RESEARCH ARTICLE

Genomewide association analysis for awn length linked to the seed shattering gene *qSH1* in rice

RISPER AUMA MAGWA[1], HU ZHAO[1], WEN YAO[1], WEIBO XIE[1], LIN YANG[1], YONGZHONG XING[1,2] and XUFENG BAI[1]*

[1]*National Key Laboratory of Crop Genetic Improvement and National Center of Plant Gene Research (Wuhan), Huazhong Agricultural University, Wuhan 430070, People's Republic of China*
[2]*Hubei Collaborative Innovation Center for Grain Industry, Yangtz University, Jinzhou 434025, People's Republic of China*

Abstract

Awn is one of the most important domesticated traits in rice (*Oryza sativa*). Understanding the genetic basis of awn length is important for grain harvest and production, because long awn length is disadvantageous for both grain harvest and milling. We investigated the awn length of 529 rice cultivars and performed a Genomewide association studies (GWAS) in the *indica* and *japonica* subpopulations, and the whole population. In total, we found 17 loci associated with awn length. Of these loci, seven were linked to previously reported quantitative trait loci, and one was linked to the awn gene *An-1*. Nine novel loci were repeatedly identified in different environments. One of the nine associations was identified in both the whole and *japonica* populations. Special interest was the detection of the most significant association SNP, sf0136352825, which was less than 95 kb from the seed shattering gene *qSH1*. These results may provide potentially favourable haplotypes for molecular breeding in rice.

[Magwa R. A., Zhao H., Yao W., Xie W., Yang L., Xing Y. and Bai X. 2016 Genomewide association analysis for awn length linked to the seed shattering gene *qSH1* in rice. *J. Genet.* **95**, 639–646]

Introduction

Asian cultivated rice (*Oryza sativa*) was domesticated from its wild ancestor, *Oryza rufipogon* (Zong *et al.* 2007; Izawa *et al.* 2009; Huang *et al.* 2011). Wild rice differs from cultivated rice in several of its traits including seed shattering, hull and pericarp colour, awn length and plant architecture. These traits were domesticated by artificial selection as the wild rice became cultivated rice. To date, two seed shattering genes, *qSH1*, a grain shattering QTL on chromosome 1, and *SH4*, a grain shattering QTL on chromosome 4, have been mapped using the populations derived from an *indica* cross to *japonica* and an *indica* cross to wild rice, respectively (Konishi *et al.* 2006; Li *et al.* 2006). *qSH1* encodes a BEL1-type homeobox-containing protein and has a single-nucleotide polymorphism (SNP) in the 5′ upstream regulatory region. This SNP has been found to be the cause of the loss of seed shattering throughout the history of rice domestication (Konishi *et al.* 2006). *SH4* has been predicted to be a transcription factor and has a nucleotide substitution of G for T in an exon, causing an amino acid substitution of lysine to asparagine in *O. sativa*. The site of this amino acid substitution has been selected for the development of nonshattering cultivars during rice domestication (Li *et al.* 2006).

The awn is a morphological characteristic found in most cereal crops, including rice, wheat (*Triticum aestivum*), barley (*Hordeum vulgare*), oats (*Avena sativa*) and sorghum (*Sorghum bicolor*). In wild plants, long awns have been reported to aid in seed dispersal, burial and protecting grains from animal predation (Elbaum *et al.* 2007). However, a long awn-like seed shattering is not a desirable trait, particularly during crop harvest because it decreases grain yield, which has not been selected during domestication. Genetic analysis has shown that the awn is a complex trait, and a few awn-related QTLs have been identified in rice (Cai and Morishima 2002; Thomson *et al.* 2003; Gu *et al.* 2005a, b; Wang *et al.* 2011). *An-1* has been isolated through map-based cloning using an advanced backcross population and has been found to encode a basic helix–loop–helix protein regulating awn length, grain size and grain number (Luo *et al.* 2013). A mutant of *TOB1* (*TONGARI-BOUSHI1*) is involved in the formation of the awn, and *TOB1* has been reported to encode

*For correspondence. E-mail: a020809@webmail.hzau.edu.cn.

Keywords. awn length; genomewide association analysis; single-nucleotide polymorphism; genetic linkage.

a YABBY protein (Tanaka *et al.* 2012). The *DL* (*DROOPING LEAF*) gene is required for the formation of the awn, whereas the loss of the expression of *OsETT2* (*OsETTIN2*) in the awn primordium has been hypothesized to be associated with a failure to initiate awn formation in *japonica* (Toriba and Hirano 2014) Recently, it has been reported that a frameshift deletion in *LABA1* (*LONG AND BARBED AWN1*), which encodes a cytokinin-activating enzyme, reduces the cytokinin concentration in the awn primordia, thus, leading to the disruption of barb formation and awn elongation in cultivated rice (Hua *et al.* 2015). These studies of awn-related QTL and genes have provided valuable insight into the genetic basis of the awn. Although these cloned awn length genes have shed some light on the awn formation in rice, they can only partially explain the variation in the natural population. A clear molecular mechanism for the transformation of the long awns of wild rice to the nonawn of cultivated rice is still not properly understood.

