Establishment and Optimization of Genomic Selection to Accelerate the Domestication and Improvement of Intermediate Wheatgrass

Xiaofei Zhang, Ahmad Sallam, Liangliang Gao, Traci Kantarski, Jesse Poland, Lee R. DeHaan, Donald L. Wyse, and James A. Anderson*

Abstract

Intermediate wheatgrass (IWG) is a perennial species and has edible and nutritious grain and desirable agronomic traits, including large seed size, high grain yield, and biomass. It also has the potential to provide ecosystem services and an economic return to farmers. However, because of its allohexaploidy and self-incompatibility, developing molecular markers for genetic analysis and molecular breeding has been challenging. In the present study, using genotyping-by-sequencing (GBS) technology, 3436 genome-wide markers discovered in a biparental population with 178 genets, were mapped to 21 linkage groups (LG) corresponding to 21 chromosomes of IWG. Genomic prediction models were developed using 3883 markers discovered in a breeding population containing 1126 representative genets from 58 half-sib families. High predictive ability was observed for seven agronomic traits using cross-validation, ranging from 0.46 for biomass to 0.67 for seed weight. Optimization results indicated that 8 to 10 genets from each half-sib family can form a good training population to predict the breeding value of their siblings, and 1600 genome-wide markers are adequate to capture the genetic variation in the current breeding population for genomic selection. Thus, with the advances in sequencing-based marker technologies, it was practical to perform molecular genetic analysis and molecular breeding on a new and challenging species like IWG, and genomic selection could increase the efficiency of recurrent selection and accelerate the domestication and improvement of IWG.

INTERMEDIATE WHEATGRASS [*Thinopyrum intermedium* (Host) Barkworth & D.R. Dewey; $2n = 6x = 42$] is a perennial grass and is genetically related to common wheat (*Triticum aestivum* L.), belonging to the Triticeae tribe of Pooideae (Mahelka et al., 2011). Producing large biomass, IWG is among the most productive cool-season forage species in the western United States (Harmoney, 2015). As a perennial species, IWG provides substantial environmental services relative to annual grain crops, including reduced soil and water erosion, reduced soil nitrate leaching, increased carbon sequestration, and reduced input of seed, tillage, energy, and pesticides (Culman et al., 2013; Glover et al., 2010; Robertson et al., 2000). In comparison with annual wheat, IWG has a more extensive root system which can capture more applied fertilizer and reduce total nitrate leaching by 86% or more (Culman et al., 2013). Compared with other perennial species, IWG also has desirable agronomic traits, including ease of threshing and harvesting, large seed size and grain yield, and edible and nutritious grain (DeHaan et al., 2014; Wagoner, 1990; Zhang et al., 2014). Thus, it is promising to domesticate and improve IWG as a perennial grain crop.

X. Zhang, L. Gao, D.L. Wyse, and J.A. Anderson, Dep. of Agronomy and Plant Genetics, Univ. of Minnesota, St. Paul, MN; A. Sallam, Dep. of Plant Pathology, Univ. of Minnesota, St. Paul, MN; T. Kantarski and J. Poland, Wheat Genetics Resource Center, Dep. of Plant Pathology, Kansas State Univ., Manhattan, KS; L.R. DeHaan, The Land Institute, Salina, KS. Received 27 July 2015. Accepted 17 Oct. 2015. *Corresponding author (ander319@umn.edu).

Abbreviations: BLUE, best linear unbiased estimate; GAUSS, Gaussian; GBS, genotyping-by-sequencing; HSD, honest significant difference; IWG, intermediate wheatgrass; LASSO, the least absolute shrinkage and selection operator; LOD, log of the odds; LD, linkage disequilibrium; LG, linkage group; QTL, quantitative trait locus; RHKS, reproducing kernel Hilbert space; RR-BLUP, ridge regression-best linear unbiased prediction; RRSV, repeated random subsampling validation; SNP, single nucleotide polymorphism; TLI, The Land Institute, Salina, KS; UNEAK, universal network-enabled analysis kit.
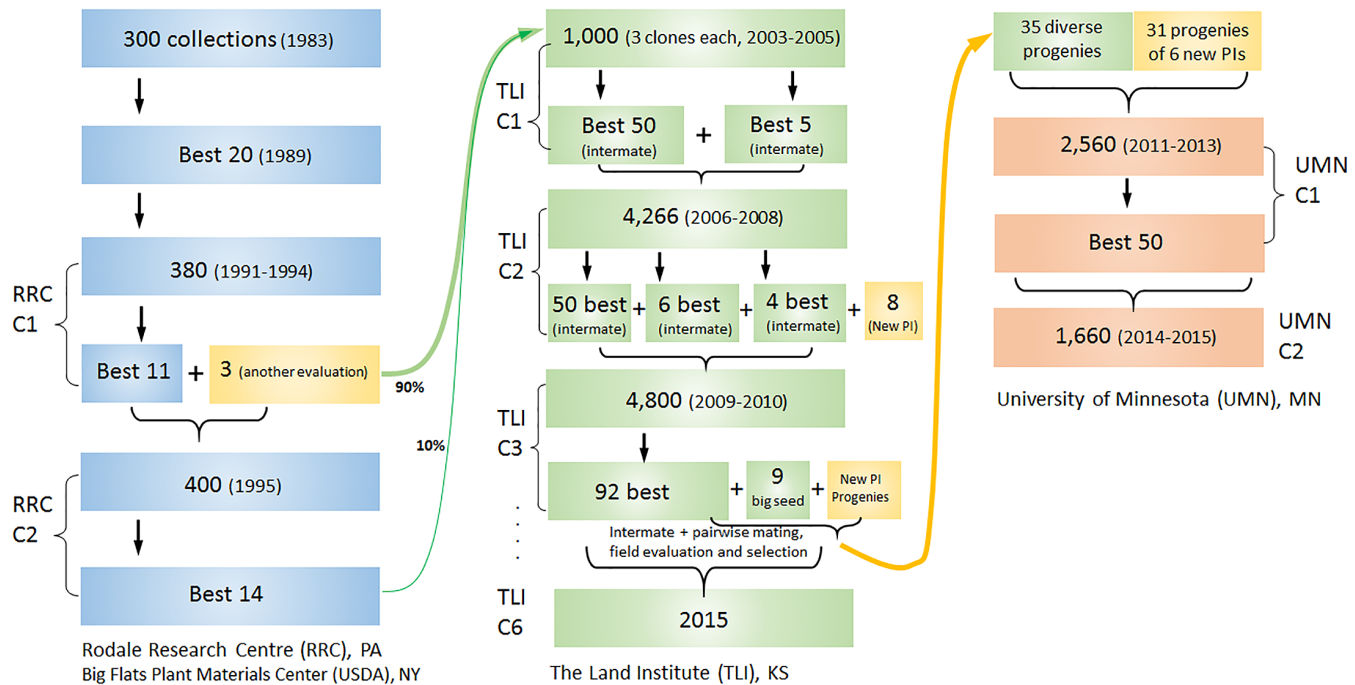
Figure 1. The history of intermediate wheatgrass domestication for grain yield. Recurrent selection was performed to improve yield per head, seed size, threshability, semidwarf stature, and early maturity. At the University of Minnesota, the breeding germplasm were derived from 66 female genets in the third selection cycle of The Land Institute (TLI-C3). The genets from the same female genet were designated as a family. Except for eight families each containing <10 genets, 58 half-sib families with 1126 genets were included in the genomic selection experiment. The term *genet* refers to a genetically unique, individual plant in an intermediate wheatgrass population. See the Materials and Methods section for details.

The domestication of IWG was initiated by Rodale Research Center, Kutztown, PA, in 1983 (Fig. 1). Intermediate wheatgrass was identified as the most promising perennial grain crop among nearly 100 species of perennial grasses. Seeds were produced with thousand-grain weight of 5.3 g on average, and the seeds can be mechanically harvested and threshed (Wagoner, 1990). The nutritional qualities of IWG are similar to wheat, but IWG has higher protein level and higher content of the sulfur containing amino acids, and whole flour of IWG grain performed well in baked products (Wagoner, 1990). After two cycles of selection performed at the Big Flats Plant Materials Center (Corning, NY), USDA Natural Resources Conservation Service, for grain yield and seed quality, the selected best plants were passed to scientists at The Land Institute, Salina, KS (TLI, Fig. 1). Since 2003, scientists at TLI have been working on the domestication of IWG by selection for improved yield per head, increased seed size, free threshing, reduced height, and early maturity. After two cycles of selection, the grain yield was increased by about 77% and seed size by about 23%, when IWG were grown in solid-seed plots (DeHaan et al., 2014). In 2011, the University of Minnesota (St. Paul, USA) and the University of Manitoba (Winnipeg, Canada), joined the domestication effort with the germplam from the third cycle of selection supplied by TLI (Fig. 1). The approach being used for domestication of IWG thus far is pedigree and phenotype-based recurrent selection (DeHaan et al., 2014).

Recent advances in sequencing based marker technologies are dramatically reducing the cost of genome-wide marker discovery, and these marker technologies can be used with any species, even those like IWG without previous genomic resources (Davey et al., 2011). Genotyping-by-sequencing is one of the most powerful marker technologies, has a simple protocol, and is suitable to species without reference genomes (Elshire et al., 2011; Lu et al., 2013; Poland et al., 2012). With GBS, cost and marker number can be easily adjusted by using 96-plex, 192-plex, and even 384-plex sequencing. Although GBS produces large amounts of missing data at these high multiplexing levels, imputation methods for ordered and unordered markers have been developed and tested (Rutkoski et al., 2013; Swarts et al., 2014). Genotyping-by-sequencing markers have been successfully used to discover genome-wide markers for developing genetic maps, mapping major quantitative trait loci (QTL), association analysis, and genomic selection (e.g., Carlson et al., 2015; Gorjanc et al., 2015; Iquira et al., 2015; Poland and Rife, 2012; Russell et al., 2014). Genome-wide markers, combined with statistical tools to associate marker variation with phenotypic variation, have the potential to revolutionize plant breeding and domestication.

Genomic selection, introduced by Meuwissen and colleagues (2001), is a marker and statistics based selection method that is being used to improve the effectiveness of breeding programs in many animal and plant species (reviewed by Hayes et al., 2009; Jannink et al.,

2010). Moderate or high cross-validation prediction accuracy was obtained for yield and some other quantitative traits in natural populations or breeding populations of maize (*Zea mays* L.), rice (*Oryza sativa* L.), wheat, oat (*Avena sativa* L.), rye (*Secale cereale* L.), barley (*Hordeum vulgare* L.), and switchgrass (*Panicum virgatum* L.; Asoro et al., 2011; Lipka et al., 2014; Rutkoski et al., 2014; Sallam et al., 2015; Spindel et al., 2015; Wang et al., 2014; Zhao et al., 2012). These studies showed that genomic selection has the capacity to improve the efficiency of breeding programs by increasing selection accuracy and reducing breeding cycle time compared with phenotypic selection.

With high-density single nucleotide polymorphism (SNP) markers, all QTL are in linkage disequilibrium (LD) with at least one marker, and breeding values can be estimated as the sum of all marker effects by regressing phenotypic values on all available markers. Thus, the predictive ability of genomic selection is affected by the marker density and the decay of LD between markers and QTL. Increasing the marker density will improve prediction accuracy by capturing more genetic variation and ensuring the conservation of marker-QTL association (Asoro et al., 2011; de Roos et al., 2009; Heffner et al., 2011b; Zhao et al., 2012). Marker density required for genomic selection depends on the rate of LD decay. Maize has a fast rate of LD decay compared with barley or wheat, so a larger number of markers is needed for maize (Heffner et al., 2011a; Zhao et al., 2012). In contrast, for a biparental population in apple (*Malus domestica* Borkh.), where 50% of adjacent markers had an LD > 0.2, the accuracy of genomic selection with only 2500 markers was 0.7 (Kumar et al., 2012).

