Novel Grain Weight Loci Revealed in a Cross between Cultivated and Wild Sorghum

Yongfu Tao,* Emma Mace, Barbara George-Jaeggli, Colleen Hunt, Alan Cruickshank, Robert Henzell, and David Jordan*

Abstract
Grain weight has increased during domestication of cereals. Together with grain number it determines yield, but the two are often negatively correlated. Understanding the genetic architecture of grain weight and its relationship with grain number is critical to enhance crop yield. Sorghum is an important food, feed, and biofuel crop well-known for its adaptation to drought and heat. This study aimed to dissect the genetic basis of thousand grain weight (TGW) in a BC1F5 population between a domesticated sorghum accession and its wild progenitor, *Sorghum bicolor* subsp. *verticilliflorum*, and investigate its relationship with grain number. Thousand grain weight, grain number, and yield were measured in field trials in two successive years. A strong negative correlation between TGW and grain number was observed in both trials. In total, 17 TGW quantitative trait loci (QTL) were identified, with 11 of them exhibiting an opposing effect on grain number, implying the correlation between TGW and grain number is due to pleiotropy. Nine grain size candidate genes were identified within 6 TGW QTL, and of these 5 showed signatures of selection during sorghum domestication. Large-effect QTL in this study that have not been identified previously in cultivated sorghum were found to contain candidate genes with domestication signal, indicating that these QTL were affected during sorghum domestication. This study sheds new light on the genetic basis of TGW, its relationship with grain number, and sorghum domestication.

Core Ideas
- Quantitative trait loci for thousand grain weight were mapped in a cross between domesticated and wild sorghum.
- The majority of thousand grain weight quantitative trait loci were negatively associated with grain number.
- Novel large-effect thousand grain weight quantitative trait loci possibly related to sorghum domestication were identified.
- Candidate genes with domestication signal were identified within the large-effect quantitative trait loci.

CEREAL CROPS, including maize, rice, wheat, barley, and sorghum are the most agronomically and economically important species, collectively feeding over two thirds of the world population (Sands et al., 2009). Sorghum provides staple food for over 500 million people in the semiarid tropics of Africa and Asia, in addition to being an important source of feed for livestock. Among the cereals, sorghum is one of the best adapted to drought and high temperatures and hence will play an increasingly important role in meeting the challenges of feeding the world’s growing population under a changing climate.
Grain yield in cereals is determined by grain number per unit land area and average grain weight. Negative correlations between grain number and grain weight have been observed in many annual crops (Griffiths et al., 2015; Jakobsson and Eriksson, 2000; Peltonen-Sainio et al., 2007; Sadras, 2007) including sorghum (Burrow et al., 2014; Heinrich et al., 1983; Yang et al., 2010). The negative association between grain weight and grain number has a profound impact on breeding strategies for increased yield. In crops that tiller, such as wheat, sorghum, and rice, grain weight has a limited range due to evolutionary and early agronomic selection, while grain number is more plastic. Therefore, grain number is the primary determinant of grain yield and the target for more recent crop yield improvement (Borrell et al., 1999; Gambin and Borras, 2012; Griffiths et al., 2015; Sadras, 2007). Intriguingly, recent studies in sorghum and wheat suggest increasing one yield component without reducing the other is possible (Boyles et al., 2016; Gambin and Borras, 2012; Griffiths et al., 2015). If the genetic and physiological basis for the negative association between grain weight and grain number could be better understood, then more effective selection for increasing both grain weight and grain number may be possible.

Understanding the domestication of sorghum can shed light on the genetic architecture of these two traits. The artificial selection process has changed grain size and grain number to adapt to human needs, as well as retention of grain in the panicle (non-shattering), increased carbohydrate content, determinate growth and reduced dormancy (Harlan, 1975). Accompanying these changes is a local reduction in genetic diversity around these major genes underlying “domestication syndrome” in the emerging domesticates, which may be further diminished by reproductive isolation of domesticates from wild populations. The QTL mapping in populations including crosses with wild relatives provides a useful tool to identify genetic regions and novel alleles to further improve traits that have lost genetic diversity during the domestication process, such as grain size.

Given its agronomic and evolutionary importance, grain weight has been a major target for genetic research and improvement practice in many crops. In sorghum, the genetic basis of grain weight has been studied in multiple linkage analysis studies (Brown et al., 2006; Feltus et al., 2006; Murray et al., 2008; Paterson et al., 1995; Pereira et al., 1995; Rami et al., 1998; Srinivas et al., 2009; Tuinstra et al., 1997) which together identified 12 unique genomic regions (Mace and Jordan, 2011). More recently, sorghum diversity panels have been used to identify loci significantly associated with grain weight and other grain yield component traits (Boyles et al., 2016; Zhang et al., 2015). However, the genetic basis underlying the change of grain size during domestication remains unclear, as these studies are predominantly focused on cultivated sorghum.

