Supporting Information

Tang et al. 10.1073/pnas.1305213110

SI Materials and Methods

Genetic Mapping of the *Sh1* Locus in *Sorghum bicolor* × *Sorghum propinquum* F_2 Population. Initial genetic mapping of *Sh1* has been described (1) in a population of 370 F_2 individuals (740 informative gametes) (Fig. 1). Recombinants in the region were validated by testing of F_3 progeny rows and determination of genotype at the genetically dominant *Sh1* locus based on progeny segregation ratios.

Sequencing, Assembly, and Annotation of *S. propinquum* Bacterial Artificial Chromosomes. DNA of selected bacterial artificial chromosomes (BACs) was isolated, sheared, end-repaired into subclones, and sequenced in an ABI 3730 capillary sequencer, and reads were pooled and assembled using Celera Assembler version 6.1 (2). The corresponding regions in *S. bicolor* and *S. propinquum* were aligned using MUMMER version 3.0 (3). Gene structures in the *S. propinquum* shattering region were predicted using a combination of similarity-based gene prediction software GENEWISE with the *S. bicolor* predicted genes (Sbi version 1.4) as the guide sequences and the ab initio prediction software FGENESH (www.softberry.com).

Compiling a Sorghum Diversity Panel for Mapping the Shattering Trait. A diversity panel of sorghum varieties suitable to study shattering were collected from Cornell University (S. Kresovich and M. Hamblin) and the USDA-ARS germplasm collection, selected to represent a wide range of geographical locations (Table S1) to minimize relatedness among genotypes and reduce false-positive associations in association mapping. The shattering phenotype for each accession was carefully validated. A simple but qualitative method is to classify phenotypes as shattering or nonshattering by vigorously shaking panicles. As a quantitative measurement, breaking tensile strength (BTS, measured in grams) was measured (4) using a digital force gauge (DPS-4; IMADA) to clasp onto the grain and measure the force required to break the pedicel while pulling the grain away from the stalk. BTS values were recorded at different developmental stages, and stable values (after maturity of the grains) were used to distinguish the shattering/nonshattering phenotype for each variety (Fig. 2*A*). Shattering varieties were often easier to distinguish because they were deciduous once the grains mature, whereas the nonshattering varieties needed to be monitored for a longer period. The breaking force for nonshattering varieties often stabilized around 50 *g* of force after maturity, whereas the force for the shattering varieties went down to small values around 10 *g*, with extreme cases requiring virtually 0 *g* of breaking forces, i.e., capable of dispersal with external force too small to measure. The final distributions of the mature BTS for the genotypes are therefore bimodal in the diversity panel. We used 25 g of mature BTS as a cutoff to distinguish the shattering or nonshattering genotypes.

Resequencing and Analyses of Polymorphic Sites Within the Region. Primers of 20–22 bp that amplify 700- to 1,000-bp amplicons were designed around the polymorphic sites of the candidate loci using PRIMER3 (5). DNA was prepared from young leaves of individual plants. PCR reactions of 15 μL per well were set up to amplify sampled regions using the following thermal cycling program: 95 °C for 30 s, 58 °C for 30 s, and 72 °C for 1 min for a total of 36 cycles and then 72 °C 10 min. The concentrations of the PCR amplicons were verified in 1% agarose gel, and excessive primers and dNTPs in the PCR reactions were removed using exo-nuclease I and shrimp alkaline phosphatase enzymatic digestion. Excessive primers and dyes in the sequencing reactions were removed using Sephadex columns before the sequencing plates were loaded onto ABI3730 capillary sequencer. The chromatograms were examined using SEQUENCHER software (version 4.1; GENECODES), and the polymorphisms were recorded in an EXCEL spreadsheet (Dataset S2). From each PCR amplicon sequence, we retained only the informative SNPs (tagging SNPs that are sufficient to reconstruct haplotype blocks), based on the observation that polymorphic sites within the same amplicon often show complete linkage disequilibrium (LD). PCR amplicons were sequenced with the DNA from the 24 accessions in the compiled shattering panel. The public genome sequence of sorghum is from a nonshattering inbred cultivar *S. bicolor* BTX623 (6); therefore, a total of 25 different genotypes are available to be compared and statistically tested. A simple general linear model (GLM) was used for testing association: $y = marker + e$, where y is the phenotype (0 for nonshattering, 1 for shattering). Because only a specific target region was searched, the risk of false-positive associations is much less than for a genomewide search, mitigating the need for inclusion of population structure parameters in the model. The association analyses were performed using TASSEL software (7).

