DOI: 10.1534/genetics.108.092239

Efficient Mapping of Plant Height Quantitative Trait Loci in a Sorghum Association Population With Introgressed Dwarfing Genes

Patrick J. Brown,*,† William L. Rooney,‡ Cleve Franks§ and Stephen Kresovich*,,[1]**

**Institute for Genomic Diversity, †Department of Plant Biology, and **Department of Plant Breeding and Genetics, Cornell University, Ithaca, New York 14853, ‡Department of Soil and Crop Sciences, Texas A&M University, College Station, Texas 77843 and §United States Department of Agriculture–Agricultural Research Service, Cropping Systems Research Laboratory, Lubbock, Texas 79415*

Manuscript received June 5, 2008
Accepted for publication July 14, 2008

ABSTRACT

Of the four major dwarfing genes described in sorghum, only *Dw3* has been cloned. We used association mapping to characterize the phenotypic effects of the *dw3* mutation and to fine map a second, epistatic dwarfing QTL on sorghum chromosome 9 (*Sb-HT9.1*). Our panel of 378 sorghum inbreds includes 230 sorghum conversion (SC) lines, which are exotic lines that have been introgressed with dwarfing quantitative trait loci (QTL) from a common parent. The causal mutation in *dw3* associates with reduced lower internode length and an elongation of the apex, consistent with its role as an auxin efflux carrier. Lines carrying the *dw3* mutation display high haplotype homozygosity over several megabases in the *Dw3* region, but most markers linked to *Dw3* do not associate significantly with plant height due to allele sharing between *Dw3* and *dw3* individuals. Using markers with a high mutation rate and the *dw3* mutation as an interaction term, significant trait associations were detected across a 7-Mb region around *Sb-HT9.1*, largely due to higher detection power in the SC lines. Conversely, the likely QTL interval for *Sb-HT9.1* was reduced to ~100 kb, demonstrating that the unique structure of this association panel provides both power and resolution for a genomewide scan.

FOUR major dwarfing genes have been reported in sorghum, *Dw1–Dw4* (Quinby 1974). Most commercial grain sorghum lines are "3-dwarf," meaning that they carry three of the four dwarfing mutations. Only *Dw3* has been cloned, and encodes a phosphoglycoprotein auxin efflux carrier orthologous to PGP1 in Arabidopsis (Multani *et al.* 2003). The *Dw2* locus is linked to a major photoperiod-sensitivity locus, *Ma1*, on chromosome 6 (Lin *et al.* 1995), whereas *Dw1* and *Dw4* have not been mapped conclusively to a linkage group and are best defined by the lines presumed to carry recessive, dwarfing alleles at these loci according to early testcross studies (Quinby 1974). In this study, we use association mapping to identify genetic polymorphisms responsible for plant height variation in sorghum, beginning with the validation of a previously cloned gene (*Dw3*) and progressing to the fine mapping of a quantitative trait locus (QTL) for plant height on sorghum chromosome 9 (*Sb-HT9.1*).

Association or linkage disequilibrium (LD) mapping was first developed for human genetics, but shows great promise for the identification of polymorphisms underlying complex traits in crop plants (Flint-Garcia *et al.* 2003). In contrast to the traditional method of linkage mapping using biparental populations, association mapping exploits the allelic diversity and rich history of recombination in a set of diverse lines. Association is consequently able to provide greater resolution than linkage mapping at the expense of reduced power, making these two approaches highly complementary (Yu and Buckler 2006). In one recent study, a major flowering time locus in maize was mapped to a 2-kb interval of noncoding DNA ~70 kb upstream of the gene whose expression it affects (Salvi *et al.* 2007). The authors then used association mapping to identify three polymorphisms in this region that associated most strongly with flowering time. A subsequent association study using a much larger set of lines not only confirmed the previous result, but also identified a previously untyped polymorphism in the *Vgt1* region that shows an even stronger trait association (Ducrocq *et al.* 2008).

While lines from a biparental linkage population are all approximately equally related to each other, lines in an association study have an unknown, complex pattern of relatedness that must be estimated from marker data and accounted for. This issue is of critical importance because a naïve association test at a given marker may be affected by a suite of correlated effects from the rest of the genome, often described as "population structure" (Veyrieras *et al.* 2007). One current solution is to use both a vector of fixed effects and a matrix of random

[1]*Corresponding author:* Institute for Genomic Diversity, 158 Biotechnology, Cornell University, Ithaca, NY 14853. E-mail: sk20@cornell.edu

effects to control for both coarse- and fine-scaled levels of relatedness, respectively (Yu *et al.* 2006; Zhao *et al.* 2007). This method is proven to work well in maize and Arabidopsis, to the extent that the resulting cumulative distributions of *P*-values are unskewed. However, because many true associations are inevitably discarded when controlling for population structure, traits that correlate closely with population structure are less conducive to association mapping. For example, a 6-bp indel in the maize *Dwarf8* gene is strongly associated with both flowering time and Northern Flint ancestry, so that the significance of the trait association varies greatly according to the germplasm and population-structure correction used (Thornsberry *et al.* 2001; Andersen *et al.* 2005; Camus-Kulandaivelu *et al.* 2006). Another problem is that of genetic heterogeneity: when the same end phenotype is produced by mutations in many independent genes and/or independent mutations in the same gene, the power to detect association is significantly reduced. For example, a massive genomewide association study for Crohn's disease in humans recently identified 32 distinct susceptibility loci, most of which explained <0.5% of the total variance and had not been identified in previous scans that included only hundreds, rather than thousands, of cases and controls (Barrett *et al.* 2008).

