# Identification of SSR markers associated with height using pool-based genome-wide association mapping in sorghum

Yi-Hong Wang · Paul Bible · Rasiah Loganantharaj · Hari D. Upadhyaya

Department of Biology, University of Louisiana, Laboratory of Bioinformatics, Center for Advanced, Computer Studies, University of Louisiana , Gene Bank, International Crops Research Institute for the Semi Arid Tropics (ICRISAT), Patancheru

**Molecular Breeding**
**Vol.** 30, **Issue**. 1, **Pages** 281-292, 2012
**DOI**: http://dx.doi.org/10.1007/s11032-011-9617-3



---

**Identification of SSR markers associated with height using pool-based genome-wide association mapping in sorghum**

Yi-Hong Wang*[1], Paul Bible[2], Rasiah Loganantharaj[2], Hari D Upadhyaya[3]

*Corresponding author.
[1] Department of Biology
University of Louisiana at Lafayette
Lafayette, LA 70504
USA
yxw9887@louisiana.edu
Phone: +1 337 482-6643
Fax: +1 337 482-5395

[2] Laboratory of Bioinformatics
Center for Advanced Computer Studies
University of Louisiana at Lafayette
Lafayette, LA 70504
USA

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

**Abstract**

Sorghum has been proposed as a potential energy crop. However, it has been traditionally bred for grain yield and forage quality, not traits related to bioenergy production. To develop tools for genetic improvement of bioenergy-related traits such as height, genetic markers associated with these traits have to be identified first. Association mapping has been extensively used in humans and in some crop plants for this purpose. However, genome-wide association mapping using the whole association panel is costly and time-consuming. A variation of this method called pool-based genome-wide association mapping has been extensively used in humans. In this variation, pools of individuals with contrasting phenotypes, instead of the whole panel, are screened with genetic markers and polymorphic markers are confirmed by screening the individuals in the pools. Here we identified several new simple sequence repeats (SSR) markers associated with height using this pool-based genome-wide association mapping in sorghum. After screening the tall and short pools of sorghum accessions from the sorghum mini core collection developed at the International Crops Research Institute for the Semi-Arid Tropics with 703 SSR markers, we have identified four markers that are closely associated with sorghum height on chromosomes 2, 6, and 9. Comparison with published maps indicates that all four markers are clustered with markers previously mapped to height or height-related traits and with candidate genes involved in regulating plant height such as *FtsZ*, *Ugt*, and *GA 2-oxidase*. The mapping method can be applied to other crop plants for which a high throughput genome-wide association mapping platform is not yet available.

**Keywords**: sorghum, pool-based genome-wide association mapping, SSR markers, molecular breeding, height

## Introduction

Sorghum (*Sorghum bicolor* (L.) Moench) has been proposed as a model crop to study C4 grasses and as a potential energy crop. It is a recognized biomass crop suitable for biofuel production because of two critical factors: high biomass production and efficient water use (Rooney et al. 2007; Saballos 2008). Sorghum biomass productivity is positively correlated with stem height and thickness (Murray et al. 2008; Zhao et al. 2009). In fact, stem is the most important organ for cellulose production and is a strong sink for soluble sugars in sorghum (Zhao et al. 2009); both can be used for bioethanol production. In addition to producing more biomass, taller varieties have other advantages despite the fact that they may mature late and some may be prone to lodging (not always; see Murray et al. 2009 and Rooney 2004): they tend to have a thicker stem with higher sugar content (Ritter et al. 2008). In addition, they may produce a better ratoon crop because height is significantly correlated with regrowth of fresh biomass with a coefficient of 0.71 (Murray et al. 2008). Height is also independent of stem structural composition, i.e., cellulose, hemicellulose and lignin content (Murray et al. 2008). This means that a variety with tall stem can be bred to contain more cellulose and less lignin. Because height of the stem defines biomass productivity of sorghum, a QTL for total dry biomass has been found to colocalize with height QTLs on chromosomes 7 and 9 (Brown et al. 2008; Murray et al. 2008). The first step to genetically improve sorghum as an energy crop is to increase its biomass yield (Murray et al. 2008). To achieve this goal, it is imperative to identify all the genes underlying the stem height and markers that can be used to select for height through marker-assisted selection.

The work has long begun: genetic studies have shown that height in sorghum is controlled only by a few major QTLs (reviewed by Salas Fernandez et al. 2009). Because of this, numerous studies have focused on height as an indirect way to identify genes related to biomass yield. Earlier genetic study has identified four loci affecting sorghum stem height: *Dw1*, *Dw2*, *Dw3* and *Dw4* (Quinby and Karper 1954). Only *Dw3* 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). *Dw3* also co-localizes with a height QTL on chromosome 7 (Brown et al. 2006) and *Dw2* is associated with a QTL on chromosome 6 (Feltus et al. 2006; Klein et al. 2008). Although *Dw1* and *Dw4* have not been mapped conclusively to a linkage group, another QTL for height has been genetically mapped on chromosome 9 (Brown et al. 2008).

Mapping of sorghum stem height was also conducted using association mapping technique which maps QTLs based on linkage disequilibrium defined as the non-random association of alleles between genetic loci (Yu and Buckler 2006). Using this method, a height QTL (Sb-HT9.1) was mapped on chromosome 9 using a panel of 377 sorghum genotypes and 49 SSR markers (Brown et al. 2008). Three significant associations for height were also found in a panel of 125 genotypes using 47 SSR and 322

SNP markers and two of them are on chromosomes 6 and 9 (Murray et al. 2009), which supports published QTL studies (Brown et al. 2006; Feltus et al. 2006; Lin et al. 1995; Pereira and Lee 1995).

A variation of the above association mapping uses a pool-based two-stage design (Sham et al. 2002). In the first stage, individuals with contrasting phenotypes (e.g., tall/short) will form two pools. Individuals in the tall pool will all be tall; individuals in the short pool will all be short. This is similar in concept to bulked segregant analysis (Giovannoni et al. 1991; Michelmore et al. 1991) except that individuals in a pool are not descendents from the same parents. DNA is then isolated from each individual, quantified, and bulked into tall or short pools. Both pools of DNA are screened with markers that cover the whole genome. Any difference in marker pattern between the two pools will be recognized as polymorphism and such markers are repeated in all individuals in both pools to ensure that it is polymorphic between individuals from both pools. This confirmatory individual genotyping is the second stage of this technique which may also be extended to other individuals not in the initial pools.

