Seed shattering in a wild sorghum is conferred by a locus unrelated to domestication

Haibao Tang, Hugo E. Cuevas, Sayan Das, Uzay U. Sezen, Chengbo Zhou, Hui Guo, Valerie H. Goff, Zhengxiang Ge, Thomas E. Clemente, and Andrew H. Paterson

*Plant Genome Mapping Laboratory, University of Georgia, Athens, GA 30605; †Center for Genomics and Biotechnology, Fujian Agriculture and Forestry University, Fuzhou 350002, Fujian Province, China; ‡J. Craig Venter Institute, Rockville, MD 20850; §Tropical Agriculture Research Station, US Department of Agriculture-Agricultural Research Service, Mayaguez, Puerto Rico 00680; and Department of Agronomy and Horticulture, University of Nebraska–Lincoln, Lincoln, NE 68588

Suppression of seed shattering was a key step during crop domestication that we have previously suggested to be convergent among independent cereal lineages. Positional, association, expression, and mutant complementation data all implicate a WRKY transcription factor, SpWRKY, in conferring shattering to a wild sorghum relative, Sorghum propinquum. We hypothesize that SpWRKY functions in a manner analogous to Arabidopsis homologs that regulate cell wall biosynthesis genes, with low expression toward the end of floral development derepressing downstream cellul wall biosynthesis genes to allow deposition of lignin that initiates the abscission zone in the seed–pedicel junction. The recent discovery of a YABBY locus that confers shattering within Sorghum bicolor and other cereals validated our prior hypothesis that some parallel domestication may have been convergent. Ironically, however, the shattering allele of SpWRKY appears to be recently evolved in S. propinquum and illustrates a case in which the genetic control of a trait in a wild relative fails to extrapolate even to closely related crops. Remarkably, the SpWRKY and YABBY loci lie only 300 kb apart and may have appeared to be a single genetic locus in some sorghum populations.

Reduction of seed dispersal by “shattering” of the mature inflorescence is a critical step during the domestication of many cereal crops that dominate human calorie consumption. For example, in wild sorghums such as Sorghum propinquum and Sorghum halepense, shattering facilitates dispersal of sexual propagules, complementing dispersal of clonal propagules by subterranean rhizomes. Shattering is thought to have been strongly selected against during domestication because humans could more efficiently harvest grains that remain attached to the plant (1), and we have previously suggested that parallel domestications of divergent cereals may have involved mutations in some corresponding genes (2).

During plant development, the shattering of seeds involves the formation of a lignified abscission layer (or dehiscence zone) and is considered a process of programmed senescence (3). The tight association of shattering with the pedicel junc
tion and the reduced expression of floral (FLO) and abscission (ABI) genes towards the end of floral development (4) likely underpins the evolution of shattering in grasses (5). This recent research hints at a possible role of the WRKY transcription factor in the abscission zone of cereals (6), which would explain the parallel evolution of shattering in divergent cereals (7). However, our understanding of the genetic basis of shattering remains incomplete.

Recently, we showed that SpWRKY, a WRKY transcription factor from S. propinquum, is capable of conferring shattering to rice (Oryza sativa ssp.) (8). The expression of SpWRKY is strongly regulated during floral development (9), and its expression is negatively regulated by two conserved WRKY transcription factors, qSH1 and qSH2 (10, 11). Orthologous genes in other cereals have been implicated in pathological shattering (12), and alleles at the shattering locus have been identified in different species (13). However, the genetic basis of shattering is poorly understood (14).

Reduction of seed dispersal by “shattering” of the mature inflorescence is a critical step during the domestication of many cereal crops that dominate human calorie consumption. For example, in wild sorghums such as Sorghum propinquum and Sorghum halepense, shattering facilitates dispersal of sexual propagules, complementing dispersal of clonal propagules by subterranean rhizomes. Shattering is thought to have been strongly selected against during domestication because humans could more efficiently harvest grains that remain attached to the plant (1), and we have previously suggested that parallel domestications of divergent cereals may have involved mutations in some corresponding genes (2).

