Published in final edited form as:
*Nat Genet*. ; 44(6): 720–724. doi:10.1038/ng.2281.

# Parallel Domestication of the *Shattering1* Genes in Cereals

**Zhongwei Lin**[1], **Xianran Li**[1], **Laura M. Shannon**[2], **Cheng-Ting Yeh**[3,4], **Ming L. Wang**[5], **Guihua Bai**[1,6], **Zhao Peng**[7], **Jiarui Li**[7], **Harold N. Trick**[7], **Thomas E. Clemente**[8], **John Doebley**[2], **Patrick S. Schnable**[3,4], **Mitchell R. Tuinstra**[9], **Tesfaye T. Tesso**[1], **Frank White**[7], and **Jianming Yu**[1]

[1]Department of Agronomy, Kansas State University, Manhattan, Kansas, USA

[2]Department of Genetics, University of Wisconsin, Madison, Wisconsin, USA

[3]Center for Plant Genomics, Iowa State University, Ames, Iowa, USA

[4]Department of Agronomy, Iowa State University, Ames, Iowa, USA

[5]US Department of Agriculture–Agricultural Research Service (USDA-ARS), Griffin, Georgia, USA

[6]USDA-ARS, Manhattan, Kansas, USA

[7]Department of Plant Pathology, Kansas State University, Manhattan, Kansas, USA

[8]Center for Plant Science Innovation, University of Nebraska, Lincoln, Nebraska, USA

[9]Department of Agronomy, Purdue University, West Lafayette, Indiana, USA

## Abstract

A key step during crop domestication is the loss of seed shattering. Here we show that seed shattering in sorghum is controlled by a single gene, *Shattering1* (*Sh1*), which encodes a YABBY transcription factor. Domesticated sorghums harbor three different mutations at the *Sh1* locus. Variants at regulatory sites in the promoter and intronic regions lead to a low level of expression, a 2.2-kb fragment deletion causes a truncated transcript that lacks the second and third exons, and a GT-to-GG splicing variant in the intron 4 results in removal of the exon 4. The distributions of these non-shattering haplotypes among sorghum landraces suggest three independent origins. The function of the rice ortholog (*OsSh1*) was subsequently validated with a shattering resistant mutant, and two maize orthologs (*ZmSh1-1* and *ZmSh1-5.1+ZmSh1-5.2*) were verified with a large mapping population. Our results indicate that *Sh1* genes for seed shattering were under parallel selection during sorghum, rice, and maize domestication.

Cereal crops, the primary calorie source for humans, were domesticated thousands of years ago[1]. During domestication, many morphological and physiological characteristics of the wild progenitors of modern crops were reshaped to meet the needs of humans by artificial selection. When ancient humans started to cultivate wild crops, one of the most notable obstacles would have been the seed shattering habit. Seeds on wild grasses shed naturally at maturity, ensuring their natural propagation. Seed shattering, however, would have caused inefficient harvesting and large losses in grain yield for ancient humans. Hence, the non-shattering trait is likely to been placed under strong selection early in domestication. Because the trait can be achieved by changes in one or two major genetic loci with large effects[2, 3, 4], non-shattering variants could have appeared in the population at discernable frequencies, leading to the fixation of the non-shattering variants in ancient domesticated

Correspondence should be addressed to J.Y. (jyu@ksu.edu).

crop populations. Selection for non-shattering crop plants would have greatly facilitated harvesting and improved production, and propagation of cereal crops would have become increasingly dependent on humans, a feature that distinguishes modern crops from their wild progenitors. Although several genes have been identified as being responsible for seed shattering in rice and wheat[5, 6, 7, 8], whether other cereals share the same molecular genetic basis for shattering had not been determined.

Sorghum is the world's fifth major crop and a new model plant with applications in bioenergy and stress management[9]. Previous genetic studies have shown that seed shattering in sorghum is governed by a single locus[10, 11]. To identify the molecular basis underlying seed shattering in sorghum, we constructed an $F_2$ population from a cross between a wild sorghum with complete seed shattering, *Sorghum virgatum* (SV), and a non-shattering domesticated sorghum line, Tx430 (Fig. 1). The $F_1$ plants showed the same complete shattering as SV. The $F_2$ segregation ratio suggested that a single gene with a complete dominance effect explained this trait, and this gene was designated *Shattering1 (Sh1)*.

An initial scan with 94 $F_2$ individuals mapped the *Sh1* gene onto sorghum chromosome 1 (Fig. 2a and Online Methods). This mapping result was subsequently verified by linkage analysis with a simple sequence repeat (SSR) marker (Xtxp302) across 286 $F_2$ plants (Fig. 2b). On the basis of the sorghum genome sequence[9], four new SSR markers (P1, P2, P3 and P4) were developed, and *Sh1* was localized within a 2.5-cM region (Fig. 2b). With three additional SSR makers (P5, P6 and P7) and one SNP1, we fine-mapped *Sh1* to a region between P6 and SNP1, using 15,000 $F_2$plants (Fig. 2c). A BAC clone (25K18) from a wild sorghum with the shattering habit, *Sorghum propinquum* (*Propinquum*), was found to cover the *Sh1* candidate region. Sequence analysis of this BAC clone eventually placed *Sh1* within a fragment of approximately 17 kb in size between P6 and SNP1 on sorghum chromosome 1 (Fig. 2d). Sequence annotation of this fragment revealed only two predicted genes: a hypothetical gene and a transcription factor gene belonging to the *YABBY* family. The sequence of the hypothetical gene from a shattering $F_2$ recombinant plant, 15G07, was identical to that of the non-shattering Tx430 parent (Fig. 2d), indicating that recombination occurred between these two genes; therefore, the *YABBY*-like gene was regarded as the functional candidate of *Sh1* in sorghum.

