**ABSTRACT**

Pearl millet is one of the most important subsistence crops grown in India and sub-Saharan Africa. In many cereal crops, reduced height is a key trait for enhancing yield, and dwarf mutants have been extensively used in breeding to reduce yield loss due to lodging under intense management. In pearl millet, the recessive d2 dwarfing gene has been deployed widely in commercial germplasm grown in India, the United States, and Australia. Despite its importance, very little research has gone into determining the identity of the d2 gene. We used comparative information, genetic mapping in two F2 populations representing a total of some 1500 progeny, and haplotype analysis of three tall and three dwarf inbred lines to delineate the d2 region by two genetic markers that, in sorghum, define a region of 410 kb with 40 annotated genes. One of the sorghum genes annotated within this region is ABCB1, which encodes a P-glycoprotein involved in auxin transport. This gene had previously been shown to underlie the economically important dw3 dwarf mutation in sorghum. The cosegregation of ABCB1 with the d2 phenotype, its differential expression in the tall inbred ICMP 451 and the dwarf inbred Tift 23DB, and the similar phenotype of stacked lower internodes in the sorghum dw3 and pearl millet d2 mutants suggest that ABCB1 is a likely candidate for d2.

**KEYWORDS**

comparative genomics  fine-mapping haplotype analysis P-glycoprotein
The precise origin of the d2 dwarf mutation is unknown. In the United States, Burton and colleagues discovered in 1939 an extremely leafy pearl millet plant with short internodes among the progeny of a plant obtained through mass selection from five introductions of pearl millet acquired a few years earlier from the Vavilov Institute of Plant Industry, Russia (Burton and Devane 1951; Hein 1953). Based on information in the Germplasm Resource Information Network (i.e., GRIN) database, the five introductions originated from Tunisia (PI 115055), Eritrea (PI 115056, PI 115058), Arabia (PI 115057), and India (PI 115059). The dwarf line was true-breeding and used in crosses with an adapted pearl millet line to form the highly successful synthetic variety ‘Starr’ (Hein 1953). Although there are no records confirming that Starr millet carried the d2 gene, the described morphology makes this a plausible hypothesis (Kumar and Andrews 1993). Around the same time, in India, Kadam et al. (1940) obtained dwarf phenotypes after inbreeding local pearl millet lines. The dwarfs had shortened internodes, overlapping leaf sheaths, and shortened peduncles and were attributed to a recessive mutation. Again, it is unclear whether any of these represented d2. In 1966, Burton and Fortson (1966) reported identification of five nonallelic dwarf mutants (D1—D5). Two of those, D1 and D2, were shown to be controlled by different single recessive genes and were assigned the gene symbols d1 and d2, respectively. The d2 gene was subsequently incorporated into Indian cultivars through backcross breeding using seed stocks provided by Dr. G. W. Burton (Bakshi et al. 1966) and is now widely used in commercial hybrids in the United States, India, and Australia (reviewed by Kumar and Andrews 1993; Gulia et al. 2007; Rai et al. 2009). The d2 gene has been mapped on pearl millet linkage group 4 to a 23.2-cM interval flanked by RFLP markers PSM84 and PSM413.2 (Azhagavel et al. 2003).

Height-reducing genes have played key roles in enhancing yield in a range of cereals. The best-known examples are the gibberellic acid (GA)-insensitive Rht-1 and GA-sensitive sd-1 dwarfing genes that were essential to the Green Revolutions in wheat and rice, respectively (Peng et al. 1999; Monna et al. 2002; Sasaki et al. 2002), but height mutants have also been widely used in other cereals. For example, in barley, the GA-sensitive sdw/denso gene located on chromosome 3H (Laurie et al. 1993) has been used extensively in feed and malt cultivars in the Western United States, Canada, Europe, and Australia (reviewed by Mickelson and Rasmusson 1994). Most commercial sorghum lines are “3-dwarf,” which indicates that they carry mutations in three of the four dwarfing genes that have been identified in this species (Schertz et al. 1974). The height-reducing gene Ddlw1 on rye chromosome 5R has been deployed in many Eastern-European and Finnish rye breeding programs to develop short-straw cultivars (Milach and Federizzi 2001). A number of these dwarfing genes have been isolated and characterized. The Rht-1 genes encode DELLA proteins that act as repressors of plant growth (Peng et al. 1999). Mutations in the DELLA domain inhibit GA-induced degradation of the DELLA proteins, which results in a GA-insensitive dwarfing phenotype (Peng et al. 1997; Dill et al. 2001; reviewed by Hedden 2003). The rice sd-1 gene is a GA 20-oxidase, which catalyzes multiple steps in the GA biosynthetic pathway (Monna et al. 2002; Sasaki et al. 2002; Spielmeyer et al. 2002), and it has recently been shown that the sdw/denso gene in barley is likely an ortholog of sd-1 in rice (Jia et al. 2009, 2011). The dw3 dwarf phenotype in sorghum is caused by an 882-bp tandem duplication in the fifth exon of the ABCC1 gene. This rearrangement results in the loss of the encoded P-glycoprotein, which modulates polar auxin transport in the stalk (Multani et al. 2003).

The aim of our research was to fine-map the d2 gene in pearl millet and to use comparative information to identify putative candidate genes for the locus. Genetic mapping of an identified candidate gene and preliminary expression analysis provide support for a model that the pearl millet d2 gene is the ortholog of sorghum dw3.

