Published online July 13, 2017

ORIGINAL RESEARCH

Multi-Locus Mixed Model Analysis Of Stem Rust Resistance In Winter Wheat

Paul D. Mihalyov, Virginia A. Nichols, Peter Bulli, Matthew N. Rouse, and Michael O. Pumphrey*

Abstract

Genome-wide association mapping is a powerful tool for dissecting the relationship between phenotypes and genetic variants in diverse populations. With the improved cost efficiency of high-throughput genotyping platforms, association mapping is a desirable method of mining populations for favorable alleles that hold value for crop improvement. Stem rust, caused by the fungus *Puccinia graminis* f. sp. *tritici,* is a devastating disease that threatens wheat (*Triticum aestivum* L.) production worldwide. Here, we explored the genetic basis of stem rust resistance in a global collection of 1411 hexaploid winter wheat accessions genotyped with 5390 single nucleotide polymorphism markers. To facilitate the development of resistant varieties, we characterized marker–trait associations underlying field resistance to North American races and seedling resistance to the races TTKSK (Ug99), TRTTF, TTTTF, and BCCBC. After evaluating several commonly used linear models, a multi-locus mixed model provided the maximum statistical power and improved the identification of loci with direct breeding application. Ten high-confidence resistance loci were identified, including SNP markers linked to *Sr8a, Sr9h, Sr28,* and *Sr31,* and at least three newly discovered resistance loci that are strong candidates for introgression into modern cultivars. In the present study, we assessed the power of multi-locus association mapping while providing an in-depth analysis for its practical ability to assist breeders with the introgression of rare alleles into elite varieties.

Core Ideas

- Multi-locus association mapping outperforms single-locus genome scans in polyploid species.
- Single nucleotide polymorphism marker IWA435 reliably detects the 1BL.1RS rye translocation in diverse germplasm.
- Markers linked to three novel stem rust resistance loci were identified.

HEXAPLOID WHEAT is a staple cereal crop that provides nearly 20% of the human population's caloric intake (Pfeifer et al., 2014). The fungal pathogen *Puccinia graminis* f. sp. *tritici* (*Pgt*), the causal agent of wheat stem rust, reduces yield and end-use quality with potential annual losses of up to US $1.12 billion globally (Pardey et al., 2013). Progressive breeding efforts successfully controlled the disease during the late twentieth century before it evolved into one of the largest biological threats to wheat production worldwide (Stokstad, 2007). Specifically, the highly virulent race Ug99 (syn. TTKSK), discovered in Uganda, overcame the previously effective resistance gene *Sr31* (Pretorius et al., 2000) and its

P.D. Mihalyov, V.A. Nichols, P. Bulli, and M.O. Pumphrey, Dep. of Crop and Soil Sciences, Washington State Univ., Pullman, WA 99164; M.N. Rouse, USDA-ARS Cereal Disease Lab. and Dep. of Plant Pathology, Univ. of Minnesota, St. Paul, MN 55108. Received 10 Jan. 2017. Accepted 18 Mar. 2017. *Corresponding author (m.pumphrey@wsu.edu).

Abbreviations: BLUE, best linear unbiased estimates; COI, coefficient of infection; GWAS, genome-wide association study; K, kinship matrix; LD, linkage disequilibrium; MAF, minor allele frequency; MAS, marker-assisted selection; MLMM, multi-locus mixed model; PC, principal component; *Pgt, Puccinia graminis* f. sp. *tritici*; Q, subpopulation membership coefficient matrix; QTL, quantitative trait locus; QTN, quantitative trait nucleotide; SNP, single nucleotide polymorphism

Published in Plant Genome
Volume 10. doi: 10.3835/plantgenome2017.01.0001

derivatives have since overcome important sources of resistance such as *Sr24, Sr36,* and *SrTmp* (Newcomb et al., 2016; Singh et al., 2011).

Chemical fungicides have been used by growers to alleviate the damage caused by *Pgt*, but several limitations such as cost, availability, safety concerns, and application methods make genetic resistance a highly favorable alternative (Oliver, 2014; Singh et al., 2016). To date, at least 65 *Pgt* resistance alleles have been identified (McIntosh et al., 2014; Yu et al., 2014), but many of them have not been characterized well enough to be used in marker-assisted selection (MAS) or are ineffective against the current widespread *Pgt* races. Thus developing reliable diagnostic markers for both known and novel *Pgt* resistance alleles can dramatically improve the discovery of new resistance genes, aid in cultivar development, and support the surveillance of pathogen variation to guide resistance gene deployment strategies.

Association mapping enables the identification of genetic markers that are closely linked to a trait of interest by taking advantage of historical recombination in diverse germplasm to provide higher resolution than traditional biparental linkage mapping. In wheat, genome-wide association studies (GWAS) have helped identify markers linked to genes controlling several traits including end-use quality (Breseghello and Sorrells, 2006; Reif et al., 2011), morphology (Rasheed et al., 2014; Zanke et al., 2014), and disease resistance (Aoun et al., 2016; Bulli et al., 2016; Jighly et al., 2015; Letta et al., 2014; Yu et al., 2011; Zhang et al., 2014).

Unbalanced allele frequencies among subpopulations (population stratification) can lead to inflated rates of both false positive and false negative marker–trait associations during association mapping (Pritchard et al., 2000). The confounding effects of population structure have been mitigated by mixed linear models that use covariables such as principal components (PCs) and population-wide kinship estimates (Patterson et al., 2006; Yu et al., 2006). During the process of mapping complex traits, additional power can be lost because of masking effects between causal loci (Würschum and Kraft, 2015). Single-locus genome scans conducted by ordinary mixed models do not adequately control for large effect loci; multi-locus models were recently proposed as a potential tool for addressing this issue (Rakitsch et al., 2013; Segura et al., 2012). The multi-locus mixed model (MLMM) introduced by Segura et al. (2012) uses stepwise regression to incorporate the most influential markers as cofactors and has been used successfully in several association mapping studies to date (Lipka et al., 2013; Sauvage et al., 2014; Vaughn et al., 2014).

Given the nature of the wheat–*Pgt* pathosystem where qualitative, semiquantitative, and quantitative genetic inheritance is common, we hypothesized that a multi-locus model applied to a diverse germplasm panel would increase the power to detect *Pgt* resistance loci compared with single-locus models. By investigating several different association mapping methodologies and cross-referencing our results with similar experiments, this study aims to reveal the utility of MLMM as a tool for wheat–*Pgt* GWAS analysis and marker discovery.

MATERIALS AND METHODS

Plant Material

A global collection of 1654 hexaploid winter wheat accessions was assembled from the USDA-ARS National Small Grains Collection for the Triticeae Coordinated Agricultural Project. To ensure genetic purity, seeds for this study were harvested from single-plant selections at the USDA-ARS Small Grains and Potato Germplasm Research Unit in Aberdeen, ID. After the genotype quality filters described below were applied, a panel of 1411 genetically unique accessions was selected for association mapping. The geographical origins of the accessions used in this study include Europe (46.6%), the Middle East (25.7%), Asia (13.7%), North America (8.2%), South America (3.9%), and a limited number from Africa and Australasia (<2%).

Genome-Wide Genotyping

All accessions were genotyped with the Infinium wheat SNP 9K iSelect assay (Illumina Inc., San Diego, CA) at the USDA-ARS genotyping lab in Fargo, ND. Illumina software GenomeStudio version 2011.1 was used to process raw SNP data and to verify allele calls manually, yielding 6758 high-quality polymorphic SNP markers. Genetic relatedness was assessed by using an identity-by-state kinship matrix computed in JMP version 6.0 (SAS Corp., Cary, NC) and pairs of accessions with >99% genetic similarity were identified. One representative accession was retained. Redundant groups of SNP markers [pairwise linkage disequilibrium (LD) $r^2 = 1.0$] were also filtered to include only one representative marker. Both individuals and SNP markers with > 10% missing genotype data were excluded and SNP markers with <5% minor allele frequency (MAF) were discarded to minimize the likelihood of Type I errors during association analyses. Application of these filtering criteria yielded 1411 accessions genotyped with 5389 polymorphic SNP markers. Markers were positioned using the 9K SNP consensus map (Cavanagh et al., 2013) with corrections made to the arm orientations of 4A, 5A, and 5B (Maccaferri et al., 2015). Additionally, fragmented linkage groups on 3D (3D1, 3D2 and 3D3), 5D [5D1cult (reversed), 5D2cult and 5D3cult], and 6D [6D1 and 6D2 (reversed)] were joined with a spacing of 20 cM.

To date, no markers on the 9K wheat SNP array have been mapped directly to rye (*Secale cereale* L.) translocations and confirmed in large germplasm collections. For additional validation and control data, we genotyped the 1411 accessions with 'SCM9', a codominant microsatellite marker which identifies the rye translocations 1AL.1RS (harboring *Sr1RS-Amigo*) and 1BL.1RS (harboring *Sr31*) (Weng et al., 2007). Marker data from this assay were then included in the genotype matrix, yielding a total of 5390 markers that were suitable for association mapping.