The panel of sorghum inbred lines used in this study is expected to have high power to detect plant height QTL. The development of this sorghum panel for association mapping, including the development of population structure covariates and a kinship matrix, has been described previously (Casa *et al.* 2008). Briefly, the panel consists of 377 inbred lines for which population structure was estimated with 47 unlinked SSRs. Sixty percent of the panel is composed of sorghum conversion (SC) lines, which are short, early-flowering lines developed from tall, photoperiod-sensitive exotics through the introgression of dwarfing and photoperiod-insensitivity alleles from a common donor (Stephens *et al.* 1967). The remaining 40% of the panel consists of elite grain and forage lines, and lines of genetic and historical importance. The genomes of the SC lines and their kin are expected to feature low-diversity, high-LD "conversion regions" linked to maturity and height loci, interspersed within a background of higher diversity and lower LD. Since the SC lines were developed quite recently, using a limited number of backcrosses to the exotic parent, the conversion regions are expected to be quite large. However, the dwarfing mutations originated much earlier, likely in the progenitors of some of the historical sorghum lines included in our panel, in which LD around the mutations is expected to be much lower. A recent study by Klein *et al.* (2008) examined the *Ma1–Dw2* conversion region on chromosome 6 in a group of >50 public inbreds that included 16 SC lines. The converted haplotype block encompassing the uncloned *Ma1–Dw2* loci was variable in size, but usually quite large, and in some cases extended across nearly the entire chromosome. Notably, several SC lines in this study did not carry the expected converted haplotype at *Ma1*. These lines could carry an alternate *ma1* allele, as the authors suggest, or could simply carry the canonical *ma1* allele in an alternate, older haplotype with lower LD.

This study consists of two components: the validation of a known gene for plant height (*Dw3*), and the fine mapping of an uncloned plant height QTL, *Sb-HT9.1*. We began by testing for phenotypic associations with *Dw3*, a presumed target of the sorghum conversion program for which the mutation is known to be a tandem duplication in the fifth exon (Multani *et al.* 2003). Since contrasting effects of the *Dw3* mutation on plant and inflorescence architecture have been reported previously (Brown *et al.* 2006), we measured multiple height-component phenotypes in addition to total plant height. Using the pattern and extent of phenotypic associations with the *Dw3* locus as guidelines, we then used association methodology to fine-map *Sb-HT9.1*. *Dw3* and *Sb-HT9.1* are consistently identified as two of the most important plant height loci in crosses between tall and dwarf sorghum (Lin *et al.* 1995; Pereira and Lee 1995). The unique genetic structure of the sorghum population used here, which combines the detection power of a linkage population with the resolution of an association panel, shows promise for the further identification of major genes for plant height and maturity in sorghum.

MATERIALS AND METHODS

Plant materials and phenotyping: The panel of 378 sorghum lines used in this study has previously been described and characterized by Casa *et al.* (2008). In brief, this panel consists of 230 lines from the sorghum conversion program (SC lines), and 148 "elite" lines, which actually constitute not only elite lines from public breeding programs, but also assorted lines of genetic and historical interest, many of which carry SC lines in their pedigrees. SC lines were developed from tall, exotic sorghum lines by crossing to a four-dwarf, elite donor line (BTx406), selfing, and selecting for short, early segregants suitable for combine harvest; the process of backcrossing to the exotic, selfing, and selecting was performed an average of four times for each SC line (Stephens *et al.* 1967). Sorghum lines were phenotyped for six plant architectural traits in three replicates in Lubbock, TX in 2006: total plant height, preflag leaf height, preflag-to-flag leaf interval, distance from flag leaf to apex, rachis length, and panicle branch length (see Figure 1). The experimental design was a randomized complete block design, with a row length of 6 m and a row spacing of 1 m. From each plot, a single representative plant from the middle of each row was selected for measurement. Repeatability values shown in Table 1 were obtained by subtracting the fraction of the total phenotypic variance attributable to variance between repetitions (Falconer and Mackay 1996) using type III sum of squares in PROC GLM in SAS 9.1 (SAS Institute, Cary, NC). For association mapping, phenotypic values were standardized within each replicate by subtracting the mean and dividing by the standard deviation, and then averaged across replicates.

Genotyping: The set of random markers used to estimate population structure and kinship in this panel has been described previously (Casa *et al.* 2008). We added markers in the genomic region around the *Dw3* locus on chromosome 7 and in the genomic region encompassing a plant height and maturity QTL on sorghum chromosome 9. Additional SSR markers were obtained by blasting SSR repeats against Phytozome (http://www.phytozome.net; Sorghum Genome Project, DoE Joint Genome Institute) and additional MITE markers were developed using Inverted Repeat Finder (http://tandem.bu.edu). SSRs were run on an ABI 3730 with fluorescently labeled primers and scored using GeneMapper. MITEs were scored on agarose gels. The presence/absence of the tandem duplication in *Dw3* was scored on agarose gels using primers 5′ (TTCAACGCGGAGCGCAAGATCAC) and 5′ (CTTGAGCAGGTGCGAGTGCGA). Seven lines amplified a larger *dw3* fragment suggesting that they carried more than two tandem copies of the duplication; these lines had similar phenotypes to the lines with two copies and were grouped with them for subsequent analyses. Heterozygous individuals comprised <5% of the data for any single marker and were treated as missing data for all analyses.

Linkage disequilibrium: For the *dw3* data set, extended haplotype homozygosity (EHH) was calculated as described by Sabeti *et al.* (2002). For the *Sb-HT9.1* data set, LD was calculated separately for the converted and elite subsets of the panel using TASSEL 2.0.1 (Bradbury *et al.* 2007).

