Common genetic control for grain filling duration and kernel weight in grain sorghum

Daniel Ajaku Otwani
d.otwani@uq.edu.au

The University of Queensland https://orcid.org/0000-0002-0652-1398

Colleen Hunt
Queensland Department of Primary Industries and Fishing - DPI&F: Queensland Department of Agriculture and Fisheries

Alan Cruickshank
Queensland Department of Primary Industries and Fishing - DPI&F: Queensland Department of Agriculture and Fisheries

Yongfu Tao
Chinese Academy of Agricultural Sciences Agricultural Genomes Institute at Shenzhen

Anna Koltunow
QAAFI: Queensland Alliance for Agriculture and Food Innovation

David Jordan
QAAFI: Queensland Alliance for Agriculture and Food Innovation

Emma Mace
QAAFI: Queensland Alliance for Agriculture and Food Innovation

Research Article

Keywords: Sorghum bicolor, GFD – Grain filling duration, Sorghum racial groups, GWAS, Single marker analysis, Haplotype

Posted Date: October 30th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-7846434/v1

Abstract

The duration of the grain filling period has been associated with yield increases in cereals including maize and sorghum. The genetic control of grain filling duration (GFD) is however not known in sorghum. This study explored the genetic variation and extent of genetic control for GFD in a diverse panel of sorghum genotypes (n = 904), in three environments across two years. A genome wide association analysis revealed 86 QTLs, 46 of which collocated with 54 previously reported grain size candidate genes in sorghum, indicating a significant enrichment. Single marker analysis revealed that genomic regions associated with grain filling duration were similarly associated with grain size. Interestingly, expression analysis of candidate genes associated with GFD revealed that GFD could be associated with processes that happen both before and after anthesis contrary to the understanding that GFD was primarily associated with processes that happen post anthesis. Haplotype analysis of *SbGS3* resolved 8 haplotypes associated with grain filling duration, 2 of which were exclusive to the guinea and Asian durra racial groups revealing opportunities for trait introgression across sorghum racial groups. These results indicate considerable opportunity to increase grain yield in sorghum, by selecting for longer GFD and diverse inter racial crosses to improve the genetic diversity for grain filling duration in sorghum. Sorghum breeders will find application of these results in diversifying trait selection to optimise yields in changing environments.

Key Message

The genetic control of grain filling duration (GFD) in sorghum intricately overlaps with that of grain size, is influenced by both pre- and post-anthesis processes, offering new opportunities for yield improvement through targeted breeding and inter-racial trait introgression.

Introduction

Grain filling, defined as the period between anthesis and physiological maturity, plays a critical role in determining maximum grain size (Egli 2006). The rate and duration of grain filling period together determine grain size, which, when combined with grain number per unit area, determines grain yield. (Boyles et al. 2016; Otwani et al. 2024; Van Oosterom and Hammer 2008). To date, genetic gain for yield in sorghum has been achieved primarily through changes in grain number, with similar trends being observed in other cereals like maize (Russell 1991) and rice (Khush 1995). Recent studies on maize yield improvements over five decades (Fernández et al. 2022; Xing et al. 2023) reveal that increases in grain weight among hybrids of different eras were largely due to extended grain filling duration, suggesting that targeting this trait could enhance yield gains in cereals. Otwani et al. (2025) report potential yield benefits in sorghum through increasing grain filling duration, suggesting that exploiting grain filling duration has potential to overcome (Gambín and Borrás 2012; Yang et al. 2010) the negative correlation between grain size and number (Sadras 2007) while breeding for increased yield.

There is limited research on the extent and genetic control of variation for grain fill duration in sorghum. Yang et al. (2009), observed that pre-anthesis ovary volume is varied across sorghum genotypes and was correlated with grain filling duration and grain size on a set of three genotypes. This result is consistent with a study by (Tao et al. 2021) indicating that grain size is limited more by the genetic potential of grain size set pre-flowering rather than assimilate supply suggesting that there may be potential to exploit genetic variation in grain size to increase grain size in sorghum. To date more than 100 quantitative trait loci (QTLs) have been identified in sorghum for grain size and weight across diverse studies (Boyles et al. 2017; Han et al. 2015; Paterson et al. 1995; Tao et al. 2018; Tao et al. 2020), some of which are in common with other grain size related traits like grain length, width, volume and grain thickness. Despite the QTLs reported for grain size, only a few predicted candidate genes have been identified in sorghum. *Sobic.001G341700* is predicted to be the causative gene for qTGW1a which acts as a negative regulator of grain size in sorghum (Zou et al. 2020) homologous to *GS3* in rice. Further, studies to explore other traits associated with grain size and their genetic control in sorghum are limited. To our knowledge, no study is available on the genetic control of grain filling duration in sorghum, however some studies in maize and rice have reported some predicted genes and transcription factors responsible for grain filling.

Three transcription factors, *NAKED ENDOSPERM1* (*NKD1*), *NKD2* and *OPAQUE2* (*O2*) have been reported to function in endosperm cellular development and promoting biosynthesis and storage of starch, proteins and lipids in the developing maize seed (Wu et al. 2024). The transactivation by *O2* of sucrose synthase1 (*Sus1*) and *Sus2* mediates endosperm filling in maize (Deng et al. 2020), and transactivation by *O2* of a *DELLA*-like transcriptional regulator, *ZmGRAS11* mediates synergistic endosperm enlargement with grain filling (Li et al. 2021). In rice, a prolonged grain filling duration mutant 1 (*gfd1*), show a longer grain filling duration, less grain number per panicle and bigger grain size. *GFD1* interacts with sugar transporters *OsSWEET4* and *OsSUT2* to mediate grain filling duration and grain size respectively and with both *OsSWEET4* and *OsSUT2* to regulate grain number (Sun et al. 2023). Some predicted gene models have been identified in sorghum through stage specific gene expression analysis including grain filling period (Costes et al. 2024; Cruet-Burgos and Rhodes 2023; Jain et al. 2024). For instance, *Sobic.002G367600*, an orthologue of *CYP78A13* in rice, a regulator of size balance between embryo and endosperm (Nagasawa et al. 2013) has been reported to be highly expressed during grain filling in sorghum (Jain et al. 2024). Carbohydrate metabolism genes have been shown to be highly expressed during the grain filling period like the *waxy* (*wx Sobic. 010G022600*) (Jain et al. 2024) and *SUGARY* (*SbSu*,*Sobic.007G204600*) which has a regulatory role in starch synthesis (Hashimoto et al. 2023). Similarly in other cereals, amylase inhibitors have been reported to accumulate from one week after anthesis through to physiological maturity in wheat (Call et al. 2021) and rice (Hakata et al. 2012) and function to improve grain quality by repressing starch degradation suggesting that these functions may be conserved in cereals.

Several studies in sorghum have attempted to explore the genotypic diversity and physiology of grain filling duration and its association with yield (Done 1986; Gambín and Borrás 2012; Otwani et al. 2025; Schaffer 1981). While these studies are pivotal in establishing a potential link between an extended grain filling duration and yield, the exploitable genetic variation for grain filling duration in diverse sorghum genotypes remains to be studied at depth, so is the genetic control and association with other yield determining traits in sorghum.

In the present study, we hypothesise (1) that genetic variation for grain filling duration is available in sorghum diverse germplasm, (2) that the grain filling duration could be extended independent of flowering time and maturity, and (3) that genetic/genomic controls of grain filling duration could be dissected by

examining the onset and progression of grain filling. We applied a population genetics approach to investigate the natural variation of grain filling duration within a diverse sorghum panel. Through genome-wide association analysis, we identified genomic regions and candidate genes that may be involved in regulating grain filling duration.

Materials and Methods

Plant materials and experiments

The sorghum diversity panel (DP, n = 904), previously described by (Otwani et al. 2025; Tao et al. 2020), was used in the current study. Three experiments were planted, two at the Hermitage Research Facility (HRF), Warwick, Queensland, Australia (28˚ 12′ S, 152˚ 5′ E, 470 m above sea level) in November 2020 and December 2021, and the third was planted at Gatton Research Facility (GAT), Gatton, Queensland, Australia, (27˚ 33′ S, 152˚ 20′ E, 94 m above sea level) in February 2021. At HRF, 881 DP genotypes were planted in a row column design with partial replication where 30% of the genotypes were replicated two or more times while the remaining 70% were in single plots in 2020/21 season (HRF1) and a fully replicated trial in 2021/22 season HRF2. At GAT, a total of 609 DP lines were planted in a fully replicated trial of two replications in a row column design. All the trials were planted during the Australian summer growing season in single row plots 4 metres long. Standard agronomic practices were employed in the trial management to ensure adequate nutrition and pest and weed control. Overall, the experiments had 598 DP genotypes in common.

Phenotypic evaluation

Single plants of each genotype were tagged in each plot at the time of head exsertion prior to onset of flowering. All measurements for timing of flowering and maturity were recorded on the tagged plant. Flowering time (DTF) was recorded as the date when the first anthers become visible at the tip of the panicle. The tagged plant was monitored throughout the season and the date of physiological maturity (DTM) was recorded as the date when a sampled grain from the tip of the panicle first showed the abscission layer (black layer) at the point of connection of the grain. Plant height was measured at HRF2 by selecting one representative plant at random from the plot and measuring the distance from the base of the plant to the tip of the panicle at physiological maturity. Single panicles were harvested at HRF2, threshed, and cleaned before grains per panicle, and thousand kernel weight (TKW) were measured using an automatic seed counter and weighing machine (Ball Coleman Gen3 seed counter). Daily weather data was recorded using a portable weather station placed within the trial to record daily maximum and minimum air temperatures for the duration of the experiment. The temperature data was used in the estimation of thermal time accumulated for respective growth and development phases as described in (Hammer and Muchow 1994). Overall, the trial at Gatton experienced lower temperatures during anthesis and post-anthesis in the grain filling period.

Statistical analyses

Linear mixed models were fitted as a multi-environment trial (MET) analysis and used to predict Best Linear Unbiased Predictions (BLUPs). The MET model was also used to estimate correlations between the study traits. All the traits were analysed using a linear mixed model and the residuals assessed for normality.

The standard representation of a linear mixed model is given by;

$$y = X\tau + Zu + e \quad (1)$$

Where y is the vector of observations with the sites stacked, X is the design matrix for fixed effects, τ is the vector of fixed effects, Z is the design matrix for random effects, u is the vector of random effects which has a normal distribution with mean 0 and variance G ($u \sim N(0, G)$), with fixed and random spatial effects included as necessary (see supplementary Table 4.1) (Gilmour et al. 1997) and e is the vector of residuals $e \sim N(0, R)$.

All the sites had significant autoregression correlations in both the column and row directions and a random effect. HRF1 had a spline effect in the column direction for both the time to flowering and duration to maturity traits but not for the grain filling duration, since the site was uneven. HRF2 had a linear trend in the column direction.

