Dominant Awn Inhibitor, which encodes an ALOG protein in sorghum, suppresses awn formation in rice

Hideki Takanashi[1], Hiromi Kajiya-Kanegae[1,†], Asuka Nishimura[1], Junko Yamada[1], Motoyuki Ishimori[1], Masaaki Kobayashi[2], Kentaro Yano[2], Hiroyoshi Iwata[1], Nobuhiro Tsutsumi[1,*], Wataru Sakamoto[3,*]

[1]Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan

[2] Department of Life Sciences, Faculty of Agriculture, Meiji University, Kawasaki, Kanagawa 214-8571, Japan

[3] Institute of Plant Science and Resources, Okayama University, Kurashiki, Okayama 710-0046, Japan

***Corresponding authors**: saka@okayama-u.ac.jp (W.S.) and atsutsu@g.ecc.u-tokyo.ac.jp (N.T.)

Footnote

[†]Present address: Research Center for Agricultural Information Technology, National Agriculture and Food Research Organization, Minato-ku, Tokyo 105-0003, Japan

Running head: *Dominant Awn Inhibitor* in sorghum

Number of black and white figures:	0
Numbers of color figures:	6
Number of tables:	1
Number of supplementary figures:	3
Numbers of supplementary tables:	2

Abstract

The awn, a needle-like structure that extends from the tip of the lemma in grass species, is believed to play a role in environmental adaptation and fitness. Wild species of cereal crops tend to have longer awns, whereas modern cultivars display a variety of awn lengths and morphologies. Awns in some crops seem to have been eliminated during domestication, and genes controlling awn formation have been reported. Interestingly, a few single dominant factors that eliminate awns are already known; in sorghum (*Sorghum bicolor*), for example, the dominant awn inhibitor has been known for a century, though its molecular identity remains uncharacterized. In this study, we conducted quantitative trait locus (QTL) analysis and a genome-wide association study of awn-related traits in sorghum. Among the three QTLs controlling awn presence and length, one of them, mapped on chromosome 3, had an extraordinary effect on eliminating awns. We provide a line of evidence demonstrating that this QTL corresponds to *Dominant Awn Inhibitor* (*DAI*), encoding an ALOG family protein that potentially acts as a transcription factor. Detailed analysis of the ALOG protein family in cereals revealed that *DAI* was considered to originate from a gene duplication of its twin paralogue on chromosome 10. Interestingly, heterologous expression of *DAI* with its own promoter in rice inhibited awn formation in the awned cultivar Kasalath. Our finding that *DAI* was born by gene duplication provides an interesting example of gain-of-function, which occurs only in sorghum but shares its functionality with rice and sorghum.

Keywords

Awn, Dominant Awn Inhibitor, genome-wide association study (GWAS), quantitative trait locus (QTL), sorghum

Introduction

The awn is a needle-like structure that forms at the tip of the lemma in grass species. Awns, by their bristle-like architecture, are known to have various important functions, such as avoiding getting fed on by animals (Grundbacher 1963, Hua et al. 2015), helping seeds to burrow into the soil (Elbaum et al. 2007), and promoting seed dispersal (Sato et al. 1996, Cavanagh et al. 2020). Moreover, awns significantly contribute to photosynthesis and thereby to grain yield, especially in crops such as barley (Kjack and Witters 1974) and wheat (Motzo and Giunta 2002, Li et al. 2006). Wild species of cereal crops tend to have long awns, and numerous genetic studies in such crops have demonstrated that multiple factors regulate awn formation and elongation. Although important in wild species, awns in cereal crops often disturb modern agriculture by hindering manual harvesting and reducing the feed value of straw for livestock (Takahashi 1955). Consequently, awns have been partially or completely eliminated during domestication (Takahashi 1955).

Awn formation and elongation have been extensively characterized in rice at the molecular level, which indicates that the awnless cultivars were established by selecting the loss-of-function allele of these genes one after another during domestication (Amarasinghe et al. 2020, Bessho-Uehara et al. 2021). *An-1*/*REGULATOR OF AWN ELONGATION 1* (*RAE1*) encodes the bHLH transcription factor (Luo et al. 2013, Furuta et al. 2015) expressed at the apex of lemma primordia, which causes continuous cell division for awn elongation. *LONG and BARBED AWN1* (*LABA1*)/*An-2* encodes a cytokinin-activating enzyme that positively regulates awn elongation (Gu et al. 2015, Hua et al. 2015). *RAE2*/*GAD1*/*GLA*, which encodes epidermal patterning factor-like protein 1 (EPFL1), which acts as a small secretory peptide in panicles (Bessho-Uehara et al. 2016, Jin et al. 2016, Yano et al. 2016, Zhang et al. 2019). Two other genes, *DROOPING LEAF* (*DL*), encoding a YABBY gene and *OsETTIN2* (*OsETT2*), encoding an auxin-responsive factor, also enhance awn elongation in rice (Toriba and Hirano 2014), although whether these genes are involved in domestication is unclear. In barley, *short awn 2* (*lks2*, *for length 2*), which encodes a SHORT INTERNODES family transcription factor, has been shown to regulate awn length and pistil morphology (Taketa et al. 2011, Yuo et al. 2012). Genes involved in brassinosteroid biosynthesis or signaling have also been shown to affect awn length with pleiotropic effects (Dockter et al. 2014).

In contrast to these ‘positive’ factors for awn development, ‘inhibitory’ factors that diminish awn are also known, including *Hooded* (*Hd*), *Tipped1* (*B1*), and *Tipped2* (*B2*) in wheat (Watkins and Ellerton 1940, Kato et al. 1998, Sourdille et al. 2002, Yoshioka et al. 2017). *B1* was recently found to encode a C2H2 zinc-finger protein with ethylene-responsive element binding factor-associated amphiphilic repression (EAR) motifs which putatively functions as a transcriptional repressor (DeWitt et al. 2020, Huang et al. 2020). Interestingly, such inhibitory factors have not been reported in rice, despite extensive studies. These results imply that awn formation is controlled by a complex and distinct genetic network, whose modification leads to varietal diversification in a species-specific manner.

Among cereal crop species, sorghum (*Sorghum bicolor*) is not an exception and displays a variety of awn morphologies among cultivars; however, unlike the aforementioned advanced studies, the genes responsible for awn development are poorly understood. Sorghum is the fifth most important cereal in the world (faostat.fao.org) and is a promising C_4 crop with higher stress tolerance than other major cereals (Tuinstra et al.

1997, Ogbaga et al. 2014). In addition, it is a suitable genetic model for C_4 grasses because of its high morphological diversity and relatively small genome (approximately 732 Mb) compared to that in other C_4 grasses (Price et al. 2005, Paterson et al. 2009, McCormick et al. 2018). In fact, our recent work showed that a next-generation sequencing approach, restriction site-associated DNA sequencing (RAD-seq), combined with the use of recombinant inbred lines (RILs), allowed the establishment of a high-density genetic map to elucidate important traits and their responsible genes, including genes associated with biomass, spikelet formation, seed morphologies, and response to chemicals (Kajiya-Kanegae et al. 2020, Jing et al. 2021, Takanashi et al. 2021). We reasoned that awn development could also be investigated, since the parents used in the RILs displayed contrasting awn morphologies. As for awn-related traits in sorghum, three quantitative trait locus (QTL) analyses (Boivin et al. 1999, Tao et al. 2000, Hart et al. 2001) and three genome-wide association studies (GWAS) (Shehzad et al. 2009, Zhang et al. 2015, Girma et al. 2019) have been previously reported, but their molecular identity remains unclear. It is noteworthy that the existence of a dominant awn inhibitor locus was reported a century ago (Vinall and Cron 1921, Sieglinger et al. 1934), suggesting the emergence of awnless cultivars in sorghum breeding.

In the present study, we conducted QTL analysis using RILs and GWAS using the sorghum germplasm for awn-related traits to enrich our limited knowledge on awn regulation in sorghum. Based on our QTL analysis for awn presence and awn length, we provide evidence that a single gene, *Dominant Awn Inhibitor* (*DAI*), located on chromosome 3, inhibits awn formation. *DAI* is likely to have originated from the duplication of its paralogue in chromosome 10, whose role remains unclear. Surprisingly, heterologous expression of sorghum *DAI* into rice was shown to make the awned rice cultivar ‘Kasalath’ awnless. Together, these results suggest that common mechanisms may exist to inhibit awn formation in cereal crops, whereas the emergence of such inhibitors differs among species.

Results

Awned and awnless: morphology used in this study

Sorghum has two types of spikelets: sessile spikelets (SSs) directly attached to the inflorescence branch, and pedicellate spikelets (PSs) attached to the inflorescence branch via a pedicel. Unlike PS, which has only incomplete floral organs, SS has a complete floret consisting of a lemma, palea, two lodicules, three anthers, and a pistil, therefore only SS can produce a seed. The awn, which is a slender extension of the lemma, is also born only in SS. Recently, we established and reported a sorghum RIL, derived from a parental cross between BTx623 and the Japanese landrace NOG (Kajiya-Kanegae et al. 2020). Since parents of this population differed in various traits (Ohnishi et al. 2019, Kajiya-Kanegae et al. 2020), we considered this RIL population to be suitable for various QTL analyses. Interestingly, we found that there was a large difference in awn length (AL) between the parents of our RILs (Fig. 1A and 1 B; BTx623 is awnless, while NOG is awned). Therefore, QTL analysis of the awn-related traits using this RILs was performed, using two generations (F_6 and F_7) cultivated in Okayama (F_6, greenhouse, in 2014) and Tokyo (F_7, field, in 2015) to validate the reproducibility.

QTL analysis of the awn-related traits identified a prominent dominant allele that inhibits awn formation

Measurement of the AL of 245 RIL individuals (F_6 generation) revealed that about half of the RILs are awnless like BTx623 (Fig. 1C). We then performed composite interval mapping of QTL using this AL as phenotypic data, with 1,789 SNP markers determined by RAD-seq (Kajiya-Kanegae et al. 2020). We detected an extraordinary high QTL on chromosome 3, termed *qAL3* [the logarithm of the odds (LOD) score 116.4 and the percentage of phenotypic variance explained (PVE) 90.6%]. An additional QTL peak was found on chromosome 9 (*qAL9*: LOD score 7.0, PVE 1.3%) (Fig. 1D, Table 1). For *qAL3*, the effect of increasing AL was attributed to the NOG allele (Fig. 1E, left). It is noteworthy that ALs are zero in almost all individuals carrying the BTx623 homozygous or heterozygous allele at *qAL3*, strongly indicating its dominant effect on inhibiting awn formation or elongation. Confirming this, spikelets of F_1 plants derived from the cross between BTx623 and NOG were shown to have no awns (Supplementary Fig. 1A). For *qAL9*, the effect for AL was not clearly resolved, most likely due to the superior effect of *qAL3* (Fig. 1E middle). Nevertheless, the NOG allele for *qAL9* appeared to increase AL, when the allele of *qAL3* was limited to only the NOG type (Fig. 1E right).

The results of our QTL analysis suggest that *qAL3* determines the awn presence (AP) rather than AL; therefore, we conducted QTL analysis using data converted from AL data to AP (re-scored as awnless = 0, awned = 1) to verify this. Again, we detected a single QTL on chromosome 3 (*qAP3*: LOD score 179.1, PVE 97.8%), which had exactly the same position as *qAL3* but with a more intense effect (Fig. 1F, Table 1). The allelic effect of *qAP3* was similar as that of *qAL3*; the BTx623 allele dominantly led to an awnless phenotype (Fig. 1G). Together, these results led us to conclude that the single locus corresponding to *qAL3/qAP3* is a dominant inhibitor of awn formation.

Existence of two QTLs controlling awn length

To enrich our knowledge on the control of AL, we performed QTL analysis using AL only in awned lines. By subtracting awnless individuals, we sought to exclude the extraordinary effect of *qAL3/qAP3* and to find other QTLs that may control AL. As expected, we were able to detect two additional QTLs on chromosome 7 (*qAL7*: LOD score 8.3, PVE 23.8%) and 9 (*qAL9*: LOD score 5.9, PVE 16.2%) (Fig. 1H, Table 1). Given that the property of *qAL9* detected here is quite similar to that of *qAL9* in the original QTL analysis with awnless individuals (Table 1), both were considered to represent the same QTL. Effects of increasing AL were attributed to the NOG alleles for both *qAL7* and *qAL9* (Fig. 1I). The QTLs presented here (*qAL3*/*qAP3*, *qAL7*, and *qAL9*) were also detected in the QTL analysis using the datasets of the 2015 cultivation in Tokyo, indicating high reproducibility for all QTLs (Supplementary Fig. 1B–H, Table 1). Based on these findings, the dominant inhibitory effect of *qAL3/AP3* is apparent, and we further characterized *qAL3/AP3* hereafter. As for *qAL7* and *qAL9*, their presumed actions are described in Discussion.

Reconfirming the effect of qAL3/qAP3 by establishing NIL

Because several RILs were shown to be heterozygous for *qAL3/AP3* by RAD-seq in the F_6 population, we next attempted to establish a near isogenic line (NIL). Of these, we selected an awnless heterozygous line at the F_6 generation, RIL205 (Fig. 2A), and followed the segregation of *qAL3/AP3* in the subsequent selfed generations. Consequently, observations of spikelets of the NILs (F_9 generation) showed that the NIL with the homogeneous BTx623 allele (NIL_B) lost awns, while that with the homogeneous NOG allele (NIL_N) had awns (Fig. 2B–E). These results suggest that there is indeed a gene in the *qAL3*/*qAP3* locus that inhibits AL or AP. Dissection of immature spikelets in NIL_N indicated that awns appear before and after the length of the sessile spikelet reaches approximately 2.0 mm (between Stage 5 and 6; Fig. 2B), implying that the gene responsible for *qAL3*/*qAP3* may function at least around stage 5.

GWAS in the sorghum germplasm highlights the widespread distribution of qAL3/qAP3

To clarify whether the extraordinary effectiveness of *qAL3/qAP3* is the only local QTL limited to our RIL population or is globally involved in awn regulation in sorghum, we conducted GWAS for AL using 289 accessions of sorghum germplasm, whose data were collected from cultivation in Tokyo for three years (2015, 2019, and 2020). Our germplasm collection was shown to exhibit a large variability in AL (Fig. 3A), and more than half of the accessions were awnless (Fig. 3B). First, we conducted GWAS using this AL data and found significant associations on chromosome 3 with a peak at Chr03:72641259 (Fig. 3C). This region was found to be consistent with that of the marker interval of *qAL3*/*qAP3*, and the peak of GWAS and the nearest marker of *qAL3*/*qAP3* were only 14 kb apart (Fig. 3D). The effect of the peak genotype was the same as that of *qAL3*/*qAP3*; ALs were zero in lines carrying the homozygous reference-type (BTx623) allele at the peak. In addition, similar results were observed for lines carrying the heterozygous allele at the peak (Fig. 3E). Next, we performed GWAS using the AP phenotype as we did in the QTL analysis, and detected similar associations on chromosome 3 as those detected in the AL (Fig. 3C insert). These results suggest that *qAL3*/*qAP3* in the QTL analysis and the peak of GWAS are likely to represent the same gene. Detected associations of GWAS presented here were also found in GWAS using datasets of the 2019 and 2020 cultivation, and these results indicate high

reproducibility for this association (Supplementary Fig. 1I and 1J). The fact that two reports have detected similar associations for AP on chromosome 3 (Zhang et al. 2015, Girma et al. 2019) by using another germplasm population also guarantees the validity of our results.

To assess the geographic distribution of alleles at the peak SNP detected by GWAS, we color-coded the individual accessions by genotype and plotted the origin of individual accessions on a world map (Fig. 3F). The results showed that the awnless accessions with a Ref-type allele (BTx623 type) were particularly abundant in south Africa and North America, while the awned accessions with an Alt-type allele (NOG type) were more common in Asia, North Africa, and Europe. Whether such geographic distribution reflects adaptation to regional climates or the result of human selection corresponding to the purpose of use is unclear.

The candidate of the Dominant Awn Inhibitor gene in sorghum

We compared the reference genome sequence of BTx623 with those from NOG that were determined by our re-sequencing data (Accession number: DRA008159). Consequently, a region of approximately 5 kb containing a single gene, *Sobic003G421300*, was shown to be missing in NOG (Supplementary Fig. 2A). Indeed, genomic PCR and Sanger sequencing verified this deletion in NOG that extended by 5.46 kb (Supplementary Fig. 2B and 2C, Fig. 4A). The peak SNP of the GWAS was located approximately 5.7 kb upstream from the transcriptional start site of *Sobic003G421300*. According to the Phytozome database (Goodstein et al. 2012), *Sobic003G421300* encodes a protein of unknown function that contains an ALOG (Arabidopsis LSH1 and Oryza G1) domain (Fig. 4B). Among the ALOG family proteins in grass, G1, TAW1, and TH1/BLS1/AFD1 in rice have been well studied. LONG STERILE LEMMA1 (G1) is thought to be involved in transcriptional regulation and is expressed in the sterile lemma primordia to prevent its development in the wild-type rice plant; the sterile lemma is abnormally enlarged like the lemma in the *g1* loss-of-function mutant (Yoshida et al. 2009). TAWAWA1 (TAW1), which is expressed in the inflorescence meristem (IM) and branch meristem, is also considered a transcription factor. TAW1 regulates the inflorescence architecture of rice by promoting IM activity and suppressing phase changes to spikelet meristem identity (Yoshida et al. 2013). For TRIANGULAR HULL1/BEAK LIKE SPIKELET1/ABNORMAL FLOWER AND DWARF1 (TH1/BLS1/ADF1), mutants showed various pleiotropic defects containing abnormal floral organs, and some results indicated that TH1/BLS1/ADF1 may influence cell sizes and cell numbers by negatively regulating the expression of cell proliferation and expansion related genes (Li et al. 2012, Ma et al. 2013, Sato et al. 2014, Ren et al. 2016, Peng et al. 2017).