Genomewide association studies (GWAS) are an alternative for gene mapping due to their evident advantage in simultaneously comparing the effects of several alleles by screening through a large number of accessions. Recently, a GWAS for awn presence or absence has been performed in rice (Zhao *et al.* 2011). The linear mixed model (LMM), also known as a mixed linear model has been used in GWAS to avoid false positive associations because of the consideration of both the structure of the population and the unequal relatedness among the individuals (Yu *et al.* 2006; Zhang *et al.* 2010; Chen *et al.* 2014; Yang *et al.* 2014). In this study, we performed a GWAS analysis of awn length in an extensive rice germplasm in two different environments using the LMM method.

Materials and methods

Plant materials and field experiments

A diverse worldwide collection of 529 *O. sativa* accessions (Chen *et al.* 2014; http://ricevarmap.ncpgr.cn) were planted in a bird-net-equipped field on the experimental farm of Huazhong Agricultural University during the winter of 2013 in Hainan and the 2014 rice growing season (summer) in Wuhan, China. Field trials were carried out following a randomized complete block design with two replicates each year. Seven 25-day old seedlings from each accession were planted in each row with 16.5 cm between the plants within a row and 26.4 cm between the rows. Field management was conducted according to standard agronomic practices. The five plants in the middle of each row were harvested individually to score awn length.

Measurement of awn length

The accessions were grown to maturity in the field and three spikelets from the five middle plants with fully filled seeds were harvested for awn length measurements. The awn lengths were measured by means of a ruler, and the mean values of the five plants for each accession were scored and used for the GWAS analysis.

SNP database

Five hundred and twenty-nine *O. sativa* landrace and elite accessions were genotyped (Chen *et al.* 2014). The detailed SNP information is available at http://ricevarmap.ncpgr.cn. The physical locations of the SNP markers were cited in TIGR Rice Genome Annotation project (ver. 6.1).

Statistical analyses

Broad-sense heritability (H^2) for awn length was calculated based on the experiments using the formula: $H^2 = \delta_g^2/(\delta_g^2+\delta_{ge}^2/n+\delta_e^2/nr)$, where δ_g^2, δ_e^2 and δ_{ge}^2 are the estimates of genetic, error variances and genotype by environment derived from the mean square expectations of a two-way ANOVA, respectively, $n = 2$ is the number of environments and $r = 2$ is the number of replicates. Linkage disequilibrium (LD) was estimated by using standardized disequilibrium coefficients (D'). Squared allele-frequency correlations (r^2) for the pairs of SNP loci were determined by using the TASSEL program. LD plots were generated in Haploview 4.2 indicating r^2 values between pairs of SNPs multiplied by 100; white, $r^2 = 0$; shades of gray, $0 < r^2 < 1$; black, $r^2 = 1$. The local genome sequences around *qSH1* were downloaded from RiceVarMap (http://ricevarmap.ncpgr.cn/). The SNP data were extracted from this dataset (http://ricevarmap.ncpgr.cn/). Nucleotide diversity was calculated using the DnaSP 5.0 program (Librado and Rozas 2009). The value of π_c/π_w was calculated by dividing the average proportion of pairwise differences per base pair in cultivated rice (π_c) by that in wild rice (π_w).

GWAS

A total of 3,916,415, 2,767,159 and 1,857,845 SNPs were used for the GWAS in the whole population, *indica* and *japonica* subpopulations, respectively. The SNPs with minor allele frequencies of $\geq$0.05 and the varieties with minor allele frequencies of $\geq$6 in a population were used. LMM was used for association analysis by running the FaST-LMM program (Lippert *et al.* 2011). Using a method described by Li *et al.* (2012), the effective number of independent SNPs was calculated as 757,578, 571,843 and 245,348 for the whole population, *indica* and *japonica*, respectively. The suggestive P values were specified as 1.3×10^{-6}, 1.8×10^{-6} and 4.1×10^{-6} for the whole population, *indica* and *japonica* subpopulations, respectively. The thresholds were set as $P = 1.0 \times 10^{-6}$ to identify significant association signals by LMM. In addition, the GLM (general linear model) was used to further validate and assess the association between the SNPs within the cloned gene *An-1* and the awn length. To obtain independent association signals, multiple SNPs exceeding the threshold in a 5-Mb sliding window were clustered by r^2 of LD $\geq$ 0.25, and the SNPs with the minimum P value in a cluster

were considered to be the lead SNPs. The detailed method has been described in a previous study (Chen *et al.* 2014; Wang *et al.* 2015).