Association testing: The MLM function in TASSEL 2.0.1 was used to perform tests of association using the population structure covariates (Q1–Q9) and kinship matrix (K) reported previously (Casa *et al.* 2008). Both fixed Q and random K effects are fitted into a mixed model to account for coarse- and fine-grained patterns of relatedness, respectively, between lines (Yu *et al.* 2006). All significant associations were confirmed in SAS 9.1 (SAS Institute). Tests that included an interaction term between *dw3* and *Sb-HT9.1* were performed in SAS. r^2 values presented in Table 1 were calculated in SAS using the correlation between the trait value and the predicted value. For the *Sb-HT9.1* data set, genotype data from the dinucleotide repeat SSRs were converted to a biallelic format: the genotype carried by BTx406 (the elite donor line used in the sorghum conversion program) was designated as the "converted" genotype, and all other genotypes were designated as "nonconverted." Five lines with >50% missing data in the *Sb-HT9.1* region were excluded from the analysis. Recognizing that some nonconverted marker genotypes might be recently derived from converted genotypes by mutation, we tested a simple formula to identify such recently derived alleles. Blocks of at least three contiguous converted alleles interrupted by a single missing data point or a single allele within 4 bp of the converted allele were changed to a single, contiguous block of converted alleles. We allowed imputation of multiple marker genotypes per line, but only if they were not adjacent. Data treated in this way yielded results very similar to those from the untreated, biallelic data set, so only the untreated biallelic data are presented. The complete phenotypic and genotypic data sets used in this study are available on-line as supplemental data.

RESULTS

Trait associations with the tandem duplication in *dw3*: The duplication in the fifth exon of *dw3* is present in 215 lines and absent in 152 lines (no amplicon was obtained for 11 of the 378 lines), for an overall frequency of 58.5%. The frequency of the *dw3* duplication is essentially identical between the elite and converted panels (58.3 *vs.* 58.7%). Presence of the *dw3* duplication associates with a reduction in total plant height ($P = 1.2e{-}6$) and preflag leaf height ($P = 4.1e{-}11$), and an increase in the distance from flag leaf to apex ($P = 1.7e{-}9$). The *dw3* mutation does not associate with inflorescence traits, which could be due to either a true lack of association or to a lack of power, since the two inflorescence traits show much higher correlation with population structure. The mixed model (Q + K) accounts for 32–39% of the phenotypic variation for vegetative traits, 54–60% of the variation for inflorescence traits, and an intermediate amount (42%) of the variation for the distance from flag to apex, which is a composite of vegetative and inflorescence organ lengths (Table 1). One reason for this difference may be the recent introgression of major dwarfing mutations such as *dw3*, which has a large phenotypic effect and does not correlate with population structure. There are almost certainly loci with similarly large effects on inflorescence traits, but since they were not targeted for introgression by the sorghum conversion program their frequencies are likely to be lower and more highly correlated with population structure.

TABLE 1

Trait repeatabilities and the proportion of the phenotypic variance explained (r^2) by various models

Trait	Repeatability	Q only	K only	Q + K	Q + K + *dw3*
Plant height	0.67	0.12	0.37	0.39	0.41[a]
Preflag leaf height	0.73	0.08	0.37	0.39	0.43[a]
Preflag to flag	0.76	0.18	0.32	0.32	0.32
Flag leaf to apex	0.73	0.14	0.42	0.42	0.44[a]
Rachis length	0.58	0.23	0.53	0.54	0.54
Branch length	0.60	0.31	0.59	0.60	0.60

[a] The effect of *dw3* on this trait is significant at $P < 0.001$.

Trait associations *vs.* haplotype homozygosity around the *Dw3* locus: To assess the degree to which polymorphisms closely linked to the *dw3* duplication might also associate with plant architecture, seven additional indel polymorphisms in the *Dw3* genomic region were genotyped, two of which were positioned within the *Dw3* gene itself (Figure 2). Significant trait associations were detected with two of the linked loci, including a MITE several kb upstream of the tandem duplication in the fourth intron, and another MITE ~300 kb downstream of the *Dw3* gene, but several other closely linked polymorphisms showed no association at all (Figure 2A). However, the lines with the *dw3* duplication had considerably higher LD and lower diversity across the entire region sampled. This is reflected in the slower decline of EHH in lines with the *dw3* duplication than in lines without the duplication (Figure 2A). For example, there is a 50% chance that two randomly sampled lines with the *dw3* duplication will carry identical haplotypes for

TABLE 2

Epistatic interaction between the *dw3* mutation and the chromosome 9 QTL

Trait	Average phenotypic values (σ) ± 95% C.I.				Model effect (σ)		
	Dw3, HT9.1 (+/+, +/+) $N = 48$ (11 SC, 37 elite)	*dw3, HT9.1* (−/−, +/+) $N = 75$ (51 SC, 24 elite)	*Dw3, ht9.1* (+/+, −/−) $N = 80$ (68 SC, 12 elite)	*dw3, ht9.1* (−/−, −/−) $N = 131$ (77 SC, 54 elite)	*dw3*	*ht9.1*	*dw3* × *ht9.1*
Plant height	1.53 ± 0.33	0.09 ± 0.15	−0.42 ± 0.13	−0.44 ± 0.09	−1.35	−1.80	+1.28
Preflag leaf height	1.65 ± 0.32	0.08 ± 0.13	−0.30 ± 0.14	−0.53 ± 0.07	−1.47	−1.74	+1.19
Preflag to flag	0.57 ± 0.26	0.11 ± 0.19	−0.19 ± 0.17	−0.21 ± 0.14	NS	NS	NS
Flag leaf to apex	−0.09 ± 0.25	0.02 ± 0.20	−0.55 ± 0.18	0.29 ± 0.15	+0.326	NS	NS

Average phenotypic values of sorghum lines carrying one, both, or neither of the dwarfing mutations are shown, along with main effects for each mutation and the interaction effect between them. Values shown are in standard deviations relative to the mean; only model effects significant at $P < 0.001$ are shown.

the entire 1.2 Mb from the most proximal sampled SSR at 57.4 Mb to the *dw3* duplication (EHH ~0.5) whereas the likelihood of this event occurring between two randomly sampled lines without the *dw3* duplication is extremely low (EHH ~0).

