^{3+} in nutrient solution at pH 4.0, as previously reported in Table S4 of Caniato et al. (2014) (3) and in Fig. 2 of Melo et al. (2013) (4). Sorghum accessions were evaluated for Al tolerance using the basal nutrient solution described in Magnavaca et al. (6) containing {0} or {27} µM Al^{3+} (values inside brackets indicate Al^{3+} activity, which was estimated with the speciation software program, GEOCHEM-PC (7). Sorghum seeds were surface-sterilized with 0.5% (w/v) NaOCl for 15 min, rinsed with ultra-pure water and allowed to germinate for 3 days at 27^{o}C on wet filter paper. Seedlings were then transferred to 8 L

containers with nutrient solution lacking Al at pH 4.0 and placed in a growth chamber with 27ºC day and 20ºC night temperatures, a light intensity of 330 µmol photons m^{-2} s^{-1} and a 12-hour photoperiod. After 24 h of acclimation, the initial length of each seedling's primary root growing in control solution (ilc) was measured. The solution was then replaced with nutrient solution of identical composition but containing either no Al or {27} µM Al^{3+} supplied as $AlK(SO_4)_2.12H_2O$. Final root lengths under Al treatment (flAl) or in control solution (flc) were obtained after five days of exposure to Al. *R*elative percent values of *N*et *R*oot *G*rowth *I*nhibition after five days of Al exposure (RNRG) were estimated by dividing the net root growth under Al treatment (flAl–ilc) by the net root growth without Al (flc–ilc).

Gene Expression Via Quantitative RT-PCR

Global SbMATE expression via TaqMan assay

Quantitative real-time analysis of gene expression was undertaken as described in Melo et al. (4) with the 7500 Fast Real-Time System (Applied Biosystems). First-strand cDNA was synthesized using 2 µg of total RNA with the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems). A series of cDNA dilutions were used for making standard curves both for *SbMATE* transcripts and for 18S *rRNA*, which was used as the reference gene. Then, the selected dilution for specific cDNA samples (10 ng for *SbMATE* transcripts and 0.01 ηg for 18S *rRNA*) were used as real-time PCR templates to quantify relative transcript levels with the TaqMan Gene Expression kit (Applied Biosystems) following the conditions recommended by the manufacturer. The forward (SbMATE_1F) and reverse (SbMATE_1R) primers and the probe (A+T)

sequences are shown in Table S2. This probe hybridizes to a monomorphic region of *SbMATE* and thus assesses the joint expression of both alleles (A+T) present in hybrid stocks. Expression of endogenous 18S *rRNA* was determined using TaqMan Ribosomal RNA Control Reagents (Applied Biosystems). Reactions with deionized water were used as negative controls. Relative quantification was calculated using the ΔΔCt method (8) and three technical replicates were adopted.

Allele-specific expression of SbMATE

An allele-specific expression assay was developed by combining the TaqMan analysis of gene expression with the TaqMan Allelic Discrimination kit (Applied Biosystems). This assay was developed based on a thymine/adenine (T/A) SNP present in the first exon of the *SbMATE* coding region (4) and differentiated expression of the A allele present in SC566 from the T allele present in the other sorghum lines. A pair of primers flanking the SNP locus (SbMATE_1F and SbMATE_1R) and a pair of probes (probe A and probe T), each specific to one *SbMATE* allele, were designed using the Primer Express 3.0 (Applied Biosystems) software with option TaqMan Allelic Discrimination (Table S2). Because both alleles are present in hybrid genotypes, unspecific probe hybridization may occur. The unspecific signal was estimated by hybridizing each probe to homozygous stocks expressing the non-complementary *SbMATE* allele for that probe and different primer and probe concentrations were tested to minimize the unspecific signal, which was always assessed and used to normalize the allele-specific signal. Allele-specific analyses of *SbMATE* expression were performed with primer and probe concentrations of 500 and 20 ηM, respectively. Under these conditions, the

nonspecific signal from the T probe ranged from 0.8% to 1.2% and was undetectable for the A probe. Allelic-specific expression was obtained using the $2^{-\Delta\Delta Ct}$ method and validated using the relative standard curve method.

SYBR Green assay for SbMATE, SbZNF1 and SbWRKY1

SbZNF1 and *SbWRKY1* expression in response to Al^{3+}, and temporal and tissue expression of these transcription factors as well as *SbMATE* were quantified using SYBR green technology with the primers detailed in Table S2. Root apices (1 cm) were collected, frozen in liquid nitrogen and total RNA was isolated using the RNeasy Plant Mini Kit (Qiagen). For assessing tissue expression, composite root samples (rest of the root excluding the root apex) and shoots were also collected after 5 days of Al^{3+} exposure. Quantitative real-time PCR was carried out in triplicate containing cDNA (5 ηg for target genes and 0.005 ηg for 18S *rRNA*), 2.5 μM of each primer and Fast SYBR Green Master Mix 1X (Applied Biosystems) in a final volume of 10 μL. First-strand cDNA synthesis and relative expression calculation were performed as described above for the TaqMan assays.