The rate of LD decay is determined by $N_e \times c$, where N_e is the effective population size and c is the recombination rate in Morgans between two markers (Gaut and Long, 2003). Domesticated crop species generally have higher LD due to smaller N_e through domestication (e.g., barley and wheat, ~50; Lorenz et al., 2011; Thuillet et al., 2005). In contrast, some plants species with outbred reproductive habit have high N_e (e.g., ryegrass [*Lolium perenne* L.] and eucalyptus [*Eucalyptus* spp.]), and prediction accuracy with genomic selection can be <0.1 (Lin et al., 2014). The relatively small N_e of wheat and barley is an advantage when applying genomic selection in breeding programs. The smaller N_e and, in turn, lower rate of LD decay, allow for higher prediction accuracies than in outbred species at the same density of markers. Another factor affecting predictive ability is size and composition of the training population. The training population provides the phenotypic and genotypic information to estimate the marker effects. In general, as the size of training population increases, the estimated accuracy of marker effects increases, and consequently, the predictive ability for selection candidates is improved (Lin et al., 2014). In addition, the relationship of the training population to validation individuals effectively determines the prediction accuracy (Habier et al., 2007; Wurschum et al.,

2013). With a stronger relationship, a higher accuracy can be obtained.

The type of genomic selection models used is another factor affecting the accuracy of genomic selection prediction accuracy. These models include ridge regression-best linear unbiased prediction (RR-BLUP), Gaussian (GAUSS), Bayes A, Bayes B, Bayes Cπ, the least absolute shrinkage and selection operator (LASSO), Bayesian ridge regression, Bayesian reproducing Kernel Hilbert Space (RKHS), Random Forest, and others (Desta and Ortiz, 2014; Endelman, 2011; Perez and de los Campos, 2014). The prediction accuracy varies among models due to their different assumptions and treatments of marker effects. RR-BLUP assigns equal variance to all markers, whereas Bayesian models allow unequal variance among markers. The RKHS regression and GAUSS models can capture both the additive effects and nonadditive interactions among loci by adding a kernel function that includes interactions among marker covariates into the model (Endelman, 2011; Perez and de los Campos, 2014). Random Forest takes advantages of all tree nodes to find the best prediction model and also capture the interactions between loci (Gonzalez-Recio and Forni, 2011). Results from various studies have shown that no single model uniformly outperforms the others across all traits. The performance of models may depend on how well their assumption and treatments of marker effects match the genetic architecture of traits and the population structure (e.g., Bao et al., 2014; Spindel et al., 2015). Many quantitative traits controlled by small numbers of major QTL are better suited to Bayesian models than to linear regression, such as RR-BLUP (Desta and Ortiz, 2014).

IWG is an allohexaploid outcrossing species. The three subgenomes and heterozygosity make it challenging to develop molecular markers, perform genetic analysis and find marker associations for agronomic traits. Molecular breeding technologies have not been available in the domestication and improvement of IWG. Here, GBS was used to discover genome-wide SNP markers and the potential of genomic selection was explored to improve the breeding efficiency of IWG, particularly for seven agronomic traits. The objectives of the present study are: (i) to develop a genetic map of IWG using a biparental population genotyped with GBS; (ii) to characterize seven agronomic traits in the current IWG breeding program in Minnesota; (iii) to discover genome-wide SNP markers of the breeding population and estimate the rate of LD decay using markers common to both the biparental population and the breeding population; (iv) to develop genomic selection models using the breeding population, estimate the predictive ability using cross-validation, and optimize the marker density and the size and composition of the training population. Finally, a genomic selection-based breeding scheme was proposed to increase the efficiency of recurrent selection in IWG breeding programs.

Materials and Methods
Terminology for IWG: Genet
The terms *line* and *variety* for inbred species are not proper to define the individual plant of IWG, an out-crossing species. *Genotype* can refer to individual plants in a population of outcrossing species. But genotype is also widely used to define the DNA sequence characteristics of individual plants, lines, or varieties. Both meanings of genotype have to be used in the genetic analysis of IWG. To avoid confusion, we use the term *genet* to refer to a plant genotype, a genetically unique, individual plant in an IWG population. Genet is commonly used by ecologists to define an organism that grows from a fertilized egg and, therefore, is genetically unique (Beeby and Brennan, 2008). Here, we would like to introduce genet to the field of agronomy and plant genetics, to refer to individual plants in an outcrossing species like IWG.

Plant Materials
An F_1 population was derived from a cross between C3-2331 and C3-2595. Both parents are from the third recurrent selection cycle at TLI. The parents have a large difference in seed weight, with C3-2331 producing large seeds. This population consists of 178 genets and was used to develop a genetic map. The breeding population of UMN-C1 was composed of 2560 genets derived from 66 female genets of the third recurrent selection cycle at TLI (Fig. 1). The genets from the same female parent were designated as a family. The UMN-C1 genets were planted on 0.9 m centers in St. Paul, MN, on 15 and 16 Sept. 2011. They were organized in five blocks (A, B, C, D, and E) in the selection nursery (Supplementary Fig. S1), and no genets were cloned. Each block had 10 genets on average (ranging from 4 to 26) from each of 66 families. Ten genets from each family were planted next to each other in the field. For the genomic selection experiment, 58 families were included in the present study, and the other eight families were excluded because of the small number of genets (<10) in each family. Fifty-eight families from Blocks A and B composed the majority of the population for genomic selection. Several families from Blocks C and D were also included. Most genets in Blocks C and D and all genets in Block E were only used for breeding but not included in the genomic selection study. In the genomic selection population, the number of genets in each family varied from 11 to 44 (Supplementary Fig. S1). The plants were grown during the 2012 and 2013 field seasons. A total of 67 kg ha^{-1} N fertilizer was applied in April each year when plants resumed growth. Weeds were controlled by spraying Dual Magnum (a.i. S-metolachlor 82.4%, Syngenta, Basel, Switzerland) at the labeled rate in April, with a second treatment in June as needed. Weather data were obtained from the Minnesota Department of Natural Resources. In general, the 2012 season had higher temperature but less precipitation, except for February, May, and July, than the 2013 growth season (Supplementary Fig. S2).

Agronomic Traits
The breeding population was evaluated for seven agronomic traits during the 2012 and 2013 field seasons. These traits included heading score, plant height, head weight, grain yield, threshability, seed weight, and biomass. Heading score, which is related to heading date, was recorded with a 1 to 6 score based on the average stage of multiple heads within a plant, when 20% of genets in the field have fully emerged seed heads: (1) no head emerging; (2) 20% emerged seed heads; (3) 40% emerged seed heads; (4) 60% emerged seed heads; (5) 80% emerged seed heads; and (6) fully emerged seed heads. Plant height was measured in centimeters from the soil surface to the tip of tallest spike at physiological maturity. For head weight, five random seed heads per plant were harvested and their weight was recorded. For grain yield, all the seed heads per plant were harvest by hand and threshed using a laboratory thresher LD 350 (Wintersteiger, Ried, Austria). The seeds were cleaned and their weight was recorded. Threshability was scored on a 1 to 4 scale based on the percentage of naked seeds by visual observation: (1) < 5% naked seeds; (2) 5–15% naked seeds; (3) 15–50% naked seeds; (4) >50% naked seeds. About 200 seeds from each plant were dehulled using a dehuller (Wintersteiger, Ried, Austria). Fifty naked seeds per plant were weighed to calculate seed weight. The biomass of individual plants was obtained by summing the dry weight of head and stem sections. All heads were harvested by hand for grain yield, and their dry weight was recorded. Three weeks after seed head harvest, the residue was cut from 5 cm to the soil surface and measured to record the fresh weight, and 10% of the genets were randomly selected to measure the dry weight and calculate the water content. For the randomly selected 141 genets, the correlation coefficient between dry weight and fresh weight was 0.97. By adjusting for average water content, the dry residue mass of all plants was obtained.

Phenotype Adjustment and Correlation Analysis
Correction for variability between 2012 and 2013 field seasons was done by calculating best linear unbiased estimates (BLUE) of each genet using the MIXED procedure in SAS (v. 9.3.1; Sallam et al., 2015). All genets were used to estimate the year effects for all traits by treating years as fixed effects in the mixed model equation. Then for all traits, BLUE for each genet in each year was calculated by correcting for the year effect estimated from the previous step. Variance components were estimated using restricted maximum likelihood (REML) in the MIXED procedure in SAS. Broad-sense heritability h^2 on a genet mean was estimated for all traits using the equation $h^2 = \sigma_g^2/(\sigma_g^2 + \sigma_e^2/n)$, where σ_g^2 is genetic variance, σ_e^2 is the error variance that includes G × E and residuals, and n is the number of years. The adjusted phenotypic data (BLUE) were also used to calculate the genetic correlation matrix for each trait using the rcorr() function in the Hmisc package of R (R Core Team, 2014).

GBS Libraries, Sequencing, and SNP Calling

Young leaf tissues were harvested from each genet from the field in St. Paul, MN. DNA was extracted using a BioSprint 96 DNA Plant Kit (Qiagen, the Netherlands). The concentration of DNA was determined using Picogreen (Promega, WI) and normalized. A total of 200 ng DNA from each sample was used to develop GBS libraries. GBS libraries were made following the protocol of Poland et al. (2012) using the two-enzyme method, with some modifications described below and discussed further in the Results and Discussion sections. Two pairs of enzymes, *Pst*I/*Msp*I and *Pst*I/*Ape*KI, were tested, and *Pst*I/*Msp*I performed better for IWG. The genomic DNA from each sample was digested with 18U of *Pst*I and 12U of *Msp*I in a reaction volume of 30 μL. After digestion, the genomic DNA of each sample was ligated with two barcode adaptors (in total 0.5 pmol) and one common adaptor (15 pmol) designed by Poland et al. (2012). Every 48 samples from the F_1 population and every 96 samples from the breeding population were pooled. Each of the two parents of the F_1 population was considered as six samples to guarantee six times sequencing coverage for both parents. After cleanup and amplification, size selection was performed using LabChip GX (Caliper Life Sciences, MA). The adaptor dimers were discarded and only DNA fragments with size ranging from 150 to 250 bp were kept for sequencing. The sequencing was performed using Illumina HiSeq 2000 at the University of Minnesota Genomics Center.

The Universal Network-Enabled Analysis Kit (UNEAK) pipeline (Lu et al., 2013) was used to discover SNP markers with a minimum call rate of 0.5 and minimum minor allele frequency of 0.05. The SNP markers were filtered with a goodness-of-fit χ^2 test ($p > 0.01$) 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. Considering the low sequencing coverage of GBS, for a heterozygote, one of the alleles might not be captured, and consequently, the genotype may mistakenly be considered a homozygote. To decrease the rate of false homozygotes, the homozygotes with sequencing count of alleles less than five for the F_1 population and four for the breeding population were considered as missing data.

Genetic Map Construction

Two parents had six times deeper sequencing coverage than their progeny. Their putative homozygous markers whose sequencing count was <10 were set to missing, and the markers with missing data for both parents were removed. According to a pseudo-testcross method, all the filtered markers were divided into three groups, nn × np, lm × ll, and hk × hk for JoinMap 4.1 (Grattapaglia and Sederoff, 1994; Van Ooijen, 2006, 2011b). The other types of markers were discarded. To decrease the effect of missing data on the order of markers, we first used the markers with <3% missing data to develop a genetic map. The minimum LOD (log of the odds) threshold for groups was determined by identifying the *grouping tree* branches with stable marker numbers over increasing consecutive LOD value. Groups with more than five markers were kept for mapping. Marker order and distances were determined using the Maximum Likelihood mapping algorithm for crossing pollinated populations in JoinMap 4.1. The markers with both distorted segregation and a distance of >15 cM to the adjacent marker were removed. The process was repeated until the distance of the adjacent markers became <15 cM. Using these markers as fixed orders, markers with <10% missing data were mapped using the same procedure. Finally, markers with <20% missing data were integrated into the genetic map. The unit of genetic map was converted from "Haldane centiMorgans" to "Kosambi centiMorgans" (Van Ooijen, 2011a). The genetic map was displayed using MapChart2.2 (Voorrips, 2002).

