

# Association Analysis of Baking and Milling Quality Traits in an Elite Soft Red Winter Wheat Population

Rupesh Gaire, Mao Huang, Clay Sneller, Carl Griffey, Gina Brown-Guedira, and Mohsen Mohammadi\*

## ABSTRACT

Although grain yield is the most important trait for growers, milling and baking industries demand high-quality and scab-free grains for various end products. To accelerate breeding of wheat (*Triticum aestivum* L.) cultivars with high grain quality, genetic dissection of quality traits is necessary. Genome-wide association studies (GWAS) were conducted to identify genomic regions responsible for milling and baking quality traits in soft red winter wheat (SRWW). Seven quality traits were evaluated in two locations and 2 yr for 270 elite SRWW lines. These traits include flour yield, softness equivalent, flour protein, and four solvent (lactose, sodium carbonate, sucrose, and water) retention capacity measurements. In this study, 27,449 single nucleotide polymorphism (SNP) markers were developed by using both genotyping-by-sequencing and 90K SNP array technologies. A linear mixed model in GWAS, accounting for population structure and kinship, was fitted to identify 18  $[-\log(P) \geq 4.0]$  genomic regions on 12 different chromosomes associated with the quality traits. Only one of these associations seems to be a previously identified quantitative trait locus, whereas others are novel associations. The most significant associations for quality traits were identified on chromosomes 1B, 2A, 2B, 4B, 5A, 7A, and 7D. Candidate gene searches, facilitated by the wheat genome assembly, led us to the identification of putative genes related to SRWW quality traits including fasciclin-like arabinogalactan. The results of this study can be used in developing high-quality SRWW varieties for the eastern region of the United States.

R. Gaire and M. Mohammadi, 915 West State St., Agronomy Dep., Purdue Univ., West Lafayette, IN, 47907; M. Huang and C. Sneller, Dep. of Horticulture and Crop Science, The Ohio State Univ., 1680 Madison Ave., Wooster, Ohio 44691; C. Griffey, 334-A Smyth Hall, Crop & Soil Environmental Sciences Dep., Virginia Tech, Blacksburg, VA, 24061-0404; G. Brown-Guedira, USDA-ARS Plant Science Research, Dep. of Crop Science, North Carolina State Univ., Raleigh, NC 27695. Received 14 Dec. 2018. Accepted 12 Feb. 2019. \*Corresponding author (mohamm20@purdue.edu). Assigned to Associate Editor Shuyu Liu.

**Abbreviations:** AACC, American Association of Cereal Chemists; DArT, diversity arrays technology; FP, flour protein; FY, flour yield; GBS, genotyping-by-sequencing; GWAS, genome-wide association study; LA, lactic acid retention capacity; LD, linkage disequilibrium; MTA, marker-trait association; PCR, polymerase chain reaction; QTL, quantitative trait locus; SE, softness equivalent; SNP, single nucleotide polymorphism; SO, sodium carbonate retention capacity; SRC, solvent retention capacity; SRWW, soft red winter wheat; SSR, simple sequence repeat; SU, sucrose retention capacity; WA, water retention capacity.

**T**HE UNITED STATES produced 9.4 Tg of soft red winter wheat (*Triticum aestivum* L., SRWW) on 2.4 million ha (6.6 million acres) of land in 2016 (USDA-NASS, 2017). The SRWW is mainly grown in the states along the Mississippi River and the eastern states including Alabama, Arkansas, Georgia, Illinois, Indiana, Kentucky, Maryland, Michigan, Missouri, North Carolina, New York, Ohio, Pennsylvania, Tennessee, Virginia, and Wisconsin (USDA-ERS, 2017). About 67% of the total production (6.3 Tg) is used for domestic consumption as food, seed, and feed (USDA-ERS, 2017), whereas ~25% (2.3 Tg) of the US SRWW was exported as grain, flour, and processed products. In the last 10 yr, on average, SRWW provided an estimated annual economic value of over US\$2.23 billion (USDA-NASS, 2017).

Although grain yield is the most important trait for growers, processing industries demand high-quality and scab-free grains for various products. Generally, scab disease (also known as Fusarium

Published in Crop Sci. 59:1085–1094 (2019).  
doi: 10.2135/cropsci2018.12.0751

© 2019 The Author(s). Re-use requires permission from the publisher.

head blight, caused by the fungus *Fusarium graminearum*) deteriorates end-use quality of wheat. In North America, SRWW is either used alone to bake biscuits and cookies or mixed with flour from hard wheat to bake cakes, doughnuts, or crackers (Gwirtz et al., 2006). Different products of SRWW require specific sets of quality parameters. For example, cookies require a protein content of 8 to 10% and water absorption of 48 to 52% (Gwirtz et al., 2006), whereas biscuits can be made using wheat with a wider range of protein (7–10%) and higher water absorption (50–54%) (Gwirtz et al., 2006). Therefore, in addition to grain yield and disease resistance, genetic dissection of quality characteristics is crucial.

End-use quality of soft wheat is defined by milling and baking characteristics. A graphical illustration of milling traits is presented in Supplemental Fig. S1. Milling quality traits of soft wheat include flour yield (FY), flour protein content (FP), and softness equivalent (SE). Flour yield is defined as the percentage of flour extracted by milling for a given amount of grain. Kernels affected by disease or environmental stress usually become shriveled, leading to a higher bran/endosperm ratio and thus reducing FY. Higher FY from a unit of grain and unit of energy used for milling is crucial for milling industries to increase their revenue. Flour softness equivalence is defined as the ratio of fine flour to the total flour, which is determined by sifting through a 94-mesh screen after a single pass through a Quadrumat Junior mill (Finney et al., 1987). Higher values of SE are vital for most soft wheat manufactured goods, particularly cakes and other high-sugar baked products. Milling industries consider grain protein as an essential parameter for soft wheat because the higher protein content can result in larger granules in the flour, whereas low protein makes it mealy and produces more fine particles (Finney et al., 1987). Lower values of FPs are preferred in SRWW, especially for making cookies.

