Natural alleles of GLA for grain length and awn development were differently domesticated in rice subspecies japonica and indica

Yanpei Zhang1,*, Zhanying Zhang1,†, Xingming Sun1, Xiaoyang Zhu1, Ben Li1, Jinjie Li1, Haifeng Guo1, Chao Chen1, Yinghua Pan2, Yuntao Liang2, Zhijian Xu3, Hongliang Zhang1 and Zichao Li1,∗†

1Key Laboratory of Crop Heterosis and Utilization, Ministry of Education/Beijing Key Laboratory of Crop Genetic Improvement, Department of Plant Genetics and Breeding, China Agricultural University, Beijing, China
2China/Guangxi Key Laboratory of Rice Genetics and Breeding, Rice Research Institute, Guangxi Academy of Agricultural Sciences, Nanning, Guangxi, China

Received 13 September 2018; revised 11 December 2018; accepted 6 January 2019.
*Correspondence (Tel +86 010 62731414; fax +86 010 62731414; email: lizichao@cau.edu.cn)
†These authors contributed equally to this work.

Summary
Rice (Oryza sativa L.) cultivars harbour morphological and physiological traits different from those of wild rice (O. rufipogon Griff.), but the molecular mechanisms underlying domestication remain controversial. Here, we show that awn and long grain traits in the near-isogenic NIL-GLA are separately controlled by variations within the GLA (Grain Length and Awn Development) gene, a new allele of GAD1/RAE2, which encodes one member of the EFPL (epidermal patterning factor-like protein) family. Haplotype analyses and transgenic studies revealed that InDel1 (variation for grain length, VGL) in the promoter region of GLA (GLA105) increases grain length by promoting transcription of GLA. Absence of InDel3 (variation for awn formation, VA) in the coding region (CDS) of GLA (GLA109) results in short awn or no awn phenotypes. Analyses of minimum spanning trees and introgression regions demonstrated that An-1, an important gene for awn formation, was preferentially domesticated and its mutation to an-1 was followed by GLA and An-2. Gene flow then occurred between the evolved japonica and indica populations. Quality analysis showed that GLA causes poor grain quality. During genetic improvement, awnlessness was selected in ssp. indica, whereas short-grained and awnless phenotypes with good quality were selected in ssp. japonica. Our findings facilitate an understanding of rice domestication and provide a favourable allele for rice breeding.

Keywords: allelic frequency, awn development, domestication, grain length, natural variation.

Introduction
Rice (Oryza sativa L.), one of the earliest domesticated, primary food crops, feeds more than half of the world population (Khush, 1997). It originated from wild rice (O. rufipogon Griff.) through domestication (Fuller et al., 2010; Huang et al., 2012; Zong et al., 2007). Wild rice has unfavourable agronomic traits, such as prostrate growth, seed shattering, a spreading panicle, long awns and low number of grains per panicle, all of which changed dramatically under domestication. Compared with its wild ancestors cultivated rice typically exhibits reduced seed shattering and dormancy, increased grain number per panicle, erect growth, closed panicles and no or short awns (Kovach et al., 2007; Sweeney and McCouch, 2007). All of these features that differ between wild rice and cultivars are known as domestication traits.

To date, some domesticated genes are functionally characterized, such as seed shattering genes Sh4 (Li et al., 2006) and qSh1 (Konishi et al., 2006; Li et al., 2006), prog1 controlling straight growth (Jin et al., 2008; Tan et al., 2008), OsLG1 regulating closed panicles (Ishii et al., 2013; Zhu et al., 2013), grain size genes GWS5/GS5 (Duan et al., 2017; Liu et al., 2017), GS3 (Mao et al., 2010), OsLG3 (Yu et al., 2017) and OsLG3b (Yu et al., 2018), seed dormancy gene Sdr4 (Sugimoto et al., 2010), seed hull colour genes Bh4 (Zhu et al., 2011) and Rc (Sweeney et al., 2006), NOG1 increasing grain production (Huo et al., 2017) and awn development genes An-1 (Luo et al., 2013), An-2/LABA1 (Gu et al., 2015; Hua et al., 2015) and GAD1/RAE2 (Bessho-Uehara et al., 2016; Jin et al., 2016).

Awn, one of the most important domestication traits in cereal crops, is common in graminaceous crops such as sorghum (Sorghum bicolor L.), wheat (Triticum aestivum L.), oat (Avena sativa L.) and barley (Hordeum vulgare L.). Long awns, extensions at the tip of the lemmas, help seed dispersal and prevent birds and mammals from preying on the grains (Elbaum et al., 2007; Hu et al., 2011; Kulic et al., 2009). In some cereal crops, such as wheat and barley, awns also have photosynthetic capability that contributes to grain yield (Abebe et al., 2010). However, rice awns have no chloroplasts and are inconvenient for harvesting and post-harvest processing (Takahashi et al., 1986). Hence, domesticated rice usually possesses no or short awns (Toriba et al., 2010).

The awn is a complex character controlled by multiple genes. Several awn-related genes in rice have been cloned. An-1, encoding a bHLH transcription factor, positively controls the formation of awn primordia, cell division and grain length in wild rice. The an-1 allele conferring shorter awns, shorter grains and higher number of grains per panicle was selected during
Results

Phenotypic characterization of NIL-GLA and NIL-gla

The backcross populations segregated at a single locus for the awn trait (Table S1). Long awns and long grains were completely associated in the BC3F4 population (Figure S1). The NIL-GLA line with long awns and grains and NIL-gla line with no awn and short grains were selected from the BC3F4 population (Figure S2). SEM of spikelet development at the Sp8I growth stage (Itoh et al., 2016) showed the awn primordia of NIL-GLA spikelets were ruptured under the action of lyase to form a mature polypeptide. RAE2/GLA regulates awn development, as well as grain number per panicle and grain length (Bessho-Uehara et al., 2016). The above reports demonstrated that genes controlling awn phenotype have pleiotropic effects on yield-related traits. Long awn and grain length are often closely associated, suggesting that alleles conferring longer grain might be neglected because of the selective sweep of awn traits. Currently, the relationship between these traits is not clear and the genetic basis and molecular mechanisms conferring the selective forces involved in awn formation and grain length during domestication need to be further examined.