Collinearity Analysis with Barley and *A. tauschii*

Barley genome sequences (cv. Morex) were downloaded from http://www.mmnt.net/db/0/0/ftp.mips.embnet.org/plants/barley/public_data/sequences (verified 23 Nov. 2015) and were converted into a local BLAST database using the BLAST command line tool makeblastdb. The physical positions of barley contigs were obtained from ftp://ftp-mips.helmholtz-muenchen.de/plants/barley/public_data/anchoring/WGS_ANCHORED_280512_AC2.FA (verified 23 Nov. 2015). The mapped marker sequences were used to align with the barley genome database using the blastall command. The threshold of e-value was set as 1×10^{-5} and the markers matching unique locations of the barley genome were retained. For the markers matching multiple locations on barley chromosomes, only the location with the smallest e-value and whose e-value was 1×10^{-5} times smaller than those of the other locations was considered as the matched location. The same procedure was used to analyze the collinearity between IWG and *Aegilops tauschii* Coss. A total of 429,891 scaffold sequences (KD499222 to KD929112) of *A. tauschii* accession AL8/78 were downloaded from GenBank and the genetic locations of scaffolds were obtained from Supplementary Table 13 of Jia et al. (2013). The collinearity results were visualized with Circos v. 0.64 (Krzywinski et al., 2009). The Spearman's rank correlation test for the matched LG and chromosomes was performed using R package pspearman (Savicky, 2014).

LD decay of the IWG Population

The markers with <30% missing data from the breeding population were converted to the database format using the BLAST command line makeblastdb. The mapped markers from the F_1 population were used to BLAST the database to identify the shared markers between the F_1 population and the breeding population. The threshold of e-value was set as 1×10^{-15}, and only the markers with one nucleotide mismatch or less from both populations were kept for subsequent analysis. The shared markers were used to estimate the LD among markers in the breeding population using Haploview (Barrett et al., 2005). The Hill and Weir formula was used to describe the LD decay of r^2 (Hill and Weir, 1988).

Genomic Selection

GBS markers were imputed with Random Forest in the R package MissForest, which was verified to be the best imputation method for the missing data of unordered markers in genomic selection (Rutkoski et al., 2013). To assess the capability of the markers to predict agronomic traits, nine genomic selection models were tested: RR-BLUP, GAUSS, Bayes A, Bayes B, Bayes Cπ, Bayesian LASSO, Bayesian ridge regression, Bayesian RKHS, and Random Forest. All statistical modeling was done in R (R Core Team, 2014). RR-BLUP and GAUSS models were calculated using function kin.BLUP in the rrBLUP package (Endelman, 2011). Bayesian models and the RKHS model were calculated using the BGLR package (Perez and de los Campos, 2014). Random Forest was performed using the package randomForest (Gonzalez-Recio and Forni, 2011). Specifically, a total of 10,000 burn-ins and 50,000 iterations of Markov-Chain Monte Carlo simulation were used in Bayesian models and RKHS model, and 500 trees and four branches were used for the Random Forest model. All other parameters in the models followed the guidelines and examples in the references and R packages instructions.

The performance of each model was assessed by means of a cross-validation scheme, repeated random subsampling validation (RRSV; Usai et al., 2009). In the breeding population, 80% of the genets were randomly selected and used as the training set to fit each prediction model, and the other 20% were used as a validation set to assess the correlation between the observed and predicted values. The sampling process without replacement was repeated 100 times. For each trait, Pearson's correlation coefficient is considered as the predictive ability, and the prediction accuracy is calculated by dividing predictive ability by the square root of heritability.

Optimization of Size and Composition of the Training Set

The 58 families were divided into training and validation sets. Different numbers of families (i.e., 5, 10, 20, 30, 40, and 50) were randomly sampled and used as the training set, and the other families were used as the validation set. For each specific number of families, samples were drawn without replacement 100 times. The Studentized range statistic, Tukey's Honest Significant Difference method (HSD, R Core Team, 2014) was used to test for significant difference in predictive ability from the six divisions of families.

All 58 families were included in both the training and the validation sets. The same number of genets from each family was randomly sampled as the training set and the other genets in the families were used as the validation set. A series of numbers of genets from each family in the training sets were tested, that is, 1, 2, 4, 6, 8, and 10. For each number of genets, samples were drawn without replacement 100 times. Tukey's HSD method was used to test the difference in predictive ability among six training sets.

Predictive Ability among Half-Siblings

The genomic selection population is mainly from Blocks A and B in the IWG field (Supplementary Fig. S1). In each block, there were 58 families and each family contains 10 genets on average. Genets in one of the two blocks were used as the training set, and the other genets were considered as the validation set. The predictive ability was estimated among half-siblings.

Optimization of Marker Density

Seven subsets of markers, 40, 200, 400, 800, 1600, 2400, and 3200, were sampled from the 3883 markers with <30% missing data. One hundred samplings were performed without replacement. The genets in Block A of the breeding population served as the training set, and the other genets were used as the validation set. The GAUSS model was used to obtain the predictive ability. Tukey's HSD method was used to test the significant difference among seven marker subsets.

Results

Substantial Phenotypic Variation among IWG Genets

As a perennial species, IWG produced much larger grain yield and biomass in the second year (2013) than the first year (2012; Fig. 2a, Supplmentary Fig. S3). The population mean of grain yield per plant (83.6 g) in 2013 almost doubled that of 2012 (43.9 g). The other five traits, seed weight, heading score, height, threshability, and head weight, were similar across years. For individual genets, each trait was significantly correlated across years; for example, seed weight, $r = 0.74$, and grain yield, $r = 0.57$ (Fig. 2b; Supplementary Fig. S3).

Within the IWG breeding population, significant variation was observed for each of the seven traits. The difference between minimum and maximum values of each trait ranged from over threefold for height and over 14-fold for head weight (Table 1). Heritability estimates were moderate for heading score, grain yield, biomass, head weight, height, and threshability (0.59 to 0.76) and high for seed weight (0.85; Table 1). The phenotypic data among 58 IWG families also showed extensive variability for all seven traits (Table 1; Supplementary Fig. S4). For example, the mean seed weight of some families was >9.0 mg, whereas some families produced small seed with a mean of 7.5 mg (Supplementary Fig. S4). But in general, the genetic variation among half-sib families was smaller than that within family (Table 1).

The genetic correlations between the seven traits were generally low (Supplemental Table S1), but high correlation was observed between grain yield and biomass ($r = 0.75$), and plant height was also significantly correlated with biomass and grain yield ($r = 0.54$ and 0.43, respectively). Also, a correlation coefficient of $r = 0.52$ was observed between head weight and grain yield ($p < 0.01$).
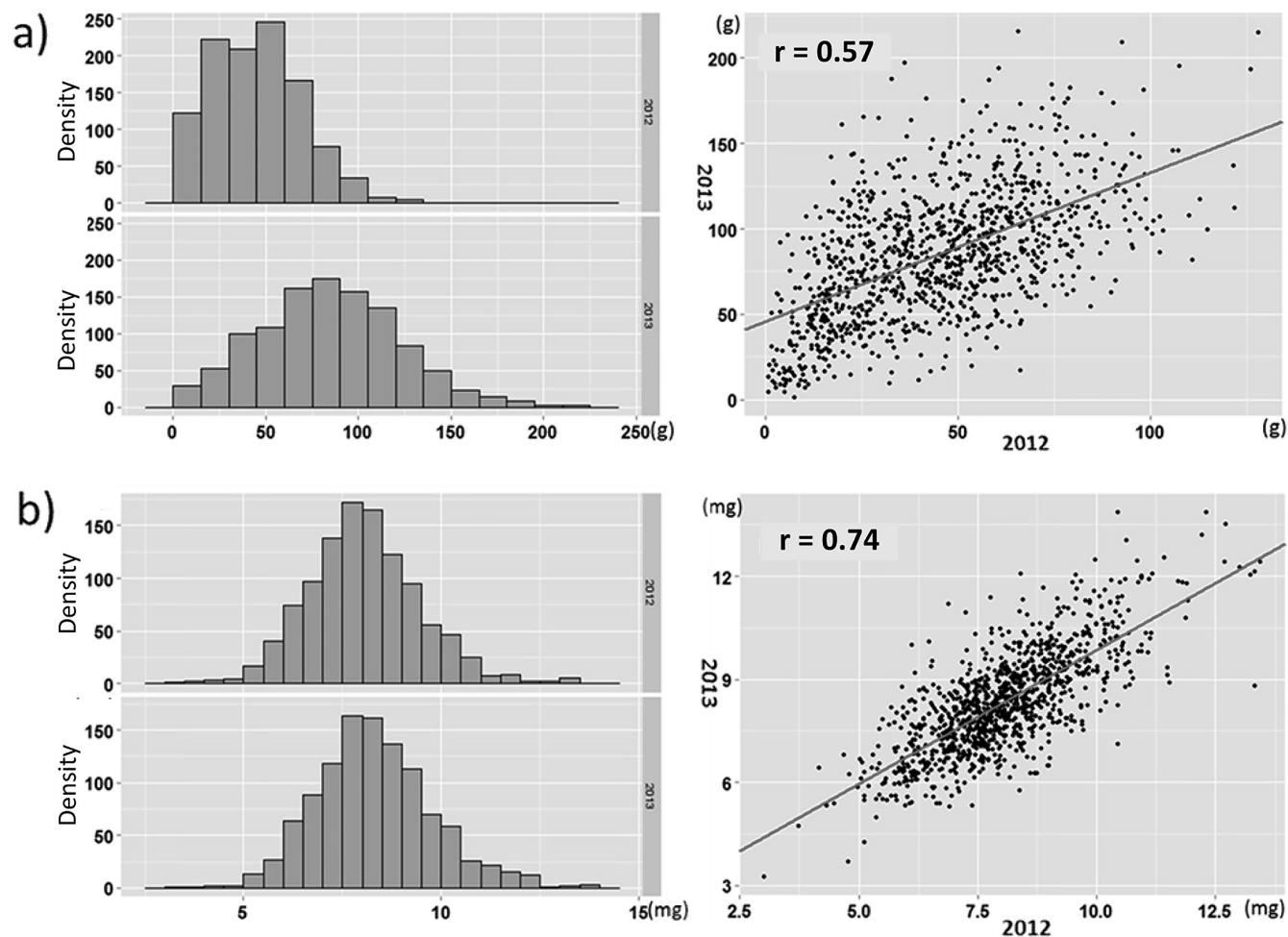
Figure 2. (a) Grain yield and (b) seed weight of the genomic selection population with 1126 genets from 58 half-sib families in 2012 and 2013. The mean of grain yield in 2012 was 43.9 g, and 83.6 g in 2013. For seed weight, the mean in 2012 was 8.0 mg, and 8.3 mg in 2013.

Table 1. Variation of seven agronomic traits in the breeding population of intermediate wheatgrass in two growth seasons, 2011–2012 and 2012–2013.

Trait	Median	Mean	Range	V_{family}[§]	$V_{genets/Family}$[§]	Total V_G[§]	V_E[§]	Heritability (h^2)
Biomass, g	689.20	685.20	187.32–1222.64	3266.44	11,992.00	15,090.00	13,745.00	0.69
Grain yield, g	84.49	84.59	22.62–191.37	140.52	385.89	521.46	499.89	0.68
Head weight, g	1.02	1.02	0.13–1.91	0.01	0.04	0.04	0.03	0.74
Seed weight, mg	8.21	8.32	3.26–13.25	0.40	1.13	1.56	0.54	0.85
Heading score[†]	4.47	4.34	1.47–6.47	0.19	0.35	0.56	0.77	0.59
Height, cm	138.84	137.67	55.81–179.32	54.76	123.57	176.84	110.26	0.76
Threshability[‡]	2.19	2.10	1.19–4.19	0.15	0.43	0.58	0.39	0.75

[†]Heading score was recorded with score 1–6 (1 is latest, 6 is earliest) when approximately 20% plants have fully emerged seed heads.

[‡]Threshability was scored as 1 (low) to 4 based (high) on the percentage of naked seeds by visual observation.

[§]Vfamily, genetic variance among families; Vgenets/family, genetic variance within family; Total VG, total genetic variance among genets; VE, the variance of residue. The calculation of genetic variances was performed by using the mixed procedure in SAS (v. 9.3) by fitting years as fixed and genets or families as random effects.