Best Linear Unbiased Estimates (BLUEs) were estimated by including the genotypes as fixed effects in model (1) (contains a main effect for genotypes at Gatton only) while BLUPs were predicted from model (1) where site x genotype was included as a random effect. The variance-covariance matrix for the site by genotype interaction (GxE) was fitted using a correlation structure (corgh). This structure allows for a different genetic variance for each site and different correlations for each pair of sites. Different models were fitted separately for each trait, with random and fixed terms included as necessary per site, see Supplementary table *4.1*. Broad sense heritability was estimated per site using the generalised heritability method as described in (Cullis et al. 2006).

All analyses were conducted in R (RCoreTeam 2024) environment version 4.04, the package ASReml-R (TheVSNiTeam 2023) was used to fit all models and the package ggplot2 (Wickham 2016) was used in visualising all figures.

Molecular marker data

Procedures for genomic DNA extraction and sequence data construction were described previously (Mace et al. 2019; Tao et al. 2018). In total, 726,309 SNPs were identified and aligned to sorghum genome assembly version v3.1.1. This diversity panel was resequenced using DArTreseq technology (Edet et al. 2018)، and conducted by Diversity Arrays Technology Pty Ltd https://www.diversityarrays.com/technology-and-resources/dartreseq/. Bulked young leaf tissue of five plants in each plot was used for DNA extraction using a modified cetyl trimethyl ammonium bromide (CTAB) method (Doyle and Doyle 1987).

The DNA samples were digested with methylation-sensitive restriction enzymes (*HpaII*, *MseI*) to remove repetitive sequences. Sequencing libraries within insertion size of 350 bp were constructed using a TruseqNano DNA HT sample preparation kit (Illumina; catalog no. FC-121-4003) following the manufacturer's recommendations. The libraries were sequenced using HiSeq 2500 (Illumina) to produce paired-end, 150-bp reads. After trimming adapters and filtering low-quality reads, the clean reads were mapped to the reference genome BTx623 (v3.1.1) (McCormick et al. 2018) with Burrows-Wheeler Alignment software (version 0.7.8) using the *mem* command (Li and Durbin 2009) to call SNPs.

GFD across sorghum racial groups

The 881 DP genotypes were allocated racial group membership based on a population structure analysis as described in (Tao et al. 2020), with a threshold of 70% genetic identity for a genotype to be allocated to a given racial group. The racial variation was analysed across each trial independently and across all trials together. Across trial data is presented.

GWAS analysis and QTL identification

The Fixed and random model Circulating Probability Unification (FarmCPU) software described in (Liu et al. 2016) was used for GWAS analysis while accounting for population structure using principal component analysis (PCA). 726,309 SNPs (minor allele frequency > 0.01) after imputation and filtering was realised and used for the GWAS analysis. For potential significant QTL identification, the package simple *.M* (Gao et al. 2008) was used to determine an estimate of the number of independent tests which was then used in the determination of the cut off P-value for the selection of effective SNPs. Thus, a cut off P-value of 1.530953e-07 for HRF1 and HRF2 analysis and 1.653423e-07 for GAT analysis was identified for significant SNPs and P-value of < 9.79e-05 for suggestive SNPs. SNPs from the three environments were collated and those sitting within 1 cM of each other within a chromosome were in the same QTL region. The QTLs were designated with letters QGFD , Q for QTL and GFD the trait and consecutive numbers starting with the chromosome number followed by a decimal point and numbers showing the number of QTLs within the chromosome (QGFD1.1, would be the first QTL in chromosome 1). Further, post GWAS analyses were conducted to compare coincidence of GFD QTLs with previously identified QTLs for grain size and other grain related traits in sorghum. To compare the overlap of GFD QTLs with previously identified grain size-related QTLs, we reviewed several studies: Takanashi et al. (2021) with 213 RILs, Zou et al. (2020) with 244 RILs, Tao et al. (2020) with 837 diversity panel lines and 1,421 BC-NAM lines, and ,(Tao et al. 2021) which involved manipulating assimilate supply. Further comparisons between detected QTLs in this study and previously reported sorghum QTLs were performed using the QTL Atlas (Mace et al. 2019) (https://aussorgm.org.au/sorghum-qtl-atlas/). To search for sorghum orthologs of rice or maize responsible genes, we used the BLASTP program in the Phytozome database (https://phytozome-next.jgi.doe.gov/). *A priori* candidate genes were further explored from previous studies that identified candidate genes associated with grain size, starch and protein content (Jain et al. 2024; Tao et al. 2018; Tao et al. 2021; Tao et al. 2020) and genes expressed from pollination to maturity in grain sorghum (Jain et al. 2024). A 1 centimorgan (cM) window was used to identify collocation of the GFD QTLs with the candidate genes.

Single marker analysis

All the significant SNPs identified for GFD across the three locations from the GWAS were collated and used for single marker analysis. The SNP effects on GFD were analysed in a linear mixed model framework with all significant SNPs included simultaneously as fixed effects in the model with GFD first as the response, then the process was repeated with TKW as the response. The random terms included genotypes and marker data in a variance model to account for kinship and structure within the genotypes. The residual term was an autoregressive model in the column and row directions. The model equation is described below;

$$Y=X\tau + Zu + e$$

Where y is the vector of observations, X is the design matrix for fixed effects, τ is the vector of fixed effects, Z is the design matrix for random effects, u is the vector of random effects which has a normal distribution with mean 0 and variance G (u~N(0, G)), with fixed and random spatial effects included as necessary (Gilmour et al. 1997) and e is the vector of residuals e~N (0, R).

Haplotype analysis for Sobic.001G341700

The haplotype analysis of the qTGW1a (*Sobic.001G341700*) an orthologue of *GS3* in rice in the sorghum diversity panel were performed using the SNPs data on genomic sequence using *Sorghum bicolor* genome v3.1.1. The package vcfR was used for the initial extraction of the genotypic information from the vcf file. The packages adegenet (Jombart 2008) (version 2.1.10) was used to convert data into a format suitable for further population analysis and clustering, while ade4 (Dray and Dufour 2007) was used for generation of principal components. The resulting haplotypes for the gene were visualized using the ggplot2 package (Wickham 2016).

Results

Phenotypic variation in grain filling duration

GFD ranged from 400 to 680-degree days with the means across the genotypes for each experiment being 510, 506 and 521-degree days for GAT, HRF1 and HRF2 respectively. Appreciable genetic variation for GFD was observed, with moderate broad sense heritability estimates ranging between 41% and 61%. Genetic correlations between sites ranged from 0.45 to 0.86 with HRF1 and HRF2 having a stronger genetic correlation and a less strong genetic correlation

reported between HRF2 and GAT A comparison of GFD across the sorghum genotypes as defined by racial grouping showed that race guinea had on average a longer GFD in comparison to all the other racial groups in each site and from the combined analysis (see figure from chapter 4).

Phenotypic correlation of GFD and other traits

A Pearson's correlation analysis for GFD and yield related traits revealed that GFD was significantly positively correlated with DTM and TKW but had a non-significant negative correlation with DTF. DTF was significantly negatively correlated with TKW while DTM had a non-significant low correlation with TKW (Figure 2).

Marker trait associations for GFD in the sorghum diversity panel

GWAS analyses for GFD conducted independently for each of the three experiments at HRF1, HRF2 and GAT identified a total of 117 significant and suggestive marker trait associations/SNPs at a significance P-value < 9.79e-05. The highest number of significant SNPs was identified from the HRF1trial (49 in total), while HRF2 and GAT identified 34 and 35 SNPs respectively. One SNP on Chromosome 5 was significant at both HRF1 and HRF2. Overall, SNPs were distributed throughout all the chromosomes with chromosome one having the most and chromosome six the least number of identified SNPs (Figure *3* Supplementary table *4.3*). For onward analysis, the 117 significant and suggestive SNPs were clustered into 86 unique QTLs based on a 1cM window around each SNP as previously described in (Tao et al. 2020) for the sorghum diversity panel. The 86 QTL regions were distributed throughout the 10 chromosomes, with HRF1, having 29, HRF2 24 and GAT 19 unique QTLs respectively. 14 QTLs were common in at least two environments, two of which were found common across all the three environments (Figure *4*).

Single marker analysis revealed common genomic regions for GFD and TKW

All significant SNPs identified in the initial GWAS analysis for GFD were subsequently tested in a single-marker analysis and found to be strongly associated with GFD, with a p-value of <0.0001. Similarly, when these SNPs were tested for association with TKW all but four were highly significantly associated at p-value <0.0001, and all but one were significant at p-value <0.05. The individual SNP effects for GFD when compared to those of TKW showed that the SNPs affected the two traits behaved in a similar fashion in terms of both the effect size and effect . The GFD SNP effects could explain up to 82% of the observed variation in the TKW SNPs effects. The largest individual SNP effect for GFD accounted for a difference of 51-degree days in GFD, equivalent to three diurnal days at ambient temperature, while the smallest SNP effect accounted for a 0.3-degree day difference in GFD. A similar trend was observed for TKW with the largest SNP effect accounting for a 10 gram difference per 1000 seeds while the smallest accounting for a 0.03 gram difference per 1000 seeds (Figure *5*).

Coincidence of GFD QTLs with expressed genes between pollination and maturity

The coincidence of 86 GFD QTLs with 938 genes exclusively expressed at each respective period from 1-2 days before pollination to physiological maturity in grain sorghum (Jain et al. 2024) revealed 59 QTLs were in linkage disequilibrium (LD) with 167 of these genes. A Chi square test (X^2 test) revealed a significant enrichment of the expressed genes for GFD QTLs at P< 0.0001. Further, 25,49,10,7 and 5 QTLs each mapped to genes expressed 1-2 days before pollination (1-2 DBP), 0-2 days after pollination (0-2 DAP), 10 days after pollination (10 DAP), 20 days after pollination (20 DAP) and 30 days after pollination (30 DAP) respectively (Supplementary table 4.2). 77% of these QTLs mapped with genes expressed exclusively early in the pre and post pollination phase of seed development (1-2 DBP and 0-2 DAP). These observations suggest that GFD could possibly be determined by plant growth and development happening early before flowering. The 27 QTLs not specifically mapped to the genes expressed exclusively in the stages above, could be commonly expressed in all the stages from pollination to maturity as well as before pollination.