Here, we isolated a multifunctional gene GLA (Grain Length and Awn Development), which is an allele of GAD1/RAE2. Using GLA-gene association analysis, haplotype analysis and transgenic studies, we analysed the functional variations of GLA for grain length and awn formation. Evolutionary studies explained the relationship among the An-1, An-2 and GLA genes during the domestication process. In addition, quality evaluation and allele utilization were used to analyse the selection histories of japonica and indica rice in the process of genetic improvement. These studies not only improve our understanding of crop domestication, but also provide a valuable genetic resource for future molecular breeding of rice.

Grain length and awn development are regulated by different functional variations of GLA

GLA was shown to affect both grain length and awn traits and the functional loci for each trait were still not clear. GLA-based association analyses were performed on the basis of 33 SNPs and 13 InDels in 358 cultivated accessions (Data S2). The results indicated that grain length and awn development were significantly associated with different variations. InDel1 and InDel3 made significant contributions to grain length, whereas SNP1 and InDel13 were significantly associated with awn formation (Figure 3a).

According to the significant variation associated with grain length, the sequences of 371 accessions including 358 cultivated and 13 wild rice varieties (Data S2) were divided into five haplotypes named GL-hap1 to GL-hap5. The wild rice accessions were mainly GL-hap1 and GL-hap2 types, indica accessions were GL-hap2, GL-hap3 and GL-hap4 and japonica accessions were GL-hap1, GL-hap2 and GL-hap4 (Figure 3b, Data S2). There was no significant difference in grain length between GL-hap1 and GL-hap2 in O. rufipogon and japonica or between GL-hap2 and GL-hap3 in indica. However, compared to GL-hap1, GL-hap2 or
GL-hap3, grain length of GL-hap4 accessions was significantly reduced in both japonica and indica (Figure 3c, Data S2). These results suggested that grain length was influenced by the InDel1 (variation for grain length, VGL) in GLA (GLAVGL).

Transient assays were performed in tobacco leaves to determine whether the promoter activity was influenced by GLAVGL, which is located in the promoter. Relative promoter activities were significantly increased in ProGLA transformants containing the functional InDel1 (Figure 3d) suggesting that InDel1 increased the promoter activity of GLA. Moreover, transcriptional levels of GLA were positively correlated with grain length in 26 rice varieties randomly selected from the 358 cultivated varieties (Figure 3e, Data S3). These results indicated that GLAVGL regulated grain length by affecting the expression of GLA.

To determine the functional variation regulating awn formation on the basis of significant SNP/InDels, we classified the GLA alleles were differently domesticated in rice
sequences of rice accessions into haplotypes An-hap1, An-hap2, An-hap3 and An-hap4. Comparing the numbers of cysteines (C) in GLA (Bessho-Uehara et al., 2016) it became evident that An-hap1 and An-hap2 were 6C type, whereas An-hap3 and An-hap4 were 7C and 4C types respectively. The rare SNP1 variant in An-hap1 occurred only in japonica. The InDel3 locus had three haplotypes. Compared with varieties of An-hap2 (five awned varieties in *O. rufipogon* and nine awned and six awnless varieties in *japonica*), most varieties with An-hap3 (54 of 63 varieties in *japonica*, 214 of 219 varieties in *indica*) and An-hap4 (33 of 37 varieties in *japonica*, 22 varieties in *indica*) were awnless (Figure 3f, Data S2). This provided further evidence that InDel3 (variation for awn, VA) in GLA (GLA<sup>V</sup>) confirmed the functional variation in awn development, a result that is consistent with the previous results for GAD1 (Bessho-Uehara et al., 2016; Jin et al., 2016).

To further confirm the functions of VGL and VA, ProGLA<sup>VGL</sup>−GLA<sup>V</sup>, ProGLA<sup>GVA</sup>−GLA<sup>V</sup>, and ProGLA<sup>GVP</sup>−GLA<sup>V</sup> complementary constructs were introduced into Nip with GLA<sup>GLA</sup> <sup>V</sup> (wildtype), the transgenic lines were named as Nip<sup>+</sup>, Nip<sup>–</sup> and Nip<sup>0</sup> respectively (Figure 4a). Compared to Nip, lines Nip<sup>−</sup>, Nip<sup>0</sup> and Nip<sup>–</sup> showed longer grains, whereas the grain lengths of lines Nip<sup>−</sup> and Nip<sup>0</sup> were unchanged (Figure 4b,d). There was no significant change in grain width between transgenic lines (Figure S7c,f). Moreover, the phenotypes for awn length and awn proportion in Nip<sup>−</sup>, Nip<sup>0</sup> and Nip<sup>–</sup> lines were restored (Figures 4c,e and S7a). This further proved that InDel1 and InDel3 in GLA were the functional sites for variation in grain length and the awn trait respectively. To further identify the function of GLA, a CRISPR-Cas9 vector (CR) was constructed and transformed into NIL-GLA, producing a CR9 transgenic line that had no awn and no effect on grain length and width (Figures 4b,c,h–j and S7b,d,e,g). Compared to Nip, lines Nip<sup>+</sup>, Nip<sup>−</sup>, Nip<sup>0</sup> and Nip<sup>–</sup> had lower grain number per panicle (Figure 4g). Compared to NIL-GLA, the grain number per panicle of line CR9 was significantly increased (Figure 4k), thus suggesting that GLA<sup>V</sup> increased grain number per panicle. All of these results collectively verified that the functional variation of GLA in regulating grain length was different from that affecting awn development and that natural alleles of GLA in the VGL and VA loci might be responsible.