QTL Mapping in a RIL Population

Genomic DNA was isolated from leaf tissue as described by Saghai-Maroof et al. (9) Genotyping-by-sequencing (GBS) (10) was performed to obtain SNP data. Sequencing was performed on the Genome Analyzer II (Illumina, Inc.) and the sequences were processed and aligned to the sorghum reference genome (*Sorghum bicolor* v1.4, https://phytozome.jgi.doe.gov/pz/portal.html) using the Burrows-Wheeler alignment (BWA) method for SNP identification. Missing data imputation was performed with the software NPUTE (11), with

haplotype window size varying from 28 to 40. Optimal window sizes were selected to maximize imputation accuracy based on known allele values. SNP filtering was undertaken with a minimum count (the minimum number of taxa in which the site must have been scored to be included in the filtered data set) of 60% and a minimum minor allele frequency (MAF) threshold of 0.4. QTL mapping based on a general linear model (GLM) was undertaken with TASSEL v3.0 (12) considering 1000 permutations (13). A total of 68,878 SNPs (GBS) (10) were tested with a GLM model.

A procedure similar to multiple interval mapping (MIM) (14), which allows fitting multiple QTLs simultaneously with main effects and epistatic interactions, was also performed. The high marker density provided by GBS precluded the need for estimating conditional probabilities of putative QTLs and fitting a mixture model. First, a single marker analysis was performed using a linear regression model. Then, a multiple regression model was fitted by adding the most significant SNP to the model based on the preceding step. This was done with a forward procedure until no more significant SNPs were found. Finally, all pairwise epistatic interactions among SNPs that were kept in the model were tested. SNP effects, partial (R_P^2) and total (R_T^2) coefficients of determination were estimated with a multi-locus regression model. R_T^2 was calculated from the full model, including all SNP effects, whereas the R_P^2 was estimated for each SNP as follows: i) we estimated the sum of squares of residuals of the reduced model (SSE_{rm}) without the SNP effect and the sum of squares of residuals (SSE_{fm}) of the full model, and ii) the R_P^2 was estimated as $\boldsymbol{1}-\left(SSE_{fm}/SSE_{rm}\right)$. This QTL mapping procedure was implemented in the

software R (R Core Team, 2015). QTL and expression QTL mapping were performed for phenotypic data of aluminum tolerance (RNRG) and *SbMATE* expression (ΔΔCt), respectively, measured in the RIL population.

Genome Wide Association Mapping

Genotypic data generation based on GBS and the procedures for imputation and data processing were the same as described for QTL mapping in RILs, except for data filtering, which in the case of the association panel eliminated SNPs with MAF < 0.05 and minimum count was 80%. GWAS was performed for both aluminum tolerance (RNRG) and *SbMATE* expression (ΔΔCt) data. A Q + K compressed mixed linear model (15, 16) was used, where Q corresponds to a population structure assignment matrix of dimension 254 x 6 that was previously estimated (17) based on 38 loci SSR. The familial relatedness or kinship (K) matrix depicting similarity between individuals (12) was estimated with TASSEL v4.0. In this matrix, each element dij is equal to the proportion of the SNPs that are different between taxon i and taxon j. The distance matrix was converted into a similarity matrix by subtracting all values from 2 and then scaled so that the minimum value in the matrix is 0 and the maximum value is 2. Association analysis was undertaken with 233,843 SNPs randomly distributed in the sorghum genome. SNPs located within the Alt_{SB} locus where *SbMATE* resides and within *SbMATE* itself, which were found to be highly associated with Al tolerance (3), were included in the genotypic dataset. GWAS was performed with TASSEL v3.0 (12) with the P3D algorithm (16) to reduce computing time.

Transactivation Assays

Plant materials and growth conditions

Arabidopsis thaliana plants were grown in soil under a 10-h-day/14-h-night cycle at 22^{o}C in a growth chamber. The light intensity was set to 100 µmol m^{-2} s^{-1} under 75% relative humidity.

Promoter constructs

The sequences of the MITE repeats for each allele were previously reported in Magalhaes et al. (Fig. 3*E* and S3) (18). Additionally, 1,749-bp of promoter DNA upstream of *SbMATE* to one of the flanking markers previously used to clone *SbMATE* (the 7-bp indel, Fig. 1 in Magalhaes et al. (18)), and a 2,010-bp fragment between the MITE repeats and the *SbMATE* start codon were also synthesized based on the sequence of the sorghum BAC 181g10 that harbors *SbMATE* (18).