# MATERIALS AND METHODS

## Mapping populations

An F2 mapping population of a few thousand seed was generated by selfing a single F1 hybrid from a cross between the d2 dwarf inbred Tift 23DB (female; Figure 1A) and the tall inbred ICMP 451 (male; Figure 1B). Tift 23DB was obtained from Wayne Hanna, University of Georgia, Tifton, GA. ICMP 451 was obtained from the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, India.

A second pearl millet mapping population, originally developed to segregate for a downy mildew resistance gene on linkage group 4, also segregated for d2. To construct this population, an F2 individual was identified from among the progeny of a genotyped F2 population derived from the cross PT 732B (d2d2) × P1449-2 (D2D2) (Qi et al. 2004) that was heterozygous at the most marker loci on linkage group 4, including the loci that spanned the region carrying both the resistance gene and the d2 gene. Twenty-two F2;3 plants were grown and analyzed with markers for the region of interest on linkage group 4, and a heterozygous F3 plant was selfed to produce a 552 progeny population. In a genetic context, this population, at least in the d2 region, behaves as an F2 population and will be referred to as such. This population was phenotyped for d2 and mapped with restriction fragment length polymorphism (RFLP) markers in early 2000 (Padi 2002). The Padi (2002) study located the d2 gene to a 2.8-cM interval between marker PSMP344 and the cosegregating markers B224CAP2 and RGR1963. Because d2 was scored as a dominant trait (dwarf, d2d2 and tall, D2D2 or D2d2), the precise position of d2 could not be determined. However, of the 19 recombination events that could be allocated, 18 events occurred between PSMP344 and d2, and 1 occurred between PSMP344 and the flanking marker.

Figure 1 Architecture of (A) inbred Tift 23DB (d2d2) and (B) inbred ICMP 451 (D2D2), the parents of the fine-mapping population at flowering time (panicle on main tiller 50% exerted). A 1-m ruler is shown for height comparison. Tift 23DB is ~50% shorter and has a greater leaf-to-stem ratio compared with ICMP 451. (C) Phenotype of the stem of Tift 23DB (left) and ICMP 451 (right) after the leaves were removed from the plants shown in (A) and (B) showing stacking of, in particular, the lower internodes in Tift 23DB compared to ICMP 451.

**Bacterial artificial chromosome (BAC) sequencing and sequence analysis**

Rice RFLP marker RGR1963 was used to screen a pearl millet BAC library (Allouis et al. 2001). Eleven positive clones were identified, of which BAC 293B22 was selected for sequencing. BAC DNA was isolated from clone 293B22 and shotgun libraries prepared as described (Dubcovsky et al. 2001). A total of 1152 subclones were sequenced from both ends using Sanger technology. PHRED, PHRAP, and CONSED were used with default parameters for base calling and quality control, sequence assembly, and contig ordering, respectively. The sequence of BAC 293B22 has been deposited in GenBank under accession number KC463796. Gene prediction was performed using FGENESH with the monocot training set (www.softberry.com). Repetitive DNA was identified by BLASTN searches against the Gramineae repeat database (http://plantrepeats.plantbiology.msu.edu/search.html).

**Markers**

B224C4P2 is a polymerase chain reaction (PCR)-based marker derived from an end-sequence of pearl millet BAC 224C4, which was one of the 11 clones identified after screening a pearl millet BAC library with the rice RFLP marker RGR1963 (Padi 2002). PSM344 and PSM305 are sequence-tagged-site markers derived from RFLP probes PSM305 and PSM344, respectively (Money et al. 1994). Three PCR-based markers, Ca_S07g023840, Ca_S07g023850, and Sb07g023860, were derived from genes identified on BAC 293B22. The prefix Ca stands for *Cenchrus americanus* and is followed by the designation of the sorghum ortholog. In addition, primer sets were developed against 40 genes located in the regions 28.17–28.44 Mb (chromosome end) on rice chromosome 8 and 58.37–59.04 Mb on sorghum chromosome 7. Based on grass comparative data, these rice and sorghum chromosome segments were expected to be syntenic to the pearl millet d2 region (Devos et al. 2000; Devos 2005). Genes selected from the rice and sorghum genomic sequences were used in BLASTN searches to find corresponding expressed sequence tags from other grass species. The genomic and expressed sequence tag sequences were aligned using the program Multalin (Corpet 1988), and primer sets were designed against conserved exon regions flanking an intron. For polymorphism screening and mapping, amplification products were separated on mutation detection enhancement (MDE) acrylamide gels to display single-strand conformation polymorphisms (Martins-Lopes et al. 2001). The primer sequences, annealing temperature and location in the sorghum, rice, and *Setaria italica* genomes for all markers are given in Supporting Information, Table S1. To facilitate interpretation of the data, all markers, irrespective of whether they were developed from rice or sorghum, were named after their sorghum ortholog.

**Genotyping**

A total of 915 F2 individuals from the cross Tift 23DB × ICMP 451 was grown in the glasshouse in batches of 100–200 plants. Genomic DNA was extracted from F2 individuals using an ultraquick DNA extraction protocol (Steiner et al. 1995) and genotyped with markers B224C4P2 and PSM305 on MDE gels to identify recombinants in the d2 region. Plants carrying a recombination event in the d2 region are referred to as “informative plants.” Informative plants were selected to produce F3 seed. High-quality DNA for further genotyping of the informative plants was obtained either from leaves of the F2 plants using a standard CTAB protocol (Murray and Thompson 1980) or from 25 bulked F3 seeds using the protocol described by Busso et al. (2000). All genotyping was done on MDE gels. DNA fragments were visualized by silver staining (Beidler et al. 1982).