The method has been used in mapping complex and heritable human conditions which are also strongly influenced by the environment. One example is cocaine addiction in twin studies. When one pool of control (normal population) and one pool of case (consists of people with addiction) were screened with markers, three chromosome regions were identified as associated with the addiction. These regions contain a number of genes that encode potential targets for anti-addiction pharmacotherapeutics (Drgon et al. 2010). Using essentially the same technique, Lind et al. (2010) identified the *ARHGAP10* gene that is associated with nicotine dependence. Addiction to heroin shows even more environmental influence because of its low heritability. But this method identified one strong association in Caucasians with rs10494334, a marker in an unannotated region of the human genome. In African Americans, the marker most significantly associated with the heroin addiction vulnerability was found in the cytosolic dual specificity phosphatase 27 gene *DUSP27* (Nielsen et al. 2010). These results indicate that markers associated with quantitative traits can be identified in genomewide screening using the pool-based association mapping and these markers can be a powerful tool in identifying these genes and in marker-assisted breeding to genetically improve sorghum as a bioenergy crop.

The most commonly used marker system for association mapping in sorghum is simple sequence repeats (SSR) (Brown et al. 2008; Casa et al. 2008; Murray et al. 2009; Shehzad 2009a). Although the more abundant single nucleotide polymorphism (SNP) markers have been developed for a number of model crop plants such as rice and maize, due to SNP's high development/detection cost, SSR remain an attractive marker system for sorghum (Wang et al. 2009a). Compared to SNP marker system, SSR markers requires just the primer sequences which are usually available once published. This is the primary reason why so much effort has been devoted to developing SSR markers (Kong et al. 2000; Li et

al 2009; Ramu et al. 2009; Srinivas et al. 2009; Yonemaru et al. 2009) and to using SSR markers in genetic mapping (Bhattramakki et al. 2000; Haussmann et al. 2002; Wu and Huang 2006), diversity studies (Abu Assar et al. 2005; Agrama and Tuinstra 2003; Ali et al. 2008; Anas and Yoshida 2004; Barnaud et al. 2007; Caniato et al. 2007; Casa et al. 2005; Deu et al. 2008; Dillon et al. 2005; Folkertsma et al. 2005; Ghebru et al. 2002; Menz et al. 2004; Pei et al. 2010; Shehzad et al. 2009b; Smith et al. 2000; Wang et al. 2009b) and molecular breeding (Knoll and Ejeta 2008) in sorghum. The objective of this study was to identify SSR markers associated with sorghum stem height using the aforementioned pool-based genome-wide association mapping technique. This mapping method has one important advantage: since only part of the population (association panel) is used for genotyping, cost is significantly reduced, a boon for resource-limited investigators. Related to this, time is also saved because not all individuals are genotyped by all markers.

## Materials and methods

### Plant materials and phenotyping

The ICRISAT (International Crops Research Institute for the Semi-Arid Tropics) height data was from Upadhyaya et al. (2009) and this will be referred to as the ICRISAT dataset. The mini core collection (10% of core, 1% of entire collection representing diversity) of sorghum was developed from over 37,000 germplasms curated at ICRISAT in India by Upadhyaya et al. (2009) following Upadhyaya and Ortiz (2001). The collection was introduced and grown in greenhouse at UL Lafayette and this will be referred to as the UL Lafayette dataset. Height of four plants was measured at harvest for each accession in the collection and the average for each accession constitutes the UL Lafayette height data. Due to variable flowering time, plants were harvested when the grain became mature. Among 216 accessions planted, 78 did not flower within five months and did not produce seeds. These plants were harvested and measured at the end of five months.

### DNA isolation and construction of pools

After two months' growth, top two leaves from one plant for each 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 slight modifications. Briefly, dried leaf tissues was cut into small pieces (about 2 $mm^2$) and filled into a 1.5 ml microcentrifuge tube. Extraction buffer (2x700 µl) containing 100 mM Tris pH 8.0, 10 mM EDTA pH 8.0, 700 mM NaCl, 12.5 mM potassium ethyl xanthogenate (PEX) was added. Samples were incubated at 65° C for 30 min with occasional mixing in a dry heat block. After incubation, samples were centrifuged at 15000 g for 5 min and 700 µl supernatant was transferred to a new 1.5-ml microcentrifuge 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 before centrifugation at 15000 g for 5 min. The supernatant containing DNA was transferred to a new tube and quantified in NanoDrop 2000 (Thermo Scientific, Waltham, MA). For PCR reaction, DNA concentration was adjusted to 40 ng/µl. DNA quality from four random samples were checked on 1% agarose gel.

To construct pools, accessions from the Mini Core collection were ranked based on height in both environments. In order for a plant to be included in a pool, that plant has to be tall or short in both environments. Based on this criterion, eight tall and eight short accessions were selected. DNA from each accession was quantified and equal amount of DNA was bulked for both tall and short pools. The final bulked DNA concentration in each pool was 80 ng/µl and these were used for SSR screening.

SSR marker screening of pools

To select the SSR markers for genome-wide mapping of genes underlying stem height, the 15,194 SSR markers identified in the genome sequence by Paterson et al. (2009) were screened for suitable primers for PCR amplification using Primer3 (Rozen and Skaletsky 2000). The settings to Primer3 were all default except the optimal length of the primer was set to 24 and the product size range was set to between 100 to 500 bp. In total, 13,397 were found to have suitable primers and can be used for marker amplification with product range between 100-500 bp. From this number, 703 SSRs were randomly selected for this study and primers were synthesized by Eurofins MWG Operon (Huntsville, AL). SSR markers are named with two numbers separated by a dash. The first number is the plate number and the second number is the sequence number. Primer sequences are available in Supplementary Table 1. Genome-wide distribution of these markers relative to genes is presented in Fig. 1.