The alveograph and baking tests have been traditionally used to determine the quality of flour and baked products, respectively (Kweon et al., 2011). However, these methods are unable to parse out the contribution of major flour functional components such as arabinoxylans, gluten strength, and damaged starch (Kweon et al., 2011). Solvent retention capacity (SRC) tests, described under Method 56-11 in the American Association of Cereal Chemists (AACC), offer alternative ways to analyze the contribution of these functional components of wheat flour. These SRC methods are commonly used in commercial baking performance tests (Slade and Levine, 1994; Gaines, 2000). A graphical illustration of SRC traits is presented in Supplemental Fig. S2. The SRC tests involve mixing flour with four different solvents (i.e., 5% lactic acid, 5% sodium carbonate, 50% sucrose, and water). Then, the amount of solvent absorbed and held as a percentage of flour weight is expressed as a proxy to the retention strength for each

solvent. Lactic acid (LA), sodium carbonate (SO), sucrose (SU), and water (WA) retentions are indicative of gluten quality, levels of starch damage, amount of nonstarch polysaccharides, and water absorption of the flour, respectively (Guttieri et al., 2004). Most soft wheat products require lower values of SO, SU, and WA. The desired level for LA depends on the type of end-use product (Behl et al., 2017). Pastry products such as cookies and cakes require weaker gluten strength and thus lower LA values, whereas crackers require greater gluten strength and therefore higher values of LA.

The genetics of quality traits of SRWW have been reported by only a few studies. In one notable study Breseghello and Sorrells (2006) applied association analysis in a population of 149 soft winter wheat cultivars for milling quality using 88 simple sequence repeat (SSR) markers. They reported significant associations for milling score, endosperm separation index, and friability on chromosome 5B. The first mapping study for SRC was performed by Smith et al. (2011) using 171 recombinant inbred lines derived from the cross ‘Foster’ × Pioneer brand ‘25R26’. This study reported major quantitative trait loci (QTLs) on chromosomes 1B and 2B, and minor ones on chromosomes 2D, 3B, and 7D. Cabrera et al. (2015) used both biparental and association mapping approaches and, consistent with Smith et al. (2011), identified QTLs controlling all the seven quality traits on chromosomes 1B and 2B. Using 33,169 diversity arrays technology (DArT) markers in a genome-wide association study (GWAS), Hoffstetter et al. (2016) mapped FY and SE on chromosome 2B, FY on chromosomes 2A and 2D, and SE on 5A.

In the current study, we identified genomic loci that are determinants of end-use quality characteristics in an elite population of SRWW using GWAS. To do this, we used field-grown grain samples at multiple locations and dense DNA marker datasets to identify loci controlling milling quality (FY, FP, and SE) and baking quality (LA, SO, SU, and WA) traits.

## MATERIALS AND METHODS

We used a panel of SRWW with 270 elite breeding lines in this study. The panel was composed of lines from nine breeding programs serving nine states: Indiana, Michigan, Ohio, Missouri, Virginia, Kentucky, Maryland, Illinois, and New York. The wheat lines were grown at Wooster, OH, and Warsaw, VA, in 2 yr (2012 and 2013). We define year–location combinations as environments, which resulted in four environments denoted as OH2012, OH2013, VA2012, and VA2013. The same germplasm was used by Huang et al. (2016) to study genomic selection for trait stability. A detailed description of experimental design and field experiments was given by Huang et al. (2016). Briefly, an augmented design with five incomplete blocks was used for each environment. Each incomplete block consisted of 64 plots (i.e., 56 test entries, and eight plots for check variety ‘Branson’).

Quality traits were evaluated in the USDA-ARS Soft Wheat Quality Research Unit located in Wooster. Fifty grams of grain from each plot was milled on a modified Brabender Quadrumat Junior mill. Flour yield was measured by passing grain through the Quadrumat break roll unit. For measuring SE, the total flour obtained from the Quadrumat Junior mill was sifted through a 94-mesh screen. Flour protein was measured using the SpectraStar near-infrared analyzer (Unity Scientific), calibrated yearly for protein by nitrogen combustion analysis using the Rapid NIII nitrogen analyzer (Elementar). Solvent retention capacity analysis using different solvents was assayed as per AACC Method 56-11.02.

Phenotypic data were first adjusted for block effects within each environment. Then, we fitted a linear model:

$$Y_{ij} = \mu + G_i + E_j + \varepsilon_{ij}$$

where  $Y_{ij}$  is phenotypic value of the  $i$ th genotype in the  $j$ th environment,  $\mu$  is grand mean,  $G_i$  is the effect of  $i$ th genotype,  $E_j$  is the effect of  $j$ th environment, and  $\varepsilon_{ij}$  is the residual term. Both environment and residual terms were treated as random effects while line was treated as fixed effects. The least square means from this analysis were used as phenotype data for GWAS analysis. For estimation of heritability, environment, line, and residual terms were treated as random effects to obtain the variance components required for calculating the broad sense heritability as  $H^2 = \sigma_g^2 / (\sigma_g^2 + \sigma_e^2 / e)$ , where  $\sigma_g^2$  is the genetic variance,  $\sigma_e^2$  is the error variance, and  $e$  is the number of environments.

The marker dataset is a combination of genotyping-by-sequencing (GBS) and the 90K single nucleotide polymorphism (SNP) array. For GBS, DNA was extracted using a Mag-Bind Plant DNA Plus kit (Omega Bio-tek). Reduced-representation genomic libraries were created using *PstI-MspI* digestion of the genomic DNA. The samples were pooled for 96-plex and sequenced using Illumina Hi-Seq 2500. The subsequent SNP calling was completed using the TASSEL 5 GBSv2 pipeline, with a  $k$ -mer length of 64 bases and a minimum  $k$ -mer count of five (Bradbury et al., 2007). Reads were aligned to the IWGSC\_WGA v0.4 wheat reference genome sequence (<https://wheat-urgi.versailles.inra.fr/Seq-Repository/Assemblies>) using the aln method of the Burrows-Wheeler aligner (version 0.7.10) (Li and Durbin, 2009). Additionally, the 90K Illumina Infinium iSelect SNP array (Wang et al., 2014) was used following the manufacturer's protocol. For SNP calling, the GenomeStudio software package (Illumina, 2017) was used. For both methods, markers with  $\geq 10\%$  missing data or  $< 5\%$  minor allele frequency were removed. Missing markers were imputed using the linkage disequilibrium (LD)  $K$ -number neighbor imputation algorithm of Tassel 5.0 (Bradbury et al., 2007). In total, 27,499 SNP markers were developed, of which 13,198 were derived from the 90K iSelect SNP array and 14,301 were derived from the GBS method.