PCR amplifications were performed in 20-µL reaction volumes containing 1× PCR buffer (Promega), 1.5 mM MgCl2, 0.25 mM of each dNTP, 0.5 µM of forward and reverse primers, 100 ng of template DNA, and 0.8 U of Taq DNA polymerase (Promega). Amplification conditions consisted of an initial denaturation at 94°C for 3 min, 38 cycles of denaturation at 94°C for 30 sec, primer annealing for 30 sec (see Table S1 for primer-specific annealing temperatures) and extension at 72°C for 1 min followed by a final extension at 72°C for 5 min. For touch-down PCR (primers indicated with a temperature range in Table S1), the annealing temperature was decreased by 1°C every two cycles until the target temperature was reached and 35 PCR cycles were done at the target temperature.

**Phenotyping**

To determine the allelic composition at the d2 locus, 13–25 F2 plants were analyzed for each of the informative F2 plants. Because of space limitations in the glasshouse, phenotyping was done in multiple batches. Plant height was measured at maturity from the basal node to the top of the panicle (Table S2). Families with a median height <90 cm consisted of plants with the dwarf phenotype, and the corresponding F2 genotype was scored as homozygous dwarf (Table 1). Because of the broad range of heights observed for the tall plants, which was caused both by the segregation of other height-affecting genes in the population and environmental effects, we were very conservative in converting the phenotypic measurements to genotypic scores for the d2 locus. Families with a median height >135 cm were considered to be derived from an F2 plant heterozygous at the d2 locus if the number of plants shorter than 110 cm was not significantly different from 25%. The height of 110 cm was chosen as threshold because less than 2% of dwarf plants were ≥110 cm. F2 genotypes giving rise to F3 families with a median height between 90 and 130 cm were considered to be either heterozygous or homozygous for the tall allele and were not further classified (Table 1).

**Comparative analyses**

The protein corresponding to the primary transcript and its location in the genome was retrieved for all annotated genes in the region between 54.31 and 64.31 Mb on sorghum chromosome 7 (JGI v.1.0) (Paterson et al. 2009). Similarly, we retrieved the protein corresponding to the primary transcript for all annotated genes in *S. italica* (35,158 loci in nine chromosomes, JGI v2.1) (Bennetzen et al. 2012), *Oryza sativa* (34,781 representative gene loci in 12 chromosomes, IRGSP build 5) (Matsumoto et al. 2005), and *Brachypodium distachyon* (23,558 loci in five chromosomes, JGI v1.0) (Vogel et al. 2010). In the first instance, a BLASTP search was carried out with the sorghum proteins as queries against the *S. italica* proteins. The top hit was recorded if the E-value was less than 1e−5 and the maximum number of hits at the threshold value was four. Homologous pairs were used to detect syntenic blocks by MCscan (multiple collinearity scan) (Tang et al. 2008) and collinear segments were identified using the empirical scoring scheme min (−log10E, 50) for one gene pair and −1 gap penalty for each 10–kb distance between any two consecutive gene pairs. Syntenic blocks with scores >300 and an E-value <1e−10 were retrieved. *S. italica* proteins located within these syntenic blocks were then used to identify the syntenic regions in rice and *B. distachyon*. The positions of orthologous gene pairs were plotted using a script in R (R Development Core Team 2005), and the dot plots were used to identify rearrangements. The precise breakpoints of the rearrangements were determined manually.
Table 1 | Median height, mean height, and SD of F3 families derived from informative F2 plants, the number of F3 plants per family with height <110 cm, and F2 genotype score