For transactivation assays, we prepared commercially synthesized (Epoch Life Science), full-length *SbMATE* promoters from the sorghum lines SC283, BR012 and Tx430 fused to the reporter gene, GUS, as well as different extensions of the *SbMATE* promoter amplified from the Al-tolerant sorghum line, SC283, which were also fused GUS.

Deletions of *SbMATE* 5' flanking sequences were obtained by PCR using SC283 DNA as template and *SbMATE* promoter-specific primers harboring EcoRI and HindIII sites, which anneal at different positions of the *SbMATE* promoter (Table S2). The amplified promoter fragments were cloned into PCR TOPO8 and transferred by recombination to the pMDC162 vector for transactivation assays in Arabidopsis protoplasts. The resulting clones were designated -2102pSbMATE-SC283::GUS, -2010pSbMATE-SC283::GUS, -

1975pSbMATE-SC283::GUS, -1360pSbMATE-SC283::GUS, -536pSbMATE-SC283:GUS, -1975to-536pSbMATE-SC283::GUS, -1360to-536pSbMATE-SC283::GUS, and contain the *SbMATE* 5′-flanking sequences delimited by the positions indicated by their numbers until the *SbMATE* translation start codon (+1). Fig. S5*B* depicts the *SbMATE* promoter deletions.

Synthetic full-length *SbMATE* promoters of the SC283 (aluminum tolerant), Tx430 and BR012 (sensitive) lines were cloned into pBlueScriptIISK (pBSK). Promoter fragments were released from the pBSK vector by digestion with EcoRI and SmaI, and cloned into the yeast expression pHIS2.1 vector (Clontech) for transactivation assays in yeast. The resulting clones contain the reporter gene *HIS3* under the control of the *SbMATE* promoter and were designated -5666pSbMATE-SC283 (SC283p), -5299pSbMATE-BR012 (BR012p), -4214pSbMATE-Tx430 (Tx430p). Two deletions of the SC283 promoter were also tested for transactivation in yeast, -2102pSbMATE-SC283 (-2102pSC283, cloned with EcoRI and HindIII) and -2010pSbMATE-SC283 (-2010pSC283, cloned with EcoRI and SmaI). This nomenclature followed the same system used for promoter deletion clones. Fig. S2*C* illustrates the full-length *SbMATE* promoters from the different sorghum lines. Full-length *SbMATE* promoters containing four and one MITE repeats, corresponding to the *SbMATE* promoter region of BR012 (-5299pBR012::GUS) and Tx430 (-4214pTx430::GUS), respectively, were also cloned into pCAMBIA1381Z with EcoRI and SmaI, which was pre-digested with the same enzymes, and tested for transactivation in Arabidopsis protoplasts.

GUS transactivation assay in Arabidopsis protoplasts

The Columbia (Col-0) ecotype of *A. thaliana* was transformed with the recombinant plasmids using *Agrobacterium tumefaciens* strain GV 3101-mediated floral dip method (19). Homozygous lines were selected for GUS transactivation analyses.

The sorghum genes *Sb09g022810* (*Sorbic.009G166200*, *SbNFY1*), *Sb09g023500* (*Sorbic.009G174300*, *SbWRKY1*) and *Sb09g021530* (*Sorbic.009G151400*, *SbZNF1*) flanked by half of the AttB1 and AttB2 recombination sequences were commercially synthesized and cloned into the pBSK vector. Transcription factor cDNAs were amplified by PCR using oligonucleotides specific for the recombination sequences, AttB1-Foward (GGGGACAAGTTTGTACA) and AttB2-Reverse (GGGGACCACTTTGTACA), cloned into pDONR201 and transferred to the pEarlyGate104 expression vector. The resulting clones, designated *35S-SbNFY1*, *35S-SbWRKY1* and *35S-SbZNF1*, harbor the respective cDNA fused to the N-terminus of YFP under the control of the 35S promoter.

For *GUS* reporter expression analyses, protoplasts were prepared from Col-O and from transgenic lines harboring the *SbMATE* promoter-GUS fusions, as described by Niyogi (20) with modifications. Briefly, the protoplasts were isolated from 15-day-old *A. thaliana* leaves by digestion for 6 h, under agitation at 80 rpm and 30 ^{o}C, with 10-15 mL of 1% (v/v) Cellulase, 0.2% (v/v) Macerozyme, 0.4 M Mannitol, 20 mM KCl and 20 mM MES pH 5.7. The digested leaves were filtered through a 65 µM nylon mesh, the pellet was washed twice in 10 mL washing solution (154 mM NaCl, 125 mM $CaCl_2$, 5 mM KCl and 2 mM MES) and resuspended in 1 mL electroporation buffer (0.6 M

Mannitol, 10 mM KCl, 15 mM $MgCl_2$ and 25 mM MES, pH 7.2). Transactivation assays were performed by electroporation (250 V, 250 F) of 10 µg of expression cassette DNA, 30 µg of sheared salmon sperm DNA into 2×10^5 to 5×10^6 protoplasts in a final volume of 0.8 ml. Protoplasts were diluted into 3 ml of 0.4 M Mannitol, 15 mM $MgCl_2$ and 4 mM MES pH 5.5. After 12 h under light followed by 24 h in the dark, the protoplasts were washed with 0.4 M Mannitol, 15 mM $MgCl_2$ and 4 mM MES pH 5.5, and frozen in liquid N_2 until use.