In the present study we used a bi-parental backcross population between an elite grain sorghum from a modern breeding program and a wild sorghum from the progenitor sub-species *Sorghum bicolor verticilliflorum*; to 1) map QTL associated with TGW; 2) investigate the relationship between grain weight and grain number; and 3) identify candidate genes within these TGW QTL regions and search for evidence of selection during domestication.

**MATERIALS AND METHODS**

**Population Development**

The population used in this study was developed by crossing an accession of the wild sub-species *S. bicolor* subsp. *verticilliflorum* with R931945–2-2, an elite grain sorghum line from the sorghum pre-breeding program (Department of Agriculture and Fisheries/University of Queensland/GRDC) and backcrossing the resulting F1 to R931945–2-2 to produce a BC1F1 population of more than 1000 individuals (Alam et al., 2014). The resulting individuals were grown out in a single row and approximately 300 individuals were selected on the basis of characteristics including plant height, maturity and absence of shattering, critical for adaptation to cultivation in Australian sorghum environments and self-pollinated. The resulting BC1F2 families were grown out in rows and a single plant was selected using the same criteria. The process of single plant to row selection was continued to the BC2F1 generation. At this stage, bulk pollen from each row was used to produce testcross hybrids on an elite cytoplasmic male sterile female (A1*F_B923171). In the end, a total of 217 F1 hybrids were produced for phenotyping. The process of backcrossing to the elite line (R931945–2-2) and selection for adaptation to a cultivated environment produced lines with an average of 80% recurrent parent genome. These lines were crossed to an elite line (A1*F_B923171) to produce a test cross hybrid. The resulting test cross hybrids had an average of 10% of their genome derived from *S. bicolor* subsp. *verticilliflorum*. Due to the relatively low proportion of the wild parent genome present in each test cross hybrid, they were generally phenotypically similar to cultivated hybrids. For example, although the wild parent exhibited very high levels of tillering, the highest tiller number per plant observed in the F1 hybrids in any of the trial years did not significantly exceed the range of tillering in the commercial hybrids that were grown in the same trial. Similar observations were also made for the traits height and maturity.

**Field Trials**

Phenotypic evaluation of F1 hybrids was conducted in two field trials in the 2003/04 and 2004/05 summer seasons at Hermitage Research Facility (28°12′S, 152°5′E, 470 m above sea level) near Warwick, Queensland, in eastern Australia. Both trials were conducted on a strongly cracking and self-mulching alluvial clay with a high montmorillonitic clay content, which has good water-holding capacity. The first field trial (HRF04) was sown on 10 December 2003 and the second field trial (HRF05) was sown on the 26 November 2004. A
population of 100,000 plants per ha was established for both trials. A row column design with two replications was used, each plot consisting of two 6-m rows with a row spacing of 0.75 m. Seed availability meant that different numbers of experimental hybrids were grown in each trial (200 hybrids for HRF04 and 179 hybrids HRF05). In addition, the hybrid of the recurrent parent R931945–2–2 was grown in each trial, as were three commercial hybrids. Standard agronomic practices and pest-control practices were applied and trials were grown on stored sub-soil moisture and in-crop rainfall only without any supplementary irrigation. In the HRF05 trial, mild to severe water stress developed due to a period of very little rainfall from the late vegetative phase until the first half of grain filling when it was relieved by rain.

**Phenotypic Measurements**

In each year the trial plots were harvested using a small-plot harvester (KEW Harvester, Kingaroy Engineering Works, Kingaroy, Australia). The harvested grain of each plot was retained and two samples of 500 seeds were counted, weighed and averaged to calculate TGW. Grain number was calculated by dividing the plot yield by the mass per seed. Grain yield was measured as machine-harvested yield expressed in t/ha.

**Statistical Analysis of the Field Trials**

Each trait was analyzed separately using a linear mixed model for the data from both trials combined. The general form of the mixed models used was:

$$ y = X\beta + Z\mu + \epsilon $$

where the response (vector $y$) is modeled by a set of fixed effects (vector $\beta$) and random effects (vector $\mu$ and $\epsilon$). The design matrices $X$ and $Z$ assign the fixed and random effects respectively to the observations.

In our model, vector $\beta$ comprised the main effect for trial, vector $\mu$ comprised the genotype effects for each trial using a correlated genetic variance structure plus Replicate and vector $\epsilon$ error.

Both trials were assessed for possible spatial effects due to extraneous field effects and neighbor effects and were included in the model as necessary.

The difference between trials for each phenotypic trait was assessed using a Wald test for the fixed trial effect in each model. Generalized heritability was calculated using the average standard error and genetic variance for each trial and trait combination following the methods proposed by Cullis et al. (2006). Best linear unbiased estimators (BLUEs) were predicted for each genotype within each trial using the same linear mixed model as above but fitting the trial × genotype term as a fixed effect. Predicted BLUEs for each trial/trait combination were correlated using a Pearson correlation.

Between-trial comparisons were made for the grain number and TGW relationship by fitting a linear regression model to assess the interaction between trial and regression slope. A series of linear regression models was also used to assess the relationship between yield and combinations of grain number and TGW. All statistical analyses were conducted using R (www.R-project.org). Linear mixed models were fitted using the ASRemL-R package (Butler et al., 2009).