Collocation of GFD QTLs with candidate genes for grain size, starch and protein content in sorghum

A comparison of GFD QTLs to 185 sorghum candidate genes associated with grain size, starch and protein content as summarised in (Jain et al. 2024; Tao et al. 2017) revealed that ~53% GFD QTLs (46 of 86) were in LD with at least one of these candidate genes, a significant enrichment at P<0.0001 X^2 test). 42 of the 54 candidate genes in LD with GFD QTLs were associated with grain size, 9 with starch content and 3 with protein content (Jain et al. 2024) (Table 2). The candidate genes associated with grain size were functional in regulating cell proliferation, elongation and division, as well as phytohormone mediated regulation of grain size (Jain et al. 2024). The QTLs QGFD1.13 ,QGFD2.11, QGFD7.1,QGFD7.4 and QGFD10.4 were within 0.2-0.6 cM of the candidate genes Sobic.001G485400,Sobic.001G481400,Sobic.002G367300,Sobic.007G193500 ,Sobic.007G054700 and Sobic.010G110100 respectively. The rice orthologues of these candidate genes except for Sobic.010G110100 have been shown to be involved positively in cell proliferation, elongation with resultant increases in grain width, grain size and grain filling (Liu et al. 2015; Lo et al. 2020; Wang et al. 2015a; Wang et al. 2012; Wang et al. 2015b) .The rice orthologue of Sobic.010G110100 suppresses cell proliferation and negatively regulate grain size and weight (Hao et al. 2021). Additionally, these candidate genes have been shown to have peak expression early in the prepollination, post fertilisation and early grain filling in sorghum (Jain et al. 2024) indicating their potential role in early embryogenesis and endosperm development revealing that grain filling could be determined early in the panicle development phases. Candidate genes associated with phytohormone signalling were collocated within 0.2-1.4 cM of QTLs QGFD1.9, QGFD1.11,QGFD1.13,QGFD1.17,QGFD3.2 , QGFD4.6 and QGFD10.4 corresponding to Sobic.001G172400,Sobic.001G120900, Sobic.001G488500, Sobic.001G109100, Sobic.003G257400 , Sobic.004G237000 and Sobic.010G111200. The rice orthologue of Sobic.001G120900 is a negative regulator of grain size through a reduction of gibberellic acid (GA) signalling (Lan et al. 2020) while mutants of the rice orthologue of Sobic.010G111200 show reduce GA and decreased grain weight and width (Shi et al. 2020). Sobic.001G109100, Sobic.001G172400 and Sobic.004G237000 have orthologues in rice that are involved in the brassinosteroid (BR) pathway signalling to reduce grain size and increase cell proliferation, expansion and grain length respectively. Sobic.001G488500 and Sobic.003G257400 were positive regulators of grain size through ethylene mediated reduction in cell proliferation with resultant increases in grain length and cell size in rice spikelets (Chen et al. 2013) and cytokinin (CK) mediated regulation of grain size respectively (Xiao et al. 2019; Yin et al. 2020) (Table 2).

Table 2: Concurrent of grain sorghum candidate genes with GFD QTLs. Corresponding genes in rice, maize and Arabidopsis have been provided with their predicted functions. The start and end predicted physical genetic and cM position for the candidate gene is provided. The cM distance from QTL position is also provided as (cM from QTL).

Gene ID	Gene name	Rice/Maize/Arabidopsis orthologue	cM	LG	START	END	cM from QTL	GFD QTL	Function
Sobic.001G056700	*O2*	Zm00001d018971	10.94	1	4275459	4279430	0.16	QGFD1.18	Regulatory protein opaque-2
Sobic.001G107100	*SRS5/TID1*	LOC_Os11g14220	23.73	1	8265620	8268721	1.02	QGFD1.17	Induces cell elongation in spikelet cells and produces longer grains
Sobic.004G245000	*AHK4*	At2g01830	107.54	4	59266365	59273133	0.13	QGFD4.6	CHASE domain containing histidine kinase protein
Sobic.001G172400	*BRD1*	LOC_Os03g40540	36.53	1	14434718	14438560	0.63	QGFD1.9	BRD1 encoded protein catalyses the C-6 oxidation step to produce active BR which increases cell proliferation and expansion in grains
Sobic.004G214100	*BC14*	Os02g0614100	104.77	4	56389819	56395087	1.44	QGFD4.6	Golgi-localized nucleotide sugar transporters
Sobic.002G367600	*BG2*	Os07g0603700	179.20	2	72744933	72746890	0.70	QGFD2.11	Cytochrome P450
Sobic.008G152800	*CBL3*	AT4G26570	139.82	8	58517448	58523171	0.59	QGFD8.3	calcineurin B-like 3
Sobic.002G272700	*EOD3/CYP78A6*	At2g46660	147.07	2	65585643	65588410	0.07	QGFD2.7	oxygen binding
Sobic.001G184900	*Expressed protein*		40.28	1	15806732	15807894	0.06	QGFD1.1	
Sobic.001G341700	*GS3/zmGS3*	Os03G0407400	130.78	1	62910779	62916258	1.03	QGFD1.16	qTGW1a encodes a G-protein subunit which negatively regulates grain size
Sobic.001G485400	*BG1*	LOC_Os03g07920	188.47	1	75624275	75627243	0.39	QGFD1.13	Overexpression positively regulates grain size due to increased cell proliferation
Sobic.010G047400	*HGW*	Os06g0160400	32.75	10	3668202	3673695	0.40	QGFD10.5	ubiquitin-associated domain protein
Sobic.001G488500	*OsFBK12*	LOC_Os03g07530	189.02	1	75844162	75849593	0.94	QGFD1.13	Acts as repressor for downstream gene *SAMS1*and reduces ethylene level and cell proliferation but increases grain length by increasing cell size in spikelet hull
Sobic.002G056000	*MET1*	At5G49160	22.90	2	5374690	5386770	0.68	QGFD2.3	methyltransferase 1
Sobic.002G054800	*O2*	Zm00001d018971	22.63	2	5243140	5247362	0.95	QGFD2.3	
Sobic.001G254200	*OsFBK12*	Os03g0171600	55.24	1	28465427	28472547	0.12	QGFD1.4	
Sobic.002G367300	*qGW7/GL7*	LOC_Os07g41200	179.12	2	72705386	72710986	0.62	QGFD2.11	Encodes a TONNEAU1-recruiting motif protein which enhances cell elongation resulting

Gene ID	Gene name	Rice/Maize/Arabidopsis orthologue	cM	LG	START	END	cM from QTL	GFD QTL	Function
									larger grains with improved quality
Sobic.008G173900	*OsPPKL3*	Os12g0617900	145.36	8	60836806	60845885	0.35	QGFD8.2	Extra large grain
Sobic.001G254100	*PGL1*	Os03g0171300	55.23	1	28215025	28215913	0.10	QGFD1.4	
Sobic.001G488400	*PGL1*	Os03g0171300	188.98	1	75826748	75828379	0.89	QGFD1.13	
Sobic.010G091700	*PGL2*	Os02g0747900	44.86	10	8099384	8100834	1.99	QGFD10.3	
Sobic.004G237000	*PGL2/BUL1*	LOC_Os02g51320	107.18	4	58488864	58490438	0.48	QGFD4.6	In BR biosynthesis pathway, *BUL1* upregulates *BDG1* which increases grain length
Sobic.001G468400	*Prol1.1*	Zm00001d028129	184.76	1	74135478	74137316	0.33	QGFD1.6	
Sobic.004G247000	*Gln-4*	Zm00001d051804	107.64	4	59472640	59476805	0.03	QGFD4.6	Glutamine synthetase isoenzymes
Sobic.007G193500	*SPL16/qGW8*	LOC_Os08g41940	130.54	7	62605971	62612183	0.25	QGFD7.4	Enhances cell proliferation which increases grain width and grain filling
Sobic.001G484200	*RGA1/D1*	Os05g0333200	188.23	1	75526130	75530742	0.14	QGFD1.13	
Sobic.003G380900	*SERF1*	Os05g0420300	146.25	3	69444094	69445096	1.97	QGFD3.11	
Sobic.009G141500	*SERF1*	Os05g0420300	87.26	9	49879082	49879738	0.03	QGFD9.6	
Sobic.009G049400	*SRS3*	Os05g0154700	54.37	9	4902297	4908926	0.64	QGFD9.3	Reduces grain length
Sobic.001G170800	*Transport protein*	-	36.26	1	14278553	14283299	0.36	QGFD1.9	
Sobic.004G133600	*ZmSWEET4c*	Zm00001d015912	71.54	4	21285737	21291316	1.65	QGFD4.3	
Sobic.010G110100	*bZIP47*	LOC_Os06g15480	55.75	10	11004657	11007712	0.23	QGFD10.4	Suppresses cell proliferation and regulates grain size and weight negatively
Sobic.010G111200	*GSR1/GW6/ GASR7*	LOC_Os06g15620	55.80	10	11197868	11198820	0.27	QGFD10.4	Mutants show reduced GA content and decreased grain width and weight
Sobic.001G482600	*TIFY 11b*	Os03g0181100	187.95	1	75415876	75416893	0.13	QGFD1.13	TIFY gene
Sobic.001G101700	*GIF1*	LOC_Os03g52320	22.82	1	7782002	7785807	0.11	QGFD1.17	Transcriptional cofactor *GIF1* interacts with *GRF4* and enhance grain size

Gene ID	Gene name	Rice/Maize/Arabidopsis orthologue	cM	LG	START	END	cM from QTL	GFD QTL	Function
Sobic.001G109100	*DLT2*	LOC_Os03g51330	24.12	1	8469856	8473335	1.41	QGFD1.17	A GRAS-family member enhances transcriptional activity of *DLT2-DLT-BZR1* complex to modulate BR pathway signalling reducing grain size
Sobic.001G120900	*SLR1*	LOC_Os03g49990	25.32	1	9381697	9384098	1.19	QGFD1.11	SLR1 negatively affects grain size via GA signalling
Sobic.001G455900	*MADS1/qLGY3*	LOC_Os03g11614	182.92	1	73188105	73199446	0.42	QGFD1.6	Encodes MADS1 TF which interacts with DEP1 and affects grain size
Sobic.001G481400	*LG3*	LOC_Os03g08470	187.70	1	75312217	75313912	0.39	QGFD1.13	Encodes TF which facilitates cell elongation and increases grain size
Sobic.002G192600	*NAC20/26*	LOC_Os01g01470	93.48	2	57932635	57934018	0.24	QGFD2.6	It positively regulates the genes involved in starch and storage protein biosynthesis.
Sobic.002G360900	*GASR9*	LOC_Os07g40240	178.27	2	72279450	72280318	0.23	QGFD2.11	Encodes a Gibberellic acid-stimulated transcript (GAST) family protein which facilitates cell elongation resulting in longer grains
Sobic.002G271200	*UGE3*	LOC_Os09g35800	146.80	2	65474918	65477505	0.34	QGFD2.7	Shows UDP-galactose/glucose epimerase activity that facilitates substrates for polymerization of polysaccharides
Sobic.002G054400	*PK2/PKpa1*	LOC_Os07g08340	22.58	2	5219182	5223489	1.00	QGFD2.3	Encodes plastidic pyruvate kinase which takes part in biosynthesis of starch in endosperm, formation of compound granule and grain filling
Sobic.003G257400	*BG3/PUP4*	LOC_Os01g48800	109.98	3	59557217	59558392	0.79	QGFD3.2	Encodes purine permease which maintains cytokinin distribution and positively regulates grain size
Sobic.003G376000	*AAP6*	LOC_Os01g65670	144.20	3	69059581	69065077	0.08	QGFD3.11	Positively regulates seed protein content and quality
Sobic.003G213800	*SBEIII*	LOC_Os06g26234	69.87	3	54790313	54793810	1.68	QGFD3.10	Involved in upregulation of starch metabolism pathway and starch biosynthesis
Sobic.003G230500	*Sh2/APL2*	LOC_Os01g44220	87.20	3	57000119	57007815	1.01	QGFD3.9	Encodes large subunit of ADP-glucose pyrophosphorylase It acts as starch biosynthetic enzyme which

