10-5-2017

Uncovering the Genetic Architecture of Seed Weight and Size in Intermediate Wheatgrass through Linkage and Association Mapping

Xiaofei Zhang
North Carolina State University

Steven R. Larson
Utah State University, steve.larson@ars.usda.gov

Liangliang Gao
Kansas State University

Soon Li Teh
University of Minnesota

Lee R. DeHaan
The Land Institute

Max Fraser
University of Minnesota

See next page for additional authors

Zhang, X., S. R. Larson, L. Gao, S. L. Teh, L. R. DeHaan, M. Fraser, A. Sallam, T. Kantarski, K. Frels, J. Poland, D. Wyse, and J. A. Anderson. 2017. Uncovering the Genetic Architecture of Seed Weight and Size in Intermediate Wheatgrass through Linkage and Association Mapping. Plant Genome 10. doi:10.3835/plantgenome2017.03.0022

Published online November 28, 2017

ORIGINAL RESEARCH

Uncovering the Genetic Architecture of Seed Weight and Size in Intermediate Wheatgrass through Linkage and Association Mapping

Xiaofei Zhang,* Steven R. Larson, Liangliang Gao, Soon Li Teh, Lee R. DeHaan, Max Fraser, Ahmad Sallam, Traci Kantarski, Katherine Frels, Jesse Poland, Donald Wyse, and James A. Anderson*

Abstract

Intermediate wheatgrass [IWG; *Thinopyrum intermedium* (Host) Barkworth & D.R. Dewey subsp. *intermedium*] is being developed as a new perennial grain crop that has a large allohexaploid genome similar to that of wheat (*Triticum aestivum* L.). Breeding for increased seed weight is one of the primary goals for improving grain yield of IWG. As a new crop, however, the genetic architecture of seed weight and size has not been characterized, and selective breeding of IWG may be more intricate than wheat because of its self-incompatible mating system and perennial growth habit. Here, seed weight, seed area size, seed width, and seed length were evaluated across multiple years, in a heterogeneous breeding population comprised of 1126 genets and two clonally replicated biparental populations comprised of 172 and 265 genets. Among 10,171 DNA markers discovered using genotyping-by-sequencing (GBS) in the breeding population, 4731 markers were present in a consensus genetic map previously constructed using seven full-sib populations. Thirty-three quantitative trait loci (QTL) associated with seed weight and size were identified using association mapping (AM), of which 23 were verified using linkage mapping in the biparental populations. About 37.6% of seed weight variation in the breeding population was explained by 15 QTL, 12 of which also contributed to either seed length or seed width. When performing either phenotypic selection or genomic selection for seed weight, we observed the frequency of favorable QTL alleles were increased to >46%. Thus, by combining AM and genomic selection, we can effectively select the favorable QTL alleles for seed weight and size in IWG breeding populations.

Core Ideas

- Twenty-three shared QTL were identified using linkage and association mapping
- Overlapped QTL explained the high genetic correlation among seed weight and size
- QTL responded positively to either phenotypic selection or genomic selection
- Combining association mapping and genomic selection would increase genetic gain

INTERMEDIATE wheatgrass ($2n = 6x = 42$) is a new perennial grain crop (Wagoner, 1990; Kantar et al., 2016). Compared with annual grain crops, it has an extended growing season and deep roots, which increase carbon sequestration and help prevent runoff and improve water quality (Glover et al., 2010; Culman et al., 2013). Moreover,

X. Zhang, Dep. of Horticultural Science, North Carolina State Univ., Raleigh; S.R. Larson, USDA-ARS, Forage and Range Research, Utah State Univ., Logan; L. Gao and J. Poland, Dep. of Plant Pathology, Kansas State Univ., Manhattan; S.L. Teh, Dep. of Horticultural Science, A. Sallam, Dep. of Plant Pathology, and M. Fraser, K. Frels, D. Wyse, and J.A. Anderson, Dep. of Agronomy and Plant Genetics, Univ. of Minnesota, St. Paul; L.R. DeHaan, The Land Institute, Salina, KS; T. Kantarski, Columbia University, New York, NY. Received 13 Mar. 2017. Accepted 10 July 2017. *Corresponding authors (ander319@umn.edu; zhan2740@umn.edu). The research was performed at the U of MN, Kansas State University, USDA-ARS, Forage and Range Research, UT, and The Land Institute, KS.

Abbreviations: AM, association mapping; BLUE, best linear unbiased estimation; GBS, genotyping-by-sequencing; IM, interval mapping; IWG, intermediate wheatgrass; LD, linkage disequilibrium; LG, linkage group; LOD, logarithm of odds; MAF, minor allele frequency; MLM, mixed linear model; MQM, multiple quantitative trait loci model; QTL, quantitative trait locus/loci; UNEAK, Universal Network-Enabled Analysis Kit.

Published in Plant Genome
Volume 10. doi: 10.3835/plantgenome2017.03.0022

the harvested grain can be used to produce a diversity of food products, such as cookies, pancakes, and bread (Wagoner, 1990; DeHaan et al., 2014; Zhang et al., 2015). In addition, IWG produces substantial biomass for animal feed or bioenergy production (Harmoney, 2015). Thus, IWG is a promising perennial grain and forage crop, providing both environmental services and economic benefits to farmers (Runck et al., 2014; Kantar et al., 2016).

The domestication of IWG was initiated in the 1980s at the Rodale Research Center, PA. Currently, The Land Institute, KS, the University of Minnesota, MN, and the University of Manitoba, Canada, are pursuing breeding efforts to improve grain yield, seed size, threshability, shattering resistance, lodging resistance, and develop cultivars for farmers (DeHaan et al., 2014; Cattani, 2016; Zhang et al., 2016). Since the inception of breeding efforts, only about 10 cycles of selection have been performed. In the breeding populations, there is substantial variation in the agronomic traits, such as height, biomass, grain yield, and seed weight. We observed that the difference between minimum and maximum values of each trait ranged from over threefold for height to over 14-fold for head weight in our breeding population (Zhang et al., 2016). Because of the large genetic variation in the breeding population, two cycles of selection resulted in a 77% increase in seed yield (DeHaan et al., 2014).

Seed weight and size are key traits in most crops. Large seeds contribute to large grain yield and are easier to harvest and process for farmers (Moles et al., 2005). Thus, seed weight was a primary target in the domestication of many crops and is selected in breeding programs, including wheat, rice (*Oryza sativa* L.), soybean [*Glycine max* (L.) Merr.], and sorghum [*Sorghum bicolor* (L.) Moench; Huang et al., 2013; Williams and Sorrells, 2014; Zhou et al., 2015]. Seed size, the major determinant of seed weight, is characterized by a combination of seed length, seed width, and seed thickness. These three parameters are positively correlated with seed weight (Tan et al., 2000). Seed weight and size have been widely accepted as a complex trait controlled by multiple genes (Tan et al., 2000; Huang et al., 2013), referred to as QTL. The identification of major QTL for seed weight and size is an important objective of crop genetic analysis and breeding programs. More than 400 QTL associated with seed traits have been identified in rice (Huang et al., 2013). In wheat, seed size QTL of varying effect were detected on all chromosomes (Williams and Sorrells, 2014). With little genetic research performed on IWG, no QTL for seed weight have been reported to date.

Compared with average seed weight of over 30 mg for bread wheat (Zanke et al., 2015), IWG has relatively small seeds. The seed weight, on average, was 4.1 mg in the initial breeding population at The Land Institute (DeHaan et al., 2014). After eight cycles of mass selection, the seed weight was doubled when grown in a spaced plant selection nursery. In the University of Minnesota breeding population, derived from the third cycle of selection from The Land Institute, the seed weight ranged from 3.26 mg to 13.25 mg, with 8.32 mg as the mean in Minnesota (Zhang et al., 2016). Conventional breeding using phenotypic selection-based recurrent selection is resulting in seed weight improvement, but it is a time and labor-consuming process.

Recent advances in sequencing technologies have been revolutionizing plant breeding and genetic research by dramatically reducing the cost of genome-wide marker discovery for any species (Davey et al., 2011). These technologies, such as GBS and restriction site associated DNA sequencing, have been successfully used to discover molecular markers for genetic map development, linkage mapping and AM, and genomic selection (e.g., Poland and Rife, 2012; Russell et al., 2014; Carlson et al., 2015; Gorjanc et al., 2015; Iquira et al., 2015). Using GBS, we developed the first consensus genetic map of IWG with 10,029 markers (Kantarski et al., 2016), and identified 3883 markers from an IWG breeding population for genomic selection (Zhang et al., 2016). A genomic selection-based breeding scheme was proposed to accelerate the domestication and improvement of IWG. Using this breeding scheme, less labor is required for planting, weeding, harvesting, threshing, and phenotyping (Zhang et al., 2016).

In contrast with grain yield, seed weight and size have high heritability (Zhang et al., 2016). Many major QTL associated with seed weight, length, and width have been fine-mapped or recently cloned; for example, *GW2*, *GS3*, and *GS5* in rice (Song et al., 2007; Mao et al., 2010; Li et al., 2011), and *ZmGS3* and *ZmGW2* in maize (*Zea mays* L.; Li et al., 2010; Mao et al., 2010). The identified QTL can be targeted for selection in a breeding population to increase selection efficiency (Lande and Thompson, 1990). Moreover, the cloning of causal genes increases our understanding of the genetic mechanisms underlying seed weight and size, which may in turn lead to improved selection methods.

Linkage mapping has been widely used to uncover genomic loci that control agronomic traits using a biparental population. The detected QTL, however, are limited to the assessment of the alleles that differ between the two parents of the population, and linkage mapping generally offers relatively low resolution, such that the markers could be 10–30 cM from the causal gene (Kearsey and Farquhar, 1998). An alternative method to map QTL is AM, also known as linkage disequilibrium (LD) mapping. The association between genotype and phenotype depends on the accumulation of historical recombination events in long-term breeding populations or natural populations (Zhu et al., 2008; Mackay et al., 2009). Hence, in AM, a larger number of markers are required to assure LD between markers and causative genes throughout the genome, thus improving the genetic resolution and enabling fine mapping. The QTL identified from the breeding population using AM can be directly used in selection for such QTL, and there is less chance of their linkage phase being disrupted by recombination. Association mapping, however, has less power to detect the effect of rare variants (Morrell et al., 2012).

Another major constraint of AM is population structure that can lead to detection of spurious associations (Myles et al., 2009). A mixed model analysis that fits population structure and genetic relationship has been widely used to statistically reduce the number of false-positive signals (Yu and Buckler, 2006; Yu et al., 2006).

To determine the genetic architecture of seed weight and size, both linkage mapping in two biparental populations and AM in a breeding population with 1126 genets were performed in the present study. The objectives of this study were (i) to understand the variation of seed weight, area size, length, and width in the IWG breeding population; (ii) to identify QTL for seed weight and size in the breeding population; (iii) to detect QTL for seed weight and size in biparental populations; (iv) to compare the QTL identified by AM with the QTL regions detected in biparental mapping populations; (v) to investigate the frequency change of favorable QTL alleles and genotypes under selection and propose breeding strategies to stack favorable QTL alleles in the breeding germplasm.

