Open Access Repository
International Crops Research Institute for Semi-Arid Tropics

Association mapping of height and maturity across five environments using the sorghum mini core collection

Hari D. Upadhyaya,[a,*] Yi-Hong Wang,[b,*] Shivali Sharma,[a] Sube Singh[a]

[a]Gene Bank, International Crops Research Institute for the Semi Arid Tropics, Patancheru 502 324, Andhra Pradesh, India.

[b]Department of Biology, University of Louisiana at Lafayette, Lafayette, LA 70504, USA.

Corresponding author: Yi-Hong Wang (e-mail: yxw9887@louisiana.edu).

Published on the web 08 June 2012.
Genome
DOI: http://dx.doi.org/10.1139/g2012-034

This is author version post print archived in the official Institutional Repository of ICRISAT www.icrisat.org

---

Association mapping of height and maturity across five environments using the sorghum mini core collection

Hari D Upadhyaya†[1], Yi-Hong Wang†*[2], Shivali Sharma[1], Sube Singh[1]

† Contribute equally to this work

[1] Gene Bank, International Crops Research Institute for the Semi Arid Tropics (ICRISAT), Patancheru 502 324, Andhra Pradesh, India.

[2] Department of Biology, University of Louisiana at Lafayette, Lafayette, LA 70504, USA.

* Corresponding author, yxw9887@louisiana.edu.

Abstract

Sorghum is a potential energy crop thanks to its high biomass productivity and low-input. Biomass yield in sorghum is defined by height and maturity. To develop molecular breeding tools for genetic improvement of these two traits, we have identified simple sequence repeat markers linked to height and maturity using a pool-based association mapping technique. The sorghum mini core collection was evaluated across five environments for height and maturity. Seven tall and seven short accessions were selected based on their height in all environments while six early- and ten late-maturing accessions were selected mostly based on their maturity in two post-rainy seasons for the two maturity pools. Two additional height pools were constructed based on phenotypes in one environment. The three pairs of pools were screened with 703 SSR markers and 39 polymorphic markers were confirmed by individual genotyping. Association mapping of the 39 markers with 242 accessions from the mini core collection identified five markers associated with maturity or height. All were clustered on chromosomes 6, 9, and 10 with previously mapped height and maturity markers or QTLs. One marker associated with both height and maturity was 84 kb from recently cloned *Ma1*. These markers will lay a foundation for identifying additional height and maturity genes in sorghum.

Key words: Sorghum, pool-based association mapping, SSR markers, molecular breeding, height, maturity

Introduction

Sorghum (*Sorghum bicolor* (L.) Moench) has been proposed as a model to study C4 grasses as a potential energy crop (Carpita and McCann 2008). It is a recognized biomass crop suitable for biofuel production because of two critical factors: high biomass production and efficient water use. Compared to other potential energy crops, sorghum can produce much more dry biomass with limited input (see Vermerris 2008). Sorghum biomass productivity is highly and positively correlated with maturity (days to 50% flowering) (Haussmann et al. 2002) and height (Murray et al. 2008*b*; Ritter et al. 2008; Zhao et al. 2009) which, in turn, is highly correlated with maturity (Kebede et al. 2001; Ali et al. 2008; Ritter et al. 2008; Upadhyaya et al. 2009). Because of such a high correlation between height and biomass yield, a quantitative trait locus (QTL) for total dry biomass has been found to co-localize with height QTLs on chromosomes 7 and 9 in sorghum (Brown et al. 2008; Murray et al. 2008*b*).

In addition to producing more biomass, taller sorghums have other advantages as well. First, they tend to have thicker stems (Ayana and Bekele 2000; Elangovan et al. 2007; Makanda 2009) with higher sugar content (Ritter et al. 2008). It is thus not surprising that a marker tightly associated with height was also tightly associated with sugar content (Murray et al. 2009). Second, under warm and humid environment during grain maturation, taller varieties are less likely to develop grain mould (Klein et al. 2001). Third, taller varieties may also produce a better ratoon crop because height is significantly correlated with regrowth fresh biomass (Murray et al. 2008*b*). And lastly, because height is also independent of stem structural composition, i.e., cellulose, hemicellulose and lignin content (Murray et al. 2008*b*), a variety with thick and tall stems can be bred to contain more cellulose and less lignin. In maize, plants do compensate low

lignin with higher hemicellulose content although drastically reduced lignin content significantly stunts plant growth (Vermerris et al. 2010). Some tall sorghums may be prone to lodging, but this is not always a case (Rooney 2004; Murray et al. 2009). To improve sorghum biomass yield through manipulation of height or maturity, it is imperative to identify genetic markers linked to or genes underlying height and maturity to facilitate molecular breeding.

Height is one of the most extensively studied traits in sorghum. Genetic linkage mapping studies have shown that sorghum height is controlled only by a few major QTLs (reviewed by Salas Fernandez et al. 2009) with high heritability (0.83-0.95; Murray et al. 2008*b*; Ritter et al. 2008). Because of this, many studies have focused on height as an indirect way to identify genes related to biomass yield (Salas Fernandez et al. 2009). An earlier genetic study has found four loci affecting sorghum height: *Dw1*, *Dw2*, *Dw3* and *Dw4* (Quinby and Karper 1954). Only *Dw3* (*SbPGP1* hereafter) has been cloned to date. It encodes a P-glycoprotein that regulates polar auxin transport and is orthologous to the maize *br2* (Multani et al. 2003). *SbPGP1* also co-localizes with a height QTL on chromosome 7 (Brown et al. 2006; Murray et al. 2008*ab*). Progress on identifying the other three has been made. *Dw2* is found to be associated with a QTL on chromosome 6 (Feltus et al. 2006; Klein et al. 2008; Mace and Jordan 2011). Although *Dw1* and *Dw4* have not been genetically mapped conclusively to a linkage group, another QTL (*Sb-HT9.1*) for height has been mapped on chromosome 9 (Brown et al. 2008). An additional height markers were also mapped to chromosome 9 (Murray et al. 2008*ab*; Wang et al. 2012) close to *Sb-HT9.1*. In total, linkage mapping identified 32 markers linked to sorghum height and maturity that can be physically localized on sorghum chromosomes (Lin et al. 1995; Kebede et al. 2001; Klein et al. 2001; Feltus et al. 2006; Ritter et al. 2008; Mace and Jordan 2011).