We compared the 7,758-nt *YABBY*-like gene region from the start codon to the stop codon in SV and Tx430 plants (Fig. 3a). No nucleotide differences were detected in 6 exons, whereas 26 SNPs were present in 5 introns. We then sequenced the *YABBY*-like gene from the 2,761-nt upstream promoter region to the 1,315-nt downstream fragment after the stop codon across 13 sorghum accessions. This set of accessions comprised three wild shattering accessions and ten non-shattering domesticated lines (Fig. 3a and Online Methods). Four haplotypes emerged from the 10 representative nucleotide variants across these 13 sorghum accessions. Three shattering wild accessions had an SV-like haplotype; two non-shattering accessions had a SC265-like haplotype harboring a GT-to-GG splice-site variant at nucleotide position 6,608; four non-shattering accessions shared a Tx430-like haplotype with two promoter variants at positions −1,194 and −1,185 and two intronic variants at positions 4,881 and 5,076; and four non-shattering accessions had a Tx623-like haplotype with a 2.2-kb deletion from 3,985 to 6,251 at the location of exons 2 and 3 (Fig. 3a, green). We also sequenced 146 sorghum accessions from around the world (Online Methods and (Supplementary Tables 1 and 2). Of note, the SV-like haplotype was conserved across all 25 shattering accessions, and the three other haplotypes were retained in almost all non-shattering accessions, except for in two that had rare recombination events (Fig. 3a and Supplementary Tables 1 and 2).

We next performed an association test across 25 shattering and 121 non-shattering accessions. Significant associations at all ten representative sites (*P* values = $6.0 \times 10^{-4}$ to $1.6 \times 10^{-11}$) were obtained. Strong signals were observed at positions −1,619, 152, 5,449 and 8,122, whereas medium-strength signals were detected at the locations of the multiple causal variants, including the four specific mutations of the Tx430-like haplotype, the 2.2-kb insertion and deletion (indel) and the splice-site variant (Fig. 3b). The distributions of the common variants at positions −1,619, 152, 5,449 and 8,122 (with an allele frequency of 83/146 for positions −1,619 and 8,122, or 84/146 for positions 152 and 5,449) were correlated with those of the presumptive causal variants in the Tx430-like (37/146) and Tx623-like (47/146) haplotypes, which occurred with lower frequencies. Because the common variants were in linkage disequilibrium (LD) with the presumptive causal variants but had higher frequencies (Supplementary Fig. 1), their synthetic association signals were stronger than those of the causal variants (Fig. 3b and Supplementary Table 2). The association signals at positions 4,881, 5,076 and 5,449 were strong because these sites contained causal alleles in the Tx430-like and Tx623-like haplotypes. When we tested three domesticated haplotypes as a group against the wild haplotype, the *YABBY*-like gene was completely associated with the shattering trait (*P* value = $1.1 \times 10^{-28}$) (Fig. 3b, red dot).

The 576-bp coding sequence of the *Sh1* gene encodes a YABBY protein consisting of 191 amino acids. The zinc finger domain is located from amino acids 11 to 52, and the YABBY domain is located from amino acids 111 to 165 (Fig. 3c).

To identify the causal polymorphisms, we conducted expression analysis via RT-PCR, focusing on the entire coding transcript. Compared with the wild shattering lines (SV and *Propinquum*), the Tx430 line showed a low level of *Sh1* transcription, whereas truncated transcripts were found in Tx623 and SC265 (Fig. 3d). Sequence analysis of the transcripts of SV, Tx430, Tx623 and SC265 further showed that Tx430 shared the same 576-bp coding region as SV, whereas Tx623 encoded a 317-bp transcript without exons 2 and 3 and SC265 encoded a 527-bp transcript missing exon 4 (Supplementary Fig. 1). Both truncated transcripts contained frameshifts that resulted in the introduction of premature stop codons. The Tx623-like haplotype therefore encodes a protein lacking the zinc finger and YABBY domains, whereas the SC265-like haplotype encodes a protein lacking only the YABBY domain (Fig. 3c).

Microscopic examination revealed that abscission layers began to form in the joint connecting the seed hull and pedicel in SV plants during flowering, whereas no abscission layer was formed in Tx430 plants (Fig. 1e,f). These results indicate that the two promoter and two intronic variants in the Tx430-like haplotype repressed the expression of the *Sh1* gene, whereas the 2.2-kb deletion in the Tx623-like haplotype and the splice-site variation from GT to GG in the SC265-like haplotype altered the *Sh1*-encoded protein, eliminating formation of the abscission layer in the joint between the hull and pedicel and thereby resulting in a loss of shattering.

Sorghum is known for its genetic diversity, and multiple domestication events have been suggested on the basis of morphology and molecular marker analyses[12, 13]. In this study, we found that the shattering haplotype remains conserved within 22 wild shattering accessions and 3 shattering *Sorghum bicolor* strains from different regions of the world (the shattering habit may have been introgressed into these 3 bicolor sorghum strains from wild sorghum), but 3 distinct haplotypes were present among 121 non-shattering domesticated sorghums. Seed shattering, vital for the propagation of wild crop progenitors in nature, should be under very strong, direct natural selection, which could explain the conserved shattering haplotype. Three ancient human groups may have cultivated wild sorghum progenitors successfully by independently selecting and fixing three different *sh1* mutations that resulted in non-

shattering in different original populations, overpowering natural selection. Indeed, among 80 non-shattering landraces, the Tx430-like haplotype group is dominated by accessions from *caudatum* (Supplementary Table 2), a race thought to be domesticated later than other races. The Tx623-like haplotype group primarily contains accessions from the *kafir* and *biocolor* races from South and East Africa. All accessions from *durra*, most from *guinea* and almost half of those from *bicolor* have the SC265-like haplotype. The lack of a dominant *sh1* haplotype among accessions from *bicolor* agrees with the wide distribution of this race.

