Presence of tannins in sorghum grains is conditioned by different natural alleles of *Tannin1*

Yuye Wu[a], Xianran Li[a], Wenwen Xiang[a], Chengsong Zhu[a], Zhongwei Lin[a], Yun Wu[a], Jiarui Li[b], Satchidanand Pandravada[a], Dustan D. Ridder[a], Guihua Bai[a,c], Ming L. Wang[d], Harold N. Trick[b], Scott R. Bean[e], Mitchell R. Tuinstra[f], Tesfaye T. Tesso[a], and Jianming Yu[a,1]

Departments of [a]Agronomy and [b]Plant Pathology, Kansas State University, Manhattan, KS 66506; [c]Hard Winter Wheat Genetics Research Unit, US Department of Agriculture-Agricultural Research Service, Manhattan, KS 66506; [d]Plant Genetic Resources Conservation Unit, US Department of Agriculture-Agricultural Research Service, Griffin, GA 30223; [e]Grain Quality and Structure Research Unit, US Department of Agriculture-Agricultural Research Service, Manhattan, KS 66502; and [f]Department of Agronomy, Purdue University, West Lafayette, IN 47907

Edited* by Ronald L. Phillips, University of Minnesota, St. Paul, MN, and approved May 23, 2012 (received for review February 6, 2012)

Sorghum, an ancient old-world cereal grass, is the dietary staple of over 500 million people in more than 30 countries in the tropics and semitropics. Its C4 photosynthesis, drought resistance, wide adaptation, and high nutritional value hold the promise to alleviate hunger in Africa. Not present in other major cereals, such as rice, wheat, and maize, condensed tannins (proanthocyanidins) in the pigmented testa of some sorghum cultivars have been implicated in reducing protein digestibility but recently have been shown to promote human health because of their high antioxidant capacity and ability to fight obesity through reduced digestion. Combining quantitative trait locus mapping, meta-quantitative trait locus fine-mapping, and association mapping, we showed that the nucleotide polymorphisms in the *Tan1* gene, coding a WD40 protein, control the tannin biosynthesis in sorghum. A 1-bp G deletion in the coding region, causing a frame shift and a premature stop codon, led to a nonfunctional allele, *tan1-a*. Likewise, a different 10-bp insertion resulted in a second nonfunctional allele, *tan1-b*. Transforming the sorghum *Tan1* ORF into a nontannin *Arabidopsis* mutant restored the tannin phenotype. In addition, reduction in nucleotide diversity from wild sorghum accessions to landraces and cultivars was found at the region that codes the highly conserved WD40 repeat domains and the C-terminal region of the protein. Genetic research in crops, coupled with nutritional and medical research, could open the possibility of producing different levels and combinations of phenolic compounds to promote human health.

domestication | food production | gene cloning | health benefit | natural selection

Diets rich in phytochemicals are beneficial to human health because of their significant antioxidant properties (1, 2). These antioxidant compounds fall into three major categories: phenolic acids, flavonoids, and tannins. Of these, tannins account for about 19% of total dietary antioxidant capacity (3). Tannins, also known as condensed tannins or proanthocyanidins (PAs), are oligomers and polymers of flavan-3-ols. Beneficial effects from diets rich in tannin-containing foodstuffs include immunomodulatory and anticancer activity; antioxidant and radical scavenging functions; antiinflammatory, cardioprotective, vasodilating, and antithrombotic effects; and UV-protective functions (1, 4). Tannins are widespread throughout the plant kingdom, with diverse biological and biochemical functions, such as protection against predation from herbivorous animals and pathogenic attack from bacteria and fungi (5). Foliar tannin concentration has been shown to be significantly correlated with community phenotypes among diverse organisms (6). Tannins in fruits, vegetables, and certain beverages contribute the bitter flavor and astringency. Interestingly, tannins are also found in grains, such as sorghum [*Sorghum bicolor* (L.) Moench] with a pigmented testa layer, some finger millets, and barley, but not in major cereal crops, such as rice, wheat, and maize (7).

Although tannin content in sorghum grains can vary considerably among different genetic backgrounds (8), it is generally much higher than in other cultivated fruits, nuts, and grains (7) (Fig. 1). Tannin sorghums are often grown in hot, humid regions of Africa for their better resistance to grain mold and bird damage, and they have been used in many traditional products, such as porridges and alcoholic beverages (9). Because tannins in sorghum grains have been shown to decrease protein digestibility and feed efficiency in humans and animals, grain sorghum production as a feedstock in the United States has been almost entirely restricted to nontannin types, exemplifying strong artificial selection against tannins in breeding and production. However, a niche market has evolved to capitalize on the benefits of tannin sorghum through developing quality bread with high antioxidant and dietary fiber levels.

Although almost all wild sorghums contain condensed tannins in their grains, both tannin and nontannin types are naturally present in cultivated sorghums. Having a low level of digestion-reducing chemical compounds or complete removal of such compounds in edible parts of crops (e.g., cereal grains) is one of several domestication syndrome traits that differentiate domesticates from their wild progenitors (10, 11). Given the function of tannins in sorghum grain's chemical defense against bird predation and bacterial and fungal attack but their digestion-reducing qualities for human and animal consumption, the coexistence of tannin and nontannin sorghums suggests that elimination of this compound from sorghum grains during domestication is incomplete, unlike in other major cereals, such as rice, wheat, and maize. Similar phenomena have been documented for various traits in different crops (10). This outcome in sorghum may be a result of a balance between the two countering forces of artificial selection for the benefit of humans vs. natural selection for the benefit of plants.