We observed that heading score was independent of head weight, seed weight, and height ($p > 0.39$), and threshability had no significant correlation with grain yield and height (p value is 0.07 and 0.68, respectively).

Genotyping-by-Sequencing Performance in IWG

The GBS protocol described by Poland et al. (2012) was optimized for IWG. First, two pairs of restriction enzymes were tested, *Pst*I/*Msp*I and *Pst*I/*Ape*KI. With a similar quantity of pass filtered reads, the library from *Pst*I/*Msp*I had a larger number of reads with good
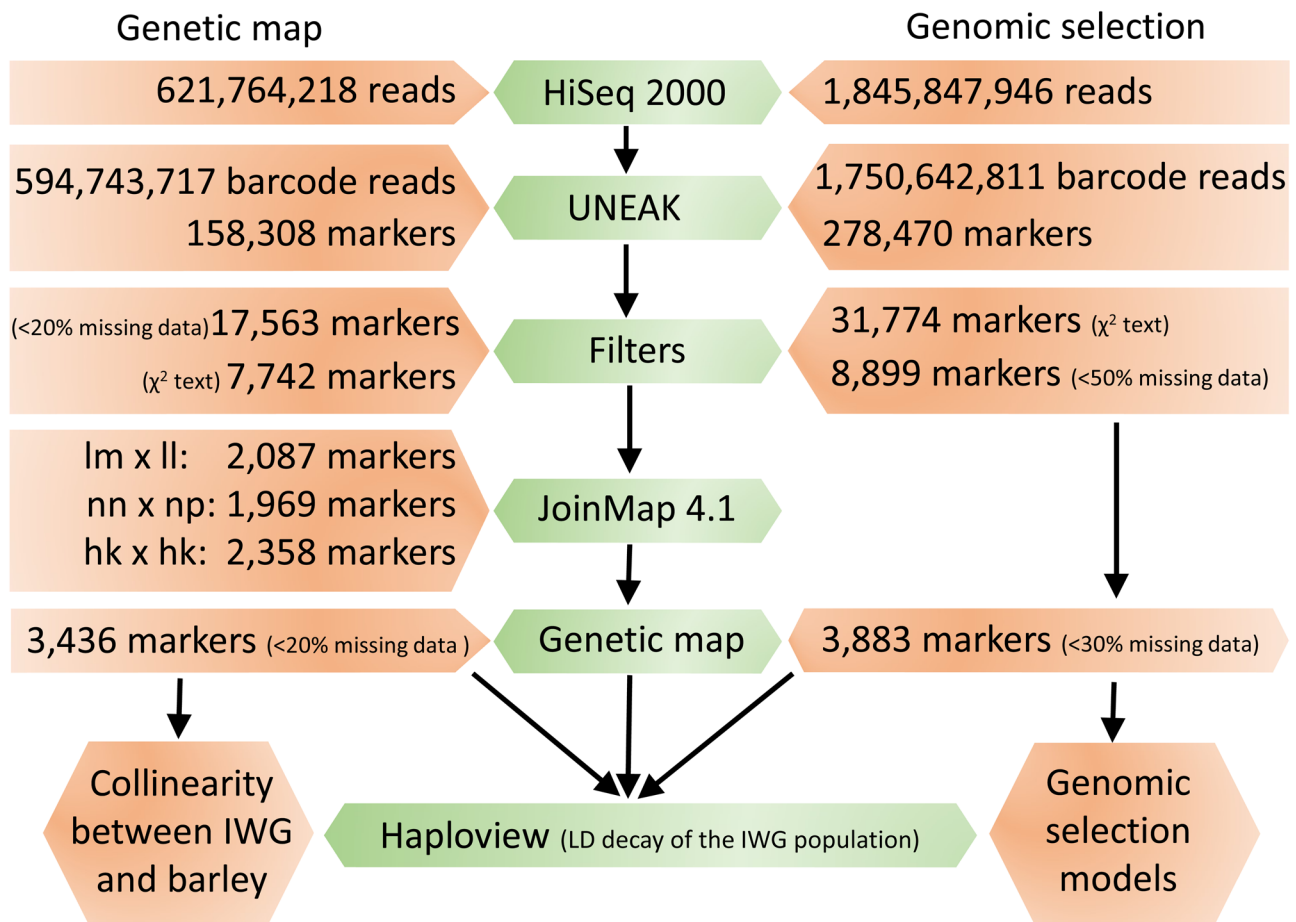
Genetic map

| 621,764,218 reads | HiSeq 2000 | 1,845,847,946 reads |

Genomic selection

594,743,717 barcode reads
158,308 markers

UNEAK

1,750,642,811 barcode reads
278,470 markers

(<20% missing data) 17,563 markers
(χ^2 text) 7,742 markers

Filters

31,774 markers (χ^2 text)
8,899 markers (<50% missing data)

lm x ll: 2,087 markers
nn x np: 1,969 markers
hk x hk: 2,358 markers

JoinMap 4.1

3,436 markers (<20% missing data)

Genetic map

3,883 markers (<30% missing data)

Collinearity between IWG and barley

Haploview (LD decay of the IWG population)

Genomic selection models

Figure 3. Flow diagram of genotypic data analysis pipeline. 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 $p > 0.01$ were kept for subsequent analysis.

barcodes, fewer tags per sample, but more reads per tag (2.60 vs. 2.15) than the library from *Pst*I/*Ape*KI (Supplementary Fig. S5a). Thus, the enzymes *Pst*I/*Msp*I were used to develop GBS libraries in the present study. Second, two barcode adaptors were used for each genet. Due to the nucleotide composition and length differences among barcode adaptors, ligation, amplification, or sequencing biases might exist. When two barcodes were used as the identifiers of one genet, the variation of read number per sample in individual GBS libraries was greatly reduced (Supplementary Fig. S5b). Third, size selection was used to remove barcode adaptor dimers. To guarantee the adequate ligation between adaptor and genome DNA, we used excessive amounts of both barcodes and common adaptors (0.5 and 15 pmol for the reaction of 50 μL, respectively). We then trimmed off the adaptor dimers by performing size selection using LabChip GX.

In general, about 150 million reads were obtained from one GBS library, one lane of HiSeq 2000. More than 95% of the reads contained good barcodes, except for one library of the breeding population (82.4%; Supplementary Fig. S6). The read number and tag number per sample were calculated, and small variations were observed among both IWG genets and GBS libraries (Supplementary Fig. S6). The four libraries of the F_1 population, with 48-plex sequencing, had twice the read number per sample, compared with 96-plex libraries (3,160,000 vs. 1,550,000 on average). But the sequence coverage (read number per tag) of 48-plex libraries was not twice as many as 96-plex libraries (3.7 vs. 2.7) because the 48-plex libraries had a much larger number of tags per sample than the 96-plex libraries (855,000 vs. 564,000; Supplementary Fig. S6).

Genotype Calling, Filtering, and Quality Control

After running the UNEAK pipeline, we obtained 158,308 putative SNPs from the F_1 population (Fig. 3). Putative homozygous markers whose sequencing count was less than five were set to missing, and markers with >20% missing data were filtered. This resulted in 17,563 markers. After running a goodness-of-fit χ^2 test ($p > 0.05$) for the two alleles in heterozygotes, we kept 7742 markers for genetic mapping (Fig. 3). The markers were converted to three classes of genotypes, lm × ll, nn × np, and hk × hk for JoinMap 4.1. The average sequence coverage for these markers was 12 reads per marker for individual genets.
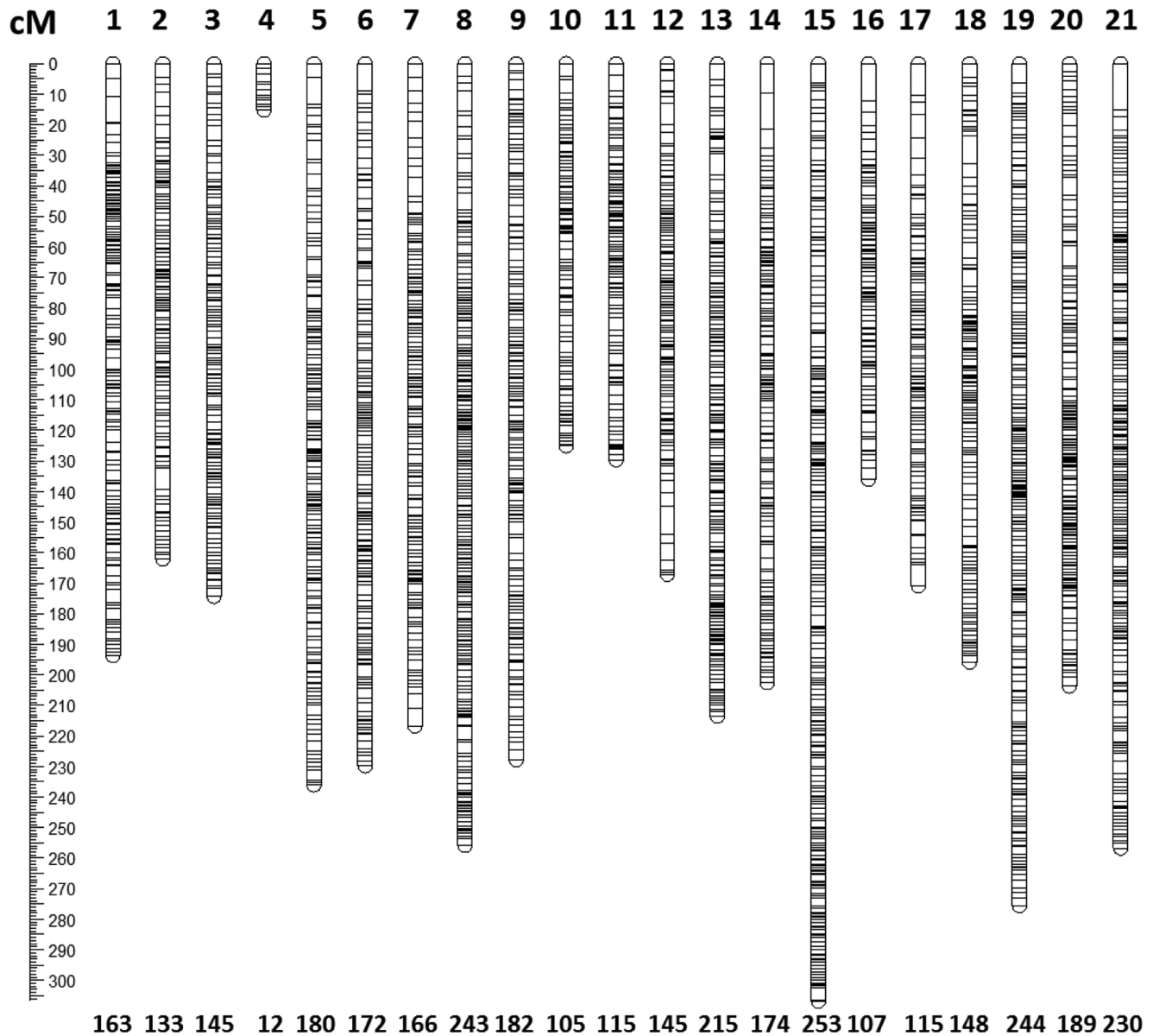
Figure 4. Genetic map of intermediate wheatgrass from an F$_1$ full-sib population with 178 genets. In total, 3436 markers with <20% missing data were ordered in the genetic map. The 21 linkage groups showed good uniformity of marker distribution. The mean distance between adjacent SNPs was 1.22 cM (SD = 1.08). The number of markers in each linkage group was listed.

From the IWG population for genomic selection, 278,470 markers were discovered after running the UNEAK pipeline (Fig. 3). The threshold of a goodness-of-fit χ^2 test for the two alleles of heterozygotes was set with $p = 1 \times 10^{-5}$. The homozygotes with less than four sequencing counts or reads were considered as missing data. After running these two filters, we obtained 8899 markers with <50% missing data and 3883 markers with <30% missing data for further genetic analysis and genomic selection (Fig. 3). The average sequence coverage for markers with <50% missing data is 5.2 reads per marker for individual genets, and 6.4 for markers with <30% missing data.