Cereal crops such as rice, maize, wheat, barley and sorghum were domesticated thousands of years ago. Although these crops were domesticated from different wild progenitors by different ancient human groups in different geographical zones, they all underwent systemic and parallel changes during the domestication process[14]. Whether these parallel changes in domestication share the same genetic basis is still vigorously debated. Although the two major genes for shattering in rice (*Qsh1* and *Sh4*) and the *Q* gene for shattering in wheat have not been found to be under selection in other crops, the newly identified *Sh1* gene in sorghum provides another entry point to test this hypothesis. A shattering quantitative trait locus (QTL) with minor genetic effect has been repeatedly mapped to a syntenic block corresponding to *Sh1* on rice chromosome 3 (refs. 15,16,17), two disarticulation QTLs were identified in the syntenic blocks on maize chromosomes 1 and 5 (refs. 11,18), and one of the two major QTLs for shattering was also mapped within the same block on foxtail millet chromosome 9 (ref. 19) (Fig. 4a).

The rice QTL for shattering in the *Sh1* syntenic region has its peak 80 kb away from the gene orthologous to *Sh1* (*OsSh1*, LOC_Os03g44710) (Fig. 4a, blue arrow, and Supplementary Note) and a non-shattering mutant that has reduced cell numbers at the abscission layer was available for further anlaysis[15, 16, 17, 20]. Using the non-shattering mutant (SR-5) and the wild-type rice breeding line (Nanjing 11), we conducted genomic DNA amplification and expression analysis and Southern blotting (Supplementary Figs. 2–4). An insertion of a >4-kb fragment was identified in intron 3 of *OsSh1*, leading to reduced levels of transcription and the shattering-resistant phenotype. Furthermore, in two recent whole-genome sequencing studies, *OsSh1* was in a list of genes shown to be under strong artificial selection[21, 22].

On maize chromosome 1, the *Sh1* orthologous gene in the reference genome of B73 (*ZmSh1-1*)[23] contains an extremely large (19.3-kb) intron 1 (Fig. 4b and Supplementary Note), similar to the key domestication gene for fruit size (*fasciated*) in tomato[24]. On maize chromosome 5, the B73 genome contains two copies of the *Sh1* orthologous gene (*ZmSh1-5.1*+*ZmSh1-5.2*) within the syntenic block (Fig. 4b). The structural change in *ZmSh1-5.1*+*ZmSh1-5.2* is present in the majority of maize inbreds but absent in all teosinte inbreds, and the insertion in *ZmSh1-1* is present in almost all maize inbreds and a few teosinte inbreds, including three with incomplete shattering (Fig. 5 and Supplementary Table 3). For two maize inbreds that did not contain the B73-type insertion in*ZmSh1-1*, sequence alignment revealed another 83-bp insertion in exon 3 of *ZmSh1-1* on chromosome 1 causing a frameshift.

To further investigate *ZmSh1* family genes, we conducted a whole-genome linkage scan with a maize-teosinte population. A major QTL for shattering with a very narrow confidence interval (genetic distance of 1.4 cM, physical distance of 0.6 Mb) was identified on chromosome 5, and another QTL for shattering was identified on chromosome 1 (2.1 cM, 2.4 Mb) (Fig. 5). These two QTL intervals correspond to the genomic regions harboring the *ZmSh1-1* and *ZmSh1-5.1*+*ZmSh1-5.2* genes. The *YABBY*-like *ZmSh1-1* gene is 1 of the 2 genes encoding transcription factors among the 59 annotated genes within the chromosome

1 interval, and the *ZmSh1-5.1*+*ZmSh1-5.2* locus is the only transcription factor gene among the 12 annotated genes within the chromosome 5 interval (Supplementary Table 4 and 5).

In summary, the identification of the *Sh1* gene in sorghum, the conserved collinearity of genomic regions containing the *Sh1* orthologs across several cereals, the identification of the rice ortholog *OsSh1* and the structural variation and QTL analyses of the two maize orthologs *ZmSh1-1* ( and *ZmSh1-5.1*+*ZmSh1-5.2*) suggest that the *Sh1* genes for seed shattering have undergone parallel selection during domestication in multiple cereals.

## Methods

### Map-based cloning

A large $F_2$ population with 15,286 plants was generated from a cross between a wild sorghum *virgatum* (*Sorghum bicolor* (L.) Moench ssp. *verticilliflorum* (Steud.) race *virgatum*) and a standard sorghum line Tx430 (*Sorghum bicolor* (L.) Moench ssp. *bicolor*). All these plants were planted in the Kansas State University greenhouse.

First, 286 $F_2$ plants were planted for genetic mapping of seed shattering. The segregation between 217 shattering and 69 non-shattering $F_2$ plants fit well with a 3:1 ratio ($\chi^2 = 0.12$, $P = 0.73$), indicating a single gene with a complete dominance effect for seed shattering. Next, 94 out of 286 random $F_2$ plants were genotyped with a 384-SNP DNA chip[25]. SNP-trait relationship was tested for 90 SNPs with polymorphisms distributed across the genome by Fisher's exact test because the shattering trait was scored as a binary trait. The significance threshold was corrected for multiple testing by Bonferroni correction, $\alpha' \approx \alpha/n = 0.05/90 = 5.6 \times 10^{-4}$, where $\alpha$ is the nominal significance threshold and $n$ is the number of SNPs. A linkage map was constructed with MAPMAKER/EXP 3.0b[26].

Subsequently, 15,000 additional $F_2$ plants were grown to the 4- or 5-leaf stage in a planting tray with 14 × 7 wells. Rapid genotypic screening was performed with two SSR markers, P1 and P2. Of the 587 recombinant $F_2$ plants transplanted for phenotyping, 38 with recombination events between SSR markers P5 and P7 were self-pollinated to produce $F_3$ families. From each of the 38 families, 20 random plants were used to confirm the corresponding $F_2$ phenotype.