Association mapping which maps QTLs based on linkage disequilibrium has been employed to map plant height in sorghum. Compared to genetic mapping using population generated from two parents, association mapping using a collection of varieties generally provides higher resolution (Collins and Morton 1998; Yu and Buckler 2006). This mapping method has been extensively used in plants and humans. Using this method in sorghum, Brown et al. (2008) mapped the afore-mentioned *Sb-HT9.1* on chromosome 9 with a panel of 377 sorghum varieties and 49 SSR markers. Five more markers (pSB0945, Xgap72, Xtxp343, Xtxp265 and pSB0301) associated with height were found by Murray et al. (2009) in a panel of 125 genotypes with 47 SSR and 322 SNP markers. Among these, pSB0945 was on chromosome 9 and Xgap72 and Xtxp265 were on chromosome 6 (Murray et al. 2009), which supports published QTL studies (Lin et al. 1995; Pereira and Lee 1995; Brown et al. 2006; Feltus et al. 2006).

A variation of the above association mapping technique uses a pool-based two-stage method first described for association mapping of complex human diseases by Arnheim et al. (1985). The method has since been extensively used in association mapping of complex human conditions such as cocaine addiction (Drgon et al. 2010), nicotine dependence (Lind et al. 2010), and addiction to heroin (Nielsen et al. 2010) and has been tested in sorghum (Wang et al. 2011, 2012). The method is similar to DNA pooling in QTL mapping using bi-parental mapping populations as effectively demonstrated by Sun et al. (2010). In this method, genotypes with contrasting phenotypes (e.g., tall or short) will form two pools. Those in the tall pool will all be tall while those in the short pool will all be short. Both pools of DNA are then screened with markers and polymorphic markers are repeated in all individual genotypes. This confirmatory individual genotyping may also be extended to other individuals not in the initial pools (Sham et

al. 2002). Confirmed markers are then used to genotype all accessions in a mapping panel to test the trait-marker association. In this study, we used this method with 703 SSR markers and height and maturity data of the sorghum mini core collection (Upadhyaya et al. 2009) from five environments to further validate the utility of the pool-based mapping method in sorghum by identifying markers associated with height and maturity. These markers will be useful to improve sorghum biomass yield using marker-assisted selection (Xu and Crouch 2008).

Materials and methods

Plant materials

Two hundred forty two accessions from the sorghum mini core collection described by Upadhyaya et al. (2009) were evaluated in three rainy season (denoted as 2008R, 2009R, and 2010R) and two post-rainy season (denoted as 2005PR and 2009PR) environments. The 2005PR data were from Upadhyaya et al. (2009). The experiment was sown in an alpha design with two replication except 2010R when three replications were used. Each plot was single-row, 4 m long, with a row-to-row spacing of 75 cm, and plant-to-plant spacing within a row of 10 cm. Ammonium phosphate was applied at the rate of 150 kg ha–1 as a basal dose, and 100 kg ha–1 of urea was applied as top dressing after 3 wk of planting. Two to three irrigations in the rainy and five in post-rainy seasons (each with 7 cm water) were provided to support the crop. For the post-rainy season environments, five irrigations (each with 7 cm water) were provided at equal intervals until grain maturity. All accessions germinated well and produced panicles. Plant height was measured of the main stalk at 50% flowering and presented as a mean of five randomly selected plants. Days to 50% flowering was used as a measure of maturity and was measured by the number of days from the 50% seedling emergence to the date when 50% plants had started flowering.

DNA isolation and genotyping

Leaves from one plant for each field-grown accession were harvested and dried at room temperature. Total genomic DNA was isolated based on Klein et al. (2000) and Williams and Ronald (1994) with modifications. Dried leaves were cut into small pieces (about 2 mm^2) and filled into a 1.5 ml microcentrifuge tube followed by adding 1.4 ml extraction buffer containing

100 mM Tris pH 8.0, 10 mM EDTA pH 8.0, 700 mM NaCl, 12.5 mM potassium ethyl xanthogenate (PEX). Samples were incubated at 65°C for 30 min with occasional mixing in a dry heat block before centrifugation at 15000 g for 5 min. After centrifugation, 700 µl supernatant was transferred to a new 1.5-ml tube containing 700 µl isopropanol and 70 µl of 3 M sodium acetate pH 5.2, mixed and centrifuged again at 15000 g for 5 min. The precipitated DNA was washed twice with 70% ethanol, air dried and resuspended in 70 µl water containing 100 ng RNase A for at least 30 min at room temperature. The samples were centrifuged again at 15000 g for 5 min to remove impurities. The supernatant containing DNA was transferred to a new tube and quantified in NanoDrop 2000 (Thermo Scientific, Waltham, MA). DNA concentration was adjusted to 40 ng/µl for all PCR reactions.

To construct pools, accessions from the mini core collection were ranked based on height and maturity in all five environments. In order for accession to be included in the tall pool, it had to rank high in all five environments. The short pool consisted of accessions ranked low in each of the five environments. Accessions for the early maturity and late maturity pools were selected only based on days to 50% flowering in the two post-rainy seasons, 2005PR and 2009PR. Again, only those ranked high or low across both environments were selected. For height pools specific to 2009PR, top and bottom nine accessions were used to construct the tall and short pools, respectively.

To screen the pools with genetic markers, 703 SSR markers were randomly selected from the 15,194 SSRs identified in the genome sequence by Paterson et al. (2009) and primers were synthesized by Eurofins MWG Operon (Huntsville, AL) (see Wang et al. 2012 for primer sequences). PCR was performed in a 5 µl reaction volume that contains 2.5 µl 2x 360 AmpliTaq

Master Mix from Applied Biosystems (ABI, Carlsbad, CA), 10 ng each of the two primers and 40 ng DNA. Thermocycling was initiated with 95°C/5 min followed by 30 cycles of 95°C/20 s, 56°C/20 s, 72°C/1 min and final extension of 72°C/7 min in either an ABI 2720 or Veriti® thermocycler. The PCR samples were electrophoresized in a LabChip Microfluidic Gel Electrophoresis System (LabChip 90) from Caliper Life Sciences (Hopkinton, MA) with 1 K DNA Assay Kit according to the manufacturer's protocol. Before loading into LabChip 90, total PCR reaction volume was adjusted to 15 µl with sterile water for all samples. The default setting for LabChip 90 was used except the sipper height was adjusted to 1 mm. PCR product size and virtual gel image were generated automatically using LabChip GX 2.3 software from Caliper. Markers polymorphic between the pools were further confirmed in all eight selected accessions that make up the two pools according to Sham et al. (2002). Genotyping of all accessions with selected SSR markers was performed following the same procedures.