Population structure was analyzed by principal component analysis of marker data. The GWAS was performed by fitting the compressed mixed linear model (Zhang et al., 2010) using the rrBLUP package version 4.5 (Endelman, 2011) in the R environment, version 3.4.0 (R Core Team, 2013). We fitted a mixed model, where markers and population structure were considered fixed effects, and polygenic effect of lines (kinship inferred from

genotype data) and residuals were treated as random effects. The least square means of the traits considered in this study were estimated across the four environments to use in GWAS. In this study, we identified marker-trait association (MTA) with  $-\log(P) \geq 4$ . When IWGSC RefSec v1.0 became available, the positions of identified SNPs were converted from v0.4 to v1.0. Pairwise LD values were estimated using TASSEL 5.0 among the significant SNP markers identified in the GWAS. The group of markers that have an LD value  $\geq 0.80$  within same chromosome were considered to represent the same genomic region, as per the tagged SNP method presented by Carlson et al. (2004). Marker effects were calculated as the differences between means of homozygous groups. Ordinary least squares regression was fitted in the R environment, version 3.4.0 (R Core Team, 2013), using identified QTL regions to estimate the amount of phenotypic variation explained. Based on the annotated wheat genome IWGSC RefSec v1.0 (Appels et al., 2018), we screened for any gene within 250-kb physical distance of the SNP representing the identified MTAs.

## RESULTS

In this study, we evaluated SRWW quality traits from grains grown in four different environments and performed GWAS analysis. The phenotypes showed reasonably high heritability, as indicated in a prior genomic selection contribution by Huang et al. (2016, 2018) (Table 1). The lowest (0.76) heritability was for FP, whereas SO showed the highest (0.92) heritability. The distribution of all traits followed a bell shape, with FY and LA slightly skewed towards the left (Fig. 1). The means, SDs and ranges of the quality traits are summarized in Table 1. The average FY was 68.2% and ranged from 62.5% in 'Catocin' to 71.7% in 'Foster'. Flour protein averaged 8.4% with a minimum of 7.1% observed in 'IL06-7550' and a maximum of 10.6% observed in '9346A1-2-5-5-2-1'. Softness equivalent averaged 55.3% with a minimum of 41.8% for '9346A1-2-5-5-2-1' and a maximum of 63.5% for 'OH08-269-58'. The LA phenotype ranged from 70.1% in '05222A1-1-2-1' to 121% in 'PIO25R26'. Sucrose retention capacity, influenced by nonstarch polysaccharides, averaged 83.6% with a minimum of 76.1% in 'Crystal' and a maximum of

**Table 1. Descriptive statistics and heritability of the milling and baking quality traits. The numbers are based on the least squares means from the four environments.**

Trait†	$h^2$	Mean	SD	Min.	Max.
		%			
FY	0.89	68.2	1.6	62.5	71.7
FP	0.76	8.4	0.5	7.2	10.6
SE	0.91	55.3	3.0	41.8	63.5
WA	0.91	54.5	1.9	49.0	59.8
SO	0.92	65.8	2.5	60.5	75.7
SU	0.91	83.6	3.3	76.1	92.5
LA	0.87	95.3	9.8	70.1	121

† FY, flour yield; FP, flour protein; SE, softness equivalent; WA, water retention capacity; SO, sodium carbonate retention capacity; SU, sucrose retention capacity; LA, lactic acid retention capacity.

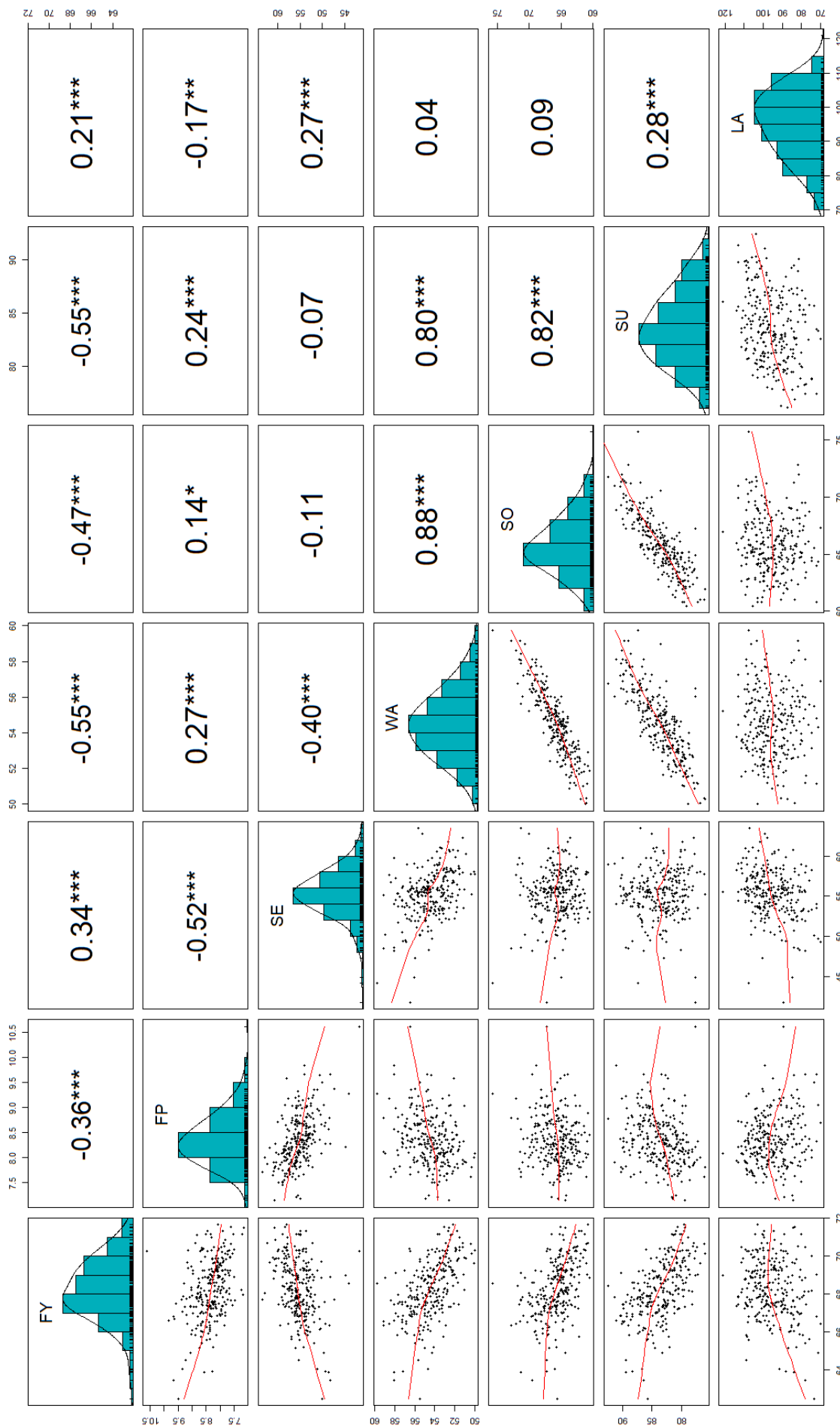


Fig. 1. Histograms (diagonals), pairwise scatterplots (lower triangle from the diagonal of histograms), and the pairwise Pearson's correlations (upper triangle from the diagonal of histograms). The line through the scatterplot is the least squares fit. The red asterisks show significance levels at 0.05 (\*), 0.01 (\*\*), and 0.001 (\*\*\*). FY, flour yield; FP, flour protein; SE, softness equivalent; WA, water retention capacity; SO, sodium carbonate retention capacity; SU, sucrose retention capacity; LA, lactic acid retention capacity.