### Association mapping

The 11.8-kb DNA fragment containing the *Sh1* gene was amplified from 146 sorghum accessions by PCR with seven pairs of primers (Supplementary Table 6) using TaKaRa LA Taq (RR002A). All PCR products were cleaned with the QIAquick PCR Purification kit (Qiagen) and sequenced by ABI 3730 (Applied Biosystems). The sequences were assembled with the CodonCode Aligner.

To confirm their phenotype, 25 shattering sorghum accessions obtained from the Germplasm Resources Information Network (GRIN) were planted in the greenhouse. These 22 shattering sorghum accessions mainly belong to wild sorghum from Africa (Supplementary Table 1), whereas the other three shattering sorghum accessions belong to domesticated sorghum. The shattering habit of wild sorghum may have introgressed into these three domesticated sorghum accessions.

Including 41 breeding lines from a *S. bicolor* panel[27] and 80 landraces with diverse origins, 121 non-shattering domesticated sorghum accessions can be divided into five domesticated races: *bicolor*, *caudatum*, *durra*, *guinea* and *kafir* (Supplementary Table 2). To represent the diversity of the *S. bicolor* panel, we selected eight domesticated sorghum accessions, Tx430, SC35, SC265, Ajabsido, Macia, SC1103, Segaolane and SC1345, from the parents of

sorghum nested association mapping (NAM) populations for sequencing of the whole *Sh1* gene.

We used ten representative common variants with allele frequency of >10% and two most closely flanking markers of the *Sh1* gene for association testing on 146 sorghum accessions, including 25 shattering and 121 non-shattering accessions, with Fisher's exact test, as Fisher's exact test is powerful for qualitative traits like shattering[28]. LD between two loci, A and B, was calculated as

$$r^2=\sum_{i=1}^{m}\sum_{j=1}^{n}p(\mathrm{A}_i\mathrm{B}_j)r_{ij}^2 \quad \text{and} \quad r_{ij}^2=\frac{(p(\mathrm{A}_i\mathrm{B}_j)-p(\mathrm{A}_i)p(\mathrm{B}_j))^2}{p(\mathrm{A}_i)(1-p(\mathrm{A}_i))p(\mathrm{B}_j)(1-p(\mathrm{B}_j))}$$

where A has $m$ alleles, B has $n$ alleles, $p(\mathrm{A}_i)$ is the frequency for $\mathrm{A}_i$, $p(\mathrm{B}_j)$ is the frequency for $\mathrm{B}_j$and $p(\mathrm{A}_i\mathrm{B}_j)$ is the frequency for $\mathrm{A}_i\mathrm{B}_j$. A triangle LD matrix was constructed with PowerMarker[29].

## Expression analysis

Total RNA was extracted by Qiagen Plant Easy RNA kit from the junction (1–2 cm) between the hull and pedicel where abscission layers were located. First-strand cDNA synthesis was performed with SuperScript II Reverse Transcriptase (18064-022, Invitrogen). *Sh1* transcript was amplified, and *Actin* transcript was amplified as an internal control (Supplementary Table 6).

## Microscopy

Longitudinal sections cut by hand were stained with- acridine orange and imaged with a Zeiss Axioplan 2 microscope. Fluorescence images were acquired with a long-pass 650-nm emission filter (red fluorescence) under excitation at 550 nm.

## Comparative genome analysis

BLASTP of SynMap on CoGe was used to conduct pairwise genome comparisons. Genome sequence data sources were selected for BLASTP analysis for the following cereals: sorghum (Tx623, id331), maize (B73, id333), rice (Nipponbare, id3) and foxtail millet (Yugu1, id32546). A genomic collinearity map was plotted in R on the basis of the BLASTP results.

## Rice mutant validation

The rice non-shattering mutant SR-5 was induced by gamma-ray irradiation from rice breeding line Nanjing 11, which belongs to the *indica* subspecies with an easy shattering habit. Scanning electron microscope (SEM) analysis of the glume-pedicel junction revealed that the surface of the glume-pedicel junction was very rough in SR-5 but smooth in Nanjing 11 when seeds were removed (Supplementary Fig. 2). The difference in the surface of the glume-pedicel junction between SR-5 and Nanjing 11 suggested development of the abscission layer was affected in the SR-5 mutant.

PCR for the *YABBY*-like *Sh1* orthologous gene (*LOC_Os03g44710*) in the rice SR-5 mutant and the wild-type Nanjing 11 line resulted in successful amplification, except for a region located in intron 3 in SR-5 (Supplementary Fig. 3 and Supplementary Table 7). Further analysis with Southern blotting showed that a >4-kb insertion was present in intron 3 of the rice ortholog of *Sh1*(*OsSh1*) in the SR-5 mutant (Supplementary Fig. 4). This >4-kb insertion greatly decreased transcription of the rice *Sh1* ortholog, which affected the

abscission layer, with seeds being retained on the heads of SR-5 plants (Supplementary Fig. 2a,b). These results suggested the *Sh1*genes were functionally conserved between two model cereal crops, sorghum and rice.

### Maize gene structure analysis and QTL validation

To determine the extent of structural variations across maize germplasm, we assembled the sequences of the *ZmSh1-1* (22 kb) and *ZmSh1-5.1+ZmSh1-5.2* (35 kb) genes across a diverse set of 27 maize and 17 teosinte inbred lines (Supplementary Table 3), using short reads from maize HapMap V2 and RNA-seq data (P.S.S., unpublished data) (Supplementary Table 3). Structural variation was scored on the basis of the coverage and abundance of the uniquely aligned reads in the corresponding region.