Association mapping

SSR markers polymorphic between accessions of the two pools were used in association mapping. Marker-trait associations were calculated using mixed linear models (MLM; Yu et al. 2006) as implemented in TASSEL (Bradbury et al. 2007; http://www.maizegenetics.net/) 3.0. MLM was controlled for population structure generated from STRUCTURE (Pritchard et al. 2000; http://pritch.bsd.uchicago.edu/ structure.html) 2.3. The number of groups/subpopulations (*k*) was tested between 1-12 when running STRUCTURE. The *k* value at which the posterior probability (Ln P(D)) plateaus as described in Casa et al. (2008) was used for association analysis. STRUCTURE was run with the admixture model, a burn-in period of 10,000 and 10,000 Markov Chain Monte Carlo repetitions as described by Evanno et al. (2005). The kinship

matrix was generated in TASSEL for MLM analysis. Significance of associations between loci and traits will be based on an F-test with p values calculated by TASSEL.

Marker localization

SSR markers associated with height or maturity were localized to the sorghum chromosomes based on Paterson et al. (2009). Markers previously identified as linked to height or maturity in sorghum were localized to the genome based sequence information of relevant markers provided in Map Viewer in the NCBI website (http://www.ncbi.nlm.nih.gov/mapview/). For pSB RFLP markers, sorghum marker sequences available from Map Viewer were searched against the sorghum genome presented in the following websites: http://www.plantgdb.org/SbGDB/, http://www.phytozome.net/sorghum, or http://genome.jgi-psf.org/ to determine their physical locations (these websites were also used to search for candidate genes close to markers identified in this study). For RFLP markers with sequence from other grasses, the sequence was used to search the sorghum nucleotide and EST database in GenBank. Top-match sorghum sequence was then used as above to place that marker. For SSR markers, if their position was not given in Map Viewer, their primer sequences were then used to search the genome database in http://www.phytozome.net/sorghum for physical localization. Alternatively, chromosome location of some SSR markers was provided by Ramu et al. (2010). Maps that also identified markers linked to height but for which sequencing information of these markers were not available in Map Viewer were not used in this study. Maps used were those in Map Viewer from Chittenden et al. (1994), Peng et al. (1999), Bhattramakki et al. (2000), Kong et al. (2000), and Bowers et al. (2003). Maps of chromosomes based on the physical distances in Mb were drawn using MapDraw 2.2 by Liu et al. (2003).

Results and discussion

Phenotypic correlation

To evaluate genotype performance of the mini core collection across all environments, Pearson's correlation was calculated for each trait among the different growing environments and between plant height and maturity. Height and maturity measured in 2008R, 2009R and 2010R were highly correlated with each other, especially maturity. Correlation between the two post-rainy seasons was lower and still lower between the rainy and post-rainy environments. The last was especially true for maturity and higher correlations were found in height (Table 1).

Height and maturity were highly correlated as expected. Except a lower but still significant correlation in 2010R, correlation between height and maturity was high and significant in all other environments (Table 2). This suggests that tall plants mature or flower late while early-maturing plants are mostly short. This result agrees with that of Ali et al. (2008) which reported a correlation of 0.83 based on the performance of 66 varieties and with that of Upadhyaya et al. (2009) which reported a correlation of 0.645 based on 2,247 varieties. However, this correlation may also depend on populations and the environment. Brown et al. (2006) evaluated 119 RILs from a cross between BTx623 and IS3620C and did not find any correlation (-0.07) between height and maturity. But upon evaluating 184 RILs from R9188 and R9403463-2-1, Ritter et al. (2008) found a correlation (0.164) significant at 0.05 level and Kebede et al. (2001) reported a correlation (0.203) significant at 0.01 level. Similarly, Murray et al. (2008a) evaluated 176 RILs from a cross between Rio and BTx623 and found the correlation (0.29) significant at 0.001 level. It seems that correlation coefficient between height and maturity

is much lower when RILs from genetic crosses are used. This may be because not all genetic variations were captured in one genetic cross.

Construction of DNA pools and marker screening

Selection criterion for the tall pool was that an accession had to be in the top 20% height across all five environments. Seven accessions were selected: IS-24139, IS-7250, IS-20625, IS-20632, IS-25301, IS-26025, and IS-7310. For the short pool, seven accessions in the bottom 10% height were also selected and they are IS-13782, IS-29582, IS-26749, IS-14090, IS-2872, IS-24462, and IS-3158. For the early- and late-maturity pools, the selection was based on data in 2005PR and 2009PR and selected accessions had to be in the top or bottom 10% maturity. The early-maturity pool contained IS-20298, IS-23992, IS-28313, IS-24462, IS-17941, and IS-28849 and the late-maturity pool consisted of IS-13893, IS-26701, IS-31557, IS-23684, IS-31446, IS-24175, IS-24218, IS-24139, IS-27034, and IS-13294. For the height pools specific to 2009PR, the top nine varieties were selected for the tall pool and the bottom nine varieties were selected for the short pools except for IS-30572. IS-30572 was not ranked among bottom nine but was in bottom 40%. It replaced IS-30838 which was ranked among bottom nine in 2009PR but top 50% in height in the four rainy season environments. The tall and short and early and late pools were then screened with the 703 SSR markers described below.

Screening of the tall and short DNA pools with the SSR markers produced 78 polymorphic markers while 82 were polymorphic between the early and late pools and 116 were between the two 2009PR-specific height pools. To conduct the confirmatory individual genotyping, four accessions were further selected from each pool based on their phenotypic values, i.e., four tallest accessions would be chosen from the tall pool and four shortest from the

short pool. The eight accessions representing the two extremes of each trait were genotyped with all polymorphic SSR markers. For height, this winnowed the 78 down to 9 markers that could clearly identify the phenotypes of at least seven of the eight genotypes. Similarly for maturity, 11 markers were identified from the 82 polymorphic SSRs. For the 2009PR-specific height pools, 19 markers were selected from the 116 polymorphic SSRs. So a total of 39 SSR markers were used to genotype all 242 accessions for association mapping.