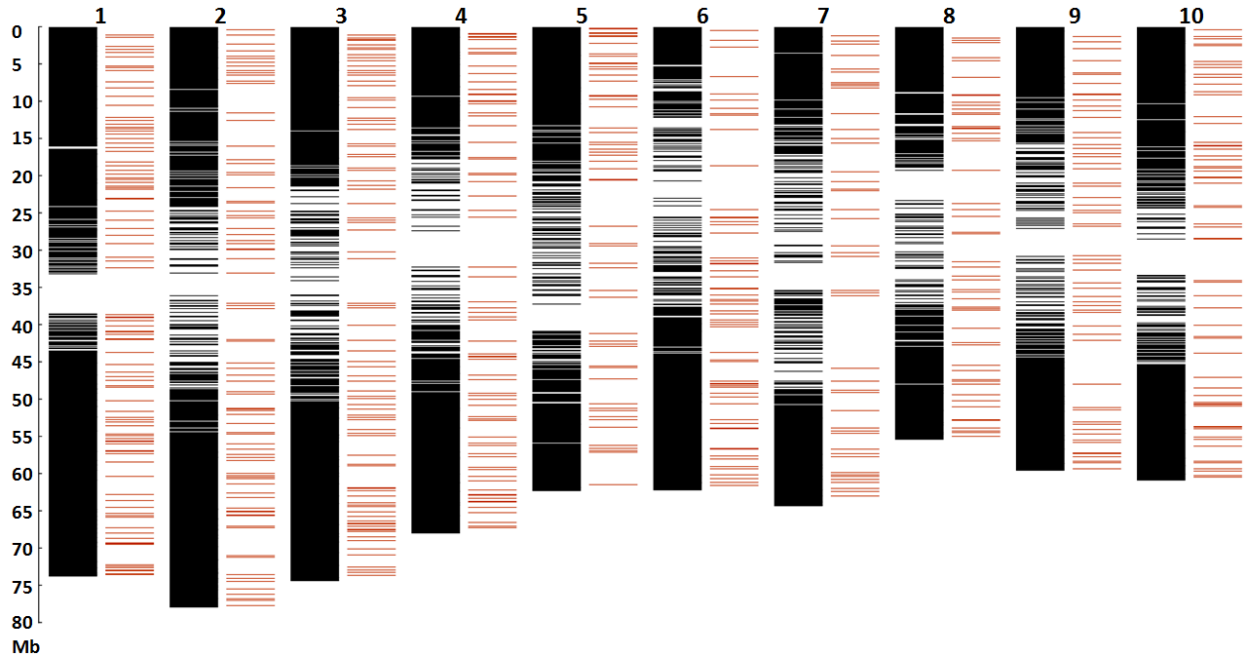



**Fig. 1** Distribution of 703 SSR markers relative to genes in the sorghum genome. For each chromosome numbered on top, Column 1 (black) represents genes; Column 2 (red) represents 703 SSR markers used in this study

The tall and short pools were screened with the 703 SSR markers which translates to 1,406 PCR reactions for the screening alone. PCR was performed in a 10 µl reaction volume that contains 5 µl 2x 360 AmpliTaq Master Mix from Applied Biosystems (ABI, Carlsbad, CA), 10 ng each of the two primers and 80 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 separated in LabChip Microfluidic Gel Electrophoresis System (LabChip 90) from Caliper Life Sciences (Hopkinton, MA) with 1 K DNA Assay Kit according 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 16 accessions that make up the two pools according to Sham et al. (2002). Genotyping of all 216 accessions with selected SSR markers was performed following the same protocols.

Association mapping

Only SSRs identified as polymorphic between the two tall and short accessions were used in association testing. Marker-trait associations were calculated using general linear models (GLM; Yu et al. 2006) as implemented in TASSEL (Bradbury et al. 2007; http://www.maizegenetics.net/) 3.0 controlled for population structure generated from STRUCTURE (Pritchard et al. 2000;

http://pritch.bsd.uchicago.edu/structure.html) 2.3. GLM was applied with the number of groups/subpopulations ($k$) set to five when running STRUCTURE, as this number maximizes the $\Delta k$, a parameter used to detect real number of groups (Evanno et al. 2005). When $\Delta k = 5$ it still allows the detection of the real number of groups with just five SSRs or 50 AFLPs as demonstrated in the simulation study (Evanno et al. 2005), in tomato association mapping (Mazzucato et al. 2008) and soybean population structure study (Lam et al. 2010). STRUCTURE was run with the admixture model, a burn-in period of 10,000 and 1000 Markov Chain Monte Carlo repetitions, and the data from 10 unlinked polymorphic SSR markers. The following equation was used for GLM analysis:

$$y = X\alpha + Q_{k=5} + e$$

where $y$ represents phenotype, $X$ represents genotype, $X\alpha$ represents fixed effects, $Q$ is population structure cofactor, and $e$ represents random effects (Bradbury et al. 2007; TASSEL 3.0 User Guide). GLM method controlled for population structure showed smallest possibility of false negatives among single trait-marker association models tested and comparable number of false positives as MLM controlled for both population structure and kinship in sorghum (Shehzad et al. 2009a). Significance of associations between loci and traits was based on an $F$-test with $p$ values calculated by TASSEL.

Marker localization

Identified SSR markers potentially associated with height were localized to the genome based on Paterson et al. (2009). Markers previously identified as linked to height 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/). Marker sequences 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/. Maps that also identified markers linked to height but for which sequencing information of these markers were not available were not used in this study. Maps used were those from Chittenden et al. (1994), Bhattramakki et al. (2000), and Bowers et al. (2003).

## Results and discussion

Phenotyping and screening of the pools

Only 216 of the original 242 accessions were grown at UL Lafayette due to quarantine reasons. Since heritability for height is high, ranging from 0.83 (Murray et al. 2008) to 0.95 (Ritter et al. 2008), we

measured height of plants grown in an enclosed environment for the UL Lafayette dataset. When this dataset was compared to the ICRISAT dataset (Upadhyaya et al. 2009), Pearson's correlation coefficient between the two datasets is 0.3328 ($p$ = 0.0000001), statistically significant yet not very high. Despite this relatively low correlation, we were able to select accessions that were relatively tall or short in both environments (Table 1 and Supplementary Table 2). Interestingly, the average stem height of the panel of 216 accessions was 2.37 meters at ICRISAT and 3.02 meters at UL Lafayette. This is because 78 of the 216 accessions from the Mini Core collection did not flower at UL Lafayette, therefore, they grew taller while all accessions flowered and produced seed at ICRISAT.