We then conducted a QTL mapping experiment using a large maize-by-teosinte population with 866 recombinant inbred lines (RILs). A maize inbred, W22, was crossed to a teosinte (*Zea mays*ssp. *parviglumis*, accession CIMMYT 8759), and the resulting $F_1$ generation was backcrossed to W22 for two generations; RILs were derived by selfing the $BC_2F_1$ for three generations. The 866 RILs were grown in two blocks during the summer of 2009 and in an additional block in the summer of 2010 at the West Madison Agricultural Research Center in Madison, Wisconsin. Randomized complete-block design was used with ten plants per plot. Shattering was scored quantitatively as the number of segments into which the mature ears fractured at harvest. Least squared means were determined for each RIL. QTL mapping was conducted with 19,838 genotyping-by-sequencing (GBS) markers of known physical positions in the maize B73 reference genome. The multiple-interval mapping method implemented in R/qtl[30] was used. The significance threshold (5.87) was determined with 10,000 permutations. Two major QTLs were identified at the genomic regions harboring *ZmSh1-1* and *ZmSh1-5.1+ZmSh1-5.2*, and two other QTLs with smaller effects were also identified (Supplementary Table 6 and 8).

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## References


1. Doebley JF, Gaut BS, Smith BD. The molecular genetics of crop domestication. Cell. 2006; 127:1309–21. [PubMed: 17190597]
2. Li C, Zhou A, Sang T. Genetic analysis of rice domestication syndrome with the wild annual species, Oryza nivara. New Phytol. 2006; 170:185–93. [PubMed: 16539615]
3. Matsui K, et al. Identification of AFLP makers linked to non-seed shattering locus (sht1) in buckwheat and conversion to STS markers for marker-assisted selection. Genome. 2004; 47:469–74. [PubMed: 15190364]
4. Nalam VJ, Vales MI, Watson CJ, Kianian SF, Riera-Lizarazu O. Map-based analysis of genes affecting the brittle rachis character in tetraploid wheat (Triticum turgidum L.). Theor Appl Genet. 2006; 112:373–81. [PubMed: 16328232]
5. Li C, Zhou A, Sang T. Rice domestication by reducing shattering. Science. 2006; 311:1936–9. [PubMed: 16527928]
6. Lin Z, et al. Origin of seed shattering in rice (Oryza sativa L.). Planta. 2007; 226:11–20. [PubMed: 17216230]
7. Konishi S, et al. An SNP caused loss of seed shattering during rice domestication. Science. 2006; 312:1392–6. [PubMed: 16614172]
8. Simons KJ, et al. Molecular characterization of the major wheat domestication gene Q. Genetics. 2006; 172:547–55. [PubMed: 16172507]

9. Paterson AH, et al. The Sorghum bicolor genome and the diversification of grasses. Nature. 2009; 457:551–6. [PubMed: 19189423]
10. Harlan JR, de Wet JMJ, Price EG. comparative evolution of cereals. Evolution. 1973; 27:15.
11. Paterson AH, et al. Convergent domestication of cereal crops by independent mutations at corresponding genetic Loci. Science. 1995; 269:1714–8. [PubMed: 17821643]
12. de Wet JMJ, Huckabay JP. the origin of sorghum bicolor. II. Distribution and Domestication. Evolution. 1967; 21:16.
13. Cui YX, Xu GW, Magill CW, Schertz KF, Hart GE. RFLP-based assay of Sorghum bicolor (L.) Moench genetic diversity. Theor Appl Genet. 1995; 90:787–796.
14. Harlan, JR. Crop and Man. Madison, WI, USA: American Society of Agronomy; 1992. p. 125-132.
15. Gu XY, Kianian SF, Hareland GA, Hoffer BL, Foley ME. Genetic analysis of adaptive syndromes interrelated with seed dormancy in weedy rice (Oryza sativa). Theor Appl Genet. 2005; 110:1108–18. [PubMed: 15782297]
16. Onishi K, Takagi K, Kontani M, Tanaka T, Sano Y. Different patterns of genealogical relationships found in the two major QTLs causing reduction of seed shattering during rice domestication. Genome. 2007; 50:757–66. [PubMed: 17893735]
17. Onishi K, Horiuchi Y, Ishigoh-Oka N, et al. A QTL cluster for plant architecture and its ecological significance in Asian wild rice. Breeding Science. 2007; 57:7–16.
18. Doebley JF, Stec A. Genetic analysis of the morphological differences between maize and teosinte. Genetics. 1991; 129:285–295. [PubMed: 1682215]
19. Devos KM, Gale MD. Genome relationships: the grass model in current research. Plant Cell. 2000; 12:637–46. [PubMed: 10810140]
20. Fukuta Y, Yagi T. Mapping of a shattering resistance gene in a mutant line SR-5 induced from an indica rice variety, Nan-jing 11. Breeding Science. 1998; 48:345–348.
21. He Z, et al. Two evolutionary histories in the genome of rice: the roles of domestication genes. PLoS Genet. 2011; 7 e1002100.
22. Xu X, et al. Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. Nat Biotechnol Online first. 2012
23. Schnable PS, Ware D, Fulton RS, Stein JC, et al. The B73 maize genome: complexity, diversity, and dynamics. Science. 2009; 326:1112–5. [PubMed: 19965430]
24. Cong B, Barrero LS, Tanksley SD. Regulatory change in YABBY-like transcription factor led to evolution of extreme fruit size during tomato domestication. Nat Genet. 2008; 40:800–4. [PubMed: 18469814]
25. Murray SC, Rooney WL, Hamblin MT, Mitchell SE, Kresovich S. Sweet sorghum genetic diversity and association mapping for brix and height. The Plant Genome. 2009; 2:48–62.
26. Lander ES, et al. MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics. 1987; 1:174–81. [PubMed: 3692487]
27. Casa AM, et al. Community resources and strategies for association mapping in sorghum. Crop science. 2008; 48:30.
28. Balding DJ. A tutorial on statistical methods for population association studies. Nat Rev Genet. 2006; 7:781–91. [PubMed: 16983374]
29. Liu K, Muse SV. PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics. 2005; 21:2128–9. [PubMed: 15705655]
30. Broman KW, Wu H, Sen S, Churchill GA. R/qtl: QTL mapping in experimental crosses. Bioinformatics. 2003; 19:889–90. [PubMed: 12724300]