Results

Phenotypic variation and heritability

The worldwide rice collection of 529 cultivars used in this study was classified into five subpopulations (Chen *et al.* 2014). *Indica* (295 cultivars) and *japonica* (156 cultivars) were the only subpopulations considered for the GWAS because they were characterized by strong population structures. Large variations in awn length were observed in the whole population and the two subpopulations, *indica* and *japonica*, in both the environments (table 1). Awn lengths ranged from 0 to 6.7 cm in the whole population. In *indica*, the awn length was on an average 0.2 cm in Hainan and 0.3 cm in Wuhan. Whereas, in *japonica*, the awn length was 0.7 cm and 1.2 cm in Hainan and Wuhan, respectively. On an average, the *japonica* awn length was three times longer than that of *indica*. Moreover, the awn length scored in Wuhan was longer than that scored in Hainan in any of the populations. The heritability (H^2) of awn length was at least 93.0% in all the populations (table 1).

Table 1. The awn length variation of the germplasm in two environments and broad-sense heritability (H^2).

	Hainan		Wuhan		
Populations	Mean ± SD	Range	Mean ± SD	Range	H^2 (%)
Whole	0.4 ± 0.9	0–5.3	0.7 ± 1.3	0–6.7	94.0
Indica	0.2 ± 0.5	0–5.3	0.3 ± 0.6	0–4.4	93.0
Japonica	0.7 ± 1.2	0–5.0	1.2 ± 1.7	0–5.8	95.0

GWAS in the three panels

There was a large phenotypic variation observed in the *indica* and *japonica* subpopulations (table 1), which could probably be attributed to the genetic variation in the intrasubpopulations. To explore this, we conducted GWAS in these two subpopulations individually and in the whole population, and used LMM to identify the significant associations (figure 1; table 2). Since the LD decay in cultivated rice has been extended from 100 to 200 kb (Mather *et al.* 2007; McNally *et al.* 2009), the association signals detected within a range of 50 kb in different environments and/or populations were considered to be associated with the same locus. In this study, 17 associations including *An-1* (tables 2 and 3) distributed across the 12 rice chromosomes except on chromosomes 2 and 3 were identified by LMM. The associations on chromosome 12 were identified in both the whole and *japonica* populations (table 2). Owing to the strong LD level ($r^2 = 0.71$)

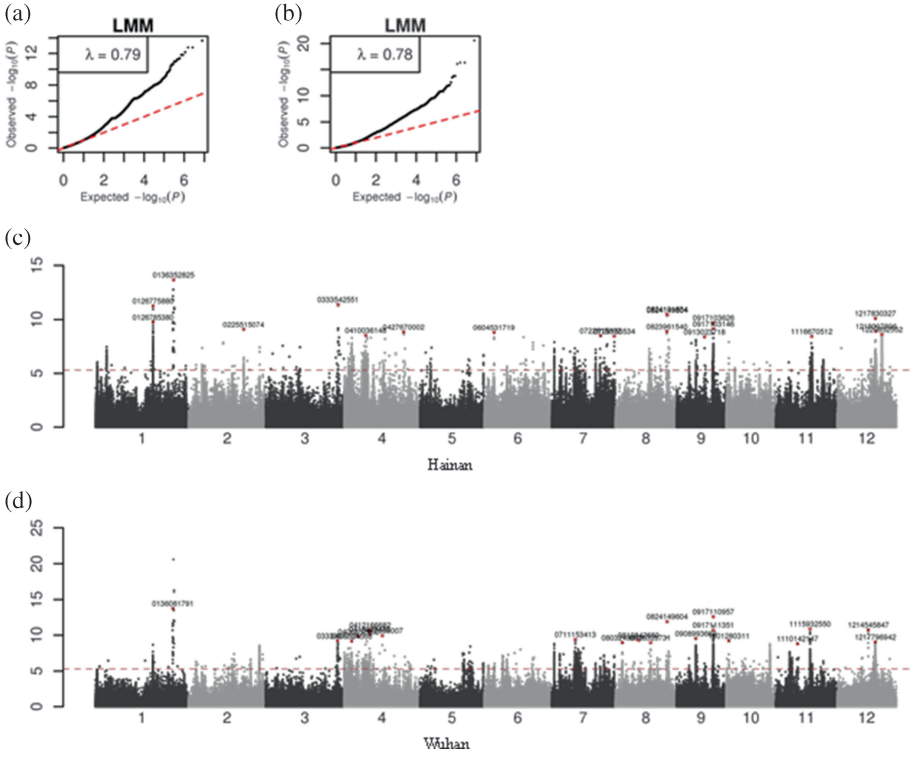

Figure 1. GWAS of awn length in the whole population. Quantile–quantile plot of linear mixed model (LMM) for awn length in the Hainan and Wuhan environments (a & b); Manhattan plots of LMM for awn length in the Hainan and Wuhan environments (c & d).