Association mapping

Associations of the 39 SSR markers with height and maturity were analyzed using TASSEL 3.0. MLM model was used in the association analysis. The true number of subpopulations (*k* value) was tested with 29 markers with *k* varied from 1 to 12. The result showed that when $k = 4$, the posterior probability (Ln P(D)) was the highest and plateaued at higher *k*. True *k* was found to be 5 in a simulation study (Evanno et al. 2005) and this number was used in tomato association mapping (Mazzucato et al. 2008) and soybean population structure study (Lam et al. 2010). Therefore, we used $k = 4$ for our MLM analysis. To be more conservative in identifying markers associated with height and maturity, *p*-values ≤ 0.01 from the MLM model were presented together with associations of markers with *p*-values higher than 0.01 but less than 0.05. This allows us to show the pattern of association across the environments (Table 3). Because of the close correlation between height and maturity in the sorghum germplasms (Table 2; Ali et al. 2008), markers identified in the height analysis can be tightly associated with maturity and vice versa.

In order for a marker to be selected, it had to be associated with a trait in at least two environments and *p*-value of at least one association had to be lower than 0.01. In all, five

markers linked to height and maturity were identified and two of them were linked to one of the traits across at least three of the five environments (Table 3). 4-162 was linked to maturity and height in four and three environments, respectively and 39-1833 to both traits in three environments each. The other three markers were associated with height in 2-3 of the five environments. Of the three markers, 40-1897 and 44-2080 were among the four markers previously identified as linked to height (Wang et al. 2012).

Some associations were stronger than others. The strongest associations were found between 39-1833 and maturity, and between 4-162 and maturity. 39-1833 was associated with maturity in 2008R with a *p*-value of 0.000062 and its association with maturity was also significant in 2009R and 2010R. Similarly, 4-162 was associated with maturity in 2008R and 2010R with *p*-values of 0.00072 and 0.00027, respectively. Association of both markers with height was less strong, with $p < 0.01$ in only one environment for each marker (Table 3).

Physical mapping

The five markers were placed onto sorghum chromosomes together with other markers linked to height and maturity in previous studies. As mentioned earlier, there were 32 markers genetically linked to height and maturity that could be placed onto sorghum chromosomes. Together with the five markers linked to height identified by Murray et al. (2009), 37 genetic markers with known chromosome location are linked to height and maturity in sorghum. These markers were not equally distributed: ten on chromosome 6, seven on chromosome 9, five on chromosome 1, four on chromosome 2, and three each on chromosomes 4, 7, and 10. The distribution of the five markers identified in this study follows similar pattern: two on chromosome 6, two on chromosome 10, and one on chromosomes 9 (Figure 1).

Most markers were reasonably close to height and maturity markers or QTLs identified in previous studies (Figure 1). On the most extensively mapped chromosome 6, 39-1833 was 1.25 Mb from Xgap72 which was one of the five markers identified by Murray et al. (2009) linked to height. Xtxp568 and Xtxp547 were linked to height, maturity, Brix, and sugar yield QTLs by Ritter et al. (2008) and were placed between Xgap72 and Xtxp265 (at 51.17 Mb, not shown in Fig. 1) by Mace and Jordan (2011)---Xtxp568 and Xtxp547 couldn't be physically placed on chromosome 6 due to lack of sequencing information. Xgap72 was also closely linked with *Ma1* on chromosome 6, which has the largest impact on maturity of all the six maturity genes (Mace and Jordan 2011). This is in agreement with Patterson et al. (1995) that *Ma1* is located between pSB095 and pSB428a but close to pSB428a and partially overlaps with the QTLs by Ritter et al. (2008). *Ma1* has been cloned recently (Murphy et al. 2011) and it was located between 40.28 and 40.29 Mb, about 1.15 Mb to Xgap72 and 84 kb to 39-1833. More interestingly, QTLs for 16 traits including biomass yield, stem and grain harvest indices, and maturity may all be mapped to this region by Murray et al. (2008*ab*) based on the position of Xtxp6. Similarly, QTLs for maturity and height mapped by Shiringani et al. (2010) is also likely within the region. This is based on their anchor marker Xtxp17 located at 58.25 Mb, close to pSB421 but not shown in Fig.1 and the fact that chromosome 6 ends at 62.20 Mb (Paterson et al. 2009). Further studies are needed to investigate these hot regions for height and maturity QTLs on chromosome 6.

On chromosome 9, 44-2080 was less than 10 kb from *Sb-HT9.1*, a height QTL mapped by Brown et al. (2008) through high-resolution association mapping using SSR markers and about 400 kb from pSB945 which was most tightly (with lowest *p*-value) linked to height among all five markers identified by Murray et al (2009). Although the region of 44-2080 was flanked

by *Sb-HT9.1* and pSB945, no other QTLs have been mapped to this region. However, the fact that three markers tightly linked to height are all located within the same region of 710 kb suggests that the underlying gene is probably close by. On chromosome 10, marker 4-191 was about 50 kb from Xrz143.3 which was linked to two overlapping maturity QTLs (Hart et al. 2001). Marker 4-162 was about 1 Mb from the maturity QTLs QDTFL3_10 and QDTFL4_10 mapped by Crasta et al. (1999) based on physical and genetic distance from Mace and Jordan (2011).

Known height or maturity genes close to linked markers

The current results demonstrate that the eight markers identified in this work are clustered with height or maturity markers or QTLs mapped in other studies. Therefore the vicinity of the markers was examined to search for gene homologs that are known to regulate plant height and maturity. Those gene homologs were found through BLAST search (Altschul et al. 1997) and have been shown to affect height and maturity in plants.