92.5% in ‘MD03W485-10-2’. Sodium carbonate retention capacity, influenced by starch damage, revealed a range of 60.5% (in ‘IL07-20743’) to 75.7% (in ‘NY91017-8080’), with a mean of 65.8%. These data indicated that there is considerable variation in the germplasm for the quality traits. The Pearson’s correlation patterns revealed two groups of traits. Whereas FY, SE, and LA in one group exhibited positive correlation coefficients with each other, they showed negative correlations with FP, WA, SO, and SU in another group (Fig. 1). In particular, WA, SO, and SU exhibited very high correlations ( $r \geq 0.8$ ) with each other. On the other hand, these three traits were negatively correlated with FY (i.e.,  $r = -0.55$ ,  $-0.47$ , and  $-0.55$ , respectively).

Principal Components 1, 2, and 3 explained 11.1, 3.7, and 3.2% of the total variation, respectively (Supplemental Fig. S3). The GWAS identified 84 MTAs with  $-\log(P) \geq 4$  for all the seven traits. The number of MTAs per trait ranged from two for LA to 60 for WA. We used LD threshold for grouping and classifying the identified MTAs into seemingly independent QTL regions. This approach allowed us to group 84 MTAs into 18 independent QTL regions on 12 of the wheat chromosomes (Table 2). For chromosomes 1B and 7A, three independent genomic regions associated with quality traits were identified. Chromosomes 3A and 4B each had two genomic regions, and the remaining

eight chromosomes each possessed a single genomic region associated with the quality traits. In this manuscript, we regard each of the 18 independent loci as one QTL region. Interestingly, it follows that one QTL region may be responsible for more than one quality attribute, as expected by the correlation patterns of the phenotypes. Of the 18 QTL regions, four markers located on chromosomes 1B, 2A, and 7D were associated with multiple quality traits. Marker S1B\_83207071 (the marker naming implies the 83,207,071th nucleotide position alongside chromosome 1B, according to assembly v1.0 reported by Appels et al., 2018) was significantly associated with FY, WA, and SO; markers IWB27057 and S7D\_394187210 were significantly associated with SO and SU; and marker S2A\_610999949 was associated with FY and SU. In all cases, the directionality of markers effects fits the direction of phenotypic correlation between the traits.

For FY, three MTAs were identified on chromosomes 1B, 2A, and 2B (Fig. 2). The FY-increasing alleles at these three loci were present in >70% of the population, and the effect size for these alleles was at least 0.9%. The three QTL together explained 19.2% of the FY variation. Marker-trait associations for FP were detected on chromosomes 5A, 6A, and 7A (Fig. 2). Based on the LD threshold, chromosome 7A showed three distinct regions around 6.1, 127.6, and 695.6 Mbp that were

**Table 2. Genome-wide association study statistics including single nucleotide polymorphism (SNP) coordinate, trait associated and the  $-\log(P)$  value, polymorphism and genotypic frequency, effect size, and percentage variation explained ( $R^2$ ) by the quantitative trait locus are presented.**

Chromosome	SNP	Trait [ $-\log(P)$ value]	Allele	Effect size	$R^2$
				%	
1A	S1A_225663298†	WA‡(4.4)§	A/T¶ (239/31)#	2.1	4.2
1B	S1B_83207071	FY (4.3), WA (7.0), SO (5.5)	T/C (192/78)	1.0, 2.0, 2.6	4.7, 7.0, 5.9
	IWB27057 (652453984)	SO (4.4), SU (6.0)	G/C (199/40)	1.3, 2.0	4.7, 6.0
	S1B_642650411	SO (4.04)	A/G (160/110)	0.5	4.3
2A	S2A_610999949	FY (5.41), SU (5.46)	A/C (251/19)	1.4, 2.5	6.0, 6.0
2B	S2B_98357082	FY(5.03)	G/A (223/47)	0.9	5.5
3A	S3A_1901220	WA (4.52)	G/C (224/46)	0.8	4.3
	IWB29612 (25493701)	SU (4.10)	C/G (194/69)	1.0	4.0
4A	S4A_612531342	SE (4.08)	A/G (241/29)	2.0	5.1
4B	IWA4662 (22854992)	SE (4.05)	C/G (232/33)	2.2	5.0
	S4B_6957401	LA (6.96)	A/G (245/25)	8.6	4.3
5A	S5A_698528352	FP (5.34)	G/A (211/59)	0.3	6.6
5B	IWA6112(531541515)	LA (4.02)	G/C (183/83)	0.7	3.4
6A	S6A_1625351	FP (4.03)	C/T (247/23)	0.4	4.8
7A	S7A_6313988	FP (5.73)	C/T (205/65)	0.3	7.1
	S7A_127662921	FP (4.73)	G/T (180/90)	0.1	5.8
	S7A_695641406	FP (4.98)	G/T (213/57)	0.3	6.1
7D	S7D_394187210	SO (5.57), SU (5.08)	T/A (195/75)	1.8, 2.3	6.0, 5.1

† The numbers after the underscore or in parentheses are SNP coordinates (bp) based on the wheat IWGSC RefSeq v1.0.

‡ FY, flour yield; FP, flour protein; SE, softness equivalent; WA, water retention capacity; SO, sodium carbonate retention capacity; SU, sucrose retention capacity; LA, lactic acid retention capacity.

§ Numbers within parenthesis are  $-\log(P)$  values associated with the highest SNP peak in the region.

¶ Two letters are the polymorphism observed at each SNP and the underlined letter is the allele conferring favorable levels of traits. For LA, the allele that reduces LA is indicated as the favorable allele.