<table>
<thead>
<tr>
<th>F2 plant ID</th>
<th>No. of F3 progeny (batch)*</th>
<th>Median plant height, cm</th>
<th>Mean plant height, cm</th>
<th>SD, cm</th>
<th>Number (percentage) of plants &lt;110 cm²</th>
<th>F2 genotype at d2 locus</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>23 (1)</td>
<td>85</td>
<td>87</td>
<td>13.99</td>
<td>23 (100%)</td>
<td>d2d2</td>
</tr>
<tr>
<td>177</td>
<td>23 (1)</td>
<td>62</td>
<td>63</td>
<td>11.82</td>
<td>23 (100%)</td>
<td>d2d2</td>
</tr>
<tr>
<td>263</td>
<td>25 (1)</td>
<td>104</td>
<td>109.1</td>
<td>16.94</td>
<td>ND</td>
<td>D2D2 or D2d2</td>
</tr>
<tr>
<td>310</td>
<td>25 (1)</td>
<td>160</td>
<td>155.4</td>
<td>22.66</td>
<td>0 (0%)</td>
<td>D2D2</td>
</tr>
<tr>
<td>320</td>
<td>24 (1)</td>
<td>150.5</td>
<td>150.5</td>
<td>21.57</td>
<td>0 (0%)</td>
<td>D2D2</td>
</tr>
<tr>
<td>344</td>
<td>23 (1)</td>
<td>85</td>
<td>86</td>
<td>17.20</td>
<td>21 (91.3%)</td>
<td>d2d2</td>
</tr>
<tr>
<td>349</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>374</td>
<td>23 (2)</td>
<td>89</td>
<td>82.3</td>
<td>18.16</td>
<td></td>
<td>d2d2</td>
</tr>
<tr>
<td>477</td>
<td>25 (1)</td>
<td>61</td>
<td>64.8</td>
<td>16.04</td>
<td>25 (100%)</td>
<td>d2d2</td>
</tr>
<tr>
<td>479</td>
<td>22 (1)</td>
<td>155</td>
<td>153</td>
<td>29.27</td>
<td>1 (4.6%)</td>
<td>ND</td>
</tr>
<tr>
<td>486</td>
<td>24 (1)</td>
<td>107</td>
<td>109.6</td>
<td>30.09</td>
<td>ND</td>
<td>D2D2 or D2d2</td>
</tr>
<tr>
<td>486</td>
<td>15 (3)</td>
<td>130</td>
<td>121.8</td>
<td>23.21</td>
<td>ND</td>
<td>D2D2 or D2d2</td>
</tr>
<tr>
<td>496</td>
<td>23 (1)</td>
<td>142</td>
<td>132.5</td>
<td>28.17</td>
<td>4 (17.4%)</td>
<td>D2D2</td>
</tr>
<tr>
<td>514</td>
<td>21 (2)</td>
<td>121</td>
<td>129.9</td>
<td>34.66</td>
<td>ND</td>
<td>D2D2 or D2d2</td>
</tr>
<tr>
<td>612</td>
<td>23 (2)</td>
<td>154</td>
<td>150.3</td>
<td>24.35</td>
<td>2 (8.7%)</td>
<td>D2D2</td>
</tr>
<tr>
<td>701</td>
<td>15 (2)</td>
<td>181</td>
<td>179.8</td>
<td>23.63</td>
<td>0 (0%)</td>
<td>D2D2</td>
</tr>
<tr>
<td>778</td>
<td>21 (2)</td>
<td>145</td>
<td>145.1</td>
<td>26.81</td>
<td>1 (4.8%)</td>
<td>D2D2</td>
</tr>
<tr>
<td>787</td>
<td>17 (3)</td>
<td>157</td>
<td>150.7</td>
<td>26.51</td>
<td>2 (11.8%)</td>
<td>D2D2</td>
</tr>
<tr>
<td>812</td>
<td>17 (2)</td>
<td>210</td>
<td>203.2</td>
<td>22.54</td>
<td>0 (0%)</td>
<td>D2D2</td>
</tr>
<tr>
<td>900</td>
<td>13 (2)</td>
<td>189</td>
<td>182</td>
<td>23.88</td>
<td>0 (0%)</td>
<td>D2D2</td>
</tr>
<tr>
<td>914</td>
<td>14 (2)</td>
<td>192.5</td>
<td>185.9</td>
<td>47.83</td>
<td>2 (14.3%)</td>
<td>D2D2</td>
</tr>
<tr>
<td>924</td>
<td>22 (2)</td>
<td>186</td>
<td>186.4</td>
<td>24.49</td>
<td>0 (0%)</td>
<td>D2D2</td>
</tr>
<tr>
<td>930</td>
<td>18 (2)</td>
<td>206</td>
<td>208.4</td>
<td>28.34</td>
<td>0 (0%)</td>
<td>D2D2</td>
</tr>
</tbody>
</table>

*a* F3 families with the same batch number were grown concurrently.

*b* Except where the percentage of plants <110 cm is 0% or 100%, SD from 25% at P ≤ 0.05 is indicated with an asterisk (*).

*c* Not determined in F3 families with a median height between 90 and 135 cm.

Expression analyses

The mapping parents ICMP 451 and Tift 23DB were grown under greenhouse conditions for 7 to 9 wk. The top-most internode and its corresponding nodal leaf were harvested when 50% of the panicle had emerged from the flag leaf sheath. The harvested material was immediately frozen in liquid nitrogen and, if needed, stored at −80°C. Total RNA was extracted from leaves and internodes with TRIzol reagent (Chomczynski and Sacchi 1987) and approximately 5−10 μg of isolated RNA was treated with DNase using the Ambion TURBO DNA-free kit. cDNA synthesis was performed on ~1 μg of DNase treated RNA using the Roche Transcriptor First Strand cDNA Synthesis kit. The manufacturer’s protocol was used for all experimental steps involving kits. PCR conditions for semi-quantitative PCR consisted of an initial denaturation at 95°C for 3 min followed by 29 cycles of denaturation at 95°C for 30 sec, annealing at 59°C for 30 sec, extension at 72°C for 1 min, and a final extension at 72°C for 3 min. Primers Ca_Sb07g023730F10 (5’-CTGGAGGCACTTCACTTGAC-3’) and Ca_Sb07g023730F10 (5’-CTGGAGGCACTTCACTTGAC-3’), designed against gene Sb07g023730 (dw3), were used to study expression of the pearl millet dw3 ortholog, whereas actin primers ActinF (5’-ACC GAGCCTCTAACC-3’) and ActinR (5’-GTATGGCTGACAC CATCACC-3’) were used as internal controls (Van den Berg et al. 2004). The expression analysis was conducted twice with samples collected from plants grown at different times of the year.

RESULTS

Identifying recombinants in the d2 region

Mapping of plant height in the PT 732B x P1449-2 population had shown the d2 gene to be located between the markers B224C4P2 and PSMP344 (Padi 2002). The sequence-tagged-site marker PSMP344 (primer set PSMP344F/R) derived from RFLP marker PS344 did not amplify from ICMP 451 and thus could only be scored as a dominant marker in the Tift 23DB × ICMP 451 population. Hence, PSMP305, which coseggregates with PSMP344 in most pearl millet maps (Qi et al. 2004), was used in combination with B224C4P2 to identify F2 plants that carried a recombination event in the d2 region. Genotyping of 915 F2 plants from the cross Tift 23DB × ICMP 451 yielded 29 recombinants, providing an estimate of 1.6 cm for the genetic distance between B224C4P2 and PSMP305. Five plants did not survive the seedling stage, and one plant was removed because of the presence of non-parental alleles, so the fine-mapping and phenotyping was performed on 23 recombinants.