RT-PCR. Plant materials were grown at the University of Georgia Plant Science Farm during the summer of 2011. *Sorghum halepense* genotype GRIF14527 (HAL) was chosen to represent the shattering category and *S. bicolor* genotype PI 658864 (BTX), a recombinant inbred line derived from a cross between BTx623 and IS3620C, was selected as a nonshattering type. Inflorescence was collected at different developmental stages by visual observation, i.e., still covered by flag leaf, just emerging from flag leaf, after anther dehiscence, and near maturity (Fig. 2 *A* and *B*). Tissue was harvested from two different individuals for each developmental stage, also collecting leaf samples to use as a control. Part of the inflorescence was used to measure BTS, and the remainder was flash frozen in liquid nitrogen and stored at −80 °C for later RNA isolation using the RNeasy plant mini kit (QIAGEN) according to the manufacturer's protocol. RNA was treated with RNase-Free DNase set (QIAGEN) to digest any residual genomic DNA and quantified using a UV spectrophotometer, with its quality and integrity confirmed on a 1% agarose gel prepared in RNase free 1× TAE buffer (Tris base, acetic acid and EDTA). First-strand cDNA was synthesized from 1 μg of total RNA using SuperScript III reverse transcriptase (Invitrogen) with 500 ng anchored oligo (dT) primers in a 20-μL reaction incubated at room temperature for 5 min before 2 h cDNA synthesis at 50 °C and 15 min at 70 °C. After cDNA synthesis, 20 μL sterile double-distilled water was added to the reaction. Each PCR consisted of 1 μL cDNA in a 20-μL reaction with 4 μL 5× GoTaq green reaction buffer, 2 μL 2 mM dNTP mix, 0.5 μL each primer (10 μM), and 0.5 Units of GoTaq DNA polymerase (Promega). The thermal profile consisted of incubation at 95 °C for 4 min, followed by 35 cycles at 95 °C for 45 s, 58 °C for 45 s, 72 °C for 45 s, and a final extension at 72 °C for 5 min. A *Sorghum* actin gene (*SbActin*) was used as an internal control, based on the following primer sequences: forward, 5′-acattgccctggactacgac-3′; reverse, 5′-aatgaaggatggctggaaga-3′.

Transformation and Confirmation of Transgene Events. Candidate genes in the high association region (*SbWRKY*, *SbTATA*) were isolated from the BAC YRL20H16, cloned in pZP211, and

transformed into nonshattering RTx430 via *Agrobacterium tumefaciens* .We confirmed the presence of the shattering allele in transformants using two pairs of primers that span the first intron, which is longer in *S. propinquum* than the corresponding sequence in *S. bicolor*. We used stringent annealing temperature and 40 PCR cycles. Among the transgenics tested, one transgenic event showed a single *S. bicolor*–sized band and therefore seemed to not be transformed (Fig. S1).

Tissue Fixation and Microscopy. We examined the microscopy of the sorghum floret from shattering sorghum (HAL; *S. halepense*) and nonshattering sorghum (BTX; *S. bicolor*) using four time series, from T1 (youngest) to T4 (fully mature). Plant tissues were fixed in standard formalin-acetic acid-alcohol (FAA) solution for 5 d and dehydrated gradually by transferring into 50%, 70%, 90%, and 99% ethanol successively for a minimum of 1 d in each. Samples were then paraffin embedded for 1-μm microtome sectioning. Thin sections were placed on glass microscope slides. Unstained images were captured by a AxioCam MRc5 camera attached on an Axioskop-2 light microscope. Total lignin was visualized on an Eclipse 80i fluorescent light microscope equipped with epifluorescence optics and Nikon UV filter (UV-2E/C). Images were captured with a Nikon DS-Ri1 camera body using NIS-Elements Basic Research software. All UV images were taken by an 800-ms exposure in auto setting. Overlapping images were stitched to create larger panoramic pictures using the Adobe Photoshop CS5 photomerge function.