During plant development, the shattering of seeds involves the formation of a lignified abscission layer (or dehiscence zone) and is considered a process of programmed senescence (3). The tight association of shattering with the pedicel junction and the reduced expression of floral (FLO) and abscission (ABI) genes towards the end of floral development (4) likely underpins the evolution of shattering in grasses (5). This recent research hints at a possible role of the WRKY transcription factor in the abscission zone of cereals (6), which would explain the parallel evolution of shattering in divergent cereals (7). However, our understanding of the genetic basis of shattering remains incomplete.

Recently, we showed that SpWRKY, a WRKY transcription factor from S. propinquum, is capable of conferring shattering to rice (Oryza sativa ssp.) (8). The expression of SpWRKY is strongly regulated during floral development (9), and its expression is negatively regulated by two conserved WRKY transcription factors, qSH1 and qSH2 (10, 11). Orthologous genes in other cereals have been implicated in pathological shattering (12), and alleles at the shattering locus have been identified in different species (13). However, the genetic basis of shattering is poorly understood (14).

Reduction of seed dispersal by shattering of the mature inflorescence is a critical step during the domestication of many cereal crops that dominate human calorie consumption. For example, in wild sorghums such as Sorghum propinquum and Sorghum halepense, shattering facilitates dispersal of sexual propagules, complementing dispersal of clonal propagules by subterranean rhizomes. Shattering is thought to have been strongly selected against during domestication because humans could more efficiently harvest grains that remain attached to the plant (1), and we have previously suggested that parallel domestications of divergent cereals may have involved mutations in some corresponding genes (2).

During plant development, the shattering of seeds involves the formation of a lignified abscission layer (or dehiscence zone) and is considered a process of programmed senescence (3). The pathway involving the formation of the abscission layer is well characterized in some eudicot species. SHATTERPROOF (SHP) genes SHP1 and SHP2 have been shown to specify valve margin cell identities in Arabidopsis (4), with their expression reinforced through negative regulation from FRUITFULL (FUL) in valve development (5) and REPLUMLESS (RPL) in the replum (6). Genes JAG, FIL, and YAB3 promote the expression of FUL and SHP but are negatively regulated by RPL (7). However, the abscission layer contributing to seed shattering in Arabidopsis is located at the valve–replum boundary and does not correspond to that of most cereal species, which is at the base of the pedicel. Therefore, it remains questionable whether orthologous genes are directly responsible for the seed dispersal mechanisms of dicots and cereals, respectively.

The genetic control of shattering varies among cereals, in many cases involving several genes. Two major genes that contribute to shattering in rice (Oryza sativa ssp.) have been identified: qSH1 and qSH4, controlling 68% and 69% of the phenotypic variance in the studied crosses, respectively (8, 9). The fixation of qSH4 occurred very early in rice domestication with the domesticated allele occurring in both subspecies indica and japonica, whereas qSH1 is much more recent, present only within temperate japonica genotypes (10, 11). In wheat, quantitative trait loci (QTLs) responsible for nonbrittle rachis are located in the homeologous regions of chromosomes 3A (Br2), 3B (Br3), and 3D (Br1) (12, 13). Comparative mapping hinted that this part of the chromosomal regions might correspond to the orthologous region in barley, controlled by two tightly linked loci, Brl1 and Brl2, but do not appear to correspond to the region in other cereals (12, 13). Indeed, although some shattering genes do appear to be in corresponding (orthologous) chromosomal locations among cereals (2), many do not; therefore, it is hypothesized that there are multiple pathways responsible for seed dispersal in the grasses (14). Steady progress in rice notwithstanding, more rice genes that control shattering are known (2, 3) but have not yet been identified; therefore, the above hypothesis remains to be tested.

Sorghum appears to be a favorable species to investigate the genetic basis of shattering, because only one genetic locus, Sh1, explains 100% of the phenotypic variance in a cross between an elite S. bicolor breeding line (BTx623), and wild S. propinquum (2). The S. bicolor genome size is ~730 Mb, has a high-quality reference sequence (15), and shows similar composition and high levels of synteny and microcolinearity with maize and rice (16) despite ~50 My of divergence (17).