For example, photoreceptors are known to regulate flowering in plants (Endo et al. 2007). At 0.17 Mb from 40-1897 on chromosome 6 was a homolog 40% identical and 64% similar to Arabidopsis *FAR1* (*far*-red-impaired response *1*) which participates phyA-mediated light response. *far1* mutants flower early, with only about half as many rosette leaves as wild type at the start of flowering (Li et al. 2011). The sorghum homolog was localized to the position based on a cDNA sequence (accession CD209023) and no corresponding gene was annotated in the genome. Another *FAR1* homolog 34% identical and 53% similar to the Arabidopsis *FAR1* was about 18 kb from 4-191 on chromosome 10. The photoperiod response gene *Ma1* (*SbPRR37*) suppresses flowering by activating the floral inhibitor CONSTANS and repressing the floral

activators Early Heading Date 1, FLOWERING LOCUS T, and *Zea mays* CENTRORADIALIS 8 (Murphy et al. 2011). *SbPRR37* is about 84 kb from 39-1833, the closest to the gene among the 703 markers analyzed in this study. A second class of genes that regulate height and maturity was related to hormone homeostasis. On chromosome 9, *GA2-oxidase*, which inactivates active gibberellins, was 50 kb from *Sb-HT9.1*, 40 kb from 44-2080, and 0.38 Mb from pSB945. In tobacco, overexpression of *GA 2-oxidase* reduced height by 86% (Biemelt et al. 2004) and 25% (Dayan et al. 2011) and in poplar the reduction was 4-fold (Busov et al. 2003). Thus close to the identified markers there are homologs of known genes that regulate height or maturity in plants. The next step is to carry out association mapping with these genes and if their association with height or maturity is confirmed then their function could be tested through genetic transformation of sorghum and Arabidopsis.

Based on a study on bi-parental mapping by Sun et al. (2010), using one marker per 2-5 cM of the genome and a population size of 250 will detect QTLs that explain over 7% variation with over 70% detection power. Although association mapping has considerably higher resolution than linkage methods (Collins and Morton 1998), markers identified in this study are most likely linked to major QTLs. To identify more minor QTLs, localized or genome-wide high-density mapping with more markers and larger panels will have to be used. Localized high-resolution association mapping with more markers has been demonstrated by Brown et al. (2008) which refined the position of *Sb-HT9.1* to 57.21 Mb based on preflag leaf height. As mentioned above, *Sb-HT9.1* is about 50 kb from *GA2-oxidase*. Since only 13 markers were used across a 7 Mb region surrounding the QTL, it is possible that more markers would further narrow down the candidate region. Doubling the population size from 250 to 500 increases detection power of

QTLs that explains over 2% variation from 50% to 70%. Localized high-resolution mapping is under way for selected regions described in this study.

Conclusion

In this study height and maturity loci in sorghum have been mapped using a pool-based association mapping strategy with with 703 SSR markers and phenotypic data from five environments. The association panel was a sorghum mini core collection developed from 22,473 landraces. In all environments, height was positively correlated with maturity at 0.001 level in the mini core collection, confirming previous studies also using sorghum germplasm collections. Five markers were identified as linked to height and maturity in more than two environments. Among the five, two were associated with height and maturity each in at least three environments. All the markers were either clustered with markers previously linked to height and maturity based on their chromosome locations or were colocalized with previously mapped height-related QTLs in sorghum. One of the two markers highly associated with height and maturity was 39-1833 which was 84 kb from the newly cloned photoperiod response gene *Ma1*/*SbPRR37*. The results further validate the utility of pool-based association mapping in sorghum. The markers identified in this study will lay a foundation for identification and cloning of height and maturity genes in the near future.

Acknowledgements

The project was supported by the University of Louisiana at Lafayette and the International Crops Research Institute for the Semi-Arid Tropics.

References

Ali, M.L., Rajewski, J.F., Baenziger, P.S., Gill, K.S., Eskridge, K.M., and Dweikat, I. 2008. Assessment of genetic diversity and relationship among a collection of US sweet sorghum germplasm by SSR markers. Mol. Breed. 21:497–509

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389-3402

Arnheim, N., Strange, C., and Erlich, H. 1985. Use of pooled DNA samples to detect linkage disequilibrium of polymorphic restriction fragments and human disease: studies of HLA class II loci. Proc. Natl. Acad. Sci. U.S.A. 82:6970–6974

Ayana, A., and Bekele, E. 2000. Geographical patterns of morphological variation in sorghum (*Sorghum bicolor* (L.) Moench) germplasm from Ethiopia and Eritrea: Quantitative characters. Euphytica, 115:91–104

Bhattramakki, D., Dong, J., Chhabra, K., and Hart, G.E. 2000. An integrated SSR and RFLP linkage map of *Sorghum bicolor* (L.) Moench. Genome, 43: 988–1002

Biemelt, S., Tschiersch, H., and Sonnewald, U. 2004. Impact of altered gibberellin metabolism on biomass accumulation, lignin biosynthesis, and photosynthesis in transgenic tobacco plants. Plant Physiol. 135: 254–265

Bowers, J.E., Abbey, C., Anderson, S., Chang, C., Draye, X., et al. 2003. A high-density genetic recombination map of sequence-tagged sites for sorghum, as a framework for comparative structural and evolutionary genomics of tropical grains and grasses. Genetics, 165:367-386

Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y., and Buckler, E.S. 2007. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics, 23:2633-2635

Brown, P.J., Klein, P.E., Bortiri, E., Acharya, C.B., Rooney, W.L., and Kresovich, S. 2006. Inheritance of inflorescence architecture in sorghum. Theor. Appl. Genet. 113:931–942

Brown, P.J., Rooney, W.L., Franks, C., and Kresovich, S. 2008. Efficient mapping of plant height quantitative trait loci in a sorghum association population with introgressed dwarfing genes. Genetics, 180:629-637

Busov, V.B., Meilan, R., Pearce, D.W., Ma, C., Rood, S.B., and Strauss, S.H. 2003. Activation tagging of a dominant gibberellin catabolism gene (*GA 2-oxidase*) from poplar that regulates tree stature. Plant Physiol. 132: 1283–1291

Carpita, N.C., and McCann, M.C. 2008. Maize and sorghum: genetic resources for bioenergy grasses. Trends Plant Sci. 13:415-420

Chittenden, L.M., Schertz, K.F., Lin, Y.R., Wing, R.A., and Paterson, A.H. 1994. A detailed RFLP map of *Sorghum bicolor* × *S. propinquum*, suitable for high-density mapping, suggests ancestral duplication of sorghum chromosomes or chromosomal segments. Theor. Appl. Genet. 87:925-933

Collins, A., and Morton, N.E. 1998. Mapping a disease locus by allelic association. Proc. Natl. Acad. Sci. U.S.A. 95:1741-1745

Crasta, O.R., Xu, W.W., Nguyen, H.T., Rosenow, D.T., and Mullet, J. 1999. Mapping of post flowering drought resistance traits in grain sorghum: association between QTLs influencing premature senescence and maturity. Mol. Gen. Genet. 262:579–588