Here, we report the cloning of the *Tannin1* (*Tan1*) gene for grain tannins in sorghum by genetic linkage mapping, fine-mapping through meta-quantitative trait locus (QTL) analysis, association validation with diverse genetic accessions, and transgenic complementation with an *Arabidopsis* nontannin mutant. Naturally occurring allelic variants in a highly conserved region of a WD40 protein cause frame shifts and premature stop

Author contributions: S.R.B., M.R.T., T.T.T., and J.Y. designed research; Yuye Wu, X.L., W.X., C.Z., Z.L., Yun Wu, and J.L. performed research; J.L., S.P., D.D.R., G.B., M.L.W., H.N.T., and S.R.B. contributed new reagents/analytic tools; Yuye Wu, X.L., W.X., C.Z., Z.L., and Yun Wu analyzed data; and Yuye Wu, X.L., G.B., and J.Y. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database [accession nos. JX122967 (*Tan1*), JX122968 (*tan1*-a), and JX122969 (*tan1*-b)].

[1]To whom correspondence should be addressed. E-mail: jyu@ksu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1201700109/-/DCSupplemental.

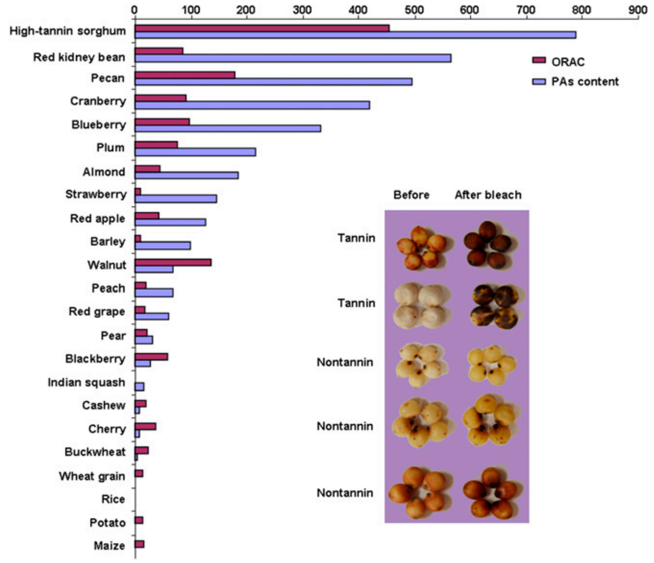

Fig. 1. PA content (milligrams of CE per 100 g) and antioxidant capacity [measured as oxygen radical absorbance capacity (ORAC), micromoles of TE per gram] across a variety of grains, nuts, and fruits. PA content is significantly (r = 0.82, P = 1.4 × 10^{-6}) correlated to ORAC value, and both parameters are higher in sorghum than in other foods. Data are from the US Department of Agriculture databases of selected foods in 2004 and 2010. (*Inset*) Bleach test for tannin determination. CE, catechin equivalents; TE, trolox equivalents.

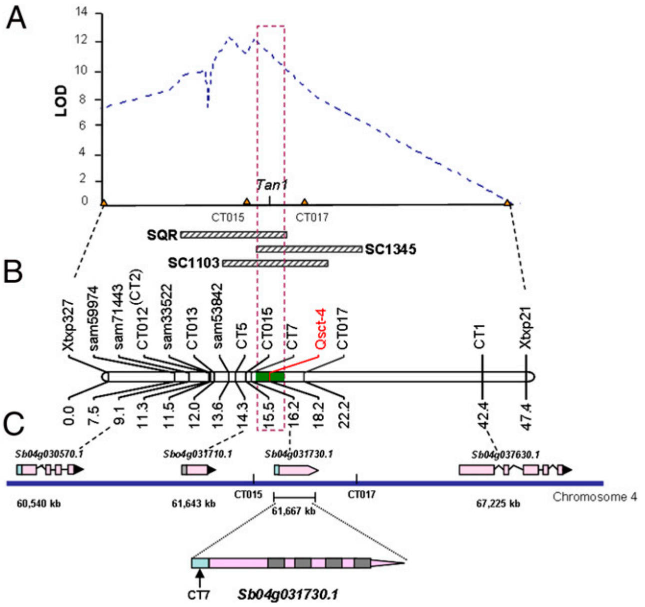

Fig. 2. Fine-mapping with the meta-QTL analysis identified a 25.4-kb genomic region that harbors the *Tan1* gene. (*A*) QTL mapping for the Tx430/ShanQuiRed population using the consensus map. (*B*) Consensus map and QTL confidence intervals (green band) identified on chromosome 4 (Qsct-4) across three sorghum mapping populations: Tx430/ShanQuiRed (SQR), Tx430/SC1345, and Tx430/SC1103. (*C*) *Sb04g031730.1* (*Tan1*) gene is the only gene identified jointly by the confidence interval of the meta-QTL analysis and the candidate gene search. LOD, logarithm of odds.

codons, resulting in truncated amino acid sequences and the absence of tannins in sorghum grains. Compared with wild relatives, sorghum landraces and cultivars exhibit reduced nucleotide diversity in the *Tan1* coding region. Our findings facilitate genetic research in relation to health, pharmaceutical, and nutritional values, as well as genetic manipulation of tannins in cereals.

Results

Genetic Cloning of *Tan1*. In the first mapping study, we identified two tannin QTLs using a recombinant inbred line (RIL) population derived from Tx430/ShanQuiRed. Tx430 is a nontannin inbred, and ShanQuiRed is a tannin inbred. These two QTLs explained 52% of the total phenotypic variation of tannin presence evaluated with the bleach test (12) (*SI Appendix*, Table S1). Both QTLs were further validated when tannin content was measured quantitatively with the vanillin-HCl test (*SI Appendix*, Fig. S1). Two additional populations were analyzed to fine-map these two QTLs, Tx430/SC1345 with 212 RILs and Tx430/SC1103 with 192 RILs. SC1345 and SC1103 are tannin parents. A different set of simple sequence repeat (SSR) markers from chromosomes 2 and 4 was then genotyped across the three populations. Composite interval mapping (CIM) was conducted independently for individual populations, followed by a meta-QTL analysis. The first consensus QTL (Qsct-2) was fine-mapped to a 797.3-kb region between common markers CT100 and CT072 on chromosome 2, and the second consensus QTL (Qsct-4) was fine-mapped to a 25.4-kb region between common markers CT015 and CT017 on chromosome 4 (Fig. 2 and *SI Appendix*, Fig. S2). We decided to pursue the chromosome 4 QTL further because Qsct-2 covered a wide range. In a complementary analysis, we identified 20 candidate genes involved in the PA biosynthetic pathway from *Arabidopsis* and other species. BLAST search of these genes against the sorghum genome sequence (13) detected four sorghum gene models: *Sb04g037630.1*, *Sb04g030570.1*, *Sb04g031710.1*, and *Sb04g031730.1* under the original chromosome 4 QTL region identified in Tx430/ShanQuiRed. These genes were predicted as homologs of *Arabidopsis TT12*, *Vitis vinifera LAR1*, *Arabidopsis LEC2* and *FUS3*, and *Arabidopsis TTG1* (*SI Appendix*, Table S2). All four genes were then mapped across three mapping populations using sequence-tagged site markers CT1, CT2, CT5, and CT7, respectively.