Recently, it was shown that a YABBY transcription factor near the Sh1 region contributes to seed shattering in crosses between different Sorghum bicolor breeding lines (18). There are three different mutations at the YABBY locus that resulted in either lower expression or produced a truncated transcript in the domesticated allele. The three different YABBY haplotypes suggested independent origins in different sorghum populations (18).

Here, we report that seed shattering of S. propinquum is conferred by a different locus than YABBY, specifically a WRKY transcription factor (SpWRKY). The expression of the WRKY gene appears to be modulated during floral and seed development and is implicated in the lignin deposition pathway. The three independent mutations in the YABBY locus validate our prior hypothesis that parallel domestications of divergent cereals may have involved mutations in some corresponding genes (2).
In comparison, the shattering gene that we report here appears recently evolved in *S. propinquum* and illustrates a case in which the genetic control of a trait in a wild relative fails to extend to closely related crops. Remarkably, the SpWRKY and YABBY loci lie only 300 kb apart and may have appeared as a single genetic locus in some sorghum populations. We further note that both genes might regulate some of the same downstream genes involved in lignin deposition and thus play important roles in the shattering pathway, however, with different relative contributions in different species.

**Results**

**Genetic Mapping of the Sh1 Locus and Genomic Alignments.** Sh1 was genetically mapped within a region flanked by two restriction fragment length polymorphism (RFLP) markers SOG0251 and SOG1273 (Fig. 1), separated by 0.42 cm (three recombinants out of 740 gametes). An *S. propinquum* bacterial artificial chromosome (BAC) library (19) was screened with overgo probes corresponding to the two flanking DNA markers, and hybridizing BACs were fingerprinted via restriction enzyme digestion. A physical contig map of uniform density (20) that spanned the two flanking markers included 10 overlapping *S. propinquum* BACs in a minimum tiling path. Sequence reads from the BACs were assembled into 30 contigs totaling 1.04 Mb (contig N50 = 63.9 kb). The *S. propinquum* BAC alignments correspond to a 1.13-Mb (11.33–12.47 Mb) region on *S. bicolor* chromosome 1. A total of 694 kb (61.3%) of this region can be aligned with the *S. bicolor* sequences. Polymorphic sites between the coding regions of the 80 genes on the basis of sequence divergence between *S. propinquum* and *S. bicolor* were tested to narrow down the target locus to be contained between base position 11,941,320 and 11,956,03 (chr1: 11.5–12.2 Mb). Four SNPs (P7E9, P3H11, P8F9, and P4C3) were significantly associated with shattering (*P < 0.001*), and two candidate genes fall inside this region: Sb01g012870 (SpWRKY) and Sb01g012880 (SpTATA).

**Association Mapping.** We identified a total of 603 SNP sites within the coding regions of the 80 genes on the basis of sequence alignment between *S. bicolor* reference and *S. propinquum* BAC sequences. Polymorphic sites between *S. bicolor* and *S. propinquum* within the Sh1 region were sampled using a diversity panel of 11 shattering and 13 nonshattering accessions (Table S1). We obtained data for a total of 67 sites, initially targeting an average of one site per gene. Among the 67 sites tested (Dataset S2), 4 were significantly (*P < 0.001*) associated with the shattering trait (P7E9, P3H11, P6F9, and P4C3; Fig. 1). The four sites were also in linkage disequilibrium with one another. However, the intermediate sites between the two peaks were not significantly associated with the shattering trait, possibly reflecting mutations that are of more recent origin than those related to shattering and therefore are not informative with regard to shattering.

The highest peak in the association plot contains two SNP markers, P7E9 (*P = 2.8 \times 10^{-5}* ) and P3H11 (*P = 2.2 \times 10^{-5}*) covering an ~50-kb genomic region. Additional primers were designed to sample more sequence sites in the ~50-kb genomic region, which extends from gene models Sb01g012870 to Sb01g012880 [gene identifiers from JGI annotation release Sbi1.4 (15)]. Based on the segregation pattern within this region (Dataset S2), we narrowed down the target locus to be contained between base position 11,941,320 and 11,956,003, because the level of association dropped to nonsignificance beyond these two flanking sites. This interval contains two candidate genes, encoding transcription factors Sb01g012870 (SpWRKY) and Sb01g012880 (SpTATA) (Fig. 1).