Materials and Methods

Plant Materials

An IWG breeding population containing 1126 IWG genets from 58 families was used as the AM panel to analyze the genetic architecture of seed weight and size. Herein, we use the term "genet" to describe a genetically unique individual for outcrossing species that can be a single plant or cloned to represent multiple plants (Zhang et al., 2016). This is a representative population of the first recurrent selection cycle at the University of Minnesota. These IWG genets were planted in a spaced-plant nursery in 2011. Although plants comprising the AM panel were not clonally replicated, repeated measurement of seed weight and size were taken in 2012 and 2013. The details about the composition, organization, and management of the population were previously described (Zhang et al., 2016). Two F_1 biparental populations, one derived from a cross between two genets, C3-2331 and C3-2595, and the other from a cross between M35 and M26, were used to map QTL for seed weight and size (Kantarski et al., 2016; Zhang et al., 2016). C3-2331 was also the parent of 34 genets in the breeding population. C3-2331 and C3-2595 were from the third recurrent selection cycle at The Land Institute. These two parents have a large difference in seed weight, with C3-2331 producing larger seeds. This biparental population has 178 genets, including the two parents. Two clonal replications were planted in St. Paul in 2013 in a completely randomized design. Seed weight and size of the seeds harvested in 2014 and 2015 field seasons were measured. The other F_1 biparental population, M35 × M26, was composed of 265 genets. Two replications of 234 genets were planted in Salina, KS, in 2012 and harvested and measured in 2013, 2014, and 2015. Three replications of 263 genets were transplanted in Providence, UT, in 2013, and three replications were harvested and measured in 2014 and two replications in 2015.

Seed Weight and Size Measurement

All the spikes on each plant were harvested by hand and threshed using a laboratory thresher (Wintersteiger LD 350, Ried, Australia). About 200 seeds from each plant were mechanically dehulled (Wintersteiger, Ried, Australia). Fifty naked seeds per plant were weighed to calculate seed weight. These fifty seeds were scanned using an HP Scanjet 4600 with a lab bench as the black background. The pictures were analyzed using SmartGrain v.1.2 following the standard manual (Tanabata et al., 2012). Seed length, width, and area size were used as parameters for seed size. For the AM panel and population C3-2331 × C3-2595, correction for variability between environments was done by calculating best linear unbiased estimation (BLUE) of each genet using the MIXED procedure in SAS (v.9.3.1; Sallam et al., 2015). The adjustment for environmental effects was described in detail in our previous study (Zhang et al., 2016). Broad-sense heritability h^2 on a genet mean was calculated using the equation $h^2 = \sigma_g^2/(\sigma_g^2 + \sigma_e^2/\eta)$, where σ_g^2 is the genetic variance, σ_e^2 is the error variance including genotype × environment and residuals, and η is the number of years. For the population M35 × M26, best linear unbiased prediction (BLUP) of each genet were calculated using the lmer function in R package lme4 (R Core Team, 2014).

GBS Libraries, Sequencing, and SNP Calling

The AM panel was genotyped as previously described (Zhang et al., 2016). To increase the sequencing coverage and the number of markers for AM, ~10 genets from each of the 58 families were sequenced one more time using Illumina HiSeq 2000 (San Diego, CA). The sequence data from the AM panel were used to call SNP with the Universal Network-Enabled Analysis Kit (UNEAK) pipeline (Lu et al., 2013). The two biparental populations were sequenced and genotyped along with five other mapping populations used for developing a consensus map as previously described (Kantarski et al., 2016). Three filters were used for the SNP markers: (i) less than 30% missing data; (ii) χ^2 test ($p > 0.05$) for heterozygotes, done to filter markers across all individuals, is based on the hypothesis that, in an allopolyploid (functional diploid) species, the sequencing counts of the two alternate tags of a SNP were equal in all heterozygotes; and (iii) the homozygote genotypes with a sequencing count of less than five tags were considered as missing data. This filtering will decrease of the rate of false homozygote genotypes (Zhang et al., 2016). Details about the genotyping of the two biparental populations were described in Zhang et al. (2016) and Kantarski et al. (2016).

Determining Marker Locations and Imputing Missing Data

The consensus genetic map, derived from seven mapping populations, was used to determine the locations of the SNP markers in the AM panel (Kantarski et al., 2016). The sequence data of the markers in the consensus map were converted into a local BLAST database, using the

BLAST command line tool makeblastdb. The marker sequences from the AM panel were aligned with the sequences of the markers in the consensus genetic map using the blastn command. Only identical markers in length and composition were considered as shared markers between the two populations and their map locations were used to analyze LD decay in the AM panel.

The missing data of GBS markers in the AM panel were imputed using LinkImpute, a software package based on a k-nearest neighbor genotype imputation method, LD-kNNi, which is designed for unordered markers (Money et al., 2015). LinkImpute provided fast and accurate genotype imputation for diverse and heterozygous accessions of apples (*Malus domestica* Borkh.), grapes (*Vitis vinifera* L.), and maize. Markers whose minor allele frequency (MAF) were larger than 0.05 were used for subsequent analysis.

LD Decay and Structure of the IWG Population

LD among mapped markers in the AM panel was estimated using Haploview 4.2 (Barrett et al., 2005). The LD decay of r^2 was described using the Hill and Weir formula (Hill and Weir, 1988). We estimated the population structure using the model-based clustering algorithm of STRUCTURE (Pritchard et al., 2000). To avoid the overestimation of subpopulation divergence due to tightly linked SNP markers, we only used 467 GBS markers at spaced at approximately 5 cM intervals. Subgroups K = 1 to 10 were tested and each was modeled 10 times with a burn-in period and number of replications equal to 10,000 using an admixture model in STRUCTURE. The optimal k was then determined using DeltaK calculated using Structure Harvester (Earl and Vonholdt, 2012). The Q matrices (K = 2 and K = 3), containing the probability of membership of each cluster for each individual, were obtained using Software CLUMPP (Jakobsson and Rosenberg, 2007) and then used as covariates in the mixed linear models (MLM).

Genome-Wide Association Study for Seed Weight and Size

The generalized linear model (GLM) and MLM in TASSEL5.0 were used to test association between the phenotypic and genotypic data (Bradbury et al., 2007). Results from MLM were further verified using the R package, Genome Association and Prediction Integrated Tool (GAPIT; Lipka et al., 2012). The MLM for AM can be specified as follows:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Qv} + \mathbf{Z\mu} + \varepsilon \qquad [1]$$

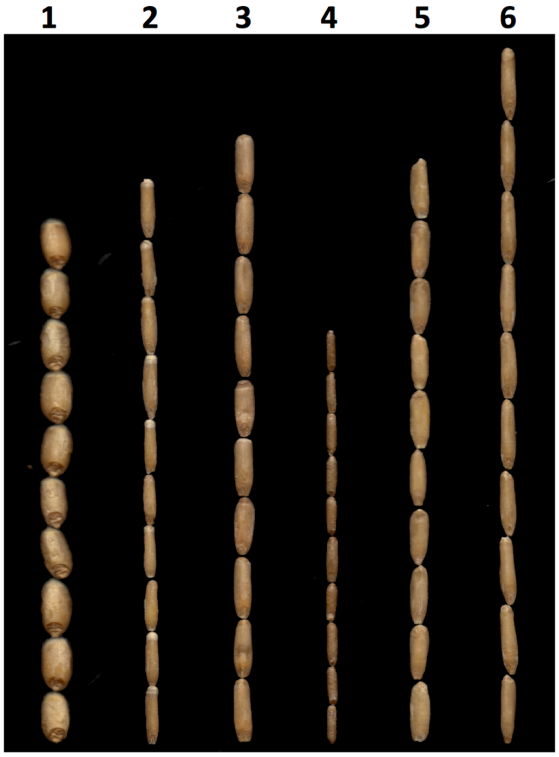
where $\mathbf{y}$ is a vector of phenotypes, $\boldsymbol{\beta}$ is a vector of fixed marker effects, $\mathbf{v}$ is a vector of the fixed effects of different subpopulations, $\boldsymbol{\mu}$ is a vector of random effects caused by relatedness, ε is a vector of residual effects, $\mathbf{X}$ is the marker matrix, $\mathbf{Q}$ is an incidence matrix containing membership proportions to cluster subpopulations identified by STRUCTURE analysis, and $\mathbf{Z}$ is the corresponding matrix that relate $\mathbf{y}$ to $\boldsymbol{\mu}$. Association mapping was conducted using filtered GBS markers with MAF $\geq$ 0.05. The population Kinship matrix was calculated using TASSEL v.5, based on the scaled identity by state method (Endelman and Jannink, 2012). A MLM (without compression) was implemented using both GAPIT and TASSEL. The *p*-values were extracted from the outputs of GAPIT and TASSEL. The false discovery rate adjusted *p*-values in GAPIT were too stringent, so relaxed *p*-value levels were used as recommended (Pasam et al., 2012; Zegeye et al., 2014). In the present study, a *p*-value of 0.0025 as the threshold of significant QTL was used. Only SNP markers showing significant *p*-values in two of the three analyses for BLUE, 2012 and 2013, were considered significant.

Multiple-linear regression was used to estimate the proportion of phenotypic variance explained by significant markers (R Core Team, 2014). The r^2 value from LD analysis or windows of 10 cM were used to define SNPs tagging a locus (Gao et al., 2016). Only the most significant SNP present within a 10 cM window was used to tag the QTL. The effects of most significant markers were extracted and considered as the effects of the corresponding QTL.

Linkage Mapping for Seed Weight and Size

For the C3-2331 × C3-2595 population, markers with <10% missing data were used to develop the genetic map (Kantarski et al., 2016; Zhang et al., 2016). This genetic map was used for QTL mapping using MapQTL6 (Van Ooijen, 2009). The integrated two-way pseudo-testcross approach was used to map QTL. The <*lmxll*> markers were used for the parent C3-2331, <*nnxnp*> markers were used for the other parent C3-2595, while <*hkxhk*> markers were not used. After splitting into a <*lmxll*> and a <*nnxnp*> dataset, we translated all markers to the population type of double haploid. Then we combined the two parental datasets into a single dataset and appended the map of the second parent to the map of the first parent. This approach allows the use of cofactors of one parent in the searching for QTL in the other parent, thus increasing the mapping power. The same procedure was used to map QTL in the M35 × M26 population.

Logarithm of odds (LOD) significance thresholds of linkage groups (LGs) were determined independently for each trait using the Permutation Test function with significance threshold ($\alpha = 0.05$) calculated using 1000 permutations in MapQTL6. Interval mapping (IM) was used to detect major QTL. The closest marker at each QTL from IM was used as the cofactor in the multiple QTL model (MQM). After initial MQM mapping, we adjusted the set of cofactors based on the updated most likely positions of QTL, and repeated the MQM mapping. Several rounds of such MQM mapping were performed to obtain the best possible final solution, where no other segregating QTL were detected. Finally, restricted MQM (rMQM) mapping was used to identify the locations and regions of QTL. The left and right markers of QTL were determined based on the confidence interval calculated by $LOD_{max} - 1$ at the estimated peak QTL position. The markers positioned in the consensus map were used to determine the locations of QTL.