Combining results from the meta-QTL analysis and the candidate-gene search revealed that CT1 (*Sb04g037630.1*, *Arabidopsis TT12*) was 5,539.3 kb from the flanking marker CT017, CT2 (*Sb04g030570.1*, *V. vinifera LAR1*) was 1,119.2 kb from CT015, and CT5 (*Sb04g031710.1*, *Arabidopsis LEC2* and *FUS3I*) was 17.2 kb from CT015 but outside the Qsct-4 region flanked by common markers. Only CT7 (*Sb04g031730.1*, *Arabidopsis TTG1*) was located right in the consensus QTL Qsct-4 region, suggesting that *Sb04g031730.1* is likely to be the causal gene. Sequence annotation indicated that *Sb04g031730.1* encodes a WD40 protein; thus, it was designated as *Tannin1* (*Tan1*).

Allelic Variation at *Tan1*. To identify whether the sequence polymorphisms in *Tan1* are responsible for the tannin presence/absence phenotype in sorghum grains, we sequenced the *Sb04g031730.1* from −1,409 nt to +1,222 nt in a 24-accession set that contains 6 tannin and 18 nontannin accessions (*SI Appendix*, Table S3). Multiple sequence alignment revealed 26 SNPs and indels. Of these polymorphisms, 23 were found in the promoter region and 3 in the exon (*SI Appendix*, Table S4). In the exon region, a G deletion at position 580 nt (194 aa) results in a frame shift in the second WD40 repeat domain. The frame shift introduces a premature stop at the fourth WD40 repeat domain, resulting in a truncation of 55 amino acid residues from the C-terminal region (Fig. 3). A G-to-T transition at position 798 nt (266 aa) is in complete linkage disequilibrium (LD) with the deletion, forming a haplotype block. However, because the deletion site precedes the transition site and the transition itself is synonymous, it is not likely to be the causal site. An earlier mutagenesis study with *Arabidopsis TTG1* indicated that the

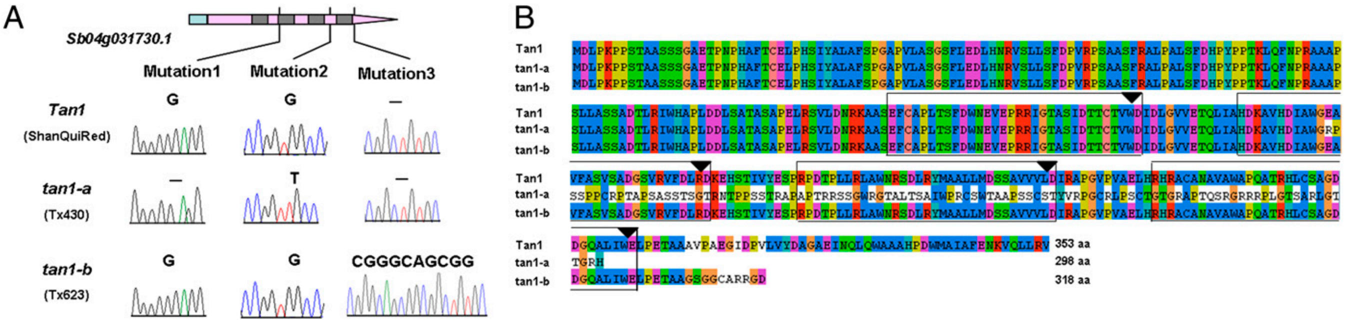

Fig. 3. Nucleotide polymorphisms in the *Tan1* gene coding region and their function prediction. (*A*) Three mutations and sequence chromatograms within the *Tan1* coding region. (*B*) Amino acid polymorphisms in the three alleles (*Tan1*, *tan1-a*, and *tan1-b*). Four WD40 repeat domains are indicated with boxes and solid triangles.

C-terminal region is vital for the structure and function of the WD40 protein (14). Accordingly, we hypothesized that the 1-bp deletion is likely the causal site in determining tannin presence/absence in the mapping populations.

All six tannin accessions have the WT *Tan1* haplotype found in ShanQuiRed, which agrees with the known complementary effect of two tannin loci from classic segregation analysis in sorghum (15). Tannin presence requires functional alleles at this locus and another locus, but one or both loci being homozygous recessive results in nontannin. The mutant haplotype (designated *tan1-a*) was found in all but three of the nontannin accessions (Tx623, Macia, and Malisor 84-7), which harbor a different mutant allele, a 10-bp insertion (CGGGCAGCGG) in the exon region (designated *tan1-b*) (Fig. 3). Similarly, this 10-bp insertion causes a frame shift at position 923 nt (309 aa) and, finally, a truncated protein with a length of only 318 aa. Predicted protein structures of different *Tan1* alleles further visualized the effect of the frame shifts in *tan1-a* and *tan1-b*, which led to disruption of the WD-40 protein structure (*SI Appendix*, Fig. S3). Further analysis of the protein sequences coded by *tan1-a* and *tan1-b* and seven independent mutations in the *Arabidopsis TTG1* gene revealed that the truncation of the C-terminal region is the common feature of the nonfunctional alleles in both species (*SI Appendix*, Fig. S4).