Population Structure and LD

Principal component analysis was performed using the complete set of genome-wide SNP markers, implemented with the R function '*prcomp*' (R Core Team, 2014), to elucidate population structure and estimate the covariables for mixed model association mapping. Additionally, STRUCTURE version 2.3.4 (Pritchard et al., 2000) was used to generate subpopulation membership coefficients (Q matrix) from a set of 1510 nonredundant tagSNPs (LD r^2 threshold for marker bins = 0.25). Assuming an admixed population structure and correlated allele frequencies, five independent runs for k = 1 to 10 hypothetical subpopulations were examined using 50,000 burn-in iterations followed by 100,000 recorded Markov chain iterations. The web-based tool STRUCTURE Harvester (Earl and vonHoldt, 2012) was used to determine the optimal number of subpopulations based on the ΔK method (Evanno et al., 2005). Both the principal component and the Q matrices were evaluated as covariables during the iterative model comparisons described in the Materials and Methods (Model Selection).

The extent of population-specific genome-wide LD was assessed by the LD squared allele frequency correlation (r^2) calculated for each intrachromosomal marker pair using JMP version 6.0; r^2 values were plotted against the map distance (cM) between each pair then fitted with a smooth loess curve using the function '*scatter.smooth*' in R. To facilitate comparisons across the literature, we report the genetic distance at which LD decays at three commonly reported critical r^2 values (0.1, 0.2, and 0.3).

Stem Rust Screening

Accessions were evaluated for both seedling and field response to *Pgt* at the USDA Cereal Disease Laboratory in St. Paul, MN. Five separate seedling resistance tests were performed under controlled greenhouse conditions (Rouse et al., 2011) by inoculating 7- to 9-d-old plants with *Pgt* races TRTTF (isolate 06YEM34–1), TTTTF (01MN84A-1–2), TTKSK (Ug99; 04KEN156/04), BCCBC (09CA115–2), and a bulk of the following six North American *Pgt* isolates: MCCFC (59KS19), QFCSC (06ND76C), QTHJC (75ND717C), RCRSC (77ND82A), RKQQC (99KS76A), and TPMKC (74MN1409)(Jin et al., 2008). After 14 d, infection types were evaluated according to the Stakman rating system (Stakman et al., 1962). Phenotypic data compatible with association analysis were generated by converting Stakman infection types to a 0 to 9 linear scale (Gao et al., 2016; Zhang et al., 2014).

Resistance was evaluated in adult plants under field conditions over 3 yr (2012, 2014, and 2015) in St. Paul, MN. Each accession was planted in 1-m rows with the checks 'Siouxland' (*Sr24+Sr31*), 'Sisson' (*Sr31+Sr36*), 'McNair' (susceptible), and 'Panola' (susceptible) planted every 100 entries. To ensure uniform disease pressure, spreader rows planted perpendicular to pairs of entry rows composed of the susceptible cultivar Panola were inoculated with urediniospores from the North American *Pgt* race bulk described above. When the majority of accessions had entered the grain-filling stage of growth (>7 on Zadok's scale; Zadoks et al., 1974), two measures were used to quantify the response to stem rust infection: (i) disease severity was scored using a modified Cobb scale (0–100%) to estimate the total coverage of rusted tissue (Peterson et al., 1948) and (ii) infection response was scored by evaluating pustule characteristics to determine the host reaction: resistant = 0.2; moderately resistant, 0.4; moderately susceptible, 0.8; susceptible, 1 (Roelfs et al., 1992). Disease scores from the two indices were multiplied to calculate a coefficient of infection (COI) on a 0 to 100 linear scale. Best linear unbiased estimates (BLUEs) across the three environments were calculated in ASReml-R (VSN International Ltd., Hemel Hempstead, UK) by fitting both genotypes and environments as fixed effects. The vector of COI-BLUE values was then used for association analysis.

Model Selection

A primary goal of applying mixed models in association mapping is to maximize the power to detect true associations with the lowest Type I error rates. The power of four common methods was evaluated using an approach described by Liu et al. (2016) to determine the most suitable model for the panel. Briefly, the '*G2P*' function from the open-source R package GAPIT (Lipka et al., 2012) was used to assign 10 causal quantitative trait nucleotides (QTNs) randomly in the genotype matrix of the NSGC panel; the QTNs were assigned geometrically distributed genetic effects and then normally distributed residual effects were simulated to obtain the desired heritability. A vector of phenotypes was generated by summing the genetic and residual effects and then each model was applied to the simulated dataset. The ability of each model to detect SNP markers correctly within 2 cM of the causal QTN was evaluated. The proportion of correctly identified causal QTNs was plotted versus the false discovery rate; for a given set of simulated phenotypes, the area under the curve was used as a measure of statistical power for each model.

The models evaluated were: (i) a general linear model with two PCs included as fixed effects (Patterson et al., 2006); (ii) a mixed linear model with two fixed-effect PCs and a random-effect kinship matrix (K matrix) (Zhao et al., 2007); (iii) a compressed mixed linear model, which is a more computationally efficient mixed linear model where relationship estimates are fitted to groups of accessions in the K matrix instead of each individual (Zhang et al., 2010b); and (iv) a multi-locus mixed-model (MLMM), which is a modified mixed linear model that begins with a mixed model then introduces forward–backward stepwise linear regression to include significant SNP markers as fixed-effect cofactors. Three variations of this model were investigated: MLMM (PC + K), MLMM (Q + K), and MLMM (K). The power of these models applied to 100 simulated datasets was compared at three levels of heritability (h^2 = 0.25, 0.50, and 0.75).

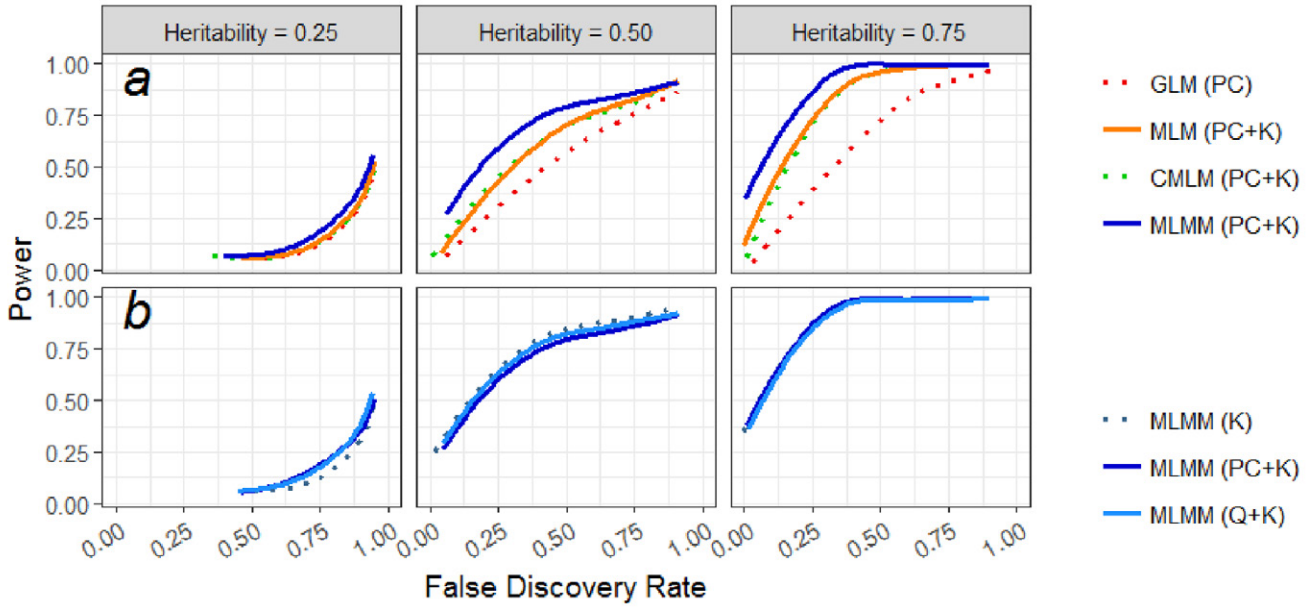

Fig. 1. Model power versus false discovery rate for genome-wide association with simulated traits exhibiting three levels of heritability (0.25, 0.5, 0.75), each derived from 10 causal single nucleotide polymorphisms (SNPs) randomly chosen from the current dataset. (a) Comparison of four unique linear models [general linear model (GLM), mixed linear model (MLM), compressed mixed linear model (CMLM), multi-locus mixed model (MLMM)] and (b) comparison of three variations to the MLMM [random-effect kinship (K), fixed-effect principal components (PC) + K, fixed-effect STRUCTURE coefficients (Q) + K]. Model power is expressed as an average of 100 simulations.

Association Analysis

Using the complete set of 5390 polymorphic markers, association analyses were conducted on the 1411 accessions across six phenotypic datasets: five seedling phenotypic datasets evaluated under controlled greenhouse conditions for their response to *Pgt* races TRTTF, TTTTF, TTKSK, BCCBC, and a North American race mixture and one dataset representing BLUEs across 3 yr of field COI resistance ratings on adult plants. On the basis of the model selection procedures, MLMM + K was chosen for conducting association analyses (Fig. 1a,b); a VanRaden kinship matrix (VanRaden, 2008) was computed in GAPIT and all remaining MLMM procedures were performed in R using the open-sourced code provided by Segura et al. (2012) (available at https://github.com/Gregor-Mendel-Institute/mlmm, accessed 8 May 2017). For each phenotypic dataset, forward linear regression was performed for 10 steps followed by a backward stepwise regression where fixed-effect SNP cofactors were removed one at a time. Further model selection was implemented by using an extended Bayesian information criterion to identify which step returned the best model fit for each trait. This method is similar to the classic BIC with an additional penalty introduced for each cofactor to minimize model space (Chen and Chen, 2008). After the optimum step was determined, *P*-values for each SNP were subjected to Bonferroni multiple-comparison adjustments; only markers exhibiting a Bonferroni *P*-value < 0.05 are reported.