# Numbers in parentheses show frequencies of lines homozygous for the contrasting alleles.

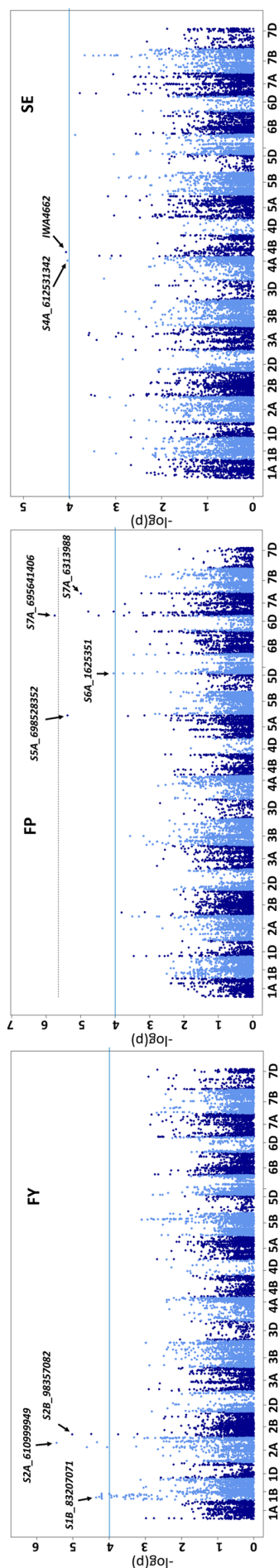


Fig. 2. Manhattan plots showing the marker-trait associations for the milling traits flour yield (FY), flour protein (FP), and softness equivalent (SE). The horizontal line in each plot corresponds to a  $-\log(P)$  value of 4. The marker with the highest significance value is shown by a black arrow, and the magnitude of their effects is shown in Table 2.

associated with FP (Table 2). The FP-increasing alleles at each of the SNPs S5A\_698528352, S6A\_1625351, S7A\_6313988, and S7A\_127662921 were present in majority ( $>0.67$ ) of the population. In contrast, the frequency of FP-increasing allele at SNP S7A\_695641406 in the population was lower (0.27). The five significant genomic regions together explained 27.3% of FP variation. For SE, MTAs were identified on chromosomes 4A and 4B (Fig. 2). The SE-increasing alleles at markers S4A\_612531342 and S4B\_22854992 revealed marker effects of 1.96 and 2.24%, respectively. The population seemed to be nearly fixed for these alleles, as indicated by allele frequencies of 0.89 and 0.88, respectively. The two QTL explained 9.8% of SE variation.

Association signals for WA were detected on chromosomes 1A, 1B, and 3A (Fig. 3, Table 2). These MTAs together explained 30.2% of variation in WA. The WA-decreasing alleles at markers S1A\_225663298 and S1B\_83207071 were present in  $>70\%$  of the population. In contrast, the WA-decreasing allele at marker S3A\_1901220 was present only in 17% of the population. For SO, the MTAs were detected on four genomic regions (three distinct regions of chromosomes 1B and one region on chromosome 7D) (Fig. 3). Of the four markers, the SO-decreasing alleles at markers S1B\_83207071 and S1B\_652453984 were present in majority of population (0.73 and 0.74, respectively). In contrast, the frequencies of SO-decreasing alleles at markers S1B\_642650411 and S7D\_394187210 were lower in frequency (0.41 and 0.28, respectively), decreasing the SO by 0.46 and 1.76%, respectively. The four genomic regions explained 34.3% of the total phenotypic variation observed in SO. Four genomic regions were identified for SU on chromosomes 1B, 2A, 3A, and 7D (Table 2, Fig. 3). The SU-lowering allele at markers S1B\_652453984, S2A\_610999949, and S3A\_25493701 showed higher frequencies (0.74, 0.93, and 0.74), and that of S7D\_394187210 was less frequent (0.28) in the population. Altogether, the four regions explained 16.8% of variation in SU. The MTAs for LA were identified on chromosome 4B and 5B (Table 2, Fig. 3). The MTAs on chromosome 4B and 5B were identified in environments OWM12 and OWM13, respectively (Table 2). The frequency of LA-decreasing alleles at markers S4B\_6957401 and IWA6112 were low (i.e., 0.09 and 0.31, respectively). The two loci explained 7% of the LA variation. As the last step, we searched for genes near the polymorphic sites and compiled them in Supplemental Table S1. In total, 96 genes were retrieved by searching the genomic regions surrounding 84 SNPs. Twenty genes were either uncharacterized or hypothetical proteins. The annotated genes were involved in several biological processes and pathways, including the auxin synthesis-response pathway, E3 ubiquitination, embryogenesis, and cell wall components.