Protein extraction and fluorometric assays for GUS activity were performed essentially as described by Jefferson et al. (21) with methylumbelliferone (MU) as a standard. For the standard assay, leaf discs were ground in 0.5 mL of GUS assay buffer (100 mM $NaH_2PO_4 \cdot H_2O$ [pH 7.0], 10 mM EDTA, 0.1% [w/v] sarcosyl, and 0.1% [v/v] Triton X-100), and 25 µL of this extract were mixed with 25 µL of GUS assay buffer containing 2 mM of the fluorescent 4-methylumbelliferone b-D glucuronide (MUG) as a substrate. The mixture was incubated at 37°C in the dark for 30 min, and GUS activity was measured using an InfiniteM200pro Microplate Reader (TECAN). The concentration of total proteins was determined with the Bradford method (22). The experiments were repeated four times with similar results.

Yeast one-hybrid assay

SbNFY1, *SbWRKY1* and *SbZNF1* cDNAs were amplified from commercially synthesized pBSK-derived clones using specific primers containing the restriction enzyme sites BamHI (forward) and XhoI (reverse), as listed in Table S2. The amplified fragments were double-digested with BamHI and XhoI and inserted into the yeast expression pGADT7-AD vector (Clontech),

creating a translational fusion between the GAL4 activation domain and the transcriptions factors.

The reporter yeast strains were constructed by transforming strain W303 (*MATa*/*MAT leu2-3,112 trp1-1 can1-100 ura3-1 ade2-1 his3-11,15* [*phi* $^{+}$]) with the *SbMATE* promoter reporter genes described in the “Promoter constructs” session using lithium acetate/polyethylene glycol method according to the manufacturer’s instructions (Clontech). The transformants were plated onto SD media (–Leu, –Trp) and incubated at 28°C for 4 days. The reporter yeasts were co-transformed with the expression cassettes *SbNFY1*, *SbWRKY1* or *SbZNF1* and plated on synthetic drop-out medium lacking Trp, Leu, Ura and His but supplemented with 10 mM 3-aminotriazole (3AT), and cultured for 3–5 days at 30°C.

Transcription Factor Effect on *SbMATE* Expression and Al Tolerance Via Haplotype Analysis in a RIL Population

SbWRKY1 and *SbZNF1* alleles from the RIL parents, SC283 and BR007, were sequenced and polymorphisms were identified: a 1-bp indel (+1A in BR007) in the second intron of *SbWRKY1* and a C (BR007) / G (SC283) SNP in *SbZNF1* intron 3. For genotyping in the RIL population, a 474-bp fragment spanning the *SbWRKY1* indel was amplified with SbWRKY1_2F and HEX-SbWRKY1_2R, digested with RsaI and resolved in the 3500xL Genetic Analyzer (Applied Biosystems). A 308-bp fragment containing the C/G SNP in *SbZNF1* was amplified with SbZNF1_2F and SbZNF1_2R and digested with BstuI. This generated a cleaved amplified polymorphic sequence (CAPS) marker that was resolved in a 2% agarose gel.

SbWRKY1 genotyping

PCR reactions were carried out in a final volume of 15 µL containing 30 ηg of DNA, 1X reaction buffer (Kapa Biosystems), 100 µM of each dNTP, 0.3 µM of each primer (SbWRKY_2F and HEX-SbWRKY_2R, Table S2), 0.5 U of Taq DNA polymerase (Kapa Biosystems). PCR thermal cycling conditions were initial denaturation at 95°C for 2 min, followed by 35 cycles of denaturing at 94°C for 30 sec, annealing at 58°C for 30 sec, extension at 72°C for 30 sec, with a final extension at 72°C for 5 min. Then, 5 µL of PCR products was digested with 1x buffer (New England Biolabs), 0.1 µg/µL BSA and 2 U of RsaI (New England Biolabs), in a total volume of 15 µL. All reactions were incubated at 37°C for 4 h and diluted 20-fold with sterile water. One µL of the diluted product was mixed with 0.2 µL of GeneScan ROX-500 size standard (Applied Biosystems and 8.8 µL of deionized formamide for resolving on a 3500xL Genetic Analyzer (Applied Biosystems). GeneMapper v5.0 (Applied Biosystems) was used to determine allele sizes.