**Expression Profile of Candidate Genes.** We used semiquantitative RT-PCR to assay expression of the two candidate genes at different floral developmental stages in shattering and nonshattering genotypes, confirming the shattering phenotype using the breaking tensile strength (BTS) method. For each stage, 10 individual florets were tested from two different panicles. In nonshattering *S. bicolor*, the BTS value actually increased from 69.8 g in the immature inflorescence to 123.1 g after anther dehiscence and remained stable in the mature inflorescence (122 g; Fig. 2). The BTS value declined rapidly in shattering *S. halepense*, from 55.1 g
the SpWRKY transgene had an average BTS of 22.3 g (SD = 18.6, \( n = 105 \)), which is a highly significant 61% reduction from the nonshattering plant (Fig. 3).

The dominant SpWRKY allele conferring shattering in S. propinquum and the recessive SpWRKY in nonshattering S. bicolor BTx623 encode proteins that differ at two amino acid positions in the aligned portions. SpWRKY proteins cannot start at the S. bicolor start, because of the ATG to ATT mutation in SpWRKY that results in a methionine (M) to isoleucine (I) substitution in the protein sequence (column 61 in Fig. 4A) and renders the SpWRKY protein 44 amino acid residues longer than SbWRKY. The second amino acid difference is a substitution of histidine (H) to glutamine (O) (column 136 in Fig. 4A). There are no amino acid differences in the two WRKY domains, particularly in the conserved (WKKY/GQK) sequence thought to be directly involved in DNA binding (23).

**Discussion**

**Evolution of SbWRKY Orthologs in Two Sorghum Species and Other Major Grass Species.** SpWRKY appears to be a gain-of-function mutation in S. propinquum after its divergence from a common ancestor shared with S. bicolor. A parsimony model based on

![Graph](https://via.placeholder.com/150)

**Fig. 3.** Transgenic progeny shattering strength estimated with breaking tensile strength for the group of individuals that lack the transgene (\( n = 38 \)) and individuals with the transgene (\( n = 105 \)). The box of each group depicts lower quartile, median, and upper quartile. Dots, if any, indicate which observations that are considered outliers.

in immature inflorescences just emerged from flag leaf to 7.5 g in mature inflorescences (Fig. 2B).

Although SbTATA closely resembled the constitutively expressed actin control, the expression of SbWRKY was heavily modulated during the course of floral development (Fig. 2C). Both SbWRKY and SpWRKY were expressed strongly in leaves, but in floral tissues, the expression gradually decreased during inflorescence development (NE, inflorescence still in flag leaf; JE, inflorescence just emerging from flag leaf). However, near the anther dehiscence stage, SbWRKY had a strong pulse of expression in the nonshattering genotype, whereas the expression of SpWRKY continued to decrease. This clear difference of gene expression suggested a correlation of increased SbWRKY expression with nonshattering (i.e., SpWRKY is a negative regulator of shattering), also suggesting that the particular developmental stage of expression is critical for manifestation of the trait.

**Mutant Complementation.** We transformed the two candidate genes SbWRKY and SbTATA from genetically dominant S. propinquum BAC YRL20H16 into recessive nonshattering sorghum cultivar Tx430, each construct including coding sequences and about 1 kb of 5' and 3'-flanking sequences (Dataset S3). In transgenic sorghums validated to carry the S. propinquum allele, the SpWRKY construct induced discernible seed dropping in some transformants, whereas SbTATA transformants and controls did not.