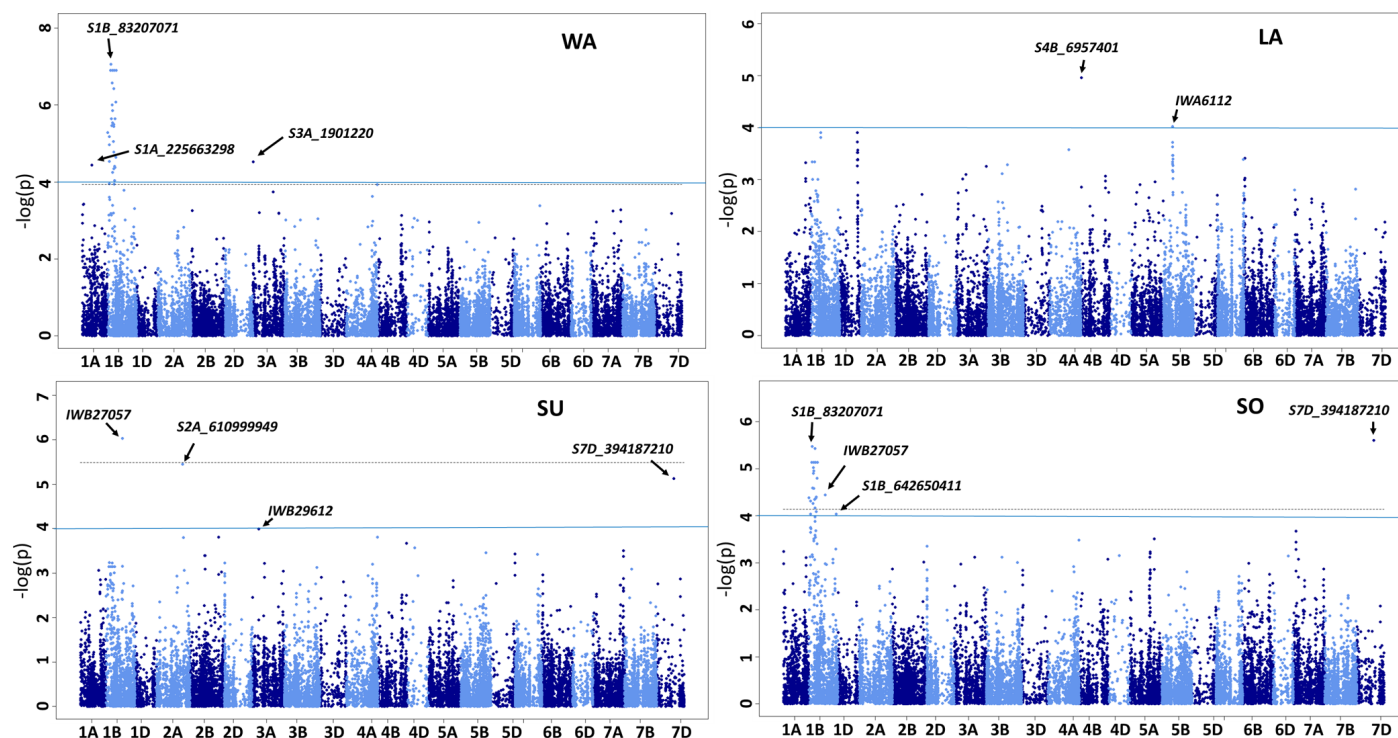


Fig. 3. Manhattan plots showing the marker-trait associations for baking traits (i.e., water solvent retention capacity [WA], sodium carbonate solvent retention capacity [SO], sucrose solvent retention capacity [SU], and lactic acid solvent retention capacity [LA]). The horizontal line in each plot corresponds to a  $-\log(P)$  value of 4. The dashed horizontal lines are automatically drawn by rrBLUP if a marker-trait association passes the threshold of 5% false discovery rate. The marker with the highest significance value is shown by a black arrow, and the magnitude of their effects is shown in Table 2.

## DISCUSSION

Wheat is a staple food to one-fifth of the world population, and almost all wheat products require flour with a specific combination of quality attributes. These attributes in soft wheat flour determine the end use of the flour. Increasing the understanding of the genetics of soft wheat quality traits will assist wheat breeders in developing cultivars that meet the demands of milling and baking industries, while simultaneously maintaining the agronomic needs of growers. Despite numerous studies involving genetic evaluation of end-use quality traits (Reif et al., 2011; Smith et al., 2011; Cabrera et al., 2015; Jiang et al., 2017), the sample size and number of genotypic markers in these studies have been generally low. Hoffstetter et al. (2016) used 33,169 DArT markers for association mapping in a relatively large collection ( $n = 470$ ) of elite lines for economically important traits that included FY and SE. Although Hoffstetter et al. (2016) had higher statistical power than other studies, it did not include baking quality traits (LA, SO, SU, and WA). This study focused on identifying genomic regions that are pertinent to the soft wheat milling and baking industry. The moderate to high heritability of these traits (Smith et al., 2011; Cabrera et al., 2015) indicates that proper selection procedure in breeding should produce genetic gains and improve profitability of the US domestic and international SRWW market.

To tailor an effective and cost-saving selection strategy, it is important to decipher the correlation among the traits under selection. Consistent with other SRWW studies (Smith et al., 2011; Cabrera et al., 2015; Jiang et al., 2017), we observed significant positive correlations among WA, SO, and SU, which are associated with dough water absorption. As SO and SU represent damaged starch and pentosans, respectively, they have the ability to absorb large amounts of water in the dough. These traits are negatively correlated with FY. The desired trait range is high FY and low water absorption (e.g., low SO, WA, and SU) in SRWW. Flour yield is directly proportional to the ease of the breakdown and separation of the endosperm. Any factor that hinders the break down and separation of endosperm would result in greater damage to the starch granules and thus greater water absorption (Smith et al., 2011). These correlations are beneficial for a selection program. For example, selecting for high FY should result in low WA, SO, and SU, and thus these traits would not need to be assessed for a secondary correlated trait. In contrast, correlation of LA with WA ( $r = 0.04$ ) and correlation of LA with SO ( $r = 0.09$ ) were low and not significant, also reinforced by Jiang et al. (2017), indicating a need to deliberately consider changes of LA when imposing selection on WA and SO. Similar to the previous study of Smith et al. (2011), a moderate and negative correlation ( $r = -0.52$ ) between FP and SE



was observed. Concurrent with this phenotypic correlation, we observed that four markers (S1B\_83207071, S1B\_652453983, S2A\_610999949, and S7D\_394187210) were associated with multiple traits, where the effect directionality of markers fits the direction of phenotypic correlation between the traits.

We have identified genomic regions that are associated with baking and milling traits in SRWW elite germplasm, with major QTL identified on the chromosomes previously reported by other research groups. Genomic regions on chromosomes 1B, 2A, and 2B were identified consistently in independent studies. Breseghello et al. (2005) identified a region on chromosome 2B for Quadrumat FY. Similarly, Smith et al. (2011) identified QTLs in two regions of chromosome 1B, one on 2A, and one on 2B. In an association genetic study involving 187 soft wheat lines, Cabrera et al. (2015) identified QTL on chromosomes 1B, 2A, and 2B using five different biparental populations. Similarly, Hoffstetter et al. (2016) mapped FY and SE on chromosome 2B using DArT markers. The primer sequences of the SSR and DArT markers shown to be associated with the QTLs in previous studies were blasted against the wheat genome assembly IWGSC v1.0 to find and compare the physical location of these markers with the regions identified in our study. If the physical locations of MTAs in our study fall between the physical locations of flanking markers or in proximity of DArT markers, it is highly likely that the MTAs we identified are the same regions already reported by previous researchers. Six QTLs (three on 1B, two on 2B, and one on 2A) reported previously by Cabrera et al. (2015) and two markers (D16 and D17) shown to be associated with FY on chromosome 2B (Hoffstetter et al., 2016) were tested. From these regions, only one QTL was identified in our study on chromosome 1B. This QTL was previously identified in two biparental populations (i.e., Foster  $\times$  Kanqueen and 91193D1  $\times$  92201D5) and flanked by *Xwmc44* and *Xbarc188* markers. These two SSR markers mark a region from 638,105,721 to 662,195,247 bp on chromosome 1B. Our MTAs were marked by SNPs at 652,453,984 (marker name is IWB27057) and 642,650,411 bp.