These studies imply that ALOG family proteins inhibit the differentiation or development of specific organs. Taking these into account, it seemed plausible that *Sobic003G421300* represents *qAL3*/*qAP3*. We named this gene *Dominant Awn Inhibitor* (*DAI*). Given ALOG family proteins acting as transcription factors, we observed the subcellular localization of DAI in order to examine whether DAI also has transcription factor-like properties. We transiently expressed green fluorescent protein (GFP) or DAI-GFP fusion protein driven by the cauliflower mosaic virus 35S promoter into onion epidermal cells by particle bombardment. While control GFP appeared to be localized to the cytoplasm and nucleus, but when magnified, GFP was present outside the nucleus, indicating that it was not actively localized to the nucleus (Fig. 4C upper). In contrast, the DAI-GFP

fusion protein was exclusively localized in the nucleus (Fig. 4C bottom). Most likely, DAI is the transcription factor negatively regulating awn formation.

Confirmation by further genotyping and complementation analysis

Next, we performed additional experiments to verify that the *DAI* corresponded to *qAL3*/*qAP3*. First, the link between the awned phenotype and *DAI* deletion was carefully examined. While the peak SNP correlated well with awn phenotypes in most of the accessions (Fig. 3E), a few accessions showed discrepant phenotypes, including the awned ones with the Ref-type allele, and the awnless ones with the Alt-type allele. Therefore, we selected some accessions and examined whether they have the 5.46-kb fragment containing *DAI* (the *DAI* fragment) and/or whether they have additional mutations in *DAI* by Sanger sequencing. The peak genotype and the presence of the *DAI* fragment matched completely, for the accessions whose awn phenotypes were consistent with the peak genotype (Fig. 4D, 10 accessions from the top). For accessions with the discrepancy between the peak genotype and the awn phenotype, our experimental reassessment clarified that three accessions (PI 656029, PI 152771, and IS 29606) had consistency between the awn phenotype and the *DAI* insertion. In contrast, IS 24453 was remarkably different because it was awned despite having the *DAI* insertion. However, the sequence analysis of *DAI* in IS 24453 revealed that a single nucleotide substitution occurred (82 bp from the translation start point; C to T), resulting in a premature stop codon truncating DAI to only 28 aa (282 aa in wild-type), which was most likely a knockout. Thus, IS 24453 has the *DAI* fragment; however, *DAI* has been mutated and lost its function. These results strongly suggested that *DAI* determines the presence of awn in sorghum.

Subsequently, we performed complementation analysis in sorghum using *Agrobacterium*-mediated transformation through tissue culture derived from immature embryos. We cloned the genomic fragment of BTx623 containing the putative promoter region (approximately 6.2 kb) and *DAI* into a binary vector (*DAI*pro::*DAI*) (Fig. 4A bottom). RIL232 is one of the awned lines in our RIL population and was used as a recipient. The obtained transformant (T_0) was found to have a single insertion based on the segregation ratio of T_1 progeny. Genotyping of T_1 plants were performed to distinguish homozygous (*DAI*/*DAI*), heterozygous (*DAI*/-), or azygous (-/-) individuals by qPCR. Five T_1 plants of each genotype were selected for the observation of spikelets and measurement of AL. We found that homozygous plants completely lost their awn, heterozygous plants were partially awned, but mostly lost their awn, and azygous plants had the same level of awn as that in the parent (Fig.4 E, F). Together, these results led us to conclude that *DAI* is indeed the gene responsible for *qAL3*/*qAP3*.

DAI may have originated from a gene duplication

DAI is present in the 5.64 kb genomic fragment that does not exist in accessions such as NOG. We questioned whether this fragment is evolutionarily derived from the deletion in the NOG ancestor or the insertion of the BTx623 ancestor. Somewhat surprisingly, the BLASTP search revealed the presence of one sorghum protein encoded by Sobic.010G223100 on chromosome 10, which showed 100% identity with DAI in the amino acid sequence. Comparison of the coding sequences (CDSs) between the two genes revealed only one

synonymous substitution at nucleotide 603-bp downstream from the translation start point (Supplementary Fig. 3). Moreover, *DAI* and *Sobic.010G223100* showed significant homology not only between CDSs but also in the regions encompassing 5.15 kb (Supplementary Fig. 3). Given that *Sobic.010G223100* is also present in the NOG genome (not shown) and that the awned predates the awnless phenotype during varietal differentiation, it is likely that the 5.46-kb fragment on chromosome 3 resulted from a duplication of *Sobic.010G223100* in the BTx623 ancestor, rather than a deletion in the NOG ancestor. Therefore, we named this *DAI* homolog on chromosome 10 DAI^{ori}. To further verify this hypothesis, synteny analysis was performed to compare the corresponding chromosomal regions among three other grass species (rice, brachypodium, and maize). There were syntenic blocks, which corresponded to the region around *DAI* on chromosome 3 of sorghum, rice chromosome 1, brachypodium chromosome 2, and maize chromosomes 3 and 8. However, there were no *DAI* homologs in their syntenic blocks (Fig. 5A and B upper). In contrast, DAI^{ori} orthologs were found in syntenic blocks of all these grass species (in rice chromosome 6, brachypodium chromosome 1, and maize chromosomes 6 and 9) (Fig. 5A and B bottom). These observations support the hypothesis that DAI^{ori} is the origin of *DAI* and might have jumped into chromosome 3 from chromosome 10 in the BTx623 ancestor.

Next, we performed BLASTP analysis on four grass species, including sorghum, using DAI/DAI^{ori} protein as a query, and constructed a phylogenetic tree using the selected related proteins. As a result, all proteins were grouped into clades represented by each rice G1 family protein (G1, G1L1–L9), although gene duplication was particularly prominent in maize (Fig. 5C). These ALOG family proteins were well-conserved, at least among these four species. DAI^{ori} was found to belong to the G1L2 clade of unknown function, which was different from TAW1 (G1L5 clade), TH1/BLS1/ADF1 (G1L6 clade), and G1 clade, whose functions have been reported. Functional analysis of genes in the G1L2 clade, including DAI^{ori}, is a topic for future research.

SbDAI can also function as an awn inhibitor in rice

As a result of sequence comparison between the surrounding regions of *DAI* and DAI^{ori}, we predicted how *DAI* was generated by gene duplication through genome rearrangement (Supplementary Fig. 3, Fig. 5D). First, a fragment of approximately 5.15 kb encompassing the intron and CDS of DAI^{ori} on chromosome 10 was duplicated and inserted into chromosome 3 (at approximately 72.6 Mb). A total of 80 bp of unidentified sequence seemed to be added to both ends during the insertion; the origin of this sequence was currently unclear, because BLAST searches against all species did not find any hits other than its own. Subsequently, Tourist1b, a transposable element, was inserted into the intron region, and the gene might have undergone further rearrangement in the UTRs and introns, resulting in the birth of the *DAI* gene in its present form. There is no obvious evidence such as transposable elements at either end of the 5.46 kb inserted fragment.

The acquisition of *DAI* as a result of gene duplication, along with DAI as a conserved ALOG family protein among grass species, raised the possibility that the whole *DAI* allele, including its promoter, may act as an awn inhibitor in species other than sorghum. To examine this possibility, we generated transformants using the awned rice cultivar Kasalath as a recipient, heterologously expressed *DAI*, and observed their spikelets. Using the pHGWFS7 vector as a backbone, the construct *SbDAI* was constructed that harbored the genomic region from BTx623, as shown in Figure 6A. As a negative control, the corresponding region from NOG was

also cloned in the construct gNOG (Fig. 6A). Each binary vector carrying *SbDAI* or gNOG was transformed into Kasalath. Consequently, the individuals with gNOG exhibited an awned phenotype similar to Kasalath, while the individuals with *SbDAI* showed a complete awnless phenotype (Fig. 6 B–D). Three independent transformants (T_0) obtained for each construct were found to be single-copy transformants, and we determined whether the individual had transgene(s) (+ in Fig. 6E) or not (azygous: - in Fig. 6E) by PCR using DNA isolated from each of the segregated T_1 populations. A T_1 plant for each genotype was selected for the measurement of AL. The results of AL measurement showed that the awns of individuals with the *SbDAI* transgene were significantly shorter (almost completely zero) than those of their azygous (Fig. 6E). This indicated that *SbDAI* (native promoter + *DAI*) could also inhibit awn elongation in heterologous rice plants, supporting the gain-of-function of *DAI* function by gene duplication.

Discussion

In this study, three QTLs were identified on chromosomes 3 (*qAL3*/*qAP3*, Chr03:71923025–72847396), 7 (*qAL7*, Chr07:58230302–Chr07:60603839), and 9 (*qAL9*, Chr09:56467317–57917278). Comparing these QTLs with those at the QTL Atlas (Mace et al. 2019), *qAL3*/*qAP3* was likely to map to the awn length or presence-related QTLs reported by previous works (Boivin et al. 1999, Tao et al. 2000, Hart et al. 2001, Mace and Jordan 2010, Girma et al. 2019, Zhang et al. 2019). In this study, we corroborated this locus on chromosome 3 and identified *DAI* as the dominant factor that inhibits awn formation.

In contrast to *qAL3/qAP3*, no QTLs related to AL have been previously reported around *qAL7* and *qAL9*. Investigation of possible genes corresponding to these QTLs identified two genes, *Dwarf1* (*Dw1*) and *Dwarf3* (*Dw3*), which are the major plant height-responsible genes in sorghum. *Dw1* (Sobic.009G230800, Chr09:57093313–57095643), which encodes a positive regulator gene of brassinosteroid signaling, regulates the length of internodes by controlling cell proliferation (Hilley et al. 2016, Yamaguchi et al. 2016, Hirano et al. 2017). *Dw3* (Sobic.007G163800, Chr07:59821905–59829910) encodes an ABCB1 auxin efflux transporter (Multani et al. 2003). Previously, we detected *qPH9* (Chr09:56.57–57.68 Mb) and *qPH7* (Chr07:59.41–60.00 Mb) as dominant plant height QTLs using the same RILs. We demonstrated that *Dw1* and *Dw3* are the genes responsible for *qPH9* and *qPH7* by sequencing analysis, respectively, highlighting that NOG has functional alleles for both genes (Kajiya-Kanegae et al. 2020). Given that *qAL7* (Chr07:58.23–60.60 Mb) and *qAL9* (Chr09:56.47–57.92 Mb) in this study contained *qPH7* and *qPH9* with high accuracy, *qAL7* and *qAL9* were thought to detect the same locus as *qPH7* and *qPH9*, respectively. These two loci were also detected as significant QTLs for the length of spikelet-related organs (e.g., sessile spikelet, pedicel, anther, and style) using the same RILs (Takanashi et al. 2021). Based on the functional aspect of *Dw1* and *Dw3* (e.g., controlling organ size) and the direction of the allelic effects, in which the length of awn and spikelet-related organs and plant height were increased in the NOG allele, *qAL7* and *qAL9* also likely corresponded to *Dw3* and *Dw1*, respectively.

A line of evidence including NILs, detailed genotyping, and complementation analysis using sorghum and rice, led us to conclude that *DAI* is the gene responsible for *qAL3*/*qAP3*, which determines the presence of

awn in sorghum dominantly. The fact that most awnless accessions are likely to have achieved the awnless phenotype using *DAI* (Fig. 3 and 4) strongly suggests that *DAI* is an important gene for the development of modern, awnless sorghum varieties. *DAI* is useful and easy to manipulate genetically as a dominant factor, for example, in generating awnless F_1 hybrids. It is therefore plausible that *DAI* has been widely selected during the course of varietal diversification. *DAI* belongs to the ALOG family and is different from any other awn-regulating genes that have been reported in other grasses. It is noteworthy that wheat *B1*, which encodes a C2H2 zinc finger protein, functions similarly in inhibiting awn formation (DeWitt et al. 2020, Huang et al. 2020). The identification of *DAI* in this study along with *B1* reinforces the occurrence of such awn inhibitors, likely the transcription factors repressing cell elongation and/or organ formation in cereal crop species. However, it is noteworthy that they encode different transcription factors. Whether *DAI* is a gene associated with sorghum domestication is still unknown and needs to be verified by comparing genomic sequences of many accessions in the future. Future studies are required to understand the mechanism by which the putative transcription factor DAI inhibits awns by affecting downstream target genes.

Our finding that *DAI* was born by gene duplication provides an interesting example of gain-of-function associated with panicle morphologies that are supplemental to plant growth. In this scenario, landing a duplicated gene into a region conferring proper expression by the novel promoter may be critically important to gain new functions. Moreover, we demonstrated in this study that the *DAI* genomic region containing its promoter can act as an awn inhibitor in rice, in which inhibitor genes have not been identified so far. Together, these results suggest that the rise of dominant awn inhibitors depends on such rare events during the diversification of cereals and is thus species-specific. In the case of *DAI*, the insertion site was within a non-coding DNA region that is 15.48 kb and 10.26 kb away from the neighboring gene both upstream (Sobic.003G421201) and downstream (Sobic.003G421400), respectively (Supplementary Fig. 2D). The upstream sequence of the insertion site in chromosome 3 was able to function as a promoter for expression in the awn or its primordium, which seems to be the reason why the DAI^{ori} paralogue was able to function as a dominant awn inhibitor, although it is still unclear which sequences were essential promoter sequences for the function of DAI. Because there was a putative long interspersed nucleotide element (LINE, Fig. 4A) upstream of the insertion site, whose epigenetic modifications have been known to lead to changes in the expression of nearby genes (Ong-Abdullah et al. 2015), it is possible that this LINE was involved in the establishment of the *DAI* function. Further elucidation of the mechanism of transcriptional regulation of *DAI* is necessary in future studies

Phylogenetic analysis showed that DAI^{ori} belongs to the G1L2 clade. Although the function of the G1L2 group is unknown, the G1L2 subfamily including DAI^{ori} may have a similar function to suppress the differentiation or development of specific tissues or organs common to these grasses, because the putative common ancestral species must have a G1L2-like gene before the differentiation of these species. Since the ancestral grasses most likely had long awns, it is unlikely that ancestors of the G1L2 group proteins had the ability to suppress awn development. The reason for this hypothesis is that the locus around G1L2 has not been detected in a vast number of researches of rice awn that have been performed. Currently, we expect that *DAI* is a novel gene that has acquired a new function to inhibit awn by inserting a copy of DAI^{ori}, which has originally

been unrelated to the regulation of awn, into chromosome 3 through gene duplication, resulting in the acquisition of a new promoter leading expression around the lemma or awn.

Conclusion

In this study, we identified the dominant awn inhibitor gene in sorghum, which has not been identified for a century since its first mention (Vinall and Cron 1921). We showed that *DAI* encodes a protein of unknown function with an ALOG domain, and it can almost eliminate the sorghum awn even in the heterologous state. Experiments using sorghum germplasm suggested that *DAI* has been widely used in the process of modification or breeding of sorghum, perhaps because of its effectiveness and convenience.

DAI is expected to be a very powerful and useful tool because an awnless phenotype can be achieved by only a single gene, and even in the heterozygous condition. For example, in the case of F_1 breeding by utilizing hybrid vigor, a high grain yield and awnless phenotype can be achieved by simply introducing *DAI* into either parent. As shown using rice, *DAI* is likely to be effective on heterologous grass species, which makes it promising for future use. Further analysis of the *DAI* gene, which is different from any other gene identified in other grass species, will shed light in a better understanding of the mechanisms underlying the awn regulation and for the birth of a novel gene via gene duplication.

Note to be added

In the course of preparing this manuscript, Zhou et al. (2021) reported a gene named *awn1*, identical to *DAI* in this study, as a candidate for an awn repressor in sorghum. Although their study did not provide direct evidence that *awn1*/*DAI* is a gene responsible for the dominant awn inhibitor, they provided several important results, including RNA sequencing and DNA affinity purification sequencing to identify the genes downstream of *awn1*. We believe that combining their report with ours will enrich our knowledge of awn regulation in sorghum.

Materials and methods

Plant materials

The recombinant inbred line (RIL) population, derived from a parental cross between BTx623 and the Japanese landrace NOG was established as described in our previous study (Kajiya-Kanegae et al. 2020). In total, 245 RILs were generated by recurrent selfing of progeny derived from a cross with BTx623 and NOG at the Institute of Plant Science and Resources (IPSR) at Okayama University, as previously reported (Kajiya-Kanegae et al. 2020). A total of 289 accessions of sorghum germplasm were obtained from public bioresources (ICRISAT sorghum collection, NARO Genebank, and USDA). Supplementary Table 1 lists these accessions. Seeds of the rice cultivar, Kasalath, were kindly provided by Dr. Yasuo Nagato of the University of Tokyo. Near isogenic lines (NILs) were obtained starting from RIL205 (awnless), one of the heterozygous lines at the genotype of the marker associated with *qAL3*/*qAP3*, in the RILs at the F_6 generation (Fig. 2A). We established NILs (F_9 generation) through PCR genotyping using insertion-deletion polymorphisms for recurrent selfed progeny of RIL205 at IPSR (primers used for genotyping are listed in Supplementary Table 2).