Allele GLA<sup>V</sup> was selected before GLA<sup>gvl</sup>

We categorized 371 rice accessions (358 cultivated and 13 wild rice) into four groups according to non-functional (−) or functional (+) alleles of GLA<sup>V</sup>GLA<sup>GVA</sup> and GLA<sup>V</sup>VGLA<sup>GVP</sup>. There was no accession in group III. Most accessions in group I (five *O. rufipogon*, 11 of 17 in *japonica*) with GLA<sup>V</sup>VGLA<sup>GVP</sup> had awned phenotypes and most accessions in groups II (53 of 63 in *japonica*, 225 of 230 in *indica*) and IV (11 in *indica*, 34 of 38 in *japonica*) with GLA<sup>V</sup>VGLA<sup>GVP</sup> were awnless (Figure 5a, Data S2). Grain lengths of wild rice accessions with GLA<sup>V</sup>VGLA<sup>GVP</sup> showed no significant difference between groups I and II. However, compared to grain lengths of group IV in *japonica* and *indica*, respectively, varieties with GLA<sup>V</sup>VGLA<sup>GVP</sup> in group I in *japonica*, group II in *japonica* and group II in *indica* had longer grains (Figure 5b, Data S2). These results indicated that the grain length and awn phenotype of various rice varieties were consistent across different haplotype combinations.

Moreover, most cultivated rice accessions belonged to group II with GLA<sup>V</sup>VGLA<sup>GVP</sup>

GLA is considered a domestication gene (Bessho-Uehara et al., 2016; Jin et al., 2016) that controls grain length and awn development. The question then was what genotypes of GLA<sup>V</sup>VGLA<sup>GVP</sup>...
were selected artificially. A nucleic acid diversity analysis was conducted and a neutrality test was carried out based on the variant sites (documented as ++). Tajima’s D test ($P < 0.01$) revealed that GLA in cultivated and wild rice deviated significantly from neutrality (Figure 5c) indicating that GLA in wild rice was subjected to balanced natural selection and that gla in cultivated rice had undergone directional artificial selection. Thus selection in wild and cultivated rice led to different phenotypes. Natural selection in wild rice favoured awned phenotypes, whereas artificial selection favoured awnlessness (GLAva) in indica and short grains plus awnlessness (GLAva) in japonica (Figure 5c).

In order to explore evolutionary relationships between GLAVGL and GLAva, 358 cultivated and 88 wild rice accessions were used to produce a minimum spanning tree. Wild rice had both GLAVGL and GLAva (documented as ++) and GLAVGLva (-++) genotypes. However, temperate japonica rice that differentiated from wild rice in China did not acquire the GLAva allele. Temperate japonica with GLAVGLva first produced the natural variant GLAva. The long grain, awnless genotype GLAVGLva was obtained by artificial selection. Subsequently, the GLAVGLva allele changed to GLAva (- -) in cultivated rice. These results indicated that the GLAva allele was selected before GLAva. Moreover, indica rice first appeared in a long-grain, awnless combination and japonica and indica had different selection histories, leading to short-grain, awnless and long-grain, awnless genotypes respectively (Figure 5d). Thus, we hypothesize that GLAva domestication occurred before divergence of the indica subspecies.

**An-1 was preferentially domesticated before GLA and An-2**

The currently cloned genes controlling awn traits are An-1, An-2 and GLA/GAD1/RAE2. To determine the evolutionary relationships among these genes during domestication a minimum spanning tree was constructed using 145 cultivated and 69 wild rice accessions. The results showed that a change in An-1 was the first step in domestication. This produced temperate japonica with an-1/An-2/GLAVGL/GLAva (symbolized --++) genotype. Then, temperate japonica with an-1 was domesticated at the functional position GLAva leading to GLAva (---). By this time, indica rice had evolved. Part of the cultivated population with an-1/An-2/GLAVGL/GLAva (---) genotype. Other cultivated rice (---) lines were domesticated by change in An-2, and an-1/An-2/GLAVGL/GLAva (---) appeared (Figure 6a). Thus An-1 was domesticated first, and was subsequently followed by changes in GLA and An-2. Variation of GLAva occurred after the GLAva domestication step.

The minimum spanning tree of GLA or three domestication genes controlling awn development clustered indica and japonica.
Figure 4  Confirmation of GLA in controlling grain length and awn development. (a) Schematic of transgenesis. VGL, variation for grain length. VA, variation for awn development. (b,c) Grain morphology of Nip, NIL-GLA, six complementation transgenic lines in Nipponbare (Nip) background, and a CRISPR/Cas9 edited line in NIL-GLA background. NipI-1 and NipI-2 are independent transgenic lines containing the ProGLAVGL::GLAVA vector. NipII-1 and NipII-2 are independent transgenic lines carrying the ProGLAVGL::GLAva construct. NipIII-1 and NipIII-2 are independent transgenic lines carrying the ProGLAvgl::GLAVA construct. CR9 is a CRISPR/Cas9 edited line. ‘+’ and ‘−’ represent functional and non-functional genotypes. Bar, 5 mm. (d–k) Comparison of grain lengths (d,h), awn lengths (e,i), awn proportions (f,j) and grain numbers per panicle among transgenic lines. Data are represented as means ± SE (n = 15), *P < 0.05, **P < 0.01 based on Student’s t-tests.
together, similar to previous studies on gene flow (Choi et al., 2017; Huang et al., 2012). To determine whether these genes were associated with gene flow of domestication alleles we performed analyses of introgressed regions. The an-1 and gla domestication alleles were transferred from japonica to the indica population, whereas the an-2 allele was transferred from indica to japonica (Figure 6b–d). These results demonstrated significant gene flow between the japonica and indica populations.