Allele sharing between *Dw3* and *dw3* alleles: To reconcile the slow decline in haplotype homozygosity with the much more rapid decline in trait associations in the *Dw3* region, we compared the incidence of two potential sources of confounding in the data: recombination, which manifests itself in lines carrying the *dw3* duplication but not carrying the converted (BTx406) allele at the marker being tested, and shared ancestry, which manifests itself in lines without the *dw3* duplication that do carry the converted allele at the marker being tested (Figure 2B). Whereas recombination increases in predictable, linear fashion with distance from the causal mutation and reaches a maximum of 30% within the region tested, shared ancestry affects a much larger proportion of lines in this data set, and represents a true source of confounding in that it does not change predictably with distance from the causal mutation. For example, the SSR locus 200 bp away from the tandem duplication is in complete linkage with the causal mutation ($D' = 1$; all the lines with the *dw3* duplication carry the same SSR allele), but a full 80% of lines not carrying the *dw3* duplication also carry this same SSR allele, $r^2 < 0.2$, and this locus shows no significant trait associations.

Fine mapping *Sb-HT9.1* using association: Results from the *Dw3* locus were used to guide the fine mapping of a second major dwarfing QTL on sorghum chromosome 9, *Sb-HT9.1*. Since the dwarfing allele at this locus has been identified as recessive (Pereira and Lee 1995), we will refer to the tall and short alleles at this QTL as *Sb-HT9.1* and *Sb-ht9.1*, respectively. The *Sb-HT9.1* QTL was not detected in the (BTx623 × IS3620C) RIL population (Brown *et al.* 2006), but the data reported here show that IS3620C has been converted at this locus, so the QTL is not segregating in that cross. This converted region in IS3620C aligns closely with previously reported QTL for plant height in sorghum (Lin *et al.* 1995; Pereira and Lee 1995). We specifically selected dinucleotide SSRs to genotype in this genomic region, on the assumption that by maximizing allele number we would minimize the incidence of shared ancestry that had confounded trait associations at the *Dw3* locus. A total of 13 SSRs were genotyped across a 7-Mb stretch of chromosome 9 that spans the entire converted region in IS3620C (Figure 3). Each marker was tested using the Q + K model, first without additional covariates, then including the *dw3* duplication as a covariate, and finally including an interaction term with the *dw3* duplication. The trait used for *Sb-HT9.1* QTL mapping is preflag leaf height, because the contrasting effects of *dw3* on different components of total plant height could complicate its use as a cofactor and interaction term in models for this trait. The same marker at 57.21 Mb consistently gives the most significant *P*-value across all analyses, and the frequency of the converted allele reaches a maximum of >60% around the same position, from 56.99 to 57.21 Mb. Inclusion of the *dw3* interaction term yielded the most significant results, and this difference became more pronounced with increasing proximity to the putative QTL.

Power and resolution of association mapping using sorghum converted lines: The difference in QTL detection power between the converted lines and the rest of the panel is manifested in their respective patterns of trait associations in the *Sb-HT9.1* QTL region (Figure 4). The converted lines have greater power to detect the presence of a linked QTL: using the *dw3* interaction model, significant *P*-values are obtained over a 2.75-Mb window between 56.31 and 59.07 Mb. In contrast, the largest contiguous block of significant *P*-values around the putative QTL in the remainder of the panel (consisting of elite, historical, and genetic stock lines) is just 200 kb for the *dw3* interaction model (from 56.99 to 57.21 Mb), a 12-fold difference. However, the elite panel provides slightly greater resolution, with a sharp peak at 57.21 Mb, whereas the converted panel more or

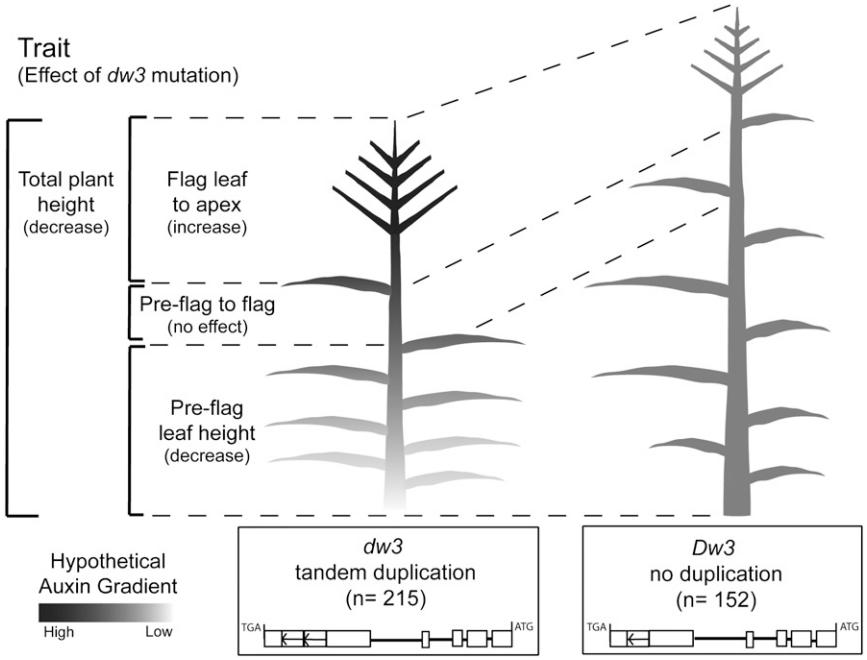

FIGURE 1.—Height component phenotypes and their associations with the *Dw3* locus. A typical *dw3* plant is portrayed at left and a typical *Dw3* plant at right. *dw3* plants carry a tandem duplication of 882 bp in the fifth exon. The three height components that make up total plant height are shown at left. Apical internodes are longer and basal internodes are shorter in a *dw3* background. Rachis length and branch length were also measured, but did not associate with *Dw3*.