In our study, associated SNPs for FP were identified on chromosomes 6A and 7A. Loci controlling grain and flour protein on chromosomes 6A and 7A have not been previously reported in SRWW but were reported in bread wheat. Nelson et al. (2006) found alleles on chromosome 7A with effects on protein in recombinant inbred lines derived from a synthetic-hexaploid (W7985)  $\times$  bread wheat (Opata 85) cross. Prasad et al. (2003) identified a QTL for grain protein content on chromosome 7A in bread wheat. Sun et al. (2008) identified QTL for grain protein content in chromosome 6A in a population of recombinant inbred lines developed from Chuan 35050  $\times$  Shannong 483 winter wheat cross. Cabrera et al. (2015)

identified MTAs for SE in two regions of chromosome 4A in the diversity panel of soft winter wheat.

In the current study, the proportion of phenotypic variation explained by the significant markers (referred to as  $R^2$ ) ranged from 3.4 to 7.1% despite the fact that all the traits have high heritability. Similar  $R^2$  values were reported in previous studies of wheat quality traits. In the association analysis for flour and dough quality traits using the global wheat core collection, Bordes et al. (2011) reported  $R^2$  values of significant markers ranging from 3.6 to 5.4%. Similarly, Hoffstetter et al. (2016) reported  $R^2$  values ranging from 1.8 to 3.5% for FY and SE in the association studies using elite SRWW lines.

We found that 13 of the 18 QTLs had high allele frequency for favorable alleles in the population, which is expected with the current breeding status of the germplasm, as the germplasm represent the core of elite breeding materials. Although these favorable alleles can still provide an opportunity for population enhancement, the remaining five QTLs can be targeted for further improvement of quality traits in the population.

Markers with larger effects can be targeted for marker-assisted selection. For example, markers associated with S4B\_6957401 are associated with large marker effects and low frequency of the favorable allele. Marker-assisted selection for this locus could potentially lead to a reduction of 7% for LA in this population. Similarly, S1B\_83207071, S2A\_610999949, S6A\_1625351, IWA4662, and S7D\_394187210 have relatively large effects and can be used in marker-assisted selection schemes. Polymerase chain reaction (PCR)-based assays such as Kompetitive Allele Specific PCR can be designed for the significant polymorphisms identified in this study using flanking sequences from the wheat reference genome. The flanking sequences of the markers are provided in Supplemental Table S2. Since the  $R^2$  values for the identified QTL are not high, the combined use of marker-assisted selection and genomic selection might be more effective (Hoffstetter et al., 2016). Huang et al. (2016) also indicated that genomic selection in this population could be effective for these traits.

Candidate gene identification was another activity in this study that became possible by the availability of the wheat genome assembly. The gene search was a secondary objective, and for the reasons stated below, explanations of gene functions in this study are more speculative because we did not functionally characterize these genes. The transition from identification of SNPs and relating the polymorphic sites to candidate genes in self-pollinating plants that are under severe selection is a difficult task because of LD across extended stretches of DNA. As a result, each polymorphic site could potentially be associated with many genes, making it very difficult to pinpoint the candidate genes. Therefore, a follow-up



study of these genes is necessary to confirm their involvement in governing the quality traits. For example, gene TraesCS1B01G129500 that is located 7.5 kb downstream of marker S1B\_159919870 (associated with FY, WA, and SO) is homologous to fasciclin-like arabinogalactan. Downregulation of the fasciclin-like arabinogalactan gene was reported to be associated with increases in FY in both hard and soft wheat (Nirmal et al., 2017). Similarly, the TraesCS7A01G013600 gene, located 8 kb downstream of marker S7A\_6218449, encodes a wall-associated receptor kinase 2-like gene. Wall-associated kinase proteins bind to the pectin in the cell wall and could play role in cell expansion (Wagner and Kohorn, 2001).

Despite the importance of quality traits in the soft wheat industry, few studies were conducted to understand the genetics of soft wheat quality traits, identify genomic regions, and deploy the knowledge learned for breeding and selection. In this paper, we studied the flour milling and baking quality of the elite soft winter wheat germplasm and performed a GWAS analysis. Altogether, we have identified 18 genomic regions associated with seven soft wheat quality traits. In this manuscript, we provide sequence information required to make breeder-friendly markers for improvement of soft wheat traits using marker-assisted selection schemes.

## Supplemental Material Available

Supplemental material is available online for this article.

## Acknowledgments

This project was supported by the USDA Agriculture and Food Research Initiative Competitive Grants 2011-68002-30029 (Triticeae Coordinated Agricultural Project [TCAP]) and 2017-67007-25939 (Wheat Coordinated Agricultural Project [WheatCAP]). We are grateful for financial support provided by Virginia Tech, The Ohio State University, and the USDA-ARS Soft Wheat Quality Laboratory. The authors would like to thank the International Wheat Genome Sequencing Consortium (IWGSC, [www.wheatgenome.org](http://www.wheatgenome.org)) for prepublication access to IWGSC RefSeq v1.0 and annotation v1.0. The corresponding author is also thankful for the financial support from Hatch Grant no. IND010810.