**Genotyping**

Genotyping of the BC$_F_i$ population was conducted based on DNA extracted from bulked young leaves of five plants of each BC$_F_i$ as described by DArT (Diversity Arrays Technology) P/L (DArT, www.diversityarrays.com). The samples were genotyped following an integrated DArT and genotyping-by-sequencing methodology involving complexity reduction of the genomic DNA to remove repetitive sequences using methylation sensitive restriction enzymes prior to sequencing on Next Generation sequencing platforms (DArT, www.diversityarrays.com). The sequence data generated were then aligned to the most recent version (v3.1.1) of the sorghum reference genome sequence (Paterson et al., 2009) to identify SNP (Single Nucleotide Polymorphism) markers and the genetic linkage location predicted based on the sorghum genetic linkage consensus map (Mace et al., 2009).

**Trait-Marker Association and QTL Analysis**

Although the population analyzed was a backcross population, the imposed selection during the development of the mapping population prevented standard bi-parental QTL mapping approaches from being applied. Instead we used a multistep process to identify TGW QTL. Single-marker analysis was conducted to calculate the significance of each marker-trait association using predicted BLUEs, followed by two strategies to identify QTL. In the first strategy, SNPs associated with TGW were identified based on a minimum $P$-value threshold of $< 0.01$ and grouped into genomic regions based on a 2-cM (centimorgan) window, while isolated markers associated with the trait were excluded. Identified genomic regions in this step were designated as high-confidence QTL. In the second strategy, markers associated with TGW were identified based on a minimum $P$-value threshold of $< 0.05$. Again, a sliding window of 2 cM was used to group identified markers into genomic regions while isolated markers were excluded. Identified regions in this strategy were then compared with association signals reported in recent association mapping studies (Supplemental Table S1) (Boyles et al., 2016; Upadhyaya et al., 2012; Zhang et al., 2015). Genomic regions with support from either of these previous studies were designated as combined QTL. Previous bi-parental QTL studies were not considered here as the majority of them used very small populations (12 with population size $< 200$ individuals, 9 with population size $< 150$ individuals), thus ended up with generally large QTL regions. These GWAS studies sampled a wide range of sorghum diversity, and identified SNPs associated with grain weight. A strict threshold of 2 cM was used to identify co-location of GWAS hits and genomic regions identified in the second strategy. As single-marker
analysis is prone to produce false positive associations due to the problem of multiple testing, only regions with multiple signal support at the \( P < 0.05 \) level and additional evidence from previous studies were considered.

The effects of QTL on TGW were analyzed using a linear mixed model with all QTL included simultaneously as fixed factors. Association of TGW QTL with grain number was tested by performing single-marker analysis of every SNP within TGW QTL. Thousand grain weight QTL with markers associated with grain number were selected and fit into a linear mixed model to calculate these TGW QTL’s effects on grain number.

**Previously Identified QTL and Candidate Genes for Grain Size in Sorghum**

Grain weight QTL were collated from 18 previously published studies, and their locations were projected onto both the sorghum reference genome v3.1.1 and the sorghum genetic linkage map following methods described in Mace and Jordan (2011). To be conservative, QTL with confidence interval > 20 cM were excluded from further analysis.

We further investigated whether any of the TGW QTL identified in this study co-located with 114 grain size candidate genes previously identified by Tao et al. (2017).

**RESULTS**

**Phenotypic analysis**

Thousand grain weight, grain number, and yield were normally distributed and showed substantial variation in both trials (Supplemental Table S2). However, the two non-irrigated field trials experienced contrasting environmental conditions due to different amounts of in-crop rain fall. While HRF04 received 513 mm of in-crop rainfall, HRF05 received only half of that amount (183 mm in total; 67 mm before flowering and 116 mm after flowering). This water limitation negatively affected all three traits in HRF04 and TGW, grain number and yield were reduced by an average of 13.34, 37.84, and 49.24%, respectively compared with HRF05 (\( p < 0.0001 \), Wald test). In the HRF04 trial, TGW varied from 16.2 g to 31.3 g, grain number from 14,430/m² to 31,920/m², and yield from 4.77 t/ha to 10.22 t/ha. In the HRF05 trial TGW ranged from 16.2 g to 27.5 g, grain number from 14,430/m² to 31,920/m² and yield from 1.68 t/ha to 5.72 t/ha (Fig. 1, Table 1).

Despite the significant difference in mean TGW between both trials, relative genotype rankings were consistent between the trials with TGW in HRF04 being highly correlated with TGW in HRF05 (\( r = 0.96 \)) (Supplemental Figure S1, Table 2), indicating a strong genetic effect on TGW variation. Heritability estimated by repeatability was generally high for all three traits. In HRF04, heritability was 0.96, 0.84, and 0.74 for TGW, grain number and yield respectively, while the HRF05 exhibited slightly lower heritability with 0.94 for TGW, 0.64 for grain number and 0.65 for yield (Table 1).