RESULTS

Genome-Wide Marker Data and LD Analysis

After quality filtering measures were implemented for the iSelect 9K SNP genotypic data, a total of 5059 biallelic SNP markers were included that spanned a genetic distance of 3557.3 cM based on the 9K consensus map (Cavanagh et al., 2013), plus an additional 330 markers lacking a map position. The B genome exhibited the greatest marker density with an average of 1.83 markers per cM, followed by 1.75 markers per cM on the A genome and 0.33 markers per cM on the D genome. The average genome-wide marker density achieved in this panel (~0.7 cM between each marker plus 330 supplemental unmapped markers) is sufficient for association mapping based on the observed rate of LD decay. Genome-wide LD decay ranged from 1.54 to 4.68 cM in this global winter wheat collection according to three critical r^2 thresholds (Fig. 2a,b), which is consistent with numerous studies that report that LD decay in hexaploid wheat ranges from 1 to 10 cM (Chao et al., 2007; Würschum et al., 2013; Yu et al., 2011; Zhang et al., 2010a).

Composition of Population Substructure

The model-based inference of population structure obtained from Structure version 2.3.4 revealed two primary subpopulations in the panel. The average membership coefficient of each subpopulation was 0.88 (Q-1) and 0.90 (Q-2), indicating a relatively low amount of admixture among the subpopulations. The first subpopulation, Q-1, consisted of 477 accessions primarily representing Asia and the Middle East, and Q-2 was composed of 934

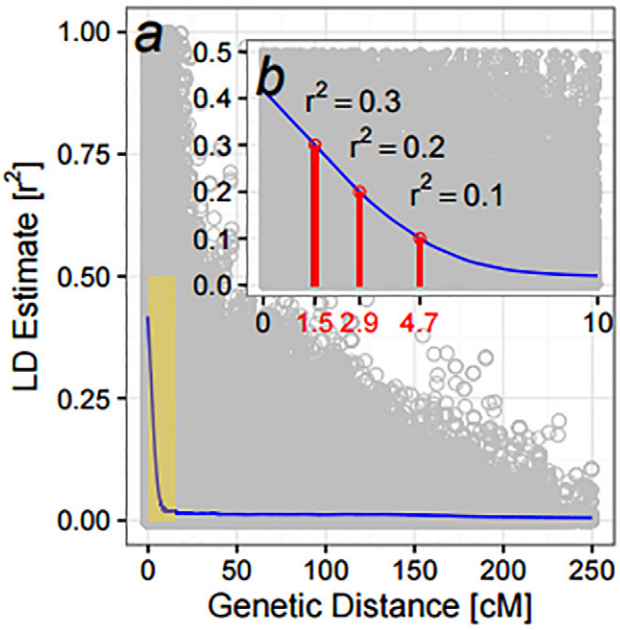

Fig. 2. (a) Genome-wide linkage disequilibrium (LD) r^2 values plotted against the genetic distance between each intrachromosomal marker pair. (b) Population-specific LD decay estimated over three critical r^2 thresholds (0.1, 0.2, 0.3).

accessions with heavy representation from Europe and the Americas (Fig. 3a,b). The first two principal components, which explained 17.3 and 5.2% of genetic variation, respectively, as well as subpopulation membership coefficients were assessed for their value as fixed-effect covariables during association mapping. For this dataset, incorporating PCs or the Q matrix as covariables did not improve power (Fig. 1b) so only the K matrix was included in the model. Several studies have documented that kinship alone may correct adequately for population stratification (Chang et al., 2016; Kumar et al., 2013).

Response to Stem Rust Infection

Among the 1411 accessions screened at the seedling stage, the frequency of accessions with a linearized infection type from 0–5 (Stakman rating ~2+ or less) ranged from 6.0 (North American race bulk) to 22.5% (BCCBC). Similarly, we observed a low frequency of resistance (COI 0–50) in field screening nursery data: 9.6 (2012), 9.6 (2014), and 8.4% (2015). These results are not unexpected, as most accessions that were included in our panel were not selected in North America, where the *Pgt* isolates used to inoculate our field studies were collected. The correlation between phenotypic data across the 3 yr was moderate (r^2 = 0.69, 0.63, and 0.54 for 2012–2014, 2012–2015, and 2014–2015, respectively), indicating environmental as well as genetic influence on the expressed phenotypes. Potential experimental error while generating phenotype data cannot be ruled out when evaluating such a large collection.

The SNP Marker IWA435 Reliably Identifies 1BL.1RS Translocations

The microsatellite marker SCM9 was used to detect 1AL.1RS and 1BL.1RS wheat–rye translocations. Out of 1411 accessions in the panel, 42 (2.98%) harbored 1BL.1RS and one harbored 1AL.1RS. After manually filtering iSelect 9K SNP markers for collinearity with SCM9, we identified IWA435 as a reliable marker for 1BL.1RS (Fig. 4). Despite already having the microsatellite marker SCM9 for 1BL.1RS, breeders generally need a consistent genotyping platform to gather large amounts of data with high throughput. The data presented here will immediately assist breeders in selecting for 1BL.1RS with more flexibility to accommodate the preferred genotyping methods.

According to SCM9 marker data, the null allele for IWA435 correctly identified 41 out of 42 (97.6%) accessions carrying 1BL.1RS. The marker results from SCM9 indicate that accession PI626798 carries 1BL.1RS, yet IWA435 showed non-1BL.1RS; however, PI626798 exhibited susceptibility to *Pgt* races TRTTF, TTTTF, and BCCBC. Given that 1BL.1RS is expected to harbor *Sr31*, SCM9 appears to have yielded a false positive. On the other hand, the null cluster for IWA435 contains two accessions that SCM9 predicts to be non-1RS: PI361785 is susceptible, and PI598319 is resistant to *Sr31*-avirulent races. These inconsistencies could be caused by a true underlying genetic deviation, experimental error in generating genotype or phenotype data, or heterogeneity that was not detected. Although IWA435 is not polymorphic for 1AL.1RS versus non-1RS genotypes, IWA435 matches the reliability of SCM9 for detecting 1BL.1RS lines.

Association Mapping Results

Genome-wide association mapping was conducted using MLMM + K for six traits: seedling resistance to individual races TRTTF, TTTTF, TTKSK, BCCBC, and a North American race bulk (MCCFC, QFCSC, QTHJC, RCRSC, RKQQC, and TPMKC) and COI-BLUE values representing 3 yr resistance ratings in artificially inoculated field conditions. These analyses revealed 15 SNP markers exhibiting a strong association (Bonferroni-adjusted $P <$ 0.05) with 10 stem rust resistance loci (Table 1). Supplementary Table S1 contains a list of all 1411 accessions used in this study, complete with phenotype information for all six traits and genotype data for all the markers described in the results and discussion.

Single-Race Seedling Tests

Three markers (SCM9, IWA7913, and IWA2722) were associated with resistance to *Pgt* race TRTTF and explained 39.8% of the total phenotypic variation. Postulations based on race specificity, infection type, chromosomal location, and previous linkage mapping studies suggest these markers are linked to *Sr31* (1BL.1RS), *Sr8a* (6AS), and a previously undocumented resistance locus on chromosome 2D. SCM9 was the first cofactor added to the regression after exhibiting the most significant

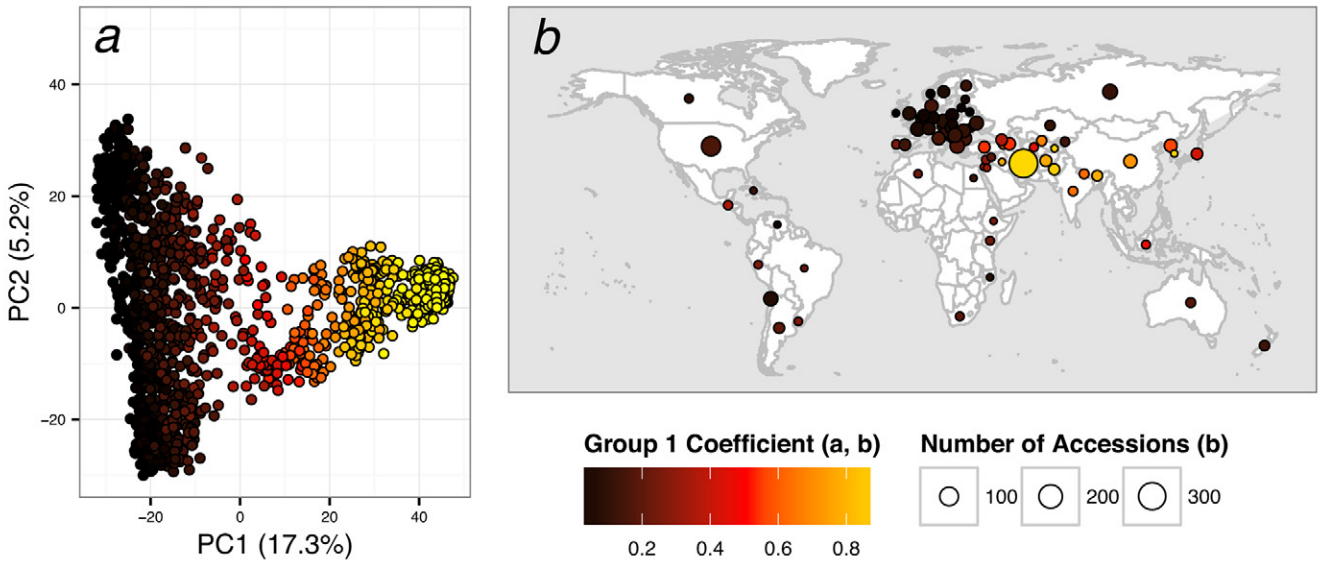

Fig. 3. (a) Scatter plot of the first two principal components, and (b) the geographical distribution of wheat accessions used in this study. Individuals or groups are color-coded by subpopulation membership coefficients derived from STRUCTURE version 2.3.4.