Dayan, J., Schwarzkopf, M., Avni, A., and Aloni, R. 2010. Enhancing plant growth and fiber production by silencing GA 2-oxidase. Plant Biotechnol. J. 8:425-435

Drgon, T., Zhang, P.W., Johnson, C., Walther, D., Hess, J., Nino, M., and Uhl, G.R. 2010. Genome wide association for addiction: Replicated results and comparisons of two analytic approaches. PLoS One, 5(1): e8832

Elangovan, M., Prabhakar, and Reddy, D.C.S. 2007. Characterization and evaluation of sorghum [*Sorghum bicolor* (L.)] germplasm from Karnataka, India. Karnataka J. Agric. Sci. 20:840-842

Endo, M., Mochizuki, N., Suzuki, T., and Nagatani, A. 2007. CRYPTOCHROME2 in vascular bundles regulates flowering in *Arabidopsis*. Plant Cell, 19:84–93

Evanno, G., Regnaut, S., and Goudet, J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. 14:2611–2620

Feltus, F.A., Hart, G.E., Schertz, K.F., Casa, A.M., Kresovich, S., Abraham, S., Klein, P.E., Brown, P.J., and Paterson, A.H. 2006. Alignment of genetic maps and QTLs between inter- and intra-specific sorghum populations. Theor. Appl. Genet. 112:1295–1305

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

Haussmann, B.I., Mahalakshmi, V., Reddy, B.V., Seetharama, N., Hash, C.T., and Geiger, H.H. 2002. QTL mapping of stay-green in two sorghum recombinant inbred populations. Theor. Appl. Genet. 106:133-1342

Kebede, H., Subadhi, P.K., Rosenow, D.T., and Nguyen, H.T. 2001. Quantitative trait loci influencing drought tolerance in grain sorghum (*Sorghum bicolor* L. Moench). Theor. Appl. Genet. 103:266–276

Klein, P.E., Klein, R.R., Cartinhour, S.W., Ulanch, P.E., Dong, J., Obert, J.A., Morshige, D.T., Schlueter, S.D., Childs, K.L., Ale, M., and Mullet, J.E. 2000. A high-throughput AFLP-based method for constructing integrated genetic and physical maps: Progress toward a sorghum genome map. Genome Res. 10: 789-807

Klein, R.R., Rodriguez-Herrera, R., Schlueter, J.A., Klein, P.E., Yu, Z.H., and Rooney, W.L. 2001. Identification of genomic regions that affect grain-mould incidence and other traits of agronomic importance in sorghum. Theor. Appl. Genet. 102:307–319

Klein, R.R., Mullet, J.E., Jordan, D.R., Miller, F.R., Rooney, W.L., Menz, M.A., Franks, C.D., and Klein, P.E. 2008. The Effect of tropical sorghum conversion and inbred development on genome diversity as revealed by high-resolution genotyping. Plant Genome, S1: S12–S26

Kong, L., Dong, L., and Hart, G.E. 2000. Characteristics, linkage-map positions, and allelic differentiation of *Sorghum bicolor* (L.) Moench DNA simple-sequence repeats (SSRs). Theor. Appl. Genet. 101: 438-448

Lam, H.M., Xu, X., Liu, X., Chen, W., Yang, G., Wong, F.L., Li, M.W., He, W., Qin, N., Wang, B., Li J., Jian, M., Wang, J., Shao, G., Wang, J., Sun, S.S., and Zhang, G. 2010. Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. Nat. Genet. 42:1053-1059

Li, G., Siddiqui, H., Teng, Y., Lin, R., Wan, X.Y., Li, J., Lau, O.S., Ouyang, X., Dai, M., Wan, J., Devlin, P.F., Deng, X.W., and Wang, H. 2011. Coordinated transcriptional regulation underlying the circadian clock in *Arabidopsis*. Nat. Cell. Biol. 13:616-622

Lin, Y.R., Schertz, K.F., and Paterson, A.H. 1995. Comparative analysis of QTLs affecting plant height and maturity across the poaceae, in referecence to an interspecific sorghum population. Genetics, 141:391-411

Lind, P.A., Macgregor, S., Vink, J.M., Pergadia, M.L., Hansell, N.K., et al. 2010. A genomewide association study of nicotine and alcohol dependence in Australian and Dutch populations. Twin Res. Hum. Genet. 13:10-29

Liu, R.H., and Meng, J.L. 2003. MapDraw: a Microsoft Excel macro for drawing genetic linkage maps based on given genetic linkage data. Heraditas (Beijing), 25: 317-321

Mace, E.S., and Jordan, D.R. 2011. Integrating sorghum whole genome sequence information with a compendium of sorghum QTL studies reveals uneven distribution of QTL and of gene-rich regions with significant implications for crop improvement. Theor. Appl. Genet. 123:169-191

Makanda, I. 2009. Combining ability and heterosis for stem sugar traits and grain yield components in dual-purpose sorghum (*Sorghum bicolor* L. Moench) germplasm. PhD Thesis. University of KwaZulu-Natal, South Africa

Mazzucato, A., Papa, R., Bitocchi, E., Mosconi, P., Nanni, L., Negri, V., Picarella, M.E., Siligato, F., Soressi, G.P., Tiranti, B., and Veronesi, F. 2008. Genetic diversity, structure and marker-trait associations in a collection of Italian tomato (*Solanum lycopersicum* L.) landraces. Theor. Appl. Genet. 116:657-669

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

Murphy, R.L., Klein, R.R., Morishige, D.T., Brady, J.A., Rooney, W.L., Miller, F.R., Dugas, D.V., Klein, P.E., and Mullet, J.E. 2011. Coincident light and clock regulation of *pseudoresponse regulator protein 37* (*PRR37*) controls photoperiodic flowering in sorghum. Proc. Natl. Acad. Sci. USA. 108:16469-16474

Murray, S.C., Sharma, A., Rooney, W.L., Klein, P.E., Mullet, J.E., Mitchell, S.E., and Kresovich, S. 2008a. Genetic improvement of sorghum as a biofuel feedstock: I. QTL for stem sugar and grain nonstructural carbohydrates. Crop Sci. 48:2165–2179

Murray, S.C., Rooney, W.L., Mitchell, S.E., Sharma, A., Klein, P.E., Mullet, J.E., and Kresovich, S. 2008b. Genetic improvement of sorghum as a biofuel feedstock: II. QTL for stem and leaf structural carbohydrates. Crop Sci. 48:2180–2193