less plateaus between 57.14 and 57.21 Mb. The full panel (Figure 3) shows the advantages conferred by both subsets, demonstrating that power and resolution are not mutually exclusive in the context of association-panel design. Contrary to our expectations, LD in the *Sb-HT9.1* region is actually higher in the elite panel than in the converted panel. Therefore, the increased resolution in the elite panel may result from just a small number of lines with low LD around *Sb-HT9.1*. LD between *dw3* and the *Sb-HT9.1* QTL is also higher in the

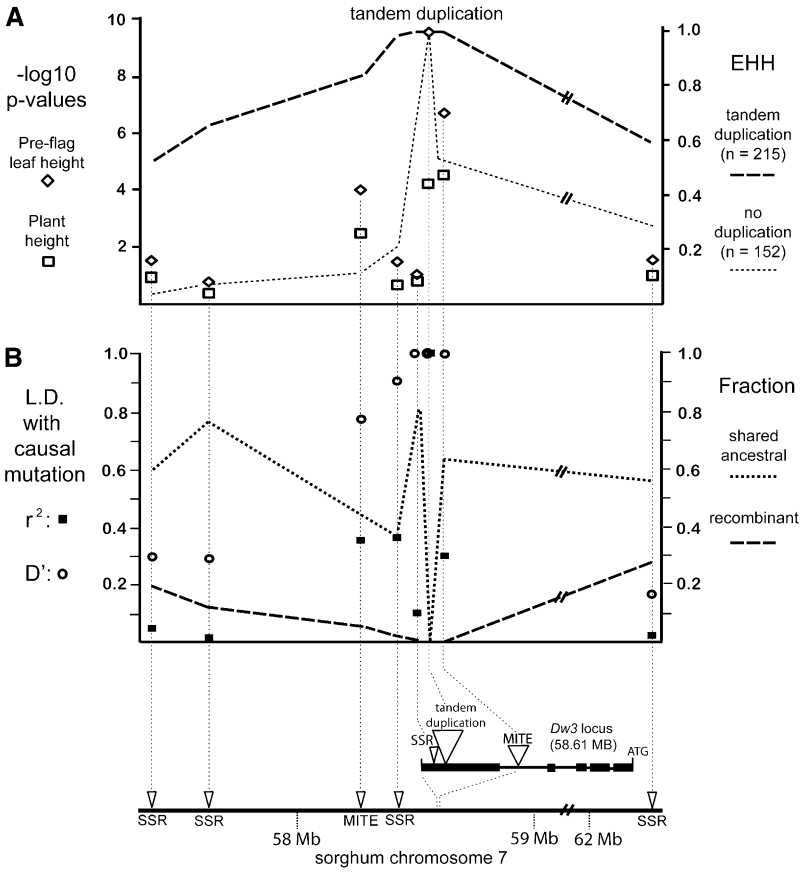

FIGURE 2.—Trait associations and patterns of linkage disequilibrium in the *Dw3* region. Eight markers in the *Dw3* region were tested, including three within the *Dw3* locus. (A) Comparison of trait associations and extended haplotype homozygosity (EHH), which provides a means of comparing LD decay between groups carrying different core haplotypes. In this case, the core haplotype is defined as the presence/absence of the tandem duplication in exon 5, which is the causal mutation. (B) Comparison of r^2 and D' decay as a result of recombination and shared ancestry for markers linked to the *dw3* mutation. The recombinant fraction is defined as the proportion of lines carrying the causal mutation that do not carry the BTx406 allele at the marker being tested. The shared ancestral fraction is defined as the proportion of lines not carrying the causal mutation that do carry the BTx406 allele at the marker being tested. D' is affected only by recombination, whereas r^2 is affected by both recombination and shared ancestry.