Gene Expression Analysis. We analyzed *Tan1* expression in different tissues: mature leaf from a flowering plant, immature panicle before heading, seed coat 15 d after pollination, and seed coat 30 d after pollination. *Tan1* was expressed in all these tissues, but the expression levels increased during panicle and seed coat development (*SI Appendix*, Fig. S5). In *Arabidopsis*, a WD40 repeat protein, together with an R2R3-MYB transcription factor and a basic helix–loop–helix domain protein, form a complex (TTG1-TT2-TT8) to regulate the late flavonoid pathway and PA-specific pathway gene *ANR* (16). To determine whether sorghum *Tan1* has a similar function, we analyzed expression of nine sorghum genes in ShanQuiRed and Tx430 (*SI Appendix*, Table S5). Among these genes, *SbLAR* is specific for the PA pathway; the other eight genes are either from the late anthocyanin pathway or involved in pathway regulation and transportation. *SbLAR* and two late anthocyanin pathway genes, *SbCHI* and *SbANS*, were not expressed in tannin-absent Tx430. Two other genes, *SbCHS* and *SbDFR*, were down-regulated in Tx430, and *SbF3H* was only weakly expressed at the seed coat 1 stage in Tx430. This result suggests that *Tan1* may have a similar regulatory function in the anthocyanin and PA pathways in sorghum seed coat.

Association Analysis Across Diverse Accessions. Because both tannin and nontannin types are present naturally in cultivated sorghum, we further expanded our examination to a population of 161 diverse sorghum accessions. This diverse panel represents major cultivated races and important US breeding lines and their progenitors (17) (*SI Appendix*, Table S6). Of the 161 sorghum accessions, 84 are tannin types and 77 are nontannin types. The frequency of the *Tan 1* allele is 78%, and that of the *tan1-a* allele is 20%. The *tan1-b* allele has a low frequency of 2%. Association of all polymorphisms found in the *Tan1* promoter and exon regions with tannin presence revealed consistent signals (Fig. 4). LD analysis further showed that many of the association signals from the promoter regions were attributable to the LD between these SNPs and the first two exonic polymorphisms (*tan1-a*), which had the strongest association ($P = 4.1 \times 10^{-11}$). The association signal at the third exonic mutation site (*tan1-b*) was not significant because of its low frequency. When three haplotypes were analyzed together, a significant association of *Tan1* and tannin phenotype was evident ($P = 2.9 \times 10^{-12}$). Notably, all 84 accessions exhibited the same sequence encoded by the *Tan1* allele in ShanQuiRed, indicating a strong conservation. Because nontannin sorghums must be homozygous recessive at one or both of the two complementary genes (15), 35 nontannin inbreds with *tan1* (*tan1_a* or *tan1_b*) alleles have the *tan1tan1_ _* genotype. Likewise, the other 42 nontannin sorghum accessions with the *Tan1* allele are expected to be of the *Tan1Tan1tan2tan2* genotype (*SI Appendix*, Table S6). Taken together, the complete conservation of the *Tan1* allele among tannin accessions, the distribution pattern of the *Tan1* and *tan1* (*tan1-a* and *tan1-b*) alleles among nontannin accession, and the strong gene-phenotype association provided further evidence that *Tan1* controls tannin development in sorghum grains.

***Tan1* Transgenic Complementation Experiment.** To validate that the identified sequence polymorphisms between *Tan1* and *tan1* are the causal sites, we introduced the ShanQuiRed *Tan1* ORF under the control of the 35S promoter into the *Arabidopsis ttg1-1* mutant in the Landsberg *erecta* (L*er*) background (Fig. 5). Among available mutants, the *ttg1-1* mutant is the best *Arabidopsis* mutant for complementation experiments because it has a C-terminal truncation of 25 aa residues, which is shorter than truncations in other mutants (*SI Appendix*, Fig. S4). The *ttg1-1* mutant exhibits four characteristics: yellow seeds attributable to the absence of tannins in seed coat, severe trichome differentiation, lack of anthocyanin pigmentation, and seed coat mucilage (18). The transgenic *Arabidopsis* containing the 35S-*Tan1-ORF* construct restored three of the four phenotypes: brown seed color from tannins in the seed coat, less trichome differentiation, and anthocyanin pigmentation (Fig. 5). Because the sorghum *Tan1* allele restored most of the phenotypes in the *ttg1-1* mutant and because protein sequence changes in the *tan1-a* and *tan1-b* are more severe than in the *ttg1-1*, this complementation experiment further confirmed that the sequence polymorphisms among *Tan1*, *tan1-a*, and *tan1-b* are responsible for the tannin phenotype in sorghum grains.

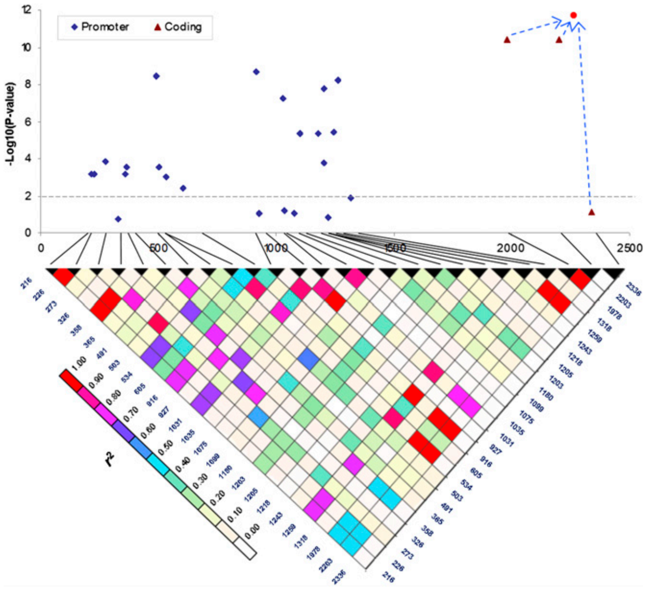

Fig. 4. Association mapping of nucleotide polymorphisms within the *Tan1* gene and tannins across a diverse set of sorghum accessions. LD between the causal sites in the coding region and other polymorphic sites in the promoter region results in the association signals at these noncausal sites. The red dot is the test with three haplotypes (*Tan1*, *tan1-a*, and *tan1-b*).