Genotype data

Genotype data of the RILs were obtained using the restriction site-associated DNA sequencing (RAD-seq) method. Detailed methods and information for DNA extraction and library construction for the RAD-seq have been described in our previous report (Kobayashi et al. 2017). The RAD-seq reads of the 245 RILs from the F_6 generation were mapped onto the reference genome Sorghum bicolor v2.1 (Paterson et al. 2009) using the bwa mem (Li and Durbin 2010) algorithm with the default options, and then the marker positions were corrected to the information of the reference genome Sorghum bicolor v3.1.1 (McCormick et al. 2018). The SNPs were called with GATK (McKenna et al. 2010) and filtered with SAMtools (Li et al. 2009) under the following conditions: min-depth 5, max-depth 30, max-missing 0.8, minimum allele frequency 0.1, remove-indels, min-alleles 2, max-alleles 2, and minQ 20. For segregation distortion analysis, we used a chi-squared test to calculate the deviation from the expected ratio (1:1) for each marker. All genotypes (BTx623/BTx623, NOG/NOG, BTx623/NOG) were used to elucidate allelic effects of QTLs in the F_6 generation. To construct the genetic map for the RILs, only homozygous SNPs were selected, then SNPs with missing rates greater than 20% and duplicated markers were removed. In total, 1,792 high-quality SNPs were selected to construct the linkage map for QTL analysis. The map covered a total length of 682.4 cM and had an average marker density of 0.4 cM.

Genotype data of sorghum germplasm were also obtained using RAD-seq. Detailed methods and information for genotyping have been described in our previous report (Yamazaki et al. 2020). SNPs with minor allele frequencies (1%) or missing rates greater than 80% were filtered out for the GWAS study. Imputation was conducted using Beagle 5.0, with default parameter settings (Browning et al. 2018). In total, we obtained data for 50,678 SNP markers for 298 accessions.

Cultivation

Field trials were performed in an open-air greenhouse at the experimental farm of IPSR (latitude: 34°35'31' 'N, longitude: 133°46'7 '' E), Kurashiki City, Okayama, Japan from June to September 2014 for the F_6 generation of the RILs. Plants were germinated in small trays filled with vermiculite for approximately 10–14 days, and then transplanted to either 30-cm diameter pots filled with soil from the IPSR field (n = 1 per line).

Field trials were performed in a green field at the University of Tokyo (latitude: 35°42'58.5' 'N, longitude: 139°45'44.5 '' E), Yayoi, Bunkyo-Ku, Tokyo, Japan, from May to October in 2015, 2019, and 2020. Plants were germinated in 200-cell trays filled with synthetic culture soil (BONSOL 1; Sumitomo Chemical Co., Ltd., Japan) for two weeks, and then transplanted into the field (n = 2 per line). The space between ridges was 60 cm, with a 15 cm distance between individuals in the same ridge. The field was fertilized using N:P:K = 10:10:10 (kg ha^{-1}) fertilizer before the trials in each year. Cultivation and phenotyping of awn-related traits for RILs were conducted in 2015 using the F_7 generation, and those for germplasm were conducted in 2015, 2019, and 2020.

Transgenic sorghum and rice plants were grown and maintained in a biohazard greenhouse at 30 °C (day) and 25 °C (night) under natural light conditions.

Phenotyping for QTL analysis and GWAS

To standardize developmental stages of spikelets among the RILs and germplasm, we sampled spikelets at the just-before-flowering stage in each line without the F_6 generation of the RILs cultivated at Okayama, which were sampled during the grain-filling stage. In the inflorescence that began flowering at the top, we defined spikelets at the just-before-flowering stage (spikelets that would flower within a day) as those attached to the primary branch just below the flowered primary branch. For each line, we picked awns from five spikelets sampled from the top, middle, and bottom of the panicle, mounted them on a paper board, and photographed with a digital camera (OM-D E-M5; Olympus, Japan). Awn length (AL) was measured using ImageJ 1.49v (Schneider et al. 2012). We also scored awn presence values on a two-step scale, with awnless lines as 0 and awned lines as 1. For all awn-related traits, the average value of five independent awns was used as the phenotypic value of an individual plant.

QTL analysis

We performed QTL analysis for awn-related traits in the RILs (AL; AP, awn presence; and AL in awned lines only). Genotype probabilities were calculated using the calc. genoprobility function with a step size of 1 cM and an assumed genotyping error probability of 0.05, using the Kosambi map function (Kosambi 1943) as implemented in the R (R Core Team 2021)/qtl package (Broman et al. 2003). QTL analysis was performed using the composite interval mapping (CIM) function of the R/qtl package with the Haley–Knott regression method (Haley and Knott 1992). Linkage analysis was performed using the R/qtl package in R version 4.1.0. The LOD significance threshold for detecting QTLs was calculated by performing 1,000 iterations using the R/qtl permutation test. Confidence intervals for QTLs were estimated based on the 2.0-LOD support interval, and the nearest flanking markers located outside the boundary of each confidence interval were defined as both

ends of the marker intervals. The additive effect and the percentage of phenotypic variance explained (PVE) by each QTL were obtained using the fitqtl function of the R/qtl package. Box plots were created using the R/ggplot2 package (Wickham 2016).

GWAS

We performed GWAS for awn-related traits in germplasm (AL and AP). GWAS was performed using a linear mixed model implemented by the "association.test" function in the gaston package ver. 1.5.7 in R (Perdry et al. 2020). A genetic relationship matrix specifying a random additive effect was computed using the 'A.mat' function of the rrBLUP package (Endelman 2011). The P values of the marker-trait associations were calculated using the Wald test. The genome-wide significance threshold was obtained based on the Benjamini–Hochberg false discovery rate [FDR; Benjamini and Hochberg (1995)] at a 0.1% level. The LD heatmaps surrounding peaks in the GWAS were constructed using the "LD" function of the gaston package. To determine the geographic distribution of alleles at the peak SNP in GWAS, the world map layout was generated using the ggplot2 package in R. Passport data of accessions were obtained from the U.S. National Plant Germplasm System (https://npgsweb.ars-grin.gov/gringlobal/search), the NARO Genebank (https://www.gene.affrc.go.jp/databases-core_collections_ws.php), and the ICRISAT sorghum collection (http://www.icrisat.org/what-we-do/crops/sorghum/Project1/Sorghum_minicore.htm).

Genotyping of the 5.46 kb fragment containing DAI for BTx623, NOG, and selected germplasm

Healthy young leaves from each line were collected, and genomic DNA was extracted using a Maxwell RSC Plant DNA Kit (Promega, USA). The presence or absence of the 5.46 kb fragment containing *DAI* was confirmed by PCR (primers 1 and 2), and its sequence information was determined by direct Sanger sequencing using the same primers.

Plasmid construction

To express GFP and DAI-GFP transiently, we made the corresponding constructs using the Gateway cloning technology (Invitrogen /Thermo Fisher Scientific, USA). The CDS of *DAI* without a stop codon was amplified by PCR using BTx623 DNA as a template (primers 3 and 4), and the amplified DNA fragment and an annealed linker oligo (for GFP: primers 5 and 6) was introduced into the pENTR/D-TOPO vector (Invitrogen; sequences were confirmed by Sanger sequencing using primers 7 and 8) and then subcloned into a pH7FWG2 vector (Karimi et al. 2002) using LR clonase (Invitrogen) to obtain 35Spro::DAI-GFP and 35Spro::GFP constructs.

For sorghum and Kasalath transgenic lines, the constructs were made using the In-Fusion HD cloning technology (Clontech/Takara Bio, Japan). For *DAI*pro::*DAI*, the pBUH3 vector (Nigorikawa et al. 2012) was linearized by PCR (primers 9 and 10). The DNA fragment containing *DAI* native promoter and the CDS of *DAI* with a stop codon were amplified by PCR using *the SbDAI* plasmid described below as a template (primers 11 and 12), and the PCR-amplified DNA fragment was cloned into linearized pBUH3 via In-Fusion HD cloning. The sequence of the *DAI*pro::*DAI* plasmid was confirmed by Sanger sequencing using primers 13 to 18. For

SbDAI, the DNA fragment containing *DAI* native promoter and the CDS of *DAI* with a stop codon were amplified by nested PCR using BTx623 DNA as a template (primers 25–28). For gNOG, the DNA fragment was amplified by nested PCR using NOG DNA as a template (primers 25, 2, 27, and 29). Amplified *DAI*pro::*DAI* and gNOG fragments were cloned into the *Nco*I/*Sac*I-digested pHGWFS7 vector (Karimi et al. 2002) via In-Fusion HD to generate *SbDAI* and gNOG plasmids, respectively. The sequence of these plasmids was confirmed by Sanger sequencing using primers 16 and 30 to 33.

Particle bombardment

GFP and DAI-GFP plasmids were transiently introduced into onion epidermal cells using a helium-driven particle accelerator (PDS/1000; Bio-Rad Laboratories, USA). Particle bombardment experiments were performed according to the conditions described by Arimura et al. (2004). Before microscopic observation, onion epidermal cells were treated with distilled water containing 2 mg L^{-1} 4',6-diamidino-2-phenylindole dihydrochloride (DAPI) to stain nuclei.

Microscopic observations

Enlarged images of secondary branches and spikelets of sorghum and Kasalath were captured under a stereomicroscope (M125; Leica, Germany) with a CCD camera (MC170 HD; Leica). Onion epidermal cells expressing GFP and DAI-GFP under control of the 35S promoter were observed using a Leica STELLARIS5 confocal microscope (Leica) with a HC PL APO CS2 20× DRY objective lens (N.A. = 0.75), HC PL APO CS2 63× OIL objective lens (N.A. = 1.40), and HyD detectors. The Leica STELLARIS5 microscope was controlled by LAS X software (Leica), DAPI and GFP were excited with a 405 nm DMDO laser and a 495 nm tuned white light laser, respectively.

Agrobacterium-mediated transformation

Agrobacterium-mediated sorghum transformation of RIL232 (one of the lines in our RILs, awned) was performed according to Wu et al. (2014) with several modifications. Immature embryos isolated from grains 15 days after pollination were pre-cultured in DBC medium at 25 °C in the dark. Pre-cultured immature embryos were collected in infection medium, pre-treated at 43 °C for 15 min, and centrifuged (20,000 ×g for 10 min). Pre-treated immature embryos were inoculated with Agrobacterium tumefaciens strain EHA105 harboring the *DAI*pro::*DAI* vector and co-cultured at 25 °C for 5 days in the dark. After 5 days of resting culture, calli were selected on a medium containing 15-20 mg L^{-1} hygromycin, and resistant calli were transferred to regeneration medium containing 10 mg L^{-1} hygromycin. Regenerated shoots were cultured on rooting medium until the roots were produced. We obtained several plantlets, but only one was confirmed as a transformant by PCR. Segregation analyses of T_1 plants derived from the self-pollinated T_0 plant by PCR using primers 19 and 20 revealed that the T_0 plant was heterozygous for a single copy of the transgene (segregation ratio was approximately 3:1). We confirmed the number of transgenes in segregated T_1 plants by qPCR using primers 21 to 24 and selected five individuals each as homozygous, heterozygous, and azygous for the measurement of AL. For each line, we picked up awns from 30 spikelets sampled from the middle portion of the panicle at the

just-before-flowering stage and measured the AL as described above.

The constructed gNOG and *SbDAI* vectors were transformed into calli induced from the scutellum of wild type rice seeds (cv. Kasalath, awned) by *Agrobacterium tumefaciens* (strain EHA105), according to Toki et al. (2006). Transgenic plants were selected on medium containing 50 mg L^{-1} hygromycin. Hygromycin-resistant plants (T_0) were transplanted into soil and grown at 30 °C (day) and 25 °C (night). Genotyping of T_0 plants was performed by PCR using primers 19, 34, and 35, and we obtained three independent transgenic lines for each construct. T_1 plants derived from the self-pollinated T_0 plants were genotyped by PCR, and individuals with/without transgene(s) were selected for the measurement of AL. For each line, we sampled panicles before flowering and measured the AL of apical spikelets on each primary and secondary branch (approximately 100 spikelets).

Synteny analysis

Synteny analysis was performed using the online SynMap2 software (Haug-Baltzell et al. 2017). Datasets were chosen for dot plots as follows: *Sorghum bicolor* (id331), *Oryza sativa* (id3), *Brachypodium distachyon* (id39836), and *Zea mays* (id38215). Only the results of chromosomes corresponding to *DAI* (Sb_Chr03) and DAI^{ori} (Sb_Chr10) were extracted from the output files of SynMap2 and dot plots were drawn using ggplot2/R. Enlarged images around *DAI* or DAI^{ori} in Figure 5B were based on the reference genomes of *Sorghum bicolor* (v3.1.1), *Oryza sativa* (v7.0), *Brachypodium distachyon* (v3.1), and *Zea mays* (RefGen_V4) in Phytozome 13 (Goodstein et al. 2012).

Phylogenetic analysis

The amino acid sequence of Sobic.003G421300/DAI was used as the query sequence for a BLASTP search against the protein databases of *Sorghum bicolor* (v3.1.1), *Oryza sativa* (v7.0), *Brachypodium distachyon* (v3.1), and *Zea mays* (RefGen_V4) using Phytozome 13. Fifty matches with an expected value $< e^{-40}$ were selected, and a phylogenetic tree was constructed using the ClustalW alignment and maximum likelihood methods based on the JTT matrix-based model (Jones et al. 1992) using default settings in the MEGA 7 package (Kumar et al. 2016). The following protein sequences from the Phytozome database were used for the phylogenetic analysis. *S. bicolor*: Sobic.001G219400, Sobic.002G029700, Sobic.003G342600, Sobic.003G421300 (DAI), Sobic.004G052800, Sobic.004G219600, Sobic.004G338400, Sobic.005G094266, Sobic.006G147500, Sobic.009G102600, Sobic.009G170100, and Sobic.010G225100 (DAI^{ori}); *O. sativa*: LOC_Os01g61310 (G1L7), LOC_Os02g07030 (G1L1), LOC_Os02g41460 (G1L3), LOC_Os02g56610 (G1L6/TH1/BLS1/ADF1), LOC_Os04g43580 (G1L4), LOC_Os05g28040 (G1L9), LOC_Os05g39500 (G1L8), LOC_Os06g46030 (G1L2), LOC_Os07g04670 (G1), and LOC_Os10g33780 (G1L5/TAW1), *B. distachyon*: Bradi1g32445, Bradi1g58070, Bradi2g22561, Bradi2g29110, Bradi2g53880, Bradi3g04960, Bradi3g28800, Bradi3g49140, Bradi3g54800, and Bradi5g15700; *Z. mays*: Zm00001d003057, Zm00001d010507, Zm00001d010596, Zm00001d012473, Zm00001d014032, Zm00001d015342, Zm00001d017263, Zm00001d018433, Zm00001d018775, Zm00001d025863, Zm00001d032696, Zm00001d036617, Zm00001d038016, Zm00001d038442, Zm00001d043011, Zm00001d046998, Zm00001d051143, and

Zm00001d051966.

Data Availability

The sequence data reported in this paper have been submitted to the DDBJ Sequence Read Archive with the accession number DRA008159.

Author Contributions

W.S. conceived the work dissecting sorghum awn development along with H.T.; material preparation, field experiments, and data analysis in QTL and GWAS was performed by H.T., H.K.-K., W.S., M.I. with supervision by M.K., K.Y., H.I., and N.T.; NIL was established by W.S and analyzed by H.T.; Sorghum transformation was performed by A.N., J.Y., and H.T. with supervision by N.T.; Rice transformation was performed by H.T.; final data were prepared for publication by H.T., H. K.-K., and W.S., and the manuscript was written by H.T. and W.S., on behalf of all authors.

Funding

This work is partly supported by the Core Research for Evolutional Science and Technology (CREST) from the Japan Science and Technology Agency (to N.T. and W.S.), KAKENHI grants (17H01457 to N.T., 18K19343 and 21H02597 to W.S., 18K05570 to H.T.) from the Japanese Society for the Promotion of Science (JSPS). We also thank the Oohara Foundation for financial support of our research group.

Disclosures

The authors have no conflicts of interest to declare.

Acknowledgments

We thank Tsuneaki Takami, Zihuan Jing, Fiona Wacera, Norikazu Ohnishi, Everlyne Omollo, and Rie Hijiya (IPSR, Okayama University) for their help with the cultivation at the IPSR, Kurashiki, Okayama. We are also grateful to the members of the Laboratory of Plant Molecular Genetics, and the Laboratory of Biometry and Bioinformatics at the University of Tokyo for their help with the cultivation in Tokyo.

References

Amarasinghe, Y.P.J., Kuwata, R., Nishimura, A., Phan, P.D.T., Ishikawa, R. and Ishii, T. (2020) Evaluation of domestication loci associated with awnlessness in cultivated rice, Oryza sativa. *Rice (N Y)* 13: 26.

Arimura, S., Yamamoto, J., Aida, G.P., Nakazono, M. and Tsutsumi, N. (2004) Frequent fusion and fission of plant mitochondria with unequal nucleoid distribution. *Proc. Natl. Acad. Sci. U. S. A.* 101: 7805-7808.

Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate - A practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* 57: 289-300.

Bessho-Uehara, K., Wang, D.R., Furuta, T., Minami, A., Nagai, K., Gamuyao, R., et al. (2016) Loss of function at RAE2, a previously unidentified EPFL, is required for awnlessness in cultivated Asian rice. *Proc. Natl. Acad. Sci. U. S. A.* 113: 8969-8974.

Bessho-Uehara, K., Yamagata, Y., Takashi, T., Makino, T., Yasui, H., Yoshimura, A., et al. (2021) Exploring the loci responsible for awn development in rice through comparative analysis of all AA genome species. *Plants (Basel)* 10.

Boivin, K., Deu, M., Rami, J.F., Trouche, G. and Hamon, P. (1999) Towards a saturated sorghum map using RFLP and AFLP markers. *Theor. Appl. Genet.* 98: 320-328.