Gene ID	Gene name	Rice/Maize/Arabidopsis orthologue	cM	LG	START	END	cM from QTL	GFD QTL	Function
									suppresses starch biosynthesis pathway
Sobic.004G238600	*SBEIII*	LOC_Os02g51070	107.25	4	58642327	58646857	0.42	QGFD4.6	Involved in upregulation of starch metabolism pathway and starch biosynthesis
Sobic.004G256800	*AAP10*	LOC_Os02g49060	116.48	4	60268993	60273180	0.00	QGFD4.1	Encodes amino acid permease which loads amino acid in endosperm
Sobic.007G051700	*ASP1*	LOC_Os08g06480	58.48	7	5282384	5291841	0.09	QGFD7.1	Encodes a transcriptional co-repressor which affecting branching and spikelet development reducing grain size
Sobic.007G054700	*NF-YC10*	LOC_Os01g24460	58.66	7	5540371	5542244	0.27	QGFD7.1	Positively regulates cell division in spikelet hull cells and endosperm increasing grain width, and grain weight
Sobic.010G022600	*Wx*	LOC_Os06g04200	26.92	10	1860964	1865278	0.07	QGFD10.6	Encodes for granule-bound starch synthase (GBSS) with a role in amylose biosynthesis in rice endosperm
Sobic.010G047700	*SSI*	LOC_Os06g06560	32.77	10	3694261	3702940	0.42	QGFD10.5	Catalyses formation of amylopectin from ADP-glucose and upregulates other enzymes involved in starch biosynthesis in endosperm
Sobic.010G072300	*Sh1*	LOC_Os06g09450	35.29	10	5859073	5867276	1.63	QGFD10.2	Downregulation of *Sh1* leads to lower production of starch. Involved in sucrose synthesis and metabolism

SbGS3 potentially has a role in moderating GFD in sorghum

Haplotype analysis for Sobic.001G341700 (*SbGS3)* a putative orthologue of GS3 in rice that is thought to function as a negative regulator of grain size (Zou et al. 2020), revealed eight different haplotype groups. A pairwise comparison of the individual haplotype groups showed that haplotype 1 had a similar effect on GFD as haplotypes 2 and 3, while haplotype 2 had similar effect on GFD as haplotypes 4,5,6 and 8. Haplotype 7 had similar effects on GFD as haplotypes 4,5 and 6 (Figure 6). Globally the haplotypes groups were significantly different from each other as shown by the Kruskal-Wallis test p-value. Interestingly, when the haplotype groups were classified based on the sorghum racial groups, the distribution of the haplotypes were more defined. While there were no significant differences within the race for the haplotype groups represented, some key haplotype groups were only present in specific sorghum races. Haplotypes 4,5 and 6 were common across all the racial groups, except for Asian Durra, while haplotype 7 was represented in Asian Durra, Caudatum and Kafir racial groups. Interestingly, haplotype 1 was only present in Asian Durra and Guinea, and haplotype 3 was only present in the Guinea race. Haplotypes 1 and 3 had longer GFD in comparison to the rest of the haplotypes, suggesting that they could be carrying a loss of function allele for *SbGS3*. Finally, no specific racial group had all the haplotype groups represented.

Discussion

Increasing cereal crop productivity in challenging production environments caused by the effects of a changing climate remains a high priority for cereal breeders. To date increases in yield have mainly been achieved by increases in grain number (Boyles et al. 2016; Khush 1995; Otwani et al. 2024; Russell

1991) which appears to have greater variability across cereals (Sadras 2007). However there does appear to be variation in grain size that could be exploited to increase yield (Fernández et al. 2022; Xing et al. 2023), studies exploring grain size related traits like GFD that could contribute to increases in grain sizes are lacking. Further, studies on the genetic underpinning of GFD and its contribution to grain size are scarce. To our knowledge, no studies are available that explore the genetic control of GFD in sorghum. This study is the first and largest of its kind to explore the genomic regions associated with GFD in sorghum. We report considerable exploitable genetic variation for GFD, its association with grain size at both phenotypic and genetic levels and propose some candidate genes for GFD in grain sorghum.

Variation for GFD in grain sorghum

The phenotypic distribution of GFD revealed a wide range of GFD across the tested genotypes and environments, suggesting that GFD is a quantitative trait controlled by multiple loci. The moderate broad sense heritability too shows that GFD could be utilised as a potential useful trait in breeding programmes. Despite accounting for the effects of temperature in the estimation of GFD, the moderate genetic correlation between HRF1 and GAT and HRF2 and GAT environments reveal that there could be other factors, genetic or environmental that influence the estimates of GFD in the tested genotypes. First, could be the contribution of other environmental factors like radiation that was not accounted for in these experiments. Another plausible reason would be the possibility that the tested genotypes could potentially have different cardinal temperature requirements for the grain filling phase as has been suggested by (Tirfessa et al. 2020) in sorghum. Further, the significant positive correlation between GFD, DTM and TKW in these diverse genotypes show that there is opportunity to manipulate GFD without penalty to DTM or TKW. DTM stability is crucial in many breeding programmes as it dictates choices for growers to target suitable varieties for specific season lengths. The positive association with TKW implies that increased grain sizes could be attained by increasing the GFD. Additionally, the lack of association of GFD with DTF, is important to guide decisions for possible trait introgression strategies. Elite breeding lines could benefit from introgression of longer GFD attribute with little or no penalty to the desired flowering window, making the trait attractive to breeders. Since many commercial sorghum breeding programmes rely almost entirely on the kafir/Caudatum crosses (Otwani et al. 2024) due to limitations imposed by the cytoplasmic male sterility system used (Jordan et al. 2011; Reddy et al. 2007), wide hybridisation across all the sorghum racial groups (Weltzien et al. 2006) would provide new opportunities for yield improvement. Targeted crosses including genotypes from the guinea race that showed consistent longer GFD and are reported to have large grain sizes (Sapkota 2021; Tao et al. 2020) would be a good starting point.

GFD is intricately linked to grain size in sorghum

Single marker analysis of the GFD SNPs, using TKW trait showed high fidelity and correspondence of these SNPs for TKW, suggesting that GFD in highly associated with TKW in the tested genotypes. Additionally, the association of GFD QTLs with candidate genes identified for grain size in sorghum (Jain et al. 2024; Tao et al. 2017) reinforces these observations. 30% of the candidate genes for grain size were in LD with 50% of the GFD QTLs. These candidate genes included a validated gene in sorghum Sobic.001G341700, whose QTL, qTGW1a encodes a G-protein subunit negatively regulating grain size (Zou et al. 2020). Haplotype analysis of GFD for this candidate gene revealed that sorghum racial group guinea, known to have large seed sizes (Sapkota 2021; Tao et al. 2020) and longer GFD had a unique haplotype not present in all the other racial groups corroborating the intricate link between GFD and TKW. Another candidate gene, Sobic.007G193500 whose rice orthologue LOC_Os08g41940, encodes *SPL16/qGW8* gene which is indicated to enhance cell proliferation increasing grain width and grain filling (Wang et al. 2015a; Wang et al. 2012) was also enriched within the GFD QTLs. Previous studies in sorghum (Yang et al. 2009), maize (Fernández et al. 2022; Xing et al. 2023), rice (Wang et al. 2008; Yang et al. 2008), wheat (Chapman et al. 2021; Xie et al. 2015) and barley (Radchuk et al. 2021) discuss potential of utilising grain filling dynamics for yield improvement. The observed links of GFD and TKW both at phenotype and genomic levels provide opportunities to explore more these observations.

GFD is dynamic and determined by mechanisms happening both before and after anthesis in sorghum

GFD is estimated from flowering to maturity in many crop species including sorghum (Gambín and Borrás 2012). The observation that 77% of expressed genes that were in LD with GFD QTLs were exclusively expressed in the early pre anthesis and post anthesis period reveal that determination of GFD like many panicle and grain associated traits (Van Oosterom and Hammer 2008) happen before anthesis. Most of these candidate genes were associated with the regulation of cell proliferation and elongation through hormone mediated pathways in rice. The rice orthologue of Sobic.010G110100, LOC_Os06g15480 encodes a gene *bZIP47* which suppresses cell proliferation early in the pre anthesis phase in rice negatively affecting grain size and weight (Hao et al. 2021).

Looking at the temporal expression profiles of the enriched candidate genes within the GFD QTLs revealed that these genes could potentially be clustered into three broad groups. First are candidate genes expressed early pre and post anthesis and appear to mediate cell division, elongation and proliferation and eventually determine the cell size and number of the panicle and floral organs (Liu et al. 2015; Lo et al. 2020; Yan et al. 2024). Similar observations of pre anthesis organ size in sorghum (Takanashi et al. 2021; Yang et al. 2009) has been associated with grain size and grain filling duration. Secondly are genes associated with phytohormone mediated regulation of cell size, cell number and GFD. The sorghum candidate genes Sobic.001G120900 and Sobic.010G111200 are orthologous to rice genes *SLR1* (Lan et al. 2020) and *GSR1/GW6/GASR7* (Shi et al. 2020) that reduce grain weight and increase grain width and weight respectively through GA signalling. These candidate genes had peak expression in the early pre anthesis and post anthesis phase in sorghum (Jain et al. 2024). Candidate genes associated with BR signalling Sobic.004G237000, Sobic.001G172400 and Sobic.001G109100 orthologous to rice candidate genes *PGL2/BU1, BRD1* and *DLT2* respectively were highly expressed early in the grain filling period 1- 10 days post anthesis. *BU1* has been reported to upregulate *BDG1* to increase grain length in rice (Heang and Sassa 2012b). The rice orthologue of Sobic.001G488500, LOC_Os03g07530 (Chen et al. 2013) is associated with ethylene mediated grain length increases in the spikelet hull. The high expression of Sobic.001G488500 during early grain filling in sorghum (Jain et al. 2024) is consistent with the role of ethylene in grain filling, fruit ripening and progression to maturity (Kim et al. 2013; Magar et al. 2024; Patterson and Bleecker 2004; Sexton and Roberts 1982). Candidate gene LOC_Os01g48800 in rice (Xiao et al. 2019; Yin et al. 2020), orthologous to Sobic.003G257400, encodes purine permease maintaining distribution of cytokinin and regulating grain size positively. Sobic.003G257400 high expression during grain filling suggests its role in grain filling consistent with reports in sorghum on the role of cytokinins in mediating grain filling and grain size (Heiniger et al. 1993), in wheat (Wheeler 1972) and maize (Xu et al. 2024). Finally, are candidate genes associated with starch biosynthesis, metabolism,

transport, storage and protein storage. These candidate genes were expressed from early anthesis through to physiological maturity in sorghum (Jain et al. 2024). The sorghum candidate genes Sobic.003G213800 and Sobic.004G238600 were highly expressed in early anthesis and start of grain filling. Their rice orthologues, LOC_Os06g26234 and LOC_Os02g51070 are involved in starch biosynthesis, metabolism and accumulation in the endosperm (Kang et al. 2013; Li et al. 2018). Starch accumulation has been shown to occur during grain filling in rice (Liu et al. 2024) suggesting that the identified candidate gene could have a role in sorghum grain filling. *Shrunken 1* and *shrunken 2*, identified in maize (Chourey and Nelson 1976; Hannah and Nelson Jr 1976) which are suppressors of starch biosynthesis, were enriched in GFD QTLs, and appear to potentially mediate GFD in sorghum and other cereals. The candidate gene Sobic.010G022600 corresponding to the *waxy* gene which encodes granule bound starch synthase and mediates amylose biosynthesis in cereals (McIntyre et al. 2008; Zhang et al. 2021a). Overall, these findings show that GFD is a complex trait and potentially determined from early in the panicle development phase in sorghum. These revelations could suggest potential interactions of GFD with grain number determination (Van Oosterom and Hammer 2008) and require further investigations to reveal the nature of such interactions and their relevance in breeding for extended GFD genotypes.