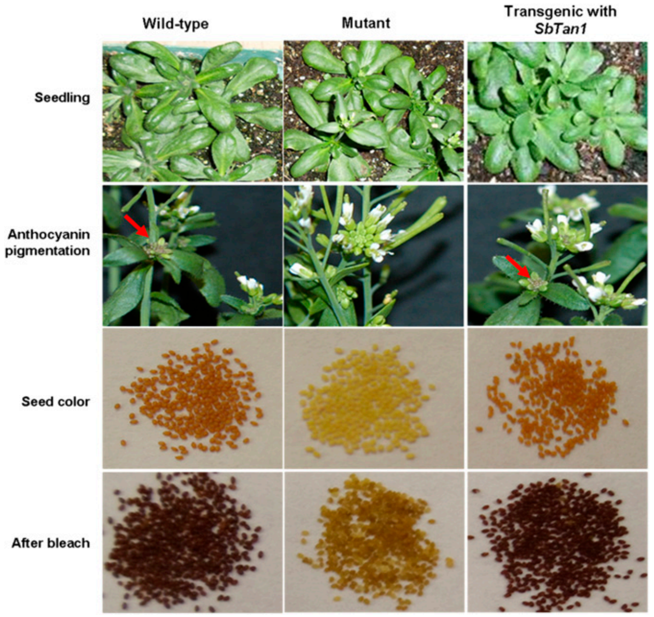

Fig. 5. Functional complementation of the *Tan1* allele to an *Arabidopsis* nontannin *ttg1* mutant. The 35S-*Tan1*-ORF construct was transformed into the mutant (CS89) and restored the WT phenotypes: anthocyanin pigmentation and tannin (indicated by seed color and bleach test). (*Left*) WT L*er* is shown. Red arrows indicate the anthocyanin pigmentation in the buds.

Phylogenetic Analysis of *Tan1* Orthologs. Because the PA pathway is present in many species, we further conducted phylogenetic analysis of the sorghum *Tan1* gene and its homologs at amino acid sequence level. Clustering of the two clades, monocot and eudicot, is evident, although SbiTAN1's homologs are highly conserved among all 23 plant species (*SI Appendix*, Fig. S6 and Table S7). The more diverged N-terminal region from positions 1–150 occupies the outer layer of the protein scaffold (*SI Appendix*, Fig. S7), which indicates that the function specificity of the WD40 protein in coordinating multiprotein complex assemblies may have been acquired after the divergence of monocot and eudicot from the ancestral copy. The more conserved WD40 repeat domains and the C-terminal region within each group (and to a lesser degree between groups) occupy the inner layer of the protein scaffold, agreeing with its critical role as the functional domain. This is consistent with the loss of function caused by the natural mutations in *tan1-a* and *tan1-b* alleles in sorghum, and the induced mutation in *ttg1-1* mutant *in Arabidopsis*. The C-terminal region was affected in all three cases.

Nucleotide Diversity Analysis. To reveal the dynamics of tannin presence and *Tan1*, we further obtained the full-length *Tan1* genomic sequences across the 18 sorghum landraces; 86 cultivars (a subset of the 161 accessions used in association mapping); and a set of 18 diverse accessions within the genus *Sorghum*, including wild (*S. bicolor* subsp. *verticilliflorum*), weedy (*S. bicolor* nothosubsp. *drummondii*), and other sorghum relatives (e.g., *Sorghum propinquum*, *Sorghum halepense,* and *S. bicolor* × *S. halepense*) (*SI Appendix*, Tables S8–S10). One weedy accession and another accession with an incomplete record were found to be nontannin (likely attributable to introgression), and both have the *tan1-a* allele.

Although similar levels of genetic diversity were observed in the *Tan1* promoter regions across three groups, both nucleotide diversity π and θ_w (an estimate of $4N_e\mu$, where N_e is the effective population size and μ is the mutation rate per nucleotide) for sorghum landraces and for cultivars, but not for the wild group, approached zero in the part of the *Tan1* coding region that codes the conserved WD40 repeat domains and the C-terminal region (Fig. 6 and *SI Appendix*, Fig. S8 and Table S11). No such differences were found in the neutral gene, *Adh1*. There are essentially no other nucleotide polymorphisms within the accessions examined in *S. bicolor* subsp. *verticilliflorum*, *S. bicolor* nothosubsp. *drummondii*, and *S. halepense*, other than the presence of *tan1-a* in one weedy *drummondii* accession. Sequence polymorphisms were present in other sorghum relatives. Across the whole set of 18 accessions within the genus *Sorghum*, Tajima's D test was significant ($P < 0.01$) for the region that codes the four WD40 repeat domains and the C-terminal region, indicating the presence of rare mutations. This agreed with the absence of nucleotide polymorphisms in this region, other than what was in *tan1-a* and *tan1-b*, across the cultivars and sorghum landraces, which is a much narrower set than the wild set.

Discussion

Tannins and Multiple Alleles of *Tan1*. Several factors make tannins an important research subject: their antioxidant capacity and relevant health benefits, their natural occurrence in a few cereal crops, and their role in sorghum production. For example, knowledge of tannins in biosynthesis pathways can be used to generate lines that produce high-content tannins in sorghum and other cereals to promote health through their antioxidant capacity and to fight obesity through reduction of digestion. Although many structural and regulatory genes have been identified in the model species *Arabidopsis* through the mutational approach (1, 19), knowledge of the genetic control of tannins remains limited in cereal crops (20). In the current study, the sequence variations in *Tan1* gene are proved to be responsible for the presence of tannins in sorghum grains by combining meta-QTL analysis, association analysis, and a functional complementation test.