Table 2. Genomewide association signals for awn length detected in different environments using linear mixed model.

Pop.	Chr.	Lead SNP	Hainan		Wuhan		Neighbouring QTL	Reference
			P_{LMM}	Var. (%)	P_{LMM}	Var. (%)		
Whole	1	sf0136352825	2.2E-14	15.1	1.8E-14	21.5	*qAL-1*	Wang *et al.* (2011)
Whole	4	sf0401888964	3.8E-07	23.0	2.2E-07	30.1		
Whole	7	sf0701260690	1.0E-07	19.8	1.6E-09	23.3	*AWL7*	Cai and Morishima (2002)
Whole	8	sf0824149604	3.9E-11	14.8	1.3E-12	19.3	*AWN8.1*	Cai and Morishima (2002)
Whole	9	sf0917103626	2.0E-10	16.5	2.8E-13	17.6	*AWL9*	Cai and Morishima (2002)
Whole	12	sf1217830327*	7.9E-11	14.9	9.1E-10	17.9		
Ind.	4	sf0401039257	4.4E-08	18.2	6.6E-08	15.0		
Ind.	4	sf0404244016	1.1E-08	8.8	7.8E-07	5.9		
Ind.	5	sf0525735689	2.8E-08	10.6	4.1E-07	7.1		
Ind.	6	sf0619349876	7.4E-08	9.2	3.4E-07	11.7		
Ind.	10	sf1010589760	9.9E-10	15.8	5.8E-07	19.1	*AWL10*	Cai and Morishima (2002)
Ind.	11	sf1123292112	1.1E-08	10.0	1.0E-07	10.6		
Ind.	12	sf1200772645	1.1E-07	11.0	1.8E-07	12.7		
Jap.	1	sf0126785380	6.6E-07	28.8	1.8E-06	37.9		
Jap.	8	sf0823997428	7.9E-12	38.1	1.6E-10	44.7	*AWN8.1*	Cai and Morishima (2002)
Jap.	8	sf0824858096	2.2E-07	11.8	5.2E-07	21.4	*AWN8.1*	Cai and Morishima (2002)
Jap.	12	sf1217836172*	5.6E-07	26.8	8.0E-07	31.5		

*Difference between the associated SNPs in the two populations was below 50 kb, thus considered to be within the same locus. Ind., *indica*; jap., *japonica*; pop., populations; chr., chromosome; var., variation.

Table 3. The associated SNPs near *An-1* identified by LMM and GLM.

Gene	Chr.	Lead SNP	Locus[a]	Pop.	*P*	Var. (%)	Env.
An-1	4	sf0416574000^{LMM*}	−11 kb	Jap.	3.9E-08	38.8	Wuhan
		sf0416570642^{GLM*}	−7.8 kb	Whole	1.6E-08	5.4	Hainan
					1.0E-10	7.7	Wuhan

*The difference between the associated SNPs in the two environments was ∼3.3 kb, and considered to be within the same locus. [a]The negative value means the associated SNP is located upstream of the 5′ UTR of *An-1*. Chr., chromosome; ind., *indica*; jap., *japonica*; pop., populations; var., variation; env., environments.

between associations sf0136352825 and sf0136081791 on chromosome 1, the associations were considered to be one despite a distance of 271 kb between them. Four associations were identified on chromosome 4, and three were identified on chromosome 8 (tables 2 and 3). Nine of the 17 associations, including two in the whole population, six in the *indica* population and one in the *japonica* population, were not located near regions with known awn-related genes and/or QTL (table 2). The six associations identified in the whole population explained the phenotypic variance of 14.8–30.1%. The association on chromosome 4 provided the largest contribution to awn length variation. The four associations on chromosomes 1, 8, 9 and 12 identified in the whole population had almost the same contribution to the phenotypic variance in Hainan (14.8–16.5%). In general, the phenotypic variances explained by these four loci in Wuhan were all larger than those in Hainan. Six of the seven associations in *indica*, in a single environment were able to explain more than 10% of the phenotypic variance. Three of the four associations in *japonica* were able to explain more than 25% of the phenotypic variance in the two environments. The associations on chromosome 8 were the largest contributors to the phenotypic variance in the *japonica* subpopulation explaining a variance of up to 44.7% (table 2).