Broman, K.W., Wu, H., Sen, S. and Churchill, G.A. (2003) R/qtl: QTL mapping in experimental crosses. *Bioinformatics* 19: 889-890.

Browning, B.L., Zhou, Y. and Browning, S.R. (2018) A one-penny imputed genome from next-generation reference panels. *Am. J. Hum. Genet.* 103: 338-348.

Cavanagh, A.M., Morgan, J.W. and Godfree, R.C. (2020) Awn morphology influences dispersal, microsite selection and burial of Australian native grass diaspores. *Front. Ecol. Evol.* 8.

DeWitt, N., Guedira, M., Lauer, E., Sarinelli, M., Tyagi, P., Fu, D.L., et al. (2020) Sequence-based mapping identifies a candidate transcription repressor underlying awn suppression at the B1 locus in wheat. *New Phytol.* 225: 326-339.

Dockter, C., Gruszka, D., Braumann, I., Druka, A., Druka, I., Franckowiak, J., et al. (2014) Induced variations in brassinosteroid genes define barley height and sturdiness, and expand the green revolution genetic toolkit. *Plant Physiol.* 166: 1912-1927.

Elbaum, R., Zaltzman, L., Burgert, I. and Fratzl, P. (2007) The role of wheat awns in the seed dispersal unit. *Science* 316: 884-886.

Endelman, J.B. (2011) Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome-Us* 4: 250-255.

Furuta, T., Komeda, N., Asano, K., Uehara, K., Gamuyao, R., Angeles-Shim, R.B., et al. (2015) Convergent loss of awn in two cultivated rice species Oryza sativa and Oryza glaberrima is caused by mutations in different loci. *G3-Genes Genom Genet. (Bethesda)* 5: 2267-2274.

Girma, G., Nida, H., Seyoum, A., Mekonen, M., Nega, A., Lule, D., et al. (2019) A large-scale genome-wide association analyses of Ethiopian sorghum landrace collection reveal loci associated with important traits. *Front. Plant Sci.* 10: 691.

Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., et al. (2012) Phytozome: A

comparative platform for green plant genomics. *Nucleic Acids Res.* 40: D1178-D1186.
Grundbacher, F.J. (1963) The physiological function of the cereal awn. *Bot. Rev.* 29: 366-381.
Gu, B.G., Zhou, T.Y., Luo, J.H., Liu, H., Wang, Y.C., Shangguan, Y.Y., et al. (2015) An-2 encodes a cytokinin synthesis enzyme that regulates awn length and grain production in rice. *Mol. Plant* 8: 1635-1650.
Haley, C.S. and Knott, S.A. (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* 69: 315-324.
Hart, G.E., Schertz, K.F., Peng, Y. and Syed, N.H. (2001) Genetic mapping of Sorghum bicolor (L.) Moench QTLs that control variation in tillering and other morphological characters. *Theor. Appl. Genet.* 103: 1232-1242.
Haug-Baltzell, A., Stephens, S.A., Davey, S., Scheidegger, C.E. and Lyons, E. (2017) SynMap2 and SynMap3D: Web-based whole-genome synteny browsers. *Bioinformatics* 33: 2197-2198.
Hilley, J., Truong, S., Olson, S., Morishige, D. and Mullet, J. (2016) Identification of Dw1, a regulator of sorghum stem internode length. *PLOS ONE* 11: e0151271.
Hirano, K., Kawamura, M., Araki-Nakamura, S., Fujimoto, H., Ohmae-Shinohara, K., Yamaguchi, M., et al. (2017) Sorghum DW1 positively regulates brassinosteroid signaling by inhibiting the nuclear localization of BRASSINOSTEROID INSENSITIVE 2. *Sci. Rep.* 7: 126.
Hua, L., Wang, D.R., Tan, L.B., Fu, Y.C., Liu, F.X., Xiao, L.T., et al. (2015) LABA1, a domestication gene associated with long, barbed awns in wild rice. *Plant Cell* 27: 1875-1888.
Huang, D.Q., Zheng, Q., Melchkart, T., Bekkaoui, Y., Konkin, D.J.F., Kagale, S., et al. (2020) Dominant inhibition of awn development by a putative zinc-finger transcriptional repressor expressed at the B1 locus in wheat. *New Phytol.* 225: 340-355.
Jin, J., Hua, L., Zhu, Z.F., Tan, L.B., Zhao, X.H., Zhang, W.F., et al. (2016) GAD1 encodes a secreted peptide that regulates grain number, grain length, and awn development in rice domestication. *Plant Cell* 28: 2453-2463.
Jing, Z., Wacera, W.F., Takami, T., Takanashi, H., Fukada, F., Kawano, Y., et al. (2021) NB-LRR-encoding genes conferring susceptibility to organophosphate pesticides in sorghum. *Sci. Rep.* in press.
Jones, D.T., Taylor, W.R. and Thornton, J.M. (1992) The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* 8: 275-282.
Kajiya-Kanegae, H., Takanashi, H., Fujimoto, M., Ishimori, M., Ohnishi, N., Wacera, W.F., et al. (2020) RAD-seq-based high-density linkage map construction and QTL mapping of biomass-related traits in sorghum using the Japanese landrace Takakibi NOG. *Plant Cell Physiol.* 61: 1262-1272.
Karimi, M., Inze, D. and Depicker, A. (2002) GATEWAY vectors for Agrobacterium-mediated plant transformation. *Trends Plant Sci.* 7: 193-195.
Kato, K., Miura, H., Akiyama, M., Kuroshima, M. and Sawada, S. (1998) RFLP mapping of the three major genes, Vrn1, Q and B1, on the long arm of chromosome 5A of wheat. *Euphytica* 101: 91-95.
Kjack, J.L. and Witters, R.E. (1974) Physiological-activity of awns in isolines of atlas barley. *Crop Sci.* 14: 243-248.
Kobayashi, M., Ohyanagi, H., Takanashi, H., Asano, S., Kudo, T., Kajiya-Kanegae, H., et al. (2017) Heap: A highly sensitive and accurate SNP detection tool for low-coverage high-throughput sequencing data. *DNA Res.*

24: 397-405.
Kosambi, D.D. (1943) The estimation of map distances from recombination values. *Ann Eugenic* 12: 172-175.
Kumar, S., Stecher, G. and Tamura, K. (2016) MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33: 1870-1874.
Li, H. and Durbin, R. (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26: 589-595.
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25: 2078-2079.
Li, X.J., Sun, L.J., Tan, L.B., Liu, F.X., Zhu, Z.F., Fu, Y.C., et al. (2012) TH1, a DUF640 domain-like gene controls lemma and palea development in rice. *Plant Mol. Biol.* 78: 351-359.
Li, X.J., Wang, H.G., Li, H.B., Zhang, L.Y., Teng, N.J., Lin, Q.Q., et al. (2006) Awns play a dominant role in carbohydrate production during the grain-filling stages in wheat (Triticum aestivum). *Physiol. Plant.* 127: 701-709.
Luo, J.H., Liu, H., Zhou, T.Y., Gu, B.G., Huang, X.H., Shangguan, Y.Y., et al. (2013) An-1 encodes a basic Helix–Loop–Helix protein that regulates awn development, grain size, and grain number in rice. *Plant Cell* 25: 3360-3376.
Ma, X.D., Cheng, Z.J., Wu, F.Q., Jin, M.N., Zhang, L.G., Zhou, F., et al. (2013) BEAK LIKE SPIKELET1 is required for lateral development of lemma and palea in rice. *Plant Mol. Biol. Rep.* 31: 98-108.
Mace, E., Innes, D., Hunt, C., Wang, X., Tao, Y., Baxter, J., et al. (2019) The Sorghum QTL Atlas: a powerful tool for trait dissection, comparative genomics and crop improvement. *Theor. Appl. Genet.* 132: 751-766.
Mace, E.S. and Jordan, D.R. (2010) Location of major effect genes in sorghum (Sorghum bicolor (L.) Moench). *Theor. Appl. Genet.* 121: 1339-1356.
McCormick, R.F., Truong, S.K., Sreedasyam, A., Jenkins, J., Shu, S.Q., Sims, D., et al. (2018) The Sorghum bicolor reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. *Plant J.* 93: 338-354.
McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20: 1297-1303.
Motzo, R. and Giunta, F. (2002) Awnedness affects grain yield and kernel weight in near-isogenic lines of durum wheat. *Aust. J. Agric. Res.* 53: 1285-1293.
Multani, D.S., Briggs, S.P., Chamberlin, M.A., Blakeslee, J.J., Murphy, A.S. and Johal, G.S. (2003) Loss of an MDR transporter in compact stalks of maize br2 and sorghum dw3 mutants. *Science* 302: 81-84.
Nigorikawa, M., Watanabe, A., Furukawa, K., Sonoki, T. and Ito, Y. (2012) Enhanced saccharification of rice straw by overexpression of rice exo-glucanase. *Rice* 5: 14.
Ogbaga, C.C., Stepien, P. and Johnson, G.N. (2014) Sorghum (Sorghum bicolor) varieties adopt strongly contrasting strategies in response to drought. *Physiol. Plant.* 152: 389-401.
Ohnishi, N., Wacera, W.F. and Sakamoto, W. (2019) Photosynthetic responses to high temperature and strong light suggest potential post-flowering drought tolerance of sorghum Japanese landrace Takakibi. *Plant Cell*

Physiol. 60: 2086-2099.
Ong-Abdullah, M., Ordway, J.M., Jiang, N., Ooi, S.E., Kok, S.Y., Sarpan, N., et al. (2015) Loss of Karma transposon methylation underlies the mantled somaclonal variant of oil palm. *Nature* 525: 533-537.
Paterson, A.H., Bowers, J.E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., et al. (2009) The Sorghum bicolor genome and the diversification of grasses. *Nature* 457: 551-556.
Peng, P., Liu, L.H., Fang, J.J., Zhao, J.F., Yuan, S.J. and Li, X.Y. (2017) The rice TRIANGULAR HULL1 protein acts as a transcriptional repressor in regulating lateral development of spikelet. *Sci. Rep.* 7: 13712.
Perdry, H., Dandine-Roulland, C., Bandyopadhyay, D. and Kettner, L. (2020) gaston: Genetic Data Handling (QC, GRM, LD, PCA) & Linear Mixed Models. R package version 1.5.7.
Price, H.J., Dillon, S.L., Hodnett, G., Rooney, W.L., Ross, L. and Johnston, J.S. (2005) Genome evolution in the genus Sorghum (Poaceae). *Ann. Bot.* 95: 219-227.
R Core Team (2021) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
Ren, D.Y., Rao, Y.C., Wu, L.W., Xu, Q.K., Li, Z.Z., Yu, H.P., et al. (2016) The pleiotropic ABNORMAL FLOWER AND DWARF1 affects plant height, floral development and grain yield in rice. *J. Integr. Plant Biol.* 58: 529-539.
Sato, D.S., Ohmori, Y., Nagashima, H., Toriba, T. and Hirano, H.Y. (2014) A role for TRIANGULAR HULL1 in fine-tuning spikelet morphogenesis in rice. *Genes Genet. Syst.* 89: 61-69.
Sato, S., Ishikawa, S., Shimono, M. and Shinjyo, C. (1996) Genetic studies on an awnness gene An-4 on chromosome 8 in rice, Oryza sativa L. *Breed. Sci.* 46: 321-327.
Schneider, C.A., Rasband, W.S. and Eliceiri, K.W. (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9: 671-675.
Shehzad, T., Iwata, H. and Okuno, K. (2009) Genome-wide association mapping of quantitative traits in sorghum (Sorghum bicolor (L.) Moench) by using multiple models. *Breed. Sci.*59: 217-227.
Sieglinger, J.B., Swanson, A.F. and Martin, J.H. (1934) Inheritance of awn development in Sorghums. *J. Agric. Res.* 49: 663-668.
Sourdille, P., Cadalen, T., Gay, G., Gill, B. and Bernard, M. (2002) Molecular and physical mapping of genes affecting awning in wheat. *Plant Breed.* 121: 320-324.
Takahashi, R. (1955) The origin and evolution of cultivated barley. *Adv. Genet.* 7: 227-266.
Takanashi, H., Shichijo, M., Sakamoto, L., Kajiya-Kanegae, H., Iwata, H., Sakamoto, W., et al. (2021) Genetic dissection of QTLs associated with spikelet-related traits and grain size in sorghum. *Sci. Rep.* 11: 9398.
Taketa, S., Yuo, T., Sakurai, Y., Miyake, S. and Ichii, M. (2011) Molecular mapping of the short awn 2 (lks2) and dense spike 1 (dsp1) genes on barley chromosome 7H. *Breed. Sci.* 61: 80-85.
Tao, Y.Z., Henzell, R.G., Jordan, D.R., Butler, D.G., Kelly, A.M. and McIntyre, C.L. (2000) Identification of genomic regions associated with stay green in sorghum by testing RILs in multiple environments. *Theor. Appl. Genet.* 100: 1225-1232.
Toki, S., Hara, N., Ono, K., Onodera, H., Tagiri, A., Oka, S., et al. (2006) Early infection of scutellum tissue with Agrobacterium allows high-speed transformation of rice. *Plant J.* 47: 969-976.

Toriba, T. and Hirano, H.Y. (2014) The DROOPING LEAF and OsETTIN2 genes promote awn development in rice. *Plant J.* 77: 616-626.
Tuinstra, M.R., Grote, E.M., Goldsbrough, P.B. and Ejeta, G. (1997) Genetic analysis of post-flowering drought tolerance and components of grain development in Sorghum bicolor (L.) Moench. *Mol. Breed.* 3: 439-448.
Vinall, H.N. and Cron, A.B. (1921) Improvement of sorghums by hybridization. *J. Hered.* 12: 435-443.
Watkins, A.E. and Ellerton, S. (1940) Variation and genetics of the awn in Triticum. *J. Genet.* 40: 243-270.
Wickham, H. (2016) ggplot2: Elegant graphics for data analysis. Springer-Verlag New York.
Wu, E., Lenderts, B., Glassman, K., Berezowska-Kaniewska, M., Christensen, H., Asmus, T., et al. (2014) Optimized Agrobacterium-mediated sorghum transformation protocol and molecular data of transgenic sorghum plants. *In Vitro Cell. Dev. Biol. Plant* 50: 9-18.
Yamaguchi, M., Fujimoto, H., Hirano, K., Araki-Nakamura, S., Ohmae-Shinohara, K., Fujii, A., et al. (2016) Sorghum Dw1, an agronomically important gene for lodging resistance, encodes a novel protein involved in cell proliferation. *Sci. Rep.* 6: 28366.
Yamazaki, K., Ishimori, M., Kajiya-Kanegae, H., Takanashi, H., Fujimoto, M., Yoneda, J., et al. (2020) Effect of salt tolerance on biomass production in a large population of sorghum accessions. *Breed. Sci.* 70: 167-175.
Yano, K., Yamamoto, E., Aya, K., Takeuchi, H., Lo, P.C., Hu, L., et al. (2016) Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. *Nat. Genet.* 48: 927-934.
Yoshida, A., Sasao, M., Yasuno, N., Takagi, K., Daimon, Y., Chen, R.H., et al. (2013) TAWAWA1, a regulator of rice inflorescence architecture, functions through the suppression of meristem phase transition. *Proc. Natl. Acad. Sci. U. S. A.* 110: 767-772.
Yoshida, A., Suzaki, T., Tanaka, W. and Hirano, H.Y. (2009) The homeotic gene long sterile lemma (G1) specifies sterile lemma identity in the rice spikelet. *Proc. Natl. Acad. Sci. U. S. A.* 106: 20103-20108.
Yoshioka, M., Iehisa, J.C.M., Ohno, R., Kimura, T., Enoki, H., Nishimura, S., et al. (2017) Three dominant awnless genes in common wheat: Fine mapping, interaction and contribution to diversity in awn shape and length. *PLOS ONE* 12: e0176148.
Yuo, T., Yamashita, Y., Kanamori, H., Matsumoto, T., Lundqvist, U., Sato, K., et al. (2012) A SHORT INTERNODES (SHI) family transcription factor gene regulates awn elongation and pistil morphology in barley. *J. Exp. Bot.* 63: 5223-5232.
Zhang, D., Kong, W., Robertson, J., Goff, V.H., Epps, E., Kerr, A., et al. (2015) Genetic analysis of inflorescence and plant height components in sorghum (Panicoidae) and comparative genetics with rice (Oryzoidae). *BMC Plant Biol.* 15: 107.
Zhang, Y.P., Zhang, Z.Y., Sun, X.M., Zhu, X.Y., Li, B., Li, J.J., et al. (2019) Natural alleles of GLA for grain length and awn development were differently domesticated in rice subspecies japonica and indica. *Plant Biotechnol. J.* 17: 1547–1559.
Zhou, L., Zhu, C., Fang, X., Liu, H., Zhong, S., Li, Y., et al. (2021) Gene duplication drove the loss of awn in sorghum. *Mol. Plant* doi: 10.1016/j.molp.2021.07.005.

Table 1. Summary of awn-related quantitative trait loci (QTLs) detected in this study.