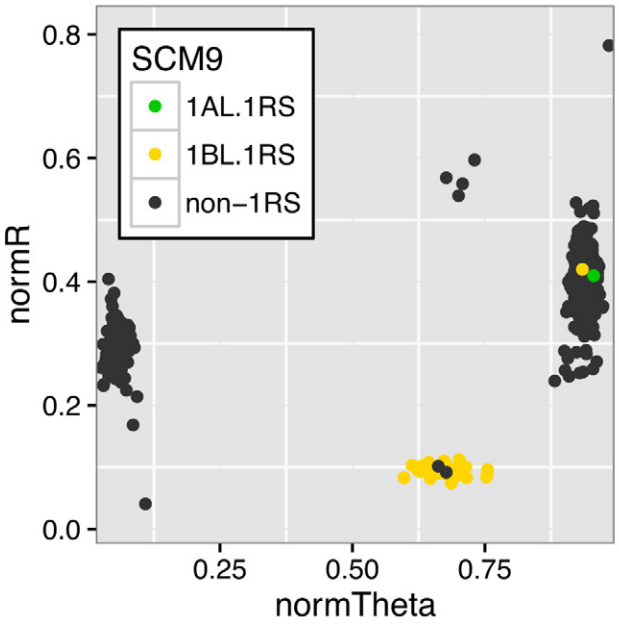

Fig. 4. Clustering pattern exhibited by assay IWA435 shows two primary wheat alleles mapped to chromosome 1BS and a null allele; data points are color-coded according to the results from microsatellite marker SCM9, showing that the null cluster represents 1BL.1RS.

association with TRTTF resistance (Bonferroni $P = 7.25 \times 10^{-76}$), which is expected given its near-perfect linkage with *Sr31*. The strongest remaining signal was from IWA7913 (*Sr8a*, Bonferroni $P = 1.32 \times 10^{-4}$), which was therefore introduced as a cofactor in Step 2; our postulation is in accordance with recent GWAS and biparental linkage mapping studies [Bajgain et al. (2015) and Guerrero-Chavez et al. (2015), respectively], which reported tight linkage of SNP marker IWA7913 to *Sr8a*. Lastly, SNP marker IWA2722, which is mapped to the centromeric region of chromosome 2D, was added to the model in Step 3 (Bonferroni $P = 0.018$). To date, *Pgt* resistance genes *Sr6*, *Sr46*, and *Sr54* have been mapped to chromosome 2D (McIntosh et al., 2014), but TRTTF exhibits virulence against each gene; furthermore, none of the previously characterized genes have been mapped near the centromere (Ghazvini et al., 2013; Tsilo et al., 2009; Yu et al., 2015). Thus, IWA2722 is likely to be linked to a newly documented resistance gene that is effective against *Pgt* race TRTTF.

Among the 20 *Pgt* resistance genes in the International Differential Set (Jin et al., 2008), only *Sr24* and *Sr31* are effective against race TTTTF. Although this North American race is avirulent against many spring wheat cultivars, it poses a greater threat to winter wheat varieties that commonly rely on *Sr6*, *Sr10*, *Sr36*, and *SrTmp* (Jin, 2005). Two genomic regions associated with TTTTF resistance were identified: (i) the 1BL.1RS rye translocation harboring *Sr31* (SCM9) and (ii) the distal end of 2BL (IWA6656, Bonferroni $P = 0.015$), where no previously characterized resistance genes are known to be effective against TTTTF. Thus IWA6656 is likely linked to a newly documented source of resistance against this virulent race. However, the presence of this gene should be confirmed in further studies.

Association mapping for resistance to *Pgt* race TTKSK (Ug99) revealed three putative loci that explained 39.6% of the total phenotypic variation. Chromosome 2B harbors numerous *Pgt* resistance genes including at least six that are effective against TTKSK (*Sr9h*, *Sr28*, *Sr36*, *Sr39*, *Sr40*, and *Sr47*). Previous efforts to develop an integrated map of these loci have been challenging, possibly because of the complex inheritance patterns observed on 2B, including segregation distortion and chromosomal rearrangements (Li et al., 2015; Zurn et al., 2014). Consequently, our ability to make postulations for chromosome 2B marker–trait associations

Table 1. Significant associations listed for each trait in order of marker inclusion into the multi-locus mixed model.

Trait	Marker	Chromosome	Position (cM)	Regression step introduced	Cumulative R^2 (%)	*P*-value (Bonferroni)	Postulation
TRTTF	SCM9	1BL.1RS	NA	1	32.6	7.25×10^{-76}	*Sr31*
	IWA7913	6A	9.54	2	36.7	1.32×10^{-4}	*Sr8a*
	IWA2722	2D	107.87	3	39.8	0.018	New
TTTTF	SCM9	1BL.1RS	NA	1	41.5	1.16×10^{-100}	*Sr31*
	IWA6656	2B	254.3	–	–	0.015	New
TTKSK	IWA840	2B	192.19	1	20.6	3.51×10^{-13}	*Sr28*
	IWA1304	2B	165.61	2	34.2	2.01×10^{-4}	*Sr9h*
	IWA1202	Unlinked (2A)	NA	3	36.2	2.74×10^{-9}	*Sr48*
	IWA574	2A	103.39	4	39.6	0.0016	*Sr48*
BCCBC	IWA8599	2B	70.19	1	11.5	5.58×10^{-15}	Unknown
	IWA6625	6A	173.3	2	13.9	1.17×10^{-4}	*Sr13*
	IWA1708	2B	192.19	3	18.3	1.81×10^{-4}	*Sr28*
	SCM9	1BL.1RS	NA	4	24.5	2.76×10^{-4}	*Sr31*
	IWA6673	6A	172.45	–	–	0.026	*Sr13*
Race bulk	SCM9	1BL.1RS	NA	1	42.8	2.31×10^{-114}	*Sr31*
	IWA4897	1A	164.28	2	44.0	0.0047	New
COI-BLUE	SCM9	1BL.1RS	NA	1	26.5	5.93×10^{-60}	*Sr31*
	IWA3819	2A	85.52	2	16.5	3.42×10^{-14}	*Sr48*
	IWA1496	2A	85.52	3	33.9	3.47×10^{-9}	*Sr48*

† COI, coefficient of infection; BLUE, best linear unbiased estimates; NA, not applicable.

is limited. IWA1304 (Bonferroni $P = 2.01 \times 10^{-4}$) is positioned on the consensus map only 1.35 cM proximal to IWA5789 (166.96 cM, 2BL), which was tightly linked to *Sr9h* in the mapping population 'SD4279' × 'Brick' (Guerrero-Chavez et al., 2015). However, *Sr9h* has also been reported ~20 cM distal to IWA1304 in the population 'LMPG-6' × 'CItr 4311' (Babiker et al., 2016). Nonetheless, the race specificity observed in this panel suggests that IWA1304 is likely to be linked to *Sr9h*. The most significant association with TTKSK resistance, IWA840 (Bonf. $P = 3.51 \times 10^{-13}$), was mapped nearly 27 cM distal to IWA1304 and is likely to be linked to *Sr28*. IWA1708, located at the same position (192.19 cM) on the 9K wheat SNP consensus map, was also associated with resistance to *Pgt* race BCCBC (Bonferroni $P = 1.81 \times 10^{-4}$); this race specificity is characteristic of *Sr28* in that avirulence is observed when plants were challenged with TTKSK and BCCBC, whereas TRTTF and TTTTF are virulent (Rouse et al., 2012).

The third SNP marker detected for TTKSK resistance, IWA1202 (Bonferroni $P = 2.74 \times 10^{-9}$), is not anchored on the consensus map, probably as a result of being monomorphic in the biparental populations used for linkage map development. Interestingly, SNP marker IWA574 on the long arm of chromosome 2A was introduced as the fourth cofactor (Bonferroni $P = 0.0016$) and exhibited strong LD with IWA1202 ($r^2 = 0.90$), suggesting that IWA1202 is located on 2AL as well. Despite their strong linkage, both markers remained highly significant after being added to the same multiple regression model. Three resistance genes on 2AL (*Sr21*, *Sr48*, and *SrTm4*) are known to confer resistance to *Pgt* race TTKSK (Bansal et al., 2009; Briggs et al., 2015; Chen et al., 2015). Both *Sr21* and *SrTm4* were originally discovered in the diploid relative *Triticum monococcum* L. and are not expected to be widespread in hexaploid wheat (McIntosh et al., 1995), thus are unlikely to be linked to IWA1202 and IWA574 in this panel. *Sr48* is found in hexaploid wheat but was mapped to the distal end of 2AL (Bansal et al., 2009) and IWA574 is near the centromere. While further research is needed to determine if the IWA1202–IWA574 locus tags *Sr48* or a novel gene, the markers presented here provide a valuable starting point to investigate the identity of this locus.