Comparison of the associated loci with those found in previous studies

This study identified a total of 17 associations in the whole population, as well as the *indica* and *japonica* subpopulations (tables 2 and 3). The lead SNP sf0416574000 discovered less than 11 kb away from the awn length gene, *An-1* (Luo *et al.* 2013) which was identified in *japonica* by LMM in the Wuhan environment. Interestingly, an association signal, sf0416570642, was identified by GLM in the whole population at the same *An-1* locus, ∼3.3 kb away from sf0416574000 in both environments (table 3). In addition to being detected near the *An-1* gene, both sf0416574000 and sf0416570642 were within the locus of an awn presence or absence QTL. In previous studies, this QTL has repeatedly been identified in two different populations in the RM307–RM185 region of chromosome 4 (Thomson *et al.* 2003;

Luo *et al.* 2013). Due to the minimal distance of ~3.3 kb between the association signals, sf0416574000 and sf0416570642 were believed to be within one locus. It was observed that the association signals detected on chromosomes 7, 8, 9 and 10 were found within the previously mapped awn presence or absence QTL, *AWL7*, *AWN8.1*, *AWL9* and *AWL10*, respectively. Finally, the association signal on chromosome 1, sf0136352825, was ~5 Mb away from *qAL1*, a previously identified major QTL for awn length (table 2).

A major locus for awn length was identified near qSH1

A significant association of sf0136352825 was identified in the whole population grown in both the Hainan and Wuhan environments (table 2; figure 2a). This association was able to explain more than 15% of the phenotypic variance and was located less than 95 kb away from the grain shattering gene, *qSH1* (Konishi *et al.* 2006). We investigated the LD level of a 416 kb region around *qSH1* on the basis of available SNPs (http://ricevarmap.ncpgr.cn). We examined the 29-kb upstream and the 386-kb downstream of *qSH1* that contained association SNPs (figure 2b). We found a significant LD between the association SNP sf0136352825 and any of the two SNPs, sf0136458144 and 0136473907, located between 13 kb and 28 kb upstream of *qSH1* (figure 2). We also investigated the nucleotide diversity of seven loci, including *qSH1*, which mapped within an ~300 kb interval. On average, ~26 SNPs per kb were detected at each locus in 183 wild-rice plants (unpublished data), but only seven SNPs were identified in the 529 cultivated rice accessions (http://ricevarmap.ncpgr.cn). The ratio of π_c/π_w for the *qSH1* locus was less than 0.4. Generally, the ratio of π_c/π_w for two loci located at 36.30 and 36.35 Mb near the association SNP, sf0136352825 (~36.35 Mb) was less than 0.2 (figure 3).