Trait	Year_location	Chr	QTL ID	Position (cM)	Nearest marker	Marker interval	LOD	PVE (%)	AE
Awn length (AL)	2014_OKY	3	*qAL3*	76.62	Chr03:72655415	Chr03:71923025–72847396	116.4	90.6	5.75
	2015_TKY	3	*qAL3*	76.62	Chr03:72655415	Chr03:71923025–72847396	100.8	88.7	4.58
	2014_OKY	9	*qAL9*	59.76	Chr09:57681027	Chr09:56467317–58331590	7.0	1.3	0.64
Awn presence (AP)	2014_OKY	3	*qAP3*	76.62	Chr03:72655415	Chr03:71923025–72847396	179.1	97.8	0.49
	2015_TKY	3	*qAP3*	76.62	Chr03:72655415	Chr03:71923025–72847396	120.8	92.7	0.48
Awn length (awned lines only)	2014_OKY	7	*qAL7*	47.27	Chr07:59733246	Chr07:58230302–60603839	8.3	23.8	1.06
	2015_TKY	7	*qAL7*	47.40	Chr07:59733246	Chr07:58230302–60603839	3.5	10.7	0.64
	2014_OKY	9	*qAL9*	59.76	Chr09:57681027	Chr09:56467317–57917278	5.9	16.2	0.87
	2015_TKY	9	*qAL9*	60.06	Chr09:57681027	Chr09:56467317–58331590	4.8	15.0	0.76

OKY, Okayama; TKY, Tokyo. Marker intervals were estimated based on confidence intervals (2.0-LOD). PVE: Phenotypic variance explained. For the additive effects (AE), positive values indicate that alleles from NOG increased the trait score.

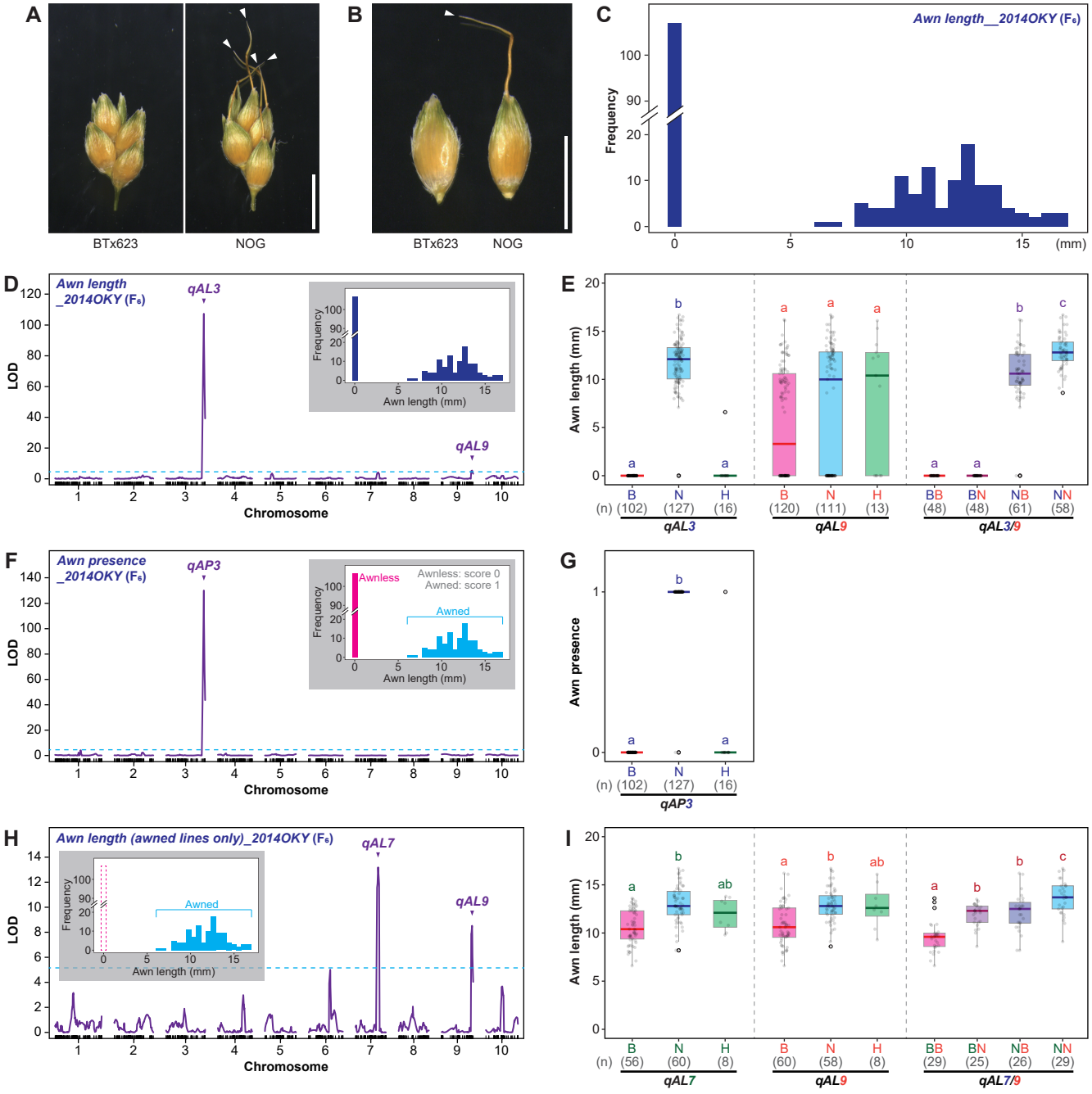


Figure 1

Quantitative trait locus (QTL) analysis for awn-related traits using the sorghum recombinant inbred line (RIL) population. (A) Secondary branches of BTx623 and NOG, parents of the RIL population. (B) Enlarged view of spikelets from BTx623 and NOG. Arrowheads indicate awn tips. (C) Frequency distribution of awn length (AL) in the RIL population (F_6 generation) cultivated at Okayama in 2014. Results of QTL analysis for the awn-related traits (D, F, H) and their allelic effects (E, G, I). (D, E) AL of all lines, (F, G) awn presence (AP), (H, I) AL of awned lines only. (D, F, H) LOD profiles obtained from composite interval mapping (CIM). Horizontal dotted lines represent a threshold of the 1,000× permutation test ($P < 0.05$). (E, G, I) Contributions of SNP genotypes of the nearest marker for significant QTLs. Box plots show effects of the nearest marker genotypes for each QTL or allelic combinations of QTLs. B, BTx623-type homogeneous allele; N, NOG-type homogeneous allele; H, heterogeneous. The number of lines that corresponded to each genotype is shown in parentheses. Different letters denote significant differences according to the Tukey-Kramer test ($P < 0.01$). Scale bars = 5 mm in (A) and (B).

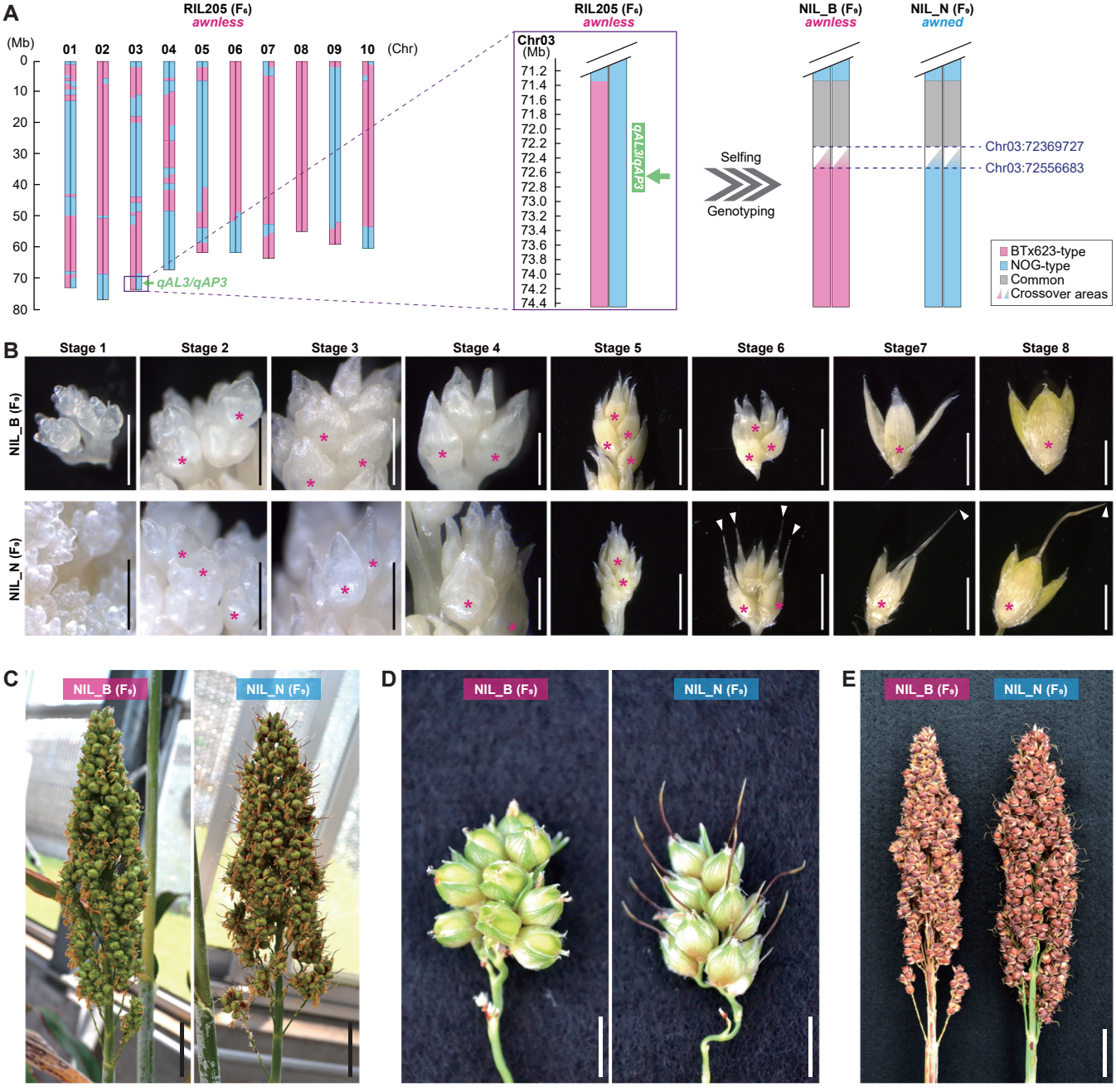


Figure 2

Reconfirming the effect of *qAL3*/*qAP3* using near isogenic lines (NILs). (A) Graphical genotypes of RIL205 (F_6, awnless), one of the heterozygous lines at the genotype of the marker associated with *qAL3*/*qAP3* (left), and magnified view of the Chr03 terminus for RIL205, NIL with the homogeneous BTx623 allele (NIL_B, F_9, awnless), and that with the homogeneous NOG allele (NIL_N, F_9, awned) at *qAL3*/*qAP3* (right). Magenta and cyan boxes indicate genomic regions from BTx623 and NOG, respectively. PCR genotyping using insertion-deletion polymorphisms revealed that recombination occurred between Chr03:72369727–Chr03:72556683 in both NILs; differences between NILs at the Chr03 terminus were approximately 1.83–2.02 Mb. "Common" indicates that the genotypes (homogeneous allele) are common between both NILs, although we could not determine whether they are BTx623-type or NOG-type by PCR genotyping. A bar and arrow in green indicate the marker interval and the position of the nearest marker of *qAL3*/*qAP3* in the QTL analysis. (B–E) Observations of spikelets and panicles of NILs at various growth stages. (B) Spikelets of young panicles. Staging of young spikelets was carried out based on the length of sessile spikelets (SSs, asterisks). Stage 1, primordial stage; Stage 2, 0.3–0.4 mm; Stage 3, 0.4–0.6 mm; Stage 4, 0.6–0.8 mm; Stage 5, 1.5–2.0 mm; Stage 6, 2.0–2.5 mm; Stage 7, 2.6–3.3 mm; and Stage 8, 3.7–4.2 mm, respectively. Arrowheads indicate awn tips. Scale bars = 0.5 mm in Stage 1–4, 2.0 mm in Stage 5–8. (C) Panicles at the grain-filling period. (D) Enlarged images of the primary branches at the grain-filling period. (E) Panicles at the mature stage. Scale bars = 3 cm in (C) and (E), 5 mm in (D).

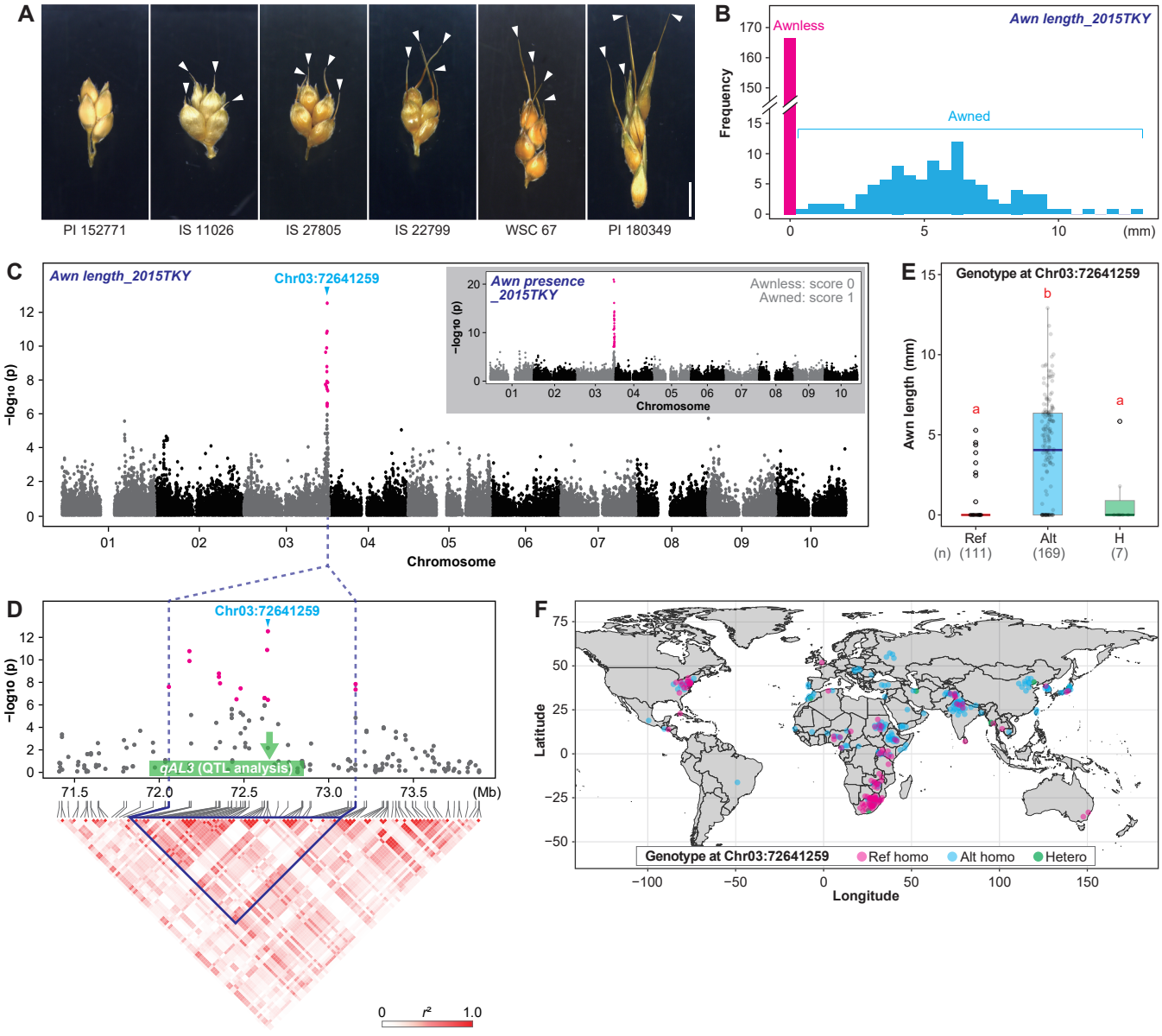
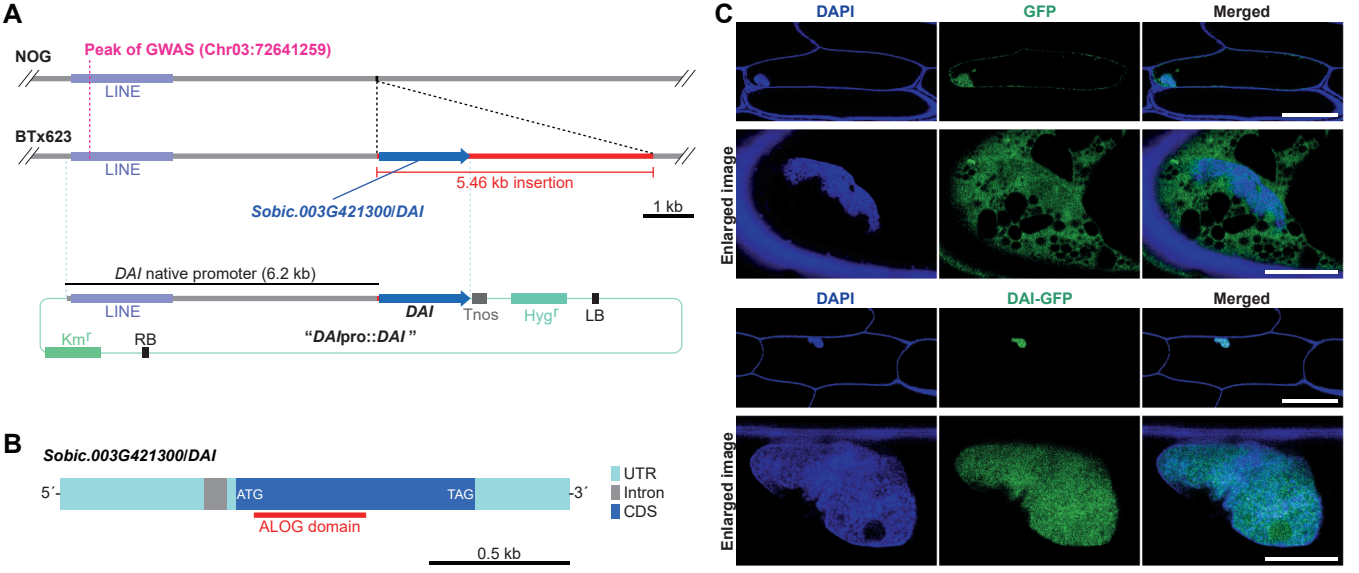


Figure 3

Genome-wide association study (GWAS) for awn-related traits using the sorghum germplasm. (A) Examples of secondary branches from the panicles of sorghum germplasm. Accession ID is shown on the bottom of each panel. Arrowheads indicate awn tips. (B) Frequency distribution of AL in the sorghum germplasm cultivated at Tokyo in 2015. (C) Manhattan plot for AL and AP (insert). Dots in magenta indicate the single-nucleotide polymorphisms (SNPs) that are significantly associated with traits (Benjamini–Hochberg-corrected significance; $P < 0.001$). (D) Local Manhattan plot (top) and LD heatmap (bottom) surrounding the peak on chromosome 3. A bar and arrow in green indicate the marker interval and the position of the nearest marker of *qAL3*/*qAP3* in the QTL analysis. (E) Box plot shows the effects of the peak SNP (Chr03:72641259) genotype for AL. Ref, reference-type homogeneous allele; Alt, altered-type homogeneous allele; H, heterogeneous. The number of lines that corresponded to each genotype is shown in parentheses. Different letters denote significant differences according to the Tukey-Kramer test ($P < 0.01$). (F) Geographic distribution of alleles at the peak SNP (Chr03:72641259), color-coded by allele (Ref homo, magenta; Alt homo, cyan; Hetero, green). Scale bar = 5 mm in (A).