SbZNF1 genotyping

PCR reactions were carried out in a final volume of 15 µL containing 30 ηg of DNA, 1X reaction buffer containing 0.8 mM $MgCl_2$ (Kapa Biosystems), 100 µM of each dNTP, 0.3 µM of each primer (SbZNF1_2F and SbZNF2_2F, Table S2), 0.5 U of Taq DNA polymerase (Kapa Biosystems). PCR thermal cycling conditions were 95°C for 2 min, followed by 40 cycles at 94°C for 30 sec, 55°C for 30 sec, 72°C for 30 sec, with a final extension at 72°C for 5 min. Then, 5 µL of PCR products were digested with 1x buffer and 2 U of BstuI (New England Biolabs), in a total volume of 10 µL. All reactions were incubated at 60°C for 4 h and digested PCR products were electrophoresed in a 2% (w/v) agarose gel at

100 V for 2 h. Upon digestion, a 308-bp fragment amplified with SbZNF1_2F and SbZNF1_2R and containing the C/G SNP in *SbZNF1* generated a cleaved amplified polymorphic sequence (CAPS) marker. The SC283 allele at the SNP locus was cleaved, yielding fragments of 81- and 227-bp, whereas the uncleaved, BR007-derived fragment was 308-bp.

Upon genotyping with *SbZNF1* (Z) and *SbWRKY1* (W) markers, RILs homozygous for the SC283 (Al-tolerant) alleles at both TF loci (Z+/W+), for the homozygous for the BR007 allele (Z-/W-), or showing alternate TF alleles (Z+/W- and Z-/W+) were identified. Ten RILs carrying each of the four derived haplotypes (i.e. combinations between TF alleles) of *SbZNF1*/*SbWRK1* were assessed for *SbMATE* expression and Al tolerance (RNRG) after 5 days in nutrient solution with +/- {27} µM Al^{3+} at pH 4.0. The mean *SbMATE* expression in RILs fixed for the SC283 allele at both TF loci (Z+/W+) was used as a reference to build orthogonal contrasts with the remaining haplotype classes. A mixed linear model considering genotype (RIL) as random and *SbZNF1* – *SbWRKY1* haplotype as fixed effect was fitted to *SbMATE* expression data (ΔΔCt). Model comparisons, confidence intervals and linear regression analysis were undertaken with the nlme (23) and lme4 (24) libraries in R. Linkage mapping of the transcription factors was undertaken with MapMaker EXP 3.0 (25).

Chromatin Immunoprecipitation (ChiP) Assay

Leaf protoplasts were isolated from *Arabidopsis thaliana* transgenic plants transformed with pMDC162 containing either an SC283 promoter fragment extending to position -2,102-bp relative to the *SbMATE* start codon

[designated -2102promoter(p)] or with pCAMBIA1381Z containing a promoter fragment isolated from the sorghum line, BR012, extending to position -5,299-bp [-5299(p)] depicted in Fig. 3 *A* and *B*. The protoplasts were then transformed with the *35S::YFP::SbWRKY1* and *35S::YFP::SbZNF1* vectors, as described in the methods for GUS transactivation assays in Arabidopsis protoplasts. Chromatin immunoprecipitation was undertaken with 1 µL of anti-GFP Polyclonal Antibody (Life Technologies - A11122; diluted 3-fold) as described in (26), or with the same volume of pre-immune IgG from rabbit used as a control.

ChIP Quantitative PCR (ChIP-qPCR)

ChiP-qPCR primers (Table S2) were first used to amplify both input and immunoprecipitated (IP) DNA to check for specific amplification of the target templates. For quantification purposes, 10 µL PCR reactions were prepared with 1 µL of either IP or input DNA diluted 3- and 10-fold in ultrapure water, respectively, 0.2 mM of each primer and 5 µL of Fast *SYBR Green PCR 2x Master mix* (Applied Biosystems).

PCR thermal cycling conditions consisted of an initial denaturation step at 95°C for 10 min and 40 cycles of denaturing at 94°C for 15 sec followed by annealing and extension at 60°C for 60 sec. Ct values of IP samples were normalized based on the input Ct. Actin (AT3G53750) was used as an endogenous control. Finally, fold enrichment over the negative control (pre-immune IgG) were calculated using the $\Delta\Delta$Ct method (8).

For statistical analysis, the Shapiro-Wilk Normality Test ($\alpha = 0.05$) was first undertaken with the R package, ExpDes.pt (27). For normally distributed data, an analysis of variance was followed by mean comparisons using the

Tukey test (α = 0.05). Otherwise, the nonparametric relative contrast effects [Nparcomp (28)] test with the same significance threshold was applied.

In silico Identification of *cis*-Elements

In silico identification of *cis*-elements within the binding sequence was undertaken with MatInspector (29).