We then evaluated the contributions of An-1, An-2 and GLA to awn development. The genotypes and phenotypes of 232 cultivated and 44 wild rice accessions revealed that after domestication of An-1, the frequency of accessions with awns decreased from 100% to 60%, and awn proportion fell to 14.04% and 3.64% after GLA and An-2 domestication respectively (Figure 6e). We concluded that An-1 made the largest contribution to awn phenotype, followed by GLA and An-2 respectively.

The selection to GLAvgl is beneficial to rice quality improvement

Grain length is closely related to grain quality (Zhao et al., 2018). We evaluated the grain quality of NILs and found that rice milled from NIL-GLA had high percentages of chalky grain and chalky area with loose and irregular starch granules, unlike that from...
NIL-gla, which displayed compactly arranged and largely sharp-edged polygonal starch granules (Figure 7a–d). These results indicated that GLA caused poor quality. The grain qualities of transgenic plants were also investigated. Compared to Nip, NipI-1, NipI-2, NipII-1 and NipII-2, transgenic plants had a higher proportion of chalky grain and small, messy starch granules, whereas Nip, the NipIII-1 and NipIII-2 lines had similar phenotypes (Figure 7a,b,e,f). Thus GLAVGL caused poor quality whereas GLAVA had no influence on rice grain quality.

Given that GLAVGL and GLAVA affected grain length, grain quality and awn development the question was how to make use of GLA in genetic improvement. We surveyed the functional allelic frequencies of GLAVGL/VA in the panel of 358 varieties. Compared with landrace groups (LAN), the non-functional allele frequencies of GLAvgl and GLAva in improved accessions (IMP) were significantly increased in japonica, unlike indica in which the functional allele GLAVA did not exist and the functional allele frequency of GLAVGL was increased (Figure 7g). These results implied that indica and japonica had different selection histories during genetic improvement. The GLAVGL allele for long grain phenotype was enriched in indica with GLAva, whereas the GLAVgl/va allele for good quality, awnlessness and short grain was utilized in japonica. Additionally, a functional molecular marker (J2) for the VGL locus was developed (Data S1). Genotyping analyses using the J2 marker in 18 cultivars further confirmed that GLAVGL was most likely responsible for long-grain phenotype (Figure 7h,i, Data S6). The J2 marker should be applied in future molecular marker assisted breeding.

Figure 7  Improved grain quality produced by null GLAVGL. (a) Comparison of milled rice in NILs, Nip and complementation lines. Bar, 1 cm. (b) SEM images of starch granules in transverse sections of grains. Bar, 10 μm. (c–f) Percentages of chalky grain (c,e) and chalky area ratio (d,f). Data are represented as means ± SE. (n = 15), *P < 0.05, **P < 0.01 based on Student’s t-tests. (g) Allele frequency analysis of GLAVGL and GLAVA. Blue and red colors represent functional (GLAVGL) and non-functional (GLAvgl) alleles of the VGL locus. Skyblue and orange colors represent functional (GLAVA) and non-functional (GLAva) alleles of the VA locus. Pie charts above and below the horizontal line represent landraces (LAN) and improved varieties (IMP). (h) Genotyping analyses of GLAVGL in 18 diverse accessions of cultivated rice. We used a pair of primers (Table S2) to amplify gene segments containing the variable InDel1 site. The length of the PCR product amplified from accessions with GLAVGL (+) was 246 bp and that from accessions with GLAVgl (−) was 200 bp. (i) Grains of the 18 cultivated rice accessions. Bar, 5 mm. *P < 0.05, **P < 0.01, ***P < 0.001 based on Student’s t-tests.
Discussion

GLA has pleiotropic effects on grain length, awn development, grain quality and grain number

Wild rice exhibits long, barbed awns that prevent granivore predation, but is detrimental for harvesting and post-harvest processing. Hence awnless varieties were produced during domestication of wild rice (Kovach et al., 2007; Sweeney and McCouch, 2007). Several genes associated with awn development have been identified. However, various genes for yield-related traits, such as grain length, grain number and yield per plant were often closely associated with awn length and hence were neglected because of the selective sweep of awn traits. Thus, the molecular mechanisms of these genes needed to be investigated individually.

In this study, we identified Os08 g37890 as GLA, a new allele of RAE2/GAD1. GLA exhibited pleiotropic effects on grain length, grain quality, grain number and awn development, and had different functional sites. Our study identified a new functional site, GLA\(^{VGL}\), that regulated grain length and grain quality and further verified the GLA\(^{VGL}\) functional site controlling awn development. In addition, GLA\(^{VGL}\) affected grain length at the transcriptional level, consistent with RNAi results in GAD1 studies (Jin et al., 2016). During genetic improvement, japonica rice underwent improved quality, shorter grain and awnlessness, whereas indica became long grained and awnless. A functional molecular marker developed for GLA\(^{VGL}\) can be used in breeding programs. On the basis of the long-grain, awnless phenotype in indica and short-grain, awnless traits in japonica, grain quality and grain length can be further improved.