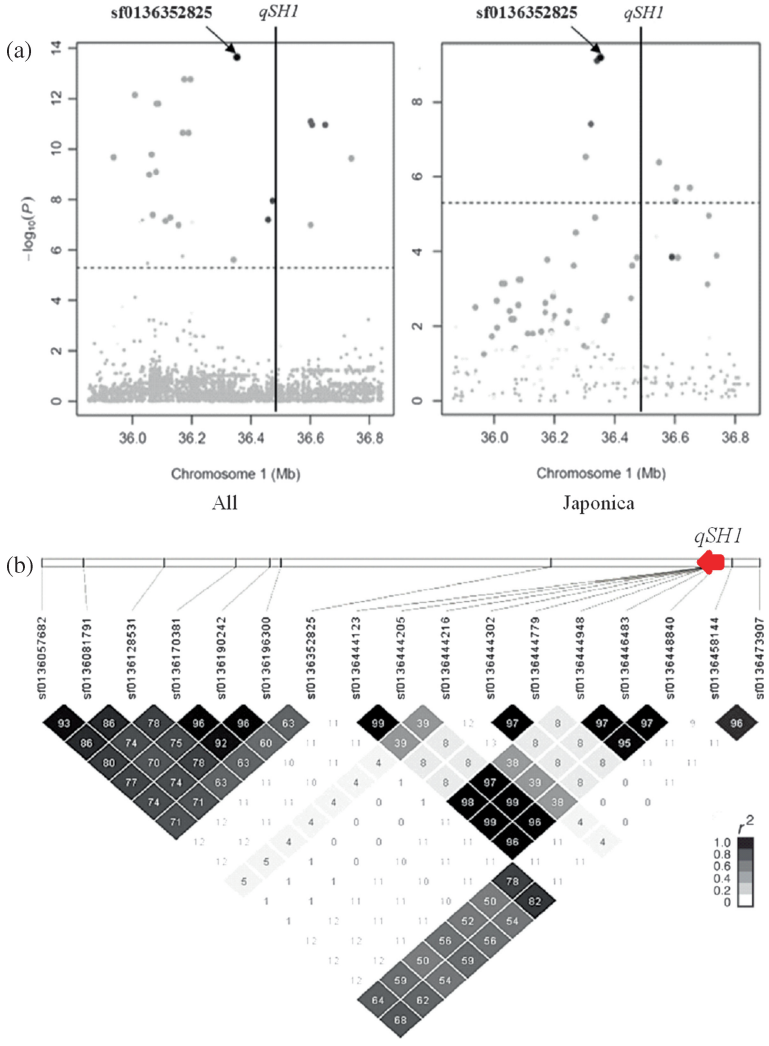

Figure 2. The associated SNPs identified near *qSH1* and LD analysis. (a) The associated SNPs identified around *qSH1* in both whole population and *japonica* population. The lead SNP, sf0136352825 is marked with a dot and indicated by an arrowhead. (b) The LD pattern around *qSH1*. The long rectangle represents the local genome of chromosome 1, the associated SNPs and the SNPs located within *qSH1* are shown below and were used to perform the LD analysis. The red arrow represents the cloned gene *qSH1*. The darkness of the colour of each box corresponds to the r^2 value according to the legend.

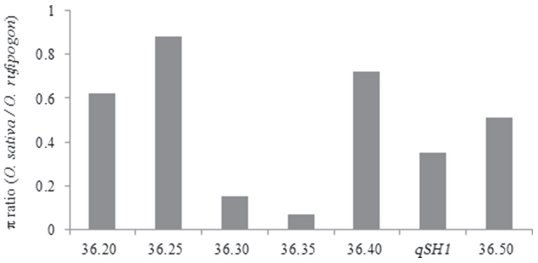

Figure 3. The relative ratio of π in cultivated rice compared to π in wild rice around *qSH1*. There was a selective sweep, ~36.30–36.35 Mb on chromosome 1, near *qSH1* in cultivated rice.

Discussion

GWAS: a powerful tool to identify loci controlling awn length

In previous studies, the recombinant inbred lines (RIL) and advanced backcross populations derived from cross combinations of *indica*, *japonica* and *O. rufipogon* have been used for awn presence or absence QTL mapping (Cai and Morishima 2002; Thomson *et al.* 2003; Gu *et al.* 2005a, b). The RIL population derived from a cross between 93-11 and Nipponbare has been used for QTL mapping and has revealed four awn length QTLs (Wang *et al.* 2011). Two loci for awn presence or absence have been identified in the BC_2F_2 population derived from the cross of tropical *Japonica* (Jefferson) and *O. rufipogon* (IRGC 105491) (Thomson *et al.* 2003). Cai and Morishima (2002) have reported a total of 13 awn presence or absence QTLs using the RIL derived from a cross between *indica* (Pei-kuh) and *O. rufipogon* (W1994). In this study, a total of 17 loci were identified in an association mapping population. Apparently, the genetic allele variations of the biparental populations were limited in comparison to an association population consisting of various accessions. This scenario may explain the detection of a larger number of association loci within the 529 cultivars in our GWAS study. It was also interesting to observe that six associated loci were found within the regions of previously mapped QTL. This indicates that these QTLs might be the candidates for these associations. The rice accessions in our study revealed a diversity of awn lengths among the different species of cultivated rice. Thus, the scoring mechanism of presence or absence of awn to detect QTL by linkage and association mapping used in the previous studies (Cai and Morishima 2002; Gu *et al.* 2005a, b; Zhao *et al.* 2011) might not be adequate. For example, Zhao *et al.* (2011) have performed a GWAS for awn presence or absence and have identified eight associations, whereas in our study, we identified 17 associations. The detection of more than twice the number of associations in our study suggests that awn length is a quantitative trait, and as such, the use of awn length for trait-marker association is more suitable than the use of the presence or absence of the awn. Thus, our study further validates the power of GWAS in identifying more genes that contribute to awn length in rice.

Different loci regulate awn length in japonica and indica

The Asian cultivar of rice, *O. sativa*, is primarily classified into two subspecies, *indica* and *japonica*. These subspecies differ in several domestication traits, such as cold sensitivity, germination and grain shape (Oka 1988). In this study, seven *indica* and four *japonica* associations for awn length were identified (table 2). The associations examined in the *indica* population were primarily distributed on chromosomes 4, 5, 6, 10, 11 and 12, whereas the associations identified in the *japonica* population were primarily located on chromosomes 1, 8 and 12. No associated locus was detected in both *indica* and *japonica* populations. For example, the *An-1* gene was identified only in *japonica* and the whole population, but it was not identified in *indica*. These results indicated that different loci regulating awn length in the *indica* and *japonica* subpopulations are probably due to divergent evolution and domestication. Indeed, some studies have suggested separate domestication events for *indica* and *japonica* rice from their ancestral species, *O. rufipogon* (Second 1982; Cheng *et al.* 2003). However, a study of rice genomic variation has suggested a divergent hypothesis that *japonica* rice was first domesticated from a specific population of *O. rufipogon*, and *indica* rice is a result of subsequent crosses between *japonica* rice and local wild rice (Huang *et al.* 2012). From these studies, it may be assumed that the two rice subpopulations will have different genes and/or alleles regulating the various rice traits, including awn length.