	1	2	3	4	5	6
Seed weight (mg)	32.53	8.32	13.87	3.23	11.95	12.12
Seed length (mm)	6.01	6.23	6.34	4.56	6.33	7.31
Seed width (mm)	3.20	1.52	1.83	1.20	1.92	1.66

Fig. 1. Typical wheat and intermediate wheatgrass seeds. (1) common wheat, 'Chinese Spring'; (2–6) intermediate wheatgrass. (2) WG117210 with the median seed size in the breeding population; (3) WG112027 with the largest seed weight; (4) WG115812 with the smallest seed weight, length, and width; (5) WG115211 with the widest seeds; (6) WG110405 with the longest seeds. The samples were collected from 2013 growth season.

Response of QTL to Phenotypic Selection and Genomic Selection

In our previous study, independent validation was used to test the effectiveness of genomic selection (Zhang et al., 2016). Of 1126 genets in the breeding population (AM panel in the present study), 494 were used to train the genomic selection model for seed weight which was used to calculate the predicted values of the remaining 632 genets. Based on the predicted values (genomic selection) and the observed phenotypic values (phenotypic selection) of 632 genets, the top 2% genets were selected. The number of QTL and the frequency of favorable genotypes and alleles were calculated. The top 0.5% genets were also selected to analyze the effect of selection intensity on the frequency of favorable QTL alleles.

Results

Variation in Seed Weight and Size in the Intermediate Wheatgrass Association Mapping Panel

Four morphometric parameters, seed weight, seed area size, seed length, and seed width were measured in a breeding population with 1126 genets (AM panel). Best linear unbiased estimation (BLUE) was used to describe the variation of seed weight and size in the present study (Supplementary Table S1). The seed weight was 8.32 mg on average, varying from 3.26 to 13.25 mg in the breeding population. Different from wheat, IWG has long and thin seeds (Fig. 1). The seed length ranged from 4.52 to 7.47 mm, seed width from 1.16 to 1.96 mm, and seed area size from 4.35 to 10.58 mm^2 (Supplementary Table S1). All these traits had normal distributions in the breeding population. High broad-sense heritability was observed for all traits (0.85–0.91), indicating that genetic effects are the major determinant of the phenotypic variance on seed weight and size in IWG.

Discovery of Genetically-Mapped Markers for Association Mapping

From 17 lanes of sequencing, we obtained 2.5 billion reads for the AM panel. A total of 10,171 markers with <30% missing data from the population were identified through the UNEAK pipeline. Among them, 4731 markers matched the mapped markers in the consensus genetic map of IWG with 100% identity (Fig. 2). These markers were located in 21 LGs, and the number of markers on each LG ranged from 101 for LG5 to 324 for LG20 (Supplementary Fig. S1). In total, 4731 markers covered the genetic map with 2655 cM (Kosambi). There

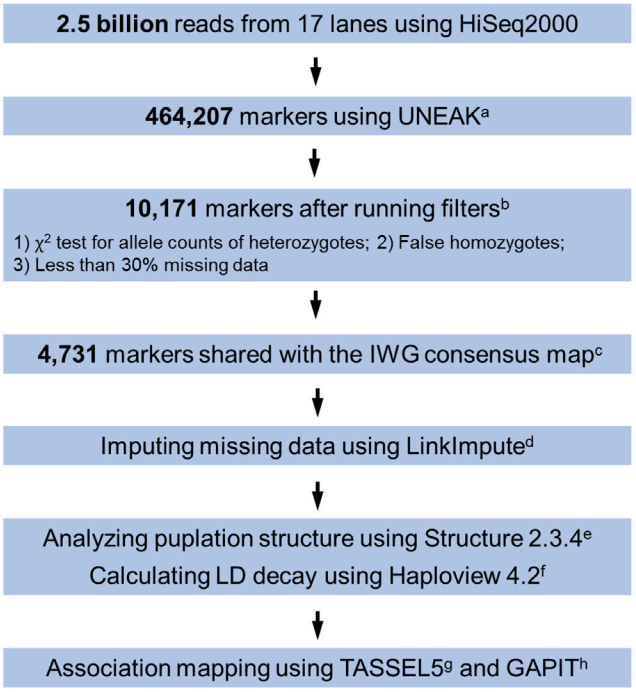


Fig. 2. Flow diagram of genotypic data analysis and association mapping pipelines. χ^2 test was performed based on the hypothesis that, in diploid species, the sequencing counts of the two paired tags of a SNP were equal in all heterozygotes. The markers with $p > 0.05$ of χ^2 test were kept for subsequent analysis. All the homozygotes whose allele sequencing counts were <5 were considered as missing data. [a]Lu et al., 2013; [b]Zhang et al., 2016; [c]Kantarski et al., 2016; [d]Money et al., 2015; [e]Pritchard et al., 2000; [f]Barrett et al., 2005; [g]Bradbury et al., 2007; [h]Lipka et al. (2012).

are three markers for every 2 cM, on average. The LD decayed (defined by conventional r^2 declining to below 0.16) at an average distance of 1 cM ($r^2 = 0.1$ at 2 cM; Supplementary Fig. S2). With an average marker density of three markers for every 2 cM, 4731 markers should be adequate to assure the LD between markers and QTL.

Association Mapping for Seed Weight and Size

We analyzed population structure using STRUCTURE. The DeltaK (K, the number of clusters) value was 80 for K = 2, 53 for K = 3, and <3 for K > 3 (Supplementary Fig. S3). The optimal number of K might be 1, 2, or 3 based on the DeltaK method. Software CLUMPP was used to obtain the Q matrices for K = 2 and K = 3. The Q matrices were used as covariates in AM using a MLM (i.e., QK model). Whether we used the Q matrix with K = 3 (3 subpopulations), K = 2 (2 subpopulations), or without Q matrix, the fit of the MLM model to the data was not improved for seed weight, seed width, seed length and area size, and the detected QTL have the same *p*-values shown in Manhattan plots (correlation coefficients, $r >$ 0.99 for seed weight, BLUE). Thus, the MLM model without a Q matrix was used to fit the AM panel and all four traits for AM in both TASSEL and GAPIT.

After fitting the MLM model, we obtained 15 QTL for seed weight using TASSEL and 19 QTL using GAPIT with both additive and dominant models. Fourteen of 15 QTL identified in TASSEL were shared with those from GAPIT (Supplementary Table S2). TASSEL has the capability of identifying additive effects and dominate effects after the additive model has been fitted. In the present study, TASSEL was used to identify QTL associated with seed weight and size. Combining the results from BLUE, 2012 and 2013, we identified 15 QTL for seed weight, 18 QTL for seed area size, 14 QTL for seed length, and 20 QTL for seed width (Table 1, Fig. 3). In total, 33 QTL were identified for IWG seed weight and size in the AM panel. Three QTL were located on LG1, 4, 5, 20, and 21, two on LG3, 8, 9, 14, and 15. No QTL was identified on LG 6, 7, or 19. The QTL were named *QSws.umn-LG1.1*, *QSws.umn-LG1.2*, *QSws.umn-LG1.3* and such (*Q* for QTL; *Sws* for seed weight and size; *umn* for the University of Minnesota; *LG1* for Linkage Group 1; and *1*, *2*, and *3* for occurrence numbers). The QTL were also assigned synonyms AM_1, 2, 3 et al., for ease of reference in this study.

Seven QTL for seed length were shared with seed weight, and nine QTL for seed width also contributed to seed weight (Fig. 4a, 4b). In total, 12 of the 15 QTL for seed weight were shared with either seed width or seed length. The shared QTL explained the high correlation between seed weight and seed length ($r = 0.69$) and seed width ($r = 0.60$). Similarly, high correlation between seed area size and seed length and seed width was observed, and area size shared 10 QTL with seed length and width, respectively. Seed weight and seed area size also showed high correlation ($r = 0.83$) and shared 10 QTL. Six QTL were shared between seed width and seed length, which also contributed to either seed weight or seed area size.

Fifteen QTL explained 37.6% of the variation for seed weight in the AM panel (Fig. 4a). All the QTL have small or moderate effects (≤0.81 mg, QTL AM_24; Table 2). Five QTL, AM_01, AM_06, AM_14, AM_29, and AM_33, showed overdominance where heterozygotes had higher effects than either homozygotes. The superior QTL alleles of AM_06, AM_14, AM_17 and AM_26 were homozygous in more than 50% of genets. The superior alleles of other QTL, however, were fixed in <15% of the population. If the superior alleles of seven QTL (AM_04, 07, 11, 18, 20, 24, and 32) are fixed, we project that the seed weight of the breeding population will be increased by 3 to 4 mg on average, based on marker effects.

Twenty QTL explained 32.0% of the variation for seed width in the AM panel (Table 2). The superior alleles of QTL, AM_03, 05, 07, 24, and 32, were homozygous in <15% of genets in the AM panel. Fourteen QTL for seed length explained 25.1% of the variance of the population. The frequency of genets with homozygous QTL AM_10 or AM_20 was <15%. Eighteen QTL for seed area size explained 38.2% of the variance of the population. The frequency of superior alleles of AM_06, 15, 19, 22, 26, and 30 is >70%. But QTL such as AM_04, 10, 18, 20, 24, and 32 need to be selected and fixed in the breeding population to increase seed weight and size.

Table 1. Identification of 33 quantitative trait loci (QTL) for seed weight, area size, length and width using association mapping.