Murray, S.C., Rooney, W.L., Mitchell, S.E., and Kresovich, S. 2009. Sweet sorghum diversity and association mapping for Brix and height. Plant Genome, 2:48-62

Nielsen, D.A., Ji, F., Yuferov, V., Ho, A., He, C., Ott, J., and Kreek, M.J. 2010. Genome-wide association study identifies genes that may contribute to risk for developing heroin addiction. Psychiatr. Genet. 20:207-214

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

Paterson, A.H., Lin, Y.R., Li, Z., Schertz, K.F., Doebley, J.F., Pinson, S.R.M., Liu, S.C., Stansel, J.W., and Irvine, J.E. 1995. Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. Science, 269: 1714-1717

Peng, Y., Schertz, K.F., Cartinhour, S., and Hart, G.E. 1999. Comparative genome mapping of *Sorghum bicolor* (L.) Moench using an RFLP map constructed in a population of recombinant inbred lines. Plant Breed. 118: 225-235

Pereira, M.G., and Lee, M. 1995. Identification of genomic regions affecting plant height in sorghum and maize. Theor. Appl. Genet. 90:380-388

Pritchard, J.K., Stephens, M., and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. Genetics, 155:945–959

Quinby, J.R., and Karper, R.E. 1954. Inheritance of height in sorghum. Agron. J. 46:211–216

Ramu, P., Deshpande, S.P., Senthilvel, S., Jayashree, B., Billot, C., Deu, M., Ananda Reddy, L., and Hash, C.T. 2010. *In silico* mapping of important genes and markers available in the public domain for efficient sorghum breeding. Mol. Breed. 26:409-418

Ritter, K.B., Jordan, D.R., Chapman, S.C., Godwin, I.D., Mace, E.S., and McIntyre, C.L. 2008. Identification of QTL for sugar-related traits in a sweet × grain sorghum (*Sorghum bicolor* L. Moench) recombinant inbred population. Mol. Breed. 22:367–384

Rooney, W. 2004. Sorghum improvement--integrating traditional and new technology to produce improved genotypes. Adv. Agron. 83: 37-109

Salas Fernandez, M.G., Becraft, P.W., Yin, Y., and Lübberstedt, T. 2009. From dwarves to giants? Plant height manipulation for biomass yield. Trends Plant Sci. 14:454-461

Sham, P., Bader, J.S., Craig, I., O'Donovan, M., and Owen, M. 2002. DNA pooling: A tool for large scale association studies. Nat. Rev. Genet. 3: 862-871

Shiringani, A.L., Frisch, M., and Friedt, W. 2010. Genetic mapping of QTLs for sugar-related traits in a RIL population of *Sorghum bicolor* L. Moench. Theor. Appl. Genet. 121:323–336

Srinivas, G., Satish, K., Madhusudhana, R., Reddy, R.N., Mohan, S.M., and Seetharama, N. 2009. Identification of quantitative trait loci for agronomically important traits and their association with genic-microsatellite markers in sorghum. Theor. Appl. Genet. 118:1439-1454

Sun, Y., Wang, J., Crouch, J.H., and Xu, Y. 2010. Efficiency of selective genotyping for genetic analysis of complex traits and potential applications in crop improvement. Mol. Breed. 26:493–511

Upadhyaya, H.D., Pundir, R.P.S., Dwivedi, S.L., Gowda, C.L.L., Reddy, V.G., and Singh, S. 2009. Developing a mini core collection of sorghum for diversified utilization of germplasm. Crop Sci. 49: 1769–1780

Vermerris, W. 2008. Genetic improvement of bioenergy crops. Springer, New York.

Vermerris, W., Sherman, D.M., and McIntyre, L.M. 2010. Phenotypic plasticity in cell walls of maize *brown midrib* mutants is limited by lignin composition. J. Exp. Bot. 61:2479-2490

Wang, Y.H., Bible, P., Loganantharaj, R., and Upadhyaya, H.D. 2012. Identification of SSR markers associated with height using pool-based genome-wide association mapping in sorghum. Mol. Breed. DOI: 10.1007/s11032-011-9617-3

Wang, Y.H., Poudel, D.D., and Hasenstein, K.H. 2011. Identification of simple sequence repeat (SSR) markers associated with saccharification yield using pool-based genome-wide association mapping in sorghum. Genome, 54:883-889

Williams, C.E., and Ronald, P.C. 1994. PCR template-DNA isolated quickly from monocot and dicot leaves without tissue homogenization. Nucleic Acids Res. 22 1917–1918

Xu, Y., and Crouch, J.H. 2008. Marker-assisted selection in plant breeding: from publications to practice. Crop Sci. 48:391–407

Yu, J., and Buckler, E.S. 2006. Genetic association mapping and genome organization of maize. Curr. Opin. Biotechnol. 17:155–160

Yu, J.M., Pressoir, G., Briggs, W.H., Bi, I.V., Yamasaki, M., Doebley, J.F., McMullen, M.D., Gaut, B.S., Nielsen, D.M., Holland, J.B., Kresovich, S., and Buckler, E.S. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. 38:203-208

Zhao, Y.L., Dolat, A., Steinberger, Y., Wang, X., Osman, A., and Xie, G.H. 2009. Biomass yield and changes in chemical composition of sweet sorghum cultivars grown for biofuel. Field Crops Res. 111:55–64

Figure 1. Chromosome location of markers linked to height/maturity in sorghum. For each chromosome, on the left are physical positions of markers/genes in Mb and on the right are SSR, RFLP markers and candidate genes. Markers previously mapped to height/maturity and genes/QTL known to control height/maturity in sorghum are cited with sources in parenthesis. H/M in parenthesis: H---marker/gene linked to height; M--- marker/gene linked to maturity; H/M--- marker/gene linked to both. Open/closed bars to the right of the chromosomes represent mapped maturity/height QTLs from previous studies. Not drawn to scale relative to the whole chromosome but to scale for the physical positions shown.

Table 1. Pearson's correlation in sorghum plant height and maturity among the five environments [†].

	2008R	2009R	2010R	2005PR	2009PR
2008R		0.919***	0.834***	0.499	0.570***
2009R	0.959***		0.830***	0.485	0.514***
2010R	0.928***	0.954***		0.476	0.590***
2005PR	0.638***	0.627***	0.677***		0.788***
2009PR	0.417***	0.383***	0.413***	0.648	

[†] Top right are plant height and lower left are maturity. *** indicates significance at 0.001 probability level in *t* test.