Genetic Map of IWG

Compared with some other marker platforms such as simple sequence repeat, diversity arrays technology, and SNP arrays, GBS produced markers with large amounts of missing data. These missing data might introduce errors to the order of markers in the genetic map. To decrease the effect of missing data, we first used the markers with <3% missing data to develop a genetic map. A total of 1119 markers were mapped into 21 LG, with the number of markers in each LG ranging from 10 to 94. Using these markers as the fixed order, 2657 markers with <10% missing data were mapped into the 21 LG. Finally, 3436 markers with <20% missing data were integrated into the genetic map. Except for LG4 with 12 markers, the number of markers in individual LG ranged from 105 to 253 (Fig. 4). The length of LG varied between
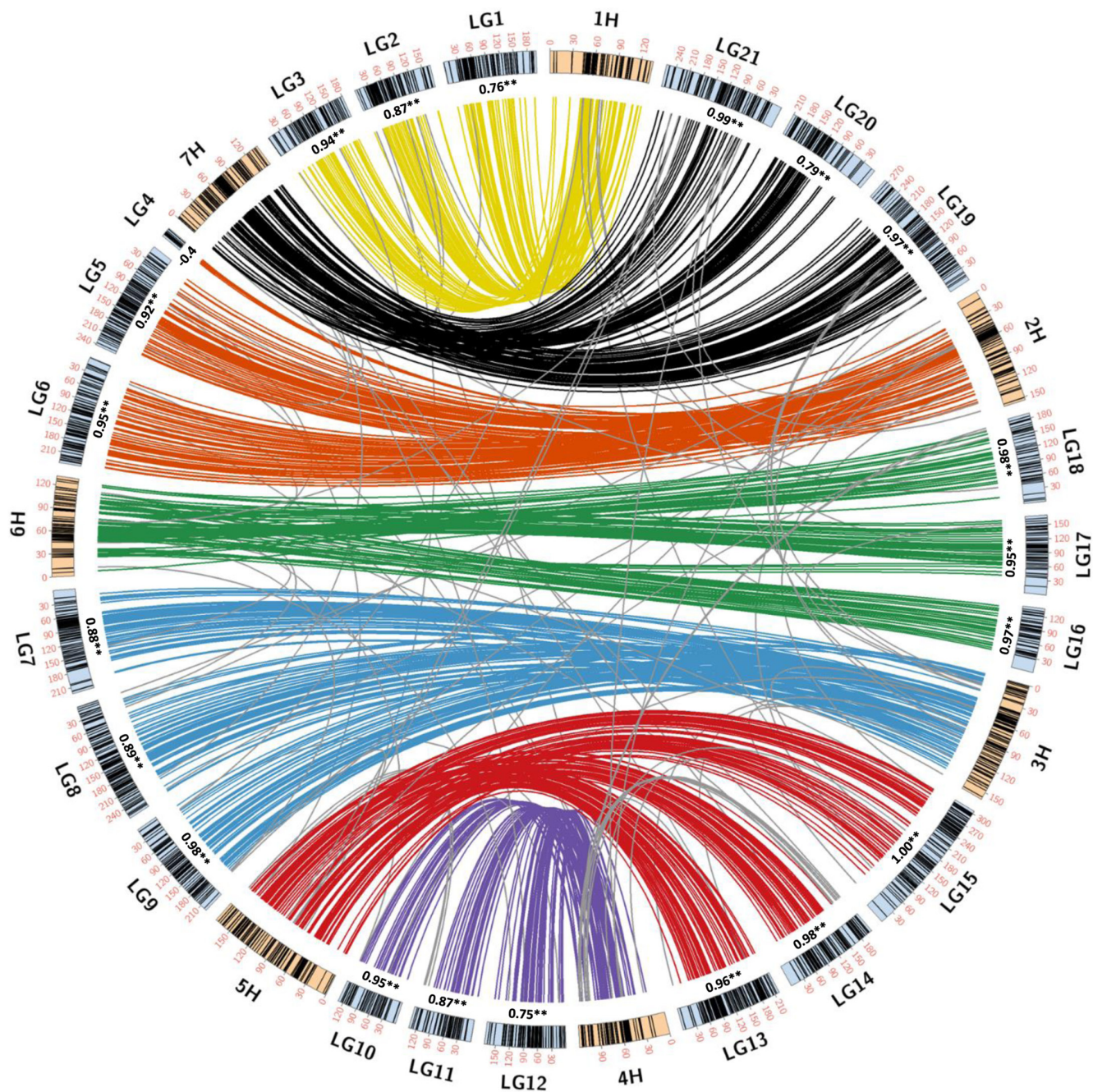
Figure 5. The collinearity of linkage groups of intermediate wheatgrass with barley chromosomes. The linkage groups were designated as LG1 to LG21, with LG1–3 corresponding to *1H*, LG4–6 to *2H*, and so on. The Spearman rank correlation coefficients (ρ) were calculated using R package 'pspearman'. ** Indicates significance at *p* < 0.01.

125 cM for LG10 and 306 cM for LG15, for a total map length of 4095 cM, which was probably expanded because of missing data. The only exception was LG4 which was 15 cM. The 21 LG showed good uniformity of marker distribution with the largest distance between adjacent markers being 15.2 cM, and only 10 intervals were larger than 8 cM. The mean distance between adjacent SNPs was 1.22 cM (SD = 1.08).

To determine the correspondence between LG and chromosomes, we performed collinearity analysis between IWG LG and reference physical or genetic maps in barley and *A. tauschii* (Jia et al., 2013; Mayer et

al., 2012). In total, 922 markers on the genetic map were uniquely matched to the barley genome. The Spearman's rank correlation coefficients (ρ) were >0.75 (*p* < 0.01) except for LG4, which indicated that IWG LG showed high synteny and collinearity with barley as three homeologous IWG LG corresponded to one barley chromosome (Fig. 5). Because a physical map of *A. tauschii* is not available, the genetic map with 13,687 scaffolds was used to perform collinearity analysis between IWG and *A. tauschii*. A total of 794 markers were shared between IWG and *A. tauschii* genetic maps. The collinearity results with *A. tauschii* were consistent with those with barley
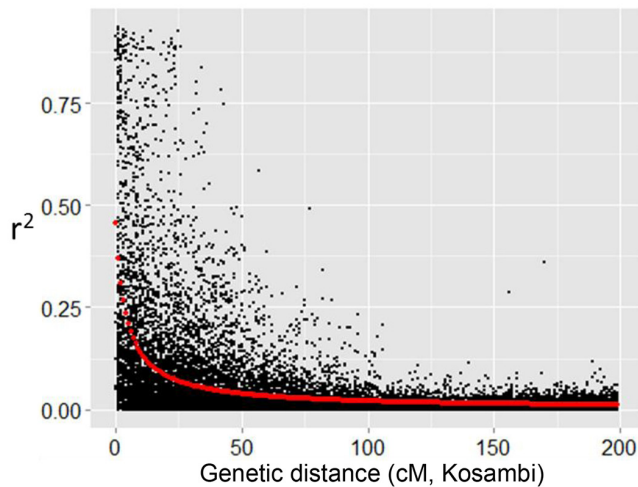
Figure 6. The rate of linkage disequilibrium (LD) decay of the genomic selection population of intermediate wheatgrass. The LD among markers in the population was estimated using Haploview (Barrett et al., 2005). The Hill and Weir formula (Hill and Weir, 1988) was used to describe the LD decay of r^2.

in that three homeologous IWG LG matched with one *A. tauschii* chromosome (Supplementary Fig. S7). Thus, the 21 LG were divided into seven subgroups. Each subgroup contained three LG, corresponding to one chromosome in barley or *A. tauschii*. Because the genetic donors of the subgenomes of IWG are still unknown, the LG could not be assigned to subgenomes. Thus, the LG were designated as LG1 to LG21, with LG1-3 corresponding to *Triticeae* homologous Group 1, LG4-6 to Group 2, and so forth.

Establishment and Optimization of Genomic Selection for IWG

Among 3883 markers (<30% missing data) from the breeding population, 1158 were shared with the markers in the genetic map. These shared markers were used to measure the LD of the breeding population. The LD decays such that, at 0.2, conventional r^2 is equal to an average distance of 5 cM (r^2 at 0.1 to 15 cM; Fig. 6). The relatively low rate of LD decay was consistent with small effective population size and limited recombination events (Fig. 1).

Two important agronomic traits, grain yield and seed weight, were used to test the effects of nine genomic prediction models on predictive ability. For seed weight, Random Forest had significantly lower predictive ability ($p < 0.01$) and the other eight models performed similarly to each other with respect to predictive ability ($p > 0.54$; Supplementary Fig. S8a). For grain yield, the nonlinear model, GAUSS, which can capture both additive and nonadditive effects of markers, showed significantly higher predictive ability than the other models ($p < 0.01$; Supplementary Fig. S8b). RR-BLUP has significantly lower predictive ability than the Bayesian models ($p < 0.01$). Subsequent comparisons of predictive ability were based on the GAUSS model.

GBS technology produced markers with high levels of missing data. To determine the amount and quality of

markers used in subsequent analysis, we used all markers with <5, 10, 15, 30, and 50% missing data and the GAUSS model to estimate the predictive ability on grain yield and seed weight. The numbers of markers with <30 and 50% missing data (3883 and 8899, respectively) were much larger than the other groups of markers, and both groups of markers showed the largest predictive ability for both grain yield and seed weight (Supplementary Fig. S9). But there was no difference in predictive ability between the two groups. Thus, the markers with <30% missing data were used for subsequent evaluation of prediction models.

Using all 1126 genets in the breeding population and RRSV, moderate to high predictive ability was obtained for all seven agronomic traits. The highest predictive ability was observed for seed weight, average $r = 0.67$. Biomass and head weight had moderate predictive ability, $r = 0.46$ and 0.47, respectively. Grain yield, height, heading score, and threshing also had high predictive ability, $r = 0.52$ to 0.62 (Fig. 7a).

Fifty-eight IWG families showed extensive variability for all seven traits (Supplementary Fig. S4). Family structure might exist, and in turn affect the predictive ability (Fig. 8a). When the number of families involved in the training set ranged from 5 to 30, larger number of families resulted in higher predictive ability. However, if >30 families were used as the training set, the ability to predict the breeding value of genets in the other families was not improved for either grain yield or seed weight, and the variation of the predictive ability increased. Moreover, the predictive ability was never higher than that of RRSV (Fig. 8a). All 58 families need to be sampled as the training set to achieve as high predictive ability as RRSV.

If all 58 families are included in the training set the number of genets from each family should be optimized. Because the smallest family only had 11 genets, the sampling number of genets per family varied from 1 to 10. In general, the training sets with large number of genets per family had high predictive ability. For grain yield, the predictive ability reached a plateau at $n = 8$ genets per family (Fig. 8b). For seed weight, when 10 genets were sampled from each family, the predictive ability was similar with that of RRSV ($p = 0.94$). The training sets containing 8 to 10 genets from each IWG family are sufficient to achieve high predictive ability.