Deploying extended GFD genotypes in sorghum breeding programmes

The results from the current study highlight plausible strategies for introgression of extended GFD trait into elite sorghum genotypes. First, since commercial hybrid breeding programmes have utilised the caudatum and kafir races in most of the current elite lines and hybrids, wide hybridization utilising the other sorghum racial groups, especially guinea would provide opportunity to develop extended GFD lines. Secondly, since the genomic regions associated with GFD and grain size have been identified, fine gene mapping could be explored in biparental populations to further understand the trait. These genomic regions may also be explored in elite breeding population to identify if indirect selection for extended GFD has happened in the breeding programs. These strategies together would contribute to quantify the value of GFD trait to underpin further resource investment.

Conclusion

This study established an intricate link between GFD and grain size in grain sorghum, reports candidate genes that potentially mediate GFD and could be targeted for breeding for longer GFD genotypes. The revelation that GFD is potentially determined by mechanisms that occur before anthesis like many other panicle and seed related traits presents opportunity to unravel associations between GFD and grain number among others and explore strategies to use them in breeding programmes. Additionally, a subset of GFD associated candidate genes could be utilised directly as a selection index together with grain size and number to improve selection outcomes in sorghum breeding.

Declarations

Acknowledgements

The authors acknowledge the contribution of the University of Queensland and Queensland Government's sorghum pre-breeding field team.

Author contributions

Conceptualization D.J., D.O and E.M.; Data curation D.O and C.H.; Formal analysis D.O. and C.H.; Funding acquisition D.J.; Investigation D.O.; Methodology D.O.; Project administration D.J and E.M.; Resources D.J., E.M. and A.C.; Software D.O. and C.H; Supervision A.K., E.M., A.C., C.H. and D.J.; Validation D.O. and C.H.; Visualization D.O.; Writing – original draft D.O.; Writing – review & editing D.O., D.J., E.M., A.K., Y.T. and C.H.

Conflict of interest

Non declared.

Funding statement

The study was funded from investments by the Queensland Government, and The University of Queensland. D.O. is a beneficiary of the University of Queensland RTP Scholarship.

Data availability

All data and code are available from the corresponding author on reasonable request.

References

1. Ashikari M, Wu J, Yano M, Sasaki T, Yoshimura A (1999) Rice gibberellin-insensitive dwarf mutant gene Dwarf 1 encodes the α-subunit of GTP-binding protein Proceedings of the National Academy of Sciences 96:10284–10289 doi:10.1073/pnas.96.18.10284
2. Boyles RE, Cooper EA, Myers MT, Brenton Z, Rauh BL, Morris GP, Kresovich S (2016) Genome-Wide Association Studies of Grain Yield Components in Diverse Sorghum Germplasm. Plant Genome 9. https://doi.org/10.3835/plantgenome2015.09.0091. :plantgenome2015.2009.0091 doi
3. Boyles RE, Pfeiffer BK, Cooper EA, Zielinski KJ, Myers MT, Rooney WL, Kresovich S (2017) Quantitative trait loci mapping of agronomic and yield traits in two grain sorghum biparental. families Crop Sci 57:2443–2456
4. Cai Y et al (2018) OsPK2 encodes a plastidic pyruvate kinase involved in rice endosperm starch synthesis, compound granule formation and grain filling. Plant Biotechnol J 16:1878–1891. 10.1111/pbi.12923

5. Call L, Haider E, D'Amico S, Reiter E, Grausgruber H (2021) Synthesis and accumulation of amylase-trypsin inhibitors and changes in carbohydrate profile during grain development of bread wheat (Triticum aestivum L). BMC Plant Biol 21:113. 10.1186/s12870-021-02886-x
6. Chapman EA, Orford S, Lage J, Griffiths S (2021) Delaying or delivering: identification of novel NAM-1 alleles that delay senescence to extend wheat grain fill duration. J Exp Bot 72:7710–7728. 10.1093/jxb/erab368
7. Chen Y, Xu Y, Luo W, Li W, Chen N, Zhang D, Chong K (2013) The F-Box Protein OsFBK12 Targets OsSAMS1 for Degradation and Affects Pleiotropic Phenotypes, Including Leaf Senescence. Rice Plant Physiol 163:1673–1685. 10.1104/pp.113.224527
8. Chen Y et al (2020) Cloning of wheat keto-acyl thiolase 2B reveals a role of jasmonic acid in grain weight determination. Nat Commun 11. 10.1038/s41467-020-20133-z
9. Chourey PS, Nelson OE (1976) The enzymatic deficiency conditioned by the shrunken-1 mutations in maize. Biochem Genet 14:1041–1055. 10.1007/BF00485135
10. Costes C, Navarro Sanz S, Calatayud C, Soriano A, Mameri H, Terrier N, Francin-Allami M (2024) Transcriptomic analysis of developing sorghum grains to detect genes related to cell wall biosynthesis and remodelling. BMC Genomic Data 25:14
11. Cruet-Burgos C, Rhodes DH (2023) Unraveling transcriptomics of sorghum grain carotenoids: a step forward for biofortification. BMC Genomics 24:233
12. Cullis BR, Smith AB, Coombes NE (2006) On the design of early generation variety trials with correlated data. J Agricultural Biol Environ Stat 11:381–393. 10.1198/108571106X154443
13. Deng Y, Wang J, Zhang Z, Wu Y (2020) Transactivation of Sus1 and Sus2 by Opaque2 is an essential supplement to sucrose synthase-mediated endosperm filling in maize. Plant Biotechnol J 18:1897–1907. https://doi.org/10.1111/pbi.13349
14. Done A (1986) Duration of grain filling and yield of grain sorghum. Agron Soc 5:175–180
15. Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities. of fresh leaf tissue Phytochemical bulletin
16. Dray S, Dufour A-B (2007) The ade4 Package: Implementing the Duality Diagram for Ecologists. J Stat Softw 22:1–20. 10.18637/jss.v022.i04
17. Eckert C et al (2014) The vacuolar calcium sensors CBL 2 and CBL 3 affect seed size and embryonic development in Arabidopsis thaliana. Plant J 78:146–156
18. Edet OU, Gorafi YS, Nasuda S, Tsujimoto H (2018) DArTseq-based analysis of genomic relationships among species of tribe. Triticeae Sci Rep 8:16397
19. Fang W, Wang Z, Cui R, Li J, Li Y (2012) Maternal control of seed size by EOD3/CYP78A6 in Arabidopsis thaliana. Plant J 70:929–939
20. Fernández JA, Messina CD, Salinas A, Prasad PV, Nippert JB, Ciampitti IA (2022) Kernel weight contribution to yield genetic gain of maize: A global review and US case studies. J Exp Bot 73:3597–3609
21. Fujita N, Yoshida M, Asakura N, Ohdan T, Miyao A, Hirochika H, Nakamura Y (2006) Function and characterization of starch synthase I using mutants in rice. Plant Physiol 140:1070–1084. 10.1104/pp.105.071845
22. Gambín BL, Borrás L (2012) Genotypic diversity in sorghum inbred lines for grain-filling patterns and other related agronomic traits. Crop Pasture Sci 62:1026–1036
23. Gao X, Starmer J, Martin ER (2008) A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. Genet Epidemiol 32:361–369. https://doi.org/10.1002/gepi.20310
24. Hakata M et al (2012) Suppression of α-amylase genes improves quality of rice grain ripened under high temperature. Plant Biotechnol J 10:1110–1117. https://doi.org/10.1111/j.1467-7652.2012.00741.x
25. Hammer GL, Muchow RC (1994) Assessing climatic risk to sorghum production in water-limited subtropical environments I. Development and testing of a simulation model. Field Crops Res 36:221–234. https://doi.org/10.1016/0378-4290(94)90114-7
26. Han L et al (2015) Fine mapping of qGW1, a major QTL for grain weight in sorghum. Theor Appl Genet 128:1813–1825
27. Hannah LC, Nelson OE Jr (1976) Characterization of ADP-glucose pyrophosphorylase from shrunken-2 and brittle-2 mutants of maize. Biochem Genet 14:547–560. 10.1007/BF00485834
28. Hao J et al (2021) The GW2-WG1-OsbZIP47 pathway controls grain size and weight in rice. Mol Plant 14:1266–1280
29. Hartings H, Maddaloni M, Lazzaroni N, Di Fonzo N, Motto M, Salamini F, Thompson R (1989) The O2 gene which regulates zein deposition in maize endosperm encodes a protein with structural homologies to transcriptional activators. EMBO J 8:2795–2801
30. Hashimoto S et al (2023) An analysis of sugary endosperm in sorghum: Characterization of mutant phenotypes depending on alleles of the corresponding starch debranching enzyme. Front Plant Sci 14:1114935
31. Heang D, Sassa H (2012a) Antagonistic actions of HLH/bHLH proteins are involved in grain length and weight in rice. PLoS ONE 7:e31325
32. Heang D, Sassa H (2012b) An atypical bHLH protein encoded by POSITIVE REGULATOR OF GRAIN LENGTH 2 is involved in controlling grain length and weight of rice through interaction with a typical bHLH protein. APG Breed Sci 62:133–141
33. Heiniger RW, Vanderlip RL, Kofoid KD (1993) Influence of pollination pattern on intrapanicle caryopsis weight in sorghum. Crop Sci 33:549–555
34. Hong Z et al (2002) Loss-of-function of a rice brassinosteroid biosynthetic enzyme, C-6 oxidase, prevents the organized arrangement and polar elongation of cells in the leaves and stem. Plant J 32:495–508. 10.1046/j.1365-313X.2002.01438.x
35. Jain R, Dhaka N, Krishnan K, Yadav G, Priyam P, Sharma MK, Sharma RA (2024) Temporal Gene Expression Profiles From Pollination to Seed Maturity in Sorghum Provide Core Candidates for Engineering Seed Traits Plant, Cell & Environment n/a. 1–29. https://doi.org/10.1111/pce.15134
36. Jang S, Li HY (2017) Oryza sativa BRASSINOSTEROID UPREGULATED1 LIKE1 induces the expression of a gene encoding a small leucine-rich-repeat protein to positively regulate lamina inclination and grain size in rice Frontiers. Plant Sci 8. 10.3389/fpls.2017.01253