Genetic linkage of awn presence and grain shattering could be due to natural selection

Generally, wild rice has shattering seeds with long awns. Both grain shattering and a long awn aid in seed dispersal and procreation (Elbaum *et al.* 2007). *qSH1* has been identified as a major QTL for grain shattering (Konishi *et al.* 2006). In the present study, sf0136352825, the lead SNP was associated with awn length (table 2; figure 2a), was found to be less than 95 kb away from *qSH1*. This suggests that a QTL or gene that controls awn length is probably linked to *qSH1*. Interestingly, it has also been reported that two QTLs for awn presence or absence and grain shattering, *qAL4-2* and *qSH4*, are located near RM252 on chromosome 4 (Gu *et al.* 2005a). Similarly, *qAL8* and *qSH8* are located near RM531 on chromosome 8 in the BC_4F_2 population (Gu *et al.* 2005a). Moreover, a QTL for grain shattering, *sh4.1* has been detected near the awn length gene *An-1* (Thomson *et al.* 2003). The association SNP sf0434102682 (P = 1E-07, *indica*, 2014WH), identified in this study, was located less than 55 kb away from the domesticated seed shattering gene, *sh4* (data not shown). Together, these data indicate the possibility of a genetic linkage between the QTL for awn length and seed shattering, probably due to the natural selection. In wild rice, seed awn length and shattering are both important factors in aiding seed dispersal, burial and in protecting grains from animal predation. The natural selection of either one could automatically lead to the selection of the

other, owing to the close genetic linkage between the two traits. This speculation can be validated only by cloning these associations and/or QTL using a map-based approach. Both long awns and grain shattering are not favourable for rice domestication. Human selection of a favorable mutant with either absence of awns or nonshattering grains could automatically lead to the transfer of unfavourable allele of the other trait. This automatic transfer is due to the close linkage of the two genes and would make it difficult to select for a plant with both absence of awns and nonshattering grains. This scenario may explain the current presence of grain shattering and/or long awns in some of the cultivated rice varieties. The loci for these two traits were among the targets of artificial selection during rice domestication as a result of their undesirable characteristics in cultivated rice (Konishi *et al.* 2006; Luo *et al.* 2013). In this study, there was a low ratio of π_c/π_w (<0.4) in the loci of both *qSH1* and the association locus for awn length (~36.35 Mb) (figure 3). Compared to wild rice, the lower diversity of SNPs in cultivated rice suggested that these two loci probably were subjected to artificial selection during rice domestication. In particular, long awns were not favourable for rice cultivation due to the difficulty experienced during grain harvest and processing, thus prompting artificial selection against long awns.

In conclusion, the association signals for awn length identified in this study may provide a rich source of knowledge in understanding the natural genetic variations underlying the evolution, domestication and breeding of *indica* and *japonica* rice in relation to awn length. A map-based approach using populations derived from biparental crosses would aid in validating some of the association loci, such as sf0136352825 detected in this study, which may be followed by fine mapping and cloning of the candidate loci. Because, a number of cultivated rice varieties in our study have a considerable awn length, the cloned loci or genes may be useful in developing awnless and/or reduced-awn length varieties by means of a marker-assisted selective breeding approach.

Acknowledgements

This work was supported by grants from the National Science Foundation of China (31300991, 91335201), the National Special Programme for Research of Transgenic Plants of China (2014ZX0800936B) and the 863 programme on the functional genomics of stress resistance and nutrient utility in rice (2012 AA10A303).

References

Cai H. W. and Morishima H. 2002 QTL clusters reflect character associations in wild and cultivated rice. *Theor. Appl. Genet.* **104**, 1217–1228.

Chen W., Gao Y. Q., Xie W. B., Gong L., Lu K., Wang W. S. *et al.* 2014 Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. *Nat. Genet.* **46**, 714–721.

Cheng C. Y., Motohashi R., Tsuchimoto S., Fckuta Y., Ohtsubo H. and Ohtsubo E. 2003 Polyphyletic origin of cultivated rice: based on the interspersion pattern of SINEs. *Mol. Biol. Evol.* **20**, 67–75.

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

Gu X. Y., Kianian S. F. and Foley M. E. 2005a Phenotypic selection for dormancy introduced a set of adaptive haplotypes from weedy into cultivated rice. *Genetics* **171**, 695–704.

Gu X. Y., Kianian S. F., Hareland G. A., Hoffer B. L. and Foley M. E. 2005b Genetic analysis of adaptive syndromes interrelated with seed dormancy in weedy rice (*Oryza sativa*). *Theor. Appl. Genet.* **110**, 1108–1118.