Origin and evolution of awn-associated domestication loci

An-1 influences the formation of awn primordia by regulating cell division (Luo et al., 2013). An-2 induces awn elongation by increasing the cytokinin concentration in awn primordial (Gu et al., 2015; Hua et al., 2015). GAD1/RAE2, an EPFL family member, regulates awn length (Bessho-Uehara et al., 2016; Jin et al., 2016). Evidence from this study indicated that An-1 was the first of these genes to undergo domestication, followed by GLA and then An-2. This also suggested that An-1 is the most important gene controlling awn development. Minimal spanning tree and introgression analyses implied that An-1 and GLA might have been domesticated preferentially in japonica before the divergence of indica, but subsequently, the An-2 allele was transferred from indica to japonica. Hence, some questions are worth pondering. Firstly, given that An-2 makes the least contribution to awn formation why did introgression of An-2 actually occur? It may be that An-2 was linked to other genes and that transfer was due to linkage drag. The next issue is whether awn-related domestication genes arose only once and were dispersed by introgressive hybridization or whether they arose more than once. Several domestication-related loci, such as OsLG3b, Ghd7, LG1 and LAB1, show introgression signals between the subspecies, but others (Sh4, qSh3, qSh1, Prog1, and Rc) lack such signals (Civan and Brown, 2018; Yu et al., 2018).

The origin of cultivated rice is a continuing, controversial issue with variable lines of evidence for single and multiple origins. Multiple origins are currently supported by most academics (Choi et al., 2017; Londo et al., 2006; Sun et al., 2018). Therefore, by combing previous studies and our results, we speculate that proto-japonica from wild rice in Southern China went through the An-1 and GLA domestication steps prior to japonica/indica differentiation and produced domesticated proto-japonica varieties with the an-1/gla genotype. After emergence of proto-indica varieties from O. rufipogon in Southeast and South Asia they underwent domestication of An-2. Later, populations of proto-japonica with an-1/gla and proto-indica with an-2 merged with each other and hybridization led to the current japonica and indica forms with genotype an-1/gla/an-2 (Figure S8). Nevertheless, the detailed evolutionary relationships based on awn traits need to be further analysed. Nonetheless, discovery of the new GLA allele will be of value for understanding the domestication and genetic improvement of rice.

Experimental procedures

Plant materials

Mapping populations were constructed from a cross between japonica cv. SYL (awned and long grains) as donor parent and japonica cv. Nipponbare (Nip, awnless and short grains) as recurrent parent. Recessive awnless and dominant awned individuals in the BC\(_2\)F\(_2\) generation with short and long grains respectively, were chosen as near isogenic lines (NILs), named NIL-gla and NIL-GLA. A segregating BC\(_2\)F\(_2\) population containing 203 individuals was used for preliminary mapping of GLA. A BC\(_2\)F\(_2\) population with 12,204 individuals was used for fine mapping. Plant materials were grown under natural paddy conditions at Beijing, or at Sanya in Hainan province.

Fine mapping and candidate gene analysis

Four hundred and seventy-seven SSR markers distributed across all 12 rice chromosomes were screened in a bulked segregant analysis (BSA) to identify the linked markers associated with the awn trait (Zhang et al., 1994) and were also used for preliminary mapping. Sequence-tagged site (STS) markers were then developed for fine mapping using DNASTAR software.

Candidate gene annotations were made on the basis of the Rice Annotation Project Database (http://rice.plantbiology.msu.edu/index.shtml). Genomic DNA fragments of four predicted genes were amplified from the DNA of NIL-gla and NIL-GLA and sequenced, respectively. Sequences were analysed using the SeqMan program in DNASTAR software. Primers are listed in Data S1.

cDNA and quantitative RT-PCR

Total RNA was extracted from 5 cm panicles of NIL-GLA and NIL-gla using Trizol reagent (Invitrogen, Carlsbad, USA). To eliminate contamination by genomic DNA, 50 \(\mu\)g of RNA was digested with Recombinant DNase I (RNase-free) (Takara, Japan) as described by the manufacturer. Two microgram of DNasel-treated RNA was reversely transcribed using M-MLV Reverse Transcriptase (Takara) with an oligo (dT18) primer. The reaction product was diluted three times as template. Quantitative RT-PCR (qRT-PCR) was carried out as previously reported (Sun et al., 2018). OsActin1 was used for internal reference (Data S1).

Vector construction and rice transformation

To make the genomic DNA complementation construct (Pro-GLA\(^{VGL}\)::GLA\(^{VGL}\)) containing functional InDel1 and InDel3 a 4717 bp genomic DNA fragment of GLA harbouring a 3292 bp region upstream of the initiation codon and a 919 bp region downstream of the start codon was amplified from NIL-GLA using...
GLA alleles were differently domesticated in rice

primers LA and cloned into the Pmel and SacI sites of the binary plant expression vector pMDC163 (Curts and Grossniklaus, 2003). Similarly, ProGLAVGLα::GLAβ and ProGLAVG::GLAβ/Vα vectors were constructed using primers LAαa and LAαα respectively. The ProGLAVGLα::GLAβ vector contained a 3292 bp corresponding region upstream of the initiation codon of GLA from NIL-GLA line and a 1421 bp region downstream of the initiation codon of GLA from NIL-gla line. The ProGLAVG::GLAβ/Vα vector consisted of a 3241 bp region upstream of the initiation codon of GLA from NIL-gla and a 1425 bp region downstream of the start codon of GLA from NIL-GLA. A CRISPR-Cas9 construct (CR) targeting the 74th–93th nt (5′-GCTGCTACAGCAAAGTCTAC-3′) of the second exon of GLA was made as previously described (Feng et al., 2013; Mao et al., 2013; Zhang et al., 2014). To generate transgenic plants, the constructs were introduced into Agrobacterium tumefaciens strain EHA105 and subsequently transformed into Nipponbare (Nip) or NIL-GLA by Agrobacterium-mediated transformation (Hiei et al., 1994).