In HRF04, a linear model including TGW, grain number and the interaction between them explained 90.43% of the variation in yield, while grain number and TGW independently explained 36.15 and 0.95%, respectively. In HRF05, a linear model accounting for TGW, grain number and the interaction between them explained 53.20% of yield variation, whereas grain number and TGW independently explained 3.09 and 8.26%, respectively. Negative correlations (\( r = -0.67 \) in HRF04, -0.74 in HRF05) between grain number and TGW were observed in both environments (Fig. 2, Table 2). The regression slope was significantly steeper in HRF05 (\( p < 0.0001 \)), indicating a higher correlation between grain number and TGW.

**TGW QTL in the BC\(_1\)F\(_5\) Population**

A total of 9747 polymorphic SNPs were identified, with the number of SNPs per chromosome varying from 559 on SBI-08 to 1733 on SBI-01 (Supplemental Table S3). The BLUEs were estimated for each genotype across two trials to exclude the effect of unbalanced environmental factors and used for trait-marker association analysis. In HRF04, 133 markers associated with TGW based on Strategy 1 were identified, and 120 of them were grouped into 9 high-confidence QTL regions (Supplemental Table S4). Four combined QTL regions were identified based on Strategy 2, co-locating with QTL reported from two recent association mapping studies; 1 of the combined QTL overlapped with a high-confidence QTL identified in strategy 1. In total, 12 unique genomic regions were identified TGW QTL from HRF04. Numbers of SNPs that were significantly associated with TGW within these QTL regions varied from 2 in q2004GW4.1 to 59 in q2004GW10.1. The effect of the 12 TGW QTL ranged from -0.56 to + 2.21 (g) (Supplemental Table S4, Fig. 3). The R931945–2-2 alleles increased TGW in the majority of the QTL (10), with the S. bicolor subsp. verticilliflorum alleles increasing TGW in 2 QTL.

In HRF05, 140 high-confidence SNPs were identified as being significantly associated with TGW using Strategy 1 and, of these, 126 SNPs were grouped into 9 high-confidence QTL (Supplemental Table S4). Two combined QTL were identified using Strategy 2, 1 of which overlapped with a high-confidence QTL, leading to 10 TGW QTL identified in HRF05. Numbers of SNPs that were significantly associated with TGW within these QTL regions varied from 2 in q2005GW6.1 to 38 in q2005GW7.2. The effect of QTL ranged from -0.21 to + 2.52 (g) with R931945–2-2 alleles increasing TGW in 9 QTL (Supplemental Table S4, Fig. 3). Five of the 10 HRF05 TGW QTL co-located with the HRF04 TGW QTL, resulting in 17 unique non-overlapping TGW QTL across both trials (Table 3, Fig. 4). Twelve TGW QTL were trial specific, including 7 unique in HRF04 and 5 unique in HRF05. However, 3 of the trial specific QTL showed a significant association at \( p\)-value < 0.05 with TGW in the alternative trial.

**Effects of TGW QTL on Grain Number**

Among the 17 TGW QTL, 11 exhibited a significant association (\( p < 0.05 \)) with grain number (Table 3) in at least one trial. The 5 QTL in common across both trials
(qGW3.3, qGW4.2, qGW7.1, qGW9.1, qGW10.1) were significantly associated with grain number variation regardless of trial, and 5 trial-specific QTL (q2004GW1.1, q2004GW5.1, q2005GW6.1, q2005GW7.2, q2004GW10.2) were significantly associated with grain number only in the trial where these QTL were detected, with the exception of q2005GW5.1, which was significantly associated with grain number in both trials. Additionally, 14 high-confidence QTL for grain number were identified using the same approach for TGW QTL identification (Supplemental Table S5), 7 of which were co-located with TGW QTL and 10 were significantly associated with TGW.

The effect on grain number of the 11 TGW QTL associated with grain number varied from −511/m^2 to −3,658.3/m^2 with R931945–2-2 alleles showing effects of reducing grain number in all cases (Supplemental Table S4). The effect direction of the R931945–2-2 alleles on grain number was opposite to the allele effects on the TGW, where the R931945–2-2 alleles resulted in increasing TGW in all 11 TGW QTL. The 5 common TGW QTL were negatively associated with grain number in both trials, however, the size of their effect varied dramatically in the different environments (Fig. 4, Supplemental Table S4).

Table 1. Variation and repeatability of TGW, grain number and yield in the BC1F5 population.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Mean</th>
<th>Range</th>
<th>Repeatability</th>
</tr>
</thead>
<tbody>
<tr>
<td>HRF04 TGW g</td>
<td>25.86</td>
<td>16.20–31.30</td>
<td>96</td>
</tr>
<tr>
<td>HRF05 TGW g</td>
<td>22.41</td>
<td>16.22–27.47</td>
<td>94</td>
</tr>
<tr>
<td>HRF04 YLD t/ha</td>
<td>8.06</td>
<td>4.77–10.22</td>
<td>74</td>
</tr>
<tr>
<td>HRF05 YLD t/ha</td>
<td>4.09</td>
<td>1.68–5.72</td>
<td>65</td>
</tr>
<tr>
<td>HRF04GN #/m^2</td>
<td>31,610</td>
<td>19,970–48,070</td>
<td>84</td>
</tr>
<tr>
<td>HRF05GN #/m^2</td>
<td>19,650</td>
<td>14,430–31,920</td>
<td>64</td>
</tr>
</tbody>
</table>

Table 2. Pearson correlation among TGW, grain number and yield across the two trials.