D

Accession ID	Origin	Genotype at Chr03:72641259	Phenotype	Presence of *DAI* (5.46 kb insertion)	Mutation in the CDS of *DAI*
PI 576320	United States of America	A (Ref homo)	Awnless	**Present**	No mutation
PI 533900	Japan	A (Ref homo)	Awnless	**Present**	No mutation
PI 295123	Australia	A (Ref homo)	Awnless	**Present**	No mutation
WSC 01	Japan	A (Ref homo)	Awnless	**Present**	No mutation
WSC 105	Zimbabwe	A (Ref homo)	Awnless	**Present**	No mutation
PI 180349	India	G (Alt homo)	Awned	**Absent**	-
PI 173120	Turkey	G (Alt homo)	Awned	**Absent**	-
PI 586554	China	G (Alt homo)	Awned	**Absent**	-
PI 179052	Turkey	G (Alt homo)	Awned	**Absent**	-
PI 196597	Taiwan	G (Alt homo)	Awned	**Absent**	-
PI 656029	United States of America	A (Ref homo)	Awned	**Absent**	-
IS 24453	South Africa	A (Ref homo)	Awned	**Present**	C to T substitution at 82nd position (**Q_{28} →Stop)**
PI 152771	Sudan	G (Alt homo)	Awnless	**Present**	No mutation
IS 29606	South Africa	G (Alt homo)	Awnless	**Present**	No mutation

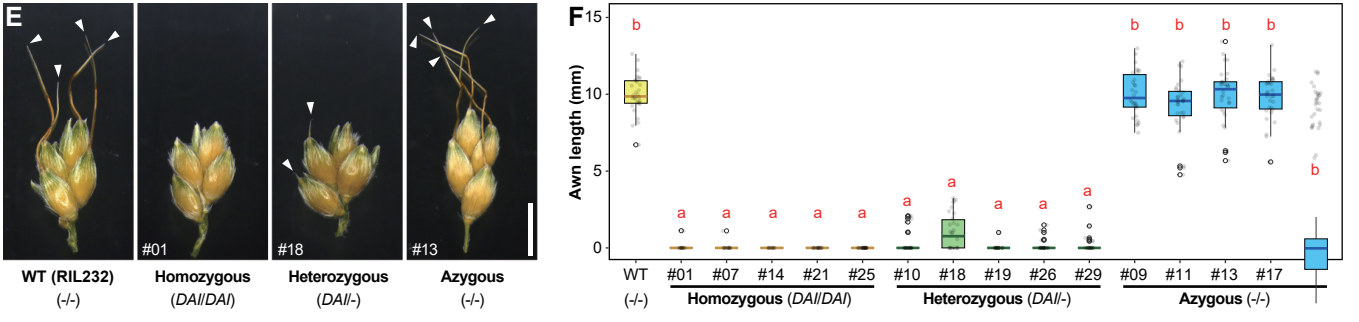


Figure 4

Identification of the *Dominant Awn Inhibitor* (*DAI*) gene. (A) Genomic structures surrounding *DAI* on chromosome 3 in the NOG and BTx623 genome (upper), and schematics of the *DAI*pro::*DAI* plasmid for making transgenic sorghum (bottom). (B) The gene structure of *DAI* in BTx623 is shown. Pale blue, gray, and blue boxes show UTR, intron, and CDS, respectively. A red line indicates a region coding the ALOG domain. (C) Localization of green florescent protein (GFP) (upper) and DAI-GFP fusion protein (bottom) in onion epidermal cells. (D) Table showing the relationship among genotype of the peak SNP in GWAS (Chr03:72641259), awn phenotype, presence of *DAI*, and mutations within the CDS of *DAI* in 14 accessions of sorghum germplasm. (E) Examples of secondary branches from panicles of the awned parental line (RIL232), homozygous transgenic plant (*DAI*/*DAI*), heterozygous transgenic plant (*DAI*/-), and azygous plant (-/-). Arrowheads indicate awn tips. (F) Box plot shows the AL of the RIL232, homozygous, heterozygous, and azygous plants. Different letters denote significant differences according to the Tukey-Kramer test ($P < 0.01$). Scale bars = 100 μm in (C; 10 μm in Enlarged images), 5 mm in (E).

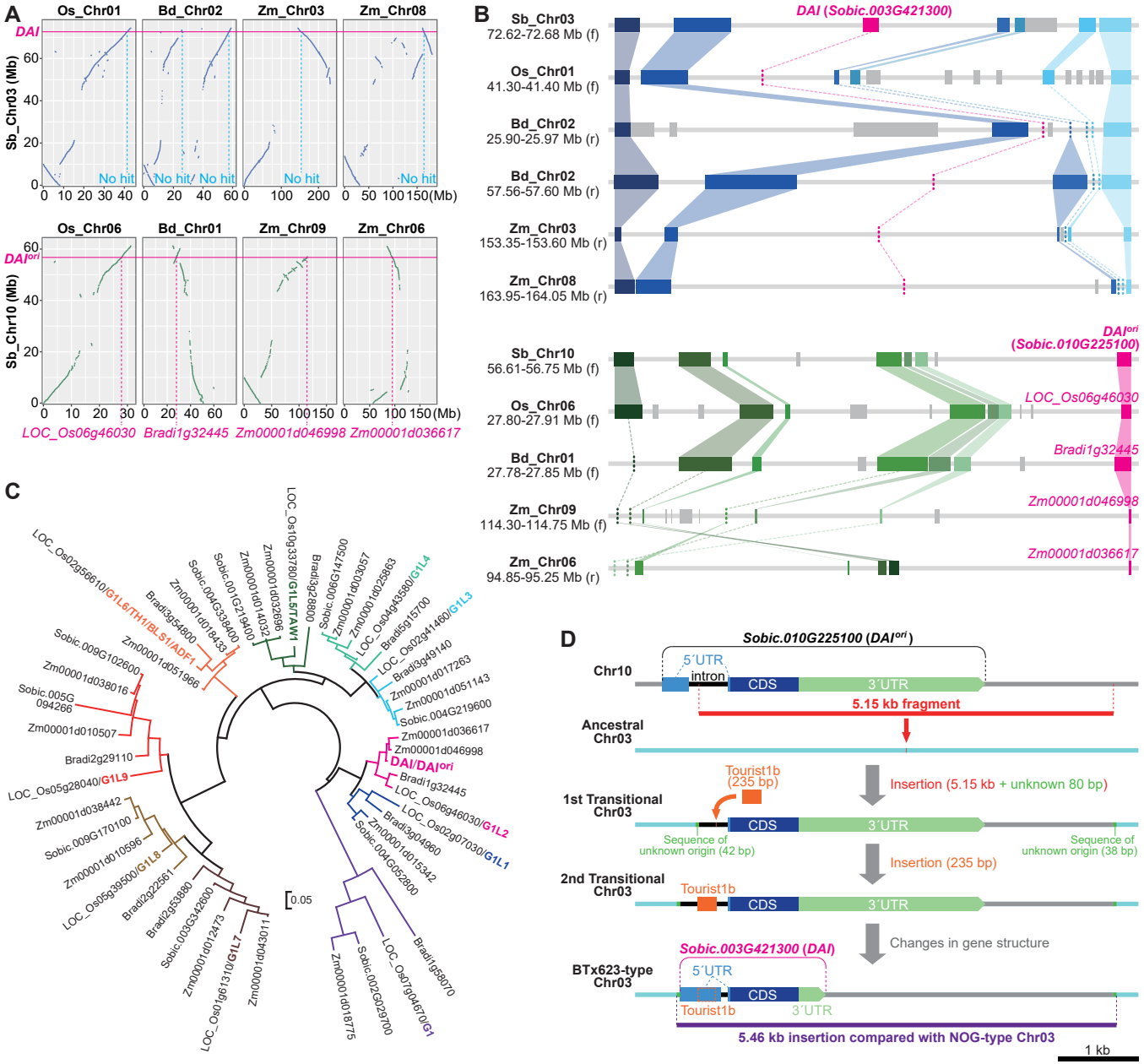


Figure 5

The birth and evolution of the *DAI* gene in sorghum. (A, B) Synteny analysis for the *DAI* gene and its paralogous gene, *DAI*^{ori} among four grass species. (A) Chromosome-level synteny analysis for sorghum chromosome 3 encoding *DAI*, and 10 encoding *DAI*^{ori}. Sb, *Sorghum bicolor*; Os, *Oryza sativa*; Bd, *Brachypodium distachyon*; and Zm, *Zea mays*. (B) Schematic diagram of the genomic structures around *SbDAI*, *SbDAI*^{ori}, and their syntenic regions in four grass species. Boxes in the same color indicate that they are orthologous genes. Genes shown in dotted lines indicate that corresponding genes do not exist in this region. 'f' and 'r' indicate forward and reverse strand, respectively. (C) Phylogenetic tree for DAI/DAI^{ori} and related proteins from the four grass species. (D) A model for the birth and evolution of the *DAI* gene. Gene structures of *DAI* and *DAI*^{ori} are based on the information of the reference genome of *Sorghum bicolor* (v3.1.1).

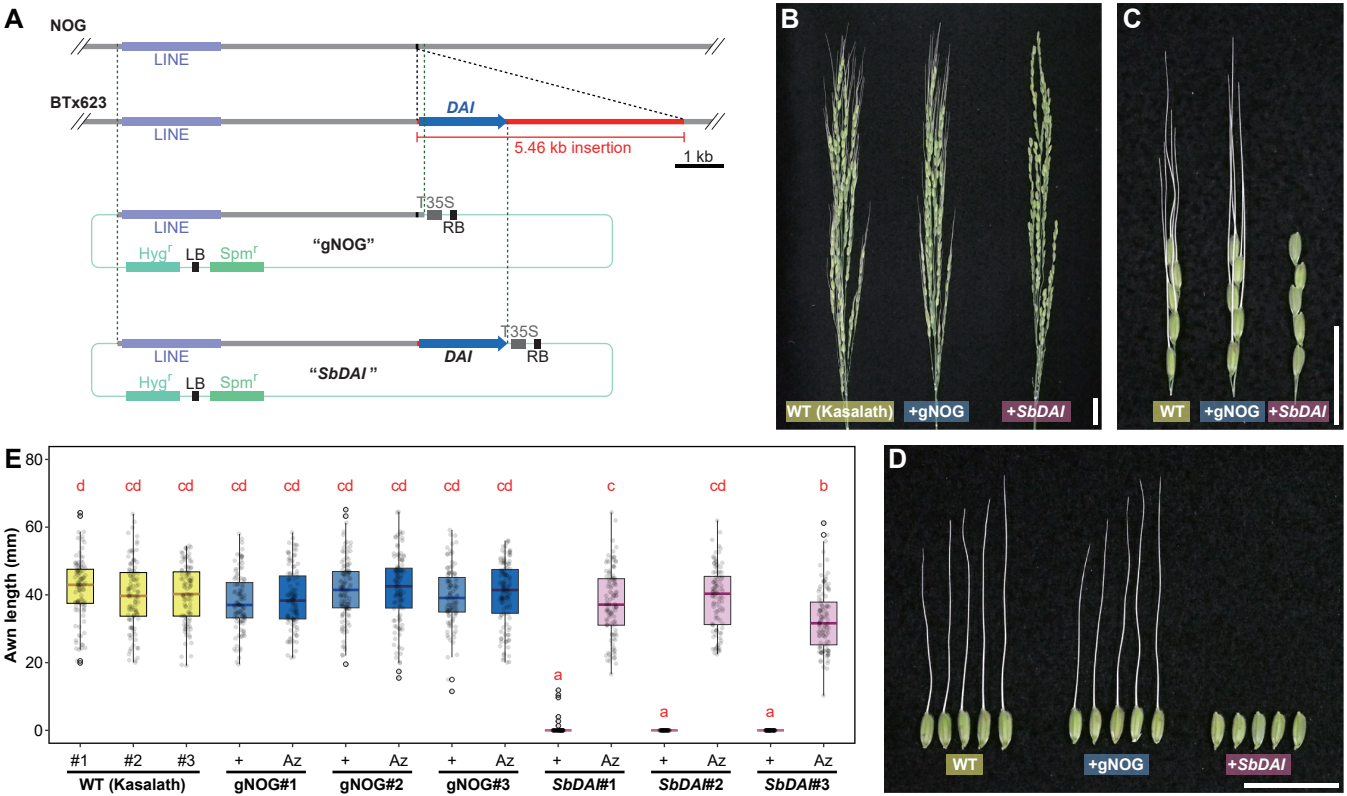
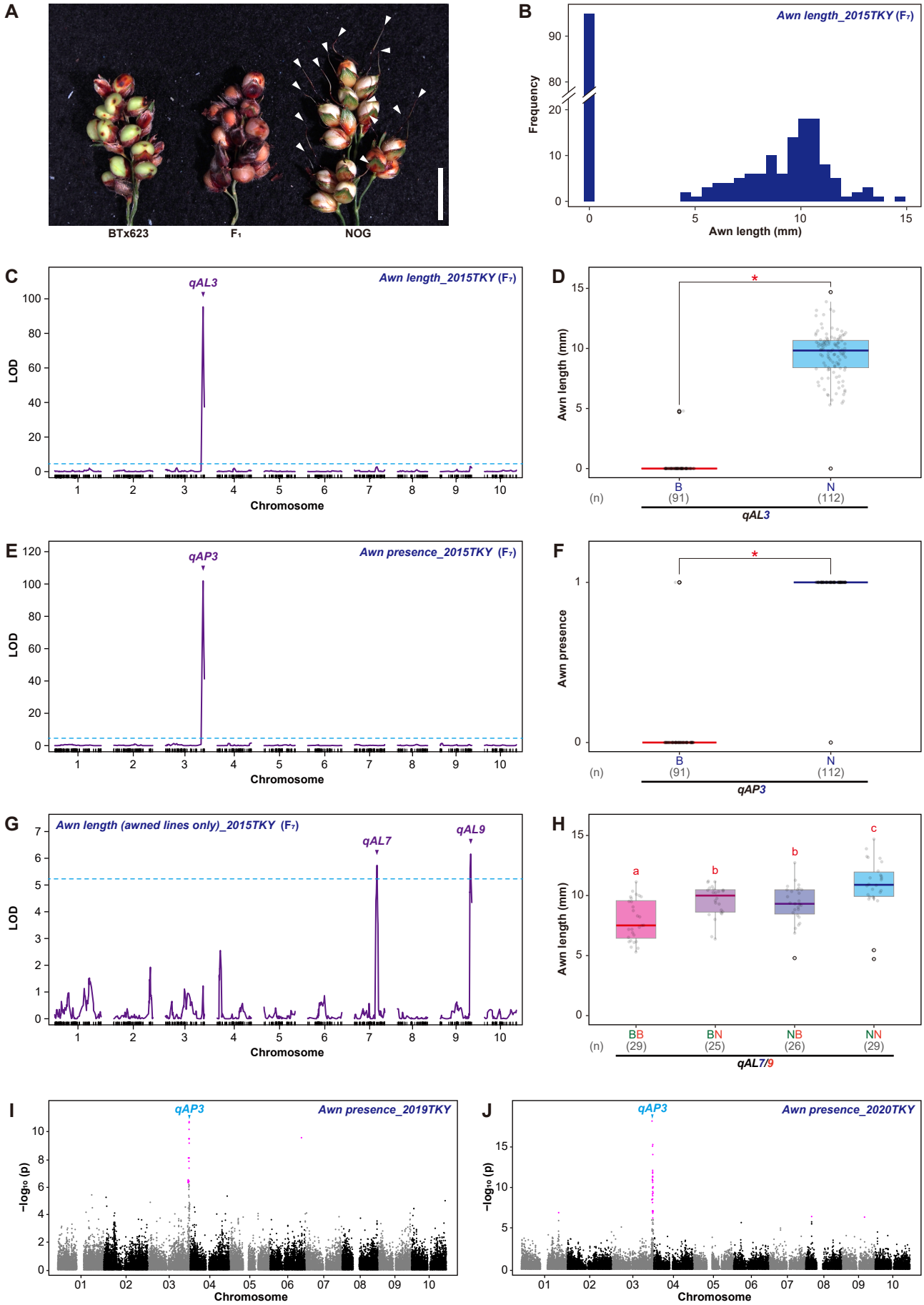