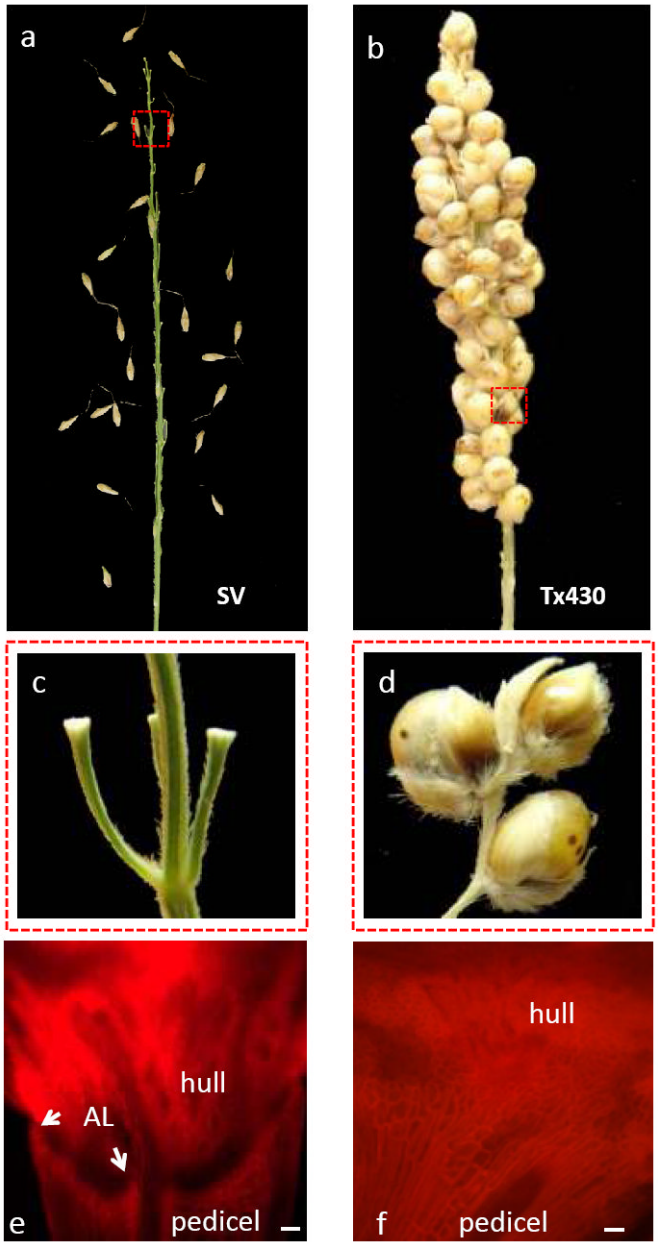


**Figure 1. Seed shattering phenotype in sorghum**

(**a**,**b**) Seeds were scattered everywhere from the top of the wild sorghum SV plant (**a**), whereas seeds were firmly retained on the head of the domesticated sorghum Tx430 plant (**b**, shown only from a panicle branch) at maturity after vigorous shaking. (**c**,**d**) Larger views of spikelets in **a** and **b** are shown for SV (**c**) and Tx430 (**d**) plants after shaking. (**e**,**f**) Abscission layers (of curved shape) were present at the junction between the hull and pedicel on SV plants (**e**), whereas no abscission layer was observed on Tx430 plants (**f**). AL, abscission layer; scale bars, 50μm.