Literature Cited

1. Hill WG, Robertson A (1966) Linkage disequilibrium in finite populations. *Theor Appl Genet* 38(6):226–231.
2. Eulgem T, Rushton PJ, Robatzek S, Somssich I (2000) The WRKY superfamily of plant transcription factors. *Trends Plant Sci* 5(5):199–206.
3. Caniato FF, et al. (2014) Association mapping provides insights into the origin and the fine structure of the sorghum aluminum tolerance locus, *Alt$_{SB}$*. *PLoS One* 9(1):e87438.
4. Melo JO, et al. (2013) Incomplete transfer of accessory loci influencing *SbMATE* expression underlies genetic background effects for aluminum tolerance in sorghum. *Plant J* 73(2):276–288.
5. Monique D, Rattunde F, Chantereau J (2006) A global view of genetic diversity in cultivated sorghums using a core collection. Genome 49(2):168–180.
6. Magnavaca R, Gardner C, Clark R (1987) Inheritance of aluminum tolerance in maize. *Genetic Aspects of Mineral Nutrition*, eds Gabelman H, Loughman B (Martinus Nijhoff Publishers, Dordrecht), pp 201–212.
7. Parker D, Norvell W, Chaney R (1995) GEOCHEM-PC: a chemical

speciation program for IBM and compatible personal computers. *Chemical Equilibrium and Reaction Models. Soil Science Society of America*, ed Loeppert R (Soil Science Society of America, Madison), pp 253–269. 42nd Ed.

8. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* 25(4):402–408.
9. Saghai-Maroof MA, Soliman KM, Jorgensen RA, Allard RW (1984) Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics. *Proc Natl Acad Sci U S A* 81(24):8014–8.
10. Elshire RJ, et al. (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6(5):e19379.
11. Roberts A, et al. (2007) Inferring missing genotypes in large SNP panels using fast nearest-neighbor searches over sliding windows. *Bioinformatics* 23(13):401–407.
12. Bradbury PJ, et al. (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23(19):2633–2635.
13. Anderson MJ, Ter Braak CJF (2003) Permutation tests for multi-factorial analysis of variance. *J Stat Comput Simul* 73(2):85–113.
14. Kao CH, Zeng ZB, Teasdale RD (1999) Multiple interval mapping for quantitative trait loci. Genetics 152(3):1203–16.
15. Yu J, et al. (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38(2):203–208.
16. Zhang Z, et al. (2010) Mixed linear model approach adapted for genome-

wide association studies. Nat Genet 42(4):355–360.

17. Caniato FF, et al. (2011) The relationship between population structure and aluminum tolerance in cultivated sorghum. *PLoS One* 6(6):e20830.
18. Magalhaes JV, et al. (2007) A gene in the multidrug and toxic compound extrusion (MATE) family confers aluminum tolerance in sorghum. *Nat Genet* 39(9):1156–1161.
19. Clough SJ, Bent AF (1999) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J* 16(6):735–743.
20. Niyogi K (2005) Arabidopsis protoplast isolation and transformation. Available at: bsw3.naist.jp/plantglobal/protocol/protoplast.pdf.
21. Jefferson RA, Kavanagh TA, Bevan MW (1987) GUS fusions: β-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J* 6(13):3901–3907.
22. Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72:248–254.
23. Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team (2017) {nlme}: Linear and nonlinear mixed effects models. Available at: https://cran.r-project.org/package=nlme.
24. Bates D, Mächler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using {lme4}. *J Stat Softw* 67(1):1–48.
25. Lander ES, et al. (1987) MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* 1(2):174–181.

26. Yamaguchi N, et al. (2014) PROTOCOL: Chromatin Immunoprecipitation from Arabidopsis Tissues. Arab B 12:e0170.

27. Ferreira EB, Cavalcanti PP, Nogueira DA (2013) Experimental Designs package. Available at: http://cran.r-project.org/package=ExpDes.pt.

28. Gao X, Alvo M, Chen J, Li G (2008) Nonparametric multiple comparison procedures for unbalanced one-way factorial designs. J Stat Plan Inference 138(8):2574–2591.

29. Cartharius K, et al. (2005) MatInspector and beyond: Promoter analysis based on transcription factor binding sites. Bioinformatics 21(13):2933–2942.

30. Machens F, Becker M, Umrath F, Hehl R (2014) Identification of a novel type of WRKY transcription factor binding site in elicitor-responsive *cis*-sequences from *Arabidopsis thaliana*. *Plant Mol Biol* 84(4–5):371–385.

SUPPLEMENTARY TABLES

Table S1: Multi-locus QTL mapping in RILs. QTLs and eQTLs identified using a multi-locus regression model for Al tolerance (relative net root growth, RNRG) and *SbMATE* expression (relative quantification, RQ) in the BR007 x SC283 RIL population.