The conservation of the *Tan1* allele among all tannin sorghums indicates a strong purifying selection without human intervention, because natural selection for self-propagation of the wild progenitors would have favored the tannin type. On the other hand, artificial selection to reduce digestion-reducing compounds would have favored the mutations that eliminate tannins in sorghum grains for better nutrient intake. The low sequence diversity within

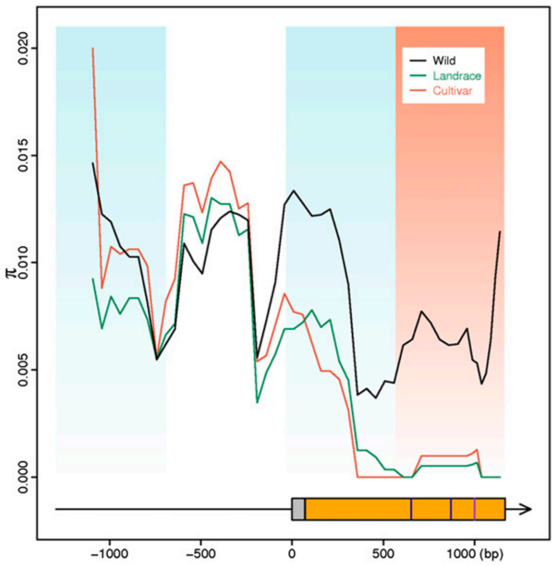

Fig. 6. Sliding-window analysis of sequence diversity at the *Tan1* locus in wild sorghums, landraces, and cultivars. The full-length *Tan1* genome sequence was obtained for this analysis. Nucleotide diversity (π) values approach zero in the *Tan1* coding region in landraces (18 accessions) and cultivars (87 accessions) but remain at 0.007 in the set (18 accessions) of wild, weedy sorghums and sorghum relatives.

the *Tan1* coding region in sorghum cultivars and landraces, the highly conserved WD40 repeat domain across species, and the allelic spectrum (*Tan1*, *tan1-a*, and *tan1-b*) support the relevance of *Tan1* to natural and artificial selection.

In the current study, we identified two different naturally occurring nonfunctional alleles, *tan1-a* and *tan1-b*, and quantified their frequencies across a set of wild sorghums, landraces, and cultivated accessions. Unlike other major cereals, including rice, maize, and wheat, the nontannin type is not completely fixed in cultivated sorghum even though it has a significantly higher frequency in sorghum cultivars and landraces than in wild sorghums. This is possibly attributable to the balance between artificial selection and natural selection. This balance may also explain why a signature of selection by the Hudson-Kreitman-Agande (HKA) test was not found in *Tan1* but the reduction in sequence diversity in the *Tan1* coding region from wild relatives to landraces and cultivars was evident. Previous analyses indicated that the selection screen approach to detect domestication-related genes may not be suitable for sorghum (11). It is well known that there is a very high level of interfertility and intermixing of the crop–weed–wild complex among domesticated sorghums, weedy types, and their wild relatives (21), which may also have contributed to this compromise.

Combining Multiple Approaches in *Tan1* Identification. Our cloning of the *Tan1* gene revealed both opportunities and challenges to identify genes underlying agronomic traits in major crops. Many genes and their functional polymorphisms have been identified in model plant *Arabidopsis* species and in model crops (22). These findings provided an inventory of genes, functional nucleotide polymorphisms, and genetic mechanisms that might be involved in crop adaptation to different agricultural or natural environments. Translating these findings into major crops and traits with agronomic, evolutionary, and medical importance requires further research. For example, previous research with artificially induced mutants has identified several WD40 genes, including petunia *An11*, *Arabidopsis TTG1*, maize *PAC1*, *Medicago trunculata* MtWD40-1, and grape WDR1 and WDR2 (23), but information about the number and frequency of different alleles cannot be obtained because of the lack of either naturally occurring tannin or nontannin types (18, 20). The presence of tannin and nontannin types in sorghum and tannins' role in feed production and human health provided a translational opportunity. Our cloning of *Tan1* in sorghum benefited from the available sorghum genome sequence, well-characterized pathways, and mutant stocks and transformation methods in *Arabidopsis*. The combination of these genomic resources with meta-QTL analysis (24) and association mapping (25) allowed us to identify the *Tan1* gene and the causal sequence polymorphisms efficiently.

On the other hand, we also demonstrated the potential challenge of allele frequency in gene cloning. Had we used a nontannin sorghum line containing the *tan1_b* allele (e.g., Tx623) in the initial crosses, validating the gene cloning results with the diversity-based association analysis might have been difficult. Without a thorough haplotype analysis and function prediction, the low-frequency *tan1_b* would have prevented us from detecting a significant signal at the 10-bp insertion site. Furthermore, unlike the induced-mutation dissection approach, the presence of multiple alleles with different natural mutation sites should be considered in gene mapping in crops. Interestingly, in another gene cloning study in sorghum, three different alleles of *PRR37* were identified to remove flower suppression on long days (26). The identification of multiple naturally occurring alleles in both studies demonstrates the wide genetic diversity in sorghum.

Recent studies have indicated that tannins from sorghum are more bioavailable and beneficial to human health (e.g., because of anticarcinogenic activity against human melanoma cells and as anticaloric agents for obese individuals) than previously thought (27, 28), although further research on the exact mechanisms is needed. Sorghum bran is a low-cost, high-content source of antioxidant and antiinflammatory compounds that fight chronic diseases, such as inflammation and high oxidative stress (29). In an ongoing study, we are attempting to identity the gene on chromosome 2 and to exploit the natural genetic variation potential of tannin content further. These results can facilitate future genetic research and breeding efforts to combine tannin compounds and agronomic properties in unique sorghum lines for potential nutraceutical applications.

Materials and Methods

QTL Mapping Populations. ShanQuiRed originated in the low-temperature and high-latitude regions of China and produces seeds with tannins. SC1345 and SC1103 were from the sorghum conversion program, and both have tannins in the grain. Tx430 was an adapted sorghum inbred line in the United States without tannins. The RIL populations were obtained by single-seed descent: Tx430/ShanQuiRed with 109 lines, Tx430/SC1345 with 212 RILs, and Tx430/SC1103 with 192 RILs. The presence of tannins in sorghum grains was performed using the standard bleach test. The seeds of three populations were obtained from Manhattan, KS, in 2004, 2008, and 2009. All bleach tests were repeated three times.