<table>
<thead>
<tr>
<th>Trait</th>
<th>TGW04</th>
<th>YLD04</th>
<th>GN04</th>
<th>TGW05</th>
<th>YLD05</th>
<th>GN05</th>
</tr>
</thead>
<tbody>
<tr>
<td>TGW04</td>
<td>1</td>
<td>0.15</td>
<td>0.34</td>
<td>0.06</td>
<td>0.34</td>
<td>0.66</td>
</tr>
<tr>
<td>TGW05</td>
<td>0.96</td>
<td>1</td>
<td>0.21</td>
<td>0.28</td>
<td>0.47</td>
<td>0.66</td>
</tr>
<tr>
<td>YLD04</td>
<td>0.15</td>
<td>0.21</td>
<td>1</td>
<td>0.47</td>
<td>0.56</td>
<td>0.16</td>
</tr>
<tr>
<td>YLD05</td>
<td>0.34</td>
<td>0.28</td>
<td>0.47</td>
<td>1</td>
<td>0.11</td>
<td>0.66</td>
</tr>
<tr>
<td>GN04</td>
<td>-0.67</td>
<td>-0.51</td>
<td>0.56</td>
<td>-0.11</td>
<td>1</td>
<td>0.66</td>
</tr>
<tr>
<td>GN05</td>
<td>-0.47</td>
<td>-0.74</td>
<td>0.13</td>
<td>0.16</td>
<td>0.66</td>
<td>1</td>
</tr>
</tbody>
</table>

Fig. 1. Frequency distributions of yield (t/ha), Thousand Grain Weight (TGW) (g) and grain number per square meter in both trials (HRF04 and HRF05).

Fig. 2. Negative correlation between Thousand Grain Weight (TGW) and grain number per square meter across the two trials. Triangles represent data from HRF05, while dots represent data from HRF04. Regression lines were fitted using a linear model.
Sorghum Grain Weight QTL Identified in Previous Studies

A total of 48 previously reported grain weight QTL corresponding to 29 unique genomic regions were extracted from 13 studies in which 15 bi-parental populations were employed to investigate the genetic basis of grain weight (Supplemental Table S6). On average, each unique region was reported in 1.67 populations, indicating a degree of consistency in the results across studies. Five of the 17 TGW QTL identified in this study were co-located with previously reported QTL from bi-parental populations, 3) the number of times the QTL co-located with previously reported QTL from a BTx623/S. propinquum population, and 4) whether a candidate gene with a signature of selection during domestication was identified within the QTL interval.

Table 3. Details of the 17 TGW QTL identified, including location, significance in the alternative year (Sig AltTrial), and significance of the QTL effect on grain number in both years (Sig GN04 and Sig GN05), and grain size candidate genes located in TGW QTL interval.