The genomic selection population used in this study was primarily composted of genets arranged in two blocks (A and B) in the field that were generated from random sampling when planted (Supplementary Fig. S1). In each block, there were about 10 genets from each family. Using Block A or B as the training set and the others as the validation set, the predictive ability for the seven traits was calculated. The results were quite similar with those from RRSV, with the highest predictive ability, 0.66, for seed weight, and the lowest value, 0.42, for biomass and head weight (Fig. 7b). This also confirmed that 8 to 10 genets from each family form the optimized training population for genomic selection, and it is feasible to use a subset of breeding materials as a training
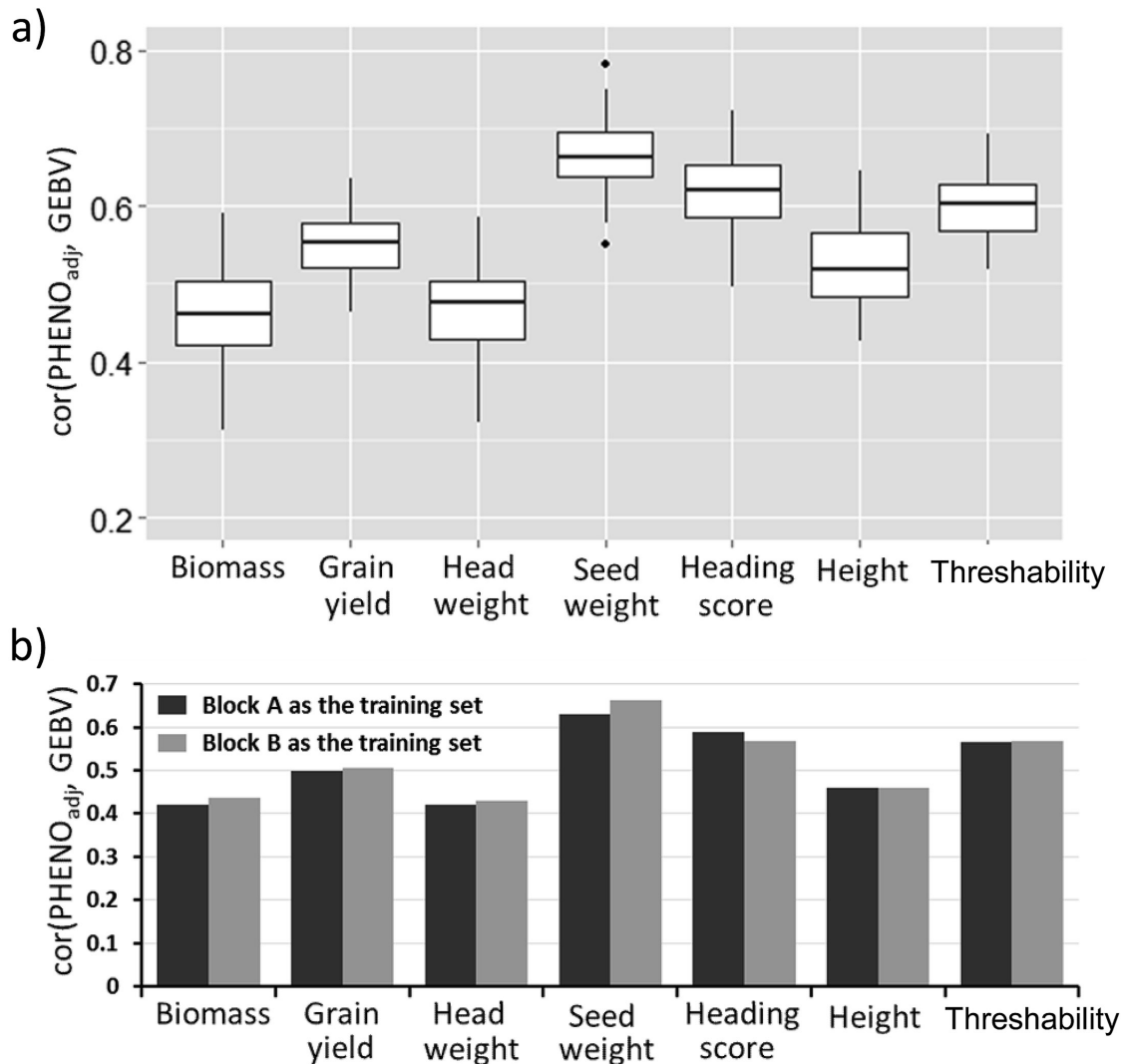
Figure 7. Predictive ability of seven agronomic traits using the nonlinear GAUSS model. (a) The predictive ability from repeated random subsampling validation. Eighty percent of the genets were randomly selected and used as the training set and the others were used as the validation set to assess the predictive ability. The sampling process without replacement was repeated 100 times. (b) The predictive ability using genets in Block A or Block B (see Supplemental Fig. S1) as the training set and the other genets as the validation set. cor is the correlation between the genomic estimated breeding values (GEBV) and the PHENO$_{adj}$ (the observed phenotypic data, best linear unbiased estimate).

population to develop a prediction model and then calculate the breeding values of their siblings.

To determine the predictive ability for genets within a half-sib population, we used the five families (i.e., C3-3471, W104_4, W111_4, C3–2925, and C3-2331), each of which has more than 20 genets in Blocks B, C, and D. The models trained with genets in Block A were used to estimate the breeding value of genes in the five families. We obtained different predictive ability across families and traits (Supplementary Fig. S10). For example, for seed weight, the predictive ability ranged from 0.50 to 0.69 within families, C3–3471, W111_4, C3–2925, and C3–2331, but the predictive ability was 0.14 within W104_4. For height, however, the predictive ability within all families was <0.22. These data indicated that although the genomic selection models trained with genets in Block A could be used to prediction the remaining

of the genomic selection population (Fig. 7b), the models might not work well for every family when used to predict the breeding value of genets within a specific half-sib family (Supplementary Fig. S10).

To estimate the marker density for future application of genomic selection in IWG breeding programs, we randomly sampled 40, 200, 400, 800, 1600, 2000, 2400, and 3200 markers from 3883 markers with <30% missing data, and genets in Block A were used as the training set and the other genets as the validation set (Fig. 8c). When the marker numbers were larger than 1600, the predictive ability reached a plateau for both grain yield and seed weight ($p > 0.10$).
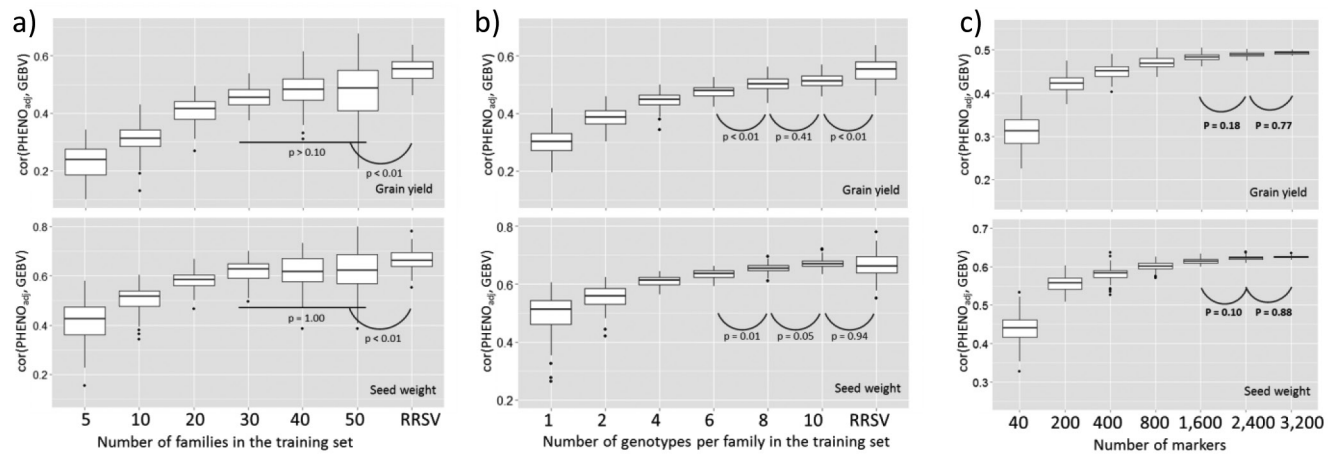
Figure 8. Optimization of the composition of the training population and marker density for the application of genomic selection in intermediate wheatgrass. (a) The predictive ability of grain yield and seed weight using a different number of families as the training set. For each specific number of families, samples were drawn without replacement 100 times; RRSV, Repeated random subsampling validation with 80% of all genets as the training set and the others as the validation set, was used as the positive control. The nonlinear model GUASS was used to develop the prediction equation. (b) The predictive ability of grain yield and seed weight using a range of genets from each of 58 families as the training set. For each number of genets, samples were drawn without replacement 100 times; (c) The predictive ability using different numbers of markers, randomly sampled from 3883 markers with less 30% missing data. The genets in Block A were used as the training set and the other genets as the validation set. cor is the correlation between the genomic estimated breeding values (GEBV) and the $PHENO_{adj}$ (the observed phenotypic data, best linear unbiased estimate).

Discussion

Efficiency of Genotyping-by-Sequencing in IWG

With the advent of next-generation sequencing, several marker technologies are now capable of discovering thousands of markers across almost any genome of interest, even for species in which little or no genetic information is available (Davey et al., 2011). As one of the most popular and powerful sequencing-based technologies for genotyping, GBS has been used to discover genome-wide markers in many species, including wheat, barley, potato, maize, and switchgrass (Glaubitz et al., 2014). In the present study, we used GBS to discover thousands of SNP markers across the genome of IWG, which is an allohexaploid and outbred species without reference genome. To perform GBS, appropriate enzymes must be chosen for the target species. We compared the sequencing results from two pairs of enzymes, *Pst*I/*Msp*I and *Pst*I/*Ape*KI. The methylation-sensitive *Ape*KI used in maize might not be appropriate for IWG because *Pst*I/*Ape*KI produced fewer reads with good barcodes and had fewer number of reads per tag than *Pst*I/*Msp*I. The enzymes pair *Pst*I/*Msp*I were also successfully used in wheat, barley, oat, and pearl millet [*Pennisetum glaucum* (L.) R. Br.; Huang et al., 2014; Moumouni et al., 2015; Poland et al., 2012] to discover SNP markers across genomes. Thus, the *Pst*I/*Msp*I may generally be a good choice when performing GBS in cereal crops. Two barcode adapters for individual samples were tested in the present study. Compared with the protocol using one barcode adaptor, the two-barcode protocol produced sequencing data with less read count variation among samples (Fig. 7b), which suggested that the bias from nucleotide composition and

length difference was decreased. Different from standard GBS protocols (Poland et al., 2012), size selection was included using LabChip GX and helped to remove the adaptor dimers, which affect the production of pass-filter reads. In the standard GBS protocol, to avoid sequencing the adaptor dimers (>0.5%), the amount of barcode adaptors has to be determined by performing adaptor titrations. And for each particular enzyme and species combination, the appropriate quantity of adapters has to be empirically determined (Elshire et al., 2011). The size selection can remove the adaptor dimers and the adaptor titrations can be omitted. Compared with standard GBS protocol (Poland et al., 2012), about five times larger amounts of barcode adaptors were used in the present study, and the efficiency of ligation between genome DNA and barcode adaptors was increased.

UNEAK is a powerful pipeline to discover GBS markers from polyploid species without a reference genome. A network filter was employed in UNEAK to remove false SNPs derived from repeats, paralogs, and tags with sequencing errors (Lu et al., 2013). Because of the low sequencing coverage of GBS and the polyploidy and heterozygosity of IWG, three more filters were required to filter out the false SNPs derived from the sampling errors during sequencing. First, some heterozygotes were miscalled as homozygotes because the low sequencing coverage of GBS leads to high amounts of missing data. A cutoff should be set to convert the miscalled homozygotes to missing data. In the present study, for the mapping population, false homozygotes can greatly affect the accuracy of marker order. We considered the homozygotes, the sequencing count number of whose alleles was less than five, as missing data. For the genomic selection population,

we set the cutoff of the count number as four, due to the high multiplex (96-plex) sequencing and high error tolerance of genomic selection (Rutkoski et al., 2013). The second filter was based on the hypothesis that, in diploid species, the tag counts of two alleles of the heterozygous loci should be equal. Using this filter, we discarded 56% of 17,563 SNPs with <20% missing data from the F_1 population. In switchgrass, after running this filter, the remaining markers of the F_1 full-sib population showed good Mendelian segregation (Lu et al., 2013). Thus, this is an important filter for SNP calling in outbred species. The third filter for the quality control of GBS markers is the amount of missing data. The threshold of the percentage of missing data depends on the study goals and the level of error tolerance of the study. For genetic mapping, we used markers with <20% missing data. Using the order of markers with <3% missing data first and progressively adding those with 10% missing data as the fixed order, we obtained a genetic map that showed high collinearity with barley and *A. tauschii*. The threshold of 20% missing data was also widely used in the genetic mapping in wheat, barley, apple, pearl millet, and cassava (*Manihot esculenta* Crantz; ICGMC, 2015; Gardner et al., 2014; Moumouni et al., 2015; Poland et al., 2012). For genomic selection, we used markers with <30% missing data. After the imputation with Random Forest, the prediction abilities from markers with 30 and 50% missing data were similar in the IWG population. Rutkoski et al. (2013) also reported that genomic selection prediction accuracy from markers with 20 and 50% missing data were similar, and a large proportion of missing data in dense marker sets was not a major concern for genomic selection.