Phenotypic evaluation

Three main panicles of each plant were collected for analysis of awn length (>1 mm), awn proportion and grains per panicle. The awn length of the panicle was represented by the average of the apical spikelet on each primary branch. Awn proportion was estimated as the number of awned spikelets per panicle. One hundred fully mature grains were randomly chosen for measurement of grain length and width. The grains were placed on a scanner for scanning. Grain length and width were calculated by analysing the scanned images (Luo et al., 2013). Grain weight per plant was represented by the dry weight of all grains from one plant. At least 15 individuals were used for phenotyping of NIL-GLA, NIL-gla and Nip transgenic plants.

Scanning electron microscopy

Young spikelets from NIL-GLA and NIL-gla were fixed in 2.5% glutaraldehyde-phosphate buffer saline fixative solution and dehydrated through an ethanol series (30, 50, 70, 80, 90, 95 and 100%). After dehydration the samples were dried with a carbon dioxide critical-point dryer. Mature grains were cleaned with 1% Tween 20 and dried in a 45 °C oven. Dried grains and spikelets were gold plated and observed using a Hitachi S-2460 scanning electron microscope at 15 kV. Cell number, cell length and cell width of lemma were quantified from scanning electron microscopy images. Quantification was represented by the average of 15 grains.

Transient expression assays of promoter activity

To make constructs ProGLA and Progla, GLA promoters with and without InDel1 were amplified using primers TEGLA and TEGla respectively (Data S1) and cloned into the Pmel and SacI sites of the pMDC162 vector. ProGLA, Progla and Pro35SS::LUC plasmid used as an internal control were co-transfected into tobacco leaves. The ratios of GUS/LUC (luciferase) were used as relative promoter activities (Sun et al., 2016; Yang et al., 2010), to obtain ProSuper::GLA-GFP and ProSuper::gla-GFP constructs respectively. The resulting vectors and CD3-1007, a plasma membrane marker (Nelson et al., 2007) were co-transfected into rice protoplasts. After culturing for 16 h at 28 °C fluorescence signals were observed as described (Zhang et al., 2011).

GLA-based association analysis and haplotype analysis

A worldwide set of 358 cultivated rice accessions selected from the 3K-Rice Project (Rice Functional Genomics and Breeding database, RFG8) (Zheng et al., 2015) and 13 Asian wild rice accessions provided by the National Germplasm Nanning Wild Rice Nursery were used for the study (Data S2). Genomic DNA fragments of GLA from different accessions were amplified using the V1, V2, V3 and V4 primers (Data S1 and S2), sequenced and analysed by DNASTAR software. Variable sites, including 33 SNPs (single nucleotide polymorphisms) and 13 InDels (insertion/deletion polymorphisms), were used for haplotype analyses (Data S2). The InDel1 and InDel3 alleles of GLA in 371 rice accessions (Data S2) were analysed from the sequencing results and then subjected to genotyping. Group analyses were carried out as previously reported (Sun et al., 2018).

Allelic frequencies and neutrality tests

The 358 accessions (143 landraces and 215 improved accessions; Data S2 and S5) were analysed for the An-1, An-2 and GLA functional loci controlling grain length or awn development and genotyped as non-functional or functional alleles (Sun et al., 2018). Eighty eight wild rice accessions provided by Dr Song Ge and 358 cultivars (241 indica and 117 japonica) varieties were used to evaluate nucleic acid diversity (π) and neutrality (Tajima’s D values) of GLA (Data S4). InDels and SNPs in GLA were identified. The data were processed by DnaSP 5.10 software (Librado and Rozas, 2009).

Analyses of minimum spanning trees

Rich sequence diversity for An-1, An-2 and GLA was present in both wild and cultivated rice (Data S5) and some haplotypes were represented by only one accession. We chose 9, 3 and 15 variable sites in An-1 An-2 and GLA respectively, and integrated haplotypes containing single varieties in developing the minimum spanning tree (Sun et al., 2018). The 358 cultivated and 88 wild rice accessions were used for analysis of the GLA region; and 145 cultivated varieties and 69 wild rice accessions were used to determine evolutionary relationships among the three genes. A minimum spanning tree was generated following a previous procedure (Liu et al., 2015; Yu et al., 2018).

Acknowledgements

We thank Professor Robert A McIntosh, University of Sydney, for critical reading of the manuscript. We thank the National Germplasm Nanning Wild Rice Nursery for providing 13 wild rice accessions, Song Ge for providing 88 wild rice accessions, and Dr Jiankang Zhu for providing the sgRNA-Cas9 vectors of the CRISPR-Cas9 system. The research was supported by grants from the Ministry of Science and Technology in China (2016YFD0100101-09, 2015BAD028B01), the National Natural Science Foundation of China (31271689).
Competing interests
The authors declare that they have no competing interests.

Author contribution
ZL, ZZ, HZ, JL and XS conceived the experiment; BL and CC produced the genetic populations; YZ and BL performed experiments; YZ, XZ, YP and XS conducted the phenotyping; YZ, XZ, XS and HG analysed the data; YZ, XZ, HG, ZZ and ZL wrote and revised the manuscript.