<table>
<thead>
<tr>
<th>QTL ID</th>
<th>Chr</th>
<th>Start*</th>
<th>End</th>
<th>Trial</th>
<th>Sig AltTrial†</th>
<th>Sig GN04</th>
<th>Sig GN05</th>
<th>Grain size candidate genes‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>q2004GW1.1</td>
<td>1</td>
<td>14917791</td>
<td>19064935</td>
<td>HRF04</td>
<td>NS</td>
<td>0.014</td>
<td>NS</td>
<td>Sobic.001G184900</td>
</tr>
<tr>
<td>q2004GW2.1</td>
<td>2</td>
<td>69872332</td>
<td>70320201</td>
<td>HRF04</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>NA</td>
</tr>
<tr>
<td>q2004GW3.1</td>
<td>3</td>
<td>58432545</td>
<td>58756700</td>
<td>HRF04</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>NA</td>
</tr>
<tr>
<td>q2004GW3.2</td>
<td>3</td>
<td>59206933</td>
<td>59746705</td>
<td>HRF04</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>NA</td>
</tr>
<tr>
<td>qGW3.3</td>
<td>3</td>
<td>70393141</td>
<td>71823632</td>
<td>Both</td>
<td>NA</td>
<td>0.047</td>
<td>0.002</td>
<td>Sobic.003G406600, Sobic.003G407300</td>
</tr>
<tr>
<td>q2004GW4.1</td>
<td>4</td>
<td>48852683</td>
<td>49136326</td>
<td>HRF04</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>NA</td>
</tr>
<tr>
<td>qGW4.2</td>
<td>4</td>
<td>58053978</td>
<td>60509497</td>
<td>Both</td>
<td>NA</td>
<td>0.047</td>
<td>0.003</td>
<td>Sobic.004G237000, Sobic.004G245000, Sobic.004G250000</td>
</tr>
<tr>
<td>q2005GW5.1</td>
<td>5</td>
<td>3468467</td>
<td>4047513</td>
<td>HRF05</td>
<td>0.013</td>
<td>0.008</td>
<td>0.043</td>
<td>NA</td>
</tr>
<tr>
<td>q2005GW5.1</td>
<td>5</td>
<td>62636349</td>
<td>6334404</td>
<td>HRF04</td>
<td>0.014</td>
<td>0.016</td>
<td>NS</td>
<td>NA</td>
</tr>
<tr>
<td>q2005GW6.1</td>
<td>6</td>
<td>3256913</td>
<td>3594223</td>
<td>HRF05</td>
<td>NS</td>
<td>NS</td>
<td>0.040</td>
<td>NA</td>
</tr>
<tr>
<td>q2005GW6.2</td>
<td>6</td>
<td>48782960</td>
<td>49088022</td>
<td>HRF05</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>NA</td>
</tr>
<tr>
<td>q2005GW6.3</td>
<td>6</td>
<td>51423868</td>
<td>51599114</td>
<td>HRF05</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>NA</td>
</tr>
<tr>
<td>qGW7.1</td>
<td>7</td>
<td>6581667</td>
<td>6789538</td>
<td>Both</td>
<td>NA</td>
<td>0.017</td>
<td>0.002</td>
<td>NA</td>
</tr>
<tr>
<td>q2005GW7.2</td>
<td>7</td>
<td>10970531</td>
<td>53856861</td>
<td>HRF05</td>
<td>NS</td>
<td>NS</td>
<td>0.004</td>
<td>Sobic.007G101500</td>
</tr>
<tr>
<td>qGW9.1</td>
<td>8</td>
<td>51052939</td>
<td>51635054</td>
<td>Both</td>
<td>NA</td>
<td>0.044</td>
<td>0.044</td>
<td>NA</td>
</tr>
<tr>
<td>qGW10.1</td>
<td>8</td>
<td>48958115</td>
<td>54115559</td>
<td>Both</td>
<td>NA</td>
<td>7.7E-04</td>
<td>0.042</td>
<td>Sobic.010G184100</td>
</tr>
<tr>
<td>q2004GW10.2</td>
<td>10</td>
<td>54952583</td>
<td>55645911</td>
<td>HRF04</td>
<td>0.046</td>
<td>5.48E-07</td>
<td>NS</td>
<td>Sobic.010G210100</td>
</tr>
</tbody>
</table>

* Genome locations are based on sorghum genome assembly (v3.1.1).
† NS indicates not significant, NA indicates not applicable.
‡ Grain size candidate genes were extracted from a list of 114 grain size candidate genes identified in Tao et al. (2017).
Candidate genes in TGW QTL

The 17 TGW QTL were investigated for co-location with 114 grain size candidate genes identified previously (Tao et al., 2017). In total 9 grain size candidate genes were located within the intervals of 6 TGW QTL identified in this study (Table 4). Among these grain size candidate genes in sorghum, 5 have been previously identified as having a purifying selection signature during sorghum domestication including Sobic.001G184900, Sobic.003G406600 (GW8 ortholog), Sobic.004G250000 (AHK4 ortholog), Sobic.007G101500 (Bt2 ortholog), and Sobic.010G21000 (GW6 ortholog) (Tao et al., 2017). Of these, two (GW8 and Bt2) have also been found to be under selection in rice or maize (Wang et al., 2012; Whitt et al., 2002).

DISCUSSION

Sorghum was first domesticated more than 5000 years ago in northern eastern Africa (Clark and Stemler, 1975; Mann et al., 1983; Wendorf et al., 1992). In common with the other important cereals, grain size was a major target of selection in early domestication with larger seeds being preferred. Larger seed size contributes to early vigor at the seedlings stage (Lafond and Baker, 1986), improved grain yield and ease of processing (Purugganan and Fuller, 2009). In part, the increase in grain size in cereals appears to have been achieved through a trade-off between grain size and grain number, as wild progenitors typically have larger numbers of small seeds compared with cultivated varieties. In this study we investigated the genetic control of grain size and its relationship with grain number in a cross between an elite Australian sorghum breeding line and an accession of the wild progenitor of the cultivated sub-species (S. bicolor subsp. verticilliflorum). The lines were grown in hybrid combination in two testing environments that strongly contrasted for abiotic stress. The nature of this cross meant that we were able to observe genetic variation that was selected during the domestication of the crop, while the contrasting testing environments permitted us to investigate the interaction between grain size, grain number, and assimilate supply as modified by drought stress.