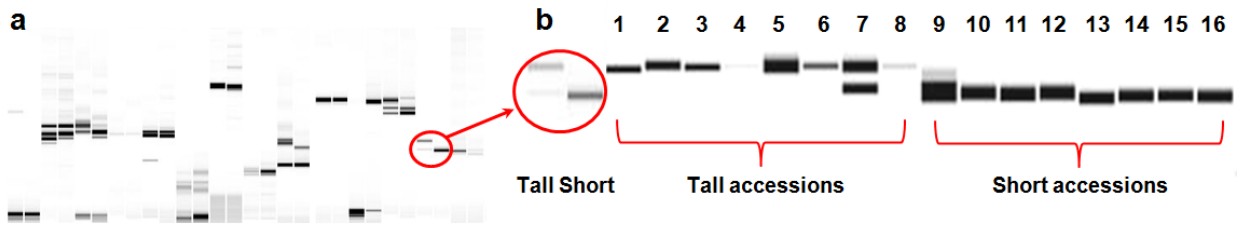
To construct the tall and short pools, the accessions were ranked based on their height in each environment. Eight tall and eight short accessions were each chosen because their heights were relatively consistent in both environments (Table 1). To show the difference in average height between the selected accessions and the whole panel, the average height in the tall/short pools was also calculated. The average height among all accessions in the panel was 3.02 m while the average height among the eight accessions in the tall pool was 4.40 m and that of short pool was only 1.44 m for the UL Lafayette dataset. For the ICRISAT dataset, the panel average height was 2.37 m while the averages for tall and short pools were 3.05 m and 1.50 m, respectively.

**Table 1** Sorghum accessions with tall/short phenotypes (in meters) based on evaluations at UL Lafayette and ICRISAT.

| Tall | | | Short | | |
|---|---|---|---|---|---|
| Accession No. | UL Lafayette | ICRISAT | Accession No. | UL Lafayette | ICRISAT |
| IS-31706 | 5.16 | 2.91 | IS-3158 | 1.11 | 1.29 |
| IS-11026 | 4.85 | 3.02 | IS-19262 | 1.24 | 1.19 |
| IS-21083 | 4.84 | 2.86 | IS-24462 | 1.37 | 1.18 |
| IS-23891 | 4.27 | 2.78 | IS-26749 | 1.40 | 1.49 |
| IS-24175 | 4.06 | 3.00 | IS-27697 | 1.52 | 2.10 |
| IS-26025 | 4.05 | 2.78 | IS-14090 | 1.53 | 1.42 |
| IS-24953 | 3.99 | 3.08 | IS-2397 | 1.68 | 1.76 |
| IS-25301 | 3.98 | 3.93 | IS-29582 | 1.72 | 1.60 |
| Pool average[a] | 4.4 | 3.05 | Pool average | 1.44 | 1.50 |
| Panel average[a] | 3.02 | 2.37 | | | |

[a] Pool average was the average height of the eight tall/short accessions. Panel average was the average height of all 216 accessions.

The tall and short pools (Table 1) were screened with 703 SSR markers covering the whole genome (Fig. 1). Approximately 97% of the markers produced scorable band. Among these, 67 showed potential polymorphisms between the pools. But when all the 67 markers were used to genotype all the 16 accessions in the pools individually, only six of them were truly polymorphic and two of the six were not completely polymorphic between the tall and short accessions. The high false positive rate was due to the



high sensitivity of LabChip 90 in detecting SSR markers. Only these four SSR markers, 37-1740, 40-1897, 23-1062, and 44-2080, were analyzed further. These SSRs contain 2-5 alleles polymorphic between the pools. Initial and confirmatory screening of 40-1897 is shown in Fig. 2 which identified accession IS-24953 as heterozygous in this locus while the other 15 accessions were homozygous. Although self-pollinated, sorghum does have relatively high out-crossing rate (Saballos 2008). Similarly, Murray et al. (2009) found 77 of 125 sorghum genotypes tested were heterozygous in one or more marker loci. This is probably the reason that oftentimes a single plant, which represents most of the plants within an accession morphologically, is sampled for analysis (Wang et al. 2009b).

**Fig. 2** Screening of tall and short sorghum pools with SSR markers and confirmation of a polymorphic marker using a LabChip 90. **a**. A partial gel image of the initial marker screening showing an SSR polymorphism between tall (first lane inside the circle) and short (second lane inside the circle) pools as circled. **b**. Confirmation of the polymorphic marker in all the accessions in the pools. The product size for short accessions is about 200 bp and that of tall accessions is 233 bp. One tall accession (Lane 7-IS-24953) is heterozygous as reflected by the pool result as circled. Accessions: Lane 1-IS-31706; Lane 2-IS-11026; Lane 3- IS-21083; Lane 4-IS-23891; Lane 5-IS-24175; Lane 6-IS-26025; Lane 7-IS-24953; Lane 8-IS-25301; Lane 9-IS-3158; Lane 10-IS-19262; Lane 11-IS-24462; Lane 12-IS-26749; Lane 13-IS-27697; Lane 14-IS-14090; Lane 15-IS-2397; Lane 16-IS-29582

Association mapping

Association of the four SSR markers was tested using TASSEL 3.0 to show markers identified using the pools were truly associated with height. The strength of association was much stronger at ICRISAT than at UL Lafayette (Table 2) probably due to the fact that the environment (photoperiod) at UL Lafayette was not conducive to reproductive growth for some accessions. For example, one third of the plants just grew vegetatively and did not produce any seeds. Despite this, association of 40-1897 and 44-2080 with height was significant at 5% level and marker effect of 40-1897 explains 13.9% of the variance in height. But at ICRISAT, association of all four markers was significant at 1% level (Table 2). Common to both environments, 40-1897 and 23-1062 both had stronger effect on height than the other two markers (Table 2).

**Table 2** Association between identified SSR markers and sorghum height in two environments.

| Marker | ICRISAT | | UL Lafayette | |
|---|---|---|---|---|
| | *p* value | % variance explained | *p* value | % variance explained |
| 37-1740 | $7 \times 10^{-11}$ | 19.4 | 0.075 | 1.5 |
| 40-1897 | $6 \times 10^{-7}$ | 26.5 | 0.022 | 13.9 |
| 23-1062 | $9 \times 10^{-12}$ | 34.6 | 0.062 | 10.7 |
| 44-2080 | $2 \times 10^{-8}$ | 14.8 | 0.003 | 4.3 |

Marker alleles with strong effect on height are presented in Fig. 3. In most cases, presence of a marker allele reduced the height and the trend was similar in both environments, except the alleles 228

and 200 of 40-1897. The 232 allele of 40-1897 increased height in both environments while alleles of 228 and 200 both reduced height at ICRISAT compared to an accession homozygous for the 224 allele. For marker 23-1062, presence of allele 244 in 29 accessions or allele 237 in 38 accessions reduced height by over 2 meters (Fig. 3) compared to an accession homozygous for the allele 264. The stronger marker effect on height explains at least partially the more significant association for the ICRISAT dataset.