The last single-race seedling screen evaluated here was for *Pgt* race BCCBC, the least virulent race in this study. After four steps of the MLMM, at least four resistance loci, tagged by the markers IWA8599, IWA6625, IWA1708, SCM9, and IWA6673, were significantly associated with BCCBC resistance. SCM9 and IWA1708 are postulated to be in linkage with *Sr31* and *Sr28*, respectively, as described above. Notably, IWA6625 and IWA6673 on the long arm of chromosome 6A exhibited a pattern similar to IWA1202 and IWA574: the first cofactor added to the regression from its respective locus did not account for the total phenotypic variation observed at the given QTL interval, resulting in a second significant marker–trait association only 0.85 cM away. Several hypotheses have been proposed to explain this incidence and will be considered later in the 'Discussion' section below. The influence of *Sr13* cannot be ruled out for this locus but its true identity would require allelism tests. Additionally, proper elucidation of the marker–trait association for IWA8599 (Bonferroni $P = 5.58 \times 10^{-15}$) was confounded by complexities in the LD patterns and the chromosomal structure of 2B.

Seedling and Field Response to Pgt *Race Mixture*

Similar to the seedling resistance loci detected for all races except TTKSK, *Sr31* was the most significant association for seedling resistance to the race mixture. Additionally, a putatively novel resistance locus linked to IWA4897 on the long arm of chromosome 1A (164.2 cM) was detected in the second step of the MLMM (Bonferroni $P = 0.0047$). This genomic region has been speculated to harbor a gene conferring adult plant resistance to Ug99 (Rouse et al., 2014), but marker–trait associations for this locus were nonsignificant in the field data, possibly indicating a difference in gene action or race specificity compared with the previously reported QTL.

Multi-locus mixed model analysis using COI-BLUE values from 3-yr field responses revealed three markers that explained 33.9% of the phenotypic variation. SCM9 was the most significant association, which documents the effectiveness of *Sr31* against North American races in field conditions. The next two markers added to the regression, IWA3819 (Bonferroni $P = 3.42 \times 10^{-14}$) and IWA1496 (Bonferroni $P = 3.47 \times 10^{-9}$), are tightly linked (LD $r^2 = 0.94$) and map to the centromeric region of 2AL. Although these markers are positioned nearly 18 cM proximal to the IWA1202–IWA574 resistance locus detected for *Pgt* race TTKSK, their strong pairwise LD ($r^2 > 0.68$) suggests they represent the same locus; furthermore, the resistant haplotype for IWA3819–IWA1496 (TT–CC) identifies 17 of the same resistant accessions represented by IWA1202–IWA574 (Supplemental Table S1). The effectiveness of this locus in 3 yr field trials and against the highly virulent race TTKSK makes it a strong candidate for follow-up studies and introgression into modern cultivars. Most accessions with low infection types and the corresponding resistant haplotype were collected from the southeastern European countries Croatia and Bulgaria. As described above, whether this resistance is conferred by *Sr48* (originally detected in the Swiss cultivar 'Arina') or a novel locus remains to be determined.

DISCUSSION

Genome-wide association studies are one of the most widely used tools for exploring the relationship between phenotypes and causal genetic variants in a target population. There is a broad range of applications for GWAS; hence, model implementation and subsequent interpretation of the results both depend heavily on the initial project aims. In this study, we pursued two primary goals: (i) characterize marker–trait relationships that have the potential to be applied in wheat breeding for stem rust resistance, and (ii) explore how GWAS can be more effectively used for trait characterization by evaluating single-locus versus multi-locus mapping approaches.

The MLMM method was used to reveal the SNP markers linked to *Pgt* resistance genes in a worldwide collection of hexaploid winter wheat. We identified 10 resistance loci associated with either qualitative race-specific seedling resistance or quantitative adult plant resistance. Notably, three newly documented loci, tagged by the SNP markers IWA2722 (for race TRTTF), IWA6656 (for race TTTTF), and IWA4897 (for NA race bulk), were associated with race-specific seedling resistance. In addition to efforts to discover new resistance loci, there are several known resistance genes that are not well characterized or the lack reliable genetic markers; genotyping a global diversity panel provided us with a unique opportunity to investigate the performance of SNP markers across extensive genetic backgrounds.

Utility of Significant Marker–Trait Associations in Broad Germplasm

Assessing genetic markers for their diagnostic ability in broad germplasm is necessary for efficient gene introgression. Among the genes present in our diversity panel, *Sr8a* is of interest because of its effectiveness against TRTTF, a highly virulent race that has been detected in Ethiopia, Yemen, and Pakistan (Mirza et al., 2010; Olivera et al., 2012). Additionally, its prevalence in modern germplasm yields an opportunity for verification of relevant marker–trait associations. IWA7913 was found to be linked with *Sr8a* in recent studies, including both association mapping and biparental linkage mapping studies (Bajgain et al., 2015; Guerrero-Chavez et al., 2015), and also displayed a significant association with TRTTF resistance in this global winter wheat panel. To evaluate its accuracy in detecting accessions potentially harboring *Sr8a*, we looked at the proportion of lines that exhibit its expected infection type (2+ or lower) among all lines homozygous for the predicted resistance allele 'G'. Disappointingly, only 74 of 298 accessions with the genotype 'GG' have an infection type of 2+ or lower, revealing an accuracy of 24.8%. Thus despite its tight linkage and strong association in select populations, its utility for MAS is limited in broad germplasm. Unfortunately, this is the fate of nearly all markers that are reported from both linkage and association mapping studies (Bernardo, 2016). This does not render mapping efforts useless, as gene discovery can be immediately beneficial for breeders regardless of marker availability, but it does highlight the need for marker verification.

Relying on single markers to detect underlying genes is limited by the extent of LD between the marker and the causal polymorphism. Dual marker haplotypes can be used to increase power in selecting for individual genes when single markers are not sufficient (Hagenblad, 2004; Lorenz et al., 2010). To investigate a candidate haplotype for *Sr8a*, we looked for the most significant marker–trait association surrounding the locus after IWA7913 was added as a cofactor. Although its association did not fall below our Bonferroni *P*-value threshold of 0.05, SNP marker IWA5781 (Bonferroni $P = 0.34$, positioned at 2.21 cM on 6A), in combination with IWA7913, revealed a haplotype with much higher diagnostic accuracy than IWA7913 alone. Among 35 accessions represented by the haplotype GG–AA (IWA7913–IWA5781), 30 expressed the expected *Sr8a* phenotype, improving the accuracy from 24.8 to 85.7% across the winter wheat diversity panel. This indicates that association mapping

with multi-locus models is a promising approach for breeders to expand the utility of GWAS.

Improving GWAS Analyses for Direct Breeding Application

The MLMM method applied, in combination with genotyping via the microsatellite marker SCM9 to detect 1AL.1RS and 1BL.1RS rye translocations, provided several valuable insights that are broadly relevant to the interpretation of GWAS analyses in hexaploid wheat. To recap, IWA435 can detect 1BL.1RS with nearly identical results to SCM9. The raw allele clustering patterns exhibited by IWA435 represent two orthologous *T. aestivum* alleles on chromosome 1BS, with an additional cluster for 1BL.1RS appearing as a null genotype (Fig. 4). Deletion mutations or, in this case, nonamplification of alien chromatin, warrant full consideration of null allele clusters; however, null marker data are statistically neutral in association mapping, resulting in potential variants being overlooked. Neither allele clustering algorithms nor GWAS models should be limited to biallelic assumptions, even in diploid organisms. The model-free density-based cluster identification algorithms DBSCAN (Ester et al., 1996) and OPTICS (Ankerst et al., 1999), which are used by the polyploid version of GenomeStudio, mark a positive step toward achieving optimum marker calls (Wang et al., 2014). However, even the most advanced clustering algorithms still require manual verification, which is a significant burden for researchers attempting to harvest data from SNP arrays with hundreds of thousands of markers.

Association mapping with models that use single-marker tests require LD-based confidence intervals to distinguish unique QTL positions. Depending on the proximity of the neighboring QTLs, the available genetic resolution may not be sufficient for delimiting unique loci. In contrast, MLMM analyses use stepwise regression to introduce significant markers as cofactors in each step of the model, thereby excluding collinear markers in strong LD with the same locus. But in some cases, a single marker added to the regression does not account for all of the phenotypic variation controlled by a given QTL interval, resulting in a second linked marker being incorporated in the model (Lipka et al., 2013; Segura et al., 2012). Several genomic regions, including SNP markers IWA1202–IWA574 and IWA3819–IWA1496 on chromosome 2A and IWA6625–IWA6673 on chromosome 6A, displayed this pattern. Several hypotheses have been proposed to describe this occurrence, such as the presence of multiple allelic variants (allelic heterogeneity) or the presence of several tightly linked genes that are indistinguishable at the available resolution. Based on observations from the current analyses, we suggest that a single causal gene may trigger this pattern as well. Without a marker diagnostic for *Sr31*, two SNP markers (IWA7385 and IWA2615) were fitted in the MLMM regression (Fig. 5); the resulting dual-marker haplotype accounted for 87.4% of the variation conferred by *Sr31* and was able to correctly identify 37 out of 42 lines carrying 1BL.1RS (confirmed with SCM9). In other words, a single imperfect marker failed to capture the total variance conferred by *Sr31* but the incorporation of a second linked marker improved detection accuracy. Extracting useful information on a genome-wide scale from multi-locus haplotypes will accelerate cultivar development through molecular breeding, especially when perfect markers are not available. Obtaining perfect markers is particularly difficult in hexaploid wheat, where genotyping complications are frequent. For example, chromosome designations in Fig. 5 are based on a consensus map derived from populations lacking 1BL.1RS, which led to mapping anomalies after alien chromatin was introduced to the panel. This is not an isolated incident, since most loci in hexaploid wheat have two homeologs with unpredictable annealing specificity to SNP probes. Here, we observed how multi-locus association mapping was able to overcome these limitations and facilitate more accurate gene detection.