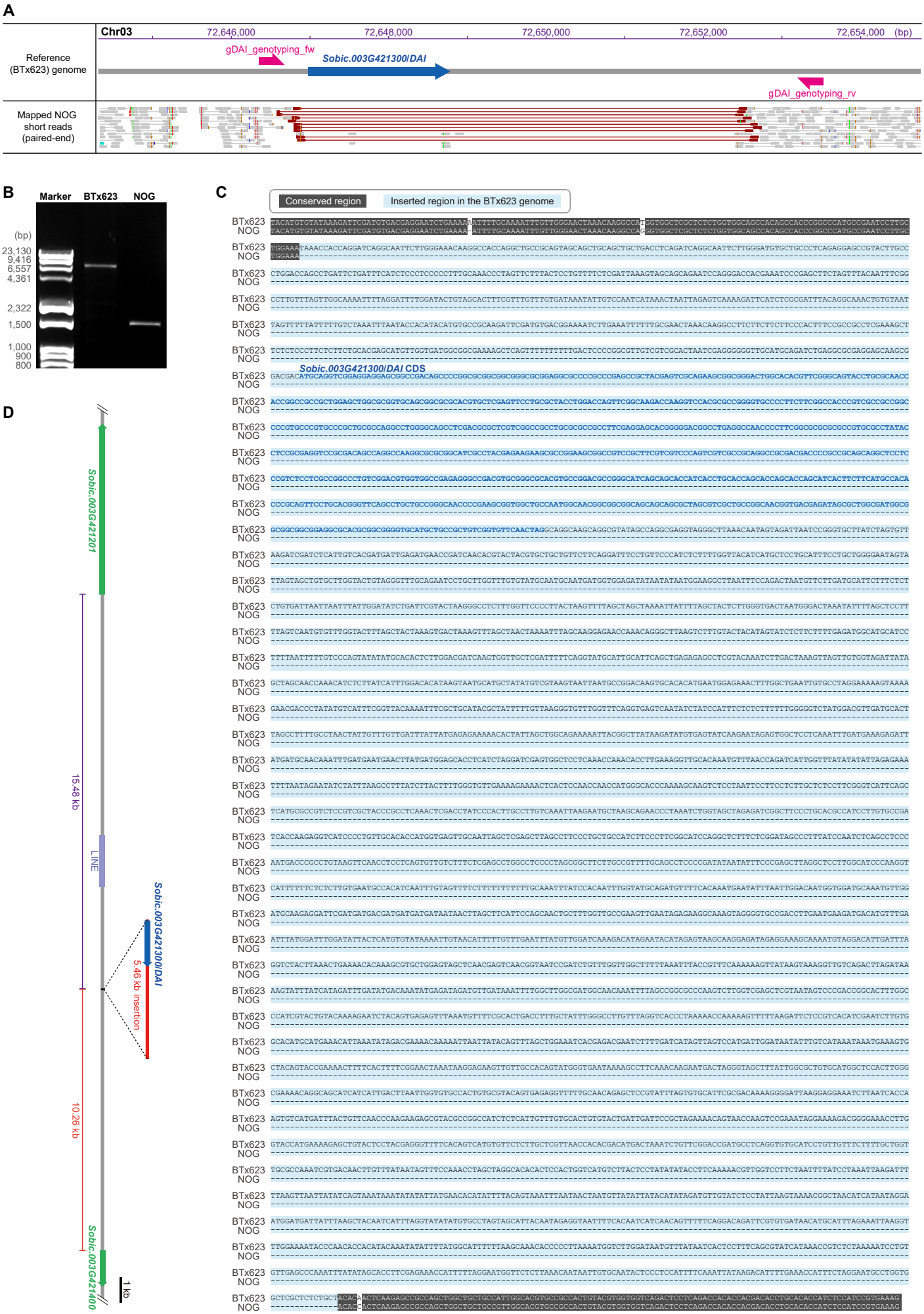
Figure 6

SbDAI can also inhibit awn elongation in Kasalath. (A) Schematics of the gNOG and *SbDAI* plasmids for making transgenic Kasalath. (B–D) Examples of panicles (B), secondary branches (C), and spikelets (D) of the wild-type Kasalath, the gNOG transgenic line, and the *SbDAI* transgenic line. (E) Box plot shows the AL of the wild-type Kasalath, gNOG transgenic lines, *SbDAI* transgenic lines, and their azygous lines. '+' or 'Az' indicate the individual has the transgene or does not, respectively. Different letters denote significant differences according to the Tukey-Kramer test ($P < 0.01$). Scale bars = 2 cm in (B–D).



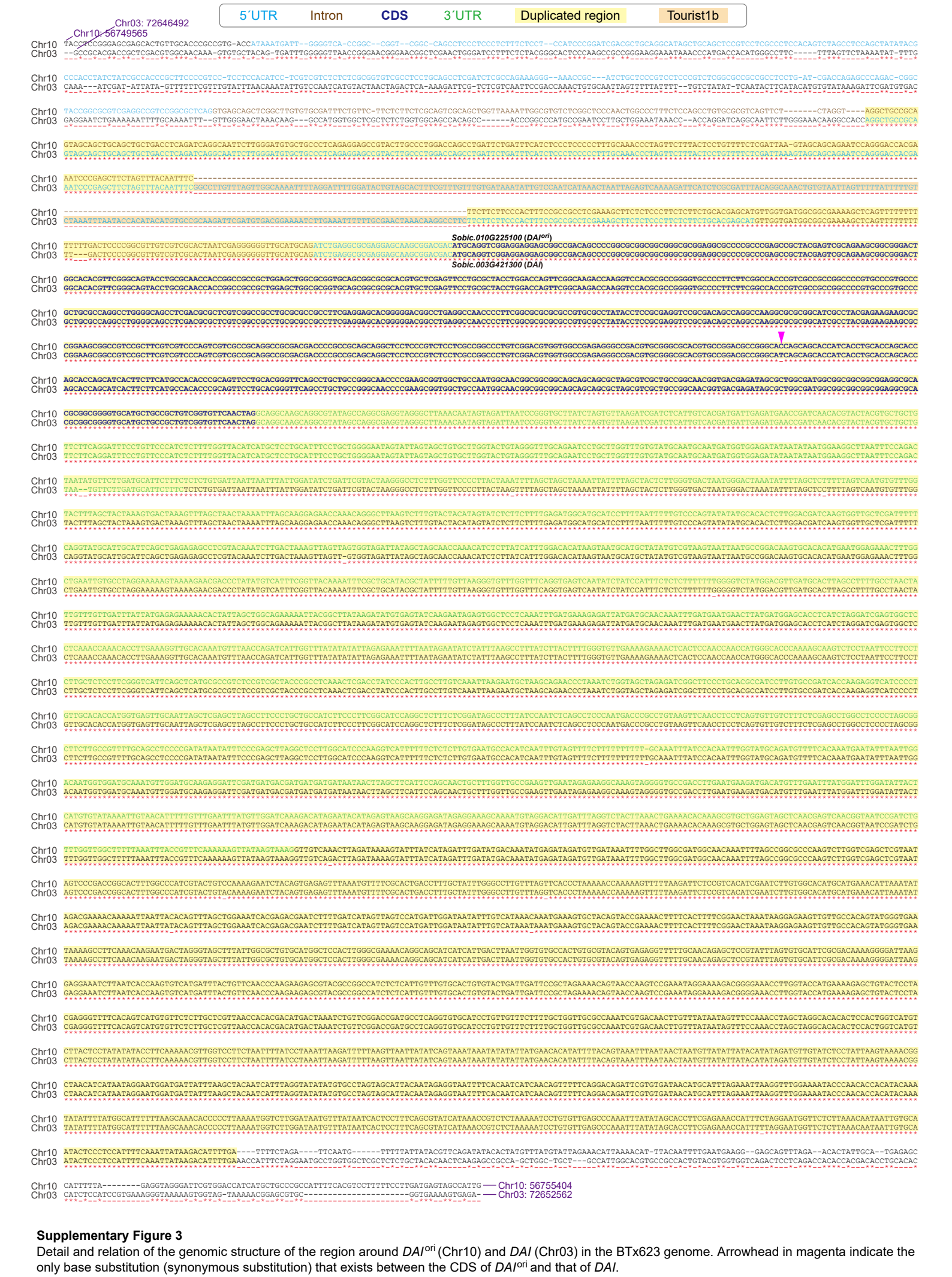
Supplementary Figure 1

Results of reproducibility check of QTL analysis and GWAS using lines cultivated in different years [2015 at Tokyo for RILs (F_7 generation), 2019 and 2020 at Tokyo for germplasm] from those in Fig. 1 and 2. (A) Awn phenotype of F_1 and its parents. Arrowheads indicate awn tips. (B) Frequency distribution of awn length in the RIL population (F_7 generation) cultivated at Tokyo in 2015. Results of QTL analysis for the awn-related traits (C, E, G) and their allelic effects (D, F, H). (C, D) Awn length of all lines, (E, F) awn presence, (G, H) awn length of awned lines only. (C, E, G) LOD profiles obtained from composite interval mapping (CIM). Horizontal dotted lines represent a threshold of the 1,000× permutation test ($P < 0.05$). (D, F, H) Contributions of SNP genotypes of the nearest marker for significant QTLs. Box plots show the effects of the nearest marker genotypes for each QTL or allelic combinations of QTLs. B, BTx623-type homogeneous allele; N, NOG-type homogeneous allele. The number of lines that corresponded to each genotype is shown in parentheses. Asterisks indicate significant differences between genotypes (Welch' s *t*-test, $P < 0.001$). Different letters denote significant differences according to the Tukey-Kramer test ($P < 0.05$). (I, J) Manhattan plots for awn presence in 2019 and 2020. Dots in magenta indicate the SNPs that were significantly associated with traits (Bonferroni-corrected significance, $P < 0.001$). Scale bar = 1 cm in (A).



Supplementary Figure 2

Difference between genomic structures of BTx623 and NOG around the *DAI* gene. (A) Re-sequencing data of NOG in the region around *DAI* visualized using Integrative Genomics Viewer. Pentagonal boxes represent each short read, and the paired reads are connected to each other by lines. The paired reads shown in red indicate that the seize of inserts are unexpectedly large, i.e., NOG genome does not have regions between them. (B) Result of genotyping of BTx623 and NOG by PCR using primers indicated in (A). (C) Details of the genomic structure of NOG and BTx623 around the *DAI* gene as revealed by Sanger sequencing of PCR products in (B). (D) Insertion position of the 5.46 kb fragment and its relationship to nearby genes.



Supplementary Figure 3

Detail and relation of the genomic structure of the region around *DAI*^{ori} (Chr10) and *DAI* (Chr03) in the BTx623 genome. Arrowhead in magenta indicate the only base substitution (synonymous substitution) that exists between the CDS of *DAI*^{ori} and that of *DAI*.

Supplementary Table 1. Sorghum germplasm examined in this study

Name	Accession ID	Origin	Race
IS 11026	IS 11026	Ethiopia	Durra
IS 11034	PI 329276	Ethiopia	Not found
BTx642	PI 656029	United States of America	Not found
IS 25301	IS 25301	Ethiopia	Durra-bicolor
IS 25249	IS 25249	Ethiopia	Durra-bicolor
Mashila	PI 576427	Ethiopia	Not found
IS 24503	IS 24503	South Africa	Bicolor
433	PI 300120	South Africa	Unknown
IS 25836	IS 25836	Mali	Durra
SC 426	PI 533900	Japan	Not found
IS 10969	IS 10969	United States of America	Guinea-caudatum
IS 20679	IS 20679	United States of America	Guinea-caudatum
IS 12706	IS 12706	United States of America	Caudatum-bicolor
IS 23586	IS 23586	Ethiopia	Guinea-caudatum
IS 6900	IS 6900	Niger	Not found
KA 24	WSC 89	Nigeria	Not found
MARIANGARIJORA MUDDAHIHAL	WSC 32	India	Not found
HEGARI MALOWAR	WSC 75/PI 570998	Sudan	Caudatum
AS 4547 JARDIRA	WSC 87	Nigeria	Not found
Juar	PI 180349	India	Durra
FETERITA HARIG 2	PI 152654	Sudan	Durra
65I 2786	PI 643625	Mexico	Not found
IS 11473	IS 11473	Ethiopia	Caudatum
AKLMOI WHITE	WSC 100	Kenya	Not found
IS 7310	IS 7310	Nigeria	Guinea
IS 15478	IS 15478	Cameroon	Guinea-caudatum
IS 30838	IS 30838	Cameroon	Guinea
IS 14131	IS 14131	Portugal	Not found
MN 4177	PI 302177	Portugal	Not found
IS 20743	IS 20743	United States of America	Bicolor
IS 2413	IS 2413	Iran	Bicolor
SOR 1	PI 576028	Romania	Not found
8493	PI 173120	Turkey	Bicolor
IS 3127	PI 211747	Afghanistan	Bicolor
IS 24953	IS 24953	Zambia	Guinea-caudatum
IS 24939	IS 24939	Zambia	Bicolor
IS 24957	PI 585445	Zambia	Not found
IS 8777	IS 8777	Uganda	Caudatum-bicolor
IS 9249	IS 9249	Uganda	Not found
IS 19670	IS 19670	Zimbabwe	Not found
IS 19669	IS 19669	Zimbabwe	Not found
IS 23684	IS 23684	Mozambique	Guinea
IS 23216	IS 23216	Zambia	Caudatum-bicolor
IS 473	IS 473	United States of America	Guinea-kafir
CE260-12-1-1	PI 656032	Senegal	Not found
IS 18712	IS 18712	United States of America	Not found
IS 21645	IS 21645	Malawi	Guinea
IS 21512	IS 21512	Malawi	Guinea
IS 21599	IS 21599	Malawi	Not found
IS 2382	IS 2382	South Africa	Caudatum
IS 5301	IS 5301	India	Guinea-caudatum
IS 32349	IS 32349	India	Guinea
IS 4951	IS 4951	India	Guinea
IS 2379	IS 2379	South Africa	Caudatum
LWALLI WHITE	PI 217837	Sudan	Caudatum
IS 10302	IS 10302	Thailand	Caudatum
ZIRA-EL-SABI	PI 569885	Sudan	Caudatum
IS 9745	IS 9745	Sudan	Caudatum
WAD YABOO 132/53	WSC 103	Zimbabwe	Not found
PI 152748 C	WSC 102	Kenya	Not found
IS 15170	IS 15170	Cameroon	Caudatum
IS 7305	IS 7305	Nigeria	Caudatum
IS 22986	IS 22986	Sudan	Caudatum
UGANDA L1	WSC 94/PI 571225	Sudan	Caudatum
E 276 FRAMIDA	WSC 93	Uganda	Not found
IS 29950	IS 29950	Zimbabwe	Guinea-caudatum
IS 20697	IS 20697	United States of America	Caudatum
RAHMETALLA GALLABAT	PI 152771	Sudan	Not found
IS 2815	IS 2815	Zimbabwe	Not found
143 DINDERAWI 1	WSC 76	Sudan	Not found
IS 8283	IS 8283	Uganda	Not found
IS 22609	IS 22609	Sri Lanka	Caudatum
IS 8916	IS 8916	Uganda	Guinea-caudatum
IS 21083	IS 21083	Kenya	Caudatum
SAP-155	PI 576437	Brazil	Not found
IS 32439	IS 32439	India	Guinea
IS 8267	PI 648213	Uganda	Not found
S. VULGARE 72-726-7	WSC 91	Uganda	Not found
E 17	WSC 58	Chad	Not found
IS 3547	IS 3547	Sudan	Not found
SC 170	PI 534157	Ethiopia	Not found
B3Tx2817	PI 593910	United States of America	Not found
TX2783	PI 656001	United States of America	Not found
IS 23521	IS 23521	Ethiopia	Guinea-caudatum
IS 23514	IS 23514	Ethiopia	Caudatum
AGIRA	PI 217855	Sudan	Caudatum
SC 112	WSC 97	Ethiopia	Not found
IS 23579	IS 23579	Ethiopia	Guinea-caudatum
RCV	PI 656008	El Salvador	Not found
IS 18758	IS 18758	Ethiopia	Not found
IS 19153	IS 19153	Sudan	Guinea-caudatum
IS 18758	IS 18758	Ethiopia	Not found
MUGBASH WHITE	WSC 82/PI 570748	Sudan	Guinea
240 WAD UMM BENEIN	WSC 81	Sudan	Not found
B-112	WSC 83	Sudan	Not found
MILO PET. 139/51 EX TANGANYIKA	WSC 74	Central African Republicn Republic	Not found
BATTANBAN	WSC 26	Cambodia	Not found
IS 24348	IS 24348	India	Caudatum
IS 17941	IS 17941	India	Caudatum
JUNELO	WSC 53	Nepal	Not found
ZA113 DAWA PAS PARA	WSC 86	Nigeria	Not found