					Seed weight			Area size			Seed length			Seed width		
ID†	QTL‡	Marker	LG§	Pos§	BLUE§	2012	2013	BLUE	2012	2013	BLUE	2012	2013	BLUE	2012	2013
AM_01	*Ti_QSws.umn_1.1*	TP38160	1	42.5	3.81¶	3.48	2.89	–	–	–	–	–	–	–	–	–
AM_02	*Ti_QSws.umn_1.2#*	TP96942	1	75.2	–	–	–	–	–	–	–	–	–	3.13	2.82	–
AM_03	*Ti_QSws.umn_1.3*	TP824692	1	137.0	–	–	–	–	–	–	–	–	–	3.45	2.72	3.19
AM_04	*Ti_QSws.umn_2.1#*	TP583965	2	51.5	2.63	–	2.72	–	–	–	–	–	–	–	–	–
		TP465797	2	52.0	–	–	–	2.67	–	2.99	–	–	–	–	–	–
		TP11430	2	55.9	–	–	–	–	–	–	–	–	–	3.96	3.58	2.91
AM_05	*Ti_QSws.umn_3.1#*	TP91425	3	56.1	–	–	–	–	–	–	–	–	–	–	–	3.01
		TP210957	3	57.3	–	–	–	3.26	3.69	–	3.12	3.45	–	–	–	–
AM_06	*Ti_QSws.umn_3.2*	TP122486	3	77.4	2.65	3.11	–	4.03	4.36	2.53	3.45	3.22	2.89	–	–	–
		TP608667	3	83.8	–	–	–	–	–	–	–	–	–	2.89	–	3.12
AM_07	*Ti_QSws.umn_4.1#*	TP693406	4	90.2	–	–	–	–	–	–	2.65	–	3.33	–	–	–
		TP88976	4	93.5	2.7	–	3.19	–	–	–	–	–	–	–	–	–
		TP361092	4	94.2	–	–	–	–	–	–	–	–	–	2.84	3.07	–
AM_08	*Ti_QSws.umn_4.2#*	TP633718	4	114.0	–	–	–	–	–	–	–	–	–	2.58	–	2.52
AM_09	*Ti_QSws.umn_4.3*	TP576833	4	129.4	–	–	–	–	–	–	–	–	–	2.57	–	3.17
AM_10	*Ti_QSws.umn_5.1*	TP238848	5	16.9	2.52	–	2.84	–	–	–	–	–	–	–	–	–
		TP889394	5	19.5	–	–	–	4.43	4.97	–	3.68	3.6	2.58	–	–	–
		TP123986	5	25.2	–	–	–	–	–	–	–	–	–	3.49	–	4.93
AM_11	*Ti_QSws.umn_5.2*	TP708348	5	32.3	3.66	–	3.82	–	–	–	–	–	–	–	–	–
		TP904298	5	37.6	–	–	–	–	–	–	–	–	–	5.22	4.77	4.50
		TP134848	5	38.1	–	–	–	3.51	2.74	3.50	–	–	–	–	–	–
AM_12	*Ti_QSws.umn_5.3*	TP892221	5	63.2	–	–	–	–	–	–	–	–	–	4.74	4.47	4.47
AM_13	*Ti_QSws.umn_8.1#*	TP867911	8	77.7	–	–	–	–	–	–	2.84	3	–	–	–	–
AM_14	*Ti_QSws.umn_8.2#*	TP227660	8	101.9	2.92	3.97	–	–	–	–	–	–	–	–	–	–
		TP70712	8	104.6	–	–	–	–	–	–	–	–	–	3.31	3.54	–
		TP285170	8	106.9	–	–	–	3.06	3.15	–	–	–	–	–	–	–
AM_15	*Ti_QSws.umn_9.1#*	TP225141	9	45.9	–	–	–	3.07	–	2.76	2.93	–	3.47	–	–	–
		TP112961	9	53.4	–	–	–	–	–	–	–	–	–	2.74	3.12	–
AM_16	*Ti_QSws.umn_9.2#*	TP604296	9	72.1	–	–	–	–	–	–	3.12	–	–	–	–	–
AM_17	*Ti_QSws.umn_10.1#*	TP898764	10	58.0	–	–	–	–	–	–	2.81	–	–	–	–	–
		TP252743	10	58.0	3.61	3.84	–	–	–	–	–	–	–	–	–	–
AM_18	*Ti_QSws.umn_11.1#*	TP299667	11	102.7	–	–	3.33	–	–	3.54	–	–	–	–	–	–
AM_19	*Ti_QSws.umn_12.1*	TP218981	12	26.5	–	–	–	3.40	3.17	–	2.88	2.74	–	–	–	–
AM_20	*Ti_QSws.umn_13.1#*	TP197489	13	55.8	3.48	–	4.2	–	–	2.68	–	–	2.61	–	–	–
		TP92513	13	57.2	–	–	–	–	–	–	–	–	–	–	–	3.00
AM_21	*Ti_QSws.umn_14.1#*	TP807369	14	88.2	–	–	–	–	–	–	–	–	–	3.27	3.93	–
AM_22	*Ti_QSws.umn_14.2*	TP60998	14	119.6	–	–	–	3.71	3.00	2.80	–	–	–	–	–	–
AM_23	*Ti_QSws.umn_15.1#*	TP547331	15	26.3	–	–	–	3.48	2.72	2.63	–	–	–	–	–	–
AM_24	*Ti_QSws.umn_15.2#*	TP508717	15	44.3	2.75	3.05	–	–	–	–	–	–	–	–	–	–
		TP791825	15	49.5	–	–	–	–	–	–	–	–	–	–	4.55	–
AM_25	*Ti_QSws.umn_16.1#*	TP184967	16	66.9	–	–	–	–	–	–	–	–	–	3.08	3.21	–
AM_26	*Ti_QSws.umn_17.1#*	TP678810	17	69.1	3.56	2.85	–	4.05	3.40	3.14	3.91	3.17	3.13	–	–	–
AM_27	*Ti_QSws.umn_18.1#*	TP510011	18	73.2	–	–	–	2.84	2.67	–	2.6	2.71	–	–	–	–
AM_28	*Ti_QSws.umn_20.1#*	TP285481	20	76.5	–	–	–	–	–	–	–	–	–	2.86	–	2.75
AM_29	*Ti_QSws.umn_20.2#*	TP883311	20	90.8	–	–	–	–	–	–	–	2.91	–	–	–	–
		TP73119	20	90.8	–	2.69	–	2.67	–	2.74	–	–	–	–	–	–
AM_30	*Ti_QSws.umn_20.3#*	TP18883	20	132.4	–	–	–	2.91	3.34	–	3.25	3.35	–	–	–	–
AM_31	*Ti_QSws.umn_21.1*	TP23563	21	106.2	–	–	–	3.88	2.75	3.23	–	–	–	4.05	2.97	3.60
AM_32	*Ti_QSws.umn_21.2#*	TP190628	21	120.4	–	–	–	3.03	3.23	–	–	–	–	–	–	–
		TP23097	21	128.8	2.65	–	–	–	–	–	–	–	–	3.20	2.69	2.68
AM_33	*Ti_QSws.umn_21.3#*	TP30866	21	141.7	2.51	–	3.44	–	–	–	–	–	–	–	–	–

† The ID of QTL in the association mapping study; AM for association mapping.

‡ The QTL were named as *QSws.umn-LG1.1*, *QSws.umn-LG1.2*, *QSws.umn-LG1.3* et al. (*Q* for QTL; *Sws* for Seed Weight and Shape; *umn* for the University of Minnesota; *LG1* for Linkage Group 1; and *1*, *2*, and *3* for occurrence numbers).

§ LG, linkage group; Pos, position in the consensus genetic map; BLUE, best linear unbiased estimation.

¶ The $-\log_{10}(p)$ values were shown in the table.

QTL shared between the linkage mapping and association mapping studies.

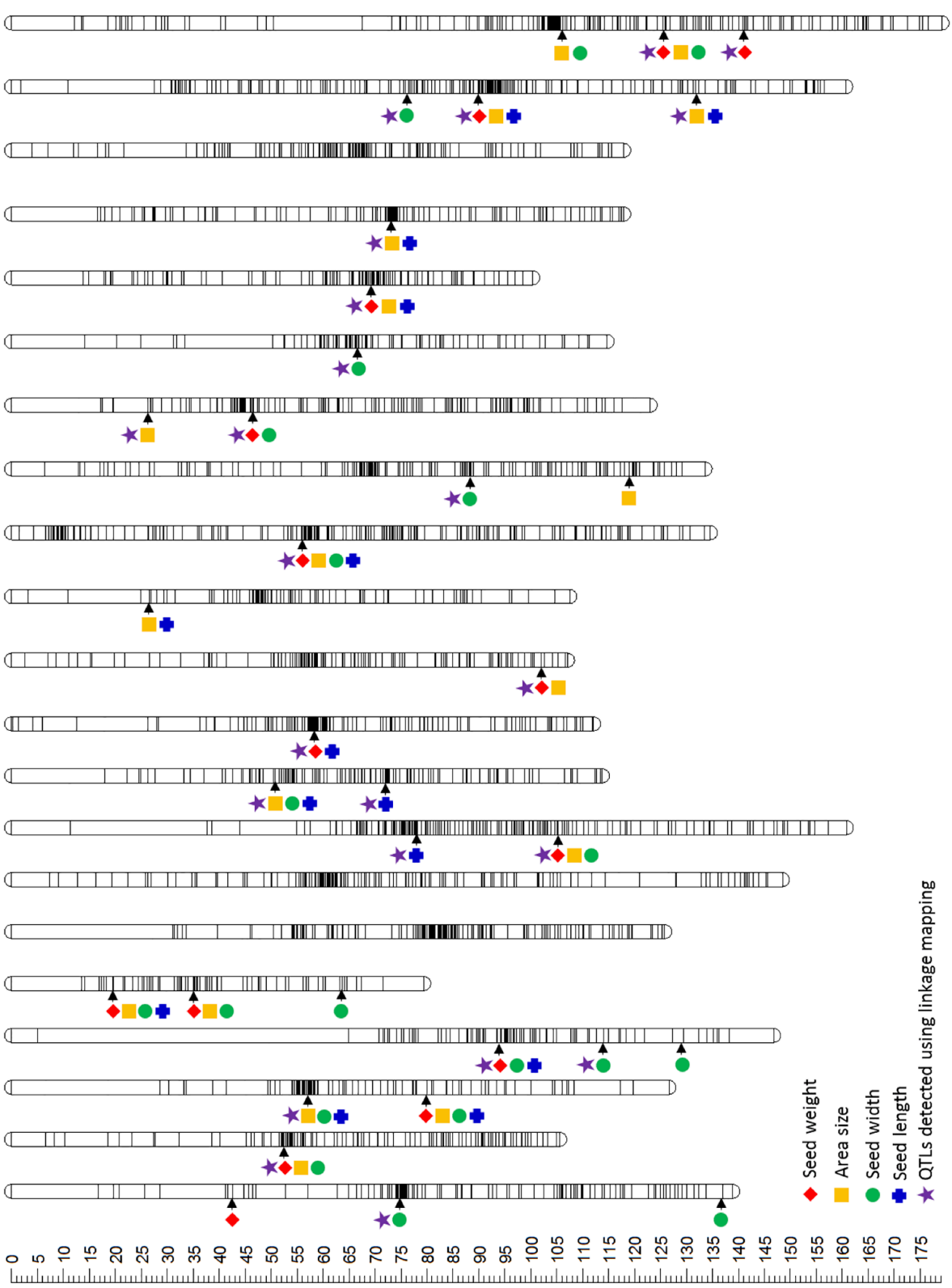


Fig. 3. Location of 33 quantitative trait loci (QTL) for seed weight, area size, length, and width in the consensus genetic map which were identified in the association mapping panel. The 4731 markers shared with the consensus genetic map were shown. Twenty-three QTL were also detected in two biparental populations using linkage mapping.