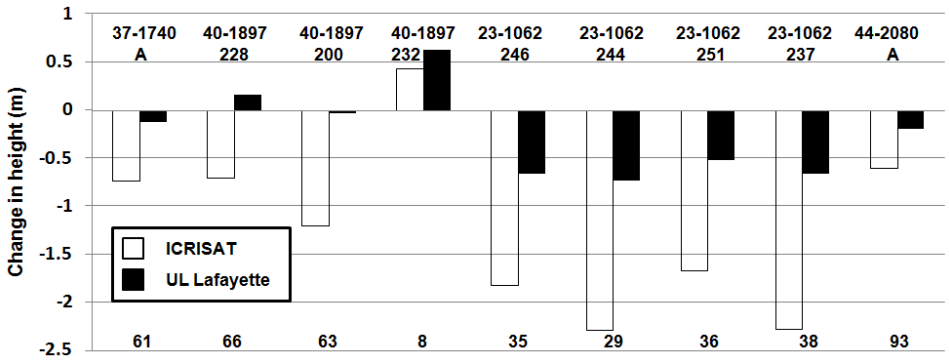



**Fig. 3** Phenotypic effect of marker alleles (in the top) at loci associated with sorghum height in two environments. Numbers in the bottom are the number of accessions with the marker allele listed on the top. The "A" alleles for 37-1740 and 44-2080 represent the presence of a marker fragment/band. All other alleles in the figure are fragment sizes in bp

Marker localization

To further demonstrate that markers identified in pool-based genome-wide association mapping were physically close to markers previous found to be genetically linked to height and height-related traits, we localized the four markers relative to the position of the linked markers on the sorghum chromosomes. The results are presented in Fig. 4. The four SSR markers were localized to three chromosomes: 2, 6 and 9. The two markers that display more significant effect on height, 23-1062 and 40-1897, were on chromosomes 2 and 6, respectively. Chromosome 2 was not as extensively mapped as chromosomes 6 and 9. But we found that 23-1062 was 385 kb from *Ugt* (UDP-glycosyltransferase/indole-3-acetate beta-glucosyltransferase). It has been shown that suppression of this gene by the constitutive expression of antisense mRNA greatly reduced the height of alfalfa and pea plants due to a twofold delay in cell cycle progression and inhibition of border cell production (Woo et al. 1999).

Two other homologous genes linked to plant height were also close to markers identified in this study. On chromosome 9, 44-2080 was about 47 kb from a gene encoding GA 2-oxidase which has been shown to regulate plant height in poplar (Busov et al. 2003) and tobacco (Biemelt et al. 2004). In tobacco, overexpression of GA 2-oxidase reduced height by 86% (Biemelt et al. 2004) and in poplar the reduction caused by GA 2-oxidase overexpression was 4-fold (Busov et al. 2003). An RFLP marker (pSB0945) identified by Murray et al. (2009) as tightly associated with sorghum height was 382 kb from GA 2-oxidase and another RFLP marker pSB416 mapped by Feltus et al. (2006) was 255 kb away (Fig. 4). On

chromosome 6, 37-1740 was 1.78 Mb from pSB520b, an RFLP marker found to be linked to sorghum height by Lin et al. (1995). Toward the other end of chromosome 6, 40-1897 was flanked by *Floricaula/leafy-like 2* and *FtsZ* (Fig. 4). *Floricaula/leafy-like 2* was mapped by Brown et al. (2006) as a candidate gene for reproductive transition/inflorescence branching. The maize homologs of this gene direct flower development, inflorescence architecture and the vegetative to reproductive phase transition (Bomblies et al. 2003). The latter could determine plant height if a transition phase is delayed. Despite the overall similar trend in the marker effect on height in the two environments, two of the three alleles (228 and 200) of 40-1897 display a slightly different pattern. While the alleles reduced height at ICRISAT, the effect of 228 was actually positive and that of 200 was close to neutral (Fig. 3). There is a possibility that *Floricaula/leafy-like 2* was involved in delaying flowering at UL Lafayette although this remains to be determined. The other flanking gene, *FtsZ*, was homologous to bacterial cell division protein. Overexpression of its homolog in tobacco reduced height by 50% to 2-fold under high or low light due to decreased chloroplast production (Jeong et al. 2002).

Physical distance in Fig. 4 can also be translated to genetic distance based on data from chromosome 6. Markers Xgap72 and Xtxp265 were identified as tightly associated with height on chromosome 6 by Murray et al. (2009). In Li's (2009) map, Xtxp265 was 10 cM from Xgap72 and 0.4 cM from Xtxp317 (used as reference point here); the respective physical distance was 10 Mb and 0.4 Mb which translate to about 1 Mb/cM. Using this estimate, 23-1062 is 0.39 cM from *Ugt* and is 2.4cM from pSB500, a marker linked to days-to-flowering QTL in chromosome 2 by Lin et al. (1995). On chromosome 6, 40-1897 was 0.44 cM from *FtsZ* and 4.89 cM from *Floricaula/leafy-like 2* and 37-1740 was 1.78 cM from pSB520b, a marker linked to height by Lin et al. (1995). And lastly on chromosome 9, 44-2080 was 0.04 cM from GA2-oxidase and was flanked by pSB416 at 0.21 cM and pSB0945 at 0.42 cM. pSB416 was linked to maturity (Feltus et al. 2006) and pSB0945 was associated with height by association mapping (Murray et al. 2009). Overall, it shows that markers identified in this study were genetically and physically clustered with markers previously identified as linked to height and height-related traits.

Although the function of the aforementioned genes is related to plant height, it remains to be determined whether they were the genes underlying height in sorghum. Increasing the genome coverage with more markers may help. For example, with 703 SSR markers, there is only one marker for every Mb on average. Using 4000 markers, the resolution will increase to one marker for every 160 kb of the genome. This is clearly doable using a high-throughput marker detection system. The LabChip 90 can process 400 PCR samples a day. Assuming screening two pools with 4000 markers, this means 8000 PCR samples to separate which will take the machine 20 days for this initial screening. This increased

coverage will provide a high-resolution physical mapping and will facilitate the ultimate identification of the candidate gene.