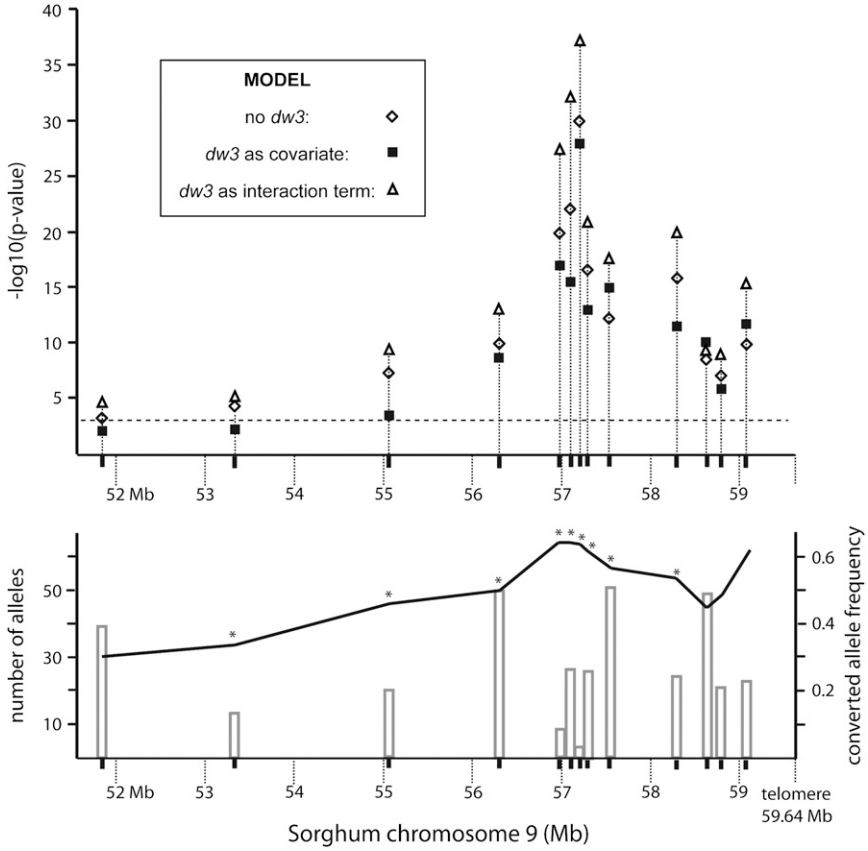

FIGURE 3.—Association mapping of *Sb-HT9.1.* Thirteen dinucleotide repeat SSRs over 7 Mb on sorghum chromosome 9 were genotyped and are indicated with thick black lines along the *x*-axis. (Top) Association results are shown: markers were tested first using the basic Q + K model (YU *et al.* 2006), then with *dw3* added as a covariate, and finally with an interaction term between *dw3* and *Sb-HT9.1.* The nominal significance threshold of $P = 0.001$ is shown as a dashed horizontal line. (Bottom) The number of alleles (shaded boxes) and the frequency of the converted allele (solid line) for each marker are shown. Markers that show evidence of conversion in IS3620C are marked with an asterisk above the solid line: this information was used to select markers to span the QTL. Results shown are for the preflag leaf-height trait; results for total plant height are very similar but slightly less significant.

elite panel, presumably because most elite grain sorghum lines carry both of these dwarfing mutations and most elite forage and sweet sorghums carry neither, whereas many more converted lines carry just one dwarfing mutation or the other. This level of cryptic population structure is not easily detected using genomewide marker information and may account for the low power to detect *Sb-HT9.1* in the elite panel.

Epistasis between *Dw3* and the chromosome 9 QTL: We used marker data from the most highly significant *Sb-HT9.1* marker, at 57.21 Mb, and the tandem duplication in *dw3* to infer whether each line in the panel carries one, both, or neither of the dwarfing QTL. The height difference between plants carrying zero and one of these mutations is significantly greater than the height difference between plants carrying one and two mutations, as shown by the positive interaction effect in (*dw3, Sb-ht9.1*) plants (Table 2). Although (*Dw3, Sb-ht9.1*) and (*dw3, Sb-ht9.1*) plants have nearly identical average plant heights, the *dw3* QTL still has a strong effect on plant height in a *Sb-ht9.1* background in at least one QTL study (BROWN *et al.* 2006). For this reason we favor a model in which the effect of each additional height mutation becomes progressively less, rather than one in which *Dw3* is completely epistatic to *Sb-HT9.1* for plant height. A similar epistatic effect between these two QTL in sorghum has been reported previously (PEREIRA and LEE 1995).

DISCUSSION

Associations with *dw3* suggest reduced auxin transport from the apex: The *dw3* duplication associates with a decrease in both plant height and preflag leaf height, whereas it associates with an increase in the distance from flag to apex (Figure 1). Given that *dw3* encodes an auxin efflux carrier, one logical hypothesis is that the elongation of apical nodes is caused by a buildup of auxin near its sites of synthesis in the apex. QTL analysis of the *Dw3* region in recombinant inbred lines revealed QTL for increased rachis length and branch length as well as decreased plant height (BROWN *et al.* 2006), but this difference in apical elongation was not reported in an analysis of *dw3/dw3* plants that reverted back to *Dw3/dw3* by unequal crossing over (MULTANI *et al.* 2003). Since this study used homozygous lines, it is possible that *Dw3/dw3* heterozygotes are indeed overdominant, with elongation of both apical and basal internodes. Alternatively, the apical node elongation associated with the *dw3* could result from the action of a linked gene. Specifically, 80 kb downstream of the *Dw3* locus there is a flavin monooxygenase of the *yucca* family, which has been implicated in auxin synthesis in Arabidopsis and rice (CHENG *et al.* 2006; YAMAMOTO *et al.* 2007); one characterized *Arabidopsis yucca* mutant displays extreme inflorescence elongation (KIM *et al.* 2007). Since only 6 of the 215 lines with the *dw3* duplication show evidence of