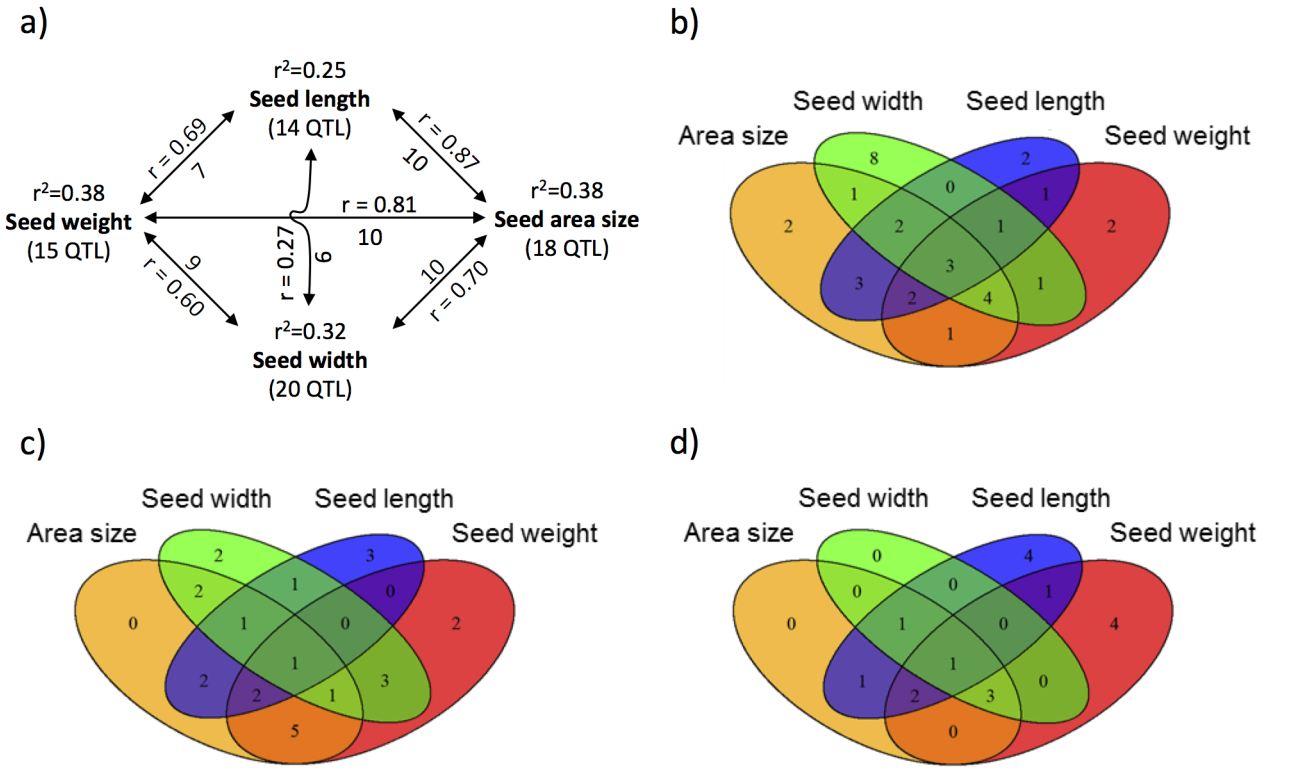


Fig. 4. High genetic overlap between seed weight, area size, length, and width. (a) Number of shared quantitative trait loci (QTL) and genetic correlation between traits in the association mapping study; (b) QTL overlap among traits in the association mapping study; (c) QTL overlap among traits in the biparental population C3-2331 × C3-2595; (d) QTL overlap among traits in the biparental population M35 × M26.

Linkage Mapping for Seed Weight and Size

The two parents, C3-2331 and C3-2595 were different in seed weight (13.20 vs. 9.79 mg, BLUE, $p < 0.01$), area size (11.53 vs. 9.09 mm^2) and seed length (7.45 vs. 6.02 mm), but similar in seed width (1.92 vs. 1.89 mm). These four traits all showed normal distribution in the population (Supplementary Fig. S4). Only two genets produced larger seeds than the superior parent C3-2331. However, 42.4% of the genets produced seeds with smaller seed weight than the inferior parent C3-2595, and 76.1% of the genets produced seeds with smaller seed width than both parents although both parents produced wide seeds (Supplementary Fig. S4).

In the C3-2331 × C3-2595 population, 1935 <*lmxll*> and <*nnxnp*> markers with less than 10% missing data were used for QTL mapping. Fourteen QTL for seed weight were detected, and the QTL LM_C01, C02, and C13 were observed in years 2014 and 2015 (Fig. 4c, Table 3, Supplementary Table S3). Fourteen QTL were identified for seed area size. Ten QTL for seed length were mapped, and two QTL, LM_C05 and C07 were observed in both environments. Eleven QTL for seed width were detected and LM_C04, C06 and C11 were significant in two environments. In total, 26 QTL were identified for seed weight and size. Among them, 19 were identified for more than one trait in at least one environment. Among 14 QTL for seed weight, nine were observed for seed area size, five for seed width, and three for seed length. All 14 QTL for seed area size were shared with other traits, but only three QTL were shared between seed length and seed width.

In the M35 × M26 population, the parents did not show large differences in seed weight or size. All the traits showed normal distribution in the population (Supplementary Fig. S4). In total, 17 QTL were identified for seed weight and size, of which 11 were detected for seed weight, eight for seed area size, 10 for seed length, and 5 for seed width, and 10 were detected for more than two traits. Among 11 QTL for seed weight, six were also detected for seed area size, four for seed length, and four for seed width. Only two QTL were shared between seed length and seed width (Table 3, Supplementary Table S4).

Thirty-eight QTL were detected from the two biparental populations, including five shared QTL. Each QTL explained only a small or moderate proportion of the phenotypic variance, varying from 2.6 to 12.7% (Table 3). Consistent with the results from AM, no major QTL were observed for seed weight and size. We observed 23 shared QTL between the biparental populations and the AM panel.

Frequency of QTL for Seed Weight was Increased by Phenotypic and Genomic Selection

A subset of the AM panel with 632 genets was used to analyze the frequency of the favorable homozygotes and the superior alleles of QTL for seed weight under the selection intensity of 2% (Fig. 5). Eleven of 15 QTL whose homozygotes showed different effect for seed weight were

Table 2. The effects and genotype frequency of quantitative trait loci (QTL) for seed weight, area size, length, and width in the breeding population.

		Seed weight					Area size					Seed length					Seed width				
			Allele1		Allele2			Allele1		Allele2			Allele1		Allele2			Allele1		Allele2	
ID	QTL	Marker	Eff†	Num†	Eff	Num	Marker	Eff	Num	Eff	Num	Marker	Eff	Num	Eff	Num	Marker	Eff	Num	Eff	Num
AM_01	*Ti_QSws.umn_1.1*	TP38160	−0.44	869/A	−0.36	32/G															
AM_02	*Ti_QSws.umn_1.2*‡																TP96942	−0.011	596/A	−0.038	69/G
AM_03	*Ti_QSws.umn_1.3*																TP824692	0.058	32/C	−0.003	821/G
AM_04	*Ti_QSws.umn_2.1*‡	TP583965	−0.34	621/A	0.18	50/G	TP465797	0.23	143/C	−0.09	138/T						TP11430	−0.008	477/A	−0.033	184/G
AM_05	*Ti_QSws.umn_3.1*‡						TP210957	−0.13	841/G	−0.52	22/T	TP210957	−0.06	841/G	−0.26	22/T	TP91425	0.025	20/A	−0.028	760/C
AM_06	*Ti_QSws.umn_3.2*	TP122486	−0.28	852/C	−0.73	12/T	TP122486	−0.20	853/C	−0.62	12/T	TP122486	−0.09	853/C	−0.30	12/T	TP608667	−0.027	151/A	0.005	301/G
AM_07	*Ti_QSws.umn_4.1*‡	TP88976	0.04	731/C	0.48	79/T						TP693406	−0.25	24/C	0.02	838/T	TP361092	0.017	154/C	−0.026	436/G
AM_08	*Ti_QSws.umn_4.2*‡																TP633718	−0.024	101/A	0.012	692/G
AM_09	*Ti_QSws.umn_4.3*																TP576833	−0.032	75/A	0.007	721/G
AM_10	*Ti_QSws.umn_5.1*	TP238848	0.30	55/C	0.28	777/T	TP889394	0.24	133/C	−0.74	6/T	TP889394	0.09	133/C	−0.41	6/T	TP123986	0.025	603/G	0.002	125/T
AM_11	*Ti_QSws.umn_5.2*	TP708348	0.10	656/C	0.61	56/T	TP134848	0.24	751/C	0.02	68/G						TP904298	0.003	92/A	0.032	650/T
AM_12	*Ti_QSws.umn_5.3*																TP892221	0.026	143/A	0.027	578/G
AM_13	*Ti_QSws.umn_8.1*‡											TP867911	0.09	219/C	0.08	164/G					
AM_14	*Ti_QSws.umn_8.2*‡	TP227660	−0.27	784/A	−0.83	10/G	TP285170	0.27	18/C	−0.22	812/T						TP70712	−0.015	418/C	−0.028	119/T
AM_15	*Ti_QSws.umn_9.1*‡						TP225141	−0.28	102/A	0.05	579/G	TP225141	−0.13	102/A	0.04	579/G	TP112961	−0.005	609/A	−0.047	44/G
AM_16	*Ti_QSws.umn_9.2*‡											TP604296	−0.10	438/A	−0.03	77/T					
AM_17	*Ti_QSws.umn_10.1*‡	TP252743	−0.32	71/A	0.48	646/G						TP898764	−0.10	803/C	−0.16	20/G					
AM_18	*Ti_QSws.umn_11.1*‡	TP299667	0.21	86/C	−0.27	264/T	TP299667	0.07	86/C	−0.17	267/T										
AM_19	*Ti_QSws.umn_12.1*						TP218981	−0.08	769/A	−0.73	11/T	TP218981	−0.04	769/A	−0.32	11/T					
AM_20	*Ti_QSws.umn_13.1*‡	TP197489	0.01	291/C	0.43	173/T	TP197489	0.00	293/C	0.19	173/T	TP197489	−0.01	293/C	0.10	173/T	TP92513	0.023	523/C	0.007	126/G
AM_21	*Ti_QSws.umn_14.1*‡																TP807369	−0.240	2/A	−0.006	956/C
AM_22	*Ti_QSws.umn_14.2*						TP60998	−0.07	702/C	−0.36	84/T										
AM_23	*Ti_QSws.umn_15.1*‡						TP547331	0.34	45/A	−0.12	491/G										
AM_24	*Ti_QSws.umn_15.2*‡	TP508717	0.67	33/C	−0.14	634/G											TP791825	−0.005	498/C	0.052	26/T
AM_25	*Ti_QSws.umn_16.1*‡																TP184967	0.001	99/A	−0.035	92/C
AM_26	*Ti_QSws.umn_17.1*‡	TP678810	−0.40	51/C	0.32	671/T	TP678810	−0.24	51/C	0.23	670/T	TP678810	−0.10	51/C	0.12	670/T					
AM_27	*Ti_QSws.umn_18.1*‡						TP510011	−0.22	14/A	−0.29	796/G	TP510011	−0.15	14/A	−0.13	796/G					
AM_28	*Ti_QSws.umn_20.1*‡																TP285481	−0.009	551/C	−0.035	84/G
AM_29	*Ti_QSws.umn_20.2*‡	TP73119	−0.31	609/T	0.00	454/W	TP73119	−0.21	608/T	0.00	457/W	TP883311	−0.21	15/C	−0.07	842/T					
AM_30	*Ti_QSws.umn_20.3*‡						TP18883	−0.66	5/G	−0.19	806/T	TP18883	−0.35	5/G	−0.10	806/T					
AM_31	*Ti_QSws.umn_21.1*						TP23563	−0.25	563/C	−0.02	62/T						TP23563	−0.032	563/C	−0.003	62/T
AM_32	*Ti_QSws.umn_21.2*‡	TP23097	−0.44	951/C	0.22	24/T	TP190628	−0.16	11/A	−0.27	860/G						TP23097	−0.042	953/C	0.000	24/T
AM_33	*Ti_QSws.umn_21.3*‡	TP30866	−0.23	289/C	−0.21	143/T															

† Eff, the effect of the homozygotes of allele1 or allele 2. The effect of the heterozygotes was considered as 0; Num, the number of genets.

‡ QTL shared between the linkage mapping and association mapping studies.

Table 3. Identification of 38 quantitative trait loci (QTL) for seed weight, area size, length, and width from two biparental populations.