Determining the genotype at the d2 locus

The height of the F2:3 plants varied both within and between F3 families and ranged from 37 to 270 cm (Table S2). Five F3 families comprised only dwarf plants, and the corresponding F2 plants were genotyped as d2d2 (Table 1). Five F3 families were segregating for plant height in a 3:1 (tall:dwarf) ratio, and the corresponding F2 plants were genotyped as d2D2 (Table 1). Nine F3 families contained either no plants <110 cm (7 families) or a single plant <110 cm (2 families) and the corresponding F2 plants were genotyped as D2D2. Integrating these data with the genotypic data for markers B224C4P2 and PSMP305 confirmed the location of d2 in the B224C4P2–PSMP305 interval.

Marker development and fine-mapping

Marker RGR1963, which corresponds to sorghum gene Sb07g023850, the two additional markers developed from BAC clone 293B22 that was identified with RGR1963, and the primers designed against selected
genes on rice chromosome 8 and sorghum chromosome 7 were tested on Tift 23DB, ICMP 451, and 4 recombinant progeny for their ability to amplify and to detect variation. This initial screen also allowed us to identify markers that would likely map to the d2 region based on their segregation pattern in the four recombinant lines. Seventy-five percent of the primer sets amplified well in pearl millet, and 30% were polymorphic in the mapping population. Four markers, Ca_Sb07g023460, Ca_Sb07g023470, Ca_Sb07g023500, and Ca_Sb07g023740 had segregation patterns in the four recombinant progeny that were inconsistent with their location in the d2 region and were not further analyzed. Eight markers corresponding to sorghum genes Sb07g023430, Sb07g023440, Sb07g023520, Sb07g023630, Sb07g023810, Sb07g023840, Sb07g023850, and Sb07g023910 were mapped in the full set of 23 informative progeny. All markers cosegregated (Figure 2A, Table S3). Following the development of a new primer set for PSMP344 (PSMP344F2/R2; Table S1) which amplified in both ICMP 451 and Tift 23DB, PSM344 was mapped between PSMP305 and the cluster of cosegregating markers. Placement of d2 relative to the fine-mapped markers showed, with the exception of one double recombination event in the d2 score, complete cosegregation of the d2 phenotype with the 8-marker cluster (Table S3). The F2 family derived from the F1 line showing the double recombination event (F1 plant with ID 612, Table 1) consisted of 23 plants, 2 of which were 2 and 4 cm shorter than the threshold of 110 cm. Furthermore, although 2:1 did not deviate significantly from a 3:1 ratio, the P-value was close to the 5% significance threshold (P = 0.071). It seems therefore likely that the genotype for this plant was D2D2 rather than D2d2. When this hypothesis was taken into consideration, we found that the d2 phenotype cosegregated with the eight-marker cluster (Figure 2A, Table S3).

In an attempt to order the newly generated markers in the pearl millet genome, we selected 16 of the 29 individuals from the PT 732B mapping population with the recombination event in the interval B224C4P2–PSMP344 for which F2 seed was available. Bulked F2 seed was grown and seedlings were used for DNA extraction. Mapping of RGR1963, Ca_Sb07g023430, Ca_Sb07g023520, Ca_Sb07g023630, Ca_Sb07g023810, and Ca_Sb07g023910 and one additional sorghum-derived marker, Ca_Sb07g024020, in the 16 informative plants identified two marker clusters (Figure 2B, Table S4). One cluster cosegregated with B224C4P2 and comprised the markers RGR1963, Ca_Sb07g023910, Ca_Sb07g024020, and Ca_Sb07g023810, which, in sorghum, are located in the interval 58.78–59.04 Mb on chromosome 7. The second cluster comprised d2, and markers Ca_Sb07g023630, Ca_Sb07g023520, and Ca_Sb07g023430, which were derived from region 58.37–58.54 Mb on sorghum chromosome 7 (Figure 2B, Table S4).

**Comparative analysis of the d2 region**

If we assume that the d2 region is completely colinear in the pearl millet and sorghum genomes, the ortholog of d2 should be present in the sorghum genome proximal to location 58.78 Mb (distal boundary). However, our mapping data did not allow us to determine the proximal boundary of the d2 region. A number of RFLP probes, including PSM344 and PSM305, had previously been end-sequenced (Money et al. 1994). The two end-sequences of PSM344, which is a 2-kb probe, mapped 24.7 kb apart on sorghum chromosome 7 (locations 12.59 Mb and 12.61 Mb). For PSM305, one end does not have homology in sorghum at an e-value threshold e<sup>-1e-09</sup> whereas the other end identifies sequences on all 10 sorghum chromosomes. We also assessed the location in sorghum of PSM364, which had previously been mapped, depending on the cross, 1.8–6.1 cM distal

![Figure 2](image-url)
of species-specific inversions were observed. The distal region of sorghum chromosome 7 (from 58.36 Mb–end) is inverted relative to the other three species (Figure 4, Table S5). This places Si0112129m, the foxtail millet ortholog of marker Ca_Sb07g023430, as the most distal marker on foxtail millet chromosome VI for which an ortholog is present on sorghum chromosome 7. In rice, the ortholog of Ca_Sb07g023430 is present on rice chromosome 12 and the rice ortholog to Si0115189m, the proximal neighbor of Si013129m, is the last marker on rice chromosome 8 with an ortholog on sorghum chromosome 7. In B. distachyon, the ortholog of Ca_Sb07g023430 marks the breakpoint of an ancestral chromosome fusion event. Other rearrangements include an inversion of the region 35.19 Mb in rice, and two inversions spanning the regions 40.52–40.80 Mb and 41.65–41.74 Mb in B. distachyon. None of these inversions correspond to the inversion that differentiates pearl millet from sorghum.