1. Paterson AH, et al. (1995) Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. *Science* 269(5231):1714–1718.
2. Myers EW, et al. (2000) A whole-genome assembly of Drosophila. *Science* 287(5461): 2196–2204.
3. Kurtz S, et al. (2004) Versatile and open software for comparing large genomes. *Genome Biol* 5(2):R12.
4. Konishi S, et al. (2006) An SNP caused loss of seed shattering during rice domestication. *Science* 312(5778):1392–1396.
5. Koressaar T, Remm M (2007) Enhancements and modifications of primer design program Primer3. *Bioinformatics* 23(10):1289–1291.
6. Paterson AH, et al. (2009) The sorghum bicolor genome and the diversification of grasses. *Nature* 457(7229):551–556.
7. Bradbury PJ, et al. (2007) TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* 23(19):2633–2635.

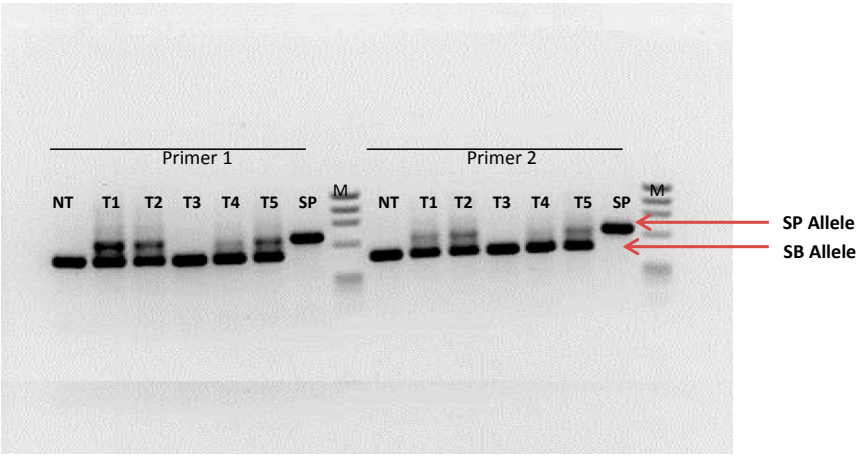

Fig. S1. Verification of the transgenic genotype. NT, nontransgenic; T1 to T5, transgenic individuals; SB, *S. bicolor*; SP, *S. propinquum*; M, marker.

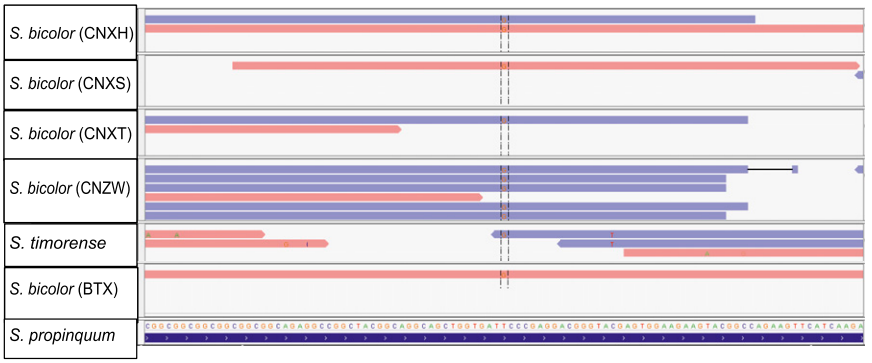

Fig. S2. Alignments of reads from sorghum resequencing data around the start codon position of *Sb01g012870*. Several wild *S. bicolor* races [CNXH, CNXS, CNXT, and CNZW; corresponding to *S. bicolor* ssp verticilliflorum race virgatum (unnamed), drummondii (IS 21691), arundinaceum (IS 14216), and aethiopicum (IS 27587), respectively] and a genomic sequence from *S. timorense* are included. The polymorphic site within the start codon location is highlighted in the middle of the panel, showing the reads are mostly ATG (*S. bicolor*-like) instead of ATT (*S. propinquum*, bottom track, which is used as the reference in the read mapping).