Concluding remarks

We have identified four SSR markers associated with sorghum stem height using the pool-based genome-wide association mapping. The identified markers were clustered with those previously identified as linked to height or height-related traits using genetic mapping. Using pools of individuals with contrasting phenotypes instead of the whole association panel significantly reduced genotyping cost and time. Furthermore, raising the number of markers for the initial screening will increase mapping resolution and this will help to pin down causal candidate genes. The increased workload can be accommodated using a high-throughput SSR marker detection system. Markers identified through this method can be effectively used in marker-assisted selection in molecular breeding, as SSR markers are already used in marker-assisted selection for early-season cold tolerance in sorghum (Knoll and Ejeta 2008). And the method may be applied to other crop plants for which a high throughput genome-wide association mapping platform is not yet available.

Genomics and genome-wide study in crop plants have progressed rather rapidly in the last two years. The technological power of the so called next generation sequencing is being harnessed to identify all genes/markers associated with major traits of interest in crop plants. By resequencing 517 rice landraces, Huang et al. (2010) identified 80 strong associations in 14 agronomic traits of rice. In soybean, Lam et al. (2010) resequenced 17 wild and 14 cultivated soybeans and found that marker-assisted breeding of soybean will be less challenging than map-based cloning. Both soybean and sorghum are self-pollinated crops although sorghum has a higher out-crossing rate. Because of this, the result will apply also to sorghum that if the right markers are identified, molecular breeding using these markers will be efficient in genetic improvement of sorghum. A method for efficient identification of markers associated with important traits in sorghum is what this study has demonstrated.

**Acknowledgements**

This work was supported by the University of Louisiana at Lafayette. We thank Jonathan A. Groomer, Chuanqin Xu and William H. Welsh for technical assistance.

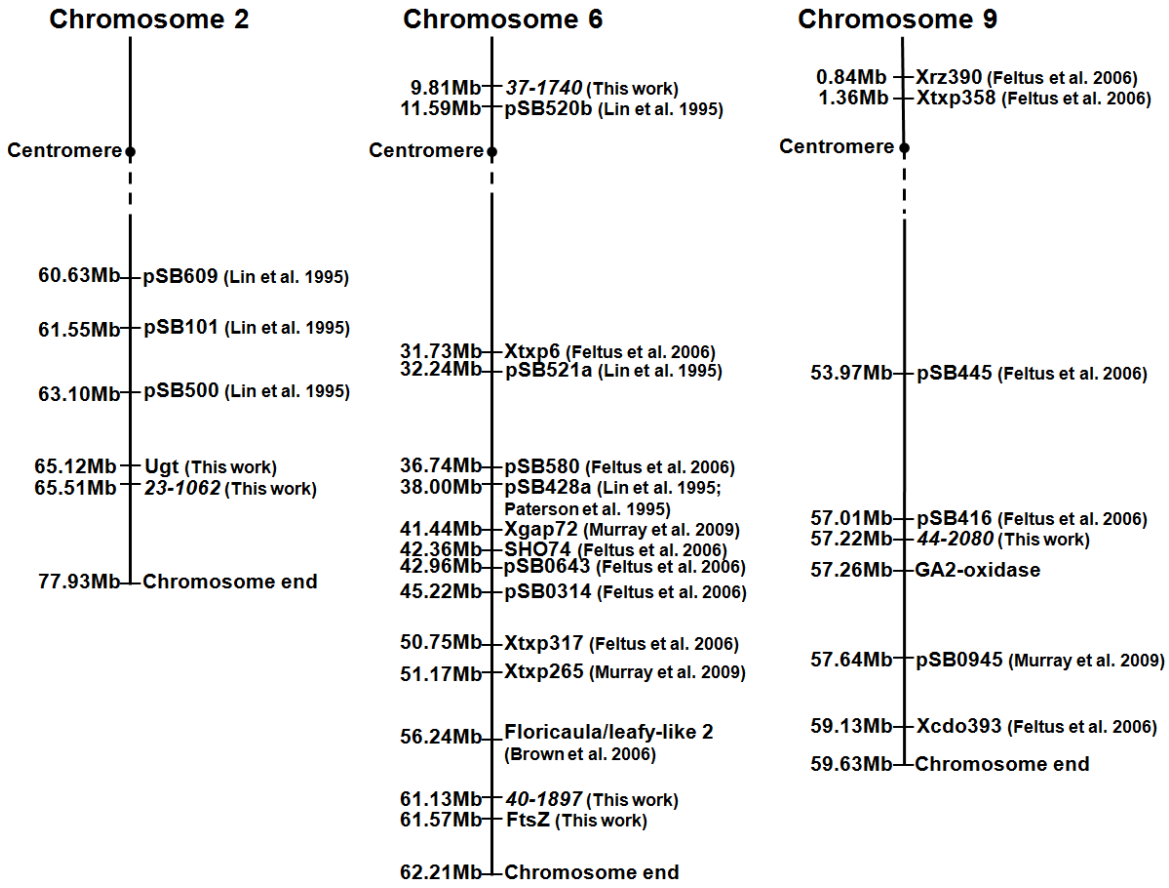



**Fig. 4** Physical localization of markers identified in this study (in italics) and markers previous mapped as linked to sorghum height and height-related traits. Source of markers is in parenthesis after each marker. For published markers, those prefixed with "pSB", "Xcdo", "Xrz" and "SHO" were RFLP markers while all others were SSR markers. Homologous genes previously shown to affect plant heights are also included (see text for details). Physical location was determined using the sorghum genome sequence from the following websites: http://www.plantgdb.org/SbGDB/, http://www.phytozome.net/sorghum, or http://genome.jgi-psf.org/. Not drawn to scale

## References


Abu Assar AH, Uptmoor R, Abdelmula AA, Salih M, Ordon F, Friedt W (2005) Genetic Variation in Sorghum Germplasm from Sudan, ICRISAT, and USA Assessed by Simple Sequence Repeats (SSRs). Crop Sci 45:1636–1644

Agrama HA, Tuinstra MR (2003) Phylogenetic diversity and relationships among sorghum accessions using SSRs and RAPDs. Afr J Biotech 2: 334-340