Distorted Segregation of Markers in IWG

In the present study, we developed the first genetic map of IWG. The markers in 21 LG showed high collinearity with seven barley and *A. tauschii* chromosomes, which verified the high quality of the genetic map and indicated the close relationship between IWG subgenomes and barley and *A. tauschii* genomes. Low marker coverage for LG4 was observed, which is homeologous to *2H* in barley and *2D* in *A. tauschii*. To obtain more markers of this linkage group, we used all the filtered markers uniquely matched to chromosome *2H* to develop the genetic map. Unfortunately, no more markers were obtained for LG4. We observed that these markers showed significant segregation distortion, that all markers are heterozygous in both parents, and >90% of the progeny are heterozygous for all the markers (data not shown). Some loci on LG4 might be related to inbreeding depression of IWG or embryo lethality of homozygotes, and prevent recombination or recovery of homozygous alleles. The lack of recombination among the loci might be the main reason that we were not able to obtain more markers for LG4 using the markers matched to chromosome *2H*. The effects of these markers of LG4 on agronomic traits will be studied when QTL mapping is performed in the near future. Five and four

of 12 markers in LG4 were uniquely matched to *2H* and *2D* chromosomes, respectively. Another linkage group, LG5 corresponding to *2H*, also has many markers showing segregation distortion. But the markers did not show as high a level of segregation distortion as markers on LG4. The other linkage group corresponding to *2H*, LG6, has few markers showing distorted segregation. Thus, we propose that LG4 is an independent linkage group, rather than a part of LG5 or LG6. Segregation distortion was also widely observed in many crop species including barley, rice, maize, wheat, and pearl millet (Moumouni et al., 2015; Pan et al., 2012; Peleg et al., 2008). Thus, we suggest keeping markers with distorted segregation for genetic mapping if three filters were applied after SNP calling: filtering false homozygotes, filtering heterozygotes with uneven count numbers of alleles, and filtering markers with a high percentage of missing data.

Optimization and Application of Genomic Selection in IWG Breeding Programs

In the present study, nine statistical models were tested. RR-BLUP and Bayesian models are linear models, but their assumptions about marker effects are different (Endelman, 2011; Perez and de los Campos, 2014). Both GAUSS and RKHS are nonlinear models, which can capture both additive and nonadditive effects. Random Forest is a machine-learning nonparametric model with the capacity to capture nonadditive sources of genetic variability, including dominance and epistasis (Desta and Ortiz, 2014). All these models, except for Random Forest, performed similarly in seed weight, but the nonlinear GAUSS model performed better than the other models in grain yield. A possible reason might be that all the genets in each family are half-siblings, and the epistatic interactions are relatively stable among siblings. Moreover, the prediction algorithm of the GAUSS model might fit the family structure of the IWG population better than the other two nonlinear models, RKHS and Random Forest. The performance of different models should be tested when used to estimate the breeding values of genets across generations in IWG breeding programs.

In the present study, a low rate of LD decay in the breeding population was observed. This observation would not be expected for an outbred species, which usually has rapid LD decay because of effective recombination in each generation of outcrossing. However, the IWG population studies result from a recent and strong selection history along with small effective population size. The IWG breeding population was derived from an initial selection of only 14 genets from the Rodale Research Centre, PA, and eight new genets introduced by TLI, KS. In addition, five cycles of strong selection have been executed (Fig. 1). Thus, the effective population size is small (~20), and the limited breeding cycles result in low recombination rate of the breeding population. The small effective population size and strong selection of limited breeding cycles are likely the main reasons for low rate of LD decay in the current breeding population. Similarly, when

applying genomic selection for outbred species, such as ryegrass, eucalyptus, and apple, artificially minimizing the effective population size was used to decrease the rate of LD decay and, in turn, improve the prediction accuracy (Kumar et al., 2012; Lin et al., 2014). The low rate of LD decay resulted in a small number of markers needed to capture the genome-wide variations. Our results also showed that 1600 markers were adequate to achieve high predictive ability. Thus, in future applications of genomic selection, GBS with 192-plex for HiSeq 2000 or 384-plex for HiSeq 2500 should be feasible to genotype thousands of IWG selection candidates.

The effect of the size and composition of the training population on predictive ability was also tested. If the genets in the training set and the validation set were from different families, the predictive ability was significantly lower than that of RRSV. This result confirmed that the genetic relationship between the training and validation sets affects predictive ability (Asoro et al., 2011; Cros et al., 2015; Hickey et al., 2014; Wientjes et al., 2013; Windhausen et al., 2012). Thus, the genets in the training set should be sampled from all families in an IWG breeding population. Eight to 10 genets from each half-sib family would form a good training population to predict the breeding value of their siblings. These results were verified by the independent validation where a high predictive ability was obtained when we used the genomic selection models trained with genets in Block A to predict the remaining genets of the genomic selection population (Fig. 7b). When using these models to predict the genets within individual families (i.e., C3-3471, W104_4, W111_4, C3–2925, and C3-2331), however, we obtained different predictive ability across families and traits (Supplementary Fig. S10). For the prediction of genets within a family, the predictive ability is mainly affected by the consistency of linkage phases within and across families (Lian et al., 2014, 2015). To achieve high predictive ability within a family, the structure and composition of the training population, and the LD within a family and among families should be investigated in a future study.

Cross validation demonstrated that moderate to high predictive ability (0.46 to 0.67) was obtained in the population of genets from the same generation. Moreover, the genomic selection model from subpopulations can be used to predict the breeding value of the other genets of the population. Thus, the genomic selection model trained with a subset of genets in a breeding population can be used to estimate the breeding value of the remaining of the breeding population (Fig. 7). In practice, the use of genomic selection in IWG breeding programs or any other species also depends on the ratio of cost to benefit. Genotyping a very large breeding population may not be cost effective, although genotyping costs are decreasing very rapidly. In conventional recurrent selection, the best genets are selected from the best families based on phenotypic data, which is an effective breeding method for improving perennial grasses (Vogel, 2013).

This selection principle can also be integrated into the genomic selection-based recurrent selection (Supplementary Fig. S11). Based on phenotypic data of the subset of genets from the breeding population, the best families (e.g., 25%) will be determined. The genets from these best families will be genotyped using GBS, and their breeding value will be estimated using the genomic selection model. The best genets from these best families will be selected and used for crossing for the next recurrent selection cycle in the greenhouse during the winter (Supplementary Fig. S11). We would not suggest to make selection within a specific family because of the variable predictive ability within a family. As an example of genomic selection based breeding scheme, an IWG breeding population with 10,000 genets (i.e., 100 genets from each of 100 original families) would all be evaluated under field conditions using only phenotypic selection. Here, only a subset of the breeding population (~1,000 genets, 10 genets per family; Fig. 8b) needs to be planted and managed in the field. And based on the phenotypic data of the subpopulation, we discard about 70% of families because of their poor performance. Only 30% of families (<3000 genets, 100 genets from each of 30 families) are actually genotyped using GBS. Substantial resources required for planting, weeding, harvesting, threshing, phenotyping, and genotyping will be saved, and the capacity of a breeding group to screen breeding genets will be greatly increased. Compared with the current breeding program (Fig. 1), the genomic selection-based recurrent selection has large genetic variations (10,000 vs. 2,000 genets) and high selection intensity (1 vs. 4%) and, in turn, gives a very high expected gain (Supplementary Fig. S11).

In the present study, genomic selection worked well to predict the breeding value of half-siblings. The marker density and the size and composition of the training population were optimized. A breeding scheme based on genomic selection was proposed and is being used in our breeding program. We are now studying the application of genomic selection between generations and across environments. After collecting the phenotypic data from the genets in the second recurrent selection cycle, we will use these genets as the validation population. The performance of the genomic selection model trained with genotypes from the parent generation (first cycle) in the present study will be tested in predicting the breeding values of the progeny generation (second cycle).

Supplemental Information Available

Supplemental information is included with this article.

Supplemental Figure S1. Organization of the IWG breeding population (UMN-C1) in the field in St. Paul.

Supplemental Figure S2. Temperature and precipitation in the two growth seasons, 2011-2012 and 2012-2013, in St. Paul.

Supplemental Figure S3. The biomass, height, head weight, heading score, and threshability of the genomic selection population with 1,126 genets from 58 half-sib families in 2012 and 2013.

Supplemental Figure S4. Trait variation among 58 families of intermediate wheatgrass in the genomic selection population with 1,126 genets.

Supplemental Figure S5. Optimization of genotyping-by-sequencing protocol for intermediate wheatgrass.

Supplemental Figure S6. The quality of 16 genotyping-by-sequencing libraries or lanes.

Supplemental Figure S7. The collinearity of linkage groups of intermediate wheatgrass with *Aegilops tauschii* chromosomes.

Supplemental Figure S8. Comparison of nine statistical models for genomic selection.

Supplemental Figure S9. The predictive ability obtained using the nonlinear GAUSS model and markers with 5, 10, 15, 30, or 50% missing data.

Supplemental Figure S10. Predictive ability on seven agronomic traits within individual half-sib families.

Supplemental Figure S11. Genomic selection-based breeding scheme.

Supplemental Table S1. The genetic correlation among seven agronomic traits of intermediate wheatgrass in two growth seasons, 2011–2012 and 2012–2013.

References

Asoro, F.G., M.A. Newell, W.D. Beavis, M.P. Scott, and J.L. Jannink. 2011. Accuracy and training population design for genomic selection on quantitative traits in elite North American oats. Plant Gen. 4:132–144. doi:10.3835/plantgenome2011.02.0007

Bao, Y., T. Vuong, C. Meinhardt, P. Tiffin, R. Denny, S.Y. Chen, H.T. Nguyen, J.H. Orf, and N.D. Young. 2014. Potential of association mapping and genomic selection to explore PI 88788 derived soybean cyst nematode resistance. Plant Gen. 7. doi:10.3835/plantgenome2013.11.0039

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

Beeby, A., and A. Brennan. 2008. First ecology: Ecological principles and environmental issues. 3rd ed. Oxford Univ. Press, UK.

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

Cros, D., M. Denis, L. Sanchez, B. Cochard, A. Flori, T. Durand-Gasselin, B. Nouy, A. Omore, V. Pomies, V. Riou, E. Suryana, and J.M. Bouvet. 2015. Genomic selection prediction accuracy in a perennial crop: Case study of oil palm (*Elaeis guineensis* Jacq.). Theor. Appl. Genet. 128:397–410. doi:10.1007/s00122-014-2439-z

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

de Roos, A.P.W., B.J. Hayes, and M.E. Goddard. 2009. Reliability of genomic predictions across multiple populations. Genetics 183:1545–1553. doi:10.1534/genetics.109.104935

DeHaan, L.R., S. Wang, S. Larson, T. Kantarski, X. Zhang, and D. Cattani. 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. p. 72–89.

Desta, Z.A., and R. Ortiz. 2014. Genomic selection: Genome-wide prediction in plant improvement. Trends Plant Sci. 19:592–601. doi:10.1016/j.tplants.2014.05.006

Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, and S.E. Mitchell. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6. doi:10.1371/journal.pone.0019379

Endelman, J.B. 2011. Ridge regression and other kernels for genomic selection with R package rrBLUP. Plant Gen. 4:250–255. doi:10.3835/plantgenome2011.08.0024

Gardner, K.M., P. Brown, T.F. Cooke, S. Cann, F. Costa, C. Bustamante, R. Velasco, M. Troggio, and S. Myles. 2014. Fast and cost-effective genetic mapping in apple using next-generation sequencing. G3: Genes Genomes Genet. 4:1681–1687. doi:10.1534/g3.114.011023

Gaut, B.S., and A.D. Long. 2003. The lowdown on linkage disequilibrium. Plant Cell 15:1502–1506. doi:10.1105/tpc.150730

Glaubitz, J.C., T.M. Casstevens, F. Lu, J. Harriman, R.J. Elshire, Q. Sun, and E.S. Buckler. 2014. TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. PLoS ONE 9. doi:10.1371/journal.pone.0090346

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

Gonzalez-Recio, O., and S. Forni. 2011. Genome-wide prediction of discrete traits using bayesian regressions and machine learning. Genet. Sel. Evol. 43. 10. doi:10.1186/1297-9686-43-7

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. doi:10.1186/s12711-015-0102-z

Grattapaglia, D., and R. Sederoff. 1994. Genetic-linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross mapping strategy and RAPD markers. Genetics 137:1121–1137.