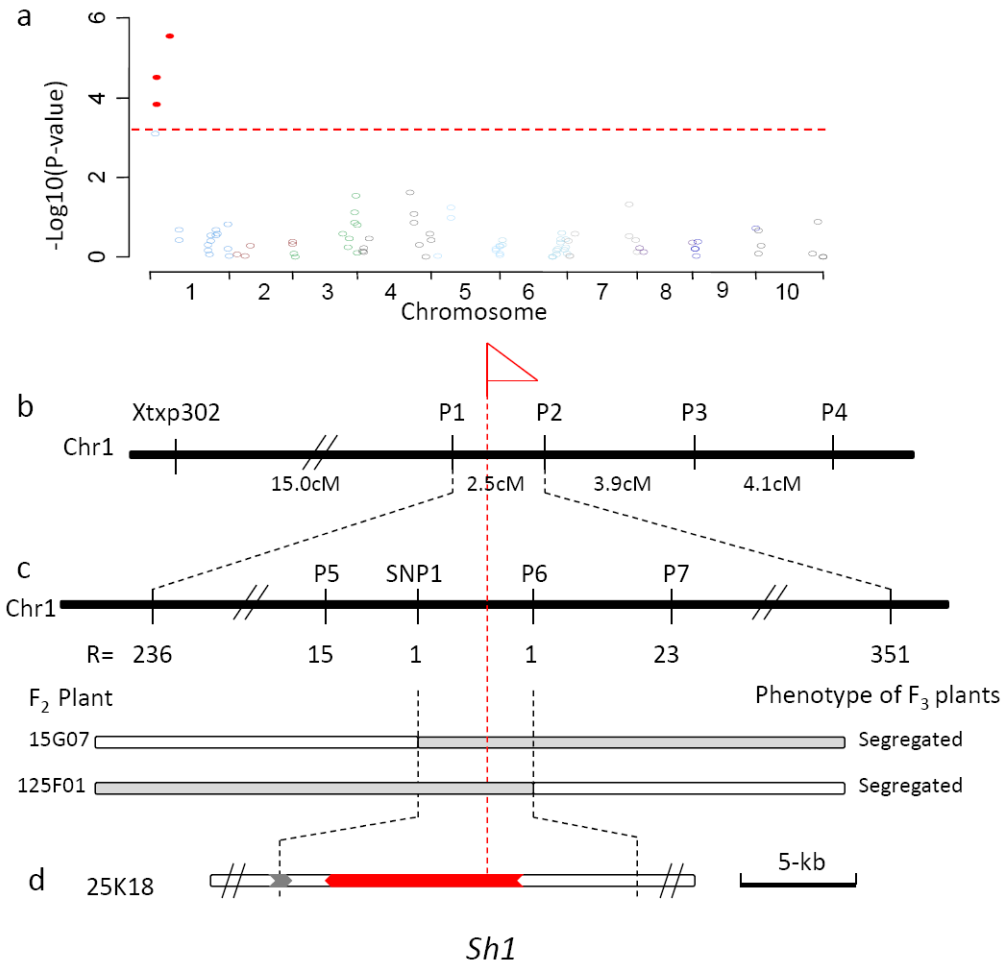


**Figure 2. Map-based cloning of *Sh1* in sorghum**

(**a**) DNA chip screening across 94 $F_2$ plants mapped the *Sh1* gene onto sorghum chromosome 1. Significant SNPs are marked as red dots; the red dashed line represents the 5% significance threshold with Bonferroni correction for 90 tests. (**b**) Genetic mapping of *Sh1* with 286 $F_2$ plants. Genetic distance between flanking pairwise molecular markers is shown. (**c**) High-resolution mapping of *Sh1* with 15,000 $F_2$ plants. The genotypes of two $F_2$ recombinant plants are shown; the $F_3$ progenies of both $F_2$ recombinant plants are segregated by phenotype. Gray bar, heterozygous region of Tx430; white bar, homozygous region of Tx430; R, recombinant plant. (**d**) The *Propinquum* BAC clone 25K18 was identified by the two flanking markers of *Sh1*, and two genes were predicted within the candidate region between these markers. Gray arrow, a hypothetical gene specifically expressed in pollen; red arrow, *YABBY*-like gene.

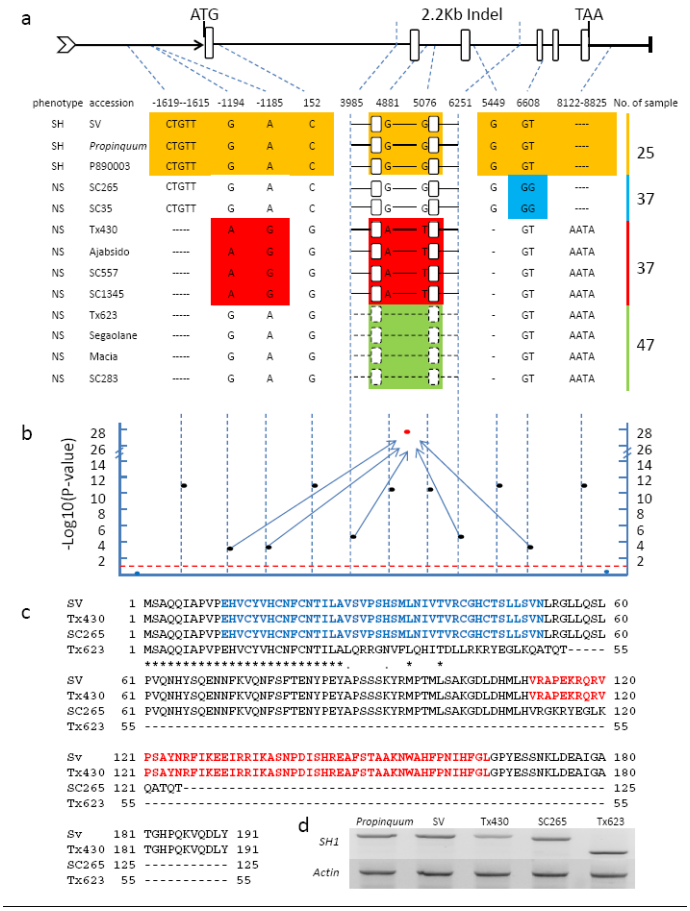


**Figure 3. Variant alleles and association mapping at *Sh1***

(**a**) Gene structure and haplotype analysis of *Sh1*. —SV-like, SC265-like, Tx430-like and Tx623-like haplotypes were identified on the basis of ten variants of the *Sh1* gene. The positions of these ten variants are shown using the *Sh1* gene sequence of SV as reference, with the start codon designated as position 0. The SV-like haplotype is highlighted in yellow. Splicing mutation from GT to GG at 6,608 in the SC265-like haplotype (blue), four specific variants at −1,194, −1,185, 4,881 and 5,076 in the Tx430-like haplotype (red) and a 2.2-kb deletion from 3,985 to 6,251 in the Tx623-like haplotype (green) are indicated. Arrow bar, promoter region; thick bar, downstream region after the stop codon; blank box, exon; thin line, intron; ATG, start codon; TAA, stop codon; adjacent unfilled boxes on a dashed line, 2.2-kb deletion; SH, shattering; NS, non-shattering. (**b**) Association testing at sites of ten variants and two fine-mapping markers P6 and SNP1. Black dots, ten variants of *Sh1*; blue dots, P6 (left) and SNP1 (right); red dot, supposed synthetic association site; red dashed line, 5% significance threshold. (**c**) Amino-acid sequence alignment for different haplotypes. Two domains, zinc finger (blue) and YABBY (red), are indicated. (**d**) There was strong transcription *of Sh1* in *Propinquum* and SV, whereas there was weak transcription in Tx430 and truncated transcripts for Tx623 and SC265.