Trait	QTL	Chr.	Physical Position (pb)	Effect	p -value	R_P^2
RNRG	1	3	71,125,567	16.73	2.94 x 10^{-15}	14.85
	1 x 2	3 x 9	71,125,567 x 48,732,977	4.66	1.63 x 10^{-03}	2.53
R_T^2						**37.90**
RQ	1	3	71,022,503	2.51	6.76 x 10^{-25}	43.75
	2	9	57,158,462	0.71	5.22 x 10^{-04}	6.31
	1 x 2	3 x 9	71,022,503 x 57,158,462	0.42	1.74 x 10^{-03}	5.17
R_T^2						**75.68**

Table S2: Primer and probe sequences

Experiment	Primer/Probe	Primer/Probe sequence (5´- 3´)	Application
SbMATE expression	SbMATE_1F	CAGCCATTGCCCATGTTCTTT	Global (A+T) expression: forward primer (TaqMan)
	SbMATE_1R	ACCAGCTTGCTCAGCATTATCA	Global (A+T) expression: reverse primer (TaqMan)
	Probe A+T	6-FAM-CCCAGTACCTGATAACGC-MGB	Global (A+T) expression: probe
	SbMATE_2F	AAGAAATCAGGGAACAGGAGGTT	Allele-specific expression: forward primer (TaqMan)
	SbMATE_2R	ACGAGGAAGACGGTCTGGAA	Allele-specific expression: reverse primer (TaqMan)
	Probe A	6-FAM-CAGG**T**GCGCGCCCACGATC-TAMRA	Allele-specific expression: probe hybridizes to the “A” allele donated by SC566
	Probe T	6-FAM-CAGG**A**GCGCGCCCACGATCA-TAMRA	Allele-specific expression: probe hybridizes to the “T” allele present in all other lines
	SbMATE_3F	GCCATTGCCCATGTTCTTTC	Temporal and tissue expression: forward primer (SYBR)
	SbMATE_3R	TCGACCTTCCTCACCAGCTT	Temporal and tissue expression: reverse primer (SYBR)
	18S_F	AATCCCTTAACGAGGATCCATTG	*18S rRNA* endogenous reference: forward primer (SYBR). It was also used in transcription factor assay
	18S_R	CGCTATTGGAGCTGGAATTACC	*18S rRNA* endogenous reference: reverse primer (SYBR). It was also used in transcription factor assay
SbWRKY1 expression	SbWRKY1_1F	CGTGCAGCAGCAAAGCAA	*SbWRKY1* expression: forward primer (SYBR)
	SbWRKY1_1R	GTCGCAGGTATGCTCGTTGA	*SbWRKY1* expression: reverse primer (SYBR)

Experiment	Primer	Primer sequence (5´- 3´)	Application
SbZNF1 expression	SbZNF1_1F	GGTCCTCGGTCGAAAGTTGA	*SbZNF1* expression: forward primer (SYBR)
	SbZNF1_1R	TCCCACTTCTACATTAACATCCTCAT	*SbZNF1* expression: reverse primer (SYBR)
SbWRKY1 and *SbZNF1* genotyping	SbWRKY1_2F	GCATTGAATCGAATATAAATGTCA	*SbWRKY1* genotyping in the RIL population: forward primer
	SbWRKY1_2R	HEX-CCGGTCGACAGTATGAACA	*SbWRKY1* genotyping in the RIL population: reverse primer
	SbZNF1_2F	CAATGCACGCGAGACAAG	*SbZNF1* genotyping in the RIL population: forward primer
	SbZNF1_2R	AACCTTAGTTAAGAACAGTG	*SbZNF1* genotyping in the RIL population: reverse primer
Transactivation assays (promoter constructs)	PALTSB1975_F*	ccccgaattcGCATGGGGCCATCAACGGTG	SC283 promoter deletions: *-1975*
	PALTSB1_R**	ccccaagcttCATGTCGGCCTAGCTACAAACC	SC283 promoter deletions: *1*
	PATTSB2838_F*	ccccgaattcGGATCCAGTGAGCTACCGTACC	SC283 promoter deletions: *-2102*
	PALTSB536_F*	ccccgaattcCCCCGTCTTGCTCTTCAGCTCT	SC283 promoter deletions: *-536*
	PALTSB1360_F*	ccccgaattcGTCGTCACGTCGACCAAACTTTC	SC283 promoter deletions: *-1360*
	PALTSB536_R***	ccccaagcttAGAGCTGAAGAGCAAGACGGG	SC283 promoter deletions: *-536*

*EcoRI sites were included in all forward primers.
**Reverse primer contains the *SbMATE* start codon followed by the HindIII site.
***Reverse primer contains the HindIII site.