To quantify the effect of SpWRKY on seed shattering, nine plants containing different transgenic events were self-pollinated, and up to 24 progenies were evaluated for each event. The transgene was segregating in eight of the nine progeny groups (Fig. S1; group T3 lacked the transgene, possibly indicating that it had not been integrated into the nucleus in the original transgenic plant). Across 136 plants from the eight validated events, reduced BTS was highly correlated with the presence of the transgene (\( r = -0.641; P = 4.2 \times 10^{-15} \)), with correlation coefficients in the individual populations (events) ranging from -0.399 to -0.946. Segregants that lacked the transgene showed an average BTS of 57.8 g (SD = 13.99, \( n = 38 \)), which was indistinguishable from that of the T3 population that lost the transgene (mean = 52.4, SD = 15.7, \( n = 17 \)). Plants containing
multiple sequence alignments favors the loss of the *SbWRKY* start codon by the ATG to ATT (methionine to isoleucine) mutation in *S. propinquum* after its divergence from a common lineage shared with *S. bicolor* (Fig. 4B). Transcriptome sequences from several wild *S. bicolor* accessions, *S. bicolor* ssp. verticilliflorum race virgatum (unnamed), *S. bicolor* (IS 21691), arundinaceum (IS 14216), and *S. bicolor* (IS 27587) have been conserved across plant species than *GRMZM2G161411* (Fig. 5). SpWRKY proteins often contain variable regions in the missing N-terminal sequence. The two maize gene copies were derived from the maize lineage genome duplication (26). The two *Setaria* genes are tandem copies adjacent to one another in the ancestral location, thus retaining genomic colinearity to the *SpWRKY* locus.

**Mutation in the Start Codon in the Shattering Allele Results in a Longer Protein.** The earlier start codon of *SpWRKY* than *SbWRKY* has been confirmed empirically, with the *S. propinquum* transcript being 132 bp longer than the *S. bicolor* transcript. Long (~140 amino acids) and short *WRKY* proteins (~100 amino acids) in sorghum also exist in other grass genomes, with the short proteins often lacking an ~40 amino acid N terminus. These extra ~40 amino acids in the long proteins are much more variable in the grasses than the (highly conserved) *WRKY* domain (Fig. 4A) or the 3′-terminal exon, contrary to general conservation patterns across plant genes. The main difference among the gene homologs is whether they have one or two additional exons in the 5′ end, which adds up to either two or three exons in total (Fig. 4C). Long *SpWRKY* proteins often contain three exons, with the only exception being *OsWRKY60* (Os03g0657400), which might have fused the first two exons. Based on the nucleotide alignments of the identified orthologs, all other orthologous genes in the related grass species have a G in the third nucleotide position of the start codon, with only *S. propinquum* showing a T. In addition, every grass ortholog in the comparison translates this specific codon to methionine with the exception of *SpWRKY* and the maize ortholog *GRMZM2G161411*, the latter with a TTG codon that translates to valine (V).

All grasses that we compared have at least one copy of the long *WRKY* protein, whereas species with two gene copies (maize and *Setaria*) also contain one short *WRKY* protein. The single orthologs in rice and *Brachypodium* are long. There are two copies in both maize and *Setaria: one short and one long*. The duplication into two copies in maize and *Setaria* occurred more recently and independently in the respective lineages after their divergence from other grasses (Fig. 4C). The QTL intervals affecting shattering on maize chromosomes 1 and 5 (2), respectively, harbor *GRMZM2G149219* (a short protein with 99 amino acids) and *GRMZM2G161411* (a long protein with 140 amino acids).

The different lengths of *SbWRKY* and *SpWRKY* proteins may impact gene function. The *WRKY* domain starts right after the first amino acid in the short *SbWRKY* protein (Fig. 4A). A sequence similarity search against GenBank using only the 44 N-terminal amino acids did not reveal any significant hits at E < 0.01. 3D structural prediction of the extended sequence in the long proteins using I-TASSER (27) suggested that part of the extended sequence folds into a β-sheet with two polar interactions with the *WRKY* domain, which may impact its ability to bind DNA (Fig. 5). We failed to detect significant signal peptide regions in the missing N-terminal sequence.

**Likely Mechanism of *SpWRKY* in the Shattering Pathway.** The *WRKY* superfamily of plant transcriptional factors has been implicated in a variety of physiological and developmental processes unique to plants, including leaf senescence (28, 29), trichome initiation (30), and embryo morphogenesis (31). The *WRKY* domain functions through direct interactions with the W-box domain in the promoter region of downstream gene targets (23). Overexpression of gene homologs results in ectopic lignin deposition in *Medicago* (32, 33), tobacco (34), and rice (35). Disruption of a *WRKY* gene in *Medicago* and *Arabidopsis* caused secondary cell wall thickening associated with the deposition
of lignin, xylan, and cellulose, implying that the gene is a negative regulator (33).