The Genetic Architecture of TGW

The mean yield of the two trials varied considerably. In 2004 the trial mean yield was 8 t/ha whereas in 2005 the mean yield was only 4 t/ha due to water limitation between the late vegetative and mid-grain fill stage. The heritability of TGW was high (> 0.9) in both trials and despite the divergent environments the ranking of genotypes for TGW in the two environments was very consistent (r = 0.96). A total of 17 QTL were identified for TGW in this study. Twelve of the 17 QTL were unreported in previous QTL analyses, which were performed mainly using bi-parental populations between cultivated sorghum varieties. The high proportion of novel QTL identified in this study is likely due to the population consisting of a cross between a wild and cultivated sorghum. The QTL that were strongly targeted by domestication are unlikely to be segregating at high frequency in cultivated germplasm, hence they may be missed in QTL studies in cultivated sorghum. To date, the only study investigating grain weight in a bi-parental population between a wild and cultivated sorghum was conducted by Paterson et al. (1995) using a cross between BTx623 and S. propinquum. In this case, the wild species used, S. propinquum, is a related inter-fertile sorghum species not thought to have contributed to the domestication of cultivated sorghum. Whereas, S. bicolor subsp. verticilliflorum used in this study is thought to be the ancestor of cultivated sorghum (Dahlgren, 2000). Therefore, novel QTL identified in the current study potentially provide a more direct insight into domestication in sorghum than previous studies.

Out of 17 TGW QTL, five high confidence QTL were detected in both trials, with three further QTL showing a significant statistical association with TGW in the alternative trial (P-value < 0.05). Not unexpectedly, given the high correlation of TGW between sites, these eight QTL included six QTL with the largest effects in HRF04 and five QTL with the largest effects in HRF05. The 4 QTL with the largest effects in HRF04 increased TGW by between 6.5 to 8.5% each compared to the mean TGW of the trial. In HRF05, the four QTL with the largest effects increased TGW by between 8 and 11.2% each compared to the mean TGW of the trial. Interestingly, none of the four QTL with the largest effects in the low-stress environment (HRF04) were previously reported in studies using cultivated bi-parental populations. Only one of the four QTL with the largest effects in HRF04, qGW3.3, co-located with a previous grain mass QTL in the population BTx623 × S. propinquum (Paterson et al., 1995). Additionally, all of the four QTL with the largest effects in HRF04 contained candidate genes for grain size exhibiting signals of domestication, indicating these QTL were targeted during sorghum domestication. This is also in line with a previous observation that domestication often targets large-effect QTL (Purugganan and Fuller, 2009). In contrast, the four QTL with the smallest effects in the low-stress environment (HRF04) were more likely to co-locate with previously reported QTL, with two of them co-locating with QTL identified in bi-parental populations of both cultivated sorghum and BTx623 × S. propinquum cross, and all four co-locating with GWAS hits in previous studies (Fig. 3). This indicates that the allele diversity of these QTL was maintained, to some extent, during sorghum domestication, possibly as a result of lower selection pressure during domestication due to their relative smaller effects.

As expected, the majority of the QTL alleles from the wild accession reduced TGW. However, three QTL on SBI-02, SBI-04, and SBI-06 were identified where alleles from the wild sorghum parental line increased grain weight, albeit with small effect size ranging between 0.21 g and 0.56 g increase in TGW (Fig. 4). The contribution of QTL alleles from the wild parent to increased grain size demonstrates that even in a trait subject to strong...
During domestication it may be possible to find favorable alleles in wild species. Similar results have been observed in maize, rice, wheat, barley, and soybean (Concibido et al., 2003; Huang et al., 2003; Swamy and Sarla, 2008; Xiao et al., 1998).

### Negative Association Between Grain Weight and Grain Number

A negative association between grain number and grain weight is widely observed in plants (Griffiths et al., 2015; Jakobsson and Eriksson, 2000; Peltonen-Sainio et al., 2007; Sadras 2007) including sorghum (Burow et al., 2014; Heinrich et al., 1983; Yang et al., 2010). However, the genetic basis of this association remains unclear. To date, very limited co-location of QTL for grain weight and number have been reported (Boyles et al., 2016; Chen et al., 2016). In this study a strong negative correlation between grain number and TGW was found in the two contrasting environments. Eleven of the 17 TGW QTL, including all but 2 of the top 5 large-effect QTL in each trial, showed statistically significant negative associations with grain number, indicating the negative association between grain number and grain weight is due to pleiotropic effect of these QTL rather than multiple genes in tight linkage.

Fig. 4. Distribution of 17 Thousand Grain Weight (TGW) QTL across the sorghum genome. Each square indicates the location of a TGW QTL. Gray squares indicate the corresponding TGW QTL is associated with grain number, while black squares indicate the corresponding TGW QTL is not associated with grain number. Each circle represents the location of a candidate gene within corresponding TGW QTL. A gray circle indicates the candidate gene was identified with a domestication selection signal in Tao et al. (2017), while a black circle indicates that a domestication selection signature was not identified in the candidate gene.