Ali ML, Rajewski JF, Baenziger PS, Gill KS, Eskridge KM, Dweiket I (2008) Assessment of genetic diversity and relationship among a collection of US sweet sorghum germplasm by SSR markers. Mol Breed 21: 497-509

Anas, Yoshida T (2004) Genetic Diversity among Japanese Cultivated Sorghum Assessed with Simple Sequence Repeats Markers. Plant Prod Sci 7:217-223

Barnaud A, Deu M, Garine E, McKey D, Joly HI (2007) Local genetic diversity of sorghum in a village in northern Cameroon: structure and dynamics of landraces. Theor Appl Genet 114:237–248

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

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

Bomblies K, Wang RL, Ambrose BA, Schmidt RJ, Meeley RB, Doebley J (2003) Duplicate *FLORICAULA/LEAFY* homologs *zfl1* and *zfl2* control inflorescence architecture and flower patterning in maize. Development 130:2385-2395

Bowers JE, Abbey C, Anderson S, Chang C, Draye X, Hoppe AH, Jessup R, Lemke C, Lennington J, Li Z, Lin YR, Liu SC, Luo L, Marler BS, Ming R, Mitchell SE, Qiang D, Reischmann K, Schulze SR, Skinner DN, Wang YW, Kresovich S, Schertz KF, Paterson AH (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 PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633-2635

Brown PJ, Klein PE, Bortiri E, Acharya CB, Rooney WL, Kresovich S (2006) Inheritance of inflorescence architecture in sorghum. Theor Appl Genet 113:931–942

Brown PJ, Rooney WL, Franks C, 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 VB, Meilan R, Pearce DW, Ma C, Rood SB, Strauss SH (2003) Activation tagging of a dominant gibberellin catabolism gene (*GA 2-oxidase*) from poplar that regulates tree stature. Plant Physiol 132: 1283–1291

Caniato FF, Guimarães CT, Schaffert RE, Alves VM, Kochian LV, Borém A, Klein PE, Magalhaes JV (2007) Genetic diversity for aluminum tolerance in sorghum. Theor Appl Genet 114:863-876

Casa AM, Mitchell SE, Hamblin MT, Sun H, Bowers JE, Paterson AH, Aquadro CF, Kresovich S (2005) Diversity and selection in sorghum: simultaneous analyses using simple sequence repeats. Theor Appl Genet 111: 23–30

Casa AM, Pressoir G, Brown PJ, Mitchell SE, Rooney WL, Tuinstra MR, Franks CD, Kresovich S (2008) Community Resources and Strategies for Association Mapping in Sorghum. Crop Sci 48:30–40

Chittenden LM, Schertz KF, Lin YR, Wing RA, Paterson AH (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

Deu M, Sagnard F, Chantereau J, Calatayud C, Hérault D, Mariac C, Pham JL, Vigouroux Y, Kapran I, Traore PS, Mamadou A, Gerard B, Ndjeunga J, Bezançon G (2008) Niger-wide assessment of in situ sorghum genetic diversity with microsatellite markers. Theor Appl Genet 116:903-913

Dillon SL, Kawrence PK, Henry RJ (2005) The new use of Sorghum bicolor-derived SSR markers to evaluate genetic diversity in 17 Australian Sorghum species. Plant Genet Resour 3: 19-28

Drgon T, Zhang PW, Johnson C, Walther D, Hess J, Nino M, Uhl GR (2010) Genome Wide Association for Addiction: Replicated Results and Comparisons of Two Analytic Approaches. PLoS One 5(1): e8832

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

Feltus FA, Hart GE, Schertz KF, Casa AM, Kresovich S, Abraham S, Klein PE, Brown PJ, Paterson AH (2006) Alignment of genetic maps and QTLs between inter- and intra-specific sorghum populations. Theor Appl Genet 112:1295–1305

Folkertsma RT, Rattunde HFW, Chandra S, Raju GS, Hash CT (2005) The pattern of genetic diversity of Guinea-race *Sorghum bicolor* (L.) Moench landraces as revealed with SSR markers. Theor Appl Genet 111:399-409

Giovannoni JJ, Wing RA, Ganal MW, Tanksley SD (1991) Isolation of molecular markers from specific chromosomal intervals using DNA pools from existing mapping populations. Nucl Acids Res 19: 6553–6558

Ghebru B, Schmidt R, Bennetzen J (2002) Genetic diversity of Eritrean sorghum landraces assessed with simple sequence repeat (SSR) markers. Theor Appl Genet 105: 229-236

Haussmann B, Hess D, Seetharama D, Welz H, Geiger H (2002) Construction of a combined sorghum linkage map from two recombinant inbred populations using AFLP, SSR, RFLP, and RAPD markers, and comparison with other sorghum maps. Theor Appl Genet 105: 629-637

Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z, Li M, Fan D, Guo Y, Wang A, Wang L, Deng L, Li W, Lu Y, Weng Q, Liu K, Huang T, Zhou T, Jing Y, Li W, Lin Z, Buckler ES, Qian Q, Zhang QF, Li J, Han B (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. Nat Genet 42:961-967

Klein PE, Klein RR, Cartinhour SW, Ulanch PE, Dong J, Obert JA, Morshige DT, Schlueter SD, Childs KL, Ale M, Mullet JE (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 RR, Mullet JE, Jordan DR, Miller FR, Rooney WL, Menz MA, Franks CD, Klein PE (2008) The Effect of Tropical Sorghum Conversion and Inbred Development on Genome Diversity as Revealed by High-Resolution Genotyping. Plant Genome S1: S12–S26

Knoll J, Ejeta G (2008) Marker-assisted selection for early-season cold tolerance in sorghum: QTL validation across populations and environments. Theor Appl Genet 116:541–553

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

Jeong WJ, Jeong SW, Min SR, Yoo OL, Liu JR (2002) Growth Retardation of Plank Transform by Overexpression of NtFtsZ1-2 in Tobacco. J Plant Biol 45: 107-111

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

Li M, Yuyama N, Luo L, Hirata M, Cai H (2009) In silico mapping of 1758 new SSR markers developed from public genomic sequences for sorghum. Mol Breed 24:41-47