Haplotype of the d2 region in three tall and three dwarf lines
We used the mapped markers to determine the allelic configuration at each of the loci in three dwarf lines, Tift 23DB, PT 732B, and 81B, and three tall lines, ICMP 451, P1449-2, and Tift red (Table 2). At all loci distal to Ca_Sb07g023430, the three dwarf lines carried the same allele, referred to as “a,” whereas the tall lines carried alleles that were different (either “b” or “c”; Table 2) than those observed in the dwarfs. The only exception was at locus Ca_Sb07g023910, where the tall Tift red appeared to have the same allele as the dwarfs. The differentiation between dwarf and tall haplotypes was lost at the three most proximal markers analyzed, Ca_Sb07g023430, PSMP344, and PSMP305.

Use of comparative information to identify a putative candidate gene for d2
Combining the mapping information with the haplotype data yielded Ca_Sb07g023810 and Ca_Sb07g023430 as the distal and proximal boundary, respectively, of the d2 region. These two markers defined a 410-kb region in sorghum in which 40 genes had been annotated (www.phytozome.net; Sh1.4 gene set). The genes, together with their location and functional annotation, are given in Table S6. Sixty-three percent of the 40 genes had no functional annotation. When we focused on the remaining 15 that had homology to characterized proteins, we found that Sb07g023730 became a likely candidate for d2 because it had previously been identified as the gene underlying the dw3 and br2 dwarf phenotypes in sorghum and maize, respectively (Multani et al. 2003). Sb07g023730, an ABC transporter of the B subfamily (member 1), encodes a P-glycoprotein that modulates auxin transport in the stalk (Multani et al. 2003; Knoller et al. 2010).

Preliminary validation of ABCB1 as a candidate for d2
Several forward and reverse primers were designed against the sequence of gene Sb07g023730. One primer combination, Ca_Sb07g023730F1/R5 (Table S1), yielded a strong amplification product in the tall inbreds ICMP 451, P1449-2 and Tift red but did not amplify in the dwarf inbreds Tift 23DB, PT 732B and 81B. Sanger sequencing of the fragment amplified from the tall inbred ICMP 451 (File S1) and BLASTX analysis of the resulting sequence to the ‘nr’ section of GenBank confirmed that the amplified fragment was derived from gene ABCB1 (96% identity with both the sorghum and maize ABCB1 protein). Mapping of this product in the informative progeny of the Tift 23DB × ICMP 451 and PT 732B × P1449-2 F₂ populations showed Ca_Sb07g023730 to fall within the cluster of markers that cosegregated with d2 (Figure 2, A and B). This finding demonstrated that the pearl millet ortholog of Sb07g023730 was located within the d2 interval.

A second primer set Ca_Sb07g023730F10/R10 gave strong amplification products in both ICMP 451 and Tift 23DB. Sequence analysis showed that the ICMP 451 and Tift 23DB fragments differed by a single synonymous SNP (File S2). Because this SNP could not be visualized by single-strand conformation polymorphism gel electrophoresis, the amplicon obtained with primer set Ca_Sb07g023730F10/R10 was not mapped. A BLASTX analysis, however, confirmed that the Ca_Sb07g023730F10/R10 amplification product corresponded to ABCB1 (92% identity with both the sorghum and maize ABCB1 protein). This primer set, which flanked introns 2 and 3 in Sb07g023730,
was used in a semiquantitative reverse transcriptase (RT)-PCR experiment and showed that Ca_Sb07g023730 was differentially expressed in both the top internode and corresponding leaf between ICMP 451 (D2D2) and Tift 23DB (d2d2) (Figure 5). The expression data provided support for our hypothesis that Sb07g023730 is the d2 gene.

DISCUSSION

Organization of the d2 region
To our knowledge, no traits have been fine-mapped in pearl millet. The d2 dwarfin gene, despite its widespread incorporation into commercial germplasm, had previously only been located to a 23.2-cM interval on pearl millet linkage group 4 (Azhaguvel et al. 2003). We initially mapped d2 to a 1.6-cM region, but attempts to further narrow the interval in an approximately 1000 progeny F2 population largely failed due to a lack of recombination. Markers developed from a 670-kb region on sorghum chromosome 7 (58.37–59.04 Mb) all cosegregated with the d2 phenotype. Recombination in pearl millet is very unevenly distributed (Liu et al. 1994; Qi et al. 2004) and it might be that the d2 region has inherently low recombination rates. An alternative explanation is that the two mapping parents, ICMP 451 and Tift 23DB, differ by an inversion in this region. Combining sequence information from a BAC clone originating from the d2 region with recombination data on the four markers that were identified on this BAC indicated the presence of an inversion in the d2 region in Tift 23DB relative to sorghum (Figure 3). Although we could not precisely determine the boundaries of the inversion, the fact that markers Ca_Sb07g023810 and PSR492 (which is orthologous to Sb07g024860) were located outside the inversion suggests that the inversion likely encompasses less than 100 genes. As the pearl millet genome is characterized by a large number of chromosomal rearrangements relative to other grass genomes (Devos et al. 2000), it was not particularly surprising to observe this inversion at the d2 locus. However, it is unclear whether the inversion is present in all pearl millet lines or if it is limited to the d2 dwarfs or, possibly, even the inbred Tift 23DB. We therefore cannot exclude the possibility that the lack of recombination seen in this region between Tift 23DB and ICMP 451 is the result of the differential presence of this rearrangement in the two parental lines. The pattern of recombination seen in the PT 732B × P1449-2 mapping population was consistent with both overall reduced recombination and the presence of an inversion in one of the parents, and hence did not provide further insights into the specific organization of the d2 region.