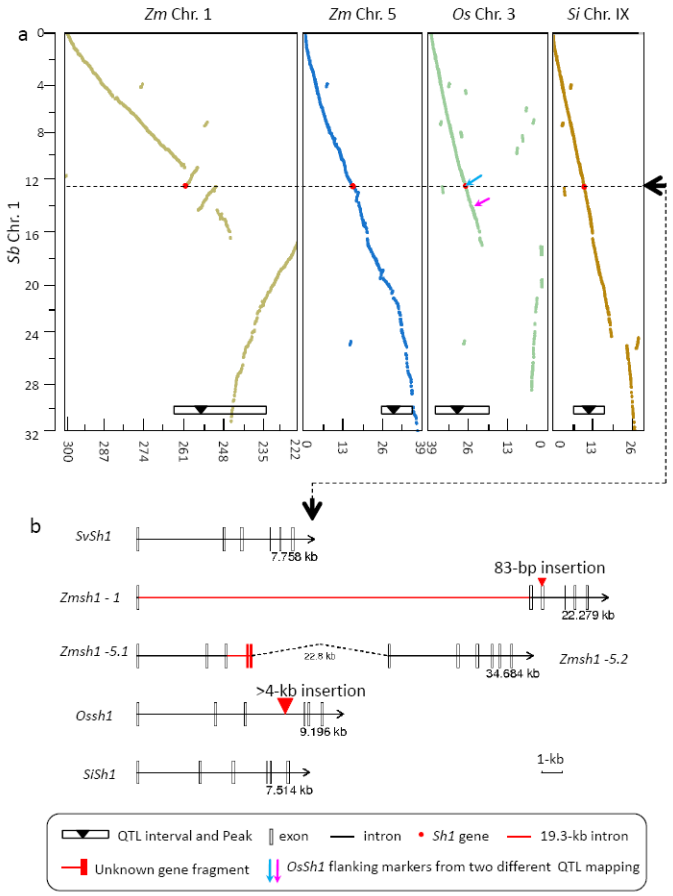


**Figure 4. Genomic regions of *Sh1* in cereals**

(**a**) Genomic regions corresponding to *Sh1* were conserved in sorghum (*Sorghum bicolor*), maize (*Zea mays*), rice (*Oryza sativa*) and foxtail millet (*Setaria italica*). The genomic collinearity map was plotted on the basis of the BLASTP result of pairwise genome analysis from CoGe; dot plot alignment indicates the collinearity of genomic regions. (**b**) *Sh1* gene structure comparison. *Sh1* gene structure is conserved, except for one extremely large intron (19.3 kb) that was present only in the *Sh1* ortholog on maize chromosome 1 (*ZmSh1-1*) and a gene fusion that occurred in one of two *Sh1* orthologs on maize chromosome 5 (*ZmSh1-5.1*).

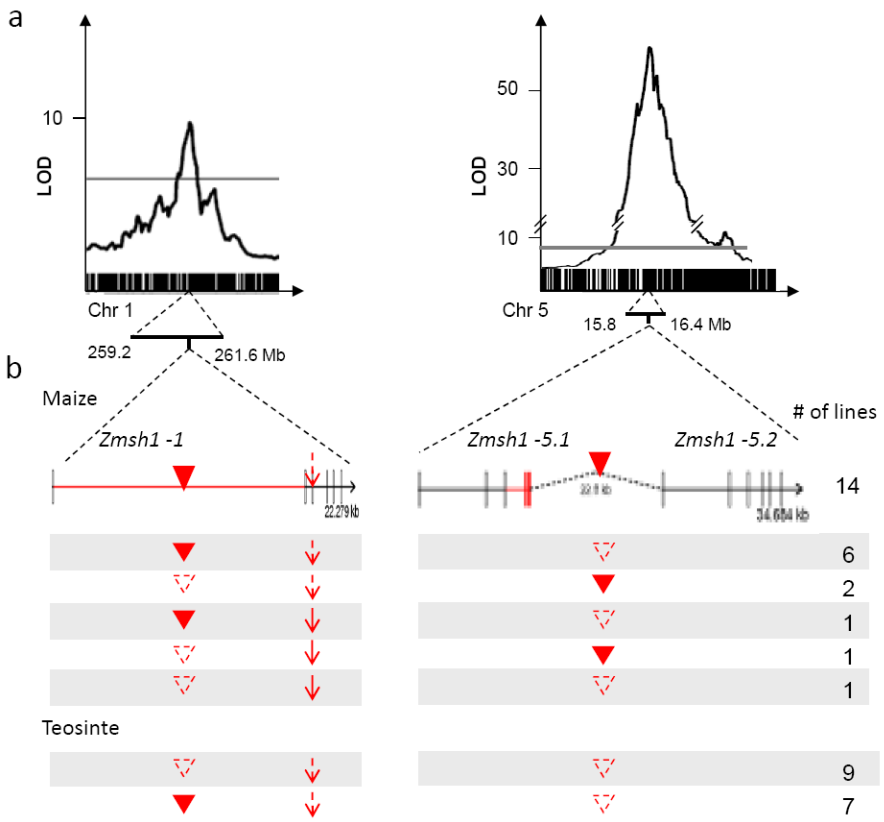


**Figure 5. Maize sh1 orthologs are located at seed shattering QTLs**

(**a**) Two QTLs explaining 3.5% and 23.1% of the phenotypic variation of shattering were detected on maize chromosomes 1 and 5, respectively, using a large mapping population. Physical positions of the QTL confidence intervals and the maize *Sh1* orthologs are indicated below the chromosome axes. LOD, logarithm of odds. (**b**) Assembly of structural variation for the *Sh1* orthologs on maize chromosomes 1 and 5. *ZmSh1-5.1* consists of a gene fusion that retains only the first three exons of the *YABBY*-like gene and two exons from an unknown gene. The replacement of exons 4–6 in *ZmSh1-5.1* causes loss of the YABBY domain. A large (23-kb) insertion is present between *ZmSh1-5.1* and *ZmSh1-5.2*. Solid/dashed red triangle, presence/absence of large insertion; solid/dashed red arrow, presence/absence of 83-bp insertion in exon 3 of *ZmSh1-1*.