37. Jia S, Xiong Y, Xiao, Wang X, Yao J (2019) OsNF-YC10, a seed preferentially expressed gene regulates grain width by affecting cell proliferation in rice. Plant Sci 280:219–227. 10.1016/j.plantsci.2018.09.021
38. Jombart T (2008) adegenet: a R package for the multivariate analysis of. genetic markers Bioinf 24:1403–1405. 10.1093/bioinformatics/btn129
39. Jordan DR, Klein RR, Sakrewski KG, Henzell RG, Klein PE, Mace ES (2011) Mapping and characterization of Rf5: a new gene conditioning pollen fertility restoration in A1 and A2 cytoplasm in sorghum (Sorghum bicolor (L.) Moench. Theor Appl Genet 123:383–396. 10.1007/s00122-011-1591-y
40. Kang G, Li S, Zhang M, Peng H, Wang C, Zhu Y, Guo T (2013) Molecular cloning and expression analysis of the starch-branching enzyme III gene from common wheat (Triticum aestivum). Biochem Genet 51:377–386. 10.1007/s10528-013-9570-4
41. Khush GS (1995) Breaking the yield frontier of rice GeoJournal 35:329–332
42. Kim J, Dotson B, Rey C, Lindsey J, Bleecker AB, Binder BM, Patterson SE (2013) New clothes for the jasmonic acid receptor COI1: delayed abscission, meristem arrest and apical dominance. PLoS ONE 8:e60505
43. Kitagawa K et al (2010) A Novel Kinesin 13 Protein Regulating Rice Seed Length Plant. Cell Physiol 51:1315–1329. 10.1093/pcp/pcq092
44. Lan J et al (2020) Small grain and semi-dwarf 3, a WRKY transcription factor, negatively regulates plant height and grain size by stabilizing SLR1 expression in rice. Plant Mol Biol 104:429–450
45. Li H, Durbin R (2009) Fast and accurate short read alignment with. Burrows–Wheeler transform Bioinf 25:1754–1760
46. Li J, Chu H, Zhang Y, Mou T, Wu C, Zhang Q, Xu J (2012) The rice HGW gene encodes a ubiquitin-associated (UBA) domain protein that regulates heading date and grain weight. PLoS ONE 7:e34231
47. Li QF et al (2018) Down-Regulation of SSSII-2 Gene Expression Results in Novel Low-Amylose Rice with Soft. Transparent Grains J Agricultural Food Chem 66:9750–9760. 10.1021/acs.jafc.8b02913
48. Li S et al (2016) The OsmiR396c-OsGRF4-OsGIF1 regulatory module determines grain size and yield in rice. Plant Biotechnol J 14:2134–2146. 10.1111/pbi.12569
49. Li X et al (2019) OsGASR9 positively regulates grain size and yield in rice (Oryza sativa). Plant Sci 286:17–27. 10.1016/j.plantsci.2019.03.008
50. Li Y et al (2021) ZmGRAS11, transactivated by Opaque2, positively regulates kernel size in maize. J Integr Plant Biol 63:2031–2037. https://doi.org/10.1111/jipb.13198
51. Liu L et al (2015) Activation of Big Grain1 significantly improves grain size by regulating auxin transport in rice Proceedings of the National Academy of Sciences 112:11102–11107
52. Liu Q et al (2018) G-protein βγ subunits determine grain size through interaction with MADS-domain transcription factors in rice. Nat Commun 9. 10.1038/s41467-018-03047-9
53. Liu X, Huang M, Fan B, Buckler ES, Zhang Z (2016) Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. PLoS Genet 12:e1005767
54. Liu X et al (2024) The evolution of granular surface structure and functional properties in rice starch during grain filling International. J Biol Macromol 279:135523. https://doi.org/10.1016/j.ijbiomac.2024.135523
55. Lo SF et al (2020) Rice Big Grain 1 promotes cell division to enhance organ development, stress tolerance and grain yield. Plant Biotechnol J 18:1969–1983
56. Mace E et al (2019) The Sorghum QTL Atlas: a powerful tool for trait dissection. Comp genomics crop Improv Theoretical Appl Genet 132:751–766. 10.1007/s00122-018-3212-5
57. Magar MM, Liu H, Yan G (2024) Wheat TaAP2/ERF Genes Regulate Heat Tolerance Through Ethylene Signaling at Grain-Filling Stage. J Plant Growth Regul :1–18
58. Martin A et al (2006) Two cytosolic glutamine synthetase isoforms of maize are specifically involved in the control of grain production. Plant Cell 18:3252–3274. 10.1105/tpc.106.042689
59. McCormick RF et al (2018) The Sorghum bicolor reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. Plant J 93:338–354
60. McIntyre C, Drenth J, Gonzalez N, Henzell R, Jordan D (2008) Molecular characterization of the waxy locus in. sorghum Genome 51:524–533
61. Mori M et al (2002) Isolation and characterization of a rice dwarf mutant with a defect in brassinosteroid biosynthesis. Plant Physiol 130:1152–1161. 10.1104/pp.007179
62. Nagasawa N et al (2013) GIANT EMBRYO encodes CYP78A13, required for proper size balance between embryo and endosperm in rice. Plant J 75:592–605
63. Otwani D, Hunt C, Cruickshank A, Powell O, Koltunow A, Mace E, Jordan D (2024) Adaptation and plasticity of yield in hybrid and inbred sorghum Crop Science 1–11 https://doi.org/10.1002/csc2.21160
64. Otwani D et al (2025) Extended grain filling has potential to improve yield in grain sorghum. J Exp Bot. 10.1093/jxb/eraf117
65. Paterson AH et al (1995) Convergent domestication of cereal crops by independent mutations at corresponding. genetic loci Sci 269:1714–1718
66. Patterson SE, Bleecker AB (2004) Ethylene-dependent and-independent processes associated with floral organ abscission in. Arabidopsis Plant Physiol 134:194–203
67. Peng B et al (2014) OsAAP6 functions as an important regulator of grain protein content and nutritional quality in rice Nature Communications. 5. 10.1038/ncomms5847

68. Radchuk V et al (2021) Grain filling in barley relies on developmentally controlled programmed cell death. Commun Biology 4:428. 10.1038/s42003-021-01953-1
69. RCoreTeam (2024) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria doi:https://www.R-project.org/
70. Reddy BVS et al (2007) Combining ability and heterosis as influenced by male-sterility inducing cytoplasms in sorghum [Sorghum bicolor (L.) Moench]. Euphytica 154:153–164. 10.1007/s10681-006-9281-6
71. Riefler M, Novak O, Strnad M, Schmülling T (2005) Arabidopsis Cytokinin Receptor Mutants Reveal Functions in Shoot Growth, Leaf Senescence, Seed Size, Germination, Root Development, and Cytokinin Metabolism. Plant Cell 18:40–54. 10.1105/tpc.105.037796
72. Russell WA (1991) Genetic Improvement of Maize Yields. In: Sparks DL (ed) Advances in Agronomy, vol 46. Academic Press, pp 245–298. https://doi.org/10.1016/S0065-2113(08)60582-9
73. Sadras VO (2007) Evolutionary aspects of the trade-off between seed size and number in crops. Field Crops Res 100:125–138. https://doi.org/10.1016/j.fcr.2006.07.004
74. Sapkota SK (2021) Genomics-Assisted Breeding for Grain Yield and Composition in Sorghum. Clemson University
75. Schaffer JA (1981) The Effect of Planting Data and Environment on the Phenology and Modelling of Grain Sorghum, *Sorghum Bicolor* (L.) Moench. Kansas State University
76. Schmidt R, Schippers JHM, Mieulet D, Watanabe M, Hoefgen R, Guiderdoni E, Mueller-Roeber B (2014) SALT-RESPONSIVE ERF1 is a negative regulator of grain filling and gibberellin-mediated seedling establishment in rice. Mol Plant 7:404–421. 10.1093/mp/sst131
77. Segami S, Kono I, Ando T, Yano M, Kitano H, Miura K, Iwasaki Y (2012) Small and round seed 5 gene encodes alphatubulin regulating seed cell elongation. rice Rice 5:1–10. 10.1186/1939-8433-5-4
78. Segami S, Takehara K, Yamamoto T, Kido S, Kondo S, Iwasaki Y, Miura K (2017) Overexpression of SRS5 improves grain size of brassinosteroid-related dwarf mutants in rice (Oryza sativa L). Breed Sci 67:393–397. 10.1270/jsbbs.16198
79. Sexton R, Roberts JA (1982) Cell biology of abscission. Annu Rev Plant Physiol 33:133–162
80. Shi CL, Dong NQ, Guo T, Ye WW, Shan JX, Lin HX (2020) A quantitative trait locus GW6 controls rice grain size and yield through the gibberellin pathway. Plant J 103:1174–1188
81. Sosso D et al (2015) Seed filling in domesticated maize and rice depends on SWEET-mediated hexose transport. Nat Genet 47:1489–1493
82. Sun C et al (2023) MATE transporter GFD1 cooperates with sugar transporters, mediates carbohydrate partitioning and controls grain-filling duration, grain size and number in rice. Plant Biotechnol J 21:621–634
83. Sunohara H, Kawai T, Shimizu-Sato S, Sato Y, Sato K, Kitano H (2009) A dominant mutation of TWISTED DWARF 1 encoding an α-tubulin protein causes severe dwarfism and right helical growth in rice Genes and. Genetic Syst 84:209–218. 10.1266/ggs.84.209
84. Takanashi H, Shichijo M, Sakamoto L, Kajiya-Kanegae H, Iwata H, Sakamoto W, Tsutsumi N (2021) Genetic dissection of QTLs associated with spikelet-related traits and grain size in sorghum Scientific reports 11:9398
85. Tang Y et al (2022) OsUGE3-mediated cell wall polysaccharides accumulation improves biomass production, mechanical strength, and salt tolerance. Plant Cell Environ 45:2492–2507. 10.1111/pce.14359
86. Tao Y, Mace E, George-Jaeggli B, Hunt C, Cruickshank A, Henzell R, Jordan D (2018) Novel Grain Weight Loci Revealed in a Cross between Cultivated and Wild Sorghum. Plant Genome 11:170089. https://doi.org/10.3835/plantgenome2017.10.0089
87. Tao Y et al (2017) Whole-Genome Analysis of Candidate genes Associated with Seed Size and Weight in Sorghum bicolor Reveals Signatures of Artificial Selection and Insights into Parallel Domestication in Cereal Crops. Front Plant Sci 8. 10.3389/fpls.2017.01237
88. Tao Y et al (2021) Manipulating assimilate availability provides insight into the genes controlling grain size in sorghum. Plant J 108:231–243
89. Tao Y et al (2020) Large-scale GWAS in sorghum reveals common genetic control of grain size among cereals. Plant Biotechnol J 18:1093–1105
90. TheVSNiTeam (2023) asreml: Fits Linear Mixed Models using REML R package version 4.2.0.302 doi:
91. Tirfessa A, McLean G, Mace E, van Oosterom E, Jordan D, Hammer G (2020) Differences in temperature response of phenological development among diverse Ethiopian sorghum genotypes are linked to racial grouping and agroecological adaptation. Crop Sci 60:977–990. https://doi.org/10.1002/csc2.20128
92. Van Oosterom E, Hammer G (2008) Determination of grain number in sorghum. Field Crops Res 108:259–268
93. Wang E et al (2008) Control of rice grain-filling and yield by a gene with a potential signature of domestication. Nat Genet 40:1370–1374
94. Wang J, Chen Z, Zhang Q, Meng S, Wei C (2020a) The NAC transcription factors OsNAC20 and OsNAC26 Regulate Starch and Storage Protein synthesis. Plant Physiol 184:1775–1791. 10.1104/pp.20.00984
95. Wang S et al (2015a) The OsSPL16-GW7 regulatory module determines grain shape and simultaneously improves rice yield and grain quality. Nat Genet 47:949–954
96. Wang S et al (2012) Control of grain size, shape and quality by OsSPL16 in rice. Nat Genet 44:950–954
97. Wang S, Yang Y, Guo M, Zhong C, Yan C, Sun S (2020b) Targeted mutagenesis of amino acid transporter genes for rice quality improvement using the CRISPR/Cas9. Syst Crop J 8:457–464. 10.1016/j.cj.2020.02.005
98. Wang Y et al (2015b) Copy number variation at the GL7 locus contributes to grain size diversity in rice. Nat Genet 47:944–948
99. Weltzien E, Rattunde HFW, Clerget B, Siart S, Touré A, Sagnard F (2006) Sorghum diversity and adaptation to drought in West Africa. In
100. Wheeler A (1972) Changes in growth-substance contents during growth of wheat grains. Ann Appl Biol 72:327–334