Comparative analysis of the d2 region
Comparative information has been used to develop markers for specific chromosome regions and, in some cases, to identify candidate genes underlying traits (Kilian et al. 1997; Yan et al. 2003, 2004). In

Table 2 Allele composition at 12 loci in three tall and three dwarf inbred lines

<table>
<thead>
<tr>
<th>Marker</th>
<th>Tall Inbreds</th>
<th>Dwarf Inbreds</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ICMP 451</td>
<td>81B</td>
</tr>
<tr>
<td></td>
<td>P1449-2</td>
<td>23DB</td>
</tr>
<tr>
<td>B224C4P2</td>
<td>b</td>
<td>b</td>
</tr>
<tr>
<td>Ca_Sb07g023840</td>
<td>b</td>
<td>b</td>
</tr>
<tr>
<td>RGR1963</td>
<td>b</td>
<td>b</td>
</tr>
<tr>
<td>Ca_Sb07g023910</td>
<td>b</td>
<td>a</td>
</tr>
<tr>
<td>Ca_Sb07g024020</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Ca_Sb07g023810</td>
<td>b</td>
<td>c</td>
</tr>
<tr>
<td>Ca_Sb07g023630</td>
<td>b</td>
<td>c</td>
</tr>
<tr>
<td>Ca_Sb07g023520</td>
<td>b</td>
<td>b</td>
</tr>
<tr>
<td>Ca_Sb07g023440</td>
<td>b</td>
<td>b</td>
</tr>
<tr>
<td>Ca_Sb07g023430</td>
<td>b</td>
<td>b</td>
</tr>
<tr>
<td>PSMP344</td>
<td>b</td>
<td>b</td>
</tr>
<tr>
<td>PSMP305</td>
<td>b</td>
<td>b</td>
</tr>
</tbody>
</table>

ND, no data.
Two of those, Sb07g023460 and Sb07g023470, are found in collinear scored pearl millet fragments and the sorghum genes used for attempt by sequence analysis to establish orthology between the ancestral grass genome. The ancestral distal position was main-
tire distal region of sorghum chromosome 7 had undergone an in-

tersection at the centromere to the telomere (Qi et al. 2004).

Although all the inversions observed were species-specific, it is interesting to note that the distal inversion in sorghum and the 23.82–23.97 Mb inversion in rice have one of their breakpoints in common, suggesting that the breakpoint might represent a region on the ancestral grass chromosome that is prone to breakage.

**ABC1 as a candidate for d2**

Our mapping data had indicated that the distal boundary for the d2 region in sorghum was at location 58.78 Mb on sorghum chromosome 7, but we had not been able to establish a proximal boundary due to cosegregation of the markers with the d2 phenotype. However, haplotype analysis of three tall and three dwarf inbred lines with the mapped markers suggested that the d2 gene was located distal to marker Ca_Sb07g023430, whose ortholog is located at position 58.37 Mb in sorghum. In the region spanned by the markers B224C4P2 and Ca_Sb07g023440, the allelic composition of the dwarf inbreds was identical and different from that of the tall inbreds, except at locus Ca_Sb07g023910 in line Tift red (Table 2). This finding suggests that the three dwarfs analyzed (Tift 23DB, PT 732B, and 81B) are derived from the same d2 source. As expected, we find different haplotypes in this region in the tall inbreds ICMP 454, P1449-2, and Tift red (Table 2). The inbred 81B is a downy mildew resistant selection from gamma-irradiation treated Tift 23DB (Rai and Hanna 1990b). Although this line maintains the dwarf haplotype in the d2 region, it likely underwent a recombination event with a tall line between markers Sb07g023430 and PSMP344 (Table 2). Tift red is a backcross line produced by the late Glenn Burton that carries a gene for purple plant color and the tall D2 allele in a Tift 23 background. Considering that no recombination was identified between Ca_Sb07g023430 and Ca_Sb07g023440 in the ~1500 progeny we analyzed from two crosses, it was surprising to see that Tift 23DB and Tift red, which are near-isogenic lines, differ from a recombination event between those two markers. The unexpected haplotype of Tift red was crucial to determining the proximal boundary of the d2 region.

The region delineated in sorghum as being orthologous to the d2 region in pearl millet contained the adenosine triphosphate–binding cassette (ABC) subfamily B1 gene, an obvious candidate for d2 because a mutation in this gene was shown to underlie the recessive sorghum dw3 dwarfing phenotype (Multani et al. 2003). A mutation in the same gene is also responsible for the brachytic2 (br2) phenotype in maize. The ABCB1 protein belongs to the multidrug resistant class of P-glycoproteins and plays a role in auxin transport in the nodal/intercalary meristem regions (Knoller et al. 2010). Consequently, in the br2/Zmpgg1 and dw3/Sbpg1 mutants, auxin accumulates in the vicinity of the nodes. Because auxin is synthesized mainly in the shoot apex and young leaves, and then transported basipetally, the lowermost internodes in the sorghum dw3 and maize br2 mutants are affected the most by the modulation of auxin transport caused by a knockout of the ABC1 gene (Multani et al. 2003; Knoller et al. 2010). The phenotype of stacked lower internodes in the pearl millet dw2 dwarf is very similar to that observed in sorghum dw3 and maize br2 mutants (Figure 1C).