Lin YR, Schertz KF, Paterson AH (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 PA, Macgregor S, Vink JM, Pergadia ML, Hansell NK, de Moor MH, Smit AB, Hottenga JJ, Richter MM, Heath AC, Martin NG, Willemsen G, de Geus EJ, Vogelzangs N, Penninx BW, Whitfield JB, Montgomery GW, Boomsma DI, Madden PA (2010) A genomewide association study of nicotine and alcohol dependence in Australian and Dutch populations. Twin Res Hum Genet 13:10-29

Mazzucato A, Papa R, Bitocchi E, Mosconi P, Nanni L, Negri V, Picarella ME, Siligato F, Soressi GP, Tiranti B, 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

Menz MA, Klein RR, Unruh NC, Rooney RL, Klein PE, Mullet JE (2004) Genetic Diversity of Public Inbreds of Sorghum Determined by Mapped AFLP and SSR Markers. Crop Sci 44:1236–1244

Michelmore RW, Paran I, Kesseli RV (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. Proc Natl Acad Sci USA 88: 9828–9832

Multani DS, Briggs SP, Chamberlin MA, Blakeslee JJ, Murphy AS, Johal GS (2003) Loss of an MDR transporter in compact stalks of maize *br2* and sorghum *dw3* mutants. Science 302:81-84

Murray SC, Rooney WL, Mitchell SE, Sharma A, Klein PE, Mullet JE, Kresovich S (2008) Genetic Improvement of Sorghum as a Biofuel Feedstock: II. QTL for Stem and Leaf Structural Carbohydrates. Crop Sci 48:2180–2193

Murray SC, Rooney WL, Mitchell SE, Kresovich S (2009) Sweet sorghum diversity and association mapping for Brix and height. Plant Genome 2:48-62

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

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Mehboob-ur-Rahman, Ware D, Westhoff P, Mayer KF, Messing J, Rokhsar DS (2009) The *Sorghum bicolor* genome and the diversification of grasses. Nature 457:551-556

Pei Z, Gao J, Chen Q, Wei j, Li Z, Luo F, Shi L, Ding B, Sun S (2010) Genetic diversity of elite sweet sorghum genotypes assessed by SSR markers. Biol Planta 54: 653-658

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

Paterson AH, Lin YR, Li Z, Schertz KF, Doebley JF, Pinson SRM, Liu SC, Stansel JW, Irvine JE (1995) Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. Science 269: 1714-1717

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Quinby JR, Karper RE (1954) Inheritance of height in Sorghum. Agron J 46:211–216

Ramu P, Kassahun B, Senthilvel S, Kumar CA, Jayashree B, Folkertsma RT, Reddy LA, Kuruvinashetti MS, Haussmann BIG, Hash CT (2009) Exploiting rice–sorghum synteny for targeted development of EST-SSRs to enrich the sorghum genetic linkage map. Theor Appl Genet 119: 1193-1204

Ritter KB, Jordan DR, Chapman SC, Godwin ID, Mace ES, McIntyre CL (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

Rooney WL, Blumenthal J, Bean B, Mullet JE (2007) Designing sorghum as a dedicated bioenergy feedstock. Biofuels Bioproducts Biorefining 1:147–157

Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, NJ, pp 365-386

Saballos A (2008) Development and utilization of sorghum as a bioenergy crop. In W Vermerris (ed.) Genetic Improvement of Bioenergy Crops. Springer, New York, pp 209-248

Salas Fernandez MG, Becraft PW, Yin Y, Lübberstedt T (2009) From dwarves to giants? Plant height manipulation for biomass yield. Trends Plant Sci 14:454-461

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

Shehzad T, Iwata H, Okuno K (2009a) Genome-wide association mapping of quantitative traits in sorghum (*Sorghum bicolor* (L.) Moench) by using multiple models. Breed Sci 59: 217-227

Shehzad T, Okuizumi H, Kawase M, Okuno K (2009b) Development of SSR-based sorghum (*Sorghum bicolor* (L.) Moench) diversity research set of germplasm and its evaluation by morphological traits. Genet Res Crop Evol 56: 809-827

Smith JSC, Kresovich S, Hopkins MS, Mitchell SE, Dean RE, Woodman RL, Lee M, Porter K (2000) Genetic Diversity among Elite Sorghum Inbred Lines Assessed with Simple Sequence Repeats. Crop Sci 40:226–232

Srinivas G, Satish K, Madhusudhana K, Seetharama N (2009) Exploration and mapping of microsatellite markers from subtracted drought stress ESTs in *Sorghum bicolor* (L.) Moench. Theor Appl Genet 118:703–717

Upadhyaya HD, Pundir RPS, Dwivedi SL, Gowda CLL, Reddy VG, Singh S (2009) Developing a mini core collection of sorghum for diversified utilization of germplasm. Crop Sci 49:1769–1780

Upadhyaya HD, Ortiz R (2001) A mini core subset for capturing diversity and promoting utilization of chickpea genetic resources in crop improvement. Theor Appl Genet 102:1292-1298

Wang ML, Barkley NA, Jenkins TM (2009a) Microsatellite Markers in Plants and Insects.
Part I: Applications of Biotechnology. Genes Genome Genomics 3: 54-67

Wang ML, Zhu C, Barkley NA, Chen Z, Erpelding JE, Murray SC, Tuinstra MR, Tesso T, Pederson GA, Yu J (2009b) Genetic diversity and population structure analysis of accessions in the US historic sweet sorghum collection. Theor Appl Genet 120:13-23

Williams CE, Ronald PC (1994) PCR template-DNA isolated quickly from monocot and dicot leaves without tissue homogenization. Nucl Acids Res 22 1917–1918

Woo HH, Orbach MJ, Hirsch AM, Hawes MC (1999) Meristem-localized inducible expression of a UDP-glycosyltransferase gene is essential for growth and development in pea and alfalfa. Plant Cell 11:2303-2315

Wu YQ, Huang Y (2006) An SSR genetic map of Sorghum bicolor (L.) Moench and its comparison to a published genetic map. Genome 50: 84-89

Yonemaru J, Ando T, Mizubayashi T, Kasuga S, Matsumoto T, Yano M (2009) Development of genome-wide simple sequence repeat markers using whole-genome shotgun sequences of sorghum (*Sorghum bicolor* (L.) Moench). DNA Res 16:187-193

Yu J, Buckler ES (2006) Genetic association mapping and genome organization of maize. Curr Opin Biotech 17:155–160

Yu JM, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203-208

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