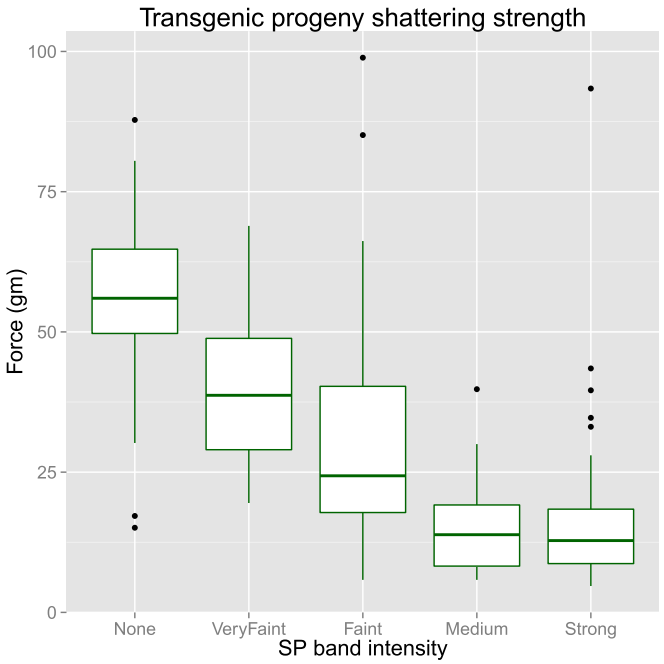

Fig. S3. Transgenic progeny shattering strength estimated with breaking tensile strength. The copy number of the introduced SP allele in the transgenic progeny is assessed with the intensity of the SP band on the agarose gel, which are scored into five discrete classes (none, very faint, faint, medium, and strong). The box of each group depicts lower quartile, median, and upper quartile. Dots, if any, indicate which observations that are considered outliers.

Table S1. Sorghum accessions selected in the shattering diversity panel

Accession ID	PGML index	Race	Origin
Complete shatterers (11 varieties)			
PI 267436	BP03 (#5)	*bicolor*	India
PI 569834	BP10 (#6)	*bicolor*	Sudan
PI 521356	BP06 (#7)	*drummondii*	Kenya
PI 365024	BP05 (#8)	*verticilliflorum*	South Africa
L-WA 27	AL03 (#10)	*verticilliflorum*	Angola
L-WA 23	AL02 (#11)	*verticilliflorum*	Angola
L-WA 13	AL01 (#12)	*verticilliflorum*	Sudan
PI 155675	BP01 (#15)	*bicolor*	Malawi
S. propinquum	SP (#20)	*S. propinquum*	—
KFS (deciduous mutant)	KFS (#21)	*bicolor*	United States
PI 570917	BP11 (#22)	*bicolor*	Sudan
Nonshatterers (13 varieties)			
PI 221607	AP02 (#1)	*bicolor*	Nigeria
PI 302115	BP04 (#2)	*verticilliflorum*	Australia
PI 152702	AP01 (#3)	*bicolor*	Sudan
NSL 87902	AN07 (#4)	*bicolor*	Cameroon
NSL77217	AN05 (#9)	*bicolor*	Chad
NSL56003	AN03 (#13)	*bicolor*	Kenya
NSL56174	AN04 (#14)	*bicolor*	Ethiopia
PI 267408	AP03 (#16)	*bicolor*	Uganda
PI 563146	BP07 (#17)	*bicolor*	Sudan
PI 267539	AP04 (#18)	*bicolor*	India
PI 563474	BP09 (#19)	*bicolor*	United States
PI 591385	BP13 (#23)	*bicolor*	India
PI 584089	BP12 (#24)	*bicolor*	Uganda

There are three accessions that did not flower. In the PGML index column, accessions with prefix (AL, AN, and AP) are from Cornell and accessions with prefix BP are from USDA-ARS. Race information was taken from the accompanying documentations shipped with the samples. PGML, Plant Genome Mapping Lab.

Other Supporting Information Files

Dataset S1 (XLS)
Dataset S2 (XLS)
Dataset S3 (DOCX)