Leaf tissue was collected at the three-leaf stage, dried in a freeze-drier (ThermoSavant) for 3 d, and ground to fine powder in a Mixer Mill (Retsch GmbH) for 5 min at 27 times per second with the aid of a 3.2-mm metal bead in each tube. DNA was extracted using a modified cetyltrimethylammonium bromide (CTAB) method. Initial QTL mapping with the Tx430/ShanQuiRed was conducted with mostly amplified fragment length polymorphism markers (30). For SSR genotyping, amplified PCR fragments were separated in the ABI Prism 3730 DNA sequencer (Applied Biosystems). SSR data were analyzed using GeneMarker software, version 1.5 (SoftGenetics LLC).

The genetic linkage map was constructed with JoinMap (31). For chromosome 2, the numbers of polymorphic SSR markers were 14 (Tx430/ShanQuiRed), 7 (Tx430/SC1345), and 10 (Tx430/SC1103); for chromosome 4, the numbers of polymorphic SSR markers were 16 (Tx430/ShanQuiRed), 10 (Tx430/SC1345), and 9 (Tx430/SC1103) (*SI Appendix*, Table S12). The CIM method in Windows QTL Cartographer (http://statgen.ncsu.edu/qtlcart/WQTLCart.htm) was used for QTL mapping in each population. With results from CIM, multiple interval mapping was used to estimate the additive and epistatic effects.

Meta-QTL Analysis. Consensus mapping was built with JoinMap using a fixed order of SSR markers according to their positions on the genome sequence. MetaQTL (32) was then used to conduct meta-QTL analysis. Command QTLproj was used to project the QTL position and confidence interval estimated for the individual mapping population onto the consensus map. Command QTLclust was used to fit the observed QTLs into the Gaussian mixture model (33) to identify the final consensus QTLs, Qsct-2 and Qsct-4.

Association Mapping and Nucleotide Diversity. For association mapping, we sequenced the *Tan1* gene across 161 diverse sorghum accessions. Multiple sequence alignment was conducted with ClustalX 2.0 (http://www.clustal.org/clustal2). Association of the SNPs and the tannin was conducted with the χ^2 test.

For nucleotide diversity analysis, we obtained the full-length *Tan1* genomic sequence across a set of 18 accessions, including wild, weedy sorghum and sorghum relatives. The full-length *Tan1* genomic sequence was also obtained across 86 sorghum cultivars (a subset of the 161 accessions was used in association analysis) and 18 sorghum landraces. Sliding-window analysis of π and θ_w was conducted using a 400-bp window with a 50-bp step.

Phylogenetic Analysis. The WD40 repeat gene family is known to be involved in diverse functions in plants. Accordingly, only the orthologs in other plants with the highest similarity score to SbiTAN1 were examined in our phylogenetic analysis. The neighbor-joining tree was built on the basis of aligned protein sequences by ClustalX 2.0.

RNA Extraction and Gene Expression Analysis. Total RNA from various organs was extracted by the RNeasy Plant Mini Kit (Qiagen), and 2 μg of RNA was used to synthesize first-strand cDNA by SuperScript II RT (Invitrogen) in a 20-μL reaction mixture. Two microliters of the RT reaction was used as a template in a 25-μL PCR solution. The sorghum *Actin* gene was used as the control. The PCR primers for *Tan1* are as follows: Sbct3 (5′-CACCAAGCTCCAGTTCAACC-3′) and Sbct4 (5′-GCCATATAGCGGAGGTCAGA-3′) (*SI Appendix*, Table S6).

Complementation Experiment. A 1.1-kb cDNA fragment of the *Tan1* coding region from ShanQuiRed was amplified by the PCR method using specific primers Sbct5 (5′-CCCGATTTCTCCACCCCATGGACCTACC-3′) and Sbct6 (5′-CACCATGGTACCAACCTTGTCAGACCCT-3′). The resulting fragment was cloned into binary expression vector PBI121 by replacing the *GUS* gene using XbaI and SacI sites. The resulting overexpressing vector PBI121-*Tan1* contained the 1.1-kb *Tan1* gene, driven by a constitutive 35S cauliflower mosaic virus (CaMV35S) promoter. After confirmation with sequencing, the construct was introduced into *Agrobacterium tumeficience* LBA4404. The *Arabidopsis ttg1-1* mutant plants were transformed with LBA4404-harboring PBI121-*Tan1* vector using the floral dip procedure (34). The *ttg1-1* stock (CS89) was obtained from the *Arabidopsis* Biological Resource Center. *Arabidopsis* plants were grown in a growth chamber with a light/dark cycle comprising 16 h of 120 $\mu E \cdot m^{-2} \cdot s^{-1}$ light and 8 h of dark with fluorescent lighting and at 22 °C. To determine the effect of *Tan1* on anthocyanin pigmentation, 5-wk-old seedings of the L*er*, CS89, and transgenic line were subjected to cold stress at 12 °C for 2 wk. The anthocyanin pigmentation of the flower buds was screened after returning the seedlings to 22 °C. In addition to the bleach test results (Fig. 5), we further confirmed the tannin phenotype restoration with the vanillin-HCl test, which quantifies the tannin content in L*er*, CS89, the transgenic line, and Columbia (Col) (*SI Appendix*, Table S13).

ACKNOWLEDGMENTS. We thank Dr. Lloyd Rooney for providing information on food and feed quality attributes of sorghum. This work was supported by Agriculture and Food Research Initiative Competitive Grant 2011-03587 from the US Department of Agriculture National Institute of Food and Agriculture; the Plant Feedstock Genomics Program (DE-SC0002259) of the US Department of Energy; and the Plant Genome Program (DBI-0820610) of the National Science Foundation, the US Department of Agriculture-Agricultural Research Service, and the National Sorghum Checkoff program.