Supplementary Table 1. Sorghum germplasm examined in this study_continued

Name	Accession ID	Origin	Race
E 1091	WSC 84	Sudan	Not found
PI 229486 VULGARE	WSC 54	Iran	Not found
CODY	WSC 63	Morocco	Not found
OOTOYO-MURA ZAIRAI	WSC 01	Japan	Not found
KANAGAWAZAIRAI	WSC 10	Japan	Not found
AS 5781 HUAN SA PHAUNG AH LPYSU	WSC 27	Myanmar	Not found
E 9	WSC 56	Chad	Not found
SCHROCK	WSC 66	Morocco	Not found
PI 220636 Q2/3/56	WSC 50	Afghanistan	Not found
IS 20727	IS 20727	United States of America	Bicolor
MN 1277 MUHEYAR	WSC 88	Nigeria	Not found
LIAOZA 1	WSC 23	China	Not found
IS 35957	IS 35957	Union of Soviet Socialist Republic	Not found
IS 16151	IS 16151	Cameroon	Caudatum-bicolor
IS 8871	IS 8871	Uganda	Not found
IS 27805	IS 27805	Hungary	Not found
IS 19397	IS 19397	Union of Soviet Socialist Republic	Not found
IS 14154	IS 14154	Union of Soviet Socialist Republic	Not found
IS 27806	IS 27806	Hungary	Not found
IS 27794	IS 27794	Hungary	Not found
IS 613	IS 613	United States of America	Not found
IS 35803	IS 35803	Union of Soviet Socialist Republic	Not found
KIKUCHI ZAIRAI	WSC 08	Japan	Not found
HIMEKI ZAIRAI	WSC 07	Japan	Not found
IS 8012	IS 8012	Japan	Bicolor
HONGKER	PI 586554	China	Caudatum-bicolor
IS 20740	IS 20740	United States of America	Bicolor
COL/PAK/1989/IBPGR/2427(5)	WSC 39	Pakistan	Not found
ITALIAN	PI 196597	Taiwan	Mixed
TOKIBI	WSC 06	Japan	Not found
TAKAKIMI	WSC 02	Japan	Not found
NOG	NOG	Japan	Not found
IKEDACHO MATSUO ZAIRAI	WSC 03	Japan	Not found
XIONG YUE 334	WSC 25	China	Not found
72-8-13	WSC 19	Taiwan	Not found
IS 30536	IS 30536	South Korea/North Korea	Caudatum-bicolor
MOCTAC LOCAL	WSC 16	South Korea/North Korea	Not found
IS 30508	IS 30508	South Korea/North Korea	Caudatum-bicolor
KOUSHUU ZAIRAISHU	WSC 13	South Korea/North Korea	Not found
IS 30417	IS 30417	China	Caudatum-bicolor
IS 20383	IS 30383	China	Caudatum-bicolor
IS 30460	IS 30460	China	Caudatum
BIG WHITE HULL	WSC 24	China	Guinea-caudatum
HANGETSUTOSUI	WSC 12	South Korea/North Korea	Not found
AI HUI	WSC 20	China	Not found
IS 30562	IS 30562	South Korea/North Korea	Bicolor
IS 30400	IS 30400	China	Caudatum-bicolor
IS 1233	IS 1233	China	Bicolor
AKAHO	WSC 09	Japan	Not found
IS 13570	IS 13570	China	Not found
IS 31681	IS 31681	Algeria	Bicolor
Dhurra	PI 532235	Oman	Bicolor
IS 27786	IS 27786	Morocco	Durra-bicolor
PHATSAI	WSC 65	Morocco	Not found
15065	PI 562402	Saudi Arabia	Not found
IS 21863	IS 21863	Syria	Bicolor
KOCADARI	PI 179052	Turkey	Not found
IS 12852	IS 12852	Turkey	Not found
IS 28747	IS 28747	Yemen	Durra-caudatum
IS 28449	IS 28449	Yemen	Guinea-caudatum
IS 28614	IS 28614	Yemen	Durra-caudatum
IS 32245	IS 32245	Yemen	Durra-caudatum
IS 13122	IS 13122	Pakistan	Not found
IS 5919	IS 5919	India	Durra
IS 22799	IS 22799	Somalia	Durra
IS 22720	IS 22720	Somalia	Durra
IS 4360	IS 4360	India	Durra
IS 3971	IS 3971	India	Durra
IS 12883	IS 12883	India	Durra
IS 8348	IS 8348	Pakistan	Durra
COL/PAK/1989/IBPGR/2420(1)	WSC 38	Pakistan	Not found
COL/PAK/1991/IBPGR/2724(2)	WSC 37	Pakistan	Not found
EC 18868	WSC 52	Nepal	Not found
GOOSENECK	WSC 31	India	Not found
IS 4060	IS 4060	India	Durra-bicolor
IS 13118	IS 13118	Pakistan	Not found
COL/PAK/1989/IBPGR/2386(2)	WSC 36	Pakistan	Not found
COL/PAK/1989/IBPGR/2416(2)	WSC 45	Pakistan	Not found
COL/PAK/1989/IBPGR/2411(1)	WSC 44	Pakistan	Not found
IS 19859	IS 19859	India	Durra
KALJANPUR	WSC 29	India	Not found
IS 4372	IS 4372	India	Guinea-durra
IS 32787	IS 32787	Somalia	Durra
IS 5094	IS 5094	India	Durra
IS 4581	IS 4581	India	Durra
IS 4631	IS 4631	India	Durra
IS 4515	IS 4515	India	Durra
IS 33844	IS 33844	India	Not found
IS 18039	IS 18039	India	Durra-bicolor
IS 17980	IS 17980	India	Durra
IS 2205	IS 2205	India	Not found
IS 2205	IS 2205	India	Not found
IS 18551	IS 18551	Ethiopia	Not found
IS 6354	IS 6354	India	Durra
IS 6351	IS 6351	India	Durra
IS 6421	IS 6421	India	Durra
IS 5667	IS 5667	India	Durra
SC 835	PI 534007	India	Not found
IS 5386	IS 5386	India	Durra
SC NO.0217 CI1197	WSC 30	India	Not found
IS 4613	IS 4613	India	Durra
IS 18729	IS 18729	United States of America	Not found
CHOONCHAN LOCAL	WSC 18	South Korea/North Korea	Not found

Supplementary Table 1. Sorghum germplasm examined in this study_continued

Name	Accession ID	Origin	Race
SC 56	PI 533910	Sudan	Not found
IS 20816	IS 20816	United States of America	Bicolor
IS 12447	IS 12447	Sudan	Durra-caudatum
DHOOTI ANEHULA	WSC 33	India	Not found
B.Tx2925	PI 629053	United States of America	Not found
B.Tx2921	PI 629045	United States of America	Not found
BTx623	PI 564163	United States of America	Not found
RTx432	PI 656000	United States of America	Not found
B.Tx2922	PI 629047	United States of America	Not found
IS 10475	IS 10475	United States of America	Not found
IS 24462	IS 24462	South Africa	Caudatum-bicolor
AS 4637 NHORONGO NENPI	WSC 106	United Republic of Tanzania	Not found
IS 10288	IS 10288	United States of America	Not found
AW 70/12 DL/59/1532	WSC 69/PI 267512	South Africa	Caudatum
IS 19262	IS 19262	Sudan	Guinea-caudatum
B2AR3043	PI 576320	United States of America	Not found
CHAL WAXY SORGHUM	WSC 14	South Korea/North Korea	Not found
B3 BTx3197	PI 598121	United States of America	Not found
PERSIAN DWARF	PI 576074	United Kingdom	Not found
E 959	WSC 101	Kenya	Not found
IS 18757	IS 18757	Australia	Not found
IS 3158	IS 3158	South Africa	Kafir
IS 13893	IS 13893	South Africa	Kafir-caudatum
IS 29335	IS 29335	Swaziland	Caudatum
IS 29314	IS 29314	Swaziland	Durra-caudatum
IS 27912	IS 27912	South Africa	Kafir-caudatum
IS 29241	IS 29241	Swaziland	Kafir-caudatum
IS 29187	IS 29187	Swaziland	Guinea-caudatum
MORABA 74	WSC 95/PI 563098	Ethiopia	Kafir
IS 13971	IS 13971	South Africa	Caudatum
IS 24492	IS 24492	South Africa	Kafir
IS 29606	IS 29606	South Africa	Kafir
IS 22294	IS 22294	Botswana	Kafir
IS 29733	IS 29733	Zimbabwe	Guinea-durra
IS 20092	IS 30092	Zimbabwe	Durra-caudatum
TSETA LOCAL NATURE TYPE 27/51	WSC 105	Zimbabwe	Not found
MN 401	WSC 90	Algeria	Not found
KOURNIANIA	WSC 64	Morocco	Not found
IS 29304	IS 29304	Swaziland	Guinea-kafir
IS 29233	IS 29233	Swaziland	Kafir
IS 13872	IS 13872	South Africa	Not found
IS 13870	IS 13870	South Africa	Not found
IS 13869	IS 13869	South Africa	Not found
IS 13868	IS 13868	South Africa	Not found
IS 13782	IS 13782	South Africa	Kafir-durra
IS 29269	IS 29269	Swaziland	Guinea-caudatum
IS 8525	IS 8525	Ethiopia	Not found
IS 29519	IS 29519	Lesotho	Kafir-caudatum
IS 19445	IS 19445	Botswana	Kafir
IS 3436	IS 3436	South Africa	Not found
CAPRICORN	PI 586438	Hungary	Kafir
IS 12965	IS 12965	Cuba	Caudatum
IS 12945	IS 12945	Nicaragua	Kafir
EAR FROM PIETESBURG DL/60/107	WSC 73	South Africa	Not found
THIMBA RED	PI 229877	South Africa	Kafir-durra
IS 19450	IS 19450	Botswana	Guinea-kafir
IS 29441	IS 29441	Lesotho	Kafir-caudatum
IS 29468	IS 29468	Lesotho	Guinea-caudatum
IS 29358	IS 29358	Lesotho	Kafir
NYAKASOBA BEST	WSC 61	Lesotho	Not found
IS 29565	IS 29565	Lesotho	Guinea-caudatum
IS 29392	IS 29392	Lesotho	Kafir
TENANT WHITE	WSC 60	Lesotho	Kafir-caudatum
IS 24463	IS 24463	South Africa	Kafir
IS 29568	IS 29568	Lesotho	Kafir-caudatum
IS 2389	IS 2389	South Africa	Kafir
RED KAFIR	WSC 71	South Africa	Not found
E 233 BARNARD RED	WSC 70	South Africa	Not found
IS 2397	IS 2397	South Africa	Kafir
IS 8774	IS 8774	South Africa	Kafir-durra
S. BASUTORUM DL/60/97	WSC 72	South Africa	Not found
IS 22616	IS 22616	Myanmar	Bicolor
ALLAKH	WSC 51	Bangladesh	Not found
COL/PAK/1989/IBPGR/2550(1)	WSC 42	Pakistan	Not found
ITALIAN	PI 295123	Australia	Kafir
COL/PAK/1989/IBPGR/2553(4)	WSC 43	Pakistan	Not found
IS 10646	IS 10646	United States of America	Not found
ESHOME	WSC 67	South Africa	Not found
IS 29627	IS 29627	South Africa	Durra-caudatum
IS 22239	IS 22239	Botswana	Kafir
IS 14290	IS 14290	Botswana	Kafir-durra
IS 29326	IS 29326	Swaziland	Caudatum-bicolor
IS 26701	IS 26701	South Africa	Caudatum-bicolor
RIO	PI 651496	-	Not found
IS 27887	IS 27887	South Africa	Caudatum-bicolor
IS 24453	IS 24453	South Africa	Caudatum-bicolor

Accessions from the United States Department of Agriculture (USDA) are identified by "PI" numbers, accessions from the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) as identified by "IS" numbers, accessions from the National Agriculture and Food Research Organization (NARO) sorghum collection as identified by "WSC" numbers in Accession ID colum.

Supplementary Table 2. Primers used in this study

No.	Primer name	Sequence (5′ -> 3′)	Comment
1	gDAI_genotyping_fw	GACCACATGGGCCTTCTTTA	For genotyping of DAI gene using BTx623, NOG, and selected germplasm genomic DNA
2	gDAI_genotyping_rv	TGGCCCTTTGATAAGTCGAG	
3	TOPO_DAI_fw	caccATGCAGGTCGGAGGAGGAGC	For constructing "Free GFP" and "DAI-GFP" plasmids
4	DAI-tag_rv	GTTGAACACCGACAGCGGCA	
5	Just_linker_fw	caccAAGGGTGGG	
6	Just_linker_rv	CCCACCCTT	
7	M13 fw	GTAAAACGACGGCCAG	
8	M13 rv	CAGGAAACAGCTATGAC	
9	Linearize_pBUH3_fw	TTGGGTACCAGCTCGAATTT	For constructing "DAIpro::DAI" plasmid for making sorghum transgenic plants
10	Linearize_pBUH3_rv	CTCTAGAGAGGGGCACGAC	
11	pBUH3_DAIpro_inF_fw	TGCCCCTCTCTAGAGGAATGAACAAGGACCGCAT	
12	pBUH3_DAI_inF_rv	CGAGCTGGTACCCAACCTACCTCGCCTGGCTAT	
13	pBUH3_inseq_fw	TCGTATGTTGTGTGGAATTG	
14	pBUH3_inseq_rv	AATGTATAATTGCGGGACTC	
15	DAI_C_fw	CAGTTCCTGCACGGGTTCAG	
16	LINE_rv	TCCATATGTCACACCAGGGG	
17	DAI_1st_fw	GATTCGTCTCGTCAATTCCG	
18	DAI_2nd_fw	CCTCAGAGGAGCCGTACTTG	
19	DAI_icheck_fw	GCACCAGCATCACTTCTTCA	For genotyping of the transgenic sorghum
20	DAI_pBUH3_icheck_rv	TTGCGGGACTCTAATCATAAAAA	
21	Cont_H2A_qfw	GCCAGGGACAACAAGAAGAC	For qPCR to determine the copy number of the transgene
22	Cont_H2A_qrv	GACACCGGACAGCAGCTT	
23	fDAI_qfw	ATTTCGGCTCCAACAATGTC	
24	fDAI_qrv	GATGTTGGCGACCTCGTATT	
25	DAIpro_1st_fw	TGATGGTATGCGGTGTTTGT	For constructing "SbDAI" and "gNOG" plasmids for making Kasalath transgenic plants
26	DAI3UTR_1st_rv	GTAGATGGCACGGGCTGTAT	
27	DAIpro_nest_inF_fw	ATGTATGATAATTCGGAATGAACAAGGACCGCATT	
28	DAItag_nest_inFrv	CTCCCGGCCGCCATGCTACCTCGCCTGGCTATACG	
29	gNOG_inF_rv	CTCCCGGCCGCCATGAAATCCACCGACGACAGAAG	
30	icheck_fw2	TGCAATAGGTCAGGCTCTCG	
31	icheck _rv	TATCTGGGAACTACTCACAC	
32	DAIseq_fw	TTCTTCTTCTTCCCACTTTC	
33	DAIseq_fw2	CTTCGTCGTCCCAGTCGT	
34	gNOG_icheck_fw	TGTTTCCCTGATGACACCTC	For genotyping of the transgenic Kasalath
35	DAI_icheck_rv	GCAGGTCACTGGATTTTGGT	
36	1stM_01_fw	CGCGATAAGGAAATTTGGAG	For genotyping to establish NILs for *qAL3*/*qAP3*
37	1stM_01_rv	GAACCTTGCCTTGCCTTACA	
38	1stM_02_fw	CTCACGCCACGAAAATATCA	
39	1stM_02_rv	GAGGGAGATGGAGGTGGTG	
40	1stM_03_fw	AGGAAAGGGAGGTGAGGTGT	
41	1stM_03_rv	AATAAAAGGCGCACACCAAG	
42	1stM_04_fw	TATACCACCCACACCCAGCA	
43	1stM_04_rv	CTACCCCTCACCACAGATCC	
44	1stM_05_fw	GGATTGCGGTAAGATCATCA	
45	1stM_05_rv	CCGTGATTGGACAATGTTTG	
46	1stM_06_fw	AACCAGACCACTGCTGGTG	
47	1stM_06_rv	ATTCTGTCGGAGGCATCATC	
48	1stM_07_fw	GGCTCCAATCAACTTGGACT	
49	1stM_07_rv	CGGCACCATGCAGTAAATAA	
50	1stM_08_fw	GGAGAGTCAATGTATCTCAGGTTTG	
51	1stM_08_rv	TTGGATGGGCTAATTCTCCA	
52	1stM_09_fw	ACCGACTCGTCGCATACTTC	
53	1stM_09_rv	CACAGTGCACACACCGTTT	
54	1stM_10_fw	TCTCCCTTCTCCCCTCTTCT	
55	1stM_10_rv	CGGAAAGGCTACCTCTCTCTC	
56	1stM_11_fw	GCCGCTATTACCCTAGCAGA	
57	1stM_11_rv	AGCAACAGCAAAGCGTAGGT	
58	1stM_07.5_fw	TTTAGGGACGGAGGGAGTAA	
59	1stM_07.5_rv	CTGATCCATCATCGCTTTCA	
60	1stM_08.2_fw	ACACGGTCATTAGGCGACAT	
61	1stM_08.2_rv	TGCGACCTTTTGGCTAGACT	
62	1stM_08.5_fw	AGCCATGAGCAAGCAAAAA	
63	1stM_08.5_rv	TCTGCGCACTCTTTCATCAC	
64	1stM_09.5_fw	AGGAACAGATGCGAGAGGAA	
65	1stM_09.5_rv	TTGTTGTGAACGATCCGACT	
66	1stM_12_fw	CCGGTAAGATCCCATTAGGG	
67	1stM_12_rv	CCAACCGTGGATTTAGGAGA	
68	1stM_13_fw	GCCTCGAGCAGACCTACG	
69	1stM_13_rv	TACGAGGACGAGGAGGGAGT	