101. Wickham H (2016) ggplot2: Elegant Graphics for Data Analysis. Springer-, New York. https://ggplot2.tidyverse.org
102. Wills DM, Whipple CJ, Takuno S, Kursel LE, Shannon LM, Ross-Ibarra J, Doebley JF (2013) From many, one: genetic control of prolificacy during maize domestication. PLoS Genet 9:e1003604
103. Wu H et al (2024) NAKED ENDOSPERM1, NAKED ENDOSPERM2, and OPAQUE2 interact to regulate gene networks in maize endosperm development. Plant Cell 36:19–39
104. Xiao W, Custard KD, Brown RC, Lemmon BE, Harada JJ, Goldberg RB, Fischer RL (2006) DNA methylation is critical for Arabidopsis embryogenesis and seed viability. Plant Cell 18:805–814
105. Xiao Y et al (2019) Big Grain3, encoding a purine permease, regulates grain size via modulating cytokinin transport in rice. J Integr Plant Biol 61:581–597. https://doi.org/10.1111/jipb.12727
106. Xie Q, Mayes S, Sparkes DL (2015) Carpel size, grain filling, and morphology determine individual grain weight in wheat. J Exp Bot 66:6715–6730. 10.1093/jxb/erv378
107. Xing G et al (2023) Changes in grain-filling characteristics of single-cross maize hybrids released in China from 1964 to 2014. J Integr Agric 22:691–700
108. Xu C et al (2024) Molecular mechanisms underlying low temperature inhibition of grain filling in maize (Zea mays L.): coordination of growth and cold responses The Plant Journal
109. Xu F et al (2015) Variations in CYP 78 A 13 coding region influence grain size and yield in rice Plant. cell Environ 38:800–811
110. Yan L et al (2024) Control of grain size and weight by the RNA-binding protein EOG1 in rice and wheat Cell Reports. 43. 10.1016/j.celrep.2024.114856
111. Yang M et al (2020) Probing early wheat grain development via transcriptomic and proteomic approaches. Funct Integr Genomics 20:63–74. 10.1007/s10142-019-00698-9
112. Yang T et al (2021) The B3 domain-containing transcription factor ZmABI19 coordinates expression of key factors required for maize seed development and grain filling. Plant Cell 33:104–128. 10.1093/plcell/koaa008
113. Yang W, Peng S, Dionisio-Sese ML, Laza RC, Visperas RM (2008) Grain filling duration, a crucial determinant of genotypic variation of grain yield in field-grown tropical irrigated rice. Field Crops Res 105:221–227
114. Yang Z, van Oosterom EJ, Jordan DR, Doherty A, Hammer GL (2010) Genetic variation in potential kernel size affects kernel growth and yield of sorghum. Crop Sci 50:685–695
115. Yang Z, Van Oosterom EJ, Jordan DR, Hammer GL (2009) Pre-anthesis ovary development determines genotypic differences in potential kernel weight in sorghum. J Exp Bot 60:1399–1408
116. Yin W et al (2020) ARGONAUTE2 enhances grain length and salt tolerance by activating BIG GRAIN3 to modulate cytokinin distribution in rice. Plant Cell 32:2292–2306
117. Yoshida A, Ohmori Y, Kitano H, Taguchi-Shiobara F, Hirano HY (2012) ABERRANT SPIKELET and PANICLE1, encoding a TOPLESS-related transcriptional co-repressor, is involved in the regulation of meristem fate in rice. Plant J 70:327–339. 10.1111/j.1365-313X.2011.04872.x
118. Yu J et al (2018) Alternative splicing of OsLG3b controls grain length and yield in japonica rice. Plant Biotechnol J 16:1667–1678. 10.1111/pbi.12903
119. Zhang B et al (2011) Golgi nucleotide sugar transporter modulates cell wall biosynthesis and plant growth in rice Proceedings of the National Academy of Sciences 108:5110–5115
120. Zhang C et al (2021a) A rare Waxy allele coordinately improves rice eating and cooking quality and grain transparency. J Integr Plant Biol 63:889–901. 10.1111/jipb.13010
121. Zhang D, Zhang M, Liang J (2021b) RGB1 Regulates Grain Development and Starch Accumulation Through Its Effect on OsYUC11-Mediated Auxin Biosynthesis in Rice Endosperm Cells Frontiers in Plant Science. 12. 10.3389/fpls.2021.585174
122. Zhang X et al (2012) Rare allele of OsPPKL1 associated with grain length causes extra-large grain and a significant yield increase in rice Proceedings of the National Academy of Sciences 109:21534–21539
123. Zou G et al (2020) Sorghum qTGW1a encodes a G-protein subunit and acts as a negative regulator of grain size. J Exp Bot 71:5389–5401
124. Zou T et al (2023) DWARF AND LOW-TILLERING 2 functions in brassinosteroid signaling and controls plant architecture and grain size in rice. Plant J 116:1766–1783. 10.1111/tpj.16464

Figures

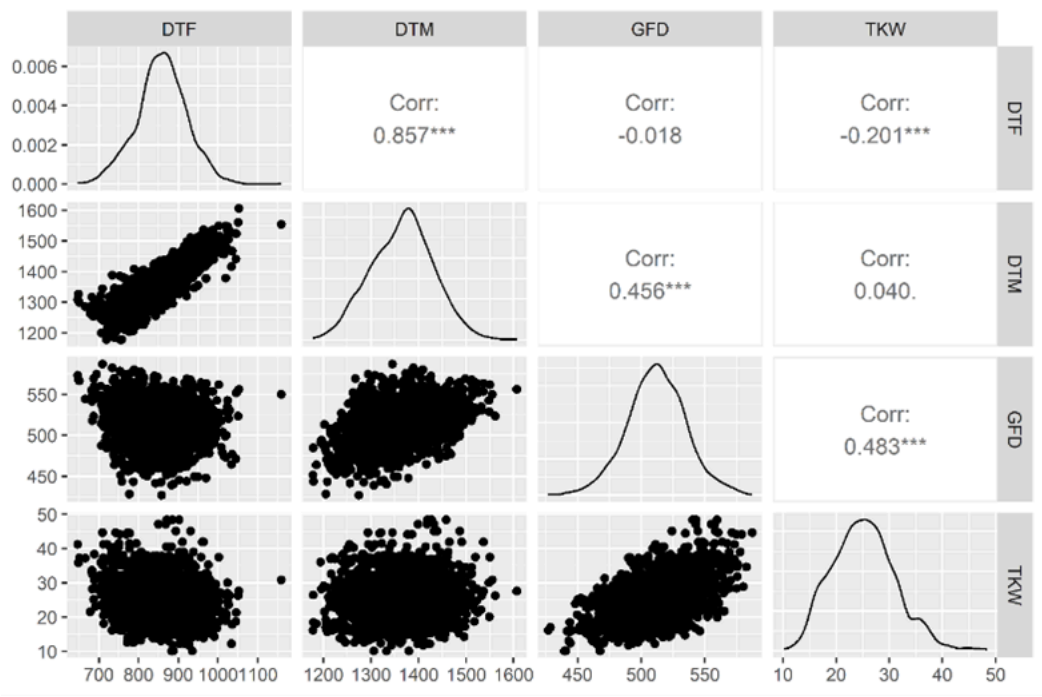

Figure 1

Figure 2: Pearson's correlation for DTF, DTM, GFD, and TKW from HRF2 data showing the association between GFD and other yield related traits in grain sorghum.