1. Dixon RA, Xie DY, Sharma SB (2005) Proanthocyanidins—A final frontier in flavonoid research? *New Phytol* 165:9–28.
2. Crozier A, Jaganath IB, Clifford MN (2009) Dietary phenolics: Chemistry, bioavailability and effects on health. *Nat Prod Rep* 26:1001–1043.
3. Floegel A, et al. (2010) Development and validation of an algorithm to establish a total antioxidant capacity database of the US diet. *Int J Food Sci Nutr* 61: 600–623.
4. Sharma SD, Meeran SM, Katiyar SK (2007) Dietary grape seed proanthocyanidins inhibit UVB-induced oxidative stress and activation of mitogen-activated protein kinases and nuclear factor-kappaB signaling in in vivo SKH-1 hairless mice. *Mol Cancer Ther* 6:995–1005.
5. Xie DY, Sharma SB, Paiva NL, Ferreira D, Dixon RA (2003) Role of anthocyanidin reductase, encoded by BANYULS in plant flavonoid biosynthesis. *Science* 299:396–399.
6. Whitham TG, et al. (2006) A framework for community and ecosystem genetics: From genes to ecosystems. *Nat Rev Genet* 7:510–523.
7. Dykes L, Rooney LW (2007) Phenolic compounds in cereal grains and their health benefits. *Cereal Foods World* 52:105–111.
8. Gu L, et al. (2004) Concentrations of proanthocyanidins in common foods and estimations of normal consumption. *J Nutr* 134:613–617.
9. Awika JM, Rooney LW (2004) Sorghum phytochemicals and their potential impact on human health. *Phytochemistry* 65:1199–1221.
10. Gepts P (2004) Crop domestication as a long-term selection experiment. *Plant Breed Rev* 24:1–44.
11. Doebley JF, Gaut BS, Smith BD (2006) The molecular genetics of crop domestication. *Cell* 127:1309–1321.
12. Waniska R, Hugo LF, Rooney LW (1992) Practical methods to determine presence oftannins in sorghum. *J Appl Poult Res* 1:122–128.
13. Paterson AH, et al. (2009) The Sorghum bicolor genome and the diversification of grasses. *Nature* 457:551–556.
14. Walker AR, et al. (1999) The TRANSPARENT TESTA GLABRA1 locus, which regulates trichome differentiation and anthocyanin biosynthesis in Arabidopsis, encodes a WD40 repeat protein. *Plant Cell* 11:1337–1350.
15. Hahn DH, Rooney LW (1986) Effect of genotype on tannins and phenols of sorghum. *Cereal Chemistry* 63:4–8.
16. Baudry A, et al. (2004) TT2, TT8, and TTG1 synergistically specify the expression of BANYULS and proanthocyanidin biosynthesis in Arabidopsis thaliana. *Plant J* 39: 366–380.
17. Casa AM, et al. (2008) Community resources and strategies for association mapping in sorghum. *Crop Sci* 48:30–40.
18. Carey CC, Strahle JT, Selinger DA, Chandler VL (2004) Mutations in the pale aleurone color1 regulatory gene of the Zea mays anthocyanin pathway have distinct phenotypes relative to the functionally similar TRANSPARENT TESTA GLABRA1 gene in Arabidopsis thaliana. *Plant Cell* 16:450–464.
19. He F, Pan QH, Shi Y, Duan CQ (2008) Biosynthesis and genetic regulation of proanthocyanidins in plants. *Molecules* 13:2674–2703.
20. Sweeney MT, Thomson MJ, Pfeil BE, McCouch S (2006) Caught red-handed: Rc encodes a basic helix-loop-helix protein conditioning red pericarp in rice. *Plant Cell* 18: 283–294.
21. Barnaud A, et al. (2009) A weed-crop complex in sorghum: The dynamics of genetic diversity in a traditional farming system. *Am J Bot* 96:1869–1879.
22. Alonso-Blanco C, et al. (2009) What has natural variation taught us about plant development, physiology, and adaptation? *Plant Cell* 21:1877–1896.
23. Hichri I, et al. (2011) Recent advances in the transcriptional regulation of the flavonoid biosynthetic pathway. *J Exp Bot* 62:2465–2483.
24. Chardon F, et al. (2004) Genetic architecture of flowering time in maize as inferred from quantitative trait loci meta-analysis and synteny conservation with the rice genome. *Genetics* 168:2169–2185.
25. Zhu C, Gore MA, Buckler ES, Yu J (2008) Status and prospects of association mapping in plants. *The Plant Genome* 1:5–20.
26. Murphy RL, et al. (2011) Coincident light and clock regulation of pseudoresponse regulator protein 37 (PRR37) controls photoperiodic flowering in sorghum. *Proc Natl Acad Sci USA* 108:16469–16474.
27. Ross JA, Kasum CM (2002) Dietary flavonoids: Bioavailability, metabolic effects, and safety. *Annu Rev Nutr* 22:19–34.
28. Gómez-Cordovés C, Bartolomé B, Vieira W, Virador VM (2001) Effects of wine phenolics and sorghum tannins on tyrosinase activity and growth of melanoma cells. *J Agric Food Chem* 49:1620–1624.
29. Burdette A, et al. (2010) Anti-inflammatory activity of select sorghum (Sorghum bicolor) brans. *J Med Food* 13:879–887.
30. Ridder DD (2005) Early-season cold tolerance in grain sorghum: The relationship with seed characteristics and the evaluation of molecular tools for breeding. Master's thesis (Kansas State University, Manhattan, KS).
31. van Ooijen JW, Voorrips RE (2001) *Joinmap v3.0, Software for the Calculation of Genetic Linkage Maps* (Plant Research International, Wageningen, The Netherlands).
32. Veyrieras JB, Goffinet B, Charcosset A (2007) MetaQTL: A package of new computational methods for the meta-analysis of QTL mapping experiments. *BMC Bioinformatics* 8:49.
33. Goffinet B, Gerber S (2000) Quantitative trait loci: A meta-analysis. *Genetics* 155: 463–473.
34. Clough SJ, Bent AF (1998) Floral dip: A simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant J* 16:735–743.