Habier, D., R.L. Fernando, and J.C.M. Dekkers. 2007. The impact of genetic relationship information on genome-assisted breeding values. Genetics 177:2389–2397. doi:10.1534/genetics.107.081190.

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

Hayes, B.J., P.J. Bowman, A.J. Chamberlain, and M.E. Goddard. 2009. Invited review: Genomic selection in dairy cattle: Progress and challenges. J. Dairy Sci. 92:433–443. doi:10.3168/jds.2008-1646

Heffner, E.L., J.L. Jannink, H. Iwata, E. Souza, and M.E. Sorrells. 2011a. Genomic selection accuracy for grain quality traits in biparental wheat populations. Crop Sci. 51:2597–2606. doi:10.2135/cropsci2011.05.0253

Heffner, E.L., J.L. Jannink, and M.E. Sorrells. 2011b. Genomic selection accuracy using multifamily prediction models in a wheat breeding program. Plant Gen. 4:65–75. doi:10.3835/plantgenome2010.12.0029

Hickey, J.M., S. Dreisigacker, J. Crossa, S. Hearne, R. Babu, B.M. Prasanna, M. Grondona, A. Zambelli, V.S. Windhausen, K. Mathews, and G. Gorjanc. 2014. Evaluation of genomic selection training population designs and genotyping strategies in plant breeding programs using simulation. Crop Sci. 54:1476–1488. doi:10.2135/cropsci2013.03.0195

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, Y.F., J.A. Poland, C.P. Wight, E.W. Jackson, and N.A. Tinker. 2014. Using genotyping-by-sequencing (GBS) for genomic discovery in cultivated oat. PLoS ONE 9. doi:10.1371/annotation/c85feb29-046d-45f6-ba12-510f9e1d9412

International Cassava Genetic Map Consortium (ICGMC). 2015. High-resolution linkage map and chromosome-scale genome assembly for cassava (*Manihot esculenta* Crantz) from 10 populations. G3: Genes Genomes Genet. 5:133–144. doi:10.1534/g3.114.015008

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

Jannink, J.L., A.J. Lorenz, and H. Iwata. 2010. Genomic selection in plant breeding: From theory to practice. Briefings in functional genomics 9:166–177. doi:10.1093/bfgp/elq001

Jia, J.Z., S.C. Zhao, X.Y. Kong, et al. 2013. *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. Nature 496:91–95. doi:10.1038/nature12028

Krzywinski, M., J. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, S.J. Jones, and M.A. Marra. 2009. Circos: An information aesthetic for comparative genomics. Genome Res. 19:1639–1645. doi:10.1101/gr.092759.109

Kumar, S., D. Chagne, M. Bink, R.K. Volz, C. Whitworth, and C. Carlisle. 2012. Genomic selection for fruit quality traits in apple (*Malus* × *domestica Borkh*.). PLoS ONE 7. doi:10.1371/journal.pone.0036674.

Lian, L., A. Jacobson, S.Q. Zhong, and R. Bernardo. 2014. Genomewide prediction accuracy within 969 maize biparental populations. Crop Sci. 54:1514–1522. doi:10.2135/cropsci2013.12.0856

Lian, L., A. Jacobson, S.Q. Zhong, and R. Bernardo. 2015. Prediction of genetic variance in biparental maize populations: Genomewide marker effects versus mean genetic variance in prior populations. Crop Sci. 55:1181–1188. doi:10.2135/cropsci2014.10.0729

Lin, Z., B.J. Hayes, and H.D. Daetwyler. 2014. Genomic selection in crops, trees and forages: A review. Crop Pasture Sci. 65:1177–1191. doi:10.1071/CP13363

Lipka, A.E., F. Lu, J.H. Cherney, E.S. Buckler, M.D. Casler, and D.E. Costich. 2014. Accelerating the switchgrass (*Panicum virgatum* L.) breeding cycle using genomic selection approaches. PLoS ONE 9. doi:10.1371/journal.pone.0112227

Lorenz, A.J., S.M. Chao, F.G. Asoro, E.L. Heffner, T. Hayashi, H. Iwata, K.P. Smith, M.E. Sorrells, and J.L. Jannink. 2011. Genomic selection in plant breeding: Knowledge and prospects. Adv. Agron. 110:77–123.

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

Mahelka, V., D. Kopecky, and L. Pastova. 2011. On the genome constitution and evolution of intermediate wheatgrass (*Thinopyrum intermedium*: Poaceae, Triticeae). BMC Evol. Biol. 11. doi:10.1186/1471-2148-11-127

Mayer, K.F.X., R. Waugh, P. Langridge, et al. 2012. A physical, genetic and functional sequence assembly of the barley genome. Nature 491:711–716. doi:10.1038/nature11543

Meuwissen, T.H.E., B.J. Hayes, and M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829.

Moumouni, K.H., B.A. Kountche, M. Jean, C.T. Hash, Y. Vigouroux, B.I.G. Haussmann, and F. Belzile. 2015. Construction of a genetic map for pearl millet, *Pennisetum glaucum* (L.) R. Br., using a genotyping-by-sequencing (GBS) approach. Mol. Breed. 35. doi:10.1007/s11032-015-0212-x

Van Ooijen, J.W. 2006. JoinMap 4. Software for the calculation of genetic linkage maps in experimental populations. Kyazma BV, Wageningen, the Netherlands.

Van Ooijen, J.W. 2011a. JoinMap 4.1. New features. Kyazma BV, Wageningen, the Netherlands.

Van Ooijen, J.W. 2011b. Multipoint maximum likelihood mapping in a full-sib family of an outcrossing species. Genet. Res. 93:343–349. doi:10.1017/S0016672311000279

Pan, Q.C., F. Ali, X.H. Yang, J.S. Li, and J.B. Yan. 2012. Exploring the genetic characteristics of two recombinant inbred line populations via high-density SNP markers in maize. PLoS ONE 7. doi:10.1371/journal.pone.0052777

Peleg, Z., Y. Saranga, T. Suprunova, Y. Ronin, M.S. Roder, A. Kilian, A.B. Korol, and T. Fahima. 2008. High-density genetic map of durum wheat × wild emmer wheat based on SSR and DArT markers. Theor. Appl. Genet. 117:103–115. doi:10.1007/s00122-008-0756-9

Perez, P., and G. de los Campos. 2014. Genome-wide regression and prediction with the BGLR statistical package. Genetics 198:483–495. doi:10.1534/genetics.114.164442

Poland, J.A., P.J. Brown, M.E. Sorrells, and J.L. Jannink. 2012. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS ONE 7. doi:10.1371/journal.pone.0032253

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

R Core Team. 2014. R: A language and environment for statistical computing. Available at http://www.R-project.org/ (verified 20 Nov. 2015). The R Foundation, Vienna, Austria.

Robertson, G.P., E.A. Paul, and R.R. Harwood. 2000. Greenhouse gases in intensive agriculture: Contributions of individual gases to the radiative forcing of the atmosphere. Science 289:1922–1925. doi:10.1126/science.289.5486.1922

Russell, J., C. Hackett, P. Hedley, H. Liu, L. Milne, M. Bayer, D. Marshall, L. Jorgensen, S. Gordon, and R. Brennan. 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

Rutkoski, J.E., J. Poland, J.L. Jannink, and M.E. Sorrells. 2013. Imputation of unordered markers and the impact on genomic selection accuracy. G3: Genes Genomes Genet. 3:427–439. doi:10.1534/g3.112.005363

Rutkoski, J.E., J.A. Poland, R.P. Singh, J. Huerta-Espino, S. Bhavani, H. Barbier, M.N. Rouse, J.L. Jannink, and M.E. Sorrells. 2014. Genomic selection for quantitative adult plant stem rust resistance in wheat. Plant Gen. 7. doi:10.3835/plantgenome2014.02.0006

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 Gen. 8. doi:10.3835/plantgenome2014.05.0020

Savicky, P. 2014. pspearman: Spearman's rank correlation test. Available at: http://CRAN.R-project.org/package=pspearman (verified 20 Nov. 2015).

Spindel, J., H. Begum, D. Akdemir, P. Virk, B. Collard, E. Redoña, G. Atlin, J.L. Jannink, and S.R. McCouch. 2015. Genomic selection and association mapping in rice (*Oryza sativa*): Effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. PLoS Genet. 11:e1004982. doi:10.1371/journal.pgen.1004982

Swarts, K., H.H. Li, J.A.R. Navarro, D. An, M.C. Romay, S. Hearne, C. Acharya, J.C. Glaubitz, S. Mitchell, R.J. Elshire, E.S. Buckler, and P.J. Bradbury. 2014. Novel methods to optimize genotypic imputation for low-coverage, next-generation sequence data in crop plants. Plant Gen. 7. doi:10.3835/plantgenome2014.05.0023

Thuillet, A.C., T. Bataillon, S. Poirier, S. Santoni, and J.L. David. 2005. Estimation of long-term effective population sizes through the history of durum wheat using microsatellite data. Genetics 169:1589–1599. doi:10.1534/genetics.104.029553

Usai, M.G., M.E. Goddard, and B.J. Hayes. 2009. LASSO with cross-validation for genomic selection. Genet. Res. 91:427–436. doi:10.1017/S0016672309990334

Vogel, K.P. 2013. Comparison of two perennial grass breeding systems with switchgrass. Crop Sci. 53:863–870. doi:10.2135/cropsci2012.09.0559

Voorrips, R.E. 2002. MapChart: Software for the graphical presentation of linkage maps and QTLs. J. Hered. 93:77–78. doi:10.1093/jhered/93.1.77

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

Wang, Y., M.F. Mette, T. Miedaner, M. Gottwald, P. Wilde, J.C. Reif, and Y.S. Zhao. 2014. The accuracy of prediction of genomic selection in elite hybrid rye populations surpasses the accuracy of marker-assisted

selection and is equally augmented by multiple field evaluation locations and test years. BMC Genomics 15. doi:10.1186/1471-2164-15-556.

Wientjes, Y.C.J., R.F. Veerkamp, and M.P.L. Calus. 2013. The effect of linkage disequilibrium and family relationships on the reliability of genomic prediction. Genetics 193:621–631. doi:10.1534/genetics.112.146290

Windhausen, V.S., G.N. Atlin, J.M. Hickey, J. Crossa, J.L. Jannink, M.E. Sorrells, B. Raman, J.E. Cairns, A. Tarekegne, K. Semagn, Y. Beyene, P. Grudloyma, F. Technow, C. Riedelsheimer, and A.E. Melchinger. 2012. Effectiveness of genomic prediction of maize hybrid performance in different breeding populations and environments. G3: Genes Genomes Genet. 2:1427–1436. doi:10.1534/g3.112.003699

Wurschum, T., J.C. Reif, T. Kraft, G. Janssen, and Y.S. Zhao. 2013. Genomic selection in sugar beet breeding populations. BMC Genet. 14. doi:10.1186/1471-2156-14-85

Zhang, X.F., L.R. DeHaan, L. Higgins, T.W. Markowski, D.L. Wyse, and J.A. Anderson. 2014. New insights into high-molecular-weight glutenin subunits and sub-genomes of the perennial crop *Thinopyrum intermedium* (Triticeae). J. Cereal Sci. 59:203–210. doi:10.1016/j.jcs.2014.01.008

Zhao, Y.S., M. Gowda, W.X. Liu, T. Wurschum, H.P. Maurer, F.H. Longin, N. Ranc, and J. Reif. 2012. Accuracy of genomic selection in European maize elite breeding populations. Theor. Appl. Genet. 124:769–776. doi:10.1007/s00122-011-1745-y