					Seed weight			Area size			Seed length			Seed width		
ID†	QTL‡	LG§	Left§	Right§	BLUE§	2014	2015	BLUE	2014	2015	BLUE	2014	2015	BLUE	2014	2015
LM_C01	*Ti_QSws.umn_2.1*¶	2	50.0	83.3	4.2#	5.3	5.2	–	3.9	–	4.8	–	–	4.8	–	–
LM_C02	*Ti_QSws.umn_2.2*	2	44.5	96.7	–	5.5	3.5	–	–	–	–	–	–	–	–	–
LM_C03	*Ti_QSws.umn_3.1*¶	3	50.7	71.1	–	–	–	–	–	–	5.2	5.9	–	–	–	–
LM_C04	*Ti_QSws.umn_4.1*¶	4	82.2	100.4	–	–	–	–	–	7.4	–	–	–	6.7	4.7	6.2
LM_C05	*Ti_QSws.umn_4.2*¶	4	95.5	117.0	–	–	–	6.8	5.2	–	10.1	4.8	8.2	5.5	–	–
LM_C06	*Ti_QSws.umn_6.1*	6	66.7	91.4	5.7	–	6.3	6.0	–	10.3	–	–	–	6.5	6.4	8.8
LM_C07	*Ti_QSws.umn_6.2*	6	79.5	127.5	–	–	–	–	–	–	8.3	5.4	3.7	4.4	–	–
LM_C08	*Ti_QSws.umn_8.1*¶	8	66.4	87.8	–	–	–	–	–	–	–	–	–	–	5.5	–
LM_C09	*Ti_QSws.umn_8.2*¶	8	98.5	124.4	–	–	–	–	5.2	–	4.0	9.0	–	–	–	–
LM_C10	*Ti_QSws.umn_9.2*¶	9	56.5	87.0	–	–	–	–	–	–	–	–	4.0	–	–	–
LM_C11	*Ti_QSws.umn_10.1*¶	10	42.2	75.8	6.5	–	7.3	6.0	–	6.5	–	–	–	9.2	6.5	9.8
LM_C12	*Ti_QSws.umn_11.2*	11	55.6	58.7	–	9.3	–	–	–	–	–	–	–	–	–	–
LM_C13	*Ti_QSws.umn_11.1*¶	11	93.7	100.4	4.5	5.9	3.3	–	–	–	–	–	–	5.7	7.3	–
LM_C14	*Ti_QSws.umn_12.2*	12	72.6	99.6	–	10.2	–	–	–	–	–	–	–	4.4	–	–
LM_C15	*Ti_QSws.umn_13.1*¶	13	56.8	77.5	–	–	–	–	–	–	6.8	–	12.6	–	–	–
LM_C16	*Ti_QSws.umn_13.2*	13	73.2	106.9	–	–	–	–	–	–	–	–	–	–	–	4.5
LM_C17	*Ti_QSws.umn_14.3*	14	6.4	64.7	6.6	–	5.1	4.7	–	6.6	–	–	–	–	–	–
LM_C18	*Ti_QSws.umn_14.1*¶	14	66.0	103.7	3.5	6.6	–	12.7	–	9.9	–	–	–	–	–	–
LM_C19	*Ti_QSws.umn_15.1*¶	15	0.0	28.9	–	4.4	–	–	–	–	–	–	–	–	–	–
LM_C20	*Ti_QSws.umn_15.3*	15	59.7	116.0	–	4.1	–	5.2	5.6	–	–	3.8	–	–	–	–
LM_C21	*Ti_QSws.umn_17.1*¶	17	25.7	60.6	5.3	–	9.4	–	5.2	–	3.5	4.6	–	–	–	–
LM_C22	*Ti_QSws.umn_18.1*¶	18	73.5	73.6	–	–	–	3.6	–	–	–	–	–	–	–	–
LM_C23	*Ti_QSws.umn_20.4*	20	0.0	41.6	7.1	–	9.8	7.2	–	–	–	–	–	–	–	–
LM_C24	*Ti_QSws.umn_20.1*¶	20	41.6	76.5	–	–	–	–	–	7.1	–	–	4.6	–	–	–
LM_C25	*Ti_QSws.umn_20.3*¶	20	95.0	135.3	4.9	–	5.8	–	–	4.9	–	–	–	–	–	–
LM_C26	*Ti_QSws.umn_21.3*¶	21	136.5	176.6	–	–	–	–	–	–	–	–	–	–	5.7	–
LM_M01	*Ti_QSws.umn_1.2*¶	1	59.8	85.5	6.5	–	–	3.3	–	–	3.4	–	–	–	–	–
LM_M02	*Ti_QSws.umn_4.4*	4	30.0	47.6	3.3	–	–	–	–	–	–	–	–	–	–	–
LM_M03	*Ti_QSws.umn_6.3*	6	34.1	78.2	–	–	–	2.9	–	–	5.0	–	–	2.6	–	–
LM_M04	*Ti_QSws.umn_6.2*ᵛ	6	92.4	102.7	3.0	–	–	–	–	–	–	–	–	–	–	–
LM_M05	*Ti_QSws.umn_6.4*	6	145.3	176.1	–	–	–	–	–	–	4.1	–	–	–	–	–
LM_M06	*Ti_QSws.umn_7.1*	7	56.2	99.6	3.8	–	–	3.8	–	–	6.5	–	–	–	–	–
LM_M07	*Ti_QSws.umn_9.1*¶	9	33.1	47.8	5.4	–	–	5.0	–	–	6.8	–	–	4.0	–	–
LM_M08	*Ti_QSws.umn_11.1*¶††	11	87.5	100.0	4.3	–	–	5.4	–	–	–	–	–	6.6	–	–
LM_M09	*Ti_QSws.umn_15.2*¶	15	43.4	52.6	5.8	–	–	–	–	–	–	–	–	–	–	–
LM_M10	*Ti_QSws.umn_16.1*¶	16	34.4	78.0	3.8	–	–	3.6	–	–	–	–	–	7.9	–	–
LM_M11	*Ti_QSws.umn_17.1*¶††	17	45.1	69.1	–	–	–	4.1	–	–	5.0	–	–	–	–	–
LM_M12	*Ti_QSws.umn_18.1*¶††	18	40.6	73.6	–	–	–	–	–	–	2.8	–	–	–	–	–
LM_M13	*Ti_QSws.umn_19.1*	19	56.5	82.7	5.6	–	–	–	–	–	4.8	–	–	–	–	–
LM_M14	*Ti_QSws.umn_20.1*¶††	20	60.5	93.1	4.1	–	–	5.2	–	–	–	–	–	4.0	–	–
LM_M15	*Ti_QSws.umn_20.2*¶	20	71.0	122.2	–	–	–	–	–	–	5.0	–	–	–	–	–
LM_M16	*Ti_QSws.umn_21.4*	21	17.7	70.2	2.7	–	–	–	–	–	–	–	–	–	–	–
LM_M17	*Ti_QSws.umn_21.2*¶	21	105.2	124.3	–	–	–	–	–	–	3.5	–	–	–	–	–

† Identification of QTL in the linkage mapping study, LM for linkage mapping, C for population from C3-2331 × C3-2595, M for population from M35 × M26.

‡ The QTL were named as *QSws.umn-LG2.1* and *QSws.umn-LG2.2*, *QSws.umn-LG2.3* et al. (Q for QTL; *Sws* for Seed Weight and Size; *umn* for the University of Minnesota; *LG2* for Linkage Group 2, respectively; and 1, 2, and 3 for occurrence numbers).

§ LG, linkage group; Left, the location of the left marker of the QTL in the consensus genetic map; Right, the location of the right marker of the QTL in the consensus genetic map; The left and right markers of QTL were determined based on confidence interval calculated by one-LOD drop from the estimated QTL position; BLUE, best linear unbiased estimation.

the percentage of variance explained by the QTL was shown in the table.

¶ QTL shared between the linkage mapping and association mapping studies.

†† QTL shared between the two biparental populations.

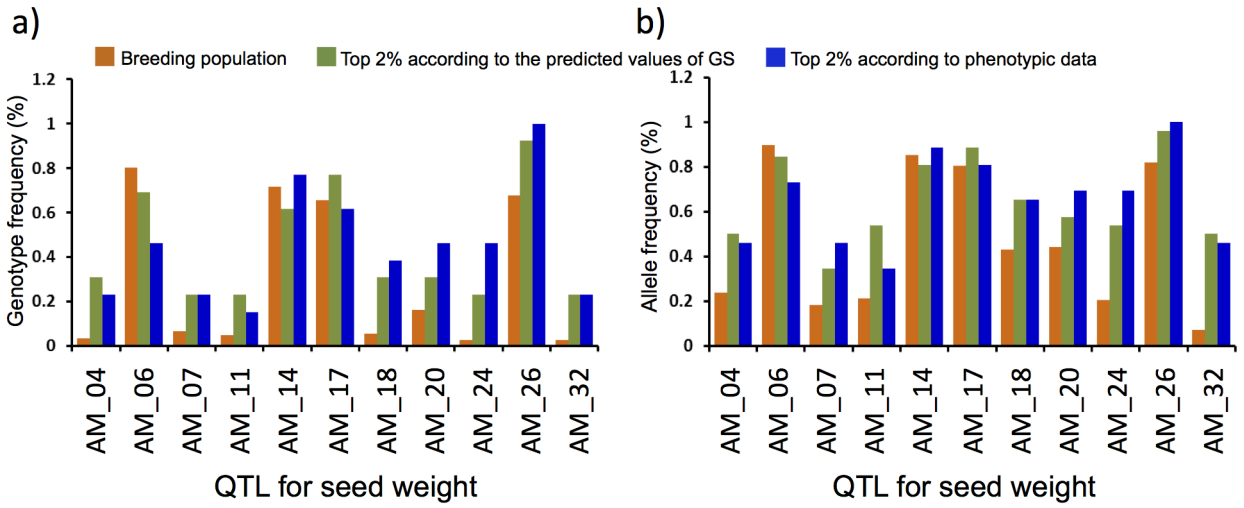


Fig. 5. The frequency of superior quantitative trait locus (QTL) genotypes or alleles for seed weight in three populations, that is, the original breeding population and the populations with the top 2% best genets from genomic selection (GS) or phenotypic selection. (a) The frequency of homozygotes with favorable QTL in the three populations. (b) The frequency of favorable QTL alleles in the three populations. Of 1126 genets in the breeding population (AM panel in the present study), 494 were used to train the genomic selection model for seed weight, which was used to calculate the predicted values of the remaining 632 genets. In the present study, the predicted values of 632 genets were used to determine the 2% top best genets. To make the data comparable, here the breeding population is the 632 genets, and the top 2% genets were also selected from the 632 genets according to the phenotypic data. Eleven of 15 QTL were analyzed, which mainly showed additive effect in the breeding population. These QTL can be fixed in the breeding population by increasing the frequency of homozygotes with favorable alleles.

included. For QTL AM_06, 14, 17, and 26, the frequency of superior homozygotes was larger than 65% in the initial breeding population (632 genets), while for the other seven QTL, the frequency of superior homozygotes was <16%. In the population of top 2% genets selected according to either predicted values or phenotypic data, the frequency of superior homozygotes for all the QTL was dramatically increased (>23%). The frequency of superior alleles was also increased to >46%. If the selected top 2% genets are used as parents under random mating (open pollination in green house), the theoretical frequency of the superior homozygotes for most QTL will be >21% (= 0.46 × 0.46) in the breeding population of the next selection cycle.