Understanding the relationship between causal gene variants and nearby polymorphisms is essential because of the inherent difficulty in genotyping a tag within the target coding region itself. Association mapping is frequently used to mine large populations for rare alleles but comes with many disadvantages. When pre-designed SNP assays are applied to GWAS, ascertainment bias limits the probability that a diagnostic genotype of such rare alleles will be acquired (Frascaroli et al., 2013). Simultaneous marker development and direct application in target populations via genotyping-by-sequencing may enhance the likelihood that diagnostic polymorphisms are discovered. However, the availability of perfect markers is not the only limiting factor; considering *Sr31*, even if we had properly designated the null cluster for IWA435 as its own allele, the genotype would have been discarded before it was used for association mapping because of the imposed 5% MAF threshold. If a causal allele is present in a GWAS panel at a lower rate than the defined cutoff for MAF, detecting a perfect marker is impossible. This scenario is probably widespread in modern plant breeding programs since many genes are introgressed into limited genetic backgrounds initially, and it may take years for allele frequencies to cross the MAF thresholds. The markers that are most suitable for MAS are diagnostic for the favorable allele in diverse germplasm collections, but lower-frequency alleles diminish statistical power. Although interrogating low-frequency alleles lends vulnerability to Type I error, discarding these markers can cause important associations to be missed. Thus the ability to include rare SNPs during association mapping will accelerate the discovery of reliable genetic markers. As a potential solution, observing multi-locus haplotypes to characterize rare variants may be able to assist in verifying questionable associations from low-frequency alleles. Access to dense SNP profiles via high-throughput genotyping opens up an opportunity to evaluate and verify rare alleles using marker–trait associations in close proximity to the target gene.

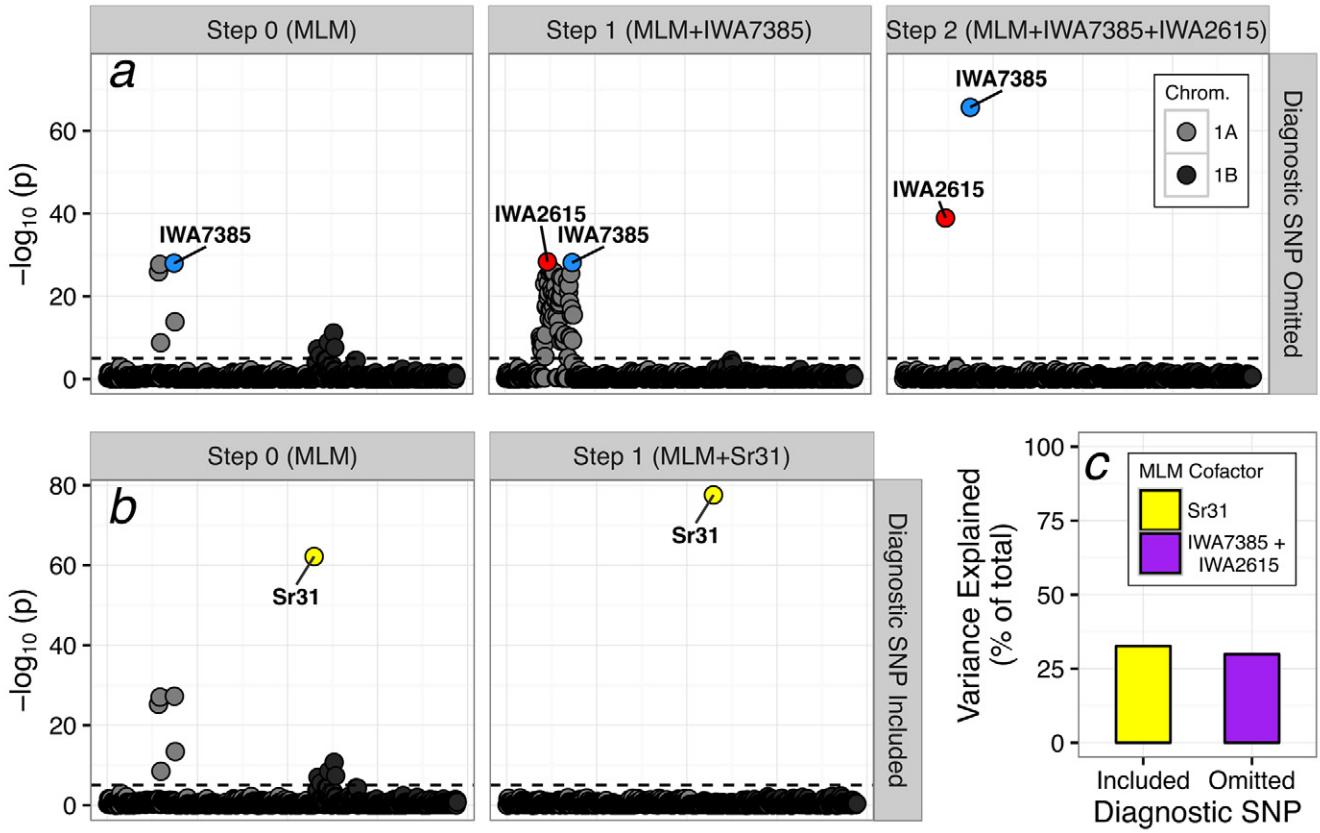

Fig. 5. The multi-locus mixed model applied to the *Pgt* race TRTTF wheat seedling resistance dataset with (a) the diagnostic SNP for *Sr31* omitted, and (b) the diagnostic SNP included; (c) the proportion of phenotypic variance explained by each of the resulting cofactors.

Conclusions

The wheat–*Puccinia* pathosystem represents perhaps one of the best characterized models for a gradient of qualitative, semiquantitative, and quantitative genetic effects. Our specific knowledge of resistance loci and the availability of some diagnostic markers presents a unique opportunity to investigate some of the challenges of GWAS analysis in polyploid species. Here we explored the power of multi-locus association mapping as a tool for characterizing stem rust resistance genes in a global population of winter wheat. We identified 10 resistance loci strongly associated with either adult plant resistance or race-specific seedling resistance; at least three of these are linked to previously uncharacterized resistance loci and warrant further characterization and selection in elite varieties. We also revealed that SNP marker IWA435 reliably detects the 1BL.1RS rye translocation, agreeing with the SCM9 results in 99.8% (1402 out of 1405) of lines tested. And lastly, we show that multi-locus association mapping can be used to identify marker combinations whose haplotypes more accurately explain the phenotypic variation conferred by individual loci than single-locus models. The 17-Gb genome of hexaploid wheat with genomic resources dwarfed by other crops such as maize (*Zea mays* L.), rice (*Oryza sativa* L.), and soybean [*Glycine max* (L.) Merr.] makes this complex issue difficult to fully address with the dataset at hand; nonetheless, sequencing and bioinformatics technologies are improving rapidly and methods in association mapping and genomic prediction stand to benefit. Obtaining application-ready diagnostic genetic markers for novel traits is normally highly resource-intensive and thus researchers at all stages of model development, genomics, and applied breeding must work together to overcome the limitations of linkage and association mapping.

Supplemental Information

Supplementary Table S1 contains a list of all 1411 accessions used in this study, complete with phenotype information for all six traits and genotype data for all markers described in the results and discussion.

Conflict of Interest Disclosure

The authors declare no conflicts of interest.

Acknowledgments

This study is part of the Triticeae Coordinated Agriculture Project (www.triticeaecap.org, accessed 8 May 2017), funded by the USDA National Institute of Food and Agriculture grant 2011-68002-30029 and 2017-67007-25939. We thank Liangliang Gao (Kansas State University) for his coding expertise and assistance with consensus map revisions. We thank Shiaoman Chao (USDA-ARS, Fargo, ND) for genotyping the National Small Grains Collection Core germplasm panels and we thank Sheri Rynearson (Washington State University) for her technical assistance with supplemental marker genotyping. And lastly, we thank Amy Fox for her careful oversight of field and greenhouse experiments.