References
Abebe, T., Melmaiea, K., Berg, V. and Wise, R.P. (2010) Drought response in the spikes of barley: gene expression in the lemma, palea, awn, and seed. Funct. Integr. Genomics, 10, 191–205.
Bessho-Uehara, K., Wang, D.R., Furuta, T., Mimami, A., Nagai, K., Gamuyao, R., Asano, K. et al. (2016) Loss of function at RAE2, a previously unidentified EFL1, is required for awnlessness in cultivated Asian rice. Proc. Natl Acad. Sci. USA, 113, 869–874.
Choi, J.Y., Platts, A.E., Fuller, D.Q., Hsing, Y.I., Wang, R.A. and Purugganan, M.D. (2017) The rice paradox: multiple origins but single domestication in Asian rice. Mol. Biol. Evol. 34, 969–979.
Civan, P. and Brown, T.A. (2018) Role of genetic introgression during the entangled history of rice. Nat. Plants, 4, 154–162.
Curtis, M.D. and Grossniklaus, U. (2003) A gateway cloning vector set for high-throughput functional analysis of genes in planta. Plant Physiol. 133, 462–469.
Duan, P., Xu, J., Zeng, D., Zhang, B., Geng, M., Zhang, G., Huang, K. et al. (2017) Natural variation in the promoter of GES5 contributes to grain size diversity in rice. Mol. Plant, 10, 685–694.
Elbaum, R., Zaltzman, L., Burgett, I. and Fratzl, P. (2007) The role of wheat awns in the seed dispersal unit. Science, 316, 884–886.
Feng, Z., Zhang, B., Ding, W., Liu, X., Yang, D.L., Wei, P., Cao, F. et al. (2013) Efficient genome editing in plants using a CRISPR/Cas system. Cell Res. 23, 1229–1232.
Fuller, D.Q., Sato, Y.I., Castillo, C., Qin, L., Weisskopf, A.R., Kingwell, B., Gu, B., Zhou, T., Luo, J., Liu, H., Wang, Y., Shangguan, Y., Zhu, J., Jin, J., Hua, L. and Zhu, Z. (2016) GAD1 encodes a secreted peptide that regulates grain number, grain length, and awn development in rice domestication. Plant Cell, 28, 2453–2463.
Khush, G.S. (1997) Origin, dispersal, cultivation and variation of rice. Plant Mol. Biol. 35, 25–34.
Konishi, S., Izawa, T., Lin, S., Ebana, K., Fukuta, Y., Sasaki, T. and Yano, M. (2006) An SNP caused loss of seed shattering during rice domestication. Science, 312, 1392–1396.
Kovach, M.J., Sweeney, M.T. and McCouch, S.R. (2007) New insights into the history of rice domestication. Trends Genet. 23, 578–587.
Kulic, I.M., Mari, M., Mohrbach, H., Thakor, R. and Mahadevan, L. (2009) Botanical ratchets. Proc. Biol. Sci. 276, 2243–2247.
Li, C., Zhou, A. and Sang, T. (2006) Rice domestication by reducing shattering. Science, 311, 1936–1939.
Librado, P. and Rozas, J. (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics, 25, 1451–1452.
Liu, H., Liu, H., Zhang, Z., Zhang, X., Wang, M., Li, H. et al. (2015) Parallel domestication of the heading date 1 gene in cereals. Mol. Biol. Evol. 32, 2726–2737.
Luo, X., Huang, M., Fan, B., Buckler, E.S. and Zhang, Z. (2016) Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. Proc Natl Acad. Sci. 12, e1005676.
Luo, J., Chen, J., Zheng, X., Wu, F., Lin, Q., Heng, Y., Tian, P. et al. (2017) GWS acts in the brassinosteroid signalling pathway to regulate grain width and weight in rice. Nat. Plants, 3, 17043.
Londo, J.P., Chiang, Y.C., Hung, K.H., Chiang, T.Y. and Schaal, B.A. (2006) Phylogeography of Asian wild rice, Oryza rufipogon, reveals multiple independent domestications of cultivated rice, Oryza sativa. Proc. Natl Acad. Sci. USA, 103, 9578–9583.
Luo, J., Liu, H., Zhou, T., Gu, B., Huang, X., Shangguan, Y., Zhu, J. 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.
Mao, H., Sun, S., Yao, J., Wang, C., Yu, S., Xu, C., Li, X. et al. (2010) Linking differential domain functions of the GSK3 protein to natural variation of grain size in rice. Proc. Natl Acad. Sci. USA, 107, 19579–19584.
Mao, Y., Zhang, H., Xu, N., Zhang, B., Gou, F. and Zhu, J. (2013) Application of the CRISPR-Cas system for efficient genome engineering in plants. Mol. Plant, 6, 2008–2011.
Nelson, B.K., Cai, X. and Nebenfuhr, A. (2007) A multicolored set of in vivo organelle markers for co-localization studies in Arabidopsis and other plants. Plant J. 51, 1126–1136.
Ni, M., Cui, D., Einstein, J., Narasimhulu, S., Vergara, C.E. and Gelnov, S.B. (1995) Strength and tissue specificity of chimeric promoters derived from the octopine and mannopine synthase genes. Plant J. 7, 661–676.
Sugimoto, K., Takeuchi, Y., Ebana, K., Miyao, A., Hirochika, H., Hara, N., Ishiyama, K. et al. (2010) Molecular cloning of Sd4, a regulator involved in seed dormancy and domestication of rice. Proc. Natl Acad. Sci. USA, 107, 5792–5797.
Sun, X., Zhang, Z., Chen, C., Wu, W., Ren, N., Jiang, C., Yu, J. et al. (2018) The C5-A gene system regulates hull pigmentation and reveals evolution of anthercyanin biosynthesis pathway in rice. J. Exp. Bot. 69, 1485–1498.
Sweeney, M. and McCouch, S. (2007) The complex history of the domestication of rice. Ann. Bot. 100, 951–957.
Sweeney, M.T., Thomson, M.J., Pfeil, B.E. and McCouch, S. (2006) Caused red-headed: Rc encodes a basic helix-loop-helix protein conditioning red pericarp in rice. Plant Cell, 18, 283–294.
Takahashi, N., Alterfa, H.A.H. and Sato, T. (1986) Significant role of awn in rice plants (1): a survey of agricultural value of rice awn. FEBS Microbiol. Lett. 301, 232–240.
Tan, L., Li, X., Liu, F., Sun, X., Li, C., Zhu, Z., Fu, Y. et al. (2008) Control of a key gene associated with long, barbed awns in wild rice. Proc. Natl Acad. Sci. USA, 105, 1360–1364.
Takaiwa, T., Suzaki, T., Yamaguchi, T., Ohmori, Y., Tsukaya, H. and Hirano, H.Y. (2010) Distinct regulation of adaxial-abaxial polarity in anther patterning in rice. Plant Cell, 22, 1452–1462.
Wang, X., Wang, H., Liu, S., Ferjani, A., Li, J., Yan, J., Yang, X. et al. (2016) Genetic variation in ZmVPP1 contributes to drought tolerance in maize seedlings. Nat. Genet. 48, 1233–1241.
Xiong, H., Yu, J., Miao, J., Li, J., Zhang, H., Wang, X., Liu, P. et al. (2018) Natural variation in OsLG3 increases drought tolerance in rice by inducing ROS scavenging. Plant Physiol. 178, 451–467.