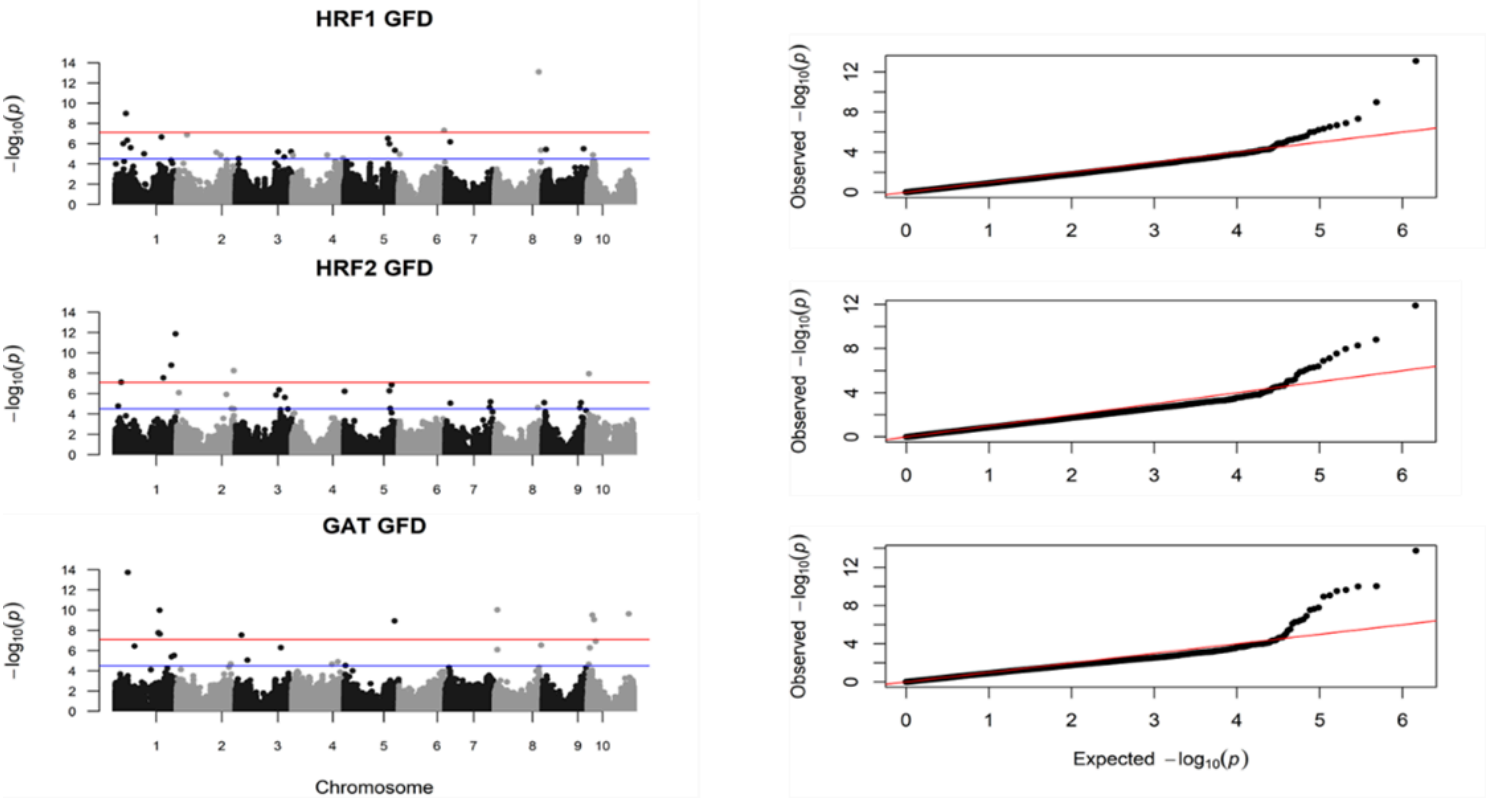

Figure 2

Figure 3: Manhattan plots and QQ plots for grain filling duration trait across the locations HFR1, HRF2 and GAT. Points above the red line on the Manhattan plot represent significant genomic regions while points above the blue horizontal line and below the red horizontal line represent suggestive genomic regions associated with grain filling duration.

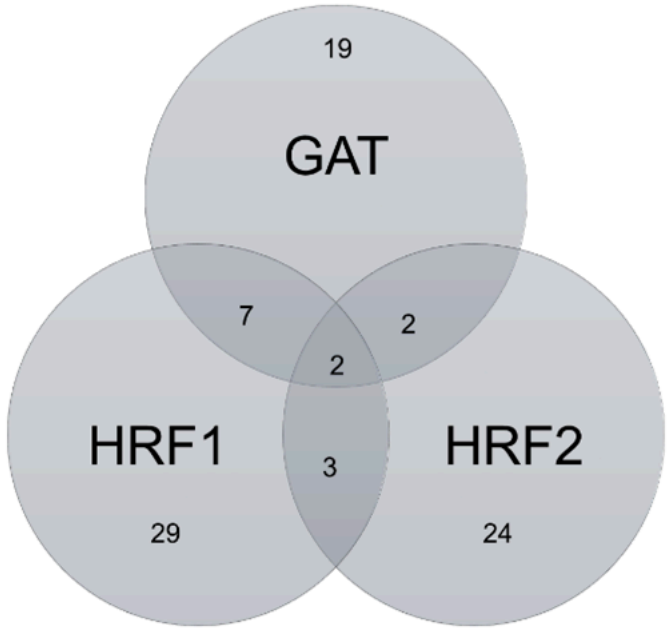

Figure 3

Figure 4: Venn diagram showing unique and common QTLs across the three environments GAT,HRF1 and HRF2. Numbers within the bigger circles denote unique QTLs per environment while numbers in the intersection of the circles denote common QTLs between those environments.

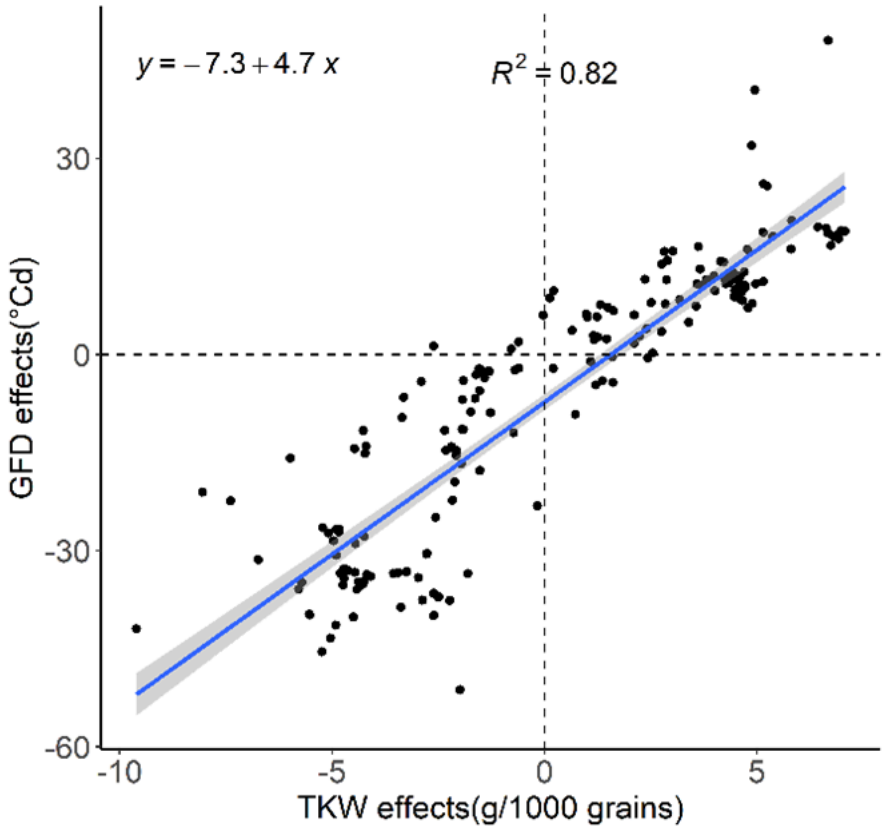

Figure 4

Figure 5: Correspondence of GFD SNP effects and TKW SNP effects when TKW trait is tested for association with GFD markers in a single marker analysis. The equation describes the relationship and the R-Squared value the magnitude of the relationship. The broken vertical and horizontal line through the origin is to show the direction of the SNP effects.

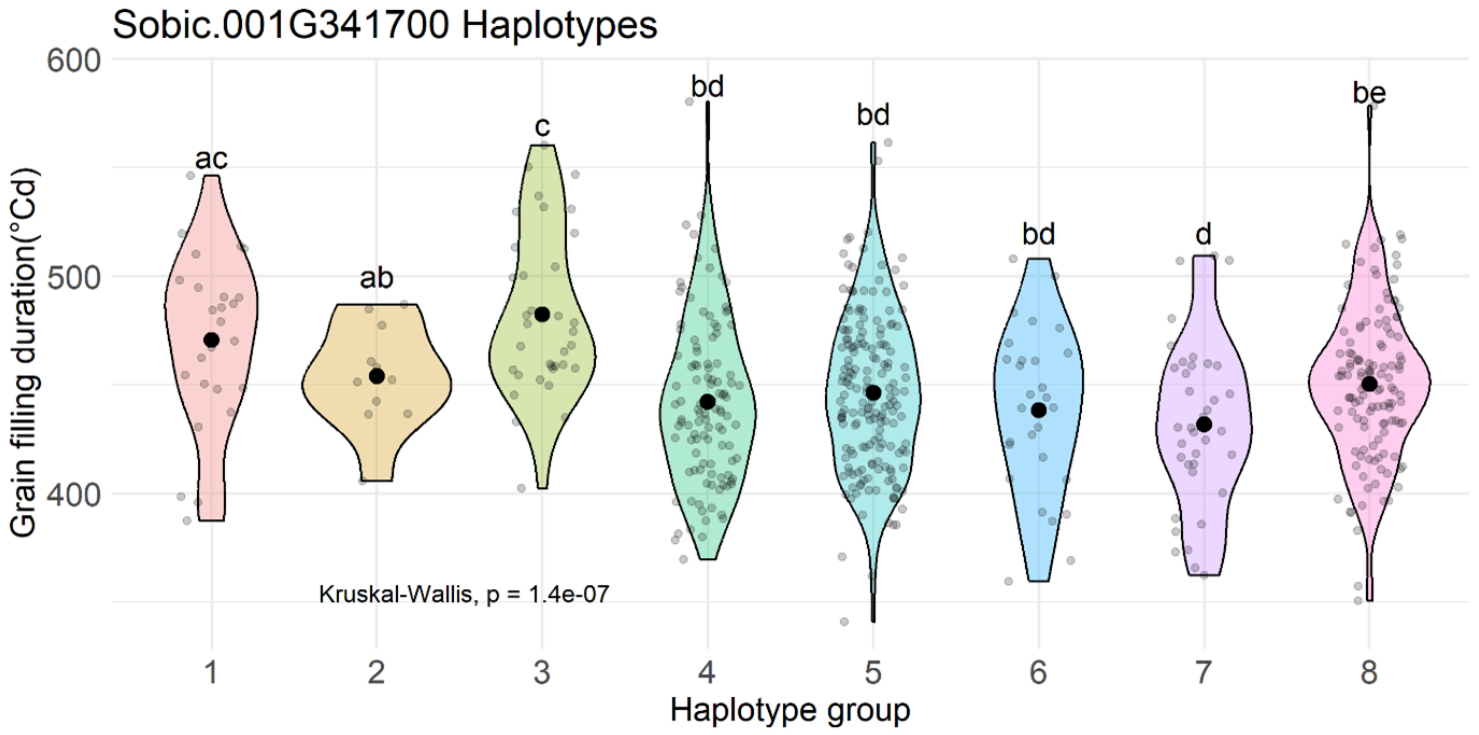

Figure 5

Figure 6: Grain filling duration across haplotype groups for Sobic.001G341700 gene in sorghum. Kruskal-Wallis P value is presented for global comparisons and Wilcoxon tests significance presented as alphabetical letters above the violin plots for differences in mean between haplotype groups at P<=0.05. Similar letters indicate no significant differences while different letters indicate significant differences.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementary2GWASCHAPTER.xlsx
- supplementrytableGWASPAPER.xlsx