Discussion

Mapping QTL in a New Perennial Grain Crop Using Association Mapping and Biparental Linkage Mapping

To develop genomic tools for IWG improvement, we optimized the GBS methods for IWG (Zhang et al., 2016), which is powerful in discovering genome-wide markers for any species. We then developed an integrated genetic map with more than 10,000 GBS markers using seven biparental populations (Kantarski et al., 2016). This genetic map provided the location of markers for AM, and facilitated the development of genetic maps for linkage mapping. In the present study, to determine the genetic architecture of seed weight and size, we used a breeding population with 1126 genets as the AM panel because substantial variation was observed in the breeding population derived from only a few selection cycles. The mapped QTL from the breeding population can be directly used to improve IWG. Linkage mapping was also performed using two biparental populations. The large number (23) of shared QTL between AM (33 QTL) and linkage mapping (38 QTL) provided validation of the QTL identified for seed weight and size.

The power of AM depends on the degree of LD between the genotyped marker and the functional variant. The resolution of mapping QTL is a function of how quickly LD decays over distance (Myles et al., 2009). Therefore, the first step of AM is to analyze the extent of LD in the mapping population. The LD (r^2 = 0.16) decayed within 1 cM in the AM panel. In total, 4731 markers were discovered with an average of three markers every 2 cM. Thus, using 4731 markers we should be able to establish the linkage between markers and functional variants. In our previous study, we reported that the LD (r^2 = 0.2) decayed within 5 cM in the same population (Zhang et al., 2016). This difference in LD decay may be due to (i) a smaller number of markers (only 1158) used for LD estimation in the previous study; (ii) a partial consensus map used in the previous study (2016 of 2891 cM of the consensus map); and (iii) the inflated genetic distance of the genetic map, 4095 cM, from one biparental population in the previous study. The missing data (20%) of GBS markers used for genetic map development may cause misinterpretation of the sites and frequency of recombination, which was the primary cause of map distance inflation. In the present study, the consensus map with 10,029 markers was used as the reference map. The

genetic distance inflation was automatically adjusted during the integration of seven genetic maps.

The mixed model approach has been widely used in AM studies to reveal important QTL. During AM there is a need to account for population structure and pair-wise genetic kinship (QK model) to minimize false positives (Yu et al., 2006). Moreover, including population structure improved control of false positives only for traits highly correlated to population structure (Yu et al., 2006). The IWG breeding population was developed from the open pollination of all parents from the third selection cycle in The Land Institute (TLI_C3). The open pollination during the improvement of IWG helped to decrease the level of population structure (Zhang et al., 2016). In the present study, we only observed a low level of population structure in the AM panel. When including population structure, the Q matrix, as fixed effects in the model, we obtained similar mapping results as without including the Q matrix ($r > 0.99$ for seed weight), which indicated that, in this breeding population, the kinship or K matrix in the mixed model would be able to capture effects of population structure. Therefore, we did not include the Q matrix in our AM.

For IWG breeding, increasing seed weight is one of the major breeding objectives. The variation in seed width, length, and area size can contribute to the variation of seed weight (Tan et al., 2000), and there are high genetic correlations among these traits. The correlations could be derived from linkage where tightly linked loci control different traits and/or pleiotropy. In the present study, due to the limited number of markers and large genome size, it was difficult to distinguish linkage from pleiotropy in the AM and linkage mapping. Thus, we considered markers within ~10 cM as the same QTL within or between the QTL mapping studies. In AM, among 15 QTL for seed weight, 13 were also detected for seed width, length, and area size. In linkage mapping, more than 64% of QTL for seed weight were shared with the other three traits. The high number of shared QTL not only explained the high genetic correlations ($r > 0.60$) between seed weight and area size, seed length, and seed width, but also provided solid evidence of the authenticity of the QTL effects. Moreover, when the top 2% of genets with largest seed weight were selected, the frequency of superior alleles of these QTL were increased dramatically, which also indicated that these QTL contributed to seed weight.

Implications for Intermediate Wheatgrass Breeding

As a new perennial grain crop, IWG has been subjected to <10 breeding cycles. The population showed significant variation for agronomic traits such as seed size, threshability, and plant height (Zhang et al., 2016). We observed high heritability ($h^2 > 0.85$) of seed weight and size in the AM panel, indicating that the observed variation in seed weight and size were mainly controlled by genetic factors. More than 30 QTL for seed weight and size were identified using either AM or linkage mapping. These QTL showed moderate or small effects. It will be difficult to obtain large seed size by selection of only a few QTL, and multiple cycles of selection are likely needed to increase the frequency of favorable QTL alleles. This observation is consistent with the finding that seed weight was doubled after eight cycles of mass selection, but that each cycle of selection gave a small progressive increase (DeHaan et al., 2014). Among the 11 QTL for seed weight we analyzed in the present study, most genets have the favorable homozygotes of only three to five QTL, with seven as the largest number in the AM panel. Several more cycles of selection will be required to fix the favorable QTL alleles in the breeding population.

The effect of selection intensity on the frequency of favorable QTL alleles was tested in the present study. We found that selecting 13 (top 2%) of 632 genets of the breeding population was sufficient to increase the frequency of favorable QTL alleles and also to include all major QTL in the breeding population (Fig. 5). Higher selection intensity, say 0.5% (3 or 4 genets), was efficient to fix some QTL in one selection cycle; however, other QTL would be lost from the breeding population. Thus, if selection was made only for seed weight, for a breeding population with 1000 to 2000 genets, ~20 best genets should be a good number to both increase the allele frequency and keep the diversity of QTL. In our breeding program, however, we usually select for improving several traits. Usually 50 to 80 best plants, which perform best in one or several traits and better than the population mean in all traits, were selected as parents for the next cycles of recurrent selection.

In the present study, the efficiency of phenotypic versus genomic selection was compared. Genomic selection, where a subset of the breeding population was used as the training population, and the remaining genets as the independent validation population, performed similarly with phenotypic selection in terms of the improvement in the frequency of favorable homozygotes and alleles (Fig. 5). None of the favorable QTL alleles were lost when using genomic selection. This finding was consistent with the high predictive ability (0.66 for seed weight) observed when using the genomic selection procedure (Zhang et al., 2016). Now, genomic-selection based recurrent selection is being used in the IWG breeding program at the University of Minnesota to improve seed weight, threshability, head weight, lodging resistance, and shattering resistance (Zhang et al., 2016).

The C3-2331 × C3-2595 population showed transgressive segregation in seed weight and size. Even though both parents produced larger seeds than mean of the breeding population, very few progeny performed better than the superior parent, C3-2331. One of the main reasons is that the parent C3-2595 provided few favorable QTL alleles, one of seven for seed area size and one of nine for seed weight (BLUE). Thus, dissecting the composition of QTL in parental genets will help to select genets with complementary QTL for biparental crossing to increase the chance to obtain genets with larger seed weight. In our

breeding program, we will perform AM to detect QTL for seed weight using the training population as part of the genomic selection process (Zhang et al., 2016). These QTL will be used to check the composition of favorable QTL alleles in parents. Based on the composition difference among parents, we will make specific biparental crosses or three-way crosses. Their progeny will be selected using the genomic selection procedure to increase the frequency of favorable alleles. Thus, combining AM and genomic selection, we should be able to accelerate selection of favorable QTL alleles for seed weight and size.

The grain yield IWG needs to increase for this crop to be attractive to growers and industry on a large scale. Further increases in IWG production will be facilitated by the identification and characterization of genes or loci controlling grain yield. As one of the major components of grain yield, seed weight, and seed size were investigated in the present study. We determined the genetic architecture of seed weight and size using AM in a breeding population with 1126 genets and linkage mapping in two biparental populations. In total, 33 QTL were identified for seed weight and size in the breeding population, of which 23 were detected in biparental populations. All these QTL were mapped on the consensus map of IWG. This study provides a reference of QTL for seed weight and size in IWG. The identification of these QTL makes marker-assisted selection feasible for IWG. Specifically, breeders can select parents with different composition of QTL, make crosses, and then select progeny with larger seed weight, which will increase selection efficiency and genetic gain in seed weight improvement. Thus, combining AM, marker-assisted selection for parents, and genomic selection for progeny, we can develop IWG varieties for large areas of commercial production within a few years.

Supplemental Information Available

Supplementary Fig. S1. The number of markers in each linkage group and the length of linkage groups. The length of linkage groups were shown in Kosambi centiMorgans.

Supplementary Fig. S2. The rate of LD decay of the breeding population of intermediate wheatgrass. The LD among markers was estimated using Haploview 4.2 (Barrett et al., 2005). The Hill and Weir formula (Hill and Weir, 1988) was used to describe the LD decay of r^2.

Supplementary Fig. S3. Population structure of intermediate wheatgrass breeding population using STRUCTURE. a) Mean L(K)(± sd) for each K value over 10 runs. b) Delta K calculated using Delta$K$ = mean(|L″(K)|)/sd(L(K)) showing a peak at the 'true' value of K, here two clusters (but notice that K = 3 also shows a high value of Delta$K$).

Supplementary Figure S4. Histogram of seed weight, area size, length, and width in biparental populations a) from C3-2331 × C3-2595; b) from M35 × M26.

Supplementary Table S1. Variation of seed weight, area size, length, and width in the breeding population of intermediate wheatgrass in two growth seasons, 2011–2012 and 2012–2013.

Supplementary Table S2. QTL for seed weight identified from the breeding population using TASSEL and GAPIT.

Supplementary Table S3. Identification of 26 QTL for seed weight, area size, length and width from C3-2331 × C3-2595 population.

Supplementary Table S4. Identification of 17 QTL for seed weight, area size, length and width from M35 × M26 population.

Conflict of Interest Disclosure

The authors declare that there is no conflict of interest.

Acknowledgments

We would like to thank The Land Institute for their generous sharing the breeding germplasm and Dr. Jim Luby's fruit breeding group at the University of Minnesota for providing MapQTL6. We also would like to thank the University of Minnesota Genomics Center for their technical support in sequencing and the Minnesota Supercomputing Institute for their resources in genomic data analysis. This work was supported by The Initiative of Renewable Energy & the Environment, University of Minnesota, grant number RL_0015-12, and The Forever Green Initiative at the University of Minnesota.

References

Barrett, J.C., B. Fry, J. Maller, and M.J. Daly. 2005. Haploview: Analysis and visualization of LD and haplotype maps. Bioinformatics 21:263–265. doi:10.1093/bioinformatics/bth457

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

Carlson, B.M., S.W. Onusko and J.B. Gross. 2015. A high-density linkage map for *Astyanax mexicanus* using genotyping-by-sequencing technology. G3: Genes Genomes Genet. 5:241–251. doi:10.1534/g3.114.015438.

Cattani, D. 2016. Selection of a perennial grain for seed productivity across years: Intermediate wheatgrass as a test species. Can. J. Plant Sci. doi:10.1139/CJPS-2016-0280

Culman, S.W., S.S. Snapp, M. Ollenburger, B. Basso, and L.R. DeHaan. 2013. Soil and water quality rapidly responds to the perennial grain Kernza wheatgrass. Agron. J. 105:735–744. doi:10.2134/agronj2012.0273

Davey, J.W., P.A. Hohenlohe, P.D. Etter, J.Q. Boone, J.M. Catchen, and M.L. Blaxter. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat. Rev. Genet. 12:499–510. doi:10.1038/nrg3012

DeHaan, L.R., S. Wang, S. Larson, D. Cattani, X. Zhang, and T. Kantarski. 2014. Current efforts to develop perennial wheat and domesticate *Thinopyrum intermedium* as a perennial grain. In: C. Batello et al., editors, Perennial Crops for Food Security: Proc. of the FAO expert workshop, Rome. 28–30 Aug. 2013. p. 39–49. FAO, Rome.