Yang, H., Shi, Y., Liu, J., Guo, L., Zhang, X. and Yang, S. (2010) A mutant CHS3 protein with TIR-NB-LRR-LIM domains modulates growth, cell death and freezing tolerance in a temperature-dependent manner in Arabidopsis. Plant J. 63, 283–296.

Yano, K., Yamamoto, E., Aya, K., Takeuchi, H., Lo, P.C., Hu, L., Yamasaki, M. 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.

Yu, J., Xiong, H., Zhu, X., Zhang, H., Li, H., Miao, J., Wang, W. et al. (2017) OsLG3 contributing to rice grain length and yield was mined by Ho-LAMap. BMC Biol. 15, 28.

Yu, J., Miao, J., Zhang, Z., Xiong, H., Zhu, X., Sun, X., Pan, Y. et al. (2018) Alternative splicing of OsLG3b controls grain length and yield in japonica rice. Plant Biotechnol. J. 16, 1667–1678.

Zhang, Q., Shen, B., Dai, X., Mei, M., Saghai Maroof, M.A. and Li, Z. (1994) Using bulked extremes and recessive class to map genes for photoperiod-sensitive genic male sterility in rice. Proc. Natl Acad. Sci. USA, 91, 8675–8679.

Zhang, Y., Su, J., Duan, S., Ao, Y., Dai, J., Liu, J., Wang, P. et al. (2011) A highly efficient rice green tissue protoplast system for transient gene expression and studying light/chloroplast-related processes. Plant Methods, 7, 30.

Zhang, H., Zhang, J., Wei, P., Zhang, B., Gou, F., Feng, Z., Mao, Y. et al. (2014) The CRISPR/Cas9 system produces specific and homozygous targeted gene editing in rice in one generation. Plant Biotechnol. J. 12, 797–807.

Zhang, Z., Li, J., Pan, Y., Li, J., Zhou, L., Shi, H., Zeng, Y. et al. (2017) Natural variation in CTB4a enhances rice adaptation to cold habitats. Nat. Commun. 8, 14788.

Zhao, D., Li, Q., Zhang, C., Zhang, C., Yang, Q., Pan, L., Ren, X. et al. (2018) G39 acts as a transcriptional activator to regulate rice grain shape and appearance quality. Nat. Commun. 9, 1240.

Zheng, T., Hong, Y., Zhang, H., Wu, Z., Wang, W., Tai, S., Lu, C. et al. (2015) Rice functional genomics and breeding database (RFGB): 3K-rice SNP and Indel sub-database. Chin. Sci. Bull. 60, 367.

Zhu, Z., Tan, L., Fu, Y., Liu, F., Cai, H., Xie, D., Wu, F. et al. (2013) Genetic control of inflorescence architecture during rice domestication. Nat. Commun. 4, 2200.

Zong, Y., Chen, Z., Innes, J., Chen, C., Wang, Z. and Wang, H. (2007) Fire and flood management of coastal swamp enabled first rice paddy cultivation in east China. Nature, 449, 459–462.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1 Co-segregation of awn and long grain traits.

Figure S2 Comparison of awn length and awn proportion between NIL-GLA and NIL-gla NILs.

Figure S3 GLA genomic sequences in NIL-GLA and NIL-gla.

Figure S4 Sequence alignment of GLA proteins of NIL-GLA and NIL-gla.

Figure S5 Comparison of the amino acid sequences in GLA alleles.

Figure S6 Sub-cellular localization of GLA protein in rice protoplasts.

Figure S7 Phenotypic analysis of transgenic plants.

Figure S8 Proposed evolutionary pathway of awns in O. sativa.

Table S1 Genetic analysis of BC3F2 and BC3F4 populations.

Table S2 Putative genes in the 26.32 kb GLA region.

Data S1 Primers used in this study.

Data S2 Information of cultivated and wild rice.

Data S3 Information of cultivated rice used in the correlation analysis between GLA mRNA levels and grain length.

Data S4 Information of cultivated rice and wild rice used in GLA analyses of nucleic acid diversity, neutral test and a minimum spanning tree.

Data S5 Information of cultivated rice and wild rice used in the phylogenetic tree and allele frequency analyses.

Data S6 Information of cultivated rice used in the genotyping analysis.