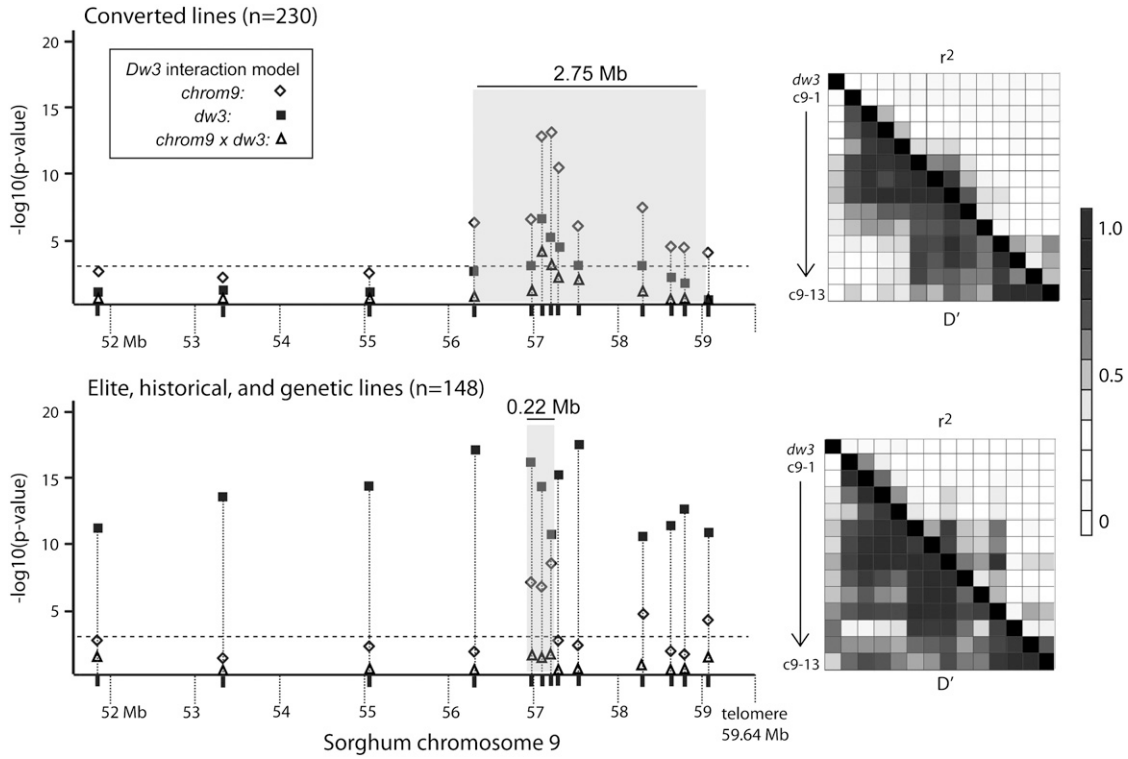

FIGURE 4.—Linkage disequilibrium and trait associations for the *Sb-HT9.1* QTL region in the sorghum converted (SC) lines ($n = 230$) *vs.* the rest of the panel ($n = 148$). The panels at left show trait associations: axes are the same as in Figure 3. Only the terms from the *Dw3* interaction model are shown. The shaded area in each panel indicates the region of contiguous significant *P*-values around the putative *Sb-HT9.1* QTL: this region extends 2.75 Mb in the converted lines *vs.* only 220 kb in the rest of the panel. The LD plots at right show both r^2 (top right) and D' (bottom left) between the *dw3* duplication and the 13 *Sb-HT9.1* markers.

recombination between *Dw3* and the linked *yucca* locus, however, we have little power to dissect the relative contributions of these two genes to the phenotype.

Detection of the *dw3* QTL: The maintenance of high EHH over several megabases in lines carrying the *dw3* duplication provides evidence that this haplotype was strongly selected for, as demonstrated previously for the *y1* locus in maize (PALAISA *et al.* 2003). Importantly, the EHH statistic measures the *potential* to obtain significant associations with linked loci, whereas the actual trait associations of individual markers are affected by circumstantial factors, such as allele frequency. Since the sorghum conversion program used phenotypic selection and relatively few backcrosses, the *dw3* locus is surrounded by a converted region of high EHH large enough to feasibly be detected in a genomewide scan. The discrepancy between high EHH and limited trait associations at the *dw3* locus could be due to demographic and/or biological factors. First, the tandem duplication in *dw3* may be a recent mutation and/or may have occurred in a haplotype that was already at high frequency, both of which would explain the high degree of allele sharing between *Dw3* and *dw3* haplotypes. Second, there may be genetic heterogeneity at the *Dw3* locus, such that additional, untyped mutations are also able to confer a *dw3* phenotype. In Figure 2, trait associations do not appear to be completely dependent on their r^2 value with the tandem duplication, as one might expect if the tandem duplication were the only causal mutation (*i.e.*, see DUCROCQ *et al.* 2008, Figure 1). The effect of the tandem duplication in *dw3* is also much higher in the elite panel than in the converted panel (Figure 4), again suggesting that there is either genetic heterogeneity in the converted panel or that the effect of the *dw3* tandem duplication is strongly background dependent.

Inferring positional information from association: One drawback to fine mapping using association methodology is that positional effects are confounded with the stochastic variation in the information content of individual markers. In this study, we attempted to minimize such variation by using markers with a high mutation rate and converting the genotype data to a biallelic format. A high mutation rate is expected to minimize the incidence of shared alleles between *Sb-HT9.1* and *Sb-ht9.1* haplotypes, and collapsing all the *Sb-HT9.1* haplotypes into a single class is expected to increase power, assuming a single origin of all dwarfing alleles at this locus. The most highly significant marker in this region, at 57.21 Mb, is also the marker with the fewest number of alleles (Figure 3). This raises the possibility of bias in our positional estimate, since for markers with fewer alleles, *Sb-ht9.1* (converted) haplotypes may

be less likely to have mutated to a *Sb-HT9.1*-like (nonconverted) state. However, we attempted to correct for such bias by inferring the presence of recently derived alleles (see MATERIALS AND METHODS) and obtained results essentially identical to those for the untreated, biallelic data set.

QTL detection power in converted *vs.* elite sorghum lines: Surprisingly, LD in the *Sb-HT9.1* region declines more rapidly in the converted lines compared to the rest of the association panel (Figure 4), so the increased power to detect the *Sb-HT9.1* QTL in the converted lines cannot simply result from increased LD. However, the structure of the converted and elite subgroups is very different: *Sb-HT9.1* in the converted lines is segregating independently of *Dw3*, whereas *Sb-HT9.1* in the elite lines is in LD with *Dw3*, with most elite lines carrying either both dwarfing mutations (*dw3, ht9.1* in Table 2) or neither (*Dw3, HT9.1*). This may also explain why the effect of *Dw3* is so much higher in the elite panel. Since a perfect marker is available for *Dw3* but not for *HT9.1* and the mutations are in LD with each other, *Dw3* "absorbs" some of the *HT9.1* effect for all markers except those most closely linked to the *HT9.1* mutation (see Figure 4). It is expected that QTL for other agronomic traits, such as flowering time, will show a similar pattern across the subsets of this panel. These results highlight the usefulness of assembling an association panel that reduces the correlation between population structure and the trait of interest.