Lignin deposition at the seed–stalk interface is closely correlated with shattering of the sorghum florescence. Lignin staining of nonshattering sorghum revealed no deposition of lignin in the seed pedicel, in contrast to the fluorescence seen in the pericarp (Fig. 6B). In shattering sorghum (HAL; S. halepense) there is heavy lignin deposition in the stalk–glume junction, stopping abruptly as the glume starts. Across four stages of seed maturity (Fig. 6 C and D), lignin deposition is tightly correlated with the emergence of the abscission layer (AL) at later stages of seed development.

We hypothesize that SpWRKY functions in a manner analogous to Medicago and Arabidopsis homologs shown to repress cell wall biosynthesis genes (33). The low expression of SpWRKY toward the end of floral development may derepress its downstream cell wall biosynthesis genes, allowing deposition of lignin hence initiating the abscission zone in the seed–pedicel junction. The general pattern of expression of ShWRKY is similar to SpWRKY through the first stages of floral development, with a spike of expression of during anther dehiscence (Fig. 2C; WA, without anthers). The high expression of ShWRKY in the floral development stage may keep cell wall biosynthesis genes repressed in S. bicolor, also repressing the shattering phenotype. This difference in the expression profile between ShWRKY and SpWRKY alleles may be related to or partially caused by the shift in position of the start codon. However, the concomitant changes in both sequences and their expression patterns may necessitate further experiments to determine the exact molecular mechanism that differentiates ShWRKY and SpWRKY with regard to shattering.

Despite conventional wisdom that shattering of sorghum is genetically dominant (2), correlation of transgene copy numbers with the level of shattering indicates that the effect of ShWRKY is to some degree dependent on dosage. In the segregating populations derived from self-pollinated sorghums containing different transformation events, we also estimated SpWRKY copy number based on the intensity of amplified bands. There is a very strong negative correlation between BTS and the inferred copy number. Plants with no SP allele had a median BTS value of 56 g. Plants that contain the transgene had median values of 38.7 g (~31% reduction) for those with very faint SpWRKY bands, 24.4 g (~56% reduction) for faint, 13.9 g (~75% reduction) for medium, and 12.8 g (~77% reduction) for strong (Fig. S3).

Relationship to the YABBY Protein and the Genetics of Shattering Trait in Sorghum Species. The YABBY protein (Sh01g013270; Chr1: 11.94 Mb) reported to account for shattering in an intraspecific sorghum population (18) is outside of the region to which we delineated Sh1 by genetic recombination but physically is a remarkably close ~300 kb to SpWRKY (Sh01g012870; Chr1: 12.29 Mb). The proximity of these two genes, each implicated in the shattering phenotype, challenges the proposition of a single controlling locus as previously assumed. Indeed, in crosses segregating for functional polymorphisms at both loci, their close genetic linkage may have interfered with being recognized as separate loci.

We showed that the SpWRKY allele evolved in S. propinquum after its divergence from S. bicolor, partially explaining how there got to be (at least) two shattering genes in the Sorghum genus. Both WRKY and YABBBY genes contain DNA-binding domains known to regulate the transcription of genes involved in cell wall biosynthesis. Similar to the role of the WRKY gene in ectopic lignin deposition (33), the fruit from the Arabidopsis double mutant fil yab3 [YAB3 is a homolog of YABBBY (18)] lacks lignin in cells normally found in the valve margin (7).

It is possible that both WRKY and YABBY might regulate some of the same downstream genes involved in lignin deposition and thus each play important roles in the shattering pathway; however, there would be different relative contributions in the two species of sorghum studied. Our transgenic introduction of the SpWRKY allele into a background (RTx430) containing the nonshattering YABBBY allele restored the shattering phenotype, suggesting another possibility of epistatic interaction between YABBBY and WRKY. It remains a subject for further investigation what the modes of interaction between the two loci are, if any. It would be interesting, for example, to see if the expressions of the two genes overlap spatially and/or temporally. Thurber et al. suggested that the time of formation and degradation of abscission layers may vary between different shattering rice species, partially explaining the fact that there are separate loci controlling the trait (3).