## References

- Appels, R., K. Eversole, C. Feuillet, N. Stein, K. Feuillet, B. Keller, et al. 2018. Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361:eaar7191. doi:10.1126/science.aar7191
- Behl, H.D., C.A. Griffey, E. Rucker, W.E. Thomason, and B.K. Chim. 2017. Small grains in 2017. CSES-198NP. Virginia Coop. Ext. <https://pubs.ext.vt.edu/CSES/CSES-198/CSES-198.html> (accessed 15 Nov. 2017).
- Bordes, J., C. Ravel, L. Le Gouis, A. Lapierre, G. Charmet, and F. Balfourier. 2011. Use of a global wheat core collection for association analysis of flour and dough quality traits. *J. Cereal Sci.* 54:137–147. doi:10.1016/j.jcs.2011.03.004
- Bradbury, P.J., Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss, and E.S. Buckler. 2007. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* 23:2633–2635. doi:10.1093/bioinformatics/btm308
- Bresegghello, F., P.L. Finney, C. Gaines, L. Andrews, J. Tanaka, G. Penner, and M.E. Sorrells. 2005. Genetic loci related to kernel quality differences between a soft and a hard wheat cultivar. *Crop Sci.* 45:1685–1695. doi:10.2135/cropsci2004.0310
- Bresegghello, F., and M.E. Sorrells. 2006. Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. *Genetics* 172:1165–1177. doi:10.1534/genetics.105.044586
- Cabrera, A., M. Guttieri, N. Smith, E. Souza, A. Sturbaum, D. Hua, et al. 2015. Identification of milling and baking quality QTL in multiple soft wheat mapping populations. *Theor. Appl. Genet.* 128:2227–2242. doi:10.1007/s00122-015-2580-3
- Carlson, C.S., M.A. Eberle, M.J. Rieder, Q. Yi, L. Kruglyak, and D.A. Nickerson. 2004. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am. J. Hum. Genet.* 74:106–120. doi:10.1086/381000
- Endelman, J.B. 2011. Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome* 4:250–255. doi:10.3835/plantgenome2011.08.0024
- Finney, K.F., W.T. Yamazaki, V.L. Youngs, and G.L. Rubenthaler. 1987. Quality of hard, soft, and durum wheats. In: E.G. Heyne, editor, *Wheat and wheat improvement*. 2nd ed. Agron. Monogr. 13. ASA, CSSA, and SSSA, Madison, WI. p. 677–748. doi:10.2134/agronmonogr13.2ed.c35
- Gaines, C.S. 2000. Collaborative study of methods for solvent retention capacity profiles (AACC Method 56-11). *Cereal Foods World* 45:303–306.
- Guttieri, M.J., C. Becker, and E.J. Souza. 2004. Application of wheat meal solvent retention capacity tests within soft wheat breeding populations. *Cereal Chem.* 81:261–266. doi:10.1094/CCHEM.2004.81.2.261
- Gwirtz, J.A., M.R. Willyard, and K.L. McFall. 2006. Wheat flour products in North America. In: L. Popper, et al., editors, *A compendium of flour improvement*. Sosland Publ., Kansas City, MO. p. 295–308.
- Hoffstetter, A., A. Cabrera, and C. Sneller. 2016. Identifying quantitative trait loci for economic traits in an elite soft red winter wheat population. *Crop Sci.* 56:547–558. doi:10.2135/cropsci2015.06.0332
- Huang, M., A. Cabrera, A. Hoffstetter, C. Griffey, D. Van Sanford, J. Costa, et al. 2016. Genomic selection for wheat traits and trait stability. *Theor. Appl. Genet.* 129:1697–1710. doi:10.1007/s00122-016-2733-z
- Huang, M., B. Ward, C. Griffey, D. Van Sanford, A. McKendry, G. Brown-Guedira, et al. 2018. The accuracy of genomic prediction between environments and populations for soft wheat traits. *Crop Sci.* 58:2274–2288. doi:10.2135/cropsci2017.10.0638
- Illumina. 2017. Introduction to GenomeStudio software. Illumina, San Diego, CA.
- Jiang, P., P.P. Zhang, X. Zhang, and H.X. Ma. 2017. Genetic diversity and association analysis for solvent retention capacity in the accessions derived from soft wheat Ningmai 9. *Int. J. Genomics* 2017:2413150. doi:10.1155/2017/2413150
- Kweon, M., L. Slade, and H. Levine. 2011. Solvent retention capacity (SRC) testing of wheat flour: Principles and value in predicting flour functionality in different wheat-based food processes and in wheat breeding—A review. *Cereal Chem.* 88:537–552. doi:10.1094/CCHEM-07-11-0092

- Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25:1754–1760. doi:10.1093/bioinformatics/btp324
- Nelson, J.C., C. Andreescu, F. Breseghello, P.L. Finney, D.G. Gualberto, C.J. Bergman, et al. 2006. Quantitative trait locus analysis of wheat quality traits. *Euphytica* 149:145–159. doi:10.1007/s10681-005-9062-7
- Nirmal, R.C., A. Furtado, P. Rangan, and R.J. Henry. 2017. Fasciclin-like arabinogalactan protein gene expression is associated with yield of flour in the milling of wheat. *Sci. Rep.* 7:12539. doi:10.1038/s41598-017-12845-y
- Prasad, M., N. Kumar, P. Kulwal, M. Röder, H. Balyan, H. Dhaliwal, and P. Gupta. 2003. QTL analysis for grain protein content using SSR markers and validation studies using NILs in bread wheat. *Theor. Appl. Genet.* 106:659–667. doi:10.1007/s00122-002-1114-y
- R Core Team. 2013. R: A language and environment for statistical computing. R Found. Stat. Comput., Vienna.
- Reif, J.C., M. Gowda, H.P. Maurer, C.F.H. Longin, V. Korzun, E. Ebmeyer, et al. 2011. Association mapping for quality traits in soft winter wheat. *Theor. Appl. Genet.* 122:961–970. doi:10.1007/s00122-010-1502-7
- Slade, L., and H. Levine. 1994. Structure–function relationships of cookie and cracker ingredients. In: H. Faridi, editor, *The science of cookie and cracker production*. Chapman & Hall, New York. p. 23–141.
- Smith, N., M. Guttieri, E. Souza, J. Shoots, M. Sorrells, and C. Sneller. 2011. Identification and validation of QTL for grain quality traits in a cross of soft wheat cultivars Pioneer Brand 25R26 and Foster. *Crop Sci.* 51:1424–1436. doi:10.2135/cropsci2010.04.0193
- Sun, H., J. Lü, Y. Fan, Y. Zhao, F. Kong, R. Li, H. Wang, and S. Li. 2008. Quantitative trait loci (QTLs) for quality traits related to protein and starch in wheat. *Prog. Nat. Sci.* 18:825–831. doi:10.1016/j.pnsc.2007.12.013
- USDA–ERS. 2017. Wheat data. USDA Econ. Res. Serv. <https://www.ers.usda.gov/data-products/wheat-data/> (accessed 15 Nov. 2017).
- USDA–NASS. 2017. Quick stats. USDA Natl. Agric. Stat. Serv. [https://quickstats.nass.usda.gov/results/9140FA3B-C859-3EA7-89BB-6E7FEC709F7E?pivot=short\\_desc](https://quickstats.nass.usda.gov/results/9140FA3B-C859-3EA7-89BB-6E7FEC709F7E?pivot=short_desc) (accessed 10 Nov. 2017).
- Wagner, T.A., and B.D. Kohorn. 2001. Wall-associated kinases are expressed throughout plant development and are required for cell expansion. *Plant Cell* 13:303–318. doi:10.1105/tpc.13.2.303
- Wang, S., D. Wong, K. Forrest, A. Allen, S. Chao, B.E. Huang, et al. 2014. Characterization of polyploid wheat genomic diversity using a high-density 90,000 single nucleotide polymorphism array. *Plant Biotechnol. J.* 12:787–796. doi:10.1111/pbi.12183
- Zhang, D., E. Ersoz, C.Q. Lai, R.J. Todhunter, H.K. Tiwari, M.A. Gore, et al. 2010. Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.* 42:355–360. doi:10.1038/ng.546