Prospects for identifying sorghum genes controlling variation in agronomic traits: This study provides a framework for cloning major genes for plant height and flowering time in sorghum. Significant trait associations in the *Sb-HT9.1* QTL region extend over 7 Mb, or nearly 1% of the physical extent of the sorghum genome, suggesting that a genomewide scan for plant-height and flowering-time genes could be performed in this panel with as few as several hundred markers. Conversely, the most likely interval for the *Sb-HT9.1* QTL, between the two most significant markers at 57.14 and 57.21 Mb, covers just 75 kb and contains just 11 predicted genes (Sorghum Genome Project, DoE Joint Genome Institute). Therefore, the paradigm of high LD, low-resolution linkage studies and low LD, high-resolution association mapping studies may be somewhat oversimplified, since association panel design is flexible enough to allow the incorporation of many different LD structures. Power to detect QTL for traits not targeted by the sorghum conversion program is likely to be considerably reduced. As more major genes for plant height and flowering time are identified in sorghum, their inclusion as model covariates should facilitate the cloning of the QTL that remain. For example, several sorghum lines with tall alleles at both *Dw3* and *Sb-HT9.1* are nevertheless quite short and are likely to carry novel dwarfing or early-flowering QTL; tall, late-flowering lines carrying both *dw3* and *Sb-ht9.1* also occur. The approach described here will not yield an exhaustive catalog of all polymorphisms that affect plant height and flowering time in sorghum, but the combination of high detection power and acceptable resolution afforded by this panel provides a simple and cost-effective means of quickly isolating genomic regions with large phenotypic effects on these important agronomic traits.

The authors thank Gael Pressoir for the SAS script for running the mixed model, Alexandra Casa for phenotyping assistance, and Gael Pressoir and two anonymous reviewers for their comments on the manuscript.

LITERATURE CITED

Andersen, J. R., T. Schrag, A. E. Melchinger, I. Zein and T. Lubberstedt, 2005 Validation of *Dwarf8* polymorphisms associated with flowering time in elite European inbred lines of maize (Zea mays L.). Theor. Appl. Genet. **111:** 206–217.

Barrett, J. C., S. Hansoul, D. L. Nicolae, J. H. Cho, R. H. Duerr *et al.*, 2008 Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. Nat. Genet. **40:** 955–962.

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

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

Camus-Kulandaivelu, L., J. B. Veyrieras, D. Madur, V. Combes, M. Fourmann *et al.*, 2006 Maize adaptation to temperate climate: relationship between population structure and polymorphism in the Dwarf8 gene. Genetics **172:** 2449–2463.

Casa, A. M., G. Pressoir, P. J. Brown, S. E. Mitchell, W. L. Rooney *et al.*, 2008 Community resources and strategies for association mapping in sorghum. Crop Sci. **48:** 30–40.

Cheng, Y., X. Dai and Y. Zhao, 2006 Auxin biosynthesis by the YUCCA flavin monooxygenases controls the formation of floral organs and vascular tissues in Arabidopsis. Genes Dev. **20:** 1790–1799.

Ducrocq, S., D. Madur, J.-B. Veyrieras, L. Camus-Kulandaivelu, M. Kloiber-Maitz *et al.*, 2008 Key impact of *Vgt1* on flowering time adaptation in maize: evidence from association mapping and ecogeographical information. Genetics **178:** 2433–2437.

Falconer, D. S., and T. F. C. Mackay, 1996 *Introduction to Quantitative Genetics*, Ed. 4. Longman, New York.

Flint-Garcia, S. A., J. M. Thornsberry and E.S. Buckler, 2003 Structure of linkage disequilibrium in plants. Annu. Rev. Plant Biol. **54:** 357–374.

Kim, J. I., A. Sharkhuu, J. B. Jin, P. Li, J. C. Jeong *et al.*, 2007 yucca6, a dominant mutation in Arabidopsis, affects auxin accumulation and auxin-related phenotypes. Plant Physiol. **145:** 722–735.

Klein, R. R., J. E. Mullet, D. R. Jordan, F. R. Miller, W. L. Rooney *et al.*, 2008 The effect of tropical sorghum conversion and inbred development on genome diversity as revealed by high-resolution genotyping. Crop Sci. **48**(S1): S12–S26.

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

Multani, D. S., S. P. Briggs, M. A. Chamberlin, J. J. Blakeslee, A. S. Murphy *et al.*, 2003 Loss of an MDR transporter in compact stalks of maize *br2* and sorghum *dw3* mutants. Science **302:** 81–84.

Palaisa, K. A., M. Morgante, M. Williams and A. Rafalski, 2003 Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci. Plant Cell **15:** 1795–1806.

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

Quinby, J. R., 1974 *Sorghum Improvement and the Genetics of Growth.* Texas A&M University Press, College Station, TX.

Sabeti, P. C., D. E. Reich, J. M. Higgins, H. Z. Levine, D. J. Richter *et al.*, 2002 Detecting recent positive selection in the human genome from haplotype structure. Nature **419:** 832–837.

Stephens, J. C., F. R. Miller and D. T. Rosenow, 1967 Conversion of alien sorghums to early combine genotypes. Crop Sci. **7:** 396.

Salvi, S., G. Sponza, M. Morgante, D. Tomes, X. Niu *et al.*, 2007 Conserved noncoding sequences associated with a flowering time quantitative trait locus in maize. Proc. Natl. Acad. Sci. USA **104:** 11376–11381.

Thornsberry, J. M., M. M. Goodman, J. Doebley, S. Kresovich, D. Nielsen *et al.*, 2001 *Dwarf8* polymorphisms associate with variation in flowering time. Nat. Genet. **28:** 286–289.

Veyrieras, J. P., L. Camus-Kulandaivelu, B. Gouesnard, D. Manicacci and A. Charcosset, 2007 Bridging genomics and genetic diversity: linkage disequilibrium structure and association mapping in maize and other cereals. Crop Sci. **47:** S60–S71.

Yamamoto, Y., N. Kamiya, Y. Morinaka, M. Matsuoka and T. Sazuka, 2007 Auxin biosynthesis by the YUCCA genes in rice. Plant Physiol. **143:** 1362–1371.

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

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

Zhao, K. M., J. Aranzana, S. Kim, C. Lister, C. Shindo *et al.*, 2007 An Arabidopsis example of association mapping in structured samples. PLoS Genet. **3:** e4.

Communicating editor: A. H. Paterson