In contrast to the hypothesis of convergent pathways for crop domestication (2, 18), the existence of two different sorghum shattering genes WRKY and YABBY supports an alternative hypothesis of multiple genetic pathways for crop domestication traits. There is a growing list of examples of major crop domestication phenotypes having diverse genetic architecture. Another related example is the seed threshability trait of the common wheat. In A- and D-genome diploid progenitors, two separate nonhomologous genes (SOG and Tg) arose by independent mutations, but both are known to contribute to the tenacity of the wheat glume (36).

Implications for Sorghum Domestication. The finding of recent and independent mutations at orthologs of the WRKY locus, both among S. bicolor genotypes (in our association study) and among divergent plant species, seems to suggest that the S. propinquum WRKY shattering locus is unstable and perhaps would have been less likely to reach high frequency during domestication than the S. bicolor YABBY locus. Given that SpWRKY is a gain-of-function mutation and thus far found only in S. propinquum, it likely represents the evolution of the genetic architecture of...
shattering in the wild species and played little or no role in the domestication of sorghum.

Nevertheless, the discovery of SpWRKY adds to our understanding of the genetics of shattering and the evolution of domestication-realted traits. Seed shattering is often under positive selection in crops and their wild relative; whereas humans tend to select nonshattering crops, seed dispersal in wild species may tend to favor maintaining or increasing the level of shattering. The seed shattering phenotype was probably the ancestral condition, whereas loss of shattering is a derived trait in the domesticate. This implies that the ancestral sorghum that predated the divergence of S. bicolor and S. propinquum contained the shattering allele of YABBY but the nonshattering allele of WRKY (Fig. 7). After the two sorghum species diverged, several nonshattering alleles of YABBY emerged in the S. bicolor population and were subsequently selected by humans (18), whereas the evolution of SpWRKY in S. propinquum perhaps reinforced or further increased the level of shattering (Fig. 7).

One finding is that the shattering phenotype is controlled by a different locus in the wild relative than in the domesticated relative is not an isolated case. All US weedy rice individuals share the same nonshattering mutation at the sh4 locus, which has been associated with loss of shattering during rice domestication, yet must have reacquired the shattering phenotype through unidentified regulatory genes other than sh4 (3). Similarly, we showed by transformation that the SpWRKY gene in the wild sorghum is fully capable of restoring the shattering phenotype even in the presence of homozygosity for the nonshattering allele of the domestication locus YABBY.

Additionally, the discovery of recent and independent duplications of maize and Setaria orthologs each followed by length mutations, together with our analysis of other sorghum species, suggests a tendency to recurring mutation at the locus. We currently only possess one valid S. propinquum accession (a second accession listed in the US Department of Agriculture National Plant Germplasm System has been substantially introgressed with S. bicolor chromatin, based on microsatellite genotyping); thus, we presently lack data about intraspecific S. propinquum sequence variation at the locus. Variation in both S. bicolor and S. propinquum would reconcile our association results with the fact that the specific allele that we cloned evolved in S. propinquum.

Materials and Methods

A diversity panel of sorghum varieties were collected from Cornell University and the US Department of Agriculture–Agricultural Research Service germplasm collection, selected to represent a wide range of geographical locations. The shattering phenotype for each accession was carefully validated by measuring BTS (measured in grams) using a digital force gauge (DPS-4; IMADA). Resequencing used BigDye terminator chemistry, and the chromatograms were examined using SEQUENCER software (version 4.1; GENECODES). The association analyses were performed using TASSEL software using a general linear model. Candidate genes in the high association region were isolated from the BAC YRL2OH16, cloned in pZP211, and transformed into nonshattering RTx430. Detailed methods and associated references are available in SI Materials and Methods.

Acknowledgments.

We thank B. Richardson for assistance with light microscopy and the late K. F. Schertz for key contributions to the early stages of this work. Financial support was provided by US Department of Agriculture Grant 01-35301-10595 (to A.H.P.).