References

Ankerst, M., M.M. Breunig, H.-P. Kriegel, and J. Sander. 1999. OPTICS: Ordering points to identify the clustering structure. SIGMOD Rec. 28:49–60. doi:10.1145/304181.304187

Aoun, M., M. Breiland, K.M. Turner, A. Loladze, S. Chao, S.S. Xu, et al. 2016. Genome-wide association mapping of leaf rust response in a durum wheat worldwide germplasm collection. Plant Genome 9. doi:10.3835/plantgenome2016.01.0008

Babiker, E.M., T.C. Gordon, S. Chao, M.N. Rouse, R. Wanyera, M. Newcomb, et al. 2016. Genetic mapping of resistance to the Ug99 race group of *Puccinia graminis* f. sp. *tritici* in a spring wheat landrace CItr 4311. Theor. Appl. Genet. 129:2161–2170. doi:10.1007/s00122-016-2764-5

Bajgain, P., M. Rouse, P. Bulli, S. Bhavani, T. Gordon, R. Wanyera, et al. 2015. Association mapping of North American spring wheat breeding germplasm reveals loci conferring resistance to Ug99 and other African stem rust races. BMC Plant Biol. 15. doi:10.1186/s12870-015-0628-9

Bansal, U.K., M.J. Hayden, B. Keller, C.R. Wellings, R.F. Park, and H.S. Bariana. 2009. Relationship between wheat rust resistance genes *Yr1* and *Sr48* and a microsatellite marker. Plant Pathol. 58:1039–1043. doi:10.1111/j.1365-3059.2009.02144.x

Bernardo, R. 2016. Bandwagons I, too, have known. Theor. Appl. Genet. 129:2323–2332. doi:10.1007/s00122-016-2772-5

Breseghello, F., and M.E. Sorrells. 2006. Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. Genetics 172:1165–1177. doi:10.1534/genetics.105.044586

Briggs, J., S. Chen, W. Zhang, S. Nelson, J. Dubcovsky, and M.N. Rouse. 2015. Genetic mapping of *SrTm4*, a recessive stem rust resistance gene from diploid wheat effective to Ug99. Phytopathology 105:1347–1354. doi:10.1094/PHYTO-12-14-0382-R

Bulli, P., J. Zhang, S. Chao, X. Chen, and M. Pumphrey. 2016. Genetic architecture of resistance to stripe rust in a global winter wheat germplasm collection. G3 (Bethesda) 6(8): 2237–2253.. doi:10.1534/g3.116.028407

Cavanagh, C.R., S. Chao, S. Wang, B.E. Huang, S. Stephen, S. Kiani, et al. 2013. Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. Proc. Natl. Acad. Sci. USA 110:8057–8062. doi:10.1073/pnas.1217133110

Chang, H.-X., P.J. Brown, A.E. Lipka, L.L. Domier, and G.L. Hartman. 2016. Genome-wide association and genomic prediction identifies associated loci and predicts the sensitivity of Tobacco ringspot virus in soybean plant introductions. BMC Genomics 17. 153. doi:10.1186/s12864-016-2487-7

Chao, S., W. Zhang, J. Dubcovsky, and M. Sorrells. 2007. Evaluation of genetic diversity and genome-wide linkage disequilibrium among U.S. wheat (*Triticum aestivum* L.) germplasm representing different market classes. Crop Sci. 47:1018–1030. doi:10.2135/cropsci2006.06.0434

Chen, J., and Z. Chen. 2008. Extended Bayesian information criteria for model selection with large model spaces. Biometrika 95:759–771. doi:10.1093/biomet/asn034

Chen, S., M.N. Rouse, W. Zhang, Y. Jin, E. Akhunov, Y. Wei, et al. 2015. Fine mapping and characterization of *Sr21*, a temperature-sensitive diploid wheat resistance gene effective against the *Puccinia graminis* f. sp. *tritici* Ug99 race group. Theor. Appl. Genet. 128:645–656. doi:10.1007/s00122-015-2460-x

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

Ester, M., H.-P. Kriegel, J. Sander, and X. Xu. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. KDD 96:226–231.

Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Mol. Ecol. 14:2611–2620. doi:10.1111/j.1365-294X.2005.02553.x

Frascaroli, E., T.A. Schrag, and A.E. Melchinger. 2013. Genetic diversity analysis of elite European maize (*Zea mays* L.) inbred lines using AFLP, SSR, and SNP markers reveals ascertainment bias for a subset of SNPs. Theor. Appl. Genet. 126:133–141. doi:10.1007/s00122-012-1968-6

Gao, L., M.K. Turner, S. 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:e0148671. doi:10.1371/journal.pone.0148671

Ghazvini, H., C.W. Hiebert, J.B. Thomas, and T. Fetch. 2013. Development of a multiple bulked segregant analysis (MBSA) method used to locate a new stem rust resistance gene (*Sr54*) in the winter wheat cultivar Norin 40. Theor. Appl. Genet. 126:443–449. doi:10.1007/s00122-012-1992-6

Guerrero-Chavez, R., K.D. Glover, M.N. Rouse, and J.L. Gonzalez-Hernandez. 2015. Mapping of two loci conferring resistance to wheat stem rust pathogen races TTKSK (Ug99) and TRTTF in the elite hard red spring wheat line SD4279. Mol. Breed. 35:8. doi:10.1007/s11032-015-0198-4

Hagenblad, J. 2004. Haplotype structure and phenotypic associations in the chromosomal regions surrounding two *Arabidopsis thaliana* flowering time loci. Genetics 168:1627–1638. doi:10.1534/genetics.104.029470

Jighly, A., B.C. Oyiga, F. Makdis, K. Nazari, O. Youssef, W. Tadesse, et al. 2015. Genome-wide DArT and SNP scan for QTL associated with resistance to stripe rust (*Puccinia striiformis* f. sp. *tritici*) in elite ICARDA wheat (*Triticum aestivum* L.) germplasm. Theor. Appl. Genet. 128:1277–1295. doi:10.1007/s00122-015-2504-2

Jin, Y. 2005. Races of *Puccinia graminis* identified in the United States during 2003. Plant Dis. 89:1125–1127. doi:10.1094/PD-89-1125

Jin, Y., L.J. Szabo, Z.A. Pretorius, R.P. Singh, R. Ward, and T. Fetch, Jr. 2008. Detection of virulence to resistance gene *Sr24* within race TTKS of *Puccinia graminis* f. sp. *tritici*. Plant Dis. 92:923–926. doi:10.1094/PDIS-92-6-0923

Kumar, S., D.J. Garrick, M.C. Bink, C. Whitworth, D. Chagné, and R.K. Volz. 2013. Novel genomic approaches unravel genetic architecture of complex traits in apple. BMC Genomics 14:393. doi:10.1186/1471-2164-14-393

Letta, T., P. Olivera, M. Maccaferri, Y. Jin, K. Ammar, A. Badebo, et al. 2014. Association mapping reveals novel stem rust resistance loci in durum wheat at the seedling stage. Plant Genome 7. doi:10.3835/plantgenome2013.08.0026

Li, C., G. Bai, S. Chao, and Z. Wang. 2015. A high-density SNP and SSR consensus map reveals segregation distortion regions in wheat. BioMed Res. Int. 2015:830618. doi:10.1155/2015/830618

Lipka, A.E., M.A. Gore, M. Magallanes-Lundback, A. Mesberg, H. Lin, T. Tiedem, et al. 2013. Genome-wide association study and pathway-level analysis of tocochromanol levels in maize grain. G3 (Bethesda) 3:1287–1299. doi:10.1534/g3.113.006148

Lipka, A.E., F. Tian, Q. 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

Liu, X., M. Huang, B. Fan, E.S. Buckler, and Z. Zhang. 2016. Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. PLoS Genet. 12:e1005767. doi:10.1371/journal.pgen.1005767

Lorenz, A.J., M.T. Hamblin, and J.-L. Jannink. 2010. Performance of single nucleotide polymorphisms versus haplotypes for genome-wide association analysis in barley. PLoS ONE 5:e14079. doi:10.1371/journal.pone.0014079

Maccaferri, M., J. Zhang, P. Bulli, Z. Abate, S. Chao, D. Cantu, et al. 2015. A genome-wide association study of resistance to stripe rust (Puccinia *striiformis* f. sp. tritici) in a worldwide collection of hexaploid spring wheat (Triticum aestivum L.). G3 (Bethesda) 5:449–465. doi:10.1534/g3.114.014563

McIntosh, R.A., J. Dubcovsky, W.J. Rogers, C. Morris, R. Appels, and X.C. Xia. 2014. Catalogue of gene symbols for wheat: 2013–14 supplement. Ann. Wheat News. 60:153–175.

McIntosh, R.A., C.R. Wellings, and R.F. Park. 1995. Wheat rusts: An atlas of resistance genes. CSIRO Publishing, Canberra, ACT.

Mirza, J.I., A. Rattu, K.A. Khanzada, I. Ahmad, and T. Fetch. 2010. Race analysis of stem rust isolates collected from Pakistan in 2008-09. In: R. McIntosh and Z. Pretorius, editors, Proceedings of BGRI 2010 Technical Workshop, St Petersburg, Russia. 30–31 May 2010. Borlaug Global Rust Initiative, Ithaca, NY. p. 30–31.