### Table 4. List of TGW candidate genes and selection signal on them during sorghum domestication

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Start</th>
<th>End*</th>
<th>Original Gene</th>
<th>Under selection</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sobic.001G184900</td>
<td>15806732</td>
<td>15807894</td>
<td>LOC_Os03g37810</td>
<td>Yes</td>
<td>Huang et al., 2012</td>
</tr>
<tr>
<td>Sobic.003G066600</td>
<td>71419053</td>
<td>71424347</td>
<td>GW8(OsPL16)</td>
<td>Yes</td>
<td>Wang et al., 2012</td>
</tr>
<tr>
<td>Sobic.003G073000</td>
<td>71489991</td>
<td>71496210</td>
<td>AHK3</td>
<td>No</td>
<td>Riefler et al., 2006</td>
</tr>
<tr>
<td>Sobic.004G237000</td>
<td>5848864</td>
<td>58490438</td>
<td>PGL2</td>
<td>No</td>
<td>Heang and Sassa, 2012</td>
</tr>
<tr>
<td>Sobic.004G245000</td>
<td>59266365</td>
<td>59273133</td>
<td>AHK4</td>
<td>Yes</td>
<td>Riefler et al., 2006</td>
</tr>
<tr>
<td>Sobic.004G247000</td>
<td>59472640</td>
<td>59476805</td>
<td>Gln-4</td>
<td>No</td>
<td>Martin et al., 2006</td>
</tr>
<tr>
<td>Sobic.007G101500</td>
<td>25707028</td>
<td>25712978</td>
<td>Bi2</td>
<td>Yes</td>
<td>Press et al., 1990</td>
</tr>
<tr>
<td>Sobic.010G184100</td>
<td>52392383</td>
<td>52396386</td>
<td>Bi1</td>
<td>No</td>
<td>Shannon et al., 1998</td>
</tr>
<tr>
<td>Sobic.010G210100</td>
<td>55370341</td>
<td>55372973</td>
<td>GW6</td>
<td>Yes</td>
<td>Song et al., 2015</td>
</tr>
</tbody>
</table>

* Start and end record physical position of cor responding gene based on sorghum genome assembly (v3.1).
Our study also revealed 6 TGW QTL that showed no evidence of being associated with grain number, offering the potential to select positively for both traits. Understanding the physiological impact and genetic control of the genes underlying these QTL may open new opportunities to simultaneously increase grain weight, grain number and potentially yield in sorghum, either through conventional breeding, gene modification, or gene editing.

**Impact of Water Limitation on Yield Components**

The two environments sampled in this study were highly divergent in the degree of water limitation they experienced. Although severe reduction of grain size can be caused by abiotic stress, it is generally regarded that a reduction in grain number, rather than a reduction in grain size, largely accounts for crop yield losses due to abiotic stress (Dolferus et al., 2011). Part of the reason is that grain number has a higher degree of plasticity than grain size, and is thus more vulnerable to abiotic stress (Bradshaw, 1965; Smith and Fretwell, 1974). Such plasticity was observed in the current study, as correlation of grain number between trials and heritability of grain number were lower than those of TGW. From an evolutionary perspective, sacrificing grain number to maintain grain size under adverse conditions is a sensible strategy as it will result in a small number of seeds with better chances of survival.

**CONCLUSIONS**

Grain size has been increased during sorghum domestication. To investigate the genetic basis underlying the change of grain size during sorghum domestication, we performed the first QTL study of TKW in a cross between an elite sorghum inbred line and its wild progenitor. The fact that large-effect TGW QTL identified in this study have not been reported in previous studies in cultivated sorghum, combined with the signals of selection observed for candidate genes within these QTL in both sorghum and other species, indicate that these QTL were the target of strong selection during sorghum domestication. Grain weight and grain number are the two major components of yield in cereal crops, but they are often negatively correlated with one another. Our results showed strong and consistent negative associations between most TGW QTL and grain number, indicating that these effects are almost certainly due to pleiotropy. However, a proportion of the QTL identified (6 out of 17), and which tended to have smaller phenotypic effects, were found not to be associated with grain number. These QTL were more likely to segregate in cultivated material and provide scope for increasing grain size and yield. The change in the relative importance of TKW and grain number on yield depending on water stress provides useful information for plant breeders developing genotypes with improved yields in water-limited environments.

**Supplemental Information**

Supplemental Table S1. List of grain weight GWAS hits identified in previous studies.

Supplemental Table S2. Phenotypic data of the mapping population.

Supplemental Table S3. Genotypic data of the mapping population.

Supplemental Table S4. List of TKW QTL identified in each trail and their effects on TKW and grain number.

Supplemental Table S5. List of grain number QTLs identified in this study.

Supplemental Table S6. List of grain weight QTLs identified in previous studies.

Supplemental Fig. S1. Correlations of measured phenotypic traits.

**Conflict of Interest Disclosure**

The authors declare that there is no conflict of interest.

**Funding**

This work was supported by Australian Research Council Discovery project DP14010250 and the Australian Grains Research and Development Corporation (www.grdc.com.au).

**Author Contribution Statement**

D.J. managed the project. D.J. and R.H. designed the experiment. B.G.J. performed the experiment. Y.T., C.H., A.C., and E.M. performed data analysis. Y.T. wrote the manuscript, E.M., D.J., and B.G.J. revised the manuscript.

**References**


D.J. and R.H. designed the experiment. Y.T., C.H., A.C., and E.M. performed data analysis. Y.T. wrote the manuscript, E.M., D.J., and B.G.J. revised the manuscript.