Experiment	Primer	Primer sequence (5´- 3´)	Application
Transactivation assays (cDNA constructs)	Sb09g022810BamHI_F	ggatcccgATGGCGGACGCTCCG	SbNFY1 cDNA amplification from pBSK: forward primer
	Sb09g022810XhoI_R	ctcgagTTA GTT TGA GAT ATC CCC ATT	SbNFY1 cDNA amplification from pBSK: reverse primer
	Sb09g023500BamHI_F	ggatccgtATG ACG CTG TCT CCA CCG CAG	SbWRKY1 cDNA amplification from pBSK: forward primer
	Sb09g023500XhoI_R	ctcgagCTA GTA CCC ATC ATA CGA AAA	SbWRKY1 cDNA amplification from pBSK: reverse primer
	Sb09g021530_F	ATG GCG CGT CCG CAG CGA GTG	SbZNF1 cDNA amplification from pBSK: forward primer
	Sb09g021530_R	CTA GTG AGT CTC TGC TAG AGA	SbZNF1 cDNA amplification from pBSK: reverse primer
Experiment	**Primer**	**Primer sequence (5´- 3´)**	**Annealing positions in the -2102 and 5299 promoters (-2012p and -5299p) relative to the *SbMATE* start codon**
Chromatin Immunoprecipiation (ChiP-qPCR)	F1	GGATCCAGTGAGCTACCGGTG	-2102p: -2102. -5299p: -2098, -2461, -2824, -3187, -3550.
	F2	AGCGACATGCCCTTTGACA	-2102p: -988. -5299p: -988.
	F3	GACCAATGTAGTGATGCTTAT	-5299p: 3785
	F4	ACGCATTTTCTCTGTAGTAGT	-5299p: 3589
	R1	TAGTCTGTAGGAGTACTCCTAGC	-5299p: -2393, -2756, -3119, -3482.
	R2	AATAAAGCAGCAAACGCCAACC	-2102p: -2011. -5299p: -2011.
	R3	CATTTCGAAAACATTGAGCTAGAG	-2102p: -923. -5299p: -923
	R4	CTGTAGGAGTATAGATGCCGC	-2102p: -2044. -5299p: -2044, -2403, -2766, -3129, -3492
	R5	TGAGTCAAGTGACACTACCGTTG	-2102p: -1963. -5299p: 1963
	ActinF	ATGTCGTGAGCCATCCTGTC	Background control
	ActinR	ACACCGGATTCGTGCGGCAT	Background control

Table S3*: In silico* analysis of *cis* elements in the 92 bp binding fragment of the *SbMATE* promoter using MatInspector (29). Position (from-to) indicates the nucleotide positions where the motifs were found in the 92 bp sequence (shown on top). Str.: DNA Strand. Matrix sim. (similarity) (29). The analysis was carried out based on the 92 bp sequence (black) plus eight nucleotides downstream (red).

92 bp region of the *SbMATE* promoter

GGATCCAGTGAGCTACCGGTGAAGGTGCTCGTTATGCGTTTAAACATTGTTCCGTCCGGCGGCATCTATACTCCTACAGACTATTAAAGTTGGTTGGCGT

1 30 60 92

ID	Detailed Matrix Information	Position (from	to)	Str.	Matrix sim.	Sequence
TRIH	Arabidopsis 6B-interacting protein 1-like 2 (AT3G14180)	11	25	(+)	0.891	agctaCCGGtgaagg
EREF	WRINKLED 1 (Ethylen respone element factors)	42	60	(-)	0.861	gCCGGacggaacaatgttt
SBPD	Squamosa promoter-binding-like protein 1	46	62	(-)	0.861	ccgccGGACggaacaat
SBPD	Squamosa promoter-binding-like protein 1	49	65	(+)	0.839	gttccGTCCggcggcat
TRIH	Trihelix-domain transcription factor VFP5 (AT5G05550)	51	65	(+)	0.904	tccgtCCGGcggcat
LBDF	Lateral organ boundaries (LOB)	53	73	(+)	0.935	cgtcCGGCggcatctatactc
LREM	RAP2.2, involved in carotenoid and tocopherol biosynthesis and in the expression of photosynthesis-related genes (Light responsive element motif)	62	72	(+)	0.873	gcATCTatact
VTBP	Avian C-type LTR TATA box	79	95	(+)	0.831	gactattAAAGttggtt
TOEF	Target of early activation tagged 1 (RAP2.7) (tertiary DNA binding preference)	82	92	(-)	0.871	caacTTTAata
DOFF	Dof1 / MNB1a - single zinc finger transcription factor	77	99	(+)	0.987	cagactattAAAGttggttggcg

Potential transcription factor binding sites were predicted using MatInspector (29) Release professional 8.4.1.
A sequence similar to the WT-box (YGACTTTT) reported by Machens et al (2014) (30), is located at 78 – 85 on the (+) strand.