Earl, D.A., and B.M. Vonholdt. 2012. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. Resour. 4:359–361. doi:10.1007/s12686-011-9548-7

Endelman, J.B. and J.L. Jannink. 2012. Shrinkage estimation of the realized relationship matrix. G3: Genes Genomes Genet. 2: 1405-1413. doi:10.1534/g3.112.004259.

Gao, L., M.K. Turner, S.M. Chao, J. Kolmer, and J.A. Anderson. 2016. Genome wide association study of seedling and adult plant leaf rust resistance in elite spring wheat breeding lines. PLoS One 11(2):e0148671. doi:10.1371/journal.pone.0148671

Glover, J.D., J.P. Reganold, L.W. Bell, J. Borevitz, E.C. Brummer, et al. 2010. Increased food and ecosystem security via perennial grains. Science 328:1638–1639. doi:10.1126/science.1188761

Gorjanc, G., M.A. Cleveland, R.D. Houston, and J.M. Hickey. 2015. Potential of genotyping-by-sequencing for genomic selection in livestock populations. Genet. Sel. Evol. 47:12–25. doi:10.1186/s12711-015-0102-z

Harmoney, K.R. 2015. Cool-season grass biomass in the southern mixed-grass prairie region of the USA. BioEnergy Res. 8:203–210. doi:10.1007/s12155-014-9514-9

Hill, W.G., and B.S. Weir. 1988. Variances and covariances of squared linkage disequilibria in finite populations. Theor. Popul. Biol. 33:54–78. doi:10.1016/0040-5809(88)90004-4

Huang, R., L. Jiang, J. Zheng, T. Wang, H. Wang, Y. Huang, et al. 2013. Genetic bases of rice grain shape: So many genes, so little known. Trends Plant Sci. 18:218–226. doi:10.1016/j.tplants.2012.11.001

Iquira, E., S. Humira, and B. Francois. 2015. Association mapping of QTLs for Sclerotinia stem rot resistance in a collection of soybean plant introductions using a genotyping by sequencing (GBS) approach. BMC Plant Biol. 15. doi:10.1186/s12870-014-0408-y

Jakobsson, M., and N.A. Rosenberg. 2007. CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801–1806. doi:10.1093/bioinformatics/btm233

Kantar, M.B., C.E. Tyl, K.M. Dorn, X. Zhang, J.M. Jungers, et al. 2016. Perennial grain and oilseed crops. Ann. Rev. Plant Biol. Vol. 67:703–729. doi:10.1146/annurev-arplant-043015-112311

Kantarski, T., S. Larson, X. Zhang, L. DeHaan, J. Borevitz, et al. 2016. Development of the first consensus genetic map of intermediate wheatgrass (*Thinopyrum intermedium*) using genotyping-by-sequencing. Theor. Appl. Genet. doi:10.1007/s00122-00016-02799-00127

Kearsey, M.J., and A.G.L. Farquhar. 1998. QTL analysis in plants; where are we now? Heredity 80:137–142. doi:10.1046/j.1365-2540.1998.00500.x

Lande, R., and R. Thompson. 1990. Efficiency of marker-assisted selection in the improvement of quantivative traits. Genetics 124:743–756.

Li, Q., X.H. Yang, G.H. Bai, M.L. Warburton, G. Mahuku, M. Gore, et al. 2010. Cloning and characterization of a putative GS3 ortholog involved in maize kernel development. Theor. Appl. Genet. 120:753–763. doi:10.1007/s00122-009-1196-x

Li, Y.B., C.C. Fan, Y.Z. Xing, Y.H. Jiang, L.J. Luo, L. Sun, et al. 2011. Natural variation in GS5 plays an important role in regulating grain size and yield in rice. Nat. Genet. 43:1266–U1134. doi:10.1038/ng.977

Lipka, A.E., F. Tian, Q.S. Wang, J. Peiffer, M. Li, P.J. Bradbury, et al. 2012. GAPIT: Genome association and prediction integrated tool. Bioinformatics 28:2397–2399. doi:10.1093/bioinformatics/bts444

Lu, F., A.E. Lipka, J. Glaubitz, R. Elshire, J.H. Cherney, M.D. Casler, et al. 2013. Switchgrass genomic diversity, ploidy, and evolution: Novel insights from a network-based SNP discovery protocol. PLoS Genet. 9(1):e1003215. doi:10.1371/journal.pgen.1003215

Mackay, T.F.C., E.A. Stone, and J.F. Ayroles. 2009. The genetics of quantitative traits: Challenges and prospects. Nat. Rev. Genet. 10:565–577. doi:10.1038/nrg2612

Mao, H.L., S.Y. Sun, J.L. Yao, C.R. Wang, S.B. Yu, et al. 2010. Linking differential domain functions of the GS3 protein to natural variation of grain size in rice. Proc. Natl. Acad. Sci. USA 107:19,579–19,584. doi:10.1073/pnas.1014419107

Moles, A.T., D.D. Ackerly, C.O. Webb, J.C. Tweddle, J.B. Dickie, and M. Westoby. 2005. A brief history of seed size. Science 307:576–580. doi:10.1126/science.1104863

Money, D., K. Gardner, Z. Migicovsky, H. Schwaninger, G.Y. Zhong and S. Myles. 2015. LinkImpute: Fast and accurate genotype imputation for nonmodel organisms. G3: Genes Genomes Genet. 5:2383–2390. doi:10.1534/g3.115.021667.

Morrell, P.L., E.S. Buckler, and J. Ross-Ibarra. 2012. Crop genomics: Advances and applications. Nat. Rev. Genet. 13:85–96. doi:10.1038/nrg3097

Myles, S., J. Peiffer, P.J. Brown, E.S. Ersoz, Z.W. Zhang, et al. 2009. Association mapping: Critical considerations shift from genotyping to experimental design. Plant Cell 21:2194–2202. doi:10.1105/tpc.109.068437

Pasam, R.K., R. Sharma, M. Malosetti, F.A. van Eeuwijk, G. Haseneyer, et al. 2012. Genome-wide association studies for agronomical traits in a world wide spring barley collection. BMC Plant Biol. 12:16–38. doi:10.1186/1471-2229-12-16

Poland, J.A., and T.W. Rife. 2012. Genotyping-by-sequencing for plant breeding and genetics. Plant Genome 5:92–102. doi:10.3835/plantgenome2012.05.0005

Pritchard, J.K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.

R Core Team. 2014. R: A language and environment for statistical computing. Available at http://www.R-project.org/. The R Foundation for Statistical Computing, Vienna, Austria.

Runck, B.C., M.B. Kantar, N.R. Jordan, J.A. Anderson, D.L. Wyse, J.O. Eckberg, et al. 2014. The reflective plant breeding paradigm: A robust system of germplasm development to support strategic diversification of agroecosystems. Crop Sci. 54:1939–1948. doi:10.2135/cropsci2014.03.0195

Russell, J., C. Hackett, P. Hedley, H. Liu, L. Milne, M. Bayer, et al. 2014. The use of genotyping by sequencing in blackcurrant (*Ribes nigrum*): Developing high-resolution linkage maps in species without reference genome sequences. Mol. Breed. 33:835–849. doi:10.1007/s11032-013-9996-8

Sallam, A.H., J.B. Endelman, J.-L. Jannink, and K.P. Smith. 2015. Assessing genomic selection prediction accuracy in a dynamic barley breeding population. Plant Genome 8. doi:10.3835/plantgenome2014.05.0020

Song, X.J., W. Huang, M. Shi, M.Z. Zhu, and H.X. Lin. 2007. A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. Nat. Genet. 39:623–630. doi:10.1038/ng2014

Tan, Y.F., Y.Z. Xing, J.X. Li, S.B. Yu, C.G. Xu, and Q.F. Zhang. 2000. Genetic bases of appearance quality of rice grains in Shanyou 63, an elite rice hybrid. Theor. Appl. Genet. 101:823–829. doi:10.1007/s001220051549

Tanabata, T., T. Shibaya, K. Hori, K. Ebana, and M. Yano. 2012. SmartGrain: High-throughput phenotyping software for measuring seed shape through image analysis. Plant Physiol. 160:1871–1880. doi:10.1104/pp.112.205120

Van Ooijen, J.M. 2009. MapQTL 6, Software for the mapping of quantitative trait loci in experimental populations of diploid species. Kyazma BV, Wageningen, the Netherlands.

Wagoner, P. 1990. Perennial grain: New use for intermediate wheatgrass. J. Soil Water Conserv. 45:81–82.

Williams, K., and M.E. Sorrells. 2014. Three-dimensional seed size and shape QTL in hexaploid wheat (*Triticum aestivum* L.) populations. Crop Sci. 54:98–110. doi:10.2135/cropsci2012.10.0609

Yu, J.M., and E.S. Buckler. 2006. Genetic association mapping and genome organization of maize. Curr. Opin. Biotechnol. 17:155–160. doi:10.1016/j.copbio.2006.02.003

Yu, J.M., G. Pressoir, W.H. Briggs, I.V. Bi, M. Yamasaki, J.F. Doebley, et al. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. 38:203–208. doi:10.1038/ng1702

Zanke, C.D., J. Ling, J. Plieske, S. Kollers, E. Ebmeyer, V. Korzun, et al. 2015. Analysis of main effect QTL for thousand grain weight in European winter wheat (*Triticum aestivum* L.) by genome-wide association mapping. Front. Plant Sci. 6. doi:10.3389/fpls.2015.00644

Zegeye, H., A. Rasheed, F. Makdis, A. Badebo, and F.C. Ogbonnaya. 2014. Genome-wide association mapping for seedling and adult plant resistance to stripe rust in synthetic hexaploid wheat. PLoS One 9(8):e105593. doi:10.1371/journal.pone.0105593

Zhang, X., J.B. Ohm, S. Haring, L.R. DeHaan, and J.A. Anderson. 2015. Towards the understanding of end-use quality in intermediate wheatgrass (*Thinopyrum intermedium*): High-molecular-weight glutenin subunits, protein polymerization, and mixing characteristics. J. Cereal Sci. 66:81–88. doi:10.1016/j.jcs.2015.10.008

Zhang, X., A. Sallam, L. Gao, T. Kantarski, J. Poland, et al. 2016. Establishment and optimization of genomic selection to accelerate the domestication and improvement of intermediate wheatgrass. Plant Genome 9. doi:10.3835/plantgenome2015.07.0059

Zhou, L., S.-B. Wang, J. Jian, Q.-C. Geng, J. Wen, Q. Song, et al. 2015. Identification of domestication-related loci associated with flowering time and seed size in soybean with the RAD-seq genotyping method. Sci. Rep. 5:9350. doi:10.1038/srep09350

Zhu, C., M. Gore, E.S. Buckler, and J.M. Yu. 2008. Status and prospects of association mapping in plants. Plant Genome 1:5–20. doi:10.3835/plantgenome2008.02.0089