Newcomb, M., P.D. Olivera, M.N. Rouse, L.J. Szabo, J. Johnson, S. Gale, et al. 2016. Kenyan isolates of *Puccinia graminis* f. sp. *tritici* from 2008 to 2014: Virulence to *SrTmp* in the Ug99 race group and implications for breeding programs. Phytopathology 106:729–736. doi:10.1094/PHYTO-12-15-0337-R

Oliver, R.P. 2014. A reassessment of the risk of rust fungi developing resistance to fungicides: Rust fungicide resistance risk. Pest Manag. Sci. 70:1641–1645. doi:10.1002/ps.3767

Olivera, P.D., Y. Jin, M. Rouse, A. Badebo, T. Fetch, Jr., R.P. Singh, et al. 2012. Races of *Puccinia graminis* f. sp. *tritici* with combined virulence to *Sr13* and *Sr9e* in a field stem rust screening nursery in Ethiopia. Plant Dis. 96:623–628. doi:10.1094/PDIS-09-11-0793

Pardey, P.G., J.M. Beddow, D.J. Kriticos, T.M. Hurley, R.F. Park, E. Duveiller, et al. 2013. Right-sizing stem-rust research. Science 340:147–148. doi:10.1126/science.122970

Patterson, N., A.L. Price, and D. Reich. 2006. Population structure and eigenanalysis. PLoS Genet. 2:e190. doi:10.1371/journal.pgen.0020190

Peterson, R.S., A.B. Campbell, and A.E. Hannah. 1948. A diagrammatic scale for estimating rust intensity on leaves and stems of cereals. Can. J. Res. 26:496–500. doi:10.1139/cjr48c-033

Pfeifer, M., K.G. Kugler, S.R. Sandve, B. Zhan, H. Rudi, T.R. Hvidsten, et al. 2014. Genome interplay in the grain transcriptome of hexaploid bread wheat. Science 345:1250091. doi:10.1126/science.1250091

Pretorius, Z.A., R.P. Singh, W.W. Wagoire, and T.S. Payne. 2000. Detection of virulence to wheat stem rust resistance gene *Sr31* in *Puccinia graminis*. f. sp. *tritici* in Uganda. Plant Dis. 84:203. doi:10.1094/PDIS.2000.84.2.203B

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. R Foundation for Statistical Computing. www.R-project.org/ (accessed 9 May 2017).

Rakitsch, B., C. Lippert, O. Stegle, and K. Borgwardt. 2013. A Lasso multi-marker mixed model for association mapping with population structure correction. Bioinformatics 29:206–214. doi:10.1093/bioinformatics/bts669

Rasheed, A., X. Xia, F. Ogbonnaya, T. Mahmood, Z. Zhang, A. Mujeeb-Kazi, et al. 2014. Genome-wide association for grain morphology in synthetic hexaploid wheats using digital imaging analysis. BMC Plant Biol. 14. doi:10.1186/1471-2229-14-128

Reif, J.C., M. Gowda, H.P. Maurer, C.F.H. Longin, V. Korzun, E. Ebmeyer, et al. 2011. Association mapping for quality traits in soft winter wheat. Theor. Appl. Genet. 122:961–970. doi:10.1007/s00122-010-1502-7

Roelfs, A.P., R.P. Singh, and E.E. Saari. 1992. Rust diseases of wheat: Concepts and methods of disease management. CIMMYT. Mexico DF, Mexico.

Rouse, M.N., I.C. Nava, S. Chao, J.A. Anderson, and Y. Jin. 2012. Identification of markers linked to the race Ug99 effective stem rust resistance gene *Sr28* in wheat (*Triticum aestivum* L.). Theor. Appl. Genet. 125:877–885. doi:10.1007/s00122-012-1879-6

Rouse, M.N., L.E. Talbert, D. Singh, and J.D. Sherman. 2014. Complementary epistasis involving *Sr12* explains adult plant resistance to stem rust in Thatcher wheat (*Triticum aestivum* L.). Theor. Appl. Genet. 127:1549–1559. doi:10.1007/s00122-014-2319-6

Rouse, M.N., R. Wanyera, P. Njau, and Y. Jin. 2011. Sources of resistance to stem rust race Ug99 in spring wheat germplasm. Plant Dis. 95:762–766. doi:10.1094/PDIS-12-10-0940

Sauvage, C., V. Segura, G. Bauchet, R. Stevens, P.T. Do, Z. Nikoloski, et al. 2014. Genome-wide association in tomato reveals 44 candidate loci for fruit metabolic traits. Plant Physiol. 165:1120–1132. doi:10.1104/pp.114.241521

Segura, V., B.J. Vilhjálmsson, A. Platt, A. Korte, Ü. Seren, Q. Long, et al. 2012. An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. Nat. Genet. 44:825–830. doi:10.1038/ng.2314

Singh, R.P., D.P. Hodson, J. Huerta-Espino, Y. Jin, S. Bhavani, P. Njau, et al. 2011. The emergence of Ug99 races of the stem rust fungus is a threat to world wheat production. Annu. Rev. Phytopathol. 49:465–481. doi:10.1146/annurev-phyto-072910-095423

Singh, R.P., P.K. Singh, J. Rutkoski, D.P. Hodson, X. He, L.N. Jørgenssen, et al. 2016. Disease impact on wheat yield potential and prospects of genetic control. Annu. Rev. Phytopathol. 54:1–20. doi:10.1146/annurev-phyto-080615-095835

Stakman, E.C., D.M. Stewart, and W.Q. Loegering. 1962. Identification of physiologic races of Puccinia graminis var. tritici. USDA-ARS E617. USDA, Washington, DC.

Stokstad, E. 2007. Deadly wheat fungus threatens world's breadbaskets. Science 315:1786–1787. doi:10.1126/science.315.5820.1786

Tsilo, T.J., S. Chao, Y. Jin, and J.A. Anderson. 2009. Identification and validation of SSR markers linked to the stem rust resistance gene *Sr6* on the short arm of chromosome 2D in wheat. Theor. Appl. Genet. 118:515–524. doi:10.1007/s00122-008-0917-x

VanRaden, P.M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91:4414–4423. doi:10.3168/jds.2007-0980

Vaughn, J.N., R.L. Nelson, Q. Song, P.B. Cregan, and Z. Li. 2014. The genetic architecture of seed composition in soybean is refined by genome-wide association scans across multiple populations. G3 (Bethesda) 4:2283–2294. doi:10.1534/g3.114.013433

Wang, S., D. Wong, K. Forrest, A. Allen, S. Chao, B.E. Huang, et al. 2014. Characterization of polyploid wheat genomic diversity using a high-density 90,000 single nucleotide polymorphism array. Plant Biotechnol. J. 12:787–796. doi:10.1111/pbi.12183

Weng, Y., P. Azhaguvel, R.N. Devkota, and J.C. Rudd. 2007. PCR-based markers for detection of different sources of 1AL.1RS and 1BL.1RS wheat–rye translocations in wheat background. Plant Breed. 126:482–486. doi:10.1111/j.1439-0523.2007.01331.x

Würschum, T., and T. Kraft. 2015. Evaluation of multi-locus models for genome-wide association studies: A case study in sugar beet. Heredity 114:281–290. doi:10.1038/hdy.2014.98

Würschum, T., S.M. Langer, C.F.H. Longin, V. Korzun, E. Akhunov, E. Ebmeyer, et al. 2013. Population structure, genetic diversity and linkage disequilibrium in elite winter wheat assessed with SNP and SSR markers. Theor. Appl. Genet. 126:1477–1486. doi:10.1007/s00122-013-2065-1

Yu, L.-X., H. Barbier, M.N. Rouse, S. Singh, R.P. Singh, S. Bhavani, et al. 2014. A consensus map for Ug99 stem rust resistance loci in wheat. Theor. Appl. Genet. 127:1561–1581. doi:10.1007/s00122-014-2326-7

Yu, L.-X., A. Lorenz, J. Rutkoski, R.P. Singh, S. Bhavani, J. Huerta-Espino, et al. 2011. Association mapping and gene–gene interaction for stem rust resistance in CIMMYT spring wheat germplasm. Theor. Appl. Genet. 123:1257–1268. doi:10.1007/s00122-011-1664-y

Yu, J., 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

Yu, G., Q. Zhang, T.L. Friesen, M.N. Rouse, Y. Jin, S. Zhong, et al. 2015. Identification and mapping of *Sr46* from *Aegilops tauschii* accession CIae 25 conferring resistance to race TTKSK (Ug99) of wheat stem rust pathogen. Theor. Appl. Genet. 128:431–443. doi:10.1007/s00122-014-2442-4

Zadoks, J.C., T.T. Chang, and C.F. Konzak. 1974. A decimal code for the growth stages of cereals. Weed Res. 14:415–421. doi:10.1111/j.1365-3180.1974.tb01084.x

Zanke, C.D., J. Ling, J. Plieske, S. Kollers, E. Ebmeyer, V. Korzun, et al. 2014. Whole genome association mapping of plant height in winter wheat (*Triticum aestivum* L.). PLoS ONE 9:e113287. doi:10.1371/journal.pone.0113287

Zhang, D., G. Bai, C. Zhu, J. Yu, and B.F. Carver. 2010a. Genetic diversity, population structure, and linkage disequilibrium in U.S. elite winter wheat. Plant Genome 3:117–127. doi:10.3835/plantgenome2010.03.0004

Zhang, D., R.L. Bowden, J. Yu, B.F. Carver, and G. Bai. 2014. Association analysis of stem rust resistance in US winter wheat. PLoS ONE 9:e103747. doi:10.1371/journal.pone.0103747

Zhang, Z., E. Ersoz, C.-Q. Lai, R.J. Todhunter, H.K. Tiwari, M.A. Gore, et al. 2010b. Mixed linear model approach adapted for genome-wide association studies. Nat. Genet. 42:355–360. doi:10.1038/ng.546

Zhao, K., M.J. Aranzana, S. Kim, C. Lister, C. Shindo, C. Tang, et al. 2007. An *Arabidopsis* example of association mapping in structured samples. PLoS Genet. 3:e4. doi:10.1371/journal.pgen.0030004

Zurn, J.D., M. Newcomb, M.N. Rouse, Y. Jin, S. Chao, J. Sthapit, et al. 2014. High-density mapping of a resistance gene to Ug99 from the Iranian landrace PI 626573. Mol. Breed. 34:871–881. doi:10.1007/s11032-014-0081-8