In an attempt to validate ABC1 as a candidate for d2, we designed multiple primer sets against the sorghum ABC1 gene and tested them in three dwarf and three tall lines. One primer set, which spanned intron 1, yielded an amplification product in all tall lines tested and in none of the dwarfs, and this polymorphism cosegregated.

**Figure 5** Semiquantitative RT-PCR with primers designed against sorghum gene Sb07g023730 showing a 564-bp fragment in cDNA extracted from leaves and internodes of ICMP 451 (D2D2) (lane a) and no/a weak fragment of the same size in cDNA extracted from leaves and internodes of Tift 23DB (d2d2) (lane b). Lane c shows the 841-bp fragment obtained using genomic DNA from the tall inbred ICMP 451. Primers homologous to an actin gene were used as an internal control.

Although gene orders were overall highly conserved between the regions orthologous to d2 in foxtail, sorghum, rice, and B. distachyon, species-specific inversions were identified in all four species. The entire distal region of sorghum chromosome 7 had undergone an inversion with a breakpoint between 58.34 and 58.36 Mb, which meant that the genes that were located immediately distal to the inversion breakpoint region in sorghum had been located near the telomere in the ancestral grass genome. The ancestral distal position was main-
tained only in foxtail millet and rice. In sorghum, B. distachyon, and pearl millet, chromosomal rearrangements had moved the ancestral telomere to an interstitial position. This almost certainly had been accompanied by a reduction in recombination, in particular in pearl millet, where a very strong recombination gradient exists along the chromosomes from the centromere to the telomere (Qi et al. 2004).
with the height phenotype in the set of informative F2 plants in both the Tift 23DB × ICMP 451, and PT 732B × P1449-2 populations. The most likely cause for the lack of amplification in the dwarf lines is either a single-nucleotide polymorphism or deletion at a primer site that prohibits primer extension or the presence of an insertion in the region between the two primer binding sites that extends the fragment to be amplified beyond the length limit of a typical PCR. More work is needed to determine whether the observed variation could be the underlying cause of the dwarf phenotype.

We also analyzed the expression of the ABCB1 gene in the topmost internode and the corresponding leaf in flowering plants of one d2 dwarf plant, Tift 23DB, and one tall plant, ICMP 451, using semi-quantitative RT-PCR. The RT-PCR yielded a strong amplification product in both tissues in ICMP 451 and a weak product in Tift 23DB, which suggests that ABCB1 is differentially expressed in the tall and dwarf inbreds in both tissues or that the stability of the ABCB1 mRNA is reduced in the dwarf mutant (Figure 5). A reduced transcript level in the dwarf mutant is in agreement with the recessive nature of the d2 mutation. In maize, ABCB1 is expressed in nodule tissue and possibly in internodes, although reports on the latter are not consistent (Maultani et al. 2003; Knoller et al. 2010). No expression was detected in maize leaves (Maultani et al. 2003). Nodes were not included in our preliminary expression analysis because of the difficulty of extracting RNA from the hard node tissue. In Arabidopsis, ABCB1 expression is highest in nodes, but the gene is also expressed in a range of other tissues (Titaipwatanakun and Murphy 2009). More detailed expression analyses are needed in pearl millet to determine precisely where the ABCB1 gene is expressed and at what levels. The greater expression in the tall compared with the dwarf inbred, however, provided some support for our hypothesis that ABCB1 is a reasonable candidate for d2.

In conclusion, using a combination of genetic mapping, haplotype analysis and comparative genomics, we have fine-mapped the pearl millet d2 dwarf phenotype to a region which, in sorghum, spans 410 kb and contains 40 annotated genes. A candidate gene, ABCB1, was identified as putatively underlying d2. Work is currently underway to isolate full-length copies of the ABCB1 gene from both a tall and a dwarf inbred to further test our candidate gene hypothesis. In the meantime, our work provides breeders with a set of markers that can be used to identify the presence of the recessive d2 gene in heterozygous condition and at the seedling stage. Phenotypically, the dwarf phenotype can only be scored in homozygous condition and at the booting stage, so the markers will enhance the efficiency of breeding programs that use tall lines as sources of novel genes for the improvement of dwarf inbreds.

ACKNOWLEDGMENTS

We thank W. Hanna for providing seeds of Tift 23DB and Tift red, and generating the Tift 23DB × ICMP 451 F2 mapping population. Early work on the PT 732B × P1449-2 cross was carried out by F.K.P., S.F., and K.M.D. at the Comparative Genomics Unit, John Innes Centre, Norwich, UK. The research was funded by fellowships from the Pioneer HiBred—Generation Challenge Program (to V.J.), the Kirkhouse Trust (to N.N.), and by a National Institute of Food and Agriculture (NIFA) 1890-CSREES Capacity Building Grants (grant 2008-38814-04740; PI: B. Singh, Fort Valley State University, GA).

LITERATURE CITED


Yan, L., A. Loukoianov, A. Blechl, G. Tranquilli, W. Ramakrishna et al., 2004 The wheat VRN2 gene is a flowering repressor down-regulated by vernalization. Science 303: 1640–1644.

Communicating editor: D. Zamir