Table 2. Pearson's correlation between sorghum plant height and days to 50% flowering in different environments.

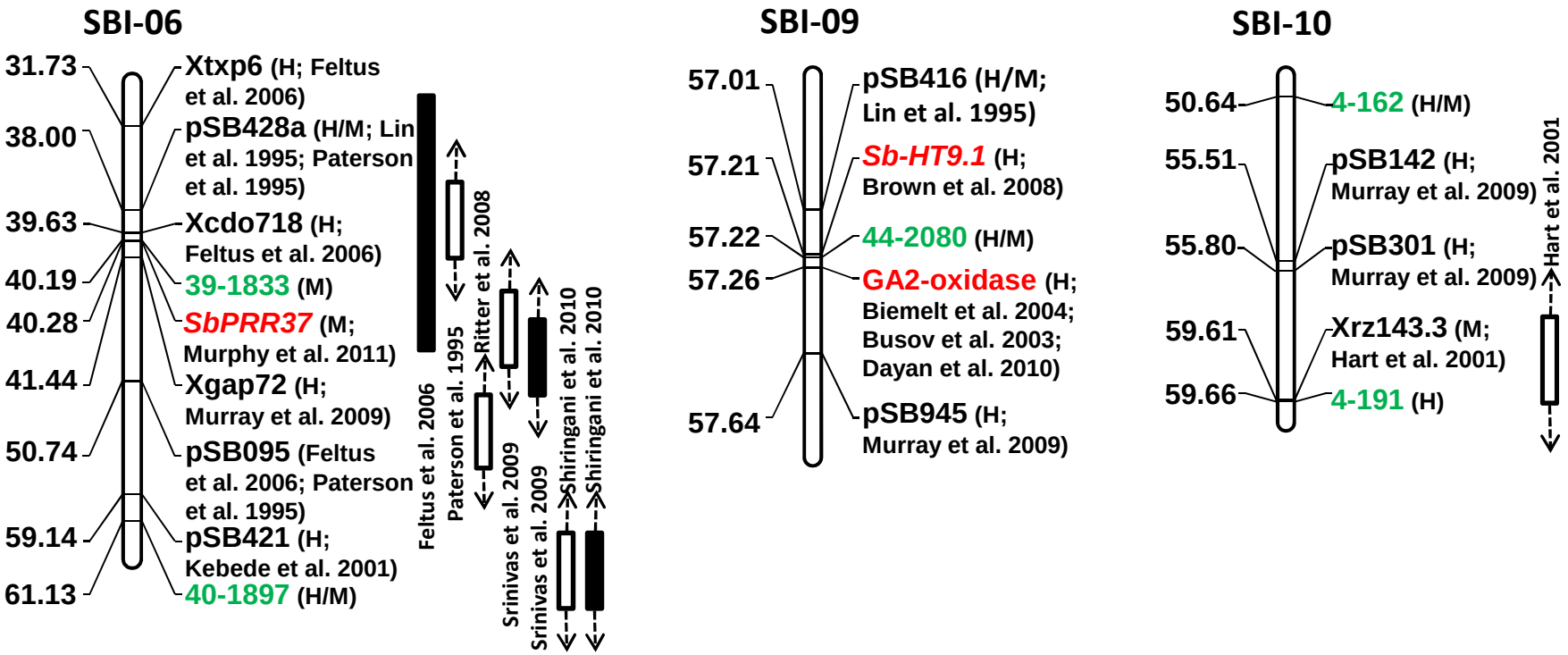
	2008R	2009R	2010R	2005PR	2009PR
Correlation coefficient	0.712***	0.798***	0.595***	0.653***	0.656***

*** indicates significance at 0.001 probability level in *t* test.

Table 3. SSR markers associated with sorghum height (H) and maturity (M) in five environments [†].

Marker	Chromosome number	Trait-environment/association *p*-value
39-1833	6	9RM/**0.0023**, 8RM/**6.19E-05**, 10RM/**0.0081**, 10RH/**0.0062**, 8RH/0.0362, 9RH/0.0133
40-1897	6	9RM/**0.0031**, 9RH/**0.0012**, 10RH/0.0212, 5PRH/0.0491
44-2080	9	9RH/0.0262, 10RH/**0.0094**
4-162	10	5PRM/0.0196, 8RM/**7.22E-04**, 9RM/0.0181, 10RM/**2.69E-04**, 5PRH/**0.0039**, 9PRH/0.049, 9RH/0.0116
4-191	10	5PRH/**0.0039**, 9RH/0.0204, 10RH/0.0420

[†] 8R, 9R, 10R, 5PR, and 9PR are abbreviations for 2008R, 2009R, 2010R, 2005PR, and 2009PR, respectively. *p*-values in bold are significant at 0.01 level.


SBI-06
31.73
38.00
39.63
40.19
40.28
41.44
50.74
59.14
61.13
Xtxp6 (H; Feltus et al. 2006)
pSB428a (H/M; Lin et al. 1995; Paterson et al. 1995)
Xcdo718 (H; Feltus et al. 2006)
39-1833 (M)
SbPRR37 (M; Murphy et al. 2011)
Xgap72 (H; Murray et al. 2009)
pSB095 (Feltus et al. 2006; Paterson et al. 1995)
pSB421 (H; Kebede et al. 2001)
40-1897 (H/M)
Feltus et al. 2006
Paterson et al. 1995
Ritter et al. 2008
Srinivas et al. 2009
Srinivas et al. 2009
Shiringani et al. 2010
Shiringani et al. 2010
SBI-09
57.01
57.21
57.22
57.26
57.64
pSB416 (H/M; Lin et al. 1995)
Sb-HT9.1 (H; Brown et al. 2008)
44-2080 (H/M)
GA2-oxidase (H; Biemelt et al. 2004; Busov et al. 2003; Dayan et al. 2010)
pSB945 (H; Murray et al. 2009)
SBI-10
50.64
55.51
55.80
59.61
59.66
4-162 (H/M)
pSB142 (H; Murray et al. 2009)
pSB301 (H; Murray et al. 2009)
Xrz143.3 (M; Hart et al. 2001)
4-191 (H)
Hart et al. 2001