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 X. H., Zhao Y., Wei X. H., Li C. Y., Wang A. H., Zhao Q. *et al.* 2011 Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nat. Genet.* **44**, 32–39.

Huang X. H., Kurata N., Wei X. H., Wang Z.-X., Wang A. H., Zhao Q. *et al.* 2012 A map of rice genome variation reveals the origin of cultivated rice. *Nature* **490**, 497–501.

Izawa T., Konishi S., Shomura A. and Yano M. 2009 DNA changes tell us about rice domestication. *Curr. Opin. Plant. Biol.* **12**, 185–192.

Konishi S., Izawa T., Lin S. Y., Ebana K., Fukuta Y., Sasaki T. *et al.* 2006 An SNP caused loss of seed shattering during rice domestication. *Science* **312**, 1392–1396.

Li C. B., Zhou A. L. and Sang T. 2006 Rice domestication by reducing shattering. *Science* **311**, 1936–1939.

Li M. X., Yeung J. M., Cherny S. S. and Sham P. C. 2012 Evaluating the effective numbers of independent tests and significant *p*-value thresholds in commercial genotyping arrays and public imputation reference datasets. *Hum. Genet.* **131**, 747–756.

Librado P. and Rozas J. 2009 DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451–1452.

Lippert C., Listgarten J., Liu Y., Kadie C. M., Davidson R. I. and Heckerman D. 2011 FaST linear mixed models for genome-wide association studies. *Nat. Methods* **8**, 833–835.

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.

Mather K. A., Caicedo A. L., Polato N. R., Olsen K. M., McCouch S. and Purugganan M. D. 2007 The extent of linkage disequilibrium in rice (*Oryza sativa* L.). *Genetics* **177**, 2223–2232.

McNally K. L., Childs K. L., Bohnert R., Davidson R. M., Zhao K., Ulat V. J. *et al.* 2009 Genome wide SNP variation reveals relationships among landraces and modern varieties of rice. *Proc. Natl. Acad. Sci. USA* **106**, 12273–12278.

Oka H. I. (Ed.) 1988 *Origin of cultivated rice.* Japan Scientific Societies Press, Tokyo, Japan.

Second G. 1982 Origin of the genic diversity of cultivated rice (*Oryza spp.*): study of the polymorphism scored at 40 isozyme loci. *Jpn. J. Genet.* **57**, 25–57.

Tanaka W., Toriba T., Ohmori Y., Yoshida A., Kawai A., Mayama-Tsuchida T. *et al.* 2012 The *YABBY* gene *TONGARI-BOUSHI1* is involved in lateral organ development and maintenance of meristem organization in the rice spikelet. *Plant Cell* **24**, 80–95.

Thomson M. J., Tai T. H., McClung A. M., Lai X.-H., Hinga M. E., Lobos K. B. *et al.* 2003 Mapping quantitative trait loci for yield, yield components and morphological traits in an advanced backcross population between *Oryza rufipogon* and the *Oryza sativa* cultivar Jefferson. *Theor. Appl. Genet.* **107**, 479–493.

Toriba T. and Hirano H.-Y. 2014 The *DROOPING LEAF* and *OsETTIN2* genes promote awn development in rice. *Plant J.* **77**, 616–626.

Wang L., Wang A. H., Huang X. H., Zhao Q., Dong G. J., Qian Q. *et al.* 2011 Mapping 49 quantitative trait loci at high resolution through sequencing-based genotyping of rice recombinant inbred lines. *Theor. Appl. Genet.* **122**, 327–340.

Wang Q. X., Xie W. B., Xing H. K., Yan J., Meng X. Z., Li X. L. *et al.* 2015 Genetic architecture of natural variation in rice chlorophyll content revealed by genome wide association study. *Mol. Plant* **8**, 946–957.

Yang W. N., Guo Z. L., Huang C. L., Duan L. F., Chen G. X., Jiang N. *et al.* 2014 Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. *Nat. Commun.* **5**, 5087.

Yu J. M., Pressoir G., Briggs W. H., Bi I. V., Yamasaki M., Doebley J. F. *et al.* 2006 A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* **38**, 203–208.

Zhang Z. W., Ersoz E., Lai C.-Q., Todhunter R. J., Tiwari H. K., Gore M. A. *et al.* 2010 Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.* **42**, 355–360.

Zhao K. Y., Tung C. W., Eizenga G. C., Wright M. H., Ali M. L., Price A. H. *et al.* 2011 Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nat. Commun.* **2**, 467.

Zong Y., Chen Z., Innes J. B., 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.

Received 5 September 2015, in final revised form 4 January 2016; accepted 8 January 2016
Unedited version published online: 13 January 2016
Final version published online: 9 August 2016